than one sample. A target polynucleotide that has an adaptor/primer oligonucleotide joined to it is considered "tagged" by the joined adaptor.

[0057] 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.

Methods of Strand-specific Selection

[0058] The compositions and methods provided herein are useful for retaining directional information in double-stranded DNA.

[0059] The terms "strand specific" or "directional", as used herein, may refer to the ability to differentiate in a double-stranded polynucleotide between the original template strand and the strand that is complementary to the original template strand. Further, methods and compositions of the invention, in various embodiments, enable adapter ligation in a strand specific manner. In various embodiments, an adapter is incorporated at a chosen end of a strand, preferably a selected strand. Further, an adapter may be incorporated in a chosen orientation. In various embodiments, strand specificity, directionality and orientation is accomplished by selecting or enriching the desired configurations or strands.

[0060] In some embodiments, the methods of the invention are used to preserve information about the direction of single-stranded nucleic acid molecules while generating double-stranded polynucleotides more suitable for molecular cloning applications. One of the strands of the double-stranded polynucleotide is synthesized so that it has at least one modified nucleotide incorporated into it along the entire length of the strand. In some embodiments, the incorporation of the modified nucleotide marks the strand for degradation or removal.

[0061] The term "first strand synthesis" refers to the synthesis of the first strand using the original nucleic acid (RNA or DNA) as a starting template for the polymerase reaction. The nucleotide sequence of the first strand corresponds to the sequence of the complementary strand.

[0062] The term "second strand synthesis" refers to the synthesis of the second strand that uses the first strand as a template for the polymerase reaction. The nucleotide sequence of the second strand corresponds to the sequence of the original nucleic acid template.

IDT Exhibit 1020.5376

[0063] The term "unmodified dNTPs" or "classic dNTPs" refers to the four deoxyribonucleotide triphosphates dATP (deoxyadenosine triphosphate), dCTP (deoxycytidine triphosphate), dGTP (deoxyguanosine triphosphate) and dTTP (deoxythymidine triphosphate) that are normally used as building blocks in the synthesis of DNA. Similarly, the term "canonical dNTP" is used to refer to the four deoxyribonucleotide triphosphates dATP, dCTP, dGTP and dTTP that are normally found in DNA. Generally, nucleotides are present in nucleoside triphosphate form in a solution for a primer extension reaction. During primer extension reactions, they are typically incorporated into a polynucleotide in nucleoside form, e.g. adenosine, thymidine, guanosine, cytidine, uridine, etc. losing two phosphates, while one of the phosphates forms part of the polynucleotide backbone. The nucleobase, e.g. adenine, guanine, thymine, cytosine, uracil etc., of the nucleotides may be removed according to various embodiments of the invention, forming an abasic site. Various methods for removing nucleobases from polynucleotides, forming abasic sites are explained in detail herein and known in the art.

[0064] The term "canonical" as used herein, refers to the nucleic acid bases adenine, cytosine, guanine and thymine that are commonly found in DNA or their deoxyribonucleotide or deoxyribonucleoside analogs. The term "noncanonical" refers to nucleic acid bases in DNA other than the four canonical bases in DNA, or their deoxyribonucleotide or deoxyribonucleoside analogs. Although uracil is a common nucleic acid base in RNA, uracil is a non-canonical base in DNA.

[0065] The term "modified nucleotide" or "modified dNTP", as used herein, refers to any molecule suitable for substituting one corresponding unmodified or classic dNTP. Such modified nucleotide must be able to undergo a base pair matching identical or similar to the classic or unmodified dNTP it replaces. The modified nucleotide or dNTP can be suitable for specific degradation in which it is selectively degraded by a suitable degrading agent, thus rendering the DNA strand containing at least one modified and degraded dNTP essentially unfit for amplification and/or hybridization. Alternatively, the modified nucleotide must mark the DNA strand containing the modified nucleotide eligible for selective removal or facilitate separation of the polynucleotide strands. Such a removal or separation can be achieved by molecules, particles or enzymes interacting selectively with the modified nucleotide, thus selectively removing or marking for removal only one polynucleotide strand.

[0066] As used in this application, the term "strand marking" refers to any method for distinguishing between the two strands of a double-stranded polynucleotide. The term

IDT Exhibit 1020.5377

"selection" refers to any method for selecting between the two strands of a double-stranded polynucleotide. The term "selective removal" or "selective marking for removal" refers to any modification to a polynucleotide strand that renders that polynucleotide strand unsuitable for a downstream application, such as amplification or hybridization.

[0067] In a one embodiment, the selection is done by incorporation of at least one modified nucleotide into one strand of a synthesized polynucleotide, and the selective removal is by treatment with an enzyme that displays a specific activity towards the at least one modified nucleotide. In a preferred embodiment, the modified nucleotide being incorporated into one strand of the synthesized polynucleotide is deoxyuridine triphosphate (dUTP), replacing dTTP in the dNTP mix, and the selective removal of the marked strand from downstream applications is carried by the nuclease Uracil-N-Glycosylase (UNG). UNG selectively degrades dUTP while it is neutral towards other dNTPs and their analogs. Treatment with UNG results in the cleavage of the N-glycosylic bond and the removal of the base portion of dU residues, forming abasic sites. In a preferred embodiment, the UNG treatment is done in the presence of an apurinic/apyrimidinic endonuclease (APE) to create nicks at the abasic sites. Consequently, a polynucleotide strand with incorporated dUTP that is treated with UNG/APE is cleaved and unable to undergo amplification by a polymerase. In another embodiment, nick generation and cleavage is achieved by treatment with a polyamine, such as N,N′-dimethylethylenediamine (DMED), or by heat treatment. In a preferred embodiment, UNG treatment is conducted in a reaction buffer containing about 32 mM DMED.

[0068] As used in this application, the term "at least one nucleotide" or "at least one modified nucleotide" refers to a plurality of dNTP molecules of the same kind or species. Thus, use of "one modified nucleotide" refers to the replacement in the dNTP mix of one of the classic dNTPs dATP, dCTP, dGTP or dTTP with a corresponding modified nucleotide species.

[0069] In a preferred embodiment, the at least one modified nucleotide is dUTP, replacing dTTP in the dNTP mix. In another embodiment, the at least one modified nucleotide is a biotinylated dNTP. In another embodiment, the at least one modified nucleotide contains a thio group. In another embodiment, the at least one modified nucleotide is an aminoallyl dNTP. In yet another embodiment, the at least one modified nucleotide is inosine, replacing dGTP in the dNTP mix.

[0070] In some embodiments, the methods of the invention are used for construction of directional cDNA libraries. Strand marking is necessary, but not sufficient for construction of directional cDNA libraries when using adaptors that are not polarity-specific, i.e. adaptors generating ligation products with two adaptor orientations. Construction of directional cDNA

IDT Exhibit 1020.5378

libraries according to the methods of invention requires strand marking of both the cDNA insert and one of the two adaptors at the ligation strand of the adaptor. A useful feature of the present invention is the ability to switch around the adaptor orientation. For example, in a duplex adaptor system where P1/P2 designates adaptor orientation resulting in sense strand selection and (optional) sequencing, and where the P2 adaptor has at least one modified nucleotide incorporated along the ligation strand of the adaptor, modification of the protocol such that the P1 adaptor (as opposed to P2 adaptor) has at least one modified nucleotide incorporated along the ligation strand allows for antisense strand selection and (optional) sequencing.

[0071] The methods of the present invention may further include a step of cleaving the input nucleic acid template. In some cases, the input nucleic acid template may be cleaved with an agent such as an enzyme. In the embodiment where the polynucleotide comprises a non-canonical nucleotide, the polynucleotide may be treated with an agent, such as an enzyme, capable of generally, specifically, or selectively cleaving a base portion of the non-canonical deoxyribonucleoside to create an abasic site. As used herein, "abasic site" encompasses any chemical structure remaining following removal of a base portion (including the entire base) with an agent capable of cleaving a base portion of a nucleotide, e.g., by treatment of a non-canonical nucleotide (present in a polynucleotide chain) with an agent (e.g., an enzyme, acidic conditions, or a chemical reagent) capable of effecting cleavage of a base portion of a non-canonical nucleotide. In some embodiments, the agent (such as an enzyme) catalyzes hydrolysis of the bond between the base portion of the non-canonical nucleotide and a sugar in the non-canonical nucleotide to generate an abasic site comprising a hemiacetal ring and lacking the base (interchangeably called "AP" site), though other cleavage products are contemplated for use in the methods of the invention. Suitable agents and reaction conditions for cleavage of base portions of non-canonical nucleotides include: N-glycosylases (also called "DNA glycosylases" or "glycosidases") including Uracil N-Glycosylase ("UNG"; specifically cleaves dUTP) (interchangeably termed "uracil DNA glyosylase"), hypoxanthine-N-Glycosylase, and hydroxy-methyl cytosine-N-glycosylase; 3-methyladenine DNA glycosylase, 3- or 7-methylguanine DNA glycosylase, hydroxymethyluracile DNA glycosylase; T4 endonuclease V. See, e.g., Lindahl, PNAS (1974) 71(9):3649-3653; Jendrisak, U.S. Pat. No. 6,190,865 B1 or any of the glycosidases provided in Table 1 or homologues thereof such as enzymes with greater than about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 99%, 99.5%, or higher homology or identity at the amino acid or nucleotide level with any of the glycosydases provided herein. In one embodiment, uracil-N-

IDT Exhibit 1020.5379

glycosylase is used to cleave a base portion of the non-canonical nucleotide. In other embodiments, the agent that cleaves the base portion of the non-canonical nucleotide is the same agent that cleaves a phosphodiester backbone at the abasic site.

Table 1: Glycosylases in bacteria, yeast and humans

E. coli	Yeast (S. cerevisiae)	Human	Type	Substrates
AlkA	Mag1		monofunctional	3-meA, hypoxanthine
UDG	Ung1	UNG	monofunctional	uracil
Fpg	Ogg1	hOGG1	bifunctional	8-oxoG, FapyG
Nth	Ntg1	hNTH1	bifunctional	Tg, hoU, hoC, urea, FapyG
	Ntg2			
Nei		hNEIL1	bifunctional	Tg, hoU, hoC, urea, FapyG, FapyA
		hNEIL2		AP site, hoU
		hNEIL3		unknown
MutY		hMYH	monofunctional	A:8-oxoG
		hSMUG1	monofunctional	U, hoU, hmU, fU
		TDG	monofunctional	T:G mispair
		MBD4	monofunctional	T:G mispair

[0072] Cleavage of base portions of non-canonical nucleotides may provide general, specific or selective cleavage (in the sense that the agent (such as an enzyme) capable of cleaving a base portion of a non-canonical nucleotide generally, specifically or selectively cleaves the base portion of a particular non-canonical nucleotide), whereby substantially all or greater than about 99.9%, 99.5%, 99%, 98.5%, 98%, about 95%, about 90%, about 85%, about 80%, about 75%, about 70%, about 65%, about 60%, about 55%, about 50%, about 45%, or about 40% of the base portions cleaved are base portions of non-canonical nucleotides. However, extent of cleavage can be less. Thus, reference to specific cleavage is exemplary. General, specific or selective cleavage is desirable for control of the fragment size in the methods of generating template polynucleotide fragments of the invention (i.e., the fragments generated

IDT Exhibit 1020.5380

by cleavage of the backbone at an abasic site). Reaction conditions may be selected such that the reaction in which the abasic site(s) are created can run to completion, or the reaction may be carried out until 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 99%, or about 100% of the non-canonical nucleotides are converted to abasic sites. In some cases, the reaction conditions may be selected such that the reaction in which abasic site(s) are created at between about 10% and about 100% of the one or more non-canonical nucleotides present in the template nucleic acid, between about 20% and about 90%, between about 30% and about 90%, between about 50% and about 90% 95%, 99%, or 100% of the non-canonical nucleotides in the template nucleic acid.

[0073] In some embodiments, the template polynucleotide comprising a non-canonical nucleotide is purified following synthesis of the template polynucleotide (to eliminate, for example, residual free non-canonical nucleotides that are present in the reaction mixture). In other embodiments, there is no intermediate purification between the synthesis of the template polynucleotide comprising the non-canonical nucleotide and subsequent steps (such as hybridization of primers, extension of primers to produce primer extension products that do not comprise non-canonical nucleotides, or do not comprise the same non-canonical nucleotides as the template nucleic acid, cleavage of a base portion of the non-canonical nucleotide and cleavage of a phosphodiester backbone at the abasic site).

[0074] It is understood that the choice of non-canonical nucleotide can dictate the choice of enzyme to be used to cleave the base portion of that non-canonical enzyme, to the extent that particular non-canonical nucleotides are recognized by particular enzymes that are capable of cleaving a base portion of the non-canonical nucleotide. In some cases, the enzyme is a glycosylase. For example, a template nucleic acid comprising non-canonical nucleotides such as dUTP, 8-oxoguanine, or a methylated purine which may be cleaved by glycosylases may be used in the methods of the present invention. Other suitable non-canonical nucleotides include deoxyinosine triphosphate (dITP), 5-hydroxymethyl deoxycytidine triphosphate (5-OH-Me-dCTP) or any of the non-canonical nucleotides provided in Table 1. See, e.g., Jendrisak, U.S. Pat. No. 6,190,865. A glycosylase such as uracil DNA glycosylase (known as UNG or UDG) which may act on dUTP to provide an abasic site, Ogg1 which may act on 8-oxoguanine to provide an abasic site, or N-methyl purine DNA glycosylase which may act on methylated purines to provide an abasic site may then be used in the methods of the present invention to act on the input nucleic acid template comprising non-canonical nucleotides to initiate a step of cleaving the input nucleic acid template. The enzymes as provided herein may provide N-glycosydic bond cleavage of the input nucleic

IDT Exhibit 1020.5381

acid template at the one or more non-canonical nucleotides provided herein to produce one or more abasic (apurinic or apyrimidic) sites.

[0075] Additional glycosylases which may be used in the methods of the present invention and their non-canonical nucleotide substrates include 5-methylcytosine DNA glycosylase (5-MCDG), which cleaves the base portion of 5-methylcytosine (5-MeC) from the DNA backbone (Wolffe et al., Proc. Nat. Acad. Sci. USA 96:5894-5896, 1999); 3-methyladenosine-DNA glycosylase I, which cleaves the base portion of 3-methyl adenosine from the DNA backbone (see, e.g. Hollis et al (2000) Mutation Res. 460: 201-210); and/or 3-methyladenosine DNA glycosylase II, which cleaves the base portion of 3-methyladenosine, 7-methylguanine, 7-methyladenosine, and/3-methylguanine from the DNA backbone. See McCarthy et al (1984) EMBO J. 3:545-550. Multifunctional and mono-functional forms of 5-MCDG have been described. See Zhu et al., Proc. Natl. Acad. Sci. USA 98:5031-6, 2001; Zhu et al., Nuc. Acid Res. 28:4157-4165, 2000; and Nedderrnann et al., J. B. C. 271:12767-74, 1996 (describing bifunctional 5-MCDG; Vairapandi & Duker, Oncogene 13:933-938, 1996; Vairapandi et al., J. Cell. Biochem. 79:249-260, 2000 (describing mono-functional enzyme comprising 5-MCDG activity). In some embodiments, 5-MCDG preferentially cleaves fully methylated polynucleotide sites (e.g., CpG dinucleotides), and in other embodiments, 5-MCDG preferentially cleaves a hemi-methylated polynucleotide. For example, mono-functional human 5-methylcytosine DNA glycosylase cleaves DNA specifically at fully methylated CpG sites, and is relatively inactive on hemimethylated DNA (Vairapandi & Duker, supra; Vairapandi et al., supra). By contrast, chick embryo 5-methylcytosine-DNA glycosylase has greater activity directed to hemimethylated methylation sites. In some embodiments, the activity of 5-MCDG is potentiated (increased or enhanced) with accessory factors, such as recombinant CpG-rich RNA, ATP, RNA helicase enzyme, and proliferating cell nuclear antigen (PCNA). See U.S. Patent Publication No. 20020197639 A1. One or more agents may be used. In some embodiments, the one or more agents cleave a base portion of the same methylated nucleotide. In other embodiments, the one or more agents cleave a base portion of different methylated nucleotides. Treatment with two or more agents may be sequential or simultaneous.

[0076] Appropriate reaction media and conditions for carrying out the cleavage of a base portion of a non-canonical nucleotide according to the methods of the invention are those that permit cleavage of a base portion of a non-canonical nucleotide. Such media and conditions are known to persons of skill in the art, and are described in various publications, such as Lindahl, PNAS (1974) 71(9):3649-3653; and Jendrisak, U.S. Pat. No. 6,190,865 BI; U.S. Pat.

IDT Exhibit 1020.5382

No. 5,035,996; and U.S. Pat. No. 5,418,149. For example, buffer conditions can be as described above with respect to polynucleotide synthesis. In one embodiment, UDG (Epicentre Technologies, Madison Wis.) is added to a nucleic acid synthesis reaction mixture, and incubated at 37° C. for 20 minutes. In one embodiment, the reaction conditions are the same for the synthesis of a polynucleotide comprising a non-canonical nucleotide and the cleavage of a base portion of the non-canonical nucleotide. In another embodiment, different reaction conditions are used for these reactions. In some embodiments, a chelating regent (e.g. EDTA) is added before or concurrently with UNG in order to prevent a polymerase from extending the ends of the cleavage products.

[0077] The polynucleotide comprising an abasic site may be labeled using an agent capable of labeling an abasic site, and, in embodiments involving fragmentation, the phosphodiester backbone of the polynucleotide comprising an abasic site may be cleaved at the site of incorporation of the non-canonical nucleotide (i.e., the abasic site by an agent capable of cleaving the phosphodiester backbone at an abasic site, such that two or more fragments are produced. In embodiments involving fragmentation, labeling can occur before fragmentation, fragmentation can occur before labeling, or fragmentation and labeling can occur simultaneously.

[0078] Agents capable of labeling (e.g., generally or specifically labeling) an abasic site, whereby a polynucleotide (or polynucleotide fragment) comprising a labeled abasic site is generated, are provided herein. In some embodiments, the detectable moiety (label) is covalently or non-covalently associated with an abasic site. In some embodiments, the detectable moiety is directly or indirectly associated with an abasic site. In some embodiments, the detectable moiety (label) is directly or indirectly detectable. In some embodiments, the detectable signal is amplified. In some embodiments, the detectable moiety comprises an organic molecule such as a chromophore, a fluorophore, biotin or a derivative thereof. In other embodiments, the detectable moiety comprises a macromolecule such as a nucleic acid, an aptamer, a peptide, or a protein such as an enzyme or an antibody. In other embodiments, the detectable signal is fluorescent. In other embodiments, the detectable signal is enzymatically generated. In some embodiments, the label is selected from, fluorescein, rhodamine, a cyanine dye, an indocyanine dye, Cy3, Cy5, an Alexa Fluor dye, phycoerythrin, 5-(((2-(carbohydrazino)-methyl)thio)acetyl)aminofluorescein, aminooxyacetyl hydrazide ("FARP"), or N-(aminooxyacetyl)-N′-(D-biotinoyl) hydrazine, trifluoroacetic acid salt (ARP).

IDT Exhibit 1020.5383

[0079] The cleavage of the input nucleic acid template comprising one or more abasic sites may further be provided by the use of enzymatic or chemical means or by the application of heat, or a combination thereof. For example the input nucleic acid template comprising one or more abasic sites may be treated with a nucleophile or a base. In some cases, the nucleophile is an amine such as a primary amine, a secondary amine, or a tertiary amine. For example, the abasic site may be treated with piperidine, moropholine, or a combination thereof. In some cases, hot piperidine (e.g., 1M at 90°C) may be used to cleave the input nucleic acid template comprising one or more abasic sites. In some cases, morpholine (e.g., 3M at 37°C or 65°C) may be used to cleave the input nucleic acid template comprising one or more abasic sites. Alternatively, a polyamine may be used to cleave the input nucleic acid template comprising one or more abasic sites. Suitable polyamines include for example spermine, spermidine, 1,4-diaminobutane, lysine, the tripeptide K-W-K, N, N-dimethylethylenediamine (DMED), piperazine, 1,2-ethylenediamine, or any combination thereof. In some cases, the input nucleic acid template comprising one or more abasic sites may be treated with a reagent suitable for carrying out a beta elimination reaction, a delta elimination reaction, or a combination thereof. In some cases, the cleavage of input nucleic acid template comprising one or more abasic sites by chemical means may provide fragments of input nucleic acid template, which fragments comprise a blocked 3' end. In some cases, the blocked 3' end lacks a terminal hydroxyl. In other cases, the blocked 3' end is phosphorylated. In still other cases, cleavage of the input nucleic acid template comprising one or more abasic sites by chemical means may provide fragments of input nucleic acid template that are not blocked. In some cases, the methods of the present invention provide for the use of an enzyme or combination of enzymes and a polyamine such as DMED under mild conditions in a single reaction mixture which does not affect the canonical nucleotides and therefore may maintain the sequence integrity of the products of the method. Suitable mild conditions may include conditions at or near neutral pH. Other suitable conditions include pH of about 4.5 or higher, 5 or higher, 5.5 or higher, 6 or higher, 6.5 or higher, 7 or higher, 7.5 or higher, 8 or higher, 8.5 or higher, 9 or higher, 9.5 or higher, 10 or higher, or about 10.5 or higher. Still other suitable conditions include between about 4.5 and 10.5, between about 5 and 10.0, between about 5.5 and 9.5, between about 6 and 9, between about 6.5 and 8.5, between about 6.5 and 8.0, or between about 7 and 8.0. Suitable mild conditions also may include conditions at or near room temperature. Other suitable conditions include a temperature of about 10°C, 11°C, 12°C, 13°C, 14°C, 15°C, 16°C, 17°C, 18°C, 19°C, 20°C, 21°C, 22°C, 23°C, 24°C, 25°C, 26°C, 27°C, 28°C, 29°C, 30°C, 31°C, 32°C, 33°C, 34°C,

IDT Exhibit 1020.5384

35°C, 36°C, 37°C, 38°C, 39°C, 40°C, 41°C, 42°C, 43°C, 44°C, 45°C, 46°C, 47°C, 48°C, 49°C, 50°C, 51°C, 52°C, 53°C, 54°C, 55°C, 56°C, 57°C, 58°C, 59°C, 60°C, 61°C, 62°C, 63°C, 64°C, 65°C, 66°C, 67°C, 68°C, 69°C, or 70°C or higher. Still other suitable conditions include between about 10°C and about 70°C , between about 15°C and about 65°C , between about 20°C and about 60°C , between about 20°C and about 55°C , between about 20°C and about 50°C, between about 20°C and about 45°C , between about 20°C and about 40°C , between about 20°C and about 35°C , or between about 20°C and about 30°C. In some cases, the use of mild cleavage conditions may provide for less damage to the primer extension products produced by the methods of the present invention. In some cases, the fewer damaged bases, the more suitable the primer extension products may be for downstream analysis such as sequencing, or hybridization. In other cases, the use of mild cleavage conditions may increase final product yields, maintain sequence integrity, or render the methods of the present invention more suitable for automation.

[0080] In embodiments involving fragmentation, the backbone of the template polynucleotide comprising the abasic site is cleaved at the abasic site, whereby two or more fragments of the polynucleotide are generated. At least one of the fragments comprises an abasic site, as described herein. Agents that cleave the phosphodiester backbone of a polynucleotide at an abasic site are provided herein. In some embodiments, the agent is an AP endonuclease such as E. coli AP endonuclease IV. In other embodiments, the agent is N,N′-dimethylethylenediamine (termed "DMED"). In other embodiments, the agent is heat, basic condition, acidic conditions, or an alkylating agent. In still other embodiments, the agent that cleaves the phosphodiester backbone at an abasic site is the same agent that cleaves the base portion of a nucleotide to form an abasic site. For example, glycosidases of the present invention may comprise both a glycosidase and a lyase activity, whereby the glycosidase activity cleaves the base portion of a nucleotide (e.g., a non-canonical nucleotide) to form an abasic site and the lyase activity cleaves the phosphodiester backbone at the abasic site so formed. In some cases, the glycosidase comprises both a glycosidase activity and an AP endonuclease activity.

[0081] Depending on the agent employed for cleaving at the abasic site of the template polynucleotide, the backbone can be cleaved 5′ to the abasic site (e.g., cleavage between the 5′-phosphate group of the abasic residue and the deoxyribose ring of the adjacent nucleotide, generating a free 3′ hydroxyl group), such that an abasic site is located at the 5′ end of the resulting fragment. In other embodiments, cleavage can also be 3′ to the abasic site (e.g., cleavage between the deoxyribose ring and 3′-phosphate group of the abasic residue and the

IDT Exhibit 1020.5385

deoxyribose ring of the adjacent nucleotide, generating a free 5′ phosphate group on the deoxyribose ring of the adjacent nucleotide), such that an abasic site is located at the 3′ end of the resulting fragment. In still other embodiments, more complex forms of cleavage are possible, for example, cleavage such that cleavage of the phosphodiester backbone and cleavage of a portion of the abasic nucleotide results. Selection of the fragmentation agent thus permits control of the orientation of the abasic site within the polynucleotide fragment, for example, at the 3′ end of the resulting fragment or the 5′ end of the resulting fragment. Selection of reaction conditions also permits control of the degree, level or completeness of the fragmentation reactions. In some embodiments, reaction conditions can be selected such that the cleavage reaction is performed in the presence of a large excess of reagents and allowed to run to completion with minimal concern about cleavage of the primer extension products of the invention. By contrast, other methods known in the art, e.g., mechanical shearing, DNase cleavage, cannot distinguish between the template polynucleotide and the primer extension products. In other embodiments, reaction conditions are selected such that fragmentation is not complete (in the sense that the backbone at some abasic sites remains uncleaved (unfragmented)), such that polynucleotide fragments comprising more than one abasic site are generated. Such fragments comprise internal (nonfragmented) abasic sites.

[0082] Following generation of an abasic site by cleavage of the base portion of the non-canonical nucleotide if present in the polynucleotide, the backbone of the polynucleotide is cleaved at the site of incorporation of the non-canonical nucleotide (also termed the abasic site, following cleavage of the base portion of the non-canonical nucleotide) with an agent capable of effecting cleavage of the backbone at the abasic site. Cleavage at the backbone (also termed "fragmentation") results in at least two fragments (depending on the number of abasic sites present in the polynucleotide comprising an abasic site, and the extent of cleavage).

[0083] Suitable agents (for example, an enzyme, a chemical and/or reaction conditions such as heat) capable of cleavage of the backbone at an abasic site include: heat treatment and/or chemical treatment (including basic conditions, acidic conditions, alkylating conditions, or amine mediated cleavage of abasic sites, (see e.g., McHugh and Knowland, Nucl. Acids Res. (1995) 23(10):1664-1670; Bioorgan. Med. Chem (1991) 7:2351; Sugiyama, Chem. Res. Toxicol. (1994) 7: 673-83; Horn, Nucl. Acids. Res., (1988) 16:11559-71), and use of enzymes that catalyze cleavage of polynucleotides at abasic sites, for example AP endonucleases (also called "apurinic, apyrimidinic endonucleases") (e.g., E. coli Endonuclease IV, available from Epicentre Tech., Inc, Madison Wis.), E. coli endonuclease

IDT Exhibit 1020.5386

III or endonuclease IV, E. coli exonuclease III in the presence of calcium ions. See, e.g. Lindahl, PNAS (1974) 71(9):3649-3653; Jendrisak, U.S. Pat. No. 6,190,865 B1; Shida, Nucleic Acids Res. (1996) 24(22):4572-76; Srivastava, J. Biol Chem. (1998) 273(13):21203-209; Carey, Biochem. (1999) 38:16553-60; Chem Res Toxicol (1994) 7:673-683. As used herein "agent" encompasses reaction conditions such as heat. In one embodiment, the AP endonuclease, E. coli endonuclease IV, is used to cleave the phosphodiester backbone at an abasic site. In another embodiment, cleavage is with an amine, such as N, N′-dimethylethylenediamine. See, e.g., McHugh and Knowland, supra.

[0084] Cleavage of the abasic site may occur between the nucleotide immediately 5′ to the abasic residue and the abasic residue, or between the nucleotide immediately 3′ to the abasic residue and the abasic residue (though, as explained herein, 5′ or 3′ cleavage of the phosphodiester backbone may or may not result in retention of the phosphate group 5′ or 3′ to the abasic site, respectively, depending on the fragmentation agent used). Cleavage can be 5′ to the abasic site (such as endonuclease IV treatment which generally results in cleavage of the backbone at a location immediately 5′ to the abasic site between the 5′-phosphate group of the abasic residue and the deoxyribose ring of the adjacent nucleotide, generating a free 3′ hydroxyl group on the adjacent nucleotide), such that an abasic site is located at the 5′ end of the resulting fragment. Cleavage can also be 3′ to the abasic site (e.g., cleavage between the deoxyribose ring and 3′-phosphate group of the abasic residue and the deoxyribose ring of the adjacent nucleotide, generating a free 5′ phosphate group on the deoxyribose ring of the adjacent nucleotide), such that an abasic site is located at the 3′ end of the resulting fragment. Treatment under basic conditions or with amines (such as N,N′-dimethylethylenediamine) results in cleavage of the phosphodiester backbone immediately 3′ to the abasic site. In addition, more complex forms of cleavage are also possible, for example, cleavage such that cleavage of the phosphodiester backbone and cleavage of (a portion of) the abasic nucleotide results. For example, under certain conditions, cleavage using chemical treatment and/or thermal treatment may comprise a β-elimination step which results in cleavage of a bond between the abasic site deoxyribose ring and its 3′ phosphate, generating a reactive α,β-unsaturated aldehyde which can be labeled or can undergo further cleavage and cyclization reactions. See, e.g., Sugiyama, Chem. Res. Toxicol . (1994) 7: 673-83; Horn, Nucl. Acids. Res., (1988) 16:11559-71. It is understood that more than one method of cleavage can be used, including two or more different methods which result in multiple, different types of cleavage products (e.g., fragments comprising an abasic site at the 3′ end, and fragments comprising an abasic site at the 5′ end).

IDT Exhibit 1020.5387

[0085] Cleavage of the backbone at an abasic site may be general, specific or selective (in the sense that the agent (such as an enzyme) capable of cleaving the backbone at an abasic site specifically or selectively cleaves the base portion of a particular non-canonical nucleotide), whereby greater than about 98%, about 95%, about 90%, about 85%, or about 80% of the cleavage is at an abasic site. However, extent of cleavage can be less. Thus, reference to specific cleavage is exemplary. General, specific or selective cleavage is desirable for control of the fragment size in the methods of generating labeled polynucleotide fragments of the invention. In some embodiments, reaction conditions can be selected such that the cleavage reaction is performed in the presence of a large excess of reagents and allowed to run to completion with minimal concern about excessive cleavage of the polynucleotide (i.e., while retaining a desired fragment size, which is determined by spacing of the incorporated non-canonical nucleotide, during the synthesis step, above). In other embodiments, extent of cleavage can be less, such that polynucleotide fragments are generated comprising an abasic site at an end and an abasic site(s) within or internal to the polynucleotide fragment (i.e., not at an end).

[0086] In embodiments involving cleavage of the phosphodiester backbone, appropriate reaction media and conditions for carrying out the cleavage of the phosphodiester backbone at an abasic site according to the methods of the invention are those that permit cleavage of the phosphodiester backbone at an abasic site. Such media and conditions are known to persons of skill in the art, and are described in various publications, such as Bioorgan. Med. Chem (1991) 7:2351; Sugiyama, Chem. Res. Toxicol . (1994) 7: 673-83; Horn, Nucl. Acids. Res., (1988) 16:11559-71); Lindahl, PNAS (1974) 71(9):3649-3653; Jendrisak, U.S. Pat. No. 6,190,865 B1; Shida, Nucleic Acids Res. (1996) 24(22):4572-76; Srivastava, J. Biol Chem. (1998) 273(13):21203-209; Carey, Biochem. (1999) 38:16553-60; Chem Res Toxicol (1994) 7:673-683.

[0087] In some cases, nucleic acids containing abasic sites are heated in a buffer solution containing an amine, for example, 25 mM Tris-HCl and 1-5 mM magnesium ions, for 10-30 minutes at 70°C. to 95°C. Alternatively, 1.0 M piperidine (a base) is added to polynucleotide comprising an abasic site which has been precipitated with ethanol and vacuum dried. The solution is then heated for 30 minutes at 90°C. and lyophilized to remove the piperidine. In another example, cleavage is effected by treatment with basic solution, e.g., 0.2 M sodium hydroxide at 37°C for 15 minutes. See Nakamura (1998) Cancer Res. 58:222-225. In yet another example, incubation at 37°C. with 100 mM N,N′-dimethylethylenediamine acetate,

IDT Exhibit 1020.5388

pH 7.4 is used to cleave. See McHugh and Knowland, (1995) Nucl. Acids Res. 23(10) 1664-1670.

[0088] The cleavage of the input nucleic acid template comprising one or more abasic sites may also be performed by enzymatic means. For example an apyrimidinic endonuclease or an apurinic endonuclease (collectively known as AP endonucleases) may be used to cleave the input nucleic acid template at the one or more abasic sites. In some cases, the input nucleic acid template comprising one or more abasic sites may be cleaved with a class I, class II, class III, or class IV AP endonuclease or a combination thereof. In some cases, the cleavage of input nucleic acid template comprising one or more abasic sites by enzymatic means may provide fragments of input nucleic acid template, which fragments comprise a blocked 3' end. In some cases, the blocked 3' end lacks a terminal hydroxyl. In other cases, the blocked 3' end is phosphorylated. In still other cases, cleavage of the input nucleic acid template comprising one or more abasic sites by enzymatic means may provide fragments of input nucleic acid template that are not blocked.

[0089] In some cases, the cleavage may be performed by use of a glycosylase and a nucleophile, or a glycosylase and an amine, or a glycosylase and an AP endonuclease such as for example UDG and DMED or UDG and an AP endonuclease at the same time. Alternatively, the input nucleic acid template comprising one or more non-canonical nucleotides may first be treated with a glycosylase to produce one or more abasic sites, and then be treated with an AP endonuclease or cleaved by chemical means. In some cases, the hybridization, and extension reactions are performed first, and then the cleavage reaction is performed after sufficient time. In other cases, the hybridization and extension reactions are performed at the same time as the cleavage reactions. In still other cases, the hybridization and extension reactions are initiated and allowed to proceed for a set period of time (e.g., 1 minute, 2 minutes, 3 minutes, 5 minutes, 10 minutes, 15 minutes, 30 minutes, 1 hour, 2 hours, 3 hours etc.) and then the cleavage reaction is initiated. In some cases, initiation of the cleavage reaction may stop the extension reaction; in other cases, the cleavage reaction and the extension reaction may then proceed concurrently.

[0090] For example, *E. coli* AP endonuclease IV may be added to reaction conditions as described above. AP Endonuclease IV can be added at the same or different time as the agent (such as an enzyme) capable of cleaving the base portion of a non-canonical nucleotide. For example, AP Endonuclease IV can be added at the same time as UNG, or at different times. Alternatively, the template nucleic acid or a reaction mixture comprising template nucleic acid may be treated with UNG and an amine at the same time. A reaction mixture suitable

IDT Exhibit 1020.5389

for simultaneous UNG treatment and N,N′-dimethylethylenediamine treatment may include about 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, 35, 40, or about 50mM DMED. Alternatively, the use of an agent that comprises both glycosidase and lyase activity may be utilized in the reaction mixture to cleave the input nucleic acid template.

[0091] Cleavage of the input nucleic acid template by chemical means, enzymatic means, or a combination thereof may provide a mixture of double stranded products, single stranded products, and partial duplexes. In some cases, the cleaved products of the cleavage reaction may be removed by one or more methods of the present invention. In some cases, the cleaved products of the cleavage reaction may be removed by purification. For example, the cleaved products of the cleavage reaction may be removed by a size-dependent purification method, or an affinity based purification method. For example, the single stranded nucleic acids may be removed by an affinity hybridization step to capture probes. In some cases, the capture probes may be hybridized to a solid substrate. In other cases, the cleaved nucleic acid products of the cleavage reaction may be removed by an affinity capture step using a ligand with affinity to a label that has been incorporated into the cleaved products of the cleavage reaction. The label, or ligand, may be incorporated prior to cleavage (e.g. during synthesis of the template nucleic acid), during cleavage, or after the cleavage step. In some cases, the label may be incorporated at the abasic site. In other cases, the cleaved nucleic acid products of the cleavage reaction may be removed by a capture step using a reactive moiety (e.g., an amine or a hydrazine) such as an immobilized reactive moiety that reacts with a reactive α,β-unsaturated aldehyde present at the abasic site of the cleaved nucleic acid product of the cleavage reaction. In some cases, the cleaved nucleic acid products of the cleavage reaction may be removed by electrophoresis or ultrafiltration.

[0092] In other cases, the single stranded products may be removed by enzymatic means. For example, a single stranded specific exonuclease or endonuclease can be used to cleave the single stranded DNA. A variety of suitable single stranded DNA specific exonucleases are suitable for the methods of the present invention such as for example exonuclease 1, and exonuclease 7. Similarly a variety of suitable single stranded DNA specific endonucleases are suitable for the methods of the present invention such as for example the single stranded DNA specific endonuclease is a S1 endonuclease or a mung bean nuclease . In some cases, any combination of single strand specific endonucleases or exonucleases known in the art such as those provided herein may be utilized to degrade or remove single stranded products,

IDT Exhibit 1020.5390

such as single stranded fragmentation products or single stranded primer extension products or a combination thereof.

[0093] In some cases, the products of the primer extension reaction generated in the methods of the present invention may be purified from the reaction mixture comprising fragmented target nucleic acid and primer extension products. For example, the primer extension step may include the use of nucleotides comprising a purification label such as for example biotin/avidin or any other suitable label (e.g. digoxin, fluorescein, an antigen, a ligand, a receptor, or any nucleotide labels provided herein). Primer extension products may therefore be understood to contain a member of the biotin/avidin ligand receptor pair or other purification label, whereas primers and template nucleic acid may not. A simple purification step may be performed to remove unincorporated nucleotides such as alcohol or polyethylene glycol precipitation, ion exchange purification, ultrafiltration, silica absorption, or reverse phase methods, and then the primer extension products may be recovered using an appropriate affinity matrix such as a matrix comprising biotin or a derivative thereof, avidin or a derivative thereof, streptavidin or a derivative thereof, an antibody or a derivative or fragment thereof, an antigen, a ligand, or a receptor in the form of particles, beads, a membrane or a column. Alternatively, the simple purification step to remove unincorporated nucleotides may be omitted or performed after the affinity purification step.

[0094] In some embodiments, the methods of the present invention further provide for the generation of one or more blunt ended double stranded products. In some embodiments the blunt ended double stranded products are produced from a template not containing any non-canonical nucleotides. In other embodiments the double stranded products are produced from a template containing one or more non-canonical nucleotides. In some cases, the extension step of the present invention directly provides blunt ended double stranded products. In other cases, the extension step of the present invention provides a mixture of blunt ended and non-blunt ended double stranded products. In still other cases, the extension step does not provide blunt ended double stranded products, or does not provide a substantial degree or amount of blunt ended double stranded products. In some cases, the non-blunt ended products of the primer extension reaction must be further treated by the methods of the present invention to produce blunt ended double stranded products, or to convert a substantial fraction of the non-blunt ended products to blunt ended products.

[0095] In some cases, the double stranded products generated by the method of the present invention may be blunt ended, when blunt end dsDNA is desirable for downstream analysis such as highly parallel sequencing, or other cloning or adaptor ligation applications, by the

IDT Exhibit 1020.5391

use of a single strand specific DNA exonuclease such as for example exonuclease 1, exonuclease 7 or a combination thereof to degrade overhanging single stranded ends of the double stranded products. Alternatively, the double stranded products may be blunt ended by the use of a single stranded specific DNA endonuclease for example but not limited to mung bean endonuclease or S1 endonuclease. Alternatively, the double stranded fragment products may be blunt ended by the use of a polymerase that comprises single stranded exonuclease activity such as for example T4 DNA polymerase, any other polymerase comprising single stranded exonuclease activity or a combination thereof to degrade the overhanging single stranded ends of the double stranded products. In some cases, the polymerase comprising single stranded exonuclease activity may be incubated in a reaction mixture that does or does not comprise one or more dNTPs. In other cases, a combination of single stranded nucleic acid specific exonucleases and one or more polymerases may be used to blunt end the double stranded products of the primer extension reaction. In still other cases, the products of the extension reaction may be made blunt ended by filling in the overhanging single stranded ends of the double stranded products. For example, the fragments may be incubated with a polymerase such as T4 DNA polymerase or Klenow polymerase or a combination thereof in the presence of one or more dNTPs to fill in the single stranded portions of the double stranded products. Alternatively, the double stranded products may be made blunt by a combination of a single stranded overhang degradation reaction using exonucleases and/or polymerases, and a fill-in reaction using one or more polymerases in the presence of one or more dNTPs.

[0096] In some embodiments, the methods of the present invention provide for generation of primer extension products comprising double stranded nucleic acids, single stranded nucleic acids, and nucleic acids comprising partial double stranded and partial single stranded portions, either from a template not comprising any non-canonical nucleotides or from a template nucleic acid comprising one or more non-canonical nucleotides; fragmentation of the template nucleic acid; optional purification of the primer extension products; and generation of double stranded products from the single stranded nucleic acid primer extension products and/or from the primer extension products comprising partial double stranded and partial single stranded portions. Methods for generation of double stranded products from partial double stranded products are provided herein including the methods for blunt ending double stranded primer extension products. Methods for generation of double stranded primer extension products from single stranded primer extension products include for example annealing one or more primers, such as any of the primers provided herein, to the

IDT Exhibit 1020.5392

single stranded primer extension product and extending the one or more annealed primers with a polymerase, such as any of the polymerases provided herein or a any suitable polymerase in a reaction mixture comprised of one or more dNTPs, including labeled dNTPs, canonical dNTPs, non-canonical dNTPs or a combination thereof. In some cases, the non-canonical nucleotides utilized in the reaction mixture for generating double stranded products from single stranded primer extension products or from partial double stranded products are different from at least one of the non-canonical nucleotides present in the template polynucleotide. Methods of generation of double stranded primer extension products from single stranded primer extension products may further include for example annealing two or more adjacent primers, such as any of the primers provided herein including random primers (e.g. pentamers, hexamers, heptamers, octamers, nonamers, decamers, undecamers, dodecamers, tridecamers etc.), to the single stranded primer extension product and ligating the adjacent primers. Methods for generating double stranded primer extension products from single stranded primer extension products may further include for example annealing one or more primers such as any of the primers provided herein including primers comprising random hybridizing portions (e.g. random pentamers, hexamers, heptamers, octamers, nonamers, decamers, undecamers, dodecamers, tridecamers etc.) to the single stranded primer extension product and extending the annealed primers. In some cases, the extension step may be performed using an enzyme (e.g., a DNA dependent DNA polymerase) comprising strand displacement activity.

[0097] In some embodiments, the methods of the present invention provide for attachment (e.g., ligation) of adaptor molecules to the double stranded DNA products of the primer extension reaction, or double stranded products generated from the single stranded or partially double stranded products of the primer extension reaction. The adaptor molecules may be ligated to double stranded DNA fragment molecules comprising single stranded overhangs, including but not limited to single, double, triple, quadruple, quintuple, sextuple, septuple, octuple, or more base overhangs, or to double stranded DNA fragment molecules comprising blunt ends. In some cases, the adaptor molecules are ligated to blunt end double stranded DNA fragment molecules which have been modified by 5' phosphorylation. In some cases, the adaptor molecules are ligated to blunt end double stranded DNA fragment molecules which have been modified by 5' phosphorylation followed by extension of the 3' end with one or more nucleotides. In some cases, the adaptor molecules are ligated to blunt end double stranded DNA fragment molecules which have been modified by 5' phosphorylation followed by extension of the 3' end with a single nucleotide (or 2, 3, 4, 5, 6,

IDT Exhibit 1020.5393

7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 or more) such as for example adenine, guanine, cytosine, or thymine. In still other cases, adaptor molecules can be ligated 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. 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.

[0098] The adaptor molecules may comprise single or double stranded nucleic acids or a combination thereof. In some cases, the adaptor molecules comprise a one, two, three, four, five, six, seven, eight, nine, ten, eleven, twelve, thirteen, fourteen, fourteen, fifteen, sixteen, seventeen, eighteen, nineteen, twenty or longer base long single stranded overhang at their 5' ends. For example, the adaptor molecules may comprise a one base long thymine, adenine, cytosine, or guanine overhang at their 5' ends. Adaptor molecule compositions are provided herein.

[0099] In some embodiments, the methods of the present invention provide for ligation or attachment of adaptor molecules to the single stranded DNA products of the extension reaction. The adaptor molecules may comprise single stranded or double stranded nucleic acids or a combination thereof. The adaptor molecules may be ligated to the single stranded DNA products of the extension reaction using T4 RNA ligase which is capable of ligating two single stranded nucleic acids (RNA or DNA) together in the absence of a template. Alternatively, a single stranded DNA specific ligase such as for example CircLigase® may be utilized in the methods of the present invention.

[00100] In some embodiments, the methods of the present invention provide for contacting an input nucleic acid template comprising one or more non-canonical nucleotides with a reaction mixture. In some cases, the reaction mixture may comprise one or more oligonucleotide primers as provided herein. For example, the reaction mixture may comprise one or more oligonucleotide primers comprising random hybridizing portions. Additionally, the reaction mixture may comprise one or more oligonucleotide primers comprising random hybridizing portions and one or more oligonucleotide primers comprising a polyT sequence.

[00101] In some cases, the reaction mixture may comprise one or more polymerases as provided herein. For example, the reaction mixture may comprise one or more polymerases

IDT Exhibit 1020.5394

comprising strand displacement activity, such as for example, Klenow polymerase, exo – Klenow polymerase, 5'-3' exo – Klenow polymerase, Bst polymerase, Bst large fragment polymerase, Vent polymerase, Deep Vent (exo-) polymerase, 9°Nm polymerase, Therminator polymerase, Therminator II polymerase, MMulV Reverse Transcriptase, phi29 polymerase, or DyNAzyme EXT polymerase, or a combination thereof. In some cases, the reaction mixture may be configured to provide double stranded products in the presence of the input nucleic acid template, the one or more oligonucleotide primers, and the one or more polymerases comprising strand displacement activity. Enzymes for use in the compositions, methods and kits of the present invention may further include any enzyme having reverse transcriptase activity. Such enzymes include, but are not limited to, retroviral reverse transcriptase, retrotransposon reverse transcriptase, hepatitis B reverse transcriptase, cauliflower mosaic virus reverse transcriptase, bacterial reverse transcriptase, E. coli DNA polymerase and klenow fragment, Tth DNA polymerase, Taq DNA polymerase (Saiki, R. K., et al., Science 239:487-491 (1988); U.S. Pat. Nos. 4,889,818 and 4,965,188), Tne DNA polymerase (WO 96/10640), Tma DNA polymerase (U.S. Pat. No. 5,374,553), C. Therm DNA polymerase from Carboxydothermus hydrogenoformans (EP0921196A1, Roche, Pleasanton, Calif., Cat. No. 2016338), ThermoScript (Invitrogen, Carsbad, Calif. Cat. No. 11731-015) and mutants, fragments, variants or derivatives thereof. As will be understood by one of ordinary skill in the art, modified reverse transcriptases may be obtained by recombinant or genetic engineering techniques that are routine and well-known in the art. Mutant reverse transcriptases can, for example, be obtained by mutating the gene or genes encoding the reverse transcriptase of interest by site-directed or random mutagenesis. Such mutations may include point mutations, deletion mutations and insertional mutations. Preferably, one or more point mutations (e.g., substitution of one or more amino acids with one or more different amino acids) are used to construct mutant reverse transcriptases of the invention. Fragments of reverse transcriptases may be obtained by deletion mutation by recombinant techniques that are routine and well-known in the art, or by enzymatic digestion of the reverse transcriptase(s) of interest using any of a number of well-known proteolytic enzymes. Mutant DNA polymerase containing reverse transcriptase activity can also be used as described in U.S. Patent Application Ser. No. 10/435,766, incorporated herein by reference.

[00102] In some cases, the reaction mixture may comprise one or more agents capable of cleaving the base portion of a non-canonical nucleotide to generate an abasic site. In some cases, the reaction mixture may contain the one or more agents capable of cleaving the base

IDT Exhibit 1020.5395

portion of a non-canonical nucleotide to generate an abasic site at the initiation of the extension reaction. In some cases, the reaction mixture may be supplemented with the one or more agents capable of cleaving the base portion of a non-canonical nucleotide to generate an abasic site after a suitable period of time (e.g., about 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 45, 60, 90, 120, 180, 240, 300, 400, 500, 600 minutes) has passed for the generation of primer extension products. Suitable agents capable of cleaving the base portion of a non-canonical nucleotide to generate an abasic site include but are not limited to UDG and MPG.

[00103] In some cases, the reaction mixture may comprise one or more agents capable of fragmenting a phosphodiester backbone at an abasic site to fragment the input nucleic acid template. In some cases, the reaction mixture may contain the one or more agents capable of fragmenting a phosphodiester backbone at an abasic site to fragment the input nucleic acid template at the initiation of the extension reaction. In some cases, the reaction mixture may be supplemented with the one or more agents capable of fragmenting a phosphodiester backbone at an abasic site to fragment the input nucleic acid template after a suitable period of time (e.g., about 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 45, 60, 90, 120, 180, 240, 300, 400, 500, 600 minutes) has passed for the generation of primer extension products. Suitable agents capable of fragmenting a phosphodiester backbone at an abasic site to fragment the input nucleic acid template include but are not limited to an amine, a primary amine, a secondary amine, a polyamine as provided herein, a nucleophile, a base (e.g. NaOH), piperidine, hot piperidine, and one or more AP endonucleases.

[00104] The methods of the present invention provide for downstream analysis of the primer extension products generated in the methods of the present invention. Said downstream analysis includes but is not limited to e.g. pyrosequencing, sequencing by synthesis, sequencing by hybridization, single molecule sequencing, nanopore sequencing, and sequencing by ligation, high density PCR, microarray hybridization, SAGE, digital PCR, and massively parallel Q-PCR; subtractive hybridization; differential amplification; comparative genomic hybridization, preparation of libraries (including cDNA and differential expression libraries); preparation of an immobilized nucleic acid (which can be a nucleic acid immobilized on a microarray), and characterizing amplified nucleic acid products generated by the methods of the invention, or a combination thereof.

Applications on single cells

[00105] Single cell sequencing and gene expression profiling is provided for a variety of suitable methods known in the art, such as disease diagnostic or prognostic applications, as well as a research tool, for example to identify novel drug targets. Diseases of interest

IDT Exhibit 1020.5396

include, without limitation, immune-mediated dysfunction, cancer, and the like. In the methods of the invention, a heterogeneous cell mixture, e.g. a tumor needle biopsy, inflammatory lesion biopsy, synovial fluid, spinal tap, etc., is divided randomly or in a certain order into spatially separated single cells, e.g. into a multiwell plate, microarray, microfluidic device, or slide. Cells are then lysed, and the contents amplified and individually analyzed for sequencing or expression of genes of interest. The cells thus analyzed can be classified according to the genetic signatures of individual cells. Such classification allows an accurate assessment of the cellular composition of a test sample, which assessment may find use, for example, in determining the identity and number of cancer stem cells in a tumor; in determining the identity and number of immune-associated cells such as the number and specificity of T cells, dendritic cells, B cells and the like.

[00106] In some embodiments, the cell sample to be analyzed is a primary sample, which may be freshly isolated, frozen, etc. However, cells to be analyzed can be cultured cells. Usually the sample is a heterogeneous mixture of cells, comprising a plurality of distinct cell types, distinct populations, or distinct subpopulations, for example 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 or more cell types, populations, or subpopulations . In some embodiments the sample is a cancer sample from a solid tumor, leukemia, lymphoma, etc., which may be a biopsy, e.g. a needle biopsy, etc., a blood sample for disseminated tumors and leukemias, and the like. Samples may be obtained prior to diagnosis, may be obtained through a course of treatment, and the like.

[00107] For isolation of cells from tissue, an appropriate solution can be used for dispersion or suspension. Such solution will generally be a balanced salt solution, e.g. normal saline, PBS, Hank's balanced salt solution, etc., conveniently supplemented with fetal calf serum or other naturally occurring factors, in conjunction with an acceptable buffer at low concentration, generally from 5-25 mM. Convenient buffers include HEPES, phosphate buffers, lactate buffers, etc. The separated cells can be collected in any appropriate medium that maintains the viability of the cells, usually having a cushion of serum at the bottom of the collection tube. Various media are commercially available and may be used according to the nature of the cells, including dMEM, HBSS, dPBS, RPMI, Iscove's medium, etc., frequently supplemented with fetal calf serum.

[00108] Systems such as Beckman MoFlo cell sorter, Becton Dickenson Influx, or Bio-Rad S3 can be used to sort heterogeneous mixtures of cells based on surface markers, size, etc. into distinct populations.

IDT Exhibit 1020.5397

[00109] In some embodiments, cells in a sample are separated on a microarray. For example, a highly integrated live-cell microarray system may utilize microwells each of which is just large enough to fit a single cell (see Tokimitsu et αl. (2007) Cytometry Part A 71k 1003:1010; and Yamamura et αl. (2005) Analytical Chemistry 77:8050; each herein specifically incorporated by reference). Prior enrichment of cells of interest - such as by FACS or other sorting - is optional and in some embodiments, cells from a sample are divided into discrete locations without any prior sorting or enrichment. For example, cells from a sample (e.g., blood sample, biopsy, solid tumor) can be individually isolated into distinct positions. Typically, for solid tissue samples, the samples are mechanically, chemically, and/or enzymatically separated (e.g., by treatment with trypsin or sonication). Cells from a sample can be placed into any cell sorting device (e.g., a microfruidic cell sorter) such that individual cells are isolated, such as at an addressable position on a planar surface. Planar surfaces can have indentations, barriers or other features ensuring isolation of individual cells. Isolated cells can then be analyzed according to the methods herein. Preferably, cells are separated into distinct positions wherein each position contains 1 or 0 cells.

[00110] Cells are optionally sorted, e.g. by flow cytometry, prior to the separation. For example, FACS sorting or size-differential sorting, can be used to increase the initial concentration of the cells of interest by at least 1,000, 10,000, 100,000, or more fold, according to one or more markers present on the cell surface. Such cells are optionally sorted according to the presence and/or absence of cell surface markers particularly markers of a population or subpopulation of interest.

Cell sorters

[00111] Where the cells are isolated into distinct positions for analysis, the cells may be sorted with a microfluidic sorter, by flow cytometry, microscopy, etc. A microfabricated fluorescence-activated cell sorter is described by Fu et al. (1999) Nature Biotechnology 17: 1109 and Fu et al. (2002) Anal. Chem. 74:2451-2457, each herein specifically incorporated by reference. A sample can be sorted with an integrated microfabricated cell sorter using multilayer soft lithography. This integrated cell sorter may incorporate various microfluidic functionalities, including peristaltic pumps, dampers, switch valves, and input and output wells, to perform cell sorting in a coordinated and automated fashion. The active volume of an actuated valve on this integrated cell sorter can be as small as 1 pL, and the volume of optical interrogation as small as 100 fL. Compared with conventional FACS machines, the microfluidic FACS provides higher sensitivity, no cross-contamination, and lower cost.

IDT Exhibit 1020.5398

[00112] Individual cells can be isolated into distinct positions (e.g., a 96-well plate or a microarray address) for further analysis and/or manipulation. For example, a cell population containing a desired cell type, such as, hematopoietic stem cells (HSCs) is sorted by FACS analysis utilizing antibodies capable of distinguishing HSCs from mature cells. The cells are sorted into 96-well plates, lysed by appropriate methods and the lysates are analyzed by qPCR, microarray analysis, and/or sequencing.

[00113] Devices for single cell isolation include a microfluidic cell sorter, which isolates live cells from cellular debris and sorts cells from a single cell suspension. Microfluidic devices can be used in combination with fluorescent signals (e.g., labeled antibodies to markers for a target population or subpopulation) from 1, 2, 3, 4, 5 or more different surface markers, and places them in individual bins for subsequent genetic studies. Other upstream steps such as digesting the tumor or cell culture to obtain a cell suspension and staining the cells with fluorescent surface markers may be incorporated in this system. The number of cells to be analyzed depends on the heterogeneity of the sample, and the expected frequency of cells of interest in the sample. Usually at least about 10^2 cells are analyzed, at least about 10^3, at least 5 x 10^3, at least about 10^4, at least about 10^5, at least about 10^6, at least about 10^7, at least about 10^8, at least about 10^9, at least about 10^{10}, at least about 10^{11}, at least about 10^{12}, at least about 10^{13}, at least about 10^{14}, at least about 10^{15}, or more cells are analyzed.

[00114] In some instances, a single cell analysis device (SCAD) is modular and can perform multiple steps, such as digestion of the tissue, separation of live cells from the debris, staining, or sorting in an integrated, fully automated fashion.

[00115] Sorted cells can be individually lysed to perform analysis of genetic (RNA, DNA) and/or protein composition of the cells. mRNA can be captured on a column of oligo-dT beads, reverse transcribed on beads, processed off chip, transferred to a macroscopic well, etc. Optionally, DNA or RNA is preamplified prior to analysis. Preamplification can be of an entire genome or transcriptome, or a portion thereof (e.g., genes/transcripts of interest). A polynucleotide sample can be transferred to a chip for analysis (e.g., by qRT-PCR) and determination of an expression profile.

[00116] A nucleic acid sample includes a plurality or population of distinct nucleic acids that can include the expression information of the phenotype determinative genes of interest of the individual cell. A nucleic acid sample can include RNA or DNA nucleic acids, e.g., mRNA, cRNA, cDNA, etc. Expression profiles can be generated by any convenient means for determining differential gene expression between two samples, e.g. quantitative

IDT Exhibit 1020.5399

hybridization of mRNA, labeled mRNA, amplified mRNA, cRNA, etc., quantitative PCR, and the like. A subject or patient sample, e.g., cells or collections thereof, e.g., tissues, is assayed. Samples are collected by any convenient method, as known in the art.

[00117] The sample can be prepared in a number of different ways, as is known in the art, e.g., by mRNA isolation from a single cell, where the isolated mRNA is used as is, amplified, employed to prepare cDNA, cRNA, etc., as is known in the differential expression art (for example, see Marcus, et al., Anal. Chem. (2006); 78(9): 3084-89). The sample can be prepared from any tissue (e.g., a lesion, or tumor tissue) harvested from a subject. Analysis of the samples can be used for any purpose (e.g., diagnosis, prognosis, classification, tracking and/or developing therapy). Cells may be cultured prior to analysis.

[00118] The expression profile may be generated from the initial nucleic acid sample using any conventional protocol. While a variety of different manners of generating expression profiles are known, such as those employed in the field of differential gene expression analysis, one representative and convenient type of protocol for generating expression profiles is quantitative PCR (QPCR, or QT-PCR). Any available methodology for performing QPCR can be utilized, for example, as described in Valera, et al., /. Neurooncol. (2007) 85(l):l-10.

Sorting of cells

[00119] Cells with selected properties, for example cells with selected surface proteins, cells with a disrupted cell membrane, cells infected with a pathogen, dying cells or dead cells can be detected in a sample by a variety of techniques well known in the art, including cell sorting, especially fluorescence-activated cell sorting (FACS), by using an affinity reagent bound to a substrate (e.g., a plastic surface, as in panning), or by using an affinity reagent bound to a solid phase particle which can be isolated on the basis of the properties of the beads (e.g., colored latex beads or magnetic particles). Naturally, the procedure used to detect the cells will depend upon how the cells have been labelled. In one example, any detectable substance which has the appropriate characteristics for the cell sorter may be used (e.g., in the case of a fluorescent dye, a dye which can be excited by the sorter's light source, and an emission spectra which can be detected by the cell sorter's detectors). In flow cytometry, a beam of laser light is projected through a liquid stream that contains cells, or other particles, which when struck by the focussed light give out signals which are picked up by detectors. These signals are then converted for computer storage and data analysis, and can provide information about various cellular properties. Cells labelled with a suitable dye are excited by the laser beam, and emit light at characteristic wavelengths. This emitted light is picked up by

IDT Exhibit 1020.5400

detectors, and these analogue signals are converted to digital signals, allowing for their storage, analysis and display.

[00120] Many larger flow cytometers are also "cell sorters", such as fluorescence-activated cell sorters (FACS), and are instruments which have the ability to selectively deposit cells from particular populations into tubes, or other collection vessels. In a particularly preferred embodiment, the cells are isolated using FACS. This procedure is well known in the art and described by, for example, Melamed et al., Flow Cytometry and Sorting, Wiley-Liss, Inc., (1990); Shapiro, Practical Flow Cytometry, 4th Edition, Wiley-Liss, Inc., (2003); and Robinson et al., Handbook of Flow Cytometry Methods, Wiley-Liss, Inc. (1993).

[00121] In order to sort cells, the instruments electronics interprets the signals collected for each cell as it is interrogated by the laser beam and compares the signal with sorting criteria set on the computer. If the cell meets the required criteria, an electrical charge is applied to the liquid stream which is being accurately broken into droplets containing the cells. This charge is applied to the stream at the precise moment the cell of interest is about to break off from the stream, then removed when the charged droplet has broken from the stream. As the droplets fall, they pass between two metal plates, which are strongly positively or negatively charged. Charged droplets get drawn towards the metal plate of the opposite polarity, and deposited in the collection vessel, or onto a microscope slide, for further examination. The cells can automatically be deposited in collection vessels as single cells or as a plurality of cells, e.g. using a laser, e.g. an argon laser (488 nm) and for example with a Flow Cytometer fitted with an Autoclone unit (Coulter EPICS Altra, Beckman- Coulter, Miami, FIa., USA). Other examples of suitable FACS machines/useful for the methods of the invention include, but are not limited to, MoFlo™ Highspeed cell sorter (Dako-Cytomation ltd), FACS Aria™ (Becton Dickinson), FACS Diva (Becton Dickinson), ALTRA™ Hyper sort (Beckman Coulter) and Cy Flow™ .sorting system (Partec GmbH).

[00122] The enrichment or sorting of desired cells and/or or precursors thereof from a sample may be accomplished using solid-phase particles. Any particle with the desired properties may be utilized. For example, large particles (e.g., greater than about 90-100 μm in diameter) may be used to facilitate sedimentation. In some cases, the particles are "magnetic particles" (i.e., particles which can be collected using a magnetic field). Labeled cells may be retained in a column (held by the magnetic field), whilst unlabelled cells pass straight through and are eluted at the other end. Magnetic particles are now commonly available from a variety of manufacturers including Dynal Biotech (Oslo, Norway) and Milteni Biotech

IDT Exhibit 1020.5401

GmbH (Germany). An example of magnetic cell sorting (MACS) is provided by Al-Mufti et al. (1999).

[00123] Laser-capture microdissection can also be used to selectively enrich labelled dendritic cells or precursors thereof on a slide using methods of the invention. Methods of using laser-capture microdissection are known in the art (see, for example, U.S. 20030227611 and Bauer et al, 2002).

Target polynucleotides

[00124] 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, grape, 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 or choloroplast DNA. The input DNA can also comprise cDNA generated from one or more of cytoplasmic, mitochondrial, or chloroplast mRNA, rRNA, or tRNA 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 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.

[00125] 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

IDT Exhibit 1020.5402

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.

[00126] 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

IDT Exhibit 1020.5403

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.

Single Cells Suitable for Analysis

[00127] Samples containing nucleic acids or single cells can be obtained from biological sources and prepared using conventional methods known in the art. In particular, DNA or RNA useful in the methods described herein can be extracted and/or amplified from any source, including bacteria, protozoa, fungi, viruses, organelles, as well higher organisms such as plants or animals, e.g., mammals, and particularly humans. Suitable nucleic acids can also be obtained from an environmental source (e.g., pond water), from man-made products (e.g., food), from forensic samples, and the like. Nucleic acids can be extracted or amplified from cells, bodily fluids (e.g., blood, a blood fraction, urine, etc.), or tissue samples by any of a variety of standard techniques. Cells may either be cultured or from primary isolates such as clinical samples. Illustrative samples include samples of plasma, serum, spinal fluid, lymph fluid, peritoneal fluid, pleural fluid, oral fluid, and external sections of the skin; samples from the respiratory, intestinal, genital, and urinary tracts; samples of tears, saliva, blood cells, stem cells, or tumors. For example, samples of fetal DNA can be obtained from an embryo (e.g., from one or a few embryonic or fetal cells) or from maternal blood. Samples can be obtained from live or dead organisms or from in vitro cultures. Illustrative samples can include single cells, paraffin-embedded tissue samples, and needle biopsies. Nucleic acids useful in the methods described herein can also be derived from one or more nucleic acid libraries, including cDNA, cosmid, YAC, BAC, P1, PAC libraries, and the like.

[00128] Samples may reflect particular states, e.g., cell proliferation, cell differentiation, cell death, disease, exposure to stimuli, and/or stages, e.g., stages of development.

IDT Exhibit 1020.5404

[00129] In particular embodiments, the methods described herein can carried out on a single cell from a preimplantation embryo, a stem cell, a suspected cancer cell, a cell from a pathogenic organism, and/or a cell obtained from a crime scene. For example, a human blastomere (e.g., from an eight-cell stage embryo or later) can be analyzed to determine whether the genome includes one or more genetic defects.

[00130] Nucleic acids of interest can be isolated using methods well known in the art, with the choice of a specific method depending on the source, the nature of nucleic acid, and similar factors. The sample nucleic acids need not be in pure form, but are typically sufficiently pure to allow the amplification steps of the methods described herein to be performed. Where the target nucleic acids are mRNA, the RNA can be reversed transcribed into cDNA by standard methods known in the art and as described in Sambrook, J., Fritsch, E. F., and Maniatis, T., Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, NY, Vol. 1, 2, 3 (1989), for example. The cDNA can then be analyzed according to the methods described herein.

[00131] In certain embodiments, a single cell can be added directly to a suitable WGA reaction mixture and WGA carried out. In other embodiments, the RNA of a single cell can be converted to DNA (e.g., cDNA) or the RNA directly amplified.

Fragmentation methods

[00132] 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 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

IDT Exhibit 1020.5405

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).

[00133] 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

IDT Exhibit 1020.5406

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.

[00134] 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.

[00135] In some embodiments each of the individual or plurality of samples comprises a single polynucleotide target or a single genome.

[00136] 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.

IDT Exhibit 1020.5407

Methods of Amplification

[00137] The methods, compositions and kits described herein can be useful to generate amplification-ready products for downstream applications such as massively parallel sequencing or hybridization platforms. Methods of amplification are well known in the art. In some embodiments, the amplification is exponential, e.g. in the enzymatic amplification of specific double stranded sequences of DNA by a polymerase chain reaction (PCR). In other embodiments the amplification method is linear. In other embodiments the amplification method is isothermal.

[00138] Thus, it is understood that 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.

[00139] 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.

Amplification

[00140] 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

IDT Exhibit 1020.5408

(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).

[00141] The methods of the present invention may further include a step of hybridizing one or more oligonucleotide primers to an input nucleic acid template. The template can optionally comprise one or more non-canonical nucleotides. In some cases the oligonucleotide primers may comprise a hybridizing portion which comprises random nucleotides, such as for example random dimers, trimers, tetramers, pentamers, hexamers, heptamers, octomers, nonomers, decamers, undecamers, dodecamers, tridecamers, tetradecamers, or longer. In other cases, the hybridizing portion may comprise a non random sequence such as a polyT sequence. In still other cases, the hybridizing portion of some of the oligonucleotide primers may comprise random nucleotides, while the hybridizing portion of some of the nucleotides comprise non-random sequences, such as polyT or "not so random sequences." In some cases, the hybridizing portion of the oligonucleotide primers may comprise "not so random sequences" such as for example a pool of sequences which randomly or pseudo-randomly prime desired sequences such as total mRNA or a substantial fraction thereof, but do not prime undesired sequences such as rRNA.

[00142] A "random primer," as used herein, can be a primer that generally comprises a sequence that is designed not necessarily based on a particular or specific sequence in a sample, but rather is based on a statistical expectation (or an empirical observation) that the sequence of the random primer is hybridizable (under a given set of conditions) to one or more sequences in the sample. A random primer can generally be an oligonucleotide or a population of oligonucleotides comprising a random sequence(s) in which the nucleotides at a given position on the oligonucleotide can be any of the four nucleotides, or any of a selected group of the four nucleotides (for example only three of the four nucleotides, or only two of the four nucleotides). In some cases all of the positions of the oligonucleotide or population of oligonucleotides can be any of the four nucleotides; in other cases, only a portion of the positions, for instance a particular region, of the oligonucleotide will comprise positions which can be any of the four bases. In some cases, the portion of the oligonucleotide which comprises positions which can be any of the four bases is about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, or about 15-20 nucleotides in length. In some cases, the portion of the oligonucleotide which comprises positions which can be any of the four bases is about 5-20, 5-15, 5-10, 4-8, 10-20, 15-20, or 10-15 nucleotides in length. In some cases, a random

IDT Exhibit 1020.5409

primer may comprise a tailed primer having a 3'-region that comprises a random sequence and a 5'-region that is a non-hybridizing sequence that comprises a specific, non-random sequence. The 3'-region may also comprise a random sequence in combination with a region that comprises poly-T sequences. The sequence of a random primer (or its complement) may or may not be naturally-occurring, or may or may not be present in a pool of sequences in a sample of interest. The amplification of a plurality of RNA species in a single reaction mixture would generally, but not necessarily, employ a multiplicity, or a large multiplicity, of random primers. As is well understood in the art, a "random primer" can also refer to a primer that is a member of a population of primers (a plurality of random primers) which collectively are designed to hybridize to a desired and/or a significant number of target sequences. A random primer may hybridize at a plurality of sites on a nucleic acid sequence. The use of random primers provides a method for generating primer extension products complementary to a target polynucleotide which does not require prior knowledge of the exact sequence of the target. In some embodiments one portion of a primer is random, and another portion of the primer comprises a defined sequence. For example, in some embodiments, a 3'-portion of the primer will comprise a random sequence, while the 5'-portion of the primer comprises a defined sequence. In some embodiments a 3'-random portion of the primer will comprise DNA, and a 5'-portion defined portion of the primer will comprise RNA; in other embodiments, both the 3' and 5'-portions will comprise DNA. In some embodiments, the 5'-portion will contain a defined sequence and the 3'-portion will comprise a poly-dT sequence that is hybridizable to a multiplicity of RNAs in a sample (such as all mRNA).

[00143] The hybridizing portion of the oligonucleotide primers may comprise a pool of hybridizing portions which hybridize to a number of sequences or fragments to be analyzed such as for example, 1; 2; 3; 4; 5; 6; 7; 8; 9; 10; 11; 12; 13; 14; 15; 16; 17; 18; 19; 20; 25; 30; 35; 40; 45; 50; 55; 60; 75; 100; 150; 200; 250; 300; 400; 500; 600; 750; 1000; 10,000; 15,000; 20,000; 25,000; 30,000; 40,000; 50,000; 60,000; 75,000; 100,000; 150,000; 200,000; 250,000 or more sequences or fragments. In some cases, each fragment may be hybridized to one primer, in other cases, each fragment is hybridized on average to 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 or more oligonucleotide primers. Oligonucleotide primers suitable for the methods of the present invention are provided herein.

[00144] The olignucleotide primers may be extended along the input nucleic acid template to which they are hybridized. In some cases, the extension may be performed with a polymerase such as for example any of the polymerases provided herein including

IDT Exhibit 1020.5410

polymerases comprising strand displacement activity. Exemplary DNA dependent DNA polymerases suitable for the methods of the present invention include but are not limited to Klenow polymerase, with or without 3'-exonuclease, Bst DNA polymerase, Bca polymerase, φ29 DNA polymerase, Vent polymerase, Deep Vent polymerase, Taq polymerase, T4 polymerase, and *E. coli* DNA polymerase 1, derivatives thereof, or mixture of polymerases. In some cases, the polymerase does not comprise a 5'-exonuclease activity. In other cases, the polymerase comprises 5' exonuclease activity. In some cases, the primer extension of the present invention may be performed using a polymerase comprising strong strand displacement activity such as for example Bst polymerase. In other cases, the primer extension of the present invention may be performed using a polymerase comprising weak or no strand displacement activity. One skilled in the art may recognize the advantages and disadvantages of the use of strand displacement activity during the primer extension step, and which polymerases may be expected to provide strand displacement activity (see e.g., New England Biolabs *Polymerases*). For example, strand displacement activity may be useful in ensuring whole genome or whole transcriptome coverage during the random priming and extension step. Strand displacement activity may further be useful in the generation of double stranded amplification products during the priming and extension step. Alternatively, a polymerase which comprises weak or no strand displacement activity may be useful in the generation of single stranded nucleic acid products during primer hybridization and extension that are hybridized to the template nucleic acid.

[00145] An "RNA-dependent DNA polymerase" or "reverse transcriptase" ("RT") is an enzyme that synthesizes a complementary DNA copy from an RNA template. All known reverse transcriptases also have the ability to make a complementary DNA copy from a DNA template; thus, they are both RNA- and DNA-dependent DNA polymerases. Reverse transcriptases may also have an RNase H activity. Some examples of reverse transcriptases are reverse transcriptase derived from Maloney murine leukemia virus (MMLV-RT), avian myeloblastosis virus, retroviral reverse transcriptase, retrotransposon reverse transcriptase, hepatitis B reverse transcriptase, cauliflower mosaic virus reverse transcriptase, bacterial reverse transcriptase, E. coli DNA polymerase and klenow fragment, and Tth DNA polymerase. A primer can be used to initiate synthesis with both RNA and DNA templates. In other examples a DNA dependent DNA polymerase may also comprise an RNA-dependent DNA polymerase such as Klenow polymerase, Bst DNA polymerase and the like.

[00146] The extension of hybridized oligonucleotide primers, at least a portion of which may comprise random hybridizing portions, non-random hybridizing portions, not-so

IDT Exhibit 1020.5411

random hybridizing portions or a combination thereof, with a polymerase comprising strand displacement activity may provide for the generation of double stranded nucleic acid product fragments. In some cases, the extension of hybridized oligonucleotide primers, at least a portion of which comprise random hybridizing portions, with a polymerase comprising strand displacement activity may produce double stranded nucleic acid products comprising a mixture of double stranded nucleic acid fragment products produced in the polymerization reaction as well as double stranded molecules comprising template nucleic acid hybridized to one or more oligonucleotide primers.

[00147] In the embodiment where the template contains one or more non-canonical nucleotides, the products of the primer extension reaction, e.g. single or double stranded, partially double stranded, or mixtures thereof, may be distinguished from the template nucleic acid in that the template nucleic acid comprises one or more non-canonical nucleotides whereas the products of the primer extension reaction do not comprise non-canonical nucleotides, or do not comprise the same one or more non-canonical nucleotides. In some cases, double stranded products of the primer extension reaction comprise a hybrid duplex of a single strand of template nucleic acid comprising one or more non-canonical nucleotides and a single strand of primer extension product that does not comprise one or more non-canonical nucleotides, or does not comprise the same one or more non-canonical nucleotides. In other cases, double stranded products of the primer extension reaction comprise two strands, of which neither strand comprises one or more non-canonical nucleotides, or of which neither strand comprises the same one or more non-canonical nucleotides as the template nucleic acid.

[00148] The extension of hybridized oligonucleotide primers may be carried out for a suitable period of time. The period of time for the extension reaction may be anywhere from seconds to minutes to hours. For example, the extension step may include incubation of the input nucleic acid template in a reaction mixture such as the reaction mixtures provided herein with one or more oligonucleotide primers at a temperature suitable for the extension reaction (e.g., 15°C-80°C) for a period of between about 5 minutes and about 24 hours. Other suitable extension times include between about 1 minute and about 8 hours, about 2 minutes and about 7 hours, about 3 minutes and about 6 hours, about 4 minutes and about 5 hours, about 5 minutes and about 4 hours, about 5 minutes and about 3 hours, about 5 minutes and about 2 hours, about 10 minutes and about 2 hours, about 15 minutes and about 2 hours, about 20 minutes and about 2 hours, about 30 minutes and about 2 hours, or between about 30 minutes and about 1 hour. Still other suitable extension times include 1 minute, 2

IDT Exhibit 1020.5412

minutes, 3 minutes, 4 minutes, 5 minutes, 6 minutes, 7 minutes, 8 minutes, 9 minutes, 10 minutes, 12 minutes, 15 minutes, 20 minutes, 30 minutes, 45 minutes, 60 minutes, 1 hour, 1.5 hours, 2 hours, 2.5 hours, 3 hours, 3.5 hours, 4 hours or more. Still other suitable extension times include about 1 minute, 2 minutes, 3 minutes, 4 minutes, 5 minutes, 6 minutes, 7 minutes, 8 minutes, 9 minutes, 10 minutes, 12 minutes, 15 minutes, 20 minutes, 30 minutes, 45 minutes, 60 minutes, 1 hour, 1.5 hours, 2 hours, 2.5 hours, 3 hours, 3.5 hours, 4 hours or more.

[00149] The extension step may be performed in a reaction mixture comprising nucleotides, labeled nucleotides or a combination thereof. For example, the hybridized oligonucleotides may be extended by one or more polymerases, such as polymerases comprising strand displacement activity or polymerases comprising weak or no strand displacement activity, along the input nucleic acid template in the presence of a mixture of dNTPs and amino allyl dNTPs. The use of amino-allyl dNTPs may allow further labeling and modification of the products of the extension reaction such as double stranded DNA fragment products. For example, the amino allyl dNTPs may provide for biotinylation, fluoresceination, labelling with Cy dyes (e.g., Cy3 or Cy5), or any other nucleic acid modification known in the art. Other modified nucleotides which are suitable for post amplification labeling by either covalent or non-covalent attachment of labels (e.g., fluorophores, chromophores, biotin, antibodies, antigens, or enzymes such as alkaline phosphatase or horse radish peroxidase) are also applicable including for example thio, phosphorothio, and amino modified nucleotides and oliognucleotides as described in U.S. Patent Nos. 6172209, 5679785, and 5623070, or any other modified nucleotides provided herein.

SPIA Amplification

[00150] 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 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.

[00151] Amplification by SPIA can occur under conditions permitting composite primer hybridization, primer extension by a DNA polymerase with strand displacement

IDT Exhibit 1020.5413

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).

[00152] 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.

[00153] 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.

[00154] In some embodiments, the composite amplification primer is generated in the amplification reaction mixture from a stem-loop chimeric pro-primer. The amplification

IDT Exhibit 1020.5414

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.

Other Amplification Methods

[00155] 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.

[00156] 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

IDT Exhibit 1020.5415

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.

[00157] 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 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.

[00158] 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.

IDT Exhibit 1020.5416

[00159] 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.

[00160] 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.

[00161] 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).

[00162] 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.

[00163] 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.

[00164] 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

IDT Exhibit 1020.5417

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.

[00165] 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.

[00166] 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.

[00167] 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.

IDT Exhibit 1020.5418

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 Whole Transcriptome Analysis

[00168] An important aspect of the invention is that the methods and compositions disclosed herein can be efficiently and cost-effectively utilized for downstream analyses, such as next generation sequencing or hybridization platforms, with minimal loss of biological material of interest. Specifically, the methods of the invention are useful for sequencing a whole transcriptome from a NGS library with depleted or reduced rRNA content.

Sequencing

[00169] 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.

[00170] 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.

IDT Exhibit 1020.5419

[00171] 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 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).

[00172] 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.

[00173] 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).

[00174] 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

IDT Exhibit 1020.5420

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 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. Accordingly, the methods of the invention 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. Directional (strand-specific) cDNA libraries are prepared using the methods of the present invention, and the selected single-stranded nucleic acid is amplified, for example, by 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

IDT Exhibit 1020.5421

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.

[00175] 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 (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).

[00176] 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.

IDT Exhibit 1020.5422

[00177] Thus, in some embodiments, the methods of the invention are useful for preparing target polynucleotides for sequencing by the sequencing by ligation methods commercialized by Applied Biosystems (e.g., SOLiD sequencing). In other embodiments, the methods are useful for preparing target polynucleotides 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 other 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 other 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.

[00178] 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 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.

IDT Exhibit 1020.5423

[00179] 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.\

[00180] 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.

[00181] 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

IDT Exhibit 1020.5424

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 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.

[00182] 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.

[00183] 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.

IDT Exhibit 1020.5425

Kits

[00184] Any of the compositions described herein may be comprised in a kit. In a non-limiting example, the kit, in a suitable container, comprises: an adaptor or several adaptors, one or more of oligonucleotide primers and reagents for ligation, primer extension and amplification. The kit may also comprise means for purification, such as a bead suspension, and nucleic acid modifying enzymes.

[00185] The containers of the kits will generally include at least one vial, test tube, flask, bottle, syringe or other containers, into which a component may be placed, and preferably, suitably aliquotted. Where there is more than one component in the kit, the kit also will generally contain a second, third or other additional container into which the additional components may be separately placed. However, various combinations of components may be comprised in a container.

[00186] 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 may 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.

[00187] In various embodiments, a kit according to the invention comprises one or more of a restriction endonuclease, e.g. BspQI, a ligase, a polymerase, e.g. a hot start polymerase such as MyTaq, a cleavage agent, a library of probes capable of acting as a primer for a primer extension reaction, and one or more non-canonical nucleotides, e.g. uracil or inosine. In some embodiments, the cleavage agent comprises one or more of a glycosylase, e.g. UNG or UDG, a primary amine, a polyamine, e.g. DMED, and endonuclease V.

[00188] In some embodiments, a kit comprises one or more of a first adapter comprising one or more non-canonical nucleotides on one strand and lacking 5' phosphates, a second adapter lacking said one or more non-canonical nucleotides and lacking 5' phosphates, and a set of primers specific to the adaptor sequences. In some embodiments, the second adapter comprises a recognition sequence for a restriction endonuclease.

[00189] In some embodiments, a kit comprises one or more of a first adapter lacking 5' phosphates, a plurality of partial duplex primers each comprising a 3' overhang and comprising a shared sequence within a double-stranded portion, and a primer that is hybridizable a sequence reverse complimentary to the adapter. In some embodiments, the first adapter comprises a recognition sequence for a restriction endonuclease. In some embodiments, the plurality of partial duplex primers comprise at least two partial duplex primers with dissimilar 3' overhang sequences.

IDT Exhibit 1020.5426

[00190] In some embodiments, the kit comprises one or more of a first adapter lacking 5' phosphates, a plurality of partial duplex primers each comprising a 3' overhang, comprising a shared sequence within a double-stranded portion, and the strand of the plurality of partial duplex primers with the 3' overhang lacking adenines in the shared sequence within the double-stranded portion, and a set of primers that are hybridizable to a sequence reverse complimentary to the adapter and the shared sequence of the partial duplex primers opposite the 3' overhang. In some embodiments, the first adapter comprises a recognition sequence for the restriction endonuclease. In some embodiments, the plurality of partial duplex primers comprise at least two partial duplex primers with dissimilar 3' overhang sequences.

[00191] A kit will preferably include instructions for employing, the kit components as well the use of any other reagent not included in the kit. Instructions may include variations that can be implemented.

[00192] 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 embodiments, the invention provides kits comprising adaptors, primers, and/or other oligonucleotides described herein. 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.

Products based on the Methods of the Invention

[00193] Products based on the methods of the invention may be commercialized by the Applicants under the trade name Encore Complete Prokaryotic RNA-Seq™. Encore is a trademark of NuGEN Technologies, Inc.

IDT Exhibit 1020.5427

EXAMPLES

Example 1 – Depletion of bacterial ribosomal RNA fragments from directional (i.e. strand-specific) whole transcriptome libraries.

[00194] This example describes the depletion of bacterial rRNA fragments from four directional cDNA libraries generated from *E.coli* total RNA, using insert-dependent adaptor cleavage (InDA-C) probes that target highly conserved prokaryotic 16S and as 23S rRNA transcript regions.

Probe design and synthesis

[00195] InDA-C probes that target prokaryotic rRNA transcripts were designed by comparing the ribosomal operons from a phylogenetically diverse set of 40 bacterial strains and 10 archaeal strains using the ClustalW multiple sequence alignment program (European Bioinformatics Institute). Candidate primer sequences were first selected from highly conserved sequences identified in 16S rRNA (9 sites) and 23S rRNA (7 sites) subunits. These conserved regions were computationally fragmented and analyzed by Primer3 (Steve Rozen and Helen J. Skaletsky (2000) Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds) Bioinformatics Methods and Protocols: Methods in Molecular Biology. Humana Press, Totowa, NJ, pp 365-386). These sequences were then filtered for optimal predicted melting temperatures ranging from 55-65° C and length. Oligonucleotides corresponding to the rRNA sense strand were synthesized individually and pooled in equimolar proportions. The final primer pool was comprised of 205 oligonucleotides ranging from 14-18 nt in length. Some primers were synthesized with one or more nucleotide analogues, such as Locked Nucleic Acid (LNA) bases, to increase their respective melting temperatures. The probe mix was diluted to 25 times the final concentration used in InDA-C depletion reactions (375 nM per species, 15 nM final).

Generation of strand-specific cDNA libraries

[00196] The Encore Complete RNA-Seq Library System (NuGEN Technologies, p/n 0311) was used to generate four strand-specific cDNA libraries from 100 ng of *E. coli* total RNA (Life Technologies, p/n AM7940) extracted from a liquid culture harvested at the mid-log phase of growth in rich media. The reverse transcription reaction was carried out according to the manufacturer's instructions except that the primer supplied in the kit was replaced with the first strand primer from the Ovation Prokaryotic RNA-Seq System (NuGEN Technologies, p/n 9030). Second strand DNA synthesis was performed as recommended in the kit and the double-stranded cDNA was sheared with a Covaris S-series device using the 200 bp sonication protocol provided with the instrument (10% duty cycle,

IDT Exhibit 1020.5428

200 cycles/burst, 5 intensity, 180 seconds). Purification of the fragmented cDNA was accomplished by adding 2 volumes of Ampure XP beads (Agencourt Genomics), washed twice with 70% ethanol and eluted with 15 μL of water. Ten microliters of each sample were prepared for ligation using the End Repair reaction as described in the kit. Ligation was performed with the reverse adaptor provided in the kit and a custom forward adaptor containing deoxyuridine and a single base substitution in the BspQI recognition site (5'-TACACTCUTTCCCUACACGACGAUCTTCCGAUCT-3'). Following the Strand Selection I reaction, samples were purified with beads as before except that elution volume was 25 μL with 18 μL of that taken forward.

Ribosomal RNA depletion

[00197] Ribosomal DNA fragments were selectively depleted from the library in three distinct steps: 1) base excision/rRNA-specific primer extension, 2) reverse adaptor cleavage and 3) PCR enrichment. The first step was performed by combining each 18 μL sample with 7 μL of mastermix containing 1 μL of InDA-C rRNA probes, 5 μL of 5X MyTaq polymerase buffer, 0.5 μL of Strand Selection II enzyme (SS4) from the Encore Complete RNA-Seq system and 0.5 μL of HS MyTaq polymerase (Bioline p/n BIO-21111). This solution was placed in a thermal cycler, heated to 37°C for 10 minutes to complete strand selection and generate single-stranded library fragments, heated to 95°C for 2 minutes to activate the hot start polymerase, cooled to 50°C for 30 seconds to anneal rRNA probes, heated to 65°C for 5 minutes to allow primer extension from insert into the reverse adaptor sequence. Samples were cooled to 4°C before adding 25 μL of adaptor cleavage mastermix containing 1X MyTaq polymerase buffer and 2.5 units of BspQI restriction enzyme (New England Biolabs p/n R0712). Reactions were carried out in a thermal cycler by heating to 55°C for 5 minutes and 95°C for 5 minutes before cooling to 4°C. Enrichment of non-rRNA fragments was accomplished by adding 50 μL of 2X PCR mastermix containing 1X MyTaq polymerase buffer, 2.5 units of HS MyTaq polymerase and 8 μL of P2 primer mix provided in the kit. Samples were placed in a thermal cycler, heated to 95°C for 2 minutes to activate the polymerase and amplified using a 2-step temperature routine: 2 cycles of 95°C for 30 seconds, 60°C for 90 seconds and 18 cycles of 95°C for 30 second, 65° C for 90 seconds. PCR products were purified using AMPure XP beads and analyzed with a 2100 Bioanalyzer (Agilent Technologies). Libraries were sequenced in single end format on an Illumina GA2X instrument. Raw data were processed using Illumina base calling software and mapped to the E. coli K-12 (substrain MG1655) reference genome (Genbank Accession #AP009048). The

IDT Exhibit 1020.5429

orientation of reads is expected to be in the sense strand orientation relative to RNA templates.

[00198] Only one of the four cDNA aliquots was converted to a library using the full complement of InDA-C components (Test4). The other three libraries were constructed with one or more of the InDA-C reagents missing (Test1, Test2 and Test3). A control library generated with random primers from the same RNA was used as a benchmark for the undepleted input sample (ctrl). The mapping statistics for the control and each of the test libraries are shown in Figure 2. A comparison of expression profiles from the four test libraries is shown in Figure 3. The targeted depletion of 16S rRNA sites by universal prokaryotic InDA-C probes is depicted in Figure 4.

Example 2 – Depletion of mitochondrial DNA fragments from a genomic DNA library.

[00199] This example describes the depletion of mitochondrial DNA fragments from a genomic DNA library, using insert-dependent adaptor cleavage (InDA-C) probes that target the mitochondrial genome.

Probe design and synthesis

[00200] InDA-C probes that anneal to both strands of the hg19 version of the human mitochondrial genome sequence were selected within mitochondrial-specific segments identified by the "Duke 20 bp uniqueness" tracks provided by the UCSC Genome Browser. These sequences were then filtered for optimal predicted melting temperatures and length. Oligonucleotides ranging from 20-25 nt in length were synthesized individually and pooled in equimolar proportions. The resulting probe mix was diluted to 25 times the final concentration used in InDA-C depletion reactions (375 nM per species, 15 nM final).

Generation of genomic DNA libraries

[00201] The Ovation Ultralow Library System (NuGEN Technologies, San Carlos, CA) was used to generate DNA libraries from 10 ng of human male DNA (Promega). The DNA was sheared with a Covaris S-series device using the 200 bp sonication protocol provided with the instrument (10% duty cycle, 200 cycles/burst, 5 intensity, 180 seconds). Purification of the fragmented DNA was accomplished by adding 2 volumes of Ampure XP beads (Agencourt Genomics), washed twice with 70% ethanol and eluted with 15 μL of water. Ten microliters of each sample were prepared for ligation using the End Repair reaction as described in the kit. Ligation was performed with a custom forward adaptor and the Illumina TruSeq reverse adaptor. The forward adaptor contained an AsiSI recognition site (5'-GCGATCGC-3') near the ligation junction (5'-AATGATACGGCGACCACCGAAGATAAGAAGAaTGAcGTcAAgTGCGATCGCAGGA

IDT Exhibit 1020.5430

TAGAT-3'). The reverse adapter contained a BspQ1 recognition site (5'-GCTCTTC-3') near the ligation junction (5'-CAAGCAGAAGACGGCATACGAGATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT-3'). Samples were purified with beads as before except that elution volume was 25 μL with 18 μL of that taken forward.

Mitochondrial DNA depletion

[00202] Mitochondrial DNA fragments were selectively depleted from the library in three distinct steps: 1) denaturation/mitochondrial-specific primer extension, 2) adaptor cleavage, and 3) PCR enrichment. The first step was performed by combining each 18 μL sample with 7 μL of mastermix containing 1 μL of InDA-C mitochondrial probes, 5 μL of 5X MyTaq polymerase buffer, and 0.5 μL of HS MyTaq polymerase (Bioline p/n BIO-21111). This solution was placed in a thermal cycler, heated to 95°C for 10 minutes to complete strand separation, generate single-stranded library fragments, and to activate the hot start polymerase, cooled to 50° C for 30 seconds to anneal rRNA probes, heated to 65° C for 5 minutes to allow primer extension from insert into the reverse adaptor sequence. Samples were cooled to 4° C before adding 25 μL of adaptor cleavage mastermix containing 1X MyTaq polymerase buffer, 2.5 units of BspQI restriction enzyme (New England Biolabs p/n R0712), and 2.5 units of AsiSI restriction enzyme (New England Biolabs p/n R0630). Reactions were carried out in a thermal cycler by heating to 40° C for 5 minutes and 95° C for 5 minutes before cooling to 4° C. Enrichment of non-mitochondrial fragments was accomplished by adding 50 μL of 2X PCR mastermix containing 1X MyTaq polymerase buffer, 2.5 units of HS MyTaq polymerase and 8 μL of 10X PCR primer mix containing 10 μM forward primer (5'-AATGATACGGCGACCACCGA-3') and 10 μM reverse primer (5'-CAAGCAGAAGACGGCATACG-3'). Samples were placed in a thermal cycler, heated to 95° C for 2 minutes to activate the polymerase and amplified using a 2-step temperature routine: 2 cycles of 95° C for 30 seconds, 60° C for 90 seconds and 18 cycles of 95°C for 30 second, 65° C for 90 seconds. PCR products were purified using AMPure XP beads and analyzed with a 2100 Bioanalyzer (Agilent Technologies). Libraries were sequenced in single end format on an Illumina GA2X instrument. Raw data were processed using Illumina base calling software and mapped to human reference genome.

Example 3 – Generation of a directional cDNA library (Figure 5).

[00203] This example describes the generation of a directional cDNA library using conventional blunt-end ligation with modified duplex adaptors and 50 ng of poly(A) + selected messenger RNA as a starting material.

IDT Exhibit 1020.5431

First strand synthesis

[00204] First strand cDNA was generated using random hexamer priming. The first strand synthesis reaction was conducted using the Invitrogen SuperScript III Reverse Transcriptase kit, with 10 μM of random hexamers, 3.0 mM $MgCl_2$ and 1.0 mM dNTPs. The cDNA synthesis reaction was carried out in 10 μL volume, incubated at 40 degrees Celsius for 60 minutes and chilled to 4 degrees Celsius.

Second strand synthesis with dUTP incorporation

[00205] Second strand synthesis was performed using the New England Biolabs NEBNext Second Strand Synthesis Module, where the Second Strand Synthesis (dNTP-free) Reaction Buffer was supplemented with dNTP mix containing 0.2 mM of dATP, dCTP and dGTP, and 0.54 mM dUTP. RNAse H-mediated nick translation was carried out by adding 65 μL of second strand synthesis master mix and incubating for one hour at 16 degrees Celsius. The reaction was stopped by adding 45 μL of 25 mM EDTA. Fragmentation and purification of cDNA fragments

[00206] The 120 μL second strand synthesis reaction was subjected to acoustic fragmentation using the Covaris S-series System according to the manufacturer's instructions, using the manufacturer recommended settings to produce fragmented DNA with an average fragment size of 150-200 bases. Fragmented DNA was concentrated using QIAquick PCR purification kit, according to the manufacturer's instructions. The fragmented and concentrated DNA was quantitated and run on Agilent Bioanalyzer DNA 1000 chip to ensure fragment distribution of 150-200 bp length.

End repair

[00207] The ends of the fragmented cDNA were repaired to generate blunt ends with 5' phosphates and 3' hydroxyls. End repair of the fragmented DNA was performed according to the Encore ™ Ultra Low Input NGS Library System I User Guide instructions using End Repair Master Mix.

Ligation with dU marked adaptors

[00208] Duplex adaptors were ligated to blunt-ended cDNA fragments according to the Encore ™ Ultra Low Input NGS Library System I User Guide Instructions, with the

IDT Exhibit 1020.5432

exception that the Ligation Adaptor Mix contained one adaptor where the ligation strand of the adaptor had at least one dU incorporated into it.

Nick repair/adaptor fill-in

[00209] Ligation of unphosphorylated adaptors leaves a single-strand nick that must be repaired prior to strand selection and amplification. To fill in the adaptor sequence and generate full-length double-stranded DNA (dsDNA), the reaction mix was heated at 72 degrees Celsius, resulting in the extension of the 3' end of the cDNA insert by *Taq* DNA polymerase (thereby filling in the adaptor sequence), and the melting of the unligated adaptor strand. The repaired dsDNA fragments with ligated adaptors were then purified using Agencourt RNAClean XP Beads , according to the Encore ™ Ultra Low Input NGS Library System I User Guide Instructions.

Strand selection with UDG/APE I treatment

[00210] Uridine digestion was performed with 1 unit of UNG and 1,000 units of APE I at 37° C for 20 minutes. Incorporation of dUTP into one strand of the cDNA insert and the ligation strand of one of the two adaptors allowed for selective removal of the products with the undesired adaptor orientation. Consequently, a polynucleotide strand with incorporated dUTP that is treated with UNG/APE I was unable to undergo amplification by a polymerase.

Library amplification

[00211] To produce a final directional cDNA library, the UNG-selected fragments were amplified by PCR according to the Library Amplification Protocol in the Encore ™ Ultra Low Input NGS Library System I User Guide.

Example 4 – Depletion of ribosomal RNA fragments from a genomic DNA library from cells sorted by size

[00212] Cells from a human blood sample are sorted on a Beckman MoFlo cell sorter, based on surface markers into distinct populations and individuals within those populations are separated and lysed using NuGEN's Prelude Direct Lysis Module according to manufacturer's recommendations.

[00213] The resulting RNA containing solution is used as input into NuGEN's Encore® Whole Blood RNA-Seq with care being taken to avoid lysis of the nucleus. Following first strand synthesis, second strand synthesis performed in the presence of dUTPs, and adapters comprising a restriction endonuclease recognition sequence are ligated and filled-in. The second strand is degraded by UNG treatment. The reaction mixture is incubated with a set of probes designed to anneal to sequences in rRNA transcripts that are converted into cDNAs.

IDT Exhibit 1020.5433

[00214] The hybridized probes are extended using a DNA polymerase all the way to the adapter sequence, generating double stranded adapters on undesired nucleic acids, comprising the restriction endonuclease recognition sequence. Adapters on nucleic acids that are not probe targets, remain single stranded. The double stranded adapter sequences are digested with a restriction enzyme to remove the adapter rendering them unable to amplify during the PCR enrichment step. PCR primers targeting the adapters, master mix and a thermophilic polymerase are added and thermal cycled 20 cycles. The resultant library is quantified and applied to an Illumina flow cell for sequencing.

Example 5 – Depletion of ribosomal RNA fragments from a genomic DNA library on a microfluidic system

[00215] CD4+CD25+ cells are sorted from a blood sample into a pool using a Becton Dickenson Influx cell sorter based on surface markers lysed using NuGEN's Prelude Direct Lysis Module according to manufacturer's recommendations.

[00216] The resulting RNA containing solution is gently introduced to Agencourt magnetic beads to a final volume of 50ul under conditions that favored RNA vs. DNA binding. Care is taken to avoid lysis of the cell nucleus. The bead containing solution is then loaded to NuGEN's Mondrian™ digital microfluidic system Encore Complete SP cartridge, the cartridge applied to the workstation and the appropriate script selected. Following first strand synthesis, second strand synthesis is performed in the presence of the suitable nucleotide analog according to manufacturer's instructions, Manufacturer's instructions are followed through fragmentation, ligation with suitable adapters comprising nucleotide analogs and a restriction endonuclease recognition sequence, and strand selection The products are retrieved from the system following strand selection and before the PCR enrichment step. The sample ~1ul in 19ul of cartridge filler fluid, was diluted to 10ul in a solution containing InDA-C probes designed to anneal to sequences in human rRNA transcripts.

[00217] The hybridized probes are extended using a DNA polymerase all the way to the adapter sequence, generating double stranded adapters on undesired nucleic acids, comprising the restriction endonuclease recognition sequence. Adapters on nucleic acids that are not probe targets, remain single stranded. The double stranded adapter sequences are digested with a restriction enzyme to remove the adapter rendering them unable to amplify during the PCR enrichment step (Figure 5). PCR primers targeting the adapters, master mix and a thermophilic polymerase are added and thermal cycled 20 cycles. The resultant library is quantified and applied to an Illumina flow cell for sequencing.

IDT Exhibit 1020.5434

Example 6 – Depletion of ribosomal RNA fragments from a genomic DNA library from single cells expressing GFP

[00218] Cells expressing GFP from a human blood sample are sorted on a FACS Vantage SE Cell sorter (BD Biosciences, San Diego, CA, http://www.bdbiosciences.com) based on color into distinct populations. Cells above a threshold GFP expression are separated into individual microwells and lysed using NuGEN's Prelude Direct Lysis Module according to manufacturer's recommendations.

[00219] The resulting RNA containing solution is primed for first strand synthesis with either N6 or USP primers (NuGEN Encore Complete first strand primer mix). The primers are extended with a Reverse Transcriptase and nucleotide solution containing dUTP and dITP at a ratio of canonical to non-canonical nucleotides for enabling fragementation to a desired size range. Following synthesis, the cDNA is fragmented by treatment with UNG (Figure 6) to generate fragments of the desired size range comprising blocked 3'-end.

[00220] The resulting cDNA product with inosines is primed with a partial duplex oligonucleotide complex comprising 33 bases of double stranded structure appended with 8 random nucleotides of single stranded DNA at one 3' end (Figure 8). A 3' extension reaction follows using the cDNA product comprising inosines as a template. Following ligation of an adapter comprising a restriction endonuclease recognition sequence to the end of the double stranded molecule and fill-in to produce blunt ends, the library is treated with Endonuclease V to remove the inosine residues and fragment the cDNA product. The resulting single stranded DNA with adapter sequences appended to each end is incubated with a set of probes designed to anneal to sequences within cDNAs corresponding to rRNA sequences.

[00221] The hybridized probes are extended using a DNA polymerase all the way to the adapter sequence, generating double stranded adapters on undesired nucleic acids, comprising the restriction endonuclease recognition sequence. Adapters on nucleic acids that are not probe targets, remain single stranded. The double stranded adapter sequences are digested with a restriction enzyme to remove the adapter rendering them unable to amplify during the PCR enrichment step (Figure 9). PCR primers targeting the adapters, master mix and a thermophilic polymerase are added and thermal cycled 20 cycles. The resultant library is quantified and applied to an Illumina flow cell for sequencing.

Example 7 – Depletion of ribosomal RNA fragments from a genomic DNA library from single cells expressing a CFP-YFP FRET system

[00222] Cells expressing a CFP-YFP FRET system are sorted on a FACS Vantage SE Cell sorter (BD Biosciences, San Diego, CA, http://www.bdbiosciences.com) based on the

IDT Exhibit 1020.5435

FRET emission signal into distinct populations. Cells above a threshold FRET emission are separated into individual microwells and lysed using NuGEN's Prelude Direct Lysis Module according to manufacturer's recommendations.

[00223] The resulting RNA containing solution is primed for first strand synthesis with either N6 or USP primers (Encore Complete first strand primer mix, NuGEN). The primers are extended with a Reverse Transcriptase and nucleotide solution containing dUTP. Following synthesis the cDNA is fragmented by treatment with UNG (Figure 7) to generate fragments of the desired size range. This cDNA product is primed with a partial duplex oligonucleotide complex library, each complex comprising 33 bases of double stranded structure appended with 8 random nucleotides of single stranded DNA as 3' overhang (Figure 8). The oligo complexes are made up of 2 strands comprising 33 nucleotides on the short strand and 41 nucleotides on the long strand, respectively. The 33 bases of the ling strand falling within the double-stranded portion lacks any adenine nucleotides.

[00224] The 8 base random sequence is annealed to the fragmented cDNA and extended with a DNA polymerase in the presence of dUTP. At the same time, the 33 base oligo is displaced by the DNA polymerase producing a blunt ended molecule. By virtue of lacking adenines in the double stranded portion of the long strand of the oligo complex, the extension product displacing the short strand does not incorporate any uracils. Following ligation of an adapter comprising a restriction endonuclease recognition sequence to the end of the double stranded molecule to the end of the double stranded molecule and fill-in to produce blunt ends, the library is treated with UNG to fragment the DNA where dUs residues are incorporated. The resulting single stranded DNA with adapter sequences appended to each end is incubated with a set of probes designed to anneal to sequences within cDNAs corresponding to rRNA sequences.

[00225] The hybridized probes are extended using a DNA polymerase all the way to the adapter sequence, generating double stranded adapters on undesired nucleic acids, comprising the restriction endonuclease recognition sequence. Adapters on nucleic acids that are not probe targets, remain single stranded. The double stranded adapter sequences are digested with a restriction enzyme to remove the adapter rendering them unable to amplify during the PCR enrichment step (Figure 9). PCR primers targeting the adapters, master mix and a thermophilic polymerase are added and thermal cycled 20 cycles. The resultant library is quantified and applied to an Illumina flow cell for sequencing.

Example 8 – Probe design for the depletion of undesired nucleic acid fragments from library.

IDT Exhibit 1020.5436

[00226] This example describes the depletion of undesired nucleic acid fragments from a library of various origins, using insert-dependent adaptor cleavage (InDA-C) probes that target the undesired nucleic acid fragments.

Probe design and synthesis

[00227] Target sequences for depletion are compiled for transcripts that might frequently be found in high abundance within a given sample type. Examples of such transcripts are ribosomal RNA (rRNAs) and mitochondrial RNAs in most sample types, globin within blood samples and chloroplast RNAs within plant samples. These sequences are compiled from public data such as RefSeq when available or from empirical data sources (Grape Genome Browser available online from Genoscope, Denoeud et al. Annotating genomes with massive-scale RNA sequencing. Genome Biology 2008, 9:R175 doi:10.1186/gb-2008-9-12-r175: http://www.genoscope.cns.fr/externe/GenomeBrowser/Vitis/) as was the case with grape, which does not have a well annotated or complete reference genome. Orientation of probes is determined based on which strand of template is to be retained following adapter ligation. Each undesired transcript is computationally "fragmented" into 70 base regions and these regions are queried using PCR primer design software such as Primer3 (Steve Rozen and Helen J. Skaletsky (2000) Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds) Bioinformatics Methods and Protocols: Methods in Molecular Biology. Humana Press, Totowa, NJ, pp 365-386). Target melting temperature is set to 60° C for human cytoplasmic and mitochondrial rRNA and human globin message and 65° C for grape cytoplasmic and mitochondrial rRNA and grape chloroplast rRNA.

[00228] Primer sequences proposed by Primer3 are BLASTed against known transcript sequences from the same organism to limit or eliminate off-target interactions. Probes determined to have off-target interactions are removed from the pool. The primer probe oligonucleotides are produced using standard phosporamidite chemistries.

Depletion of RNA and DNA sequences

[00229] The designed primer probes specific for undesired nucleic acid sequences, such as human cytoplasmic and mitochondrial rRNA, human globin mRNA, grape cytoplasmic and mitochondrial rRNA, grape chloroplast rRNA, are utilized in depleting the undesired sequences in one of the ways described herein, such as one of the methods exemplified in Examples 1, 2, 4, 5, 6, or 7 (Figures 1, 5-7, and 9). Lower annealing and extension temperatures may be used for more aggressive strand depletion conditions. Briefly,

IDT Exhibit 1020.5437

single stranded nucleic acids in various adapter configurations are hybridized with a set of designed primer probes for depleting undesired nucleic acids. The nucleic acid is prepared with a restriction endonuclease recognition sequence supplied on the 5' end. The primer probes are extended resulting in a double-stranded structure around the restriction endonuclease recognition sequence. Cleaving the nucleic acid at the restriction endonuclease recognition site further destroys a primer annealing sequence targeted by a subsequent amplification reaction, e.g. PCR. Thus, nucleic acids targeted by the primer probes are unavailable for amplification, enriching the remainder of the nucleic acids in a sample.

IDT Exhibit 1020.5438

CLAIMS

WHAT IS CLAIMED IS:

1. A method for depleting or reducing specific non-desired nucleic acid sequences from a nucleic acid library, the method comprising:

a. generating a nucleic acid library comprising single-stranded DNA fragments with adaptors of fixed orientation attached to each end of each DNA fragment;

b. annealing sequence-specific oligonucleotide probes to the single-stranded DNA fragments with adaptors of fixed orientation attached at each end, wherein the sequence-specific oligonucleotide probes are designed to be complementary to the non-desired nucleic acid sequences, and wherein at least one of the two adaptors comprises a recognition sequence for a restriction endonuclease specific for double-stranded DNA;

c. extending the sequence-specific oligonucleotide probes with a DNA polymerase, thereby creating double-stranded DNA fragments comprising at least a portion of the non-desired nucleic acid sequences;

d. treating the population of DNA fragments comprising double-stranded and single-stranded DNA with a restriction endonuclease specific for double-stranded DNA, thereby cleaving double-stranded DNA fragments at the restriction endonuclease site; and

e. performing PCR with a set of primers specific to the adaptor sequences, thereby amplifying the DNA fragments comprising the desired nucleic acid sequences.

2. The method of claim 1, further comprising an additional step of sequencing the amplified products.

3. The method of claim 1, wherein the nucleic acid library originates from a population of sorted cells.

4. The method of claim 3, wherein the nucleic acid library originates from a single cell.

5. The method of claim 3, further comprising sorting cells into a multiwell plate, microarray, microfluidic device, or slide and thereby generating the population of sorted cells.

6. The method of claim 5, wherein the sorting is performed according to a cell surface marker.

7. The method of claim 5, wherein the sorting is performed according to optical properties of the cells.

8. The method of claim 5, wherein the sorting is performed according to cell size.

IDT Exhibit 1020.5439

9. The method of claim 1, wherein the non-desired nucleic acid sequences comprise bacterial ribosomal RNA, mitochondrial DNA, human globin mRNA, human cytoplasmic rRNA, human mitochondrial rRNA, grape cytoplasmic rRNA, grape mitochondrial rRNA, or grape chloroplast rRNA.

10. The method of claim 1, wherein the restriction endonuclease of step d. is BspQI.

11. The method of claim 1, wherein the DNA polymerase comprises a hot start polymerase.

12. The method of claim 11, wherein the DNA polymerase is MyTaq polymerase.

13. The method of claim 1, wherein step a. comprises the following steps:

i. reverse transcribing a RNA sample;

ii. generating double-stranded cDNA from the reverse transcribed RNA sample, wherein at least one of the four dNTPs dATP, dCTP, dGTP or dTTP is replaced by a non-canonical dNTP during second strand synthesis and incorporated into the second strand;

iii. performing end repair on the double-stranded cDNA;

iv. ligating adaptors to a 5' end of the double-stranded cDNA, wherein one of the adaptor strands has the non-canonical nucleotide incorporated into a ligation strand of the adaptor;

v. performing gap repair; and

vi. selectively removing the second strand by a cleavage agent.

14. The method of claim 5, wherein the non-canonical nucleotide comprises uridine or inosine.

15. The method of claim 13, wherein step vi comprises cleaving a base portion of the one or more non-canonical nucleotides, thereby forming an abasic site.

16. The method of claim 13, wherein the cleavage agent comprises a glycosylase.

17. The method of claim 16, wherein the glycosylase is UNG or UDG.

18. The method of claim 13, wherein the cleavage agent comprises a primary amine.

19. The method of claim 13, wherein the cleavage agent comprises a polyamine.

20. The method of claim 19, wherein the polyamine is DMED.

21. The method of claim 13, wherein the cleavage agent comprises a glycosylase and a polyamine.

22. The method of claim 13, wherein the cleavage agent comprises endonuclease V.

23. A method of adapter ligation to a pool of nucleic acids, comprising:

a. ligating a nucleic acid comprising a first nucleic acid strand comprising a 5' phosphate, a second nucleic acid strand comprising a 5' phosphate and one or more non-

IDT Exhibit 1020.5440

canonical nucleotides, with at least a first adapter comprising a first adapter strand lacking a 5' phosphate and a second adapter strand lacking a 5' phosphate and one or more non-canonical nucleotides;

b. performing a 3' extension reaction; and

c. performing a cleavage reaction with an agent comprising one or more cleavage reagents, thereby cleaving at least one nucleic acid strand comprising one or more non-canonical nucleotides; wherein one of the one or more cleavage agents is specific for nucleic acid strands comprising the one or more non-canonical nucleotides.

24. The method of claim 23, further comprising ligating the nucleic acid with a second adapter comprising a third adapter strand lacking a 5' phosphate and a fourth adapter strand lacking a 5' phosphate and one or more non-canonical nucleotides, wherein the first and the second adapters are different.

25. The method of claims 23 or 24, wherein the nucleic acid is ligated with a first or second adapter at each end.

26. The method of claim 23, wherein the non-canonical nucleotide is selected from uracil and inosine.

27. The method of claim 23, wherein step c comprises cleaving a base portion of the one or more non-canonical nucleotides, thereby forming an abasic site.

28. The method of claim 23, wherein the one or more cleavage reagents comprises a glycosylase.

29. The method of claim 28, wherein the glycosylase is UNG or UDG.

30. The method of claim 23, wherein the one or more cleavage reagents comprises a primary amine.

31. The method of claim 23, wherein the one or more cleavage reagents comprises a polyamine.

32. The method of claim 31, wherein the polyamine is DMED.

33. The method of claim 23, wherein the one or more cleavage reagents comprises a glycosylase and a polyamine.

34. The method of claim 23, wherein the one or more cleavage reagents comprises endonuclease V.

35. The method of claim 24, further comprising performing an amplification reaction comprising a first primer and second primer, wherein the first primer is hybridizable to the first adapter strand and the second primer is hybridizable to the fourth adapter strand, thereby generating amplified products.

IDT Exhibit 1020.5441

36. The method of any one of claims 23-35, wherein the first adapter comprises a recognition sequence for a restriction endonuclease specific for double-stranded DNA.

37. The method of claim 36, further comprising,

d. hybridizing a probe to a sequence on the first nucleic acid strand,

e. extending the probe with a DNA polymerase, thereby producing a partial duplex nucleic acid, and

f. treating the partial duplex nucleic acid with a restriction endonuclease specific for double-stranded DNA, thereby cleaving double-stranded DNA fragments at the recognition sequence.

38. The method of claim 37, further comprising performing PCR with a set of primers specific to the adaptor sequences, thereby amplifying at least a second nucleic acid in the pool of nucleic acids.

39. The method of claim 38, wherein the second nucleic acid lacks the sequence in step d.

40. The method of claim 38, further comprising sequencing a portion of the second nucleic acid.

41. The method of claim 23, wherein the nucleic acid is generated by

i. performing a first strand synthesis on an RNA, thereby forming a first strand synthesis product; and

ii. performing a second strand synthesis on the first strand in the presence of a non-canonical nucleotide, thereby forming a second strand synthesis product.

42. The method of claim 41, further comprising selectively cleaving the RNA.

43. The method of claim 42, wherein selectively cleaving the RNA comprises treatment with RNAse H.

44. The method of claim 41, further comprising

iii. fragmenting the first and second strand synthesis products, thereby generating fragmented first and second strand synthesis products;

iv. performing end repair; and

v. performing 5' phosphorylation.

45. The method of claim 23, wherein the pool of nucleic acids originates from a population of sorted cells.

46. The method of claim 45, wherein the pool of nucleic acids originates from a single cell.

47. The method of claim 45, further comprising sorting cells into a multiwell plate, microarray, microfluidic device, or slide and thereby generating the population of sorted cells.

IDT Exhibit 1020.5442

48. The method of claim 47, wherein the sorting is performed according to a cell surface marker.

49. The method of claim 47, wherein the sorting is performed according to optical properties of the cells.

50. The method of claim 47, wherein the sorting is performed according to cell size.

51. The method of claim 23, wherein the pool of nucleic acids comprises bacterial ribosomal RNA, mitochondrial DNA, human globin mRNA, human cytoplasmic rRNA, human mitochondrial rRNA, grape cytoplasmic rRNA, grape mitochondrial rRNA, or grape chloroplast rRNA.

52. The method of claim 36, wherein the restriction endonuclease is BspQI.

53. The method of claim 23, wherein the 3' extension reaction is performed using a hot start polymerase.

54. The method of claim 53, wherein the 3' extension reaction is performed using MyTaq polymerase.

55. A method of adapter ligation to create a strand retained library of nucleic acids comprising desired and undesired nucleic acids, comprising:

a. mixing pool of templates comprising an undesired nucleic acid comprising one or more non-canonical nucleotides and a desired nucleic acid comprising one or more non-canonical nucleotides with a plurality of partial duplex primers each comprising a 3' overhang,

b. annealing to the templates the plurality of partial duplex primers;

c. performing primer extension reaction along the templates, thereby forming double-stranded nucleic acids each comprising a primer extension product;

d. ligating an adapter to at least one 5' end of the primer extension products; and

e. cleaving the templates from the double-stranded nucleic acids with a cleavage agent that is specific for nucleic acids comprising the one or more nucleotides.

56. The method of claim 55, wherein the plurality of partial duplex primers comprise at least two partial duplex primers with dissimilar 3' overhang sequences.

57. The method of claim 55 or 56, wherein the plurality of partial duplex primers comprise a shared sequence within a double-stranded portion.

58. The method of claim 55, further comprising a step comprising performing a primer extension reaction along the adapter.

59. The method of claim 55, wherein step e comprises cleaving a base portion of the one or more non-canonical nucleotides, thereby forming an abasic site.

IDT Exhibit 1020.5443

60. The method of claim 55, wherein the cleavage agent comprises a glycosylase.

61. The method of claim 60, wherein the glycosylase is UNG or UDG.

62. The method of claim 55, wherein the cleavage agent comprises a primary amine.

63. The method of claim 55, wherein the cleavage agent comprises a polyamine.

64. The method of claim 63, wherein the polyamine is DMED.

65. The method of claim 55, wherein the cleavage agents comprises a glycosylase and a polyamine.

66. The method of claim 55, wherein the cleavage agent comprises endonuclease V.

67. The method of claim 55, wherein the one or more non-canonical nucleotides comprise uracil or inosine.

68. The method of claim 55, wherein the one or more non-canonical nucleotides comprise uracil and inosine.

69. The method of claim 55, wherein the pool of templates is generated by

i. performing a first strand synthesis on an RNA in the presence of the one or more non-canonical nucleotides, thereby forming a first strand synthesis product;

ii. and performing a fragmentation reaction.

70. The method of claim 69, further comprising selectively cleaving the RNA.

71. The method of claim 70, wherein selectively cleaving the RNA comprises treatment with RNAse H.

72. The method of claim 69, wherein the fragmentation reaction comprises utilizing a cleavage agent targeting the one or more non-canonical nucleotides.

73. The method of claim 72, wherein fragmentation reaction comprises cleaving a base portion of the one or more non-canonical nucleotides, thereby forming an abasic site.

74. The method of claim 72, wherein the cleavage agent comprises a glycosylase.

75. The method of claim 74, wherein the glycosylase is UNG or UDG.

76. The method of claim 72, wherein the cleavage agent comprises a primary amine.

77. The method of claim 72, wherein the cleavage agent comprises a polyamine.

78. The method of claim 77, wherein the polyamine is DMED.

79. The method of claim 72, wherein the cleavage agents comprises a glycosylase and a polyamine.

80. The method of claim 72, wherein the cleavage agent comprises endonuclease V.

81. The method of any one of claims 55-80, wherein the adapter comprises a recognition sequence for a restriction endonuclease specific for double-stranded DNA.

82. The method of claim 81, further comprising,

IDT Exhibit 1020.5444

f. hybridizing a probe to a sequence of the primer extension product;

g. extending the probe with a DNA polymerase, thereby producing a partial duplex nucleic acid,

h. treating the partial duplex nucleic acid with a restriction endonuclease specific for double-stranded DNA, thereby cleaving double-stranded DNA fragments at the recognition sequence.

83. The method of claim 82, further comprising performing PCR with a primer that is hybridizable a sequence reverse complimentary to the adapter, thereby amplifying a desired nucleic acid in the pool of templates.

84. The method of claim 83, further comprising sequencing a portion of the desired nucleic acid.

85. The method of claim 55, wherein the pool of templates originates from a population of sorted cells.

86. The method of claim 85, wherein the pool of templates originates from a single cell.

87. The method of claim 85, further comprising sorting cells into a multiwell plate, microarray, microfluidic device, or slide and thereby generating the population of sorted cells.

88. The method of claim 87, wherein the sorting is performed according to a cell surface marker.

89. The method of claim 87, wherein the sorting is performed according to optical properties of the cells.

90. The method of claim 87, wherein the sorting is performed according to cell size.

91. The method of claim 55, wherein the pool of templates comprises bacterial ribosomal RNA, mitochondrial DNA, human globin mRNA, human cytoplasmic rRNA, human mitochondrial rRNA, grape cytoplasmic rRNA, grape mitochondrial rRNA, or grape chloroplast rRNA.

92. The method of claim 81, wherein the restriction endonuclease is BspQI.

93. The method of claim 55, wherein the primer extension reaction is performed using a hot start polymerase.

94. The method of claim 93, wherein the primer extension reaction is performed using MyTaq polymerase.

95. A method of adapter ligation to create a strand retained library of nucleic acids with desired and undesired nucleic acids, comprising:

IDT Exhibit 1020.5445

a. mixing a pool of templates comprising an undesired nucleic acid and a desired nucleic acid with a plurality of partial duplex primers each comprising a 3' overhang;

b. annealing to the templates the plurality of partial duplex primers;

c. performing primer extension reaction along the templates, thereby forming double-stranded nucleic acids each comprising a primer extension product;

d. ligating an adapter to at least one 5' end of the primer extension product; and

e. cleaving the primer extension products from the double-stranded nucleic acids with a cleavage agent that is specific for nucleic acids comprising the one or more nucleotides.

96. The method of claim 95, wherein the plurality of partial duplex primers comprise at least two partial duplex primers with dissimilar 3' overhang sequences.

97. The method of claim 95 or 96, wherein the plurality of partial duplex primers comprise a shared sequence within a double-stranded portion.

98. The method of claim 97, wherein the strand of the plurality of partial duplex primers with the 3' overhang lacks adenines in the shared sequence within the double-stranded portion.

99. The method of claim 95, further comprising a step comprising performing a primer extension reaction along the adapter.

100. The method of claim 99, wherein the primer extension reaction is performed in the presence of one or more non-canonical nucleotides.

101. The method of claim 95, wherein step e comprises cleaving a base portion of the one or more non-canonical nucleotides, thereby forming an abasic site.

102. The method of claim 95, wherein the cleavage agent comprises a glycosylase.

103. The method of claim 102, wherein the glycosylase is UNG or UDG.

104. The method of claim 95, wherein the cleavage agent comprises a primary amine.

105. The method of claim 95, wherein the cleavage agent comprises a polyamine.

106. The method of claim 105, wherein the polyamine is DMED.

107. The method of claim 95, wherein the cleavage agents comprises a glycosylase and a polyamine.

108. The method of claim 95, wherein the cleavage agent comprises endonuclease V.

109. The method of claim 95 or 100, wherein the one or more non-canonical nucleotides comprise uracil or inosine.

110. The method of claim 95, wherein the pool of templates is generated by

i. performing a first strand synthesis on an RNA in the presence of the one or more non-canonical nucleotides, thereby forming a first strand synthesis product;

IDT Exhibit 1020.5446

ii. and performing a fragmentation reaction.

111. The method of claim 110, further comprising selectively cleaving the RNA.

112. The method of claim 111, wherein selectively cleaving the RNA comprises treatment with RNAse H.

113. The method of claim 110, wherein the fragmentation reaction comprises utilizing a cleavage agent targeting the one or more non-canonical nucleotides.

114. The method of claim 113, wherein the fragmentation reaction comprises cleaving a base portion of the one or more non-canonical nucleotides, thereby forming an abasic site.

115. The method of claim 113, wherein the cleavage agent comprises a glycosylase.

116. The method of claim 115, wherein the glycosylase is UNG or UDG.

117. The method of claim 113, wherein the cleavage agent comprises a primary amine.

118. The method of claim 113, wherein the cleavage agent comprises a polyamine.

119. The method of claim 118, wherein the polyamine is DMED.

120. The method of claim 113, wherein the cleavage agents comprises a glycosylase and a polyamine.

121. The method of claim 113, wherein the cleavage agent comprises endonuclease V.

122. The method of any one of claims 95-121, wherein the adapter comprises a recognition sequence for a restriction endonuclease specific for double-stranded DNA.

123. The method of claim 122, further comprising,

f. hybridizing a probe to a sequence of the undesired nucleic acid;

g. extending the probe with a DNA polymerase, thereby producing a partial duplex nucleic acid,

h. treating the partial duplex nucleic acid with a restriction endonuclease specific for double-stranded DNA, thereby cleaving double-stranded DNA fragments at the recognition sequence.

124. The method of claim 123, further comprising performing PCR with a set of primers that are hybridizable to a sequence reverse complimentary to the adapter and the shared sequence of the partial duplex primers opposite the 3' overhang, thereby amplifying a desired nucleic acid in the pool of templates.

125. The method of claim 124, further comprising sequencing a portion of the desired nucleic acid.

126. The method of claim 95, wherein the pool of templates originates from a population of sorted cells.

127. The method of claim 126, wherein the pool of templates originates from a single cell.

IDT Exhibit 1020.5447

128. The method of claim 126, further comprising sorting cells into a multiwell plate, microarray, microfluidic device, or slide and thereby generating the population of sorted cells.

129. The method of claim 128, wherein the sorting is performed according to a cell surface marker.

130. The method of claim 128, wherein the sorting is performed according to optical properties of the cells.

131. The method of claim 128, wherein the sorting is performed according to cell size.

132. The method of claim 95, wherein the pool of templates comprises bacterial ribosomal RNA, mitochondrial DNA, human globin mRNA, human cytoplasmic rRNA, human mitochondrial rRNA, grape cytoplasmic rRNA, grape mitochondrial rRNA, or grape chloroplast rRNA.

133. The method of claim 122, wherein the restriction endonuclease is BspQI.

134. The method of claim 95, wherein the primer extension reaction is performed using a hot start polymerase.

135. The method of claim 134, wherein the primer extension reaction is performed using MyTaq polymerase

136. A kit comprising;

a. a restriction endonuclease;

b. a first adapter comprising one or more non-canonical nucleotides on one strand and lacking 5' phosphates;

c. a second adapter lacking said one or more non-canonical nucleotides and lacking 5' phosphates;

d. a ligase;

e. a polymerase;

f. a cleavage agent;

g. a library of probes; and

h. a set of primers specific to the adaptor sequences;

wherein the second adapter comprises a recognition sequence for the restriction endonuclease.

137. A kit comprising;

a. a restriction endonuclease;

b. a first adapter lacking 5' phosphates;

IDT Exhibit 1020.5448

c. a plurality of partial duplex primers each comprising a 3' overhang and comprising a shared sequence within a double-stranded portion;

d. a ligase;

e. a polymerase;

f. a cleavage agent;

g. a library of probes capable of acting as a primer for a primer extension reaction; and

h. a primer that is hybridizable a sequence reverse complimentary to the adapter;

wherein the first adapter comprises a recognition sequence for the restriction endonuclease; and

wherein the plurality of partial duplex primers comprise at least two partial duplex primers with dissimilar 3' overhang sequences.

138. A kit comprising;

a. a restriction endonuclease;

b. a first adapter lacking 5' phosphates;

c. a plurality of partial duplex primers each comprising a 3' overhang, comprising a shared sequence within a double-stranded portion, and the strand of the plurality of partial duplex primers with the 3' overhang lacking adenines in the shared sequence within the double-stranded portion

d. a ligase;

e. a polymerase;

f. a cleavage agent;

g. a library of probes capable of acting as a primer for a primer extension reaction; and

h. a set of primers that are hybridizable to a sequence reverse complimentary to the adapter and the shared sequence of the partial duplex primers opposite the 3' overhang;

wherein the first adapter comprises a recognition sequence for the restriction endonuclease; and

wherein the plurality of partial duplex primers comprise at least two partial duplex primers with dissimilar 3' overhang sequences.

139. The kit as in one of claims 136-138, wherein the restriction endonuclease is BspQI.

140. The kit as in one of claims 136-138, wherein the polymerase is a hot start polymerase

141. The kit of claim 140, wherein the polymerase is MyTaq.

IDT Exhibit 1020.5449

142. The kit as in one of claims 136-138, further comprising one or more non-canonical nucleotides.

143. The kit as in one of claims 136-138, wherein the one or more non-canonical nucleotides comprise uracil or inosine.

144. The kit as in one of claims 136-138, wherein the cleavage agent comprises a glycosylase.

145. The kit as in claim 144, wherein the glycosylase is UNG or UDG.

146. The kit as in one of claims 136-138, wherein the cleavage agent comprises a primary amine.

147. The kit as in one of claims 136-138, wherein the cleavage agent comprises a polyamine.

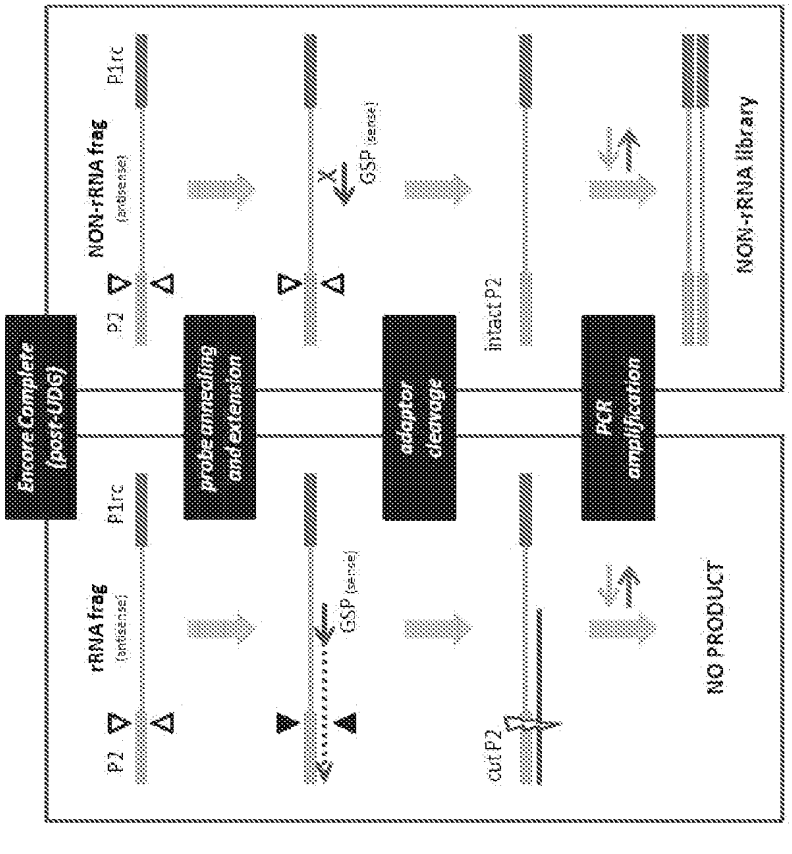
148. The kit of claim 147, wherein the polyamine is DMED.

149. The kit as in one of claims 136-138, wherein the cleavage agent comprises a glycosylase and a polyamine.

150. The kit as in one of claims 136-138, wherein the cleavage agent comprises endonuclease V.

IDT Exhibit 1020.5450

Figure 1. rRNA depletion from a nucleic acid library using insert-dependent adaptor cleavage (InDA-C).



[1] Encore Complete RNA-Seq library system is used to generate strand-specific cDNA libraries from 100 ng of *E.coli* total RNA.

Figure 2. Summary of results - Example 1.

library	probe	BspQ1	total number of reads	% of reads mapping to reference	% of mapped reads in ORFs	% of mapped reads in rRNA	% of rRNA reads mapping to sense strand	% of rRNA reads mapping to antisense strand
ctrl	-	-	27,161,879	99.3	1.0	99.0	99.7	0.3
test1	-	-	7,510,963	97.3	5.6	92.9	99.1	0.9
test2	+	-	8,679,815	97.5	5.6	92.9	99.1	0.9
test3	-	+	9,331,682	97.4	5.2	93.4	99.2	0.8
test4	+	+	9,355,578	95.1	11.7	86.1	98.1	1.9

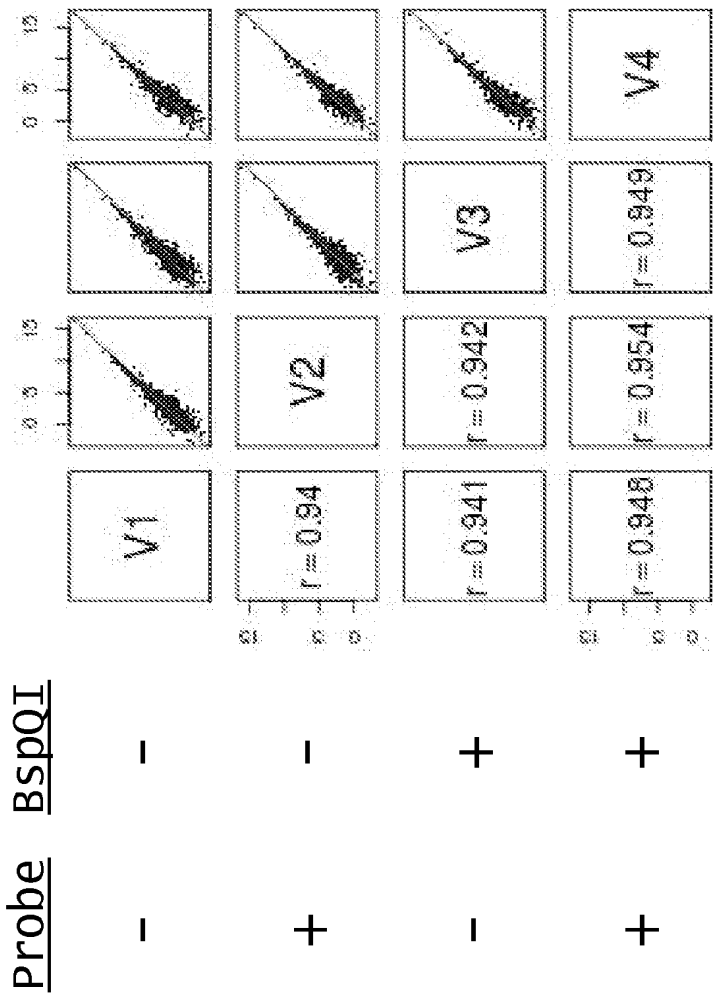


Figure 3. Comparison of expression profiles from the test libraries 1-4 in Example 1. InDA-C does not perturb transcript abundance.

Figure 4

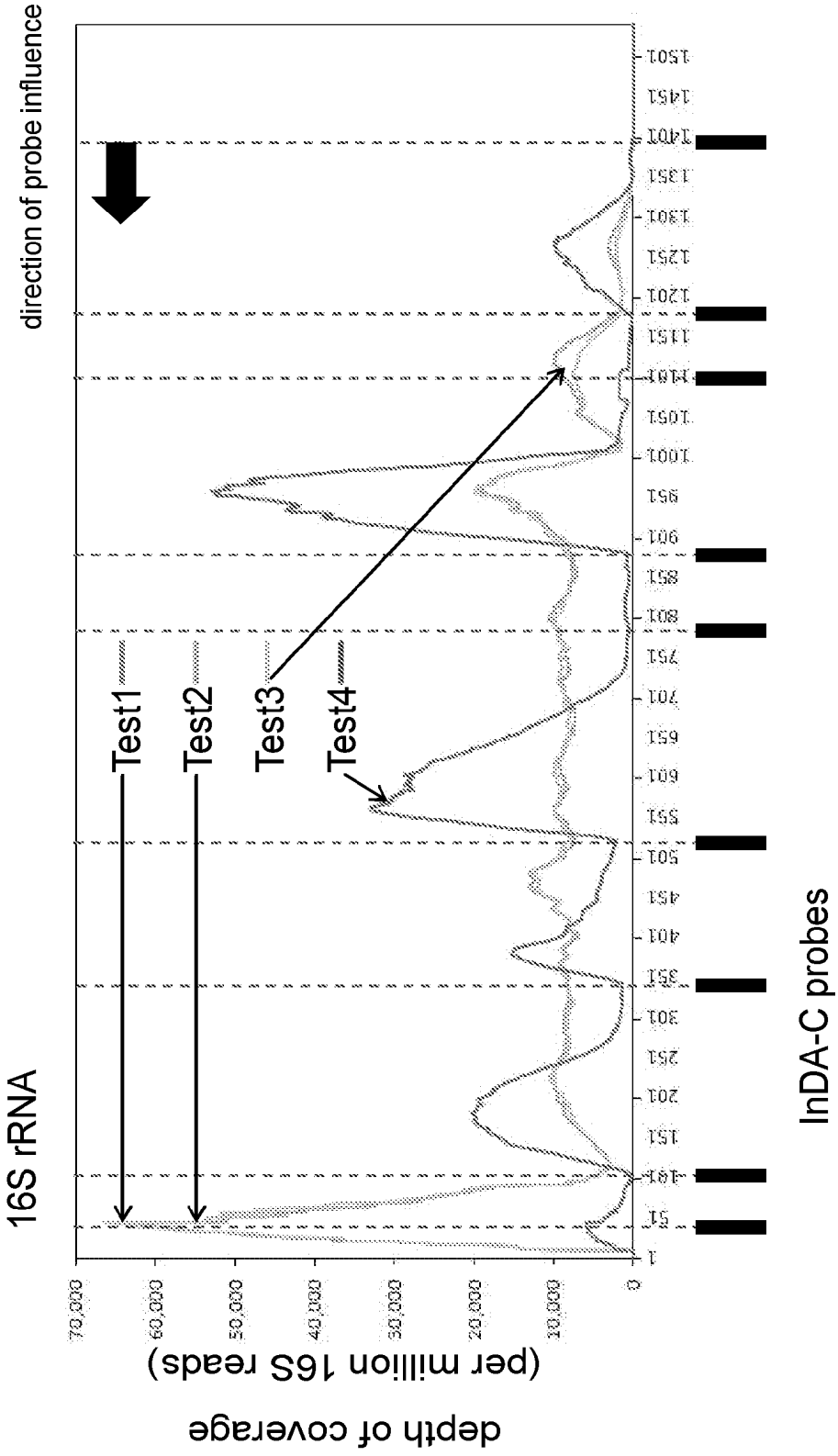

direction of probe influence
16S rRNA
Test1
Test2
Test3
Test4
InDA-C probes
depth of coverage
(per million 16S reads)
70,000
60,000
50,000
40,000
30,000
20,000
10,000
0
1
51
101
151
201
251
301
351
401
451
501
551
601
651
701
751
801
851
901
951
1001
1051
1101
1151
1201
1251
1301
1351
1401
1451
1501

Figure 5: Directional Library Construction

Figure 6: InDA-C with Single cell, cDNA double hydrolysis

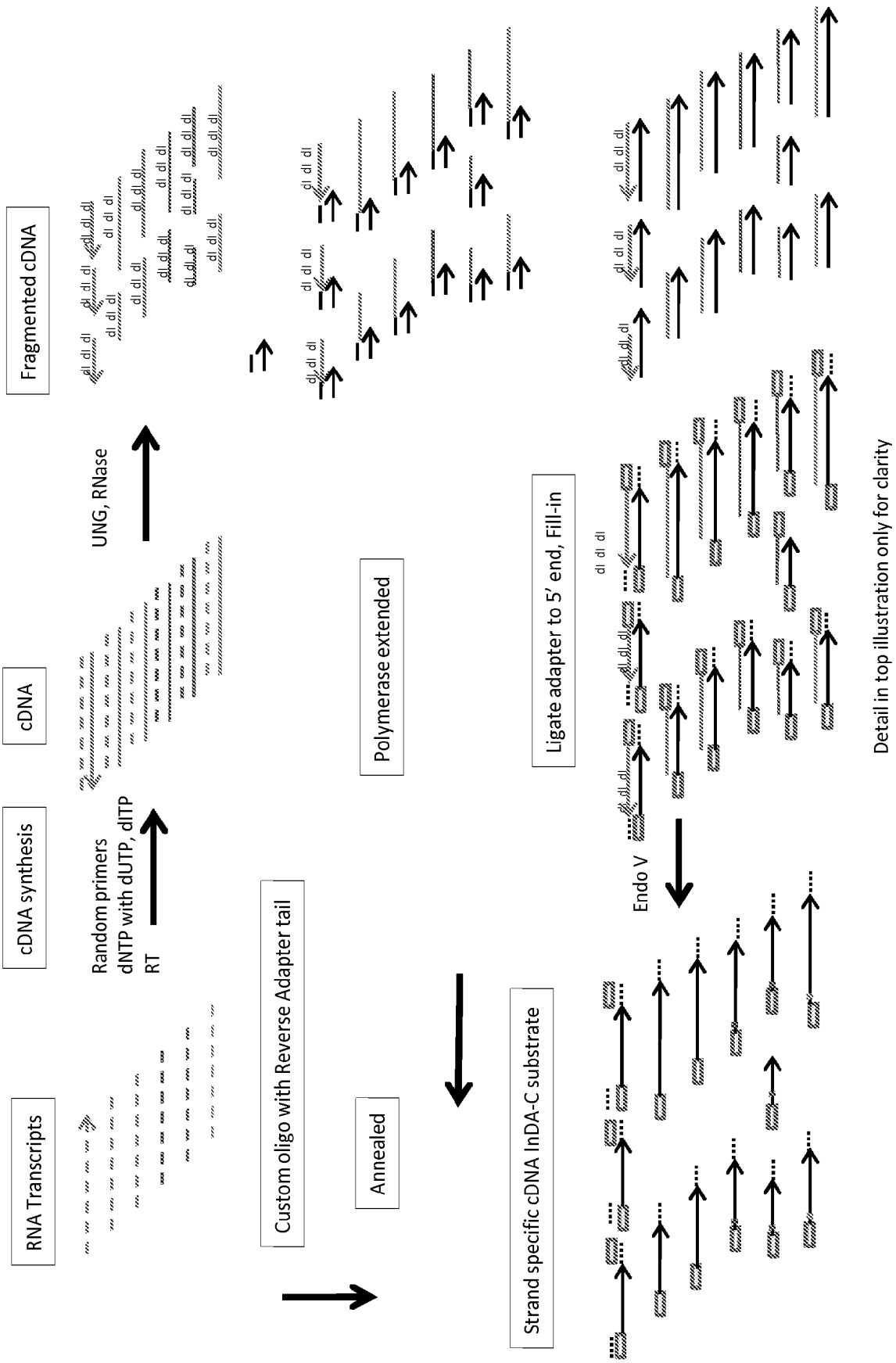


Figure 7: InDA-C with Single cell, 3 base code in second strand primer

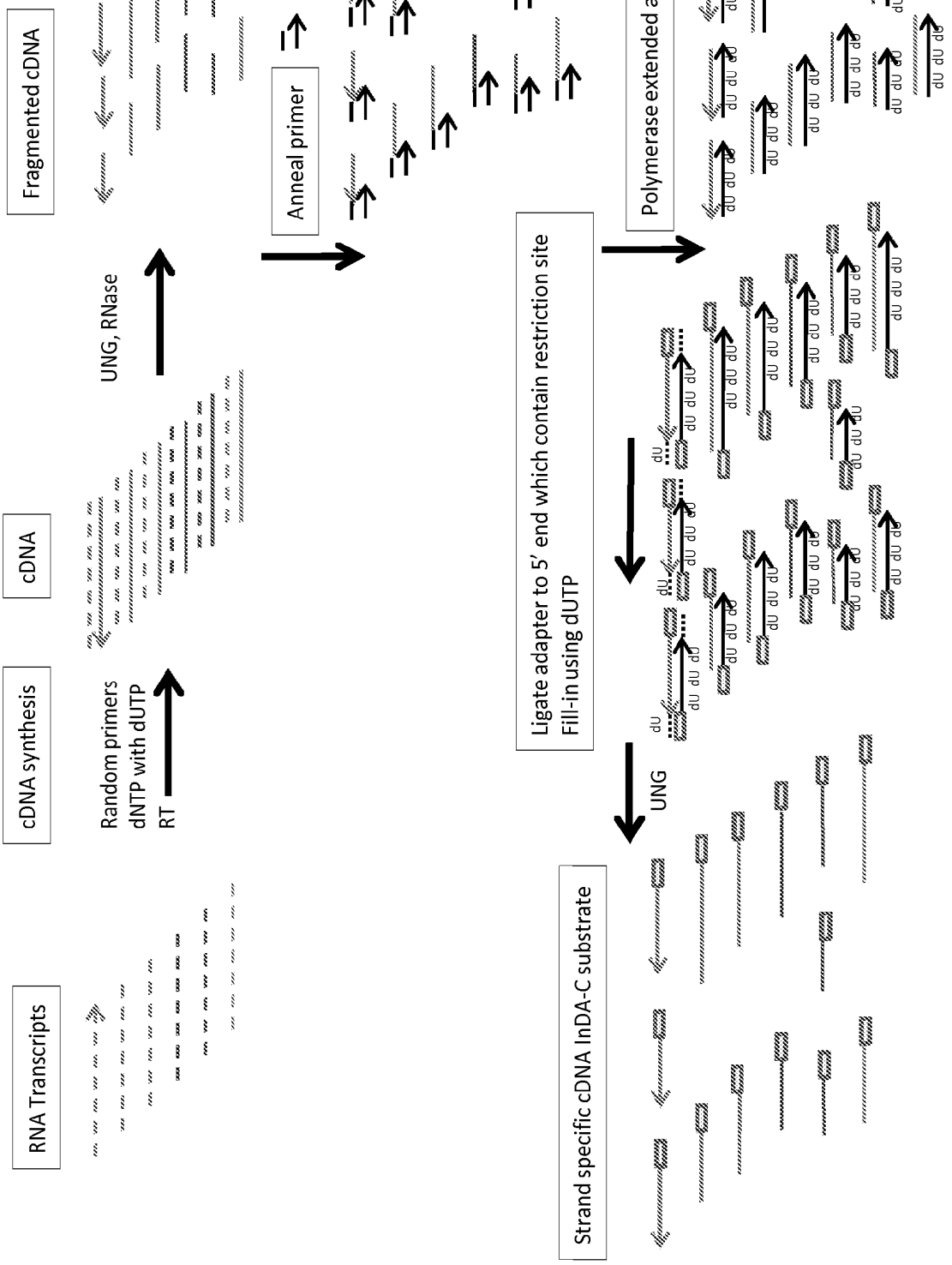
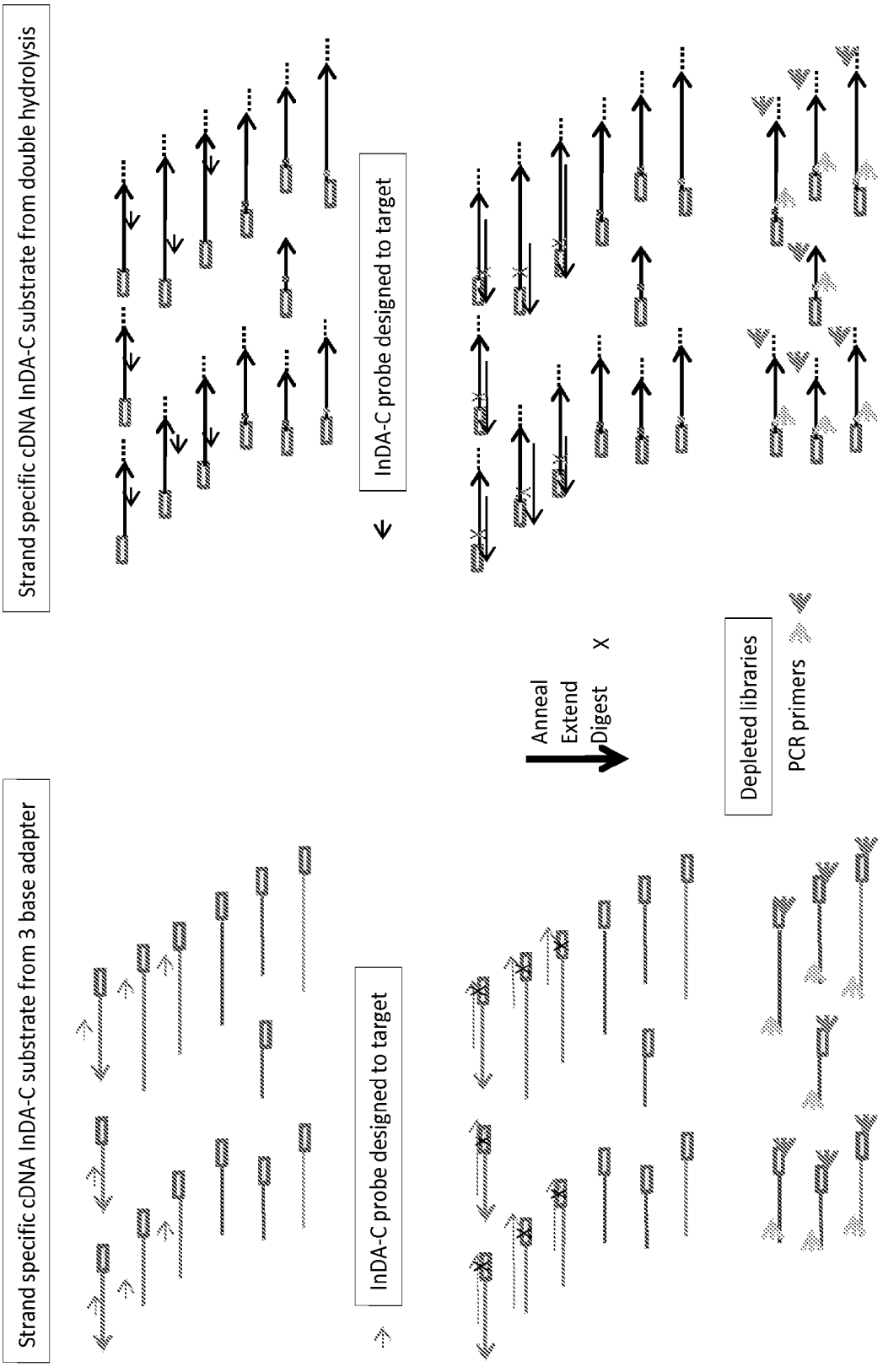


Figure 8: Primer construct

Partial duplex primer according to Example 7:
33 base double stranded region with 8 random nucleotides on lower strand
Double stranded region of lower strand is 3 base code, there are no Adenosine residues
33 base region contains sequence amenable to specific priming during PCR library enrichment

Partial duplex primer according to Example 6:
33 base double stranded region with 8 random nucleotides on lower strand
33 base region contains Type II restriction enzyme recognition site
33 base region contains sequence amenable to specific priming during PCR library enrichment

Figure 9


Strand specific cDNA InDA-C substrate from 3 base adapter
InDA-C probe designed to target
Strand specific cDNA InDA-C substrate from double hydrolysis
InDA-C probe designed to target
Anneal
Extend
Digest X
Depleted libraries
PCR primers

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
26 April 2012 (26.04.2012)

PCT

(10) International Publication Number
WO 2012/054873 A2

(51) International Patent Classification:
C12Q 1/68 (2006.01) *C12Q 1/48* (2006.01)
C12N 15/11 (2006.01)

(21) International Application Number:
PCT/US2011/057350

(22) International Filing Date:
21 October 2011 (21.10.2011)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

61/406,067	22 October 2010 (22.10.2010)	US
61/510,579	22 July 2011 (22.07.2011)	US
13/278,333	21 October 2011 (21.10.2011)	US

(71) Applicant *(for all designated States except US)*: **COLD SPRING HARBOR LABORATORY** [US/US]; One Bungtown Road, Cold Spring Harbor, NY 11724 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **HICKS, James** [US/US]; 3 Great Meadow Road, Locust Valley, New York 11560 (US). **NAVIN, Nicholas** [US/US]; 1621 Wichita St., Houston, TX 77004 (US). **TROGE, Jennifer** [US/US]; 100 Simcoe St., Oyster Bay, NY 11771 (US). **WANG, Zihua** [CN/US]; 13 Meadowlawn St., Huntington, NY 11743 (US). **WIGLER, Michael** [US/US]; 52 Snake Hill Road, Cold Spring Harbor, NY 11724 (US).

(74) Agent: **GERSHIK, Gary, J.**; Cooper & Dunham Llp, 30 Rockefeller Plaza, New York, NY 10112 (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:

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

WO 2012/054873 A2

(54) **Title**: VARIETAL COUNTING OF NUCLEIC ACIDS FOR OBTAINING GENOMIC COPY NUMBER INFORMATION

(57) **Abstract**: A method for obtaining from genomic material genomic copy number information unaffected by amplification distortion, comprising obtaining segments of the genomic material, tagging the segments with substantially unique tags to generate tagged nucleic acid molecules, such that each tagged nucleic acid molecule comprises one segment of the genomic material and a tag, subjecting the tagged nucleic acid molecules to amplification by polymerase chain reaction (PCR), generating tag associated sequence reads by sequencing the product of the PCR reaction, assigning each tagged nucleic acid molecule to a location on a genome associated with the genomic material by mapping the subsequence of each tag associated sequence read corresponding to a segment of the genomic material to a location on the genome, and counting the number of tagged nucleic acid molecules having a different tag that have been assigned to the same location on the genome, thereby obtaining genomic copy number information unaffected by amplification distortion.

IDT Exhibit 1020.5460

VARIETAL COUNTING OF NUCLEIC ACIDS FOR OBTAINING GENOMIC COPY NUMBER INFORMATION

This application claims priority of U.S. Serial No. 13/278,333, filed October 21, 2011 and U.S. Provisional Applications Nos. 61/510,579, filed July 22, 2011, and 61/406,067, filed October 22, 2010 the contents of each of which are hereby incorporated by reference in their entirety.

The invention disclosed herein was supported by the government under grant No. W81XWH-09-1-0591 from the Department of Defense. Accordingly, the U. S. Government has certain rights to this invention.

Throughout this application, various publications are referenced by numbers in parentheses. Full citations for these references may be found at the end of the specification immediately preceding the claims. The disclosures of these publications in their entireties are hereby incorporated by reference into this application to more fully describe the state of the art to which this invention pertains.

Background of the Invention

Genomic copy number information is commonly obtained using whole genome amplification (WGA). The endemic problem with the WGA method is over-sampling of certain regions, yielding a non-uniform amplification of the genome (1). WGA methods begin with the step that initiates the process, a polymerase (Phi 29) makes a strand from genomic DNA utilizing a random primer coupled to an adaptor for subsequent PCR (Figure 1). If the input DNA strands are referred to as the "0-th derivative", and the first synthesized strand as a "first derivative," subsequent strands are called the (n+1)-th derivative if their template was an n-th derivative. Only strands that are 2-nd derivative or higher become amplified in the PCR step, resulting in a 'stacking' over the regions 'chosen' by the polymerase for the first derivative.

Coverage of the genome by sequencing WGA of single cell DNA is limited by the stacking phenomenon (Figure 2). Thus it is difficult to obtain single cell measurements, particularly when based on WGA, due to distortions that originate from stochastic sampling and amplification steps. Moreover, the current method of WGA is a black box, with the unspecified reagents purchased from a vendor, which hampers optimization. Moreover, the WGA method does not extend to a method usable for single cell RNA profiling.

Ligation-mediated PCR was developed in an attempt to solve the above-identified problems inherent in WGA. In this method, adaptors are ligated to an *Mse*I restriction endonuclease digest of genomic DNA from a single cell, followed by PCR amplification using primers complementary to the adaptors. The amplified DNA is then used for CGH or DNA sequencing (2,3). However, like WGA, the method still requires an amplification step.

Parameswaran et al. (2007) and U.S. Patent No. 7,622,281 describe methods of labeling nucleic acid molecules with barcodes for the purpose of identifying the source of the nucleic acid molecules, thereby allowing for high-throughput sequencing of multiple samples (4,5). Eid et al. (2009) describe a single molecule sequencing method wherein single-molecule real time sequencing data is obtained from a DNA polymerase performing uninterrupted template-directed synthesis using four distinguishable fluorescently labeled dNTPs (6). However, these methods do not provide genomic information unaffected by amplification distortion.

Miner et al. (2004) describe a method of molecular barcoding to label template DNA prior to PCR amplification, and report that the method allows for the identification of contaminant and redundant sequences by counting only distinctly tagged sequences (22). U.S. Patent No. 7,537,897 describes methods for molecular counting by labeling molecules of an input sample with unique oligonucleotide tags and subsequently amplifying and counting the number of different tags (23). Miner et al. and U.S. Patent No. 7,537,897 both describe labeling of input nucleic acid molecules by ligation, which has been found to be an inefficient reaction.

McCloskey et al. (2007) describe a method of molecular encoding which does not use ligation but instead uses template specific primers to barcode template DNA molecules prior to PCR amplification (24). However, such a method requires that template specific primers be made for each species of template DNA molecule studied.

As described herein, obtaining accurate genomic copy number information by high-throughput sequencing of genomic DNA prepared by WGA methods is hampered by the copy number distortions introduced by non-uniform amplification of genomic DNA. Thus, there exists a need for a method that allows for copy number determination free of distortions caused by amplification steps and which allows for accurate and efficient copy number determination of complex samples. Such a method should also be robust using existing methodologies for high volume, massively parallel sequencing.

SUMMARY OF THE INVENTION

A method is provided for obtaining from genomic material genomic copy number information unaffected by amplification distortion, comprising:

a) obtaining segments of the genomic material;

b) tagging the segments with substantially unique tags to generate tagged nucleic acid molecules, such that each tagged nucleic acid molecule comprises one segment of the genomic material from step (a) and a tag;

c) subjecting the tagged nucleic acid molecules to amplification by polymerase chain reaction (PCR);

d) generating tag associated sequence reads by sequencing the product of step (c);

e) assigning each tagged nucleic acid molecule to a location on a genome associated with the genomic material by mapping the subsequence of each tag associated sequence read corresponding to a segment of the genomic material to a location on the genome; and

f) counting the number of tagged nucleic acid molecules having a different tag that have been assigned to the same location on the genome,

thereby obtaining genomic copy number information unaffected by amplification distortion.

Also provided is a method for obtaining from mRNA transcripts mRNA copy number information unaffected by amplification distortion, comprising:

a) obtaining cDNA from the mRNA transcripts;

b) optionally obtaining segments of the cDNA;

c) tagging the cDNA or segments of cDNA with substantially unique tags to generate tagged nucleic acid molecules, such that each tagged nucleic acid molecule comprises

i) one cDNA from step (a) or one segment of the cDNA from step (b), and

ii) a tag;

d) subjecting the tagged nucleic acid molecules to amplification by polymerase chain reaction (PCR);

e) generating tag associated sequence reads by sequencing the product of step (d);

f) assigning each tagged nucleic acid molecule to a location on a cDNA library associated with the mRNA transcripts by mapping the subsequence of each tag associated sequence read corresponding to a cDNA or a segment of cDNA to a location on the cDNA library; and

g) counting the number of tagged nucleic acid molecules having a different tag that have been assigned to the same location on the cDNA library,

thereby obtaining mRNA copy number information unaffected by amplification distortion.

Also provided is a method for obtaining from mRNA transcripts mRNA copy number information unaffected by amplification distortion, comprising:

a) generating tagged nucleic acid molecules, comprising:

i) subjecting the mRNA transcripts to a polymerase reaction in the presence of primers capable of hybridizing to the polyA tail of the mRNA transcripts under conditions that promote the formation of only one complement, thereby generating first order derivative strands;

ii) adding a polynucleotide tail to the first order derivative strands; and

iii) subjecting the first order derivative strands to a polymerase reaction in the presence of primers capable of hybridizing to the polynucleotide tail added in step (ii) under conditions that promote the formation of only one complement, thereby generating second order derivative strands,

wherein the primers of at least one of steps (i) and (iii) comprise substantially unique primers, thereby generating tagged nucleic acid molecules;

b) subjecting the tagged nucleic acid molecules to polymerase chain reaction (PCR) amplification;

c) generating tag associated sequence reads by sequencing the product of step (b);

d) assigning each tagged nucleic acid molecule to a location on a cDNA library associated with the mRNA transcripts by mapping the subsequence of each tag associated sequence read corresponding to a mRNA transcript to a location on the cDNA library; and

e) counting the number of tagged nucleic acid molecules having a different tag that have been assigned to the same location on the cDNA library,

thereby obtaining mRNA copy number information unaffected by amplification distortion.

Also provided is a method for obtaining from genomic material DNA methylation information unaffected by amplification distortion, comprising:

a) obtaining segments of the genomic material;

b) adding a polynucleotide tail to the ends of the segments of the genomic material to generate zero-th order derivative strands;

c) subjecting the zero-th order derivative strands to a polymerase reaction in the presence essentially unique primers, wherein the essentially unique primers are capable of hybridizing to the polynucleotide tail of the zero-th order strands, under conditions that promote the formation of only one complement, thereby generating tagged nucleic acid molecules;

d) separating the tagged nucleic acid molecules into a group consisting of hemi-methylated tagged nucleic acid molecules and a group consisting of unmethylated tagged nucleic acid molecules;

e) subjecting each group of step (d) to amplification by polymerase chain reaction (PCR);

f) generating tag associated sequence reads by sequencing the product of step (e);

g) assigning each tagged nucleic acid molecule to a location on a genome associated with the genomic material by mapping the subsequence of each tag associated sequence read corresponding to a segment of the genomic material to a location on the genome; and

h) counting the number of tagged nucleic acid molecules having a different tag that have been assigned to the same location on the genome,

thereby obtaining DNA methylation information unaffected by amplification distortion.

Also provided is a composition of matter derived from genomic material comprising tagged nucleic acid molecules, said tagged nucleic acid molecules being produced by a process comprising:

a) obtaining segments of the genomic material;

b) adding a polynucleotide tail to the ends of the segments of the genomic material to generate zero-th order derivative strands;

c) subjecting the zero-th order derivative strands of step (b) to a polymerase reaction in the presence of primers capable of hybridizing to the polynucleotide tail of the zero-th order strands under conditions that promote the formation of only one complement, thereby generating first order derivative strands;

d) adding a polynucleotide tail to the first order derivative strands;

e) subjecting the first order derivative strands to a polymerase reaction in the presence of primers capable of hybridizing to the polynucleotide tail of the first order derivative strands under conditions that promote the formation of only one complement, thereby generating second order derivative strands,

wherein the primers of at least one of steps (c) and (e) comprise substantially unique primers, thereby generating tagged nucleic acid molecules.

Also provided is a composition of matter derived from genomic material comprising tagged nucleic acid molecules, said tagged nucleic acid molecules being produced by a process comprising:

a) obtaining segments of genomic material;

b) adding a polynucleotide tail to the ends of the segments of genomic material to generate zero-th order derivative strands,

c) subjecting the zero-th order derivative strands of step (b) to a ligation reaction in the presence of primers capable of hybridizing to the polynucleotide tail of the zero-th order strands added in step (b) under conditions that promote the ligation of a primer to the 5' ends of the zero-th order derivative strands,

d) subjecting the product of step (c) to a polymerase reaction in the presence of primers capable of hybridizing to the polynucleotide tail added in step (b) under conditions which promote the formation of only one complement, wherein the primers of step (d) have different

nucleotide sequences than the primers of step (c), and wherein the polymerase has 3'-5' proofreading activity,

wherein the primers of at least one of steps (c) and (d) comprise substantially unique primers, thereby generating tagged nucleic acid molecules.

Also provided is a composition of matter derived from a mRNA transcripts comprising tagged nucleic acid molecules, said tagged nucleic acid molecules being produced by a process comprising:

a) obtaining cDNA from the mRNA transcripts;

b) optionally obtaining segments of the cDNA; and

c) tagging the cDNA or segments of cDNA with substantially unique tags to generate tagged nucleic acid molecules, wherein each tagged nucleic acid molecule comprises

i) one cDNA from step (a) or one segment of the cDNA from step (b) and

ii) a tag.

Also provided is a composition of matter derived from mRNA transcripts comprising tagged nucleic acid molecules, said tagged nucleic acid molecules being produced by a process comprising:

a) obtaining mRNA transcripts;

b) subjecting the mRNA transcripts to a polymerase reaction in the presence of primers capable of hybridizing to the polyA tail of the mRNA transcripts under conditions that promote the formation of only one complement, thereby generating first order derivative strands;

c) adding a polynucleotide tail to the first order derivative strands;

d) subjecting the first order derivative strands to a polymerase reaction in the presence of primers capable of hybridizing to the

polynucleotide tail added in step (c) under conditions that promote the formation of only one complement, thereby generating second order derivative strands,

wherein the primers of at least one of steps (b) and d) comprise substantially unique primers, thereby generating tagged nucleic acid molecules.

Also provided is a kit for determining nucleic acid copy number information unaffected by amplification distortion comprising:

a) a terminal transferase, and

b) plurality of substantially unique primers,

wherein the substantially unique primers comprise substantially unique tags, and wherein the substantially unique primers are capable of hybridizing to the polynucleotide tail of a nucleic acid molecule which has a polynucleotide tail added by the terminal transferase.

BRIEF DESCRIPTION OF THE FIGURES

Figure 1. Whole Genome Amplification (WGA).

Figure 2. Genome coverage comparison.

Genomic coverage comparison of 7 sequencing lanes from one single cell library versus 7 lanes from a million cell library. Stacking causes the single cell reads to be concentrated over the first derivative reads from WGA. More reads yield diminishing returns.

Figure 3. Sequencing read counts for CpG dinucleotides within selected CpG islands

Normal cell line SKN-1 (A,C,E) and breast cancer line MDA-MB-231 (B,D,F). Orange indicates T residues (unmethylated). Blue indicates C residues (methylated). Counts of the plus strand are above the X-axis, and counts on the minus strand are below.

Figure 4. Comparison of array CGH with 'sequence counting' for analyzing copy number data.

Graph shows a portion of chromosome 17 from FFPE sample JZ33 including raw (gray) ratio and segmented (green) data from array CGH, and raw sequence count data (orange) and its segmentation (blue).

Figure 5. ROMA CGH of sKBR3 compared with profiles of 3 independent single cell isolates derived from sparse DNA sequencing.

Figure 6. Sequencing across the breakpoints of a homozygous deletion in cell on chromosome 5 line T47D.

Figure 7. Flowchart for genomic profiling by representational sequence counting

Figure 8. Scheme for varietal counting. Varietal tags are shown as random nucleotide sequences (N_{10})

Flowchart of one embodiment of the invention.

Figure 9. Nick ligation protocol

The sequence of primer A includes universal primer A (PrA), varietal tag (B), and oligo dT+CATG, the sequence of primer B includes universal primer B (PrB) and oligo

dT+CATG, PrA and PrB are universal primers, and cPrA, cPrB, and cB are complementary to PrA, PrB and B, respectively.

Figure 10. Schema for targeted sequencing of regional markers

UfBPrA is forward primer, B is tag, UrPrR is reverse primer, PrB is gene-specific reverse primer, Uf is universal forward primer, PrA is gene specific forward primer, Ur is universal reverse primer.

Figure 11. Comparison of Varietal Counting with Sequence Counting

Figure 11A shows genome wide copy number data. Figure 11B shows partial genome copy number data.

DETAILED DESCRIPTION OF THE INVENTION

A method is provided for obtaining from genomic material genomic copy number information unaffected by amplification distortion, comprising:

a) obtaining segments of the genomic material;

b) tagging the segments with substantially unique tags to generate tagged nucleic acid molecules, such that each tagged nucleic acid molecule comprises one segment of the genomic material from step (a) and a tag;

c) subjecting the tagged nucleic acid molecules to amplification by polymerase chain reaction (PCR);

d) generating tag associated sequence reads by sequencing the product of step (c);

e) assigning each tagged nucleic acid molecule to a location on a genome associated with the genomic material by mapping the subsequence of each tag associated sequence read corresponding to a segment of the genomic material to a location on the genome; and

f) counting the number of tagged nucleic acid molecules having a different tag that have been assigned to the same location on the genome,

thereby obtaining genomic copy number information unaffected by amplification distortion.

In an embodiment, the method further comprises estimating a genomic copy number of a region of the genome comprising more than one location on the genome by assigning as the copy number of the region the highest count obtained in step (f) for the locations within the region.

In an embodiment, the method further comprises comparing a count obtained in step (f) for a location on the genome to a count for the same location obtained from a reference sample, thereby estimating a relative genomic copy number of the location.

In an embodiment, the method further comprises

a) summing the counts obtained in step (f) for locations on the genome which comprise a first region of the genome, wherein the first region comprises more than one location;

b) summing the counts obtained in step (f) for locations on the genome which comprise a second region of the genome, wherein the second region is comprised of a number of locations which is comparable to the number of locations of the first region;

c) comparing the value obtained in step (a) to the value obtained in step (b),

thereby estimating the relative genomic copy number of the first region of the genome to the genomic copy number of the second region of the genome.

Step (b) of the above embodiment may further comprise

i) summing the counts obtained in step (f) for locations on the genome which comprise a third region of the genome, wherein the third region is comprised of a number of locations which is comparable to the number of locations of the first region; and

ii) obtaining an average of the sum of the counts obtained in step (f) for locations which comprise the second region and the sum of the counts obtained in step (f) for locations which comprise the third region.

In an embodiment of the method, the second region of the genome comprises a centromere.

In an embodiment, the method further comprises summing the counts obtained in step (f) for locations which comprise a region of the genome, and comparing the sum to a sum obtained from a reference sample for the same region of the genome, thereby estimating a relative genomic copy number of the region of the genome.

Also provided is a method for obtaining from mRNA transcripts mRNA copy number information unaffected by amplification distortion, comprising:

a) obtaining cDNA from the mRNA transcripts;

b) optionally obtaining segments of the cDNA;

c) tagging the cDNA or segments of cDNA with substantially unique tags to generate tagged nucleic acid molecules, such that each tagged nucleic acid molecule comprises

i) one cDNA from step (a) or one segment of the cDNA from step (b), and

ii) a tag;

d) subjecting the tagged nucleic acid molecules to amplification by polymerase chain reaction (PCR);

e) generating tag associated sequence reads by sequencing the product of step (d);

f) assigning each tagged nucleic acid molecule to a location on a cDNA library associated with the mRNA transcripts by mapping the subsequence of each tag associated sequence read corresponding to a cDNA or a segment of cDNA to a location on the cDNA library; and

g) counting the number of tagged nucleic acid molecules having a different tag that have been assigned to the same location on the cDNA library,

thereby obtaining mRNA copy number information unaffected by amplification distortion.

Also provided is a method for obtaining from mRNA transcripts mRNA copy number information unaffected by amplification distortion, comprising:

a) generating tagged nucleic acid molecules, comprising:

i) subjecting the mRNA transcripts to a polymerase reaction in the presence of primers capable of hybridizing to the polyA tail of the mRNA transcripts under conditions that promote the

formation of only one complement, thereby generating first order derivative strands;

ii) adding a polynucleotide tail to the first order derivative strands; and

iii) subjecting the first order derivative strands to a polymerase reaction in the presence of primers capable of hybridizing to the polynucleotide tail added in step (ii) under conditions that promote the formation of only one complement, thereby generating second order derivative strands,

wherein the primers of at least one of steps (i) and (iii) comprise substantially unique primers, thereby generating tagged nucleic acid molecules;

a) subjecting the tagged nucleic acid molecules to polymerase chain reaction (PCR) amplification;

b) generating tag associated sequence reads by sequencing the product of step (b);

c) assigning each tagged nucleic acid molecule to a location on a cDNA library associated with the mRNA transcripts by mapping the subsequence of each tag associated sequence read corresponding to a mRNA transcript to a location on the cDNA library; and

d) counting the number of tagged nucleic acid molecules having a different tag that have been assigned to the same location on the cDNA library,

thereby obtaining mRNA copy number information unaffected by amplification distortion.

Also provided is a method for obtaining from genomic material DNA methylation information unaffected by amplification distortion, comprising:

a) obtaining segments of the genomic material;

b) adding a polynucleotide tail to the ends of the segments of the genomic material to generate zero-th order derivative strands;

c) subjecting the zero-th order derivative strands to a polymerase reaction in the presence essentially unique primers, wherein the essentially unique primers are capable of hybridizing to the polynucleotide tail of the zero-th order strands, under conditions that promote the formation of only one complement, thereby generating tagged nucleic acid molecules;

d) separating the tagged nucleic acid molecules into a group consisting of hemi-methylated tagged nucleic acid molecules and a group consisting of unmethylated tagged nucleic acid molecules;

e) subjecting each group of step (d) to amplification by polymerase chain reaction (PCR);

f) generating tag associated sequence reads by sequencing the product of step (e);

g) assigning each tagged nucleic acid molecule to a location on a genome associated with the genomic material by mapping the subsequence of each tag associated sequence read corresponding to a segment of the genomic material to a location on the genome; and

h) counting the number of tagged nucleic acid molecules having a different tag that have been assigned to the same location on the genome,

thereby obtaining DNA methylation information unaffected by amplification distortion.

In an embodiment of the methods, tagging the segments to generate tagged nucleic acid molecules comprises:

a) adding a polynucleotide tail to the ends of the segments of the genomic material to generate zero-th order derivative strands;

b) subjecting the zero-th order derivative strands of step (a) to a polymerase reaction in the presence of primers capable of hybridizing to the polynucleotide tail of the zero-th order strands under conditions that promote the formation of only one complement, thereby generating first order derivative strands;

c) adding a polynucleotide tail to the first order derivative strands;

d) subjecting the first order derivative strands to a polymerase reaction in the presence of primers capable of hybridizing to the polynucleotide tail of the first order derivative strands under conditions that promote the formation of only one complement, thereby generating second order derivative strands,

wherein the primers of at least one of steps (b) and (d) comprise substantially unique primers, thereby generating tagged nucleic acid molecules.

In an embodiment of the methods, tagging the segments to generate tagged nucleic acid molecules comprises:

a) adding a polynucleotide tail to the ends of the segments of genomic material to generate zero-th order derivative strands,

b) subjecting the zero-th order derivative strands of step (a) to a ligation reaction in the presence of primers capable of hybridizing to the polynucleotide tail of the zero-th order strands added in step (a) under conditions that promote the ligation of a primer to the 5' ends of the zero-th order derivative strands,

c) subjecting the product of step (b) to a polymerase reaction in the presence of primers capable of hybridizing to the polynucleotide tail added in step (a) under conditions which promote the formation of only one complement, wherein the primers of step (c) have different nucleotide sequences than the primers of step (b), and wherein the polymerase has 3'-5' proofreading activity,

wherein the primers of at least one of steps (b) and (c) comprise substantially unique primers, thereby generating tagged nucleic acid molecules.

In an embodiment of the methods, adding a polynucleotide tail comprises the use of a terminal transferase.

In an embodiment of the methods, tagging the segments to generate tagged nucleic acid molecules comprises ligation of adaptors comprising the tags to at least one end of the segments of the genomic material.

In an embodiment of the methods, the adaptors comprising the tags are ligated to only one end of the segments of the genomic material.

In an embodiment of the methods, the tags comprise a sequence that aids PCR amplification.

In an embodiment of the methods, each tagged nucleic acid molecule comprises one tag.

In an embodiment of the methods, each tagged nucleic acid molecule comprises more than one tag.

In an embodiment of the methods, segments of the genomic material are produced by restriction endonuclease digestion, mechanical shearing, heating, or sonication.

In an embodiment of the methods, segments of the cDNA are produced by restriction endonuclease digestion, mechanical shearing, heating, or sonication.

In an embodiment of the methods, the maximum copy number of a location on a cDNA library is not less than the number of tagged nucleic acid molecules having a different tag that have been assigned to the same location on the cDNA library.

An embodiment of the above methods further comprises analyzing mRNA copy number.

In an embodiment of the methods, separating the tagged nucleic acid molecules into a group consisting of hemi-methylated tagged nucleic acid molecules and a group consisting of unmethylated tagged nucleic acid molecules is by cleavage with methylation sensitive restriction enzymes, partitioning with antibodies, or partitioning with methyl-C binding proteins directed to methylated or hydroxymethylated cytosine.

In an embodiment of the methods, the tagged nucleic acid molecules are subject to hybrid capture prior to PCR or prior to sequencing.

In an embodiment of the methods, each tagged nucleic acid molecule differs at more than one nucleotide.

In an embodiment of the methods, the tag sequences further comprise a sample tag.

In an embodiment of the methods, the tagged nucleic acid molecules are pooled with a plurality of tagged nucleic acid molecules having a different sample tag prior to PCR amplification or prior to sequencing.

An embodiment of the methods further comprises deconvoluting the tag associated sequence reads by grouping the tag associated sequence reads according to sample tag.

In an embodiment of the methods, the tagged nucleic acid molecules are generated from a single species.

In an embodiment of the methods, the tagged nucleic acid molecules are generated from a single organism.

In an embodiment of the methods, tagged nucleic acid molecules are generated from a single cell.

In an embodiment, the single cell is from a needle aspirate of suspected cancer lesions.

In an embodiment, the single cell is from a core biopsy of suspected cancer lesions.

In an embodiment of the methods, the tagged nucleic acid molecules are generated from two or more organisms.

In an embodiment, the tagged nucleic acid molecules are generated from a single cell of each organism.

In an embodiment of the methods, the single species is human.

In an embodiment of the methods, the tagged nucleic acid molecules are generated from two or more species.

In an embodiment, the tagged nucleic acid molecules are generated from a population of microbes.

A further embodiment comprises comparing the genomic copy number information obtained for different species of the population to determine the relative count of those different species in the population.

Also provided is a composition of matter derived from genomic material comprising tagged nucleic acid molecules, said tagged nucleic acid molecules being produced by a process comprising:

a) obtaining segments of the genomic material;

b) adding a polynucleotide tail to the ends of the segments of the genomic material to generate zero-th order derivative strands;

c) subjecting the zero-th order derivative strands of step (b) to a polymerase reaction in the presence of primers capable of hybridizing to the polynucleotide tail of the zero-th order strands under conditions that promote the formation of only one complement, thereby generating first order derivative strands;

d) adding a polynucleotide tail to the first order derivative strands;

e) subjecting the first order derivative strands to a polymerase reaction in the presence of primers capable of hybridizing to the polynucleotide tail of the first order derivative strands under conditions that promote the formation of only one complement, thereby generating second order derivative strands,

wherein the primers of at least one of steps (c) and (e) comprise substantially unique primers, thereby generating tagged nucleic acid molecules.

Also provided is a composition of matter derived from genomic material comprising tagged nucleic acid molecules, said tagged nucleic acid molecules being produced by a process comprising:

a) obtaining segments of genomic material;

b) adding a polynucleotide tail to the ends of the segments of genomic material to generate zero-th order derivative strands,

c) subjecting the zero-th order derivative strands of step (b) to a ligation reaction in the presence of primers capable of hybridizing to the polynucleotide tail of the zero-th order strands added in step (b) under conditions that promote the ligation of a primer to the 5' ends of the zero-th order derivative strands,

d) subjecting the product of step (c) to a polymerase reaction in the presence of primers capable of hybridizing to the polynucleotide tail added in step (b) under conditions which promote the formation of only one complement, wherein the primers of step (d) have different nucleotide sequences than the primers of step (c), and wherein the polymerase has 3'-5' proofreading activity,

wherein the primers of at least one of steps (c) and (d) comprise substantially unique primers, thereby generating tagged nucleic acid molecules.

Also provided is a composition of matter derived from a mRNA transcripts comprising tagged nucleic acid molecules, said tagged nucleic acid molecules being produced by a process comprising:

a) obtaining cDNA from the mRNA transcripts;

b) optionally obtaining segments of the cDNA; and

c) tagging the cDNA or segments of cDNA with substantially unique tags to generate tagged nucleic acid molecules, wherein each tagged nucleic acid molecule comprises

i) one cDNA from step (a) or one segment of the cDNA from step (b) and

ii) a tag.

Also provided is a composition of matter derived from mRNA transcripts comprising tagged nucleic acid molecules, said tagged nucleic acid molecules being produced by a process comprising:

a) obtaining mRNA transcripts;

b) subjecting the mRNA transcripts to a polymerase reaction in the presence of primers capable of hybridizing to the polyA tail of the mRNA transcripts under conditions that promote the formation of only one complement, thereby generating first order derivative strands;

c) adding a polynucleotide tail to the first order derivative strands;

d) subjecting the first order derivative strands to a polymerase reaction in the presence of primers capable of hybridizing to the polynucleotide tail added in step (c) under conditions that promote the formation of only one complement, thereby generating second order derivative strands,

wherein the primers of at least one of steps (b) and d) comprise substantially unique primers, thereby generating tagged nucleic acid molecules.

In an embodiment of the above compositions, adding a polynucleotide tail comprises the use of a terminal transferase.

In an embodiment of the above compositions, each tagged nucleic acid molecule comprises one tag.

In an embodiment of the above compositions, each tagged nucleic acid molecule comprises more than one tag.

In an embodiment of the above compositions, segments of the genomic material are produced by restriction endonuclease digestion, mechanical shearing, heating, or sonication.

In an embodiment of the above compositions, segments of the cDNA are produced by restriction endonuclease digestion, mechanical shearing, heating, or sonication.

In an embodiment of the above compositions, the complexity of the tagged nucleic acid molecules is reduced by hybrid capture.

In an embodiment of the above compositions, each tagged nucleic acid molecules differs at more than one nucleotide.

In an embodiment of the above compositions, the tag sequences further comprise a sample tag.

Also provided is a composition of matter comprising at least two pools of tagged nucleic acid molecules as described herein.

In an embodiment of the above compositions, the tagged nucleic acid molecules are generated from a single cell.

In an embodiment, the single cell is from a needle aspirate of suspected cancer lesions.

In an embodiment, the single cell is from a core biopsy of suspected cancer legions.

In an embodiment of the above compositions, the tagged nucleic acid molecules are generated from two or more organisms.

Also provided is a kit for determining nucleic acid copy number information unaffected by amplification distortion comprising:

a) a terminal transferase, and

b) plurality of substantially unique primers,

wherein the substantially unique primers comprise substantially unique tags, and wherein the substantially unique primers are capable of hybridizing to the polynucleotide tail of a nucleic acid molecule which has a polynucleotide tail added by the terminal transferase.

In an embodiment, the kit further comprises a DNA polymerase having 3'-5' proofreading activity.

In an embodiment of the kits, the plurality of substantially unique primers comprises 10^n primers, wherein n is an integer from 2 to 9.

In an embodiment, the kits further comprise a second terminal transferase and a primer capable of hybridizing to the polynucleotide tail of a nucleic acid molecule which has a polynucleotide tail added by the second terminal transferase.

In an embodiment of the kits, the tags are six nucleotides long.

In an embodiment of the kits, the tags are 15 nucleotides long.

In an embodiment, the kits further comprise sample tags.

In an embodiment, the sample tags are 2 nucleotides long.

In an embodiment, the sample tags are 4 nucleotides long.

In an embodiment, the sample tags comprise a sample tag set.

In an embodiment of the above kits, the substantially unique tags comprise the sample tags.

Terms

For the purpose of this invention, different words and phrases are defined as follows:

As used herein, the term "adaptor" refers to an oligonucleotide or nucleic acid fragment or segment that can be ligated to a nucleic acid molecule of interest. For the purposes of this invention adaptors may, as options, comprise primer binding sites, recognition sites for endonucleases, common sequences, promoters, tag sequences, and sample tag sequences. Preferably, adaptors are positioned to be located on both sides (flanking) a particular nucleic acid molecule of interest. In accordance with the invention, adaptors may be added to nucleic acid molecules of interest by standard recombinant techniques (e.g. restriction digest and ligation). For example, adaptors may be added to a population of linear molecules, (e.g. a genomic DNA which has been cleaved or digested) to form a population of linear molecules containing adaptors at one and preferably both termini of all or a substantial portion of molecules. The adaptor may be entirely or substantially double stranded or entirely single stranded. A double stranded adaptor may comprise two oligonucleotides that are at least partially complementary. The adaptor may be phosphorylated or unphosphorylated on one or both strands. Adaptors may be used for DNA sequencing. Adaptors may also incorporate modified nucleotides that modify the properties of the adaptor sequence. For example, methylated cytosines may be substituted for cytosines. In an embodiment of this invention the

adaptors ligated to genomic DNA to enable cluster generation on the sequencer contain cytosines which were all methylated. This modification protects such adaptors from bisulfite conversion, and is taken into account in the downstream applications and analysis of this invention.

As used herein, the term "amplifying" refers to the process of synthesizing nucleic acid molecules that are complementary to one or both strands of a template nucleic acid. Amplifying a nucleic acid molecule typically includes denaturing the template nucleic acid, annealing primers to the template nucleic acid at a temperature that is below the melting temperatures of the primers, and enzymatically elongating from the primers to generate an amplification product. The denaturing, annealing and elongating steps each can be performed once. Generally, however, the denaturing, annealing and elongating steps are performed multiple times (e.g., polymerase chain reaction (PCR)) such that the amount of amplification product is increasing, often times exponentially, although exponential amplification is not required by the present methods. Amplification typically requires the presence of deoxyribonucleoside triphosphates, a DNA polymerase enzyme and an appropriate buffer and/or co-factors for optimal activity of the polymerase enzyme. The term "amplification product" refers to the nucleic acid sequences, which are produced from the amplifying process as defined herein.

As used herein, the term "amplification distortion" refers to non-uniform amplification of template nucleic acid molecules.

As used herein, the term "bisulfite-treatment" refers to the treatment of nucleic acid with a reagent used for the bisulfite conversion of cytosine to uracil. Examples of bisulfite conversion reagents include but are not limited to treatment with a bisulfite, a disulfite or a hydrogensulfite compound.

As used herein, the term "bisulfite-converted material" refers to a nucleic acid that has been contacted with bisulfite ion in an amount appropriate for bisulfite conversion protocols known in the art. Thus, the term "bisulfite-converted material" includes nucleic acids that have been contacted with, for example, magnesium bisulfite or sodium bisulfite, prior to treatment with base.

As used herein, the term "sequence capture" or "hybrid capture" refers to a process of hybridizing a "capture probe" to a nucleic acid having a sequence which is complementary to

the sequence of the capture probe. A capture probe may be immobilized to a substrate on a solid phase microarray, wherein "substrate" refers to short nucleic acid sequences which are known. Their location on the solid phase microarray may or may not be predetermined. The capture probe comprising a "sequence complementary to the substrate" may be immobilized to the solid phase microarray by hybridizing to its complementary "substrate sequence."

As used herein, the term "clamp" refers to a nucleotide sequence of CC, CG, GC or GG which may be located at the first two nucleotide positions at the 5' end of a primer, tag, or sample tag, or the last two nucleotide positions at the 3' end of a primer, tag, or sample tag.

As used herein, the term "comparable number of locations" means that a first genomic region has a number of locations on the genome which is within 25% of the number of locations of a second genomic region. In an embodiment, the first genomic region which has a comparable number of locations to the second genomic region has a number of locations which is within 20%, 15%, 10%, 5%, 1%, or exactly the same as the number of locations of the second genomic location.

As used herein, a subsequence of a tag associated sequence read is "corresponding to" a tag when the subsequence is identical to the nucleotide sequence of the tag.

As used herein, a subsequence of a tag associated sequence read is "corresponding to" a species of a nucleic acid molecule when the subsequence is substantially identical to or substantially complementary to at least about 10, 12, 14, 16, 18, 20, or more nucleotides of the sequence of the species of nucleic acid molecule. In an embodiment, the subsequence is identical to or fully complementary to at least about 10, 12, 14, 16, 18, 20, or more nucleotides of the sequence of the species of nucleic acid molecule.

The term "CpG site" refers to the CpG dinucleotide.

As used here, the term "CpG island" refers to a region of DNA with a high G+C content and a high frequency of CpG dinucleotides relative to the bulk genome, as further defined at the UCSC Genome Bioinformatics Site at genome.ucsc.edu/index.html?org=Human&db=hg19&hgsid=171216665, and in particular, by annotation in the USCS Genome Bioinformatics Site database (CpG Islands Track or Table: cpgIslandExt).

As used herein, the term "fully complementary" refers to the reverse complement of a nucleic acid sequence.

As used herein, the term "library" refers to a collection of nucleic acid molecules (circular or linear). In one preferred embodiment, a library is representative of all of the DNA content of an organism (such a library is referred to as a "genomic" library), or a set of nucleic acid molecules representative of all of the expressed genes (such a library is referred to as a cDNA library) in a cell, tissue, organ or organism. A library may also comprise random sequences made by de novo synthesis, mutagenesis of one or more sequences and the like. A library may be contained in one vector.

As used herein, the term "mapping" refers to identifying a location on a genome or cDNA library that has a sequence which is substantially identical to or substantially fully complementary to the subsequence of a tag associated sequence read corresponding to a species of nucleic acid molecule, and assigning the tag associated sequence read or the tagged nucleic acid molecule generating the tag associated sequence read to the location. The nucleic acid molecule may be, but is not limited to the following: a segment of genomic material, a cDNA, a mRNA, or a segment of a cDNA.

As used herein, the term "methylation" refers to the covalent attachment of a methyl group at the C5-position of the nucleotide base cytosine within the CpG dinucleotides of genomic region of interest. The term "methylation state" or refers to the presence or absence of 5-methyl-cytosine ("5-Me") at one or a plurality of CpG dinucleotides within a DNA sequence. A methylation site is a sequence of contiguous linked nucleotides that is recognized and methylated by a sequence specific methylase. A methylase is an enzyme that methylates (i. e., covalently attaches a methyl group) one or more nucleotides at a methylation site.

As used herein, the term "mRNA transcript" refers to the nucleic acid molecule resulting from transcription of DNA.

As used herein, the term "pool" means to combine one plurality of nucleic acid molecules with at least one other plurality of nucleic acid molecules. In an embodiment, tagged nucleic acid molecules are pooled. Pooling may be done after any step in which tagged nucleic acid molecules are generated. In an embodiment, pooling is done prior to PCR and/or prior to sequencing. In an embodiment, tagged nucleic acid molecules generated from a single cell

are pooled with tagged nucleic acid molecules generated from a different single cell.

As used herein, the term "probe array" refers to an array of N different nucleic acid molecules deposited on a reaction substrate which serve to interrogate mixtures of target molecules or multiple sites on a single target molecule administered to the surface of the array.

As used herein, the term "read" or "sequence read" refers to the nucleotide or base sequence information of a nucleic acid that has been generated by any sequencing method. A read therefore corresponds to the sequence information obtained from one strand of a nucleic acid fragment. For example, a DNA fragment where sequence has been generated from one strand in a single reaction will result in a single read. However, multiple reads for the same DNA strand can be generated where multiple copies of that DNA fragment exist in a sequencing project or where the strand has been sequenced multiple times. A read therefore corresponds to the purine or pyrimidine base calls or sequence determinations of a particular sequencing reaction.

As used herein, the term "reference genome" refers to a genome of the same species as that being analyzed for which genome the sequence information is known.

As used herein, the term "region of the genome" refers to a continuous genomic sequence comprising multiple discrete locations.

As used herein, the term "representation of DNA" or "representation of a genome" refers to a sampling of the DNA or genome produced by a restriction endonuclease digestion of genomic or other DNA, followed by linkage of adaptors and then amplification with primers complementary to the adaptors.

As used herein, the term "sample tag" refers to a nucleic acid having a sequence no greater than 1000 nucleotides and no less than two that may be covalently attached to each member of a plurality of tagged nucleic acid molecules or tagged reagent molecules. A "sample tag" may comprise part of a "tag."

As used herein, the term "sample tag set" refers to a plurality of unique sample tags of the same length, wherein the nucleotide sequence of each unique sample tag differs from the nucleotide sequence of any other unique sample tag in the sample tag set at two or more

corresponding positions in the sequence. A unique sample tag selected from the sample tag set may be covalently attached to each member of a unique sample consisting of a plurality of tagged nucleic acid molecules in order to identify a tagged nucleic acid molecule as a member of that unique sample.

As used herein, the term "segments of cDNA," refers to the nucleic acid molecules resulting from fragmentation of cDNA.

As used herein, the term "segments of genomic material" refers to the nucleic acid molecules resulting from fragmentation of genomic DNA.

A nucleic acid molecule which contains a sequence of nucleotides identical to a segment of a cDNA or a segment of genomic material comprises a "segment of cDNA" or a "segment of genomic material," respectively.

As used herein, the "sequence complexity" or "complexity" with regards to a population of polynucleotides refers to the number of different species of polynucleotides present in the population.

As used herein, "substantially identical" or "substantially fully complementary" sequences have at least about 80% sequence identity or complementarity, respectively, to a nucleotide sequence. Substantially identical sequences or substantially fully complementary sequences may have at least about 85%, 90%, 95% or 100% sequence identity or complementarity, respectively.

As used herein, the term "substantially unique primers" refers to a plurality of primers, wherein each primer comprises a tag, and wherein at least 50% of the tags of the plurality of primers are unique. Preferably, the tags are at least 60%, 70%, 80%, 90%, or 100% unique tags.

As used herein, the term "substantially unique tags" refers to tags in a plurality of tags, wherein at least 50% of the tags of the plurality are unique to the plurality of tags. Preferably, substantially unique tags will be at least 60%, 70%, 80%, 90%, or 100% unique tags.

As used herein, the term "tag" refers to a nucleic acid having a sequence no greater than 1000 nucleotides and no less than two that may be covalently attached to a nucleic acid molecule or reagent molecule. A tag may comprise a part of an adaptor or a primer.

As used herein, a "tagged nucleic acid molecule" refers to a nucleic acid molecule which is covalently attached to a "tag."

Where a range of values is provided, it is understood that each intervening value, to the tenth of the unit of the lower limit unless the context clearly dictates otherwise, between the upper and lower limit of that range, and any other stated or intervening value in that stated range, is encompassed within the invention. The upper and lower limits of these smaller ranges may independently be included in the smaller ranges, and are also encompassed within the invention, subject to any specifically excluded limit in the stated range. Where the stated range includes one or both of the limits, ranges excluding either or both of those included limits are also included in the invention.

All combinations of the embodiments of the invention described herein are within the scope of the invention unless clearly indicated otherwise.

Single Cell Copy Number Profiling by 'Sequence Counting' and Sample Tagging.

High throughput sequencers are in effect single molecule sequencers. Because the human genome is sequenced and assembled, a single molecule sequencer is also a single molecule identifier. Therefore, sequencing is also counting, as illustrated in Figure 4 which illustrates how low coverage sequencing can yield an estimate of copy number which is comparable to array hybridization. The distinction for this is that 'sequence counting' can be done on single cells, and, using the multiplex system disclosed herein, relatively inexpensively, in effect converting the high throughput sequencer into a detector of single cells with abnormal genomes. To the sequencing technology two additional features were added: limiting dilution and sample tagging. For single-cell sequencing, the single cells or the nuclei of the single cells are first deposited into separate wells or are kept physically separated by other means, i.e. a gel. The single cells or the nuclei of cells can be deposited in a well by a cell sorter. Alternatively a collection of cells or nuclei can be diluted into a fluid that is then made semi-solid. A unique sample tag is prepared for each well for sequencing, with a separate sample tag per well. After this step, the wells are pooled and sequenced. The sample tag is used to deconvolute the information, allowing a copy number profile to be made for each individual cell.

Protocol: Isolate single tumor cells or single nuclei of the tumor cell, tag and optionally sample tag the DNA of each cell or nuclei, and amplify the tagged DNA to microgram

quantities to generate tagged libraries for Solexa sequencing. The isolation of single tumor cells will occur by Fluoresent Activated Cell Sorting by total genomic DNA content, to deposit single nuclei into individual wells.

Alternatively, tumor cells can be isolated by FACS sorting with fluorescent antibodies against breast epithelium markers, including cytokeratin 8, 18 or 19. Cells are lysed *in situ* and the genome is fragmented by heating. Libraries are constructed from the fragmented genomic DNA. Preliminary experiments were conducted using high-resolution ROMA microarray to quantify whole-genome copy number on DNA amplified using WGA from single SK-BR-3 cells (Figure 5). These experiments demonstrate the feasibility and efficiency of amplifying 6 pg of DNA into microgram quantities to determine copy number by microarray methods, which suggests that the DNA is also suitable for determining copy number by sequencing.

Since sequencing whole tumor genomes from single cells is costly and time consuming, two methods are proposed to stratify the genome and sequence for copy number: Representation Sequencing and Bin Sequencing.

Representation Sequencing

Representation sequencing involves restriction digesting genomic DNA into fragments and sequencing only the stratified population of restriction fragments to determine copy number. This method has the advantage of generating precise restriction fragments that can be reliably prepared as libraries from different individual tumor cells. In this protocol genomic DNA is amplified from a single tumor cell, digested with the DpnII, a second digestion is preformed with a AluI enzyme to remove highly repetitive "Alu SINE elements," sequencing adaptors (Illumina GA, Solexa) with a unique 4 nucleotide sample tag are ligated and the ligated fragments are PCR amplified to generate a sequencing library. *In silico* modeling predicts that the double digest will increase the fraction of unique sequencing reads to 75% from the 55% normally seen in unselected human genome sequencing. The sample tagged DNA libraries are then pooled, and applied to the Solexa flow-cells for cluster amplification and single read sequencing. Single-end reads are used because they are faster and cheaper than paired-end reads and provide all of the information necessary for 'sequence counting'. Mapping algorithms then align each sequence read with a high quality score (with a read length in excess of 35bp) and an unambiguous sample tag to a specific location on the DpnII restriction map of the human genome. The result is a count of reads at places in the genome,

similar to the process demonstrated in Figure 4, except that the data is segregated by sample tag, so that each well receives a copy number profile.

Bin Sequencing

It is clear that sequencing of random size-selected fragment from the genome is a viable alternative. In this method, the DNA from single tumor cells is amplified to generate fragments from 200-2000bp. Blunt-end cloning is then used to ligate adaptors to these fragments, followed by PCR amplification to generate libraries (standard Solexa protocol for single read genomic DNA). However, in addition to this protocol adaptor sequences with 4 nucleotide sample tags are ligated to generate unique libraries from each tumor cell genome. The sample tagged DNA libraries are then pooled, and applied to the Solexa flow-cells for cluster amplification and single read sequencing. The 35bp sequences are then mapped to specific regions of the human genome and deconvoluted using the unique sample tag to determine which sequence reads originated from which tumor cells. The crucial step in this procedure that is different from representation methods is to artificially divide the human genome into "bins" of several kilobases. Break analysis in sequence count data has another degree of freedom compared to what is practiced for array data, namely that the boundaries of the bins can be adjusted so as to minimize variance. Each bin contains a number of random genomic sequence reads from which a mean value is computed to determine genomic copy number at low sequence coverage.

The particular use of the WGA method does not depend on absolutely uniform amplification across the genome. Useful data can be obtained even if only half of the potential fragments from a single cell were amplified in a given reaction. Additionally, redundancy of data gained from segmentation benefits in identifying breakpoints.

Representation Sequencing vs. Bin Sequencing

A comparison of the two methods reveals the number of samples that can be profiled per sequencing run. Of the 3,149,324 Dpn II fragments in size range 200-800 (size optimal for sequencing) there are 618,629 not cut with AluI and thus available to be amplified. Approximately 75% of the fragment ends, or about 450,000, yield unique mappings and thus provide locations for counting. The overall process is presented as a flowchart in Figure 7.

Assuming 15 million quality reads per Solexa sequencing lane, random sequencing yields about 55% or 8.25 million unique reads per lane. The DpnII/AluI representation should

increase the number of unique reads to 11.25 million. Modeling a single flow cell lane with 100kb bins distributed across the genome, the projected results are as follows:

Table 1

# samples	avg reads/bin/sample (random)	avg.reads/bin/sample (Dpn/Alu)
1 sample per lane	267	363
2 samples per lane	133	181
12 samples per lane	22	30

The number of samples (individual cells) to be sequenced per lane, is between 4 and ten. With 8 lanes on a flowcell, 30-80 samples could be profiled at the level of Figure 4 relatively inexpensively.

Finally, the whole genomic copy number (generated by sequencing) will be determined for multiple, individual tumor cells in a single solid tumor. The data will be segmented into sections that minimize variance of copy number, subject to Kolmogorov-Smirnov significance tests, to generate individual tumor copy number profiles. As stated above, an additional degree of freedom, which our algorithms do not yet incorporate, is selecting the "boundary" of the bins. The single cell tumor genome profiles will be compared to study tumor heterogeneity in single solid tumors which may elucidate the progression of genomic events during tumorigenesis.

Varietal counting of nucleic acids

Measuring the absolute or relative concentration of nucleic acid sequences in a sample often suffers from severe stochastic distortions due to differential amplifications or losses during processing and manipulation of the nucleic acids. There is never more information available than the number of molecules of the starting material in the sample itself. The method described herein is designed to capture the information at the earliest stage, in a form that is then resistant to distortion by further amplification.

Input nucleic acid molecules (0-th order derivatives), or first round copies of them (1-st order derivatives), present in the sample, can be randomly tagged using a large choice of distinct tags. Every tagged molecule becomes essentially unique, the combination of the information in the tag and the information in the nucleic acid, which can later be read by sequencing. The combination of a tag with a 0-th order, 1-st order, or subsequent N-th order nucleic acid molecule is called a "tagged nucleic acid molecule." After amplification or selection or other processing, the number of different tags is counted.

Within a pool of tagged nucleic acid molecules, each tagged nucleic acid molecule is likely unique in the pool when a sufficiently large number of distinct tags is used. Then in subsequent amplifications these unique tagged nucleic acid molecules are amplified to facilitate detection, but no new species of tagged nucleic acid molecules are created. Thus, distorted counts are not created provided that the number of different tagged nucleic acid molecules is counted.

In the disclosed method, each sequence read generated from a tagged nucleic acid molecule comprises two parts, or subsequences. The first part corresponds to the tag, which identifies the tag (and the sample tag if samples are pooled). The second part corresponds to a nucleic acid in the sample which is mapped to a location in the genome (or to a transcript if RNA molecules are being counted). After deconvoluting by sample tag, the number of different tagged nucleic acid molecules mapped to each location is counted. There are two separate methods of counting, and each can be combined for greater accuracy and determination of confidence.

The first counting method is called "maximum tagged nucleic acid molecule number." The number of different tagged nucleic acid molecules at a given location cannot exceed the true (absolute) copy number. Therefore, for a given region comprising multiple locations, the maximum copy number of the region is not less than the maximum number of different tagged nucleic acid molecules mapped to any location in that region. The maximum number

of different tagged nucleic acid molecules centered about a moving window of a fixed number of locations can be taken as the measurement of true copy number of the window. This will provide at worst an underestimate of the maximum true copy number for that window, and never an overestimate. The set of maximum number of different tagged nucleic acid molecules is one type of copy number profile.

The counting method described above is most accurate when the processing efficiency is excellent and the number of molecules mapped to each location is less than the number of tags. Under such conditions virtually each molecule in the sample is tagged and counted, and the true count can be derived. But there is a second method.

The second counting method is called "total different tagged nucleic acid molecules," and is useful, for example, when tagging efficiency is low. The total number of different tagged nucleic acid molecules within a region of a fixed number of locations will be, within a statistical counting error, a monotonic function of true copy number, regardless of the efficiency of tagging. Low efficiency of tagging merely increases the statistical error. The relative copy number of a region can be estimated by comparing the total number of different tagged nucleic acid molecules mapped to the region to, for example, the mean value of tagged nucleic acid molecules mapped to other regions of the genome which have a comparable number of locations.

If the number of molecules at a location on a genome or cDNA library exceeds the number of tags, and tagging is too efficient, copy number information can not be recovered. Rather, the total number of tags at each location would merely be measured. There are at least three ways to remedy this condition: 1) diminish the efficiency of the first step of the process, for example by reducing the reaction time; 2) reduce the amount of sample; and 3) increase the total number of tags. For example, if the length of the tag were N nucleotides, there are 4^N possible tags. For N = 15, there are approximately 10^9 tags, easily in excess of the number of molecules per location.

The collection of reads lend themselves to modeling the power of inference. For each location, there exists the number of tagged nucleic acid molecules observed, and for each species of tagged nucleic acid molecule, the number of times it was observed. With a good estimate of the input, the number of genomes, critical parameters can be accurately deduced because the vast majority of the copy numbers are two per genome. The limiting case of single cell analysis lends itself to the most accurate modeling, with the proportion of null

reads enabling the inference of the probability, theta, that a single molecule at a location is detected. Theta is a function of the efficiency of tagging, the efficiency of amplification and processing, and the read depth.

Using the language invented of the WGA reaction, in the specific protocol described here, all first order derivative DNA molecules (i.e., first copies) are tagged, with no new tags introduced when subsequent order derivatives are created. This enables the "counting" of original molecules, alleles in the case of genome analysis. This method is referred to herein as "varietal counting." As long as sequence depth is sufficient, subsequent amplification or methods of trapping that enrich for regions of interest, or the combination, cannot distort the count. Pooling from multiple cells becomes a natural extension of this method, which improves efficiency, increases throughput and decreases costs.

Tags

The tags are nucleotide sequences with constant portions and variable portions. The length of the constant portion and the length of the variable portion may be the same or different. The length of the tags within a plurality of tags is typically, but not necessarily, the same for all tags. In an embodiment, the total length of a tag is less than 100 nucleotides.

The constant portions are used to aid in the tagging of nucleic acid molecules. For example, the constant portion may contain a nucleotide sequence which allows hybridization to a poly-N tail added by a terminal transferase. The constant portion may also be used to aid in manipulation of the tagged nucleic acid molecules. For example, the constant portion may contain a sequence which is a primer binding site useful for PCR amplification. The constant portion may also be used to identify a sample if there are multiple samples being processed in parallel, i.e., the constant portion may contain a sample tag.

The sequence of individual tags does not need to be known prior to generating tagged nucleic acid molecules. A plurality of tags may have variable portions whose sequences are elucidated for the first time during sequencing. The variable portion of a tag may be a sequence of nucleotides where each nucleotide is individually one of A, T, C, or G, or one of any two of A, T, C, and G, or one of any three of A, T, C, and G, i.e. the variable portion of a tag may consist of only two or three species of nucleotides. Tags may be designed such that the variable portion has a random sequence. Tags may be designed such that the variable portions are constructed from a set of dinucleotides or trinucleotides.

Tags may be associated with nucleic acid molecules by using, singly or in combination, ligation, nick-ligation, terminal transferase, hybridization, priming, and the like. For example, nucleic acid molecules may be cut with a restriction endonuclease(s), treated with a terminal transferase to add a poly-N tail, hybridized to a tag molecule, and then nick ligated to form a tagged nucleic acid molecule. Specific methods of generating tagged nucleic acid molecules are discussed in detail below.

Tagging with the aid of a terminal transferase

Figure 8 illustrates one embodiment of this method, where the N_{10} designates an oligonucleotide having a tag component and a sample tag component. The first six nucleotides (4096 possible sequences), randomly generated, provide the tag component, and the last four nucleotides (256 choices) are chosen to encode the micro-well, and, hence, a sample tag. The lengths of the tag and the sample tag can be changed to suit needs, and are not limited to a total of ten nucleotides. The lengths are chosen for illustration. The sample tag allows pooling, with deconvolution occurring afterwards. The combination of a tag with a nucleic acid molecule from the input sample provides an essentially unique tagged nucleic acid molecule for the first derivative strand, that can only with very low probability be replicated by chance from another like molecule (a chance of roughly 1/N where N is the number of available tags). In the end, multiple copies of sequence-identical molecules in the input sample are each individually counted. In applications to mRNA counting, the length of the tag component can be increased to accommodate possibly thousands of transcripts from the same gene.

Although there are multiple embodiments of the basic method besides the one illustrated in Figure 8, the embodiment illustrated in Figure 8 has several useful design features. First, it avoids the use of ligases (an alternate method for associating a tag with a nucleic acid molecule), which has been found to be an inefficient reaction. Instead, this embodiment uses terminal transferase (TTase), which is a robust enzyme from which high efficiency is expected. Second, once the tag and sample tag are incorporated into the first derivative DNA strand, micro-wells can be pooled and the subsequent steps carried out more efficiently in larger volumes with larger numbers of molecules, possibly with the addition of carrier nucleic acids. Because of a clean-up step, removing free adaptor, no more unique first order derivative strands with tags can be created, and second order strands can be arithmetically expanded if found to be useful. Third, the method extends immediately from DNA to

mRNA. Rather than cleaving and adding a polynucleotide tail with terminal transferase, the naturally occurring polyA tail of mRNAs can be utilized. Fourth, it is easy to see how to combine analysis of both RNA and DNA from the same cell with this schema. Instead of tailing DNA with A, it can be tailed with, for example, C, thus distinguishing the DNA from the RNA. DNA and RNA derivatives can either be read together in the same sequencing runs, or separately by amplification using their differing PCR primer adaptors. Fifth, by using a longer sample tag, sequences belonging to a particular cell can be preferentially amplified for deeper study when warranted.

This method can be applied to assessing DNA methylation. After the first order derivative strands are made, the duplex DNA can be pooled from multiple samples. These molecules will be hemi-methylated, and still retain epigenetic marks. Therefore they can then be separated into methylated and unmethylated samples either by cleavage with methylation sensitive restriction enzymes, or by partitioning with antibodies or methyl-C binding proteins directed to methylated (or hydroxyl methylated) cytosine. Error rates will be measured based on expectations from regions of known methylation state.

Tagging by ligation

In another embodiment of the invention, adaptors with tag(s) can be ligated to sample DNA or RNA that has been converted to DNA. In the second step these molecules can be pooled (or not) and amplified for sequencing. In the third step the numbers of each different tagged nucleic acid molecule of each locus (or mappable sequence read) are compiled, giving statistics from which copy number can be inferred.

In another embodiment, to measure RNA, RNA is first converted to DNA using methods that produce a single double stranded molecule for each RNA in the sample. (For example, prime with oligo dT and reverse transcriptase, followed with random priming and Klenow polymerase.) The procedures for analyzing DNA can then be followed.

In another embodiment, the sample DNA can be sheared and the ends prepared for ligation. However, there are three major advantages to cleaving with one or more restriction endonucleases (RE). The first is that the expected number of tagged nucleic acid molecules at each location can be precisely predicted from the starting number of cells in the sample, or conversely the number of cells in the sample inferred from the expected number of tagged nucleic acid molecules per location. The second is that the expected complexity of the

product is reduced, enabling copy number determination from fewer sequence reads. The third are certain advantages in the flexibility and functionality that can be built into the tags.

In an embodiment, sample nucleic acid is cleaved with a specific restriction enzyme and sample molecules are ligated to tagged (and sample tagged if appropriate) oligonucleotide adaptors which contain sequences that allow PCR amplification. There are various designs for these adaptors, and also alternatives to PCR (such as WGA), based on the functionalities built into the tags. Some of these additional functionalities are discussed below.

The ligated molecules now are essentially unique, the ligated molecules each being the combination of a substantially unique tag and DNA sequence to which they were ligated. Further amplification may distort the yields of molecules, but will not create new tags (although some may be lost by statistical processes which are readily modeled during analysis).

Through sequence error during processing, the tag may very occasionally mutate, creating the appearance of a new tagged nucleic acid molecule. But if the set of tags are sufficiently large, the tagged nucleic acid molecules will generally differ at more than one nucleotide, and so tagged nucleic acid molecules that differ by a single nucleotide can (optionally) be ignored when counting.

For single cells, individual cells are loaded at most one per microwell. The nucleic acid is then prepared in wells, and the molecules tagged. Once tagged they can be pooled for further processing, which greatly improves the uniformity and efficiency of DNA amplification methods.

In yet another embodiment, DNA is cut, and adaptors are ligated to both ends of each fragment, enabling them to be amplifiable by PCR. To avoid adaptor-adaptor self ligation, double stranded adaptors without 5' phosphorylation can be used, so an adaptor is ligated on one strand only at each end of the cleaved sample DNA. Then the molecules are pooled, and treated as for representations: the shorter single strand adaptor is melted off at elevated temperature but well below the Tm for the sample duplex, the ends of the double stranded DNA filled in with Klenow polymerase, unligated adaptors removed, and then the sample is PCR amplified.

In the preceding embodiment, adaptors are used not only for tagging but also for amplification. Adaptors are for identifying each molecule individually. They can also be used

for amplification, as described above, or for other uses (see below). Because in the above embodiment adaptor ligation at both ends of sample fragments is required for amplification, the yield may be diminished. To improve yield, the method can work with molecules that have an adaptor at one end only.

In one embodiment, the reactions are performed as before, but tags do not contain sequences enabling PCR. Prior to PCR, but after an optional pooling, fresh adaptors (different composition and without tags) are ligated so that molecules with only a single tag adaptor can be amplified and eventually counted. If the original tag has a selection sequence, molecules with the tag adaptor before sequencing can be optionally enriched by hybrid selection.

In another embodiment, initial adaptors are designed with a phosphorylated 5' end and a 3' dideoxy end, so that the initial ligation occurs to the opposite strand as described in the embodiment above. (Instead of using an RE with a 5' overhang, one which leaves a 3' overhang may be preferred.) After removal of unligated adaptors, there follows several rounds of arithmetic amplification, followed by tailing molecules with dT using terminal transferase, and amplification of the processed sample DNA by PCR.

In another embodiment, DNA from sample is labeled as before, pooled if required, and unligated adaptors removed from the reaction. The reaction is subject to whole genome amplification (WGA), and the product of WGA is used to make sequencing libraries. For this embodiment, the initial adaptors may also contain: (A) cleavage sites for ease of making sequencing libraries; and (B) a sequence large enough for optional enrichment by hybrid selection prior to making libraries for sequencing.

In the embodiments just described, it is assumed that the tag is double stranded. The production of a great variety of double stranded molecules is not completely straightforward. In one embodiment, single strand DNA is synthesized as type I and type II. Type II has three sets of functionalities A, B, C, and D. Functionalities B contain the tag sequence, C the adhesive ends with a RE site, and D the buffer. The B will contain some length of runs with random sequence, for example 10 nucleotides with a total complexity of 4^{10}, or about 10^6 possible sequences. Type I oligo will be complementary to type II at functionality A, the two joined and annealed, and then extended with polymerase to create a double stranded molecule. After cleavage with the RE site, the resulting double stranded molecule is the tag used for the subsequent tagging of the sample.

Additional methods for producing single and double stranded oligonucleotides suitable for use as tags are known in the art, for example, in U.S. Patent Nos. 5,639,603 and 7,622,281, U.S. Patent Application Publication No. 2006/0073506 and Parameswaran et al. (2007) (7,5,8,4).

Nick-ligation

In this embodiment of the invention, 3'OH ends of nucleic acid molecules, e.g., segments of genomic material or mRNA transcripts, are first extended with terminal transferase, adding a poly-N tail. Primer adaptors are then annealed to the poly-N tails and the resulting duplex with a nick is sealed with DNA ligase. This series of reactions is much more efficient than a single step double stranded ligation. The resulting product is amplified by PCR. This method has been successfully used with as little as 1ng of DNA. Tagged nucleic acid molecules generated using the nick-ligation method can be generated from a single cell DNA, and then pooled prior to PCR. Pooling from 100 cells yields the DNA concentration needed for successful amplification. Sample tags can be added in the nick ligation step, and this is readily achieved (see Figure 9).

Technology for inexpensive assay of regional markers

The above description is of methodology designed to measure copy number and mutational loads throughout much of the genome in single cells. But the application of discoveries, once made, of predictive markers for outcome or therapeutic response will not require whole genome information. Our present outcome markers are in fact copy number markers at very specific loci. Essentially the same methodology as disclosed hereinabove can be employed to make these assessments very affordable. Consider, for example, the case where it is desirable to assess a large study for the predicative power of a set of N markers, and these markers are copy number markers that can be assessed by varietal counting. In principle, sample coding, pooling, and parallel capture and sequencing can be used to assay a large number of samples in a single sequencing lane. The major difference is that we are not restricted to minute quantities of DNA, so WGA is not needed, and assays will be more robust. We have shown in our autism work that excellent copy number data can be obtained from read depth after capture. Thus we can use direct library preparation with just sample tags. The target area for capture is likely to be different and smaller than the exome. Thus we can pool more samples per sequencing lane, and the sequencing costs for assaying hundreds of markers in hundreds of samples can be reduced. Of course, there are costs associated with the development and

testing of the capture reagents, and the cost of the capture itself, and these are not negligible, but if a test were to come into widespread clinical use, it would be inexpensive, and this is one of our major objectives. There is a clear alternative, namely creating a bank of region specific PCR primers with tags and sample tags. The tags would be necessary because random unique ends created by shearing would not be available for counting based methods. This alternative plan could be less expensive and more robust than capture (see Figure 10).

Single cell RNA measurements

Several methods have now been published whereby researchers have profiled the RNA content of single cells by sequencing (Tang et al., 2009). This is potentially a very powerful method for analysis of tumor cells. A varietal counting method could be very useful for RNA profiling from single cells, and for the same reasons that it is useful with DNA. When variety is not introduced by nucleic acid processing varietal counting provides a way to avoid the distortions created by PCR. RNA and DNA can be analyzed from the same cells. The DNA analysis enables us to cluster cells into stromal and tumor subpopulations, and this in turn facilitates the interpretation of the RNA profiling. As an example of utility, suppose one is interested in searching thousands of circulating cells, or cells from a biopsy, or from the margins of a tumor, for rare tumor cells. DNA copy number can facilitate its identification, and the RNA from that cell can then be reliably used to infer additional properties, such as the tissue of origin. This combination is especially powerful for the early detection of the new incidence or recurrence of cancer.

Examples

Methylation sequencing

Currently, high throughput sequencing requires physical purification of sub-genomic regions of interest in order to obtain high coverage of those regions. The method of sequence capture has been applied to examine methylation in large sets of CpG islands at the sequence level. Methylation sequencing is based on the conversion of C residues to T using bisulfite reagents, with methylated C's detected as those protected from C to T conversion. This means either that sequence capture must take place before bisulfite treatment, which requires more starting DNA than is available for patient specimens, or that sequence capture must be designed to accommodate the C to T conversion. The bisulfite based sequencing method used herein adopts the latter and so is suitable for high throughput single molecule

sequencing by either the Illumina or Roche 454 methods, using as little as 100 ng of input DNA. The sequencing method depends on hybrid capture of bisulfite treated DNA followed by sequencing through the Illumina Solexa sequencer and mapping back to the genome using an algorithm developed Dr. Andrew Smith (19). It was observed that CpG islands can exhibit multiple different levels of methylation, from imprinted (allelic) to randomly partially methylated, to completely methylated or unmethylated. In particular, one observation that is important for this invention is that many islands switch from completely unmethylated in normal tissues to virtually completely methylated in matched tumor tissue ("DNA-methylation switch domains"). Figure 3 shows three examples of the simplest cases in normal cells versus the breast cancer cell line MDA-MB-231: stably unmethylated (A and B); the ALX gene, switched from unmethylated to completely methylated (C and D); and a case, SSTR4, where only a segment of the island is switched (E and F). This complete switch makes it possible to detect rare tumor cell DNA in bodily fluids or other cell populations using nested PCR of bisulfite treated DNA and opens up the possibility of a creating a method for the early detection of recurrence in the form of circulating metastatic cells.

Comparative genomic hybridization (CGH) by high throughput sequencing

Until now the most economical method for high resolution genomic analysis has employed some form of microarray. The rapid rise of 'next generation' DNA sequencing is changing that landscape. Both the Illumina (Solexa) and Roche 454 instruments can provide millions of individual sequence reads randomly distributed across any input DNA (Craig et al., 2008). The depth of coverage of these technologies depends on the total length of the input DNA, from whole genome (large breadth, lower depth) to narrowly focused regions achieved by hybrid capture of specific sequences (narrow breadth, higher depth) (Albert et al., 2007; Hodges et al., 2007; Okou et al., 2007). Converting sequencing reads to copy number analysis is a matter of mapping reads and then put them into 'bins' of a certain size, depending on the purpose, counting up the numbers in the bins and performing appropriate statistical analysis to define copy number change breakpoints.

Sequence counting was applied to the degraded formalin-fixed paraffin-embedded (FFPE) DNA that was available from retrospective clinical trials with very encouraging results. Figure 4 displays a comparison of microarray and sequence counting results in a region around Her2 on chromosome 17 for a 10 year old paraffin-embedded clinical trial sample JZ33. The gray and orange points are (respectively) the raw ratio data for 244K Agilent

microarray and one lane of Solexa sequencing reads divided into 30K bins. Note that the sequencing (orange) track shows significantly less noise. The green and blue lines are the result of segmenting each data type. The segments are generally concordant, but the sequence counting method reveals a somewhat more detailed pattern. Another feature of note from Figure 4 is the robustness of our segmenting program, irrespective of the variation of noise in the two sets of raw data.

The efficacy of sequence counting on FFPE DNA and the rapidly decreasing cost of DNA sequencing in general provides an opportunity to convert our previous microarray methods to a highly flexible and scalable platform for breast cancer genome diagnostics. Both the breadth of DNA to be sequenced and the size and location (and therefore resolution) of individual bins can be easily adjusted to fit the particular loci to be assayed.

Single cell microarray and sequencing

A representation of DNA amplified from single cells was used for genomic profiling by 'sequence counting'. As a preliminary step in developing that technology, DNA was amplified from batches of 1, 10 and 100 cells of the cell line SKBR3 using the Whole Genome Amplification (WGA) Kit from Sigma-Aldrich and compared these amplified representations to equivalent microarrays on the same format done with a standard SKBR3 DNA prep. Individual cells were picked from adherent culture dishes using a cell transfer micropipette. After random WGA1 amplification the DNA preps were digested with DpnII and AluI, adaptors added and only DpnII fragments without AluI sites were amplified by PCR. These representations were labeled and applied to ROMA custom microarrays with 390,000 probes designed to detect DpnII fragments of appropriate size across the genome. The idea was to test how reproducible this two step process would be when only 2-6 chromosomes are present in each single cell preparation (SKBR3 is roughly triploid).

The results are shown in Figure 5, presenting a comparison of the segmented profiles of a batch preparation of SKBR3 DNA (Figure 5A) against our normal male reference cell line (SKN-1) and three independent single cell DNAs (Figure 5 B,C,D). A visual inspection of these profiles makes it clear that after segmentation, the results for single cells are virtually identical to a standard DNA, leading us to have confidence that sequencing DNA from a representation will yield useful results. Although the raw ratio data is not presented here, we note that the standard deviation of the ratio data is much higher for the three single cell experiments (0.706, 0.725, 0.676) than for a batch DNA preparation (0.313). The extra noise

does not adversely affect segmentation, although in certain regions very fine detail may be lost (red arrow in Figure 5A). This is comparable to other WGA1 experiments such as the one depicted in gray in Figure 4. In this light it is also worth noting that our particular use of the WGA1 method does not depend on absolutely uniform amplification across the genome. We benefit from the redundancy of data that we gain from segmentation to identify breakpoints. Even if only half of the potential fragments from a single cell were amplified in a given reaction, we would still obtain useful data.

Sequencing across chromosome breakpoints

One of the methods for the detection of rare tumor cells, in lymph nodes, blood or other bodily fluids is the amplification or sequencing of unique breakpoints resulting from deletions and/or translocations that occur during breast tumor development. Although events such as deletions and duplications are apparent from array CGH, the actual structure underlying the copy number changes is not.

The DNA sequences surrounding 22 breakpoints from two breast cancer cell lines were enriched by hybrid capture on custom microarrays as described by Hodges et al (20). Following hybrid capture, the enriched DNA was amplified and prepared for 'paired end' sequencing on the Illumina GA2 instrument. Paired end sequencing yields approximately 36 bp of sequence from each end of a single DNA molecule (after preparation for sequencing, each molecule is approximately 200 bp in length), and those sequences are then mapped back to the consensus genome. For a molecule from a normal, non-rearranged region the two matched reads should map approximately 130 bp apart. In contrast, at the edges of a deletion, the two reads from a single molecule would be separated on the map by the width of the deletion, or in the case of a translocation, would map to another chromosome. These are referred to as 'bridging fragments' because they connect regions not normally adjacent to one another. Examples of two such bridging fragments are shown at the top of Figure 6, along with the original CGH pattern, and graphs of the distance between paired ends (blue) and the total number of reads (red) across the capture region (X axis). By mapping the exact location of a breakpoint, highly sensitive and specific probes that will distinguish tumor cells from large numbers of normal cell can be prepared. Of 22 regions displaying deletions captured from cell lines T47D and MDA-MB-436, 15 behaved as simple deletions from which unique 'bridging fragments' could be identified as in Figure 6. Two additional sites 'jumped' to different chromosomes indicating translocations. The other 5 sites involved repeated

sequences on one of the paired for which the identification of the exact breakpoint was not possible and which would not be suitable for creating unique probes.

Copy number profiling using varietal tags on restriction enzyme fragments.

As one illustration of varietal counting, copy number information was obtained using a "total different tagged nucleic acid molecules method.", DNA fragments were first associated with tags by using terminal transferase and nick-ligation (Figure 9). Genome DNA was obtained from the cell line SKBR3. DNA (2μg) was digested by NlaIII, resulting DNA fragments with 4bp, "CATG", 3' overhangs. This 3'OH end was extended with terminal transferase by adding a poly-A tail. One of primer adaptors was designed to have tags, 12 Ts and CATG at the 3' end, so it could be annealed to the 3' ends of the DNA fragments with a nick site (Figure 9, Product 3). This nick site was ligated by DNA ligase, forming duplex DNA fragments with tags. Another primer adaptor was then added to these duplex DNA fragments by one cycle of PCR (Figure 9, Step 5), allowing the resulting products of this reaction to be further amplified by PCR, followed by sequencing (Illumina/Solexa). Then, by mapping the subsequence of each tag associated sequence read to a location on the genome, the number of DNA fragments (assigned to the same genomic location) with different tags was counted.

We have previously obtained copy number profiles of SKBR3 genomic DNA from a million cell library, by sequencing and sequence counting. Figure 11 compares the use of varietal counting (a "total different tagged nucleic acid molecules method") with 'sequence counting' for analyzing copy number data, genome-wide (Figure 11A) and partial-genome (Figure 11B). The essentially equivalent result obtained using varietal counting illustrates that by adding varietal tags to DNA fragments before PCR one can use varietal counting as described herein to reliably obtain copy number information unaffected by amplification distortion.

Data analysis for the procedure described above employed a modification of the 'bin-counting' method because the NlaIII restriction sites are not distributed evenly in the genome. Therefore, in order to ensure equal weight for each of the 50K bins, the bin boundaries for this experiment were designed to encompass and equal number of NlaIII fragments sized between 100bp and 250bp, the size range from which the Illumina libraries were made.

Discussion

Single cell data is much more informative than data obtained from multiple cells, for example from array hybridization of multiple sections of a primary tumor. First, one can afford to analyze many more genomes by multiplex single cell sequencing than by array hybridization (on the order of a hundred single cells versus a dozen sections). Second, one is able to directly assess the exact proportion of cells in a region of a tumor that are malignant. Third, one is able to make a far more detailed analysis of the temporal order of genetic lesions, since one will not be looking at the average state of a population. Fourth, one is able to assess whether tumor heterogeneity is of the anatomically segregated or the co-habitation type, or both.

The invention described herein may be used to determine the original genome copy number (or the gene expression profile) of a sample without amplification distortion. All measurements of copy number or gene expression, especially involving multiple samples, are potential applications. Samples can be nucleic acids derived from single cells or populations of cells. The invention can be used to study cellular heterogeneity in tumors or normal populations of cells.

The invention may be used to generate a genomic profile of a cell, e.g., a cancer cell, or population of cells. From the genomic profile it can be determined if specific genes are amplified or deleted. When the sample is derived from expressed RNA, and the regions are expressed genes, the relative counts comprise an expression profile. From the expression profile the tissue of origin of a cancer cell can be identified. When the sample is a mixture of genomes from different species, and the genomic regions distinguish the different species, the relative counts comprise a census of those different species. For example, from a census we can determine the population of microbial flora and relate the same to disease.

One of the advantages of the use of tags is that it allows the use of restriction endonuclease fragments in genome analysis. Counting methods for genome copy number determination use counts of independent reads. Two reads are not considered independent if they are the result of DNA amplification occurring during sample processing. Typically, this entails shearing sample DNA fragments, and the position of ends of the fragments are used to disambiguate PCR duplicates. If two DNA reads have the same ends, they could be PCR duplicates, and not independent, so they are counted as only one true read. One cannot disambiguate restriction fragments this way, because two reads of the same fragment will have the same ends. However, by the addition of tags, independent reads can be distinguished, if they are

added before any amplification step, as they will have different tags. By contrast, if two reads of the same genome region have the same tag, they are highly likely to be PCR duplicates, and not independent.

The ability to use sequence of restriction fragments in genome analysis can be exploited to reduce the cost of determining copy number profile, much as performing ROMA facilitates genome analysis. By size selection of fragments, one can sample the genome at greater depth at fewer loci, and still obtain a high resolution copy number profile but with fewer reads and therefore at lower cost. This is the essence of representation, and results in complexity reduction, the central element of ROMA. In ROMA, the analysis of the representation is done by hybridization to arrays, and here the representation is analyzed by sequencing.

Further complexity reduction can be achieved by sequence capture. The sequence capture can be designed specifically for the fragments that can be predicted by in silico methods to map to the regions where one wants copy number mutation. In this way one can target special regions of the genome, such as the exome, or known cancer genes. Very little sequence of the sample is then needed for copy number determination, and samples can be multiplexed at a great cost saving.

The invention disclosed herein may also be used to determine the proportion of components in a mixture. For example, suppose one possesses a set of reagents, such as antibody molecules, that bind to a set of substrates, such as different proteins, and one desires to detect the relative proportion of those substrates in a sample, such as blood. First, one covalently combines a tag having a sample tag with each reagent, where the sample tag portion identifies the reagent, and the varietal tag will be used in counting. The tagged reagents are then combined with the sample, and reagents that combine with substrates are physically separated. For example, the substrates may be biotinylated before being mixed with reagents, then reagents combining with substrates separated by avidin affinity chromatography. The tags are amplified by PCR and sequenced, effectively yielding relative counts for each reagent species and therefore each substrate. In this case, the constant portions of the tags include the sample identifier and nucleotides used for amplification.

The invention provides the following advantages: (1) the method is simple; (2) with statistical analysis the confidence of a measurement can be determined with great accuracy; (3) this is a sequence based method, which in the long run will be cheaper than arrays; (4) long read lengths are not necessary, providing savings; (5) because tagged nucleic acid molecules are

counted, many processing methods that create distortion in yields can be applied without major loss in the tagged nucleic acid molecule count; (6) among these processing methods are hybrid selection (aka capture microarrays), which restrict the sequence information to regions of importance, and thus reduce cost by reducing the total number of reads per sample needed for determination of the desired information; (7) the molecules from one sample can be combined with tags and a sample tag, creating sample tagged nucleic acid molecules which can then be pooled, and after processing and sequencing, the sample tags decoded, giving a profile for each sample; (8) pooling adds efficiency and uniformity; (9) the samples can be individual cells, yielding large numbers of single cell measurements cheaply.

The invention described herein has clinical applications as described above, including: (1) determination of the genomic heterogeneity of tumors; (2) detection of tumor cells in needle biopsies, in particular an assessment of the heterogeneity of the cells in a biopsy; (3) assessment of the spread of a cancer to regional lymph nodes; and (4) detection of malignant cells in sites such as blood and bone marrow. The presence of cells with the molecular signature of the primary in the blood or bone marrow opens up the possibility of using more general single cell/molecule assays as a first line of defense in the early detection of cancer, i.e. breast cancer in women at risk.

When there is radiological evidence of a breast mass, the patient most commonly undergoes fine needle biopsy, which is examined by the pathologist who makes an assessment of the malignancy of the mass. Because needle aspiration destroys the histology of the biopsy, one of the pathologist's canonical tools for assessing malignancy is not available. Single cell sequencing would not only more accurately predict the presence of malignant cells, but also give what we might call "genomic staging." Is the tumor a simplex type of complex type? Is it mongenomic or polygenomic? Are there critical loci that are amplified or deleted that might immediately suggest a preoperative therapy?

The single cell/molecule approach allows the assessment of the spread of cancer to regional lymph nodes and to establish the DNA-methylation states at switch regions.

Successful application of single cell/molecule detection in the above contexts will lead to application to the detection of cancer recurrence, and the measurement of cancer load. Detection of recurrence at the earliest possible time may prolong survival for a large class of patients by earlier intervention than they would otherwise receive. Moreover, if cancer load were monitored in the patient, even before the site of the cancer growth is discernible by

imaging, then one could measure the response of the cancer to a variety of therapeutic measurements. That is, the most effective choice of therapy could be determined empirically in the patient.

Additional applications exist in the diagnosis and selection of treatment for patients with auto-immune diseases, and in vitro fertilization (IVF) screening of pre-implant embryos for genetic defects such as trisomy-21.

References

The disclosures of the following references in their entireties are hereby incorporated by reference into this application.

1. Pinard R, de Winter A, Sarkis GJ, Gerstein MB, Tartaro KR, Plant RN, et al. Assessment of whole genome amplification-induced bias through high-throughput, massively parallel whole genome sequencing. BMC Genomics. 2006; 7:216. PMCID: PMC1560136.

2. Klein, C. et al. Comparative genomic hybridization, loss of heterozygosity, and DNA sequence analysis of single cells. PNAS. 1999; 96, pp. 4494–4499.

3. Stoecklein, N. et al. SCOMP Is Superior to Degenerated Oligonucleotide Primed-Polymerase Chain Reaction for Global Amplification of Minute Amounts of DNA from Microdissected Archival Tissue Samples. American Journal of Pathology. 2002; 161(1):43-51.

4. Parameswaran et al. A pyrosequencing-tailored nucleotide barcode design unveils opportunities for large-scale sample multiplexing. Nucleic Acids Research. 2007; 35(19): e130.

5. U.S. Patent No. 7,622,281, Ronaghi et al.

6. Eid et al. Real-Time DNA Sequencing from Single Polymerase Molecules. Science. 2009; 323:133-138.

7. U.S. Patent No. 5,639,603, Dower et al.

8. U.S. Patent Application Publication No. 2006/0073506, Christians et al.

9. Navin N, Krasnitz A, Rodgers L, Cook K, Meth J, Kendall J, et al. Inferring tumor progression from genomic heterogeneity. Genome Res. 2010;20(1):68-80. PMCID: 2798832.

10. Fisher B, Redmond CK, Fisher ER. Evolution of knowledge related to breast cancer heterogeneity: a 25-year retrospective. J Clin Oncol. 2008;26(13):2068-71.

11. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009;10(3):R25. PMCID: PMC2690996.

12. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol. 1987;4(4):406-25.

13. De Baetselier P, Roos E, Brys L, Remels L, Gobert M, Dekegel D, et al. Nonmetastatic tumor cells acquire metastatic properties following somatic hybridization with normal cells. Cancer Metastasis Rev. 1984;3(1):5-24.

14. Duelli DM, Padilla-Nash HM, Berman D, Murphy KM, Ried T, Lazebnik Y. A virus causes cancer by inducing massive chromosomal instability through cell fusion. Curr Biol. 2007;17(5):431-7.

15. Jorgensen HF, Adie K, Chaubert P, Bird AP. Engineering a high-affinity methyl-CpG-binding protein. Nucleic Acids Res. 2006;34(13):e96. PMCID: PMC1540740.

16. Meehan RR, Lewis JD, Bird AP. Characterization of MeCP2, a vertebrate DNA binding protein with affinity for methylated DNA. Nucleic Acids Res. 1992;20(19):5085-92. PMCID: PMC334288.

17. Huang J., Pang J., Watanabe T., Ng HK, Ohgaki H. Whole genome amplification for array comparative genomic hybridization using DNA extracted from formalin-fixed, paraffin-embedded histological sections. J Mol Diagn. 2009 Mar;11(2):109-16. Epub 2009 Feb 5.

18. Talseth-Palmer BA, Bowden NA, Hill A, Meldrum C, Scott RJ. Whole genome amplification and its impact on CGH array profiles. BMC Res Notes. 2008 Jul 29;1:56.

19. Hodges E, Smith AD, Kendall J, Xuan Z, Ravi K, Rooks M, Zhang MQ, Ye K, Bhattacharjee A, Brizuela L, McCombie WR, Wigler M, Hannon GJ, Hicks JB. High definition profiling of mammalian DNA methylation by array capture and single molecule bisulfite sequencing. Genome Res. 2009 Sep;19(9):1593-605. Epub 2009 Jul 6.

20. Hodges E, Xuan Z, Balija V, Kramer M, Molla MN, Smith SW, Middle CM, Rodesch MJ, Albert TJ, Hannon GJ, McCombie WR. Genome-wide in situ exon capture for selective resequencing. Nat Genet. 2007 Dec;39(12):1522-7. Epub 2007 Nov 4.

21. genome.ucsc.edu/index.html?org=Human&db=hg19&hgsid=171216665

22. Miner B.E., Stoger, R.J., Burden, A.F., Laird, C.D., Hansen R.S.. Molecular barcodes detect redundancy and contamination in hairpin-bisulfite PCR. Nucleic Acids Research, 2004:32(17):e135.

23. U.S. Patent No. 7,537,897, issued May 26, 2009 (Brenner et al.).

24. McCloskey M.L., Stoger, R., Hansen, R.S., Laird, C.D. Encoding PCR Products with Batch-stamps and Barcodes. Biochem. Genet. 2007:45:761-767.

What is claimed is:

1. A method for obtaining from genomic material genomic copy number information unaffected by amplification distortion, comprising:

a) obtaining segments of the genomic material;

b) tagging the segments with substantially unique tags to generate tagged nucleic acid molecules, such that each tagged nucleic acid molecule comprises one segment of the genomic material from step (a) and a tag;

c) subjecting the tagged nucleic acid molecules to amplification by polymerase chain reaction (PCR);

d) generating tag associated sequence reads by sequencing the product of step (c);

e) assigning each tagged nucleic acid molecule to a location on a genome associated with the genomic material by mapping the subsequence of each tag associated sequence read corresponding to a segment of the genomic material to a location on the genome; and

f) counting the number of tagged nucleic acid molecules having a different tag that have been assigned to the same location on the genome,

thereby obtaining genomic copy number information unaffected by amplification distortion.

2. The method of claim 1, further comprising estimating a genomic copy number of a region of the genome comprising more than one location on the genome by assigning as the copy number of the region the highest count obtained in step (f) for the locations within the region.

3. The method of claim 1, further comprising comparing a count obtained in step (f) for a location on the genome to a count for the same location obtained from a reference sample, thereby estimating a relative genomic copy number of the location.

4. The method of claim 1, further comprising

a) summing the counts obtained in step (f) for locations on the genome which comprise a first region of the genome, wherein the first region comprises more than one location;

b) summing the counts obtained in step (f) for locations on the genome which comprise a second region of the genome, wherein the second region is comprised of a number of locations which is comparable to the number of locations of the first region;

c) comparing the value obtained in step (a) to the value obtained in step (b),

thereby estimating the relative genomic copy number of the first region of the genome to the genomic copy number of the second region of the genome.

5. The method of claim 4, wherein step (b) further comprises

i) summing the counts obtained in step (f) for locations on the genome which comprise a third region of the genome, wherein the third region is comprised of a number of locations which is comparable to the number of locations of the first region; and

ii) obtaining an average of the sum of the counts obtained in step (f) for locations which comprise the second region and the sum of the counts obtained in step (f) for locations which comprise the third region.

6. The method of any one of claims 4-5, wherein the second region of the genome comprises a centromere.

7. The method of claim 1, further comprising summing the counts obtained in step (f) for locations which comprise a region of the genome, and comparing the sum to a sum

obtained from a reference sample for the same region of the genome, thereby estimating a relative genomic copy number of the region of the genome.

8. A method for obtaining from mRNA transcripts mRNA copy number information unaffected by amplification distortion, comprising:

a) obtaining cDNA from the mRNA transcripts;

b) optionally obtaining segments of the cDNA;

c) tagging the cDNA or segments of cDNA with substantially unique tags to generate tagged nucleic acid molecules, such that each tagged nucleic acid molecule comprises

i) one cDNA from step (a) or one segment of the cDNA from step (b), and

ii) a tag;

d) subjecting the tagged nucleic acid molecules to amplification by polymerase chain reaction (PCR);

e) generating tag associated sequence reads by sequencing the product of step (d);

f) assigning each tagged nucleic acid molecule to a location on a cDNA library associated with the mRNA transcripts by mapping the subsequence of each tag associated sequence read corresponding to a cDNA or a segment of cDNA to a location on the cDNA library; and

g) counting the number of tagged nucleic acid molecules having a different tag that have been assigned to the same location on the cDNA library,

thereby obtaining mRNA copy number information unaffected by amplification distortion.

9. A method for obtaining from mRNA transcripts mRNA copy number information unaffected by amplification distortion, comprising:

a) generating tagged nucleic acid molecules, comprising:

i) subjecting the mRNA transcripts to a polymerase reaction in the presence of primers capable of hybridizing to the polyA tail of the mRNA transcripts under conditions that promote the formation of only one complement, thereby generating first order derivative strands;

ii) adding a polynucleotide tail to the first order derivative strands; and

iii) subjecting the first order derivative strands to a polymerase reaction in the presence of primers capable of hybridizing to the polynucleotide tail added in step (ii) under conditions that promote the formation of only one complement, thereby generating second order derivative strands,

wherein the primers of at least one of steps (i) and (iii) comprise substantially unique primers, thereby generating tagged nucleic acid molecules;

b) subjecting the tagged nucleic acid molecules to polymerase chain reaction (PCR) amplification;

c) generating tag associated sequence reads by sequencing the product of step (b);

d) assigning each tagged nucleic acid molecule to a location on a cDNA library associated with the mRNA transcripts by mapping the subsequence of each tag associated sequence read corresponding to a mRNA transcript to a location on the cDNA library; and

e) counting the number of tagged nucleic acid molecules having a different tag that have been assigned to the same location on the cDNA library,

thereby obtaining mRNA copy number information unaffected by amplification distortion.

10. A method for obtaining from genomic material DNA methylation information unaffected by amplification distortion, comprising:

a) obtaining segments of the genomic material;

b) adding a polynucleotide tail to the ends of the segments of the genomic material to generate zero-th order derivative strands;

c) subjecting the zero-th order derivative strands to a polymerase reaction in the presence essentially unique primers, wherein the essentially unique primers are capable of hybridizing to the polynucleotide tail of the zero-th order strands, under conditions that promote the formation of only one complement, thereby generating tagged nucleic acid molecules;

d) separating the tagged nucleic acid molecules into a group consisting of hemi-methylated tagged nucleic acid molecules and a group consisting of unmethylated tagged nucleic acid molecules;

e) subjecting each group of step (d) to amplification by polymerase chain reaction (PCR);

f) generating tag associated sequence reads by sequencing the product of step (e);

g) assigning each tagged nucleic acid molecule to a location on a genome associated with the genomic material by mapping the subsequence of each tag associated sequence read corresponding to a segment of the genomic material to a location on the genome; and

h) counting the number of tagged nucleic acid molecules having a different tag that have been assigned to the same location on the genome,

thereby obtaining DNA methylation information unaffected by amplification distortion.

11. The method of any one of claims 1-7, wherein tagging the segments to generate tagged nucleic acid molecules comprises:

a) adding a polynucleotide tail to the ends of the segments of the genomic material to generate zero-th order derivative strands;

b) subjecting the zero-th order derivative strands of step (a) to a polymerase reaction in the presence of primers capable of hybridizing to the polynucleotide tail of the zero-th order strands under conditions that promote the formation of only one complement, thereby generating first order derivative strands;

c) adding a polynucleotide tail to the first order derivative strands;

d) subjecting the first order derivative strands to a polymerase reaction in the presence of primers capable of hybridizing to the polynucleotide tail of the first order derivative strands under conditions that promote the formation of only one complement, thereby generating second order derivative strands,

wherein the primers of at least one of steps (b) and (d) comprise substantially unique primers, thereby generating tagged nucleic acid molecules.

12. The method of any one of claims 1-7, wherein tagging the segments to generate tagged nucleic acid molecules comprises:

a) adding a polynucleotide tail to the ends of the segments of genomic material to generate zero-th order derivative strands,

b) subjecting the zero-th order derivative strands of step (a) to a ligation reaction in the presence of primers capable of hybridizing to the polynucleotide tail of the zero-th order strands added in step (a) under conditions that promote the ligation of a primer to the 5' ends of the zero-th order derivative strands,

c) subjecting the product of step (b) to a polymerase reaction in the presence of primers capable of hybridizing to the polynucleotide tail added in step (a) under conditions which promote the formation of only one complement, wherein the primers of step (c) have different nucleotide sequences than the primers of step (b), and wherein the polymerase has 3'-5' proofreading activity,

wherein the primers of at least one of steps (b) and (c) comprise substantially unique primers, thereby generating tagged nucleic acid molecules.

13. The method of any one of claims 9-12, wherein adding a polynucleotide tail comprises the use of a terminal transferase.

14. The method of any one of claims 1-7, wherein tagging the segments to generate tagged nucleic acid molecules comprises ligation of adaptors comprising the tags to at least one end of the segments of the genomic material.

15. The method of claim 14, wherein the adaptors comprising the tags are ligated to only one end of the segments of the genomic material.

16. The method of any one of claims 1-15, wherein the tags comprise a sequence that aids PCR amplification.

17. The method of any one of claims 1-16, wherein each tagged nucleic acid molecule comprises one tag.

18. The method of any one of claims 1-17, wherein each tagged nucleic acid molecule comprises more than one tag.

19. The method of any one of claims 1-7 or 10-18, wherein segments of the genomic material are produced by restriction endonuclease digestion, mechanical shearing, heating, or sonication.

20. The method of claim 8, wherein segments of the cDNA are produced by restriction endonuclease digestion, mechanical shearing, heating, or sonication.

21. The method of any one of claims 8-9, 13, 17-18 or 20, wherein the maximum copy number of a location on a cDNA library is not less than the number of tagged nucleic

acid molecules having a different tag that have been assigned to the same location on the cDNA library.

22. The method of any one of claims 1-7 or 10-19, further comprising analyzing mRNA copy number.

23. The method of any one of claims 10, 13, 16-19, or 22, wherein separating the tagged nucleic acid molecules into a group consisting of hemi-methylated tagged nucleic acid molecules and a group consisting of unmethylated tagged nucleic acid molecules is by cleavage with methylation sensitive restriction enzymes, partitioning with antibodies, or partitioning with methyl-C binding proteins directed to methylated or hydroxymethylated cytosine.

24. The method of any one of claims 1-23, wherein the tagged nucleic acid molecules are subject to hybrid capture prior to PCR or prior to sequencing.

25. The method of any one of claims 1-24, wherein each tagged nucleic acid molecule differs at more than one nucleotide.

26. The method of any one of claims 1-25, wherein the tag sequences further comprise a sample tag.

27. The method of claim 26, wherein the tagged nucleic acid molecules are pooled with a plurality of tagged nucleic acid molecules having a different sample tag prior to PCR amplification or prior to sequencing.

28. The method of claim 27, further comprising deconvoluting the tag associated sequence reads by grouping the tag associated sequence reads according to sample tag.

29. The method of any one of claims 1-28, wherein the tagged nucleic acid molecules are generated from a single species.

30. The method of claim 29, wherein the tagged nucleic acid molecules are generated from a single organism.

31. The method of any one of claims 1-30, wherein tagged nucleic acid molecules are generated from a single cell.

32. The method of claim 31, wherein the single cell is from a needle aspirate of suspected cancer lesions.

33. The method of claim 31, wherein the single cell is from a core biopsy of suspected cancer lesions.

34. The method of claim 29, wherein the tagged nucleic acid molecules are generated from two or more organisms.

35. The method of claim 34, wherein the tagged nucleic acid molecules are generated from a single cell of each organism.

36. The method of any one of claims 30-35, wherein the single species is human.

37. The method of any one of claims 1-28, wherein the tagged nucleic acid molecules are generated from two or more species.

38. The method of claim 37, wherein the tagged nucleic acid molecules are generated from a population of microbes.

39. The method of claim 38, comprising comparing the genomic copy number information obtained for different species of the population to determine the relative count of those different species in the population.

40. A composition of matter derived from genomic material comprising tagged nucleic acid molecules, said tagged nucleic acid molecules being produced by a process comprising:

a) obtaining segments of the genomic material;

b) adding a polynucleotide tail to the ends of the segments of the genomic material to generate zero-th order derivative strands;

c) subjecting the zero-th order derivative strands of step (b) to a polymerase reaction in the presence of primers capable of hybridizing to the polynucleotide tail of the zero-th order strands under conditions that promote the formation of only one complement, thereby generating first order derivative strands;

d) adding a polynucleotide tail to the first order derivative strands;

e) subjecting the first order derivative strands to a polymerase reaction in the presence of primers capable of hybridizing to the polynucleotide tail of the first order derivative strands under conditions that promote the formation of only one complement, thereby generating second order derivative strands,

wherein the primers of at least one of steps (c) and (e) comprise substantially unique primers, thereby generating tagged nucleic acid molecules.

41. A composition of matter derived from genomic material comprising tagged nucleic acid molecules, said tagged nucleic acid molecules being produced by a process comprising:

a) obtaining segments of genomic material;

b) adding a polynucleotide tail to the ends of the segments of genomic material to generate zero-th order derivative strands,

c) subjecting the zero-th order derivative strands of step (b) to a ligation reaction in the presence of primers capable of hybridizing to the polynucleotide tail of the zero-th order strands added in step (b) under conditions that promote the ligation of a primer to the 5' ends of the zero-th order derivative strands,

d) subjecting the product of step (c) to a polymerase reaction in the presence of primers capable of hybridizing to the polynucleotide tail added in step (b) under conditions which promote the formation of only one complement, wherein the primers of step (d) have different nucleotide sequences than the primers of step (c), and wherein the polymerase has 3'-5' proofreading activity,

wherein the primers of at least one of steps (c) and (d) comprise substantially unique primers, thereby generating tagged nucleic acid molecules.

42. A composition of matter derived from mRNA transcripts comprising tagged nucleic acid molecules, said tagged nucleic acid molecules being produced by a process comprising:

a) obtaining cDNA from the mRNA transcripts;

b) optionally obtaining segments of the cDNA; and

c) tagging the cDNA or segments of cDNA with substantially unique tags to generate tagged nucleic acid molecules, wherein each tagged nucleic acid molecule comprises

iii) one cDNA from step (a) or one segment of the cDNA from step (b) and

iv) a tag.

43. A composition of matter derived from mRNA transcripts comprising tagged nucleic acid molecules, said tagged nucleic acid molecules being produced by a process comprising:

a) obtaining mRNA transcripts;

b) subjecting the mRNA transcripts to a polymerase reaction in the presence of primers capable of hybridizing to the polyA tail of the mRNA transcripts under conditions that promote the formation of only one complement, thereby generating first order derivative strands;

c) adding a polynucleotide tail to the first order derivative strands;

d) subjecting the first order derivative strands to a polymerase reaction in the presence of primers capable of hybridizing to the polynucleotide tail added in step (c) under conditions that promote the formation of only one complement, thereby generating second order derivative strands,

wherein the primers of at least one of steps (b) and d) comprise substantially unique primers, thereby generating tagged nucleic acid molecules.

44. The composition of matter of any one of claims 40-41 or 43, wherein adding a polynucleotide tail comprises the use of a terminal transferase.

45. The composition of matter of any one of claims 40-44, wherein each tagged nucleic acid molecule comprises one tag.

46. The composition of matter of any one of claims 40-44, wherein each tagged nucleic acid molecule comprises more than one tag.

47. The composition of matter of any one of claims 40-41 or 44-46, wherein segments of the genomic material are produced by restriction endonuclease digestion, mechanical shearing, heating, or sonication.

48. The composition of matter of any one of claims 42-46, wherein segments of the cDNA are produced by restriction endonuclease digestion, mechanical shearing, heating, or sonication.

49. The composition of matter of any one of claims 40-48, wherein the complexity of the tagged nucleic acid molecules is reduced by hybrid capture.

50. The composition of matter of any one of claims 40-49, wherein each tagged nucleic acid molecules differs at more than one nucleotide.

51. The composition of matter of any one of claims 40-50, wherein the tag sequences further comprise a sample tag.

52. A composition of matter comprising at least two pools of tagged nucleic acid molecules according to claim 51.

53. The composition of matter of any one of claims 40-52, wherein the tagged nucleic acid molecules are generated from a single cell.

54. The composition of matter of claim 53, wherein the single cell is from a needle aspirate of suspected cancer lesions.

55. The composition of matter of claim 53, wherein the single cell is from a core biopsy of suspected cancer legions.

56. The composition of matter of any one of claims 40-52, wherein the tagged nucleic acid molecules are generated from two or more organisms.

57. A kit for determining nucleic acid copy number information unaffected by amplification distortion comprising:

a) a terminal transferase, and

b) plurality of substantially unique primers,

wherein the substantially unique primers comprise substantially unique tags, and wherein the substantially unique primers are capable of hybridizing to the polynucleotide tail of a nucleic acid molecule which has a polynucleotide tail added by the terminal transferase.

58. The kit of claim 57, further comprising a DNA polymerase having 3'-5' proofreading activity.

59. The kit of any one of claims 57-58, wherein the plurality of substantially unique primers comprises 10^n primers, wherein n is an integer from 2 to 9.

60. The kit of any one of claims 57-59, further comprising a second terminal transferase and a primer capable of hybridizing to the polynucleotide tail of a nucleic acid molecule which has a polynucleotide tail added by the second terminal transferase.

61. The kit of any one of claims 57-60, wherein the tags are six nucleotides long.

62. The kit of any one of claims 57-60, wherein the tags are 15 nucleotides long.

63. The kit of any one of claims 57-62, further comprising sample tags.

64. The kit of claim 63, wherein the sample tags are 2 nucleotides long.

65. The kit of claim 63, wherein the sample tags are 4 nucleotides long.

66. The kit of any one of claims 63-65, wherein the sample tags comprise a sample tag set.

67. The kit of any one of claims 63-66, wherein the substantially unique tags comprise the sample tags.

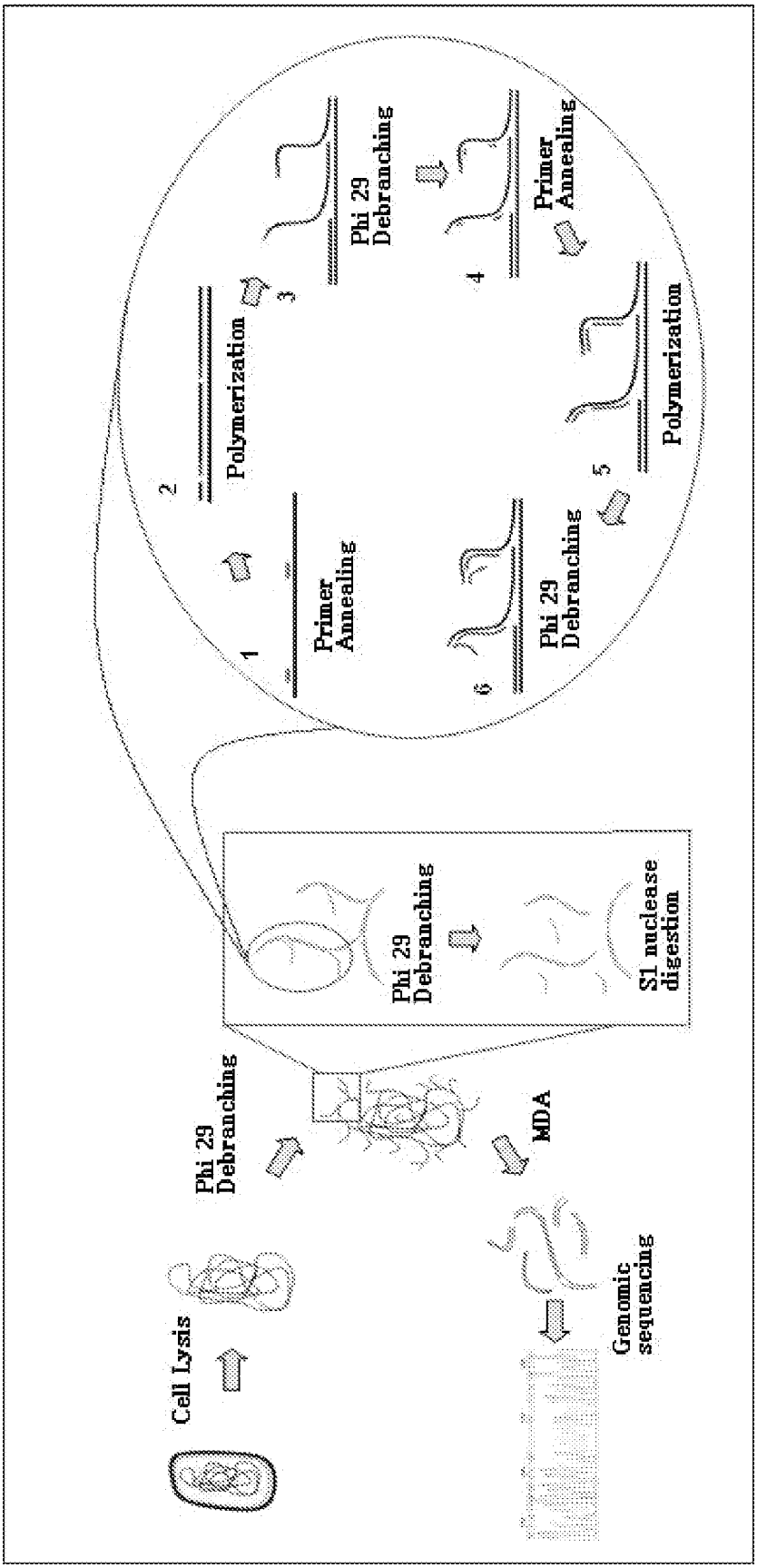

Cell Lysis
Phi 29 Debranching
MDA
Genomic sequencing
Phi 29 Debranching
S1 nuclease digestion
1
Primer Annealing
2
Polymerization
3
Phi 29 Debranching
4
Primer Annealing
5
Polymerization
6
Phi 29 Debranching

Fig. 1

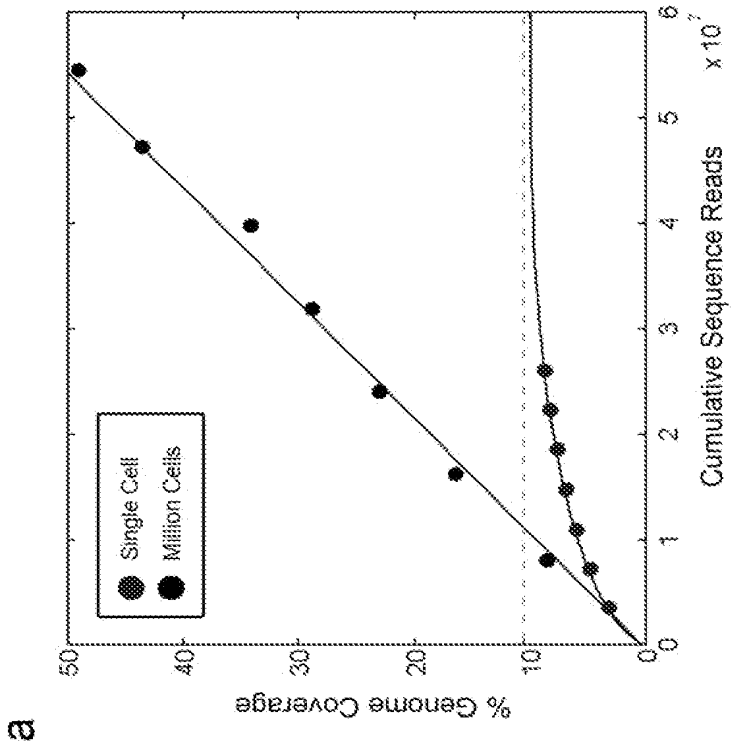

a
% Genome Coverage
0
10
20
30
40
50
Single Cell
Million Cells
0
1
2
3
4
5
6
Cumulative Sequence Reads
x 10^7

Fig. 2

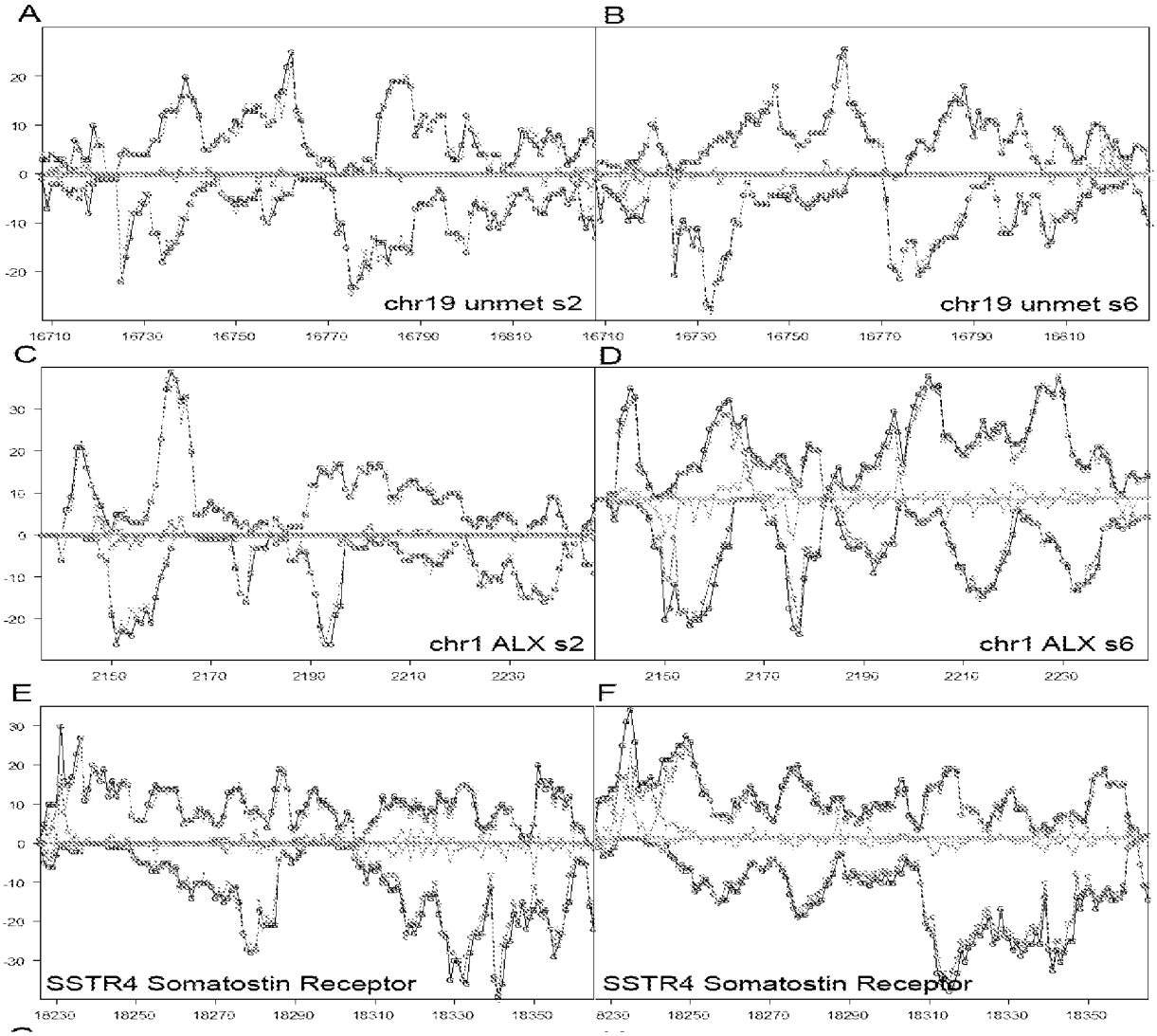

A
B
chr19 unmet s2
chr19 unmet s6
C
D
chr1 ALX s2
chr1 ALX s6
E
F
SSTR4 Somatostin Receptor
SSTR4 Somatostin Receptor

Fig. 3

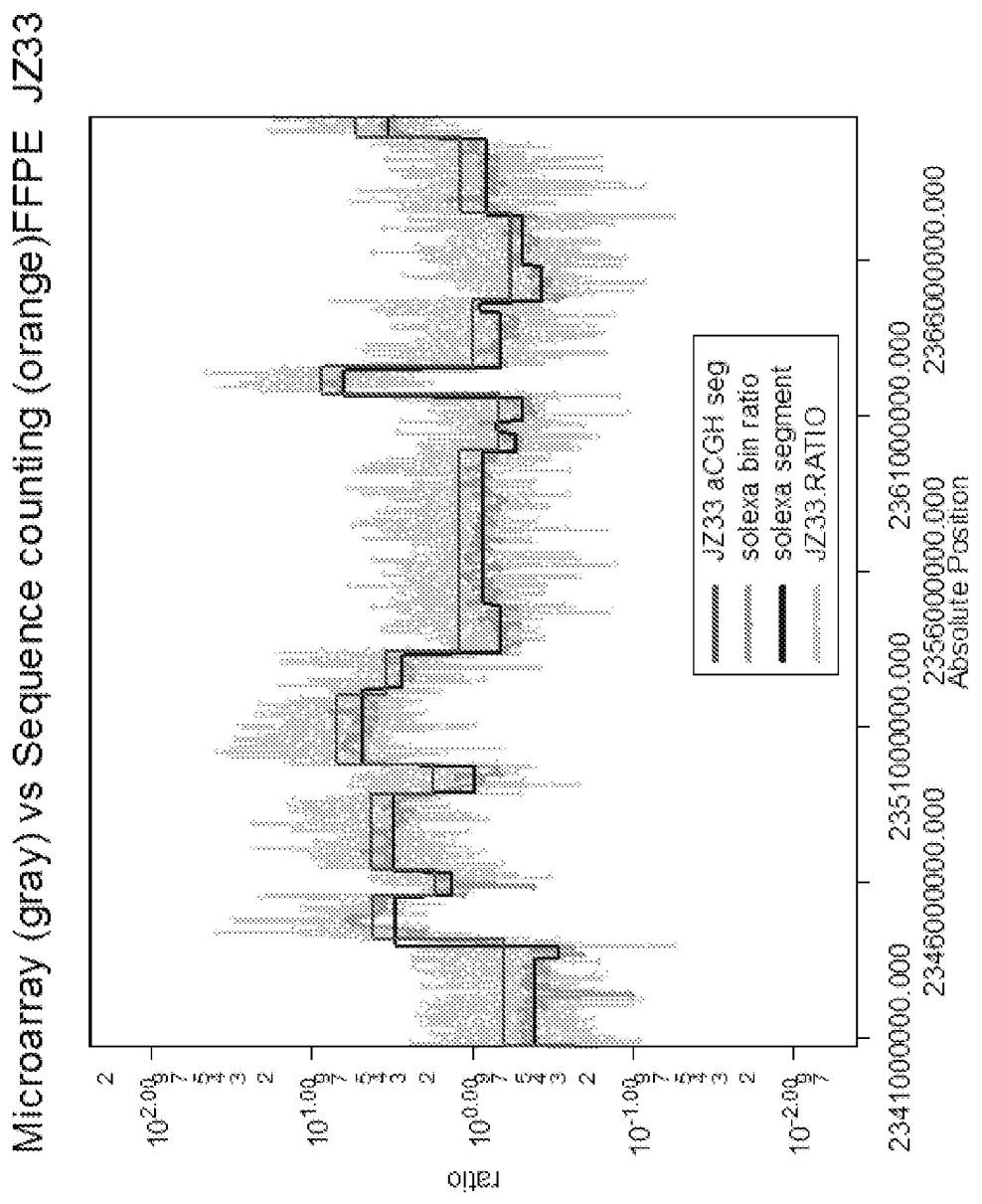

Microarray (gray) vs Sequence counting (orange)FFPE JZ33
ratio
$10^{2.00}$
$10^{1.00}$
$10^{0.00}$
$10^{-1.00}$
$10^{-2.00}$
2341000000.000
2346000000.000
2351000000.000
2356000000.000
2361000000.000
2366000000.000
Absolute Position
JZ33 aCGH seg
solexa bin ratio
solexa segment
JZ33.RATIO

Fig. 4

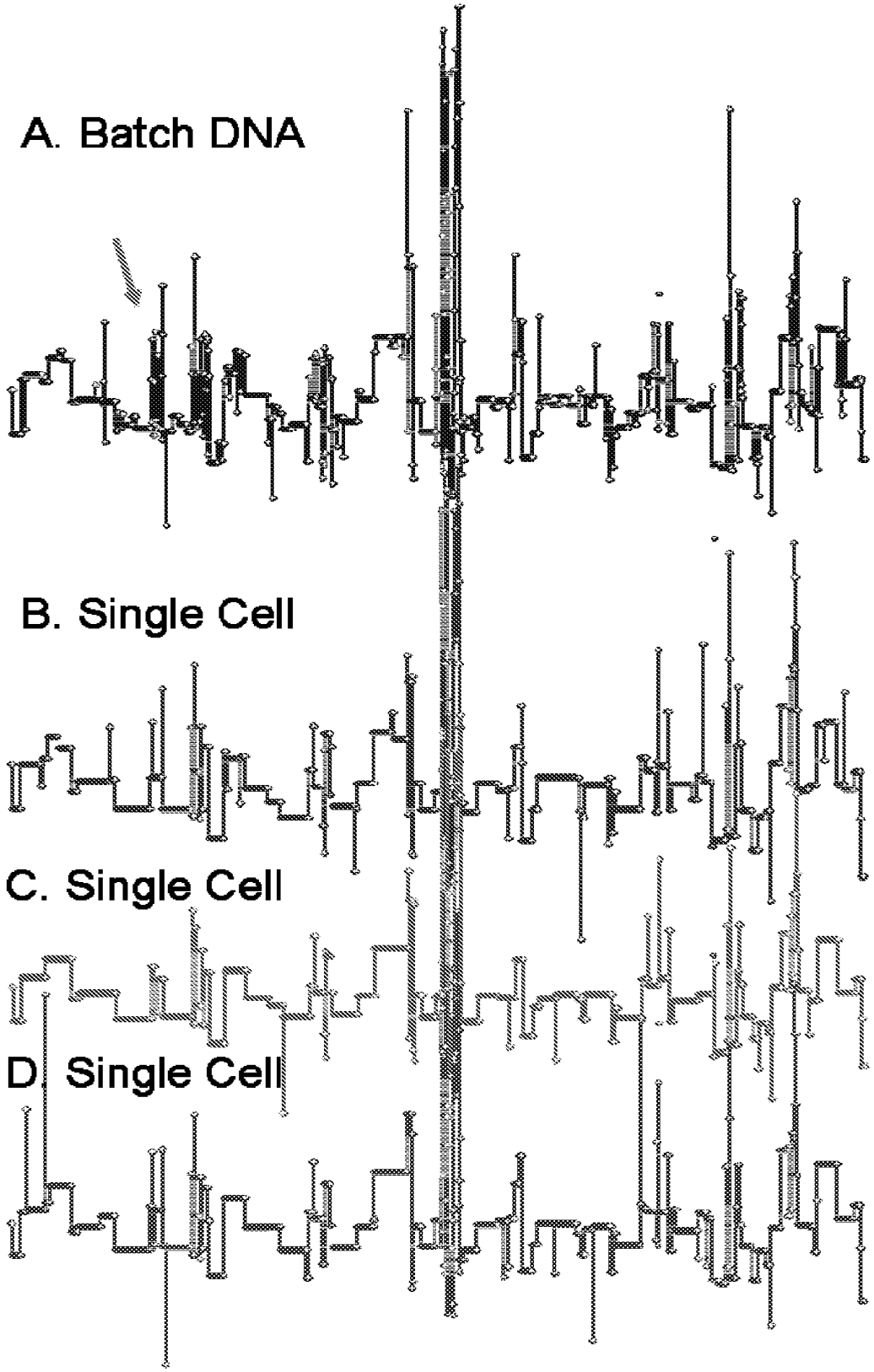


Fig. 5

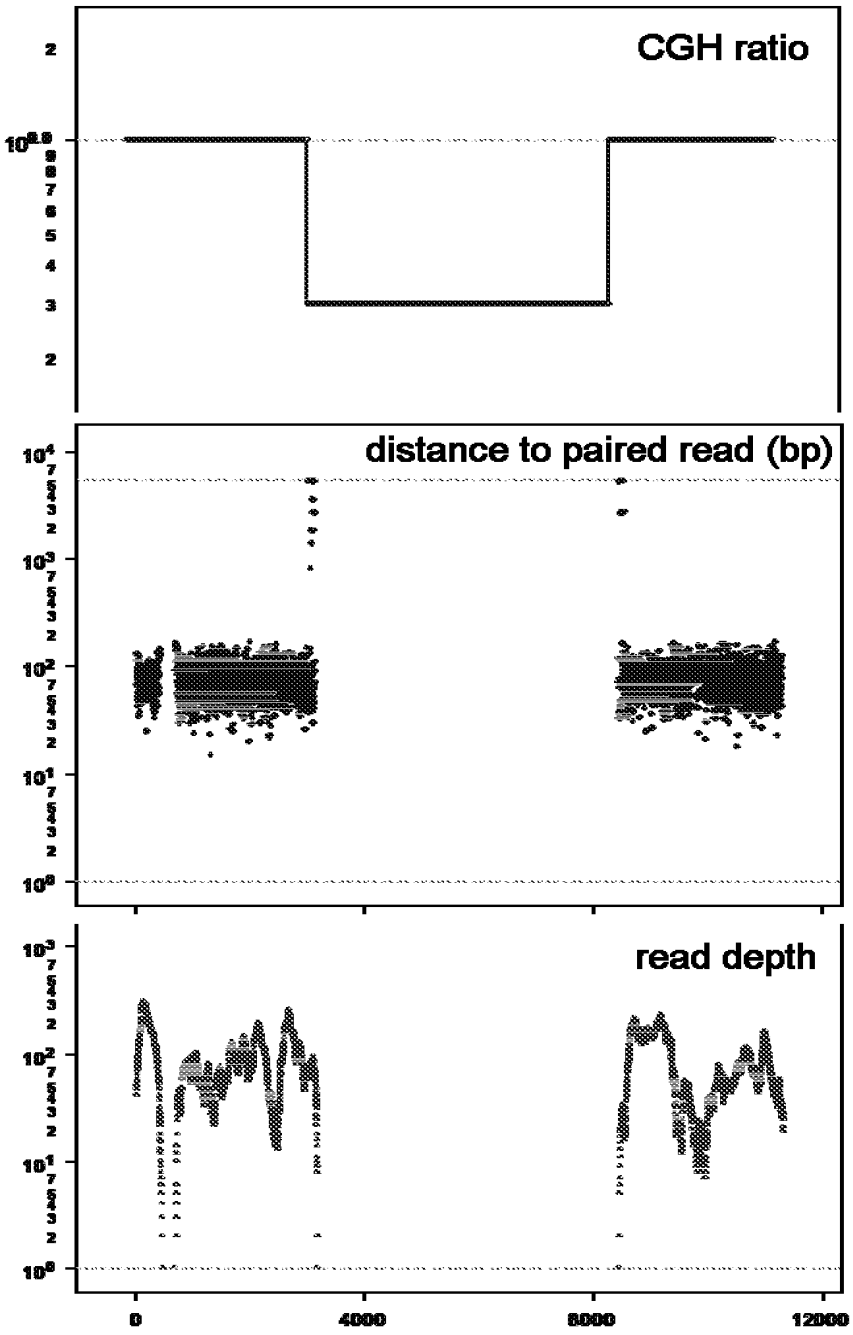

CGH ratio
distance to paired read (bp)
read depth
0
4000
8000
12000

Fig. 6

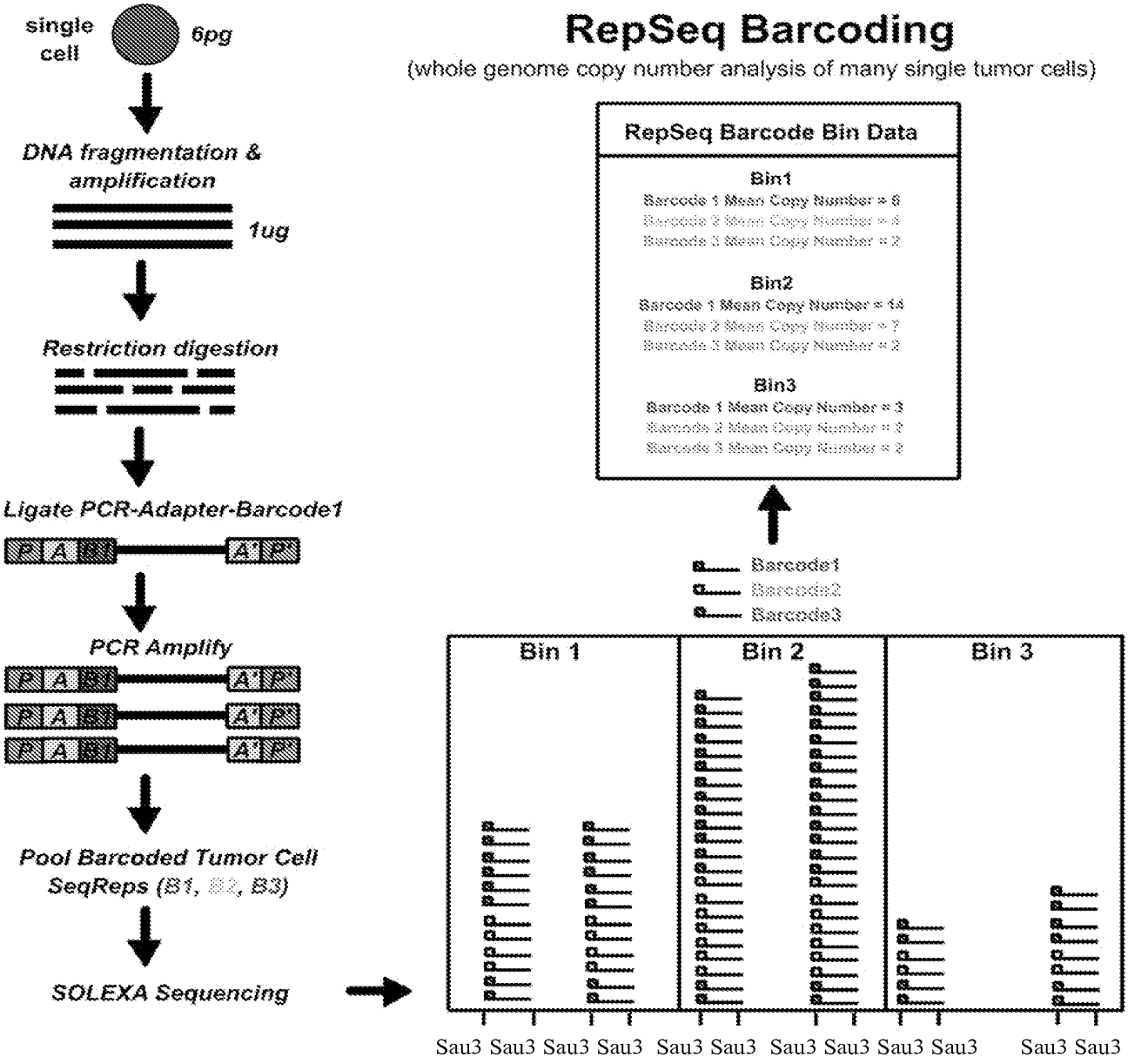

RepSeq Barcoding
(whole genome copy number analysis of many single tumor cells)
single cell
6pg
DNA fragmentation & amplification
1ug
Restriction digestion
Ligate PCR-Adapter-Barcode1
PCR Amplify
Pool Barcoded Tumor Cell SeqReps (B1, B2, B3)
SOLEXA Sequencing
RepSeq Barcode Bin Data
Bin1
Barcode 1 Mean Copy Number = 6
Barcode 2 Mean Copy Number = 4
Barcode 3 Mean Copy Number = 2
Bin2
Barcode 1 Mean Copy Number = 14
Barcode 2 Mean Copy Number = 7
Barcode 3 Mean Copy Number = 2
Bin3
Barcode 1 Mean Copy Number = 3
Barcode 2 Mean Copy Number = 2
Barcode 3 Mean Copy Number = 2
Barcode1
Barcode2
Barcode3
Bin 1
Bin 2
Bin 3
Sau3 Sau3 Sau3 Sau3 Sau3 Sau3 Sau3 Sau3 Sau3 Sau3 Sau3 Sau3
Chromosome 17q12

Fig. 7

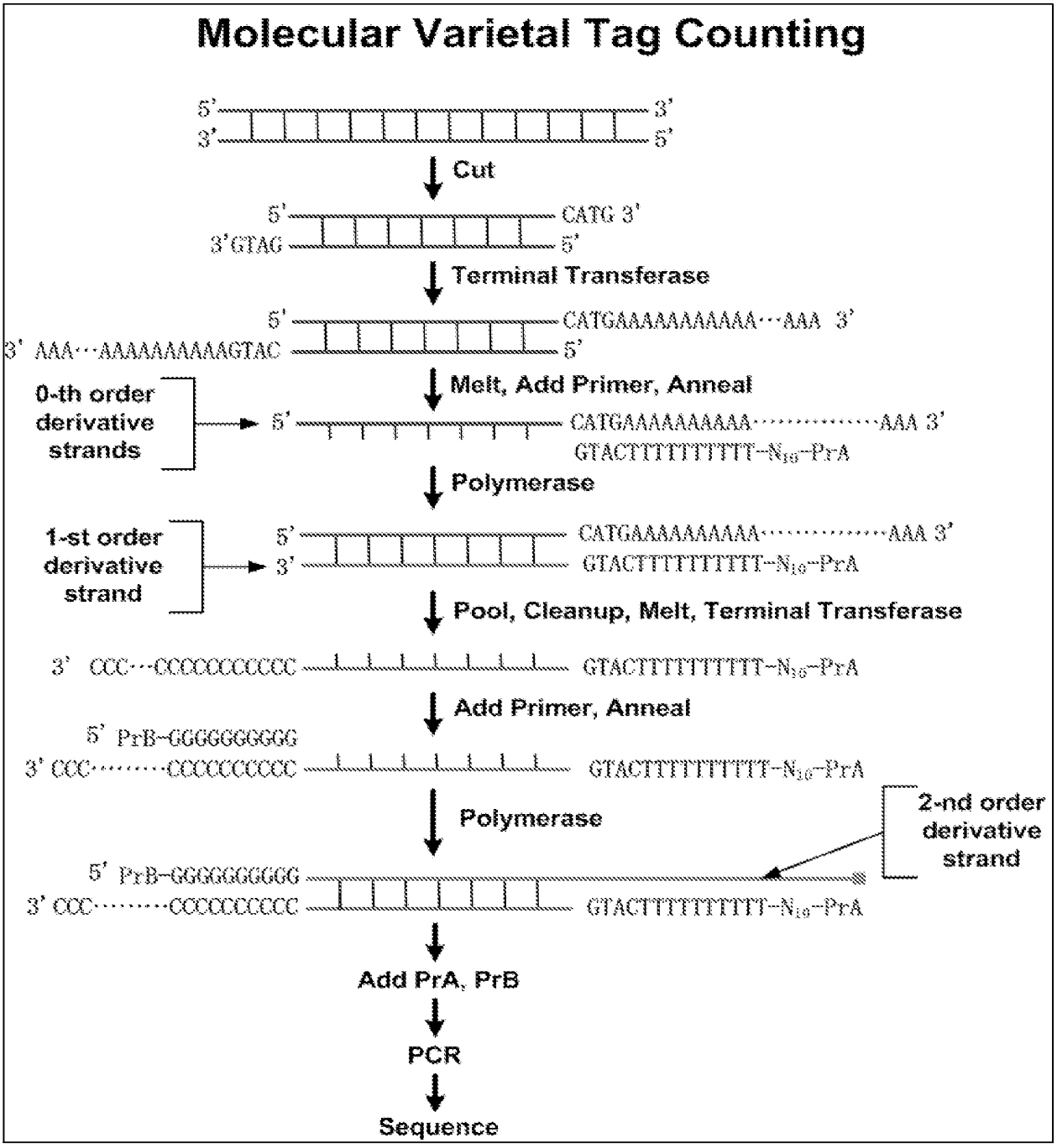

Molecular Varietal Tag Counting
5'
3'
3'
5'
Cut
5'
CATG 3'
3'GTAG
5'
Terminal Transferase
5'
CATGAAAAAAAAAA···AAA 3'
3' AAA···AAAAAAAAAGTAC
5'
Melt, Add Primer, Anneal
0-th order derivative strands
5'
CATGAAAAAAAAAA·····AAA 3'
GTACTTTTTTTTTT-N10-PrA
Polymerase
1-st order derivative strand
5'
3'
CATGAAAAAAAAAA·····AAA 3'
GTACTTTTTTTTTT-N10-PrA
Pool, Cleanup, Melt, Terminal Transferase
3' CCC···CCCCCCCCCC
GTACTTTTTTTTTT-N10-PrA
Add Primer, Anneal
5' PrB-GGGGGGGGG
3' CCC·····CCCCCCCCC
GTACTTTTTTTTTT-N10-PrA
Polymerase
2-nd order derivative strand
5' PrB-GGGGGGGGG
3' CCC·····CCCCCCCCC
GTACTTTTTTTTTT-N10-PrA
Add PrA, PrB
PCR
Sequence

Fig. 8

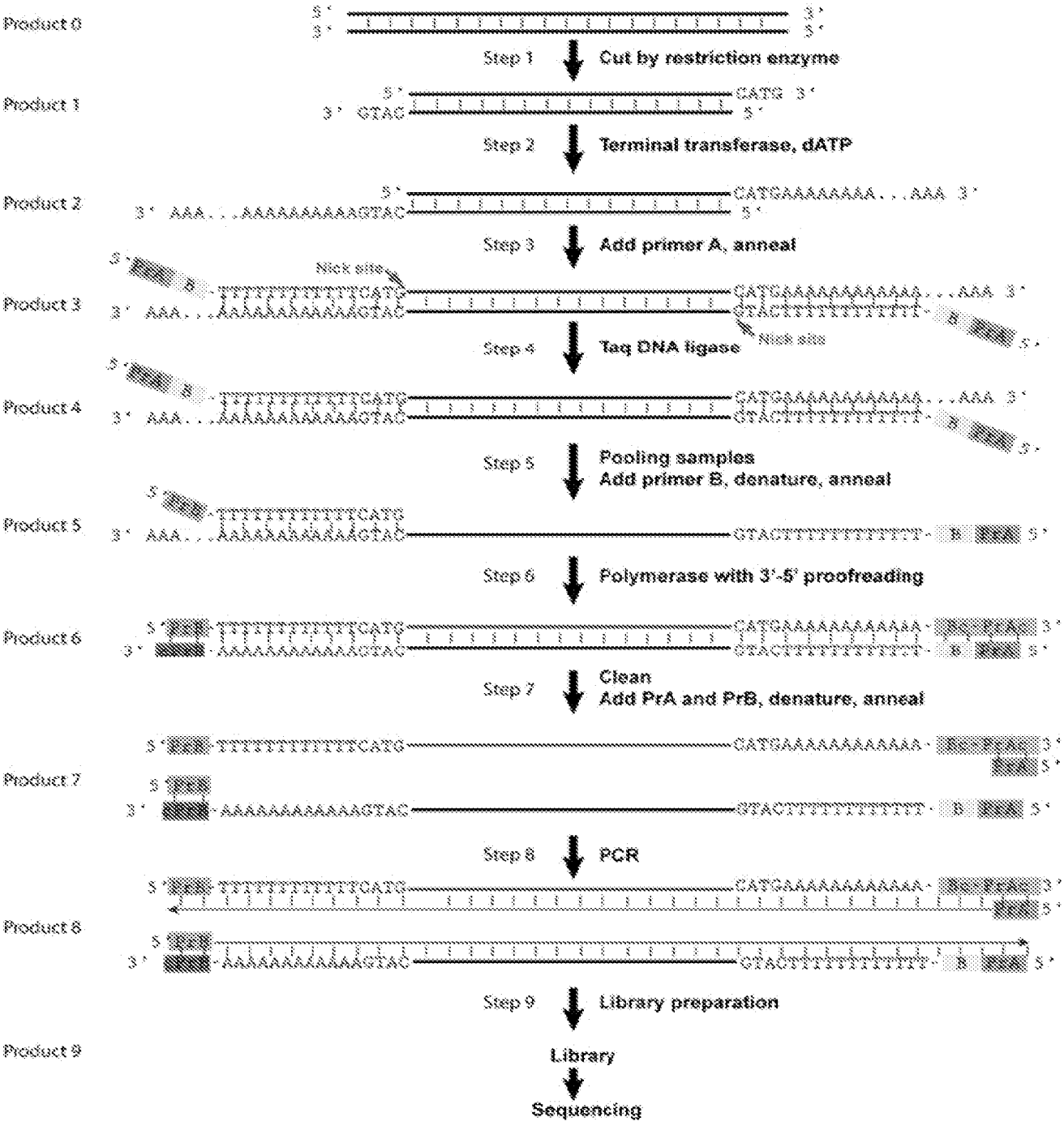

Product 0
Step 1
Cut by restriction enzyme
Product 1
Step 2
Terminal transferase, dATP
Product 2
Step 3
Add primer A, anneal
Nick site
Product 3
Step 4
Taq DNA ligase
Nick site
Product 4
Step 5
Pooling samples
Add primer B, denature, anneal
Product 5
Step 6
Polymerase with 3'-5' proofreading
Product 6
Step 7
Clean
Add PrA and PrB, denature, anneal
Product 7
Step 8
PCR
Product 8
Step 9
Library preparation
Product 9
Library
Sequencing

Fig. 9

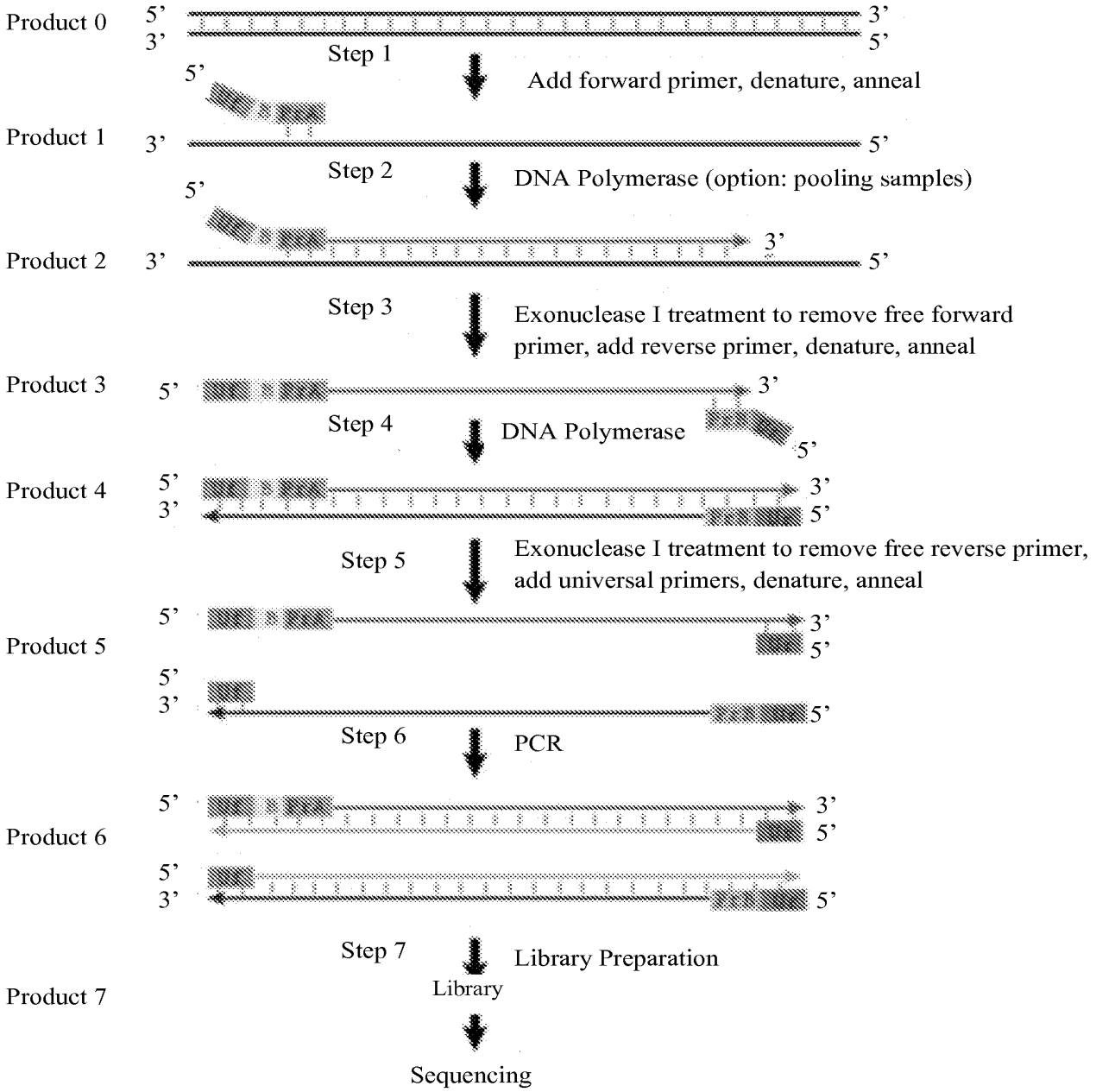

Product 0
5'
3'
3'
5'
Step 1
Add forward primer, denature, anneal
5'
Product 1
3'
5'
Step 2
DNA Polymerase (option: pooling samples)
5'
Product 2
3'
3'
5'
Step 3
Exonuclease I treatment to remove free forward primer, add reverse primer, denature, anneal
Product 3
5'
3'
Step 4
DNA Polymerase
5'
Product 4
5'
3'
3'
5'
Step 5
Exonuclease I treatment to remove free reverse primer, add universal primers, denature, anneal
5'
Product 5
3'
5'
5'
3'
5'
Step 6
PCR
5'
Product 6
3'
5'
5'
3'
5'
Step 7
Library Preparation
Product 7
Library
Sequencing

Fig. 10

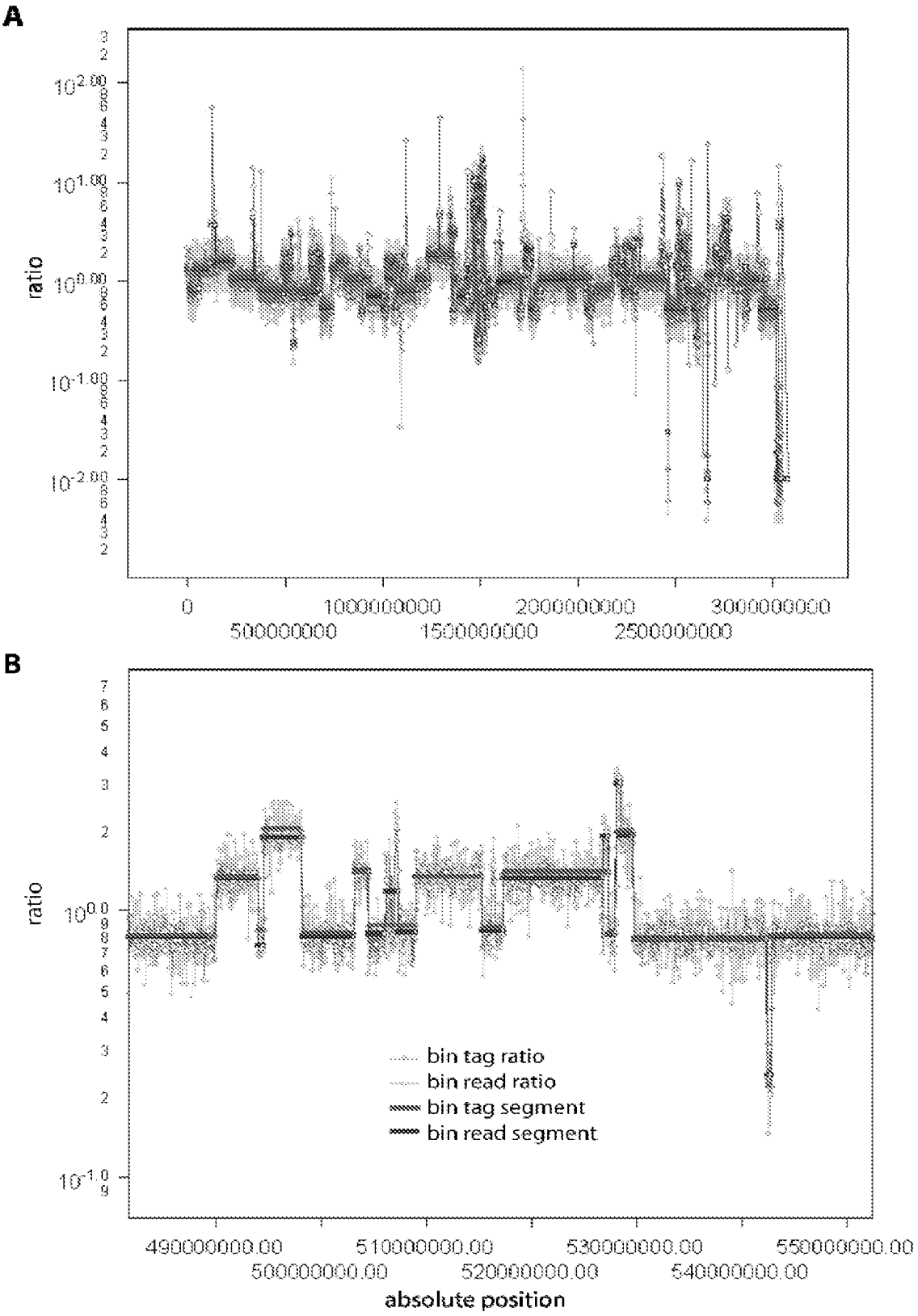

A
ratio
0
500000000
1000000000
1500000000
2000000000
2500000000
3000000000
B
ratio
bin tag ratio
bin read ratio
bin tag segment
bin read segment
490000000.00
500000000.00
510000000.00
520000000.00
530000000.00
540000000.00
550000000.00
absolute position

Fig. 11

Espacenet

Bibliographic data: WO2010064893 (A1) — 2010-06-10

METHOD FOR THE REDUCTION OF REPETITIVE SEQUENCES IN ADAPTER-LIGATED RESTRICTION FRAGMENTS

Inventor(s): TREBBI DANIELE [IT]; VAN EIJK MICHAEL JOSEPHUS THERESIA [NL] ± (TREBBI, DANIELE, ; VAN EIJK, MICHAEL, JOSEPHUS, THERESIA)

Applicant(s): KEYGENE NV [NL]; TREBBI DANIELE [IT]; VAN EIJK MICHAEL JOSEPHUS THER [NL] ± (KEYGENE N.V, ; TREBBI, DANIELE, ; VAN EIJK, MICHAEL, JOSEPHUS, THERESIA)

Classification: **- international:** ***C12N15/00***
- cooperative: **C12N15/1093**

Application number: WO2009NL00232 20091125 Global Dossier

Priority number (s): US20080119864P 20081204

Also published as: CN102239258 (A) EP2367937 (A1) EP2367937 (B1) JP2012510810 (A) US2011287492 (A1)

Abstract of WO2010064893 (A1)

The present invention relates to a method for the reduction of repetitive sequences (or the improvement in low-copy sequences) in DNA samples by a combination of restriction endonuclease treatment, followed by adapter ligation, renaturation kinetics-based fractionation, optionally coupled with duplex sequence nucleases and further restriction endonuclease treatment, followed by adapter ligation. The low-copy enriched fractions can be used in further DNA analysis.

IDT Exhibit 1020.5539

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
10 June 2010 (10.06.2010)

PCT

(10) International Publication Number
WO 2010/064893 A1

(51) International Patent Classification:
C12N 15/00 (2006.01)

(21) International Application Number:
PCT/NL2009/000232

(22) International Filing Date:
25 November 2009 (25.11.2009)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/119,864 4 December 2008 (04.12.2008) US

(71) Applicant *(for all designated States except US)*: **KEYGENE N.V.** [NL/NL]; P.O. Box 216, 6700 AE Wageningen (NL).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **TREBBI, Daniele** [IT/IT]; Via G. Leopardi 16/1, Tavagnacco fraz. Branco, I-33010 Udine (IT). **VAN EIJK, Michael, Josephus, Theresia** [NL/NL]; 12, Pastoor Strijboschstraat, NL-5373 EJ Herpen (NL).

(74) Agent: **DE LANG, R.-J.**; Exter Polak & Charlouis B.V., P.O. Box 3241, NL-2280 GE Rijswijk (NL).

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

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK, 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 2010/064893 A1

(54) **Title**: METHOD FOR THE REDUCTION OF REPETITIVE SEQUENCES IN ADAPTER-LIGATED RESTRICTION FRAGMENTS

(57) **Abstract**: The present invention relates to a method for the reduction of repetitive sequences (or the improvement in low-copy sequences) in DNA samples by a combination of restriction endonuclease treatment, followed by adapter ligation, renaturation kinetics-based fractionation, optionally coupled with duplex sequence nucleases and further restriction endonuclease treatment, followed by adapter ligation. The low-copy enriched fractions can be used in further DNA analysis.

IDT Exhibit 1020.5540

Title: Method for the reduction of repetitive sequences in adapter-ligated restriction fragments

The present invention relates to methods for reducing the amount repetitive sequences in genomes and to methods for enriching samples for low-copy sequences. The invention further relates to the use of renaturation kinetics based fractionation for the enrichment of adapter-ligated restriction fragments for low-copy sequences. The invention further relates to the use of double strand-specific nuclease (DSN) for the enrichment of adapter-ligated restriction fragments for low-copy sequences. The invention further relates to the use of samples enriched for low-copy fragments in fingerprint analysis, SNP mining and marker development.

Background of the invention

Many genomes, in particular plants, contain repetitive sequences, i.e. parts, whether large or smaller, of the DNA that are repeated at several positions in a genome. In particular when genome sizes increase and in particular in plants, where genomes can have large sizes (e.g. Arabidopsis thaliana 130 Mb, Oryza sativa 430 Mb, Zea Mays 2400Mb, Barley 5300 Mb and Onion 15000 Mb per haploid genome). Some genomes such as pepper and maize are said to contain up to 80 % of repetitive sequences.

AFLP (EP534858, Vos et al 1995) is a successful technology for the development of markers and the identification of SNPs that are linked to a particular trait or QTL. To subsequently develop such AFLP markers or SNPs to PCR markers, i.e. markers that can be used with a normal PCR assay, the markers or SNPs have to be converted. The conversion is a laborious process that in many occasions is not successful (Brugmans et al. Nucleic Acids Research 2003 31(10):e55) in particular due to the presence of repetitive sequences in the genome. Repetitive sequences do not yield useful markers or SNPs that can be linked to a trait or a quantitative trait locus. That in itself does not necessarily pose problem, but the presence of repetitive sequences hampers the subsequent conversion of AFLP markers to PCR markers. Therefore, there is a need to enrich restriction fragment mixtures obtained from DNA samples for low-copy sequences, because this will increase the probability for successful conversion to PCR markers and thereby improve the effectiveness of genetic marker development for traits

IDT Exhibit 1020.5541

Summary of the invention

The present inventors have now found that a combination of elements of the AFLP technology, DNA reassociation kinetics and DSN normalisation provides an efficient and reliable way of enriching a DNA sample for low copy sequences. The enrichment for low copy sequences is achieved, according to the invention by providing adapter-ligated restriction fragments from a DNA sample using a first restriction endonuclease, subjecting the adapter-ligated restriction fragments to renaturation kinetics based fractionation and subjecting the fractions to a DSN, restricting the remaining fragments with a second restriction endonuclease and ligating a second adapter to the restriction fragment to thereby obtain a low-copy enriched fraction of the starting DNA.

Definitions:

In the following description and examples, a number of terms are used. In order to provide a clear and consistent understanding of the specification and claims, including the scope to be given such terms, the following definitions are provided. Unless otherwise defined herein, all technical and scientific terms used have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. The disclosures of all publications, patent applications, patents and other references are incorporated herein in their entirety by reference.

AFLP: AFLP refers to a method for selective amplification of nucleic acids based on digesting a nucleic acid with one or more restriction endonucleases to yield restriction fragments, ligating adaptors to the restriction fragments and amplifying the adaptor-ligated restriction fragments with at least one primer that is (part) complementary to the adaptor, (part) complementary to the remains of the restriction endonuclease, and that further contains at least one randomly selected nucleotide from amongst A,C, T, or G (or U as the case may be) at the 3'end of the primer. AFLP does not require any prior sequence information and can be performed on any starting DNA. In general, AFLP comprises the steps of:

(a) digesting a nucleic acid, in particular a DNA or cDNA, with one or more specific restriction endonucleases, to fragment the DNA into a corresponding series of restriction fragments;

(b) ligating the restriction fragments thus obtained with a double-stranded synthetic oligonucleotide adaptor, one end of which is compatible with one or both of the ends of the restriction fragments, to thereby produce adaptor-ligated, restriction fragments of the starting DNA;

(c) contacting the adaptor-ligated, restriction fragments under hybridizing conditions with one or more oligonucleotide primers that contain selective nucleotides at their 3'-end;

IDT Exhibit 1020.5542

(d) amplifying the adaptor-ligated, restriction fragment hybridised with the primers by PCR or a similar technique so as to cause further elongation of the hybridised primers along the restriction fragments of the starting DNA to which the primers hybridised; and

(e) detecting, identifying or recovering the amplified or elongated DNA fragment thus obtained.

AFLP thus provides a reproducible subset of adaptor-ligated fragments. AFLP is described in EP 534858, US 6045994 and in Vos et al 1995. AFLP: a new technique for DNA fingerprinting. Nucleic Acids Research 23(21):4407-4414. Reference is made to these publications for further details regarding AFLP. The AFLP is commonly used as a complexity reduction technique and a DNA fingerprinting technology. Within the context of the use of AFLP as a fingerprinting technology, the concept of an AFLP marker has been developed.

AFLP marker: An AFLP marker is an amplified adaptor-ligated restriction fragment that is different between two samples that have been amplified using AFLP (fingerprinted), using the same set of primers. As such, the presence or absence of this amplified adaptor-ligated restriction fragment can be used as a marker that is linked to a trait or phenotype. With conventional slab gel electrophoresis , an AFLP marker shows up as a band in the gel with a certain mobility. Other electrophoretic techniques such as capillary electrophoresis may not refer to this as a band, but the concept remains the same, i.e. a nucleic acid with a certain length and mobility. Absence or presence of the band may be indicative of (or associated with) the presence or absence of the phenotype. AFLP markers typically involve polymorphisms in the restriction site of the endonuclease (type 1) or the in selective nucleotides (type 2). Occasionally, AFLP markers may involve indels (insertion-deletions in the restriction fragment (type 3).

Selective base or selective nucleotide: Located at the 3' end of the primer that contains a part that is complementary to the adaptor and a part that is complementary to the remains of the restriction site, the selective base is randomly selected from amongst A, C, T or G (or U as the case may be). By extending a primer with a selective base, the subsequent amplification will yield only a reproducible subset of the adaptor- ligated restriction fragments, i.e. only the fragments that can be amplified using the primer carrying the selective base. Selective nucleotides can be added to the 3'end of the primer in a number varying between 1 and 10. Typically, 1-4 suffice. Both primers may contain a varying number of selective bases. With each added selective base, the subset reduces the amount of amplified adaptor-ligated restriction fragments in the subset by a factor of about 4. Typically, the number of selective bases used in AFLP is indicated by +N/+M, wherein one primer carries N selective nucleotides and the other primers carries M selective nucleotides. Thus, an EcoRI/MseI +1/+2 AFLP is shorthand for the digestion of the starting DNA with EcoRI and MseI, ligation of appropriate adaptors and amplification with one primer directed to the EcoRI restricted

IDT Exhibit 1020.5543

position carrying one selective base and the other primer directed to the MseI restricted site carrying 2 selective nucleotides. A primer used in AFLP that carries at least one selective nucleotide at its 3' end is also depicted as an AFLP-primer. Primers that do not carry a selective nucleotide at their 3' end and which in fact are complementary to the adaptor and the remains of the restriction site are sometimes indicated as AFLP+0 primers. The term selective nucleotide is also used for nucleotides of the target sequence that are located adjacent to the adaptor section and that have been identified by the use of selective primer as a consequence of which, the nucleotide has become known.

Sequencing: The term sequencing refers to determining the order of nucleotides (base sequences) in a nucleic acid sample, e.g. DNA or RNA. Many techniques are available such as Sanger sequencing and high-throughput sequencing technologies such as offered by 454 technologies (Roche Applied Science), Illumina Inc. and Applied BioSystems, Helicos and others..

Restriction endonuclease: a restriction endonuclease or restriction enzyme is an enzyme that recognizes a specific nucleotide sequence (target site) in a double-stranded DNA molecule, and will cleave both strands of the DNA molecule at or near every target site, leaving a blunt or a staggered end.

Frequent cutters and rare cutters: Restriction enzymes typically have recognition sequences that vary in number of nucleotides from 3, 4 (such as MseI) to 6 (EcoRI) and even 8 (NotI). The restriction enzymes used can be frequent and rare cutters. The term 'frequent' in this respect is typically used in relation to the term 'rare'. Frequent cutting endonucleases (aka frequent cutters) are restriction endonucleases that have a relatively short recognition sequence. Frequent cutters typically have 3-5 nucleotides that they recognise and subsequently cut. Thus, a frequent cutter on average cuts a DNA sequence every 64-512 nucleotides. Rare cutters are restriction endonucleases that have a relatively long recognition sequence. Rare cutters typically have 6 or more nucleotides that they recognise and subsequently cut. Thus, a rare 6-cutter on average cuts a DNA sequence every 1024 nucleotides, leading to longer fragments. It is observed again that the definition of frequent and rare is relative to each other, meaning that when a 4 bp restriction enzyme, such as MseI, is used in combination with a 5-cutter such as AvaII, AvaII is seen as the rare cutter and MseI as the frequent cutter.

Restriction fragments: the DNA molecules produced by digestion with a restriction endonuclease are referred to as restriction fragments. Any given genome (or nucleic acid, regardless of its origin) will be digested by a particular restriction endonuclease into a discrete set of restriction fragments. The DNA fragments that result from restriction endonuclease cleavage can be further used in a variety of techniques and can for instance be detected by gel electrophoresis.

IDT Exhibit 1020.5544

Ligation: the enzymatic reaction catalyzed by a ligase enzyme in which two double-stranded DNA molecules are covalently joined together is referred to as ligation. In general, both DNA strands are covalently joined together, but it is also possible to prevent the ligation of one of the two strands through chemical or enzymatic modification of one of the ends of the strands. In that case the covalent joining will occur in only one of the two DNA strands.

Synthetic oligonucleotide: single-stranded DNA molecules having preferably from about 10 to about 50 bases, which can be synthesized chemically are referred to as synthetic oligonucleotides. In general, these synthetic DNA molecules are designed to have a unique or desired nucleotide sequence, although it is possible to synthesize families of molecules having related sequences and which have different nucleotide compositions at specific positions within the nucleotide sequence. The term synthetic oligonucleotide will be used to refer to DNA molecules having a designed or desired nucleotide sequence.

Adaptors: short double-stranded DNA molecules with a limited number of base pairs, e.g. about 10 to about 30 base pairs in length, which are designed such that they can be ligated to the ends of restriction fragments. Adaptors are generally composed of two synthetic oligonucleotides which have nucleotide sequences which are partially complementary to each other. When mixing the two synthetic oligonucleotides in solution under appropriate conditions, they will anneal to each other forming a double-stranded structure. After annealing, one end of the adaptor molecule is designed such that it is compatible with the end of a restriction fragment and can be ligated thereto; the other end of the adaptor can be designed so that it cannot be ligated, but this need not be the case (double ligated adaptors).

Adaptor-ligated restriction fragments: restriction fragments that have been capped by adaptors.

Primers: in general, the term primers refer to synthetic DNA molecules which can prime the synthesis of DNA. DNA polymerase cannot synthesize DNA *de novo* without a primer: it can only extend an existing DNA strand in a reaction in which the complementary strand is used as a template to direct the order of nucleotides to be assembled. We will refer to the synthetic oligonucleotide molecules which are used in a polymerase chain reaction (PCR) as primers.

DNA amplification: the term DNA amplification will be typically used to denote the *in vitro* multiplication of DNA molecules, using PCR yielding double stranded DNA or other other amplification methods such as GenomiPhi / RepliG yielding single stranded DNA. It is noted that yet other amplification methods exist and they may be used in the present invention without departing from the gist.

Repetitive sequence: a repetitive sequence is a sequence that is repeated many times, but at least two or more times in a DNA or a genome. Highly-repetitive sequence: a highly-repetitive sequence is a sequence that is repeated typically hundreds to thousands of times

IDT Exhibit 1020.5545

in a DNA or a genome. Examples are Tandem repeats such as satellite DNA, minisatellites and microsatellite and Interspersed repeats such as SINEs (Short INterspersed Elements) and LINEs (Long INterspersed Elements), (retro)transposons.

Low-copy sequence, low copy-number sequence, unique sequence: sequence that occurs only in a limited number of copies (or only in one copy) in a genome and can be used as a marker linked to a phenotype. Low-copy fraction: fraction of a DNA sample that is reduced in repetitive sequences and hence is enriched in low-copy sequences, i.e. a fraction from which repetitive sequences have been removed.

DSN (double strand-specific nuclease): a nuclease that specifically cuts double stranded DNA sequences (Shagin et al 2002 Genome Research 12 1935-1945, Zhulidov et al 2004 NAR 32 3 e37).

Renaturation kinetics based fractionation: a technique wherein low-copy sequences are separated from high-copy or repetitive sequences in a DNA sample . The procedure involves heating a sample of genomic DNA until it denatures into the single stranded-form, and then slowly cooling it, so the strands can pair back together. While the sample is cooling, measurements can be taken of how much of the DNA is base paired at each temperature.

The amount of single and double-stranded DNA can be measured by rapidly diluting the sample, which slows reassociation, and then binding the DNA to a hydroxylapatite column. The column can be eluted to elute the single-stranded DNA first and followed by elution of the double stranded DNA. The amount of DNA in these two solutions can be measured using a spectrophotometer. Since a sequence of single-stranded DNA needs to find its complementary strand to reform a double helix, common sequences renature more rapidly than rare sequences. Indeed, the rate at which a sequence will reassociate is proportional to the number of copies of that sequence in the DNA sample. A sample with a highly-repetitive sequence will renature rapidly, while complex sequences will renature slowly.

However, instead of simply measuring the percentage of double-stranded DNA versus time, the amount of renaturation is usually measured relative to a C_0t value. The C_0t value is the product of C_0 (the initial concentration of DNA), t (time in seconds), and a constant that depends on the concentration of cations in the buffer. Repetitive DNA will renature at low C_0t values, while complex and unique DNA sequences will renature at high C_0t values.

C_0t filtration: C_0t filtration is a technique that uses the principles of DNA renaturation kinetics to separate the repetitive DNA sequences that dominate many eukaryotic genomes from "gene-rich" single/low-copy sequences. This allows DNA sequencing to concentrate on the parts of the genome that are most informative and interesting, which will speed up the discovery of new genes and make the process more efficient (Peterson DG, Wessler SR, Paterson AH (2002). "Efficient capture of unique sequences from eukaryotic genomes". Trends Genet. 18 (11): 547–50, LLamoureux D, Peterson DG, Li W, Fellers JP, Gill BS

IDT Exhibit 1020.5546

(2005). "The efficacy of C_0t-based gene enrichment in wheat (Triticum aestivum L.)". Genome 48 (6): 1120–6. , Yuan Y, SanMiguel PJ, Bennetzen JL (2003). "High-C_0t sequence analysis of the maize genome". Plant J. 34 (2): 249–55.

Detailed description of the invention

Thus, in a first embodiment the invention relates to a method for the reduction of repetitive sequences (enhancing the (relative) amount of low-copy sequences) in adapter-ligated restriction fragments comprising the steps of:

a. restricting a starting DNA in a sample with a first restriction endonuclease and ligating a first adapter to the restriction fragments to obtain adapter-ligated restriction fragments;
b. performing a renaturation kinetics based fractionation on the adapter-ligated restriction fragments obtained in step (a);
c. subjecting the renatured fractions of step (b) to a duplex specific nuclease (DSN);
d. restricting the restriction fragments obtained from step (c) with a second restriction endonuclease and ligating a second adapter to the restriction fragments;
e. obtaining a low-copy enriched fraction of the starting DNA.

The method according to the invention combines certain elements of AFLP (creation of adapter- ligated restriction fragments using a restriction enzyme and ligation of adapters) with certain elements of renaturation kinetics based fracturation (allowing high copy sequences to renature to become double stranded but low copy sequences to remain single-stranded), removing the double stranded sequences by DSN and, again AFLP (further creation of adapter-ligated restriction fragments using a second restriction enzyme. The results is that a fraction of the starting DNA is obtained that is enriched for the presence of low-copy sequences and reduced in the number of highly repetitive sequences.

In step (a) of the method according to the invention, a starting DNA is treated with a restriction endonuclease. The starting DNA can be any DNA such as genomic DNA, cDNA, BAC DNA, mitochondrial DNA, chloroplast DNA or mixtures of plant genomic DNA containing mitochondrial and chloroplast DNA . The first restriction endonuclease digests the starting DNA to yield restriction fragments. The first restriction endonuclease is preferably a rare cutter, such as a hexacutter, such as EcoRI, EcoRII, BamHI, HindIII, PstI leading to restriction fragments which typically range in size between 1-4 kb in case an AT-rich endonuclease sich as EcoRI (recognition sequence GAATTC) is applied in and AT-rich organism such as most plant species. To these restriction fragments, first adapters are ligated to yield adapter-ligated restriction fragments.

IDT Exhibit 1020.5547

In step (b) of the method according to the invention, the adapter-ligated restriction fragments, are subjected to a renaturation kinetics-based fractionation. Thus, the adapter-ligated restriction fragments are heated until denaturation into the single stranded-form, and then slowly cooled down, so the strands can pair back together. As the adapter-ligated restriction fragments are typically within the same length range due to the use of the resection endonuclease, the repetitive sequences will re-anneal within a determinable amount of time while the low-copy sequences will remain single stranded for a longer period of time. After cooling has proceeded long enough, the re-annealing is stopped, for instance by forced cooling. In renaturation kinetics-based fractionation disclosed in the art, the amount of single and double-stranded DNA is usually measured by rapidly diluting the sample, which slows reassociation, and then binding the DNA to a hydroxylapatite column after which the single stranded and double stranded fractions can be eluted.

In a preferred embodiment of the invention, the use of the column and the intermediate separation into fractions that essentially contain single stranded DNA or double stranded DNA, respectively, can be avoided by (directly) treating the sample with a double stranded nuclease in step (c) of the method. The double stranded nuclease specifically cuts double stranded DNA sequences (Shagin et al 2002 Genome Research 12 1935-1945, Zhulidov et al 2004 NAR 32 3 e37). As most of the repetitive sequences are now in double stranded form, the sample is reduced for adapter-ligated restriction fragments that are derived from repetitive sequences (or enriched for adapter-ligated restriction fragments that are derived from low-copy or unique sequences).

The renatured fractions of step (b), whether or not DSN treated in optional step (c), are subjected in step (d) of the method of the invention to a second digestion with a restriction endonuclease, followed by ligation of the adapter corresponding to the endonuclease. The resulting adapter-ligated restriction fragments, that, in a preferred embodiment, on each side of the restriction fragment carry a different adapter. Thus the result of the method of the current invention is that a set of adapter-ligated restriction fragments is obtained that is enriched for low-copy DNA (or reduced significantly in relative occurrence of repetitive DNA).

The obtained adapter-ligated restriction fragments can be used in subsequent fingerprinting technologies such as AFLP, for SNP mining etc.. Thus in one embodiment, the obtained low-copy DNA sample can be selectively amplified and fingerprinted to obtain higher quality fingerprints.

The C_0t methodology is in itself well known for the normalisation of cDNA libraries, also in combination with DSN. Typically, C_0t-fragments are obtained by shearing or otherwise by fragmenting DNA in a more or less random manner. Prior to C_0t analysis, fragments are typically size selected to reduce the effect that larger fragments renature quicker than shorter fragments, which could be detrimental to the desired result. By now performing a first

IDT Exhibit 1020.5548

endonuclease restriction and adapter ligation before C_ot treatment and further fragmentation, the size of the fragments are more manageable and generally lie within a more limited size range. On the other hand, performing endonuclease restriction using two enzymes (and hence two adapters), would lead to a too large spread of lengths (in the area of 900-50 bp). By executing the method with a first restriction step, then a C_ot treatment and a subsequent restriction step, the C_ot treatment is performed on a set of adapter-ligated restriction fragments that have renaturation kinetics that are more homogeneous than those of fragment mixtures containing a wide range of large and small adapter-ligated fragments.

The method of the present invention is particularly useful to increase the success rate of conversion of AFLP markers (or other markers) to, for instance PCR markers and in the discovery of useful SNPs which can be turned into assays applicable on whole genomic DNA.

Description of the Figures

Figure 1 Pepper results for Pepper E33/M49 and Pepper P14/M60 are

Figure 2 Maize results for Maize E35/M48 and Maize P12/M50

Figure 3 C_ot-DSN Protocol

Examples

The invention will now be explained by the following examples. It will be clear to the skilled man that variations on this protocol are possible without departing from the gist of the invention,

1. Agarose gel

For DNA quality/integrity check, load 5µl DNA (RNAse treated!) on a 1% agarose gel.

2. Nanodrop quantification

For accurate quantification of DNA, dry and resuspend/ Speedvacuum concentrate the DNA to 1000 ng/µL if necessary (minimum concentration of 114 ng/µL)

3. EcoRI DNA digestion (60 µg DNA)

Reagent	Conc.	Sample
60 µg DNA	1000 ng/µL	60 µL
EcoRI	50 U/µL	6 µL
One-Phor-All	10x	60 µL
BSA	10 mg/µL	3 µL
DTT	1 M	3 µL
Milli Q water		468 µL
Total volume		600 µL

Treat samples at 37°C for 6 hours.

- Use 5U of EcoRI per µg DNA [standard AFLP protocol requires 25U/µg DNA]

IDT Exhibit 1020.5549

- Amount of DNA can vary from a minimum of 20 µg to more than 100 µg (scalable)
- Size-selection step between restriction and ligation is not performed

4. EcoRI adapter ligation

Reagent	Conc.	Sample
EcoRI Adaptor	5 µM	300
One-Phor-All 10x		100
BSA	10 mg/µL	2
DTT	1 M	2
ATP	10 mM	300
T4 ligase	1 U/µL	300
Total volume		1004

The 1004 µL are added to the restricted 600 µL and treated at 37°C overnight.

- Use 40 ng DNA/pmole adaptor [like in the standard AFLP protocol]

5. Agarose gel

Check quality of the EcoRI digestion. Expect a smear between 1-10 kb with highest intensity at 2-3 kb.

6. Size-selection

Treat samples at 80°C for 20 min. in a water-bath.

7. Fill-in adaptors (ds adaptors synthesis)

In order to make ds-adaptor the following 300 µL solution is added to each tube containing 1.604mL R/L

Reagent	Conc.	Sample
10xPCR buffer ROCHE		190
dNTPs	5 mM	70
Taq Polymerase ROCHE	5 U/µL	7
Milli Q water		33
Total volume (1x PCR)		300

Treat samples (1.904 mL) at 72°C for 30 min. in a water-bath, then chill tubes on ice and EtOH precipitate.

8. DNA EtOH precipitation of ds-RL samples

Transfer the 1.904 mL R/L in a 50 mL tube (blue cap, V bottom).

Add the same volume (1.9 mL) of 7.5 M Ammonium Acetate and mix.

Add three volumes (11.4 mL) of 100% ethanol and let DNA precipitate at -20°C overnight.

Centrifuge samples at 4800 rpm (large rotor) for 20 min. at 4°C.

Wash pellet with 1.0 mL cold 70% ethanol and carefully (!) transfer the pellet into a 1.5 mL eppendorf tube.

IDT Exhibit 1020.5550

Centrifuge tubes at 14000 rpm (micro centrifuge) for 20 minutes at 4°C.

Carefully discard supernatant and let EtOH evaporate from pellets at 37°C for 15-20 minutes.

9. DNA resuspension

Re-suspend the pellet in 100 µL Milli Q water or EB buffer (QIAGEN) leaving tubes at RT for 3 hours or at 60°C for 20 min. and mixing slowly by pipetting.

10. Ampure Bead Purification of fragments > 300 bp (See Ampure Protocol)

At the 100 µL of the resuspended samples add 400 µL EB buffer (QIAGEN) and mix.

Add 350µL of Ampure beads and vortex (Ratio Beads/Sample = 0.7 to have cutoff ~ 300 bp).

Incubate samples for 5 min at RT.

Insert tubes in the (Magnetic Particle Concentrator) MPC for 5 min. to precipitate the beads.

Remove supernatant by pipetting (leave tubes in the MPC).

Keep tubes in the MPC and wash beads with 1 mL 70% EtOH and discard supernatant (twice).

Dry pellets at 37°C for 10-15 minutes.

Remove tubes from the MPC and elute DNA with 200 µL Milli Q water (twice) by adding Milli Q water, vortexing/pipetting to resuspend all beads and leaving tubes in MPC for 5 min for beads precipitation.

Keep the supernatant (with DNA!) and place it in a new 1.5 mL tube (400 µL/sample).

11. Agarose gel

Check quality of purification (load 2µL on a 1% agarose gel).

12. Nanodrop quantification

For accurate DNA quantification.

13. SpeedVac concentration to 1000 ng/µL

Samples are concentrated to obtain ds-R/L purified DNA at a concentration of 1000 ng/µL.

(Alternatively dry the sample and re-suspend it in the correct volume of Milli Q water).

14. C_0t treatment

C_0t treatment is performed at the following conditions:

1. 500 ng/µL DNA
2. 0.3 M NaCl
3. renaturation at 63□C

At these conditions C_0t-320 takes 60 h 16m

Hybridization is performed as follow (in 200µL tubes) :

Sample	
DNA (at 1000 ng/µL)	2 µL
2x C_0t Hybridization Buffer (0.6 M NaCl)	
Note: composition - 30% of 4x trimmer hybridization buffer at 2 M NaCl	
- 70% MQW	2 µL

IDT Exhibit 1020.5551

Total volume 4 µL

Cover the 4 µL sample with 20 µL mineral oil and centrifuge (make sure the 4 µL are covered by oil).

Treat samples at 95°C for 10 min (PCR thermocycler) and let hybridize at 63°C for the time required for the C_0t treatment (C_0t-320 = 60h 16m).

15. DSN treatment

Preheat the DSN Master Buffer at 63°C for >5 min.

C_0t treatment is stopped (after 60h 16m 00s) placing the sample on ice for 30 seconds.

Place samples at 63°C for 1 min.

Add 5 µL of preheated DSN Master Buffer to the 4 µL of renaturating solutions, mix quickly by pipetting

and incubate at 63°C for additional 10 min.

Add 1 µL DSN Enzyme, mix quickly by pipetting and centrifuge (make sure all the 10 µL are on the bottom of the tube) and treat at 68°C for 25 min.

Add 10µL of Stop Solution, mix and centrifuge and keep at 68°C for additional 5 min before placing tubes on ice.

Add 20 µL sterile water [total volume after DSN treatment = 40 µL] and purify the sample.

16. Qiagen purification

Follow Qiagen purification instruction and elute in 150 µL Milli Q water (three 50 µL elutions).

17. SpeedVac concentration

Concentrate the 150µL to 31 µL (~40 min. at 35°C). Adjust volume to 31 µL if necessary.

18. EcoRI+0 ds-synthesis of the DSN-treated ss-DNA samples using Advantage 2 PCR kit

30 µL (of the 31 µL eluted from Qiagen purification) is +0 amplified to make ds-DNA as follow.

The remaining 1 µL (Not +0 amplified) will be used for agarose analysis

Sample	
Purified ss DNA from DSN treatment	30
EcoRI+0 E00L (50 ng/µL)	1.5
10x Ad2 buffer	5
50x dNTP	1
50x Ad2 Polymerase mix	1
MQW	11.5
total	50

The samples are treated at 94°C for 30 sec. then 24 cycles at 94°C for 7 sec., 56°C for 30 sec. and 72°C for 10 min., with last step at 72°C for 10 min and stored at 4°C.

19. Agarose gel

Load 1µL of "Not +0 amplified" and 1.6 µL "+0 amplified" for each sample.

IDT Exhibit 1020.5552

20. Qiagen purification

Follow Qiagen purification instruction and elute in 150 µL Milli Q water (three 50 µL elutions).

21. Speedvacuum concentration

Concentrate the 150µL to 40-50 µL (minimum DNA concentration of 14 ng/µL).

22. Nanodrop quantification

For accurate DNA quantification.

23. MseI Restriction (Standard AFLP protocol)

Reagent	Conc.	Sample
C_ot treated DNA (500 ng) + Milli Q water		35.6
MseI	40 U/µL	0.03
One-Phor-All	10x	4
BSA (µL)	10 mg/µL	0.2
DTT (µL)	1 M	0.2
Total volume		40

Samples are treated at 37°C for 1.0 h

Size-selection step between restriction and ligation is not performed

24. MseI adapter Ligation (Standard AFLP protocol)

Reagent	Conc.	Sample
MseI Adapter	50 µM	1
One-Phor-All	10x	1
BSA	10 mg/µL	0.05
DTT	1 M	0.05
ATP	10 mM	1
T4 ligase	1 U/µL	1
MilliQ Water		5.9
Total volume		10

The 10 µL are added to the restricted 40 µL and treated at 37°C overnight.

Starting from the C_ot-treated R/L samples, the standard protocols (i.e. for AFLP, CRoPS, etc. can be followed)

The reduction of repetitive sequences was performed by selecting samples from pepper and maize (both highly repeated genomes). The composition of convention R/L mixes (restriction-ligation mixtures) prior to any C_ot treatment was

EcoRI/MseI	100-1000 bp
PstI/MseI	100-1000 bp
MseI-only with partial digestion	1000 bp average
EcoRI-only	2000-3000 bp average
PstI-only	4000-5000 bp average

IDT Exhibit 1020.5553

C_0t parameter ranging from 0-320 (0, 1, 2, 5, 10, 20, 40, 80, 160, 320)

Pepper results for Pepper E33/M49 and Pepper P14/M60 are displayed in Fig 1.

The various restriction enzyme digests and C_0t denaturation times were tested on two pepper samples that were restricted with EcoRI (pepper E33/M49) and amplified using primer E33 (E-AAG) and M49 (M-CAG) and with PstI (pepper P14/M60) and amplified using primer P14 (P-AT) and M60 (M-CTC). From Fig 1 it is clear that repetitive forms are reduced and some hitherto undiscovered low-copy bands appear

Maize results for Maize E35/M48 and Maize P12/M50 are displayed in Fig 2.

The various restriction enzyme digests and C_0t denaturation times were tested on two Maize samples that were restricted with EcoRI (Maize E35/M48) and amplified using primer E35 (E-ACA) and M48 (M-CAC) and with PstI (Maize P12/M50) and amplified using primer P12 (P-AC) and M50 (M-CAT). From Fig 2 it is clear that repetitive forms are reduced and some hitherto undiscovered low-copy bands appear

From these results it was concluded that

Pre-treatment	C_0t effect
EcoRI/MseI digestion	No
PstI/MseI digestion	No
MseI-partial	No
EcoRI only	Yes
Pst-only	Yes
C_0t time	320

	No. Bands AFLP std.	No. bandsat C_0t 320	New bandsat C_0t 320	% Reduction of bands by C_0t treatment	No. variable bands by C_0t treatment(variable intensity)
Pepper E35/M48	82	34	0 (0%)	58%	9 (27%)
Pepper E33/M49	81	49	3 (4%)	43%	16 (33%)
Maize E35/M48	39	30	3 (8%)	31%	9 (30%)
Maize E33/M49	30	18	2 (7%)	47%	3 (17%)
***C_0t* effect**			**~ 5%**	**40 to 50%**	**20 to 30%**

IDT Exhibit 1020.5554

These results indicate that the C_0t treatment determined the disappearance of 40-50 % of the standard AFLP bands and the appearance of 5% on new bands. A total of 20-30 % of the bands that remained after the C_0t treatment presented a variation (increased or decreased) of band intensity.

Validation on qRT-PCR Pepper

	STS single copy gene	Ribosomal repetitive	Chloroplast repetitive
gDNA	1	47.8	45.1
R/L before treatment (EcoRI-only)	1	49.8	43.8
Normalized gDNA - No C_0t	1	60.5	67.4
Normalized gDNA - C_0t 320	1	13.6	23.7
Cot effect		- 4.5 fold	- 2.9 fold
R/L post treatment - No C_0t	1	13.5	34.1
R/L post treatment - C_0t 320	1	5.4	5
Cot effect		- 2.5 fold	- 6.8 fold
+1/+2 - No C_0t	1	67.0	115.9
+1/+2 - C_0t 320	1	16.3	41.8
C_0t effect		- 4.1 fold	- 2.8 fold

From these results is concluded that sequences that are ~ 40 fold more abundant than single copy genes are reduced ~ 4 fold by the C_0t treatment (40:1 to 10:1)

Further validation is achieved by Sanger sequencing.

IDT Exhibit 1020.5555

CLAIMS

1. Method for the reduction of repetitive sequences (enhancing the (relative) amount of low-copy sequences) in adapter-ligated restriction fragments comprising the steps of:
   a. restricting a starting DNA in a sample with a first restriction endonuclease and ligating a first adapter to the restriction fragments to obtain adapter-ligated restriction fragments;
   b. performing a renaturation kinetics based fractionation on the adapter-ligated restriction fragments obtained in step (a);
   c. optionally, subjecting the renatured fractions of step (b) to a duplex specific nuclease (DSN);
   d. restricting the restriction fragments obtained from step (b) or (c) with a second restriction endonuclease and ligating a second adapter to the restriction fragments;
   e. obtaining a low-copy enriched fraction of adapter-ligated restriction fragments from the starting DNA.

2. Method according to claim 1, wherein the first restriction endonuclease is a rare cutter.

3. Method according to claim 1 or 2, wherein the second restriction endonuclease is a frequent cutter.

4. Method according to claims 1-3, wherein the renaturation kinetics based fractionation is C_0t

5. Method according to claim 4 wherein the C_0t value is 320

6. Method according to claims 1-5, wherein the low-copy fraction of the adaptor-ligated restriction fragments obtained from starting DNA is amplified using primers directed to the adapters having optionally 1-10 selective nucleotides at the 3'end

7. Method according to any of the previous claims, wherein the starting DNA is genomic DNA, cDNA, BAC DNA, mitochondrial DNA, chloroplast DNA or mixtures of plant genomic DNA with mitochondrial DNA and chloroplast DNA

IDT Exhibit 1020.5556

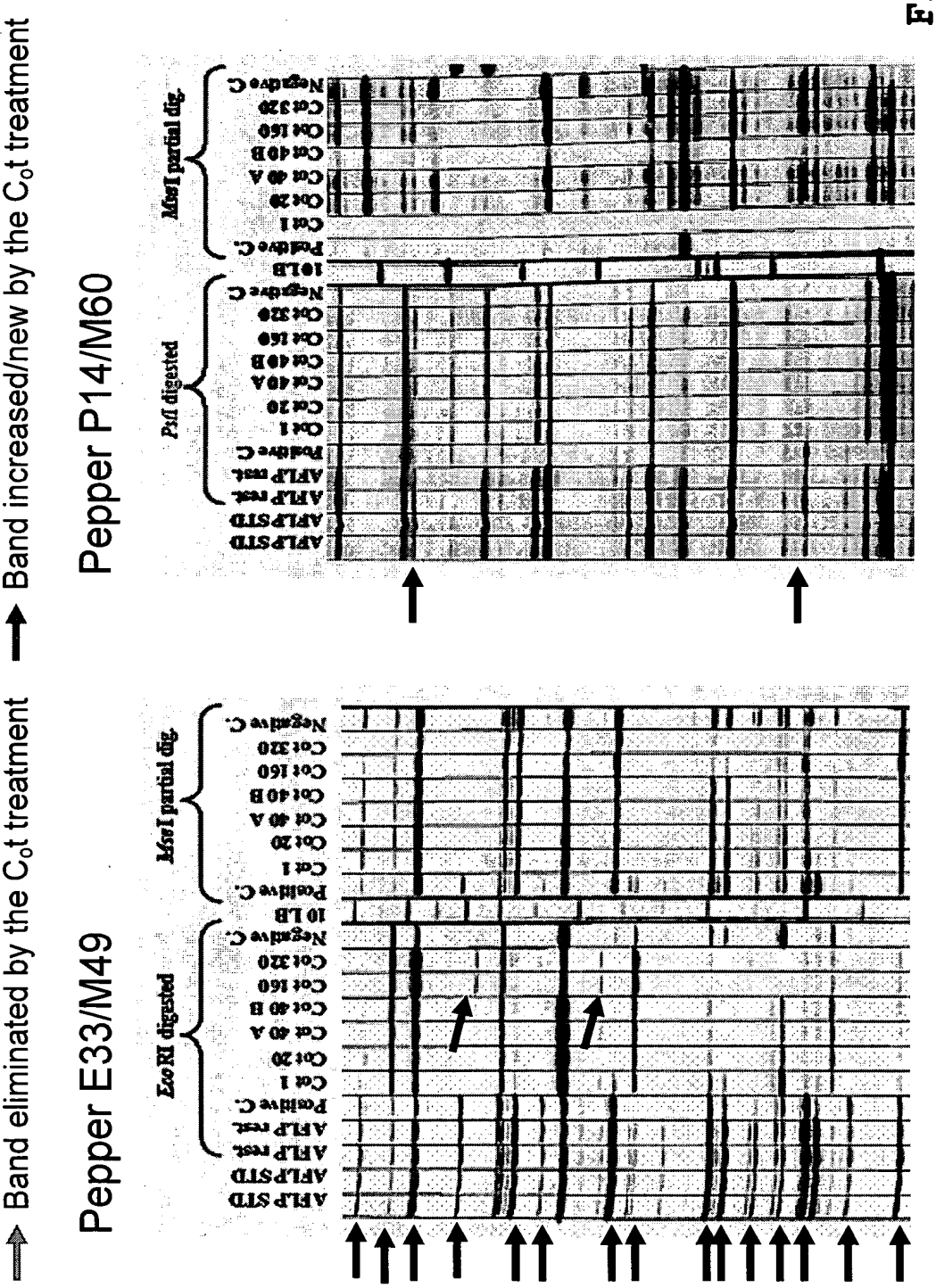
8. Method according to any of the previous claims, wherein DSN normalisation is performed using a treatment comprising a duplex specific nuclease.

9. Use of the method as defined in claims 1-8 to generate AFLP fingerprints enriched for low-copy sequences.

10. Use of the method as defined in claims 1-8 as a treatment of a DNA sample prior to marker conversion or SNP mining.

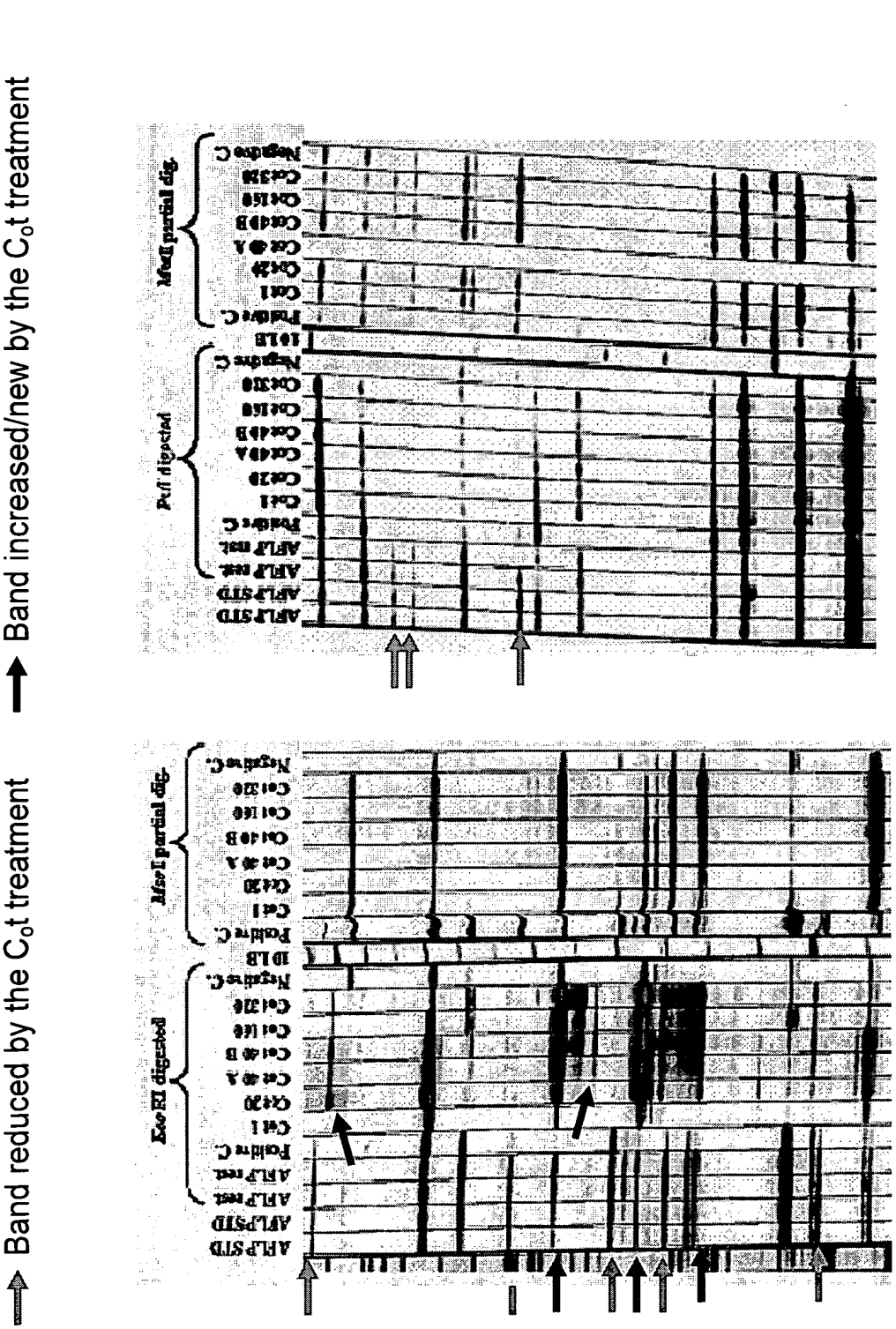
IDT Exhibit 1020.5557

Fig 1



IDT Exhibit 1020.5558

Fig 2



IDT Exhibit 1020.5559

C_0t-DSN Protocol –

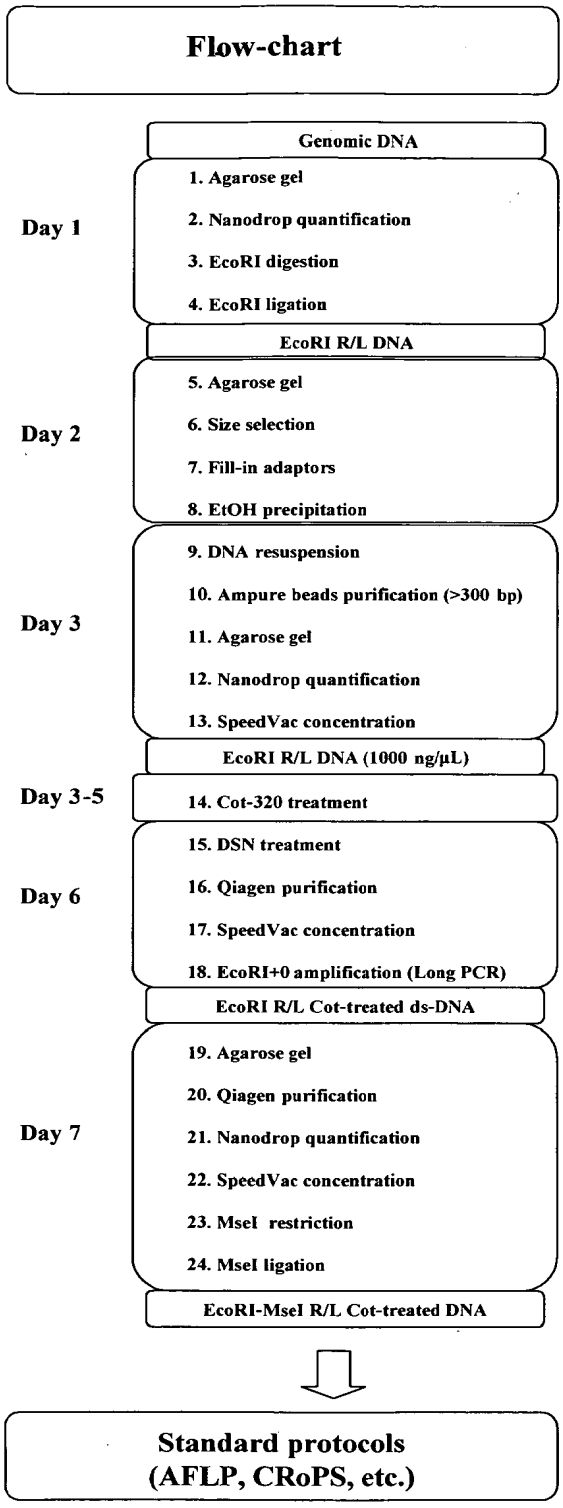


Fig 3

IDT Exhibit 1020.5560

INTERNATIONAL SEARCH REPORT

International application No
PCT/NL2009/000232

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12N15/00

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)
C12N

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

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

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	MEUDT ET AL: "Almost Forgotten or Latest Practice? AFLP applications, analyses and advances" TRENDS IN PLANT SCIENCE, ELSEVIER SCIENCE, OXFORD, GB, vol. 12, no. 3, 16 March 2007 (2007-03-16) , pages 106-117, XP005920449 ISSN: 1360-1385 the whole document	1-10
Y	EP 0 534 858 A1 (KEYGENE NV [NL]) 31 March 1993 (1993-03-31) the whole document	1-10
Y	WO 2008/070223 A2 (CYTOTREND BIOTECH ENGINEERING [US]; CHEN ZHONG [US]; LIU NING [CN]) 12 June 2008 (2008-06-12) the whole document	1-10

-/--

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

[X] See patent family annex.

* Special categories of cited documents :

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

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

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

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

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

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

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

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

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
4 February 2010	05/03/2010

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

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

IDT Exhibit 1020.5561

INTERNATIONAL SEARCH REPORT

International application No
PCT/NL2009/000232

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WO 01/06014 A1 (AMERICAN LAB TECHNOLOGIES INC [US]) 25 January 2001 (2001-01-25) the whole document	1-10
Y	WO 2008/119084 A1 (CHILDREN S MERCY HOSPITAL [US]; NEWKIRK HEATHER [US]; BI CHENGPENG [US) 2 October 2008 (2008-10-02) the whole document	1-10
A	PETERSON D G ET AL: "Efficient capture of unique sequences from eukaryotic genomes" TRENDS IN GENETICS, ELSEVIER SCIENCE PUBLISHERS B.V. AMSTERDAM, NL, vol. 18, no. 11, 1 November 2002 (2002-11-01), pages 547-550, XP004389596 ISSN: 0168-9525 the whole document	1-10

IDT Exhibit 1020.5562

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/NL2009/000232

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 0534858 A1	31-03-1993	AT 191510 T	15-04-2000
		AT 382094 T	15-01-2008
		AU 672760 B2	17-10-1996
		CA 2119557 A1	01-04-1993
		CZ 9400669 A3	15-12-1994
		CZ 291877 B6	18-06-2003
		CZ 298187 B6	18-07-2007
		DE 69230873 D1	11-05-2000
		DE 69230873 T2	09-11-2000
		DE 69233719 T2	02-01-2009
		DK 0534858 T3	11-09-2000
		DK 0969102 T3	13-05-2008
		WO 9306239 A1	01-04-1993
		EP 0969102 A2	05-01-2000
		ES 2147550 T3	16-09-2000
		ES 2299229 T3	16-05-2008
		FI 941360 A	24-05-1994
		FI 20031526 A	17-10-2003
		GR 3033895 T3	30-11-2000
		HU 68504 A2	28-06-1995
		IL 103267 A	25-07-2004
		JP 6510668 T	01-12-1994
		JP 3236295 B2	10-12-2001
		JP 4088403 B2	21-05-2008
		JP 2001061486 A	13-03-2001
		JP 4385072 B2	16-12-2009
		JP 2008131947 A	12-06-2008
		NO 941064 A	20-05-1994
		PT 534858 E	29-09-2000
		PT 969102 E	25-03-2008
		RU 2182176 C2	10-05-2002
		ZA 9207323 A	30-08-1993
WO 2008070223 A2	12-06-2008	US 2010021971 A1	28-01-2010
WO 0106014 A1	25-01-2001	AU 6350800 A	05-02-2001
WO 2008119084 A1	02-10-2008	EP 2129800 A1	09-12-2009
		US 2008274558 A1	06-11-2008

IDT Exhibit 1020.5563

Espacenet

Bibliographic data: WO2011019964 (A1) — 2011-02-17

METHODS, COMPOSITIONS, AND KITS FOR GENERATING NUCLEIC ACID PRODUCTS SUBSTANTIALLY FREE OF TEMPLATE NUCLEIC ACID

Inventor(s): KURN NURITH [US]; WANG SHENGLONG [US] ± (KURN, NURITH, ; WANG, SHENGLONG)

Applicant(s): NUGEN TECHNOLOGIES INC [US]; KURN NURITH [US]; WANG SHENGLONG [US] ± (NUGEN TECHNOLOGIES, INC, ; KURN, NURITH, ; WANG, SHENGLONG)

Classification: - international: *C12P19/34*
- cooperative: **C12P19/34**; **C12Q1/6806**; **C12Q1/6853** more

Application number: WO2010US45384 20100812 Global Dossier

Priority number (s): US20090242706P 20090915 ; US20090239749P 20090903 ; US20090233441P 20090812

Also published as: WO2011019964 (A9) EP2464738 (A1) EP2464738 (A4) US2011224105 (A1)

Abstract of WO2011019964 (A1)

Methods, kits, and compositions which enable the generation of double stranded DNA products from various substrates such as low abundance RNAs or whole transcriptosomes are provided. The double stranded DNA products are suitable for downstream applications involving microarray analysis or sequencing.

IDT Exhibit 1020.5564

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
17 February 2011 (17.02.2011)

PCT

(10) International Publication Number
WO 2011/019964 A1

(51) International Patent Classification:
C12P 19/34 (2006.01)

(21) International Application Number:
PCT/US2010/045384

(22) International Filing Date:
12 August 2010 (12.08.2010)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

61/233,441	12 August 2009 (12.08.2009)	US
61/239,749	3 September 2009 (03.09.2009)	US
61/242,706	15 September 2009 (15.09.2009)	US

(71) Applicant *(for all designated States except US)*: **NUGEN TECHNOLOGIES, INC.** [US/US]; 821 Industrial Road, 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). **WANG, Shenglong** [CA/US]; 8248 Saturn Park Drive, San Ramon, CA 94582 (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, 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 2011/019964 A1

(54) **Title**: METHODS, COMPOSITIONS, AND KITS FOR GENERATING NUCLEIC ACID PRODUCTS SUBSTANTIALLY FREE OF TEMPLATE NUCLEIC ACID

(57) **Abstract**: Methods, kits, and compositions which enable the generation of double stranded DNA products from various substrates such as low abundance RNAs or whole transcriptosomes are provided. The double stranded DNA products are suitable for downstream applications involving microarray analysis or sequencing.

IDT Exhibit 1020.5565

METHODS, COMPOSITIONS, AND KITS FOR GENERATING NUCLEIC ACID PRODUCTS SUBSTANTIALLY FREE OF TEMPLATE NUCLEIC ACID

CROSS-REFERENCE

[0001] This application claims the benefit of U.S. Provisional Application No. 61/233,441, filed August 12, 2009; U.S. Provisional Application No. 61/239,749, filed September 3, 2009; and U.S. Provisional Application No. 61/242,706, filed September 15, 2009, which applications are incorporated herein by reference in their entireties.

BACKGROUND OF THE INVENTION

[0002] Large scale analysis of nucleic acids often requires amplification and further generation of products such as single stranded DNA (ss DNA) or double stranded DNA (ds DNA) that are either labeled, ready for labeling, or ready for ligation of desired adapters. These large scale analysis methods are evolving and comprise the use of various high or low density microarrays, high throughput sequencing, and other high throughput or highly parallel techniques. The limitations of current nucleic acid preparation techniques impact the ability to carry out large scale analysis of multiple parameters, as is required for, for example, the genotyping of multiple loci in the study of complex diseases, detecting the presence or absence of specific nucleic acid species in a sample, large scale sequencing and the like. Moreover, it is well accepted that molecular analysis determination of genomic instability in various pathological conditions such as cancer is most precisely carried out in well defined cell populations, such as those obtained by laser capture micro-dissection or cell sorting. Nucleic acid preparation technologies that provide robust amplification of very small polynucleotide samples, for example, from one or a very few cells, may provide a solution to the limited starting materials generally available for analysis.

[0003] Likewise, the ability to amplify ribonucleic acid (RNA) is an important aspect of efforts to elucidate biological processes. Total cellular mRNA represents gene expression activity at a defined time. Gene expression is affected by cell cycle progression, developmental regulation, response to internal and external stimuli and the like. The profile of expressed genes for any cell type in an organism reflects normal or disease states, response to various stimuli, developmental stages, cell differentiation, and the like. Non-coding RNAs have been shown to be of great importance in regulation of various cellular functions and in certain disease pathologies. Such RNAs are often present in very low levels. Thus, amplification methods capable of amplifying low abundance RNAs, as well as methods capable of amplifying whole transcriptome RNA are of great importance.

[0004] Therefore, there is a need for improved methods of preparing DNA and RNA samples, including methods which can globally or specifically amplify DNA or RNA polynucleotide targets. The invention described herein fulfills this need.

SUMMARY OF THE INVENTION

[0005] Accordingly, the present invention provides improved methods for generation of ds DNA fragment molecules suitable for downstream applications including but not limited to large scale analysis.

[0006] In one embodiment, the present invention provides a method comprising: a) providing an input nucleic acid template (e.g. DNA or RNA); b) hybridizing one or more oligonucleotide primers to the input nucleic acid template; c) extending the one or more oligonucleotide primers along the input nucleic acid template with a polymerase comprising strand displacement activity, wherein the hybridizing and extending produces primer extension products comprising one or more double stranded products; and d)cleaving the input nucleic acid template with an agent comprising one or more cleavage reagents. In one aspect, steps (b) and (c) are performed simultaneously. In another

IDT Exhibit 1020.5566

aspect, steps (b) and (c) are performed sequentially. In a specific embodiment, the method further comprises generating one or more blunt ended double stranded products from the one or more double stranded products. The blunt ended double stranded products can be generated using enzymatic or non enzymatic methods. In the specific embodiment where enzymatic methods are utilized, the enzyme can be an exonuclease, an endonuclease, or a combination thereof. In the specific embodiment where the enzyme is an endonuclease, the endonuclease can be, but is not limited to, an S1 endonuclease, mung bean endonuclease, or a combination thereof. Such blunt ended double stranded products created using the methods provided herein, can be used to generate a library of blunt ended double stranded products. Such blunt ended double stranded products can be analyzed by next generation sequencing.

[0007] In a related embodiment, the present invention provides a method comprising: a) providing an input nucleic acid template (e.g., DNA or RNA) comprising one or more non-canonical nucleotides (e.g., dUTP) in a reaction mixture; b) hybridizing one or more oligonucleotide primers to the input nucleic acid template; c) extending the one or more oligonucleotide primers along the input nucleic acid template with a polymerase comprising strand displacement activity, wherein the hybridizing and extending produces primer extension products comprising one or more double stranded products; and d) cleaving the input nucleic acid template with an agent comprising one or more cleavage reagents. In one aspect, steps (b) and (c) are performed simultaneously. In another aspect, steps (b) and (c) are performed sequentially.

[0008] In a separate but related embodiment, the present invention provides a method comprising: a) contacting an input template nucleic acid (e.g., DNA or RNA) comprising one or more non-canonical nucleotides (e.g., dUTP) in a reaction mixture comprising: i) one or more oligonucleotide primers; ii) a polymerase comprising strand displacement activity; iii) an agent capable of cleaving a base portion of a non-canonical nucleotide, whereby an abasic site is generated; and iv) an agent capable of fragmenting a phosphodiester backbone at the abasic site;whereby double stranded DNA fragments are generated.

[0009] In a separate but related embodiment, the present invention provides a method comprising: a) providing an input nucleic acid template comprising one or more non-canonical nucleotides (e.g., dUTP); b) amplifying the input nucleic acid template to produce amplification products in a reaction mixture comprising oligonucleotide primers comprising random hybridizing portions, and an enzyme comprising strand displacement activity; and c) fragmenting the input nucleic acid template, wherein the fragmenting step is performed by adding to the reaction mixture comprising amplification products an agent that cleaves a base portion of a non-canonical nucleotide to generate an abasic site and an agent that cleaves a phosphodiester backbone of a nucleic acid at an abasic site. In one aspect of the embodiment, the method may further comprise d) optionally separating the amplification products from the fragmentation products; e) treating the amplification products with one or more agents to produce double stranded blunt-ended amplification products; and f) phosphorylating the 5' ends of the double stranded blunt-ended amplification products. In one further aspect, step (d) is performed after step (f). In another further aspect, step (d) is performed before step (e). In yet another further aspect, step (d) is not performed. In another aspect of the embodiment, the method further comprises ligating the one or more phosphorylated double stranded blunt-ended amplification products with adaptor nucleic acid molecules. In some cases, the amplification products are analyzed by next generation sequencing.

[0010] In one aspect of any one of the foregoing embodiments, the present invention provides for cleaving a base portion of the non-canonical nucleotide, thereby forming an abasic site. In some cases the agent comprises a glycosylase such as UNG or UDG. In some cases, the agent comprises a primary amine, a polyamine, or DMED. In some cases, the agent is a glycosylase and a polyamine.

IDT Exhibit 1020.5567

[0011] In one aspect of any one of the foregoing embodiments, the present invention provides a method wherein the one or more oligonucleotide primers comprise a label. For example, the one or more oligonucleotide primers may comprise an amino-allyl label.

[0012] In another aspect of any one of the foregoing embodiments, the one or more oligonucleotide primers comprise a random hexamer, heptamer, octomer, nonamer, decamer, undecamer, dodecamer, or tridecamer. In another aspect, the one or more oligonucleotide primers comprise a poly T sequence. In yet another aspect of any one of the foregoing embodiments, the one or more oligonucleotide primers comprise a mixture of random hexamers, heptamers, octomers, nonamers, decamers, undecamers, dodecamers, or tridecamers; and a poly T sequence.

[0013] In another aspect of any one of the foregoing embodiments, the method further comprises degrading single stranded DNA in the reaction mixture, such as with an ssDNA specific exonuclease, an ssDNA specific endonuclease, or a combination thereof. In some cases, the exonuclease is exonuclease 1, exonuclease 7 or a combination thereof. In some cases, the endonuclease is S1 endonuclease, mung bean endonuclease, or a combination thereof.

[0014] In another aspect of any one of the foregoing embodiments, the strand displacing polymerase comprises Klenow polymerase, exo – Klenow polymerase, 5'-3' exo – Klenow polymerase, Bst polymerase, Bst large fragment polymerase, Vent polymerase, , Vent polymerase, Deep Vent (exo-) polymerase, 9°Nm polymerase, Therminator polymerase, Therminator II polymerase, MMulV Reverse Transcriptase, phi29 polymerase, or DyNAzyme EXT polymerase, or a combination thereof.

[0015] In another aspect of any one of the foregoing embodiments, the present invention further provides for phosphorylating the 5' ends of the one or more blunt ended double stranded products.

[0016] In another aspect of any one of the foregoing embodiments, the present invention further provides for extending the 3' end of the one or more blunt ended double stranded products.

[0017] In another aspect of any one of the foregoing embodiments, the present invention further provides for ligating the double stranded products with one or more double stranded adapter oligonucleotides.

[0018] In another aspect of any one of the foregoing embodiments, the present invention comprises providing the input nucleic acid template comprising one or more non-canonical nucleotides by performing PCR, strand displacement amplification (SDA), multiple displacement amplification (MDA), rolling circle amplification (RCR), single primer isothermal amplification (SPIA), Ribo-SPIA, ligase chain reaction (LCR), Nucleic Acid Sequence Based Amplification (NASBA), Q-Beta Replicase amplification, Self-sustained sequence replication (3SR) or ligation activated transcription (LAT).

[0019] In a separate but related embodiment, the present invention provides a kit comprising: a) a glycosylase; b) a polyamine, an AP endonuclease, or a combinatin thereof; and c) one or more double stranded adapter oligonucleotides. In one aspect of this embodiment, the present invention provides a kit comprising instructions for the use of said kit. In aspect, the instructions comprise a method for generating dsDNA fragments, said method comprising: a) providing an input nucleic acid template comprising one or more non-canonical nucleotides in a reaction mixture; b) hybridizing and extending one or more oligonucleotide primers to the input nucleic acid template with an enzyme comprising strand displacement activity, wherein the hybridizing and extending produces one or more double stranded products; and c) cleaving the input nucleic acid template.

[0020] In one aspect of the embodiment, the kit further comprises one or more oligonucleotide primers. In another aspect of this embodiment, the one or more oligonucleotide primers comprise random hybridizing portions. In

IDT Exhibit 1020.5568

another aspect of this embodiment, the one or more oligonucleotide primers comprise polyT hybridizing portions. In another aspect of this embodiment, the one or more oligonucleotide primers comprise a mixture of random hybridizing portions and polyT portions. In some cases, at least one or at least two of the one or more oligonucleotide primers is a composite primer.

[0021] In a separate but related embodiment, the present invention provides a composition comprising: a) an input template nucleic acid comprising one or more non-canonical nucleotides; b) one or more oligonucleotide primers comprising randomized hybridizing portions; c) a polymerase comprising strand-displacement activity; d) a glycosylase; and one or more double stranded DNA product molecules.

[0022] In another aspect, a kit is provided comprising: a) an input template nucleic acid, b) one or more oligonucleotide primers comprising randomized hybridizing portions; c) a polymerase comprising strand-displacement activity; and d) an endonuclease.

[0023] In another aspect, a method is provided comprising: a) providing an input RNA template in a reaction mixture; b) hybridizing a first primer to the input RNA template; c) reverse transcribing the input RNA template in the presence of one or more non-canonical nucleotides to generate a first strand cDNA; d) cleaving the input RNA template; e) performing second strand synthesis to generate a double stranded cDNA; and f) cleaving the first strand cDNA to generate a single stranded second strand cDNA.

[0024] In one embodiment, the reverse transcribing comprises use of an RNA-dependent DNA polymerase. In another embodiment, the reaction mixture further comprises an inhibitor of DNA-dependent DNA polymerase activity. In another embodiment, the inhibitor of DNA-dependent DNA polymerase activity inhibits the DNA-dependent DNA polymerase activity of the RNA-dependent DNA polymerase. In another embodiment, the inhibitor is actinomycin D. In another embodiment, the inhibitor is removed prior to the performing second strand synthesis. In another embodiment, the first primer comprises a 5'-tail sequence that does not hybridize to the input RNA template. In another embodiment, the 5'-tail sequence comprises DNA. In another embodiment, the 3'-end of the first primer hybridizes to the input RNA template. In another embodiment, the reverse transcribing comprises extension of the first primer hybridized to the input RNA template. In another embodiment, the non-canonical nucleotide is dUTP. In another embodiment, the second strand synthesis comprises primer extension of a second primer hybridized to the first strand cDNA. In another embodiment, the second primer comprises a 3'-sequence hybridizable to the first strand cDNA. In another embodiment, the second primer comprises a 5'-tail that is not complementary to the first strand cDNA. In another embodiment, the second strand synthesis is carried out by DNA polymerase. In another embodiment, the second strand synthesis is carried out in the absence of non-canonical nucleotide triphosphates. In another embodiment, cleaving the input RNA template in a complex with the first primer extension product or products comprises exposing the input RNA template to an RNase H with or without other RNases, or cleaving the input RNA template following first primer extension reaction by heat or chemical treatment or combination thereof. In another embodiment, cleaving the first strand cDNA comprises combining the reaction mixture with an enzyme that cleaves the base portion of the non-canonical nucleotide to generate an abasic site.

[0025] In another embodiment, the enzyme that cleaves the base portion of the non-canonical nucleotide to generate an abasic site is a glycosylase. In another embodiment, the glycosylase is UNG or UDG. In another embodiment, the reaction mixture further comprises an amine. In another embodiment, the amine is DMED. In another embodiment, the cleaving the first strand cDNA generates fragments of the first strand cDNA with blocked 3'-ends.

IDT Exhibit 1020.5569

[0026] In another embodiment, the method further comprises sequencing said single stranded second strand cDNA. In another embodiment, the sequencing comprises next generation sequencing. In another embodiment, the single stranded second strand cDNA comprises sequence homologous to the input RNA template flanked by 3' and/or 5' sequences compatible for use in DNA sequencing. In another embodiment, said single stranded second strand cDNA comprises sequence homologous to the input RNA template flanked by 3' and/or 5' comprising sequences that function as barcodes. In another embodiment, said single stranded second strand cDNA comprises sequence homologous to the input RNA template flanked by 3' and/or 5' sequences that comprise recognition sequence for one or more restriction enzymes. In another embodiment, said single stranded second strand cDNA comprises sequence homologous to the input RNA template flanked by 3' and/or 5' sequences that enables circularization by hybridizing an oligonucleotide complementary to the end sequences and ligation of the ends. In another embodiment, the method further comprises amplification of the cDNA by rolling circle amplification. In another embodiment, the method further comprises amplifying the single stranded cDNA by single primer isothermal amplification.

[0027] In another embodiment, the input RNA template is from a biological sample. In another the input RNA template is from a sample lysate. In another embodiment, the input RNA template is from a cell free fluid. In another embodiment, the cell free fluid is plasma or serum. In another embodiment, the input RNA template is fragmented RNA. In another embodiment, the input RNA template is from an FFPE sample. In another embodiment, the input RNA template is fragmented by treatment with heat in the presence of multivalent cations.

INCORPORATION BY REFERENCE

[0028] 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

[0029] 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:

[0030] **Figure 1a**: depicts methods and compositions in one embodiment of the invention for generating double stranded nucleic acid products from one or more template polynucleotides utilizing a plurality of oligonucleotides comprising random hybridizing portions and a polymerase comprising strand displacement activity. In this embodiment the products are generated from a template that contains one or more non-canonical nucleotides (dUTP). The figure further depicts methods and compositions for fragmenting the template polynucleotide and removing undesired single stranded nucleic acids.

[0031] **Figure 1b**: depicts methods and compositions in another embodiment of the invention for generating double stranded nucleic acid products from one or more template polynucleotides utilizing a plurality of oligonucleotides comprising random hybridizing portions and a polymerase comprising strand displacement activity. In this embodiment the template nucleotide does not contain any non-canonical nucleotides. Methods and compositions for fragmenting the template polynucleotide and removing undesired single stranded nucleic acids can be used.

IDT Exhibit 1020.5570

[0032] **Figure 2**: depicts methods and compositions for preparing double stranded nucleic acid for adaptor ligation.

[0033] **Figure 3**: depicts methods and compositions for generating labeled double stranded nucleic acid products from one or more template polynucleotides utilizing a plurality of oligonucleotides comprising one or more labels and random hybridizing portions and a polymerase comprising strand displacement activity. The figure further predicts methods and compositions for fragmenting the template polynucleotide. Undesired single stranded nucleic acids can be removed.

[0034] **Figure 4**: depicts BioAnalyzer profiles of input amplified cDNA, random primed products mixture and random primed products following cleavage of the input amplified cDNA. Whole transcriptome amplified cDNA from HeLa total RNA. The x-axis represents product length in bp. The concentration of the starting material is 172ng/μL. The concentration of random primed amplified DNA is 200ng/μL. The concentration of the random primed and fragmented input nucleic acid is 100ng/μL.

[0035] **Figure 5**: depicts BioAnalyzer profiles of input amplified cDNA, random primed products mixture and random primed products following cleavage of the input amplified cDNA. Whole transcriptome amplified cDNA from total RNA from a biological sample and brain. The x-axis represents product length in bp. The profile of the random priming products following removal of the input amplified cDNA is very reproducible for the two samples which represent different sample source.

[0036] **Figure 6:** depicts an individual using a kit to prepare a blunt-ended sample from a whole transcriptome which is then put onto a next-generation nucleic acid analysis platform, such as a high-throughput sequencing apparatus. The performing and monitoring of the sequencing reactions may be automated such that the sequencing reactions are performed via robotics. In addition, the sequencing information and data obtained may be provided to a personal computer, a personal digital assistant, a cellular phone, a video game system, or a television so that a user can monitor the progress of the sequencing reactions remotely.

[0037] **Figure 7:** illustrates the generation of second strand cDNA with defined sequences at 3'- and 5'-ends.

[0038] **Figure 8:** illustrates that second strand cDNA with defined 5'- and 3'- ends can be amplified with single primer isothermal amplification (SPIA).

DETAILED DESCRIPTION OF THE INVENTION

I. Introduction

[0039] Large scale analysis of nucleic acids often requires amplification and generation of products which are single stranded (ss DNA or RNA) or double stranded (ds DNA) and are either labeled, ready for labeling, ready for ligation of desired adaptors and the like. These large scale analysis methods are evolving and comprise, e.g. pyrosequencing, sequencing by synthesis, sequencing by hybridization, single molecule sequencing, nanopore sequencing, and sequencing by ligation, high density PCR, microarray hybridization, array comparative genomic hybridization (CGH), serial analysis of gene expression (SAGE), digital PCR, and massively parallel Q-PCR; subtractive hybridization; differential amplification; preparation of libraries (including cDNA and differential expression libraries); preparation of an immobilized nucleic acid (which can be a nucleic acid immobilized on a microarray), and characterizing amplified nucleic acid products generated by the methods of the invention.

[0040] Various methods for global amplification of DNA target molecules (e.g., whole genome amplification) have been described, including methods based on the polymerase chain reaction (PCR). See, e.g., U.S. Pat. Nos. 5,731,171; 6,365,375; Daigo et al., (2001) Am. J. Pathol. 158 (5):1623-1631; Wang et al, (2001); Cancer Res. 61:4169-4174; Zheng et al, (2001) Cancer Epidemiol. 10:697-700; Dietmaier et al (1999) Am. J. Pathol. 154 (1) 83-

IDT Exhibit 1020.5571

95; Stoecklein et al (2002) Am. J. Pathol. 161 (1):43-51; U.S. Pat. Nos. 6,124,120; 6,280,949; Dean et al (2002) PNAS 99 (8):5261-5266. However, PCR-based global amplification methods, such as whole genome amplification (WGA), may generate non-specific amplification artifacts, give incomplete coverage of loci, or generate DNA of insufficient length that cannot be used in many applications. PCR-based methods also suffer from the propensity of the PCR reaction to generate products that are preferentially amplified, and thus resulting in biased representation of genomic sequences in the products of the amplification reaction. Methods of global amplification of DNA using composite primers have been described. See, e.g. U.S. Patent application No. 10/824,829.

[0041] Additionally, a number of methods for the analysis of gene expression have been developed in recent years. See, for example, U.S. Pat. Nos. 6,251,639; 6,692,918; 6,686,156; 5,744,308; 6,143,495; 5,824,517; 5,829,547; 5,888,779; 5,545,522; 5,716,785; 5,409,818; EP 0971039A2; EP0878553A2; and U.S. Patent Application Publication Nos. 2002/0115088, 2003/0186234, 2003/0087251, and 2004/0023271. These include quantification of specific mRNAs and the simultaneous quantification of a large number of mRNAs as well as the detection and quantification of patterns of expression of known and unknown genes. RNA amplification is most commonly performed using the reverse transcriptase-polymerase chain reaction (RT-PCR) method and variations thereof. These methods are based on replication of RNA by reverse transcriptase to form single stranded DNA complementary to the RNA (cDNA), which is followed by polymerase chain reaction (PCR) amplification to produce multiple copies of double stranded DNA. RNA amplification also includes methods based on *in vitro* transcription such as, for example, SP6, T3, or T7 transcription methods. These methods are based on replication of RNA by reverse transcriptase to form DNA (cDNA), which may then include an optional amplification step such as any of the amplification steps provided herein including, but not limited to, PCR or linear amplification. Multiple copies of the RNA can then be made using a DNA dependent RNA polymerase such as T7, T3 or SP6 polymerase which acts on a promoter that has been incorporated during the step of reverse transcription or during a subsequent amplification step. Methods of amplification from RNA templates have been described, for example, in U.S. Patent Application Publication Nos. 2003/0096349, 20040081978 and U.S. Patent Nos. 5,989,812 and 6,946,251.

[0042] Sequencing of nucleic acids continues to be a useful way to analyze DNA and RNA samples. Recent developments have made possible highly parallel high throughput sequencing. Many of these approaches use an in vitro cloning step to generate many copies of each individual molecule. Emulsion PCR is one method, which can involve isolating individual DNA fragment molecules along with primer-coated beads in aqueous bubbles within an oil phase. A polymerase chain reaction (PCR) then coats each bead with clonal copies of the isolated library molecule and these beads are subsequently immobilized for later sequencing. See, e.g., WO04069849A2 and WO05010145A2. In other cases, surface methods of clonal amplification have been developed, for example, by the use of bridge PCR where fragments are amplified upon primers attached to a solid surface. In some cases, bridge PCR requires the ligation of nucleic acid adaptor molecules to double stranded fragment molecules to allow efficient capture of the fragments on the solid surface. These methods produce many physically isolated locations which each contain many copies of a single fragment. While these methods have provided improvements in sequencing throughput, there is a continuing need to improve the methods of obtaining samples appropriate for sequencing, and of handling, and amplifying such samples to produce suitable double or single stranded nucleic acid. In particular, there is a need to improve methods for obtaining suitable nucleic acid product molecules from complex samples such as whole genome or transcriptome samples, or samples comprising a large number of genomic fragments or RNA fragments.

IDT Exhibit 1020.5572

[0043] The generation of the ds DNA products may be based on random priming of the target nucleic acids, which may represent whole genome, whole transcriptome, total RNA or total DNA in a sample. Random primers and corresponding DNA polymerases suitable for this reaction are provided herein. Additional random primers and suitable polymerases are known in the art. Random hexamers, nonamers and longer oligonucleotide primers are suitable for the methods of the present invention, as well as various combinations of random, not-so random, and non random sequences. The primers may be labeled, may contain components which enhance hybridization, groups or links which prevent degradation by a proof reading polymerase and other modification designed for the specific applications. The generation of the ds DNA products may also be based on non-random priming or "not so random" priming of the target nucleic acids in a sample. For example, a pool of primers may be utilized that randomly hybridizes to total mRNA or a substantial fraction thereof but does not substantially hybridize to rRNA. Alternatively, a pool of primers may be utilized that hybridizes to a large number of genomic sequences, while not hybridizing to undesired genomic sequences. Non random primers which are designed to randomly prime along the desired template nucleic acids in the sample have been described for example in international patent application publication number WO 2009/055732.

[0044] DNA polymerases useful for random primed DNA synthesis can be selected from a group comprising polymerases with strand displacement activity. Both thermostable and thermolabile polymerases useful for these reactions include Klenow polymerase, with or without 3'-exonuclease, Bst DNA polymerase, φ29 DNA polymerase and other polymerases known in the art. In some cases, the polymerase does not comprise a 5'-exonuclease activity.

[0045] Random priming of the input nucleic acid template with a DNA polymerase comprising strand displacement activity may result in priming along the template as well as the displaced single stranded primer extension products leading to the generation of ds DNA products comprising primer extension products as well as ds DNA species comprising the original template, or templates, and a complement primer extension primer. Thus the products of random priming reaction as described above are non homogenous with respect to their origin, that is to say, a mixture of the input nucleic acid template and primer extension products, the last comprise both ds DNA and single stranded DNA. The non homogeneous nature of the mixture of nucleic acids at the end of the random priming reaction with a polymerase comprising stand displacement activity also includes differences in size, in so far as the input nucleic acid is longer. Additionally, the non homogenous nature may include differences in composition; for example, the random priming product may be synthesized in the presence of modified nucleotide tri-phosphates such as labeled nucleotide tri phosphate, when random priming is carried out for the generation of labeled products, such as required for example for array based analysis.

[0046] It is desirable to generate a homogenous population of random priming products for downstream analysis, such as array based analysis, for example, array based comparative genome analysis (array CGH), or next generation high throughput sequencing (for example Illumina Genome Analysis). In the first example random priming is carried out for the generation of labeled target libraries from the respective test and reference genomic DNA. Only the random priming products are labeled by incorporation of labeled dNTP or employing a labeled random primer, while the input nucleic acid to be analyzed is not labeled. It is desirable to separate the non labeled, much longer, input nucleic acid from the labeled shorter random priming products for improved analysis. For example, the longer input nucleic acid may provide cross hybridization resulting in erroneous results. Additionally, it is possible that the determination of label specific activity of the random priming products applied to the array will be inaccurately determined in a mixture of labeled random priming products and non labeled input nucleic acid.

IDT Exhibit 1020.5573

[0047] Similarly, the generation of ds DNA libraries from nucleic acid to be sequenced using the newly developed second generation high throughput sequencing platforms, such as the commercially available Genome Analyzer (Illumina), will benefit from increasing the homogeneity of the random priming products used for ligation dependent adaptor ligation required for subsequent clonal amplification (cluster formation).

[0048] Large scale analysis of biological samples is often limited by the scarcity of the nucleic acid, RNA or DNA, samples. The amount of nucleic acids in biological samples is often much lower than that required for large scale analysis. The increased recognition of the importance of analyzing defined cell populations in biological samples, as little as single cells, imposes further limitation of direct analysis of the sample nucleic acid. Thus, the genome wide analysis of biological samples benefits from the development of methods and reagents for linear amplification of the sample nucleic acids (RNA and/or DNA). The present invention provides methods, composition and kits for generation of linearly amplified products which are specifically marked to be employed as templates for target generation by, for example, random priming, and to be distinguished from the random priming products described, for further fragmentation and removal from the desired products. The methods are also applicable to other amplification procedures and reagents. In general, the methods of the invention apply to any method in which the sample nucleic acid is first replicated or amplified employing DNA synthesis in the presence of non canonical dNTP for marking of the synthesized nucleic acid which will in turn serve as an input for further preparation of targets for downstream analysis, such as generation of double stranded or single stranded products by random priming. In some cases, the replication or amplification is performed in a limited fashion such as a single round of amplification or a small number of rounds of amplification (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40 etc.) such as may be used during the synthesis of cDNA to avoid amplification bias. In other cases, the invention provides methods for generating a first copy of the template RNA in the sample, employing an RNA dependent DNA polymerase, in the presence of non canonical nucleotide and an inhibitor of DNA dependent DNA polymerase, thus generating a marked first strand cDNA which can serve as the input for generating second strand cDNA by DNA dependent DNA polymerase in the absence of the non canonical nucleotides and the inhibitor of DNA dependent DNA polymerase, thus generating first and second cDNA strands which are distinguishable by the presence of the non canonical nucleotides in the said first but not second strand cDNA. The said first strand cDNA can be degraded by the methods of the inventions (fragmentation by glycosylase and DMED for example) without affecting the second strand cDNA which can be subjected to down stream analysis such highly parallel sequencing in a strand specific manner.

II. Exemplary Methods of the Invention

[0049] The methods of the present invention may include a step of providing an input nucleic acid template in a reaction mixture. Suitable input nucleic acid templates include, but are not limited to, e.g., DNA or RNA templates. In some embodiments the templates do not comprise non-canonical nucleotides. In other embodiments the templates comprise one or more non-canonical nucleotides. The input nucleic acid template may comprise DNA or RNA or a combination thereof, which template may optionally further comprise one or more non-canonical nucleotides. In some cases, the input nucleic acid template is DNA such as, for example, whole genomic DNA or a portion thereof. In some cases, the input nucleic acid template is DNA that has been previously selected. For example DNA, such as genomic DNA, may be hybridized to capture probes (e.g., capture probes on an array or on beads), washed to remove un-hybridized sequences corresponding to undesired fragments of nucleic acid and then collected for the methods of the present invention. Alternatively, DNA may be previously selected by non-random

IDT Exhibit 1020.5574

or "not so random" priming and amplification of a sample of genomic DNA. In some cases, the non-random or not-so random priming and amplification may be performed in the presence of one or more non-canonical nucleotides such that they are incorporated into the template nucleic acid. In some cases the DNA derived from the RNA is generated by reverse transcriptase in the presence of non canonical nucleotide (e.g., dUTP) and an inhibitor of the DNA dependent DNA polymerase activity of the reverse transcriptase (e.g., actinomycin D). In related cases, the non-random or not-so-random priming and amplification may be performed not in the presence of non-canonical nucleotides. In some cases the template is RNA such as total RNA, mRNA, viral RNA, small RNA fraction and the like. In other cases, the input nucleic acid template may comprise DNA that is derived from RNA such as by reverse transcription. Alternatively, the input nucleic acid template may comprise RNA that is derived from DNA such as by *in vitro* transcription. In some cases, the reverse transcription or *in vitro* transcription may be performed in the presence of one or more non-canonical nucleotides which are thus incorporated into the template nucleic acid. In other cases, non-canonical nucleotides are not incorporated into the template nucleic acid. The nucleic acid in the sample may also represent a mixture of genomic DNAs, such as from a microbial population.

[0050] Reaction mixtures suitable for the methods of the present invention include aqueous solutions or water in oil emulsions as provided herein. Various buffers, detergents, salts, chelators, chaotropes, azotropes, solvents, cations, divalent cations such as magnesium or manganese, and anions are suitable for storage and downstream manipulation (e.g., hybridization, polymerization, etc.) of nucleic acids. Reaction mixture compositions suitable for the methods of the present invention are provided herein.

[0051] In some cases, the input nucleic acid template is RNA, such as for example, whole transcriptome RNA, or a portion thereof (e.g., mRNA or miRNA). In some cases the input nucleic acid template is RNA that does not comprise any non-canonical nucleotides. In some cases, the input nucleic acid template is RNA that further comprises one or more non-canonical nucleotides. In some cases, the input nucleic acid template is RNA that has been previously selected. For example, RNA may be hybridized to capture probes (e.g., capture probes on an array), washed to remove un-hybridized sequences corresponding to undesired fragments of RNA and then collected for the methods of the present invention. Alternatively, RNA may be previously selected by non-random or "not so random" priming and amplification of a sample of RNA. In some cases, the RNA may be previously selected by non-random or not so random priming and amplification in the presence of one or more non-canonical nucleotides such that they are incorporated into the template nucleic acid.

[0052] The input nucleic acid template may be provided by first performing an amplification reaction on a sample comprising nucleic acid. In some cases, the amplification reaction may be performed in the presence of a non-canonical nucleotide. In some cases, the amplification reaction may be performed not in the presence of a non-canonical nucleotide. Conditions for limited and/or controlled incorporation of a non-canonical nucleotide are provided herein. See, e.g., Jendrisak, U.S. Pat. No. 6,190,865; Mol. Cell Probes (1992) 251-6; Anal. Biochem. (1993) 211:164-9; see also Sambrook (1989) "Molecular Cloning: A Laboratory Manual", second edition; Ausebel (1987, and updates) "Current Protocols in Molecular Biology." The frequency (or spacing) of non-canonical nucleotides in the resulting polynucleotide comprising a non-canonical nucleotide, and thus the average size of fragments generated using the methods of the invention (i.e., following cleavage of a base portion of a non-canonical nucleotide, and cleavage of a phosphodiester backbone at a non-canonical nucleotide), is controlled by variables including: frequency of nucleotide(s) corresponding to the non-canonical nucleotide(s) in the template (or other measures of nucleotide content of a sequence, such as average G-C content), ratio of canonical to non-canonical nucleotide present in the reaction mixture; ability of the polymerase to incorporate the non-canonical

IDT Exhibit 1020.5575

nucleotide, relative efficiency of incorporation of non-canonical nucleotide verses canonical nucleotide, and the like. It is understood that the average fragmentation size also relates to the reaction conditions used during fragmentation, as is further discussed herein. The reaction conditions can be empirically determined, for example, by assessing average fragment size generated using the methods of the invention taught herein. The level of labeling at an abasic site also relates to the frequency of incorporation of non-canonical nucleotides, as is further discussed herein.

[0053] In the embodiment where the template nucleic acid contains one or more non-canonical nucleotides, the template nucleic acid comprising one or more non-canonical nucleotides may be synthesized so that it comprises a sufficient and predictable density of non canonical nucleotides to provide for sufficient and predictable fragmentation, and when used with one or more agents capable of cleaving at the non-canonical nucleotides (e.g., a glycosidase, a glycosidase and an amine, or a glycosidase and an AP endonuclease) to further generate fragments of desirable size range. In some cases, the one or more agents employed for cleaving at the non-canonical nucleotides of the template nucleic acid may provide fragmentation products with a blocked 3'-end. Generally, a non-canonical base can be incorporated at about every 5, 10, 15, 20, 25, 30, 40, 50, 65, 75, 85, 100, 123, 150, 175, 200, 225, 250, 300, 350, 400, 450, 500, 550, 600, 650 or more nucleotides apart in the resulting polynucleotide comprising a non-canonical nucleotide. In one embodiment, the non-canonical nucleotide is incorporated about every 200 nucleotides, about every 100 nucleotide, or about every 50, 25, 20, 15, 10, 9, 8, 7, 6, 5, or fewer nucleotides. In another embodiment, the non-canonical nucleotide is incorporated about every 50 to about 200 nucleotides. In some embodiments, a 1:1, 1:2, 1:3, 1:4, 1:5, 1:6, 1:10, 1:15, 1:20 or higher ratio of non-canonical to canonical nucleotide may be used in the reaction mixture. In some cases, a 1:1, 1:2, 1:3, 1:4, 1:5, 1:6, 1:10, 1:15, 1:20 or higher ratio of the non-canonical nucleotide dUTP to canonical nucleotide dTTP is used in the reaction mixture.

[0054] The input nucleic acid template or the nucleic acid starting material from which template nucleic acid comprising one or more non-canonical nucleotides is provided for the methods of the present invention may include double-stranded, partially double-stranded, and single-stranded nucleic acids from any source in purified or unpurified form, which can be DNA (dsDNA and ssDNA) or RNA, including tRNA, mRNA, rRNA, mitochondrial DNA and RNA, chloroplast DNA and RNA, DNA-RNA hybrids, or mixtures thereof, genes, chromosomes, plasmids, the genomes of biological material such as microorganisms, e.g., bacteria, yeasts, viruses, viroids, molds, fungi, plants, animals, humans, and fragments thereof. Obtaining and purifying nucleic acids use standard techniques in the art. RNAs can be obtained and purified using standard techniques in the art. Starting material comprising DNA (including genomic DNA) can be transcribed into RNA form, which can be achieved using methods disclosed in Kurn, U.S. Pat. No. 6,251,639, and by other techniques, such as expression systems. RNA copies of genomic DNA would generally include untranscribed sequences generally not found in mRNA, such as introns, regulatory and control elements, etc. DNA copies of an RNA starting material can be synthesized using methods described in Kurn, U.S. Patent Publication No. 2003/0087251 or other techniques known in the art.

[0055] Synthesis of template polynucleotide comprising one or more non-canonical nucleotides from a DNA-RNA hybrid can be accomplished by for example denaturation of the hybrid to obtain ssDNA and/or RNA, cleavage with an agent capable of cleaving RNA from an RNA/DNA hybrid. The nucleic acid starting material can be only a minor fraction of a complex mixture such as a biological sample and can be obtained from various biological materials by procedures well known in the art. The starting material nucleic acid can be known or unknown and may contain more than one desired specific nucleic acid sequence of interest, each of which may be the same or different from each other. Therefore, the methods of the invention are useful not only for producing one specific polynucleotide suitable for downstream analysis, but also for producing simultaneously more than one different

IDT Exhibit 1020.5576

specific polynucleotides from one or more than one different template nucleic acid comprising a non-canonical nucleotide. The nucleic acid starting material can be a sub-population of nucleic acids, for example, a subtractive hybridization probe, total genomic DNA, restriction fragments, a cDNA library, cDNA prepared from total mRNA, a cloned library, or amplification products of any of the templates described herein. In some cases, the initial step of the synthesis of the complement of a portion of a nucleic acid starting material sequence is template denaturation. The denaturation step may be thermal denaturation or any other method, such as alkali treatment.

[0056] For simplicity, the template polynucleotide either comprising a non-canonical nucleotide or not comprising a non-canonical nucleotide is described as a single nucleic acid. It is understood that the polynucleotide can be a single polynucleotide, or a population of polynucleotides (from a few to a multiplicity to a very large multiplicity of polynucleotides). It is further understood that a template polynucleotide can be a multiplicity (from small to very large) of different polynucleotide molecules. Such populations can be related in sequence (e.g., member of a gene family or superfamily) or extremely diverse in sequence (e.g., generated from all mRNA, generated from all genomic DNA, etc.). Polynucleotides can also correspond to single sequence (which can be part or all of a known gene, for example a coding region, genomic portion, etc.). Methods, reagents, and reaction conditions for generating specific polynucleotide sequences and multiplicities of polynucleotide sequences are known in the art.

[0057] Suitable methods of synthesis of a polynucleotide comprising a non-canonical nucleotide are generally template-dependent (in the sense that polynucleotide comprising a non-canonical nucleotide is synthesized along a polynucleotide template, as generally described herein). It is understood that non-canonical nucleotides can be incorporated into a polynucleotide as a result of template-independent methods. For example, one or more primer(s) can be designed to comprise one or more non-canonical nucleotides. See, e.g., Richards, U.S. Pat. Nos. 6,037,152, 5,427,929, and 5,876,976. As discussed herein, inclusion of at least one non-canonical nucleotide in a primer results in cleavage of a base-portion of a non-canonical nucleotide and labeling at the abasic site (i.e., following generation of an abasic site, as described herein), thus generating a polynucleotide fragment comprising a portion of the primer.

[0058] Suitable amplification reactions further include any DNA or RNA amplification reaction, including but not limited to polymerase chain reaction (PCR), strand displacement amplification (SDA), multiple displacement amplification (MDA), rolling circle amplification (RCR), single primer isothermal amplification (SPIA, see e.g. US Patent No. 6251639), RNA single primer isothermal amplification (Ribo-SPIA, see e.g. US Patent Application Publication No. 2003/0087251), ligase chain reaction (LCR), nucleic acid sequence based amplification (NASBA), Q-beta replicase amplification, self-sustained sequence replication (3SR), ligation activated transcription (LAT), *in vitro* transcription (IVT), or a combination thereof. In some cases, the amplification methods for providing the template nucleic acid may be performed under limiting conditions such that only a few rounds of amplification (e.g., 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 etc.), such as for example as is commonly done for cDNA generation. The number of rounds of amplification can be about 1-30, 1-20, 1-15, 1-10, 5-30, 10-30, 15-30, 20-30, 10-30, 15-30, 20-30, or 25-30.

[0059] PCR is an *in vitro* amplification procedure 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.

IDT Exhibit 1020.5577

[0060] LCR 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.

[0061] NASBA 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.

[0062] The Q-beta-replicase method 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.

[0063] 3SR 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).

[0064] TMA, used by Gen-Probe, 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.

[0065] SDA (Westin et al 2000, Nature Biotechnology, 18, 199-202; Walker et al 1992, Nucleic Acids Research, 20, 7, 1691-1696), 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.

[0066] RCA (Lizardi et al. 1998, Nature Genetics, 19:225-232) 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.

[0067] 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.

IDT Exhibit 1020.5578

[0068] Some aspects of the invention utilize 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.

[0069] 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).

[0070] The methods of the present invention may further include a step of hybridizing one or more oligonucleotide primers to an input nucleic acid template. The template can optionally comprise one or more non-canonical nucleotides. In some cases the oligonucleotide primers may comprise a hybridizing portion which comprises random nucleotides, such as for example random dimers, trimers, tetramers, pentamers, hexamers, heptamers, octomers, nonomers, decamers, undecamers, dodecamers, tridecamers, tetradecamers, or longer. In other cases, the hybridizing portion may comprise a non random sequence such as a polyT sequence. In still other cases, the hybridizing portion of some of the oligonucleotide primers may comprise random nucleotides, while the hybridizing portion of some of the nucleotides comprise non-random sequences, such as polyT or "not so random sequences." In some cases, the hybridizing portion of the oligonucleotide primers may comprise "not so random sequences" such as for example a pool of sequences which randomly or pseudo-randomly prime desired sequences such as total mRNA or a substantial fraction thereof, but do not prime undesired sequences such as rRNA.

[0071] A "random primer," as used herein, can be a primer that generally comprises a sequence that is designed not necessarily based on a particular or specific sequence in a sample, but rather is based on a statistical expectation (or an empirical observation) that the sequence of the random primer is hybridizable (under a given set of conditions) to one or more sequences in the sample. A random primer can generally be an oligonucleotide or a population of oligonucleotides comprising a random sequence(s) in which the nucleotides at a given position on the oligonucleotide can be any of the four nucleotides, or any of a selected group of the four nucleotides (for example only three of the four nucleotides, or only two of the four nucleotides). In some cases all of the positions of the oligonucleotide or population of oligonucleotides can be any of the four nucleotides; in other cases, only a portion of the positions, for instance a particular region, of the oligonucleotide will comprise positions which can be any of the

IDT Exhibit 1020.5579

four bases. In some cases, the portion of the oligonucleotide which comprises positions which can be any of the four bases is about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, or about 15-20 nucleotides in length. In some cases, the portion of the oligonucleotide which comprises positions which can be any of the four bases is about 5-20, 5-15, 5-10, 4-8, 10-20, 15-20, or 10-15 nucleotides in length. In some cases, a random primer may comprise a tailed primer having a 3'-region that comprises a random sequence and a 5'-region that is a non-hybridizing sequence that comprises a specific, non-random sequence. The 3'-region may also comprise a random sequence in combination with a region that comprises poly-T sequences. The sequence of a random primer (or its complement) may or may not be naturally-occurring, or may or may not be present in a pool of sequences in a sample of interest. The amplification of a plurality of RNA species in a single reaction mixture would generally, but not necessarily, employ a multiplicity, or a large multiplicity, of random primers. As is well understood in the art, a "random primer" can also refer to a primer that is a member of a population of primers (a plurality of random primers) which collectively are designed to hybridize to a desired and/or a significant number of target sequences. A random primer may hybridize at a plurality of sites on a nucleic acid sequence. The use of random primers provides a method for generating primer extension products complementary to a target polynucleotide which does not require prior knowledge of the exact sequence of the target. In some embodiments one portion of a primer is random, and another portion of the primer comprises a defined sequence. For example, in some embodiments, a 3'-portion of the primer will comprise a random sequence, while the 5'-portion of the primer comprises a defined sequence. In some embodiments a 3'-random portion of the primer will comprise DNA, and a 5'-portion defined portion of the primer will comprise RNA; in other embodiments, both the 3' and 5'-portions will comprise DNA. In some embodiments, the 5'-portion will contain a defined sequence and the 3'-portion will comprise a poly-dT sequence that is hybridizable to a multiplicity of RNAs in a sample (such as all mRNA).

[0072] The hybridizing portion of the oligonucleotide primers may comprise a pool of hybridizing portions which hybridize to a number of sequences or fragments to be analyzed such as for example, 1; 2; 3; 4; 5; 6; 7; 8; 9; 10; 11; 12; 13; 14; 15; 16; 17; 18; 19; 20; 25; 30; 35; 40; 45; 50; 55; 60; 75; 100; 150; 200; 250; 300; 400; 500; 600; 750; 1000; 10,000; 15,000; 20,000; 25,000; 30,000; 40,000; 50,000; 60,000; 75,000; 100,000; 150,000; 200,000; 250,000 or more sequences or fragments. In some cases, each fragment may be hybridized to one primer, in other cases, each fragment is hybridized on average to 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 or more oligonucleotide primers. Oligonucleotide primers suitable for the methods of the present invention are provided herein.

[0073] The olignucleotide primers may be extended along the input nucleic acid template to which they are hybridized. In some cases, the extension may be performed with a polymerase such as for example any of the polymerases provided herein including polymerases comprising strand displacement activity. Exemplary DNA dependent DNA polymerases suitable for the methods of the present invention include but are not limited to Klenow polymerase, with or without 3'-exonuclease, Bst DNA polymerase, Bca polymerase, φ29 DNA polymerase, Vent polymerase, Deep Vent polymerase, Taq polymerase, T4 polymerase, and *E. coli* DNA polymerase 1, derivatives thereof, or mixture of polymerases. In some cases, the polymerase does not comprise a 5'-exonuclease activity. In other cases, the polymerase comprises 5' exonuclease activity. In some cases, the primer extension of the present invention may be performed using a polymerase comprising strong strand displacement activity such as for example Bst polymerase. In other cases, the primer extension of the present invention may be performed using a polymerase comprising weak or no strand displacement activity. One skilled in the art may recognize the advantages and disadvantages of the use of strand displacement activity during the primer extension step, and which polymerases

IDT Exhibit 1020.5580

may be expected to provide strand displacement activity (see e.g., New England Biolabs *Polymerases*). For example, strand displacement activity may be useful in ensuring whole genome or whole transcriptome coverage during the random priming and extension step. Strand displacement activity may further be useful in the generation of double stranded amplification products during the priming and extension step. Alternatively, a polymerase which comprises weak or no strand displacement activity may be useful in the generation of single stranded nucleic acid products during primer hybridization and extension that are hybridized to the template nucleic acid.

[0074] An "RNA-dependent DNA polymerase" or "reverse transcriptase" ("RT") is an enzyme that synthesizes a complementary DNA copy from an RNA template. All known reverse transcriptases also have the ability to make a complementary DNA copy from a DNA template; thus, they are both RNA- and DNA-dependent DNA polymerases. Reverse transcriptases may also have an RNase H activity. Some examples of reverse transcriptases are reverse transcriptase derived from Maloney murine leukemia virus (MMLV-RT), avian myeloblastosis virus, retroviral reverse transcriptase, retrotransposon reverse transcriptase, hepatitis B reverse transcriptase, cauliflower mosaic virus reverse transcriptase, bacterial reverse transcriptase, E. coli DNA polymerase and klenow fragment, and Tth DNA polymerase. A primer can be used to initiate synthesis with both RNA and DNA templates. In other examples a DNA dependent DNA polymerase may also comprise an RNA-dependent DNA polymerase such as Klenow polymerase, Bst DNA polymerase and the like.

[0075] The extension of hybridized oligonucleotide primers, at least a portion of which may comprise random hybridizing portions, non-random hybridizing portions, not-so random hybridizing portions or a combination thereof, with a polymerase comprising strand displacement activity may provide for the generation of double stranded nucleic acid product fragments. In some cases, the extension of hybridized oligonucleotide primers, at least a portion of which comprise random hybridizing portions, with a polymerase comprising strand displacement activity may produce double stranded nucleic acid products comprising a mixture of double stranded nucleic acid fragment products produced in the polymerization reaction as well as double stranded molecules comprising template nucleic acid hybridized to one or more oligonucleotide primers.

[0076] In the embodiment where the template contains one or more non-canonical nucleotides, the products of the primer extension reaction, e.g. single or double stranded, partially double stranded, or mixtures thereof, may be distinguished from the template nucleic acid in that the template nucleic acid comprises one or more non-canonical nucleotides whereas the products of the primer extension reaction do not comprise non-canonical nucleotides, or do not comprise the same one or more non-canonical nucleotides. In some cases, double stranded products of the primer extension reaction comprise a hybrid duplex of a single strand of template nucleic acid comprising one or more non-canonical nucleotides and a single strand of primer extension product that does not comprise one or more non-canonical nucleotides, or does not comprise the same one or more non-canonical nucleotides. In other cases, double stranded products of the primer extension reaction comprise two strands, of which neither strand comprises one or more non-canonical nucleotides, or of which neither strand comprises the same one or more non-canonical nucleotides as the template nucleic acid.

[0077] The extension of hybridized oligonucleotide primers may be carried out for a suitable period of time. The period of time for the extension reaction may be anywhere from seconds to minutes to hours. For example, the extension step may include incubation of the input nucleic acid template in a reaction mixture such as the reaction mixtures provided herein with one or more oligonucleotide primers at a temperature suitable for the extension reaction (e.g., 15°C-80°C) for a period of between about 5 minutes and about 24 hours. Other suitable extension times include between about 1 minute and about 8 hours, about 2 minutes and about 7 hours, about 3 minutes and

IDT Exhibit 1020.5581

about 6 hours, about 4 minutes and about 5 hours, about 5 minutes and about 4 hours, about 5 minutes and about 3 hours, about 5 minutes and about 2 hours, about 10 minutes and about 2 hours, about 15 minutes and about 2 hours, about 20 minutes and about 2 hours, about 30 minutes and about 2 hours, or between about 30 minutes and about 1 hour. Still other suitable extension times include 1 minute, 2 minutes, 3 minutes, 4 minutes, 5 minutes, 6 minutes, 7 minutes, 8 minutes, 9 minutes, 10 minutes, 12 minutes, 15 minutes, 20 minutes, 30 minutes, 45 minutes, 60 minutes, 1 hour, 1.5 hours, 2 hours, 2.5 hours, 3 hours, 3.5 hours, 4 hours or more. Still other suitable extension times include about 1 minute, 2 minutes, 3 minutes, 4 minutes, 5 minutes, 6 minutes, 7 minutes, 8 minutes, 9 minutes, 10 minutes, 12 minutes, 15 minutes, 20 minutes, 30 minutes, 45 minutes, 60 minutes, 1 hour, 1.5 hours, 2 hours, 2.5 hours, 3 hours, 3.5 hours, 4 hours or more.

[0078] The extension step may be performed in a reaction mixture comprising nucleotides, labeled nucleotides or a combination thereof. For example, the hybridized oligonucleotides may be extended by one or more polymerases, such as polymerases comprising strand displacement activity or polymerases comprising weak or no strand displacement activity, along the input nucleic acid template in the presence of a mixture of dNTPs and amino allyl dNTPs. The use of amino-allyl dNTPs may allow further labeling and modification of the products of the extension reaction such as double stranded DNA fragment products. For example, the amino allyl dNTPs may provide for biotinylation, fluoresceination, labelling with Cy dyes (e.g., Cy3 or Cy5), or any other nucleic acid modification known in the art. Other modified nucleotides which are suitable for post amplification labeling by either covalent or non-covalent attachment of labels (e.g., fluorophores, chromophores, biotin, antibodies, antigens, or enzymes such as alkaline phosphatase or horse radish peroxidase) are also applicable including for example thio, phosphorothio, and amino modified nucleotides and oliognucleotides as described in U.S. Patent Nos. 6172209, 5679785, and 5623070, or any other modified nucleotides provided herein.

[0079] The methods of the present invention may further include a step of cleaving the input nucleic acid template. In some cases, the input nucleic acid template may be cleaved with an agent such as an enzyme. In the embodiment where the polynucleotide comprises a non-canonical nucleotide, the polynucleotide may be treated with an agent, such as an enzyme, capable of generally, specifically, or selectively cleaving a base portion of the non-canonical deoxyribonucleoside to create an abasic site. As used herein, "abasic site" encompasses any chemical structure remaining following removal of a base portion (including the entire base) with an agent capable of cleaving a base portion of a nucleotide, e.g., by treatment of a non-canonical nucleotide (present in a polynucleotide chain) with an agent (e.g., an enzyme, acidic conditions, or a chemical reagent) capable of effecting cleavage of a base portion of a non-canonical nucleotide. In some embodiments, the agent (such as an enzyme) catalyzes hydrolysis of the bond between the base portion of the non-canonical nucleotide and a sugar in the non-canonical nucleotide to generate an abasic site comprising a hemiacetal ring and lacking the base (interchangeably called "AP" site), though other cleavage products are contemplated for use in the methods of the invention. Suitable agents and reaction conditions for cleavage of base portions of non-canonical nucleotides include: N-glycosylases (also called "DNA glycosylases" or "glycosidases") including Uracil N-Glycosylase ("UNG"; specifically cleaves dUTP) (interchangeably termed "uracil DNA glyosylase"), hypoxanthine-N-Glycosylase, and hydroxy-methyl cytosine-N-glycosylase; 3-methyladenine DNA glycosylase, 3- or 7-methylguanine DNA glycosylase, hydroxymethyluracile DNA glycosylase; T4 endonuclease V. See, e.g., Lindahl, PNAS (1974) 71(9):3649-3653; Jendrisak, U.S. Pat. No. 6,190,865 B1 or any of the glycosidases provided in Table 1 or homologues thereof such as enzymes with greater than about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 99%, 99.5%, or higher homology or identity at the amino acid or nucleotide level with any of the glycosydases provided herein. In one embodiment, uracil-N-

IDT Exhibit 1020.5582

glycosylase is used to cleave a base portion of the non-canonical nucleotide. In other embodiments, the agent that cleaves the base portion of the non-canonical nucleotide is the same agent that cleaves a phosphodiester backbone at the abasic site.

Table 1: Glycosylases in bacteria, yeast and humans

E. coli	Yeast (S. cerevisiae)	Human	Type	Substrates
AlkA	Mag1		monofunctional	3-meA, hypoxanthine
UDG	Ung1	UNG	monofunctional	uracil
Fpg	Ogg1	hOGG1	bifunctional	8-oxoG, FapyG
Nth	Ntg1	hNTH1	bifunctional	Tg, hoU, hoC, urea, FapyG
	Ntg2			
Nei		hNEIL1	bifunctional	Tg, hoU, hoC, urea, FapyG, FapyA
		hNEIL2		AP site, hoU
		hNEIL3		unknown
MutY		hMYH	monofunctional	A:8-oxoG
		hSMUG1	monofunctional	U, hoU, hmU, fU
		TDG	monofunctional	T:G mispair
		MBD4	monofunctional	T:G mispair

[0080] Cleavage of base portions of non-canonical nucleotides may provide general, specific or selective cleavage (in the sense that the agent (such as an enzyme) capable of cleaving a base portion of a non-canonical nucleotide generally, specifically or selectively cleaves the base portion of a particular non-canonical nucleotide), whereby substantially all or greater than about 99.9%, 99.5%, 99%, 98.5%, 98%, about 95%, about 90%, about 85%, about 80%, about 75%, about 70%, about 65%, about 60%, about 55%, about 50%, about 45%, or about 40% of the base portions cleaved are base portions of non-canonical nucleotides. However, extent of cleavage can be less. Thus, reference to specific cleavage is exemplary. General, specific or selective cleavage is desirable for control of the fragment size in the methods of generating template polynucleotide fragments of the invention (i.e., the fragments generated by cleavage of the backbone at an abasic site). Reaction conditions may be selected such that the reaction in which the abasic site(s) are created can run to completion, or the reaction may be carried out until 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 99%, or about 100% of the non-canonical nucleotides are converted to abasic sites. In some cases, the reaction conditions may be selected such that the reaction in which abasic site(s) are created at between about 10% and about 100% of the one or more non-canonical nucleotides present in the template nucleic acid, between about 20% and about 90%, between about 30% and about 90%, between about 50% and about 90% 95%, 99%, or 100% of the non-canonical nucleotides in the template nucleic acid.

[0081] In some embodiments, the template polynucleotide comprising a non-canonical nucleotide is purified following synthesis of the template polynucleotide (to eliminate, for example, residual free non-canonical nucleotides that are present in the reaction mixture). In other embodiments, there is no intermediate purification

IDT Exhibit 1020.5583

between the synthesis of the template polynucleotide comprising the non-canonical nucleotide and subsequent steps (such as hybridization of primers, extension of primers to produce primer extension products that do not comprise non-canonical nucleotides, or do not comprise the same non-canonical nucleotides as the template nucleic acid, cleavage of a base portion of the non-canonical nucleotide and cleavage of a phosphodiester backbone at the abasic site).

[0082] It is understood that the choice of non-canonical nucleotide can dictate the choice of enzyme to be used to cleave the base portion of that non-canonical enzyme, to the extent that particular non-canonical nucleotides are recognized by particular enzymes that are capable of cleaving a base portion of the non-canonical nucleotide. In some cases, the enzyme is a glycosylase. For example, a template nucleic acid comprising non-canonical nucleotides such as dUTP, 8-oxoguanine, or a methylated purine which may be cleaved by glycosylases may be used in the methods of the present invention. Other suitable non-canonical nucleotides include deoxyinosine triphosphate (dITP), 5-hydroxymethyl deoxycytidine triphosphate (5-OH-Me-dCTP) or any of the non-canonical nucleotides provided in Table 1. See, e.g., Jendrisak, U.S. Pat. No. 6,190,865. A glycosylase such as uracil DNA glycosylase (known as UNG or UDG) which may act on dUTP to provide an abasic site, Ogg1 which may act on 8-oxoguanine to provide an abasic site, or N-methyl purine DNA glycosylase which may act on methylated purines to provide an abasic site may then be used in the methods of the present invention to act on the input nucleic acid template comprising non-canonical nucleotides to initiate a step of cleaving the input nucleic acid template. The enzymes as provided herein may provide N-glycosydic bond cleavage of the input nucleic acid template at the one or more non-canonical nucleotides provided herein to produce one or more abasic (apurinic or apyrimidic) sites.

[0083] Additional glycosylases which may be used in the methods of the present invention and their non-canonical nucleotide substrates include 5-methylcytosine DNA glycosylase (5-MCDG), which cleaves the base portion of 5-methylcytosine (5-MeC) from the DNA backbone (Wolffe et al., Proc. Nat. Acad. Sci. USA 96:5894-5896, 1999); 3-methyladenosine-DNA glycosylase I, which cleaves the base portion of 3-methyl adenosine from the DNA backbone (see, e.g. Hollis et al (2000) Mutation Res. 460: 201-210); and/or 3-methyladenosine DNA glycosylase II, which cleaves the base portion of 3-methyladenosine, 7-methylguanine, 7-methyladenosine, and/3-methylguanine from the DNA backbone. See McCarthy et al (1984) EMBO J. 3:545-550. Multifunctional and mono-functional forms of 5-MCDG have been described. See Zhu et al., Proc. Natl. Acad. Sci. USA 98:5031-6, 2001; Zhu et al., Nuc. Acid Res. 28:4157-4165, 2000; and Nedderrnann et al., J. B. C. 271:12767-74, 1996 (describing bifunctional 5-MCDG; Vairapandi & Duker, Oncogene 13:933-938, 1996; Vairapandi et al., J. Cell. Biochem. 79:249-260, 2000 (describing mono-functional enzyme comprising 5-MCDG activity). In some embodiments, 5-MCDG preferentially cleaves fully methylated polynucleotide sites (e.g., CpG dinucleotides), and in other embodiments, 5-MCDG preferentially cleaves a hemi-methylated polynucleotide. For example, mono-functional human 5-methylcytosine DNA glycosylase cleaves DNA specifically at fully methylated CpG sites, and is relatively inactive on hemimethylated DNA (Vairapandi & Duker, supra; Vairapandi et al., supra). By contrast, chick embryo 5-methylcytosine-DNA glycosylase has greater activity directed to hemimethylated methylation sites. In some embodiments, the activity of 5-MCDG is potentiated (increased or enhanced) with accessory factors, such as recombinant CpG-rich RNA, ATP, RNA helicase enzyme, and proliferating cell nuclear antigen (PCNA). See U.S. Patent Publication No. 20020197639 A1. One or more agents may be used. In some embodiments, the one or more agents cleave a base portion of the same methylated nucleotide. In other embodiments, the one or more agents cleave a base portion of different methylated nucleotides. Treatment with two or more agents may be sequential or simultaneous.

IDT Exhibit 1020.5584

[0084] Appropriate reaction media and conditions for carrying out the cleavage of a base portion of a non-canonical nucleotide according to the methods of the invention are those that permit cleavage of a base portion of a non-canonical nucleotide. Such media and conditions are known to persons of skill in the art, and are described in various publications, such as Lindahl, PNAS (1974) 71(9):3649-3653; and Jendrisak, U.S. Pat. No. 6,190,865 BI; U.S. Pat. No. 5,035,996; and U.S. Pat. No. 5,418,149. For example, buffer conditions can be as described above with respect to polynucleotide synthesis. In one embodiment, UDG (Epicentre Technologies, Madison Wis.) is added to a nucleic acid synthesis reaction mixture, and incubated at 37° C. for 20 minutes. In one embodiment, the reaction conditions are the same for the synthesis of a polynucleotide comprising a non-canonical nucleotide and the cleavage of a base portion of the non-canonical nucleotide. In another embodiment, different reaction conditions are used for these reactions. In some embodiments, a chelating regent (e.g. EDTA) is added before or concurrently with UNG in order to prevent a polymerase from extending the ends of the cleavage products.

[0085] The polynucleotide comprising an abasic site may be labeled using an agent capable of labeling an abasic site, and, in embodiments involving fragmentation, the phosphodiester backbone of the polynucleotide comprising an abasic site may be cleaved at the site of incorporation of the non-canonical nucleotide (i.e., the abasic site by an agent capable of cleaving the phosphodiester backbone at an abasic site, such that two or more fragments are produced. In embodiments involving fragmentation, labeling can occur before fragmentation, fragmentation can occur before labeling, or fragmentation and labeling can occur simultaneously.

[0086] Agents capable of labeling (e.g., generally or specifically labeling) an abasic site, whereby a polynucleotide (or polynucleotide fragment) comprising a labeled abasic site is generated, are provided herein. In some embodiments, the detectable moiety (label) is covalently or non-covalently associated with an abasic site. In some embodiments, the detectable moiety is directly or indirectly associated with an abasic site. In some embodiments, the detectable moiety (label) is directly or indirectly detectable. In some embodiments, the detectable signal is amplified. In some embodiments, the detectable moiety comprises an organic molecule such as a chromophore, a fluorophore, biotin or a derivative thereof. In other embodiments, the detectable moiety comprises a macromolecule such as a nucleic acid, an aptamer, a peptide, or a protein such as an enzyme or an antibody. In other embodiments, the detectable signal is fluorescent. In other embodiments, the detectable signal is enzymatically generated. In some embodiments, the label is selected from, fluorescein, rhodamine, a cyanine dye, an indocyanine dye, Cy3, Cy5, an Alexa Fluor dye, phycoerythrin, 5-(((2-(carbohydrazino)-methyl)thio)acetyl)aminofluorescein, aminooxyacetyl hydrazide ("FARP"), or N-(aminooxyacetyl)-N′-(D-biotinoyl) hydrazine, trifluoroacetic acid salt (ARP).

[0087] The cleavage of the input nucleic acid template comprising one or more abasic sites may further be provided by the use of enzymatic or chemical means or by the application of heat, or a combination thereof. For example the input nucleic acid template comprising one or more abasic sites may be treated with a nucleophile or a base. In some cases, the nucleophile is an amine such as a primary amine, a secondary amine, or a tertiary amine. For example, the abasic site may be treated with piperidine, moropholine, or a combination thereof. In some cases, hot piperidine (e.g., 1M at 90°C) may be used to cleave the input nucleic acid template comprising one or more abasic sites. In some cases, morpholine (e.g., 3M at 37°C or 65°C) may be used to cleave the input nucleic acid template comprising one or more abasic sites. Alternatively, a polyamine may be used to cleave the input nucleic acid template comprising one or more abasic sites. Suitable polyamines include for example spermine, spermidine, 1,4-diaminobutane, lysine, the tripeptide K-W-K, N, N-dimethylethylenediamine (DMED), piperazine, 1,2-ethylenediamine, or any combination thereof. In some cases, the input nucleic acid template comprising one or more abasic sites may be treated with a reagent suitable for carrying out a beta elimination reaction, a delta

IDT Exhibit 1020.5585

elimination reaction, or a combination thereof. In some cases, the cleavage of input nucleic acid template comprising one or more abasic sites by chemical means may provide fragments of input nucleic acid template, which fragments comprise a blocked 3' end. In some cases, the blocked 3' end lacks a terminal hydroxyl. In other cases, the blocked 3' end is phosphorylated. In still other cases, cleavage of the input nucleic acid template comprising one or more abasic sites by chemical means may provide fragments of input nucleic acid template that are not blocked. In some cases, the methods of the present invention provide for the use of an enzyme or combination of enzymes and a polyamine such as DMED under mild conditions in a single reaction mixture which does not affect the canonical nucleotides and therefore may maintain the sequence integrity of the products of the method. Suitable mild conditions may include conditions at or near neutral pH. Other suitable conditions include pH of about 4.5 or higher, 5 or higher, 5.5 or higher, 6 or higher, 6.5 or higher, 7 or higher, 7.5 or higher, 8 or higher, 8.5 or higher, 9 or higher, 9.5 or higher, 10 or higher, or about 10.5 or higher. Still other suitable conditions include between about 4.5 and 10.5, between about 5 and 10.0, between about 5.5 and 9.5, between about 6 and 9, between about 6.5 and 8.5, between about 6.5 and 8.0, or between about 7 and 8.0. Suitable mild conditions also may include conditions at or near room temperature. Other suitable conditions include a temperature of about 10°C, 11°C, 12°C, 13°C, 14°C, 15°C, 16°C, 17°C, 18°C, 19°C, 20°C, 21°C, 22°C, 23°C, 24°C, 25°C, 26°C, 27°C, 28°C, 29°C, 30°C, 31°C, 32°C, 33°C, 34°C, 35°C, 36°C, 37°C, 38°C, 39°C, 40°C, 41°C, 42°C, 43°C, 44°C, 45°C, 46°C, 47°C, 48°C, 49°C, 50°C, 51°C, 52°C, 53°C, 54°C, 55°C, 56°C, 57°C, 58°C, 59°C, 60°C, 61°C, 62°C, 63°C, 64°C, 65°C, 66°C, 67°C, 68°C, 69°C, or 70°C or higher. Still other suitable conditions include between about 10°C and about 70°C , between about 15°C and about 65°C , between about 20°C and about 60°C , between about 20°C and about 55°C , between about 20°C and about 50°C, between about 20°C and about 45°C , between about 20°C and about 40°C , between about 20°C and about 35°C , or between about 20°C and about 30°C. In some cases, the use of mild cleavage conditions may provide for less damage to the primer extension products produced by the methods of the present invention. In some cases, the fewer damaged bases, the more suitable the primer extension products may be for downstream analysis such as sequencing, or hybridization. In other cases, the use of mild cleavage conditions may increase final product yields, maintain sequence integrity, or render the methods of the present invention more suitable for automation.

[0088] In embodiments involving fragmentation, the backbone of the template polynucleotide comprising the abasic site is cleaved at the abasic site, whereby two or more fragments of the polynucleotide are generated. At least one of the fragments comprises an abasic site, as described herein. Agents that cleave the phosphodiester backbone of a polynucleotide at an abasic site are provided herein. In some embodiments, the agent is an AP endonuclease such as E. coli AP endonuclease IV. In other embodiments, the agent is N,N′-dimethylethylenediamine (termed "DMED"). In other embodiments, the agent is heat, basic condition, acidic conditions, or an alkylating agent. In still other embodiments, the agent that cleaves the phosphodiester backbone at an abasic site is the same agent that cleaves the base portion of a nucleotide to form an abasic site. For example, glycosidases of the present invention may comprise both a glycosidase and a lyase activity, whereby the glycosidase activity cleaves the base portion of a nucleotide (e.g., a non-canonical nucleotide) to form an abasic site and the lyase activity cleaves the phosphodiester backbone at the abasic site so formed. In some cases, the glycosidase comprises both a glycosidase activity and an AP endonuclease activity.

[0089] Depending on the agent employed for cleaving at the abasic site of the template polynucleotide, the backbone can be cleaved 5′ to the abasic site (e.g., cleavage between the 5′-phosphate group of the abasic residue and the deoxyribose ring of the adjacent nucleotide, generating a free 3′ hydroxyl group), such that an abasic site is

IDT Exhibit 1020.5586

located at the 5′ end of the resulting fragment. In other embodiments, cleavage can also be 3′ to the abasic site (e.g., cleavage between the deoxyribose ring and 3′-phosphate group of the abasic residue and the deoxyribose ring of the adjacent nucleotide, generating a free 5′ phosphate group on the deoxyribose ring of the adjacent nucleotide), such that an abasic site is located at the 3′ end of the resulting fragment. In still other embodiments, more complex forms of cleavage are possible, for example, cleavage such that cleavage of the phosphodiester backbone and cleavage of a portion of the abasic nucleotide results. Selection of the fragmentation agent thus permits control of the orientation of the abasic site within the polynucleotide fragment, for example, at the 3′ end of the resulting fragment or the 5′ end of the resulting fragment. Selection of reaction conditions also permits control of the degree, level or completeness of the fragmentation reactions. In some embodiments, reaction conditions can be selected such that the cleavage reaction is performed in the presence of a large excess of reagents and allowed to run to completion with minimal concern about cleavage of the primer extension products of the invention. By contrast, other methods known in the art, e.g., mechanical shearing, DNase cleavage, cannot distinguish between the template polynucleotide and the primer extension products. In other embodiments, reaction conditions are selected such that fragmentation is not complete (in the sense that the backbone at some abasic sites remains uncleaved (unfragmented)), such that polynucleotide fragments comprising more than one abasic site are generated. Such fragments comprise internal (nonfragmented) abasic sites.

[0090] Following generation of an abasic site by cleavage of the base portion of the non-canonical nucleotide if present in the polynucleotide, the backbone of the polynucleotide is cleaved at the site of incorporation of the non-canonical nucleotide (also termed the abasic site, following cleavage of the base portion of the non-canonical nucleotide) with an agent capable of effecting cleavage of the backbone at the abasic site. Cleavage at the backbone (also termed "fragmentation") results in at least two fragments (depending on the number of abasic sites present in the polynucleotide comprising an abasic site, and the extent of cleavage).

[0091] Suitable agents (for example, an enzyme, a chemical and/or reaction conditions such as heat) capable of cleavage of the backbone at an abasic site include: heat treatment and/or chemical treatment (including basic conditions, acidic conditions, alkylating conditions, or amine mediated cleavage of abasic sites, (see e.g., McHugh and Knowland, Nucl. Acids Res. (1995) 23(10):1664-1670; Bioorgan. Med. Chem (1991) 7:2351; Sugiyama, Chem. Res. Toxicol. (1994) 7: 673-83; Horn, Nucl. Acids. Res., (1988) 16:11559-71), and use of enzymes that catalyze cleavage of polynucleotides at abasic sites, for example AP endonucleases (also called "apurinic, apyrimidinic endonucleases") (e.g., E. coli Endonuclease IV, available from Epicentre Tech., Inc, Madison Wis.), E. coli endonuclease III or endonuclease IV, E. coli exonuclease III in the presence of calcium ions. See, e.g. Lindahl, PNAS (1974) 71(9):3649-3653; Jendrisak, U.S. Pat. No. 6,190,865 B1; Shida, Nucleic Acids Res. (1996) 24(22):4572-76; Srivastava, J. Biol Chem. (1998) 273(13):21203-209; Carey, Biochem. (1999) 38:16553-60; Chem Res Toxicol (1994) 7:673-683. As used herein "agent" encompasses reaction conditions such as heat. In one embodiment, the AP endonuclease, E. coli endonuclease IV, is used to cleave the phosphodiester backbone at an abasic site. In another embodiment, cleavage is with an amine, such as N, N′-dimethylethylenediamine. See, e.g., McHugh and Knowland, supra.

[0092] Cleavage of the abasic site may occur between the nucleotide immediately 5′ to the abasic residue and the abasic residue, or between the nucleotide immediately 3′ to the abasic residue and the abasic residue (though, as explained herein, 5′ or 3′ cleavage of the phosphodiester backbone may or may not result in retention of the phosphate group 5′ or 3′ to the abasic site, respectively, depending on the fragmentation agent used). Cleavage can be 5′ to the abasic site (such as endonuclease IV treatment which generally results in cleavage of the backbone at a

IDT Exhibit 1020.5587

location immediately 5′ to the abasic site between the 5′-phosphate group of the abasic residue and the deoxyribose ring of the adjacent nucleotide, generating a free 3′ hydroxyl group on the adjacent nucleotide), such that an abasic site is located at the 5′ end of the resulting fragment. Cleavage can also be 3′ to the abasic site (e.g., cleavage between the deoxyribose ring and 3′-phosphate group of the abasic residue and the deoxyribose ring of the adjacent nucleotide, generating a free 5′ phosphate group on the deoxyribose ring of the adjacent nucleotide), such that an abasic site is located at the 3′ end of the resulting fragment. Treatment under basic conditions or with amines (such as N,N′-dimethylethylenediamine) results in cleavage of the phosphodiester backbone immediately 3′ to the abasic site. In addition, more complex forms of cleavage are also possible, for example, cleavage such that cleavage of the phosphodiester backbone and cleavage of (a portion of) the abasic nucleotide results. For example, under certain conditions, cleavage using chemical treatment and/or thermal treatment may comprise a β-elimination step which results in cleavage of a bond between the abasic site deoxyribose ring and its 3′ phosphate, generating a reactive α,β-unsaturated aldehyde which can be labeled or can undergo further cleavage and cyclization reactions. See, e.g., Sugiyama, Chem. Res. Toxicol . (1994) 7: 673-83; Horn, Nucl. Acids. Res., (1988) 16:11559-71. It is understood that more than one method of cleavage can be used, including two or more different methods which result in multiple, different types of cleavage products (e.g., fragments comprising an abasic site at the 3′ end, and fragments comprising an abasic site at the 5′ end).

[0093] Cleavage of the backbone at an abasic site may be general, specific or selective (in the sense that the agent (such as an enzyme) capable of cleaving the backbone at an abasic site specifically or selectively cleaves the base portion of a particular non-canonical nucleotide), whereby greater than about 98%, about 95%, about 90%, about 85%, or about 80% of the cleavage is at an abasic site. However, extent of cleavage can be less. Thus, reference to specific cleavage is exemplary. General, specific or selective cleavage is desirable for control of the fragment size in the methods of generating labeled polynucleotide fragments of the invention. In some embodiments, reaction conditions can be selected such that the cleavage reaction is performed in the presence of a large excess of reagents and allowed to run to completion with minimal concern about excessive cleavage of the polynucleotide (i.e., while retaining a desired fragment size, which is determined by spacing of the incorporated non-canonical nucleotide, during the synthesis step, above). In other embodiments, extent of cleavage can be less, such that polynucleotide fragments are generated comprising an abasic site at an end and an abasic site(s) within or internal to the polynucleotide fragment (i.e., not at an end).

[0094] In embodiments involving cleavage of the phosphodiester backbone, appropriate reaction media and conditions for carrying out the cleavage of the phosphodiester backbone at an abasic site according to the methods of the invention are those that permit cleavage of the phosphodiester backbone at an abasic site. Such media and conditions are known to persons of skill in the art, and are described in various publications, such as Bioorgan. Med. Chem (1991) 7:2351; Sugiyama, Chem. Res. Toxicol . (1994) 7: 673-83; Horn, Nucl. Acids. Res., (1988) 16:11559-71); Lindahl, PNAS (1974) 71(9):3649-3653; Jendrisak, U.S. Pat. No. 6,190,865 B1; Shida, Nucleic Acids Res. (1996) 24(22):4572-76; Srivastava, J. Biol Chem. (1998) 273(13):21203-209; Carey, Biochem. (1999) 38:16553-60; Chem Res Toxicol (1994) 7:673-683.

[0095] In some cases, nucleic acids containing abasic sites are heated in a buffer solution containing an amine, for example, 25 mM Tris-HCl and 1-5 mM magnesium ions, for 10-30 minutes at 70°C. to 95°C. Alternatively, 1.0 M piperidine (a base) is added to polynucleotide comprising an abasic site which has been precipitated with ethanol and vacuum dried. The solution is then heated for 30 minutes at 90°C. and lyophilized to remove the piperidine. In another example, cleavage is effected by treatment with basic solution, e.g., 0.2 M sodium hydroxide at 37°C for 15

IDT Exhibit 1020.5588

minutes. See Nakamura (1998) Cancer Res. 58:222-225. In yet another example, incubation at 37°C. with 100 mM N,N′-dimethylethylenediamine acetate, pH 7.4 is used to cleave. See McHugh and Knowland, (1995) Nucl. Acids Res. 23(10) 1664-1670.

[0096] The cleavage of the input nucleic acid template comprising one or more abasic sites may also be performed by enzymatic means. For example an apyrimidinic endonuclease or an apurinic endonuclease (collectively known as AP endonucleases) may be used to cleave the input nucleic acid template at the one or more abasic sites. In some cases, the input nucleic acid template comprising one or more abasic sites may be cleaved with a class I, class II, class III, or class IV AP endonuclease or a combination thereof. In some cases, the cleavage of input nucleic acid template comprising one or more abasic sites by enzymatic means may provide fragments of input nucleic acid template, which fragments comprise a blocked 3' end. In some cases, the blocked 3' end lacks a terminal hydroxyl. In other cases, the blocked 3' end is phosphorylated. In still other cases, cleavage of the input nucleic acid template comprising one or more abasic sites by enzymatic means may provide fragments of input nucleic acid template that are not blocked.

[0097] In some cases, the cleavage may be performed by use of a glycosylase and a nucleophile, or a glycosylase and an amine, or a glycosylase and an AP endonuclease such as for example UDG and DMED or UDG and an AP endonuclease at the same time. Alternatively, the input nucleic acid template comprising one or more non-canonical nucleotides may first be treated with a glycosylase to produce one or more abasic sites, and then be treated with an AP endonuclease or cleaved by chemical means. In some cases, the hybridization, and extension reactions are performed first, and then the cleavage reaction is performed after sufficient time. In other cases, the hybridization and extension reactions are performed at the same time as the cleavage reactions. In still other cases, the hybridization and extension reactions are initiated and allowed to proceed for a set period of time (e.g., 1 minute, 2 minutes, 3 minutes, 5 minutes, 10 minutes, 15 minutes, 30 minutes, 1 hour, 2 hours, 3 hours etc.) and then the cleavage reaction is initiated. In some cases, initiation of the cleavage reaction may stop the extension reaction; in other cases, the cleavage reaction and the extension reaction may then proceed concurrently.

[0098] For example, *E. coli* AP endonuclease IV may be added to reaction conditions as described above. AP Endonuclease IV can be added at the same or different time as the agent (such as an enzyme) capable of cleaving the base portion of a non-canonical nucleotide. For example, AP Endonuclease IV can be added at the same time as UNG, or at different times. Alternatively, the template nucleic acid or a reaction mixture comprising template nucleic acid may be treated with UNG and an amine at the same time. A reaction mixture suitable for simultaneous UNG treatment and N,N′-dimethylethylenediamine treatment may include about 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, 35, 40, or about 50mM DMED. Alternatively, the use of an agent that comprises both glycosidase and lyase activity may be utilized in the reaction mixture to cleave the input nucleic acid template.

[0099] Cleavage of the input nucleic acid template by chemical means, enzymatic means, or a combination thereof may provide a mixture of double stranded products, single stranded products, and partial duplexes. In some cases, the cleaved products of the cleavage reaction may be removed by one or more methods of the present invention. In some cases, the cleaved products of the cleavage reaction may be removed by purification. For example, the cleaved products of the cleavage reaction may be removed by a size-dependent purification method, or an affinity based purification method. For example, the single stranded nucleic acids may be removed by an affinity hybridization step to capture probes. In some cases, the capture probes may be hybridized to a solid substrate. In other cases, the cleaved nucleic acid products of the cleavage reaction may be removed by an affinity capture step

IDT Exhibit 1020.5589

using a ligand with affinity to a label that has been incorporated into the cleaved products of the cleavage reaction. The label, or ligand, may be incorporated prior to cleavage (e.g. during synthesis of the template nucleic acid), during cleavage, or after the cleavage step. In some cases, the label may be incorporated at the abasic site. In other cases, the cleaved nucleic acid products of the cleavage reaction may be removed by a capture step using a reactive moiety (e.g., an amine or a hydrazine) such as an immobilized reactive moiety that reacts with a reactive α,β-unsaturated aldehyde present at the abasic site of the cleaved nucleic acid product of the cleavage reaction. In some cases, the cleaved nucleic acid products of the cleavage reaction may be removed by electrophoresis or ultrafiltration.

[00100] In other cases, the single stranded products may be removed by enzymatic means. For example, a single stranded specific exonuclease or endonuclease can be used to cleave the single stranded DNA. A variety of suitable single stranded DNA specific exonucleases are suitable for the methods of the present invention such as for example exonuclease 1, and exonuclease 7. Similarly a variety of suitable single stranded DNA specific endonucleases are suitable for the methods of the present invention such as for example the single stranded DNA specific endonuclease is a S1 endonuclease or a mung bean nuclease . In some cases, any combination of single strand specific endonucleases or exonucleases known in the art such as those provided herein may be utilized to degrade or remove single stranded products, such as single stranded fragmentation products or single stranded primer extension products or a combination thereof.

[00101] In some cases, the products of the primer extension reaction generated in the methods of the present invention may be purified from the reaction mixture comprising fragmented target nucleic acid and primer extension products. For example, the primer extension step may include the use of nucleotides comprising a purification label such as for example biotin/avidin or any other suitable label (e.g. digoxin, fluorescein, an antigen, a ligand, a receptor, or any nucleotide labels provided herein). Primer extension products may therefore be understood to contain a member of the biotin/avidin ligand receptor pair or other purification label, whereas primers and template nucleic acid may not. A simple purification step may be performed to remove unincorporated nucleotides such as alcohol or polyethylene glycol precipitation, ion exchange purification, ultrafiltration, silica absorption, or reverse phase methods, and then the primer extension products may be recovered using an appropriate affinity matrix such as a matrix comprising biotin or a derivative thereof, avidin or a derivative thereof, streptavidin or a derivative thereof, an antibody or a derivative or fragment thereof, an antigen, a ligand, or a receptor in the form of particles, beads, a membrane or a column. Alternatively, the simple purification step to remove unincorporated nucleotides may be omitted or performed after the affinity purification step.

[00102] In some embodiments, the methods of the present invention further provide for the generation of one or more blunt ended double stranded products. In some embodiments the blunt ended double stranded products are produced from a template not containing any non-canonical nucleotides. In other embodiments the double stranded products are produced from a template containing one or more non-canonical nucleotides. In some cases, the extension step of the present invention directly provides blunt ended double stranded products. In other cases, the extension step of the present invention provides a mixture of blunt ended and non-blunt ended double stranded products. In still other cases, the extension step does not provide blunt ended double stranded products, or does not provide a substantial degree or amount of blunt ended double stranded products. In some cases, the non-blunt ended products of the primer extension reaction must be further treated by the methods of the present invention to produce blunt ended double stranded products, or to convert a substantial fraction of the non-blunt ended products to blunt ended products.

IDT Exhibit 1020.5590

[00103] In some cases, the double stranded products generated by the method of the present invention may be blunt ended, when blunt end dsDNA is desirable for downstream analysis such as highly parallel sequencing, or other cloning or adaptor ligation applications, by the use of a single strand specific DNA exonuclease such as for example exonuclease 1, exonuclease 7 or a combination thereof to degrade overhanging single stranded ends of the double stranded products. Alternatively, the double stranded products may be blunt ended by the use of a single stranded specific DNA endonuclease for example but not limited to mung bean endonuclease or S1 endonuclease. Alternatively, the double stranded fragment products may be blunt ended by the use of a polymerase that comprises single stranded exonuclease activity such as for example T4 DNA polymerase, any other polymerase comprising single stranded exonuclease activity or a combination thereof to degrade the overhanging single stranded ends of the double stranded products. In some cases, the polymerase comprising single stranded exonuclease activity may be incubated in a reaction mixture that does or does not comprise one or more dNTPs. In other cases, a combination of single stranded nucleic acid specific exonucleases and one or more polymerases may be used to blunt end the double stranded products of the primer extension reaction. In still other cases, the products of the extension reaction may be made blunt ended by filling in the overhanging single stranded ends of the double stranded products. For example, the fragments may be incubated with a polymerase such as T4 DNA polymerase or Klenow polymerase or a combination thereof in the presence of one or more dNTPs to fill in the single stranded portions of the double stranded products. Alternatively, the double stranded products may be made blunt by a combination of a single stranded overhang degradation reaction using exonucleases and/or polymerases, and a fill-in reaction using one or more polymerases in the presence of one or more dNTPs.

[00104] In some embodiments, the methods of the present invention provide for generation of primer extension products comprising double stranded nucleic acids, single stranded nucleic acids, and nucleic acids comprising partial double stranded and partial single stranded portions, either from a template not comprising any non-canonical nucleotides or from a template nucleic acid comprising one or more non-canonical nucleotides; fragmentation of the template nucleic acid; optional purification of the primer extension products; and generation of double stranded products from the single stranded nucleic acid primer extension products and/or from the primer extension products comprising partial double stranded and partial single stranded portions. Methods for generation of double stranded products from partial double stranded products are provided herein including the methods for blunt ending double stranded primer extension products. Methods for generation of double stranded primer extension products from single stranded primer extension products include for example annealing one or more primers, such as any of the primers provided herein, to the single stranded primer extension product and extending the one or more annealed primers with a polymerase, such as any of the polymerases provided herein or a any suitable polymerase in a reaction mixture comprised of one or more dNTPs, including labeled dNTPs, canonical dNTPs, non-canonical dNTPs or a combination thereof. In some cases, the non-canonical nucleotides utilized in the reaction mixture for generating double stranded products from single stranded primer extension products or from partial double stranded products are different from at least one of the non-canonical nucleotides present in the template polynucleotide. Methods of generation of double stranded primer extension products from single stranded primer extension products may further include for example annealing two or more adjacent primers, such as any of the primers provided herein including random primers (e.g. pentamers, hexamers, heptamers, octamers, nonamers, decamers, undecamers, dodecamers, tridecamers etc.), to the single stranded primer extension product and ligating the adjacent primers. Methods for generating double stranded primer extension products from single stranded primer extension products may further include for example annealing one or more primers such as any of the primers provided herein including

IDT Exhibit 1020.5591

primers comprising random hybridizing portions (e.g. random pentamers, hexamers, heptamers, octamers, nonamers, decamers, undecamers, dodecamers, tridecamers etc.) to the single stranded primer extension product and extending the annealed primers. In some cases, the extension step may be performed using an enzyme (e.g., a DNA dependent DNA polymerase) comprising strand displacement activity.

[00105] In some embodiments, the methods of the present invention provide for attachment (e.g., ligation) of adaptor molecules to the double stranded DNA products of the primer extension reaction, or double stranded products generated from the single stranded or partially double stranded products of the primer extension reaction. The adaptor molecules may be ligated to double stranded DNA fragment molecules comprising single stranded overhangs, including but not limited to single, double, triple, quadruple, quintuple, sextuple, septuple, octuple, or more base overhangs, or to double stranded DNA fragment molecules comprising blunt ends. In some cases, the adaptor molecules are ligated to blunt end double stranded DNA fragment molecules which have been modified by 5' phosphorylation. In some cases, the adaptor molecules are ligated to blunt end double stranded DNA fragment molecules which have been modified by 5' phosphorylation followed by extension of the 3' end with one or more nucleotides. In some cases, the adaptor molecules are ligated to blunt end double stranded DNA fragment molecules which have been modified by 5' phosphorylation followed by extension of the 3' end with a single nucleotide (or 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 or more) such as for example adenine, guanine, cytosine, or thymine. In still other cases, adaptor molecules can be ligated 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. 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.

[00106] The adaptor molecules may comprise single or double stranded nucleic acids or a combination thereof. In some cases, the adaptor molecules comprise a one, two, three, four, five, six, seven, eight, nine, ten, eleven, twelve, thirteen, fourteen, fourteen, fifteen, sixteen, seventeen, eighteen, nineteen, twenty or longer base long single stranded overhang at their 5' ends. For example, the adaptor molecules may comprise a one base long thymine, adenine, cytosine, or guanine overhang at their 5' ends. Adaptor molecule compositions are provided herein.

[00107] In some embodiments, the methods of the present invention provide for ligation or attachment of adaptor molecules to the single stranded DNA products of the extension reaction. The adaptor molecules may comprise single stranded or double stranded nucleic acids or a combination thereof. The adaptor molecules may be ligated to the single stranded DNA products of the extension reaction using T4 RNA ligase which is capable of ligating two single stranded nucleic acids (RNA or DNA) together in the absence of a template. Alternatively, a single stranded DNA specific ligase such as for example CircLigase® may be utilized in the methods of the present invention.

[00108] In some embodiments, the methods of the present invention provide for contacting an input nucleic acid template comprising one or more non-canonical nucleotides with a reaction mixture. In some cases, the reaction mixture may comprise one or more oligonucleotide primers as provided herein. For example, the reaction mixture may comprise one or more oligonucleotide primers comprising random hybridizing portions. Additionally, the reaction mixture may comprise one or more oligonucleotide primers comprising random hybridizing portions and one or more oligonucleotide primers comprising a polyT sequence.

IDT Exhibit 1020.5592

[00109] In some cases, the reaction mixture may comprise one or more polymerases as provided herein. For example, the reaction mixture may comprise one or more polymerases comprising strand displacement activity, such as for example, Klenow polymerase, exo – Klenow polymerase, 5'-3' exo – Klenow polymerase, Bst polymerase, Bst large fragment polymerase, Vent polymerase, Deep Vent (exo-) polymerase, 9°Nm polymerase, Therminator polymerase, Therminator II polymerase, MMulV Reverse Transcriptase, phi29 polymerase, or DyNAzyme EXT polymerase, or a combination thereof. In some cases, the reaction mixture may be configured to provide double stranded products in the presence of the input nucleic acid template, the one or more oligonucleotide primers, and the one or more polymerases comprising strand displacement activity. Enzymes for use in the compositions, methods and kits of the present invention may further include any enzyme having reverse transcriptase activity. Such enzymes include, but are not limited to, retroviral reverse transcriptase, retrotransposon reverse transcriptase, hepatitis B reverse transcriptase, cauliflower mosaic virus reverse transcriptase, bacterial reverse transcriptase, E. coli DNA polymerase and klenow fragment, Tth DNA polymerase, Taq DNA polymerase (Saiki, R. K., et al., Science 239:487-491 (1988); U.S. Pat. Nos. 4,889,818 and 4,965,188), Tne DNA polymerase (WO 96/10640), Tma DNA polymerase (U.S. Pat. No. 5,374,553), C. Therm DNA polymerase from Carboxydothermus hydrogenoformans (EP0921196A1, Roche, Pleasanton, Calif., Cat. No. 2016338), ThermoScript (Invitrogen, Carsbad, Calif. Cat. No. 11731-015) and mutants, fragments, variants or derivatives thereof. As will be understood by one of ordinary skill in the art, modified reverse transcriptases may be obtained by recombinant or genetic engineering techniques that are routine and well-known in the art. Mutant reverse transcriptases can, for example, be obtained by mutating the gene or genes encoding the reverse transcriptase of interest by site-directed or random mutagenesis. Such mutations may include point mutations, deletion mutations and insertional mutations. Preferably, one or more point mutations (e.g., substitution of one or more amino acids with one or more different amino acids) are used to construct mutant reverse transcriptases of the invention. Fragments of reverse transcriptases may be obtained by deletion mutation by recombinant techniques that are routine and well-known in the art, or by enzymatic digestion of the reverse transcriptase(s) of interest using any of a number of well-known proteolytic enzymes. Mutant DNA polymerase containing reverse transcriptase activity can also be used as described in U.S. Patent Application Ser. No. 10/435,766, incorporated herein by reference.

[00110] In some cases, the reaction mixture may comprise one or more agents capable of cleaving the base portion of a non-canonical nucleotide to generate an abasic site. In some cases, the reaction mixture may contain the one or more agents capable of cleaving the base portion of a non-canonical nucleotide to generate an abasic site at the initiation of the extension reaction. In some cases, the reaction mixture may be supplemented with the one or more agents capable of cleaving the base portion of a non-canonical nucleotide to generate an abasic site after a suitable period of time (e.g., about 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 45, 60, 90, 120, 180, 240, 300, 400, 500, 600 minutes) has passed for the generation of primer extension products. Suitable agents capable of cleaving the base portion of a non-canonical nucleotide to generate an abasic site include but are not limited to UDG and MPG.

[00111] In some cases, the reaction mixture may comprise one or more agents capable of fragmenting a phosphodiester backbone at an abasic site to fragment the input nucleic acid template. In some cases, the reaction mixture may contain the one or more agents capable of fragmenting a phosphodiester backbone at an abasic site to fragment the input nucleic acid template at the initiation of the extension reaction. In some cases, the reaction mixture may be supplemented with the one or more agents capable of fragmenting a phosphodiester backbone at an abasic site to fragment the input nucleic acid template after a suitable period of time (e.g., about 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 45, 60, 90, 120, 180, 240, 300, 400, 500, 600 minutes) has passed for the generation of primer extension

IDT Exhibit 1020.5593

products. Suitable agents capable of fragmenting a phosphodiester backbone at an abasic site to fragment the input nucleic acid template include but are not limited to an amine, a primary amine, a secondary amine, a polyamine as provided herein, a nucleophile, a base (e.g. NaOH), piperidine, hot piperidine, and one or more AP endonucleases.

[00112] The methods of the present invention provide for downstream analysis of the primer extension products generated in the methods of the present invention. Said downstream analysis includes but is not limited to e.g. pyrosequencing, sequencing by synthesis, sequencing by hybridization, single molecule sequencing, nanopore sequencing, and sequencing by ligation, high density PCR, microarray hybridization, SAGE, digital PCR, and massively parallel Q-PCR; subtractive hybridization; differential amplification; comparative genomic hybridization, preparation of libraries (including cDNA and differential expression libraries); preparation of an immobilized nucleic acid (which can be a nucleic acid immobilized on a microarray), and characterizing amplified nucleic acid products generated by the methods of the invention, or a combination thereof.

[00113] Large scale sequencing methods provided by the present invention include methods such as those provided by the Genome Analyzer by Illumina, the Genome Sequencer by 454/Roche Lifesciences, the SMRT platform by Pacific Biosciences, the HeliScope by Helicos Biosciences, the Polonator by Dover Systems, and semiconductor sequencing (Personal Genome Machine (PGM™) by Ion Torrent.

[00114] In some embodiments, the methods are useful for preparing target polynucleotide(s) for sequencing by synthesis using the methods commercialized by 454/Roche Lifesciences 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; 7,323,305. In general, polynucleotides may be prepared by primer extension from a template comprising non-canonical nucleotides according to the methods of the present invention, which methods may further include cleavage of the template nucleic acid, purification of the primer extension products, adaptor ligation, further amplification, or a combination thereof. The primer extension products may then be immobilized onto beads, and compartmentalized in a water-in-oil emulsion suitable for amplification by PCR. In some cases, alternative amplification methods other than PCR may 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 may include a step of rendering the nucleic acid bound to the beads single stranded or partially single stranded. The beads may 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 results 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.

[00115] 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. Patent Application Ser. No. 11/167,046, 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, polynucleotides may be prepared by primer extension from a template comprising non-canonical nucleotides according to the methods of the present invention, which methods may further include cleavage of template nucleic acid, purification of the primer extension products, adaptor ligation, further amplification, or a combination thereof. The primer extension products may then be immobilized onto a flow-cell surface. The methods may 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

IDT Exhibit 1020.5594

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.

[00116] 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, polynucleotides may be prepared by primer extension from a template comprising non-canonical nucleotides according to the methods of the present invention, which methods may further include cleavage of template nucleic acid, purification of the primer extension products, adaptor ligation, further amplification, or a combination thereof. The primer extension products may then be incorporated into a water in oil emulsion along with polystyrene beads and amplified by for example PCR. In some cases, alternative amplification methods may 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 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 may include a step of rendering the nucleic acid bound to the beads 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 form 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.

[00117] In some embodiments, the methods are useful for preparing target polynucleotide(s) for sequencing by the sequencing by ligation methods commercialized by Dover Systems. In general, polynucleotides may be prepared by primer extension from a template comprising non-canonical nucleotides according to the methods of the present invention, which methods may further include cleavage of template nucleic acid, purification of the primer extension products, adaptor ligation, further amplification, or a combination thereof. The polynucleotides may then amplified in an emulsion in the presence of magnetic beads. Any amplification methods may 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 may 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 cells, 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 being 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

IDT Exhibit 1020.5595

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.

[00118] In some embodiments, the methods are useful for preparing target polynucleotide(s) for sequencing by the methods commercialized by Illumina as described U.S. Patent Nos. 5,750,341; 6,306,597; and 5,969,119. In general, polynucleotides may be prepared by primer extension from a template comprising non-canonical nucleotides according to the methods of the present invention, which methods may further include cleavage of template nucleic acid, purification of the primer extension products, adaptor ligation, further amplification, or a combination thereof. The methods of the present invention may provide 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. The initiate the first base sequencing cycle, four labeled reversible terminators, primers, and DNA polymerase are added. After laser excitation, florescence form 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.

[00119] 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. Patent Nos. 7462452; 7476504; 7405281; 7170050; 7462468; 7476503; 7315019; 7302146; 7313308; and U.S. Patent Application Publication Nos. US20090029385; US20090068655; US20090024331; and US20080206764. In general, polynucleotides may be prepared by primer extension from a template comprising non-canonical nucleotides according to the methods of the present invention, which methods may further include cleavage of template nucleic acid, purification of the primer extension products, adaptor ligation, further amplification, or a combination thereof. The polynucleotides may 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.

[00120] In some embodiments, the methods of the present invention are useful for preparing target polynucleotide(s) for sequencing by the methods commercialized by Ion Torrent Systems. Such methods are described, for example, in U.S. Patent Application Publication Nos. 20100197507, 20100188073, 20100137143, 20100035252, 20090127589, and 20090026082. Ion Torrent Systems technology can use chemical-sensitive field

IDT Exhibit 1020.5596

effect transistors (FETs). Ion Torrent Systems technology can include use of a semiconductor chip that comprises multiple layers, e.g., a layer with micro-machined wells, an ion-sensitive layer, and an ion sensor layer. Nucleic acids attached to beads can be introduced into the micro-machined wells. A clonal population of single nucleic acids can be attached to a single bead. One type of deoxyribonucleotide (e.g., dATP, dCTP, dGTP, or dTTP) can be introduced into the micro-machined wells to initiate sequencing of the nucleic acids on the beads. Upon incorporation of nucleotides by DNA polymerase, protons 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 micro-machined wells of a semiconductor chip.

[00121] In some embodiments, the methods of the present invention are useful for preparing target polynucleotide(s) for analysis by array comparative genomic hybridization as described for example in U.S. Patent Nos. 5856097, 5965362, 5976790, 5665549, 6335167, 6159685, 7238484, 7534567, 7537895; and U.S. Patent Application Publication Nos. US20050118634, US20050260645, US20050282227, US20060292608, US20070087355, US20070134676, US20070160988, US20070219727, and US20090069191. In general, polynucleotides may be prepared by primer extension from a template comprising non-canonical nucleotides according to the methods of the present invention, which methods may further include cleavage of template nucleic acid, purification of the primer extension products, adaptor ligation, further amplification, or a combination thereof. Specific cleavage of the template nucleic acid and optional purification of the primer extension products from the cleaved template nucleic acid may provide for elimination or reduction of potential cross-hybridization of array probes with non-labeled input template, as well as increased specific activity of the labeled targets for hybridization to array probes. The primer extension products may further be labeled by any methods known in the art including methods provided herein including by incorporation of labeled nucleotides by a polymerase during primer extension, during further rounds of amplification of the primer extension products, or by end labeling. In some cases, incorporation of labeled nucleotides by a polymerase during primer extension may be performed by random priming of template nucleic acid comprising one or more non-canonical nucleotides, not so random priming of template nucleic acid comprising one or more non-canonical nucleotides, or non random priming of template nucleic acid comprising oen or more non-canonical nucleotides. In some cases, incorporation of labeled nucleotides may be provided by using primers which are labeled. In still other cases, incorporation of labeled nucleotides may be provided by performing further rounds of amplification such as any of the amplification means provided herein in the presence of one or more labeled nucleotides or labeled primers. End labeling reactions may be performed using a T4 RNA ligase in the presence of a suitable label such as a biotinylated or fluorophore conjugated nucleotide. Alternatively, end-labeling reactions may be performed with a terminal deoxynucleotide transferase in the presence of a suitable label such as biotinylated ddUTP or a fluorophore conjugated ddUTP.

[00122] In some cases, primer extension products prepared from different biological samples such as samples from two different organisms or samples from suspected or known normal tissue and samples from suspected or known disease or tumor tissue may be differentially labeled. For example, nucleic acid primer extension products provided from samples comprising normal tissue may be labeled with one detectable identifying characteristic (e.g. fluorophore, chromophore, biotin, avidin, antigen, antibody, protein, enzyme, alkaline phosphatase or horse radish peroxidase), while primer extension products from samples comprising disease or tumor tissue may be labeled with a different detectable identifying characteristic. In some cases, the differentially labeled primer extension products represent a control set of primer extension products and a test set of primer extension products. Insome cases, after

IDT Exhibit 1020.5597

differential labeling of the primer extension products of the present invention, the products may be mixed and contacted with an array of positionally addressable immobilized oligonucleotide probes. The degree of hybridization may be measured by standard methods known in the art such as fluorescence detection, and the data may be normalized, and analyzed to detect differences in the nucleic acids such as differences in copy number of nucleic acid sequences, or the presence or absence of SNPs or other polymorphisms such as variable nucleotide repeats, inserts, and deletions.

[00123] In the methods of the invention, the steps may be carried out in the order listed or, in some cases, may be carried out in a different order. In some methods a later step depends on the formation of a product from an earlier step, in which case such steps must be carried out in the order listed. One of ordinary skill in the art will understand which steps should be carried out in the order listed, and which steps can be carried out in a different order.

[00124] In the methods of the invention described herein, the analysis of the nucleic acid products generated may be automated such that the downstream reactions are performed via robotics. Once performed, data obtained may be provided to a personal computer, a personal digital assistant, a cellular phone, a video game system, or a television so that a user can monitor the progress of the sequencing reactions remotely. This process is illustrated, for example, in **Figure 6**. The performing, monitoring and obtaining of results of the downstream reactions can be done from a place other than where the apparatus is located, for example from a physically separate room, building, city, state, country or the like. Likewise, the results can be transmitted to the remote user from a physically separate room, building, city, state, country or the like.

[00125] In some embodiments, the components provided herein are added simultaneously at the initiation of the amplification process. In another embodiment, components are added in any order prior to or after appropriate time points during the amplification process, as required and/or permitted by the amplification reaction. Such time points, some of which are noted below, can be readily identified by a person of skill in the art. The enzymes used for nucleic acid amplification according to the methods of the invention can be added to the reaction mixture either prior to a target nucleic acid denaturation step, following a denaturation step, or following hybridization of the primer to the target nucleic acid, as determined by their thermal stability and/or other considerations known to the person of skill in the art. The oligonucleotide primer extension reactions can be performed consecutively or simultaneously with the other method steps of the present invention. In these embodiments, the reaction conditions and components may be varied between the different reactions.

[00126] In some embodiments, the methods can be stopped at various time points, and resumed at a later time. Said time points can be readily identified by a person of skill in the art. One time point is at the end of the oligonucleotide primer extension reaction. Another time point is at the end of the cleavage reaction. Another time point is at the end of any purification step. Methods for stopping the reactions include for example, cooling the reaction mixture to a temperature that inhibits enzyme activity, heating the reaction mixture to a temperature that destroys an enzyme, or purifying the desired nucleic acid or fragment thereof from the reaction mixture. Methods for resuming the reactions include, for example, raising the temperature of the reaction mixture to a temperature that permits enzyme activity, replenishing a destroyed (depleted) enzyme, or contacting the nucleic acid with a suitable reaction mixture. In some embodiments, one or more of the components of the reactions is replenished prior to, at, or following the resumption of the reactions. Alternatively, the reaction can be allowed to proceed (i.e., from start to finish) without interruption.

[00127] In some embodiments the reaction can be allowed to proceed without purification of intermediate complexes, for example, to remove primer, or single stranded cleavage fragments. Products can be purified at various time points, which can be readily identified by a person of skill in the art. One time point is at the end of the

IDT Exhibit 1020.5598

primer extension reaction. Another time point is at the end of input nucleic acid template cleavage. Yet another time point is at the end of ligation of adaptor molecules to the double stranded DNA fragments. In some embodiments, the removal of primers and/or template at the end of a defined step by enzymes with appropriate nuclease activities are also useful, for example, cleavage of the input nucleic acid template comprising one or more non-canonical nucleotides with a glycosylase and an amine.

[00128] **Strand specific (strand tracking) RNA amplification and direct cDNA sequencing (RNA-seq)**

[00129] One embodiment of a method for the generation of second strand cDNA with defined sequences at 3'- and 5'-ends is shown in **Figure 7** (sequence A' at the 3'-end and sequence B at the 5'-end).

[00130] In **Figure 7**, sequence A is the 5'-tail sequence of a first tailed primer, employed for the generation of first strand cDNA. The first primer comprises a 3'-sequence that can hybridize to an input RNA target (template) and a 5'-tail sequence A. The tail sequence A can be DNA.

[00131] Reverse transcription of the input RNA target (template) can be carried out by extension of a first primer or primers hybridized to the RNA target by an RNA-dependent DNA polymerase. The reaction can be carried out in the presence of actinomycin D, an inhibitor of the DNA-dependent DNA polymerase activity of the reverse transcriptase. The presence of actinomycin D can be used to establish extension of the hybridized primer to RNA targets only and prevent extension along a DNA target which may be in the sample comprising the RNA target (template). In addition, presence of actinomycin D can prevent primer extension along the first primer extension product, or products, to generate second strand synthesis in the reverse transcriptase first strand synthesis, thus establishing the strand specificity of the primer extension products generated in the reverse transcription reaction. There are several known DNA-dependent DNA polymerase inhibitors known in the art that can be used in the methods of the provided invention. For example, Actinomycin-D acts as a DNA-dependent DNA polymerase inhibitor by binding to DNA and preventing initiation of replication (Guy and Taylor (1978) PNAS 75:6088-92.) Other examples of DNA-dependent DNA polymerase inhibitors include, but are not limited to, actinomycin (dactinomycin), alpha-amanitin, aphidicolin (Cozad and Warner (1982) Gamete Research 6:155-60; Gonzcol and Plotkin (1985) Arch Virology 84:129-34; Haraguchi et al. (1983) Nucleic Acids Research 11:1197-1209), BP5, novobiocin (Schneck and Staudenbauer (1977) Nuc Acids Res 4:2057-64), rifampicin, rifamycin (Frolova et al. (1977) Nuc Acids Res 4:523-8), sulfoquinovosylmonoacylglycerol, sulfoquinovosyldiacylglycerol (Ohta et al. (2000) Mutat Res 467:139-52; Ohta et al. (1999) Biol Pharm Bull 22:111-16), ursane, oleanane triterpenoids, ursolic acid, oleanolic acid (Deng et al. (1999) J Nat Prod 62:1624-6), mikanolide, dihydromikanolide (U.S. Pat. No. 6,767,561), dehydroaltenusin (Mizushina et al. (2000) J Biol Chem 275:33957-61), catapol (Pungitore et al. (2004) J Nat Prod 67:357-61), taxinine, cephalomanninine (Oshige et al. (2004) Bioorganic and Medicinal Chem 12:2597-601), dipeptide alcohols (Kato et al. (2005) Int J Mol Med 16:653-9), corylifolin; bakuchiol; resveratrol; Neobavaisoflavone; daidzein; bakuchicin (Sun et al. (1998) J Nat Prod 61:362-6), levodopa, dopamine (Wick (1980) Cancer Research 40:1414-8), anacardic acid and oleic acid (Chen et al. (1998) Chem Comm 24:2769-70).

[00132] The DNA-dependent DNA polymerase inhibitors of the present invention can be added in an amount effective to inhibit DNA-dependent DNA polymerase activity. As a result, replication of DNA, when present in the sample in combination with RNA target or targets, is inhibited. The amount of a DNA-dependent DNA polymerase inhibitor that should be added to achieve the desired inhibition is well known and understood in the art. For example, the DNA-dependent DNA polymerase inhibitor Actinomycin D may be added in a concentration from about 0.1 μg/ml to about 100 μg/ml. In a preferred embodiment, 50% of DNA replication is inhibited by addition of a DNA-dependent DNA polymerase inhibitor. In another preferred embodiment, 60% of DNA replication is

IDT Exhibit 1020.5599

inhibited by addition of a DNA-dependent DNA polymerase inhibitor. In yet other preferred embodiments, 70%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% or 100% of DNA replication is inhibited by addition of a DNA-dependent DNA polymerase inhibitor.

[00133] Combinations of DNA-dependent DNA polymerase inhibitors may be used to achieve the desired inhibition. Use of combinations of DNA-dependent DNA polymerase inhibitors may be desirable to minimize unwanted effects on the reaction conditions by decreasing the concentration of each inhibitor in the reaction mixture. Alternatively, the combination of inhibitors may be able to achieve a greater amount of inhibition, such as by inhibiting DNA-dependent DNA polymerase activity at various stages.

[00134] A first strand cDNA synthesis reaction mixture can comprise a non-canonical nucleotide. For example, a first strand cDNA synthesis reaction mixture can further comprise dUTP and the four canonical dNTPs to generate primer extension products comprising dU. Other non canonical nucleotide triphosphates described herein can similarly be employed to generate tailed first strand synthesis comprising non canonical nucleotides.

[00135] Second strand synthesis can be carried out following cleavage of the RNA targets and a clean-up step which can remove the actinomycin D from the reaction products. RNA in the complex with first strand cDNA can be cleaved, for example, with an RNase which cleaves RNA in the RNA-DNA heteroduplex. RNA can also be cleaved by heating the reaction mixture, for example, heating the reaction mixture at 75ºC for 5 to 15 minutes.

[00136] Second strand synthesis can be carried out following hybridization of a second tailed primer to the first strand cDNA generated as described above. The second primer can comprise a 3'-sequence that can hybridize to the first cDNA comprising dU and a 5'-tail A. The second primer can further comprise a 5'-tail sequence B which is not complementary to the first strand cDNA. Primer extension along the first strand cDNA template can be carried out by DNA polymerase to produce a double stranded cDNA comprising a complex of the first and second primer extension products. The second strand cDNA synthesis can be carried out in the absence of the non canonical nucleotide tri phosphate (dUTP, for example) and can comprise a complement of the 5'-tail sequence of the first primer, sequence A', at the 3'-end.

[00137] Cleavage of the first strand cDNA in the complex can be achieved by combining the reaction mixture with an enzyme that cleaves the base portion of the non canonical nucleotide to generate an abasic site, and an amine, for example, DMED, which cleaves the backbone at the abasic site. For example, the combined action of the N-glycosylase (e.g., UNG or UDG) and DMED can result in the fragmentation of the first strand cDNA comprising the non canonical nucleotide (dU) to generate fragments comprising a blocked 3'-end, without affecting the second strand cDNA, which does not comprise non canonical nucleotides.

[00138] The single stranded second strand cDNA produced by the above reaction can comprise sequence B at the 5'-end, a sequence homologous to the RNA target, or a fraction thereof, and a sequence A' at the 3'-end. Sequences B and A' can be known and can be employed for further manipulation of the single stranded second strand cDNA. For example, the single stranded second strand cDNA can be rendered dsDNA by hybridization of an oligonucleotide comprising sequence A or part thereof, and extension of the hybridized oligonucleotide by DNA polymerase along the second strand cDNA template. Furthermore, the second strand cDNA can be amplified using primers that can hybridize to the end sequences, for example, by PCR. Solution phase PCR, emulsion PCR, surface bound PCR, such as Bridging-PCR and the like, are possible using the said single stranded DNA product or products. The primers can further comprise tail sequences which render the amplification product ends suitable for analysis on various analytical platforms, such as highly parallel sequencing (next generation sequencing, NGS). Any of the next generation sequencing techniques described herein can be used to sequence the second strand cDNA

IDT Exhibit 1020.5600

(e.g., Genome Analyzer by Illumina, the Genome Sequencer by 454/Roche Lifesciences, the SMRT platform by Pacific Biosciences, the HeliScope by Helicos Biosciences, the Polonator by Dover Systems, and semiconductor sequencing (Personal Genome Machine (PGM™) by Ion Torrent). The specific desired sequences can be appended at the end of the dsDNA product by ligation.

[00139] Direct application of the single stranded cDNA with desired 3'- and 5'-end sequences is also possible, when the end sequences are compatible with the chosen sequencing system (next generation sequencing (NGS) for example).

[00140] The end sequences flanking the sequence homologous to the RNA target can comprise molecular barcodes, when used for multiplex analysis, wherein specific barcodes can be designed for generation of the above described single stranded cDNA from individual samples followed by pooling of a number of products derived from individual samples.

[00141] The end sequence can comprise recognition sequences for restriction enzymes, thus enabling cleavage of the ends by the desired restriction enzyme, leading to the generation of ends suitable for further manipulations. For example, double stranded cDNA comprising desired overhangs at the two ends can be generated.

[00142] The defined known ends flanking the single stranded cDNA can further enable circularization by hybridizing an oligonucleotide complementary to the end sequences and ligation of the ends. The circular cDNA can be further amplified by rolling circle amplification (RCA).

[00143] The second strand cDNA with the defined 5'- and 3'- ends can be also amplified by single primer isothermal amplification (SPIA) (**Figure 8**) (see, e.g., U.S. Patent Application Publication No. 20090130721). For example, the single stranded second strand cDNA can be hybridized to a first composite primer, wherein the first composite primer comprises a 3' portion (e.g., DNA) which can anneal to the single stranded second strand cDNA and a 5' portion, which can be RNA, e.g., which does not hybridize to the single stranded second strand cDNA. The first composite primer can be extended with DNA polymerase. The second strand cDNA can be extended, using the RNA portion of the first composite primer as a template, with an RNA dependent DNA polymerase, to form a DNA/RNA heteroduplex. An agent that cleaves single-stranded RNA, e.g., RNase I, can be used to degrade the first composite primer that did not anneal to the second strand cDNA. This degradation can be used to prevent competition with other primers for subsequent annealing and to increase the efficiency of the subsequent single primer isothermal amplification reaction. The enzyme that is capable of cleaving single-stranded RNA may be, for example, RNase I, RNase T, RNase A, or combination thereof. Generally, an enzyme with little or no sequence specificity, and an enzyme that does not cleave the RNA in an RNA/DNA heteroduplex, can be used to cleave single-stranded RNA. In some embodiments, the enzyme capable of cleaving single-stranded RNA is RNase I. In one embodiment, the RNase I is inactivated by heat.

[00144] The RNA component of the RNA/DNA heteroduplex can be degraded by an enzyme that degrades RNA in an RNA/DNA heteroduplex, e.g., RNase H, to generate a 3' overhang. The enzymes that degrade RNA can be inactivated. Next, a composite amplification primer can hybridize to the 3' overhang of the second strand cDNA, wherein the composite amplification primer can comprises an RNA portion and a 3' DNA portion. The composite amplification primer hybridized to the 3' overhang of the second strand cDNA can be extended with at least one enzyme comprising DNA-dependent DNA polymerase activity. The first primer extension product can be displaced, RNA can be cleaved from the composite amplification primer and another composite amplification primer can hybridize such that primer extension and strand displacement are repeated, and whereby multiple copies of a polynucleotide sequence complementary to the RNA sequence of interest are generated.

IDT Exhibit 1020.5601

[00145] An input RNA target (template) can be total RNA from a biological sample. The RNA can be purified from a sample lysate or can be purified from cell free fluids, such as, e.g., plasma or serum. The sample can be cell lysate comprising RNA and DNA.

[00146] The RNA target can be fragmented RNA. RNA fragmentation can result from purification from a fixed sample, such as a Formalin-Fixed, Paraffin-Embedded (FFPE) sample, or can be carried out intentionally, for example, by heating in the presence of multivalent cation. The RNA can come from a fresh, frozen sample. An input template nucleic acid can be derived, e.g., from whole blood, frozen sample, sorted cells, FFPE sample, a tissue biopsy, serum, or skin.

[00147] The products of a linear amplification, or the primer extension to produce double stranded cDNA with desired sequence at the two ends, can be produced in reaction mixtures comprising non canonical nucleotide triphosphate, for example dUTP, thus generating a strand which is cleavable (for example, by combined reaction of UNG and DMED). Thus, following further manipulation of the products the synthesized strand can be removed by cleavage.

[00148] **Figure 1a** depicts methods and compositions in one embodiment of the invention for generating double stranded nucleic acid products from one or more template polynucleotides utilizing a plurality of oligonucleotides comprising random hybridizing portions and a polymerase comprising strand displacement activity. In this embodiment the products are generated from a template that contains one or more non-canonical nucleotides (dUTP). The figure further depicts methods and compositions for fragmenting the template polynucleotide and removing undesired single stranded nucleic acids.

[00149] **Figure 1b** depicts methods and compositions in another embodiment of the invention for generating double stranded nucleic acid products from one or more template polynucleotides utilizing a plurality of oligonucleotides comprising random hybridizing portions and a polymerase comprising strand displacement activity. In this embodiment the template nucleotide does not contain any non-canonical nucleotides. Methods and compositions for fragmenting the template polynucleotide and removing undesired single stranded nucleic acids can be used.

[00150] **Figure 2** depicts methods and compositions for preparing double stranded nucleic acid for adaptor ligation.

[00151] **Figure 3** depicts methods and compositions for generating labeled double stranded nucleic acid products from one or more template polynucleotides utilizing a plurality of oligonucleotides comprising one or more labels and random hybridizing portions and a polymerase comprising strand displacement activity. The figure further predicts methods and compositions for fragmenting the template polynucleotide. Undesired single stranded nucleic acids can be removed.

[00152] **Figure 4** depicts BioAnalyzer profiles of input amplified cDNA, random primed products mixture and random primed products following cleavage of the input amplified cDNA. Whole transcriptome amplified cDNA from HeLa total RNA. The x-axis represents product length in bp. The concentration of the starting material is 172ng/μL. The concentration of random primed amplified DNA is 200ng/μL. The concentration of the random primed and fragmented input nucleic acid is 100ng/μL.

[00153] **Figure 5** depicts BioAnalyzer profiles of input amplified cDNA, random primed products mixture and random primed products following cleavage of the input amplified cDNA. Whole transcriptome amplified cDNA from total RNA from a biological sample and brain is used. The x-axis represents product length in bp. The profile of the random priming products following removal of the input amplified cDNA is very reproducible for the two samples which represent different sample source.

IDT Exhibit 1020.5602

III. Compositions of the Invention

[00154] Appropriate reaction media and conditions for carrying out the methods of the invention are those that permit nucleic acid preparation according to the methods of the invention. Such media and conditions are known to persons of skill in the art, and are described in various publications, such as U.S. Pat. Nos. 5,554,516; 5,716,785; 5,130,238; 5,194,370; 6,090,591; 5,409,818; 5,554,517; 5,169,766; 5,480,784; 5,399,491; 5,679,512; and PCT Pub. No. WO 99/42618. For example, a buffer may be Tris buffer, although other buffers can also be used as long as the buffer components are non-inhibitory to enzyme components of the methods of the invention. The pH is preferably from about 5 to about 11, more preferably from about 6 to about 10, even more preferably from about 7 to about 9, and most preferably from about 7.5 to about 8.5. The reaction medium can also include bivalent metal ions such as Mg2+ or Mn2+, at a final concentration of free ions that is within the range of from about 0.01 to about 15 mM, and most preferably from about 1 to 10 mM. The reaction medium can also include other salts, such as KCl or NaCl, that contribute to the total ionic strength of the medium. For example, the range of a salt such as KCl is preferably from about 0 to about 125 mM, more preferably from about 0 to about 100 mM, and most preferably from about 0 to about 75 mM. The reaction medium can further include additives that could affect performance of the amplification or primer extension reactions, but that are not integral to the activity of the enzyme components of the methods. Such additives include proteins such as BSA, single strand binding proteins (e.g., T4 gene 32 protein), and non-ionic detergents such as NP40 or Triton. Reagents, such as DTT, that are capable of maintaining enzyme activities can also be included. Such reagents are known in the art. Where appropriate, an RNase inhibitor (such as RNasin) that does not inhibit the activity of an RNase employed in the method can also be included. Any aspect of the methods of the invention can occur at the same or varying temperatures. Amplification or primer extension reactions may be performed isothermally, which avoids the cumbersome thermocycling process, or may be performed by standard thermocycling techniques such as PCR. The amplification or primer extension reaction may be carried out at a temperature that permits hybridization of the oligonucleotides primers of the invention to the template polynucleotide and primer extension products, and that does not substantially inhibit the activity of the enzymes employed. The temperature can be in the range of preferably about 25ºC to about 85ºC, more preferably about 30ºC to about 80ºC, and most preferably about 37ºC to about 75ºC. In some embodiments that include RNA transcription, the temperature for the transcription steps is lower than the temperature(s) for the preceding steps. In these embodiments, the temperature of the transcription steps can be in the range of preferably about 25ºC to about 85ºC, more preferably about 30ºC to about 75ºC, and most preferably about 37ºC to about 70ºC.

[00155] Nucleotide and/or nucleotide analogs, such as deoxyribonucleoside triphosphates, that can be employed for synthesis of the primer extension products or for amplification in the methods of the invention are provided in the amount of from preferably about 50 to about 2500 μM, more preferably about 100 to about 2000 μM, even more preferably about 200 to about 1700 μM, and most preferably about 250 to about 1500 μ M. In some embodiments, a nucleotide or nucleotide analog whose presence in the primer extension strand enhances displacement of the strand (for example, by causing base pairing that is weaker than conventional AT, CG base pairing) is included. Such nucleotide or nucleotide analogs include deoxyinosine and other modified bases, all of which are known in the art. Nucleotides and/or analogs, such as ribonucleoside triphosphates, that can be employed for synthesis of the RNA transcripts in the methods of the invention are provided in the amount of from preferably about 0.25 to about 6 mM, more preferably about 0.5 to about 5 mM, even more preferably about 0.75 to about 4 mM, and most preferably about 1 to about 3 mM. Nucleotide analogues may further include non-canonical nucleotides such as for example methylated purine nucleotides, or dUTP. Additionally, nucleotide analogues may include labeled nucleotides such

IDT Exhibit 1020.5603

as amino-allyl labled nucleotides, digoxin labeled nucleotides, fluorophore (e.g. Cy3, Cy5, fluoroscein, rhodamine) labeled nucleotides, biotin labeled nucleotides, or protein labeled nucleotides (e.g. avidin, streptavidin, neutravidin, horseradish peroxidase, alkaline phosphatase).

[00156] Modified nucleotides can be canonical nucleotide or non-canonical (cleavable) nucleotides. It is understood, however, that modified nucleotides that are not non-canonical (cleavable) nucleotide under the reaction conditions used in the methods of the invention, if present, generally should not affect the ability of the polynucleotide to undergo cleavage of a base portion of non-canonical nucleotide, such that an abasic site is generated, and/or cleavage of a phosphodiester backbone at an abasic site, such that fragments are generated, as described herein. If present, modification to the nucleotide structure, such as methylated nucleotides 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. Other types of modifications include, for example, "caps", substitution of one or more of the naturally occurring nucleotides with an analog, internucleotide modifications such as, for example, those with uncharged linkages (e.g., methyl phosphonates, phosphotriesters, phosphoamidates, carbamates, etc.) and with charged linkages (e.g., phosphorothioates, phosphorodithioates, etc.), those containing pendant moieties, such as, for example, proteins (e.g., nucleases, toxins, antibodies, signal peptides, ply-L-lysine, etc.), those with intercalators (e.g., acridine, psoralen, etc.), those containing chelators (e.g., metals, radioactive metals, boron, oxidative metals, etc.), those containing alkylators, those with modified linkages (e.g., alpha anomeric nucleic acids, etc.), as well as unmodified forms of the polynucleotide(s). It is understood that internucleotide modifications may, e.g., alter the efficiency and/or kinetics of cleavage of the phosphodiester backbone (as when, for example a phosphodiester backbone is cleaved at an abasic site, as described herein). Further, any of the hydroxyl groups ordinarily present in the sugars may be replaced, for example, by phosphonate groups, phosphate groups, protected by standard protecting groups, or activated to prepare additional linkages to additional nucleotides. The 5′ and 3′ terminal OH can be phosphorylated or substituted with amines or organic capping groups moieties of from 1 to 20 carbon atoms. Other hydroxyls may also be derivatized to standard protecting groups. Polynucleotides can also contain analogous forms of ribose or deoxyribose sugars that are generally known in the art, including, for example, 2′—O-methyl-, 2′-O-allyl, 2′-fluoro- or 2′-azido-ribose, carbocyclic sugar analogs, α-anomeric sugars, epimeric sugars such as arabinose, xyloses or lyxoses, pyranose sugars, furanose sugars, sedoheptuloses, acyclic analogs and abasic nucleoside analogs. One or more phosphodiester linkages may be replaced by alternative linking groups. These alternative linking groups include, but are not limited to, embodiments wherein phosphate is replaced by P(O)S("thioate"), P(S)S ("dithioate"), "(O)NR_2("amidate"), P(O)R, P(O)OR′, CO or CH_2("formacetal"), in which each R or R′ is independently H or substituted or unsubstituted alkyl (1-20 C) optionally containing an ether (—O—) linkage, aryl, alkenyl, cycloalkyl, cycloalkenyl or araldyl. Not all linkages in a polynucleotide need be identical. The preceding description applies to all polynucleotides referred to herein, including DNA. It is understood, however, that modified nucleotides and/or internucleotide linkages and/or, if present, generally should not affect the ability of the polynucleotide to undergo cleavage of a base portion of a non-canonical nucleotide, such that an abasic site is generated, and/or the ability of a polynucleotide to undergo cleavage of a phosphodiester backbone at an abasic site, such that fragments are generated, as described herein.

[00157] In order to prepare primer extension products including single and double stranded DNA products, oligonucleotide primers may be used that bind to the input nucleic acid template and are subsequently extended along the template by one or more enzymes such as for example ligases or polymerases including but not limited to

IDT Exhibit 1020.5604

polymerases comprising strand displacement activity. The oligonucleotide components of the amplification reactions of the invention are generally in excess of the number of target nucleic acid sequence to be amplified. They can be provided at about or at least about any of the following: 1, 5, 10, 10^2, 10^4, 10^6, 10^8, 10^{10}, 10^{12} times the amount of target nucleic acid. Oligonucleotide primers, composite primers and pro-promoter template oligonucleotides (PTO) can each be provided at about or at least about any of the following concentrations: 50 nM, 100 nM, 500 nM, 1000 nM, 2500 nM, 5000 nM. The primer may be complementary to at least a portion of the input nucleic acid template and may comprise DNA, RNA, labeled RNA or DNA, or mixtures thereof. The primers may comprise specific sequences, multiple sequences, or random sequences. Various embodiments of the primers used in the methods of the invention are described herein. For the methods described herein, one or more primers can be used.

[00158] In one embodiment, the oligonucleotide primers are oligonucleotides that are supplied externally. The primer in the methods of the invention comprises a sequence (which may or may not be the whole of the primer) that is hybridizable (under a given set of conditions) to the template nucleic acid such that any extension products would be homologous to the template nucleic acid. In another embodiment, the primer is a primer comprising DNA (in some embodiments, consisting of DNA).

[00159] To achieve hybridization to a template nucleic acid (which, as is well known and understood in the art, depends on other factors such as, for example, ionic strength and temperature), the sequence of the primer that is hybridizable to the template nucleic acid is preferably of at least about 60%, more preferably at least about 75%, even more preferably at least about 90%, and most preferably at least about 95% or 100% complementarity to the template nucleic acid.

[00160] Suitable primers in the methods of the invention are long enough such that they do not dissociate or do not dissociate substantially from the template nucleic acid in an uninduced manner (e.g. they do not dissociate in the absence of heat or chemical denaturation or strand displacement activity). Suitable primers are preferably from about 3 to about 200, more preferably from about 5 or 6 to about 150, even more preferably from about 10 to about 100, and most preferably from about 12 to about 50, nucleotides in length, or about 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87 , 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100 nucleotides in length or longer.

[00161] In some embodiments, the one or more primers comprise random sequences. In other embodiments, the primer comprises a portion (for example, a 3′ portion) that comprises a random sequence (i.e., a sequence designed to be hybridizable (under a given set of conditions) to one or more sequences in the sample). In some cases the one or more primers comprise a polyT or a polyA sequence. In some cases, the primers comprise a mixture of random sequences and specific sequences such as polyT sequences or polyA sequences. In some cases, the primers are generated as a pool of random sequences and are further selected by hybridization or subtractive hybridization to remove sequences which hybridize to undesired sequences. In some embodiments, the primers comprise a pool of primers synthesized to comprise all desired sequences (e.g. designed to hybridize to all possible or substantially all possible desired sequences and not to hybridize to undesired sequences in the samples such as for example rRNA sequences.) This pool of sequences comprises a pool of semi-random or not so random sequences.

[00162] In some cases, non-standard bases which exhibit base pairing interactions with more than one partner such as inosine which can base pair with uracil, adenine, and cytosine may be used. The non-standard bases with greater

IDT Exhibit 1020.5605

propensity for multiple base pairing partners are referred to as wobble bases. In other cases, specific annealing sequences may be used in combination with such wobble bases. In some cases, primers comprise specific annealing sequences in which one or more positions within the annealing sequence are randomized. Such a combination of specific sequences with randomized positions may be used for example to amplify a set of related or homologous nucleic acid sequences. In still other cases, primers comprise specific annealing sequences, or a set of specific sequences.

[00163] In some embodiments, the primer comprises a 5′ portion that is non-hybridizable to a template nucleic acid (tailed primer). In some cases, the 5'-region of the primer further comprises defined or universal sequences. In some cases, a primer of the present invention may comprise a linker region that is not a hybridizable portion and is not a tail. Primers may also comprise a ligand. The ligand may or may not be comprised of nucleic acid or a peptide. In some cases, the ligand is a small organic molecule such as for example biotin or a fluorophore. The primers may be synthesized by a number of common methods known in the art. In some cases, the methods of synthesizing primers of the present invention include solid phase synthesis methods such as provided in US Patent 5,623,068.

[00164] In some aspects of the present invention, composite primers are utilized, such as for example during generation of a template nucleic acid comprising one or more non-canonical nucleotide or for amplification or labeling of the products of the primer extension step of the present invention. Composite primers of the present invention include primers that comprise an RNA segment and a DNA segment. The RNA segment may be generally located at the 5’ end of the composite primer and the DNA segment may generally be located at the 3’ end of the composite c primer. In some cases the RNA and DNA segments are adjacent. In other cases, there is an intervening element between the RNA and the DNA segment. The intervening element may comprise additional nucleic acid sequence. In one embodiment, a portion of the DNA segment comprises a template annealing sequence. The template annealing sequence as described primers provided herein may be target specific, a set of target specific sequences, non-random, not-so random, or a random sequence.

[00165] In some embodiments of the present invention, the DNA-RNA composite primer annealing sequences may be between about 2 nucleotides in length to about 100 nucleotides in length. In some cases, annealing sequences may be between about 3 nucleotides in length to about 50 nucleotides in length. In still other cases, annealing sequences may be about 4 to about 30 nucleotides in length. For example, but without limitation, DNA-RNA composite primer annealing sequences may be about 5, 6, 7, 8, 9, 10, 12, 14, 16, 18, 20, or 25 nucleotides in length. In some cases, random annealing sequences comprise random pentamers, hexamers, heptamers, octomers, decamers, undecamers, dodecamers, tirdecamers, or larger.

[00166] In some embodiments of the present invention, a portion of the RNA segment of the composite primer comprises a sequence tail. The sequence tail may in some cases comprise the same sequence as the sequence tail used for a previous primer extension reaction or a previous amplification reaction. In other cases, the sequence tail may comprise a different sequence as compared to the sequence tail used for a previous primer extension reaction or amplification reaction. In still other cases, the sequence tail may be complementary to a sequence tail used in a previous primer extension reaction or a previous amplification reaction..

[00167] In some embodiments of the present invention, a composite primer comprising a 5’-RNA segment comprising a sequence tail and a 3’ DNA segment comprising an annealing sequence (e.g. a random annealing sequence) is annealed to a primer extension product (e.g. a single stranded primer extension product, or a partial double stranded product comprising a portion that is double stranded and a portion that is single stranded), a polymerase is added to the annealing reaction, and a double stranded product is created comprising the primer

IDT Exhibit 1020.5606

extension product and a composite primer extension product. In some cases, the polymerase may comprise stand displacement activity. The double stranded product may be denatured and another primer may be annealed to the composite primer extension product and extended with a polymerase to create a double stranded product comprising an RNA/DNA heteroduplex. In some embodiments of the present invention, the double stranded product comprising an RNA /DNA heteroduplex is used as a substrate for amplification via SPIA using a chimeric amplification primer as provided herein. Alternatively, the double stranded product may be amplified using any other amplification method provided herein.

[00168] In some cases, the RNA portion of the RNA/DNA heteroduplex may be cleaved by an agent that cleaves the RNA portion of an RNA/DNA heteroduplex. The agent may be for example an RNase H or a polymerase or other enzyme that comprises RNase H-like cleavage activity. In some cases an amplification primer such as a chimeric amplification primer (e.g. an RNA DNA composite primer suitable for use as an amplification primer) may be annealed to the 3' ssDNA portion provided by the step of cleaving the RNA portion of the RNA/DNA heteroduplex. In some cases, a DNA polymerase, such as a DNA polymerase with strand displacement activity may be used to extend the hybridized chimeric amplification primer in a reaction mixture to generate a double stranded product comprising an RNA/DNA heteroduplex and a single stranded product of SPIA amplification. In some cases, the reaction mixture may comprise a chimeric amplification primer, an agent that cleaves the RNA portion of an RNA/DNA heteroduplex and a DNA polymerase comprising strand displacement activity and various buffers, salts, and reagents (e.g. dNTPs) for amplification of the double stranded SPIA substrate comprising an RNA/DNA heteroduplex as provided herein. In some cases, the amplification of the SPIA substrate may be performed isothermally.

[00169] Amplification primers of the present invention include chimeric amplification primers. Chimeric amplification primers are RNA/DNA composite primers that can be used to create multiple copies of (amplify) a polynucleotide sequence isothermally using RNA cleavage, and DNA polymerase activity with strand displacement. Amplification with such primers is described, for example in U.S. Patent Nos. 6,251,639, 6,692,918, and 6,946,251. The composite amplification primer comprises sequences capable of hybridizing to a portion of a DNA template, and most often comprises sequences hybridizable to a defined 3'-portion of the DNA. Amplification primers of the present invention may also comprise all DNA canonical or non canonical nucleotides.

[00170] In some embodiments, a chimeric amplification primer comprises at least one RNA portion that is capable of binding (hybridizing) to a sequence on a DNA template and at least one DNA portion that is capable of binding (hybridizing) to a sequence on a DNA temlate, wherein the RNA portion is susceptible to being cleaved with an RNase H when hybridized to the DNA template. The chimeric amplification primers bind to the DNA template to form a partial heteroduplex in which only the RNA portion of the primer is cleaved upon contact with a ribonuclease such as RNase H, while the DNA template remains intact, thus enabling annealing of another chimeric primer. In some aspects of the invention, the 5' RNA portion of the chimeric amplification primer may be about 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, or about 40 nucleotides in length.

[00171] The chimeric amplification primers may also comprise a 3'-DNA portion that is capable of hybridization to a sequence on the DNA template such that its hybridization to the DNA is favored over that of the nucleic acid strand that is displaced from the DNA template by the DNA polymerase. Such primers can be rationally designed based on well known factors that influence nucleic acid binding affinity, such as sequence length and/or identity, as well as hybridization conditions. In some aspects of the present invention, the 3' DNA portion of a composite amplification primer may be about 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 nucleotides in length.

IDT Exhibit 1020.5607

[00172] In some embodiments of the present invention chimeric oligonucleotides are utilized. In some cases, the chimeric oligonucleotides may be used to provide a defined end sequence for the SPIA amplified DNA products. The chimeric oligonucleotides may comprise a 5' segment comprising RNA and a 3' segment comprising DNA. In some embodiments, a portion of the 3' segment of the chimeric oligonucleotide is substantially the same as the RNA sequence removed from the DNA-RNA heteroduplex of a SPIA substrate by RNase H.

[00173] Primers of the present invention may be useful for the primer extension step of the present invention for producing primer extension products that do not comprise non-canonical nucleotides (e.g. single or double stranded nucleic acids or a combination thereof) or that do not comprise the same non-canonical nucleotides as a template nucleic acid, and that may be differentially separated from template nucleic acid comprising one or more non-canonical nucleotides via a step of cleaving the template nucleic acid. Primers of the present invention may also be useful for the generation of template nucleic acid comprising one or more non-canonical nucleotides, such as for example by an amplification or polymerization step including any of the amplification or polymerization methods provided herein including PCR, SPIA, IVT, reverse transcription or a combination thereof, or by ligation of a plurality of adjacent primers, or a combination thereof. In some cases, primers are useful for the generation of template nucleic acid comprising non-canonical nucleotides via a limited amplification method using a single round of amplification or a small number of rounds (e.g. 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, etc.) such as in the generation of cDNA. Primers of the present invention may also be useful for amplifying or labeling the products of the primer extension reaction. For example, primers of the present invention may be used for further amplification of the products of the primer extension reaction after the cleavage of the template nucleic acid, purification or a combination thereof, such as by PCR, SPIA, IVT, any amplification method provided herein, or a combination thereof.

[00174] Cleavage agents of the present invention include enzymes such as for example glycosylases (e.g. UDG or MPG) and endonucleases (e.g. AP endonucleases), and chemical agents such as for example bases, amines, primary amines, secondary amines, tertiary amines, polyamines (e.g. spermine, spermidine, 1,4-diaminobutane, lysine, the tripeptide K-W-K, N, N-dimethylethylenediamine (DMED), piperazine, or 1,2-ethylenediamine) and nucleophiles or any combination thereof. Cleavage agents comprising enzymes may be used in a reaction mixture at various concentrations suitable for the methods of the present invention including between about 1nM and about 1mM, or between about 5nM and about 0.5mM, about 10nM and about 0.25mM, about 25nM and about 0.1mM, about 50nM and about 0.05mM, about 100nM and about 0.025mM, about 150nM and about 0.0125mM, about 200nM and about 0.01mM, about 300nM and about 5μM, about 500nM and about 2 μM, or between about about 0.75 μM and about 1 μM. In some cases, cleavage agents comprising enzymes may be used in a reaction mixture at about 1nM, 2nM, 3nM, 4nM, 5nM, 7.5nM, 10nM, 15nM, 20nM, 25nM, 30nM, 35nM, 50nM, 75nM, 100nM, 150nM, 200nM, 250nM, 500nM, 750nM, 1 μM, 2 μM, 3 μM, 5 μM, 7.5 μM, 10 μM, 25 μM, 50 μM, 100 μM or more. Cleavage agents comprising chemical agents may be used in a reaction mixture at various suitable for the methods of the present invention including between about 1nM and about 1mM, or between about 5nM and about 0.5mM, about 10nM and about 0.25mM, about 25nM and about 0.1mM, about 50nM and about 0.05mM, about 100nM and about 0.025mM, about 150nM and about 0.0125mM, about 200nM and about 0.01mM, about 300nM and about 5μM, about 500nM and about 2 μM, or between about about 0.75 μM and about 1 μM. In some cases, cleavage agents comprising chemical agents may be used in a reaction mixture at about 1nM, 2nM, 3nM, 4nM, 5nM, 7.5nM, 10nM, 15nM, 20nM, 25nM, 30nM, 35nM, 50nM, 75nM, 100nM, 150nM, 200nM, 250nM, 500nM, 750nM, 1 μM, 2 μM, 3 μM, 5 μM, 7.5 μM, 10 μM, 25 μM, 50 μM, 100 μM or more.

IDT Exhibit 1020.5608

[00175] The initial step of the methods of the present invention may be providing the template nucleic acid, which template comprises one or more non-canonical nucleotides. The non-canonical nucleotides may be incorporated by synthetic (e.g. chemical synthesis) or enzymatic means (e.g. incorporation by ligase or polymerase), or a combination thereof.

[00176] The input nucleic acid template is generally a polymeric nucleotide, which in the intact natural state can have about 30 to 5,000,000 to about $3x10^9$ or more nucleotides and in an isolated state can have about 20 to 100,000 or more nucleotides, usually about 100 to 50,000 nucleotides, more frequently 500 to 20,000 nucleotides, or about 500 to 10,000 nucleotides. The template includes RNAs or DNAs from any source and or/species, including human, animals, plants, and microorganisms such as bacteria, yeasts, viruses, viroids, molds, fungi, plants, and fragments thereof. Exemplary templates can be obtained and purified using standard techniques in the art and includes RNAs and DNAs in purified or unpurified form, which include, but are not limited to, mRNAs, tRNAs, snRNAs, rRNAs, retroviruses, small non-coding RNAs, microRNAs, polysomal RNAs, pre-mRNAs, intronic RNA, viral RNA, DNA-RNA heteroduplexes, DNA, genomic DNA, and fragments thereof. Use of a DNA template in the methods of the present invention (including genomic DNA target) may require initial reverse transcription of an RNA target into DNA form, which can be achieved using methods known in the art. Preparation of a RNA/DNA heteroduplex, double stranded RNA, sequences prone to stem loop formation, or double stranded DNA may require denaturation of the duplex to obtain a single stranded nucleic acid, or denaturation followed by reverse transcription of the RNA strand to obtain an DNA. The template may be only a minor fraction of a complex mixture such as a biological sample and may be obtained from various biological materials by procedures known in the art.

[00177] The template nucleic acid can be known or unknown and may contain more than one desired or suspected specific nucleic acid sequence of interest, each of which may be the same or different from each other. Therefore, the methods provided herein are useful not only for preparing double stranded DNA fragments from one specific nucleic acid sequence, but also for preparing double stranded DNA fragments simultaneously more than one different specific nucleic acid sequence located on the same or different nucleic acid molecules.

[00178] Enzymes of the present invention include but are not limited to one or more polymerases, one or more endonucleases, one or more exonucleases, one or more ligases, one or more nucleotide kinases, one or more glycosidases, one or more RNases, including but not limited to RNase1 or RNase H, or any combination thereof. Polymerases of the present invention include RNA dependent and DNA dependent polymerases. Polymerases further include RNA and DNA dependent DNA polymerases as well as RNA and DNA dependent RNA polymerases. Accordingly, polymerases include reverse transcriptases, and transcriptases. In some embodiments, polymerases of the present invention comprise both RNA dependent and DNA dependent DNA polymerase activity. In some cases, polymerases of the present invention include proof reading activity. Some of the polymerases of the present invention may be useful for degradation of 3' or 5' ends or both. Some of the polymerases of the present invention may be useful for filling in of single stranded overhangs of double stranded nucleic acids. Some of the polymerases of the present invention may be useful for forming blunt ends from double stranded nucleic acids comprising single stranded overhangs from one or more of their 3' and 5' ends either by trimming back the overhangs (degrading the single stranded portion), or by filling in the overhang (polymerizing a sequence complementary and hybridized to the single stranded overhang).

[00179] Polymerases may be used in a reaction mixture at various concentrations suitable for the methods of the present invention including between about 1nM and about 1mM, or between about 5nM and about 0.5mM, about 10nM and about 0.25mM, about 25nM and about 0.1mM, about 50nM and about 0.05mM, about 100nM and about

IDT Exhibit 1020.5609

0.025mM, about 150nM and about 0.0125mM, about 200nM and about 0.01mM, about 300nM and about 5μM, about 500nM and about 2 μM, or between about about 0.75 μM and about 1 μM. In some cases, polymerases may be used in a reaction mixture at about 1nM or more, 2nM or more, 3nM or more, 4nM or more, 5nM or more, 7.5nM or more, 10nM or more, 15nM or more, 20nM or more, 25nM or more, 30nM or more, 35nM or more, 50nM or more, 75nM or more, 100nM or more, 150nM or more, 200nM or more, 250nM or more, 500nM or more, 750nM or more, 1 μM or more, 2 μM or more, 3 μM or more, 5 μM or more, 7.5 μM or more, 10 μM or more, 25 μM or more, 50 μM or more, 100 μM or more.

[00180] Exemplary polymerase include mesophilic and thermophilic polymerases including but not limited to Klenow polymerase, exo- Klenow polymerase, 3' 5' exo- Klenow polymerase, 3' exo- Klenow polymerase, 5' exo- Klenow polymerase, DNA polymerase 1, Bst polymerase, Bst large fragment polymerase, Bca Polymerase and derivatives thereof, Vent polymerase, Vent polymerase, Deep Vent (exo-) polymerase, 9°Nm polymerase, Therminator polymerase, Therminator II polymerase, MMulV Reverse Transcriptase, phi29 polymerase, T4 DNA polymerase, T4 RNA polymerase, or DyNAzyme EXT polymerase, Phusion Polymerase, or any other polymerases known in the art or any combination thereof. An "RNA-dependent DNA polymerase" or "reverse transcriptase" ("RT") is an enzyme that synthesizes a complementary DNA copy from an RNA template. All known reverse transcriptases also have the ability to make a complementary DNA copy from a DNA template; thus, they are both RNA- and DNA-dependent DNA polymerases. Reverse transcriptases may also have an RNase H activity. Some examples of reverse transcriptases are reverse transcriptase derived from Maloney murine leukemia virus (MMLV-RT), avian myeloblastosis virus, retroviral reverse transcriptase, retrotransposon reverse transcriptase, hepatitis B reverse transcriptase, cauliflower mosaic virus reverse transcriptase, bacterial reverse transcriptase, E. coli DNA polymerase and klenow fragment, and Tth DNA polymerase. A primer can be used to initiate synthesis with both RNA and DNA templates. In other examples a DNA dependent DNA polymerase may also comprise an RNA-dependent DNA polymerase such as Klenow polymerase, Bst DNA polymerase and the like.

[00181] Ligases of the present invention include thermophilic and mesophilic ligases including but not limited to T4 DNA ligase, T4 RNA ligase, Taq ligase, *E. coli* ligase, Tfi ligase, and Ampligase®, or any other ligases known in the art or combinations therof. Ligases may be used in a reaction mixture at various concentrations suitable for the methods of the present invention including between about 1nM and about 1mM, or between about 5nM and about 0.5mM, about 10nM and about 0.25mM, about 25nM and about 0.1mM, about 50nM and about 0.05mM, about 100nM and about 0.025mM, about 150nM and about 0.0125mM, about 200nM and about 0.01mM, about 300nM and about 5μM, about 500nM and about 2 μM, or between about 0.75 μM and about 1 μM. In some cases, ligases may be used in a reaction mixture at about 1nM or more, 2nM or more, 3nM or more, 4nM or more, 5nM or more, 7.5nM or more, 10nM or more, 15nM or more, 20nM or more, 25nM or more, 30nM or more, 35nM or more, 50nM or more, 75nM or more, 100nM or more, 150nM or more, 200nM or more, 250nM or more, 500nM or more, 750nM or more, 1 μM or more, 2 μM or more, 3 μM or more, 5 μM or more, 7.5 μM or more, 10 μM or more, 25 μM or more, 50 μM or more, 100 μM or more.

[00182] Kinases of the present invention include thermophilic and mesophilic nucleotide kinases including but not limited to T4 polynucleotide kinase or any other nucleotide kinases known in the art or any combination thereof. Nucleotide kinases may be used in a reaction mixture at various concentrations suitable for the methods of the present invention including between about 1nM and about 1mM, or between about 5nM and about 0.5mM, about

10nM and about 0.25mM, about 25nM and about 0.1mM, about 50nM and about 0.05mM, about 100nM and about 0.025mM, about 150nM and about 0.0125mM, about 200nM and about 0.01mM, about 300nM and about 5μM, about 500nM and about 2 μM, or between about about 0.75 μM and about 1 μM. In some cases, nucleotide kinases may be used in a reaction mixture at about 1nM or more, 2nM or more, 3nM or more, 4nM or more, 5nM or more, 7.5nM or more, 10nM or more, 15nM or more, 20nM or more, 25nM or more, 30nM or more, 35nM or more, 50nM or more, 75nM or more, 100nM or more, 150nM or more, 200nM or more, 250nM or more, 500nM or more, 750nM or more, 1 μM or more, 2 μM or more, 3 μM or more, 5 μM or more, 7.5 μM or more, 10 μM or more, 25 μM or more, 50 μM or more, 100 μM or more.

[00183] Exonucleases of the present invention include thermophilic and mesophilic exonucleases including but not limited to single strand specific exonucleases, single stranded DNA specific exonucleases, RNA exonucleases, exonuclease 1, exonuclease 7, any other exonucleases known in the art, or any combination thereof. Exonucleases may be used in a reaction mixture at various concentrations suitable for the methods of the present invention including between about 1nM and about 1mM, or between about 5nM and about 0.5mM, about 10nM and about 0.25mM, about 25nM and about 0.1mM, about 50nM and about 0.05mM, about 100nM and about 0.025mM, about 150nM and about 0.0125mM, about 200nM and about 0.01mM, about 300nM and about 5μM, about 500nM and about 2 μM, or between about 0.75 μM and about 1 μM. In some cases, exonucleases may be used in a reaction mixture at about 1nM or more, 2nM or more, 3nM or more, 4nM or more, 5nM or more, 7.5nM or more, 10nM or more, 15nM or more, 20nM or more, 25nM or more, 30nM or more, 35nM or more, 50nM or more, 75nM or more, 100nM or more, 150nM or more, 200nM or more, 250nM or more, 500nM or more, 750nM or more, 1 μM or more, 2 μM or more, 3 μM or more, 5 μM or more, 7.5 μM or more, 10 μM or more, 25 μM or more, 50 μM or more, 100 μM or more.

[00184] Endonucleases of the present invention include thermophilic and mesophilic endonucleases including but not limited to single strand specific endonucleases, single stranded DNA specific endonucleases, RNA endonucleases, S1 endonuclease, mung bean endonucleases, and any other endonucleases known in the art, or any combination thereof. Endonucleases may be used in a reaction mixture at various concentrations suitable for the methods of the present invention including between about 1nM and about 1mM, or between about 5nM and about 0.5mM, about 10nM and about 0.25mM, about 25nM and about 0.1mM, about 50nM and about 0.05mM, about 100nM and about 0.025mM, about 150nM and about 0.0125mM, about 200nM and about 0.01mM, about 300nM and about 5μM, about 500nM and about 2 μM, or between about 0.75 μM and about 1 μM. In some cases, exonucleases may be used in a reaction mixture at about 1nM or more, 2nM or more, 3nM or more, 4nM or more, 5nM or more, 7.5nM or more, 10nM or more, 15nM or more, 20nM or more, 25nM or more, 30nM or more, 35nM or more, 50nM or more, 75nM or more, 100nM or more, 150nM or more, 200nM or more, 250nM or more, 500nM or more, 750nM or more, 1 μM or more, 2 μM or more, 3 μM or more, 5 μM or more, 7.5 μM or more, 10 μM or more, 25 μM or more, 50 μM or more, 100 μM or more.

[00185] Glycosylases of the present invention include but are not limited to glycosylases capable of cleaving a nucleotide to generate an abasic site in a polynucleotide substrate. In some embodiments, glycosylases of the present invention provide for specific cleavage of incorporated non-canonical nucleotides. Exemplary glycosylases include but are not limited to UNG/UDG, MPG, any other glycosylases such as those described in international application publication number WO /2008/005459, or US application numbers 10/441,663, and 11/026,280, or any combination thereof. Glycosidases may be used in a reaction mixture at various concentrations suitable for the methods of the present invention including between about 1nM and about 1mM, or between about 5nM and about

IDT Exhibit 1020.5611

0.5mM, about 10nM and about 0.25mM, about 25nM and about 0.1mM, about 50nM and about 0.05mM, about 100nM and about 0.025mM, about 150nM and about 0.0125mM, about 200nM and about 0.01mM, about 300nM and about 5μM, about 500nM and about 2 μM, or between about about 0.75 μM and about 1 μM. In some cases, glycosidases may be used in a reaction mixture at about 1nM or more, 2nM or more, 3nM or more, 4nM or more, 5nM or more, 7.5nM or more, 10nM or more, 15nM or more, 20nM or more, 25nM or more, 30nM or more, 35nM or more, 50nM or more, 75nM or more, 100nM or more, 150nM or more, 200nM or more, 250nM or more, 500nM or more, 750nM or more, 1 μM or more, 2 μM or more, 3 μM or more, 5 μM or more, 7.5 μM or more, 10 μM or more, 25 μM or more, 50 μM or more, 100 μM or more. In some cases, one glycosidase such as UDG or any glycosidase provided herein may be used to cleave the base portion of one or more non-canonical nucleotides in the template nucleic acid. In other cases more than one glycosidase such as any combination of the glycosidases provided herein may be used to cleave the base portion of one or more non-canonical nucleotides in the template nucleic acid. In some cases, the glycosidase used to cleave the base portion of one or more non-canonical nucleotides in the template nucleic acid to generate one or more abasic sites may also comprise a lyase activity for cleaving the phosphodiester backbone at the one or more abasic sites or a portion thereof.

IV. Kits of the Invention

[00186] The present invention provides kits containing one or more compositions of the present invention and other suitable reagents suitable for carrying out the methods of the present invention. The invention provides, e.g., diagnostic kits for clinical or criminal laboratories, or nucleic acid amplification or analysis kits for general laboratory use. The present invention thus includes kits which include some or all of the reagents necessary to carry out the methods of the present invention, e.g., sample preparation reagents, oligonucleotides, binding molecules, stock solutions, nucleotides, polymerases, inhibitors, enzymes, positive and negative control oligonucleotides and target sequences, test tubes or plates, labeling reagents, fragmentation reagents, detection reagents, purification matrices, and an instruction manual. In some embodiments, the kit of the present invention contains a non-canonical nucleotide. Suitable non-canonical nucleotides include any nucleotides provided herein including but not limited to dUTP, or a methylated purine.

[00187] In some embodiments, the kit may contain one or more reaction mixture components, or one or more mixtures of reaction mixture components. In some cases, the reaction mixture components or mixtures thereof may be provided as concentrated stocks, such as 1.1x, 1.5x, 2x, 2.5x, 3x, 4x, 5x, 6x, 7x, 10x, 15x, 20x, 25x, 33x, 50x, 75x, 100x or higher concentrated stock. The reaction mixture components may include any of the compositions provided herein including but not limited to buffers, salts, divalent cations, azeotropes, chaotropes, dNTPs, labeled nucleotides, non-canonical nucleotides, dyes, fluorophores, biotin, enzymes (such as endonucleases, exonucleases, glyosylases), or any combination thereof.

[00188] In some embodiments, the kit may contain one or more oligonucleotide primers, such as the oligonucleotide primers provided herein. For example, the kit may contain one or more oligonucleotide primers comprising random hybridizing portions. Alternatively, the kit may contain oligonucleotide primers comprising polyT hybridizing portions. In some cases, the kit may contain oligonucleotide primers that comprise random hybridizing portions and primers comprising polyT hybidizing portions. In still other cases, the kit may contain "not so random" primers that have been pre-selected to hybridize to desired nucleic acids, but not hybridize to undesired nucleic acids. In some cases the kit may contain tailed primers comprising a 3'-portion hybridizable to the target nucleic acid and a 5'-portion which is not hybridizable to the target nucleic acid. In some cases, the kit may contain

IDT Exhibit 1020.5612

chimeric primers comprising an RNA portion and a DNA portion. In some cases, the kit may contain primers comprising non-canonical nucleotides.

[00189] In some embodiments, the kit of the present invention may contain one or more polymerases or mixtures thereof. In some cases, the one or more polymerases or mixtures thereof may comprise strand displacment activity. Suitable polymerases include any of the polymerases provided herein. The kit may further contain one or more polymerase substrates such as for example dNTPs, non-canonical nucleotides, or labeled nucleotides.

[00190] In some embodiments, the kit of the present invention may contain one or more agents capable of inhibiting the DNA-dependent DNA polymerase activity or the RNA-dependent DNA polymerase (reverse transcriptase), such as actinomycin D (dactinomycin), alpha-amanitin, aphidicolin, BP5, novobiocin, rifampicin, rifamycin, sulfoquinovosylmonoacylglycerol, sulfoquinovosyldiacylglycerol, ursane, oleanane triterpenoids, ursolic acid, oleanolic acid, mikanolide, dihydromikanolide, dehydroaltenusin, catapol, taxinine, cephalomanninine, dipeptide alcohols, corylifolin; bakuchiol; resveratrol; Neobavaisoflavone; daidzein; bakuchicin, anacardic acid and oleic acid.

[00191] In some embodiments, the kit of the present invention may contain one or more means for purification of the nucleic acid products, removing of the fragmented products from the desired products and removing the inhibitor of DNA-dependent DNA polymerase activity, or combination of the above. Suitable means for the purification of the nucleic acid products include but are not limited to single stranded specific exonucleases, affinity matrices, nucleic acid purification columns, spin columns, ultrafiltration or dialysis reagents, or electrophoresis reagents including but not limited acrylamide or agarose, or any combination thereof

[00192] In some embodiments, the kit of the present invention may contain one or more agents capable of cleaving the base portion of a non-canonical nucleotide to generate an abasic site. In some cases, this agent may comprise one or more glycosylases. Suitable glycosylsases include any glycosylases provided herein including but not limited to UDG, or MPG.

[00193] In some embodiments, the kit of the present invention may contain one or more agents capable of fragmenting a phosphodiester backbone at an abasic site to fragment the input nucleic acid template. In some cases, this agent may comprise one or more amines, primary amines, secondary amines, polyamines, piperidine, AP endonucleases, or any combination thereof.

[00194] In some embodiments, the kit of the present invention may contain one or more compositions or reagents for removing single stranded nucleic acid from the reaction mixture or purifying the double stranded fragments produced in the methods of the present invention. Suitable compositions or reagents include but are not limited to single stranded specific exonucleases, affinity matrices, nucleic acid purification columns, spin columns, ultrafiltration or dialysis reagents, or electrophoresis reagents including but not limited acrylamide or agarose, or any combination thereof.

[00195] In some embodiments, the kits of the present invention may contain one or more reagents for end labeling of the primer extension products such as terminal transferases and labeled nucleotides or labeled terminator nucleotides such as dideoxynucleotides or acyclonucleotides.

[00196] In some embodiments, the kit of the present invention may contain one or more reagents for producing blunt ends from the double stranded products generated by the extension reaction. For example, the kit may contain one or more of single stranded DNA specific exonucleases including but not limited to exonuclease 1 or exonuclease 7;a single stranded DNA specific endonucleases such as mung bean exonuclease or S1 exonuclease, one or more polymerases such as for example T4 DNA polymerase or Klenow polymerase, or any mixture thereof. Alternatively, the kit may contain one or more single stranded DNA specific exonucleases, endonucleases and one

IDT Exhibit 1020.5613

or more polymerases, wherein the reagents are not provided as a mixture. Additionally, the reagents for producing blunt ends may comprise dNTPs.

[00197] In some embodiments, the kit of the present invention may contain one or more reagents for preparing the double stranded products for ligation to adaptor molecules. For example, the kit may contain dATP, dCTP, dGTP, dTTP, or any mixture thereof. In some cases, the kit may contain a polynucleotide kinase, such as for example T4 polynucleotide kinase. Additionally, the kit may contain a polymerase suitable for producing a 3' extension from the blunt ended double stranded DNA fragments. Suitable polymerases are included, for example, exo- Klenow polymerase.

[00198] In some embodiments, the kit of the present invention may contain one or more adaptor molecules such as any of the adaptor molecules provided herein. Suitable adaptor molecules include single or double stranded nucleic acid (DNA or RNA) molecules or derivatives thereof, stem-loop nucleic acid molecules, double stranded molecules comprising one or more single stranded overhangs of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 bases or longer, proteins, peptides, aptamers, organic molecules, small organic molecules, or any adaptor molecules known in the art that may be covalently or non-covalently attached, such as for example by ligation, to the double stranded DNA fragments.

[00199] The kit may further contain instructions for the use of the kit. For example, the kit may contain constructions for generating double stranded fragments of nucleic acids useful for large scale analysis including but not limited to e.g., microarray based analysis such as comparative genomic hybridization, highly parallel SNP analysis, pyrosequencing, sequencing by synthesis, sequencing by hybridization, single molecule sequencing, nanopore sequencing, and sequencing by ligation, high density PCR, microarray hybridization, SAGE, digital PCR, and massively parallel Q-PCR; subtractive hybridization; differential amplification; preparation of libraries (including cDNA and differential expression libraries); preparation of an immobilized nucleic acid (which can be a nucleic acid immobilized on a microarray), and characterizing amplified nucleic acid products generated by the methods of the invention, or any combination thereof. In some cases, the kit may contain instructions for providing or generating an input nucleic acid template comprising one or more non-canonical nucleotides. The kit may further contain instructions for mixing the one or more reaction mixture components to generate one or more reaction mixtures suitable for the methods of the present invention. The kit may further contain instructions for hybridizing the one or more oligonucleotide primers to an input nucleic acid template. The kit may further contain instructions for extending the one or more oligonucleotide primers with for example a polymerase, including but not limited to a polymerase comprising strand displacement activity. The kit may further contain instructions for cleaving the base portion of a non-canonical nucleotide to generate an abasic site, with for example a glycosylase. The kit may further contain instructions for fragmenting a phosphodiester backbone at an abasic site to fragment the input nucleic acid template, with for example any of the suitable agents provided herein such as a polyamine. The kit may further contain instructions for purification of any of the products provided by any of the steps of the methods provided herein. The kit may further contain instructions for degradation of single stranded nucleic acids present within the reaction mixture with for example a single stranded DNA specific exonuclease. The kit may further contain instructions for producing blunt ended fragments, for example by removing single stranded overhangs or filling in single stranded overhangs, with for example single stranded DNA specific exonucleases, polymerases, or any combination thereof. The kit may further contain instructions for phosphorylating the 5' ends of the double stranded DNA fragments produced by the methods of the present invention. The kit may further contain instructions for ligating one or more adaptor molecules to the double stranded DNA fragments of the present invention. The kit may

IDT Exhibit 1020.5614

further contain instructions for use of the double stranded DNA fragments for downstream analysis such as hybridization, array hybridization, microarray hybridization, large-scale sequencing, or polymorphism detection.

V. Examples

[00200] *Example 1: Whole transcriptome amplification followed by random priming to generate dsDNA products with reduced (degradation) of input amplified products:*

[00201] Linear Whole Transcriptome Amplification is carried out using WT-Ovation Pico RNA Amplification kit (NuGEN Technologies Inc, San Carlos) following the manufacturer instructions (http://www.nugeninc.com/tasks/sites/nugen/assets/File/user_guides/userguide_wt_ov_pico.pdf) to generate template nucleic acid. The input total RNA for each reaction is 10ng of total RNA samples (HeLa total RNA and Brain total RNA from Ambion and total RNA from a biological sample). The amplification is performed in the presence of all four dNTPs as well as dUTP to render the template nucleic acid susceptible to fragmentation by the combined action of UDG and DMED.

[00202] The amplified cDNA is purified using *QIAquick® PCR Purification Kit, Cat. #28104* as describe in the User Guide (reference above).

[00203] Random priming and extension of the amplified cDNA (i.e. template nucleic acid) is carried out using the WT-Ovation Exon Module (http://www.nugeninc.com/nugen/index.cfm/products/target-prep-modules/wt-ovation-exon-module/) with a modified buffer E2. The buffer is prepared with all four dNTPs without dUTP, so as to render the random priming products resistant to fragmentation by the combined action of UDG and DMED. 1μg of WT-Ovation Pico amplified and purified DNA (i.e. template nucleic acid) is combined with 12μL water, 20μL of the modified E2 buffer (no dUTP) to a total volume of 32μL. The mixture is heated at 98°C for 3 min. and cooled down to 4°C. Following this denaturation step, 3μL each of random primer and enzyme reagent (E1 and E3) are added to the mixture and the reaction mixture is incubated for the following temperature and durations: 4°C for 1 min., 30°C for 10 min., 42°C fro 15 and 75°C for 10 min. (enzyme inactivation). The primer extension reaction products (including the input template nucleic acid) are purified using Agencourt RNAClean® magenetic beads. The primer extension products (including the input template nucleic acid) are eluted from the beads in 20μL water. Yield is measured using Nanodrop and electrophoretic mobility is determined using BioAnanlyzer (Agilent).

[00204] 4-5 μg of purified primer extension products, including the input template nucleic acid comprising dUTP (in 16μl), is mixed with 3μL of fragmentation buffer (DMED, NuGEN's FL-Ovation Kit (http://www.nugeninc.com/nugen/index.cfm/products/target-prep-modules/fl-ovation-biotin-v21) and 1μL of UDG (2U/μL; NEB). The mixture is incubated at 37°C for 30 min.

[00205] Following fragmentation of the template nucleic acid, the primer extension products are purified using Agencourt RNAClean magnetic beads (following manufacturer instruction) and the purified products are eluted in 20μL water. Yield and electrophoretic mobility profile are determined as above.

[00206] Results: The BioAnalyzer profiles for the linearly amplified DNA, the product of random-primed primer extension reaction on the amplified DNA, and the product of random priming followed by cleavage of the input amplified DNA generated for the various total RNA samples are shown on **Figures 4** and **5**, below.

[00207] Amplified cDNA (the product of linear whole transcriptome amplification from the various total RNA samples) generated in the presence of non canonical nucleotide dUTP, can be fragmented by the combined action of UDG and DMED, while the random primed primer extension product generated in the absence of the non canonical nucleotide is not susceptible to cleavage by these combined reagent. Thus, the input amplified cDNA is

IDT Exhibit 1020.5615

distinguishable from the random priming products. Cleavage of the input amplified cDNA enables clean-up of the mixture of the random priming products away from the input, as is demonstrated by the reduced size distribution as indicated by the BioAnanlyzer (Agilent) profiles, which reflect change in the length of the multiplicity of species in the analyzed sample.

[00208] The sample input on to the Bioanalyzer chips is 1.5μL containing various amount of DNA: Input amplified cDNA is about 200ng/μL, random priming product (prior to cleavage step) is about 170ng/μL and the random priming product following cleavage of the input amplified cDNA is about 100ng/μL.

[00209] *Example 2: Preparation of double stranded products of the present invention for high throughput sequencing:*

[00210] The double stranded amplification products of Example 1 are blunt ended using a single stranded exonuclease such as exo1 or exo7 for degrading single stranded overhangs. Alternatively a polymerase such as T4 DNA polymerase or Klenow DNA polymerase is utilized for filling in single stranded overhangs, or a combination of exonuclease and polymerase. The blunt ended double stranded products are then phosphorylated with T4 polynucleotide kinase. The phosphorylated double stranded nucleic acid amplification products are then ligated to adaptor nucleic acids. The amplification products are then encapsulated in a water-in-oil emulsion and amplified in a clonal fashion in the presence of particles comprising sequences complementary to the adaptor nucleic acids. The resulting particles comprising clonally amplified amplification products are then sequenced using a Genome Sequencer manufactured by 454/Roche LifeSciences.

[00211] The above-described method may be automated such that the sequencing reactions are performed via robotics. In addition, the sequencing information and data obtained may be provided to a personal computer, a personal digital assistant, a cellular phone, a video game system, or a television so that a user can monitor the progress of the sequencing reactions remotely. This process is illustrated, for example, in **Figure 6**. The performing, monitoring and obtaining of results of the sequencing reactions can be done from a place other than where the sequencing apparatus is located, for example from a physically separate room, building, city, state, country or the like. Likewise, the results can be transmitted to the remote user from a physically separate room, building, city, state, country or the like.

[00212] 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 1020.5616

CLAIMS

WHAT IS CLAIMED IS:

1. A method comprising:

(a) providing an input nucleic acid template in a reaction mixture;

(b) hybridizing one or more oligonucleotide primers to the input nucleic acid template;

(c) extending the one or more oligonucleotide primers along the input nucleic acid template with a polymerase comprising strand displacement activity, wherein the hybridizing and extending produces primer extension products comprising one or more double stranded products; and

(d) cleaving the input nucleic acid template with an agent comprising one or more cleavage reagents.

2. The method of claim 1, wherein the input nucleic acid template comprises one or more non-canonical nucleotides.

3. The method of claim 2, wherein the method further comprises generating one or more blunt ended double stranded products from the one or more double stranded products.

4. The method of claim 3 wherein the generating one or more blunt ended double stranded products comprises using an enzyme.

5. The method of claim 4 wherein the enzyme is an exonuclease, an endonuclease, or a combination thereof.

6. The method of claim 5 wherein the enzyme is an endonuclease.

7. The method of claim 6 wherein the endonuclease is S1 endonuclease, mung bean endonuclease, or a combination thereof.

8. The method of claim 6 wherein the method further comprises generating a library of blunt ended double stranded products.

9. The method of claim 6 wherein the method further comprises analyzing the blunt ended double stranded products by next generation sequencing.

10. The method of claim 2 wherein the non-canonical nucleotides comprise dUTP.

11. The method of claim 2, wherein the cleaving step cleaves a base portion of the non-canonical nucleotide, thereby forming an abasic site.

12. The method of claim 2, wherein the agent comprises a glycosylase.

13. The method of claim 12, wherein the glycosylase is UNG or UDG.

14. The method of claim 2, wherein the agent comprises a primary amine.

15. The method of claim 2, wherein the agent comprises a polyamine.

16. The method of claim 15, wherein the polyamine is DMED.

17. The method of claim 2, wherein the agent comprises a glycosylase and a polyamine.

IDT Exhibit 1020.5617

18. The method of claim 1, wherein the input template nucleic acid comprises DNA or RNA, or complements thereof or products of amplification of the input DNA or RNA template.

19. The method of claim 1, wherein steps (b) and (c) are performed simultaneously.

20. The method of claim 1, wherein steps (b) and (c) are performed sequentially.

21. The method of claim 1, wherein the one or more oligonucleotide primers comprise a label.

22. The method of claim 1, wherein the one or more oligonucleotide primers comprise an amino-allyl label.

23. The method of claim 1, wherein the one or more oligonucleotide primers comprise a random hexamer, heptamer, octomer, nonamer, decamer, undecamer, dodecamer, or tridecamer.

24. The method of claim 1, wherein the one or more oligonucleotide primers comprise a poly T sequence.

25. The method of claim 1, wherein the one or more oligonucleotide primers comprise a mixture of random hexamers, heptamers, octomers, nonamers, decamers, undecamers, dodecamers, or tridecamers; and a poly T sequence.

26. The method of claim 1 wherein one or more oligonucleotide primers comprise a template hybridizing portion and a tailed, non template hybridizing, portion.

27. The method of claim 1 wherein one or more oligonucleotide primers comprise chimeric primers.

28. The method of claim 27 wherein the chimeric primers comprise a DNA portion and a 5'-RNA portion.

29. The method of claim 1, wherein the method further comprises degrading single stranded DNA in the reaction mixture.

30. The method of claim 29, wherein the method further comprises degrading single stranded DNA in the reaction mixture with an ssDNA specific exonuclease, an ss DNA specific endonuclease, or a combination thereof.

31. The method of claim 30, wherein the exonuclease is exonuclease 1, exonuclease 7 or a combination thereof.

32. The method of claim 20, wherein the endonuclease is S1 endonuclease, mung bean endonuclease, or a combination thereof.

33. The method of claim 1, wherein the strand displacing polymerase comprises Klenow polymerase, exo – Klenow polymerase, 5'-3' exo – Klenow polymerase, Bst polymerase, Bst large fragment polymerase, Vent polymerase, , Vent polymerase, Deep Vent (exo-) polymerase, 9°Nm polymerase, Therminator polymerase, Therminator II polymerase, MMulV Reverse Transcriptase, phi29 polymerase, or DyNAzyme EXT polymerase, or a combination thereof.

34. The method of claim 1, wherein the method further comprises phosphorylating the 5' ends of the one or more blunt ended double stranded products.

35. The method of claim 1, wherein the method further comprises extending the 3' end of the one or more blunt ended double stranded products.

IDT Exhibit 1020.5618

36. The method of claim 35, wherein the method further comprises ligating the double stranded products with one or more double stranded adapter oligonucleotides.

37. The method of claim 1, further comprising subjecting the input nucleic acid template to PCR, strand displacement amplification (SDA), multiple displacement amplification (MDA), rolling circle amplification (RCR), single primer isothermal amplification (SPIA), Ribo-SPIA, ligase chain reaction (LCR), Nucleic Acid Sequence Based Amplification (NASBA), Q-Beta Replicase amplification, Self-sustained sequence replication (3SR) or ligation activated transcription (LAT).

38. A method comprising:

(a) contacting an input template nucleic acid comprising one or more non-canonical nucleotides in a reaction mixture comprising:

(i) one or more oligonucleotide primers;

(ii) a polymerase comprising strand displacement activity;

(iii) an agent capable of cleaving a base portion of a non-canonical nucleotide, whereby an abasic site is generated; and

(iv) an agent capable of fragmenting a phosphodiester backbone at the abasic site;

whereby double stranded DNA fragments are generated.

39. A method comprising:

(a) providing an input nucleic acid template comprising one or more non-canonical nucleotides;

(b) amplifying the input nucleic acid template to produce amplification products in a reaction mixture comprising oligonucleotide primers comprising random hybridizing portions, and an enzyme comprising strand displacement activity; and

(c) fragmenting the input nucleic acid template, wherein the fragmenting step is performed by adding to the reaction mixture comprising amplification products an agent that cleaves a base portion of a non-canonical nucleotide to generate an abasic site and an agent that cleaves a phosphodiester backbone of a nucleic acid at an abasic site.

40. The method of claim 38 further comprising:

(d) optionally separating the amplification products from the fragmentation products;

(e) treating the amplification products with one or more agents to produce double stranded blunt-ended amplification products; and

(f) phosphorylating the 5' ends of the double stranded blunt-ended amplification products.

41. The method of claim 40, wherein step (d) is performed after step (f).

42. The method of claim 40, wherein step (d) is performed before step (e).

43. The method of claim 40, wherein step (d) is not performed.

IDT Exhibit 1020.5619

44. The method of claim 40 further comprising ligating the one or more phosphorylated double stranded blunt-ended amplification products with adaptor nucleic acid molecules.

45. The method of claim 44, wherein the amplification products are analyzed by next generation sequencing.

46. A kit comprising:

(a) a glycosylase;

(b) a polyamine, an AP endonuclease, or a combinatin thereof; and

(c) one or more double stranded adapter oligonucleotides.

47. The kit of claim 46, wherein the kit further comprises instructions for the use of said kit.

48. The kit of claim 47, wherein the instructions comprise a method for generating dsDNA fragments, said method comprising:

(a) providing an input nucleic acid template comprising one or more non-canonical nucleotides in a reaction mixture;

(b) hybridizing and extending one or more oligonucleotide primers to the input nucleic acid template with an enzyme comprising strand displacement activity, wherein the hybridizing and extending produces one or more double stranded products; and

(c) cleaving the input nucleic acid template.

49. The kit of claim 46, wherein the kit further comprises one or more oligonucleotide primers.

50. The kit of claim 49, wherein the one or more oligonucleotide primers comprise random hybridizing portions.

51. The kit of claim 49, wherein the one or more oligonucleotide primers comprise polyT hybridizing portions.

52. The kit of claim 49, wherein the one or more oligonucleotide primers comprise a mixture of random hybridizing portions and polyT portions.

53. The kit of claim 49, wherein at least one of the one or more oligonucleotide primers is a composite primer.

54. The kit of claim 49, wherein at least two of the one or more oligonucleotide primers is a composite primer.

55. A composition comprising:

(a) an input template nucleic acid comprising one or more non-canonical nucleotides;

(b) one or more oligonucleotide primers comprising randomized hybridizing portions;

(c) a polymerase comprising strand-displacement activity;

(d) a glycosylase; and

IDT Exhibit 1020.5620

(e) one or more double stranded DNA product molecules.

56. A kit comprising:

(a) an input template nucleic acid;

(b) one or more oligonucleotide primers comprising randomized hybridizing portions;

(c) a polymerase comprising strand-displacement activity; and

(d) an endonuclease.

57. A method comprising:

(a) providing an input RNA template in a reaction mixture;

(b) hybridizing a first primer to the input RNA template;

(c) reverse transcribing the input RNA template in the presence of one or more non-canonical nucleotides to generate a first strand cDNA;

(d) cleaving the input RNA template;

(e) performing second strand synthesis to generate a double stranded cDNA; and

(f) cleaving the first strand cDNA to generate a single stranded second strand cDNA.

58. The method of claim 57, wherein said reverse transcribing comprises use of an RNA-dependent DNA polymerase.

59. The method of claim 58, wherein the reaction mixture further comprises an inhibitor of DNA-dependent DNA polymerase activity.

60. The method of claim 59, wherein the inhibitor of DNA-dependent DNA polymerase activity inhibits the DNA-dependent DNA polymerase activity of the RNA-dependent DNA polymerase.

61. The method of claim 59, wherein said inhibitor is actinomycin D.

62. The method of claim 59, wherein said inhibitor is removed prior to the performing second strand synthesis.

63. The method of claim 57, wherein the first primer comprises a 5'-tail sequence that does not hybridize to the input RNA template.

64. The method of claim 63, wherein the 5'-tail sequence comprises DNA.

65. The method of claim 57, wherein the 3'-end of the first primer hybridizes to the input RNA template.

66. The method of claim 57, wherein the reverse transcribing comprises extension of the first primer hybridized to the input RNA template.

67. The method of claim 57, wherein the non-canonical nucleotide is dUTP.

IDT Exhibit 1020.5621

68. The method of claim 57, wherein said second strand synthesis comprises primer extension of a second primer hybridized to the first strand cDNA.

69. The method of claim 68, wherein said second primer comprises a 3'-sequence hybridizable to the first strand cDNA.

70. The method of claim 68, wherein said second primer comprises a 5'-tail that is not complementary to the first strand cDNA.

71. The method of claim 57, wherein said second strand synthesis is carried out by DNA polymerase.

72. The method of claim 57, wherein said second strand synthesis is carried out in the absence of non-canonical nucleotide triphosphates.

73. The method of claim 57, wherein cleaving the input RNA template in a complex with the first primer extension product or products comprises exposing the input RNA template to an RNase H with or without other RNases, or cleaving the input RNA template following first primer extension reaction by heat or chemical treatment or combination thereof.

74. The method of claim 57, wherein cleaving the first strand cDNA comprises combining the reaction mixture with an enzyme that cleaves the base portion of the non-canonical nucleotide to generate an abasic site.

75. The method of claim 74, wherein said enzyme that cleaves the base portion of the non-canonical nucleotide to generate an abasic site is a glycosylase.

76. The method of claim 74, wherein said glycosylase is UNG or UDG.

77. The method of claim 74, wherein the reaction mixture further comprises an amine.

78. The method of claim 77, wherein said amine is DMED.

79. The method of claim 57, wherein said cleaving the first strand cDNA generates fragments of the first strand cDNA with blocked 3'-ends.

80. The method of claim 57, further comprising sequencing said single stranded second strand cDNA.

81. The method of claim 80, wherein said sequencing comprises next generation sequencing.

82. The method of claim 57, wherein said single stranded second strand cDNA comprises sequence homologous to the input RNA template flanked by 3' and/or 5' sequences compatible for use in DNA sequencing.

83. The method of claim 57, wherein said single stranded second strand cDNA comprises sequence homologous to the input RNA template flanked by 3' and/or 5' comprising sequences that function as barcodes.

IDT Exhibit 1020.5622

84. The method of claim 57, wherein said single stranded second strand cDNA comprises sequence homologous to the input RNA template flanked by 3' and/or 5' sequences that comprise recognition sequence for one or more restriction enzymes.

85. The method of claim 57, wherein said single stranded second strand cDNA comprises sequence homologous to the input RNA template flanked by 3' and/or 5' sequences that enables circularization by hybridizing an oligonucleotide complementary to the end sequences and ligation of the ends.

86. The method of claim 57, further comprising amplification of the cDNA by rolling circle amplification.

87. The method of claim 57, further comprising amplifying the single stranded cDNA by single primer isothermal amplification.

88. The method of claim 57, wherein the input RNA template is from a biological sample.

89. The method of claim 57, wherein the input RNA template is from a sample lysate.

90. The method of claim 57, wherein the input RNA template is from a cell free fluid.

91. The method of claim 57, wherein the cell free fluid is plasma or serum.

92. The method of claim 57, wherein the input RNA template is fragmented RNA.

93. The method of claim 92, wherein the input RNA template is from an FFPE sample.

94. The method of claim 92, wherein the input RNA template is fragmented by treatment with heat in the presence of multivalent cations.

IDT Exhibit 1020.5623

Figure 1-a

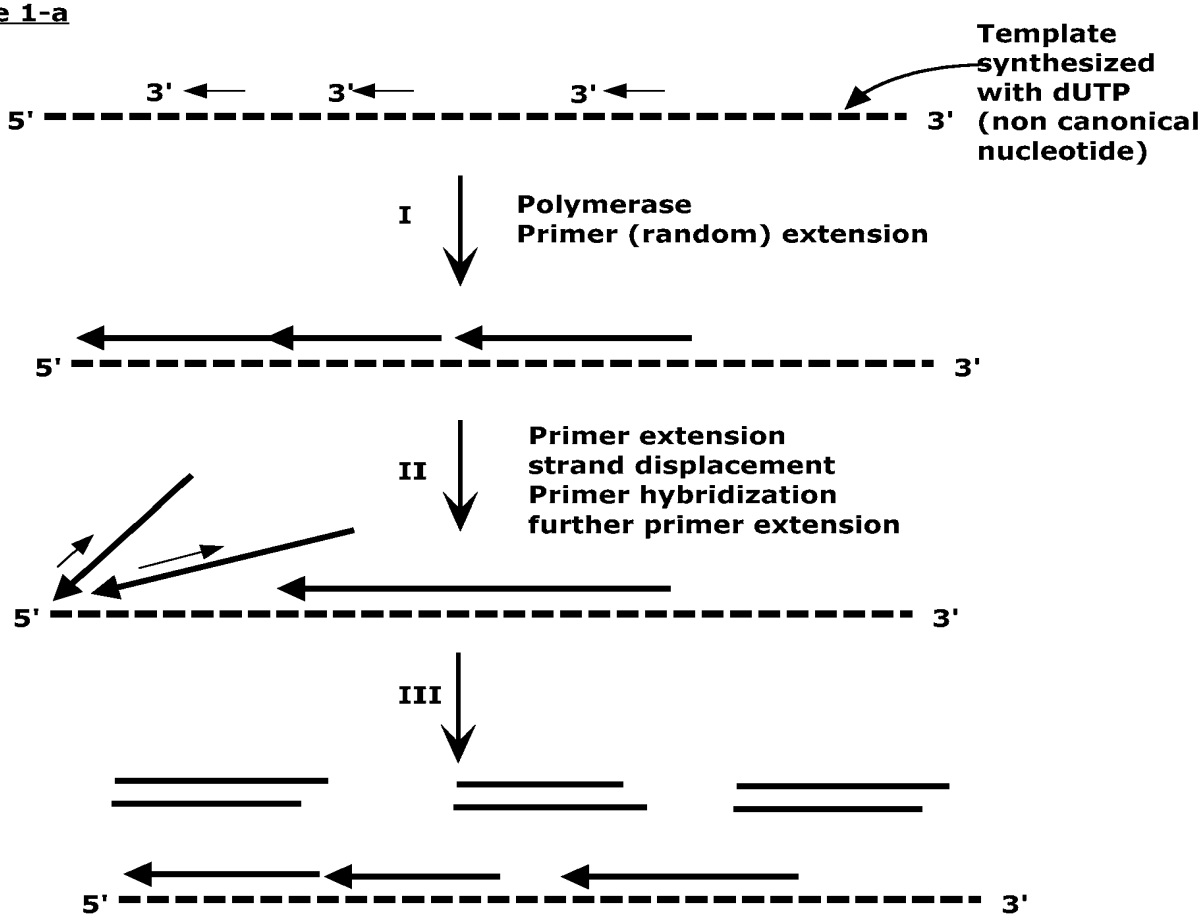
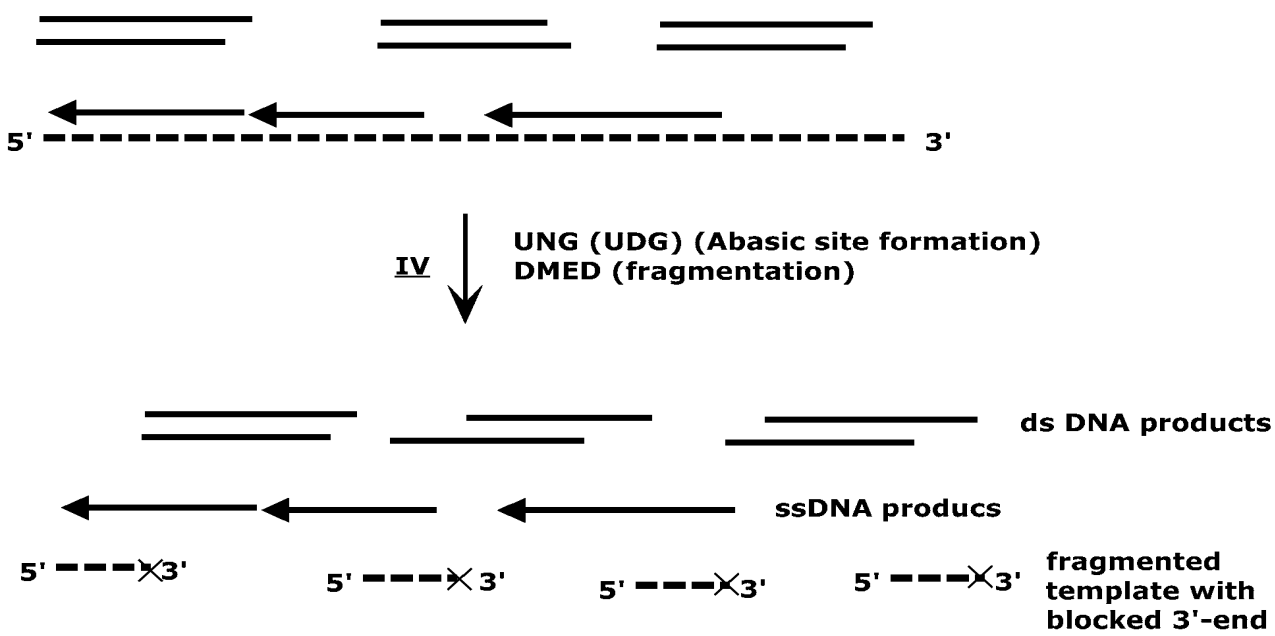
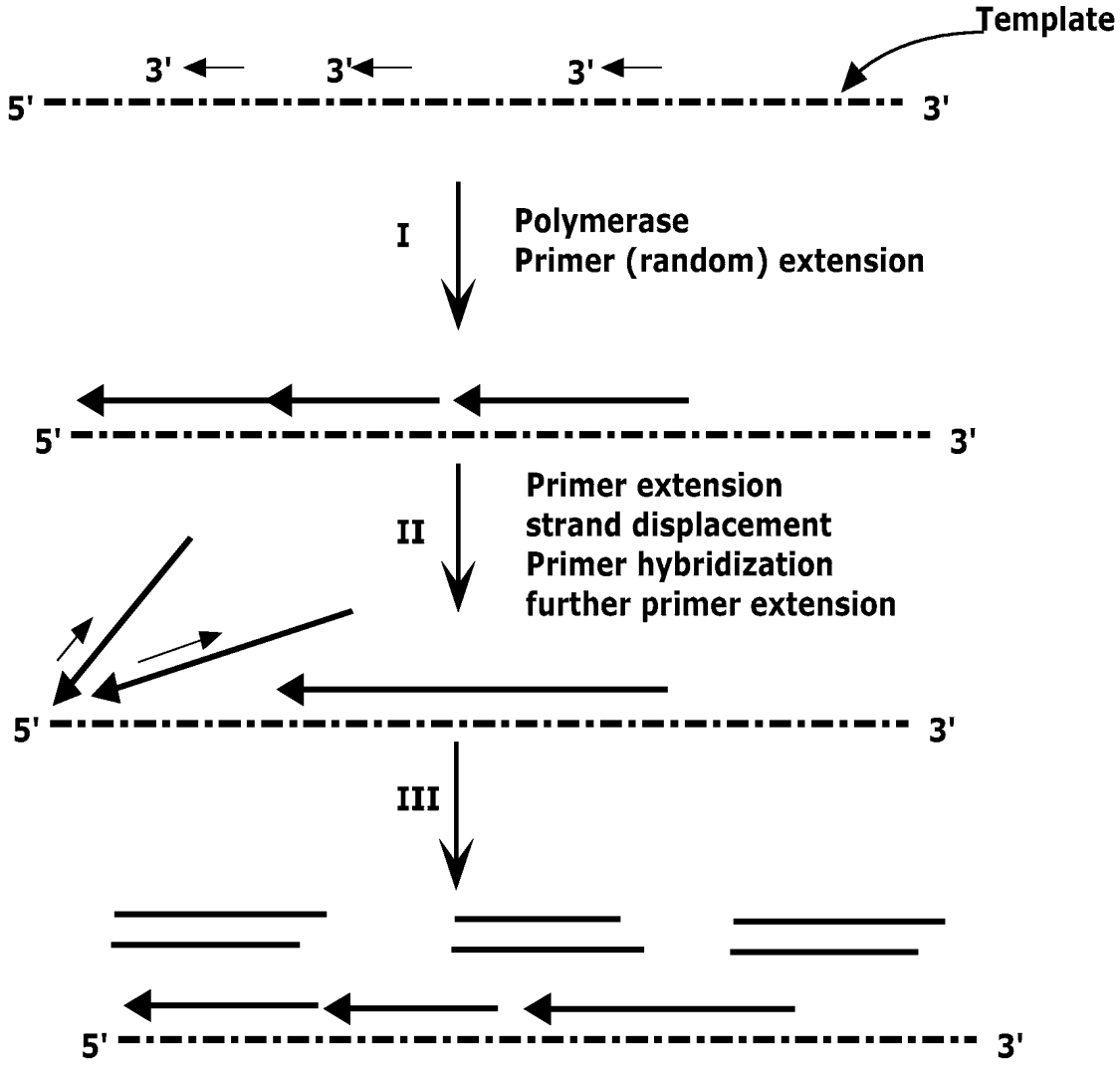

Template synthesized with dUTP (non canonical nucleotide)
3'
3'
3'
5'
3'
I
Polymerase
Primer (random) extension
5'
3'
II
Primer extension
strand displacement
Primer hybridization
further primer extension
5'
3'
III
5'
3'

Figure 1-a continued



ds products can be blunt ended as in Figure 2.
The fragmented template and other single stranded products can be purified out
The 3'-end of the fragmented product is blocked

Figure 1-b



Blunt end ds products can be generated as in Figure 2.

Figure 2: Generation of ds products ready for adaptor ligation as required, for example, for generation of libraries for Next Generation Sequencing

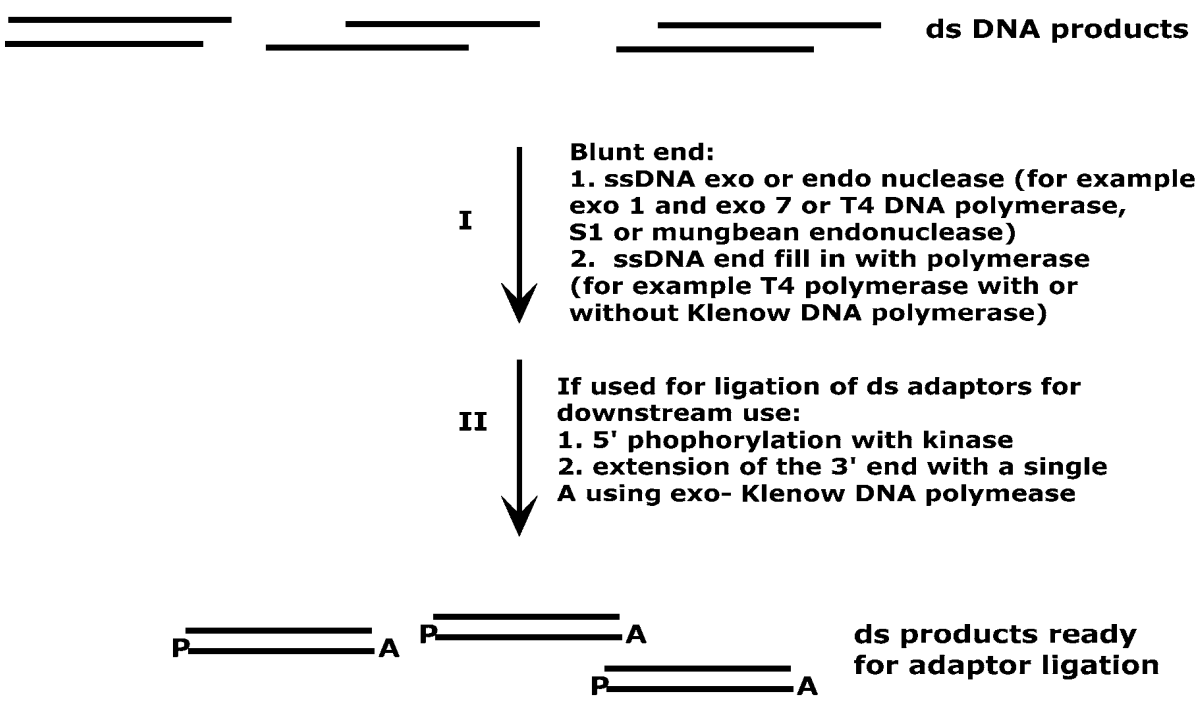


Figure 3

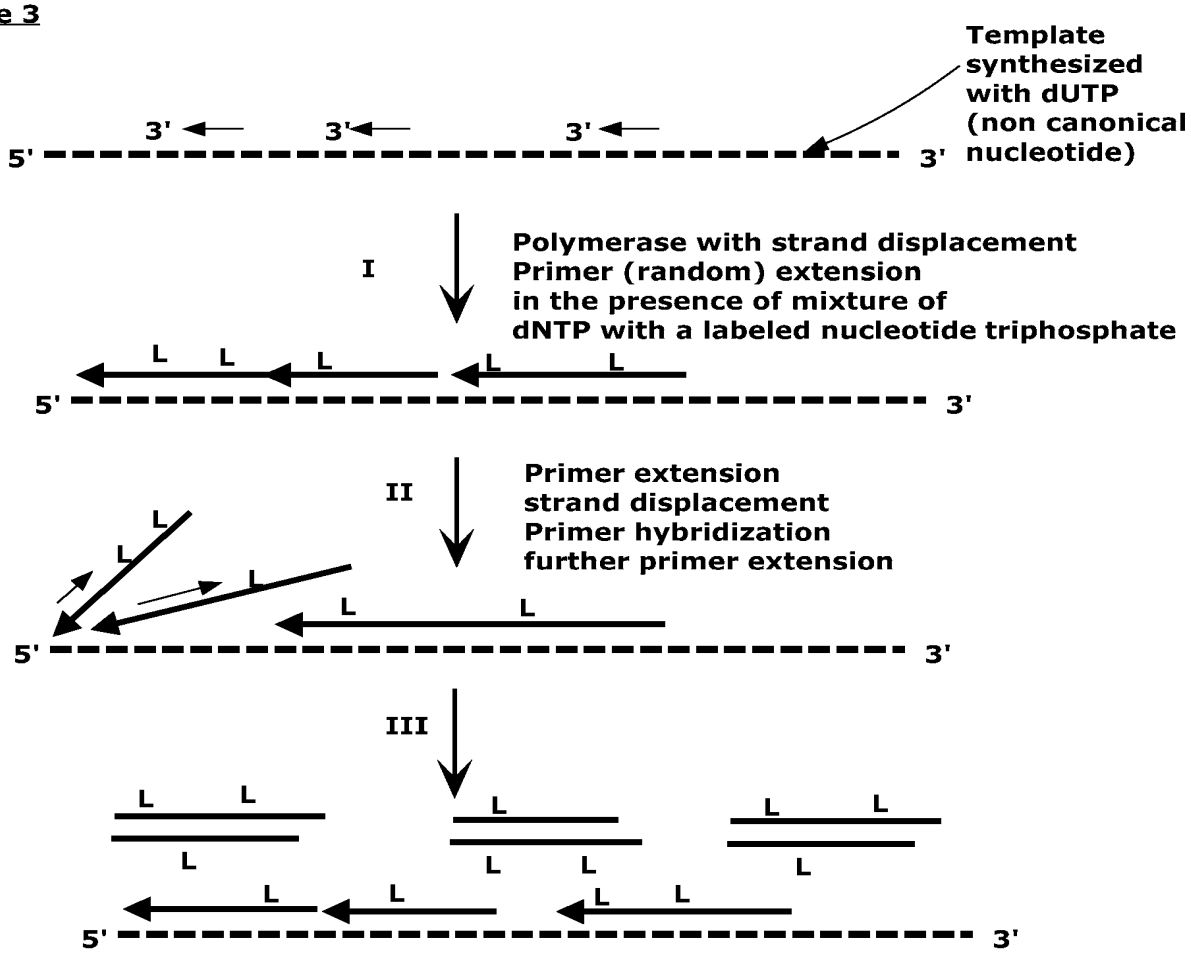

Template synthesized with dUTP (non canonical nucleotide)
3'
5'
I
Polymerase with strand displacement
Primer (random) extension
in the presence of mixture of
dNTP with a labeled nucleotide triphosphate
L
II
Primer extension
strand displacement
Primer hybridization
further primer extension
III

Figure 3 continued

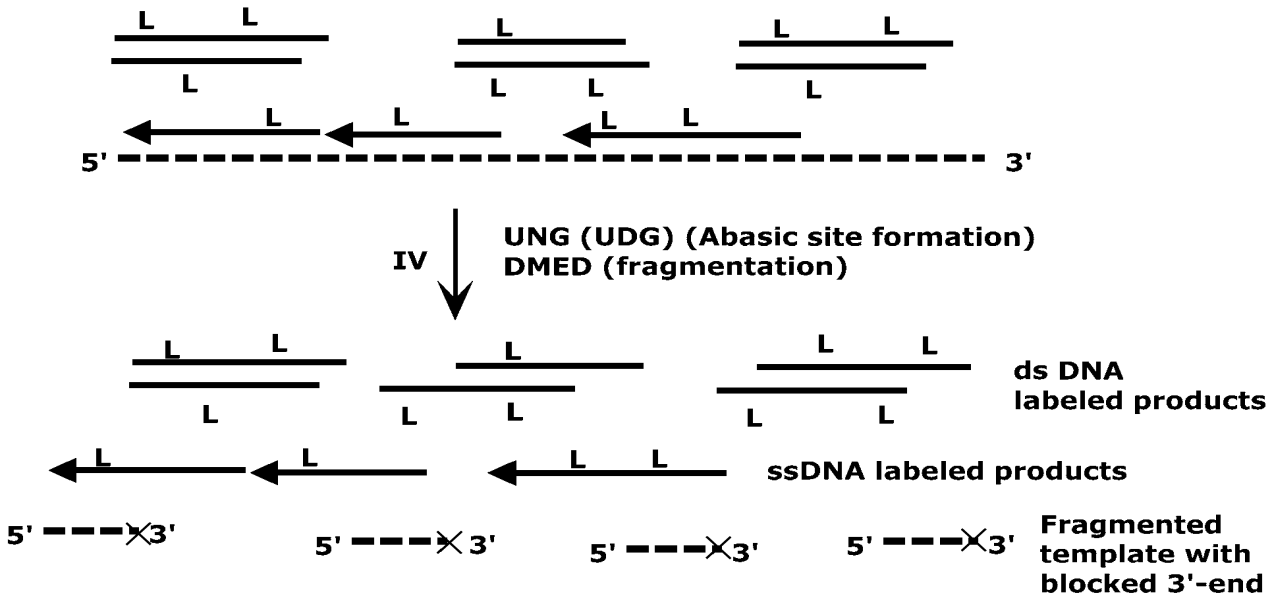


Figure 4

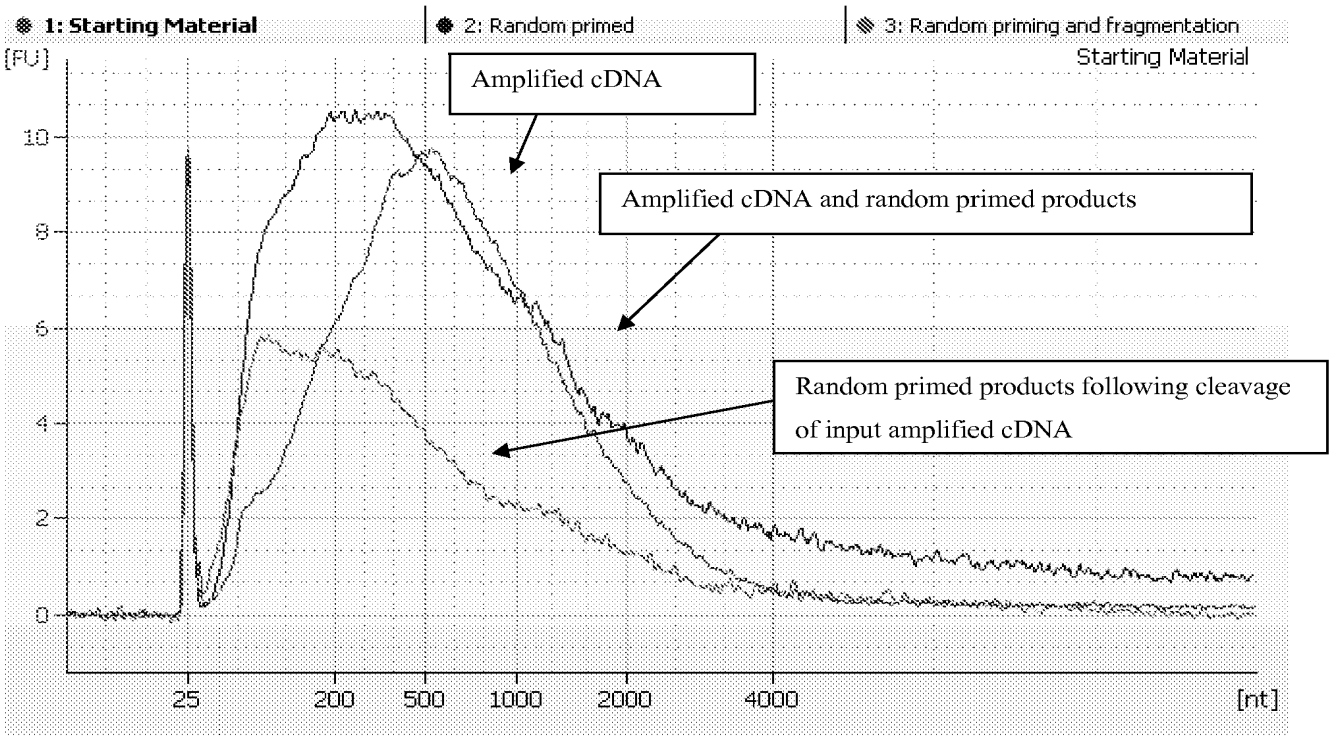

1: Starting Material
2: Random primed
3: Random priming and fragmentation
Starting Material
[FU]
10
8
6
4
2
0
25
200
500
1000
2000
4000
[nt]
Amplified cDNA
Amplified cDNA and random primed products
Random primed products following cleavage of input amplified cDNA

Figure 5

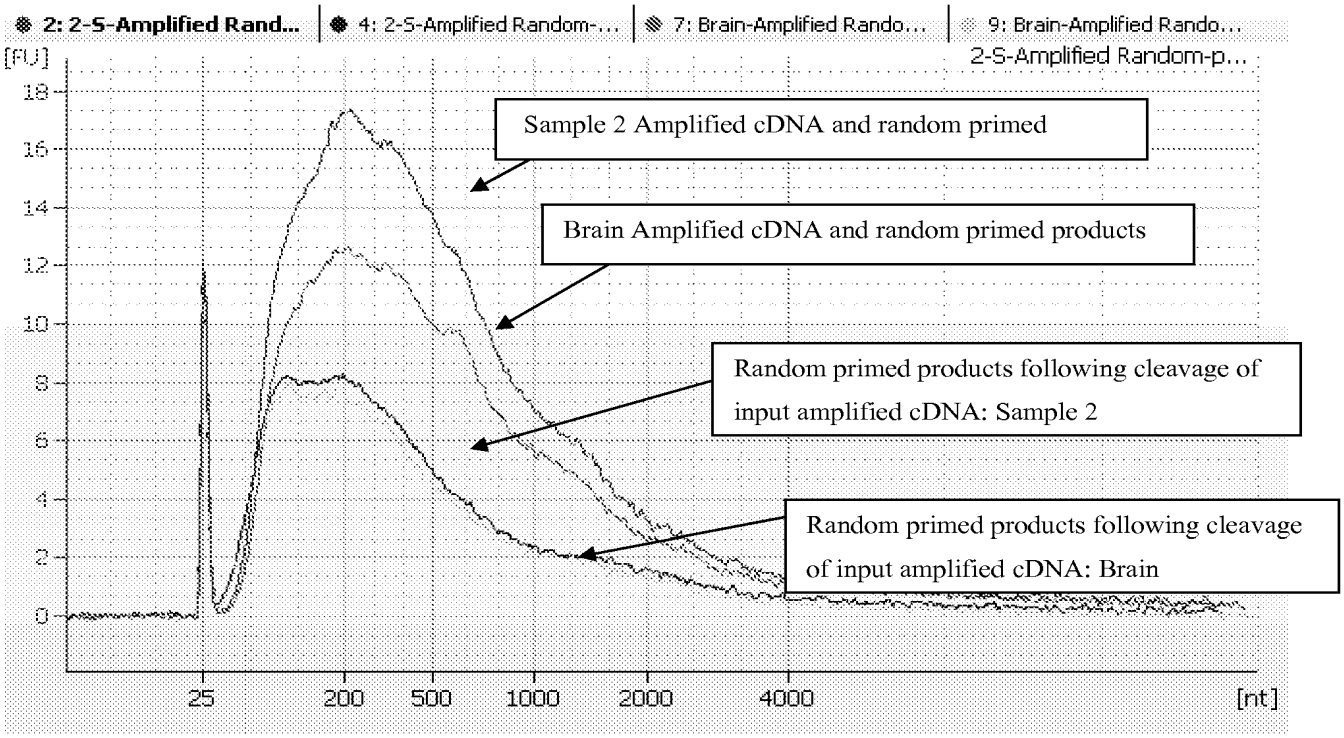

2: 2-S-Amplified Rand...
4: 2-S-Amplified Random-...
7: Brain-Amplified Rando...
9: Brain-Amplified Rando...
2-S-Amplified Random-p...
[FU]
18
16
14
12
10
8
6
4
2
0
25
200
500
1000
2000
4000
[nt]
Sample 2 Amplified cDNA and random primed
Brain Amplified cDNA and random primed products
Random primed products following cleavage of input amplified cDNA: Sample 2
Random primed products following cleavage of input amplified cDNA: Brain

Figure 6

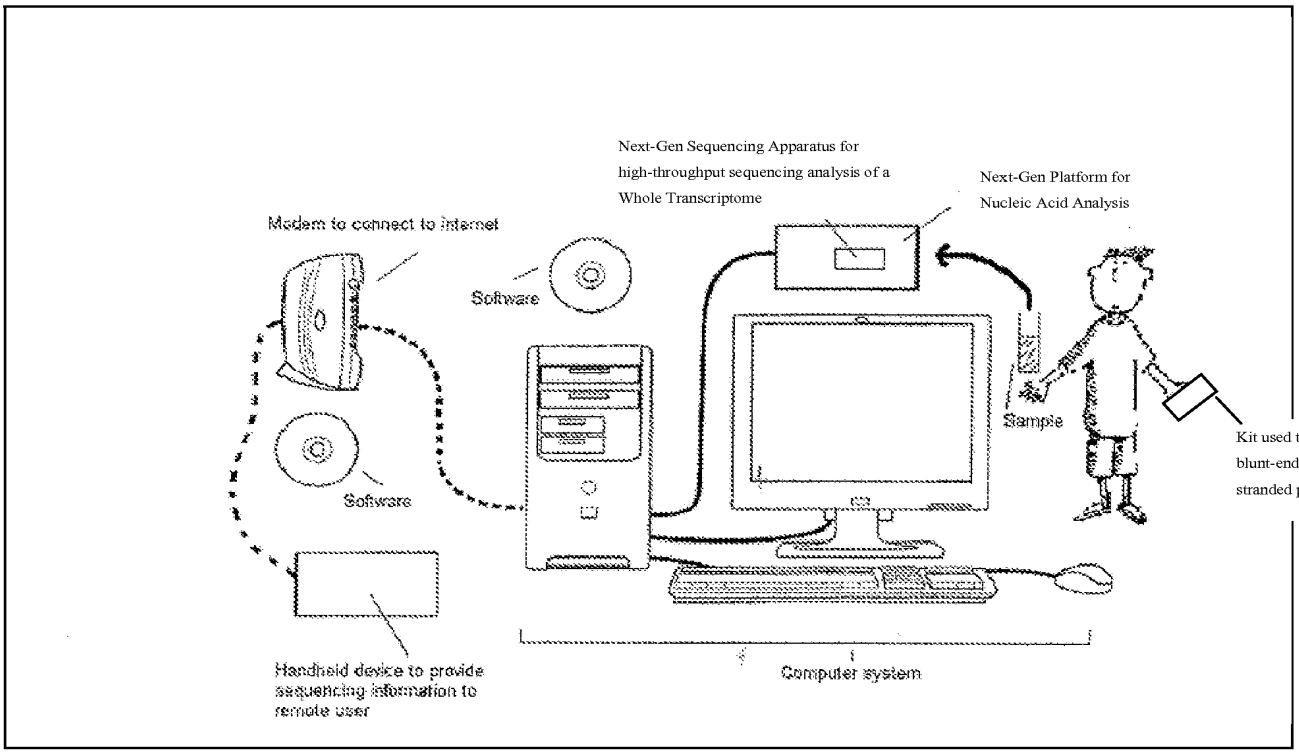

Next-Gen Sequencing Apparatus for high-throughput sequencing analysis of a Whole Transcriptome
Next-Gen Platform for Nucleic Acid Analysis
Modem to connect to Internet
Software
Software
Sample
Kit used to prepare blunt-ended double stranded product
Handheld device to provide sequencing information to remote user
Computer system

Figure 7

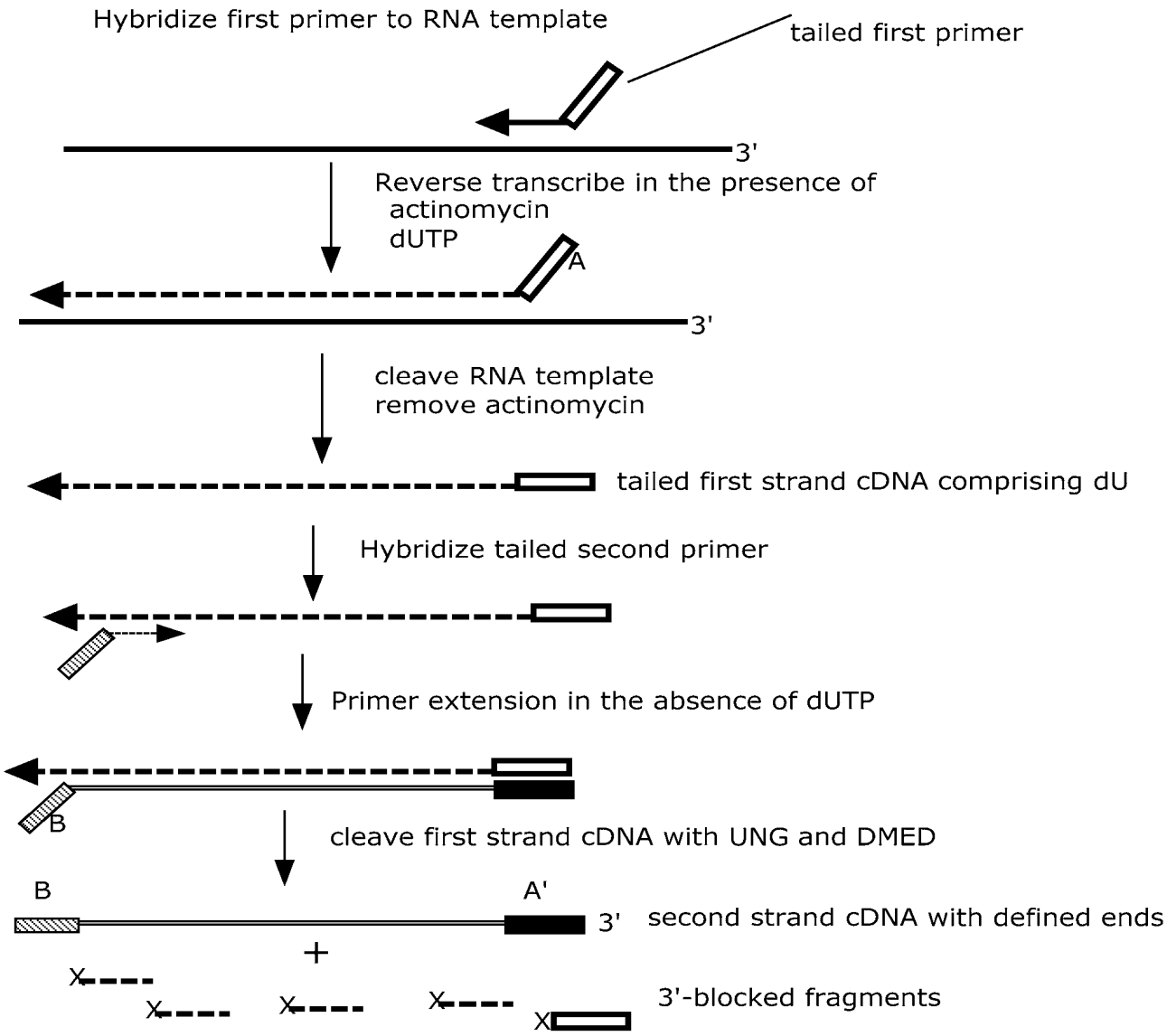
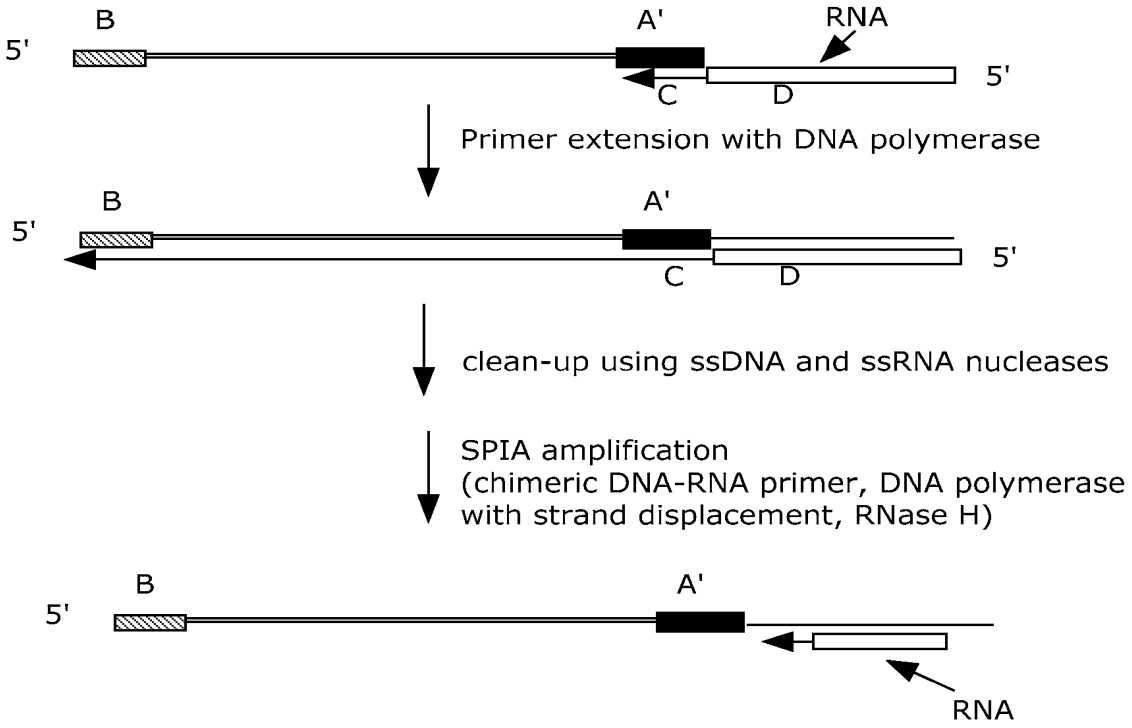

Hybridize first primer to RNA template
tailed first primer
3'
Reverse transcribe in the presence of
actinomycin
dUTP
A
3'
cleave RNA template
remove actinomycin
tailed first strand cDNA comprising dU
Hybridize tailed second primer
Primer extension in the absence of dUTP
B
cleave first strand cDNA with UNG and DMED
B
A'
3'
second strand cDNA with defined ends
3'-blocked fragments

Figure 8


B
A'
RNA
5'
5'
C
D
Primer extension with DNA polymerase
B
A'
5'
5'
C
D
clean-up using ssDNA and ssRNA nucleases
SPIA amplification
(chimeric DNA-RNA primer, DNA polymerase
with strand displacement, RNase H)
B
A'
5'
RNA

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 10/45384

A. CLASSIFICATION OF SUBJECT MATTER
IPC(8) - C12P 19/34 (2010.01)
USPC - 435/91.5, 435/91.51
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: 435/91.5, 435/91.51

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
USPC: 435/91.5, 435/91.51(text search)

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
Electronic data bases: PubWEST (PGPB, USPT, EPAB, JPAB); Google Scholar
Search terms: Nucleic acid template, non-canonical, double stranded product, uracil glycosylase or uracil DNA glycosylase, chimeric primer, nucleic acid free of template nucleic acid, blunt ended product, double stranded adapter nucleic acid

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 2009/0130721 A1 (KURN et al). 21 May 2009 (21.05.2009). Especially para [0009[, [0040], [0051], [0053], [0128-0131],[0159], [0179],[0200], sheet 1 fig 1A.	1, 18-28, 33-35, 37 ---------- 2-17, 29-32, 36, 57-94
Y	KARATA et al. Construction of a circular single-stranded DNA template containing a defined lesion. DNA Repair 4 July 2009 8 No 7 Pages 852?856. Especially abstract, pg 3 para 3,pg 7 fig 1.	2-17, 29-32
Y	US 2004/0005614 A1 (KURN et al.) 8 Janaury 2004 (08.01.2004). Especially para [0011], [0012], [0036], [0071], [0118], [0123], [0248], [0287].	14-17, 57-94
Y	US 6,197,557 B1 (MAKAROV et al.) 6 March 2001 (06.03.2001). Especially col 41 ln 15-17, col 42 ln 49-52	3-9, 36
Y	BUCHMAN et al. Selective RNA amplification: a novel method using dUMP-containing primers and uracil DNA glycosylase. PCR Meth Appl August 1993 Vol 3 No 1 Pages 28-31. Especially pg 29 left col para 2 and 3, center col para 1.	59-62

☐ 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
12 November 2010 (12.11.2010)	03 DEC 2010
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 10/45384

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 1-37, directed to a method comprising: a) providing an input nucleic acid template in a reaction mixture; b)hybridizing one or more oligonucleotide primers to the nucleic acid; c) extending the one or more oligonucleotide primers with a polymerase containing strand displacement activity; and d) cleaving the input nucleic acid template with an agent comprising one or more cleavage agents.

- Please see extra sheet for continuation -

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.:
1-37 and 57-94

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 1020.5636

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 10/45384

Continuation of Box III: Lack of Unity of Invention

Group II: claims 38-45, directed to a method comprising: a) contacting an input template nucleic acid comprising one or more non-canonical nucleotides in a reaction mixture comprising contacting an input template nucleic acid comprising one or more non-canonical nucleotides in a reaction mixture comprising: (i) one or more oligonucleotide primers; (ii) a polymerase comprising strand displacement activity; (iii) an agent capable of cleaving a base portion of a non-canonical nucleotide to generate an abasic site; and (iv) an agent capable of fragmenting a phosphodiester backbone at an abasic site; whereby double stranded DNA fragments are generated.
Group III: claims 46-54, directed to a kit comprising a glycosylase, a polyamine or an AP endonuclease; and one or more double stranded adapter oligonucleotides.
Group IV: claims 55 and 56, directed to a composition or kit comprising (a) an input template nucleic acid comprising one or more non-canonical nucleotides; (b) one or more oligonucleotide primers comprising randomized hybridizing portions; (c) a polymerase comprising strand-displacement activity; and (d) a glycosylase or endonuclease.
Group V: claims 57-94, directed to a method comprising: a) providing an input RNA template in a reaction mixture; b) hybridizing a first primer to the RNA template; c) reverse transcribing the RNA template in the presence of one or more non-canonical nucleotides to generate a first strand cDNA; d) cleaving the input RNA template; e) performing second strand synthesis to generate double stranded cDNA; and f) cleaving the first strand cDNA to generate a single stranded second strand cDNA.

The inventions listed as Groups I - V do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons:

The special technical feature of the Group I claims is a method comprising: a) providing an input nucleic acid template in a reaction mixture; b)hybridizing one or more oligonucleotide primers to the nucleic acid; c) extending the one or more oligonucleotide primers with a polymerase containing strand displacement activity; and d) cleaving the input nucleic acid template with an agent comprising one or more cleavage agents. The special technical feature of the Group II claims is a method comprising: a) contacting an input template nucleic acid comprising one or more non-canonical nucleotides in a reaction mixture comprising contacting an input template nucleic acid comprising one or more non-canonical nucleotides in a reaction mixture comprising: (i) one or more oligonucleotide primers; (ii) a polymerase comprising strand displacement activity; (iii) an agent capable of cleaving a base portion of a non-canonical nucleotide to generate an abasic site; and (iv) an agent capable of fragmenting a phosphodiester backbone at an abasic site; whereby double stranded DNA fragments are generated. The special technical features of the Group III claims is a kit comprising a glycosylase, a polyamine or an AP endonuclease; and one or more double stranded adapter oligonucleotides. The special technical feature of the Group IV claims is a composition or kit comprising (a) an input template nucleic acid comprising one or more non-canonical nucleotides; (b) one or more oligonucleotide primers comprising randomized hybridizing portions; (c) a polymerase comprising strand-displacement activity; and (d) a glycosylase or endonuclease. The special technical feature of the Group V claims is a method comprising: a) providing an input RNA template in a reaction mixture; b) hybridizing a first primer to the RNA template; c) reverse transcribing the RNA template in the presence of one or more non-canonical nucleotides to generate a first strand cDNA; d) cleaving the input RNA template; e) performing second strand synthesis to generate double stranded cDNA; and f) cleaving the first strand cDNA to generate a single stranded second strand cDNA.

The only common technical element shared by the above groups is that they are related to synthesis and cleavage of nucleic acids. Groups I, II and V share the common technical elements of being methods comprising providing an input template nucleic acid, primer oligonucleotides, and a DNA polymerizing enzyme. Groups I and II share the additional common technical element of the polymerase having strand displacement activity. These common technical elements do not represent an improvement over the prior art of US 6,197,557 B1 to Makarov et al. (see abstract, col 46, ln 33-44; col 69, ln 48 - 60; col 64, ln 31-36) Groups II and IV share an additional common technical element, wherein the template nucleic acid comprises a non-canonical nucleotide. This common technical element does not improve upon the prior art of WO 2008/033442 A2 to Frederiksson et al. (see pg. 3, ln 19-21; pg. 9, ln 5-9; pg. 17; ln 21-27). Therefore, the inventions of Groups I-V lack unity of invention under PCT Rule 13 because they do not share a same or corresponding special technical feature.

IDT Exhibit 1020.5637

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
17 March 2011 (17.03.2011)

PCT

(10) International Publication Number
WO 2011/032053 A1

(51) International Patent Classification:
C12P 19/34 (2006.01)

(21) International Application Number:
PCT/US2010/048541

(22) International Filing Date:
10 September 2010 (10.09.2010)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/241,837 11 September 2009 (11.09.2009) US

(71) Applicant *(for all designated States except US)*: **NUGEN TECHNOLOGIES, INC.** [US/US]; 821 Industrial Road, 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). **WANG, Shenglong** [CA/US]; 8248 Saturn Park Drive, San Ramon, CA 94582 (US). **TURNER, Leah, R.** [US/US]; 939 Rosette Court, Sunnyvale, CA 94086 (US). **SEMENTCHENKO, Victor** [US/US]; 2436 Glendenning Avenue, Santa Clara, CA 95050 (US).

(74) Agents: **WILLIS, Michael, A.** 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, 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 2011/032053 A1

(54) **Title**: COMPOSITIONS AND METHODS FOR WHOLE TRANSCRIPTOME ANALYSIS

(57) **Abstract**: The present invention provides methods and compositions, including kits, for the generation of cDNA from mRNA with reduced ribosomal RNA representation.

IDT Exhibit 1020.5638

COMPOSITIONS AND METHODS FOR WHOLE TRANSCRIPTOME ANALYSIS

CROSS-REFERENCE

[0001] This application claims the benefit of U.S. Provisional Application No. 61/241837, filed September 11, 2009, which application is incorporated herein by reference.

BACKGROUND OF THE INVENTION

[0002] High throughput next generation genome-wide transcriptome analysis methods have been developed in recent years, enabling highly parallel analysis of all transcripts in a given sample, employing a variety of platforms. These platforms include high and low density microarrays and massively parallel sequencing. Analysis of the transcriptome from total RNA samples usually entails reduction or near complete elimination of ribosomal RNA (rRNA) species from the samples. rRNA species represent the majority of RNA in the samples, when compared to all other transcripts. The presence of rRNA may be detrimental to many downstream applications, especially in amplified RNA samples.

[0003] Various methods have been applied for rRNA removal. For example Invitrogen has introduced the RiboMinus procedure for removal of rRNA by binding the rRNA to a solid surface by hybridization followed by separation of the solid surface from the samples. Affymetrix Inc. has introduced a procedure for removing rRNA by enzymatic digestions. Others have analyzed only polyA mRNA by purification of the polyA tailed mRNA. These methods, however, are collectively highly inefficient, time consuming and/or expensive and can result in degradation and loss of the mRNA transcripts in the sample. Furthermore, these methods are difficult to perform when analyzing gene expression profiles in minute biological samples.

[0004] It is desirable to obtain either amplified or non amplified cDNA from total mRNA in which the amount of rRNA is reduced relative to the original total RNA in the samples, so as to increase the content of non ribosomal RNA transcripts, representing the expressed genes.

SUMMARY OF THE INVENTION

[0005] The present invention provides novel methods and compositions for designing, selecting for, screening, and using specific combinations of reagents that allow for selective cDNA synthesis from an RNA sample, useful for downstream applications such as sequencing and array analyses. Specifically, an important aspect of this invention is the compositions and methods that allow for the reduction or substantial elimination of ribosomal RNA (rRNA) representation in a cDNA sample prepared from RNA.

[0006] In one aspect, the invention provides a method for differentially reducing the reverse transcription of ribosomal RNA (rRNA) from an RNA sample. In some embodiments, the method comprises: (a) providing one or more primers of known sequence; (b) combining the one or more primers with a reverse transcriptase (RT); and (c) reverse transcribing the RNA sample, thereby producing reverse transcribed products, wherein the reverse transcription of rRNA is reduced. In some embodiments, the reverse transcription of rRNA is reduced by at least about 10 fold. In some embodiments, the method further

IDT Exhibit 1020.5639

comprises step (d) amplifying said reverse transcribed products, thereby producing amplified products. In some embodiments, the method further comprises the step of sequencing the reverse transcribed products. In some embodiments, the reverse transcribed products are amplified prior to sequencing. In some embodiments, the method does not include the step of purification of one or more RNA sequences from the sample.

[0007] In one aspect, the invention provides a method of identifying a primer sequence that differentially reduces the reverse transcription of ribosomal RNA (rRNA) from an RNA sample. In some embodiments, the method comprises (a) providing one or more primers of known sequence; (b) combining the one or more primers with a reverse transcriptase (RT) to reverse transcribe the RNA sample, thereby producing reverse transcribed products; (c) optionally generating double-stranded cDNA products from said reverse transcribed products; (d) optionally amplifying said double-stranded cDNA products; and (e) analyzing the reverse transcribed products to determine if the reverse transcription of rRNA is reduced by the primer sequence.

[0008] In one aspect, the invention provides a method of identifying a reverse transcriptase enzyme (RT) that differentially reduces the reverse transcription of rRNA from an RNA sample. In some embodiments, the method comprises (a) providing one or more primers of known sequence; (b) using the primer with an RT to reverse transcribe the RNA sample, thereby producing reverse transcribed products; and (c) optionally generating double-stranded cDNA products from said reverse transcribed products; (d) optionally amplifying said double-stranded cDNA products; and (e) analyzing the reverse transcribed products to determine if the reverse transcription of rRNA is reduced by the RT.

[0009] In one aspect, the invention provides for a method for whole transcriptome sequencing comprising providing a RNA sample, reverse transcribing the sample, amplifying the RNA sample using one or more primers to produce amplified products, and performing sequencing on the products. In some embodiments, the method does not include a) pre-treatment of the RNA sample, b) prior selection of the RNA sample, c) selection of sequences from the RNA sample comprising a polyA sequence, d) prior removal of ribosomes from the RNA sample, or d) substantial loss of the RNA in the sample. In some embodiments, the method does not include the step of purification of one or more RNA sequences from the sample.

[0010] In one aspect, the invention provides for a collection of chimeric primers, each comprising RNA and DNA, and a RT enzyme, wherein the RT is used in combination with the chimeric primers to reverse transcribe a whole transcriptome. In such an embodiment, no more than 5%, 10%, 15%, 20%, 25%, 30%, 40%, 50%, 60%, 70%, 75%, or 80% of the resulting amplicons are rRNA sequences.

[0011] In some embodiments, the one or more primers are tailed primers comprising a 3' portion hybridizable to one or more target RNAs in the RNA sample and a 5' portion that is not hybridizable to the one or more target RNAs in the RNA sample. In some embodiments, the one or more primers are chimeric primers comprising a DNA portion and an RNA portion. In some embodiments, the chimeric primers comprise a 3'-DNA portion hybridizable to one or more target RNAs in the RNA sample and a 5'-RNA portion that is not hybridizable to the one or more target RNAs in the RNA sample. The

IDT Exhibit 1020.5640

hybridizable 3' portion of the primers can comprise a randomly generated sequence. The hybridizable 3' portion of the primer can comprise 5-15 nucleotides, for example 5, 6, 7 ,8, 9, 10, 11, 12, 13, 14, or 15 nucleotides. The non-hybridizable 5' portion of the primers can comprise a promoter-specific sequence. In some embodiments, the primer is one of a plurality of primers of known sequence, each primer comprising a 3' portion hybridizable to the sample RNA and a 5' portion that is not hybridizable to the sample RNA, wherein each of the non-hybridizable 5' portions of the primers comprise the same sequence.

[0012] In some embodiments, the RT is selected from the group consisting of: a Moloney murine leukemia virus RT, a human immunodeficiency virus RT, a rous sarcoma virus RT, an avian myeloblastosis virus RT, a rous associated virus RT, a myeloblastosis associated virus RT, an avian sarcoma-leukosis virus RT, an RT lacking RNase H activity, modified RTs derived therefrom, and combinations thereof.

[0013] In some embodiments, the reverse transcription of rRNA is reduced by at least about 20%, at least about 30%, at least about 40%, or at least about 50%. In some embodiments, the reverse transcription of rRNA is reduced by at least about 2-fold. In some embodiments, the sequence that reduces the reverse transcription of ribosomal RNA is identified based on varying the non-hybridizable 5' of the one or more primers. In some embodiments, the primer sequence that reduces the reverse transcription of ribosomal RNA is identified based on varying the RT used in step (c).

[0014] In some embodiments, the RNA sample comprises mRNA and rRNA. In some embodiments, the RNA is selected from the group consisting of total RNA, mitochondrial RNA, chloroplast RNA, DNA-RNA hybrids, viral RNA, cell free RNA, and mixtures thereof.

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 the effect of first strand tailed primers on non amplified cDNA generation by RTa.

[0018] Figure 2 depicts the effect of first strand tailed primers on rRNA content in non-amplified cDNA generation by RTb.

[0019] Figure 3 depicts that mitochondrial rRNA and transcripts are affected to a lesser extent by choice of first strand tailed primer: RTa.

IDT Exhibit 1020.5641

[0020] Figure 4 depicts that mitochondrial rRNA and transcripts are affected to a lesser extent by choice of first strand tailed primer: RTb.

[0021] Figure 5 depicts non amplified cDNA, where delta Ct represents Ct for RTa minus Ct for RTb.

[0022] Figure 6 depicts non amplified cDNA, where delta CT represents Ct for RTa minus Ct for RTb.

[0023] Figure 7 depicts non amplified cDNA generated by RTa: rRNA avg. Ct..

[0024] Figure 8 depicts non-amplified cDNA generated by RTb: rRNA avg. Ct.

[0025] Figure 9 depicts non-amplified cDNA generated by RTa: mt rRNA avg. Ct.

[0026] Figure 10 depicts non-amplified cDNA generated by RTb: mt rRNA avg. Ct.

[0027] Figure 11 depicts quantification of rRNA in amplified cDNA.

[0028] Figure 12 depicts quantification of various mRNAs in amplified cDNA (average rRNA Ct is given as a reference).

[0029] Figure 13 depicts quantification of mitochondrial RNA in amplified cDNA.

[0030] Figure 14 depicts quantification of mitochondrial transcripts in amplified cDNA.

[0031] Figure 15 depicts a comparison of cDNA prepared with RTa or RTb and SPIA-amplified using WGA-Ovation.

[0032] Figure 16 depicts a comparison of cDNA generated by RTa or RTb and SPIA-amplified using Pico-Ovation SPIA.

[0033] Figure 17 depicts a comparison of SPIA amplification reagents: cDNA generated with RTa and amplified with Pico- or WGA-Ovation SPIA.

[0034] Figure 18 depicts a comparison of SPIA amplification reagents: cDNA generated with RTb and amplified with Pico- or WGA-Ovation SPIA.

[0035] Figure 19 depicts differences in rRNA representation by comparison of depth of coverage.

[0036] Figure 20 depicts differences in rRNA representation by comparison of depth of coverage.

DETAILED DESCRIPTION OF THE INVENTION

General

[0037] In one embodiment, the present invention provides methods and compositions for selectively amplifying a target population of nucleic acid molecules (e.g. all mRNA molecules expressed in a cell except for the most highly expressed mRNA species). In another embodiment, the present invention provides methods and compositions for screening for and/or selecting suitable primers and reverse transcriptases (RTs) which are capable of selectively reducing the cDNA generation of rRNA, while maintaining the representation of non-rRNA transcripts. The methods further enable generation of cDNA which can be further amplified by a variety of amplification methods, for example, the single primer isothermal linear amplification method (SPIA).

Primers

[0038] A "primer," as used herein, refers to a nucleotide sequence, generally with a free 3'-OH group, that hybridizes with a template sequence (such as one or more target RNAs, or a primer extension

IDT Exhibit 1020.5642

product) and is capable of promoting polymerization of a polynucleotide complementary to the template. A "primer" can be, for example, an oligonucleotide. It can also be, for example, a sequence of the template (such as a primer extension product or a fragment of the template created following RNase cleavage of a template-DNA complex) that is hybridized to a sequence in the template itself (for example, as a hairpin loop), and that is capable of promoting nucleotide polymerization. Thus, a primer can be an exogenous (e.g., added) primer or an endogenous (e.g., template fragment) primer. A primer may contain a non-hybridizing sequence that constitutes a tail on the primer. A primer may still be hybridizing even though its sequences are not completely complementary to the target.

[0039] The primers of the invention are usually oligonucleotide primers. A primer is generally an oligonucleotide that is employed in an extension by a polymerase along a polynucleotide template such as in, for example, PCR or SPIA. The oligonucleotide primer is often a synthetic polynucleotide that is single stranded, containing a sequence at its 3'-end that is capable of hybridizing with a sequence of the target polynucleotide. Normally, the 3' region of the primer that hybridizes with the target nucleic acid has at least 80%, preferably 90%, more preferably 95%, most preferably 100%, complementarity to a sequence or primer binding site. "Complementary", as used herein, refers to complementarity to all or only to a portion of a sequence. The number of nucleotides in the hybridizable sequence of a specific oligonucleotide primer should be such that stringency conditions used to hybridize the oligonucleotide primer will prevent excessive random non-specific hybridization. Usually, the number of nucleotides in the hybridizing portion of the oligonucleotide primer will be at least as great as the defined sequence on the target polynucleotide that the oligonucleotide primer hybridizes to, namely, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 11, at least 12, at least 13, at least 14, at least 15, at least about 20, and generally from about 6 to about 10 or 6 to about 12 of 12 to about 200 nucleotides, usually about 20 to about 50 nucleotides. In general, the target polynucleotide is larger than the oligonucleotide primer or primers as described previously. In some cases, the hybridizable sequence of an oligonucleotide primer is a random sequence. Oligonucleotide primers comprising random sequences may be referred to as random primers, as described herein. In other cases, an oligonucleotide primer such as a first primer or a second primer comprises a set of primers such as for example a set of first primers or a set of second primers. In some cases, the set of first or second primers may comprise a mixture of primers designed to hybridize to a plurality (e.g. 2, 3, 4, about 6, 8, 10, 20, 40, 80, 100, 125, 150, 200, 250, 300, 400, 500, 600, 800, 1000, 1500, 2000, 2500, 3000, 4000, 5000, 6000, 7000, 8000, 10,000, 20,000, 25,000 or more) of target sequences. In some cases, the plurality of target sequences may comprise a group of related sequences, random sequences, a whole transcriptome or fraction (e.g. substantial fraction) thereof, or any group of sequences such as mRNA.

[0040] Tailed primers are employed in certain embodiments of the invention. In general, a tailed primer comprises a 3' portion that is hybridizable to one or more target polynucleotides, such as one or more target RNAs in an RNA sample, and a 5' portion that is not hybridizable to the one or more target polynucleotides. In general, the non-hybridizable 5' portion does not hybridize to the one or more target polynucleotides under conditions in which the hybridizable 3' portion of the tailed primer hybridizes to

IDT Exhibit 1020.5643

the one or more target polynucleotides. In some embodiments, the non-hybridizable 5' portion comprises a promoter-specific sequence. Generally, a promoter-specific sequence comprises a single-stranded DNA sequence region which, in double-stranded form is capable of mediating RNA transcription. Examples of promoter-specific sequences are known in the art, and include, without limitation, T7, T3, or SP6 RNA polymerase promoter sequences. When the tailed primer is extended with a DNA polymerase, a primer extension product with a 5' portion comprising a defined sequence can be created. This primer extension product can then have a second primer anneal to it, which can be extended with a DNA polymerase to create a double stranded product comprising a defined sequence at one end. In some embodiments, where the non-hybridizable 5' portion of one or more tailed primers comprises a promoter-specific sequence, creation of a double-stranded product comprising a defined sequence at one end generates a double-stranded promoter sequence that is capable of mediating RNA transcription. In some embodiments, a double-stranded promoter sequence is generated by hybridizing to the promoter-specific sequence an oligonucleotide comprising a sequence complementary to the promoter-specific sequence. In some embodiments, formation of a double-stranded promoter is followed by the generation of single-stranded RNA by RNA transcription of sequence downstream of the double-stranded promoter, generally in a reaction mixture comprising all necessary components, including but not limited to ribonucleoside triphosphates (rNTPs) and a DNA-dependent RNA polymerase. Tailed primers can comprise DNA, RNA, or both DNA and RNA. In some embodiments, the tailed primer consists of DNA.

[0041] Composite primers are employed in certain embodiments of the invention. Composite primers are primers that are composed of RNA and DNA portions. In some aspects, the composite primer is a tailed composite primer comprising, for example, a 3'-DNA portion and a 5'-RNA portion. In the tailed composite primer, a 3'-portion, all or a portion of which comprises DNA, is complementary to a polynucleotide; and a 5'-portion, all or a portion of which comprises RNA, is not complementary to the polynucleotide and does not hybridize to the polynucleotide under conditions in which the 3'-portion of the tailed composite primer hybridizes to the polynucleotide target. When the tailed composite primer is extended with a DNA polymerase, a primer extension product with a 5'-RNA portion comprising a defined sequence can be created. This primer extension product can then have a second primer anneal to it, which can be extended with a DNA polymerase to create a double stranded product with an RNA/DNA heteroduplex comprising a defined sequence at one end. The RNA portion can be selectively cleaved from the partial heteroduplex to create a double-stranded DNA with a 3'-single-stranded overhang which can be useful for various aspects of the present invention including allowing for isothermal amplification using a composite amplification primer.

[0042] In other aspects, the composite primer is an amplification composite primer (interchangeably called composite amplification primer). In the amplification composite primer, both the RNA and the DNA portions are generally complementary and hybridize to a sequence in the polynucleotide 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.

IDT Exhibit 1020.5644

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 sequence of the polynucleotide. Composite primers are described in more detail below.

[0043] A "random primer," as used herein, is a primer that generally comprises a sequence that is designed not necessarily based on a particular or specific sequence in a sample, but rather is based on a statistical expectation (or an empirical observation) that the sequence of the random primer is hybridizable (under a given set of conditions) to one or more sequences in the sample. A random primer will generally be an oligonucleotide or a population of oligonucleotides comprising a random sequence(s) in which the nucleotides at a given position on the oligonucleotide can be any of the four nucleotides, or any of a selected group of the four nucleotides (for example only three of the four nucleotides, or only two of the four nucleotides). In some cases all of the positions of the oligonucleotide or population of oligonucleotides can be any of two or more nucleotides. In other cases, only a portion of the oligonucleotide, for instance a particular region, will comprise positions which can be any of two or more bases. In some cases, the portion of the oligonucleotide which comprises positions which can be any of two or more bases is about 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, or about 15-20 nucleotides in length. In some cases, a random primer may comprise a tailed primer having a 3'-region that comprises a random sequence and a 5'-region that is a non-hybridizing sequence that comprises a specific, non-random sequence. The 3'-region may also comprise a random sequence in combination with a region that comprises poly-T sequences. The sequence of a random primer (or its complement) may or may not be naturally-occurring, or may or may not be present in a pool of sequences in a sample of interest. The amplification of a plurality of RNA species in a single reaction mixture can employ, but not necessarily employ, a multiplicity, preferably a large multiplicity, of random primers. As is well understood in the art, a "random primer" can also refer to a primer that is a member of a population of primers (a plurality of random primers) which collectively are designed to hybridize to a desired and/or a significant number of target sequences. A random primer may hybridize at a plurality of sites on a nucleic acid sequence. The use of random primers provides a method for generating primer extension products complementary to a target polynucleotide which does not require prior knowledge of the exact sequence of the target. In some embodiments one portion of a primer is random, and another portion of the primer comprises a defined sequence. For example, in some embodiments, a 3'-portion of the primer will comprise a random sequence, while the 5'-portion of the primer comprises a defined sequence. In some embodiments a 3'-random portion of the primer will comprise DNA, and a 5'- defined portion of the primer will comprise RNA, in other embodiments, both the 3' and 5'-portions will comprise DNA. In some embodiments, the 5'-portion will contain a defined sequence and the 3'-portion will comprise a poly-dT sequence that is

IDT Exhibit 1020.5645

hybridizable to a multiplicity of RNAs in a sample (such as all mRNA). In some embodiments, a "random primer," or primer comprising a randomly generated sequence, comprises a collection of primers comprising one or more nucleotides selected at random from two or more different nucleotides, such that all possible sequence combinations of the nucleotides selected at random may be represented in the collection. In some embodiments, generation of one or more random primers does not include a step of excluding or selecting certain sequences or nucleotide combinations from the possible sequence combinations in the random portion of the one or more random primers.

[0044] In one embodiment, the preferred primers of the invention are tailed primers. In this embodiment, the 5'-tail can comprise RNA and is non hybridizable to the RNA in the sample. The 3'-end of the first primer(s) can be hybridizable to the RNA in the sample, comprise DNA and comprise a random sequence, enabling hybridization across the whole transcriptome. The first primer may also comprise a mixture of primers. The mixture of first primers may also include a first primer comprising a 3'-DNA sequence hybridizable to the 3'-poly A tail of mRNA, in addition to the first primers comprising a random sequence at the 3'-ends.

RNA-Dependent DNA Polymerases

[0045] RNA-dependent DNA polymerases for use in the methods and compositions of the invention are capable of effecting extension of a primer according to the methods of the invention. Accordingly, a preferred RNA-dependent DNA polymerase is one that is capable of extending a nucleic acid primer along a nucleic acid template that is comprised at least predominantly of ribonucleotides. Suitable RNA-dependent DNA polymerases for use in the methods and compositions of the invention include reverse transcriptases (RTs). RTs are well known in the art. Examples of RTs include, but are not limited to, Moloney murine leukemia virus (M- MLV) reverse transcriptase, human immunodeficiency virus (HIV) reverse transcriptase, rous sarcoma virus (RSV) reverse transcriptase, avian myeloblastosis virus (AMV) reverse transcriptase, rous associated virus (RAV) reverse transcriptase, and myeloblastosis associated virus (MAV) reverse transcriptase or other avian sarcoma-leukosis virus (ASLV) reverse transcriptases, and modified RTs derived therefrom. See e.g. US7056716. Many reverse transcriptases, such as those from avian myeoloblastosis virus (AMV-RT), and Moloney murine leukemia virus (MMLV-RT) comprise more than one activity (for example, polymerase activity and ribonuclease activity) and can function in the formation of the double stranded cDNA molecules. However, in some instances, it is preferable to employ a RT which lacks or has substantially reduced RNase H activity. RTs devoid of RNase H activity are known in the art, including those comprising a mutation of the wild type reverse transcriptase where the mutation eliminates the RNase H activity. Examples of RTs having reduced RNase H activity are described in US20100203597. In these cases, the addition of an RNase H from other sources, such as that isolated from E. coli, can be employed for the formation of the double stranded cDNA. Combinations of RTs are also contemplated, including combinations of different non-mutant RTs, combinations of different mutant RTs, and combinations of one or more non-mutant RT with one or more mutant RT.

IDT Exhibit 1020.5646

Methods of Amplification

[0046] Methods of nucleic acid amplification are well known in the art.

[0047] In some aspects of the invention, the amplification method that is used to amplify the marked DNA is a single primer isothermal amplification (SPIA) using a complex comprising a RNA/DNA partial heteroduplex as a template. In this method a complex comprising a RNA/DNA partial heteroduplex is a substrate for further amplification as follows: an enzyme (such as RNase H), which cleaves RNA sequence from an RNA/DNA heteroduplex, cleaves the RNA sequence (A) from the partial heteroduplex, leaving a partially double stranded polynucleotide complex comprising an exposed 3' single-stranded DNA sequence. The 3' single-stranded sequence (formed by cleavage of RNA in the complex comprising a RNA/DNA partial heteroduplex) is generally the complement of a composite amplification primer, and thus forms a specific binding site for the composite amplification primer. This composite amplification primer is the third primer used in this invention and also comprises a unique 5'-sequence. Extension of a bound composite amplification primer by a DNA-dependent DNA polymerase with strand displacement activity produces a primer extension product, which displaces the previously bound cleaved primer extension product, whereby single stranded DNA product accumulates. The single stranded DNA product is a copy of the complement of the target RNA (or "antisense" DNA). The cycle repeats with the removal of the unique sequence on the 5'-tail of the primer extension product, exposing the 3'-end of the second primer extension product for another cycle of amplification. This linear amplification is referred to as "SPIA" (for Single Primer Linear Isothermal Amplification), and is described in Kurn et al., U.S. Pat. No. 6,251,639 and 6,692,918.

[0048] Amplification using a complex comprising a RNA/DNA partial heteroduplex as a template for further amplification by SPIA generally occurs 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.

[0049] In some embodiments, the methods of the invention result in amplification of a multiplicity, a large multiplicity, or a very large multiplicity of target RNA. In some embodiments, essentially all of the RNA present in the initial sample (e.g., all of the mRNA) substantially free of rRNA is amplified. In other embodiments, at least 1, at least 5, at least 10, at least 20, at least 50, at least 100, at least 200, at least 300, up to at least 10,000 or more distinct sequences (such as other sub-segments of a RNA) are amplified (also substantially free of rRNA), as assessed, e.g., by analysis of marker sequences known to be present in the template sample under analysis, using methods known in the art. Target RNA sequences that are amplified may be present on the same RNA strand or on different RNA strands. It will be understood by

IDT Exhibit 1020.5647

those of skill in the art that the global amplification methods of the invention are suitable for amplification of any pool or subset of RNA.

[0050] In some embodiments, the methods of the invention are used to globally amplify a single stranded RNA target or the double stranded DNA that is initially produced from the RNA target using methods described herein. In these cases, the amplification product will generally be a copy of either the target RNA (sense copy) or of the complement to the target RNA (antisense copy). Whether the sense or antisense copy is produced will depend on the method, as will be understood by one of ordinary skill in the art. In some embodiments, the amplification product of different senses can be annealed to form a double-stranded (or partially double-stranded) complex. In other embodiments, they can be prevented from annealing (or subsequently denatured) to produce a mixture of single stranded amplification products. The amplified products may be of differing lengths.

[0051] As illustrated in these embodiments, 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 may 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).

[0052] For simplicity, the isothermal amplification methods of the invention are described as two distinct steps or phases, above. It is understood that the two phases may occur simultaneously in some embodiments (for example, if the enzyme that cleaves RNA from RNA/DNA heteroduplex is included in the first reaction mixture).

[0053] Although generally only one composite amplification primer is described above, it is further understood that the amplification methods may 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.

Methods of Screening for Reagents for Selective cDNA Synthesis

[0054] The compositions and methods provided herein are useful for screening for a combination of reagents that will yield a desired result.

[0055] By way of an example, an optimal RT can be screened for to achieve a reduction in the reverse transcription and/or subsequent amplification of rRNA by at least about 5%, at least about 10%, at least about 20%, at least about 25%, at least about 30%, at least about 40%, at least about 50%, at least about 60%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 99%, or more. By way of further example, an optimal RT can be screened for to achieve a reduction in the reverse transcription and/or subsequent amplification of rRNA by at least about 2-fold, 3-fold, 4-fold, 5-fold, 6-fold, 7-fold, 8-fold, 10-fold, 100-fold, 200-fold, 300-fold, 400-fold, 500-fold, 600-fold, 700-fold, 800-fold, 900-fold, 1000-fold, 10,000-fold, or more.

IDT Exhibit 1020.5648

[0056] By way of another example, an optimal reverse transcription primer sequence, such as a tailed primer, a chimeric primer comprising RNA and DNA sequences, or a tailed chimeric primer, can be screened for to achieve a reduction in the reverse transcription and/or subsequent amplification of rRNA by at least about 5%, at least about 10%, at least about 20%, at least about 25%, at least about 30%, at least about 40%, at least about 50%, at least about 60%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 99%, or more. By way of further example, an optimal reverse transcription primer sequence can be screened for to achieve a reduction in the reverse transcription and/or subsequent amplification of rRNA by at least about 2-fold, 3-fold, 4-fold, 5-fold, 6-fold, 7-fold, 8-fold, 10-fold, 100-fold, 200-fold, 300-fold, 400-fold, 500-fold, 600-fold, 700-fold, 800-fold, 900-fold, 1000-fold, 10,000-fold, or more.

[0057] By way of yet another example, an optimal combination of reverse transcription primer and RT can be screened for to achieve a reduction in the reverse transcription and/or subsequent amplification of rRNA by at least about 5%, at least about 10%, at least about 20%, at least about 25%, at least about 30%, at least about 40%, at least about 50%, at least about 60%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, or at least about 99%. In some embodiments, reverse transcription and/or subsequent amplification of rRNA is reduced by at least about 2-fold, 3-fold, 4-fold, 5-fold, 6-fold, 7-fold, 8-fold, 10-fold, 100-fold, 200-fold, 300-fold, 400-fold, 500-fold, 600-fold, 700-fold, 800-fold, 900-fold, 1000-fold, 10,000-fold, or more.

[0058] In some embodiments, a reduction in the reverse transcription and/or subsequent amplification of rRNA is with respect to an amount, fraction, or representation of rRNA in the starting sample prior to reverse transcription and/or amplification. In some embodiments, a reduction in the reverse transcription and/or subsequent amplification of rRNA is with respect to an amount, fraction, or representation of rRNA in a control sample, wherein RNA is reverse transcribed or subsequently amplified using one or more non-selective agents, such as a pool of random hexamer, heptamter, octomer, 9-mer, or N-mer primers (e.g. with each position of every primer in the pool selected at random, preferably representing all combinations of sequences of length N, the length of the random primer).

[0059] In some embodiments, the method does not include a step for the purification of one or more RNA sequences from the sample, such as the removal of a non-target RNA, or the separation of target RNA. Non-target RNA can include rRNA.

Downstream Applications for Whole Transcriptome Analysis

[0060] An important aspect of the invention is that the methods and compositions disclosed herein can be efficiently and cost-effectively utilized for downstream analyses, such as next-generation sequencing or hybridization platforms, with minimal loss of biological material of interest. In one embodiment, a purpose of the embodiments herein is to reduce rRNA content from total RNA, i.e. from the whole transcriptome.

IDT Exhibit 1020.5649

[0061] The methods of the invention are useful, for example, for efficient sequencing of a polynucleotide sequence of interest. Specifically the methods of the invention are useful for massively parallel sequencing of a product substantially free of rRNA sequences.

[0062] In one embodiment, the invention provides for a method for whole transcriptome sequencing comprising providing a RNA sample, providing one or more primers of known sequence, combining the one or more primers with a reverse transcriptase, reverse transcribing the sample, amplifying the RNA sample using one or more primers to produce amplified products, and performing sequencing on the products. In this embodiment, the method preferably does not include one or more of: (a) pre-treatment of RNA sample, (b) prior selection of the RNA sample, (c) selection of sequences from the RNA sample comprising a polyA sequence, (d) prior removal of ribosomes from the RNA sample, (e) substantial loss of the RNA in the sample, (f) purification of target RNA from the sample, and (g) purification of non-target RNA from the sample. In some embodiments, sequencing is performed on single- or double-stranded cDNA generated by the reverse transcription of an RNA sample without amplifying the RNA sample. In some embodiments, the starting amount of RNA is .01ng to 100mg. The primers used for reverse transcription and/or amplification can be tailed primers, chimeric primers, or tailed and chimeric primers.

[0063] In one embodiment, a collection of tailed primers, and a RT enzyme is provided, wherein the RT is used in combination with the tailed primers to reverse transcribe a whole transcriptome. In one embodiment, a collection of chimeric primers, each comprising RNA and DNA, and a RT enzyme is provided, wherein the RT is used in combination with the chimeric primers to reverse transcribe a whole transcriptome. In some embodiment, no more than 5%, 10%, 15%, 20%, 25%, 30%, 40%, 50%, 60%, 70%, 75%, 80% of the resulting products are rRNA sequences.

[0064] Known methods for sequencing include, for example, those described in:, Sanger, F. et al. , Proc. Natl. Acad. Sci. U. S. A. 75,5463-5467 (1977); Maxam, A. M. & Gilbert, W. Proc Natl Acad Sci USA 74,560-564 (1977); Ronaghi, M. et al. , Science 281, 363,365 (1998); Lysov, 1. et al., Dokl Akad Nauk SSSR 303, 1508-1511 (1988); Bains W. & Smith G. C. J. Theor Biol 135,303-307 (1988); Drnanac, R. et al. , Genomics 4,114-128 (1989); Khrapko, K. R. et al. , FEBS Lett 256.118-122 (1989); Pevzner P. A. J Biomol Struct Dyn 7, 63-73 (1989); and Southern, E. M. et al., Genomics 13, 1008-1017 (1992). Pyrophosphate-based sequencing reaction as described, e. g., in US patent 6,274, 320,6258, 568 and 6,210, 891), may also be used. In some cases, the methods above require that the nucleic acid attached to the solid surface be single stranded. In such cases, the unbound strand may be melted away using any number of commonly known methods such as addition of NaOH, application of low ionic (e. g., salt) strength solution, enzymatic degradation or displacement of the second strand, or heat processing. Where the solid surface comprises a plurality of beads, following this strand removal step, the beads can be pelleted and the supernatant discarded. The beads can then be resuspended in a buffer, and a sequencing primer or other non-amplification primer can be added. The primer is annealed to the single stranded amplification product. This can be accomplished by using an appropriate annealing buffer and temperature conditions, e. g., as according to standard procedures in the art.

IDT Exhibit 1020.5650

[0065] The methods of the invention are useful, for example, for sequencing of an RNA sequence of interest. The sequencing process can be carried out by processing and amplifying a target RNA containing the sequence of interest by any of the methods described herein. Addition of nucleotides during primer extension can be analyzed using methods known in the art, for example, incorporation of a terminator nucleotide, sequencing by synthesis (e.g. pyrosequencing), or sequencing by ligation.

[0066] In embodiments wherein the end product is in the form of DNA primer extension products, in addition to the nucleotides, such as natural deoxyribonucleotide triphosphates (dNTPs), that are used in the amplification methods, appropriate nucleotide triphosphate analogs, which may be labeled or unlabeled, that upon incorporation into a primer extension product effect termination of primer extension, may be added to the reaction mixture. Preferably, the dNTP analogs are added after a sufficient amount of reaction time has elapsed since the initiation of the amplification reaction such that a desired amount of second primer extension product or fragment extension product has been generated. Said amount of the time can be determined empirically by one skilled in the art.

[0067] Suitable dNTP analogs include those commonly used in other sequencing methods and are well known in the art. Examples of dNTP analogs include dideoxyribonucleotides. Examples of rNTP analogs (such as RNA polymerase terminators) include 3'-dNTP. Sasaki et al., Biochemistry (1998) 95:3455-3460. These analogs may be labeled, for example, with fluorochromes or radioisotopes. The labels may also be labels which are suitable for mass spectroscopy. The label may also be a small molecule which is a member of a specific binding pair, and can be detected following binding of the other member of the specific binding pair, such as biotin and streptavidin, respectively, with the last member of the binding pair conjugated to an enzyme that catalyzes the generation of a detectable signal that could be detected by methods such as colorimetry, fluorometry or chemiluminescence. All of the above examples are well known in the art. These are incorporated into the primer extension product or RNA transcripts by the polymerase and serve to stop further extension along a template sequence. The resulting truncated polymerization products are labeled. The accumulated truncated products vary in length, according to the site of incorporation of each of the analogs, which represent the various sequence locations of a complementary nucleotide on the template sequence.

[0068] Analysis of the reaction products for elucidation of sequence information can be carried out using any of various methods known in the art. Such methods include gel electrophoresis and detection of the labeled bands using appropriate scanner, sequencing gel electrophoresis and detection of the radiolabeled band directly by phosphorescence, capillary electrophoresis adapted with a detector specific for the labels used in the reaction, and the like. The label can also be a ligand for a binding protein which is used for detection of the label in combination with an enzyme conjugated to the binding protein, such as biotin-labeled chain terminator and streptavidin conjugated to an enzyme. The label is detected by the enzymatic activity of the enzyme, which generates a detectable signal. As with other sequencing methods known in the art, the sequencing reactions for the various nucleotide types (A, C, G, T or U) are carried out either in a single reaction vessel, or in separate reaction vessels (each representing one of the various nucleotide types). The choice of method to be used is dependent on practical considerations readily apparent to one

IDT Exhibit 1020.5651

skilled in the art, such as the nucleotide tri phosphate analogs and/or label used. Thus, for example, when each of the analogs is differentially labeled, the sequencing reaction can be carried out in a single vessel. The considerations for choice of reagent and reaction conditions for optimal performance of sequencing analysis according to the methods of the invention are similar to those for other previously described sequencing methods. The reagent and reaction conditions should be as described above for the nucleic acid amplification methods of the invention.

[0069] 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.

[0070] 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).

[0071] 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. In general, preferred aspects, particularly in accordance with the invention provide for immobilization of the complex 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 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.

IDT Exhibit 1020.5652

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).

[0072] 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 (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).

[0073] 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.

IDT Exhibit 1020.5653

Kits

[0074] 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 primers, a reverse transcription enzyme and optionally reagents for amplification.

[0075] The containers of the kits will generally include at least one vial, test tube, flask, bottle, syringe or other containers, into which a component may be placed, and preferably, suitably aliquotted. Where there is more than one component in the kit, the kit also will generally contain a second, third or other additional container into which the additional components may be separately placed. However, various combinations of components may be comprised in a container.

[0076] 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 may 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.

[0077] A kit will preferably include instructions for employing the kit components as well the use of any other reagent not included in the kit. Instructions may include variations that can be implemented.

EXAMPLES

Example 1

Reaction Conditions and Reagents:

[0078] In this example cDNA was generated from total RNA (10ng total Brain RNA, Ambion). First strand cDNA was generated using a variety of tailed first primers, which differ in the 5'-tail sequence. Two different reverse transcriptases (RTa and RTb) were used with the various first strand primers. The reaction conditions, both buffer compositions and incubation conditions, were selected for the given RT according to the manufacturer's suggestions. For RTa in this and the following example, first and second cDNA reagents and reaction conditions were those of Ovation Pico (NuGEN Technologies). The RTb reagents and reaction conditions were those of Ovation One Direct RNA amplification (NuGEN Technologies). The various first strand primers employed in the example comprise a 3'-end random hexamer sequence. First strand primers A, B, and C were tested, as well as a mixture of primer A comprising random hexamer at the 3'-end and a primer comprising a poly dT sequence at the 3'-end and the same 5-RNA tail as the A first strand primers. Primers A, B, and C are chimeric primers comprising a RNA tail. cDNA was also generated with tailed all DNA first primers which are the same sequence composition as the A or B or C. cDNA was also generated using N6 or N9 primers, representing a hexamer or a nine-mer of a random DNA sequence. The cDNA generated using these non tailed primers equally represented the total RNA as generated by the corresponding RT. These non-tailed random primers are commonly used for random priming to generate cDNA from RNA.

[0079] Second strand synthesis was carried out as per the manufacturer's instruction (Ribo-SPIA second strand synthesis).

IDT Exhibit 1020.5654

[0080] Further amplification of the cDNA was carried out following generation of the double stranded cDNA by SPIA, employing chimeric amplification 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 reaction conditions and components were the same for all reactions, regardless of RT selection. SPIA amplification was carried out according to manufacturer instructions, with chimeric amplification primer corresponding to the first primer(s) utilized in the first strand synthesis.

[0081] Non-amplified (double stranded cDNA) and amplified cDNA quantification was carried out by real-time qPCR with SYBR green, using specific primer pairs for the various RNA components (18S rRNA, 28S rRNA, various mRNA, 12S mitochondrial RNA (mtrRNA), 16S mtrRNA, and various mitochondrial transcripts) as shown. PCR primers were designed to interrogate/quantify different sequence regions along the test transcripts. Real-time PCR was carried out following the dilution of aliquots taken at the end of second strand synthesis (non-amplified cDNA) or at the end of linear amplification (amplified cDNA). The aliquots were diluted in Tris-EDTA and 2ul were added to the real time PCR reactions.

Results:

[0082] Real time PCR quantification of a) non-amplified cDNA of rRNA (18S and 28S), b) mRNA (specific mRNAs as stated), c) mitochondrial rRNA (mt rRNA, 12S and 16S) and d) mitochondrial gene transcripts demonstrated reduced reverse transcription of rRNA as compared to other transcripts. This reduced reverse transcription of rRNA is affected in part by the choice of RT and in part by the choice of the sequence of tailed first strand random primers, as shown in the figures.

[0083] *Figures 1-4*: The effect of the choice of tailed first primer on the efficiency of reverse transcription of rRNA was reduced efficiency of reverse transcription (higher delta Ct of cDNA normalized to N9), dependent on the choice of the first primer.

[0084] *Figures 5-6*: The efficiency of reverse transcription of rRNA was affected by the choice of RT. The effect of the choice of RT on the efficiency of transcription was determined by the delta Ct for the various non-amplified cDNA (real time PCR) generated with the same first primers and different RTs (RTa or RTb). Here, delta Ct equals Ct for RTa minus Ct for RTb

[0085] *Figures 7-10*: Another result of the effect of the choice of the tailed first strand cDNA primer and the RT employed for the generation of the non-amplified cDNA, is shown. Specifically, shown is the quantification of the various regions on rRNA as was carried out by real time PCR. The average Ct value across all the regions for a given first strand tailed primer and RT (avg. Ct), and standard deviation across these values as well as the non tailed random primers, N9 and N6, is shown. The rRNA content is affected by choice of first strand primer, and to a lesser extent by the RT. The effect on mitochondrial rRNA is minimal by both. The overall effect on the content of rRNA in the non amplified cDNA, and subsequently in the amplified cDNA, is greater than a 10-fold reduction (delta Ct greater than 3), thus

IDT Exhibit 1020.5655

rendering the methods of the inventions most effective relative to other methods such as rRNA removal by enzymatic digestions, hybridization, or selective transcription of only mRNAs with polyA sequences.

[0086] *Figures 11-14*: Real time PCR quantification of amplified cDNA of rRNA (18S and 28S), mRNA (specific mRNAs as stated), mitochondrial rRNA (mt rRNA, 12S and 16S) and mitochondrial gene transcripts, was carried out following single primer linear isothermal amplification (SPIA). The results, shown in the figures, demonstrate that the pattern of relative reduction of the content of rRNA shown above in the corresponding non-amplified cDNA is maintained in the amplified cDNA. The results further demonstrate that, as shown for non-amplified cDNA, there is a minimal effect of primer selection or RT on the content of mrtRNA in the amplified cDNA generated from total RNA.

Example 2

[0087] This example describes reduction in the representation of rRNA in amplified cDNA from C. elegans total RNA. Non-amplified cDNA was prepared priming with Primer A (3' & random) or Primer B (3' & random) and RTa or RTb and using first and second strand reagents and protocol. Total C. elegans RNA input was 2ng. The cDNA was further amplified with two different reagent and protocol systems: SPIA amplification using reagents and protocols from the Ovation Pico kit (NUGEN) or reagents and protocol from WGA-Ovation kit (NuGEN). rRNA and non rRNA cDNA in the double stranded cDNA generated by the different primer sets, RT and protocols were quantified by qPCR as described in Example 1.

[0088] Figures 15 and 16 represent the delta CT (dCT) of quantification of the same transcript loci in amplified cDNA prepared with different first strand RT and amplified with the same SPIA amplification reagents and protocols. Similar differential reduction of rRNA, but not non rRNA or mRNA, is observed in products generated by the same SPIA amplification reagents and protocol.

[0089] Figures 17 and 18 illustrate quantification of rRNA and mRNA in amplified cDNA prepared by cDNA synthesis using the same RT and cDNA reagents and protocol (RTa in Figure 17 and RTb in Figure 18) and amplified by different SPIA reagents (SPIA amplification from Ovation Pico kit and SPIA amplification from WGA-Ovation kit). rRNA and mRNA representation is not differentially affected by the SPIA amplification system.

[0090] The results in this example indicate that the reduced representation of rRNA in the amplified cDNA is dependent on the combination of primer set and cDNA synthesis (RTa and RTb).

Example 3

[0091] This example describes differential reduction in rRNA representation in amplified cDNA from C. elegans total RNA as measured by mass normalized quantification by qPCR and massively parallel sequencing (RNA-Seq) using Illumina's Genome Analyzer.

[0092] Quantification of rRNA and mRNA in purified amplification products generated by RTa (Ovation Pico cDNA) with two primer sets demonstrate the reduction of rRNA representation in the products generated with one of the two primer sets relative to the other. The effect is independent of the

IDT Exhibit 1020.5656

SPIA amplification reagent and conditions, as demonstrated by results from products generated using the reagents and conditions for SPIA amplification reaction from the Ovation Pico or WG-Ovation kits.

[0093] Quantification by qPCR of rRNA was carried out with 10pg input (amplified products) and that of the mRNA transcripts with 10ng input (amplified products). The corrected dCt reflects the correction of the rRNA values to 10ng input (mass normalized).

[0094] The two data sets represented in Table 1 are from two separate experiments each carried out in duplicate, and demonstrate the reproducibility of the amplification reactions. The corrected dCt equals the average Ct for mRNA minus the average Ct for rRNA, for products generated using each of the defined conditions. The delta Ct for products generated by the Primer C set and RTa (cDNA synthesis with Ovation Pico reagents and reaction conditions) in both experiments is about 4 Ct lower than that obtained with the primer A set. This difference, about 10 fold, is reflected in the representation of the rRNA detected by sequencing, Table 3. The data obtained from transcripts quantified by qPCR provides a rough estimate of the representation of the rRNA in the total product of the amplified cDNA.

[0095] A more accurate value of the rRNA representation in the amplified cDNA can be achieved from massively parallel sequencing enabled by the Next Generation Sequencing technologies and platforms, as represented by the RNA-Seq data using Illumina's Genome Analyzer. Sequencing data were obtained for two C. elegans libraries generated from amplified cDNA prepared by the primer sets and protocols described above. For sequencing ready library preparations, purified amplified cDNA was made double stranded by random priming (NuGEN Exon kit). Library preparation for sequence analysis on Illumina Genome Analyzer was carried out using Illumina's protocols and reagents. The sequence data was generated from a single lane run according to the manufacturer instructions. The two libraries were prepared at different times and by different laboratories. In so far as the rRNA sequencing reads are the highest abundant species the sequencing data for both libraries are suitable for the estimation of the effect of rRNA representation in the amplified cDNA. The reduced representation of rRNA in the total amplified cDNA prepared with primer C set, RTa cDNA synthesis, and WGA-Ovation SPIA amplification, as compared to that prepared with primer set A, RTa cDNA synthesis and Ovation-Pico SPIA amplification, is shown in table 2. Specifically, Table 2 provides a summary of C. elegans rRNA representation in RNA-Seq data from Illumina Genome Analyzer IIx sequencing data generated with two libraries. Column A provides results for a library prepared from amplified cDNA generated with primer A (3' & random), RTa cDNA (Ovation Pico), and SPIA amplification using the Ovation Pico reagents. Column C provides results for a library prepared from amplified cDNA generated with primer C (3' & random), RTa cDNA (Ovation Pico), and SPIA amplification using the WG-Ovation reagents.

[0096] Further indication of the dependence of the effect of the efficiency of first strand cDNA priming and reverse transcription on the rRNA representation is demonstrated from the analysis of alignment of the rRNA sequencing reads to the rRNA sequences, as shown in Figures 18 and 19. The solid line corresponds to the depth of coverage that is derived from the library prepared from amplified cDNA using Primer A (3' & random), RTa and SPIA amplification using the reagents and protocols from Ovation Pico kit, and the dashed line corresponds to the depth of coverage that is derived from the library

IDT Exhibit 1020.5657

prepared from amplified cDNA using primer C (3' & random), RTa and SPIA amplification using the reagents and protocols from WGA-Ovation kit.

Table 1:

	Avg. rRNA Ct (10pg)	Avg. non rRNA Ct (10ng)	dCt corrected	ddCt
Primer A (3' & random) Pico SPIA 1	13.4	21.5	14.1	
Primer A (3' & random) WGA SPIA 2	13.3	21.4	14.1	
Primer C (3' & random) Pico SPIA 1	16.7	21.3	10.5	3.6
Primer C (3' & random) WGA SPIA 2	16.1	20.0	9.9	4.1

Table 2:

	A	C
Reads	15722172	10922087
rRNA	10008595	3646359
ws190 aligned reads	10319515	4343998
% non-ribosomal WS190 aligned reads	3.00%	16.10%
% rRNA	97.00%	83.90%
18S	2123935	2028797
23S	1376060	1471198

Table 3: Expected differential reduction of % rRNA representation relative to non-ribosomal transcripts.

			Calculated rRNA		
	Non ribosomal RNA	rRNA	% of total	Fold reduction	delta Ct (relative to A)
A	3	97	97%		
B	3	9.7	76%	10	3.3
C	3	3	50%	32.3	5.0
D	3	0.3	9%	323.3	8.3

Example 4

[0097] This example describes reduction in rRNA representation in amplified cDNA from different human total RNA samples, measured by massively parallel sequencing using Illumina Genome Analyzer IIx (RNA-Seq).

[0098] Reduced representation of rRNA in amplified cDNA is demonstrated by RNA-Seq data obtained for different amplified cDNA prepared from different human total RNA samples. The samples analyzed in this example represent both high quality total RNA (human brain reference RNA) and low quality total

IDT Exhibit 1020.5658

RNA derived from FFPE sample (colon tumor). The details of the samples used, input into the amplification using various NuGEN's RNA amplification kits, reagents and protocols for first strand cDNA synthesis and SPIA amplification, and library preparation (Encore; NuGEN Technologies) for the massively parallel sequencing are summarized in Table 4.

[0099] The rRNA representation in the total amplified cDNA obtained from the sequencing data for the various preparations and samples is summarized in Table 5. In Table 5, 103803-1 and -2 are replicate amplifications from same RNA; and 6957-1 and -2 are replicate libraries prepared from a single amplification. rRNA representation is shown to be affected by the overall reagents and protocols used for the preparation of amplified cDNA, and marked reduction in rRNA representation is shown for all samples, including those derived from poor quality RNA derived from FFPE samples.

Table 4:

Sample	RNA	Input	1st strand cDNA primer	1st strand cDNA protocol	SPIA protocol	Library prep
FFPE	FFPE colon tumor	100ng	RNA-SEQ	FFPE	FFPE	Encore
RNA-SEQ$_m$	FFPE colon tumor	100ng	RNA-SEQ	FFPE	Pico	Encore
RNA-SEQ	Human brain reference	10ng	RNA-SEQ	Pico	Pico	Encore

Table 5:

		rRNA
FFPE	103803-1	31.4%
	103803-2	27.6%
	6957-1	30.6%
	6957-2	29.0%
RNA-SEQ$_m$	103803-1	13.8%
	103803-2	14.9%
	6957-1	15.2%
	6957-2	15.9%
RNA-SEQ	Human brain RNA reference	5.8%

[00100] Table 6 provides sequences of example primers for reverse transcription ("random"), cDNA second strand synthesis ("3'"), and SPIA primers useful in the methods and compositions of the invention. Ribonucleotides appear in lower case and deoxyribonucleotides appear in upper case. "N" represents a random nucleotide.

Table 6:

Primer ID	Sequence	Length
A SPIA	*cuucuauaguuuagguaacuuug* TGTTTGA	30
A random	*cuucuauaguuuagguaacuuuguguuuga* NNNNNN	
A 3'	*cuucuauaguuuagguaacuuuguguuuga* TTTTTTTTTTTTNN	

IDT Exhibit 1020.5659

B SPIA	*gguaauacgacucacuau* AGGCAGA	25
B random	*gguaauacgacucacuauaggcaga* NNNNNN	
B 3'	*gguaauacgacucacuauaggcaga* TTTTTTTTTTTTNN	
C SPIA	*gacggaugcggucu* CCAGTGT	21
C random	*gacggaugcggucuccagugu* NNNNNN	
C 3'	*gacggaugcggucuccagugu* TTTTTTTTTTTTNN	
D SPIA	*cccuccaaggctccc* CAGTATC	22
D random	*cccuccaaggctccccaguauc* NNNNNN	
D 3'	*cccuccaaggctccccaguauc* TTTTTTTTTTTTNN	
E SPIA	*agggugagaaaggcc* GGAGACA	22
E random	*agggugagaaaggccggagaca* NNNNNN	
E 3'	*agggugagaaaggccggagaca* TTTTTTTTTTTTNN	
F SPIA	*acacauacgauuuagguga* CACTAT	25
F random	*acacauacgauuuaggugacacuau* NNNNNN	
F 3'	*acacauacgauuuaggugacacuau* TTTTTTTTTTTTNN	
G SPIA	*cguuaaguaugaacccag* ATGTATT	25
G random	*cguuaaguaugaacccagauguauu* NNNNNN	
G 3'	*cguuaaguaugaacccagauguauu* TTTTTTTTTTTTNN	
H SPIA	*gaagacagauggugc* AGCCACA	22
H random	*gaagacagauggugcagccaca* NNNNNN	
H 3'	*gaagacagauggugcagccaca* TTTTTTTTTTTTNN	
I SPIA	*cguauucugacgacguacuc* TCAGCCT	27
I random	*cguauucugacgacguacucucagccu* NNNNNN	
I 3'	*cguauucugacgacguacucucagccu* TTTTTTTTTTTTNN	

[00101] 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 1020.5660

CLAIMS

WHAT IS CLAIMED IS:

1. A method for differentially reducing the reverse transcription of ribosomal RNA (rRNA) from an RNA sample comprising:

(a) providing one or more primers of known sequence;

(b) combining the one or more primers with a reverse transcriptase (RT); and

(c) reverse transcribing the RNA sample, thereby producing reverse transcribed products, wherein the reverse transcription of rRNA is reduced.

2. The method of claim 1, wherein the reverse transcription of rRNA is reduced by at least about 10 fold.

3. The method of claim 1, further comprising step (d) amplifying said reverse transcribed products, thereby producing amplified products.

4. The method of claim 1, further comprising sequencing said reverse transcribed products.

5. The method of claim 4, wherein said reverse transcribed products are amplified prior to sequencing.

6. The method of claim 1, wherein the one or more primers are tailed primers comprising a 3' portion hybridizable to one or more target RNAs in the RNA sample and a 5' portion that is not hybridizable to the one or more target RNAs in the RNA sample.

7. The method of claim 1, wherein the one or more primers are chimeric primers comprising a DNA portion and an RNA portion.

8. The method of claim 7, wherein the chimeric primers comprise a 3'-DNA portion hybridizable to one or more target RNAs in the RNA sample and a 5'-RNA portion that is not hybridizable to the one or more target RNAs in the RNA sample.

9. The method of claim 6, wherein the hybridizable 3' portion of the primers comprise a randomly generated sequence.

10. The method of claim 6, wherein the hybridizable 3' portion of the primers comprise 5-15 nucleotides.

11. The method of claim 6, wherein the hybridizable 3' portion of the primers comprise 6 nucleotides.

12. The method of claim 6, wherein the non-hybridizable 5' portion of the primers comprise a promoter-specific sequence.

13. The method of claim 6, wherein the primer is one of a plurality of primers of known sequence, each primer comprising a 3' portion hybridizable to the sample RNA and a 5' portion that is not hybridizable to the sample RNA, wherein each of the non-hybridizable 5' portions of the primers comprise the same sequence.

14. The method of claim 1, wherein the RNA sample comprises mRNA and rRNA.

IDT Exhibit 1020.5661

15. The method of claim 1, wherein the RNA is selected from the group consisting of total RNA, mitochondrial RNA, chloroplast RNA, DNA-RNA hybrids, viral RNA, cell free RNA, and mixtures thereof.

16. The method of claim 1, wherein the method does not include purification of one or more RNA sequences from the sample.

17. The method of claim 1, wherein the RT is selected from the group consisting of: a Moloney murine leukemia virus RT, a human immunodeficiency virus RT, a rous sarcoma virus RT, an avian myeloblastosis virus RT, a rous associated virus RT, a myeloblastosis associated virus RT, an avian sarcoma-leukosis virus RT, an RT lacking RNase H activity, modified RTs derived therefrom, and combinations thereof.

18. A method of identifying a primer sequence that differentially reduces the reverse transcription of ribosomal RNA (rRNA) from an RNA sample comprising:

(a) providing one or more primers of known sequence;

(b) combining the one or more primers with a reverse transcriptase (RT) to reverse transcribe the RNA sample, thereby producing reverse transcribed products;

(c) optionally generating double-stranded cDNA products from said reverse transcribed products;

(d) optionally amplifying said double-stranded cDNA products; and

(e) analyzing the reverse transcribed products to determine if the reverse transcription of rRNA is reduced by the primer sequence.

19. The method of claim 1, wherein the one or more primers are tailed primers comprising a 3' portion hybridizable to one or more target RNAs in the RNA sample and a 5' portion that is not hybridizable to the one or more target RNAs in the RNA sample.

20. The method of claim 18, wherein the one or more primers are chimeric primers comprising a DNA portion and an RNA portion.

21. The method of claim 18, wherein the reverse transcription of rRNA is reduced by at least about 20%, at least about 30%, at least about 40%, or at least about 50%.

22. The method of claim 18, wherein the reverse transcription of rRNA is reduced by at least about 2-fold.

23. The method of claim 20, wherein the chimeric primers comprise a 3'-DNA portion hybridizable to one or more target RNAs in the RNA sample and a 5'-RNA portion that is not hybridizable to the one or more target RNAs in the RNA sample.

24. The method of claim 19, wherein the hybridizable 3' portion of the primers comprise a randomly generated sequence.

25. The method of claim 19, wherein the hybridizable 3' portion of the primers comprise 5-15 nucleotides.

26. The method of claim 19, wherein the hybridizable 3' portion of the primers comprise 6 nucleotides.

IDT Exhibit 1020.5662

27. The method of claim 19, wherein the non-hybridizable 5' portion of the primers comprise a promoter-specific sequence.

28. The method of claim 19, wherein the primer is one of a plurality of primers of known sequence, each primer comprising a 3' portion hybridizable to the sample RNA and a 5' portion that is not hybridizable to the sample RNA, wherein each of the non-hybridizable 5' portions of the primers comprise the same sequence.

29. The method of claim 19, wherein the primer sequence that reduces the reverse transcription of ribosomal RNA is identified based on varying the non-hybridizable 5' portion of the one or more primers.

30. The method of claim 19, wherein the primer sequence that reduces the reverse transcription of ribosomal RNA is identified based on varying the RT used in step (e).

31. The method of claim 18, wherein the RNA sample comprises mRNA and rRNA.

32. The method of claim 18, wherein the RNA is selected from the group consisting of total RNA, mitochondrial RNA, chloroplast RNA, DNA-RNA hybrids, viral RNA, cell free RNA, and mixtures thereof.

33. The method of claim 18, wherein the RT is selected from the group consisting of: a Moloney murine leukemia virus RT, a human immunodeficiency virus RT, a rous sarcoma virus RT, an avian myeloblastosis virus RT, a rous associated virus RT, a myeloblastosis associated virus RT, an avian sarcoma-leukosis virus RT, an RT lacking RNase H activity, modified RTs derived therefrom, and combinations thereof.

34. A method of identifying a reverse transcriptase enzyme (RT) that differentially reduces the reverse transcription of rRNA from an RNA sample comprising:

(a) providing one or more primers of known sequence;

(b) combining the one or more primers with an RT to reverse transcribe the RNA sample, thereby producing reverse transcribed products;

(c) optionally generating double-stranded cDNA products from said reverse transcribed products;

(d) optionally amplifying said double-stranded cDNA products; and

(e) analyzing the reverse transcribed products to determine if the reverse transcription of rRNA is reduced by the RT.

35. The method of claim 34, wherein the one or more primers are tailed primers comprising a 3' portion hybridizable to one or more target RNAs in the RNA sample and a 5' portion that is not hybridizable to the one or more target RNAs in the RNA sample.

36. The method of claim 34, wherein the one or more primers are chimeric primers comprising a DNA portion and an RNA portion.

37. The method of claim 34, wherein the reverse transcription of rRNA is reduced by at least about 20%, at least about 30%, at least about 40%, or at least about 50%.

IDT Exhibit 1020.5663

38. The method of claim 34, wherein the reverse transcription of rRNA is reduced by at least about 2-fold.

39. The method of claim 36, wherein the chimeric primers comprise a 3'-DNA portion hybridizable to one or more target RNAs in the RNA sample and a 5'-RNA portion that is not hybridizable to the one or more target RNAs in the RNA sample.

40. The method of claim 35, wherein the hybridizable 3' portion of the primers comprise a randomly generated sequence.

41. The method of claim 35, wherein the hybridizable 3' portion of the primers comprise 5-15 nucleotides.

42. The method of claim 35, wherein the hybridizable 3' portion of the primers comprise 6 nucleotides.

43. The method of claim 35, wherein the non-hybridizable 5' portion of the primers comprise a promoter-specific sequence.

44. The method of claim 35, wherein the primer is one of a plurality of primers of known sequence, each primer comprising a 3' portion hybridizable to the sample RNA and a 5' portion that is not hybridizable to the sample RNA, wherein each of the non-hybridizable 5' portions of the primers comprise the same sequence.

45. The method of claim 35, wherein the primer sequence that reduces the reverse transcription of ribosomal RNA is identified based on varying the non-hybridizable 5' portion of the one or more primers.

46. The method of claim 35, wherein the primer sequence that reduces the reverse transcription of ribosomal RNA is identified based on varying the RT used in step (e).

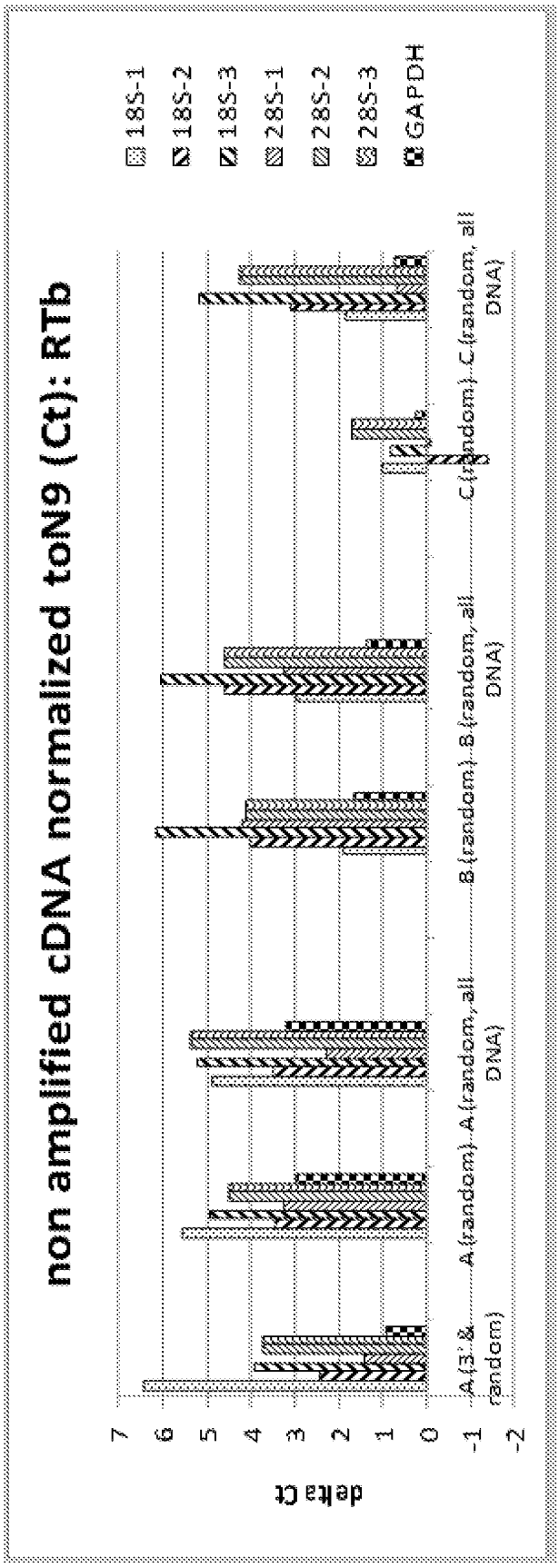
47. The method of claim 34, wherein the RNA sample comprises mRNA and rRNA.

48. The method of claim 34, wherein the RNA is selected form the group consisting of total RNA, mitochondrial RNA, chloroplast RNA, DNA-RNA hybrids, viral RNA, cell free RNA, and mixtures thereof.

49. The method of claim 34, wherein the RT is selected from the group consisting of: a Moloney murine leukemia virus RT, a human immunodeficiency virus RT, a rous sarcoma virus RT, an avian myeloblastosis virus RT, a rous associated virus RT, a myeloblastosis associated virus RT, an avian sarcoma-leukosis virus RT, an RT lacking RNase H activity, modified RTs derived therefrom, and combinations thereof.

IDT Exhibit 1020.5664

Figure 1:


non amplified cDNA normalized toN9 (Ct): RTb
delta Ct
7
6
5
4
3
2
1
0
-1
-2
18S-1
18S-2
18S-3
28S-1
28S-2
28S-3
GAPDH
A (3' & random)
A (random)
A (random, all DNA)
B (random)
B (random, all DNA)
C (random)
C (random, all DNA)

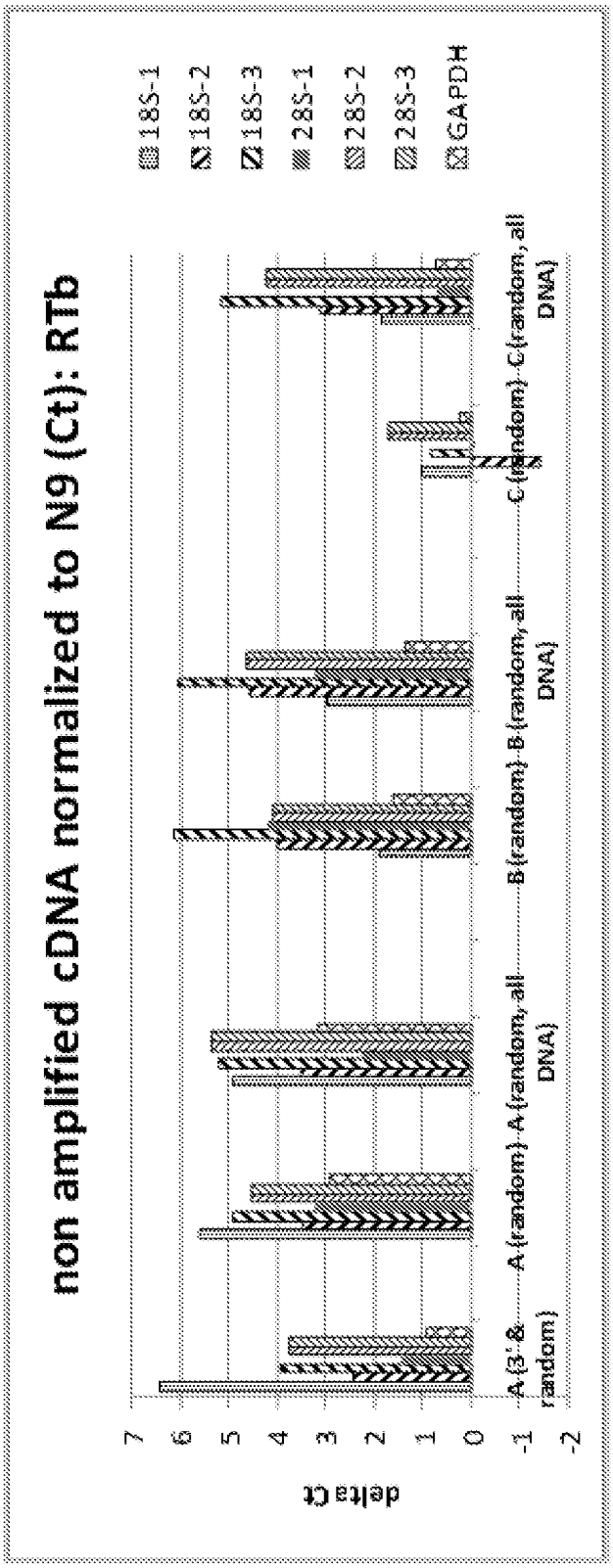

non amplified cDNA normalized to N9 (Ct): RTb
18S-1
18S-2
18S-3
28S-1
28S-2
28S-3
GAPDH
delta ct
7
6
5
4
3
2
1
0
-1
-2
A (3' & random)
A (random)
A (random, all DNA)
B (random)
B (random, all DNA)
C (random)
C (random, all DNA)

Figure 2:

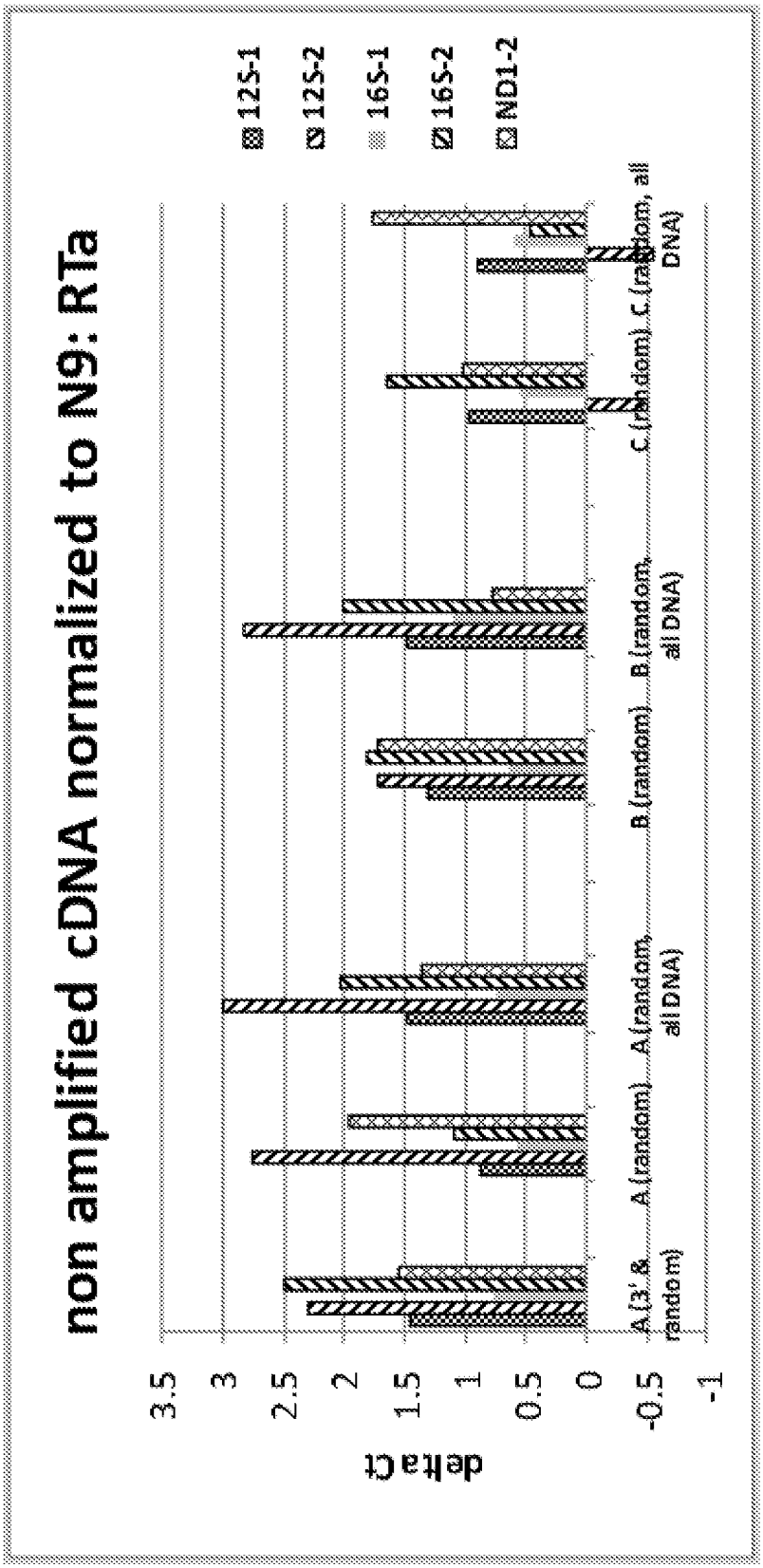

non amplified cDNA normalized to N9: RTa
12S-1
12S-2
16S-1
16S-2
ND1-2
deltaCt
3.5
3
2.5
2
1.5
1
0.5
0
-0.5
-1
A.(3' & random)
A.(random)
A.(random, all DNA)
B.(random)
B.(random, all DNA)
C.(random)
C.(random, all DNA)

Figure 3:

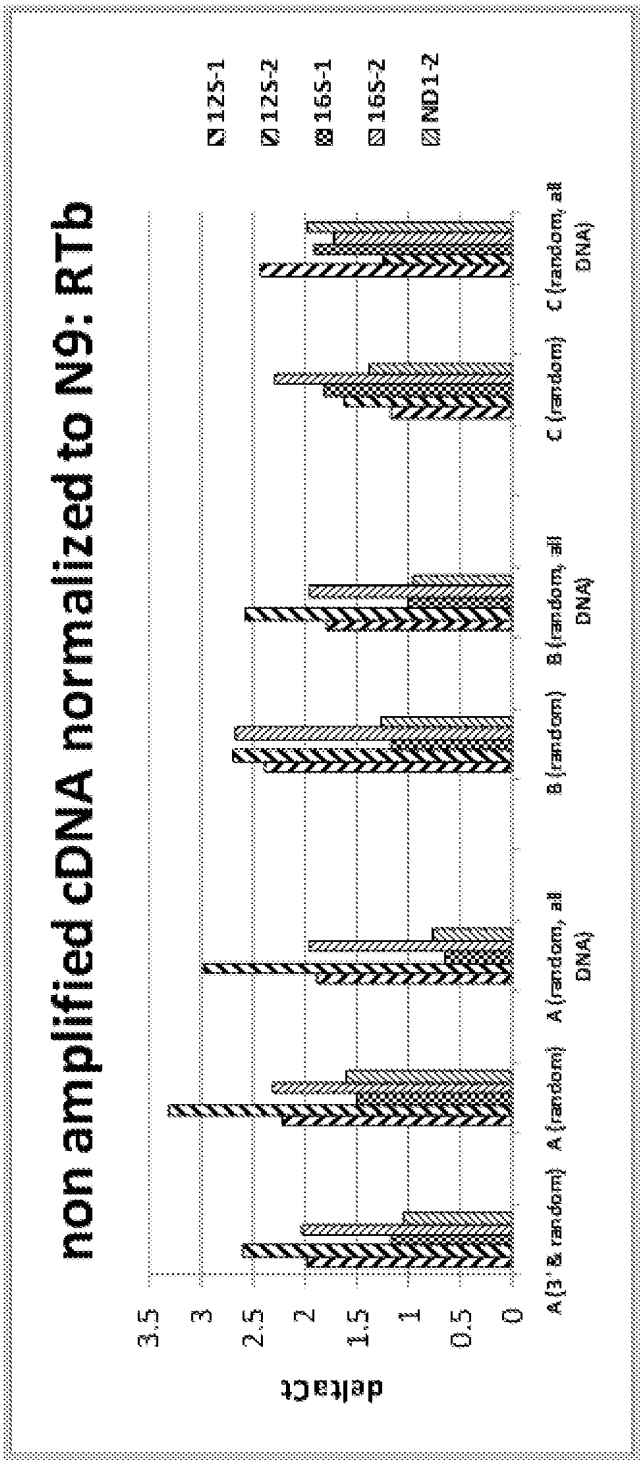

non amplified cDNA normalized to N9: RTb
deltaCt
3.5
3
2.5
2
1.5
1
0.5
0
12S-1
12S-2
16S-1
16S-2
ND1-2
A (3' & random)
A (random)
A (random, all DNA)
B (random)
B (random, all DNA)
C (random)
C (random, all DNA)

Figure 4:

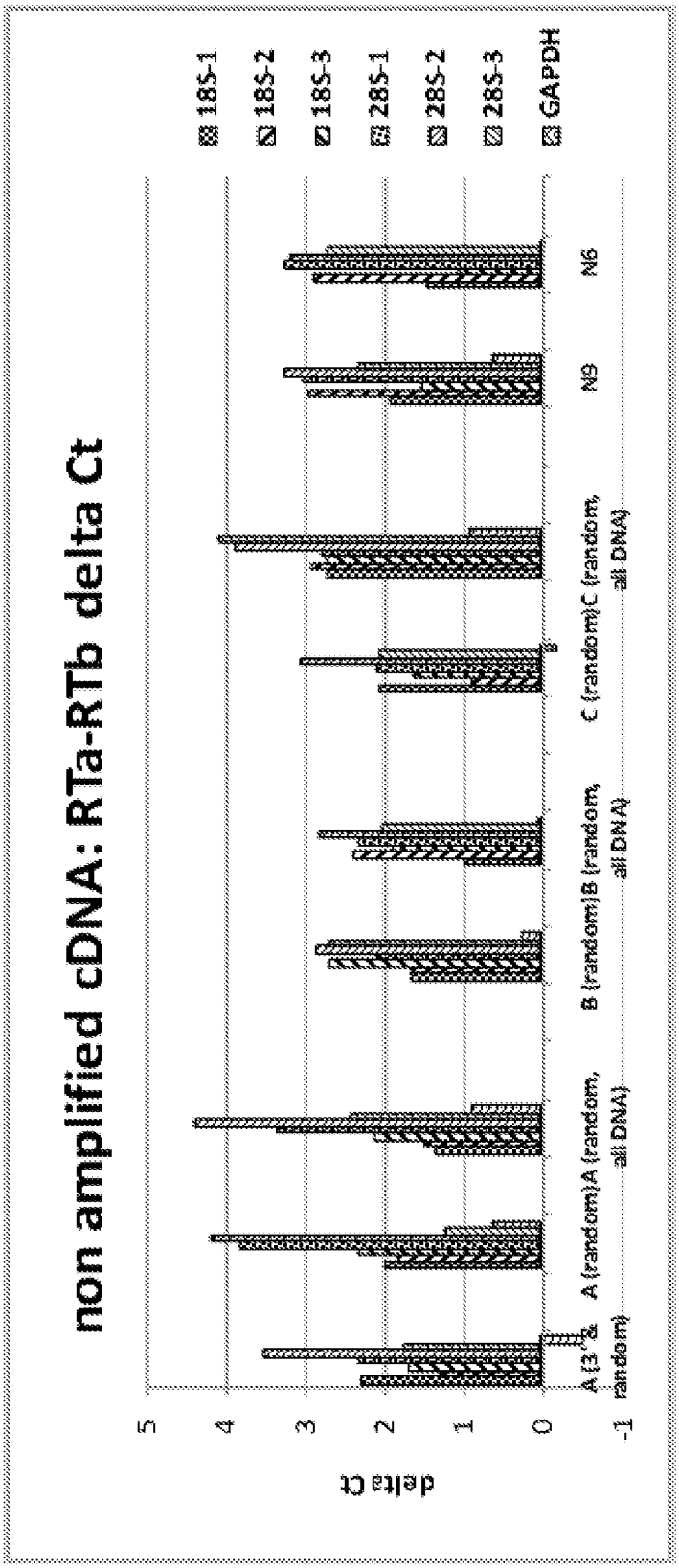

non amplified cDNA: RTa-RTb delta Ct
18S-1
18S-2
18S-3
28S-1
28S-2
28S-3
GAPDH
delta Ct
5
4
3
2
1
0
-1
A (3' & random)
A (random, all DNA)
B (random)
B (random, all DNA)
C (random)
C (random, all DNA)
N9
N6

Figure 5:

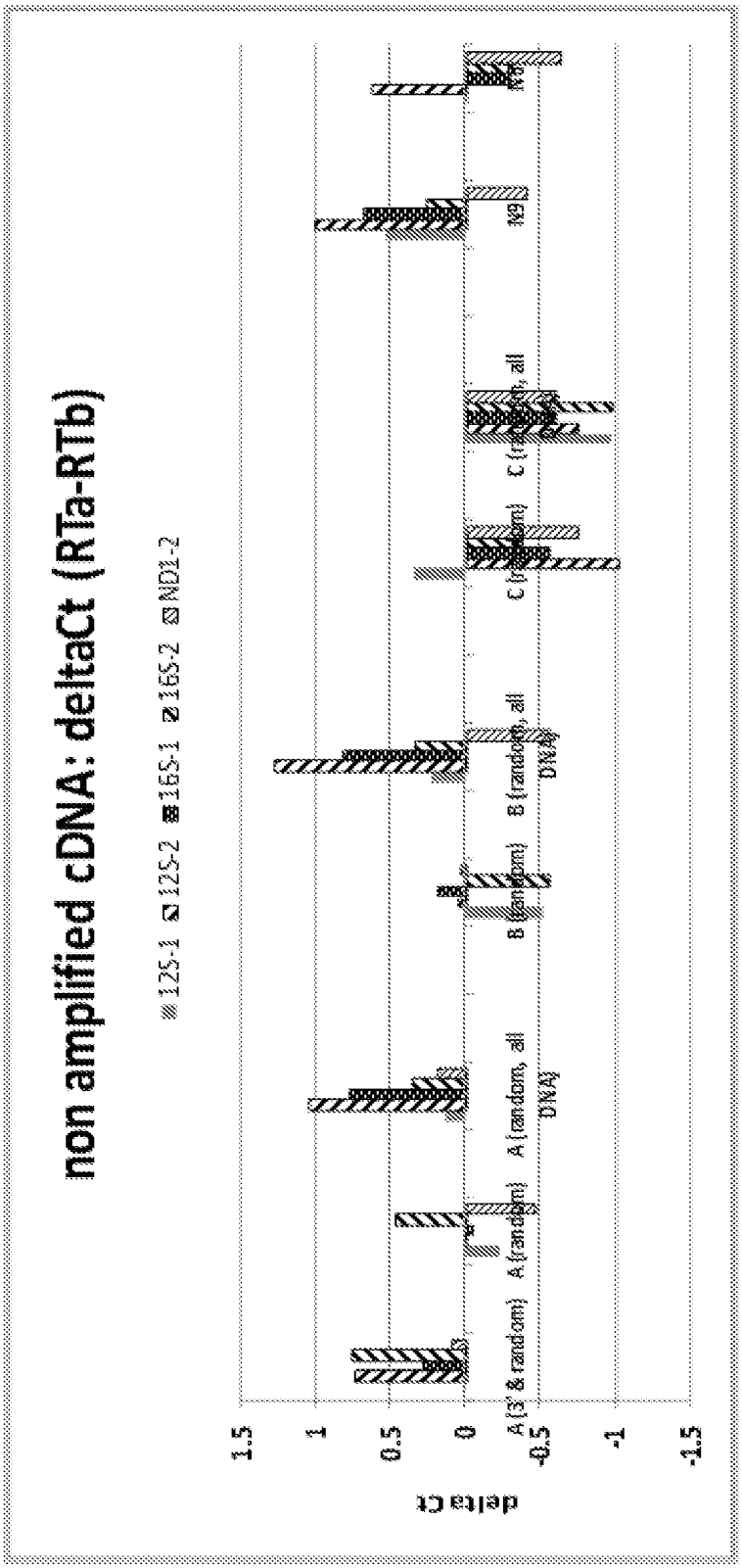

non amplified cDNA: deltaCt (RTa-RTb)
12S-1
12S-2
16S-1
16S-2
ND1-2
delta Ct
1.5
1
0.5
0
-0.5
-1
-1.5
A (3' & random)
A (random)
A (random, all DNA)
B (random)
B (random, all DNA)
C (random)
C (random, all
ND1

Figure 6:

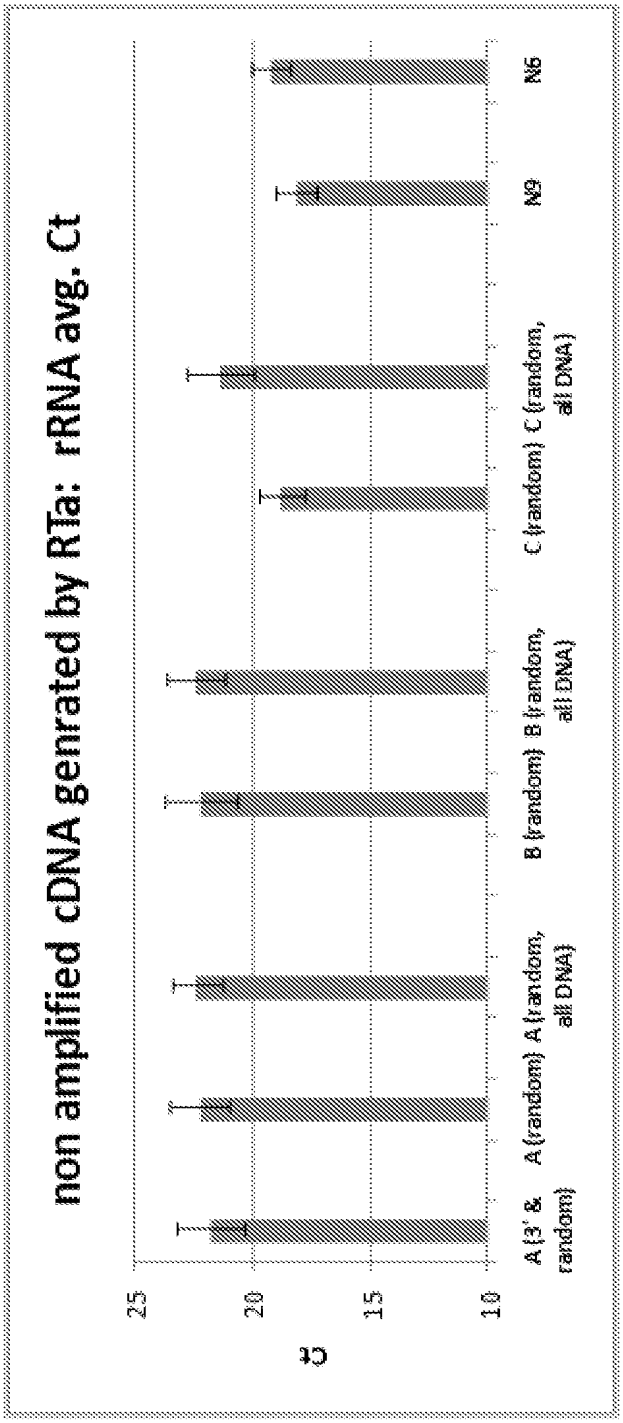

non amplified cDNA genrated by RTa: rRNA avg. Ct
Ct
25
20
15
10
A (3' & random)
A (random)
A (random, all DNA)
B (random)
B (random, all DNA)
C (random)
C (random, all DNA)
N9
N6

Figure 7:

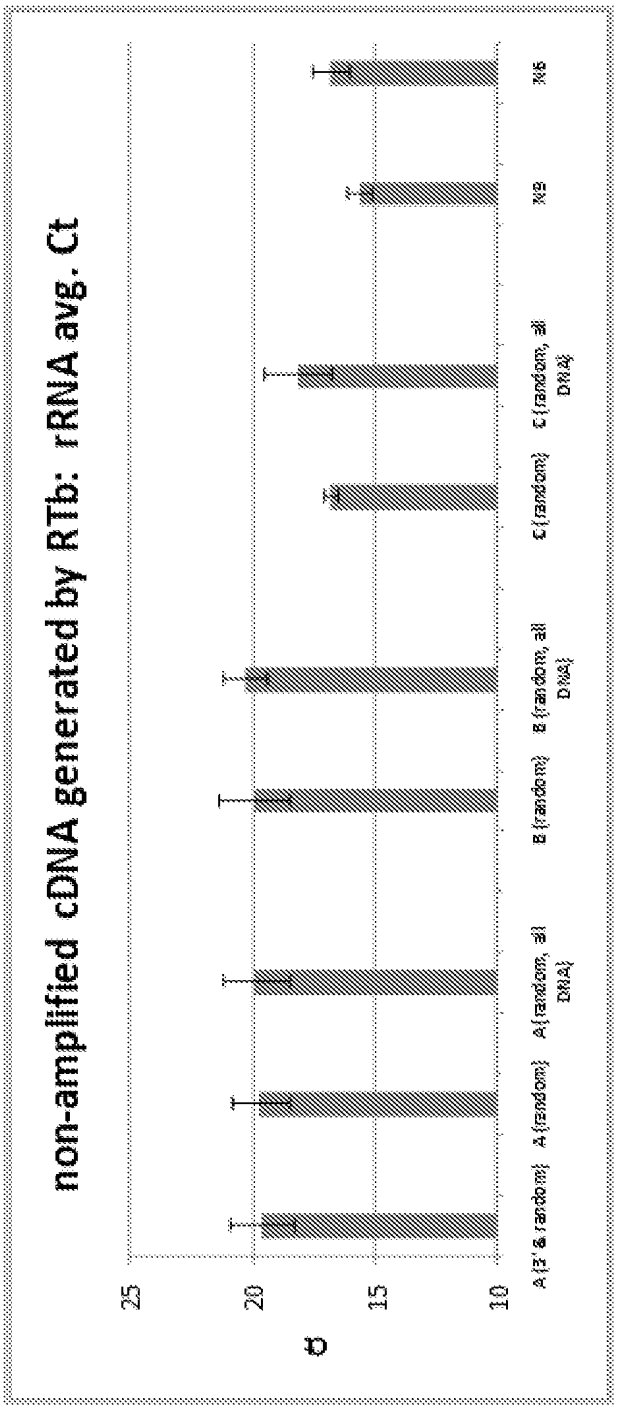

non-amplified cDNA generated by RTb: rRNA avg. Ct
Ct
25
20
15
10
A (3' & random)
A (random)
A (random, all DNA)
B (random)
B (random, all DNA)
C (random)
C (random, all DNA)
NS
NS

Figure 8:

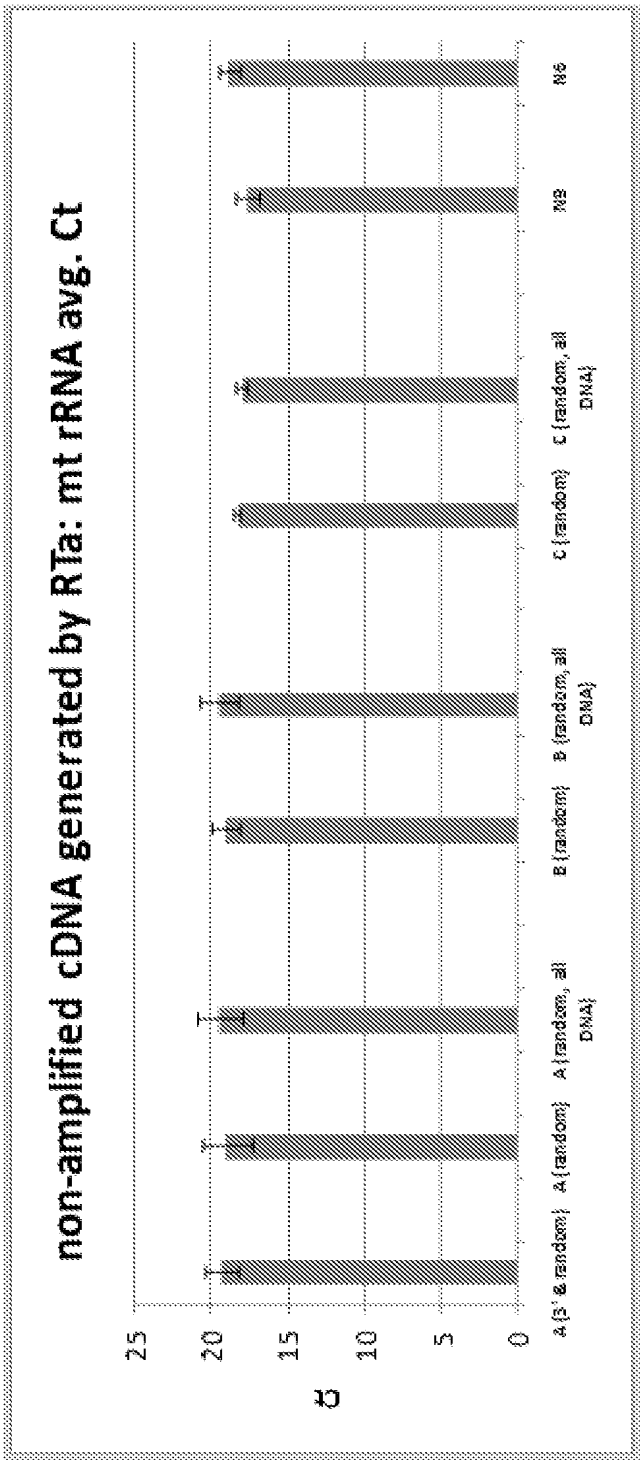

non-amplified cDNA generated by RTa: mt rRNA avg. Ct
Ct
25
20
15
10
5
0
A (3' & random)
A (random)
A (random, all DNA)
B (random)
B (random, all DNA)
C (random)
C (random, all DNA)
N2
N6

Figure 9:

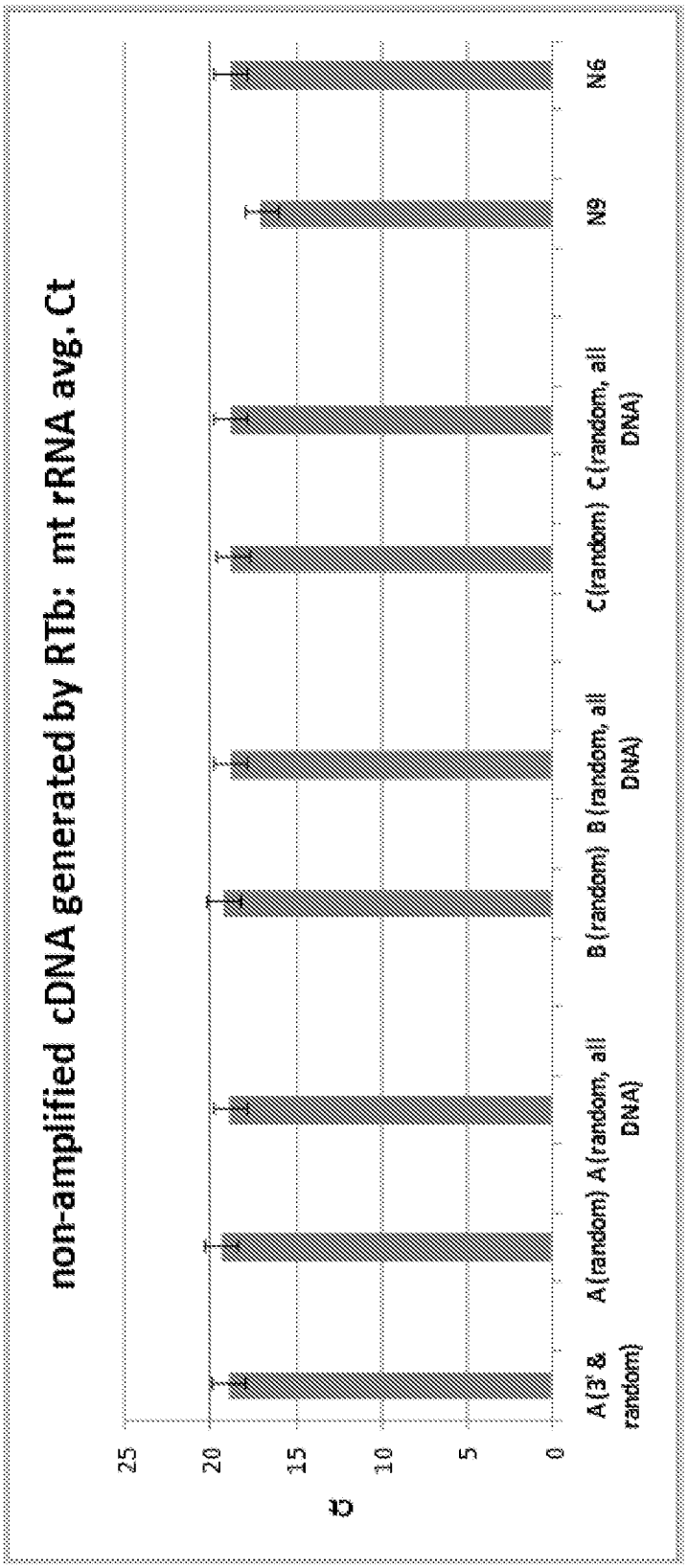

non-amplified cDNA generated by RTb: mt rRNA avg. Ct
Ct
25
20
15
10
5
0
A (3' & random)
A (random)
A (random, all DNA)
B (random)
B (random, all DNA)
C (random)
C (random, all DNA)
N9
N6

Figure 10:

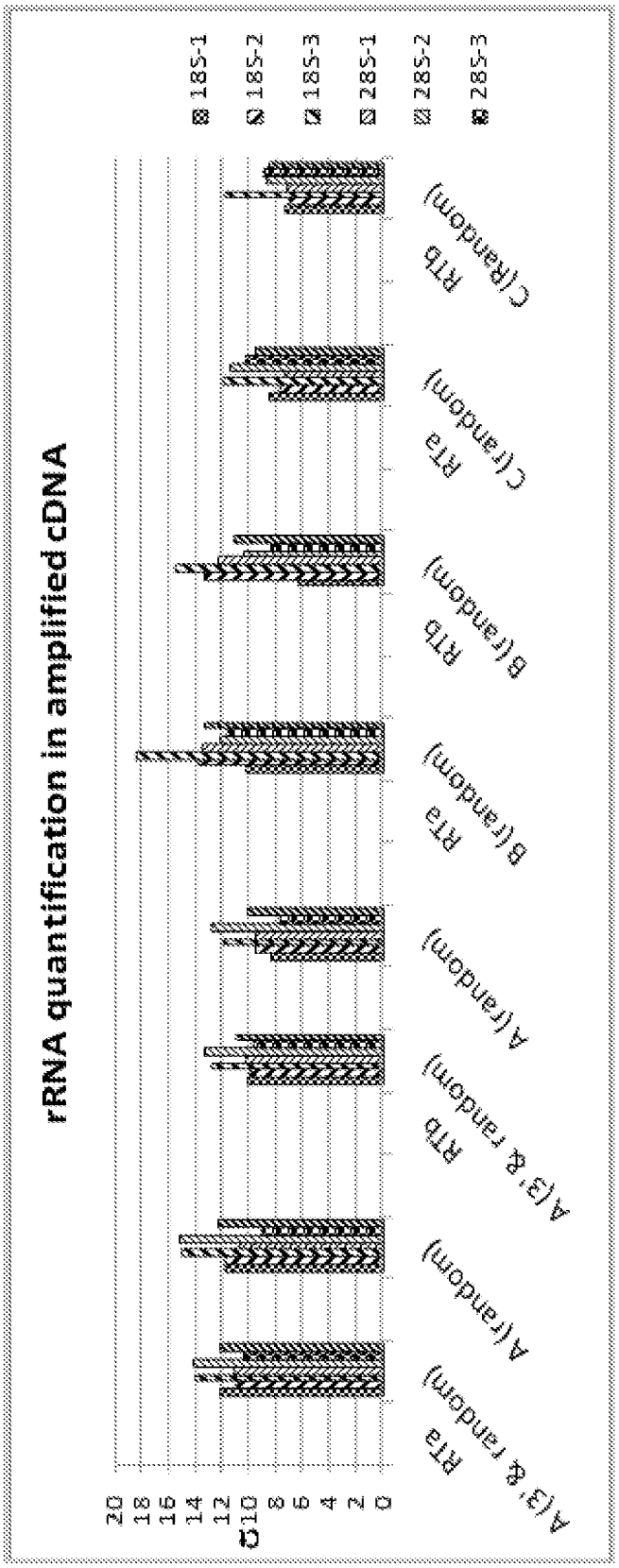

rRNA quantification in amplified cDNA
18S-1
18S-2
18S-3
28S-1
28S-2
28S-3
Ct
20
18
16
14
12
10
8
6
4
2
0
A(3' & random) RTa
A(3' & random) RTb
A(random)
B(random) RTa
B(random) RTb
C(random) RTa
C(random) RTb

Figure 11:

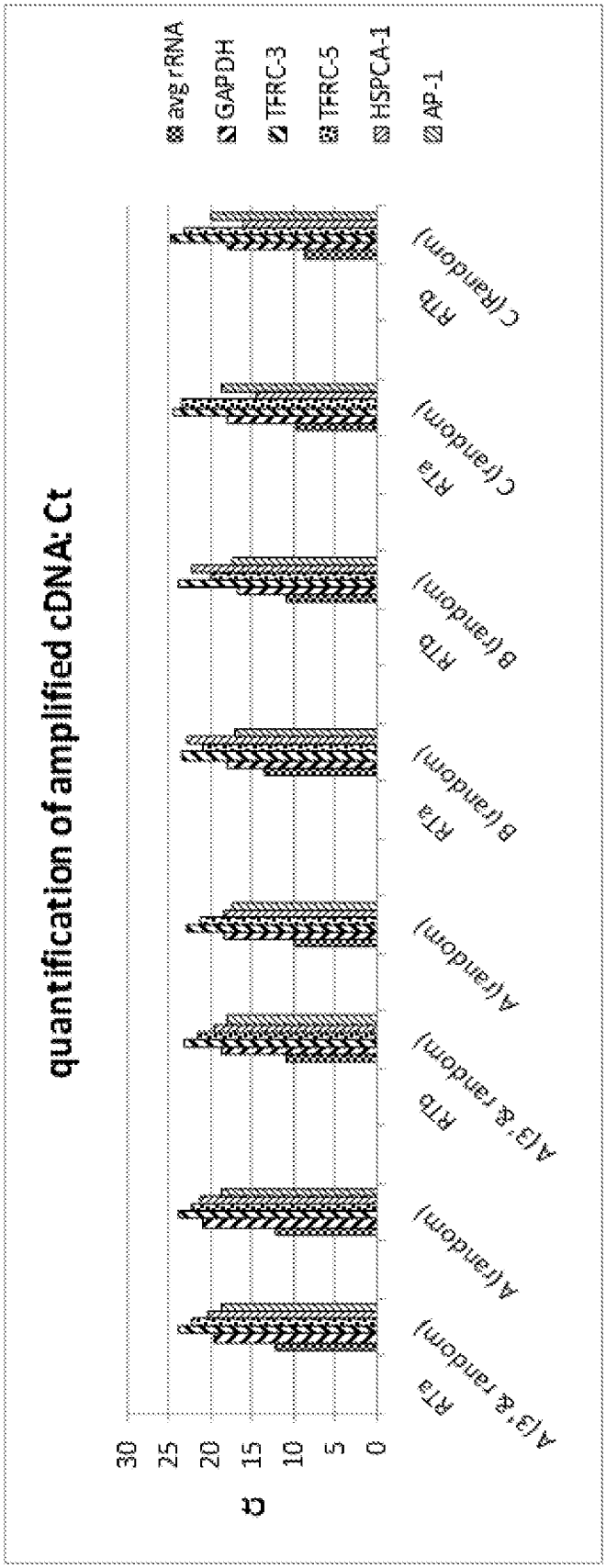

quantification of amplified cDNA: Ct
avg rRNA
GAPDH
TFRC-3
TFRC-5
HSPCA-1
AP-1
Ct
30
25
20
15
10
5
0
A(3'& random)
RTa
A(3'& random)
RTb
A(random)
RTa
A(random)
RTb
B(random)
RTa
B(random)
RTb
C(random)
RTa
C(random)
RTb

Figure 12:

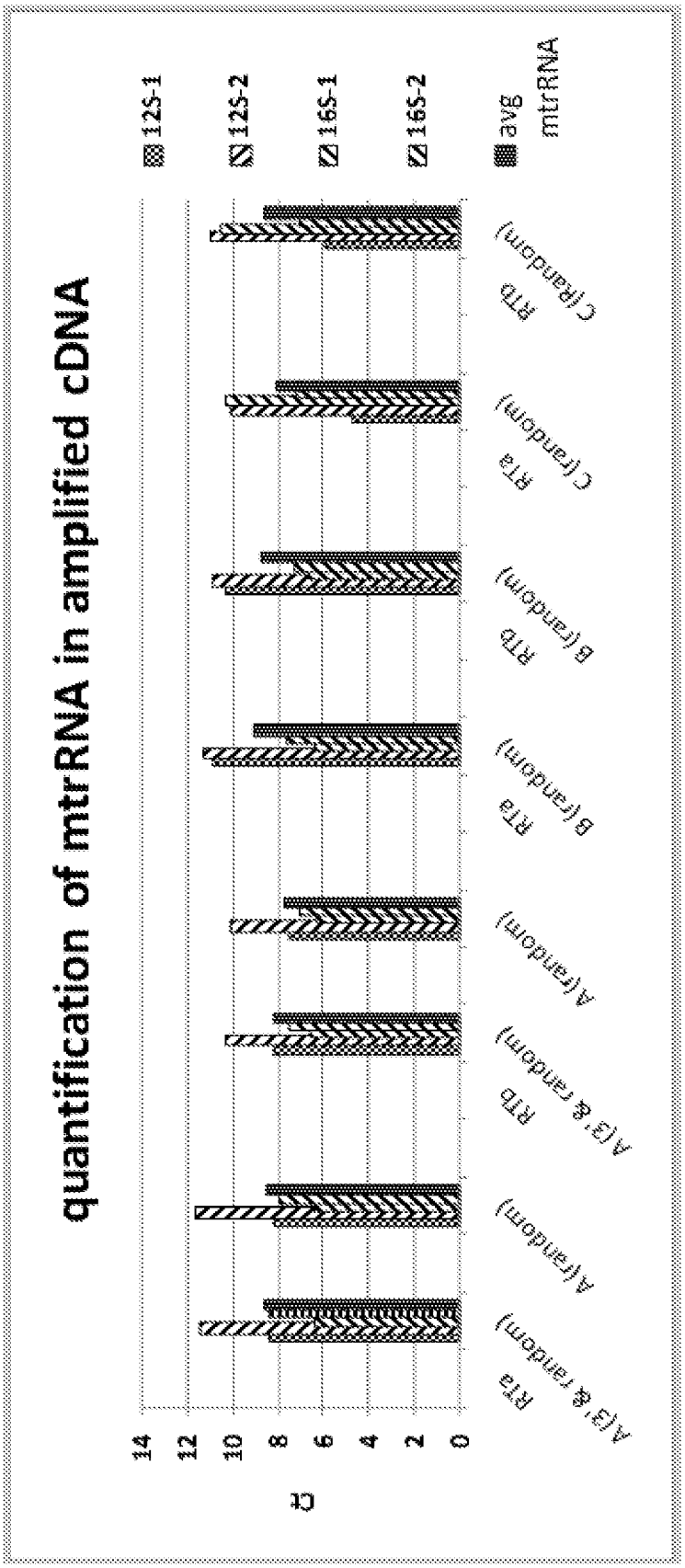

quantification of mtrRNA in amplified cDNA
12S-1
12S-2
16S-1
16S-2
avg mtrRNA
Ct
14
12
10
8
6
4
2
0
A(3' & random)
RTa
A(random)
RTb
A(3' & random)
RTb
A(random)
RTa
B(random)
RTa
B(random)
RTb
C(random)
RTa
C(random)
RTb

Figure 13:

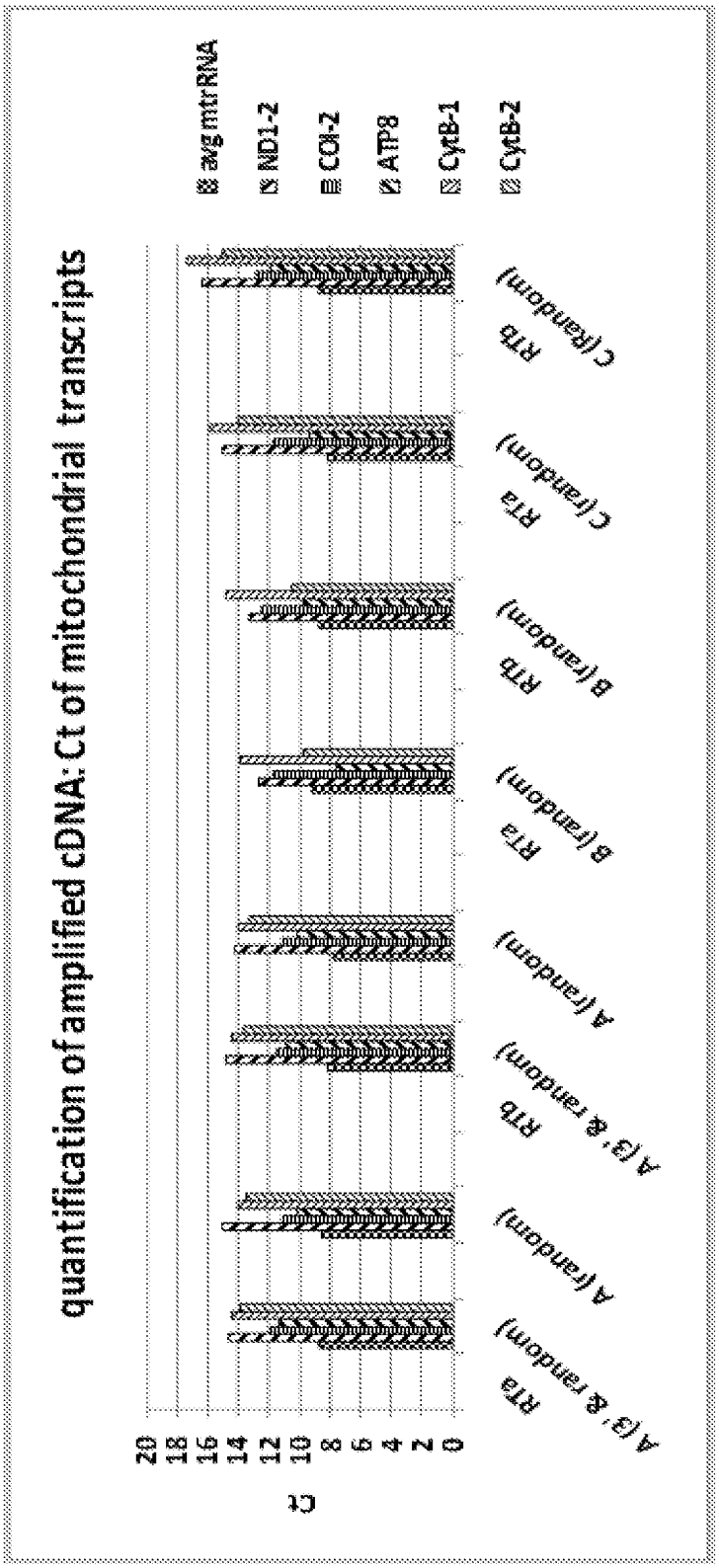

quantification of amplified cDNA: Ct of mitochondrial transcripts
avg mtrRNA
ND1-2
COI-2
ATP8
CytB-1
CytB-2
Ct
20
18
16
14
12
10
8
6
4
2
0
RTa
RTb
A (3', & random)
A (random)
B (random)
C (random)

Figure 14:

Figure 15:

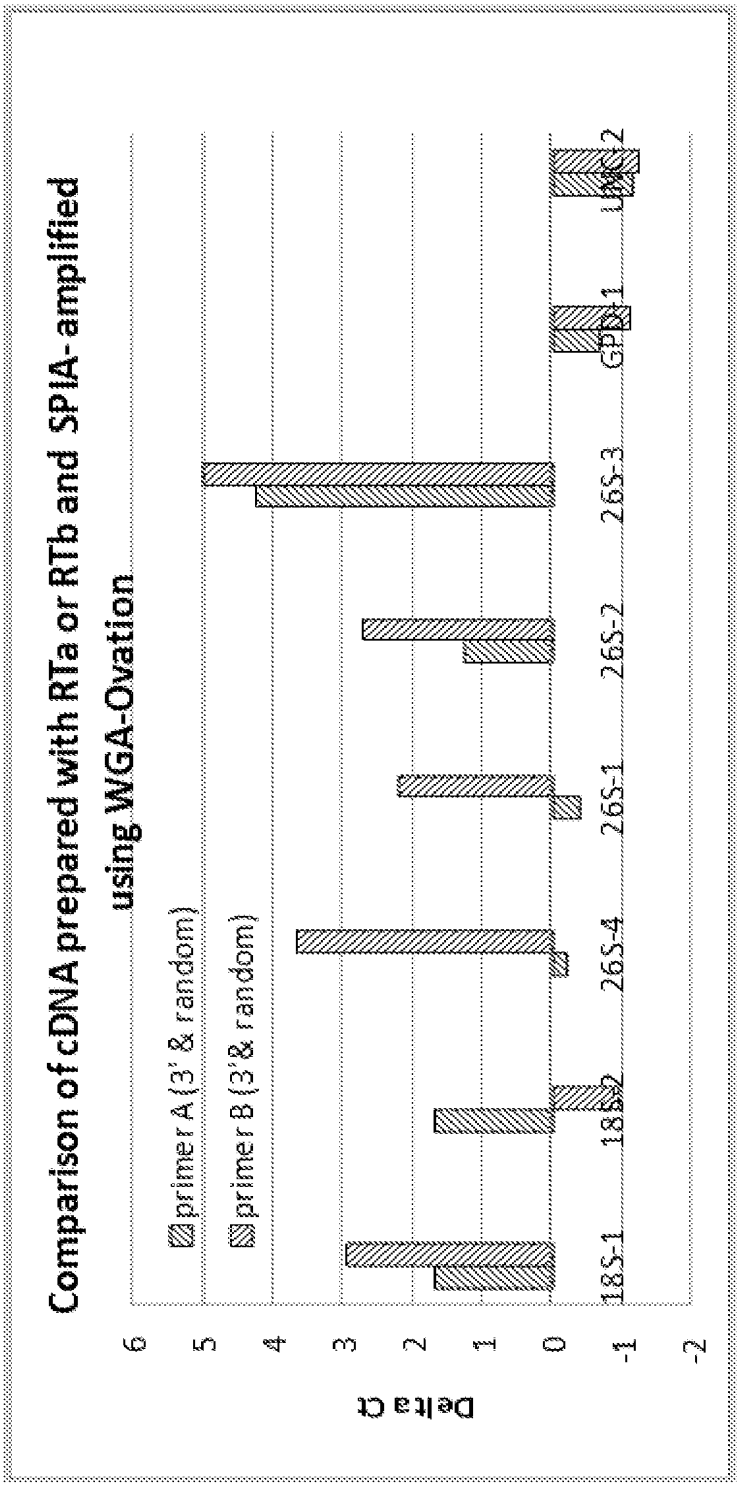

Comparison of cDNA prepared with RTa or RTb and SP1A-amplified using WGA-Ovation
primer A (3' & random)
primer B (3'& random)
Delta Ct
6
5
4
3
2
1
0
-1
-2
18S-1
18S-2
26S-4
26S-1
26S-2
26S-3
GPD-1
UBC-2

Figure 16:

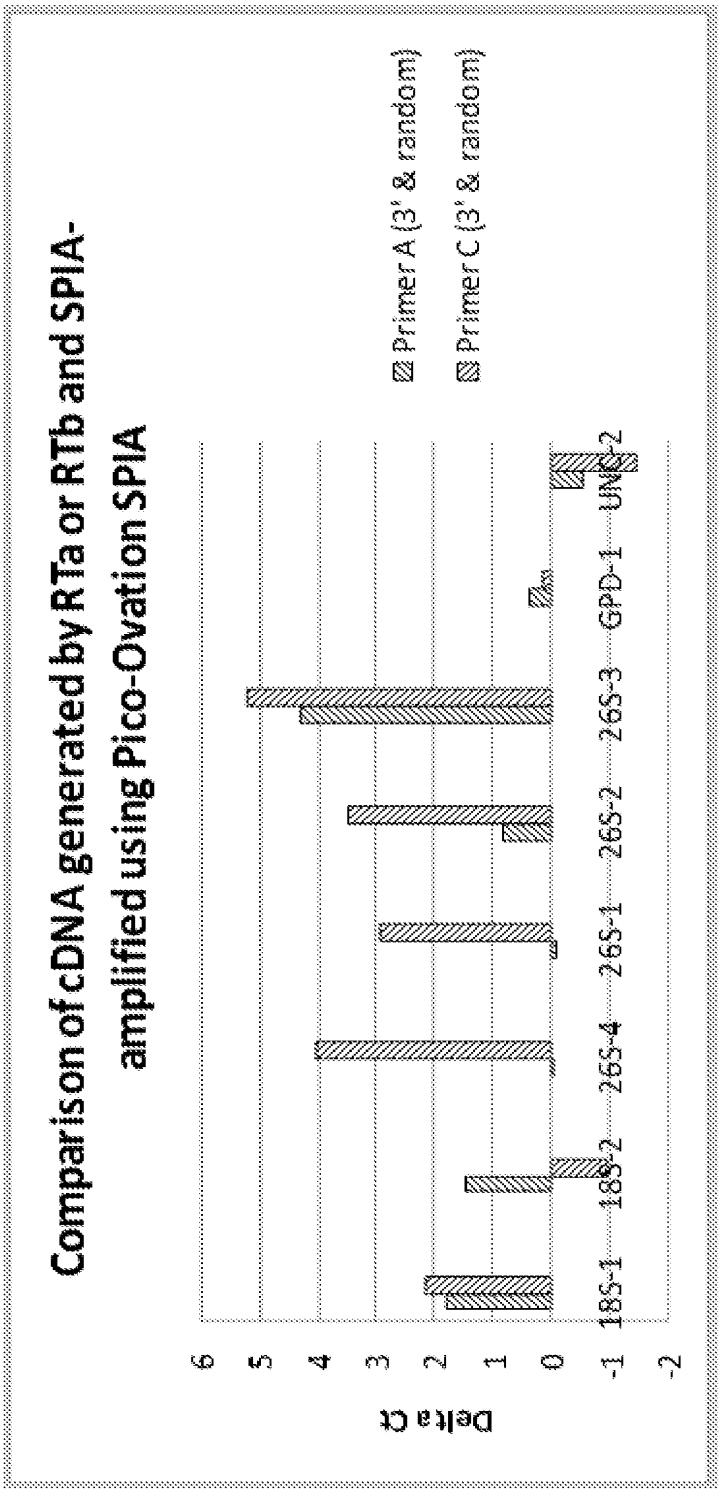

Comparison of cDNA generated by RTa or RTb and SPIA-amplified using Pico-Ovation SPIA
Primer A (3' & random)
Primer C (3' & random)
Delta ct
6
5
4
3
2
1
0
-1
-2
18S-1
18S-2
26S-4
26S-1
26S-2
26S-3
GPD-1
UNC-2

Figure 17:

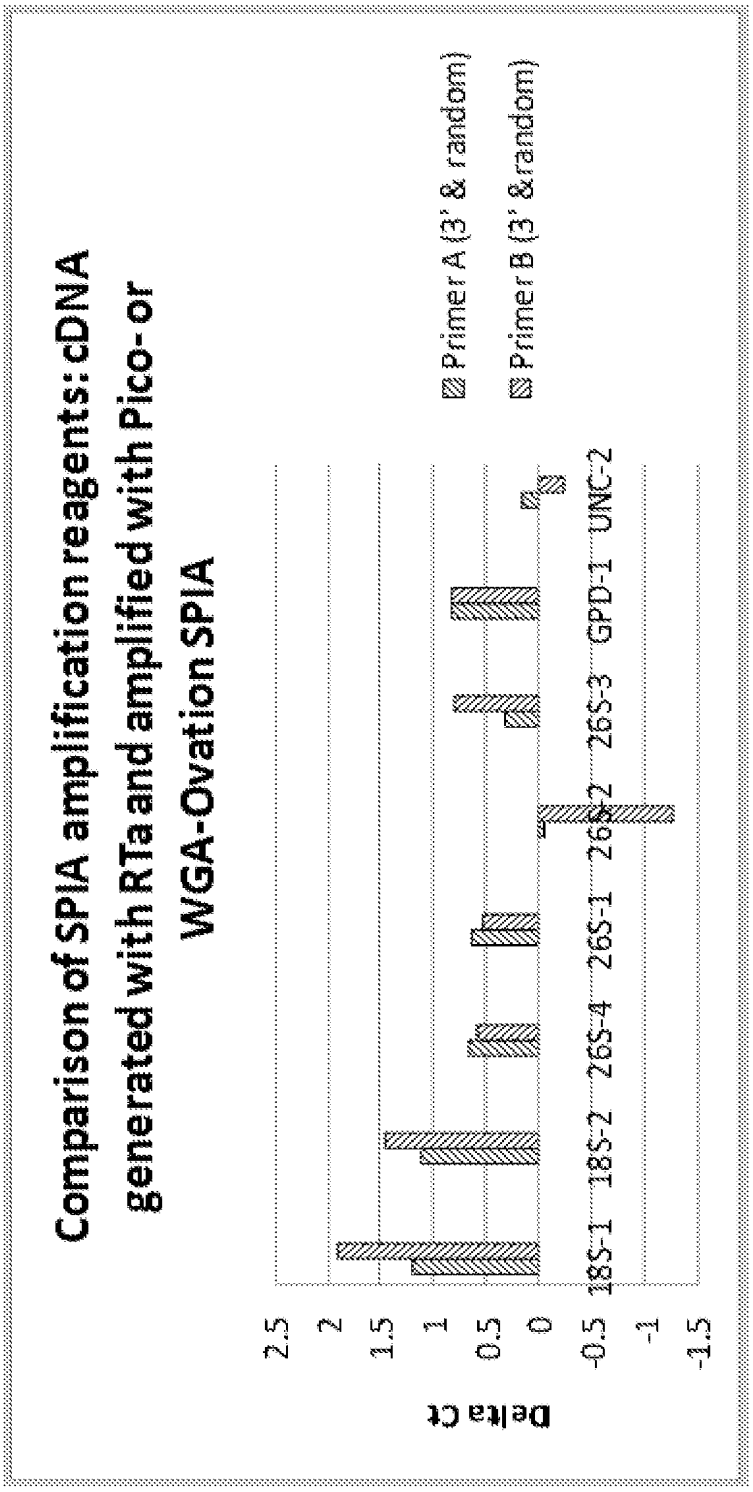
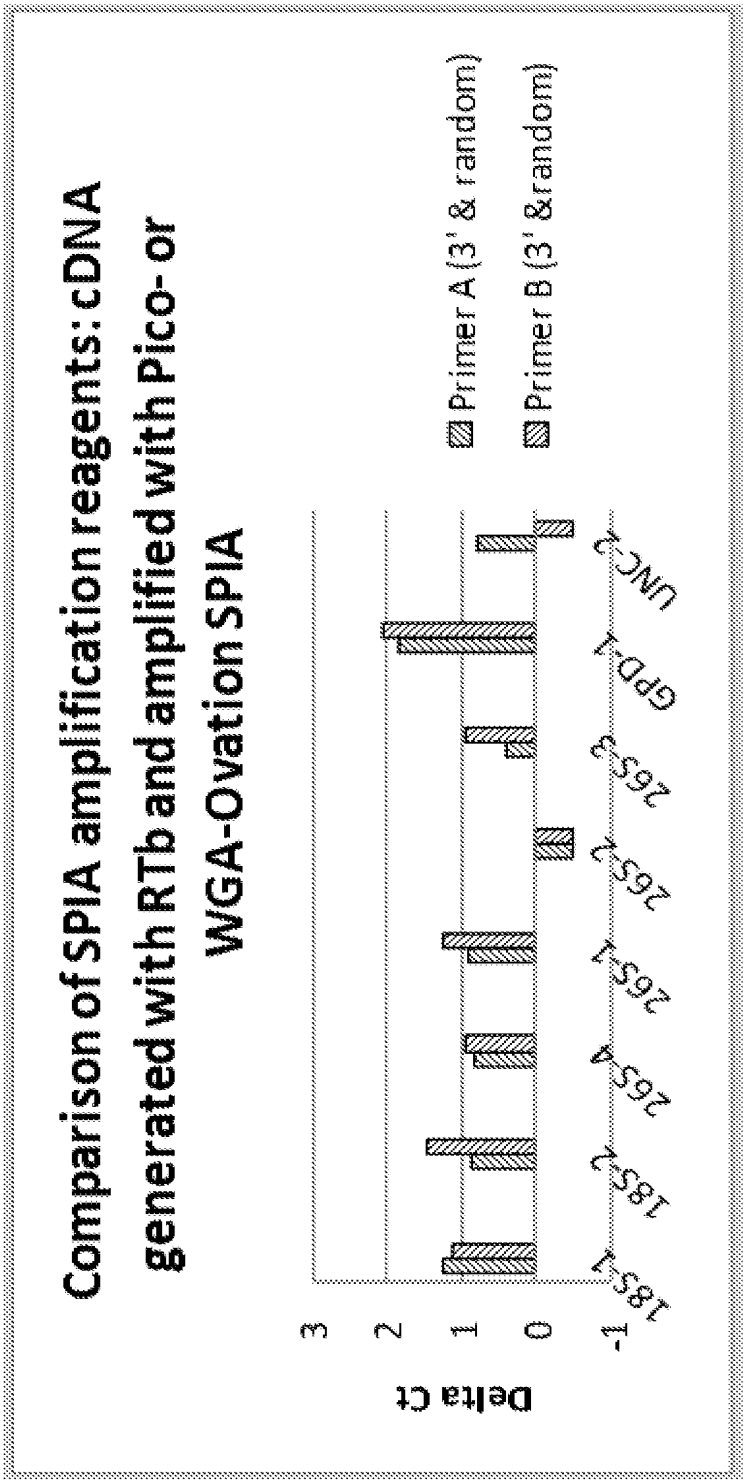

Comparison of SPIA amplification reagents: cDNA generated with RTa and amplified with Pico- or WGA-Ovation SPIA
Primer A (3' & random)
Primer B (3' &random)
Delta Ct
2.5
2
1.5
1
0.5
0
-0.5
-1
-1.5
18S-1
18S-2
26S-4
26S-1
26S-2
26S-3
GPD-1
UNC-2

Figure 18:


Comparison of SPIA amplification reagents: cDNA generated with RTb and amplified with Pico- or WGA-Ovation SPIA
Primer A (3' & random)
Primer B (3' &random)
Delta Ct
3
2
1
0
-1
18S-1
18S-2
26S-1
26S-2
26S-3
GPD-1
UNC-2

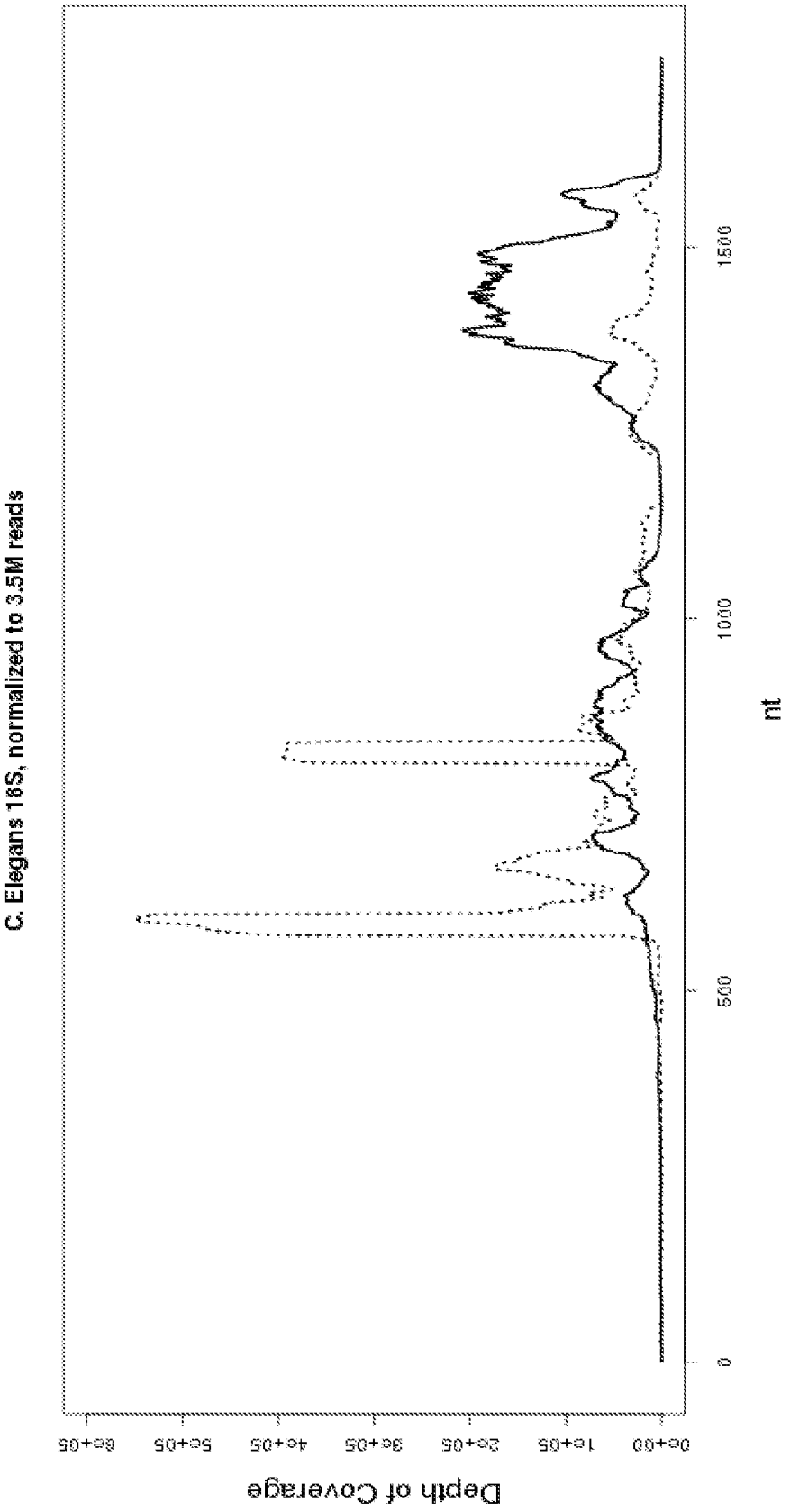
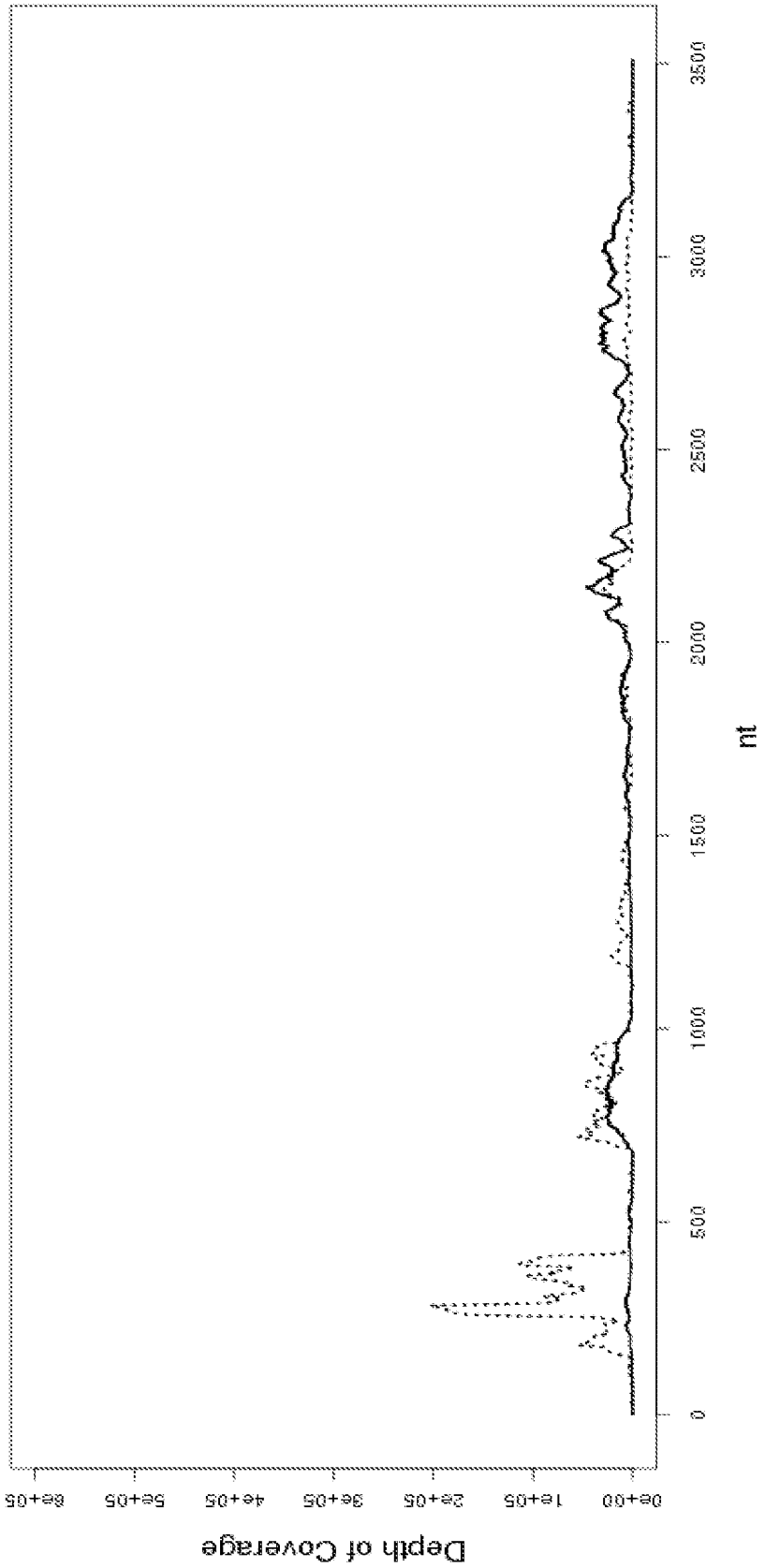

C. Elegans 16S, normalized to 3.5M reads
Depth of Coverage
0e+00
1e+05
2e+05
3e+05
4e+05
5e+05
6e+05
0
500
1000
1500
nt

Figure 19:

Figure 20:


C. Elegans 23S, normalized to 3.5M reads
Depth of Coverage
0e+00
1e+05
2e+05
3e+05
4e+05
5e+05
6e+05
0
500
1000
1500
2000
2500
3000
3500
nt

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 10/48541

A. CLASSIFICATION OF SUBJECT MATTER
IPC(8) - C12P 19/34 (2010.01)
USPC - 435/91.2; 435/91.51
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) - C12P 19/34 (2010.01)
USPC - 435/91.2; 435/91.51

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
(Text Search)

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 Scholar and PubMed.
Search Terms: transcriptome, eliminat$, reduc$, decreas$, rRNA, ribosomal RNA, reverse transcriptase, RT, chimera primer, chimeric primer, primer, tailed primer, SPIA, reverse, transcrip$, sequenced, sequencing, amplif$, PCR, identif$.

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 2008/0187969 A1 (CASTLE et al.) 07 August 2008 (07.08.2008) table 2; para [0007], [0008], [0012], [0078], [0029], [0030], [0077]-[0086], [0095]-[0099], [0089]-[0091], [0095]-[0099].	1-3, 6, 9-19, 21-22 and 24-33 ---------- 4-5, 7-8, 20, 23 and 34-49
Y	US 2007/0031865 A1 (WILLOUGHBY) 08 February 2007 (08.02.2007) para [0020], [0031], [0048], [0049], [0053], [0053], [0055].	4-5
Y	US 2008/0085515 A1 (REMACLE et al.) 10 April 2008 (10.04.2008) para [0277]-[0280].	7-8, 20, 23, 36 and 39
Y	US 2006/0281092 A1 (WILLE et al.) 14 December 2006 (14.12.2006) para [0103].	34-49

☐ 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
22 November 2010 (22.11.2010)	06 DEC 2010
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)

Espacenet

Bibliographic data: WO2013059746 (A1) — 2013-04-25

COMPOSITIONS AND METHODS FOR DIRECTIONAL NUCLEIC ACID AMPLIFICATION AND SEQUENCING

Inventor(s): ARMOUR CHRIS [US] ± (ARMOUR, CHRIS)

Applicant(s): NUGEN TECHNOLOGIES INC [US]; ARMOUR CHRIS [US] ± (NUGEN TECHNOLOGIES, INC, ; ARMOUR, CHRIS)

Classification: - international: *C40B50/06*
- cooperative: **C12N15/1017**; **C12N15/1093**; **C12N15/1096**; **C12N15/66**; **C12Q1/6806**; **C12Q1/6848**; **C12Q1/6869**; **C12Q1/6874** more

Application number: WO2012US61218 20121019 Global Dossier

Priority number (s): US201161549162P 20111019

Also published as: CA2852949 (A1) CN104080958 (A) EP2769007 (A1) EP2769007 (A4) EP2769007 (B1) more

Abstract of WO2013059746 (A1)

The invention provides methods and compositions for directional nucleic acid amplification and sequencing. The invention further provides methods and compositions for the construction of directional eDNA libraries. The method may comprise: a) reverse transcribing a RNA sample to generate a first strand eDNA; b) generating a second strand eDNA from the rrst strand eDNA, wherein at least one of the four dNTPs dATP, dCTP, dGTP or dTTP is replaced by a modified dNTP during second strand synthesis and incorporated into the second strand, thereby generating a double-stranded eDNA; c) performing end repair on the double-stranded eDNA; d) ligating adaptors to the double-stranded eDIMA, wherein only one of the adaptors has the modified dNTP incorporated into a ligation strand of the adaptor; e) performing gap repair; and f) selectively cleaving the second strand and the ligation strand of the adaptor that has the modified dNTP by a suitable cleavage agent.

IDT Exhibit 1020.5686

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
25 April 2013 (25.04.2013)

WIPO | PCT

(10) International Publication Number
WO 2013/059746 A1

(51) International Patent Classification:
C40B 50/06 (2006.01)

(21) International Application Number:
PCT/US2012/061218

(22) International Filing Date:
19 October 2012 (19.10.2012)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/549,162 19 October 2011 (19.10.2011) US

(71) Applicant *(for all designated States except US)*: **NUGEN TECHNOLOGIES, INC.** [US/US]; 821 Industrial Road, San Carlos, CA 95050 (US).

(72) Inventor; and
(71) Applicant *(for US only)*: **ARMOUR, Chris** [US/US]; 13531 104th Place Ne, Kirkland, WA 98034 (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, 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).

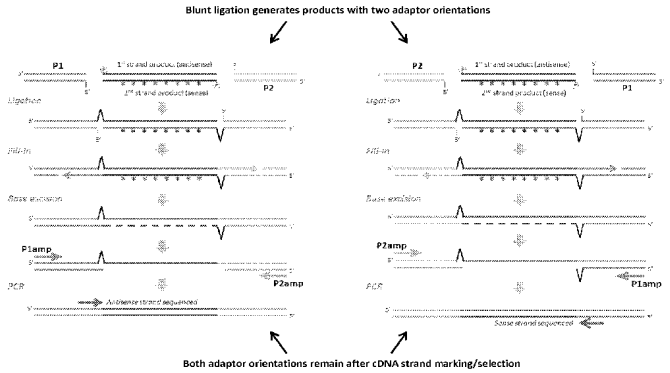
Published:

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

— *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))*

(54) **Title**: COMPOSITIONS AND METHODS FOR DIRECTIONAL NUCLEIC ACID AMPLIFICATION AND SEQUENCING

Figure 1. Conventional system: Strand marking of cDNA inserts is not sufficient for directional library construction.



The asterisks in the diagram (*) indicate dUTP incorporation.

(57) **Abstract**: The invention provides methods and compositions for directional nucleic acid amplification and sequencing. The invention further provides methods and compositions for the construction of directional eDNA libraries. The method may comprise: a) reverse transcribing a RNA sample to generate a first strand eDNA; b) generating a second strand eDNA from the rrrst strand eDNA, wherein at least one of the four dNTPs dATP, dCTP, dGTP or dTTP is replaced by a modified dNTP during second strand synthesis and incorporated into the second strand, thereby generating a double-stranded eDNA; c) performing end repair on the double-stranded eDNA; d) Iigating adaptors to the double-stranded eDIMA, wherein only one of the adaptors has the modified dNTP incorporated into a ligation strand of the adaptor; e) performing gap repair; and f) selectively cleaving the second strand and the ligation strand of the adaptor that has the modified dNTP by a suitable cleavage agent.

WO 2013/059746 A1

IDT Exhibit 1020.5687

COMPOSITIONS AND METHODS FOR DIRECTIONAL NUCLEIC ACID AMPLIFICATION AND SEQUENCING

CROSS-REFERENCE

[0001] This application claims the benefit of U.S. Provisional Application No. 61/549,162, filed Oct. 19, 2011, which application is incorporated herein by reference.

BACKGROUND OF THE INVENTION

[0002] Recent advances in high-throughput, next generation sequencing technologies have enabled whole genome sequencing and new approaches to functional genomics, including comprehensive analysis of any transcriptome. One of these next generation sequencing methods involves direct sequencing of complementary DNA (cDNA) generated from messenger and structural RNAs (RNA-Seq). RNA-Seq provides several key advantages over traditional sequencing methods. First, it allows for high resolution study of all expressed transcripts, annotating the 5' and 3' ends and splice junctions of each transcript. Second, RNA-Seq allows for quantification of the relative number of transcripts in each cell. Third, RNA-Seq provides a way to measure and characterize RNA splicing by measuring the levels of each splice variant. Together, these advancements have provided new insights into individual cell function.

[0003] One drawback of performing standard RNA-Seq is the lack of information on the direction of transcription. Standard cDNA libraries constructed for RNA-Seq consist of randomly primed double-stranded cDNA. Non-directional ligation of adaptors containing universal priming sites prior to sequencing leads to a loss of information as to which strand was present in the original RNA template. Although strand information can be inferred in some cases by subsequent analysis, for example, by using open reading frame (ORF) information in transcripts that encode for a protein, or by assessing splice site information in eukaryotic genomes, direct information on the originating strand is highly desirable. For example, direct information on which strand was present in the original RNA sample is needed to assign the sense strand to a non-coding RNA, and when resolving overlapping transcripts.

[0004] Several methods have recently been developed for strand-specific RNA-Seq. These methods can be divided into two main classes. The first class utilizes distinct adaptors in a known orientation relative to the 5' and 3' end of the RNA transcript. The end result is a cDNA library where the 5' and 3' end of the original RNA are flanked by two distinct adaptors. A disadvantage of this method is that only the ends of the cloned molecules preserve directional information. This can be problematic for strand-specific manipulations of long clones, and can lead to loss of directional information when there is fragmentation.

[0005] The second class of strand-specific RNA-Seq methods marks one strand of either the original RNA (for example, by bisulfite treatment) or the transcribed cDNA (for example, by

IDT Exhibit 1020.5688

incorporation of modified nucleotides), followed by degradation of the unmarked strand. Strand marking by bisulfite treatment of RNA is labor intensive and requires alignment of the sequencing reads to reference genomes that have all the cytosine bases converted to thymines on one of the two strands. The analysis is further complicated due to the fact that base conversion efficiency during bisulfite treatment is imperfect, i.e. less than 100%.

[0006] Strand marking by modification of the second strand of cDNA has become the preferred approach for directional cDNA cloning and sequencing (Levin *et al.,* 2010). However, cDNA second strand marking approaches, such as the one described in WO 2011/003630, are not sufficient to preserve directionality information when using conventional blunt-end ligation and cDNA library construction strategies with duplex adaptors, where two universal sequencing sites are introduced by two separate adaptors. The marking approach described in WO 2011/000360 utilizes a four-step process, consisting of 1) incorporation of a cleavable nucleotide into one strand of the cDNA insert, 2) end repair of the cDNA insert, 3) non-directional ligation of adaptors containing universal sequencing sites and 4) selective hydrolysis of library fragments with undesired adaptor orientation. To preserve directionality information, the method requires that the 5' and 3' ends of the strand selected for amplification are marked differentially, which can be achieved, for example, by ligation of directional (i.e. polarity-specific) adaptors, or by use of a specialized forked adaptor where each strand of a double-stranded polynucleotide is covalently attached to two distinct universal sequencing sites, one sequencing site at each end of the strand. Application of the methodology described in WO 2011/000360 does not result in directional sequencing libraries when using conventional duplex adaptors because the marked strand, i.e. the strand with incorporated cleavable nucleotides, is not differentially labeled at its 5' and 3' ends.

[0007] There is a need for improved methods for directional cDNA sequencing from cDNA libraries constructed with conventional duplex adaptors. The invention described herein fulfills this need.

SUMMARY OF THE INVENTION

[0008] The present invention provides novel methods, compositions, and kits for construction of directional cDNA libraries and directional cDNA sequencing. Specifically, an important aspect of this invention is the methods and compositions that allow for directional cDNA cloning and strand retention using duplex adaptors and blunt-end ligation, thereby generating ligation products with two adaptor orientations. In one aspect, the invention provides a method for cloning cDNA while retaining the directionality and strand information of the original RNA sample. In some embodiments, the method comprises: a) reverse transcribing a RNA sample to generate a first strand cDNA; b) generating a second strand cDNA from the first strand cDNA, wherein at least one of the four dNTPs dATP, dCTP, dGTP or dTTP is replaced by a modified dNTP during second strand synthesis and incorporated into the second strand, thereby generating a double-stranded cDNA; c) performing end repair on the double-stranded cDNA; d) ligating adaptors to the double-stranded cDNA, wherein only

IDT Exhibit 1020.5689

one of the adaptors has the modified dNTP incorporated into a ligation strand of the adaptor; e) performing gap repair; and f) selectively cleaving the second strand and the ligation strand of the adaptor that has the modified dNTP by a suitable cleavage agent, thereby generating a directional cDNA library comprising the first strand cDNA In afurther aspect, the method optionally comprises fragmenting the double stranded cDNA prior to performing end repair on the double-stranded cDNA. In a specific embodiment, the method further comprises amplifying the remaining cDNA strand or the cDNA strand that does not comprise the modified nucleotide, thereby generating amplified products. In another specific embodiment, the method further comprises sequencing the remaining cDNA strand or the amplified products.

[0009] In another aspect, the invention provides for a method for selective removal of cDNA constructs in the undesired orientation.

[0010] In yet another aspect, the invention provides a method for whole transcriptome directional sequencing, comprising: a) reverse transcribing a RNA sample to generate a first strand cDNA; b) generating a second strand cDNA from the first strand cDNA, wherein at least one of the four dNTPs dATP, dCTP, dGTP or dTTP is replaced by a modified dNTP during second strand synthesis and incorporated into the second strand, thereby generating a double-stranded cDNA; c) performing end repair on the double-stranded cDNA; d) ligating adaptors to the double-stranded cDNA, wherein only one of the adaptors has the modified dNTP incorporated into a ligation strand of the adaptor; e) performing gap repair; f) selectively cleaving the second strand and the ligation strand of the adaptor that has the modified dNTP by a suitable cleavage agent, thereby generating a directional cDNA library comprising the first strand cDNA and h) amplification and/or sequencing of the directional cDNA library. In a further aspect, the method optionally comprises fragmenting the double stranded cDNA prior to performing end repair on the double-stranded cDNA.

[0011] In one aspect of any one of the foregoing aspects, the present invention provides for cleaving a base portion of the modified nucleotide thereby forming an abasic site. In a preferred embodiment, the modified nucleotide comprises dUTP. In some cases the cleavage agent comprises a glycosylase. In a preferred embodiment, the glycosylase comprises UNG. In some cases, the cleavage agent comprises a glycosylase and an endonuclease. In some cases, the endonuclease comprises an apurinic/apyrimidinic endonuclease (APE). In some cases, the cleavage agent comprises a glycosylase and a APE. In some cases, the cleavage agent comprises a UNG and a APE. In some cases, the cleavage agent comprises a glycosylase and a polyamine. In a preferred embodiment, the polyamine is N,N-dimethylethylenediamine (DMED). In some cases, the cleavage agent comprises a glycosylase and DMED. In some cases, the cleavage agent comprises a UNG and DMED.

[0012] In another aspect of any one of the foregoing aspects, the method further comprises creating nicks in a phosphodiester backbone at an abasic site with an enzyme, chemical agent, and/or heat following removal of a base portion of the modified nucleotide. In some cases, cleaving a phosphodiester backbone at an abasic site following removal of a base portion of the modified

IDT Exhibit 1020.5690

nucleotide comprises using an enzyme. In a preferred embodiment, the enzyme is an endonuclease. In some cases, the endonuclease comprises an apurinic/apyrimidinic endonuclease (APE). In some cases, creating nicks at an abasic site following removal of a base portion of the modified nucleotide comprises using a chemical agent. In some cases, the chemical agent comprises a primary amine or a polyamine. In a preferred embodiment, the polyamine is N,N-dimethylethylenediamine (DMED).

[0013] In another aspect of any of the foregoing aspects, the method further comprises cleaving the RNA sample following reverse transcription of the RNA sample. In some cases, cleaving the RNA sample comprises exposing the RNA sample to an RNase. In a preferred embodiment, the RNase is RNase H. In some cases, cleaving the RNA sample comprises exposing the RNA sample to heat or chemical treatment or a combination thereof.

[0014] In another aspect of any of the foregoing aspects, the method further comprises reducing or depleting non-desired nucleic acid sequences. In some cases, the non-desired nucleic acid is ribosomal RNA (rRNA).

[0015] In another aspect of any of the foregoing aspects, the amplification of the remaining cDNA strand comprises polymerase chain reaction (PCR), strand displacement amplification (SDA), multiple displacement amplification (MDA), rolling circle amplification (RCA), single primer isothermal amplification (SPIA), or ligase chain reaction (LCR). In some cases, the amplification comprises PCR. In some case, the amplification comprises SPIA.

[0016] In another aspect of any of the foregoing aspects, the sequencing of the remaining cDNA strand or amplified products from the remaining cDNA strand comprises next generation sequencing.

[0017] Kits for performing any of the methods described herein are another feature of the invention. Such kits may include reagents, enzymes and platforms for amplification, cloning and sequencing of nucleic acids. In one embodiment, a kit is provided comprising: a) one or more primers; b) a reverse transcription enzyme, c) a glycosylase and d) an adaptor or several adaptors wherein one of the adaptors comprises at least one modified nucleotide in a ligation strand of said adaptor. In another embodiment, the kit further comprises reagents for amplification. In another embodiment, the kit further comprises a polyamine, an APE, or a combination thereof. In another embodiment, the kit further comprises at least one modified nucleotide or dNTP. In some cases, the modified nucleotide comprises dUTP. In some cases, the glycosylase comprises UNG. In yet another embodiment, the kit further comprises reagents for sequencing. A kit will preferably include instructions for employing the kit components as well as the use of any other reagent not included in the kit.

INCORPORATION BY REFERENCE

[0018] 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.

IDT Exhibit 1020.5691

BRIEF DESCRIPTION OF THE DRAWINGS

[0019] 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 description that sets forth illustrative embodiments, in which the principles of the invention are utilized, and the accompanying drawings of which:

[0020] Figure 1 depicts construction of a non-directional cDNA library with conventional duplex adaptors (i.e. where ligation products have two adaptor orientations) and strand marking.

[0021] Figure 2 depicts construction of directional (strand-specific) cDNA libraries with conventional duplex adaptors using the methods of the invention.

[0022] Figure 3 depicts a table summarizing strand retention efficiency data using the methods of the invention.

[0023] Figure 4 depicts a flow diagram illustrating the steps for generating a directional cDNA library using the methods of the invention.

DETAILED DESCRIPTION OF THE INVENTION

General

[0024] Reference will now be made in detail to exemplary embodiments of the invention. While the disclosed methods and compositions will be described in conjunction with the exemplary embodiments, it will be understood that these exemplary embodiments are not intended to limit the invention. On the contrary, the invention is intended to encompass alternatives, modifications and equivalents, which may be included in the spirit and scope of the invention.

[0025] In one embodiment, the present invention provides methods and compositions for construction of directional cDNA libraries. The methods described herein enable directional cDNA cloning and strand retention using conventional duplex adaptors and blunt-end ligation. The methods further enable generation of strand-specific cDNA which can be further amplified using a variety of amplification methods. In another embodiment, the present invention provides methods for whole transcriptome directional sequencing. In yet another embodiment, the present invention provides methods and compositions for generation of a directional, rRNA-depleted cDNA library.

[0026] One aspect involves a method for generation of a directional cDNA library. The first step in the method entails use of an RNA sample or RNA template from which a first strand cDNA can be generated through reverse transcription. The RNA sample can be derived from any number of sources known in the art including, but not limited to, messenger RNA (mRNA) or ribosomal RNA (rRNA) in either purified or unpurified forms and reverse transcription can be performed using any number of RNA dependent DNA polymerases known in the art. In one embodiment, the RNA template can be derived from DNA including, but not limited to, genomic DNA wherein the DNA is converted to RNA using methods known in the art including, but not limited to, transcription. In a preferred embodiment, as exemplified in FIG. 4, the RNA can be poly A+ RNA. Reverse transcription of the RNA sample can be performed using primers comprising sequence complementary to known

IDT Exhibit 1020.5692

sequences or comprising random sequences. In one embodiment, the primers used in the methods described herein can be composite primers comprising both DNA and RNA. In a preferred embodiment, the RNA sample can be reverse transcribed using random hexamer primers.

[0027] The second step in the method described herein for the generation of a directional cDNA library entails generating a second strand cDNA from the first strand cDNA in order to form a double stranded cDNA. Second strand synthesis can be performed in the presence of a modified dNTP. In a preferred embodiment, second strand synthesis can be performed in the presence of dATP, dCTP, dGTP and dUTP in place of dTTP. Second strand synthesis in the presence of dUTP causes incorporation of at least one dUTP in the strand of the second strand cDNA. The dUTP in the second strand cDNA serves to mark the second strand since in this context dUTP is a modified or non-canonical dNTP. Second strand synthesis can be performed using any number of second strand synthesis protocols known in the art including, but not limited to, those that utilize RNase H mediated nick translation in combination with a DNA dependent DNA polymerase such as DNA polymerase I (not Klenow Fragment). Second strand synthesis can also be performed using commercially available kits such as New England Biolabs NEBNext Second Strand Synthesis Module The second strand cDNA product produced during second strand synthesis can also be referred to as the sense strand product since the sequence of the second strand cDNA comprises the sequence found in the template RNA, while the first strand cDNA can be also be referred to as the antisense strand product. In another embodiment, second strand synthesis can be performed following removal of the RNA template after first strand synthesis. Removal of the RNA template can be achieved using enzymes, heat denaturation, or chemical denaturation. Enzymatic mediated removal of the RNA template can be performed with an RNase, preferably RNase H, or a combination of enzymes, such as RNase H and RNase1 As a further aspect to this embodiment, second strand synthesis can be performed using a primer comprising sequence that is complementary to sequence present in the first strand product in conjunction with the use of a DNA dependent DNA polymerase.

[0028] In one embodiment, second strand synthesis can be followed by end repair of the double stranded cDNA generated following second strand synthesis. End repair can include the generation of blunt ends, non-blunt ends (i.e sticky or cohesive ends), or single base overhangs such as the addition of a single dA nucleotide to the 3′-end of the double-stranded DNA product, by a polymerase lacking 3′-exonuclease activity. In a preferred embodiment, end repair can be performed on the double stranded cDNA to produce blunt ends wherein the double stranded cDNA contains 5' phosphates and 3' hydroxyls. End repair can be performed using any number of enzymes and/or methods known in the art including, but not limited to, commercially available kits such as the Encore ™ Ultra Low Input NGS Library System I.

[0029] In one embodiment, end repair can be performed after the double-stranded cDNA has been fragmented. Fragmentation of the double-stranded products can be achieved through methods known in the art. Fragmentation can be through physical fragmentation methods and/or enzymatic

IDT Exhibit 1020.5693

fragmentation methods. Physical fragmentation methods can include nebulization, sonication, and/or hydrodynamic shearing. In a preferred embodiment, the fragmentation of the double-stranded cDNA is performed by sonication. Reagents for carrying out enzymatic fragmentation reactions are commercially available (e.g, from New England Biolabs).

[0030] Following end repair of the double stranded cDNA, the methods described herein for generating directional cDNA libraries involve ligating adaptors to the double-stranded cDNA. The adaptors can be any type of adaptor known in the art including, but not limited to, conventional duplex or double stranded adaptors. In a preferred embodiment, the adaptors can be double stranded DNA adaptors. In an embodiment, the adaptors can be oligonucleotides of known sequence and, thus, allow generation and/or use of sequence specific primers for amplification and/or sequencing of any polynucleotides to which the adaptor(s) is appended or attached. Preferably, the adaptors can be any adaptors that can be marked and selected for by methods known in the art. In a preferred embodiment, the adaptors can be marked via incorporation of at least one modified dNTP. In a preferred embodiment, one and only one of the adaptors comprises a modified dNTP while the other or any other adaptor(s) does not comprise the modified dNTP. In a preferred embodiment one and only one of the adaptors comprises a modified dNTP in a ligation strand of said adaptor, while the other or any other adaptor(s) does not comprise the modified dNTP in a ligation strand of said adaptor(s). In one embodiment, the modified dNTP is dUTP. In a preferred embodiment, the adaptors can be appended to the double-stranded product in multiple orientations. In a preferred embodiment, the methods described herein can involve the use of two conventional duplex adaptors comprising double stranded DNA of known sequence that are blunt ended and can bind to the double stranded product in one of two orientations, wherein one of the adaptors comprises a modified dNTP incorporated into the ligation strand while the other adaptor does not contain the modified dNTP in the ligation strand. In a preferred embodiment, the modified dNTP is dUTP.

[0031] According to the methods described herein, the adaptors can be ligated to the double-stranded cDNA by blunt end ligation in either of two orientations. One of the adaptors comprises a modified dNTP (preferably dUTP) incorporated into the ligation strand while the other adaptor does not comprise the modified dNTP (preferably dUTP) incorporated into the ligation strand. In one embodiment, the ligation of the adaptors to the double stranded cDNA creates a gap between the non-ligation strand of either of the adaptors and a strand of the double-stranded cDNA, whereby the non-ligation strand of the respective adaptor is not bound to a strand of the double stranded cDNA. As such, a gap repair or fill-in reaction can be performed using any number of methods known in the art including, but not limited to, the use of a DNA dependent DNA polymerase with weak or no strand displacement activity.

[0032] Following ligation of the adaptors and, optionally, gap repair, a double stranded cDNA/adaptor complex is generated. The complex can then be subjected to strand selection. Strand selection can entail base excision of the modified dNTP that is incorporated into the second strand of

IDT Exhibit 1020.5694

the double stranded cDNA and the ligation strand of one and only one of the adaptors ligated to the double stranded cDNA. Base excision of the modified dNTP can be performed using an enzyme, chemical agent, and/or heat and creates an abasic site wherever the modified dNTP is incorporated in a nucleotide sequence. In addition to base excision, the methods of the present invention can also entail cleavage of the phoshodiester bond at the abasic site. The phosphodiester bond can also be referred to as the phosphodiester backbone or DNA backbone. Cleavage of the DNA backbone can be performed using any number of agents including anenzyme, chemical agent, heat, or any combination thereof.

[0033] Base excision and/or cleavage of the DNA backbone leads to the cleavage or removal of the marked second strand that comprises a modified dNTP as well as the ligation strand of the one adaptor that comprises the modified dNTP, while the unmarked strand and the ligation strand of the adaptor that does not contain the modified dNTP remain intact. In this instance, the remaining strand of the double strand cDNA following base excision and/or DNA backbone cleavage is the first strand cDNA or the antisense strand product. In addition, base excision and/or DNA backbone cleavage also cleaves or removes the ligation strand of the one adaptor that comprises the modified dNTP, regardless of which strand of the double stranded cDNA said ligation strand is ligated to. Amplification of the remaining strand can be preferentially performed using a primer whose sequence is complementary to the ligation strand of the adaptor that does not comprise the modified dNTP. Amplification using a primer whose sequence is complementary to the ligation strand of the adaptor that comprises the modified dNTP produces no product since that ligation strand has been cleaved and/or removed following base excision and/or cleavage of the modified dNTP. Amplification of the remaining strand can be performed using any number of amplification techniques known in the art including, but not limited to, polymerase chain reaction (PCR). Following amplification of the remaining strand, sequencing of the amplified products can be performed using primers complementary to sequences present in the ligation strand of the adaptor that does not comprise the modified dNTP, which ensures sequencing of only the unmarked first strand cDNA or the antisense strand product that remained after strand selection. Sequencing can be performed on the remaining cDNA directly and/or on the products resulting from amplification of the remaining strand. Sequencing can be performed using any of sequencing methods known in the art including, but not limited to, next generation sequencing methods. The methods of the present invention as described above result in the generation of directional cDNA libraries that comprises cDNAs of the antisense orientation or first strand cDNA due to the marking and cleavage or removal of the sense strand product (second strand cDNA).

[0034] In another embodiment, the cleaved or degraded marked strands can be removed prior to amplification and/or sequencing of the remaining strand. In a preferred embodiment, the cleaved second strand and cleaved ligation strand of the one adaptor that comprises the modified dNTP can be removed prior to amplification and/or sequencing of the remaining first strand cDNA or antisense

IDT Exhibit 1020.5695

product. In yet another preferred embodiment, the remaining first strand cDNA (antisense product) can be purified prior to amplification and/or sequencing. Purification of the remaining strand can be performed using methods known in the art for purification of cDNA such as kits commercially available from Qiagen and/or Roche.

[0035] In an alternative embodiment, the methods described herein can be used to generate a directional cDNA library that comprises cDNAs in the sense orientation or second strand cDNA. In this embodiment, the methods described herein can be performed as described above with the exceptions that the first strand synthesis from the RNA template, as opposed to the second strand synthesis as described above, can be performed in the presence of a modified dNTP, and second strand synthesis is performed in the presence of unmodified dNTPs or classic dNTPs.In one aspect of this embodiment, the dNTPs including any and all modified dNTPs used during first strand synthesis can be removed, washed away, or replaced with unmodified dNTPs prior to second strand synthesis. As a further aspect of this embodiment, unmodified dNTPs can be used during second strand synthesis. In a preferred embodiment, the modified dNTP comprises dUTP. The antisense strand product (first strand cDNA) marked with a modified dNTP and ligated to the one and only one adaptor that comprises the modified dNTP in the ligation strand of said adaptor can be selectively cleaved or removed. As such, the remaining strand and ligated adaptor available for downstream amplification and/or sequencing comprises the sense strand product ligated to the adaptor that does not comprise the modified dNTP in the ligation strand of said adaptor.

[0036] A schematic of a preferred embodiment of the methods described herein for generating and sequencing a directional strand specific cDNA.library is illustrated in FIG. 2. Overall, the method illustrated in FIG. 2 allows determination of the strand orientation of a template RNA used to generate cDNA with improved efficiency over conventional methods as illustrated in FIG 1. The methods illustrated in FIGs. 1 and 2 both use strand marking of the cDNA and blunt end ligation of conventional duplex adaptors to the cDNA generated from template RNA as means for determining strand orientation. In both FIGs. 1 and 2, the method involves blunt end ligating double-stranded duplex adaptors (P1/P2 in FIGs. 1 and 2) to a double stranded cDNA complex formed from an RNA sample wherein the second strand product, which is also referred to as the sense strand product since it is complementary to and of the same strand orientation as the RNA template, is marked via incorporation of dUTP during second strand synthesis. In a preferred embodiment as illustrated in both FIGs. 1 and 2, the duplex adaptors do not contain free 5' phosphate groups. As such, both adaptors (P1/P2) contain a strand (the ligation strand) that ligates with the free 5'phosphate on the double-stranded cDNA and a strand that does not ligate (non-ligation strand) to the double-stranded cDNA.. Ligation can be facillitated through the use of enzymes (i.e. T4 DNA ligase) and methods known in the art, including, but not limited to, commercially available kits such as the Encore ^{TM} Ultra Low Input NGS Library System. In a preferred embodiment of the present invention as depicted in FIG. 2, the ligation strand of one and only one of the adaptors (P2) is marked via incorporation of

IDT Exhibit 1020.5696

dUTP. As depicted in FIGs. 1 and 2, ligation of the duplex adaptors can occur in one of two orientations. In the schematic on the left side of FIGs. 1 and 2, the ligation strand of the P2 adaptor is ligated to the marked sense strand (second strand product). In the schematic on the right side of FIG. 2, the ligation strand of the P2 adaptor is ligated to the unmarked antisense strand (first strand product).

[0037] In the methods illustrated in both FIGs. 1 and 2, the duplex adaptors are unphosphorylated and thus do not contain free 5' phosphate groups. As such, both adaptors (P1/P2) contain a strand (the ligation strand) that will ligate with the free 5'phosphate on the double-stranded cDNA and a strand that does not ligate (non-ligation strand) to the double-stranded cDNA and thus leaves a gap. As such, in either orientation, the double-stranded cDNA containing the ligated adaptors is subjected to gap or fill-in repair (preferably with a DNA dependent DNA polymerase such as *Taq* DNA polymerase) in order to fill-in the gap through DNA dependent DNA polymerase mediated synthesis of the sequence of the non-ligation strand of the duplex adaptors using the respective ligation strand as template.

[0038] In both FIGs. 1 and 2, gap repair is followed by base excision via treatment with a cleavage agent, which can be an enzyme. In one embodiment as shown in FIGs. 1 and 2, base excision can be performed with an enzyme such as UNG. In a preferred embodiment, base excision can be followed by cleavage of the phosphodiester or DNA backbone using an enzyme, chemical agent, and/or heat at the site where the base was cleaved. In FIG. 1, base excision leads to the cleavage of the marked sense strand product, while both adaptors remain intact. In FIG. 2, base excision leads to the cleavage of both the marked sense strand product and the one adaptor that has dUTP incorporated into the ligation strand of said adaptor, while the adaptor that does not have dUTP incorporated into the ligation strand remains intact. In both FIGs. 1 and 2, the marked sense strand product or second strand product can be cleaved and thus only the antisense strand product or first strand product remains following base excision.

[0039] In contrast to FIG. 1 wherein both adaptor orientations remain intact following base excision, the schematic on the left side of FIG. 2 shows that the marked sense strand product can be cleaved along with the marked ligation strand of the P2 adaptor that is ligated to the sense strand product. As such, only the antisense strand product and the ligation strand of the adaptor ligated to the antisense strand (P1 in FIG. 2) remain intact and available for downstream processing. In a preferred embodiment downstream processing entails amplification of the remaining cDNA strand or antisense strand product. In contrast to FIG. 1 wherein amplification of the antisense strand product can be performed using primers complementary to sequence contained in either the ligation strand of the P1 adaptor (P1amp) or P2 adaptor (P2amp), FIG. 2 shows that amplification of the antisense strand product can only be performed using primers complementary to sequence contained in the ligation strand of the P1 adaptor (P1amp).. In a preferred embodiment, downstream processing can also entail sequencing of the antisense strand product (first strand product) and/or the amplified products. In FIG. 1, downstream sequencing using primers complementary to sequence in the ligation strand of the P1

IDT Exhibit 1020.5697

adaptor will sequence either the sense or antisense strands relative to the RNA template. In FIG. 2, downstream sequencing using primers complementary to sequence in the ligation strand of the P1 adaptor will sequence only the antisense strand relative to the RNA template.

[0040] FIG. 4 illustrates a flow chart depicting one embodiment of the method for generating a directional strand specific cDNA library. The method involves the steps of generating first strand cDNA by performing random-primed reverse transcription on polyA+ RNA; generating 2nd strand cDNA using a DNA-dependent DNA polymerase using dATP, dCTP, dGTP, and dUTP in place of dTTP; fragmenting the double-stranded cDNA using sonication; end-repairing the purified fragmented double-stranded cDNA to generate blunt ends; ligating duplex adaptors wherein one of the ligation strands of one of the duplex adaptors is marked via incorporation of dUTP; nick repairing the ligation products to generate double-stranded cDNA containing the ligated adaptors; performing strand selection of the purified double-stranded cDNA containing the ligated adaptors using an enzyme and/or chemical agent; amplifying the remaining cDNA strand using PCR.

[0041] In an aspect to any of the embodiments above, the directional cDNA libraries created by the methods described herein can be depleted of non-desired nucleic acid sequences. In one embodiment, the non-desired nucleic acid comprises RNA. In a preferred embodiment, the non-desired nucleic acid comprises ribosomal or rRNA. Removal or depletion of rRNA from the directional cDNA libraries generated by the methods of the present invention can be performed by any of the methods known in the art including, but not limited to, removal of rRNA from the starting population, differential priming using oligo dT primers (i.e. priming polyadenylated transcripts only), and/or differential priming where primers complementary to rRNA sequences are specifically eliminated (or under-represented) in a primer pool (Not-So-Random or NSR primer approach).

[0042] In general, the methods described herein can be used to create nucleic acid libraries preferentially populated with nucleic acids of specific strand orientations relative to the nucleic acid template from which the library was generated. The nucleic acid libraries generated by the methods described herein can be used to ascertain the directionality and strand orientation of the nucleic acid template. In one embodiment, the nucleic acid template can be an RNA template and the nucleic acid library can be a cDNA library. In a preferred embodiment, the RNA template can be non-rRNA. In yet another preferred embodiment, the RNA template can be rRNA. In an aspect of the methods described herein, the cDNA library can be a directional cDNA library that retains the directionality and strand information pertaining to the original RNA template or sample, that is to say, the directional library of the methods of the invention represents products generated from first strand cDNA, or reverse transcription of the template RNA, or the second strand cDNA (a copy of the first strand cDNA). The methods of the invention provide means for exclusive retention of either first strand cDNA products or second strand cDNA products, thus enabling assigning the directionality of

IDT Exhibit 1020.5698