Other suitable amplification methods include, but are not limited to, ligase chain reaction (LCR) (see Wu and Wallace (1989) *Genomics* 4: 560, Landegren, *et al.* (1988) *Science* 241:1077, and Barringer et al. (1990) *Gene* 89: 117), transcription amplification (Kwoh, *et al.* (1989) *Proc. Natl. Acad. Sci. USA* 86: 1173), self-sustained sequence replication (Guatelli, *et al.* (1990) *Proc. Nat. Acad. Sci. USA* 87: 1874), dot PCR, and linker adapter PCR, *etc.*

Nucleic Acid Samples

A variety of tissue samples can be the source of the nucleic acid samples used in the present methods. Genomic or subgenomic DNA fragments can be isolated from a subject's sample (*e.g.*, a tumor sample, a normal adjacent tissue (NAT), a blood sample or any normal control)). In certain embodiments, the tissue sample is preserved as a frozen sample or as formaldehyde- or paraformaldehyde-fixed paraffin-embedded (FFPE) tissue preparation. For example, the sample can be embedded in a matrix, *e.g.*, an FFPE block or a frozen sample. The isolating step can include flow-sorting of individual chromosomes; and/or micro-dissecting a subject's sample (*e.g.*, a tumor sample, a NAT, a blood sample).

Protocols for DNA isolation from a tissue sample are known in the art. Additional methods to isolate nucleic acids (*e.g.*, DNA) from formaldehyde- or paraformaldehyde-fixed, paraffin-embedded (FFPE) tissues are disclosed, *e.g.*, in Cronin M. *et al.*, (2004) *Am J Pathol.* 164(1):35–42; Masuda N. *et al.*, (1999) *Nucleic Acids Res.* 27(22):4436–4443; Specht K. *et al.*, (2001) *Am J Pathol.* 158(2):419–429, Ambion RecoverAll™ Total Nucleic Acid Isolation Protocol (Ambion, Cat. No. AM1975, September 2008), and QIAamp® DNA FFPE Tissue Handbook (Qiagen, Cat. No. 37625, October 2007). RecoverAll™ Total Nucleic Acid Isolation Kit uses xylene at elevated temperatures to solubilize paraffin-embedded samples and a glass-fiber filter to capture nucleic acids. QIAamp® DNA FFPE Tissue Kit uses QIAamp® DNA Micro technology for purification of genomic and mitochondrial DNA.

The isolated nucleic acid samples (*e.g.*, genomic DNA samples) can be fragmented or sheared by practicing routine techniques. For example, genomic DNA can be fragmented by physical shearing methods, enzymatic cleavage methods, chemical cleavage methods, and other methods well known to those skilled in the art. The nucleic

IDT Exhibit 1020.3033

acid library can contain all or substantially all of the complexity of the genome. The term "substantially all" in this context refers to the possibility that there can in practice be some unwanted loss of genome complexity during the initial steps of the procedure. The methods described herein also are useful in cases where the nucleic acid library is a portion of the genome, *i.e.*, where the complexity of the genome is reduced by design. In some embodiments, any selected portion of the genome can be used with the methods described herein. In certain embodiments, the entire exome or a subset thereof is isolated.

Methods can further include isolating a nucleic acid sample to provide a library (*e.g.*, a nucleic acid library). In certain embodiments, the nucleic acid sample includes whole genomic, subgenomic fragments, or both. The isolated nucleic acid samples can be used to prepare nucleic acid libraries. Thus, in one embodiment, the methods featured in the invention further include isolating a nucleic acid sample to provide a library (*e.g.*, a nucleic acid library as described herein). Protocols for isolating and preparing libraries from whole genomic or subgenomic fragments are known in the art (*e.g.*, Illumina's genomic DNA sample preparation kit). In certain embodiments, the genomic or subgenomic DNA fragment is isolated from a subject's sample (*e.g.*, a tumor sample, a normal adjacent tissue (NAT), a blood sample or any normal control)). In one embodiment, the sample (*e.g.*, the tumor or NAT sample) is a preserved. For example, the sample is embedded in a matrix, *e.g.*, an FFPE block or a frozen sample. In certain embodiments, the isolating step includes flow-sorting of individual chromosomes; and/or microdissecting a subject's sample (*e.g.*, a tumor sample, a NAT, a blood sample). In certain embodiments, the nucleic acid sample used to generate the nucleic acid library is less than 5, less than 1 microgram, less than 500 ng, less than 200 ng, less than 100 ng, less than 50 ng or less than 20 ng (*e.g.*, 10 ng or less).

In still other embodiments, the nucleic acid sample used to generate the library includes RNA or cDNA derived from RNA. In some embodiments, the RNA includes total cellular RNA. In other embodiments, certain abundant RNA sequences (*e.g.*, ribosomal RNAs) have been depleted. In some embodiments, the poly(A)-tailed mRNA fraction in the total RNA preparation has been enriched. In some embodiments, the cDNA is produced by random-primed cDNA synthesis methods. In other embodiments, the cDNA synthesis is initiated at the poly(A) tail of mature mRNAs by priming by

IDT Exhibit 1020.3034

oligo(dT)-containing oligonucleotides. Methods for depletion, poly(A) enrichment, and cDNA synthesis are well known to those skilled in the art.

The method can further include amplifying the nucleic acid sample (*e.g.*, DNA or RNA sample) by specific or non-specific nucleic acid amplification methods that are well known to those skilled in the art. In some embodiments, certain embodiments, the nucleic acid sample is amplified, *e.g.*, by whole-genome amplification methods such as random-primed strand-displacement amplification.

In other embodiments, the nucleic acid sample is fragmented or sheared by physical or enzymatic methods and ligated to synthetic adapters, size-selected (*e.g.*, by preparative gel electrophoresis) and amplified (*e.g.*, by PCR). In other embodiments, the fragmented and adapter-ligated group of nucleic acids is used without explicit size selection or amplification prior to hybrid selection.

In other embodiments, the isolated DNA (*e.g.*, the genomic DNA) is fragmented or sheared. In some embodiments, the library includes less than 50% of genomic DNA, such as a subfraction of genomic DNA that is a reduced representation or a defined portion of a genome, *e.g.*, that has been subfractionated by other means. In other embodiments, the library includes all or substantially all genomic DNA.

In some embodiments, the library includes less than 50% of genomic DNA, such as a subfraction of genomic DNA that is a reduced representation or a defined portion of a genome, *e.g.*, that has been subfractionated by other means. In other embodiments, the library includes all or substantially all genomic DNA. Protocols for isolating and preparing libraries from whole genomic or subgenomic fragments are known in the art (*e.g.*, Illumina's genomic DNA sample preparation kit). Alternative DNA shearing methods can be more automatable and/or more efficient (*e.g.*, with degraded FFPE samples). Alternatives to DNA shearing methods can also be used to avoid a ligation step during library preparation.

The methods described herein can be performed using a small amount of nucleic acids, *e.g.*, when the amount of source DNA is limiting (*e.g.*, even after whole-genome amplification). In one embodiment, the nucleic acid comprises less than about 5 μg, 4 μg, 3 μg, 2 μg, 1 μg, 0.8 μg, 0.7 μg, 0.6 μg, 0.5 μg, or 400 ng, 300 ng, 200 ng, 100 ng, 50 ng, or 20 ng or less of nucleic acid sample. For example, to prepare 500 ng of hybridization-ready nucleic acids, one typically begins with 3 μg of genomic DNA. One

IDT Exhibit 1020.3035

can start with less, however, if one amplifies the genomic DNA (*e.g.*, using PCR) before the step of solution hybridization. Thus it is possible, but not essential, to amplify the genomic DNA before solution hybridization.

In some embodiments, a library is generated using DNA (*e.g.*, genomic DNA) from a sample tissue, and a corresponding library is generated with RNA (or cDNA) isolated from the same sample tissue.

Design of Baits

A bait can be a nucleic acid molecule, *e.g.*, a DNA or RNA molecule, which can hybridize to (*e.g.*, be complementary to), and thereby allow capture of a target nucleic acid. In one embodiment, a bait is an RNA molecule. In other embodiments, a bait includes a binding entity, *e.g.*, an affinity tag, that allows capture and separation, *e.g.*, by binding to a binding entity, of a hybrid formed by a bait and a nucleic acid hybridized to the bait. In one embodiment, a bait is suitable for solution phase hybridization.

Baits can be produced and used by methods and hybridization conditions as described in US 2010/0029498 and Gnirke, A. *et al.* (2009) *Nat Biotechnol.* 27(2):182-189, and USSN 61/428,568, filed December 30, 2010, incorporated herein by reference. For example, biotinylated RNA baits can be produced by obtaining a pool of synthetic long oligonucleotides, originally synthesized on a microarray, and amplifying the oligonucleotides to produce the bait sequences. In some embodiments, the baits are produced by adding an RNA polymerase promoter sequence at one end of the bait sequences, and synthesizing RNA sequences using RNA polymerase. In one embodiment, libraries of synthetic oligodeoxynucleotides can be obtained from commercial suppliers, such as Agilent Technologies, Inc., and amplified using known nucleic acid amplification methods.

Each bait sequence can include a target-specific (*e.g.*, a member-specific) bait sequence and universal tails on each end. As used herein, the term "bait sequence" can refer to the target-specific bait sequence or the entire oligonucleotide including the target-specific "bait sequence" and other nucleotides of the oligonucleotide. In one embodiment, a target-specific bait hybridizes to a nucleic acid sequence comprising a nucleic acid sequence in intron 26 of DCTN1, in intron 19 of ALK, or a fusion junction joining introns 26 of DCTN1 and 19 of ALK. In another embodiment, a target-specific

IDT Exhibit 1020.3036

bait hybridizes to a nucleic acid sequence comprising a nucleic acid sequence in intron 2 of LMNA, in intron 10 of NTRK1, or a fusion junction joining intron 2 of LMNA and intron 10 of NTRK1.

In one embodiment, the bait is an oligonucleotide about 200 nucleotides in length, of which 170 nucleotides are target-specific "bait sequence". The other 30 nucleotides (*e.g.*, 15 nucleotides on each end) are universal arbitrary tails used for PCR amplification. The tails can be any sequence selected by the user. For example, the pool of synthetic oligonucleotides can include oligonucleotides of the sequence of 5'-ATCGCACCAGCGTGTN$_{170}$CACTGCGGCTCCTCA-3' with N$_{170}$ indicating the target-specific bait sequences.

The bait sequences described herein can be used for selection of exons and short target sequences. In one embodiment, the bait is between about 100 nucleotides and 300 nucleotides in length. In another embodiment, the bait is between about 130 nucleotides and 230 nucleotides in length. In yet another embodiment, the bait is between about 150 nucleotides and 200 nucleotides in length. The target-specific sequences in the baits, *e.g.*, for selection of exons and short target sequences, are between about 40 nucleotides and 1000 nucleotides in length. In one embodiment, the target-specific sequence is between about 70 nucleotides and 300 nucleotides in length. In another embodiment, the target-specific sequence is between about 100 nucleotides and 200 nucleotides in length. In yet another embodiment, the target-specific sequence is between about 120 nucleotides and 170 nucleotides in length.

Sequencing

The invention also includes methods of sequencing nucleic acids. In one embodiment, any of a variety of sequencing reactions known in the art can be used to directly sequence at least a portion of a DCTN1-ALK or LMNA-NTRK1 fusion. In one embodiment, the DCTN1-ALK or LMNA-NTRK1fusion sequence is compared to a corresponding reference (control) sequence.

In one embodiment, the sequence of the fusion nucleic acid molecule is determined by a method that includes one or more of: hybridizing an oligonucleotide, *e.g.*, an allele specific oligonucleotide for one alteration described herein to said nucleic acid; hybridizing a primer, or a primer set (*e.g.*, a primer pair), that amplifies a region

IDT Exhibit 1020.3037

comprising the mutation or a fusion junction of the allele; amplifying, *e.g.*, specifically amplifying, a region comprising the mutation or a fusion junction of the allele; attaching an adapter oligonucleotide to one end of a nucleic acid that comprises the mutation or a fusion junction of the allele; generating an optical, *e.g.*, a colorimetric signal, specific to the presence of the one of the mutation or fusion junction; hybridizing a nucleic acid comprising the mutation or fusion junction to a second nucleic acid, *e.g.*, a second nucleic acid attached to a substrate; generating a signal, *e.g.*, an electrical or fluorescent signal, specific to the presence of the mutation or fusion junction; and incorporating a nucleotide into an oligonucleotide that is hybridized to a nucleic acid that contains the mutation or fusion junction.

In another embodiment, the sequence is determined by a method that comprises one or more of: determining the nucleotide sequence from an individual nucleic acid molecule, *e.g.*, where a signal corresponding to the sequence is derived from a single molecule as opposed, *e.g.*, from a sum of signals from a plurality of clonally expanded molecules; determining the nucleotide sequence of clonally expanded proxies for individual nucleic acid molecules; massively parallel short-read sequencing; template-based sequencing; pyrosequencing; real-time sequencing comprising imaging the continuous incorporation of dye-labeling nucleotides during DNA synthesis; nanopore sequencing; sequencing by hybridization; nano-transistor array based sequencing; polony sequencing; scanning tunneling microscopy (STM) based sequencing; or nanowire-molecule sensor based sequencing.

Any method of sequencing known in the art can be used. Exemplary sequencing reactions include those based on techniques developed by Maxam and Gilbert (*Proc. Natl Acad Sci USA* (1977) 74:560) or Sanger (Sanger *et al.* (1977) *Proc. Nat. Acad. Sci* 74:5463). Any of a variety of automated sequencing procedures can be utilized when performing the assays (*Biotechniques* (1995) 19:448), including sequencing by mass spectrometry (see, for example, U.S. Patent Number 5,547,835 and international patent application Publication Number WO 94/16101, entitled *DNA Sequencing by Mass Spectrometry* by H. Köster; U.S. Patent Number 5,547,835 and international patent application Publication Number WO 94/21822 entitled *DNA Sequencing by Mass Spectrometry Via Exonuclease Degradation* by H. Köster), and U.S. Patent Number 5,605,798 and International Patent Application No.

IDT Exhibit 1020.3038

PCT/US96/03651 entitled *DNA Diagnostics Based on Mass Spectrometry* by H. Köster; Cohen *et al.* (1996) *Adv Chromatogr* 36:127-162; and Griffin *et al.* (1993) *Appl Biochem Biotechnol* 38:147-159).

Sequencing of nucleic acid molecules can also be carried out using next-generation sequencing (NGS). Next-generation sequencing includes any sequencing method that determines the nucleotide sequence of either individual nucleic acid molecules or clonally expanded proxies for individual nucleic acid molecules in a highly parallel fashion (*e.g.*, greater than 10^5 molecules are sequenced simultaneously). In one embodiment, the relative abundance of the nucleic acid species in the library can be estimated by counting the relative number of occurrences of their cognate sequences in the data generated by the sequencing experiment. Next generation sequencing methods are known in the art, and are described, *e.g.*, in Metzker, M. (2010) *Nature Biotechnology Reviews* 11:31-46, incorporated herein by reference.

In one embodiment, the next-generation sequencing allows for the determination of the nucleotide sequence of an individual nucleic acid molecule (*e.g.*, Helicos BioSciences' HeliScope Gene Sequencing system, and Pacific Biosciences' PacBio RS system). In other embodiments, the sequencing method determines the nucleotide sequence of clonally expanded proxies for individual nucleic acid molecules (*e.g.*, the Solexa sequencer, Illumina Inc., San Diego, Calif; 454 Life Sciences (Branford, Conn.), and Ion Torrent). *e.g.*, massively parallel short-read sequencing (*e.g.*, the Solexa sequencer, Illumina Inc., San Diego, Calif.), which generates more bases of sequence per sequencing unit than other sequencing methods that generate fewer but longer reads. Other methods or machines for next-generation sequencing include, but are not limited to, the sequencers provided by 454 Life Sciences (Branford, Conn.), Applied Biosystems (Foster City, Calif.; SOLiD sequencer), and Helicos BioSciences Corporation (Cambridge, Mass.).

Platforms for next-generation sequencing include, but are not limited to, Roche/454's Genome Sequencer (GS) FLX System, Illumina/Solexa's Genome Analyzer (GA), Life/APG's Support Oligonucleotide Ligation Detection (SOLiD) system, Polonator's G.007 system, Helicos BioSciences' HeliScope Gene Sequencing system, and Pacific Biosciences' PacBio RS system.

IDT Exhibit 1020.3039

NGS technologies can include one or more of steps, *e.g.*, template preparation, sequencing and imaging, and data analysis.

Template preparation

Methods for template preparation can include steps such as randomly breaking nucleic acids (*e.g.*, genomic DNA or cDNA) into smaller sizes and generating sequencing templates (*e.g.*, fragment templates or mate-pair templates). The spatially separated templates can be attached or immobilized to a solid surface or support, allowing massive amounts of sequencing reactions to be performed simultaneously. Types of templates that can be used for NGS reactions include, *e.g.*, clonally amplified templates originating from single DNA molecules, and single DNA molecule templates.

Methods for preparing clonally amplified templates include, *e.g.*, emulsion PCR (emPCR) and solid-phase amplification.

EmPCR can be used to prepare templates for NGS. Typically, a library of nucleic acid fragments is generated, and adapters containing universal priming sites are ligated to the ends of the fragment. The fragments are then denatured into single strands and captured by beads. Each bead captures a single nucleic acid molecule. After amplification and enrichment of emPCR beads, a large amount of templates can be attached or immobilized in a polyacrylamide gel on a standard microscope slide (*e.g.*, Polonator), chemically crosslinked to an amino-coated glass surface (*e.g.*, Life/APG; Polonator), or deposited into individual PicoTiterPlate (PTP) wells (*e.g.*, Roche/454), in which the NGS reaction can be performed.

Solid-phase amplification can also be used to produce templates for NGS. Typically, forward and reverse primers are covalently attached to a solid support. The surface density of the amplified fragments is defined by the ratio of the primers to the templates on the support. Solid-phase amplification can produce hundreds of millions spatially separated template clusters (*e.g.*, Illumina/Solexa). The ends of the template clusters can be hybridized to universal sequencing primers for NGS reactions.

Other methods for preparing clonally amplified templates also include, *e.g.*, Multiple Displacement Amplification (MDA) (Lasken R. S. *Curr Opin Microbiol.* 2007; 10(5):510-6). MDA is a non-PCR based DNA amplification technique. The reaction involves annealing random hexamer primers to the template and DNA synthesis by high

IDT Exhibit 1020.3040

fidelity enzyme, typically Φ29 at a constant temperature. MDA can generate large sized products with lower error frequency.

Template amplification methods such as PCR can be coupled with NGS platforms to target or enrich specific regions of the genome (*e.g.*, exons). Exemplary template enrichment methods include, *e.g.*, microdroplet PCR technology (Tewhey R. *et al., Nature Biotech.* 2009, 27:1025-1031), custom-designed oligonucleotide microarrays (*e.g.*, Roche/NimbleGen oligonucleotide microarrays), and solution-based hybridization methods (*e.g.*, molecular inversion probes (MIPs) (Porreca G. J. *et al., Nature Methods*, 2007, 4:931-936; Krishnakumar S. *et al., Proc. Natl. Acad. Sci. USA*, 2008, 105:9296-9310; Turner E. H. *et al., Nature Methods*, 2009, 6:315-316), and biotinylated RNA capture sequences (Gnirke A. *et al.*, *Nat. Biotechnol.* 2009; 27(2):182-9)

Single-molecule templates are another type of templates that can be used for NGS reaction. Spatially separated single molecule templates can be immobilized on solid supports by various methods. In one approach, individual primer molecules are covalently attached to the solid support. Adapters are added to the templates and templates are then hybridized to the immobilized primers. In another approach, single-molecule templates are covalently attached to the solid support by priming and extending single-stranded, single-molecule templates from immobilized primers. Universal primers are then hybridized to the templates. In yet another approach, single polymerase molecules are attached to the solid support, to which primed templates are bound.

Sequencing and imaging

Exemplary sequencing and imaging methods for NGS include, but are not limited to, cyclic reversible termination (CRT), sequencing by ligation (SBL), single-molecule addition (pyrosequencing), and real-time sequencing.

CRT uses reversible terminators in a cyclic method that minimally includes the steps of nucleotide incorporation, fluorescence imaging, and cleavage. Typically, a DNA polymerase incorporates a single fluorescently modified nucleotide corresponding to the complementary nucleotide of the template base to the primer. DNA synthesis is terminated after the addition of a single nucleotide and the unincorporated nucleotides are washed away. Imaging is performed to determine the identity of the incorporated labeled nucleotide. Then in the cleavage step, the terminating/inhibiting group and the

IDT Exhibit 1020.3041

fluorescent dye are removed. Exemplary NGS platforms using the CRT method include, but are not limited to, Illumina/Solexa Genome Analyzer (GA), which uses the clonally amplified template method coupled with the four-color CRT method detected by total internal reflection fluorescence (TIRF); and Helicos BioSciences/HeliScope, which uses the single-molecule template method coupled with the one-color CRT method detected by TIRF.

SBL uses DNA ligase and either one-base-encoded probes or two-base-encoded probes for sequencing. Typically, a fluorescently labeled probe is hybridized to its complementary sequence adjacent to the primed template. DNA ligase is used to ligate the dye-labeled probe to the primer. Fluorescence imaging is performed to determine the identity of the ligated probe after non-ligated probes are washed away. The fluorescent dye can be removed by using cleavable probes to regenerate a 5'-PO_4 group for subsequent ligation cycles. Alternatively, a new primer can be hybridized to the template after the old primer is removed. Exemplary SBL platforms include, but are not limited to, Life/APG/SOLiD (support oligonucleotide ligation detection), which uses two-base-encoded probes.

Pyrosequencing method is based on detecting the activity of DNA polymerase with another chemiluminescent enzyme. Typically, the method allows sequencing of a single strand of DNA by synthesizing the complementary strand along it, one base pair at a time, and detecting which base was actually added at each step. The template DNA is immobile, and solutions of A, C, G, and T nucleotides are sequentially added and removed from the reaction. Light is produced only when the nucleotide solution complements the first unpaired base of the template. The sequence of solutions which produce chemiluminescent signals allows the determination of the sequence of the template. Exemplary pyrosequencing platforms include, but are not limited to, Roche/454, which uses DNA templates prepared by emPCR with 1-2 million beads deposited into PTP wells.

Real-time sequencing involves imaging the continuous incorporation of dye-labeled nucleotides during DNA synthesis. Exemplary real-time sequencing platforms include, but are not limited to, Pacific Biosciences platform, which uses DNA polymerase molecules attached to the surface of individual zero-mode waveguide (ZMW) detectors to obtain sequence information when phospholinked nucleotides are being

IDT Exhibit 1020.3042

incorporated into the growing primer strand; Life/VisiGen platform, which uses an engineered DNA polymerase with an attached fluorescent dye to generate an enhanced signal after nucleotide incorporation by fluorescence resonance energy transfer (FRET); and LI-COR Biosciences platform, which uses dye-quencher nucleotides in the sequencing reaction.

Other sequencing methods for NGS include, but are not limited to, nanopore sequencing, sequencing by hybridization, nano-transistor array based sequencing, polony sequencing, scanning tunneling microscopy (STM) based sequencing, and nanowire-molecule sensor based sequencing.

Nanopore sequencing involves electrophoresis of nucleic acid molecules in solution through a nano-scale pore which provides a highly confined space within which single-nucleic acid polymers can be analyzed. Exemplary methods of nanopore sequencing are described, *e.g.*, in Branton D. *et al.*, *Nat Biotechnol.* 2008; 26(10):1146-53.

Sequencing by hybridization is a non-enzymatic method that uses a DNA microarray. Typically, a single pool of DNA is fluorescently labeled and hybridized to an array containing known sequences. Hybridization signals from a given spot on the array can identify the DNA sequence. The binding of one strand of DNA to its complementary strand in the DNA double-helix is sensitive to even single-base mismatches when the hybrid region is short or is specialized mismatch detection proteins are present. Exemplary methods of sequencing by hybridization are described, *e.g.*, in Hanna G.J. *et al.*, *J. Clin. Microbiol.* 2000; 38 (7): 2715–21; and Edwards J.R. *et al.*, *Mut. Res.* 2005; 573 (1-2): 3–12.

Polony sequencing is based on polony amplification and sequencing-by-synthesis via multiple single-base-extensions (FISSEQ). Polony amplification is a method to amplify DNA *in situ* on a polyacrylamide film. Exemplary polony sequencing methods are described, *e.g.*, in US Patent Application Publication No. 2007/0087362.

Nano-transistor array based devices, such as Carbon NanoTube Field Effect Transistor (CNTFET), can also be used for NGS. For example, DNA molecules are stretched and driven over nanotubes by micro-fabricated electrodes. DNA molecules sequentially come into contact with the carbon nanotube surface, and the difference in current flow from each base is produced due to charge transfer between the DNA

IDT Exhibit 1020.3043

molecule and the nanotubes. DNA is sequenced by recording these differences. Exemplary Nano-transistor array based sequencing methods are described, *e.g.*, in U.S. Patent Application Publication No. 2006/0246497.

Scanning tunneling microscopy (STM) can also be used for NGS. STM uses a piezo-electric-controlled probe that performs a raster scan of a specimen to form images of its surface. STM can be used to image the physical properties of single DNA molecules, *e.g.*, generating coherent electron tunneling imaging and spectroscopy by integrating scanning tunneling microscope with an actuator-driven flexible gap. Exemplary sequencing methods using STM are described, *e.g.*, in U.S. Patent Application Publication No. 2007/0194225.

A molecular-analysis device which is comprised of a nanowire-molecule sensor can also be used for NGS. Such device can detect the interactions of the nitrogenous material disposed on the nanowires and nucleic acid molecules such as DNA. A molecule guide is configured for guiding a molecule near the molecule sensor, allowing an interaction and subsequent detection. Exemplary sequencing methods using nanowire-molecule sensor are described, *e.g.*, in U.S. Patent Application Publication No. 2006/0275779.

Double ended sequencing methods can be used for NGS. Double ended sequencing uses blocked and unblocked primers to sequence both the sense and antisense strands of DNA. Typically, these methods include the steps of annealing an unblocked primer to a first strand of nucleic acid; annealing a second blocked primer to a second strand of nucleic acid; elongating the nucleic acid along the first strand with a polymerase; terminating the first sequencing primer; deblocking the second primer; and elongating the nucleic acid along the second strand. Exemplary double ended sequencing methods are described, *e.g.*, in U.S. Patent Serial No. 7,244,567.

Data analysis

After NGS reads have been generated, they can be aligned to a known reference sequence or assembled *de novo*.

For example, identifying genetic variations such as single-nucleotide polymorphism and structural variants in a sample (*e.g.*, a tumor sample) can be accomplished by aligning NGS reads to a reference sequence (*e.g.*, a wild-type

IDT Exhibit 1020.3044

sequence). Methods of sequence alignment for NGS are described *e.g.*, in Trapnell C. and Salzberg S.L. *Nature Biotech.*, 2009, 27:455-457.

Examples of *de novo* assemblies are described, *e.g.*, in Warren R. *et al., Bioinformatics*, 2007, 23:500-501; Butler J. *et al., Genome Res.*, 2008, 18:810-820; and Zerbino D.R. and Birney E., *Genome Res.*, 2008, 18:821-829.

Sequence alignment or assembly can be performed using read data from one or more NGS platforms, *e.g.*, mixing Roche/454 and Illumina/Solexa read data.

Algorithms and methods for data analysis are described in USSN 61/428,568, filed December 30, 2010, incorporated herein by reference.

Fusion Expression Level

In certain embodiments, DCTN1-ALK or LMNA-NTRK1 fusion expression level can also be assayed. Fusion expression can be assessed by any of a wide variety of methods for detecting expression of a transcribed molecule or protein. Non-limiting examples of such methods include immunological methods for detection of secreted, cell-surface, cytoplasmic, or nuclear proteins, protein purification methods, protein function or activity assays, nucleic acid hybridization methods, nucleic acid reverse transcription methods, and nucleic acid amplification methods.

In certain embodiments, activity of a particular gene is characterized by a measure of gene transcript (*e.g.,* mRNA), by a measure of the quantity of translated protein, or by a measure of gene product activity. DCTN1-ALK or LMNA-NTRK1 fusion expression can be monitored in a variety of ways, including by detecting mRNA levels, protein levels, or protein activity, any of which can be measured using standard techniques. Detection can involve quantification of the level of gene expression (*e.g.,* genomic DNA, cDNA, mRNA, protein, or enzyme activity), or, alternatively, can be a qualitative assessment of the level of gene expression, in particular in comparison with a control level. The type of level being detected will be clear from the context.

Methods of detecting and/or quantifying the fusion gene transcript (mRNA or cDNA made therefrom) using nucleic acid hybridization techniques are known to those of skill in the art (see Sambrook *et al. supra*). For example, one method for evaluating the presence, absence, or quantity of cDNA involves a Southern transfer as

IDT Exhibit 1020.3045

described above. Briefly, the mRNA is isolated (*e.g.*, using an acid guanidinium-phenol-chloroform extraction method, Sambrook *et al. supra.*) and reverse transcribed to produce cDNA. The cDNA is then optionally digested and run on a gel in buffer and transferred to membranes. Hybridization is then carried out using the nucleic acid probes specific for the DCTN1-ALK fusion cDNA or the LMNA-NTRK1 fusion cDNA, *e.g.*, using the probes and primers described herein.

In other embodiments, DCTN1-ALK or LMNA-NTRK1 expression is assessed by preparing genomic DNA or mRNA/cDNA (*i.e.*, a transcribed polynucleotide) from cells in a subject sample, and by hybridizing the genomic DNA or mRNA/cDNA with a reference polynucleotide which is a complement of a polynucleotide comprising the DCTN1-ALK fusion or the LMNA-NTRK1 fusion, and fragments thereof. cDNA can, optionally, be amplified using any of a variety of polymerase chain reaction methods prior to hybridization with the reference polynucleotide. Expression of a fusion as described herein can likewise be detected using quantitative PCR (QPCR) to assess the level of expression.

Detection of Fusion Polypeptide

The activity or level of a fusion polypeptide (*e.g.*, DCTN1-ALK or LMNA-NTRK1) can also be detected and/or quantified by detecting or quantifying the expressed polypeptide. The fusion polypeptide can be detected and quantified by any of a number of means known to those of skill in the art. These can include analytic biochemical methods such as electrophoresis, capillary electrophoresis, high performance liquid chromatography (HPLC), thin layer chromatography (TLC), hyperdiffusion chromatography, and the like, or various immunological methods such as fluid or gel precipitin reactions, immunodiffusion (single or double), immunoelectrophoresis, radioimmunoassay (RIA), enzyme-linked immunosorbent assays (ELISAs), immunofluorescent assays, Western blotting, immunohistochemistry (IHC) and the like. A skilled artisan can adapt known protein/antibody detection methods.

Another agent for detecting a fusion polypeptide is an antibody molecule capable of binding to a polypeptide corresponding to a marker, *e.g.*, an antibody with a detectable label. Techniques for generating antibodies are described herein. The

IDT Exhibit 1020.3046

term "labeled", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (*i.e.*, physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with another reagent that is directly labeled. Examples of indirect labeling include detection of a primary antibody using a fluorescently labeled secondary antibody and end-labeling of a DNA probe with biotin such that it can be detected with fluorescently labeled streptavidin.

In another embodiment, the antibody is labeled, *e.g.,* a radio-labeled, chromophore-labeled, fluorophore-labeled, or enzyme-labeled antibody. In another embodiment, an antibody derivative (*e.g.,* an antibody conjugated with a substrate or with the protein or ligand of a protein-ligand pair {*e.g.,* biotin-streptavidin}), or an antibody fragment (*e.g.,* a single-chain antibody, an isolated antibody hypervariable domain, *etc.*) which binds specifically with a DCTN1-ALK or LMNA-NTRK1 fusion protein, is used.

Fusion polypeptides from cells can be isolated using techniques that are known to those of skill in the art. The protein isolation methods employed can, for example, be such as those described in Harlow and Lane (Harlow and Lane, 1988, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York).

Means of detecting proteins using electrophoretic techniques are well known to those of skill in the art (see generally, R. Scopes (1982) *Protein Purification*, Springer-Verlag, N.Y.; Deutscher, (1990) *Methods in Enzymology* Vol. 182: *Guide to Protein Purification*, Academic Press, Inc., N.Y.).

In another embodiment, Western blot (immunoblot) analysis is used to detect and quantify the presence of a polypeptide in the sample.

In another embodiment, the polypeptide is detected using an immunoassay. As used herein, an immunoassay is an assay that utilizes an antibody to specifically bind to the analyte. The immunoassay is thus characterized by detection of specific binding of a polypeptide to an anti-antibody as opposed to the use of other physical or chemical properties to isolate, target, and quantify the analyte.

The fusion polypeptide is detected and/or quantified using any of a number of immunological binding assays (see, *e.g.,* U.S. Patent Nos. 4,366,241; 4,376,110;

IDT Exhibit 1020.3047

4,517,288; and 4,837,168). For a review of the general immunoassays, see also Asai (1993) *Methods in Cell Biology* Volume 37: *Antibodies in Cell Biology*, Academic Press, Inc. New York; Stites & Terr (1991) *Basic and Clinical Immunology* 7th Edition.

Kits

In one aspect, the invention features, a kit, *e.g.*, containing an oligonucleotide having a mutation described herein, e.g., a DCTN1-ALK fusion or a LMNA-NTRK1 fusion. Optionally, the kit can also contain an oligonucleotide that is the wildtype counterpart of the mutant oligonucleotide.

A kit featured in the invention can include a carrier, *e.g.*, a means being compartmentalized to receive in close confinement one or more container means. In one embodiment the container contains an oligonucleotide, *e.g.*, a primer or probe as described above. The components of the kit are useful, for example, to diagnose or identify a mutation in a tumor sample in a patient. The probe or primer of the kit can be used in any sequencing or nucleotide detection assay known in the art, *e.g.*, a sequencing assay, *e.g.*, an NGS method, RT-PCR, or *in situ* hybridization.

In some embodiments, the components of the kit are useful, for example, to diagnose or identify a DCTN1-ALK fusion in a tumor sample in a patient, and to accordingly identify an appropriate therapeutic agent to treat the cancer.

A kit featured in the invention can include, *e.g.*, assay positive and negative controls, nucleotides, enzymes (*e.g.*, RNA or DNA polymerase or ligase), solvents or buffers, a stabilizer, a preservative, a secondary antibody, *e.g.*, an anti-HRP antibody (IgG) and a detection reagent.

An oligonucleotide can be provided in any form, *e.g.*, liquid, dried, semi-dried, or lyophilized, or in a form for storage in a frozen condition.

Typically, an oligonucleotide, and other components in a kit are provided in a form that is sterile. An oligonucleotide, *e.g.*, an oligonucleotide that contains an ALK mutation, *e.g.*, a DCTN1-ALK fusion, described herein, or an oligonucleotide complementary to a LMNA-NTRK1 fusion described herein, is provided in a liquid solution, the liquid solution generally is an aqueous solution, *e.g.*, a sterile aqueous solution. When the oligonucleotide is provided as a dried form, reconstitution generally

IDT Exhibit 1020.3048

is accomplished by the addition of a suitable solvent. The solvent, *e.g.*, sterile buffer, can optionally be provided in the kit.

The kit can include one or more containers for the composition containing an oligonucleotide in a concentration suitable for use in the assay or with instructions for dilution for use in the assay. In some embodiments, the kit contains separate containers, dividers or compartments for the oligonucleotide and assay components, and the informational material. For example, the oligonucleotides can be contained in a bottle or vial, and the informational material can be contained in a plastic sleeve or packet. In other embodiments, the separate elements of the kit are contained within a single, undivided container. For example, an oligonucleotide composition is contained in a bottle or vial that has attached thereto the informational material in the form of a label. In some embodiments, the kit includes a plurality (*e.g.*, a pack) of individual containers, each containing one or more unit forms (*e.g.*, for use with one assay) of an oligonucleotide. For example, the kit includes a plurality of ampoules, foil packets, or blister packs, each containing a single unit of oligonucleotide for use in sequencing or detecting a mutation in a tumor sample. The containers of the kits can be air tight and/or waterproof. The container can be labeled for use.

For antibody-based kits, the kit can include: (1) a first antibody (e.g., attached to a solid support) which binds to a fusion polypeptide; and, optionally, (2) a second, different antibody which binds to either the polypeptide or the first antibody and is conjugated to a detectable agent.

In one embodiment, the kit can include informational material for performing and interpreting the sequencing or diagnostic. In another embodiment, the kit can provide guidance as to where to report the results of the assay, *e.g.*, to a treatment center or healthcare provider. The kit can include forms for reporting the results of a sequencing or diagnostic assay described herein, and address and contact information regarding where to send such forms or other related information; or a URL (Uniform Resource Locator) address for reporting the results in an online database or an online application (*e.g.*, an app). In another embodiment, the informational material can include guidance regarding whether a patient should receive treatment with a particular chemotherapeutic drug, depending on the results of the assay.

IDT Exhibit 1020.3049

The informational material of the kits is not limited in its form. In many cases, the informational material, *e.g.*, instructions, is provided in printed matter, *e.g.*, a printed text, drawings, and/or photographs, *e.g.*, a label or printed sheet. However, the informational material can also be provided in other formats, such as computer readable material, video recording, or audio recording. In another embodiment, the informational material of the kit is contact information, *e.g.*, a physical address, email address, website, or telephone number, where a user of the kit can obtain substantive information about the sequencing or diagnostic assay and/or its use in the methods described herein. The informational material can also be provided in any combination of formats.

In some embodiments, a biological sample is provided to an assay provider, *e.g.*, a service provider (such as a third party facility) or a healthcare provider, who evaluates the sample in an assay and provides a read out. For example, in one embodiment, an assay provider receives a biological sample from a subject, such as a blood or tissue sample, *e.g.*, a biopsy sample, and evaluates the sample using an assay described herein, *e.g.*, a sequencing assay or *in situ* hybridization assay, and determines that the sample contains a DCTN1-ALK fusion. The assay provider, *e.g.*, a service provider or healthcare provider, can then conclude that the subject is, or is not, a candidate for a particular drug or a particular cancer treatment regimen.

The assay provider can provide the results of the evaluation, and optionally, conclusions regarding one or more of diagnosis, prognosis, or appropriate therapy options to, for example, a healthcare provider, or patient, or an insurance company, in any suitable format, such as by mail or electronically, or through an online database. The information collected and provided by the assay provider can be stored in a database.

The invention is further illustrated by the following example, which should not be construed as further limiting.

EXAMPLES

Example 1: DCTN1-ALK fusion

The following exemplifies the use of massively parallel sequencing assays to identify novel alterations, such as DCTN1-ALK fusions. Based on the results shown herein, additional alterations, *e.g.*, ALK translocations, can be screened using, *e.g.*, either qRT-PCR analysis of cDNA prepared from a pre-selected tumor sample.

IDT Exhibit 1020.3050

A pan-cancer diagnostic assay based on massively parallel sequencing technology was used to identify genetic alterations in a series of colorectal cancer (CRC), non-small cell lung cancer (NSCLC) and melanoma samples. A novel ALK translocation was identified in one melanoma sample. The resulting DCNT1-ALK fusion is the product of a balanced translocation between two parts of chromosome 2. A snapshot of the sequencing reads is presented in FIG. 5. Both breakpoints are in introns and the predicted mRNA is in frame. ALK ("Anaplastic Lymphoma receptor tyrosine Kinase" RefSeq No. NM_004304, GenBank Record dated October 2, 2011) is a known oncogene. DCTN1 ("dynactin 1" NM_004082, RefSeq No. NM_004082, GenBank record dated August 14, 2011) is involved in intracellular transport. The ALK breakpoint is in the canonical intron 19, and the DCTN1 breakpoint is in intron 26. The translocation juxtaposes DCTN1 (exons 1 to 26) with the ALK kinase domain (exons 20-26) (FIG. 1). The nucleotide sequence (SEQ ID NO:6) and polypeptide sequence (SEQ ID NO:7) of the fusion protein are shown in FIGs. 2A to 2F. This fusion may permit the use of a kinase inhibitor or an ALK inhibitor in the treatment of melanoma or other cancers.

A BLAT sequence was performed on the fused construct to identify the coordinates of the ends within both partners. It was determined that the end coordinates are at chromosome 2 position 29,448,216 in the DCTN1 gene and at chromosome 2 position 74,591,795 of the ALK gene. The sequence of the DCTN1-ALK fusion breakpoint (SEQ ID NO:5) is shown in FIG. 1.

A recipricol fusion construct (ALK-DCTN1) caused by the translocation was also identified. This fusion is illustrated in FIG. 3, and the nucleotide sequence (SEQ ID NO:11, and the amino acid sequence (SEQ ID NO:12) are presented in FIGs. 4A-4D. The sequence of the ALK-DCTN1 breakpoint (SEQ ID NO:13) is shown in FIG. 3. It is not been determined whether the ALK-DCTN1 polypeptide is expressed, and it is not expected that the ALK-DCTN1 polypeptide is relevant to oncogenesis.

The ALK segment of the DCTN1-ALK fusion starts at the exon 19 recombination site previously reported for the majority of *ALK* fusion genes and contains the same intracellular domain as other *ALK* rearrangements including *EML4-ALK*, known to be ALK kinase inhibitor sensitive (Soda, M., *et al. Nature* 448,

IDT Exhibit 1020.3051

561-566 (2007); Kwak, E.L., *et al. N Engl J Med* 363, 1693-1703 (2010)). Clinical detection of *ALK* rearrangements is commonly performed using fluorescence in situ hybridization (FISH) with *ALK* break-apart probes (Kwak, E.L., *et al. N Engl J Med* 363, 1693-1703 (2010)) or by RT-PCR (Takeuchi, K., *et al. Clin Cancer Res* 14, 6618-6624 (2008)) using primers for specific *ALK* rearrangement partners.

Translocations in the region of ALK are previously known to be associated with certain types of cancer. For example, an inversion resulting in an EML4 (Echinoderm Microtubule associated protein Like 4)-ALK fusion is associated with lung cancer (*e.g.*, non-small cell lung cancer), and these cancers respond to treatment with the ALK inhibitor crizotinib. A translocation that results in an NPM (Nucleophosmin)-ALK fusion is associated with lymphoma, and translocations that result in TPM3 (Tropomyosin3)-ALK and TPM4-ALK fusions are associated with myofibroblastic tumors.

Expression of DCTN1-ALK was detected in a Spitz tumor, which is a family of neoplasms that can range from benign Spitz nevi to Spitzoid melanomas. Expression of DCTN1-ALK was also detected in a non-Langerhans cell histiocytosis.

These findings suggest that Spitz tumor patients (*e.g.*, patients with melanocytic neoplasms) and patients with histiocytosis, as well as patients with melanomas or other cancers carrying the identified rearrangement may be candidates for therapy with crizotinib or other ALK kinase inhibitors.

Example 2: LMNA-NTRK1 fusion

A novel fusion between LMNA and NTRK1 was also found in a single melanoma sample. The fusion is the result of a ~750 kb deletion between intron 2 of LMNA and intron 10 of NTRK1, and the fusion product is in-frame (FIG. 6). The nucleotide sequence (SEQ ID NO:9) and polypeptide sequence (SEQ ID NO:10) of the fusion protein are shown in FIGs. 7A-7B. The sequence at the fusion junction (SEQ ID NO:8) is presented in FIG. 6. A snapshot of the sequencing reads is presented in FIG. 8.

LMNA ("Lamin A/C," RefSeq NM_005572, GenBank Record dated October 16, 2011) is a gene involved in the formation of the nuclear lamina, a two dimensional matrix of proteins known to be involved in nuclear stability, chromatin

IDT Exhibit 1020.3052

structure, and gene expression. LMNA has previously been shown to be overexpressed in cancer.

NTRK1 ("neurotrophic tyrosine kinase" RefSeq NM_002529; GenBank record dated October 16, 2011) is a receptor tyrosine kinase known to be an occasional target of re-arrangement in papillary thyroid carcinomas. The kinase domain is preserved in the fusion gene described in FIGs. 7A-7B, and is in the region of NTRK1 where fusions have previously been observed (*see e.g.,* Greco, A. *et al., Mol Cell Endocrin* (2010) 321(1): 44-49).

NTRK1 fusions have been observed across a relatively large expanse of NTRK1 (*see e.g.,* Greco, A. *et al., Mol Cell Endocrin* (2010) 321(1): 44-49). We observed one fusion event on the shorter side, but other versions are possible. Some LMNA-NTRK1 fusions may have different length pieces of LMNA attached at the 5' end of the fusion gene, and different length pieces of NTRK1 attached at the 3' end of the fusion gene, provided that the tyrosine kinase domain or a functional portion thereof is preserved.

Expression of LMNA-NTRK1 was detected in a Spitz tumor, which is a family of neoplasms that can range from benign Spitz nevi to Spitzoid melanomas.

We also observed an NTRK1 mutation in an unrelated colorectal sample set using similar NGS sequencing methods.

Incorporated by reference herein in their entirety are any polynucleotide and polypeptide sequences which reference an accession number correlating to an entry in a public database, such as those maintained by the COSMIC database, available on the worldwide web at sanger.ac.uk/genetics/CGP/cosmic/; and the Institute for Genomic Research (TIGR) on the world wide web at tigr.org and/or the National Center for Biotechnology Information (NCBI) on the world wide web at ncbi.nlm.nih.gov.

Equivalents

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments described herein. Such equivalents are intended to be encompassed by the following claims.

IDT Exhibit 1020.3053

1. An isolated or purified nucleic acid molecule that encodes a DCTN1-ALK fusion or an LMNA-NTRK1 fusion, or a breakpoint comprising fragment thereof, wherein:

(i) the DCTN1-ALK fusion comprises an in-frame fusion of one or more of exons 1-26 of the nucleotide sequence of DCTN1 of SEQ ID NO:6 or a fragment thereof, or a sequence at least 85% identical thereto, and one or more of exons 20-29 of the nucleotide sequence of ALK of SEQ ID NO:6, or a fragment thereof, or a sequence at least 85% identical thereto; or

(ii) the LMNA-NTRK1 fusion comprises an in-frame fusion of one or more of exons 1-2 of the nucleotide sequence of LMNA of SEQ ID NO:9 or a fragment thereof, or a sequence at least 85% identical thereto, and one or more of exons 11-19 of the nucleotide sequence of NTRK1 of SEQ ID NO:9, or a fragment thereof, or a sequence at least 85% identical thereto.

2. The nucleic acid molecule of claim 1, wherein the DCTN1-ALK fusion comprises the nucleotide sequence of 1-3195 of SEQ ID NO:6 or a fragment thereof, or a sequence at least 85% identical thereto, and the nucleotide sequence of 3196-4907 of SEQ ID NO:6 or a fragment thereof, or a sequence at least 85% identical thereto.

3. The nucleic acid molecule of claim 1, wherein the LMNA-NTRK1 fusion comprises the nucleotide sequence of 1-513 of SEQ ID NO:9 or a fragment thereof, or a sequence at least 85% identical thereto, and the nucleotide sequence of 514-1740 of SEQ ID NO:9 or a fragment thereof, or a sequence at least 85% identical thereto.

4. A nucleic acid molecule that is capable of hybridizing to a DCTN1-ALK fusion comprising the nucleotide sequence of SEQ ID NO:6, or a fragment thereof comprising a breakpoint.

IDT Exhibit 1020.3054

5. A nucleic acid molecule that is capable of hybridizing to an LMNA-NTRK1 fusion comprising the nucleotide sequence of SEQ ID NO:9, or a fragment thereof comprising a breakpoint.

6. A fragment of the nucleic acid molecule of any of claims 1-5, wherein said fragment comprises oligonucleotides between 10 and 25 nucleotides in length, or between 100 to 300 nucleotides in length.

7. The fragment of claim 6, which is a probe or primer that includes an oligonucleotide between about 5 and 25 nucleotides in length.

8. The fragment of claim 6, which is a bait that comprises an oligonucleotide between about 100 to 300 nucleotides, 130 and 230 nucleotides, or 150 and 200 nucleotides, in length.

9. The nucleic acid molecule of any of claims 1-5, wherein said breakpoint comprises the nucleotide sequence of SEQ ID NO:5 or SEQ ID NO:13, or a fragment thereof.

10. A nucleic acid molecule of any of claims 1-5 suitable as probe, primer, bait or library member that specifically binds to the DCTN1-ALK fusion or the LMNA-NTRK1 fusion.

11. The isolated or purified nucleic acid molecule of any of claims 1-5, which is operatively linked to a native or a heterologous regulatory sequence.

12. An isolated or purified vector comprising a nucleic acid molecule of any of claims 1-5.

13. A host cell comprising a vector of claim 12.

IDT Exhibit 1020.3055

14. A nucleic acid molecule that specifically reduces or inhibits the expression of the nucleic acid molecule of any of claims 1-3.

15. The nucleic acid molecule of claim 14, which is chosen from an antisense molecule, ribozyme, siRNA, or triple helix molecule.

16. An isolated or purified a DCTN1-ALK fusion or an LMNA-NTRK1 fusion polypeptide or breakpoint containing fragment thereof, wherein:

(i) the DCTN1-ALK fusion comprises the amino acid sequence of amino acids 1-1065 of DCTN1 of SEQ ID NO:7 or a fragment thereof, or a sequence at least 85% identical thereto, and the amino acid sequence of 1066-1640 of ALK of SEQ ID NO:7, or a fragment thereof, or a sequence at least 85% identical thereto; or

(ii) the LMNA-NTRK1 fusion comprises the amino acid sequence of amino acids 1-175 of SEQ ID NO:10 or a fragment thereof, or a sequence at least 85% identical thereto, and the amino acid sequence of 168-580 of NTRK1 of SEQ ID NO:10, or a fragment thereof, or a sequence at least 85% identical thereto.

17. The isolated or purified fusion polypeptide of claim 16, having a kinase activity, and/or a dimerizing or multimerizing activity.

18. The isolated or purified DCTN1-ALK fusion or LMNA-NTRK1 fusion polypeptide of claim 17, wherein:

(i) the DCTN1-ALK fusion comprises the amino acid sequence of SEQ ID NO:7 or a fragment thereof, or a sequence at least 85% identical thereto; or

(ii) the LMNA-NTRK1 fusion comprises the amino acid sequence of SEQ ID NO:10 or a fragment thereof, or a sequence at least 85% identical thereto.

19. An isolated or purified antibody molecule that specifically binds to the DCTN1-ALK fusion or the LMNA-NTRK1 fusion polypeptide of any of claims 16-18.

IDT Exhibit 1020.3056

20. A reaction mixture comprising:

a detection reagent, or purified or isolated preparation thereof; and

a target nucleic acid derived from a neoplasm or a cancer,

wherein said detection reagent can distinguish a reference sequence from a mutation chosen from: a nucleic acid, or amino acid sequence, having a DCTN1-ALK breakpoint, *e.g.*, as identified in FIG. 1 (*e.g.*, SEQ ID NO:5); or an LMNA-NTRK1 breakpoint, *e.g.*, as identified in FIG. 6 (*e.g.*, SEQ ID NO:8), or an associated mutation.

21. The reaction mixture of claim 20, wherein the detection reagent specifically distinguishes a wild type ALK or another ALK fusion (or DCTN1) from DCTN1-ALK nucleic acid, *e.g.*, as described herein in FIG. 2 (SEQ ID NO:6); or DCTN1-ALK polypeptide (*e.g.*, as described herein in FIG. 2 (SEQ ID NO:7), or an associated mutation.

22. The reaction mixture of claim 20, wherein the detection reagent specifically distinguishes a wild type NTRK1 or another NTRK1 fusion (or LMNA) from LMNA-NTRK1 nucleic acid, *e.g.*, as described herein in FIG. 7 (SEQ ID NO:9); or LMNA-NTRK1 polypeptide, *e.g.*, as described herein in FIG. 7 (SEQ ID NO:10), or an associated mutation.

23. The reaction mixture of any of claims 20-22, wherein the detection reagent comprises a DNA, RNA or mixed DNA/RNA, molecule which is complementary with a nucleic acid sequence on a target nucleic acid (the detection reagent binding site) wherein the detection reagent binding site is disposed in relationship to the interrogation position such that binding of the detection reagent to the detection reagent binding site allows differentiation of mutant and reference sequences for the mutant.

24. A method of making a reaction mixture comprising:

combining a detection reagent, or purified or isolated preparation thereof, with

IDT Exhibit 1020.3057

a target nucleic acid derived from neoplasm or cancer, wherein said detection reagent can distinguish a reference sequence from a mutation chosen from: a nucleic acid, or amino acid sequence, having a DCTN1-ALK breakpoint, *e.g.*, as identified in FIG. 1 (*e.g.*, SEQ ID NO:5); or an LMNA-NTRK1 breakpoint, *e.g.*, as identified in FIG. 6 (*e.g.*, SEQ ID NO:8), or an associated mutation.

25. The method of claim 24, wherein the detection reagent specifically distinguishes a wild type ALK or another ALK fusion (or DCTN1) from DCTN1-ALK nucleic acid, *e.g.*, as described herein in FIG. 2 (SEQ ID NO:6); or DCTN1-ALK polypeptide (*e.g.*, as described herein in FIG. 2 (SEQ ID NO:7), or an associated mutation.

26. The method of claim 24, wherein the detection reagent specifically distinguishes a wild type NTRK1 or another NTRK1 fusion (or LMNA) from LMNA-NTRK1 nucleic acid, *e.g.*, as described herein in FIG. 7 (SEQ ID NO:9); or LMNA-NTRK1 polypeptide, *e.g.*, as described herein in FIG. 7 (SEQ ID NO:10), or an associated mutation.

27. A purified or isolated preparation of a fusion nucleic acid molecule from a neoplasm or cancer disposed in a sequencing device, or a sample holder for use in such a device, wherein said mutation is chosen from: a nucleic acid having a DCTN1-ALK breakpoint, *e.g.*, as identified in FIG. 1 (*e.g.*, SEQ ID NO:5); or an LMNA-NTRK1 breakpoint, *e.g.*, as identified in FIG. 6 (*e.g.*, SEQ ID NO:8), or an associated mutation.

28. The preparation of claim 27, wherein the mutation comprises a DCTN1-ALK nucleic acid, *e.g.*, as described herein in FIG. 2 (SEQ ID NO:6) or a fragment thereof; or a DCTN1-ALK polypeptide, *e.g.*, as described herein in FIG. 2 (SEQ ID NO:7), or a fragment thereof, or an associated mutation.

IDT Exhibit 1020.3058

29. The method of claim 27, wherein the mutation comprises an LMNA-NTRK1 nucleic acid, *e.g.*, as described herein in FIG. 7 (SEQ ID NO:9) or a fragment thereof; or an LMNA-NTRK1 polypeptide, *e.g.*, as described herein in FIG. 7 (SEQ ID NO:10), or a fragment thereof, or an associated mutation.

30. A purified or isolated preparation of a fusion nucleic acid molecule from a neoplasm or cancer disposed in a device for determining a physical or chemical property, *e.g.*, stability of a duplex, *e.g.*, T_m or a sample holder for use in such a device, wherein said, wherein said mutation is chosen from: a nucleic acid having a DCTN1-ALK breakpoint, *e.g.*, as identified in FIG. 1 (*e.g.*, SEQ ID NO:5); or an LMNA-NTRK1 breakpoint, *e.g.*, as identified in FIG. 6 (*e.g.*, SEQ ID NO:8), or an associated mutation.

31. A detection reagent comprising a DNA, RNA or mixed DNA/RNA molecule, comprising a nucleotide sequence which is complementary with a nucleic acid sequence on a target nucleic acid in which the detection reagent binding site is disposed in relationship to the interrogation position such that binding (or in embodiments, lack of binding) of the detection reagent to the detection reagent binding site allows differentiation of a mutant and a reference sequence and said target nucleic acid is derived from a neoplasm or cancer,

wherein said mutation is chosen from: a nucleic acid having a DCTN1-ALK breakpoint, *e.g.*, as identified in FIG. 1 (*e.g.*, SEQ ID NO:5); or an LMNA-NTRK1 breakpoint, *e.g.*, as identified in FIG. 6 (*e.g.*, SEQ ID NO:8), or an associated mutation.

32. A purified or isolated preparations of a fusion nucleic acid molecule, *e.g.*, DNA, *e.g.*, genomic DNA or cDNA, or RNA, containing an interrogation position useful for determining if a mutation is present, wherein said nucleic acid molecule is derived from a neoplasm or cancer, wherein said mutation is chosen from: a nucleic acid having a DCTN1-ALK breakpoint, *e.g.*, as identified in FIG. 1 (*e.g.*, SEQ ID

IDT Exhibit 1020.3059

NO:5); or an LMNA-NTRK1 breakpoint, *e.g.*, as identified in FIG. 6 (*e.g.*, SEQ ID NO:8), or an associated mutation.

33. A reaction mixture, comprising:

a detection reagent, or purified or isolated preparation thereof, e.g., a substrate, e.g., a substrate for phosphorylation or other activity, or an antibody; and

a target fusion protein derived from a neoplasm or a cancer, wherein the detection reagent is specific for a DCTN1-ALK breakpoint encoded by the nucleotide sequence in FIG. 1 (*e.g.*, SEQ ID NO:5); or an LMNA-NTRK1 breakpoint encoded by the nucleotide sequence in FIG. 6 (*e.g.*, SEQ ID NO:8).

34. A method of making a reaction mixture comprising:

combining a detection reagent, or purified or isolated preparation thereof, e.g., a substrate, e.g., a substrate for phosphorylation or other activity, or an antibody, described herein with a target fusion protein derived from a neoplasm or cancer, wherein the detection reagent is specific for a DCTN1-ALK breakpoint encoded by the nucleotide sequence in FIG. 1 (*e.g.*, SEQ ID NO:5); or an LMNA-NTRK1 breakpoint encoded by the nucleotide sequence in FIG. 6 (*e.g.*, SEQ ID NO:8).

35. A kit comprising a detection reagent of claim 31.

36. A method of reducing an activity of a DCTN1-ALK fusion or an LMNA-NTRK1 fusion, comprising:

optionally, acquiring knowledge of the presence of the fusion; and

contacting the fusion, or a fusion-expressing cell, with an agent that inhibits an activity or expression of the fusion.

37. The method of claim 36, wherein:

(i) the DCTN1-ALK fusion comprises the amino acid sequence of SEQ ID NO:7 or a fragment thereof, or a sequence at least 85% identical thereto; or

IDT Exhibit 1020.3060

(ii) the LMNA-NTRK1 fusion comprises the amino acid sequence of SEQ ID NO:10 or a fragment thereof, or a sequence at least 85% identical thereto.

38. The method of claim 37, wherein the contacting step is effected *in vitro* or cells in culture.

39. The method of claim 37, wherein the contacting step is effected in a subject.

40. The method of claim 37, wherein the contacting step is effected in an animal subject.

41. A method of treating a subject having cancer, comprising:

acquiring knowledge of the presence of a DCTN1-ALK fusion or an LMNA-NTRK1 fusion in said subject; and

administering to the subject an effective amount of an anti cancer agent, thereby treating cancer in the subject,

wherein:

(i) the DCTN1-ALK fusion comprises the amino acid sequence of SEQ ID NO:7 or a fragment thereof, or a sequence at least 85% identical thereto; or

(ii) the LMNA-NTRK1 fusion comprises the amino acid sequence of SEQ ID NO:10 or a fragment thereof, or a sequence at least 85% identical thereto.

42. The method of claim 41, wherein said anti-cancer agent is a kinase inhibitor.

43. The method of claim 42, wherein the kinase inhibitor is administered responsive to the determination of presence of the fusion in a tumor sample from said subject.

IDT Exhibit 1020.3061

44. The method of claim 43, wherein said administration is administered responsive to acquiring knowledge or information of the presence of the fusion in said subject.

45. The method of claim 43, wherein said administration is administered responsive to acquiring knowledge or information of the presence of the fusion in said subject from another party.

46. The method of claim 43, wherein the method comprises receiving a communication of the presence of the fusion in a subject.

47. The method of claim 43, wherein said administration is administered responsive to an identification of the presence of the fusion in a subject, wherein said identification arises from collaboration with another party.

48. The method of claim 43, comprising determining the presence of the fusion by sequencing.

49. The method of claim 43, wherein said cancer is chosen from a Spitz tumor, melanoma, colorectal cancer, lung cancer, pancreatic cancer, esophageal-gastric cancer, thyroid cancer, or adenocarcinoma.

50. The method of claim 49, wherein said cancer is a Spitz tumor, melanoma.

51. The method of claim 49, wherein said lung cancer is chosen from: small cell lung cancer (SCLC), adenocarcinoma of the lung, bronchogenic carcinoma, or a combination thereof.

52. The method of claim 49, wherein the lung cancer is non-small cell lung cancer (NSCLC).

IDT Exhibit 1020.3062

53. The method of claim 49, wherein the lung cancer is squamous cell carcinoma (SCC).

54. The method of claim 41, wherein the anti-cancer agent is a multi-kinase inhibitor.

55. The method of claim 41, wherein the anti-cancer agent is an ALK-specific inhibitor.

56. The method of claim 41, wherein the anti-cancer agent is TAE-684 or PF02341066 (crizotinib).

57. The method of claim 41, wherein the anti-cancer agent is AF-802, LDK-378, or ASP-3026.

58. The method of claim 41, wherein the anti-cancer agent is CEP-37440, CEP-28122, or CEP-108050.

59. The method of claim 41, wherein the anti-cancer agent is AP26113.

60. The method of claim 43, wherein the kinase inhibitor inhibits the kinase activity of the fusion.

61. The method of claim 43, wherein the kinase inhibitor is selected from antisense molecules, ribozymes, RNAi, triple helix molecules that hybridize to a nucleic acid encoding the fusion, or a transcription regulatory region that blocks or reduces mRNA expression of the fusion.

62. The method of claim 43, wherein the kinase inhibitor is administered in combination with a second therapeutic agent or a different therapeutic modality.

IDT Exhibit 1020.3063

63. A method for screening for an agent that modulates, *e.g.*, inhibits, the expression or activity of a DCTN1-ALK fusion or an LMNA-NTRK1 fusion, comprising:

optionally, determining if the fusion is present;

contacting the fusion (or a host cell expressing the fusion) with a candidate agent; and

detecting a change in a parameter associated with the fusion.

64. The method of claim 63, wherein said parameter is the expression or an activity of the fusion.

65. The method of claim 63, further comprising comparing a value for the parameter to a reference value.

66. The method of claim 63, comprising comparing a parameter obtained from a sample with the candidate agent to a parameter obtained from a sample without the candidate agent.

67. The method of claim 63, further comprising, if a decrease in expression or activity of the fusion is detected, indentifying or classifying the candidate agent as an inhibitor.

68. The method of claim 63, wherein said contacting occurs in a cell-free system.

69. The method of claim 63, wherein said contacting is effected in a cell in culture.

70. The method of claim 63, wherein said parameter is selected from one or more of:

(i) direct binding of the candidate agent to the fusion polypeptide;

(ii) a change in kinase activity;

IDT Exhibit 1020.3064

(iii) a change in an activity of a cell containing the fusion, *e.g.*, a change in proliferation, morphology or tumorigenicity of the cell;

(iv) a change in tumor present in an animal subject, *e.g.*, size, appearance, proliferation, of the tumor; or

(v) a change in the level, *e.g.*, expression level, of the fusion polypeptide or nucleic acid molecule.

71. A method of determining the presence of a DCTN1-ALK fusion or an LMNA-NTRK1 fusion, comprising:

directly acquiring knowledge that a fusion nucleic acid molecule of any of claims 1-3, or a polypeptide of any of claims 16-18 is present in a sample from a subject.

72. The method of claim 71, wherein said sample comprised, fluid, cells, tissue, e.g., a tumor tissue.

73. The method of claim 71, wherein the sample is a nucleic acid sample.

74. The method of claim 71, wherein the sample is a protein sample.

75. The method of claim 71, wherein the sample is acquired from a subject.

76. The method of claim 71, wherein the sample comprises a tumor biopsy or a circulating tumor cell or nucleic acid.

77. The method of claim 71, wherein the tumor is chosen from a Spitz tumor, a melanoma, a colorectal cancer, a lung cancer, an esophageal-gastric cancer, a thyroid cancer, or an adenocarcinoma.

78. The method of claim 71, wherein the tumor is from a lung cancer.

IDT Exhibit 1020.3065

79. The method of claim 71, wherein the tumor is a NSCLC, a SCLC, a SCC, or a combination thereof.

80. The method of claim 71, wherein the fusion is detected in a nucleic acid molecule.

81. The method of claim 80, comprising sequencing by a method chosen from one or more of: nucleic acid hybridization assay, amplification-based assays, PCR-RFLP assay, real-time PCR, sequencing, screening analysis, FISH, spectral karyotyping or MFISH, comparative genomic hybridization), *in situ* hybridization, SSP, HPLC or mass-spectrometric genotyping.

82. The method of claim 71, comprising acquiring a sequence for a position in a nucleic acid molecule by sequencing at least one nucleotide of the nucleic acid molecule, thereby determining that the fusion is present in the nucleic acid molecule.

83. The method of claim 82, wherein the sequence acquired is compared to a reference sequence, or a wild type reference sequence.

84. The method of claim 71, wherein the fusion polypeptide is detected.

85. The method of claim 71, comprising: contacting a protein sample with a reagent which specifically binds to a fusion polypeptide; and detecting the formation of a complex of the fusion polypeptide and the reagent.

86. The method of claim 85, wherein the reagent is labeled with a detectable group to facilitate detection of the bound and unbound reagent.

87. The method of claim 86, wherein the reagent is an antibody molecule.

IDT Exhibit 1020.3066

88. The method of claim 71, wherein the level or activity the fusion is evaluated.

89. The method of claim 71, wherein the fusion is detected prior to initiating, during, or after, a treatment in a subject.

90. The method of claim 71, wherein the fusion is detected at the time of diagnosis with a cancer.

91. The method of claim 71, wherein the fusion is detected at a pre-determined interval, *e.g.*, a first point in time and at least at a subsequent point in time.

92. The method of claim 71, comprising, responsive to a determination of the presence of the fusion, one or more of:

(1) stratifying a patient population;

(2) identifying or selecting the subject as likely or unlikely to respond to a treatment;

(3) selecting a treatment option; or

(4) prognosticating the time course of the disease in the subject.

93. The method of claim 92, wherein the treatment option comprises administering an ALK kinase inhibitor.

94. The method of claim 93, wherein the kinase inhibitor is TAE-684 or PF02341066 (crizotinib).

95. The method of claim 93, wherein the kinase inhibitor is AF-802, LDK-378, or ASP-3026.

96. The method of claim 93, wherein the kinase inhibitor is vandetanib (ZACTIMA®, ZD6474).

IDT Exhibit 1020.3067

97. The method of claim 93, wherein the kinase inhibitor is CEP-37440, CEP-28122, CEP-108050 or AP26113.

98. A method of evaluating a patient, comprising:

identifying, selecting, or obtaining information or knowledge that the patient has participated in a clinical trial or has been treated for cancer;

acquiring genotype information that identifies a DCTN1-ALK fusion or an LMNA-NTRK1 fusion as being in the patient,

wherein the presence of the fusion is identifies the patient as having an increased risk for, or having, a cancer associated with the fusion; and

treating the subject with a kinase inhibitor.

99. The method of claim 98, further comprising providing a report to another party.

100. The method of claim 99, wherein the party is the patient, a caregiver, a physician, an oncologist, a hospital, clinic, third-party payor, insurance company or government office.

101. The method of claim 99, wherein said report is in electronic, web-based, or paper form.

102. The method of claim 99, wherein a report is generated which identifies the presence or absence of the mutation, and optionally includes an identifier for the patient from which the sequence was obtained.

103. The method of claim 102, wherein said report comprises;

information on the role of the fusion, or wildtype sequence, in disease;

information on prognosis, resistance, or potential or suggested therapeutic options;

IDT Exhibit 1020.3068

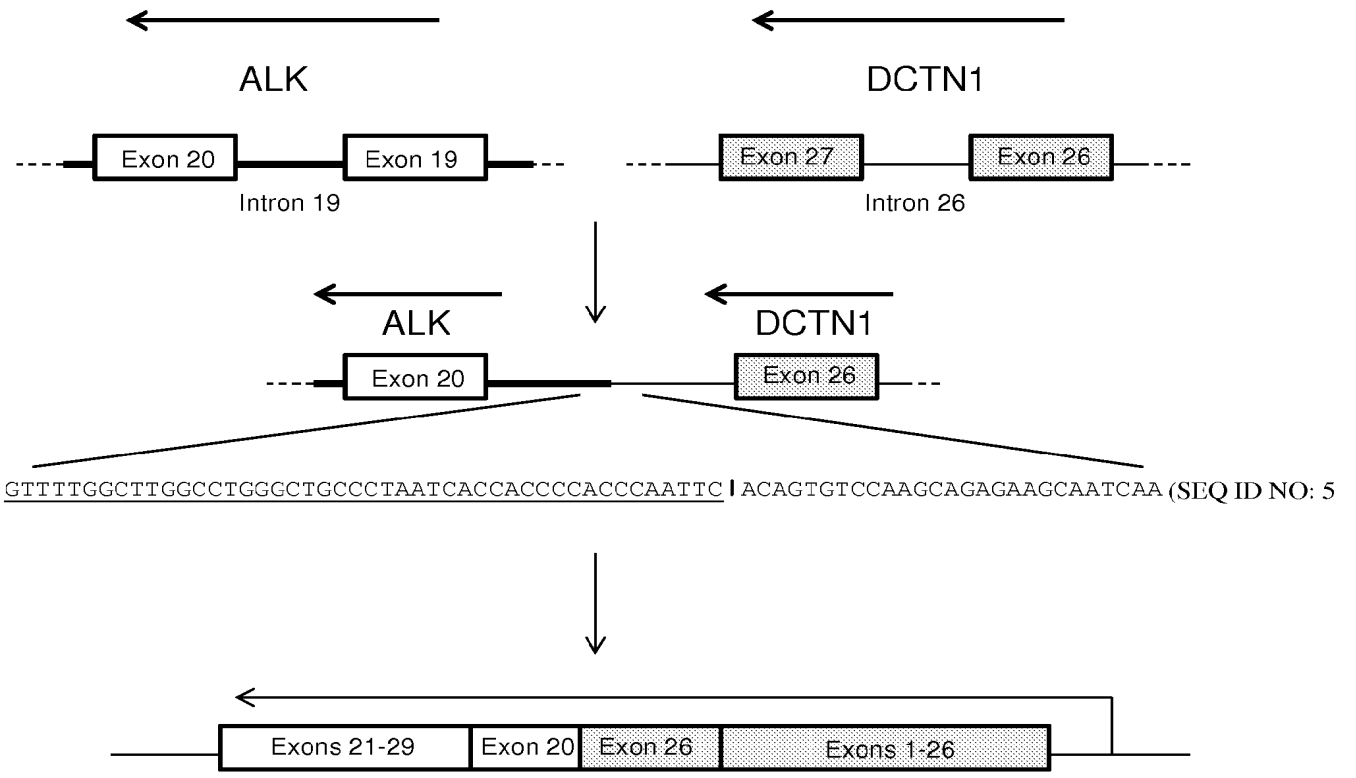
information on the likely effectiveness of a therapeutic option, the acceptability of a therapeutic option, or the advisability of applying the therapeutic option to a patient; or

information, or a recommendation on, the administration of a drug.

104. A method for generating a personalized cancer treatment report, by obtaining a sample from a subject, detecting a DCTN1-ALK fusion or an LMNA-NTRK1 fusion in the sample, and selecting a treatment based on the mutation identified.

IDT Exhibit 1020.3069

FIG. 1



IDT Exhibit 1020.3070

FIG. 2A

```
atg gca cag agc aag agg cac gtg tac agc cgg acg ccc agc ggc agc agg atg agt gcg   60
  M   A   Q   S   K   R   H   V   Y   S   R   T   P   S   G   S   R   M   S   A  20

gag gca agc gcc cgg cct ctg cgg gtg ggc tcc cgt gta gag gtg att gga aaa ggc cac  120
  E   A   S   A   R   P   L   R   V   G   S   R   V   E   V   I   G   K   G   H  40

cga ggc act gtg gcc tat gtt gga gcc aca ctg ttt gcc act ggc aaa tgg gta ggc gtg  180
  R   G   T   V   A   Y   V   G   A   T   L   F   A   T   G   K   W   V   G   V  60

att ctg gat gaa gca aag ggc aaa aat gat gga act gtt caa ggc agg aag tac ttc act  240
  I   L   D   E   A   K   G   K   N   D   G   T   V   Q   G   R   K   Y   F   T  80

tgt gat gaa ggg cat ggc atc ttt gtg cgc cag tcc cag atc cag gta ttt gaa gat gga  300
  C   D   E   G   H   G   I   F   V   R   Q   S   Q   I   Q   V   F   E   D   G 100

gca gat act act tcc cca gag aca cct gat tct tct gct tca aaa gtc ctc aaa aga gag  360
  A   D   T   T   S   P   E   T   P   D   S   S   A   S   K   V   L   K   R   E 120

gga act gat aca act gca aag act agc aaa ctg cgg gga ctg aag cct aag aag gca ccg  420
  G   T   D   T   T   A   K   T   S   K   L   R   G   L   K   P   K   K   A   P 140

aca gcc cga aag acc aca act cgg cga ccc aag ccc acg cgc cca gcc agt act ggg gtg  480
  T   A   R   K   T   T   T   R   R   P   K   P   T   R   P   A   S   T   G   V 160

gct ggg gcc agt agc tcc ctg ggc ccc tct ggc tca gcg tca gca ggt gag ctg agc agc  540
  A   G   A   S   S   S   L   G   P   S   G   S   A   S   A   G   E   L   S   S 180

agt gag ccc agc acc ccg gct cag act ccg ctg gca gca ccc atc atc ccc acg ccg gtc  600
  S   E   P   S   T   P   A   Q   T   P   L   A   A   P   I   I   P   T   P   V 200

ctc acc tct cct gga gca gtc ccc ccg ctt cct tcc cca tcc aag gag gag gag gga cta   660
  L   T   S   P   G   A   V   P   P   L   P   S   P   S   K   E   E   E   G   L 220

agg gct cag gtg cgg gac ctg gag gag aaa cta gag acc ctg aga ctg aaa cgg gca gaa   720
  R   A   Q   V   R   D   L   E   E   K   L   E   T   L   R   L   K   R   A   E 240

gac aaa gca aag cta aaa gag ctg gag aaa cac aaa atc cag ctg gag cag gtg cag gaa   780
  D   K   A   K   L   K   E   L   E   K   H   K   I   Q   L   E   Q   V   Q   E 260

tgg aag agc aaa atg cag gag cag cag gcc gac ctg cag cgg cgc ctc aag gag gcg aga   840
  W   K   S   K   M   Q   E   Q   Q   A   D   L   Q   R   R   L   K   E   A   R 280

aag gaa gcc aag gag gcg ctg gag gca aag gaa cgc tat atg gag gag atg gct gat act   900
  K   E   A   K   E   A   L   E   A   K   E   R   Y   M   E   E   M   A   D   T 300

gct gat gcc att gag atg gcc act ttg gac aag gag atg gct gaa gag cgg gct gag tcc   960
  A   D   A   I   E   M   A   T   L   D   K   E   M   A   E   E   R   A   E   S 320
```


IDT Exhibit 1020.3071

FIG. 2B

```
ctg cag cag gag gtg gag gca ctg aag gag cgg gtg gac gag ctc act act gac tta gag  1020
  L   Q   Q   E   V   E   A   L   K   E   R   V   D   E   L   T   T   D   L   E   340

atc ctc aag gct gag att gaa gag aag ggc tca gat ggc gct gca tcc agt tat cag ctc  1080
  I   L   K   A   E   I   E   E   K   G   S   D   G   A   A   S   S   Y   Q   L   360

aag cag ctt gag gag cag aat gcc cgc ctg aag gat gcc ctg gtg agg atg cgg gat ctt  1140
  K   Q   L   E   E   Q   N   A   R   L   K   D   A   L   V   R   M   R   D   L   380

tct tcc tca gag aag cag gag cat gtg aag ctc cag aag ctc atg gaa aag aag aac caa  1200
  S   S   S   E   K   Q   E   H   V   K   L   Q   K   L   M   E   K   K   N   Q   400

gag ctg gaa gtt gtg agg caa cag cgg gag cgt ctg cag gag gag cta agc cag gca gag  1260
  E   L   E   V   V   R   Q   Q   R   E   R   L   Q   E   E   L   S   Q   A   E   420

agc acc att gat gag ctc aag gag cag gtg gat gct gct ctg ggt gct gag gag atg gtg  1320
  S   T   I   D   E   L   K   E   Q   V   D   A   A   L   G   A   E   E   M   V   440

gag atg ctg aca gat cgg aac ctg aat ctg gaa gag aaa gtg cgc gag ttg agg gag act  1380
  E   M   L   T   D   R   N   L   N   L   E   E   K   V   R   E   L   R   E   T   460

gtg gga gac ttg gaa gcg atg aat gag atg aac gat gag ctg cag gag aat gca cgt gag  1440
  V   G   D   L   E   A   M   N   E   M   N   D   E   L   Q   E   N   A   R   E   480

aca gaa ctg gag ctg cgg gag cag ctg gac atg gca ggc gcg cgg gtt cgt gag gcc cag  1500
  T   E   L   E   L   R   E   Q   L   D   M   A   G   A   R   V   R   E   A   Q   500

aag cgt gtg gag gca gcc cag gag acg gtt gca gac tac cag cag acc atc aag aag tac  1560
  K   R   V   E   A   A   Q   E   T   V   A   D   Y   Q   Q   T   I   K   K   Y   520

cgc cag ctg acc gcc cat cta cag gat gtg aat cgg gaa ctg aca aac cag cag gaa gca  1620
  R   Q   L   T   A   H   L   Q   D   V   N   R   E   L   T   N   Q   Q   E   A   540

tct gtg gag agg caa cag cag cca cct cca gag acc ttt gac ttc aaa atc aag ttt gct  1680
  S   V   E   R   Q   Q   Q   P   P   P   E   T   F   D   F   K   I   K   F   A   560

gag act aag gcc cat gcc aag gca att gag atg gaa ttg agg cag atg gag gtg gcc cag  1740
  E   T   K   A   H   A   K   A   I   E   M   E   L   R   Q   M   E   V   A   Q   580

gcc aat cga cac atg tcc ctg ctg aca gcc ttc atg cct gac agc ttc ctt cgg cca ggt  1800
  A   N   R   H   M   S   L   L   T   A   F   M   P   D   S   F   L   R   P   G   600

ggg gac cat gac tgc gtt ctg gtg ctg ttg ctc atg cct cgt ctc att tgc aag gca gag  1860
  G   D   H   D   C   V   L   V   L   L   L   M   P   R   L   I   C   K   A   E   620

ctg atc cgg aag cag gcc cag gag aag ttt gaa cta agt gag aac tgt tca gag cgg cct  1920
L   I   R   K   Q   A   Q   E   K   F   E   L   S   E   N   C   S   E   R   P     640
```


IDT Exhibit 1020.3072

FIG. 2C

```
ggg ctg cga gga gct gct ggg gag caa ctc agc ttt gct gct gga ctg gtg tac tcg ctg   1980
  G   L   R   G   A   A   G   E   Q   L   S   F   A   A   G   L   V   Y   S   L   660

agc ctg ctg cag gcc acg cta cac cgc tat gag cat gcc ctc tct cag tgc agt gtg gat   2040
  S   L   L   Q   A   T   L   H   R   Y   E   H   A   L   S   Q   C   S   V   D   680

gtg tat aag aaa gtg ggc agc ctg tac cct gag atg agt gcc cat gag cgc tcc ttg gat   2100
  V   Y   K   K   V   G   S   L   Y   P   E   M   S   A   H   E   R   S   L   D   700

ttc ctc att gaa ctg ctg cac aag gat cag ctg gat gag act gtc aat gtg gag cct ctc   2160
  F   L   I   E   L   L   H   K   D   Q   L   D   E   T   V   N   V   E   P   L   720

acc aag gcc atc aag tac tat cag cat ctg tac agc atc cac ctt gcc gaa cag cct gag   2220
  T   K   A   I   K   Y   Y   Q   H   L   Y   S   I   H   L   A   E   Q   P   E   740

gac tgt act atg cag ctg gct gac cac att aag ttc acg cag agt gct ctg gac tgc atg   2280
  D   C   T   M   Q   L   A   D   H   I   K   F   T   Q   S   A   L   D   C   M   760

agt gtg gag gta gga cgg ctg cgt gcc ttc ttg cag ggt ggg cag gag gct aca gat att   2340
  S   V   E   V   G   R   L   R   A   F   L   Q   G   G   Q   E   A   T   D   I   780

gcc ctc ctg ctc cgg gat ctg gaa act tca tgc agt gac atc cgc cag ttc tgc aag aag   2400
  A   L   L   L   R   D   L   E   T   S   C   S   D   I   R   Q   F   C   K   K   800

atc cga agg cga atg cca ggg aca gat gct cct ggg atc cca gct gca ctg gcc ttt gga   2460
  I   R   R   R   M   P   G   T   D   A   P   G   I   P   A   A   L   A   F   G   820

cca cag gta tct gac acg ctc cta gac tgc agg aaa cac ttg acg tgg gtc gtg gct gtg   2520
  P   Q   V   S   D   T   L   L   D   C   R   K   H   L   T   W   V   V   A   V   840

ctg cag gag gtg gca gct gct gct gcc cag ctc att gcc cca ctg gca gag aat gag ggg   2580
  L   Q   E   V   A   A   A   A   A   Q   L   I   A   P   L   A   E   N   E   G   860

cta ctt gtg gct gct ctg gag gaa ctg gct ttc aaa gca agc gag cag atc tat ggg acc   2640
  L   L   V   A   A   L   E   E   L   A   F   K   A   S   E   Q   I   Y   G   T   880

ccc tcc agc agc ccc tat gag tgt ctg cgc cag tca tgc aac atc ctc atc agt acc atg   2700
  P   S   S   S   P   Y   E   C   L   R   Q   S   C   N   I   L   I   S   T   M   900

aac aag ctg gcc aca gcc atg cag gag ggg gag tat gat gca gag cgg ccc ccc agc aag   2760
  N   K   L   A   T   A   M   Q   E   G   E   Y   D   A   E   R   P   P   S   K   920

cct cca ccg gtt gaa ctg cgg gct gct gcc ctt cgt gca gag atc aca gat gct gaa ggc   2820
  P   P   P   V   E   L   R   A   A   A   L   R   A   E   I   T   D   A   E   G   940
```


IDT Exhibit 1020.3073

FIG. 2D

```
ctg ggt ttg aag ctc gaa gat cga gag aca gtt att aag gag ttg aag aag tca ctc aag  2880
  L   G   L   K   L   E   D   R   E   T   V   I   K   E   L   K   K   S   L   K  960

att aag gga gag gag cta agt gag gcc aat gtg cgg ctg agc ctc ctg gag aag aag ttg  2940
  I   K   G   E   E   L   S   E   A   N   V   R   L   S   L   L   E   K   K   L  980

gac agt gct gcc aag gat gca gat gag cgc atc gag aaa gtc cag act cgg ctg gag gag 3000
  D   S   A   A   K   D   A   D   E   R   I   E   K   V   Q   T   R   L   E   E 1000

acc cag gca ctg ctg cga aag aag gag aaa gag ttt gag gag aca atg gat gca ctc cag 3060
  T   Q   A   L   L   R   K   K   E   K   E   F   E   E   T   M   D   A   L   Q 1020

gct gac atc gac cag ctg gag gca gag aag gca gaa cta aag cag cgt ctg aac agc cag 3120
  A   D   I   D   Q   L   E   A   E   K   A   E   L   K   Q   R   L   N   S   Q 1040

tcc aaa cgc acg att gag gga ctc cgg ggc cct cct cct tca ggc att gct act ctg gtc 3180
S   K   R   T   I   E   G   L   R   G   P   P   P   S   G   I   A   T   L   V  1060

tct ggc att gct ggt gtg tac cgc cgg aag cac cag gag ctg caa gcc atg cag atg gag 3240
  S   G   I   A   G   V   Y   R   R   K   H   Q   E   L   Q   A   M   Q   M   E 1080

ctg cag agc cct gag tac aag ctg agc aag ctc cgc acc tcg acc atc atg acc gac tac 3300
  L   Q   S   P   E   Y   K   L   S   K   L   R   T   S   T   I   M   T   D   Y 1100

aac ccc aac tac tgc ttt gct ggc aag acc tcc tcc atc agt gac ctg aag gag gtg ccg 3360
  N   P   N   Y   C   F   A   G   K   T   S   S   I   S   D   L   K   E   V   P 1120

cgg aaa aac atc acc ctc att cgg ggt ctg ggc cat ggc gcc ttt ggg gag gtg tat gaa 3420
  R   K   N   I   T   L   I   R   G   L   G   H   G   A   F   G   E   V   Y   E 1140

ggc cag gtg tcc gga atg ccc aac gac cca agc ccc ctg caa gtg gct gtg aag acg ctg 3480
  G   Q   V   S   G   M   P   N   D   P   S   P   L   Q   V   A   V   K   T   L 1160

cct gaa gtg tgc tct gaa cag gac gaa ctg gat ttc ctc atg gaa gcc ctg atc atc agc 3540
  P   E   V   C   S   E   Q   D   E   L   D   F   L   M   E   A   L   I   I   S 1180

aaa ttc aac cac cag aac att gtt cgc tgc att ggg gtg agc ctg caa tcc ctg ccc cgg 3600
  K   F   N   H   Q   N   I   V   R   C   I   G   V   S   L   Q   S   L   P   R 1200

ttc atc ctg ctg gag ctc atg gcg ggg gga gac ctc aag tcc ttc ctc cga gag acc cgc  3660
  F   I   L   L   E   L   M   A   G   G   D   L   K   S   F   L   R   E   T   R 1220

cct cgc ccg agc cag ccc tcc tcc ctg gcc atg ctg gac ctt ctg cac gtg gct cgg gac  3720
  P   R   P   S   Q   P   S   S   L   A   M   L   D   L   L   H   V   A   R   D 1240

att gcc tgt ggc tgt cag tat ttg gag gaa aac cac ttc atc cac cga gac att gct gcc  3780
  I   A   C   G   C   Q   Y   L   E   E   N   H   F   I   H   R   D   I   A   A 1260
```


IDT Exhibit 1020.3074

FIG. 2E

```
aga aac tgc ctc ttg acc tgt cca ggc cct gga aga gtg gcc aag att gga gac ttc ggg 3840
  R   N   C   L   L   T   C   P   G   P   G   R   V   A   K   I   G   D   F   G 1280

atg gcc cga gac atc tac agg gcg agc tac tat aga aag gga ggc tgt gcc atg ctg cca 3900
  M   A   R   D   I   Y   R   A   S   Y   Y   R   K   G   G   C   A   M   L   P 1300

gtt aag tgg atg ccc cca gag gcc ttc atg gaa gga ata ttc act tct aaa aca gac aca 3960
  V   K   W   M   P   P   E   A   F   M   E   G   I   F   T   S   K   T   D   T 1320

tgg tcc ttt gga gtg ctg cta tgg gaa atc ttt tct ctt gga tat atg cca tac ccc agc 4020
  W   S   F   G   V   L   L   W   E   I   F   S   L   G   Y   M   P   Y   P   S 1340

aaa agc aac cag gaa gtt ctg gag ttt gtc acc agt gga ggc cgg atg gac cca ccc aag 4080
  K   S   N   Q   E   V   L   E   F   V   T   S   G   G   R   M   D   P   P   K 1360

aac tgc cct ggg cct gta tac cgg ata atg act cag tgc tgg caa cat cag cct gaa gac 4140
  N   C   P   G   P   V   Y   R   I   M   T   Q   C   W   Q   H   Q   P   E   D 1380

agg ccc aac ttt gcc atc att ttg gag agg att gaa tac tgc acc cag gac ccg gat gta 4200
  R   P   N   F   A   I   I   L   E   R   I   E   Y   C   T   Q   D   P   D   V 1400

atc aac acc gct ttg ccg ata gaa tat ggt cca ctt gtg gaa gag gaa gag aaa gtg cct 4260
  I   N   T   A   L   P   I   E   Y   G   P   L   V   E   E   E   E   K   V   P 1420

gtg agg ccc aag gac cct gag ggg gtt cct cct ctc ctg gtc tct caa cag gca aaa cgg 4320
  V   R   P   K   D   P   E   G   V   P   P   L   L   V   S   Q   Q   A   K   R 1440

gag gag gag cgc agc cca gct gcc cca cca cct ctg cct acc acc tcc tct ggc aag gct 4380
  E   E   E   R   S   P   A   A   P   P   P   L   P   T   T   S   S   G   K   A 1460

gca aag aaa ccc aca gct gca gag atc tct gtt cga gtc cct aga ggg ccg gcc gtg gaa 4440
  A   K   K   P   T   A   A   E   I   S   V   R   V   P   R   G   P   A   V   E 1480

ggg gga cac gtg aat atg gca ttc tct cag tcc aac cct cct tcg gag ttg cac aag gtc 4500
  G   G   H   V   N   M   A   F   S   Q   S   N   P   P   S   E   L   H   K   V 1500

cac gga tcc aga aac aag ccc acc agc ttg tgg aac cca acg tac ggc tcc tgg ttt aca 4560
  H   G   S   R   N   K   P   T   S   L   W   N   P   T   Y   G   S   W   F   T 1520

gag aaa ccc acc aaa aag aat aat cct ata gca aag aag gag cca cac gac agg ggt aac 4620
  E   K   P   T   K   K   N   N   P   I   A   K   K   E   P   H   D   R   G   N 1540

ctg ggg ctg gag gga agc tgt act gtc cca cct aac gtt gca act ggg aga ctt ccg ggg 4700
  L   G   L   E   G   S   C   T   V   P   P   N   V   A   T   G   R   L   P   G 1560
```


IDT Exhibit 1020.3075

FIG. 2F

```
gcc tca ctg ctc cta gag ccc tct tcg ctg act gcc aat atg aag gag gta cct ctg ttc  4760
  A   S   L   L   L   E   P   S   S   L   T   A   N   M   K   E   V   P   L   F  1580

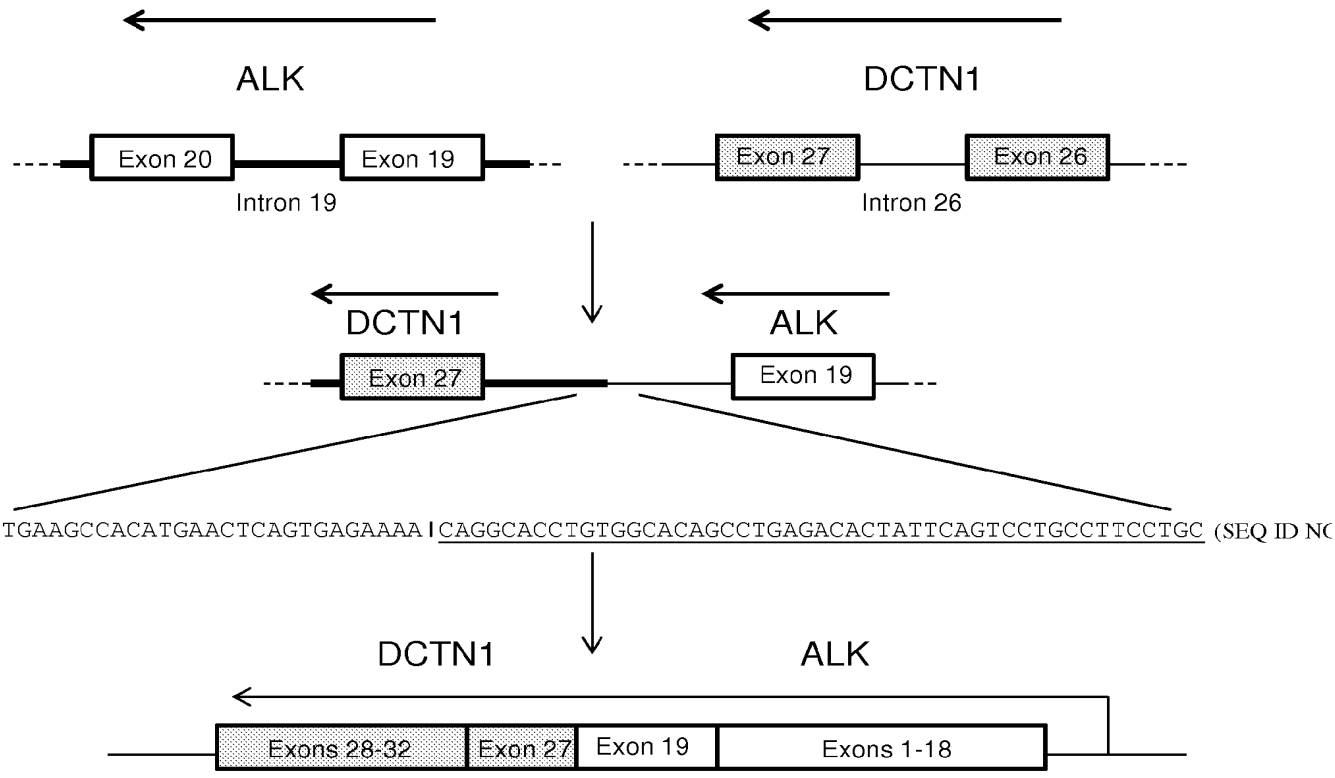
agg cta cgt cac ttc cct tgt ggg aat gtc aat tac ggc tac cag caa cag ggc ttg ccc  4820
  R   L   R   H   F   P   C   G   N   V   N   Y   G   Y   Q   Q   Q   G   L   P  1600

tta gaa gcc gct act gcc cct gga gct ggt cat tac gag gat acc att ctg aaa agc aag  4880
  L   E   A   A   T   A   P   G   A   G   H   Y   E   D   T   I   L   K   S   K  1620

aat agc atg aac cag cct ggg ccc tga  4907 (SEQ ID NO: 6)
  N   S   M   N   Q   P   G   P   *  1640 (SEQ ID NO: 7)
```


IDT Exhibit 1020.3076

FIG. 3



IDT Exhibit 1020.3077

FIG. 4A

```
atg gga gcc atc ggg ctc ctg tgg ctc ctg ccg ctg ctg ctt tcc acg gca gct gtg ggc  60
  M   G   A   I   G   L   L   W   L   L   P   L   L   L   S   T   A   A   V   G  20

tcc ggg atg ggg acc ggc cag cgc gcg ggc tcc cca gct gcg ggg ccg ccg ctg cag ccc 120
  S   G   M   G   T   G   Q   R   A   G   S   P   A   A   G   P   P   L   Q   P  40

cgg gag cca ctc agc tac tcg cgc ctg cag agg aag agt ctg gca gtt gac ttc gtg gtg 180
  R   E   P   L   S   Y   S   R   L   Q   R   K   S   L   A   V   D   F   V   V  60

ccc tcg ctc ttc cgt gtc tac gcc cgg gac cta ctg ctg cca cca tcc tcc tcg gag ctg 240
  P   S   L   F   R   V   Y   A   R   D   L   L   L   P   P   S   S   S   E   L  80

aag gct ggc agg ccc gag gcc cgc ggc tcg cta gct ctg gac tgc gcc ccg ctg ctc agg 300
  K   A   G   R   P   E   A   R   G   S   L   A   L   D   C   A   P   L   L   R 100

ttg ctg ggg ccg gcg ccg ggg gtc tcc tgg acc gcc ggt tca cca gcc ccg gca gag gcc 360
  L   L   G   P   A   P   G   V   S   W   T   A   G   S   P   A   P   A   E   A 120

cgg acg ctg tcc agg gtg ctg aag ggc ggc tcc gtg cgc aag ctc cgg cgt gcc aag cag 420
  R   T   L   S   R   V   L   K   G   G   S   V   R   K   L   R   R   A   K   Q 140

ttg gtg ctg gag ctg ggc gag gag gcg atc ttg gag ggt tgc gtc ggg ccc ccc ggg gag 480
  L   V   L   E   L   G   E   E   A   I   L   E   G   C   V   G   P   P   G   E 160

gcg gct gtg ggg ctg ctc cag ttc aat ctc agc gag ctg ttc agt tgg tgg att cgc caa  540
  A   A   V   G   L   L   Q   F   N   L   S   E   L   F   S   W   W   I   R   Q  180

ggc gaa ggg cga ctg agg atc cgc ctg atg ccc gag aag aag gcg tcg gaa gtg ggc aga  600
  G   E   G   R   L   R   I   R   L   M   P   E   K   K   A   S   E   V   G   R  200

gag gga agg ctg tcc gcg gca att cgc gcc tcc cag ccc cgc ctt ctc ttc cag atc ttc  660
  E   G   R   L   S   A   A   I   R   A   S   Q   P   R   L   L   F   Q   I   F  220

ggg act ggt cat agc tcc ttg gaa tca cca aca aac atg cct tct cct tct cct gat tat  720
  G   T   G   H   S   S   L   E   S   P   T   N   M   P   S   P   S   P   D   Y  240

ttt aca tgg aat ctc acc tgg ata atg aaa gac tcc ttc cct ttc ctg tct cat cgc agc  780
  F   T   W   N   L   T   W   I   M   K   D   S   F   P   F   L   S   H   R   S  260

cga tat ggt ctg gag tgc agc ttt gac ttc ccc tgt gag ctg gag tat tcc cct cca ctg  840
  R   Y   G   L   E   C   S   F   D   F   P   C   E   L   E   Y   S   P   P   L  280

cat gac ctc agg aac cag agc tgg tcc tgg cgc cgc atc ccc tcc gag gag gcc tcc cag  900
  H   D   L   R   N   Q   S   W   S   W   R   R   I   P   S   E   E   A   S   Q  300

atg gac ttg ctg gat ggg cct ggg gca gag cgt tct aag gag atg ccc aga ggc tcc ttt  960
  M   D   L   L   D   G   P   G   A   E   R   S   K   E   M   P   R   G   S   F  320
```


IDT Exhibit 1020.3078

FIG. 4B

```
ctc ctt ctc aac acc tca gct gac tcc aag cac acc atc ctg agt ccg tgg atg agg agc  1020
  L   L   L   N   T   S   A   D   S   K   H   T   I   L   S   P   W   M   R   S   340

agc agt gag cac tgc aca ctg gcc gtc tcg gtg cac agg cac ctg cag ccc tct gga agg  1080
  S   S   E   H   C   T   L   A   V   S   V   H   R   H   L   Q   P   S   G   R   360

tac att gcc cag ctg ctg ccc cac aac gag gct gca aga gag atc ctc ctg atg ccc act  1140
  Y   I   A   Q   L   L   P   H   N   E   A   A   R   E   I   L   L   M   P   T   380

cca ggg aag cat ggt tgg aca gtg ctc cag gga aga atc ggg cgt cca gac aac cca ttt  1200
  P   G   K   H   G   W   T   V   L   Q   G   R   I   G   R   P   D   N   P   F   400

cga gtg gcc ctg gaa tac atc tcc agt gga aac cgc agc ttg tct gca gtg gac ttc ttt  1260
  R   V   A   L   E   Y   I   S   S   G   N   R   S   L   S   A   V   D   F   F   420

gcc ctg aag aac tgc agt gaa gga aca tcc cca ggc tcc aag atg gcc ctg cag agc tcc  1320
  A   L   K   N   C   S   E   G   T   S   P   G   S   K   M   A   L   Q   S   S   440

ttc act tgt tgg aat ggg aca gtc ctc cag ctt ggg cag gcc tgt gac ttc cac cag gac  1380
  F   T   C   W   N   G   T   V   L   Q   L   G   Q   A   C   D   F   H   Q   D   460

tgt gcc cag gga gaa gat gag agc cag atg tgc cgg aaa ctg cct gtg ggt ttt tac tgc  1440
  C   A   Q   G   E   D   E   S   Q   M   C   R   K   L   P   V   G   F   Y   C   480

aac ttt gaa gat ggc ttc tgt ggc tgg acc caa ggc aca ctg tca ccc cac act cct caa  1500
  N   F   E   D   G   F   C   G   W   T   Q   G   T   L   S   P   H   T   P   Q   500

tgg cag gtc agg acc cta aag gat gcc cgg ttc cag gac cac caa gac cat gct cta ttg  1560
  W   Q   V   R   T   L   K   D   A   R   F   Q   D   H   Q   D   H   A   L   L   520

ctc agt acc act gat gtc ccc gct tct gaa agt gct aca gtg acc agt gct acg ttt cct  1620
  L   S   T   T   D   V   P   A   S   E   S   A   T   V   T   S   A   T   F   P   540

gca ccg atc aag agc tct cca tgt gag ctc cga atg tcc tgg ctc att cgt gga gtc ttg  1680
  A   P   I   K   S   S   P   C   E   L   R   M   S   W   L   I   R   G   V   L   560

agg gga aac gtg tcc ttg gtg cta gtg gag aac aaa acc ggg aag gag caa ggc agg atg  1740
  R   G   N   V   S   L   V   L   V   E   N   K   T   G   K   E   Q   G   R   M   580

gtc tgg cat gtc gcc gcc tat gaa ggc ttg agc ctg tgg cag tgg atg gtg ttg cct ctc  1800
  V   W   H   V   A   A   Y   E   G   L   S   L   W   Q   W   M   V   L   P   L   600

ctc gat gtg tct gac agg ttc tgg ctg cag atg gtc gca tgg tgg gga caa gga tcc aga  1860
  L   D   V   S   D   R   F   W   L   Q   M   V   A   W   W   G   Q   G   S   R   620

gcc atc gtg gct ttt gac aat atc tcc atc agc ctg gac tgc tac ctc acc att agc gga  1920
  A   I   V   A   F   D   N   I   S   I   S   L   D   C   Y   L   T   I   S   G   640
```


IDT Exhibit 1020.3079

FIG. 4C

```
gag gac aag atc ctg cag aat aca gca ccc aaa tca aga aac ctg ttt gag aga aac cca  1980
  E   D   K   I   L   Q   N   T   A   P   K   S   R   N   L   F   E   R   N   P   660

aac aag gag ctg aaa ccc ggg gaa aat tca cca aga cag acc ccc atc ttt gac cct aca  2040
  N   K   E   L   K   P   G   E   N   S   P   R   Q   T   P   I   F   D   P   T   680

gtt cat tgg ctg ttc acc aca tgt ggg gcc agc ggg ccc cat ggc ccc acc cag gca cag  2100
  V   H   W   L   F   T   T   C   G   A   S   G   P   H   G   P   T   Q   A   Q   700

tgc aac aac gcc tac cag aac tcc aac ctg agc gtg gag gtg ggg agc gag ggc ccc ctg  2160
  C   N   N   A   Y   Q   N   S   N   L   S   V   E   V   G   S   E   G   P   L   720

aaa ggc atc cag atc tgg aag gtg cca gcc acc gac acc tac agc atc tcg ggc tac gga  2220
  K   G   I   Q   I   W   K   V   P   A   T   D   T   Y   S   I   S   G   Y   G   740

gct gct ggc ggg aaa ggc ggg aag aac acc atg atg cgg tcc cac ggc gtg tct gtg ctg  2280
  A   A   G   G   K   G   G   K   N   T   M   M   R   S   H   G   V   S   V   L   760

ggc atc ttc aac ctg gag aag gat gac atg ctg tac atc ctg gtt ggg cag cag gga gag  2340
  G   I   F   N   L   E   K   D   D   M   L   Y   I   L   V   G   Q   Q   G   E   780

gac gcc tgc ccc agt aca aac cag tta atc cag aaa gtc tgc att gga gag aac aat gtg  2400
  D   A   C   P   S   T   N   Q   L   I   Q   K   V   C   I   G   E   N   N   V   800

ata gaa gaa gaa atc cgt gtg aac aga agc gtg cat gag tgg gca gga ggc gga gga gga  2460
  I   E   E   E   I   R   V   N   R   S   V   H   E   W   A   G   G   G   G   G   820

ggg ggt gga gcc acc tac gta ttt aag atg aag gat gga gtg ccg gtg ccc ctg atc att  2520
  G   G   G   A   T   Y   V   F   K   M   K   D   G   V   P   V   P   L   I   I   840

gca gcc gga ggt ggt ggc agg gcc tac ggg gcc aag aca gac acg ttc cac cca gag aga  2580
  A   A   G   G   G   G   R   A   Y   G   A   K   T   D   T   F   H   P   E   R   860

ctg gag aat aac tcc tcg gtt cta ggg cta aac ggc aat tcc gga gcc gca ggt ggt gga  2640
  L   E   N   N   S   S   V   L   G   L   N   G   N   S   G   A   A   G   G   G   880

ggt ggc tgg aat gat aac act tcc ttg ctc tgg gcc gga aaa tct ttg cag gag ggt gcc  2700
  G   G   W   N   D   N   T   S   L   L   W   A   G   K   S   L   Q   E   G   A   900

acc gga gga cat tcc tgc ccc cag gcc atg aag aag tgg ggg tgg gag aca aga ggg ggt  2760
  T   G   G   H   S   C   P   Q   A   M   K   K   W   G   W   E   T   R   G   G   920

ttc gga ggg ggt gga ggg ggg tgc tcc tca ggt gga gga ggc gga gga tat ata ggc ggc  2820
  F   G   G   G   G   G   G   C   S   S   G   G   G   G   G   G   Y   I   G   G   940

aat gca gcc tca aac aat gac ccc gaa atg gat ggg gaa gat ggg gtt tcc ttc atc agt  2880
  N   A   A   S   N   N   D   P   E   M   D   G   E   D   G   V   S   F   I   S   960

cca ctg ggc atc ctg tac acc cca gct tta aaa gtg atg gaa ggc cac ggg gaa gtg aat  2940
  P   L   G   I   L   Y   T   P   A   L   K   V   M   E   G   H   G   E   V   N   980
```


IDT Exhibit 1020.3080

FIG. 4D

```
att aag cat tat cta aac tgc agt cac tgt gag gta gac gaa tgt cac atg gac cct gaa  3000
  I   K   H   Y   L   N   C   S   H   C   E   V   D   E   C   H   M   D   P   E  1000

agc cac aag gtc atc tgc ttc tgt gac cac ggg acg gtg ctg gct gag gat ggc gtc tcc  3060
  S   H   K   V   I   C   F   C   D   H   G   T   V   L   A   E   D   G   V   S  1020

tgc att gtg tca ccc acc ccg gag cca cac ctg cca ctc tcg ctg atc ctc tct gtg gtg  3120
  C   I   V   S   P   T   P   E   P   H   L   P   L   S   L   I   L   S   V   V  1040

acc tct gcc ctc gtg gcc gcc ctg gtc ctg gct ttc tcc ggc atc atg att gaa gaa cag  3180
  T   S   A   L   V   A   A   L   V   L   A   F   S   G   I   M   I   E   E   Q  1060

cag cga gga gcc atc cct ggg cag gct cca ggg tct gtg cca ggc cca ggg ctg gtg aag  3240
  Q   R   G   A   I   P   G   Q   A   P   G   S   V   P   G   P   G   L   V   K  1080

gac tca cca ctg ctg ctt cag cag atc tct gcc atg agg ctg cac atc tcc cag ctc cag  3300
  D   S   P   L   L   L   Q   Q   I   S   A   M   R   L   H   I   S   Q   L   Q  1100

cat gag aac agc atc ctc aag gga gcc cag atg aag gca tcc ttg gca tcc ctg ccc cct  3360
  H   E   N   S   I   L   K   G   A   Q   M   K   A   S   L   A   S   L   P   P  1120

ctg cat gtt gca aag cta tcc cat gag ggc cct ggc agt gag tta cca gct gga gcg ctg  3420
  L   H   V   A   K   L   S   H   E   G   P   G   S   E   L   P   A   G   A   L  1140

tat cgt aag acc agc cag ctg ctg gag aca ttg aat caa ttg agc aca cac acg cac gta  3480
  Y   R   K   T   S   Q   L   L   E   T   L   N   Q   L   S   T   H   T   H   V  1160

gta gac atc act cgc acc agc cct gct gcc aag agc ccg tcg gcc caa ctt atg gag caa  3540
  V   D   I   T   R   T   S   P   A   A   K   S   P   S   A   Q   L   M   E   Q  1180

gtg gct cag ctt aag tcc ctg agt gac acc gtc gag aag ctc aag gat gag gtc ctc aag  3600
  V   A   Q   L   K   S   L   S   D   T   V   E   K   L   K   D   E   V   L   K  1200

gag aca gta tct cag cgc cct gga gcc aca gta ccc act gac ttt gcc acc ttc cct tca  3660
  E   T   V   S   Q   R   P   G   A   T   V   P   T   D   F   A   T   F   P   S  1220

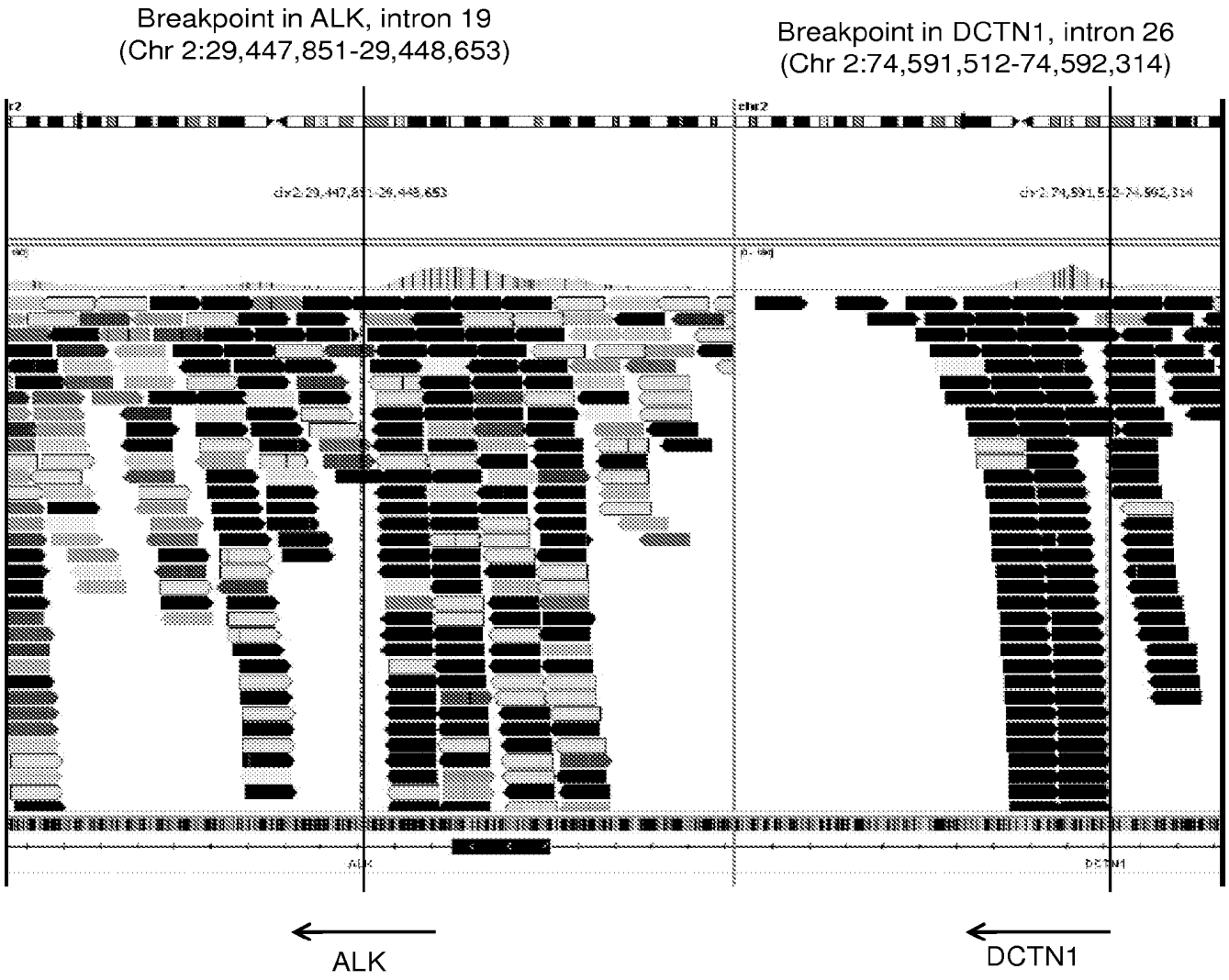
tca gcc ttc ctc agg gcc aag gag gag cag cag gat gac aca gtc tac atg ggc aaa gtg  3720
  S   A   F   L   R   A   K   E   E   Q   Q   D   D   T   V   Y   M   G   K   V  1240

acc ttc tca tgt gcg gct ggt ttt gga cag cga cac cgg ctg gtg ctg acc cag gag cag  3780
  T   F   S   C   A   A   G   F   G   Q   R   H   R   L   V   L   T   Q   E   Q  1260

ctg cac cag ctt cac agt cgc ctc atc tcc taa      3813 (SEQ ID NO: 11)
  L   H   Q   L   H   S   R   L   I   S   *      1270 (SEQ ID NO: 12)
```

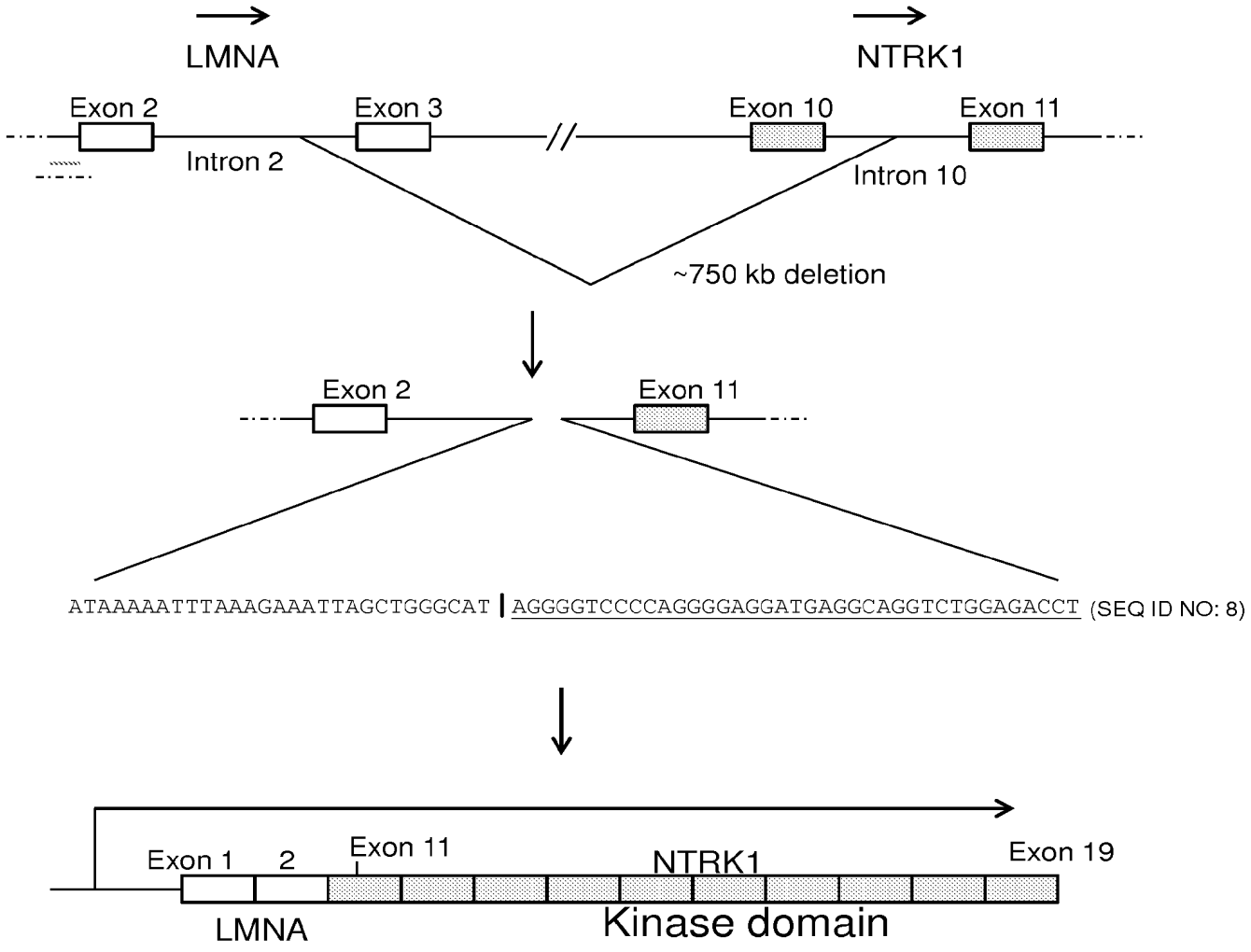

IDT Exhibit 1020.3081

FIG. 5



IDT Exhibit 1020.3082

FIG. 6



IDT Exhibit 1020.3083

FIG. 7A

```
atg gag acc ccg tcc cag cgg cgc gcc acc cgc agc ggg gcg cag gcc agc tcc act ccg  60
  M   E   T   P   S   Q   R   R   A   T   R   S   G   A   Q   A   S   S   T   P  20

ctg tcg ccc acc cgc atc acc cgg ctg cag gag aag gag gac ctg cag gag ctc aat gat 120
  L   S   P   T   R   I   T   R   L   Q   E   K   E   D   L   Q   E   L   N   D  40

cgc ttg gcg gtc tac atc gac cgt gtg cgc tcg ctg gaa acg gag aac gca ggg ctg cgc 180
  R   L   A   V   Y   I   D   R   V   R   S   L   E   T   E   N   A   G   L   R  60

ctt cgc atc acc gag tct gaa gag gtg gtc agc cgc gag gtg tcc ggc atc aag gcc gcc 240
  L   R   I   T   E   S   E   E   V   V   S   R   E   V   S   G   I   K   A   A  80

tac gag gcc gag ctc ggg gat gcc cgc aag acc ctt gac tca gta gcc aag gag cgc gcc 300
  Y   E   A   E   L   G   D   A   R   K   T   L   D   S   V   A   K   E   R   A 100

cgc ctg cag ctg gag ctg agc aaa gtg cgt gag gag ttt aag gag ctg aaa gcg cgc aat 360
  R   L   Q   L   E   L   S   K   V   R   E   E   F   K   E   L   K   A   R   N 120

acc aag aag gag ggt gac ctg ata gct gct cag gct cgg ctg aag gac ctg gag gct ctg 420
  T   K   K   E   G   D   L   I   A   A   Q   A   R   L   K   D   L   E   A   L 140

ctg aac tcc aag gag gcc gca ctg agc act gct ctc agt gag aag cgc acg ctg gag ggc 480
  L   N   S   K   E   A   A   L   S   T   A   L   S   E   K   R   T   L   E   G 160

gag ctg cat gat ctg cgg ggc cag gtg gcc aag gtc tcc ttc tcg ccg gtg gac act aac 540
  E   L   H   D   L   R   G   Q   V   A   K   V   S   F   S   P   V   D   T   N 180

agc aca tct gga gac ccg gtg gag aag aag gac gaa aca cct ttt ggg gtc tcg gtg gct 600
  S   T   S   G   D   P   V   E   K   K   D   E   T   P   F   G   V   S   V   A 200

gtg ggc ctg gcc gtc ttt gcc tgc ctc ttc ctt tct acg ctg ctc ctt gtg ctc aac aaa  660
  V   G   L   A   V   F   A   C   L   F   L   S   T   L   L   L   V   L   N   K 220

tgt gga cgg aga aac aag ttt ggg atc aac cgc ccg gct gtg ctg gct cca gag gat ggg  720
  C   G   R   R   N   K   F   G   I   N   R   P   A   V   L   A   P   E   D   G 240

ctg gcc atg tcc ctg cat ttc atg aca ttg ggt ggc agc tcc ctg tcc ccc acc gag ggc  780
  L   A   M   S   L   H   F   M   T   L   G   G   S   S   L   S   P   T   E   G 260

aaa ggc tct ggg ctc caa ggc cac atc atc gag aac cca caa tac ttc agt gat gcc tgt  840
  K   G   S   G   L   Q   G   H   I   I   E   N   P   Q   Y   F   S   D   A   C 280

gtt cac cac atc aag cgc cgg gac atc gtg ctc aag tgg gag ctg ggg gag ggc gcc ttt  900
  V   H   H   I   K   R   R   D   I   V   L   K   W   E   L   G   E   G   A   F 300
```


IDT Exhibit 1020.3084

FIG. 7B

```
ggg aag gtc ttc ctt gct gag tgc cac aac ctc ctg cct gag cag gac aag atg ctg gtg  960
  G   K   V   F   L   A   E   C   H   N   L   L   P   E   Q   D   K   M   L   V  320

gct gtc aag gca ctg aag gag gcg tcc gag agt gct cgg cag gac ttc caa cgt gag gct 1020
  A   V   K   A   L   K   E   A   S   E   S   A   R   Q   D   F   Q   R   E   A  340

gag ctg ctc acc atg ctg cag cac cag cac atc gtg cgc ttc ttc ggc gtc tgc acc gag 1080
  E   L   L   T   M   L   Q   H   Q   H   I   V   R   F   F   G   V   C   T   E  360

ggc cgc ccc ctg ctc atg gtc ttc gag tat atg cgg cac ggg gac ctc aac cgc ttc ctc 1140
  G   R   P   L   L   M   V   F   E   Y   M   R   H   G   D   L   N   R   F   L  380

cga tcc cat gga ccc gat gcc aag ctg ctg gct ggt ggg gag gat gtg gct cca ggc ccc 1200
  R   S   H   G   P   D   A   K   L   L   A   G   G   E   D   V   A   P   G   P  400

ctg ggt ctg ggg cag ctg ctg gcc gtg gct agc cag gtc gct gcg ggg atg gtg tac ctg 1260
  L   G   L   G   Q   L   L   A   V   A   S   Q   V   A   A   G   M   V   Y   L  420

gcg ggt ctg cat ttt gtg cac cgg gac ctg gcc aca cgc aac tgt cta gtg ggc cag gga 1320
  A   G   L   H   F   V   H   R   D   L   A   T   R   N   C   L   V   G   Q   G  440

ctg gtg gtc aag att ggt gat ttt ggc atg agc agg gat atc tac agc acc gac tat tac 1380
  L   V   V   K   I   G   D   F   G   M   S   R   D   I   Y   S   T   D   Y   Y  460

cgt gtg gga ggc cgc acc atg ctg ccc att cgc tgg atg ccg ccc gag agc atc ctg tac 1440
  R   V   G   G   R   T   M   L   P   I   R   W   M   P   P   E   S   I   L   Y  480

cgt aag ttc acc acc gag agc gac gtg tgg agc ttc ggc gtg gtg ctc tgg gag atc ttc 1500
  R   K   F   T   T   E   S   D   V   W   S   F   G   V   V   L   W   E   I   F  500

acc tac ggc aag cag ccc tgg tac cag ctc tcc aac acg gag gca atc gac tgc atc acg 1560
  T   Y   G   K   Q   P   W   Y   Q   L   S   N   T   E   A   I   D   C   I   T  520

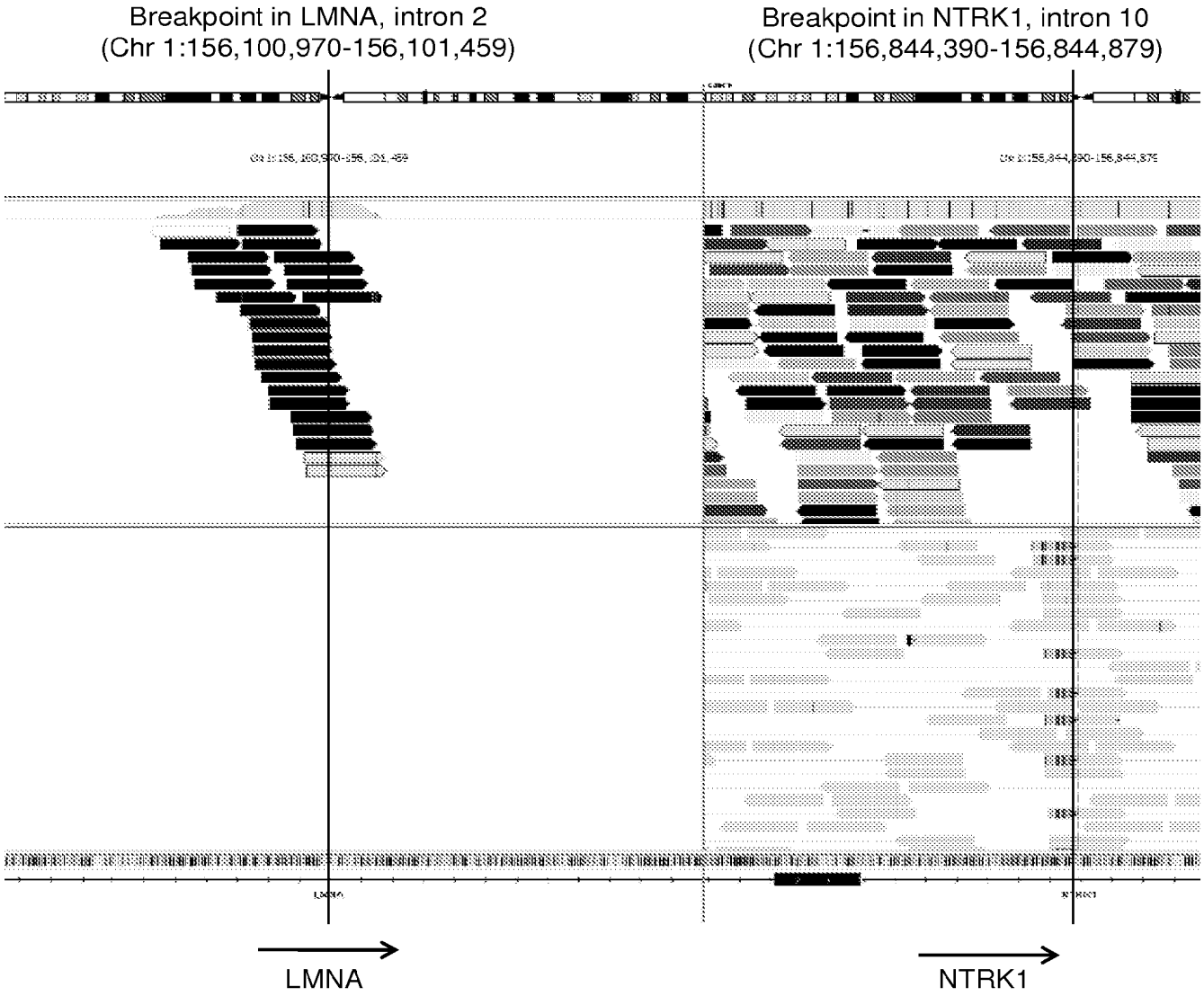
cag gga cgt gag ttg gag cgg cca cgt gcc tgc cca cca gag gtc tac gcc atc atg cgg 1620
  Q   G   R   E   L   E   R   P   R   A   C   P   P   E   V   Y   A   I   M   R  540

ggc tgc tgg cag cgg gag ccc cag caa cgc cac agc atc aag gat gtg cac gcc cgg ctg 1680
  G   C   W   Q   R   E   P   Q   Q   R   H   S   I   K   D   V   H   A   R   L  560

caa gcc ctg gcc cag gca cct cct gtc tac ctg gat gtc ctg ggc tag (SEQ ID NO: 9)  1740
  Q   A   L   A   Q   A   P   P   V   Y   L   D   V   L   G   * (SEQ ID NO: 10)  580
```


IDT Exhibit 1020.3085

FIG. 8



IDT Exhibit 1020.3086

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 12/61211

A. CLASSIFICATION OF SUBJECT MATTER
IPC(8) - C12P 21/02; C07H 21/04 (2013.01)
USPC - 435/69.7, 536/23.4, 536/24.3
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
USPC: 435/69.7, 536/23.4, 536/24.3

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
USPC: 435/69.7, 69.1; 536/23.4, 23.1, 24.3 (text search)

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
Electronic data bases: PatBase; Google Scholar; GenCore sequence search (NT, AA)
Search terms: anaplastic lymphoma kinase (ALK) ,ALK fusion, dynactin 1 (DCTN1),DCTN1-ALKm neurotrophic tyrosine kinase receptor type 1 (NTRK1), NTRK1 fusion, lamin A/C (LMNA, LMNC, LMN1), LMNA-NTRK1, fusion junction, breakpoint, antibody, ALK inhibitor, N

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X --- A	CAMIDGE et al. Optimizing the detection of lung cancer patients harboring anaplastic lymphoma kinase (ALK) gene rearrangements potentially suitable for ALK inhibitor treatment. Clin Cancer Res 14 November 2010 Vol 16 No 22 Pages 5581-5590. Especially pg 5586 col 2 para 2-3.	36 ------ 37-40
X	EP 0698096 B1 (GRECO et al.) 26 March 1997 (26.03.1997). Especially col 5 ln 45 continued to col 6 ln 35, col 14 ln 46 continued to col 15 ln 15.	63-70
A	KELLEHER et al. The emerging pathogenic and therapeutic importance of the anaplastic lymphoma kinase gene. Eur J Cancer September 2010 Vol 46 No 13 Pages 2357-2368. Especially pg 2365 table 6.	1-3, 6-35, 41-62, 71-104
A	US 2002/0197679 A1 (TANG et al.) 26 December 2002 (26.12.2002). Especially SEQ ID NO: 18	1-4, 6-15, 71-97
A	US 2009/0156475 A1 (RIKOVA et al.) 18 June 2009 (18.06.2009). Especially SEQ ID NOs: 20, 21	1-4, 6-19, 37-62, 71-97
Y, P	WANG et al. Fusion of dynactin 1 to anaplastic lymphoma kinase in inflammatory myofibroblastic tumor. Hum Pathol ePub 1 June 2012 Vol 43 No 11 Pages 2047-2052. Especially abstract.	1-3, 6-19, 41-62, 71-104
A	GRECO et al. Rearrangements of NTRK1 gene in papillary thyroid carcinoma. Mol Cell Endocrinol 28 May 2010 Vol 321 No 1 Pages 44-49. Especially pg 46 col 2 para 3.	1-3, 6-35, 41-62, 71-104

☒ Further documents are listed in the continuation of Box C. ☐

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

Date of the actual completion of the international search	Date of mailing of the international search report
15 January 2013 (15.01.2013)	05 FEB 2013
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-3201	Authorized officer: Lee W. Young PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

Form PCT/ISA/210 (second sheet) (July 2009)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 12/61211

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2005/0287541 A1 (NAKAGAWARA et al.) 29 December 2005 (29.12.2005). Especially SEQ ID NO: 4	1-3, 5-15, 71-97
A	US 2008/0131375 A1 (GORDON et al.) 5 June 2008 (05.06.2008). Especially SEQ ID NO: 27	1-3, 5-15, 71-97
A	US 2007/0010657 A1 (KLOKE et al.) 11 June 2007 (11.06.2007). Especially SEQ ID NO: 68	16-19, 37-62, 71-97
A	US 2011/0150893 A1 (CHO et al.) 23 June 2011 (23.06.2011). Especially SEQ ID NO: 5	16-19, 37-62, 71-97
A	US 2008/0226664 A1 (OLD et al.) 18 September 2008 (18.09.2008) Especially SEQ ID NO: 25	16-19, 37-62, 71-97
A	TAKEUCHI et al. Multiplex reverse transcription-PCR screening for EML4-ALK fusion transcripts. Clin Cancer Res 15 October 2008 Vol 14 No 20 Pages 6618-6624. Especially pg 6619 col 1 para .	20-35
A	US 2008/0051462 A1 (FRITZ et al.) 28 February 2008 (28.02.2008). Especially SEQ ID NO: 136	20-35
A	US 2007/0059710 A1 (LUKE et al.) 15 March 2007 (15.03.2007). Especially SEQ ID NO: 51978	20-35

IDT Exhibit 1020.3088

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
5 May 2011 (05.05.2011)

PCT

(10) International Publication Number
WO 2011/053987 A1

(51) International Patent Classification:
C12Q 1/68 (2006.01) *C12P 19/34* (2006.01)

(21) International Application Number:
PCT/US2010/055137

(22) International Filing Date:
2 November 2010 (02.11.2010)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/257,241 2 November 2009 (02.11.2009) US

(71) Applicant *(for all designated States except US)*: **NUGEN TECHNOLOGIES, INC.** [US/US]; 821 Industrial Road, San Carlos, CA 94070 (US).

(72) Inventor; and
(75) Inventor/Applicant *(for US only)*: **KURN, Nurith** [IL/US]; 2876 Ramona Street, Palo Alto, CA 94306 (US).

(74) Agents: **SEIDEL, Jeffrey, J.** et al.; Wilson Sonsini Goodrich & Rosati, 650 Page Mill Road, Palo Alto, CA 94304-1050 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

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

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

WO 2011/053987 A1

(54) Title: COMPOSITIONS AND METHODS FOR TARGETED NUCLEIC ACID SEQUENCE SELECTION AND AMPLIFICATION

(57) Abstract: The present invention provides novel methods, compositions, and kits for the production of amplification-ready, sequence-specific, target region-specific, and strand-specific regions of interest directly from samples containing complex DNA. The methods, composition, and kits provided herein are useful for selective target generation, genome partitioning, or user-selected enrichment of desired regions of interest. The invention described herein will enable multiplexing for genome -wide analysis with increased efficiency and is amenable to automation.

IDT Exhibit 1020.3089

COMPOSITIONS AND METHODS FOR TARGETED NUCLEIC ACID SEQUENCE SELECTION AND AMPLIFICATION

CROSS REFERENCE

[0001] This application claims the benefit of U.S. Provisional Application No. 61/257,241 filed November 2, 2009, which is herein incorporated by reference in its entirety.

BACKGROUND OF THE INVENTION

[0002] The development of massively parallel sequencing technologies and platforms enable whole genome sequencing and targeted re-sequencing of defined organisms or mixtures of organisms. Although the cost of generating the newer sequencing information is decreasing and throughput of these technologies and platforms is increasing, it is recognized that focused target enrichment from high complexity genomic DNA will improve sequencing at high depth, enabling the sequencing or targeted re-sequencing of a larger number of samples as required for various fundamental biological studies of normal and disease development and pathogenesis.

[0003] Various methods for selective enrichment of a multiplicity of targets from genomic DNA, commonly referred to as "genome partitioning," were developed in recent years. Some of these methods are based on selective hybridization to oligonucleotides designed to hybridize to the user-selected genomic regions. The hybridization can be to oligonucleotides immobilized on high or low density microarrays or solution phase hybridization to oligonucleotides modified with a ligand which can be subsequently immobilized to a solid surface, such as a bead. Other methods employ sequence-specific amplification (e.g., PCR) to amplify specific genomic regions in a droplet, allowing clonal amplification of defined regions for downstream sequencing.

[0004] A major drawback of the recently developed methods is the requirement to sequester regions of interest for hybridization to the specific oligonucleotides or to individual droplets. A related drawback to hybridization of surface-immobilized oligonucleotides is the low efficiency of the reaction, requiring large quantities of input DNA. These drawbacks lead to the need to fragment the original genomic DNA and in some instances further amplify the genomic DNA or library of fragments prior to selective enrichment by hybridization or sequestration to individual droplets

[0005] Therefore, there is a need for improved methods of selectively enriching target regions of genomic DNA for downstream next-generation applications such as massively parallel sequencing. The invention described herein fulfills this need.

SUMMARY OF THE INVENTION

[0006] The present invention provides novel methods, compositions, and kits for the production of amplification-ready DNA of target sequence regions of interest partitioned from complex DNA, such as double stranded DNA, genomic DNA, or mixed DNA from more than one organism. The invention

IDT Exhibit 1020.3090

provides methods, compositions and kits for directly generating the DNA of target sequence regions of interest that can be used for selective target generation, genome partitioning, or user-selected enrichment of desired regions of interest. The invention described herein will enable multiplexing for genome-wide analysis with increased efficiency and is amenable to automation. The invention described herein will allow for selective enrichment of strand-specific sequence regions of interest, derived from a selected strand of highly complex double-stranded DNAs. The methods, compositions and kits further enable amplification of selective regions of interest, and specifically selective strands of interest by a variety of amplification methods, for example, the single primer isothermal linear amplification method (SPIA).

[0007] In one aspect, a method for selectively partitioning a plurality of target regions of interest from complex DNA is provided comprising a) hybridizing one or more oligonucleotides to the target regions of interest to form a plurality of target-oligonucleotide complexes; b) tethering a nucleic acid-modifying enzyme to a target-oligonucleotide complex via the oligonucleotide; and c) cleaving the target region of interest in the complex, thereby releasing the target region of interest from the enzyme.

[0008] In one embodiment, the one or more oligonucleotides hybridized to the target regions of interest are extended along the target region of interest by DNA polymerase prior to step b. In another embodiment, the complex DNA comprises double-stranded DNA. In another embodiment, the complex DNA comprises genomic DNA. In another embodiment, the genomic DNA comprises a mixture of genomic DNA from more than one organism. In another embodiment, the complex DNA comprises cDNA. In another embodiment, the cDNA is generated from a mixture of DNAs from more than one organism. In another embodiment, the selective partitioning is strand-specific. In another embodiment, the enzyme is a DNA duplex-specific endonuclease. In another embodiment, the enzyme is a restriction enzyme. In another embodiment, the method further comprises denaturing the complex DNA prior to hybridization of the oligonucleotides. In another embodiment, the method further comprises the formation of partial triplexes. In another embodiment, the method further comprises ligating adapters to the target regions of interest once released from the enzyme. In another embodiment, the adapter is selected from a group consisting of a double stranded adapter with an overhang at one end, a double stranded adapter with a 3' single stranded overhang, an adapter that comprises a RNA-DNA heteroduplex, an adapter that comprises a chimeric DNA-RNA oligonucleotide and a stem-loop adapter. In another embodiment, the selective partitioning is carried out directly on the complex DNA. In another embodiment, said method does not involve amplifying the complex DNA prior to selective partitioning. In another embodiment, the method further comprises amplifying the partitioned target regions of interest thereby enriching for the target regions of interest. In another embodiment, the amplifying comprises single primer isothermal amplification. In another embodiment, the method further comprises sequencing of the amplified products. In another embodiment, the sequencing is performed using a massively parallel sequencing method. In another embodiment, the enzyme is synthetic, semisynthetic, or recombinant. In another embodiment, the ligand is selected from the group consisting of a small molecule, an antigen, an antibody, hybrid antibody or antibody fragment, an siRNA, an antisense RNA,

IDT Exhibit 1020.3091

an aptamer, a peptide and a first member of a specific binding pair. In another embodiment, the ligand-binding component is selected from the group consisting of a small molecule, an antigen, an antibody, hybrid antibody or antibody fragment, an siRNA, an antisense RNA, an aptamer, a protein therapeutic, or a peptide, a ligand binding protein, and a second member of specific binding pair. In another embodiment, at least 100 different regions of interest are selectively partitioned. In another embodiment, at least 100 oligonucleotides are used for selectively partitioning the target region of interest from complex DNA. In another embodiment, each oligonucleotide is coupled to the same ligand. In another embodiment, the ligand is biotin and the ligand-binding component is avidin or streptavidin.

[0009] In another aspect, a method of preparing amplification-ready selectively targeted regions of interest from double-stranded DNA is provided wherein the method comprises: a) denaturing the double-stranded DNA thereby generating single-stranded DNA; b) hybridizing to the single-stranded target regions of interest one or more oligonucleotides to form partial duplexes, each oligonucleotide being coupled to a ligand; c) contacting the ligands with a ligand-binding component coupled to a nucleic acid-modifying enzyme, wherein the enzyme cleaves the target regions of interest thereby obtaining products comprising the target regions of interest free of the enzyme; and d) ligating adapters to the products, whereby the targeted regions of interest are amplification-ready. In one embodiment, said method does not involve amplifying the double-stranded DNA until after step d.

[0010] In another aspect, a method of preparing amplification-ready selectively targeted regions of interest from double-stranded DNA is provided wherein the method comprises: a) hybridizing to the double-stranded target regions of interest one or more oligonucleotides to form partial triplexes, each oligonucleotide being coupled to a ligand; b) contacting the ligands with a ligand-binding component coupled to a nucleic acid-modifying enzyme, wherein the enzyme cleaves the target regions of interest thereby obtaining products comprising the target regions of interest free of the enzyme; and c) ligating adapters to the products, whereby the targeted regions of interest are amplification-ready. In one embodiment, said method does not involve amplifying the double-stranded DNA until after step c.

[0011] In one embodiment, the double-stranded DNA comprises genomic DNA. In another embodiment, the double-stranded DNA comprises cDNA. In another embodiment, the cDNA is generated from a mixture of DNAs from more than one organism. In another embodiment, the genomic DNA comprises a mixture of genomic DNA from more than one organism. In another embodiment, the method is strand-specific. In another embodiment, the enzyme is a DNA-duplex specific endonuclease. In another embodiment, the enzyme is a restriction enzyme. In another embodiment, the adapter is selected from a group consisting of a double stranded adapter with an overhang at one end, a double stranded adapter with a 3' single stranded overhang, an adapter that comprises a RNA-DNA heteroduplex, an adapter that comprises a chimeric DNA-RNA oligonucleotide and a stem-loop adapter. In another embodiment, the method is carried out directly on the double-stranded DNA. In another embodiment, the method further comprises amplifying the target regions of interest, thereby enriching for the target sequence regions of interest. In another embodiment, amplifying comprises single primer

IDT Exhibit 1020.3092

isothermal amplification. In another embodiment, the method further comprises sequencing of the amplified products. In another embodiment, the sequencing is performed using a massively parallel sequencing method. In another method, the enzyme is synthetic, semisynthetic or recombinant. In another embodiment, the ligand is selected from the group consisting of a small molecule, an antigen, an antibody, hybrid antibody or antibody fragment, an siRNA, an antisense RNA, an aptamer, a peptide and a first member of a specific binding pair. In another embodiment, the ligand-binding component is selected from the group consisting of a small molecule, an antigen, an antibody, hybrid antibody or antibody fragment, an siRNA, an antisense RNA, an aptamer, a protein therapeutic, or a peptide, a ligand binding protein, and a second member of specific binding pair. In another embodiment, the ligand is biotin and the ligand-binding component is avidin or streptavidin. In another embodiment, at least 100 different regions of interest are prepared for amplification. In another embodiment, at least 100 oligonucleotides are used for preparing the target regions of interest for amplification. In another embodiment, each oligonucleotide is coupled to the same ligand.

[0012] In another aspect, a DNA complex is provided comprising: a) DNA comprising at least one DNA strand; b) an oligonucleotide hybridized to the DNA wherein the oligonucleotide is coupled to a ligand; and c) a ligand-binding component coupled to a nucleic acid-modifying enzyme, wherein the ligand and the ligand-biding component are further coupled to each other. In one embodiment, the DNA comprises genomic DNA. In another embodiment, the DNA is double stranded and the complex is a partial triplex. In another embodiment, the DNA is single stranded and the complex is a partial duplex. In another embodiment, the nucleic acid-modifying enzyme is synthetic, semisynthetic, or recombinant. In another embodiment, the ligand is selected from the group consisting of a small molecule, an antigen, an antibody, hybrid antibody or antibody fragment, an siRNA, an antisense RNA, an aptamer, a peptide and a first member of a specific binding pair. In another embodiment, the ligand binding component is selected from the group consisting of a small molecule, an antigen, an antibody, hybrid antibody or antibody fragment, an siRNA, an antisense RNA, an aptamer, a protein therapeutic, or a peptide, a ligand binding protein, and a second member of specific binding pair. In another embodiment, the ligand is biotin and the ligand binding component is avidin or streptavidin. In another embodiment, the oligonucleotide is further extended with a DNA polymerase.

[0013] In another aspect, a kit is provided comprising: a) one or more oligonucleotides, each coupled to a ligand; and b) a ligand-binding component that selectively binds to said ligand, wherein said ligand-binding component is coupled to a nucleic acid-modifying enzyme. In one embodiment, the kit further comprises reagents for amplification. In another embodiment, the kit further comprises an adapter for amplification. In another embodiment, the adapter is selected from a group consisting of a double stranded adapter with an overhang at one end, a double stranded adapter with a 3' single stranded overhang, an adapter that comprises a RNA-DNA heteroduplex, an adapter that comprises a chimeric DNA-RNA oligonucleotide and a stem-loop adapter. In another embodiment, the kit further comprises reagents for sequencing. In another embodiment, the sequencing reagents comprise reagents for a

IDT Exhibit 1020.3093

massively parallel sequencing method. In another embodiment, the reagents are reagents for performing single primer isothermal amplification. In another embodiment, the kit further comprises a DNA polymerase. In another embodiment, the kit further comprises a plurality of oligonucleotides. In another embodiment, the plurality of oligonucleotides are coupled to different ligands. In another embodiment, the ligand is selected from the group consisting of a small molecule, an antigen, an antibody, hybrid antibody or antibody fragment, an siRNA, an antisense RNA, an aptamer, a peptide and a first member of a specific binding pair. In another embodiment, the ligand binding component is selected from the group consisting of a small molecule, an antigen, an antibody, hybrid antibody or antibody fragment, an siRNA, an antisense RNA, an aptamer, a protein therapeutic, or a peptide, a ligand binding protein, and a second member of specific binding pair. In another embodiment, the ligand is biotin and the ligand-binding component is avidin or streptavidin. In another embodiment, the nucleic acid modifying enzyme is synthetic, semisynthetic, or recombinant. In another embodiment, the nucleic acid modifying enzyme is a DNA duplex-specific endonuclease. In another embodiment, the nucleic acid modifying enzyme is a restriction endonuclease.

[0014] In another aspect, a plurality of amplification-ready partial DNA duplexes generated from genomic DNA is provided, each duplex comprising at least one DNA strand comprising the sequence of a target region of interest wherein the duplex is further ligated to an adaptor for single primer isothermal amplification.

INCORPORATION BY REFERENCE

[0015] All publications, patents, and patent applications mentioned in this specification are herein incorporated by reference to the same extent as if each individual publication, patent, or patent application was specifically and individually indicated to be incorporated by reference.

BRIEF DESCRIPTION OF THE DRAWINGS

[0016] The novel features of the invention are set forth with particularity in the appended claims. A better understanding of the features and advantages of the present invention will be obtained by reference to the following detailed description that sets forth illustrative embodiments, in which the principles of the invention are utilized, and the accompanying drawings of which:

[0017] **Figure 1** depicts strand-specific selective hybridization.

[0018] **Figure 2** depicts selective tethering of a NA-modifying enzyme at target sequence regions of interest.

[0019] **Figure 3** depicts selective cleavage at target sequence regions of interest.

[0020] **Figure 4** depicts adapter ligation and selective amplification of target sequence regions of interest.

[0021] **Figure 5** depicts adapter ligation and selective amplification of target sequence regions of interest.

IDT Exhibit 1020.3094

[0022] **Figure 6** depicts adapter ligation and selective amplification of target sequence regions of interest.

[0023] **Figure 7** depicts adapter ligation and selective amplification of target sequence regions of interest.

[0024] **Figure 8** depicts selective triplex formation and tethering of NA-modifying enzyme at target sequence regions of interest.

[0025] **Figure 9** depicts selective cleavage of dsDNA target sequence regions of interest.

DETAILED DESCRIPTION OF THE INVENTION

General Features

[0026] Selective enrichment of defined sequences from complex DNA according to the methods of the invention can be carried out in solution and can be based on the principle of tethering nucleic acid (NA) modifying enzymes to the specific regions of interest and subsequently generating amplification-ready products of the defined regions. The selection of regions of interest can be defined by the study objectives or any other user needs. Thus, for example, the regions of interest can represent all known coding regions, the entire exome, selected regions of coding genomic regions representing selected pathways, selected genomic regions known to comprise genomic variation related to altered phenotype, entire or selected regions of a specific chromosome, and the like.

[0027] The methods of the invention can be directed at selective, highly multiplex enrichment of genomic regions of interest from high complexity nucleic acid samples such as whole genomes or mixtures thereof. The selective enrichment can be carried out in solution comprising the high complexity nucleic acid samples and can employ sequence-specific hybridization, tethering of nucleic acid modifying (NA-modifying) enzyme (or enzymes) at the desired sequence of interest to enable generation of amplification-ready products of the defined sequence regions of interest. The methods can employ a multiplicity of oligonucleotides designed to hybridize to a multiplicity of target sequence regions of interest. The methods described herein can be used to obtain a plurality of different regions of interest from a complex nucleic acid sample containing complex DNA wherein the method can include incubating the nucleic acid sample with a plurality of different oligonucleotides, wherein the oligonucleotides selectively hybridize to each of the plurality of different regions of interest, and wherein each of the plurality of oligonucleotides can be indirectly coupled to a NA-modifying enzyme such as a DNA duplex-specific nuclease. The indirect coupling can be via binding between specific binding pairs such as a ligand and a ligand binding component.

[0028] Tethering the nucleic acid (NA)-modifying enzyme to the region(s) of interest can be achieved by hybridization of one or more oligonucleotides to the specific region(s) of interest. The target-specific oligonucleotide(s) can be modified with at least one ligand (referred to throughout interchangeably as a label or member of a specific-binding pair) which can serve as an anchor to bind a ligand binding component (for example, ligand binding protein, antibody, aptamer, or a second member of specific binding pair) conjugated to the NA-modifying enzyme.

IDT Exhibit 1020.3095

[0029] The solution-based processes of the current invention can be more efficient than hybridization to immobilized oligonucleotides as commonly used in the recently developed genome partitioning methods. In addition to the efficient solution phase hybridization, the methods of the invention can employ various nucleic acid modifying enzymes as well as polymerases, which can be suitable for manipulation and amplification of very low input nucleic acid samples.

[0030] Insofar as the methods of the invention do not require fragmentation and sequestration of the input nucleic acid sample, the methods can be suitable to enrich sequence regions of interest from very low input nucleic acid samples. The invention described herein can be used to generate amplification-ready DNA of a target region of interest directly from a sample containing complex DNA without any prior manipulation of the sample such as fragmentation, amplification or sequestration.

[0031] Amplification-ready DNA of a target sequence region of interest can be generated directly from a sample containing complex DNA in a single reaction tube. In another embodiment, the methods described herein can be used to generate amplification-ready DNA of a target region of interest directly from a sample containing complex DNA in less than 20 minutes, 30 minutes, 60 minutes, 90 minutes, 120 minutes, 180 minutes, or 240 minutes, or 300 minutes. In another embodiment, the methods described herein can be used to generate amplification-ready DNA in about 20-300 minutes, 30-240 minutes, 60-180 minutes, 90 to 120 minutes, 20-240 minutes, 20-180 minutes, 20-120 minutes, 20-90 minutes, 20-60 minutes, or 20-30 minutes.

[0032] The methods described herein can be used to generate a collection of at least 25, 50, 75, 100, 500, 1000, 2500, 5000, 10,000, 25,000, 50,000, 100,000, 500,000, or 1,000,000 target sequence regions of interest directly from a sample of complex DNA using a plurality of oligonucleotides. The methods described herein can be used to generate a collection of about 1000 to 1,000,000, 2500 to 1,000,000, 5000 to 1,000,000, 10,000 to 1,000,000, 25,000 to 1,000,000, 50,000 to 1,000,000, 100,000 to 1,000,000, or 500,000 to 1,000,000 target sequence regions of interest directly from a sample of complex DNA using a plurality of oligonucleotides.

Input Nucleic Acid Sample

[0033] The input is a nucleic acid. The input nucleic acid can be DNA, or complex DNA, for example genomic DNA. The input DNA may also be cDNA. The cDNA can be generated from RNA, e.g., mRNA. The input DNA can be of a specific species, for example, human, rat, mouse, other animals, specific plants, bacteria, algae, viruses, and the like. The input complex also can be from a mixture of genomes of different species such as host-pathogen, bacterial populations and the like. The input DNA can be cDNA made from a mixture of genomes of different species. Alternatively, the input nucleic acid can be from a synthetic source. The input DNA can be mitochondrial DNA. The input DNA can be cell-free DNA. The cell-free DNA can be obtained from, e.g., a serum or plasma sample. The input DNA can comprise one or more chromosomes. For example, if the input DNA is from a human, the DNA can comprise one or more of chromosome 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21,

IDT Exhibit 1020.3096

22, X, or Y. The DNA can be from a linear or circular genome. The DNA can be plasmid DNA, cosmid DNA, bacterial artificial chromosome (BAC), or yeast artificial chromosome (YAC). The input DNA can be from more than one individual or organism. The input DNA can be double stranded or single stranded. The input DNA can be part of chromatin. The input DNA can be associated with histones.

[0034] In so far as the methods of the invention do not require fragmentation and sequestration of the input nucleic acid sample, the methods can be suitable for enrichment of desired sequence regions of interest from very low input nucleic acid samples. In certain embodiments, input as low as a picogram or a nanogram of nucleic acid are sufficient. The methods of the current invention can be suitable for enrichment of selected sequence regions of interest from complex genomic nucleic acid samples. The amount of input nucleic acid can be, e.g., less than 1 pg, 10 pg, 100 pg, 1 ng, 10 ng, 100 ng, 1 µg, 10 µg, or 100 µg. The amount of input nucleic acid can be, e.g., about 1 pg to 1 ng, 1 pg to 1 µg, or 1 ng to 1 µg.

[0035] Insofar as no prior denaturation of the input nucleic acid sample is required, the method of the invention can be suitable for *in situ* selective enrichment and amplification of sequence regions of interest.

[0036] In some embodiments, the oligonucleotides targeting the selected sequence regions of interest are designed to hybridize to ssDNA targets. In the case where the input nucleic acid sample comprises genomic DNA or other dsDNA, the input nucleic acid sample can be first denatured to render the target single stranded and enable hybridization of the oligonucleotides to the desired sequence regions of interest. In these embodiments, the methods and compositions described herein can allow for region-specific enrichment and amplification of sequence regions of interest.

[0037] In other embodiments, the oligonucleotides targeting the selected sequence regions of interest are designed to hybridize to dsDNA target, without denaturation of the dsDNA. In these embodiments, the oligonucleotides targeting the selected sequence regions of interest are designed to form a triple helix (triplex) at the selected sequence regions of interest. The hybridization of the oligonucleotides to the dsDNA sequence regions of interest can be carried out without prior denaturation of the double stranded nucleic acid sample. In such embodiments, the methods and compositions described herein can allow for region-specific enrichment as well as strand-specific enrichment and amplification of sequence regions of interest. This method can be useful for generation of copies of strand specific sequence regions of interest from complex nucleic acid without the need to denature the dsDNA input DNA, thus enabling enrichment and analysis of multiplicity of sequence regions of interest in the native complex nucleic acid sample. The method can find use for studies and analyses carried out *in situ*, enable studies and analysis of complex genomic DNA in single cells or collection of very small well defined cell population, as well as permit the analysis of complex genomic DNA without disruption of chromatin structures.

Oligonucleotides

[0038] In one embodiment of the invention, the oligonucleotides of the invention hybridize to single stranded targets to form duplexes, as is exemplified in Figure 1. This hybridization is then followed by

IDT Exhibit 1020.3097

tethering a NA-modifying enzyme at the sequence region of interest and selective cleavage of dsDNA sequence region of interest, as is depicted in Figures 2-3.

[0039] In another embodiment of the invention the oligonucleotides of the invention hybridize to double stranded targets to form a triplex nucleic acid, as is exemplified in Figure 8. In this embodiment, this hybridization is then followed by tethering of a NA-modifying enzyme at the sequence region of interest and selective cleavage of a dsDNA sequence region of interest, as is depicted in Figures 8-9.

[0040] In some embodiment of the invention, the hybridization of the oligonucleotide or oligonucleotides to the target dsDNA can be carried out without prior denaturation by recombinase mediated hybridization, for example RecA mediated specific hybridization, for example as described in US patents 5965361 and 7468244 and references therein. The best known RecA-like recombinase is the RecA protein from *E. coli*, which is available commercially (Roche Applied Science, Indianapolis, Ind. USA).

[0041] RecA orthologues have also been identified in a wide variety of prokaryotic genera, including *Bacillus* (*B. halodurans*, GenBank accession no. NP_243249), *Streptomyces* (*S. agalactiae*, GenBank accession no. NP_689079; *S. pyogenes* MGAS315, GenBank accession no. NP_665604); *Staphylococcus* (*S. aureus* subsp. *Aureus* MW2, GenBank accession no. NP_645985); *Brucella* (*B. melitensis*, GenBank accession no. NP_539704); *Helicobacter* (*H. pylori*, GenBank accession no. NP_206952); *Corynebacterium* (*C. glutamicum*, GenBank accession no. NP_601162); *Bordetella* (*B. hinzii*, GenBank accession no. AAM92267); *Bacteroides* (*B. fragilis*, GenBank accession no. AAK58827); *Haemophilus* (*H. influenzae*, GenBank accession no. AAM91954); *archaebacteria* (Reich et al., *Extremophiles* 5(4):265-75 (2001); and others.

[0042] In eukaryotes, RecA orthologues are typically members of the Rad51 family (Shibata et al., *Proc. Natl. Acad. Sci. USA* 98:8425-8432 (2001)), members of which have been identified in a wide variety of eukaryotic organisms including yeasts, such as *Saccharomyces cerevisiae* (GenBank accession no. NP_011021), *Drosophila melanogaster* (GenBank accession no. Q27297), *Caenorhabditis elegans* (GenBank accession no. BAA24982), and *homo sapiens* (Yoshimura et al., "Cloning and sequence of the human RecA-like gene cDNA," *Nucl. Acids Res.* 21(7):1665 (1993); GenBank accession no. NP_002866).

[0043] Allelic variants and engineered mutants (collectively, muteins) of RecA and of its orthologues have also been described, including recA-803 (Madiraju et al., *Proc. Natl. Acad. Sci. USA* 85(18):6592-6 (1988).

[0044] In some embodiments, multiple oligonucleotides are employed to hybridize to a multiplicity of genomic regions of interest. The multiplicity of oligonucleotides can be directed to hybridize to a large genomic region of interest, allowing selective enrichment of a large genome region. The multiplicity of oligonucleotides can be directed to hybridize to a large genomic region of interest, allowing selective enrichment of a large genome region directed to hybridize to a multiplicity of genomic regions of interest on a single chromosome or multiplicity of chromosomes. The chromosomes can be of a specific species (for example, human, rat, mouse, other animals, specific plants; bacteria, algae, viruses, and the like) or a

IDT Exhibit 1020.3098

mixture of chromosomes of different species (such as host-pathogen, bacterial populations, and the like). Insofar as the genomic regions of interest may be very long, multiple oligonucleotides can be designed to hybridize to different sequence regions within the genomic region of interest. Tiling of oligonucleotides specific for hybridization across targeted genomic regions of interest is also envisioned.

[0045] The length of the oligonucleotides designed to carry out the methods of the invention can depend on the application, or user-defined criteria, and can be determined for maximum specificity. In addition, when a multiplicity of oligonucleotides are employed, the sequence/composition (e.g., % GC content) and length of the oligonucleotides can be designed to ensure specific and efficient hybridization at the same stringency for hybridization to all genomic regions of interest, thus enabling highly multiplex target enrichment process. It is clear to persons skilled in the art that the sequence content and oligonucleotide length can be adjusted to provide efficient and specific hybridization to the defined genomic regions of interest under a defined stringency as determined by the solution composition and temperature.

[0046] The oligonucleotides of the invention can be synthetic polynucleotides that are single stranded and contain a sequence that is capable of hybridizing with a sequence of the target input DNA. The region of an oligonucleotide that hybridizes with the target nucleic acid can be at least 80%, 90%, 95%, or 100% complementary to the target sequence. The number of nucleotides in the hybridizable sequence of a specific oligonucleotide can be such that stringency conditions used to hybridize the oligonucleotide can prevent excessive random non-specific hybridization. Usually, the number of nucleotides in the hybridizing portion of an oligonucleotide can be at least as great as the defined sequence on the target polynucleotide that the oligonucleotide hybridizes to, namely, at least 10, at least 15, at least 20, at least 25, or at least 30 nucleotides, or generally from about 25 to about 100 nucleotides, or from about 25 to about 60 nucleotides, or from about 25 to about 50 nucleotides, or from about 25 to about 35 nucleotides, or from about 20 to about 30 nucleotides. The melting temperature (T_m) of oligonucleotides with different sequences in a reaction can be similar or different.

[0047] Large sets of oligonucleotides can be synthesized by various methods. Methods for highly parallel synthesis of a large number of oligonucleotides on microarrays or other solid surface configuration (e.g., micro fluidic array platform) were developed in recent years (for example, by Affymetrix, Nimblegen, Agilent, Febit, LC Science and others). The number of oligonucleotides with different sequences that can be used in the methods of provided invention can be at least 10, 25, 50, 75, 100, 250, 500, 750, 1000, 2500, 5000, 7500, 10,000, 25,000, 50,000, 75,000, 100,000, 250,000, 500,000, 750,000, or 1,000,000. The number of oligonucleotides with different sequences that can be used in the methods of the provided invention can be about 100-1000, about 1000-10,000, about 10,000 to 100,000, or about 100,000 to 1,000,000.

[0048] In some embodiments, the hybridized oligonucleotides are extended along the sequence region of interest by a DNA polymerase.

IDT Exhibit 1020.3099

Ligands, Ligand Binding Components, and Specific Binding Pairs

[0049] The oligonucleotides used for carrying out the methods of the invention can comprise at least one label or ligand at one end. In some embodiments the label or ligand is attached to the oligonucleotides at the 5'-end. In other embodiments, the oligonucleotides have a label or ligand attached at both 3'- and 5'-ends. In another embodiment, the label or ligand is attached to only the 3' end of the oligonucleotide. Each end may be modified with different labels or ligands, enabling targeted tethering of defined NA-modifying enzymes to each end. Alternatively, the labels or ligands at each end of an oligonucleotide can be the same.

[0050] Generally, a ligand can be any compound for which a ligand-binding component or receptor naturally exists or can be prepared. A ligand can be a compound that is designed and engineered to interact with a receptor. Similarly, a ligand-binding component or receptor or anti-ligand can be generally any compound or composition capable of recognizing a particular spatial and polar organization of a ligand molecule, e.g., epitopic or determinant site.

[0051] The ligand and ligand- binding component can be considered to be members of a specific binding pair. Generally a specific binding pair member can be one of two different molecules (either the ligand or the ligand-binding component) having an area on the surface or in a cavity which specifically binds to and is thereby defined as complementary with a particular spatial and polar organization of the other molecule. The members of the specific binding pair can be referred to as a ligand and ligand-binding component. Exemplary embodiments include an immunological pair such as antigen-antibody, operator-repressor, nuclease-nucleotide, biotin-avidin, hormone-hormone receptor, IgG-protein A, DNA-DNA, DNA-RNA, and the like. Additional exemplary embodiments of ligand and ligand-binding components include ligand-receptor, substrate-enzyme (e.g., substrate-kinase, substrate-protease, substrate-phosphatase), protein domain homo-oligomers (e.g., homo-dimers, homo-trimers, homo-tetramers, or homo-polymers), protein domain hetero-oligomers (e.g., hetero-dimers, hetero-trimers, hetero-tetramers, or hetero-polymers), nucleic acid-protein nucleic acid binding domain (e.g., DNA-protein DNA binding domain, RNA-protein RNA binding domain), phosphate containing molecule-phospho binding protein, methylation containing molecule-methylation binding protein, acetyl-lysine containing molecule-bromodomain, phospho-Y-containing peptide-SH2 domain, calmodulin binding peptide (CBP)-calmodulin, or histidine tag-nickel, etc. The specific binding pair can be a pair of proteins that interact as found in the Database of Interacting Proteins (DIP) (http://dip.doe-mbi.ucla.edu/dip/Main.cgi) or the Database of Ligand-Receptor Partners (DLRP) (http://dip.doe-mbi.ucla.edu/dip/DLRP.cgi). An interacting partner can be a protein interaction domain described at http://pawsonlab.mshri.on.ca/index.php?option=com_content&task=view&id=30&Itemid=63.

[0052] In exemplary embodiments a ligand can be selected from but not limited to a small molecule, an antigen, an antibody, hybrid antibody or antibody fragment, an siRNA, an antisense RNA, an aptamer, a polypeptide, a peptide, and a first member of a specific binding pair such as biotin or avidin/streptavidin.

IDT Exhibit 1020.3100

[0053] In exemplary embodiments a ligand binding component can be selected from but not limited to a ligand binding protein, a small molecule, an antigen, an antibody, hybrid antibody or antibody fragment, an siRNA, an antisense RNA, an aptamer, a protein, a polypeptide, a peptide, and a second member of a specific binding pair such as avidin/streptavidin or biotin. Related exemplary ligand-binding components include naturally occurring and synthetic receptors, antibodies, enzymes, Fab fragments, lectins, nucleic acids, repressors, oligonucleotides, protein A, complement component C1q, or DNA binding proteins and the like.

Nucleic Acid Modifying Enzymes

[0054] The NA-modifying enzyme can be DNA-specific modifying enzyme. The NA-modifying enzyme can be a duplex-specific modifying enzyme.

[0055] The NA-modifying enzyme can be selected for specificity for dsDNA. The enzyme can be a duplex-specific endonuclease, a blunt-end frequent cutter restriction enzyme, or other restriction enzyme. Examples of blunt-end cutters include DraI or SmaI. The NA-modifying enzyme can be an enzyme provided by New England Biolabs (www.neb.com). The NA-modifying enzyme can be a homing endonuclease (a homing endonuclease can be an endonuclease that does not have a stringently-defined recognition sequence). The NA-modifying enzyme can be a nicking endonuclease (a nicking endonuclease can be an endonuclease that can cleave only one strand of DNA in a double-stranded DNA substrate). The NA-modifying enzyme can be a high fidelity endonuclease (a high fidelity endonuclease can be an engineered endonuclease that has less "star activity" than the wild-type version of the endonuclease). The NA-modifying enzyme can be an endonuclease that has a methylated recognition site (e.g., DpnI). The NA-modifying enzyme can be a 3-base cutter, 4-base cutter, 5-base cutter, 6-base cutter, 7-base cutter, 8-base cutter, 9-base cutter, 10-base cutter, 11-base cutter, 12-base cutter, 13-base cutter, 14-base cutter, 15-base cutter, 16-base cutter, 17-base cutter, 18-base cutter, 19-base cutter, or 20-base cutter endonuclease. The NA-modifying enzyme can be a thermophilic restriction enzyme. A restriction enzyme can recognize a palindromic sequence.

[0056] The NA-modifying enzyme can be a Type I endonuclease, a Type II endonuclease, a Type III endonuclease, or endoribonuclease. The NA-modifying enzyme can be a structure specific nuclease (e.g., a flap endonuclease (FEN)), a sequence specific nuclease, an endonuclease that can generate a "blunt end," an endonuclease that can generate a 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 base 5' overhang, an endonuclease that can generate a 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 base 3' overhang, a micrococcal nuclease (e.g., S7 Nuclease or MNase), mung bean nuclease, an S1 nuclease, nuclease Bal-31, or a P1 nuclease.

[0057] The NA-modifying enzyme can be an exonuclease (e.g., 3' to 5' exonuclease (e.g., exonuclease I, exonuclease III, exonuclease T), a 5' to 3' exonuclease (e.g., Lambda exonuclease, T7 exonuclease), a poly(A)-specific 3' to 5' exonuclease, a 3' to 5' exoribonuclease), a polymerase (e.g., a DNA polymerase, an RNA polymerase), a ligase (e.g., a DNA ligase), a phosphatase (e.g., a DNA phosphatase, a 3' phosphatase, alkaline phosphatase), a kinase (e.g., a DNA kinase, e.g., T4 polynucleotide kinase), a DNA

IDT Exhibit 1020.3101

methyltransferase, DNA gyrase, uracil-DNA glycosylase (UDG), a topoisomerase, a recombinase (e.g., Cre Recombinase), or a helicase.

[0058] Other NA modification enzymes are also envisioned to be useful for carrying out the methods of the invention.

[0059] The NA-modifying enzyme employed can require interaction with solution phase subunit. Various duplex-specific restriction enzymes require dimer formation for cleavage of both strands of a duplex. This required dimerization (the interaction of the tethered enzyme and a solution phase enzyme) can increase the specificity of cleavage by limiting it to the site of the sequence regions of interest.

[0060] The NA-modifying enzymes can be natural, recombinant, synthetic, or semi-synthetic. The NA-modifying enzymes can be chemical enzyme mimics. For example, such enzymes can be isolated from specific organisms or recombinantly generated. Exemplary NA-modifying enzymes include but are not limited to the duplex-specific nuclease from the hepatopancreas of the Kamchatka crab, *Drosophila* zinc-finger nucleases, and targeting individual subunits of the FokI restriction endonuclease to specific DNA strands. The action of the NA-modifying enzyme at selective sites in the sequence regions of interest can result in marking the sites for subsequent reaction, or lead to dsDNA cleavage which can render the sites suitable for appending desired adapters for a variety of downstream applications such as amplification, library generation, and sequencing.

Linkers

[0061] The label or ligand can be linked to the oligonucleotide by a linker, the synthesis of which is known in the art. The length and flexibility of the linker can affect the flexibility and reach of the tethered NA-modifying enzyme, as will be explained in the following. Various linkers are known in the art and are useful for carrying out the methods of the invention. The linker can comprise at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 33, 34, 35, 36, 37, 38, 39, 40 or more covalent bonds. The linker can comprise at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 or more carbons.

[0062] The flexibility and length of linkers used in the method of the invention can be designed to influence the reach of the NA-modifying enzyme to act on the desired sequence region of interest, or act on proximal sequences which can be within reach. Thus, linkers of defined flexibility and length can be used for attaching the label or ligand to the oligonucleotides, as well as the linkers used to attach the NA-modifying enzyme to the member of the specific binding pair (for binding to the label or ligand).

[0063] The attachment of the NA-modifying enzyme to the ligand member of the specific binding pair can be engineered so that the attachment is directed to specific moieties of either or both the member of the specific binding pair and the NA-modifying enzyme. Furthermore, the attachment can be genetically engineered to incorporate the desired moieties into either the member of the specific binding pair or the NA-modifying enzyme or both. A genetically engineered fused product of the member of the specific binding pair and the NA-modifying enzyme is also envisioned. Flexible linkers are well known in the art,

IDT Exhibit 1020.3102

for example PEG (polyethylene glycol) or PEO (hydrophilic polyethylene oxide) spacer arms of various lengths (for example as provided by Thermo Scientific (http://www.piercenet.com/products/browse.cfm?fldID=062CC432-FBF8-4E9D-8930-458385BCED1D).

Adapters

[0064] Various adapter designs are envisioned which are suitable for generation of amplification-ready products of target sequence regions/strands of interest. Ligation of adapters at the desired end of the sequence regions of interest is suitable for carrying out the methods of the invention. Various ligation modalities are envisioned, dependent on the choice of NA-modifying enzymes and the resulting dsDNA cleavage. For example, when a blunt end product comprising the target region/sequence of interest is generated, blunt end ligation can be suitable. Alternatively, where the cleavage is carried out using a restriction enzyme of known sequence specificity, leading to the generation of cleavage sites with known sequence overhangs, suitable ends of the adapters can be designed to enable hybridization of the adapter to the cleavage site of the sequence region of interest and subsequent ligation. Reagents and methods for efficient and rapid ligation of adapters are commercially available and are known in the art.

[0065] Various adapter designs are envisioned. Exemplary adapter designs are shown in Figures 4-7.

[0066] The adapters can be designed to comprise a double-stranded portion which can be ligated to the dsDNA (or dsDNA with overhang) products. Various ligation processes and reagents are known in the art and can be useful for carrying out the methods of the invention. For example, blunt ligation can be employed. Similarly, a single dA nucleotide can be added to the 3'-end of the dsDNA product, by a polymerase lacking 3'-exonuclease activity and can anneal to an adapter comprising a dT overhang (or the reverse). This design allows the hybridized components to be subsequently ligated (e.g., by T4 DNA ligase). Other ligation strategies and the corresponding reagents and known in the art and kits and reagents for carrying out efficient ligation reactions are commercially available (e.g, from New England Biolabs, Roche).

[0067] A common feature of the adapter designs depicted in Figures 4 to 7 is the 3'-single stranded DNA comprising a unique sequence (the sequence of an amplification primer) which can hybridize to an amplification primer employed in a subsequent amplification step, thus rendering the ligation product suitable for subsequent amplification of the sequence regions of interest.

[0068] A novel stem-loop adapter design is shown in Figure 7. The stem portion of the adapter comprises an RNA-DNA heteroduplex formed by hybridization of the 3'-portion of the oligonucleotide, comprising RNA sequence, to its 5'-end portion. The 3'-end of the oligonucleotide can comprise a dNTP to facilitate ligation when required for efficient ligation using the ligase of choice. The stem may or may not be blunt ended, as required for efficient hybridization to the cleavage products of the sequence sequences of interest generated by the methods of the invention. In this embodiment, the ligation products comprise a loop at one end attached to the cleavage product via an RNA-DNA heteroduplex. The RNA portion of the heteroduplex is susceptible to cleavage by RNase H. The adapter further

IDT Exhibit 1020.3103

comprises a loop sequence that is hybridizable to an amplification primer used for the single primer amplification of the sequence regions of interest. Cleavage of the RNA portion of the stem by RNase H results is the formation of ssDNA 3'-overhang suitable which is the primer binding site for downstream amplification.

[0069] The dsDNA portion of the adapters can further comprise indexing or bar-coding sequences designed to mark either the samples or sequences of interest. The ability to incorporate marking and/or bar-coding information in the adapter design can further expand the high multiplexing ability of the methods of the invention.

Methods of Amplification

[0070] The methods, compositions and kits described herein can be useful to generate amplification-ready products directly from genomic DNA for downstream applications such as massively parallel sequencing (Next Generation Sequencing methods), multiplexed quantification of large sets of sequence regions of interest, such as by high density qPCR arrays and other highly parallel quantification platforms (selective massively parallel target pre-amplification), as well as generation of libraries with enriched population of sequence regions of interest. The methods described herein can be used to generate a collection of at least 25, 50, 75, 100, 500, 1000, 2500, 5000, 10,000, 25,000, 50,000, 100,000, 500,000, or 1,000,000 amplification-ready target sequence regions of interest directly from a sample of complex DNA using a plurality of oligonucleotides.

[0071] Methods of nucleic acid amplification are well known in the art. In some embodiments, the amplification method is isothermal. In other embodiments the amplification method is linear. In other embodiments the amplification is exponential.

SPIA Amplification

[0072] Amplification of the sequence regions of interest employing a linear amplification method such as the single primer isothermal amplification (SPIA) can be used. SPIA enables generation of multiple copies of the strand specific sequence regions of interest and employs a single amplification primer, thus reducing the complexity associated with multiple oligonucleotide design and manufacturing, enables the use of a generic amplification primer, and is linear. The fidelity of quantification of the copy number of the sequence regions of interest in the complex genomic NA sample is a highly desirable feature of the presented methods of the invention.

[0073] Amplification by SPIA can occur under conditions permitting composite primer hybridization, primer extension by a DNA polymerase with strand displacement activity, cleavage of RNA from a RNA/DNA heteroduplex and strand displacement. In so far as the composite amplification primer hybridizes to the 3'-single-stranded portion (of the partially double stranded polynucleotide which is formed by cleaving RNA in the complex comprising a RNA/DNA partial heteroduplex) comprising, generally, the complement of at least a portion of the composite amplification primer sequence,

IDT Exhibit 1020.3104

composite primer hybridization may be under conditions permitting specific hybridization. In SPIA, all steps are isothermal (in the sense that thermal cycling is not required), although the temperatures for each of the steps may or may not be the same. It is understood that various other embodiments can be practiced given the general description provided above. For example, as described and exemplified herein, certain steps may be performed as temperature is changed (e.g., raised, or lowered).

[0074] Although generally only one composite amplification primer is described above, it is further understood that the SPIA amplification methods can be performed in the presence of two or more different first and/or second composite primers that randomly prime template polynucleotide. In addition, the amplification polynucleotide products of two or more separate amplification reactions conducted using two or more different first and/or second composite primers that randomly prime template polynucleotide can be combined.

[0075] The composite amplification primers are primers that are composed of RNA and DNA portions. In the amplification composite primer, both the RNA and the DNA portions are generally complementary and can hybridize to a sequence in the amplification-ready product to be copied or amplified. In some embodiments, a 3'-portion of the amplification composite primer is DNA and a 5'-portion of the composite amplification primer is RNA. The composite amplification primer is designed such that the primer is extended from the 3'-DNA portion to create a primer extension product. The 5'-RNA portion of this primer extension product in a RNA/DNA heteroduplex is susceptible to cleavage by RNase H, thus freeing a portion of the polynucleotide to the hybridization of an additional composite amplification primer. The extension of the amplification composite primer by a DNA polymerase with strand displacement activity releases the primer extension product from the original primer and creates another copy of the sequence of the polynucleotide. Repeated rounds of primer hybridization, primer extension with strand displacement DNA synthesis, and RNA cleavage create multiple copies of the strand-specific sequence of the polynucleotide.

Other Amplification Methods

[0076] Some aspects of the invention comprise the amplification of polynucleotide molecules or sequences within the polynucleotide molecules. Amplification generally refers to a method that can result in the formation of one or more copies of a nucleic acid or polynucleotide molecule or in the formation of one or more copies of the complement of a nucleic acid or polynucleotide molecule. Amplifications can be used in the invention, for example, to amplify or analyze a polynucleotide bound to a solid surface. The amplifications can be performed, for example, after archiving the samples in order to analyze the archived polynucleotide.

[0077] In some aspects of the invention, exponential amplification of nucleic acids or polynucleotides is used. These methods often depend on the product catalyzed formation of multiple copies of a nucleic acid or polynucleotide molecule or its complement. The amplification products are sometimes referred to as "amplicons." One such method for the enzymatic amplification of specific double stranded sequences

IDT Exhibit 1020.3105

of DNA is polymerase chain reaction (PCR). This *in vitro* amplification procedure is based on repeated cycles of denaturation, oligonucleotide primer annealing, and primer extension by thermophilic template dependent polynucleotide polymerase, resulting in the exponential increase in copies of the desired sequence of the polynucleotide analyte flanked by the primers. The two different PCR primers, which anneal to opposite strands of the DNA, are positioned so that the polymerase catalyzed extension product of one primer can serve as a template strand for the other, leading to the accumulation of a discrete double stranded fragment whose length is defined by the distance between the 5' ends of the oligonucleotide primers. Other amplification techniques that can be used in the methods of the provided invention include, e.g., AFLP (amplified fragment length polymorphism) PCR (see e.g. Vos et al. 1995. AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research* 23: 4407-14), allele-specific PCR (see e.g., Saiki RK, Bugawan TL, Horn GT, Mullis KB, Erlich HA (1986). Analysis of enzymatically amplified beta-globin and HLA-DQ alpha DNA with allele-specific oligonucleotide probes *Nature* 324: 163-166), Alu PCR, assembly PCR (see e.g., Stemmer WP, Crameri A, Ha KD, Brennan TM, Heyneker HL (1995). Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides *Gene* 164: 49-53), assymetric PCR (see e.g., Saiki RK *supra*), colony PCR, helicase dependent PCR (see e.g., Myriam Vincent, Yan Xu and Huimin Kong (2004). Helicase-dependent isothermal DNA amplification *EMBO reports* 5 (8): 795–800), hot start PCR, inverse PCR (see e.g., Ochman H, Gerber AS, Hartl DL. *Genetics*. 1988 Nov;120(3):621-3), *in situ* PCR, intersequence-specific PCR or ISSR PCR, digital PCR, linear-after-the-exponential-PCR or Late PCR (see e.g., Pierce KE and Wangh LJ (2007). Linear-after-the-exponential polymerase chain reaction and allied technologies Real-time detection strategies for rapid, reliable diagnosis from single cells *Methods Mol Med.* 132: 65-85), long PCR, nested PCR, real-time PCR, duplex PCR, multiplex PCR, quantitative PCR, or single cell PCR.

[0078] Another method for amplification involves amplification of a single stranded polynucleotide using a single oligonucleotide primer. The single stranded polynucleotide that is to be amplified contains two non-contiguous sequences that are complementary to one another and, thus, are capable of hybridizing together to form a stem-loop structure. This single stranded polynucleotide already may be part of a polynucleotide analyte or may be created as the result of the presence of a polynucleotide analyte.

[0079] Another method for achieving the result of an amplification of nucleic acids is known as the ligase chain reaction (LCR). This method uses a ligase enzyme to join pairs of preformed nucleic acid probes. The probes hybridize with each complementary strand of the nucleic acid analyte, if present, and ligase is employed to bind each pair of probes together resulting in two templates that can serve in the next cycle to reiterate the particular nucleic acid sequence.

[0080] Another method for achieving nucleic acid amplification is the nucleic acid sequence based amplification (NASBA). This method is a promoter-directed, enzymatic process that induces *in vitro* continuous, homogeneous and isothermal amplification of a specific nucleic acid to provide RNA copies

IDT Exhibit 1020.3106

of the nucleic acid. The reagents for conducting NASBA include a first DNA primer with a 5'-tail comprising a promoter, a second DNA primer, reverse transcriptase, RNAse-H, T7 RNA polymerase, NTP's and dNTP's.

[0081] Another method for amplifying a specific group of nucleic acids is the Q-beta-replicase method, which relies on the ability of Q-beta-replicase to amplify its RNA substrate exponentially. The reagents for conducting such an amplification include "midi-variant RNA" (amplifiable hybridization probe), NTP's, and Q-beta-replicase.

[0082] Another method for amplifying nucleic acids is known as 3SR and is similar to NASBA except that the RNAse-H activity is present in the reverse transcriptase. Amplification by 3SR is an RNA specific target method whereby RNA is amplified in an isothermal process combining promoter directed RNA polymerase, reverse transcriptase and RNase H with target RNA. See for example Fahy et al. PCR Methods Appl. 1:25-33 (1991).

[0083] Another method for amplifying nucleic acids is the Transcription Mediated Amplification (TMA) used by Gen-Probe. The method is similar to NASBA in utilizing two enzymes in a self-sustained sequence replication. See U.S. Patent No. 5,299,491 herein incorporated by reference.

[0084] Another method for amplification of nucleic acids is Strand Displacement Amplification (SDA) (Westin et al 2000, *Nature Biotechnology*, 18, 199-202; Walker et al 1992, *Nucleic Acids Research*, 20, 7, 1691-1696), which is an isothermal amplification technique based upon the ability of a restriction endonuclease such as HincII or BsoBI to nick the unmodified strand of a hemiphosphorothioate form of its recognition site, and the ability of an exonuclease deficient DNA polymerase such as Klenow exo minus polymerase, or Bst polymerase, to extend the 3'-end at the nick and displace the downstream DNA strand. Exponential amplification results from coupling sense and antisense reactions in which strands displaced from a sense reaction serve as targets for an antisense reaction and vice versa.

[0085] Another method for amplification of nucleic acids is Rolling Circle Amplification (RCA) (Lizardi et al. 1998, *Nature Genetics*, 19:225-232). RCA can be used to amplify single stranded molecules in the form of circles of nucleic acids. In its simplest form, RCA involves the hybridization of a single primer to a circular nucleic acid. Extension of the primer by a DNA polymerase with strand displacement activity results in the production of multiple copies of the circular nucleic acid concatenated into a single DNA strand.

[0086] In some embodiments of the invention, RCA is coupled with ligation. For example, a single oligonucleotide can be used both for ligation and as the circular template for RCA. This type of polynucleotide can be referred to as a "padlock probe" or a "RCA probe." For a padlock probe, both termini of the oligonucleotide contain sequences complementary to a domain within a nucleic acid sequence of interest. The first end of the padlock probe is substantially complementary to a first domain on the nucleic acid sequence of interest, and the second end of the padlock probe is substantially complementary to a second domain, adjacent to the first domain near the first domain. Hybridization of the oligonucleotide to the target nucleic acid results in the formation of a hybridization complex. Ligation

IDT Exhibit 1020.3107

of the ends of the padlock probe results in the formation of a modified hybridization complex containing a circular polynucleotide. In some cases, prior to ligation, a polymerase can fill in the gap by extending one end of the padlock probe. The circular polynucleotide thus formed can serve as a template for RCA that, with the addition of a polymerase, results in the formation of an amplified product nucleic acid. The methods of the invention described herein can produce amplified products with defined sequences on both the 5'-and 3'-ends. Such amplified products can be used as padlock probes.

[0087] Some aspects of the invention utilize the linear amplification of nucleic acids or polynucleotides. Linear amplification generally refers to a method that involves the formation of one or more copies of the complement of only one strand of a nucleic acid or polynucleotide molecule, usually a nucleic acid or polynucleotide analyte. Thus, the primary difference between linear amplification and exponential amplification is that in the latter process, the product serves as substrate for the formation of more product, whereas in the former process the starting sequence is the substrate for the formation of product but the product of the reaction, i.e. the replication of the starting template, is not a substrate for generation of products. In linear amplification the amount of product formed increases as a linear function of time as opposed to exponential amplification where the amount of product formed is an exponential function of time.

[0088] In some embodiments, amplification methods can be solid-phase amplification, polony amplification, colony amplification, emulsion PCR, bead RCA, surface RCA, surface SDA, etc., as will be recognized by one of skill in the art. In some embodiments, amplification methods that results in amplification of free DNA molecules in solution or tethered to a suitable matrix by only one end of the DNA molecule can be used. Methods that rely on bridge PCR, where both PCR primers are attached to a surface (see, e.g., WO 2000/018957 and Adessi et al., *Nucleic Acids Research* (2000): 28(20): E87) can be used. In some cases the methods of the invention can create a "polymerase colony technology," or "polony." referring to a multiplex amplification that maintains spatial clustering of identical amplicons (see Harvard Molecular Technology Group and Lipper Center for Computational Genetics website). These include, for example, in situ polonies (Mitra and Church, *Nucleic Acid Research* 27, e34, Dec. 15, 1999), in situ rolling circle amplification (RCA) (Lizardi et al., *Nature Genetics* 19, 225, July 1998), bridge PCR (U.S. Pat. No. 5,641,658), picotiter PCR (Leamon et al., *Electrophoresis* 24, 3769, November 2003), and emulsion PCR (Dressman et al., *PNAS* 100, 8817, Jul. 22, 2003). The methods of the invention provide new methods for generating and using polonies.

Downstream Applications for Amplified Products

[0089] An important aspect of the invention is that the methods and compositions disclosed herein can be efficiently, cost-effectively, and with the minimal loss of biological material, utilized for downstream analyses. The amplified and enriched copies of the selected sequence regions of interest are useful for massively parallel sequencing (Next Generation Sequencing methods), multiplexed quantification of large sets of sequence regions of interest, such as by high density qPCR arrays and other highly parallel

IDT Exhibit 1020.3108

quantification platforms (selective massively parallel target pre-amplification), as well as generation of libraries with enriched population of sequence regions of interest.

Sequencing

[0090] In one embodiment, the invention provides for products ready for amplification in preparation for sequencing.

[0091] In some embodiments, the methods are useful for preparing target polynucleotide(s) for sequencing by the sequencing by ligation methods commercialized by Applied Biosystems (e.g., SOLiD sequencing). In general, double stranded fragment polynucleotides can be prepared by the methods of the present invention, and then incorporated into a water-in-oil emulsion along with polystyrene beads and amplified, for example by PCR. In some cases, alternative amplification methods can be employed in the water-in-oil emulsion such as any of the methods provided herein. The amplified product in each water microdroplet formed by the emulsion can interact, bind, or hybridize with the one or more beads present in that microdroplet leading to beads with a plurality of amplified products of substantially one sequence. When the emulsion is broken, the beads float to the top of the sample and are placed onto an array. The methods can include a step of rendering the nucleic acid bound to the beads single- stranded or partially single stranded. Sequencing primers are then added along with a mixture of four different fluorescently labeled oligonucleotide probes. The probes bind specifically to the two bases in the polynucleotide to be sequenced immediately adjacent and 3' of the sequencing primer to determine which of the four bases are at those positions. After washing and reading the fluorescence signal from the first incorporated probe, a ligase is added. The ligase cleaves the oligonucleotide probe between the fifth and sixth bases, removing the fluorescent dye from the polynucleotide to be sequenced. The whole process is repeated using a different sequence primer until all of the intervening positions in the sequence are imaged. The process allows the simultaneous reading of millions of DNA fragments in a 'massively parallel' manner. This 'sequence-by-ligation' technique uses probes that encode for two bases rather than just one allowing error recognition by signal mismatching, leading to increased base determination accuracy.

[0092] In some embodiments, the methods are useful for preparing target polynucleotide(s) for sequencing by synthesis using the methods commercialized by 454/Roche Life Sciences including but not limited to the methods and apparatus described in Margulies et al., *Nature* (2005) 437:376-380 (2005); and US Patent Nos. 7,244,559; 7,335,762; 7,211,390; 7,244,567; 7,264,929; and 7,323,305. In general, double stranded fragment polynucleotides can be prepared by the methods of the present invention, immobilized onto beads, and compartmentalized in a water-in-oil PCR emulsion. In some cases, alternative amplification methods can be employed in the water-in-oil emulsion such as any of the methods provided herein. When the emulsion is broken, amplified fragments remain bound to the beads. The methods can include a step of rendering the nucleic acid bound to the beads single stranded or partially single stranded. The beads can be enriched and loaded into wells of a fiber optic slide so that there is approximately 1 bead in each well. Nucleotides are flowed across and into the wells in a fixed

IDT Exhibit 1020.3109

order in the presence of polymerase, sulfhydrolase, and luciferase. Addition of nucleotides complementary to the target strand can result in a chemiluminescent signal that is recorded, such as by a camera. The combination of signal intensity and positional information generated across the plate allows software to determine the DNA sequence.

[0093] In some embodiments, the methods are useful for preparing target polynucleotide(s) for sequencing by the methods commercialized by Helicos BioSciences Corporation (Cambridge, MA) as described in U.S. application Ser. No. 11/167046, and U.S. Patent Nos. 7501245; 7491498; 7,276,720; and in U.S. Patent Application Publication Nos. US20090061439; US20080087826; US20060286566; US20060024711; US20060024678; US20080213770; and US20080103058. In general, double stranded fragment polynucleotides can be prepared by the methods of the present invention, and then immobilized onto a flow-cell surface. The methods can include a step of rendering the nucleic acid bound to the flow-cell surface stranded or partially single stranded. Polymerase and labeled nucleotides are then flowed over the immobilized DNA. After fluorescently labeled nucleotides are incorporated into the DNA strands by a DNA polymerase, the surface is illuminated with a laser, and an image is captured and processed to record single molecule incorporation events to produce sequence data.

[0094] In some embodiments, the methods are useful for preparing target polynucleotide(s) for sequencing by the sequencing by ligation methods commercialized by Dover Systems. Generally, double stranded fragment polynucleotides can be prepared by the methods of the present invention. The polynucleotides can then be amplified in an emulsion in the presence of magnetic beads. Any amplification methods can be employed in the water-in-oil emulsion such as any of the methods provided herein. The resulting beads with immobilized clonal polynucleotide polonies are then purified by magnetic separation, capped, amine functionalized, and covalently immobilized in a series of flow cells. The methods can include a step of rendering the nucleic acid bound to the flow-cell surface stranded or partially single stranded. Then, a series of anchor primers are flowed through the cell, where they hybridize to the synthetic oligonucleotide sequences at the 3' or 5' end of proximal or distal genomic DNA tags. Once an anchor primer is hybridized, a mixture of fully degenerate nonanucleotides ('nonamers') and T4 DNA ligase is flowed into the cell; each of the nonamer mixture's four components is labeled with one of four fluorophores, which correspond to the base type at the query position. The fluorophore-tagged nonamers selectively ligate onto the anchor primer, providing a fluorescent signal that identifies the corresponding base on the genomic DNA tag. Once the probes are ligated, fluorescently labeling the beads, the array is imaged in four colors. Each bead on the array will fluoresce in only one of the four images, indicating whether there is an A, C, G, or T at the position being queried. After imaging, the array of annealed primer-fluorescent probe complex, as well as residual enzyme, are chemically striped using guanidine HCl and sodium hydroxide. After each cycle of base reads at a given position have been completed, and the primer-fluorescent probe complex has been stripped, the anchor primer is replaced, and a new mixture of fluorescently tagged nonamers is introduced, for which the query position is shifted one base further into the genomic DNA tag. Seven bases are queried in this fashion, with the

IDT Exhibit 1020.3110

sequence performed from the 5' end of the proximal tag, followed by six base reads with a different anchor primer from the 3' end of the proximal tag, for a total of 13 base pair reads for this tag. This sequence is then repeated for the 5' and 3' ends of the distal tag, resulting in another 13 base pair reads. The ultimate result is a read length of 26 bases (thirteen from each of the paired tags). However, it is understood that this method is not limited to 26 base read lengths.

[0095] In some embodiments, the methods are useful for preparing target polynucleotide(s) from selectively enriched populations of specific sequence regions of interest in a strand-specific manner for sequencing by the methods well known in the art and further described below.

[0096] For example the methods are useful for sequencing by the method commercialized by Illumina as described US Patent Nos. 5,750,341; 6,306,597; and 5,969,119. In general, double stranded fragment polynucleotides can be prepared by the methods of the present invention to produce amplified nucleic acid sequences tagged at one (e.g., (A)/(A') or both ends (e.g., (A)/(A') and (C)/(C')). In some cases, single stranded nucleic acid tagged at one or both ends is amplified by the methods of the present invention (e.g., by SPIA or linear PCR). The resulting nucleic acid is then denatured and the single stranded amplified polynucleotides are randomly attached to the inside surface of flow-cell channels. Unlabeled nucleotides are added to initiate solid-phase bridge amplification to produce dense clusters of double-stranded DNA. To initiate the first base sequencing cycle, four labeled reversible terminators, primers, and DNA polymerase are added. After laser excitation, fluorescence from each cluster on the flow cell is imaged. The identity of the first base for each cluster is then recorded. Cycles of sequencing are performed to determine the fragment sequence one base at a time. For paired-end sequencing, such as for example, when the polynucleotides are labeled at both ends by the methods of the present invention, sequencing templates can be regenerated in-situ so that the opposite end of the fragment can also be sequenced.

[0097] In some embodiments, the methods are useful for preparing target polynucleotide(s) for sequencing by the methods commercialized by Pacific Biosciences as described in US Patent Nos. 7462452; 7476504; 7405281; 7170050; 7462468; 7476503; 7315019; 7302146; 7313308; and US Application Publication Nos. US20090029385; US20090068655; US20090024331; and US20080206764. In general, double stranded fragment polynucleotides can be prepared by the methods of the present invention. The polynucleotides can then be immobilized in zero mode waveguide arrays. The methods may include a step of rendering the nucleic acid bound to the waveguide arrays single stranded or partially single stranded. Polymerase and labeled nucleotides are added in a reaction mixture, and nucleotide incorporations are visualized via fluorescent labels attached to the terminal phosphate groups of the nucleotides. The fluorescent labels are clipped off as part of the nucleotide incorporation. In some cases, circular templates are utilized to enable multiple reads on a single molecule.

[0098] Another example of a sequencing technique that can be used in the methods of the provided invention is nanopore sequencing (see e.g. Soni GV and Meller A. (2007) *Clin Chem* 53: 1996-2001). A nanopore can be a small hole of the order of 1 nanometer in diameter. Immersion of a nanopore in a

IDT Exhibit 1020.3111

conducting fluid and application of a potential across it can result in a slight electrical current due to conduction of ions through the nanopore. The amount of current that flows is sensitive to the size of the nanopore. As a DNA molecule passes through a nanopore, each nucleotide on the DNA molecule obstructs the nanopore to a different degree. Thus, the change in the current passing through the nanopore as the DNA molecule passes through the nanopore can represent a reading of the DNA sequence.

[0099] Another example of a sequencing technique that can be used in the methods of the provided invention is semiconductor sequencing provided by Ion Torrent (e.g., using the Ion Personal Genome Machine (PGM)). Ion Torrent technology can use a semiconductor chip with multiple layers, e.g., a layer with micro-machined wells, an ion-sensitive layer, and an ion sensor layer. Nucleic acids can be introduced into the wells, e.g., a clonal population of single nucleic can be attached to a single bead, and the bead can be introduced into a well. To initiate sequencing of the nucleic acids on the beads, one type of deoxyribonucleotide (e.g., dATP, dCTP, dGTP, or dTTP) can be introduced into the wells. When one or more nucleotides are incorporated by DNA polymerase, protons (hydrogen ions) are released in the well, which can be detected by the ion sensor. The semiconductor chip can then be washed and the process can be repeated with a different deoxyribonucleotide. A plurality of nucleic acids can be sequenced in the wells of a semiconductor chip. The semiconductor chip can comprise chemical-sensitive field effect transistor (chemFET) arrays to sequence DNA (for example, as described in U.S. Patent Application Publication No. 20090026082). Incorporation of one or more triphosphates into a new nucleic acid strand at the 3' end of the sequencing primer can be detected by a change in current by a chemFET. An array can have multiple chemFET sensors.

Genetic Analysis

[00100] The methods of the present invention can be used in the analysis of genetic information of selective genomic regions of interest as well as genomic regions which may interact with the selective region of interest. Amplification methods as disclosed herein can be used in the devices, kits, and methods known to the art for genetic analysis, such as, but not limited to those found in U.S. Patent Nos. 6449562, 6287766, 7361468, 7414117, 6225109, and 6110709. In some cases, amplification methods of the present invention can be used to amplify target nucleic acid for DNA hybridization studies to determine the presence or absence of polymorphisms. The polymorphisms, or alleles, can be associated with diseases or conditions such as genetic disease. In other cases the polymorphisms can be associated with susceptibility to diseases or conditions, for example, polymorphisms associated with addiction, degenerative and age related conditions, cancer, and the like. In other cases, the polymorphisms can be associated with beneficial traits such as increased coronary health, or resistance to diseases such as HIV or malaria, or resistance to degenerative diseases such as osteoporosis, Alzheimer's or dementia.

Additional uses of the methods of the invention:

IDT Exhibit 1020.3112

[00101] The methods of the invention can also find use in the analysis of genomic DNA modifications such as methylation. It is envisioned that directed hybridization of multiplicity of oligonucleotides useful for carrying out the methods of the invention can be designed to specifically hybridize to regions of interest and the action of the tethered NA-modifying enzymes can specifically act, or be prevented from acting on modified nucleotides in the complex genomic DNA sample, as compared to the non modified canonical nucleotides. It is also envisioned that the selected tethered NA-modifying enzyme can specifically recognize nucleotide modifications and mark the sequence for subsequent cleavage.

[00102] The tethered NA-modifying enzyme can also act on sequence regions in the complex genomic DNA as these come in contact, are come to close proximity, to the selected sequence regions of interest.

Kits

[00103] Any of the compositions described herein may be comprised in a kit. In a non-limiting example the kit, in suitable container means, comprises: one oligonucleotide or a plurality of oligonucleotides, each coupled with at least one ligand, and a ligand-binding agent coupled to a NA-modifying enzyme. The kit can further contain adapters and/or reagents useful for ligation. The kit can further optionally contain a DNA-polymerase. The kit can further optionally contain reagents for amplification, for example reagents useful for single primer isothermal amplification methods. The kit can further optionally contain reagents for sequencing, for example, reagents useful for next-generation massively parallel sequencing methods.

[00104] The containers of the kits can generally include at least one vial, test tube, flask, bottle, syringe or other containers, into which a component may be placed, and preferably, suitably aliquoted. Where there is more than one component in the kit, the kit also can generally contain a second, third or other additional container into which the additional components can be separately placed. However, various combinations of components can be comprised in a container.

[00105] When the components of the kit are provided in one or more liquid solutions, the liquid solution can be an aqueous solution. However, the components of the kit can be provided as dried powder(s). When reagents and/or components are provided as a dry powder, the powder can be reconstituted by the addition of a suitable solvent.

[00106] A kit can include instructions for employing the kit components as well the use of any other reagent not included in the kit. Instructions can include variations that can be implemented.

EXAMPLES

Example 1 - Method 1, sequence-specific enrichment of a target region of interest from denatured double-stranded genomic DNA

[00107] Genomic DNA sample is mixed with a solution comprising oligonucleotides designed to specifically hybridize to the sequence regions of interest to be enriched and amplified. The oligonucleotides of the invention comprise a label linked to one end, preferably the 5'-end. The label can

IDT Exhibit 1020.3113

be any member of a specific binding pair, for example biotin. The ligand (label) can be attached to the end of the oligonucleotides by a linker. The length and composition of the linker will depend on the desired flexibility. Various linkers and methods for synthesis are well known in the art. The length and composition (%GC) of the oligonucleotide are selected to provide efficient hybridization to the multiplicity of sequence regions of interest at a predetermined stringency. The design requirements are well known in the art. Methods used for the design and selection of oligonucleotides employed for analyses employing high density microarrays are well known in the art and are suitable for the design of the multiplicity of oligonucleotides suitable for use in the methods of the current invention. The multiplicity of oligonucleotides for use in a method of the invention can be synthesized by any of the known methods, such as provided by synthesis employing microarrays as provided by various venders including Agilent, LC Sciences, Febit and the like. The number of oligonucleotides selected for the enrichment of a given sequence region of interest can vary from a single oligonucleotide to numerous oligonucleotides dependent on the length of the sequence region of interest.

[00108] The mixture is heated to denature the double stranded genomic DNA (dsDNA) and is cooled to allow hybridization of the multiplicity of oligonucleotides comprising a label or ligand to the specific sequence regions of interest (Figure 1).

[00109] The hybridized oligonucleotides are optionally extended along the sequence region of interest by DNA polymerase.

[00110] An example of the process is further illustrated in Figure 2. A mixture comprising nucleic acid modifying enzyme conjugated to a second member of the specific binding pair, for example avidin or streptavidin when the label or ligand is biotin, is added to the reaction mixture. Binding of the second member of the specific binding pair to the ligand attached to the 5'-end of the oligonucleotide results in anchoring the NA-modifying enzyme to the hybridized oligonucleotide or the extended oligonucleotide.

[00111] The NA-modifying enzyme is selected for specificity for dsDNA. The enzyme may be a duplex-specific endonuclease or a blunt end frequent cutter restriction enzyme, or other restriction enzymes. Other NA modification enzymes are also envisioned to be useful for carrying out the methods of the invention.

[00112] The NA-modifying enzyme employed may require interaction with solution phase subunit. Various duplex specific restriction enzymes require dimer formation for cleavage of both strands of the duplex. The required dimerization (e.g., the interaction of the tethered enzyme and a solution phase enzyme) will increase the specificity of cleavage by limiting it to the site of the sequence regions of interest.

[00113] The cleavage of both strands of the duplex formed by the hybridized (and optionally extended) oligonucleotide and the sequence region of interest results in the formation of a double stranded product (may be a partial double stranded product) without the tethered NA modification enzyme, and a partial double stranded product with the tethered NA-modifying enzyme (Figure 3). The latter comprises a short segment of the oligonucleotide which may be too short for stable hybridization, thus preventing ligation

IDT Exhibit 1020.3114

of the adapters to that product. Alternatively, adapter ligation to the desired end is controlled by polarity and sequence content of the ligation ends. The double stranded, or partially dsDNA products, without the label or ligand bearing segment of the oligonucleotide, are substrates for ligation.

[00114] Various adapter designs are envisioned. Some examples are shown in Figures 4, 5, 6 and 7. The adapters are designed to comprise a double-stranded portion which can be ligated to the dsDNA (or dsDNA with overhang) products. Various ligation process and reagents are known in the art and are useful for carrying out the methods of the invention. For example blunt ligation can be employed. Similarly, a single dA nucleotide can be added to the 3'-end of the dsDNA product, by a polymerase lacking 3'-exonuclease, and the adapter designed to comprise a dT overhang (or the reverse). This design allows hybridization of the two components to be subsequently ligated. Other ligation strategies and the corresponding reagents are known in the art, and kits and reagents for carrying out efficient ligation reactions are commercially available.

[00115] A common feature of the adapter designs depicted in Figures 4 to 7 is the 3'-single stranded DNA comprising a unique sequence which is hybridizable to an amplification primer employed in subsequent amplification step. As shown in the above mentioned figures, strand specific amplification of the sequence regions of interest is carried out by the single primer isothermal amplification (SPIA). A chimeric RNA-DNA amplification primer, DNA polymerase with strand displacement activity, and RNase H are employed in the isothermal linear amplification (see e.g., U.S. Patent Nos. 7402386, 7354717, 7351557, 7771946, and 7771934). The single amplification primer comprises a 3'-DNA sequence and a 5'-RNA sequence. The primer sequence is hybridizable to the 3'-ssDNA portion of the ligation product, and can be partially hybridizable to the sequence of the dsDNA portion of the adapter. The dsDNA portion of the adapters can further comprise indexing or bar-coding sequences designed to mark either the samples or sequences of interests. The ability to incorporate marking and/or bar-coding information in the adapter design further expands the high multiplexing ability of the methods of the invention.

[00116] A novel stem-loop adapter design is shown in Figure 7. The stem portion of the adapter comprises an RNA-DNA heteroduplex formed by hybridization of the 3'-portion of the oligonucleotide, comprising RNA sequence, to its 5'-end portion. The 3'-end of the oligonucleotide can comprise a dNTP to facilitate ligation, when required for efficient ligation using the ligase of choice. The stem may or may not be blunt ended, as required for efficient hybridization to the cleavage products of the sequence sequences of interest generated by the methods of the invention. The ligation products comprise a loop at one end attached to the cleavage product via an RNA-DNA heteroduplex. The RNA portion of the heteroduplex is susceptible to cleavage by RNase H. The adapter further comprises a loop sequence that is hybridizable to the amplification primer used for the single primer amplification of the sequence regions of interest. Cleavage of the RNA portion of the stem by RNase H results in the formation of a ssDNA 3'-overhang which is the primer binding site. Amplification of the strand specific sequence region of interest can be carried out by repeated primer extension steps or preferably by the isothermal

IDT Exhibit 1020.3115

single primer linear amplification (SPIA). Following ligation, the ligation product is added to amplification reaction mixture comprising the chimeric amplification primer, RNase H and DNA polymerase with strand displacement activity. The RNA portion of the stem heteroduplex is cleaved by RNase H to open the stem loop. Multiple copies of the strand specific selected sequence regions of interest is carried out by hybridization of the chimeric RNA-DNA chimeric primer, extension of the primer along the target strand by DNA polymerase, cleavage of the RNA portion of the hybridized primer by RNase H, and hybridization of a new primer at the free primer binding site. Repeated cycles of this process generate multiple copies of the ligated target strand.

[00117] The entire population of ligation products, comprising the multiplicity of selected sequence regions of interest, is amplified using the generic chimeric amplification primer, thus generating multiple copies of the strand-specific selected sequence regions of interest. All copies comprise the copies of the targeted sequence regions of interest and can further comprise the stem sequence of the adapter which in turn can comprise marking and/or bar coding information.

[00118] The products of the selected enrichment and amplification are useful for quantification of the various targeted sequence regions of interest in the complex genomic nucleic acid sample, for example by qPCR. Massively parallel quantification platforms have been developed and can be useful for genome-wide quantification of sequences of interest. The products can further be useful for generation of libraries of sequence regions of interest which can be analyzed by sequencing with or without prior cloning. The products can be rendered suitable for massively parallel sequencing using any of various methods and platforms developed in recent years. In addition, the marked and bar coded products generated from various samples (e.g., individually) can be combined and analyzed in a mixture. Comparative analysis of the libraries obtained from various samples can be used for the elucidation of the effect of sequence, structural or copy number variations on phenotype, development, differentiation and the like.

Example 2 - Method 2, sequence-specific enrichment of a target region of interest from non-denatured double-stranded genomic DNA

[00119] Targeted tethering of a NA-modifying enzyme can be achieved by sequence-specific triplex formation as shown in Figures 8 and 9. Oligonucleotides comprising a label or ligand are designed to specifically hybridize to dsDNA sequence regions of interest in a complex genomic DNA sample to form triplex. The targeted tethering of the NA-modifying enzyme is carried out in a manner similar to that of Method 1. The modifying enzyme is selected to directly or indirectly cleave the dsDNA target upstream of the triplex, and the targeted cleavage site is rendered suitable for ligation of adapters as described for Method 1. The generation of enriched and amplified copies of the multiplicity of sequence regions of interest is carried out similar to that described above (Method 1).

[00120] The method is useful for generation of copies of strand specific sequence regions of interest from complex nucleic acid without the need to denature the dsDNA input DNA, thus enabling enrichment and

IDT Exhibit 1020.3116

analysis of a multiplicity of sequence regions of interest in the native complex nucleic acid sample. It is envisioned that the method will find use for studies and analyses carried out *in situ*, enable studies and analysis of complex genomic DNA in single cells or collection of very small well defined cell population, as well as the analysis of complex genomic DNA without disruption of chromatin structures.

Example 3: Use of sequence-specific enrichment of a target region of interest from denatured double-stranded genomic DNA for diagnosing cancer

[00121] Genomic DNA is isolated from a patient sample (e.g., a tumor sample) and mixed with oligonucleotides that comprise a biotin linked to the 5' end through a 12 carbon spacer linker. The genomic DNA is denatured, and the biotin linked oligonucleotides are hybridized to the genomic DNA and extended with a DNA polymerase. A "blunt-cutter" endonuclease conjugated to streptavidin is added to the reaction mixture, and the endonuclease binds the biotin linked oligonucleotides via the conjugated streptavidin. The endonuclease cleaves both strands of the duplex formed by the hybridized and extended oligonucleotide. Adapters are added to the end of cleaved DNA. The adaptors permit binding of chimeric primers for subsequent linear amplification of the cleaved DNA. The adaptors also provide binding sites for primers that permit Next Generation Sequencing (e.g., Illumina sequencing). The results of the sequencing are analyzed to determine the presence of absence of mutations associated with cancer.

[00122] While preferred embodiments of the present invention have been shown and described herein, it will be obvious to those skilled in the art that such embodiments are provided by way of example only. Numerous variations, changes, and substitutions will now occur to those skilled in the art without departing from the invention. It should be understood that various alternatives to the embodiments of the invention described herein may be employed in practicing the invention. It is intended that the following claims define the scope of the invention and that methods and structures within the scope of these claims and their equivalents be covered thereby.

IDT Exhibit 1020.3117

CLAIMS

WHAT IS CLAIMED IS:

1. A method for selectively partitioning a plurality of target regions of interest from complex DNA comprising:
    a. hybridizing one or more oligonucleotides to the target regions of interest to form a plurality of target-oligonucleotide complexes;
    b. tethering a nucleic acid-modifying enzyme to a target-oligonucleotide complex via the oligonucleotide; and
    c. cleaving the target region of interest in the complex, thereby releasing the target region of interest from the enzyme.
2. The method of claim 1 wherein the one or more oligonucleotides hybridized to the target regions of interest are extended along the target region of interest by DNA polymerase prior to step b.
3. The method of claim 1 wherein the complex DNA comprises double-stranded DNA.
4. The method of claim 1 wherein the complex DNA comprises genomic DNA.
5. The method of claim 4 wherein the genomic DNA comprises a mixture of genomic DNA from more than one organism.
6. The method of claim 1 wherein the complex DNA comprises cDNA.
7. The method of claim 6 wherein the cDNA is generated from a mixture of DNAs from more than one organism.
8. The method of claim 1 wherein the selective partitioning is strand-specific.
9. The method of claim 1 wherein the enzyme is a DNA duplex-specific endonuclease.
10. The method of claim 1 wherein the enzyme is a restriction enzyme.
11. The method of claim 1 wherein the method further comprises denaturing the complex DNA prior to hybridization of the oligonucleotides.
12. The method of claim 1 wherein the method further comprises the formation of partial triplexes.
13. The method of claim 1 wherein the method further comprises ligating adapters to the target regions of interest once released from the enzyme.
14. The method of claim 13 wherein the adapter is selected from a group consisting of a double stranded adapter with an overhang at one end, a double stranded adapter with a 3' single stranded overhang, an adapter that comprises a RNA-DNA heteroduplex, an adapter that comprises a chimeric DNA-RNA oligonucleotide and a stem-loop adapter.

IDT Exhibit 1020.3118

15. The method of claim 1 wherein the selective partitioning is carried out directly on the complex DNA.

16. The method of claim 1 wherein said method does not involve amplifying the complex DNA prior to selective partitioning.

17. The method of claim 1 wherein the method further comprises amplifying the partitioned target regions of interest thereby enriching for the target regions of interest.

18. The method of claim 17 wherein the amplifying comprises single primer isothermal amplification.

19. The method of claim 17 wherein the method further comprises sequencing of the amplified products.

20. The method of claim 19 wherein the sequencing is performed using a massively parallel sequencing method.

21. The method of claim 1 wherein the enzyme is synthetic, semisynthetic, or recombinant.

22. The method of claim 1 wherein the ligand is selected from the group consisting of a small molecule, an antigen, an antibody, hybrid antibody or antibody fragment, an siRNA, an antisense RNA, an aptamer, a peptide and a first member of a specific binding pair.

23. The method of claim 1 wherein the ligand-binding component is selected from the group consisting of a small molecule, an antigen, an antibody, hybrid antibody or antibody fragment, an siRNA, an antisense RNA, an aptamer, a protein therapeutic, or a peptide, a ligand binding protein, and a second member of specific binding pair.

24. The method of claim 1 wherein at least 100 different regions of interest are selectively partitioned.

25. The method of claim 1 wherein at least 100 oligonucleotides are used for selectively partitioning the target region of interest from complex DNA.

26. The method of claim 25 wherein each oligonucleotide is coupled to the same ligand.

27. The method of claim 1 wherein the ligand is biotin and the ligand-binding component is avidin or streptavidin.

28. A method of preparing amplification-ready selectively targeted regions of interest from double-stranded DNA wherein the method comprises:

    a. denaturing the double-stranded DNA thereby generating single-stranded DNA;

    b. hybridizing to the single-stranded target regions of interest one or more oligonucleotides to form partial duplexes, each oligonucleotide being coupled to a ligand;

IDT Exhibit 1020.3119

c. contacting the ligands with a ligand-binding component coupled to a nucleic acid-modifying enzyme, wherein the enzyme cleaves the target regions of interest thereby obtaining products comprising the target regions of interest free of the enzyme; and

d. ligating adapters to the products, whereby the targeted regions of interest are amplification-ready.

29. A method of preparing amplification-ready selectively targeted regions of interest from double-stranded DNA wherein the method comprises:

a. hybridizing to the double-stranded target regions of interest one or more oligonucleotides to form partial triplexes, each oligonucleotide being coupled to a ligand;

b. contacting the ligands with a ligand-binding component coupled to a nucleic acid-modifying enzyme, wherein the enzyme cleaves the target regions of interest thereby obtaining products comprising the target regions of interest free of the enzyme; and

c. ligating adapters to the products, whereby the targeted regions of interest are amplification-ready.

30. The method of claims 28 or 29 wherein the double-stranded DNA comprises genomic DNA.

31. The method of claims 28 or 29 wherein the double-stranded DNA comprises cDNA.

32. The method of claim 31 wherein the cDNA is generated from a mixture of DNAs from more than one organism.

33. The method of claim 30 wherein the genomic DNA comprises a mixture of genomic DNA from more than one organism.

34. The method of claims 28 or 29 wherein the method is strand-specific.

35. The method of claims 28 or 29 wherein the enzyme is a DNA-duplex specific endonuclease.

36. The method of claims 28 or 29 wherein the enzyme is a restriction enzyme.

37. The method of claims 28 or 29 wherein the adapter is selected from a group consisting of a double stranded adapter with an overhang at one end, a double stranded adapter with a 3' single stranded overhang, an adapter that comprises a RNA-DNA heteroduplex, an adapter that comprises a chimeric DNA-RNA oligonucleotide and a stem-loop adapter.

38. The method of claims 28 or 29 wherein the method is carried out directly on the double-stranded DNA.

39. The method of claim 28 wherein said method does not involve amplifying the double-stranded DNA until after step d.

IDT Exhibit 1020.3120

40. The method of claim 29 wherein said method does not involve amplifying the double-stranded DNA until after step c.

41. The method of claims 28 or 29 wherein the method further comprises amplifying the target regions of interest, thereby enriching for the target sequence regions of interest.

42. The method of claim 41 wherein the amplifying comprises single primer isothermal amplification.

43. The method of claim 41 wherein the method further comprises sequencing of the amplified products.

44. The method of claim 43 wherein the sequencing is performed using a massively parallel sequencing method.

45. The method of claims 28 or 29 wherein the enzyme is synthetic, semisynthetic or recombinant.

46. The method of claims 28 or 29 wherein the ligand is selected from the group consisting of a small molecule, an antigen, an antibody, hybrid antibody or antibody fragment, an siRNA, an antisense RNA, an aptamer, a peptide and a first member of a specific binding pair.

47. The method of claims 28 or 29 wherein the ligand-binding component is selected from the group consisting of a small molecule, an antigen, an antibody, hybrid antibody or antibody fragment, an siRNA, an antisense RNA, an aptamer, a protein therapeutic, or a peptide, a ligand binding protein, and a second member of specific binding pair.

48. The method of claims 28 or 29 wherein the ligand is biotin and the ligand-binding component is avidin or streptavidin.

49. The method of claims 28 or 29 wherein at least 100 different regions of interest are prepared for amplification.

50. The method of claims 28 or 29 wherein at least 100 oligonucleotides are used for preparing the target regions of interest for amplification.

51. The method of claims 28 or 29 wherein each oligonucleotide is coupled to the same ligand.

52. A DNA complex comprising:

    a. DNA comprising at least one DNA strand;

    b. an oligonucleotide hybridized to the DNA wherein the oligonucleotide is coupled to a ligand; and

    c. a ligand-binding component coupled to a nucleic acid-modifying enzyme, wherein the ligand and the ligand-biding component are further coupled to each other.

53. The complex of claim 52 wherein the DNA comprises genomic DNA.

IDT Exhibit 1020.3121

54. The complex of claim 52 wherein the DNA is double stranded and the complex is a partial triplex.

55. The complex of claim 52 wherein the DNA is single stranded and the complex is a partial duplex.

56. The complex of claim 52 wherein the nucleic acid-modifying enzyme is synthetic, semisynthetic, or recombinant.

57. The complex of claim 52 wherein the ligand is selected from the group consisting of a small molecule, an antigen, an antibody, hybrid antibody or antibody fragment, an siRNA, an antisense RNA, an aptamer, a peptide and a first member of a specific binding pair.

58. The complex of claim 52 wherein the ligand binding component is selected from the group consisting of a small molecule, an antigen, an antibody, hybrid antibody or antibody fragment, an siRNA, an antisense RNA, an aptamer, a protein therapeutic, or a peptide, a ligand binding protein, and a second member of specific binding pair.

59. The complex of claim 52 wherein the ligand is biotin and the ligand binding component is avidin or streptavidin.

60. The complex of claim 52 wherein the oligonucleotide is further extended with a DNA polymerase.

61. A kit comprising:

    a. one or more oligonucleotides, each coupled to a ligand; and

    b. a ligand-binding component that selectively binds to said ligand, wherein said ligand-binding component is coupled to a nucleic acid-modifying enzyme.

62. The kit of claim 61 wherein the kit further comprises reagents for amplification.

63. The kit of claim 62 wherein the kit further comprises an adapter for amplification.

64. The kit of claim 63 wherein the adapter is selected from a group consisting of a double stranded adapter with an overhang at one end, a double stranded adapter with a 3' single stranded overhang, an adapter that comprises a RNA-DNA heteroduplex, an adapter that comprises a chimeric DNA-RNA oligonucleotide and a stem-loop adapter.

65. The kit of claim 62 wherein the kit further comprises reagents for sequencing.

66. The kit of claim 65 wherein the sequencing reagents comprise reagents for a massively parallel sequencing method.

67. The kit of claim 62 wherein the reagents are reagents for performing single primer isothermal amplification.

68. The kit of claim 61 wherein the kit further comprises a DNA polymerase.

IDT Exhibit 1020.3122

69. The kit of claim 61 wherein the kit further comprises a plurality of oligonucleotides.

70. The kit of claim 69 wherein the plurality of oligonucleotides are coupled to different ligands.

71. The kit of claim 61 wherein the ligand is selected from the group consisting of a small molecule, an antigen, an antibody, hybrid antibody or antibody fragment, an siRNA, an antisense RNA, an aptamer, a peptide and a first member of a specific binding pair.

72. The kit of claim 61 wherein the ligand binding component is selected from the group consisting of a small molecule, an antigen, an antibody, hybrid antibody or antibody fragment, an siRNA, an antisense RNA, an aptamer, a protein therapeutic, or a peptide, a ligand binding protein, and a second member of specific binding pair.

73. The kit of claim 61 wherein the ligand is biotin and the ligand-binding component is avidin or streptavidin.

74. The kit of claim 61 wherein the nucleic acid modifying enzyme is synthetic, semisynthetic, or recombinant.

75. The kit of claim 61 wherein the nucleic acid modifying enzyme is a DNA duplex-specific endonuclease.

76. The kit of claim 61 wherein the nucleic acid modifying enzyme is a restriction endonuclease.

77. A plurality of amplification-ready partial DNA duplexes generated from genomic DNA, each duplex comprising at least one DNA strand comprising the sequence of a target region of interest wherein the duplex is further ligated to an adaptor for single primer isothermal amplification.

IDT Exhibit 1020.3123

Figure 1: Strand specific selective hybridization

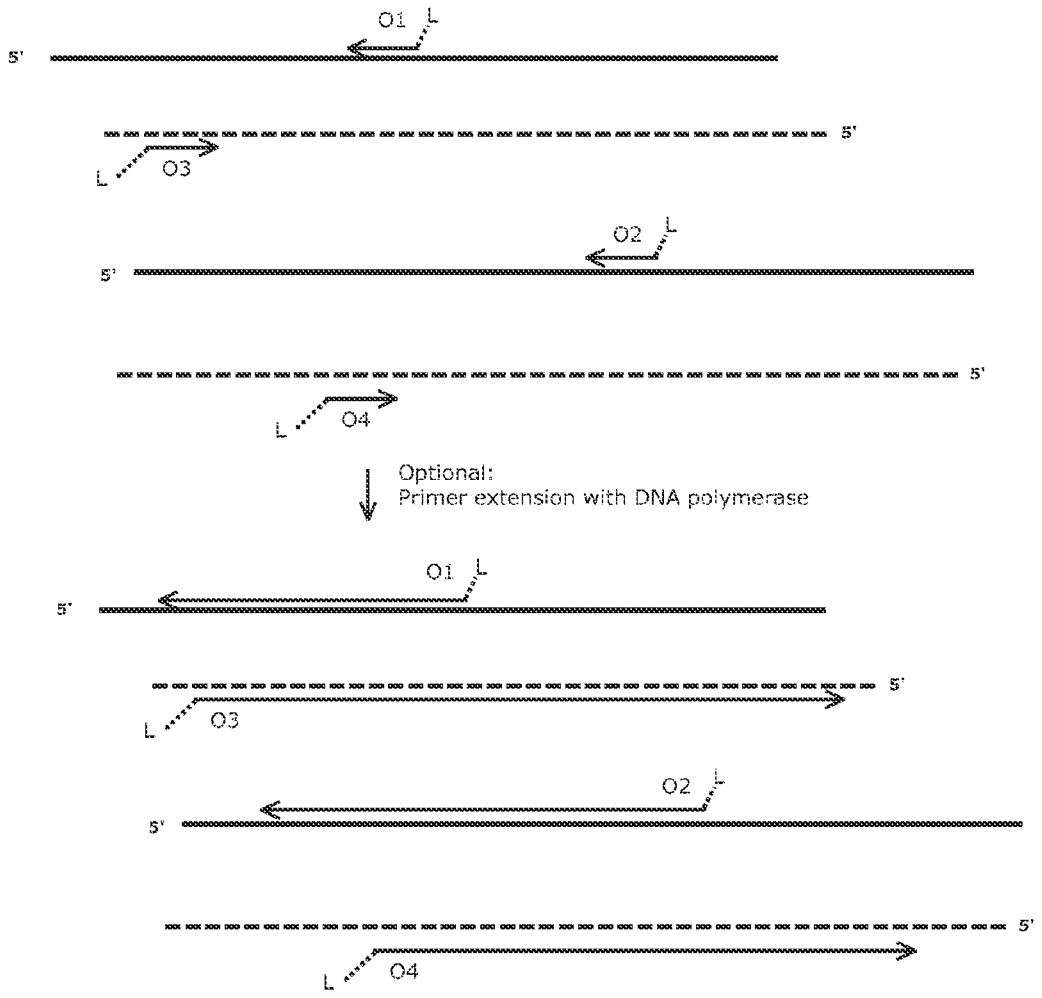


Figure 2: Selective tethering of NA-modifying enzyme at target sequence regions of interest

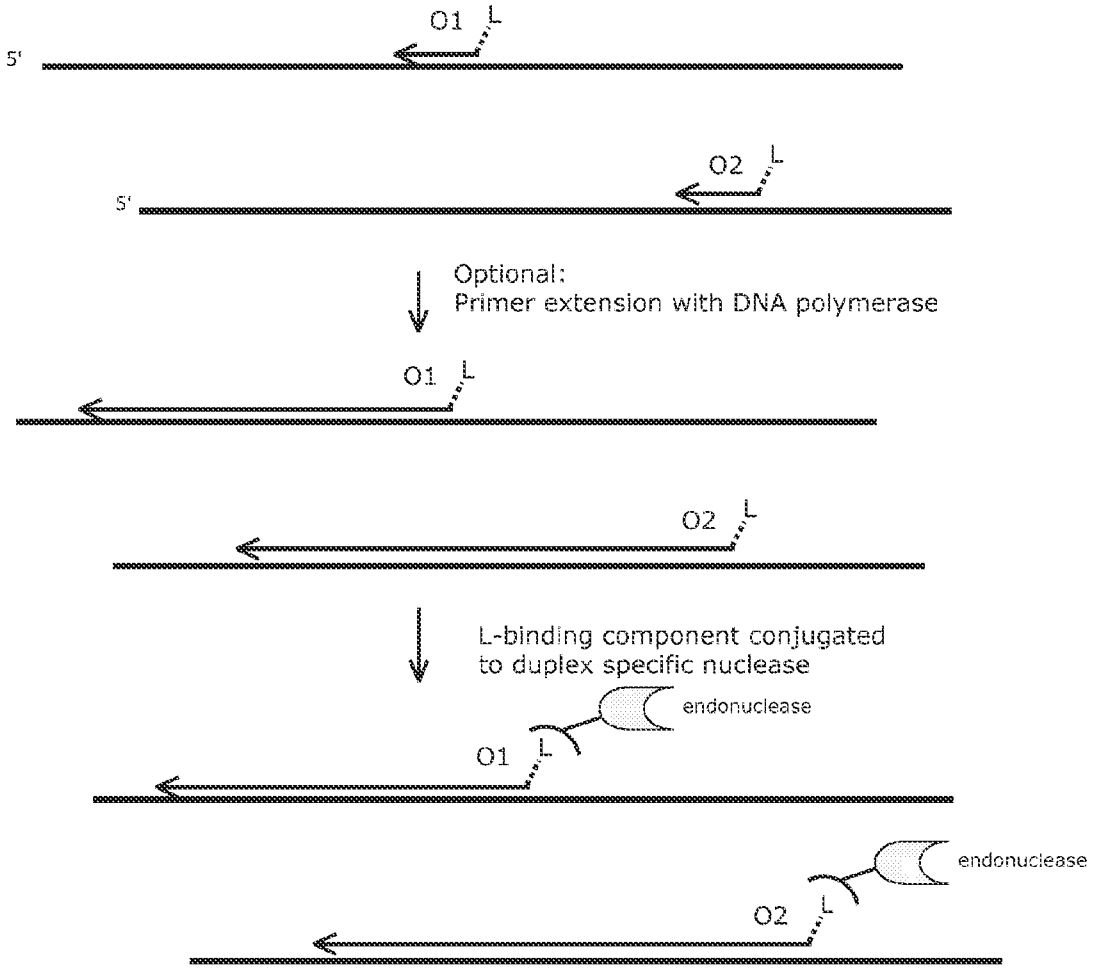


Figure 3: Selective cleavage at target sequence regions of interest

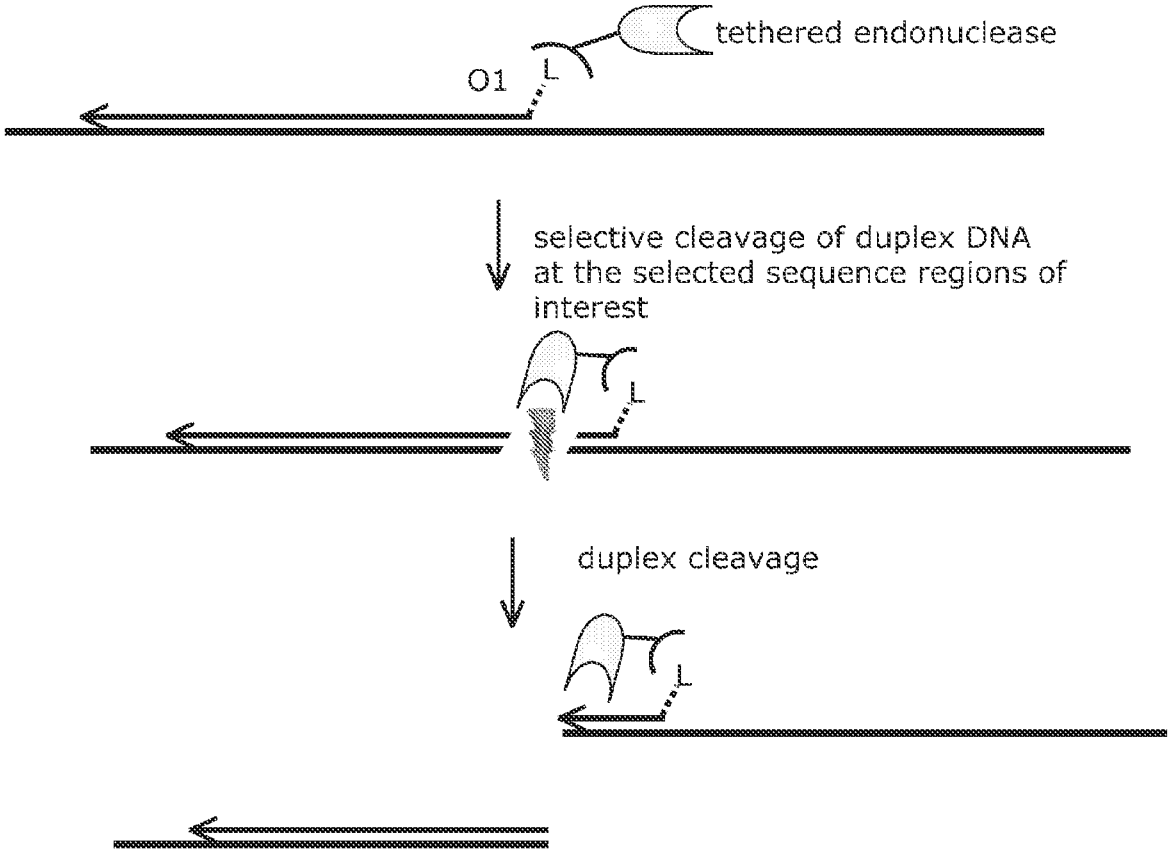


Figure 4: Adapter ligation and selective amplification of target sequence regions of interest

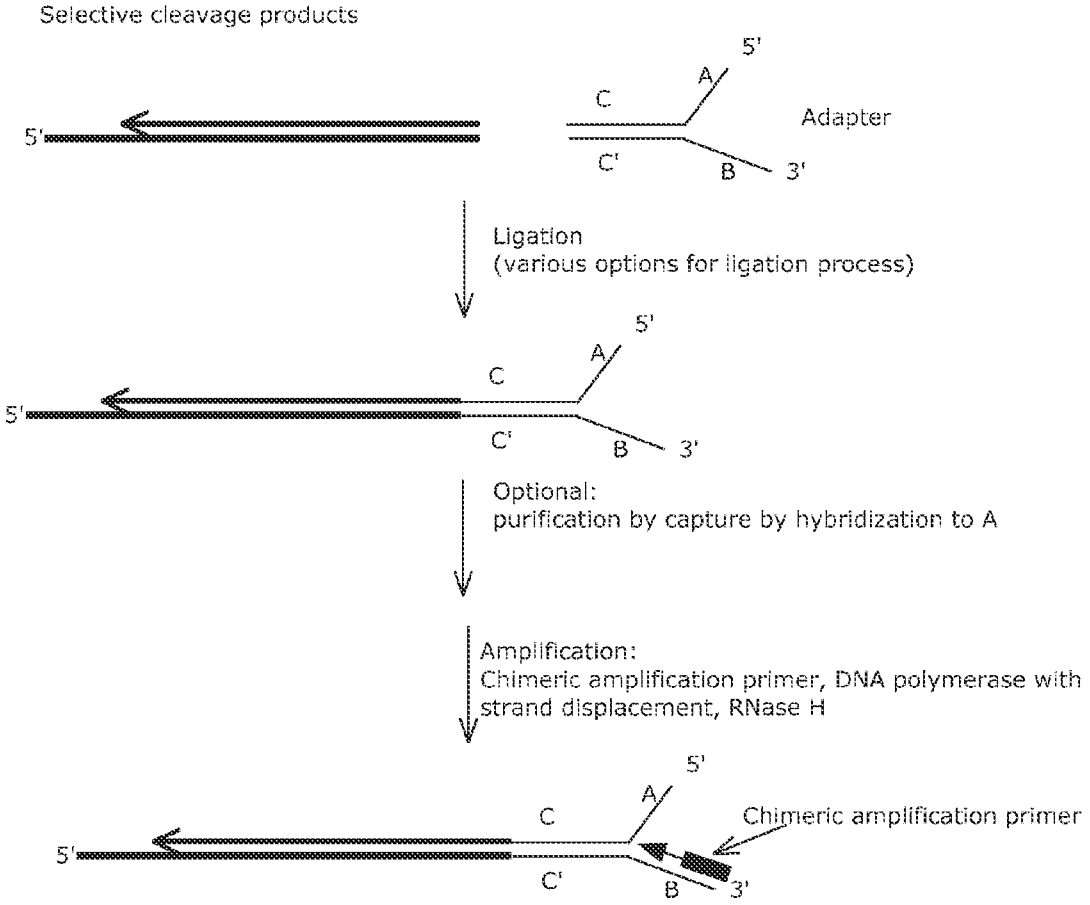


Figure 5: Adapter ligation and selective amplification of target sequence regions of interest

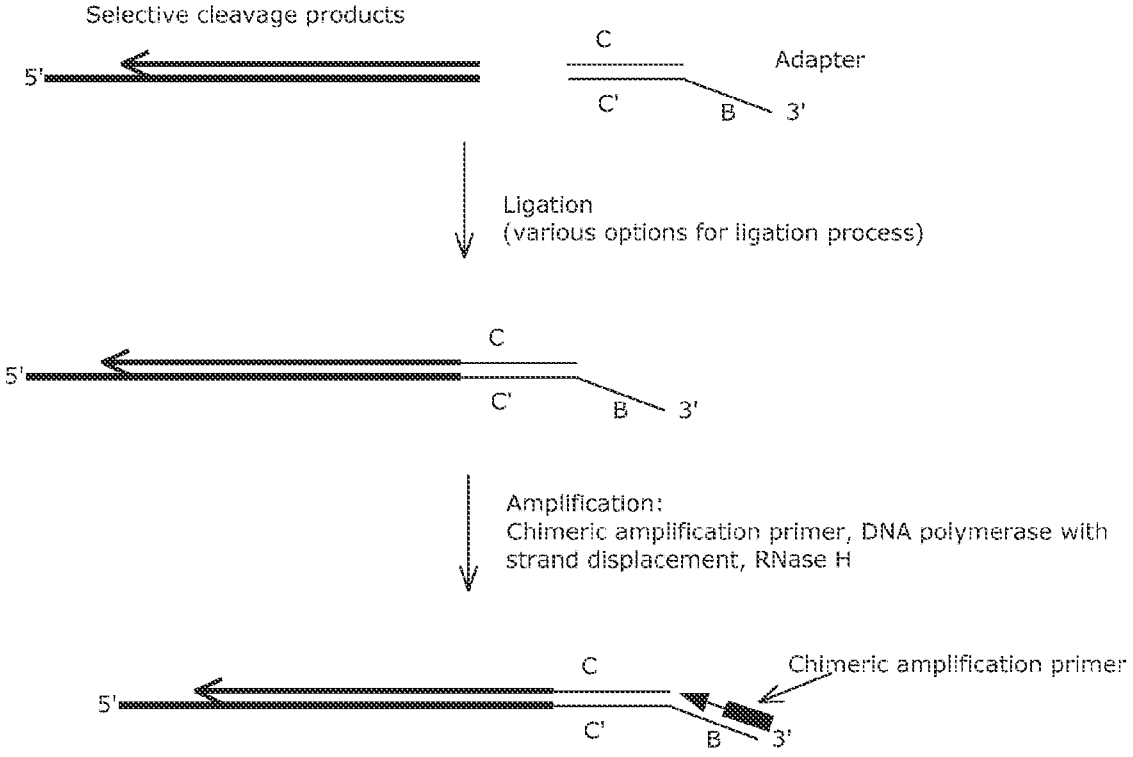


Figure 6: adapter ligation and selective amplification of target sequence regions of interest

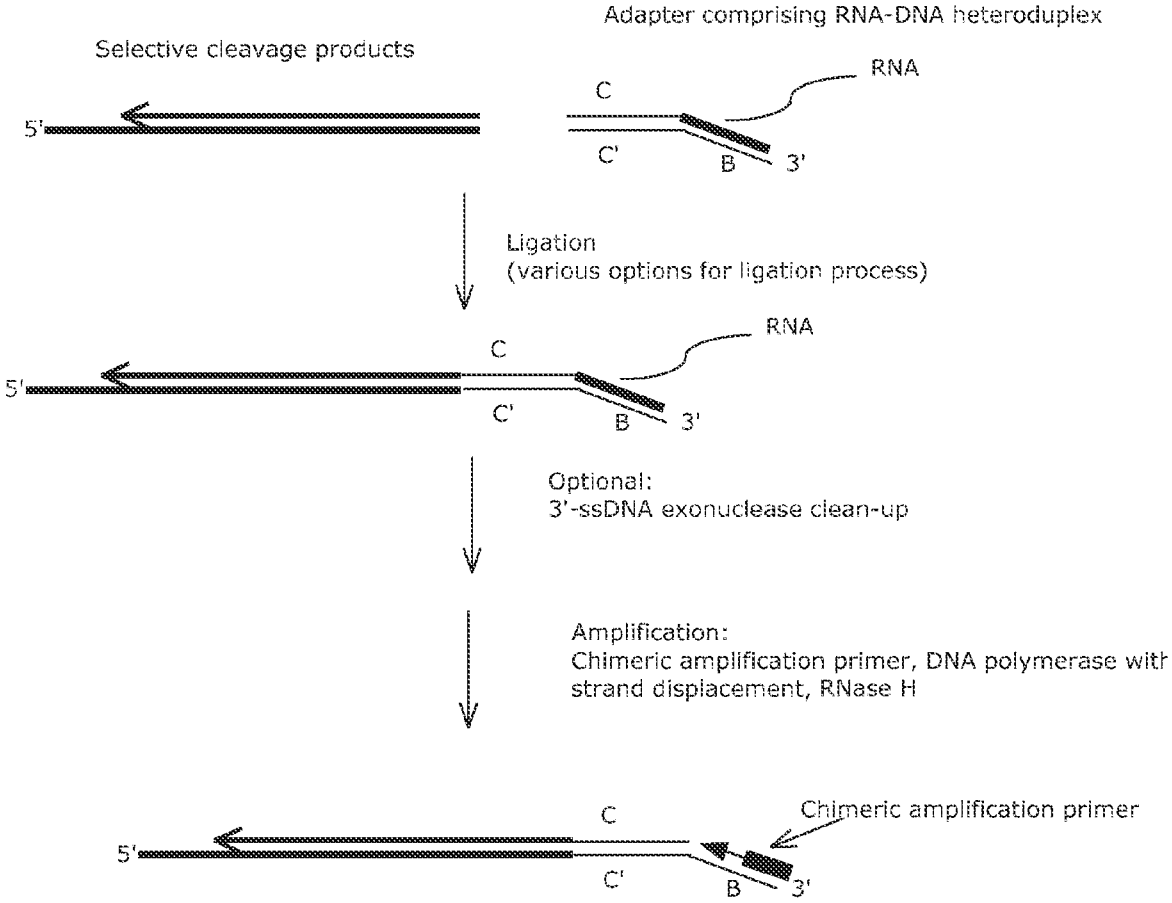


Figure 7: Adapter ligation and selective amplification of target sequence regions of interest

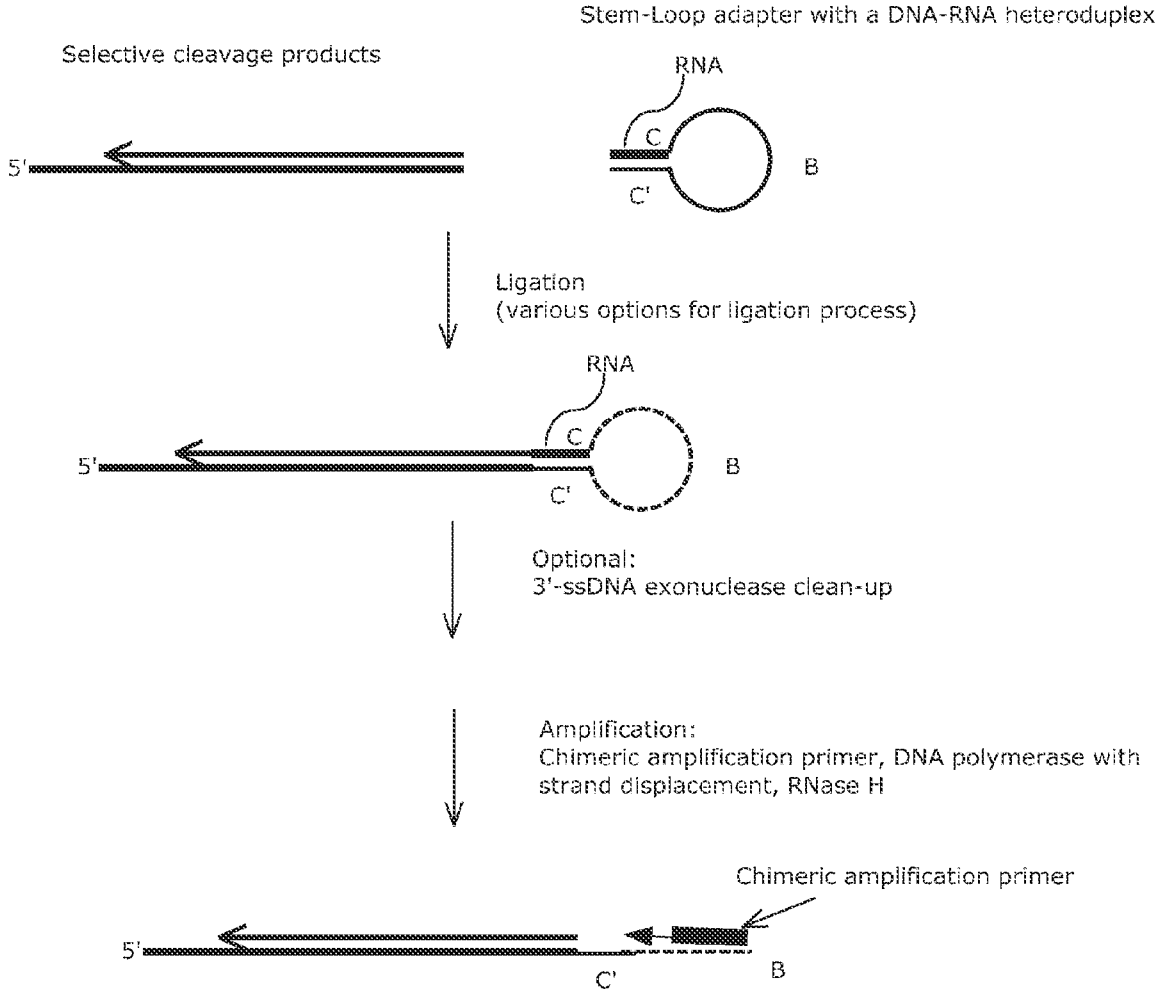


Figure 8: Selective triplex formation and tethering of NA-modifying enzyme at target sequence regions of interest.

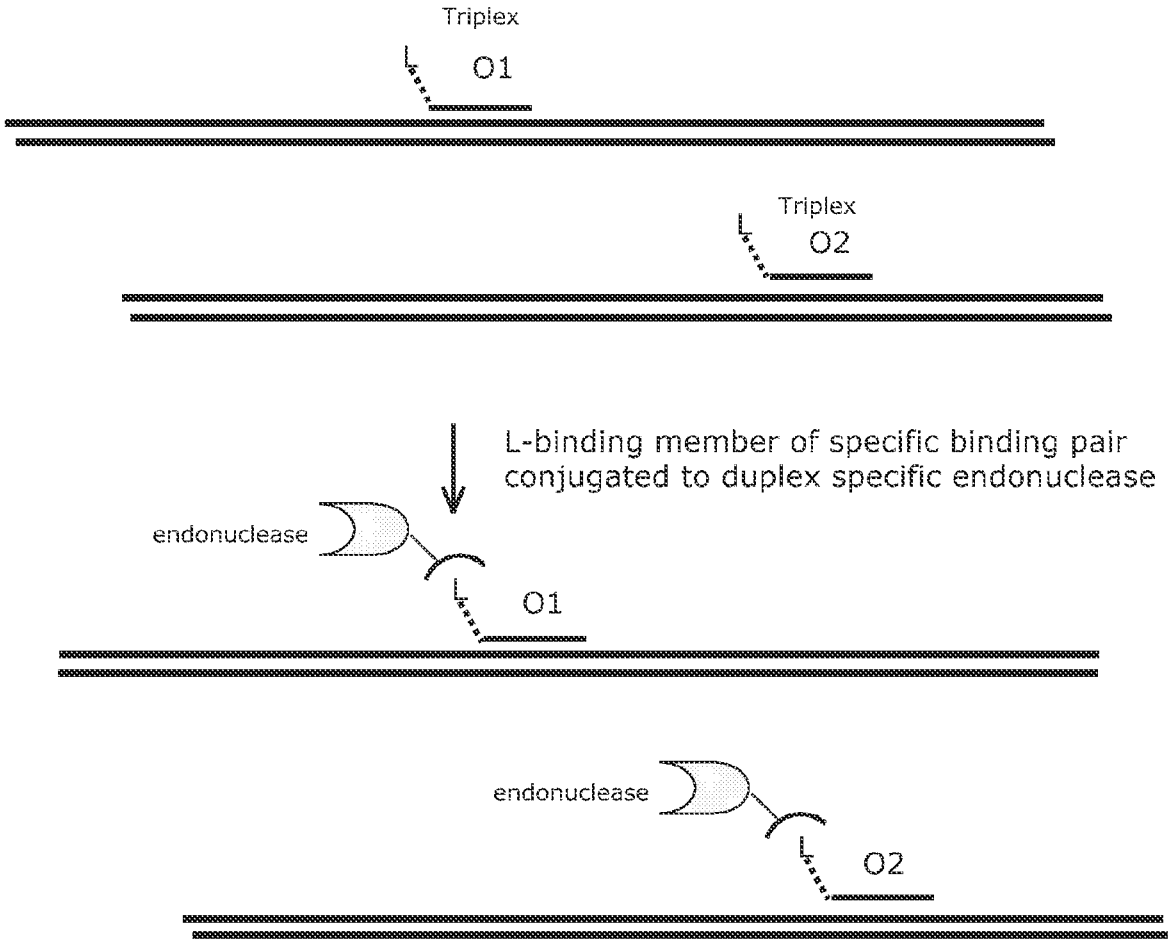
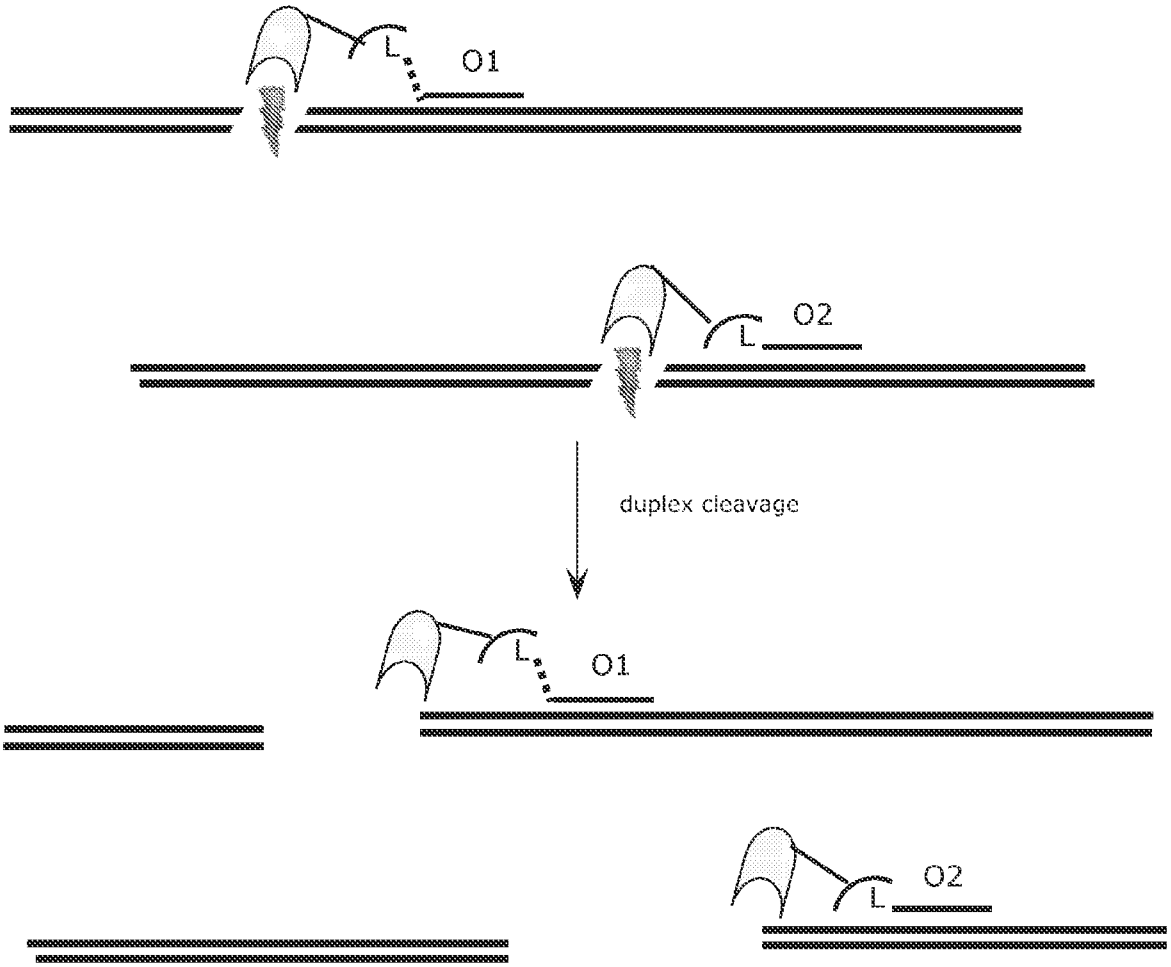


Figure 9: Selective cleavage of dsDNA target sequence regions of interest



INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 10/55137

A. CLASSIFICATION OF SUBJECT MATTER
IPC(8) - C12Q 1/68; C12P 19/34 (2011.01)
USPC - 435/6; 435/91.1
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
USPC: 435/6; 435/91.1

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
USPC: 435/6, 91.1, 91.2, 91.5
(keyword limited; terms below)

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
PubWEST (PGPB,USPT,USOC,EPAB,JPAB); Google; PubMed
Search terms: genome partitioning, complex DNA, restriction, enzyme, endonuclease, tethered, linked, biotin, streptavidin, single primer isothermal amplification, partial triplex, primer, covalent

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X --- Y	US 2003/0175780 A1 (JONES) 18 September 2003 (18.09.2003) para [0009]-[0010], [0014]-[0015], [0054], [0058], [0080], [0090], [0099], [0101], [0103], [0105]; Figs. 1-4	1-11, 13-17, 19-27 --- 12, 18
Y	FUJIWARA et al. Direct probing: covalent attachment of probe DNA to double-stranded target DNA. Nucleic Acids Res. 15 December 1998 (15.12.1998), Vol. 26, No. 24, pages 5728-5733; abstract	12
Y	US 2002/0058270 A1 (KURN) 16 May 2002 (16.05.2002) para [0044], [0046]-[0048]	18
A	WO 2000/55364 A2 (REEVE et al.) 21 September 2000 (21.09.2000)	1-27
A	US 2002/0164634 A1 (PATIL et al.) 7 November 2002 (07.11.2002)	1-27
A	US 2006/0035274 A1 (DONG) 16 February 2006 (16.02.2006)	1-27
A	US 2006/0281082 A1 (ZHU) 14 December 2006 (14.12.2006)	1-27

☐ Further documents are listed in the continuation of Box C. ☐

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

Date of the actual completion of the international search	Date of mailing of the international search report
9 February 2011 (09.02.2011)	24 FEB 2011
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-3201	Authorized officer: Lee W. Young PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

Form PCT/ISA/210 (second sheet) (July 2009)

IDT Exhibit 1020.3133

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 10/55137

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

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

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

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

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:
This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees must be paid.

Group I claims 1-27 are directed to a method for selectively partitioning a plurality of target regions of interest from complex DNA.

Group II claims 28-51 are directed to a method of preparing amplification-ready selectively targeted regions of interest from double-stranded DNA.

Group III claims 52-60 are directed to a DNA complex.

Group IV claims 61-76 are directed to a kit comprising one or more oligonucleotides, each coupled to a ligand; and a ligand-binding component that selectively binds to said ligand, wherein said ligand binding component is coupled to a nucleic acid-modifying enzyme.

----------(Continued on Extra Sheet)----------

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

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

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
1-27

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.

IDT Exhibit 1020.3134

Continuation of Box III - Observations where unity of invention is lacking:

Group V claim 77 is directed to a plurality of amplification-ready partial DNA duplexes generated from genomic DNA, each duplex comprising at least one DNA strand comprising the sequence of a target region of interest wherein the duplex is further ligated to an adaptor for single primer isothermal amplification.

Groups I-IV have the general shared technical features of a DNA comprising at least one DNA strand, an oligonucleotide hybridized to the DNA wherein the oligonucleotide is coupled to a ligand and a ligand-binding component coupled to a nucleic acid-modifying enzyme, wherein the ligand and the ligand-biding component are further coupled to each other. (Please note that it is assumed that claim 1 sub(b) is in partial error, omitting reference to a ligand to which the nucleic-acid modifying enzyme is tethered.). However, this is not an improvment over the prior art of US 2003/0175780 A1 to Jones that teaches DNA comprising at least one DNA strand (paras [0016]-[0029] indicating Figs. 1-4]), an oligonucleotide hybridized to the DNA wherein the oligonucleotide is coupled to a nucleic acid-modifying enzyme (Figs. 1-4; para [0099], [0101] - "A variety of enzymes have been covalently attached to oligonucleotides....Use of a double-stranded recognition domain with the class-IIS restriction endonuclease attached to it could be used to target cutting to the cleavage domain adjacent to the ligated adaptor's recognition domain, so long as buffer conditions during the prior ligation do not permit cutting.") wherein the ligand and the ligand-biding component are further coupled to each other (para [0080], [0090] biotin/streptavidin binding component). Jones further teaches kits of said composition and instructions for practicing various embodiments of the disclosure (para [0110]). While Jones does not specifically teach that that the restriction endonuclease is attached to the oligonucleotide via a ligand coupled to the olionucleotide and a ligand-binding component coupled to the enzyme, Jones does disclose the requirement that conditions be impermissible for cutting during a prior ligation step (para [0101]), and further teaches that such impermissible conditions may be achieved via impermissible buffer conditions (para [0101]). It would have been obvious that modifying the buffer conditions might not be desirable, and that achieving conditions impermissible for cutting could be achieved via other means, including physical separation of the oligonucleotide and the enzyme during the ligation step, with subsequent targeting of the enzyme to the oligonucleotide at the time cutting is desired. Such targeting could be achieved via the well-known biotin/streptavidin ligand/ligand-binding component, which is taught by Jones as a means to bind the oligonucleotide to another component (para [0080], [0090]). It thus would have been obvious to one of skill in the art to attach the ligand (biotin) to the oligonucleotide in the disclosed DNA complex, to attach the ligand-binding component (streptavidin) to the restriction enzyme, and to assemble a DNA complex as claimed, wherein the restriction enzyme is attached to the oligonucleotide via the ligand coupled to the olionucleotide and the ligand-binding component coupled to the enzyme, and the ligand and ligand-binding component further coupled to each other. Apart from the general and well known shared technical feature of at least one DNA strand shared with Groups I-IV as disclosed above by Jones (paras [0016]-[0029] indicating Figs. 1-4]) and an amplification ready target shared with Group II as also disclosed by Jones (para [0022]), Group V has no other shared technical feature with Groups I-IV. Therefore, Groups I-V are not so linked as to form a single general inventive concept under PCT Rule 13.1. The first Group to be searched will be Group I, claims 1-27. In order for all inventions to be examined, the appropriate additional examination fees must be paid.

IDT Exhibit 1020.3135

Espacenet

Bibliographic data: WO2013177220 (A1) — 2013-11-28

METHODS OF SAMPLE PREPARATION

Inventor(s): HEAD STEVEN ROBERT [US]; ORDOUKHANIAN PHILLIP T [US]; SALOMON DANIEL R [US] ± (HEAD, STEVEN, ROBERT, ; ORDOUKHANIAN, PHILLIP, T, ; SALOMON, DANIEL, R)

Applicant(s): SCRIPPS RESEARCH INST [US]; HEAD STEVEN ROBERT [US]; ORDOUKHANIAN PHILLIP T [US]; SALOMON DANIEL R [US] ± (THE SCRIPPS RESEARCH INSTITUTE, ; HEAD, STEVEN, ROBERT, ; ORDOUKHANIAN, PHILLIP, T, ; SALOMON, DANIEL, R)

Classification: **- international:** ***C12P19/34; C12Q1/68; G06F19/18***
- cooperative: **B01J19/0046**; **C12Q1/6874**; B01J2219/0059; B01J2219/00709; B01J2219/00722; C40B40/06 more

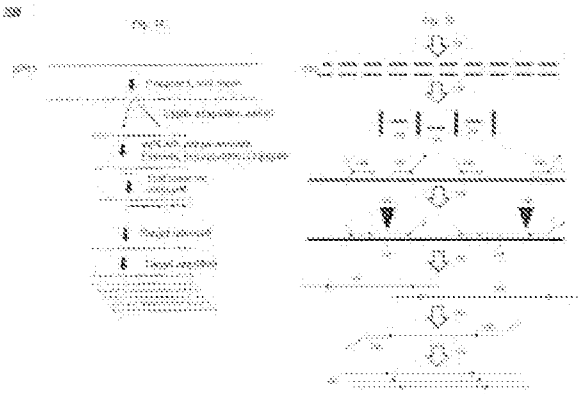
Application number: WO2013US42106 20130521 Global Dossier

Priority number(s): US201261649836P 20120521 ; US201261654389P 20120601 ; US201261716378P 20121019 ; US201361749871P 20130107 ; US201361763424P 20130211 ; US201361763441P 20130211

Also published as: AU2013266394 (A1) CA2874413 (A1) CN104736722 (A) EP2852687 (A1) EP2852687 (A4) more

Abstract of WO2013177220 (A1)

The present disclosure provides methods, compositions, and kits for methods that can improve techniques nucleic acid analysis, and can allow for more reliable and accurate targeted, multiplexed, high throughput sequencing. The methods, compositions, and kits can be used for sequencing target loci of nucleic acid. The methods, compositions, and kits disclosed herein can be used for assisted de novo targeted sequencing. The methods, compositions, and kits disclosed herein can also be used for library labeling for de novo sequencing and phasing.



IDT Exhibit 1020.3136

IDT Exhibit 1020.3137

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
28 November 2013 (28.11.2013)

WIPO | PCT

(10) International Publication Number
WO 2013/177220 A1

(51) International Patent Classification:
C12Q 1/68 (2006.01) *G06F 19/18* (2011.01)
C12P 19/34 (2006.01)

(21) International Application Number:
PCT/US2013/042106

(22) International Filing Date:
21 May 2013 (21.05.2013)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

61/649,836	21 May 2012 (21.05.2012)	US
61/654,389	1 June 2012 (01.06.2012)	US
61/716,378	19 October 2012 (19.10.2012)	US
61/749,871	7 January 2013 (07.01.2013)	US
61/763,424	11 February 2013 (11.02.2013)	US
61/763,441	11 February 2013 (11.02.2013)	US

(71) Applicant *(for all designated States except US)*: **THE SCRIPPS RESEARCH INSTITUTE** [US/US]; 10550 North Torrey Pines Road, La Jolla, CA 92037 (US).

(72) Inventors; and

(71) Applicants *(for US only)*: **HEAD, Steven, Robert** [US/US]; 11673 Valle Vista Road, Lakeside, CA 92040 (US). **ORDOUKHANIAN, Phillip, T.** [US/US]; 6557 Thornwood Street, San Diego, CA 92111 (US). **SALOMON, Daniel, R.** [US/US]; 4728 Reedley Terrace, San Diego, CA 92130 (US).

(74) Agents: **EASTERDAY, Mathew** et al.; Wilson Sonsini Goodrich & Rosati, 650 Page Mill Road, Palo Alto, CA 94304-1050 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

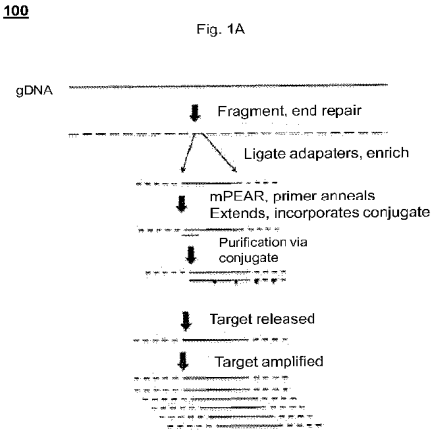
(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

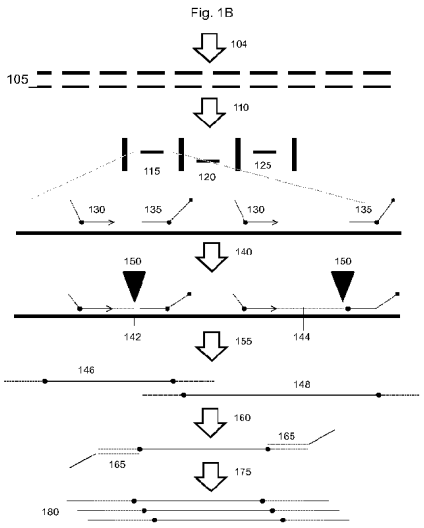
Published:

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

[Continued on next page]

(54) **Title**: METHODS OF SAMPLE PREPARATION





(57) **Abstract**: The present disclosure provides methods, compositions, and kits for methods that can improve techniques nucleic acid analysis, and can allow for more reliable and accurate targeted, multiplexed, high throughput sequencing. The methods, compositions, and kits can be used for sequencing target loci of nucleic acid. The methods, compositions, and kits disclosed herein can be used for assisted de novo targeted sequencing. The methods, compositions, and kits disclosed herein can also be used for library labeling for de novo sequencing and phasing.

WO 2013/177220 A1

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

METHODS OF SAMPLE PREPARATION

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] The present application is a non-provisional of, and claims the benefit of US Provisional Patent Application No. 61/763,441 (Attorney Docket No. 44013-703.101) filed on February 11, 2013, US Provisional Patent Application No. 61/763,424 (Attorney Docket No. 44013-701.105) filed on February 11, 2013, US Provisional Patent Application No. 61/749,871 (Attorney Docket No. 44013-701.104) filed on January 7, 2013, US Provisional Patent Application No. 61/716,378 (Attorney Docket No. 44013-701.103) filed on October 19, 2012, US Provisional Patent Application No. 61/654,389 (Attorney Docket No. 44013-701.102) filed on June 1, 2012, and US Provisional Patent Application No. 61/649,836 (Attorney Docket No. 44013-701.101) filed on May 21, 2012, the entire contents of each of which are incorporated herein by reference.

BACKGROUND OF THE INVENTION

[0002] Several biological applications involve nucleic acid sequencing, including next-generation sequencing. Next-generation sequencing can amplify clonal errors. Additionally, data-analysis of next generation sequencing can require the use of a reference genome. Methods for targeting specific regions of a genome for sequencing analysis are needed.

[0003] Current next generation sequencing platforms pose a problem for genome assembly and sequencer read alignment for longer sequences. Regions of repetitive sequence, homologous sequence, and variable sequence are not reliably mapped. A strategy of aligning to a reference is required for these short read sequencers, and this aligning step can dramatically increase bias and the computational steps required to obtain reliable sequencing results for longer read lengths. There is a need in the art for improved methods and systems for targeting sequences of interest and preparing them for sequencing reactions of a longer read length.

SUMMARY OF THE INVENTION

[0004] In some embodiments, the disclosure provides for a method comprising spatially separating nucleic acid fragments of a nucleic acid; generating one or more amplicons, wherein the amplicons are generated by hybridizing a primer and a probe to a common strand of the nucleic acid fragment; performing a primer extension reaction; ligating a product of the primer extension reaction with the probe to form the amplicons; associating the one or more amplicons with an identifier; and obtaining a sequence of the amplicon.

[0005] In some embodiments, amplicons may be amplified. Amplification of amplicons may be performed by linear amplification, non-linear amplification or rolling circle amplification.

IDT Exhibit 1020.3140

[0006] In some embodiments, at least one amplicon is ligated to at least one different amplicon. In some cases, an indentifier may comprise a molecular barcode, a nucleic acid sequence, a nucleic acid not A, T, C or G. In some cases, the identifier is located at the 5' end of the amplicon. In some cases, the identifier is located at the 3' end of the amplicons. In some cases, the identifier is associated with the amplicon through amplification of the amplicon.

[0007] In some embodiments, the nucleic acid is selected from a group consisting of: DNA, RNA, cDNA and genomic DNA. In some cases, the nucleic acid is fragmented by method from a group consisting of: sonication, enzymatic digestion, heat, exposure to UV light, repetitive pipetting, and nebulization.

[0008] In some embodiments, spatially separating the nucleic acid fragments is performed in partitions. In some cases, spatially separating the nucleic acid fragments is performed by tethering the nucleic acid to a solid or semi-solid support. In some cases the nucleic acid fragment is hybridized to a primer tethered to a solid or semi-solid support, wherein the primer comprises the identifier. In some cases the nucleic acid fragment is hybridized to a probe tethered to a solid or semi-solid support, wherein the probe comprises the identifier. In some cases, the solid or semi-solid support is addressed.

[0009] In some embodiments the probe comprises the identifier. In some cases the probe comprises one or more adapter sequences. In some cases the probe hybridizes to a target nucleic acid. In some cases the probe comprises degenerate sequence. In some cases the probe comprises a synthetic nucleotide. In some cases the probe comprises a primer.

[0010] In some embodiments the primer comprises the identifier. In some cases the primer comprises one or more adapter sequences. In some cases the primer hybridizes to a target nucleic acid. In some cases the primer comprises degenerate sequence. In some cases the primer comprises a synthetic nucleotide. In some cases the amplicons are associated with a unique identifier. In some cases the identifier represents an individual partition.

[0011] In some embodiments the sequence of the amplicons is performed by massively parallel sequencing. In some cases a computing device is used to generate a consensus sequence of all or part of the nucleic acid fragment, from sequence reads comprising the identifier.

[0012] In some embodiments a computing device is used to generate a consensus sequence of all or part of the nucleic acid, from sequence reads comprising the identifier. In some cases a computing device is used to generate a consensus sequence without comparing the consensus sequence to a reference.

[0013] In some embodiments the consensus sequence has at least 1X , 5X, 10X, or 50X depth of coverage.

[0014] In some embodiments method of the disclosure provides for multiplexed analysis for multiple nucleic acid fragments. In some cases, the method is multiplexed for at least 2 nucleic acid fragments, at least 10 nucleic acid fragments, at least 100 nucleic acid fragments, at least 10000

IDT Exhibit 1020.3141

nucleic acid fragments, at least 100000 nucleic acid fragments, or at least 1000000 nucleic acid fragments.

[0015] In some embodiments the product of the primer extension reaction is at least 100 nucleotides, at least 1000 nucleotides, or at least 10000 nucleotides.

[0016] In some embodiments, the disclosure provides for transmitting sequencing data generated, receiving sequencing data generated, storing sequencing data generated, comprising comparing or analyzing sequencing data generated, transmitting a report related to sequencing data generated, receiving a report related to sequencing data generated, storing a report related to sequencing data generated, storing a report related to sequencing data generated, comparing or analyzing a report related to sequencing data generated by methods of this disclosure. In some cases, the disclosure provides for transforming sequencing data to a report related to sequencing data using a computing device comprising non transitory computer readable media.

[0017] In some embodiments, the primer or probe is specific to one or more regions of the nucleic acid fragment. In some cases, the primer or probe is at least 50% complementary to one or more regions of the nucleic acid fragment. In some cases, the primer or probe is at least 75% complementary to one or more regions of the nucleic acid fragment. In some cases, the primer or probe is at least 90% complementary to one or more regions of the nucleic acid fragment.

[0018] In some embodiments, one or more amplicons are linked to form a contiguous sequence. In some cases the disclosure provides for performing a primer extension reaction comprises addition of a strand displacing polymerase. In some cases the disclosure provides for performing a primer extension reaction to form a primer extension product, wherein the primer extension product comprises an affinity conjugate and wherein the primer extension product comprises a target sequence.

[0019] In some embodiments the disclosure provides for performing a primer extension reaction to form a primer extension product and performing affinity purification of the primer extension product using the affinity conjugate. In some cases the affinity conjugate is biotin. In some cases the affinity purification of the primer extension product is performed using streptavidin.

[0020] In some embodiments, the disclosure provides for a method comprising: obtaining a nucleic acid, wherein the nucleic acid comprises a target sequence; hybridizing a TELA primer and a TELA probe to a common strand of the nucleic acid; performing a primer extension reaction; ligating a product of the primer extension reaction with the TELA probe to form a ligation product comprising the target sequence; and sequencing the target sequence.

[0021] In some embodiments the target sequence is at least 30% of the ligation product. In some cases, one or more ligation products are linked to form a contiguous sequence.

[0022] In some embodiments, the disclosure provides for a method comprising: obtaining a nucleic acid library; ligating an adapter sequence to one or more nucleic acids of the nucleic acid library;

IDT Exhibit 1020.3142

hybridizing a primer to the adapter sequence, wherein the primer comprises a spacer region and a locus specific region; performing a primer extension reaction to form a primer extension product, wherein the primer extension product comprises an affinity conjugate and wherein the primer extension product comprises a target sequence; performing affinity purification of the primer extension product using the affinity conjugate.

[0023] In some embodiments the nucleic acid library is fragmented gDNA. In some cases, the nucleic acid library is expressed sequences. In some cases, the nucleic acid library is epigenetically sorted. In some embodiments the primer extension product is sequenced.

[0024] In some cases the nucleic acid library comprises at least 2, at least 10, at least 100, at least 10000, at least 100000, or at least 1000000 nucleic acid fragments.

[0025] In some cases, the nucleic acid is fragmented by method from a group consisting of: sonication, enzymatic digestion, heat, exposure to UV light, repetitive pipetting, and nebulization. The method of claim 1 further comprising amplifying a nucleic acid to generate the nucleic acid library.

[0026] In some embodiments, the disclosure provides for methods comprising amplifying the adaptor ligated nucleic acid library. In some cases, the amplifying is linear. In some cases, the amplifying is performed by rolling circle amplification. In some cases, the amplifying is non-linear. In some cases the primer extension product comprises an identifier.

[0027] In some embodiments the primer extension product comprises a molecular barcode. In some cases, the primer extension product comprises a nucleic acid sequence. In some cases, the primer extension product comprises a nucleic acid not A, T, C or G. In some cases the adaptor is located at the 5' end of the nucleic acid of the nucleic acid library. In some cases, the primer extension product is located at the 3' end of the nucleic acid of the nucleic acid library. In some cases the nucleic acid is selected from a group consisting of: DNA, RNA, cDNA and genomic DNA.

[0028] In some embodiments, the disclosure provides for methods in which the spacer region is degenerate or random sequence. In some cases, the spacer region comprises at least 1 nucleotide, at least 10 nucleotides, or 100 nucleotides.

[0029] In some cases, the spacer region comprises a molecular barcode. In some cases, the spacer region comprises a nucleic acid sequence. In some cases, the spacer region comprises a nucleic acid not A, T, C or G. In some cases,the spacer region comprises an enzymatic target sequence.

[0030] In some embodiments, the locus specific region is at least 50% complementary to a locus of the nucleic acid. In some cases, the locus specific region is at least 70% complementary to a locus of the nucleic acid. In some cases,the locus specific region is at least 80% complementary to a locus of the nucleic acid. In some cases, the locus specific region is at least 90% complementary to a locus of the nucleic acid. In some cases, the locus specific region is at least 99% complementary to a locus of the nucleic acid. In some cases, the locus specific region binds a locus of the nucleic acid upstream of the target sequence.

IDT Exhibit 1020.3143

[0031] In some embodiments, the affinity conjugate is biotin. In some cases, the affinity purification of the primer extension product is performed using streptavidin.

[0032] In some embodiments, the disclosure provides for sequencing the primer extension product. In some cases the identifier represents a source of the sample from which the nucleic acid library was generated.

[0033] In some cases the sequence of the primer extension reaction is obtained by performing massively parallel sequencing. In some cases a computing device is used to generate a consensus sequence of all or part of the target sequence, from sequence reads comprising the identifier.

[0034] In some embodiments the consensus sequence has at least 1X, 5X, 10X, or 50X depth of coverage.

[0035] In some embodiments the disclosure provides for a method that is multiplexed for multiple samples or target sequences. In some cases the method is multiplexed for at least 2, 10, 100, 10000, 100000, 1000000, 1000000 samples or target sequences.

[0036] In some embodiments the product of the primer extension reaction is at least 100 nucleotides, at least 1000 nucleotides, or at least 10000 nucleotides.

[0037] In some embodiments, one or more primer extension products are linked to form a contiguous sequence.

INCORPORATION BY REFERENCE

[0038] All publications, patents, and patent applications mentioned in this specification are herein incorporated by reference to the same extent as if each individual publication, patent, or patent application was specifically and individually indicated to be incorporated by reference.

BRIEF DESCRIPTION OF THE DRAWINGS

[0039] The novel features of the invention are set forth with particularity in the appended claims. A better understanding of the features and advantages of the present invention will be obtained by reference to the following detailed description that sets forth illustrative embodiments, in which the principles of the invention are utilized, and the accompanying drawings of which:

[0040] **FIG. 1A** depicts a schematic representation of a workflow involving mPEAR.

[0041] **FIG. 1B** depicts a schematic representation of a workflow involving targeted sequencing.

[0042] **FIG. 2A** depicts examples of a donor primer (D).

[0043] **FIG. 2B** depicts examples of an acceptor probe (A).

[0044] **FIG. 3A** depicts a representation of a general TELA or mPEAR extension reaction.

[0045] **FIG. 3B** depicts a representation of an affinity purification reaction of TELA or mPEAR labeled products.

IDT Exhibit 1020.3144

[0046] **FIG. 3C** depicts a representation of release of TELA or mPEAR labeled products from an immobilized surface.

[0047] **FIG. 4** depicts a representation of RAPEL methods.

[0048] **FIG. 5** depicts a representation of use of sequencer platform adapter with targeted sequencing products.

[0049] **FIG. 6** depicts multiple target sequences flanked by adapter sequences and molecular labels or bar codes.

[0050] **FIG. 7** depicts a mPEAR or TELA primer and a target sequence.

[0051] **FIG. 8** depicts the attachment of adapater sequences and barcodes for de novo sequencing of a target sequence.

[0052] **FIG. 9** depicts an example read structure of sequencing generated from mPEAR, TELA, RAPEL or other products of the methods and compositions of this disclosure.

[0053] **FIG. 10** depicts multiple adapter sequences, barcodes and primer sites within a mPEAR, TELA, RAPEL or other products of the methods and compositions of this disclosure.

[0054] **FIG. 11** depicts a computer readable stored media platform and example means for transmission of data s generated by the composition and methods of this disclosure

[0055] **FIG. 12** depicts a schematic representation of compositions and methods of the disclosure using rolling circle amplification to assemble multiple contiguous sequence targets.

[0056] **FIG. 13** depicts a schematic of physical separation of target sequences and de novo assembly of sequences based on bar code sequences.

[0057] **FIG. 14** depicts examples of sequences, including bar codes and adapter sequences.

DETAILED DESCRIPTION OF THE INVENTION

[0058] The present disclosure provides methods, compositions, and kits for methods that can improve techniques nucleic acid analysis, and can allow for more reliable and accurate targeted, multiplexed, high throughput sequencing. The methods, compositions, and kits can be used for sequencing target loci of nucleic acid. The methods, compositions, and kits disclosed herein can be used for assisted de novo targeted sequencing. The methods, compositions, and kits disclosed herein can also be used for library DNA/RNA labeling for true de novo sequencing and phasing.

I. Definitions

[0059] "Affinity conjugate" as described herein provides use of specific interactions between two molecules for the purification of a target molecule. An affinity ligand having affinity for a target molecule may be attached to an insoluble support and functions as bait for capturing a target molecule. The target molecule may be covalently or non covalently to a conjugate molecule that

IDT Exhibit 1020.3145

interacts or binds to the affinity ligand. The affinity ligand can be any molecule that will bind the target without also binding other molecules in the solution.

[0060] "Amplified nucleic acid" or "amplified polynucleotide" can be any nucleic acid or poly nucleotide molecule whose amount has been increased at least two fold by any nucleic acid amplification or replication method performed in vitro as compared to its starting amount. For example, an amplified nucleic acid can be obtained from a polymerase chain reaction (PCR) which can, in some instances, amplify DNA in an exponential manner (e.g. 2^n). Amplified nucleic acid can also be obtained from a linear amplification. Amplified nucleic acid can be obtained by primer elongations. identify

[0061] "Amplification product" can refer to a product resulting from an amplification reaction.

[0062] An "amplicon" can be a polynucleotide or nucleic acid that is the source and/or product of natural or artificial amplification or replication events.

[0063] The term "biological sample" or "sample" generally refers to a sample or part isolated from a biological entity. The biological sample may show the nature of the whole and examples include, without limitation, bodily fluids, dissociated tumor specimens, cultured cells, and any combination thereof. Biological samples can come from one or more individuals. One or more biological samples can come from the same individual. One non limiting example would be if one sample came from an individual's blood and a second sample came from an individual's tumor biopsy. Examples of biological samples can include but are not limited to, blood, serum, plasma, nasal swab or nasopharyngeal wash, saliva, urine, gastric fluid, spinal fluid, tears, stool, mucus, sweat, earwax, oil, glandular secretion, cerebral spinal fluid, tissue, semen, vaginal fluid, interstitial fluids, including interstitial fluids derived from tumor tissue, ocular fluids, spinal fluid, throat swab, breath, hair, finger nails, skin, biopsy, placental fluid, amniotic fluid, cord blood, emphatic fluids, cavity fluids, sputum, pus, micropiota, meconium, breast milk and/or other excretions. The samples may include nasopharyngeal wash. Examples of tissue samples of the subject may include but are not limited to, connective tissue, muscle tissue, nervous tissue, epithelial tissue, cartilage, cancerous or tumor sample, or bone. The sample may be provided from a human or animal. The sample may be provided from a mammal, vertebrate, such as murines, simians, humans, farm animals, sport animals, or pets. The sample may be collected from a living or dead subject. The sample may be collected fresh from a subject or may have undergone some form of pre-processing, storage, or transport.

[0064] "Bodily fluid" generally can describe a fluid or secretion originating from the body of a subject. In some instances, bodily fluids can be a mixture of more than one type of bodily fluid mixed together. Some non limiting examples of bodily fluids can be: blood, urine, bone marrow, spinal fluid, pleural fluid, lymphatic fluid, amniotic fluid, ascites, sputum, or a combination thereof.

[0065] "Complementary" or "complementarity" can refer to nucleic acid molecules that are related by base-pairing. Complementary nucleotides are, generally, A and T (or A and U), or C and G. Two single stranded RNA or DNA molecules are said to be substantially complementary when the

IDT Exhibit 1020.3146

nucleotides of one strand, optimally aligned and with appropriate nucleotide insertions or deletions, pair with at least about 90% to about 95% complementarity, and more preferably from about 98% to about 100% complementarity, and even more preferably with 100% complementarity. Alternatively, substantial complementarity exists when an RNA or DNA strand will hybridize under selective hybridization conditions to its complement. Selective hybridization conditions include, but are not limited to, stringent hybridization conditions. Hybridization temperatures are generally at least about 2° C to about 6° C lower than melting temperatures (T_m).

[0066] A "barcode" or "molecular barcode" can be a material for labeling. The barcode can label a molecule such as a nucleic acid or a polypeptide. The material for labeling can be associated with information. A barcode can be called a sequence identifier (i.e. a sequence-based barcode or sequence index). A barcode can be a particular nucleotide sequence. A barcode can be used as an identifier. A barcode can be a different size molecule or different ending points of the same molecule. Barcodes can include a specific sequence within the molecule and a different ending sequence. For example, a molecule that is amplified from the same primer and has 25 nucleotide positions is different than a molecule that is amplified and has 27 nucleotide positions. The addition positions in the 27mer sequence can be considered a barcode. A barcode can be incorporated into a polynucleotide. A barcode can be incorporated into a polynucleotide by many methods. Some non-limiting methods for incorporating a barcode can include molecular biology methods. Some non-limiting examples of molecular biology methods to incorporate a barcode are through primers (e.g. tailed primer elongation), probes (i.e. elongation with ligation to a probe), or ligation (i.e. ligation of known sequence to a molecule).

[0067] A barcode can be incorporated into any region of a polynucleotide. The region can be known. The region can be unknown. The barcode can be added to any position along the polynucleotide. The barcode can be added to the 5' end of a polynucleotide. The barcode can be added to the 3' end of the polynucleotide. The barcode can be added in between the 5' and 3' end of a polynucleotide. A barcode can be added with one or more other known sequences. One non limiting example is the addition of a barcode with a sequence adapter.

[0068] Barcodes can be associated with information. Some non-limiting examples of the type of information a barcode can be associated with information include: the source of a sample; the orientation of a sample; the region or container a sample was processed in; the adjacent polynucleotide; or any combination thereof.

[0069] In some cases, barcodes can be made from combinations of sequences (different from combinatorial barcoding)and can be used to identify a sample or a genomic coordinate and a different template molecule or single strand the molecular label and copy of the strand was obtained from. In some cases a sample identifier, a genomic coordinate and a specific label for each biological molecule may be amplified together.

IDT Exhibit 1020.3147

[0070] Barcodes can be added before pooling of samples. When the sequences are determined of the pooled samples, the barcode can be sequenced along with the rest of the polynucleotide. The barcode can be used to associate the sequenced fragment with the source of the sample.

[0071] Barcodes can also be used to identify the orientation of a sample. One or more barcodes can be used together. Two or more barcodes can be adjacent to one another, not adjacent to one another, or any combination thereof.

[0072] Barcodes can be used for combinatorial labeling.

[0073] "Combinatorial labeling" can be a method by which two or more barcodes are used to label. The two or more barcodes can label a polynucleotide. The barcodes, each, alone can be associated with information. The combination of the barcodes together can be associated with information. In some cases a combination of barcodes can be used together to determine in a randomly amplified molecule that the amplification occurred from the original sample template and not a synthetic copy of that template

[0074] "Degenerate" can refer to a nucleic acid or nucleic acid region that is comprised of random bases. The terms "degenerate" and "random" can be used interchangeably when referring to nucleic acid sequences (e.g. "degenerate primers" or "random primers" or "degenerate probes" or "random probes"). The degenerate region can be of variable length. The degenerate region can comprise some portion of the whole nucleic acid (e.g. a semi-degenerate primer). The degenerate region can comprise the whole nucleic acid (e.g. a "degenerate primer"). A degenerate nucleic acid mix, or semi-degenerate nucleic acid mix may be comprised of every possible combination of base pairs, less than every possible combination of base pairs, or some combination of base pairs, a few combinations of base pairs, or a single base pair combination. A degenerate primer mix, or semi-degenerate primer mix can comprise

[0075] "Double stranded" can refer to two polynucleotide strands that have annealed through complementary base-pairing.

[0076] "Known oligonucleotide sequence" or "known oligonucleotide" or "known sequence" can refer to a polynucleotide sequence that is known. A known oligonucleotide sequence can correspond to an oligonucleotide that has been designed, e.g. a universal primer for next generation sequencing platforms (e.g., Illumina, 454), a probe, an adaptor, a tag, a primer, a molecular barcode sequence, an identifier. A known sequence can comprise part of a primer. A known oligonucleotide sequence may not actually be known by a particular user but can be constructively known, for example, by being stored as data which may be accessible by a computer. A known sequence may also be a trade secret that is actually unknown or a secret to one or more users but may be known by the entity who has designed a particular component of the experiment, kit, apparatus or software that the user is using.

[0077] "Library" can refer a collection of nucleic acid. A library can contain one or more target fragments. In some instances the target fragments can be amplified nucleic acids. In other instances, the target fragments can be nucleic acid that is not amplified. A library can contain nucleic acid that

IDT Exhibit 1020.3148

has one or more known oligonucleotide sequence(s) added to the 3' end, the 5' end or both the 3' and 5' end. The library may be prepared so that the fragments can contain a known oligonucleotide sequence that identifies the source of the library (e.g a molecular identification barcode identifying a patient or DNA source). In some instances, two or more libraries can be pooled to create a library pool. Libraries may also be generated with other kits and techniques such as transpon mediated labeling, or "fragmentation" as known in the art. Kits may be commercially available such as the Illumina Nextera kit.

[0078] "Locus specific" or "loci specific" can refer to one or more loci corresponding to a location in a nucleic acid molecule (e.g. a location within a chromosome or genome). In some instances, a loci can be associated with genotype. In some instances loci may be directly isolated and enriched from the sample, e.g., based on hybridization and/or other sequence-based techniques, or they may be selectively amplified using the sample as a template prior to detection of the sequence. In some instances, loci may be selected on the basis of DNA level variation between individuals, based upon specificity for a particular chromosome, based on CG content and/or required amplification conditions of the selected loci, or other characteristics that will be apparent to one skilled in the art upon reading the present disclosure. A locus may also refer to a specific genomic coordinate or location in a genome as denoted by the reference sequence of that genome.

[0079] "Long nucleic acid" can refer to a polynucleotide longer than 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 kilobases.

[0080] The term "melting temperature" or "T_m" commonly refers to the temperature at which a population of double-stranded nucleic acid molecules becomes half dissociated into single strands. Equations for calculating the T_m of nucleic acids are well known in the art. One equation that gives a simple estimate of the T_m value is as follows: T_m=81.5+16.6(log 10[Na^+])0.41(%[G+C])−675/n−1.0 m, when a nucleic acid is in aqueous solution having cation concentrations of 0.5 M or less, the (G+C) content is between 30% and 70%, n is the number of bases, and m is the percentage of base pair mismatches (see, e.g., Sambrook J et al., *Molecular Cloning, A Laboratory Manual,* 3rd Ed., Cold Spring Harbor Laboratory Press (2001)). Other references can include more sophisticated computations, which take structural as well as sequence characteristics into account for the calculation of T_m.

[0081] "Nucleotide" can refer to a base-sugar-phosphate combination. Nucleotides are monomeric units of a nucleic acid sequence (e.g. DNA and RNA). The term nucleotide includes ribonucleoside triphosphates ATP, UTP, CTG, GTP and deoxyribonucleoside triphosphates such as dATP, dCTP, dITP, dUTP, dGTP, dTTP, or derivatives thereof. Such derivatives can include, for example, [αS]dATP, 7-deaza-dGTP and 7-deaza-dATP, and nucleotide derivatives that confer nuclease resistance on the nucleic acid molecule containing them. The term nucleotide as used herein also refers to dideoxyribonucleoside triphosphates (ddNTPs) and their derivatives. Illustrative examples of dideoxyribonucleoside triphosphates include, but are not limited to, ddATP, ddCTP, ddGTP, ddITP,

IDT Exhibit 1020.3149

and ddTTP. A nucleotide may be unlabeled or detectably labeled by well known techniques. Fluorescent labels and their attachment to oligonucleotides are described in many reviews, including Haugland, *Handbook of Fluorescent Probes and Research Chemicals,* 9th Ed., Molecular Probes, Inc., Eugene Oreg. (2002); Keller and Manak, *DNA Probes,*2nd Ed., Stockton Press, New York (1993); Eckstein, Ed., *Oligonucleotides and Analogues: A Practical Approach*, IRL Press, Oxford (1991); Wetmur, *Critical Reviews in Biochemistry and Molecular Biology,* 26:227-259 (1991); and the like. Other methodologies applicable to the invention are disclosed in the following sample of references: Fung et al., U.S. Pat. No. 4,757,141; Hobbs, Jr., et al., U.S. Pat. No. 5,151,507; Cruickshank, U.S. Pat. No. 5,091,519; Menchen et al., U.S. Pat. No. 5,188,934; Begot et al., U.S. Pat. No. 5,366,860; Lee et al., U.S. Pat. No. 5,847,162; Khanna et al., U.S. Pat. No. 4,318,846; Lee et al., U.S. Pat. No. 5,800,996; Lee et al., U.S. Pat. No. 5,066,580: Mathies et al., U.S. Pat. No. 5,688,648; and the like. Labeling can also be carried out with quantum dots, as disclosed in the following patents and patent publications: U.S. Pat. Nos. 6,322,901; 6,576,291; 6,423,551; 6,251,303; 6,319,426; 6,426,513; 6,444,143; 5,990,479; 6,207,392; 2002/0045045; and 2003/0017264. Detectable labels include, for example, radioactive isotopes, fluorescent labels, chemiluminescent labels, bioluminescent labels and enzyme labels. Fluorescent labels of nucleotides may include but are not limited fluorescein, 5-carboxyfluorescein (FAM), 2′7′-dimethoxy-4′5-dichloro-6-carboxyfluorescein (JOE), rhodamine, 6-carboxyrhodamine (R6G), N,N,N′,N′-tetramethyl-6-carboxyrhodamine (TAMRA), 6-carboxy-X-rhodamine (ROX), 4-(4′dimethylaminophenylazo) benzoic acid (DABCYL), Cascade Blue, Oregon Green, Texas Red, Cyanine and 5-(2′-aminoethyl)aminonaphthalene-1-sulfonic acid (EDANS). Specific examples of fluorescently labeled nucleotides include [R6G]dUTP, [TAMRA]dUTP, [R110]dCTP, [R6G]dCTP, [TAMRA]dCTP, [JOE]ddATP, [R6G]ddATP, [FAM]ddCTP, [R110]ddCTP, [TAMRA]ddGTP, [ROX]ddTTP, [dR6G]ddATP, [dR110]ddCTP, [dTAMRA]ddGTP, and [dROX]ddTTP available from Perkin Elmer, Foster City, Calif. FluoroLink DeoxyNucleotides, FluoroLink Cy3-dCTP, FluoroLink Cy5-dCTP, FluoroLink Fluor X-dCTP, FluoroLink Cy3-dUTP, and FluoroLink Cy5-dUTP available from Amersham, Arlington Heights, Ill.; Fluorescein-15-dATP, Fluorescein-12-dUTP, Tetramethyl-rodamine-6-dUTP, IR770-9-dATP, Fluorescein-12-ddUTP, Fluorescein-12-UTP, and Fluorescein-15-2′-dATP available from Boehringer Mannheim, Indianapolis, Ind.; and Chromosomee Labeled Nucleotides, BODIPY-FL-14-UTP, BODIPY-FL-4-UTP, BODIPY-TMR-14-UTP, BODIPY-TMR-14-dUTP, BODIPY-TR-14-UTP, BODIPY-TR-14-dUTP, Cascade Blue-7-UTP, Cascade Blue-7-dUTP, fluorescein-12-UTP, fluorescein-12-dUTP, Oregon Green 488-5-dUTP, Rhodamine Green-5-UTP, Rhodamine Green-5-dUTP, tetramethylrhodamine-6-UTP, tetramethylrhodamine-6-dUTP, Texas Red-5-UTP, Texas Red-5-dUTP, and Texas Red-12-dUTP available from Molecular Probes, Eugene, Oreg. Nucleotides can also be labeled or marked by chemical modification. A chemically-modified single nucleotide can be, e.g., biotin-dNTP. Some non-limiting examples of biotinylated dNTPs can include, biotin-dATP

IDT Exhibit 1020.3150

(e.g., bio-N6-ddATP, biotin-14-dATP), biotin-dCTP (e.g., biotin-11-dCTP, biotin-14-dCTP), and biotin-dUTP (e.g. biotin-11-dUTP, biotin-16-dUTP, biotin-20-dUTP).

[0082] "Polymerase" can refer to an enzyme that links individual nucleotides together into a strand, using another strand as a template.

[0083] "Polymerase chain reaction" or "PCR" can refer to a technique for replicating a specific piece of selected DNA in vitro, even in the presence of excess non-specific DNA. Primers are added to the selected DNA, where the primers initiate the copying of the selected DNA using nucleotides and, typically, Taq polymerase or the like. By cycling the temperature, the selected DNA is repetitively denatured and copied. A single copy of the selected DNA, even if mixed in with other, random DNA, can be amplified to obtain thousands, millions, or billions of replicates. The polymerase chain reaction can be used to detect and measure very small amounts of DNA and to create customized pieces of DNA.

[0084] The term "polynucleotides" may include but is not limited to various DNA, RNA molecules, derivatives or combination thereof. These may include species such as dNTPs , ddNTPs, DNA, RNA, peptide nucleic acids, cDNA, dsDNA, ssDNA, plasmid DNA, cosmid DNA, chromosomal DNA, genomic DNA, viral DNA, bacterial DNA, mtDNA (mitochondrial DNA), mRNA, rRNA, tRNA, nRNA, siRNA, snRNA, snoRNA, scaRNA, microRNA, dsRNA, ribozyme, riboswitch and viral RNA.

[0085] A "primer" generally refers to an oligonucleotide used to, e.g., prime nucleotide extension, ligation and/or synthesis, such as in the synthesis step of the polymerase chain reaction or in the primer extension techniques used in certain sequencing reactions. A primer may also be used in hybridization techniques as a means to provide complementarity of a locus to a capture oligonucleotide for detection of a specific nucleic acid region.

[0086] "Primer extension product" can refer to the product resulting from an primer extension reaction using a contiguous polynucleotide as a template, and a complementary or partially complementary primer to the contiguous sequence.

[0087] "Sequencing", "sequence determination" and the like generally refers to any and all biochemical methods that may be used to determine the order of nucleotide bases in a nucleic acid.

[0088] "Streptavidin" can refer to a protein or peptide that can bind to biotin and can include: native egg-white avidin, recombinant avidin, deglycosylated forms of avidin, bacterial streptavidin, recombinant streptavidin, truncated streptavidin, and/or any derivative thereof.

[0089] A "subject" generally refers to an organism that is currently living or an organism that at one time was living or an entity with a genome that can replicate. The methods, kits, and/or compositions of the disclosure can be applied to one or more single-celled or multi-cellular subjects, including but not limited to microorganisms such as bacterium and yeast; insects including but not limited to flies, beetles, and bees; plants including but not limited to corn, wheat, seaweed or algae; and animals including, but not limited to: humans; laboratory animals such as mice, rats, monkeys,

IDT Exhibit 1020.3151

and chimpanzees; domestic animals such as dogs and cats, agricultural animals such as cows, horses, pigs, sheep, goats; and wild animals such as bears, pandas, lions, tigers, leopards, elephants, zebras, giraffes, gorillas, dolphins, and whales. The methods of this disclosure can also be applied to germs or infectious agents, such as viruses or virus particles or one or more cells that have been infected by one or more viruses.

[0090] A "support" can be solid, semisolid, a bead, a surface. The support can mobile in a solution or can be immobile.

[0091] Unique identifier can mean a molecular bar code. Can be a percentage of a nucleic acid in a mix, such as dUTP.

II. Polynucleotide

[0092] A polynucleotide molecule can be treated. For example, a polynucleotide can be treated by chemical, physical, and/or enzymatic forces. The nucleic acid can be obtained from a subject or biological specimen. In some embodiments, the nucleic acid is DNA. The DNA can be of genomic origin or a cDNA library generated from a subject's RNA, or cell free DNA.

[0093] In some cases the polynucleotide may represent the entire genetic complement of an organism or subject. The polynucleotide can be genomic DNA molecules from a eukaryote which can include both intron and exon sequences (coding sequence), as well as non-coding regulatory sequences such as promoter and enhancer sequences. In some cases (e.g., DNA), molecules may comprise sub-sets of polynucleotide sequences of genomic DNA, such as, for example, particular chromosomes or fragments of chromosomes. The polynucleotide can be RNA or a combination of RNA and DNA, single stranded or double stranded. Sometimes, the sequence of the primary polynucleotide molecules can be unknown. In some embodiments the polynucleotide molecules are human genomic DNA molecules. In some embodiments, the polynucleotide is not genomic and can be from a mitochondria, a chloroplast, a plasmid a bacterium an/or a virus. In some cases the polynucleotide molecules are chromosomal or genomic DNA molecules from an organism that has been infected by a virus; in some instances, the viral infection may have caused alterations or insertions into the DNA.

[0094] The DNA molecules can be treated chemically or. In some cases polynucleotide molecules may be treated prior to, or subsequent to any fragmentation processes, and prior to or subsequent to the ligation of the adaptor sequences.

[0095] In many cases, fragmentation of polynucleotides, such as through mechanical shearing or enzymatic digestion results in fragments with a heterogeneous mix of blunt and 3′- and 5′- overhanging ends. In some cases, the methods can provide for repair of the fragment ends using methods or kits (e.g. Lucigen DNA terminator End Repair Kit) known in the art to generate ends that are designed for insertion, for example, into blunt sites of cloning vectors. In some cases, the methods

IDT Exhibit 1020.3152

of the disclosure provide for blunt ended fragment ends of the population of nucleic acids sequenced. Further, in some cases, the blunt ended fragment may also be phosphorylated or dephosphorylated to facilitate ligation. The phosphate moiety can be introduced via enzymatic treatment, for example, using a kinase, (e.g. T4 polynucleotide kinase) or the phosphate moiety can be dephosphorylated using an alkaline phosphatase.

[0096] In other cases, polynucleotide sequences are prepared with single overhanging nucleotides by, for example, activity of certain types of DNA polymerase such as Taq polymerase or Klenow exo minus polymerase which has a nontemplate-dependent terminal transferase activity that adds a single deoxynucleotide, for example, deoxyadenosine (A) to the 3′ ends of, for example, PCR products. Such enzymes can be utilized to add a single nucleotide 'A' to the blunt ended 3′ terminus of each strand of the target polynucleotide duplexes. Thus, an 'A' could be added to the 3′ terminus of each end repaired duplex strand of the target polynucleotide duplex by reaction with Taq or Klenow exo minus polymerase, whilst the adaptor polynucleotide construct could be a T-construct with a compatible 'T' overhang present on the 3′ terminus of each duplex region of the adaptor construct. This end modification also prevents self-ligation of both adaptor and target such that there is a bias towards formation of the combined ligated adaptor-target sequences.

[0097] Polynucleotides may be derived from a variety of sources including any species containing genetic material. In some cases samples may be derived from human, mammal, non-human mammal, ape, monkey, chimpanzee, reptilian, amphibian, avian, insect or various invertebrate sources. Samples may also be derived from microorganisms which may include but are not limited to unicellular organisms or multi-cellular organisms, bacteria, parasites, fungi, protists, algae, larvae, nematodes, worms, viruses and any combination thereof.

[0098] Further, samples may be extracted from variety of tissues and tissue types. Polynucleotides may be fetal in origin (e.g., fluid taken from a pregnant subject), or may be derived from tissue of the subject itself. Polynucleotides may also be found as cell free, or in a state not contained within cells. Polynucleotides can be extracted from, for example, a bodily fluid or tissue.

[0099] After collection of tissue or bodily fluids containing polynucleotides, samples may be treated. For example, the nucleic acid can be fragmented, purified, partially purified, and/or mixed with different polynucleotides from different sources, or any combination thereof. The starting material nucleic acid can comprise DNA of known origin or unknown origin, or a combination thereof. The starting material nucleic acid can comprise DNA of known origin and be mixed with DNA of known sequence. In some instances, the DNA of known sequence can act as a control or a sample reference.

[00100] Polynucleotide samples may be treated by any methods herein. Samples may be isolated and extracted using a variety of techniques known in the art. Isolation and purification of polynucleotides may be accomplished using any means, including, but not limited to, the use of commercial kits and protocols provided by companies such as Sigma Aldrich, Life Technologies,

IDT Exhibit 1020.3153

Promega, Affymetrix, IBI or the like. Kits and protocols may also be non-commercially available. In some cases, polynucleotides such as DNA may be isolated, extracted and prepared using commercially available kits such as Qiagen Qiamp® Circulating Nucleic Acid Kit protocol, Qiagen Qubit ™ dsDNA HS Assay kit protocol, Agilent™ DNA 1000 kit, or TruSeq™ Sequencing Library Preparation; Low-Throughput (LT) protocols. Polynucleotide samples may be purified from a bodily fluid, such as blood, by using a Ficoll reagent, such as Ficoll-Paque PLUS GE Healthcare Life Sciences.

[00101] A plurality of polynucleotide sequences, such as that from a genome, may be first fragmented before subsequent steps as described herein. The size of the polynucleotide fragments, described in terms of length, may vary depending on the source of the target polynucleotide, the method used for fragmentation, and the desired application. In some cases one or more fragmentation steps may be used. For example, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, or more fragmentation steps may be used.

[00102] Fragments may be about 1-10, 10-20, 20-50, 50-100, 50-200, 100-200, 200-300, 300-400, 400-500, 500-1000, 1000-5000, 5000-10000, 10000-100000, 100000-250000, or 250000-500000 nucleotides in length. Fragments can be at least about 10, 20, 100, 200, 300, 400, 500, 1000, 5000, 10000, 100000, 250000, 500000, or more nucleotides in length. Fragments can be less than about 10, 20, 100, 200, 300, 400, 500, 1000, 5000, 10000, 100000, 250000, 500000, nucleotides in length.

[00103] Numerous fragmentation methods are described herein and known in the art. For example, fragmentation may be performed through physical, mechanical or enzymatic methods. Physical fragmentation may include exposing a target polynucleotide to heat or to ultraviolet (UV) light. Mechanical disruption may be used to mechanically shear a target polynucleotide into fragments of the desired range. Mechanical shearing may be accomplished through a number of methods known in the art, including repetitive pipetting of the target polynucleotide, sonication and nebulization. Target polynucleotides may also be fragmented using enzymatic methods. In some cases, enzymatic digestion may be performed using enzymes such as using restriction enzymes.

[00104] Restriction enzymes may be used to perform specific or non-specific fragmentation of target polynucleotides. The methods may use one or more types of restriction enzymes, generally described as Type I enzymes, Type II enzymes, and/or Type III enzymes. Type II and Type III enzymes are generally commercially available and well known in the art. Type II and Type III enzymes recognize specific sequences of nucleotide within a double stranded polynucleotide sequence (a "recognition sequence" or "recognition site"). Upon binding and recognition of these sequences, Type II and Type III enzymes cleave the polynucleotide sequence. In some cases, cleavage will result in a polynucleotide fragment with a portion of overhanging single stranded DNA, called a "sticky end." In other cases, cleavage will not result in a fragment with an overhang, creating a "blunt end." The methods may comprise use of restriction enzymes that generate either sticky ends or blunt ends.

IDT Exhibit 1020.3154

[00105] Restriction enzymes may recognize a variety of recognition sites in the target polynucleotide. Some restriction enzymes ("exact cutters") recognize only a single recognition site (e.g., GAATTC). Other restriction enzymes are more promiscuous, and recognize more than one recognition site, or a variety of recognition sites. Some enzymes cut at a single position within the recognition site, while others may cut at multiple positions. Some enzymes cut at the same position within the recognition site, while others cut at variable positions.

[00106] A polynucleotide may be exposed to two or more restriction enzymes simultaneously or sequentially. This may be accomplished by, for example, adding more than one restriction enzyme to a partition, or by adding one restriction enzyme to a partition, performing the digestion, deactivating the restriction enzyme (e.g., by heat treatment) and then adding a second restriction enzyme.

[00107] In some embodiments, the present invention can use dilution and spatial separation of target nucleic acid. In some instances, long fragments of nucleic acid are diluted before being spatially separated. Dilution can be accomplished by any method known in the art, such as by the addition of a diluent, such as water, or a suitable buffer. An exemplary method of dilution involves determining the concentration of the nucleic acid before dilution and calculating how much diluent to add so that the diluted sample can be partitioned into quantities that contain sub-genomic quantities of DNA (i.e. so that one sample contains less than one whole genome). In another exemplary method, dilution can be calculated so that the sample can be partitioned in a way that each partition contains approximately 1, 2, 3, 5, 10, 20, 50, 80, 100, 150, 200, 400, 500, 1000, 1500, 5,000, 10,000, 100,000, 1,000,000, 10,000,000, 100,000,000, 1,000,000,000 fragments of nucleic acid. In another exemplary method, dilution is accomplished to facilitate partitioning sample so that approximately 1%, 2%, 5%, 10%, 15%, 20%, 30%, 50%, 70%, 80%, 95% or 100% of one copy of the genome can be separated.

[00108] Spatial separation can be accomplished by many methods known in the art, such as pipetting, micropipeting, or microfluidics. Partitions can be made of any known methods in the art, including plates (e.g. 96-well), mirofluidic chambers, micro-droplets, or simple spatial separation on solid surfaces such as silicon chips or beads or semi-solid surfaces. Oil and/or emulsions can be used for spatial separation.

[00109] In an exemplary method, dilution and spatial separation is conducted so that the there is a low probability that two partitions contain the same locus of DNA from each parental chromosome, or that multiple fragments from the same genomic locus will be extremely rare.

III. Targeted Sequencing

[00110] Targeted sequencing can includes the ability to detect complex variation, avoiding clonal errors, and analysis that is less computationally burdensome (e.g. de novo sequencing). There are several embodiments of targeted sequencing. Some examples can include Primer Extension with Sequence Specific primers (PELA), Targeted RAndom Primer Extension Ligation and labeling

IDT Exhibit 1020.3155

(targeted RAPELLing). Targeted RAPELL can comprise a targeting step with the RAPELLing methods disclosed herein. Other targeted sequencing methods can include highly multiplexed PCR that can use biotin dUTP for long or short-range PCR. Another embodiment of targeted sequencing is Targeted Elongation Ligation and Amplification (TELA). Another embodiment of targeted sequencing is can be multiplexed primer extension and affinity reaction (mPEAR). Targeted sequencing methods can comprise a circularized rolling circle amplification (CRCA). In some cases, "targeted sequencing"? refers to any methods for the isolation and amplification of biologically relevant genomic locations for DNA sequencing. In some cases amplification is performed on conserved or functional elements of the genome that are relavant to assay by DNA sequencing. In some cases, this may include epigenetic information, such as methylation of nucleic acids, such as methylated DNA.

A. Targeted primer extension

[00111] Primer extension can be targeted by using primers that comprise a region that can hybridize to a known sequence. In some embodiments, the known sequence is within a target locus. In some embodiments, the known sequence is outside a target locus.

[00112] Primers can be designed to be tiled. A primer tiling strategy can be accomplished by using a plurality of unpaired or paired primers such that each primer can generate an amplicon. The primers can be designed such that multiple primers can generate multiple amplicons that can "tile" a loci. Tile can mean that the amplicons can overlap each other. The primers can be designed such that the amplicons generated essentially cover the entire target loci 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 2000, 5000, 10,000, 50,000, 100,000, 500,000, 1,000,000 times or more. In some instances, the locus can covered by amplifying regions of a longer target region. In some embodiments, the loci is a target loci. The primers can each hybridize to a region surrounding and including the target loci. The furthest upstream primer can be designed such that it is able to hybridize to a region starting approximately 30, 40, 50, 60, 70, 80, 90, 100, 120, 130, 140, 160, 180, 200, 250, 300, 350 400 or more nucleotides upstream of the target loci. In some embodiments, the starting region of the furthest upstream primer is between approximately 100 to approximately 200 nucleotides upstream of the target loci. The furthest downstream primer can be designed such that it is able to hybridize to a region starting approximately 30, 40, 50, 60, 70, 80, 90, 100, 120, 130, 140, 160, 180, 200, 250, 300, 350, 400 or more nucleotides downstream of the target loci. In some embodiments, the starting region of the furthest downstream primer is between approximately 100 to approximately 200 nucleotides downstream of the target loci.

[00113] In some cases primers may be designed to be complementary to a reference genome. The reference genome used may be a standard reference or an ethnic or population specific reference such

IDT Exhibit 1020.3156

as a reference that may include the major allele at each polymorphic position. In the case of SNPs, degenerate or N or every possible nucleotide of the four nucleotides may be used in synthesis of that position in the primer. A reference design may include assigning a "window" which may comprise a 100 bp sequence for the target location. In some cases, a window may be at least 1, 10, 20, 30, 40, 50 , 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000. 9000, 10000, 250000, or 500000 base pairs. In some cases, a window may be at most 1, 10, 20, 30, 40, 50 , 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000. 9000, 10000, 250000, or 500000 base pairs. Tiled windows may be assigned to a target locus such that the best performing probe based on length TM and specificity may be chosen from each window. Windows may be adjusted to account for any bias from defined windows. In some cases windows may overlap.

[00114] In some embodiments, the overlapping amplicons can cover essentially all of the target loci. In some embodiments, the overlapping amplicons can cover essentially all of a target loci and regions flanking the target region. The number of amplicons to cover a target locus can be approximately 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 150, 200, 250, 300, 350, 400, 500, 600, 700, 800, 900, 1000, 1500, 2000, 2500, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10,000 or more. The number of amplicons to cover a region can depend on the length of the region, the sequence variability within the region, the number of tandem repeats within the region, the percentage of CG bases in the region or other sequence or structural variations within the region that can affect sequencing quality. Typically, it may be preferable to have a larger number of amplicons covering a region to increase the depth and accuracy of the sequencing.

[00115] In some cases windows or tiles may be designed such that specific sequence motifs, such as CCG polynucleotides may cause sequencing chemistries to generate errors. The position of these polynucleotide sequences within a read may affect sequence quality. In some cases, if the sequence motifs are located near a solid surface, they may have a differential effect on the quality of the read, than if they were located at the end of a read. In some cases, adjusting the position at which these motifs are within the read produces a consensus and random error profile that may be corrected with the redundant read processing. This is an attribute that may not be available in PCR based approaches with identical read structure. PCR approaches may also incorporate the least troublesome sequence combination to avoid stuttering of polymerase during sequencing. Such a result may cause a CGG sequence interpretation during cluster amplification instead of the correct CCG configuration which is considered a systematic error.

[00116] Tiled strategies can also modify the error profile across a read at each position of the target (e.g. by tiled amplicons having different sequencing start sites). For example, a sequence, such as "CCG," can be more difficult to accurately sequence through. However, in some instances, the position of the read where the CCG sequence occurs can directly impact the ability of the chemistry to sequence through the read. Therefore, a tiling strategy can adjust the position of the CCG position in

IDT Exhibit 1020.3157

the read, which can allow for an increased chance that the chemistry will accurately read through the CCG sequence motif.

[00117] In some embodiments, the amplified regions can comprise target regions. The amplicons can be approximately 50, 60, 70, 80, 90, 100, 110, 120, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 270, 290, 300, 320, 340, 360, 380, 400, 450, 500, 550, 600 or more nucleotides in length. In some embodiments, each amplicon is generally between approximately 100 nucleotides to approximately 200 nucleotides.

[00118] The tiled amplicons can comprise nucleotides that overlap one another. The overlap can be approximately 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 110 120, 130, 140, 150, 160, 170, 180, 190, 200, 225, 250, 275, 300 or more nucleotides.

[00119] Primers can be designed such that they are optimized. Some non limiting examples of ways to optimize primers are by: variation in length or sequence to account for an optimal Tm; based on their specificity to the target hybridization location; to avoid dimerization with one another and/or to avoid common polymorphisms. In some embodiments, primers can be designed to avoid areas with common known polymorphisms, i.e. where the minor allele frequency is more than 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10% or more. Primers can comprise a 3'OH group that can be extended by a polymerase.

[00120] The primers can be labeled by using nucleotides. The primers can be labeled at the 3' end, the 5' end, and/or in the middle. The primers can comprise labeled nucleotides (e.g. nucleotides conjugated to biotin or fluorescent moieties). Methods for labeling nucleotides have been described herein. The primers can comprise a 5' nucleotide tail label. The primers can comprise a nucleotide tail and/ or a labeled nucleotide. The 5' nucleotide tail can comprise a known sequence. The known sequence added in the 5' nucleotide tail can make the amplicon useful for downstream reactions. The known sequence can comprise an adapter, a molecular barcode, and/or or other known sequence.

[00121] An adapter or sequencing adapter or adapter sequence can comprise a sequence that corresponds to an adaptor for the sequencing platform used in the reaction. An adapter, for example, can be added to or near the 5' end. In some embodiments, the 5' adapter is referred to as an A adapter. Adapter sequences are described further herein. The molecular barcode can correspond to the sample source, the direction of the elongation, and/or the target region.

[00122] The primer can be hybridized to a nucleic acid and elongation can occur. A strand displacing thermostable polymerase can be used to extend. A copy of the template nucleic acid can be generated by polymerase during elongation. The polymerase can also serve to displace any nucleic acid(s) that are hybridized to the template. The process of amplicon generation can repeated or cycled. The amplicon generation process can comprise: heat denaturation, primer annealing, and primer extension.

IDT Exhibit 1020.3158

[00123] The primers can be designed such that only the forward strand of the template is copied. In other embodiments, the primers can be designed such that only the reverse strand of the template is copied. In other embodiments, the primers can be designed such that both the forward and reverse strand of the template are both copied. The primers can comprise a molecular barcode that comprises two or more nucleotides that can be associated with information about the orientation or template strand being copied.

[00124] Forward and reverse reactions can be performed together. Forward and reverse reactions can be performed separately. When forward and reverse reactions are performed separately, they can be combined in later steps. The amplicons generated from the separate forward and reverse reactions can be mixed before library generation. The libraries created from separate forward and reverse reactions can be pooled prior to sequencing. The forward and reverse reactions can be sequenced separately and the data can be combined in silico.

[00125] The forward and reverse strand amplicons can be used for error correction. The forward and reverse strand amplicons can be used for refining mapping. The forward and reverse strand amplicons can be used for distance analysis.

[00126] The amplified products or amplicons can be size-processed in such a matter as to reduce or control the total length. In some instances, the size-processing can be fragmentation, in other instances, the size-processing can be to halt elongation. The size-processing can result in amplicons that are of a size that can be optimal for sequencing. Size-processing of the amplicons may occur by any enzymatic or physical means known in the art.

[00127] Enzymatic fragmentation can occur by using enzymes that can cut (e.g. hydrolyze) nucleic acid bonds. Some non limiting examples of enzymes that can cut nucleic acids include: hydrolases, nucleases, ribonucleases, deoxyribonucleases, phosphoesterases, topoisomerases, endonucleases, restriction enzymes, type II restriction endonucleases, or type I restriction nucleases.

[00128] In some instances, nucleic acid fragmentation can occur by physical or mechanical force. Some non limiting methods of physically fragmenting nucleic acids can include: sonification, nebulization, or hydroshearing.

[00129] The size-processing of the amplicons can also occur at the 3' end during elongation. In some instances, the size-processing accomplished by generating random 3' ends. 3' size-processing can occur by using nucleotides that lack a 3' OH group, biotin-ddNTPs, dUTP followed by UDG/APE1, methyl C, or other modified nucleotides. Incorporation of such nucleotides can halt or terminate elongation. The frequency of incorporation of such nucleotides can be altered by titrating the amount of such nucleotides. In some embodiments, amplicons terminated using a biotinylated ddNTP, can be further isolated by using streptavidin bead purification.

[00130] Nucleotides can be added to the 3' end of the amplicon; this can be called a 3' nucleotide tail. The 3' nucleotide tail can comprise a known sequence. The known sequence can make the amplicon useful for downstream reactions. The 3' nucleotide tail can comprise an adapter, a

IDT Exhibit 1020.3159

molecular barcode, and/or a known sequence for amplification. The adapter can comprise a sequence that corresponds to an adaptor that can be used with a sequencing platform. In some embodiments, this 3' adapter can be referred to as an B-adapter. The molecular barcode can correspond to the sample source, the direction of the elongation, and/or the target region. The second known sequence can comprise known nucleic acids. The methods used to add a 3' nucleotide tail can vary based on the method(s) used for size processing. In some cases the 3' tail can be an extension of further nucleotides, differentiating one copy and 5' labeled molecule from another copy and 5' labeled molecule.

[00131] In the presence of a 3' OH (e.g. if enzymatic or physical fragmentation is used as a method of size-processing), the 3' nucleotide tail can be added by ligation or by an additional primer extension step. The 3' nucleotide tail can be added through ligation, the steps comprising: end repair and ligation of double stranded construct with a random overhang. The 3' nucleotide tail can be added through primer extension and elongation on the single stranded template.

[00132] If a terminating nucleotide is used for size-processing, the 3' end of the molecule may not be available for primer extension or ligation due to the lack of a 3' OH. In the absence of a 3' OH, a 3' nucleotide tail can be added by hybridizing a primer comprising a random nucleotide sequence on the 3' end and a complementary sequence to the 3' nucleotide tail on the 5' end (e.g. 5'-known sequence-random sequence...-3'). The random sequence of the primer can hybridize to the amplicon. The random sequence can be comprised of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, or more nucleotides. A mixture of primers that comprise a random sequence can comprise a mixture of all possible sequences. In one non limiting example, if the random sequence comprises 5 nulceotides than there can be 4^5 or 1,024 possible combinations of A, T, C, and G bases. In another non limiting example, if the random sequence comprises 6 nucleotides, than there can be 4^6 or 4,096 combinations of A, T, C, and G bases and the primer mix can comprise a mixture of all or essentially all of the possible combinations.

[00133] In some embodiments, the amplicon has been purified or captured through incorporation of a biotin-ddNTP binding to streptavidin. A strand displacing polymerase can be used to extend 5-3'. The random sequence at the end of the biotinylated template can extend and displace all other randomly associated primers, thereby becoming the only extended nucleic acid on the template amplicon. In one non-limiting example, the complex can comprise: a 5' A adapter - locus specific primer - target - terminating ddNTP - biotin 3'; the reverse strand can comprise: 5'- B adapter - random sequence - target - locus specific sequence - A adapter - 3'. The complex can be double stranded. The complex can be further isolated by washing the streptavidin beads and removing the supernatant. One strand of the complex can be released from the strepavidin beads by heat denaturation and the other strand can remain covalently bound to the straptavidin through the biotin moiety.

IDT Exhibit 1020.3160

[00134] The processed amplicon can be amplified through linear amplification or PCR. To amplify the nucleic acid, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14,1 5, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30 or more cycles of PCR can be performed with primers that are complementary to the A and B adapter sequences. In an exemplary embodiment, approximately 9 to approximately 12 cycles of PCR are performed.

[00135] The amplification step can incorporate more known nucleotides at the 5' and 3' ends (e.g. by using PCR primers that contain known sequences at the 5' end). The molecular barcodes can be added during this step. In some embodiments, the known nucleotide sequences at the 5' and/or 3' end may not comprise the full length sequencing adaptors that can be necessary for a next generation sequencing reaction. In these instances, full length sequencing adapters can be incorporated during the PCR step. In some cases such as ILMN sequnceing the most 3' ends of the adapter sequences may be identical. In some cases an extra base or a few extra bases may need to be added to insure directionality of A and B adapter sequences on the same strand of the library molecule.

[00136] In an exemplary embodiment, the resulting sequencer ready library can consist of double stranded molecules in the following format: 5'- A adapter - synthetic primer - amplicon - B adapter - 3'. In some embodiments, the resulting sequencer ready library can comprise 1, 2, or 3, or more molecular bar codes. The A adaptor and B adapter can correspond to adapter sequences that are used with the sequencing platform; adapter sequences are discussed herein. Any method of sequencing can be used to analyze the sequencer ready library; a summary of sequencing methods can be found herein. The data resulting from such a sequencing reaction can be stored, transmitted and/or analyzed by any method known in the art; methods of data storage and transmittal can be found herein.

[00137] Sequencing reads can be analyzed through data analysis. The data can be analyzed using software. Software can trim adaptors off of the reads. The sample can be identified if optional molecular barcodes were incorporated. Duplicate reads can be removed. The genomic coordinate of the read can be identified by the known synthetic sequence at the beginning of the read. Reads corresponding to the same known genomic coordinate can be binned together and a consensus sequence can be generated. The consensus sequence can be generated without the use of a reference genome. Consensus sequences can be compared to a reference genome. Reads that do not form a consensus can be removed from analysis. Each contiguous segment of target nucleic acid can be considered a singular 'target' and all primers corresponding to that target are considered that targets primer set. This can produce an in silico read length equivalent to the full length of the target, regardless of target size.

[00138] After a consensus sequence is determined, known haplotypes for the target region can be queried. A haplotype can be a member of a polymorphic set. Haplotype data can be information concerning the haplotype of a sample. Match haplotypes that are known to either cause disease or not can be identified. For consensus sequences that do not match a known haplotype, the de novo sequence can be used to determine novel haplotypes, haplogroups and/or structural variation.

IDT Exhibit 1020.3161

[00139] Advantages of targeted sequencing can include data analysis that may not require alignment of each read to a reference genome (i.e. the synthetic sequence at the beginning of each read may identify the genomic position and the remainder of the read can be de novo or reference free). Sequencing variants can be more reliably distinguished from genomic variants. For example, several amplicons that cover the same genomic region having a variant sequence can indicate that the variant is genomic. Primer sets can be binned for each contiguous target, allowing for assembly. Complex variation can be detected. Random 3' end of reads can avoid clonal errors (i.e. can show different templates with low frequency variation such as somatic mutations). Other advantages of the targeted sequencing method include: ddNTP termination can provide random fragmentation that can involve less clean-up. Synthetic sequences in the chimeric library molecule may be used for identification or sample and genomic coordinate. In some cases, they may be removed from the read and assembly analysis such that only sample derived sequences is used for assisted de novo assembly. Random 3' ends of reads can ensure that the sequenced molecule is not "clonal", thereby possibly providing a dramatic reduction in errors and allowing greater sensitivity to detect somatic variation. Duplicate reads can be removed in silico, whereas traditional PCR-based targeted sequencing does not allow for removal of duplicates. Tiling probe design can allow for redundant sampling and consensus read lengths. This can avoid target drop out because of SNPs or novel biology under one primer site. The incorporation of biotin during primer extension can simplify clean up. Random priming of isolated molecules can allow incorporation of B-adaptors without ligation or additional clean up or end repair. Synthetic sequences at the beginning of each read can allow for dramatically reduced computational burden by avoiding reference genome mapping. Long consensus reads can allow for detection of complex variation. Linear consumption of target specific synthetic primers can reduce cost per sample and increases volume per production lot.

B. Highly multiplexed PCR

[00140] Targeted sequencing methods can comprise highly multiplexed PCR with or without biotin. A target region can be isolated from a genomic DNA sample. Target regions can comprise a plurality of target genes. Target genes have been disclosed herein. Target regions can be isolated using biotinylated capture of PCR products. The PCR products size can range from 10-1000, 100-10,000, 100-20,000, 1,000-20,000, 2,000-15,000, 10,000-15,000, 10,000- 20,000, 10,000-100,000, or 10 and 200,000 nucleotides in length.

[00141] Primers can be designed to flank the one or more areas of interest (i.e. target regions or target loci). The one or more areas of interest can be genes. Primers can hybridize to the target sequence. An amplification reaction can be performed. The amplification reaction can be PCR. The PCR can be long range PCR. The amplification reaction can be highly multiplexed. The amplification reaction can be low copy or low cycle. The PCR of one or more loci can be

IDT Exhibit 1020.3162

multiplexed. The PCR can be highly multiplexed PCR. Highly multiplexed can me that more than 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 120, 140, 160, 180, 200, 250, 300, 400, 500, 600, 700, 800, 900, 1000, 1500, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000 or more target loci. The amplification of a loci can be performed independently. The one or more amplification products can be pooled.

[00142] Biotin-conjugated dNTPs can be incorporated. Biotin-conjugated dNTPs can be incorporated during amplification. Amplification can occur with a dNTP mixture that comprises approximately 0.1%, 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 12%, 15%, 20%, 25% or more biotin conjugated dNTPs. In some embodiments the biotin conjugated dNTP can be dUTP.

[00143] The PCR amplified product can be run on a gel. The gel can be an agarose gel. The gel can be a low-melt agarose gel. A marker ladder can be loaded into an adjacent well. The portion of the gel that contains PCR product can be excised. In some embodiments, the portion to be excised will be guided by the length expected from primer design. In some embodiments, the gel portion corresponding to PCR product in the 10-15K range is the gel portion to be excised.

[00144] The amplificatiom product can be purified from the agarose. The gel-excised agarose can be dissolved. The gel-excised agarose can be dissolved in dissolving buffer. The purified or semi-purified polynucleotide can be further isolated. The polynucleotide can be bound to a column. The column can be a genomic DNA binding column. The column can be washed. The polynucleotide can be eluted from the column. The eluted polynucleotide can be fragmented. Methods of fragmentation have been disclosed herein. The polynucleotide can be fragmented in to fragment ranges that range from approximately 50-1000, 100-1000, 150-700, 200-660, 100-800, or 10-1500 nucleotides.

[00145] Target amplification product can be separated. Target amplification product can be separated based on size selection. Target amplification product can be separated based on size via gel electrophoresis/gel purification, size exclusion columns/column clean-up, and/or solid phase reversible immobilization (SPIR) beads optimized for size selection.

[00146] Target purification can be separated by affinity purification (e.g. with streptavidin if biotin dNTPs have been incorporated). The polynucleotide comprising biotin can be exposed to streptavidin. The polynucleotide can be exposed to streptavidin before or after it has been fragmented. The streptavidin can comprise straptavidn-coated beads. The streptavidin-biotin-fragment complex can be purified or semi-purified. The streptavidin-biotin-fragment complex can be washed.

[00147] The polynucleotide fragments can be treated. Methods of polynucleotide treatment are disclosed herein. The polynucleotide fragments can be end-repaired. Methods of end-repair are disclosed herein. The polynucleotide fragments can be adapter tailed. Methods of adapter tailing are disclosed herein. The polynucleotide fragments can be amplified. Methods of amplification are disclosed herein. The polynucleotide fragments can be amplified in the presence of straptavidin.

IDT Exhibit 1020.3163

[00148] Amplified product can be sequencer-ready and may comprise a sequence library. Amplified product may be further processed. Methods of polynucleotide processing are disclosed herein. Amplified product may be run on an agarose gel. Amplified product may be gel excised and purified before sequencing.

[00149] Methods of sequencing a sequence library are disclosed herein. Data can be produced from sequencing a library. Methods for storing and transmitting data that has been produced from a sequencing reaction are disclosed herein. The data can be analyzed or processed. Methods for analyzing or processing of data have been disclosed herein. The analyzed data can be used. Methods of using analyzed sequencing data have been disclosed herein.

C. Targeted elongation and ligation adapter (TELA)

[00150] Target molecules may be tagged or barcoded with alternative methods, herein referred to as TELA. The TELA methods, as shown in **FIG. 10** generally provide for attachment of barcodes in a multi step process involving hybridization of primers comprising: locus specific sequences; universal adapter sequences, and barcode sequences. Primers with this configuration are herein referred to as TELA primers. Primer extension is then performed followed by ligation of products to form a contiguous sequence. In another embodiment, universal adapter sequences, contiguous with barcode sequences may be hybridized to polynucleotide sequences containing other universal adapter sequences and amplified together. In other cases, universal adapter sequences, contiguous with bar code sequences, may be hybridized to polynucleotide sequences containing other adapter sequences which may be suitable for high throughput sequencing platforms or other applications, such as ligation of a locus specific primer ot a random tailed primer.

D. Design and Annealing of TELA primers

[00151] In some embodiments, target polynucleotides are first hybridized to TELA primers formed from locus specific sequences, or second and third probe domains, (which may be of known, partially known or unknown sequence) and universal adapter sequences and barcodes, or first and 4th probe domains. Hybridization of TELA primers to target molecules form primer-target constructs which are then used to carry out an initial primer extension reaction in which extension products complementary to the template strand of each individual adaptor-target construct are formed as shown in **FIG. 7**. The resulting primer extension products are then ligated and may be amplified to collectively provide a library of tagged or barcoded template polynucleotides as shown in **FIG. 6**. The term library refers to the collection of target fragments containing known common sequences at their 3′ and 5′ ends, and may also be referred to as a 3′ and 5′ modified library. In some cases, a locus specific primer and a random primer may be used in a similar strategy, whereby either a forward or reverse locus specific primer is used in combination with a corresponding (i.e forward/reverse) random primer. In some

IDT Exhibit 1020.3164

cases, the use of a combination of locus specific primer and random primer may generate overlapped amplicons. In some cases, this may generate longer target lengths for sequencing.

[00152] In some cases, universal adapter sequences for all samples or one or both strands of the duplexes may carry the tag sequence to barcode or track the identity of the samples. In some cases, a barcode is not included.

[00153] An important feature of the TELA primer sequence, as shown in **FIG. 7** is that a portion of the sequence may not fully anneal to the target sequence, **710**. Generally, this portion of the sequence comprises a separate primer site, **730**, which is contiguous with sequence that is complementary to a locus of interest, **700**. TELA sequences are generally formed by the annealing of two partially complementary polynucleotide strands so as to provide, when the two strands are annealed, at least one double-stranded region and at least one non-complementary single-stranded region. In some cases, the locus specific region may be linked to a random spacer sequence, **720**. In some cases, spacer sequences may be less than 20, 30, 40 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700 800, 900 or 1000 nucleotides in length. In other cases, spacer sequences may be greater than 20, 30, 40 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700 800, 900 or 1000 nucleotides in length.

[00154] In general the double stranded region of the adapter comprises or more consecutive nucleotides, formed by annealing of the two partially complementary polynucleotide strands, generally between complementary sequences of locus specific sequence. Double stranded, as provided herein generally refers to two strands that have annealed and does not refer to any particular structural DNA feature. Additionally, the double stranded region may also refer to the locus specific sequence, as the sequence is complementary to a sequence in the target polynucleotide.

[00155] Generally, the locus specific region of TELA primers, as in mPEAR primers, may be designed to be as short as possible without loss of function. In this context, 'function' refers to the ability of the double-stranded region to form a stable duplex under standard reaction conditions for an enzyme-catalyzed nucleic acid ligation reaction, (e.g. incubation at a temperature in the range of 4° C. to 60° C. in a primer extension buffer appropriate for the enzyme), such that the two strands forming an adaptor remain partially annealed during extension of the primer on the target molecule.

[00156] Identical adaptors are ligated to both ends of each target polynucleotide. The target sequence in each adaptor-target construct will be flanked by complementary sequences derived from the double-stranded region of the primers. The longer the double-stranded region, and hence the complementary sequences derived therefrom in the adaptor-target constructs, the greater the possibility that the TELA primer-target construct is able to fold back and base-pair to itself in these regions of internal self-complementarity under the annealing conditions used in primer extension and/or PCR. It is, therefore, generally the double-stranded region can be less than 100, 90, 80, 70, 60, 50, 40 30, 20, 10 nucleotides in length. In some instances, the double-stranded region may be greater than 100, 90, 80, 70, 60, 50, 40 30, 20, 10 nucleotides in length in order to reduce this effect. The stability of the double-stranded region may be increased, and hence its length potentially reduced, by

IDT Exhibit 1020.3165

the inclusion of non-natural nucleotides which exhibit stronger base-pairing than standard Watson-Crick base pairs.

[00157] The TELA primer may vary in the percentage of complementarity in the locus specific sequence. In some cases it may be 100% complementary in the double-stranded region. In other cases it may greater than 1%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% complementary. In other cases it may be less than 1%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% complementary. One or more nucleotide mismatches may be tolerated within the double-stranded region, provided that the two strands are capable of forming a stable duplex under standard ligation conditions.

[00158] Sequences of universal adaptors for use in the methods may generally include a double-stranded region forming the 'ligatable' end of the adaptor, i.e. the end that is joined to a target polynucleotide in the ligation reaction. The ligatable end of the adaptor may be blunt or, in other embodiments, short 5′ or 3′ overhangs of one or more nucleotides may be present to facilitate/promote ligation. The 5′ terminal nucleotide at the ligatable end of the adaptor is phosphorylated to enable phosphodiester linkage to a 3′ hydroxyl group on the target polynucleotide.

[00159] Another feature can include a region of a universal adapter sequence where the two polynucleotide strands forming the adaptor exhibit a degree of non-complementarity such that the two strands are not capable of fully annealing to each other under standard annealing conditions for a primer extension or PCR reaction. In some instances, this region may undergo annealing under standard reaction conditions for an enzyme-catalyzed ligation reaction, provided that the two strands revert to single stranded form under annealing conditions in an amplification reaction.

[00160] Generally, regions may be designed to prevent annealing in a variety of ways as described herein.

[00161] In terms of length, regions that may not anneal may be determined by function, for example, the need to provide a suitable sequence for binding of a primer for primer extension, PCR and/or sequencing. Generally, the length of such a region may extend for any number of nucleotides. In many cases it is preferred to minimize the overall length of the adaptor, for example, in order to facilitate separation of unbound adaptors from adaptor-target constructs following the ligation step. Therefore, it is generally preferred that the unannealed regions should be less than 20, 30, 40 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700 800, 900 or 1000 nucleotides in length. In other cases, unannealed regions should be greater than 20, 30, 40 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700 800, 900 or 1000 nucleotides in length. In some cases non complementary regions may destabilize the 5' end of the primer. In some cases, minimizing the length of this primer may also minimize the chance that the tail sequence coud partially bind to another region of the template DNA.

[00162] The actual nucleotide sequence of the TELA primers may be any sequence and length suitable. TELA primers may be selected by the user such that the desired sequence elements are ultimately included in the common sequences of the library of templates derived from the adaptors to,

IDT Exhibit 1020.3166

for example, provide binding sites for particular sets of universal amplification primers and/or sequencing primers. Additional sequence elements may be included, for example, to provide binding sites for sequencing primers which will ultimately be used in sequencing of template molecules in the library, or products derived from amplification of the template library, for example on a solid support in sequencing applications.

[00163] TELA primer sequences may comprise two strands of DNA, but may also include any nucleotides or nucleotide derivatives that may be suitable. Alternative nucleotides may include mixtures of natural and non-natural nucleotides (e.g. one or more ribonucleotides) linked by a mixture of phosphodiester and non-phosphodiester backbone linkages. Other non-nucleotide modifications may be included such as, for example, biotin moieties, blocking groups and capture moieties for attachment, such as biotinylated nucleotides.

[00164] Further adapter sequences may also comprise exonuclease resistant modifications such as phosphorothioate linkages. Such modifications reduce the number of adaptor-dimers present in the library, since the two adaptors cannot undergo ligation without removal of their non complementary overhangs. The adaptors can be treated with an exonuclease enzyme prior to the ligation reaction with the target, to ensure that the overhanging ends of the strands cannot be removed during the ligation process. Treatment of the adaptors in this manner reduces the formation of the adaptor-dimers at the ligation step.

E. Primer Extension and Ligation

[00165] After annealing TELA primers, a primer extension reaction may be performed using any suitable polymerase as described herein (see Section II-F). Further, one or more products may be ligated together to form a contiguous sequence using ligations methods as described herein.

[00166] Additionally, one or more contiguous primer extension-ligation products may be further assembled into longer fragments for downstream analysis as shown in **FIG. 6**. In some cases, universal adapter sites in the primers may be used as hybridization sites to anneal one or more products together. Using PCR or further primer-extension reactions, multiple products may be assembled into longer contiguous strands.

[00167] The combined ligated polynucleotide sequences and unligated adaptor polynucleotide constructs may be purified from any components of the ligation reaction, such as enzymes, buffers, salts and the like. Suitable purification methods are known in the art and utilize standard methods (Sambrook and Russell, Molecular Cloning, A Laboratory Manual, third edition).

[00168] Additionally, after ligation of the first primer extension product, adapters may be attached to the flanking regions of the ligation product. Adapter primers, comprising an adapter region and a universal priming region, complementary or partially complementary to universal adapter sites in the ligation product may be used to generated polynucleotides with attached adapters. The general

IDT Exhibit 1020.3167

protocol for using TELA primers (i.e hybridization followed by primer extension and ligation), and general design of TELA primers may apply to use of adapter primers.

[00169] In some cases, adapter sequences may be useful for downstream applications such as sequencing as described herein.

F. Amplification Techniques

[00170] Numerous amplification methods and techniques are known in the art. Any suitable methods may be used in the methods of this disclosure, so as to increase the quantity or amount of polynucleotides, while maintaining the initial content of sequence information of the original sample or ligation product. One or more amplification methods may be used and in one or more combinations.

[00171] Examples of amplification methods may include but are not limited to, polymerase chain reaction (PCR) (U.S. Pat. Nos. 4,683,195; and 4,683,202; PCR Technology: Principles and Applications for DNA Amplification, ed. H. A. Erlich, Freeman Press, NY, N.Y., 1992), ligase chain reaction (LCR) (Wu and Wallace, Genomics 4:560, 1989; Landegren et al., Science 241:1077, 1988), strand displacement amplification (SDA) (U.S. Pat. Nos. 5,270,184; and 5,422,252), transcription-mediated amplification (TMA) (U.S. Pat. No. 5,399,491), linked linear amplification (LLA) (U.S. Pat. No. 6,027,923), and the like, self-sustained sequence replication (Guatelli et al., Proc. Nat. Acad. Sci. USA, 87, 1874 (1990) and WO90/06995), selective amplification of target polynucleotide sequences (U.S. Pat. No. 6,410,276), consensus sequence primed polymerase chain reaction (CP-PCR) (U.S. Pat. No. 4,437,975), arbitrarily primed polymerase chain reaction (AP-PCR) (U.S. Pat. Nos. 5,413,909, 5,861,245) and nucleic acid based sequence amplification (NASBA). (See, U.S. Pat. Nos. 5,409,818, 5,554,517, and 6,063,603, each of which is incorporated herein by reference). Other amplification methods that may be used include: Qbeta Replicase, described in PCT Patent Application No. PCT/US87/00880, isothermal amplification methods such as SDA, described in Walker et al., Nucleic Acids Res. 20(7):1691-6 (1992), and rolling circle amplification, described in U.S. Pat. No. 5,648,245. Other amplification methods that may be used are described in, U.S. Pat. Nos. 5,242,794, 5,494,810, 4,988,617 and in U.S. Ser. No. 09/854,317 and US Pub. No. 20030143599, each of which is incorporated herein by reference. In some aspects DNA is amplified by multiplex locus-specific PCR. In a preferred aspect the DNA is amplified using adaptor-ligation and single primer PCR. Other available methods of amplification, such as balanced PCR (Makrigiorgos, et al., Nature Biotech, 20:936-9 (2002)) and isothermal amplification methods such as nucleic acid sequence based amplification (NASBA) and self-sustained sequence replication (Guatelli et al., PNAS USA 87:1874 (1990)). Based on such methodologies, a person skilled in the art readily can design primers in any suitable regions to be amplified.

IDT Exhibit 1020.3168

G. Amplification Products and Conditions

[00172] In general, any suitable amplification products and conditions to produce products may be used in the methods of this disclosure. Various amplification lengths, cycle times, hybridization, annealing and extension conditions may be used, as appropriate for various amplification techniques and sequences.

i. Amplification Lengths

[00173] Generally, the length of an amplified product may be any length and contain any sequence that may be useful in the enumeration of sequences. Generally, an amplified polynucleotide may be at least about 5 bp, 10 bp, 20 bp, 30 bp, 40 bp, 50 bp, 60 bp, 70 bp, 80 bp, 90 bp, 100 bp, 200 bp bp, 300 bp, 400 bp, 500 bp, 600 bp, 700 bp, 800 bp, 900 bp, 1 kb, 2 kb, 3, kb, 4 kb, 5 kb, 6 kb, 7 kb, 8 kb, 9 kb, 10kb, 20 kb, 30 kb, 40 kb, 50 kb, 75 kb, or 100 kb. Generally, an amplified polynucleotide may be at most about 5 bp, 10 bp, 20 bp, 30 bp, 40 bp, 50 bp, 60 bp, 70 bp, 80 bp, 90 bp, 100 bp, 200 bp bp, 300 bp, 400 bp, 500 bp, 600 bp, 700 bp, 800 bp, 900 bp, 1 kb, 2 kb, 3, kb, 4 kb, 5 kb, 6 kb, 7 kb, 8 kb, 9 kb, 10kb, 20 kb, 30 kb, 40 kb, 50 kb, 75 kb, or 100 kb.

ii. Amplification Conditions

[00174] In general any suitable amplification conditions may be used, for either selective or universal amplification. In some cases, amplification may be linear. In some cases, amplification may be logarithmic. Since the methods of the disclosure provide for enumeration of one or more sequences, which may be amplified, it may be suitable to control amplification in various steps to control variability between samples.

[00175] For example, in some cases, a limited number of amplification cycles may be used in either a selective or universal amplification step. This may be particularly suitable for selective amplification wherein different primer sets for different loci or barcodes may behave differently under multiplex conditions wherein a plurality of loci or barcodes are used. Primers in different primer sets may differ in their ability to hybridize to template, and thus yield differences in amplification efficiency between primer sets. Each set of primers for a given locus may behave differently based on sequence context of the primer and sample DNA, buffer conditions, and other conditions. A universal DNA amplification for a multiplexed assay system may generally introduce less bias and variability.

IDT Exhibit 1020.3169

[00176] To minimize amplification variation between one or more loci or barcodes, for example, amplification may be performed using a linear amplification method, followed by logarithmic universal amplification. In some cases, the number of cycles is limited between 1-50 cycles, such that amplification is linear or near linear. In some cases amplification cycles for linear amplification may be at least about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40 or 50 cycles. In some cases amplification cycles for linear amplification may be at most about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30 ,40 or 50 cycles. In some cases, after linear selective amplification of sequences from ligation products, a logarithmic universal amplification step may be performed as described herein. Universal amplification, wherein common primer sets may be used for a plurality of loci or barcode amplification products may further reduce amplification variability, while producing increasing amounts of sample.

[00177] In other cases, logarithmic amplification may be used before linear amplification. In some cases amplification cycles for logarithmic amplification may be at least about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30 ,40 or 50 cycles. In some cases amplification cycles for logarithmic amplification may be at most about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30 ,40 or 50 cycles.

[00178] Generally, any suitable number of primer sets may be used for amplification. In some cases, amplification primer sets may be about equal to the number of loci tested. In some cases, primers sets may be at least about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80 , 90 100, 125, 150, 175, 200, 300, 400, 500, 600, 700 800, 900 or 1000 primer sets. In some cases, primers sets may be at most about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80 , 90 100, 125, 150, 175, 200, 300, 400, 500, 600, 700 800, 900 or 1000 primer sets.

IV. Sequencing Methods

[00179] Numerous methods of sequence determination are compatible with the systems and methods of the disclosures. Exemplary methods for sequence determination include, but are not limited to, hybridization-based methods, such as disclosed in Drmanac, U.S. Pat. Nos. 6,864,052; 6,309,824; and 6,401,267; and Drmanac et al, U.S. patent publication 2005/0191656, which are incorporated by reference, sequencing by synthesis methods, e.g., Nyren et al, U.S. Pat. Nos. 7,648,824, 7,459,311 and 6,210,891; Balasubramanian, U.S. Pat. Nos. 7,232,656 and 6,833,246; Quake, U.S. Pat. No. 6,911,345; Li et al, Proc. Natl. Acad. Sci., 100: 414-419 (2003); pyrophosphate sequencing as described in Ronaghi et al., U.S. Pat. Nos. 7,648,824, 7,459,311, 6,828,100, and 6,210,891; and ligation-based sequencing determination methods, e.g., Drmanac et al., U.S. Pat. Appl. No. 20100105052, and Church et al, U.S. Pat. Appin Nos. 20070207482 and 20090018024.

[00180] Sequence information may be determined using methods that determine many (typically thousands to billions) nucleic acid sequences in an intrinsically parallel manner, where many sequences are read out preferably in parallel using a high throughput serial process. Such methods include but are not limited to pyrosequencing (for example, as commercialized by 454 Life Sciences,

IDT Exhibit 1020.3170

Inc., Branford, Conn.); sequencing by ligation (for example, as commercialized in the SOLiD™ technology, Life Technology, Inc., Carlsbad, Calif.); sequencing by synthesis using modified nucleotides (such as commercialized in TruSeq™ and HiSeq™ technology by Illumina, Inc., San Diego, Calif., HeliScope™ by Helicos Biosciences Corporation, Cambridge, Mass., and PacBio RS by Pacific Biosciences of California, Inc., Menlo Park, Calif.), sequencing by ion detection technologies (Ion Torrent, Inc., South San Francisco, Calif.); sequencing of DNA nanoballs (Complete Genomics, Inc., Mountain View, Calif.); nanopore-based sequencing technologies (for example, as developed by Oxford Nanopore Technologies, LTD, Oxford, UK), and like highly parallelized sequencing methods.

V. Kits

[00181] Kits may be used in preparing libraries of tagged polynucleotides using the method using either mPEAR, Rappel, targeted Rappel or TELA methods.

[00182] A kit may comprise at least a supply of mPEAR, TELA primers, universal adapters or a combination thereof, as defined herein, plus a supply of at least one amplification primer which is capable of annealing to the adaptor primer and priming synthesis of an extension product, which extension product would include any target sequence ligated to the adaptor when the adaptor is in use.

[00183] In some cases, features of the adaptor sequence for inclusion in the kit are as described elsewhere herein in relation to other aspects of the invention. The structure and properties of amplification primers are well known to those skilled in the art. Suitable primers of appropriate nucleotide sequence for use with the adaptors included in the kit can be readily prepared using standard automated nucleic acid synthesis equipment and reagents in routine use in the art. The kit may include a supply of one single type of primer or separate supplies (or even a mixture) of two different primers, for example a pair of amplification primers suitable for PCR or isothermal amplification of templates modified with the adaptor sequences in solution phase and/or on a suitable solid support (i.e. solid-phase amplification). The kit may comprise a double stranded adapter for ligation to a sample of interest, plus at least two different amplification primers that carry a different tag sequence, where the tag sequence does not hybridize to the adapter. This kit can be used to amplify at least two different samples where each sample is amplified using a single tagged primer, and then pooled after the individual amplification reactions.

[00184] Adaptors and/or primers may be supplied in the kits ready for use, or more preferably as concentrates requiring dilution before use, or even in a lyophilized or dried form requiring reconstitution prior to use. If required, the kits may further include a supply of a suitable diluent for dilution or reconstitution of the primers. Optionally, the kits may further comprise supplies of reagents, buffers, enzymes, dNTPs, etc., for use in carrying out PCR or isothermal amplification. Suitable (but non-limiting) examples of such reagents are as described in the Materials and Methods

IDT Exhibit 1020.3171

sections of the accompanying Examples. Further components which may optionally be supplied in the kit include 'universal' sequencing primers suitable for sequencing templates prepared using the mismatched adaptors and primers.

D. Capture of targeted library molecules via primer extension (mPEAR)

[00185] mPEAR is a sample preparation technique that can be used for targeted sequencing. It can involve several steps as shown in **FIG. 3**.

i. Fragmentation

[00186] A polynucleotide sample may be first fragmented before subsequent steps. Fragmentation methods are described herein. The size of the polynucleotide fragments, described in terms of length, may vary depending on the source of the target polynucleotide, the method used for fragmentation, and the desired application. In some cases one or more fragmentation steps may be used. For example, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, or more fragmentation steps may be used.

ii. Polynucleotide Strand End Repair

[00187] Fragmentation of polynucleotides, such as through mechanical shearing or enzymatic digestion, can result in fragments with a heterogeneous mix of blunt and 3′- and 5′-overhanging ends. In some embodiments DNA fragments can be repaired or treated using methods or kits (i.e. Lucigen DNA terminator End Repair Kit) known in the art to generate ends that are designed for insertion, for example, into blunt sites of cloning vectors. Blunt ended fragment ends of the population of nucleic acids can be sequenced. Further, in some cases, the blunt ended fragment may also be phosphorylated. The phosphate moiety can be introduced via enzymatic treatment, for example, using a kinase, (i.e. shrimp alkaline kinase). Blunt ended fragments may be dephosphorylated by using a phosphatase. Sticky ended fragments may be trimmed using a nuclease. Overhangs may be added to blunt ends by any method known in the art.

[00188] Polynucleotide sequences can be prepared with single overhanging nucleotides by, for example, activity of certain types of DNA polymerase such as Taq polymerase or Klenow exo minus polymerase which has a nontemplate-dependent terminal transferase activity that adds a single deoxynucleotide, for example, deoxyadenosine (A) to the 3′ ends of polynucleotides. Such enzymes can be utilized to add a single nucleotide 'A' to the blunt ended 3′ terminus of each strand of the target polynucleotide duplexes. Thus, an 'A' could be added to the 3′ terminus of each end repaired duplex strand of the target polynucleotide duplex by reaction with Taq or Klenow exo minus polymerase,

IDT Exhibit 1020.3172

whilst the adaptor polynucleotide construct could be a T-construct with a compatible 'T' overhang present on the 3′ terminus of each duplex region of the adaptor construct. This end modification also prevents self-ligation of both adapter and target such that there is a bias towards formation of the combined ligated adaptor-target sequences.

iii. Barcoding

[00189] identifier sequences or molecular barcodes can be used, **360**. These sequences provide for a characteristic marker of the source of particular target molecules that maybe identified in downstream application such as sequencing. Often, the unique identifier is a bar-code oligonucleotide of known sequence that is used to tag the target molecules. mPEAR methods can comprise attaching oligonucleotide barcodes to nucleic acid target molecules through an enzymatic reaction such as a ligation reaction. For example, the ligase enzyme may covalently attach a DNA bar code to fragmented DNA.

[00190] Another exemplary method for adding a molecular barcode can comprise using oligonucleotide primers containing a bar code sequences for use in an amplification reactions (e.g., PCR, or linear amplification etc.).

[00191] Often, as described herein, an identifier may be an oligonucleotide barcode sequence that is contiguous with the first or second probe in a probe set. In some cases, however, different identifiers may be used. An identifier, as with barcode sequences, may be unique or non-unique. For example, in some cases, the unique identifier may be a hybridization probe. In one example, a hybridization probe may comprise an oligonucleotide sequence and an additional component such as fluorescent element (i.e. nanoparticle, nanoprobe, quantum dot, etc). In some cases, one or more fluorescent elements may be described as barcodes as well. For example, fluorescent elements of varying wavelengths or colors may be arrayed in unique or non unique patterns or sequences. In other cases, the identifier is a dye, in which case the attachment may comprise intercalation of the dye into the analyte molecule (such as intercalation into DNA or RNA) or binding to a probe labeled with the dye. In still other cases, the identifier may be a nucleic acid oligonucleotide, in which case the attachment to the polynucleotide sequences may comprise a ligation reaction between the oligonucleotide and the sequences or incorporation through PCR. In other cases, the reaction may comprise addition of a metal isotope, in which either the first or second probe is labeled with the isotope.

IDT Exhibit 1020.3173

[00192] The unique identifiers (e.g., oligonucleotide barcodes, probes, etc.) may be attached to polynucleotide sequences in a variety of ways. Barcodes may comprise different lengths. In some cases they may comprise about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 50, 100, 500, or 1,000 nucleotides in length. In some cases, molecular barcodes may be less than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 50, 100, 500, or 1000 nucleotides in length. In some cases, multiple barcodes may be attached to a polynucleotide. In some cases about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 50, 100, 500, or 1,000 barcodes may be attached to a single polynucleotide. In some cases, a polynucleotide may be attached with less than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 50, 100, 500, or 1,000 barcodes.

iv. Attachment of known sequences

[00193] Known sequences can be attached to the ends of the fragmented polynucleotide sample. Known sequences can comprise a molecular barcode, an adaptor for sequencing, or any other sequence such as a universal primer sequence. A universal primer sequence can comprise, for example, a known sequence that primers can hybridize to, e.g. for a PCR amplification reaction. In some embodiments, a ligation reaction ligation can be used to covalently join a known sequence to a fragment. Ligation methods can utilize a ligase enzyme such as a DNA ligase to join the ends of the polynucleotide strands (e.g. the fragment and the known sequence) such that covalent linkages are formed. A 5′-phosphate moiety may facilitate ligation to the target 3′-OH. Joining can mean covalent linkage of polynucleotide strands which were not previously covalently linked. In some embodiments, joining can involve formation of a phosphodiester linkage between the two polynucleotide strands, but other means of covalent linkage (e.g. non-phosphodiester backbone linkages) may be used. The known sequences can incorporate a marker that can facilitate separation (e.g. biotinylated nucleotides or nucleotides that are attached to a moiety that can be separated by antibody purification methods).

[00194] The fragments with adjoined known sequences can be amplified and/or purified. Purification can be based on size of the polynucleotides. Purification can be based on separation techniques such as biotin/streptavidin or antibody-based separation techniques. Purification can mean that the polynucleotide is separated from one or more components of the ligation reaction, such as enzymes, buffers, salts and the like. Suitable purification methods are known in the art and utilize standard methods. The fragments with adjoined, known sequences can be amplified. Amplification can be accomplished by any means known in the art, including PCR and/or linear amplification. During the amplification step, primers may be used that have tails with known sequences which can add, for example, molecular bar codes and/or adaptor sequences.

[00195] In some cases, unique barcodes may be formed from the combination of various sequences. In some cases, non-unique barcodes may be linked with additional sequences such as probe sequences, portions of probe sequences or additional sequences linked to the probe to form a unique barcode sequence. For example, the formation of a unique sequence may be formed at the beginning

IDT Exhibit 1020.3174

(start) and end (stop) portions of the probe sequences when used, alone or in combination, with a non-unique bar code sequence. The combination of sequences (i.e. probe sequence and non-unique barcode sequence), may provide unique identifying sequences. For example, in some cases a barcode may be designed with a general structure, 5'XXXXYYYY, wherein X is a variable length region complementary to one sequence selected from sequences including but not limited to first and/or second probe/hybridization sequences, adapter sequences, universal priming sequences, or linker sequences. Y may be selected from a non-unique barcode sequence of variable length. In some cases, Y sequences may be common to all probe sets in a sample. In other cases, Y sequences may be unique one locus, or a plurality of loci, such as a whole chromosome, or loci associated with a particular disease or genotype. In some cases, the length, or number of nucleotides defined as either X or Y may be about 1-20, 20-50, 50-75, 75-100, 100-150, 150-200, 200- 300, 300- 400 or 400-500 nucleotides. In some cases, the length, or number of nucleotides defined as either X or Y may be at least about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 75, 100, 125, 150, 175, 200, 300, 400 or 500 nucleotides. In some cases, the length, or number of nucleotides defined as either X or Y may be at most about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 75, 100, 125, 150, 175, 200, 300, 400 or 500 nucleotides.

[00196] In alternative configurations, barcodes may be formed through the combination of at least about 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10 sequences. Barcodes may be formed the combination of at most about 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10 non-unique sequences.

v. mPEAR primers

[00197] An mPEAR primer can anneal to known sequences universal priming sites in the known sequences that have been joined to the fragments. An exemplary mPEAR primer, is diagrammed in **FIG. 3, 350.** An mPEAR primer can comprise: a universal amplification sequence, a target loci specific sequence, a molecular bar code, a spacer sequence and/or other known sequence(s). A universal amplification sequence may hybridize to a universal adapter sequence attached to the fragment. A spacer sequence may be of a variable length and can comprise degenerate nucleotides, known nucleotides or any combination thereof. A locus specific sequence may hybridize to a target locus, **300** or may hybridize to a region just upstream or downstream to a region of interest, **310**. An mPEAR primer may be designed to anneal to a site upstream or downstream from the region of interest. In some embodiments, this can allow for increased specificity, e.g. pseudogenes and gene families with similar sequence homology can be avoided, thereby reducing false positives.

[00198] Multiple mPEAR primers may be used to target the same or different regions of interest. Two or more mPEAR primers may be used that target the same region of interest. Two or more mPEAR primers may be designed such that they tile a region of interest. Two mPEAR primers may

IDT Exhibit 1020.3175

be designed such that they target the same region but from inverse directions. The use of multiple mPEAR primers can allow for analysis of several regions of interest simultaneously.

[00199] mPEAR primers can hybridize to a target fragment DNA that has been joined or ligated to a known sequence. In some embodiments, the two separate regions that bind increase the specificity of the mPEAR primer.

[00200] In some embodiments, the universal 5' end of the mPEAR primer can serve to improve downstream sequencing. For example, the universal 5' end may stabilize the synthetic oligonucleotides toward the end of DNA library fragments. The universal 5' end may increase sequencer efficiency. The universal 5' end may keep a sequence, such as an anchor sequence, toward the beginning of the read. In downstream applications such as sequencing, this can, in some instances, allow a sequencer to appropriately position the target sequence without wasted sequencer capacity.

[00201] In some embodiments, blocking polynucleotides, such as blocking oligonucleotides or blocking oligos, can be used. In some embodiments, the primers can bind and cover the universal sequence at the 5' end thereby optionally reducing or eliminating the need for additional blocking oligos.

[00202] The spacer sequence consisting of degenerate nucleotides may be synthesized adjacent to the universal sequence. The number of degenerate nucleotides can be variable. Degenerate nucleotides can allow for some flexibility in the DNA library start and stop positions during DNA sequencing. This can allow for flexibility in the design of the locus specific priming sites. The presence of variable sequence start sites may help avoid systematic errors in the sequencing step, and can allow for a randomized error profile across the reads of the redundant DNA library fragments.

[00203] The locus specific priming site can be designed to recognize DNA sequences that are upstream of the actual target sequence. The locus specific region can be designed to be as short as possible without loss of function. In this context, 'function' refers to the ability of the double-stranded region to form a stable duplex under standard reaction conditions for an enzyme-catalyzed nucleic acid primer extension reaction, (e.g. incubation at a temperature in the range of 4° C to 60° C in an annealing buffer appropriate for the enzyme), such that the two strands forming the adaptor remain partially annealed during extension of the primer to a target molecule.

[00204] mPEAR primer regions may be designed to prevent self-annealing in a variety of ways. In some cases, the mPEAR primer can take the form such that either the universal priming site or the locus specific site may be longer than the one other. In such cases there is a single stranded region on one of the strands, or a sequence selected such that the two strands do not hybridize, and thus form a continuous single stranded primer. In some cases, sequence may be designed such that they anneal in 'bubbles' conformations, wherein both ends of the mPEAR primer construct(s) are capable of hybridizing to each other and forming a duplex, but the central region cannot for a duplex. The portion of the strand(s) forming the central region are not annealed under conditions in which other

IDT Exhibit 1020.3176

portions of the same two strands are annealed to form one or more double-stranded regions. In some instances, the longer the length of the mPEAR primer can correlate with the possibility that the mPEAR is able to base-pair to itself. Therefore, in some embodiments, the length can be reduced in order to reduce this effect. In some embodiments, the stability can also be increased by the inclusion of non-natural nucleotides which exhibit stronger base-pairing than standard Watson-Crick base pairs.

[00205] The actual nucleotide sequence of the mPEAR primers may be any suitable and may be selected by the user such that the desired sequence elements are ultimately included in the common sequences of the library of templates derived from the primers to, for example, provide binding sites for particular sets of universal amplification primers and/or sequencing primers. Additional sequence elements may be included, for example, to provide binding sites for sequencing primers which will ultimately be used in sequencing of template molecules in the library, or products derived from amplification of the template library, for example on a solid support in sequencing applications.

[00206] Generally mPEAR sequences may comprise DNA, but may also include any nucleotides or nucleotide derivatives that may be suitable. Alternative nucleotides may include mixtures of natural and non-natural nucleotides (e.g. one or more ribonucleotides) linked by a mixture of phosphodiester and non-phosphodiester backbone linkages. Other non-nucleotide modifications may be included such as, for example, biotin moieties, blocking groups and capture moieties for attachment, such as biotinylated nucleotides.

vi. mPEAR Elongation

[00207] A hybridized mPEAR primer can be extended to amplify all or a portion of the target fragment. A primer extension reaction may be performed using any suitable polymerase, **330,** as described herein. Primer extension reactions are well known in the art and may include any suitable reagents for reaction. Selection of polymerases can be based on different criteria, including length of primer extension, enzyme fidelity, speed, turnover rate and the like. In some cases, Klenow or Klenow fragments may be suitable for a primer extension reaction.

[00208] Primer extension reaction, **370** conditions may vary and any combination of cycles, times and temperature may be used to perform one or more reactions. Reaction conditions may generally vary based on various parameters of primer design, including melting temperatures, predicted dimer-dimer formations, average extension length and the like.

[00209] In some embodiments, the extension reaction can be carried out in the presence of marked nucleotides. In some embodiments, some portion of the nucleotides in the reaction mix are marked with an affinity conjugate such as biotin, **340**. The extension can occur through an elongation reaction using, e.g. a polymerase. The extension reaction can produce a complement to the targeted fragment. In some embodiments, the complement to the targeted fragment is marked with biotinylated nucleotides. The resulting primer extension products can comprise a library of template

IDT Exhibit 1020.3177

polynucleotides. In some cases, the primer itself may be biotinylated. In some embodiments, the library of template polynucleotides can be separated from other fragments in the mix using streptavidin.an immobilized surface with an agent used to bind the affinity conjugate for purification of the target molecule as shown in **FIG. 3B**. In some cases, the immobilized surface may comprise streptavidin to affinity purify the conjugate biotin. In some cases, the immobilized surfaces may be streptavidin coated beads which may be purified via used of a magnet, **380**. In some cases the primer extension with affinity conjugate provides for more efficient or more specific capture of the probe than with current techniques in the art. In some cases current methods rely on affinity via hydrogen bonds of synthetic sequence to patient derived sequence. The synthetic sequence is designed by using the reference genome. In some cases patient derived sequence may be different than the synthetic sequence and in some cases this limits the binding efficiency.

vii. Library

[00210] The library of template polynucleotides can be amplified (e.g. by PCR or linear amplification reactions). Amplification can occur using primers that can hybridize to the known sequence at the 5' and 3' ends.

[00211] The contents of an amplification reaction are generally known in the art and may include appropriate substrates (such as dNTPs), enzymes (e.g. a DNA polymerase) and buffer components required for an amplification reaction. PCR amplification reactions can require at two amplification primers, often denoted 'forward' and 'reverse' primers (primer oligonucleotides) that are capable of annealing specifically to a part of the polynucleotide sequence to be amplified under conditions encountered in the primer annealing step of each cycle of an amplification reaction. In certain embodiments the forward and reverse primers may be identical. In linear amplification, one primer can be required.

[00212] During the amplification step, amplification may be performed using degenerate or universal primers for all samples or a forward primer that has sequences specific for the target polynucleotide, i.e. locus specific sequences).

[00213] Amplification methods include but are not limited to, polymerase chain reaction (PCR) (U.S. Pat. Nos. 4,683,195; and 4,683,202; PCR Technology: Principles and Applications for DNA Amplification, ed. H. A. Erlich, Freeman Press, NY, N.Y., 1992), ligase chain reaction (LCR) (Wu and Wallace, Genomics 4:560, 1989; Landegren et al., Science 241:1077, 1988), strand displacement amplification (SDA) (U.S. Pat. Nos. 5,270,184; and 5,422,252), transcription-mediated amplification (TMA) (U.S. Pat. No. 5,399,491), linked linear amplification (LLA) (U.S. Pat. No. 6,027,923), and the like, self-sustained sequence replication (Guatelli et al., Proc. Nat. Acad. Sci. USA, 87, 1874 (1990) and WO90/06995), selective amplification of target polynucleotide sequences (U.S. Pat. No. 6,410,276), consensus sequence primed polymerase chain reaction (CP-PCR) (U.S. Pat. No.

IDT Exhibit 1020.3178

4,437,975), arbitrarily primed polymerase chain reaction (AP-PCR) (U.S. Pat. Nos. 5,413,909, 5,861,245) and nucleic acid based sequence amplification (NASBA). (See, U.S. Pat. Nos. 5,409,818, 5,554,517, and 6,063,603, each of which is incorporated herein by reference). Other amplification methods that may be used include: Qbeta Replicase, described in PCT Patent Application No. PCT/US87/00880, isothermal amplification methods such as SDA, described in Walker et al., Nucleic Acids Res. 20(7):1691-6 (1992), and rolling circle amplification, described in U.S. Pat. No. 5,648,245. Other amplification methods that may be used are described in, U.S. Pat. Nos. 5,242,794, 5,494,810, 4,988,617 and in U.S. Ser. No. 09/854,317 and US Pub. No. 20030143599, each of which is incorporated herein by reference. In some aspects DNA is amplified by multiplex locus-specific PCR. In some aspects the DNA is amplified using adaptor-ligation and single primer PCR. Other available methods of amplification, such as balanced PCR (Makrigiorgos, et al., Nature Biotechnol, 20:936-9 (2002)) and isothermal amplification methods such as nucleic acid sequence based amplification (NASBA) and self-sustained sequence replication (Guatelli et al., PNAS USA 87:1874 (1990)). Based on such methodologies, a person skilled in the art readily can design primers in any suitable regions 5′ and 3′ to a locus of interest. Such primers may be used to amplify DNA of any length so long that it contains the locus of interest in its sequence.

[00214] An amplification reaction may be performed with more than two amplification primers. In order to prevent the amplification of mPEAR dimers, the amplification primers can be modified to contain nucleotides that hybridize across the whole of the primer extension product and into the target molecule template (or the dNTP's attached to the 3′ end thereof). A first amplification primer can be modified and treated to help prevent exonuclease digestion of the strands. A first amplification primer that is universal can amplify all samples rather than modifying and treating each of the tagged primers separately. A tagged primer can be introduced as a sample specific third primer in the amplification reaction, but does not need to be specially modified and treated to reduce exonuclease digestion. A third amplification primer that carries a tag can comprise a sequence that is the same as at least a portion of the first amplification primer such that it can be used to amplify the duplex resulting from extension of the first amplification primer.

[00215] A primer extension may be performed on one or both strands of the target molecule template. Primer extensions and subsequent amplifications can run through the end of the DNA library molecule. The use of an enzymatic primer extension on both strands of DNA can be adventageous. Two reactions targeting the same sequence can increase specificity and can reduce failure rates.

[00216] Amplification primers can be different lengths. In the case of nested PCR, the three or more amplification primers can be designed to be longer than the primer used to amplify the previous amplicon, so the length of the added nucleotides is fully controllable and may be hundreds of nucleotides if desired.

IDT Exhibit 1020.3179

[00217] The forward and reverse primers can be of sufficient length to hybridize to the whole of the universal adaptor sequence and at least one base of the target sequence (or the nucleotide dNTP added as a 3′-overhang on the target strands). The forward and reverse primers can also contain a region that can extend beyond the adaptor construct. In some embodiments, the amplification primers can be at least 10, 20, 30, 40, 50, 60, 70, 80, 90 100, 150, 200, 300, 400 or 500 bases in length. In other embodiments, amplification primers may be at most 10, 20, 30, 40, 50, 60, 70, 80, 90 100, 150, 200, 300, 400 or 500 bases in length. The forward and reverse primers can be of significantly different lengths. In some embodiments, a first primer may be 20-40 bases, whereas a second primer may be 40-100 bases in length. The nucleotide sequences of the adaptor-target specific portions of the forward and reverse primers can be selected to achieve specific hybridization to the adaptor-target sequences to be amplified under the conditions of the annealing steps of the amplification reaction, while minimizing non-specific hybridization to any other target sequences present.

[00218] Amplification primers are generally single stranded polynucleotide structures. They can contain a mixture of natural and non-natural bases and also natural and non-natural backbone linkages.

[00219] Primers can contain sequences specific for capture on various platforms. In some cases, a sequence may be incorporated to allow hybridization to a known sequence in various high through platform kits, such as one provided by Illumina. Incorporation of hybridization sequences for parallel loading of samples onto a surface platform for sequencing is known in the art.

[00220] Primers can comprise non-nucleotide chemical modifications, for example phosphorothioates to increase exonuclease resistance, again provided such that modifications do not prevent primer function. Modifications may, for example, facilitate attachment, of the primer to a solid support, for example a biotin moiety. Certain modifications may themselves improve the function of the molecule as a primer, or may provide some other useful functionality, such as providing a site for cleavage to enable the primer (or an extended polynucleotide strand derived therefrom) to be cleaved.

[00221] Amplifications can be carried out on either the pooled or unpooled samples. Tags can be part of the amplification primers. In some embodiments, each sample can be amplified independently prior to pooling. The pooled nucleic acid samples can be processed for sequencing.

[00222] Amplification steps can be used to produce high quantities of higher quality sample. Amplification steps can be used to incorporate additional barcodes or adapter sequences with target polynucleotide sequences.

viii. Sample capture

[00223] The pooled or unpooled sample can be captured in preparation for sequencing. Sequencing can be performed as an array of single captured targets. The amplification products can be attached on

IDT Exhibit 1020.3180

a planar surface, or on a pool of beads. The pool of beads can be isolated into an emulsion with a single bead in each 'partition' of the emulsion. At a concentration of only one template per 'partition', only a single template is amplified on each bead. In some embodiments, the methods of mPEAR targeting can be used in combination with the methods of genome RAPELLing.

[00224] Any chemical means for capture of amplified product may be suitable. In some embodiments, a single point covalent attachment to a solid support at or near the 5′ end of the primer, leaving the template-specific portion of the primer free to anneal to its cognate template and the 3′ hydroxyl group free for primer extension is suitable for capture. Any suitable covalent attachment means known in the art may be used for this purpose. The chosen attachment chemistry will depend on the nature of the solid support, and any derivative or functionalization applied to it. The primer itself may include a moiety, which may be a non-nucleotide chemical modification, to facilitate attachment. In a particular embodiment, the primer may include a sulphur-containing nucleophile, such as phosphorothioate or thiophosphate, at the 5′ end.

[00225] In other embodiments, capture may be achieved through biotin-streptavidin, or streptavidin derivatives, interactions. Amplified products that comprise biotin may be incubated with surfaces with streptavidin or streptavidin derivatives, thus allowing products to be immobilized as shown in **FIG. 3B**. Purification methods, known in the art may be used to retrieve amplification products, such as through the use of magnets and/or streptavidin coated beads. In some embodiments, additional wash steps can be used. Amplification products can be eluted by flushing with excess concentrations of biotin or biotin related compounds as shown in **FIG. 3C**. Methods known in the art for biotin-streptavidin affinity purification of polynucleotides (see US5405746, US5500356 and US5759778) are incorporated by reference in their entirety.

ix. Sequencing and data analysis

[00226] mPEAR samples can be sequenced by any method known in the art, several non-limiting examples are disclosed herein. Sequencing can produce data. The data can be stored, processed, and transmitted as disclosed herein.

[00227] The read structure, as depicted in **FIG. 9** of the captured molecules can offer significant advantages in speed and quality of data analysis. Because the locus specific primer recognition site can be designed to hybridize upstream of the target sequence, that locus specific site may be used to identify the genomic location of the read. This can be generally referred to as "anchored read alignment" and can greatly reduce the amount of data processing and statistical analysis. In one non-limiting example, instead of taking short sequence reads and in silico aligning them to the entire reference genome, reads can be automatically binned or associated to their correct genomic position. This can greatly reduce the time for data analysis, increase accuracy, and decrease the computational power needed for such analysis.

IDT Exhibit 1020.3181

[00228] In addition the mPEAR method can allow for true de novo sequencing of polynucleotide targets. Thus, by "anchoring" the read with the known sequence, elongation product can be assembled without the use of a reference genome. Thus mPEAR allows for greater sensitivity of genomic loci that greatly differ from the reference genome. Longer insertions, larger deletions, and repeats that are clinically relevant can be detected with greater sensitivity and accuracy. Viral insertion sites and/or mobile polynucleotide elements can be detected and localized and this process can be multiplexed for greater efficiency.

[00229] The optional addition of one or more barcodes can allow for, e.g., binning of the sample to the sequence read. Barcoding can be useful for a variety of applications, including tracking of individual polynucleotide molecules, as shown in **FIG. 13**. In some embodiments, DNA molecule reads can be associated to a sample. In some embodiments, mPEAR barcoding can be used to phase information, in which individual molecules may be identified as paternally or maternally inherited from a single individual.

[00230] E. Single end adapter library and rolling circle amplification

[00231] Rolling circle amplification can be used with targeting methods. Rolling circle amplification can be used to generate a linear amplification reaction.

i. Fragmentation

[00232] A polynucleotide sample, **1200** may be first fragmented, **1210** before subsequent steps as shown in **FIG. 12**. Fragmentation methods have been described herein. The size of the polynucleotide fragments, described in terms of length, may vary depending on the source of the target polynucleotide, the method used for fragmentation, and the desired application. In some cases one or more fragmentation steps may be used. For example, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, or more fragmentation steps may be used.

[00233] In some cases nucleic acids may be fragmented into sizes atleast 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500 , 600, 700, 800, 900, 1000, 2000, 3000, 5000 base pairs in length. In some cases nucleic acids may be fragmented into sizes at most 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500 , 600, 700, 800, 900, 1000, 2000, 3000, 5000 base pairs in length.

[00234] Numerous fragmentation methods are described herein and known in the art. For example, fragmentation may be performed through physical, mechanical or enzymatic methods. Physical fragmentation may include exposing a target polynucleotide to heat or to UV light. Mechanical disruption may be used to mechanically shear a target polynucleotide into fragments of the desired range. Mechanical shearing may be accomplished through a number of methods known in the art, including repetitive pipetting of the target

IDT Exhibit 1020.3182

polynucleotide, sonication and nebulization. Target polynucleotides may also be fragmented using enzymatic methods. In some cases, enzymatic digestion may be performed using enzymes such as using restriction enzymes.

[00235] Restriction enzymes may be used to perform specific or non-specific fragmentation of target polynucleotides. The methods of the present disclosure may use one or more types of restriction enzymes, generally described as Type I enzymes, Type II enzymes, and/or Type III enzymes. Type II and Type III enzymes are generally commercially available and well known in the art. Type II and Type III enzymes recognize specific sequences of nucleotide base pairs within a double stranded polynucleotide sequence (a "recognition sequence" or "recognition site"). Upon binding and recognition of these sequences, Type II and Type III enzymes cleave the polynucleotide sequence. In some cases, cleavage will result in a polynucleotide fragment with a portion of overhanging single stranded DNA, called a "sticky end." In other cases, cleavage will not result in a fragment with an overhang, creating a "blunt end." The methods of the present disclosure may comprise use of restriction enzymes that generate either sticky ends or blunt ends.

[00236] Restriction enzymes may recognize a variety of recognition sites in the target polynucleotide. Some restriction enzymes ("exact cutters") recognize only a single recognition site (e.g., GAATTC). Other restriction enzymes are more promiscuous, and recognize more than one recognition site, or a variety of recognition sites. Some enzymes cut at a single position within the recognition site, while others may cut at multiple positions. Some enzymes cut at the same position within the recognition site, while others cut at variable positions.

ii. Nucleic Acid Strand End Repair

[00237] In many cases, fragmentation of nucleic acids, such as through mechanical shearing or enzymatic digestion results in fragments with a heterogeneous mix of blunt and 3′- and 5′-overhanging ends. In some cases, the compositions and methods of the disclosure provide for repair of the fragment ends using methods or kits (i.e. Lucigen DNA terminator End Repair Kit) known in the art to generate ends that are designed for insertion, for example, into blunt sites of cloning vectors. In some cases, the compositions and methods of the disclosure provide for blunt ended fragment ends of the population of DNAs sequenced. Further, in some cases, the blunt ended fragment may also be phosphorylated. The phosphate moiety can

IDT Exhibit 1020.3183

be introduced via enzymatic treatment, for example, using a kinase, (i.e. shrimp alkaline kinase).

[00238] In other cases, polynucleotide sequences are prepared with single overhanging nucleotides by, for example, activity of certain types of DNA polymerase such as Taq polymerase or Klenow exo minus polymerase which has a nontemplate-dependent terminal transferase activity that adds a single deoxynucleotide, for example, deoxyadenosine (A) to the 3′ ends of, for example, PCR products. Such enzymes can be utilized to add a single nucleotide 'A' to the blunt ended 3′ terminus of each strand of the target polynucleotide duplexes. Thus, an 'A' could be added to the 3′ terminus of each end repaired duplex strand of the target polynucleotide duplex by reaction with Taq or Klenow exo minus polymerase, whilst the adaptor polynucleotide construct could be a T-construct with a compatible 'T' overhang present on the 3′ terminus of each duplex region of the adaptor construct. This end modification also prevents self-ligation of both adapter and target such that there is a bias towards formation of the combined ligated adaptor-target sequences.

[00239] In some cases, Nextera kits such as provided by Illumina/Epicentre, which use a tn5 transposase to simultaneously fragment the double-stranded DNA and ligate adapters to the ends of the fragments may be used. For example, the amplified cDNA is 'tagmentated' at 55 °C for 5 min in a 20-µl reaction with 0.25 µl of transposase and 4 µl of 5× HMW Nextera reaction buffer (containing Illumina-compatible adapters). To strip the transposase off the DNA, 35 µl of PB is then added the tagmentation reaction mix, and the tagmentated DNA was purified with 88 µl of SPRI XP beads (sample to beads ratio of 1:1.6). The reagents for this method are available in Nextera DNA sample kits (Epicentre/Illumina). Alternative kits may also be used, such as provided by Roche FLX and Titanium sequencing systems.

[00240] In some cases, cDNA fragmentation may not be performed. Rather, RNA molecules, before reverse transcription to cDNA, may be fragmented using any suitable method, including applicable techniques described herein and as described by Hashimony et al Hashimshony, 2012.

[00241] In some cases, the fragmented DNA is size-selected using agarose gel methods such as SizeSelect™ Gels (Life Technologies) or Pippin Prep™ kits or beads such as AMPure XP (Beckman Coulter). In other embodiments, fragmented DNA is end repaired or polynucleotide tailed for subsequent steps of library preparation.

[00242] Fragmentation of polynucleotides, such as through mechanical shearing or enzymatic digestion, can result in fragments, **1230** with a heterogeneous mix of blunt and 3′- and 5′-overhanging

IDT Exhibit 1020.3184

ends. In some cases, the fragment ends can be repaired or treated, **1220**. Such methods for polynucleotide strand end repair have been described herein.

iii. Attachment of known sequence

[00243] Known sequences can be attached to the ends of the fragmented polynucleotide sample. Methods for attachment of known sequences (e.g. ligation) have been described herein.

[00244] The known sequence, **1235,** can comprise a molecular barcode, one or more adapter(s), or any other sequence such as a universal primer sequence. In some embodiments, the known sequence comprises two adapters, an A adapter **1240**, and a B adapter, **1245** that can sit in an "inverted" matter. The "inverted" manner can mean that the 5' end of the A adapter can be linked to the 5' end of the B adapter. In some embodiments, a restriction enzyme or other nuclease site may be engineered between the A and B adapters.

iv. Ligation

[00245] The fragmented polynucleotide attached to the known sequence can be circularized. Circularization can be accomplished through ligation, **1250**. In some embodiments, the 5' end of the fragment is ligated to the 3' end of the known sequence. In other embodiments, the 3' end of the fragment is ligated to the 5' end of the known sequence.

v. Amplification

[00246] The circularized polynucleotide can be amplified through a rolling circle amplification process. In such a process, a primer, **1255** can hybridize to of the circularized polynucleotide, **1260**. A polymerase can elongate and copy the circularized template, i.e. rolling circle amplification. The polymerase can copy the circularized template 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500 times or more. This can result in a linear copy that comprises one or more sequential linear copies, **1265**. In some embodiments there can be 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500 copies in a linear polynucleotide. These copies can be cleaved by a restriction enzyme or nuclease to create a library for sequencing. In some embodiments, PCR amplification methods can be used to add full length adapter sequences to the 5' and 3' ends, thereby producing a library for sequencing. Methods for using PCR primers to add known sequences to an amplified product have been disclosed herein.

[00247] In some cases, tiled locus specific primers can be used for amplifying circular molecules that contain the sequences of interest. Primers may be designed upstream of targets and tiled across the full length of the target. Primers may be designed in both directions as well. Double stranded

IDT Exhibit 1020.3185

DNA "circles" may be de natured and locus specific primers bind (both directions). As rolling circle amplification occurs, amplified product may be displaced from the template molecule. Excess primers may bind to the growing strand and amplify the extended copy in the alternate direction to form long double stranded DNA molecules with repetitive sequence. The repetitive sequence may include A and B adapter sequences flanking the target sequence. This can now be used as a template for PCR to amplify in the full length sequencing adapters. In other cases a restriction site in between the adapters becomes active when the ends of the molecule are joined.

vi. Sequencing and data analysis

[00248] Rolling circle amplification samples can be sequenced by any method known in the art; several non-limiting examples are disclosed herein. Sequencing can produce data. The data can be stored, processed, and transmitted as disclosed herein.

II. DNA LABELING FOR DE NOVO SEQUENCING AND LABELING

A. Random Primer Extension, Ligation and Labeling (RAPELL)

[00249] Methods of random primer extension ligation and labeling (RAPELL) can be used for obtaining sequence information of long fragments of nucleotides using short read systems. RAPELL methods can comprise: polynucleotide isolation, purification, dilution, and spatial separation before amplification and incorporation of a molecular label and finally sequencing. **FIG 1** depicts an exemplary method according to the disclosure. **FIG 1A** shows a process **100** for obtaining sequence information of long fragments of nucleotides using short read systems using the mPEAR system, as described herein. The process **100**comprises obtaining a sample of long nucleic acid **105** of high molecular weight (over 5 kilobases). **FIG 1B** shows a process for further processing of high molecular weight nucleic acid. The long nucleic acid **105** is diluted and spatially separated **110** in sub genome quantities into several partitions (e.g. **115**, **120**, **125**). Each partition **115**, **120**, **125** may contain a long nucleic acid **105**. In each partition, the long nucleic acid fragment **105** comes into contact with a polymerase/ligase mixture, containing primers **130**, and adapters **135**. The primers **130** can comprise a random sequence at the three prime end to allow for random binding along the long nucleic acid fragment **105** and a region of know sequence (represented by a circle and a dashed line and a 3' cap (represented by a diamond shape). A polymerase reaction **140** extends the primers **130** are along various random areas of the template nucleic acids until the elongation product (e.g. **142**, **144**) reaches a downstream adaptor **135**. A ligase **150** ligates **155** the elongation product to the downstream adaptor and an amplicon library (e.g. **146**, **148**) is created. A second set of primers (e.g.

IDT Exhibit 1020.3186

165, **170**) can be added **160** for a polymerase chain reaction (PCR) **175**, thereby producing a library suitable for sequencing **180**.

i. Sample Acquisition

[00250] RAPELL methods can use a polynucleotide of long molecular length. The sample can come from a nucleic acid library, such as a cDNA library. The sample can come from genomic DNA. The nucleic acid can also be isolated from one or more subjects. In some exemplary methods, the nucleic acid is deoxynucleic acid (DNA) of high molecular weight. High molecular weight can refer to, for example approximately more than 0.5, 1, 3, 4, 5, 10, 15, 20, 30, 35, 40, 50, 60, 70, 80, 90, 100, 200 or more kilobases. The nucleic acid sample can contain nucleic acid fragments that range in size from approximately 1-50 kb, 5-50 kb, 5-100 kb, 20-90 kb, 50- 100 kb, 5- 200 kb. The nucleic acid can be largely purified from cellular components. The sample of nucleic acid may be in chromosomal form. In some instances, the polynucleotide may be fragmented in to smaller sizes. Methods for polynucleotide fragmentation have been disclosed herein. In some instances, methods known in the art (e.g. physical shearing or enzymatic digestion) may be used to fragment the chromosomal nucleic acid into sizes such as approximately 1, 3, 4, 5, 10, 15, 20, 30, 35, 40, 50, 60, 70, 80, 90, 100, 200 kb or into sizes that range from -50 kb, 5-50 kb, 5-100 kb, 20-90 kb, 50- 100 kb, 5- 200 kb.

ii. Dilution and spatial separation

[00251] The present invention uses dilution and spatial separation of sample nucleic acid. In some instances, long fragments of nucleic acid are diluted before being spatially separated. Dilution can be accomplished by any method known in the art, such as by the addition of a dilute, such as water, or a suitable buffer. An exemplary method of dilution involves determining the concentration of the nucleic acid before dilution and calculating how much dilute to add so that the diluted sample can be partitioned into quantities that contain sub-genomic quantities of DNA (i.e. so that one sample contains less than one whole genome). In another exemplary method, dilution can be calculated so that the sample can be partitioned in a way that each partition contains approximately 1, 2, 3, 5, 10, 20, 50, 80, 100, 150, 200, 400, 500, 1000, 1500, 5,000, 10,000, 100,000, 1,000,000, 10,000,000, 100,000,000, 1,000,000,000 fragments of nucleic acid. In another exemplary method, dilution is accomplished to facilitate partitioning sample so that approximately 1%, 2%, 5%, 10%, 15%, 20%, 30%, 50%, 70%, 80%, 95% or 100% of one copy of the genome can be separated.

[00252] Spatial separation can be accomplished by many methods known in the art, such as pipetting, micropipeting, or microfluidics. Partitions can be made of any known methods in the art, including plates (e.g. 96-well), mirofluidic chambers, micro-droplets, or simple spatial separation on solid surfaces such as silicon chips or beads.

IDT Exhibit 1020.3187

[00253] In an exemplary method, dilution and spatial separation is conducted so that the there is a low probability that two partitions contain the same locus of DNA from each parental chromosome, or that multiple fragments from the same genomic locus will be extremely rare.

iii. Primer extension, ligation and amplification

[00254] In a partition, the diluted polynucleotide can come into contact with a mixture of synthetic oligonucleotids, native dNTPs, polymerase (or polymerase fragments), ligase and associated buffers sufficient for primer extension and ligation. The said mixture of synthetic oligonucleotides is comprised of a donor primer and an acceptor probe.

a. Donor primer

[00255] The present invention can use a donor primer to generate complementary regions of the diluted template. **FIG 22A** is an illustration diagramming an exemplary donor primer **200**. A donor primer may comprise: an adapter sequence **205**, a molecular label (i.e. bar code) **210** and a region of random primer **215**. A donor primer can be made of nucleotides comprising DNA nucleotides, RNA nucleotides, or any combination thereof.

[00256] An adapter sequence **205** can be located near the 5' end of the donor primer **200**. An adapter sequence can have a length of approximately 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, 55, 60, 75, 80, 85, 90, 95, 100, 125, 130, 150, 200 or more nucleotides. In some instances the adapter sequence **205** may be designed with such a sequence that secondary structure may form such as a hairpin, or stem-loop. To facilitate secondary structure formation and release one or more uracil bases can be added. An adaptor that can form a secondary structure may be used in some instances to reduce binding of donor primers with RAPELL products. The secondary structure may be designed such that it can be selectively eliminated by cutting. Selective cutting of the hairpin may be accomplished by the use of an enzyme, such as a nuclease, e.g. Drosha.

[00257] A molecular barcode **210** can be designed to designate which partition the reaction takes place in. Therefore the number of molecular barcodes may equal the number of partitions used for the reaction. In one non limiting example, 96 different donor primers could be used, each with a different molecular label if nucleotide fragments were separated into 96-well plate partitions. The molecular label can be 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 or more nucleotides in length.

[00258] The 3' prime end of the donor primer can comprise a random sequence that can act as a random primer. Random primers can be short segments of nucleic acid that can hybridize to a template and prime a reaction. The random region can comprise 6, 7, 8, 9, 10, or more nucleotides. In one non-limiting example, the region can be 6 nucleotides long (i.e. a hexamer) and therefore the

IDT Exhibit 1020.3188

mixture of primers can contain every possible combination of bases (4^6=4,096 total possible combinations). In another non-limiting example, the region can be 8 nucleotides long (i.e. an octamer) and the mixture of donor primers can contain every possible combination of bases (4^8 = 65,563 total possible combinations).

b. Acceptor probe

[00259] The present invention can use an acceptor probe **240** that can bind to the template. **FIG 2B** is an illustration diagramming an exemplary acceptor probe **240**. An acceptor probe may comprise: an adapter sequence **230**, a molecular barcode **225** and a region of random primer **220**. An acceptor probe can be made of nucleotides comprising DNA nucleotides, RNA nucleotides, or any combination thereof. An acceptor probe **240** may contain a 3' cap **235** on the 3'-most nucleotide to prevent extension or elongation from the acceptor probe **240**.

[00260] The 5' prime end of the acceptor probe **240** comprises a random primer **230**. Random primers **230** can be short segments of nucleic acid that consist of every possible combination of bases. The random region can comprise 6, 7, 8, 9, 10, or more nucleotides. In one non-limiting example, the region can be 6 nucleotides long (i.e. a hexamer) and therefore the mixture of primers can contain every possible combination of bases (4^6=4,096 total possible combinations). In another non-limiting example, the region can be 8 nucleotides long (i.e. an octamer) and the mixture of acceptor probes can contain every possible combination of bases (4^8 = 65,563 total possible combinations).

[00261] A molecular bar code **310** can be designed to designate which partition the reaction takes place in. Therefore the number of molecular barcodes **310** needed for each acceptor probe **300** may equal the number of partitions used for the reaction. In one non limiting example, 96 different donor primers could be used, each with a different molecular label if nucleotide fragments were separated into 96-well plate partitions. The molecular label can be 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 or more nucleotides in length. The adaptor probe **240** and donor primer **200** with the same molecular label may paired and used in the same partition. In some embodiments, the adaptor probe **300240** and the donor primer **200** will be paired so that their molecular labels are different and used in the same partition.

[00262] An adapter sequence **230** can be located near the 3' end of the adapter probe **300**. An adapter sequence may have a length of approximately 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, 55, 60, 75, 80, 85, 90, 95, 100, 125, 130, 150, 200 or more nucleotides. In some instances the adapter sequence **230** may be designed with such a sequence that secondary structure may form such as a hairpin, or stem-loop. To facilitate secondary structure formation and release one or more uracil bases may be added. An adaptor that can form a secondary structure may be used in some instances to reduce binding of acceptor probes **240** with RAPELL products. The secondary structure may be designed such that it can be selectively

IDT Exhibit 1020.3189

eliminated by cutting. Selective cutting of the hairpin may be accomplished by the use of an enzyme, such as Drosha.

iv. Binding and elongation

[00263] **FIG 4** illustrates an exemplary process of binding and elongation **400** of donor primers **405** and acceptor probes **410**. The random primer region of the acceptor probe **410** and the donor primer **405** can bind to regions along the template. Since the mixture of donor primers **405** and acceptor probes **410** may have, for example, every combination of hexamer or octamer bases, the acceptor probes and donor primers can anneal throughout the template **415** in a manner that can produce statistically random RAPELL products **430**. Once the donor primer **405** and the acceptor probe **410** have annealed, a non-strand displacing polymerase **420** can extend the donor primer **405** until it reaches the acceptor probe **410**. A non-strand displacing polymerase can be used to fully extend. Once the extension has reached the acceptor probe **410**, a ligase **425** may be used to ligate the extension product and the acceptor probe **410** to create a RAPELL product **430**. The methods of binding and elongation can be repeated in the same partition to produce a plurality of RAPELL products **430**. In some instances the plurality of RAPELL products can comprise hundreds, thousands, or millions, of short copies of the template fragment. The resulting RAPELL products **430** can be of variable length and can be released from the template by melting.

[00264] In some examples a preamplification step may be used. For instance by ligating adapters to the ends of long DNA fragments, spatially separating the fragments and then performing long range PCR. Then the random RAPELL primers are used as described. Targeted long range PCR may also be performed by using the multiplexed PCR preamplification previously described.

v. RAPELL product

[00265] The methods disclosed herein can generate a RAPELL product **430**. The RAPELL product can be comprised of: the donor primer **405** and the acceptor probe **410**. The RAPELL product can have adaptor regions on the 5' and 3' ends. In some instances, the RAPELL product can be processed, if, for example, a secondary structure has been generated on the 5' or 3' end. The secondary structure may be selectively eliminated by cutting. Selective cutting of a secondary structure, such as a hairpin, can be accomplished by the use of an enzyme, such as Drosha.

vi. Amplification

[00266] **FIG 5** illustrates an exemplary method **500** of how a RAPELL product **530** can be amplified by polymerase chain reaction (PCR). To amplify a RAPELL product **530** primers complementary to the adaptor sequences (e.g. integrated into the RAPELL product through the donor primer **505** and

IDT Exhibit 1020.3190

acceptor probe **510**) can be used (e.g. Sequencer primer A **515** and Sequencer primer B **520**). In some instances a RAPELL product **530** can be amplified to produce a plurality of RAPELL products for a sequencing library **525**. A plurality of RAPELL products **530** can be amplified in the same reaction chamber using. In some embodiments, RAPELL products **530** that exist in separate partitions (see **FIG 1**) may be pooled rather than running a reaction in each separate partition.

[00267] The Sequencer primers (e.g. Sequencer primer A **515** and Sequencer primer B **520**) can comprise: a first region that can bind to the adaptor sequences (e.g. 505 and 510) and a region suitable for a specific sequencing platform adapter sequence (e.g. an Illumina sequence). Cycling conditions of the amplification reaction can be optimized or adjusted to produce a desired fragment size, or fragment size range optional for sequencer performance. A size selection and quantification can be used to achieve optimal sequencer performance.

vii. Sequencing and data analysis

[00268] Each copy can contain a identifier specific to a single partition of the reaction. Each sequencer read can be anchored to a single partition and thus can be stitched to a longer template fragment. Long read lengths are achieved by combining overlapping sequences from the same molecular barcode (i.e. the same partition).

III. SEQUENCING

[00269] Numerous methods of sequence determination can be used. Exemplary methods for sequence determination include, but are not limited to, hybridization-based methods, such as disclosed in Drmanac, U.S. Pat. Nos. 6,864,052; 6,309,824; and 6,401,267; and Drmanac et al, U.S. patent publication 2005/0191656, which are incorporated by reference, sequencing by synthesis methods, e.g., Nyren et al, U.S. Pat. Nos. 7,648,824, 7,459,311 and 6,210,891; Balasubramanian, U.S. Pat. Nos. 7,232,656 and 6,833,246; Quake, U.S. Pat. No. 6,911,345; Li et al, Proc. Natl. Acad. Sci., 100: 414-419 (2003); pyrophosphate sequencing as described in Ronaghi et al., U.S. Pat. Nos. 7,648,824, 7,459,311, 6,828,100, and 6,210,891; and ligation-based sequencing determination methods, e.g., Drmanac et al., U.S. Pat. Appin No. 20100105052, and Church et al, U.S. Pat. Appin Nos. 20070207482 and 20090018024.

[00270] Sequence determination may also be determined using methods that determine many (typically thousands to billions) of nucleic acid sequences in an intrinsically parallel manner, where many sequences are read out preferably in parallel using a high throughput serial process. Such methods include but are not limited to pyrosequencing (for example, as commercialized by 454 Life Sciences, Inc., Branford, Conn.); sequencing by ligation (for example, as commercialized in the SOLiD™ technology, Life Technology, Inc., Carlsbad, Calif.); sequencing by synthesis using modified nucleotides (such as commercialized in TruSeq™ and HiSeq™ technology by Illumina, Inc.,

IDT Exhibit 1020.3191

San Diego, Calif., HeliScope™ by Helicos Biosciences Corporation, Cambridge, Mass., and PacBio RS by Pacific Biosciences of California, Inc., Menlo Park, Calif.), sequencing by ion detection technologies (Ion Torrent, Inc., South San Francisco, Calif.); sequencing of DNA nanoballs (Complete Genomics, Inc., Mountain View, Calif.); nanopore-based sequencing technologies (for example, as developed by Oxford Nanopore Technologies, LTD, Oxford, UK), and like highly parallelized sequencing methods.

[00271] Some methods of sequencing require an adapter sequence, sometimes referred to as an "adapter" or a "sequence adapter". Adapter sequence can be platform specific. Adapters can comprise an anchor. Adapters can comprise a sequencing sequence. Adapters can comprise an amplification sequence. In some embodiments, an adapter sequence can comprise an anchor, a sequencing sequence, and an amplification sequence. An adapter sequence can be added to the 5' end. An adapter sequence can be added to the 3' end. Adapter sequences can be added to both the 3' and 5' end. Adapter sequences can facilitate sequencing.

IV. SYSTEM FOR DATA TRANSMITTAL AND STORAGE

[00272] Another aspect of the invention provides a system that is configured to implement the methods of the disclosure. The system can include a computer server ("server") that is programmed to implement the methods described herein. **FIG 11** depicts a system adapted to enable a user to store, analyze, and process sequence information. The system includes a central computer server that is programmed to implement exemplary methods described herein. The server includes a central processing unit (CPU, also "processor") which can be a single core processor, a multi core processor, or plurality of processors for parallel processing. The server also includes memory (e.g. random access memory, read-only memory, flash memory); electronic storage unit (e.g. hard disk); communications interface (e.g. network adaptor) for communicating with one or more other systems; and peripheral devices which may include cache, other memory, data storage, and/or electronic display adaptors. The memory, storage unit, interface, and peripheral devices can be in communication with the processor through a communications bus (solid lines), such as a motherboard. The storage unit can be a data storage unit for storing data. The server is operatively coupled to a computer network ("network") with the aid of the communications interface. The network can be the Internet, an intranet and/or an extranet, an intranet and/or extranet that is in communication with the Internet, a telecommunication or data network. The network in some cases, with the aid of the server, can implement a peer-to-peer network, which may enable devices coupled to the server to behave as a client or a server. In some embodiments, the computing resources can be configured into a cloud-service model.

[00273] The storage unit can store files, such as sequence data, sample data, molecular barcodes, software, or any aspect of data associated with the invention. The data storage unit may be coupled

IDT Exhibit 1020.3192

with data that can bin sample sequence with the sample source or other information contained in a molecular barcode.

[00274] The server can communicate with one or more remote computer systems through the network. The one or more remote computer systems may be, for example, personal computers, laptops, tablets, telephones, smart phones, or personal digital assistants. The remote computer systems may, for example, be used to transmit patient data to a caregiver. The data or hardware or system, for example, may be encrypted or modified (e.g. to comply with HIPPA rules and standards).

[00275] In some situations the system includes a single server. In other situations, the system includes multiple servers in communication with one another through an intranet, extranet and/or the Internet.

[00276] The server can be adapted to store sample information, such as, for example, sample source, date, orientation, sequence, statistical data, or any other information of potential relevance. Such information can be stored on the storage unit or the server and such data can be transmitted through a network.

[00277] Methods as described herein can be implemented by way of machine (or computer processor) executable code (or software) stored on an electronic storage location of the server, such as, for example, on the memory, or electronic storage unit. During use, the code can be executed by the processor. In some cases, the code can be retrieved from the storage unit and stored on the memory for ready access by the processor. In some situations, the electronic storage unit can be precluded, and machine-executable instructions are stored on memory. Alternatively, the code can be executed on a second computer system.

[00278] Aspects of the systems and methods provided herein, such as the server, can be embodied in programming. Various aspects of the technology may be thought of as "products" or "articles of manufacture" typically in the form of machine (or processor) executable code and/or associated data that is carried on or embodied in a type of machine readable medium. Machine-executable code can be stored on an electronic storage unit, such memory (e.g., read-only memory, random-access memory, flash memory) or a hard disk. "Storage" type media can include any or all of the tangible memory of the computers, processors or the like, or associated modules thereof, such as various semiconductor memories, tape drives, disk drives and the like, which may provide non-transitory storage at any time for the software programming. All or portions of the software may at times be communicated through the Internet or various other telecommunication networks. Such communications, for example, may enable loading of the software from one computer or processor into another, for example, from a management server or host computer into the computer platform of an application server. Thus, another type of media that may bear the software elements includes optical, electrical, and electromagnetic waves, such as used across physical interfaces between local devices, through wired and optical landline networks and over various air-links. The physical elements that carry such waves, such as wired or wireless likes, optical links, or the like, also may be

IDT Exhibit 1020.3193

considered as media bearing the software. As used herein, unless restricted to non-transitory, tangible "storage" media, terms such as computer or machine "readable medium" refer to any medium that participates in providing instructions to a processor for execution.

[00279] Hence, a machine readable medium, such as computer-executable code, may take many forms, including but not limited to, tangible storage medium, a carrier wave medium, or physical transmission medium. Non-volatile storage media can include, for example, optical or magnetic disks, such as any of the storage devices in any computer(s) or the like, such may be used to implement the system. Tangible transmission media can include: coaxial cables, copper wires, and fiber optics (including the wires that comprise a bus within a computer system). Carrier-wave transmission media may take the form of electric or electromagnetic signals, or acoustic or light waves such as those generated during radio frequency (RF) and infrared (IR) data communications. Common forms of computer-readable media therefore include, for example: a floppy disk, a flexible disk, hard disk, magnetic tape, any other magnetic medium, a CD-ROM, DVD, DVD-ROM, any other optical medium, punch cards, paper tame, any other physical storage medium with patterns of holes, a RAM, a ROM, a PROM and EPROM, a FLASH-EPROM, any other memory chip or cartridge, a carrier wave transporting data or instructions, cables, or links transporting such carrier wave, or any other medium from which a computer may read programming code and/or data. Many of these forms of computer readable media may be involved in carrying one or more sequences of one or more instructions to a processor for execution.

[00280] The results of sequencing can be presented to a user with the aid of a user interface, such as a graphical user interface.

V. PATIENT CARE

[00281] The targeted sequencing technology may be used in sequencing of subjects in research or clinical settings. In an exemplary embodiment, the sequencing of one or more known disease causing genes can be done in a clinical setting for carrier screening. Clinical testing panels may be developed and potentially a targeted sequencing panel including interpretable locations in the genome. The technology can also be used to pick up mobile elements such as viral insertion sites or be used to identify SNPs, mutations, allelic variation or genetic modifications.

[00282] The invention described herein can provide data that can be used for a medical professional or caregiver to make care giving decisions. In some embodiments, the identification of one or more alleles, gene variants, SNPs or other genomic modifications can identify an animal, including a human, as a disease carrier. The patient status as a disease carrier can direct patient care, fertility decisions, prognosis, theranosis, monitoring, diagnosis and/or treatment of a patient or subject.

[00283] Diagnoses can comprise determining the condition of a patient. Diagnosis can be conducted at one time point or on an ongoing basis. For example, a patient can be diagnosed as being infected

IDT Exhibit 1020.3194

with a virus based on identification of viral insertion points in genomic DNA. A patient can be diagnosed with a condition based on the presence or absence of a genetic sequence.

[00284] A patient can be identified as a carrier for an autosomal recessive mutation or allele or chromosomal variant. A patient's status as a carrier can affect the patient's contraceptive choices, the patient's choices for preventative care, or the like. In some cases, carrier testing can be performed on individuals. Fertility and/or contraceptive decisions may be made based on the carrier status of an individual or couple. In some cases, measurement of fetal samples (e.g. prenatal testing) can influence treatment decisions for the mother or fetus.

[00285] Prognosis can comprise determining the outcome of a patient's disease, the chance of recovery, or how the disease will progress. For example, identifying chromosomal abnormalities can provide information upon which a prognosis may be based. For example, the presence or absences of a chromosomal variant may predict a cancer survival rate.

[00286] Monitoring can comprise the serial testing of a patient to determine the presence of a disease or to monitor the disease progression. For example, an individual who has been infected by a virus can undergo serial monitoring to determine if antiviral treatments are preventing the further infection of the individual.

[00287] Theranosis can comprise determining a therapy treatment. For example, a patient's therapy treatment can be determined, in part or in whole, based on the presence or absence of certain genomic elements. For example, the presence or absence of certain genomic variants within an individual's cytochrome p450 genes may affect drug metabolism rates and therefore can affect the type of treatment for an individual.

[00288] Clinical testing can be done using the methods described herein. In some instances, one or more protocols may be developed to comply with Clinical Laboratory Improvement Amendments (CLIA) or Food and Drug Administration (FDA) regulations.

VI. CLINICAL OR LABORATORY RESEARCH

[00289] Methods, kits and/or compositions may be used in clinical or laboratory research settings to study the genetic basis of disease, e.g. identify new genetic variations that can contribute to disease, likelihood of disease, or other conditions. Methods can be used to study viral infection, vaccine effectiveness. Methods, kits and/or compositions can be used to study regions of the genome that have traditionally been difficult to accurately sequence, such as highly polymorphic or variable regions within the genome.

VII. KITS

[00290] The reagents for carrying out the methods and assays of the present invention are optionally provided in a kit form to facilitate the application of these assays for the user. Such kits also typically

IDT Exhibit 1020.3195

include instructions for carrying out the subject assay, and may optionally include the fluid receptacle, e.g., the cuvette, multiwell plate, microfluidic device, etc. in which the reaction is to be carried out.

[00291] These kit reagents of the disclosure may be provided in vials for measuring by the user, or in pre-measured vials or ampoules which are simply combined to yield an appropriate reaction mixture. The reagents may be provided in liquid and/or lyophilized form and may optionally include appropriate buffer solutions for dilution and/or rehydration of the reagents. Generally, all of the reagents and instructions are co-packaged in a single box, pouch or the like that is ready for use.

VIII. TARGET GENES

[00292] The methods provided herein can be used to target all or a part of a disease related gene. For example nucleic acids containing a genes or gene segment related to one or more of the following diseases can be processed from a sample using the disclosed methods: ABCC8-Related Hyperinsulinism, Achromatopsia, Alkaptonuria, Alpha-1 Antitrypsin Deficiency, Alpha-Mannosidosis, Andermann Syndrome, ARSACS, Aspartylglycosaminuria, Ataxia With Vitamin E Deficiency, Ataxia-Telangiectasia, Autosomal Recessive Polycystic Kidney Disease, Bardet-Biedl Syndrome, BBS1-Related, Bardet-Biedl Syndrome, BBS10-Related, Biotinidase Deficiency, Bloom Syndrome, Canavan Disease, Carnitine Palmitoyltransferase IA Deficiency, Carnitine Palmitoyltransferase II Deficiency, Cartilage-Hair Hypoplasia, Choroideremia, Citrullinemia Type 1, CLN3-Related Neuronal Ceroid Lipofuscinosis, CLN5-Related Neuronal Ceroid Lipofuscinosis, Cohen Syndrome, Congenital Disorder of Glycosylation Type Ia, Congenital Disorder of Glycosylation Type Ib, Congenital Finnish Nephrosis, Costeff Optic Atrophy Syndrome, Cystic Fibrosis, Cystinosis, D-Bifunctional Protein Deficiency, Factor V Leiden Thrombophilia, Factor XI Deficiency, Familial Dysautonomia, Familial Mediterranean Fever, Fanconi Anemia Type C, Fragile X Syndrome, Galactosemia, Gaucher Disease, GJB2-Related DFNB 1 Nonsyndromic Hearing Loss and Deafness, Glucose-6-Phosphate Dehydrogenase Deficiency, Glutaric Acidemia Type 1, Glycogen Storage Disease Type Ia, Glycogen Storage Disease Type Ib, Glycogen Storage Disease Type III, Glycogen Storage Disease Type V, GRACILE Syndrome, Hb Beta Chain-Related Hemoglobinopathy (Including Beta Thalassemia and Sickle Cell Disease), Hereditary Fructose Intolerance, Hereditary Thymine-Uraciluria, Herlitz Junctional Epidermolysis Bullosa (LAMA3-Related), Herlitz Junctional Epidermolysis Bullosa (LAMB3-Related), Herlitz Junctional Epidermolysis Bullosa (LAMC2-Related), Hexosaminidase A Deficiency (Including Tay-Sachs Disease), HFE-Associated Hereditary Hemochromatosis, Homocystinuria Caused by Cystathionine Beta-Synthase Deficiency, Hurler Syndrome, Hypophosphatasia (Autosomal Recessive), Inclusion Body Myopathy 2, Isovaleric Acidemia, Joubert Syndrome 2, Krabbe Disease, Limb-Girdle Muscular Dystrophy Type 2D, Limb-Girdle Muscular Dystrophy Type 2E, Lipoamide Dehydrogenase Deficiency, Long Chain 3-Hydroxyacyl-CoA Dehydrogenase Deficiency, Maple Syrup Urine Disease Type 1B, Medium Chain

IDT Exhibit 1020.3196

Acyl-CoA Dehydrogenase Deficiency, Megalencephalic Leukoencephalopathy With Subcortical Cysts, Metachromatic Leukodystrophy, MTHFR Deficiency, Mucolipidosis IV, Muscle-Eye-Brain Disease, NEB-Related Nemaline Myopathy, Niemann-Pick Disease Type C, Niemann-Pick Disease (SMPD1-Associated), Nijmegen Breakage Syndrome, Northern Epilepsy, Pendred Syndrome, PEX1-Related Zellweger Syndrome Spectrum, Phenylalanine Hydroxylase Deficiency, Polyglandular Autoimmune Syndrome Type 1, Pompe Disease, PPT1-Related Neuronal Ceroid Lipofuscinosis, Primary Carnitine Deficiency, Primary Hyperoxaluria Type 1, Primary Hyperoxaluria Type 2, PROP1-Related Combined Pituitary Hormone Deficiency, Prothrombin Thrombophilia, Pseudocholinesterase Deficiency, Pycnodysostosis, Rhizomelic Chondrodysplasia Punctata Type 1, Salla Disease, Segawa Syndrome, Short Chain Acyl-CoA Dehydrogenase Deficiency, Sjogren-Larsson Syndrome, Smith-Lemli-Opitz Syndrome, Spinal Muscular Atrophy, Steroid-Resistant Nephrotic Syndrome, Sulfate Transporter-Related Osteochondrodysplasia, TPP1-Related Neuronal Ceroid Lipofuscinosis, Tyrosinemia Type I, Usher Syndrome Type 1F, Usher Syndrome Type 3, Very Long Chain Acyl-CoA Dehydrogenase Deficiency (Wilson Disease), and X-Linked Juvenile Retinoschisis.

[00293] The disease can be a cancer. Cancer can be, e.g., a tumor, a leukemia such as acute leukemia, acute t-cell leukemia, acute lymphocytic leukemia, acute myelocytic leukemia, myeloblastic leukemia, promyelocytic leukemia, myelomonocytic leukemia, monocytic leukemia, erythroleukemia, chronic leukemia, chronic myelocytic (granulocytic) leukemia, or chronic lymphocytic leukemia, polycythemia vera, lymphomas such as Hodgkin's lymphoma, follicular lymphoma or non-Hodgkin's lymphoma, multiple myeloma, Waldenström's macroglobulinemia, heavy chain disease, solid tumors, sarcomas, carcinomas such as, e.g., fibrosarcoma, myxosarcoma, liposarcoma, chondrosarcoma, osteogenic sarcoma, lymphangiosarcoma, mesothelioma, Ewing's tumor, leiomyosarcoma, rhabdomyosarcoma, colon carcinoma, colorectal cancer, pancreatic cancer, breast cancer, ovarian cancer, prostate cancer, including castration resistant prostate cancer, squamous cell carcinoma, basal cell carcinoma, adenocarcinoma, sweat gland carcinoma, sebaceous gland carcinoma, papillary carcinoma, papillary adenocarcinomas, cystadenocarcinoma, medullary carcinoma, bronchogenic, carcinoma, renal cell carcinoma, hepatoma, bile duct carcinoma, choriocarcinoma, seminoma, embryonal carcinoma, Wilms' tumor, cervical cancer, uterine cancer, testicular tumor, lung carcinoma, small cell lung carcinoma, bladder carcinoma, epithelial carcinoma, glioma, craniopharyngioma, ependymoma, pinealoma, hemangioblastoma, acoustic neuroma, oligodendroglioma, meningioma, melanoma, neuroblastoma, retinoblastoma, endometrial cancer, non small cell lung cancer, head and neck cancer, or kidney cancer.

[00294] The disease can be an autoimmune disease. The autoimmune disease can be a disease or disorder arising from and directed against an individual's own tissues or a co-segregate or manifestation thereof or resulting condition therefrom. Examples of autoimmune diseases or disorders include, but are not limited to: Acute Disseminated Encephalomyelitis (ADEM), Arthritis,

IDT Exhibit 1020.3197

Acute necrotizing hemorrhagic leukoencephalitis, Addison's disease, Agammaglobulinemia, Alopecia areata, Amyloidosis, Ankylosing spondylitis, Anti-GBM/Anti-TBM nephritis, Antiphospholipid syndrome (APS), Autoimmune angioedema, Autoimmune aplastic anemia, Autoimmune dysautonomia, Autoimmune hepatitis, Autoimmune hyperlipidemia, Autoimmune immunodeficiency, Autoimmune inner ear disease (AIED), Autoimmune myocarditis, Autoimmune pancreatitis, Autoimmune retinopathy, Autoimmune thrombocytopenic purpura (ATP), Autoimmune thyroid disease, Autoimmune urticaria, Axonal & neuronal neuropathies, Balo disease, Behcet's disease, Bullous pemphigoid, Cardiomyopathy, Castleman disease, Celiac disease, Chagas disease, Chronic fatigue syndrome, Chronic inflammatory demyelinating polyneuropathy (CIDP), Chronic recurrent multifocal ostomyelitis (CRMO),-Strauss syndrome, Cicatricial pemphigoid/benign mucosal pemphigoid, Crohn's disease, Cogans syndrome, Cold agglutinin disease, Congenital heart block, Coxsackie myocarditis, CREST disease, mixed cryoglobulinemia, neuropathies, Dermatitis herpetiformis, Dermatomyositis, Devic's disease (neuromyelitis optica) ,Discoid lupus, Dressler's syndrome ,Endometriosis, Eosinophilic esophagitis, Eosinophilic fasciitis, Erythema nodosum, Experimental allergic encephalomyelitis, Evans syndrome, Fibromyalgia, Fibrosing alveolitis, Giant cell arteritis (temporal arteritis), Glomerulonephritis, Goodpasture's syndrome, Granulomatosis with Polyangiitis (GPA), Graves' disease, Guillain-Barre syndrome, Hashimoto's encephalitis, Hashimoto's thyroiditis, Hemolytic anemia, Henoch-Schonlein purpura, Herpes gestationis, Hypogammaglobulinemia, Idiopathic thrombocytopenic purpura (ITP), IgA nephropathy, IgG4-related sclerosing disease, Immunoregulatory lipoproteins, Inclusion body myositis, Insulin-dependent diabetes (type1), Interstitial cystitis, Juvenile arthritis, Type I diabetes, Kawasaki syndrome, Lambert-Eaton syndrome, Leukocytoclastic vasculitis, Lichen planus, Lichen sclerosus, Ligneous conjunctivitis, Linear IgA disease (LAD), Lupus (SLE), Lyme disease, Meniere's disease, Microscopic polyangiitis, Mixed connective tissue disease (MCTD), Mooren's ulcer, Mucha-Habermann disease, Multiple sclerosis, Myasthenia gravis, Myositis, Narcolepsy, Neuromyelitis optica (Devic's), Neutropenia, Ocular cicatricial pemphigoid, Optic neuritis, Palindromic rheumatism, PANDAS (Pediatric Autoimmune Neuropsychiatric Disorders Associated with Streptococcus), Paraneoplastic cerebellar degeneration, Paroxysmal nocturnal hemoglobinuria (PNH), Parry Romberg syndrome, Parsonnage-Turner syndrome, Pars planitis (peripheral uveitis), Pemphigus, Peripheral neuropathy, Perivenous encephalomyelitis, Pernicious anemia, POEMS syndrome, Polyarteritis nodosa,Type I, II, or III autoimmune polyglandular syndromes, Polymyalgia rheumatic, Polymyositis, Postmyocardial infarction syndrome, Postpericardiotomy syndrome, Progesterone dermatitis, Primary biliary cirrhosis, Primary sclerosing cholangitis, Psoriasis, Psoriatic arthritis, Idiopathic pulmonary fibrosis, Pyoderma gangrenosum, Pure red cell aplasia, Raynauds phenomenon, Reflex sympathetic dystrophy, Reiter's syndrome, Relapsing polychondritis, Restless legs syndrome, Retroperitoneal fibrosis, Rheumatic fever, Rheumatoid arthritis, Sarcoidosis, Schmidt syndrome, Scleritis, Scleroderma, Sjogren's syndrome, Sperm & testicular autoimmunity, Stiff person syndrome,

IDT Exhibit 1020.3198

Subacute bacterial endocarditis (SBE), Susac's syndrome, Sympathetic ophthalmia, Takayasu's arteritis, Temporal arteritis/Giant cell arteritis, Thrombocytopenic purpura (TTP), Tolosa-Hunt syndrome, Transverse myelitis, Ulcerative colitis, Undifferentiated connective tissue disease (UCTD), Uveitis, Vasculitis, Vesiculobullous dermatosis, Vitiligo, and Wegener's granulomatosis.

[00295] The methods provided herein can be used to prepare samples comprising genes or parts of genes related to cancer. For example, the methods provided herein can prepare samples for sequencing somatic mutations related to cancer. Somatic mutations related to cancer can be found, for example, in the COSMIC database maintained by the Wellcome Trust Sanger Institute, which is incorporated herein by reference.

X. PROPHETIC EXAMPLES

EXAMPLE 1: Sample Protocol for mPEAR-generated libraries

[00296] Obtain input polynucleotide. Generally at least 50 ng of high molecular weight genomic DNA (gDNA). Fragment the gDNA.

[00297] Fragmentation:

[00298] Multiple fragmentation methods are suitable. Fragmentation by shearing can be used (e.g. Covaris). The mean fragment size can be approximately 100, 200, 300, 400 or more nucleotides but may vary depending on sequencer platform used. The fragment size can be greater for third generation sequencing technology.

[00299] Re-suspend DNA in 75µL 1 x Tris-EDTA (TE) buffer. Add resuspension mixture to glass Covaris tube. The following settings can be used: duty cycle 10%, intensity 5, cycles/burst 200, Time 120s.

End Repair and Adaptor ligation:

[00300] Blunt ends can be generated. A-tailing can be done for Illumina library generation. For other sequencing platforms, A-tailing may be optional.

[00301] End Repair

[00302] Add ligase buffer first to beads. Make master mix.

[00303] Prepare the following reaction mixture in a 0.5mL low DNA binding tube:

[00304]	H_20	75 µL
[00305]	T4 DNA ligase buffer with 10mM ATP	10 µL
[00306]	10mM dNTP mix	4 µL
[00307]	T4 DNA Polymerase	5 µL
[00308]	Klenow Fragment 5U/µL	1 µL
[00309]	T4 Polynucleotide kinase	5 µL

IDT Exhibit 1020.3199

[00310] TOTAL: 100 µL

[00311] Resuspend the beads in a end repair mix. Incubate the samples in a heat block at 20°C for 30 minutes. Place on magnetic particle consolidator (MPC) and remove supernatant. Using the MPC, wash the immobilized library 3 times with 200 µL of 1X SSC buffer, no tRNA. Mix well between each wash.

A-tailing mix:

[00312] Add buffer first to beads or make master mix. Prepare the following reaction mixture in a 0.5 mL low DNA binding tube:

[00313] H_20 32 µL

[00314] 10X buffer B011 5 µL

[00315] 1mM dATP 10 µL

[00316] Klenow Exo- 3 µL

[00317] Total: 50 µL

[00318] Resuspend the beads in the A tailing mix. Incubate the samples in a heat block at 37°C for 30 minutes. Place on MPC and remove supernatant. Using the MPC, wash the immobilized library 3 times with 200 µL 1X SSC with tRNA and 1 time with 1X SSC (1X = 0.150 M Sodium Chloride, 0.015 M Sodium Citrate) buffer, no tRNA. Mix well between each wash.

Ligation:

[00319] Add 2X ligation buffer to beads. Resuspend the beads such that the final adaptor concentration is 0.3 µM. Dilute stock 1:10 with annealing solution. A different barcode adaptors can be used for 10, 15, and 20 cycles from above.

[00320] dH20 18 µL

[00321] 2X Rapid Ligation Buffer (Enzymatics B101L) 25 µL

[00322] TruSeq barcoded Adaptors (1:10 dilution bc10,11,12) 3 µL

[00323] Total: 45 µL

[00324] Add 5 µL of T4 DNA Ligase (Rapid) and mix by pipetting up and down. Incubate the samples at 20°C for 15 minutes, with mixing. Using magnet, wash the beads 3 times with 200 µL 1X SSC (with 10ng/uL tRNA) and 1 times with 200 uL 1X SSC (no tRNA). Resuspend the beads in 23 µL H2O. Transfer to thin walled PCR tube, rinse tube with another 23 uL water.

PCR Enrichment:

[00325] This step may be optional.

[00326] Prepare the following PCR reaction mix in a 500 µL thin wall PCR tube

IDT Exhibit 1020.3200

[00327]	Beads resuspended in H2O	23 µL
[00328]	2x HiFi KAPA Master Mix	50 µL
[00329]	Truseq PCR primer 1	2 µL
[00330]	Truseq PCR primer 2	2 µL
[00331]	Nuclease water, from above	23 µL

[00332] Run the following PCR protocol 10 and 15 cycles:

[00333] 45 sec at 98°C

[00334] Remove 50 µL after 10 cycles then do 10 more cycles.

[00335] 98°C, 15 s

[00336] 60°C, 30 s

[00337] 72°C, 30 s

[00338] 1 min at 72°C

[00339] Hold at 4°C

[00340] Purify the DNA (e.g. with Zymo (25) PCR Purification Kit). Elute in 30 µL dH_20. Purify 10 cycle and 20 cycle product on 4% agarose gel.

EXAMPLE 2: Exemplary adapter design for mPEAR

[00341] Adaptor sequences can consist of a universal sequence at the 5' end, a 3 nucleotide sequence for directional information, and a 4 to 6 nucleotide molecular barcode at the 3' end. Incorporating molecular barcodes can allow for multiplexed target enrichment and sequencing. DNA samples can be individually fragmented, end repaired, and adaptors ligated. Because each sample can have a barcode, the mPEAR step may be multiplexed, dramatically reducing cost and increasing throughput by eliminating individual sample processing steps post library generation. mPEAR library generation with Genome RAPELLing is amenable to all sequencing platforms. Sequencer platform specific adaptors are incorporated through a low cycle PCR reaction after elution from streptavidin beads. Shorter adaptor sequences allow for greater ligation efficiency and more precise size selection of library molecules. Examples of suitable sequence are shown in **FIG.14**

EXAMPLE 3: mPEAR

[00342] Prepare libraries via 12 cycles PCR with barcoded samples. Transfer 30 µL DNA sample to 0.5 mL LoBind tubes. Add 1 nmol of each 3' blocked adaptor blocker oligonucleotides (10 µL each 100 µM, universal blocker 1, universal blocker 2, blocker 1, and blocker 2) and dry (e.g speedvac).

[00343] Resuspend:

[00344] 5 µL 10X std. Taq buffer	(60%) 5X biotin-dNTP
[00345] 10 uL 10X 60% biotin-dNTP	6.25 µL dA, dC, dG (20 mM)
[00346] 1 µL 60 primer mix (25 µM)	15 µL bio-dUTP (5 mM)

IDT Exhibit 1020.3201

[00347] 32 µL nuclease-free water 2.5 µL dT (20 mM)

[00348] 1 µL 100 mM soln. $MgCl_2$ 13.75 µL nH_2O

[00349] 1 µL Taq (5 U/µL) 1 µL

[00350] Total: 50 µL

[00351] Heat to 98°C, for 2 minutes. Either slow cool using ramp on thermocycler to 47°C or fast cool to 47°C, then hold at 47°C for 4 minutes, then 72°C fast ramp for 10 minutes, then add 1 µL of 0.5 M EDTA to quench. Place on ice.

Universal blocker 1:

GATCGGAAGAGCACACGTCTGAACTCCAGTCAC555555ATCTCGTATGCCGTCTTCTGCTTGX

Universal blocker 2:

CAAGCAGAAGACGGCATACGAGAT555555GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCX

Library Immobilization

[00352] Wash streptavidin beads (e.g. Dynal M280 Streptavidin beads) with: B&W-100uL, 1M Tris, pH 7.5, 20 µL 0.5 M EDTA, 4 mL 5 M NaCl fill to10 mL with nH_20. Transfer 25 µL beads to a new 0.5 mL LoBind tube. Wash streptavidin magnetic beads with B&W buffer (with 10ng/µL tRNA) 3 times, 200 µL each wash (pipette mixing). The last wash can be 5 minutes. Use a MPC, to pellet the beads and remove the buffer. Add 50 µL of B&W buffer (no tRNA). Transfer 50 µL of DNA to the 50 µL of Dynal M-280 streptavidin beads or vice versa can use an additional 20 µL B&W buffer to rinse tube. Mix well and incubate at room temperature, for 1 hour, mixing every 15 minutes. Using a magnet, wash the immobilized library 3 times with 200 µL of B&W buffer. Wash 2 times with 200 µL with 1X SSC. Transfer beads to PCR tube using last wash. Remove all remaining 1X SSC buffer of last wash. Beads should now be in PCR tube ready for amplification.

[00353] Enrich by PCR – 15 and 25 cycles

[00354] Add to beads: 98°C, 45 sec - hold

[00355] 50 µL 2X HiFi KAPA ready mix 98°C, 15 sec

[00356] 2 µL PCR 1 60°C, 30 sec

[00357] 2 µL PCR 2 72°C, 30 sec

[00358] 46 µL nH2O 72°C, 60 sec - hold

[00359] 100 µL 4°C - hold

IDT Exhibit 1020.3202

[00360] Clean DNA (e.g. using Zymo-25 column). Elute with 30 uL and speedvac dry. Resuspend in 20 μL. Load onto 2% agarose gel and run with ladder. Cut out region from 320-400 nucleotides. Excised DNA can be processed and sequenced.

Data Analysis:

[00361] After standard data quality metrics are determined, sequence reads can be binned or associated by molecular barcode.

Anchored Read Alignment:

[00362] On target sequencer reads will contain the mPEAR primer-annealing site. The annealing site is used as an "anchor" to position the read in the genome. The sequence downstream of the "anchor" can be treated as de novo sequence. This is a significant advantage over shot-gun exome sequencing methods that require the read to be similar to the reference genome being used. Longer insertions and deletions are detected through de-novo sequencing but missed by resequencing as they are too divergent from the reference genome to be detected by most alignment methods. Translocations, viral insertion sites and other mobile elements that flank known mPEAR annealing sites are also detected through anchored read alignment.

Detached-mate pair sequencing:

[00363] mPEAR allows for detached mate pair sequencing. Both strands are targeted with different mPEAR annealing sites. The annealing sites may be paired in silico. The overlap between the reads improves sequence quality and the distance between the mPEAR annealing sites may be used to determine copy number or chromosomal rearrangements. In some cases, both mPEAR annealing sites will be contained in a single read. This allows for the determination of long repetitive sequences such as the tandem tri-nucleotide repeat predictive of Fragile X syndrome. Longer target sequences may require multiple anchor alignment.

EXAMPLE 3: Genotyping Diagnostics

[00364] TELA reactions are prepared on a human DNA sample (e.g a population of human skin cells, suspected of being cancerous). Cells can be lysed using detergent and heat and approximately 15,000 copies of diploid DNA are precipitated via cholorform/ethanol extraction. A resuspension of DNA can be collected with approximately 10,000 copies of haploid DNA. A library of 100 TELA primer sets can be applied to the sample of DNA. Primer sets within the library can contain locus

IDT Exhibit 1020.3203

specific sequences for various oncogenes and tumor suppressors known to be associated with skin cancers. Each primer also can contains a barcode sequence. Primer extension reactions are conducted with similar reaction conditions as described herein with Klenow fragment polymerase. Subsequent PCR amplification, using universal priming sites (via the TELA primers) and degenerate primers can increase barcoded DNA yields to 10 ng.

[00365] The sample can be sequenced to sufficient coverage (e.g. 500) using a multiplex sequencing strategy. Bar coding of individual DNA strands can allow for sequencing information to be gained from individual strands rather than as an average of entire sample of DNA. Based upon the number of DNA strands sequenced and bar codes assigned, SNP phasing/haplotyping information is gained and many repetitive regions of DNA can be resolved. In addition, a substantial boost in accuracy can be gained by discarding mutations that appear randomly with respect to haplotypes, as those are likely to be sequencing errors. SNP phasing/haplotyping information provides genetics clues to type of skin cancer that may exist in the cells. A report can be generated for a medical practitioner for follow-up and evaluation. Further, these reports can be submitted and accessed electronically via the internet. Analysis of sequence data occurs at a site other than the location of the subject. The report is generated and transmitted to the user/medical practitioner's location. Via an internet enabled computer, the medical practitioner accesses the reports reflecting the analysis of the suspected cancer as shown in **FIG. 11**.

EXAMPLE 4: RAPELL Protocol

[00366] Add gDNA (e.g. human brain DNA), NEB4 buffer, nH_20 and denature 10 min at 96°C.

[00367] 1 µL human brain DNA (HBD) 50 ng

[00368] 2 µL RAPELL 1 (10 µM)

[00369] 2 µL RAPELL 2 (10 µM)

[00370] 1.2 µL 10 mM dNTP

[00371] 4 µL 10X NEB4 buffer

[00372] 17.8 µL nH_2O

[00373] 2 µL T4 DNA polymerase (no strand displacement)

[00374] 2 µL DNA ligase (600U/µL)

[00375] 3.6 µL 10 mM ATP

[00376] 35.6 µL reaction.

[00377] Incubate 20°C for 0.5 hour. Denature 96°C for 2 minutes. Add 2 µL T4 DNA polymerase. Add2 µL DNA ligase. Incubate at 20°C for 0.5 hours. Cycle 5 times between heat denature and adding enzymes, 2nd cycle add 2 µL 10X NEB4 buffer. Incubate at 75°C for 20 minutes to kill enzymes.

[00378] Use 53.6 µL from above. Add the following:

IDT Exhibit 1020.3204

[00379] 6 μL UDG

[00380] 6 μL APE

[00381] Total: 65.6 μL

[00382] Incubate 37°C for 1hour. Clean DNA (e.g. Zymo-5 clean-up, with 5 vol DNA bind buffer (320uL)). Elute with 20 μL elutant.

[00383] PCR between 15 and 25 cycles	Cycling
[00384] 20 μL DNA from above	98°C, 45 s
[00385] 2 μL PCR1 (25 μM)	98°C, 15 s
[00386] 2 μL PCR2 (25 μM)	60°C, 30 s
[00387] 50 μL 2X HiFi KAPA mix	72°C, 30 s
[00388] 26 μL nH_2O	72°C, 60 s
[00389] 100 μL	4°C - hold

[00390] Clean DNA (e.g. Zymo - 25 clean up). Elute with 30 μL elutant and dry (e.g. with speedvac). Run on 2% agarose gel. Excise DNA for further processing, sequencing and data analysis.

EXAMPLE 5: Exemplary oligonucleotides for RAPELLing

[00391] **FIG. 18** discloses exemplary oligonucleotides that can be used to perform the methods of the invention.

EXAMPLE 6:

[00392] Sample method of obtaining long fragment sequencing reads with mPEAR-generated libraries. Polynucleotide is obtained from a source. The polynucleotide is fragmented, end repaired, and subjected to mPEAR-mediated generation of an amplified target library. The target library is then subjected to the method of RAPELLing where the target library is fragmented into long polynucleotide fragments and subjected to partitioning, copying, amplifying, and sequencing. A consensus sequence is then assembled.

EXAMPLE 7:

[00393] 1. Primer design: A tiling strategy can be used for comprehensive coverage of targeted loci. Starting from 100-200 nucleotides upstream of the target loci and continuing through the loci to extend 100-200 nucleotides past the loci. The target can be broken up into 100-200 nucleotide windows wherein the best performing primer or primers are chosen. Primers can be variable in length. Primers can be optimized to account for an optimal Tm. Primers can be designed for maximum specificity to the target location. Primers can be designed to avoid dimerization. The 5'

IDT Exhibit 1020.3205

end of the primer can be "tailed" with a sequence corresponding to the sequencer platform used in the reaction. This sequence can be referred to as an "A adapter". The primer can be referred to as an "A adapter tailed primer". Primers can have a 3'OH group that can be extended by polymerase.

[00394] 2. Addition of an A adapter: an A adapter tailed primers can hybridize to genomic DNA ("gDNA"). A strand displacing thermostable polymerase can used to extend. Extension can occur in the 5'-3' direction. A copy of the template DNA can be made by the polymerase while displacing any DNA strands that are already hybridized to the template. The process can be repeated or cycled by heat denaturing, primer annealing and primer extension. Only the forward strand is copied in this manner.

[00395] 3. Fragmentation: Fragmentation of the amplified products can be accomplished by enzymatic or physical means to generate random three prime ends. The preferred method for generating random 3' ends is to incorporate a biotinylated ddNTP by using a low ratio of biotin ddNTP/dNTP in the primer extension reaction. The ddNTPs can randomly terminate the extended molecule. Other fragmentation methods can include shearing such as with sonication or enzymatic fragmentation. By using a biotinylated ddNTP, the biotinylated molecules can be isolated by streptavidin bead purification.

[00396] 4. Addition of B adapter: a B adapter can be added for sequencing reactions. If enzymatic or physical fragmentation is used, the B adapter can be added by ligation or an additional primer extension step. When terminating ddNTP is used for fragment generation in the previous step, the 3' end of the molecule is not available for primer extension or ligation due to the lack of a 3' OH. In this instance a B adapter can be added by hybridizing a random primer with the B adapter tail on or near the 5' end of the isolated molecule. In one non limiting example, the random primer can comprise: 5'-B adapter-NNNN...-3' wherein "NNNN" is representative of a random nucleic acid sequence of one or more nucleotides. The random segment of the primer can hybridize along the captured molecule. A strand displacing polymerase can be used to extend. The extension can occur in the 5' to 3'direction. The random primer at the end of the biotinylated template can extend and displace all other primers. After extending and displacing, it can be the only molecule remaining on the biotinylated template. The double stranded complex can consist of a strand containing a 5' A adapter - locus specific primer - target - terminating ddNTP - biotin 3'. The reverse strand can contain 5'- B adapter - random sequence - target - locus specific sequence - A Adapter - 3' . The complex can again be isolated by washing the streptavidin beads and removing the supernatant.

[00397] 5. Release: The complex with both an A adapter and a B adapter can be released from the beads by heat denaturation. 9-12 cycles of PCR is performed with primers complimentary to the A and B adapter sequences. If they have not already been incorporated, a full length sequencing adapters can be incorporated during the PCR step by using tailed primers.

IDT Exhibit 1020.3206

[00398] Optionally, molecular barcodes for sample multiplex during sequencing can be added with the original A adapter primer extension, the B adapter primer extension, and/or during the final PCR amplification.

[00399] 6. Sequencer ready library: the resulting sequencer ready library can consist of double stranded molecules in the following format: 5'- A adapter – synthetic primer – target – random end of target – b adapter – 3'. The library can be sequenced and the data can be stored and/or transmitted for analysis.

[00400] 7. Data analysis: during data analysis the adapters can be trimmed off of the reads. Samples can be identified based on any optional barcodes. Duplicate reads can be removed. Genomic coordinates can be identified by the known synthetic sequence at the beginning of the read. Reads corresponding to the same known genomic coordinate can be binned together. A consensus sequence can be generated without the use of a reference genome. Reads that do not form a consensus can be removed from analysis as off target. Each contiguous segment of target DNA can be considered a singular target and all primers corresponding to that target can be considered that targets primer set. This can produce an in silico read length that can be equivalent to the full length of the target. Generation of an in silico read length can be performed regardless of target size. After consensus sequence is determined, haplotypes for the target region can queried. For consensus sequences that do not match a known haplotype, the de novo sequence can be used to determine novel haplotypes and/or structural variation.

[00401] 8. Use: the technology may be used in clinical sequencing of known disease causing genes such as carrier testing. Other testing panels may be developed and potentially a targeted sequencing panel including interpretable locations in the genome. In addition, because the sequencing can go from known to unknown, the technology can be used to identify viral insertion sites.

EXAMPLE 8:

[00402] 1. Primer design: multiple primers can be designed to hybridize to specific target sequences within or nearby a selected target region. The spacing of each primer can be variable. The variable spacing can result in a high level of coverage along the entire length of the selected target region.

[00403] 2. Addition of an A adapter: the single direction primers also can contain a 5' sequence specific for the sequencing platform to be used (e.g. "A" adapter sequence on 5' end).

[00404] 3. Elongation: a primer mix is hybridized to target DNA and extended with a polymerase using a mix of dNTPs spiked with a low concentration of dUTP.

[00405] 4. Size-control/fragmentation: extended primers can be cut to generate a nested set of single stranded products. One non-limiting example of a way to cut is to use uracil DNA glycosolase ("UDG") and/or human apurinic/apyrimidinic endonuclease I (APE I) to generate nested set of single stranded products. These products can be anchored at the 5' end by the A adapter sequence. An

IDT Exhibit 1020.3207

alternative fragmentation strategy could use a mix of dNTPs spiked with a low concentration of methyl-dCTP followed by restriction enzyme digest with a four base cutter that will not cut sites with methyl-C incorporated.

[00406] 5. Addition of a B adapter: a B sequencing adapter is can be added to the 3' end by ligation or primer extension using a double stranded construct containing the B sequencing adapter with a 5'-random base overhang. After the B adapter is added, the library can be PCR amplified using A and B specific PCR primers. If the adapters that have been added are not full length, full length adapters can be added through primer tails.

[00407] Optionally, molecular barcodes for sample multiplex during sequencing can be added with the original A adapter primer extension, the B adapter primer extension, and/or during the final PCR amplification.

[00408] 6. Sequencer ready library: the resulting sequencer ready library can consist of double stranded molecules in the following format: 5'- A adapter – synthetic primer – target – random end of target – b adapter – 3'. The library can be sequenced and the data can be stored and/or transmitted for analysis.

[00409] 7. Data analysis: during data analysis the adapters can be trimmed off of the reads. Samples can be identified based on any optional barcodes. Duplicate reads can be removed. Genomic coordinates can be identified by the known synthetic sequence at the beginning of the read. Reads corresponding to the same known genomic coordinate can be binned together. A consensus sequence can be generated without the use of a reference genome. Reads that do not form a consensus can be removed from analysis as off target. Each contiguous segment of target DNA can be considered a singular target and all primers corresponding to that target can be considered that targets primer set. This can produce an in silico read length that can be equivalent to the full length of the target. Generation of an in silico read length can be performed regardless of target size. After consensus sequence is determined, haplotypes for the target region can queried. For consensus sequences that do not match a known haplotype, the de novo sequence can be used to determine novel haplotypes and/or structural variation.

[00410] 8. Use: the technology may be used in clinical sequencing setting. One non-limiting example is to test for alleles that are known to be associated with disease, e.g. carrier testing. Other testing panels may be developed and potentially a targeted sequencing panel including interpretable locations in the genome. In addition, because the sequencing can go from known to unknown, the technology can be used to identify viral insertion sites. Because this targeted method provides sequence alignment location while also allowing de novo assembly of reads, and because the target is sequenced by primer extension and not based on purification or recognition by hybridization, it can enrich for sequencing libraries that contain insertions, deletions, and/or other genetic anomalies that other hybridization based target capture methods have difficulty retaining. In one non-limiting

IDT Exhibit 1020.3208

example, this approach can be particularly suitable sequencing the Human Histocompatibility Antigen (HLA) region.

[00411] This method also describes a useful method for targeted, semi-targeted, and whole genome amplification procedures including whole genome phasing. In the case of whole genome phasing the strategy would be to start by a highly diluted, sub-genomic quantity of DNA in multiple reactions. The DNA could be in relatively large fragment (e.g. 10-40kb). In this iteration of the protocol, the majority of the sequence generated in each reaction is derived from a single copy of the subject's genome and thus, the majority of detected variations could be homozygous.

EXAMPLE 9:

[00412] 1. Library preparation: a library can be prepared using platform-specific library preparation method or kit. The method or kit can be commercially available and can generate a sequencer-ready library. Platform-specific library preparation methods can add a known sequence to the end of nucleic acid molecules; the known sequence can be referred to as an adapter sequence. Optionally, the library preparation method can incorporate one or more molecular barcodes.

[00413] 2. Targeting: DNA molecules from the sequencer-ready library can be selected (i.e. targeted) using a pool of one or more primers (i.e. mPEAR primers). A mPEAR primer can hybridize to a target library molecule or fragment. The hybridized mPEAR primers can be extended using a polymerase. mPEAR primer can comprise a universal or common 5' end, a spacer sequence, and a target or locus-specific sequence. The universal or common end can hybridize to the universal adapter sequences from the previous library generation step. This can serve to stabilize the synthetic oligonucleotides toward the end of DNA library fragments. Stabilization toward the end of the library fragments can allow the sequencer read to appropriately position the target sequence without wasted sequencer capacity. The spacer sequence can comprise a variable number of degenerate nucleotides. The degenerate nucleotides allow for length flexibility in the DNA library start and stop positions during DNA sequencing. Having variable sequence start sites can reduce systematic errors in the sequencing step. Having variable sequence start sites can allows for a randomized error profile across the reads of the redundant DNA library fragments. Finally, a locus-specific binding site localized near the three prime end of the mPEAR primer. The locus specific priming site is designed to recognize a DNA sequence that is upstream of the actual target sequence. One or more mPEAR primers can be used in tandem to target one or more regions of interest. Targeting sequences upstream of the target allows for increased specificity as pseudogenes and gene families with similar sequence homology can be avoided, reducing false positives in the data. Optionally, a second mPEAR primer can be designed to bind to the opposite strand. Optionally, the mPEAR primer can incorporate a molecular barcode.

IDT Exhibit 1020.3209

[00414] 2. Extension: the 3' end of the mPEAR primer can be available for primer extension. A polymerase can be used to extend the molecule. The extension can occur in a 5' to 3' direction. Biotinylated dNTPs can be incorporated (i.e. a mixture of native and biotinylated dNTPs can be used in the extension reaction). Optionally, the mPEAR primer extension can occur on both strands. The mPEAR primer extension can run through the end of the DNA library molecules. The optional use of a mPEAR primer extension occuring on both strands of DNA can be advantageous; having two reactions targeting the same sequence can increase specificity and/or can reduce failures (e.g. if one of the mPEAR primers were to fail, for example, by not hybridizing).

[00415] 3. Separation: the newly synthesized, biotinylated, DNA library/capture molecule hybrid can be incubated with streptavidin (e.g. streptavidin coated magnetic beads). The target, biotinylated DNA molecules can be isolated through magnetic bead purification. One or more washes can be performed with a suitable buffer. Optionally, DNA library molecules can be eluted from the capture molecules. Alternatively, an amplification reaction can be performed while the magnetic beads are still in solution (i.e. with no elution steo). Amplification can occur through PCR with the appropriate primers. During PCR amplification, full length sequencing platform specific adapter sequences can be incorporated. The resulting amplified molecules can be sequencer-ready or can be further purified through any means known in the art before sequencing).

[00416] 4. Sequencing: The library can be sequenced and the data can be stored and/or transmitted for analysis.

[00417] 5. Data analysis: during data analysis the adapters can be trimmed off of the reads. Samples can be identified based on any optional barcodes. Duplicate reads can be removed. Genomic coordinates can be identified by the known synthetic sequence at the beginning of the read. Reads corresponding to the same known genomic coordinate can be binned together. A consensus sequence can be generated without the use of a reference genome. Reads that do not form a consensus can be removed from analysis as off target. The read structure of the captured molecules can have significant advantages in speed and quality of data analysis. Because the locus specific primer recognition site is upstream of the target DNA sequence, that locus specific site is used to identify the genomic location of the read. A reference genome is not strictly needed; by "seeding" the read with the known sequence, the remainder can be assembled without the use of a reference genome. True de novo sequencing of the DNA targets can be accomplished. De novo sequencing can allow for greater sensitivity of genomic loci that greatly differ from the reference genome. Detection of longer insertions, deletions, repeats that are clinically relevant, and potentially viral insertion sites or mobile elements that disrupt gene function are detected with greater sensitivity. No reference score, or similarity to the reference is strictly needed to measure quality.

EXAMPLE 10:

IDT Exhibit 1020.3210

[00418] 1. Primer Design: primers can be designed upstream of the target sequence or loci. A first primer can comprise: a locus specific sequence and a 5' sequence wherein the 5' sequence can comprise all or part of a first adapter sequence of the sequencing platform being used (i.e. a TELA primer). A second primer (i.e. "a probe") can comprise: a random sequence consisting of 8 nucleotides and a 3' sequence, wherein the 3' sequence comprises all or part of a second adapter sequence of the sequencing platform being used (i.e. a TELA probe).

[00419] 2. Hybridization: the TELA primer can hybridize to a specific sequence. The specific sequence can be near the loci of interest. In some embodiments, the specific sequence can be just outside the loci of interest. The TELA probe can hybridize to random sequences across the entire genome. The spacing of the TELA probe hybridization can be adjusted by adjusting TELA probe concentration. The TELA primer and the TELA probe can hybridize to the same template strand.

[00420] 3. Elongation: a non-displacing polymerase can be used to extend the first primer until it reaches the second primer on the DNA template. A DNA ligase can join (i.e. connect or ligate) the TELA primer to the TELA probe. The resulting product can consists of a single stranded copy of the DNA template flanked by adapter tails.

[00421] 4. Amplification: PCR can be used to amplify the product. In cases where the product does not contain complete (i.e. full length) adapter sequences, tailed primer amplification can incorporate the remainder of the first and second sequencer adapters (i.e. sometimes referred to as an A adapter and a B adapter). The amplification can produce a sequencer-ready library. Optionally, the product can be purified or further processed to be sequencer-ready.

[00422] Optionally, one or more molecular barcodes can be added with the TELA primer, the TELA probe and/or during the final PCR tailed amplification.

[00423] 5. Sequencing: the library can be sequenced and the data can be stored and/or transmitted for analysis.

[00424] 6. Data analysis: the data from sequencing can be analyzed. An advantage of this approach includes the ability to determine the genomic position of the sequencer read from the synthetic sequence at the beginning of each read. In addition, clonal errors can be avoided or reduced because the 3' end of the sequencer read is randomly generated (i.e. through the randomly binding TELA probe). If a variant is detected from multiple sequencer reads with different 3' ends, it is likely a true, or genomic, variant as opposed to a sequencer or read variant. If an error occurs during amplification, it may only appear in reads with the same 3' end.

EXAMPLE 11:

[00425] Sequencing libraries were prepped according to standard library preparation methods. Two samples were barcoded with Truseq barcodes #5 and #6.

IDT Exhibit 1020.3211

[00426] Sample A was prepared by: fragmentation (i.e. shearing) using the Covaris shearing method. The fragmented sample was: end-repaired, A-tailed, and adapter ligated. Then magnetic bead purification 2 times – index 5, took all of prepped sample into protocol.

[00427] Sample B: Fragmented on covaris, end-repaired, A-tailed, and adapter ligated then magnetic bead purification 2 times, then PCR cycled this material 12 times, then magnetic bead purified before being purified further on 2% agarose gel, bands at approx. 350-420bp were cut, excised, and isolated by agarose dissolving buffer and zymo-25 column– index 6, approx.. 200 ng input

[00428] Transfer all of DNA sample to 0.5 mL LoBind tube (Eppendorf). Add 1 nano mole of each 3' blocked adapter blocker oligos. (10 uL each 100 uM, universal blocker 1, universal blocker 2, blocker 1, and blocker 2). Speedvac dry and resuspend.

[00429] 5 µL 10X standard Taq buffer	(60%) 5X biotin-dNTP
[00430] 10 µL 10X 60% biotin-dNTP	6.25 µL dA, dC, dG (20 mM)
[00431] 1 µL 60 primer mix (25 µM)	15 µl bio-dUTP (5 mM)
[00432] 32 µL nuclease-free water	2.5 µL dT (20 mM)
[00433] 1 µL 100 mM solution $MgCl_2$	13.75 uL nH_2O
[00434] 1 µL Taq (5 U/µL)	
[00435] TOTAL:	50 uL

[00436] Heat to 98°C, for 2 minutes. Slow cool using ramp on thermocycler to 47°C then to 72°C for 10 minutes, then add 1 µL of 0.5 M EDTA to quench and place on ice.

[00437] Library Immobilization

[00438] Wash Dynal M280 Streptavidin beads (B&W-100µL, 1M Tris pH7.5, 20 µL 0.5 M EDTA, 4 mL 5 M NaC up to 10 mL with nH_2O). Transfer 25µL beads to a new 0.5 mL LoBind tube. Wash streptavidin magnetic beads with B&W buffer (with 10ng/uL tRNA), 3 times, 200 µL each wash (pipette mixing). Last wash wait 5 minutes before removing. (Use a MPC, to pellet the beads and remove the buffer). Add 50 µL of B&W buffer (no tRNA). Transfer 50 µL of DNA to the 50 µL of Dynal M-280 streptavidin beads or vice versa (can use an additional 20 µL B&W buffer to rinse tube). Mix well and incubate at room temperature, for 1 hour, mixing every 15 minutes. Using magnet, wash the immobilized library 3 times with 200 µL of B&W buffer and 2 times with 200 µL wash with 1X saline sodium citrate (SSC). Transfer beads to PCR tube using last wash. Remove all remaining 1X SSC buffer of last wash. Beads should now be in PCR tube ready for amplification.

[00439] **Enrich by PCR**: between 15 and 25 cycles (take out 50 µL at 15 cycles and let rest finish to 25 cycles).

	Cycles
[00440] Add to beads:	98°C, 45 sec - hold
[00441] 50 µL 2X HiFi KAPA ready mix (KAPA Biosystems	98°C, 15 sec
[00442] 2 µL PCR 1	60°C, 30 sec
[00443] 2 µL PCR 2	72°C, 30 sec
[00444] 46 µL nH_2O	72°C, 60 sec - hold

IDT Exhibit 1020.3212

[00445] 100 µL 4°C - hold

[00446] Clean DNA (e.g. Zymo-25 column), elute with 30 µL elutants, speedvac dry, resuspend 20 µL load onto 2% agarose gel. Cut 320-400 nucleotides band. The material cycled 25 times produced enough DNA product for sequencing. Amplified DNA was purified on 2% agarose and bands at approximately 350-420 nucleotides were cut, excised, and isolated by agarose dissolving buffer and zymo-25 column.

EXAMPLE 12:

[00447] Libraries were prepped: (12 cycles PCR). Two samples barcode 5,735 ng and barcode 6, 684 ng. Transfer 30 µL DNA sample to 0.5 mL LoBind tube. Add 1 nmol of each 3' blocked adapter blocker oligonucleotides. (10 µL each 100 µM, universal blocker 1, universal blocker 2, blocker 1, and blocker 2). Speedvac dry.

[00448] Resuspend:

[00449] 5 µL 10X std. Taq buffer (60%) 5X biotin-dNTP

[00450] 10 µL 10X 60% biotin-dNTP 6.25 µL dA, dC, dG (20 mM)

[00451] 1 µL 60 primer mix (25 uM) 15 µl bio-dUTP (5 mM)

[00452] 32 µL nuclease-free water 2.5 µL dT (20 mM)

[00453] 1 µL 100 mM soln. MgCl2 13.75 µL nH_2O

[00454] 1 µL Taq (5 U/uL)

[00455] Total: 50 µL

[00456] Heat to 98°C, for 2 minutes. Either slow cool using ramp on thermocycler to 47°C or fast cool to 47°C, then hold at 47°C for 4 minutes, then 72°C fast ramp for 10 minutes, then add 1 µL of 0.5 M EDTA to quench. And place on ice.

[00457] Library Immobilization

[00458] Wash Dynal M280 Streptavidin beads:

[00459] (B&W-100uL, 1M Tris pH7.5, 20 uL 0.5 M EDTA, 4 mL 5 M NaCl…up to 10 mL with n-water). Transfer 25µL beads to a new 0.5 mL LoBind micro centrifuge tube. Wash streptavidin magnetic beads with B&W buffer (with 10ng/uL tRNA), 3 times, 200 µL each wash (pipette mixing). Last wash wait 5 minutes before removing. Use a MPC, to pellet the beads and remove the buffer. Add 50 µL of B&W buffer (no tRNA). Transfer 50 µL of DNA to the 50 µL of Dynal M-280 streptavidin beads or vice versa. An additional 20 µL B&W buffer can be used to rinse tube. Mix well and incubate at room temperature, for 1 hour, mixing every 15 minutes. Using magnet, wash the immobilized library 3 times with 200 µL of B&W buffer and 2 times with 200 µL wash with 1X SSC. Transfer beads to PCR tube using last wash. Remove all remaining 1X SSC buffer of last wash. Beads should now be in PCR tube ready for amplification.

[00460] Enrich by PCR – 15 and 25 cycles

IDT Exhibit 1020.3213

[00461] Add to beads: 98°C, 45 sec - hold

[00462] 50 µL 2X HiFi KAPA ready mix 98°C, 15 sec

[00463] 2 µL PCR 1 60°C, 30 sec

[00464] 2 µL PCR 2 72°C, 30 sec

[00465] 46 µL nH_2O 72°C, 60 sec - hold

[00466] 100 µL 4C - hold

[00467] Clean DNA (e.g. Zymo-25 column), elute 30 µL, speedvac dry, resuspend 20 µL load onto 2% agarose gel. Cut 320-400 nucleotide area and gel extract DNA. DNA can be further processed and sequenced.

[00468] Universal blocker1:
GATCGGAAGAGCACACGTCTGAACTCCAGTCAC555555ATCTCGTATGCCGTCTTCTGCTTGX

[00469] Universal blocker 2:
CAAGCAGAAGACGGCATACGAGAT555555GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCX

[00470] The method can be used with approximately 10 kb region of the human CCLS gene. The method can also be performed with highly multiplexed short PCR reactions.

[00471] EXAMPLE 13

[00472] A PCR reaction can be performed 2 times. 1 µL ~1.2 µg gDNA (e.g. human brain DNA). Two uL of a 60 primer mix (25 uM stock) can be added with:

[00473] 25 µL 2X KAPA HiFi

[00474] 22 µL nH_2O

[00475] Total: 50 µL

[00476] PCR – 10 cycles

[00477] 98°C, 45 s

[00478] 98°C, 15 s

[00479] 47°C, 30 s

[00480] 72°C, 30 s

[00481] 72°C, 1 min

[00482] 4°C - hold

[00483] Clean DNA (e.g. Zymo-25 clean-up), elute with 30 µL, and dry (e.g. using a speed vacuum). Load DNA on a 2% agarose, two wells, gel purify fragments between 300 – 600 nucleotides. Elute from gel, isolate using, e.g. Zymo-25, elute with 30 µL water, speedvac to dry.

[00484] Fragmentation: resuspend all DNA in 50µL 1X TE (with 2 ng/uL tRNA). Add to glass Covaris tube. Use settings: Duty Cycle 10%, Intensity 5, Cycles/Burst 200, time 120 s to shear DNA.

IDT Exhibit 1020.3214

[00485] 50 uL Fragmented DNA, 20 uL End Repair mix (8 uL water, 7 uL 10X KAPA end-repair, 5 uL KAPA enzymes). Incubate 20°C for 30 minutes. Clean-up, add 120 µL AmpureXP beads to 70 uL = 190uL. Mix well, incubate 10min let DNA bind. Move tube to magnet, remove liquid, 3 minutes. Wash beads with 200 µL 80% ethanol, wait 30 seconds, remove, repeat, 2 washes total. Allow beads dry 10 min.

[00486] Add beads from above. Allow beads to rehydrate for 3 minutes. 50 µL A-tailing master mix (42 uL water, 5 uL 10X KAPA A-tail, 3 uL KAPA A-tail enzyme). Mix well and incubate 30 °C for 30 minutes. Clean by adding 90 µL 20% PEG8000/2.5M NaCl solution. The total volume can be 140 µL, mix thoroughly via pipetting. Incubate 10 min let DNA bind. Move tube to magnet, remove liquid. Wash beads with 200 uL 80% ethanol, wait 30 seconds, remove, repeat, 2 washes total. Allow beads dry for 5 minutes. Ligate adapters. Add beads from above, allow beads rehydrate for 3 min.

[00487] 45 µL Ligation master mix (30 µL water, 10 µL 5X KAPA Lig., 5 µL T4 DNA ligase).

[00488] 3 µL adapters (1:10 dilution of standard) (final adapter conc. in reaction = 0.3 uM)

[00489] Mix well and incubate 20 °C for 15 minutes. Wash 2 times. Add 50 µL 20%PEG8000/2.5M NaCl solution. Total volume 100 µL, mix thoroughly via pipetting. Incubate 10 minutes let DNA bind. Move tube to magnet, remove liquid. Wash beads with 200 µL 80% ethanol, wait 30 seconds, remove, repeat, 2 washes. Allow beads dry for 5 minutes. Resuspend beads in 50 µL water, wait 5 minutes. Add 50 µL 20%PEG8000/2.5M NaCl solution. Incubate 10 min let DNA bind. Move tube to magnet, remove liquid. Wash beads with 200 µL 80% ethanol, wait 30 seconds, remove, repeat, 2 washes total. Allow beads dry for 5 minutes. Resuspend beads 23 µL Tris pH 8, allow rehydrate 3 minutes. Collect DNA from elute and transfer to PCR tube. Enrich by PCR. Prepare the following PCR reaction mix in a 500 µL thin wall PCR tube.

[00490]	Use DNA from above	23 µL
[00491]	2x HiFi KAPA Master Mix	50 µL
[00492]	Truseq PCR primer 1	2 µL
[00493]	Truseq PCR primer 2	2 µL
[00494]	Nuclease water	23 µL

[00495] Run the following PCR protocol 10 cycles:

[00496] 45 sec at 98°C

[00497] 10 cycles of:

[00498] 98 °C, 15 sec

[00499] 60 °C, 30 sec

[00500] 72 °C, 30 sec

[00501] 1 min at 72 °C

[00502] Hold at 4 °C

[00503] Purify the DNA (e.g. with Zymo (25) PCR Purification Kit) and elute in 30 µL dH_2O. Purify 10 cycle and 20 cycle material on, 2 wells per PCR, 4% agarose gel.

IDT Exhibit 1020.3215

[00504] While preferred embodiments of the present invention have been shown and described herein, it will be obvious to those skilled in the art that such embodiments are provided by way of example only. Numerous variations, changes, and substitutions will now occur to those skilled in the art without departing from the invention. It should be understood that various alternatives to the embodiments of the invention described herein may be employed in practicing the invention. It is intended that the following claims define the scope of the invention and that methods and structures within the scope of these claims and their equivalents be covered thereby.

[00505] While preferred embodiments of the present invention have been shown and described herein, it will be obvious to those skilled in the art that such embodiments are provided by way of example only. Numerous variations, changes, and substitutions will now occur to those skilled in the art without departing from the invention. It should be understood that various alternatives to the embodiments of the invention described herein may be employed in practicing the invention. It is intended that the following claims define the scope of the invention and that methods and structures within the scope of these claims and their equivalents be covered thereby.

IDT Exhibit 1020.3216

WHAT IS CLAIMED IS:

1. A method comprising:
    a. Spatially separating nucleic acid fragments of a nucleic acid;
    b. generating one or more amplicons, wherein the amplicons are generated by:
        i. hybridizing a primer and a probe to a common strand of the nucleic acid fragment;
        ii. performing a primer extension reaction;
        iii. ligating a product of the primer extension reaction with the probe to form the amplicons;
    c. associating the one or more amplicons with an identifier;
    d. obtaining a sequence of the amplicon.
2. The method of claim 1, further comprising amplifying the amplicons.
3. The method of claim 2, wherein the amplifying is linear.
4. The method of claim 3, wherein the amplifying is performed by rolling circle amplification.
5. The method of claim 2, wherein at least one amplicon is ligated to at least one different amplicon.
6. The method of claim 2, wherein the amplifying is non-linear.
7. The method of claim 1, wherein the identifier comprises a molecular barcode.
8. The method of claim 1, wherein the identifier comprises a nucleic acid sequence.
9. The method of claim 1, wherein the identifier comprises a nucleic acid not A, T, C, or G.
10. The method of claim 1, wherein the identifier is located at the 5' end of the amplicon.
11. The method of claim 1, wherein the identifier is located at the 3' end of the amplifcon.
12. The method of claim 1, wherein the identifier is associated with the amplicon through amplification of the amplicon.
13. The method of claim 1, wherein the nucleic acid is selected from a group consisting of: DNA, RNA, cDNA and genomic DNA.
14. The method of claim 1, wherein fragmenting of the nucleic acid is performed by method from a group consisting of: sonication, enzymatic digestion, heat, exposure to UV light, repetitive pipetting, and nebulization.
15. The method of claim 1, wherein spatially separating the nucleic acid fragments is performed in partitions.
16. The method of claim 1, wherein spatially separating the nucleic acid fragments is performed by tethering the nucleic acid to a solid or semi-solid support.
17. The method of claim 16, wherein the nucleic acid fragment is hybridized to a primer tethered to a solid or semi-solid support, wherein the primer comprises the identifier.

IDT Exhibit 1020.3217

18. The method of claim 16, wherein the nucleic acid fragment is hybridized to a probe tethered to a solid or semi-solid support, wherein the probe comprises the identifier.
19. The method of claim 17, wherein the solid or semi-solid support is addressed.
20. The method of claim 1, wherein the probe comprises the identifier.
21. The method of claim 1, wherein the probe comprises one or more adapter sequences.
22. The method of claim 1, wherein the probe hybridizes to a target nucleic acid.
23. The method of claim 1, wherein the probe comprises degenerate sequence.
24. The method of claim 1, wherein the probe comprise a synthetic nucleotide.
25. The method of claim 1, wherein the probe comprises a primer.
26. The method of claim 1, wherein the primer comprises the identifier.
27. The method of claim 1, wherein the primer comprises one or more adapter sequences.
28. The method of claim 1, wherein the primer hybridizes to a target nucleic acid.
29. The method of claim 1, wherein the primer comprises degenerate sequence.
30. The method of claim 1, wherein the primer comprise a synthetic nucleotide.
31. The method of claim 1, wherein the amplicons are associated with a unique identifier.
32. The method of claim 15, wherein the identifier represents an individual partition.
33. The method of claim 1, wherein obtaining the sequence of the amplicons is performed by massively parallel sequencing.
34. The method of claim 1, wherein a computing device is used to generate a consensus sequence of all or part of the nucleic acid fragment, from sequence reads comprising the identifier.
35. The method of claim 1, wherein a computing device is used to generate a consensus sequence of all or part of the nucleic acid, from sequence reads comprising the identifier.
36. The method of claims 34 and 35, wherein a computing device is used to generate a consensus sequence without comparing the consensus sequence to a reference.
37. The method of claims 34 and 35, wherein the consensus sequence has at least 1X depth of coverage.
38. The method of claims 34 and 35, wherein the consensus sequence has at least 5X depth of coverage.
39. The method of claims 34 and 35, wherein the consensus sequence has at least 10X depth of coverage.
40. The method of claims 34 and 35, wherein the consensus sequence has at least 50X depth of coverage.
41. The method of claim 1, wherein the method is multiplexed for multiple nucleic acid fragments.
42. The method of claim 35, wherein the method is multiplexed for at least 2 nucleic acid fragments.

IDT Exhibit 1020.3218

43. The method of claim 35, wherein the method is multiplexed for at least 10 nucleic acid fragments.
44. The method of claim 35, wherein the method is multiplexed for at least 100 nucleic acid fragments.
45. The method of claim 35, wherein the method is multiplexed for at least 10000 nucleic acid fragments.
46. The method of claim 35, wherein the method is multiplexed for at least 100000 nucleic acid fragments.
47. The method of claim 35, wherein the method is multiplexed for at least 1000000 nucleic acid fragments.
48. The method of claim 1, wherein the product of the primer extension reaction is at least 100 nucleotides.
49. The method of claim 1, wherein the product of the primer extension reaction is at least 1000 nucleotides.
50. The method of claim 1, wherein the product of the primer extension reaction is at least 10000 nucleotides.
51. A method comprising transmitting sequencing data generated by the method of claim 1.
52. A method comprising receiving sequencing data generated by the method of claim 1.
53. A method comprising storing sequencing data generated by the method of claim 1.
54. A method comprising comparing or analyzing sequencing data generated by the method of claim 1.
55. A method comprising transmitting a report related to sequencing data generated by the method of claim 1.
56. A method comprising receiving a report related to sequencing data generated by the method of claim 1.
57. A method comprising storing a report related to sequencing data generated by the method of claim 1.
58. A method comprising comparing or analyzing a report related to sequencing data generated by the method of claim 1.
59. A method of claim 1 further comprising transforming sequencing data to a report related to sequencing data using a computing device comprising non transitory computer readable media.
60. The method of claim 1, wherein the primer or probe is specific to one or more regions of the nucleic acid fragment.
61. The method of claim 1, wherein the primer or probe is at least 50% complementary to one or more regions of the nucleic acid fragment.

IDT Exhibit 1020.3219

62. The method of claim 1, wherein the primer or probe is at least 75% complementary to one or more regions of the nucleic acid fragment.
63. The method of claim 1, wherein the primer or probe is at least 90% complementary to one or more regions of the nucleic acid fragment.
64. The method of claim 1, wherein one or more amplicons are linked to form a contiguous sequence.
65. The method of claim 1 wherein the performing a primer extension reaction comprises addition of a strand displacing polymerase.
66. The method of claim 1 further comprising performing a primer extension reaction to form a primer extension product, wherein the primer extension product comprises an affinity conjugate and wherein the primer extension product comprises a target sequence.
67. The method of claim 1 further comprising performing a primer extension reaction to form a primer extension product and performing affinity purification of the primer extension product using the affinity conjugate.
68. The method of claim 66 or 67 wherein the affinity conjugate is biotin.
69. The method of claim 68 wherein the affinity purification of the primer extension product is performed using streptavidin.

70. A method comprising:
    a. obtaining a nucleic acid, wherein the nucleic acid comprises a target sequence;
    b. hybridizing a TELA primer and a TELA probe to a common strand of the nucleic acid;
    c. performing a primer extension reaction;
    d. ligating a product of the primer extension reaction with the TELA probe to form a ligation product comprising the target sequence; and
    e. sequencing the target sequence.

71. The method of claim 70, wherein the target sequence is at least 30% of the ligation product.

72. A method comprising transmitting sequencing data generated by the method of claim70.

73. A method comprising receiving sequencing data generated by the method of claim 70.

74. A method comprising storing sequencing data generated by the method of claim 70.

75. A method comprising comparing or analyzing sequencing data generated by the method of claim 70.

76. A method comprising transmitting a report related to sequencing data generated by the method of claim 70.

77. A method comprising receiving a report related to sequencing data generated by the method of claim 70.

78. A method comprising storing a report related to sequencing data generated by the method of claim 70.

IDT Exhibit 1020.3220

79. A method comprising comparing or analyzing a report related to sequencing data generated by the method of claim 70.

80. A method of claim 70 further comprising transforming sequencing data to a report related to sequencing data using a computing device comprising non transitory computer readable media.

81. The method of claim 70, wherein one or more ligation products are linked to form a contiguous sequence.

82. A method comprising:
    a. obtaining a nucleic acid library;
    b. ligating an adapter sequence to one or more nucleic acids of the nucleic acid library;
    c. hybridizing a primer to the adapter sequence, wherein the primer comprises a spacer region and a locus specific region;
    d. performing a primer extension reaction to form a primer extension product, wherein the primer extension product comprises an affinity conjugate and wherein the primer extension product comprises a target sequence;
    e. performing affinity purification of the primer extension product using the affinity conjugate.

83. The method of claim 82, wherein the nucleic acid library is fragmented genomic DNA.

84. The method of claim 82, wherein the nucleic acid library is expressed sequences.

85. The method of claim 82, wherein the nucleic acid library is epigenetically sorted.

86. The method of claim 82, wherein the primer extension product is sequenced.

87. The method of claim 82, wherein the nucleic acid library comprises at least 2 nucleic acid fragments.

88. The method of claim 82, wherein the nucleic acid library comprises at least 10 nucleic acid fragments.

89. The method of claim 82, wherein the nucleic acid library comprises at least 100 nucleic acid fragments.

90. The method of claim 82, wherein the nucleic acid library comprises at least 10,000 nucleic acid fragments.

91. The method of claim 82, wherein the nucleic acid library comprises at least 100,000 nucleic acid fragments.

92. The method of claim 82, wherein the nucleic acid library comprises at least 1,000,000 nucleic acid fragments.

93. The method of claim 82, wherein the nucleic acid library has been fragmented.

94. The method of claim 93 wherein the fragmenting was performed by method from a group consisting of: sonication, enzymatic digestion, heat, exposure to UV light, repetitive pipetting, and nebulization.

95. The method of claim 82 further comprising amplifying a nucleic acid to generate the nucleic acid library.

IDT Exhibit 1020.3221

96. The method of claim 82 further comprising amplifying the adaptor ligated nucleic acid library.

97. The method of claim 95 or 96, wherein the amplifying is linear.

98. The method of claim 97, wherein the amplifying is performed by rolling circle amplification.

99. The method of claim 96, wherein the amplifying is non-linear.

100. The method of the claim 82, wherein the primer extension product comprises an identifier.

101. The method of the claim 82, wherein the primer extension product comprises a molecular barcode.

102. The method of claim 82, wherein the primer extension product comprises a nucleic acid sequence.

103. The method of claim 82, wherein the primer extension product comprises a nucleic acid not A, T, C or G.

104. The method of the claim 82, wherein the adaptor is located at the 5' end of the nucleic acid of the nucleic acid library.

105. The method of claim 82, wherein the primer extension product is located at the 3' end of the nucleic acid of the nucleic acid library.

106. The method of claim 82, wherein the nucleic acid is selected from a group consisting of: DNA, RNA, cDNA and genomic DNA.

107. The method of claim 82, wherein the spacer region is degenerate or random sequence.

108. The method of claim 82, wherein the spacer region comprises at least 1 nucleotide.

109. The method of claim 82, wherein the spacer region comprises at least 10 nucleotides.

110. The method of claim 82, wherein the spacer region comprises at least 100 nucleotides.

111. The method of claim 82, wherein the spacer region comprises a molecular barcode.

112. The method of claim 82, wherein the spacer region comprises a nucleic acid sequence.

113. The method of claim 82, wherein the spacer region comprises a nucleic acid not A, T, C or G.

114. The method of claim 82, wherein the spacer region comprises an enzymatic target sequence.

115. The method of claim 82, wherein the locus specific region is at least 50% complementary to a locus of the nucleic acid.

116. The method of claim 82, wherein the locus specific region is at least 70% complementary to a locus of the nucleic acid.

117. The method of claim 82, wherein the locus specific region is at least 80% complementary to a locus of the nucleic acid.

IDT Exhibit 1020.3222

118. The method of claim 82, wherein the locus specific region is at least 90% complementary to a locus of the nucleic acid.

119. The method of claim 82, wherein the locus specific region is at least 99% complementary to a locus of the nucleic acid.

120. The method of claim 82, wherein the locus specific region binds a locus of the nucleic acid upstream of the target sequence.

121. The method of claim 82, wherein the affinity conjugate is biotin.

122. The method of claim 82, wherein the affinity purification of the primer extension product is performed using streptavidin.

123. The method of claim 82, further comprising sequencing the primer extension product.

124. The method of claim 100, wherein the identifier represents a source of the sample from which the nucleic acid library was generated.

125. The method of claim 123, wherein the sequencing of the primer extension product is obtained by performing massively parallel sequencing.

126. The method of claim 100, wherein a computing device is used to generate a consensus sequence of all or part of the target sequence, from sequence reads comprising the identifier.

127. The method of claim 100, wherein a computing device is used to generate a consensus sequence of all or part of the target sequence, from sequence reads comprising the identifier.

128. The method of claim 100, wherein a computing device is used to generate a consensus sequence without comparing the consensus sequence to a reference.

129. The method of claim 126-128, wherein the consensus sequence has at least 1X depth of coverage.

130. The method of claim 126-128, wherein the consensus sequence has at least 5X depth of coverage.

131. The method of claim 126-128, wherein the consensus sequence has at least 10X depth of coverage.

132. The method of claim 126-128 wherein the consensus sequence has at least 50X depth of coverage.

133. The method of claim 82, wherein the method is multiplexed for multiple samples or target sequences.

134. The method of claim 82, wherein the method is multiplexed for at least 2 samples or target sequences.

135. The method of claim 82, wherein the method is multiplexed for at least 10 samples or target sequences.

IDT Exhibit 1020.3223

136. The method of claim 82, wherein the method is multiplexed for at least 100 samples or target sequences.

137. The method of claim 82, wherein the method is multiplexed for at least 10000 samples or target sequences.

138. The method of claim 82, wherein the method is multiplexed for at least 100000 samples or target sequences.

139. The method of claim 82, wherein the method is multiplexed for at least 1000000 samples or target sequences.

140. The method of claim 82, wherein the product of the primer extension reaction is at least 100 nucleotides.

141. The method of claim 82, wherein the product of the primer extension reaction is at least 1000 nucleotides.

142. The method of claim 82, wherein the product of the primer extension reaction is at least 10000 nucleotides.

143. A method comprising transmitting sequencing data generated by the method of claim 82.

144. A method comprising receiving sequencing data generated by the method of claim 82.

145. A method comprising storing sequencing data generated by the method of claim 82.

146. A method comprising comparing or analyzing sequencing data generated by the method of claim 82.

147. A method comprising transmitting a report related to sequencing data generated by the method of claim 82.

148. A method comprising receiving a report related to sequencing data generated by the method of claim 82.

149. A method comprising storing a report related to sequencing data generated by the method of claim 82.

150. A method comprising comparing or analyzing a report related to sequencing data generated by the method of claim 82.

151. A method of claim 82 further comprising transforming sequencing data to a report related to sequencing data using a computing device comprising non transitory computer readable media.

152. The method of claim 82, wherein one or more primer extension products are linked to form a contiguous sequence.

IDT Exhibit 1020.3224

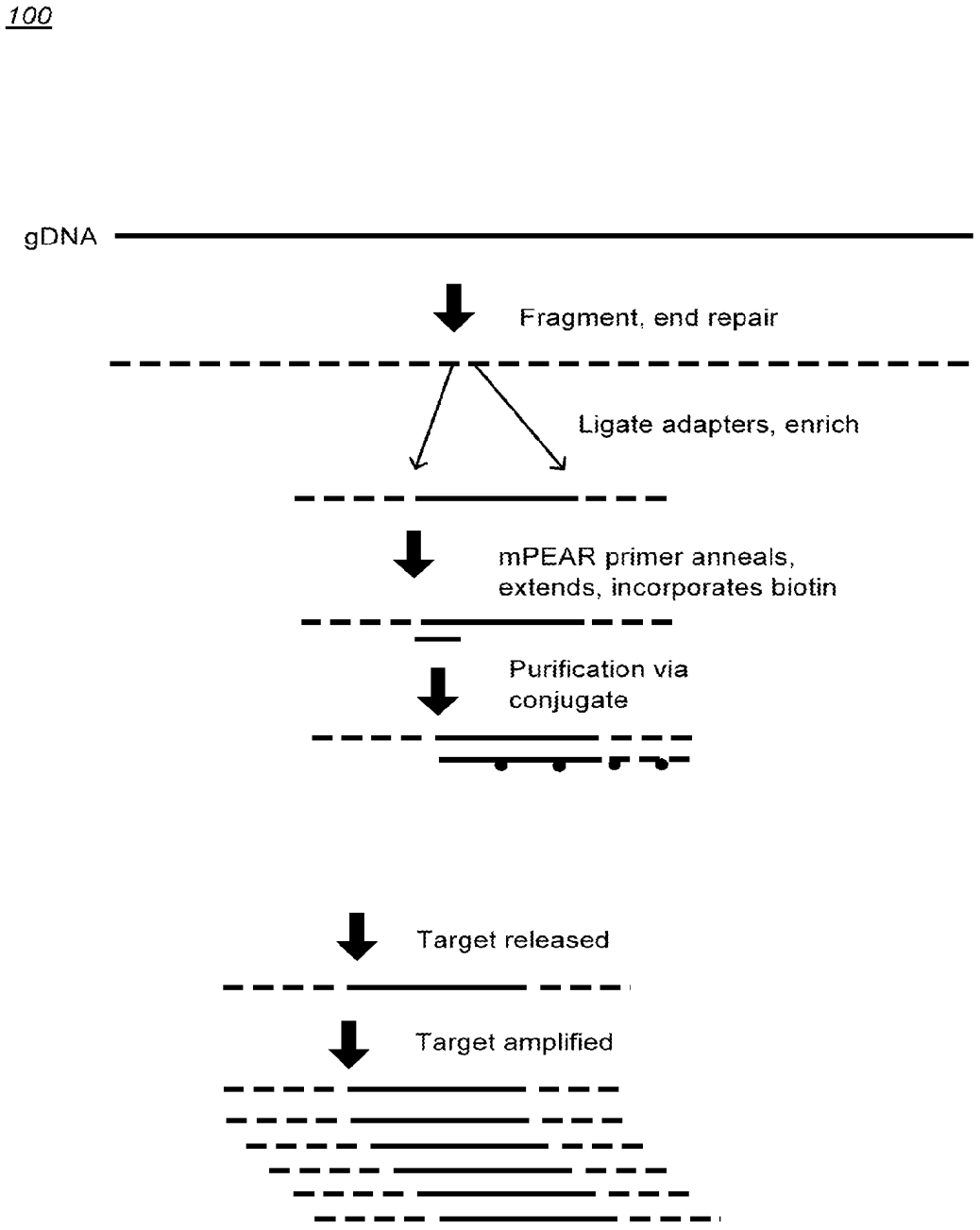


FIG. 1A

IDT Exhibit 1020.3225

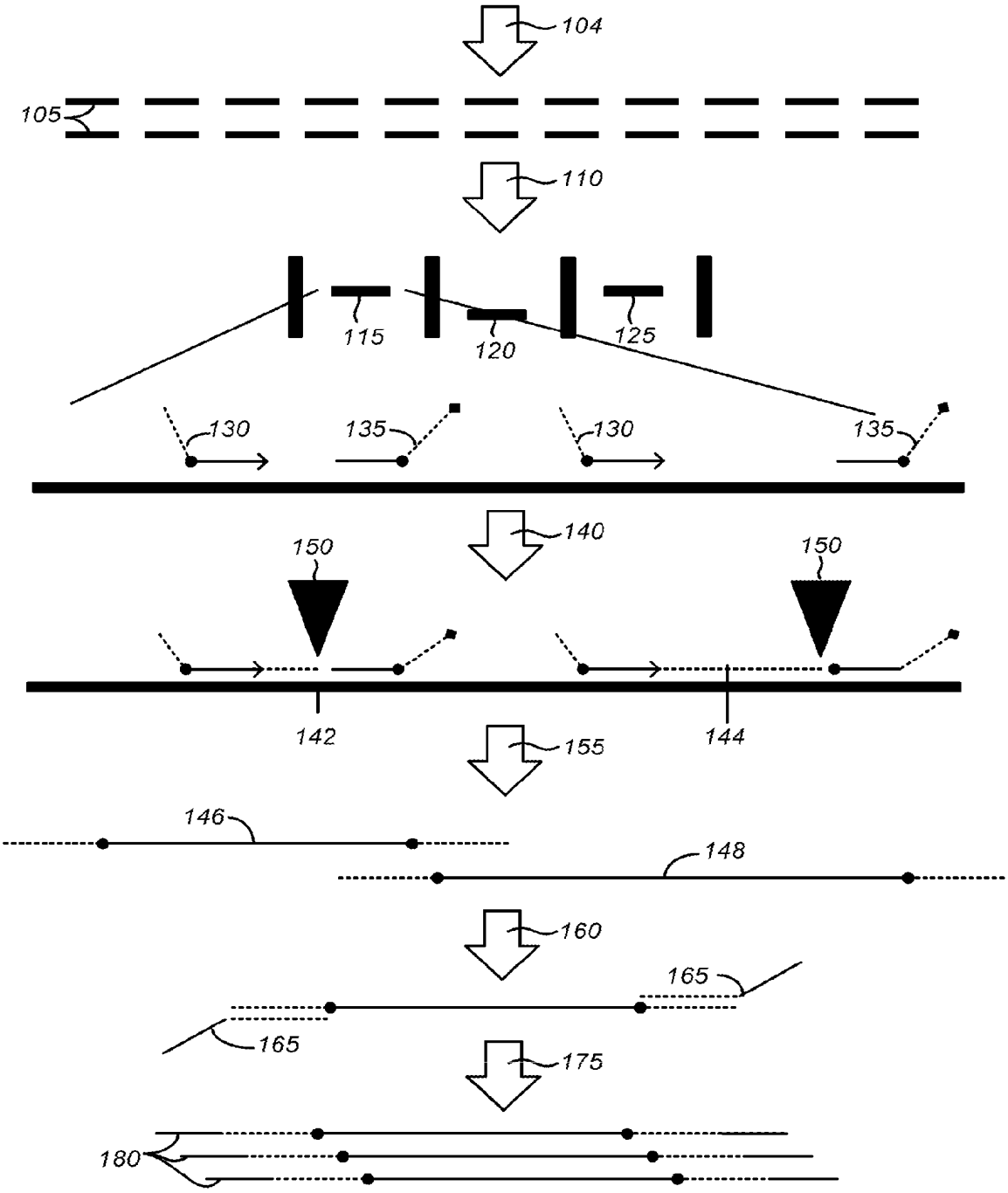


FIG. 1B

IDT Exhibit 1020.3226

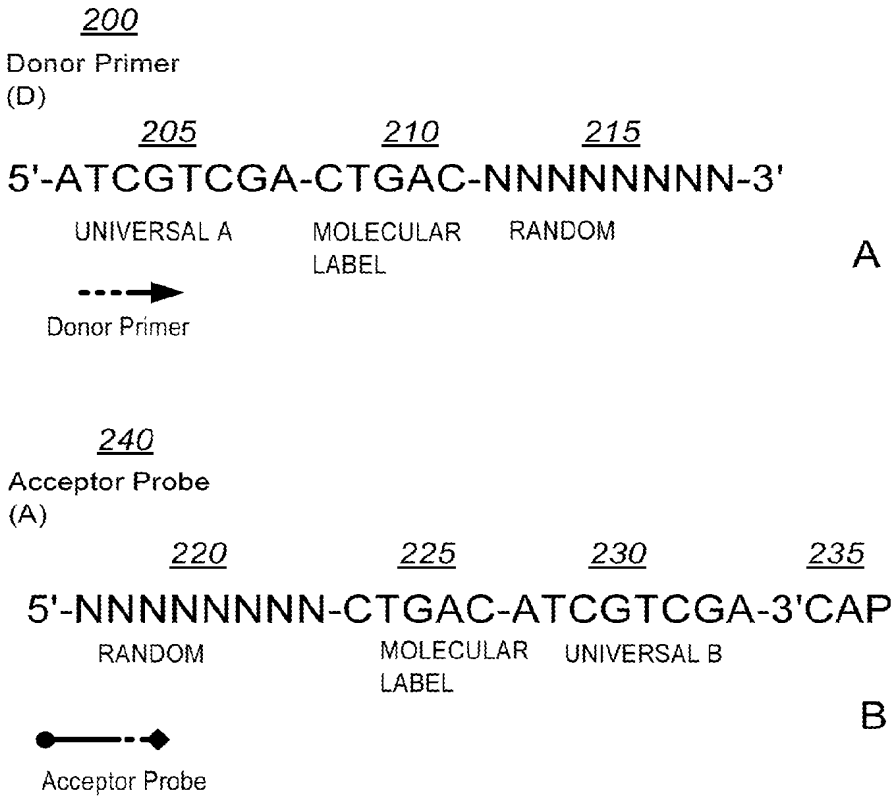


FIG. 2

IDT Exhibit 1020.3227

300

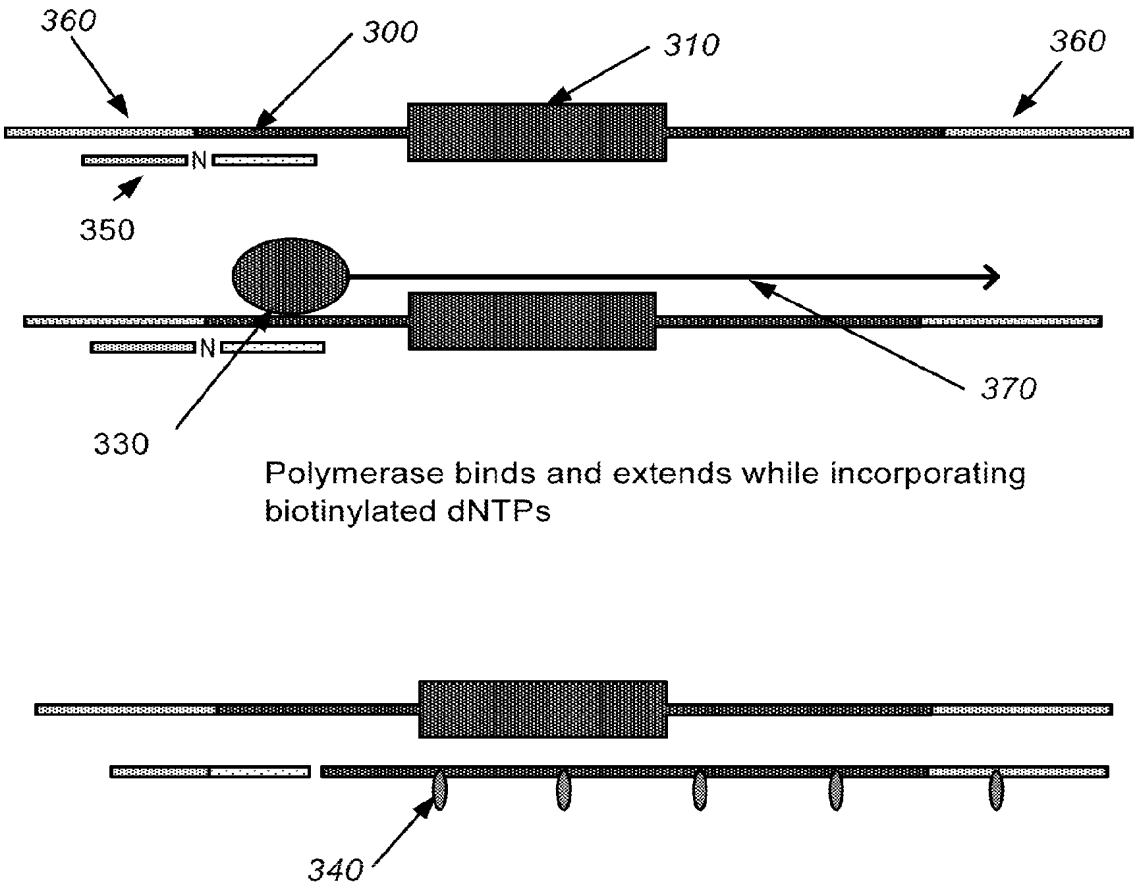


FIG. 3A

IDT Exhibit 1020.3228

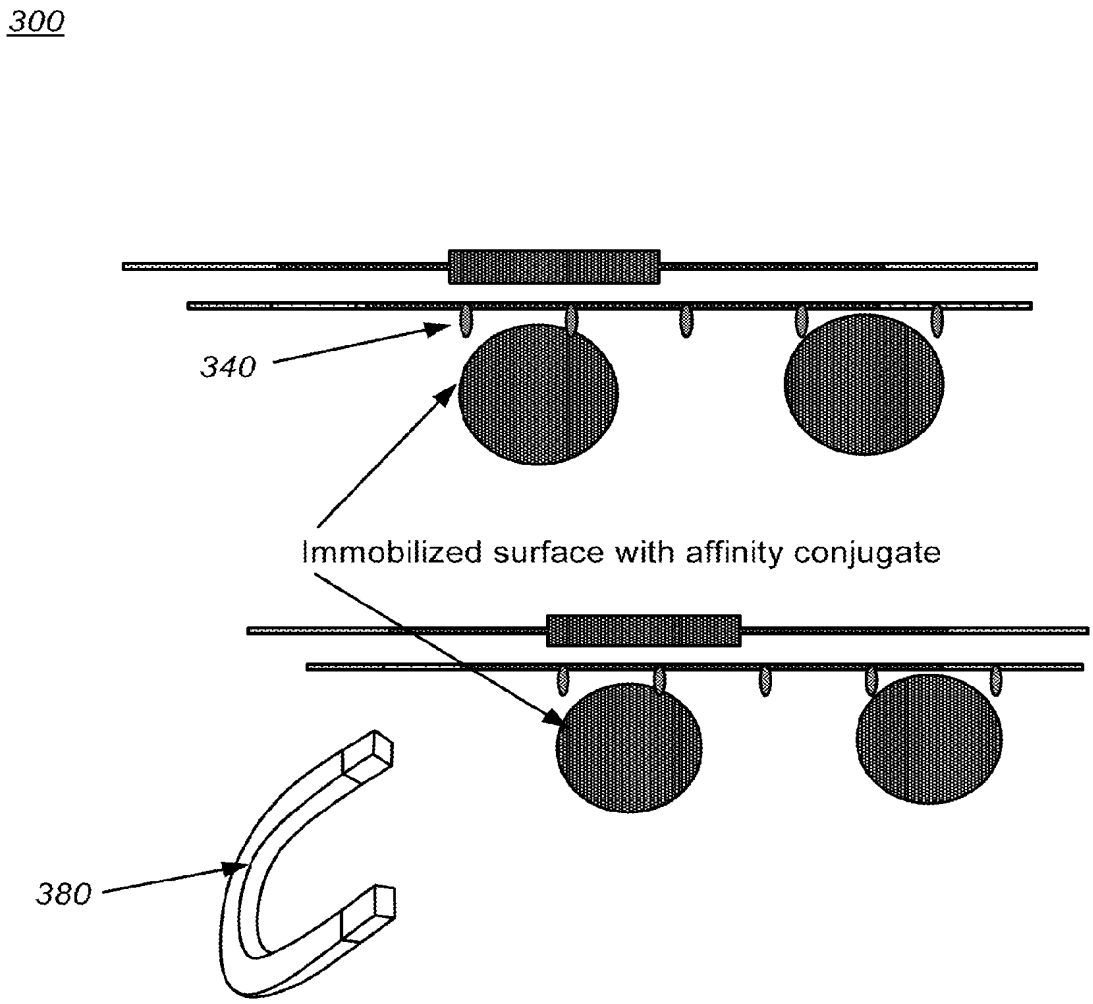


FIG. 3B

IDT Exhibit 1020.3229

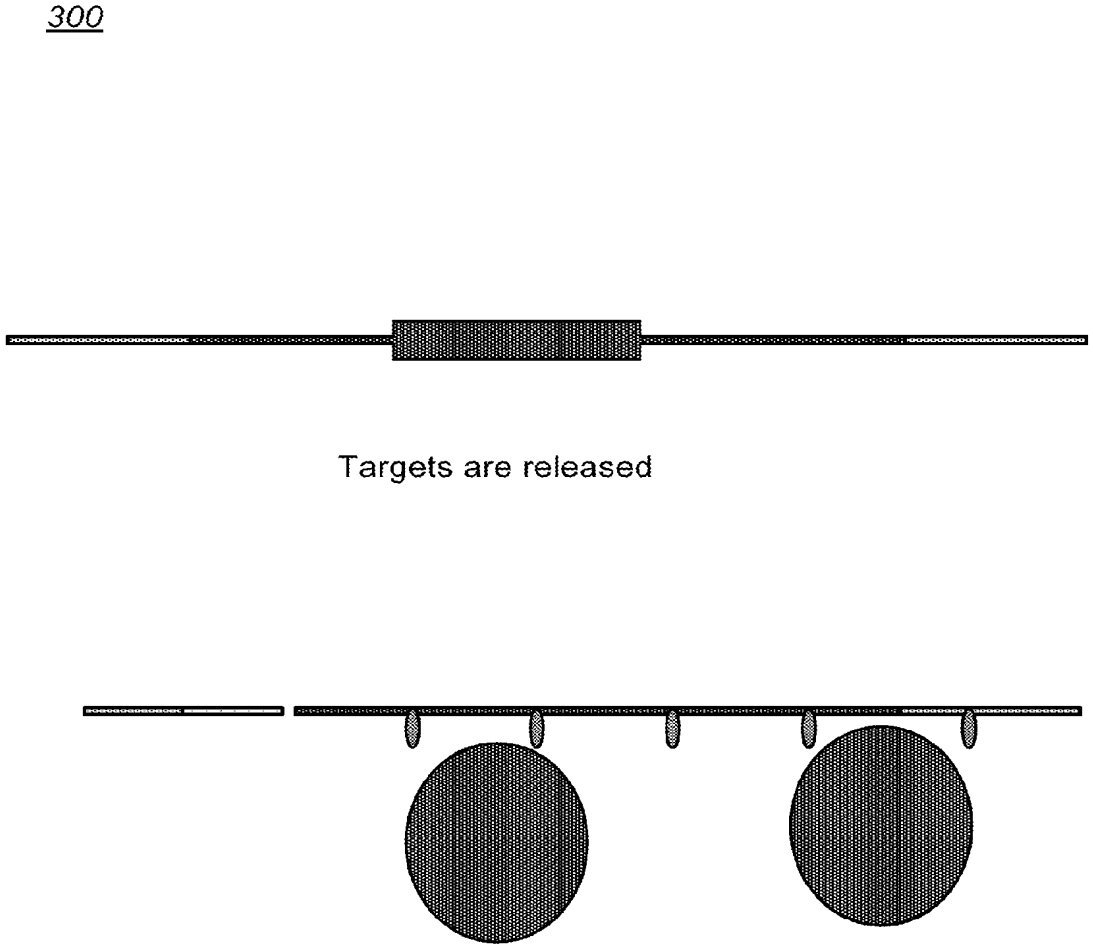


FIG. 3C

IDT Exhibit 1020.3230

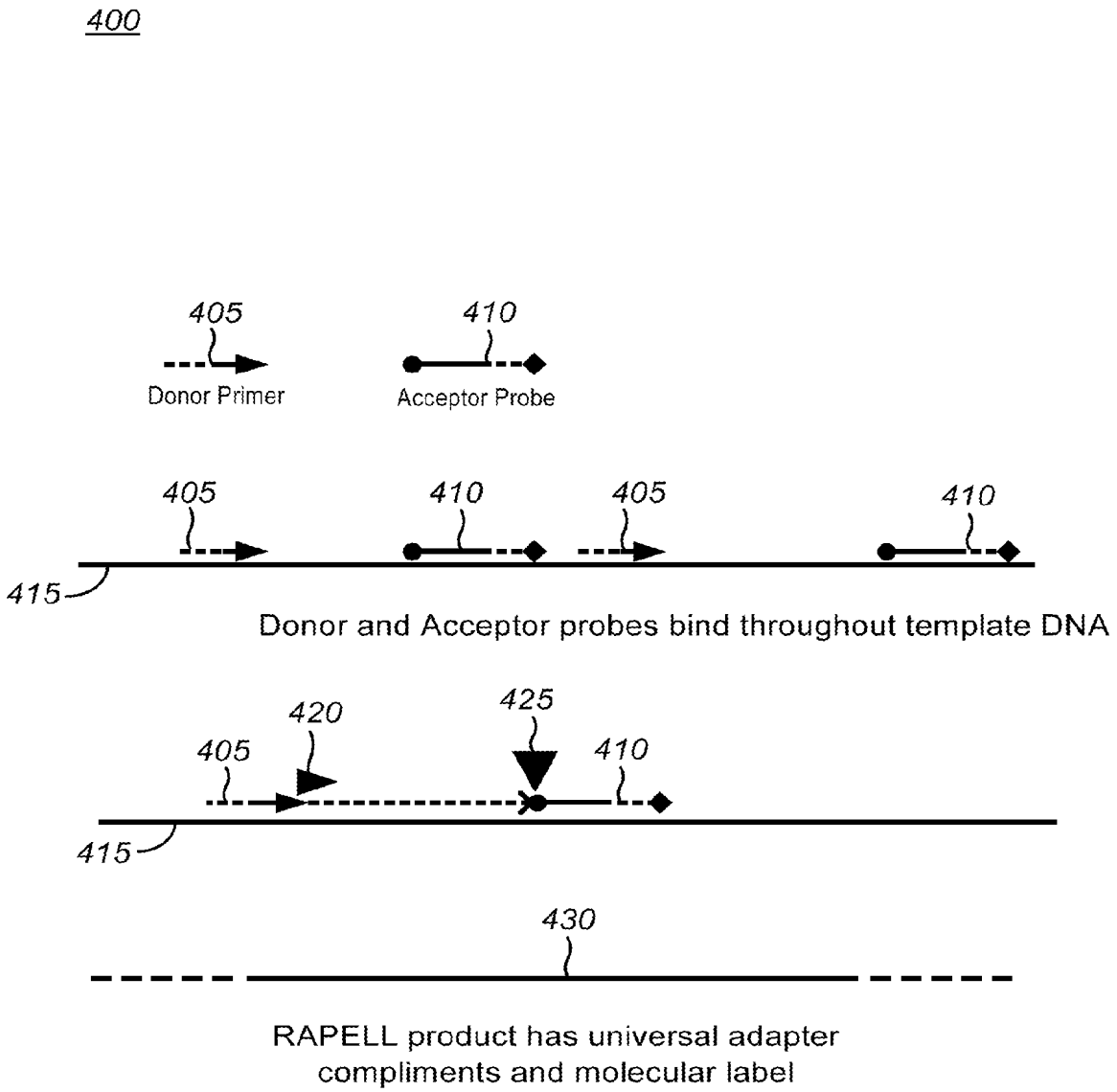


FIG. 4

IDT Exhibit 1020.3231

500

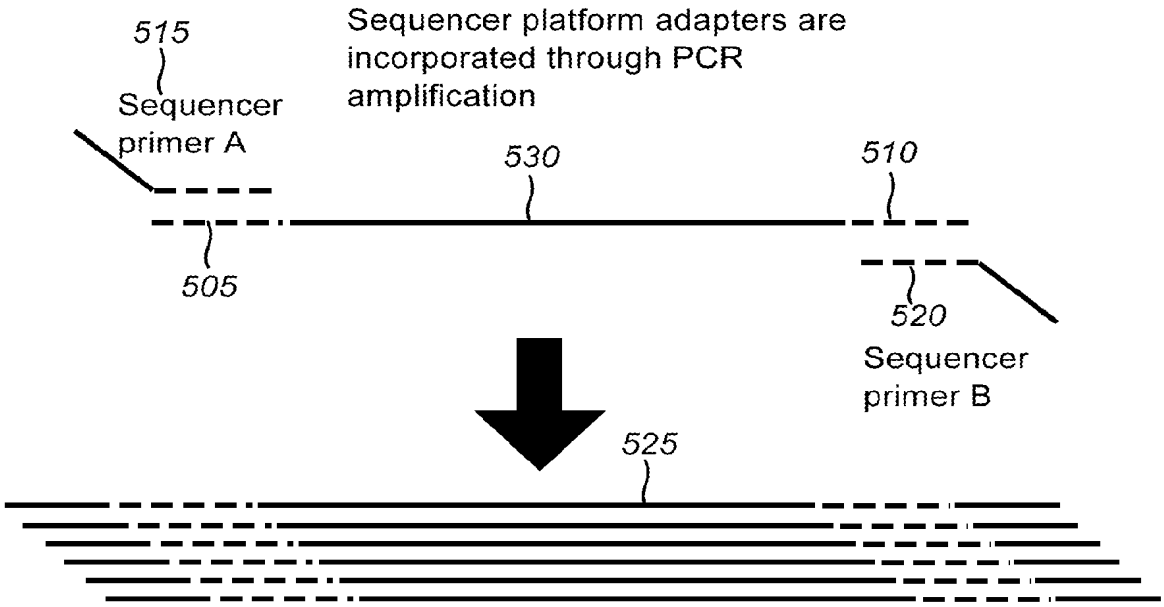


FIG. 5

IDT Exhibit 1020.3232

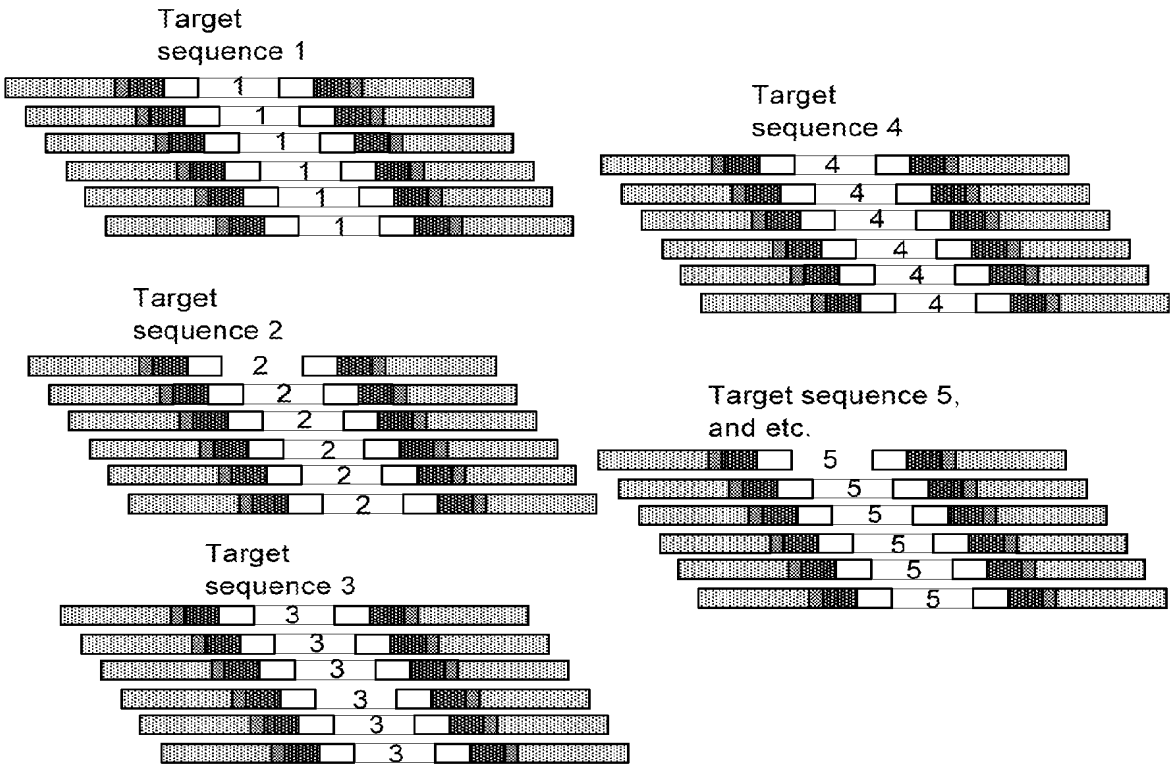


FIG. 6

IDT Exhibit 1020.3233

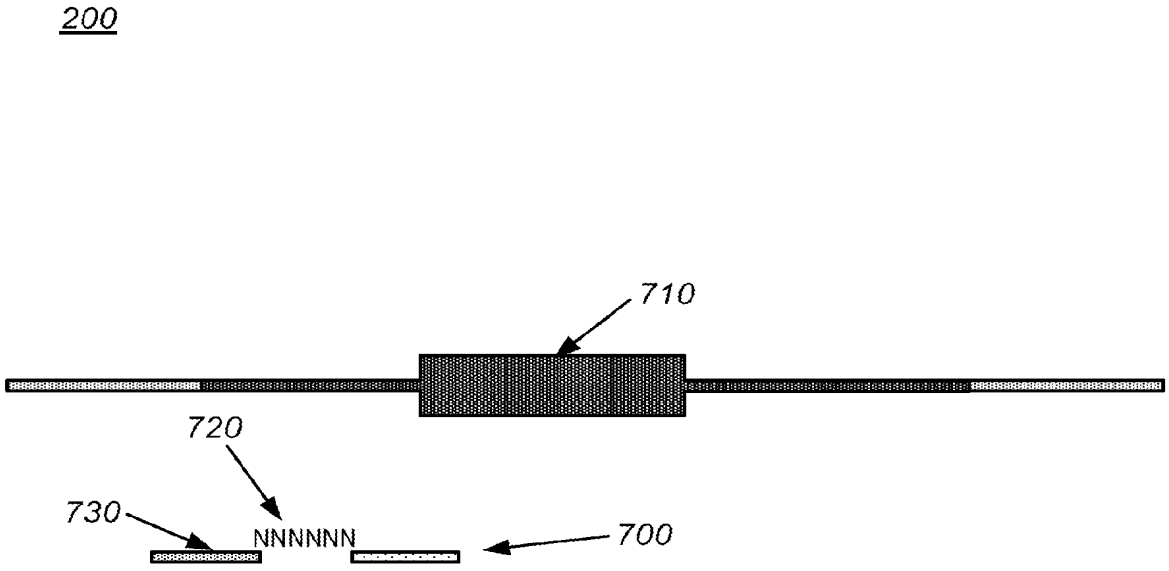


FIG. 7

IDT Exhibit 1020.3234

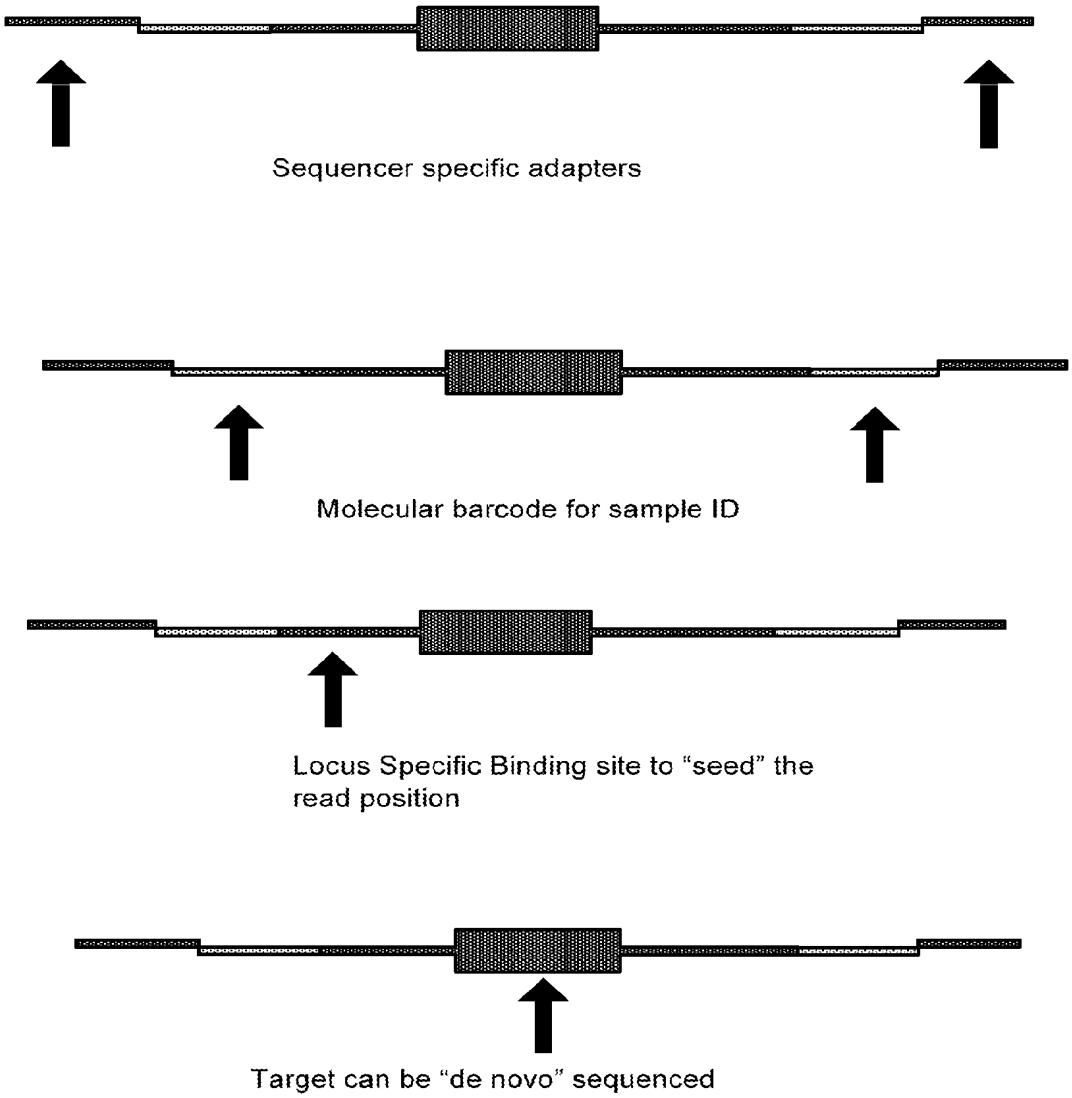


FIG. 8

IDT Exhibit 1020.3235

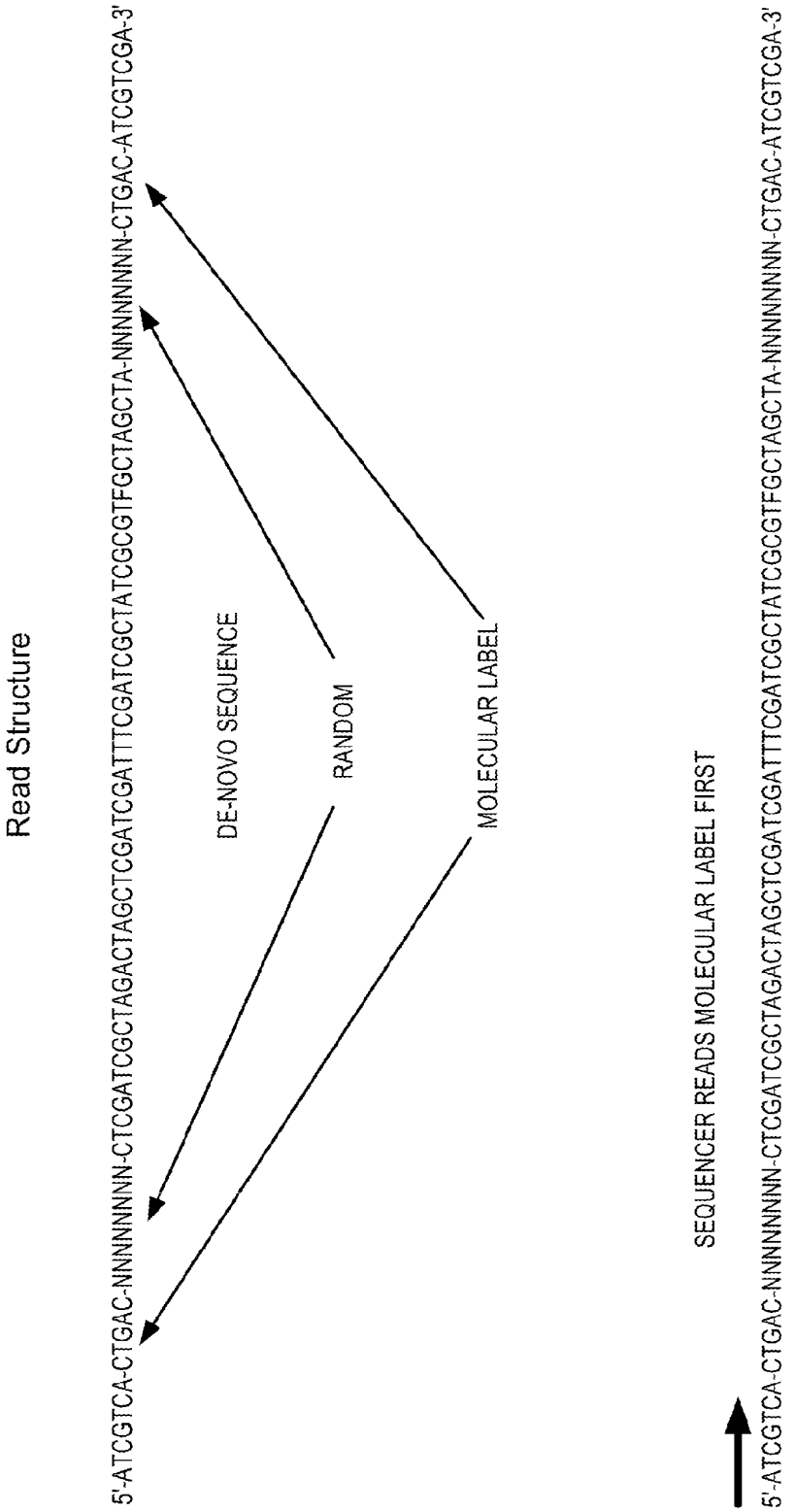


FIG. 9

IDT Exhibit 1020.3236

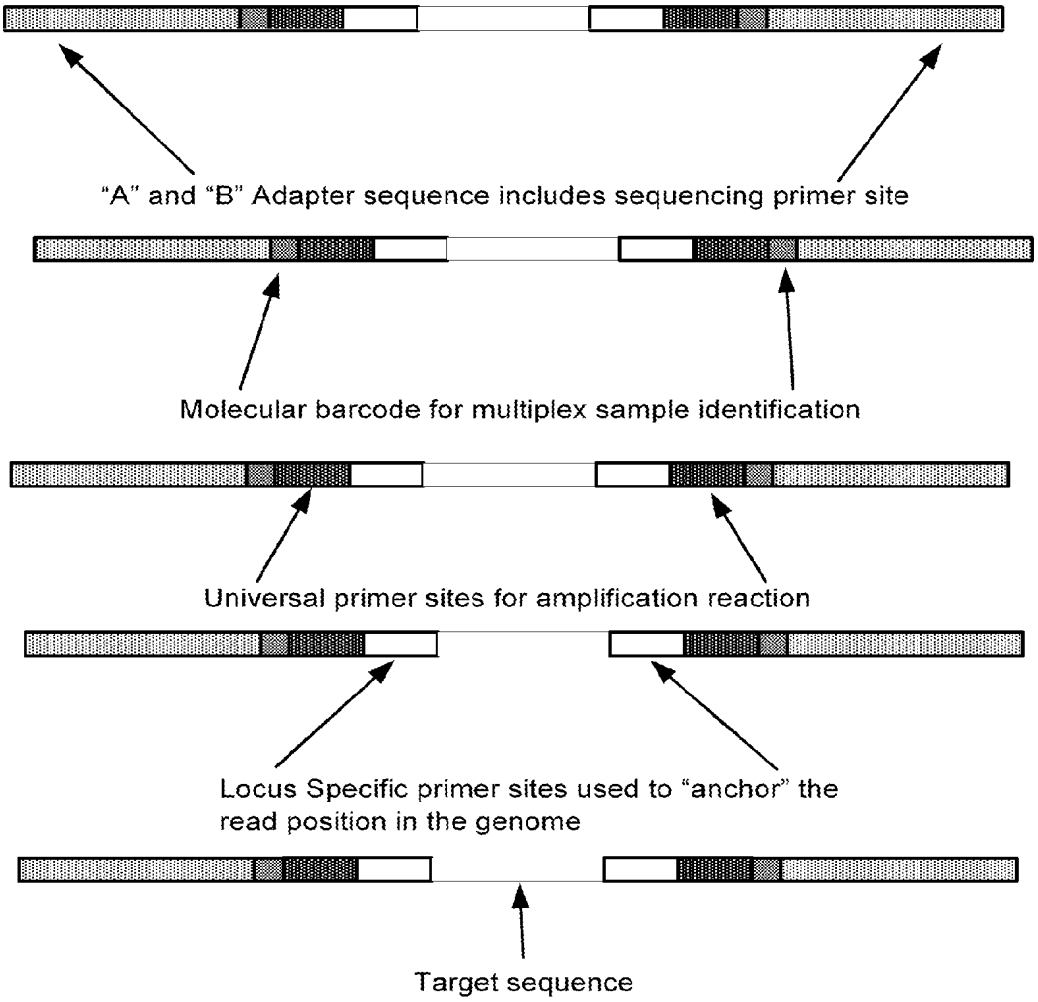


FIG. 10

IDT Exhibit 1020.3237

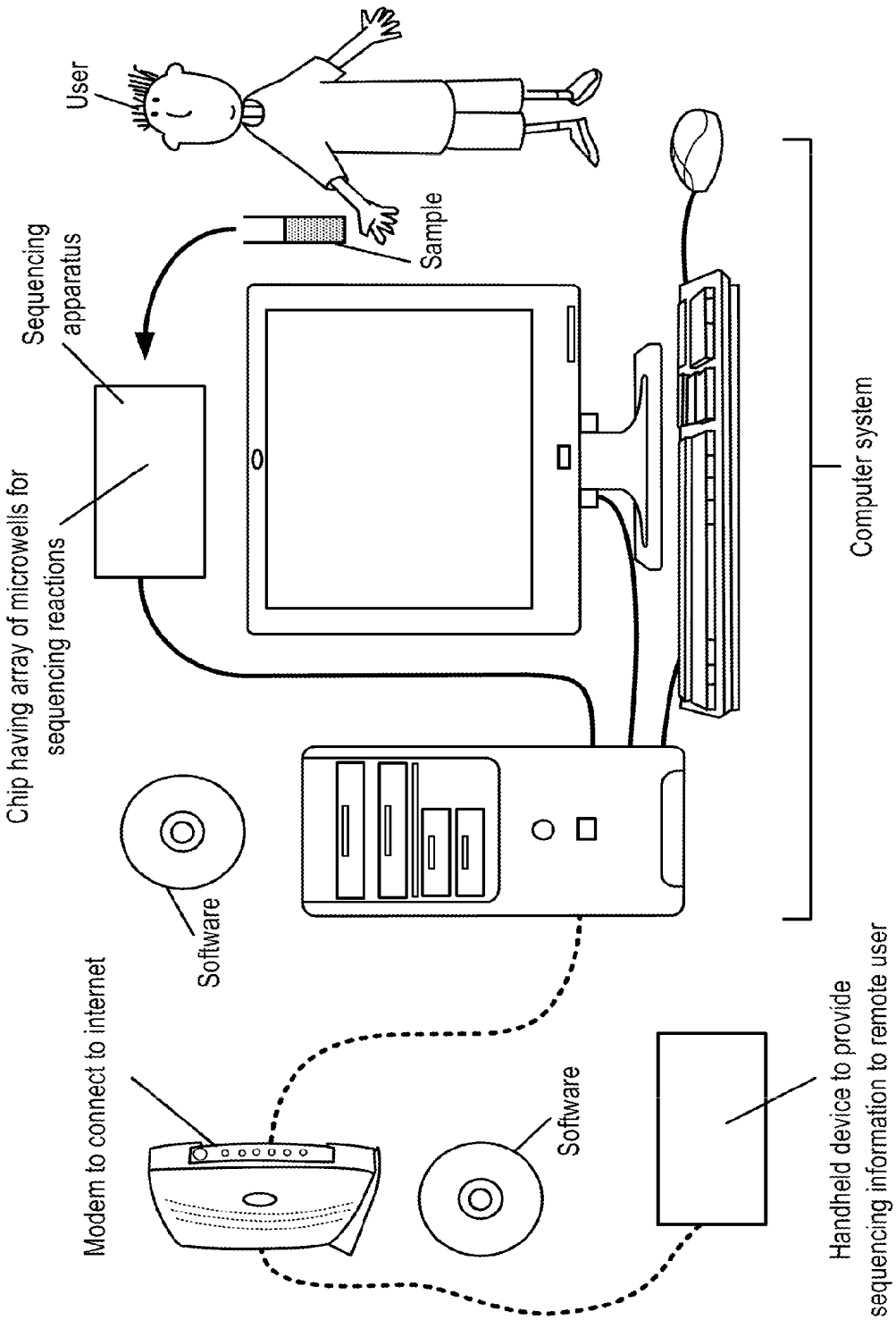


FIG. 11

IDT Exhibit 1020.3238

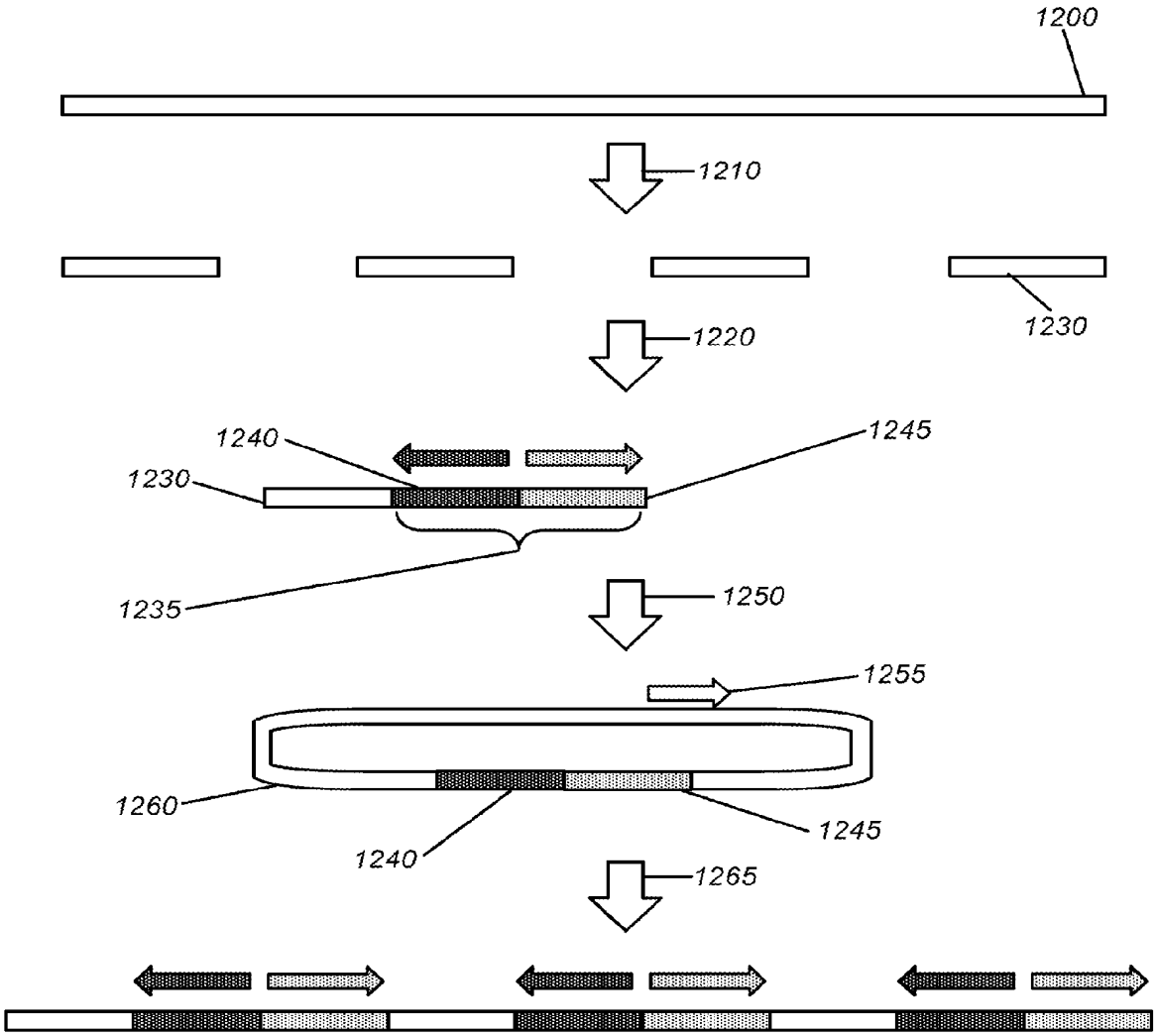


FIG. 12

IDT Exhibit 1020.3239

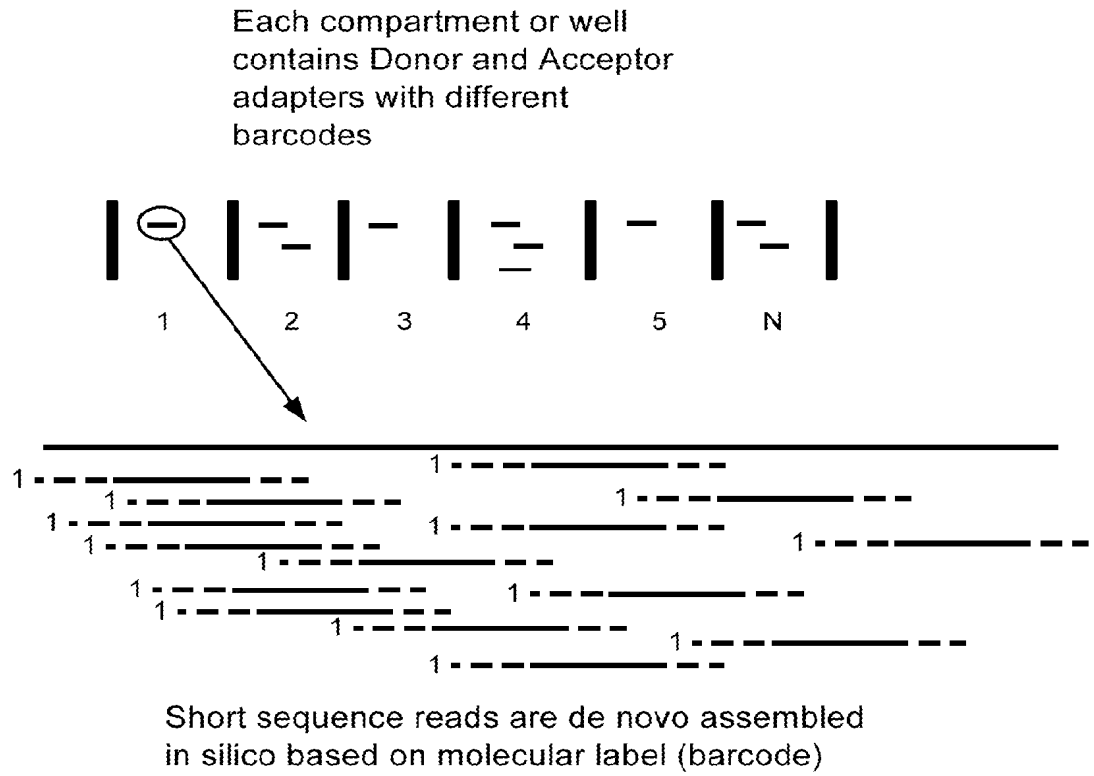


FIG. 13

IDT Exhibit 1020.3240

Truseq standard construct – multiplex (Red= barcode)

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT-XXX-AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC

TTACTATGCCGCTGGTGGCTCTAGATGTGAGAAAGGGATGTGCTGCGAGAAGGCTAGA-XXX-TCTAGCCTTCTCGTGTGCAGACTTGAGGTCAGTG

NNNNNNNATCTCGTATGCCGTCTTCTGCTTG

NNNNNNNTAGAGCATACGGCAGAAGACGAAC

Three oligos to order: Rapel 1 & 2 and PCR1 (PCR2 we already have).

Rapel-1:

barcode 13

UUT-TCCCTACACGACGCTCTTCCGATCT-AGTCAA-N8-3'

-UT-AGGGATGTGCTGCGAGAAGGCTAGA-5'

Rapel-2:

barcode 17

P-N8-AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC-GTAGAG-

ATCTCGTATGCCGTCTTCTGCTTG-amino-3'

PCR1

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT

PCR2

CAAGCAGAAGACGGCATACGAGAT

IDT order

Rapel 1: AGATCGGAAGAGCGTCGTGTAGGGAT/ideoxyU//ideoxyU//TTCCCTACACGACGCTCTTCCGATCTAGTC

AANNNNNNNN

Rapel 2: /5Phos/NNNNNNNNAGATCGGAAGAGCACACGTCTGAACTCCAGTCACGTAGAGATCTCGTATGCCGTCTTCTGCTTG

PCR 1: AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT

FIG. 14

IDT Exhibit 1020.3241

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US13/42106

A. CLASSIFICATION OF SUBJECT MATTER
IPC(8) - C12Q 1/68; C12P 19/34; G06F 19/18 (2013.01)
USPC - 435/91.2; 435/6.12; 506/26
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC(8) - C12Q 1/68; C12P 19/34; C12N 9/00; G01N 33/52, 33/53; G06F 19/18; C40B 50/06, 50/08, 40/04, 40/10 (2013.01)
USPC - 435/91.2, 6, 6.1, 6.12, 4, 89, 85, 84, 72, 41, 91.2, 91.52, 183, 91.1; 506/15, 18, 26, 27; 536/23.1, 22.1, 18.7, 1.11, 24.3, 24.33

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

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
MicroPatent (US-G, US-A, EP-A, EP-B, WO, JP-bib, DE-C,B, DE-A, DE-T, DE-U, GB-A, FR-A); DialogPRO (Derwent, INSPEC, NTIS, PASCAL, *Current Contents Search, Dissertation Abstracts Online, Inside Conferences*); *ScienceDirect; Google; Google Scholar;* separate, 'nucleic acid,' 'DNA,' 'RNA,' amplicon, primer, probe, affinity, barcode, adapter, biotin, massively, library, sequence

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X --- Y	WO 2011/142836 A9 (PIEPRZYK, M et al.) November 17, 2011; figures 8A, 8B; paragraphs [0039], [0041], [0042], [0052], [0055], [0056], [0060], [0064], [0091], [0096], [0111], [0141], [0150], [0177], [0185], [0187], [0193], [0200]-[0204], [0208], [0502]-[0508]	1-4, 6-8, 12, 13, 15-18, 20-33, 41, 48, 60-63, 66, 67, 68/66, 68/67, 69/68/66, 69/68/67 --- 5, 9-11, 14, 19, 34, 35, 42-47, 49-59, 64, 65, 70-96, 97/95, 97/96, 98/97/95, 98/97/96, 99-128, 133-152
Y	US 2009/0036325 A1 (MCKERNAN, KJ et al.) February 5, 2009; paragraphs [0002], [0003], [0007], [0016], [0019]	5, 64, 81, 152
Y	US 2007/0264642 A1 (WILLIS, TD et al.) November 15, 2007; paragraphs [0020], [0149], [0150], [0161], [0226], [0227], [0235], [0258]-[0262], [0384], [0385]	9, 103, 108-110, 113, 114
Y	US 2010/0273219 A1 (MAY, A et al.) October 28, 2010; paragraphs [0013], [0061]	10, 11
Y	US 7670810 B2 (GUNDERSON, K et al.) March 2, 2010; column 2, lines 6-25; column 3, lines 13-25; column 9, lines 63-67 to column 10, lines 1-3; column 11, lines 25-42; column 28, lines 29-48; column 37, lines 27-20; column 38, lines 8-17; column 82, lines 37-46	14, 65, 70-81
Y	US 8168385 B2 (BRENNER, S) May 1, 2012; column 1, lines 33-39; column 4, lines 8-15; column 6, lines 31-40; column 6, lines 59-67; column 14, lines 1-21	19

☒ Further documents are listed in the continuation of Box C. ☐

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

Date of the actual completion of the international search	Date of mailing of the international search report
07 October 2013 (07.10.2013)	17 OCT 2013
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-3201	Authorized officer: Shane Thomas PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

Form PCT/ISA/210 (second sheet) (July 2009)

IDT Exhibit 1020.3242

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US13/42106

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	MEYER, M et al. Targeted High-Throughput Sequencing Of Tagged Nucleic Acid Samples. Nucleic Acids Research. 01 August 2007, Vol. 35, No. 15, e97; abstract; page 3, left column, fourth paragraph to page 3, right column, fourth paragraph; page 5, left column, third paragraph. DOI:10.1093/nar/gkm566.	34, 35, 42-47, 94, 126-128
Y	US 8114978 B2 (JONES, KW et al.) February 14, 2012; column 1, lines 45-48; column 16, lines 44-65; column 23, lines 6-12; column 11, lines 2-7, 63-66; column 12, lines 62-67 to column 13, lines 1-3	45-47, 49, 50, 91, 92, 120, 137-139, 141, 142
Y	KIENZLER, R et al. Large-Scale DNA Sequence Analysis In The Cloud: A Stream-Based Approach. Euro-Par 2011: Parallel Processing Workshops Lecture Notes in Computer Science. January 2012, Vol. 7156, pp 467-476; pages 6 to 9; figures 3-5	51-59, 72-80, 143-151
Y	US 2010/0222238 A1 (SMITH, DR et al.) September 2, 2010; figure 14; paragraphs [0007], [0008], [0024], [0124]	82-96, 97/95, 97/96, 98/97/95, 98/97/96, 99-128,133-152
Y	CIPRIANY, B et al. Real-Time Analysis And Selection Of Methylated DNA By Fluorescence-Activated Single Molecule Sorting In A Nanofluidic Channel. Proceedings of the National Acadamy of Sciences. 29 May 2012 (published online before print 14 May 2012), Vol. 109, No. 22, pp. 8477-8482; abstract; page 8481, left column, second paragraph to page 8481, right column, second paragraph. DOI:10.1073/pnas.1117549109.	85
Y	HAMADY, M et al. Error-Correcting Barcoded Primers For Pyrosequencing Hundreds Of Samples In Multiplex. Nat Methods. 10 February 2008, Vol. 5, No. 3, pp. 235-237 and online supplement; abstract; page 235, right column, third paragraph; page 237, left column, first paragraph; Supplementary Data 1, page 1. DOI:10.1038/NMETH.1184.	107

IDT Exhibit 1020.3243

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US13/42106

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

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

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

3. ☒ Claims Nos.: 36-40, 129-132
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

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

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

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

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

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

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.

IDT Exhibit 1020.3244

Espacenet

Bibliographic data: WO2012103154 (A1) — 2012-08-02

STEM-LOOP COMPOSITE RNA-DNA ADAPTOR-PRIMERS: COMPOSITIONS AND METHODS FOR LIBRARY GENERATION, AMPLIFICATION AND OTHER DOWNSTREAM MANIPULATIONS

Inventor(s): KURN NURITH [US]; STAPLETON MARK [US]; WANG SHENLONG [US] ± (KURN, NURITH, ; STAPLETON, MARK, ; WANG, SHENLONG)

Applicant(s): NUGEN TECHNOLOGIES INC [US]; KURN NURITH [US]; STAPLETON MARK [US]; WANG SHENLONG [US] ± (NUGEN TECHNOLOGIES, INC, ; KURN, NURITH, ; STAPLETON, MARK, ; WANG, SHENLONG)

Classification: **- international:** *C12P19/34*
- cooperative: **C12Q1/6806** more
Global Dossier

Application number: WO2012US22448 20120124

Priority number (s): US201161435730P 20110124

Abstract of WO2012103154 (A1)

The present invention concerns manipulation of nucleic acid molecules, such as amplification, or adaptor-ligation. RNA/DNA chimeric stem-loop adaptor/primers are utilized during target nucleic acid manipulation. Methods controlling the activation of stem-loop oligonucleotides are within the scope of the invention. Methods combining multiple nucleic acid reactions in one reaction volume are envisioned in various embodiments of the invention.

IDT Exhibit 1020.3245

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
2 August 2012 (02.08.2012)

WIPO | PCT

(10) International Publication Number
WO 2012/103154 A1

(51) International Patent Classification:
C12P 19/34 (2006.01)

(21) International Application Number:
PCT/US2012/022448

(22) International Filing Date:
24 January 2012 (24.01.2012)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/435,730 24 January 2011 (24.01.2011) US

(71) Applicant *(for all designated States except US)*: **NUGEN TECHNOLOGIES, INC.** [US/US]; 201 Industrial Road, Suite 310, San Carlos, CA 94070 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **KURN, Nurith** [IL/US]; 2876 Ramona Street, Palo Alto, CA 94306 (US). **STAPLETON, Mark** [US/US]; 3170 Hambletonian Lane, Walnut Creek, CA 94598 (US). **WANG, Shenlong** [CA/US]; 8248 Saturn Park Drive, San Ramon, CA 94582 (US).

(74) Agents: **SKUBATCH, Maya** et al.; Wilson Sonsini Goodrich & Rosati, 650 Page Mill Road, Palo Alto, CA 94304-1050 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

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

WO 2012/103154 A1

(54) **Title**: STEM-LOOP COMPOSITE RNA-DNA ADAPTOR-PRIMERS: COMPOSITIONS AND METHODS FOR LIBRARY GENERATION, AMPLIFICATION AND OTHER DOWNSTREAM MANIPULATIONS

(57) **Abstract**: The present invention concerns manipulation of nucleic acid molecules, such as amplification, or adaptor-ligation. RNA/DNA chimeric stem-loop adaptor/primers are utilized during target nucleic acid manipulation. Methods controlling the activation of stem-loop oligonucleotides are within the scope of the invention. Methods combining multiple nucleic acid reactions in one reaction volume are envisioned in various embodiments of the invention.

IDT Exhibit 1020.3246

STEM-LOOP COMPOSITE RNA-DNA ADAPTOR-PRIMERS: COMPOSITIONS AND METHODS FOR LIBRARY GENERATION, AMPLIFICATION AND OTHER DOWNSTREAM MANIPULATIONS

CROSS-REFERENCE

[0001] This application claims the benefit of U.S. Provisional Application No. 61/435,730, filed January 24, 2011, which application is incorporated herein by reference.

BACKGROUND OF THE INVENTION

[0002] Advances in the study of biological molecules have been led, in part, by improvement in technologies used to characterizee the molecules or their biological reactions. In particular, the study of the nucleic acids DNA and RNA has benefited from developing technologies used for manipulating and analyzing nucleic acids, such as amplification, quantification, genotyping, and sequence analysis. Expanding use of nucleic acid based techniques demand employment of new methods and composition that exceed speed, efficiency, range and/or automation of current techniques. The instant application describes methods and compositions that improve currently available techniques for nucleic acid manipulation.

SUMMARY OF THE INVENTION

[0003] In one aspect, the invention provides methods of preparing a target nucleic acid molecule for further processing, comprising: attaching an RNA/DNA chimeric stem-loop oligonucleotide having a portion of its stem and at least a portion of its loop comprise of RNA to said target nucleic acid molecule thereby producing a stem-loop-target complex; and contacting said stem-loop-target complex with an RNase that cleaves RNA from a RNA/DNA heteroduplex, thereby preparing the stem-loop-target complex for further processing. In some embodiments, the RNase is RNase H. In some embodiments, further processing comprises nucleic acid sequencing. In some embodiments, sequencing comprises single molecule sequencing. In some embodiments, sequencing comprises sequencing by synthesis. In some embodiments, further processing comprises nucleic acid amplification. In some embodiments, a portion of the stem-loop oligonucleotide is further used to prime an amplification reaction. In some embodiments, the amplification is isothermal amplification. In some embodiments, the amplification is PCR. In some embodiments, the nucleic acid is produced by fragmenting a longer nucleic acid. In some embodiments, the target nucleic acid is blunt ended. In some embodiments, the target nucleic acid comprises a 3' overhang. In some embodiments, the target nucleic acid comprises a 5' overhang. In some embodiments, the stem-loop oligonucleotide further

IDT Exhibit 1020.3247

comprises a 3' overhang. In some embodiments, the stem-loop oligonucleotide further comprises a 5' overhang. In some embodiments, the stem-loop oligonucleotide overhang and nucleic acid overhang are at least partially complementary to each other. In some embodiments, the attaching step comprises attaching a first RNA/DNA chimeric stem-loop oligonucleotide to a first end of said target nucleic acid and a second RNA/DNA chimeric stem-loop oligonucleotide to a second end of said target nucleic acid. In some embodiments, said first and said second RNA/DNA chimeric stem-loop oligonucleotides are the same. In some embodiments, the attaching step comprises the use of a ligase. In some embodiments, the ligase is T4 ligase. In some embodiments, the stem-loop oligonucleotide lacks 5' phosphate. In some embodiments, the target nucleic acid is further extended along the cleaved stem-loop oligonucleotide sequence, thereby forming a target nucleic acid extension product, wherein a portion of the stem-loop oligonucleotide acts as a template. In some embodiments, the stem-loop oligonucleotide that is in duplex with the nucleic acid extension is further cleaved using an RNase that cleaves RNA from a RNA/DNA heteroduplex, thereby forming a priming ready construct. In some embodiments, the priming ready construct is further amplified. In some embodiments, the amplification comprises single primer isothermal amplification (SPIA). In some embodiments, at least one loop of the stem-loop oligonucleotide further comprises one or more sequence elements selected from the group consisting of a primer binding site, a sequencing primer binding site, a barcode, a promoter, an adaptor-group sequence, a restriction enzyme recognition site, a target complementary overhang, a probe complementary overhang, a probe binding site, a random sequence, and a complementary sequence to any of the sequence elements within the group. In some embodiments, at least one stem of the stem-loop oligonucleotide further comprises one or more sequence elements selected from the group consisting of a primer binding site, a sequencing primer binding site, a barcode, a promoter, an adaptor-group sequence, a restriction enzyme recognition site, a target complementary overhang, a probe complementary overhang, a probe binding site, a random sequence, and a complementary sequence to any of the sequence elements within the group. In some embodiments, at least one overhang of the stem-loop oligonucleotide further comprises one or more sequence elements selected from the group consisting of a primer binding site, a sequencing primer binding site, a barcode, a promoter, an adaptor-group sequence, a restriction enzyme recognition site, a target complementary overhang, a probe

IDT Exhibit 1020.3248

complementary overhang, a probe binding site, a random sequence, and a complementary sequence to any of the sequence elements within the group.

[0004] In another aspect, the invention provides methods for amplifying a nucleic acid molecule, comprising: contacting in a single mixture an RNA/DNA stem-loop adaptor having a portion of its stem and at least a portion of its loop comprise of RNA; a double-stranded nucleic acid with a 3' overhang; RNase that cleaves RNA from a RNA/DNA heteroduplex; a DNA polymerase with strand displacement activity. Some embodiments, further comprise forming an extension product by priming the double-stranded nucleic acid at its 3' overhang using at least a portion of the stem-loop oligonucleotide. In some embodiments, the loop-RNA in the extension product is further cleaved by the RNase, thereby forming a second double stranded nucleic acid with at least one 3' overhang. In some embodiments, the loop-RNA in the extension product is further cleaved by the RNase, thereby forming a second double stranded nucleic acid with at least one 3' overhang. In some embodiments, the RNase comprises RNase H. Some embodiments further comprise at least a second amplification mixture with at least a second stem-loop oligonucleotide, wherein each stem-loop oligonucleotide further comprises a barcode sequence. In some embodiments, the 3' overhang is produced by forming a product, wherein the stem-loop oligonucleotide is attached to a double stranded nucleic acid by one strand; extending the double stranded nucleic acid along the stem-loop oligonucleotide; and cleaving the stem-loop oligonucleotide using an RNase that cleaves RNA sequences in an RNA/DNA heteroduplex. In some embodiments, the 3' overhang is produced during single primer isothermal amplification.

[0005] In yet another aspect, the invention provides kits for amplifying nucleic acids, comprising: a stem-loop oligonucleotide, wherein at least a portion of the stem comprises RNA and wherein at least a portion of the loop comprises RNA; an RNase that cleaves RNA sequences in an RNA/DNA duplex; a ligase; and a DNA polymerase.

[0006] In a further aspect, the invention provides methods of amplifying a double stranded target nucleic acid molecule, comprising: (a) attaching a strand of a first stem-loop oligonucleotide to said double stranded target nucleic acid molecule thereby producing an oligonucleotide-attached nucleic acid molecule; and (b) cleaving part of the stem-loop oligonucleotide using an RNase; and using a second stem-loop oligonucleotide as a primer in amplifying the double stranded target nucleic acid molecule. In some embodiments, the second stem-loop oligonucleotide is an exact copy of the first stem-loop oligonucleotide. In some embodiments, the RNase comprises RNase H. In some

IDT Exhibit 1020.3249

embodiments, the stem-loop oligonucleotide comprises an RNA/DNA heteroduplex in a stem. In some embodiments, a strand of the oligonucleotide-attached nucleic acid molecule is extended. In some embodiments, the extension yields a second RNA/DNA heteroduplex. In some embodiments, cleaving part of the stem-loop oligonucleotide comprises cleaving the second RNA/DNA heteroduplex. In some embodiments, cleaving part of the stem-loop oligonucleotide generates a priming ready construct. In some embodiments, the priming ready construct comprises a single stranded region that is 3' to the double stranded target nucleic acid. In some embodiments, the amplification comprises PCR. In some embodiments, the amplification comprises SPIA. In some embodiments, the attaching comprises ligation. In some embodiments, T4 ligase is utilized in ligation. In some embodiments, the stem-loop oligonucleotide lacks 5' phosphate. In some embodiments, the target nucleic acid is produced by fragmenting a longer nucleic acid. In some embodiments, the target nucleic acid is blunt ended. In some embodiments, the target nucleic acid comprises a 3' overhang. In some embodiments, the target nucleic acid comprises a 5' overhang. In some embodiments, the stem-loop oligonucleotide further comprises a 3' overhang. In some embodiments, the stem-loop oligonucleotide further comprises a 5' overhang. In some embodiments, the stem-loop oligonucleotide overhang and nucleic acid overhang are at least partially complementary to each other. In some embodiments, at least one loop of the stem-loop oligonucleotide further comprises one or more sequence elements selected from the group consisting of a primer binding site, a sequencing primer binding site, a barcode, a promoter, an adaptor-group sequence, a restriction enzyme recognition site, a target complementary overhang, a probe complementary overhang, a probe binding site, a random sequence, and a complementary sequence to any of the sequence elements within the group. In some embodiments, at least one stem of the stem-loop oligonucleotide further comprises one or more sequence elements selected from the group consisting of a primer binding site, a sequencing primer binding site, a barcode, a promoter, an adaptor-group sequence, a restriction enzyme recognition site, a target complementary overhang, a probe complementary overhang, a probe binding site, a random sequence, and a complementary sequence to any of the sequence elements within the group. In some embodiments, at least one overhang of the stem-loop oligonucleotide further comprises one or more sequence elements selected from the group consisting of a primer binding site, a sequencing primer binding site, a barcode, a promoter, an adaptor-group sequence, a restriction enzyme recognition site, a target complementary overhang, a probe

IDT Exhibit 1020.3250

complementary overhang, a probe binding site, a random sequence, and a complementary sequence to any of the sequence elements within the group.

INCORPORATION BY REFERENCE

[0007] All publications and patent applications mentioned in this specification are herein incorporated by reference to the same extent as if each individual publication or patent application was specifically and individually indicated to be incorporated by reference.

BRIEF DESCRIPTION OF THE DRAWINGS

[0008] The novel features of the invention are set forth with particularity in the appended claims. A better understanding of the features and advantages of the present invention will be obtained by reference to the following detailed description that sets forth illustrative embodiments, in which the principles of the invention are utilized, and the accompanying drawings of which:

[0009] Figure 1 illustrates examples of stem-loop oligonucleotides according to embodiments of the invention.

[0010] Figure 2 A-B illustrate two different methods of ligating stem-loop oligonucleotides to a double stranded target nucleic acid followed by nucleic acid extension and RNA cleavage generating priming ready constructs.

[0011] Figure 3 illustrates a method for appending disparate sequences to the two ends of a target nucleic acid according to the embodiments of the invention.

[0012] Figure 4 A-B illustrate methods for performing single primer isothermal amplification (SPIA) using stem-loop oligonucleotide (A) or linear (B) primers according to embodiments of the invention.

[0013] Figure 5 illustrates the use of stem-loop oligonucleotides as pro-primers and primers according to some embodiments of the invention.

[0014] Figure 6 illustrates in μg, the yields of whole genome amplification (WGA) reactions utilizing either the stem-loop chimeric amplification primer or the corresponding linear chimeric amplification primer.

[0015] Figure 7 illustrates the BioAnalyzer Profile of whole genome amplification (WGA) products demonstrating the florescent units as a function of size in nucleotides. Line 1, stem-loop primer with 100ng FFPE sample; line 2, linear primer with 50 ng FFPE sample; and line 3, linear primer with 100ng FFPE sample.

IDT Exhibit 1020.3251

DEFINITIONS

[0016] The term "nucleic acid" refers to a nucleotide polymer, and unless otherwise limited, includes known analogs of natural nucleotides that can function in a similar manner (e.g., hybridize) to naturally occurring nucleotides.

[0017] The terms "polynucleotide", "nucleotide", "nucleotide sequence", "nucleic acid" and "oligonucleotide" are used interchangeably. They refer to a polymeric form of nucleotides of any length, either deoxyribonucleotides or ribonucleotides, or analogs thereof. Polynucleotides may have any three dimensional structure, and may perform any function, known or unknown. The following are non limiting examples of polynucleotides: coding or non-coding regions of a gene or gene fragment, intergenic DNA, loci (locus) defined from linkage analysis, exons, introns, messenger RNA (mRNA), transfer RNA, ribosomal RNA, short interfering RNA (siRNA), short-hairpin RNA (shRNA), micro-RNA (miRNA), small nucleolar RNA, ribozymes, complementary DNA (cDNA), which is a DNA representation of mRNA, usually obtained by reverse transcription of messenger RNA (mRNA) or by amplification; DNA molecules produced synthetically or by amplification, genomic DNA, recombinant polynucleotides, branched polynucleotides, plasmids, vectors, isolated DNA of any sequence, isolated RNA of any sequence, nucleic acid probes, and primers. A polynucleotide may comprise modified nucleotides, such as methylated nucleotides and nucleotide analogs. If present, modifications to the nucleotide structure may be imparted before or after assembly of the polymer. The sequence of nucleotides may be interrupted by non nucleotide components. A polynucleotide may be further modified after polymerization, such as by conjugation with a labeling component. Polynucleotide sequences, when provided, are listed in the 5' to 3' direction, unless stated otherwise.

[0018] The term nucleic acid encompasses double- or triple-stranded nucleic acids, as well as single-stranded molecules. In double- or triple-stranded nucleic acids, the nucleic acid strands need not be coextensive (i.e., a double-stranded nucleic acid need not be double-stranded along the entire length of both strands).

[0019] The term nucleic acid also encompasses any chemical modification thereof, such as by methylation and/or by capping. Nucleic acid modifications can include addition of chemical groups that incorporate additional charge, polarizability, hydrogen bonding, electrostatic interaction, and functionality to the individual nucleic acid bases or to the nucleic acid as a whole. Such modifications may include base modifications such as 2′-position sugar modifications, 5-position pyrimidine modifications, 8-position purine

modifications, modifications at cytosine exocyclic amines, substitutions of 5-bromo-uracil, backbone modifications, unusual base pairing combinations such as the isobases isocytidine and isoguanidine, and the like.

[0020] More particularly, in certain embodiments, nucleic acids, can include polydeoxyribonucleotides (containing 2-deoxy-D-ribose), polyribonucleotides (containing D-ribose), and any other type of nucleic acid that is an N- or C-glycoside of a purine or pyrimidine base, as well as other polymers containing normucleotidic backbones, for example, polyamide (e.g., peptide nucleic acids (PNAs)) and polymorpholino (commercially available from the Anti-Virals, Inc., Corvallis, Oreg., as Neugene) polymers, and other synthetic sequence-specific nucleic acid polymers providing that the polymers contain nucleobases in a configuration which allows for base pairing and base stacking, such as is found in DNA and RNA. The term nucleic acid also encompasses linked nucleic acids (LNAs), which are described in U.S. Pat. Nos. 6,794,499, 6,670,461, 6,262,490, and 6,770,748, which are incorporated herein by reference in their entirety for their disclosure of LNAs.

[0021] The nucleic acid(s) can be derived from a completely chemical synthesis process, such as a solid phase-mediated chemical synthesis, from a biological source, such as through isolation from any species that produces nucleic acid, or from processes that involve the manipulation of nucleic acids by molecular biology tools, such as DNA replication, PCR amplification, reverse transcription, or from a combination of those processes.

[0022] As used herein, the term "target polynucleotide" or “target nucleic acid” refer to a nucleic acid molecule or polynucleotide in a starting population of nucleic acid molecules having a target sequence whose presence, amount, and/or nucleotide sequence, or changes in these, are desired to be determined. In general, a target polynucleotide is a double-stranded nucleic acid molecule, and may be derived from any source of or process for generating double-stranded nucleic acid molecules.

[0023] As used herein, the term "target sequence" refers generally to a nucleic acid sequence on a single strand of nucleic acid. The target sequence may be a portion of a gene, a regulatory sequence, genomic DNA, cDNA, RNA including mRNA, miRNA, and rRNA, or others. The target sequence may be a target sequence from a sample or a secondary target such as a product of an amplification reaction. The “target sequence” can be embedded in a molecule that includes the nucleotide sequence of a target nucleic acid, such as, for example, the amplification product obtained by amplifying a target nucleic acid or the cDNA produced upon reverse transcription of an RNA target nucleic acid.

IDT Exhibit 1020.3253

[0024] A "nucleotide probe" or "probe" refers to a polynucleotide used for detecting or identifying its corresponding target polynucleotide in a hybridization reaction. Thus, a probe is hybridizable to one or more target polynucleotides. Probes can be perfectly complementary to one or more target polynucleotides in a sample, or contain one or more nucleotides that are not complemented by a corresponding nucleotide in the one or more target polynucleotides in a sample.

[0025] As used herein, the term "complementary" refers to the capacity for precise pairing between two nucleotides. I.e., if a nucleotide at a given position of a nucleic acid is capable of hydrogen bonding with a nucleotide of another nucleic acid, then the two nucleic acids are considered to be complementary to one another at that position. Complementarity between two single-stranded nucleic acid molecules may be "partial," in which only some of the nucleotides bind, or it may be complete when total complementarity exists between the single-stranded molecules. The degree of complementarity between nucleic acid strands has significant effects on the efficiency and strength of hybridization between nucleic acid strands.

[0026] "Hybridization" and "annealing" refer to a reaction in which one or more polynucleotides react to form a complex that is stabilized via hydrogen bonding between the bases of the nucleotide residues. The hydrogen bonding may occur by Watson Crick base pairing, Hoogstein binding, or in any other sequence specific manner. The complex may comprise two strands forming a duplex structure, three or more strands forming a multi stranded complex, a single self hybridizing strand, or any combination of these. A hybridization reaction may constitute a step in a more extensive process, such as the initiation of a PCR or other amplification reactions, or the enzymatic cleavage of a polynucleotide by a ribozyme. A first sequence that can be stabilized via hydrogen bonding with the bases of the nucleotide residues of a second sequence is said to be "hybridizable" to said second sequence. In such a case, the second sequence can also be said to be hybridizable to the first sequence.

[0027] In general, a "complement" of a given sequence is a sequence that is fully or substantially complementary to and hybridizable to the given sequence. In general, a first sequence that is hybridizable to a second sequence or set of second sequences is specifically or selectively hybridizable to the second sequence or set of second sequences, such that hybridization to the second sequence or set of second sequences is preferred (e.g. thermodynamically more stable under a given set of conditions, such as stringent conditions commonly used in the art) to hybridization with non-target sequences during a

IDT Exhibit 1020.3254

hybridization reaction. Typically, hybridizable sequences share a degree of sequence complementarity over all or a portion of their respective lengths, such as between 25%-100% complementarity, including at least about 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, and 100% sequence complementarity.

[0028] The term "hybridized" as applied to a polynucleotide refers to a polynucleotide in a complex that is stabilized via hydrogen bonding between the bases of the nucleotide residues. The hydrogen bonding may occur by Watson Crick base pairing, Hoogstein binding, or in any other sequence specific manner. The complex may comprise two strands forming a duplex structure, three or more strands forming a multi stranded complex, a single self hybridizing strand, or any combination of these. The hybridization reaction may constitute a step in a more extensive process, such as the initiation of a PCR reaction, or the enzymatic cleavage of a polynucleotide by a ribozyme. A sequence hybridized with a given sequence is referred to as the "complement" of the given sequence.

[0029] "Specific hybridization" refers to the binding of a nucleic acid to a target nucleotide sequence in the absence of substantial binding to other nucleotide sequences present in the hybridization mixture under defined stringency conditions. Those of skill in the art recognize that relaxing the stringency of the hybridization conditions allows sequence mismatches to be tolerated.

[0030] As used herein, the term "complementary" refers to the capacity for precise pairing between two nucleotides. I.e., if a nucleotide at a given position of a nucleic acid is capable of hydrogen bonding with a nucleotide of another nucleic acid, then the two nucleic acids are considered to be complementary to one another at that position. Complementarity between two single-stranded nucleic acid molecules may be "partial," in which only some of the nucleotides bind, or it may be complete when total complementarity exists between the single-stranded molecules. The degree of complementarity between nucleic acid strands has significant effects on the efficiency and strength of hybridization between nucleic acid strands.

[0031] The term "oligonucleotide" is generally used to refer to a nucleic acid that is relatively short, generally shorter than 200 nucleotides, more particularly, shorter than 100 nucleotides, most particularly, shorter than 50 nucleotides. Typically, oligonucleotides are single-stranded DNA molecules.

IDT Exhibit 1020.3255

[0032] The term "primer" refers to an oligonucleotide that is capable of hybridizing (also termed "annealing") with a nucleic acid and serving as an initiation site for nucleotide (RNA or DNA) polymerization under appropriate conditions (i.e., in the presence of four different nucleoside triphosphates and an agent for polymerization, such as DNA or RNA polymerase or reverse transcriptase) in an appropriate buffer and at a suitable temperature. The appropriate length of a primer depends on the intended use of the primer, but primers are typically at least 7 nucleotides long and, more typically range from 10 to 30 nucleotides, or even more typically from 15 to 30 nucleotides, in length. Other primers can be somewhat longer, e.g., 30 to 50 nucleotides long. In this context, "primer length" refers to the portion of an oligonucleotide or nucleic acid that hybridizes to a complementary "target" sequence and primes nucleotide synthesis. Short primer molecules generally require cooler temperatures to form sufficiently stable hybrid complexes with the template. A primer need not reflect the exact sequence of the template but must be sufficiently complementary to hybridize with a template. The term "primer site" or "primer binding site" refers to the segment of the target nucleic acid to which a primer hybridizes. A construct with presenting a primer binding site is often referred to as a "priming ready construct" or "amplification ready construct".

[0033] A primer is said to anneal to another nucleic acid if the primer, or a portion thereof, hybridizes to a nucleotide sequence within the nucleic acid. The statement that a primer hybridizes to a particular nucleotide sequence is not intended to imply that the primer hybridizes either completely or exclusively to that nucleotide sequence.

[0034] As used herein, "expression" refers to the process by which a polynucleotide is transcribed into mRNA and/or the process by which the transcribed mRNA (also referred to as "transcript") is subsequently being translated into peptides, polypeptides, or proteins. The transcripts and the encoded polypeptides are collectively referred to as "gene product." If the polynucleotide is derived from genomic DNA, expression may include splicing of the mRNA in a eukaryotic cell.

DETAILED DESCRIPTION OF THE INVENTION

[0035] The practice of the present invention employs, unless otherwise indicated, conventional techniques of immunology, biochemistry, chemistry, molecular biology, microbiology, cell biology, genomics and recombinant DNA, which are within the skill of the art. See Sambrook, Fritsch and Maniatis, MOLECULAR CLONING: A LABORATORY MANUAL, 2nd edition (1989); CURRENT PROTOCOLS IN MOLECULAR

IDT Exhibit 1020.3256

BIOLOGY (F. M. Ausubel, et al. eds., (1987)); the series METHODS IN ENZYMOLOGY (Academic Press, Inc.): PCR 2: A PRACTICAL APPROACH (M.J. MacPherson, B.D. Hames and G.R. Taylor eds. (1995)), Harlow and Lane, eds. (1988) ANTIBODIES, A LABORATORY MANUAL, and ANIMAL CELL CULTURE (R.I. Freshney, ed. (1987)).

[0036] The present invention relates to methods and compositions relating to the use of stem-loop oligonucleotides as adaptors and/or primers. In various embodiments, the stem-loop oligonucleotides are chimeras of RNA and DNA. In some embodiments, the stem-loop oligonucleotide is activated through the cleavage of an RNA/DNA heteroduplex stem (e.g. by degradation of RNA by RNase H). In some embodiments, the stem-loop oligonucleotide comprises an RNA sequence in at least a portion of its loop and/or in its stem. Methods of replication and amplification using the stem-loop oligonucleotides described herein can also generate RNA/DNA heteroduplexes. Further, methods and compositions in various embodiments of the invention relate to the use of a stem-loop oligonucleotide as a dual purpose agent acting as an adaptor and primer at once.

[0037] In some embodiments, the present invention allows for the amplification of molecules having at least one double stranded region by using adaptors that reduce the propensity to form adaptor dimers. In certain aspects, the present invention provides an inert oligonucleotide (e.g. a stem-loop oligonucleotide) for attachment to a double stranded molecule such that it renders the oligonucleotide-ligated molecule capable of being modified, such as amplified, for example by strand displacement, single primer isothermal amplification (SPIA) or polymerase chain reaction. For example, upon cleavage of the inert adaptor, the oligonucleotide becomes active and suitable for providing at least in part one or more sequences employable for adaptor dependent nucleic acid extension or amplification. For another example, during SPIA amplification or polymerase chain reaction, non-cleaved adaptor/primer (e.g. stem-loop composite oligonucleotide) cannot prime effectively due to duplex formation along its stem. Its use as a primer (before it is cleaved) can be mildly or strictly limited due to the internal stem formation competing with target hybridization.

[0038] In addition to the advantages provided by the adaptors/primers of the invention, the invention further provides novel conditions for modification of nucleic acid molecules with the composite stem-loop adaptors, and subsequent amplification with said activated oligonucleotides. In most cases in the art, the process of nucleic acid analysis involves multiple enzymatic reactions that are performed in a sequential manner, frequently with

IDT Exhibit 1020.3257

intermediate purification steps between at least some of the reactions. For example, preparation of a library from genomic or viral DNA may involve fragmentation, purification, "polishing" (end-repair) of fragments, a second purification or heat inactivation, adaptor ligation and subsequent extension reactions in separate steps that often need a considerable amount of time, equipment use and human intervention, creating major obstacles for high throughput and diagnostic-type applications. The process becomes more complicated when library preparation involves additional enzymatic steps, such as DNA or library digestion with methylation-sensitive or methylation-specific endonucleases (for example, in the preparation of Methylome libraries or restriction nuclease cleavage within the adaptor sequence to produce sticky ends for DNA cloning.

[0039] In specific aspects, the present invention introduces a concept of multiplexing two or more enzymatic processes in one reaction, and teaches how to optimize a highly multiplexed enzymatic process. In particular embodiments, the present invention is directed to compositions and methods for simultaneous processing of DNA molecules with a combination of enzymes in a one-step-one-tube reaction and producing either a collection of molecules suitable for further processing, for example for amplification, or in some cases, the methods and compositions of the invention result in amplified DNA molecules. In addition, the methods and compositions of the present invention provides composite adaptor/primer oligonucleotides and methods of using the same that eliminate the possible formation of stable adaptor dimers, thus eliminating potential difficulties regularly introduced by other commonly employed methods and reagents. In particular embodiments, the methods of the invention can be easily applied to any type of fragmented double stranded DNA including but not limited to, for example, free DNA isolated from plasma, serum, and/or urine; apoptotic DNA from cells and/or tissues; DNA fragmented enzymatically in vitro (for example, by DNase I and/or restriction endonuclease); and/or DNA fragmented by mechanical forces (hydro-shear, sonication, nebulization, etc.). Additional suitable methods and compositions of producing nucleic acid molecules comprising stem-loop oligonucleotides are further described in detail in US Patent No. 7,803,550, which is herein incorporated by reference in its entirety.

[0040] In other embodiments, the invention can be easily applied to any high molecular weight double stranded DNA including, for example, DNA isolated from tissues, cell culture, bodily fluids, animal tissue, plant, bacteria, fungi, viruses, etc.

IDT Exhibit 1020.3258

[0041] Disclosed are compositions and methods for the generation of libraries using chimeric RNA-DNA stem loop oligonucleotides from any nucleic acid molecules. In various embodiments, the chimeric stem-loop oligonucleotides comprise heteroduplex stems. The stem-loop oligonucleotides described herein can be used as adaptors, primers or both. In various embodiments, the nucleic acid molecules are double stranded DNA molecules. In some embodiments, the unique structure of the stem loop chimeric oligonucleotides provide for the generation of partial duplex DNA from fragments of the sample dsDNA comprising a dsDNA fragment insert and 3'-ssDNA at one or both ends, which comprise a sequence complementary to a sequence of the stem-loop composite chimeric oligonucleotide.

[0042] The partial duplex product thus generated can be a substrate for amplification, for example, PCR or isothermal amplification. In some embodiments, the amplification reaction employs the same chimeric RNA-DNA stem-loop oligonucleotide or a different stem-loop oligonucleotide as a primer. For example, an isothermal amplification reaction comprising a chimeric RNA-DNA stem-loop oligonucleotide or a portion thereof, DNA polymerase with strand displacement, and RNase H can be performed for nucleic acid amplification using as substrate said partial duplex product.

[0043] In some embodiments, the disclosed stem-loop chimeric adaptor-pro-primer, is used as a primer for amplification, for example isothermal amplification of nucleic acids. Such pro-primer molecules can be activated to a primer in the reaction mixture. Primer activation can be carried out by an RNase targeting RNA in a RNA/DNA duplex, such as, RNase H. The RNase can be included as a component of the amplification reaction mixture. The concentration of the pro-primer, the amount of RNase activity and other reaction conditions can be adjusted to achieve activation of the pro-primer at a desired rate. For example, the pro-primer can be activated in the same reaction mixture, at a rate that is comparable to the overall duration of the amplification reaction allowing for gradual activation of stem-loop oligonucleotides over time. The pro-primer chimeric RNA-DNA stem-loop oligonucleotide can be useful for the isothermal amplification, for example SPIA, of nucleic acids employing any other method for the generation of substrates for SPIA amplification such as described for RNA amplification or whole genome amplification, as described in US. Patent No.s 6,251,639 and 6,946,251, which are both herein incorporated by reference in their entirety. Primer activation from pro-primer stem-loop chimeric oligonucleotide serves as a "slow release" of an active primer in the reaction. In general, formation of side products, for example due to high primer concentration in the initial period of an amplification reaction, can be mitigated leading

IDT Exhibit 1020.3259

to increase in desired amplification products (specificity) and higher yield, for example, formation of stem-loop oligonucleotide dimers, which are formed from the attachment of two stem-loop oligonucleotides, can be targeted by RNase activity cleaving the RNA in the RNA/DNA duplex of the stem, thus releasing remaining portions of the two oligonucleotides from each other.

Stem-loop chimeric RNA-DNA adaptor/pro-primer:

[0044] In various embodiments, a chimeric/composite stem-loop oligonucleotide comprises any oligonucleotide having a sequence that can be joined to a target polynucleotide. In many embodiments, at least a portion of the oligonucleotide sequence is known. Stem-loop oligonucleotides can comprise DNA, RNA, nucleotide analogues, non-canonical nucleotides, labeled nucleotides, modified nucleotides, or combinations thereof. Stem-loop oligonucleotides can comprise one or more of single-stranded and double-stranded regions, giving rise to stems, loops and overhangs. For example, a stem-loop oligonucleotide can have one or more loops, stems and optionally one or more overhangs. In a single strand stem-loop oligonucleotide, multiple overhangs can be achieved by lack of complementarity in the 5' and 3' distal regions. Any one of stems, loops, and overhangs can comprise DNA or RNA sequences. In general, a partial-duplex oligonucleotide comprises one or more single-stranded regions and one or more double-stranded regions. Double-stranded compositions can comprise two separate oligonucleotides hybridized to one another (also referred to as an "oligonucleotide duplex"), and hybridization may leave one or more blunt ends, one or more 3' overhangs, one or more 5' overhangs, one or more bulges resulting from mismatched and/or unpaired nucleotides, or any combination of these. In some embodiments, a single-stranded adaptor comprises two or more sequences that are able to hybridize with one another. When two such hybridizable sequences are contained in a single-stranded oligonucleotide, hybridization yields a hairpin structure or a stem-loop structure. When two hybridized regions of an oligonucleotide are separated from one another by a non-hybridized region, a "bubble" structure results. Oligonucleotide comprising a bubble structure can consist of a single strand comprising internal hybridizations, or may comprise two or more strands hybridized to one another. Internal sequence hybridization, such as between two hybridizable sequences in a strand, can produce a double-stranded structure in a single-stranded adaptor oligonucleotide. Oligonucleotides of different kinds can be used in combination, such as a hairpin oligonucleotide, stem-

loop oligonucleotide and a double-stranded oligonucleotide, or oligonucleotides of different sequences. Hybridizable sequences of the adaptors and primers described herein may or may not include one or both ends of the oligonucleotide. When neither of the ends are included in the hybridizable sequences, both ends are "free" or "overhanging." When only one end is hybridizable to another sequence in the adaptor, the other end forms an overhang, such as a 3' overhang or a 5' overhang. When both the 5'-terminal nucleotide and the 3'-terminal nucleotide are included in the hybridizable sequences, such that the 5'-terminal nucleotide and the 3'-terminal nucleotide are complementary and hybridize with one another, the end is referred to as "blunt." Different adaptors can be joined to target polynucleotides in sequential reactions or simultaneously. For example, the first and second adaptors can be added to the same reaction. Adaptors can be manipulated prior to combining with target polynucleotides. For example, terminal phosphates can be added or removed.

[0045] Stem-loop oligonucleotides, according to the various embodiments of the invention vary in size. Stem-loop oligonucleotides are typically at least 10 nucleotides long and, more typically range from 14 to 60 nucleotides, or even more typically from 18, 21, 24, 27 to 40, 45 or 50 nucleotides, in length. Other stem-loop oligonucleotides can be somewhat longer, e.g., 50 to 80 nucleotides long or longer. Stem-loop oligonucleotides can have any of the length ranges capped by any of the limits above, e.g. 18 to 60 nucleotides.

[0046] In some embodiments, one of the hybridizable sequences in a single-stranded stem-loop structure comprises RNA. For example, a stem-loop oligonucleotide can comprise a 5' end comprising sequence A' and a 3' end comprising sequence A, where A is hybridizable to A', one of A or A' comprises DNA, and the other of A or A' comprises RNA. Examples of such stem-loop oligonucleotides are illustrated in Figure 1. Similarly, a stem-loop oligonucleotide can comprise a 5' end comprising sequence B and a 3' end comprising sequence B', where B is hybridizable to B', one of B or B' comprises DNA, and the other of B or B' comprises RNA. In some embodiments, one of A or A' consists entirely of DNA, and/or one of A or A' consists entirely of RNA. In some embodiment, one of B or B' consists entirely of DNA, and/or one of B or B' consists entirely of RNA. Sequence A can be the same as or different from sequence B and/or B'. Sequence A' can be the same as or different from sequence B and/or B'. In some embodiments, the stem comprising RNA (e.g. A, A', B, or B') further comprises one or more terminal DNA residues (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, or more terminal DNA residues), such that the sequence comprising RNA is flanked by DNA residues at both

IDT Exhibit 1020.3261

ends (i.e. both the 5' end and the 3' end of the sequence comprising RNA). Hybridization of a sequence comprising RNA to a sequence comprising DNA creates an RNA-DNA heteroduplex. In some embodiments, RNA is cleaved by an enzyme that cleaves RNA from an RNA-DNA heteroduplex, such as enzymes comprising ribonuclease activity. Preferably, the enzyme comprising ribonuclease activity cleaves ribonucleotides in an RNA/DNA heteroduplex regardless of the identity and type of nucleotides adjacent to the ribonucleotide to be cleaved. It is preferred that the ribonuclease cleaves independent of sequence identity. Examples of suitable enzymes comprising ribonuclease activity for the methods and compositions of the invention are well known in the art, including ribonuclease H (RNase H) and enzymes comprising RNase H activity, e.g., Hybridase. In some embodiments, cleavage of RNA from an RNA-DNA heteroduplex removes all double-stranded character from a single-stranded stem-loop oligonucleotide, such that extension by a polymerase that uses the oligonucleotide as template requires no strand displacement step or strand displacement activity. In some embodiments, one or both ends of a stem-loop oligonucleotide comprising RNA in its stem are joined to a target polynucleotide, such that cleavage of the RNA from the RNA-DNA hetero duplex produces a 5' overhang or a 3' overhang. In some embodiments, a stem-loop oligonucleotide comprises a tag/barcode sequence. Reactions of the stem-loop oligonucleotide with a target nucleic acid, such as ligation, extension, priming etc., can impart the target oligonucleotide tagged with the barcode sequence. Blunt or sticky ended stem-loop oligonucleotides can be joined with blunt or sticky ended target polynucleotides. In some embodiments, an end comprising a 5' overhang produced by cleavage of RNA from an RNA-DNA heteroduplex is filled in by the extension of the produced 3' end using the 5' overhang as template.

[0047] Various exemplary configurations of the stem-loop chimeric RNA-DNA adaptor/pro-primer are shown in Figure 1. In some embodiments, the stem-loop chimeric RNA-DNA oligonucleotides are composed of sequences that are complementary at their distal ends (A and A') and form a DNA-RNA heteroduplex stem. In various embodiments, the stem-loop structures comprise a single stranded loop region (B). All or part of the loop region can comprise RNA (loop-RNA). The complement of the DNA sequence at the 3'-region as well as the RNA loop sequence can be removed exposing a priming site for nucleic acid amplification. In some embodiments, an isothermal amplification comprising an RNA-DNA chimeric primer utilizes said priming site. In some embodiments the RNA/DNA chimeric primer is a stem-loop chimeric pro-primer. In

IDT Exhibit 1020.3262

some embodiments the stem-loop chimeric adaptor and the stem-loop chimeric pro-primer are the same.

[0048] In some embodiments, the 3' or 5'-region of the chimeric RNA-DNA stem loop oligonucleotide of the current invention comprises a 3' or 5'-single stranded overhang, comprising DNA or RNA or combination of RNA and DNA. The 3' or 5'-single stranded overhangs may be utilized for hybridization to sticky-ended template fragments, such as those generated by fragmentation of the template dsDNA by restriction enzymes, for example, or when fragments are tailed by a single nucleotide, such as for example in the case of AT cloning procedures as described in detail in US Patent Pub. US 20100167954 and Marchuk et al. (Construction of T-vectors, a rapid and general system for direct cloning of unmodified PCR products. Marchuk et al., *Nucleic Acids Res.* 1991 March 11; 19(5):1154), which are both incorporated herein in their entirety. In these cases, the overhang sequence can be designed to be sufficiently complementary to a 3' or 5'-overhang on a template fragment allowing hybridization under the reaction conditions.

[0049] The sequence in the 3' or 5'-region comprising the stem of the stem-loop chimeric oligonucleotide of the invention may be all RNA or may comprise RNA and DNA nucleotides. The length of the RNA sequence in the stem region can be designed to optimize the rate of cleavage of the RNA portion by an RNase H that cleaves the RNA sequences in a DNA-RNA heteroduplex. In some embodiments the chimeric stem-loop oligonucleotide adaptor and pro-primer are the same. In some embodiments, the stem sequence is designed for optimal performance of the chimeric primer generated by cleavage of the RNA portion of the stem heteroduplex by RNase H.

[0050] In various embodiments of the invention, a plurality of stem-loop oligonucleotides are used as adaptors and/or primers. For example, a first stem-loop oligonucleotide and a second stem-loop oligonucleotide can be used to attach to two different ends of a target nucleic acid. For another example, a first stem-loop oligonucleotide and a second stem-loop oligonucleotide can be used to prime from two different ends of a target nucleic acid. For yet another example, a first stem-loop oligonucleotide can be used as an adaptor and a second stem-loop oligonucleotide can be used as a primer. A first and a second stem-loop oligonucleotide can be 25%-100% similar to each other, including at least about 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, and 100% sequence similarity. A first and a second stem-loop oligonucleotide can be 25%-100% complementarity, including at least about 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%,

IDT Exhibit 1020.3263

65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, and 100% sequence complementarity.

[0051] In some embodiments, a second stem-loop oligonucleotide interacts with a product generated using a first stem-loop oligonucleotide. For example, a first stem-loop oligonucleotide can be attached to a target nucleic acid via the oligonucleotide's 3' end and the target nucleic acid can be extended complementary to the attached oligonucleotide, as described below (Figure 2a). A portion of the attached oligonucleotide can be cleaved generating a priming ready construct. A second stem-loop oligonucleotide that is hybridizable to the priming ready construct, such as one that has a sequence similarity with the first stem-loop oligonucleotde, can further prime an amplification reaction. For another example, a first stem-loop oligonucleotide can be attached to a target nucleic acid via the oligonucleotide's 5' end. Upon cleavage of a heteroduplex stem region, a priming ready construct can be generated comprising a portion of the stem-loop oligonucleotide in a 3' overhang. A second stem-loop oligonucleotide, that is hybridizable to the priming ready construct, such as one that has a sequence complementarity with the first stem-loop oligonucleotde, can further prime an amplification reaction.

[0052] The sequence of the stem-loop chimeric RNA-DNA adaptor/pro-primer may further comprise one or more unique identifier sequences, or molecular barcodes, which will be useful for multiplex amplification and analysis of pooled samples.

[0053] Further features and uses of adaptor oligonucleotides are described in US Patent Publications US 2011-0319290 and US 2011-0105364, which are both herein incorporated by reference in their entirety.

Library generation using stem-loop chimeric RNA-DNA adaptor/pro-primer

[0054] In various embodiments, library generation of target nucleic acids is carried out by ligation of stem-loop chimeric RNA-DNA adaptor/pro-primer to the ends of the said target nucleic acids. The target nucleic acids can be dsDNA or fragments thereof. Cleavage of the RNA portion of the heteroduplex DNA-RNA stem can be achieved using an RNase targeting said RNA, such as RNase H. Replication of the released stem and or loop sequences that is attached to a first strand of the target nucleic acid can be achieved, for example, by extending a complementary strand along the released sequences. The released sequences can be directly adjacent to the 3'-DNA sequence of the stem or include the same. Synthesis of a DNA sequence along the stem-loop oligonucleotide sequences, for example by DNA polymerase, can yield a DNA-RNA heteroduplex comprising one or more RNA sequences from the stem-loop oligonucleotide and the synthesized DNA that is complementary. Cleavage of the RNA sequence(s) in the RNA-DNA heteroduplex, for example with RNase H, can result in a 3' overhang that is

IDT Exhibit 1020.3264

complementary to at least a portion of the stem-loop oligonucleotide. In some embodiments, the RNA sequence(s) forming a duplex with the synthesized DNA originate from the loop of the stem-loop oligonucleotide. In some embodiments the DNA polymerase is a DNA dependent DNA polymerase with RNA dependent DNA polymerase activity. In some embodiment the DNA dependent DNA polymerase and the RNA dependent DNA polymerase are two different enzymes.

[0055] Figures 2a and 2b illustrate a series of reaction steps for the generation of nucleic acid libraries, according to some embodiments of the invention. The two figures differ in the first step of attachment of the stem-loop chimeric RNA-DNA adaptor/ pro-primer to the fragments of the dsDNA target, yielding two different kinds of stem-loop-target complexes. Adaptor attachment through ligation may be carried out by ligation to one of the strands of the dsDNA fragment (Figure 2a), resulting in a stem-loop-target complex with a nick or gap or to both strand of the dsDNA fragment (Figure 2b), resulting in a stem-loop-target complex with strand continuity, without a nick or a gap between the stem-loop oligonucleotide and the target. The nature of the ligation may depend on the nature of the ends of the dsDNA and the stem-loop adaptor as well as the choice of the ligase activity. In many cases, ligation depends on the phosphorylation on the 5' end. In these cases, ligation to both strands of the dsDNA can be achieved using a phosphorylated 5'-end on both the target nucleic acid and the stem-loop chimeric RNA-DNA adaptor/pro-primer. When only one of the components comprises a 5'-phosphate, only one of the strands of the target nucleic acids can be ligated to the adaptor/pro-primer. In some embodiments, a strand of the target nucleic acid is ligated to the double stranded portion of the stem-loop. When one end, e.g. the 5'-end, of the chimeric stem-loop RNA-DNA adaptor/pro-primer comprises RNA, employment of a ligase capable of ligating a DNA sequence to an RNA sequence can be beneficial. The ligase can further comprise DNA-DNA ligation activity allowing for a DNA end of the stem-loop oligonucleotide to ligate to a DNA target nucleic acid.

[0056] After ligation to either a single strand of the target nucleic acid, for example a double stranded target nucleic acid (Figure 2a) or to both strands (Figure 2b), a complementary sequence can be synthesized, for example using 3'-extension along the stem-loop oligonucleotide sequence. Cleaving the stem of the adaptor using an RNase, for example RNase H, can linearize the stem loop structure and release it from one of the strand of the target nucleic acid at the same time. The stem loop sequences can be replicated by a DNA polymerase through extension of a free 3'-end of the target nucleic acid, forming a target nucleic acid extension product. Target nucleic acid-adaptor constructs with a single strand attachment can be replicated without the activity of

IDT Exhibit 1020.3265

an RNase. Extension of the fragment strand along the ligated chimeric DNA-RNA oligonucleotide can be carried out with a DNA polymerase comprising DNA- and RNA-dependent DNA polymerase, such as a reverse transcriptase or DNA polymerase comprising both activities, such as Bst, or a combination thereof. The target nucleic acid extension product resulting from the extension may comprise a DNA-RNA heteroduplex comprising an RNA sequence originating from the adaptor, e.g. from the loop, from any other adaptor sequence, or from a combination adaptor locations, and a DNA extension product. The RNA portion of the heteroduplex can be cleaved by an RNase, such as RNase H, in the same reaction mixture combining the product of the extension by DNA polymerase and RNase H. In various embodiments, the product of this reaction comprises a library of partial duplexes comprising dsDNA inserts of the target nucleic acids flanked by 3'-single stranded DNA extensions (or overhangs) comprising sequences complementary to the A and B sequences of the stem-loop chimeric RNA-DNA adaptor/pro-primer (Figure 2a and 2b).

[0057] The partial duplex products described above are suitable for further manipulation, for example generation of libraries comprising dsDNA fragments and 5'- and/or 3'-single stranded DNA overhangs as illustrated in Figure 3 or isothermal amplification using the chimeric RNA-DNA primers, as illustrated in Figures 4a and 4b.

Generation of libraries of dsDNA fragments with 5'- and 3'-single stranded non-complementary DNA comprising defined sequences

[0058] In various embodiments, libraries comprising dsDNA fragment inserts and known overhang sequences at one or both ends are generated from the partial duplexes described above. A partial duplex with a 3' overhang can be combined with an oligonucleotide comprising a 3'-sequence sufficiently complementary to a sequence of the single stranded 3' overhang portion of the partial duplex and a 5'-tail sequence. The complementary sequence can be directly adjacent to the double stranded region. The 5' tail sequence can be any desired sequence that is unable to efficiently hybridize to the 3' overhang based on lack of complementarity, under conditions which allow hybridization of the 3' end of the oligonucleotide to the partial duplex . The 3'-hybridized portion of the hybridized oligonucleotide can be ligated to the 5' end of the partial duplex downstream (Figure 3), thereby generating terminal Y (or forked) products comprising a dsDNA portion and non-complementary 3'- and 5'-single stranded sequences at a given end. Methods and compositions relating to terminal Y products are further described in US Patent No. 7,741,463, which is herein incorporated in its entirety.

[0059] A library comprising nucleic acids with greater dsDNA content or fully double stranded (ds) nucleic acids with desired sequences at both ends can be generated. The desired sequences can

IDT Exhibit 1020.3266

be different on each end of a strand. An oligonucleotide comprising a sequence complementary to a single stranded 3'-end of a terminal Y product like the ones described above can be hybridized to the terminal Y product. Extension of the hybridized oligonucleotide (primer) by DNA polymerase can generate dsDNA products comprising a target insert and a distinct sequence at each end, as shown in Figure 3. This process provides for the generation of libraries with asymmetric end sequences without the need to append/ligate a different sequence to each end. The described process of appending sequences is symmetric with increased efficiency for the library generation.

[0060] Further amplification of the various libraries described herein is possible, when desired, by various methods including, for example, PCR employing suitable forward and reverse primers or isothermal amplification/SPIA with one or more suitable primer(s). In some embodiments, the amplification primers may further comprise tails that are not hybridizable to the unique sequences of the library, thus extending the common sequences used in the library as may be desired for further manipulation of the libraries, such as for sequencing.

Isothermal amplification using chimeric RNA-DNA primers

[0061] Partial duplex products comprising dsDNA insert of the fragmented dsDNA template described above (Figures 2a and 2b), can be used as substrates for isothermal amplification employing chimeric RNA-DNA primers (SPIA), as shown in Figures 4a and 4b. Amplification may be carried out with stem-loop chimeric RNA-DNA pro-primer or a linear, non stem-loop chimeric RNA-DNA primer. The chimeric primer may be the stem-loop chimeric adaptor/pro-primer employed for the generation of the said partial duplex.

[0062] Amplification can be carried out in reaction mixtures comprising a partial duplex, a chimeric RNA-DNA primer, DNA polymerase with strand displacement and RNase, such as RNase H, which cleaves RNA in a RNA-DNA duplex. The chimeric RNA-DNA primer may be generated in the reaction mixture from the stem-loop chimeric RNA-DNA pro-primer through cleavage of the RNA portion of the stem RNA-DNA heteroduplex. The generation of an active primer from the pro-primer in the reaction mixture reduces the potential accumulation of undesired products that may be generated in the presence of high concentration of active primer. Active primer can be generated from the stem loop chimeric RNA-DNA pro-primer by cleavage of the RNA in the RNA-DNA heteroduplex stem portion by RNase H as shown in Figure 4a (insert).

[0063] The sequence of the chimeric RNA-DNA primer can comprise a sequence that is the same as sequence A and or B (the loop sequence of the stem-loop chimeric pro-primer; Figure 4a), or may comprise sequences hybridizable to the complement of the loop sequence (Figure 4b).

IDT Exhibit 1020.3267

[0064] Figure 5 illustrates the use of stem-loop oligonucleotides as pro-primers and primers according to some embodiments of the invention. Accordingly, a composite amplification primer can be generated in the amplification reaction mixture from a stem-loop chimeric pro-primer. The amplification reaction mixture can comprise one or more target partial duplex nucleic acid(s), for example a partial duplex DNA with a 3' overhang as illustrated in the figure, one or more chimeric stem-loop pro-primer(s), one or more DNA polymerase(s) with strand displacement activity, and one or more RNases targeting RNA in RNA/DNA heteroduplexes, for example RNase H. The RNA portion of the RNA/DNA heteroduplex at the stem of the chimeric stem-loop pro-primer can be cleaved by RNase H to generate, for example, a linear composite primer comprising a 3'-DNA and 5'-RNA. The linearized amplification primer can hybridize to a 3'-single stranded DNA portion (overhang) of a target partial duplex and can be extended by the DNA polymerase with strand displacement activity. The RNA portion of the hybridized primer in a heteroduplex can be cleaved by RNase H to free a portion of the primer binding site. A second linear composite amplification primer can hybridize to the freed primer binding site, and can be extended along the target DNA strand. The previously synthesized primer extension product (amplification product) can be displaced by the newly extended primer. Repeated cycles of primer hybridization, primer extension by strand displacement DNA polymerase, and cleavage of the RNA portion of the hybridized primer can generate multiple copies of a target nucleic acid (Figure 5).

Target polynucleotides

[0065] In various embodiments of the invention, nucleic acid are used as substrates for further manipulation. The input nucleic acid can be DNA, or complex DNA, for example genomic DNA. The input DNA may also be cDNA. The cDNA can be generated from RNA, e.g., mRNA. The input DNA can be of a specific species, for example, human, rat, mouse, other animals, plants, bacteria, algae, viruses, and the like. The input nucleic acid also can be from a mixture of genomes of different species such as host-pathogen, bacterial populations and the like. The input DNA can be cDNA made from a mixture of genomes of different species. Alternatively, the input nucleic acid can be from a synthetic source. The input DNA can be mitochondrial DNA. The input DNA can be cell-free DNA. The cell-free DNA can be obtained from, e.g., a serum or plasma sample. The input DNA can comprise one or more chromosomes. For example, if the input DNA is from a human, the DNA can comprise one or more of chromosome 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, X, or Y. The DNA can be from a linear

IDT Exhibit 1020.3268

or circular genome. The DNA can be plasmid DNA, cosmid DNA, bacterial artificial chromosome (BAC), or yeast artificial chromosome (YAC). The input DNA can be from more than one individual or organism. The input DNA can be double stranded or single stranded. The input DNA can be part of chromatin. The input DNA can be associated with histones. The methods described herein can be applied to high molecular weight DNA, such as is isolated from tissues or cell culture, for example, as well as highly degraded DNA, such as cell-free DNA from blood and urine and/or DNA extracted from formalin-fixed, paraffin-embedded tissues, for example.

[0066] The different samples from which the target polynucleotides are derived can comprise multiple samples from the same individual, samples from different individuals, or combinations thereof. In some embodiments, a sample comprises a plurality of polynucleotides from a single individual. In some embodiments, a sample comprises a plurality of polynucleotides from two or more individuals. An individual is any organism or portion thereof from which target polynucleotides can be derived, non-limiting examples of which include plants, animals, fungi, protists, monerans, viruses, mitochondria, and chloroplasts. Sample polynucleotides can be isolated from a subject, such as a cell sample, tissue sample, or organ sample derived therefrom, including, for example, cultured cell lines, biopsy, blood sample, or fluid sample containing a cell. The subject may be an animal, including but not limited to, an animal such as a cow, a pig, a mouse, a rat, a chicken, a cat, a dog, etc., and is usually a mammal, such as a human. Samples can also be artificially derived, such as by chemical synthesis. In some embodiments, the samples comprise DNA. In some embodiments, the samples comprise genomic DNA. In some embodiments, the samples comprise mitochondrial DNA, chloroplast DNA, plasmid DNA, bacterial artificial chromosomes, yeast artificial chromosomes, oligonucleotide tags, or combinations thereof. In some embodiments, the samples comprise DNA generated by primer extension reactions using any suitable combination of primers and a DNA polymerase, including but not limited to polymerase chain reaction (PCR), reverse transcription, and combinations thereof. Where the template for the primer extension reaction is RNA, the product of reverse transcription is referred to as complementary DNA (cDNA). Primers useful in primer extension reactions can comprise sequences specific to one or more targets, random sequences, partially random sequences, and combinations thereof. Reaction conditions suitable for primer extension reactions are known in the art. In general, sample polynucleotides

comprise any polynucleotide present in a sample, which may or may not include target polynucleotides.

[0067] Methods for the extraction and purification of nucleic acids are well known in the art. For example, nucleic acids can be purified by organic extraction with phenol, phenol/chloroform/isoamyl alcohol, or similar formulations, including TRIzol and TriReagent. Other non-limiting examples of extraction techniques include: (1) organic extraction followed by ethanol precipitation, e.g., using a phenol/chloroform organic reagent (Ausubel et al., 1993), with or without the use of an automated nucleic acid extractor, e.g., the Model 341 DNA Extractor available from Applied Biosystems (Foster City, Calif.); (2) stationary phase adsorption methods (U.S. Pat. No. 5,234,809; Walsh et al., 1991); and (3) salt-induced nucleic acid precipitation methods (Miller et al., (1988), such precipitation methods being typically referred to as "salting-out" methods. Another example of nucleic acid isolation and/or purification includes the use of magnetic particles to which nucleic acids can specifically or non-specifically bind, followed by isolation of the beads using a magnet, and washing and eluting the nucleic acids from the beads (see e.g. U.S. Pat. No. 5,705,628). In some embodiments, the above isolation methods may be preceded by an enzyme digestion step to help eliminate unwanted protein from the sample, e.g., digestion with proteinase K, or other like proteases. See, e.g., U.S. Pat. No. 7,001,724. If desired, RNase inhibitors may be added to the lysis buffer. For certain cell or sample types, it may be desirable to add a protein denaturation/digestion step to the protocol. Purification methods may be directed to isolate DNA, RNA, or both. When both DNA and RNA are isolated together during or subsequent to an extraction procedure, further steps may be employed to purify one or both separately from the other. Sub-fractions of extracted nucleic acids can also be generated, for example, purification by size, sequence, or other physical or chemical characteristic. In addition to an initial nucleic isolation step, purification of nucleic acids can be performed after any step in the methods of the invention, such as to remove excess or unwanted reagents, reactants, or products.

Fragmentation methods

[0068] In some embodiments, sample polynucleotides are fragmented into a population of fragmented insert DNA molecules of one or more specific size range(s). In some embodiments, fragments are generated from at least about 1, 10, 100, 1000, 10000, 100000, 300000, 500000, or more genome-equivalents of starting DNA. Fragmentation may be accomplished by methods known in the art, including chemical, enzymatic, and

IDT Exhibit 1020.3270

mechanical fragmentation. In some embodiments, the fragments have an average length from about 10 to about 10,000 nucleotides. In some embodiments, the fragments have an average length from about 50 to about 2,000 nucleotides. In some embodiments, the fragments have an average length from about 100-2,500, 10-1,000, 10-800, 10-500, 50-500, 50-250, or 50-150 nucleotides. In some embodiments, the fragments have an average length less than 500 nucleotides, such as less than 400 nucleotides, less than 300 nucleotides, less than 200 nucleotides, or less than 150 nucleotides. In some embodiments, the fragmentation is accomplished mechanically comprising subjecting sample polynucleotides to acoustic sonication. In some embodiments, the fragmentation comprises treating the sample polynucleotides with one or more enzymes under conditions suitable for the one or more enzymes to generate double-stranded nucleic acid breaks. Examples of enzymes useful in the generation of polynucleotide fragments include sequence specific and non-sequence specific nucleases. Non-limiting examples of nucleases include DNase I, Fragmentase, restriction endonucleases, variants thereof, and combinations thereof. For example, digestion with DNase I can induce random double-stranded breaks in DNA in the absence of Mg++ and in the presence of Mn++. In some embodiments, fragmentation comprises treating the sample polynucleotides with one or more restriction endonucleases. Fragmentation can produce fragments having 5' overhangs, 3' overhangs, blunt ends, or a combination thereof. In some embodiments, such as when fragmentation comprises the use of one or more restriction endonucleases, cleavage of sample polynucleotides leaves overhangs having a predictable sequence. In some embodiments, the method includes the step of size selecting the fragments via standard methods such as column purification or isolation from an agarose gel. Combination of fragmentation methods can be utilized, such as a combination enzymatic and chemical methods. In a particular example, an abasic site can be generated, e.g. using a glycosylase (Uracil-DNA glycosylase, Thymine-DNA glycosylase etc.), and the abasic site can be cleaved using a chemical method, such as by contacting the abasic site with dimethylethylenediamine (DMED).

[0069] In some embodiments, the 5' and/or 3' end nucleotide sequences of fragmented DNA are not modified prior to ligation with one or more adaptor oligonucleotides. For example, fragmentation by a restriction endonuclease can be used to leave a predictable overhang, followed by ligation with one or more adaptor oligonucleotides comprising an overhang complementary to the predictable overhang on a DNA fragment. In another example, cleavage by an enzyme that leaves a predictable blunt end can be followed by ligation of

blunt-ended DNA fragments to adaptor oligonucleotides comprising a blunt end. In some embodiments, the fragmented DNA molecules are blunt-end polished (or "end repaired") to produce DNA fragments having blunt ends, prior to being joined to adaptors. The blunt-end polishing step may be accomplished by incubation with a suitable enzyme, such as a DNA polymerase that has both 3' to 5' exonuclease activity and 5' to 3' polymerase activity, for example T4 polymerase. In some embodiments, end repair is followed by an addition of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 or more nucleotides, such as one or more adenine, one or more thymine, one or more guanine, or one or more cytosine, to produce an overhang. DNA fragments having an overhang can be joined to one or more adaptor oligonucleotides having a complementary overhang, such as in a ligation reaction. For example, a single adenine can be added to the 3' ends of end repaired DNA fragments using a template independent polymerase, followed by ligation to one or more adaptors each having a thymine at a 3' end. In some embodiments, adaptor oligonucleotides can be joined to blunt end double-stranded DNA fragment molecules which have been modified by extension of the 3' end with one or more nucleotides followed by 5' phosphorylation. In some cases, extension of the 3' end may be performed with a polymerase such as for example Klenow polymerase or any of the suitable polymerases provided herein, or by use of a terminal deoxynucleotide transferase, in the presence of one or more dNTPs in a suitable buffer containing magnesium. In some embodiments, target polynucleotides having blunt ends are joined to one or more adaptors comprising a blunt end. Phosphorylation of 5' ends of DNA fragment molecules may be performed for example with T4 polynucleotide kinase in a suitable buffer containing ATP and magnesium. The fragmented DNA molecules may optionally be treated to dephosphorylate 5' ends or 3' ends, for example, by using enzymes known in the art, such as phosphatases.

[0070] In some embodiments, each of the plurality of independent samples comprises at least about 1pg, 10pg, 100pg, 1ng, 10ng, 20ng, 30ng, 40ng, 50ng, 75ng, 100ng, 150ng, 200ng, 250ng, 300ng, 400ng, 500ng, 1μg, 1.5μg, 2μg, or more of nucleic acid material. In some embodiments, each of the plurality of independent samples comprises less than about 1pg, 10pg, 100pg, 1ng, 10ng, 20ng, 30ng, 40ng, 50ng, 75ng, 100ng, 150ng, 200ng, 250ng, 300ng, 400ng, 500ng, 1μg, 1.5μg, 2μg, or more of nucleic acid.

[0071] In some embodiments each of the individual or plurality of samples comprises a single polynucleotide target or a single genome.

IDT Exhibit 1020.3272

[0072] In another aspect, the invention provides compositions that can be used in the above described methods. Compositions of the invention can comprise any one or more of the elements described herein. In one embodiment, the composition comprises a plurality of target polynucleotides, each target polynucleotide comprising one or more barcode sequences selected from a plurality of barcode sequences, wherein said target polynucleotides are from two or more different samples, and further wherein the sample from which each of said polynucleotides is derived can be identified in a combined sequencing reaction with an accuracy of at least 95% based on a single barcode contained in the sequence of said target polynucleotide. In some embodiments, the composition comprises a plurality of first adaptor/primer oligonucleotides, wherein each of said first adaptor/primer oligonucleotides comprises at least one of a plurality of barcode sequences, wherein each barcode sequence of the plurality of barcode sequences differs from every other barcode sequence in said plurality of barcode sequences at at least three nucleotide positions.

Features of stem-loop oligonucleotides

[0073] Adaptors/primers can contain one or more of a variety of sequence elements, including but not limited to, one or more amplification primer annealing sequences or complements thereof, one or more sequencing primer annealing sequences or complements thereof, one or more transcriptional promoter sequences (e.g. T7, T3, SP6, etc.) or complements thereof, one or more barcode sequences, one or more common sequences shared among multiple different adaptors or subsets of different adaptors (adaptor group sequence), one or more restriction enzyme recognition sites, one or more overhangs complementary to one or more target polynucleotide overhangs, one or more probe binding sites (e.g. for attachment to a sequencing platform, such as a flow cell for massive parallel sequencing, such as developed by Illumina, Inc.), one or more random or near-random sequences (e.g. one or more nucleotides selected at random from a set of two or more different nucleotides at one or more positions, with each of the different nucleotides selected at one or more positions represented in a pool of adaptors comprising the random sequence), and combinations thereof. Two or more sequence elements can be non-adjacent to one another (e.g. separated by one or more nucleotides), adjacent to one another, partially overlapping, or completely overlapping. For example, an amplification primer annealing sequence can also serve as a sequencing primer annealing sequence. Sequence elements can be located at or near the 3' end, at or near the 5' end, or in the

IDT Exhibit 1020.3273

interior of the adaptor oligonucleotide. In stem-loop oligonucleotides comprising stem, loop and/or overhang regions , sequence elements can be located partially or completely outside the secondary structure, partially or completely inside the secondary structure, or in between sequences participating in the secondary structure. For example, , sequence elements can be located partially or completely inside or outside the hybridizable sequences (the "stem"), including in the sequence between the hybridizable sequences (the "loop"). In some embodiments, the first adaptor/primer oligonucleotides in a plurality of first adaptor/primer oligonucleotides having different barcode sequences comprise a sequence element common among all first adaptor oligonucleotides in the plurality. In some embodiments, all second adaptor/primer oligonucleotides comprise a sequence element common among all second adaptor/primer oligonucleotides that is different from the common sequence element shared by the first adaptor/primer oligonucleotides. A difference in sequence elements can be any such that at least a portion of different adaptors /primers do not completely align, for example, due to changes in sequence length, deletion or insertion of one or more nucleotides, or a change in the nucleotide composition at one or more nucleotide positions (such as a base change or base modification). In some embodiments, an adaptor/primer oligonucleotide comprises a 5' overhang, a 3' overhang, or both that is complementary to one or more target polynucleotides. Complementary overhangs can be one or more nucleotides in length, including but not limited to 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, or more nucleotides in length. Complementary overhangs may comprise a fixed sequence. Complementary overhangs may comprise a random sequence of one or more nucleotides, such that one or more nucleotides are selected at random from a set of two or more different nucleotides at one or more positions, with each of the different nucleotides selected at one or more positions represented in a pool of adaptors with complementary overhangs comprising the random sequence. In some embodiments, an adaptor/primer overhang is complementary to a target polynucleotide overhang produced by restriction endonuclease digestion. In some embodiments, an adaptor overhang consists of an adenine or a thymine.

[0074] Adaptor/primer oligonucleotides can have any suitable length, at least sufficient to accommodate the one or more sequence elements of which they are comprised. In some embodiments, adaptor/primers are about, less than about, or more than about 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 90, 100, 200, or more nucleotides in length. In some embodiments, the stem of a stem-loop adaptor/primer is about, less than about,

IDT Exhibit 1020.3274

or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50, 75, 100, or more nucleotides in length. Stems may be designed using a variety of different sequences that result in hybridization between the complementary regions on a stem-loop chimeric adaptor, resulting in a local region of double-stranded DNA. For example, stem sequences may be utilized that are from 15 to 18 nucleotides in length with equal representation of G:C and A:T base pairs. Such stem sequences are predicted to form stable dsDNA structures below their predicted melting temperatures of ~45°C. Sequences participating in the stem of the stem-loop adaptor can be perfectly complementary, such that each base of one region in the stem hybridizes via hydrogen bonding with each base in the other region in the stem according to Watson-Crick base-pairing rules. Alternatively, sequences in the stem may deviate from perfect complementarity. For example, there can be mismatches and or bulges within the stem structure created by opposing bases that do not follow Watson-Crick base pairing rules, and/or one or more nucleotides in one region of the stem that do not have the one or more corresponding base positions in the other region participating in the stem. Mismatched sequences may be cleaved using enzymes that recognize mismatches. The stem and/or loop of a stem-loop can comprise DNA, RNA, or both DNA and RNA. In some embodiments, the stem and/or loop of a stem-loop, or one or both of the hybridizable sequences forming the stem of a stem-loop, comprise nucleotides, bonds, or sequences that are substrates for cleavage, such as by an enzyme, including but not limited to endonucleases and glycosylases. The composition of a stem may be such that only one of the hybridizable sequences forming the stem is cleaved. For example, one of the sequences forming the stem may comprise RNA while the other sequence forming the stem consists of DNA, such that cleavage by an enzyme that cleaves RNA in an RNA-DNA duplex, such as RNase H, cleaves only the sequence comprising RNA. The stem and/or loop of a stem-loop oligonucleotide can comprise non-canonical nucleotides (e.g. uracil), and/or methylated nucleotides. In some embodiments, the loop sequence of a hairpin adaptor is about, less than about, or more than about 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, or more nucleotides in length.

Barcodes

[0075] As used herein, the term "barcode" refers to a known nucleic acid sequence that allows some feature of a polynucleotide with which the barcode is associated to be identified. In some embodiments, the feature of the polynucleotide to be identified is the sample from

IDT Exhibit 1020.3275

which the polynucleotide is derived. In some embodiments, barcodes are at least 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, or more nucleotides in length. In some embodiments, barcodes are shorter than 10, 9, 8, 7, 6, 5, or 4 nucleotides in length. In some embodiments, barcodes associated with some polynucleotides are of different length than barcodes associated with other polynucleotides. In general, barcodes are of sufficient length and comprise sequences that are sufficiently different to allow the identification of samples based on barcodes with which they are associated. In some embodiments, a barcode, and the sample source with which it is associated, can be identified accurately after the mutation, insertion, or deletion of one or more nucleotides in the barcode sequence, such as the mutation, insertion, or deletion of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more nucleotides. In some embodiments, each barcode in a plurality of barcodes differ from every other barcode in the plurality at at least three nucleotide positions, such as at least 3, 4, 5, 6, 7, 8, 9, 10, or more positions. In some embodiments, both a first adaptor/primer and a second adaptor/primer comprise at least one of a plurality of barcode sequences. In some embodiments, barcodes for second adaptor/primer oligonucleotides are selected independently from barcodes for first adaptor/primer oligonucleotides. In some embodiments, first adaptor/primer oligonucleotides and second adaptor/primer oligonucleotides having barcodes are paired, such that adaptor/primers of the pair comprise the same or different one or more barcodes. In some embodiments, the methods of the invention further comprise identifying the sample from which a target polynucleotide is derived based on a barcode sequence to which the target polynucleotide is joined. In general, a barcode comprises a nucleic acid sequence that when joined to a target polynucleotide serves as an identifier of the sample from which the target polynucleotide was derived.

[0076] In some embodiments, the plurality of barcode sequences from which barcode sequences are selected includes sequences selected from the group consisting of: AAA, TTT, CCC, GGG. In some embodiments, the plurality of barcode sequences from which barcode sequences are selected includes sequences selected from the group consisting of: AAAA, CTGC, GCTG, TGCT, ACCC, CGTA, GAGT, TTAG, AGGG, CCAT, GTCA, TATC, ATTT, CACG, GGAC, and TCGA. In some embodiments, the plurality of barcode sequences from which barcode sequences are selected includes sequences selected from the group consisting of: AAAAA, AACCC, AAGGG, AATTT, ACACG, ACCAT, ACGTA, ACTGC, AGAGT, AGCTG, AGGAC, AGTCA, ATATC, ATCGA, ATGCT, ATTAG, CAACT, CACAG, CAGTC, CATGA, CCAAC, CCCCA, CCGGT, CCTTG,

CGATA, CGCGC, CGGCG, CGTAT, CTAGG, CTCTT, CTGAA, CTTCC, GAAGC, GACTA, GAGAT, GATCG, GCATT, GCCGG, GCGCC, GCTAA, GGAAG, GGCCT, GGGGA, GGTTC, GTACA, GTCAC, GTGTG, GTTTT, TAATG, TACGT, TAGCA, TATAC, TCAGA, TCCTC, TCGAG, TCTCT, TGACC, TGCAA, TGGTT, TGTGG, TTAAT, TTCCG, TTGGC, and TTTTA.

Attachment of Adaptors

Ligation

[0077] The terms "joining" and "ligation" as used herein, with respect to two polynucleotides, such as a stem-loop adaptor/primer oligonucleotide and a target polynucleotide, refers to the covalent attachment of two separate polynucleotides to produce a single larger polynucleotide with a contiguous backbone. Methods for joining two polynucleotides are known in the art, and include without limitation, enzymatic and non-enzymatic (e.g. chemical) methods. Examples of ligation reactions that are non-enzymatic include the non-enzymatic ligation techniques described in U.S. Pat. Nos. 5,780,613 and 5,476,930, which are herein incorporated by reference. In some embodiments, an adaptor oligonucleotide is joined to a target polynucleotide by a ligase, for example a DNA ligase or RNA ligase. Multiple ligases, each having characterized reaction conditions, are known in the art, and include, without limitation NAD^+-dependent ligases including tRNA ligase, Taq DNA ligase, *Thermus filiformis* DNA ligase, Escherichia coli DNA ligase, Tth DNA ligase, *Thermus scotoductus* DNA ligase (I and II), thermostable ligase, Ampligase thermostable DNA ligase, VanC-type ligase, 9° N DNA Ligase, Tsp DNA ligase, and novel ligases discovered by bioprospecting; ATP-dependent ligases including T4 RNA ligase, T4 DNA ligase, T3 DNA ligase, T7 DNA ligase, Pfu DNA ligase, DNA ligase 1, DNA ligase III, DNA ligase IV, and novel ligases discovered by bioprospecting; and wild-type, mutant isoforms, and genetically engineered variants thereof. Ligation can be between polynucleotides having hybridizable sequences, such as complementary overhangs. Ligation can also be between two blunt ends. Generally, a 5' phosphate is utilized in a ligation reaction. The 5' phosphate can be provided by the target polynucleotide, the adaptor oligonucleotide, or both. 5' phosphates can be added to or removed from polynucleotides to be joined, as needed. Methods for the addition or removal of 5' phosphates are known in the art, and include without limitation enzymatic and chemical processes. Enzymes useful in the addition and/or removal of 5' phosphates include kinases, phosphatases, and polymerases. In some embodiments, both of the two ends joined in a ligation reaction (e.g. an adaptor end and a target polynucleotide end)

IDT Exhibit 1020.3277

provide a 5' phosphate, such that two covalent linkages are made in joining the two ends. In some embodiments, only one of the two ends joined in a ligation reaction (e.g. only one of an adaptor end and a target polynucleotide end) provides a 5' phosphate, such that only one covalent linkage is made in joining the two ends. In some embodiments, only one strand at one or both ends of a target polynucleotide is joined to an adaptor oligonucleotide. In some embodiments, both strands at one or both ends of a target polynucleotide are joined to an adaptor oligonucleotide. In some embodiments, 3' phosphates are removed prior to ligation. In some embodiments, an adaptor oligonucleotide is added to both ends of a target polynucleotide, wherein one or both strands at each end are joined to one or more adaptor oligonucleotides. When both strands at both ends are joined to an adaptor oligonucleotide, joining can be followed by a cleavage reaction that leaves a 5' overhang that can serve as a template for the extension of the corresponding 3' end, which 3' end may or may not include one or more nucleotides derived from the adaptor oligonucleotide. In some embodiments, a target polynucleotide is joined to a first adaptor oligonucleotide on one end and a second adaptor oligonucleotide on the other end. In some embodiments, the target polynucleotide and the adaptor to which it is joined comprise blunt ends. In some embodiments, separate ligation reactions are carried out for each sample, using a different first adaptor oligonucleotide comprising at least one barcode sequence for each sample, such that no barcode sequence is joined to the target polynucleotides of more than one sample. A target polynucleotide that has an adaptor/primer oligonucleotide joined to it is considered "tagged" by the joined adaptor.

[0078] In some embodiments, joining of an adaptor/primer to a target polynucleotide produces a joined product polynucleotide having a 3' overhang comprising a nucleotide sequence derived from the adaptor/primer. In some embodiments, a primer oligonucleotide comprising a sequence complementary to all or a portion of the 3' overhang is hybridized to the overhang and extended using a DNA polymerase to produce a primer extension product hybridized to one strand of the joined product polynucleotide. The DNA polymerase may comprise strand displacement activity, such that one strand of the joined product polynucleotide is displaced during primer extension.

Pooling

[0079] In some embodiments, after joining at least one adaptor oligonucleotide to a target polynucleotide, the 3' end of one or more target polynucleotides is extended using the one or more joined adaptor oligonucleotides as template. For example, an adaptor

IDT Exhibit 1020.3278

comprising two hybridized oligonucleotides that is joined to only the 5' end of a target polynucleotide allows for the extension of the unjoined 3'end of the target using the joined strand of the adaptor as template, concurrently with or following displacement of the unjoined strand. If both strands of an adaptor comprising two hybridized oligonucleotides are joined to a target polynucleotide such that the joined product has a 5' overhang, the complementary 3' end can be extended using the 5' overhang as template. As a further example, a stem-loop adaptor oligonucleotide can be joined to the 5' end of a target polynucleotide. While double-stranded in secondary structure, such a stem-loop adaptor remains single-stranded, and is thus a 5' overhang appended to the target polynucleotide (e.g. when the 5' end of the stem-loop adaptor is not joined to the target polynucleotide). Removal of the secondary structure, either prior to (e.g. thermal denaturing, or degradation) or concurrently with (e.g. strand displacement) the activity of a polymerase, provides a template for the extension of the 3' end of the complementary strand of the target polynucleotide. In some embodiments, the 3' end of the target polynucleotide that is extended comprises one or more nucleotides from an adaptor oligonucleotide. For target polynucleotides to which adaptors are joined on both ends, extension can be carried out for both 3' ends of a double-stranded target polynucleotide having 5' overhangs. This 3' end extension, or "fill-in" reaction, generates a complementary sequence, or "complement," to the adaptor oligonucleotide template that is hybridized to the template, thus filling in the 5' overhang to produce a double-stranded sequence region. Where both ends of a double-stranded target polynucleotide have 5' overhangs that are filled in by extension of the complementary strands' 3' ends, the product is completely double-stranded. Extension can be carried out by any suitable polymerase known in the art, such as a DNA polymerase, many of which are commercially available. DNA polymerases can comprise DNA-dependent DNA polymerase activity, RNA-dependent DNA polymerase activity, or DNA-dependent and RNA-dependent DNA polymerase activity. DNA polymerases can be thermostable or non-thermostable. Examples of DNA polymerases include, but are not limited to, Taq polymerase, Tth polymerase, Tli polymerase, Pfu polymerase, Pfutubo polymerase, Pyrobest polymerase, Pwo polymerase, KOD polymerase, Bst polymerase, Sac polymerase, Sso polymerase, Poc polymerase, Pab polymerase, Mth polymerase, Pho polymerase, ES4 polymerase, VENT polymerase, DEEPVENT polymerase, EX-Taq polymerase, LA-Taq polymerase, Expand polymerases, Platinum Taq polymerases, Hi-Fi polymerase, Tbr polymerase, Tfl polymerase, Tru polymerase, Tac polymerase, Tne

IDT Exhibit 1020.3279

polymerase, Tma polymerase, Tih polymerase, Tfi polymerase, Klenow fragment, and variants, modified products and derivatives thereof. 3' end extension can be performed before or after pooling of target polynucleotides from independent samples.

[0080] In some embodiments, the fill-in reaction is followed by or performed as part of amplification of one or more target polynucleotides using a first primer and/or a second primer, wherein the first primer comprises a sequence that is hybridizable to at least a portion of the complement of one or more of a first adaptor/primer oligonucleotides, and further wherein the second primer comprises a sequence that is hybridizable to at least a portion of the complement of one or more of a second adaptor/primer oligonucleotides. Each of the first and second primers may be of any suitable length, such as about, less than about, or more than about 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 90, 100, or more nucleotides, any portion or all of which may be complementary to the corresponding target sequence (e.g. about, less than about, or more than about 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, or more nucleotides). "Amplification" refers to any process by which the copy number of a target sequence is increased. Methods for primer-directed amplification of target polynucleotides are known in the art, and include without limitation, methods based on the polymerase chain reaction (PCR). Conditions favorable to the amplification of target sequences by PCR are known in the art, can be optimized at a variety of steps in the process, and depend on characteristics of elements in the reaction, such as target type, target concentration, sequence length to be amplified, sequence of the target and/or one or more primers, primer length, primer concentration, polymerase used, reaction volume, ratio of one or more elements to one or more other elements, and others, some or all of which can be altered. In general, PCR involves the steps of denaturation of the target to be amplified (if double stranded), hybridization of one or more primers to the target, and extension of the primers by a DNA polymerase, with the steps repeated (or "cycled") in order to amplify the target sequence. Steps in this process can be optimized for various outcomes, such as to enhance yield, decrease the formation of spurious products, and/or increase or decrease specificity of primer annealing. Methods of optimization are well known in the art and include adjustments to the type or amount of elements in the amplification reaction and/or to the conditions of a given step in the process, such as temperature at a particular step, duration of a particular step, and/or number of cycles. In some embodiments, an amplification reaction comprises at least 5, 10, 15, 20, 25, 30, 35, 50, or more cycles. In some embodiments, an amplification reaction comprises no more than 5, 10, 15, 20, 25, 35, 50, or more cycles. Cycles can

contain any number of steps, such as 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more steps. Steps can comprise any temperature or gradient of temperatures, suitable for achieving the purpose of the given step, including but not limited to, 3' end extension (e.g. adaptor fill-in), primer annealing, primer extension, and strand denaturation. Steps can be of any duration, including but not limited to about, less than about, or more than about 1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 70, 80, 90, 100, 120, 180, 240, 300, 360, 420, 480, 540, 600, or more seconds, including indefinitely until manually interrupted. Cycles of any number comprising different steps can be combined in any order. In some embodiments, different cycles comprising different steps are combined such that the total number of cycles in the combination is about, less that about, or more than about 5, 10, 15, 20, 25, 30, 35, 50, or more cycles. In some embodiments, amplification is performed following the fill-in reaction. Amplification can be performed before or after pooling of target polynucleotides from independent samples.

Methods

Methods of Amplification

[0081] The methods, compositions and kits described herein can be useful to generate amplification-ready products directly from genomic DNA or whole or partial transcriptome RNA for downstream applications such as massively parallel sequencing (Next Generation Sequencing methods), multiplexed quantification of large sets of sequence regions of interest, such as by high density qPCR arrays and other highly parallel quantification platforms (selective massively parallel target pre-amplification), as well as generation of libraries with enriched population of sequence regions of interest. The methods described herein can be used to generate a collection of at least 25, 50, 75, 100, 500, 1000, 2500, 5000, 10,000, 25,000, 50,000, 100,000, 500,000, or 1,000,000 amplification-ready target sequence regions of interest directly from a sample of complex DNA using a plurality of oligonucleotides.

[0082] Methods of nucleic acid amplification are well known in the art. In some embodiments, the amplification method is isothermal. In other embodiments the amplification method is linear. In other embodiments the amplification is exponential.

[0083] **SPIA Amplification**

[0084] Amplification of the sequence regions of interest employing a linear amplification method such as the single primer isothermal amplification (SPIA) can be used. SPIA enables generation of multiple copies of the strand specific sequence regions of interest

IDT Exhibit 1020.3281

and employs a single amplification primer, thus reducing the complexity associated with multiple oligonucleotide design and manufacturing, enables the use of a generic amplification primer, and can be linear. The fidelity of quantification of the copy number of the sequence regions of interest in the complex genomic NA sample is a highly desirable feature of the presented methods of the invention.

[0085] Amplification by SPIA can occur under conditions permitting composite primer hybridization, primer extension by a DNA polymerase with strand displacement activity, cleavage of RNA from a RNA/DNA heteroduplex and strand displacement. In so far as the composite amplification primer hybridizes to the 3′-single-stranded portion (of the partially double stranded polynucleotide which is formed by cleaving RNA in the complex comprising a RNA/DNA partial heteroduplex) comprising, generally, the complement of at least a portion of the composite amplification primer sequence, composite primer hybridization may be under conditions permitting specific hybridization. In SPIA, all steps are isothermal (in the sense that thermal cycling is not required), although the temperatures for each of the steps may or may not be the same. It is understood that various other embodiments can be practiced given the general description provided above. For example, as described and exemplified herein, certain steps may be performed as temperature is changed (e.g., raised, or lowered).

[0086] Although generally only one composite amplification primer is described above, it is further understood that the SPIA amplification methods can be performed in the presence of two or more different first and/or second composite primers that randomly prime template polynucleotide. In addition, the amplification polynucleotide products of two or more separate amplification reactions conducted using two or more different first and/or second composite primers that randomly prime template polynucleotide can be combined.

[0087] The composite amplification primers are primers that are composed of RNA and DNA portions. In the amplification composite primer, both the RNA and the DNA portions are generally complementary and can hybridize to a sequence in the amplification-ready product to be copied or amplified. In some embodiments, a 3′-portion of the amplification composite primer is DNA and a 5′-portion of the composite amplification primer is RNA. The composite amplification primer is designed such that the primer is extended from the 3′-DNA portion to create a primer extension product. The 5′-RNA portion of this primer extension product in a RNA/DNA heteroduplex is susceptible to cleavage by RNase H, thus freeing a portion of the polynucleotide to the hybridization of an additional

IDT Exhibit 1020.3282

composite amplification primer. The extension of the amplification composite primer by a DNA polymerase with strand displacement activity releases the primer extension product from the original primer and creates another copy of the sequence of the polynucleotide. Repeated rounds of primer hybridization, primer extension with strand displacement DNA synthesis, and RNA cleavage create multiple copies of the strand-specific sequence of the polynucleotide.

[0088] In some embodiments, the composite amplification primer is generated in the amplification reaction mixture from a stem-loop chimeric pro-primer. The amplification reaction mixture can comprise a target partial duplex nucleic acid, for example a target partial duplex DNA, a chimeric stem-loop pro-primer, DNA polymerase with strand displacement activity, and an RNase targeting RNA in a RNA/DNA heteroduplex, for example RNase H. The RNA portion of the RNA/DNA heteroduplex at the stem of the chimeric stem-loop pro-primer can be cleaved by RNase H to generate, for example, a linear composite primer comprising a 3'-DNA and 5'-RNA. The linearized amplification primer can hybridize to a 3'-single stranded DNA portion (overhang) of a target partial duplex and can be extended by the DNA polymerase with strand displacement activity. The RNA portion of the hybridized primer in a heteroduplex can be cleaved by RNase H to free a portion of the primer binding site. A second linear composite amplification primer can hybridize to the freed primer binding site, and can be extended along the target DNA strand. The previously synthesized primer extension product (amplification product) can be displaced by the newly extended primer. Repeated cycles of primer hybridization, primer extension by strand displacement DNA polymerase, and cleavage of the RNA portion of the hybridized primer can generate multiple copies of a target nucleic acid (Figure 5).

[0089] **Other Amplification Methods**

[0090] Some aspects of the invention comprise the amplification of polynucleotide molecules or sequences within the polynucleotide molecules. Amplification generally refers to a method that can result in the formation of one or more copies of a nucleic acid or polynucleotide molecule or in the formation of one or more copies of the complement of a nucleic acid or polynucleotide molecule. Amplifications can be used in the invention, for example, to amplify or analyze a polynucleotide bound to a solid surface. The amplifications can be performed, for example, after archiving the samples in order to analyze the archived polynucleotide.

IDT Exhibit 1020.3283

[0091] In some aspects of the invention, exponential amplification of nucleic acids or polynucleotides is used. These methods often depend on the product catalyzed formation of multiple copies of a nucleic acid or polynucleotide molecule or its complement. The amplification products are sometimes referred to as "amplicons." One such method for the enzymatic amplification of specific double stranded sequences of DNA is polymerase chain reaction (PCR). This in vitro amplification procedure is based on repeated cycles of denaturation, oligonucleotide primer annealing, and primer extension by thermophilic template dependent polynucleotide polymerase, resulting in the exponential increase in copies of the desired sequence of the polynucleotide analyte flanked by the primers. The two different PCR primers, which anneal to opposite strands of the DNA, are positioned so that the polymerase catalyzed extension product of one primer can serve as a template strand for the other, leading to the accumulation of a discrete double stranded fragment whose length is defined by the distance between the 5′ ends of the oligonucleotide primers. Other amplification techniques that can be used in the methods of the provided invention include, e.g., AFLP (amplified fragment length polymorphism) PCR (see e.g.: Vos et al. 1995. AFLP: a new technique for DNA fingerprinting. Nucleic Acids Research 23: 4407-14), allele-specific PCR (see e.g., Saiki R K, Bugawan T L, Horn G T, Mullis K B, Erlich H A (1986). Analysis of enzymatically amplified beta-globin and HLA-DQ alpha DNA with allele-specific oligonucleotide probes Nature 324: 163-166), Alu PCR, assembly PCR (see e.g., Stemmer W P, Crameri A, Ha K D, Brennan T M, Heyneker H L (1995). Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides Gene 164: 49-53), assymetric PCR (see e.g., Saiki R K supra), colony PCR, helicase dependent PCR (see e.g., Myriam Vincent, Yan Xu and Huimin Kong (2004). Helicase-dependent isothermal DNA amplification EMBO reports 5 (8): 795-800), hot start PCR, inverse PCR (see e.g., Ochman H, Gerber A S, Hartl D L. Genetics. 1988 November; 120(3):621-3), in situ PCR, intersequence-specific PCR or IS SR PCR, digital PCR, linear-after-the-exponential-PCR or Late PCR (see e.g., Pierce K E and Wangh L T (2007). Linear-after-the-exponential polymerase chain reaction and allied technologies Real-time detection strategies for rapid, reliable diagnosis from single cells Methods Mol. Med. 132: 65-85), long PCR, nested PCR, real-time PCR, duplex PCR, multiplex PCR, quantitative PCR, or single cell PCR.

[0092] Another method for amplification involves amplification of a single stranded polynucleotide using a single oligonucleotide primer. The single stranded polynucleotide that is to be amplified contains two non-contiguous sequences that are substantially or

IDT Exhibit 1020.3284

completely complementary to one another and, thus, are capable of hybridizing together to form a stem-loop structure. This single stranded polynucleotide already may be part of a polynucleotide analyte or may be created as the result of the presence of a polynucleotide analyte.

[0093] Another method for achieving the result of an amplification of nucleic acids is known as the ligase chain reaction (LCR). This method uses a ligase enzyme to join pairs of preformed nucleic acid probes. The probes hybridize with each complementary strand of the nucleic acid analyte, if present, and ligase is employed to bind each pair of probes together resulting in two templates that can serve in the next cycle to reiterate the particular nucleic acid sequence.

[0094] Another method for achieving nucleic acid amplification is the nucleic acid sequence based amplification (NASBA). This method is a promoter-directed, enzymatic process that induces in vitro continuous, homogeneous and isothermal amplification of a specific nucleic acid to provide RNA copies of the nucleic acid. The reagents for conducting NASBA include a first DNA primer with a 5′-tail comprising a promoter, a second DNA primer, reverse transcriptase, RNase-H, T7 RNA polymerase, NTP's and dNTP's.

[0095] Another method for amplifying a specific group of nucleic acids is the Q-beta-replicase method, which relies on the ability of Q-beta-replicase to amplify its RNA substrate exponentially. The reagents for conducting such an amplification include "midi-variant RNA" (amplifiable hybridization probe), NTP's, and Q-beta-replicase.

[0096] Another method for amplifying nucleic acids is known as 3SR and is similar to NASBA except that the RNase-H activity is present in the reverse transcriptase. Amplification by 3SR is an RNA specific target method whereby RNA is amplified in an isothermal process combining promoter directed RNA polymerase, reverse transcriptase and RNase H with target RNA. See for example Fahy et al. PCR Methods Appl. 1:25-33 (1991).

[0097] Another method for amplifying nucleic acids is the Transcription Mediated Amplification (TMA) used by Gen-Probe. The method is similar to NASBA in utilizing two enzymes in a self-sustained sequence replication. See U.S. Pat. No. 5,299,491 herein incorporated by reference.

[0098] Another method for amplification of nucleic acids is Strand Displacement Amplification (SDA) (Westin et al 2000, Nature Biotechnology, 18, 199-202; Walker et al 1992, Nucleic Acids Research, 20, 7, 1691-1696), which is an isothermal amplification technique based upon the ability of a restriction endonuclease such as HincII or BsoBI to nick the unmodified strand of a hemiphosphorothioate form of its recognition site, and

IDT Exhibit 1020.3285

the ability of an exonuclease deficient DNA polymerase such as Klenow exo minus polymerase, or Bst polymerase, to extend the 3′-end at the nick and displace the downstream DNA strand. Exponential amplification results from coupling sense and antisense reactions in which strands displaced from a sense reaction serve as targets for an antisense reaction and vice versa.

[0099] Another method for amplification of nucleic acids is Rolling Circle Amplification (RCA) (Lizardi et al. 1998, Nature Genetics, 19:225-232). RCA can be used to amplify single stranded molecules in the form of circles of nucleic acids. In its simplest form, RCA involves the hybridization of a single primer to a circular nucleic acid. Extension of the primer by a DNA polymerase with strand displacement activity results in the production of multiple copies of the circular nucleic acid concatenated into a single DNA strand.

[00100] In some embodiments of the invention, RCA is coupled with ligation. For example, a single oligonucleotide can be used both for ligation and as the circular template for RCA. This type of polynucleotide can be referred to as a "padlock probe" or a "RCA probe." For a padlock probe, both termini of the oligonucleotide contain sequences complementary to a domain within a nucleic acid sequence of interest. The first end of the padlock probe is substantially complementary to a first domain on the nucleic acid sequence of interest, and the second end of the padlock probe is substantially complementary to a second domain, adjacent to the first domain near the first domain. Hybridization of the oligonucleotide to the target nucleic acid results in the formation of a hybridization complex. Ligation of the ends of the padlock probe results in the formation of a modified hybridization complex containing a circular polynucleotide. In some cases, prior to ligation, a polymerase can fill in the gap by extending one end of the padlock probe. The circular polynucleotide thus formed can serve as a template for RCA that, with the addition of a polymerase, results in the formation of an amplified product nucleic acid. The methods of the invention described herein can produce amplified products with defined sequences on both the 5′- and 3′-ends. Such amplified products can be used as padlock probes.

[00101] Some aspects of the invention utilize the linear amplification of nucleic acids or polynucleotides. Linear amplification generally refers to a method that involves the formation of one or more copies of the complement of only one strand of a nucleic acid or polynucleotide molecule, usually a nucleic acid or polynucleotide analyte. Thus, the primary difference between linear amplification and exponential amplification is that in the latter process, the product serves as substrate for the formation of more product,

IDT Exhibit 1020.3286

whereas in the former process the starting sequence is the substrate for the formation of product but the product of the reaction, i.e. the replication of the starting template, is not a substrate for generation of products. In linear amplification the amount of product formed increases as a linear function of time as opposed to exponential amplification where the amount of product formed is an exponential function of time.

[00102] In some embodiments, amplification methods can be solid-phase amplification, polony amplification, colony amplification, emulsion PCR, bead RCA, surface RCA, surface SDA, etc., as will be recognized by one of skill in the art. In some embodiments, amplification methods that results in amplification of free DNA molecules in solution or tethered to a suitable matrix by only one end of the DNA molecule can be used. Methods that rely on bridge PCR, where both PCR primers are attached to a surface (see, e.g., WO 2000/018957 and Adessi et al., Nucleic Acids Research (2000): 28(20): E87) can be used. In some cases the methods of the invention can create a "polymerase colony technology," or "polony." referring to a multiplex amplification that maintains spatial clustering of identical amplicons (see Harvard Molecular Technology Group and Lipper Center for Computational Genetics website). These include, for example, in situ polonies (Mitra and Church, Nucleic Acid Research 27, e34, Dec. 15, 1999), in situ rolling circle amplification (RCA) (Lizardi et al., Nature Genetics 19, 225, July 1998), bridge PCR (U.S. Pat. No. 5,641,658), picotiter PCR (Leamon et al., Electrophoresis 24, 3769, November 2003), and emulsion PCR (Dressman et al., PNAS 100, 8817, Jul. 22, 2003). The methods of the invention provide new methods for generating and using polonies.

Other Downstream Applications

[00103] An important aspect of the invention is that the methods and compositions disclosed herein can be efficiently, cost-effectively, and with the minimal loss of biological material, utilized for downstream analyses.

[00104] In particular embodiments, the oligonucleotide-attached nucleic acid molecule is further modified, such as by cloning, including by incorporation of the modified molecule into a vector, said incorporation occurring at ends in the modified molecule generated by endonuclease cleavage within the inverted repeat.

[00105] In specific embodiments, the oligonucleotide-attached nucleic acid molecule is immobilized to a solid support, such as non-covalently or covalently.

[00106] The libraries of amplified, non-amplified, non-targeted nucleic acids, as well as enriched copies of selected sequence regions of interest are useful for massively parallel

IDT Exhibit 1020.3287

sequencing (Next Generation Sequencing methods), multiplexed quantification of large sets of sequence regions of interest, such as by high density qPCR arrays and other highly parallel quantification platforms (selective massively parallel target pre-amplification), as well as generation of libraries with enriched population of sequence regions of interest.

Sequencing

[00107] In one embodiment, the invention provides for products ready for amplification in preparation for sequencing. In some embodiments, the target polynucleotides are pooled followed by sequencing one or more polynucleotides in the pool. Sequencing methods utilizing adaptor incorporated sequences are well known in the art and are further described, for example, in US Patent No.s 8,053,192 and 8,017,335.

[00108] Sequencing processes are generally template dependent. Nucleic acid sequence analysis that employs template dependent synthesis identifies individual bases, or groups of bases as they are added during a template mediated synthesis reaction, such as a primer extension reaction, where the identity of the base is complementary to the template sequence to which the primer sequence is hybridized during synthesis. Other such processes include ligation driven processes, where oligonucleotides or polynucleotides are complexed with an underlying template sequence, in order to identify the sequence of nucleotides in that sequence. Typically, such processes are enzymatically mediated using nucleic acid polymerases, such as DNA polymerases, RNA polymerases, reverse transcriptases, and the like, or other enzymes such as in the case of ligation driven processes, e.g., ligases.

[00109] Sequence analysis using template dependent synthesis can include a number of different processes. For example, in the ubiquitously practiced four-color Sanger sequencing methods, a population of template molecules is used to create a population of complementary fragment sequences. Primer extension is carried out in the presence of the four naturally occurring nucleotides, and with a sub-population of dye labeled terminator nucleotides, e.g., dideoxyribonucleotides, where each type of terminator (ddATP, ddGTP, ddTTP, ddCTP) includes a different detectable label. As a result, a nested set of fragments is created where the fragments terminate at each nucleotide in the sequence beyond the primer, and are labeled in a manner that permits identification of the terminating nucleotide. The nested fragment population is then subjected to size based separation, e.g., using capillary electrophoresis, and the labels associated with each different sized fragment is identified to identify the terminating nucleotide. As a result, the sequence of labels moving past a detector in the separation system provides a direct

IDT Exhibit 1020.3288

readout of the sequence information of the synthesized fragments, and by complementarity, the underlying template (See, e.g., U.S. Pat. No. 5,171,534, incorporated herein by reference in its entirety for all purposes).

[00110] Other examples of template dependent sequencing methods include sequence by synthesis processes, where individual nucleotides are identified iteratively, as they are added to the growing primer extension product.

[00111] Pyrosequencing is an example of a sequence by synthesis process that identifies the incorporation of a nucleotide by assaying the resulting synthesis mixture for the presence of by-products of the sequencing reaction, namely pyrophosphate. In particular, a primer/template/polymerase complex is contacted with a single type of nucleotide. If that nucleotide is incorporated, the polymerization reaction cleaves the nucleoside triphosphate between the α and β phosphates of the triphosphate chain, releasing pyrophosphate. The presence of released pyrophosphate is then identified using a chemiluminescent enzyme reporter system that converts the pyrophosphate, with AMP, into ATP, then measures ATP using a luciferase enzyme to produce measurable light signals. Where light is detected, the base is incorporated, where no light is detected, the base is not incorporated. Following appropriate washing steps, the various bases are cyclically contacted with the complex to sequentially identify subsequent bases in the template sequence. See, e.g., U.S. Pat. No. 6,210,891, incorporated herein by reference in its entirety for all purposes).

[00112] In related processes, the primer/template/polymerase complex is immobilized upon a substrate and the complex is contacted with labeled nucleotides. The immobilization of the complex may be through the primer sequence, the template sequence and/or the polymerase enzyme, and may be covalent or noncovalent. For example, immobilization of the complex can be via a linkage between the polymerase or the primer and the substrate surface. A variety of types of linkages are useful for this attachment, including, e.g., provision of biotinylated surface components, using e.g., biotin-PEG-silane linkage chemistries, followed by biotinylation of the molecule to be immobilized, and subsequent linkage through, e.g., a streptavidin bridge. Other synthetic coupling chemistries, as well as non-specific protein adsorption can also be employed for immobilization. In alternate configurations, the nucleotides are provided with and without removable terminator groups. Upon incorporation, the label is coupled with the complex and is thus detectable. In the case of terminator bearing nucleotides, all four different nucleotides, bearing individually identifiable labels, are contacted with the

IDT Exhibit 1020.3289

complex. Incorporation of the labeled nucleotide arrests extension, by virtue of the presence of the terminator, and adds the label to the complex. The label and terminator are then removed from the incorporated nucleotide, and following appropriate washing steps, the process is repeated. In the case of non-terminated nucleotides, a single type of labeled nucleotide is added to the complex to determine whether it will be incorporated, as with pyrosequencing. Following removal of the label group on the nucleotide and appropriate washing steps, the various different nucleotides are cycled through the reaction mixture in the same process. See, e.g., U.S. Pat. No. 6,833,246, incorporated herein by reference in its entirety for all purposes). For example, the Illumina Genome Analyzer System is based on technology described in WO 98/44151, hereby incorporated by reference, wherein DNA molecules are bound to a sequencing platform (flow cell) via an anchor probe binding site (otherwise referred to as a flow cell binding site) and amplified in situ on a glass slide. The DNA molecules are then annealed to a sequencing primer and sequenced in parallel base-by-base using a reversible terminator approach. Typically, the Illumina Genome Analyzer System utilizes flow-cells with 8 channels, generating sequencing reads of 18 to 36 bases in length, generating >1.3 Gbp of high quality data per run (see www.illumina.com).

[00113] In yet a further sequence by synthesis process, the incorporation of differently labeled nucleotides is observed in real time as template dependent synthesis is carried out. In particular, an individual immobilized primer/template/polymerase complex is observed as fluorescently labeled nucleotides are incorporated, permitting real time identification of each added base as it is added. In this process, label groups are attached to a portion of the nucleotide that is cleaved during incorporation. For example, by attaching the label group to a portion of the phosphate chain removed during incorporation, i.e., a β, γ, or other terminal phosphate group on a nucleoside polyphosphate, the label is not incorporated into the nascent strand, and instead, natural DNA is produced. Observation of individual molecules typically involves the optical confinement of the complex within a very small illumination volume. By optically confining the complex, one creates a monitored region in which randomly diffusing nucleotides are present for a very short period of time, while incorporated nucleotides are retained within the observation volume for longer as they are being incorporated. This results in a characteristic signal associated with the incorporation event, which is also characterized by a signal profile that is characteristic of the base being added. In related aspects, interacting label components, such as fluorescent resonant energy transfer

IDT Exhibit 1020.3290

(FRET) dye pairs, are provided upon the polymerase or other portion of the complex and the incorporating nucleotide, such that the incorporation event puts the labeling components in interactive proximity, and a characteristic signal results, that is again, also characteristic of the base being incorporated (See, e.g., U.S. Pat. Nos. 6,056,661, 6,917,726, 7,033,764, 7,052,847, 7,056,676, 7,170,050, 7,361,466, 7,416,844 and Published U.S. Patent Application No. 2007-0134128, the full disclosures of which are hereby incorporated herein by reference in their entirety for all purposes).

[00114] In some embodiments, the nucleic acids in the sample can be sequenced by ligation. This method uses a DNA ligase enzyme to identify the target sequence, for example, as used in the polony method and in the SOLiD technology (Applied Biosystems, now Invitrogen). In general, a pool of all possible oligonucleotides of a fixed length is provided, labeled according to the sequenced position. Oligonucleotides are annealed and ligated; the preferential ligation by DNA ligase for matching sequences results in a signal corresponding to the complementary sequence at that position.

[00115] In some embodiments, sequencing comprises extension of a sequencing primer comprising a sequence hybridizable to at least a portion of the complement of the first adaptor oligonucleotide. In some embodiments, sequencing comprises extension of a sequencing primer comprising a sequence hybridizable to at least a portion of the complement of the second adaptor oligonucleotide. A sequencing primer may be of any suitable length, such as about, less than about, or more than about 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 90, 100, or more nucleotides, any portion or all of which may be complementary to the corresponding target sequence (e.g. about, less than about, or more than about 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, or more nucleotides). In some embodiments, sequencing comprises a calibration step, wherein the calibration is based on each of the nucleotides at one or more nucleotide positions in the barcode sequences. Calibration can be useful in processing the sequencing data, for example, by facilitating or increasing the accuracy of identifying a base at a given position in the sequence.

[00116] In some embodiments, accurate identification of the sample from which a target polynucleotide is derived is based on at least a portion of the sequence obtained for the target polynucleotide and is at least 90%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.8%, 99.85%, 99.9%, 99.95%, 99.99%, or more accurate. In some embodiments, the sample source of a target polynucleotide is identified based on a single barcode contained in the sequence. In some embodiments, accuracy can be increased by identifying the source of

IDT Exhibit 1020.3291

a target polynucleotide using two or more barcodes contained in the sequence. Multiple barcodes can be joined to a target polynucleotide by the incorporation of multiple barcodes into a single adaptor/primer to which a target polynucleotide is joined, and/or by joining two or more adaptors/primers having one or more barcodes to a target polynucleotide. In some embodiments, the identity of the sample source of a target polynucleotide comprising two or more barcode sequences may be accurately determined using only one of the barcode sequences that it comprises. In general, accurate identification of a sample from which a target polynucleotide is derived comprises correct identification of a sample source from among two or more samples in a pool, such as about, less than about, or more than about 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 16, 20, 24, 28, 32, 36, 40, 50, 60, 70, 80, 90, 100, 128, 192, 384, 500, 1000 or more samples in a pool.

[00117] In some embodiments, the methods are useful for preparing target polynucleotide(s) for sequencing by the sequencing by ligation methods commercialized by Applied Biosystems (e.g., SOLiD sequencing). In general, double stranded fragment polynucleotides can be prepared by the methods of the present invention, and then incorporated into a water-in-oil emulsion along with polystyrene beads and amplified, for example by PCR. In some cases, alternative amplification methods can be employed in the water-in-oil emulsion such as any of the methods provided herein. The amplified product in each water microdroplet formed by the emulsion can interact, bind, or hybridize with the one or more beads present in that microdroplet leading to beads with a plurality of amplified products of substantially one sequence. When the emulsion is broken, the beads float to the top of the sample and are placed onto an array. The methods can include a step of rendering the nucleic acid bound to the beads single-stranded or partially single stranded. Sequencing primers are then added along with a mixture of four different fluorescently labeled oligonucleotide probes. The probes bind specifically to the two bases in the polynucleotide to be sequenced immediately adjacent and 3′ of the sequencing primer to determine which of the four bases are at those positions. After washing and reading the fluorescence signal from the first incorporated probe, a ligase is added. The ligase cleaves the oligonucleotide probe between the fifth and sixth bases, removing the fluorescent dye from the polynucleotide to be sequenced. The whole process is repeated using a different sequence primer until all of the intervening positions in the sequence are imaged. The process allows the simultaneous reading of millions of DNA fragments in a ‘massively parallel’ manner. This ‘sequence-by-ligation’ technique uses probes that encode for two bases rather than just one allowing

IDT Exhibit 1020.3292

error recognition by signal mismatching, leading to increased base determination accuracy.

[00118] In some embodiments, the methods are useful for preparing target polynucleotide(s) for sequencing by synthesis using the methods commercialized by 454/Roche Life Sciences including but not limited to the methods and apparatus described in Margulies et al., Nature (2005) 437:376-380 (2005); and U.S. Pat. Nos. 7,244,559; 7,335,762; 7,211,390; 7,244,567; 7,264,929; and 7,323,305. In general, double stranded fragment polynucleotides can be prepared by the methods of the present invention, immobilized onto beads, and compartmentalized in a water-in-oil PCR emulsion. In some cases, alternative amplification methods can be employed in the water-in-oil emulsion such as any of the methods provided herein. When the emulsion is broken, amplified fragments remain bound to the beads. The methods can include a step of rendering the nucleic acid bound to the beads single stranded or partially single stranded. The beads can be enriched and loaded into wells of a fiber optic slide so that there is approximately 1 bead in each well. Nucleotides are flowed across and into the wells in a fixed order in the presence of polymerase, sulfhydrolase, and luciferase. Addition of nucleotides complementary to the target strand can result in a chemiluminescent signal that is recorded, such as by a camera. The combination of signal intensity and positional information generated across the plate allows software to determine the DNA sequence.

[00119] In some embodiments, the methods are useful for preparing target polynucleotide(s) for sequencing by the methods commercialized by Helicos BioSciences Corporation (Cambridge, Mass.) as described in U.S. application Ser. No. 11/167,046, and U.S. Pat. Nos. 7,501,245; 7,491,498; 7,276,720; and in U.S. Patent Application Publication Nos. US20090061439; US20080087826; US20060286566; US20060024711; US20060024678; US20080213770; and US20080103058. In general, double stranded fragment polynucleotides can be prepared by the methods of the present invention, and then immobilized onto a flow-cell surface. The methods can include a step of rendering the nucleic acid bound to the flow-cell surface stranded or partially single stranded. Polymerase and labeled nucleotides are then flowed over the immobilized DNA. After fluorescently labeled nucleotides are incorporated into the DNA strands by a DNA polymerase, the surface is illuminated with a laser, and an image is captured and processed to record single molecule incorporation events to produce sequence data.

IDT Exhibit 1020.3293

[00120] In some embodiments, the methods are useful for preparing target polynucleotide(s) for sequencing by the sequencing by ligation methods commercialized by Dover Systems. Generally, double stranded fragment polynucleotides can be prepared by the methods of the present invention. The polynucleotides can then be amplified in an emulsion in the presence of magnetic beads. Any amplification methods can be employed in the water-in-oil emulsion such as any of the methods provided herein. The resulting beads with immobilized clonal polynucleotide polonies are then purified by magnetic separation, capped, amine functionalized, and covalently immobilized in a series of flow cells. The methods can include a step of rendering the nucleic acid bound to the flow-cell surface stranded or partially single stranded. Then, a series of anchor primers are flowed through the cell, where they hybridize to the synthetic oligonucleotide sequences at the 3′ or 5′ end of proximal or distal genomic DNA tags. Once an anchor primer is hybridized, a mixture of fully degenerate nonanucleotides ('nonamers') and T4 DNA ligase is flowed into the cell; each of the nonamer mixture's four components is labeled with one of four fluorophores, which correspond to the base type at the query position. The fluorophore-tagged nonamers selectively ligate onto the anchor primer, providing a fluorescent signal that identifies the corresponding base on the genomic DNA tag. Once the probes are ligated, fluorescently labeling the beads, the array is imaged in four colors. Each bead on the array will fluoresce in only one of the four images, indicating whether there is an A, C, G, or T at the position being queried. After imaging, the array of annealed primer-fluorescent probe complex, as well as residual enzyme, are chemically striped using guanidine HCl and sodium hydroxide. After each cycle of base reads at a given position have been completed, and the primer-fluorescent probe complex has been stripped, the anchor primer is replaced, and a new mixture of fluorescently tagged nonamers is introduced, for which the query position is shifted one base further into the genomic DNA tag. Seven bases are queried in this fashion, with the sequence performed from the 5′ end of the proximal tag, followed by six base reads with a different anchor primer from the 3′ end of the proximal tag, for a total of 13 base pair reads for this tag. This sequence is then repeated for the 5′ and 3′ ends of the distal tag, resulting in another 13 base pair reads. The ultimate result is a read length of 26 bases (thirteen from each of the paired tags). However, it is understood that this method is not limited to 26 base read lengths.

[00121] In some embodiments, the methods are useful for preparing target polynucleotide(s) from selectively enriched populations of specific sequence regions of

IDT Exhibit 1020.3294

interest in a strand-specific manner for sequencing by the methods well known in the art and further described below.

[00122] For example the methods are useful for sequencing by the method commercialized by Illumina as described U.S. Pat. Nos. 5,750,341; 6,306,597; and 5,969,119. In general, double stranded fragment polynucleotides can be prepared by the methods of the present invention to produce amplified nucleic acid sequences tagged at one (e.g., (A)/(A′) or both ends (e.g., (A)/(A′) and (C)/(C′)). In some cases, single stranded nucleic acid tagged at one or both ends is amplified by the methods of the present invention (e.g., by SPIA or linear PCR). The resulting nucleic acid is then denatured and the single stranded amplified polynucleotides are randomly attached to the inside surface of flow-cell channels. Unlabeled nucleotides are added to initiate solid-phase bridge amplification to produce dense clusters of double-stranded DNA. To initiate the first base sequencing cycle, four labeled reversible terminators, primers, and DNA polymerase are added. After laser excitation, fluorescence from each cluster on the flow cell is imaged. The identity of the first base for each cluster is then recorded. Cycles of sequencing are performed to determine the fragment sequence one base at a time. For paired-end sequencing, such as for example, when the polynucleotides are labeled at both ends by the methods of the present invention, sequencing templates can be regenerated in-situ so that the opposite end of the fragment can also be sequenced.

[00123] In some embodiments, the methods are useful for preparing target polynucleotide(s) for sequencing by the methods commercialized by Pacific Biosciences as described in U.S. Pat. Nos. 7,462,452; 7,476,504; 7,405,281; 7,170,050; 7,462,468; 7,476,503; 7,315,019; 7,302,146; 7,313,308; and US Application Publication Nos. US20090029385; US20090068655; US20090024331; and US20080206764. In general, double stranded fragment polynucleotides can be prepared by the methods of the present invention. The polynucleotides can then be immobilized in zero mode waveguide arrays. The methods may include a step of rendering the nucleic acid bound to the waveguide arrays single stranded or partially single stranded. Polymerase and labeled nucleotides are added in a reaction mixture, and nucleotide incorporations are visualized via fluorescent labels attached to the terminal phosphate groups of the nucleotides. The fluorescent labels are clipped off as part of the nucleotide incorporation. In some cases, circular templates are utilized to enable multiple reads on a single molecule.

[00124] Another example of a sequencing technique that can be used in the methods of the provided invention is nanopore sequencing (see e.g. Soni G V and Meller A. (2007) Clin

IDT Exhibit 1020.3295

Chem 53: 1996-2001). A nanopore can be a small hole of the order of 1 nanometer in diameter. Immersion of a nanopore in a conducting fluid and application of a potential across it can result in a slight electrical current due to conduction of ions through the nanopore. The amount of current that flows is sensitive to the size of the nanopore. As a DNA molecule passes through a nanopore, each nucleotide on the DNA molecule obstructs the nanopore to a different degree. Thus, the change in the current passing through the nanopore as the DNA molecule passes through the nanopore can represent a reading of the DNA sequence.

[00125] Another example of a sequencing technique that can be used in the methods of the provided invention is semiconductor sequencing provided by Ion Torrent (e.g., using the Ion Personal Genome Machine (PGM)). Ion Torrent technology can use a semiconductor chip with multiple layers, e.g., a layer with micro-machined wells, an ion-sensitive layer, and an ion sensor layer. Nucleic acids can be introduced into the wells, e.g., a clonal population of single nucleic can be attached to a single bead, and the bead can be introduced into a well. To initiate sequencing of the nucleic acids on the beads, one type of deoxyribonucleotide (e.g., dATP, dCTP, dGTP, or dTTP) can be introduced into the wells. When one or more nucleotides are incorporated by DNA polymerase, protons (hydrogen ions) are released in the well, which can be detected by the ion sensor. The semiconductor chip can then be washed and the process can be repeated with a different deoxyribonucleotide. A plurality of nucleic acids can be sequenced in the wells of a semiconductor chip. The semiconductor chip can comprise chemical-sensitive field effect transistor (chemFET) arrays to sequence DNA (for example, as described in U.S. Patent Application Publication No. 20090026082). Incorporation of one or more triphosphates into a new nucleic acid strand at the 3′ end of the sequencing primer can be detected by a change in current by a chemFET. An array can have multiple chemFET sensors.

Genetic Analysis

[00126] The methods of the present invention can be used in the analysis of genetic information of selective genomic regions of interest as well as genomic regions which may interact with the selective region of interest. Amplification methods as disclosed herein can be used in the devices, kits, and methods known to the art for genetic analysis, such as, but not limited to those found in U.S. Pat. Nos. 6,449,562, 6,287,766, 7,361,468, 7,414,117, 6,225,109, and 6,110,709. In some cases, amplification methods of the present invention can be used to amplify target nucleic acid for DNA hybridization studies to

IDT Exhibit 1020.3296

determine the presence or absence of polymorphisms. The polymorphisms, or alleles, can be associated with diseases or conditions such as genetic disease. In other cases the polymorphisms can be associated with susceptibility to diseases or conditions, for example, polymorphisms associated with addiction, degenerative and age related conditions, cancer, and the like. In other cases, the polymorphisms can be associated with beneficial traits such as increased coronary health, or resistance to diseases such as HIV or malaria, or resistance to degenerative diseases such as osteoporosis, Alzheimer's or dementia.

Kits

[00127] Any of the compositions described herein may be comprised in a kit. In a non-limiting example the kit, in suitable container means, comprises: one or more RNA DNA stem-loop oligonucleotides. The kit can further contain enzymes and/or reagents useful for ligation, cleavage and or amplification. The kit can contain a DNA-polymerase. The kit can contain reagents for amplification, for example reagents useful for single primer isothermal amplification methods. The kit can further optionally contain reagents for sequencing, for example, reagents useful for next-generation massively parallel sequencing methods.

[00128] The containers of the kits can generally include at least one vial, test tube, flask, bottle, syringe or other containers, into which a component may be placed, and preferably, suitably aliquoted. Where there is more than one component in the kit, the kit also can generally contain a second, third or other additional container into which the additional components can be separately placed. However, various combinations of components can be comprised in a container.

[00129] When the components of the kit are provided in one or more liquid solutions, the liquid solution can be an aqueous solution. However, the components of the kit can be provided as dried powder(s). When reagents and/or components are provided as a dry powder, the powder can be reconstituted by the addition of a suitable solvent.

[00130] A kit can include instructions for employing the kit components as well the use of any other reagent not included in the kit. Instructions can include variations that can be implemented.

[00131] In one aspect, the invention provides kits containing any one or more of the elements disclosed in the above methods and compositions. In some embodiments, a kit comprises a composition of the invention, in one or more containers. In some

IDT Exhibit 1020.3297

embodiments, the invention provides kits comprising adaptors, primers, and/or other oligonucleotides described herein. In some embodiments, the kit further comprises one or more of: (a) a DNA ligase, (b) a DNA-dependent DNA polymerase, (c) an RNA-dependent DNA polymerase, (d) random primers, (e) primers comprising at least 4 thymidines at the 3' end, (f) a DNA endonuclease, (g) a DNA-dependent DNA polymerase having 3' to 5' exonuclease activity, (h) a plurality of primers, each primer having one of a plurality of selected sequences, (i) a DNA kinase, (j) a DNA exonuclease, (k) magnetic beads, (l) an enzyme comprising RNase H activity, (m) an RNA ligase, and (n) one or more buffers suitable for one or more of the elements contained in said kit. The adaptors, primers, other oligonucleotides, and reagents can be, without limitation, any of those described above. Elements of the kit can further be provided, without limitation, in any suitable amounts and/or using any of the combinations (such as in the same kit or same container) described above or any other suitable combination known in the art. The kits may further comprise additional agents, such as those described above, for use according to the methods of the invention. The kit elements can be provided in any suitable container, including but not limited to test tubes, vials, flasks, bottles, ampules, syringes, or the like. The agents can be provided in a form that may be directly used in the methods of the invention, or in a form that requires preparation prior to use, such as in the reconstitution of lyophilized agents. Agents may be provided in aliquots for single-use or as stocks from which multiple uses, such as in a number of reaction, may be obtained.

EXAMPLES

Example 1: *Whole transcriptome amplification using stem-loop chimeric amplification primer and the corresponding linear (non stem-loop) chimeric amplification primer*

[00132] In this example cDNA was generated from total RNA (50pg total HeLa total RNA) and control sample without added total RNA (no template control, NTC). The reagents and protocols for first and second strand cDNA were those of Ovation Pico (NuGEN Technologies). The chimeric first strand primer used for this example comprises a combination of a chimeric primer with a random sequence at the 3'-end (Primer A random, in Table 1) and a chimeric primer with an oligo dT at the 3'-portion (Primer A dT, in Table 1) and further comprises the same 5'-RNA tail. The 5'-RNA tail of the first strand cDNA primer comprises the same sequence as the single chimeric

IDT Exhibit 1020.3298

amplification primer used in the isothermal single primer amplification step, as shown below.

[00133] The double stranded cDNA with an RNA/DNA heteroduplex at one end generated by the first and second cDNA synthesis steps above was purified as per the manufacturer instruction, and subjected to the single primer isothermal amplification (SPIA) step, using the reagents and conditions of the Ovation Pico kit (NuGEN Technologies) and either the stem-loop chimeric amplification primer (Stem-Loop amplification primer A, Table 1) or the corresponding linear chimeric amplification primer (Linear amplification primer A, Table1). The amplification primer generated from the stem-loop amplification primer as well as the linear amplification primers are chimeric primers comprising a DNA sequence at the 3'-end and an RNA sequence at the 5'-end, which are the same sequence as the corresponding tail of the first strand primers. SPIA amplification was carried out according to manufacturer instructions. SPIA amplification products were purified using MinElute columns (QIAGEN) using the manufacturer instructions and reagents.

Table 1

Primer ID	Sequence
Primer A random	*cuucuauaguuuagguaacuuuguguuuga*NNNNNN
Primer A dT	*cuucuauaguuuagguaacuuuguguuuga*TTTTTTTTTTTTNN
Stem-loop amplification primer A	*ucaaacacuucuauaguuuagguaacuuug*TGTTTGA
Linear amplification primer A	*cuucuauaguuuagguaacuuug* TGTTTGA

[00134] Table 1 illustrates the primer sequences with ribonucleotides appearing in lower case and deoxyribonucleotides appearing in upper case. "N" represents a random nucleotide.

[00135] The yields of transcriptome amplification employing the stem-loop chimeric amplification and the corresponding linear chimeric amplification primer using the method of the invention are summarized in Table 2.

Table 2.

Sample	*Amplification primer*	*µg*	*Avg µg (Std Dev)*
50 pg RNA	Linear amplification primer A	4.87	**4.50 (0.32)**
50 pg RNA	Linear amplification primer A	4.31	
50 pg RNA	Linear amplification primer A	4.32	
NTC	Linear amplification primer A	0.73	1.36 (1.25)
NTC	Linear amplification primer A	1.99	
50 pg RNA	Stem-loop amplification primer A	5.61	**5.51 (0.11)**
50 pg RNA	Stem-loop amplification primer A	5.52	
50 pg RNA	Stem-loop amplification primer A	5.38	
NTC	Stem-loop amplification primer A	0.85	0.90 (0.11)
NTC	Stem-loop amplification primer A	0.96	

[00136] The results demonstrate the higher amplification yields obtained in reactions employing the stem-loop chimeric amplification primer of the invention as compared to the corresponding linear chimeric amplification primer. This is likely due to employing a pro-primer which is activated during the initial reaction thus leading to lower production of non-specific products as indicated by the lower yield of amplification product in the no target control.

Example 2: *Whole genome amplification (genomic DNA derived from FFPE sample) using stem-loop chimeric amplification primer and the corresponding linear (non stem-loop) chimeric amplification primer*

[00137] Whole genome amplification (WGA) was performed as described in the Ovation WGA FFPE System Users Guide (Part #6200, NuGEN Technologies Inc). Briefly, 50 or 100 ng of isolated FFPE genomic DNA was fragmented, end-repaired and adenosine-tailed to generate 5′ phosphorylated/3′ A-overhang fragments. Adaptors suitable for amplification using the method of the invention were utilized for this example (Adaptor

B, Table 3). Adaptor B comprises two oligonucleotides that when annealed form a single T-overhang to aid in ligation efficiency and a long 3' overhang that comprises the priming site for the SPIA chimeric amplification primer. After the adaptors were ligated to the fragments, the reactions were purified to remove excess adaptors using Agencourt® RNAClean® XP beads. The resulting adaptor ligated fragments were isothermally amplified using the enzyme mixture and protocol from the kit and purified using an appropriate purification method, such as Qiagen® spin columns. Either the stem-loop chimeric amplification primer (Stem-loop amplification primer B, Table 3) or the corresponding linear chimeric amplification primer (Linear amplification primer B, Table 3), were employed in the isothermal single primer amplification step.

[00138] Amplification products were purified as per the manufacturer instructions and the yields of whole genome amplification reactions were determined using Nanodrop (Thermo Scientific; Figure 6). The size-distribution of the amplification products from these reactions were determined using BioAnalyze (Agilent) (Figure 7).

Table 3

Oligonucleotide	Sequence
Adaptor B	5'-cgtagactcaTctgcctatagtgagtcgtattacc-3' |||||||||||||||| 3'-tgcatctgagtAgacgga-5'
Stem-loop chimeric amplification primer B	5'-rUrCrUrGrCrCrUrGrGrUrArArUrArCrGrArCrUrCrArCrUrArUAGGCAGA-3'
Linear chimeric amplification primer B	5'-rGrGrUrArArUrArCrGrArCrUrCrArCrUrArUAGGCAGA-3'

[00139] While preferred embodiments of the present invention have been shown and described herein, it will be obvious to those skilled in the art that such embodiments are provided by way of example only. Numerous variations, changes, and substitutions will now occur to those skilled in the art without departing from the invention. It should be understood that various alternatives to the embodiments of the invention described herein may be employed in practicing the invention. It is intended that the following claims define the scope of the invention and that methods and structures within the scope of these claims and their equivalents be covered thereby.

IDT Exhibit 1020.3301

CLAIMS

WHAT IS CLAIMED IS:

1. A method of preparing a target nucleic acid molecule for further processing, comprising:

(a) attaching an RNA/DNA chimeric stem-loop oligonucleotide having a portion of its stem and at least a portion of its loop comprise of RNA to said target nucleic acid molecule thereby producing a stem-loop-target complex; and

(b) contacting said stem-loop-target complex with an RNase that cleaves RNA from a RNA/DNA heteroduplex, thereby preparing the stem-loop-target complex for further processing.

2. The method of claim 1, wherein the RNase is RNase H.

3. The method of claims 1-2, wherein further processing comprises nucleic acid sequencing.

4. The method of claim 3, wherein sequencing comprises single molecule sequencing.

5. The method of claim 3, wherein sequencing comprises sequencing by synthesis.

6. The method of claims 1-5, wherein further processing comprises nucleic acid amplification.

7. The method of claim 6, wherein a portion of the stem-loop oligonucleotide is further used to prime an amplification reaction.

8. The method of claim 6-7, wherein the amplification is isothermal amplification.

9. The method of claim 6-7, wherein the amplification is PCR.

10. The method of claims 1-9, wherein the nucleic acid is produced by fragmenting a longer nucleic acid.

11. The method of claims 1-10, wherein the target nucleic acid is blunt ended.

12. The method of claims 1-11, wherein the target nucleic acid comprises a 3' overhang.

IDT Exhibit 1020.3302

13. The method of claims 1-12, wherein the target nucleic acid comprises a 5' overhang.

14. The method of claim 12, wherein the stem-loop oligonucleotide further comprises a 3' overhang.

15. The method of claim 13, wherein the stem-loop oligonucleotide further comprises a 5' overhang.

16. The method of claims 14-15, wherein the stem-loop oligonucleotide overhang and nucleic acid overhang are at least partially complementary to each other.

17. The method of claims 1-16, wherein the attaching step comprises attaching a first RNA/DNA chimeric stem-loop oligonucleotide to a first end of said target nucleic acid and a second RNA/DNA chimeric stem-loop oligonuclotide to a second end of said target nucleic acid.

18. The method of claim 17, wherein said first and said second RNA/DNA chimeric stem-loop oligonucleotides are the same.

19. The method of claims 1-18, wherein the attaching step comprises the use of a ligase.

20. The method of claim 19, wherein the ligase is T4 ligase.

21. The method of claims 1-20, wherein the stem-loop oligonucleotide lacks 5' phosphate.

22. The method of claims 1-21, wherein the target nucleic acid is further extended along the cleaved stem-loop oligonucleotide sequence, thereby forming a target nucleic acid extension product, wherein a portion of the stem-loop oligonucleotide acts as a template.

23. The method of claim 22, wherein the stem-loop oligonucleotide that is in duplex with the nucleic acid extension is further cleaved using an RNase that cleaves RNA from a RNA/DNA heteroduplex, thereby forming a priming ready construct.

24. The method of claim 23, wherein the priming ready construct is further amplified.

25. The method of claim 24, wherein the amplification comprises single primer isothermal amplification (SPIA).

IDT Exhibit 1020.3303

26. The method of claims 1-25, wherein at least one loop of the stem-loop oligonucleotide further comprises one or more sequence elements selected from the group consisting of a primer binding site, a sequencing primer binding site, a barcode, a promoter, an adaptor-group sequence, a restriction enzyme recognition site, a target complementary overhang, a probe complementary overhang, a probe binding site, a random sequence, and a complementary sequence to any of the sequence elements within the group.

27. The method of claims 1-26, wherein at least one stem of the stem-loop oligonucleotide further comprises one or more sequence elements selected from the group consisting of a primer binding site, a sequencing primer binding site, a barcode, a promoter, an adaptor-group sequence, a restriction enzyme recognition site, a target complementary overhang, a probe complementary overhang, a probe binding site, a random sequence, and a complementary sequence to any of the sequence elements within the group.

28. The method of claims 1-27, wherein at least one overhang of the stem-loop oligonucleotide further comprises one or more sequence elements selected from the group consisting of a primer binding site, a sequencing primer binding site, a barcode, a promoter, an adaptor-group sequence, a restriction enzyme recognition site, a target complementary overhang, a probe complementary overhang, a probe binding site, a random sequence, and a complementary sequence to any of the sequence elements within the group.

29. A method for amplifying a nucleic acid molecule, comprising:

contacting in a single mixture

(a) an RNA/DNA stem-loop adaptor having a portion of its stem and at least a portion of its loop comprise of RNA;

(b) a double-stranded nucleic acid with a 3' overhang;

(c) RNase that cleaves RNA from a RNA/DNA heteroduplex;

(d) a DNA polymerase with strand displacement activity.

30. The method of claim 29, further comprising forming an extension product by priming the double-stranded nucleic acid at its 3' overhang using at least a portion of the stem-loop oligonucleotide.

IDT Exhibit 1020.3304

31. The method of claim 30, wherein the loop-RNA in the extension product is further cleaved by the RNase, thereby forming a second double stranded nucleic acid with at least one 3' overhang.

32. The method of claims 29-31, wherein the RNase comprises RNase H.

33. The method of claims 29-32, further comprising at least a second amplification mixture with at least a second stem-loop oligonucleotide, wherein each stem-loop oligonucleotide further comprises a barcode sequence.

34. The method of claims 29-33, wherein the 3' overhang is produced by

(a) forming a product, wherein the stem-loop oligonucleotide is attached to a double stranded nucleic acid by one strand;

(b) extending the double stranded nucleic acid along the stem-loop oligonucleotide; and

(c) cleaving the stem-loop oligonucleotide using an RNase that cleaves RNA sequences in an RNA/DNA heteroduplex.

35. The method of claims 29-33, wherein the 3' overhang is produced during single primer isothermal amplification.

36. A kit for amplifying nucleic acids, comprising:

(a) a stem-loop oligonucleotide, wherein at least a portion of the stem comprises RNA and wherein at least a portion of the loop comprises RNA;

(b) an RNase that cleaves RNA sequences in an RNA/DNA duplex;

(c) a ligase; and

(d) a DNA polymerase.

37. A method of amplifying a double stranded target nucleic acid molecule, comprising:

(a) attaching a strand of a first stem-loop oligonucleotide to said double stranded target nucleic acid molecule thereby producing an oligonucleotide-attached nucleic acid molecule; and

(b) cleaving part of the stem-loop oligonucleotide using an RNase; and using a second stem-loop oligonucleotide as a primer in amplifying the double stranded target nucleic acid molecule.

IDT Exhibit 1020.3305

38. The method of claim 37, wherein the second stem-loop oligonucleotide is an exact copy of the first stem-loop oligonucleotide.

39. The method of claims 37-38, wherein the RNase comprises RNase H.

40. The method of claims 37-39, wherein the stem-loop oligonucleotide comprises an RNA/DNA heteroduplex in a stem.

41. The method of claims 37-40, wherein a strand of the oligonucleotide-attached nucleic acid molecule is extended.

42. The method of claim 41, wherein the extension yields a second RNA/DNA heteroduplex.

43. The method of claim 42, wherein cleaving part of the stem-loop oligonucleotide comprises cleaving the second RNA/DNA heteroduplex.

44. The method of claim 43, wherein cleaving part of the stem-loop oligonucleotide generates a priming ready construct.

45. The method of claim 43, wherein the priming ready construct comprises a single stranded region that is 3' to the double stranded target nucleic acid.

46. The method of claims 37-45, wherein the amplification comprises PCR.

47. The method of claims 37-45, wherein the amplification comprises SPIA.

48. The method of claims 37-47, wherein the attaching comprises ligation.

49. The method of claim 48, wherein T4 ligase is utilized in ligation.

50. The method of claims 37-49, wherein the stem-loop oligonucleotide lacks 5' phosphate.

51. The method of claims 37-50, wherein the target nucleic acid is produced by fragmenting a longer nucleic acid.

52. The method of claims 37-51, wherein the target nucleic acid is blunt ended.

53. The method of claims 37-52, wherein the target nucleic acid comprises a 3' overhang.

IDT Exhibit 1020.3306

54. The method of claims 37-53, wherein the target nucleic acid comprises a 5' overhang.

55. The method of claim 53, wherein the stem-loop oligonucleotide further comprises a 3' overhang.

56. The method of claim 54, wherein the stem-loop oligonucleotide further comprises a 5' overhang.

57. The method of claims 55-56, wherein the stem-loop oligonucleotide overhang and nucleic acid overhang are at least partially complementary to each other.

58. The method of claims 37-57, wherein at least one loop of the stem-loop oligonucleotide further comprises one or more sequence elements selected from the group consisting of a primer binding site, a sequencing primer binding site, a barcode, a promoter, an adaptor-group sequence, a restriction enzyme recognition site, a target complementary overhang, a probe complementary overhang, a probe binding site, a random sequence, and a complementary sequence to any of the sequence elements within the group.

59. The method of claims 37-58, wherein at least one stem of the stem-loop oligonucleotide further comprises one or more sequence elements selected from the group consisting of a primer binding site, a sequencing primer binding site, a barcode, a promoter, an adaptor-group sequence, a restriction enzyme recognition site, a target complementary overhang, a probe complementary overhang, a probe binding site, a random sequence, and a complementary sequence to any of the sequence elements within the group.

60. The method of claims 37-59, wherein at least one overhang of the stem-loop oligonucleotide further comprises one or more sequence elements selected from the group consisting of a primer binding site, a sequencing primer binding site, a barcode, a promoter, an adaptor-group sequence, a restriction enzyme recognition site, a target complementary overhang, a probe complementary overhang, a probe binding site, a random sequence, and a complementary sequence to any of the sequence elements within the group.

IDT Exhibit 1020.3307

Figure 1

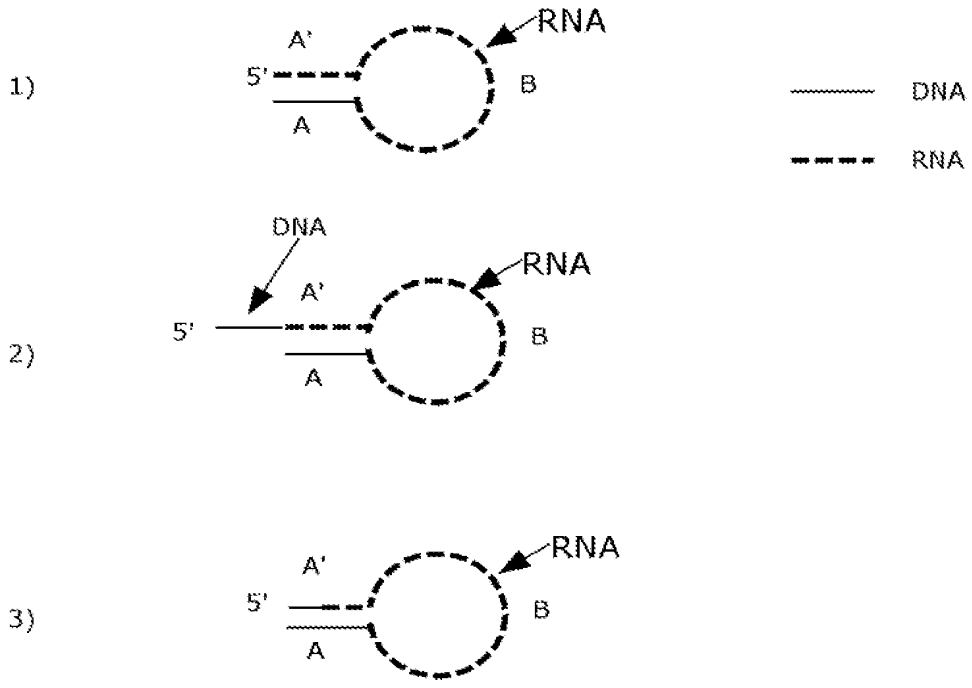


Figure 2a

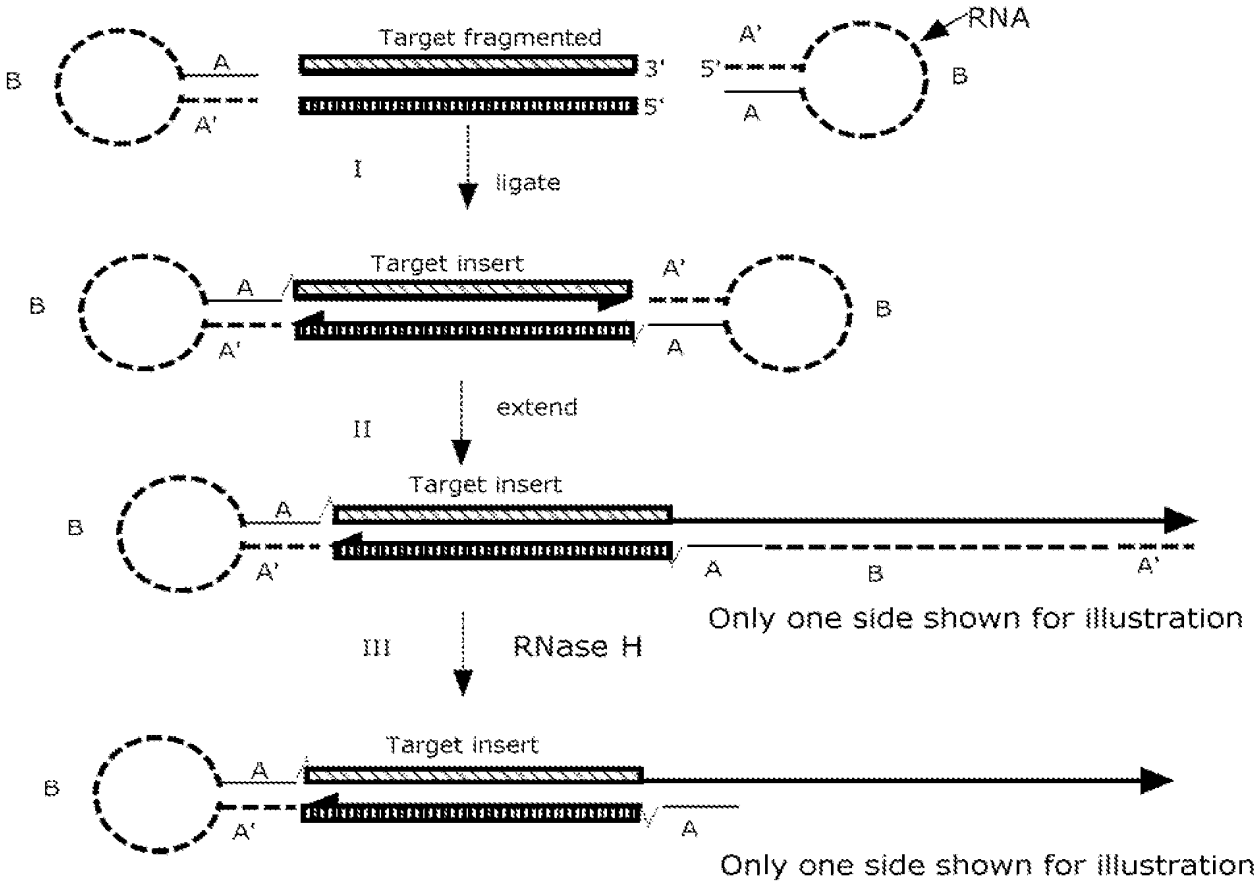

Target fragmented
3'
5'
A'
5'
RNA
B
A
A'
B
I
ligate
Target insert
A'
B
A
A'
A
B
II
extend
Target insert
B
A
A'
A
B
A'
Only one side shown for illustration
III
RNase H
Target insert
B
A
A'
A
Only one side shown for illustration

Figure 2b

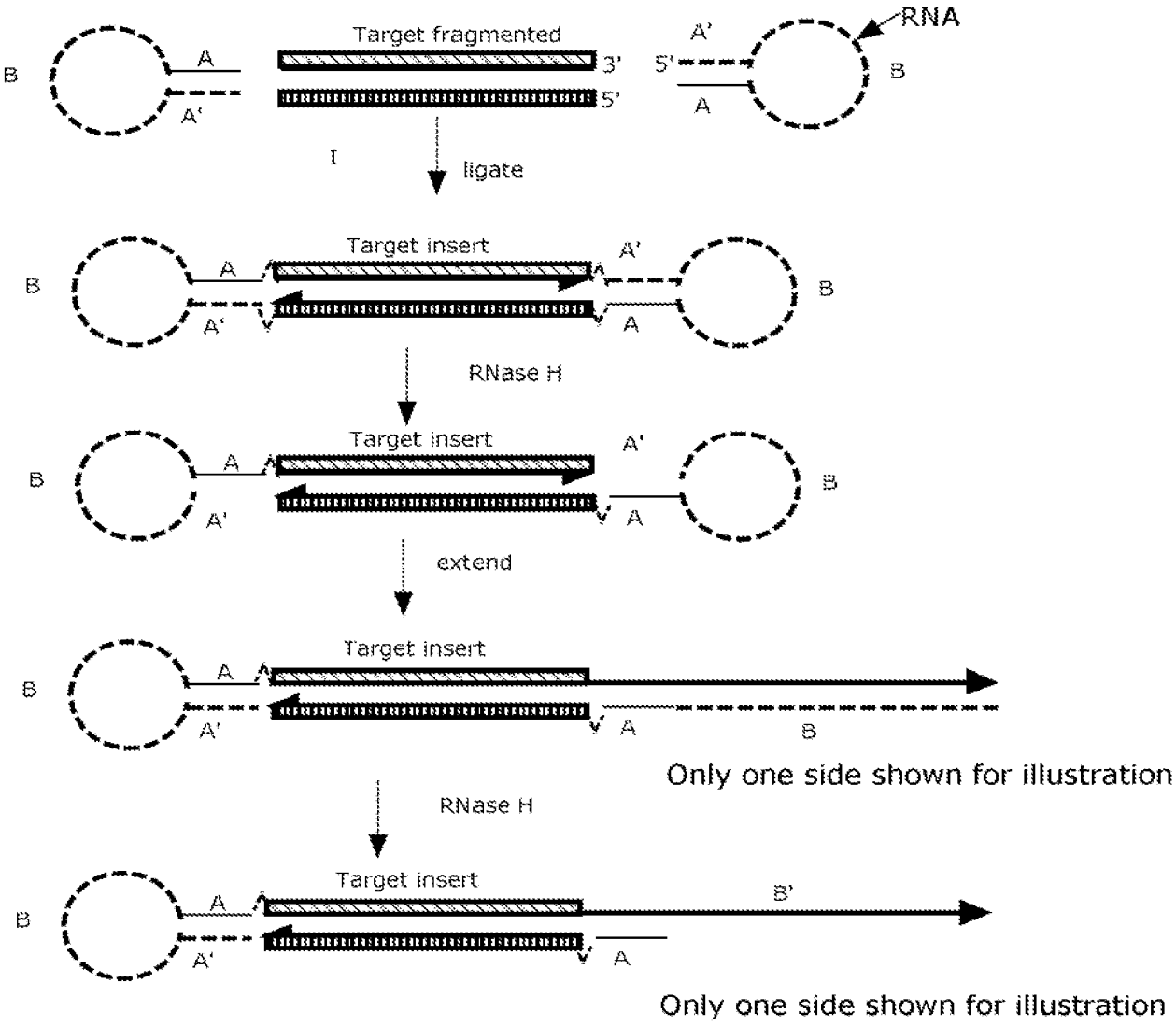

Target fragmented
RNA
A'
A
B
3'
5'
I
ligate
Target insert
RNase H
extend
Only one side shown for illustration
B'

Figure 3

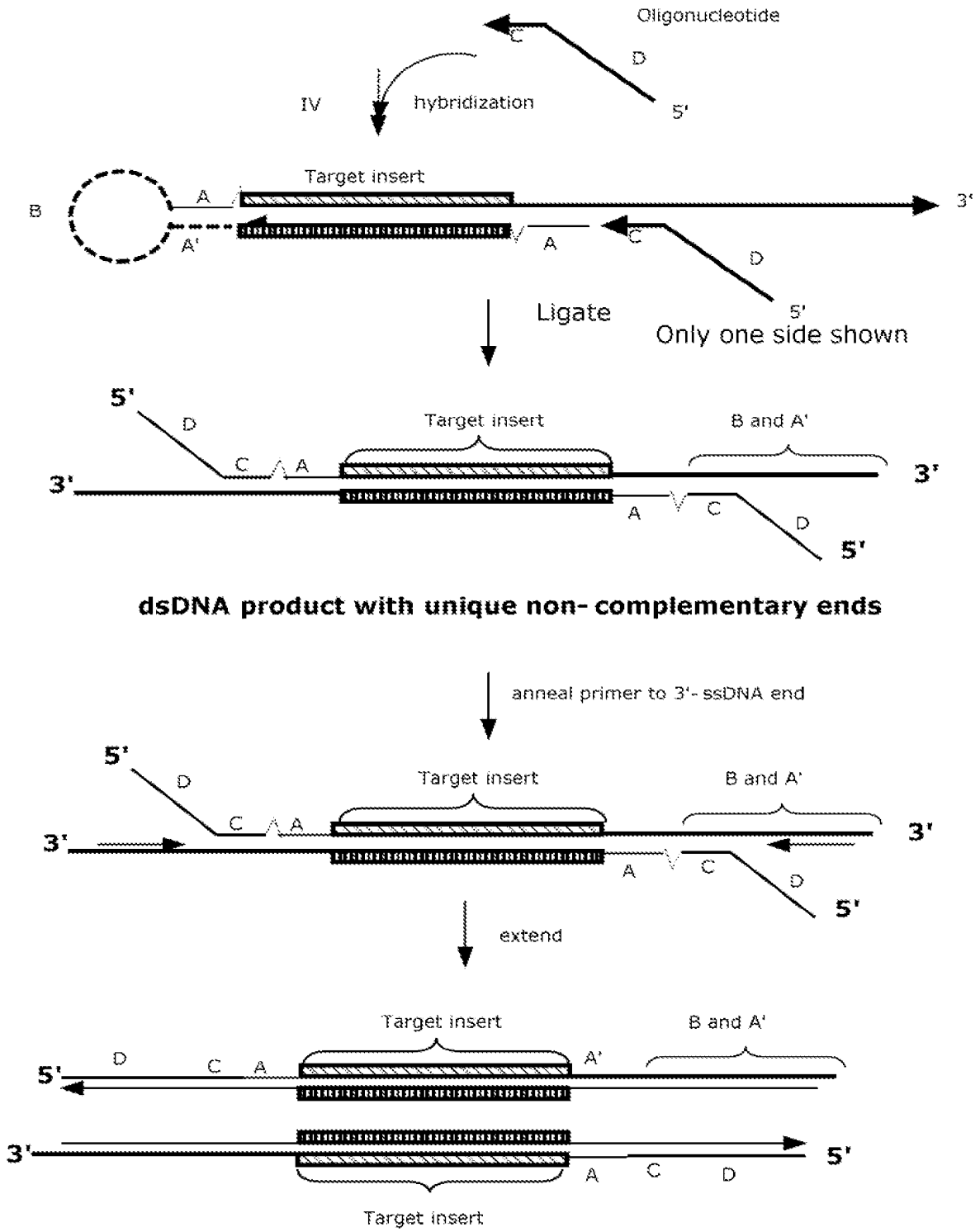

Oligonucleotide
C
D
5'
IV
hybridization
Target insert
B
A
A'
3'
A
C
D
5'
Ligate
Only one side shown
5'
D
C
A
Target insert
B and A'
3'
3'
A
C
D
5'
dsDNA product with unique non- complementary ends
anneal primer to 3'- ssDNA end
5'
D
C
A
Target insert
B and A'
3'
3'
A
C
D
5'
extend
Target insert
B and A'
D
C
A
A'
5'
3'
5'
A
C
D
Target insert

Figure 4a

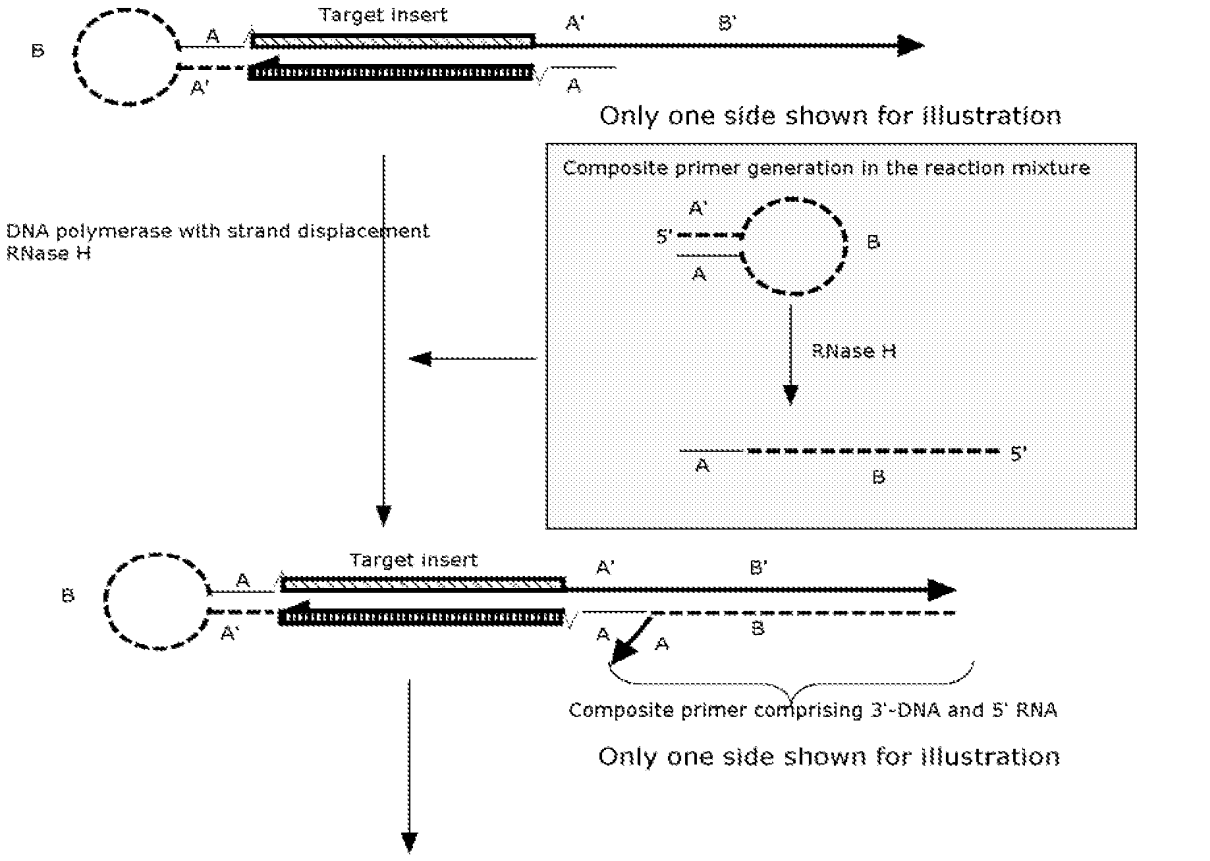

Target insert
A
A'
B'
B
A'
A
Only one side shown for illustration
Composite primer generation in the reaction mixture
A'
5'
B
A
RNase H
A
5'
B
DNA polymerase with strand displacement
RNase H
Target insert
A
A'
B'
B
A'
A
A
B
Composite primer comprising 3'-DNA and 5' RNA
Only one side shown for illustration
SPIA amplification
Both target insert strands are amplified
Generate copies of sense and antisense strands comprising sequence A or A' at the 5'-region

Figure 4b

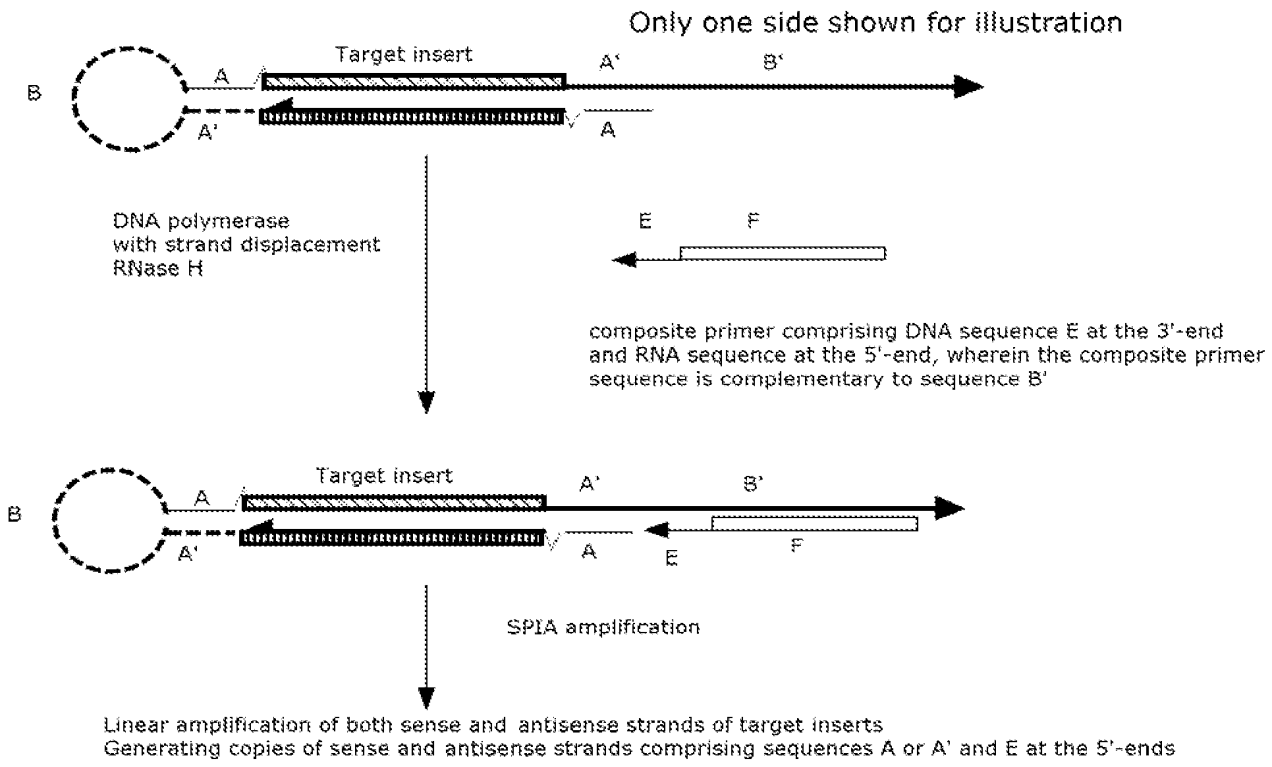

Only one side shown for illustration
Target insert
A
A'
B'
B
A'
A
DNA polymerase
with strand displacement
RNase H
E
F
composite primer comprising DNA sequence E at the 3'-end
and RNA sequence at the 5'-end, wherein the composite primer
sequence is complementary to sequence B'
Target insert
A
A'
B'
B
A'
A
E
F
SPIA amplification
Linear amplification of both sense and antisense strands of target inserts
Generating copies of sense and antisense strands comprising sequences A or A' and E at the 5'-ends

Figure 5

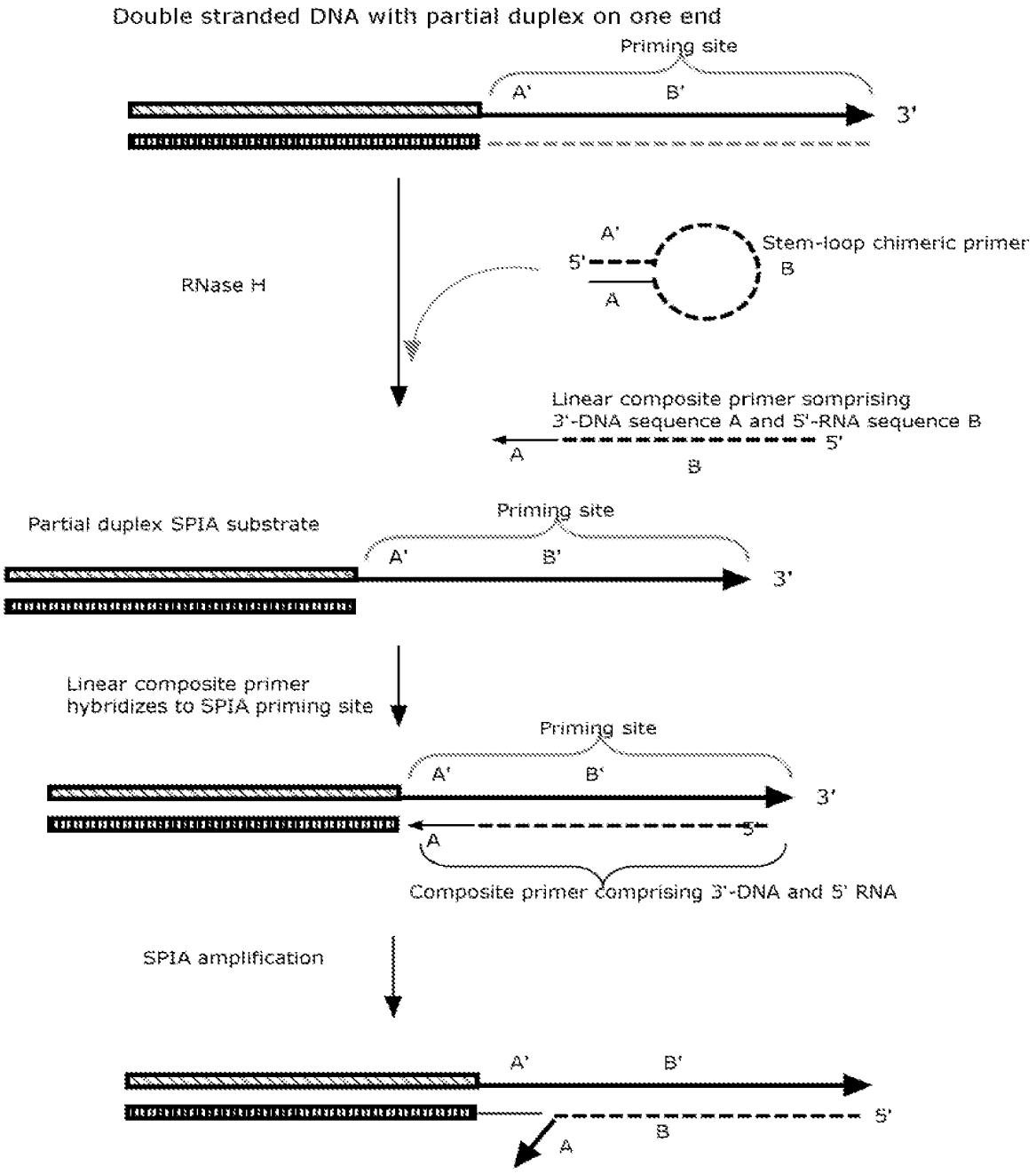

Double stranded DNA with partial duplex on one end
Priming site
A'
B'
3'
RNase H
A'
5'
A
Stem-loop chimeric primer
B
Linear composite primer somprising
3'-DNA sequence A and 5'-RNA sequence B
5'
A
B
Partial duplex SPIA substrate
Priming site
A'
B'
3'
Linear composite primer
hybridizes to SPIA priming site
Priming site
A'
B'
3'
5'
A
Composite primer comprising 3'-DNA and 5' RNA
SPIA amplification
A'
B'
5'
B
A

Figure 6

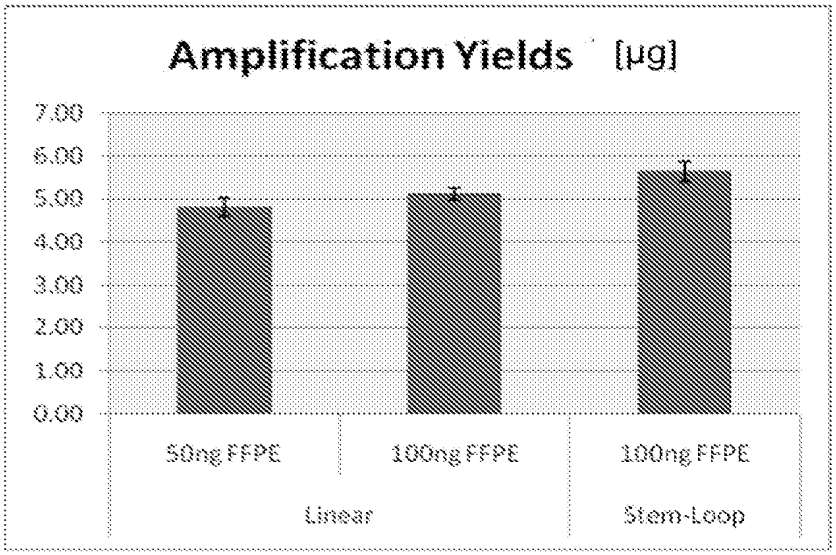
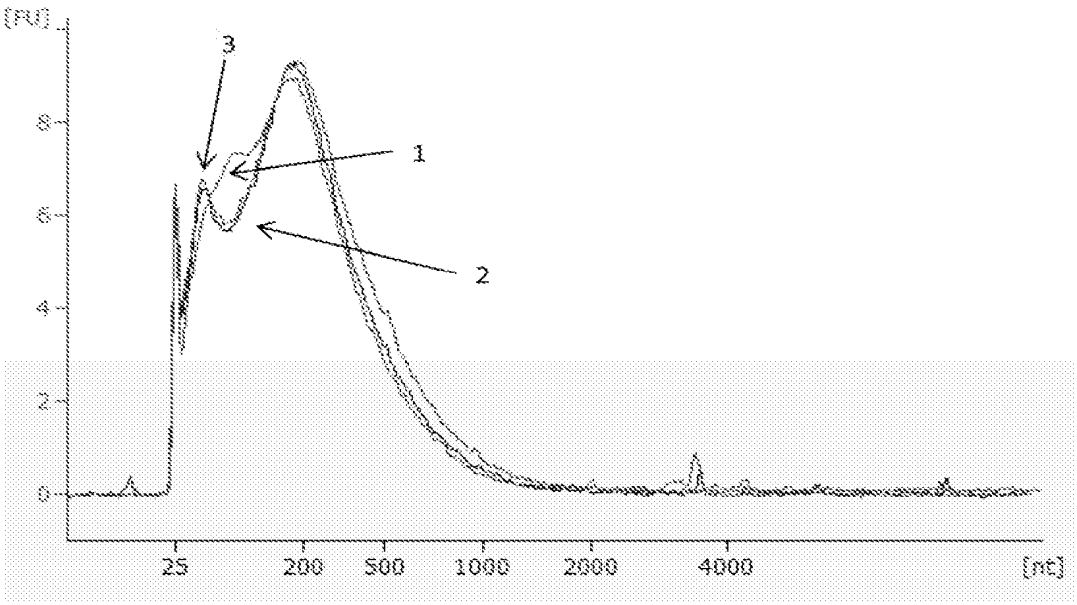

Amplification Yields [µg]
7.00
6.00
5.00
4.00
3.00
2.00
1.00
0.00
50ng FFPE
100ng FFPE
100ng FFPE
Linear
Stem-Loop

Figure 7



INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 12/22448

A. CLASSIFICATION OF SUBJECT MATTER
IPC(8) - C12P 19/34 (2012.01)
USPC - 435/91.2
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
USPC: 435/91.2

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
USPC: 435/91.2; 435/91.51; 536/23.1 (text search)

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
Electronic data bases: PubWEST (EPAB, USPT, JPAB, PGPB); Google Scholar
Search terms: Stem?loop structure, (i.e. hairpin loop), RNA/DNA chimera, adaptor, ligation, primer, nucleic acid manipulation, amplification, RNase H, Next Gen sequencing

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X --- Y	US 7,803,550 B2 (MAKAROV et al.) 28 September 2010 (28.09.2010). Especially col 2 ln 6-18, col 4 ln 8-49, col 11 ln 17-35, sheet 5 fig 3, sheet 9 fig 6B.	1, 2, 29-32, 36-39 ---------- 3-5
Y	METZKER, Sequencing technologies - the next generation. Nature Rev Genetics, January 2010, Vol 11, No 1, Pages 31-46. Especially pg 32.	3-5
A	US 2009/0203085 A1 (KURN et al.) 13 August 2009 (13.08.2009). Especially sheet 3 fig 3.	29-32, 36-39
A	US 2006/0216724 A1 (CHRISTIANS et al.). 28 September 2006 (28.09.2006). Especially sheet 4 fig 4.	29-32
A	BROUDE, Stem-loop oligonucleotides: a robust tool for molecular biology and biotechnology. Trends Biotechnol, June 2002, Vol 20, No 6, Pages 249-256. Especially pg 250 fig 1, pg 252 fig 2.	1-5, 29-32, 36-39

☐ Further documents are listed in the continuation of Box C. ☐

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

Date of the actual completion of the international search	Date of mailing of the international search report
17 April 2012 (17.04.2012)	10 MAY 2012
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-3201	Authorized officer: Lee W. Young PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

Form PCT/ISA/210 (second sheet) (July 2009)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 12/22448

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

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

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

3. ☒ Claims Nos.: 6-28, 33-35, and 40-60
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

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

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

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

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

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

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.

IDT Exhibit 1020.3318

Espacenet

Bibliographic data: WO2007052006 (A1) — 2007-05-10

METHOD OF PREPARING LIBRARIES OF TEMPLATE POLYNUCLEOTIDES

Inventor(s): GORMLEY NIALL ANTHONY [GB]; SMITH GEOFFREY PAUL [GB]; BENTLEY DAVID [GB]; RIGATTI ROBERTO [GB] ± (GORMLEY, NIALL, ANTHONY, ; SMITH, GEOFFREY, PAUL, ; BENTLEY, DAVID, ; RIGATTI, ROBERTO)

Applicant(s): SOLEXA LTD [GB]; GORMLEY NIALL ANTHONY [GB]; SMITH GEOFFREY PAUL [GB]; BENTLEY DAVID [GB]; RIGATTI ROBERTO [GB] ± (SOLEXA LIMITED, ; GORMLEY, NIALL, ANTHONY, ; SMITH, GEOFFREY, PAUL, ; BENTLEY, DAVID, ; RIGATTI, ROBERTO)

Classification: - international: *C12Q1/68*
- cooperative: **C12N15/1068**; **C12N15/1096**; **C12Q1/6855**; **C12Q1/686**; C40B40/06; Y10T436/143333 more

Application number: WO2006GB04055 20061031 Global Dossier

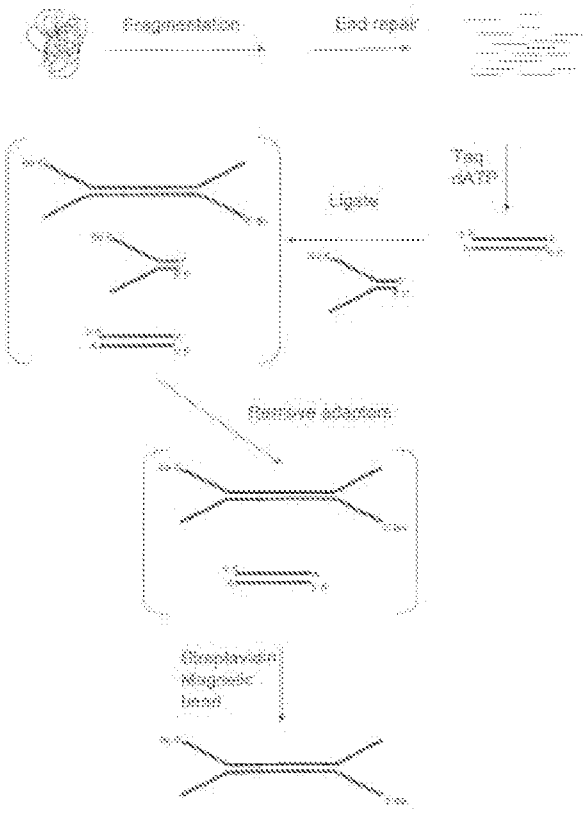
Priority number(s): GB20050022310 20051101 ; US20060486953 20060714

Also published as: DK1954818 (T3) EP1954818 (A1) EP1954818 (B1) EP2423325 (A1) ES2450047 (T3) more

Abstract of WO2007052006 (A1)

The invention provides a method of preparing a library of template polynucleotides which have common sequences at their 5' ends and at their 3' ends, and also the use of the library of templates in methods of solid-phase nucleic acid amplification.

IDT Exhibit 1020.3319



IDT Exhibit 1020.3320

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
10 May 2007 (10.05.2007)

PCT

(10) International Publication Number
WO 2007/052006 A1

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

(21) International Application Number:
PCT/GB2006/004055

(22) International Filing Date: 31 October 2006 (31.10.2006)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
0522310.2 1 November 2005 (01.11.2005) GB
11/486,953 14 July 2006 (14.07.2006) US

(71) Applicant *(for all designated States except US)*: **SOLEXA LIMITED** [GB/GB]; Chesterford Research Park, Little Chesterford, Saffron Walden, Essex CB10 1XL (GB).

(72) Inventors; and

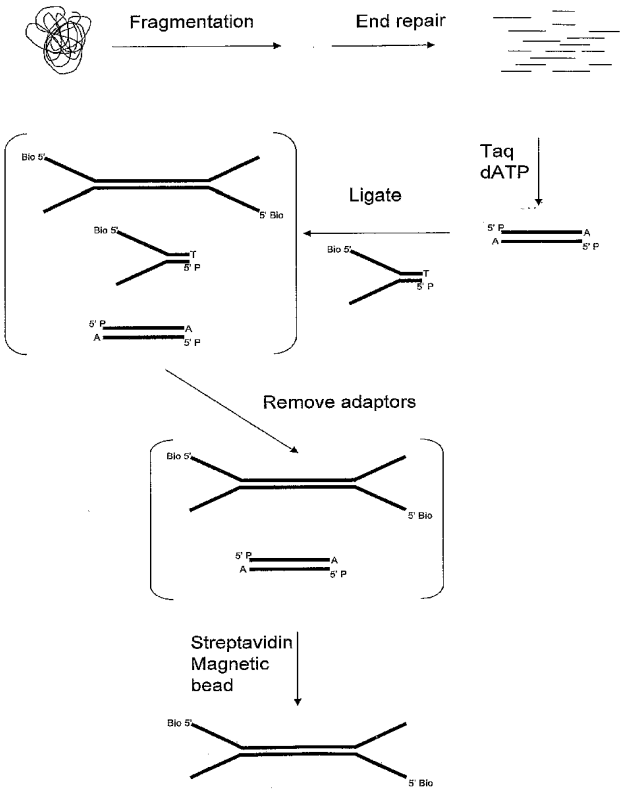
(75) Inventors/Applicants *(for US only)*: **GORMLEY, Niall, Anthony** [IE/GB]; c/o SOLEXA LIMITED, Chesterford Research Park, Little Chesterford, Saffron Walden, Essex CB10 1XL (GB). **SMITH, Geoffrey, Paul** [GB/GB]; c/o SOLEXA LIMITED, Chesterford Research Park, Little Chesterford, Saffron Walden, Essex CB10 1XL (GB). **BENTLEY, David** [GB/GB]; c/o SOLEXA LIMITED, Chesterford Research Park, Little Chesterford, Saffron Walden, Essex CB10 1XL (GB). **RIGATTI, Roberto** [IT/GB]; c/o SOLEXA LIMITED, Chesterford Research Park, Little Chesterford, Saffron Walden, Essex CB10 1XL (GB).

(74) Agents: **WHITE, Nina, Louise** et al.; BOULT WADE TENNANT, Verulam Gardens, 70 Gray's Inn Road, London WC1X 8BT (GB).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LV, LY, MA, MD, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

[Continued on next page]

(54) Title: METHOD OF PREPARING LIBRARIES OF TEMPLATE POLYNUCLEOTIDES



(57) Abstract: The invention provides a method of preparing a library of template polynucleotides which have common sequences at their 5' ends and at their 3' ends, and also the use of the library of templates in methods of solid-phase nucleic acid amplification.

WO 2007/052006 A1

IDT Exhibit 1020.3321

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report*

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

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

Method of preparing libraries of template polynucleotides

Field of the invention

The invention relates to a method of preparing a library of template polynucleotides and also the use of the library of templates in methods of solid-phase nucleic acid amplification. In particular, the invention relates to a method of preparing a library of template polynucleotides which have common sequences at their 5' ends and at their 3' ends.

Background to the invention

Molecular biology and pharmaceutical drug development now make intensive use of nucleic acid analysis. The most challenging areas are whole genome sequencing, single nucleotide polymorphism detection, screening and gene expression monitoring, which typically require analysis of large amounts of nucleic acid.

One area of technology which revolutionised the study of nucleic acids was the development of nucleic acid amplification techniques, such as the polymerase chain reaction (PCR). Amplification reactions, such as PCR, can enable the user to specifically and selectively amplify a particular target nucleic acid of interest from a complex mixture of nucleic acids. However, there is also an ongoing need for nucleic acid amplification techniques which enable simultaneous amplification of complex mixtures of templates of diverse sequence, such as genomic DNA fragments (e.g. "whole genome" amplification) or cDNA libraries, in a single amplification reaction.

PCR amplification cannot occur in the absence of annealing of forward and reverse amplification primers to

primer binding sequences in the template to be amplified under the conditions of the annealing steps of the PCR reaction, i.e. if there is insufficient complementarity between primers and template. Some prior knowledge of the sequence of the template is therefore required before one can carry out a PCR reaction to amplify a *specific* template, unless random primers are used with a consequential loss of specificity. The user must usually know the sequence of at least the primer-binding sites in the template in advance so that appropriate primers can be designed, although the remaining sequence of the template may be unknown. The need for prior knowledge of the sequence of the template increases the complexity and cost of PCR amplification of complex mixtures of templates, such as genomic DNA fragments.

WO 98/44151 and WO 00/18957 both describe methods of forming polynucleotide arrays based on "solid-phase" nucleic acid amplification, which is analogous to a polymerase chain reaction wherein the amplification products are immobilised on a solid support in order to form arrays comprised of nucleic acid clusters or "colonies". Each cluster or colony on such an array is formed from a plurality of identical immobilised polynucleotide strands and a plurality of identical immobilised complementary polynucleotide strands. The arrays so-formed are generally referred to herein as "clustered arrays" and their general features will be further understood by reference to WO 98/44151 or WO 00/18957, the contents of both documents being incorporated herein in their entirety by reference.

As aforesaid, the solid-phase amplification methods of WO 98/44151 and WO 00/18957 are essentially a form of the polymerase chain reaction carried out on a solid support.

Like any PCR reaction these methods require the use of forward and reverse amplification primers (which may be identical or different) capable of annealing to a template to be amplified. In the methods of WO 98/44151 and WO 00/18957 both primers are immobilised on the solid support at the 5′ end. Other forms of solid-phase amplification are known in which only one primer is immobilised and the other is present in free solution (Mitra, R.D and Church, G.M., Nucleic Acids Research, 1999, Vol.27, No.24).

In common with all PCR techniques, solid-phase PCR amplification requires the use of forward and reverse amplification primers which include "template-specific" nucleotide sequences which are capable of annealing to sequences in the template to be amplified, or the complement thereof, under the conditions of the annealing steps of the PCR reaction. The sequences in the template to which the primers anneal under conditions of the PCR reaction may be referred to herein as "primer-binding" sequences.

Certain embodiments of the methods described in WO 98/44151 and WO 00/18957 make use of "universal" primers to amplify templates comprising a variable template portion that it is desired to amplify flanked 5' and 3' by common or "universal" primer binding sequences. The "universal" forward and reverse primers include sequences capable of annealing to the "universal" primer binding sequences in the template construct. The variable template portion may itself be of known, unknown or partially known sequence. This approach has the advantage that it is not necessary to design a specific pair of primers for each template to be amplified; the same primers can be used for amplification of different templates provided that each template is modified by addition of the same universal primer-binding sequences

to its 5' and 3' ends. The variable template sequence can therefore be any DNA fragment of interest. An analogous approach can be used to amplify a mixture of templates, such as a plurality or library of template nucleic acid molecules (e.g. genomic DNA fragments), using a single pair of universal forward and reverse primers, provided that each template molecule in the mixture is modified by the addition of the same universal primer-binding sequences.

Such "universal primer" approaches to PCR amplification, and in particular solid-phase PCR amplification, are advantageous since they enable multiple template molecules of the same or different, known or unknown sequence to be amplified in a single amplification reaction, which may be carried out on a solid support bearing a single pair of "universal" primers. Simultaneous amplification of a mixture of templates of different sequences by PCR would otherwise require a plurality of primer pairs, each pair being complementary to each unique template in the mixture. The generation of a plurality of primer pairs for each individual template is not a viable option for complex mixtures of templates.

The addition of universal priming sequences onto the ends of templates to be amplified by PCR can be achieved by a variety of methods known to those skilled in the art. For example, a universal primer consisting of a universal sequence at its 5′ end and a degenerate sequence at its 3′ end can be used in a PCR (DOP-PCR, eg PNAS 1996 vol 93 pg 14676-14679) to amplify fragments randomly from a complex template or a complex mixture of templates. The degenerate 3′ portion of the primer anneals at random positions on DNA and can be extended to generate a copy of the template that has the universal sequence at its 5′ end.

Alternatively, adapters that contain universal priming sequences can be ligated onto the ends of templates. The adapters may be single-stranded or double-stranded. If double-stranded, they may have overhanging ends that are complementary to overhanging ends on the template molecules that have been generated with a restriction endonuclease. Alternatively, the double-stranded adapters may be blunt, in which case the templates are also blunt ended. The blunt ends of the templates may have been formed during a process to shear the DNA into fragments, or they may have been formed by an end repair reaction, as would be well known to those skilled in the art.

A single adapter or two different adapters may be used in a ligation reaction with templates. If a template has been manipulated such that its ends are the same, i.e. both are blunt or both have the same overhang, then ligation of a single compatible adapter will generate a template with that adapter on both ends. However, if two compatible adapters, adapter A and adapter B, are used, then three permutations of ligated products are formed: template with adapter A on both ends, template with adapter B on both ends, and template with adapter A on one end and adapter B on the other end. This last product is, under some circumstances, the only desired product from the ligation reaction and consequently additional purification steps are necessary following the ligation reaction to purify it from the ligation products that have the same adapter at both ends.

The current invention presented herein is a method that uses a single adapter in a ligation reaction to generate a library of template polynucleotides each of which have common, but different, universal primer sequences at their 5′ and 3′ ends. The method can be applied to preparing

simple or complex mixes of templates for amplification, for example a solid surface, using primer sequences, with no prior knowledge of the template sequences. The invention is applicable to the preparation of templates from complex samples such as whole genomes or mixtures of cDNAs, as well as mono-template applications.

Summary of the invention

In a first aspect the invention provides a method of generating a library of template polynucleotide molecules which have common sequences at their 5' ends and common sequences at their 3' ends, the method comprising:

ligating identical mismatched adapter polynucleotides to both ends of each of one or more target polynucleotide duplexes to form one or more adapter-target constructs, wherein each mismatched adapter is formed from two annealed polynucleotide strands that form a bimolecular complex comprising at least one double-stranded region and an unmatched region, and

carrying out an initial primer extension reaction in which a primer oligonucleotide is annealed to an adapter portion of each of the adapter-target constructs and extended by sequential addition of nucleotides to form extension products complementary to at least one strand of each of the adapter-target constructs,

wherein the extension products, and optionally amplification products derived therefrom, collectively provide a library of template polynucleotide molecules which have common

sequences at their 5' ends and common sequences at their 3' ends.

A second aspect of the invention relates to use of a library of template polynucleotide molecules prepared according to the method of the first aspect of the invention as a template for solid-phase PCR amplification. Thus, in a particular embodiment the invention provides a method of solid-phase nucleic acid amplification of template polynucleotide molecules which comprises:
preparing a library of template polynucleotide molecules which have common sequences at their 5' and 3' ends using the method according to the first aspect of the invention and carrying out a solid-phase nucleic acid amplification reaction wherein said template polynucleotide molecules are amplified.

A third aspect of the invention relates to use of a library of template polynucleotide molecules prepared according to the method of the first aspect of the invention as a template for whole genome amplification. Thus, in a particular embodiment the invention provides a method of whole genome amplification which comprises:
using the method according to the first aspect of the invention to prepare a library of template polynucleotide molecules which have common sequences at their 5' and 3' ends starting from a complex mixture of whole genome fragments and carrying out a whole genome amplification reaction wherein said template polynucleotide molecules are amplified.

In a fourth aspect the invention provides a kit for use in preparing a library of template polynucleotide molecules which have common sequences at their 5' and 3' ends wherein the common sequence at the 5' end of each individual template in the library is not identical and not fully complementary to the common sequence at the 3' end of said template, the kit comprising mismatched adapter polynucleotides as defined herein in relation to the first aspect of the invention and oligonucleotide primers capable of annealing to the mismatched adapter polynucleotides.

Brief description of the drawings

Figure 1 illustrates several examples of forked mismatched adapters for use in the method of the invention, specifically depicting different overhanging or blunt end structures permissible at the "ligatable" end of the adapter. Figure 1(e) schematically illustrates the sequence components of the two partially complementary strands (denoted oligo A and oligo B) which form the forked adapter when annealed. The 5′ end of oligo B is complementary (COMP) to a part of the SEQ PRIMER sequence in oligo A. Oligo A includes a single "T" nucleotide overhang at the 3′ end. The 5′ end of oligo A is phosphorylated. P represents a terminal phosphate group; X and Y represent surface capture functionalities.

Figure 2 illustrates one embodiment of the method of the invention based on use of the forked adapters illustrated in Figure 1. Figure 2(a) depicts the steps of fragmenting a complex sample such as genomic DNA to generate a plurality of target duplex fragments, ligation of the target duplex

fragments to mismatched (forked) adapters to generate adapter-template constructs and removal of unbound adapters. The forked adapter includes a biotin group at the 5' end which is not ligated to the target fragment to facilitate solid-phase capture of the adapter-target constructs, e.g. onto streptavidin magnetic beads. Figure 2(b) depicts an initial primer extension reaction in which primers are annealed to mismatched adapter regions on each strand of an adapter-target construct and extended to generate extension products complementary to each strand of the adapter-target construct. For simplicity the ligation and primer extension steps are illustrated for a single adapter-target construct.

Figure 3 illustrates an alternative embodiment of the invention in which adapter-target constructs are subjected to multiple rounds of primer annealing and extension to generate multiple single-stranded copies of each adapter-target construct. For simplicity the primer extension steps are illustrated for a single adapter-target construct.

Figure 4 illustrates a still further embodiment of the invention in which adapter-target constructs are subjected to PCR amplification to generate multiple double-stranded copies of each adapter-target construct. For simplicity PCR amplification illustrated for a single adapter-target construct.

Figure 5 illustrates an embodiment of the invention, depicting steps of fragmenting a complex sample such as genomic DNA to generate a plurality of target fragments, ligation of the target fragments to mismatched (forked) adapters to generate adapter-template constructs and

subsequent removal of unbound adapters, wherein the adapters do not include a biotin group at the 5' end. The resulting adapter-target constructs may be subjected to PCR amplification to generate multiple double-stranded copies of each adapter-target construct. For simplicity the ligation steps are illustrated for a single adapter-target construct.

Figure 6 illustrates further examples of forked mismatched adapters for use in the method of the invention, again depicting the permissible blunt or overhang formats at the "ligatable" end of the adapter. Figure 6(e) schematically illustrates the component sequences present in the two strands (denoted oligo C and oligo B) which form the adapter when annealed. P represents a terminal phosphate group; X and Y represent surface capture functionalities.

Figure 7 illustrates a still further embodiment of the invention based on use of the forked adapters illustrated in Figure 6. Figure 7(a) depicts fragmentation and ligation steps substantially similar to those illustrated in Figure 5. Figure 7(b) depicts subsequent PCR amplification using "tailed" PCR primers which have 3′ end sequences complementary to a sequence at the 5′ end of the adapter, and schematically illustrates the sequence composition of the double-stranded amplification products formed in the PCR reaction. For simplicity the ligation and PCR amplification steps are illustrated for a single adapter-target construct.

Figure 8 illustrates alternative embodiments of mismatched adapters for use in the method of the invention. P represents a terminal phosphate group; X and Y represent

surface capture functionalities; X and Z represent modifications to prevent ligation.

Figure 9a illustrates a still further embodiment of the invention based on use of the alternative adapters illustrated in Figure 8. Figure 9(a) depicts fragmentation, ligation and subsequent removal of unbound adapters. Figure 9(b) depicts annealing of identical amplification primers to a double-stranded region of the adapter on each strand of the adapter-target construct. The adapter-target constructs can be amplified by PCR using this single primer species. For simplicity the ligation steps and primer annealing are illustrated for a single adapter-target construct.

Detailed description of the invention

In its first aspect the invention provides a method of generating a library of template polynucleotide molecules which have common sequences at their 5' and 3' ends. In this context the term "common" is interpreted as meaning common to all templates in the library. As explained in further detail below, all templates within the library will contain regions of common sequence at (or proximal to) their 5′ and 3′ ends, wherein the common sequence at the 5' end of each individual template in the library is not identical and not fully complementary to the common sequence at the 3' end of said template.

The term "library" merely refers to a collection or plurality of template molecules which share common sequences at their 5′ ends and common sequences at their 3′ ends. Use of the term "library" to refer to a collection or plurality of template molecules should not be taken to imply that the templates making up the library are derived from a

particular source, or that the "library" has a particular composition. By way of example, use of the term "library" should not be taken to imply that the individual templates within the library must be of different nucleotide sequence or that the templates be related in terms of sequence and/or source.

In it's various embodiments the invention encompasses formation of so-called "monotemplate" libraries, which comprise multiple copies of a single type of template molecule, each having common sequences at their 5′ ends and their 3′ ends, as well as "complex" libraries wherein many, if not all, of the individual template molecules comprise different target sequences (as defined below), although all share common sequences at their 5′ ends and 3′ ends. Such complex template libraries may be prepared using the method of the invention starting from a complex mixture of target polynucleotides such as (but not limited to) random genomic DNA fragments, cDNA libraries etc. The invention also extends to "complex" libraries formed by mixing together several individual "monotemplate" libraries, each of which has been prepared separately using the method of the invention starting from a single type of target molecule (i.e. a monotemplate). In preferred embodiments more than 50%, or more than 60%, or more than 70%, or more than 80%, or more than 90%, or more than 95% of the individual polynucleotide templates in a complex library may comprise different target sequences, although all templates in a given library will share common sequence at their 5′ ends and common sequence at their 3′ ends.

Use of the term "template" to refer to individual polynucleotide molecules in the library merely indicates that one or both strands of the polynucleotides in the

library are capable of acting as templates for template-dependent nucleic acid polymerisation catalysed by a polymerase. Use of this term should not be taken as limiting the scope of the invention to libraries of polynucleotides which are actually used as templates in a subsequent enzyme-catalysed polymerisation reaction.

The library is formed by ligating identical adapter polynucleotide molecules ("mismatched adapters", the general features of which are defined below) to the 5' and 3' ends of one or more target polynucleotide duplexes (which may be of known, partially known or unknown sequence) to form adapter-target constructs and then carrying out an initial primer extension reaction in which extension products complementary to both strands of each individual adapter-target construct are formed. The resulting primer extension products, and optionally amplified copies thereof, collectively provide a library of template polynucleotides.

Each strand of each template molecule in the library formed in the primer extension reaction will therefore have the following structure, when viewed as a single strand:

5'-[common sequence I]-[target sequence]-[common sequence II]-3'

wherein "common sequence I" represents a sequence derived from copying a first strand of the mismatched adapter and is common to all template molecules in the library generated in the initial primer extension reaction;
"target" represents a sequence derived from one strand of the target polynucleotide duplex and may be different in different individual template molecules within the library; and

"common sequence II" represents a sequence derived from copying of a second strand of the mismatched adapter and is also common to all template molecules in the library generated in the initial primer extension reaction.

Since "common sequence I" and "common sequence II" are common to all template strands in the library they may include "universal" primer-binding sequences, enabling all templates in the library to be ultimately amplified in a solid-phase PCR procedure using universal primers.

It is a key feature of the invention, however, that the common 5' and 3' end sequences denoted "common sequence I" and "common sequence II" are not *fully* complementary to each other, meaning that each individual template strand can contain different (and non-complementary) universal primer sequences at its 5' and 3' ends.

It is generally advantageous for complex libraries of templates to be amplified by PCR (e.g. whole genome amplification) whether performed in solution or on a solid support, to include regions of "different" sequence at their 5' and 3' ends, which are nevertheless common to all template molecules in the library, especially if the amplification products are to be ultimately sequenced. For example, the presence of common unique sequence at one end only of each template in the library can provide a binding site for a sequencing primer, enabling one strand of each template in the amplified form of the library to be sequenced in a single sequencing reaction using a single type of sequencing primer.

Typically "common sequence I" and "common sequence II" will consist of no more than 100, or no more than 50, or no more than 40 consecutive nucleotides at the 5' and 3' ends, respectively, of each strand of each template

polynucleotide. The precise length of the two sequences may or may not be identical. The nucleotide sequences of "common sequence I" and "common sequence II" in the template polynucleotides will be determined in part by the sequences of the adapter strands ligated to the target polynucleotides and in part by the sequence of the primer used in the initial primer extension reaction, and any subsequent rounds of nucleic acid amplification.

In embodiments wherein the initial primer extension product is subjected to further amplification by conventional PCR, then the products of the amplification reaction will be double-stranded polynucleotides, one strand of which has the structure:

5'-[common sequence I]-[target sequence]-[common sequence II]-3'

It will be appreciated that "common sequence II" in the amplification products may differ somewhat to the "common sequence II" present in the products of the initial primer extension reaction, since the former will be determined in part by the sequence of the PCR primer used to prime synthesis of a polynucleotide strand complementary to the initial primer extension product, whereas the latter will be determined solely by copying of the adapter sequences at the 3′ ends of the adapter-template constructs in the initial primer extension. Nevertheless, since the PCR primer is designed to anneal to a sequence in the initial extension products which is complementary to the 3′ adapter, the two forms of "common sequence II" will contain identical sequence, at least at the 3′ end. Additional sequence may be included at the 5′ end of "common sequence II" in the

amplified products, for example by the use of "tailed" PCR primers, as described in detail below. In other embodiments the common sequences present in the amplification products may actually be shorter than the common sequences including in the adapters originally ligated to the target.

The precise nucleotide sequences of the common regions of the template molecules in the library are generally not material to the invention and may be selected by the user. The common sequences must at least comprise "primer-binding" sequences which enable specific annealing of amplification primers when the templates are in use in a solid-phase amplification reaction. The primer-binding sequences are thus determined by the sequence of the primers to be ultimately used for solid-phase amplification. The sequence of these primers in turn is advantageously selected to avoid or minimise binding of the primers to the target portions of the templates within the library under the conditions of the amplification reaction, but is otherwise not particularly limited. By way of example, if the target portions of the templates are derived from human genomic DNA, then the sequences of the primers to be used in solid phase amplification should ideally be selected to minimise non-specific binding to any human genomic sequence.

The adapter polynucleotides used in the method of the invention are referred to herein as "mismatched" adapters because, as will be explained in detail herein, it is essential that the adapters include a region of sequence mismatch, i.e. they must not be formed by annealing of fully complementary polynucleotide strands.

Mismatched adapters for use in the invention are formed by annealing of two partially complementary polynucleotide strands so as to provide, when the two strands are annealed,

at least one double-stranded region and at least one unmatched region.

The "double-stranded region" of the adapter is a short double-stranded region, typically comprising 5 or more consecutive base pairs, formed by annealing of the two partially complementary polynucleotide strands. This term simply refers to a double-stranded region of nucleic acid in which the two strands are annealed and does not imply any particular structural conformation.

Generally it is advantageous for the double-stranded region to be as short as possible without loss of function. By "function" in this context is meant that the double-stranded region form a stable duplex under standard reaction conditions for an enzyme-catalysed nucleic acid ligation reaction, which will be well known to the skilled reader (e.g. incubation at a temperature in the range of from 4°C to 25°C in a ligation buffer appropriate for the enzyme), such that the two strands forming the adapter remain partially annealed during ligation of the adapter to a target molecule. It is not absolutely necessary for the double-stranded region to be stable under the conditions typically used in the annealing steps of primer extension or PCR reactions.

Since identical adapters are ligated to both ends of each template molecule, the target sequence in each adapter-target construct will be flanked by complementary sequences derived from the double-stranded region of the adapters. The longer the double-stranded region, and hence the complementary sequences derived therefrom in the adapter-target constructs, the greater the possibility that the adapter-target construct is able to fold back and base-pair to itself in these regions of internal self-complementarity

under the annealing conditions used in primer extension and/or PCR. Generally it is preferred for the double-stranded region to be 20 or less, 15 or less, or 10 or less base pairs in length in order to reduce this effect. The stability of the double-stranded region may be increased, and hence its length potentially reduced, by the inclusion of non-natural nucleotides which exhibit stronger base-pairing than standard Watson-Crick base pairs.

It is preferred, but not absolutely essential, for the two strands of the adapter to be 100% complementary in the double-stranded region. It will be appreciated that one or more nucleotide mismatches may be tolerated within the double-stranded region, provided that the two strands are capable of forming a stable duplex under standard ligation conditions.

Adapters for use in the invention will generally include a double-stranded region adjacent to the "ligatable" end of the adapter, i.e. the end that is joined to a target polynucleotide in the ligation reaction. The ligatable end of the adapter may be blunt or, in other embodiments, short 5′ or 3′ overhangs of one or more nucleotides may be present to facilitate/promote ligation. The 5′ terminal nucleotide at the ligatable end of the adapter should be phosphorylated to enable phosphodiester linkage to a 3′ hydroxyl group on the target polynucleotide.

The term "unmatched region" refers to a region of the adapter wherein the sequences of the two polynucleotide strands forming the adapter exhibit a degree of non-complementarity such that the two strands are not capable of annealing to each other under standard annealing conditions for a primer extension or PCR reaction. The two strands in the unmatched region may exhibit some degree of annealing

under standard reaction conditions for a enzyme-catalysed ligation reaction, provided that the two strands revert to single stranded form under annealing conditions.

The conditions encountered during the annealing steps of a PCR reaction will be generally known to one skilled in the art, although the precise annealing conditions will vary from reaction to reaction (see Sambrook *et al.*, 2001, *Molecular Cloning, A Laboratory Manual*, 3rd Ed, Cold Spring Harbor Laboratory Press, Cold Spring Harbor Laboratory Press, NY; Current Protocols, eds Ausubel *et al.*). Typically such conditions may comprise, but are not limited to, (following a denaturing step at a temperature of about 94°C for about one minute) exposure to a temperature in the range of from 40°C to 72°C (preferably 50-68°C) for a period of about 1 minute in standard PCR reaction buffer.

Different annealing conditions may be used for a single primer extension reaction not forming part of a PCR reaction (again see Sambrook *et al.*, 2001, *Molecular Cloning, A Laboratory Manual*, 3rd Ed, Cold Spring Harbor Laboratory Press, Cold Spring Harbor Laboratory Press, NY; Current Protocols, eds Ausubel *et al.*). Conditions for primer annealing in a single primer extension include, for example, exposure to a temperature in the range of from 30 to 37°C in standard primer extension buffer. It will be appreciated that different enzymes, and hence different reaction buffers, may be used for a single primer extension reaction as opposed to a PCR reaction. There is no requirement to use a thermostable polymerase for a single primer extension reaction.

It is to be understood that the "unmatched region" is provided by different portions of the same two polynucleotide strands which form the double-stranded

region(s). However, the portions of the two strands forming the unmatched region are not annealed under conditions in which other portions of the same two strands are annealed to form one or more double-stranded regions. For avoidance of doubt it is to be understood that a single-stranded or single base overhang at the 5' or 3' end of a polynucleotide duplex does not constitute an "unmatched region" in the context of this invention.

The portions of the two strands forming the double-stranded region typically comprise at least 10, or at least 15, or at least 20 consecutive nucleotides on each strand. The lower limit on the length of the unmatched region will typically be determined by function, for example the need to provide a suitable sequence for binding of a primer for primer extension, PCR and/or sequencing. Theoretically there is no upper limit on the length of the unmatched region, except that it general it is advantageous to minimise the overall length of the adapter, for example in order to facilitate separation of unbound adapters from adapter-target constructs following the ligation step. Therefore, it is preferred that the unmatched region should be less than 50, or less than 40, or less than 30, or less than 25 consecutive nucleotides in length on each strand.

The overall length of the two strands forming the adapter will typically in the range of from 25 to 100 nucleotides, more typically from 30 to 55 nucleotides.

The portions of the two strands forming the unmatched region should preferably be of similar length, although this is not absolutely essential, provided that the length of each portion is sufficient to fulfil its desired function (e.g. primer binding). The inventors' have shown by experiment that the portions of the two strands forming the

unmatched region may differ by up to 25 nucleotides without unduly affecting adapter function.

Most preferably the portions of the two polynucleotide strands forming the unmatched region will be completely mismatched, or 100% non-complementary. However, skilled readers will be appreciate that some sequence "matches", i.e. a lesser degree of non-complementarity may be tolerated in this region without affecting function to a material extent. As aforesaid, the extent of sequence mismatching or non-complementarity must be such that the two strands in the unmatched region remain in single-stranded form under annealing conditions as defined above.

The precise nucleotide sequence of the adapters is generally not material to the invention and may be selected by the user such that the desired sequence elements are ultimately included in the common sequences of the library of templates derived from the adapters, for example to provide binding sites for particular sets of universal amplification primers and/or sequencing primers. Additional sequence elements may be included, for example to provide binding sites for sequencing primers which will ultimately be used in sequencing of template molecules in the library, or products derived from amplification of the template library, for example on a solid support. The adapters may further include "tag" sequences, which can be used to tag or mark template molecules derived from a particular source. The general features and use of such tag sequences is described in the applicant's pending application published as WO 05/068656.

Although the precise nucleotide sequence of the adapter is generally non-limiting to the invention, the sequences of the individual strands in the unmatched region should be

such that neither individual strand exhibits any internal self-complementarity which could lead to self-annealing, formation of hairpin structures etc. under standard annealing conditions. Self-annealing of a strand in the unmatched region is to be avoided as it may prevent or reduce specific binding of an amplification primer to this strand.

The mismatched adapters are preferably formed from two strands of DNA, but may include mixtures of natural and non-natural nucleotides (e.g. one or more ribonucleotides) linked by a mixture of phosphodiester and non-phosphodiester backbone linkages. Other non-nucleotide modifications may be included such as, for example, biotin moieties, blocking groups and capture moieties for attachment to a solid surface, as discussed in further detail below.

The one or more "target polynucleotide duplexes" to which the adapters are ligated may be any polynucleotide molecules that it is desired to amplify by solid-phase PCR, generally with a view to sequencing. The target polynucleotide duplexes may originate in double-stranded DNA form (e.g. genomic DNA fragments) or may have originated in single-stranded form, as DNA or RNA, and been converted to dsDNA form prior to ligation. By way of example, mRNA molecules may be copied into double-stranded cDNAs suitable for use in the method of the invention using standard techniques well known in the art. The precise sequence of the target molecules is generally not material to the invention, and may be known or unknown. Modified DNA molecules including non-natural nucleotides and/or non-natural backbone linkages could serve as the target, provided that the modifications do not preclude adapter ligation and/or copying in a primer extension reaction.

Although the method could in theory be applied to a single target duplex (i.e. one individual double-stranded molecule), it is preferred to use a mixture or plurality of target polynucleotide duplexes. The method of the invention may be applied to multiple copies of the same target molecule (so-called monotemplate applications) or to a mixture of different target molecules which differ from each other with respect to nucleotide sequence over all or a part of their length, e.g. a complex mixture of templates. The method may be applied to a plurality of target molecules derived from a common source, for example a library of genomic DNA fragments derived from a particular individual. In a preferred embodiment the target polynucleotides will comprise random fragments of human genomic DNA. The fragments may be derived from a whole genome or from part of a genome (e.g. a single chromosome or sub-fraction thereof), and from one individual or several individuals. The DNA target molecules may be treated chemically or enzymatically either prior to, or subsequent to the ligation of the adaptor sequences. Techniques for fragmentation of genomic DNA include, for example, enzymatic digestion or mechanical shearing.

"Ligation" of adapters to 5' and 3' ends of each target polynucleotide involves joining of the two polynucleotide strands of the adapter to double-stranded target polynucleotide such that covalent linkages are formed between both strands of the two double-stranded molecules. In this context "joining" means covalent linkage of two polynucleotide strands which were not previously covalently linked. Preferably such "joining" will take place by formation of a phosphodiester linkage between the two polynucleotide strands but other means of covalent linkage

(e.g. non-phosphodiester backbone linkages) may be used. However, it is an essential requirement that the covalent linkages formed in the ligation reactions allow for read-through of a polymerase, such that the resultant construct can be copied in a primer extension reaction using primers which binding to sequences in the regions of the adapter-target construct that are derived from the adapter molecules.

The ligation reactions will preferably be enzyme-catalysed. The nature of the ligase enzyme used for enzymatic ligation is not particularly limited. Non-enzymatic ligation techniques (e.g. chemical ligation) may also be used, provided that the non-enzymatic ligation leads to the formation of a covalent linkage which allows read-through of a polymerase, such that the resultant construct can be copied in a primer extension reaction.

The desired products of the ligation reaction are adapter-target constructs in which identical adapters are ligated at both ends of each target polynucleotide, given the structure adapter-target-adapter. Conditions of the ligation reaction should therefore be optimised to maximise the formation of this product, in preference to targets having an adapter at one end only.

The products of the ligation reaction may be subjected to purification steps in order to remove unbound adapter molecules before the adapter-target constructs are processed further. Any suitable technique may be used to remove excess unbound adapters, preferred examples of which will be described in further detail below.

Adapter-target constructs formed in the ligation reaction are then subject to an initial primer extension reaction in which a primer oligonucleotide is annealed to an

adapter portion of each of the adapter-target constructs and extended by sequential addition of nucleotides to the free 3' hydroxyl end of the primer to form extension products complementary to at least one strand of each of the adapter-target constructs.

The term "initial" primer extension reaction refers to a primer extension reaction in which primers are annealed directly to the adapter-target constructs, as opposed to either complementary strands formed by primer extension using the adapter-target construct as a template or amplified copies of the adapter-target construct. It is a key feature of the method of the invention that the initial primer extension reaction is carried out using a "universal" primer which binds *specifically* to a cognate sequence within an adapter portion of the adapter-target construct, and is not carried out using a target-specific primer or a mixture of random primers. The use of an adapter-specific primer for the initial primer extension reaction is key to formation of a library of templates which have common sequence at the 5′ *and* common sequence at the 3′ end.

The primers used for the initial primer extension reaction will be capable of annealing to each individual strand of adapter-target constructs having adapters ligated at both ends, and can be extended so as to obtain two separate primer extension products, one complementary to each strand of the construct. Thus, in the most preferred embodiment the initial primer extension reaction will result in formation of primer extension products complementary to each strand of each adapter-target

In a preferred embodiment the primer used in the initial primer extension reaction will anneal to a primer-

binding sequence (in one strand) in the unmatched region of the adapter.

The term "annealing" as used in this context refers to sequence-specific binding/hybridisation of the primer to a primer-binding sequence in an adapter region of the adapter-target construct under the conditions to be used for the primer annealing step of the initial primer extension reaction.

The products of the primer extension reaction may be subjected to standard denaturing conditions in order to separate the extension products from strands of the adapter-target constructs. Optionally the strands of the adapter-target constructs may be removed at this stage. The extension products (with or without the original strands of the adapter-target constructs) collectively form a library of template polynucleotides which can be used as templates for solid-phase PCR.

If desired, the initial primer extension reaction may be repeated one or more times, through rounds of primer annealing, extension and denaturation, in order to form multiple copies of the same extension products complementary to the adapter-target constructs.

In other embodiments the initial extension products may be amplified by convention solution-phase PCR, as described in further detail below. The products of such further PCR amplification may be collected to form a library of templates comprising "amplification products derived from" the initial primer extension products. In a preferred embodiment both primers used for further PCR amplification will anneal to different primer-binding sequences on opposite strands in the unmatched region of the adapter. Other embodiments may, however, be based on the use of a

single type of amplification primer which anneals to a primer-binding sequence in the double-stranded region of the adapter. In embodiments of the method based on PCR amplification the "initial" primer extension reaction occurs in the first cycle of PCR.

Inclusion of the initial primer extension step (and optionally further rounds of PCR amplification) to form complementary copies of the adapter-target constructs (prior to whole genome or solid-phase PCR) is advantageous, for several reasons. Firstly, inclusion of the primer extension step, and subsequent PCR amplification, acts as an enrichment step to select for adapter-target constructs with adapters ligated at both ends. Only target constructs with adapters ligated at both ends provide effective templates for whole genome or solid-phase PCR using common or universal primers specific for primer-binding sequences in the adapters, hence it is advantageous to produce a template library comprising only double-ligated targets prior to solid-phase or whole genome amplification.

Secondly, inclusion of the initial primer extension step, and subsequent PCR amplification, permits the length of the common sequences at the 5' and 3' ends of the target to be increased prior to solid-phase or whole genome PCR. As outlined above, it is generally advantageous for the length of the adapter molecules to be kept as short as possible, to maximise the efficiency of ligation and subsequent removal of unbound adapters. However, for the purposes of whole genome or solid-phase PCR it may be an advantage to have longer sequences common or "universal" sequences at the 5' and 3' ends of the templates to be amplified. Inclusion of the primer extension (and subsequent amplification) steps means that the length of the

common sequences at one (or both) ends of the polynucleotides in the template library can be increased after ligation by inclusion of additional sequence at the 5' ends of the primers used for primer extension (and subsequent amplification). The use of such "tailed" primers is described in further detail below.

Various non-limiting specific embodiments of the method of the invention will now be described in further detail with reference to the accompanying drawings. Features described as being preferred in relation to one specific embodiment of the invention apply *mutatis mutandis* to other specific embodiments of the invention unless stated otherwise.

Figure 1 illustrates several embodiments of a particular type of mismatched adapter for use in the method of the invention. The adapter is formed by annealing two single-stranded oligonucleotides, herein referred to as "oligo A" and "oligo B". Oligo A and oligo B may be prepared by conventional automated oligonucleotide synthesis techniques in routine use in the art. The oligonucleotides are partially complementary such that the 3′ end of oligo A is complementary to the 5′ end of oligo B. The 5′ end of oligo A and the 3′ end of oligo B are not complementary to each other. When the two strands are annealed, the resulting structure is double stranded at one end (the double-stranded region) and single stranded at the other end (the unmatched region) and is referred to herein as a "forked adapter" (Fig. 1a). The double-stranded region of the forked adapter may be blunt-ended (Fig. 1b) or it may have an overhang. In the latter case, the overhang may be a 3′ overhang (Fig. 1c)

or a 5′ overhang (Fig. 1d), and may comprise a single nucleotide or more than one nucleotide.

The 5′ end of the double-stranded part of the forked adapter is phosphorylated, i.e. the 5′ end of oligo B (Fig. 1a-d). The presence of the 5' phosphosphate group identifies this as the "ligatable" end of the adapter. The 5′ end of oligo A may be biotinylated or bear another functionality (represented by X) that enables it to be captured on a surface, such as a bead. Such alternative functionalities other than biotin are known to those skilled in the art. The 3′ end of oligo B may also be biotinylated or bear another functionality (represented by Y) that enables it to be captured on a surface (Fig. 1d).

The phosphodiester bonds that comprise the back-bone of the oligonucleotides may be replaced with non-enzymatically cleavable bonds such as phosphorothioate bonds. Preferably only the last, or last and penultimate, phosphodiester bonds at both the 3′ and 5′ ends of the oligonucleotides will be substituted with phosphorothioate bonds. In the most preferred embodiment of the invention, oligo A contains a biotin group on its 5′ end, oligo B is phosphorylated at its 5′ end and the double-stranded portion of the duplex contains a single base 3′ overhang comprising a 'T' nucleotide. Oligo A consists of two sequences: a sequence at the 5′ end which is identical to that of a universal primer to be used for PCR amplification, referred to herein as "PRIMER 1" sequence, and at its 3′ end a sequence identical to that of a universal sequencing primer, herein referred to herein as "SEQ PRIMER" sequence, plus an additional 'T' nucleotide on the 3′ end. Oligo B also consists of two sequences: a sequence at its 5′ end that is complementary to only part of the 3′ end of the SEQ PRIMER sequence in Oligo

A, excluding the 'T' overhang of Oligo A, and a sequence complementary to that of a universal PCR amplification primer, herein referred to as "comp- PRIMER 2" at its 3' end (Fig. 1e).

Figure 2 illustrates one embodiment of the method of the invention based on use of the forked adapters illustrated in Figure 1. A mixture of target DNA molecules of different sequence may be prepared by mixing a number, greater than one, of individual DNA molecules. In the preferred procedure, genomic DNA is fragmented into small molecules, preferably less than 1000 base pairs in length. Fragmentation of DNA may be achieved by a number of methods including: enzymatic digestion, chemical cleavage, sonication, nebulisation, or hydroshearing, preferably nebulisation.

Fragmented DNA may be made blunt-ended by a number of methods known to those skilled in the art. In the preferred method, the ends of the fragmented DNA are end repaired with T4 DNA polymerase and Klenow polymerase, a procedure well known to those skilled in the art, and then phosphorylated with a polynucleotide kinase enzyme. A single 'A' deoxynucleotide is then added to both 3' ends of the DNA molecules using Taq polymerase enzyme, producing a one-base 3' overhang that is complementary to the one-base 3' 'T' overhang on the double-stranded end of the forked adapter.

A ligation reaction between the forked adapter and the DNA fragments is then performed using a suitable ligase enzyme (e.g. T4 DNA ligase) which joins two copies of the adapter to each DNA fragment, one at either end, to form adapter-target constructs. The products of this reaction can be purified from unligated adapter by a number of means, including size-inclusion chromatography, preferably by

electrophoresis through an agarose gel slab followed by excision of a portion of the agarose that contains the DNA greater in size that the size of the adapter (Fig. 2a).

After the excess adapter has been removed, unligated target DNA remains in addition to ligated adapter-target constructs and this can be removed by selectively capturing only those target DNA molecules that have adapter attached. The presence of a biotin group on the 5' end of Oligo A of the adapter enables any target DNA ligated to the adapter to be captured on a surface coated with streptavidin, a protein that selectively and tightly binds biotin. Streptavidin can be coated onto a surface by means known to those skilled in the art. In the preferred method, commercially available magnetic beads that are coated in streptavidin can be used to capture ligated adapter-target constructs. The application of a magnet to the side of a tube containing these beads immobilises them such that they can be washed free of the unligated target DNA molecules (Fig. 2a).

An oligonucleotide, herein referred to as PRIMER 2, which hybridises to the "comp- PRIMER 2" sequence on the oligo B strand of the adapter-target constructs can be used in an initial primer extension reaction to generate a complementary copy of the adapter-target strand attached to the bead. The resulting primer extension product forms a double-stranded duplex with its complementary adapter-target strand attached to the bead and it can then be isolated and purified from adapter-target strand on the bead by denaturation (Fig. 2b).

There are several standard methods for separating the strand of a DNA duplex by denaturation, including thermal denaturation, or preferably chemical denaturation in 100 mM sodium hydroxide solution. The pH of a solution of single-

stranded DNA in a sodium hydroxide collected from the supernatant of a suspension of magnetic beads can be neutralised by adjusting with an appropriate solution of acid, or preferably by buffer-exchange through a size-exclusion chromatography column pre-equilibrated in a buffered solution. The resulting solution contains a library of single-stranded DNA template molecules all of which comprise in order: 5' PRIMER 2 sequence, target DNA, the complement of SEQ PRIMER sequence, then the complement of PRIMER 1 sequence. This template library can then be used for solid-phase PCR amplification using immobilised PRIMER 1 and PRIMER 2 oligonucleotides. It will be appreciated, however, that the utility of the library is not limited to solid-phase PCR but extends to any type of PCR amplification, including whole-genome amplification performed entirely in solution.

Figure 3 illustrates an alternative embodiment of the invention in which adapter-target constructs prepared as described above with reference to Figure 2 are subjected to multiple rounds of primer annealing and extension to generate multiple single-stranded copies of each adapter-target construct. In this embodiment of the invention, the initial primer extension reaction on the bead-immobilised adapter-template molecules with PRIMER 2 is in effect replaced with an asymmetric PCR amplification with the PRIMER 2 oligonucleotide (Fig. 3), this being equivalent to multiple rounds of the same primer extension reaction. In this embodiment, multiple single-stranded copies of the bead-immobilised strands are generated in the supernatant of the bead suspension due to PCR thermocycling, hence a separate denaturation step is not necessary to recover the newly synthesised complementary copies of the bead-

immobilised adapter-target strands; the copies can be purified from the supernatant by standard methods, known to those skilled in the art.

In another embodiment of the invention, illustrated in Figure 4, the initial primer extension reaction on the bead-immobilised adapter-target constructs with PRIMER 2 forms part of a standard (symmetric) PCR amplification with the PRIMER 2 and PRIMER 1 oligonucleotides. In this embodiment, multiple double-stranded copies of the bead-immobilised strands are generated in the supernatant of the bead suspension due to PCR thermocycling, hence a separate denaturation step is not necessary to recover the newly synthesised complementary copies of the bead-immobilised adapter-target strands; the copies can be purified from the supernatant by standard methods, known to those skilled in the art.

In another embodiment of the invention, illustrated in Figure 5, the forked adapter does not contain a biotin group at the 5′ end of the Oligo A strand. In this embodiment, fragmented DNA may be made blunt-ended by a number of methods know to those skilled in the art. In the preferred method, the ends of the fragmented are end repaired with T4 DNA polymerase and Klenow polymerase, and then phosphorylated with polynucleotide kinase enzyme. A single 'A' deoxynucleotide is then added to both 3′ ends of the DNA molecules with Taq polymerase enzyme, producing a one-base 3′ overhang that is complementary to the one-base 3′ 'T' overhang on the double-stranded "ligatable" end of the forked adapter. A ligation reaction between the forked adapter and the DNA fragments is then performed, e.g. using T4 DNA ligase enzyme, which joins two copies of the adapter to each DNA template molecule, one at either end.

The products of the ligation reaction can be purified from unligated adapter by a number of means, including size-inclusion chromatography, preferably by electrophoresis through an agarose gel slab followed by excision of a portion of the agarose that contains the DNA greater in size that the size of the adapter. An aliquot of the purified DNA is then used in a PCR amplification with the PRIMER 2 and PRIMER 1 oligonucleotides (Fig. 5). The first PCR cycle will involve an initial primer extension reaction with primer 2 (not illustrated). The primers selectively amplify those template DNA molecules that have adapters ligated on both ends. The product of the reaction is a library of double-stranded template molecules, each of which comprise in order on one of the duplex strands: 5′ PRIMER 2 sequence, target DNA, the complement of SEQ PRIMER sequence, then the complement of PRIMER 1 sequence. This library can then be used on a solid-phase PCR platform that contains immobilised PRIMER 1 and PRIMER 2 oligonucleotides.

Figure 6 illustrates further examples of forked mismatched adapters for use in the method of the invention. In this embodiment the forked adapter is formed by annealing two single-stranded oligonucleotides, herein referred to as "oligo C" and "oligo B". The oligonucleotides are partially complementary such that the 3′ end of oligo C is complementary to the 5′ end of oligo B. The 5′ end of oligo C and the 3′ end of oligo B are not complementary to each other. When the two oligos are annealed the resulting structure is double-stranded at one end (double-stranded region) and single-stranded at the other end (unmatched region) (Figure 6a). The double-stranded region of the forked adapter may be blunt-ended (Fig. 6d) or it may have an overhang. In the latter case, the overhang may be a 3′

overhang (Fig. 6c) or a 5'overhang (Fig. 6b), and may comprise a single base or more than one base.

The 5' end of the double-stranded region of the forked adapter is phosphorylated i.e. the 5' end of 'oligo B' (Figure 6a-d) to provide a "ligatable" end. The 5' end of oligo C may be biotinylated or bear another functionality (X) that enables it to be captured on a surface, such as a bead. The 3' end of oligo B may also be biotinylated or bear another functionality (Y) that enables it to be captured on a surface (Fig. 6d).

The phosphodiester bonds that comprise the back-bone of the oligonucleotides may be replaced with non-enzymatically cleavable bonds such as phosphorothioate bonds. Preferably only the last, or last and penultimate, phosphodiester bonds at both the 3' and 5' ends of the oligonucleotides will be substituted with phosphorothioate bonds. Oligo C consists of only one sequence: a sequence identical to that of a universal sequencing primer denoted "SEQ PRIMER" (or identical to part of the 3' end of the "SEQ PRIMER" sequence), plus an additional 'T' nucleotide on the 3' end. Oligo B consists of two sequences: a sequence at its 5' end that is complementary to only part of the 3' end of the SEQ PRIMER sequence in Oligo C, excluding the 'T' overhang of 'Oligo C', and a sequence at its 3' end which is complementary to that of a universal PCR amplification primer, herein referred to as the "comp-PRIMER 2" sequence, (Fig. 6e).

Figure 7 illustrates a further embodiment of the invention based on use of the forked adapters illustrated in Figure 6. In this embodiment, adapter-target constructs are prepared substantially as described above with reference to

Figure 5, except that the adapters illustrated in Figure 6 are used (Fig. 7a).

An aliquot of the purified adapter-target constructs is used in a standard solution-phase PCR amplification with "tailed" primer oligonucleotides. Tailed primers are primers that only hybridize via their 3′ end to a target sequence, leaving a 5′ non-hybridised tail. The length and precise sequence of the non-hybridised tail is non-limiting, but may typically comprise from 10 to 50 nucleotides, more typically from 15 to 30 nucleotides. When used in amplifications by PCR, the initial round of PCR amplification (i.e. the first and second primer extension reactions) rely on binding of the 3' ends of the tailed primers to cognate primer-binding sequences in the adapter regions of the adapter-target constructs. The 5′ non-hybridising tails derived from the tailed primers act as templates in subsequent PCR cycles and are therefore copied into the resultant double-stranded PCR products.

In the current embodiment, either one or both of the primers used in the amplification reaction can be "tailed" primers. In one embodiment, the primers used are denoted PRIMER 3 and PRIMER 4, where PRIMER 3 consists of a 5′ tail sequence, and a 3′ sequence that is complementary to the "comp PRIMER 2" sequence in the forked adapter; and PRIMER 4 consists of a 5′ tail sequence, and a 3′ sequence that is identical to the 5′ end of the SEQ PRIMER sequence present in the unmatched region of the forked adapter. Following amplification by PCR, the tail sequences are incorporated into the copies of the adapter-target DNA construct.

In one embodiment of the invention, the tail sequences on PRIMER 3 and PRIMER 4 are non-identical sequences. The tail sequence of PRIMER 3 and the tail sequence of PRIMER 4

can then be used to form the sequence of surface-immobilised primers to be used on a solid-phase DNA amplification platform (Fig. 7b).

In another embodiment of the invention, the tail sequences on PRIMER 3 and PRIMER 4 are identical sequences. The products of the solution-phase PCR will thus have the same sequence at their ends: the common tail sequence of PRIMER 3 and PRIMER 4. This common tail sequence can then be used to form the sequence of a single surface-immobilised primer on a solid-phase DNA amplification platform. Surface amplification of the library of templates may thus be performed using a single PCR primer immobilised on the surface.

Figure 8 illustrates alternative embodiments of mismatched adapters for use in the method of the invention. These "modified" forked adapters may be designed to enable solid-phase amplification of templates using a single surface bound primer. The adapter is formed by annealing two single-stranded oligonucleotides, herein referred to as "oligo D" and "oligo E". The oligonucleotides are partially complementary such that the 3′ end of oligo D is complementary to the 5′ end of oligo E, and the 5′ end of oligo D is complementary to the 3′ end of oligo E, however the central portions of oligo D and oligo E are non-complementary. When the oligo D and oligo E are annealed the resulting structure is double stranded at both ends (double-stranded regions) and single stranded in the middle (unmatched region) and is referred to herein as the "modified Forked adapter" (Fig. 8a).

One end of the modified forked adapter is modified to prevent ligation of a DNA molecule to this end. Such modifications are known to those skilled in the art and may

include, for example, the presence of a 5′ or 3′ overhang. The other "ligatable" end may be blunt-ended (Fig. 8d) or may have an overhang. In the latter case, the overhang may be a 3′ overhang (Fig. 8c) or a 5′overhang (Fig. 8b), and may comprise a single base or more than one base. The 5′ strand of the ligatable end is phosphorylated i.e. the 5′ end of oligo E (Figure 8a-d). The 5′ end of oligo D may be biotinylated or bear another functionality that enables it to be captured on a surface, such as a bead. The 3′ end of oligo E may be biotinylated or bear another functionality that enables it to be captured on a surface (Fig. 8d). The modifications to prevent ligation (Z,W) may be the same as or different to the surface capture functionalities (X,Y).

The phosphodiester bonds that comprise the back-bone of the oligonucleotides may be replaced with non-enzymatically cleavable bonds such as phosphorothioate bonds. Preferably only the last, or last and penultimate, phosphodiester bonds at both the 3′ and 5′ ends of the oligonucleotides will be substituted with phosphorothioate bonds.

In the preferred embodiment of the invention, oligo E is phosphorylated at its 5′ end and the 3′ end of oligo D contains a single base 3′ overhang comprising a "T" nucleotide. Oligo D consists of two sequences: a sequence at its 5′ end which is identical to that of a universal PCR amplification primer, referred to herein as "PRIMER 5" sequence, next to a sequence identical to that of a universal sequencing primer denoted "SEQ PRIMER" sequence plus the additional "T" nucleotide on the 3′ end. Oligo E consists of three sequences: a sequence at its 5′ end that is complementary to only part of the 3′ end of the SEQ PRIMER sequence in Oligo D, excluding the 'T' overhang of Oligo D, a central sequence non-complementary to any part of

Oligo D, and a 3′ end that is complementary to the "PRIMER 5" sequence of Oligo D (Fig. 8e).

Figure 9 illustrates a still further embodiment of the invention based on use of the alternative adapters illustrated in Fig. 8. In this embodiment adapter-target constructs may be prepared substantially as described above in relation to Figure 5, except that the modified forked adapters illustrated in Figure 8 are used. An aliquot of the purified adapter-target constructs is used in a solution-phase PCR amplification using PRIMER 5 oligonucleotide to selectively amplify those ligation products that have the modified adapter on both ends (Fig. 9b). The product of the solution-phase PCR can then be purified and amplified on a solid-phase PCR platform with a single immobilised primer, e.g. PRIMER 5. Inclusion of the mismatch sequence in oligo E ensures that all products of this solid-phase amplification will contain common sequencing primer binding sequences on one strand only, enabling sequencing using a universal sequencing primer which anneals to this common sequence.

Use of the template library

Template libraries prepared according to the method of the invention may be used in essentially any method of nucleic acid analysis which requires further amplification of the templates and/or sequencing of the templates or amplification products thereof. Exemplary uses of the template libraries include, but are not limited to, providing templates for whole genome amplification and also solid-phase PCR amplification (of either monotemplate or complex template libraries). A particularly preferred use

is in whole-genome amplification carried out on a solid-support.

Whole-genome amplification

Template libraries prepared according to the method of the invention starting from a complex mixture of genomic DNA fragments representing a whole or substantially whole genome provide suitable templates for so-called "whole-genome" amplification. The term "whole-genome amplification" refers to a nucleic acid amplification reaction (e.g. PCR) in which the template to be amplified comprises a complex mixture of nucleic acid fragments representative of a whole (or substantially whole genome).

Solid-phase amplification

Once formed, the library of templates prepared according to the methods described above can be used for solid-phase nucleic acid amplification.

Thus, in further aspects the invention provides a method of solid-phase nucleic acid amplification of template polynucleotide molecules which comprises:
preparing a library of template polynucleotide molecules which have common sequences at their 5' and 3' ends using a method according to the first aspect of the invention described herein and carrying out a solid-phase nucleic acid amplification reaction wherein said template polynucleotide molecules are amplified.

The term "solid-phase amplification" as used herein refers to any nucleic acid amplification reaction carried out on or in association with a solid support such that all or a portion of the amplified products are immobilised on the solid support as they are formed. In particular, the

term encompasses solid-phase polymerase chain reaction (solid-phase PCR), which is a reaction analogous to standard solution phase PCR, except that one or both of the forward and reverse amplification primers is/are immobilised on the solid support.

Although the invention encompasses "solid-phase" amplification methods in which only one amplification primer is immobilised (the other primer usually being present in free solution), it is preferred for the solid support to be provided with both the forward and the reverse primers immobilised. In practice, there will be a "plurality" of identical forward primers and/or a "plurality" of identical reverse primers immobilised on the solid support, since the PCR process requires an excess of primers to sustain amplification. References herein to forward and reverse primers are to be interpreted accordingly as encompassing a "plurality" of such primers unless the context indicates otherwise.

As will be appreciated by the skilled reader, any given PCR reaction requires at least one type of forward primer and at least one type of reverse primer specific for the template to be amplified. However, in certain embodiments the forward and reverse primers may comprise template-specific portions of identical sequence, and may have entirely identical nucleotide sequence and structure (including any non-nucleotide modifications). In other words, it is possible to carry out solid-phase amplification using only one type of primer, and such single-primer methods are encompassed within the scope of the invention. Other embodiments may use forward and reverse primers which contain identical template-specific sequences but which differ in some other structural features. For example one

type of primer may contain a non-nucleotide modification which is not present in the other.

In other embodiments of the invention the forward and reverse primers may contain template-specific portions of different sequence.

In all embodiments of the invention, amplification primers for solid-phase PCR are preferably immobilised by covalent attachment to the solid support at or near the 5′ end of the primer, leaving the template-specific portion of the primer free for annealing to it's cognate template and the 3′ hydroxyl group free for primer extension. Any suitable covalent attachment means known in the art may be used for this purpose. The chosen attachment chemistry will depend on the nature of the solid support, and any derivatisation or functionalisation applied to it. The primer itself may include a moiety, which may be a non-nucleotide chemical modification, to facilitate attachment. In one particularly preferred embodiment the primer may include a sulphur-containing nucleophile, such as phosphorothioate or thiophosphate, at the 5' end. In the case of solid-supported polyacrylamide hydrogels (as described below), this nucleophile will bind to a "C" group present in the hydrogel. The most preferred means of attaching primers and templates to a solid support is via 5' phosphorothioate attachment to a hydrogel comprised of polymerised acrylamide and *N*-(5-bromoacetamidylpentyl) acrylamide (BRAPA).

It is preferred to use the library of templates prepared according to the first aspect of the invention to prepare clustered arrays of nucleic acid colonies, analogous to those described in WO 00/18957 and WO 98/44151, by solid-phase PCR amplification. The terms "cluster" and "colony" are used interchangeably herein to refer to a discrete site on a solid support comprised of a plurality of identical

immobilised nucleic acid strands and a plurality of identical immobilised complementary nucleic acid strands. The term "clustered array" refers to an array formed from such clusters or colonies. In this context the term "array" is not to be understood as requiring an ordered arrangement of clusters.

Use in sequencing/methods of sequencing

The invention also encompasses methods of sequencing amplified nucleic acids generated by whole genome or solid-phase amplification. Thus, the invention provides a method of nucleic acid sequencing comprising amplifying a library of nucleic acid templates using whole genome or solid-phase amplification as described above and carrying out a nucleic acid sequencing reaction to determine the sequence of the whole or a part of at least one amplified nucleic acid strand produced in the whole genome or solid-phase amplification reaction.

Sequencing can be carried out using any suitable "sequencing-by-synthesis" technique, wherein nucleotides are added successively to a free 3' hydroxyl group, resulting in synthesis of a polynucleotide chain in the 5' to 3' direction. The nature of the nucleotide added is preferably determined after each nucleotide addition.

The initiation point for the sequencing reaction may be provided by annealing of a sequencing primer to a product of the whole genome or solid-phase amplification reaction. In this connection, one or both of the adapters added during formation of the template library may include a nucleotide sequence which permits annealing of a sequencing primer to

amplified products derived by whole genome or solid-phase amplification of the template library.

The products of solid-phase amplification reactions wherein both forward and reverse amplification primers are covalently immobilised on the solid surface are so-called "bridged" structures formed by annealing of pairs of immobilised polynucleotide strands and immobilised complementary strands, both strands being attached to the solid support at the 5' end. Arrays comprised of such bridged structures provide inefficient templates for nucleic acid sequencing, since hybridisation of a conventional sequencing primer to one of the immobilised strands is not favoured compared to annealing of this strand to its immobilised complementary strand under standard conditions for hybridisation.

In order to provide more suitable templates for nucleic acid sequencing it is preferred to remove substantially all or at least a portion of one of the immobilised strands in the "bridged" structure in order to generate a template which is at least partially single-stranded. The portion of the template which is single-stranded will thus be available for hybridisation to a sequencing primer. The process of removing all or a portion of one immobilised strand in a "bridged" double-stranded nucleic acid structure may be referred to herein as "linearisation".

Bridged template structures may be linearised by cleavage of one or both strands with a restriction endonuclease or by cleavage of one strand with a nicking endonuclease. Other methods of cleavage can be used as an alternative to restriction enzymes or nicking enzymes, including *inter alia* chemical cleavage (e.g. cleavage of a diol linkage with periodate), cleavage of abasic sites by

cleavage with endonuclease, or by exposure to heat or alkali, cleavage of ribonucleotides incorporated into amplification products otherwise comprised of deoxyribonucleotides, photochemical cleavage or cleavage of a peptide linker.

It will be appreciated that a linearization step may not be essential if the solid-phase amplification reaction is performed with only one primer covalently immobilised and the other in free solution.

In order to generate a linearised template suitable for sequencing it is necessary to remove "unequal" amounts of the complementary strands in the bridged structure formed by amplification so as to leave behind a linearised template for sequencing which is fully or partially single stranded. Most preferably one strand of the bridged structure is substantially or completely removed.

Following the cleavage step, regardless of the method used for cleavage, the product of the cleavage reaction may be subjected to denaturing conditions in order to remove the portion(s) of the cleaved strand(s) that are not attached to the solid support. Suitable denaturing conditions will be apparent to the skilled reader with reference to standard molecular biology protocols (Sambrook *et al.*, 2001, *Molecular Cloning, A Laboratory Manual*, 3rd Ed, Cold Spring Harbor Laboratory Press, Cold Spring Harbor Laboratory Press, NY; Current Protocols, eds Ausubel *et al.*).

Denaturation (and subsequent re-annealing of the cleaved strands) results in the production of a sequencing template which is partially or substantially single-stranded. A sequencing reaction may then be initiated by hybridisation of a sequencing primer to the single-stranded portion of the template.

Thus, the nucleic acid sequencing reaction may comprise hybridising a sequencing primer to a single-stranded region

of a linearised amplification product, sequentially incorporating one or more nucleotides into a polynucleotide strand complementary to the region of amplified template strand to be sequenced, identifying the base present in one or more of the incorporated nucleotide(s) and thereby determining the sequence of a region of the template strand.

One preferred sequencing method which can be used in accordance with the invention relies on the use of modified nucleotides that can act as chain terminators. Once the modified nucleotide has been incorporated into the growing polynucleotide chain complementary to the region of the template being sequenced there is no free 3'-OH group available to direct further sequence extension and therefore the polymerase can not add further nucleotides. Once the nature of the base incorporated into the growing chain has been determined, the 3' block may be removed to allow addition of the next successive nucleotide. By ordering the products derived using these modified nucleotides it is possible to deduce the DNA sequence of the DNA template. Such reactions can be done in a single experiment if each of the modified nucleotides has attached a different label, known to correspond to the particular base, to facilitate discrimination between the bases added at each incorporation step. Alternatively, a separate reaction may be carried out containing each of the modified nucleotides separately.

The modified nucleotides may carry a label to facilitate their detection. Preferably this is a fluorescent label. Each nucleotide type may carry a different fluorescent label. However the detectable label need not be a fluorescent label. Any label can be used which allows the detection of an incorporated nucleotide.

One method for detecting fluorescently labelled nucleotides comprises using laser light of a wavelength specific for the labelled nucleotides, or the use of other suitable sources of illumination. The fluorescence from the label on the nucleotide may be detected by a CCD camera or other suitable detection means.

The invention is not intended to be limited to use of the sequencing method outlined above, as essentially any sequencing methodology which relies on successive incorporation of nucleotides into a polynucleotide chain can be used. Suitable alternative techniques include, for example, Pyrosequencing™, FISSEQ (fluorescent in situ sequencing), MPSS (massively parallel signature sequencing) and sequencing by ligation-based methods.

The target polynucleotide to be sequenced using the method of the invention may be any polynucleotide that it is desired to sequence. Using the template library preparation method described in detail herein it is possible to prepare template libraries starting from essentially any double or single-stranded target polynucleotide of known, unknown or partially known sequence. With the use of clustered arrays prepared by solid-phase amplification it is possible to sequence multiple targets of the same or different sequence in parallel.

Kits

The invention also relates to kits for use in preparing libraries of template polynucleotides using the method of the first aspect of the invention.

Preferred embodiments of the kit comprise at least a supply of a mismatched adapter as defined herein, plus a supply of at least one amplification primer which is capable

of annealing to the mismatched adapter and priming synthesis of an extension product, which extension product would include any target sequence ligated to the adapter when the adapter is in use.

The preferred features of the "mismatch" adapters for inclusion in the kit are as described elsewhere herein in relation to other aspects of the invention. The structure and properties of amplification primers will be well known to those skilled in the art. Suitable primers of appropriate nucleotide sequence for use with the adapters included in the kit can be readily prepared using standard automated nucleic acid synthesis equipment and reagents in routine use in the art. The kit may include as supply of one single type of primer or separate supplies (or even a mixture) of two different primers, for example a pair of PCR primers suitable for PCR amplification of templates modified with the mismatched adapter in solution phase and/or on a suitable solid support (i.e. solid-phase PCR).

In one embodiment the kit may include supplies of different primer-pairs for use in solution phase and solid phase PCR. In this context the "different" primer-pairs may be of substantially identical nucleotide sequence but differ with respect to some other feature or modification, such as for example surface-capture moieties, etc. In other embodiments the kit may include a supply of primers for use in an initial primer extension reaction and a different primer-pair (or pairs) for solution and/or solid phase PCR amplification.

Adapters and/or primers may be supplied in the kits ready for use, or more preferably as concentrates requiring dilution before use, or even in a lyophilised or dried form requiring reconstitution prior to use. If required, the

kits may further include a supply of a suitable diluent for dilution or reconstitution of the primers. Optionally, the kits may further comprise supplies of reagents, buffers, enzymes, dNTPs etc for use in carrying out PCR amplification. Suitable (but non-limiting) examples of such reagents are as described in the Materials and Methods sections of the accompanying Examples. Further components which may optionally be supplied in the kit include "universal" sequencing primers suitable for sequencing templates prepared using the mismatched adapters and primers,

The invention will be further understood with reference to the following non-limiting experimental example:

Example

Experimental overview

The following experimental details describe the complete exposition of one embodiment of the invention as described above. The DNA source used is purified Human cell line DNA supplied by the Coriell Cell Repositories, Camden, NJ 08103 USA, catalog no. NA07055. The DNA is first prepared for ligation to forked adapters by: fragmentation of the DNA by nebulisation, end repair of the DNA ends to make them blunt-ended and phosphorylation, then the addition of a single 'A' nucleotide onto the 3' ends of the human DNA fragments. The ligation reaction is performed with the prepared fragmented DNA and adapters pre-formed by annealing 'Oligo A' and 'Oligo B' (sequences given below). The product of the reaction is isolated/purified from unligated adapter by gel electrophoresis. Finally, the product of the ligation

reaction is subject to cycles of PCR to selectively amplify ligated product that contains adapter at both ends of the fragments.

Materials and methods

Nebulization

Materials:

- Human genomic DNA (1mg/ml) Coriell NA07055
- Buffer (glycerol 53.1 ml, water 42.1 ml, 1 M TrisHCl pH7.5 3.7 ml, 0.5 M EDTA 1.1 ml)
- Nebulizer Invitrogen (#K7025-05)
- Qiagen columns PCR purification kit (#28104)

Mix: 25µl (5 micrograms) of DNA
725µl Buffer

Procedure:

Chilled DNA solution was fragmented in the nebulizer on ice for 5 to 6 minutes under at least 32psi of pressure. The recovered volume (usually somewhere between 400 and 600µl) was split into 3 aliquots and purified with a Qiagen PCR-purification kit, but using only one column, and finally eluted in 30µl of EB (Qiagen).

End-repair

Materials:

- T4 DNA Polymerase NEB #M0203S
- 10xNEB 2 buffer NEB #M7002S
- 100x BSA NEB #M9001S

- dNTPs mix (10mM each) NEB #N0447S
- E. coli DNA Pol I large fragment (Klenow, NEB #M0210S)
- T4 polynucleotide kinase NEB #M0201S
- T4 PNK buffer NEB #M0201S
- 100 mM ATP
- Qiagen columns PCR purification kit (#28104)

End repair mix was assembled as follows:

DNA	30µl
Water	12µl
10xNEB2	5µl
100xBSA	0.5µl
10mM dNTPs	2µl
T4 DNA pol (3U/µl)	5µl
	50µl total

The reaction was incubated for 15 min at room temperature, then 1µl of E. coli DNA Pol I large fragment (Klenow) added and the reaction incubated for a further 15 min at room temperature. The DNA was purified from enzymes, buffer, etc by loading the reaction mix on a Qiagen column, finally eluting in 30µl EB. The 5' ends of the DNA were then phosphorylated using polynucleotide kinase as follows:

DNA	30µl
Water	9.5µl
10xPNK buffer	5µl
100mM ATP	0.5µl
T4 PNK (10U/µl)	5µl
	50µl total

The reaction was incubated for 30 min at 37°C, then heat inactivated at 65°C for 20 min. DNA was then purified from enzymes, buffer, etc by loading the reaction mix on a Qiagen column, finally eluting in 30µl EB. Three separate tubes were pooled to give 90µl total.

A- tailing reaction

Materials:

- Taq DNA polymerase NEB #M0267L
- 10x thermopol buffer NEB #B9004S
- 1mM dATP Amersham-Pharmacia #272050
- Qiagen MinElute column PCR purification kit (#28004)

The following reaction mix was assembled:

DNA	30µl
10x thermopol buffer	5µl
1mM dATP	10µl
Taq pol (5U/µl)	3µl
	~50µl total

The reaction was incubated for 30 min at 70°C, then the DNA purified from enzymes, buffer, etc by loading the reaction mix on a Qiagen MinElute column, finally eluting in 10µl EB.

Anneal forked adapter

Materials:

- 'Oligo A' and 'Oligo B'
- 50mM Tris/50mM NaCl pH7
- PCR machine

100µM Oligo A 20µl

100µm Oligo B	20µl
Tris/NaCl	10µl
	50µl at 40µM duplex in 10mM Tris/10mM NaCl pH7.5

Oligo A: 5'ACACTCTTTCCCTACACGACGCTCTTCCGATCxT (x = phosphorothioate bond) (SEQ ID NO:1)

Oligo B: 5'Phosphate-GATCGGAAGAGCTCGTATGCCGTCTTCTGCTTG (SEQ ID NO:2)

The adapter strands were annealed in a PCR machine programmed as follows:

Ramp at 0.5 °C/sec to 97.5 °C

Hold at 97.5 °C for 150 sec

Then a step of 97.5 °C for 2 sec with a temperature drop of 0.1 °C/cycle for 775 cycles

Ligation reaction

Materials:

- 40µM forked adapter
- A-tailed genomic DNA
- Quick Ligase NEB #M2200L
- Quick Ligase 2x buffer NEB #M2200L
- PCR machine
- Qiagen columns PCR purification kit (#28104)

Reaction mix was assembled as follows:

DNA	10µl
2x buffer	25µl
40 µM adapter	10µl
Quick Ligase	5µl

~50µl total

The reaction was incubated for 20 min at room temperature then the DNA purified from enzymes, buffer, etc by loading the reaction mix on a Qiagen column, finally eluting in 30µl EB.

Gel purification

Materials:

- Agarose Biorad #161-3101
- 100 base pair ladder NEB #N3231L
- TAE
- Loading buffer (50 mM Tris pH8, 40 mM EDTA, 40% w/v sucrose)
- Ethidium bromide
- Gel trays and tank. Electrophoresis unit

The entire sample from the purified ligation reaction was loaded into one lane of a 2% agarose gel containing ethidium bromide and run at 120V for 50 min. The gel was then viewed on a 'White-light' box and fragments from above 300bp to at least 750bp excised and purified with a Qiagen Gel purification kit, eluting in 30µl EB. For large gel slices two minElute columns were used, eluting each in 15µl EB and pool.

PCR amplification

Materials:

- Ligated DNA
- PRIMER 1: AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGA (SEQ ID NO:3)
- PRIMER 2: CAAGCAGAAGACGGCATACGA (SEQ ID NO:4)

- 2x Jump Start RedTaq PCR mix Sigma #P0982
- PCR machine
- Qiagen MinElute columns Qiagen(#28004)

The purified ligated DNA was diluted 25 fold, then a PCR reaction mix prepared as follows:

DNA	1µl
2x Jump Start RedTaq mix	25µl
100 µM Primer 1	0.5µl
100 µM Primer 2	0.5µl
Water	23µl
	~50µl total

Thermocycling was carried out in a PCR machine under the following conditions:

- 2 min @ 94°C
- [45 sec@ 94°C, 45 sec @ 65°C, 2 min @ 70°C] 16 cycles
- 5 min @ 70°C
- Hold @ 4°C

PCR products were purified from enzymes, buffer, etc on a Qiagen MinElute column, eluting in 10µl EB. The resulting DNA library is ready for amplification on a surface PCR platform.

Validation of library

1 µl of the DNA library was cloned into a plasmid vector and plated out on agar. 19 colonies were picked, miniprepped and the cloned inserts sequenced by conventional Sanger sequencing. The sequence data obtained was as follows:

Clone 1 (SEQ ID NO:5)

TGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTGTGG
GGACCGTCCTGTGCATTGTAGGGTGTTCAACAGCATCCCTGACCTCCACCTACAAGATGC
CAGTAGCGAATCCCCTCAGCCCTCATCTCCTTGCCATAGTTGTGTCAACCAAAATCATCT
CCACACATTGTTAGATGTTTACTGGGAGGCAGACTCACTCCCACTTGAGAACCACTGTAC
TAGAAATATCACCAAGAGAATGAGATCGGAAGAGCTCGTATGCCGTCTTCTGCTTG

Clone 2 (SEQ ID NO:6)

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTGT
GGGCTTTGTTCTTTGAGAGGTTGCAGTCAACATGATTCTTTAAGACCAGAACCCTGCACA
CTTCTTGGGCTGTATTTCTTACATTCCTTTTCTATTTTAACCATATCCCATCTTACCTAC
TTCCAGCATAGTGGTCATATTTAATTTTTACAAAACCATTTTGCCACTTGCTGCCAACTA
TGTTCTTTATAAAGCAGACTTTGAGATGGAGGCTAGTGTTCAGAGGGGATGCTTAGGAGA
ACTTTGGAGATTAATACTTATGGCAGGTAAGGGAAGGAAGCAGGATTAGACAGAAATATT
GAACTGTGATACAAAGTCAGCAAAGACTTTAGTCAATAGATCGGAAGAGCTCGTATGCCG
TCTTCTGCTTG

Clone 3 (SEQ ID NO:7)

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTTT
CGATTCCCTTCAATGATTATTCCATTCGAGTACATTCGATGATTCCATTCGATTCTATAT
GATGATGATTGCATTCGAGTCCGTGTATTATTCCATTCCATTCCATTAGATGATTCCATT
CGAGTCCATTCGATGATTCTCTTCGATTCCGTTCGATAATTACGCTTGATTCCGTTTGAT
GTTGATTCCATTCGAGTCCATTCAATGTTAATTCCATTCGATTCTAAGCGATGATTCCAT
TCCTTTCCATTAGAAGATGATTCCATTCGAGACCATTCGATGATTGCATTCAACTCATTC
GATGACGATTCCATTCAATTCTGTTCAATGATTCCATTAGATTCCATTTGATGATGAGTC
CATTCGATTCCATTTGATGATGATTCCATGCGATTCCATTAGATGATGACCCCTTTCATT
TCCAAGATCGGAAGAGCTCGTATGCCGTCTTCTGCTTG

Clone 4 (SEQ ID NO:8)

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTTT
AAATGCTAGGCATATTGTGTACCCCACATTGGTTTGTAGCCAGCTCTATGTCATAGGGCC
CTTACCCTTTACCTATTTATTGTTAGTATAATGTCCATAAACAAGCCAATGGCTCAGCAT

GAACTGATGCTAAAGAAAGCTCATGCCTGAGTGATAAATTAAGTGACCTCAGCTATTTCT
CTTCAGTGTTGTGAAAGTTATTTTTAACAGTAGGTTTCCTGGTAGATTCTCTAACCACTC
GGTATTTCACATGGCCCAACTTGGTTAACTCGACTGGTTACGGCAAATGCTGAAGATCGG
AAGAGCTCGTATGCCGTCTTCTGCTTG

Clone 5 (SEQ ID NO:9)
AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGAAGCTCTTCCGATCTTA
AGGAAGTTAGGTAGATAATTTTTGTTTAGGCCATATAGCTTTGATTTTCTGATAACAATT
TTATAAACTTAGAAATTTTCATGTAAGATACAGGAATACTGGAAGCAAAAAAAAGAAGGT
GCTTTAACCTTAGGGATTGAAAAAATAGTAATTTAGGTTGAAAATGCTGCTTGAAAGTTA
ATGCTGATAGCATTACTACACATGATGATTTTTTCTGGAAGGAAAGCTTTATCTGGGCCT
TCAATTTAGGAATTTTTCTCTTTGGTTTTTAAAAGCTGCCATATTCACTTGAGCTTCATG
GGAAAGATGCAAATAACTAAAACAAATGAACAAAAACCATGTTGAGGTCAGGAACTTATT
TCAAGAAAGCAAGTTCTAGGTTTTCTTTTAAAGTGACAGTAGAGCCTTAGGCCTCAAACC
ATCTACAACCATGTTAACAGTAAGATCGGAAGAGCTCGTATGCCGTCTTCTGCTTG

Clone 6 (SEQ ID NO:10)
AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCAGATCTTC
CTGCCTCAGCCTCCCGAGTAGTTGGGATTGCAGGCATGTGCCACCATGCCCTGCTAATTT
TTGTATTTTTAACTAAAGGAGGGGTTTTGCCATGTTGGCCAGGCTGGTCTTGAATGCCTG
ACCTCAGGTGATCCGCCCACCTCAGCCTCCCAGTGCTGGGATTACAGGTGTGAGCCACTG
CGCCCAGCTGAGGGTAACTATTTTTAATGTGGCTGATGAATGTAACTATCCTGTCCCATG
TCTCTGTCCCCAGCTGCAGAGCCCTCGTCGAGATCGGAAGAGCTCGTATGCCGTCTTCTG
CTTG

Clone 7 (SEQ ID NO:11)
AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTAA
GGTTTAAATTTTCTATATAGCCTGAAAAGTTTGAATGTTTAATTCAAACTAATTTATTGA
GCAATGGCATTTAAGAAAATGGAAAGATACAAAGGGACTTTCATCAGATGATAAGTGGAT
AAGAGAGAAAAATGCAGACAGATGAGCCAGAGTTGTGTAAAAGCTGGGAGGCTAGCAGGG
CCTTGTAGATAGCCAAGCTGACTGGGGAACAGAGATAAATGGGAGCAGAGATCAGAAAGT

TCATCCTTACCCTGCTGCCCGTGGTGAAAGGAGACTTGCAAGATCGGAAGAGCTCGTATG
CCGTCTTCTGCTTG

Clone 8 (SEQ ID NO:12)

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTAA
GGAATCTTTATTTTCTACATTTGAGTTTGGAAAACTGAGCTAGCACATCTAAATCCATCT
AATTTTGGTCATTGGTTTTAACAAGTTCATCTTATTTTTTTAAACATCTGATCTTTATTT
TATAGAATAGACTACACAAAGTCTTTTGGAAAATTAAAATATTTTAACTTCCAACAATTT
TCAGATTTTACTTATAAAAAAATTTAAAATCCTCTACTTTACTCGCATCTTTATTATTTC
TGACTTTCTAGCTACTTAAAGTTAAGGAGGAAATTAACCTCTCTAAGATCGGAAGAGCTC
GTATGCCGTCTTCTGCTTG

Clone 9 (SEQ ID NO:13)

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTCG
CCATAACCACAGCCCAGGCCTCCGTAGCCACAGCACCCATAGCCATGGCCACCATTGTAG
TTTCTGTAGTAGCTGCCACACATGGTGTTGGTTGTTGAGGTCATCCTTGGGTAGGAAGGA
GTGTAGGTGACTTCAGTATGGACACTTCTCCGCAGAGGGCCTTTTATATGCCTCAGTGAA
TCAGAACATAGCGTGCCCCTGCAAAAATATCTCTAAAGGCCTTTCATTGTGCTGAGAAGT
TCTGGCCCTTACGTATCTCTCTGATTTCATATCCTGCTACTCTCCTCCATTTATCTATAA
TGCTCAAACTCTGCTGGCTTTTTGTCTTTTAAAATGCAGCAGGTTTCTTCTCACAATAAG
GAGATCGGAAGAGCTCGTATGCCGTCTTCTGCTTG

Clone 10 (SEQ ID NO:14)

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTAG
AGAGAAGTTATTTAAGACAAGTAAGGTATCAGGTTGTGACTCAAATACCACAGAATCTAG
CTATTGTTAGCAATATTAAGTATATTTTCTTAAATGAAGGATTCTCCATTTACCATATGC
CCCTTGGATAATTTCCAGAGAATTTAATTTTTTAAAAGGAATTTTCACCAATTAAATTAT
TGTTTTGATCAAAAGAGGACCCACTGAACACCTTATTCATTATTAAAATGTATCATAAAA
CTTAATTATGGAGCTGGGTACAGGGGCTCATGCCTGTAATCCCAGCACTTTGGGAGGCTG
AGGCAGGAGGACTGCTTGAGTCCAGGAGTTTGAGACCAGCCTGGGTAACATGGTGAAACC
CTGTCTCTACAAAAAATACAAAAAATTAGCCAGGTGTGGTGAGATCGGAAGAGCTCGTAT
GCCGTCTTCTGCTTG

Clone 11 (SEQ ID NO:15)

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTCC
CAGGGAAGCCAAAAGATTGGACACCCCTCTTTCAAACTATAAATTCCTCCCATAGTTAGT
TTGGCCTATGCTGGGAAATGAACAAGGGTGGCTTTGAGGTTAGAAGCAAAATGGAGTCAG
TTAGGTCAGACTTTTTTTCACTATCATACTTTTTCTATGTCAGATTTATCTCACTTGTAA
TTTTTGCAAGGGTGGTTTCAGAGCCACTAAGCTTGTGGTAATTTTTTACTGCAATAGAAA
ACTAATGCATTAGGTAACCCTCTTTTTTTCCCTCTGATTGCTTGCTCTGGGGTAAGCCAG
CTGCCATGTTGAATTTTTCATTTTGTTACTGAGCAATCAATGGGCTCACTGCCCGACGTG
CATAGAGGCCAATACTGTGGCACTAGTTTTTGAGAAAAGATCGGAAGAGCTCGTATGCCG
TCTTCTGCTTG

Clone 12 (SEQ ID NO:16)

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTTT
TCAGATTTTATATGTATTAACTCAGAACACACACCTCTTATCACACATATTTTTTCATGT
AATTTATCTAAATCTTATAGAAAAGGGTCCATTTGCATTTTCTCTTATTAGACTCCTGAT
TTCAAATAATATATTACTTATGAGTATTTTTCTGTGCTGTAGTTATTCATTCTTATAGAT
ATGTAACATAATTCCTTTTGCAAAGGTAAAAATTGAGCTATCTCTTGTTGAGGATTTGTT
GATCTCTGTCTAAAGTTTCAAAAATAAAGAACTTTAAAAGCAAAATGTAAATTCCTTTCA
AGTTTTAGTAAAATTACTTCAAACTTAGTAGCTTAAACAATACAGATTTATTATGTTACA
GTTCTGTAAGACAGAAATCTGACTTGATCACACCATGGTAAAACCAAGATACTGCCAGGG
TTGGTTTTTTCTTGGGGGGGGTCTGTGGGAAGAGTTTGTTTCCTTTGGTTTTCCACAGCC
CAGAGGCTGCTTGCATTCCTTTGATCACTAGATCGGAAGAGCTCGTATGCCGTCTTCTGC
TTG

Clone 13 (SEQ ID NO:17)

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTTA
TTGTTTAGGGAATAATGACAAGGAAAAAAAGTCTGTAGATATTCAGTACAGAGGCACCCA
TCTTTTTAAATTTCTGAAGATTTTTTACTCATGCTTGGTTGAATCCACAGATGCAGAACC
CATAGGTTCAGAGGGCCAGCTGTGCTTTGAAAATATTAGCTTGTGTTTTTATTAGAAAGA
AAACTCTGAGGCCAGGCACGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCTGAG
GTGGGCGGATCACAAGGTGAGGAGATCGAGACCATTCTGGCTAACATGGTGAAACCCTGT