

METHODS FOR THE ANALYSIS OF CIRCULATING MICROPARTICLES

TECHNICAL FIELD

5 The present invention relates to the analysis of cell free biomolecules (e.g. cell free nucleic acid molecules and cell free polypeptides). In particular, it relates to the analysis of cell free biomolecules contained within or derived from circulating microparticles. Provided are reagents and methods for analysing biomolecules of circulating microparticles including reagents and methods for analysing biomolecules of single circulating microparticles.

10 BACKGROUND

Cell-free DNA (cfDNA) in the circulation is typically fragmented (typically in the range of 100-200 base pairs in length), and thus methods for cfDNA analysis have traditionally focused upon biological signals that can be found with these short DNA fragments. For example, detecting single-nucleotide variants within individual molecules, or performing 'molecular counting' across a large number of sequenced fragments to indirectly infer the presence of large-scale chromosomal abnormalities e.g. tests for foetal chromosomal trisomies that assess foetal DNA within the maternal circulation (a form of so-called 'non-invasive prenatal testing', or NIPT).

A large variety of methods to analyse circulating cell-free DNA have been described previously. Depending upon the specific application area, these assays may employ different terminology for a broadly similar set of sample types and technical methods, such as circulating tumour DNA (ctDNA), cell-free foetal DNA (cffDNA), and/or liquid biopsy, or non-invasive prenatal testing. In general, these methods comprise a laboratory protocol to prepare samples of circulating cell-free DNA for sequencing, a sequencing reaction itself, and then an informatic framework to analyse the resulting sequences to detect a relevant biologic signal. The methods involve a DNA purification and isolation step prior to sequencing, which means that the subsequent analysis must rely solely on the information contained in the DNA itself. Following sequencing, such methods generally employ one or more informatic or statistical frameworks to analyse various aspects of the sequence data, such as detecting specific mutations therein, and/or detecting selective enrichment or selective depletion of particular chromosomes or sub-chromosomal regions (for example, which might be indicative of a chromosomal aneuploidy in a developing foetus).

Many of these methods are for use in NIPT (e.g. in US patents 6258540 B1, 8296076 B2, 8318430 B2, 8195415 B2, 9447453 B2, and 8442774 B2). The most common methods for performing non-invasive prenatal testing for the detection of foetal chromosomal abnormalities (such as trisomies, and/or sub-chromosomal abnormalities such as microdeletions) involve sequencing a large number of molecules of cfDNA, mapping the resulting sequences to the genome (i.e. to determine which chromosome and/or which part of a given chromosome the

- 2 -

sequence derive from), and then, for one or more such chromosomal or sub-chromosomal regions, determining the amount of sequence that maps thereto (e.g. in the form of absolute numbers of reads or relative numbers of reads) and then comparing this to one or more normal or abnormal threshold or cutoff values, and/or performing a statistical test, to determine whether said region(s) may be overrepresented in amount of sequence (which may, for example, correspond to a chromosomal trisomy) and/or whether said region(s) may be underrepresented in amount of sequence (which may, for example, correspond to a microdeletion).

A variety of additional or modified approaches to analyzing cell free DNA using data from unlinked, individual molecules have also been described (e.g. WO2016094853 A1, US2015344970 A1 and US20150105267 A1).

Despite the existence of such a wide range of methods, there remains a need for new methods of analysing cfDNA that would allow the reliable detection of long-range genetic information (e.g. phasing) and also for methods with greater sensitivity. For example, in the case of NIPT, foetal cfDNA only represents a minor fraction of the overall cfDNA in pregnant individuals (the majority of circulating DNA being normal maternal DNA). Therefore, a considerable technical challenge for NIPT revolves around differentiating foetal cfDNA from maternal DNA. Similarly, in a patient with cancer, cfDNA only represents a tiny fraction of the overall circulating DNA. Therefore, a similar technical challenge exists in relation to the use of cfDNA analysis for the diagnosis or monitoring of cancer.

Separately, methods have also been described that allow the isolation of cell type-specific apoptotic bodies by fluorescence-activated cell sorting (FACS) (Atkin-Smith et al., 2017. Scientific Reports 7, 39846) and that allow the multiplexed profiling of protein markers in single extracellular vesicles (Lee et al., 2018. ACS Nano. 23, 12(1), 494-503).

DESCRIPTION

The invention provides methods for the analysis of samples comprising circulating microparticles (or samples derived from circulating microparticles) such as apoptotic bodies. The invention is based on multi-parametric measurement of different types of biomolecules comprised within or derived from single circulating microparticles. In particular, the invention allows the measurement of linked signals corresponding to the presence, absence and/or level of two or more types of target biomolecule in the same circulating microparticle. As illustrated in Figure 30, signals corresponding to the levels of fragments of genomic DNA may be produced (e.g. by partitioning, barcoding and sequencing) and a signal corresponding to the level of a target polypeptide may be produced (e.g. using a barcoded affinity probe). In addition, a signal corresponding to the level of a modified nucleotide (e.g. a nucleotide comprising 5-methylcytosine) may be produced (e.g. by an affinity-based enrichment approach such as one that uses an enrichment probe that is specific

for or preferentially binds 5-methylcytosine in fragments of genomic DNA). These measurements and associated techniques thus produce a series of linked signals corresponding to the physical and biological state of a circulating microparticle.

- 5 The multi-parametric methods provided herein adds additional layers of information to the the earlier inventions provided by the inventor in PCT/GB2017/053820, PCT/GB2017/053812, and PCT/GB2017/053816.

10 In PCT/GB2017/053820, the inventor previously provided methods for the analysis of nucleic acid fragments in circulating microparticles (or microparticles originating from blood). That invention is based on a linked-fragment approach in which fragments of nucleic acid from a single microparticle are linked together. This linkage enables the production of a set of linked sequence reads (i.e. set of linked signals) corresponding to the sequences of fragments from a single microparticle.

15 The linked-fragment approach provides highly sensitive cfDNA analysis and also enables the detection of long-range genetic information. The approach is based on a combination of insights. Firstly, the methods take advantage of the insight that individual circulating microparticles (for example, an individual circulating apoptotic body) will contain a number of fragments of genomic DNA that have been generated from the same individual cell (somewhere in the body) which has undergone apoptosis. Secondly, a fraction of such fragments of genomic DNA within an individual microparticle will preferentially comprise sequences from one or more specific chromosomal regions. Cumulatively, such circulating microparticles thus serve as a data-rich and multi-feature 'molecular stethoscope' to observe what may be quite complex genetic events occurring in a limited somatic tissue space somewhere in the body; importantly, since such microparticles in large part enter the circulation prior to clearance or metabolism, they may be detected noninvasively. The invention describes experimental and informatic methods of using these 'stethoscopes' i.e. sets of linked fragments and linked sequence reads (either in the form of single, individual microparticles, or, in many embodiments, complex samples comprising a large number of single circulating microparticles) to perform analytic and diagnostic tasks.

30 The present invention advances the concept of the 'molecular stethoscope' by harnessing the data provided by the co-localisation of, for example, non-nucleic acid molecules (e.g. target polypeptides) with nucleic acid molecules (e.g. fragments of genomic DNA) in single circulating microparticles. This advance is based on the discovery that rather than being singular and freely diffusible in the blood, many biomolecules (e.g. nucleic acid molecules and polypeptides) comprised within the circulation are biophysically retained within circulating microparticles. The invention exploits this rich source of information by measuring signals corresponding to the presence, absence and/or level of a plurality of target biomolecules of a circulating microparticle

- 4 -

to produce a set of (informatically) linked signals for the circulating microparticle. In addition, by including in this set one or more signals corresponding to one or more target biomolecules that is/are characteristic of a particular cell or tissue type, the cellular origin of a particular set of linked signals, derived from a single circulating microparticle, can be determined. This provides the set of linked signals with a 'cellular context' providing a much richer source of information than currently available methods. In so doing, the invention provides methods of analysis with high accuracy, sensitivity, and precision. Such methods have clear applications in a wide range of diagnostic and monitoring applications including cancer diagnosis and monitoring, and NIPT.

The inventor has previously provided reagents and methods related to barcoding. In WO2016/207639, the inventor provided a wide range of reagents, kits and methods for molecular barcoding including multimeric barcoding reagents. In PCT/GB2017/053812, the inventor provided further methods and reagents for molecular barcoding. In PCT/GB2017/053816, the inventor provided reagents and methods for molecular barcoding of nucleic acids of single cells.

The entire content of WO2016/207639, PCT/GB2017/053812, PCT/GB2017/053816 and PCT/GB2017/053820 is incorporated herein by reference.

The invention provides a method of analysing a sample comprising a circulating microparticle or a sample derived from a circulating microparticle, wherein the circulating microparticle comprises at least two target molecules, wherein the at least two target molecules are biomolecules, and wherein the method comprises measuring a signal corresponding to the presence, absence and/or level of each of the target molecules to produce a set of at least two (informatically) linked signals for the circulating microparticle, wherein at least one of the linked signals corresponds to the presence, absence and/or level of a first biomolecule in the sample and at least one of the linked signals corresponds to the presence, absence and/or level of a second biomolecule in the sample.

The invention provides a method of analysing a sample comprising a circulating microparticle or a sample derived from a circulating microparticle, wherein the circulating microparticle comprises at least two target molecules, wherein the at least two target molecules are biomolecules, and wherein the method comprises measuring a signal corresponding to the presence, absence and/or level of each of the target molecules to produce a single signal for the circulating microparticle, wherein the single signal corresponds to the presence, absence and/or level of the biomolecules in the sample.

The first biomolecule may be a fragment of a target nucleic acid (e.g. a fragment of genomic DNA) and second biomolecule may be a target (or predefined) non-nucleic acid biomolecule (e.g.

– 5 –

a target polypeptide). Optionally, the fragment of a target nucleic acid may comprise at least one modified nucleotide or nucleobase.

5 The target molecules may comprise at least one or, preferably, at least two fragments of of a target nucleic acid (e.g. genomic DNA).

10 The first biomolecule may be a polypeptide and the second target biomolecule may be a fragment of a target nucleic acid (e.g. genomic DNA) comprising an epigenetic modification (e.g. 5-hydroxy-methylcytosine DNA or 5-methylcytosine DNA).

The first biomolecule may be 5-hydroxy-methylcytosine DNA and the second target biomolecule may be a fragment of RNA.

15 The first biomolecule may be 5-methylcytosine DNA and the second target biomolecule may be a fragment of RNA.

The first biomolecule may be 5-hydroxy-methylcytosine DNA and the second target biomolecule may be a biomolecule selected from Biomolecule group 1.

20 The first biomolecule may be 5-methylcytosine DNA and the second target biomolecule may be a biomolecule selected from Biomolecule group 1.

The first and second biomolecules may be selected from Biomolecule group 1.

25 The invention provides a method of analysing a sample comprising a circulating microparticle or a sample derived from a circulating microparticle, wherein the circulating microparticle comprises at least three target molecules, wherein at least two of the target molecules are fragments of genomic DNA and at least one of the target molecules is a fragment of RNA, and wherein the method comprises measuring a signal corresponding to the presence, absence and/or level of
30 each of the target molecules to produce a set of at least two (informatically) linked signals for the circulating microparticle, wherein at least one of the linked signals corresponds to the presence, absence and/or level of the fragments of genomic DNA in the sample and at least one of the linked signals corresponds to the presence, absence and/or level of the fragment of RNA in the sample.

35 The invention provides a method of analysing a sample comprising a circulating microparticle or a sample derived from a circulating microparticle, wherein the circulating microparticle comprises at least three target molecules, wherein at least two of the target molecules are fragments of genomic DNA and at least one of the target molecules is a fragment of RNA, and wherein the

method comprises measuring a signal corresponding to the presence, absence and/or level of each of the target molecules to produce a single signal for the circulating microparticle, wherein the single signal corresponds to the presence, absence and/or level of the fragments of genomic DNA and the fragment of RNA in the sample.

5

The invention provides a method of analysing a sample comprising a circulating microparticle or a sample derived from a circulating microparticle, wherein the circulating microparticle comprises at least three target molecules, wherein at least two of the target molecules are fragments of a target nucleic acid (e.g. genomic DNA) and at least one of the target molecules is a target biomolecule (e.g. a target polypeptide), and wherein the method comprises measuring a signal corresponding to the presence, absence and/or level of each of the target molecules to produce a set of at least three (informatically) linked signals for the circulating microparticle, wherein each of at least two of the linked signals corresponds to the presence, absence and/or level of one of the fragments of the target nucleic acid (e.g. genomic DNA) in the sample and at least one of the linked signals corresponds to the presence, absence and/or level of the target biomolecule (e.g. the target polypeptide) in the sample.

The invention provides a method of analysing a sample comprising a circulating microparticle or a sample derived from a circulating microparticle, wherein the circulating microparticle comprises at least three target molecules, wherein at least two of the target molecules are fragments of a target nucleic acid (e.g. genomic DNA) and at least one of the target molecules is a target biomolecule (e.g. a target polypeptide), and wherein the method comprises measuring a signal corresponding to the presence, absence and/or level of each of the target molecules to produce a set of at least two (informatically) linked signals for the circulating microparticle, wherein at least one of the linked signals corresponds to the presence, absence and/or level of the fragments of the target nucleic acid (e.g. genomic DNA) in the sample and at least one of the linked signals corresponds to the presence, absence and/or level of the target biomolecule (e.g. the target polypeptide) in the sample.

The invention provides a method of analysing a sample comprising a circulating microparticle or a sample derived from a circulating microparticle, wherein the circulating microparticle comprises at least three target molecules, wherein at least two of the target molecules are fragments of a target nucleic acid (e.g. genomic DNA) and at least one of the target molecules is a target biomolecule (e.g. a target polypeptide), and wherein the method comprises measuring a signal corresponding to the presence, absence and/or level of each of the target molecules to produce a single signal for the circulating microparticle, wherein the single signal corresponds to the presence, absence and/or level of the fragments of the target nucleic acid (e.g. genomic DNA) and the target biomolecule (e.g. the target polypeptide) in the sample.

The fragments of the target nucleic acid (e.g. genomic DNA) may comprise a specific sequence of nucleotides and/or the fragments of the target nucleic acid (e.g. genomic DNA) may comprise at least one modified nucleotide or nucleobase. The fragments of the target nucleic acid may not comprise a specific sequence of nucleotides. The fragments of the target nucleic acid may
5 comprise untargeted and/or unknown and/or randomly-selected and or randomly-sampled sequences of nucleotides. For example, the modified nucleotide or nucleobase may be 5-methylcytosine or 5-hydroxy-methylcytosine. The fragments of the target nucleic acid (e.g. genomic DNA) may comprise one or microsatellite sequences and/or microsatellite genomic regions (i.e. short tandem repeats).

10

A target polypeptide may comprise a specific amino acid sequence and/or the target polypeptide may comprise a post-translational modification. For example, the target polypeptide may comprise an acetylated amino acid residue and/or a methylated amino acid residue (for example, a specific acetylated amino acid residue on/within a specific polypeptide and/or a specific
15 methylated amino acid residue on/within a specific polypeptide).

15

The method may comprise measuring the signal corresponding to the presence, absence and/or level of each of the target molecules of the circulating microparticle to produce a set of at least three (informatically) linked signals for the circulating microparticle, wherein one of the linked
20 signals corresponds to the presence, absence and/or level of a first fragment of a target nucleic acid (e.g. genomic DNA) of the circulating microparticle, one of the linked signals corresponds to the presence, absence and/or level of a second fragment of a target nucleic acid (e.g. genomic DNA) of the circulating microparticle, and one of the linked signals corresponds to the presence, absence and/or level of the target biomolecule (e.g. the target polypeptide) of the circulating
25 microparticle.

25

The step of measuring a signal corresponding to the presence, absence and/or level of the fragments of a target nucleic acid (e.g. genomic DNA) may comprise analysing a sequence of each of at least two of the at least two fragments of the target nucleic acid (e.g. genomic DNA),
30 optionally wherein the step of measuring a signal corresponding to the presence, absence and/or level of the fragments of the target nucleic acid (e.g. genomic DNA) comprises sequencing at least a portion of each of at least two of the at least two fragments of the target nucleic acid (e.g. genomic DNA) to produce at least two (informatically) linked sequence reads.

30

The step of measuring a signal corresponding to the presence, absence and/or level of the fragments of a target nucleic acid (e.g. genomic DNA) may comprise: (a) linking at least two of the at least two fragments of the target nucleic acid (e.g. genomic DNA) to produce a set of at least two linked fragments of the target nucleic acid (e.g. genomic DNA); and, optionally, (b) analysing a sequence of each of at least two of the linked fragments in the set. Step (b) may comprise

35

sequencing at least a portion of each of at least two of the linked fragments in the set to produce at least two (informatically) linked sequence reads.

5 The step of measuring a signal corresponding to the presence, absence and/or level of the fragments of a target nucleic acid (e.g. genomic DNA) may comprise: (a) appending each of at least two of the at least two fragments of the target nucleic acid (e.g. genomic DNA) of the circulating microparticle to a barcode sequence to produce a set of linked fragments of the target nucleic acid (e.g. genomic DNA); and, optionally, (b) analysing a sequence of each of at least two of the linked fragments in the set. Step (b) may comprise sequencing at least a portion of each of
10 at least two of the linked fragments in the set to produce at least two (informatically) linked sequence reads, wherein the at least two linked sequence reads are linked by the barcode sequence. Optionally, each of at least two of the at least two fragments of the target nucleic acid may comprise the same barcode sequence.

15 The step of measuring a signal corresponding to the presence, absence and/or level of the fragments of a target nucleic acid (e.g. genomic DNA) may comprise: (a) appending each of at least two of the at least two fragments of the target nucleic acid (e.g. genomic DNA) of the circulating microparticle to a different barcode sequence of a set of barcode sequences to produce a set of linked fragments of the target nucleic acid (e.g. genomic DNA); and, optionally,
20 (b) analysing a sequence of each of at least two of the linked fragments in the set. Step (b) may comprise sequencing at least a portion of each of at least two of the linked fragments in the set to produce at least two (informatically) linked sequence reads. The at least two linked sequence reads may be linked by the set of barcode sequences (i.e. the barcode sequence appended to a first fragment of the target nucleic acid and the barcode sequence appended to a second
25 fragment of the target nucleic acid link the two sequence reads to each other by being present within the same set of barcode sequences).

The step of measuring a signal corresponding to the presence, absence and/or level of the fragments of a target nucleic acid (e.g. genomic DNA) may comprise: (a) appending a first
30 barcode sequence to a first fragment of the target nucleic acid (e.g. genomic DNA) to produce a first barcoded target nucleic acid molecule, and appending a second barcode sequence to a second fragment of the target nucleic acid (e.g. genomic DNA) to produce a second barcoded target nucleic acid molecule, wherein the first and second barcode sequences each comprise the same barcode sequence, or each comprise a different barcode sequence of a set of barcode
35 sequences; and, optionally, (b) analysing a sequence of each of the first and second barcoded target nucleic acid molecules. Step (b) may comprise sequencing at least a portion of each of the first and second barcoded target nucleic acid molecules to produce at least two (informatically) linked sequence reads. The at least two linked sequence reads may be linked by the same barcode sequence or the set of barcode sequences. Step (b) may comprise sequencing all or at

least a portion of each of the first and second barcode sequences appended to the first and second fragments of the target nucleic acid.

- 5 The step of measuring a signal corresponding to the presence, absence and/or level of the fragments of a target nucleic acid (e.g. genomic DNA) may comprise: (a) appending (e.g. annealing or ligating) a first barcoded oligonucleotide to a first fragment of the target nucleic acid (e.g. genomic DNA) to produce a first barcoded target nucleic acid molecule, and appending (e.g. annealing or ligating) a second barcoded oligonucleotide to a second fragment of the target nucleic acid (e.g. genomic DNA) to produce a second barcoded target nucleic acid molecule, 10 wherein the first and second barcoded oligonucleotides each comprise the same barcode sequence, or each comprise a different barcode sequence of a set of barcode sequences; and, optionally, (b) analysing a sequence of each of the first and second barcoded target nucleic acid molecules. Step (b) may comprise sequencing at least a portion of each of the first and second barcoded target nucleic acid molecules to produce at least two (informatically) linked sequence 15 reads. The at least two linked sequence reads may be linked by the same barcode sequence or the set of barcode sequences. Step (b) may comprise sequencing all or at least a portion of each of the first and second barcoded oligonucleotides appended to the first and second fragments of the target nucleic acid.
- 20 The step of measuring a signal corresponding to the presence, absence and/or level of the fragments of a target nucleic acid (e.g. genomic DNA) may comprise: (a) contacting the sample with a multimeric barcoding reagent, wherein the multimeric barcoding reagent comprises first and second barcode regions linked together, wherein each barcode region comprises a nucleic acid sequence; and (b) appending barcode sequences to each of first and second fragments of 25 the target nucleic acid of the microparticle to produce first and second barcoded target nucleic acid molecules for the microparticle, wherein the first barcoded target nucleic acid molecule comprises the nucleic acid sequence of the first barcode region and the second barcoded target nucleic acid molecule comprises the nucleic acid sequence of the second barcode region. The first and second barcode regions may each comprise the same barcode sequence, or the first and 30 second barcode regions may comprise a different barcode sequence of a set of barcode sequences. The method may further comprise (c) analysing a sequence of each of the first and second barcoded target nucleic acid molecules. Step (c) may comprise sequencing at least a portion of each of the first and second barcoded target nucleic acid molecules to produce at least two (informatically) linked sequence reads. The at least two linked sequence reads may be linked 35 by the same barcode sequence or by the set of barcode sequences.

The step of measuring a signal corresponding to the presence, absence and/or level of the fragments of a target nucleic acid (e.g. genomic DNA) may comprise: (a) contacting the sample with a multimeric barcoding reagent, wherein the multimeric barcoding reagent comprises first

- 10 -

and second barcoded oligonucleotides linked together, and wherein the barcoded oligonucleotides each comprise a barcode region; and (b) appending (e.g. annealing or ligating) the first and second barcoded oligonucleotides to first and second fragments of the target nucleic acid of the microparticle to produce first and second barcoded target nucleic acid molecules. The
5 barcode regions of the first and second barcoded oligonucleotides may each comprise the same barcode sequence, or the barcode regions of the first and second barcoded oligonucleotides may each comprise a different barcode sequence of a set of barcode sequences. The method may further comprise (c) analysing a sequence of each of the first and second barcoded target nucleic acid molecules. Step (c) may comprise sequencing at least a portion of each of the first and
10 second barcoded target nucleic acid molecules to produce at least two (informatically) linked sequence reads. The at least two linked sequence reads may be linked by the same barcode sequence or the set of barcode sequences.

The fragments of the target nucleic acid (e.g. genomic DNA) may comprise at least one
15 epigenetic modification (e.g. a modified nucleotide or nucleobase) and the step of measuring a signal corresponding to the presence, absence and/or level of the fragments of the target nucleic acid (e.g. genomic DNA) may comprise measuring a signal corresponding to the presence, absence and/or level of the epigenetic modification (e.g. the modified nucleotide or nucleobase) of the fragments of the target nucleic acid (e.g. genomic DNA). For example, the modified
20 nucleotide or nucleobase may comprise 5-methylcytosine or 5-hydroxy-methylcytosine.

The invention provides a method of analysing a sample comprising a circulating microparticle or a sample derived from a circulating microparticle, wherein the circulating microparticle comprises at least two target molecules, wherein at least one of the target molecules is a fragment of a target
25 nucleic acid (e.g. genomic DNA) comprising an epigenetic modification and at least one of the target molecules is a target biomolecule (e.g. a target polypeptide), and wherein the method comprises measuring a signal corresponding to the presence, absence and/or level of each of the target molecules to produce a set of at least two (informatically) linked signals for the circulating microparticle, wherein at least one of the linked signals corresponds to the presence, absence
30 and/or level of the epigenetic modification in the sample and at least one of the linked signals corresponds to the presence, absence and/or level of the target biomolecule (e.g. the target polypeptide) in the sample.

The invention provides a method of analysing a sample comprising a circulating microparticle or a
35 sample derived from a circulating microparticle, wherein the circulating microparticle comprises at least two target molecules, wherein at least one of the target molecules is a fragment of a target nucleic acid (e.g. genomic DNA) comprising an epigenetic modification and at least one of the target molecules is a target biomolecule (e.g. a target polypeptide), and wherein the method comprises measuring a signal corresponding to the presence, absence and/or level of each of the

target molecules to produce a single signal for the circulating microparticle, wherein the single signal corresponds to the presence, absence and/or level of the fragment of the epigenetic modification and the target biomolecule (e.g. the target polypeptide) in the sample.

5 The method may comprise the step of analysing the sequence of the target nucleic acid (e.g. genomic DNA) comprising an epigenetic modification. Alternatively, the method may not comprise the step of analysing the sequence of the target nucleic acid (e.g. genomic DNA) comprising an epigenetic modification.

10 An epigenetic modification may comprise a modified nucleotide e.g. a modified gDNA nucleotide or a modified RNA nucleotide. The modified nucleotide may comprise a modified base. The modified base may be a methylated base e.g. 5-methylcytosine or 5-hydroxy-methylcytosine. The fragment of a target nucleic acid (e.g. genomic DNA) comprising an epigenetic modification may comprise 5-methylcytosine DNA or 5-hydroxy-methylcytosine DNA.

15 A signal corresponding to the presence, absence and/or level of the epigenetic modification (e.g. the modified DNA or RNA nucleotide) may be measured using a barcoded affinity probe. The barcoded affinity probe may comprise at least one affinity moiety linked to a barcoded oligonucleotide, wherein the barcoded oligonucleotide comprises at least one nucleotide (i.e. wherein the barcoded oligonucleotide comprises a nucleotide sequence at least one nucleotide in length), and wherein the affinity moiety is capable of binding to a target biomolecule (i.e. capable of binding to the epigenetic modification). The signal may be measured by determining the presence, absence and/or level of the barcoded oligonucleotide of the barcoded affinity probe (e.g. by sequencing or PCR).

25 A signal corresponding to the presence, absence and/or level of the epigenetic modification (e.g. the modified DNA or RNA nucleotide) may be measured by flow cytometry and/or fluorescence-activated cell sorting using an optically-labelled affinity probe and/or a fluorescently-labelled affinity probe. The optically-labelled affinity probe and/or fluorescently-labelled affinity probe may be measured and/or detected using optical microscopy and/or fluorescence microscopy visualisation. For example, using a fluorescence microscope, and/or using fluorescent laser-based detection, and/or using a fluorescence-activated cell sorting (FACS) instrument. The optically-labelled affinity probe and/or fluorescently-labelled affinity probe may be measured and/or detected using a sorting process e.g. using fluorescence-activated cell sorting (FACS).

30 A signal corresponding to the presence, absence and/or level of the epigenetic modification (e.g. a modified DNA or RNA nucleotide) may be measured using a method comprising a molecular conversion step. In the case of a modified nucleotide (i.e. a nucleotide comprising a modified base such as 5-methylcytosine or 5-hydroxy-methylcytosine), the molecular conversion step may

be performed to convert said modified base(s) into a different modified or unmodified nucleotide which may be detected (e.g. using PCR or sequencing), providing the signal corresponding to the presence, absence and/or level of the epigenetic modification. This conversion step may comprise a bisulfite conversion step, an oxidative bisulfite conversion step, or any other molecular conversion step. The methods may be used to measure 5-methylcytosine in fragments of genomic DNA of a circulating microparticle.

The method may further comprise one or more steps of partitioning a sample comprising one or more circulating microparticles (or a sample derived from one or more circulating microparticles). Additionally or alternatively, the method may further comprise one or more steps of appending any one or more barcode sequences and/or partition barcode sequences and/or barcoded oligonucleotides to one or more fragments of a target nucleic acid. The one or more barcode sequences and/or barcoded oligonucleotides may be provided by and/or comprised within one or more multimeric barcoding reagents as described herein.

A signal corresponding to the presence, absence and/or level of the non-nucleic acid biomolecule (e.g. target polypeptide) may be measured using a barcoded affinity probe. The barcoded affinity probe may comprise at least one affinity moiety linked to a barcoded oligonucleotide, wherein the barcoded oligonucleotide comprises at least one nucleotide (i.e. wherein the barcoded oligonucleotide comprises a nucleotide sequence at least one nucleotide in length), and wherein the affinity moiety is capable of binding to a target biomolecule (i.e. the target non-nucleic acid biomolecule (e.g. target polypeptide)). The signal may be measured by determining the presence, absence and/or level of the barcoded oligonucleotide of the barcoded affinity probe (e.g. by sequencing or PCR).

A signal corresponding to the presence, absence and/or level of the non-nucleic acid biomolecule (e.g. target polypeptide) may be measured by flow cytometry and/or fluorescence-activated cell sorting using an optically-labelled affinity probe and/or a fluorescently-labelled affinity probe. The optically-labelled affinity probe and/or fluorescently-labelled affinity probe may be measured and/or detected using optical microscopy and/or fluorescence microscopy visualisation. For example, using a fluorescence microscope, and/or using fluorescent laser-based detection, and/or using a fluorescence-activated cell sorting (FACS) instrument. The optically-labelled affinity probe and/or fluorescently-labelled affinity probe may be measured and/or detected using a sorting process e.g. using fluorescence-activated cell sorting (FACS).

A signal corresponding to the presence, absence and/or level of the non-nucleic acid biomolecule (e.g. target polypeptide) may be measured by supports labelled with an affinity probe. The supports labelled with an affinity probe may comprise beads (such as magnetic beads) labelled with affinity probes, for example labelled with antibodies specific for a target polypeptide. The

presence, absence and/or level of the non-nucleic acid biomolecule (e.g. target polypeptide within a circulating microparticle) may be measured by incubating and/or binding said non-nucleic acid biomolecule to said affinity probe(s) on said supports, optionally wherein the support-bound fraction (ie the microparticle(s) comprising, and/or comprising high levels of the said non-nucleic acid biomolecule) is further isolated and/or processed (such as partitioned and/or barcoded and/or analysed by nucleic acid sequencing), and optionally wherein the support-unbound fraction (ie the microparticle(s) not comprising, and/or comprising low levels of the said non-nucleic acid biomolecule) is further isolated and/or processed (such as partitioned and/or barcoded and/or analysed by nucleic acid sequencing).

The signal corresponding to the presence, absence and/or level of the non-nucleic acid biomolecule (e.g. target polypeptide) may be measured separately from the signal corresponding to the presence, absence and/or level of the nucleic acid biomolecule. For example, the signal corresponding to the presence, absence and/or level of the non-nucleic acid biomolecule (e.g. target polypeptide) may be measured by FACS and the signal corresponding to the presence, absence and/or level of the nucleic acid biomolecule may be measured by sequencing.

In the methods, a set of linked signals may be measured for the (or for each) circulating microparticle corresponding to the presence, absence and/or level of fragments of a target nucleic acid (e.g. genomic DNA), an epigenetic modification (e.g. a modified nucleotide such as a modified nucleotide comprising 5-methylcytosine and/or 5-hydroxymethylcytosine) and a target non-nucleic acid biomolecule (e.g. the target polypeptide).

For example, in the methods, the target molecules of the circulating microparticle may comprise at least 2 (different) fragments of a target nucleic acid (e.g. genomic DNA), at least one fragment of a target nucleic acid (e.g. genomic DNA) comprising an epigenetic modification, and at least one target non-nucleic acid biomolecule (e.g. a target polypeptide). The method may comprise measuring a signal corresponding to the presence, absence and/or level of each of the target molecules to produce a set of linked signals for the circulating microparticle. The method may provide a (different) linked signal for each of the target molecules. In the method, each of at least two of the linked signals may correspond to the presence, absence and/or level of one of the fragments of the target nucleic acid (e.g. genomic DNA); at least one of the linked signals may correspond to the presence, absence and/or level of the epigenetic modification (e.g. a modified nucleotide such as a modified nucleotide comprising 5-methylcytosine and/or 5-hydroxymethylcytosine); and at least one of the linked signals may correspond to the presence, absence and/or level of the target non-nucleic acid biomolecule (e.g. the target polypeptide).

The circulating microparticle may comprise at least 3, at least 4, at least 5, at least 10, at least 50, at least 100, at least 500, at least 1000, at least 5000, at least 10,000, at least 100,000, or at least

- 14 -

1,000,000 (different) target molecules, and optionally wherein the method comprises producing a set of at least 3, at least 4, at least 5, at least 10, at least 50, at least 100, at least 500, at least 1000, at least 5000, at least 10,000, at least 100,000, or at least 1,000,000 (different) linked signals for the circulating microparticle (i.e. a (different) linked signal for each of the target molecules of the circulating microparticle).

The target molecules of the circulating microparticle may comprise at least 2, at least 3, at least 4, at least 9, at least 49, at least 99, at least 499, at least 999, at least 4999, at least 9,999, at least 99,999, or at least 999,999 (different) fragments of a target nucleic acid (e.g. genomic DNA), and at least one target non-nucleic acid biomolecule (e.g. a target polypeptide), optionally wherein the method comprises producing a set of at least 3, at least 4, at least 5, at least 10, at least 50, at least 100, at least 500, at least 1000, at least 5000, at least 10,000, at least 100,000, or at least 1,000,000 (different) linked signals for the circulating microparticle (i.e. a (different) linked signal for each of the target molecules of the circulating microparticle).

The target molecules of the circulating microparticle may comprise at least 2, at least 3, at least 4, at least 9, at least 49, at least 99, at least 499, at least 999, at least 4999, at least 9,999, at least 99,999, or at least 999,999 (different) target polypeptides, and at least one fragment of a target nucleic acid (e.g. genomic DNA), optionally wherein the method comprises producing a set of at least at least 3, at least 4, at least 5, at least 10, at least 50, at least 100, at least 500, at least 1000, at least 5000, at least 10,000, at least 100,000, or at least 1,000,000 (different) linked signals for the circulating microparticle (i.e. a (different) linked signal for each of the target molecules of the circulating microparticle).

The sample may comprise first and second circulating microparticles, wherein each circulating microparticle comprises target molecules (e.g. at least 2 or at least 3 target molecules), and wherein the method comprises performing the step of measuring (as described herein) to produce a set of linked signals for the first circulating microparticle and performing the step of measuring as described herein to produce a set of linked signals for the second circulating microparticle.

For example, the step of measuring a signal corresponding to the presence, absence and/or level of the fragments of a target nucleic acid (e.g. genomic DNA) may comprise: (a) contacting the sample with a library comprising at least two multimeric barcoding reagents, wherein each multimeric barcoding reagent comprises first and second barcode regions linked together, wherein each barcode region comprises a nucleic acid sequence and wherein the first and second barcode regions of a first multimeric barcoding reagent are different to the first and second barcode regions of a second multimeric barcoding reagent of the library; and (b) appending barcode sequences to each of first and second fragments of the target nucleic acid of the first microparticle to produce first and second barcoded target nucleic acid molecules for the

- 15 -

first microparticle, wherein the first barcoded target nucleic acid molecule comprises the nucleic acid sequence of the first barcode region of the first multimeric barcoding reagent and the second barcoded target nucleic acid molecule comprises the nucleic acid sequence of the second barcode region of the first multimeric barcoding reagent, and appending barcode sequences to each of first and second fragments of the target nucleic acid of the second microparticle to produce first and second barcoded target nucleic acid molecules for the second microparticle, wherein the first barcoded target nucleic acid molecule comprises the nucleic acid sequence of the first barcode region of the second multimeric barcoding reagent and the second barcoded target nucleic acid molecule comprises the nucleic acid sequence of the second barcode region of the second multimeric barcoding reagent.

For example, the step of measuring a signal corresponding to the presence, absence and/or level of the fragments of a target nucleic acid (e.g. genomic DNA) may comprise: (a) contacting the sample with a library comprising at least two multimeric barcoding reagents, wherein each multimeric barcoding reagent comprises first and second barcoded oligonucleotides linked together, wherein the barcoded oligonucleotides each comprise a barcode region and wherein the barcode regions of the first and second barcoded oligonucleotides of a first multimeric barcoding reagent of the library are different to the barcode regions of the first and second barcoded oligonucleotides of a second multimeric barcoding reagent of the library; and (b) appending (e.g. annealing or ligating) the first and second barcoded oligonucleotides of the first multimeric barcoding reagent to first and second fragments of the target nucleic acid of the first microparticle to produce first and second barcoded target nucleic acid molecules, and appending (e.g. annealing or ligating) the first and second barcoded oligonucleotides of the second multimeric barcoding reagent to first and second fragments of the target nucleic acid of the second microparticle to produce first and second barcoded target nucleic acid molecules.

The sample may comprise n circulating microparticles, wherein each circulating microparticle comprises target molecules (e.g. at least 2 or at least 3 target molecules), and wherein the method comprises performing the step of measuring (as described herein) for each circulating microparticle to produce a set of linked signals for each circulating microparticle, optionally wherein n is at least 3, at least 5, at least 10, at least 50, at least 100, at least 1000, at least 10,000, at least 100,000, at least 1,000,000, at least 10,000,000, or at least 100,000,000 circulating microparticles.

The methods may further comprise a step of determining the identity of the cell of origin and/or tissue of origin of the target biomolecules from which the the set of linked signals is derived. The step of determining the identity of the cell of origin and/or tissue of origin may comprise identifying in the set of linked signals one or more signature signals. A signature signal may be a signal corresponding to the presence, absence and/or level of a signature target biomolecule, wherein a

signature target biomolecule is a target biomolecule that is characteristic of a particular cell and/or tissue.

5 A signature signal may be a combinatoric signature signal, corresponding to the presence, absence and/or level of any two or more signature target biomolecules, wherein said signature target biomolecules are target biomolecules that are characteristic of a particular cell and/or tissue (e.g. wherein said target biomolecules together are characteristic of a particular cell and/or tissue). For example, a combinatoric signature signal may correspond to the presence, absence and/or level of any two or more biomolecules from Biomolecule Group 1; optionally, a
10 combinatoric signature signal may correspond to the the presence, absence and/or level of any two or more biomolecules from Biomolecule Group 1, as well as any one or more reference sequences, as well as any one or more epigenetic signals (such as one or more signals corresponding to 5-methylcytosine, and/or one or more signals corresponding to 5-hydroxymethylcytosine). A signature signal may be a combinatoric signature signal,
15 corresponding to the presence, absence and/or level of any number of signature target biomolecules, such as at least 3, at least 4, at least 5, at least 10, at least 20, at least 30, or at least 50 signature target biomolecules (and/or lists or groups thereof, such as lists or groups of reference sequences, and/or lists or groups of signals corresponding to 5-methylcytosine and/or 5-hydroxymethylcytosine).

20 The cell of origin may be from a specific subject (e.g. a fetal cell, a maternal cell or a paternal cell). The cell of origin may be a lung cell, a liver cell, an ovarian cell, a kidney cell, a pancreas cell, a uterine cell, a skin cell, an epithelial cell, an endothelial cell, a brain cell, a bladder cell, a blood cell, a lymphocyte cell, a prostate cell, a breast cell, a colorectal cell, a brain cell, a uterine
25 cell, a heart cell, a vascular cell (such as an arterial cell or a venous cell), and/or any other type of cell.

The cell of origin may be a cancerous cell or a malignant cell. The cell of origin may be a lung cancer cell, a breast cancer cell, an ovarian cancer cell, a prostate cancer cell, a kidney cancer
30 cell, a liver cancer cell, a blood cancer cell, a leukaemia cell, a lymphoma cell, a colorectal cancer cell, a pancreatic cancer cell, a brain cancer cell, a uterine cancer cell, a bile duct cancer cell, a skin cancer cell, a melanoma cell, a bladder cancer cell, an oesophageal cancer cell, an oral cancer cell, a pharyngeal cancer cell, and/or any other type of cancer cell.

35 The tissue of origin may be from a specific subject (e.g. a fetal tissue, a maternal tissue or a paternal tissue). The tissue of origin may be a lung tissue, a liver tissue, an ovarian tissue, a cardiac tissue, a vascular tissue, an endovascular tissue, an endovascular plaque tissue, a stable endovascular plaque tissue, an unstable and/or vulnerable endovascular plaque tissue, an atherosclerotic tissue, a thrombosis tissue, an embolism tissue, a cerebrovascular tissue, an

endocarditis tissue, a myocarditis tissue, a peripheral artery tissue, a brain tissue, a cardiomyopathy tissue, and/or any other tissue.

5 The tissue of origin may be cancerous tissue or malignant tissue. The tissue of origin may be cancerous lung tissue, cancerous liver tissue, cancerous ovarian tissue, cancerous breast tissue, cancerous prostate tissue, cancerous blood tissue, cancerous leukaemia tissue, cancerous lymphoma tissue, cancerous colorectal tissue, cancerous pancreatic tissue, cancerous brain tissue, cancerous skin tissue, cancerous melanoma tissue, cancerous bladder tissue, cancerous oesophageal tissue, and/or any other cancerous tissue.

10 A signature signal may comprise a signal corresponding to the presence, absence and/or level of a first signature biomolecule and a signal corresponding to the presence, absence and/or level of a second signature biomolecule. The first and second signature biomolecules may take any of the forms described herein for target biomolecules. For example, a signature signal may
15 comprise a signal corresponding to the presence, absence and/or level of any one or more biomolecules listed in biomolecule group 1.

A signature biomolecule may be a polypeptide that is only expressed in a specific cell type or tissue type (e.g. a cancer cell or a fetal cell). A signature biomolecule may be a polypeptide that is
20 preferentially expressed in a specific cell type or tissue type (e.g. a cancer cell or a fetal cell). A signature biomolecule may be a nucleic acid (such as an mRNA molecule or a microRNA molecule) that is only expressed (or is preferentially expressed) in a specific cell type or tissue type (e.g. a cancer cell or a fetal cell, or an endovascular tissue such as an endovascular plaque). For example, a signature biomolecule may comprise any one or more biomolecules listed in
25 biomolecule group 1.

A signature biomolecule may be epigenetic modification e.g. genomic DNA fragments comprising 5-hydroxymethylcytosine. Genomic DNA fragments comprising 5-hydroxymethylcytosine may provide a signature signal for cancerous and/or malignant cells or tissues.

30 A signature biomolecule may be a polypeptide or RNA encoding the polypeptide e.g. TTF-1 (also known as NK2 Homeobox 1) or TTF-1 RNA. TTF-1 (or TTF-1 RNA) may provide a signature signal for lung cells and/or tissue.

35 A signature signal for lung cancer may be provided by measuring a signal corresponding to the presence, absence and/or level of genomic DNA fragments comprising 5-hydroxymethylcytosine (a first signature biomolecule) and a signal corresponding to the presence, absence and/or level of TTF-1 or TTF-1 RNA (as second signature biomolecule).

The invention provides a method of analysing a sample comprising a circulating microparticle or a sample derived from a circulating microparticle, and wherein the method comprises: (a) contacting the sample with a barcoded affinity probe, wherein the barcoded affinity probe comprises at least one affinity moiety linked to a barcoded oligonucleotide, wherein the barcoded oligonucleotide comprises at least one nucleotide (i.e. wherein the barcoded oligonucleotide comprises a nucleotide sequence at least one nucleotide in length), and wherein the affinity moiety is capable of binding to a target biomolecule; (b) forming a reaction mixture, wherein the step of forming the reaction mixture comprises binding the affinity moiety to the target molecule, if present, to form a barcoded biomolecule complex comprising the barcoded affinity probe and the target biomolecule; and (c) determining the presence, absence and/or level of the target biomolecule in the sample by measuring the presence, absence and/or level of the barcoded oligonucleotide in the reaction mixture.

The invention provides a method of analysing a sample comprising a circulating microparticle or a sample derived from a circulating microparticle, wherein the circulating microparticle comprises a target biomolecule, and wherein the method comprises: (a) contacting the sample with a barcoded affinity probe, wherein the barcoded affinity probe comprises at least one affinity moiety linked to a barcoded oligonucleotide, wherein the barcoded oligonucleotide comprises at least one nucleotide (i.e. wherein the barcoded oligonucleotide comprises a nucleotide sequence at least one nucleotide in length), and wherein the affinity moiety is capable of binding to the target biomolecule; (b) forming a reaction mixture, wherein the step of forming the reaction mixture comprises binding the affinity moiety to the target biomolecule to form a barcoded biomolecule complex comprising the barcoded affinity probe and the target biomolecule; and (c) determining the level of the target biomolecule in the sample by measuring the level of the barcoded oligonucleotide in the reaction mixture.

The invention provides a method of analysing a sample comprising a circulating microparticle or a sample derived from a circulating microparticle, and wherein the method comprises: (a) contacting the sample with at least one affinity moiety, and wherein the affinity moiety is capable of binding to a target biomolecule; (b) forming a reaction mixture, wherein the step of forming the reaction mixture comprises (i) binding the affinity moiety to the target biomolecule, if present, and (ii) contacting the sample with a barcoded oligonucleotide and linking the barcoded oligonucleotide to the affinity moiety to form a barcoded biomolecule complex comprising a barcoded affinity probe and the target biomolecule, wherein the barcoded affinity probe comprises at least one affinity moiety linked to the barcoded oligonucleotide, and wherein the barcoded oligonucleotide comprises at least one nucleotide (i.e. wherein the barcoded oligonucleotide comprises a nucleotide sequence at least one nucleotide in length); and (c) determining the presence, absence and/or level of the target biomolecule in the sample by measuring the presence, absence and/or level of the barcoded oligonucleotide in the reaction mixture.

The invention provides a method of analysing a sample comprising a circulating microparticle or a sample derived from a circulating microparticle, wherein the circulating microparticle comprises a target biomolecule, and wherein the method comprises: (a) contacting the sample with at least one affinity moiety, and wherein the affinity moiety is capable of binding to the target biomolecule; (b) forming a reaction mixture, wherein the step of forming the reaction mixture comprises (i) binding the affinity moiety to the target biomolecule and (ii) contacting the sample with a barcoded oligonucleotide and linking the barcoded oligonucleotide to the affinity moiety to form a barcoded biomolecule complex comprising a barcoded affinity probe and the target biomolecule, wherein the barcoded affinity probe comprises at least one affinity moiety linked to the barcoded oligonucleotide, and wherein the barcoded oligonucleotide comprises at least one nucleotide (i.e. wherein the barcoded oligonucleotide comprises a nucleotide sequence at least one nucleotide in length); and (c) determining the level of the target biomolecule in the sample by measuring the level of the barcoded oligonucleotide in the reaction mixture.

The step of forming a reaction mixture may comprise incubating the reagents under conditions suitable for binding of the affinity moiety to the target biomolecule.

Prior to the step of measuring the presence, absence and/or level of the barcoded oligonucleotide in the sample, the method may comprise removing or depleting barcoded affinity probes and/or barcoded oligonucleotides that are not part of barcoded biomolecule complexes.

Measuring the level of the barcoded oligonucleotide in the reaction mixture may comprise quantifying the level of the barcoded oligonucleotide in the reaction mixture.

A barcoded oligonucleotide may be linked to an affinity moiety directly or indirectly (e.g. via one or more linker molecules). A barcoded oligonucleotide may be linked to an affinity moiety via a linker molecule, wherein said linker molecule is appended to and/or linked to and/or bound to (covalently or non-covalently) both at least one affinity moiety, and at least one barcoded oligonucleotide. A barcoded oligonucleotide may be linked to any affinity moiety by one or more covalent linkage(s) (or bond(s)) (e.g. by a covalent bond such as a bond created by the LightningLink® antibody labelling kit, Innova Biosciences), one or more non-covalent linkages (or bond(s)) (e.g. a protein-protein interaction or a streptavidin-biotin linkage e.g. an affinity moiety may comprise a streptavidin domain and a barcoded oligonucleotide may comprise a biotin moiety) or a nucleic acid hybridization linkage. Any one or more linker molecule may be a biopolymer (e.g. a nucleic acid molecule) or a synthetic polymer. Any one or more linker molecule may comprise one or more units of ethylene glycol and/or poly(ethylene) glycol (e.g. hexa-ethylene glycol or penta-ethylene glycol). Any one or more linker molecule may comprise

– 20 –

one or more ethyl groups, such as one or more C3 (three-carbon) spacers, C6 spacers, C12 spacers, or C18 spacers.

5 A sample may be contacted with a library of at least 2, at least 3, at least 5, at least 10, at least 20, or at least 30 different barcoded affinity probes.

10 A barcoded affinity probe may comprise an aptamer, optionally wherein the barcoded affinity probe is an aptamer. The aptamer may provide both the affinity moiety and barcoded oligonucleotide of the barcoded affinity probe.

15 An aptamer may comprise at least one affinity moiety linked to a barcoded oligonucleotide, wherein the barcoded oligonucleotide comprises at least one nucleotide, and wherein the affinity moiety is capable of binding to a target biomolecule. The aptamer may comprise a barcode sequence. Any or all of the nucleic acid sequence of the aptamer may be associated with, and/or serve to identify, the affinity moiety of the aptamer, and/or identify the target biomolecule for which the affinity moiety of the aptamer is capable of binding.

20 An affinity moiety may be capable of binding to a target biomolecule. The affinity moiety may be capable of specifically binding to a target biomolecule. The affinity moiety may bind to a target biomolecule. The affinity moiety may bind specifically bind to a target biomolecule. The affinity moiety may have a high affinity for a target biomolecule.

25 An affinity moiety may comprise one or more of: an antibody, an antibody fragment, a light chain antibody fragment, a single-chain variable fragment (scFv), a peptide, a cell penetrating peptide, an aptamer, a DNA aptamer, and/or an RNA aptamer.

An affinity moiety may comprise an antibody or fragment thereof and the target molecule may be a polypeptide.

30 An affinity moiety may comprise an antibody or fragment thereof and the target molecule may be a fragment of a nucleic acid.

35 An affinity moiety may comprise an antibody or fragment thereof and the target molecule may be a fragment of a nucleic acid comprising an epigenetic modification e.g 5-methylcytosine or 5-hydroxy-methylcytosine.

An affinity moiety may comprise an aptamer and the target molecule may be a polypeptide.

An affinity moiety may comprise an aptamer and the target molecule may be a fragment of a nucleic acid.

5 An affinity moiety may comprise aptamer and the target molecule may be a fragment of a nucleic acid comprising an epigenetic modification e.g 5-methylcytosine or 5-hydroxy-methylcytosine.

10 The barcoded affinity probe may comprise an aptamer, wherein said aptamer is comprised within an aptamer sequence within an affinity oligonucleotide. The barcoded affinity probe may comprise an aptamer, wherein said aptamer is comprised within an aptamer sequence within an affinity oligonucleotide, wherein said affinity oligonucleotide comprises a barcode sequence. The barcoded affinity probe may comprise an aptamer, wherein said aptamer is comprised within an aptamer sequence within an affinity oligonucleotide, wherein said affinity oligonucleotide comprises a barcode sequence, wherein all or part of said barcode sequence is partially or fully comprised of said aptamer sequence. The aptamer and/or aptamer sequence and/or affinity
15 oligonucleotide and/or barcode sequence may comprise one or more DNA nucleotides. Optionally, any said aptamer and/or aptamer sequence and/or affinity oligonucleotide and/or barcode sequence may comprise one or more RNA nucleotides.

20 A barcoded affinity probe may comprise at least two affinity moieties. The barcoded affinity probe may comprise at least first and second affinity moieties, wherein said first affinity moiety is capable of binding to a first target biomolecule, and wherein said second affinity moiety is capable of binding to a second target biomolecule, wherein said first and second target biomolecules are different.

25 A barcoded affinity probe may comprise at least 3, at least 4, at least 5, or at least 10 different affinity moieties. Optionally, each of the affinity moieties is capable of binding to a different target biomolecule.

30 A barcoded affinity probe may comprise at least two affinity moieties that are linked directly or indirectly. The at least two affinity moieties of a barcoded affinity probe may be linked to a support (e.g. a solid support), a molecular support, or a macromolecular support.

35 A barcoded affinity probe may comprise at least two affinity moieties. Each of the affinity moieties may comprise an aptamer. The at least two affinity moieties of a barcoded affinity probe may be comprised within a single aptamer. The at least two affinity moieties of a barcoded affinity probe may be comprised within a single contiguous nucleic acid sequence (e.g. a DNA sequence and/or an RNA sequence).

A barcoded affinity probe may comprise at least two different barcoded oligonucleotides.

A barcoded oligonucleotide comprises at least one nucleotide. A barcoded oligonucleotide may comprise a barcode sequence. The barcoded oligonucleotide comprises a barcode sequence of at least 2, at least 3, at least 5, at least 10, at least 20, or at least 30 nucleotides.

5

A barcoded oligonucleotide may comprise a barcode sequence associated with and/or identifying of the affinity moiety to which it is linked. Each of the barcoded oligonucleotides linked with the same affinity moiety (e.g., the same antibody specific for the same protein target) may comprise the same sequence (e.g. the same barcode sequence). Each of the barcoded oligonucleotides linked with the same affinity moiety comprise different sequences (e.g. two or more different barcode sequences). Optionally, each of the barcoded oligonucleotides linked with a different affinity moiety may comprise different sequences (e.g. two or more different barcode sequences).

10

A barcoded oligonucleotide may comprise an adapter and/or coupling sequence, wherein said sequence is at least 1, at least 2, at least 3, at least 5, at least 10, at least 20, or at least 30 nucleotides in length. An adapter and/or coupling sequence of a barcoded oligonucleotide may comprise a sequence complementary to a target region of a barcoded oligonucleotide comprised within any multimeric barcoding reagent and/or library thereof. An adapter and/or coupling sequence of a barcoded oligonucleotide may comprise a poly(A) sequence of 2 or more nucleotides in length. An adapter and/or coupling sequence within a barcoded oligonucleotide may be comprised within the 3' end, and/or within the 5' end, of said barcoded oligonucleotide.

15

20

A barcoded affinity probe may comprise one or more secondary barcoded oligonucleotides, wherein said secondary barcoded oligonucleotide comprises a sequence at least partially complementary to all or part of one or more (non-secondary) barcoded oligonucleotides. A secondary barcoded oligonucleotide may be fully or partially annealed (i.e. hybridised) to any one or more (non-secondary) barcoded oligonucleotides. A secondary barcoded oligonucleotide may be fully or partially annealed (i.e. hybridised) to any one or more (non-secondary) barcoded oligonucleotide(s) in a secondary barcoded oligonucleotide annealing reaction. A secondary barcoded oligonucleotide annealing reaction may take place prior to, and/or after, and/or during any of steps (a), (b) or (c). A secondary barcoded oligonucleotide may comprise one or more nucleotides of a barcode sequence, wherein said barcode sequence is associated with and/or identifying of the affinity moiety to which it is linked within a barcoded affinity probe.

25

30

A barcoded affinity probe may comprise one or more affinity moieties, and one or more primary barcoded oligonucleotides, and one or more secondary barcoded oligonucleotides.

35

The sample may comprise one or more circulating microparticles and/or the sample may be derived from one or more circulating microparticles.

A biomolecule may be a polypeptide (e.g. a protein), a carbohydrate, a lipid, or a nucleic acid. A biomolecule may be a metabolite.

- 5 The sample may comprise a first circulating microparticle and a second circulating microparticle, or wherein the sample is derived from a first circulating microparticle and a second circulating microparticle, wherein step (b) comprises forming at least one barcoded biomolecule complex comprising a barcoded affinity probe and a target biomolecule of the first circulating microparticle, and forming at least one barcoded biomolecule complex comprising a barcoded affinity probe and
10 a target biomolecule of the second circulating microparticle. The sample may further comprise a fragment of a target nucleic acid of the first circulating microparticle and a fragment of a target nucleic acid of the second circulating microparticle.

- In step (a), (b) and/or (c) the barcoded affinity probes may be at any concentration, for example at
15 concentrations of at least 100 nanomolar, at least 10 nanomolar, at least 1 nanomolar, at least 100 picomolar, at least 10 picomolar, at least 1 picomolar, at least 100 femtomolar, at least 10 femtomolar, or at least 1 femtomolar. The concentrations may be 1 picomolar to 100 nanomolar, 10 picomolar to 10 nanomolar, or 100 picomolar to 1 nanomolar.

- 20 Optionally, in any one or more steps of any of the methods (such as any step of appending coupling sequences and/or coupling molecules, any step of appending barcode sequences such as any step of appending and/or linking and/or connecting barcoded oligonucleotides (such as any step of appending/linking/connecting barcode sequences comprised within barcoded oligonucleotides), the step(s) and/or method(s) may be performed in a high-viscosity solution.
- 25 Optionally, such a high-viscosity solution may be comprised of a poly (ethylene) glycol (PEG) solution, such as one or more of: PEG 400, PEG 1000, PEG 2000, PEG 4000, PEG 5000, PEG 8000, PEG 10000, and/or PEG 20,000. Optionally, such a solution may comprise at least 5% poly (ethylene) glycol, at least 10% poly (ethylene) glycol, at least 20% poly (ethylene) glycol, at least 25% poly (ethylene) glycol, at least 30% poly (ethylene) glycol, at least 40% poly (ethylene)
30 glycol, or at least 50% poly (ethylene) glycol by weight or by volume; optionally, such a solution may comprise any two or more PEG molecules wherein each such two or more PEG molecules are present at one of these said concentrations by weight or volume. Optionally, such a high-viscosity solution may comprise the solution employed during any step of annealing barcoded oligonucleotides to target nucleic acids. Optionally, such a high-viscosity solution may have a
35 dynamic viscosity of at least 1.0 centipoise, at least 1.1 centipoise, at least 1.2 centipoise, at least 1.5 centipoise, at least 2.0 centipoise, at least 5.0 centipoise, at least 10.0 centipoise, at least 20.0 centipoise, at least 50.0 centipoise, at least 100.0 centipoise, or at least 200.0 centipoise (e.g. at 25 degrees Celsius at standard sea-level pressure). Preferably, such a high-viscosity solution will have a dynamic viscosity of at least 1.5 centipoise. The use of a high-viscosity

solution may slow the diffusion of reagents (such as barcoded oligonucleotides and/or multimeric barcoding reagents) to prevent or retard diffusion away from their target molecules such as target nucleic acids.

- 5 Optionally, in any one or more steps of any of the methods (such as any step of appending coupling sequences and/or coupling molecules, any step of appending barcode sequences such as any step of appending and/or linking and/or connecting barcoded oligonucleotides (such as any step of appending/linking/connecting barcode sequences comprised within barcoded oligonucleotides)), the step(s) and/or method(s) may be performed in a solution comprising one or
 10 more molecular crowding reagents, i.e. wherein said molecular crowding reagent(s) have the effect of increasing the effective concentration of target molecules and/or barcoded oligonucleotides and/or multimeric barcoding reagents and/or other constituents in said step. Optionally, any one or more molecular crowding reagents may comprise beads and/or other solid supports of any size, such as micron-scale beads (such as beads with a diameter of at least 1.0,
 15 at least 2.0, at least 3.0, at least 5.0, at least 10, at least 20, at least 50, or at least 100 micrometres) and/or nanometre-sized beads (such as beads with a diameter of at least 1.0, at least 2.0, at least 3.0, at least 5.0, at least 10, at least 20, at least 50, or at least 100 nanometres).

- 20 One or more steps of removing and/or depleting unbound barcoded affinity probes may be performed during and/or after any step of binding one or more barcoded affinity probes to one or more biomolecules from any one or more circulating microparticles.

- 25 Optionally, any method of measuring a biomolecule from a circulating microparticle may comprise measurement with a single barcoded affinity probe. Optionally, any method of measuring a biomolecule from a circulating microparticle may comprise measurement with a single barcoded affinity probe, wherein said single barcoded affinity probe comprises an oligonucleotide at least a single nucleotide in length.

- 30 Optionally, any nucleotide and/or oligonucleotide sequence at least a single nucleotide in length may be considered to be a barcode and/or a barcode sequence (and/or a barcoded oligonucleotide) within a barcoded affinity probe. Said nucleotide and/or oligonucleotide sequence at least a single nucleotide in length does not need to be different to any other nucleotide and/or oligonucleotide sequence within said barcoded affinity probe and/or to any other nucleotide and/or
 35 oligonucleotide sequence within any other barcoded affinity probe.

Step (c) of the method may comprise measuring the presence, absence and/or level of the barcoded oligonucleotide by analysing a nucleotide sequence of the barcoded oligonucleotide, optionally wherein the sequence is analysed by sequencing (wherein at least a portion of the

- 25 -

barcoded oligonucleotide is sequenced) or PCR (wherein at least a portion of the barcoded oligonucleotide is amplified).

5 Step (c) may comprise measuring the presence, absence and/or level of the barcoded oligonucleotide by a primer-extension and/or PCR reaction and/or a quantitative or semi-quantitative PCR reaction (such as a real-time PCR reaction).

10 Step (c) may comprise measuring the presence, absence and/or level of the barcoded oligonucleotide by a primer-extension and/or PCR reaction and/or a quantitative or semi-quantitative PCR reaction (such as a real-time PCR reaction), wherein at least one primer in the reaction is specific for and/or at least partially complementary to (and/or at least partially identical to) at least part of said barcoded oligonucleotide.

15 In the methods, step (b) or step (c) may comprise linking together at least two barcoded biomolecule complexes of the first circulating microparticle and linking together at least two barcoded biomolecule complexes of the second circulating microparticle.

20 A sample comprising one or more circulating microparticle may be chemically crosslinked (e.g. with formaldehyde). The circulating microparticles may be chemically crosslinked prior to step (a), (b) and/or (c).

25 A sample comprising one or more circulating microparticles may be permeabilised (e.g. with chemical surfactant). The circulating microparticles may be permeabilised prior to step (a) and/or (b).

A sample comprising one or more circulating microparticles may be chemically crosslinked (e.g. with formaldehyde) and then permeabilised (e.g. with a chemical surfactant) prior to step (a) and/or (b).

30 The method may (optionally as part of step (c)) comprise: (i) contacting the reaction mixture with a multimeric barcoding reagent, wherein the multimeric barcoding reagent comprises first and second barcode regions linked together, wherein each barcode region comprises a nucleic acid sequence; (ii) appending a barcode sequence of a barcode region of the multimeric barcoding reagent to the barcoded oligonucleotide of the at least one barcoded biomolecule complex of the circulating microparticle; and (iii) measuring the presence, absence and/or level of the barcoded oligonucleotide in the reaction mixture by analysing the appended barcode sequence of the barcode region of the multimeric barcoding reagent.

35

The reaction mixture may further comprise a fragment of a target nucleic acid of the circulating microparticle, and wherein the method: (i) contacting the reaction mixture with a multimeric barcoding reagent, wherein the multimeric barcoding reagent comprises first and second barcode regions linked together, wherein each barcode region comprises a nucleic acid sequence; (ii) 5 appending a barcode sequence of a first barcode region of the multimeric barcoding reagent to the barcoded oligonucleotide (i.e. a first fragment of a target nucleic acid) of the at least one barcoded biomolecule complex of the circulating microparticle to produce a first barcoded target nucleic acid molecule, and appending a barcode sequence of a second barcode region of the multimeric barcoding reagent to the fragment of the target nucleic acid (i.e. a second fragment of 10 a target nucleic acid) to produce a second barcoded target nucleic acid molecule; and (iii) analysing a sequence of each of the first and second barcoded target nucleic acid molecules.

The step of analysing a sequence of each of the first and second barcoded target nucleic acid molecules may be performed by sequencing at least a portion of each of the first and second 15 barcoded target nucleic acid molecules.

The method may further comprise sequencing at least a portion of each of the first and second barcoded target nucleic acid molecules of the first circulating microparticle. The method may comprise producing a sequence read for the first barcoded target nucleic acid molecule, wherein 20 the sequence read comprises at least a portion of the sequence of the first barcode region of the multimeric barcoding reagent and at least a portion of the sequence of the first fragment of a target nucleic acid of the circulating microparticle. The method may comprise producing a sequence read for the second barcoded target nucleic acid molecule, wherein the sequence read comprises at least a portion of the sequence of the second barcode region of the multimeric 25 barcoding reagent and at least a portion of the sequence of the second fragment of a target nucleic acid of the circulating microparticle.

The method may (optionally as part of step (c)) comprise partitioning the reaction mixture into at least first and second partitions and analysing the nucleotide sequences of the barcoded 30 oligonucleotides of the barcoded biomolecule complexes in each of the first and second partition.

The method may comprise partitioning the reaction mixture into at least 3, at least 4, at least 5, at least 10, at least 100, at least 1000, at least 10,000, at least 100,000, at least 1,000,000, at least 10,000,000, at least 100,000,000, or at least 1,000,000,000 partitions. Preferably, the method 35 comprises partitioning the reaction mixture into at least 1000 partitions.

A target nucleic acid molecule may comprise a barcoded oligonucleotide of a barcoded biomolecule complex of a circulating microparticle. The barcoded oligonucleotide of the barcoded

biomolecule complex of the circulating microparticle may be present within the barcoded biomolecule complex or derived from the barcoded biomolecule complex.

5 Two or more target nucleic acid molecules may comprise both a fragment of a target nucleic acid of a microparticle and a barcoded oligonucleotide of a barcoded biomolecule complex of a circulating microparticle.

10 Two or more target nucleic acid molecules may comprise both a fragment of a target nucleic acid (e.g. genomic DNA) of a microparticle and a barcoded oligonucleotide of a barcoded biomolecule complex of the circulating microparticle.

15 Two or more target nucleic acid molecules may comprise both a fragment of a target nucleic acid (e.g. RNA) of a microparticle and a barcoded oligonucleotide of a barcoded biomolecule complex of the circulating microparticle.

20 The step of analysing the nucleotide sequences of the barcoded oligonucleotides of the barcoded biomolecule complexes may comprise appending a first partition barcode sequence to at least one barcoded oligonucleotide (a first fragment of a target nucleic acid of the first partition) partitioned into said first partition (to produce a first barcoded target nucleic acid molecule of the first partition), wherein the at least one barcoded oligonucleotide partitioned into said first partition is comprised in or derived from a barcoded biomolecule complex, and appending a second partition barcode sequence to at least one barcoded oligonucleotide (a first fragment of a target nucleic acid of the second partition) partitioned into said second partition (to produce a first barcoded target nucleic acid molecule of the second partition), wherein the at least one barcoded oligonucleotide partitioned into said first partition is comprised in or derived from a barcoded biomolecule complex. Preferably, the first and second partitions each comprise a barcoded oligonucleotide comprised in or derived from a barcoded biomolecule complex.

30 The first and second partition barcode sequences may be different. The first partition barcode sequence may be comprised within a first set of partition barcode sequences, and the second partition barcode sequence may be comprised within a second set of partition barcode sequences, wherein said first and second sets of partition barcode sequences are different. The first partition barcode sequence may be the nucleic acid sequence of a barcode region of a first multimeric barcoding reagent and the second partition barcode sequence may be the nucleic acid sequence of a second multimeric barcoding reagent, wherein the first and second multimeric barcoding reagents each comprise two or more barcode regions linked together;

The step of analysing the nucleotide sequences of the barcoded oligonucleotides of the barcoded biomolecule complexes may further comprise analysing appended partition barcode sequences from each of said first and second partitions.

- 5 A fragment of a target nucleic acid (e.g. gDNA or RNA) from a circulating microparticle (a second fragment of a target nucleic acid of the first partition) may also be appended to a said first partition barcode sequence of said first partition (to produce a second barcoded target nucleic acid molecule of the first partition), and/or a fragment of a target nucleic acid (e.g. gDNA or RNA) from
10 a different circulating microparticle (a second fragment of a target nucleic acid of the second partition) may also be appended to a said second partition barcode sequence of said second partition (to produce a second barcoded target nucleic acid molecule of the second partition).

- The step of analysing a sequence of each of the first and second barcoded target nucleic acid molecules may be performed by sequencing at least a portion of each of the first and second
15 barcoded target nucleic acid molecules.

- The method may further comprise analysing a sequence of each of the first and second barcoded target nucleic acid molecules of the first partition, and analysing a sequence of each of the first and second barcoded target nucleic acid molecules of the second partition. Optionally, the step
20 of analysing a sequence is performed by sequencing at least a portion of each of the first and second barcoded target nucleic acid molecules.

- The method may further comprise sequencing at least a portion of each of the first and second barcoded target nucleic acid molecules of the first partition. The method may comprise producing
25 a sequence read for the first barcoded target nucleic acid molecule, wherein the sequence read comprises at least a portion of the sequence of the first partition barcode and at least a portion of the sequence of the first fragment of a target nucleic acid of the first partition. The method may comprise producing a sequence read for the second barcoded target nucleic acid molecule, wherein the sequence read comprises at least a portion of the first partition barcode and at least a
30 portion of the sequence of the second fragment of a target nucleic acid of the first partition.

- The method may further comprise sequencing at least a portion of each of the first and second barcoded target nucleic acid molecules of the second partition. The method may comprise
35 producing a sequence read for the first barcoded target nucleic acid molecule, wherein the sequence read comprises at least a portion of the sequence of the second partition barcode and at least a portion of the sequence of the first fragment of a target nucleic acid of the second partition. The method may comprise producing a sequence read for the second barcoded target nucleic acid molecule, wherein the sequence read comprises at least a portion of the sequence of

the second partition barcode and at least a portion of the sequence of the second fragment of a target nucleic acid of the second partition.

A sequence read may comprise at least 5, at least 10, at least 25, at least 50, at least 100, at least 250, at least 500, at least 1000, at least 2000, at least 5000, or at least 10,000 nucleotides from the target nucleic acid (e.g. genomic DNA). Preferably, each sequence read comprises at least 5 nucleotides from the target nucleic acid. By "at least a portion of the sequence" herein is meant at least 2, at least 3, at least 4, at least 5, at least 10, at least 25, at least 50, at least 100, at least 250, at least 500, at least 1000, at least 2000, or at least 5000 nucleotides of the relevant sequence. Preferably, by "at least a portion of the sequence" herein is meant at least 2 nucleotides of the relevant sequence.

The method may comprise a step of amplifying the signal from one or more barcoded affinity probes (i.e. a signal-amplification step or process). The signal-amplification process may comprise one or more strand-displacement amplification reactions and/or one or more multiple-displacement amplification reactions. The signal-amplification process may comprise an in vitro transcription reaction. The signal-amplification process may comprise a step of appending and/or binding and/or annealing (i.e. hybridising) one or more secondary barcoded oligonucleotides to said barcoded affinity probe, such as to a (non-secondary) barcoded oligonucleotide within a barcoded affinity probe. The signal-amplification process may comprise a step of appending and/or binding one or more secondary affinity moieties to said barcoded affinity probe (for example, binding a secondary antibody to a (non-secondary) antibody within a barcoded affinity probe). Optionally, any number of at least 2, at least 3, at least 5, or at least 10 secondary barcoded oligonucleotides and/or secondary affinity moieties may be appended and/or bound and/or annealed to any barcoded affinity probe. The method of appending and/or annealing and/or binding 2 or more secondary barcoded oligonucleotides and/or secondary affinity moieties to a barcoded affinity probe may be performed in separate sequential steps of appending and/or annealing and/or binding each thereof, or may be performed in a single parallel step.

A barcoded oligonucleotide, and/or a secondary barcoded oligonucleotide, may comprise a template for an in vitro transcription reaction. A barcoded oligonucleotide, and/or a secondary barcoded oligonucleotide, may contain a promoter region for an in vitro transcription reaction, such as a promoter for T7 RNA polymerase.

A barcoded oligonucleotide, and/or a secondary barcoded oligonucleotide, may comprise a circular (e.g. a circularised) oligonucleotide (such as a circular DNA oligonucleotide or a circular RNA oligonucleotide). A circular barcoded oligonucleotide may comprise one or more complementary primer oligonucleotides at least one nucleotide in length, wherein said complementary primer oligonucleotides is/are annealed to a sequence (or sequences) within said circular barcoded oligonucleotide(s). A circular barcoded oligonucleotide may be employed as a

- template for one or more strand-displacement amplification reactions and/or one or more multiple-displacement amplification reactions, such as reactions using a strand displacing polymerase, such as a phi29 DNA polymerase (optionally wherein one or more complementary primer oligonucleotides are employed as primers for such amplification reactions). A strand-displacement amplification reaction and/or a multiple-displacement amplification reaction may take place before and/or after and/or during any step of binding any one or more barcoded affinity probes to any target biomolecule from a sample. A product of any one or more said strand-displacement amplification reactions and/or one or more said multiple-displacement amplification reactions may comprise a target nucleic acid molecule for any method described herein. A product of any one or more said strand-displacement amplification reactions and/or one or more said multiple-displacement amplification reactions may be appended to any barcode sequence (such as any partition barcode sequence, any barcoded oligonucleotide, any barcode sequence and/or barcoded oligonucleotide comprised within any multimeric barcoding reagent).
- The method may (optionally as part of step (c)) comprise: (i) contacting the reaction mixture with a library comprising at least two multimeric barcoding reagents, wherein each multimeric barcoding reagent comprises first and second barcode regions linked together, wherein each barcode region comprises a nucleic acid sequence and wherein the first and second barcode regions of a first multimeric barcoding reagent are different to the first and second barcode regions of a second multimeric barcoding reagent of the library; and (ii) appending barcode sequences to each of a first fragment of a target nucleic acid and a second fragment of a target nucleic acid of the first microparticle to produce first and second barcoded target nucleic acid molecules for the first microparticle, wherein the first barcoded target nucleic acid molecule comprises the nucleic acid sequence of the first barcode region of the first multimeric barcoding reagent and the second barcoded target nucleic acid molecule comprises the nucleic acid sequence of the second barcode region of the first multimeric barcoding reagent, and appending barcode sequences to each of a first fragment of a target nucleic acid and a second fragment of a target nucleic acid of the second microparticle to produce first and second barcoded target nucleic acid molecules for the second microparticle, wherein the first barcoded target nucleic acid molecule comprises the nucleic acid sequence of the first barcode region of the second multimeric barcoding reagent and the second barcoded target nucleic acid molecule comprises the nucleic acid sequence of the second barcode region of the second multimeric barcoding reagent.
- The first fragment of a target nucleic acid of the first microparticle may be the barcoded oligonucleotide of the at least one barcoded biomolecule complex of the first circulating microparticle, and wherein the first fragment of a target nucleic acid of the second microparticle may be the barcoded oligonucleotide of the at least one barcoded biomolecule complex of the second circulating microparticle.

The reaction mixture may further comprise a fragment of a target nucleic acid of the first circulating microparticle and wherein the second fragment of a target nucleic acid of the first circulating microparticle is the fragment of the target nucleic acid of the first circulating microparticle.

5

The reaction mixture may further comprise a fragment of a target nucleic acid of the second circulating microparticle and wherein the second fragment of a target nucleic acid of the second circulating microparticle is the fragment of the target nucleic acid of the second circulating microparticle.

10

The step of contacting the reaction mixture with a library of multimeric barcoding reagents may be performed in a single contiguous aqueous volume. Step (c) may be performed in a single contiguous aqueous volume, optionally wherein steps (b) and (c) are performed in a single contiguous aqueous volume, optionally wherein steps (a), (b) and (c) are performed in a single contiguous aqueous volume.

15

The method may further comprise analysing a sequence of each of the first and second barcoded target nucleic acid molecules of the first circulating microparticle, and analysing a sequence of each of the first and second barcoded target nucleic acid molecules of the second circulating microparticle. Optionally, the step of analysing a sequence is performed by sequencing at least a portion of each of the first and second barcoded target nucleic acid molecules.

20

The method may further comprise sequencing at least a portion of each of the first and second barcoded target nucleic acid molecules of the first circulating microparticle. The method may comprise producing a sequence read for the first barcoded target nucleic acid molecule, wherein the sequence read comprises at least a portion of the sequence of the first barcode region of the first multimeric barcoding reagent and at least a portion of the sequence of the first fragment of a target nucleic acid of the first circulating microparticle. The method may comprise producing a sequence read for the second barcoded target nucleic acid molecule, wherein the sequence read comprises at least a portion of the sequence of the second barcode region of the first multimeric barcoding reagent and at least a portion of the sequence of the second fragment of a target nucleic acid of the first circulating microparticle.

25

30

The method may further comprise sequencing at least a portion of each of the first and second barcoded target nucleic acid molecules of the second circulating microparticle. The method may comprise producing a sequence read for the first barcoded target nucleic acid molecule, wherein the sequence read comprises at least a portion of the sequence of the first barcode region of the second multimeric barcoding reagent and at least a portion of the sequence of the first fragment of a target nucleic acid of the second circulating microparticle. The method may comprise

35

producing a sequence read for the second barcoded target nucleic acid molecule, wherein the sequence read comprises at least a portion of the sequence of the second barcode region of the second multimeric barcoding reagent and at least a portion of the sequence of the second fragment of a target nucleic acid of the second circulating microparticle.

5

A sequence read may comprise at least 5, at least 10, at least 25, at least 50, at least 100, at least 250, at least 500, at least 1000, at least 2000, at least 5000, or at least 10,000 nucleotides from the target nucleic acid (e.g. genomic DNA). Preferably, each sequence read comprises at least 5 nucleotides from the target nucleic acid. By "at least a portion of the sequence" herein is meant at least 2, at least 3, at least 4, at least 5, at least 10, at least 25, at least 50, at least 100, at least 250, at least 500, at least 1000, at least 2000, or at least 5000 nucleotides of the relevant sequence. Preferably, by "at least a portion of the sequence" herein is meant at least 2 nucleotides of the relevant sequence.

10

15 The method may further comprise partitioning the sample or reaction mixture into at least first and second partitions and analysing the nucleotide sequences of the barcoded oligonucleotides in each of the first and second partitions, wherein the first partition comprises at least one barcoded oligonucleotide comprised in or derived from the at least one barcoded biomolecule complex of the first circulating microparticle, and wherein the second partition comprises at least one
20 barcoded oligonucleotide comprised in or derived from the at least one barcoded biomolecule complex of the second circulating microparticle. The step of partitioning may be performed prior to step (a), prior to step (b) and/or prior to step (c).

25

The method may comprise partitioning the sample into at least 3, at least 4, at least 5, at least 10, at least 100, at least 1000, at least 10,000, at least 100,000, at least 1,000,000, at least 10,000,000, at least 100,000,000, or at least 1,000,000,000 partitions. Preferably, method comprises partitioning the sample into at least 1000 partitions.

30

The step of analysing the nucleotide sequences of the barcoded oligonucleotides of the barcoded biomolecule complexes may comprise: (i) appending a first partition barcode sequence to the at least one barcoded oligonucleotide of the first partition; and (ii) appending a second partition barcode sequence to the at least one barcoded oligonucleotide of the second partition.

The first and second partition barcode sequences may be different.

35

The first partition barcode sequence may be from a first set of partition barcode sequences, and the second partition barcode sequence may be from a second set of partition barcode sequences, and wherein the first and second sets of partition barcode sequences are different.

The first partition barcode sequence may be the nucleic acid sequence of a barcode region of a first multimeric barcoding reagent, and the second partition barcode sequence may be the nucleic acid sequence of a barcode region of a second multimeric barcoding reagent, and wherein the first and second multimeric barcoding reagents each comprise two or more barcode regions
5 linked together.

The first partition may further comprise a fragment of a target nucleic acid of the first circulating microparticle, and wherein the second partition may further comprise a fragment of a target nucleic acid of the second circulating microparticle.

10 The step of analysing the nucleotide sequences of the barcoded oligonucleotides of the barcoded biomolecule complexes may comprise: (i) appending a first partition barcode sequence to at least one barcoded oligonucleotide of the first partition and appending the first partition barcode sequence to at least one fragment of a target nucleic acid of the first circulating microparticle; (ii)
15 appending a second partition barcode sequence to at least one barcoded oligonucleotide of the second partition and appending the second partition barcode sequence to at least one fragment of a target nucleic acid of the second circulating microparticle; and wherein said first and second partition barcode sequences are different.

20 The step of analysing the nucleotide sequences of the barcoded oligonucleotides of the barcoded biomolecule complexes may comprise: (i) appending a first partition barcode sequence of a first set of partition barcode sequences to at least one barcoded oligonucleotide of the first partition and appending a second partition barcode sequence of the first set of partition barcode sequences to at least one fragment of a target nucleic acid of the first circulating microparticle;
25 and (ii) appending a first partition barcode sequence of a second set of partition barcode sequences to at least one barcoded oligonucleotide of the second partition and appending a second partition barcode sequence of the second set of partition barcode sequences to at least one fragment of a target nucleic acid of the second circulating microparticle; and wherein the first and second sets of partition barcode sequences are different.

30 The first and second partition barcode sequences of the first set of partition barcode sequences may be the nucleic acid sequences of first and second barcode regions of a first multimeric barcoding reagent, and wherein the first and second partition barcode sequences of the second set of partition barcode sequences may be the nucleic acid sequences of first and second
35 barcode regions of a second multimeric barcoding reagent, and wherein the first and second multimeric barcoding reagents each comprise two or more barcode regions linked together.

The first partition may further comprise a fragment of a target nucleic acid and wherein the second partition may further comprise a fragment of a target nucleic acid, and wherein the step of

analysing the nucleotide sequences of the barcoded oligonucleotides of the barcoded biomolecule complexes comprises: (i) appending a first partition barcode sequence to the at least one barcoded oligonucleotide of the first partition and appending the first partition barcode sequence to at least one fragment of a target nucleic acid of the first partition; (ii) appending a second partition barcode sequence to the at least one barcoded oligonucleotide of the second partition and appending the second partition barcode sequence to at least one fragment of a target nucleic acid of the second partition; wherein said first and second partition barcode sequences are different. Alternatively, the step of analysing the nucleotide sequences of the barcoded oligonucleotides of the barcoded biomolecule complexes comprises: (i) appending a first partition barcode sequence of a first set of partition barcode sequences to the at least one barcoded oligonucleotide of the first partition and appending a second partition barcode sequence of the first set of partition barcode sequences to at least one fragment of a target nucleic acid of the first partition; (ii) appending a first partition barcode sequence of a second set of partition barcode sequences to the at least one barcoded oligonucleotide of the second partition and appending a second partition barcode sequence of the second set of partition barcode sequences to at least one fragment of a target nucleic acid of the second partition; wherein the first and second sets of partition barcode sequences are different.

The first and second partition barcode sequences of the first set of partition barcode sequences may be the nucleic acid sequences of first and second barcode regions of a first multimeric barcoding reagent, and wherein the first and second partition barcode sequences of the second set of partition barcode sequences may be the nucleic acid sequences of first and second barcode regions of a second multimeric barcoding reagent, and wherein the first and second multimeric barcoding reagents each comprise two or more barcode regions linked together.

The invention provides the use of a barcoded affinity probe to determine the presence, absence and/or level of a target biomolecule in a circulating microparticle or in a sample derived therefrom, wherein the barcoded affinity probe comprises at least one affinity moiety linked to a barcoded oligonucleotide, wherein the barcoded oligonucleotide comprises at least one nucleotide and wherein the affinity moiety is capable of binding to the target biomolecule.

The invention provides a barcoded affinity probe for determining the presence, absence and/or level of a target biomolecule, wherein the barcoded affinity probe comprises at least one affinity moiety linked to a barcoded oligonucleotide, wherein the barcoded oligonucleotide comprises at least one nucleotide and wherein the affinity moiety is capable of binding to the target biomolecule.

- 35 -

The barcoded affinity probe, target biomolecule, affinity moiety and barcoded oligonucleotide may take any of the forms described herein. In particular, they may take any of the forms described herein in relation to the methods.

- 5 The invention provides a library of barcoded affinities probes for determining the presence, absence and/or level of at least two target biomolecules, wherein the library comprises: (i) a first barcoded affinity probe comprising at least one affinity moiety linked to a barcoded oligonucleotide, wherein the barcoded oligonucleotide comprises at least one nucleotide and wherein the affinity moiety is capable of binding to a first target biomolecule; and (ii) a second
10 barcoded affinity probe comprising at least one affinity moiety linked to a barcoded oligonucleotide, wherein the barcoded oligonucleotide comprises at least one nucleotide and wherein the affinity moiety is capable of binding to a second target biomolecule; and wherein the first target biomolecule and the second target biomolecule are different.

- 15 The library of barcoded affinities probes, barcoded affinity probes, target biomolecules, affinity moieties and barcoded oligonucleotides may take any of the forms described herein. In particular, they may take any of the forms described herein in relation to the methods.

- The first target biomolecule may be a polypeptide and the second target biomolecule may be a
20 barcoded oligonucleotide or a fragment of a target nucleic acid (e.g. genomic DNA).

- The first target biomolecule may be a polypeptide and the second target biomolecule may be a fragment of a target nucleic acid (e.g. genomic DNA) comprising an epigenetic modification (e.g. 5-hydroxy-methylcytosine DNA or 5-methylcytosine DNA).
25

- The first target biomolecule may be 5-hydroxy-methylcytosine DNA and the second target biomolecule may be a biomolecule selected from Biomolecule group 1.

- The first target biomolecule may be 5-methylcytosine DNA and the second target biomolecule
30 may be a biomolecule selected from Biomolecule group 1.

The first and second target biomolecules may be selected from Biomolecule group 1.

- 35 Optionally, any library of two or more barcoded affinity probes may comprise a single, mixed solution comprising said two or more barcoded affinity probes. Optionally, any library of two or more barcoded affinity probes may comprise two or more separate solutions, wherein each solution comprises a solution of one of said two or more barcoded affinity probes. Optionally, any library of two or more barcoded affinity probes may be provided in the form of a kit, wherein said

kit is comprised of two or more separate solutions, wherein each solution comprises a solution one of said two or more barcoded affinity probes.

- 5 The sample may be contacted with a library of at least 2, at least 3, at least 5, at least 10, at least 20, or at least 30 different barcoded affinity probes. Preferably, the library comprises at least 2 different barcoded affinity probes. Each of the barcoded affinity probes may comprise at least one affinity moiety linked to a barcoded oligonucleotide, wherein the barcoded oligonucleotide comprises at least one nucleotide, and wherein the affinity moiety is capable of binding to a target biomolecule. The affinity moiety of each of the different barcoded affinity probes in the library
10 may be capable of binding to a different target biomolecule. The library of barcoded affinity probes may be capable of binding at least 2, at least 3, at least 5, at least 10, at least 20, or at least 30 different target biomolecules. Preferably, the library of barcoded affinity probes is capable of binding at least 2 different target biomolecules.
- 15 Optionally, in any library of two or more barcoded affinity probes, barcoded affinity probes comprising the same affinity moiety (and/or comprising affinity moieties capable of binding to the same target biomolecule) may comprise identical barcoded oligonucleotides. Optionally, in any library of barcoded affinity probes, barcoded affinity probes comprising the same affinity moiety (and/or comprising affinity moieties with affinity for the same target biomolecule) may comprise
20 different barcoded oligonucleotides or different barcode sequences from a set of two or more different barcode sequences, and/or from a set of at least 10 different barcode sequences, and/or from a set of at least 100 different barcode sequences, and/or from a set of at least 1000 different barcode sequences, and/or from a set of at least 10,000 different barcode sequences, and/or from a set of at least 1,000,000 different barcode sequences.
- 25 Optionally, in any library of two or more different barcoded affinity probes, each barcoded affinity probe may comprise a set of two or more different affinity moieties (for example, each barcoded affinity probe may comprise two or more different affinity moieties, each capable of binding to a different target biomolecule). Optionally, in any library of barcoded affinity probes, barcoded
30 affinity probes comprising the same set of two or more different affinity moieties (and/or comprising sets of affinity moieties capable of binding to the same target biomolecule(s)) may comprise identical barcoded oligonucleotides. Optionally, in any library of barcoded affinity probes, barcoded affinity probes comprising the same set of two or more different affinity moieties (and/or comprising sets of affinity moieties capable of binding to the same target biomolecule(s))
35 may comprise different barcode sequences or different barcode sequences from a set of two or more different barcode sequences, and/or from a set of at least 10 different barcode sequences, and/or from a set of at least 100 different barcode sequences, and/or from a set of at least 1000 different barcode sequences, and/or from a set of at least 10,000 different barcode sequences, and/or from a set of at least 1,000,000 different barcode sequences.

A library of two or more different barcoded affinity probes may comprise barcoded affinity probes each comprising one or more affinity moieties, and one or more primary barcoded oligonucleotides, and one or more secondary barcoded oligonucleotides, wherein each primary
5 barcoded oligonucleotide in said library comprises an identical sequence, and wherein each secondary barcoded oligonucleotides in said library comprises a different sequence.

There is provided an optically-labelled and/or fluorescently-labelled affinity probe, wherein said optically-labelled and/or fluorescently-labelled affinity probe comprises at least one affinity moiety
10 with affinity and/or specificity for any one or more biomolecules (or target biomolecules) selected from Biomolecule group 1. There is provided an optically-labelled and/or fluorescently-labelled affinity probe, wherein said optically-labelled and/or fluorescently-labelled affinity probe comprises at least one affinity moiety with affinity and/or specificity for any one or more biomolecules (or target biomolecules) selected from Biomolecule group 1, and comprises at least one optical
15 and/or fluorescent label.

There is provided a library of two or more optically-labelled and/or fluorescently-labelled affinity probes, comprising at least a first and a second affinity probe for at least first and second biomolecules (or target biomolecules) selected from Biomolecule group 1, wherein each optically-
20 labelled and/or fluorescently-labelled affinity probe comprises at least one optical and/or fluorescent label. There is provided a library of two or more optically-labelled and/or fluorescently-labelled affinity probes, comprising a first optically-labelled and/or fluorescently-labelled affinity probe with affinity and/or specificity for 5-methylcytosine DNA or for 5-hydroxy-methylcytosine DNA, and at least a second optically-labelled and/or fluorescently-labelled affinity probe with
25 affinity and/or specificity for any one or more biomolecules (or target biomolecules) selected from Biomolecule group 1.

There is provided one or more oligonucleotides, wherein said oligonucleotide comprises a sequence identical to and/or complementary to any of the DNA and/or RNA sequences of any of
30 the biomolecules of Biomolecule group 1. There is provided one or more primers, wherein said primer comprises a sequence identical to and/or complementary to any of the DNA and/or RNA sequences of any of the biomolecules of Biomolecule group 1. There is provided one or more oligonucleotide probes for an in situ hybridisation (ISH) process, wherein said oligonucleotide probe comprises a sequence identical to and/or complementary to any of the DNA and/or RNA
35 sequences of any of the biomolecules of Biomolecule group 1. There is provided one or more oligonucleotide probes for a fluorescence in situ hybridisation (FISH) process, wherein said oligonucleotide probe comprises a sequence identical to and/or complementary to any of the DNA and/or RNA sequences of any of the biomolecules of Biomolecule group 1. Optionally, any said oligonucleotide, and/or primer, and/or oligonucleotide probe may comprise an optical and/or

fluorescent label. Optionally, any said oligonucleotide, and/or primer, and/or oligonucleotide probe may comprise an adapter sequence and/or a coupling sequence. Optionally, any said oligonucleotide, and/or primer, and/or oligonucleotide probe may be employed in a reverse transcription process, and/or a primer-extension process; and/or a PCR process, and/or an in situ hybridisation (ISH) process, and/or a fluorescence in situ hybridisation (FISH) process. There is provided a library of two or more oligonucleotides, wherein each said oligonucleotide comprises a sequence identical to and/or complementary to any of the DNA and/or RNA sequences of any of the biomolecules of Biomolecule group 1.

10 In the method, the circulating microparticle may contain at least two fragments of a target nucleic acid, and wherein the method comprises: (a) preparing the sample for sequencing comprising linking at least two of the at least two fragments of the target nucleic acid to produce a set of at least two linked fragments of the target nucleic acid; and (b) sequencing at least two of the linked fragments in the set to produce at least two (informatically) linked sequence reads.

15 In the method, the circulating microparticle may contain at least two fragments of a target nucleic acid, and wherein the method comprises: (a) preparing the sample for sequencing comprising linking at least two of the at least two fragments of the target nucleic acid to produce a set of at least two linked fragments of the target nucleic acid; and (b) sequencing at least two of the linked fragments in the set to produce at least two (informatically) linked sequence reads.

20 In the method, the circulating microparticle contain at least two fragments of genomic DNA, and wherein the method comprises: (a) preparing the sample for sequencing comprising linking at least two of the at least two fragments of genomic DNA to produce a set of at least two linked fragments of genomic DNA; and (b) sequencing at least two of the linked fragments in the set to produce at least two linked sequence reads.

25 In the method, the circulating microparticle may contain at least two fragments of genomic DNA, and wherein the method comprises: (a) preparing the sample for sequencing comprising linking at least two of the at least two fragments of genomic DNA to produce a set of at least two linked fragments of genomic DNA; and (b) sequencing at least two of the linked fragments in the set to produce at least two linked sequence reads.

30 In the methods, at least 3, at least 4, at least 5, at least 10, at least 50, at least 100, at least 500, at least 1000, at least 5000, at least 10,000, at least 100,000, or at least 1,000,000 fragments of the target nucleic acid of the microparticle may be linked as a set and then sequenced to produce at least 3, at least 4, at least 5, at least 10, at least 50, at least 100, at least 500, at least 1000, at least 5000, at least 10,000, at least 100,000, or at least 1,000,000 linked sequence reads.

Preferably, at least 5 fragments of the target nucleic acid of the microparticle may be linked as a set and then sequenced to produce at least 5 linked sequence reads.

5 In the methods, each of the linked sequence reads may provide the sequence of at least 1 nucleotide, at least 5 nucleotides, at least 10 nucleotides, at least 20 nucleotides, at least 30 nucleotides, at least 50 nucleotides, at least 100 nucleotides, at least 200 nucleotides, at least 500 nucleotides, at least 1000 nucleotides, or at least 10,000 nucleotides of a linked fragment. Preferably, each of the linked sequence reads may provide the sequence of at least 20
10 nucleotides of a linked fragment.

10 In the methods, a total of at least 2, at least 10, at least 100, at least 1000, at least 10,000, at least 100,000, at least 1,000,000, at least 10,000,000, at least 100,000,000, at least 1,000,000,000, at least 10,000,000,000, at least 100,000,000,000, or at least 1,000,000,000,000 sequence reads may be produced. Preferably, a total of at least 500,000 sequence reads are
15 produced.

A sequence read may comprise at least 5, at least 10, at least 25, at least 50, at least 100, at least 250, at least 500, at least 1000, at least 2000, at least 5000, or at least 10,000 nucleotides from the target nucleic acid (e.g. genomic DNA). Preferably, each sequence read comprises at
20 least 5 nucleotides from the target nucleic acid.

A sequence read may comprise a raw sequence read, or portion thereof, generated from a sequencing instrument e.g. a 50-nucleotide long sequence raw sequence read generated from an Illumina sequence instrument. A sequence read may comprise a merged sequence from both
25 reads of a paired-end sequencing run e.g. concatenated or merged sequences from both a first and second read of a paired-end sequencing run on an Illumina sequencing instrument. A sequence read may comprise a portion of a raw sequence read generated from a sequencing instrument e.g. 20 contiguous nucleotides within a raw sequence read of 150 nucleotides generated by an Illumina sequencing instrument. A single raw sequence read may comprise the
30 at least two linked sequence reads produced by the methods of the invention.

Sequence reads may be produced by any method known in the art. For example, by chain-termination or Sanger sequencing. Preferably, sequencing is performed by a next-generation sequencing method such as sequencing by synthesis, sequencing by synthesis using reversible
35 terminators (e.g. Illumina sequencing), pyrosequencing (e.g. 454 sequencing), sequencing by ligation (e.g. SOLiD sequencing), single-molecule sequencing (e.g. Single Molecule, Real-Time (SMRT) sequencing, Pacific Biosciences), or by nanopore sequencing (e.g. on the Minion or Promethion platforms, Oxford Nanopore Technologies). Most preferably, sequence reads are produced by sequencing by synthesis using reversible terminators (e.g. Illumina sequencing).

The methods may comprise a further step of mapping each of the linked sequence reads to a reference genomic sequence. The linked sequence reads may comprise sequences mapped to the same chromosome of the reference genomic sequence or sequences mapped to two or more
 5 different chromosomes of the reference genomic sequence.

The microparticle may have a diameter of at least 100 nm, at least 110 nm, at least 125 nm, at least 150 nm, at least 175 nm, at least 200 nm, at least 250 nm or at least 500 nm. Preferably, the microparticle has a diameter of at least 200 nm. The diameter of the microparticle may be
 10 100-5000 nm. The diameter of the microparticle may be 10-10,000 nm (e.g. 100-10,000 nm, 110-10,000 nm), 50-5000 nm, 75-5,000 nm, 100-3,000 nm. The diameter of the microparticle may be 10-90 nm, 50-100 nm, 90-200 nm, 100-200 nm, 100-500 nm, 100-1000 nm, 1000-2000 nm, 90-5000 nm, or 2000-10,000 nm. Preferably, the microparticle diameter is between 100 and 5000 nm. Most preferably, the microparticle has a diameter that is between 200 and 5000 nm. The
 15 sample may include microparticles of at least two different sizes, or at least three different sizes, or a range of different sizes.

The linked fragments of genomic DNA may originate from a single genomic DNA molecule.

20 The methods may further comprise the step of estimating or determining the genomic sequence length of the linked fragments of genomic DNA. Optionally, this step may be performed by sequencing substantially an entire sequence of a linked fragment (i.e. from its approximate 5' end to its approximate 3' end) and counting the number of nucleotides sequenced therein. Optionally, this may be performed by sequencing a sufficient number of nucleotides at the 5' end of the
 25 sequence of the linked fragment to map said 5' end to a locus within a reference genome sequence (e.g. human genome sequence), and likewise sequencing a sufficient number of nucleotides at the 3' end of the linked fragment to map said 3' end to a locus within the reference genome sequence, and then determining the genomic sequence length of the linked fragment using the reference genome sequence (i.e. the number of nucleotides sequenced at the 3' end of
 30 the linked fragment + the number of nucleotides sequenced at the 5' end of the linked fragment + the number of nucleotides between these sequences in the reference genome (i.e. the unsequenced portion)).

35 In the methods, the sample may comprise first and second circulating microparticles, wherein each microparticle contains at least two fragments of a target nucleic acid (e.g. genomic DNA), and wherein the method comprises performing step (a) to produce a first set of linked fragments of the target nucleic acid for the first microparticle and a second set of linked fragments of the target nucleic acid for the second microparticle, and performing step (b) to produce a first set of

linked sequence reads (i.e. set of linked signals) for the first microparticle and a second set of linked sequence reads (i.e. set of linked signals) for the second microparticle.

5 In the methods, the set of linked sequence reads (i.e. set of linked signals) produced for the first microparticle may be distinguishable from the set of linked sequence reads (i.e. set of linked signals) produced for the second microparticle.

10 In the methods, the sample may comprise n microparticles originating from blood, wherein each microparticle contains at least two fragments of a target nucleic acid (e.g. genomic DNA), and wherein the method comprises performing step (a) to produce n sets of linked fragments of the target nucleic acid, one set for each of the n microparticles, and performing step (b) to produce n sets of linked sequence reads (i.e. sets linked signals), one for each of the n microparticles.

15 In the methods, n may be at least 3, at least 5, at least 10, at least 50, at least 100, at least 1000, at least 10,000, at least 100,000, at least 1,000,000, at least 10,000,000, at least 100,000,000, at least 1,000,000,000, at least 10,000,000,000, or at least 100,000,000,000. Preferably, n is at least 100,000 microparticles.

20 In the methods, the sample may comprise at least 3, at least 5, at least 10, at least 50, at least 100, at least 1000, at least 10,000, at least 100,000, at least 1,000,000, at least 10,000,000, at least 100,000,000, at least 1,000,000,000, at least 10,000,000,000, or at least 100,000,000,000 microparticles (and/or a sample derived from at least 3, at least 5, at least 10, at least 50, at least 100, at least 1000, at least 10,000, at least 100,000, at least 1,000,000, at least 10,000,000, at least 100,000,000, at least 1,000,000,000, at least 10,000,000,000, or at least 100,000,000,000 microparticles), wherein said microparticles (and/or a sample derived therefrom) are/is comprised within a single contiguous aqueous volume during any step of the method, such as any step of contacting the sample with a library of multimeric barcoding reagents, and/or any step of appending and/or linking and/or connecting barcode sequences (such as barcoded oligonucleotides) to target nucleic acids, and/or any step of appending coupling sequences to target nucleic acids, and/or any step of appending and/or linking and/or connecting coupling molecules to target nucleic acids or other target biomolecules, and/or any step of crosslinking or permeabilising.

35 The set of linked sequence reads (i.e. set of linked signals) produced for each microparticle may be distinguishable from the sets of linked sequence reads produced for the other microparticles.

The methods may further comprise, prior to step (a), the step of partitioning the sample into at least two different reaction volumes.

In the present invention, two sequences or sequence reads (e.g. as determined by a sequencing reaction) may be linked informatically by any means that allows such sequences to be related or interrelated to each other in any way, within a computer system, within an algorithm, or within a dataset. Such linking may be comprised of, and/or established by, and/or represented by a discrete identifying link, or by a shared property, or by any indirect method linking, interrelating, or correlating two or more such sequences.

The linking may be comprised of, and/or established by, and/or represented by a sequence within a sequencing reaction itself (e.g. in the form of a barcode sequence determined through the sequencing reaction, or in the form of two different parts or segments of a single determined sequence which together comprise a first and a second linked sequence), or established, comprised, or represented independent of such sequences (such as established by merit of being comprised within the same flowcell, or within the same lane of a flowcell, or within the same compartment or region of a sequencing instrument, or comprised within the same sequencing run of a sequencing instrument, or comprised with a degree of spatial proximity within a biological sample, and/or with a degree of spatial proximity within a sequencing instrument or sequencing flowcell. Linking may be comprised of, and/or established by, and/or represented by a measure or parameter corresponding to a physical location or partition within a sequencing instrument, such as a pixel or pixel location within an image and/or within a multi-pixel camera or a multi-pixel charge-coupled device, and/or such as a nanopore or location of a nanopore within a nanopore sequencing instrument or nanopore membrane.

Linking may be absolute (i.e., two sequences are either linked or unlinked, with no quantitative, semi-quantitative, or qualitative/categorical relationships outside of this). Linking may also be relative, probabilistic, or established, comprised, or represented in terms of a degree, a probability, or an extent of linking, for example relative to (or represented by) one or more parameters that may hold one of a series of quantitative, semi-quantitative, or qualitative/categorical values. For example, two (or more) sequences may be linked informatically by a quantitative, semi-quantitative, or qualitative/categorical parameter, which represents, comprises, estimates, or embodies the proximity of said two (or more) sequences within a sequencing instrument, or the proximity of said two (or more) sequences within a biological sample.

For any analysis involving two or more sequences that are linked informatically by any such way, the existence (or lack thereof) of linking may be employed as a parameter in any analysis or evaluation step or any algorithm for performing same. For any analysis involving two or more sequences that are linked informatically by any such way, the degree, probability, or extent of linking may be employed as a parameter in any analysis or evaluation step or any algorithm for performing same.

In one version of such linking, a given set of two or more linked sequences may be associated with a specific identifier, such as an alphanumeric identifier, or a barcode, or a barcode sequence. In one further version a given set of two or more linked sequences may be associated with or a
5 barcode, or a barcode sequence, wherein said barcode or barcode sequence is comprised within a sequence determined by the sequencing reaction. For example, each sequence determined in a sequencing reaction may comprise both a barcode sequence and a sequence corresponding to a genomic DNA sequence. Optionally, certain sequences or linked sequences may be represented by or associated with two or more barcodes or identifiers.

10 In another version of linking, two or more linked sequences may be kept within discrete partitions within a computer, or computer network, within a hard drive, or any sort of storage medium, or any other means of storing sequence data. Optionally, certain sequences or linked sequences may be kept in two or more partitions within such a computer or data medium.

15 Sequences that are linked informatically may comprise one or more sets of informatically linked sequences. Sequences in a linked set of sequences may all share the same linking function or representation thereof; for example, all sequences within a linked set may be associated with the same barcode or with the same identifier, or may be comprised within the same partition within a
20 computer or storage medium; all sequences may share any other form of linking, interrelation, and/or correlation. One or more sequences in a linked set may be exclusive members of said set, and thus not members of any other set. Alternatively, one or more sequences in a linked set may be non-exclusive members of said set, and thus said sequences may be represented by and/or associated with two or more different linked sets of sequences.

25 The invention provides a method of analysing a sample comprising at least two circulating microparticles or a sample derived from at least two circulating microparticles, wherein the method comprises: (i) partitioning the sample into at least two partitions, wherein each partition comprises, on average, less than n circulating microparticles; and (ii) determining the presence,
30 absence and/or level of at least two target biomolecules in each of at least two of the at least two partitions. Optionally, wherein n is 1000, 500, 200, 100, 50, 40, 30, 20, 10, 5, 4, 3, 2, 1, 0.5, 0.4, 0.3, 0.2, 0.1, 0.05, 0.04, 0.03, 0.02, 0.01, 0.005, 0.001, 0.0005, or 0.0001. Preferably, wherein n is 0.5. Optionally, wherein step (i) comprises partitioning the sample into at least 3 partitions, at least 5 partitions, at least 10 partitions, at least 100 partitions, at least 1000 partitions, at least
35 10,000 partitions, at least 100,000 partitions, at least 1,000,000 partitions, at least 10,000,000 partitions, at least 100,000,000 partitions, or at least 1,000,000,000 partitions. Preferably, wherein step (i) comprises partitioning the sample into at least 1000 partitions.

- 44 -

The step of (ii) determining the presence, absence and/or level of at least two target biomolecules may be performed for each of at least two of the at least two partitions by a method of analysing a sample (i.e. the sample in the partition) comprising a circulating microparticle or a sample derived from a circulating microparticle, wherein the circulating microparticle comprises at least two target molecules, wherein the at least two target molecules are biomolecules, and wherein the method comprises measuring a signal corresponding to the presence, absence and/or level of each of the target molecules to produce a set of at least two (informatically) linked signals for the circulating microparticle (i.e. a set of at least two (informatically) linked signals for the partition), wherein at least one of the linked signals corresponds to the presence, absence and/or level of a first biomolecule in the sample (i.e. the sample in the partition) and at least one of the linked signals corresponds to the presence, absence and/or level of a second biomolecule in the sample (i.e. the sample in the partition). The method may be performed by any of the methods provided herein that comprise producing a set of at least two linked signals of a microparticle. The method may produce a set of at least two linked signals for each of at least two of the at least two partitions.

The invention provides a method of analysing a sample comprising at least two circulating microparticles or a sample derived from at least two circulating microparticles, wherein the method comprises: (i) partitioning the sample into at least two partitions, wherein a first partition comprises at least first and second target biomolecules of a first circulating microparticle and a second partition comprises at least first and second target biomolecules of a second circulating microparticle, and wherein each partition of at least two of the at least two partitions comprises, on average, less than [X] total mass of DNA; and (ii) determining the presence, absence and/or level of at least two target biomolecules in each of at least two of the at least two partitions. Optionally, wherein [X] is 1.0 attogram of DNA, 10 attograms of DNA, 100 attograms of DNA, 1.0 femtogram of DNA, 10 femtograms of DNA, 100 femtograms of DNA, 1.0 picogram of DNA, 10 picograms of DNA, 100 picograms of DNA, or 1.0 nanogram of DNA. Preferably, wherein [X] is 100 femtograms of DNA.

The step of (ii) determining the presence, absence and/or level of at least two target biomolecules may be performed for each of at least two of the at least two partitions by a method of analysing a sample (i.e. the sample in the partition) comprising a circulating microparticle or a sample derived from a circulating microparticle, wherein the circulating microparticle comprises at least two target molecules, wherein the at least two target molecules are biomolecules, and wherein the method comprises measuring a signal corresponding to the presence, absence and/or level of each of the target molecules to produce a set of at least two (informatically) linked signals for the circulating microparticle (i.e. a set of at least two (informatically) linked signals for the partition), wherein at least one of the linked signals corresponds to the presence, absence and/or level of a first biomolecule in the sample (i.e. the sample in the partition) and at least one of the linked signals

corresponds to the presence, absence and/or level of a second biomolecule in the sample (i.e. the sample in the partition). The method may be performed by any of the methods provided herein that comprise producing a set of at least two linked signals of a microparticle. The method may produce a set of at least two linked signals for each of at least two of the at least two
5 partitions.

The invention provides a method of analysing a sample comprising at least two circulating microparticles or a sample derived from at least two circulating microparticles, wherein the method comprises: (i) partitioning the sample into at least two partitions, wherein a first partition
10 comprises at least first and second target biomolecules of a first circulating microparticle and a second partition comprises at least first and second target biomolecules of a second circulating microparticle, and wherein each partition of at least two of the at least two partitions comprises, on average, less than [Y] total mass of polypeptide; and (ii) determining the presence, absence and/or level of at least two target biomolecules in each of at least two of the at least two partitions.
15 Optionally, wherein [Y] is 1.0 attogram of polypeptide, 10 attograms of polypeptide, 100 attograms of polypeptide, 1.0 femtogram of polypeptide, 10 femtograms of polypeptide, 100 femtograms of polypeptide, 1.0 picogram of polypeptide, 10 picograms of polypeptide, 100 picograms of polypeptide, or 1.0 nanogram of polypeptide. Preferably, wherein [Y] is 100 femtograms of polypeptide.

20 The step of (ii) determining the presence, absence and/or level of at least two target biomolecules may be performed for each of at least two of the at least two partitions by a method of analysing a sample (i.e. the sample in the partition) comprising a circulating microparticle or a sample derived from a circulating microparticle, wherein the circulating microparticle comprises at least two target
25 molecules, wherein the at least two target molecules are biomolecules, and wherein the method comprises measuring a signal corresponding to the presence, absence and/or level of each of the target molecules to produce a set of at least two (informatically) linked signals for the circulating microparticle (i.e. a set of at least two (informatically) linked signals for the partition), wherein at least one of the linked signals corresponds to the presence, absence and/or level of a first
30 biomolecule in the sample (i.e. the sample in the partition) and at least one of the linked signals corresponds to the presence, absence and/or level of a second biomolecule in the sample (i.e. the sample in the partition). The method may be performed by any of the methods provided herein that comprise producing a set of at least two linked signals of a microparticle. The method may produce a set of at least two linked signals for each of at least two of the at least two
35 partitions.

The method may further comprise analysing the sequence of at least two target nucleic acid molecules that have been partitioned into each of said first and second partitions.

The method may comprise partitioning the sample into at least 3, at least 4, at least 5, at least 10, at least 100, at least 1000, at least 10,000, at least 100,000, at least 1,000,000, at least 10,000,000, at least 100,000,000, or at least 1,000,000,000 partitions. Preferably, the method comprises partitioning the sample into at least 1000 partitions.

5

The first target biomolecule may be a polypeptide and the second target biomolecule may be a barcoded oligonucleotide or a fragment of a target nucleic acid (e.g. genomic DNA).

10

The first target biomolecule may be a polypeptide and the second target biomolecule may be a fragment of a target nucleic acid (e.g. genomic DNA) comprising an epigenetic modification (e.g. 5-hydroxy-methylcytosine DNA or 5-methylcytosine DNA).

15

The first target biomolecule may be 5-hydroxy-methylcytosine DNA and the second target biomolecule may be a biomolecule selected from Biomolecule group 1.

The first target biomolecule may be 5-methylcytosine DNA and the second target biomolecule may be a biomolecule selected from Biomolecule group 1.

20

The first and second target biomolecules may be selected from Biomolecule group 1.

25

Any one or more steps of determining (or measuring) the presence, absence and/or level of a target biomolecule (or measuring a signal corresponding to the presence, absence and/or level of a target biomolecule) may be performed using one or more barcoded affinity probes (as provided herein) e.g. by binding a barcoded affinity probe to the target biomolecule. Any one or more steps of determining (or measuring) the presence, absence and/or level of a target biomolecule (or measuring a signal corresponding to the presence, absence and/or level of a target biomolecule) may be performed in accordance with any of the methods comprising contacting a sample with a barcoded affinity probe (as provided herein). Optionally, the method comprises binding at least one barcoded affinity probe to a target biomolecule, wherein a barcode sequence from a multimeric barcoding reagent is appended to the barcoded oligonucleotide of the barcoded affinity probe. Optionally, wherein the measurement is made by analysing the barcode sequence of the multimeric barcoding reagent and/or by analysing the sequence of a barcode from the barcoded oligonucleotide of the barcoded affinity probe.

35

Any one or more steps of determining (or measuring) the presence, absence and/or level of a target biomolecule (or measuring a signal corresponding to the presence, absence and/or level of a target biomolecule) may be performed using one or more optical and/or fluorescent/fluorescence measurement processes e.g. using one or more optically-labelled and/or fluorescently-labelled affinity probes. For example, the step of measuring may be performed

using one or more optically-labelled and/or fluorescently-labelled affinity probes, wherein at least one optically-labelled and/or fluorescently-labelled affinity probe is bound to a target biomolecule, and wherein said measurement is made using at least one optical-measurement step or at least one fluorescence-detection step (e.g., wherein said measurement is made by measuring an
5 optical and/or fluorescent signal from said optically-labelled and/or fluorescently-labelled affinity probes).

Optionally, any one or more optical and/or fluorescent/fluorescence measurement processes may comprise an optical and/or fluorescent measurement of a sample comprising one or more
10 circulating microparticles and/or comprising biomolecules from one or more circulating microparticles, wherein said sample is comprised within an aqueous volume and/or an aqueous droplet (such as a droplet analysed with a fluorescence activated cell sorting (FACS) instrument). Optionally, any such optical and/or fluorescent measurement process may further comprise a sorting and/or selection process, for example wherein any one or more optical and/or fluorescent
15 measurement(s) of a circulating microparticle are employed to sort and/or select any given circulating microparticle and/or any group and/or subset of two or more circulating microparticles (for example, to sort a sample comprising circulating microparticles into a first subset of circulating microparticles exhibiting high levels of a particular target biomolecule, and into a second subset of circulating microparticles exhibiting high levels of said particular target biomolecule)

20
Optionally, any one or more optical and/or fluorescent/fluorescence measurement processes may comprise an optical and/or fluorescent measurement of a sample comprising one or more circulating microparticles and/or comprising biomolecules from one or more circulating microparticles, wherein said sample is comprised upon a planar surface (such as a planar glass
25 surface such as a microscope slide, or any other planar surface). Optionally, any one or more optical and/or fluorescent/fluorescence measurement processes may comprise an optical and/or fluorescent measurement of a sample comprising one or more circulating microparticles and/or comprising biomolecules from one or more circulating microparticles, wherein said sample is visualised with an optical microscope and/or a fluorescence microscope.

30
Optionally, any one or more fluorescently-labelled affinity probes may comprise a fluorophore with a particular absorption spectrum and/or emission spectrum. Optionally, any one or more fluorescently-labelled affinity probes comprised within a pool and/or library and/or set of two or more fluorescently-labelled affinity probes may comprise a fluorophore with an absorption
35 spectrum and/or emission spectrum different to at least one and/or at least 2 of the other fluorescently-labelled affinity probes within said pool and/or library and/or set.

Optionally, all fluorescently-labelled affinity probes with affinity for the same target biomolecule comprised within a pool and/or library and/or set of two or more fluorescently-labelled affinity

probes may comprise a fluorophore with the same absorption spectrum and/or emission spectrum. Optionally, all fluorescently-labelled affinity probes with affinity for the same target biomolecule comprised within a pool and/or library and/or set of two or more fluorescently-labelled affinity probes may comprise the same fluorophore. Optionally, fluorescently-labelled affinity probes with affinity for the same target biomolecule comprised within a pool and/or library and/or set of two or more fluorescently-labelled affinity probes may comprise two or more different fluorophores (e.g. two or more different fluorophores comprising two or more different absorption spectra and/or emission spectra). Optionally, fluorescently-labelled affinity probes within a pool and/or library and/or set of two or more fluorescently-labelled affinity probes may each comprise a fluorophore from a set of two or more different fluorophores (e.g. two or more different fluorophores comprising two or more different absorption spectra and/or emission spectra), wherein all said fluorescently-labelled affinity probes with affinity for the same target biomolecule share the same fluorophore, optionally wherein each fluorophore identifies and/or is associated with the target biomolecule of said fluorescently-labelled affinity probes. Optionally, in any pool and/or library and/or set of two or more fluorescently-labelled affinity probes, a number of different fluorophores (e.g. any number of different fluorophores comprising different absorption spectra and/or emission spectra) may be used, such as at least 2, at least 3, at least 4, at least 5, at least 10, at least 15, at least 20, or at least 50.

Optionally, in any method of analysing a sample comprising at least one circulating microparticle, any sample, and/or solution, and/or reaction or reaction mixture, and/or aqueous volume, and/or mixture comprising any number or concentration of circulating microparticles, and/or any number or concentration of biomolecules from one or more circulating microparticles, and/or any number or concentration of (identical or different) barcodes, and/or any number or concentration of (identical or different) barcode molecules, and/or any number or concentration of (identical or different) barcode sequences, and/or any number or concentration of (identical or different) barcoded oligonucleotides, and/or any number or concentration of (identical or different) multimeric barcoding reagents, and/or any number or concentration of (identical or different) affinity moieties, and/or any number or concentration of (identical or different) barcoded affinity probes, and/or any number or concentration of (identical or different) adapter oligonucleotides, and/or any number or concentration of (identical or different) coupling sequences, and/or any number or concentration of (identical or different) enrichment probes, and/or any number or concentration of (identical or different) primers, and/or any number or concentration of (identical or different) hybridisation probes, and/or any number or concentration of (identical or different) fluorescence in situ hybridisation probes, may be comprised within a single partition, or at least first and second partitions (e.g. partitioned or divided into first and second partitions), or comprised within (e.g. partitioned or divided into) any number of partitions, such as at least 3 partitions, at least 4 partitions, at least 5 partitions, at least 10 partitions, at least 100 partitions, at least 1000 partitions, at least 10,000 partitions, at least 100,000 partitions, at least 1,000,000

partitions, at least 10,000,000 partitions, at least 100,000,000 partitions, or at least 1,000,000,000 partitions, during, and/or before, and/or after any one or more steps of said method.

5 Optionally, in any of the methods, any one or more target biomolecules may be measured and/or analysed by a process of optical measurement and/or optical quantitation. Optionally, in any of the methods, any one or more target biomolecules may be measured and/or analysed with an optically labelled and/or fluorescently labelled affinity probe, wherein said affinity probe has affinity and/or specificity for said target biomolecule.

10 Optionally, any method of measuring and/or analysing a biomolecule may comprise one or more steps of direct detection. Optionally, any method of measuring and/or analysing a biomolecule may comprise one or more steps of indirect detection.

15 For avoidance of doubt, in the present invention and in any methods herein, any term referring to any one or more biomolecule being 'in' a circulating microparticle, and/or 'within' a circulating microparticle, and/or 'of' a circulating microparticle, and/or 'from' a circulating microparticle, and/or 'comprised in' a circulating microparticle, and/or 'comprised within' a circulating microparticle, refers broadly to said biomolecule being found (and/or potentially found) fully or partially within any form or location of said circulating microparticle (including fully or partially enclosed within a
20 membrane, and/or fully or partially on the outer surface and/or on the inner surface of a membrane, and/or fully or partially embedded within a membrane).

Optionally, in any of the methods, any step of analysing the sequence of one or more target nucleic acid molecules may be performed by a primer-extension reaction. Optionally, in any of the
25 methods, any step of analysing the sequence of one or more target nucleic acid molecules may be performed by polymerase chain reaction (PCR), optionally with use of a primer set providing amplification (and thus measurement and detection) of a specific target sequence (such as a specific DNA, RNA, or cDNA target sequence). Optionally, in any of the methods, any step of analysing the sequence of one or more target nucleic acid molecules may be performed by a
30 reverse-transcription reaction, optionally with one or more subsequent primer-extension or PCR steps.

Optionally, in any of the methods, any step of analysing the sequence of one or more target nucleic acid molecules may be performed by an in situ hybridisation (ISH) process, such as a
35 fluorescence in situ hybridisation (FISH) process.

The methods of the invention may be deterministic (e.g. one barcode sequence may be used to identify sequence reads from a single microparticle) or probabilistic (e.g. one barcode sequence may be used to identify sequence reads likely to be from a single microparticle). As a further

example, in the methods, the step of partitioning may aim to achieve an average of just 1 circulating microparticle per partition. However, this is an intrinsically statistical process, it cannot be guaranteed that each partition will contain only biomolecules from a single microparticle; therefore, it cannot be guaranteed that the set of linked signals corresponding to the biolecules
5 from a particular partition will correspond to biomolecules from a single microparticle. For example, if a particular partition contains two different microparticles, the set of linked signals may correspond to the two microparticles.

The invention further comprises systems and apparatuses for analysing a sample comprising one
10 or more circulating microparticles (or two or more such samples, e.g. each comprising one or more circulating microparticles). Optionally, such a system may comprise at least one algorithm or a part of an algorithm and/or computer programme (such as an algorithm or part of an algorithm and/or computer programme comprised within a computer system and/or comprised within a web- or internet-based computer storage system) for analysing one or more sets of linked signals
15 derived from measurement of at least one circulating microparticle (e.g. one or more sets of linked sequences derived from measurement of at least one circulating microparticle) (such as any one or more algorithms and/or computer programmes configured to calculate any one or parameter value, such as any parameter values described herein), and/or at least one reference sequence and/or set of reference sequences (such as one or more reference sequences comprised within a
20 computer system and/or a computer data-storage system such as a server and/or hard disc), and/or at least one set of barcoded oligonucleotides, and/or at least one multimeric barcoding reagent and/or a library thereof, and/or at least one physical apparatus comprising one or more partitions (such as one or more tubes, each comprising a partition; and/or one or more plates comprising wells, wherein each well comprises a partition; and/or one or more apparatus
25 comprising two or more partitions wherein each such partition comprises a droplet, such as a microfluidic device comprising or capable of generating microfluidic droplets (such as a Chromium system as provided by 10X Genomics), or a planar surface comprising one or more droplets thereupon), and/or at least one enzyme or enzyme solution capable of appending barcode sequences to target nucleic acids (such as any ligase enzyme, polymerase enzyme, and/or
30 transposase enzyme), and/or at least one algorithm and/or computer programme configured to report the results of any one or more analyses herein (such as the results of any one or more methods of diagnosing and/or diagnostic tests based upon analysing two or more linked signals from a sample comprising one or more circulating microparticles) to a physician and/or other healthcare worker and/or patient. For example, a system for analysing a sample comprising one
35 or more circulating microparticles may comprise at least one algorithm or a part thereof for analysing one or more sets of linked signals derived from measurement of at least one circulating microparticle, and at least one set of barcoded oligonucleotides (configured to be appended to target biomolecules comprised with or derived from a circulating microparticle), and at least one physical apparatus comprising one or more partitions; alternatively such a system may comprise

at least one algorithm or a part thereof for analysing one or more sets of linked signals derived from measurement of at least one circulating microparticle, and at least one library of multimeric barcoding reagents; alternatively such a system may comprise at least one algorithm or a part thereof for analysing one or more sets of linked signals derived from measurement of at least one circulating microparticle, and at least one library of multimeric barcoding reagents, and at least one physical apparatus comprising one or more partitions; alternatively such a system may comprise at least one algorithm or a part thereof for analysing one or more sets of linked signals derived from measurement of at least one circulating microparticle, and at least one set of barcoded oligonucleotides (configured to be appended to target biomolecules comprised with or derived from a circulating microparticle), and at least one algorithm and/or computer programme configured to report the results of any one or more analyses herein.

1. SAMPLES OF CIRCULATING MICROPARTICLES

A sample for use in the methods of the invention may comprise at least one circulating microparticle (i.e. a microparticle originating from blood (e.g. human blood)) and/or a sample for use in the methods of the invention may be derived from at least one circulating microparticle. The microparticle(s) may originate from maternal blood. The microparticle(s) may originate from the blood of a patient with a disease (e.g. cancer). The sample may, for example, be a blood sample, a plasma sample or a serum sample. The sample may be a mammalian sample. Preferably, the sample is a human sample.

The circulating microparticle(s) may be one or more of a variety of cell-free microparticles that have been found in blood, plasma, and/or serum from humans and/or other animals (Orozco et al, Cytometry Part A (2010). 77A: 502 514, 2010). "Cell-free" refers to the fact that such microparticles are not cells. Instead, the microparticles are derived from cells e.g. by secretion or following apoptosis. These microparticles are diverse in the tissues and cells from which they originate, as well as the biophysical processes underlying their formation, as well as their respective sizes and molecular structures and compositions. The microparticle may comprise one or more components from a cell membrane (e.g. incorporating phospholipid components) and one or more intracellular and/or cell-nuclear components. The microparticle(s) may be selected from one or more of exosomes, apoptotic bodies (also known as apoptotic vesicles) and/or extracellular microvesicles.

A microparticle may be defined as a membranous vesicle containing at least two fragments of a target nucleic acid (e.g. genomic DNA). A microparticle may have a diameter of 100-5000 nm. Preferably, the microparticle has a diameter of 100-3000 nanometers.

Exosomes are amongst the smallest circulating microparticles, are typically in the range of 50 to 100 nanometers in diameter, and are thought derive from the cell membrane of viable, intact cells, and contain both protein and RNA components (including both mRNA molecules and/or degraded mRNA molecules, and small regulatory RNA molecules such as microRNA molecules) contained
5 within an outer phospholipid component. Exosomes are thought to be formed by exocytosis of cytoplasmic multivesicular bodies (Gyorgy et al, Cell. Mol. Life Sci. (2011) 68:2667–2688). Exosomes are thought to play varied roles in cell-cell signaling as well as extracellular functions (Kanada et al, PNAS (2015) 1418401112). Techniques for quantitating or sequencing the microRNA and/or mRNA molecules found in exosomes have been described previously (e.g. US
10 patent application 13/456,121, European application EP2626433 A1).

Microparticles also include apoptotic bodies (also known as apoptotic vesicles) and extracellular microvesicles, which altogether can range up to 1 micron or even 2 to 5 microns in diameter, and are generally thought to be larger than 100 nanometers in diameter (Lichtenstein et al, Ann N Y
15 Acad Sci. (2001); 945:239-49). All classes of circulating microparticles are thought to be generated by a large number and variety of cells in the body (Thierry et al, Cancer Metastasis Rev 35 (3), 347-376. 9 (2016) /s10555-016-9629-x).

Preferably, the microparticle is not an exosome e.g. the microparticle is any microparticle having a
20 larger diameter than an exosome.

Samples for use in the methods may include a sample comprising at least one circulating microparticle as well as a sample derived from at least one circulating microparticle. For example, the step of measuring a signal or measuring a reagent (e.g. a barcoded oligonucleotide)
25 may be performed on a sample comprising at least one intact circulating microparticle (e.g. wherein the sample or reaction mixture comprises an intact circulating microparticle at the time of measuring the signal or measuring the reagent). Alternatively, the step of measuring a signal or measuring a reagent (e.g. a barcoded oligonucleotide) may be performed on a sample comprising biomolecules derived from a circulating microparticle (e.g. biomolecules purified and/or processed
30 and/or fractionated and/or isolated from a circulating microparticle). The sample may not comprise an intact circulating microparticle at the time of measuring a signal or measuring a reagent.

A sample may comprise at least 2, at least 3, at least 4, at least 5, at least 7, at least 10, at least
35 15, at least 20, at least 30, at least 40, at least 50, at least 100, at least 200, at least 500, at least 1000, at least 5000, at least 10,000, at least 20,000, at least 50,000, at least 100,000, at least 1,000,000, at least 10,000,000, at least 100,000,000, at least 1,000,000,000, or at least 100,000,000,000 different target biomolecules and/or target epitopes. Preferably, a sample comprises at least 100 target biomolecules and/or target epitopes.

In the sample, the fragments of nucleic acid (e.g. genomic DNA) may be at a concentration of less than 1.0 picograms of DNA per microliter, less than 10 picograms of DNA per microliter, less than 100 picograms of DNA per microliter, less than 1.0 nanograms of DNA per microliter, less than 10
5 nanograms of DNA per microliter, less than 100 nanograms of DNA per microliter, or less than 1000 nanograms of DNA per microliter.

A sample may comprise (or be derived from) at least 2, at least 3, at least 4, at least 5, at least 7, at least 10, at least 50, at least 100, at least 500, at least 1000, at least 5000, at least 10,000, at
10 least 50,000, at least 100,000, at least 1,000,000, at least 10,000,000, or at least 100,000,000 circulating microparticles. Preferably, a sample comprises (or is derived from) at least 100 circulating microparticles.

In a sample, the microparticles may be at a concentration of less than 0.001 microparticles per
15 microliter, less than 0.01 microparticles per microliter, less than 0.1 microparticles per microliter, less than 1.0 microparticles per microliter, less than 10 microparticles per microliter, less than 100 microparticles per microliter, less than 1000 microparticles per microliter, less than 10,000 microparticles per microliter, less than 100,000 microparticles per microliter, less than 1,000,000 microparticles per microliter, less than 10,000,000 microparticles per microliter, or less than
20 100,000,000 microparticles per microliter.

A circulating microparticle may comprise at least 2, at least 3, at least 4, at least 5, at least 7, at least 10, at least 15, at least 20, at least 30, at least 40, at least 50, at least 100, at least 200, at least 500, at least 1000, at least 5000, at least 10,000, at least 20,000, at least 100,000, at least
25 500,000, at least 1,000,000, or at least 10,000,000 different target biomolecules and/or target epitopes. Preferably, a circulating microparticle comprises at least 10 target biomolecules and/or target epitopes.

In the methods of the invention, any number of one or more different target biomolecules and/or
30 target epitopes may be measured and/or analysed. Optionally, in any of the methods, a group of at least 2, at least 3, at least 4, at least 5, at least 7, at least 10, at least 15, at least 20, at least 30, at least 40, at least 50, at least 100, at least 200, at least 500, at least 1000, at least 5000, at least 10,000, or at least 20,000 different target biomolecules and/or target epitopes may be measured and/or analysed. Preferably, a group of at least 3 different target biomolecules and/or
35 target epitopes are measured and/or analysed.

In the methods, the same target biomolecule (and/or target epitope), and/or the same group of 2 or more target biomolecules (and/or target epitopes), may be measured and/or analysed for all circulating microparticles (or partitions) within a sample. Optionally, in any of the methods, a

particular target biomolecule (and/or target epitope), and/or a particular group of 2 or more target biomolecules (and/or target epitopes), may be measured and/or analysed for a subset of circulating microparticles within a sample. Optionally, in any of the methods, a sample of circulating microparticles may be divided into any number of two or more sub-samples, wherein a
5 different particular target biomolecule (and/or target epitope), and/or a different particular group of 2 or more target biomolecules (and/or target epitopes), may be measured and/or analysed for each said sub-sample.

Optionally, in any of the methods, two or more different target epitopes of the same biomolecule
10 may be measured and/or analysed. For example, two or more different affinity probes (such as two or more different antibodies) with affinity or specificity for two or more different epitopes within a target biomolecule (such as a target protein) may be used to measure or analyse said target biomolecule.

15 A biomolecule (also referred to herein as a target biomolecule) may be a chemical or molecular species present in or derived from a circulating microparticle. A biomolecule may be a macromolecule. A biomolecule may be a macromolecule. A biomolecule may be a polypeptide (e.g. a protein), a carbohydrate molecule, a lipid molecule, or a nucleic acid molecule. A biomolecule may be a metabolite. Preferably, the biomolecule is a human biomolecule.

20 A target biomolecule may have a predetermined (or predefined) sequence e.g. a target polypeptide may have a predetermined (or predefined) amino acid sequence or epitope. Similarly, a fragment of a target nucleic acid may have a predetermined (or predefined) nucleotide sequence. The methods may comprise measuring a signal corresponding to the presence,
25 absence and/or level of the predetermined (or predefined) sequence or epitope using a target specific reagent e.g. a barcoded affinity probe or affinity probe.

A biomolecule may be a nucleic acid biomolecule or a non-nucleic acid biomolecule.

30 As used herein, the term "polypeptide" includes a chain of at least two amino acid monomers linked by a peptide bond, a peptide and a protein e.g. a post-translationally modified protein such as a glycoprotein. One or more biomolecules may be one or more protein isoforms.

A biomolecule may comprise an epitope of an antigen present in or derived from a circulating
35 microparticle. For example, the epitope may be an epitope of a polypeptide or protein. A biomolecule may comprise a specific epitope e.g. a specific protein epitope and/or a specific epitope generated by a post-translational modification of a protein (such as a lysine methylation modification). A biomolecule may comprise a specific nucleic acid epitope, such as a specific nucleic acid modification (such as a 5-methylcytosine DNA epitope and/or a 5-hydroxy-

methylcytosine DNA epitope). A biomolecule may comprise a specific epitope recognised by one or more affinity probes (e.g. a barcoded affinity probe), such as a specific epitope recognised by an antibody.

- 5 A biomolecule may comprise an epitope that is not a nucleic acid epitope. A biomolecule may not be a 5-methylcytosine DNA molecule (i.e. a biomolecule may be an epitope that is not a 5-methylcytosine DNA epitope) and/or a biomolecule may not be a 5-hydroxy-methylcytosine DNA molecule (i.e. a biomolecule may be an epitope that is not a 5-hydroxy-methylcytosine DNA epitope).

10

The biomolecule may be a DNA-binding protein. Optionally, the biomolecule is not a DNA-binding protein.

- 15 The biomolecule may be a histone protein (e.g. histone H1, histone H2A, histone H2B, histone H3, and/or histone H4, and/or any histone variant). The histone protein may be a post-translationally modified histone protein (e.g. histone H3 lysine 4 trimethylation, histone H3 lysine 27 trimethylation, and/or any histone acetylation modification). Optionally, the biomolecule is not a histone protein.

- 20 The biomolecule may be a chromatin protein. Optionally, the biomolecule is not a chromatin protein.

- 25 The biomolecule may be a membrane protein or polypeptide. Optionally, the biomolecule is not a membrane protein or polypeptide. The biomolecule may be a polypeptide or protein that immunoprecipitates with DNA. Optionally, the biomolecule is not a polypeptide or protein that immunoprecipitates with DNA.

- 30 The biomolecule may be a biomolecule that binds DNA. Optionally, the biomolecule is not a biomolecule that binds DNA. The biomolecule may be a membrane biomolecule or membrane-associated biomolecule. Optionally, the biomolecule is not a membrane biomolecule or membrane-associated biomolecule. The biomolecule may be a biomolecule that immunoprecipitates with DNA. Optionally, the biomolecule is not a biomolecule that immunoprecipitates with DNA.

- 35 A biomolecule may be comprised fully or partially on the inner surface and/or on the outer surface of a membrane (such as a lipid bilayer membrane of a circulating microparticle) of a circulating microparticle. A biomolecule may be comprised fully or partially enclosed within a membrane of a circulating microparticle (such as enclosed within a lipid bilayer of a circulating microparticle). A biomolecule may be comprised within and/or across a membrane of a circulating microparticle,

and/or any combination thereof. A biomolecule may be comprised fully or partially embedded within a membrane (such as fully or partially embedded within a lipid bilayer membrane of a circulating microparticle) of a circulating microparticle.

- 5 A biomolecule may be derived from the inner surface and/or on the outer surface of a circulating microparticle, and/or derived from within a circulating microparticle (such as derived from within a membrane of a circulating microparticle), and/or derived from within and/or across a membrane of a circulating microparticle, and/or any combination thereof.
- 10 A biomolecule may be DNA (e.g. double-stranded DNA (dsDNA) or single-stranded DNA (ssDNA)), RNA (e.g. double-stranded RNA (dsRNA) or single-stranded RNA (ssRNA)), or a fragment thereof. A biomolecule may be genomic DNA or RNA (e.g. mRNA), or a fragment thereof.
- 15 One or more biomolecules (or target biomolecules) may be a DNA fragment, RNA fragment and/or polypeptide selected from (or encoding) Biomolecule group 1, which comprises:
 - Plasma-based protein markers of cancer and/or cancer aggressiveness, including Prostate-Specific Antigen (PSA), and CA-125;
 - 20 • Cell surface and immune-cell-type markers, including CD3, CD4, CD8, CD19, CD20, CD20, CD41, CD45, CD61, CD62, CD146, CD235a, and CD326;
 - Genes and proteins involved in oncogenesis and malignant transformation, and genes used as immunocytochemical markers for assessing cancer cell type and sub-type, - 25 including Antigen KI-67 (Ki-67), NK2 Homeobox 1 (TTF-1), B-cell Lymphoma 2 (BCL2), BRAF, C-kit/CD117, c-Myc, c-Raf, Ras, Survivin, Vascular Endothelial Growth Factor Receptor (VEGFR), Tumor-Associated Glycoprotein 72 (TAG-72), Epidermal Growth Factor Receptor (EGFR), Estrogen Receptor, Programmed Death Ligand 1 (PD-L1), - 30 Cyclin B1, Epithelial Cell Adhesion Molecule (EpCAM), HER2/Neu, Progesterone Receptor, K-ras, NRAS, Beta-2 Microglobulin (B2M), Calcitonin, CA19-9, CA15-3/CA27.29, Chromogranin A (CgA), Neuron-Specific Enolase, Lactate Dehydrogenase, Thyroglobulin, Claudin-1 (CLDN1), HE4, Platelet-Derived Growth Factor Receptor (PDGF-R), Nuclear Matrix Protein 22, Cytokeratin 8 (CK-8), Cytokeratin 18 (CK-18), - 35 Cytokeratin Fragment 21-1, and OVX1;
 - Markers (i.e. plasma protein markers) associated with pregnancy (such as foetal markers or placental markers) or associated with complications of pregnancy, including Alpha-

– 57 –

fetoprotein (AFP), Beta-human Chorionic Gonadotropin (Beta-hCG), and Toll-like Receptor 4 (TLR4);

- 5 • Proteins associated with circulating lipoprotein particles and/or endovascular plaques, including Annexin V, Apolipoprotein A1 (Apo A-1), Plasminogen Activator Inhibitor (PAI-1), CD31, CD144, and Urokinase Plasminogen Activator (uPA);
- 10 • MicroRNA molecules (miRNAs) associated with (and/or differentially expressed within) endovascular plaques, including miR-1, miR-19b, miR-21, miR-22, miR-29b, miR-92a, miR-99a, miR-100, miR-126, miR-127, miR-133a, miR-133b, miR-143, miR-145, miR-199a, miR-210, and let-7f;
- 15 • Markers of lymphocytes and/or other immune cells, including LY6G6D and Immunoglobulin; and
- And other target biomolecules, including Transthyretin, C-reactive protein (CRP), and troponin.

20 The biomolecules (or target biomolecules) provided above are collectively known herein as “Biomolecule group 1”.

Such a DNA fragment may include all or part of a DNA sequence (for example, a genomic sequence, exonic region sequence, intronic region sequence, promoter region sequence, and/or terminator region sequence) of one or more of the protein encoding genes. Such an RNA
25 fragment may include all or part of an RNA sequence (for example, an exonic RNA sequence, an intronic RNA sequence, a 5'- untranslated region sequence, and/or a 3' untranslated region sequence) of one or more of the protein encoding genes. Such a polypeptide may include all or part of one or more of the proteins. Such a polypeptide may include one or more post-translationally modified forms of said polypeptide (for example, wherein said polypeptide has
30 been acetylated, or methylated, at any one or more amino acid residues). Preferably, the biomolecule is a human biomolecule (e.g. human Ki-67).

A biomolecule may comprise an epigenetic modification. An epigenetic modification may comprise a modified nucleotide e.g. a modified gDNA nucleotide or a modified RNA nucleotide.
35 The modified nucleotide may comprise a modified base. The modified base may be a methylated base e.g. 5-methylcytosine or 5-hydroxy-methylcytosine. A biomolecule (such as a fragment of a target nucleic acid (e.g. genomic DNA)) may comprise 5-methylcytosine (i.e., may comprise 5-methylcytosine DNA and/or may comprise a 5-methylcytosine DNA nucleotide). A biomolecule (such as a fragment of a target nucleic acid (e.g. genomic DNA)) may comprise 5-hydroxy-

5 methylcytosine (i.e., may comprise 5-hydroxy-methylcytosine DNA and/or may comprise a 5-hydroxy-methylcytosine DNA nucleotide). An epigenetic modification may comprise a post-translational modification of a protein. The post-translation modification may be methylation, phosphorylation, acetylation, ubiquitylation and/or sumoylation. The post-translationally modified polypeptide may be a histone protein. For example, a post-translationally modified histone protein (e.g. histone H3 lysine 4 trimethylation, histone H3 lysine 27 trimethylation, and/or any histone acetylation modification).

10 A biomolecule may comprise an exogenously-administered molecule, such an exogenously-administered polypeptide (such as an exogenously-administered antibody), and/or an exogenously-administered nucleic acids (such as an exogenously-administered oligonucleotide, such as an exogenously-administered barcode sequences e.g. a barcoded oligonucleotide).

15 A biomolecule may comprise a barcoded oligonucleotide (or a barcode sequence thereof) of a barcoded affinity probe.

20 At least two of the biomolecules of a circulating microparticle may be fragments of a target nucleic acid (e.g. molecules of fragmented genomic DNA). These molecules of fragmented genomic DNA, and/or sequences comprised within these molecules of fragmented genomic DNA, may be linked by any method described herein.

25 The fragments of the target nucleic acid may be fragments of DNA (e.g. molecules of fragmented genomic DNA) or fragments of RNA (e.g. fragments of mRNA). Preferably, the fragments of the target nucleic acid are fragments of genomic DNA.

The fragments of DNA may be fragments of mitochondrial DNA. The fragments of DNA may be fragments of mitochondrial DNA from a maternal cell or tissue. The fragments of DNA may be fragments of mitochondrial DNA from a foetal or placental tissue. The fragments of DNA may be fragments of mitochondrial DNA from a diseased and/or cancer tissue.

30 A microparticle may comprise a platelet. A microparticle may comprise a tumour-educated platelet. A target nucleic acid may comprise platelet RNA (e.g., fragments of platelet RNA, and/or fragments of a tumour-educated platelet RNA). A sample comprising one or more platelets may comprise platelet-rich plasma (for example, platelet-rich plasma comprising tumour-educated platelets).

The fragments of the target nucleic acid may comprise double-stranded or single stranded nucleic acids. The fragments of genomic DNA may comprise double-stranded DNA or single-stranded

DNA. The fragments of the target nucleic acid may comprise partially double-stranded nucleic acids. The fragments of genomic DNA may comprise partially double-stranded DNA.

5 The fragments of the target nucleic acid may be fragments originating from a single nucleic acid molecule, or fragments originating from two or more nucleic acid molecules. For example, the fragments of genomic DNA may originate from a single genomic DNA molecule.

10 As would be appreciated by the skilled person, as used herein the term fragments of a target nucleic acid refers to the original fragments present in the microparticle and to copies or amplicons thereof. For example, the term fragments of gDNA refers to the original gDNA fragments present in the microparticle and, for example, to DNA molecules that may be prepared from the original genomic DNA fragments by a primer-extension reaction. As a further example, the term fragments of mRNA refers to the original mRNA fragments present in the microparticle and, for example, to cDNA molecules that may be prepared from the original mRNA fragments by
15 reverse transcription.

The fragments of the target nucleic acid (e.g. genomic DNA) may be at least 10 nucleotides, at least 15 nucleotides, at least 20 nucleotides, at least 25 nucleotides or at least 50 nucleotides. The fragments of the target nucleic acid (e.g. genomic DNA) may be 15 to 100,000 nucleotides, 20
20 20 to 50,000 nucleotides, 25 to 25,000 nucleotides, 30 to 10,000 nucleotides, 35-5,000 nucleotides, 40-1000 nucleotides or 50-500 nucleotides. The fragments of the target nucleic acid (e.g. genomic DNA) may be 20 to 200 nucleotides in length, 100 to 200 nucleotides in length, 200 to 1000 nucleotides in length, 50 to 250 nucleotides in length, 1000 to 10,000 nucleotides in length, 10,000 to 100,000 nucleotides in length, or 50 to 100,000 nucleotides in length.
25 Preferably, the molecules of fragmented genomic DNA are 50 to 500 nucleotides in length.

Optionally, any method of analysing a sample comprising one or more circulating microparticle(s) (and/or a sample derived from one or more circulating microparticles), may comprise a combinatoric measurement (e.g. measurement of the presence, absence, and/or level)
30 comprising measurement of any combination of any two or more different biomolecules (e.g. any two or more different target biomolecules). For example, any such method may comprise measurement of linked fragments of genomic DNA (such as by barcoding and/or sequencing), optionally wherein measurement of linked fragments of genomic DNA further comprises measurement and/or estimation of the genomic or nucleotide sequence length(s) of said
35 fragments of genomic DNA, and optionally wherein measurement of linked fragments of genomic DNA further comprises measurement and/or estimation of the genomic coordinates (or genomic position) of the 3' end(s) and/or 5' ends of linked fragments of genomic DNA, and measurement of one or more modified nucleotide or nucleobase (such as measurement of 5-methylcytosine, and measurement of 5-hydroxy-methylcytosine), and measurement of one or more polypeptide

– 60 –

biomolecules (such as measurement of any 1 or more biomolecules from Biomolecule Group 1). Optionally, any such combinatoric measurement may (further) comprise measurement of one or more plasma-based protein markers of cancer and/or cancer aggressiveness, and one or more cell surface or immune-cell-type markers, and one or more proteins involved in oncogenesis and malignant transformation or immunocytochemical markers for assessing cancer cell type and cell type, and one or more markers associated with pregnancy or associated with complications of pregnancy, and one or more proteins associated with circulating lipoprotein particles and/or endovascular plaques, and one or more microRNA molecules (such as any such markers as provided within the lists comprised within Biomolecule Group 1). For example, a combinatoric measurement may comprise measurement of linked fragments of genomic DNA, optionally wherein measurement of linked fragments of genomic DNA further comprises measurement and/or estimation of the genomic or nucleotide sequence length(s) of said fragments of genomic DNA, and optionally wherein measurement of linked fragments of genomic DNA further comprises measurement and/or estimation of the genomic coordinates (or genomic position) of the 3' end(s) and/or 5' ends of linked fragments of genomic DNA, and measurement of one or more modified nucleotide or nucleobase (such as measurement of 5-methylcytosine, and measurement of 5-hydroxy-methylcytosine), and measurement of PSA, and CA-125, and CD4, and CD8, and Ki-67, and BCL2, and EGFR; optionally such a combinatoric measurement may further comprise measurement of TTF-1 and/or Ras and/or c-Myc and/or PD-L1 and/or estrogen receptor and/or cyclin B1.

Optionally, any combinatoric measurement may comprise a separate such combinatoric measurement of two or more samples from a single individual (e.g. a single patient) wherein said two or more samples are taken/made from the same individual but separated by one or more durations of time (such as as at least 1 month, at least 3 months, at least 6 months, at least 12 months, at least 18 months, at least 2 years, at least 3 years, at least 4 years, at least 5 years, and/or at least 10 years, and/or any other duration of time). For example a particular combinatoric measurement (of any sort described herein) may be performed on a first sample taken from an individual, and separately performed on a second sample take from said individual at a later period of time. Any number of such sequential (time-separated) samples from an individual may be so analysed, such as at least 3, at least 4, at least 5, at least 6, at least 8, at least 10, at least 15, at least 20, at least 25, or at least 30 sequential samples, or any greater or similar number.

2. ISOLATING SAMPLES OF CIRCULATING MICROPARTICLES

A large number of methods for isolating circulating microparticles (and/or particular subsets, categories, or fractions of circulating microparticles) have been described previously. European patent(s) ES2540255 (B1) and US patent 9005888 B2 describe methods of isolating particular circulating microparticles such as apoptotic bodies based upon centrifugation procedures. A large

number of methods for isolating different types of cell-free microparticles by centrifugation, ultracentrifugation, and other techniques have been well described and developed previously (Gyorgy et al, Cell. Mol. Life Sci. (2011) 68:2667–2688).

- 5 The methods may further comprise isolating a sample comprising one or more circulating microparticles from blood, plasma or serum. The microparticle(s) may be isolated from blood, plasma or serum. The method may further comprise a step of isolating the microparticle(s) from blood, plasma or serum.
- 10 The microparticle(s) may be isolated by centrifugation, size exclusion chromatography and/or filtering.

- 15 The step of isolating may comprise centrifugation. The microparticle(s) may be isolated by pelleting with a centrifugation step and/or an ultracentrifugation step, or a series of two or more centrifugation steps and/or ultracentrifugation steps at two or more different speeds, wherein the pellet and/or the supernatant from one centrifugation/ultracentrifugation step is further processed in a second centrifugation/ultracentrifugation step, and/or a differential centrifugation process

- 20 The centrifugation or ultracentrifugation step(s) may be performed at a speed of 100-500,000 G, 100-1000 G, 1000-10,000 G, 10,000-100,000 G, 500-100,000 G, or 100,000-500,000 G. The centrifugation or ultracentrifugation step may be performed for a duration of at least 5 seconds, at least 10 seconds, at least 30 seconds, at least 60 seconds, at least 5 minutes, at least 10 minutes, at least 30 minutes, at least 60 minutes, or at least 3 hours.

- 25 The step of isolating may comprise size exclusion chromatography e.g. a column-based size exclusion chromatography process, such as one including a column comprising a sepharose-based matrix, or a sephacryl-based matrix.

- 30 The size exclusion chromatography may comprise using a matrix or filter comprising pore sizes at least 50 nanometers, at least 100 nanometers, at least 200 nanometers, at least 500 nanometers, at least 1.0 micrometer, at least 2.0 micrometers, or at least 5.0 micrometers in size or diameter.

- 35 The step of isolating may comprise filtering the sample. The filtrate may provide the microparticle(s) analysed in the methods. Optionally, the filter is used to isolate microparticles below a certain size, and wherein the filter preferentially or completely removes particles greater than 100 nanometers in size, greater than 200 nanometers in size, greater than 300 nanometers in size, greater than 500 nanometers in size, greater than 1.0 micrometer in size, greater than 2.0 micrometers in size, greater than 3.0 micrometers in size, greater than 5.0 micrometers in size, or greater than 10.0 micrometers in size. Optionally, two or more such filtering steps may be

performed, using filters with the same size-filtering parameters, or with different size-filtering parameters. Optionally, the filtrate from one or more filtering steps comprises microparticles, and linked sequence reads are produced therefrom.

5 **3. PREPARATION OF SAMPLES OF CIRCULATING MICROPARTICLES FOR ANALYSIS**

In the methods, any one or more target biomolecules may be measured and/or analysed whilst the circulating microparticle is intact. Optionally, any one or more target biomolecules may be measured and/or analysed whilst the circulating microparticle is not intact (i.e. after the one or more biomolecules have been released from the circulating microparticle).

A sample comprising one or more circulating microparticle may be chemically crosslinked (e.g. with formaldehyde). A sample comprising one or more circulating microparticles may be permeabilised (e.g. with chemical surfactant). A sample comprising one or more circulating microparticles may be chemically crosslinked (e.g. with formaldehyde). The step(s) of chemical crosslinking and/or permeabilisation may be performed prior to measuring and/or analysing the target biomolecule(s) of the one or more circulating microparticles.

The cross-linking step may be performed with a chemical crosslinking agent e.g. formaldehyde, paraformaldehyde, glutaraldehyde, disuccinimidyl glutarate, ethylene glycol bis(succinimidyl succinate), a homobifunctional crosslinker, or a heterobifunctional crosslinker. Any such crosslinking step may further be ended by a quenching step, such as quenching a formaldehyde-crosslinking step by mixing with a solution of glycine. Any such crosslinks may be removed prior to specific subsequent steps of the protocol, such as prior to a primer-extension, PCR, or nucleic acid purification step. A step of crosslinking by a chemical crosslinking agent serves the purpose of holding biomolecules (e.g. fragments of genomic DNA and/or polypeptides) within each microparticle in physical proximity to each other, such that the sample may be manipulated and processed whilst retaining the basic structural nature of the microparticles (i.e., whilst retaining physical proximity of genomic DNA fragments and/or polypeptides derived from the same microparticle).

The microparticle(s) may be permeabilised with an incubation step. The incubation step may be performed in the presence of a chemical surfactant (e.g. Triton X-100 ($C_{14}H_{22}O(C_2H_4O)_n$ (n=9-10)), NP-40, Tween 20, Tween 80, Saponin, Digitonin, or Sodium dodecyl sulfate). The incubation step may be performed at a temperature of at least 20 degrees Celsius, at least 30 degrees Celsius, at least 37 degrees Celsius, at least 45 degrees Celsius, at least 50 degrees Celsius, at least 60 degrees Celsius, at least 65 degrees Celsius, at least 70 degrees Celsius, or at least 80 degrees Celsius. The incubation step may be at least 1 second long, at least 5 seconds long, at

least 10 seconds long, at least 30 seconds long, at least 1 minute long, at least 5 minutes long, at least 10 minutes long, at least 30 minutes long, at least 60 minutes long, or at least 3 hours long.

Any one or more target biomolecules may be measured and/or analysed following a step of transferring any one or more of the reagents described herein (e.g. barcoded oligonucleotides, multimeric barcoding reagents, affinity probes, barcoded affinity probes etc.) into one or more microparticles. The methods may comprise a step of transferring any one or more of the reagents described herein (e.g. barcoded oligonucleotides, multimeric barcoding reagents, affinity probes, barcoded affinity probes etc.) into one or more circulating microparticles.

10

In the methods, any one or more of the reagents described herein may be transferred into one or more circulating microparticles by complexation with a transfection reagent or lipid carrier (e.g. a liposome or a micelle). The transfection reagent may be a lipid transfection reagent e.g. a cationic lipid transfection reagent. Optionally, said cationic lipid transfection reagent comprises at least two alkyl chains. Optionally, said cationic lipid transfection reagent may be a commercially available cationic lipid transfection reagent such as Lipofectamine.

15

In the methods, the reagents for analysing a first circulating microparticle may be comprised within a first lipid carrier, and the reagents for analysing a second circulating microparticle may be comprised within a second lipid carrier. The lipid carrier may be a liposome or a micelle.

20

Prior to the step of transferring, the method may comprise step of cross-linking the biomolecules (e.g. the fragments of genomic DNA and/or target polypeptides) in the microparticle. Prior to the step of transferring, and optionally after the step of cross-linking, the method may further comprise the step of permeabilising the microparticle

25

Any one or more target biomolecules may be measured and/or analysed following a step of releasing the target biomolecules from the one or more circulating microparticles. The one or more target biomolecules may be released from the circulating microparticle(s) by a step of dissolving, permeabilising and/or lysing the circulating microparticle(s). The methods of the invention may comprise releasing the target biomolecules from the one or more circulating microparticles (e.g. by dissolving, permeabilising and/or lysing the one or more circulating microparticles). This release step may be performed with a high-temperature incubation step, and/or via incubation with a molecular solvent or chemical surfactant

30

35

Any one or more target biomolecules may be measured and/or analysed following a step of purifying and/or isolating and/or processing any one or more target biomolecules from one or more circulating microparticles. The methods of the invention may comprise one or more steps of processing, purifying, fractionating, and/or isolating any or all target biomolecules and/or other

constituents of said circulating microparticle(s), prior to, and/or during, and/or following any step of analysing said sample. The methods may comprise a step of purifying and/or isolating nucleic acids (such as DNA molecules and/or RNA molecules). The methods may comprise a step of purifying and/or isolating polypeptides (such as proteins and/or post-translationally modified proteins).

Any one or more target biomolecules may be measured and/or analysed following a step of binding and/or appending any one or more said target biomolecules and/or target nucleic acid molecules to a support, such as a solid support, and/or a semi-solid support, and/or a gel support.

The methods may comprise a step of appending one or more molecules (such as any one or more nucleic acid molecules, such as DNA molecules and/or RNA molecules, and/or any polypeptide molecules such as proteins or post-translationally modified proteins) to a support. Any number or fraction of such molecules from a sample comprising one or more circulating microparticles may be appended to one or more supports; optionally, at least 0.01%, at least 0.1%, at least 1%, at least 10%, at least 50% or 100% of such molecules may be appended to one or more supports.

Any one or more such molecules may be linked to any form of support (e.g. a macromolecule, solid support or semi-solid support, or a dendrimer). Any support may be a bead (e.g. a gel bead, an agarose bead, a silica bead, a styrofoam bead, a gel bead (such as those available from 10x Genomics®), an antibody conjugated bead, an oligo-dT conjugated bead, a streptavidin bead or a magnetic bead (e.g. a superparamagnetic bead). Any bead may be of any size and/or molecular structure (such as 10 nanometres to 100 microns in diameter, 100 nanometres to 10 microns in diameter, or 1 micron to 5 microns in diameter). The molecules may be linked to the support directly or indirectly (e.g. via a linker molecule). The molecules may be linked by being bound to the support and/or by being bound or annealed to linker molecules that are bound to the support. The molecules may be bound to the support (or to the linker molecules) by covalent linkage, non-covalent linkage (e.g. a protein-protein interaction or a streptavidin-biotin bond) or nucleic acid hybridization. The linker molecule may be a biopolymer (e.g. a nucleic acid molecule) or a synthetic polymer. The linker molecule may comprise one or more units of ethylene glycol and/or poly(ethylene) glycol (e.g. hexa-ethylene glycol or penta-ethylene glycol). The linker molecule may comprise one or more ethyl groups, such as a C3 (three-carbon) spacer, C6 spacer, C12 spacer, or C18 spacer. Any support may be functionalised to enable attachment of two or more molecules. This functionalisation may be enabled through the addition of chemical moieties (e.g. carboxylated groups, alkynes, azides, acrylate groups, amino groups, sulphate groups, or succinimide groups), and/or protein-based moieties (e.g. streptavidin, avidin, or protein G) to the support.

- The molecules may be linked by a macromolecule by being bound to the macromolecule and/or by being annealed to the macromolecule. The macromolecule may be a nucleic acid comprising two or more nucleotides each capable of binding to a barcode molecule. Additionally or alternatively, the nucleic acid may comprise two or more regions each capable of hybridizing to a barcode molecule. The macromolecule may be a synthetic polymer (e.g. a dendrimer) or a biopolymer such as a nucleic acid (e.g. a single-stranded nucleic acid such as single-stranded DNA), a peptide, a polypeptide or a protein (e.g. a multimeric protein). The dendrimer may comprise at least 2, at least 3, at least 5, or at least 10 generations.
- 5 The methods may comprise appending one or more circulating microparticles to a support by a method comprising: (a) appending coupling molecules comprising one or more biotin moieties to target molecules (such as target nucleic acid molecules, or target polypeptide molecules) by any method, and/or appending biotin-conjugated affinity probes to target molecules, to create biotin-conjugated target molecules and (b) appending said biotin-conjugated target molecules to one or more streptavidin-conjugated supports (such as one or more streptavidin-conjugated beads).
10 Optionally, prior to and/or during step (b), partitioning said biotin-conjugated target molecules into two or more partitions.
- 15 Any one or more target biomolecules may be measured and/or analysed following a step of partitioning the sample into two or more partitions. The methods may comprise partitioning the sample into two or more partitions. Optionally, each partition may comprise one or more supports, wherein molecules from the microparticle(s) partitioned into each partition are appended to supports comprised within the same partitions respectively. Optionally, a sample comprising any number of microparticles (such as at least 1000 microparticles, at least 1,000,000
20 microparticles, or at least 100,000,000 microparticles) may be appended by such a process. Optionally, any number and/or average number of microparticles may be partitioned into each partition (for example, an average of less than 100, less than 10, less than 1, less than 0.5, less than 0.2, less than 0.1, less than 0.05, less than 0.01, less than 0.001, less than 0.0001, less than 0.00001, or less than 0.000001 microparticles may be partitioned into each partition). Each
25 partition may contain, or on average contain, any number of supports, such as an average of 0.1 supports, an average of 0.5 supports, an average of 1 support, an average of 2 supports, an average of 5 supports, an average of 10 supports, or an average of 100 supports. Optionally, following any process of appending molecules from a sample comprising two or more circulating microparticles to supports within partitions, all or any fraction of the solution comprised within any
30 fraction and/or all partitions may be merged together to form a single, de-partitioned support-appended reaction mixture, wherein said de-partitioned support-appended reaction mixture comprises supports to which molecules from the sample have been so appended. Optionally, said de-partitioned support-appended reaction mixture may then be employed for any process of analysing the sample comprising two or more circulating microparticles, such as any method of
35

measuring fragments of genomic DNA, any method of measuring modified nucleotides or nucleobases, and/or any method of measuring one or more target polypeptides. Optionally, two or more target molecules appended to a support within a de-partitioned support-appended reaction mixture (e.g., two or more molecules from the same circulating microparticle, such as two or more fragments of genomic DNA, and/or two or more polypeptides bound to a barcoded affinity probe) may be appended to the same barcode sequence, or to different barcode sequences from a set of barcode sequences, to link the said two or more target molecules. Optionally, any said process of appending barcode sequences may comprise appending two or more barcoded oligonucleotides from a multimeric barcoding reagent to two or more target molecules appended to the same support within a de-partitioned support-appended reaction mixture. Optionally, any said process of appending barcode sequences may comprise contacting a de-partitioned support-appended reaction mixture with a library of at least 2, at least 100, at least 1000, at least 10,000, at least 1,000,000, at least 10,000,000, or at least 1,000,000,000 multimeric barcoding reagents, and appending barcoded oligonucleotides comprised within said multimeric barcoding reagents to target molecules that have been appended to supports within said de-partitioned support-appended reaction mixture. Any one or more de-partitioned support-appended reaction mixture(s) may comprise a sample derived from one or more circulating microparticle(s) for use with any one or more method(s) described herein. Optionally, for any such method, any number of partitions (such as at least 10, at least 1000, at least 1,000,000, or at least 1,000,000,000 partitions), any type of partitions (such as reaction tubes, or aqueous droplets, or aqueous droplets within an emulsion), and/or any volume of partitions (such as less than or greater than 100 femtoliters, less than or greater than 1.0, 10.0, or 100.0 picoliters, less than or greater than 1.0, 10.0, or 100.0 nanoliters, or less than or greater than 1.0, 10.0, or 100.0 microliters) may be used, such as any number, type, or volume of partition described herein and/or in PCT/GB2017/053820, the content of which is incorporated herein by reference.

4. LINKING BY BARCODING

The invention provides a method of preparing a sample for sequencing, wherein the sample comprise a circulating microparticle (or microparticle originating from blood), wherein the microparticle contains at least two fragments of a target nucleic acid (e.g. genomic DNA), and wherein the method comprises appending the at least two fragments of the target nucleic acid of the microparticle to a barcode sequence, or to different barcode sequences of a set of barcode sequences, to produce a set of linked fragments of the target nucleic acid.

The invention provides a method of preparing a sample for sequencing, wherein the sample comprise a circulating microparticle, wherein the circulating microparticle contains at least two fragments of a target nucleic acid (e.g. genomic DNA), and wherein the method comprises appending the at least two fragments of the target nucleic acid of the circulating microparticle to a

barcode sequence, or to different barcode sequences of a set of barcode sequences, to produce a set of linked fragments of the target nucleic acid.

- 5 Prior to the step of appending the at least two fragments of the target nucleic acid of the microparticle to a barcode sequence, or to different barcode sequences of a set of barcode sequences, the method may comprise appending a coupling sequence to each of the fragments of the target nucleic acid (e.g. genomic DNA) of the microparticle, wherein the coupling sequences are then appended to the barcode sequence, or to different barcode sequences of a set of barcode sequences, to produce the set of linked fragments of the target nucleic acid.
- 10 In the method, the sample may comprise first and second microparticles originating from blood, wherein each microparticle contains at least two fragments of a target nucleic acid (e.g. genomic DNA), and wherein the method may comprise appending the at least two fragments of the target nucleic acid of the first microparticle to a first barcode sequence, or to different barcode sequences of a first set of barcode sequences, to produce a first set of linked fragments of the target nucleic acid and appending the at least two fragments of the target nucleic acid of the second microparticle to a second barcode sequence, or to different barcode sequences of a second set of barcode sequences, to produce a second set of linked fragments of the target nucleic acid.
- 15 20 The first barcode sequence may be different to the second barcode sequence. The barcode sequences of the first set of barcode sequences may be different to the barcode sequences of the second set of barcode sequences.
- 25 In the methods, the sample may comprise n microparticles originating from blood, wherein each microparticle contains at least two fragments of a target nucleic acid (e.g. genomic DNA), and wherein the method comprises performing step (a) to produce n sets of linked fragments of the target nucleic acid, one set for each of the n microparticles.
- 30 In the methods, n may be at least 3, at least 5, at least 10, at least 50, at least 100, at least 1000, at least 10,000, at least 100,000, at least 1,000,000, at least 10,000,000, at least 100,000,000, at least 1,000,000,000, at least 10,000,000,000, or at least 100,000,000,000. Preferably, n is at least 100,000 microparticles.
- 35 Preferably, each set of linked sequence reads (i.e. set of linked signals) is linked by a different barcode sequence or a different set of barcode sequences. Each barcode sequence of a set of barcode sequences may be different to the barcode sequences of at least 1, at least 4, at least 9, at least 49, at least 99, at least 999, at least 9,999, at least 99,999, at least 999,999, at least 9,999,999, at least 99,999,999, at least 999,999,999, at least 9,999,999,999, at least

99,999,999,999, or at least 999,999,999,999 other sets of barcode sequences in the library. Each barcode sequence of a set of barcode sequences may be different to the barcode sequences of all of the other sets of barcode sequences in the library. Preferably, each barcode sequence in a set of barcode sequences is different to the barcode sequences at least 9 other sets of barcode sequences in the library.

The invention provides a method of analysing a sample comprising a microparticle originating from blood, wherein the microparticle contains at least two fragments of a target nucleic acid, and wherein the method comprises: (a) preparing the sample for sequencing comprising appending the at least two fragments of a target nucleic acid (e.g. genomic DNA) of the microparticle to a barcode sequence to produce a set of linked fragments of the target nucleic acid; and (b) sequencing each of the linked fragments in the set to produce at least two linked sequence reads, wherein the at least two linked sequence reads are linked by the barcode sequence.

A barcode sequence may contain a unique sequence. Each barcode sequence may comprise at least 5, at least 10, at least 15, at least 20, at least 25, at least 50 or at least 100 nucleotides. Preferably, each barcode sequence comprises at least 5 nucleotides. Preferably each barcode sequence comprises deoxyribonucleotides, optionally all of the nucleotides in a barcode sequence are deoxyribonucleotides. One or more of the deoxyribonucleotides may be a modified deoxyribonucleotide (e.g. a deoxyribonucleotide modified with a biotin moiety or a deoxyuracil nucleotide). The barcode sequence may comprise one or more degenerate nucleotides or sequences. The barcode sequence may not comprise any degenerate nucleotides or sequences.

In the method, prior to the step of appending the at least two fragments of the target nucleic acid of the microparticle to a barcode sequence, the method may comprise appending a coupling sequence to each of the fragments of the nucleic acid of the microparticle, wherein the coupling sequences are then appended to the barcode sequence to produce the set of linked fragments.

In the methods, the sample may comprise first and second microparticles originating from blood, wherein each microparticle contains at least two fragments of a target nucleic acid (e.g. genomic DNA), and wherein the method comprises performing step (a) to produce a first set of linked fragments of the target nucleic acid for the first microparticle and a second set of linked fragments of the target nucleic acid for the second microparticle, and performing step (b) to produce a first set of linked sequence reads (i.e. set of linked signals) for the first microparticle and a second set of linked sequence reads (i.e. set of linked signals) for the second microparticle, wherein the at least two linked sequence reads for the first microparticle are linked by a different barcode sequence to the at least two linked sequence reads of the second microparticle.

The first set of linked fragments may be linked by a different barcode sequence to the second set of linked fragments.

5 In the methods, the sample may comprise n microparticles originating from blood, wherein each microparticle contains at least two fragments of a target nucleic acid (e.g. genomic DNA), and wherein the method comprises performing step (a) to produce n sets of linked fragments of the target nucleic acid, one set for each of the n microparticles, and performing step (b) to produce n sets of linked sequence reads (i.e. sets linked signals), one for each of the n microparticles.

10 In the methods, n may be at least 3, at least 5, at least 10, at least 50, at least 100, at least 1000, at least 10,000, at least 100,000, at least 1,000,000, at least 10,000,000, at least 100,000,000, at least 1,000,000,000, at least 10,000,000,000, or at least 100,000,000,000. Preferably, n is at least 100,000 microparticles.

15 Preferably, each set of linked sequence reads (i.e. set of linked signals) is linked by a different barcode sequence.

In the methods, the different barcode sequences may be provided as a library of barcode sequences. The library used in the methods may comprise at least 2, at least 5, at least 10, at
20 least 50, at least 100, at least 1000, at least 10,000, at least 100,000, at least 1,000,000, at least 10,000,000, at least 100,000,000, at least 1,000,000,000, at least 10,000,000,000, at least 100,000,000,000, or at least 1,000,000,000,000 different barcode sequences. Preferably, the library used in the methods comprises at least 1,000,000 different barcode sequences.

25 In the methods, each barcode sequence of the library may be appended only to fragments from a single microparticle.

The methods may be deterministic i.e. one barcode sequence may be used to identify sequence reads from a single microparticle or probabilistic i.e. one barcode sequence may be used to
30 identify sequence reads likely to be from a single microparticle. In certain embodiments, one barcode sequence may be appended to fragments of genomic DNA from two or more microparticles.

The method may comprise: (a) preparing the sample for sequencing comprising appending each
35 of the at least two fragments of a target nucleic acid (e.g. genomic DNA) of the microparticle to a different barcode sequence of a set of barcode sequences to produce a set of linked fragments of the target nucleic acid; and (b) sequencing each of the linked fragments in the set to produce at least two linked sequence reads, wherein the at least two linked sequence reads are linked by the set of barcode sequences.

In the methods, prior to the step of appending each of the at least two fragments of the target nucleic acid of the microparticle to a different barcode sequence, the method may comprise appending a coupling sequence to each of the fragments of the target nucleic acid of the microparticle, wherein each of the at least two fragments of the target nucleic acid of the microparticle is appended to a different barcode sequence of the set of barcode sequences by its coupling sequence.

In the methods, the sample may comprise first and second microparticles originating from blood, wherein each microparticle contains at least two fragments of a target nucleic acid (e.g. genomic DNA), and wherein the method may comprise performing step (a) to produce a first set of linked fragments of the target nucleic acid for the first microparticle and a second set of linked fragments of the target nucleic acid for the second microparticle, and performing step (b) to produce a first set of linked sequence reads (i.e. set of linked signals) for the first microparticle and a second set of linked sequence reads (i.e. set of linked signals) for the second microparticle, wherein the first set of linked sequence reads are linked by a different set of barcode sequences to the second set of linked sequence reads.

In the methods, the sample may comprise n microparticles originating from blood, wherein each microparticle contains at least two fragments of a target nucleic acid (e.g. genomic DNA), and wherein the method may comprise performing step (a) to produce n sets of linked fragments of the target nucleic acid, one set for each of the n microparticles, and performing step (b) to produce n sets of linked sequence reads (i.e. sets linked signals), one for each of the n microparticles.

In the methods, n may be at least 3, at least 5, at least 10, at least 50, at least 100, at least 1000, at least 10,000, at least 100,000, at least 1,000,000, at least 10,000,000, at least 100,000,000, at least 1,000,000,000, at least 10,000,000,000, or at least 100,000,000,000. Preferably, n is at least 100,000 microparticles.

Preferably, each set of linked sequence reads (i.e. set of linked signals) is linked by a different set of barcode sequences.

In the methods, the different sets of barcode sequences may be provided as a library of sets of barcode sequences. The library used in the methods may comprise at least 2, at least 5, at least 10, at least 50, at least 100, at least 1000, at least 10,000, at least 100,000, at least 1,000,000, at least 10,000,000, at least 100,000,000, at least 1,000,000,000, at least 10,000,000,000, at least 100,000,000,000, or at least 1,000,000,000,000 different sets of barcode sequences. Preferably, the library used in the methods comprises at least 1,000,000 different sets of barcode sequences.

Each barcode sequence of a set of barcode sequences may be different to the barcode sequences of at least 1, at least 4, at least 9, at least 49, at least 99, at least 999, at least 9,999, at least 99,999, at least 999,999, at least 9,999,999, at least 99,999,999, at least 999,999,999, at
5 least 9,999,999,999, at least 99,999,999,999, or at least 999,999,999,999 other sets of barcode sequences in the library. Each barcode sequence in a set of barcode sequences may be different to the barcode sequences of all of the other sets of barcode sequences in the library. Preferably, each barcode sequence in a set of barcode sequences is different to the barcode sequences at
10 least 9 other sets of barcode sequences in the library.

In the methods, barcode sequences from a set of barcode sequences of the library may be appended only to fragments from a single microparticle.

The methods may be deterministic i.e. one set of barcode sequences may be used to identify
15 sequence reads from a single microparticle or probabilistic i.e. one set of barcode sequences may be used to identify sequence reads likely to be from a single microparticle.

The method may comprise preparing first and second samples for sequencing, wherein each sample comprises at least one microparticle originating from blood, wherein each microparticle
20 contains at least two fragments of a target nucleic acid (e.g. genomic DNA), and wherein the barcode sequences each comprise a sample identifier region, and wherein the method comprises: (i) performing step (a) for each sample, wherein the barcode sequence(s) appended to the fragments of the target nucleic acid from the first sample have a different sample identifier region to the barcode sequence(s) appended to the fragments of the target nucleic acid from the
25 second sample; (ii) performing step (b) for each sample, wherein each linked sequence read comprises the sequence of the sample identifier region; and (iii) determining the sample from which each linked sequence read is derived by its sample identifier region.

In the methods, before, during, and/or after the step(s) of appending barcode sequences and/or
30 coupling sequences, the method may comprise the step of cross-linking the fragments of genomic DNA in the microparticle(s).

In the methods, before, during, and/or after the step(s) of appending barcode sequences and/or coupling sequences, and/or optionally after the step of cross-linking the fragments of genomic
35 DNA in the microparticle(s), the method may comprise the step of permeabilising the microparticle(s). Prior to the step of transferring, and optionally after the step of cross-linking, the method comprises permeabilising the microparticle.

Barcode sequences may be comprised within barcoded oligonucleotides in a solution of barcoded oligonucleotides; such barcoded oligonucleotides may be single-stranded double-stranded, or single-stranded with one or more double-stranded regions. The barcoded oligonucleotides may be ligated to the fragments of the target nucleic acid in a single-stranded or double-stranded
5 ligation reaction. The barcoded oligonucleotide may comprise a single-stranded 5' or 3' region capable of ligating to a fragment of the target nucleic acid. Each barcoded oligonucleotide may be ligated to a fragment of the target nucleic acid in a single-stranded ligation reaction. Alternatively, barcoded oligonucleotides may comprise a blunt, recessed, or overhanging 5' or 3' region capable of ligating to a fragment of the target nucleic acid. Each barcoded oligonucleotide
10 may be ligated to a fragment of the target nucleic acid in a double-stranded ligation reaction.

In certain methods, the ends of fragments of the target nucleic acid may be converted into blunt double-stranded ends in a blunting reaction and the barcoded oligonucleotides may comprise a blunt double-stranded end. Each barcoded oligonucleotide may be ligated to a fragment of the
15 target nucleic in a blunt-end ligation reaction. In certain methods, the ends of fragments of the target nucleic acid may have their ends converted into blunt double-stranded ends in a blunting reaction, and then have their ends converted into a form with single 3' adenosine overhangs, and wherein the barcoded oligonucleotides comprise a double-stranded end with a single 3' thymine overhang capable of annealing to the single 3' adenosine overhangs of the fragments of the
20 target nucleic acid. Each barcoded oligonucleotide may be ligated to a fragment of the target nucleic acid in a double-stranded A/T ligation reaction.

In certain methods, barcoded oligonucleotides comprise a target region on their 3' or 5' end capable of annealing to a target region in a target nucleic acid and/or coupling sequence, and
25 barcode sequences may be appended to target nucleic acids by annealing barcoded oligonucleotides to said target nucleic acid and/or coupling sequence, and optionally extending and/or ligating the barcoded oligonucleotide to a nucleic acid target and/or coupling sequence.

In certain methods, a coupling sequence may be appended to fragments of genomic DNA prior to
30 appending a barcoded oligonucleotide.

The method may comprise, prior to the step of appending, the step of partitioning the nucleic acid sample into at least two different reaction volumes.

35 5. LINKING BY BARCODING USING MULTIMERIC BARCODING REAGENTS

The invention provides a method of preparing a sample for sequencing, wherein the sample comprises a circulating microparticle (i.e. a microparticle originating from blood), and wherein the microparticle contains at least two fragments of a target nucleic acid (e.g. genomic DNA), and wherein the method comprises the steps of: (a) contacting the sample with a multimeric barcoding

reagent, wherein the multimeric barcoding reagent comprises first and second barcode regions linked together, wherein each barcode region comprises a nucleic acid sequence; and (b) appending barcode sequences to each of first and second fragments of the target nucleic acid of the microparticle to produce first and second barcoded target nucleic acid molecules for the microparticle, wherein the first barcoded target nucleic acid molecule comprises the nucleic acid sequence of the first barcode region and the second barcoded target nucleic acid molecule comprises the nucleic acid sequence of the second barcode region.

The invention provides a method of preparing a sample for sequencing, wherein the sample comprises a microparticle originating from blood, and wherein the microparticle contains at least two fragments of a target nucleic acid (e.g. genomic DNA), and wherein the method comprises the steps of: (a) contacting the sample with the multimeric barcoding reagent, wherein the multimeric barcoding reagent comprises first and second barcoded oligonucleotides linked together, and wherein the barcoded oligonucleotides each comprise a barcode region; and (b) annealing or ligating the first and second barcoded oligonucleotides to first and second fragments of the target nucleic acid of the microparticle to produce first and second barcoded target nucleic acid molecules.

The invention provides a method of preparing a sample for sequencing, wherein the sample comprises first and second microparticles originating from blood, and wherein each microparticle contains at least two fragments of a target nucleic acid (e.g. genomic DNA), and wherein the method comprises the steps of: (a) contacting the sample with a library comprising at least two multimeric barcoding reagents, wherein each multimeric barcoding reagent comprises first and second barcode regions linked together, wherein each barcode region comprises a nucleic acid sequence and wherein the first and second barcode regions of a first multimeric barcoding reagent are different to the first and second barcode regions of a second multimeric barcoding reagent of the library; and (b) appending barcode sequences to each of first and second fragments of the target nucleic acid of the first microparticle to produce first and second barcoded target nucleic acid molecules for the first microparticle, wherein the first barcoded target nucleic acid molecule comprises the nucleic acid sequence of the first barcode region of the first multimeric barcoding reagent and the second barcoded target nucleic acid molecule comprises the nucleic acid sequence of the second barcode region of the first multimeric barcoding reagent, and appending barcode sequences to each of first and second fragments of the target nucleic acid of the second microparticle to produce first and second barcoded target nucleic acid molecules for the second microparticle, wherein the first barcoded target nucleic acid molecule comprises the nucleic acid sequence of the first barcode region of the second multimeric barcoding reagent and the second barcoded target nucleic acid molecule comprises the nucleic acid sequence of the second barcode region of the second multimeric barcoding reagent.

- 74 -

The invention provides a method of preparing a sample for sequencing, wherein the sample comprises first and second microparticles originating from blood, and wherein each microparticle contains at least two fragments of a target nucleic acid (e.g. genomic DNA), and wherein the method comprises the steps of: (a) contacting the sample with a library comprising at least two
5 multimeric barcoding reagents, wherein each multimeric barcoding reagent comprises first and second barcoded oligonucleotides linked together, wherein the barcoded oligonucleotides each comprise a barcode region and wherein the barcode regions of the first and second barcoded oligonucleotides of a first multimeric barcoding reagent of the library are different to the barcode regions of the first and second barcoded oligonucleotides of a second multimeric barcoding
10 reagent of the library; and (b) annealing or ligating the first and second barcoded oligonucleotides of the first multimeric barcoding reagent to first and second fragments of the target nucleic acid of the first microparticle to produce first and second barcoded target nucleic acid molecules, and annealing or ligating the first and second barcoded oligonucleotides of the second multimeric barcoding reagent to first and second fragments of the target nucleic acid of the second
15 microparticle to produce first and second barcoded target nucleic acid molecules.

The barcoded oligonucleotides may be ligated to the fragments of the target nucleic acid in a single-stranded or double-stranded ligation reaction.

20 In the methods, the barcoded oligonucleotide may comprise a single-stranded 5' or 3' region capable of ligating to a fragment of the target nucleic acid. Each barcoded oligonucleotide may be ligated to a fragment of the target nucleic acid in a single-stranded ligation reaction.

In the methods, the barcoded oligonucleotides may comprise a blunt, recessed, or overhanging 5'
25 or 3' region capable of ligating to a fragment of the target nucleic acid. Each barcoded oligonucleotide may be ligated to a fragment of the target nucleic acid in a double-stranded ligation reaction.

In the methods, the ends of fragments of the target nucleic acid may be converted into blunt
30 double-stranded ends in a blunting reaction and the barcoded oligonucleotides may comprise a blunt double-stranded end. Each barcoded oligonucleotide may be ligated to a fragment of the target nucleic in a blunt-end ligation reaction.

In the methods, the ends of fragments of the target nucleic acid may have their ends converted
35 into blunt double-stranded ends in a blunting reaction, and then have their ends converted into a form with single 3' adenosine overhangs, and wherein the barcoded oligonucleotides comprise a double-stranded end with a single 3' thymine overhang capable of annealing to the single 3' adenosine overhangs of the fragments of the target nucleic acid. Each barcoded oligonucleotide may be ligated to a fragment of the target nucleic acid in a double-stranded A/T ligation reaction.

In the methods, the ends of fragments of the target nucleic acid may be contacted with a restriction enzyme, wherein the restriction enzyme digests each fragment at restriction sites to create ligation junctions at these restriction sites, and wherein the barcoded oligonucleotides
5 comprise an end compatible with these ligation junctions. Each barcoded oligonucleotide may be ligated to a fragment of the target nucleic acid at said ligation junctions in a double-stranded ligation reaction. Optionally, said restriction enzyme may be EcoRI, HindIII, or BglII.

10 In the methods, prior to the step of annealing or ligating the first and second barcoded oligonucleotides to first and second fragments of the target nucleic acid, the method may comprise appending a coupling sequence to each of the fragments of the target nucleic acid, wherein the first and second barcoded oligonucleotides are then annealed or ligated to the coupling sequences of the first and second fragments of the target nucleic acid..

15 In the methods, step (b) may comprise: (i) annealing the first and second barcoded oligonucleotides of the first multimeric barcoding reagent to first and second fragments of the target nucleic acid of the first microparticle, and annealing the first and second barcoded oligonucleotides of the second multimeric barcoding reagent to first and second fragments of the target nucleic acid of the second microparticle; and
20 (ii) extending the first and second barcoded oligonucleotides of the first multimeric barcoding reagent to produce first and second different barcoded target nucleic acid molecules and extending the first and second barcoded oligonucleotides of the second multimeric barcoding reagent to produce first and second different barcoded target nucleic acid molecules, wherein each of the barcoded target nucleic acid molecules comprises at least one nucleotide synthesised
25 from the fragments of the target nucleic acid as a template.

The method may comprise: (a) contacting the sample with a library comprising at least two multimeric barcoding reagents, wherein each multimeric barcoding reagent comprises first and second barcoded oligonucleotides linked together, wherein the barcoded oligonucleotides each
30 comprise in the 5' to 3' direction a target region and a barcode region, wherein the barcode regions of the first and second barcoded oligonucleotides of a first multimeric barcoding reagent of the library are different to the barcode regions of the first and second barcoded oligonucleotides of a second multimeric barcoding reagent of the library, and wherein the sample is further contacted with first and second target primers for each multimeric barcoding reagent;
35 and (b) performing the following steps for each microparticle (i) annealing the target region of the first barcoded oligonucleotide to a first sub-sequence of a first fragment of the target nucleic acid (e.g. genomic DNA) of the microparticle, and annealing the target region of the second barcoded oligonucleotide to a first sub-sequence of a second fragment of the target nucleic acid (e.g. genomic DNA) of the microparticle, (ii) annealing the first target primer to a second sub-sequence

of the first fragment of the target nucleic acid of the microparticle, wherein the second sub-sequence is 3' of the first sub-sequence, and annealing the second target primer to a second sub-sequence of the second fragment of the target nucleic acid of the microparticle, wherein the second sub-sequence is 3' of the first sub-sequence, (iii) extending the first target primer using the first fragment of the target nucleic acid of the microparticle as template until it reaches the first sub-sequence to produce a first extended target primer, and extending the second target primer using the second fragment of the target nucleic acid of the microparticle until it reaches the first sub-sequence to produce a second extended target primer, and (iv) ligating the 3' end of the first extended target primer to the 5' end of the first barcoded oligonucleotide to produce a first barcoded target nucleic acid molecule, and ligating the 3' end of the second extended target primer to the 5' end of the second barcoded oligonucleotide to produce a second barcoded target nucleic acid molecule, wherein the first and second barcoded target nucleic acid molecules are different and each comprises at least one nucleotide synthesised from the target nucleic acid as a template.

The multimeric barcoding reagents may each comprise: (i) first and second hybridization molecules linked together, wherein each of the hybridization molecules comprises a nucleic acid sequence comprising a hybridization region; and (ii) first and second barcoded oligonucleotides, wherein the first barcoded oligonucleotide is annealed to the hybridization region of the first hybridization molecule and wherein the second barcoded oligonucleotide is annealed to the hybridization region of the second hybridization molecule.

The multimeric barcoding reagents may each comprise: (i) first and second barcode molecules linked together, wherein each of the barcode molecules comprises a nucleic acid sequence comprising a barcode region; and (ii) first and second barcoded oligonucleotides, wherein the first barcoded oligonucleotide comprises a barcode region annealed to the barcode region of the first barcode molecule, and wherein the second barcoded oligonucleotide comprises a barcode region annealed to the barcode region of the second barcode molecule.

The invention provides a method of preparing a sample for sequencing, wherein the sample comprises at least two microparticles originating from blood, wherein each microparticle comprises at least two fragments of a target nucleic acid, and wherein the method comprises the steps of: (a) contacting the sample with a library comprising first and second multimeric barcoding reagents, wherein each multimeric barcoding reagent comprises first and second barcode molecules linked together, wherein each of the barcode molecules comprises a nucleic acid sequence comprising, optionally in the 5' to 3' direction, a barcode region and an adapter region; (b) appending a coupling sequence to first and second fragments of the target nucleic acid (e.g. genomic DNA) of first and second microparticles; (c) for each of the multimeric barcoding reagents, annealing the coupling sequence of the first fragment to the adapter region of the first

- barcode molecule, and annealing the coupling sequence of the second fragment to the adapter region of the second barcode molecule; and (d) for each of the multimeric barcoding reagents, appending barcode sequences to each of the at least two fragments of the target nucleic acid of the microparticle to produce first and second different barcoded target nucleic acid molecules, wherein the first barcoded target nucleic acid molecule comprises the nucleic acid sequence of the barcode region of the first barcode molecule and the second barcoded target nucleic acid molecule comprises the nucleic acid sequence of the barcode region of the second barcode molecule.
- 10 In the method, each of the barcode molecules may comprise a nucleic acid sequence comprising, in the 5' to 3' direction, a barcode region and an adapter region, and wherein step (d) comprises, for each of the multimeric barcoding reagents, extending the coupling sequence of the first fragment using the barcode region of the first barcode molecule as a template to produce a first barcoded target nucleic acid molecule, and extending the coupling sequence of the second
- 15 fragment using the barcode region of the second barcode molecule as a template to produce a second barcoded target nucleic acid molecule, wherein the first barcoded target nucleic acid molecule comprises a sequence complementary to the barcode region of the first barcode molecule and the second barcoded target nucleic acid molecule comprises a sequence complementary to the barcode region of the second barcode molecule.
- 20 In the method, each of the barcode molecules may comprise a nucleic acid sequence comprising, in the 5' to 3' direction, an adapter region and a barcode region, wherein step (d) comprises, for each of the multimeric barcoding reagents, (i) annealing and extending a first extension primer using the barcode region of the first barcode molecule as a template to produce a first barcoded
- 25 oligonucleotide, and annealing and extending a second extension primer using the barcode region of the second barcode molecule as a template to produce a second barcoded oligonucleotide, wherein the first barcoded oligonucleotide comprises a sequence complementary to the barcode region of the first barcode molecule and the second barcoded oligonucleotide comprises a sequence complementary to the barcode region of the second barcode molecule,
- 30 (ii) ligating the 3' end of the first barcoded oligonucleotide to the 5' end of the coupling sequence of the first fragment to produce a first barcoded target nucleic acid molecule and ligating the 3' end of the second barcoded oligonucleotide to the 5' end of the coupling sequence of the second fragment to produce a second barcoded target nucleic acid molecule.
- 35 In the method, each of the barcode molecules may comprise a nucleic acid sequence comprising, in the 5' to 3' direction, an adapter region, a barcode region and a priming region wherein step (d) comprises, for each of the multimeric barcoding reagents, (i) annealing a first extension primer to the priming region of the first barcode molecule and extending the first extension primer using the barcode region of the first barcode molecule as a template to produce a first barcoded

oligonucleotide, and annealing a second extension primer to the priming region of the second barcode molecule and extending the second extension primer using the barcode region of the second barcode molecule as a template to produce a second barcoded oligonucleotide, wherein the first barcoded oligonucleotide comprises a sequence complementary to the barcode region of the first barcode molecule and the second barcoded oligonucleotide comprises a sequence complementary to the barcode region of the second barcode molecule, and (ii) ligating the 3' end of the first barcoded oligonucleotide to the 5' end of the coupling sequence of the first fragment to produce a first barcoded target nucleic acid molecule and ligating the 3' end of the second barcoded oligonucleotide to the 5' end of the coupling sequence of the second fragment to produce a second barcoded target nucleic acid molecule.

The method may comprise: (a) contacting the sample with a library comprising first and second multimeric barcoding reagents, wherein each multimeric barcoding reagent comprises first and second barcode molecules linked together, wherein each of the barcode molecules comprises a nucleic acid sequence comprising, in the 5' to 3' direction, a barcode region and an adapter region, and wherein the sample is further contacted with first and second adapter oligonucleotides for each of the multimeric barcoding reagents, wherein the first and second adapter oligonucleotides each comprise an adapter region, and; (b) ligating the first and second adapter oligonucleotides for the first multimeric barcoding reagent to first and second fragments of the target nucleic acid of the first microparticle, and ligating the first and second adapter oligonucleotides for the second multimeric barcoding reagent to first and second fragments of the target nucleic acid of the second microparticle; (c) for each of the multimeric barcoding reagents, annealing the adapter region of the first adapter oligonucleotide to the adapter region of the first barcode molecule, and annealing the adapter region of the second adapter oligonucleotide to the adapter region of the second barcode molecule; and (d) for each of the multimeric barcoding reagents, extending the first adapter oligonucleotide using the barcode region of the first barcode molecule as a template to produce a first barcoded target nucleic acid molecule, and extending the second adapter oligonucleotide using the barcode region of the second barcode molecule as a template to produce a second barcoded target nucleic acid molecule, wherein the first barcoded target nucleic acid molecule comprises a sequence complementary to the barcode region of the first barcode molecule and the second barcoded target nucleic acid molecule comprises a sequence complementary to the barcode region of the second barcode molecule.

The method may comprise the steps of: (a) contacting the sample with a library comprising first and second multimeric barcoding reagents, wherein each multimeric barcoding reagent comprises: (i) first and second barcode molecules linked together, wherein each of the barcode molecules comprises a nucleic acid sequence comprising, optionally in the 5' to 3' direction, an adapter region and a barcode region, and (ii) first and second barcoded oligonucleotides, wherein the first barcoded oligonucleotide comprises a barcode region annealed to the barcode

region of the first barcode molecule, wherein the second barcoded oligonucleotide comprises a barcode region annealed to the barcode region of the second barcode molecule, and wherein the barcode regions of the first and second barcoded oligonucleotides of the first multimeric barcoding reagent of the library are different to the barcode regions of the first and second barcoded oligonucleotides of the second multimeric barcoding reagent of the library; wherein the sample is further contacted with first and second adapter oligonucleotides for each of the multimeric barcoding reagents, wherein the first and second adapter oligonucleotides each comprise an adapter region; (b) annealing or ligating the first and second adapter oligonucleotides for the first multimeric barcoding reagent to first and second fragments of the target nucleic acid (e.g. genomic DNA) of the first microparticle, and annealing or ligating the first and second adapter oligonucleotides for the second multimeric barcoding reagent to first and second fragments of the target nucleic acid (e.g. genomic DNA) of the second microparticle; (c) for each of the multimeric barcoding reagents, annealing the adapter region of the first adapter oligonucleotide to the adapter region of the first barcode molecule, and annealing the adapter region of the second adapter oligonucleotide to the adapter region of the second barcode molecule; and (d) for each of the multimeric barcoding reagents, ligating the 3' end of the first barcoded oligonucleotide to the 5' end of the first adapter oligonucleotide to produce a first barcoded target nucleic acid molecule and ligating the 3' end of the second barcoded oligonucleotide to the 5' end of the second adapter oligonucleotide to produce a second barcoded target nucleic acid molecule.

In the method, step (b) may comprise annealing the first and second adapter oligonucleotides for the first multimeric barcoding reagent to first and second fragments of the target nucleic acid (e.g. genomic DNA) of the first microparticle, and annealing the first and second adapter oligonucleotides for the second multimeric barcoding reagent to first and second fragments of the target nucleic acid (e.g. genomic DNA) of the second microparticle, and wherein either: (i) for each of the multimeric barcoding reagents, step (d) comprises ligating the 3' end of the first barcoded oligonucleotide to the 5' end of the first adapter oligonucleotide to produce a first barcoded-adapter oligonucleotide and ligating the 3' end of the second barcoded oligonucleotide to the 5' end of the second adapter oligonucleotide to produce a second barcoded-adapter oligonucleotide, and extending the first and second barcoded-adapter oligonucleotides to produce first and second different barcoded target nucleic acid molecules each of which comprises at least one nucleotide synthesised from the fragments of the target nucleic acid as a template, or (ii) for each of the multimeric barcoding reagents, before step (d), the method comprises extending the first and second adapter oligonucleotides to produce first and second different target nucleic acid molecules each of which comprises at least one nucleotide synthesised from the fragments of the target nucleic acid as a template.

In the methods, prior to the step of annealing or ligating the first and second adapter oligonucleotides to first and second fragments of the target nucleic acid, the method may comprise appending a coupling sequence to each of the fragments of the target nucleic acid, wherein the first and second adapter oligonucleotides are then annealed or ligated to the coupling sequences of the first and second fragments of the target nucleic acid.

In any method described herein, the method may comprise a step of cross-linking the fragments of the target nucleic acid (e.g. genomic DNA) in the microparticle(s). The step may be performed with a chemical crosslinking agent e.g. formaldehyde, paraformaldehyde, glutaraldehyde, disuccinimidyl glutarate, ethylene glycol bis(succinimidyl succinate), a homobifunctional crosslinker, or a heterobifunctional crosslinker. This step may be performed before any permeabilisation step, after any permeabilisation step, before any partitioning step, before any step of appending coupling sequences, after any step of appending coupling sequences, before any step of appending barcode sequences (e.g. before a step (b)), after any step of appending barcode sequences (e.g. after a step (d)), whilst appending barcode sequences, or any combination thereof. For example, prior to contacting a sample comprising microparticles with a library of two or more multimeric barcoding reagents, the sample comprising microparticles may be crosslinked. Any such crosslinking step may further be ended by a quenching step, such as quenching a formaldehyde-crosslinking step by mixing with a solution of glycine. Any such crosslinks may be removed prior to specific subsequent steps of the protocol, such as prior to a primer-extension, PCR, or nucleic acid purification step.

In the methods, during step (b), (c) and/or (d) (i.e. the steps of appending the barcode sequences), the microparticles and/or fragments of the target nucleic acid may be contained within a gel or hydrogel, such as an agarose gel, a polyacrylamide gel, or any covalently crosslinked gel, such as a covalently crosslinked poly (ethylene glycol) gel, or a covalently crosslinked gel comprising a mixture of a thiol-functionalised poly (ethylene glycol) and an acrylate-functionalised poly (ethylene glycol).

In any method described herein, optionally after any step of cross-linking, the method may comprise permeabilising the microparticle(s). The microparticles may be permeabilised with an incubation step. The incubation step may be performed in the presence of a chemical surfactant. Optionally this permeabilisation step may take place before appending barcode sequences (e.g. before step (b)), after appending barcode sequences (e.g. after step (d)), or both before and after appending barcode sequences. The incubation step may be performed at a temperature of at least 20 degrees Celsius, at least 30 degrees Celsius, at least 37 degrees Celsius, at least 45 degrees Celsius, at least 50 degrees Celsius, at least 60 degrees Celsius, at least 65 degrees Celsius, at least 70 degrees Celsius, or at least 80 degrees Celsius. The incubation step may be at least 1 second long, at least 5 seconds long, at least 10 seconds long, at least 30 seconds

long, at least 1 minute long, at least 5 minutes long, at least 10 minutes long, at least 30 minutes long, at least 60 minutes long, or at least 3 hours long. This step may be performed after any crosslinking step, before any permeabilisation step, after any permeabilisation step, before any partitioning step, before any step of appending coupling sequences, after any step of appending coupling sequences, before any step of appending barcode sequences (e.g. before step (b)), after any step of appending barcode sequences (e.g. after step (d)), whilst appending barcode sequences, or any combination thereof. For example, prior to contacting a sample comprising microparticles with a library of two or more multimeric barcoding reagents, the sample comprising microparticles may be crosslinked, and then permeabilised in the presence of a chemical surfactant.

In any of the methods described herein, the sample of microparticles may be digested with a proteinase digestion step, such as a digestion with a Proteinase K enzyme. Optionally, this proteinase digestion step may be at least 10 seconds long, at least 30 seconds long, at least 60 seconds long, at least 5 minutes long, at least 10 minutes long, at least 30 minutes long, at least 60 minutes long, at least 3 hours long, at least 6 hours long, at least 12 hours long, or at least 24 hours long. This step may be performed after any crosslinking step, before any permeabilisation step, after any permeabilisation step, before any partitioning step, before any step of appending coupling sequences, after any step of appending couplings sequences, before any step of appending barcode sequences (e.g. before step (b)), after any step of appending barcode sequences (e.g. after step (d)), whilst appending barcode sequences, or any combination thereof. For example, prior to contacting a sample comprising microparticles with a library of two or more multimeric barcoding reagents, the sample comprising microparticles may be crosslinked, and then partially digested with a Proteinase K digestion step.

In the methods, steps (a) and (b), and optionally (c) and (d), may be performed on the at least two microparticles in a single reaction volume.

The method may further comprise, prior to step (b), the step of partitioning the nucleic acid sample into at least two different reaction volumes.

The invention provides a method of analysing a sample comprising a microparticle originating from blood, wherein the microparticle contains at least two fragments of a target nucleic acid (e.g. genomic DNA), and wherein the method comprises: (a) preparing the sample for sequencing comprising: (i) contacting the sample with a multimeric barcoding reagent comprising first and second barcode regions linked together, wherein each barcode region comprises a nucleic acid sequence, and (ii) appending barcode sequences to each of the at least two fragments of the target nucleic acid of the microparticle to produce first and second different barcoded target

- 82 -

nucleic acid molecules, wherein the first barcoded target nucleic acid molecule comprises the nucleic acid sequence of the first barcode region and the second barcoded target nucleic acid molecule comprises the nucleic acid sequence of the second barcode region; and (b) sequencing each of the barcoded target nucleic acid molecules to produce at least two linked sequence
5 reads.

In the methods, prior to the step of appending barcode sequences to each of the at least two fragments of genomic DNA of the microparticle, the method may comprise appending a coupling sequence to each of the fragments of genomic DNA of the microparticle, wherein a barcode
10 sequence is then appended to the coupling sequence of each of the at least two fragments of genomic DNA of the microparticle to produce the first and second different barcoded target nucleic acid molecules.

During step (a) the microparticles and/or fragments of the target nucleic acid may be contained
15 within a gel or hydrogel, such as an agarose gel, a polyacrylamide gel, or any covalently crosslinked gel, such as a covalently crosslinked poly (ethylene glycol) gel, or a covalently crosslinked gel comprising a mixture of a thiol-functionalised poly (ethylene glycol) and an acrylate-functionalised poly (ethylene glycol).

The sample of microparticles may be digested with a proteinase digestion step, such as a digestion with a Proteinase K enzyme. Optionally, this proteinase digestion step may be at least 10 seconds long, at least 30 seconds long, at least 60 seconds long, at least 5 minutes long, at least 10 minutes long, at least 30 minutes long, at least 60 minutes long, at least 3 hours long, at least 6 hours long, at least 12 hours long, or at least 24 hours long. This step may be performed
20 before permeabilisation, after permeabilisation, before appending barcode sequences (e.g. before step (a)(ii)), after appending barcode sequences (e.g. after step (a)(ii)), whilst appending barcode sequences, or any combination thereof.
25

Step (a) of the method may be performed by any of the methods of preparing a sample (or nucleic acid sample) for sequencing described herein.
30

The method may comprise preparing first and second samples for sequencing, wherein each sample comprises at least one microparticle originating from blood, wherein each microparticle contains at least two fragments of a target nucleic acid (e.g. genomic DNA), and wherein the barcode sequences each comprise a sample identifier region, and wherein the method
35 comprises: (i) performing step (a) for each sample, wherein the barcode sequence(s) appended to the fragments of the nucleic acid from the first sample have a different sample identifier region to the barcode sequence(s) appended to the fragments of the target nucleic acid from the second sample; (ii) performing step (b) for each sample, wherein each sequence read comprises the

sequence of the sample identifier region; and (iii) determining the sample from which each sequence read is derived by its sample identifier region.

5 The method may comprise analysing a sample comprising at least two microparticles originating from blood, wherein each microparticle contains at least two fragments of a target nucleic acid (e.g. genomic DNA), and wherein the method comprises the steps of: (a) preparing the sample for sequencing comprising: (i) contacting the sample with a library of multimeric barcoding reagents comprising a multimeric barcoding reagent for each of the two or more microparticles, wherein each multimeric barcoding reagent is as defined herein; and (ii) appending barcode
10 sequences to each of the at least two fragments of the target nucleic acid of each microparticle, wherein at least two barcoded target nucleic acid molecules are produced from each of the at least two microparticles, and wherein the at least two barcoded target nucleic acid molecules produced from a single microparticle each comprise the nucleic acid sequence of a barcode region from the same multimeric barcoding reagent; and (b) sequencing each of the barcoded
15 target nucleic acid molecules to produce at least two linked sequence reads for each microparticle.

The barcode sequences may be appended to the fragments of genomic DNA of the microparticles in a single reaction volume i.e. step (a) of the method may be performed in a single reaction
20 volume.

Prior to the step of appending (step (a)(ii)), the method may further comprise the step of partitioning the sample into at least two different reaction volumes.

25 In any of the methods, prior to the step of appending barcode sequences, the multimeric barcoding reagents may separate, fractionate, or dissolve into two or more constituent parts e.g. releasing barcoded oligonucleotides.

In any of the methods, the multimeric barcoding reagents may be at a concentration of less than
30 1.0 femtomolar, less than 10 femtomolar, less than 100 femtomolar, less than 1.0 picomolar, less than 10 picomolar, less than 100 picomolar, less than 1 nanomolar, less than 10 nanomolar, less than 100 nanomolar, or less than 1.0 micromolar.

6. LINKING BY LINKING FRAGMENTS TOGETHER

35 The invention provides a method of analysing a sample comprising a microparticle originating from blood, wherein the microparticle contains at least two fragments of a target nucleic acid (e.g. genomic DNA), and wherein the method comprises: (a) preparing the sample for sequencing comprising linking together at least two fragments of the target nucleic acid of the microparticle to produce a single nucleic acid molecule comprising the sequences of the at least two fragments of

– 84 –

the target nucleic acid; and (b) sequencing each of the fragments in the single nucleic acid molecule to produce at least two linked sequence reads.

5 The at least two fragments of the target nucleic acid (e.g. genomic DNA) may be contiguous in the single nucleic acid molecule.

The at least two linked sequence reads may be provided within a single raw sequence read.

10 The method may comprise, prior to the step of linking, appending a coupling sequence to at least one of the fragments of the target nucleic acid (e.g. genomic DNA) and then linking together the at least two fragments of the target nucleic acid by the coupling sequence.

15 The fragments of the target nucleic acid (e.g. genomic DNA) may be linked together by a solid support, wherein two or more fragments are linked to the same solid support (directly or indirectly e.g. via a coupling sequence). Optionally, the solid support is a bead, such as a Styrofoam bead, a superparamagnetic bead, or an agarose bead.

20 The fragments of the target nucleic acid (e.g. genomic DNA) may be linked together by a ligation reaction e.g. a double-stranded ligation reaction or a single-stranded ligation reaction

The ends of fragments of a target nucleic acid may be converted into blunt, ligatable double-stranded ends in a blunting reaction, and the method may comprise ligating two or more of the fragments to each other by a blunt-end ligation reaction.

25 The ends of fragments of a target nucleic acid may be contacted with a restriction enzyme, wherein the restriction enzyme digests the fragments at restriction sites to create ligation junctions at these restriction sites, and wherein the method may comprise ligating two or more of the fragments to each other by a ligation reaction at the ligation junctions. Any target nucleic acid may be contacted with a restriction enzyme, wherein the restriction enzyme digests the fragments at
30 restriction sites to create ligation junctions at these restriction sites, and wherein the method may comprise ligating two or more of the fragments to each other by a ligation reaction at the ligation junctions. Optionally, said restriction enzyme may be EcoRI, HindIII, or BglII.

35 A coupling sequence may be appended to two or more fragments of a target nucleic acid prior to linking together the fragments. Optionally, two or more different coupling sequences are appended to a population of fragments of the target nucleic acid.

The coupling sequence may comprise a ligation junction on at least one end, and wherein a first coupling sequence is appended to a first fragment of the target nucleic acid, and wherein a

second coupling sequence is appended to a second fragment of the target nucleic acid, and wherein the two coupling sequences are ligated to each other, thus linking together the two fragments of the target nucleic acid.

- 5 The coupling sequence may comprise an annealing region on at least one 3' end, and wherein a first coupling sequence is appended to a first fragment of the target nucleic acid, and wherein a second coupling sequence is appended to a second fragment of the target nucleic acid, and wherein the two coupling sequences are complementary to and annealed to each other along a segment at least one nucleotide in length, and wherein a DNA polymerase is used to extend at
10 least one of the 3' ends of a first coupling sequence at least one nucleotide into the sequence of the second fragment of the target nucleic acid, thus linking together the two fragments of the target nucleic acid (e.g. genomic DNA).

- 15 Prior to linking together the at least two fragments, the method may further comprise a step of cross-linking the microparticles e.g. with a chemical crosslinking agent, such as formaldehyde, paraformaldehyde, glutaraldehyde, disuccinimidyl glutarate, ethylene glycol bis(succinimidyl succinate), a homobifunctional crosslinker, or a heterobifunctional crosslinker.

- 20 Prior to linking together the at least two fragments, the method may further comprise partitioning the microparticles into two or more partitions.

- 25 The method may further comprise permeabilizing the microparticles during an incubation step. This step may be performed before partitioning (if performed), after partitioning (if performed), before linking together the fragments and/or after linking together the fragments.

- 30 The incubation step may be performed in the presence of a chemical surfactant, such as Triton X-100 ($C_{14}H_{22}O(C_2H_4O)_n$ ($n=9-10$)), NP-40, Tween 20, Tween 80, Saponin, Digitonin, or Sodium dodecyl sulfate.

- 35 The incubation step is performed at a temperature of at least 20 degrees Celsius, at least 30 degrees Celsius, at least 37 degrees Celsius, at least 45 degrees Celsius, at least 50 degrees Celsius, at least 60 degrees Celsius, at least 65 degrees Celsius, at least 70 degrees Celsius, at least 80 degrees Celsius, at least 90 degrees Celsius, or at least 95 degrees Celsius.

- 40 The incubation step may be at least 1 second long, at least 5 seconds long, at least 10 seconds long, at least 30 seconds long, at least 1 minute long, at least 5 minutes long, at least 10 minutes long, at least 30 minutes long, at least 60 minutes long, or at least 3 hours long.

The method may comprise digesting the sample of microparticles with a proteinase digestion step, such as a digestion with a Proteinase K enzyme. Optionally, this proteinase digestion step may be at least 10 seconds long, at least 30 seconds long, at least 60 seconds long, at least 5 minutes long, at least 10 minutes long, at least 30 minutes long, at least 60 minutes long, at least 3 hours long, at least 6 hours long, at least 12 hours long, or at least 24 hours long. This step may be performed before partitioning (if performed), after partitioning (if performed), before linking together the fragments and/or after linking together the fragments.

The method may comprise amplifying (original) fragments of a target nucleic acid, and then linking together two or more of the resulting nucleic acid molecules.

The step of linking together the fragments may create a concatamerised nucleic acid molecule, comprising at least 3, at least 5, at least 10, at least 50, at least 100, at least 500, or at least 1000 nucleic acid molecules that have been appended to each other into single, contiguous nucleic acid molecules.

The method may be used to produce linked sequence reads for at least 3 microparticles, at least 5 microparticles, at least 10 microparticles, at least 50 microparticles, at least 100 microparticles, at least 1000 microparticles, at least 10,000 microparticles, at least 100,000 microparticles, at least 1,000,000 microparticles, at least 10,000,000 microparticles, at least 100,000,000 microparticles, at least 1,000,000,000 microparticles, at least 10,000,000,000 microparticles, or at least 100,000,000,000 microparticles.

The sample may comprise at least two microparticles originating from blood, wherein each microparticle contains at least two fragments of a target nucleic acid (e.g. genomic DNA), and wherein the method comprises performing step (a) to produce a single nucleic acid molecule comprising the sequences of the at least two fragments of the target nucleic acid for each microparticle, and performing step (b) to produce linked sequence reads for each microparticle.

Before, during, and/or after the step of linking together at least two fragments of the target nucleic acid (e.g. genomic DNA), the method may comprise the step of cross-linking the fragments of the target nucleic acid in the microparticle(s). The cross-linking step may be performed with a chemical crosslinking agent e.g. formaldehyde, paraformaldehyde, glutaraldehyde, disuccinimidyl glutarate, ethylene glycol bis(succinimidyl succinate), a homobifunctional crosslinker, or a heterobifunctional crosslinker.

Before, during, and/or after the step of linking together at least two fragments of the target nucleic acid (e.g. genomic DNA), and/or optionally after the step of cross-linking the fragments of the

target nucleic acid in the microparticle(s), the method comprises the step of permeabilising the microparticle(s).

5 Prior to step (a), the method may further comprises the step of partitioning the nucleic acid sample into at least two different reaction volumes.

10 In one embodiment of a method of linking together at least two fragments of the target nucleic acid of a circulating microparticle to produce a single nucleic acid molecule comprising the sequences of at least two fragments of the target nucleic acid, a sample comprising at least one circulating microparticle (e.g. wherein said sample is obtained and/or purified by any method disclosed herein) is crosslinked at room temperature in a solution of 1% formaldehyde for 10 minutes, and then the formaldehyde crosslinking step is quenched with glycine. The microparticles are pelleted with a centrifugation step (e.g. at 3000xG for 5 minutes) and resuspended in 1X NEBuffer 2 (New England Biolabs) with 1.0% sodium dodecyl sulfate (SDS),
15 and incubated at 45 degrees Celsius for 10 minutes to permeabilise the microparticle(s). The SDS is quenched by addition of Triton X-100, and the solution is incubated with AluI (New England Biolabs) at 37 degrees Celsius overnight to create blunt, ligatable ends. The enzyme is inactivated by addition of SDS to a final concentration of 1.0% and incubation at 65 degrees Celsius for 15 minutes. The SDS is quenched by addition of Triton X-100, and the solution is
20 diluted at least 10-fold in 1X buffer for T4 DNA Ligase, and to a total concentration of DNA of at most 1.0 nanogram of DNA per microliter. The diluted solution is incubated with T4 DNA Ligase overnight at 16 degrees Celsius to ligate together fragments from circulating microparticles. Crosslinks are then reversed and protein components degraded by incubation overnight at 65 degrees Celsius in a solution of Proteinase K. Ligated DNA is then purified (e.g. with a Qiagen spin-column PCR Purification Kit, and/or Ampure XP beads). Illumina sequencing adapter
25 sequences are then appended with a Nextera in vitro transposition method (Illumina; as per manufacturer's protocol), an appropriate number of PCR cycles are performed to amplify the ligated material; and then amplified and purified size-appropriate DNA is sequenced on an Illumina sequencer (e.g. an Illumina NextSeq 500, or a MiSeq) with paired-end reads of at least
30 50 bases each. Each end of the paired-end sequences is mapped independently to the reference human genome to elucidate linked sequence reads (e.g. reads wherein the two ends comprise sequences from different fragments of genomic DNA from a single circulating microparticle).

35 A method of linking together at least two fragments of the target nucleic acid of a microparticle to produce a single nucleic acid molecule comprising the sequences of the at least two fragments of the target nucleic acid may have a variety of unique properties and features that make it desirable as a method for linking sequences from one or more circulating microparticles. In one respect, such methods enable the linking of sequences from circulating microparticles without complex instrumentation (e.g. microfluidics for partitioning-based approaches). Furthermore, the approach

is (broadly) able to be performed in single, individual reactions that could comprise a large number of circulating microparticles (e.g. hundreds, or thousands, or greater numbers), and thus is able to process a large number of circulating microparticles without the need for multiple reactions that may otherwise be necessary, for example, in a combinatorial indexing approach.

- 5 Furthermore, since the method does not necessarily require the use of barcodes and/or multimeric barcoding reagents, it is not limited by the size of barcode libraries (and/or multimeric barcoding reagent libraries) to achieve useful molecular measurement of linked sequences from circulating microparticles.

10 **7. LINKING BY PARTITIONING**

The methods may be performed on a nucleic acid sample comprising at least two microparticles that has been partitioned into at least two different reaction volumes (or partitions).

- 15 In any of the methods, a nucleic acid sample comprising at least two microparticles may be partitioned into at least two different reaction volumes (or partitions). The different reaction volumes (or partitions) may be provided by different reaction vessels (or different physical reaction vessels). The different reaction volumes (or partitions) may be provided by different aqueous droplets e.g. different aqueous droplets within an emulsion or different aqueous droplets on a solid support (e.g. a slide).

- 20 For example, a nucleic acid sample may be partitioned prior to appending barcode sequences to fragments of the target nucleic acid of a microparticle. Alternatively, a nucleic acid sample may be partitioned prior to linking together at least two fragments of the target nucleic acid of a microparticle.

- 25 For any method involving a partitioning step, any steps of the method subsequent to said partitioning step may be performed independently upon each partition, such as any step of appending barcode sequences or appending coupling sequences, or any step of ligating, annealing, primer-extension, or PCR. Reagents (such as oligonucleotides, enzymes, and buffers) may be added directly to each partition. In methods wherein partitions comprise aqueous droplets in an emulsion, such addition steps may be performed via a process of merging aqueous droplets within the emulsion, such as with a microfluidic droplet-merger conduit, and optionally using a mechanical or thermal mixing step.

- 35 The partitions comprise different droplets of aqueous solution within an emulsion, and wherein the emulsion is a water-in-oil emulsion, and wherein droplets are generated by a physical shaking or a vortexing step, or wherein the droplets are generated by the merger of an aqueous solution with an oil solution within a microfluidic conduit or junction.

For methods wherein partitions comprise aqueous droplets within an emulsion, such a water-in-oil emulsion may be generated by any method or tool known in the art. Optionally, this may include commercially available microfluidic systems such as the Chromium system or other systems available from 10X Genomics Inc, digital droplet generators from Raindance Technologies or Bio-
 5 Rad, as well as component-based systems for microfluidic generation and manipulation such as Drop-Seq (Macosko et al., 2015, Cell 161, 1202–1214) and inDrop (Klein et al., 2015, Cell 161, 1187–1201).

The partitions may comprise different physically non-overlapping spatial volumes within a gel or
 10 hydrogel, such as an agarose gel, a polyacrylamide gel, or any covalently crosslinked gel, such as a covalently crosslinked poly (ethylene glycol) gel, or a covalently crosslinked gel comprising a mixture of thiol-functionalised poly (ethylene glycol) molecules and acrylate-functionalised poly (ethylene glycol) molecules.

15 The sample of microparticles may be separated into a total of at least 10, at least 100, at least 1000, at least 10,000, at least 100,000, at least 1,000,000, at least 10,000,000, at least 100,000,000, or at least 1,000,000,000 partitions. Preferably, the solution of microparticles is separated into a total of at least 1000 partitions.

20 The sample of microparticles may be separated into partitions such that an average of less than 0.0001 microparticles, less than 0.001 microparticles, less than 0.01 microparticles, less than 0.1 microparticles, less than 1.0 microparticle, less than 10 microparticles, less than 100 microparticles, less than 1000 microparticles, less than 10,000 microparticles, less than 100,000 microparticles, less than 1,000,000 microparticles, less than 10,000,000 microparticles, or less
 25 than 100,000,000 microparticles are present per partition. Preferably, an average of less than 1.0 microparticle is present per partition.

The solution of microparticles may be separated into partitions such that an average of less than 1.0 attogram of DNA, less than 10 attograms of DNA, less than 100 attograms of DNA, less than
 30 1.0 femtogram of DNA, less than 10 femtograms of DNA, less than 100 femtograms of DNA, less than 1.0 picogram of DNA, less than 10 picograms of DNA, less than 100 picograms of DNA, or less than 1.0 nanogram of DNA is present per partition. Preferably, less than 10 picograms of DNA are present per partition.

35 The partitions may be less than 100 femtoliters, less than 1.0 picoliter, less than 10 picoliters, less than 100 picoliters, less than 1.0 nanoliter, less than 10 nanoliters, less than 100 nanoliters, less than 1.0 microliter, less than 10 microliters, less than 100 microliters, or less than 1.0 milliliter in volume.

Barcode sequences may be provided in each partition. For each of the two or more partitions comprising barcode sequences, the barcode sequences contained therein may comprise multiple copies of the same barcode sequence, or comprise different barcode sequences from the same set of barcode sequences.

5

After the microparticles have been separated into two or more partitions, the microparticles may permeabilised with an incubation step by any of the methods described herein.

10 The sample of microparticles may be digested with a proteinase digestion step, such as a digestion with a Proteinase K enzyme. Optionally, this proteinase digestion step may be at least 10 seconds long, at least 30 seconds long, at least 60 seconds long, at least 5 minutes long, at least 10 minutes long, at least 30 minutes long, at least 60 minutes long, at least 3 hours long, at least 6 hours long, at least 12 hours long, or at least 24 hours long. This step may be performed before partitioning, after partitioning, before appending barcode sequences, after appending
15 barcode sequences and/or whilst appending barcode sequences.

Appending Sequences by Combinatorial Barcoding Processes

A method of appending barcode sequences may comprise at least two steps of a combinatorial barcoding process, wherein a first barcoding step is performed wherein a sample of microparticles
20 is partitioned into two or more partitions, wherein each partition comprises a different barcode sequence or a different set of barcode sequences that are then appended to sequences from fragments of target nucleic acid (e.g. genomic DNA) of microparticles contained within that partition, and wherein the barcoded nucleic acid molecules of at least two partitions are then merged into a second sample mixture, and wherein this second sample mixture is then partitioned
25 into two or more new partitions, wherein each new partition comprises a different barcode sequence or different set of barcode sequences that are then appended to sequences from fragments of the target nucleic acid (e.g. genomic DNA) of microparticles contained within the two or more new partitions.

30 Optionally, a combinatorial barcoding process may comprise a first barcoding step, wherein: A) a first sample mixture comprising at least first and second circulating microparticles is partitioned into at least first and second original partitions (for example, wherein at least a first circulating microparticle from the sample is partitioned into the first original partition, and wherein at least a second circulating microparticle from the sample is partitioned into the second original partition),
35 wherein the first original partition comprises a barcode sequence (or a set of barcode sequences) different to a barcode sequence (or a set of barcode sequences) comprised within the second original partition, and wherein a barcode sequence (or barcode sequences from a set of barcode sequences) comprised within the first original partition is appended to at least first and second fragments of the target nucleic acid of the first circulating microparticle, and wherein a barcode

sequence (or barcode sequences from a set of barcode sequences) comprised within the second original partition is appended to at least first and second fragments of the target nucleic acid of the second circulating microparticle; and wherein at least one circulating microparticle comprised within the first original partition and at least one circulating microparticle comprised within the second original partition are merged to produce a second sample mixture, and a second barcoding step, wherein: B) microparticles comprised within the second sample mixture are partitioned into at least first and second new partitions (for example, wherein at least a first circulating microparticle from the second sample mixture is partitioned into the first new partition, and wherein at least a second circulating microparticle from the second sample mixture is partitioned into the second new partition), wherein the first new partition comprises a barcode sequence (or a set of barcode sequences) different to a barcode sequence (or the set of barcode sequences) comprised within the second new partition, and wherein a barcode sequence (or barcode sequences from a set of barcode sequences) comprised within the first new partition is appended to at least first and second fragments of the target nucleic acid of the first circulating microparticle, and wherein a barcode sequence (or barcode sequences from a set of barcode sequences) comprised within the second new partition is appended to at least first and second fragments of the target nucleic acid of the second circulating microparticle.

Alternative processes for combinatorial barcoding processes are described in PCT/GB2017/053820 which is incorporated herein by reference. Optionally, in any combinatorial barcoding process, one or more steps of chemical crosslinking may be performed, prior to and/or after any step in any combinatorial barcoding process.

Optionally, in any combinatorial barcoding process, in a step following a chemical crosslinking step, crosslinked microparticles may be permeabilised. Further details are provided in PCT/GB2017/053820, which is incorporated herein by reference.

Optionally, in any combinatorial barcoding process, in any one or more step(s) following a chemical crosslinking step, the crosslinks may be partially or fully reversed. Further details are provided in PCT/GB2017/053820, which is incorporated herein by reference.

Optionally, in any combinatorial barcoding process, barcode sequences may be appended by any one or more methods described herein (such as single-stranded ligation, double-stranded ligation, blunt-ended ligation, A-tailed ligation, sticky-end-mediated ligation, hybridisation, hybridisation and extension, hybridisation and extension and ligation, and/or transposition).

Optionally, during any step of any combinatorial barcoding process, at least 2, at least 3, at least 5, at least 10, at least 20, at least 50, at least 100, at least 200, at least 500, at least 1000, at least 2000, at least 5000, at least 10,000, at least 50,000, at least 100,000, at least 500,000, or at least

1,000,000 circulating microparticles may be comprised within a partition (and/or within each of at least first and second partitions; and/or within any larger number of partitions). Preferably, at least 50 circulating microparticles may be comprised within a partition (and/or within each of at least first and second partitions; and/or within any larger number of partitions).

5

Optionally, during any step of any combinatorial barcoding process, at least 2, at least 3, at least 5, at least 10, at least 20, at least 50, at least 100, at least 200, at least 500, at least 1000, at least 2000, at least 5000, at least 10,000, at least 50,000, at least 100,000, at least 500,000, at least 1,000,000, at least 10,000,000, or at least 100,000,000 partitions may be employed (e.g.

10 circulating microparticles may be partitioned into said number(s) of partitions). Preferably, during any step of any combinatorial barcoding process, at least 24 partitions may be employed (e.g. circulating microparticles may be partitioned into said number(s) of partitions).

Optionally, during any step of any combinatorial barcoding process, a sample of microparticles
15 may be separated into partitions such that an average of less than 0.0001 microparticles, less than 0.001 microparticles, less than 0.01 microparticles, less than 0.1 microparticles, less than 1.0 microparticle, less than 10 microparticles, less than 100 microparticles, less than 1000 microparticles, less than 10,000 microparticles, less than 100,000 microparticles, less than 1,000,000 microparticles, less than 10,000,000 microparticles, or less than 100,000,000
20 microparticles are present per partition. Preferably, an average of less than 1.0 microparticle is present per partition.

Optionally, during any step of any combinatorial barcoding process, a solution of microparticles may be separated into partitions such that an average of less than 1.0 attogram of DNA, less than
25 10 attograms of DNA, less than 100 attograms of DNA, less than 1.0 femtogram of DNA, less than 10 femtograms of DNA, less than 100 femtograms of DNA, less than 1.0 picogram of DNA, less than 10 picograms of DNA, less than 100 picograms of DNA, or less than 1.0 nanogram of DNA is present per partition. Preferably, less than 10 picograms of DNA are present per partition.

30 Optionally, during any step of any combinatorial barcoding process, partitions may be less than 100 femtoliters, less than 1.0 picoliter, less than 10 picoliters, less than 100 picoliters, less than 1.0 nanoliter, less than 10 nanoliters, less than 100 nanoliters, less than 1.0 microliter, less than 10 microliters, less than 100 microliters, or less than 1.0 milliliter in volume.

35 Optionally, any combinatorial barcoding process may comprise at least 2, at least 3, at least 4, at least 5, at least 10, at least 20, at least 30, at least 40, at least 50, at least 100, at least 500, or at least 1000 different barcoding steps. Each of the barcoding steps may be as described herein for the first and second barcoding steps.

Optionally, in any combinatorial barcoding process, any one or more partitioning step may comprise stochastic character – for example, an estimated number (rather than an exact or precise number) of circulating microparticles may be partitioned into one or more partitions; i.e., said number(s) of circulating microparticles per partition may be subject to statistical or probabilistic uncertainty (such as subject to Poisson loading and/or distribution statistics).

Optionally, in any combinatorial barcoding process, the set of barcodes appended to a particular sequence (e.g. appended to a sequence of a fragment of genomic DNA; e.g. a set comprising a first barcode appended to said sequence during a first barcoding step and a second barcode appended to said sequence during a second barcoding step) may be employed to link sequences from a single microparticle and/or to link sequences from a set of two or more microparticles. Optionally, in any combinatorial barcoding process, the same set of two (or more than two) barcodes may be appended to a particular sequence (e.g. appended to a sequence of a fragment of genomic DNA) from two or more circulating microparticles (e.g., wherein said two or more circulating microparticles are partitioned into the same series of first and second partitions during the first and second barcoding steps respectively). Optionally, in any combinatorial barcoding process, the same set of two (or more than two) barcodes may be appended to a particular sequence (e.g. appended to a sequence of a fragment of genomic DNA) from only one circulating microparticle (e.g., wherein only one circulating microparticle is partitioned into a specific series of first and second partitions during the first and second barcoding steps respectively).

Optionally, in any combinatorial barcoding process, the number of partitions employed in any one or more barcoding steps, and the number of different barcoding steps, may combinatorically combine such that, on average, each set of two (or more) barcodes is appended to sequences from only one circulating microparticle. Further details are provided in PCT/GB2017/053820, which is incorporated herein by reference.

A combinatorial barcoding process could provide advantages over alternative barcoding processes in the form of reducing the requirement for sophisticated and/or complex equipment to achieve a high number of potential identifying barcode sets for the purposes of appending barcodes to sequences (e.g. from fragments of genomic DNA) from circulating microparticles. For example, a combinatorial barcoding process employing 96 different partitions (as, for example, would be easily implemented with standard 96-well plates used broadly within molecular biology) across two different barcoding steps could achieve a net of $(96 \times 96 =)$ 9216 different barcode sets; which considerably reduces the amount of partitions that would be required to perform such indexing compared with alternative, non-combinatoric approaches. Considerably higher levels of combinatoric indexing resolution could furthermore be achieved by increasing the number of barcoding steps, and/or increasing the number of partitions employed at one or more such barcoding steps. Furthermore, combinatorial barcoding processes may obviate the need for

complex instrumentation – such as, for example, microfluidic instrumentation (such as the 10X Genomics Chromium System) – that is employed for alternative barcoding processes.

8. LINKING BY SPATIAL SEQUENCING OR IN-SITU SEQUENCING OR IN-SITU LIBRARY CONSTRUCTION

5 The invention provides a method of preparing a sample for sequencing, wherein the sample comprises a microparticle originating from blood, and wherein the microparticle contains at least two fragments of a target nucleic acid (e.g. genomic DNA), and wherein the method comprises:
10 (a) preparing the sample for sequencing, wherein the at least two fragments of the target nucleic acid of the microparticle are linked by their proximity to each other on a sequencing apparatus to produce a set of at least two linked fragments of the target nucleic acid; and (b) sequencing each of the linked fragments of the target nucleic acid using the sequencing apparatus to produce at least two linked sequence reads.

15 The nucleic acid sample may comprise at least two microparticles originating from blood, wherein each microparticle contains at least two fragments of a target nucleic acid (e.g. genomic DNA), and wherein the method comprises performing step (a) to produce a set of linked fragments of the target nucleic acid for each microparticle and wherein the fragments of the target nucleic acid of each microparticle are spatially distinct on the sequencing apparatus, and performing step (b) to
20 produce linked sequence reads for each microparticle.

The at least two fragments from a microparticle may hold physical proximity to each other within or on the sequencing apparatus itself, and wherein this physical proximity is known or can be determined or observed by the sequencing apparatus or by or during its operation, and wherein
25 this measure of physical proximity serves to link the at least two sequences.

The methods may comprise sequencing using an in situ library construction process. In the methods, intact or partially intact microparticles from a sample may be placed onto the sequencer, and wherein two or more fragments of the target nucleic acid (e.g. genomic DNA) are processed
30 into sequencing-ready templates within the sequencer i.e. sequencing using an in situ library construction process. In situ library construction is described in Schwartz et al (2012) PNAS 109(46):18749-54).

The methods may comprise in situ sequencing. In the methods, the sample may remain intact (e.g. largely or partially intact), and fragments of the target nucleic acid (e.g. genomic DNA) within microparticles are sequenced directly e.g. using 'FISSEQ' fluorescent in situ sequencing
35 technique method as described in Lee et al. (2014) Science, 343, 6177, 1360-1363).

Optionally, samples of microparticles may be crosslinked with a chemical crosslinker, and then placed within or upon the sequencing apparatus, and then retained in physical proximity to each other. Optionally, two or more fragments of target nucleic acid (e.g. genomic DNA) from a microparticle placed within or upon the sequencing apparatus may then have all or part of their sequence determined by a sequencing process. Optionally, such fragments may be sequenced by a fluorescent in situ sequencing technique, wherein sequences of said fragments are determined by an optical sequencing process. Optionally, one or more coupling, adapter, or amplification sequence may be appended to said fragments of the target nucleic acid. Optionally, said fragments may be amplified in an amplification process, wherein the amplified products remain in physical proximity or in physical contact of the fragments from which they were amplified. Optionally, these amplified products are then sequenced by an optical sequencing process. Optionally, said amplified products are appended to a planar surface, such as a sequencing flowcell. Optionally, said amplified products generated from single fragments each make up a single cluster within a flowcell. Optionally, in any method as above, the distance between any two or more sequenced molecules is known a priori by configuration within the sequencing apparatus, or may be determined or observed during the sequencing process. Optionally, each sequenced molecule is mapped within a field of clusters, or within an array of pixels, wherein the distance between any two or more sequenced molecules is determined by the distance between said clusters or pixels. Optionally, any measure or estimation of distance or proximity may be used to link any two or more determined sequences.

Optionally, sequences determined by any method as above may be further evaluated, wherein a measure of distance or proximity between two or more sequenced molecules is compared to one or more cutoff or threshold values, and only molecules within a particular range, or above or below a particular threshold or cutoff value, are determined to be linked informatically. Optionally, a set of two or more such cutoff or threshold values or ranges thereof may be employed, such that different degrees and/or classes and/or categories of linking for any two or more sequenced molecules may be determined.

9. LINKING BY SEPARATE SEQUENCING PROCESSES

The invention provides a method of preparing a sample for sequencing, wherein the sample comprises a microparticle originating from blood, and wherein the microparticle contains at least two fragments of a target nucleic acid (e.g. genomic DNA), and wherein the method comprises: (a) preparing the sample for sequencing, wherein the at least two fragments of a target nucleic acid (e.g. genomic DNA) of each microparticle are linked by being loaded into a separate sequencing process to produce a set of at least two linked fragments the target nucleic acid; and (b) sequencing each of the linked fragments of the target nucleic acid using the sequencing apparatus to produce a set of at least two linked sequence reads (i.e. a set of at least two linked signals).

The sample may comprise at least two microparticles originating blood, wherein each microparticle contains at least two fragments of a target nucleic acid (e.g. genomic DNA), and the method may comprise performing step (a) to produce linked fragments of the target nucleic acid
5 for each microparticle wherein the at least two fragments of the target nucleic acid of each microparticle are linked by being loaded into a separate sequencing process, and performing step (b) for each sequencing process to produce linked sequence reads for each microparticle.

10 In the methods, fragments of a first single microparticle (or group of microparticles) may be sequenced independently of the fragments of other microparticles, and the resulting sequence reads are linked informatically; fragments contained within a second single microparticle (or group of microparticles) are sequenced independently of the first microparticle or group of microparticles, and the resulting sequence reads are linked informatically.

15 Optionally, first and the second sequencing processes (of all sequencing processes) are conducted with different sequencing instruments, and/or conducted with the same sequencing instrument but at two different times or within two different sequencing processes. Optionally, the first and the second sequencing processes are conducted with the same sequencing instrument but within two different regions, partitions, compartments, conduits, flowcells, lanes, nanopores,
20 microsc scaffold, array of microsc scaffolds, or integrated circuit of the sequencing instrument. Optionally, 3 or more, 10 or more, 1000 or more, 1,000,000 or more, or 1,000,000,000 or more microparticles or groups of microparticles may be linked by the above method.

10. AMPLIFYING ORIGINAL FRAGMENTS PRIOR TO LINKING

25 As would be appreciated by the skilled person, as used herein the term 'fragments' (e.g. 'fragments of genomic DNA', or 'fragments of a target nucleic acid', or 'fragments of genomic DNA of/from a microparticle') refers to the original fragments present in the microparticle, as well as to portions, copies, or amplicons thereof, including copies of only a part of an original fragment (e.g. an amplicon thereof), as well as to modified fragments or copies (e.g. fragments to which a
30 coupling sequence has been appended). For example, the term fragments of genomic DNA refers to the original genomic DNA fragments present in the microparticle and, for example, to DNA molecules that may be prepared from the original genomic DNA fragments by a primer-extension reaction. As a further example, the term fragments of mRNA refers to the original mRNA fragments present in the microparticle and, for example, to cDNA molecules that may be
35 prepared from the original mRNA fragments by reverse transcription. As used herein, 'fragments of a target nucleic acid' also refers to barcoded oligonucleotides (e.g. barcoded oligonucleotides of barcoded affinity probes) and other nucleic acid reagents described herein.

The methods may, prior to the step of appending barcode sequences, further comprise a step of amplifying the original fragments of the target nucleic of a microparticle e.g. by a primer-extension step or a polymerase chain reaction step. Barcode sequences may then be appended to the amplicons or copies of the original fragments of the target nucleic acid using any of the methods described herein.

The primer-extension step or polymerase chain reaction step may be performed using one or more primers that contain a segment of one or more degenerate bases.

The primer-extension step or polymerase chain reaction step may be performed using one or more primers that are specific for a particular target nucleic acid sequence (e.g. a particular target genomic DNA sequence).

The amplification step may be performed by a strand displacing polymerase, such as Phi29 DNA polymerase, or a Bst polymerase or a Bsm polymerase, or modified derivatives of phi29, Bst, or Bsm polymerases. The amplification may be performed by a multiple-displacement amplification reaction and a set of primers containing a region of one or more degenerate bases. Optionally, random hexamer, random heptamer, random octamer, random nonamer, or random decamer primers are used.

The amplification step may comprise extension by a DNA polymerase of a single-stranded nick in a fragment of an original target nucleic acid. The nick may be generated by an enzyme with single-stranded DNA cleavage behaviour, or by a sequence-specific nicking restriction endonuclease.

The amplification step may comprise incorporating at least one or more dUTP nucleotides into a DNA strand synthesized by replicating or amplifying at least a portion of one or more fragments of genomic DNA by a DNA polymerase, and wherein a nick is generated by a uracil-excising enzyme such as a uracil DNA glycosylase enzyme.

The amplification step may comprise the generation of priming sequences upon a nucleic acid comprising a fragment of genomic DNA, wherein the priming sequences are generated by a primase enzyme, such as a Thermus Thermophilus PrimPol polymerase or a TthPrimPol polymerase, and wherein a DNA polymerase is used to copy at least one nucleotide of a sequence of a fragment of genomic DNA using this priming sequence as a primer.

The amplification step may be performed by a linear amplification reaction, such as an RNA amplification process performed through an in vitro transcription process.

The amplification step may be performed by a primer-extension step or a polymerase chain reaction step, and wherein the primer or primers used therefor are universal primers corresponding to one or more universal priming sequence(s). The universal priming sequence(s) may be appended to fragments of genomic DNA by a ligation reaction, by a primer-extension or
5 polymerase chain reaction, or by an in vitro transposition reaction.

11. APPENDING COUPLING SEQUENCES TO FRAGMENTS PRIOR TO LINKING

In any of the methods, barcode sequences may be appended directly or indirectly (e.g. by
10 annealing or ligation) to fragments of a target nucleic acid (e.g. gDNA) of a microparticle. The barcode sequences may be appended to coupling sequences (e.g. synthetic sequences) that are appended to the fragments.

In methods comprising linking together at least two fragments of the target nucleic acid of the
15 microparticle to produce a single nucleic acid molecule, a coupling sequence may first be appended to each of the at least two fragments and the fragments may then be linked together by the coupling sequence.

A coupling sequence may be appended to an original fragment of target nucleic acid of a
20 microparticle or to a copy or amplicon thereof.

A coupling sequence may be added to the 5' end or 3' end of two or more fragments of the nucleic acid sample. In this method, the target regions (of the barcoded oligonucleotides) may
25 comprise a sequence that is complementary to the coupling sequence.

A coupling sequence may be comprised within a double-stranded coupling oligonucleotide or within a single-stranded coupling oligonucleotide. A coupling oligonucleotide may be appended to the target nucleic acid by a double-stranded ligation reaction or a single-stranded ligation reaction. A coupling oligonucleotide may comprise a single-stranded 5' or 3' region capable of
30 ligating to a target nucleic acid and the coupling sequence may be appended to the target nucleic acid by a single-stranded ligation reaction.

A coupling oligonucleotide may comprise a blunt, recessed, or overhanging 5' or 3' region capable of ligating to a target nucleic acid and the coupling sequence may be appended to the
35 target nucleic acid a double-stranded ligation reaction.

The end(s) of a target nucleic acid may be converted into blunt double-stranded end(s) in a blunting reaction, and the coupling oligonucleotide may comprise a blunt double-stranded end,

– 99 –

and wherein the coupling oligonucleotide may be ligated to the target nucleic acid in a blunt-end ligation reaction.

5 The end(s) of a target nucleic acid may be converted into blunt double-stranded end(s) in a blunting reaction, and then converted into a form with (a) single 3' adenosine overhang(s), and wherein the coupling oligonucleotide may comprise a double-stranded end with a single 3' thymine overhang capable of annealing to the single 3' adenosine overhang of the target nucleic acid, and wherein the coupling oligonucleotide is ligated to the target nucleic acid in a double-stranded A/T ligation reaction

10

The target nucleic acid may be contacted with a restriction enzyme, wherein the restriction enzyme digests the target nucleic acid at restriction sites to create (a) ligation junction(s) at the restriction site(s), and wherein the coupling oligonucleotide comprises an end compatible with the ligation junction, and wherein the coupling oligonucleotide is then ligated to the target nucleic acid in a double-stranded ligation reaction.

15

A coupling oligonucleotide may be appended via a primer-extension or polymerase chain reaction step.

20 A coupling oligonucleotide may be appended via a primer-extension or polymerase chain reaction step, using one or more oligonucleotide(s) that comprise a priming segment including one or more degenerate bases.

25 A coupling oligonucleotide may be appended via a primer-extension or polymerase chain reaction step, using one or more oligonucleotide(s) that further comprise a priming or hybridisation segment specific for a particular target nucleic acid sequence.

30 A coupling sequence may be added by a polynucleotide tailing reaction. A coupling sequence may be added by a terminal transferase enzyme (e.g. a terminal deoxynucleotidyl transferase enzyme). A coupling sequence may be appended via a polynucleotide tailing reaction performed with a terminal deoxynucleotidyl transferase enzyme, and wherein the coupling sequence comprises at least two contiguous nucleotides of a homopolymeric sequence.

35 A coupling sequence may comprise a homopolymeric 3' tail (e.g. a poly(A) tail). Optionally, in such methods, the target regions (of the barcoded oligonucleotides) comprise a complementary homopolymeric 3' tail (e.g. a poly(T) tail).

A coupling sequence may be comprised within a synthetic transposome, and may be appended via an in vitro transposition reaction.

A coupling sequence may be appended to a target nucleic acid, and wherein a barcode oligonucleotide is appended to the target nucleic acid by at least one primer-extension step or polymerase chain reaction step, and wherein said barcode oligonucleotide comprises a region of at least one nucleotide in length that is complementary to said coupling sequence. Optionally, this region of complementarity is at the 3' end of the barcode oligonucleotide. Optionally, this region of complementarity is at least 2 nucleotides in length, at least 5 nucleotides in length, at least 10 nucleotides in length, at least 20 nucleotides in length, or at least 50 nucleotides in length.

12. COUPLING MOLECULES AND METHODS OF EMPLOYING COUPLING MOLECULES FOR MICROPARTICLE ANALYSIS

The methods may comprise: (a) appending one or more coupling molecule(s) to one or more target biomolecule(s) of or from said circulating microparticle(s) to create one or more appended coupling molecule(s), and (b) linking one or more barcode sequence(s) to said appended coupling molecule(s) to create one or more barcoded appended coupling molecule(s). Optionally, any such step of linking one or more barcode sequence(s) to said appended coupling molecule(s) may comprise appending one or more barcoded oligonucleotide(s) to said appended coupling molecule(s), optionally wherein said barcoded oligonucleotide(s) are comprised within one or more multimeric barcoding reagents (such as a library of two or more multimeric barcoding reagents).

The methods may comprise: (a) performing one or more step(s) of crosslinking said sample, (b) performing one or more steps of appending one or more coupling molecule(s) to one or more target biomolecule(s) of or from said circulating microparticle(s) to create one or more appended coupling molecule(s), and (c) linking one or more barcode sequence(s) (e.g. barcoded oligonucleotides, such as barcoded oligonucleotides comprised within one or more multimeric barcoding reagents) to said appended coupling molecule(s) to create one or more barcoded appended coupling molecule(s). Optionally, following any step of crosslinking, one or more steps of permeabilising the sample and/or microparticles may be performed. Optionally, following any step of crosslinking, one or more steps of partially or fully reversing the crosslinks may be performed. Optionally, following any step of crosslinking, one or more steps of partially or fully proteinase-digesting the sample may be performed.

Optionally, following any one or more steps of creating one or more barcoded appended coupling molecule(s), the process may optionally further comprise one or more barcode-connecting steps, wherein one or more barcode sequence(s) are appended to one or more target nucleic acid molecule(s). Optionally, any such one or more barcode-connecting steps may comprise a process of annealing and/or ligating one or more barcode sequence(s) within one or more barcoded appended coupling molecule(s) to one or more target nucleic acid molecule(s) within said

barcoded appended coupling molecule(s). Optionally, any one or more barcode-connecting steps may be performed following one or more steps of crosslinking the sample of one or more microparticles, and/or be performed following one or more steps of partially or fully reversing crosslinks and/or be performed following one or more steps of partial or full proteinase digestion.

5

The methods may comprise: (a) performing one or more step(s) of crosslinking said sample, and (optionally) then performing one or more steps of permeabilising said sample, (b) one or more steps of appending one or more coupling molecule(s) to one or more target biomolecules of or from said circulating microparticle(s) to create one or more (singly and/or doubly and/or multiply) appended coupling molecule(s), wherein one or more such target biomolecule(s) comprise a target nucleic acid molecule, (c) one or more steps of linking at least one barcode sequence (e.g. one or more steps of linking at least one barcoded oligonucleotide, such as at least one barcoded oligonucleotide comprised within one or more multimeric barcoding reagents) to said appended coupling molecule(s) to create one or more barcoded appended coupling molecule(s), and (d) performing one or more barcode-connecting steps, wherein a barcode sequence within a barcoded appended coupling molecule(s) is appended to a target nucleic acid molecule within said barcoded appended coupling molecule(s), optionally wherein one or more steps of reversing the crosslinking and/or one or more steps of proteinase digestion are performed prior to and/or during the step (d) of performing one or more barcode-connecting steps, and optionally wherein one or more said barcode-connecting steps comprises one or more steps of annealing and/or ligating one or more barcode sequence(s) within a barcoded appended coupling molecule to one or more target nucleic acid molecules within said barcoded appended coupling molecule.

The methods may comprise two or more steps of appending one or more coupling molecule(s) to one or more target biomolecules of or from said circulating microparticle(s) to create one or more appended coupling molecule(s). The method may comprise one or more steps of appending two or more coupling molecule(s) to each of one or more target biomolecules of or from said circulating microparticle(s) to create one or more multiply-appended coupling molecule(s) (i.e. one or more appended coupling molecules). The method may comprise a first step of appending a first coupling molecule to each of one or more target biomolecules of or from said circulating microparticle(s) to create one or more singly-appended coupling molecule(s), and then a second step of appending a second coupling molecule to each of said singly-appended coupling molecule(s) to create one or more doubly-appended coupling molecule(s) (i.e. one or more appended coupling molecules). Any number of (sequential or simultaneous) steps of appending a coupling molecule to singly and/or doubly and/or multiply-appended coupling molecule(s) may be performed to create one or more multiply-appended coupling molecule(s) (i.e. one or more appended coupling molecules), optionally then followed by one or more steps of reversing crosslinks, and/or optionally then followed by any one or more barcode-connecting steps. Any

step of appending a coupling molecule may comprise appending a coupling molecule directly or indirectly to an appended coupling molecule.

5 The methods may comprise one or more steps of diluting the sample and/or the derived sample and/or any solution and/or reaction mixture, wherein the concentration of nucleic acids (such as DNA and/or RNA) and/or the concentration of polypeptides in the sample, is/are reduced to or reduced below a certain concentration, such as a concentration of less than 1.0 picograms of DNA (and/or RNA and/or protein) per microliter, less than 10 picograms of DNA (and/or RNA and/or protein) per microliter, less than 100 picograms of DNA (and/or RNA and/or protein) per microliter, less than 1.0 nanograms of DNA (and/or RNA and/or protein) per microliter, less than 10 nanograms of DNA (and/or RNA and/or protein) per microliter, less than 100 nanograms of DNA (and/or RNA and/or protein) per microliter, or less than 1000 nanograms of DNA (and/or RNA and/or protein) per microliter. Optionally, any such step(s) of diluting may be performed prior to and/or during and/or following any one or more step(s) and/or process(es) during any method of analysing a sample comprising one or more circulating microparticle(s) and/or a sample derived from one or more circulating microparticle(s). Optionally, any such step of diluting may be performed following any one or more steps of partially or fully reversing crosslinks, and/or following any one or more steps of proteinase digestion, and/or prior to any one or more barcode-connecting steps.

20 Any step(s) of appending one or more coupling molecule(s) to one or more target biomolecule(s) and/or to one or more (singly- and/or doubly- and/or multiply-appended) coupling molecules of or from said circulating microparticle(s), may be performed upon one, or two, or more than two, or all, or any number and/or fraction and/or part of said target biomolecule(s) and/or said (singly- and/or doubly- and/or multiply-appended) coupling molecules. Any step(s) of linking one or more barcode sequence(s) to any one or more appended coupling molecule(s) may be performed upon one, or two, or more than two, or all, or any number and/or fraction and/or part of said appended coupling molecule(s). Any barcode-connecting step(s) (wherein one or more barcode sequence(s) are appended to one or more target nucleic acid molecule(s)) may be performed upon one, or two, or more than two, or all, or any number and/or fraction and/or part of said target nucleic acid molecule(s). Any barcode-connecting step(s) (wherein one or more barcode sequence(s) within one or more barcoded appended coupling molecule(s) is annealed and/or ligated to one or more target nucleic acid molecule(s) within said barcoded appended coupling molecule(s)) may be performed upon one, or two, or more than two, or all, or any number and/or fraction and/or part of said barcoded appended coupling molecule(s).

In any method of appending a coupling molecule, any one or more target biomolecule(s) may comprise any type of target nucleic acid molecule, such as a fragment of genomic DNA, an mRNA molecule or fragment thereof, a microRNA molecule, and/or a barcoded oligonucleotide

(such as a barcoded oligonucleotide within a barcoded affinity probe), and/or any other type of target nucleic acid molecule. Optionally, in any method of appending a coupling molecule, one or more target biomolecule(s) may comprise both one or more fragments of genomic DNA, and one or more oligonucleotides appended to an affinity moiety (i.e. one or more barcoded
 5 oligonucleotides within a barcoded affinity probe).

Optionally, any one or more barcoded appended coupling molecule(s) created during any method of analysing a sample comprising one or more circulating microparticle(s) and/or a sample derived from one or more circulating microparticle(s) may comprise: one or more target
 10 biomolecule(s) (such as a target nucleic acid sequence), one or more (first) coupling molecule(s) appended to said target biomolecule(s) (optionally where said first coupling molecules may each comprise one or more coupling sequences), one or more second or further coupling molecule(s) appended to said (first) coupling molecules (optionally where said second or further coupling molecules may each comprise one or more coupling sequences), one or more linker moieties
 15 (and/or linker molecules) optionally comprised within each of one or more coupling molecules, one or more binding moieties (and/or linker molecules) optionally comprised within each of one or more coupling molecules, and one or more barcode sequences (such as one or more barcoded oligonucleotides) linked to any one or more first, second, and/or further coupling molecules.

20 Optionally, any one or more coupling molecule(s) may comprise one or more coupling sequences. Optionally, any one or more coupling molecule(s) may comprise one or more binding moieties. Optionally, any one or more coupling molecule(s) may comprise one or more linker molecules and/or linker moieties (for example, any one or more linker molecules disposed between a coupling sequence and a binding moiety, or disposed between two different coupling sequences,
 25 or disposed between two different binding moieties). Optionally, any one or more coupling molecule(s) may comprise one or more adapter sequences. Optionally, any one or more coupling molecule(s) may comprise one or more barcode sequences. Optionally, any one or more coupling molecule(s) may comprise one or more barcoded oligonucleotides. Any linker molecule and/or linker moiety may comprise a biopolymer (e.g. a nucleic acid molecule) or a synthetic polymer.

30 Any linker molecule and/or linker moiety may comprise one or more units of ethylene glycol and/or poly(ethylene) glycol (e.g. hexa-ethylene glycol or penta-ethylene glycol). Any linker molecule and/or linker moiety may comprise one or more ethyl groups, such as a C3 (three-carbon) spacer, C6 spacer, C12 spacer, or C18 spacer. Any linker molecule may comprise a sequence or chain (such as a concatenated and/or linear molecular sequence or chain) of two or
 35 more linker moieties in series, such as two or more poly(ethylene) glycol linker moieties, or two or more C12 or C18 spacers; optionally, any linker molecule may comprise at least 3, at least 4, at least 5, at least 10, or at least 20 linker moieties in a chain and/or linear sequence.

Optionally, any one or more coupling molecule(s) may comprise one or more coupling sequences, and further comprise one or more binding moieties, and further comprise one or more linker molecules and/or linker moieties.

- 5 Optionally, any one or more coupling molecule(s) may comprise at least first and second coupling sequences, wherein said first and second coupling sequences are connected to each other by one or more linker molecule(s).

- 10 Optionally, in any method, any one or more target biomolecule (e.g. any one or more target nucleic acid molecule) may have at least 1, at least 2, at least 3, at least 5, at least 10, at least 50, at least 100, or at least 1000 coupling molecule(s) appended and/or linked to it, either directly and/or indirectly, either in linear sequence and/or at multiple sites comprised within said target biomolecule, and optionally involving a method comprising multiple, sequential and/or independent steps of appending/linking coupling molecules to each other (such as at least 2, at
15 least 5, or at least 10 sequential steps of appending/linking a second coupling molecule to one or more first, previously-appended/linked coupling molecule(s)).

- 20 Optionally, a coupling molecule may comprise an oligonucleotide sequence (e.g. a coupling sequence) and a binding moiety, wherein said oligonucleotide sequence and binding moiety are linked covalently or non-covalently. Optionally, a coupling molecule may comprise an oligonucleotide sequence and a binding moiety, wherein said oligonucleotide sequence and binding moiety are linked by a linker moiety (e.g. a linker molecule). Optionally, a coupling molecule may comprise a first oligonucleotide sequence, connected in physical sequence to a linker moiety, and then connected in subsequent physical sequence to a binding moiety.
25 Optionally, a coupling molecule may comprise a first binding moiety, connected in physical sequence to a linker moiety, and then connected in subsequent physical sequence to a second binding moiety.

- 30 Optionally, a coupling molecule may comprise at least a first oligonucleotide sequence and a second oligonucleotide sequence, wherein said at least first and second oligonucleotide sequences are linked by a linker moiety. Optionally, a coupling molecule may comprise an oligonucleotide sequence and two or more binding moieties, wherein said oligonucleotide sequence and binding moieties are linked by a branched linker moiety (such as a branched linker molecule comprising two or more ethyl groups, such as two or more spacer moieties, such as two
35 or more C3 (three-carbon) spacers, and/or C6 spacers, and/or C12 spacers, and/or C18 spacers. Optionally, a coupling molecule may comprise three or more binding moieties, wherein said binding moieties and binding moieties are linked by a branched or multiply-branched linker moiety.

Optionally, any step of appending and/or linking (such as any step of appending a coupling molecule, and/or any step of appending a barcode sequence) may be performed by any method of attachment and/or binding, such as any method of covalent or non-covalent binding, any method of annealing or hybridisation (such as any method of annealing two complementary oligonucleotide sequences to each other, such as annealing a first coupling sequence to a second coupling sequence, or annealing a sequence comprised within a barcoded oligonucleotide to a coupling sequence and/or to an adapter sequence), any method of ligation (such as single-stranded ligation or double-stranded ligation, such as blunt or overhang-mediated double-stranded ligation), or any method of binding a biotin moiety to a streptavidin moiety or streptavidin-related moiety, or any method of binding an affinity moiety to a moiety for which it has affinity (such as any method of binding an antibody to its target epitope), or any click-chemistry-related method (such as any copper(I)-catalysed azide-alkyne cycloaddition (CuAAC) reaction, a strain-promoted azide-alkyne cycloaddition (SPAAC) reaction, a strain-promoted alkyne-nitrone cycloaddition (SPANC) reaction, or an alkene and tetrazole photoclick reaction).

Any one or more binding moiety may comprise any molecule and/or class of molecule and/or macromolecule (and or any one or more parts thereof, e.g. any one or more parts of a molecule and/or macromolecule) that is capable of binding to, and/or has a preferential and/or thermodynamic and/or chemical potential to bind to and/or bind with any one or more other molecule(s) or parts thereof (such as any other binding moiety, or any part(s) of any other binding moiety).

Any one or more binding moieties may comprise any of the following: a biotin moiety, a streptavidin moiety (and/or any moiety comprising a derivative of streptavidin, such as neutravidin or avidin), an azide moiety, an alkyne moiety, an amine moiety (such as a primary amine), an alkene moiety, a trans-cyclooctene moiety, a dibenzocyclooctyne moiety, a tetrazine moiety, a hapten moiety (such as a small-molecule hapten moiety, such as digoxigenin), any form of affinity moiety (such as an antibody, antibody fragment, aptamer such as a DNA aptamer or RNA aptamer), and/or any epitope to which any affinity moiety has any affinity and/or preferential affinity for, an I-Linker (from Integrated DNA Technologies), and/or an acrydite moiety.

13. OPTIONAL ADDITIONAL STEPS OF THE METHODS

The methods may comprise determining the presence or absence of at least one modified nucleotide or nucleobase in one or more fragments of genomic DNA from a sample comprising one or more circulating microparticles. The methods may comprise measurement of the modified nucleotide or nucleobase (e.g. measuring the modified nucleotide or nucleobase) in fragments of genomic DNA of a circulating microparticle. The measured value may be a total value of the analysed fragments of genomic DNA (i.e. linked fragments of genomic DNA) of a circulating microparticle and/or the measured value may be a value for each analysed fragment of genomic

DNA. The modified nucleotide or nucleobase may be 5-methylcytosine or 5-hydroxy-methylcytosine.

- Measurement(s) of modified nucleotides or nucleobases in one or more fragments of genomic DNA from circulating microparticles enables a variety of molecular and informatic analyses that may complement measurement of the sequence of said fragments themselves. In one respect, measurement of so-called 'epigenetic' marks (i.e. measurement of the 'epigenome') within fragments of genomic DNA from circulating microparticles enables comparison to (and/or mapping against) reference epigenetic sequences and/or lists of reference epigenetic sequences. This enables an 'orthogonal' form of analysing sequences from fragments of genomic DNA from circulating microparticles in comparison to measurement only of the standard 4 (unmodified) bases and/or their traditional 'genetic' sequences. Furthermore, measurement of modified nucleotides and/or nucleobases may enable more precision determination and/or estimation of the types of cells and/or tissues from which one or more circulating microparticles have arisen. Since different cell types within the body exhibit different epigenetic signatures, measurement of the epigenome of fragments of genomic DNA from circulating microparticles may therefore allow more precise such microparticle-to-cell type mapping. In the methods, epigenetic measurements from fragments of genomic DNA from circulating microparticles may be compared with (e.g. mapped to) a list (or lists) of reference epigenetic sequences corresponding to methylation and/or hydroxymethylation within particular specific tissues. This may enable the elucidation of and/or enrichment for microparticles (e.g. linked sets of sequences from particular microparticles) from a particular tissue type and/or a particular healthy and/or diseased tissue (e.g. cancer tissue). For example, the measurement of a modified nucleotide or nucleobase in fragments of genomic DNA of a circulating microparticle may enable the identification of linked sequences (or linked sequence reads) of fragments of genomic DNA originating from cancer cells. In a further example, the measurement of a modified nucleotide or nucleobase in fragments of genomic DNA of a circulating microparticle may enable the identification of linked sequences (or linked sequence reads) of fragments of genomic DNA originating from foetal cells. The absolute amount of a particular modified nucleotide or nucleobase may correlate with health and/or disease within a particular tissue. For example, the level of 5-hydroxy-methylcytosine is strongly altered in cancerous tissue compared with normal healthy tissues; measurement of 5-hydroxy-methylcytosine in fragments of genomic DNA from circulating microparticles may therefore enable more precise detection and/or analysis of circulating microparticles originating from cancer cells.
- The methods may comprise measurement of 5-methylcytosine in fragments of genomic DNA of a circulating microparticle (e.g., measuring 5-methylcytosine in fragments of genomic DNA of a circulating microparticle). The methods may comprise measurement of 5-hydroxy-methylcytosine in fragments of genomic DNA of a circulating microparticle (e.g., measuring 5-hydroxy-methylcytosine in fragments of genomic DNA of a circulating microparticle).

The methods may comprise measurement of 5-methylcytosine in fragments of genomic DNA of a circulating microparticle (e.g., measuring 5-methylcytosine in fragments of genomic DNA of a circulating microparticle), wherein said measurement is performed using an enrichment probe that is specific for or preferentially binds 5-methylcytosine in fragments of genomic DNA compared with other modified or unmodified bases. The methods may comprise measurement of 5-hydroxy-methylcytosine in fragments of genomic DNA of a circulating microparticle (e.g., measuring 5-hydroxy-methylcytosine in fragments of genomic DNA of a circulating microparticle), wherein said measurement is performed using an enrichment probe that is specific for or preferentially binds 5-hydroxy-methylcytosine in fragments of genomic DNA compared with other modified or unmodified bases.

The methods may comprise measurement of 5-methylcytosine in fragments of genomic DNA of two or more circulating microparticles (e.g., measuring 5-methylcytosine in fragments of genomic DNA of a first circulating microparticle and measuring 5-methylcytosine in fragments of genomic DNA of a second circulating microparticle). The methods may comprise measurement of 5-hydroxy-methylcytosine in fragments of genomic DNA of two or more circulating microparticles (e.g., measuring 5-hydroxy-methylcytosine in fragments of genomic DNA of a first circulating microparticle and measuring 5-hydroxy-methylcytosine in fragments of genomic DNA of a second circulating microparticle).

The methods may comprise measurement of 5-methylcytosine in fragments of genomic DNA of two or more circulating microparticles (e.g., measuring 5-methylcytosine in fragments of genomic DNA of a first circulating microparticle and measuring 5-methylcytosine in fragments of genomic DNA of a second circulating microparticle), wherein said measurement is performed using an enrichment probe that is specific for or preferentially binds 5-methylcytosine in fragments of genomic DNA compared with other modified or unmodified bases. The methods may comprise measurement of 5-hydroxy-methylcytosine in fragments of genomic DNA of two or more circulating microparticles (e.g., measuring 5-hydroxy-methylcytosine in fragments of genomic DNA of a first circulating microparticle and measuring 5-hydroxy-methylcytosine in fragments of genomic DNA of a second circulating microparticle), wherein said measurement is performed using an enrichment probe that is specific for or preferentially binds 5-hydroxy-methylcytosine in fragments of genomic DNA compared with other modified or unmodified bases.

The methods may comprise measurement of 5-methylcytosine in fragments of genomic DNA of a circulating microparticle (e.g., measuring 5-methylcytosine in fragments of genomic DNA of a circulating microparticle), wherein said measurement is performed using a bisulfite conversion process or an oxidative bisulfite conversion process. The methods may comprise measurement of 5-hydroxy-methylcytosine in fragments of genomic DNA of a circulating microparticle (e.g.,

measuring 5-hydroxy-methylcytosine in fragments of genomic DNA of a circulating microparticle), wherein said measurement is performed using a bisulfite conversion process or an oxidative bisulfite conversion process.

- 5 The methods may comprise measurement of 5-methylcytosine in fragments of genomic DNA of two or more circulating microparticles (e.g., measuring 5-methylcytosine in fragments of genomic DNA of a first circulating microparticle and measuring 5-methylcytosine in fragments of genomic DNA of a second circulating microparticle), wherein said measurement is performed using a bisulfite conversion process or an oxidative bisulfite conversion process. The methods may
- 10 comprise measurement of 5-hydroxy-methylcytosine in fragments of genomic DNA of two or more circulating microparticles (e.g., measuring 5-hydroxy-methylcytosine in fragments of genomic DNA of a first circulating microparticle and measuring 5-hydroxy-methylcytosine in fragments of genomic DNA of a second circulating microparticle), wherein said measurement is performed using a bisulfite conversion process or an oxidative bisulfite conversion process.

- 15 Optionally, sequences from two or more constituent parts of a sample comprising one or more circulating microparticles may be determined as relates to determining the presence or absence of at least one modified nucleotide or nucleobase in one or more fragments of genomic DNA from said sample. For example, an enrichment step may be performed to enrich for fragments of
- 20 genomic DNA within a sample containing a modified base (such as 5-methylcytosine, or 5-hydroxy-methylcytosine), wherein a first constituent part of the sample comprising fragments of genomic DNA that have been enriched by said enrichment step may be sequenced, and a second constituent part of the sample comprising fragments of genomic DNA that have not been enriched by said enrichment step may also be sequenced (e.g. sequenced in a separate sequencing
- 25 reaction). Optionally said second constituent part of the sample may comprise a non-enriched and/or supernatant fraction (e.g. a fraction not bound by an enrichment probe or affinity probe during an enrichment process) produced during the enrichment process. Optionally the original sample may be divided into first and second sub-samples, wherein the first sub-sample is employed to perform an enrichment step to produce the first constituent part of the sample, and
- 30 wherein the said second constituent part of the sample may comprise the second, non-enriched sub-sample. Any combination of two or more enriched and/or unenriched and/or converted (e.g. bisulfite-converted, and/or oxidative bisulfite-converted) and/or unconverted constituent parts of a sample may be sequenced. For example, a sample comprising one or more circulating microparticles may be used to produce three constituent parts, such as a constituent part
- 35 enriched for 5-methylcytosine DNA (or alternatively, a constituent part that has been bisulfite-converted), a constituent part enriched for 5-hydroxy-methylcytosine (or alternatively, a constituent part that has been oxidative-bisulfite-converted), and an unenriched (and/or unconverted) constituent part. Optionally, any such two or more constituent parts of a sample may be sequenced individually in separate sequencing reactions (such as within separate flowcells, or

within separate lanes of a single flowcell). Optionally, any such two or more parts of a sample may be appended to identifying barcode sequences (e.g. which identify a given sequence as being within an enriched or unenriched constituent part of a sample) and then sequenced within the same sequencing process (such as within the same flowcell or lane of a flowcell).

5

Optionally, any method of linking sequences as described herein (for example, by appending barcode sequences, such as by appending barcode sequences from a multimeric barcoding reagent or by appending barcode sequences from a library of two or more multimeric barcoding reagents) may be performed before any such enrichment and/or molecular conversion step (for example, wherein such a linking process is performed on the original sample comprising at least one circulating microparticle, or at least two circulating microparticles, wherein the linked sequences are then used as input sequences for an enrichment or molecular conversion process).

10

15

For example, a sample comprising two or more circulating microparticles may be appended to barcode sequences from a library of two or more multimeric barcoding reagents, wherein first and second barcode sequences from a first multimeric barcoding reagent are appended to first and second fragments of genomic DNA from a first circulating microparticle, and wherein first and second barcode sequences from a second multimeric barcoding reagent are appended to first and second fragments of genomic DNA from a second circulating microparticle, and wherein the resulting barcode-appended fragments of genomic DNA are enriched for 5-methylcytosine (and/or 5-hydroxy-methylcytosine), and wherein the enriched fragments of genomic DNA are then sequenced, wherein the barcode sequences are then used to determine which enriched fragments were appended to barcodes from the same multimeric barcoding reagent(s), and thereby predict (or determine) which enriched fragments were comprised within the same circulating microparticle(s). In this example, a second sequencing reaction may also be performed on unenriched fragments of genomic DNA (for example, by sequencing fragments of genomic DNA within the supernatant fraction (i.e. the non-captured, non-enriched fraction) of the enrichment step, wherein the barcode sequences are then used to determine which unenriched fragments were appended to barcodes from the same multimeric barcoding reagent(s), and thereby predict (or determine) which unenriched fragments were comprised within the same circulating microparticle(s). In this example, if both enriched and unenriched fragments of genomic DNA are so sequenced, it may therefore be predicted (or determined) both which enriched and which unenriched fragments were appended to barcodes from the same multimeric barcoding reagent(s), and thereby be predicted (or determined) both which enriched and which unenriched fragments were comprised within the same circulating microparticle(s). Methods similar to this example may also be employed, for example by employing one or more molecular conversion methods, and/or for example by preparing, analysing, or sequencing three or more

20

25

30

35

- 110 -

constituent parts of a sample (for example, a constituent part enriched for 5-methylcytosine, a constituent part enriched for 5-hydroxy-methylcytosine, and an unenriched constituent part).

Optionally, any method of linking sequences as described herein (for example, by appending
5 barcode sequences, such as by appending barcode sequences from a multimeric barcoding reagent or a library of two or more multimeric barcoding reagents) may be performed after any such enrichment and/or molecular conversion step (for example, wherein an enrichment step is performed to enrich for fragments of genomic DNA containing 5-methylcytosine, or containing 5-hydroxy-methylcytosine, and wherein the fragments of genomic DNA enriched through this
10 process are then linked by any method described herein).

The methods may comprise determining the presence or absence of at least one modified nucleotide or nucleobase in the fragments of genomic DNA, wherein an enrichment step is performed to enrich for fragments of genomic DNA containing said modified base. Such modified
15 base may comprise one or more of 5-methylcytosine, or 5-hydroxy-methylcytosine, or any other modified base. Such an enrichment step may be performed by an enrichment probe, such as an antibody, enzyme, enzyme fragment, or other protein, or an aptamer, or any other probe, that is specific for or preferentially binds with said modified base compared with other modified or unmodified bases. Such an enrichment step may be performed by an enzyme capable of
20 enzymatically modifying DNA molecules containing a modified base, such as a glucosyltransferase enzyme, such as a 5-hydroxymethylcytosine glucosyltransferase enzyme. Optionally, the presence of 5-hydroxymethylcytosine within a fragment of genomic DNA may be determined with a 5-hydroxymethylcytosine glucosyltransferase enzyme, wherein the 5-hydroxymethylcytosine glucosyltransferase enzyme is used to transfer a glucose moiety from
25 uridine diphosphoglucose to the modified base within the fragment of genomic DNA to produce a glucosyl-5-hydroxymethylcytosine base, optionally wherein said glucosyl-5-hydroxymethylcytosine base is then detected, such as being detected with a glucosyl-5-hydroxymethylcytosine-sensitive restriction enzyme, wherein fragments of genomic DNA resistant to digestion by said glucosyl-5-hydroxymethylcytosine-sensitive restriction enzyme are considered to contain a modified 5-hydroxymethylcytosine base; optionally, said fragments of genomic DNA resistant to digestion
30 may be sequenced to determine their sequence(s) by any method described herein. Optionally, if barcode sequences are appended, this enrichment step may be performed before the step of appending barcode sequences or after the step of appending barcode sequences. Optionally, if two or more sequences of fragments of genomic DNA from a microparticle are appended to each other, this enrichment step may be performed before the step of appending such sequences to each other or after the step of appending such sequences to each other. Any method of measuring at least one modified nucleotide or nucleobase in the fragments of genomic DNA using an enrichment probe may be performed with commercially available enrichment probes or other products such as commercially available antibodies, such as the anti-5-hydroxy-methylcytosine

- 111 -

antibody ab178771 (Abcam), or such as the anti-5-methylcytosine antibody ab10805 (Abcam). Furthermore, commercially available products and/or kits may also be used for additional step(s) of such methods, such as Protein A or Protein G Dynabeads (ThermoFisher) for binding, recovery, and processing/washing of antibodies and/or fragments bound thereto.

5

The methods may comprise determining the presence or absence of at least one modified nucleotide or nucleobase in the fragments of genomic DNA, wherein a molecular conversion step is performed to convert said modified base(s) into a different modified or unmodified nucleotide which may be detected during the process of determining a nucleic acid sequence. This conversion step may comprise a bisulfite conversion step, an oxidative bisulfite conversion step, or any other molecular conversion step. Optionally, if barcode sequences are appended, this enrichment step may be performed before the step of appending barcode sequences or after the step of appending barcode sequences. Optionally, if two or more sequences of fragments of genomic DNA from a microparticle are appended to each other, this enrichment step may be performed before the step of appending such sequences to each other or after the step of appending such sequences to each other. Any method of measuring at least one modified nucleotide or nucleobase in the fragments of genomic DNA using a molecular conversion step may be performed with commercially available molecular conversion kits, such as the EpiMark Bisulfite Conversion Kit (New England Biolabs), or the TruMethyl Seq Oxidative Bisulfite Sequencing Kit (Cambridge Epigenetix).

In any method of performing a molecular conversion step, one or more adapter oligonucleotide(s) may be appended to one or both ends of a fragment of genomic DNA (and/or a collection of fragments of genomic DNA within a sample) following the molecular conversion process. For example, a single-stranded adapter oligonucleotide (for example, comprising a binding site for a primer used for amplification, such as by PCR amplification) may be ligated with a single-stranded ligase enzyme to one or both ends of the converted fragment of genomic DNA (and/or a collection of fragments of genomic DNA within a sample). Optionally, a barcode sequence and/or adapter sequence (such as within a barcoded oligonucleotide) may be appended to one end of a fragment of genomic DNA (and/or a collection of fragments of genomic DNA within a sample) prior to a molecular conversion step, and then an adapter oligonucleotide may be appended to a second end of the fragment(s) of genomic DNA following a molecular conversion process. Optionally, said second end may comprise an end created during the molecular conversion process (i.e. wherein the fragment(s) of genomic DNA has/have undergone a fragmentation process, thus creating one or more new ends of said fragment(s) relative to their corresponding original fragment(s). Such methods of appending adapter oligonucleotides may have the benefit of allowing fragments of genomic DNA that have been fragmented and/or degraded during a molecular conversion process to be further amplified and/or analysed and/or sequenced.

In any method of performing a molecular conversion step, any adapter oligonucleotide, and/or barcoded oligonucleotide, and/or barcode sequence, and/or any coupling sequence and/or any coupling oligonucleotide, may comprise one or more synthetic 5-methylcytosine nucleotides. Optionally, any adapter oligonucleotide, and/or barcoded oligonucleotide, and/or barcode sequence, and/or any coupling sequence and/or any coupling oligonucleotide, may be configured such that any or all cytosine nucleotides contained therein are synthetic 5-methylcytosine nucleotides. Optionally, any adapter oligonucleotide, and/or barcoded oligonucleotide, and/or barcode sequence, and/or any coupling sequence and/or any coupling oligonucleotide, comprising one or more synthetic 5-methylcytosine nucleotides, may be appended to fragment(s) of genomic DNA prior to a molecular conversion step; alternatively and/or additionally, they may be appended to fragment(s) of genomic DNA subsequent to a molecular conversion step. Such synthetic 5-methylcytosine nucleotides within said adapter(s) and/or oligonucleotide(s) and/or sequence(s) may have a benefit of reducing or minimising their degradation and/or fragmentation during a molecular conversion process (such as a bisulfite conversion process), due to their resistance to degradation during such a process.

The methods may comprise determining the presence or absence of at least one modified nucleotide or nucleobase in the fragments of genomic DNA, wherein said modified nucleotide or nucleobase (such as 5-methylcytosine or 5-hydroxy-methylcytosine) is determined or detected by a sequencing reaction. Optionally, said sequencing reaction may be performed by a nanopore-based sequencing instrument, such as a Minion, a Gridion X5, a Promethion, and/or a Smidgion sequencing instrument produced by Oxford Nanopore Technologies, wherein the presence of modified nucleotide(s) or nucleobase(s) is determined during the process of translocating a fragment of genomic DNA through a nanopore within the sequencing instrument and by analysing the current signal through the nanopore apparatus during said translocation of the fragment of genomic DNA. Optionally, said sequencing reaction may be performed by a zero-mode-waveguide-based sequencing instrument, such as a Sequel or RSII sequencing instrument produced by Pacific Biosciences, wherein the presence of modified nucleotide(s) or nucleobase(s) is determined during the process of synthesising a copy of at least part of a fragment of genomic DNA within a zero-mode waveguide within the sequencing instrument and by analysing the optical signal derived from said zero-mode waveguide during said process of copying at least a part of the fragment of genomic DNA.

In any method of performing an enrichment step and/or a molecular-conversion step, said enrichment and/or conversion may be incomplete and/or less than 100% efficient. For example, a molecular conversion process may be performed such that less than 100% of a particular class of targeted modified nucleotide (such as 5-methylcytosine, or 5-hydroxy-methylcytosine) are converted with a molecular conversion process (such as bisulfite conversion or oxidative bisulfite conversion). For example, approximately 99%, or approximately 95%, or approximately 90%, or

- 113 -

approximately 80%, or approximately 70%, or approximately 60%, or approximately 50%, or approximately 40%, or approximately 25%, or approximately 10% of such targeted modified nucleotide(s) may be converted during such a molecular conversion process. This incomplete molecular conversion process may be performed by limiting the duration of time for which the molecular conversion process is conducted (e.g., by making said duration of time shorter than the standard time employed to achieve full or near-full efficiency of the molecular conversion process), such that, on average, said target conversion efficiencies are achieved. Such incomplete molecular conversion processes may have a benefit of reducing the amount of sample degradation/fragmentation and/or sample loss that, for example, is characteristic of many molecular conversion processes such as bisulfite conversion.

Similarly, in any method of performing an enrichment step, said enrichment may be incomplete and/or less than 100% efficient. For example, an enrichment step for 5-methylcytosine (and/or 5-hydroxy-methylcytosine) may be performed wherein approximately 99%, or approximately 95%, or approximately 90%, or approximately 80%, or approximately 70%, or approximately 60%, or approximately 50%, or approximately 40%, or approximately 25%, or approximately 10% of fragments of genomic DNA containing such targeted modified nucleotide(s) are captured and recovered during an enrichment step (such as an enrichment step using an affinity probe such as an antibody specific for said targeted modified nucleotide(s)). Optionally, said incomplete enrichment may be performed by limiting and/or reducing the amount and/or concentration of the affinity probe used in the enrichment process (for example, by empirically testing the efficiency of such capture by using different amounts and/or concentrations of said affinity probes, and optionally by using DNA sequences comprising known modified nucleotide profiles as evaluation metrics for said empirical testing). Optionally, said incomplete enrichment may be performed by limiting and/or reducing the duration of time wherein the affinity probe is used to bind and/or capture the target fragments of genomic DNA within the enrichment process (i.e. by using different incubation times wherein the affinity probe is able to interact with potential target fragments of genomic DNA within a sample); for example, by empirically testing the efficiency of such capture by using different durations of incubation, and optionally by using DNA sequences comprising known modified nucleotide profiles as evaluation metrics for said empirical testing). Such incomplete enrichment may have a benefit of reducing false-positive molecular signals (e.g., wherein fragments of genomic DNA are captured during an enrichment process but where said fragments do not have the desired target modified nucleotide). Additionally, said incomplete enrichment may have a benefit of reducing the cost and complexity of the enrichment process(es) themselves.

The methods may comprise performing a sequence-enrichment or sequence-capture step, in which one or more specific genomic DNA sequences are enriched from the fragments of genomic DNA. This step may be performed by any method of performing sequence enrichment, such as

using DNA oligonucleotides complementary to said sequences, or RNA oligonucleotides complementary to said sequences, or by a step employing a primer-extension target-enrichment step, or by a step employing a molecular inversion probe set or a by a step employing a padlock probe set. Optionally, if barcode sequences are appended, this enrichment step may be performed before the step of appending barcode sequences or after the step of appending barcode sequences. Optionally, if two or more sequences of fragments of genomic DNA from a microparticle are appended to each other, this enrichment step may be performed before the step of appending such sequences to each other or after the step of appending such sequences to each other.

The methods may comprise performing a sequence-depletion or sequence-removal step, in which one or more specific genomic DNA sequences (and/or specific RNA sequences) are depleted and/or removed from the fragments of genomic DNA (and/or from the fragments or molecules of RNA). This step may be performed by any method of performing sequence depletion or removal, such as using DNA oligonucleotides complementary to said sequences, or RNA oligonucleotides complementary to said sequences. Optionally, any such depletion and/or removal step may comprise depletion or removal of ribosomal RNA sequences.

The method may comprise enriching at least 1, at least 5, at least 10, at least 50, at least 100, at least 500, at least 1000, at least 5000, at least 10,000, at least 100,000, at least 1,000,000, or at least 10,000,000 different fragments of genomic DNA.

In the methods, each unique input molecule may be sequenced within the sequencing reaction on average at least 1.0 times, on average at least 1.5 times, on average at least 2.0 times, on average at least 3.0 times, on average at least 5.0 times, on average at least 10.0 times, on average at least 20.0 times, on average at least 50.0 times, or on average at least 100 times. Optionally, unique input molecules that are sequenced at least two times within the sequencing reaction (i.e. redundantly sequenced with at least two sequence reads) are used to detect and/or remove errors or inconsistencies in sequencing between said at least two sequence reads made by the sequencing reaction.

Prior to performing a sequencing reaction, and/or prior to performing an amplification reaction, a nucleotide repair reaction may be performed, in which damaged and/or excised bases or oligonucleotides are removed and/or repaired. Optionally, said repair reaction may be performed in the presence of one or more of the following: *Thermus aquaticus* DNA Ligase, *e. coli* Endonuclease IV, *Bacillus stearothermophilus* DNA Polymerase, *e. coli* formamidopyrimidine [fapy]-DNA glycosylase, *e. coli* Uracil-DNA Glycosylase, T4 Endonuclease V, and *e. coli* Endonuclease VIII.

- 115 -

In the methods, a universal adapter sequence (e.g. one or two universal adapter sequences) may be appended prior to a sequencing step, and/or prior to an amplification step such as a PCR amplification step. Optionally, one or more such universal adapter sequences may be added by a random-primed or gene-specific primer extension step, by an in vitro transposition reaction wherein one or more said universal adapter sequences are comprised within a synthetic transposome, by a double-stranded or single-stranded ligation reaction (with or without a preceding fragmentation step, such as a chemical fragmentation step, an acoustic or mechanical fragmentation step, or an enzymatic fragmentation step; and optionally with or without a blunting, and/or 3' A-tailing step).

Barcode Sequences Comprising Enzymatically-Produced Copies or Enzymatically-Produced Complements

One or more barcode sequences may be comprised within oligonucleotides (e.g. comprised within barcoded oligonucleotides) comprising enzymatically-produced copies or enzymatically-produced complements of a barcode sequence.

Optionally, one or more barcode sequences may be comprised within a barcoded oligonucleotide, wherein the barcode region of the barcoded oligonucleotide comprises an enzymatically-produced copy or enzymatically-produced complement of a barcode sequence. Optionally, one or more barcode sequences may be comprised within a barcoded oligonucleotide, wherein the barcode region of the barcoded oligonucleotide comprises an enzymatically-produced complement of a barcode sequence comprised within a barcode molecule. Optionally, one or more barcode sequences may be comprised within a barcoded oligonucleotide, wherein the barcode region of the barcoded oligonucleotide comprises an enzymatically-produced copy of a barcode sequence comprised within a barcode molecule.

Optionally, one or more barcode sequences may be comprised within a barcoded oligonucleotide, wherein the barcode region of the barcoded oligonucleotide comprises an enzymatically-produced complement of a barcode sequence comprised within a multimeric barcode molecule. Optionally, one or more barcode sequences may be comprised within a barcoded oligonucleotide, wherein the barcode region of the barcoded oligonucleotide comprises an enzymatically-produced copy of a barcode sequence comprised within a multimeric barcode molecule.

Optionally, one or more barcode sequences may be comprised within a first barcoded oligonucleotide, wherein the barcode region of the barcoded oligonucleotide comprises an enzymatically-produced complement of a barcode sequence comprised within a second barcoded oligonucleotide. Optionally, one or more barcode sequences may be comprised within a first barcoded oligonucleotide, wherein the barcode region of the barcoded oligonucleotide comprises

an enzymatically-produced copy of a barcode sequence comprised within a second barcoded oligonucleotide.

- Any enzymatic process used for copying, replicating, and/or synthesising nucleic acid sequences may be employed to produce enzymatically-produced copies or enzymatically-produced complements of a barcode sequence. Optionally, a primer-extension process may be employed. Optionally, a primer-extension process may be employed, wherein a barcode sequence comprised within a barcode molecule (and/or comprised within a multimeric barcode molecule, and/or comprised within a barcoded oligonucleotide) is copied within a primer-extension step, and wherein the resulting primer-extension product of the primer-extension step comprises all or part of a barcode sequence (e.g. comprises all or part of a barcoded oligonucleotide) which is then appended to the sequence of a nucleic acid from a circulating microparticle (e.g., appended to the sequence of a fragment of genomic DNA from a circulating microparticle).
- Optionally, a polymerase chain reaction (PCR) process may be employed. Optionally, a polymerase chain reaction (PCR) process may be employed, wherein a barcode sequence comprised within a barcode molecule (and/or comprised within a multimeric barcode molecule, and/or comprised within a barcoded oligonucleotide) is copied within a PCR extension step, and wherein the resulting extension product of the PCR extension step comprises all or part of a barcode sequence (e.g. comprises all or part of a barcoded oligonucleotide) which is then appended to the sequence of a nucleic acid from a circulating microparticle (e.g., appended to the sequence of a fragment of genomic DNA from a circulating microparticle). Optionally, a polymerase chain reaction (PCR) process may be employed, wherein a barcode sequence comprised within a barcode molecule (and/or comprised within a multimeric barcode molecule, and/or comprised within a barcoded oligonucleotide) is copied with at least two sequential PCR extension steps (e.g. copied with at least a first PCR cycle and then a second PCR cycle), and wherein at least two resulting PCR extension products each comprise all or part of a barcode sequence (e.g. comprises all or part of a barcoded oligonucleotide) which is then appended to the sequence of a nucleic acid from a circulating microparticle (e.g., appended to the sequence of a fragment of genomic DNA from a circulating microparticle).

- Optionally, a rolling-circle amplification (RCA) process may be employed. Optionally, a rolling-circle amplification (RCA) process may be employed, wherein a barcode sequence comprised within a barcode molecule (and/or comprised within a multimeric barcode molecule, and/or comprised within a barcoded oligonucleotide) is copied within a rolling-circle amplification step. For example, as illustrated in Figure 7. Further details of such methods are provided in PCT/GB2017/053820, which is incorporated herein by reference.

- 117 -

Optionally, any such process of producing enzymatically-produced copies or enzymatically-produced complements of a barcode sequence may be performed in a single reaction volume.

Optionally, any such process of producing enzymatically-produced copies or enzymatically-produced complements of a barcode sequence may be performed in two or more different

5 reaction volumes (i.e., performed in two or more different partitions). Optionally, any such process of producing enzymatically-produced copies or enzymatically-produced complements of a barcode sequence may be performed in at least 3, at least 5, at least 10, at least 50, at least 100, at least 500, at least 1000, at least 10,000, at least 100,000, at least 1,000,000, at least 10,000,000, or at least 100,000,000 different reaction volumes (and/or partitions).

10

Optionally, any such process of producing enzymatically-produced copies or enzymatically-produced complements of a barcode sequence may be performed in a reaction volume comprising sequences of nucleic acids from one or more circulating microparticles (e.g., in a reaction volume comprising one or more circulating microparticles). Optionally, a process of

15 producing enzymatically-produced copies or enzymatically-produced complements of a barcode sequence may be performed in a first reaction volume comprising sequences of nucleic acids of a first circulating microparticle from a sample (e.g., comprising fragments of genomic DNA of a first circulating microparticle from a sample, and/or comprising a first circulating microparticle from a sample) and performed in a second reaction volume comprising sequences of nucleic acids of a
20 second circulating microparticle from the sample (e.g., comprising fragments of genomic DNA of a second circulating microparticle from the sample, and/or comprising a second circulating microparticle from the sample).

Optionally, a process of producing enzymatically-produced copies or enzymatically-produced
25 complements of a barcode sequence may be performed in N different reaction volumes, wherein each such reaction volume comprises at least one barcode sequence and further comprises sequences of nucleic acids of a circulating microparticle from a sample (e.g., further comprises fragments of genomic DNA of a circulating microparticle from a sample, and/or further comprises a circulating microparticle from a sample), wherein N is at least 2, at least 3, at least 5, at least 10,
30 at least 50, at least 100, at least 500, at least 1000, at least 10,000, at least 100,000, at least 1,000,000, at least 10,000,000, or at least 100,000,000. Optionally, the barcode sequences comprised across the N different reaction volumes may together comprise at least 2, at least 3, at least 5, at least 10, at least 50, at least 100, at least 500, at least 1000, at least 10,000, at least 100,000, at least 1,000,000, at least 10,000,000, or at least 100,000,000 different barcode
35 sequences.

Optionally, a process of producing enzymatically-produced copies or enzymatically-produced complements of a barcode sequence may be performed in a first reaction volume comprising a first barcode sequence and further comprising sequences of nucleic acids of a first circulating

microparticle of a sample (e.g., further comprising fragments of genomic DNA of a first circulating microparticle from a sample, and/or further comprising a first circulating microparticle from a sample) and performed in a second reaction volume comprising a second barcode sequence and further comprising sequences of nucleic acids of a second circulating microparticle of the sample (e.g., further comprising fragments of genomic DNA of a second circulating microparticle from the sample, and/or further comprising a second circulating microparticle from the sample), wherein the first barcode sequence is different to the second barcode sequence.

Optionally, a process of producing enzymatically-produced copies or enzymatically-produced complements of a barcode sequence may be performed in at first reaction volume comprising sequences of nucleic acids of a first circulating microparticle of a sample (e.g., comprising fragments of genomic DNA of a first circulating microparticle of a sample) wherein at least first and second enzymatically-produced copies or enzymatically-produced complements of a barcode sequence from the first reaction volume are appended to sequences of nucleic acids of the first circulating microparticle of the sample, and performed in at second reaction volume comprising sequences of nucleic acids of a second circulating microparticle of the sample (e.g., comprising fragments of genomic DNA of a second circulating microparticle of the sample) wherein at least first and second enzymatically-produced copies or enzymatically-produced complements of a barcode sequence from the second reaction volume are appended to sequences of nucleic acids of the second circulating microparticle of the sample.

Optionally, any process of producing enzymatically-produced copies or enzymatically-produced complements of a barcode sequence may be performed for (and/or performed on or with) a library comprising two or more barcode sequences. Optionally, any process of producing enzymatically-produced copies or enzymatically-produced complements of a barcode sequence may be performed for (and/or performed on or with) a library comprising two or more barcode molecules. Optionally, any process of producing enzymatically-produced copies or enzymatically-produced complements of a barcode sequence may be performed for (and/or performed on or with) a library comprising two or more multimeric barcode molecules. Optionally, any process of producing enzymatically-produced copies or enzymatically-produced complements of a barcode sequence may be performed for (and/or performed on or with) a library comprising two or more multimeric barcoding reagents. Optionally, any process of producing enzymatically-produced copies or enzymatically-produced complements of a barcode sequence may be performed for (and/or performed on or with) a library comprising two or more barcoded oligonucleotides.

Optionally, any process of producing enzymatically-produced copies or enzymatically-produced complements of a barcode sequence may further comprise appending any one or more enzymatically-produced copies or enzymatically-produced complements of a barcode sequence to each of one or more sequences of nucleic acids of a circulating microparticle (e.g. to fragments

of genomic DNA of a circulating microparticle) in an appending step. Optionally, any one or more such appending step may comprise a step of hybridisation (e.g. a step of hybridising a barcoded oligonucleotide to a nucleic acid sequence), a step of hybridisation and extension hybridisation (e.g. a step of hybridising a barcoded oligonucleotide to a nucleic acid sequence and then
5 extending the hybridised barcoded oligonucleotide with a polymerase), and/or a step of ligation (e.g. a step of ligating a barcoded oligonucleotide to a nucleic acid sequence). Following any one or more such appending steps, the nucleic acid sequences comprising barcode sequences and the sequences of nucleic acids from circulating microparticle(s) to which they have been appended, may then be subject to a sequencing step.

10 Optionally, any process of producing enzymatically-produced copies or enzymatically-produced complements of a barcode sequence may further comprise appending any one or more enzymatically-produced copies or enzymatically-produced complements of a barcode sequence to each of one or more sequences of nucleic acids of a circulating microparticle (e.g. to fragments
15 of genomic DNA of a circulating microparticle), wherein said sequences of nucleic acids of a circulating microparticle further comprise a coupling sequence. Any coupling sequence and/or method(s) of appending coupling sequences, and/or methods of appending barcode sequences to coupling sequences (and/or to oligonucleotides comprising coupling sequences) described herein may be employed.

20 Optionally, any process of producing enzymatically-produced copies or enzymatically-produced complements of a barcode sequence and further comprising appending any one or more enzymatically-produced copies or enzymatically-produced complements of a barcode sequence to sequences of nucleic acids of a circulating microparticle, may further comprise a step of
25 chemically crosslinking a circulating microparticle (and/or chemically crosslinking a sample comprising two or more circulating microparticles). Optionally, said step of chemical crosslinking may be performed prior to and/or after a step of partitioning circulating microparticles and/or barcode molecules into two or more different partitions. Optionally, said step of chemical crosslinking may be followed by a step of reversing said crosslinks, for example with a high-
30 temperature thermal incubation step. Optionally, any process of producing enzymatically-produced copies or enzymatically-produced complements of a barcode sequence and further comprising appending any one or more enzymatically-produced copies or enzymatically-produced complements of a barcode sequence to sequences of nucleic acids of a circulating microparticle, may further comprise a step of permeabilising said circulating microparticle(s), for example with a
35 high-temperature incubation step and/or with a chemical surfactant.

Optionally, any process of producing enzymatically-produced copies or enzymatically-produced complements of a barcode sequence may be performed with any number and/or type and/or volume of partition described herein. Optionally, any process of producing enzymatically-

- 120 -

produced copies or enzymatically-produced complements of a barcode sequence in one or more partitions may comprise one or more partitions comprising any number of circulating microparticles as described herein. Optionally, any process of producing enzymatically-produced copies or enzymatically-produced complements of a barcode sequence in one or more partitions may comprise one or more partitions comprising any number (or average number) of circulating microparticles as described herein. Optionally, any process of producing enzymatically-produced copies or enzymatically-produced complements of a barcode sequence in one or more partitions may comprise one or more partitions comprising any mass (or average mass) of nucleic acids (e.g. any mass of fragments of genomic DNA) from circulating microparticles as described herein.

Processes of producing enzymatically-produced copies and/or enzymatically-produced complements of a barcode sequence may have a variety of desirable features and characteristics for the purposes of analysing linked sequences from circulating microparticles. In the first case, producing enzymatically-produced copies and/or enzymatically-produced complements of a barcode sequence enables the production of a large absolute mass of barcode sequences (e.g. a large absolute mass of barcode molecules or barcoded oligonucleotides), using only a small amount of starting barcode sequence material (e.g., PCR and RCA processing can produce vast exponential amplification of input material for subsequent use and manipulation).

Furthermore, producing enzymatically-produced copies and/or enzymatically-produced complements of barcode sequences wherein such barcode sequences are comprised within libraries (e.g. comprised within libraries of barcode molecules, libraries of multimeric barcode molecules, libraries of multimeric barcoding reagents, and/or libraries of barcoded oligonucleotides) enables the production of a large absolute mass of barcode sequences of defined sequence character (e.g. wherein the large absolute mass of barcode sequences comprise sequences from the previously-established and/or previously-characterised library or libraries).

Furthermore, many enzymatic copying and amplification processes (such as rolling circle amplification by the phi29 polymerase, and primer-extension and/or PCR amplification by thermostable polymerases such as Phusion polymerase) exhibit high molecular accuracy during said copying (in terms of the rate of error production within newly copied sequence), and thus exhibit favourable accuracy profiles of the resulting barcode sequences (e.g. the resulting barcode molecules, multimeric barcode molecules, and/or barcoded oligonucleotides) in comparison with non-enzymatic approaches (e.g. in comparison with standard chemical oligonucleotide synthesis procedures, such a phosphoramidite oligonucleotide synthesis).

Furthermore, enzymatic copying and amplification processes (e.g. primer-extension and PCR processes) are highly amenable to subsequent steps of modification, processing, and

functionalisation of said sequences, which also may have the further benefit of themselves being achievable on large absolute masses of substrate in relatively straightforward fashion. For example, primer-extension products are readily configured and/or configurable for subsequent ligation processes (e.g., as in a primer-extension and ligation process, as for example may be performed to produce barcoded oligonucleotides and/or multimeric barcoding reagents). And for further example, the direct products of enzymatic-copying processes themselves (e.g. wherein a complement/copy of a barcode sequence is annealed to the barcode sequence itself) may have desirable functional and/or structural properties. For example, a barcoded oligonucleotide produced through an enzymatic primer-extension process is retained structurally tethered (through the annealed nucleotide sequence) to the barcode molecule (e.g. multimeric barcode molecules) along which it was produced, in a singular macromolecular complex that may then be further processed and/or functionalised as a singular, intact reagent in solution.

14. GENERAL PROPERTIES OF MULTIMERIC BARCODING REAGENTS

Use of multimeric barcoding reagents exhibits a variety of useful features and functionalities to link sequences from circulating microparticles. In the first case, such reagents (and/or libraries thereof) can comprise very well-defined, well-characterised sets of barcodes, which can inform and enhance subsequent bioinformatic analysis (for example, as relates to use of multimeric barcode molecules and/or multimeric barcoding reagents of known and/or empirically determined sequence). Additionally, such reagents enable extremely easy partitioning and/or other molecular or biophysical processes of multiple barcode sequences at once (i.e., since multiple barcode sequences are comprised within each such reagent, they automatically 'move together' within solution and during liquid handling and/or processing steps). Furthermore, the proximity between multiple barcode sequences of such reagents itself can enable novel functional assay forms, such as crosslinking circulating microparticles and then appending sequences from such multimeric reagents to the fragments of genomic DNA contained therein (including e.g. within solution-phase reactions thereof, i.e. with two or more microparticles within a single partition).

The invention provides multimeric barcoding reagents for labelling one or more target nucleic acids. A multimeric barcoding reagent comprises two or more barcode regions are linked together (directly or indirectly).

Each barcode region comprises a nucleic acid sequence. The nucleic acid sequence may be single-stranded DNA, double-stranded DNA, or single stranded DNA with one or more double-stranded regions.

Each barcode region may comprise a sequence that identifies the multimeric barcoding reagent. For example, this sequence may be a constant region shared by all barcode regions of a single multimeric barcoding reagent. Each barcode region may contain a unique sequence which is not

- present in other regions, and may thus serve to uniquely identify each barcode region. Each barcode region may comprise at least 5, at least 10, at least 15, at least 20, at least 25, at least 50 or at least 100 nucleotides. Preferably, each barcode region comprises at least 5 nucleotides. Preferably each barcode region comprises deoxyribonucleotides, optionally all of the nucleotides
- 5 in a barcode region are deoxyribonucleotides. One or more of the deoxyribonucleotides may be a modified deoxyribonucleotide (e.g. a deoxyribonucleotide modified with a biotin moiety or a deoxyuracil nucleotide). The barcode regions may comprise one or more degenerate nucleotides or sequences. The barcode regions may not comprise any degenerate nucleotides or sequences.
- 10 The multimeric barcoding reagent may comprise at least 5, at least 10, at least 20, at least 25, at least 50, at least 75, at least 100, at least 200, at least 500, at least 1000, at least 5000, or at least 10,000 barcode regions. Preferably, the multimeric barcoding reagent comprises at least 5 barcode regions.
- 15 The multimeric barcoding reagent may comprise at least 2, at least 3, at least 4, at least 5, at least 10, at least 20, at least 25, at least 50, at least 75, at least 100, at least 200, at least 500, at least 1000, at least 5000, at least 10^4 , at least 10^5 , or at least 10^6 unique or different barcode regions. Preferably, the multimeric barcoding reagent comprises at least 5 unique or different barcode regions.
- 20 A multimeric barcoding reagent may comprise: first and second barcode molecules linked together (i.e. a multimeric barcode molecule), wherein each of the barcode molecules comprises a nucleic acid sequence comprising a barcode region.
- 25 The barcode molecules of a multimeric barcode molecule may be linked on a nucleic acid molecule. The barcode molecules of a multimeric barcode molecule may be comprised within a (single) nucleic acid molecule. A multimeric barcode molecule may comprise a single, contiguous nucleic acid sequence comprising two or more barcode molecules. A multimeric barcode molecule may be a single-stranded nucleic acid molecule (e.g. single-stranded DNA), a double-
- 30 stranded-stranded nucleic acid molecule or a single stranded molecule comprising one or more double-stranded regions. A multimeric barcode molecule may comprise one or more phosphorylated 5' ends capable of ligating to 3' ends of other nucleic acid molecules. Further details of the multimeric barcode molecules and multimeric barcoding reagents are provided in PCT/GB2017/053820, which is incorporated herein by reference.
- 35 The barcode molecules may be linked by a support e.g. a macromolecule, solid support or semi-solid support. The sequences of the barcode molecules linked to each support may be known. The barcode molecules may be linked to the support directly or indirectly (e.g. via a linker molecule). The barcode molecules may be linked by being bound to the support and/or by being

- 123 -

bound or annealed to linker molecules that are bound to the support. The barcode molecules may be bound to the support (or to the linker molecules) by covalent linkage, non-covalent linkage (e.g. a protein-protein interaction or a streptavidin-biotin bond) or nucleic acid hybridization. The linker molecule may be a biopolymer (e.g. a nucleic acid molecule) or a synthetic polymer. The linker molecule may comprise one or more units of ethylene glycol and/or poly(ethylene) glycol (e.g. hexa-ethylene glycol or penta-ethylene glycol). The linker molecule may comprise one or more ethyl groups, such as a C3 (three-carbon) spacer, C6 spacer, C12 spacer, or C18 spacer. The linker molecule may comprise at least 2, at least 3, at least 4, at least 5, at least 10, or at least 20 sequential repeating units of any individual linker (such as a sequential linear series of at least 2, at least 5, or at least 10 C12 spacers or C18 spacers). The linker molecule may comprise a branched linker molecule, wherein 2 or more barcode molecules are linked to a support by a single linker molecule.

The barcode molecules may be linked by a macromolecule by being bound to the macromolecule and/or by being annealed to the macromolecule.

The barcode molecules may be linked to the macromolecule directly or indirectly (e.g. via a linker molecule). The barcode molecules may be linked by being bound to the macromolecule and/or by being bound or annealed to linker molecules that are bound to the macromolecule. The barcode molecules may be bound to the macromolecule (or to the linker molecules) by covalent linkage, non-covalent linkage (e.g. a protein-protein interaction or a streptavidin-biotin bond) or nucleic acid hybridization. The linker molecule may be a biopolymer (e.g. a nucleic acid molecule) or a synthetic polymer. The linker molecule may comprise one or more units of ethylene glycol and/or poly(ethylene) glycol (e.g. hexa-ethylene glycol or penta-ethylene glycol). The linker molecule may comprise one or more ethyl groups, such as a C3 (three-carbon) spacer, C6 spacer, C12 spacer, or C18 spacer.

The macromolecule may be a synthetic polymer (e.g. a dendrimer) or a biopolymer such as a nucleic acid (e.g. a single-stranded nucleic acid such as single-stranded DNA), a peptide, a polypeptide or a protein (e.g. a multimeric protein).

The dendrimer may comprise at least 2, at least 3, at least 5, or at least 10 generations.

The macromolecule may be a nucleic acid comprising two or more nucleotides each capable of binding to a barcode molecule. Additionally or alternatively, the nucleic acid may comprise two or more regions each capable of hybridizing to a barcode molecule.

The nucleic acid may comprise a first modified nucleotide and a second modified nucleotide, wherein each modified nucleotide comprises a binding moiety (e.g. a biotin moiety, or an alkyne

moiety which may be used for a click-chemical reaction) capable of binding to a barcode molecule. Optionally, the first and second modified nucleotides may be separated by an intervening nucleic acid sequence of at least one, at least two, at least 5 or at least 10 nucleotides.

5

The nucleic acid may comprise a first hybridisation region and a second hybridisation region, wherein each hybridisation region comprises a sequence complementary to and capable of hybridizing to a sequence of at least one nucleotide within a barcode molecule. The complementary sequence may be at least 5, at least 10, at least 15, at least 20, at least 25 or at least 50 contiguous nucleotides. Preferably, the complementary sequence is at least 10 contiguous nucleotides. Optionally, the first and second hybridisation regions may be separated by an intervening nucleic acid sequence of at least one, at least two, at least 5 or at least 10 nucleotides.

10

15

The macromolecule may be a protein such as a multimeric protein e.g. a homomeric protein or a heteromeric protein. For example, the protein may comprise streptavidin e.g. tetrameric streptavidin.

20

The support may be a solid support or a semi-solid support. The support may comprise a planar surface. The support may be a slide e.g. a glass slide. The slide may be a flow cell for sequencing. If the support is a slide, the first and second barcode molecules may be immobilized in a discrete region on the slide. Optionally, the barcode molecules of each multimeric barcoding reagent in a library are immobilized in a different discrete region on the slide to the barcode molecules of the other multimeric barcoding reagents in the library. The support may be a plate comprising wells, optionally wherein the first and second barcode molecules are immobilized in the same well. Optionally, the barcode molecules of each multimeric barcoding reagent in library are immobilized in a different well of the plate to the barcode molecules of the other multimeric barcoding reagents in the library.

25

30

Preferably, the support is a bead (e.g. a gel bead). The bead may be an agarose bead, a silica bead, a styrofoam bead, a gel bead (such as those available from 10x Genomics®), an antibody conjugated bead, an oligo-dT conjugated bead, a streptavidin bead or a magnetic bead (e.g. a superparamagnetic bead). The bead may be of any size and/or molecular structure. For example, the bead may be 10 nanometres to 100 microns in diameter, 100 nanometres to 10 microns in diameter, or 1 micron to 5 microns in diameter. Optionally, the bead is approximately 10 nanometres in diameter, approximately 100 nanometres in diameter, approximately 1 micron in diameter, approximately 10 microns in diameter or approximately 100 microns in diameter. The bead may be solid, or alternatively the bead may be hollow or partially hollow or porous. Beads of certain sizes may be most preferable for certain barcoding methods. For example, beads less

35

than 5.0 microns, or less than 1.0 micron, may be most useful for barcoding nucleic acid targets within individual cells. Preferably, the barcode molecules of each multimeric barcoding reagent in a library are linked together on a different bead to the barcode molecules of the other multimeric barcoding reagents in the library.

5

The support may be functionalised to enable attachment of two or more barcode molecules. This functionalisation may be enabled through the addition of chemical moieties (e.g. carboxylated groups, alkynes, azides, acrylate groups, amino groups, sulphate groups, or succinimide groups), and/or protein-based moieties (e.g. streptavidin, avidin, or protein G) to the support. The barcode

10

molecules may be attached to the moieties directly or indirectly (e.g. via a linker molecule).

Functionalised supports (e.g. beads) may be brought into contact with a solution of barcode molecules under conditions which promote the attachment of two or more barcode molecules to each bead in the solution (generating multimeric barcoding reagents).

15

In a library of multimeric barcoding reagents, the barcode molecules of each multimeric barcoding reagent in a library may be linked together on a different support to the barcode molecules of the other multimeric barcoding reagents in the library.

20

The multimeric barcoding reagent may comprise: at least 2, at least 3, at least 4, at least 5, at least 10, at least 20, at least 25, at least 50, at least 75, at least 100, at least 200, at least 500, at least 1000, at least 5000, at least 10^4 , at least 10^5 , or at least 10^6 barcode molecules linked together, wherein each barcode molecule is as defined herein; and a barcoded oligonucleotide annealed to each barcode molecule, wherein each barcoded oligonucleotide is as defined herein.

25

Preferably, the multimeric barcoding reagent comprises at least 5 barcode molecules linked together, wherein each barcode molecule is as defined herein; and a barcoded oligonucleotide annealed to each barcode molecule, wherein each barcoded oligonucleotide is as defined herein.

30

The multimeric barcoding reagent may comprise: at least 2, at least 3, at least 4, at least 5, at least 10, at least 20, at least 25, at least 50, at least 75, at least 100, at least 200, at least 500, at least 1000, at least 5000, at least 10^4 , at least 10^5 , or at least 10^6 unique or different barcode molecules linked together, wherein each barcode molecule is as defined herein; and a barcoded oligonucleotide annealed to each barcode molecule, wherein each barcoded oligonucleotide is as defined herein. Preferably, the multimeric barcoding reagent comprises at least 5 unique or

35

different barcode molecules linked together, wherein each barcode molecule is as defined herein; and a barcoded oligonucleotide annealed to each barcode molecule, wherein each barcoded oligonucleotide is as defined herein.

A multimeric barcoding reagent may comprise two or more barcoded oligonucleotides as defined herein, wherein the barcoded oligonucleotides each comprise a barcode region. A multimeric barcoding reagent may comprise: at least 2, at least 3, at least 4, at least 5, at least 10, at least 20, at least 25, at least 50, at least 75, at least 100, at least 200, at least 500, at least 1000, at least 5000, at least 10,000, at least 100,000, or at least 1,000,000 unique or different barcoded oligonucleotides. Preferably, the multimeric barcoding reagent comprises at least 5 unique or different barcoded oligonucleotides.

The barcoded oligonucleotides of a multimeric barcoding reagent are linked together (directly or indirectly). The barcoded oligonucleotides of a multimeric barcoding reagent are linked together by a support e.g. a macromolecule, solid support or semi-solid support, as described herein. The multimeric barcoding reagent may comprise one or more polymers to which the barcoded oligonucleotides are annealed or attached. For example, the barcoded oligonucleotides of a multimeric barcoding reagent may be annealed to a multimeric hybridization molecule e.g. a multimeric barcode molecule. Alternatively, the barcoded oligonucleotides of a multimeric barcoding reagent may be linked together by a macromolecule (such as a synthetic polymer e.g. a dendrimer, or a biopolymer e.g. a protein) or a support (such as a solid support or a semi-solid support e.g. a gel bead). Additionally or alternatively, the barcoded oligonucleotides of a (single) multimeric barcoding reagent may be linked together by being comprised within a (single) lipid carrier (e.g. a liposome or a micelle).

The barcoded oligonucleotides of a multimeric barcoding reagent may comprise: a first barcoded oligonucleotide comprising, optionally in the 5' to 3' direction, a barcode region, and a target region capable of annealing or ligating to a first fragment of the target nucleic acid; and a second barcoded oligonucleotide comprising, optionally in the 5' to 3' direction, a barcode region, and a target region capable of annealing or ligating to a second fragment of the target nucleic acid.

The barcoded oligonucleotides of a multimeric barcoding reagent may comprise: a first barcoded oligonucleotide comprising a barcode region, and a target region capable of ligating to a first fragment of the target nucleic acid; and a second barcoded oligonucleotide comprising a barcode region, and a target region capable of ligating to a second fragment of the target nucleic acid.

The barcoded oligonucleotides of a multimeric barcoding reagent may comprise: a first barcoded oligonucleotide comprising, in the 5' to 3' direction, a barcode region, and a target region capable of annealing to a first fragment of the target nucleic acid; and a second barcoded oligonucleotide comprising, in the 5' to 3' direction, a barcode region, and a target region capable of annealing to a second fragment of the target nucleic acid.

15. GENERAL PROPERTIES OF BARCODED OLIGONUCLEOTIDES

A barcoded oligonucleotide comprises a barcode region. The barcoded oligonucleotides may comprise, optionally in the 5' to 3' direction, a barcode region and a target region. The target region is capable of annealing or ligating to a fragment of the target nucleic acid. Alternatively, a barcoded oligonucleotide may consist essentially of or consist of a barcode region.

The 5' end of a barcoded oligonucleotide may be phosphorylated. This may enable the 5' end of the barcoded oligonucleotide to be ligated to the 3' end of a target nucleic acid. Alternatively, the 5' end of a barcoded oligonucleotide may not be phosphorylated.

A barcoded oligonucleotide may be a single-stranded nucleic acid molecule (e.g. single-stranded DNA). A barcoded oligonucleotide may comprise one or more double-stranded regions. A barcoded oligonucleotide may be a double-stranded nucleic acid molecule (e.g. double-stranded DNA).

The barcoded oligonucleotides may comprise or consist of deoxyribonucleotides. One or more of the deoxyribonucleotides may be a modified deoxyribonucleotide (e.g. a deoxyribonucleotide modified with a biotin moiety or a deoxyuracil nucleotide). The barcoded oligonucleotides may comprise one or more degenerate nucleotides or sequences. The barcoded oligonucleotides may not comprise any degenerate nucleotides or sequences.

The barcode regions of each barcoded oligonucleotide may comprise different sequences. Each barcode region may comprise a sequence that identifies the multimeric barcoding reagent. For example, this sequence may be a constant region shared by all barcode regions of a single multimeric barcoding reagent. The barcode region of each barcoded oligonucleotide may contain a unique sequence which is not present in other barcoded oligonucleotides, and may thus serve to uniquely identify each barcoded oligonucleotide. Each barcode region may comprise at least 5, at least 10, at least 15, at least 20, at least 25, at least 50 or at least 100 nucleotides. Preferably, each barcode region comprises at least 5 nucleotides. Preferably each barcode region comprises deoxyribonucleotides, optionally all of the nucleotides in a barcode region are deoxyribonucleotides. One or more of the deoxyribonucleotides may be a modified deoxyribonucleotide (e.g. a deoxyribonucleotide modified with a biotin moiety or a deoxyuracil nucleotide). The barcode regions may comprise one or more degenerate nucleotides or sequences. The barcode regions may not comprise any degenerate nucleotides or sequences.

The target regions of each barcoded oligonucleotide may comprise different sequences. Each target region may comprise a sequence capable of annealing to only a single fragment of a target nucleic acid within a sample of nucleic acids (i.e. a target specific sequence). Each target region may comprise one or more random, or one or more degenerate, sequences to enable the target region to anneal to more than one fragment of a target nucleic acid. Each target region may

comprise at least 5, at least 10, at least 15, at least 20, at least 25, at least 50 or at least 100 nucleotides. Preferably, each target region comprises at least 5 nucleotides. Each target region may comprise 5 to 100 nucleotides, 5 to 10 nucleotides, 10 to 20 nucleotides, 20 to 30 nucleotides, 30 to 50 nucleotides, 50 to 100 nucleotides, 10 to 90 nucleotides, 20 to 80 nucleotides, 30 to 70 nucleotides or 50 to 60 nucleotides. Preferably, each target region comprises 30 to 70 nucleotides. Preferably each target region comprises deoxyribonucleotides, optionally all of the nucleotides in a target region are deoxyribonucleotides. One or more of the deoxyribonucleotides may be a modified deoxyribonucleotide (e.g. a deoxyribonucleotide modified with a biotin moiety or a deoxyuracil nucleotide). Each target region may comprise one or more universal bases (e.g. inosine), one or modified nucleotides and/or one or more nucleotide analogues.

The target regions may be used to anneal the barcoded oligonucleotides to fragments of target nucleic acids, and then may be used as primers for a primer-extension reaction or an amplification reaction e.g. a polymerase chain reaction. Alternatively, the target regions may be used to ligate the barcoded oligonucleotides to fragments of target nucleic acids. The target region may be at the 5' end of a barcoded oligonucleotide. Such a target region may be phosphorylated. This may enable the 5' end of the target region to be ligated to the 3' end of a fragment of a target nucleic acid.

The barcoded oligonucleotides may further comprise one or more adapter region(s). An adapter region may be between the barcode region and the target region. A barcoded oligonucleotide may, for example, comprise an adapter region 5' of a barcode region (a 5' adapter region) and/or an adapter region 3' of the barcode region (a 3' adapter region). Optionally, the barcoded oligonucleotides comprise, in the 5' to 3' direction, a barcode region, an adapter region and a target region.

The adapter region(s) of the barcoded oligonucleotides may comprise a sequence complementary to an adapter region of a multimeric barcode molecule or a sequence complementary to a hybridization region of a multimeric hybridization molecule. The adapter region(s) of the barcoded oligonucleotides may enable the barcoded oligonucleotides to be linked to a macromolecule or support (e.g. a bead). The adapter region(s) may be used for manipulating, purifying, retrieving, amplifying, or detecting barcoded oligonucleotides and/or target nucleic acids to which they may anneal or ligate.

The adapter region of each barcoded oligonucleotide may comprise a constant region. Optionally, all adapter regions of barcoded oligonucleotides of each multimeric barcoding reagent are substantially identical. The adapter region may comprise at least 1, at least 2, at least 3, at least 4, at least 5, at least 6, at least 8, at least 10, at least 15, at least 20, at least 25, at least 50,

- 129 -

at least 100, or at least 250 nucleotides. Preferably, the adapter region comprises at least 4 nucleotides. Preferably each adapter region comprises deoxyribonucleotides, optionally all of the nucleotides in an adapter region are deoxyribonucleotides. One or more of the deoxyribonucleotides may be a modified deoxyribonucleotide (e.g. a deoxyribonucleotide modified with a biotin moiety or a deoxyuracil nucleotide). Each adapter region may comprise one or more universal bases (e.g. inosine), one or modified nucleotides and/or one or more nucleotide analogues.

Optionally, a barcoded oligonucleotide may comprise one or more binding moieties, and/or one or more linker moieties (such as any barcoded oligonucleotide comprised within a multimeric barcoding reagent). Optionally, any barcoded oligonucleotide may be linked and/or appended to any one or more coupling molecules.

The barcoded oligonucleotides may be synthesized by a chemical oligonucleotide synthesis process. The barcoded oligonucleotides synthesis process may include one or more step of an enzymatic production process, an enzymatic amplification process, or an enzymatic modification procedure, such as an in vitro transcription process, a reverse transcription process, a primer-extension process, or a polymerase chain reaction process.

These general properties of barcoded oligonucleotides are applicable to any of the multimeric barcoding reagents described herein.

16. GENERAL PROPERTIES OF LIBRARIES OF MULTIMERIC BARCODING REAGENTS

The invention provides a library of multimeric barcoding reagents comprising first and second multimeric barcoding reagents as defined herein, wherein the barcode regions of the first multimeric barcoding reagent are different to the barcode regions of the second multimeric barcoding reagent.

The library of multimeric barcoding reagents may comprise at least 5, at least 10, at least 20, at least 25, at least 50, at least 75, at least 100, at least 250, at least 500, at least 10^3 , at least 10^4 , at least 10^5 , at least 10^6 , at least 10^7 , at least 10^8 or at least 10^9 multimeric barcoding reagents as defined herein. Preferably, the library comprises at least 10 multimeric barcoding reagents as defined herein. Preferably, the first and second barcode regions of each multimeric barcoding reagent are different to the barcode regions of at least 9 other multimeric barcoding reagents in the library.

The first and second barcode regions of each multimeric barcoding reagent may be different to the barcode regions of at least 4, at least 9, at least 19, at least 24, at least 49, at least 74, at least 99, at least 249, at least 499, at least 999 (i.e. 10^3-1), at least 10^4-1 , at least 10^5-1 , at least

- 130 -

10⁶-1, at least 10⁷-1, at least 10⁸-1 or at least 10⁹-1 other multimeric barcoding reagents in the library. The first and second barcode regions of each multimeric barcoding reagent may be different to the barcode regions of all of the other multimeric barcoding reagents in the library. Preferably, the first and second barcode regions of each multimeric barcoding reagent are different to the barcode regions of at least 9 other multimeric barcoding reagents in the library.

The barcode regions of each multimeric barcoding reagent may be different to the barcode regions of at least 4, at least 9, at least 19, at least 24, at least 49, at least 74, at least 99, at least 249, at least 499, at least 999 (i.e. 10³-1), at least 10⁴-1, at least 10⁵-1, at least 10⁶-1, at least 10⁷-1, at least 10⁸-1 or at least 10⁹-1 other multimeric barcoding reagents in the library. The barcode regions of each multimeric barcoding reagent may be different to the barcode regions of all of the other multimeric barcoding reagents in the library. Preferably, the barcode regions of each multimeric barcoding reagent are different to the barcode regions of at least 9 other multimeric barcoding reagents in the library.

The invention provides a library of multimeric barcoding reagents comprising first and second multimeric barcoding reagents as defined herein, wherein the barcode regions of the barcoded oligonucleotides of the first multimeric barcoding reagent are different to the barcode regions of the barcoded oligonucleotides of the second multimeric barcoding reagent.

Different multimeric barcoding reagents within a library of multimeric barcoding reagents may comprise different numbers of barcoded oligonucleotides.

The library of multimeric barcoding reagents may comprise at least 5, at least 10, at least 20, at least 25, at least 50, at least 75, at least 100, at least 250, at least 500, at least 10³, at least 10⁴, at least 10⁵, at least 10⁶, at least 10⁷, at least 10⁸ or at least 10⁹ multimeric barcoding reagents as defined herein. Preferably, the library comprises at least 10 multimeric barcoding reagents as defined herein. Preferably, the barcode regions of the first and second barcoded oligonucleotides of each multimeric barcoding reagent are different to the barcode regions of the barcoded oligonucleotides of at least 9 other multimeric barcoding reagents in the library.

The barcode regions of the first and second barcoded oligonucleotides of each multimeric barcoding reagent may be different to the barcode regions of the barcoded oligonucleotides of at least 4, at least 9, at least 19, at least 24, at least 49, at least 74, at least 99, at least 249, at least 499, at least 999 (i.e. 10³-1), at least 10⁴-1, at least 10⁵-1, at least 10⁶-1, at least 10⁷-1, at least 10⁸-1 or at least 10⁹-1 other multimeric barcoding reagents in the library. The barcode regions of the first and second barcoded oligonucleotides of each multimeric barcoding reagent may be different to the barcode regions of the barcoded oligonucleotides of all of the other multimeric barcoding reagents in the library. Preferably, the barcode regions of the first and second

- 131 -

barcoded oligonucleotides of each multimeric barcoding reagent are different to the barcode regions of the barcoded oligonucleotides of at least 9 other multimeric barcoding reagents in the library.

- 5 The barcode regions of the barcoded oligonucleotides of each multimeric barcoding reagent may be different to the barcode regions of the barcoded oligonucleotides of at least 4, at least 9, at least 19, at least 24, at least 49, at least 74, at least 99, at least 249, at least 499, at least 999 (i.e. 10^3-1), at least 10^4-1 , at least 10^5-1 , at least 10^6-1 , at least 10^7-1 , at least 10^8-1 or at least 10^9-1 other multimeric barcoding reagents in the library. The barcode regions of the barcoded
10 oligonucleotides of each multimeric barcoding reagent may be different to the barcode regions of the barcoded oligonucleotides of all of the other multimeric barcoding reagents in the library. Preferably, the barcode regions of the barcoded oligonucleotides of each multimeric barcoding reagent are different to the barcode regions of the barcoded oligonucleotides of at least 9 other multimeric barcoding reagents in the library.

- 15 These general properties of libraries of multimeric barcoding reagents are applicable to any of the multimeric barcoding reagents described herein.

17. MULTIMERIC BARCODING REAGENTS COMPRISING BARCODED 20 OLIGONUCLEOTIDES ANNEALED TO A MULTIMERIC BARCODE MOLECULE

- The invention provides a multimeric barcoding reagent for labelling a target nucleic acid, wherein the reagent comprises: first and second barcode molecules linked together (i.e. a multimeric barcode molecule), wherein each of the barcode molecules comprises a nucleic acid sequence comprising a barcode region; and first and second barcoded oligonucleotides, wherein the first
25 barcoded oligonucleotide comprises, optionally in the 5' to 3' direction, a barcode region annealed to the barcode region of the first barcode molecule and a target region capable of annealing or ligating to a first fragment of the target nucleic acid, and wherein the second barcoded oligonucleotide comprises, optionally in the 5' to 3' direction, a barcode region annealed to the barcode region of the second barcode molecule and a target region capable of annealing or
30 ligating to a second fragment of the target nucleic acid.

- The invention provides a multimeric barcoding reagent for labelling a target nucleic acid, wherein the reagent comprises: first and second barcode molecules linked together (i.e. a multimeric barcode molecule), wherein each of the barcode molecules comprises a nucleic acid sequence
35 comprising a barcode region; and first and second barcoded oligonucleotides, wherein the first barcoded oligonucleotide comprises a barcode region annealed to the barcode region of the first barcode molecule and a target region capable of ligating to a first fragment of the target nucleic acid, and wherein the second barcoded oligonucleotide comprises a barcode region annealed to

the barcode region of the second barcode molecule and a target region capable of ligating to a second fragment of the target nucleic acid.

5 The invention provides a multimeric barcoding reagent for labelling a target nucleic acid, wherein the reagent comprises: first and second barcode molecules linked together (i.e. a multimeric barcode molecule), wherein each of the barcode molecules comprises a nucleic acid sequence comprising a barcode region; and first and second barcoded oligonucleotides, wherein the first barcoded oligonucleotide comprises in the 5' to 3' direction a barcode region annealed to the barcode region of the first barcode molecule and a target region capable of annealing to a first
10 fragment of the target nucleic acid, and wherein the second barcoded oligonucleotide comprises in the 5' to 3' direction a barcode region annealed to the barcode region of the second barcode molecule and a target region capable of annealing to a second fragment of the target nucleic acid.

The invention provides a multimeric barcoding reagent for labelling a target nucleic acid, wherein
15 the reagent comprises: first and second barcode molecules linked together (i.e. a multimeric barcode molecule), wherein each of the barcode molecules comprises a nucleic acid sequence comprising a barcode region; and first and second barcoded oligonucleotides, wherein the first barcoded oligonucleotide comprises a barcode region annealed to the barcode region of the first barcode molecule and capable of ligating to a first fragment of the target nucleic acid, and
20 wherein the second barcoded oligonucleotide comprises a barcode region annealed to the barcode region of the second barcode molecule and capable of ligating to a second fragment of the target nucleic acid.

Each barcoded oligonucleotide may consist essentially of or consist of a barcode region.

25 Preferably, the barcode molecules comprise or consist of deoxyribonucleotides. One or more of the deoxyribonucleotides may be a modified deoxyribonucleotide (e.g. a deoxyribonucleotide modified with a biotin moiety or a deoxyuracil nucleotide). The barcode molecules may comprise one or more degenerate nucleotides or sequences. The barcode molecules may not comprise
30 any degenerate nucleotides or sequences.

The barcode regions may uniquely identify each of the barcode molecules. Each barcode region may comprise a sequence that identifies the multimeric barcoding reagent. For example, this sequence may be a constant region shared by all barcode regions of a single multimeric
35 barcoding reagent. Each barcode region may comprise at least 5, at least 10, at least 15, at least 20, at least 25, at least 50 or at least 100 nucleotides. Preferably, each barcode region comprises at least 5 nucleotides. Preferably each barcode region comprises deoxyribonucleotides, optionally all of the nucleotides in a barcode region are deoxyribonucleotides. One or more of the deoxyribonucleotides may be a modified deoxyribonucleotide (e.g. a deoxyribonucleotide

– 133 –

modified with a biotin moiety or a deoxyuracil nucleotide). The barcode regions may comprise one or more degenerate nucleotides or sequences. The barcode regions may not comprise any degenerate nucleotides or sequences.

- 5 Preferably, the barcode region of the first barcoded oligonucleotide comprises a sequence that is complementary and annealed to the barcode region of the first barcode molecule and the barcode region of the second barcoded oligonucleotide comprises a sequence that is complementary and annealed to the barcode region of the second barcode molecule. The complementary sequence of each barcoded oligonucleotide may be at least 5, at least 10, at least 15, at least 20, at least
10 25, at least 50 or at least 100 contiguous nucleotides.

The target regions of the barcoded oligonucleotides (which are not annealed to the multimeric barcode molecule(s)) may be non-complementary to the multimeric barcode molecule(s).

- 15 The barcoded oligonucleotides may comprise a linker region between the barcode region and the target region. The linker region may comprise one or more contiguous nucleotides that are not annealed to the multimeric barcode molecule and are non-complementary to the fragments of the target nucleic acid. The linker may comprise 1 to 100, 5 to 75, 10 to 50, 15 to 30 or 20 to 25 non-complementary nucleotides. Preferably, the linker comprises 15 to 30 non-complementary
20 nucleotides. The use of such a linker region enhances the efficiency of the barcoding reactions performed using the multimeric barcoding reagents.

- Barcode molecules may further comprise one or more nucleic acid sequences that are not complementary to barcode regions of barcoded oligonucleotides. For example, barcode
25 molecules may comprise one or more adapter regions. A barcode molecule, may, for example, comprise an adapter region 5' of a barcode region (a 5' adapter region) and/or an adapter region 3' of the barcode region (a 3' adapter region). The adapter region(s) (and/or one or more portions of an adapter region) may be complementary to and anneal to oligonucleotides e.g. the adapter regions of barcoded oligonucleotides. Alternatively, the adapter region(s) (and/or one or more
30 portions of an adapter region) of barcode molecule may not be complementary to sequences of barcoded oligonucleotides. The adapter region(s) may be used for manipulating, purifying, retrieving, amplifying, and/or detecting barcode molecules.

- 35 The multimeric barcoding reagent may be configured such that: each of the barcode molecules comprises a nucleic acid sequence comprising in the 5' to 3' direction an adapter region and a barcode region; the first barcoded oligonucleotide comprises, optionally in the 5' to 3' direction, a barcode region annealed to the barcode region of the first barcode molecule, an adapter region annealed to the adapter region of the first barcode molecule and a target region capable of annealing to a first fragment of the target nucleic acid; and the second barcoded oligonucleotide

comprises, optionally in the 5' to 3' direction, a barcode region annealed to the barcode region of the second barcode molecule, an adapter region annealed to the adapter region of the second barcode molecule and a target region capable of annealing to a second fragment of the target nucleic acid.

5

The adapter region of each barcode molecule may comprise a constant region. Optionally, all adapter regions of a multimeric barcoding reagent are substantially identical. The adapter region may comprise at least 1, at least 2, at least 3, at least 4, at least 5, at least 6, at least 8, at least 10, at least 15, at least 20, at least 25, at least 50, at least 100, or at least 250 nucleotides.

10

Preferably, the adapter region comprises at least 4 nucleotides. Preferably each adapter region comprises deoxyribonucleotides, optionally all of the nucleotides in an adapter region are deoxyribonucleotides. One or more of the deoxyribonucleotides may be a modified deoxyribonucleotide (e.g. a deoxyribonucleotide modified with a biotin moiety or a deoxyuracil nucleotide). Each adapter region may comprise one or more universal bases (e.g. inosine), one or modified nucleotides and/or one or more nucleotide analogues.

15

The barcoded oligonucleotides may comprise a linker region between the adapter region and the target region. The linker region may comprise one or more contiguous nucleotides that are not annealed to the multimeric barcode molecule and are non-complementary to the fragments of the target nucleic acid. The linker may comprise 1 to 100, 5 to 75, 10 to 50, 15 to 30 or 20 to 25 non-complementary nucleotides. Preferably, the linker comprises 15 to 30 non-complementary nucleotides. The use of such a linker region enhances the efficiency of the barcoding reactions performed using the multimeric barcoding reagents.

20

25

The barcode molecules of a multimeric barcode molecule may be linked on a nucleic acid molecule. Such a nucleic acid molecule may provide the backbone to which single-stranded barcoded oligonucleotides may be annealed. Alternatively, the barcode molecules of a multimeric barcode molecule may be linked together by any of the other means described herein.

30

The multimeric barcoding reagent may comprise: at least 2, at least 3, at least 4, at least 5, at least 10, at least 20, at least 25, at least 50, at least 75, at least 100, at least 200, at least 500, at least 1000, at least 5000, or at least 10,000 barcode molecules linked together, wherein each barcode molecule is as defined herein; and a barcoded oligonucleotide annealed to each barcode molecule, wherein each barcoded oligonucleotide is as defined herein. Preferably, the multimeric barcoding reagent comprises at least 5 barcode molecules linked together, wherein each barcode molecule is as defined herein; and a barcoded oligonucleotide annealed to each barcode molecule, wherein each barcoded oligonucleotide is as defined herein.

35

- 135 -

The multimeric barcoding reagent may comprise: at least 2, at least 3, at least 4, at least 5, at least 10, at least 20, at least 25, at least 50, at least 75, at least 100, at least 200, at least 500, at least 1000, at least 5000, at least 10^4 , at least 10^5 , or at least 10^6 unique or different barcode molecules linked together, wherein each barcode molecule is as defined herein; and a barcoded oligonucleotide annealed to each barcode molecule, wherein each barcoded oligonucleotide is as defined herein. Preferably, the multimeric barcoding reagent comprises at least 5 unique or different barcode molecules linked together, wherein each barcode molecule is as defined herein; and a barcoded oligonucleotide annealed to each barcode molecule, wherein each barcoded oligonucleotide is as defined herein.

10

The multimeric barcoding reagent may comprise: at least 5, at least 10, at least 20, at least 25, at least 50, at least 75, at least 100, at least 200, at least 500, at least 1000, at least 5000, or at least 10,000 barcode regions, wherein each barcode region is as defined herein; and a barcoded oligonucleotide annealed to each barcode region, wherein each barcoded oligonucleotide is as defined herein. Preferably, the multimeric barcoding reagent comprises at least 5 barcode regions, wherein each barcode region is as defined herein; and a barcoded oligonucleotide annealed to each barcode region, wherein each barcoded oligonucleotide is as defined herein.

15

The multimeric barcoding reagent may comprise: at least 2, at least 3, at least 4, at least 5, at least 10, at least 20, at least 25, at least 50, at least 75, at least 100, at least 200, at least 500, at least 1000, at least 5000, at least 10^4 , at least 10^5 , or at least 10^6 unique or different barcode regions, wherein each barcode region is as defined herein; and a barcoded oligonucleotide annealed to each barcode region, wherein each barcoded oligonucleotide is as defined herein. Preferably, the multimeric barcoding reagent comprises at least 5 unique or different barcode regions, wherein each barcode region is as defined herein; and a barcoded oligonucleotide annealed to each barcode region, wherein each barcoded oligonucleotide is as defined herein.

20

25

Figure 1 shows a multimeric barcoding reagent, including first (D1, E1, and F1) and second (D2, E2, and F2) barcode molecules, which each include a nucleic acid sequence comprising a barcode region (E1 and E2). These first and second barcode molecules are linked together, for example by a connecting nucleic acid sequence (S). The multimeric barcoding reagent also comprises first (A1, B1, C1, and G1) and second (A2, B2, C2, and G2) barcoded oligonucleotides. These barcoded oligonucleotides each comprise a barcode region (B1 and B2) and a target region (G1 and G2).

30

35

The barcode regions within the barcoded oligonucleotides may each contain a unique sequence which is not present in other barcoded oligonucleotides, and may thus serve to uniquely identify each such barcode molecule. The target regions may be used to anneal the barcoded

- 136 -

oligonucleotides to fragments of target nucleic acids, and then may be used as primers for a primer-extension reaction or an amplification reaction e.g. a polymerase chain reaction.

Each barcode molecule may optionally also include a 5' adapter region (F1 and F2). The
5 barcoded oligonucleotides may then also include a 3' adapter region (C1 and C2) that is complementary to the 5' adapter region of the barcode molecules.

Each barcode molecule may optionally also include a 3' region (D1 and D2), which may be comprised of identical sequences within each barcode molecule. The barcoded oligonucleotides
10 may then also include a 5' region (A1 and A2) which is complementary to the 3' region of the barcode molecules. These 3' regions may be useful for manipulation or amplification of nucleic acid sequences, for example sequences that are generated by labeling a nucleic acid target with a barcoded oligonucleotide. The 3' region may comprise at least 4, at least 5, at least 6, at least 8, at least 10, at least 15, at least 20, at least 25, at least 50, at least 100, or at least 250
15 nucleotides. Preferably, the 3' region comprises at least 4 nucleotides. Preferably each 3' region comprises deoxyribonucleotides, optionally all of the nucleotides in an 3' region are deoxyribonucleotides. One or more of the deoxyribonucleotides may be a modified deoxyribonucleotide (e.g. a deoxyribonucleotide modified with a biotin moiety or a deoxyuracil nucleotide). Each 3' region may comprise one or more universal bases (e.g. inosine), one or
20 modified nucleotides and/or one or more nucleotide analogues.

The invention provides a library of multimeric barcoding reagents comprising at least 10 multimeric barcoding reagents for labelling a target nucleic acid for sequencing, wherein each multimeric barcoding reagent comprises: first and second barcode molecules comprised within a
25 (single) nucleic acid molecule, wherein each of the barcode molecules comprises a nucleic acid sequence comprising a barcode region; and first and second barcoded oligonucleotides, wherein the first barcoded oligonucleotide comprises, optionally in the 5' to 3' direction, a barcode region complementary and annealed to the barcode region of the first barcode molecule and a target region capable of annealing or ligating to a first fragment of the target nucleic acid, and wherein
30 the second barcoded oligonucleotide comprises, optionally in the 5' to 3' direction, a barcode region complementary and annealed to the barcode region of the second barcode molecule and a target region capable of annealing or ligating to a second fragment of the target nucleic acid. Preferably, the barcode regions of the first and second barcoded oligonucleotides of each multimeric barcoding reagent are different to the barcode regions of the barcoded
35 oligonucleotides of at least 9 other multimeric barcoding reagents in the library.

18. MULTIMERIC BARCODING REAGENTS COMPRISING BARCODED OLIGONUCLEOTIDES LINKED BY A MACROMOLECULE

The invention provides a multimeric barcoding reagent for labelling a target nucleic acid, wherein the reagent comprises first and second barcoded oligonucleotides linked together by a
5 macromolecule, and wherein the barcoded oligonucleotides each comprise a barcode region.

Further details of the barcoded oligonucleotides are provided in PCT/GB2017/053820, which is incorporated herein by reference.

10 The barcoded oligonucleotides may be linked by a macromolecule by being bound to the macromolecule and/or by being annealed to the macromolecule.

The barcoded oligonucleotides may be linked to the macromolecule directly or indirectly (e.g. via a linker molecule). The barcoded oligonucleotides may be linked by being bound to the
15 macromolecule and/or by being bound or annealed to linker molecules that are bound to the macromolecule. The barcoded oligonucleotides may be bound to the macromolecule (or to the linker molecules) by covalent linkage, non-covalent linkage (e.g. a protein-protein interaction or a streptavidin-biotin bond) or nucleic acid hybridization. The linker molecule may be a biopolymer (e.g. a nucleic acid molecule) or a synthetic polymer. The linker molecule may comprise one or
20 more units of ethylene glycol and/or poly(ethylene) glycol (e.g. hexa-ethylene glycol or penta-ethylene glycol). The linker molecule may comprise one or more ethyl groups, such as a C3 (three-carbon) spacer, C6 spacer, C12 spacer, or C18 spacer.

The macromolecule may be a synthetic polymer (e.g. a dendrimer) or a biopolymer such as a
25 nucleic acid (e.g. a single-stranded nucleic acid such as single-stranded DNA), a peptide, a polypeptide or a protein (e.g. a multimeric protein).

The dendrimer may comprise at least 2, at least 3, at least 5, or at least 10 generations.

30 The macromolecule may be a nucleic acid comprising two or more nucleotides each capable of binding to a barcoded oligonucleotide. Additionally or alternatively, the nucleic acid may comprise two or more regions each capable of hybridizing to a barcoded oligonucleotide.

The nucleic acid may comprise a first modified nucleotide and a second modified nucleotide,
35 wherein each modified nucleotide comprises a binding moiety (e.g. a biotin moiety, or an alkyne moiety which may be used for a click-chemical reaction) capable of binding to a barcoded oligonucleotide. Optionally, the first and second modified nucleotides may be separated by an intervening nucleic acid sequence of at least one, at least two, at least 5 or at least 10 nucleotides.

The nucleic acid may comprise a first hybridisation region and a second hybridisation region, wherein each hybridisation region comprises a sequence complementary to and capable of hybridizing to a sequence of at least one nucleotide within a barcoded oligonucleotide. The complementary sequence may be at least 5, at least 10, at least 15, at least 20, at least 25 or at least 50 contiguous nucleotides. Optionally, the first and second hybridisation regions may be separated by an intervening nucleic acid sequence of at least one, at least two, at least 5 or at least 10 nucleotides.

The macromolecule may be a protein such as a multimeric protein e.g. a homomeric protein or a heteromeric protein. For example, the protein may comprise streptavidin e.g. tetrameric streptavidin.

Libraries of multimeric barcoding reagents comprising barcoded oligonucleotides linked by a macromolecule are also provided. Such libraries may be based on the general properties of libraries of multimeric barcoding reagents described herein. In the libraries, each multimeric barcoding reagent may comprise a different macromolecule.

19. MULTIMERIC BARCODING REAGENTS COMPRISING BARCODED OLIGONUCLEOTIDES LINKED BY A SOLID SUPPORT OR A SEMI-SOLID SUPPORT

The invention provides a multimeric barcoding reagent for labelling a target nucleic acid, wherein the reagent comprises first and second barcoded oligonucleotides linked together by a solid support or a semi-solid support, and wherein the barcoded oligonucleotides each comprise a barcode region.

The first barcoded oligonucleotide may further comprise a target region capable of annealing or ligating to a first fragment of the target nucleic acid, and the second barcoded oligonucleotide may further comprise a target region capable of annealing or ligating to a second fragment of the target nucleic acid.

The first barcoded oligonucleotide may comprise in the 5'-3' direction a barcode region and a target region capable of annealing to a first fragment of the target nucleic acid, and the second barcoded oligonucleotide may comprise in the 5'-3' direction a barcode region and a target region capable of annealing to a second fragment of the target nucleic acid.

The barcoded oligonucleotides may further comprise any of the features described herein.

The barcoded oligonucleotides may be linked by a solid support or a semi-solid support. The barcoded oligonucleotides may be linked to the support directly or indirectly (e.g. via a linker

- molecule). The barcoded oligonucleotides may be linked by being bound to the support and/or by being bound or annealed to linker molecules that are bound to the support. The barcoded oligonucleotides may be bound to the support (or to the linker molecules) by covalent linkage, non-covalent linkage (e.g. a protein-protein interaction or a streptavidin-biotin bond) or nucleic acid hybridization. The linker molecule may be a biopolymer (e.g. a nucleic acid molecule) or a synthetic polymer. The linker molecule may comprise one or more units of ethylene glycol and/or poly(ethylene) glycol (e.g. hexa-ethylene glycol or penta-ethylene glycol). The linker molecule may comprise one or more ethyl groups, such as a C3 (three-carbon) spacer, C6 spacer, C12 spacer, or C18 spacer. The linker molecule may comprise at least 2, at least 3, at least 4, at least 5, at least 10, or at least 20 sequential repeating units of any individual linker (such as a sequential linear series of at least 2, at least 5, or at least 10 C12 spacers or C18 spacers). The linker molecule may comprise a branched linker molecule, wherein 2 or more barcode molecules are linked to a support by a single linker molecule.
- The support may comprise a planar surface. The support may be a slide e.g. a glass slide. The slide may be a flow cell for sequencing. If the support is a slide, the first and second barcoded oligonucleotides may be immobilized in a discrete region on the slide. Optionally, the barcoded oligonucleotides of each multimeric barcoding reagent in a library are immobilized in a different discrete region on the slide to the barcoded oligonucleotides of the other multimeric barcoding reagents in the library. The support may be a plate comprising wells, optionally wherein the first and second barcoded oligonucleotides are immobilized in the same well. Optionally, the barcoded oligonucleotides of each multimeric barcoding reagent in library are immobilized in a different well of the plate to the barcoded oligonucleotides of the other multimeric barcoding reagents in the library.
- Preferably, the support is a bead (e.g. a gel bead). The bead may be an agarose bead, a silica bead, a styrofoam bead, a gel bead (such as those available from 10x Genomics®), an antibody conjugated bead, an oligo-dT conjugated bead, a streptavidin bead or a magnetic bead (e.g. a superparamagnetic bead). The bead may be of any size and/or molecular structure. For example, the bead may be 10 nanometres to 100 microns in diameter, 100 nanometres to 10 microns in diameter, or 1 micron to 5 microns in diameter. Optionally, the bead is approximately 10 nanometres in diameter, approximately 100 nanometres in diameter, approximately 1 micron in diameter, approximately 10 microns in diameter or approximately 100 microns in diameter. The bead may be solid, or alternatively the bead may be hollow or partially hollow or porous. Beads of certain sizes may be most preferable for certain barcoding methods. For example, beads less than 5.0 microns, or less than 1.0 micron, may be most useful for barcoding nucleic acid targets within individual cells. Preferably, the barcoded oligonucleotides of each multimeric barcoding reagent in a library are linked together on a different bead to the barcoded oligonucleotides of the other multimeric barcoding reagents in the library.

5 The support may be functionalised to enable attachment of two or more barcoded oligonucleotides. This functionalisation may be enabled through the addition of chemical moieties (e.g. carboxylated groups, alkynes, azides, acrylate groups, amino groups, sulphate groups, or succinimide groups), and/or protein-based moieties (e.g. streptavidin, avidin, or protein G) to the support. The barcoded oligonucleotides may be attached to the moieties directly or indirectly (e.g. via a linker molecule).

10 Functionalised supports (e.g. beads) may be brought into contact with a solution of barcoded oligonucleotides under conditions which promote the attachment of two or more barcoded oligonucleotides to each bead in the solution (generating multimeric barcoding reagents).

15 Libraries of multimeric barcoding reagents comprising barcoded oligonucleotides linked by a support are also provided. Such libraries may be based on the general properties of libraries of multimeric barcoding reagents described herein. In the libraries, each multimeric barcoding reagent may comprise a different support (e.g. a differently labelled bead). In a library of multimeric barcoding reagents, the barcoded oligonucleotides of each multimeric barcoding reagent in a library may be linked together on a different support to the barcoded oligonucleotides of the other multimeric barcoding reagents in the library.

20

20. METHODS OF PREPARING A NUCLEIC ACID SAMPLE FOR SEQUENCING

The methods of preparing a nucleic acid sample for sequencing may comprise (i) contacting the nucleic acid sample with a multimeric barcoding reagent comprising first and second barcode regions linked together, wherein each barcode region comprises a nucleic acid sequence, and
25 (ii) appending barcode sequences to first and second fragments of a target nucleic acid to produce first and second different barcoded target nucleic acid molecules, wherein the first barcoded target nucleic acid molecule comprises the nucleic acid sequence of the first barcode region and the second barcoded target nucleic acid molecule comprises the nucleic acid sequence of the second barcode region.

30

In methods in which the multimeric barcoding reagent comprises first and second barcoded oligonucleotides linked together, the barcode sequences may be appended to first and second fragments of the target nucleic acid by any of the methods described herein.

35 The first and second barcoded oligonucleotides may be ligated to the first and second fragments of the target nucleic acid to produce the first and second different barcoded target nucleic acid molecules. Optionally, prior to the ligation step, the method comprises appending first and second coupling sequences to the target nucleic acid, wherein the first and second coupling

sequences are the first and second fragments of the target nucleic acid to which the first and second barcoded oligonucleotides are ligated.

The first and second barcoded oligonucleotides may be annealed to the first and second fragments of the target nucleic acid extended to produce the first and second different barcoded target nucleic acid molecules. Optionally, prior to the annealing step, the method comprises appending first and second coupling sequences to the target nucleic acid, wherein the first and second coupling sequences are the first and second fragments of the target nucleic acid to which the first and second barcoded oligonucleotides are annealed.

The first and second barcoded oligonucleotides may be annealed at their 5' ends to the first and second sub-sequences of the target nucleic acid and first and second target primers may be annealed to third and fourth sub-sequences of the target nucleic acid, respectively, wherein the third subsequence is 3' of the first subsequence and wherein the fourth sub-sequence is 3' of the second subsequence. The method further comprises extending the first target primer using the target nucleic acid as template until it reaches the first sub-sequence to produce a first extended target primer, and extending the second target primer using the target nucleic acid as template until it reaches the second sub-sequence to produce a second extended target primer, and ligating the 3' end of the first extended target primer to the 5' end of the first barcoded oligonucleotide to produce a first barcoded target nucleic acid molecule, and ligating the 3' end of the second extended target primer to the 5' end of the second barcoded oligonucleotide to produce a second barcoded target nucleic acid molecule, wherein the first and second barcoded target nucleic acid molecules are different and each comprises at least one nucleotide synthesised from the target nucleic acid as a template. Optionally, prior to either or both annealing step(s), the method comprises appending first and second, and/or third and fourth, coupling sequences to the target nucleic acid, wherein the first and second coupling sequences are the first and second sub-sequences of the target nucleic acid to which the first and second barcoded oligonucleotides are annealed, and/or wherein the third and fourth coupling sequences are the third and fourth sub-sequences of the target nucleic acid to which the first and second target primers are annealed.

As described herein, prior to annealing or ligating a multimeric hybridization molecule, multimeric barcode molecule, barcoded oligonucleotide, adapter oligonucleotide or target primer to a target nucleic acid, a coupling sequence may be appended to the target nucleic acid. The multimeric hybridization molecule, multimeric barcode molecule, barcoded oligonucleotide, adapter oligonucleotide or target primer may then be annealed or ligated to the coupling sequence.

- 142 -

A coupling sequence may be added to the 5' end or 3' end of two or more target nucleic acids of the nucleic acid sample. In this method, the target regions (of the barcoded oligonucleotides) may comprise a sequence that is complementary to the coupling sequence.

5 A coupling sequence may be comprised within a double-stranded coupling oligonucleotide or within a single-stranded coupling oligonucleotide. A coupling oligonucleotide may be appended to the target nucleic acid by a double-stranded ligation reaction or a single-stranded ligation reaction. A coupling oligonucleotide may comprise a single-stranded 5' or 3' region capable of
10 ligating to a target nucleic acid and the coupling sequence may be appended to the target nucleic acid by a single-stranded ligation reaction.

A coupling oligonucleotide may comprise a blunt, recessed, or overhanging 5' or 3' region capable of ligating to a target nucleic acid and the coupling sequence may be appended to the target nucleic acid a double-stranded ligation reaction.

15 The end(s) of a target nucleic acid may be converted into blunt double-stranded end(s) in a blunting reaction, and the coupling oligonucleotide may comprise a blunt double-stranded end, and wherein the coupling oligonucleotide may be ligated to the target nucleic acid in a blunt-end ligation reaction.

20 The end(s) of a target nucleic acid may be converted into blunt double-stranded end(s) in a blunting reaction, and then converted into a form with (a) single 3' adenosine overhang(s), and wherein the coupling oligonucleotide may comprise a double-stranded end with a single 3' thymine overhang capable of annealing to the single 3' adenosine overhang of the target nucleic
25 acid, and wherein the coupling oligonucleotide is ligated to the target nucleic acid in a double-stranded A/T ligation reaction

The target nucleic acid may be contacted with a restriction enzyme, wherein the restriction enzyme digests the target nucleic acid at restriction sites to create (a) ligation junction(s) at the
30 restriction site(s), and wherein the coupling oligonucleotide comprises an end compatible with the ligation junction, and wherein the coupling oligonucleotide is then ligated to the target nucleic acid in a double-stranded ligation reaction.

35 A coupling oligonucleotide may be appended via a primer-extension or polymerase chain reaction step.

A coupling oligonucleotide may be appended via a primer-extension or polymerase chain reaction step, using one or more oligonucleotide(s) that comprise a priming segment including one or more degenerate bases.

A coupling oligonucleotide may be appended via a primer-extension or polymerase chain reaction step, using one or more oligonucleotide(s) that further comprise a priming or hybridisation segment specific for a particular target nucleic acid sequence.

5

A coupling sequence may be added by a polynucleotide tailing reaction. A coupling sequence may be added by a terminal transferase enzyme (e.g. a terminal deoxynucleotidyl transferase enzyme). A coupling sequence may be appended via a polynucleotide tailing reaction performed with a terminal deoxynucleotidyl transferase enzyme, and wherein the coupling sequence

10

comprises at least two contiguous nucleotides of a homopolymeric sequence.

A coupling sequence may comprise a homopolymeric 3' tail (e.g. a poly(A) tail). Optionally, in such methods, the target regions (of the barcoded oligonucleotides) comprise a complementary homopolymeric 3' tail (e.g. a poly(T) tail).

15

A coupling sequence may be comprised within a synthetic transposome, and may be appended via an in vitro transposition reaction.

A coupling sequence may be appended to a target nucleic acid, and wherein a barcode oligonucleotide is appended to the target nucleic acid by at least one primer-extension step or polymerase chain reaction step, and wherein said barcode oligonucleotide comprises a region of at least one nucleotide in length that is complementary to said coupling sequence. Optionally, this region of complementarity is at the 3' end of the barcode oligonucleotide. Optionally, this region of complementarity is at least 2 nucleotides in length, at least 5 nucleotides in length, at least 10

20

25

nucleotides in length, at least 20 nucleotides in length, or at least 50 nucleotides in length.

In methods in which an adapter oligonucleotide is appended (e.g. ligated or annealed) to a target nucleic acid, the adapter region of the adapter oligonucleotide provides a coupling sequence capable of hybridizing to the adapter region of a multimeric hybridization molecule or a multimeric barcode molecule.

30

The invention provides a method of preparing a nucleic acid sample for sequencing comprising the steps of: (a) appending a coupling sequence to first and second fragments of a target nucleic acid; (b) contacting the nucleic acid sample with a multimeric barcoding reagent comprising first and second barcode molecules linked together, wherein each of the barcode molecules comprises a nucleic acid sequence comprising (in the 5' to 3' or 3' to 5' direction), a barcode region and an adapter region; (c) annealing the coupling sequence of the first fragment to the adapter region of the first barcode molecule, and annealing the coupling sequence of the second fragment to the adapter region of the second barcode molecule; and (d) appending barcode

35

- 144 -

sequences to each of the at least two fragments of the target nucleic acid to produce first and second different barcoded target nucleic acid molecules, wherein the first barcoded target nucleic acid molecule comprises the nucleic acid sequence of the barcode region of the first barcode molecule and the second barcoded target nucleic acid molecule comprises the nucleic acid sequence of the barcode region of the second barcode molecule.

In the method, each of the barcode molecules may comprise a nucleic acid sequence comprising, in the 5' to 3' direction, a barcode region and an adapter region, and step (d) may comprise extending the coupling sequence of the first fragment of the target nucleic acid using the barcode region of the first barcode molecule as a template to produce a first barcoded target nucleic acid molecule, and extending the coupling sequence of the second fragment of the target nucleic acid using the barcode region of the second barcode molecule as a template to produce a second barcoded target nucleic acid molecule, wherein the first barcoded target nucleic acid molecule comprises a sequence complementary to the barcode region of the first barcode molecule and the second barcoded target nucleic acid molecule comprises a sequence complementary to the barcode region of the second barcode molecule.

In the method, each of the barcode molecules may comprise a nucleic acid sequence comprising, in the 5' to 3' direction, an adapter region and a barcode region, and step (d) may comprise (i) annealing and extending a first extension primer using the barcode region of the first barcode molecule as a template to produce a first barcoded oligonucleotide, and annealing and extending a second extension primer using the barcode region of the second barcode molecule as a template to produce a second barcoded oligonucleotide, wherein the first barcoded oligonucleotide comprises a sequence complementary to the barcode region of the first barcode molecule and the second barcoded oligonucleotide comprises a sequence complementary to the barcode region of the second barcode molecule, (ii) ligating the 3' end of the first barcoded oligonucleotide to the 5' end of the coupling sequence of the first fragment of the target nucleic acid to produce a first barcoded target nucleic acid molecule and ligating the 3' end of the second barcoded oligonucleotide to the 5' end of the coupling sequence of the second fragment of the target nucleic acid to produce a second barcoded target nucleic acid molecule.

In the method, each of the barcode molecules may comprise a nucleic acid sequence comprising, in the 5' to 3' direction, an adapter region, a barcode region and a priming region wherein step (d) comprises (i) annealing a first extension primer to the priming region of the first barcode molecule and extending the first extension primer using the barcode region of the first barcode molecule as a template to produce a first barcoded oligonucleotide, and annealing a second extension primer to the priming region of the second barcode molecule and extending the second extension primer using the barcode region of the second barcode molecule as a template to produce a second barcoded oligonucleotide, wherein the first barcoded oligonucleotide comprises a sequence

- 145 -

complementary to the barcode region of the first barcode molecule and the second barcoded oligonucleotide comprises a sequence complementary to the barcode region of the second barcode molecule, (ii) ligating the 3' end of the first barcoded oligonucleotide to the 5' end of the coupling sequence of the first fragment of the target nucleic acid to produce a first barcoded target nucleic acid molecule and ligating the 3' end of the second barcoded oligonucleotide to the 5' end of the coupling sequence of the second fragment of the target nucleic acid to produce a second barcoded target nucleic acid molecule.

The methods for preparing a nucleic acid sample for sequencing may be used to prepare a range of different nucleic acid samples for sequencing. The target nucleic acids may be DNA molecules (e.g. genomic DNA molecules) or RNA molecules (e.g. mRNA molecules). The target nucleic acids may be from any sample. For example, an individual cell (or cells), a tissue, a bodily fluid (e.g. blood, plasma and/or serum), a biopsy or a formalin-fixed paraffin-embedded (FFPE) sample.

The sample may comprise at least 10, at least 100, or at least 10^3 , at least 10^4 , at least 10^5 , at least 10^6 , at least 10^7 , at least 10^8 or at least 10^9 target nucleic acids

The method may comprise producing at least 2, at least 5, at least 10, at least 20, at least 25, at least 50, at least 75, at least 100, at least 250, at least 500, at least 10^3 , at least 10^4 , at least 10^5 , at least 10^6 , at least 10^7 , at least 10^8 or at least 10^9 different barcoded target nucleic acid molecules. Preferably, the method comprises producing at least 5 different barcoded target nucleic acid molecules.

Each barcoded target nucleic acid molecule may comprise at least 1, at least 5, at least 10, at least 25, at least 50, at least 100, at least 250, at least 500, at least 1000, at least 2000, at least 5000, or at least 10,000 nucleotides synthesised from the target nucleic acid as template. Preferably, each barcoded target nucleic acid molecule comprises at least 20 nucleotides synthesised from the target nucleic acid as template.

Alternatively, each barcoded target nucleic acid molecule may comprise at least 5, at least 10, at least 25, at least 50, at least 100, at least 250, at least 500, at least 1000, at least 2000, at least 5000, or at least 10,000 nucleotides of the target nucleic acid. Preferably, each barcoded target nucleic acid molecule comprises at least 5 nucleotides of the target nucleic acid.

A universal priming sequence may be added to the barcoded target nucleic acid molecules. This sequence may enable the subsequent amplification of at least 5, at least 10, at least 20, at least 25, at least 50, at least 75, at least 100, at least 250, at least 500, at least 10^3 , at least 10^4 , at

least 10^5 , at least 10^6 , at least 10^7 , at least 10^8 , or at least 10^9 different barcoded target nucleic acid molecules using one forward primer and one reverse primer.

Optionally, in any method of analysing a sample comprising a circulating microparticle or a sample derived from a circulating microparticle wherein the method comprises appending and/or linking and/or connecting barcode sequences comprised within multimeric barcoding reagents to target molecules such as target nucleic acid molecules (e.g. wherein the method comprises appending and/or linking and/or connecting barcoded oligonucleotides comprised within multimeric barcoding reagents to target molecules such as target nucleic acid molecules), barcode sequences (e.g. barcoded oligonucleotides) from or comprised within any number of different multimeric barcoding reagents may be so appended and/or linked and/or connected. For example, barcoded oligonucleotides from at least 2, at least 3, at least 5, at least 10, at least 50, at least 100, or at least 1000 different multimeric barcoding reagents may be appended and/or linked and/or connected to target nucleic acid molecules comprised within or derived from a single circulating microparticle; optionally such ratios of multimeric barcoding reagents-per-circulating microparticle may be true on average for any or all circulating microparticles within a sample of circulating microparticles. Optionally, in any method wherein barcoded oligonucleotides from 2 or more multimeric barcoding reagents are appended and/or linked and/or connected to target nucleic acid molecules comprised within or derived from a single circulating microparticle, any number of 1 or more barcode sequences from any first such multimeric barcoding reagent may be appended to any number of 1 or more barcode sequences from any second such multimeric barcoding reagent (a 'cross-barcoding reaction'), in such manner that the resulting barcode-to-barcode appended molecules may be sequenced with a sequencing reaction, in such manner that said 2 or more multimeric barcoding reagents may be identified as having participated in a 'cross-barcoding reaction' with each other and thus co-localised (e.g. occupied physically close spatial proximity and/or occupied nearby or overlapping (or partially overlapping) physical volumes in solution) and co-labelled (i.e. co-barcoded) the same (physically close or nearby) single circulating microparticle (or sample or target biomolecules comprised therein and/or derived therefrom); optionally any or all target nucleic acid molecules appended to barcode sequences (e.g. barcoded oligonucleotides) comprised within any first multimeric barcoding reagent that has participated in a 'cross-barcoding reaction' may be considered or found to be linked to any or all target nucleic acid molecules appended to barcode sequences (e.g. barcoded oligonucleotides) comprised within any second multimeric barcoding reagent that has participated in the same 'cross-barcoding reaction'; optionally this 'cross-barcoding reaction' method may be employed for any or all circulating microparticles and/or any or all multimeric barcoding reagents and/or any or all target nucleic acid molecules within any method(s) described herein.

Optionally, in any method wherein barcoded oligonucleotides from 2 or more multimeric barcoding reagents are appended and/or linked and/or connected to target nucleic acid molecules

– 147 –

comprised within or derived from a single circulating microparticle, any number of 1 or more barcode sequences (e.g. comprised within barcoded oligonucleotides) from each of first and second (or more) such multimeric barcoding reagents may be appended to molecular identifier sequences comprised within a single synthetic DNA template (e.g. a single-stranded synthetic DNA template) to create 'barcode-to-molecular-identifier-sequence' molecules, wherein said synthetic DNA template comprises at least 2 copies (e.g. at least 2 tandemly-repeated copies) of a molecular identifier sequence, wherein said molecular identifier sequence comprises an identifier sequence at least 1 nucleotide in length (or at least 2, at least 5, at least 10, at least 15, at least 20, at least 30, or at least 50 nucleotides in length), and wherein said identifier sequence is the same (i.e. identical in sequence) for all molecular identifier sequences within a (and/or each) single synthetic DNA template (and optionally, wherein said identifier sequence is different between each of two or more different synthetic DNA templates).

Optionally, each such molecular identifier sequence may comprise (at the 5' end and/or at the 3' end) one or more adapter sequences (the one or more adapter sequences may be of any length); optionally any one or more such adapter sequences may be the same for all molecular identifier sequences and/or synthetic DNA templates (such as within a library of different synthetic DNA templates). Optionally, any one or more such adapter sequences may be partially or fully complementary to any target sequence(s) within barcoded oligonucleotides (e.g. barcoded oligonucleotides within a library of multimeric barcoding reagents). Optionally, a library of 2 or more different synthetic DNA templates may be employed, wherein the identifier sequence is the same (i.e. identical in sequence) for all molecular identifier sequences within a single synthetic DNA template, but wherein the molecular identifier sequence is different between 2 or more different single synthetic DNA templates. Optionally, a library of synthetic DNA templates may comprise at least 10, at least 100, at least 1000, at least 1,000,000, at least 10,000,000, at least 100,000,000, at least 1,000,000,000, or at least 100,000,000,000 different synthetic DNA templates (e.g. wherein each synthetic DNA template within said library comprises a different identifier sequence). Optionally each such individual (different) synthetic DNA template may be present at any concentration (e.g. at 2 or more copies) within a library and/or solution. Methods of synthesising and using synthetic DNA templates and/or libraries thereof are described in Methods 5, 6, and 7 herein.

Optionally, a sample comprising or derived from one or more circulating microparticle (of any sort, and at any concentration, as described herein), may be combined to form a solution (e.g. within a contiguous aqueous volume) with a library of multimeric barcoding reagents (of any sort and at any concentration described herein) and with a library of 2 or more synthetic DNA templates (of any sort and at any concentration described herein), and barcode sequences (e.g. barcoded oligonucleotides) from said multimeric barcoding reagents may then be appended and/or linked and/or connected (by any one or methods described herein) to target nucleic acid molecules

comprised within or derived from said circulating microparticles and also appended and/or linked and/or connected (by any one or methods described herein) to molecular identifier sequences comprised within said library of synthetic DNA templates (optionally wherein all such appending and/or linking and/or connecting takes place in a single and/or simultaneous step), optionally in such manner that barcode molecules from any 2 or more different multimeric barcoding reagents (e.g. from barcoded oligonucleotides from any 2 or more multimeric barcoding reagents) may be appended to molecular identifier sequences comprised within a single synthetic DNA template (e.g. to a single synthetic DNA template in physical proximity to said multimeric barcoding reagents within said solution), and optionally wherein the resulting barcode-to-molecular-identifier-sequence molecules are then sequenced with a sequencing reaction, in such manner that barcode molecules from any 2 or more different multimeric barcoding reagents appended to the same molecular identifier sequence (i.e. to a molecular identifier sequence comprised within a single synthetic DNA templates) may be identified as having participated in a 'cross-barcoding reaction' with each other and thus identified as having co-localised and co-labelled (i.e. co-barcoded) target molecules from the same single circulating microparticle (or sample derived therefrom); optionally any or all target nucleic acid molecules appended to barcode sequences comprised within any multimeric barcoding reagent that has participated in such a 'cross-barcoding reaction' may be considered or found to be linked to any or all target nucleic acid molecules appended to barcode sequences comprised within any other multimeric barcoding reagent that has participated in said 'cross-barcoding reaction' (e.g. other multimeric barcoding reagents that have had one or more constituent barcode sequences thereof appended to the same molecular identifier sequence). Optionally, any number or total number of 'barcode-to-molecular-identifier-sequence' molecules (e.g. as determined from a sequencing reaction) may be counted and/or quantified (e.g. by counting the number of reads, and/or counting the number of unique reads, resulting from a sequencing reaction, wherein each read comprises any given pairing of: 1) all or part of a barcode sequence/barcoded oligonucleotide sequence (or complement thereof) from a multimeric barcoding reagent, and 2) all or part of a molecular identifier sequence (or complement thereof) from a synthetic DNA template; optionally the total number of reads (and/or unique reads) resulting from any such sequencing reaction comprising all or part of any barcode sequence/barcoded oligonucleotide sequence (or complement thereof) from a first multimeric barcoding reagent (i.e. within a library of multimeric barcoding reagents) and also comprising all or part of a specific, single molecular identifier sequence (or complement thereof) from a synthetic DNA template may be counted to create a first labelling count, and the total number of reads (and/or unique reads) resulting from said sequencing reaction comprising all or part of any barcode sequence/barcoded oligonucleotide sequence (or complement thereof) from a second multimeric barcoding reagent (i.e. within a library of multimeric barcoding reagents) and also comprising all or part of said specific, single molecular identifier sequence (or complement thereof) from said synthetic DNA template may be counted to create a second labelling read count. Optionally the sum of said first and second labelling read counts may be

considered as a weighting value to determine a degree of connectedness and/or linking and/or degree of physical proximity and/or probability of linking between said first and second multimeric barcoding reagents. Optionally, each of said first and second labelling read counts may be compared with a count cutoff or threshold value, such that in reactions wherein both the first and second labelling read counts are equal to or greater than said count cutoff or threshold value, said first and second multimeric barcoding reagents may be considered to be linked (and, by extension, any target biomolecules such as target nucleic acid molecules that are labelled by barcode sequence(s)/barcoded oligonucleotide(s) from either said of said first or second multimeric barcoding reagents may also be considered to be linked). Potential such count cutoff or threshold values include 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 30, 50, 100, 200, 500, or 1000 reads. Optionally, any such labelling read counts and/or analyses thereof may be performed for any or all pairwise comparisons of two different multimeric barcoding reagents within a library of multimeric barcoding reagents. Optionally, any such labelling read counts and/or analyses thereof may be performed higher-order comparisons of sets comprising three or more different multimeric barcoding reagents within a library of multimeric barcoding reagents (such as sets of at least 5, at least 10, at least 20, at least 50, at least 100, or at least 1000 different multimeric barcoding reagents); optionally, all multimeric barcoding reagents within any such set of different multimeric barcoding reagents may be considered linked to each other (i.e. considered to have participated in a 'cross-barcoding reaction') if the labelling read count for each multimeric barcoding reagent within said set corresponding to any single, specific molecular identifier sequence is equal to or greater than any particular count cutoff or threshold values. Optionally, any labelling read count for a particular multimeric barcoding reagent may be divided by the total number of reads (and/or unique reads) in the sequencing reaction comprising (all or part of) any barcode sequences/barcoded oligonucleotide sequences from said multimeric barcoding reagent, to calculate a normalised labelling read count; optionally said normalised labelling read count may then be compared with a normalised count cutoff or threshold value, such that all multimeric barcoding reagents within any set of different multimeric barcoding reagents may be considered linked to each other if the labelling read count for each multimeric barcoding reagent within said set corresponding to any single, specific molecular identifier sequence is equal to or greater than any particular normalised count cutoff or threshold value. Potential such normalised count cutoff or threshold values include 0.00000001, 0.0000001, 0.000001, 0.00001, 0.0001, 0.001, 0.002, 0.003, 0.004, 0.005, 0.0075, 0.01, 0.015, 0.02, 0.03, 0.04, 0.05, 0.075, 0.10, 0.15, 0.20, 0.25, or 0.30.

Optionally, prior to and/or during any step of appending and/or linking and/or connecting barcode sequences to target nucleic acid molecules and/or molecular identifier sequences comprised within a library of synthetic DNA templates, the synthetic DNA templates within said library of synthetic DNA templates may be dissolved (e.g. freely-floating and diffusible) within the solution (i.e. within the reaction solution and/or contiguous aqueous volume). Optionally, prior to and/or

- 150 -

during any step of appending and/or linking and/or connecting barcode sequences to target nucleic acid molecules and/or molecular identifier sequences comprised within a library of synthetic DNA templates, the synthetic DNA templates within said library of synthetic DNA templates may be appended to one or more circulating microparticles and/or molecules

5 comprised within or upon said circulating microparticle(s) (e.g. within or from a sample comprising one or more circulating microparticles); optionally said synthetic DNA templates may comprise one or more sequences that is complementary to one or more coupling sequences and/or adapter sequences (such as coupling sequences within coupling molecules) wherein said coupling sequences are first appended to target biomolecules (such as target nucleic acids) within said

10 circulating microparticle(s) and wherein said synthetic DNA templates are then appended to the complementary sequences within said coupling sequences; optionally, any step of appending synthetic DNA templates to one or more circulating microparticles and/or molecules comprised within or upon said circulating microparticle(s) may further comprise a further or simultaneous step of appending multimeric barcoding reagents to said circulating microparticles and/or

15 molecules comprised within or upon said circulating microparticle(s) (for example, wherein barcoded oligonucleotides within said multimeric barcoding reagents comprise a sequence (such as a sequence within their target region) complementary to coupling sequences that are first appended to target biomolecules such as target nucleic acids comprised within said circulating microparticle(s)). Optionally, any synthetic DNA templates may be comprised within (i.e. comprise part of) any coupling molecule.

20

Optionally, prior to and/or during any step of appending and/or linking and/or connecting barcode sequences from a library of multimeric barcoding reagents to target nucleic acid molecules and/or molecular identifier sequences comprised within a library of synthetic DNA templates, the

25 multimeric barcoding reagents within said library of multimeric barcoding reagents may be dissolved (e.g. freely-floating and diffusible) within the solution (i.e. within the reaction solution and/or contiguous aqueous volume). Optionally, prior to and/or during any step of appending and/or linking and/or connecting barcode sequences to target nucleic acid molecules and/or molecular identifier sequences comprised within a library of synthetic DNA templates, the

30 multimeric barcoding reagents may be bound to one or more circulating microparticles and/or molecules comprised within or upon said circulating microparticle(s) (e.g. within or from a sample comprising one or more circulating microparticles) by a 'multimeric barcoding reagent binding step'; optionally said multimeric barcoding reagents may comprise one or more sequences (e.g. comprised within their constituent barcoded oligonucleotides) that is complementary to one or

35 more coupling sequences and/or adapter sequences (such as coupling sequences within coupling molecules) wherein said coupling sequences are first appended to target biomolecules (such as target nucleic acids) within said circulating microparticle(s) and wherein said multimeric barcoding reagents are then annealed to the complementary sequences within said coupling sequences. Optionally, any 'multimeric barcoding reagent binding step' (such as any process of

- 151 -

annealing multimeric barcoding reagents to coupling sequences within one or more circulating microparticles and/or molecules comprised within or upon said circulating microparticle(s)) may precede any subsequent process of appending barcode sequences to target biomolecules (such as any process described herein). Optionally, any 'multimeric barcoding reagent binding step' may precede any subsequent dissociation process, wherein said dissociation process comprises dissociating barcoded oligonucleotides from the barcode molecules to which they are annealed in a dissociation process, such as through a heat-denaturation (i.e. duplex-melting) step, and/or any other type of dissociation process as described herein; optionally, any such dissociation process may then be further followed by any process or method of appending barcode sequences (such as appending barcode sequences and/or barcoded oligonucleotides through an annealing process).

Optionally, prior to and/or during any step of appending and/or linking and/or connecting barcode sequences (in the form of barcoded oligonucleotides) to target nucleic acid molecules (such as any 'cross-barcoding reaction' process and/or step), barcoded oligonucleotides may be dissociated from the barcode molecules to which they are annealed in a dissociation process, such as through a heat-denaturation (i.e. duplex-melting) step; optionally such a dissociation process may be at least 1 second in length, at least 5 seconds in length, at least 10 seconds in length, at least 15 seconds in length, at least 20 seconds in length, at least 30 seconds in length, at least 45 seconds in length, at least 60 seconds in length, at least 90 seconds in length, at least 2 minutes in length, at least 3 minutes in length, at least 5 minutes in length, or at least 10 minutes in length; optionally such a dissociation process may be conducted at any temperature such as at least 45 degrees Celsius, at least 50 degrees Celsius, at least 55 degrees Celsius, at least 60 degrees Celsius, at least 65 degrees Celsius, at least 70 degrees Celsius, or at least 70 degrees Celsius; optionally such a dissociation process may be conducted in the presence of nucleic acid denaturant such as DMSO and/or betaine, optionally wherein said nucleic acid denaturant is at a concentration of at least 5% by weight or volume, at least 10% by weight or volume, at least 15% by weight or volume, at least 20% by weight or volume, at least 25% by weight or volume, at least 30% by weight or volume, at least 35% by weight or volume, at least 40 % by weight or volume, or at least 50% by weight or volume; optionally such a dissociation process and/or heat-denaturation step may be followed immediately by an annealing process, wherein barcoded oligonucleotides are annealed to target nucleic acids, optionally wherein said annealing process comprises a process of lowering the temperature of the solution to a temperature conducive to said annealing; optionally any such dissociation process and/or annealing step may be performed in a high-viscosity solution (such as any high-viscosity solution described herein).

Optionally, prior to and/or during and/or following any step of appending and/or linking and/or connecting barcode sequences (such as in the form of barcoded oligonucleotides, such as in the

- 152 -

form of barcoded oligonucleotides comprised within multimeric barcoding reagents) to target nucleic acid molecules, and/or prior to and/or during and/or following any step of proteinase digestion, and/or any step of crosslink reversal (e.g. the reversal of formaldehyde crosslinks), and/or any step of purifying barcoded target nucleic acid molecules, barcoded nucleic acid molecules from and/or comprised within and/or derived from two or more different samples (such as samples from two or more different patients) may be combined (i.e. merged together) into a 'pooled sample solution'. Optionally, any such pooled sample solution may be further processed in any way, such as further prepared and/or modified and/or amplified for high-throughput sequencing, and/or processed in any enrichment step, such as any enrichment process comprising enrichment for a modified nucleotide such as 5-methylcytosine, or 5-hydroxy-methylcytosine (such as wherein said enrichment is performed using an enrichment probe that is specific for or preferentially binds 5-methylcytosine or 5-hydroxy-methylcytosine in fragments of genomic DNA compared with other modified or unmodified bases; optionally, one or more sequencing processes may then be performed to analyse said enriched (i.e. enrichment-probe-bound) barcoded nucleic acids and/or the resulting modified-nucleotide-depleted (i.e. non-enrichment-probe-bound) barcoded nucleic acids.

The method may comprise preparing two or more independent nucleic acid samples for sequencing, wherein each nucleic acid sample is prepared using a different library of multimeric barcoding reagents (or a different library of multimeric barcode molecules), and wherein the barcode regions of each library of multimeric barcoding reagents (or multimeric barcode molecules) comprise a sequence that is different to the barcode regions of the other libraries of multimeric barcoding reagents (or multimeric barcode molecules). Following the separate preparation of each of the samples for sequencing, the barcoded target nucleic acid molecules prepared from the different samples may be pooled and sequenced together. The sequence read generated for each barcoded target nucleic acid molecule may be used to identify the library of multimeric barcoding reagents (or multimeric barcode molecules) that was used in its preparation and thereby to identify the nucleic acid sample from which it was prepared.

In any method of preparing a nucleic acid sample for sequencing, the target nucleic acid molecules may be present at particular concentrations within the nucleic acid sample, for example at concentrations of at least 100 nanomolar, at least 10 nanomolar, at least 1 nanomolar, at least 100 picomolar, at least 10 picomolar, at least 1 picomolar, at least 100 femtomolar, at least 10 femtomolar, or at least 1 femtomolar. The concentrations may be 1 picomolar to 100 nanomolar, 10 picomolar to 10 nanomolar, or 100 picomolar to 1 nanomolar. Preferably, the concentrations are 10 picomolar to 1 nanomolar.

In any method of preparing a nucleic acid sample for sequencing, the multimeric barcoding reagents may be present at particular concentrations within the nucleic acid sample, for example

- 153 -

at concentrations of at least 100 nanomolar, at least 10 nanomolar, at least 1 nanomolar, at least 100 picomolar, at least 10 picomolar, at least 1 picomolar, at least 100 femtomolar, at least 10 femtomolar, or at least 1 femtomolar. The concentrations may be 1 picomolar to 100 nanomolar, 10 picomolar to 10 nanomolar, or 100 picomolar to 1 nanomolar. Preferably, the concentrations are 1 picomolar to 100 picomolar.

In any method of preparing a nucleic acid sample for sequencing, the multimeric barcode molecules may be present at particular concentrations within the nucleic acid sample, for example at concentrations of at least 100 nanomolar, at least 10 nanomolar, at least 1 nanomolar, at least 100 picomolar, at least 10 picomolar, at least 1 picomolar, at least 100 femtomolar, at least 10 femtomolar, or at least 1 femtomolar. The concentrations may be 1 picomolar to 100 nanomolar, 10 picomolar to 10 nanomolar, or 100 picomolar to 1 nanomolar. Preferably, the concentrations are 1 picomolar to 100 picomolar.

In any method of preparing a nucleic acid sample for sequencing, the barcoded oligonucleotides may be present at particular concentrations within the nucleic acid sample, for example at concentrations of at least 100 nanomolar, at least 10 nanomolar, at least 1 nanomolar, at least 100 picomolar, at least 10 picomolar, at least 1 picomolar, at least 100 femtomolar, at least 10 femtomolar, or at least 1 femtomolar. The concentrations may be 1 picomolar to 100 nanomolar, 10 picomolar to 10 nanomolar, or 100 picomolar to 1 nanomolar. Preferably, the concentrations are 100 picomolar to 100 nanomolar.

21. METHODS OF PREPARING A NUCLEIC ACID SAMPLE FOR SEQUENCING USING MULTIMERIC BARCODING REAGENTS

The invention provides a method of preparing a nucleic acid sample for sequencing, wherein the method comprises the steps of: contacting the nucleic acid sample with a multimeric barcoding reagent as defined herein; annealing the target region of the first barcoded oligonucleotide to a first fragment of a target nucleic acid, and annealing the target region of the second barcoded oligonucleotide to a second fragment of the target nucleic acid; and extending the first and second barcoded oligonucleotides to produce first and second different barcoded target nucleic acid molecules, wherein each of the barcoded target nucleic acid molecules comprises at least one nucleotide synthesised from the target nucleic acid as a template.

In any method of preparing a nucleic acid sample for sequencing, either the nucleic acid molecules within the nucleic acid sample, and/or the multimeric barcoding reagents, may be present at particular concentrations within the solution volume, for example at concentrations of at least 100 nanomolar, at least 10 nanomolar, at least 1 nanomolar, at least 100 picomolar, at least 10 picomolar, or at least 1 picomolar. The concentrations may be 1 picomolar to 100 nanomolar,

10 picomolar to 10 nanomolar, or 100 picomolar to 1 nanomolar. Alternative higher or lower concentrations may also be used.

- The method of preparing a nucleic acid sample for sequencing may comprise contacting the nucleic acid sample with a library of multimeric barcoding reagents as defined herein, and wherein: the barcoded oligonucleotides of the first multimeric barcoding reagent anneal to fragments of a first target nucleic acid and first and second different barcoded target nucleic acid molecules are produced, wherein each barcoded target nucleic acid molecule comprises at least one nucleotide synthesised from the first target nucleic acid as a template; and the barcoded oligonucleotides of the second multimeric barcoding reagent anneal to fragments of a second target nucleic acid and first and second different barcoded target nucleic acid molecules are produced, wherein each barcoded target nucleic acid molecule comprises at least one nucleotide synthesised from the second target nucleic acid as a template.
- In the method the barcoded oligonucleotides may be isolated from the nucleic acid sample after annealing to the fragments of the target nucleic acid and before the barcoded target nucleic acid molecules are produced. Optionally, the barcoded oligonucleotides are isolated by capture on a solid support through a streptavidin-biotin interaction.
- Additionally or alternatively, the barcoded target nucleic acid molecules may be isolated from the nucleic acid sample. Optionally, the barcoded target nucleic acid molecules are isolated by capture on a solid support through a streptavidin-biotin interaction.
- The step of extending the barcoded oligonucleotides may be performed while the barcoded oligonucleotides are annealed to the barcode molecules.

Figure 3 shows a method of preparing a nucleic acid sample for sequencing, in which a multimeric barcoding reagent defined herein (for example, as illustrated in Figure 1) is used to label and extend two or more nucleic acid sub-sequences in a nucleic acid sample. In this method, a multimeric barcoding reagent is synthesised which incorporates at least a first (A1, B1, C1, and G1) and a second (A2, B2, C2, and G2) barcoded oligonucleotide, which each comprise both a barcode region (B1 and B2) and a target region (G1 and G2 respectively).

A nucleic acid sample comprising a target nucleic acid is contacted or mixed with the multimeric barcoding reagent, and the target regions (G1 and G2) of two or more barcoded oligonucleotides are allowed to anneal to two or more corresponding sub-sequences within the target nucleic acid (H1 and H2). Following the annealing step, the first and second barcoded oligonucleotides are extended (e.g. with the target regions serving as primers for a polymerase) into the sequence of the target nucleic acid, such that at least one nucleotide of a sub-sequence is incorporated into

the extended 3' end of each of the barcoded oligonucleotides. This method creates barcoded target nucleic acid molecules, wherein two or more sub-sequences from the target nucleic acid are labeled by a barcoded oligonucleotide.

- 5 Alternatively, the method may further comprise the step of dissociating the barcoded oligonucleotides from the barcode molecules before annealing the target regions of the barcoded oligonucleotides to sub-sequences of the target nucleic acid.

10 Figure 4 shows a method of preparing a nucleic acid sample for sequencing, in which a multimeric barcoding reagent described herein (for example, as illustrated in Figure 1) is used to label and extend two or more nucleic acid sub-sequences in a nucleic acid sample, but wherein the barcoded oligonucleotides from the multimeric barcoding reagent are dissociated from the barcode molecules prior to annealing to (and extension of) target nucleic acid sequences. In this method, a multimeric barcoding reagent is synthesised which incorporates at least a first (A1, B1, C1, and G1) and a second (A2, B2, C2, and G2) barcoded oligonucleotide, which each comprise a barcode region (B1 and B2) and a target region (G1 and G2), which is capable of annealing to a sub-sequence within the target nucleic acid (H1 and H2). The method of Figure 4 is described in detail in PCT/GB2017/053820, which is incorporated herein by reference.

- 20 A universal priming sequence may be added to the barcoded target nucleic acid molecules. This sequence may enable the subsequent amplification of at least 5, at least 10, at least 20, at least 25, at least 50, at least 75, at least 100, at least 250, at least 500, at least 10^3 , at least 10^4 , at least 10^5 , at least 10^6 , at least 10^7 , at least 10^8 , or at least 10^9 different barcoded target nucleic acid molecules using one forward primer and one reverse primer.

25 Prior to contacting the nucleic acid sample with a multimeric barcoding reagent, or library of multimeric barcoding reagents, as defined herein, a coupling sequence may be added to the 5' end or 3' end of two or more target nucleic acids of the nucleic acid sample. In this method, the target regions may comprise a sequence that is complementary to the coupling sequence. The target regions may comprise a sequence that is complementary to the coupling sequence. The coupling sequence may comprise a homopolymeric 3' tail (e.g. a poly(A) tail). The coupling sequence may be added by a terminal transferase enzyme. In the method in which the coupling sequence comprises a poly(A) tail, the target regions may comprise a poly(T) sequence. Such coupling sequences may be added following a high-temperature incubation of the nucleic acid sample, to denature the nucleic acids contained therein prior to adding a coupling sequence.

30 Alternatively, a coupling sequence could be added by digestion of a target nucleic acid sample with a restriction enzyme, in which case a coupling sequence may be comprised of one or more nucleotides of a restriction enzyme recognition sequence. In this case, a coupling sequence may be at least partially double-stranded, and may comprise a blunt-ended double-stranded DNA

- 156 -

sequence, or a sequence with a 5' overhang region of 1 or more nucleotides, or a sequence with a 3' overhang region of 1 or more nucleotides. In these cases, target regions in multimeric barcoding reagents may then comprise sequences that are either double-stranded and blunt-ended (and thus able to ligate to blunt-ended restriction digestion products), or the target regions
5 may contain 5' or 3' overhang sequences of 1 or more nucleotides, which make them cohesive (and thus able to anneal with and ligate to) against said restriction digestion products.

The method may comprise preparing two or more independent nucleic acid samples for sequencing, wherein each nucleic acid sample is prepared using a different library of multimeric
10 barcoding reagents (or a different library of multimeric barcode molecules), and wherein the barcode regions of each library of multimeric barcoding reagents (or multimeric barcode molecules) comprise a sequence that is different to the barcode regions of the other libraries of multimeric barcoding reagents (or multimeric barcode molecules). Following the separate preparation of each of the samples for sequencing, the barcoded target nucleic acid molecules
15 prepared from the different samples may be pooled and sequenced together. The sequence read generated for each barcoded target nucleic acid molecule may be used to identify the library of multimeric barcoding reagents (or multimeric barcode molecules) that was used in its preparation and thereby to identify the nucleic acid sample from which it was prepared.

20 The invention provides a method of preparing a nucleic acid sample for sequencing, wherein the method comprises the steps of: (a) contacting the nucleic acid sample with a multimeric barcoding reagent, wherein each barcoded oligonucleotide comprises in the 5' to 3' direction a target region and a barcode region, and first and second target primers; (b) annealing the target region of the first barcoded oligonucleotide to a first sub-sequence of a target nucleic acid and annealing the
25 target region of the second barcoded oligonucleotide to a second sub-sequence of the target nucleic acid; (c) annealing the first target primer to a third sub-sequence of the target nucleic acid, wherein the third sub-sequence is 3' of the first sub-sequence, and annealing the second target primer to a fourth sub-sequence of the target nucleic acid, wherein the fourth sub-sequence is 3' of the second sub-sequence; (d) extending the first target primer using the target nucleic acid as
30 template until it reaches the first sub-sequence to produce a first extended target primer, and extending the second target primer using the target nucleic acid as template until it reaches the second sub-sequence to produce a second extended target primer; and (e) ligating the 3' end of the first extended target primer to the 5' end of the first barcoded oligonucleotide to produce a first barcoded target nucleic acid molecule, and ligating the 3' end of the second extended target
35 primer to the 5' end of the second barcoded oligonucleotide to produce a second barcoded target nucleic acid molecule, wherein the first and second barcoded target nucleic acid molecules are different, and wherein each of the barcoded target nucleic acid molecules comprises at least one nucleotide synthesised from the target nucleic acid as a template.

In the method, steps (b) and (c) may be performed at the same time.

22. METHODS OF PREPARING A NUCLEIC ACID SAMPLE FOR SEQUENCING USING MULTIMERIC BARCODING REAGENTS AND ADAPTER OLIGONUCLEOTIDES

5 The methods provided below may be performed with any of the kits defined herein.

The invention further provides a method of preparing a nucleic acid sample for sequencing, wherein the method comprises the steps of: (a) contacting the nucleic acid sample with a first and second adapter oligonucleotide as defined herein; (b) annealing or ligating the first adapter
10 oligonucleotide to a first fragment of a target nucleic acid, and annealing or ligating the second adapter oligonucleotide to a second fragment of the target nucleic acid; (c) contacting the nucleic acid sample with a multimeric barcoding reagent as defined herein; (d) annealing the adapter region of the first adapter oligonucleotide to the adapter region of the first barcode molecule, and annealing the adapter region of the second adapter oligonucleotide to the adapter region of the
15 second barcode molecule; and (e) ligating the 3' end of the first barcoded oligonucleotide to the 5' end of the first adapter oligonucleotide to produce a first barcoded-adapter oligonucleotide and ligating the 3' end of the second barcoded oligonucleotide to the 5' end of the second adapter oligonucleotide to produce a second barcoded-adapter oligonucleotide.

20 The invention further provides a method of preparing a nucleic acid sample for sequencing, wherein the method comprises the steps of: (a) contacting the nucleic acid sample with a first and second adapter oligonucleotide as defined herein; (b) the first adapter oligonucleotide to a first fragment of a target nucleic acid, and ligating the second adapter oligonucleotide to a second fragment of the target nucleic acid; (c) contacting the nucleic acid sample with a multimeric
25 barcoding reagent as defined herein; (d) annealing the adapter region of the first adapter oligonucleotide to the adapter region of the first barcode molecule, and annealing the adapter region of the second adapter oligonucleotide to the adapter region of the second barcode molecule; and (e) extending the first adapter oligonucleotide using the barcode region of the first barcode molecule as a template to produce a first barcoded target nucleic acid molecule, and
30 extending the second adapter oligonucleotide using the barcode region of the second barcode molecule as a template to produce a second barcoded target nucleic acid molecule, wherein the first barcoded target nucleic acid molecule comprises a sequence complementary to the barcode region of the first barcode molecule and the second barcoded target nucleic acid molecule comprises a sequence complementary to the barcode region of the second barcode molecule.

35 The invention further provides a method of preparing a nucleic acid sample for sequencing, wherein the method comprises the steps of: (a) contacting the nucleic acid sample with a first and second adapter oligonucleotide as defined herein; (b) annealing the target region of the first adapter oligonucleotide to a first fragment of a target nucleic acid, and annealing the target region

- 158 -

of the second adapter oligonucleotide to a second fragment of the target nucleic acid; (c) contacting the nucleic acid sample with a multimeric barcoding reagent as defined herein; (d) annealing the adapter region of the first adapter oligonucleotide to the adapter region of the first barcode molecule, and annealing the adapter region of the second adapter oligonucleotide to the adapter region of the second barcode molecule; and (e) ligating the 3' end of the first barcoded oligonucleotide to the 5' end of the first adapter oligonucleotide to produce a first barcoded-adapter oligonucleotide and ligating the 3' end of the second barcoded oligonucleotide to the 5' end of the second adapter oligonucleotide to produce a second barcoded-adapter oligonucleotide.

10 In the method the first and second barcoded-adapter oligonucleotides may be extended to produce first and second different barcoded target nucleic acid molecules each of which comprises at least one nucleotide synthesised from the target nucleic acid as a template.

15 Alternatively, the first and second adapter oligonucleotides may be extended to produce first and second different target nucleic acid molecules each of which comprises at least one nucleotide synthesised from the target nucleic acid as a template. In this method, step (f) produces a first barcoded target nucleic acid molecule (i.e. the first barcoded oligonucleotide ligated to the extended first adapter oligonucleotide) and a second barcoded target nucleic acid molecule (i.e. the second barcoded oligonucleotide ligated to the extended second adapter oligonucleotide).

20 The step of extending the adapter oligonucleotides may be performed before step (c), before step (d) and/or before step (e), and the first and second adapter oligonucleotides may remain annealed to the first and second barcode molecules until after step (e).

25 The method may be performed using a library of multimeric barcoding reagents as defined herein and an adapter oligonucleotide as defined herein for each of the multimeric barcoding reagents. Preferably, the barcoded-adapter oligonucleotides of the first multimeric barcoding reagent anneal to fragments of a first target nucleic acid and first and second different barcoded target nucleic acid molecules are produced, wherein each barcoded target nucleic acid molecule comprises at least one nucleotide synthesised from the first target nucleic acid as a template; and the barcoded-adapter oligonucleotides of the second multimeric barcoding reagent anneal to fragments of a second target nucleic acid and first and second different barcoded target nucleic acid molecules are produced, wherein each barcoded target nucleic acid molecule comprises at least one nucleotide synthesised from the second target nucleic acid as a template.

35 The method may be performed using a library of multimeric barcoding reagents as defined herein and an adapter oligonucleotide as defined herein for each of the multimeric barcoding reagents. Preferably, the adapter oligonucleotides of the first multimeric barcoding reagent anneal to fragments of a first target nucleic acid and first and second different target nucleic acid molecules are produced, wherein each target nucleic acid molecule comprises at least one nucleotide

synthesised from the first target nucleic acid as a template; and the adapter oligonucleotides of the second multimeric barcoding reagent anneal to fragments of a second target nucleic acid and first and second different target nucleic acid molecules are produced, wherein each target nucleic acid molecule comprises at least one nucleotide synthesised from the second target nucleic acid as a template.

The barcoded-adapter oligonucleotides may be isolated from the nucleic acid sample after annealing to the fragments of the target nucleic acid and before the barcoded target nucleic acid molecules are produced. Optionally, the barcoded-adapter oligonucleotides are isolated by capture on a solid support through a streptavidin-biotin interaction.

The barcoded target nucleic acid molecules may be isolated from the nucleic acid sample. Optionally, the barcoded target nucleic acid molecules are isolated by capture on a solid support through a streptavidin-biotin interaction.

Figure 5 shows a method of preparing a nucleic acid sample for sequencing using a multimeric barcoding reagent. In the method first (C1 and G1) and second (C2 and G2) adapter oligonucleotides are annealed to a target nucleic acid in the nucleic acid sample, and then used in a primer extension reaction. Each adapter oligonucleotide is comprised of an adapter region (C1 and C2) that is complementary to, and thus able to anneal to, the 5' adapter region of a barcode molecule (F1 and F2). Each adapter oligonucleotide is also comprised of a target region (G1 and G2), which may be used to anneal the barcoded oligonucleotides to target nucleic acids, and then may be used as primers for a primer-extension reaction or a polymerase chain reaction. These adapter oligonucleotides may be synthesised to include a 5'-terminal phosphate group.

The adapter oligonucleotides, each of which has been extended to include sequence from the target nucleic acid, are then contacted with a multimeric barcoding reagent which comprises a first (D1, E1, and F1) and second (D2, E2, and F2) barcode molecule, as well as first (A1 and B1) and second (A2 and B2) barcoded oligonucleotides, which each comprise a barcode region (B1 and B2), as well as 5' regions (A1 and A2). The first and second barcode molecules each comprise a barcode region (E1 and E2), an adapter region (F1 and F2), and a 3' region (D1 and D2), and are linked together, in this embodiment by a connecting nucleic acid sequence (S).

After contacting the primer-extended nucleic acid sample with a multimeric barcoding reagent, the 5' adapter regions (C1 and C2) of each adapter oligonucleotides are able to anneal to a 'ligation junction' adjacent to the 3' end of each barcoded oligonucleotide (J1 and J2). The 5' end of the extended adapter oligonucleotides are then ligated to the 3' end of the barcoded oligonucleotides within the multimeric barcoding reagent, creating a ligated base pair (K1 and K2) where the

- 160 -

ligation junction was formerly located. The solution may subsequently be processed further or amplified, and used in a sequencing reaction.

5 This method, like the methods illustrated in Figures 3 and 4, creates barcoded target nucleic acid molecules, wherein two or more fragments from the nucleic acid sample are labeled by a barcoded oligonucleotide. In this method a multimeric barcoding reagent does not need to be present for the step of annealing target regions to fragments of the target nucleic acid, or the step of extending the annealed target regions using a polymerase. This feature may hold advantages in certain applications, for example wherein a large number of target sequences are of interest, and the target regions are able to hybridise more rapidly to target nucleic acids when they are not
10 constrained molecularly by a multimeric barcoding reagent.

23. METHODS OF PREPARING A NUCLEIC ACID SAMPLE FOR SEQUENCING USING MULTIMERIC BARCODING REAGENTS, ADAPTER OLIGONUCLEOTIDES AND EXTENSION PRIMERS
15

Methods of preparing a nucleic acid sample for sequencing using multimeric barcoding reagents, adapter oligonucleotides and extension primers are described in PCT/GB2017/053820, which is incorporated herein by reference.

24. METHODS OF PREPARING A NUCLEIC ACID SAMPLE FOR SEQUENCING USING MULTIMERIC BARCODING REAGENTS, ADAPTER OLIGONUCLEOTIDES AND TARGET PRIMERS
20

Methods of preparing a nucleic acid sample for sequencing using multimeric barcoding reagents, adapter oligonucleotides and target primers are described in PCT/GB2017/053820, which is
25 incorporated herein by reference. Figure 6 illustrates one way in which this method may be performed. In this method, the target nucleic acid is genomic DNA. It will be appreciated that the target nucleic acid may be another type of nucleic acid e.g. an RNA molecule such as an mRNA molecule.

25. METHODS OF PREPARING A NUCLEIC ACID SAMPLE FOR SEQUENCING USING MULTIMERIC BARCODING REAGENTS AND TARGET PRIMERS
30

Methods of preparing a nucleic acid sample for sequencing using multimeric barcoding reagents and target primers are described in PCT/GB2017/053820, which is incorporated herein by
35 reference.

26. METHODS OF SYNTHESISING A MULTIMERIC BARCODING REAGENT

The invention further provides a method of synthesising a multimeric barcoding reagent for labelling a target nucleic acid comprising: (a) contacting first and second barcode molecules with first and second extension primers, wherein each of the barcode molecules comprises a single-stranded nucleic acid comprising in the 5' to 3' direction an adapter region, a barcode region and a priming region; (b) annealing the first extension primer to the priming region of the first barcode molecule and annealing the second extension primer to the priming region of the second barcode molecule; and (c) synthesising a first barcoded extension product by extending the first extension primer and synthesising a second barcoded extension product by extending the second extension primer, wherein the first barcoded extension product comprises a sequence complementary to the barcode region of the first barcode molecule and the second barcoded extension product comprises a sequence complementary to the barcode region of the second barcode molecule, and wherein the first barcoded extension product does not comprise a sequence complementary to the adapter region of the first barcode molecule and the second barcoded extension product does not comprise a sequence complementary to the adapter region of the second barcode molecule; and wherein the first and second barcode molecules are linked together.

The method may further comprise the following steps before the step of synthesising the first and second barcoded extension products: (a) contacting first and second barcode molecules with first and second blocking primers; and (b) annealing the first blocking primer to the adapter region of the first barcode molecule and annealing the second blocking primer to the adapter region of the second barcode molecule; and wherein the method further comprises the step of dissociating the blocking primers from the barcode molecules after the step of synthesising the barcoded extension products.

In the method, the extension step, or a second extension step performed after the synthesis of an extension product, may be performed, in which one or more of the four canonical deoxyribonucleotides is excluded from the extension reaction, such that the second extension step terminates at a position before the adapter region sequence, wherein the position comprises a nucleotide complementary to the excluded deoxyribonucleotide. This extension step may be performed with a polymerase lacking 3' to 5' exonuclease activity.

The barcode molecules may be provided by a single-stranded multimeric barcode molecule as defined herein.

The barcode molecules may be synthesised by any of the methods defined herein. The barcode regions may uniquely identify each of the barcode molecules. The barcode molecules may be linked on a nucleic acid molecule. The barcode molecules may be linked together in a ligation

- 162 -

reaction. The barcode molecules may be linked together by a further step comprising attaching the barcode molecules to a solid support.

5 The first and second barcode molecules may be assembled as a double-stranded multimeric barcode molecule by any of the methods defined herein prior to step (a) defined above (i.e. contacting first and second barcode molecules with first and second extension primers). The double-stranded multimeric barcode molecule may be dissociated to produce single-stranded multimeric barcode molecules for use in step (a) defined above (i.e. contacting first and second barcode molecules with first and second extension primers).

10

The method may further comprise the steps of: (a) annealing an adapter region of a first adapter oligonucleotide to the adapter region of the first barcode molecule and annealing an adapter region of a second adapter oligonucleotide to the adapter region of the second barcode molecule, wherein the first adapter oligonucleotide further comprises a target region capable of annealing to
15 a first sub-sequence of the target nucleic acid and the second adapter oligonucleotide further comprises a target region capable of annealing to a second sub-sequence of the target nucleic acid; and (b) ligating the 3' end of the first barcoded extension product to the 5' end of the first adapter oligonucleotide to produce a first barcoded oligonucleotide and ligating the 3' end of the second barcoded extension product to the 5' end of the second adapter oligonucleotide to
20 produce a second barcoded oligonucleotide. Optionally, the annealing step (a) may be performed before the step of synthesising the first and second barcoded extension products and wherein the step of synthesising the first and second barcoded extension products is conducted in the presence of a ligase enzyme that performs the ligation step (b). The ligase may be a thermostable ligase. The extension and ligation reaction may proceed at over 37 degrees
25 Celsius, over 45 degrees Celsius, or over 50 degrees Celsius.

The target regions may comprise different sequences. Each target region may comprise a sequence capable of annealing to only a single sub-sequence of a target nucleic acid within a sample of nucleic acids. Each target region may comprise one or more random, or one or more
30 degenerate, sequences to enable the target region to anneal to more than one sub-sequence of a target nucleic acid. Each target region may comprise at least 5, at least 10, at least 15, at least 20, at least 25, at least 50 or at least 100 nucleotides. Preferably, each target region comprises at least 5 nucleotides. Each target region may comprise 5 to 100 nucleotides, 5 to 10
35 nucleotides, 10 to 20 nucleotides, 20 to 30 nucleotides, 30 to 50 nucleotides, 50 to 100 nucleotides, 10 to 90 nucleotides, 20 to 80 nucleotides, 30 to 70 nucleotides or 50 to 60 nucleotides. Preferably, each target region comprises 30 to 70 nucleotides. Preferably each target region comprises deoxyribonucleotides, optionally all of the nucleotides in a target region are deoxyribonucleotides. One or more of the deoxyribonucleotides may be a modified deoxyribonucleotide (e.g. a deoxyribonucleotide modified with a biotin moiety or a deoxyuracil

nucleotide). Each target region may comprise one or more universal bases (e.g. inosine), one or modified nucleotides and/or one or more nucleotide analogues.

- 5 The adapter region of each adapter oligonucleotide may comprise a constant region. Optionally, all adapter regions of adapter oligonucleotides that anneal to a single multimeric barcoding reagent are substantially identical. The adapter region may comprise at least 4, at least 5, at least 6, at least 8, at least 10, at least 15, at least 20, at least 25, at least 50, at least 100, or at least 250 nucleotides. Preferably, the adapter region comprises at least 4 nucleotides. Preferably each adapter region comprises deoxyribonucleotides, optionally all of the nucleotides
- 10 in an adapter region are deoxyribonucleotides. One or more of the deoxyribonucleotides may be a modified deoxyribonucleotide (e.g. a deoxyribonucleotide modified with a biotin moiety or a deoxyuracil nucleotide). Each adapter region may comprise one or more universal bases (e.g. inosine), one or modified nucleotides and/or one or more nucleotide analogues.
- 15 For any of the methods involving adapter oligonucleotides, the 3' end of the adapter oligonucleotide may include a reversible terminator moiety or a reversible terminator nucleotide (for example, a 3'-O-blocked nucleotide), for example at the 3' terminal nucleotide of the target region. When used in an extension and/or extension and ligation reaction, the 3' ends of these adapter oligonucleotides may be prevented from priming any extension events. This may
- 20 minimize mis-priming or other spurious extension events during the production of barcoded oligonucleotides. Prior to using the assembled multimeric barcoding reagents, the terminator moiety of the reversible terminator may be removed by chemical or other means, thus allowing the target region to be extended along a target nucleic acid template to which it is annealed.
- 25 Similarly, for any of the methods involving adapter oligonucleotides, one or more blocking oligonucleotides complementary to one or more sequences within the target region(s) may be employed during extension and/or extension and ligation reactions. The blocking oligonucleotides may comprise a terminator and/or other moiety on their 3' and/or 5' ends such that they are not able to be extended by polymerases. The blocking oligonucleotides may be designed such that
- 30 they anneal to sequences fully or partially complementary to one or more target regions, and are annealed to said target regions prior to an extension and/or extension and ligation reaction. The use of blocking primers may prevent target regions from annealing to, and potentially mis-priming along, sequences within the solution for which such annealing is not desired (for example, sequence features within barcode molecules themselves). The blocking oligonucleotides may be
- 35 designed to achieve particular annealing and/or melting temperatures. Prior to using the assembled multimeric barcoding reagents, the blocking oligonucleotide(s) may then be removed by, for example, heat-denaturation and then size-selective cleanup, or other means. The removal of the blocking oligonucleotide(s) may allow the target region to be extended along a target nucleic acid template to which it is annealed.

The method may comprise synthesising a multimeric barcoding reagent comprising at least 5, at least 10, at least 20, at least 25, at least 50, at least 75 or at least 100 barcode molecules, and wherein: (a) each barcode molecule is as defined herein; and (b) a barcoded extension product is synthesised from each barcode molecule according to any method defined herein; and, optionally, (c) an adapter oligonucleotide is ligated to each of the barcoded extension products to produce barcoded oligonucleotides according to any of the methods defined herein.

The invention further provides a method of synthesising a library of multimeric barcoding reagents, wherein the method comprises repeating the steps of any of the methods defined herein to synthesise two or more multimeric barcoding reagents. Optionally, the method comprises synthesising a library of at least 5, at least 10, at least 20, at least 25, at least 50, at least 75, at least 100, at least 250, at least 500, at least 10^3 , at least 10^4 , at least 10^5 , at least 10^6 , at least 10^7 , at least 10^8 , at least 10^9 or at least 10^{10} multimeric barcoding reagents as defined herein. Preferably, the library comprises at least 5 multimeric barcoding reagents as defined herein. Preferably, the barcode regions of each of the multimeric barcoding reagents may be different to the barcode regions of the other multimeric barcoding reagents.

Figure 8 illustrates a method of synthesizing a multimeric barcoding reagent for labeling a target nucleic acid. In this method, first (D1, E1, and F1) and second (D2, E2, and F2) barcode molecules, which each include a nucleic acid sequence comprising a barcode region (E1 and E2), and which are linked by a connecting nucleic acid sequence (S), are denatured into single-stranded form. To these single-stranded barcode molecules, a first and second extension primer (A1 and A2) is annealed to the 3' region of the first and second barcode molecules (D1 and D2), and a first and second blocking primer (R1 and R2) is annealed to the 5' adapter region (F1 and F2) of the first and second barcode molecules. These blocking primers (R1 and R2) may be modified on the 3' end such that they cannot serve as a priming site for a polymerase.

A polymerase is then used to perform a primer extension reaction, in which the extension primers are extended to make a copy (B1 and B2) of the barcode region of the barcode molecules (E1 and E2). This primer extension reaction is performed such that the extension product terminates immediately adjacent to the blocking primer sequence, for example through use of a polymerase which lacks strand displacement or 5'-3' exonuclease activity. The blocking primers (R1 and R2) are then removed, for example through high-temperature denaturation.

This method thus creates a multimeric barcoding reagent containing a first and second ligation junction (J1 and J2) adjacent to a single-stranded adapter region (F1 and F2). This multimeric barcoding reagent may be used in the method illustrated in Figure 5.

- 165 -

The method may further comprise the step of ligating the 3' end of the first and second barcoded oligonucleotides created by the primer-extension step (the 3' end of B1 and B2) to first (C1 and G1) and second (C2 and G2) adapter oligonucleotides, wherein each adapter oligonucleotide comprises an adapter region (C1 and C2) which is complementary to, and thus able to anneal to, the adapter region of a barcode molecule (F1 and F2). The adapter oligonucleotides may be synthesised to include a 5'-terminal phosphate group.

Each adapter oligonucleotide may also comprise a target region (G1 and G2), which may be used to anneal the barcoded oligonucleotides to target nucleic acids, and may separately or subsequently be used as primers for a primer-extension reaction or a polymerase chain reaction. The step of ligating the first and second barcoded oligonucleotides to the adapter oligonucleotides produces a multimeric barcoding reagent as illustrated in Figure 1 that may be used in the methods illustrated in Figure 3 and/or Figure 4.

Figure 9 shows a method of synthesizing multimeric barcoding reagents (as illustrated in Figure 1) for labeling a target nucleic acid. In this method, first (D1, E1, and F1) and second (D2, E2, and F2) barcode molecules, which each include a nucleic acid sequence comprising a barcode region (E1 and E2), and which are linked by a connecting nucleic acid sequence (S), are denatured into single-stranded form. To these single-stranded barcode molecules, a first and second extension primer (A1 and A2) is annealed to the 3' region of the first and second barcode molecules (D1 and D2), and the adapter regions (C1 and C2) of first (C1 and G1) and second (C2 and G2) adapter oligonucleotides are annealed to the 5' adapter regions (F1 and F2) of the first and second barcode molecules. These adapter oligonucleotides may be synthesised to include a 5'-terminal phosphate group.

A polymerase is then used to perform a primer extension reaction, in which the extension primers are extended to make a copy (B1 and B2) of the barcode region of the barcode molecules (E1 and E2). This primer extension reaction is performed such that the extension product terminates immediately adjacent to the adapter region (C1 and C2) sequence, for example through use of a polymerase which lacks strand displacement or 5'-3' exonuclease activity.

A ligase enzyme is then used to ligate the 5' end of the adapter oligonucleotides to the adjacent 3' end of the corresponding extension product. In an alternative embodiment, a ligase enzyme may be included with the polymerase enzyme in one reaction which simultaneously effects both primer-extension and ligation of the resulting product to the adapter oligonucleotide. Through this method, the resulting barcoded oligonucleotides may subsequently be used as primers for a primer-extension reaction or a polymerase chain reaction, for example as in the method shown in Figure 3 and/or Figure 4.

27. METHODS OF SEQUENCING AND/OR PROCESSING SEQUENCING DATA

The invention provides a method of sequencing a target nucleic acid of a circulating microparticle, wherein the circulating microparticle contains at least two fragments of a target nucleic acid, and
5 wherein the method comprises: (a) preparing a sample for sequencing comprising linking at least two of the at least two fragments of the target nucleic acid to produce a set of at least two linked fragments of the target nucleic acid; and (b) sequencing each of the linked fragments in the set to produce at least two (informatically) linked sequence reads.

10 The invention provides a method of sequencing genomic DNA of a circulating microparticle, wherein the circulating microparticle contains at least two fragments of genomic DNA, and wherein the method comprises: (a) preparing a sample for sequencing comprising linking at least two of the at least two fragments of genomic DNA to produce a set of at least two linked fragments of genomic DNA; and (b) sequencing each of the linked fragments in the set to produce
15 at least two (informatically) linked sequence reads.

The invention provides a method of sequencing a target nucleic acid of a circulating microparticle comprising: (a) linking at least two fragments of the target nucleic acid from a (single) circulating microparticle to produce a set of at least two linked fragments of the target nucleic acid; and (b)
20 sequencing each of the linked fragments in the set to produce at least two (informatically) linked sequence reads.

The invention provides a method of sequencing circulating microparticle genomic DNA comprising: (a) linking at least two fragments of genomic DNA from a (single) circulating
25 microparticle to produce a set of at least two linked fragments of circulating microparticle genomic DNA; and (b) sequencing each of the linked fragments in the set to produce at least two (informatically) linked sequence reads.

The invention further provides a method of sequencing a sample, wherein the sample has been
30 prepared by any one of the methods of preparing a nucleic acid sample for sequencing as defined herein. The method of sequencing the sample comprises the steps of: isolating the barcoded target nucleic acid molecules, and producing a sequence read from each barcoded target nucleic acid molecule that comprises the barcode region, the target region and at least one additional nucleotide from the target nucleic acid. Each sequence read may comprise at least 5, at least 10,
35 at least 25, at least 50, at least 100, at least 250, at least 500, at least 1000, at least 2000, at least 5000, or at least 10,000 nucleotides from the target nucleic acid. Preferably, each sequence read comprises at least 5 nucleotides from the target nucleic acid.

- 167 -

The methods may produce a sequence read from one or more barcoded target nucleic acid molecule produced from at least 10, at least 100, or at least 10^3 , at least 10^4 , at least 10^5 , at least 10^6 , at least 10^7 , at least 10^8 or at least 10^9 different target nucleic acids.

- 5 Sequencing may be performed by any method known in the art. For example, by chain-termination or Sanger sequencing. Preferably, sequencing is performed by a next-generation sequencing method such as sequencing by synthesis, sequencing by synthesis using reversible terminators (e.g. Illumina sequencing), pyrosequencing (e.g. 454 sequencing), sequencing by ligation (e.g. SOLiD sequencing), single-molecule sequencing (e.g. Single Molecule, Real-Time
10 (SMRT) sequencing, Pacific Biosciences), or by nanopore sequencing (e.g. on the Minion or Promethion platforms, Oxford Nanopore Technologies).

- The invention further provides a method for processing sequencing data obtained by any of the methods defined herein. The method for processing sequence data comprises the steps of: (a)
15 identifying for each sequence read the sequence of the barcode region and the sequence from the target nucleic acid; and (b) using the information from step (a) to determine a group of sequences from the target nucleic acid that were labelled with barcode regions from the same multimeric barcoding reagent.

- 20 The method may further comprise the step of determining a sequence of a target nucleic acid by analysing the group of sequences to identify contiguous sequences, wherein the sequence of the target nucleic acid comprises nucleotides from at least two sequence reads.

- The invention further provides an algorithm for processing (or analysing) sequencing data
25 obtained by any of the methods defined herein. The algorithm may be configured to perform any of the methods for processing sequencing data defined herein. The algorithm may be used to detect the sequence of a barcode region within each sequence read, and also to detect the sequence within a sequence read that is derived from a target nucleic acid, and to separate these into two associated data sets.

- 30 The invention further provides a method of generating a synthetic long read from a target nucleic acid comprising the steps of: (a) preparing a nucleic acid sample for sequencing according to any of the methods defined herein; (b) sequencing the sample, optionally wherein the sample is sequenced by any of the methods defined herein; and (c) processing the sequence data obtained
35 by step (b), optionally wherein the sequence data is processed according to any of the methods defined herein; wherein step (c) generates a synthetic long read comprising at least one nucleotide from each of the at least two sequence reads.

- 168 -

The method may enable the phasing of a target sequence of a target nucleic acid molecule i.e. it may enable the determination of which copy of a chromosome (i.e. paternal or maternal) the sequence is located. The target sequence may comprise a specific target mutation, translocation, deletion or amplification and the method may be used to assign the mutation, translocation, deletion or amplification to a specific chromosome. The phasing two or more target sequences may also enable the detection of aneuploidy.

The synthetic long read may comprise at least 50, at least 100, at least 250, at least 500, at least 750, at least 1000, at least 2000, at least 10^4 , at least 10^5 , at least 10^6 , at least 10^7 or at least 10^8 nucleotides. Preferably, the synthetic long read comprises at least 50 nucleotides.

The invention further provides a method of sequencing two or more co-localised target nucleic acids comprising the steps of: (a) preparing a nucleic acid sample for sequencing according to any of the methods defined herein; (b) sequencing the sample, optionally wherein the sample is sequenced by any of the methods defined herein; and (c) processing the sequence data obtained by step (b), optionally wherein the sequence data is processed according to any of the methods defined herein; wherein step (c) identifies at least two sequence reads comprising nucleotides from at least two target nucleic acids co-localised in the sample.

Any method of analysing barcoded or linked nucleic acid molecules by sequencing may comprise a redundant sequencing reaction, wherein target nucleic acid molecules (e.g. that have been barcoded in a barcoding reaction) are sequenced two or more times within a sequencing reaction. Optionally, each such molecule prepared from a sample may be sequenced, on average, at least twice, at least 3 times, at least 5 times, at least 10 times, at least 20 times, at least 50 times, or at least 100 times.

In any method of analysing barcoded nucleic acid molecules by sequencing, an error correction process may be employed. This process may comprise the steps of: (i) determining two or more sequence reads from a sequencing dataset comprising the same barcode sequence, and (ii) aligning the sequences from said two or more sequence reads to each other. Optionally, this error correction process may further comprise a step of (iii) determining a majority and/or most common and/or most likely nucleotide at each position within the sequence read and/or at each position within the sequence of the target nucleic acid molecule. This step may optionally comprise establishing a consensus sequence of each target nucleic acid sequence by any process of error correction, error removal, error detection, error counting, or statistical error removal. This step may further comprise the step of collapsing multiple sequence reads comprising the same barcode sequence into a representation comprising a single, error-corrected read. Optionally, any step of determining two or more sequence reads from a sequencing dataset comprising the same barcode sequence, may comprise determining sequence reads comprising

barcode sequences with at least a certain extent of identical nucleotides and/or sequence similarity, for example at least 70%, at least 80%, at least 90%, or at least 95% sequence similarity (for example, allowing for mismatches and/or insertions or deletions at any point between to barcode sequences).

5

In any method of using analysing barcoded nucleic acid molecules by sequencing, an alternative error correction process may be employed, comprising the steps of: (i) determining two or more sequence reads from a sequencing dataset that comprise the same target nucleic acid sequence, wherein said two or more sequence reads further comprise two or more different barcode sequences, wherein the barcode sequences are from the same multimeric barcode molecule and/or multimeric barcoding reagent, and (ii) aligning the sequences from said two or more sequence reads to each other. Optionally, this error correction process may further comprise a step of (iii) determining a majority and/or most common and/or most likely nucleotide at each position within the sequence of the target nucleic acid molecule. This step may optionally

10

comprise establishing a consensus sequence of the target nucleic acid molecule by any process of error correction, error removal, error detection, error counting, or statistical error removal. This step may further comprise the step of collapsing multiple sequence reads comprising the same target nucleic acid molecule into a representation comprising a single, error-corrected read. The target nucleic acid molecule may comprise, for example, a genomic DNA sequence. Optionally,

15

any step of comparing two barcode sequences, and/or comparing a sequenced barcode sequence and a reference barcode sequence, may comprise determining sequences comprising at least a certain extent of identical nucleotides and/or sequence similarity, for example at least 70%, at least 80%, at least 90%, or at least 95% sequence similarity (for example, allowing for mismatches and/or insertions or deletions at any point between to barcode sequences).

20

25

28. METHODS FOR DETERMINING AND ANALYSING SETS OF LINKED SEQUENCE READS (I.E. SETS LINKED SIGNALS) FROM MICROPARTICLES

The invention provides a method of determining a set of linked sequence reads (i.e. a set of linked signals) of fragments of a target nucleic acid (e.g. genomic DNA) from a single

30

microparticle, wherein the method comprises: (a) analyzing a sample according to any of the methods described herein; and (b) determining a set of two or more linked sequence reads.

The set of two or more linked sequence reads may be determined by identifying sequence reads comprising the same barcode sequence.

35

The set of two or more linked sequence reads may be determined by identifying sequence reads comprising different barcode sequences from the same set of barcode sequences.

- 170 -

The set of two or more linked sequence reads may be determined by identifying sequence reads comprising barcode sequences of barcode regions from the same multimeric barcoding reagent.

Two or more linked sequence reads may be determining by identifying sequence reads
5 comprised within two or more non-overlapping segments of the same sequenced molecule

The set of two or more linked sequence reads may be determined by identifying their spatial proximity within the sequencing instrument used for their sequencing. Optionally this spatial proximity is determined through the use of a cutoff or threshold value, or determined through a
10 non-random or above-average proximity. Optionally, this spatial proximity is represented as a quantitative, semi-quantitative, or categorical value corresponding to different degrees of spatial proximity within the sequencing instrument.

The method may comprise determining at least 3, at least 5, at least 10, at least 50, at least 100,
15 at least 1000, at least 10,000, at least 100,000, at least 1,000,000 sets of linked sequence reads (i.e. sets of linked signals).

The invention provides a method of determining the total number of sets of linked sequence reads (i.e. sets of linked signals) within a sequence dataset comprising: (a) analyzing a sample
20 according to any of the methods described herein; and (b) determining the number of sets of linked sequence reads.

The number of sets of linked sequence reads (i.e. sets of linked signals) may determined by counting the number of sequence reads comprising different barcode sequences.
25

The number of sets of linked sequence reads (i.e. sets of linked signals) may be determined by counting the sets of barcode sequences that have a barcode sequence in a sequence read.

The number of sets of linked sequence reads (i.e. sets of linked signals) may be determined by counting the number of multimeric barcoding reagents that have a barcode region that barcode sequence of which is in a sequence read.
30

Optionally, only barcode sequences represented at least 2 times, at least 3 times, at least 5 times, at least 10 times, at least 20 times, at least 50 times, or at least 100 times within the sequence
35 dataset are included in these counting processes. Optionally, sequence reads and/or barcode sequences are processed through an error-correction process prior to said counting processes. Optionally, technical duplicate reads represented more than once in the overall sequence dataset are collapsed into single de-duplicated reads in a de-duplication process prior to said counting processes.

The method may comprise counting or estimating a total number of sets of linked sequence reads (i.e. sets of linked signals), wherein two or more nucleic acid sequences comprising fragments of a target nucleic acid (e.g. genomic DNA) from a microparticle are appended to each other within sequences comprising said sequence dataset, and the number of sequence reads from said sequence dataset comprising at least two different segments of the target nucleic acid are counted, thus determining the number of sets of linked sequence reads within the sequence dataset. Optionally, the total number of sequenced molecules within said sequence dataset are counted, thus determining the number of sets of linked sequence reads within the sequence dataset. Optionally, only sequenced molecules comprising at least 3 different segments of the target nucleic acid, comprising at least 5 different segments of the target nucleic acid, comprising at least 10 different segments of the target nucleic acid, or comprising at least 50 different segments of the target nucleic acid are counted.

The method may comprise counting or estimating a total number of sets of linked sequence reads (i.e. sets of linked signals), wherein sets of sequences are linked informatically by spatial proximity within the sequencing instrument, and wherein the total number of sequenced molecules within said sequence dataset are counted, thus determining the number of sets of linked sequence reads within the sequence dataset. Optionally, the total number of sequenced molecules within said sequence dataset are counted and then divided by an invariant normalization factor, thus determining the number of sets of linked sequence reads within the sequence dataset.

The invention provides a method of determining a parameter value from a set of linked sequence reads (i.e. a set of linked signals), wherein the method comprises: (a) determining a set of linked sequence reads according to any of the methods described herein; and (b) mapping (at least a portion of) each sequence of the set of linked sequence reads to one or more reference nucleotide sequences; and (c) determining the parameter value by counting or identifying the presence of one or more reference nucleotide sequences within the set of linked sequence reads.

Optionally, this reference sequence may comprise an entire genome, an entire chromosome, a part of a chromosome, a gene, a part of a gene, any other part or parts of a genome, or any other synthetic or actual sequence. The reference sequence may comprise a transcript, a part of a transcript, a transcript isoform, or a part of a transcript isoform; the reference sequence may comprise a splice junction of a transcript. The reference sequence may be from the human genome. The reference sequence may be from one or more different reference human genome sequences, such as different reference sequences from a library of two or more different reference human genome sequences, or from a library of two or more different haplotype-phased

reference human genome sequences (for example, different genome sequences from the International HapMap Project, and/or the 100 Genomes Project).

Further options for reference sequences are described in PCT/GB2017/053820, which is
5 incorporated herein by reference.

Optionally, one or more reference sequence(s) may comprise a sequence that is present exclusively within, or found preferentially within, or found at high and/or above-average levels within particular tissues (i.e. particular cell types) and/or within particular specific diseased tissue.
10 Optionally, one or more reference sequence(s) may be present exclusively within, or found preferentially within, or found at high and/or above-average levels within, non-maternal and/or paternal tissues. Optionally, one or more reference sequence(s) may be present exclusively within, or found preferentially within, or found at high and/or above-average levels within, maternal tissues. Optionally, one or more reference sequence(s) may be present exclusively within, or found
15 preferentially within, or found at high and/or above-average levels within, one or more particular tissue types (for example, a lung tissue, or a pancreas tissue, or a lymphocyte). Optionally, one or more reference sequence(s) may be present exclusively within, or found preferentially within, or found at high and/or above-average levels within, a particular type of diseased tissue (such as a cancer tissue, such as a lung cancer tissue or a colorectal cancer tissue, or from a non-cancer
20 diseased tissue such as an infarcted myocardial tissue, or a diseased cerebrovascular tissue, or a placental tissue undergoing eclampsia or pre-eclampsia). Optionally, one or more reference sequence(s) may be present exclusively within, or found preferentially within, or found at high and/or above-average levels within, a particular type of tissue (such as a lung tissue, or a pancreas tissue, or a lymphocyte). Optionally, one or more reference sequence(s) may be
25 present exclusively within, or found preferentially within, or found at high and/or above-average levels within, a particular type of healthy tissue (such as a healthy lung tissue, or a healthy pancreas tissue, or a healthy lymphocyte).

Optionally, any one or more reference sequence(s) that comprise a sequence that is present
30 exclusively within, or found preferentially within, or found at high and/or above-average levels within particular tissues (i.e. particular cell types) and/or within particular specific diseased tissue, may be established by an empirical measurement and/or evaluation process. Further options are provided in PCT/GB2017/053820, which is incorporated herein by reference.

35 Optionally, one or more reference sequence(s) may comprise a sequence comprised within a barcoded affinity probe, wherein the target molecule of said barcoded affinity probe (e.g. a protein for which said barcoded affinity probe has affinity) is present exclusively within, or found preferentially within, or found at high and/or above-average levels within particular tissue(s) (i.e. particular cell types) and/or within particular specific diseased tissue(s). Optionally, one or more

- 173 -

reference sequence(s) may comprise a sequence comprised within a barcoded affinity probe, wherein the target molecule of said barcoded affinity probe (e.g. a protein for which said barcoded affinity probe has affinity) is absent within, or preferentially absent within, or found at low and/or below-average levels within particular tissue(s) (i.e. particular cell types) and/or within particular specific diseased tissue(s).

The reference nucleotide sequence may comprise a sequence corresponding to a chromosome or a portion of a chromosome. Optionally this sequence is at least 1 nucleotide in length, at least 10 nucleotides in length, at least 100 nucleotides in length, at least 1000 nucleotides in length, at least 10,000 nucleotides in length, at least 100,000 nucleotides in length, at least 1,000,000 nucleotides in length, at least 10,000,000 nucleotides in length, or at least 100,000,000 nucleotides in length.

The reference nucleotide sequence may comprise two or more sequences corresponding to two or more chromosomes, or to sequences corresponding to two or more portions of one or more chromosomes. Optionally these sequences are each at least 1 nucleotide in length, at least 10 nucleotides in length, at least 100 nucleotides in length, at least 1000 nucleotides in length, at least 10,000 nucleotides in length, at least 100,000 nucleotides in length, at least 1,000,000 nucleotides in length, at least 10,000,000 nucleotides in length, or at least 100,000,000 nucleotides in length. Optionally, this reference sequence may comprise an entire genome sequence.

The reference nucleotide sequence may comprise one or more sliding windows, wherein each window comprises a span of a genomic region of a finite length, and wherein two or more windows are offset a certain finite number of nucleotides along said genomic region. Optionally, these sliding windows may be partially overlapping, immediately adjacent to each other, or separated by a span of a certain number of nucleotides.

The reference nucleotide sequence may comprise a repeat sequence. Optionally this repeat sequence comprises a dinucleotide repeat, a trinucleotide repeat, a tetranucleotide repeat, or a pentanucleotide repeat. Optionally, the reference nucleotide sequence comprises a series of two or more immediately adjacent copies of the same repeat unit, such as 2 immediately adjacent copies, 5 immediately adjacent copies, 8 immediately adjacent copies, 10 immediately adjacent copies, 15 immediately adjacent copies, 20 immediately adjacent copies, 30 immediately adjacent copies, 40 immediately adjacent copies, 50 immediately adjacent copies, or 100 immediately adjacent copies.

Optionally, any one or more reference sequences may be employed to analyse sequences determined by any method described herein. Any one or more reference sequences may be

- employed to analyse sequences of fragments of genomic DNA. Any one or more reference sequences may be employed to analyse sequences of RNA. Any one or more reference sequences may be employed to analyse sequences of fragments of genomic DNA wherein a measurement of a modified nucleotide or nucleobase is performed upon one or more said
- 5 fragment(s) of genomic DNA (as one such example, any one or more reference sequences may be employed to analyse sequences of fragments of genomic DNA that have been enriched by an enrichment process for a modified nucleotide such as 5-methylcytosine, or 5-hydroxy-methylcytosine; as another such example, any one or more reference sequences may be employed to analyse sequences of fragments of genomic DNA that have had at least one
- 10 nucleotide contained therein converted by a molecular-conversion process, such as a bisulfite conversion process, or an oxidative bisulfite conversion process, wherein said conversion process is employed to detect one or more modified nucleotides such as 5-methylcytosine, or 5-hydroxy-methylcytosine).
- 15 Optionally any one or more reference sequence(s) may comprise one or more differentially methylated regions (DMRs) (e.g a DMR at least 20, at least 30, at least 50, at least 80, at least 100, at least 120, at least 150, at least 200, at least 300, or at least 500 nucleotides in length), for example DMRs differentially methylated between any two cell types and/or tissue types, and/or DMRs preferentially methylated (or preferentially demethylated) in one or more specific tissue
- 20 types and/or cell types and/or diseased tissue types.

- Optionally, any one or more reference sequences may be employed to analyse sequences of fragments of genomic DNA, wherein the 5'-most and/or 3'-most nucleotides of any such fragments of genomic DNA (and/or nucleotides near to the 5'-most and/or 3'-most nucleotides,
- 25 such as nucleotides within the nearest 2, 3, 4, or 5 nucleotides of the 5'-most and/or 3'-most nucleotides) are mapped to said reference sequences. Further options are provided in PCT/GB2017/053820, which is incorporated herein by reference. Optionally, reference sequences and/or lists thereof may comprise sequences of chromatin accessibility and/or openness of chromatin (for example, as measure by an ATAC-seq assay and/or a DNase accessibility assay)
- 30 (for example, in any one or more specific tissues and/or diseased tissues and/or healthy tissues), optionally wherein a weighting value corresponding to each such reference sequence is generated corresponding to the extent and/or likelihood of chromatin accessibility and/or openness of chromatin for each such reference sequence (e.g. within any one or more specific tissues and/or diseased tissues and/or healthy tissues).

35

The parameter value may be a quantitative or semi-quantitative value and is determined by counting the number of sequence reads within the set of sequences that are determined to comprise a sequence originating from the said reference nucleotide sequence or sequences.

Further options are provided in PCT/GB2017/053820, which is incorporated herein by reference.

The parameter value may be a binary value and may be determined by detecting whether at least one sequence read within the set of sequence reads comprises a sequence originating from the said reference nucleotide sequence or sequences. Further options are provided in PCT/GB2017/053820, which is incorporated herein by reference.

Optionally, each reference sequence within a list and/or group of two or more reference sequences may be associated with a weighting and/or association value. Optionally, this weighting and/or association value may correspond to a likelihood or probability that a given sequence is non-maternal or paternal, or correspond to a likelihood or probability that a given sequence is maternal. Optionally, this weighting and/or association value may correspond to a likelihood or probability that a given sequence is from a particular tissue type (for example, a lung tissue, or a pancreas tissue, or a lymphocyte). Optionally, this weighting and/or association value may correspond to a likelihood or probability that a given sequence is from a particular type of diseased tissue (such as a cancer tissue such as a lung cancer tissue or a colorectal cancer tissue, or from a non-cancer diseased tissue such as an infarcted myocardial tissue, or a diseased cerebrovascular tissue, or a placental tissue undergoing eclampsia or pre-eclampsia).

Optionally, any such weighting and/or association value for any one or more reference sequences may be established by an empirical measurement and/or evaluation process. Optionally, a weighting and/or association value for any one or more reference sequences may be established by measuring the expression (e.g. RNA levels) of two or more transcripts in two or more different tissue types (for example, a diseased tissue and a healthy tissue), and then the absolute and/or relative expression level(s) of said two or more transcripts within the first and second tissue types may be established empirically as said weighting and/or association value(s) for said first and second tissue types respectively. Optionally, any weighting and/or association value for any one or more reference sequences may be established by measuring the level of 5-methylcytosine (or, similarly, 5-hydroxy-methylcytosine) of two or more genomic regions (for example, two or more genes, or two or more gene promoter regions) in two or more different tissue types (for example, a diseased tissue and a healthy tissue), and then the absolute and/or relative 5-methylcytosine level(s) of said two or more genes (or promoters) within the first and second tissue types may be established empirically as said weighting and/or association value(s) for said first and second tissue types respectively. Further options are provided in PCT/GB2017/053820, which is incorporated herein by reference.

Optionally, any such weighting and/or association value for any one or more reference sequences may be established by an empirical measurement and/or evaluation process, wherein said empirical measurement and/or evaluation process employs one or more samples comprising one or more circulating microparticles as input samples for said empirical measurement and/or

evaluation process (for example, wherein first and second sequences of fragments of genomic DNA from a circulating microparticle are linked, such as by any method(s) described herein). Optionally, any said one or more circulating microparticles each comprise at least first and second fragments of genomic DNA. Optionally, any said one or more samples comprising one or more

5 circulating microparticles may be obtained from patients with one or more particular diseases, such as cancer (such as lung cancer, or pancreatic cancer), or such as cancer at a particular stage (such as stage I, stage II, stage III, stage IV) or such as cancer with particular clinical characteristics (such as benign cancer, such as malignant cancer, such as local cancer, such as metastatic cancer, or such as treatment-resistant cancer). Optionally, said one or more samples

10 comprising one or more circulating microparticles may be from patients who do not have any such one or more particular diseases. Optionally, said one or more samples comprising one or more circulating microparticles may be from patients who are considered to be healthy. Optionally, any said one or more samples comprising one or more circulating microparticles may comprise at least first and second samples from the same individual, wherein the first sample is made from

15 the individual at an earlier time, and the second sample is made from the individual at a later time, separated by a duration of time between the first and second samples (such as an hour, or a day, or a week, or a month, or 3 months, or 6 months, or 12 months, or 2 years, or 3 years, or 5 years, or 10 years). Optionally, any such weighting and/or association value for any one or more reference sequences may be established by an empirical measurement and/or evaluation

20 process, wherein said empirical measurement and/or evaluation process employs at least one sample (comprising one or more circulating microparticles) from a patient with a disease, and at least one sample (comprising one or more circulating microparticles) from a person without said disease (for example, wherein the amount and/or signal corresponding to said reference sequence within the sample(s) from the person(s) with the disease is compared to the amount

25 and/or signal corresponding to said reference sequence within the sample(s) from the person(s) without the disease, for example wherein the ratio of said two measures is employed as said weighting and/or association value). Optionally, any such weighting and/or association value for any one or more reference sequences may be established by an empirical measurement and/or evaluation process, wherein said empirical measurement and/or evaluation process employs

30 samples (comprising one or more circulating microparticles) from a group of at least two patients with a disease, and samples (comprising one or more circulating microparticles) from a group of at least two people without said disease. Optionally, any said groups of patients with a disease (or groups of persons without said disease) may each comprise at least 3, at least 5, at least 10, at least 20, at least 50, at least 100, at least 200, at least 500, at least 1000, at least 2000, at least

35 10,000, at least 20,000, at least 50,000, at least 100,000, at least 500,000, at least 1,000,000, or at least 10,000,000 individuals. Optionally, any patients within said groups of patients with a disease (or any persons within said groups of persons without said disease) may each provide two or more samples comprising circulating microparticles, wherein each sample is obtained at a different time point (such as time points separated by at least a day, by at least a week, by at

least a month, by at least 2 months, by at least 6 months, by at least a year, by at least 2 years, or by at least 5 years).

5 Optionally, in any method wherein one or more samples comprising one or more circulating microparticles are employed as input samples to establish any weighting and/or association value for any one or more reference sequences by an empirical measurement and/or evaluation process, said weighting and/or association value(s) may relate to a 5-methylcytosine level (for example they may relate to a 5-methylcytosine level within a particular healthy or particular diseased tissue), or optionally may relate to a 5-hydroxy-methylcytosine level (for example they
10 may relate to a 5-hydroxy-methylcytosine level within a particular healthy or particular diseased tissue). Further options are provided in PCT/GB2017/053820, which is incorporated herein by reference.

15 Optionally, the method may comprise counting the number of reference sequences from one or more list(s) of reference sequences in a set of linked sequence reads (i.e. a set of linked signals). Optionally, this counting process may be performed for all sets of linked sequence reads in a sample, or any one or more subsets thereof. Optionally, each reference sequence may be associated with a weighting and/or association value, such that the counting process comprises a weighted counting process, wherein a weighted sum of reference sequences within a set of linked
20 sequence reads is determined. Optionally, this weighting value may correspond to a likelihood or probability that a given sequence is non-maternal or paternal, or correspond to a likelihood or probability that a given sequence is maternal, or correspond to a likelihood or probability that a given sequence is from a particular tissue of origin (such as a lung tissue, or a pancreas tissue, or a lymphocyte), or correspond to a likelihood or probability that a given sequence is from a
25 particular healthy tissue of origin (such as a healthy lung tissue, or a healthy pancreas tissue, or a healthy lymphocyte), or correspond to a likelihood or probability that a given sequence is from a particular diseased tissue of origin (such as a diseased lung tissue, or a diseased pancreas tissue, or a diseased lymphocyte), or correspond to a likelihood or probability that a given sequence is from a particular cancerous tissue of origin (such as a cancerous lung tissue, or a
30 cancerous pancreas tissue, or a cancerous lymphocyte),

Optionally, any sum or weighted sum of reference sequences from a set of linked sequence reads may be compared to one or more threshold values, and wherein sets of linked sequence reads (i.e. sets of linked signals) comprising a number of reference sequences greater than said
35 threshold value(s) are determined and/or suspected to be from a particular tissue of origin. Optionally, any process of determining any such said sum and comparing with one or more threshold may be performed for all sets of linked sequence reads in the sample, and/or any one or more subsets thereof. Further options are provided in PCT/GB2017/053820, which is incorporated herein by reference.

Optionally, any one or more sets of linked sequences (or, for example, all sets of linked sequence reads (i.e. sets of linked signals) in a sample) may be analysed by and/or compared with two or more different lists of reference sequences. Optionally, sets of linked sequence reads in a sample may be analysed with a first list of reference sequences that correspond to a first particular tissue type, and also analysed with a second list of reference sequences that correspond to a second particular tissue type. Optionally, sets of linked sequence reads in a sample may be analysed with a first list of reference sequences that correspond to a particular healthy tissue type, and also analysed with a second list of reference sequences that correspond to a particular diseased tissue type. Optionally, sets of linked sequence reads in a sample may be analysed with a first list of reference sequences that correspond to a particular healthy tissue type, and also analysed with a second list of reference sequences that correspond to a cancerous tissue of the same tissue type. Further options are provided in PCT/GB2017/053820, which is incorporated herein by reference.

The sequence reads from the set of linked sequence reads (i.e. a set of linked signals) may be mapped to two or more reference nucleotide sequences corresponding to the same genomic region or genomic regions, wherein each reference nucleotide sequence comprises a different mutated allele or different set of mutated alleles within said genomic region or genomic regions, and said parameter value may be determined by the presence of one or more reference nucleotide sequences within said set of linked sequence reads.

The lengths of said fragments of a target nucleic acid (e.g. genomic DNA) may be determined or estimated, and the parameter may comprise a mean, media, mode, maximum, minimum, or any other single representative value of said determined or estimated lengths. Optionally, the lengths of genomic DNA sequence within each sequenced fragment is determined by sequencing substantially an entire sequence of a fragment of genomic DNA (i.e. from its approximate 5' end to its approximate 3' end) and counting the number of nucleotides sequenced therein. Optionally, this is performed by sequencing a sufficient number of nucleotides at the 5' end of the sequence of fragmented genomic DNA to map said 5' end to a locus within a reference human genome sequence, and likewise sequencing a sufficient number of nucleotides at the 3' end of the sequence of fragmented genomic DNA to map said 3' end to a locus within a reference human genome sequence, and by then calculating the total span in nucleotides comprising said 5' segment within the reference human genome sequence, said 3' segment within the reference human genome sequence, as well as any un-sequenced human genome sequence contained between the two sequenced portions.

The parameter value may be determined for at least 2, at least 10, at least 100, at least 1000, at least 10,000, at least 100,000, at least 1,000,000, at least 10,000,000, at least 100,000,000, or at least 1,000,000,000 sets of linked sequence reads (i.e. sets of linked signals).

The parameter value may be determined for at least 2 sets of linked sequence reads (i.e. sets of linked signals), and the parameter value may be evaluated by determining the number of sets of linked sequence reads where the parameter value is equal to a specific parameter value, equal to one of a set of two or more parameter values, less than a specific parameter value, greater than a specific parameter value, or within at least one range of values for the said parameter, or within one of two or more ranges of values for the said parameter. Optionally, the fraction or proportion of sets of linked sequence reads determined to meet one or more of the above conditions out of all evaluated sets of linked sequence reads is determined. Optionally, a parameter value is determined for at least 2 sets of linked sequence reads, and the mean, average, mode, or median parameter value across the group of parameter values is determined.

The parameter value is determined for a group of at least 2 sets of linked sequence reads (i.e. sets of linked signals), and the parameter values may be evaluated by comparing the group of parameter values with a second group of parameter values. Optionally, said second group of parameter values may correspond to an expected normal distribution of parameter values, or to an expected abnormal distribution of parameter values. Optionally, these parameter values may be derived from synthetic data, from randomized data, or from experimental data generated from one or more separate samples of circulating microparticles representing one or more normal or abnormal conditions. Optionally, at least 1, at least 10, at least 100, at least 1000, at least 10,000, at least 100,000, or at least 1,000,000 further groups of parameter values may be determined and further compared with the first group of parameter values. Further options are provided in PCT/GB2017/053820, which is incorporated herein by reference.

At least two different parameter values may determined for the set of linked sequence reads (i.e. a set of linked signals). Optionally, at least 3, at least 10, at least 100, at least 1000, at least 10,000, at least 100,000, at least 1,000,000, at least 10,000,000, or at least 100,000,000 different parameter values are determined.

The invention provides a method of determining a group of sets of linked sequence reads (i.e. sets of linked signals) comprising: (a) determining a parameter value for each of two or more sets of linked sequence reads, wherein the parameter value for each set of linked sequence reads is determined according to any method described herein; and (b) comparing the parameter values for the sets of linked sequence reads to identify a group of two or more sets of linked sequence reads.

The group of sets of linked sequence reads (i.e. sets of linked signals) may be determined by identifying sets of linked sequence reads having a parameter value equal to a specific parameter value, equal to one of a set of two or more parameter values, less than a specific parameter

value, greater than a specific parameter value, or within at least one range of values for the said parameter value, or within one of two or more ranges of values for the said parameter value.

Optionally, the number of sets of linked sequence reads within the group is determined, thus determining the size of the group.

5

The method may comprise further evaluating a group of sets of linked sequence reads (i.e. sets of linked signals), wherein the group of sets of linked sequence reads is further analysed by a second analysis step. Optionally, this second analysis step comprises determining and/or evaluating a second parameter value for the group of sets of linked sequence reads. Optionally, this second analysis step comprises determining the presence or absence of specific alleles within the sequences comprised within the group of sets of linked sequence reads. Optionally, this second analysis step comprises determining the presence or absence of chromosomal abnormalities such as one or more aneuploidies, or microdeletions, or copy number variations, or a loss-of-heterozygosity, or a rearrangement or translocation event, a single-nucleotide variant, a de novo mutation, or any other genomic feature or mutation.

15

The method may comprise further evaluating the group of sets of linked sequence reads (i.e. sets of linked signals) by a second analysis step, wherein the second analysis step comprises determining the number of sequence reads within each set of linked sequence reads within the group of sets of linked sequence reads that map to one or more reference nucleotide sequences. Optionally, this reference sequence or reference sequences may comprise an entire genome, an entire chromosome, a part of a chromosome, a gene, a part of a gene, any other part or parts of a genome, or any other synthetic or actual sequence. Optionally, this second analysis step comprises counting the total number of sequence reads within the group that map within a reference sequence, and then dividing this number of sequence reads by the total number of sets within the group, to estimate a relative number of sequence reads within the reference sequence per set. This may thus form an estimate of the relative number of sequence reads within the reference sequence per microparticle within the original sample of microparticles corresponding to the group of sets of linked sequence reads. Optionally, this second analysis step may further comprise a step of comparing this estimated relative number to a threshold value, wherein an estimated relative number greater than said threshold value, or alternatively an estimated relative number lesser than said threshold value may indicate the presence or absence of a specific medical or genetic condition, such as a chromosomal aneuploidy or microdeletion.

20

25

30

35

29. METHODS FOR DETERMINING AND ANALYSING SETS OF LINKED SIGNALS FROM MICROPARTICLES

Optionally, for any method described herein, any number of one or more parameter values may be determined and/or calculated and/or estimated (and then optionally further analysed and/or evaluated and/or compared with any method and/or reference value(s) and or control

parameter(s)), wherein any one or more parameter values are derived from and/or related to and/or are associated with any measurement(s) of any signal(s) and/or any signal(s) themselves (for example, any signal(s) from a set of at least two linked signals, such as a set of at least two linked signals from measurements of a circulating microparticle), wherein said measurement(s) and/or signal(s) are derived from and/or relate to and/or are associated with any type of molecule and/or biomolecule and/or target molecule and/or target biomolecule, such as any one or more fragments of genomic DNA, any one or more RNA sequences and/or RNA molecules, any one or more modified nucleotides and/or modified nucleobases, any one or more polypeptides (such as any one or more proteins and/or target proteins, and/or any one or more post-translationally modified proteins), such as any level, and/or any presence, and/or any absence, of any one or more such molecule(s) and/or biomolecule(s). Optionally, any such parameter value(s) may be compared to one or more control parameter value(s), optionally wherein one or more such control parameter value(s) are determined from one or more second and/or different signals (such as from one or more signal(s) from a second, different set of linked signals, such as from a second set of linked signals from measurement(s) of a second, different circulating microparticle). Any parameter value(s) and/or control parameter value(s) may be determined for at least 2, at least 10, at least 100, at least 1000, at least 10,000, at least 100,000, at least 1,000,000, at least 10,000,000, at least 100,000,000, or at least 1,000,000,000 sets of linked signals. At least two different parameter values may determined for any set of linked signals. Optionally, at least 3, at least 10, at least 100, at least 1000, at least 10,000, at least 100,000, at least 1,000,000, at least 10,000,000, or at least 100,000,000 different parameter values may be determined. Options for methods involving and relevant to calculation, derivation, establishment, analysis and/or use of any such parameter value(s) and/or control parameter value(s) are provided in PCT/GB2017/053820, which is incorporated herein by reference.

Optionally, any number of one or more signal(s) corresponding to a level (and/or any estimated level and/or predicted level and/or measured level) of any molecule and/or biomolecule and/or target molecule and/or target biomolecule (such as any level of a modified nucleotide and/or modified nucleobase, or any level of a target polypeptide or target post-translationally modified polypeptide) may comprise a parameter value and/or control parameter value. Options for methods involving and relevant to any such parameter value(s) and/or control parameter value(s) are provided in PCT/GB2017/053820, which is incorporated herein by reference.

Optionally, any number of one or more signal(s) corresponding to the presence, and/or comprising the absence (and/or any predicted or measured presence or absence) of any molecule and/or biomolecule and/or target molecule and/or target biomolecule (such as any level of a modified nucleotide and/or modified nucleobase, or any level of a target polypeptide or target post-translationally modified polypeptide) may comprise a parameter value and/or control parameter value, such as a qualitative or categorical parameter value and/or control parameter

value. Options for methods involving and relevant to any such parameter value(s) and/or control parameter value(s) are provided in PCT/GB2017/053820, which is incorporated herein by reference.

- 5 Optionally, in any method(s) wherein a sample comprising circulating microparticles (and/or a sample derived from circulating microparticles) is divided into at least two subsets and/or sub-populations (e.g. into a first subset of circulating microparticles and a second subset of circulating microparticles, for example wherein a sample is sorted such as FACS sorted into a first subset of circulating microparticles exhibiting high levels of a particular target biomolecule, and into a
10 second subset of circulating microparticles exhibiting low levels of said particular target biomolecule), membership within any one or more subsets and/or sub-populations of circulating microparticles may comprise a parameter value, such as a qualitative and/or categorical value.

- Optionally, in any method(s) involving use of one or more barcoded affinity probes, any one or
15 more reference sequences (e.g. any reference sequence(s) employed to analyse one or more sets and/or groups of linked sequences and/or linked sequence reads and/or linked signals) may comprise one or more oligonucleotide sequences comprised within said one or more barcoded affinity probes (e.g. any one or more reference sequences may comprises sequences of oligonucleotides, such as sequences of barcoded oligonucleotides, comprised within any one or
20 more barcoded affinity probe(s)). Optionally, in any method(s) involving use of one or more barcoded affinity probes wherein said barcoded affinity probes have affinity for a polypeptide encoded in the human genome, each sequence from any one or more set(s) of linked sequence reads comprising a sequence within a barcoded affinity probe may be considered (e.g. may informatically be considered) to map (e.g. to synthetically or artificially map) to a reference
25 sequence comprising all or part of the human genome sequence corresponding to the gene of the protein to which each said barcoded affinity probe(s) have affinity. Optionally, any method(s) involving the generation, prediction, calculation, and/or analysis or use of parameter values related to reference sequence(s) may employ reference sequences associated in any way with any one or more barcoded affinity probe(s). Any such one or more reference sequence(s) may be
30 associated with a weighting and/or association value, optionally wherein any such weighting and/or association value(s) may be established by any empirical measurement and/or evaluation process(es) (such as by any empirical measurement and/or evaluation process(es) involving one or more samples from one or more individuals or groups of individuals, such as groups of healthy individuals and/or groups of individuals with one or more diseases or conditions; optionally
35 wherein said samples may comprise circulating microparticles, and/or optionally wherein said samples may comprise other samples such as tissue and/or biopsy samples). Options for methods involving and relevant to any such reference sequences and/or parameter value(s) and/or values and/or weighting and/or association value(s) and/or empirical measurement and/or

evaluation process(es) are provided in PCT/GB2017/053820, which is incorporated herein by reference.

5 For any analysis involving two or more signals that are linked informatically by any such way, the existence (or lack thereof) of linking may be employed as a parameter (such as a parameter value and/or control parameter value) in any analysis or evaluation step or any algorithm for performing same. For any analysis involving two or more signals that are linked informatically by any such way, the degree, probability, extent or level of linking may be employed as a parameter in any analysis or evaluation step or any algorithm for performing same.

10 The invention provides a method of determining a parameter value from a set of linked signals wherein the method comprises: (a) determining a set of linked signals according to any of the methods described herein; and (b) determining the parameter value by counting or identifying the presence of one or more reference nucleotide sequences within the set of linked signals.

15 Any parameter value may be a quantitative or semi-quantitative value and may be determined by counting the number of sequence reads within a set of linked sequences that are determined to comprise a sequence originating from the any reference nucleotide sequence or sequences. Further options are provided in PCT/GB2017/053820, which is incorporated herein by reference.

20 Any parameter value(s) and/or control parameter value(s) may be determined for at least 2 sets of linked signals, and the parameter value may be evaluated by determining the number of sets of linked signals where the parameter value is equal to a specific (e.g control) parameter value, equal to one of a set of two or more parameter values, less than a specific parameter value, greater than a specific parameter value, or within at least one range of values for the said parameter, or within one of two or more ranges of values for the said parameter. Optionally, the fraction or proportion of sets of linked signals determined to meet one or more of the above conditions out of all evaluated sets of linked signals is determined. Optionally, a parameter value is determined for at least 2 sets of linked signals, and the mean, average, mode, or median parameter value across the group of parameter values is determined.

30. METHODS FOR TRANSFORMING LINKED SEQUENCE READ DATA FOR ANALYSIS BY ALGORITHMS

35 The invention provides methods for transforming linked sequence data into forms representative thereof that may be more readily or more comprehensively analysed by analytic or statistical tools. Of particular importance, the methods may be used to analyse particular samples of circulating microparticles for the presence of structural abnormalities (for exempling, translocations, or large-scale copy number variations), but wherein the specific nature, genomic

location, or size of said structural abnormalities is not known previously, and furthermore, where such factors may not be of direct importance to the particular biological measurement.

Sequences from microparticles may be used to detect the presence of structural abnormalities that may indicate the presence of cancer within the body of the person from whom the sample was derived. The presence and/or burden of a certain number of structural abnormalities itself may be indicative of cancer (or indicative of a risk thereof), but the genomic locations of such potential abnormalities may be neither known prospectively nor relevant to the cancer risk assessment; thus transforming linked microparticle sequence data into a form more readily analysable with informatic or statistical tools may enhance the sensitivity and specificity of this method. Of particular importance, the transformation methods may enable analysis of such microparticle linked-sequence data with a particular family of numeric tools that typically require some transformation of the data for effective analysis, such as deep learning and/or machine learning approaches, as well as neural network/recurrent neural network approaches.

The invention provides a method of transforming linked sequence data generated from a sample of microparticles, wherein a first set of linked sequence reads (i.e. a first set of linked signals) is generated from fragments of a target nucleic acid of a first circulating microparticle, and wherein a second set of linked sequence reads (i.e. a second set of linked signals) is generated from fragments of a target nucleic acid of a second circulating microparticle.

The first and second sets of linked sequence reads (i.e. sets linked signals) may be mapped to a reference genome sequence, and wherein each sequence read is transformed into a representation comprising the chromosome to which it was mapped, and an index function, wherein said index function comprises its linkage to another at least 1 sequence from the same set of linked sequence reads. Optionally, said index function may be a unique identifier that identifies the corresponding set of linked sequence reads.

31. METHODS FOR DETERMINING GENOMIC REARRANGMENTS, TRANSLOCATIONS, STRUCTURAL VARIANTS, OR GENOMIC LINKAGES

The invention provides a method of determining the presence of a genomic rearrangement or structural variant within a set of linked sequence reads (i.e. set of linked signals) of fragments of a target nucleic acid (e.g. genomic DNA) from a single microparticle, wherein the method comprises: (a) determining a set of linked sequence reads according to any of the methods described herein; and (b) mapping (at least a portion of) each sequence of the set of linked sequence reads to a first reference nucleotide sequence comprising a first genomic region, and mapping (at least a portion of) each sequence of the set of linked sequence reads to a second reference nucleotide sequence comprising a second genomic region; and (c) counting the number

of sequence reads from the set of linked sequence reads that are found to map within the first genomic region, and counting the number of sequence reads from the set of linked sequence reads that are found to map within the second genomic region.

- 5 The genomic rearrangement or structural variant may be any type of genomic-structural phenomenon e.g. a genomic copy number variation (including a copy number gain or a copy number loss), a microdeletion, or any sort of rearrangement (e.g. an inversion), a translocation such a chromosomal translocation (e.g. an intra-chromosomal translocation or an inter-chromosomal translocation).

10

In the methods, the numbers of counted number of sequence reads may then be used in a further evaluation step or statistical analysis to determine whether a genomic linkage (i.e. a connection along the same stretch of a chromosome) may exist between the first genomic region and the second genomic region. The method may be conducted for a single set of linked sequence reads (i.e. set of linked signals), and it may also be conducted for a group of two or more sets of linked sequence reads, as well as conducted for all sets of linked sequence reads within a sample of microparticles, or a subgroup thereof.

15

- 20 Optionally, the total number of sequence reads within the set of linked sequence reads (i.e. set of linked signals) is also determined. The first and the second genomic regions may be located within the same chromosome, and if so then may be immediately adjacent to each other or may be separated by any number of nucleotides. Alternatively, the first and the second genomic regions may be located within two different chromosomes. The first and second genomic regions may each be any number of nucleotides in length, from 1 nucleotide to the length of a chromosome arm or an entire chromosome.

25

- 30 Optionally, an evaluation is performed wherein the number of sequence reads within the first genomic region are compared with a first threshold value, and the number of sequence reads within the second genomic region compared with a second threshold value, wherein the first number being equal to or above the first threshold value and the second number being equal to or above the second threshold value determines or indicates the presence of a genomic linkage between the first genomic region and the second genomic region and/or the presence of a rearrangement or translocation event involving the first and the second genomic regions.

30

- 35 Further options are are provided in PCT/GB2017/053820, which is incorporated herein by reference.

32. METHODS FOR PHASING VARIANTS OR VARIANT ALLELES

The invention provides methods for phasing alleles that are distributed across a chromosomal region. These analyses may be geared towards any application or task where the presence of two nucleic acid variants on the same chromosome or on two different chromosomes may have biological or medical significance. For example, wherein two different variant sites may be found within a single gene (the case of compound heterozygosity), it can be highly relevant whether a mutation in the first site is located within the same copy of the gene within an individual's genome as a mutation in the second site, or if, by contrast, they are each located on one of the two different copies of the gene within the individual's genome – for example, if two mutations are inactivating mutations, then their being located on the same copy of the gene will still allow for one active, functioning copy of the gene, whereas if the two inactivating mutations are each located on one of the two copies of the gene, then neither copy of the gene will be active.

The invention provides a method of phasing two variant alleles, wherein a first variant allele is comprised within a first genomic region, and wherein a second variant allele is comprised within a second genomic region, and wherein each variant allele has at least two variants or potential variants, wherein the method comprises: (a) determining a set of linked sequence reads (i.e. set of linked signals) according to any of the methods described herein; and (b) determining whether a sequence comprising each potential variant from the first variant allele is present within the set of linked sequence reads, and determining whether a sequence comprising each potential variant from the second variant allele is present within the same set of linked sequence reads.

The variant allele may comprise a single nucleotide, or a region of two or more nucleotides, or insertions and/or deletions of one or more nucleotides. Optionally, a further evaluation step is performed in which the presence of a first variant of a first allele is detected, and wherein the presence of a first variant of a second allele is detected, and wherein these two alleles being found within the same set of linked sequence reads (i.e. set of linked signals) indicates or estimates a probability that the two alleles are in the same chromosomal phase as each other, and/or linked along the same chromosome or haplotype or haplotype block.

The method may be repeated for two or more pairs of variant alleles, comprising any potential variant allele, and any potential variant within an allele or a variant allele site, and any combination thereof of any two or more different such variant alleles.

The method may be performed on a single set of linked sequence reads (i.e. set of linked signals) from a microparticle, or it may be performed on a group of two or more sets of linked sequence reads. It may also be performed on all sets of linked sequence reads from a particular sample, and it may also be performed on one or more particular groups of sets of linked sequence reads.

Further options are provided in PCT/GB2017/053820, which is incorporated herein by reference.

5 Optionally, the method may be used to phase three or more variant alleles. Optionally, this may be performed by phasing all said three or more variant alleles simultaneously within a single step, or may be performed by a sequence of two or more sequential steps.

10 Optionally, the method may be used to phase variant alleles (e.g. at least 2, at least 5, at least 10, at least 25, at least 50, at least 100, at least 500, at least 1000, at least 10,000, or at least 100,000 variant alleles) across a genomic span. The genomic span may be at least 100 kilobases, at least 1 megabase, at least 10 megabases, or an entire chromosome arm or an entire chromosome. Further options are provided in PCT/GB2017/053820, which is incorporated herein by reference.

15 The variant allele may be any sort of genetic variant, including single-nucleotide variant or single-nucleotide polymorphism, a variant that is two or more nucleotides in length, an insertion or deletion of one or more nucleotides, a de novo mutation, a loss-of-heterozygosity, a rearrangement or translocation event, a copy number variation, or any other genomic feature or mutation.

20 The method may comprise or be extended to comprise a genetic imputation process. Optionally, a list of one or more alleles or variant alleles from a set of linked sequence reads (i.e. set of linked signals) from a microparticle is determined to perform a genetic imputation process; optionally this list may be determined from a group of two or more sets of linked sequence reads, or from a particular sub-group of sets of linked sequence reads. A genetic imputation process may be performed in which one or more such lists are compared with one or more previously known haplotypes or haplotype blocks from a human population, to phase or to estimate the phase of the alleles or variant alleles within said lists, or to determine or estimate a haplotype or haplotype block for a portion of the genome from which said sequences were derived. Optionally, two or more alleles or variant alleles may be phased prior to performing a genetic imputation process. Optionally, the phasing of such two or more alleles or variant alleles may be performed through any process as above. Optionally, a combined and/or iterative process of phasing and/or genetic imputation and/or haplotype estimation may be performed, wherein any such step or component may be repeated one, two or a greater number of times.

35 Any tools and/or methods and/or informatic approaches to performing genetic imputation and/or haplotype estimation and/or phasing and/or variant estimation may be employed. Optionally, SHAPEIT2, MaCH, Minimac, IMPUTE2, and/or Beagle may be employed.

Optionally, a genetic imputation process may be employed to generate one or more reference sequences (e.g. to generate one or more lists of reference sequences). Optionally, a genetic imputation process may be employed concurrently to and/or along with a haplotype-estimation process. Further options for a genetic imputation process are provided in PCT/GB2017/053820, which is incorporated herein by reference.

Optionally, a genetic imputation process may employ an input list of sequences and/or alleles (e.g. a list of single-nucleotide polymorphisms), wherein said input list is derived from sequences of fragments of genomic DNA from circulating microparticles. Optionally, said input list may be derived from linked sequences of fragments of genomic DNA from circulating microparticles. Further options for said input list are provided in PCT/GB2017/053820, which is incorporated herein by reference. Optionally, said input list may be derived from a subset of (linked or unlinked) sequences of fragments of genomic DNA from circulating microparticles, wherein said subset of sequences comprises sequences contained within, and/or likely to be contained within, and/or enriched within, and/or suspected to be enriched within, a cancer genome.

Any an input list of sequences and/or alleles (e.g. a list of single-nucleotide polymorphisms), and/or any one or more reference sequences (e.g. one or more lists of reference sequences) and/or any subset thereof may be generated by any method described herein.

Optionally, a genetic imputation process may be employed to generate, determine, or estimate a haplotype or haplotype block for a portion of a genome. Further options for a genetic imputation process are provided in PCT/GB2017/053820, which is incorporated herein by reference.

Optionally, a genetic imputation process may employ a catalogue of two or more previously known (and/or previously predicted or created) haplotypes or haplotype blocks from a human population. Optionally, a haplotype or haplotype block may relate to a genomic region at least 2 nucleotides, at least 10, at least 100, at least 1000, at least 10,000, at least 100,000, at least 1,000,000, at least 10,000,000, or at least 100,000,000 nucleotides in length; optionally, a haplotype or haplotype block may relate to a chromosome arm, a full chromosome, and/or a full genome.

Optionally, a genetic imputation process may employ a catalogue of at least 2, at least 3, at least 5, at least 10, at least 50, at least 100, at least 500, at least 1000, at least 5000, at least 10,000, at least 50,000, at least 100,000, at least 500,000, or at least 1,000,000 more previously known (and/or previously predicted or created) haplotypes or haplotype blocks.

The method may be conducted for a single set of linked sequence reads (i.e. set of linked signals), and it may also be conducted for a group of two or more sets of linked sequence reads,

as well as conducted for all sets of linked sequence reads within a sample of microparticles, or a subgroup thereof.

33. METHODS FOR DETERMINING AND ANALYSING LINKED SEQUENCE

5 READS OF FOETAL ORIGIN

The invention provides methods for analyzing linked sequence data wherein said data is generated from a sample from a pregnant female (thus the sample may comprise a mixture of microparticles of maternal origin, i.e. from normal somatic maternal tissues, and microparticles of foetal (and/or placental) origin). The methods may be used to detect the presence of a foetal
10 chromosomal abnormality, such as a foetal trisomy, or a foetal chromosomal microdeletion. Several such methods may be performed on the same set of foetal sequences, thus enabling multiplexed and sensitive detection of foetal genetic conditions.

The invention provides a method of determining a set of linked sequence reads (i.e. set of linked
15 signals) of foetal origin, wherein the method comprises: (a) determining a set of linked sequence reads according to any of the methods described herein, wherein the sample comprises microparticles originating from maternal blood; and (b) comparing (at least a portion of) each sequence read of the set of linked sequence reads to a reference list of sequences present in the foetal genome; and (c) identifying a set of linked sequence reads of foetal origin by the presence
20 of one or more sequences from the reference list within one or more sequence reads of the set of linked sequence reads.

A set of linked sequence reads (i.e. set of linked signals) of foetal origin may comprise, consist of or consist essentially of sequence reads of fragments of a target nucleic acid originating from a
25 foetus. Optionally, a set of linked sequence reads of foetal origin may comprise or consist of sequence reads of fragments of a target nucleic acid originating from a foetus, and also comprise or consist of sequence reads of fragments of a target nucleic acid originating from one or more maternal tissues and/or maternal cells.

The reference list of sequences (or sequence variants) present in the foetal genome may
30 comprise, consist of, or consist essentially of, sequences enriched in the foetal genome. The reference list of sequences present in the foetal genome may comprise, consist of, or consist essentially of, sequences enriched in the foetal genome (compared to the maternal genome). Further options for the reference list of sequences present in the foetal genome Further are
35 provided in PCT/GB2017/053820, which is incorporated herein by reference.

The microparticles may originate from the maternal blood of a pregnant individual. Optionally, the microparticles may originate from the maternal blood of a pregnant individual wherein the individual is pregnant with at least two developing fetuses (e.g. the individual is pregnant with

- 190 -

twins, or triplets, or any larger number of developing fetuses). Optionally, the microparticles may originate from the maternal blood of a pregnant individual wherein the pregnancy has been generated through an in vitro fertilisation. Optionally, any in vitro fertilisation process may further comprise any step of pre-implantation genetic screening, pre-implantation genetic diagnosis, pre-implantation embryo evaluation, and/or pre-implantation embryo selection.

34. METHODS FOR DIAGNOSIS AND MONITORING

The invention provides methods of diagnosis and monitoring based on any of the methods described herein.

10

The invention provides a method of diagnosing a disease or condition in a test subject, wherein the method comprises: (a) determining a parameter value for a first set of linked sequence reads (i.e. set of linked signals) determined from a test sample from the subject, wherein the parameter value is determined according to any of the methods described herein; and (b) comparing the parameter value for the set of linked sequence reads determined from the test sample to a control parameter value.

15

The control parameter value may be determined from a second set of linked sequence reads (i.e. set of linked signals) determined from the test sample from the subject, wherein the control parameter value is determined according to any of the methods described herein.

20

The control parameter value may be determined from a set of linked sequence reads (i.e. set of linked signals) determined from a control sample, wherein the control parameter value is determined according to any of the methods described herein.

25

The disease or condition may be cancer, a chromosomal aneuploidy, or a chromosomal microdeletion, a genomic copy number variation (e.g. a copy number gain or a copy number loss), a loss-of-heterozygosity, a rearrangement or translocation event, a single-nucleotide variant, or a de novo mutation.

30

The invention provides a method of monitoring a disease or condition in a test subject, wherein the method comprises: (a) determining a parameter value for a first set (of sets) of linked sequence reads determined from a test sample from the subject, wherein the parameter value is determined according to any of the methods described herein; and (b) comparing the parameter value for the set of linked sequence reads (i.e. set of linked signals) to a control parameter value.

35

The control parameter value may be determined from a second set of linked sequence reads (i.e. set of linked signals) determined from a control sample obtained from the same subject at an

– 191 –

earlier time point than the test sample. The time interval between the control and test samples being obtained may be at least 1 day, at least 1 week, at least 1 month or at least 1 year.

Any method of determining a parameter value and/or performing a second analysis step
5 described herein may be performed independently on linked sets of sequences from two or more different samples from a subject separated by a time interval, where the two or more different samples are from the same subject, wherein the time interval is at least 1 day, at least 1 week, at least 1 month at least 1 year, at least 2 years, or at least 3 years. Any such parameter value and/or result of a second analysis step may be compared between any two or more such different
10 samples. The absolute or relative difference between such parameter value and/or result of a second analysis step may be determined by such a comparison step. Optionally, such absolute or relative differences may be normalised to and/or divided by the length of the time interval between the two samples. Optionally, such absolute or relative differences and/or associated normalised values may be compared with one or more threshold values, wherein a value above such a
15 threshold value may indicate a disease or a condition, such as cancer or a heightened risk of cancer development.

The disease or condition may be cancer.

20 The invention provides a method of diagnosing a disease or condition in a subject, wherein the method comprises: (a) determining a set of linked sequence reads (i.e. set of linked signals) according to any of the methods described herein, wherein the sample comprises a microparticle originating from blood; and (b) comparing (at least a portion of) each sequence read of the set of linked sequence reads to a reference list of sequences present in cells of the disease, wherein
25 the presence of one or more sequences from the reference list within one or more sequence reads of the set of linked sequence reads indicates the presence of the disease.

The disease or condition may be cancer.

30 The invention provides a method of determining a set of linked sequence reads (i.e. set of linked signals) of diseased cell (e.g. tumour cell) origin, wherein the method comprises: (a) determining a set of linked sequence reads according to any of the methods described herein, wherein the sample comprises a microparticle originating from blood; and (b) comparing (at least a portion of) each sequence read of the set of linked sequence reads to a reference list of sequences present
35 in cells of the disease (e.g. cells of a tumour); and (c) identifying a set of linked sequence reads of diseased cell (e.g. tumour cell) origin by the presence of one or more sequences from the reference list within one or more sequence reads of the set of linked sequence reads.

- 192 -

The invention provides a method of determining a tumour genotype comprising: (a) determining a set of linked sequence reads (i.e. set of linked signals) of tumour origin according to any of the methods described herein; and (b) determining the tumour genotype from the set of linked sequence reads of tumour origin.

5

The sample may comprise a microparticle (or microparticles) originating from blood from a patient diagnosed with the disease (e.g. cancer). The sample may comprise a microparticle (or two or more microparticles) originating from blood from a patient suspected of having the disease (e.g. cancer).

10

Optionally, in any method(s) of diagnosing and/or estimating or predicting the risk of and/or monitoring any one or more disease(s) and/or conditions, the method(s) may comprise a further step (i.e. a result-communication step) wherein any one or more result(s) of the method (e.g. any one or more diagnostic result(s) and/or readout(s), and/or any one or more prognostic result(s) and/or readout(s), and/or any one or more risk-stratification result(s) and/or readout(s) and/or any one or more risk-estimation result(s) and/or readout(s) and/or measurement(s)) is/are communicated to the patient (i.e. to the patient from which any one or more samples comprising one or more circulating microparticles had been derived) and/or said patient's representative and/or family member, and/or any one or more physician(s), nurse(s), and/or any other healthcare provider(s) and/or institution or organisation providing healthcare services to said patient.

15

20

Optionally, any result-communication step may comprise the last step of any method described herein. Optionally, any result-communication step may comprise communication of any such result(s) via electronic media such as email, internet-based communications and/or internet-based interface and/or any electronic messaging system and/or any telephone-based method such as phone calling and/or text messaging; and/or any paper-based method such as post; and/or any in-person method such as in-person conversation and/or disclosure. Optionally, in any such result-communication step, at least one such result may be communicated, and/or any two or more such result(s) may be communicated, and/or all such result(s) may be communicated, and/or any fraction or number of all such results may be communicated.

25

30

35. COMBINED MICROPARTICLE-BASED AND NON-MICROPARTICLE-BASED ANALYSIS

The methods of analysing a sample comprising one or more circulating microparticle(s) and/or a sample derived from one or more circulating microparticle(s) (for example, a method of diagnosing and/or monitoring and/or predicting any disease and/or condition and or genetic sequence and/or genetic mutation and/or genetic status or chromosomal or structural abnormality), may further comprise measurement and/or consideration of one or more non-microparticle measurements or factors measured from and/or associated with the same individual

35

from whom said circulating microparticle(s) were acquired and/or derived to perform a combined microparticle-based and non-microparticle-based analysis.

The methods of analysing a sample comprising one or more circulating microparticle(s) and/or a sample derived from one or more circulating microparticle(s), may be combined with one or more non-microparticle factors (such as personal factors, demographic factors, clinical/medical factors, molecular or biochemical factors, genetic factors, and/or any other form of health-related or health-history-related factors) from the same individual, such as weight, body-mass index (BMI), obesity status, gender, age, ethnicity and/or ethnic background, current and/or previous and/or historical smoking status, diabetes status (such as type I diabetes status and/or type II diabetes status), a history of one or more previous strokes, a history of one or more previous transient ischaemic attacks, a history of one or more previous pregnancies, a family history of any form of disease (such as any form of heart disease, and/or cardiovascular disease, and/or cancer, and/or any specific cancer type (such as breast and/or ovarian cancer), the results of any blood, plasma, and/or serum test or measurement (such as any blood count such as a complete blood count (CBC), and/or such as prostate specific antigen (PSA) level, and/or PSA velocity (over a period of months and/or years, and/or CA-125 levels and/or CA-125 velocity, and/or any metabolite measurements (such as a basic metabolic panel (BMP), and/or systolic and/or diastolic blood pressure, and/or blood cholesterol level and/or high blood cholesterol level status, and/or C reactive protein levels, and/or the results and/or interpreted results of any one or more electrocardiogram (ECG) tests, and/or the results and/or interpreted results of any one or more tissue biopsies or tissue aspirates (such as a lung biopsy, a heart biopsy, a liver biopsy, and/or a kidney biopsy, optionally wherein any such biopsy material is assessed by any molecular-pathologic process or technique, such as any immunohistochemistry technique, such as any in situ hybridisation technique (to analyse DNA and/or RNA molecules) and/or any cell-based or morphology-based techniques, and/or the presence of any one or more pre-existing conditions (such as any lung disease, any heart disease, any liver disease, any kidney disease, any neurologic disease, and/or any psychologic or psychiatric disease or condition), the results and/or interpreted results of any one or more medical imaging test(s) (such as any computed tomography scan, any spiral computed tomography scan, any low-dose computed tomography scan, any magnetic resonance imaging scan, any positron emission tomography scan, any ultrasound scan, and/or any optical coherence tomography scan), and/or the presence or absence of any one or more monogenic risk alleles (such as any breast cancer or ovarian cancer susceptibility or predisposition gene), and/or any polygenic risk scores or risk estimates, and/or any aforementioned and/or other measurement wherein said measurements are made and/or tracked longitudinally over time (such as on a monthly basis or a yearly basis, optionally wherein at least two such longitudinal measurements are made, or at least 3, or at least 5, or at least 10, or at least 20, or at least 100 longitudinal measurements are made). Optionally, any combination of two or more such non-microparticle factors (e.g PSA level and CA-125 level) may be measured

and/or determined and then analysed in conjunction with any method of analysing a sample comprising one or more circulating microparticle(s) (and/or a sample derived from one or more circulating microparticles) described herein; optionally any two or more such non-microparticle factors may be measured and/or determined from one or more patient blood sample(s), wherein said patient blood sample(s) also provides said sample(s) comprising one or more circulating microparticle(s). Optionally, any one or more non-microparticle factors may be compared with any one or more cutoffs and/or thresholds and/or normal (i.e. healthy) ranges and/or diseased (i.e. unhealthy) ranges, such as wherein any such non-microparticle factor being above any such threshold, below any such threshold, within any such range, and/or outside of any such range may indicate a health status (i.e. indicate healthiness for a particular disease or condition in said patient, i.e. a 'health status readout'), and/or may indicate a disease status (i.e. indicate the presence or a risk of a disease, i.e. a 'disease status readout') for a particular disease or condition; optionally any method of analysing one or more circulating microparticle(s) (and/or a sample derived from one or more circulating microparticles) may be analysed in conjunction with any number of (one or more) 'health status readout(s)' and/or 'disease status readout(s)' to create a combinatoric diagnostic, and/or prognostic, and/or risk-stratification and/or risk-estimation readout and/or measurement; optionally any such combinatoric diagnostic, and/or prognostic, and/or risk-stratification and/or risk-estimation readout and/or measurement may further comprise analysis by an algorithm and/or computer program (i.e. software), for example to generate and/or calculate one or more categorical scores or results (such as a high score or a low score, or a positive result or a negative result), and/or one or more quantitative or numeric scores (such as 1,2 or 3, or a number on a scale from 1 to 10 or 1 to 100, or a percentage or risk or likelihood rating), wherein said scores may optionally be associated with or indicative of a diagnosis, prognosis, risk estimate or likelihood and/or risk factor and/or risk category for any disease, condition, or syndrome.

36. METHODS AND USES FOR DIAGNOSIS, PROGNOSIS, AND/OR RISK-STRATIFICATION OR RISK-ESTIMATION

The methods of the invention may comprise a step of analysis by or in conjunction with one or more algorithms (such as a manual algorithm and/or an automated algorithm such as a computer-based and/or quantitative algorithm), and optionally or further may be employed to produce or estimate any diagnostic, and/or prognostic, and/or risk-stratification and/or risk-estimation readout and/or measurement. Any one or more such diagnostic, and/or prognostic, and/or risk-stratification and/or risk-estimation readouts and/or measurements may comprise one or more categorical scores or results (such as a high score or a low score, or a positive result or a negative result), and/or one or more quantitative or numeric scores (such as 1,2 or 3, or a number on a scale from 1 to 10 or 1 to 100, or a percentage or risk or likelihood rating), wherein said scores may optionally be associated with or indicative of a diagnosis, prognosis, risk

estimate or likelihood and/or risk factor and/or risk category for any disease, condition, or syndrome.

- Optionally, any such disease, condition, or syndrome may comprise any one or more cancers or pre-malignant conditions (such as any lung cancer, or any breast cancer, or any ovarian cancer, or any prostate cancer, or any kidney cancer, or any liver cancer, or any blood cancer, or any leukaemia, or any lymphoma, or any colorectal cancer, or any pancreatic cancer, or any brain cancer, or any uterine cancer, or any bile duct cancer, or any skin cancer, or any melanoma, or any bladder cancer, or any oesophageal cancer, or any oral cancer, or any pharyngeal cancer).
- Optionally, any such cancers or pre-malignant conditions may further comprise a diagnosis or estimate of cancer or pre-cancer stage and/or grade (such as stage 1, 2, 3, or 4), and/or any measure of aggressiveness, and/or any measurement or prediction or prognosis of metastasis or metastatic potential.
- Optionally, any such disease, condition, or syndrome may comprise any one or more cardiac or vascular diseases and/or conditions, such as myocardial infarction, atherosclerosis, cardiomyopathy (such as hypertrophic cardiomyopathy or dilated cardiomyopathy), heart failure, venous thrombosis, deep vein thrombosis, embolism, thrombosis, stroke (such as an ischaemic stroke or a haemorrhagic stroke), coronary artery disease, cerebrovascular disease, peripheral artery disease, endovascular plaques, stable endovascular plaques, unstable or vulnerable endovascular plaques, valvular heart disease, aneurisms, endocarditis, or myocarditis.
- Optionally, any such disease, condition, or syndrome may comprise any one or more diseases or conditions or complications associated with pregnancy, such as pre-eclampsia, eclampsia, gestational diabetes, preterm labour, hypertension, deep vein thrombosis, ectopic pregnancy, or any foetal genetic and/or chromosomal abnormality, such as one or more aneuploidies, or microdeletions, or copy number variations, or a loss-of-heterozygosity, or a rearrangement or translocation event, a single-nucleotide variant, a de novo mutation, or any other genomic feature or mutation. Optionally, any such disease, condition, or syndrome may comprise trisomy of chromosome 21 (i.e. Down Syndrome) in a developing foetus, and/or trisomy of chromosome 13 (i.e. Patau Syndrome) in a developing foetus, and/or trisomy of chromosome 18 (i.e. Edwards Syndrome) in a developing foetus, and/or trisomy of chromosome 9 in a developing foetus, and/or trisomy of chromosome 8 in a developing foetus, and/or Triple X Syndrome, and/or Klinefelter Syndrome. Optionally, any such disease, condition, or syndrome may comprise a genomic microdeletion, such as microdeletion syndrome, such as DiGeorge Syndrome, and/or Prader-Willi Syndrome, and/or Angelman Syndrome, and/or Neurofibromatosis Type I and/or Type II, and/or Williams Syndrome, and/or Miller-Dieker Syndrome.

Optionally, any such disease, condition, or syndrome may comprise any monogenic disease or mongenic disease predisposition, such as any monogenic disease or mongenic disease predisposition exhibiting a dominant inheritance pattern, and/or any monogenic disease or mongenic disease predisposition exhibiting a recessive inheritance pattern, and/or any

5 monogenic disease or mongenic disease predisposition exhibiting an X-linked inheritance pattern. Optionally, any such any monogenic disease or mongenic disease predisposition may comprise a Thalassaemia disease, and/or sickle cell anaemia, and/or haemophilia, and/or Tay Sachs disease, and/or cystic fibrosis, and/or Huntington's disease, and/or fragile-X syndrome.

Optionally, any such monogenic disease or mongenic disease predisposition may comprise a

10 foetal such monogenic disease or mongenic disease predisposition (i.e. present in a foetal genome, such as present in foetal nucleic acids comprised within a pregnant maternal blood sample).

Optionally, any method of analysing a sample comprising one or more circulating microparticle(s)

15 (and/or a sample derived from one or more circulating microparticles), may comprise a diagnostic, and/or prognostic, and/or risk-stratification and/or risk-estimation readout and/or measurement for a combined disease set of any two or more diseases, conditions, or syndromes (such as any combination of two or more diseases, conditions, or syndromes described herein). For example, any such method may comprise a diagnostic, and/or prognostic, and/or risk-stratification and/or

20 risk-estimation readout and/or measurement for each member of a combined disease set, for example a combined disease set comprising: lung cancer and breast cancer; or a combined disease set comprising: lung cancer and prostate cancer; or a combined disease set comprising: lung cancer and breast cancer and colorectal cancer; or a combined disease set comprising: lung cancer and prostate cancer and colorectal cancer; or a combined disease set comprising: lung

25 cancer and prostate cancer and colorectal cancer and pancreatic cancer; or a combined disease set comprising: lung cancer and breast cancer and colorectal cancer and pancreatic cancer; or a combined disease set comprising: lung cancer and breast cancer and colorectal cancer and pancreatic cancer and ovarian cancer; or a combined disease set comprising: lung cancer and breast cancer and colorectal cancer and pancreatic cancer and ovarian cancer and uterine

30 cancer; or a combined disease set comprising: prostate cancer and colorectal cancer and pancreatic cancer; or a combined disease set comprising: breast cancer and colorectal cancer and pancreatic cancer and ovarian cancer; or a combined disease set comprising: colorectal cancer and pancreatic cancer; or a combined disease set comprising: colorectal cancer and pancreatic cancer and ovarian cancer; or a combined disease set comprising: colorectal cancer and pancreatic cancer and ovarian cancer and uterine; optionally any preceeding combined

35 disease set may further comprise a diagnostic, and/or prognostic, and/or risk-stratification and/or risk-estimation readout and/or measurement for any cancer (i.e. a diagnostic, and/or prognostic, and/or risk-stratification and/or risk-estimation readout and/or measurement for any cancer of any type and/or any stage ((and/or any combined disease set comprising any two or more cancers),

wherein the specific cancer (i.e. the specific type of cancer, such as the specific type of cancer within a combined disease set) is not known and/or not diagnosed).

Optionally, any method of analysing a sample comprising one or more circulating microparticle(s) (and/or a sample derived from one or more circulating microparticles), may comprise a diagnostic, and/or prognostic, and/or risk-stratification and/or risk-estimation readout and/or measurement for any one or more cancers or pre-malignant conditions (such as any combined disease set comprising any two or more cancers), wherein said diagnostic, and/or prognostic, and/or risk-stratification and/or risk-estimation readout and/or measurement comprises an estimate of cancer or pre-cancer stage and/or grade (such as stage 1, 2, 3, or 4), and/or a measure of aggressiveness, and/or a measurement or prediction or prognosis of metastasis (and/or a risk or likelihood of metastasis) or metastatic potential.

Optionally, any method of analysing a sample comprising one or more circulating microparticle(s) (and/or a sample derived from one or more circulating microparticles), comprising a diagnostic, and/or prognostic, and/or risk-stratification and/or risk-estimation readout and/or measurement for any cancer of any type and/or any stage (and/or any combined disease set comprising any two or more cancers), wherein the specific cancer (i.e. the specific type of cancer) is not known and/or not diagnosed, may further comprise a 'cancer-ranking' process, wherein said ranking process comprises creation of an ordered list of the individual diseases comprised within a combined diseased set (such as a combined disease set comprising: lung cancer and prostate cancer and colorectal cancer and pancreatic cancer; or a combined disease set comprising: lung cancer and breast cancer and colorectal cancer and pancreatic cancer; or a combined disease set comprising: lung cancer and breast cancer and colorectal cancer and pancreatic cancer and ovarian cancer). Optionally, said ranking process may comprise a process wherein said individual diseases are ordered based upon one or more pairwise comparisons (i.e. individual disease-to-individual disease comparisons), wherein each such more pairwise comparisons evaluates which of the two individual diseases is more likely and/or more severe (e.g. based upon said analysing a sample comprising one or more circulating microparticle(s) (and/or a sample derived from one or more circulating microparticles)).

Optionally, any method of analysing a sample comprising one or more circulating microparticle(s) (and/or a sample derived from one or more circulating microparticles), comprising a diagnostic, and/or prognostic, and/or risk-stratification and/or risk-estimation readout and/or measurement may comprise an estimate and/or readout of a likelihood of death from any one or more diseases (such as an estimate and/or readout of a likelihood of death from any cancer, and/or from any specific cancer, and/or from (one of) any two or more different specific cancers (e.g. any two or more different specific cancers comprised within any combined disease set comprising two or more different specific cancers); optionally, any method of generating an such estimate and/or

- readout of a likelihood of death may be configured to estimate and/or readout a likelihood of death within a specific period of time from the time at which said sample was taken from a person; optionally such specific period of time may comprise any one or more of the following: 3 months, 6 months, 9 months, 12 months, 18 months, 2 years, 3 years, 4 years, 5 years, 6 years, 8 years, 10 years, 12 years, 15 years, 20 years, 25 years, 30 years, 35 years, 40 years, and/or 50 years; optionally, any method of generating an such estimate and/or readout of a likelihood of death (such as within a specific period of time) may be configured to estimate and/or readout a likelihood of death in the event that the associated disease (i.e. the associated disease providing a likelihood of death) remains untreated (i.e. wherein the patient is not treated with therapy and/or surgery for said disease); optionally, any method of generating an such estimate and/or readout of a likelihood of death (such as within a specific period of time) may be configured to estimate and/or readout a likelihood of death in the event that the associated disease (i.e. the associated disease providing a likelihood of death) is treated (i.e. wherein the patient is treated with therapy and/or surgery for said disease); optionally, any likelihood of death from a disease calculated based upon a patient receiving treatment for said disease may be compared with the associated likelihood of death from said disease calculated based upon the patient not receiving treatment for said disease (e.g. said likelihoods may be divided by one or another, e.g. to calculate or estimate an expected or potential survival benefit in the event the patient is treated for said disease).
- Any combined disease set may comprise a combined foetal genetic disease set, for example a combined foetal genetic disease set comprising: Down Syndrome and Patau Syndrome in a developing foetus; or a combined foetal genetic disease set comprising: Down Syndrome and Edwards Syndrome in a developing foetus; or a combined foetal genetic disease set comprising: Down Syndrome and Patau Syndrome and Edwards Syndrome in a developing foetus; or a combined foetal genetic disease set comprising: Down Syndrome and Patau Syndrome and Edwards Syndrome and trisomy of chromosome 9 in a developing foetus; or a combined foetal genetic disease set comprising: Down Syndrome and Patau Syndrome and Edwards Syndrome and trisomy of chromosome 9 in a developing foetus and one or more microdeletion syndromes; or or a combined foetal genetic disease set comprising: Down Syndrome and Patau Syndrome and Edwards Syndrome and trisomy of chromosome 9 in a developing foetus and one or more microdeletion syndromes and one or more foetal monogenic diseases or foetal monogenic disease predispositions (such as Thalassaemia, and/or sickle cell anaemia, and/or haemophilia, and/or Tay Sachs disease, and/or cystic fibrosis, and/or Huntington's disease, and/or fragile-X syndrome, and/or any combination of at least two, at least three, or at least four members thereof).

Optionally, in any methods of analysing a sample comprising one or more circulating microparticle(s) (and/or a sample derived from one or more circulating microparticles), any measurements of any two or more biomolecules from any circulating microparticle(s), and/or any

two or more linked signals corresponding to any such measurement(s), may be used to identify and/or predict circulating microparticle(s) (and/or any set of two or more linked signals associated with and/or derived from any such circulating microparticle(s)) that are derived from tissues and/or cells associated with any one or more of the conditions and/or diseases and/or tissue types disclosed previously and/or herein. Optionally, in any methods of analysing a sample comprising one or more circulating microparticle(s) (and/or a sample derived from one or more circulating microparticles), one or more parameter values may be used to identify and/or predict circulating microparticle(s) that are derived from tissues and/or cells associated with any one or more of the conditions and/or diseases and/or tissue types disclosed previously and/or herein; for example, any one or more parameter values may be compared to one or more control parameter values, wherein any such parameter being above a particular specific control parameter value, below a particular specific control parameter value, within a specific range of control parameter values, and/or outside of a specific range of control parameter values indicates and/or predicts and/or estimates the tissue and/or cell type from which the associated circulating microparticle(s) (and/or the associated set of two or more linked signals associated with and/or derived from such circulating microparticle(s)) are derived. Optionally, any such method of identifying tissue and/or cell type associated with circulating microparticle(s) and/or associated with a linked set of signals may further comprise counting the total number (and/or proportion) of all linked sets of signals (and/or the total number of circulating microparticle(s)) identified and/or predicted to derive from any (and/or all) particular tissue and/or cell type; optionally said total number (and/or said proportion) may be compared within one or more threshold number(s) and/or ranges, wherein any such total number (and/or proportion) being above a particular threshold number, below a particular threshold number, within a specific range of threshold numbers, and/or outside of a specific range of threshold numbers indicates and/or predicts and/or estimates and/or provides a diagnosis, prognosis, risk estimate or likelihood and/or risk factor and/or risk category for any disease, condition, or syndrome.

37. LIBRARIES AND KITS FOR PERFORMING THE METHODS OF THE INVENTION

The invention further provides libraries comprising one or more of the reagents defined herein. The invention also provides libraries specifically adapted for performing any of the methods defined herein.

The invention further provides kits comprising one or more of the components defined herein. The invention also provides kits specifically adapted for performing any of the methods defined herein.

Kits for labelling a target nucleic acid are described in PCT/GB2017/053820, which is incorporated herein by reference.

- 200 -

The invention further provides a kit for labelling a target nucleic acid molecule and a target biomolecule, wherein the kit comprises a multimeric barcoding reagent as defined herein and a barcoded affinity probe as defined herein. Preferably, the target biomolecule is a non-nucleic acid target biomolecule (e.g. a target polypeptide).

The invention further provides a kit for labelling a target nucleic acid molecule and a target biomolecule, wherein the kit comprises: (a) a multimeric barcoding reagent, wherein the multimeric barcoding reagent comprises first and second barcode regions linked together, wherein each barcode region comprises a nucleic acid sequence; and (b) a barcoded affinity probe, wherein the barcoded affinity probe comprises at least one affinity moiety linked to a barcoded oligonucleotide, wherein the barcoded oligonucleotide comprises at least one nucleotide, and wherein the affinity moiety is capable of binding to the target biomolecule.

The invention further provides a kit for labelling a target nucleic acid and a target biomolecule, wherein the kit comprises: (a) a multimeric barcoding reagent comprising (i) first and second barcode molecules linked together (i.e. a multimeric barcode molecule), wherein each of the barcode molecules comprises a nucleic acid sequence comprising, optionally in the 5' to 3' direction, an adapter region and a barcode region, and (ii) first and second barcoded oligonucleotides, wherein the first barcoded oligonucleotide comprises a barcode region annealed to the barcode region of the first barcode molecule, and wherein the second barcoded oligonucleotide comprises a barcode region annealed to the barcode region of the second barcode molecule; and (b) first and second adapter oligonucleotides, wherein the first adapter oligonucleotide comprises, optionally in the 5' to 3' direction, an adapter region capable of annealing to the adapter region of the first barcode molecule and a target region capable of annealing or ligating to a first fragment of the target nucleic acid, and wherein the second adapter oligonucleotide comprises, optionally in the 5' to 3' direction, an adapter region capable of annealing to the adapter region of the second barcode molecule and a target region capable of annealing or ligating to a second fragment of the target nucleic acid; and (c) a barcoded affinity probe, wherein the barcoded affinity probe comprises at least one affinity moiety linked to a barcoded oligonucleotide, wherein the barcoded oligonucleotide comprises at least one nucleotide, and wherein the affinity moiety is capable of binding to the target biomolecule.

The kits may comprise an affinity probe instead of (or in addition to) a barcoded affinity probe. The affinity probe may take any of the forms described herein. The affinity probe may comprise at least one affinity moiety, wherein the affinity moiety is capable of binding to the target biomolecule.

- 201 -

The target regions of each adapter oligonucleotide may comprise different sequences. Each target region may comprise a sequence capable of annealing to only a single fragment of a target nucleic acid within a sample of nucleic acids. Each target region may comprise one or more random, or one or more degenerate, sequences to enable the target region to anneal to more than one fragment of a target nucleic acid. Each target region may comprise at least 5, at least 10, at least 15, at least 20, at least 25, at least 50 or at least 100 nucleotides. Preferably, each target region comprises at least 5 nucleotides. Each target region may comprise 5 to 100 nucleotides, 5 to 10 nucleotides, 10 to 20 nucleotides, 20 to 30 nucleotides, 30 to 50 nucleotides, 50 to 100 nucleotides, 10 to 90 nucleotides, 20 to 80 nucleotides, 30 to 70 nucleotides or 50 to 60 nucleotides. Preferably, each target region comprises 30 to 70 nucleotides. Preferably each target region comprises deoxyribonucleotides, optionally all of the nucleotides in a target region are deoxyribonucleotides. One or more of the deoxyribonucleotides may be a modified deoxyribonucleotide (e.g. a deoxyribonucleotide modified with a biotin moiety or a deoxyuracil nucleotide). Each target region may comprise one or more universal bases (e.g. inosine), one or modified nucleotides and/or one or more nucleotide analogues.

The target regions may be used to anneal the adapter oligonucleotides to fragments of target nucleic acids, and then may be used as primers for a primer-extension reaction or an amplification reaction e.g. a polymerase chain reaction. Alternatively, the target regions may be used to ligate the adapter oligonucleotides to fragments of target nucleic acids. The target region may be at the 5' end of an adapter oligonucleotide. Such a target region may be phosphorylated. This may enable the 5' end of the target region to be ligated to the 3' end of a fragment of a target nucleic acid.

The adapter oligonucleotides may comprise a linker region between the adapter region and the target region. The linker region may comprise one or more contiguous nucleotides that are not annealed to the first and second barcode molecules (i.e. the multimeric barcode molecule) and are non-complementary to the fragments of the target nucleic acid. The linker may comprise 1 to 100, 5 to 75, 10 to 50, 15 to 30 or 20 to 25 non-complementary nucleotides. Preferably, the linker comprises 15 to 30 non-complementary nucleotides. The use of such a linker region enhances the efficiency of the barcoding reactions performed using the kits described herein.

Each of the components of the kit may take any of the forms defined herein.

The components may be provided in the kit as physically separated components.

The kit may comprise: (a) a multimeric barcoding reagent comprising at least 5, at least 10, at least 20, at least 25, at least 50, at least 75 or at least 100 barcode molecules linked together, wherein each barcode molecule is as defined herein; and (b) an adapter oligonucleotide capable

- 202 -

of annealing to each barcode molecule, wherein each adapter oligonucleotide is as defined herein.

Figure 2 shows a kit comprising a multimeric barcoding reagent and adapter oligonucleotides for labelling a target nucleic acid. In more detail, the kit comprises first (D1, E1, and F1) and second (D2, E2, and F2) barcode molecules, with each incorporating a barcode region (E1 and E2) and also a 5' adapter region (F1 and F2). These first and second barcode molecules are linked together, in this embodiment by a connecting nucleic acid sequence (S).

The kit further comprises first (A1 and B1) and second (A2 and B2) barcoded oligonucleotides, which each comprise a barcode region (B1 and B2), as well as 5' regions (A1 and A2). The 5' region of each barcoded oligonucleotide is complementary to, and thus may be annealed to, the 3' regions of the barcode molecules (D1 and D2). The barcode regions (B1 and B2) are complementary to, and thus may be annealed to, the barcode regions (E1 and E2) of the barcode molecules.

The kit further comprises first (C1 and G1) and second (C2 and G2) adapter oligonucleotides, wherein each adapter oligonucleotide comprises an adapter region (C1 and C2) that is complementary to, and thus able to anneal to, the 5' adapter region of a barcode molecule (F1 and F2). These adapter oligonucleotides may be synthesised to include a 5'-terminal phosphate group. Each adapter oligonucleotide also comprises a target region (G1 and G2), which may be used to anneal the barcoded-adapter oligonucleotides (A1, B1, C1 and G1, and A2, B2, C2 and G2) to target nucleic acids, and then may be used as primers for a primer-extension reaction or a polymerase chain reaction.

The kit may comprise a library of two or more multimeric barcoding reagents, wherein each multimeric barcoding reagent is as defined herein, and adapter oligonucleotides for each of the multimeric barcoding reagents, wherein each adapter oligonucleotide is as defined herein. The barcode regions of the first and second barcoded oligonucleotides of the first multimeric barcoding reagent are different to the barcode regions of the first and second barcoded oligonucleotides of the second multimeric barcoding reagent.

The kit may comprise a library comprising at least 5, at least 10, at least 20, at least 25, at least 50, at least 75, at least 100, at least 250, at least 500, at least 10^3 , at least 10^4 , at least 10^5 , at least 10^6 , at least 10^7 , at least 10^8 or at least 10^9 multimeric barcoding reagents as defined herein. Preferably, the kit comprises a library comprising at least 10 multimeric barcoding reagents as defined herein. The kit may further comprise adapter oligonucleotides for each of the multimeric barcoding reagents, wherein each adapter oligonucleotide may take the form of any of the adapter oligonucleotides defined herein. Preferably, the barcode regions of the first and second

- 203 -

barcoded oligonucleotides of each multimeric barcoding reagent are different to the barcode regions of the barcoded oligonucleotides of at least 9 other multimeric barcoding reagents in the library.

- 5 The barcode regions of the first and second barcoded oligonucleotides of each multimeric barcoding reagent may be different to the barcode regions of the barcoded oligonucleotides of at least 4, at least 9, at least 19, at least 24, at least 49, at least 74, at least 99, at least 249, at least 499, at least 999 (i.e. 10^3-1), at least 10^4-1 , at least 10^5-1 , at least 10^6-1 , at least 10^7-1 , at least 10^8-1 or at least 10^9-1 other multimeric barcoding reagents in the library. The barcode regions of
10 the first and second barcoded oligonucleotides of each multimeric barcoding reagent may be different to the barcode regions of the barcoded oligonucleotides of all of the other multimeric barcoding reagents in the library. Preferably, the barcode regions of the first and second barcoded oligonucleotides of each multimeric barcoding reagent are different to the barcode regions of the barcoded oligonucleotides of at least 9 other multimeric barcoding reagents in the
15 library.

- The barcode regions of the barcoded oligonucleotides of each multimeric barcoding reagent may be different to the barcode regions of the barcoded oligonucleotides of at least 4, at least 9, at least 19, at least 24, at least 49, at least 74, at least 99, at least 249, at least 499, at least 999 (i.e. 10^3-1), at least 10^4-1 , at least 10^5-1 , at least 10^6-1 , at least 10^7-1 , at least 10^8-1 or at least 10^9-1
20 other multimeric barcoding reagents in the library. The barcode regions of the barcoded oligonucleotides of each multimeric barcoding reagent may be different to the barcode regions of the barcoded oligonucleotides of all of the other multimeric barcoding reagents in the library. Preferably, the barcode regions of the barcoded oligonucleotides of each multimeric barcoding reagent are different to the barcode regions of the barcoded oligonucleotides of at least 9 other
25 multimeric barcoding reagents in the library.

- Preferably, the kit comprises at least two different barcoded affinity probes, wherein each of at least two of the at least two different barcoded affinity probes are capable of binding to a different
30 target biomolecule.

The invention is further defined in the following set of numbered clauses:

1. A method of analysing a sample comprising a circulating microparticle or a sample derived
35 from a circulating microparticle, and wherein the method comprises:
(a) contacting the sample with a barcoded affinity probe, wherein the barcoded affinity probe comprises at least one affinity moiety linked to a barcoded oligonucleotide, wherein the barcoded oligonucleotide comprises at least one nucleotide, and wherein the affinity moiety is capable of binding to a target biomolecule;

- 204 -

- (b) forming a reaction mixture, wherein the step of forming the reaction mixture comprises binding the affinity moiety to the target molecule, if present, to form a barcoded biomolecule complex comprising the barcoded affinity probe and the target biomolecule; and
 - 5 (c) determining the presence, absence and/or level of the target biomolecule in the sample by measuring the presence, absence and/or level of the barcoded oligonucleotide in the reaction mixture.
- 2. A method of analysing a sample comprising a circulating microparticle or a sample derived from a circulating microparticle, wherein the circulating microparticle comprises a target
10 biomolecule, and wherein the method comprises:
 - (a) contacting the sample with a barcoded affinity probe, wherein the barcoded affinity probe comprises at least one affinity moiety linked to a barcoded oligonucleotide, wherein the barcoded oligonucleotide comprises at least one nucleotide, and wherein the affinity moiety is capable of binding to the target biomolecule;
 - 15 (b) forming a reaction mixture, wherein the step of forming the reaction mixture comprises binding the affinity moiety to the target biomolecule to form a barcoded biomolecule complex comprising the barcoded affinity probe and the target biomolecule; and
 - (c) determining the level of the target biomolecule in the sample by measuring the level of the barcoded oligonucleotide in the reaction mixture.
- 20 3. A method of analysing a sample comprising a circulating microparticle or a sample derived from a circulating microparticle, and wherein the method comprises:
 - (a) contacting the sample with at least one affinity moiety, and wherein the affinity moiety is capable of binding to a target biomolecule;
 - (b) forming a reaction mixture, wherein the step of forming the reaction mixture comprises (i)
25 binding the affinity moiety to the target biomolecule, if present, and (ii) contacting the sample with a barcoded oligonucleotide and linking the barcoded oligonucleotide to the affinity moiety to form a barcoded biomolecule complex comprising a barcoded affinity probe and the target biomolecule, wherein the barcoded affinity probe comprises at least one affinity moiety linked to the barcoded oligonucleotide, and wherein the barcoded
30 oligonucleotide comprises at least one nucleotide; and
 - (c) determining the presence, absence and/or level of the target biomolecule in the sample by measuring the presence, absence and/or level of the barcoded oligonucleotide in the reaction mixture.
- 4. A method of analysing a sample comprising a circulating microparticle or a sample derived from a circulating microparticle, wherein the circulating microparticle comprises a target
35 biomolecule, and wherein the method comprises:
 - (a) contacting the sample with at least one affinity moiety, and wherein the affinity moiety is capable of binding to the target biomolecule;

– 205 –

- (b) forming a reaction mixture, wherein the step of forming the reaction mixture comprises (i) binding the affinity moiety to the target biomolecule and (ii) contacting the sample with a barcoded oligonucleotide and linking the barcoded oligonucleotide to the affinity moiety to form a barcoded biomolecule complex comprising a barcoded affinity probe and the target biomolecule, wherein the barcoded affinity probe comprises at least one affinity moiety linked to the barcoded oligonucleotide, and wherein the barcoded oligonucleotide comprises at least one nucleotide; and
- (c) determining the level of the target biomolecule in the sample by measuring the level of the barcoded oligonucleotide in the reaction mixture.
5. The method of any one of clauses 1-4, wherein the sample is contacted with at least two different barcoded affinity probes.
6. The method of any one of clauses 1-5, wherein the barcoded affinity probe comprises an aptamer, optionally wherein the barcoded affinity probe is an aptamer.
7. The method of any one of clauses 1-5, wherein the affinity moiety is an antibody or an aptamer.
8. The method of any one of clauses 1-7, wherein the barcoded affinity probe comprises at least two affinity moieties.
9. The method of any one of clauses 1-8, wherein the barcoded affinity probe comprises at least two different barcoded oligonucleotides.
10. The method of any one of clauses 1-9, wherein the barcoded oligonucleotide comprises a barcode sequence associated with and/or identifying of the affinity moiety to which it is linked.
11. The method of any one of clauses 1-10, wherein the barcoded oligonucleotide comprises a barcode sequence of at least 2, at least 3, at least 5, at least 10, at least 20, or at least 30 nucleotides.
12. The method of any one of clauses 1-11, wherein step (c) comprises analysing a nucleotide sequence of the barcoded oligonucleotide, optionally wherein the sequence is analysed by sequencing or PCR.
13. The method of any one of clauses 1-12, wherein step (b) or step (c) comprises linking together at least two barcoded biomolecule complexes of the first circulating microparticle and linking together at least two barcoded biomolecule complexes of the second circulating microparticle.
14. The method of any one of clauses 1-13, wherein the target biomolecule is a polypeptide or a fragment of a target nucleic acid.
15. The method of any one of clauses 1-14, wherein the sample further comprises a fragment of target nucleic acid of the microparticle.
16. The method of any one of clauses 1-15, wherein the sample comprises a first circulating microparticle and a second circulating microparticle, or wherein the sample is derived from a first circulating microparticle and a second circulating microparticle, wherein step (b) comprises forming at least one barcoded biomolecule complex comprising a barcoded affinity

- 206 -

probe and a target biomolecule of the first circulating microparticle, and forming at least one barcoded biomolecule complex comprising a barcoded affinity probe and a target biomolecule of the second circulating microparticle.

17. The method of any one of clauses 1-16, wherein the sample further comprises a fragment of a target nucleic acid of the first circulating microparticle and a fragment of a target nucleic acid of the second circulating microparticle.
18. The method of clause 16 or clause 17, wherein step (c) comprises: (i) contacting the reaction mixture with a library comprising at least two multimeric barcoding reagents, wherein each multimeric barcoding reagent comprises first and second barcode regions linked together, wherein each barcode region comprises a nucleic acid sequence and wherein the first and second barcode regions of a first multimeric barcoding reagent are different to the first and second barcode regions of a second multimeric barcoding reagent of the library; and (ii) appending barcode sequences to each of a first fragment of a target nucleic acid and a second fragment of a target nucleic acid of the first microparticle to produce first and second barcoded target nucleic acid molecules for the first microparticle, wherein the first barcoded target nucleic acid molecule comprises the nucleic acid sequence of the first barcode region of the first multimeric barcoding reagent and the second barcoded target nucleic acid molecule comprises the nucleic acid sequence of the second barcode region of the first multimeric barcoding reagent, and appending barcode sequences to each of a first fragment of a target nucleic acid and a second fragment of a target nucleic acid of the second microparticle to produce first and second barcoded target nucleic acid molecules for the second microparticle, wherein the first barcoded target nucleic acid molecule comprises the nucleic acid sequence of the first barcode region of the second multimeric barcoding reagent and the second barcoded target nucleic acid molecule comprises the nucleic acid sequence of the second barcode region of the second multimeric barcoding reagent.
19. The method of clause 18, wherein the first fragment of a target nucleic acid of the first microparticle is the barcoded oligonucleotide of the at least one barcoded biomolecule complex of the first circulating microparticle, and wherein the first fragment of a target nucleic acid of the second microparticle is the barcoded oligonucleotide of the at least one barcoded biomolecule complex of the second circulating microparticle.
20. The method of clause 18 or clause 19, wherein the reaction mixture further comprises a fragment of a target nucleic acid of the first circulating microparticle and wherein the second fragment of a target nucleic acid of the first circulating microparticle is the fragment of the target nucleic acid of the first circulating microparticle.
21. The method of any one of clauses 18-20, wherein the reaction mixture further comprises a fragment of a target nucleic acid of the second circulating microparticle and wherein the second fragment of a target nucleic acid of the second circulating microparticle is the fragment of the target nucleic acid of the second circulating microparticle.

22. The method of any one of clauses 18-21, wherein the step of contacting the reaction mixture with a library of multimeric barcoding reagents is performed in a single contiguous aqueous volume.
23. The method of any one of clauses 18-22, wherein step (c) is performed in a single contiguous aqueous volume, optionally wherein steps (b) and (c) are performed in a single contiguous aqueous volume, optionally wherein steps (a), (b) and (c) are performed in a single contiguous aqueous volume
24. The method of clause 16 or clause 17, wherein the method further comprises partitioning the sample or reaction mixture into at least first and second partitions and analysing the nucleotide sequences of the barcoded oligonucleotides in each of the first and second partitions, wherein the first partition comprises at least one barcoded oligonucleotide comprised in or derived from the at least one barcoded biomolecule complex of the first circulating microparticle, and wherein the second partition comprises at least one barcoded oligonucleotide comprised in or derived from the at least one barcoded biomolecule complex of the second circulating microparticle.
25. The method of clause 24, wherein the step of analysing the nucleotide sequences of the barcoded oligonucleotides of the barcoded biomolecule complexes comprises: (i) appending a first partition barcode sequence to the at least one barcoded oligonucleotide of the first partition; (ii) appending a second partition barcode sequence to the at least one barcoded oligonucleotide of the second partition.
26. The method of clause 25, wherein said first and second partition barcode sequences are different.
27. The method of clause 25, wherein the first partition barcode sequence is from a first set of partition barcode sequences, and the second partition barcode sequence is from a second set of partition barcode sequences, and wherein the first and second sets of partition barcode sequences are different.
28. The method of any one of clauses 25-27, wherein said first partition barcode sequence is the nucleic acid sequence of a barcode region of a first multimeric barcoding reagent, and the second partition barcode sequence is the nucleic acid sequence of a second multimeric barcoding reagent, and wherein the first and second multimeric barcoding reagents each comprise two or more barcode regions linked together.
29. The method of any one of clauses 25-28, wherein the first partition further comprises a fragment of a target nucleic acid of the first circulating microparticle, and wherein the second partition further comprises a fragment of a target nucleic acid of the second circulating microparticle.
30. The method of clause 29, wherein the step of analysing the nucleotide sequences of the barcoded oligonucleotides of the barcoded biomolecule complexes comprises: (i) appending a first partition barcode sequence to at least one barcoded oligonucleotide of the first partition and appending the first partition barcode sequence to at least one fragment of a target nucleic

- acid of the first circulating microparticle; (ii) appending a second partition barcode sequence to at least one barcoded oligonucleotide of the second partition and appending the second partition barcode sequence to at least one fragment of a target nucleic acid of the second circulating microparticle; and wherein said first and second partition barcode sequences are different.
- 5
31. The method of clause 29, wherein the step of analysing the nucleotide sequences of the barcoded oligonucleotides of the barcoded biomolecule complexes comprises: (i) appending a first partition barcode sequence of a first set of partition barcode sequences to at least one barcoded oligonucleotide of the first partition and appending a second partition barcode
- 10
- sequence of the first set of partition barcode sequences to at least one fragment of a target nucleic acid of the first circulating microparticle; and (ii) appending a first partition barcode sequence of a second set of partition barcode sequences to at least one barcoded oligonucleotide of the second partition and appending a second partition barcode sequence of the second set of partition barcode sequences to at least one fragment of a target nucleic acid
- 15
- of the second circulating microparticle; and wherein the first and second sets of partition barcode sequences are different.
32. The method of clause 31, wherein the first and second partition barcode sequences of the first set of partition barcode sequences are the nucleic acid sequences of first and second barcode regions of a first multimeric barcoding reagent, and wherein the first and second
- 20
- partition barcode sequences of the second set of partition barcode sequences are the nucleic acid sequences of first and second barcode regions of a second multimeric barcoding reagent, and wherein the first and second multimeric barcoding reagents each comprise two or more barcode regions linked together.
33. Use of a barcoded affinity probe to determine the presence, absence and/or level of a target
- 25
- biomolecule in a circulating microparticle or in a sample derived therefrom, wherein the barcoded affinity probe comprises at least one affinity moiety linked to a barcoded oligonucleotide, wherein the barcoded oligonucleotide comprises at least one nucleotide and wherein the affinity moiety is capable of binding to the target biomolecule.
34. A barcoded affinity probe for determining the presence, absence and/or level of a target
- 30
- biomolecule, wherein the barcoded affinity probe comprises at least one affinity moiety linked to a barcoded oligonucleotide, wherein the barcoded oligonucleotide comprises at least one nucleotide and wherein the affinity moiety is capable of binding to the target biomolecule.
35. A library of barcoded affinities probes for determining the presence, absence and/or level of at least two target biomolecules, wherein the library comprises: (i) a first barcoded affinity probe comprising at least one affinity moiety linked to a barcoded oligonucleotide, wherein the
- 35
- barcoded oligonucleotide comprises at least one nucleotide and wherein the affinity moiety is capable of binding to a first target biomolecule; and (ii) a second barcoded affinity probe comprising at least one affinity moiety linked to a barcoded oligonucleotide, wherein the barcoded oligonucleotide comprises at least one nucleotide and wherein the affinity moiety is

capable of binding to a second target biomolecule; and wherein the first target biomolecule and the second target biomolecule are different.

36. A method of analysing a sample comprising at least two circulating microparticles or a sample derived from at least two circulating microparticles, wherein the method comprises: (i) partitioning the sample into at least two partitions, wherein each partition comprises, on average, less than n circulating microparticles; and (ii) determining the presence, absence and/or level of at least two target biomolecules in each of at least two of the at least two partitions. Optionally, wherein n is 1000, 500, 200, 100, 50, 40, 30, 20, 10, 5, 4, 3, 2, 1, 0.5, 0.4, 0.3, 0.2, 0.1, 0.05, 0.04, 0.03, 0.02, 0.01, 0.005, 0.001, 0.0005, or 0.0001.
37. A method of analysing a sample comprising at least two circulating microparticles or a sample derived from at least two circulating microparticles, wherein the method comprises: (i) partitioning the sample into at least two partitions, wherein a first partition comprises at least first and second target biomolecules of a first circulating microparticle and a second partition comprises at least first and second target biomolecules of a second circulating microparticle, and wherein each partition comprises, on average, less than $[X]$ total mass of DNA; and (ii) determining the presence, absence and/or level of at least two target biomolecules in each of at least two of the at least two partitions. Optionally, wherein $[X]$ is 1.0 attogram of DNA, 10 attograms of DNA, 100 attograms of DNA, 1.0 femtogram of DNA, 10 femtograms of DNA, 100 femtograms of DNA, 1.0 picogram of DNA, 10 picograms of DNA, 100 picograms of DNA, or 1.0 nanogram of DNA.
38. A method of analysing a sample comprising at least two circulating microparticles or a sample derived from at least two circulating microparticles, wherein the method comprises: (i) partitioning the sample into at least two partitions, wherein a first partition comprises at least first and second target biomolecules of a first circulating microparticle and a second partition comprises at least first and second target biomolecules of a second circulating microparticle, and wherein each partition comprises, on average, less than $[Y]$ total mass of protein; and (ii) determining the presence, absence and/or level of at least two target biomolecules in each of at least two of the at least two partitions. Optionally, wherein $[Y]$ is 1.0 attogram of protein, 10 attograms of protein, 100 attograms of protein, 1.0 femtogram of protein, 10 femtograms of protein, 100 femtograms of protein, 1.0 picogram of protein, 10 picograms of protein, 100 picograms of protein, or 1.0 nanogram of protein.
39. The method of any one of clauses 36-38, wherein the method further comprises analysing the sequence of at least two target nucleic acid molecules that have been partitioned into each of said first and second partitions.

The invention is further defined in the following set of numbered clauses:

1. A method of analysing a sample comprising a circulating microparticle or a sample derived from a circulating microparticle, wherein the circulating microparticle comprises at least three

- 210 -

- target molecules, wherein at least two of the target molecules are fragments of genomic DNA and at least one of the target molecules is a target polypeptide, and wherein the method comprises measuring a signal corresponding to the presence, absence and/or level of each of the target molecules to produce a set of at least two linked signals for the circulating microparticle, wherein at least one of the linked signals corresponds to the presence, absence and/or level of the fragments of genomic DNA in the sample and at least one of the linked signals corresponds to the presence, absence and/or level of the target polypeptide in the sample.
- 5
- 10
- 15
- 20
- 25
- 30
- 35
2. The method of clause 1, wherein the fragments of genomic DNA comprise a specific sequence of nucleotides and/or wherein the fragments of genomic DNA comprise at least one modified nucleotide or nucleobase, optionally wherein the modified nucleotide or nucleobase is 5-methylcytosine or 5-hydroxy-methylcytosine.
 3. The method of clause 1 or clause 2, wherein the target polypeptide comprises a specific amino acid sequence and/or wherein the target polypeptide comprises a post-translational modification, optionally wherein the target polypeptide comprises an acetylated amino acid residue and/or a methylated amino acid residue.
 4. The method of any one of clauses 1-3, wherein the method comprises measuring the signal corresponding to the presence, absence and/or level of each of the target molecules of the circulating microparticle to produce a set of at least three linked signals for the circulating microparticle, wherein one of the linked signals corresponds to the presence, absence and/or level of a first fragment of genomic DNA of the circulating microparticle, one of the linked signals corresponds to the presence, absence and/or level of a second fragment of genomic DNA of the circulating microparticle, and one of the linked signals corresponds to the presence, absence and/or level of the target polypeptide of the circulating microparticle.
 5. The method of any one of clauses 1-4, wherein the step of measuring a signal corresponding to the presence, absence and/or level of the fragments of genomic DNA comprises analysing a sequence of each of at least two of the at least two fragments of genomic DNA, optionally wherein the step of measuring a signal corresponding to the presence, absence and/or level of the fragments of genomic DNA comprises sequencing at least a portion of each of at least two of the at least two fragments of genomic DNA.
 6. The method of any one of clauses 1-5, wherein the step of measuring a signal corresponding to the presence, absence and/or level of the fragments of genomic DNA comprises:
(a) linking at least two of the at least two fragments of genomic DNA to produce a set of at least two linked fragments of genomic DNA; and, optionally,

- 211 -

- (b) sequencing at least a portion of each of at least two of the linked fragments in the set to produce at least two linked sequence reads.
7. The method of any one of clauses 1-6, wherein the step of measuring a signal corresponding to the presence, absence and/or level of the fragments of genomic DNA comprises:
- 5 (a) appending each of at least two of the at least two fragments of genomic DNA of the circulating microparticle to a barcode sequence to produce a set of linked fragments of genomic DNA; and, optionally,
- 10 (b) sequencing at least a portion of each of at least two of the linked fragments in the set to produce at least two linked sequence reads, wherein the at least two linked sequence reads are linked by the barcode sequence.
8. The method of any one of clauses 1-6, wherein the step of measuring a signal corresponding to the presence, absence and/or level of the fragments of genomic DNA comprises:
- 15 (a) appending each of at least two of the at least two fragments of genomic DNA of the circulating microparticle to a different barcode sequence of a set of barcode sequences to produce a set of linked fragments of genomic DNA; and, optionally,
- 20 (b) sequencing at least a portion of each of at least two of the linked fragments in the set to produce at least two linked sequence reads, wherein the at least two linked sequence reads are linked by the set of barcode sequences.
9. The method of any one of clauses 1-8, wherein the fragments of genomic DNA comprise at least one modified nucleotide or nucleobase and wherein the step of measuring a signal corresponding to the presence, absence and/or level of the fragments of genomic DNA
- 25 comprises measuring a signal corresponding to the presence, absence and/or level of the modified nucleotide or nucleobase of the fragments of genomic DNA, optionally wherein the modified nucleotide or nucleobase is 5-methylcytosine or 5-hydroxy-methylcytosine.
10. The method of clause 9, wherein the signal corresponding to the presence, absence and/or level of the modified nucleotide or nucleobase is measured using (i) a barcoded affinity probe, wherein the barcoded affinity probe comprises at least one affinity moiety linked to a barcoded oligonucleotide, wherein the barcoded oligonucleotide comprises at least one nucleotide, and wherein the affinity moiety is capable of binding to the modified nucleotide or nucleobase, optionally wherein the signal is measured by determining the presence, absence
- 30 and/or level of the barcoded oligonucleotide by sequencing; and/or (ii) an optically-labelled affinity probe and/or a fluorescently-labelled affinity probe, optionally wherein the signal is measured by flow cytometry and/or fluorescence-activated cell sorting.
- 35

11. The method of any one of clauses 1-10, wherein the signal corresponding to the presence, absence and/or level of the target polypeptide is measured using (i) a barcoded affinity probe, wherein the barcoded affinity probe comprises at least one affinity moiety linked to a barcoded oligonucleotide, wherein the barcoded oligonucleotide comprises at least one nucleotide, and wherein the affinity moiety is capable of binding to the target polypeptide, optionally wherein the signal is measured by determining the presence, absence and/or level of the barcoded oligonucleotide by sequencing; and/or (ii) an optically-labelled affinity probe and/or a fluorescently-labelled affinity probe, optionally wherein the signal is measured by flow cytometry and/or fluorescence-activated cell sorting.
12. The method of any one of clauses 1-11, wherein the circulating microparticle comprises at least 3, at least 4, at least 5, at least 10, at least 50, at least 100, at least 500, at least 1000, at least 5000, at least 10,000, at least 100,000, or at least 1,000,000 target molecules, and wherein the method comprises producing a set of at least 3, at least 4, at least 5, at least 10, at least 50, at least 100, at least 500, at least 1000, at least 5000, at least 10,000, at least 100,000, or at least 1,000,000 linked signals for the circulating microparticle.
13. The method of any one of clauses 1-12, wherein the target molecules comprise at least 2, at least 3, at least 4, at least 9, at least 49, at least 99, at least 499, at least 999, at least 4999, at least 9,999, at least 99,999, or at least 999,999 fragments of genomic DNA, and optionally wherein the method comprises producing a set of at least 3, at least 4, at least 5, at least 10, at least 50, at least 100, at least 500, at least 1000, at least 5000, at least 10,000, at least 100,000, or at least 1,000,000 linked signals for the circulating microparticle.
14. The method of any one of clauses 1-13, wherein the target molecules comprise at least 2, at least 3, at least 4, at least 9, at least 49, at least 99, at least 499, at least 999, at least 4999, at least 9,999, at least 99,999, or at least 999,999 target polypeptides, and optionally wherein the method comprises producing a set of at least at least 3, at least 4, at least 5, at least 10, at least 50, at least 100, at least 500, at least 1000, at least 5000, at least 10,000, at least 100,000, or at least 1,000,000 linked signals for the circulating microparticle.
15. The method of any one of clauses 1-14, wherein the sample comprises first and second circulating microparticles, wherein each circulating microparticle comprises at least three target molecules as defined in any one of clauses 1-14, and wherein the method comprises performing the step of measuring in accordance with any one of clauses 1-14 to produce a set of linked signals for the first circulating microparticle and performing the step of measuring in accordance with any one of clauses 1-14 to produce a set of linked signals for the second circulating microparticle; optionally wherein the sample comprises n circulating microparticles, wherein each circulating microparticle comprises at least three target molecules as defined in

- 213 -

any one of clauses 1-14, and wherein the method comprises performing the step of measuring in accordance with any one of clauses 1-14 for each circulating microparticle to produce a set of linked signals for each circulating microparticle, optionally wherein n is at least 3, at least 5, at least 10, at least 50, at least 100, at least 1000, at least 10,000, at least 100,000, at least 1,000,000, at least 10,000,000, or at least 100,000,000 circulating microparticles.

BRIEF DESCRIPTION OF THE DRAWINGS

The invention, together with further objects and advantages thereof, may best be understood by making reference to the description taken together with the accompanying drawings, in which:

Figure 1 illustrates a multimeric barcoding reagent that may be used in the method illustrated in Figure 3 or Figure 4.

Figure 2 illustrates a kit comprising a multimeric barcoding reagent and adapter oligonucleotides for labelling a target nucleic acid.

Figure 3 illustrates a first method of preparing a nucleic acid sample for sequencing using a multimeric barcoding reagent.

Figure 4 illustrates a second method of preparing a nucleic acid sample for sequencing using a multimeric barcoding reagent.

Figure 5 illustrates a method of preparing a nucleic acid sample for sequencing using a multimeric barcoding reagent and adapter oligonucleotides.

Figure 6 illustrates a method of preparing a nucleic acid sample for sequencing using a multimeric barcoding reagent, adapter oligonucleotides and target oligonucleotides.

Figure 7 illustrates a method of assembling a multimeric barcode molecule using a rolling circle amplification process.

Figure 8 illustrates a method of synthesizing multimeric barcoding reagents for labeling a target nucleic acid that may be used in the methods illustrated in Figure 3, Figure 4 and/or Figure 5.

Figure 9 illustrates an alternative method of synthesizing multimeric barcoding reagents (as illustrated in Figure 1) for labeling a target nucleic acid that may be used in the method illustrated in Figure 3 and/or Figure 4.

Figure 10 is a graph showing the total number of nucleotides within each barcode sequence.

Figure 11 is a graph showing the total number of unique barcode molecules in each sequenced
5 multimeric barcode molecule.

Figure 12 shows representative multimeric barcode molecules that were detected by the analysis script.

10 Figure 13 is a graph showing the number of unique barcodes per molecular sequence identifier against the number of molecular sequence identifiers following the barcoding of synthetic DNA templates of known sequence with multimeric barcoding reagents containing barcoded oligonucleotides.

15 Figure 14 is a graph showing the number of unique barcodes per molecular sequence identifier against the number of molecular sequence identifiers following the barcoding of synthetic DNA templates of known sequence with multimeric barcoding reagents and separate adapter oligonucleotides.

20 Figure 15 is a table showing the results of barcoding genomic DNA loci of three human genes (BRCA1, HLA-A and DQB1) with multimeric barcoding reagents containing barcoded oligonucleotides.

Figure 16 is a schematic illustration of a sequence read obtained from barcoding genomic DNA
25 loci with multimeric barcoding reagents containing barcoded oligonucleotides.

Figure 17 is a graph showing the number of barcodes from the same multimeric barcoding reagent that labelled sequences on the same synthetic template molecule against the number of synthetic template molecules.

30 Figure 18 illustrates a method in which two or more sequences from a microparticle are determined and linked informatically.

Figure 19 illustrates a method in which sequences from a particular microparticle are linked by a
35 shared identifier.

Figure 20 illustrates a method in which molecular barcodes are appended to fragments of genomic DNA within microparticles that have been partitioned, and wherein said barcodes provide a linkage between sequences derived from the same microparticle.

Figure 21 illustrates a specific method in which molecular barcodes are appended to fragments of genomic DNA within microparticles by multimeric barcoding reagents, and wherein said barcodes provide a linkage between sequences derived from the same microparticle.

5

Figure 22 illustrates a method in which fragments of genomic DNA within individual microparticles are appended to each other, and wherein the resulting molecules are sequenced, such that sequences from two or more fragments of genomic DNA from the same microparticle are determined from the same sequenced molecule, thereby establishing a linkage between

10

Figure 23 illustrates a method in which individual microparticles (and/or small groups of microparticles) from a large sample of microparticles are sequenced in two or more separate, individual sequencing reactions, and the sequences determined from each such sequencing reaction are thus determined to be linked informatically and thus predicted to derive from the same individual microparticle (and/or small group of microparticles).

15

Figure 24 illustrates a specific method in which fragments of genomic DNA within individual microparticles are appended to a discrete region of a sequencing flow cell prior to sequencing, and wherein the proximity of fragments sequenced on said flow cell provides a linkage between sequences derived from the same microparticle.

20

Figure 25 illustrates the linkage of sequences of fragments of genomic DNA within a circulating microparticle, as produced by a method of appending barcoded oligonucleotides (from the 'Variant A' version of the example protocol). Shown is the density of sequence reads across all chromosomes in the human genome, with clear clustering of reads within singular chromosomal segments.

25

Figure 26 illustrates the linkage of sequences of fragments of genomic DNA within a circulating microparticle, as produced by a method of appending barcoded oligonucleotides (from the 'Variant B' version of the example protocol). Shown is the density of sequence reads across all chromosomes in the human genome, with clear clustering of reads within singular chromosomal segments.

30

Figure 27 illustrates the linkage of sequences of fragments of genomic DNA within a circulating microparticle, as produced by a method of appending barcoded oligonucleotides (from the 'Variant B' version of the example protocol). Shown is the density of sequence reads zoomed in within a specific chromosomal segment, to show the focal, high-density nature of these linked reads.

35

- Figure 28 illustrates the linkage of sequences of fragments of genomic DNA within a circulating microparticle, as produced by a method of appending barcoded oligonucleotides (from the 'Variant C' version of the example protocol). Shown is the density of sequence reads across all chromosomes in the human genome, with clear clustering of reads within singular chromosomal segments, though with such segments being larger in chromosomal span than in the other Variant methods (due to the larger microparticles being pelleted within Variant C compared with Variants A or B).
- Figure 29 illustrates a negative-control experiment, wherein fragments of genomic DNA are purified (i.e. therefore being unlinked) before being appended to barcoded oligonucleotides. No clustering of reads is observed at all, validating that circulating microparticles comprise fragments of genomic DNA from focal, contiguous genomic regions.
- Figure 30 illustrates the concept of multi-parametric measurement of target molecules of a single circulating microparticle.
- Figure 31 illustrates a method in which target biomolecules are measured using barcoded affinity probes and a step of partitioning.
- Figure 32 illustrates a method in which target biomolecules are measured using barcoded affinity probes and multimeric barcoding reagents.
- Figure 33 illustrates a method (and associated experimental results) of analysing a sample comprising a circulating microparticle, wherein the circulating microparticle comprises fragments of genomic DNA and a protein, and wherein the method comprises measurement of the protein using an antibody-conjugated bead-based approach, and subsequent barcoding and sequencing of fragments of genomic DNA.
- Figure 34 illustrates a method (and associated experimental results) of analysing a sample comprising a circulating microparticle, wherein the circulating microparticle comprises fragments of genomic DNA and a protein, and wherein the method comprises measurement of the protein using an antibody-conjugated bead-based approach, and wherein a step of measuring modified nucleobases is also performed, and then with subsequent barcoding and sequencing of fragments of genomic DNA.
- Figure 35 illustrates a method (and associated experimental results) of analysing a sample comprising a circulating microparticle, wherein the circulating microparticle comprises fragments of genomic DNA and a first protein and a second protein, and wherein the method comprises

measurement of the first protein using an antibody-conjugated bead-based approach, and measurement of the second protein with a barcoded affinity probe, and subsequent barcoding and sequencing of fragments of genomic DNA and sequences from barcoded affinity probes. A detailed description of each of Figures 18-35 is provided below.

5

Figure 18 illustrates a method in which two or more sequences from a microparticle are determined and linked informatically. In the method, a microparticle, comprised within or derived from a blood, plasma, or serum sample, comprises two or more fragments of genomic DNA. The sequences of at least parts of these fragments of genomic DNA is determined; and furthermore, through one or more methods, an informatic linkage is established such that the first and second sequences from a microparticle are linked.

This linkage may take any form, such as a shared identifier (which could, for example, derive from a shared barcode that may be appended to said first and second genomic DNA sequences during a molecular barcoding process); any other shared property may also be used to link the two sequences; the data comprising the sequences themselves may be comprised within a shared electronic storage medium or partition thereof. Furthermore, the linkage may comprise a non-binary or relative value, for example representing the physical proximity of the two fragments within a spatially-metered sequencing reaction, or representing an estimated likelihood or probability that the two sequences may derive from fragments of genomic DNA comprised within the same microparticle.

Figure 19 illustrates a method in which sequences from a particular microparticle are linked by a shared identifier. In the method, a number of sequences from fragments of genomic DNA comprised within two different microparticles (e.g. two different microparticles derived from a single blood, plasma, or serum sample) are determined, e.g. by a nucleic acid sequencing reaction. Sequences corresponding to fragments of genomic DNA from the first microparticle are each assigned to the same informatic identifier (here, the identifier '0001'), and sequences corresponding to fragments of genomic DNA from the second microparticle are each assigned to the same, different informatic identifier (here, the identifier '0002'). This information of sequences and corresponding identifiers thus comprises informatic linkages between sequences derived from the same microparticle, with the set of different identifiers serving the function of informatic linkage.

Figure 20 illustrates a method in which molecular barcodes are appended to fragments of genomic DNA within microparticles that have been partitioned, and wherein said barcodes provide a linkage between sequences derived from the same microparticle. In the method, microparticles from a sample of microparticles are partitioned into two or more partitions, and then the fragments of genomic DNA within the microparticles are barcoded within the partitions, and

then sequences are determined in such a way that the barcodes identify from which partition the sequence was derived, and thereby link the different sequences from individual microparticles.

5 In the first step, microparticles are partitioned into two or more partitions (which could comprise, for example, different physical reaction vessels, or different droplets within an emulsion). The fragments of genomic DNA are then released from the microparticles within each partition (i.e., the fragments are made physically accessible such that they can then be barcoded). This release step may be performed with a high-temperature incubation step, and/or via incubation with a molecular solvent or chemical surfactant. Optionally (but not shown here), an amplification step
10 may be performed at this point, prior to appending barcode sequences, such that all or part of a fragment of genomic DNA is replicated at least once (e.g. in a PCR reaction), and then barcode sequences may be subsequently appended to the resulting replication products.

Barcode sequences are then appended to the fragments of genomic DNA. The barcode
15 sequences may take any form, such as primers which comprise a barcode region, or barcoded oligonucleotides within multimeric barcoding reagents, or barcode molecules within multimeric barcode molecules. The barcode sequences may also be appended by any means, for example by a primer-extension and/or PCR reaction, or a single-stranded or double-stranded ligation reaction, or by in vitro transposition. In any case, the process of appending barcode sequences
20 produces a solution of molecules within each partition wherein each such molecule comprises a barcode sequence, and then all or part of a sequence corresponding to a fragment of genomic DNA from a microparticle that was partitioned into said partition.

The barcode-containing molecules from different partitions are then merged together into a single
25 reaction, and then a sequencing reaction is performed on the resulting molecules to determine sequences of genomic DNA and the barcode sequences to which they have been appended. The associated barcode sequences are then used to identify the partitions from which each sequence was derived, and thereby link sequences determined in the sequencing reaction that were derived from fragments of genomic DNA comprised within the same microparticle or group of
30 microparticles.

Figure 21 illustrates a specific method in which molecular barcodes are appended to fragments of genomic DNA within microparticles by multimeric barcoding reagents, and wherein said barcodes provide a linkage between sequences derived from the same microparticle. In the method,
35 microparticles from a sample of microparticles are crosslinked and then permeabilised, and then the fragments of genomic DNA comprised within the microparticles are barcoded by multimeric barcoding reagents, and then sequences are determined in such a way that the barcodes identify by which multimeric barcoding reagent each sequence was barcoded, and thereby link the different sequences from individual microparticles.

In the first step, microparticles from a sample of microparticles are crosslinked by a chemical crosslinking agent. This step serves the purpose of holding fragments of genomic DNA within each microparticle in physical proximity to each other, such that the sample may be manipulated and processed whilst retaining the basic structural nature of the microparticles (i.e., whilst retaining physical proximity of genomic DNA fragments derived from the same microparticle). In a second step, the crosslinked microparticles are permeabilised (i.e., the fragments of genomic DNA are made physically accessible such that they can then be barcoded in a barcoding step); this permeabilisation may for example be performed by incubation with a chemical surfactant such as a non-ionic detergent.

Barcode sequences are then appended to fragments of genomic DNA, wherein barcode sequences comprised within a multimeric barcoding reagent (and/or multimeric barcode molecule) are appended to fragments within the same crosslinked microparticle. The barcode sequences may be appended by any means, for example by a primer-extension reaction, or by a single-stranded or double-stranded ligation reaction. The process of appending barcode sequences is conducted such that a library of many multimeric barcoding reagents (and/or multimeric barcode molecules) is used to append sequences to a sample comprising many crosslinked microparticles, under dilution conditions such that each multimeric barcoding reagent (and/or multimeric barcode molecule) typically will only barcode sequences comprised within a single microparticle.

A sequencing reaction is then performed on the resulting molecules to determine sequences of genomic DNA and the barcode sequences to which they have been appended. The associated barcode sequences are then used to identify by which multimeric barcoding reagent (and/or multimeric barcode molecule) each sequence was barcoded, and thereby link sequences determined in the sequencing reaction that were derived from fragments of genomic DNA comprised within the same microparticle.

Figure 22 illustrates a method in which fragments of genomic DNA within individual microparticles are appended to each other, and wherein the resulting molecules are sequenced, such that sequences from two or more fragments of genomic DNA from the same microparticle are determined from the same sequenced molecule, thereby establishing a linkage between fragments within the same microparticle. In the method, fragments of genomic DNA within individual microparticle are crosslinked to each other, and then blunted, and then the resulting blunted fragments of genomic DNA are ligated to each other into contiguous, multi-part sequences. The resulting molecules are then sequenced, such that sequences from two or more fragments of genomic DNA comprised within the same sequenced molecule are thus determined to be linked as deriving from the same microparticle.

In the first step, microparticles from a sample of microparticles are crosslinked by a chemical crosslinking agent. This step serves the purpose of holding fragments of genomic DNA within each microparticle in physical proximity to each other, such that the sample may be manipulated and processed whilst retaining the basic structural nature of the microparticles (i.e., whilst retaining physical proximity of genomic DNA fragments derived from the same microparticle). In a second step, the crosslinked microparticles are permeabilised (i.e., the fragments of genomic DNA are made physically accessible such that they can then be barcoded in a barcoding step); this permeabilisation may for example be performed by incubation with a chemical surfactant such as a non-ionic detergent.

In a next step, the ends of fragments of genomic DNA within each microparticle are blunted (i.e. any overhangs are removed and/or ends are filled-in) such that the ends are able to be appended to each other in a double-stranded ligation reaction. A double-stranded ligation reaction is then performed (e.g. with T4 DNA Ligase), wherein the blunted ends of molecules comprised within the same microparticles are ligated to each other into contiguous, multi-part double-stranded sequences. This ligation reaction (or any other step) may be performed under dilution conditions such that spurious ligation products between sequences comprised within two or more different microparticles are minimised.

A sequencing reaction is then performed on the resulting molecules to determine sequences of genomic DNA within each multi-part molecule. The resulting molecules are then evaluated, such that sequences from two or more fragments of genomic DNA comprised within the same sequenced molecule are thus determined to be linked as deriving from the same microparticle.

Figure 23 illustrates a method in which individual microparticles (and/or small groups of microparticles) from a large sample of microparticles are sequenced in two or more separate, individual sequencing reactions, and the sequences determined from each such sequencing reaction are thus determined to be linked informatively and thus predicted to derive from the same individual microparticle (and/or small group of microparticles). In the method, microparticles from a sample of microparticles are divided into two or more separate sub-samples of microparticles. Each sub-sample may comprise one or more individual microparticles, but in any case will comprise only a fraction of the original sample of microparticles.

The fragments of genomic DNA within each sub-sample are then released and processed into a form such that they may be sequenced (e.g., they may be appended to sequencing adapters such as Illumina sequencing adapters, and optionally amplified and purified for sequencing). This method may or may not include a step of appending barcode sequences; optionally the sequenced molecules do not comprise any barcode sequences.

- 221 -

Fragments of genomic DNA (and/or replicated copies thereof) from each individual sub-sample are then sequenced in separate, independent sequencing reactions. For example, molecules from each sub-sample may be sequenced on a separate sequencing flowcell, or may be sequenced
5 within a different lane of a flowcell, or may be sequenced within a different port or flowcell of a nanopore sequencer.

The resulting sequenced molecules are then evaluated, such that sequences from the same individual sequencing reaction are thus determined to be linked as deriving from the same
10 microparticle (and/or from the same small group of microparticles).

Figure 24 illustrates a specific method in which fragments of genomic DNA within individual microparticles are appended to a discrete region of a sequencing flowcell prior to sequencing, and wherein the proximity of fragments sequenced on said flowcell comprises a linkage between
15 sequences derived from the same microparticle. In the method, microparticles from a sample of microparticles are crosslinked and then permeabilised, and then fragments of genomic DNA comprised within individual microparticles are appended to a sequencing flowcell, such that two or more fragments from the same individual microparticle are appended to the same region of the flowcell. The appended molecules are then sequenced, and the proximity of the resulting
20 sequences on the flowcell comprises a linking value, wherein sequences within close proximity on the flowcell may be predicted to derive from the same individual microparticle within the original sample.

In the first step, microparticles from a sample of microparticles are crosslinked by a chemical
25 crosslinking agent. This step serves the purpose of holding fragments of genomic DNA within each microparticle in physical proximity to each other, such that the sample may be manipulated and processed whilst retaining the basic structural nature of the microparticles (i.e., whilst retaining physical proximity of genomic DNA fragments derived from the same microparticle). In a second step, the crosslinked microparticles are permeabilised (i.e., the fragments of genomic
30 DNA are made physically accessible such that they can then be appended to a flowcell); this permeabilisation may for example be performed by incubation with a chemical surfactant such as a non-ionic detergent.

In a next step, fragments of genomic DNA from microparticles are then appended to the flowcell
35 of a sequencing apparatus, such that two or more fragments crosslinked within the same microparticle are appended to the same discrete region of the flowcell. This may be performed in a multi-part reaction involving adapter molecules; for example, an adapter molecule may be appended to fragments of genomic DNA within microparticles, and said adapter molecule may comprise a single-stranded portion that is complementary to single-stranded primers on the

- 222 -

flowcell. Sequences from a crosslinked microparticle may then be allowed to diffuse and anneal to different primers within the same region of the flowcell.

5 The resulting sequenced molecules are then sequenced, such that the proximity of the resulting sequences on the flowcell provides a linking value, wherein sequences within close proximity on the flowcell (e.g. within a certain discrete region and/or proximity value) may be predicted to derive from the same individual microparticle within the original sample.

10 The advantages of the invention may be illustrated, by way of example only, by reference to possible applications in NIPT and cancer detection:

By way of example, in the field of oncology, the invention may enable a powerful new framework to screen for the early detection of cancer. Several groups are seeking to develop cfDNA assays which can detect low levels of circulating DNA from early tumours (so-called 'circulating tumour
15 DNA' or ctDNA) prior to metastatic conversion. One of the chief approaches taken to delineate cancerous from non-cancerous specimens is by detecting 'structural variants' (genetic amplifications, deletions, or translocations) that are a near-universal hallmark of malignancies; however, detection of such large-scale genetic events through the current 'molecular counting' framework requires ultra-deep sequencing of cfDNA to achieve statistically meaningful detection,
20 and even then requires that a sufficient amount of ctDNA be present in the plasma to generate a sufficient absolute molecular signal even with hypothetically unlimited sequencing depth.

By contrast, the current invention may enable direct molecular assessment of structural variation, with potential single-molecule sensitivity: any structural variation that includes a 'rearrangement
25 site' (for example, a point on one chromosome that has been translocated with and thus attached to another chromosome, or a point where a gene or other chromosomal segment has been amplified or deleted within a single chromosome) may be detectable directly by this method, since circulating microparticles containing DNA of the rearrangement may include a population of DNA fragments flanking both sides of the rearrangement site itself, which by this method can then be
30 linked with each other to informatically deduce both the location of the rearrangement itself, and the bound of the two participating genomic loci on each end thereof.

To conceptualise how this may improve both the cost-effectiveness and the absolute analytic sensitivity of a universal cancer screen, the example can be given of a hypothetical single
35 circulating microparticle, which contains a chromosomal translocation from an early cancer cell, and which contains a total of 1 megabase of DNA spanning the left and right halves of this translocation, with this DNA being fragmented as 10,000 different, 100-nucleotide-long individual fragments that cumulatively span the entire 1 megabase segment. To detect the presence of this translocation event using the current, unlinked-fragment-only approach, the single, 100-base-pair

– 223 –

fragment that itself contains the exact site of translocation would need to be sequenced, and sequenced across its entire length to detect the actual translocation site itself. This test method would thus need to both: 1) efficiently convert all of the 10,000 fragments into a format that can be read on a sequencer (i.e., the majority of the 10,000 fragments must be successfully processed and retained throughout the entire DNA purification and sequencing sample-preparation process), and then 2) all of the 10,000 fragments must be sequenced at least once by a DNA sequencing process to reliably sequence the one that includes the translocation site (i.e., at least 1 megabase of sequencing must be performed, even assuming a theoretical uniform sampling of all input molecules into the sequencing step). Thus, 1 megabase of sequencing would need to be performed to detect the translocation event.

By contrast, to detect the presence of the translocation with a high degree of statistical confidence but using the linked-fragment approach, only a small number of input fragments from each side of the translocation site itself would need to be sequenced (to distinguish a 'confident' translocation event from e.g. statistical noise or mis-mapping errors). To provide a high degree of statistical confidence, on the order of 10 fragments from each side of the translocation could be sequenced; and since they need only be mapped to a location in the genome and not sequenced across their entire length to observe the actual translocation itself, on the order of only 50 base pairs from each fragment need be sequenced. Taken together, this generates a total sequencing requirement of 1000 base pairs to detect the presence of the translocation – a 1000-fold reduction from the 1,000,000 base pairs required by current state-of-the-art.

In addition to this considerable benefit in terms of relative sequencing throughput and cost, a linked-read approach may also increase the absolute achievable sensitivity of these cancer-screening tests. Since, for early-stage (and thus potentially curable) cancers, the absolute amount of tumour DNA in the circulation is low, the loss of sample DNA during the sample processing and preparation process for sequencing could significantly impede test efficacy, even with theoretically limitless sequencing depth. In keeping with the above example, using current approaches, the single DNA fragment containing the translocation site itself would need to be retained and successfully processed throughout the entire sample collection, processing, and sequencing-preparation protocol and then be successfully sequenced. However, all of these steps result in a certain fraction of 'input' molecules thereto being either physically lost from the processed sample (e.g. during a centrifugation or cleanup step), or otherwise simply not successfully processed/modified for subsequent steps (e.g., not successfully amplified prior to placement on a DNA sequencer). In contrast, since the linked-read approach of the invention need only involve sequencing of a small proportion of actual 'input' molecules, this type of sample loss may have a considerably reduced impact upon the ultimate sensitivity of the final assay.

In addition to its applications in oncology and cancer screening, this invention may also enable considerable new tools in the domain of noninvasive prenatal testing (NIPT). A developing foetus (and the placenta in which it is contained) shed fragmented DNA into the maternal circulation, a proportion of which is contained within circulating microparticles. Analogous to the problem of cancer screening from ctDNA, circulating foetal DNA only represents a minor fraction of the overall circulating DNA in pregnant individuals (the majority of circulating DNA being normal maternal DNA). A considerable technical challenge for NIPT revolves around differentiating actual foetal DNA from maternal DNA fragments (which will share the same nucleotide sequence since they are the source of inheritance for half of the foetal genome). An additional technical challenge for NIPT involves the detection of long-range genomic sequences (or mutations) from the short fragments of foetal DNA present in the circulation.

Analysis of linked fragments originating from the same individual circulating microparticle presents a powerful framework for substantially addressing both of these technical challenges for NIPT. Since (approximately) half of the foetal genome will be identical in sequence to the (approximately) half of the maternal genome which the developing foetus has inherited, it is difficult to distinguish whether a given sequenced fragment with a maternal sequence may have been generated by normal maternal tissues, or rather by developing foetal tissues. By contrast, for the (approximately) half of the foetal genome which has been paternally inherited (inherited from the father), the presence of sequence variants (e.g. single nucleotide variants or other variants) present in the paternal genome but not in the maternal genome serves as a molecular marker to identify these paternally-inherited foetal fragments (since the only paternal DNA sequences in circulation will be those from the pregnancy itself).

The ability to sequence multiple fragments from single circulating foetal microparticles that happen to contain both maternal and paternal sequences (e.g. sequences from one particular maternally-inherited foetal chromosome, together with sequences from a second foetal chromosome that has been paternally inherited) thus presents a method for direct recognition of which maternal sequences have been inherited by the developing foetus: maternal sequences that are found co-localised within microparticles that also contain paternal sequences can be predicted to be foetally-inherited maternal sequences, and, in contrast, maternal sequences that are not found co-localised with paternal sequences can be predicted to represent the maternal sequences which were not inherited by the foetus. By this technique, the large majority of circulating DNA that is comprised of normal maternal DNA may be specifically filtered out of the processed sequence dataset, and only sequences evidenced as being true foetal sequences may be isolated informatically for further analysis.

Since 'foetal fractions' (the fraction of all circulating DNA which has been generated by the foetus itself) for NIPT assays are frequently below 10%, and for some clinical specimens between 1%

- and 5%, and since this paternal-sequence-derived 'informatic-gating' step produces an 'effective foetal fraction' of 100% (assuming minimal mis-mapping errors), this linked-fragment approach has the potential to improve the signal-to-noise ratio for NIPT tests by one to two orders of magnitude. Therefore, the invention has the potential to improve the overall analytic sensitivity and specificity of NIPT tests, as well as considerably reduce the amount of sequencing required for the process, and also enable NIPT tests to be performed earlier in pregnancy (time points at which foetal fractions are sufficiently low that current tests have unacceptable false-positive and false-negative rates).
- Importantly, the present invention provides a novel, orthogonal dimensionality within sequence data from circulating DNA in the form of informatically linked sequences, upon which analysis algorithms, computations, and/or statistical tests may be performed directly to generate considerably more sensitive and specific genetic measurements. For example, rather than evaluating overall amounts of sequence between two chromosomes across an entire sample to measure a foetal chromosomal aneuploidy, linked sequences (and/or sets or subsets thereof) can be assessed directly to examine, for example, the number of sequences per informatically-linked set that map to a particular chromosome or chromosome portion. Comparisons and/or statistical tests may be performed to compare linked sets of sequences of different presumed cellular origin (for example, comparison between foetal sequences and maternal sequences, or between presumed healthy tissues and presumed cancerous or malignant tissues), or to evaluate sequence features or numeric features which only exist at the level of linked sets of sequences (and which do not exist at the level of individual, unlinked sequences), such as specific chromosomal distribution patterns, or cumulative enrichments of particular sequences or sequence sets.
- In addition to its application for detection of foetal microparticle sequences, this method has the potential to detect long-range genetic sequences or sequence mutations present in the foetal genome. Much in the same manner as described for cancer genome rearrangements, if several DNA fragments from a foetal microparticle are sequenced that span and/or flank a genomic rearrangement site (e.g. a translocation or amplification or deletion), then these classes of rearrangements may be informatically detected even without directly sequencing rearrangement sites themselves. In addition, outside of genomic rearrangement events, this method has the potential to detect 'phasing' information within individual genomic regions. For example, if two single-nucleotide variants are found at different points within a specific gene but separated by several kilobases of genomic distance, this method may enable assessment of whether these two single nucleotide variants are located on the same, single copy of the gene in the foetal genome, or whether they are each located on a different one of the two copies of the gene present in the foetal genome (i.e. whether they are located within the same haplotype). This function may have particular clinical utility for the genetic assessment and prognosis of de novo single nucleotide

mutations in foetal genomes, which comprise a large fraction of major developmental disorders with genetic etiology.

Figure 31 illustrates a method in which target biomolecules are measured using
5 barcoded affinity probes and a step of partitioning. In the method, barcoded affinity probes are incubated with microparticles from a sample of microparticles and allowed to bind to target polypeptides (i.e. target biomolecules) within or upon said microparticles. The barcoded affinity probes comprise an affinity moiety capable of binding to the target polypeptides and a barcoded oligonucleotide that identifies the barcoded affinity probe. The microparticles are then partitioned
10 into two or more partitions, and then the fragments of genomic DNA within the microparticles and the barcoded oligonucleotides from bound barcoded affinity probes are barcoded within the partitions, and then sequences are determined in such a way that the barcodes identify from which partition the sequence was derived, and thereby link the different sequences from individual microparticles.

15 Following a step of binding barcoded affinity probes to target polypeptides, microparticles are partitioned into two or more partitions (which could comprise, for example, different physical reaction vessels, or different droplets within an emulsion). The fragments of genomic DNA and barcoded oligonucleotides from barcoded affinity probes are then released from the microparticles
20 within each partition (i.e., the fragments are made physically accessible such that they can then be barcoded). This release step may be performed with a high-temperature incubation step, and/or via incubation with a molecular solvent or chemical surfactant. Optionally (but not shown here), an amplification step may be performed at this point, prior to appending barcode sequences, such that all or part of a fragment of genomic DNA is replicated at least once (e.g. in
25 a PCR reaction), and then barcode sequences may be subsequently appended to the resulting replication products.

Barcode sequences are then appended to the fragments of genomic DNA (or amplified products thereof) and barcoded oligonucleotides (or amplicons thereof) from barcoded affinity probes (i.e.
30 barcode sequences are appended to the "target nucleic acid molecules". The barcode sequences may take any form, such as primers which comprise a barcode region, or barcoded oligonucleotides within multimeric barcoding reagents, or barcode molecules within multimeric barcode molecules. The barcode sequences may also be appended by any means, for example by a primer-extension and/or PCR reaction, or a single-stranded or double-stranded ligation
35 reaction, or by in vitro transposition. In any case, the process of appending barcode sequences produces a solution of molecules within each partition wherein each such molecule comprises a barcode sequence, and then all or part of a sequence corresponding to a fragment of genomic DNA or barcoded oligonucleotide from a barcoded affinity probe from a microparticle that was partitioned into said partition.

The barcode-containing molecules from different partitions are then merged together into a single reaction, and then a sequencing reaction is performed on the resulting molecules to determine sequences of genomic DNA and/or sequences from barcoded affinity probes and the barcode sequences to which they have been appended. The associated barcode sequences are then used to identify the partitions from which each sequence was derived, and thereby link sequences determined in the sequencing reaction that were derived from target biomolecules comprised within the same microparticle or group of microparticles. The sequence of the barcoded oligonucleotide identifies the linked affinity moiety and thereby the target polypeptide to which the affinity moiety binds. Therefore, the sequencing data identifies genomic DNA fragments and one or more target polypeptides likely to have been co-localised within the same circulating microparticle.

Figure 32 illustrates a method in which target biomolecules are measured using barcoded affinity probes and multimeric barcoding reagents. In the method, barcoded affinity probes are incubated with microparticles from a sample of microparticles and allowed to bind to target polypeptides (i.e. target biomolecules) within or upon said microparticles. The barcoded affinity probes comprise an affinity moiety capable of binding to the target polypeptides and a barcoded oligonucleotide that identifies the barcoded affinity probe. Microparticles from a sample of microparticles are then crosslinked and then permeabilised, and then target nucleic acid molecules (i.e. the fragments of genomic DNA and the barcoded oligonucleotides from barcoded affinity probes comprised within the microparticles) are barcoded by multimeric barcoding reagents, and then sequences are determined in such a way that the barcodes identify by which multimeric barcoding reagent each sequence was barcoded, and thereby link the different sequences from individual microparticles.

Following a step of binding barcoded affinity probes to target polypeptides, microparticles from a sample of microparticles are crosslinked by a chemical crosslinking agent. This step serves the purpose of holding fragments of genomic DNA and barcoded oligonucleotides from barcoded affinity probes within each microparticle in physical proximity to each other, such that the sample may be manipulated and processed whilst retaining the basic structural nature of the microparticles (i.e., whilst retaining physical proximity of genomic DNA fragments and barcoded oligonucleotides from barcoded affinity probes derived from the same microparticle). In a second step, the crosslinked microparticles are permeabilised (i.e., the fragments of genomic DNA are made physically accessible such that they can then be barcoded in a barcoding step); this permeabilisation may for example be performed by incubation with a chemical surfactant such as a non-ionic detergent. Optionally, a (first or second) step of binding barcoded affinity probes to target polypeptides may be performed following any such step of crosslinking, and/or following any such step of permeabilisation.

– 228 –

Barcode sequences are then appended to fragments of genomic DNA and to barcoded oligonucleotides comprised with barcoded affinity probes, wherein barcode sequences comprised within a multimeric barcoding reagent (and/or multimeric barcode molecule) are appended to fragments within or bound to the same crosslinked microparticle. The barcode sequences may be
5 appended by any means, for example by a primer-extension reaction, or by a single-stranded or double-stranded ligation reaction. The process of appending barcode sequences is conducted such that a library of many multimeric barcoding reagents (and/or multimeric barcode molecules) is used to append sequences to a sample comprising many crosslinked microparticles, under dilution conditions such that each multimeric barcoding reagent (and/or multimeric barcode
10 molecule) typically will only barcode target nucleic acid molecules comprised within a single microparticle. Optionally, any method of appending one or more coupling molecules to target nucleic acid molecules (e.g. to fragments of genomic DNA and/or to barcoded oligonucleotides from barcoded affinity probes) may be performed prior to and/or during any step of appending barcode sequences, and then (optionally) barcode sequences from multimeric barcoding reagents
15 may be linked to said coupling molecules, optionally with a subsequent barcode-connecting step wherein said barcode sequences are appended to said target nucleic acid molecules.

A sequencing reaction is then performed on the resulting molecules to determine sequences of genomic DNA and barcoded oligonucleotides from barcoded affinity probes and the barcode
20 sequences to which they have been appended. The associated barcode sequences are then used to identify by which multimeric barcoding reagent (and/or multimeric barcode molecule) each sequence was barcoded, and thereby link sequences determined in the sequencing reaction that were derived from fragments of genomic DNA and barcoded oligonucleotides from barcoded affinity probes comprised within or bound to the same microparticle. The sequence of the
25 barcoded oligonucleotide identifies the linked affinity moiety and thereby the target polypeptide to which the affinity moiety binds. Therefore, the sequencing data identifies genomic DNA fragments and one or more target polypeptides likely to have been co-localised within the same circulating microparticle.

30 EXAMPLES

EXAMPLE 1

MATERIALS AND METHODS

35

Method 1 – Synthesis of a Library of Nucleic Acid Barcode Molecules

Synthesis of Double-Stranded Sub-Barcode Molecule Library

– 229 –

In a PCR tube, 10 microliters of 10 micromolar BC_MX3 (an equimolar mixture of all sequences in SEQ ID NO: 18 to 269) were added to 10 microliters of 10 micromolar BC_ADD_TP1 (SEQ ID NO: 1), plus 10 microliters of 10X CutSmart Buffer (New England Biolabs) plus 1.0 microliter of 10 millimolar deoxynucleotide triphosphate nucleotide mix (Invitrogen) plus 68 microliters H₂O, to final volume of 99 microliters. The PCR tube was placed on a thermal cycler and incubated at 75°C for 5 minutes, then slowly annealed to 4°C, then held 4°C, then placed on ice. 1.0 microliter of Klenow polymerase fragment (New England Biolabs; at 5 U/uL) was added to the solution and mixed. The PCR tube was again placed on a thermal cycler and incubated at 25°C for 15 minutes, then held at 4°C. The solution was then purified with a purification column (Nucleotide Removal Kit; Qiagen), eluted in 50 microliters H₂O, and quantitated spectrophotometrically.

Synthesis of Double-Stranded Downstream Adapter Molecule

In a PCR tube, 0.5 microliters of 100 micromolar BC_ANC_TP1 (SEQ ID NO: 2) were added to 0.5 microliters of 100 micromolar BC_ANC_BT1 (SEQ ID NO: 3), plus 20 microliters of 10X CutSmart Buffer (New England Biolabs) plus 178 microliters H₂O, to final volume of 200 microliters. The PCR tube was placed on a thermal cycler and incubated at 95°C for 5 minutes, then slowly annealed to 4°C, then held 4°C, then placed on ice, then stored at -20°C.

Ligation of Double-Stranded Sub-Barcode Molecule Library to Double-Stranded Downstream Adapter Molecule

In a 1.5 milliliter Eppendorf tube, 1.0 microliter of Double-Stranded Downstream Adapter Molecule solution was added to 2.5 microliters of Double-Stranded Sub-Barcode Molecule Library, plus 2.0 microliters of 10X T4 DNA Ligase buffer, and 13.5 microliters H₂O to final volume of 19 microliters. 1.0 microliter of T4 DNA Ligase (New England Biolabs; high concentration) was added to the solution and mixed. The tube was incubated at room temperature for 60 minutes, then purified with 1.8X volume (34 microliters) Ampure XP Beads (Agencourt; as per manufacturer's instructions), and eluted in 40 microliters H₂O.

PCR Amplification of Ligated Library

In a PCR tube, 2.0 microliters of Ligated Library were added to 2.0 microliters of 50 micromolar BC_FWD_PR1 (SEQ ID NO: 4), plus 2.0 microliters of 50 micromolar BC_REV_PR1 (SEQ ID NO: 5), plus 10 microliters of 10X Taq PCR Buffer (Qiagen) plus 2.0 microliter of 10 millimolar deoxynucleotide triphosphate nucleotide mix (Invitrogen) plus 81.5 microliters H₂O, plus 0.5 microliters Qiagen Taq Polymerase (at 5U/uL) to final volume of 100 microliters. The PCR tube was placed on a thermal cycler and amplified for 15 cycles of: 95°C for 30 seconds, then 59°C for 30 seconds, then 72°C for 30 seconds; then held at 4°C. The solution was then purified with 1.8X volume (180 microliters) Ampure XP Beads (Agencourt; as per manufacturer's instructions), and eluted in 50 microliters H₂O.

– 230 –

Uracil Glycosylase Enzyme Digestion

To an eppendorf tube 15 microliters of the eluted PCR amplification, 1.0 microliters H₂O, plus 2.0 microliters 10X CutSmart Buffer (New England Biolabs), plus 2.0 microliter of USER enzyme solution (New England Biolabs) was added and mixed. The tube was incubated at 37°C for 60 minutes, then the solution was purified with 1.8X volume (34 microliters) Ampure XP Beads (Agencourt; as per manufacturer's instructions), and eluted in 34 microliters H₂O.

MlyI Restriction Enzyme Cleavage

To the eluate from the previous (glycosylase digestion) step, 4.0 microliters 10X CutSmart Buffer (New England Biolabs), plus 2.0 microliter of MlyI enzyme (New England Biolabs, at 5U/uL) was added and mixed. The tube was incubated at 37°C for 60 minutes, then the solution was purified with 1.8X volume (72 microliters) Ampure XP Beads (Agencourt; as per manufacturer's instructions), and eluted in 40 microliters H₂O.

Ligation of Sub-Barcode Library to MlyI-Cleaved Solution

In a 1.5 milliliter Eppendorf tube, 10 microliter of MlyI-Cleaved Solution solution was added to 2.5 microliters of Double-Stranded Sub-Barcode Molecule Library, plus 2.0 microliters of 10X T4 DNA Ligase buffer, and 4.5 microliters H₂O to final volume of 19 microliters. 1.0 microliter of T4 DNA Ligase (New England Biolabs; high concentration) was added to the solution and mixed. The tube was incubated at room temperature for 60 minutes, then purified with 1.8X volume (34 microliters) Ampure XP Beads (Agencourt; as per manufacturer's instructions), and eluted in 40 microliters H₂O.

Repeating Cycles of Sub-Barcode Addition

The experimental steps of : 1) Ligation of Sub-Barcode Library to MlyI-Cleaved Solution, 2) PCR Amplification of Ligated Library, 3) Uracil Glycosylase Enzyme Digestion, and 4) MlyI Restriction Enzyme Cleavage were repeated, in sequence, for a total of five cycles.

Synthesis of Double-Stranded Upstream Adapter Molecule

In a PCR tube, 1.0 microliters of 100 micromolar BC_USO_TP1 (SEQ ID NO: 6) were added to 1.0 microliters of 100 micromolar BC_USO_BT1 (SEQ ID NO: 7), plus 20 microliters of 10X CutSmart Buffer (New England Biolabs) plus 178 microliters H₂O, to final volume of 200 microliters. The PCR tube was placed on a thermal cycler and incubated at 95°C for 60 seconds, then slowly annealed to 4°C, then held 4°C, then placed on ice, then stored at -20°C.

Ligation of Double-Stranded Upstream Adapter Molecule

In a 1.5 milliliter Eppendorf tube, 3.0 microliters of Upstream Adapter solution were added to 10.0 microliters of final (after the fifth cycle) MlyI-Cleaved solution, plus 2.0 microliters of 10X T4 DNA Ligase buffer, and 5.0 microliters H₂O to final volume of 19 microliters. 1.0 microliter of T4 DNA

– 231 –

Ligase (New England Biolabs; high concentration) was added to the solution and mixed. The tube was incubated at room temperature for 60 minutes, then purified with 1.8X volume (34 microliters) Ampure XP Beads (Agencourt; as per manufacturer's instructions), and eluted in 40 microliters H₂O.

5

PCR Amplification of Upstream Adapter-Ligated Library

In a PCR tube, 6.0 microliters of Upstream Adapter-Ligated Library were added to 1.0 microliters of 100 micromolar BC_CS_PCR_FWD1 (SEQ ID NO: 8), plus 1.0 microliters of 100 micromolar BC_CS_PCR_REV1 (SEQ ID NO: 9), plus 10 microliters of 10X Taq PCR Buffer (Qiagen) plus 2.0 microliter of 10 millimolar deoxynucleotide triphosphate nucleotide mix (Invitrogen) plus 73.5 microliters H₂O, plus 0.5 microliters Qiagen Taq Polymerase (at 5U/μL) to final volume of 100 microliters. The PCR tube was placed on a thermal cycler and amplified for 15 cycles of: 95°C for 30 seconds, then 61°C for 30 seconds, then 72°C for 30 seconds; then held at 4°C. The solution, containing a library of amplified nucleic acid barcode molecules, was then purified with 1.8X volume (180 microliters) Ampure XP Beads (Agencourt; as per manufacturer's instructions). The library of amplified nucleic acid barcode molecules was then eluted in 40 microliters H₂O.

10

15

The library of amplified nucleic acid barcode molecules synthesised by the method described above was then used to assemble a library of multimeric barcode molecules as described below.

20

Method 2 – Assembly of a Library of Multimeric Barcode Molecules

A library of multimeric barcode molecules was assembled using the library of nucleic acid barcode molecules synthesised according to the methods of Method 1.

25 Primer-Extension with Forward Termination Primer and Forward Splinting Primer

In a PCR tube, 5.0 microliters of the library of amplified nucleic acid barcode molecules were added to 1.0 microliters of 100 micromolar CS_SPLT_FWD1 (SEQ ID NO: 10), plus 1.0 microliters of 5 micromolar CS_TERM_FWD1 (SEQ ID NO: 11), plus 10 microliters of 10X Thermopol Buffer (NEB) plus 2.0 microliter of 10 millimolar deoxynucleotide triphosphate nucleotide mix (Invitrogen) plus 80.0 microliters H₂O, plus 1.0 microliters Vent Exo-Minus Polymerase (New England Biolabs, at 2U/μL) to final volume of 100 microliters. The PCR tube was placed on a thermal cycler and amplified for 1 cycle of: 95°C for 30 seconds, then 53°C for 30 seconds, then 72°C for 60 seconds, then 1 cycle of: 95°C for 30 seconds, then 50°C for 30 seconds, then 72°C for 60 seconds, then held at 4°C. The solution was then purified a PCR purification column (Qiagen), and eluted in 85.0 microliters H₂O.

30

35

Primer-Extension with Reverse Termination Primer and Reverse Splinting Primer

In a PCR tube, the 85.0 microliters of forward-extension primer-extension products were added to 1.0 microliters of 100 micromolar CS_SPLT_REV1 (SEQ ID NO: 12), plus 1.0 microliters of 5

– 232 –

micromolar CS_TERM_REV1 (SEQ ID NO: 13), plus 10 microliters of 10X Thermopol Buffer (NEB) plus 2.0 microliter of 10 millimolar deoxynucleotide triphosphate nucleotide mix (Invitrogen), plus 1.0 microliters Vent Exo-Minus Polymerase (New England Biolabs, at 2U/uL) to final volume of 100 microliters. The PCR tube was placed on a thermal cycler and amplified for 1 cycle of: 95°C for 30 seconds, then 53°C for 30 seconds, then 72°C for 60 seconds, then 1 cycle of: 95°C for 30 seconds, then 50°C for 30 seconds, then 72°C for 60 seconds, then held at 4°C. The solution was then purified a PCR purification column (Qiagen), and eluted in 43.0 microliters H₂O.

10 **Linking Primer-Extension Products with Overlap-Extension PCR**

In a PCR tube were added the 43.0 microliters of reverse-extension primer-extension products, plus 5.0 microliters of 10X Thermopol Buffer (NEB) plus 1.0 microliter of 10 millimolar deoxynucleotide triphosphate nucleotide mix (Invitrogen), plus 1.0 microliters Vent Exo-Minus Polymerase (New England Biolabs, at 2U/uL) to final volume of 50 microliters. The PCR tube was placed on a thermal cycler and amplified for 5 cycles of: 95°C for 30 seconds, then 60°C for 60 seconds, then 72°C for 2 minutes; then 5 cycles of: 95°C for 30 seconds, then 60°C for 60 seconds, then 72°C for 5 minutes; then 5 cycles of: 95°C for 30 seconds, then 60°C for 60 seconds, then 72°C for 10 minutes; then held at 4°C. The solution was then purified with 0.8X volume (80 microliters) Ampure XP Beads (Agencourt; as per manufacturer's instructions), and eluted in 40 microliters H₂O.

Amplification of Overlap-Extension Products

In a PCR tube were added 2.0 microliters of Overlap-Extension PCR solution, plus 1.0 microliters of 100 micromolar CS_PCR_FWD1 (SEQ ID NO: 14), plus 1.0 microliters of 100 micromolar CS_PCR_REV1 (SEQ ID NO: 15), plus 10 microliters of 10X Thermopol Buffer (NEB) plus 2.0 microliter of 10 millimolar deoxynucleotide triphosphate nucleotide mix (Invitrogen), plus 1.0 microliters Vent Exo-Minus Polymerase (New England Biolabs, at 2U/uL), plus 83.0 microliters H₂O to final volume of 100 microliters. The PCR tube was placed on a thermal cycler and amplified for 15 cycles of: 95°C for 30 seconds, then 58°C for 30 seconds, then 72°C for 10 minutes; then held at 4°C. The solution was then purified with 0.8X volume (80 microliters) Ampure XP Beads (Agencourt; as per manufacturer's instructions), and eluted in 50 microliters H₂O, and quantitated spectrophotometrically.

Gel-Based Size Selection of Amplified Overlap-Extension Products

Approximately 250 nanograms of Amplified Overlap-Extension Products were loaded and run on a 0.9% agarose gel, and then stained and visualised with ethidium bromide. A band corresponding to 1000 nucleotide size (plus and minus 100 nucleotides) was excised and purified with a gel extraction column (Gel Extraction Kit, Qiagen) and eluted in 50 microliters H₂O.

– 233 –

Amplification of Overlap-Extension Products

- In a PCR tube were added 10.0 microliters of Gel-Size-Selected solution, plus 1.0 microliters of 100 micromolar CS_PCR_FWD1 (SEQ ID NO: 14), plus 1.0 microliters of 100 micromolar CS_PCR_REV1 (SEQ ID NO: 15), plus 10 microliters of 10X Thermopol Buffer (NEB) plus 2.0 microliter of 10 millimolar deoxynucleotide triphosphate nucleotide mix (Invitrogen), plus 1.0 microliters Vent Exo-Minus Polymerase (New England Biolabs, at 2U/uL) plus 75.0 microliters H₂O to final volume of 100 microliters. The PCR tube was placed on a thermal cycler and amplified for 15 cycles of: 95°C for 30 seconds, then 58°C for 30 seconds, then 72°C for 4 minutes; then held at 4°C. The solution was then purified with 0.8X volume (80 microliters) Ampure XP Beads (Agencourt; as per manufacturer's instructions), and eluted in 50 microliters H₂O, and quantitated spectrophotometrically.

Selection and Amplification of Quantitatively Known Number of Multimeric Barcode Molecules

- Amplified gel-extracted solution was diluted to a concentration of 1 picogram per microliter, and then to a PCR tube was added 2.0 microliters of this diluted solution (approximately 2 million individual molecules), plus 0.1 microliters of 100 micromolar CS_PCR_FWD1 (SEQ ID NO: 14), plus 0.1 microliters of 100 micromolar CS_PCR_REV1 (SEQ ID NO: 15), plus 1.0 microliter 10X Thermopol Buffer (NEB) plus 0.2 microliter of 10 millimolar deoxynucleotide triphosphate nucleotide mix (Invitrogen), plus 0.1 microliters Vent Exo-Minus Polymerase (New England Biolabs, at 2U/uL) plus 6.5 microliters H₂O to final volume of 10 microliters. The PCR tube was placed on a thermal cycler and amplified for 11 cycles of: 95°C for 30 seconds, then 57°C for 30 seconds, then 72°C for 4 minutes; then held at 4°C.
- To the PCR tube was added 1.0 microliters of 100 micromolar CS_PCR_FWD1 (SEQ ID NO: 14), plus 1.0 microliters of 100 micromolar CS_PCR_REV1 (SEQ ID NO: 15), plus 9.0 microliters of 10X Thermopol Buffer (NEB) plus 2.0 microliter of 10 millimolar deoxynucleotide triphosphate nucleotide mix (Invitrogen), plus 1.0 microliters Vent Exo-Minus Polymerase (New England Biolabs, at 2U/uL) plus 76.0 microliters H₂O to final volume of 100 microliters. The PCR tube was placed on a thermal cycler and amplified for 10 cycles of: 95°C for 30 seconds, then 57°C for 30 seconds, then 72°C for 4 minutes; then held at 4°C. The solution was then purified with 0.8X volume (80 microliters) Ampure XP Beads (Agencourt; as per manufacturer's instructions), and eluted in 50 microliters H₂O, and quantitated spectrophotometrically.

Method 3: Production of Single-Stranded Multimeric Barcode Molecules by In Vitro Transcription and cDNA Synthesis

This method describes a series of steps to produce single-stranded DNA strands, to which oligonucleotides may be annealed and then barcoded along. This method begins with four identical reactions performed in parallel, in which a promoter site for the T7 RNA Polymerase is

- 234 -

appended to the 5' end of a library of multimeric barcode molecules using an overlap-extension PCR amplification reaction. Four identical reactions are performed in parallel and then merged to increase the quantitative amount and concentration of this product available. In each of four identical PCR tubes, approximately 500 picograms of size-selected and PCR-amplified multimeric barcode molecules (as produced in the 'Selection and Amplification of Quantitatively Known Number of Multimeric Barcode Molecules' step of Method 2) were mixed with 2.0 microliters of 100 micromolar CS_PCR_FWD1_T7 (SEQ ID NO. 270) and 2.0 microliters of 100 micromolar CS_PCR_REV4 (SEQ ID NO. 271), plus 20.0 microliters of 10X Thermopol PCR buffer, plus 4.0 microliters of 10 millimolar deoxynucleotide triphosphate nucleotide mix, and 2.0 microliters Vent Exo Minus polymerase (at 5 units per microliter) plus water to a total volume of 200 microliters. The PCR tube was placed on a thermal cycler and amplified for 22 cycles of: 95°C for 60 seconds, then 60°C for 30 seconds, then 72°C for 3 minutes; then held at 4°C. The solution from all four reactions was then purified with a gel extraction column (Gel Extraction Kit, Qiagen) and eluted in 52 microliters H₂O.

Fifty (50) microliters of the eluate was mixed with 10 microliters 10X NEBuffer 2 (NEB), plus 0.5 microliters of 10 millimolar deoxynucleotide triphosphate nucleotide mix, and 1.0 microliters Vent Exo Minus polymerase (at 5 units per microliter) plus water to a total volume of 100 microliters. The reaction was incubated for 15 minutes at room temperature, then purified with 0.8X volume (80 microliters) Ampure XP Beads (Agencourt; as per manufacturer's instructions), and eluted in 40 microliters H₂O, and quantitated spectrophotometrically.

A transcription step is then performed, in which the library of PCR-amplified templates containing T7 RNA Polymerase promoter site (as produced in the preceding step) is used as a template for T7 RNA polymerase. This comprises an amplification step to produce a large amount of RNA-based nucleic acid corresponding to the library of multimeric barcode molecules (since each input PCR molecule can serve as a template to produce a large number of cognate RNA molecules). In the subsequent step, these RNA molecules are then reverse transcribed to create the desired, single-stranded multimeric barcode molecules. Ten (10) microliters of the eluate was mixed with 20 microliters 5X Transcription Buffer (Promega), plus 2.0 microliters of 10 millimolar deoxynucleotide triphosphate nucleotide mix, plus 10 microliters of 0.1 millimolar DTT, plus 4.0 microliters SuperAselN (Ambion), and 4.0 microliters Promega T7 RNA Polymerase (at 20 units per microliter) plus water to a total volume of 100 microliters. The reaction was incubated 4 hours at 37°C, then purified with an RNEasy Mini Kit (Qiagen), and eluted in 50 microliters H₂O, and added to 6.0 microliters SuperAselN (Ambion).

The RNA solution produced in the preceding in vitro transcription step is then reverse transcribed (using a primer specific to the 3' ends of the RNA molecules) and then digested with RNase H to create single-stranded DNA molecules corresponding to multimeric barcode molecules, to which

– 235 –

oligonucleotides maybe be annealed and then barcoded along. In two identical replicate tubes, 23.5 microliters of the eluate was mixed with 5.0 microliters of 10 millimolar deoxynucleotide triphosphate nucleotide mix, plus 3.0 microliters SuperAseIn (Ambion), and 10.0 microliters of 2.0 micromolar CS_PCR_REV1 (SEQ ID NO. 272) plus water to final volume of 73.5 microliters. The reaction was incubated on a thermal cycler at 65°C for 5 minutes, then 50°C for 60 seconds; then held at 4°C. To the tube was added 20 microliters 5X Reverse Transcription buffer (Invitrogen), plus 5.0 microliters of 0.1 millimolar DTT, and 1.75 microliters Superscript III Reverse Transcriptase (Invitrogen). The reaction was incubated at 55°C for 45 minutes, then 60°C for 5 minutes; then 70°C for 15 minutes, then held at 4°C., then purified with a PCR Cleanup column (Qiagen) and eluted in 40 microliters H₂O.

Sixty (60) microliters of the eluate was mixed with 7.0 microliters 10X RNase H Buffer (Promega), plus 4.0 microliters RNase H (Promega. The reaction was incubated 12 hours at 37°C, then 95°C for 10 minutes, then held at 4°C, then purified with 0.7X volume (49 microliters) Ampure XP Beads (Agencourt; as per manufacturer's instructions), and eluted in 30 microliters H₂O, and quantitated spectrophotometrically.

Method 4: Production of Multimeric Barcoding Reagents Containing Barcoded Oligonucleotides

This method describes steps to produce multimeric barcoding reagents from single-stranded multimeric barcode molecules (as produced in Method 3) and appropriate extension primers and adapter oligonucleotides.

In a PCR tube, approximately 45 nanograms of single-stranded RNase H-digested multimeric barcode molecules (as produced in the last step of Method 3) were mixed with 0.25 microliters of 10 micromolar DS_ST_05 (SEQ ID NO. 273, an adapter oligonucleotide) and 0.25 microliters of 10 micromolar US_PCR_Prm_Only_03 (SEQ ID NO. 274, an extension primer), plus 5.0 microliters of 5X Isothermal extension/ligation buffer, plus water to final volume of 19.7 microliters. In order to anneal the adapter oligonucleotides and extension primers to the multimeric barcode molecules, in a thermal cycler, the tube was incubated at 98°C for 60 seconds, then slowly annealed to 55°C, then held at 55°C for 60 seconds, then slowly annealed to 50°C then held at 50°C for 60 seconds, then slowly annealed to 20°C at 0.1°C/sec, then held at 4°C. To the tube was added 0.3 microliters (0.625U) Phusion Polymerase (NEB; 2 U/uL) 2.5 microliters (100 U) Taq DNA Ligase (NEB; 40 U/uL); and 2.5 microliters 100 millimolar DTT. In order to extend the extension primer(s) across the adjacent barcode region(s) of each multimeric barcode molecule, and then to ligate this extension product to the phosphorylated 5' end of the adapter oligonucleotide annealed to the downstream thereof, the tube was then incubated at 50°C for 3 minutes, then held at 4°C. The reaction was then purified with a PCR Cleanup column (Qiagen) and eluted in 30 microliters H₂O, and quantitated spectrophotometrically.

Method 5: Production of Synthetic DNA Templates of Known Sequence

This method describes a technique to produce synthetic DNA templates with a large number of tandemly-repeated, co-linear molecular sequence identifiers, by circularizing and then tandemly
5 amplifying (with a processive, strand-displacing polymerase) oligonucleotides containing said molecular sequence identifiers. This reagent may then be used to evaluate and measure the multimeric barcoding reagents described herein.

In a PCR was added 0.4 microliters of 1.0 micromolar Syn_Temp_01 (SEQ ID NO. 275) and 0.4
10 microliters of 1.0 micromolar ST_Splint_02 (SEQ ID NO. 276) and 10.0 microliters of 10X NEB CutSmart buffer. On a thermal cycler, the tube was incubated at 95°C for 60 seconds, then held at 75°C for 5 minutes, then slowly annealed to 20°C then held at 20°C for 60 seconds, then held at 4°C. To circularize the molecules through an intramolecular ligation reaction, the tube was then added 10.0 microliters ribo-ATP and 5.0 microliters T4 DNA Ligase (NEB; High Concentration).
15 The tube was then incubated at room temperature for 30 minutes, then at 65°C for 10 minutes, then slowly annealed to 20°C then held at 20°C for 60 seconds, then held at 4°C. To each tube was then added 10X NEB CutSmart buffer, 4.0 microliters of 10 millimolar deoxynucleotide triphosphate nucleotide mix, and 1.5 microliters of diluted phi29 DNA Polymerase (NEB; Diluted 1:20 in 1X CutSmart buffer) plus water to a total volume of 200 microliters. The reaction was
20 incubated at 30°C for 5 minutes, then held at 4°C, then purified with 0.7X volume (140 microliters) Ampure XP Beads (Agencourt; as per manufacturer's instructions), and eluted in 30 microliters H₂O, and quantitated spectrophotometrically.

Method 6: Barcoding Synthetic DNA Templates of Known Sequence with Multimeric**Barcoding Reagents Containing Barcoded Oligonucleotides**

In a PCR tube were added 10.0 microliters 5X Phusion HF buffer (NEB), plus 1.0 microliters 10 millimolar deoxynucleotide triphosphate nucleotide mix, plus 2.0 microliters (10 nanograms) 5.0 nanogram/ microliters Synthetic DNA Templates of Known Sequence (as produced by Method 5), plus water to final volume of 42.5 microliters. The tube was then incubated at 98°C for 60
30 seconds, then held at 20°C. To the tube was added 5.0 microliters of 5.0 picogram/microliter Multimeric Barcoding Reagents Containing Barcoded Oligonucleotides (as produced by Method 4). The reaction was then incubated at 70°C for 60 seconds, then slowly annealed to 60°C, then 60°C for five minutes, then slowly annealed to 55°C, then 55°C for five minutes, then slowly annealed to 50°C, then 50°C for five minutes, then held at 4°C. To the reaction was added 0.5
35 microliters of Phusion Polymerase (NEB), plus 2.0 microliters 10 uM SynTemp_PE2_B1_Short1 (SEQ ID NO. 277, a primer that is complementary to part of the extension products produced by annealing and extending the multimeric barcoding reagents created by Method 4 along the synthetic DNA templates created by Method 5, serves as a primer for the primer-extension and then PCR reactions described in this method). Of this reaction, a volume of 5.0 microliters was

– 237 –

- added to a new PCR tube, which was then incubated for 30 seconds at 55°C, 30 seconds 60°C, and 30 seconds 72°C, then followed by 10 cycles of: 98°C then 65°C then 72°C for 30 seconds each, then held at 4°C. To each tube was then added 9.0 microliters 5X Phusion buffer, plus 1.0 microliters 10 millimolar deoxynucleotide triphosphate nucleotide mix, plus 1.75 microliters 10 uM
- 5 SynTemp_PE2_B1_Short1 (SEQ ID NO. 277), plus 1.75 microliters 10 uM US_PCR_Prm_Only_02 (SEQ ID NO. 278, a primer partially complementary to the extension primer employed to generate the multimeric barcoding reagents as per Method 4, and serving as the 'forward' primer in this PCR amplification reaction), plus 0.5 microliters Phusion Polymerase (NEB), plus water to final volume of 50 microliters. The PCR tube was placed on a thermal cycler
- 10 and amplified for 24 cycles of: 98°C for 30 seconds, then 72°C for 30 seconds; then held at 4°C, then purified with 1.2X volume (60 microliters) Ampure XP Beads (Agencourt; as per manufacturer's instructions), and eluted in 30 microliters H₂O, and quantitated spectrophotometrically.
- 15 The resulting library was then barcoded for sample identification by a PCR-based method, amplified, and sequenced by standard methods using a 150-cycle, mid-output NextSeq flowcell (Illumina), and demultiplexed informatically for further analysis.

Method 7: Barcoding Synthetic DNA Templates of Known Sequence with Multimeric

20 Barcoding Reagents and Separate Adapter Oligonucleotides

- To anneal and extend adapter oligonucleotides along the synthetic DNA templates, in a PCR tube were added 10.0 microliters 5X Phusion HF buffer (NEB), plus 1.0 microliters 10 millimolar deoxynucleotide triphosphate nucleotide mix, plus 5.0 microliters (25 nanograms) 5.0 nanogram/ microliters Synthetic DNA Templates of Known Sequence (as produced by Method 5), plus 0.25
- 25 microliters of 10 micromolar DS_ST_05 (SEQ ID NO. 273, an adapter oligonucleotide), plus water to final volume of 49.7 microliters. On a thermal cycler, the tube was incubated at 98°C for 2 minutes, then 63°C for 1 minute, then slowly annealed to 60°C then held at 60°C for 1 minute, then slowly annealed to 57°C then held at 57°C for 1 minute, then slowly annealed to 54°C then held at 54°C for 1 minute, then slowly annealed to 50°C then held at 50°C for 1 minute, then
- 30 slowly annealed to 45°C then held at 45°C for 1 minute, then slowly annealed to 40°C then held at 40°C for 1 minute, then held at 4°C. To the tube was added 0.3 microliters Phusion Polymerase (NEB), and the reaction was incubated at 45°C for 20 seconds, then 50°C for 20 seconds, then 55°C for 20 seconds, 60°C for 20 seconds, then 72°C for 20 seconds, then held at 4°C; the reaction was then purified with 0.8X volume (40 microliters) Ampure XP Beads (Agencourt; as per
- 35 manufacturer's instructions), and eluted in 30 microliters H₂O, and quantitated spectrophotometrically.

In order to anneal adapter oligonucleotides (annealed and extended along the synthetic DNA templates as in the previous step) to multimeric barcode molecules, and then to anneal and then

- 238 -

- extend extension primer(s) across the adjacent barcode region(s) of each multimeric barcode molecule, and then to ligate this extension product to the phosphorylated 5' end of the adapter oligonucleotide annealed to the downstream thereof, to a PCR tube was added 10 microliters of the eluate from the previous step (containing the synthetic DNA templates along which the adapter oligonucleotides have been annealed and extended), plus 3.0 microliters of a 50.0 nanomolar solution of RNase H-digested multimeric barcode molecules (as produced in the last step of Method 3), plus 6.0 microliters of 5X Isothermal extension/ligation buffer, plus water to final volume of 26.6 microliters. On a thermal cycler, the tube was incubated at 70°C for 60 seconds, then slowly annealed to 60°C, then held at 60°C for 5 minutes, then slowly annealed to 55°C then held at 55°C for 5 minutes, then slowly annealed to 50°C at 0.1°C/sec then held at 50°C for 30 minutes, then held at 4°C. To the tube was added 0.6 microliters 10 uM US_PCR_Prm_Only_02 (SEQ ID NO: 278, an extension primer), and the reaction was incubated at 50°C for 10 minutes, then held at 4°C. To the tube was added 0.3 microliters (0.625U) Phusion Polymerase (NEB; 2 U/uL) 2.5 microliters (100 U) Taq DNA Ligase (NEB; 40 U/uL); and 2.5 microliters 100 millimolar DTT. The tube was then incubated at 50°C for 5 minutes, then held at 4°C. The reaction was then purified with 0.7X volume (21 microliters) Ampure XP Beads (Agencourt; as per manufacturer's instructions), and eluted in 30 microliters H₂O, and quantitated spectrophotometrically.
- To a new PCR tube was add 25.0 microliters of the eluate, plus 10.0 microliters 5X Phusion HF buffer (NEB), plus 1.0 microliters 10 millimolar deoxynucleotide triphosphate nucleotide mix, plus 2.0 microliters 10 uM SynTemp_PE2_B1_Short1 (SEQ ID NO: 277; a primer that is complementary to part of the extension products produced by the above steps; serves as a primer for the primer-extension and then PCR reactions described here), plus 0.5 uL Phusion Polymerase (NEB), plus water to final volume of 49.7 microliters. Of this reaction, a volume of 5.0 microliters was added to a new PCR tube, which was then incubated for 30 seconds at 55°C, 30 seconds 60°C, and 30 seconds 72°C, then followed by 10 cycles of: 98°C then 65°C then 72°C for 30 seconds each, then held at 4°C. To each tube was then added 9.0 microliters 5X Phusion buffer, plus 1.0 microliters 10 millimolar deoxynucleotide triphosphate nucleotide mix, plus 1.75 microliters 10 uM SynTemp_PE2_B1_Short1 (SEQ ID NO: 277), plus 1.75 microliters 10 uM US_PCR_Prm_Only_02 (SEQ ID NO: 278), plus 0.5 microliters Phusion Polymerase (NEB), plus water to final volume of 50 microliters. The PCR tube was placed on a thermal cycler and amplified for 24 cycles of: 98°C for 30 seconds, then 72°C for 30 seconds; then held at 4°C, then purified with 1.2X volume (60 microliters) Ampure XP Beads (Agencourt; as per manufacturer's instructions), and eluted in 30 microliters H₂O, and quantitated spectrophotometrically.

The resulting library was then barcoded for sample identification by a PCR-based method, amplified, and sequenced by standard methods using a 150-cycle, mid-output NextSeq flowcell (Illumina), and demultiplexed informatically for further analysis.

Method 9: Barcoding Genomic DNA Loci with Multimeric Barcoding Reagents Containing Barcoded Oligonucleotides

- This method describes a framework for barcoding targets within specific genomic loci (e.g. barcoding a number of exons within a specific gene) using multimeric barcoding reagents that contain barcoded oligonucleotides. First, a solution of Multimeric Barcode Molecules was produced by In Vitro Transcription and cDNA Synthesis (as described in Method 3). Then, solutions of multimeric barcoding reagents containing barcoded oligonucleotides was produced as described in Method 4, with a modification made such that instead of using an adapter oligonucleotide targeting a synthetic DNA template (i.e. DS_ST_05, SEQ ID NO: 273, as used in Method 4), adapter oligonucleotides targeting the specific genomic loci were included at that step. Specifically, a solution of multimeric barcoding reagents containing appropriate barcoded oligonucleotides was produced individually for each of three different human genes: BRCA1 (containing 7 adapter oligonucleotides, SEQ ID NOs 279-285), HLA-A (containing 3 adapter oligonucleotides, SEQ ID NOs 286-288), and DQB1 (containing 2 adapter oligonucleotides, SEQ ID NOs 289-290). The process of Method 4 was conducted for each of these three solutions as described above. These three solutions were then merged together, in equal volume, and diluted to a final, total concentration all barcoded oligonucleotides of approximately 50 nanomolar.
- In a PCR tube were plus 2.0 microliters 5X Phusion HF buffer (NEB), plus 1.0 microliter of 100 nanogram/microliter human genomic DNA (NA12878 from Coriell Institute) to final volume of 9.0 microliters. In certain variant versions of this protocol, the multimeric barcoding reagents (containing barcoded oligonucleotides) were also added at this step, prior to the high-temperature 98°C incubation. The reaction was incubated at 98°C for 120 seconds, then held at 4°C. To the tube was added 1.0 microliters of the above 50 nanomolar solution of multimeric barcode reagents, and then the reaction was incubated for 1 hour at 55°C, then 1 hour at 50°C, then 1 hour at 45°C, then held at 4°C. (Note that for certain samples, this last annealing process was extended to occur overnight, for a total of approximately 4 hours per temperature step).
- In order to add a reverse universal priming sequence to each amplicon sequence (and thus to enable subsequent amplification of the entire library at once, using just one forward and one reverse amplification primer), the reaction was diluted 1:100, and 1.0 microliter of the resulting solution was added in a new PCR tube to 20.0 microliters 5X Phusion HF buffer (NEB), plus 2.0 microliters 10 millimolar deoxynucleotide triphosphate nucleotide mix, plus 1.0 microliters a reverse-primer mixture (equimolar concentration of SEQ ID Nos 291-303, each primer at 5 micromolar concentration), plus 1.0 uL Phusion Polymerase (NEB), plus water to final volume of 100 microliters. The reaction was incubated at 53°C for 30 seconds, 72°C for 45 seconds, 98°C for 90 seconds, then 68°C for 30 seconds, then 64°C for 30 seconds, then 72°C for 30 seconds; then held at 4°C. The reaction was then purified with 0.8X volume (80 microliters) Ampure XP

– 240 –

Beads (Agencourt; as per manufacturer's instructions), and eluted in 30 microliters H₂O, and quantitated spectrophotometrically.

The resulting library was then barcoded for sample identification by a PCR-based method, amplified, and sequenced by standard methods using a 150-cycle, mid-output NextSeq flowcell (Illumina), and demultiplexed informatically for further analysis.

Method 10 – Sequencing the Library of Multimeric Barcode Molecules

10 Preparing Amplified Selected Molecules for Assessment with High-Throughput Sequencing

To a PCR tube was added 1.0 microliters of the amplified selected molecule solution, plus 1.0 microliters of 100 micromolar CS_SQ_AMP_REV1 (SEQ ID NO: 16), plus 1.0 microliters of 100 micromolar US_PCR_Prm_Only_02 (SEQ ID NO: 17), plus 10 microliters of 10X Thermopol Buffer (NEB) plus 2.0 microliter of 10 millimolar deoxynucleotide triphosphate nucleotide mix (Invitrogen), plus 1.0 microliters Vent Exo-Minus Polymerase (New England Biolabs, at 2U/uL) plus 84.0 microliters H₂O to final volume of 100 microliters. The PCR tube was placed on a thermal cycler and amplified for 3 cycles of: 95°C for 30 seconds, then 56°C for 30 seconds, then 72°C for 3 minutes; then held at 4°C. The solution was then purified with 0.8X volume (80 microliters) Ampure XP Beads (Agencourt; as per manufacturer's instructions), and eluted in 85 microliters H₂O.

This solution was then added to a new PCR tube, plus 1.0 microliters of 100 micromolar Illumina_PE1, plus 1.0 microliters of 100 micromolar Illumina_PE2, plus 10 microliters of 10X Thermopol Buffer (NEB) plus 2.0 microliter of 10 millimolar deoxynucleotide triphosphate nucleotide mix (Invitrogen), plus 1.0 microliters Vent Exo-Minus Polymerase (New England Biolabs, at 2U/uL) to final volume of 100 microliters. The PCR tube was placed on a thermal cycler and amplified for 4 cycles of: 95°C for 30 seconds, then 64°C for 30 seconds, then 72°C for 3 minutes; then 18 cycles of: 95°C for 30 seconds, then 67°C for 30 seconds, then 72°C for 3 minutes; then held at 4°C. The solution was then purified with 0.8X volume (80 microliters) Ampure XP Beads (Agencourt; as per manufacturer's instructions), and eluted in 40 microliters H₂O.

High-throughput Illumina sequencing was then performed on this sample using a MiSeq sequencer with paired-end, 250-cycle V2 sequencing chemistry.

Method 11 – Assessment of Multimeric Nature of Barcodes Annealed and Extended Along Single Synthetic Template DNA Molecules

– 241 –

A library of barcoded synthetic DNA templates was created using a solution of multimeric barcoding reagents produced according to a protocol as described generally in Method 3 and Method 4, and using a solution of synthetic DNA templates as described in Method 5, and using a laboratory protocol as described in Method 6; the resulting library was then barcoded for sample
 5 identification by a PCR-based method, amplified, and sequenced by standard methods using a 150-cycle, mid-output NextSeq flowcell (Illumina), and demultiplexed informatically for further analysis. The DNA sequencing results from this method were then compared informatically with data produced from Method 10 to assess the degree of overlap between the multimeric barcoding of synthetic DNA templates and the arrangement of said barcodes on individual multimeric
 10 barcoding reagents (the results are shown in Figure 17).

RESULTS

Structure and Expected Sequence Content of Each Sequence Multimeric Barcoding

15 Reagent Molecule

The library of multimeric barcode molecules synthesised as described in Methods 1 to 3 was prepared for high-throughput sequencing, wherein each molecule sequenced includes a contiguous span of a specific multimeric barcode molecule (including one or more barcode sequences, and one or more associate upstream adapter sequences and/or downstream adapter
 20 sequences), all co-linear within the sequenced molecule. This library was then sequenced with paired-end 250 nucleotide reads on a MiSeq sequencer (Illumina) as described. This yielded approximately 13.5 million total molecules sequenced from the library, sequenced once from each end, for a total of approximately 27 million sequence reads.

25 Each forward read is expected to start with a six nucleotide sequence, corresponding to the 3' end of the upstream adapter: TGACCT

This forward read is followed by the first barcode sequence within the molecule (expected to be 20 nt long).

30 This barcode is then followed by an 'intra-barcode sequence' (in this case being sequenced in the 'forward' direction (which is 82 nucleotides including both the downstream adapter sequence and upstream adapter sequence in series):

ATACCTGACTGCTCGTCAGTTGAGCGAATTCCGTATGGTGGTACACACCTACACTACTCGGA
 CGCTCTCCGATCTTGACCT

35

Within the 250 nucleotide forward read, this will then be followed by a second barcode, another intra-barcode sequence, and then a third barcode, and then a fraction of another intra-barcode sequence.

– 242 –

Each reverse read is expected to start with a sequence corresponding to the downstream adapter sequence: GCTCAACTGACGAGCAGTCAGGTAT

This reverse read is then followed by the first barcode coming in from the opposite end of the molecule (also 20 nucleotides long, but sequenced from the opposite strand of the molecule and thus of the inverse orientation to those sequenced by the forward read)

This barcode is then followed by the 'intra-barcode sequence' but in the inverse orientation (as it is on the opposite strand):

AGGTCAAGATCGGAAGAGCGTCCGAGTAGTGTAGGTGTGTACCACCATACGGAATTCGCTC
AACTGACGAGCAGTCAGGTAT

Likewise this 250 nucleotide reverse read will then be followed by a second barcode, another intra-barcode sequence, and then a third barcode, and then a fraction of another intra-barcode sequence.

Sequence Extraction and Analysis

With scripting in Python, each associated pair of barcode and flanking upstream-adapter and downstream-adapter sequence were isolated, with each individual barcode sequence of each barcode molecule then isolated, and each barcode sequence that was sequenced within the same molecule being annotated as belonging to the same multimeric barcode molecule in the library of multimeric barcode molecules. A simple analysis script (Networkx; Python) was employed to determine overall multimeric barcode molecule barcode groups, by examining overlap of barcode-barcode pairs across different sequenced molecules. Several metrics of this data were made, including barcode length, sequence content, and the size and complexity of the multimeric barcode molecules across the library of multimeric barcode molecules.

Number of Nucleotides within Each Barcode Sequence

Each individual barcode sequence from each barcode molecule, contained within each Illumina-sequenced molecule was isolated, and the total length of each such barcode was determined by counting the number of nucleotides between the upstream adapter molecule sequence, and the downstream adapter molecule sequence. The results are shown in Figure 10.

The overwhelming majority of barcodes are 20 nucleotides long, which corresponds to five additions of our four-nucleotide-long sub-barcode molecules from our double-stranded sub-barcode library. This is thus the expected and desired result, and indicates that each 'cycle' of:

Ligation of Sub-Barcode Library to MlyI-Cleaved Solution, PCR Amplification of the Ligated Library, Uracil Glycosylase Enzyme Digestion, and MlyI Restriction Enzyme Cleavage, was successful and able to efficiently add new four-nucleotide sub-barcode molecules at each cycle, and then was successfully able to amplify and carry these molecules forward through the protocol

for continued further processing, including through the five total cycles of sub-barcode addition, to make the final, upstream-adaptor-ligated libraries.

We also used this sequence analysis method to quantitate the total number of unique barcodes in total, across all sequenced multimeric barcode molecules: this amounted to 19,953,626 total unique barcodes, which is essentially identical to the 20 million barcodes that would be expected, given that we synthesised 2 million multimeric barcode molecules, each with approximately 10 individual barcode molecules.

Together, this data and analysis thus shows that the methods of creating complex, combinatoric barcodes from sub-barcode sequences is effective and useful for the purpose of synthesising multimeric barcode molecules.

Total Number of Unique Barcode Molecules in Each Multimeric Barcode Molecule

Figure 11 shows the results of the quantification of the total number of unique barcode molecules (as determined by their respective barcode sequences) in each sequenced multimeric barcode molecule. As described above, to do this we examined, in the first case, barcode sequences which were present and detected within the same individual molecules sequenced on the sequencer. We then employed an additional step of clustering barcode sequences further, wherein we employed a simple network analysis script (Networkx) which can determine links between individual barcode sequences based both upon explicit knowledge of links (wherein the barcodes are found within the same, contiguous sequenced molecule), and can also determine 'implicit' links, wherein two or more barcodes, which are not sequenced within the same sequenced molecule, instead both share a direct link to a common, third barcode sequence (this shared, common link thus dictating that the two first barcode sequences are in fact located on the same multimeric barcode molecule).

This figure shows that the majority of multimeric barcode molecules sequenced within our reaction have two or more unique barcodes contained therein, thus showing that, through our Overlap-Extension PCR linking process, we are able to link together multiple barcode molecules into multimeric barcode molecules. Whilst we would expect to see more multimeric barcode molecules exhibiting closer to the expected number of barcode molecules (10), we expect that this observed effect is due to insufficiently high sequencing depth, and that with a greater number of sequenced molecules, we would be able to observe a greater fraction of the true links between individual barcode molecules. This data nonetheless suggest that the fundamental synthesis procedure we describe here is efficacious for the intended purpose.

Representative Multimeric Barcode Molecules

– 244 –

Figure 12 shows representative multimeric barcode molecules that have been detected by our analysis script. In this figure, each 'node' is a single barcode molecule (from its associated barcode sequence), each line is a 'direct link' between two barcode molecules that have been sequenced at least once in the same sequenced molecule, and each cluster of nodes is an individual multimeric barcode molecule, containing both barcodes with direct links and those within implicit, indirect links as determined by our analysis script. The inset figure includes a single multimeric barcode molecule, and the sequences of its constituent barcode molecules contained therein.

This figure illustrates the our multimeric barcode molecule synthesis procedure: that we are able to construct barcode molecules from sub-barcode molecule libraries, that we are able to link multiple barcode molecules with an overlap-extension PCR reaction, that we are able to isolate a quantitatively known number of individual multimeric barcode molecules, and that we are able to amplify these and subject them to downstream analysis and use.

Barcoding Synthetic DNA Templates of Known Sequence with (i) Multimeric Barcoding Reagents Containing Barcoded Oligonucleotides, and (ii) Multimeric Barcoding Reagents and Separate Adapter Oligonucleotides

Sequence Extraction and Analysis

With scripting in Python and implemented in an Amazon Web Services (AWS) framework, for each sequence read following sample-demultiplexing, each barcode region from the given multimeric barcode reagent was isolated from its flanking upstream-adapter and downstream-adapter sequence. Likewise, each molecular sequence identifier region from the given synthetic DNA template molecule was isolated from its flanking upstream and downstream sequences. This process was repeated for each molecule in the sample library; a single filtering step was performed in which individual barcodes and molecular sequence identifiers that were present in only a single read (thus likely to represent either sequencing error or error from the enzymatic sample-preparation process) were censored from the data. For each molecular sequence identifier, the total number of unique (ie with different sequences) barcode regions found associated therewith within single sequence reads was quantitated. A histogram plot was then created to visualize the distribution of this number across all molecular sequence identifiers found in the library.

Discussion

Figure 13 shows the results of this analysis for Method 6 (Barcoding Synthetic DNA Templates of Known Sequence with Multimeric Barcoding Reagents Containing Barcoded Oligonucleotides). This figure makes clear that the majority of multimeric barcoding reagents are able to successfully label two or more of the tandemly-repeated copies of each molecular sequence identifier with

which they are associated. A distribution from 1 to approximately 5 or 6 'labelling events' is observed, indicating that there may be a degree of stochastic interactions that occur with this system, perhaps due to incomplete enzymatic reactions, or steric hindrance at barcode reagent/synthetic template interface, or other factors.

5

Figure 14 shows the results of this same analysis conducted using Method 7 (Barcoding Oligonucleotides Synthetic DNA Templates of Known Sequence with Multimeric Barcode Molecules and Separate Adapter Oligonucleotides). This figure also clearly shows that the majority of multimeric barcoding reagents are able to successfully label two or more of the tandemly-repeated copies of each molecular sequence identifier with which they are associated, with a similar distribution to that observed for the previous analysis.

10

Together, these two figures show that this framework for multimeric molecular barcoding is an effective one, and furthermore that the framework can be configured in different methodologic ways. Figure 13 shows results based on a method in which the framework is configured such that the multimeric barcode reagents already contain barcoded oligonucleotides, prior to their being contacted with a target (synthetic) DNA template. In contrast, Figure 14 shows results based on an alternative method in which the adapter oligonucleotides first contact the synthetic DNA template, and then in a subsequent step the adapter oligonucleotides are barcoded through contact with a multimeric barcode reagent. Together these figures demonstrate both the multimeric barcoding ability of these reagents, and their versatility in different key laboratory protocols.

15

To analyse whether, and the extent to which, individual multimeric barcoding reagents successfully label two or more sub-sequences of the same synthetic DNA template, the groups of different barcodes on each individual multimeric barcoding reagent in the library (as predicted from the Networkx analysis described in the preceding paragraph and as illustrated in Figure 12) was compared with the barcodes annealed and extended along single synthetic DNA templates (as described in Method 11). Each group of barcodes found on individual multimeric barcoding reagents was given a numeric 'reagent identifier label'. For each synthetic DNA template molecular sequence identifier (i.e., for each individual synthetic DNA template molecule) that was represented in the sequencing data of Method 11 by two or more barcodes (i.e., wherein two or more sub-sequences of the synthetic template molecule were annealed and extended by a barcoded oligonucleotide), the corresponding 'reagent identifier label' was determined. For each such synthetic template molecule, the total number of multimeric barcodes coming from the same, single multimeric barcoding reagent was then calculated (i.e., the number of different sub-sequences in the synthetic template molecule that were labeled by a *different* barcoded oligonucleotide but from the *same*, single multimeric barcoding reagent was calculated). This analysis was then repeated and compared with a 'negative control' condition, in which the

25

30

35

barcodes assigned to each 'reagent identifier label' were randomized (i.e. the same barcode sequences remain present in the data, but they no longer correspond to the actual molecular linkage of different barcode sequences across the library of multimeric barcoding reagents).

- 5 The data from this analysis is shown in Figure 17, for both the actual experimental data and for the control data with randomized barcode assignments (note the logarithmic scale of the vertical axis). As this figure shows, though the number of unique barcoding events per target synthetic DNA template molecule is small, they overlap almost perfectly with the known barcode content of individual multimeric barcoding reagents. That is, when compared with the randomized barcode data (which contains essentially no template molecules that appear to be 'multivalently barcoded'), the overwhelming majority (over 99.9%) of template molecules in the actual experiment that appear to be labeled by multiple barcoded oligonucleotides from the same, individual multimeric barcoding reagent, are in fact labeled multiply by the same, single reagents in solution. By contrast, if there were no non-random association between the different barcodes that labelled individual synthetic DNA templates (that is, if Figure 17 showed no difference between the actual experimental data and the randomized data), then this would have indicated that the barcoding had not occurred in a spatially-constrained manner as directed by the multimeric barcoding reagents. However, as explained above, the data indicates convincingly that the desired barcoding reactions did occur, in which sub-sequences found on single synthetic DNA templates interacted with (and were then barcoded by) only single, individual multimeric barcoding reagents.

Barcoding Genomic DNA Loci with Multimeric Barcoding Reagents Containing Barcoded Oligonucleotides

25

Sequence Extraction and Analysis

- As with other analysis, scripting was composed in Python and implemented in an Amazon Web Services (AWS) framework. For each sequence read following sample-demultiplexing, each barcode region from the given multimeric barcode reagent was isolated from its flanking upstream-adaptor and downstream-adaptor sequence and recorded independently for further analysis. Likewise, each sequence to the 3' end of the downstream region (representing sequence containing the barcoded oligonucleotide, and any sequences that the oligonucleotide had primed along during the experimental protocol) was isolated for further analysis. Each downstream sequence of each read was analysed for the presence of expected adapter oligonucleotide sequences (i.e. from the primers corresponding to one of the three genes to which the oligonucleotides were directed) and relevant additional downstream sequences. Each read was then recorded as being either 'on-target' (with sequence corresponding to one of the expected, targeted sequence) or 'off-target'. Furthermore, for each of the targeted regions, the total number of unique multimeric barcodes (i.e. with identical but duplicate barcodes merged into

a single-copy representation) was calculated. A schematic of each expected sequence read, and the constituent components thereof, is shown in Figure 16.

Discussion

5 Figure 15 shows the results of this analysis for this method, for four different independent samples. These four samples represent a method wherein the process of annealing the multimeric barcode reagents took place for either 3 hours, or overnight (approximately 12 hours). Further, for each of these two conditions, the method was performed either with the multimeric barcode reagents retained intact as originally synthesized, or with a modified protocol in which the
10 barcoded oligonucleotides are first denatured away from the barcode molecules themselves (through a high-temperature melting step). Each row represents a different amplicon target as indicated, and each cell represents the total number of unique barcode found associated with each amplicon in each of the four samples. Also listed is the total proportion of on-target reads, across all targets summed together, for each sample.

15 As seen in the figure, the majority of reads across all samples are on-target; however there is seen a large range in the number of unique barcode molecules observed for each amplicon target. These trends across different amplicons seem to be consistent across the different experimental conditions, and could be due to different priming (or mis-priming) efficiencies of the
20 different oligonucleotides, or different amplification efficiencies, or different mapping efficiencies, plus potential other factors acting independently or in combination. Furthermore, it is clear that the samples that were annealed for longer have a larger number of barcodes observed, likely due to more complete overall annealing of the multimeric reagents to their cognate genomic targets. And furthermore, the samples where the barcoded oligonucleotides were first denatured from the
25 barcode molecules show lower overall numbers of unique barcodes, perhaps owing to an avidity effect wherein fully assembled barcode molecules can more effectively anneal clusters of primers to nearby genomic targets at the same locus. In any case, taken together, this figure illustrates the capacity of multimeric reagents to label genomic DNA molecules, across a large number of molecules simultaneously, and to do so whether the barcoded oligonucleotides remain bound on
30 the multimeric barcoding reagents or whether they have been denatured therefrom and thus potentially able to diffuse more readily in solution.

EXAMPLE 2

35 MATERIALS AND METHODS FOR LINKING SEQUENCES FROM MICROPARTICLES

All experimental steps are conducted in a contamination-controlled laboratory environment, including the use of standard physical laboratory separations (E.g. pre-PCR and post-PCR laboratories).

Protocol for Isolating a Microparticle Specimen

A standard blood sample (e.g. 5-15 mL in total) is taken from a subject, and processed with a blood fractionation method using EDTA-containing tubes to isolate the plasma fraction, using centrifugation at 800 x G for 10 minutes. Then a cellular plasma fraction is then carefully isolated and centrifuged at 800 x G for 10 minutes to pellet remaining intact cells. The supernatant is then carefully isolated for further processing. The supernatant is then centrifuged at 3000 x G for 30 minutes to pellet a microparticle fraction (a high-speed centrifugation mode at 20,000 x G for 30 minutes is used to pellet a higher-concentration microparticle specimen); then the resulting supernatant is carefully removed, and the pellet is resuspended in an appropriate buffer for the following processing step. An aliquot from the resuspended pellet is taken and used to quantitate the concentration of DNA in the resuspended pellet (e.g. using a standard fluorescent nucleic acid staining method such as PicoGreen, ThermoFisher Scientific). The specimen is adjusted in volume to achieve an appropriate concentration for subsequent processing steps.

Protocol for Partitioning and PCR-Amplification

Following the process of isolating a microparticle specimen as above, the pellet is resuspended in a PCR buffer comprising a full solution of 1X PCR buffer, PCR polymerase enzyme, dNTPs, and a set of primer pairs; a polymerase and PCR buffer appropriate for direct PCR is employed. This resuspending step is performed such that each 5 microliters of the resuspended solution contains approximately 0.1 picograms of DNA from the microparticle specimen itself. A panel of 5-10 primer pairs (a greater number is used for larger amplicon panels) covering one or more gene targets is designed using a multiplex PCR design algorithm (e.g. PrimerPlex; PREMIER Biosoft) to minimise cross-priming and to achieve approximately equal annealing temperatures across all primers; each amplicon length is locked between 70 and 120 nucleotides; each forward primer has a constant forward adapter sequence at its 5' end, and each reverse primer has a constant reverse adapter sequence at its 5' end, and the primers are included in the polymerase reaction at equimolar concentrations. The resuspended sample is then spread across a set of PCR tubes (or individual wells in a 384-well plate format) with 5.0 microliters of the reaction solution included in each tube/well; up to 384 or more individual reactions are performed as the total amount of DNA in the microparticle specimen allows; 10-15 PCR cycles are performed for subsequent barcoding with barcoded oligonucleotides; 22-28 PCR cycles are performed for subsequent barcoding with multimeric barcoding reagents.

Protocol for Barcoding with Barcoded Oligonucleotides

Following the protocol of PCR amplification as above, barcoded oligonucleotides are added to each well, with each forward barcoded oligonucleotide comprising the forward adapter sequence at its 3' end, a forward (read 1) Illumina sequencing primer sequence on its 5' end, and a 6-nucleotide barcode sequence between the two; a reverse primer containing a reverse (read 2) Illumina amplification sequence on its 5' end and the reverse adapter sequence at its 3' end is

– 249 –

used. A different single barcoded oligonucleotide (i.e. containing a different barcode sequence) is used for each well. The PCR reaction volume is adjusted to 50 microliters to dilute the target-specific primers, and 8-12 PCR cycles are performed to append barcode sequences to the sequences within each tube/well. The amplification products from each well are purified using a SPRI cleanup/size-selection step (Agencourt Ampure XP, Beckman-Coulter Genomics), and the resulting purified products from all wells are merged into a single solution. A final PCR reaction using the full-length Illumina amplification primers (PE PCR Primer 1.0/2.0) is performed for 7-12 cycles to amplify the merged products to the appropriate concentration for loading onto an Illumina flowcell, and the resulting reaction is SPRI purified/size-selected and quantitated.

Protocol for Barcoding with Multimeric Barcoding Reagents

To append barcode sequences with multimeric barcoding reagents, following the process of PCR amplification as above, PCR amplification products from individual wells are purified with a SPRI purification step, and then resuspended in 1X PCR reaction buffer (with dNTPs) in individual wells without merging or cross-contaminating the samples from different wells. From a library of at least 10 million different multimeric barcoding reagents, an aliquot containing approximately 5 multimeric barcoding reagents is then added to each well, wherein each multimeric barcoding reagent is a contiguous multimeric barcode molecule made of 10-30 individual barcode molecules, with each barcode molecule comprising a barcode region with a different sequence from the other barcode molecules, and with a barcoded oligonucleotide annealed to each barcode molecule. Each barcoded oligonucleotide contains a forward (read 1) Illumina sequencing primer sequence on its 5' end, and the forward adapter sequence (also contained in the forward PCR primers) at its 3' end, with its barcode sequence within the middle section. A reverse primer containing a reverse (read 2) Illumina amplification sequence on its 5' end and the reverse adapter sequence at its 3' end is also included in the reaction mixture. A hot-start polymerase is used for this barcode-appending reaction. The polymerase is first activated at its activation temperature, and then 5-10 PCR cycles are performed with the annealing step performed at the forward/reverse adapter annealing temperature to extend the barcoded oligonucleotides along the PCR-amplified products, and to extend the reverse Illumina amplification sequence to these primer-extension products. The resulting products from each well are purified using a SPRI cleanup/size-selection, and the resulting purified products from all wells are merged into a single solution. A final PCR reaction using the full-length Illumina amplification primers (PE PCR Primer 1.0/2.0) is performed for 7-12 cycles to amplify the merged products to the appropriate concentration for loading onto an Illumina flowcell, and the resulting reaction is SPRI purified/size-selected and quantitated.

Protocol for Sequencing and Informatic Analysis

Following barcoding and amplification protocols, amplified samples are quantitated and sequenced on Illumina sequencers (e.g. HiSeq 2500). Prior to loading, samples are combined

– 250 –

with sequencer-ready phiX genomic DNA libraries such that phiX molecules comprise 50-70% of the final molar fraction of the combined libraries. Combined samples are then each loaded onto one or more lanes of the flowcell at the recommended concentration for clustering. Samples are sequenced to a read depth wherein each individual barcoded sequence is sequenced on average
5 by 5-10 reads, using paired-end 2x100 sequencing cycles. Raw sequences are then quality-trimmed and length-trimmed, constant adapter/primer sequences are trimmed away, and the genomic DNA sequences and barcode sequences from each retained sequence read are isolated informatically. Linked sequences are determined by detecting genomic DNA sequences that are appended to the same barcode sequence, or appended to different barcode sequences from the
10 same set of barcode sequences (i.e. from the same multimeric barcoding reagent).

Protocol for Barcoding Fragments of Genomic DNA using Barcoded Oligonucleotides

To isolate circulating microparticles from whole blood, 1.0 milliliters of whole human blood (collected with K2 EDTA tubes) were added to each of two 1.5 milliliter Eppendorf DNA Lo-Bind
15 tubes, and centrifuged in a desktop microcentrifuge for 5 minutes at 500xG; the resulting top (supernatant) layer (approximately 400 microliters from each tube) were then added to new 1.5 milliliter Eppendorf DNA Lo-Bind tubes, and again centrifuged in a desktop microcentrifuge for 5 minutes at 500xG; the resulting top (supernatant) layer (approximately 300 microliters from each tube) were then added to new 1.5 milliliter Eppendorf DNA Lo-Bind tubes, and centrifuged in a
20 desktop microcentrifuge for 15 minutes at 3000xG; the resulting supernatant layer was fully and carefully aspirated, and the pellet in each tube was resuspend in 10 microliters Phosphate-Buffered Saline (PBS) and then the two 10 microliter resuspended samples were merged into a single 20 microliter sample (producing the sample for 'Variant A' of the present method).

25 In a related variant of the method ('Variant C'), an aliquot of this original 20 microliter sample was transferred to a new 1.5 milliliter Eppendorf DNA Lo-Bind tube, and centrifuged for 5 minutes at 1500xG, with the resulting pellet then resuspended in PBS and aliquoted into low-concentration solutions as described below.

30 Circulating microparticles within the aforementioned 20 microliter sample (and/or from the resuspend 'Variant C' sample) were then partitioned prior to appending barcoded oligonucleotides. To partition low numbers of circulating microparticles per partition, the 20-microliter sample was aliquoted into solutions containing lower microparticle concentrations; 8 solutions with different concentrations were used, with the first being the original (undiluted) 20-
35 microliter sample, and each of the subsequent 7 solutions having a 2.5-fold lower microparticle concentration (in PBS) relative to the preceding solution. A 0.5 microliter aliquot of each solution was then added to 9.5 microliters of 1.22X 'NEBNext Ultra II End Prep Reaction Buffer' (New England Biolabs) in H₂O in 200 microliter PCR tubes (Flat cap; from Axygen) and mixed gently. To permeabilise the microparticles, tubes were heated at 65 degrees Celsius for 30 minutes on a

- 251 -

thermal cycler with a heated lid. To each tube was added 0.5 microliters 'NEBNext Ultra II End Prep Enzyme Mix' and mixed the solutions were mixed gently; the solutions were incubated at 20 degrees Celsius for 30 minutes and then 65 degrees Celsius for 30 minutes on a thermal cycler.

- 5 To each tube was added 5.0 microliters 'NEBNext Ultra II Ligation Master Mix', and 0.33 microliters 0.5X (in H₂O) 'NEBNext Ligation Enhancer', and 0.42 microliters 0.04X (in 0.1X NEBuffer 3) 'NEBNext Adapter', and the solutions were mixed gently; the solutions were then incubated at 20 degrees Celsius for 15 minutes (or for 2 hours in "Variant B" of this method) on a thermal cycler with the heated lid turned off. To each tube was added 0.5 microliters 'NEBNext
10 USER Enzyme', and the solutions were mixed gently; the solutions were then incubated at 20 degrees Celsius for 20 minutes at 37 degrees Celsius for 30 minutes on a thermal cycler with a heated lid set to 50 degrees Celsius, and then held at 4 degrees Celsius. Each reaction was then purified with 1.1X-volume Ampure XP SPRI beads (Agencourt; as per manufacturer's instructions) and eluted in 21.0 microliters H₂O. This process of ligating 'NEBNext Adapter' sequences to
15 fragments of genomic DNA from partitioned circulating microparticles provides a process of appending a coupling sequence to said fragments (wherein the 'NEBNext Adapter' itself, which comprises partially double-stranded and partially single-stranded sequences, comprises said coupling sequences, wherein the process of appending coupling sequence is performed with a ligation reaction). In a subsequent step of the process, barcoded oligonucleotides are appended
20 to fragments of genomic DNA from partitioned circulating microparticles with an annealing and extension process (performed via a PCR reaction).

- In 'Variant B' of this method, following the above USER enzyme step but prior to Ampure XP purification, the USER-digested samples were added to 50.0 microliters 'NEBNext Ultra II Q5
25 Master Mix', and 2.5 microliters 'Universal PCR Primer for Illumina', and 2.5 microliters of a specific 'NEBNext Index Primer' [from NEBNext Multiplex Oligos Index Primers Set 1 or Index Primers Set 2], and 28.2 microliters H₂O, and the solutions were mixed gently, and then amplified by 5 cycles PCR in a thermal cycler, with each cycle being: 98 degrees Celsius for 20 seconds, and 65 degrees Celsius for 3 minutes. Each reaction was then purified with 0.95X-volume
30 Ampure XP SPRI beads (Agencourt; as per manufacturer's instructions) and eluted in 21.0 microliters H₂O.

- Ampure XP-purified solutions (either following USER-digestion or following the initial PCR amplification process for 'Variant B' of the methods) (20.0 microliters each) were then added to
35 25.0 microliters 'NEBNext Ultra II Q5 Master Mix', and 2.5 microliters 'Universal PCR Primer for Illumina', and 2.5 microliters of a specific 'NEBNext Index Primer', and the solutions were mixed gently, and then amplified by 28 (Or 26 cycles for Variant B) cycles PCR in a thermal cycler, with each cycle being: 98 degrees Celsius for 10 seconds, and 65 degrees Celsius for 75 seconds; with a single final extension step of 75 degrees Celsius for 5 minutes. Each reaction was then

– 252 –

purified with 0.9X-volume Ampure XP SPRI beads (Agencourt; as per manufacturer's instructions) and eluted in 25.0 microliters H₂O. These steps of PCR append barcode sequences to the sequences of fragments of genomic DNA from circulating microparticles, wherein the barcode sequences are comprised within barcoded oligonucleotides (i.e. comprised within the specific 'NEBNext Index Primer' employed within each PCR reaction). In each primer-binding and extension step of the PCR reactions, the barcoded oligonucleotides hybridise to coupling sequences (e.g. the sequences within the 'NEBNext Adapter') and then are used to prime an extension step, wherein the 3' end of the barcoded oligonucleotide is extended to produce a sequence comprising both the barcode sequence and a sequence of a fragment of genomic DNA from a circulating microparticle. One barcoded oligonucleotide (and thus one barcode sequence) was employed per PCR reaction, with different barcode sequences used for each of the different PCR reactions. Therefore, sequences of fragments of genomic DNA from circulating microparticles in each partition were appended to a single barcode sequence, which links the set of sequences from the partition. The set of sequences in each of the partitions was linked by a different barcode sequence.

To create a negative-control sample, a separate 20-microliter sample of circulating microparticles was prepared as in the first paragraph above, but then the fragments of genomic DNA therein were isolated and purified with a Qiagen DNEasy purification kit (using the spin-column and centrifugation protocol as per the Qiagen manufacturer's instructions), and eluted in 50 microliters H₂O, and then being processed with the NEBNext End Prep, Ligation, USER, and PCR processing steps as described above. This negative-control sample was employed to analyse the sequencing signals and readouts wherein fragments of genomic DNA from a very large number of circulating microparticles are analysed (i.e. wherein no linking of sequences from one or a small number of circulating microparticles has been performed).

Following the above steps of centrifuging and partitioning circulating microparticles, and then appending coupling sequences, appending barcode sequences, and PCR amplification and purification, several barcoded libraries comprising sequences from fragments of genomic DNA from circulating microparticles were then merged and sequenced on a Mid-Output Illumina NextSeq 500 flowcell for 150 cycles performed with paired-end reads (100x50), plus a separate (forward-direction) Index Read (to determine the barcode sequences appended with the barcoded oligonucleotides). Typically, between 6 and 12 barcoded libraries (i.e. comprising one barcoded set of linked sequences per library) were merged and sequenced per flowcell; coverage of at least several million total reads were achieved per barcoded library. Sequence reads were demultiplexed according to the barcode within the index read, sequences from each barcoded partition were mapped with Bowtie2 to the reference human genome sequence (hg38), and then mapped (and de-duplicated) sequences were imported into Seqmonk (version 1.39.0) for visualisation, quantitation, and analysis. In typical representative analyses, reads were mapped

into sliding windows of 500 Kb along each human chromosome and then the total number of reads across each such window were quantitated and visualised.

5 Key experimental results of these barcoded oligonucleotide methods are shown in figures 25-29, and described in further detail here:

10 Figure 25 illustrates the linkage of sequences of fragments of genomic DNA within a representative circulating microparticle, as produced by a method of appending barcoded oligonucleotides (from the 'Variant A' version of the example protocol). Shown is the density of sequence reads across all chromosomes in the human genome within 500 kilobase (Kb) sliding windows tiled across each chromosome. Two clear, self-contained clusters of reads are observed, approximately 200 Kb and 500 Kb in total span respectively. Notably, both of the two read clusters are on the same chromosome, and furthermore are from nearby portions of the same chromosome arm (on chromosome 14), thus confirming the suspicion that, indeed, multiple
15 intramolecular chromosomal structures may be packaged into singular circulating microparticles, whereupon fragments of genomic DNA derived therefrom circulate within the human vasculature.

20 Figure 26 also illustrates the linkage of sequences of fragments of genomic DNA within a circulating microparticle, but as produced by a variant method of appending barcoded oligonucleotides (from the 'Variant B' version of the example protocol) wherein the duration of ligation is increased relative to 'Variant A'. Shown again is the density of sequence reads across all chromosomes in the human genome, with clear clustering of reads within singular chromosomal segments (on chromosome 1 and chromosome 12 respectively). It is possible that the partition employed in this experiment comprised two different microparticles, in which case it is likely that one read cluster arose from each microparticle; alternatively, it is possible that a
25 single microparticle contained a read cluster from each of chromosomes 1 and 12, which would thus demonstrate that inter-molecular chromosomal structures may also be packaged into singular circulating microparticles which then circulate through the blood.

30 Figure 27 illustrates the linkage of sequences of fragments of genomic DNA within a circulating microparticle, as produced by a method of appending barcoded oligonucleotides (from the 'Variant B' version of the example protocol). Shown are the actual sequence reads (of the read cluster from chromosome 12 from figure 26) zoomed in within a large and then within a small chromosomal segment, to show the focal, high-density nature of these linked reads, and to
35 demonstrate the fact that the read clusters comprise clear, contiguous clusters of sequences from individual chromosome molecules from single cells, even down to the level of demonstrating immediately adjacent, non-overlapping, nucleosomally-positioned fragments.

– 254 –

Figure 28 illustrates the linkage of sequences of fragments of genomic DNA within a circulating microparticle, as produced by a method of appending barcoded oligonucleotides (from the 'Variant C' version of the example protocol). In contrast to Variant A and Variant B, this Variant C experiment employed a lower-speed centrifugation process to isolate a different, larger population of circulating microparticles compared with the other two variants. Shown is the density of sequence reads across all chromosomes in the human genome, from this experiment, again with clear clustering of reads observed within singular chromosomal segments. However, such segments are clearly larger in chromosomal span than in the other Variant methods (due to the larger microparticles being pelleted within Variant C compared with Variants A or B).

Figure 29 illustrates a negative-control experiment, wherein fragments of genomic DNA are purified with a cleanup kit (Qiagen DNEasy Spin Column Kit) (i.e. therefore being unlinked) before being appending to barcoded oligonucleotides as in the 'Variant A' protocol. As would be expected given the input sample of unlinked reads, no clustering of reads is observed at all (rather, what reads do exist are dispersed randomly and essentially evenly throughout all chromosomal regions of the genome), validating that circulating microparticles comprise fragments of genomic DNA from focal, contiguous genomic regions within individual chromosomes. Even with further random sampling/sub-sampling of reads from said control library, no read clusters are observed.

EXAMPLE 3

MATERIALS AND METHODS FOR MEASURING SETS OF LINKED SIGNALS FROM TARGET BIOMOLECULES

Protocol for CD2 Protein Measurement and Selection

To measure CD2 protein levels on circulating microparticles, microparticles were isolated and resuspended in phosphate buffered saline (PBS) as described above, and were then incubated with 10 uL washed CD2 Dynabeads (Invitrogen, catalogue number 11159D) for 20 minutes at 4 degrees Celcius. Following bead-sample incubation and binding, the reaction mixture was bound by a magnet and the resulting supernatant (bead-unbound) phase containing 'CD2-negative' circulating microparticles was aspirated and transferred to a new tube, and the beads with bound 'CD2-positive' circulating microparticles was released from the magnet and resuspended in PBS. The CD2-negative and CD2-positive were then partitioned and aliquoted into low-concentration solutions as described above and then individual aliquots were barcoded and prepared for sequencing with a NEBNext sample-preparation kit as described above; a fraction of the CD2-negative was also then further processed for methylation and PMCA measurement as described below.

Protocol for Measurement and Enrichment of 5-Methylcytosine-Modified DNA

To measure 5-methylcytosine-modified DNA within fragments of genomic DNA within circulating microparticles, CD2-negative microparticles were isolated as described above, and then
5 partitioned and aliquoted as described above, and then fragments of genomic DNA from the aliquoted and partitioned microparticles were released from said microparticles by incubation at 65 degrees Celsius for 30 minutes as described above, and then the ends of the fragments of genomic DNA were end-repaired, A-tailed, ligated to adapters and then digested with USER
10 enzyme with a NEBNext sample-preparation kit as described above, and then samples were diluted 5-fold by volume in 1X CutSmart buffer (New England Biolabs) and then digested at 37 degrees Celsius for 30 minutes with 1.0 uL HpaII enzyme (New England Biolabs), which digests unmethylated DNA at CCGG sites but which is inhibited from digesting by methylated CCGG sites, thus enriching for fragments of DNA comprising methylated CCGG sequences compared with unmethylated CCGG sequences. The resulting samples were then PCR-amplified with
15 partition barcodes using a 'NEBNext Ultra II Q5 Master Mix' and 'NEBNext Index Primers' and then cleaned up with Ampure XP beads as described previously. Resulting barcoded and amplified samples were quantitated, pooled, and sequenced on a V2 2x25 basepair MiSeq flowcell (Illumina) such that each individual barcoded sample produced approximately 1 million total sequence reads; data was mapped with Bowtie2 (in the Galaxy cloud-based informatics
20 suite) to the human reference sequence and analysed further in SeqMonk genomics software as described previously.

Synthesis of Barcoded Affinity Probes

To synthesise barcoded affinity probes against PMCA (Plasma membrane calcium ATPase protein), two complementary oligonucleotides were synthesised (PolyT_5AM_3dT_1 and
25 PolyT_5AM_3dT_COMPL1 by Integrated DNA Technologies), with each comprising outer forward and reverse sequences for the NEBNext Index primers and an internal synthetic barcode sequence, and each blocked on the 3' end with an inverted dT base, and with PolyT_5AM_3dT_1 comprising a 5' C12 amino modifier (for activation and conjugation to an
30 antibody). The oligonucleotides were annealed to each other using a slow primer-annealing cycle on a thermal cycler, cleaned up with 2.8X Ampure XP beads, and resuspended in H₂O. and then 100 microliters of 42 micromolar purified, annealed oligonucleotide was conjugated to 100 micrograms of an affinity-purified monoclonal antibody against human PMCA protein (ab2783, Abcam) with the ThunderLink PLUS Oligo Conjugation System (Expedeon, catalogue number
35 425-0300) as per manufacturer's directions, with activated oligo material conjugated to activated antibody material at a 1:2 volumetric ratio, and then diluted 1:400 in PBS, and then used as a barcoded affinity probe for PMCA measurement as below.

- 256 -

PolyT_5AM_3dT_1:

/5AmMC12/TTCCCTACACGACGCTCTTCCGATCTCAGTTAGATACAACGTGACCTGA
GCAGTCTTAGCG AGATCGGAAGAGCACACGTCTGAACT*C*/3InvdT/

5

PolyT_5AM_3dT_COMPL1:

G*A*GTTTCAGACGTGTGCTCTTCCGATCTCGCTAAGACTGCTCAGGTCACGTTGTAT
CTAACTGAGATCGGAAGAGCGTCGTGTAGGGA*A*/3InvdT/

10

In the above sequences:

* = phosphorothioate bond

/5AmMC12/ = 5-prime terminal amino modifier with C12 linker

15

/3InvdT/ = 3-prime terminal inverted dT base

Protocol for PMCA Protein Measurement

To measure PMCA protein levels on circulating microparticles, CD2-negative microparticles were isolated as described above, and then 20 microliters CD2-negative microparticles were incubated with 1.0 microliter of 1:400 diluted barcoded affinity probe against PMCA for 30 minutes at 4 degrees Celsius. The sample was then centrifuged at 3000xG for 15 minutes at room temperature, the supernatant was aspirated (with care taken not to disturb the pellet), and the pellet was washed with 300 microliters PBS and then again centrifuged at 3000xG for 15 minutes at room temperature, with the supernatant again aspirated (with care again taken not to disturb the pellet), and the resulting washed, barcoded affinity probe-bound microparticle sample was resuspended in 25 microliters PBS. The resulting microparticle sample was then partitioned and aliquoted into low-concentration solutions as described above and then individual aliquots were barcoded and prepared for sequencing with a NEBNext sample-preparation kit as described above. The resulting xsamples were then PCR-amplified with partition barcodes using a 'NEBNext Ultra II Q5 Master Mix' and 'NEBNext Index Primers' and then cleaned up with Ampure XP beads as described previously. Resulting barcoded and amplified samples were quantitated, pooled, and sequenced on a V2 2x25 basepair MiSeq flowcell (Illumina) such that each individual barcoded sample produced approximately 1 million total sequence reads; data was mapped with Bowtie 2 (in the Galaxy cloud-based informatics suite) to the human reference sequence and analysed further in SeqMonk genomics software as described previously. Reads comprising the internal synthetic barcode sequences from PMCA barcoded affinity probes were detected, quantitated and analysed separately for each barcoded library.

In Figure 33, at the top of the figure is shown the schematic of an experimental method wherein a sample of circulating microparticles is generated and then incubated with a solution of beads, wherein the beads are conjugated to antibodies for the CD2 protein (which is found on the membrane of a subset of immune cells and on microparticles that will derive therefrom). Following a process of allowing CD2-positive microparticles (ie microparticles with a high concentration of

- 257 -

CD2 protein on their surface) to bind to the anti-CD2 beads, a magnet is used to collect the beads and the microparticles bound thereto (thus performing a measurement of and selection for CD2 protein comprised on the beads). The supernatant (comprising CD2-negative microparticles) and the bead-bound fraction (containing CD2-positive microparticles) are then diluted and partitioned into partitions, and the nucleic acid content (i.e. fragments of genomic DNA) comprised within each partition is appended to a partition-associated barcode, and then barcoded nucleic acids across several partitions are pooled and sequenced.

At the bottom of the figure is shown sequences of fragments of genomic DNA within two representative circulating microparticle partitions, as produced by a method of appending barcoded oligonucleotides, and taken from the CD2-positive pool (left) and from the CD2-negative pool. Shown is the density of sequence reads across all chromosomes in the human genome within 2 Megabase (Mb) sliding windows tiled across each chromosome. Clear, self-contained clusters of reads are observed, of varying but large sizes, showing that measurement of a target polypeptide (CD2 in this example) from circulating microparticles, combined with measurement of many linked fragments of genomic DNA, is achievable by these experimental methods.

In Figure 34, at the top of the figure is shown the schematic of an experimental method wherein a sample of circulating microparticles is generated and then incubated with a solution of beads, wherein the beads are conjugated to antibodies for the CD2 protein (which is found on the membrane of a subset of immune cells and on microparticles that will derive therefrom). Following a process of allowing CD2-positive microparticles (ie microparticles with a high concentration of CD2 protein on their surface) to bind to the anti-CD2 beads, a magnet is used to collect the beads and the microparticles bound thereto (thus performing a measurement of and selection for CD2 protein comprised on the beads). The supernatant (comprising CD2-negative microparticles) fraction is then diluted and partitioned into partitions, and the nucleic acid content (i.e. fragments of genomic DNA) comprised within each partition is then digested with a 5-methylcytosine-sensitive restriction enzyme (HpaII, which digests at unmethylated CCGG DNA sites but which is inhibited by cytosine methylation), to thus enrich for fragments of genomic DNA which are unmethylated at CCGG sites (thus performing a measurement of 5-methylcytosine-modified DNA). The resulting un-digested, non-methylated-enriched DNA fragments are then appended to a partition-associated barcode, and then barcoded nucleic acids across several partitions are pooled and sequenced.

At the bottom left of the figure is shown sequences of fragments of genomic DNA within a representative circulating microparticle partition, as produced by a method of appending barcoded oligonucleotides, and taken from the CD2-negative pool following depletion of unmethylated DNA fragments by HpaII digestion. Shown is the density of sequence reads across all chromosomes in the human genome within 2 Megabase (Mb) sliding windows tiled across each chromosome. At

right is a plot of the percentage of sequence reads containing CCGG sequences, within 4 control (undigested) libraries and 4 HpaII-digested libraries (enriched for methylated CCGG DNA). As expected, the digested libraries exhibit a small but clear depletion of CCGG sequences fractionally within the library, which will correspond to the molecular depletion of unmethylated CCGG-containing fragments in the HpaII samples, thus showing that the methods are cumulatively able to measure polypeptides, and fragments of genomic DNA, and modified DNA nucleotides, from circulating microparticles.

In Figure 35, at the top of the figure is shown the schematic of an experimental method wherein a sample of circulating microparticles is generated and then incubated with a solution of beads, wherein the beads are conjugated to antibodies for the CD2 protein (which is found on the membrane of a subset of immune cells and on microparticles that will derive therefrom). Following a process of allowing CD2-positive microparticles (ie microparticles with a high concentration of CD2 protein on their surface) to bind to the anti-CD2 beads, a magnet is used to collect the beads and the microparticles bound thereto (thus performing a measurement of and selection for CD2 protein comprised on the beads). The supernatant (comprising CD2-negative microparticles) fraction is then incubated with a solution of barcoded affinity probes comprising an antibody against PMCA (Plasma membrane calcium ATPase) protein and a barcoded oligonucleotide. The resulting barcoded affinity probe-bound microparticles are then pelleted by a centrifugation step and washed with PBS to remove unbound barcoded affinity probes. The resulting barcoded affinity probe-bound microparticles are then resuspended in PBS and diluted and partitioned into partitions, and the nucleic acid content (i.e. fragments of genomic DNA and sequences from barcoded affinity probes) comprised within each partition is then appended to a partition-associated barcode, and then barcoded nucleic acids across several partitions are pooled and sequenced.

At the bottom left of the figure is shown sequences of fragments of genomic DNA within a representative circulating microparticle partition, as produced by a method of appending barcoded oligonucleotides, and taken from the CD2-negative pool and then incorporating measurement of PMCA with barcoded affinity probes. Shown is the density of sequence reads across all chromosomes in the human genome within 2 Megabase (Mb) sliding windows tiled across each chromosome. At right is shown the number of sequence reads in each of 4 control samples (without barcoded affinity probe labelling) and 2 samples (i.e. circulating microparticle partitions) following a process of labelling with PMCA-targeted barcoded affinity probes. No sequence reads from the barcoded affinity probe are found in the control samples, but large quantitative amounts of sequences from the barcoded affinity probe are observed in each of the positive samples. Cumulatively these results illustrate that the methods are able to measure multiple polypeptides (including via use of barcoded affinity probes) and fragments of genomic DNA from circulating microparticles.

Various publications are cited herein, the disclosures of which are incorporated by reference in their entireties.

CLAIMS

1. A method of analysing a sample comprising a circulating microparticle or a sample derived from a circulating microparticle, wherein the circulating microparticle is a membranous vesicle, wherein the circulating microparticle comprises at least three target molecules, wherein at least two of the target molecules are fragments of genomic DNA and at least one of the target molecules is a target polypeptide, and wherein the method comprises measuring a signal corresponding to the presence, absence and/or level of each of the target molecules to produce a set of at least two linked signals for the circulating microparticle, wherein at least one of the linked signals corresponds to the presence, absence and/or level of the fragments of genomic DNA in the sample and at least one of the linked signals corresponds to the presence, absence and/or level of the target polypeptide in the sample, and wherein the step of measuring a signal corresponding to the presence, absence and/or level of the fragments of genomic DNA comprises linking at least two of the at least two fragments of genomic DNA to produce a set of at least two linked fragments of genomic DNA.
2. The method of claim 1, wherein the fragments of genomic DNA comprise a specific sequence of nucleotides and/or wherein the fragments of genomic DNA comprise at least one modified nucleotide or nucleobase, optionally wherein the modified nucleotide or nucleobase is 5-methylcytosine or 5-hydroxy-methylcytosine.
3. The method of claim 1 or claim 2, wherein the target polypeptide comprises a specific amino acid sequence and/or wherein the target polypeptide comprises a post-translational modification, optionally wherein the target polypeptide comprises an acetylated amino acid residue and/or a methylated amino acid residue.
4. The method of any one of claims 1-3, wherein the method comprises measuring the signal corresponding to the presence, absence and/or level of each of the target molecules of the circulating microparticle to produce a set of at least three linked signals for the circulating microparticle, wherein one of the linked signals corresponds to the presence, absence and/or level of a first fragment of genomic DNA of the circulating microparticle, one of the linked signals corresponds to the presence, absence and/or level of a second fragment of genomic DNA of the circulating microparticle, and one of the linked signals corresponds to the presence, absence and/or level of the target polypeptide of the circulating microparticle.
5. The method of any one of claims 1-4, wherein the step of measuring a signal corresponding to the presence, absence and/or level of the fragments of genomic DNA comprises analysing a sequence of each of at least two of the at least two fragments of genomic DNA, optionally wherein the step of measuring a signal corresponding to the presence, absence and/or level

- 261 -

of the fragments of genomic DNA comprises sequencing at least a portion of each of at least two of the at least two fragments of genomic DNA.

- 5 6. The method of any one of claims 1-5, wherein the step of measuring a signal corresponding to the presence, absence and/or level of the fragments of genomic DNA comprises sequencing at least a portion of each of at least two of the linked fragments in the set to produce at least two linked sequence reads.
- 10 7. The method of any one of claims 1-6, wherein the step of measuring a signal corresponding to the presence, absence and/or level of the fragments of genomic DNA comprises:
 - (a) appending each of at least two of the at least two fragments of genomic DNA of the circulating microparticle to a barcode sequence to produce a set of linked fragments of genomic DNA; and, optionally,
 - 15 (b) sequencing at least a portion of each of at least two of the linked fragments in the set to produce at least two linked sequence reads, wherein the at least two linked sequence reads are linked by the barcode sequence.
- 20 8. The method of any one of claims 1-6, wherein the step of measuring a signal corresponding to the presence, absence and/or level of the fragments of genomic DNA comprises:
 - (a) appending each of at least two of the at least two fragments of genomic DNA of the circulating microparticle to a different barcode sequence of a set of barcode sequences to produce a set of linked fragments of genomic DNA; and, optionally,
 - 25 (b) sequencing at least a portion of each of at least two of the linked fragments in the set to produce at least two linked sequence reads, wherein the at least two linked sequence reads are linked by the set of barcode sequences.
- 30 9. The method of any one of claims 1-8, wherein the fragments of genomic DNA comprise at least one modified nucleotide or nucleobase and wherein the step of measuring a signal corresponding to the presence, absence and/or level of the fragments of genomic DNA comprises measuring a signal corresponding to the presence, absence and/or level of the modified nucleotide or nucleobase of the fragments of genomic DNA, optionally wherein the modified nucleotide or nucleobase is 5-methylcytosine or 5-hydroxy-methylcytosine.
- 35 10. The method of claim 9, wherein the signal corresponding to the presence, absence and/or level of the modified nucleotide or nucleobase is measured using (i) a barcoded affinity probe, wherein the barcoded affinity probe comprises at least one affinity moiety linked to a barcoded oligonucleotide, wherein the barcoded oligonucleotide comprises at least one nucleotide, and wherein the affinity moiety is capable of binding to the modified nucleotide or nucleobase, optionally wherein the signal is measured by determining the presence, absence

- 262 -

and/or level of the barcoded oligonucleotide by sequencing; and/or (ii) an optically-labelled affinity probe and/or a fluorescently-labelled affinity probe, optionally wherein the signal is measured by flow cytometry and/or fluorescence-activated cell sorting.

- 5 11. The method of any one of claims 1-10, wherein the signal corresponding to the presence, absence and/or level of the target polypeptide is measured using (i) a barcoded affinity probe, wherein the barcoded affinity probe comprises at least one affinity moiety linked to a barcoded oligonucleotide, wherein the barcoded oligonucleotide comprises at least one nucleotide, and wherein the affinity moiety is capable of binding to the target polypeptide, optionally wherein the signal is measured by determining the presence, absence and/or level of the barcoded oligonucleotide by sequencing; and/or (ii) an optically-labelled affinity probe and/or a fluorescently-labelled affinity probe, optionally wherein the signal is measured by flow cytometry and/or fluorescence-activated cell sorting.
- 10
12. The method of any one of claims 1-11, wherein the circulating microparticle comprises at least 3, at least 4, at least 5, at least 10, at least 50, at least 100, at least 500, at least 1000, at least 5000, at least 10,000, at least 100,000, or at least 1,000,000 target molecules, and wherein the method comprises producing a set of at least 3, at least 4, at least 5, at least 10, at least 50, at least 100, at least 500, at least 1000, at least 5000, at least 10,000, at least 100,000, or at least 1,000,000 linked signals for the circulating microparticle.
- 15
13. The method of any one of claims 1-12, wherein the target molecules comprise at least 2, at least 3, at least 4, at least 9, at least 49, at least 99, at least 499, at least 999, at least 4999, at least 9,999, at least 99,999, or at least 999,999 fragments of genomic DNA, and optionally wherein the method comprises producing a set of at least 3, at least 4, at least 5, at least 10, at least 50, at least 100, at least 500, at least 1000, at least 5000, at least 10,000, at least 100,000, or at least 1,000,000 linked signals for the circulating microparticle.
- 20
14. The method of any one of claims 1-13, wherein the target molecules comprise at least 2, at least 3, at least 4, at least 9, at least 49, at least 99, at least 499, at least 999, at least 4999, at least 9,999, at least 99,999, or at least 999,999 target polypeptides, and optionally wherein the method comprises producing a set of at least at least 3, at least 4, at least 5, at least 10, at least 50, at least 100, at least 500, at least 1000, at least 5000, at least 10,000, at least 100,000, or at least 1,000,000 linked signals for the circulating microparticle.
- 25
15. The method of any one of claims 1-14, wherein the sample comprises first and second circulating microparticles, wherein each circulating microparticle comprises at least three target molecules as defined in any one of claims 1-14, and wherein the method comprises performing the step of measuring in accordance with any one of claims 1-14 to produce a set
- 30
- 35

- 263 -

of linked signals for the first circulating microparticle and performing the step of measuring in accordance with any one of claims 1-14 to produce a set of linked signals for the second circulating microparticle; optionally wherein the sample comprises n circulating microparticles, wherein each circulating microparticle comprises at least three target molecules as defined in any one of claims 1-14, and wherein the method comprises performing the step of measuring in accordance with any one of claims 1-14 for each circulating microparticle to produce a set of linked signals for each circulating microparticle, optionally wherein n is at least 3, at least 5, at least 10, at least 50, at least 100, at least 1000, at least 10,000, at least 100,000, at least 1,000,000, at least 10,000,000, or at least 100,000,000 circulating microparticles.

5

10

15