transcription from the genomic DNA. The directionality of transcription is inferred from the knowledge of which of the cDNA strand (first or second strand) is represented in the sequence information. The directionality and strand information of the RNA template can refer to the strand of genomic DNA from which the RNA template was derived or transcribed. As a further aspect of the methods described herein, the directional cDNA library can be used to determine the directionality of transcription by comparing the sequence of cDNAs in the directional cDNA library to the RNA template and/or genomic DNA. Methods of comparing nucleotide sequences are known in the art and can include well known nucleotide sequence alignment programs or algorithms such as the BLAST algorithm from NCBI.

[0043] Based on the methods described herein, the retention of the directionality and strand information of the RNA template can be determined with greater than 50% efficiency. The efficiency of retention of directionality and strand orientation using the methods described herein can be > 50%, >55%, > 60%, >65%, >70%, >75%, > 80%, >85%, > 90%, or >95%. The efficiency of retention of directionality and strand orientation can be > 99%.The methods described herein can be used to generate directional cDNA libraries wherein greater than 50% of the cDNAs in the cDNA library comprise a specific strand orientation. The retention of a specific strand orientation using the methods described herein can be > 50%, >55%, > 60%, >65%, >70%, >75%, > 80%, >85%, > 90%, or >95%. The retention of specific strand orientation of cDNAs in the directional cDNA library can be > 99%.As illustrated in FIG. 3, the methods of the present invention were used to generate directional cDNA libraries designed to retain the antisense strand product or first strand cDNA. As shown in FIG. 3, >97% of the sequence reads that mapped to the coding exons of human mRNAs from which the cDNAs were derived where in the antisense orientation.

[0044] Unless otherwise specified, terms and symbols of genetics, molecular biology, biochemistry and nucleic acid used herein follow those of standard treatises and texts in the field, e.g. Kornberg and Baker, *DNA Replication*, Second Edition (W.H. Freeman, New York, 1992); Lehninger, *Biochemistry*, Second Edition (Worth Publishers, New York, 1975); Strachan and Read, Human *Molecular Genetics*, Second Edition (Wiley-Liss, New York, 1999); Eckstein, editor, *Oligonucleotides and Analogs: A Practical Approach* (Oxford University Press, New York, 1991); Gait, editor, *Oligonucleotide Synthesis: A Practical Approach* (IRL Press, Oxford, 1984); and the like.

RNA Sample

[0045] The RNA sample of the present invention can be double-stranded, partially double-stranded; and single-stranded nucleic acids from any source including, but not limited to, synthetic or semisynthetic nucleic acids 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,

IDT Exhibit 1020.5699

humans, and fragments thereof. Exemplary 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. Exemplary RNA samples can be obtained and purified using standard techniques in the art and includes RNAs 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, cell free RNA and fragments thereof. In one embodiment, the RNA sample provided for the methods of the present invention includes a whole transcriptome which can include tRNA, mRNA, rRNA, and non-coding RNA. The non-coding RNA, or ncRNA may include snoRNAs, microRNAs, siRNAs, piRNAs and long nc RNAs. In a preferred embodiment, the RNA sample has the rRNA content reduced or removed using standard techniques in the art. In a most preferred embodiment, the RNA sample is mRNA.

Primers

[0046] The term "primer", as used herein, can refer to a nucleotide sequence, generally with a free 3' hydroxyl group, that is capable of hybridizing with a template (such as one or more target polynucleotides, one or more target DNAs, one or more target RNAs or a primer extension product) and is also 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 [i.e. RNase H] 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 of the primer. A primer may still be hybridizing to a target even though its sequences are not fully complementary to the target.

[0047] The primers of the invention are generally oligonucleotides that are employed in an extension reaction by a polymerase along a polynucleotide template, such as in PCR, SPIA or cDNA synthesis, for example. The oligonucleotide primer can be 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.

[0048] "Complementary", as used herein, can refer 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

IDT Exhibit 1020.5700

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 or 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.

[0049] In some cases, the identity of the investigated target polynucleotide sequence is known, and hybridizable primers can be synthesized precisely according to the antisense sequence of the aforesaid target polynucleotide sequence. In other cases, when the target polynucleotide sequence is unknown, 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 yet 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.

[0050] Tailed primers can be 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 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

IDT Exhibit 1020.5701

sequence can be 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 can be 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.

[0051] Composite primers can be 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 can be 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.

[0052] In other aspects, the composite primer can be 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. 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.

[0053] 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

IDT Exhibit 1020.5702

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 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.

IDT Exhibit 1020.5703

[0054] In one embodiment, the primers of the invention can be 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.

[0055] In certain embodiments of the invention, the polynucleotide template for the polymerase reaction can be a RNA molecule with a poly(A) tail. In such cases, it is preferred that the primers are oligo(dT), oligo(dU) or oligo(U) primers, or, alternatively, composite primers with oligo d(T), oligo(dU) or oligo(U) region on the 3' end of the primer.

[0056] In another embodiment of the invention, the polynucleotide template for the polymerase reaction can be a RNA molecule without a poly(A) tail. In such cases, it is preferred that the primers are random primers, or, alternatively, composite primers with a random sequence that is hybridizable to the RNA in the sample on the 3' end of the primer.

[0057] In certain other embodiments of the invention, the polynucleotide template for the polymerase reaction can be a cDNA molecule. In such cases, it is preferred that the primers are random primers, or, alternatively, composite primers such as the amplication composite primers described herein with a random sequence that is hybridizable to a portion of the cDNA template on the 3' end of the primer. In yet another embodiment, the polynucleotide template for the polymerase reaction is a cDNA molecule whose sequence is known. In such cases, it is preferred that the primers contain sequences complementary to all or a portion of the known sequence of the target polynucleotide or, alternatively, composite primers such as the amplication composite primers described herein with a sequence that is complementary to a portion of the cDNA template whose sequence is known on the 3' end of the primer.

RNA-Dependent DNA Polymerases

[0058] 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 polymerasecan be 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

IDT Exhibit 1020.5704

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 degradation of the starting RNA sample and 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.

DNA-Dependent DNA Polymerases

[0059] DNA-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 DNA-dependent DNA polymerase can be one that is capable of extending a nucleic acid primer along a first strand cDNA in the presence of the RNA template or after selective removal of the RNA template. 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, .phi.29 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 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.

[0060] In one embodiment, the double stranded products generated by the methods of the present invention can be end repaired to produce blunt ends for the adaptor ligation applications of the present

IDT Exhibit 1020.5705

invention. Generation of the blunt ends on the double stranded products may be generated 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 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.

[0061] In another embodiment, the adaptor ligation applications of the present invention can leave a gap between a non-ligation strand of the adaptors and a strand of the double stranded product of the present invention. In these instances, a gap repair or fill-in reaction may be necessary to append the double stranded product with the sequence of the non-ligation strand of the adaptor. Gap repair can be performed with any number of DNA dependent DNA polymerase described herein. In one embodiment, gap repair can be performed with a DNA dependent DNA polymerase with strand displacement activity. In one embodiment, gap repair can be performed using a DNA dependent DNA polymerase with weak or no strand displacement activity. In one embodiment, the ligation strand of the adaptor can serve as the template for the gap repair or fill-in reaction. In a preferred embodiment, gap repair can be performed using *Taq* DNA polymerase.

Methods of Strand-specific Selection

[0062] The compositions and methods provided herein are useful for retaining directional information in double-stranded DNA.

[0063] The term “strand specific” or “directional”, as used herein, can 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.

IDT Exhibit 1020.5706

[0064] In some embodiments, the methods of the invention can be 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 can be 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.

[0065] The term "first strand synthesis" can refer 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.

[0066] The term "second strand synthesis" can refer 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.

[0067] The term "unmodified dNTPs" or "classic dNTPs" can refer 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" or "canonical nucleotide" can be used to refer to the four deoxyribonucleotide triphosphates dATP, dCTP, dGTP and dTTP that are normally found in DNA.

[0068] The term "canonical", as used herein, can refer to the nucleic acid bases adenine, cytosine, guanine and thymine that are commonly found in DNA or their deoxyribonucleotide or deoxyribonucleoside analogs. The term "non-canonical" can refer 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.

[0069] The term "modified nucleotide" or "modified dNTP", as used herein, can refer to any molecule suitable for substituting one corresponding unmodified or classic dNTP. Such modified nucleotides 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 must be suitable for specific degradation or cleavage in which it is selectively degraded or cleaved by a suitable degrading or cleavageagent, thus rendering the DNA strand containing at least one modified and degraded or cleaved dNTP essentially unfit for amplification, sequencing, and/or hybridization. Alternatively, the modified nucleotide must mark the DNA strand containing the modified nucleotide eligible for selective removal or cleavage or facilitate separation of the polynucleotide strands. Such a removal or cleavage or separation can be achieved by molecules, particles or enzymes interacting selectively with the modified nucleotide, thus selectively removing or marking for removal or cleaving only one polynucleotide strand.

[0070] As used in this application, the term "strand marking" can refer to any method for distinguishing between the two strands of a double-stranded polynucleotide. The term "selection" can

IDT Exhibit 1020.5707

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

[0071] The selective removal or cleavage of a marked strand in the present invention can be achieved through the use of enzymatic treatment of the marked strand. Enzymes that can be used for selective removal or cleavage of the marked strand according to the methods of the present invention can include glycosylases such as Uracil-N-Glycosylase (UNG), which selectively degrades the base portion of dUTP from the DNA backbone. Additional glycosylases which can be used in the methods of the present invention and their non-canonical or modified 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 Neddermann 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.

[0072] In some embodiments of the present invention the generation of an abasic site in the DNA backbone through the removal or cleavage of the base portion of at least one modified nucleotide (i.e. dUTP) can be followed by fragmentation or cleavage of the backbone at the abasic site. Suitable agents (for example, an enzyme, a chemical and/or reaction conditions such as heat) capable of

IDT Exhibit 1020.5708

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/ or the 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 or phosphodiester bond at an abasic site. In another embodiment, cleavage is with an amine, such as N,N'-dimethylethylenediamine (DMED). See, e.g., McHugh and Knowland, supra.

[0073] In some cases, the nucleic acid 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 90^0 C) may be used to cleave the nucleic acid comprising one or more abasic sites. In some cases, morpholine (e.g., 3M at 37^0 C or 65^0C) may be used to cleave the nucleic acid comprising one or more abasic sites. Alternatively, a polyamine may be used to cleave the nucleic acid comprising one or more abasic sites. Suitable polyamines include for example spermine, spermidine, 1,4-diaminobutane, lysine, the tripeptide K--W--K, DMED, piperazine, 1,2-ethylenediamine, or any combination thereof. In some cases, the nucleic acid 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 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 or unmodified 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 10^0 C, 11^0C, 12^0C, 13^0C, 14^0C, 15^0C, 16^0C, 17^0C, 18^0C, 19^0C, 20^0C, 21^0C, 22^0C, 23^0C,

IDT Exhibit 1020.5709

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. 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.

[0074] In embodiments involving fragmentation, the backbone of the 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 or phosphodiester bonds 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 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, glycosylases of the present invention may comprise both a glycosylase and a lyase activity, whereby the glycosylase activity cleaves the base portion of a nucleotide (e.g., a modified nucleotide) to form an abasic site and the lyase activity cleaves the phosphodiester backbone at the abasic site so formed. In some cases, the glycosylase comprises both a glycosylase activity and an AP endonuclease activity.

[0075] Appropriate reaction media and conditions for carrying out the cleavage of a base portion of a non-canonical or modified nucleotide according to the methods of the invention are those that permit cleavage of a base portion of a non-canonical or modified 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 B1; U.S. Pat. No. 5,035,996; and U.S. Pat. No. 5,418,149. 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 or modified nucleotide and the cleavage of a base portion of the non-canonical or modified nucleotide. In another embodiment, different reaction conditions are used for

IDT Exhibit 1020.5710

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.

[0076] 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 out by 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 DMED, or by heat treatment. In a preferred embodiment, UNG treatment is conducted in a reaction buffer containing 32 mM DMED.

[0077] 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.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 in an aminoallyl dNTP. In yet another embodiment, the at least one modified nucleotide is inosine, replacing dGTP in the dNTP mix.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 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.

IDT Exhibit 1020.5711

[0078] In an embodiment where the second strand and one of the adaptors contains at least one modified nucleotide, the second strand and the one of the adaptors may be synthesized so that each comprises a sufficient and predictable density of modified nucleotides to provide for sufficient and predictable fragmentation, and when used with one or more agents capable of cleaving at the modified nucleotides (e.g., a glycosylase, a glycosylase and an amine, a glycosylase and heat, or a glycosylase and an AP endonuclease) to further generate fragments of desirable size range. Generally, a modified 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 modified nucleotide. In one embodiment, the modified 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 modified 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 modified to non-modified 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 modified nucleotide dUTP to non-modified nucleotide dTTP is used in the reaction mixture.

[0079] The term "adaptor", as used herein, refers to an oligonucleotide of known sequence, the ligation of which to a target polynucleotide or a target polynucleotide strand of interest enables the generation of amplification-ready products of the target polynucleotide or the target polynucleotide strand of interest. Various adaptor designs are envisioned. 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. The adaptors can be designed to comprise a double-stranded portion which can be ligated to double-stranded DNA (or double-stranded DNA with overhang) products. Various ligation processes and reagents are known in the art and can be useful for carrying out the methods of the invention. For example, blunt ligation can be employed. Similarly, a single dA nucleotide can be added to the 3′-end of the double-stranded DNA product, by a polymerase lacking 3′-exonuclease activity and can anneal to an adapter comprising a dT overhang (or the reverse). This design allows the hybridized components to be subsequently ligated (e.g., by T4 DNA ligase). Other ligation strategies and the corresponding reagents are known in the art and kits and reagents for carrying out efficient ligation reactions are commercially available (e.g, from New England Biolabs, Roche). The double-stranded DNA portion of the adaptors can further comprise indexing or bar-coding sequences designed to mark either the samples or sequences of interest.

[0080] Blunt-end ligation with conventional duplex adaptors can be employed in the present invention, meaning that the adaptors are capable of ligation at either end of the target polynucleotide

IDT Exhibit 1020.5712

strand, thereby generating ligation products with two adaptor orientations. In a preferred embodiment, one of the two adaptors has at least one modified nucleotide incorporated along the ligation strand of the adaptor.

Methods of Amplification

[0081] The methods, compositions and kits described herein can be useful to generate amplification-ready products for downstream applications such as massively parallel sequencing (i.e. next generation sequencing methods) or hybridization platforms. Methods of amplification are well known in the art. Suitable amplification reactions can include any DNA amplification reaction, including but not limited to polymerase chain reaction (PCR), strand displacement amplification (SDA), linear amplification, multiple displacement amplification (MDA), rolling circle amplification (RCA), single primer isothermal amplification (SPIA, see e.g. U.S. Pat. No. 6,251,639), Ribo-SPIA, 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.

[0082] 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.

[0083] 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.

[0084] 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.

IDT Exhibit 1020.5713

[0085] 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.

[0086] 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.

Downstream Applications for Whole Transcriptome Analysis

[0087] 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 cDNA library or a whole transcriptome while retaining information on which strand was present in the original RNA sample. 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 or unknown sequence, combining the one or more primers with a reverse transcriptase, reverse transcribing the sample, 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 modified dNTP during second strand synthesis and incorporated into the second strand, performing end repair on the double-stranded cDNA, ligating adaptors to the double-stranded cDNA, wherein one of the adaptors has the modified dNTP incorporated into a ligation strand of the adaptor, performing gap repair, selectively removing or marking for removal the second strand by a suitable degrading agent, amplifying the RNA sample using one or more primers to produce amplified products, and performing sequencing on the products. In some embodiments, sequencing is performed on single-stranded cDNA as generated by the methods of the present invention without amplifying the RNA sample following selective removal of the marked second strand. 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.

[0088] 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

IDT Exhibit 1020.5714

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 embodiments, no more than 5%, 10%, 15%, 20%, 25%, 30%, 40%, 50%, 60%, 70%, 75%, 80% of the resulting products are rRNA sequences.

[0089] The methods of the invention can be useful for sequencing by the method commercialized by Illumina, as described in 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 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.

[0090] In some embodiments, the methods of the invention can be 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 can be 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 can be 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.

[0091] 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

IDT Exhibit 1020.5715

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.

[0092] 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.

Kits

[0093] Any of the compositions described herein may be comprised in a kit. In a non-limiting example, the kit, in a suitable container, comprises: one or more primers, a reverse transcription enzyme, and optionally reagents for amplification.

[0094] The containers of the kits can generally include at least one vial, test tube, flask, bottle, syringe or other containers, into which a component may be placed, and preferably, suitably 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.

[0095] 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.

[0096] 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,

IDT Exhibit 1020.5716

polymerases, enzymes, positive and negative control oligonucleotides and target sequences, test tubes or plates, fragmentation reagents, detection reagents, purification matrices, and an instruction manual. In some embodiments, the kit of the present invention contains a non-canonical or modified nucleotide. Suitable non-canonical or modified nucleotides include any nucleotides provided herein including but not limited to dUTP, or a methylated purine.

[0097] 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, 1.5 x, 2 x, 2.5 x, 3 x, 4 x, 5 x, 6 x, 7 x, 10 x, 15 x, 20 x, 25 x, 33 x, 50 x, 75 x, 100 x 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 or modified nucleotides, dyes, fluorophores, biotin, enzymes (such as endonucleases, exonucleases, glycosylases), or any combination thereof.

[0098] 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 hybridizing 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 chimeric primers comprising an RNA portion and a DNA portion. In some cases, the kit may contain primers comprising non-canonical or modified nucleotides.

[0099] 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 displacement 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 or modified nucleotides.

[00100] 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, 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.

[00101] 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

IDT Exhibit 1020.5717

cases, this agent may comprise one or more glycosylases. Suitable glycosylases include any glycosylases provided herein including but not limited to UDG, or MPG.

[00102] 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 such as DMED, piperidine, AP endonucleases, or any combination thereof.

[00103] 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 or more polymerases, wherein the reagents are not provided as a mixture. Additionally, the reagents for producing blunt ends may comprise dNTPs.

[00104] 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.

[00105] 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.

[00106] In some embodiments, the kit of the present invention may contain one or more reagents for performing gap or fill-in repair on the ligation complex formed between the adaptors and the double stranded products of the present invention. The kit may contain a polymerase suitable for performing gap repair. Suitable polymerases are included, for example, *Taq* DNA polymerase.

[00107] The kit may further contain instructions for the use of the kit. For example, the kit may contain instructions for generating directional cDNA libraries or directional cDNA libraries representing a whole transcriptome useful for large scale analysis including but not limited to e.g., , pyrosequencing, sequencing by synthesis, sequencing by hybridization, single molecule sequencing, nanopore sequencing, and sequencing by ligation, high density PCR, digital PCR, massively parallel

IDT Exhibit 1020.5718

Q-PCR, 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 generating a second strand comprising one or more modified 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 a nucleic acid template. The kit may further contain instructions for extending the one or more oligonucleotide primers with for example a polymerase. The kit may further contain instructions for cleaving the base portion of a modified 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 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.

[00108] 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.

Products based on the Methods of the Invention

[00109] Products based on the methods of the invention may be commercialized by the Applicants under the trade name Encore ™, Ultra-low Encore ™ or Encore Eukaryotic Stranded RNA-Seq. Encore is a trademark of NuGEN Technologies, Inc.

EXAMPLES

Example 1 – Generation of a directional cDNA library.

[00110] 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. An overview of the end-end work-flow for the generation of the directional cDNA library is shown in FIG. 4.

[00111] First strand synthesis

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

IDT Exhibit 1020.5719

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.

[00112] Second strand synthesis with dUTP incorporation

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.

[00113] Fragmentation and purification of cDNA fragments

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.

[00114] End repair

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.

[00115] Ligation with dU marked adaptors

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

[00116] Nick repair/adaptor fill-in

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.

[00117] Strand selection with UDG/APE I treatment

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.

IDT Exhibit 1020.5720

Consequently, a polynucleotide strand with incorporated dUTP that is treated with UNG/APE I was unable to undergo amplification by a polymerase.

[00118] Library amplification

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 2 – RNA Strand Retention Efficiency.

[00119] In this example, strand retention efficiency using the methods of the invention was validated experimentally by assessing the strand bias of sequence reads that map to the coding exons of human mRNAs. A directional cDNA library, as described in the invention, and a non-directional cDNA library (control) were generated from poly (A)+ RNA isolated from human whole brain. Single end 40 nucleotide reads were generated using the Illumina Genome Analyzer II. Strand retention efficiency was measured by comparing the strand biases of sequence reads from the directional library and the non-directional control library. The results are presented in Figure 3. After dUTP incorporation and UNG/APE I digestion of the strand with the undesired P2/P1 adaptor orientation, 98% of reads from the directional cDNA library aligned to the correct (antisense) strand orientation, as compared to approximately 50% of reads in a non-directional control cDNA library.

Example 3 – RNA Strand Retention Efficiency.

[00120] In this example, strand retention efficiency using the methods of the invention was validated experimentally by assessing the strand bias of sequence reads that map to the 5' UTR and 3' UTR regions of human mRNAs. Strand retention efficiency was measured as described in Example 2. The corresponding strand retentions for the directional library were 95% and 98% in the 5' UTR and 3' UTR regions, respectively, and 39% and 50% for the non-directional library.

Example 4 – Generation of a directional cDNA library.

[00121] 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.

[00122] First strand synthesis with dUTP incorporation

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 supplemented with dNTP mix containing dATP, dCTP, dGTP, and dUTP in place of dTTP. 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. After first strand synthesis, non-incorporated dNTPs were removed prior to second strand synthesis.

[00123] Second strand synthesis

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

IDT Exhibit 1020.5721

supplemented with dNTP mix containing dATP, dCTP, dGTP, and dTTP. 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.

[00124] Fragmentation and purification of cDNA fragments

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.

[00125] End repair

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.

[00126] Ligation with dU marked adaptors

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

[00127] Nick repair/adaptor fill-in

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.

[00128] Strand selection with UDG/APE I treatment

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.

[00129] Library amplification

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.

IDT Exhibit 1020.5722

CLAIMS

WHAT IS CLAIMED IS:

1. A method for construction of a directional cDNA library 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; and
    f. selectively cleaving the second strand and the adaptor that has the modified dNTP by a suitable cleavage agent, thereby generating a directional cDNA library comprising the first strand cDNA.
2. The method of claim 1, wherein the modified dNTP comprises dUTP.
3. The method of claim 1, wherein the method, optionally, further comprises fragmenting the double-stranded cDNA prior to step (c).
4. The method of claim 1, wherein step (f) comprises cleaving a base portion of the modified nucleotide, thereby forming an abasic site.
5. The method of claim 4, further comprising cleaving a phosphodiester backbone at the abasic site.
6. The method of claim 4, wherein the cleaving the base portion of the modified nucleotide comprises use of an enzyme.
7. The method of claim 5, wherein the cleaving the phosphodiester backbone at the abasic site comprises use of an enzyme, chemical agent, and/or heat.
8. The method of claim 7, wherein the enzyme is an endonuclease.
9. The method of claim 8, wherein the endonuclease is an apurinic/apyrimidinic endonuclease (APE).
10. The method of claim 6, wherein the enzyme comprises a glycosylase.
11. The method of claim 10, wherein the glycosylase is uracil-N-glycosylase (UNG) or uracil DNA glycosylase (UDG).
12. The method of claim 7, wherein the chemical agent comprises a primary amine.
13. The method of claim 7, wherein the chemical agent comprises a polyamine.
14. The method of claim 13, wherein the polyamine comprises N,N-dimethylethylenediamine (DMED).

IDT Exhibit 1020.5723

15. The method of claim 1, wherein the cleavage agent comprises a glycosylase and a polyamine.
16. The method of claim 1, wherein the cleavage agent comprises a glycosylase and a APE.
17. The method of claim 1, further comprising amplifying the directional cDNA library, thereby generating amplified products.
18. The method of claim 17, wherein the amplification of the directional cDNA library 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).
19. The method of claim 18, wherein the amplification comprises PCR.
20. The method of claim 18, wherein the amplification comprises SPIA.
21. The method of claim 17, further comprising an additional step of sequencing the amplified products.
22. The method of claim 21, wherein the sequencing comprises next generation sequencing.
23. The method of claim 1, further comprising cleaving the RNA sample following reverse transcription.
24. The method of claim 23, wherein cleaving the RNA sample comprises exposing the RNA sample to an RNase following reverse transcription.
25. The method of claim 24, wherein the RNase is RNase H.
26. The method of claim 23, wherein cleaving the RNA sample comprises exposing the RNA sample to heat or chemical treatment or a combination thereof.
27. The method of claim 1 further comprising reducing or depleting non-desired nucleic acid sequences from the directional cDNA library.
28. The method of claim 27, wherein the non-desired nucleic acid is ribosomal RNA (rRNA).
29. 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 adaptor that has the modified dNTP, by a suitable cleavage agent, thereby generating a directional cDNA library comprising the first strand cDNA; and

IDT Exhibit 1020.5724

g. amplifying and/or sequencing the directional cDNA library.

30. The method of claim 29, wherein the modified dNTP comprises dUTP.

31. The method of claim 29, wherein the method, optionally, further comprises fragmenting the double-stranded cDNA prior to step (c).

32. The method of claim 29, wherein step (f) comprises cleaving a base portion of the modified nucleotide, thereby forming an abasic site.

33. The method of claim 32, further comprising cleaving a phosphodiester backbone at the abasic site.

34. The method of claim 32, wherein the cleaving of the base portion of the modified nucleotide comprises use of an enzyme.

35. The method of claim 33, wherein the cleaving the phosphodiester backbone at the abasic site comprises use of an enzyme, chemical agent, and/or heat.

36. The method of claim 35, wherein the enzyme is an endonuclease.

37. The method of claim 36, wherein the endonuclease is an apurinic/apyrimidinic endonuclease (APE).

38. The method of claim 34, wherein the enzyme comprises a glycosylase.

39. The method of claim 38, wherein the glycosylase is uracil-N-glycosylase (UNG) or uracil DNA glycosylase (UDG).

40. The method of claim 35, wherein the chemical agent comprises a primary amine.

41. The method of claim 35, wherein the chemical agent comprises a polyamine.

42. The method of claim 41, wherein the polyamine comprises N,N-dimethylethylenediamine (DMED).

43. The method of claim 29, wherein the -cleavage agent comprises a glycosylase and a polyamine.

44. The method of claim 29, wherein the -cleavage agent comprises a glycosylase and a APE.

45. The method of claim 29, wherein 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).

46. The method of claim 45, wherein the amplification comprises PCR.

47. The method of claim 45, wherein the amplification comprises single primer isothermal amplification (SPIA).

48. The method of claim 29, wherein the sequencing comprises next generation sequencing.

49. The method of claim 29, further comprising cleaving the RNA sample following reverse transcription.

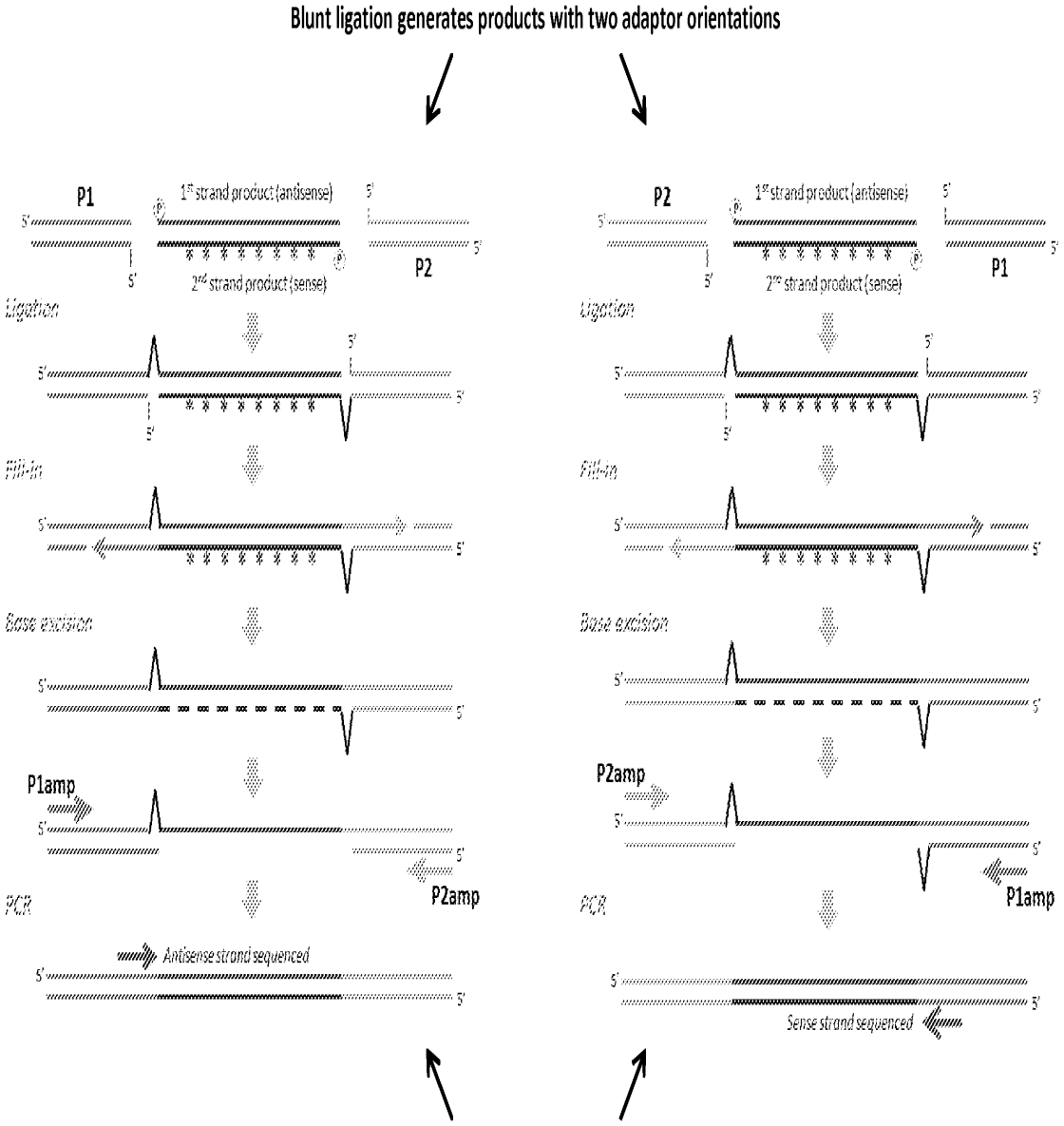
50. The method of claim 49, wherein cleaving the RNA sample comprises exposing the RNA sample to an RNase following reverse transcription.

IDT Exhibit 1020.5725

51. The method of claim 50, wherein the RNase is RNase H.
52. The method of claim 49, wherein cleaving the RNA sample comprises exposing the RNA sample to heat or chemical treatment or a combination thereof.
53. The method of claim 49 further comprising reducing or depleting non-desired nucleic acid sequences from the directional cDNA library.
54. The method of claim 53, wherein the non-desired nucleic acid is ribosomal RNA (rRNA).
55. A kit comprising:
    a. One or more primers;
    b. a reverse transcription enzyme;
    c. a glycosylase;
    d. one or more adaptors wherein one of the adaptors comprises at least one modified nucleotide in a ligation strand of the adaptor.
56. The kit of claim 55, further comprising written instructions for use of the kit.
57. The kit of claim 55, further comprising at least one modified nucleotide or dNTP.
58. The kit of claim 55, further comprising reagents for amplification.
59. The kit of claim 55, further comprising reagents for sequencing.
60. The kit of claim 55, further comprising a polyamine, an APE, or a combination thereof.
61. The kit of claim 55, wherein the modified nucleotide comprises dUTP, and wherein the glycosylase comprises UNG.
62. The kit of claim 57, wherein the modified nucleotide comprises dUTP.
63. A method for construction of a directional cDNA library comprising:
    a. synthesizing first and second strand cDNA from template RNA, wherein either first strand cDNA or second strand cDNA synthesis comprises incorporating a modified dNTP, to generate double stranded cDNA;
    b. performing end repair on the double-stranded cDNA;
    c. 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;
    d. performing gap repair; and
    e. selectively cleaving the first or the second strand cDNA and the adaptor that has the modified dNTP by a suitable cleavage agent, thereby generating a directional cDNA library comprising the first or the second strand cDNA.

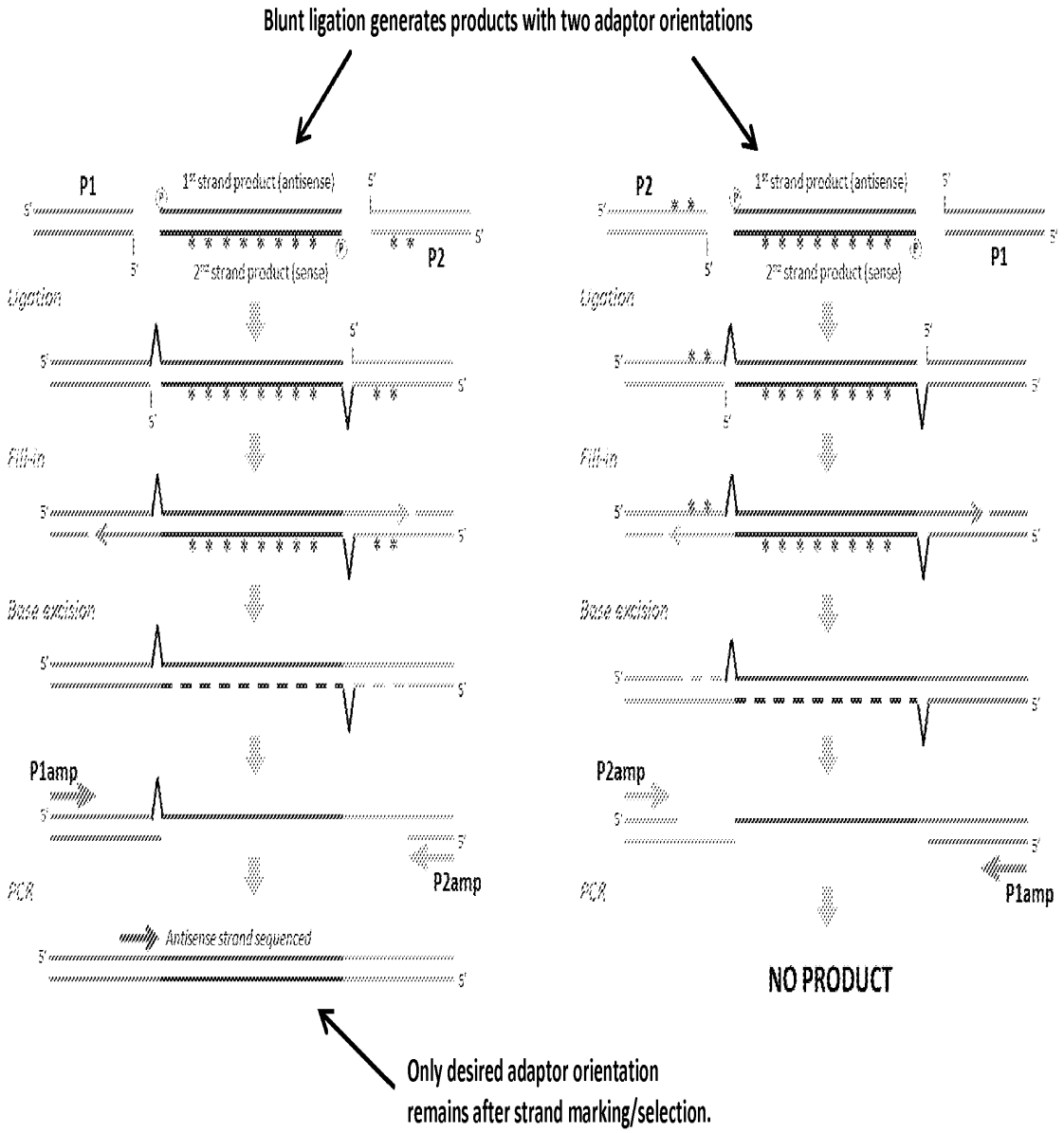
IDT Exhibit 1020.5726

Figure 1. Conventional system: Strand marking of cDNA inserts is not sufficient for directional library construction.



The asterisks in the diagram (*) indicate dUTP incorporation.

Figure 2. Construction of directional (strand-specific) cDNA libraries using duplex adaptors: Strand marking of both cDNA and one adaptor results in directional library construction.

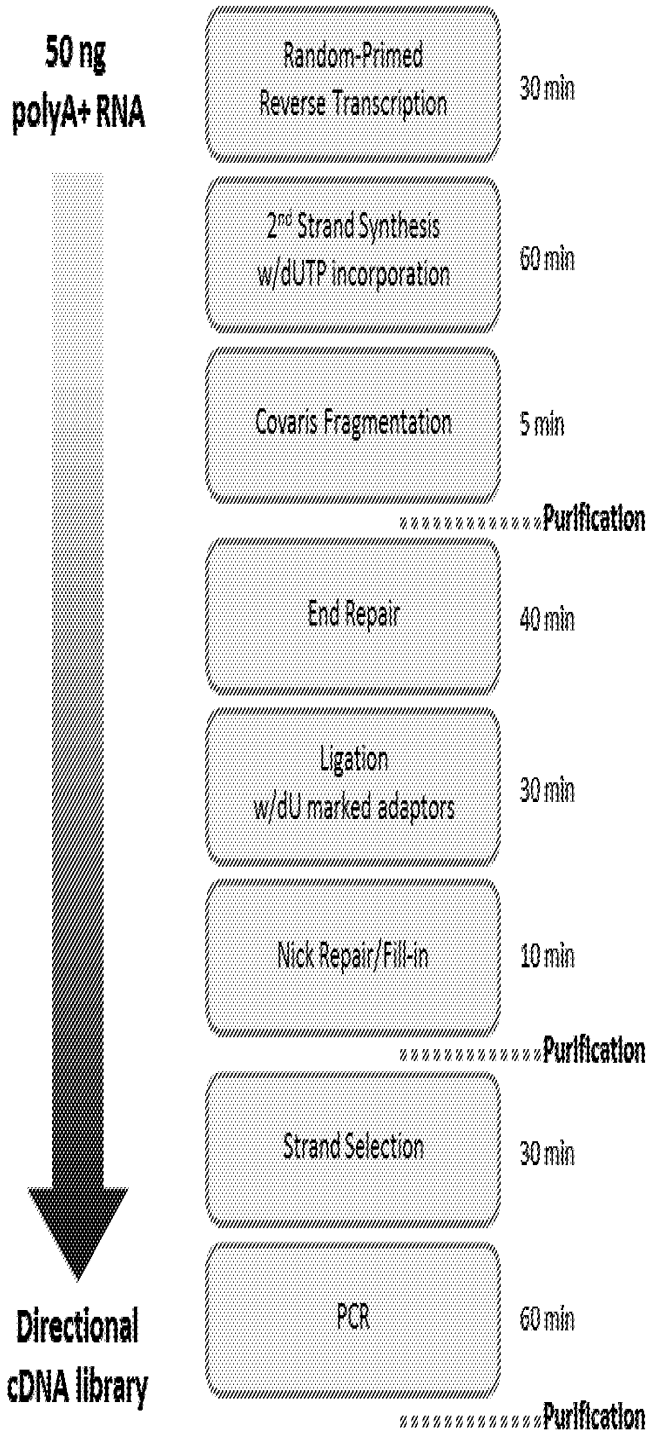


The asterisks in the diagram (*) indicate dUTP incorporation.

Figure 3. Summary of strand retention efficiency data for sequence reads mapping to coding exons of human mRNAs.

Library modifications				% of all reads			% of reads mapping to coding exons	
bases used in 2ss reaction	bases used in P2 adaptor	base excision	total reads	mapped	not mapped	strand orientation expected	strand orientation observed - sense*	strand orientation observed - antisense*
dA,dC,cG,dT	dA,dC,cG,dT	none	3261009	89.8	10.2	sense and antisense	52.3	47.7
dA,dC,cG,dT	dA,dC,cG,dT	UNG+APE	4111334	89.1	10.9	sense and antisense	51.1	48.9
dA,dC,cG,dU	dA,dC,cG,dU	UNG+APE	4829979	93.8	6.2	antisense only	2.1	97.9
dA,dC,cG,dT	dA,dC,cG,dT	none	3673934	89.6	10.4	sense and antisense	52.5	47.5
dA,dC,cG,dT	dA,dC,cG,dT	UNG+DMED	3977383	89.2	10.8	sense and antisense	51.2	48.8
dA,dC,cG,dU	dA,dC,cG,dU	UNG+DMED	4616560	93.8	6.2	antisense only	1.9	98.1

Figure 4. End-to-end workflow for construction of a directional cDNA library.



INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2012/061218

A. CLASSIFICATION OF SUBJECT MATTER
IPC(8) - C40B 50/06 (2013.01)
USPC - 506/26
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC(8) - C12P 19/34; C12Q 1/68; C40B 40/06, 40/08, 50/06 (2013.01)
USPC - 435/91.2, 91.5, 91.51, 91.52, 91.53; 506/16, 17, 26

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
CPC Class/Subclass(es): C12N 15/10; C12P 19/34; C12Q 1/6806, 1/6844, 1/6853, 1/6855

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
MicroPatent, Orbit.com, Google Patents, Google Scholar

C. DOCUMENTS CONSIDERED TO BE RELEVANT.

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X ----- Y	BORODINA et al. "Chapter five – A Strand-Specific Library Preparation Protocol for RNA Sequencing," Methods in Enzymology, 20 September 2011 (20.09.2011), Vol. 500, Pgs. 79-98. entire document	1-4, 6, 10, 11, 17-19, 21-25, 27-32, 34, 38, 39, 45, 46, 48-51, 53-59, 61-63 ---------- 5, 7-9, 12-16, 20, 26, 33, 35-37, 40-44, 47, 52, 60
Y	US 2011/0224105 A1 (KURN et al) 15 September 2011 (15.09.2011) entire document	5, 7-9, 12-16, 20, 26, 33, 35-37, 40-44, 47, 52, 60
A	EP 2,272,976 A1 (SOLDATOV et al) 12 January 2011 (12.01.2011) entire document	1-63
A	PARKHOMCHUK et al. "Transcriptome analysis by strand specific sequencing of complementary DNA," Nucleic Acids Research, 20 July 2009 (20.07.2009), Vol. 37, No. 18, Pg. e123. entire document	1-63

☐ 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
18 January 2013	12 FEB 2013
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-3201	Authorized officer: Blaine R. Copenheaver 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/US2012/061218

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing filed or furnished:

   a. (means)

   ☐ on paper

   ☐ in electronic form

   b. (time)

   ☐ in the international application as filed

   ☐ together with the international application in electronic form

   ☐ subsequently to this Authority for the purposes of search

2. ☐ In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that in the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

3. Additional comments:

The application does not contain a sequence listing complying with the relevant standard and could not be used for the purposes of the International Search.

IDT Exhibit 1020.5732

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
30 May 2014 (30.05.2014)

WIPO | PCT

(10) International Publication Number
WO 2014/082032 A1

(51) International Patent Classification:
C40B 20/04 (2006.01)

(21) International Application Number:
PCT/US2013/071656

(22) International Filing Date:
25 November 2013 (25.11.2013)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/729,853 26 November 2012 (26.11.2012) US
61/730,463 27 November 2012 (27.11.2012) US
61/784,394 14 March 2013 (14.03.2013) US

(71) Applicant: **THE UNIVERSITY OF TOLEDO** [US/US]; 2801 Bancroft St., Ms 218, Toledo, OH 43606 (US).

(72) Inventors: **WILLEY, James, C.**; 4235 Deepwood Lane, Toledo, OH 43614 (US). **BLOMQUIST, Thomas**; 19 Karis St., Waterville, OH 43566 (US). **CRAWFORD, Erin**; 544 Bruns Dr., Rossford, OH 43460 (US).

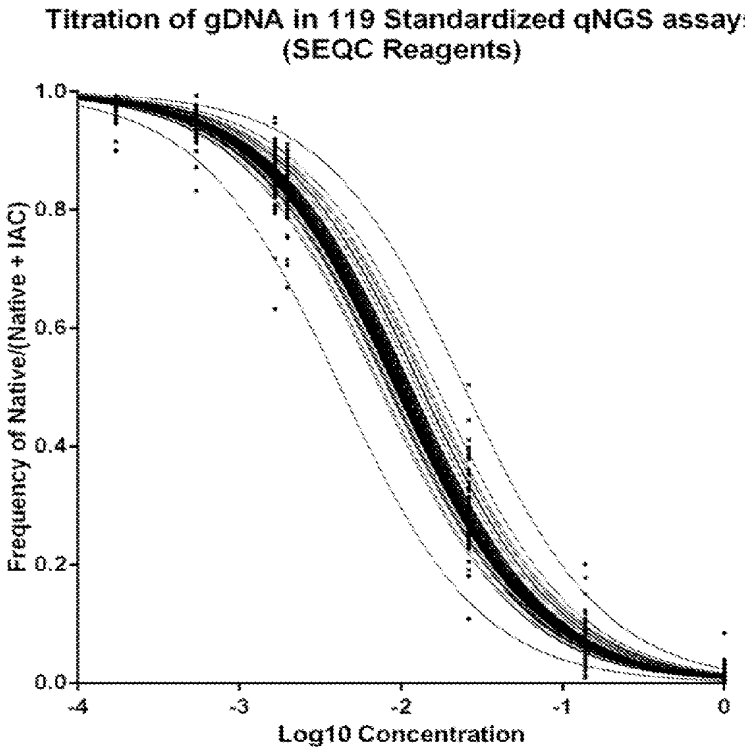
(74) Agent: **MARTINEAU, Catherine, B.**; Macmillan, Sobanski & Todd, LLC, One Maritime Plaza, 5th Floor, 720 Water Street, Toledo, OH 43604 (US).

(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, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV,

[Continued on next page]

(54) Title: METHODS FOR STANDARDIZED SEQUENCING OF NUCLEIC ACIDS AND USES THEREOF



(57) Abstract: Methods for standardized sequencing of nucleic acids and uses thereof are described. The identification of genetic information is becoming a key piece of information for the diagnosis and treatment of many diseases. In order to make such diagnostic tool readily available, it is desired that this identification be as efficient and as inexpensive as possible.

WO 2014/082032 A1

IDT Exhibit 1020.5733

MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

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

TITLE

METHODS FOR STANDARDIZED SEQUENCING OF NUCLEIC ACIDS AND USES THEREOF

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of United States Provisional Application Ser. Nos. 61/729,853 filed November 26, 2012, 61/730,463 filed November 27, 2012, and 61/784,394 filed March 14, 2013, the entire disclosures of which are expressly incorporated herein by reference.

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH

[0002] The invention was made with U.S. Government support under Grant Numbers CA138397 and HL108016 awarded by the National Institutes of Health. The United States Government has certain rights in the invention.

FIELD OF THE INVENTION

[0003] The present invention relates methods for standardized sequencing of nucleic acids and uses thereof.

BACKGROUND

[0004] The identification of genetic information is becoming a key piece of information for the diagnosis and treatment of many diseases. In order to make such diagnostic tool readily available, it is desired that this identification be as efficient and as inexpensive as possible. For diagnostic, medical, regulatory and ethical aspects, this identification should be as accurate as possible in order to rule out false measurements.

[0005] In addition to the desire to acquire human genetic material information, there is great interest in acquiring genetic information on, for example, mitochondria, pathogens and organisms that cause diseases.

[0006] One method for acquiring information is the Sanger sequencing method of genome analysis. Other methods are becoming available which provide an improved performance when compared with the Sanger sequencing method. These methods include a short high density parallel sequencing technology, next generation sequencing (i.e., NextGen or "NGS"), which are attempting to provide a more comprehensive and accurate view of RNA in biological samples than the Sanger sequence method.

[0007] Next-generation sequencing (NGS) is useful in a multitude of clinical applications by virtue of its automated and highly parallelized analysis of nucleic acid templates. However, the limit of clinical questions that NGS can address is largely determined by: i) the upstream source of nucleic acid template (e.g., human tissue, microbial sample, etc.), and ii) whether the clinically

IDT Exhibit 1020.5735

relevant biological variation in the nucleic acid template is greater than the technical variation (which is often introduced by such variants as workflow for sample preparation, sequencing and/or data analysis).

[0008] The workflow for NGS library preparation varies widely, but can broadly be grouped into one of two approaches: 1) digestion or fragmentation of the nucleic acid sample with subsequent ligation to a universal adaptor sequence, or 2) PCR with target specific primers that incorporate a universal adaptor sequence at their 5' ends. In both approaches, if a nucleic acid template is RNA, a reverse transcription step is used to create the requisite DNA template for sequencing.

[0009] One concern with NGS is that these quantitative sequencing methods have high intra-lab and inter-lab variation. This problem thus reduces the value of any results, and has prevented the use of these sequencing methods in molecular diagnostics.

[0010] For example, non-systematic (i.e., non-reproducible) biases (i.e., errors), are often inadvertently introduced during preparation of the sequencing library. These non-systemic biases are a major roadblock to implementing NGS as a reliable and efficient routine measurement of nucleic acid abundance (quantification) in the clinical setting.

[0011] The most likely source of non-systematic bias (thus preventing inter-laboratory comparison, and hence routine clinical use, of quantitative NGS data) stems from issues arising from nucleic acid fragmentation, adaptor ligation and PCR.

[0012] Also, although not explicitly required, the FDA has issued guidance and industry recommendations that PCR-based in vitro diagnostic (IVD) devices should contain internal amplification controls (IAC) to control for interfering substances and verify that a negative result for a sample is not caused by inhibitors.

[0013] In addition, in order to avoid stochastic sampling error and ensure reliable measurements, it is necessary to sequence (i.e., read) a sufficient number of copies of the analyte being measured. One problem is that the range of transcript representation following library preparation often remains very high, typically one million-fold or greater, imposing high cost. This is because the transcripts from each gene must be sequenced at least 10 times (ensure 10 "reads"). To ensure 10 reads for the least represented genes, it is necessary to read a gene represented at one million fold higher level at least 10 million times.

[0014] Thus, a NGS method that reduces inter-experimental and inter-laboratory variation in measurement of nucleic acid copy number in samples will be of great use to both research and clinical applications.

SUMMARY OF THE INVENTION

[0015] Described herein is a method for providing reproducibility in measurement of nucleic acid copy number in samples, comprising measuring a proportional relationship of at least one

IDT Exhibit 1020.5736

native target sequencing event of at least one nucleic acid in a sample to the respective competitive internal amplification control (IAC) for that nucleic acid.

[0016] Also described herein is a method where the at least one event comprises: an observation, a count and/or a read between the native target and its respective IAC.

[0017] Also described herein is a method for controlling for non-systematic error in PCR-based NGS library preparation, comprising sharing identical priming sites to a native nucleic acid template of interest so as to mimic the kinetics of the native target in the PCR reaction, and thus control for target-specific variation in PCR efficiency.

[0018] Also described herein is the use of a competitive IAC method to provide for the convergence of target analyte representation in a sample, while retaining quantitative information of the original representation of both low- and high-abundance target, enabling quantitative measurement of original representation with a low number of sequencing reads.

[0019] In a particular embodiment, described herein is a method for determining an amount of a first nucleic acid, comprising:

providing a series of serially-diluted standardized mixtures comprising a competitive template for said first nucleic acid and a competitive template for a second nucleic acid present in a number of samples comprising said first nucleic acid,

wherein said competitive templates are at known concentrations relative to each other; combining one of said samples comprising said first nucleic acid with a first one of said serially-diluted standardized mixtures;

co-amplifying said first nucleic acid and said competitive template for said first nucleic acid to produce amplified product thereof; obtaining a first relationship,

said first relationship comparing said amplified product of said first nucleic acid to said amplified product of said competitive template for said first nucleic acid; determining whether said first relationship is within about 1:10 to about 10:1;

if not, repeating said combining, co-amplifying, obtaining and determining steps with a second one of said serially-diluted standardized mixtures;

co-amplifying said second nucleic acid and said competitive template for said second nucleic acid to produce amplified product thereof;

obtaining a second relationship, said second relationship comparing said amplified product of said second nucleic acid to said amplified product of said competitive template for said second nucleic acid; and

comparing said first and said second relationships.

[0020] In one embodiment, the method includes comparing said amplified project of said first nucleic acid to said amplified project of said competitive template for said first nucleic,

determining whether said first relationship is within about 1:100 to about 100:1, or 1:1000 to about 1000:1, or 1:10,000 to about 10,000 to 1,

IDT Exhibit 1020.5737

if not, repeating said combining, co-amplifying, obtaining and determining steps with a second one of said serially-diluted standardized mixtures;

co-amplifying said second nucleic acid and said competitive template for said second nucleic acid to produce amplified product thereof; obtaining a second relationship;

said second relationship comparing said amplified product of said second nucleic acid to said amplified project of said competitive template for said second nucleic acid; and

comparing said first and said second relationships.

[0021] Further, in certain embodiments, the products from the series of co-amplification reactions in claim 2 are combined and amplified in a second round using primer pairs that recognize each NT and CT product from first round of amplification and that also have gene specific barcode primer and universal primer at 5' end to facilitate sequencing.

[0022] Other systems, methods, features, and advantages of the present invention will be or will become apparent to one with skill in the art upon examination of the following drawings and detailed description. It is intended that all such additional systems, methods, features, and advantages be included within this description, be within the scope of the present invention, and be protected by the accompanying claims.

BRIEF DESCRIPTION OF THE DRAWINGS

[0023] The patent or application file may contain one or more drawings executed in color and/or one or more photographs. Copies of this patent or patent application publication with color drawing(s) and/or photograph(s) will be provided by the Patent Office upon request and payment of the necessary fee.

[0024] **FIG. 1:** A549 gDNA titrated relative to ICA mixture. Graph showing titration of a mixture of approximately equimolar competitive internal amplification controls (IAC) relative to a fixed amount of genomic DNA (gDNA) input of 100,000 copies into each Multiplex-PCR. Plotted on the X-axis is the initially estimated amount of each IAC in equimolar mixture. The Y-axis represents the frequency of observed sequencing events (reads) for each native template divided by the sum of read frequencies for the native template and its respective competitive IAC.

[0025] **FIGS. 2A-2F:** Graphs showing titration of a mixture of approximated equimolar competitive internal amplification controls (IAC) relative to: **FIG. 2A**) a fixed input amount of 100,000 copies genomic DNA (gDNA) into Multiplex-PCR, or 11 ng of Reverse Transcribed RNA cDNA material from SEQC Samples: **FIG. 2B**) A-RT1, **FIG. 2C**) A-RT2, **FIG. 2D**) B, **FIG. 2E**) C and **FIG. 2F)** D), input into each Multiplex-PCR. Plotted on the X-axis is the initially estimated amount of each IAC in the equimolar mixture. The Y-axis represents the frequency of observed sequencing events (reads) divided by the sum of read frequencies for the native template and its respective competitive IAC.

[0026] **FIG. 3:** Graphs showing analysis of assay accuracy. The values measured for SEQC

IDT Exhibit 1020.5738

Sample C-RT1 were compared to the values expected (% difference) based on measured SEQC Sample A and B values. The percent difference between the predicted signal C′ and the actual assay signal C was used as an indication of relative assay accuracy (RA). An RA score ΔC for a target gene was defined as (C–C′/C′), respectively. The mean RA value (line), median quartiles (box), standard deviation (whiskers), and outliers are presented. The mean RA value was very close to the estimated value. Only certain assays has RA greater than 25% difference from the mean.

[0027] **FIG. 4A:** Graph showing gene targets (n=88) measured between Samples SEQC A-RT1, B, C evaluated for inter-gene differential expression (DE); i.e., whether their DE was between 1.5 to 3.0, 2 to 3, 3 to 5, 5 to 10, or greater than 10-fold in change. Control for false positive or negative change was assessed by comparing SEQC A-RT1 versus SEQC A-RT2.

[0028] **FIG. 4B: Table 1** presenting the summary statistics for differential expression in Sample C compared to expected based on Samples A and B.

[0029] **FIG. 5:** Graphs showing assay reproducibility between Reverse Transcriptions. Two reverse transcriptions of SEQC Sample A (RT1 versus RT2) were measured for expression of 119 gene targets that successfully passed performance criteria in **FIG. 1**.

[0030] **FIG. 6:** Graph showing same data as in **FIG. 3** and **FIG. 4**. Expected measurements in Sample C (x-axis) versus observed measurements (y-axis). Reverse transcription of SEQC Sample C were measured for expression of 119 gene targets that successfully passed performance criteria in **FIG. 1**. Of the 119 gene targets, 88 had $R^2 > 0.95$ for curve fit of the hill equation to determine equivalence point and concentration of each target (y-axis) (**FIG. 2**).

[0031] **FIG. 7:** Graph showing convergence and increased uniformity of 97 gene targets from **FIG. 5**. Plotted on the X-axis is the data from **FIG. 5** in proportion to highest abundance template. On the Y-axis is the actual proportion of sequencing reads that gene-target is in proportion to highest sequence template.

[0032] **FIG. 8:** Graph showing ROC curve to detect >3 fold change for RNA-Sequencing using Illumina platform from Bullard et al. BMC Bioinformatics 2010, 11:94.

[0033] **FIG. 9:** Provides a schematic illustration of a PCR Master mix with a mixture of internal Amplification Controls (IAC).

[0034] **FIGS. 10-11:** Graphs showing the titration of a mixture of Internal Amplification Controls against gDNA and SEQC cDNA.

[0035] **FIG. 12:** Graph showing the same Library Preparation Replicate Sequencing (Intra-Site), where X-axis = 1.8 million sequencing reads and Y-axis = 3.0 million sequencing reads.

[0036] **FIG. 13:** Graph showing the separate Library Preparation Sequenced (Intra-Site), where X-axis = 2.6 million sequencing reads and Y-axis = 4.8 million sequencing reads.

[0037] **FIGS. 14A-14B:** Graphs showing predicting measurement of Sample C and D based on measurements of Samples A and B (intra-site); where X-axis = 15.2 million sequencing reads

IDT Exhibit 1020.5739

and Y-axis = 4.9 million sequencing reads.

[0038] **FIG. 15:** Graph showing inter-laboratory comparison of measurements (inter-site), that is Separate Library Preparations Sequenced at Different Sites (Inter-Site) where X-axis = 2.6 million sequencing reads and Y-axis = 0.4 million sequencing reads.

[0039] **FIGS. 16A-16B:** Graphs showing Receiver Curve to accurately detect fold changes based on **FIG. 13 (Results 4)**, showing Receiver Curve to Call Differential Expression Based on FIG. 14 - Results 4.

[0040] **FIG. 17:** Graph showing PCR-Driven Library Preparation Converges Native Target Concentrations Reducing Required Read Depth.

[0041] **FIGS. 18A-18B:** Standardized RNA Sequencing (STARSEQ) Workflow and Data Analysis

[0042] **FIGS. 19A-19C:** STARSEQ reduces oversampling without signal compression..

[0043] **FIGS. 20A-20B:** STARSEQ reduces required sequencing reads up to 10,000-fold.

[0044] **FIGS. 21A-21E:** Performance of STARSEQ with ERCC Reference Materials.

[0045] **FIGS. 22A-22F:** Performance of STARSEQ with endogenous cDNA targets.

[0046] **FIGS. 23A-23B:** Cross-platform comparison of STARSEQ with TaqMan qPCR and Illumina RNA-Sequencing.

[0047] **FIG. 24:** Difference plots between TaqMan and STARSEQ measurements.

[0048] **FIG. 25:** Difference plots between Illumina RNA-Sequencing and STARSEQ measurements.

[0049] **FIG. 26:** Assay Performance.

[0050] **FIG. 27:** STARSEQ "true negative" versus Taqman and RNA-sequencing.

[0051] **Fig. 28:** Standard Deviation of ERCC measurements.

DETAILED DESCRIPTION

[0052] Throughout this disclosure, various publications, patents and published patent specifications are referenced by an identifying citation. The disclosures of these publications, patents and published patent specifications are hereby incorporated by reference into the present disclosure to more fully describe the state of the art to which this invention pertains.

[0053] Described herein are methods for evaluating nucleic acids, and applications and business methods employing such compositions and methods. Some aspects of the present disclosure relate to improvements upon the Willey and Willey et al. U.S. Pat. Nos. 5,043,390; 5,639,606; 5,876,978 and 7,527,930, which are expressly incorporated herein by reference.

[0054] *Methods for Assessing a Nucleic Acid*

[0055] Described herein are methods for assessing amounts of a nucleic acid in a sample. In some embodiments, the method allows measurement of small amounts of a nucleic acid, for example, where the nucleic acid is expressed in low amounts in a specimen, where small amounts

of the nucleic acid remain intact and/or where small amounts of a specimen are provided.

[0056] "Specimen" as used herein can refer to material collected for analysis, e.g., a swab of culture, a pinch of tissue, a biopsy extraction, a vial of a bodily fluid e.g., saliva, blood and/or urine, etc. that is taken for research, diagnostic or other purposes from any biological entity.

[0057] Specimen can also refer to amounts typically collected in biopsies, e.g., endoscopic biopsies (using brush and/or forceps), needle aspirate biopsies (including fine needle aspirate biopsies), as well as amounts provided in sorted cell populations (e.g., flow-sorted cell populations) and/or micro-dissected materials (e.g., laser captured micro-dissected tissues). For example, biopsies of suspected cancerous lesions in the lung, breast, prostate, thyroid, and pancreas, commonly are done by fine needle aspirate (FNA) biopsy, bone marrow is also obtained by biopsy, and tissues of the brain, developing embryo, and animal models may be obtained by laser captured micro-dissected samples.

[0058] "Biological entity" as used herein can refer to any entity capable of harboring a nucleic acid, including any species, e.g., a virus, a cell, a tissue, an in vitro culture, a plant, an animal, a subject participating in a clinical trial, and/or a subject being diagnosed or treated for a disease or condition.

[0059] "Sample" as used herein can refer to specimen material used for a given assay, reaction, run, trial and/or experiment. For example, a sample may comprise an aliquot of the specimen material collected, up to and including all of the specimen. As used herein the terms assay, reaction, run, trial and/or experiment can be used interchangeably

[0060] In some embodiments, the specimen collected may comprise less than about 100,000 cells, less than about 10,000 cells, less than about 5,000 cells, less than about 1,000 cells, less than about 500 cells, less than about 100 cells, less than about 50 cells, or less than about 10 cells.

[0061] In some embodiments, assessing, evaluating and/or measuring a nucleic acid can refer to providing a measure of the amount of a nucleic acid in a specimen and/or sample, e.g., to determine the level of expression of a gene. In some embodiments, providing a measure of an amount refers to detecting a presence or absence of the nucleic acid of interest. In some embodiments, providing a measure of an amount can refer to quantifying an amount of a nucleic acid can, e.g., providing a measure of concentration or degree of the amount of the nucleic acid present. In some embodiments, providing a measure of the amount of nucleic acid refer to enumerating the amount of the nucleic acid, e.g., indicating a number of molecules of the nucleic acid present in a sample. The "nucleic acid of interest" may be referred to as a "target" nucleic acid, and/or a "gene of interest," e.g., a gene being evaluated, may be referred to as a target gene. The number of molecules of a nucleic acid can also be referred to as the number of copies of the nucleic acid found in a sample and/or specimen.

[0062] As used herein, "nucleic acid" can refer to a polymeric form of nucleotides and/or nucleotide-like molecules of any length. In certain embodiments, the nucleic acid can serve as a

IDT Exhibit 1020.5741

template for synthesis of a complementary nucleic acid, e.g., by base-complementary incorporation of nucleotide units. For example, a nucleic acid can comprise naturally occurring DNA, e.g., genomic DNA; RNA, e.g., mRNA, and/or can comprise a synthetic molecule, including but not limited to cDNA and recombinant molecules generated in any manner. For example the nucleic acid can be generated from chemical synthesis, reverse transcription, DNA replication or a combination of these generating methods. The linkage between the subunits can be provided by phosphates, phosphonates, phosphoramidates, phosphorothioates, or the like, or by nonphosphate groups, such as, but not limited to peptide-type linkages utilized in peptide nucleic acids (PNAs). The linking groups can be chiral or achiral. The polynucleotides can have any three-dimensional structure, encompassing single-stranded, double-stranded, and triple helical molecules that can be, e.g., DNA, RNA, or hybrid DNA/RNA molecules.

[0063] A nucleotide-like molecule can refer to a structural moiety that can act substantially like a nucleotide, for example exhibiting base complementarity with one or more of the bases that occur in DNA or RNA and/or being capable of base-complementary incorporation. The terms "polynucleotide," "polynucleotide molecule," "nucleic acid molecule," "polynucleotide sequence" and "nucleic acid sequence," can be used interchangeably with "nucleic acid" herein. In some specific embodiments, the nucleic acid to be measured may comprise a sequence corresponding to a specific gene.

[0064] In some embodiments the specimen collected comprises RNA to be measured, e.g., mRNA expressed in a tissue culture. In some embodiments the specimen collected comprises DNA to be measured, e.g., cDNA reverse transcribed from transcripts. In some embodiments, the nucleic acid to be measured is provided in a heterogeneous mixture of other nucleic acid molecules.

[0065] The term "native template" as used herein can refer to nucleic acid obtained directly or indirectly from a specimen that can serve as a template for amplification. For example, it may refer to cDNA molecules, corresponding to a gene whose expression is to be measured, where the cDNA is amplified and quantified.

[0066] The term "primer" generally refers to a nucleic acid capable of acting as a point of initiation of synthesis along a complementary strand when conditions are suitable for synthesis of a primer extension product.

[0067] *General Description of Method*

[0068] The preparation of a sequencing library involves some combination, or all, of the following steps: 1) nucleic acid fragmentation; 2) in vivo cloning, which serves to attach flanking nucleic acid adaptor sequences; 3) in vitro adaptor ligation; 4) PCR based adaptor addition; and, 5) unimolecular inversion probe type technology with, or without, polymerase fill-in, and ligation of probe to capture the sequence by circularization, with adaptor contained within the probe sequence.

IDT Exhibit 1020.5742

[0069] The definition of "nucleic acid adaptor" is that the "nucleic acid adaptor" can serve as any or all of the following: a) sequencing primer recognition site, b) barcode sequence of nucleotides to deconvolute the sample that was prepared for sequencing during analysis, and c) universal nucleic acid site which allows for multi-template amplification, or further addition of fusion-tail sequences through amplification.

[0070] The prepared sequencing library from one or more of steps 1-5 above is then analyzed on a sequencing instrument, and a representative sampling of the library is sequenced. The number of times that each unique nucleic acid target is observed then is counted, and the relative proportion between each unique nucleic acid targets' counts is assessed. This relative proportion, however, does not represent the true proportionality of abundance between each unique nucleic acid target in the original sample.

[0071] This loss of original representation is a technical artifact (e.g., error, bias) of steps 1-5. Further, this error is non-systematic, i.e., not the same amount of bias, between at least the following errors: i) library preparation steps (1-5 above); ii) preparation sequencing library replicates; iii) different time of replicates;, iv) different technicians preparing the library; and/or, v) preparing the library in a different laboratory.

[0072] As this non-systematic error in proportion of nucleic acid, in effect, targets the errors (i-v), any comparison of results between library preparations for the same sample are prone to error, thus limiting the application of sequencing as a tool for cost-effectively as well as reliably measuring nucleic acid copies.

[0073] One embodiment described herein is a method which utilizes a mixture of a known number (i.e., abundance, concentration and/or amount) of internal standard nucleic acid molecules corresponding to unique nucleic acid targets (also defined as 'native target" or NT) which are to be mixed in a nucleic acid sample prior to preparation of library for sequencing, or prior to sequencing (if library preparation is not required).

[0074] Each nucleic acid target is similar to its respective internal standard, with the exception of one or more changes to the nucleic acid sequence. These differences between native target and internal standard are identifiable with sequencing, and can include deletions, additions, or alteration to the ordering or composition of nucleotides used.

[0075] By introducing internal standards in a sample of nucleic acid targets prior to library preparation, the non-systematic error introduced by steps 1-5 (as well as sequencer instrument specific bias) is experienced by both the native target and the internal standard target similarly.

[0076] At the end of the sequencing, the proportion of sequencing events, (i.e., observations, counts, reads) between the native target and its respective internal standard is assessed, along with the original number of internal standard nucleic acid molecules input into the sample prior to library preparation, in order to quantifiably determine the original amount of each native target in the original sample prior to library preparation and sequencing.

IDT Exhibit 1020.5743

[0077] As the inclusion of internal standard thus controls for error and relative changes in proportion between native targets during steps 1-5 and subsequent sequencing, the method described herein also enables low-abundance native targets to be preferentially amplified (i.e., enriched) relative to higher-abundance native targets during library preparation. This preferential amplification or enrichment can be harnessed so that at the end of sequencing library preparation, the relative proportion between each unique native target will converge towards an equimolar (i.e., uniform) abundance in the library. This results in more equal coverage of sequencing depth between native targets. And, since the internal standard experiences the preferential amplification or enrichment as well, this method allows the original amount of each native target in the original sample to be quantifiably determined prior to library preparation.

[0078] In one non-limiting example, for every 10-fold reduction in depth of proportion between native targets, an approximate 10-fold reduction in direct sequencing cost is achieved, because 10-fold fewer sequencing reads are required.

[0079] The addition of a mixture of nucleic acid standards prior to sequencing library preparation (or prior to sequencing if library preparation is not required) thus provides an accurate quantification of native targets at end point with sequencing.

[0080] The use of a standardized mixture of nucleic acid internal standards enables a direct comparison of results between laboratories for nucleic acid molecular diagnostics and other quantitative sequencing results.

[0081] Also, in certain embodiment, the further addition of internal standards enables the convergence of native target abundance, thus reducing the direct sequencing costs by the fold-proportion native target abundances are normalized towards each other.

[0082] The inclusion of a mixture of internal standard nucleic acid of known amount (i.e., abundance, concentration and/or number) during library preparation provides certain advantages. Since it might not be known which of the steps 1-5 or sequencing might introduce error, the present method reduces this bias, thus enabling inter-library and inter-laboratory comparison of results, and at the same time, provides the ability to reduce direct-sequencing cost through nucleic acid target convergence of concentration.

[0083] In certain embodiments, the method described herein includes a known number of internal standard molecules for each gene to be measured in nucleic acid sample prior to sequencing, or prior to preparation of library for sequencing.

[0084] Also, in certain embodiments, the preparation of a standardized mixture of internal standards can be used by multiple laboratories, thereby increasing reliability of measurement of each targeted gene and increasing inter-experimental and inter-laboratory reproducibility of measurement. The measurement of copy number for each nucleic acid relative to a known number of copies of its respective internal standard molecules within a standardized mixture of internal standards, and use of the same SMIS across experiments and laboratories, thus increases the

IDT Exhibit 1020.5744

reliability and quality control by controlling for variation introduced by preparation of sequencing library.

[0085] In certain embodiments, the method described herein uses a gene specific reverse transcription and/or a PCR for library preparation for quantification by sequencing. In certain embodiments, the optimization of PCR enables the multiplexing of up 100, 300, 500, 1000, or more genes to yield sufficient PCR product of each targeted gene for quantification by sequencing. The optimization of PCR can bring about a convergence of initial inter-gene transcript representation by 10-fold, 100-fold, 1000-fold, 10,000-fold, or greater while maintaining ability to quantify initial relative transcript representation through measurement of each gene relative to its respective internal standard. Thus, the inclusion of known number of copies of internal standards in sample prior to library preparation (or prior to sequencing if library preparation is not required) controls for subsequent changes in transcript representation. It is now possible to optimize inter-gene convergence without losing the information regarding initial representation. For example, in certain embodiments, there can be a convergence of more than 1000-fold, resulting in a reduction of "read" requirement from 10,000,000 to 10,000.

[0086] Also, in certain embodiments where each chip for a typical next generation sequencer enables 10 million reads, this result enables increasing the number of samples analyzed/chip from 1 to 1,000. Currently, since a chip for a typical sequencer costs $1,000, the chip cost/sample is thereby reduced from about $1,000 to about $1.00.

[0087] In addition, rare transcripts can be measured with statistical significance. For example, the number of copies of a nucleic acid corresponding to a gene transcript can be determined, e.g., the number of copies/cell, where the gene is expressed in low copy number. Enumerating less than about 1,000 molecules can allow measurement of less than about 10 copies/cell of at least 100 different gene transcripts in a small biological specimen. The methods are capable of measuring and/or enumerating less than about 10 copies/cell of at least 100 different gene transcripts in a small biological specimen.

[0088] In still some embodiments, more measurements can be obtained from a given specimen and/or sample, e.g., of the size typically used to measure that few copies of a nucleic acid corresponding to one gene. For example, practice of some embodiments can measure and/or enumerate less than about 100, less than about 50, less than about 20, less than about 10, less than about 8, or less than about 5 copies/cell of at least about 20, at least about 50, at least about 80, at least about 100, at least about 120, at least about 150, or at least about 200 different nucleic acids in a sample, e.g., corresponding to different gene transcripts.

[0089] The expressed material may be endogenous to the biological entity, e.g., transcripts of a gene naturally expressed in a given cell type, or the expressed material to be measured may be of an exogenous nature. For example, the methods can be used to quantify transfected genes following gene therapy and/or a reporter gene in transient transfection assays, e.g., to determine

IDT Exhibit 1020.5745

the efficiency of transfection.

[0090] EXAMPLES

[0091] The methods and embodiments described herein are further defined in the following Examples, in which all parts and percentages are by weight and degrees are Celsius, unless otherwise stated. Certain embodiments of the present invention are defined in the Examples herein. It should be understood that these Examples, while indicating preferred embodiments of the invention, are given by way of illustration only. From the discussion herein and these Examples, one skilled in the art can ascertain the essential characteristics of this invention and without departing from the spirit and scope thereof, can make various changes and modifications of the invention to adapt it to various usages and conditions.

[0092] The PCR-based NGS library preparation method incorporates competitive internal amplification controls (IAC). This method both controls for the majority of non-systematic errors introduced during NGS library preparation, and enables inter-laboratory comparison of quantitative NGS data.

[0093] The competitive IAC thus controls for non-systematic error in PCR-based NGS library preparation by sharing identical priming sites to a native nucleic acid template of interest so as to mimic the kinetics of the native target in the PCR reaction, and thus control for target-specific variation in PCR efficiency.

[0094] In methods described herein, since the competitive IAC experiences the same kinetics as a native nucleic acid template, the proportional relationship of native target sequencing reads to its respective competitive IAC does not change during NGS library preparation.

[0095] Moreover, when the concentration of the competitive IAC placed into the sample preparation is known, it is now possible to accurately calculate the original abundance of native nucleic acid molecules that was present at the start of NGS library preparation.

[0096] As one example, when multiple laboratories used the same mixture of competitive IAC in multiple different studies, each of the multiple laboratories has shown that its results are concordant.

[0097] Thus, use of competitive IAC in PCR-based NGS library preparation enables cost-effective highly multiplexed analyses of multiple nucleic acid targets across multiple samples with a high degree of accuracy and reproducibility.

[0098] An additional benefit of incorporating competitive IAC is that protocols which result in normalization (i.e., convergence) of each native target toward equimolar (i.e., uniform) concentration, such as Multiplex PCR, can be implemented using such method.

[0099] It is to be understood that with normalization of template concentrations, it is now possible that abundance between native nucleic acid templates may vary greater than one million fold. In the past, the most highly represented native nucleic acid template would be unnecessarily oversampled and sequenced ten million times in order to sequence the least represented nucleic

IDT Exhibit 1020.5746

acid template (e.g., at least ten times to accurately detect a 2-fold change (Power=80%; Type 1 error rate=0.05)). However, use of the competitive IAC method described herein provides for the normalization in representation of the target analytes, yet still retains quantitative information of the original representation of both low- and high-abundance targets with a low number of sequencing reads. The reduction in oversampling of overrepresented nucleic acid targets thus results in reduced cost and stochastic sampling error associated with deep sequencing.

[00100] *Example 1*

[00101] *Multiplex-PCR with Competitive IAC for NGS Library Preparation and Subsequent Measurement of Nucleic Acid Abundance*

[00102] Reference material RNA titration pools used in the FDA-sponsored Sequencing Quality Control (SEQC) project (which already have nucleic acid abundance measured by multiple qPCR, Microarray and NGS platforms under a variety of conditions) were obtained.

[00103] NGS libraries were prepared from the reverse transcribed reference material using Multiplex-PCR in the presence of primers and competitive IAC for 150-gene targets.

[00104] The NGS library preparation was evaluated using Multiplex-PCR with competitive IAC for reproducibility of nucleic acid abundance measurement within individual test sites, between laboratories, and across different nucleic acid measurement platforms.

[00105] The costs and advantages of Multiplex-PCR with competitive IAC for NGS library preparation were compared to a commonly employed Illumina-based NGS library preparation protocol and Taqman qPCR for accurately measuring nucleic acid abundance in a clinical setting.

[00106] *Methods and Results*

[00107] Forward and reverse primers were designed that correspond to 101 basepair regions (i.e., amplicon) for each of 150 uniquely transcribed genes in the human genome. Each primer was designed with a uniform 68°C melting temperature. Each primer also contained a universal tail sequence that can be used for multi-template PCR, such as used in the addition of barcode and sequencing adaptor sequences after initial Multiplex-PCR. These primers were synthesized by Integrated DNA Technologies (IDT) and combined in equimolar ratio, and diluted to a final working concentration of 50 nMolar of each primer. A corresponding mixture of 150 competitive internal amplification controls (IAC) each 101 bases in length was synthesized by Integrated DNA Technologies (IDT). Each of the competitive IAC contained identical target-specific priming sites to their respective native nucleic acid template targets. Internal to these identical forward and reverse priming sites were six nucleotide changes in the internal portion of the sequence, so as to be able to differentiate a competitive IAC from its corresponding native target during post-sequencing data analysis.

[00108] Each competitive IAC was combined into a mixture at approximately equimolar concentration relative to each other by IDT. Because the mixing of competitive IAC may not have been in an exact 1:1 ratio, the absolute abundance of copies of each of the competitive IAC and

IDT Exhibit 1020.5747

their proportion in relation to each other for each of the 150 competitive IAC were determined by titration relative to a known amount of genomic DNA (gDNA) reference material. Genomic DNA reference material can serve as a normalizing reagent, because between each of the unique genomic sequences exist a one-to-one proportion to each other throughout the genome. Thus, perceived differences in competitive IAC concentration when titrated against gDNA, actually indicates a systematic difference in proportion that exists between competitive IAC in the mixture. This systematic difference is determined by the titration against a fixed amount of gDNA and is always applied to future calculations and measurements obtained using that particular lot or mixture of IAC (FIG. 1).

[00109] FIG. 1 shows the titration of a mixture of internal amplification controls (IAC) relative to a fixed amount of gDNA input of 100,000 copies into each Multiplex-PCR. Plotted on the Y-axis is the frequency or proportion of observed native reads divided by the sum of both native reads and its respective competitive IAC reads. On the X-axis is the initially estimated amount of each target in an approximately equimolar mixture of competitive IAC. 10 dilutions ranging from 10,000,000 copies of each IAC (Log10 Concentration = 0) to 1,000 (Log10 Concentration = -2) copies were input into each of 10 reactions to generate the curve shown. Of the 150 designed primer sets, competitive IAC, and respective native targets, 119 titrated with a goodness of fit (R^2 >0.95). Greater than 95% of the competitive IAC were within 10-fold of the expected equivalence point, (Native)/(Native + IAC) = 0.5, when diluted to 100,000 copies (10,000,000 starting IAC copies diluted to 100,000, or Log10 Concentration -2; i.e., 100-fold from 10,000,000). The new concentration served as the actual concentration for each of the 119 assays in the mixture of the competitive IAC and served as a reference of absolute accuracy (i.e., true accuracy) µg.

[00110] Thus, after testing the 150 assays against titrated mixture of IAC with fixed amount of gDNA, it was determined that 119 of the 150 assays had sufficient performance characteristics (Hill Plot $R^2 > 0.95$). These corrections were subsequently applied to all future measurements made with this mixture of IAC.

[00111] Phase III of the MAQC project, also known as the Sequencing Quality Control (SEQC) project, generated four pools from two RNA sample types: Universal Human Reference RNA (UHRR) from Stratagene and a Human Brain Reference RNA (HBRR) from Ambion. The four pools included the two reference RNA samples as well as two mixtures of the original samples: Sample A, 100% UHRR; Sample B, 100% HBRR; Sample C, 75% UHRR: 25% HBRR; and Sample D, 25% UHRR: 75% HBRR. This combination of biologically different RNA sources and known titration differences provided a method for assessing the accuracy of a platform based on the differentially expressed genes detected. Ten (10) µg aliquots of these RNA pools were used for Samples A, B, C and D.

[00112] Each of the RNA titration pool reference materials (Samples A, B, C, D) were reverse transcribed, as described in Canales et al., 2006, with the exception that Superscript III reverse

IDT Exhibit 1020.5748

transcriptase from Invitrogen was used in place of MMLV reverse transcriptase and 1 μg of RNA was placed in each reverse transcription reaction. In addition, Sample A was reverse transcribed twice from two separate preparations of reverse transcription master mix, so as to determine variance introduced by reverse transcription on sequencing library preparation.

[00113] One (1) μL from each of these 5 reverse transcribed RNA Titration Pools of cDNA (Samples A-RT1, A-RT2, B, C and D), was spiked into 1 of 12 Multiplex-PCR reactions containing serially diluted mixture of competitive internal amplification control (IAC) mixture representing 150 targets. These 12 serial dilutions of mixtures of competitive IAC range from 107 copies loaded, all the way to 103. A total of 12 μL of each sample was consumed during Multiplex PCR, corresponding to ~133 ng of RNA in total for each sample.

[00114] The concentration where native material was in equal concentration to competitive IAC (i.e., the equivalence point) for each gene target was determined in each reverse transcribed reference material (Samples SEQC A-RT1, A-RT2, B, C and D), and was determined using the Hill equation (FIG. 2).

[00115] The graph in FIG. 2 shows the titration of a mixture of internal amplification controls (IAC) relative to a fixed amount of 100,000 copies gDNA, or 11 ng of Reverse Transcribed RNA cDNA material (Samples SEQC A-RT1, A-RT2, B, C and D), input into each Multiplex-PCR. Plotted on the Y-axis is the frequency or proportion of observed native reads divided by the sum of both native reads and its respective competitive IAC reads. On the X-axis is the initially estimated amount of each target in an approximately equimolar mixture of competitive IAC. Dilutions ranging from 10,000,000 copies of each IAC (Log10 Concentration = 0) to 1,000 (Log10 Concentration = -2) copies were input into each of 10 reactions to generate the curve above.

[00116] Since Samples C and D represent a known cross titration between Samples A and B, the accuracy of the platform differentially expressed genes was assessed (FIG. 3). The values measured for SEQC Sample C-RT1 were compared to the values expected (% difference) based on measured SEQC Sample A and B values. The percent difference between the predicted signal C′ and the actual assay signal C was used as an indication of relative assay accuracy (RA). An RA score ΔC for a target gene was defined as (C–C′/C′), respectively. The distribution of percent difference from expected RA score for each gene is presented in a box plot for Standardized qNGS (n=88) and Standardized RT-PCR (n=201). Box plot components are: horizontal line, median; box, interquartile range; whiskers, 1.5× interquartile range; black squares, outliers.

[00117] Gene targets (n=88) measured between Samples SEQC A-RT1, B, C, were evaluated for inter-gene differential expression (DE) (FIG. 4A); DE was between 1.5 to 3.0, 2 to 3, 3 to 5, 5 to 10, or greater than 10-fold in change. Control for false positive or negative change was assessed by comparing SEQC A-RT1 versus SEQC A-RT2. The summary statistics showing differential expression in Sample C compared to expected based on Samples A and B is shown in FIG. 4B – Table 1.

IDT Exhibit 1020.5749

[00118] The assay reproducibility between reverse transcriptions is shown in FIG. 5. Two reverse transcriptions of SEQC Sample A (RT1 versus RT2) were measured for expression of 119 gene targets that successfully passed performance criteria in FIG. 1. Of the 119 gene targets, 97 had $R^2 > 0.95$ for curve fit of the Hill equation to determine equivalence point and concentration of each target (FIG. 2).

[00119] The same data as in FIG. 3 and FIG. 4 are depicted in FIG. 6. Expected measurements in Sample C (x-axis) versus observed measurements (y-axis). Reverse transcription of SEQC Sample C was measured for expression of 119 gene targets that successfully passed performance criteria in FIG. 1. Of the 119 gene targets, 88 had $R^2 > 0.95$ for curve fit of the hill equation to determine equivalence point and concentration of each target (y-axis) (FIG. 2).

[00120] The convergence and increased uniformity of 97 gene targets from FIG. 5 is shown in FIG. 7. Plotted on the X-axis is the data from FIG. 5 in proportion to highest abundance template. On the Y-axis is the actual proportion of sequencing reads that gene-target is in proportion to highest sequence template. Note that measurement and accuracy is not compressed (FIGS. 4-6), yet 75% of gene targets are within 10-fold sequencing read abundance to each other. That is, the sequencing depth decreased from approximately 1000-fold, down to 10-fold. This represents a 100-fold decrease in direct sequencing cost.

[00121] The ROC curve to detect >3 fold change for RNA-Sequencing using Illumina platform from Bullard et al. BMC Bioinformatics 2010, 11:94 is shown in FIG. 8. Compared to FIG. 4B, this ROC curve represents an ~ 75% accuracy of RNA-seq to detect a >3 fold change. Whereas standardized qNGS described herein has a greater than 97% accuracy (FIG. 4B). It is to be noted that the standardized qNGS method utilized a 10-fold sequencing depth in order to accurately detect a 3 fold change over a 1000-fold proportion difference between native targets. In contrast, traditional RNA-sequencing would require 100-fold more reads to arrive at a similar accuracy. In one example, 5 million sequencing reads were utilized to accurately quantify 97 genes using the standardized qNGS method. In comparison, traditional RNA-sequencing would have required well over 500 million reads for accurate quantification.

[00122] *Example 2*

[00123] *Quantitative Sequencing Following PCR-driven Library Preparation with Internal Standard Mixtures Has Improved Analytical Performance and Lower Cost*

[00124] Non-systematic biases introduced during preparation of next-generation sequencing (NGS) libraries as the primary source of technical variation have prevented application of NGS to measuring nucleic acid abundance in the clinical setting.

[00125] The costs of current qPCR clinical diagnostics are fixed to the cost of the chemistry (usually fluorescent) they use, and linearly associated with the number of nucleic acid targets they are interrogating. Further, each assay target requires a separate reaction vessel and multiple controls, which can become prohibitively expensive. These cumulative costs prevent the

IDT Exhibit 1020.5750

emergence of more complex clinical diagnostics based on measurement of multiple nucleic acid targets. More cost effective alternatives for multiplexed nucleic acid target abundance measurement are not as flexible in bringing new assay targets online cost-effectively without disrupting existing gene panels, or are not amenable to standardization and inter-site reproducibility of quantitative data. While NGS is amenable to cost-effective highly multiplexed quantitative analysis of multiple patient samples and nucleic acid targets there is a need for an efficient way to enable comparison of quantitative NGS results between sites, and to avoid the need for deep sequencing to accurately measure nucleic acid abundance.

[00126] In this example; a PCR-based NGS library preparation protocol that incorporated competitive internal amplification control (IAC) mixtures (i.e. internal standards) controlled for the majority of bias introduced during NGS library preparation, enabling clinical laboratories to offer cost effective moderately complex diagnostic panels from quantitative NGS data.

[00127] Reference material RNA titration pools used in the FDA-sponsored Sequencing Quality Control (SEQC) project were obtained (Samples A, B, C and D). Because the SEQC project RNA Samples C and D represent a known cross titration between SEQC project RNA Samples A and B, it is possible to compare SEQC expression value to measured and expected values for expression to determine accuracy of the method. Using Multiplex-PCR with primers and competitive IAC for 150-gene targets NGS libraries were prepared from: 1) gDNA to test general analytical performance, and 2) cDNA from reverse transcribed SEQC reference material to determine accuracy.

[00128] *Results:*

[00129] Using gDNA mixed with serially titrated competitive IAC mixtures as input, a linear dynamic range over 10^6 orders of magnitude was observed, with an average $R^2 = 0.995$ (0.993 – 0.997; 95% CI). The correlation coefficient of expected versus observed for Sample C was $R^2 = 0.96$, and Sample D was $R^2 = 0.94$, with an ROC curve-determined accuracy to detect a 3-fold change of 97% (95 – 99%; 95% CI). Inter-site correlation coefficient of measurements based on only 400,000 sequencing reads was $R^2 = 0.92$ across a linear dynamic range of ~10^5 orders of abundance between native targets.

[00130] The method described herein overcomes key sources of non-systematic bias introduced during NGS library preparation. This enables reproducible inter-laboratory and inter-platform quantitative NGS results, and a clear path to regulatory approval for clinical diagnostic applications.

[00131] The method described herein (an NGS with Internal Amplification Controls (IAC)) provides intra-site and inter-site reproducibility of quantitative next-generation sequencing (NGS) data. The method described herein also reduces need for deep sequencing, and hence direct sequencing cost, by converging the number of reads required to adequately sequence both rare and high abundance nucleic acid targets.

IDT Exhibit 1020.5751

[00132] FIG. 9 provides a schematic illustration of a PCR Master mix with a mixture of internal Amplification Controls (IAC). The IAC serves as a inter-library inter-site reference. The IAC is stable for a lengthy period of time (e.g., for years). The mixture of IACs control for PCR bias, and are present at a known concentration. The mixture of target specific primers includes hundreds of targets per reactions. The target specific primers contain a universal tail.

[00133] FIGS. 10A-10B are graphs showing the titration of a mixture of Internal Amplification Controls against gDNA and SEQC cDNA. The plot is in the format of a Dose-Response Curve for inhibition of an Enzymatic System. Taq polymerase is the enzyme. Inhibitor is the concentration of competitive Internal Amplification Controls (IAC). The dose-response is measured as the proportion of sequencing reads observed for the native genomic DNA (gDNA) or native complementary DNA (cDNA) target versus the sum of native and IAC sequencing reads. gDNA Plot represents 119 of 150 designed gene targets (~80% assay design success rate). The average correlation coefficient for fitting a three parameter fixed slope Hill Equation to each of the 119 assays was $R^2 = 0.995$ (0.993 – 0.997; 95% CI).

[00134] The average IC50 (Inhibitory Concentration 50%) was $\mathbf{10^{4.98}}$ with number of gDNA copies input being 10^5. Thus, titration of a mixture of internal amplification controls provides a true, not relative, accuracy of measuring copies of complex mixture of nucleic acids.

[00135] In FIGS. 10-11 (*Result 1*), the cDNA Plot represents 110 of 119 working gene target assays. Nine (9) assays had insufficient read depth of at least 1 sequencing read for both the native target as well as internal amplification control. The average IC50 (Inhibitory Concentration 50%) was determined for each nucleic acid target in SEQC Samples A, B, C and D under a variety of conditions and used in subsequent examples, Comparison of results were performed for:

[00136] FIG. 12 (*Result 2*) showing the same Library Preparation Replicate Sequencing (Intra-Site), where X-axis = 1.8 million sequencing reads and Y-axis = 3.0 million sequencing reads.

[00137] FIG. 13 (*Result 3*) showing the separate Library Preparation Sequenced (Intra-Site), where X-axis = 2.6 million sequencing reads and Y-axis = 4.8 million sequencing reads.

[00138] FIGS. 14A-14B (*Result 4*) showing predicting measurement of Sample C and D based on measurements of Samples A and B (intra-site); where X-axis = 15.2 million sequencing reads and Y-axis = 4.9 million sequencing reads.

[00139] FIG. 15 (*Result 5*) showing inter-laboratory comparison of measurements (inter-site), that is Separate Library Preparations Sequenced at Different Sites (Inter-Site) where X-axis = 2.6 million sequencing reads and Y-axis = 0.4 million sequencing reads.

[00140] FIGS. 16A-16B (*Result 6*) showing Receiver Curve to accurately detect fold changes based on FIG. 13 (Results 4), showing Receiver Curve to Call Differential Expression Based on FIG. 14 - Results 4.

[00141] FIG. 17 (*Result 7*) showing PCR-Driven Library Preparation Converges Native Target Concentrations Reducing Required Read Depth. The convergence of Native Template Amplicon

IDT Exhibit 1020.5752

Concentrations during PCR driven library preparation reduces number of sequencing reads to adequately sequence all targets. The Internal Amplification Controls provide the necessary reference point at the beginning of PCR driven sequencing library preparation to accurately measure each nucleic acid target despite convergence of template concentration (See FIGS. 12-16 - Results 2-6). In this example, direct sequencing depth is reduced by 1000-fold, and all targets are within 100-fold of each other.

It is also to be understood that it is within the contemplated scope of the present discolsure that the methods described herein include the use of moderate complexity clinical panels based on PCR-driven NGS library preparation with Internal Amplification Controls. Non-limiting examples include panels for: Lung Cancer Risk Test (15 gene); Lung Cancer Diagnostic Test (4 gene); Lung Cancer Chemo-Resist Test (20 gene), and BCR-ABL Fusion Transcript test (2 genes).

[00142] *Example 3*

[00143] *STAndardized RNA SEQuencing (STARSEQ)*

[00144] STAndardized RNA SEQuencing (STARSEQ) was assessed using two separate reference materials: 1) genomic DNA (gDNA) derived from the blood of a phenotypically normal individual (de-identified sample 723) at the University of Toledo Medical Center (UTMC) according to a protocol approved by the UTMC institutional review board, and 2) four reference RNA samples (A, B, C and D) provided by the FDA sponsored Sequencing Quality Control (SEQC) project (previously MAQC consortium). Sample A consists of Universal Human Reference RNA obtained from Stratagene. Sample B consists of Human Brain Reference RNA obtained from Ambion. For the SEQC project, samples A and B were then combined with Ambion External RNA Controls Consortium (ERCC) Spike-In Control RNA Mixes 1 and 2, respectively, so as to achieve a final concentration of 2% in samples A and B based on Total RNA concentration.

[00145] Each spike-in mix of ERCC RNA controls contains the same controls spanning a dynamic range greater than 10^6, but in different formulations. Within each formulation mix are 4 subgroups that exhibit known fold differences in abundance between mix 1 and 2; 0.5x, 0.67x, 1.0x and 4.0x-fold difference. Samples A and B were then combined in 3:1 and 1:3 proportional mixtures to create samples C and D, respectively. The gDNA "reference" material represents a sample where the majority of endogenous targets are in a very close 1:1 proportion to each other. Whereas, samples A-D represent a complex mixture of synthetic (ERCC controls) and endogenous RNA targets in known proportions that can be used as ground truth benchmarks for assessing a method's analytical performance characteristics across a greater than 10^6 fold dynamic range of abundances.

[00146] *Reverse transcription of RNA reference materials*

[00147] Ten micrograms each of samples A-D reference RNA materials at a concentration of 1 µg/µL were obtained from the FDA sponsored SEQC project (fda.gov/ScienceResearch

IDT Exhibit 1020.5753

/BioinformaticsTools /Microarray Quality Control Project). For each sample two 2 μg aliquots of RNA were reverse transcribed. Each reverse transcription reaction took place in a 90 μL volume using manufacturer's protocol for Superscript III reverse transcription (Life Technologies) and oligo(dT) priming. After reverse transcription, the two 90 μL cDNA products for each sample were combined into a single 180 μL volume (reverse transcription 1; RT1). For sample A, an additional set of two 2 μg aliquots of RNA were reverse transcribed using a separate master mix (reverse transcription 2; RT2).

[00148] *STARSEQ assay target selection*

[00149] The MicroArray Quality Control (MAQC) consortium previously selected a list of 1,297 genes to evaluate performance of multiple qPCR and microarray platforms). From this list, 150 endogenous targets were selected to develop STARSEQ assays. These 150 assays were chosen, in part, because the gene targets they represent are expressed over a greater than 10^6 dynamic range. These reagents were used to measure absolute as well as relative proportion of each gene target in gDNA and reverse transcribed reference RNA samples A-D. In addition, 28 of 92 External RNA Control Consortium (ERCC) targets were also selected to develop STARSEQ assays.

[00150] *STARSEQ primer design and synthesis*

[00151] Forward and reverse PCR primers were designed to corresponding 101-bp amplicon regions for each of 150 uniquely transcribed genes in the human genome and 28 ERCC targets. Each forward and reverse primer set was designed with a uniform 68°C melting temperature using Primer3 software (Untergasser et al, NAR, 2012). In order to minimize off-target priming, primer pair specificity was verified using GenomeTester 1.3 to identify any additional amplicons less than 1000 bp in size. Each primer also contains a universal tail sequence not present in the human genome, which can be used for multi-template PCR addition of barcode and platform specific sequencing adapters. The forward universal tails are identical in sequence adapters used for arrayed primer extension (APEX-2), while the reverse tail sequence is the same as the forward with the exception of the last four 3' bases, enabling directionality during sequencing. Target specific primers with universal tails for the 150 endogenous targets and 28 ERCC targets were synthesized by Integrated DNA Technologies (IDT) and Life technologies, respectively. A primer pool for endogenous or ERCC targets was created by combining synthesized primers in equimolar ratio, and diluting to a final working concentration of 50 nM for each primer in dilute Tris-EDTA buffer.

[00152] *STARSEQ competitive internal standard mixture design and synthesis*

[00153] Each 101-bp competitive internal standard (IS) was designed to retain identical target specific priming sites to their respective native nucleic acid target (FIGS. 18A-18B). Internal to these identical priming sites are six nucleotide changes, so as to be able to differentiate a competitive IS from its corresponding native target during post-sequencing data analysis. The 150

IDT Exhibit 1020.5754

competitive IS corresponding to the endogenous targets were synthesized by Integrated DNA Technologies (IDT), and the 28 competitive IS corresponding to ERCC targets were synthesized by Life Technologies.

[00154] For the 150 competitive IS templates corresponding to endogenous targets, concentration was measured by optical density at IDT, and subsequently combined in a 1:1 stochiometric molar ratio based on these measurements. Concentration of each IS was determined empirically by cross-titrating the mixture relative to a fixed gDNA input of 100,000 copies (ID 723). In gDNA from a phenotypically healthy individual, it is now believed that the majority of loci would be in a 1:1 proportion to each other, providing a reasonable and cost-effective reference material to determine the actual concentration for each competitive IS template.

[00155] For the 28 competitive IS templates corresponding to ERCC targets, no such reference material exists for normalization. Thus, each standard was separately amplified with forward and reverse primers (without universal sequences), column purified (QIAquick PCR purificiation kit), visualized and quantified for only a single peak at 101-bases on an Agilent 2100 Bioanalyzer using DNA Chips with DNA 1000 Kit reagents according to manufacturer's protocol (Agilent Technologies Deutschland GmbH, Waldbronn, Germany). Quantified standards were then combined in a 1:1 stoichiometric molar ratio, to create a stock concentrated mixture of internal standards (IS). Both the endogenous and ERCC target mixtures of competitive IS were then serially diluted to working concentrations and used in all subsequent experiments as a reference mixture for quantifying absolute copies of each transcript in samples A-D (FIGS. 18A-18B).

[00156] *Multiplex competitive PCR with universally tailed target specific primers*

[00157] For each multiplex competitive polymerase chain reaction (PCR), a 10 μL reaction volume was prepared containing: 1 μL of native templates, 1 μL of competitive IS mixture at varying input concentrations, 1 μL of corresponding primer-mix, 1 μL of 2 mM dNTPs, 1 μL of 10x Idaho Technology reaction buffer with 30 mM $MgCl_2$, 0.1 μL of Promega GoTaq Hot Start Taq polymerase (5u/μL) and 4.9 μL of RNAse free water (FIG. 18A). Genomic DNA was spiked into 10 separate multiplex-PCR reactions containing serially diluted mixture of competitive IS mixture representing 150 endogenous targets. These 10 dilutions represent a series of 3-fold dilutions of IS mixture ranging in abundance from $2x10^6$ - 10^3 copies loaded. Sample A-D cDNA for RT1 was spiked into 5 separate multiplex-PCR reactions containing serially diluted mixture of competitive IS mixture representing 28 ERCC targets. These 5 dilutions represent a series of dilutions of IS mixture; 10^6, 10^5, 10^4, 10^3 and 300 copies loaded. Reverse transcribed RNA for samples A (RT1 and RT2), B, C and D were spiked into 12 separate multiplex-PCR reactions containing serially diluted mixture of competitive IS mixture representing 150 endogenous targets. These 12 dilutions represent a series of 3-fold dilutions of IS mixture ranging in abundance from $6x10^7$ - $3.4x10^2$ copies loaded. A total of 17 μL of each cDNA sample was consumed during multiplex competitive PCR, corresponding to ~377 ng of RNA for each sample.

IDT Exhibit 1020.5755

[00158] *Standardized RNA Sequencing (STARSEQ) Workflow and Data Analysis*

[00159] FIG. 18A NT = Native Target (e.g. cDNA, gDNA, etc.); IS = Internal Standard, a ssDNA or dsDNA molecule that a) is homologous to a specific native target at the primer sequences and therefore competes for amplification with the native target, but b) contains one or more base substitutions internal to the primer sites and therefore is distinguishable from the native target. The IS template for each gene is in a fixed relationship relative to the IS for the other genes in an internal standard mixture.

[00160] FIG. 18B shows the proportional relationship among native targets in the original sample is preserved during amplification and sequencing because a) the competition between each NT and its respective IS preserves the original concentration for each NT, and b) the IS are in a fixed relationship relative to each other. Determining the abundance of native target in the original sample is obtained by multiplying the ratio of sequencing counts for NT and IS (NT:IS) by the concentration of internal standard (IS) loaded into the amplicon library preparation (i.e., equivalence point determination). Native targets for which values could not be measured across at least three dilution points are not shown. FIG. 18B - upper panel: shows tinearity of cross titrating competitive Internal Standard Mixture with constant amount of genomic DNA (gDNA) for 123 targets. Dotted lines represent 95% prediction interval for NT:IS ratio values. FIG. 18B - middle panel: shows the linearity of cross titrating competitive Internal Standard Mixture with constant amount of 26 ERCC native targets from samples A, B, C and D. Each ERCC target is at a different concentration spanning a greater than 10^6 dynamic range in abundance. FIG. 18B - bottom panel shows the linearity of cross titrating competitive Internal Standard Mixture with constant amount of endogenous cDNA native targets from samples A, B, C and D (same targets as assessed in gDNA; upper panel).

[00161] *Touchdown PCR with multiplex competitive PCR*

[00162] Increasing level of multiplex PCR requires a commensurate decrease in the concentration of primers used. Decreasing primer concentration has two predominant effects in multiplex PCR: **1)** reduces formation of primer-dimer products, and **2)** plateaus amplicon product formation early preventing dNTPs from becoming a limited reagent (less primer method). This latter effect is important as it enables all target templates to reach plateau phase, and in the presence of competitive IS drastically reduces oversampling/sequencing of high abundance targets without signal compression (FIGS. 19A-19C, FIGS. 20A-20B).

[00163] *STARSEQ reduces oversampling without signal compression*

[00164] FIG. 19A depicts two native targets (NT) within a hypothetical cDNA sample. One native target is in high abundance, 10^8 copies ("Abundant" NT), while another is in low abundance, 10^2 copies ("Rare" NT), representing a one-million fold difference in abundance between targets. This hypothetical cDNA sample is combined with a mixture of internal standards (IS) with a fixed relationship of concentrations at 10^5 copies.

IDT Exhibit 1020.5756

[00165] FIG. 19B depicts the multiplex competitive PCR library preparation for FIG. 19A. The PCR amplification plots for both the "Abundant" and "Rare" NT are separated for purposes of clarity, but occur in the same reaction. During multiplex competitive PCR, each native target competes equally with its respective competitive internal standard for dNTPs, polymerase and a limiting concentration of primers. Because the starting concentration of each target's primer-pair is the same, each competitive reaction will plateau around the same end-point concentration (~10^9 copies).

[00166] In FIG. 19C, the equal competition between each NT and respective IS preserves the proportional relationship between Native Targets in the original sample, allowing for measurement of native target abundance without signal compression. Yet, a 10^6 fold range of templates is reduced to 10^3 after multiplex competitive PCR library preparation resulting in a 1,000 fold reduction in oversampling of the high abundance target.

[00167] Mixing a sample of native targets in multiple ratios with IS mixture (FIG. 18A) results in a greater degree of uniformity in template concentration than what is obtainable with only one internal standard spike-in (FIG. 19A).

[00168] *STARSEQ reduces required sequencing reads up to 10,000-fold*

[00169] FIG. 20A shows the actual proportional sequencing data for ERCC (n=104) and endogenous (n=400) cDNA targets. X-axis represents the proportional abundance of each target in a library preparation normalized to the lowest abundance target (set to 10^0). Y-axis is in units of proportional sequencing reads (coverage) required to sequence the lowest abundance target at least once.

[00170] FIG. 20B is a tabular summary of FIG. 20A where the number of sequencing reads represents the sum of all sequencing reads to observe all targets at least once. Required number of traditional RNA-sequencing reads are calculated based on an assumed 1:1 relationship between target copies present in the library, and sequencing coverage required. Fold reduction in required sequencing reads by STARSEQ is the quotient of traditional RNA-sequencing and STARSEQ sequencing reads.

[00171] However, there is a limit to which one can dilute primers and still successfully amplify targets of interest. This limit can be pushed lower through several approaches: 1) increase primer melting temperature, and 2) increase the time during which annealing occurs to allow for eventual primer binding. Both of these solutions can exacerbate off-target priming. This apparent obstacle can is now shown herein to be able to be remedied by use of a modified touchdown PCR protocol. In this protocol, high annealing temperatures are incorporated during initial cycles of PCR to increase stringency of primer binding reducing off-target priming. In subsequent cycles annealing temperature is gradually lowered resulting in increased yield once sufficient specific product has formed during earlier high stringency cycles. Using this framework, the following protocol was developed: Each multiplex competitive reaction mixture was cycled in an air thermocycler

IDT Exhibit 1020.5757

(RapidCycler (Idaho Technology, Inc. Idaho Falls, Idaho) under modified touchdown PCR conditions with low primer concentration: 95°C/3 min (Taq activation); 5 cycles of 94°C/30 s (denaturation), 72°C/4 min (annealing), and 72°C/15 sec (extension); repeat 5 cycles with annealing temperature decreased 1°C to 71°C; iterate 1°C decrease and 5 cycles until annealing temperature is 64°C (total of 45 cycles).

[00172] In particular embodiments, Hot Start Taq polymerase is used, as off-target priming and enzymatic activity is sufficiently high during reaction preparation that only primer-dimer product will otherwise be seen.

[00173] *Performance of STARSEQ with ERCC Reference Materials*

[00174] FIG. 21A shows the measured signal abundance of ERCC targets in samples A, B, C and D. Points represent the median of ERCC measurements from those library preparations with at least 15 sequencing reads for both the NT and IS. X-axis units are derived from Ambion product literature for the known concentration of ERCC spike-in controls, SEQC project material preparation protocols, and an assumed 100% reverse transcription yield for each target.

[00175] FIG. 21B shows difference plots of data in FIG. 21A ordered numerically by ERCC ID. Each ERCC target depicted was measured at least once in all four samples A-D. For purposes of clarity, ERCC-170 is highlighted orange in FIG 21A and FIG. 21B.

[00176] In FIG. 21C, samples C and D represent a 3:1 and 1:3 mixture, respectively, of Total RNA from samples A and B. These ratios were used to calculate expected measurements for samples C and D (x-axis) from measurements of A and B, and plotted against actual measurements of samples C and D (y-axis) (n=52).

[00177] In FIG. 21D, points represent standard deviation in measurements of ERCC targets in SEQC A, B, C and D, for those assays with at least two IS dilution points that had at least 15 sequencing reads for both the NT and IS. The red line depicts the expected standard deviation based on a Poisson sampling distribution plus a baseline 0.08 technical replicate standard deviation.

[00178] FIG. 21E shows ROC curves to detect fold change with corresponding area under the curve (AUC) with 95% confidence intervals. ROC curves are derived from the comparison of differential ratio subpools of ERCC targets in samples: A vs B, A vs C, A vs D, B vs C, B vs D and C vs D. Results for 1.1-fold change represent a range of differential ratio subpools [1.05 – 1.174] (controls n=100, tests n=96); 1.25 [1.175 – 1.374] (controls n=163, tests n=163); 1.5 [1.375 – 1.74] (controls n=229, tests n=227); 2.0 [1.75 – 2.49] (controls n=229, tests n=223); ≥4.0 [2.5 – 10.0] (controls n=286, tests n=290).

[00179] *Performance of STARSEQ with endogenous cDNA targets*

[00180] Absolute signal abundance of cDNA targets in sample A in units of copies per library preparation were measured on separate days, different sites (OU = Ohio University; UTMC = University of Toledo Medical Center), and between different reverse transcription preparations

IDT Exhibit 1020.5758

(RT1 and RT2). Points represent the median of ERCC measurements from those library preparations with at least 15 sequencing reads for both the NT and IS. FIG. 22A shows the inter-day effect (n=88). FIG. 22B shows the inter-day and Inter-site effect (n=81). FIG. 22C shows the inter-day and Inter-library effect (n=92). FIG. 22D shows the inter-day, Inter-site and Inter-library effect (n=80). FIGs. 22E-22F show that samples C and D represent a 3:1 and 1:3 mixture, respectively, of Total RNA from samples A and B. These ratios were used to calculate expected measurements for samples C and D (x-axis) from measurements of A and B, and plotted against actual measurements of samples C (n=86) and D (n=90) (y-axis).

[00181] *Cross-platform comparison of STARSEQ with TaqMan qPCR and Illumina RNA-Sequencing.*

[00182] The average of differences for measurements of samples A and B between STARSEQ and TaqMan qPCR (FIG. 24 showing difference plots between TaqMan and STARSEQ measurements) or Illumina RNA-sequencing (FIG. 25 showing difference plots between Illumina RNA-Sequencing and STARSEQ measurements) was determined for each endogenous target. This difference was subtracted from TaqMan qPCR or Illumina RNA-sequencing measurements for samples C and D and plotted (x-axis) against STARSEQ measurements of C and D (y-axis).

[00183] STARSEQ measurements represent the median measurement from library preparations that had at least 15 sequencing reads for both the NT and IS. FIG. 26A shows a comparison of TaqMan qPCR with STARSEQ (n=292). FIG. 26B shows a comparison of Illumina RNA-Sequencing with STARSEQ (n=340).

[00184] *Assay Performance*

[00185] Assay measurement performance as assessed in SEQC samples A, B, C and D for ERCC as well as endogenous cDNA targets, as shown in FIG. 26. Endogenous targets were also assessed against gDNA control (see FIG. 18B).

[00186] True negative measurements occur when sufficient number of competitive internal standard was sequenced (sequenced at least 15 times), but insufficient native template was observed across all spike-in concentrations of internal standard. An upper limit of expression for these assays can still be calculated as [1/(IS sequencing counts)] x concentration IS loaded into the library preparation with the lowest IS concentration present. These measurements represent true negative measurements and the lower limit of accurate quantification can be determined from these data.

[00187] Failed assays are measurements where "sequencing depth was too low" for both the NT and IS. These represent true assay failures (neither native nor internal standard was sequenced at least 15 times). In this way, competitive IS mixtures can control for false negative reporting.

[00188] *Addition of barcodes and sequencing adaptors*

[00189] A set of fusion primers were designed with their 3'-end complementary to the universal APEX-2 sequence tails added during multiplex competitive PCR. These fusion primers

IDT Exhibit 1020.5759

are tailed with a four nucleotide index/barcode sequence and 5' to that, a forward or reverse ion torrent amplicon sequencing adapter (FIG. 26). Both forward and reverse sequencing primers were intentionally barcoded to dual index each sample and reduce likelihood of false-indexing a sequence read; both barcodes must match. For each barcoding reaction, a 10 μL reaction volume was prepared containing: 1 μL of multiplex competive PCR product, 1 μL of 1 μM forward and reverse barcoding primer, 1 μL of 2 mM dNTPs, 1 μL of 10x Idaho Technology reaction buffer with 30 mM $MgCl_2$, 0.1 μL of Promega GoTaq Hot Start Taq polymerase (5u/μL) and 4.9 μL of RNAse free water. Each barcoding reaction was cycled in an air thermocycler (RapidCycler (Idaho Technology, Inc. Idaho Falls, Idaho) under the following conditions: 95°C/3 min (Taq activation); 15 cycles of 94°C/5 s (denaturation), 58°C/10 sec (annealing), and 72°C/15 sec (extension). Reaction vessels are immediately removed and kept at 4°C during all subsequent steps. The goal during this step is to prevent heterodimerization of barcoded product. Depending on the type of heterodimerization, post-sequencing alignment errors can arise from false sequencing base calls with resultant decrease in measurement precision and accuracy. Newly barcoded multiplex competitive PCR sequencing libraries are then individually quantified on an Agilent 2100 Bioanalyzer using DNA Chips with DNA 1000 Kit reagents according to manufacturer's protocol (Agilent Technologies Deutschland GmbH, Waldbronn, Germany). Uniquely barcoded sequencing libraries are then mixed in a known stoichiometric ratio so as to optimize the percentage of sequencing reads that each library will eventually receive; in most cases 1:1 is used.

[00190] *STARSEQ "true negative" versus Taqman and RNA-sequencing*

[00191] 26 STARSEQ measurements had sufficient data to report back a less than measurement. Of the 26 measurements, TaqMan reported not detected (ND) for 14, and RNA-Seq reported ND for 1 (see FIG. 27). Because STARSEQ could detect IS, but not accurately quantify NT present, these represent False Negative detections for TaqMan and RNA-Seq. Less than measurements were calculated as [1/(IS sequencing counts)] x concentration IS loaded into the library preparation.

[00192] *Standard Deviation of ERCC measurements.*

[00193] FIG. 28 shows the SD of differences is calculated from data presented in FIG. 21. Intra-assay Intra-sample SD is calculated from the median of intra-assay SD within each sample A-D. Intra-assay Inter-sample SD is calculated from the median of intra-assay SD across samples A-D. Inter-assay Inter-sample SD is calculated from the median of inter-assay SD of residuals across samples A-D. Since the SD is reported in Log_{10} values, it is roughly equivalent to reporting of the coefficient of variation (CV).

[00194] *Product Purification and Sequencing*

[00195] In certain embodiments, it is necessary during the purification of barcoded sequencing libraries that a system does not use strong denaturants or chaotropic salts, such as guanidine

IDT Exhibit 1020.5760

hydrochloride or thiocyanate. These agents result in downstream template heterodimerization, false sequencing base calls and post-sequencing alignment errors. For this reason, each mixture of barcoded sequencing libraries were purified using Life Technologies E-Gel SizeSelect 2% Agarose gels, which does not report the use of denaturants or chaotropic salts, and can be run in a refrigerated room to prevent heat denaturation during electrophoretic separation. Purified sequencing libraries were then quantified using the KAPA Library Quantification Kit for Ion Torrent Sequencing Platforms (Kapa Biosystems). Based on this quantification, libraries were diluted appropriately and prepared for Ion Torrent PGM Sequencing service according to manufacturer's recommendations at the University of Toledo Medical Center (UTMC), Toledo, OH and Ohio University (OU), Athens, OH.

[00196] *FASTQ file processing*

[00197] Raw sequencing data from an NGS service were provided back in FASTQ format. Sequencing reads were extracted and each sequencing read was parsed into 3 separate FASTQ files: 1) forward (*query-barcode.fastq*) and 2) reverse barcode (*query-revbarcode.fastq*) regions, as well as 3) central portion of the amplicon (*query-subject.fastq*) corresponding to the region internal to target specific priming sites where six nucleotide substitutions should exist between NT and matching competitive IS.

[00198] *BFAST of sequences against subject database*

[00199] Each of the three FASTQ files were aligned with known reference FASTA databases corresponding to whether it was a barcode (*barcode.fa*) or amplicon region (*subject.fa*) using the BLAT-like fast, accurate search tool (BFAST, version 0.7.0a), with file output in sequence alignment/map (SAM) format. BFAST match against the index databases and SAM file output was performed for the trimmed FASTQ files containing 1) forward barcode, 2) reverse barcode and 3) captured amplicon subject sequences.

[00200] *Binning of sequence counts*

[00201] Each of the three SAM files from 1) forward and 2) reverse barcode, and 3) amplicon region were then merged into a practical extraction and reporting language (PERL) hash table using the sequence read ID as a key for matching (http://www.perl.org/). Based on barcode and amplicon alignment, each sequencing read was binned into an array corresponding to the IS input concentration for a given sample preparation, and whether it was called as an NT or IS by BFAST alignment. If the forward and reverse barcode alignment calls did not match, the sequence read was not binned. The resulting hash table of binned sequencing reads is output in comma delimited format and processed as outlined in the Statistical Methods section.

[00202] *Measuring relative abundance*

[00203] At least 14 sequencing reads were required for each of the NT and IS. Correct fold-dilution was determined based on change in NT:IS ratio across multiple assay targets and across multiple serially diluted internal standard spike-ins. The dilution of internal standard was then

IDT Exhibit 1020.5761

multiplied by NT:IS ratio. Each assay had multiple measurements per assay because of multiple dilutions of internal standard. If the STDEV of these measurements is less than 10-fold in variance, the median of these measurements was accepted. Correct measurements were based on predetermined assay systematic bias of internal standard concentrations. The population of these measurements was normalized to a population median.

[00204] *STARSEQ measurement inclusion/exclusion criteria*

[00205] Each native target (gDNA or cDNA) was measured relative to its respective internal standard within a cross-titrated concentration of the ISM (FIG. 18). An empirical threshold of at least 15 sequencing reads each for native target (NT) and respective competitive internal standard (IS) was the optimal inclusion/exclusion criterion to consider a NT:IS ratio valid (power>80%; type 1 error rate<0.05; to detect 2-fold NT:IS ratio change) (FIG. 18). For those assays with more than one measurement that met criteria above, a coefficient of variation (CV) of >1000% between measurements triggered exclusion for that assay measurement in that particular sample.

[00206] *Statistical Methods: Estimate of Native Target Concentration*

[00207] For each gene target and technical replicate with input concentration of each IS mixture indexed with the subscript *i*, an estimate of the concentration of the native target (NC_i) was calculated based on the observed/binned sequence counts of both the native target (NT_i) and internal standard (IS_i), as well as the known starting concentration (in units of template copies per library preparation) of the internal standard (SC_i):

$$\log_{10} NC_i = \log_{10} \frac{NT_i}{IS_i} + \log_{10} SC_i$$

[00208] The empirically determined optimal method and QC parameter for estimating the summarization quantity was, 1) the median (NC_{median}) of NC_i technical replicate measures that have, 2) at least 15 sequencing counts for both NT_i as well as IS_i, and 3) coefficient of variation (CV) across NC_i of less than 1.00 on a base 10 logarithm scale. This was selected so as to have sufficient sampling of a given target to enable the detection of a 2-fold change in abundance between targets with a type 1 error rate of less than 0.05, and a type 2 error rate less than 0.20.

[00209] *Example 4*

[00210] *Non-limiting Examples of Applications*

[00211] In some embodiments, a method for obtaining a numerical index that indicates a biological state comprises providing 2 samples corresponding to each of a first biological state and a second biological state; measuring and/or enumerating an amount of each of 2 nucleic acids in each of the 2 samples; providing the amounts as numerical values that are directly comparable between a number of samples; mathematically computing the numerical values corresponding to

IDT Exhibit 1020.5762

each of the first and second biological states; and determining a mathematical computation that discriminates the two biological states. First and second biological states as used herein correspond to two biological states of to be compared, such as two phenotypic states to be distinguished. Non-limiting examples include, e.g., non-disease (normal) tissue vs. disease tissue; a culture showing a therapeutic drug response vs. a culture showing less of the therapeutic drug response; a subject showing an adverse drug response vs. a subject showing a less adverse response; a treated group of subjects vs. a non-treated group of subjects, etc.

[00212] A "biological state" as used herein can refer to a phenotypic state, for e.g., a clinically relevant phenotype or other metabolic condition of interest. Biological states can include, e.g., a disease phenotype, a predisposition to a disease state or a non-disease state; a therapeutic drug response or predisposition to such a response, an adverse drug response (e.g. drug toxicity) or a predisposition to such a response, a resistance to a drug, or a predisposition to showing such a resistance, etc. In preferred embodiments, the numerical index obtained can act as a biomarker, e.g., by correlating with a phenotype of interest. In some embodiments, the drug may be and anti-tumor drug. In certain embodiments, the use of the method described herein can provide personalized medicine.

[00213] In certain embodiments, the biological state corresponds to a normal expression level of a gene. Where the biological state does not correspond to normal levels, for example falling outside of a desired range, a non-normal, e.g., disease condition may be indicated.

[00214] A numerical index that discriminates a particular biological state, e.g., a disease or metabolic condition, can be used as a biomarker for the given condition and/or conditions related thereto. For example, in some embodiments, the biological state indicated can be at least one of an angiogenesis-related condition, an antioxidant-related condition, an apoptosis-related condition, a cardiovascular-related condition, a cell cycle-related condition, a cell structure-related condition, a cytokine-related condition, a defense response-related condition, a development-related condition, a diabetes-related condition, a differentiation-related condition, a DNA replication and/or repair-related condition, an endothelial cell-related condition, a hormone receptor-related condition, a folate receptor-related condition, an inflammation-related condition, an intermediary metabolism-related condition, a membrane transport-related condition, a neurotransmission-related condition, a cancer-related condition, an oxidative metabolism-related condition, a protein maturation-related condition, a signal transduction-related condition, a stress response-related condition, a tissue structure-related condition, a transcription factor-related condition, a transport-related condition, and a xenobiotic metabolism-related condition. In other specific embodiments, antioxidant and xenobiotic metabolism enzyme genes can be evaluated in human cells; micro-vascular endothelial cell gene expression; membrane transport genes expression; immune resistance; transcription control of hormone receptor expression; and gene expression patterns with drug resistance in carcinomas and tumors.

IDT Exhibit 1020.5763

[00215] In some embodiments, one or more of the nucleic acids to be measured are associated with one of the biological states to a greater degree than the other(s). For example, in some embodiments, one or more of the nucleic acids to be evaluated is associated with a first biological state and not with a second biological state.

[00216] A nucleic acid may be said to be "associated with" a particular biological state where the nucleic acid is either positively or negatively associated with the biological state. For example, a nucleic acid may be said to be "positively associated" with a first biological state where the nucleic acid occurs in higher amounts in a first biological state compared to a second biological state. As an illustration, genes highly expressed in cancer cells compared to non-cancer cells can be said to be positively associated with cancer. On the other hand, a nucleic acid present in lower amounts in a first biological state compared to a second biological state can be said to be negatively associated with the first biological state.

[00217] The nucleic acid to be measured and/or enumerated may correspond to a gene associated with a particular phenotype. The sequence of the nucleic acid may correspond to the transcribed, expressed, and/or regulatory regions of the gene (e.g., a regulatory region of a transcription factor, e.g., a transcription factor for co-regulation).

[00218] In some embodiments, expressed amounts of more than 2 genes are measured and used in to provide a numerical index indicative of a biological state. For example, in some cases, expression patterns of multiple genes are used to characterize a given phenotypic state, e.g., a clinically relevant phenotype. In some embodiments, expressed amounts of at least about 5 genes, at least about 10 genes, at least about 20 genes, at least about 50 genes, or at least about 70 genes may be measured and used to provide a numerical index indicative of a biological state. In some embodiments of the instant invention, expressed amounts of less than about 90 genes, less than about 100 genes, less than about 120 genes, less than about 150 genes, or less than about 200 genes may be measured and used to provide a numerical index indicative of a biological state.

[00219] Determining which mathematic computation to use to provide a numerical index indicative of a biological state may be achieved by any methods known in the arts, e.g., in the mathematical, statistical, and/or computational arts. In some embodiments, determining the mathematical computation involves a use of software. For example, in some embodiments, a machine learning software can be used.

[00220] Mathematically computing numerical values can refer to using any equation, operation, formula and/or rule for interacting numerical values, e.g., a sum, difference, product, quotient, log power and/or other mathematical computation. In some embodiments, a numerical index is calculated by dividing a numerator by a denominator, where the numerator corresponds to an amount of one nucleic acid and the denominator corresponds to an amount the another nucleic acid. In certain embodiments, the numerator corresponds to a gene positively associated with a given biological state and the denominator corresponds to a gene negatively associated with the

IDT Exhibit 1020.5764

biological state. In some embodiments, more than one gene positively associated with the biological state being evaluated and more than one gene negatively associated with the biological state being evaluated can be used. For example, in some embodiments, a numerical index can be derived comprising numerical values for the positively associated genes in the numerator and numerical values for an equivalent number of the negatively associated genes in the denominator. In such balanced numerical indices, the reference nucleic acid numerical values cancel out. In some embodiments, balanced numerical values can neutralize effects of variation in the expression of the gene(s) providing the reference nucleic acid(s). In some embodiments, a numerical index is calculated by a series of one or more mathematical functions.

[00221] In some embodiments, more than 2 biological states can be compared, e.g., distinguished. For example, in some embodiments, samples may be provided from a range of biological states, e.g., corresponding to different stages of disease progression, e.g., different stages of cancer. Cells in different stages of cancer, for example, include a non-cancerous cell vs. a non-metastasizing cancerous cell vs. a metastasizing cell from a given patient at various times over the disease course. Cancer cells of various types of cancer may be used, including, for example, a bladder cancer, a bone cancer, a brain tumor, a breast cancer, a colon cancer, an endocrine system cancer, a gastrointestinal cancer, a gynecological cancer, a head and neck cancer, a leukemia, a lung cancer, a lymphoma, a metastases, a myeloma, neoplastic tissue, a pediatric cancer, a penile cancer, a prostate cancer, a sarcoma, a skin cancer, a testicular cancer, a thyroid cancer, and a urinary tract cancer. In preferred embodiments, biomarkers can be developed to predict which chemotherapeutic agent can work best for a given type of cancer, e.g., in a particular patient.

[00222] A non-cancerous cell may include a cell of hematoma and/or scar tissue, as well as morphologically normal parenchyma from non-cancer patients, e.g., non-cancer patients related or not related to a cancer patient. Non-cancerous cells may also include morphologically normal parenchyma from cancer patients, e.g., from a site close to the site of the cancer in the same tissue and/or same organ; from a site further away from the site of the cancer, e.g., in a different tissue and/or organ in the same organ-system, or from a site still further away e.g., in a different organ and/or a different organ-system.

[00223] Numerical indices obtained can be provided as a database. Numerical indices and/or databases thereof can find use in diagnoses, e.g. in the development and application of clinical tests.

[00224] *Diagnostic Applications*

[00225] In some embodiments, a method of identifying a biological state is provided. In some embodiments, the method comprises measuring and/or enumerating an amount of each of 2 nucleic acids in a sample, providing the amounts as numerical values; and using the numerical values to provide a numerical index, whereby the numerical index indicates the biological state.

IDT Exhibit 1020.5765

[00226] A numerical index that indicates a biological state can be determined as described above in accordance with various embodiments. The sample may be obtained from a specimen, e.g., a specimen collected from a subject to be treated. The subject may be in a clinical setting, including, e.g., a hospital, office of a health care provider, clinic, and/or other health care and/or research facility. Amounts of nucleic acid(s) of interests in the sample can then be measured and/or enumerated.

[00227] In certain embodiments, where a given number of genes are to be evaluated, expression data for that given number of genes can be obtained simultaneously. By comparing the expression pattern of certain genes to those in a database, a chemotherapeutic agent that a tumor with that gene expression pattern would most likely respond to can be determined.

[00228] In some embodiments, the methods can be used to quantify exogenous normal gene in the presence of mutated endogenous gene. Using primers that span the deleted region, one can selectively amplify and quantitate expression from a transfected normal gene and/or a constitutive abnormal gene.

[00229] In some embodiments, methods described herein can be used to determine normal expression levels, e.g., providing numerical values corresponding to normal gene transcript expression levels. Such embodiments may be used to indicate a normal biological state, at least with respect to expression of the evaluated gene.

[00230] Normal expression levels can refer to the expression level of a transcript under conditions not normally associated with a disease, trauma, and/or other cellular insult. In some embodiments, normal expression levels may be provided as a number, or preferably as a range of numerical values corresponding to a range of normal expression of a particular gene, e.g., within +/-a percentage for experimental error. Comparison of a numerical value obtained for a given nucleic acid in a sample, e.g., a nucleic acid corresponding to a particular gene, can be compared to established-normal numerical values, e.g., by comparison to data in a database provided herein. As numerical values can indicate numbers of molecules of the nucleic acid in the sample, this comparison can indicate whether the gene is being expressed within normal levels or not.

[00231] In some embodiments, the method can be used for identifying a biological state comprising assessing an amount a nucleic acid in a first sample, and providing said amount as a numerical value wherein said numerical value is directly comparable between a number of other samples. In some embodiments, the numerical value is potentially directly comparable to an unlimited number of other samples. Samples may be evaluated at different times, e.g., on different days; in the same or different experiments in the same laboratory; and/or in different experiments in different laboratories.

[00232] *Therapeutics*

[00233] Some embodiments provide a method of improving drug development. For example, use of a standardized mixture of internal standards, a database of numerical values and/or a

IDT Exhibit 1020.5766

database of numerical indices may be used to improve drug development.

[00234] In some embodiments, modulation of gene expression is measured and/or enumerated at one or more of these stages, e.g., to determine effect a candidate drug. For example, a candidate drug (e.g., identified at a given stage) can be administered to a biological entity. The biological entity can be any entity capable of harboring a nucleic acid, as described above, and can be selected appropriately based on the stage of drug development. For example, at the lead identification stage, the biological entity may be an in vitro culture. At the stage of a clinical trial, the biological entity can be a human patient.

[00235] The effect of the candidate drug on gene expression may then be evaluated, e.g., using various embodiments of the instant invention. For example, a nucleic acid sample may be collected from the biological entity and amounts of nucleic acids of interest can be measured and/or enumerated. For example, amounts can be provided as numerical value and/or numerical indices. An amount then may be compared to another amount of that nucleic acid at a different stage of drug development; and/or to a numerical values and/or indices in a database. This comparison can provide information for altering the drug development process in one or more ways.

[00236] Altering a step of drug development may refer to making one or more changes in the process of developing a drug, preferably so as to reduce the time and/or expense for drug development. For example, altering may comprise stratifying a clinical trial. Stratification of a clinical trial can refer to, e.g., segmenting a patient population within a clinical trial and/or determining whether or not a particular individual may enter into the clinical trial and/or continue to a subsequent phase of the clinical trial. For example, patients may be segmented based on one or more features of their genetic makeup determined using various embodiments of the instant invention. For example, consider a numerical value obtained at a pre-clinical stage, e.g., from an in vitro culture that is found to correspond to a lack of a response to a candidate drug. At the clinical trial stage, subjects showing the same or similar numerical value can be exempted from participation in the trial. The drug development process has accordingly be altered, saving time, and costs.

[00237] *Kits*

[00238] The internal amplification control (IAC)/competitive internal standards (IS) described herein may be assembled and provided in the form of kits. In some embodiments, the kit provides the IAC and reagents necessary to perform a PCR, including Multiplex-PCR and next-generation sequencing (NGS). The IAC may be provided in a single, concentrated form where the concentration is known, or serially diluted in solution to at least one of several known working concentrations.

[00239] The kits may include IS of 150 identified endogenous targets, as described herein, or IS of 28 ERCC targets, as described herein, or both. These IS may be provided in solution

IDT Exhibit 1020.5767

allowing the IS to remain stable for up to several years.

[00240] The kits may also provide primers designed specifically to amplify the IS of 150 endogenous targets, the IS of 28 ERCC targets, and their corresponding native targets. The kits may also provide one or more containers filled with one or more necessary PCR reagents, including but not limited to dNTPs, reaction buffer, Taq polymerase, and RNAse-free water. Optionally associated with such container(s) is a notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of IAC and associated reagents, which notice reflects approval by the agency of manufacture, use or sale for research use.

[00241] The kits may include appropriate instructions for preparing, executing, and analyzing PCR, including Multiplex-PCR and NGS, using the IS included in the kit. The instructions may be in any suitable format, including, but not limited to, printed matter, videotape, computer readable disk, or optical disc.

[00242] All publications, including patents and non-patent literature, referred to in this specification are expressly incorporated by reference herein. Citation of the any of the documents recited herein is not intended as an admission that any of the foregoing is pertinent prior art. All statements as to the date or representation as to the contents of these documents is based on the information available to the applicant and does not constitute any admission as to the correctness of the dates or contents of these documents.

[00243] While the invention has been described with reference to various and preferred embodiments, it should be understood by those skilled in the art that various changes may be made and equivalents may be substituted for elements thereof without departing from the essential scope of the invention. In addition, many modifications may be made to adapt a particular situation or material to the teachings of the invention without departing from the essential scope thereof.

[00244] Therefore, it is intended that the invention not be limited to the particular embodiment disclosed herein contemplated for carrying out this invention, but that the invention will include all embodiments falling within the scope of the claims.

IDT Exhibit 1020.5768

CLAIMS

What is claimed is:

1. A method for controlling for non-systematic error in an amplification-based next generation sequencing (NGS) library preparation, comprising:

including of an internal amplification control (IAC) sharing identical priming sites to a native nucleic acid target template of interest in a NGS library preparation;

mimicing the kinetics of the native nucleic acid target in the amplification reaction, and

controling for sample-, platform-, experiment-, operator- and/or target-specific variation in amplification efficiency.

2. The method of claim 1, wherein the amplification-based NGS comprises a PCR-based next generation sequencing (NGS).

3. A method for controlling abundance of target analytes while retaining quantitative information regarding original abundance of the target analytes, comprising:

including of one or more internal amplification control (IAC) sharing identical priming sites to one or more native nucleic acid templates of interest;

causing convergence of abundance among target analytes while retaining an original proportion of each native nucleic acid target template relative to its respective IAC; and,

retaining quantitative information regarding the original abundance representation of each target analyte and enabling quantification with a low number of sequencing reads.

4. A method for reducing oversampling of overrepresented native nucleic acid targets and stochastic sampling error associated with deep sequencing, comprising:

i) preparing a mixture comprising a known number of internal amplification control (IAC) nucleic acid molecules corresponding to each native nucleic acid target; and

ii) mixing the IAC mixture of step i) with a native nucleic acid target-containing sample prior to preparation of a library for sequencing, or prior to sequencing if library preparation is not required;

wherein each native nucleic acid target is similar to its respective IAC, with the exception of one or more changes to the nucleic acid sequence that are identifiable with sequencing, and wherein such changes can include one or more of deletions, additions, or alteration to the ordering or composition of nucleotides used;

iii) assessing the proportion of sequencing events between the native nucleic acid target and its respective IAC, along with the known number of IAC nucleic acid molecules input into the sample prior to library preparation; and

IDT Exhibit 1020.5769

iv) quantifiably determining the original amount of each native nucleic acid target in the original sample prior to library preparation and sequencing.

5. The method of claim 4, wherein the known number in step i) comprises one or more of: abundance, concentration and amount.

6. The method of claim 4, wherein the sequencing events in step iii) comprises one or more of: observations, counts and reads.

7. A method for quantifiably determining a native nucleic acid target copy number of a native nucleic acid target in a sample, comprising:

introducing an internal amplification control (IAC) in a sample of nucleic acid targets prior to library preparation such that non-systematic error and/or sequencer instrument specific bias is experienced by both the native nucleic acid target and IAC similarly.

8. The method of claim 7, wherein at an end of sequencing library preparation, the relative proportion between each unique native target will converge towards an equimolar abundance in the library.

9. The method of claim 8, wherein equimolar abundance comprises a uniform abundance in the library.

10. The method of any one of the preceding claims, wherein, for every 10-fold reduction in range of abundance among native nucleic acid targets, 10-fold fewer sequencing reads are required.

11. The method of any one of the preceding claims, further comprising:

conducting an inter-laboratory comparison of clinical molecular diagnostic results generated by nucleic acid quantification using sequencing that employs internal amplification control (IAC).

12. A method comprising:

i) assessing a proportion of sequencing events between at least one native nucleic acid target and its respective internal amplification control (IAC) standard;

ii) assessing the original number of IAC molecules input into the sample prior to library preparation; and,

iii) determining the original number of molecules for each native nucleic acid target in the

IDT Exhibit 1020.5770

sample prior to library preparation and sequencing by multiplying the native nucleic acid target/IAC proportion times the original IAC input number.

13. The method of claim 12, wherein the method includes using the same mixture of competitive IAC molecules in multiple different tests.

14. The method of any one of the preceding claims, further comprising:

conducting highly multiplexed analyses of multiple native nucleic acid targets across multiple samples, experiments and/or platforms.

15. The method of any one of the preceding claims, wherein abundance among native nucleic acid target templates varies greater than one million fold

16. The method of any one of the preceding claims, wherein abundance among native nucleic acid target templates varies greater than one hundred thousand fold.

17. The method of any one of the preceding claims, wherein abundance among native nucleic acid target templates varies greater than ten thousand fold.

18. The method of any one of the preceding claims, wherein abundance among native nucleic acid target templates varies greater than one thousand fold.

19. The method of any one of the preceding claims, further comprising determining an amount of a first native nucleic acid target in a first sample, comprising:

i) providing a standardized mixture comprising a competitive template for said first nucleic acid and a competitive template for a second nucleic acid in said first sample, wherein said competitive templates are at known concentrations relative to each other;

ii) combining said first sample with said standardized mixture;

iii) co-amplifying said first nucleic acid and said competitive template for said first nucleic acid to produce first amplified product thereof;

iv) diluting said first amplified product;

v) further co-amplifying said diluted first amplified product of said first nucleic acid and of said competitive template for said first nucleic acid, to produce second amplified product thereof; and

vi) co-amplifying said second nucleic acid and said competitive template for said second nucleic acid to produce first amplified product thereof.

IDT Exhibit 1020.5771

20. A method for determining an amount of a first native nucleic acid target, comprising:

i) providing a series of serially-diluted standardized mixtures comprising a competitive internal amplification control (IAC) for said first native nucleic acid template and a competitive IAC for a second native nucleic acid target present in a number of samples comprising said first native nucleic acid target, wherein said competitive IACs are at known concentrations relative to each other;

ii) combining one of said samples comprising said first native nucleic acid target with a first one of said serially-diluted standardized mixtures;

iii) co-amplifying said first native nucleic acid target and said competitive IAC for said first native nucleic acid target to produce amplified product thereof;

iv) obtaining a first relationship, said first relationship comparing said amplified product of said first native nucleic acid target to said amplified product of said competitive IAC for said first native nucleic acid target;

v) determining whether said first relationship is within about 1:10 to about 10:1; if not, repeating said combining, co-amplifying, obtaining and determining steps with a second one of said serially-diluted standardized mixtures;

vi) co-amplifying said second native nucleic acid target and said competitive IAC for said second native nucleic acid target to produce amplified product thereof;

vii) obtaining a second relationship, said second relationship comparing said amplified product of said second native nucleic acid target to said amplified product of said competitive IAC for said second native nucleic acid target; and

viii) comparing said first and said second relationships.

21. The method of claim 21, wherein in comparing said amplified product of said first native nucleic acid target to said amplified product of said competitive IAC for said first native nucleic acid sequence, it is determined whether said first relationship is within about 1:100 to about 100:1, or 1:1000 to about 1000:1, or 1:10,000 to about 10,000 to 1;

if not, repeating said combining, co-amplifying, obtaining and determining steps with a second one of said serially-diluted standardized mixtures;

co-amplifying said second nucleic acid and said competitive template for said second nucleic acid to produce amplified product thereof;

obtaining a second relationship, said second relationship comparing said amplified product of said second nucleic acid to said amplified project of said competitive template for said second nucleic acid; and

comparing said first and said second relationships.

IDT Exhibit 1020.5772

22. The method of claim 21, wherein the products from the co-amplification reactions are combined and amplified in a second round of amplification using primer pairs that recognize each native nucleic acid target and competitive IAC product from a first round of amplification and that also have a gene-specific barcode primer and a universal primer at 5' end to facilitate sequencing.

23. The method of any one of the preceding claims, further comprising: diluting said amplified product of said first native nucleic acid target and said competitive IAC for said first native nucleic acid target; and further co-amplifying said diluted amplified product to produce further amplified product thereof.

24. The method of any one of the preceding claims, further comprising: diluting said amplified product of said second native nucleic acid target and said competitive IAC for said second native nucleic acid target; and further co-amplifying said diluted amplified product to produce further amplified product thereof.

25. The method of any one of the preceding claims, further comprising: performing a series of serial dilutions of a mixture comprising said second native nucleic acid target.

26. The method of any one of the preceding claims, wherein one of said serial dilutions is selected to provide said second native nucleic acid target approximately calibrated to said competitive IAC for said second native nucleic acid target in said first one of said serially-diluted standardized mixtures.

27. The method of any one of the preceding claims, wherein said second native nucleic acid target serves as a first reference nucleic acid.

28. The method of any one of the preceding claims, wherein said first reference nucleic acid is a control for loading.

29. The method of any one of the preceding claims, wherein said first reference nucleic acid corresponds to at least one gene selected from GADP, ACTB and β-actin.

30. The method of any one of the preceding claims, wherein said first reference nucleic acid is present at two different concentrations in two of said serially-diluted standardized mixtures.

IDT Exhibit 1020.5773

31. The method of any one of the preceding claims, wherein said series of serially-diluted standardized mixtures further comprises sufficient amounts of said competitive IAC for assessing said first native nucleic acid target in at least or about 10^6 samples.

32. The method of any one of the preceding claims, wherein said series of serially-diluted standardized mixtures further comprises a number of other competitive IAC molecules for other native nucleic acid target(s) wherein said competitive IAC molecules are at known concentrations relative to one another, thereby allowing assessment of said other native nucleic acid target(s).

33. The method of any one of the preceding claims, wherein at least one of said other native nucleic acid targets serves as a second reference nucleic acid.

34. The method of any one of the preceding claims, wherein said second reference nucleic acid corresponds to at least one gene selected from CADP, ACTB and β-actin.

35. The method of any one of the preceding claims, wherein said assessing further comprises:

i) co-amplifying said second reference nucleic acid and said competitive IAC for said second reference nucleic acid to produce amplified product thereof;

ii) obtaining a third relationship, said third relationship comparing said amplified product of said second reference nucleic acid to said amplified product of said competitive template for said second reference nucleic acid; and,

iii) comparing said first and said third relationships.

36. The method of any one of the preceding claims, wherein said series of serially-diluted standardized mixtures further comprises sufficient amounts of said number of other competitive IAC(s) for assessing said other native nucleic acid target(s) in more than, or about, 10^6 samples.

37. The method of any one of the preceding claims, wherein said first native nucleic acid target and said other native nucleic acid target(s) vary in amount over a range of more than, or about, two orders of magnitude.

38. The method of any one of the preceding claims, wherein said competitive IAC for said first native nucleic acid target is present at a series of concentrations relative to said competitive IAC for said second native nucleic acid target.

IDT Exhibit 1020.5774

39. The method of any one of the preceding claims, wherein obtaining said first or said second relationship comprises a use of microfluidic device, capillary electrophoresis, an oligonucleotide array, mass spectrometry, or chromatography.

40. The method of any one of the preceding claims, wherein said standardized mixtures control for at least two sources of variation selected from cDNA loading, intra-nucleic acid amplification efficiency, inter-nucleic acid amplification efficiency, inter-specimen amplification efficiency, inter-sample amplification efficiency, and intra-sample amplification efficiency.

41. The method of any one of the preceding claims, wherein at least one of said standardized mixtures of said series can enumerate less than, or about, 1,000 molecules of said first native nucleic acid target in one of said samples.

42. The method of any one of the preceding claims, wherein said method is computer implemented.

43. The method of any one of the preceding claims, wherein said computer implementation comprises instructing a robotic handler to select said first one of said serially - diluted standardized mixtures for combining.

44. The method of any one of the preceding claims, wherein said computer implementation comprises obtaining said first relationship.

45. The method of any one of the preceding claims, wherein said obtaining said first relationship involves determining an area under a curve.

46. The method of any one of the preceding claims, wherein said computer implementation comprises instructing said robotic handler to select said second one of said serially-diluted standardized mixtures based on said first relationship.

47. The method of any one of the preceding claims, wherein said first nucleic acid comprises an RNA molecule.

48. The method of any one of the preceding claims wherein said first native nucleic acid target comprises a DNA molecule.

IDT Exhibit 1020.5775

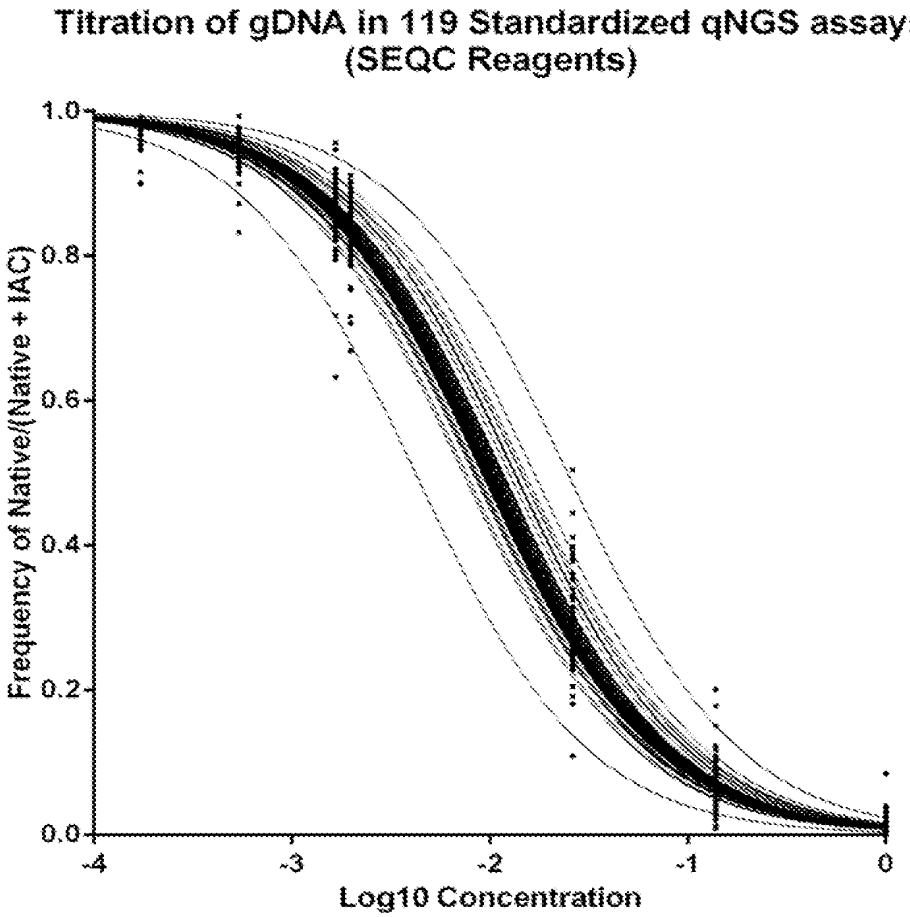

Titration of gDNA in 119 Standardized qNGS assays
(SEQC Reagents)
Frequency of Native/(Native + IAC)
1.0
0.8
0.6
0.4
0.2
0.0
-4
-3
-2
-1
0
Log10 Concentration

FIG. 1

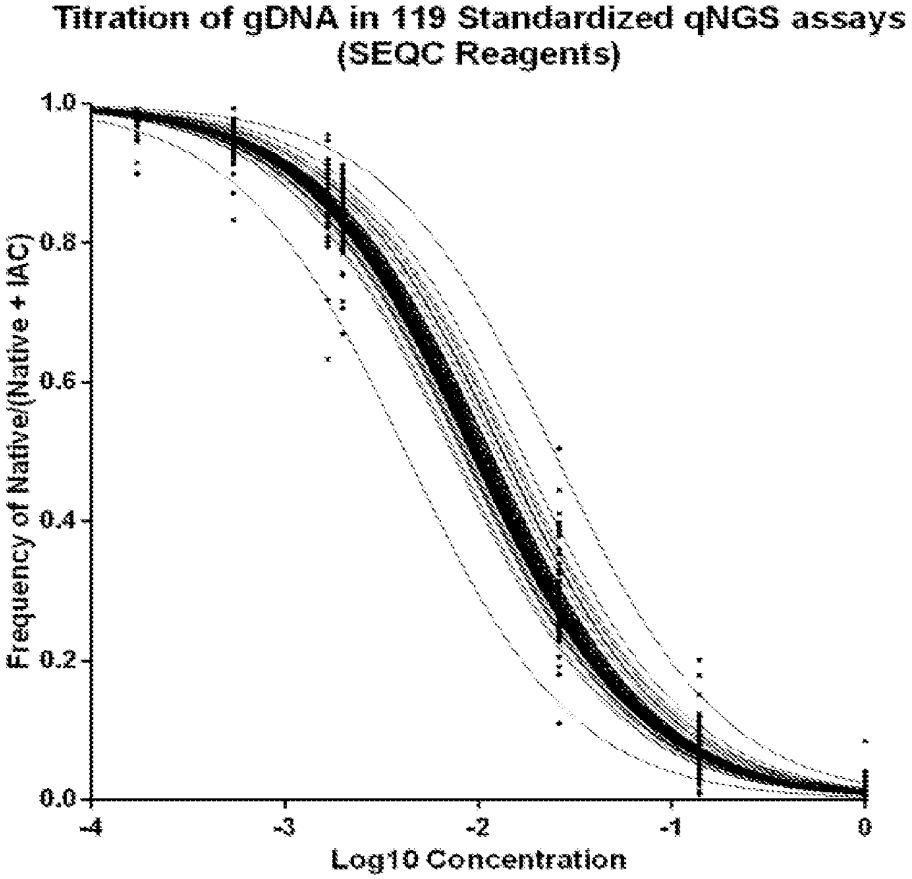

Titration of gDNA in 119 Standardized qNGS assays
(SEQC Reagents)
Frequency of Native/(Native + IAC)
1.0
0.8
0.6
0.4
0.2
0.0
-4
-3
-2
-1
0
Log10 Concentration

FIG. 2A

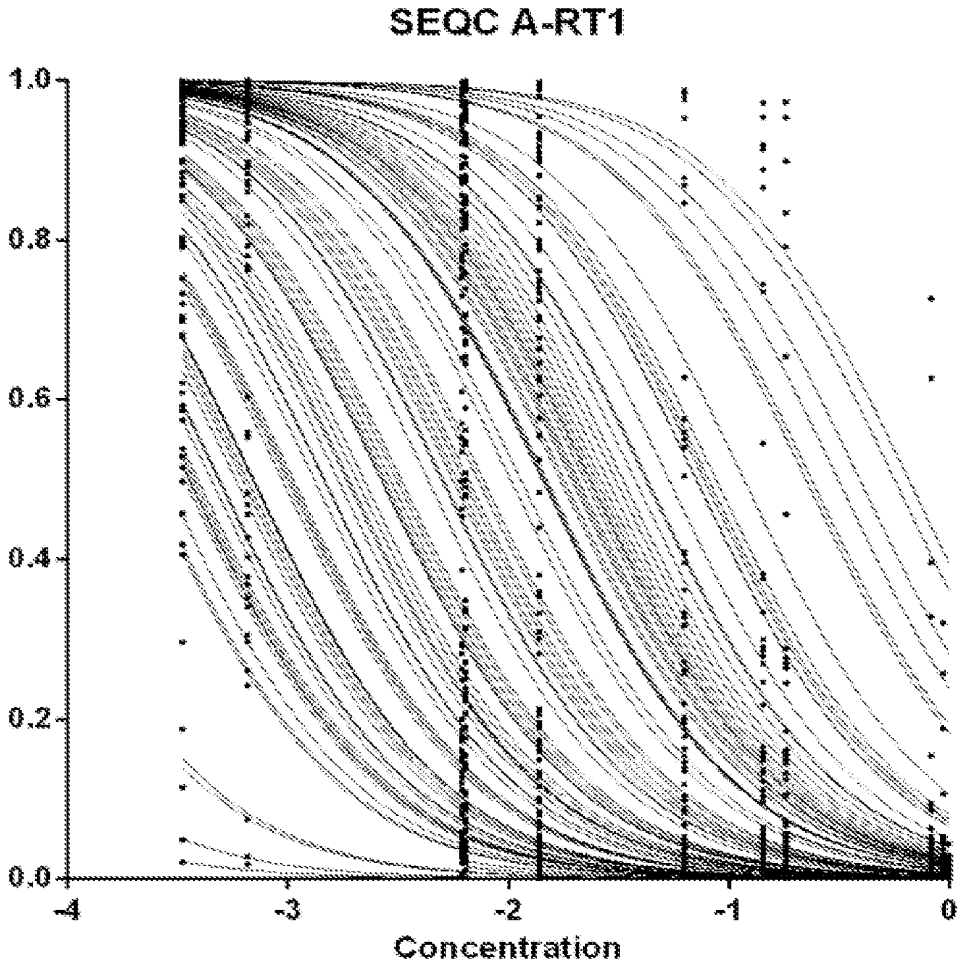


FIG. 2B

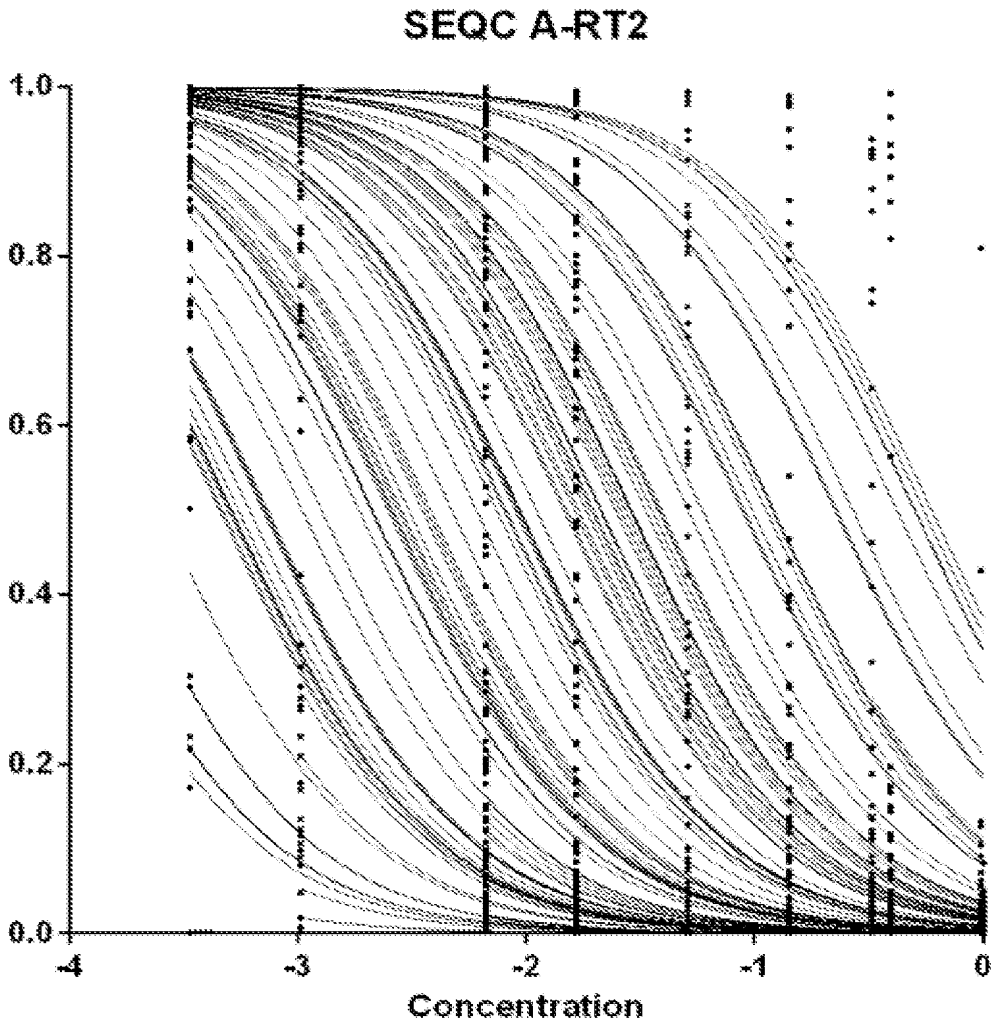


FIG. 2C

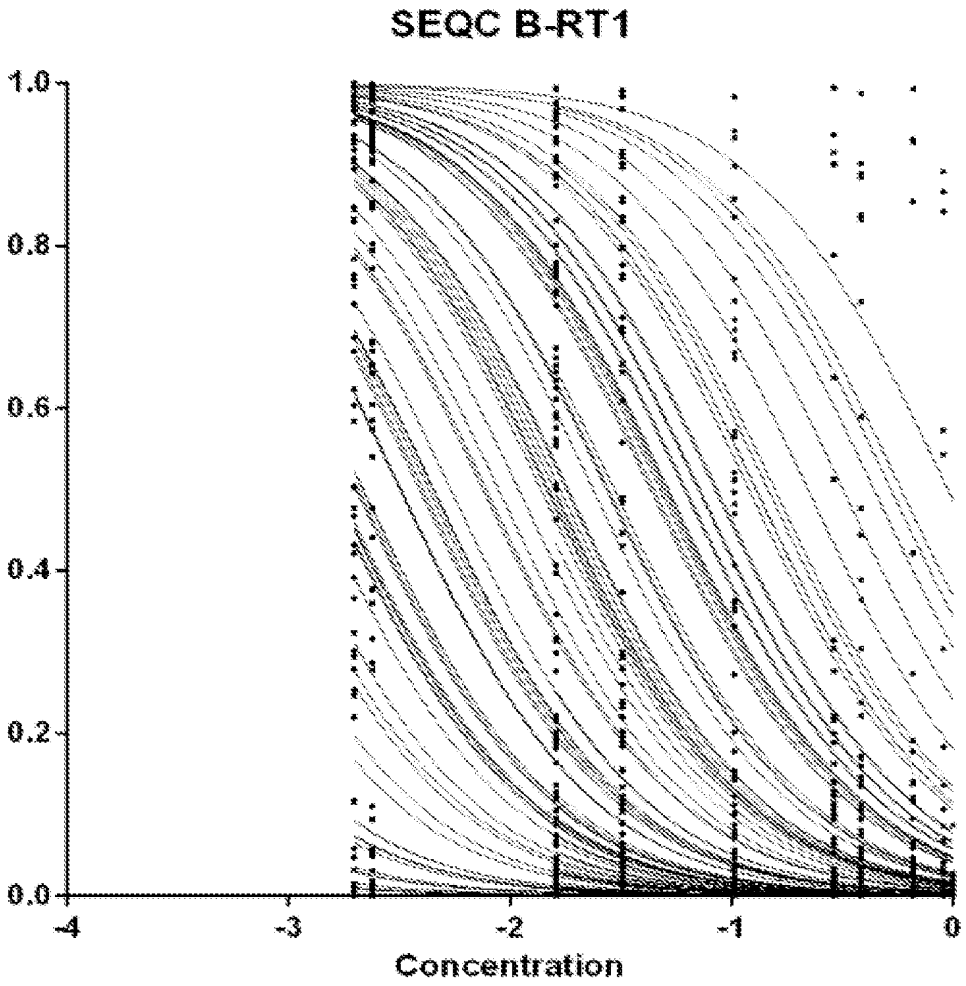


FIG. 2D

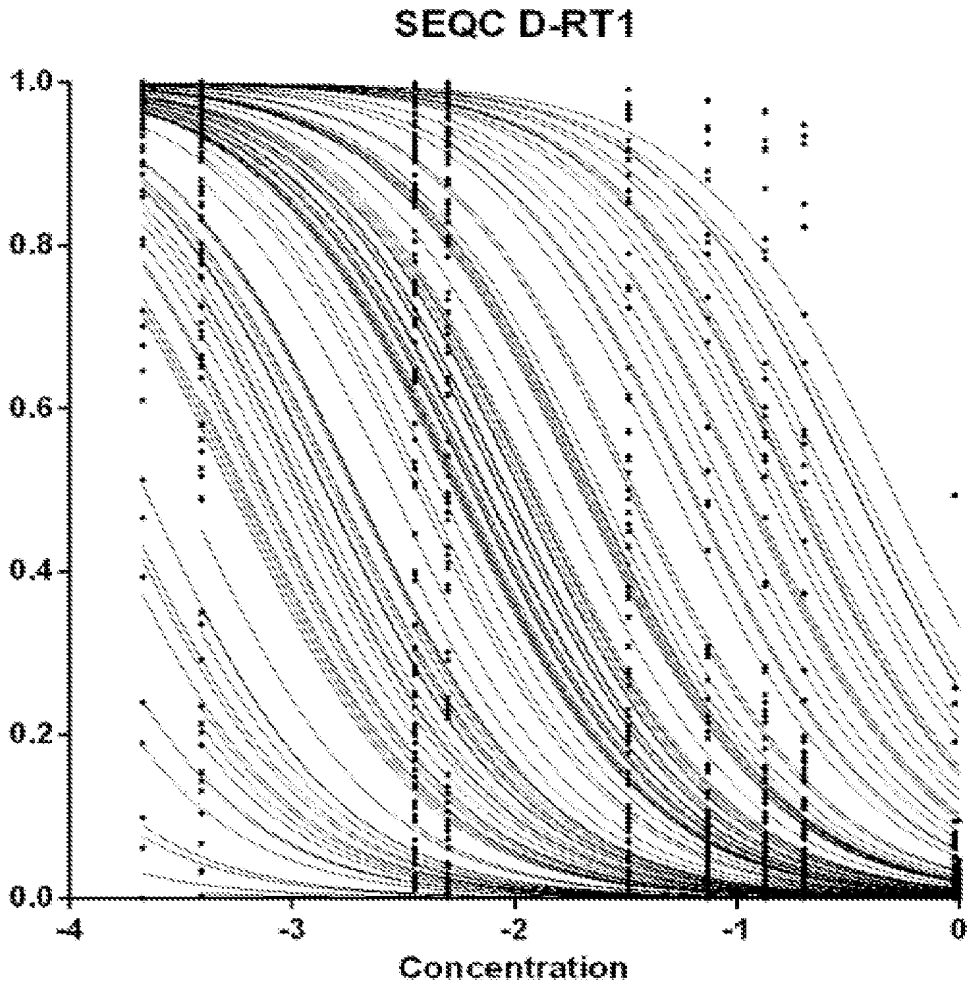


FIG. 2E

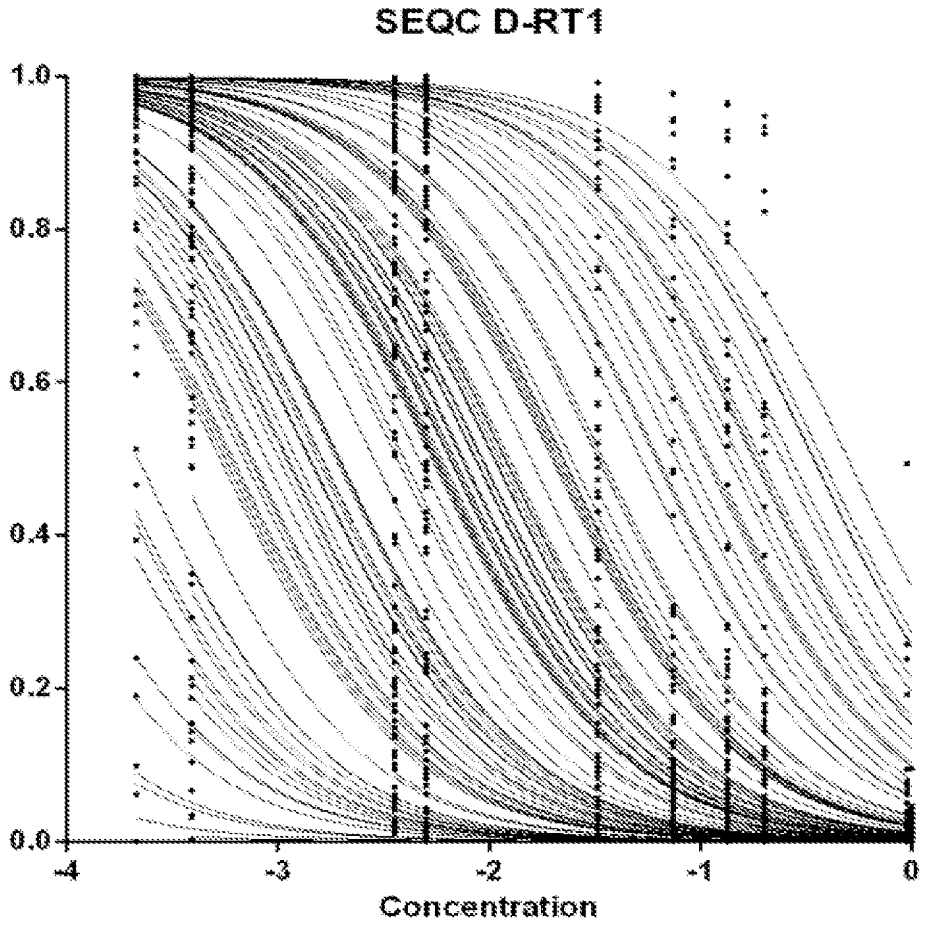


FIG. F

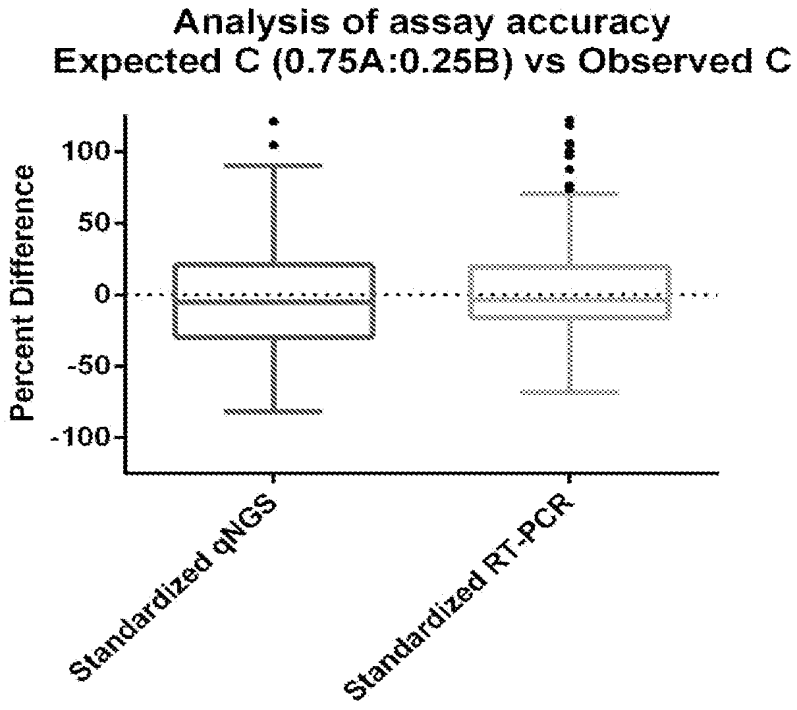

Analysis of assay accuracy
Expected C (0.75A:0.25B) vs Observed C
Percent Difference
100
50
0
-50
-100
Standardized qNGS
Standardized RT-PCR

FIG. 3A

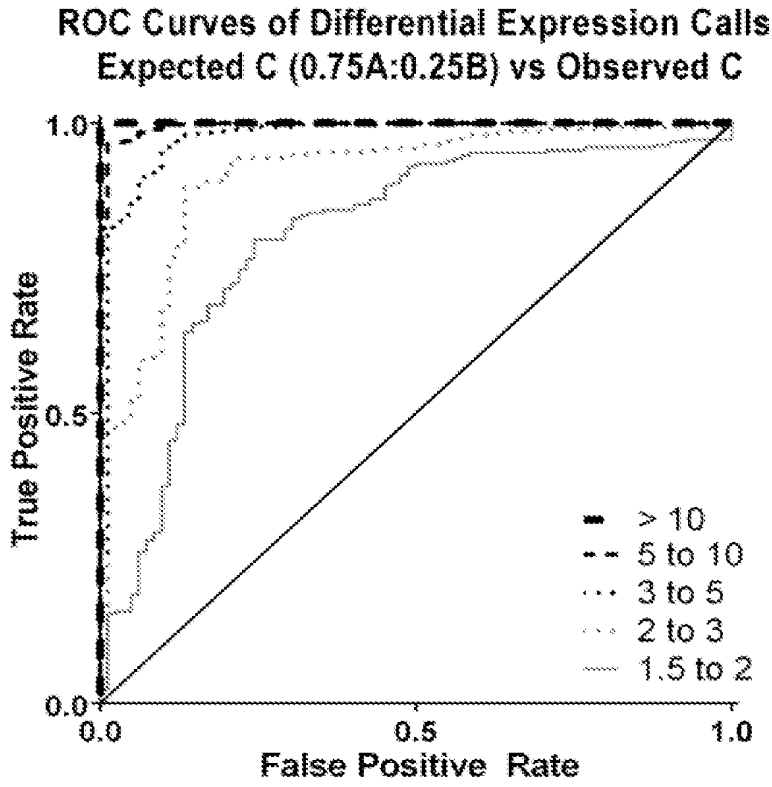


FIG. 4A

Detect Fold Change	Accuracy	95% CI
> 10	0.99	(0.99- 1.00)
5 to 10	0.99	(0.99- 1.00)
3 to 5	0.97	(0.95-0.99)
2 to 3	0.91	(0.87-0.94)
1.5 to 3	0.80	(0.74-0.86)

FIG. 4B

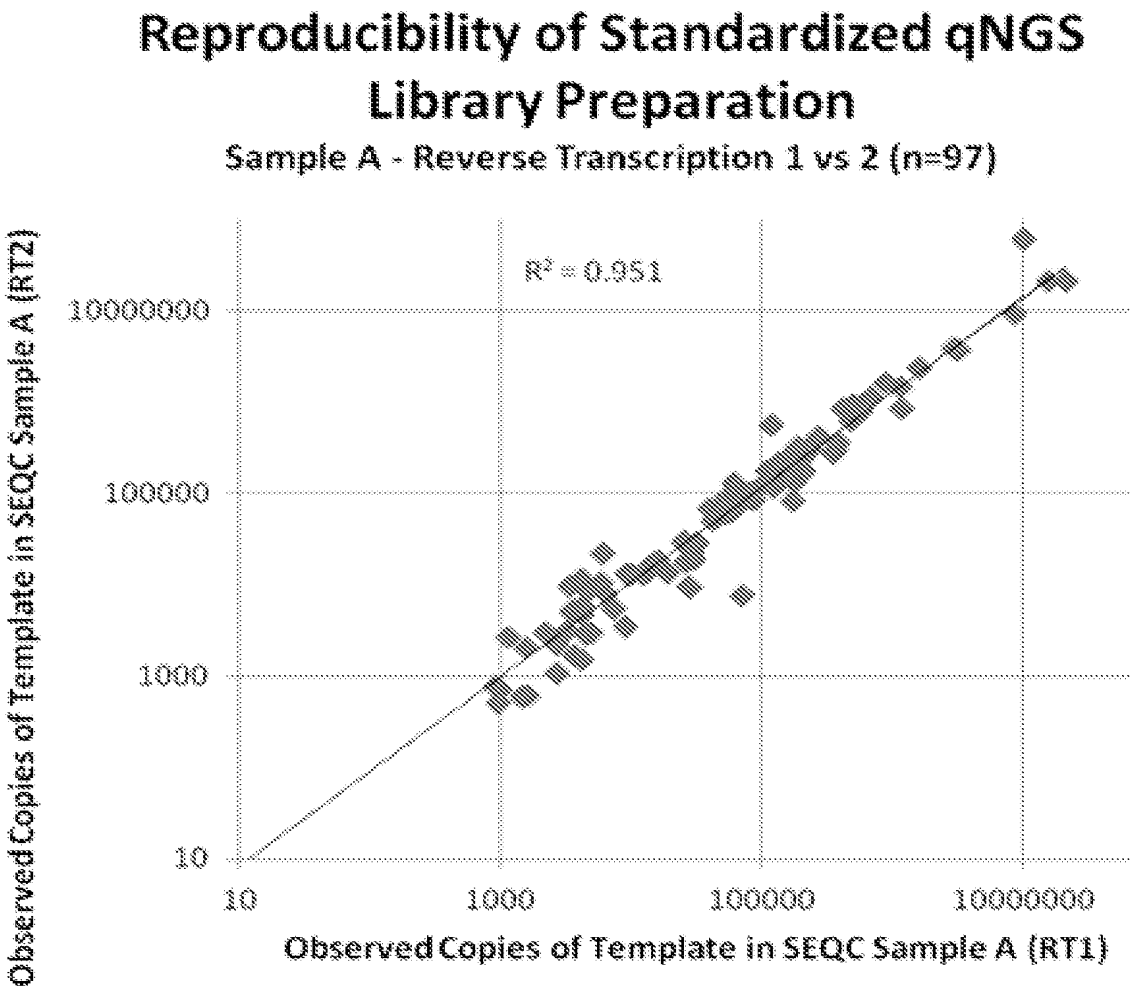

Reproducibility of Standardized qNGS Library Preparation
Sample A - Reverse Transcription 1 vs 2 (n=97)
R² = 0.951
Observed Copies of Template in SEQC Sample A (RT2)
10000000
100000
1000
10
10
1000
100000
10000000
Observed Copies of Template in SEQC Sample A (RT1)

FIG. 5

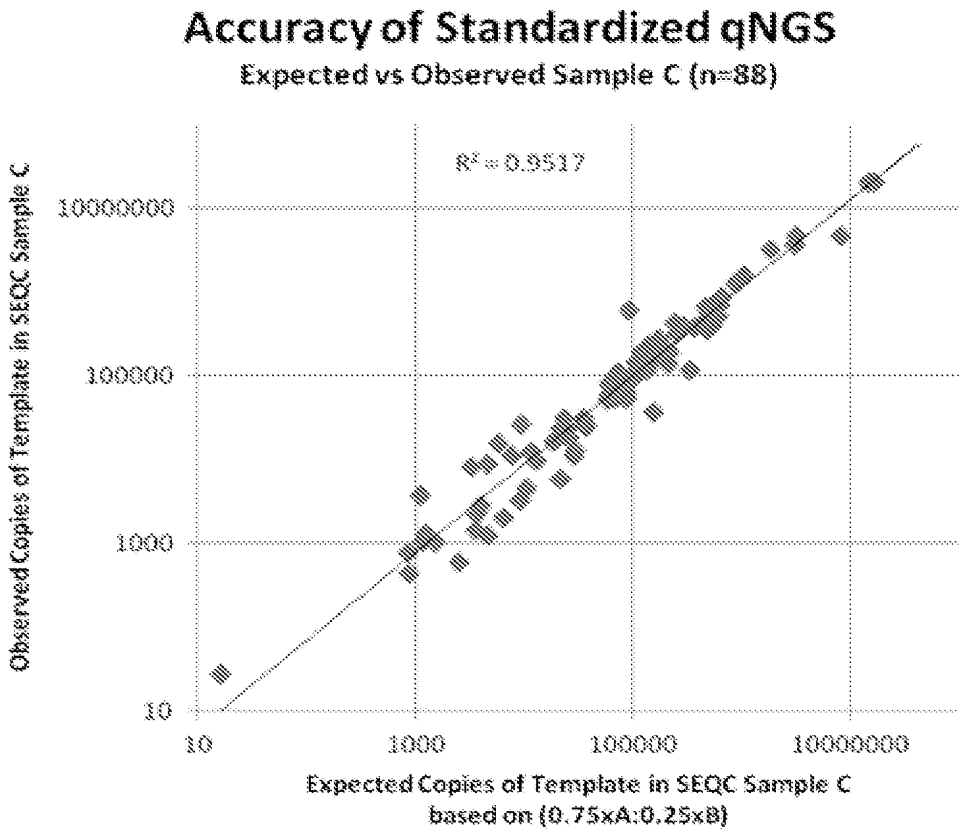

Accuracy of Standardized qNGS
Expected vs Observed Sample C (n=88)
R² = 0.9517
Observed Copies of Template in SEQC Sample C
10000000
100000
1000
10
10
1000
100000
10000000
Expected Copies of Template in SEQC Sample C
based on (0.75xA:0.25xB)

FIG. 6

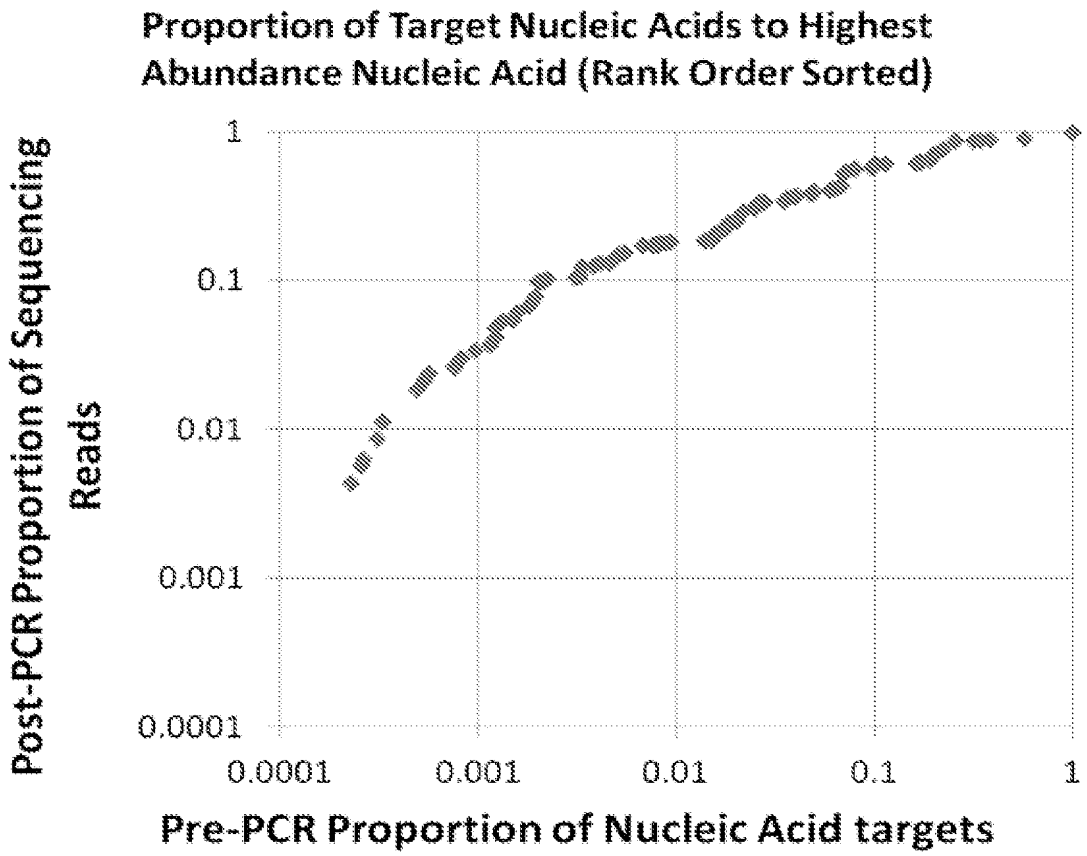

Proportion of Target Nucleic Acids to Highest Abundance Nucleic Acid (Rank Order Sorted)
Post-PCR Proportion of Sequencing Reads
1
0.1
0.01
0.001
0.0001
0.0001
0.001
0.01
0.1
1
Pre-PCR Proportion of Nucleic Acid targets

FIG. 7

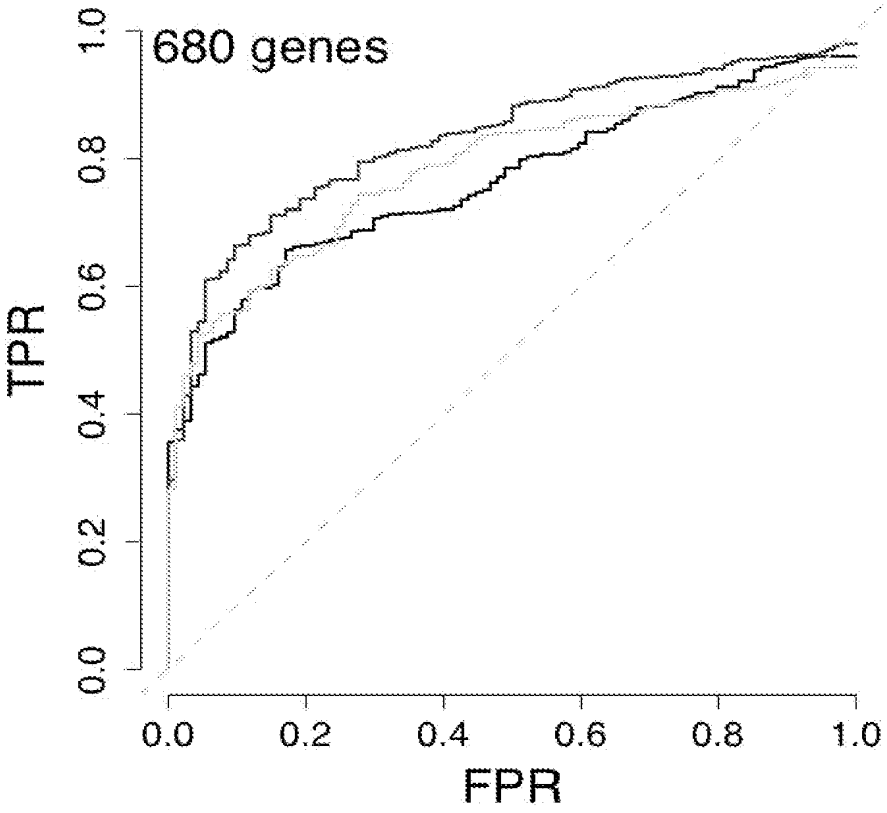


FIG. 8

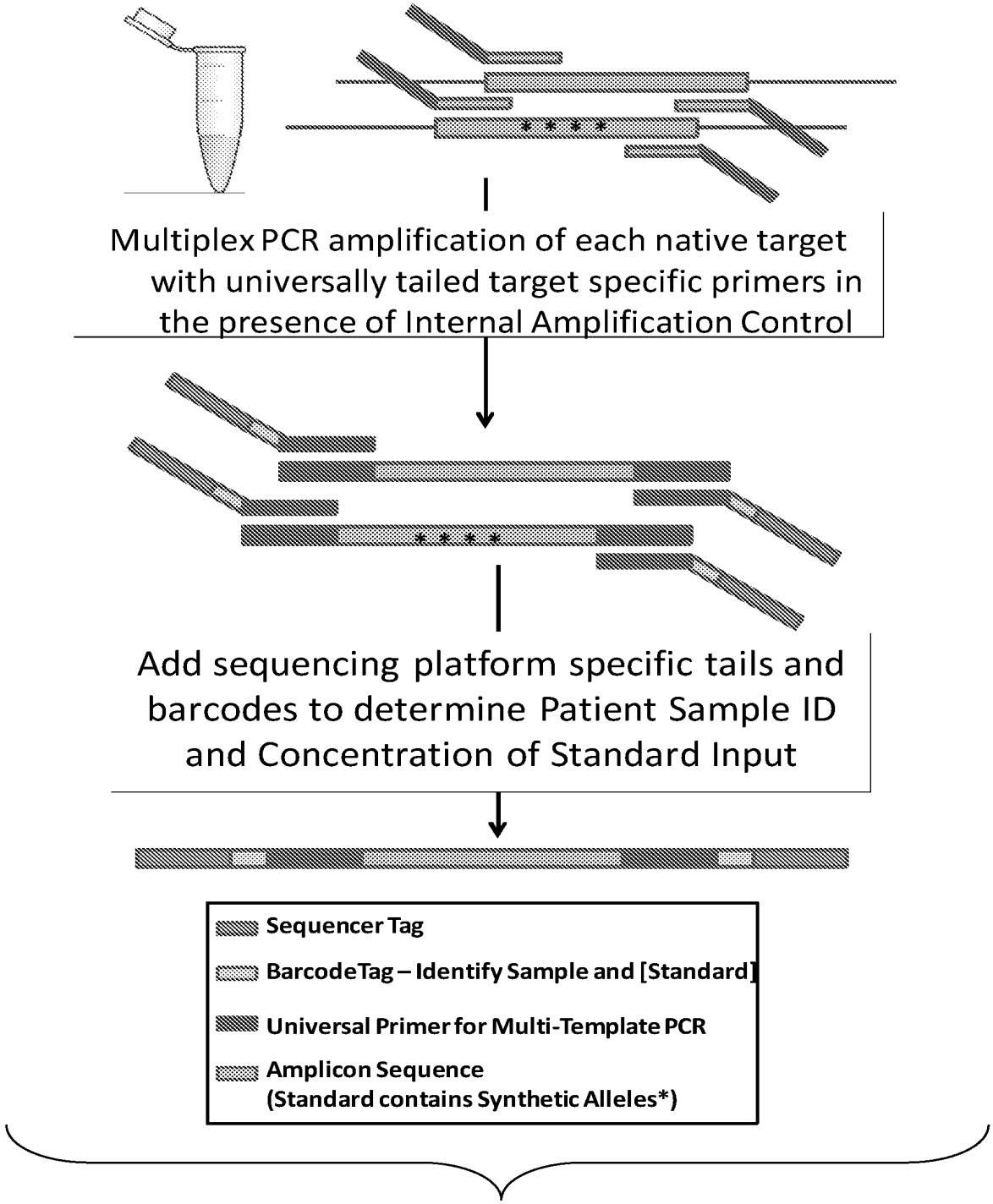

Multiplex PCR amplification of each native target with universally tailed target specific primers in the presence of Internal Amplification Control
Add sequencing platform specific tails and barcodes to determine Patient Sample ID and Concentration of Standard Input
Sequencer Tag
BarcodeTag – Identify Sample and [Standard]
Universal Primer for Multi-Template PCR
Amplicon Sequence
(Standard contains Synthetic Alleles*)

FIG. 9

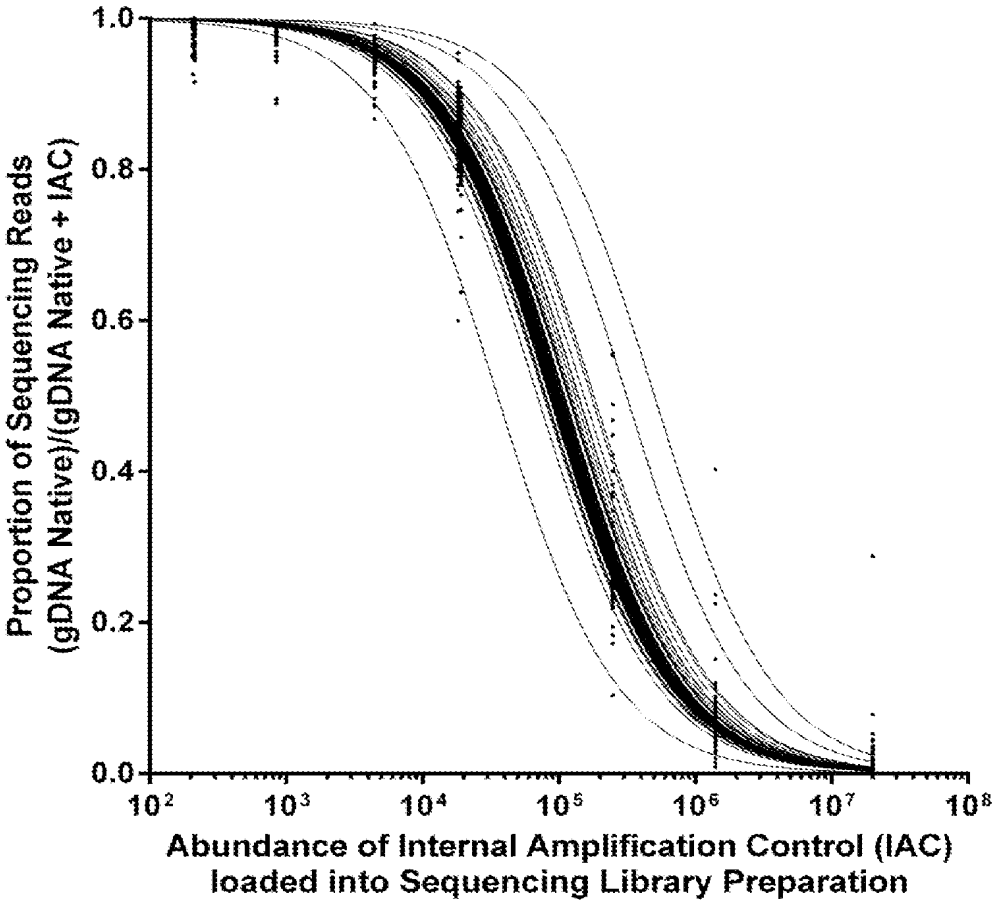

Proportion of Sequencing Reads
(gDNA Native)/(gDNA Native + IAC)
1.0
0.8
0.6
0.4
0.2
0.0
10²
10³
10⁴
10⁵
10⁶
10⁷
10⁸
Abundance of Internal Amplification Control (IAC)
loaded into Sequencing Library Preparation

FIG. 10

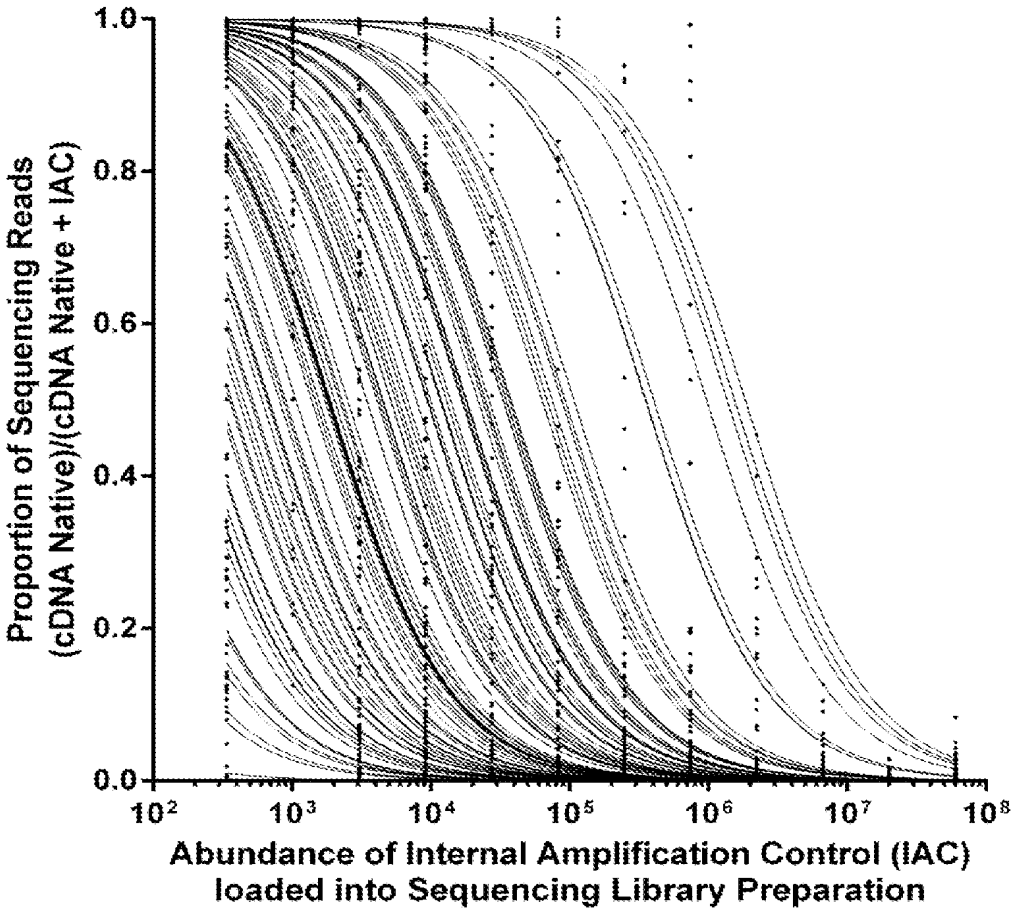

Proportion of Sequencing Reads
(cDNA Native)/(cDNA Native + IAC)
1.0
0.8
0.6
0.4
0.2
0.0
10^2
10^3
10^4
10^5
10^6
10^7
10^8
Abundance of Internal Amplification Control (IAC)
loaded into Sequencing Library Preparation

FIG. 11

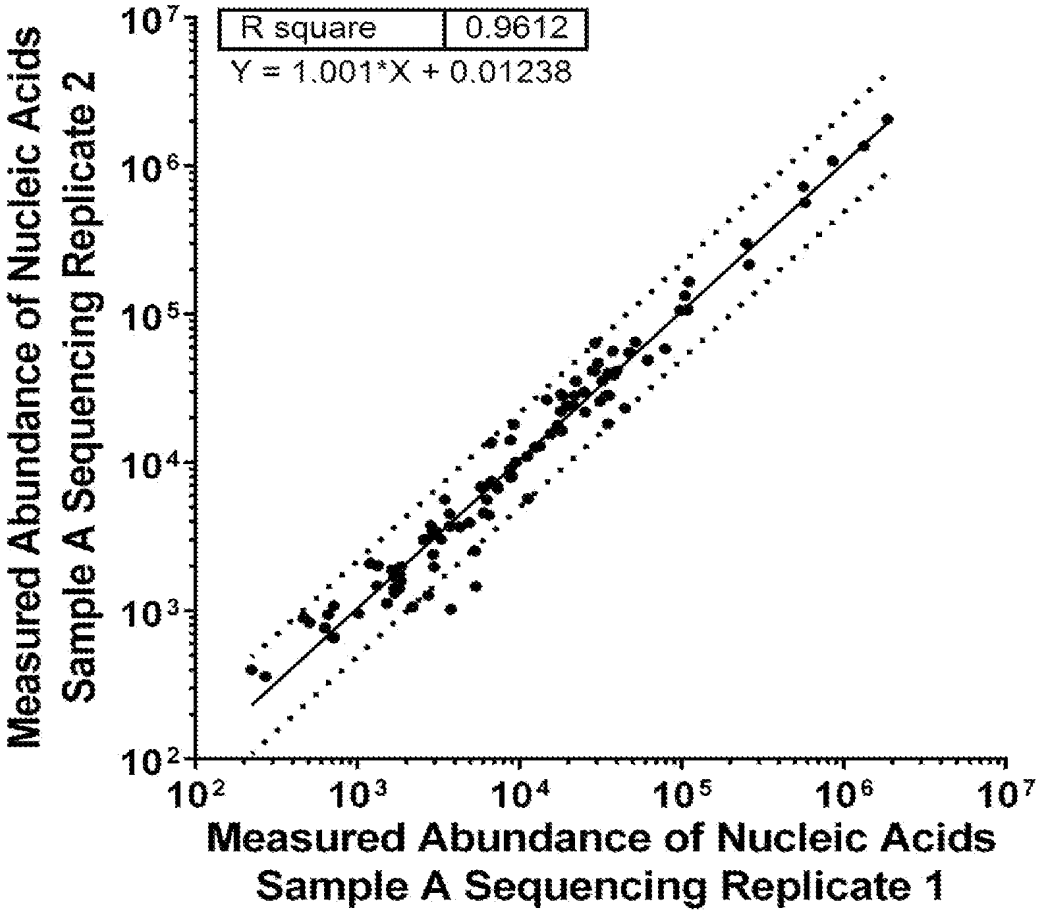

R square 0.9612
Y = 1.001*X + 0.01238
Measured Abundance of Nucleic Acids Sample A Sequencing Replicate 2
Measured Abundance of Nucleic Acids Sample A Sequencing Replicate 1
10^2
10^3
10^4
10^5
10^6
10^7

FIG. 12

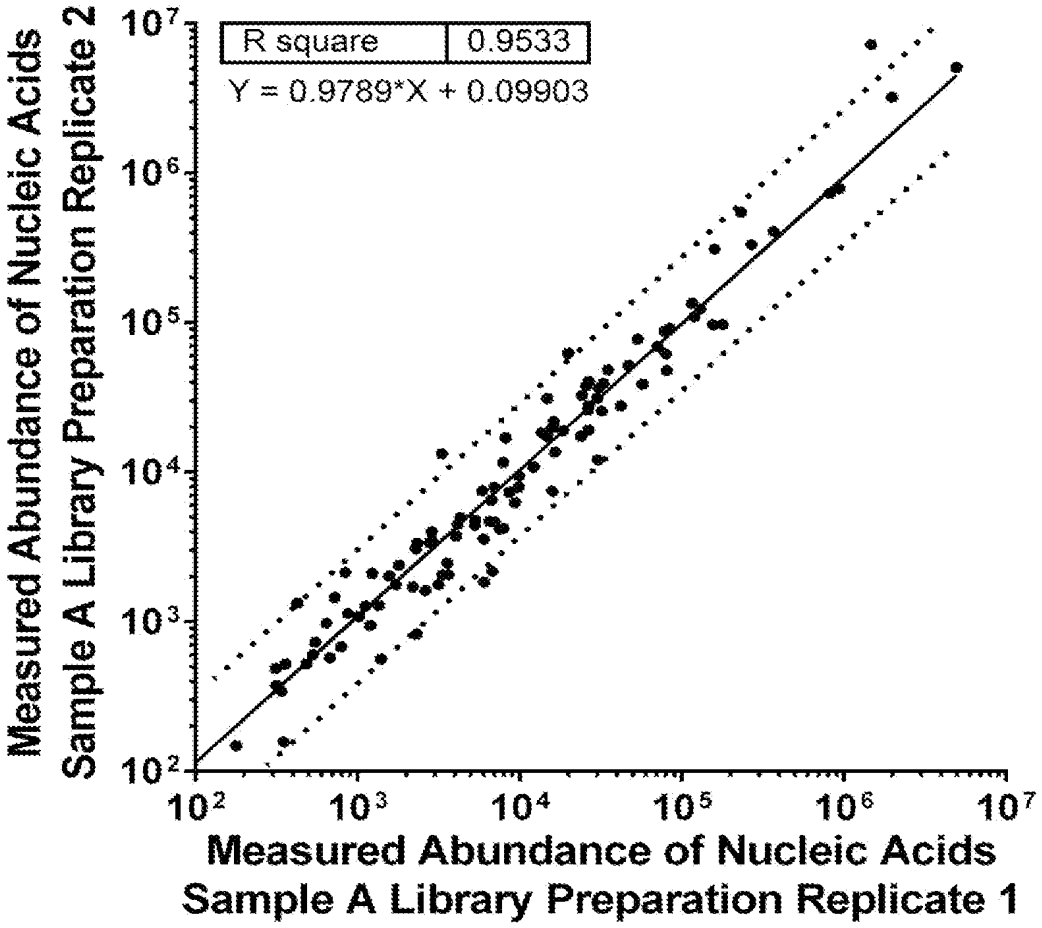

R square 0.9533
Y = 0.9789*X + 0.09903
Measured Abundance of Nucleic Acids
Sample A Library Preparation Replicate 2
Measured Abundance of Nucleic Acids
Sample A Library Preparation Replicate 1
10^2
10^3
10^4
10^5
10^6
10^7

FIG. 13

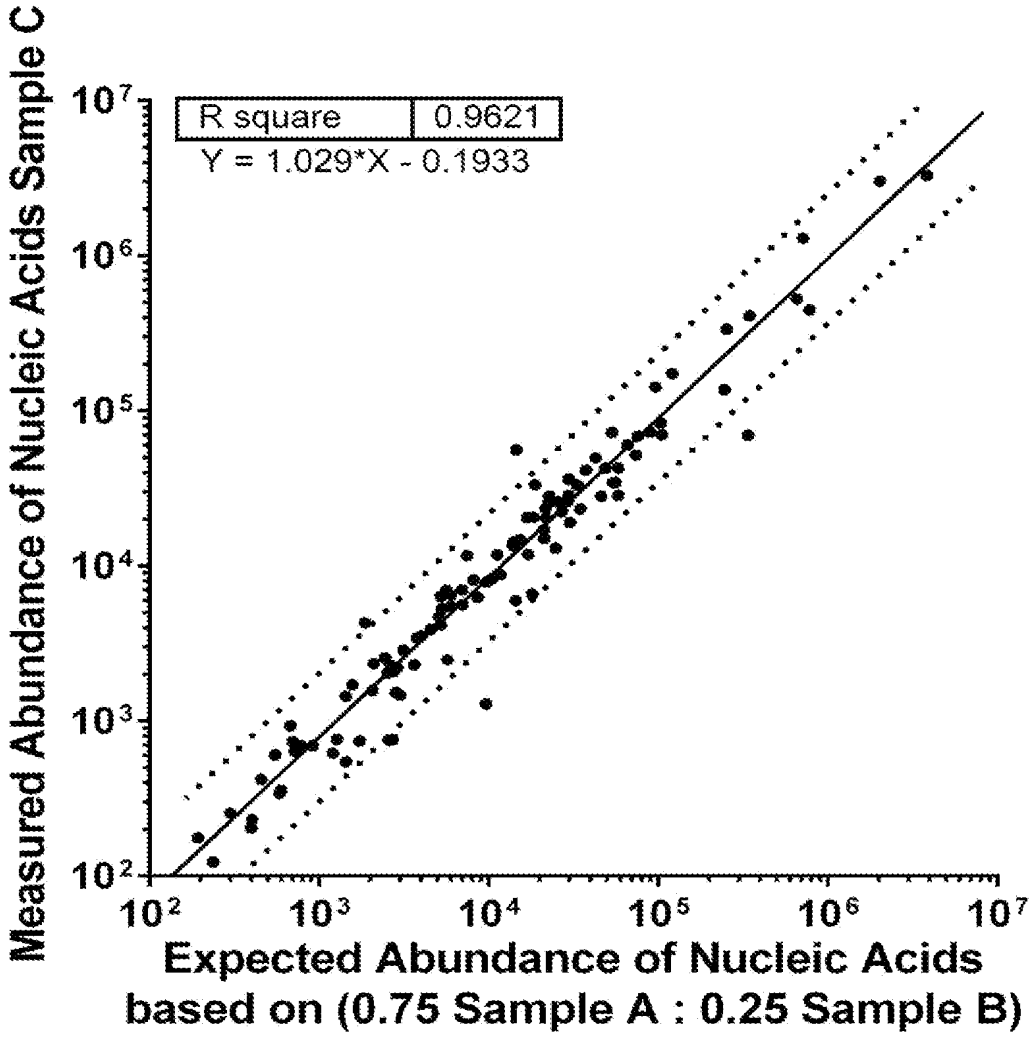

R square 0.9621
Y = 1.029*X - 0.1933
Measured Abundance of Nucleic Acids Sample C
Expected Abundance of Nucleic Acids based on (0.75 Sample A : 0.25 Sample B)
10^2
10^3
10^4
10^5
10^6
10^7

FIG. 14A

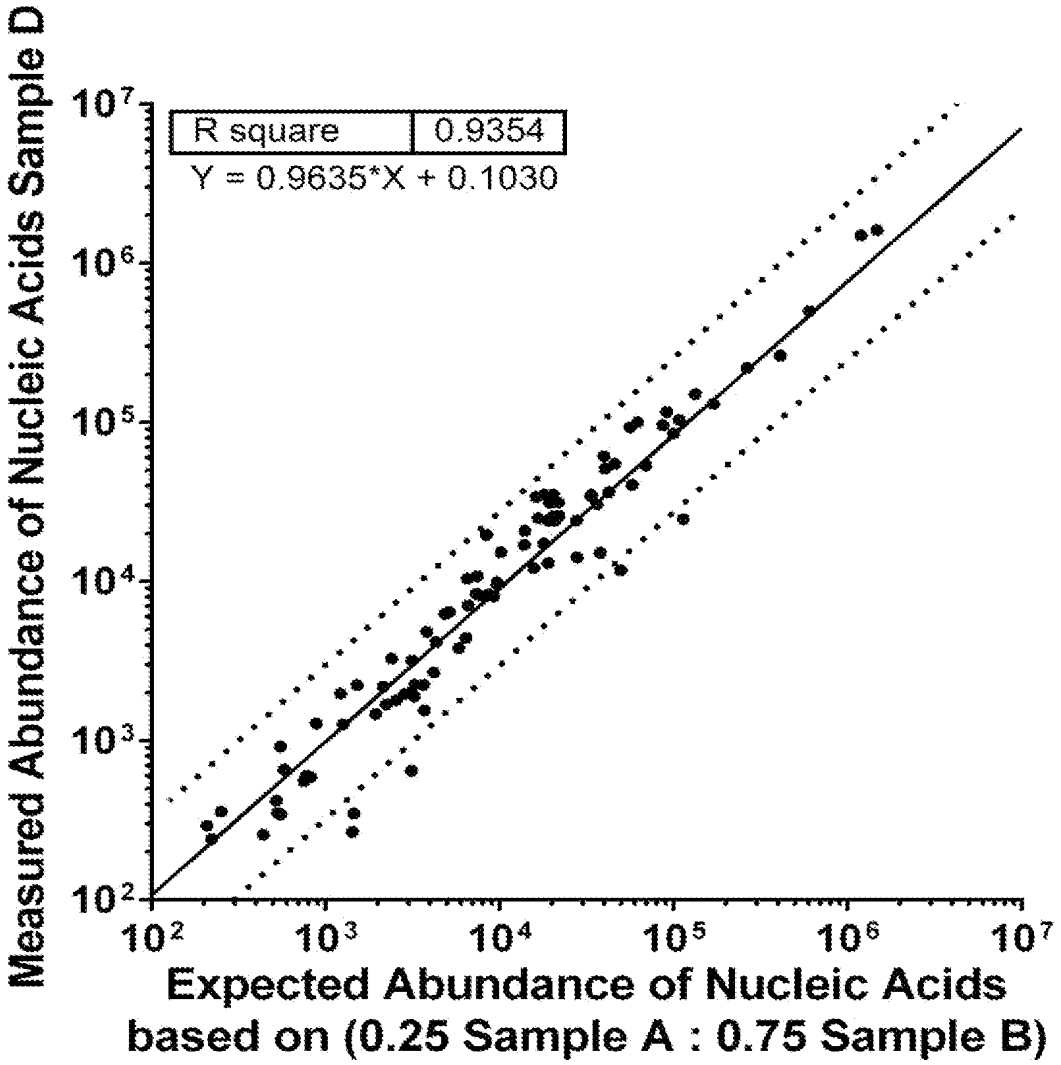

R square 0.9354
Y = 0.9635*X + 0.1030
Measured Abundance of Nucleic Acids Sample D
Expected Abundance of Nucleic Acids based on (0.25 Sample A : 0.75 Sample B)
10^2
10^3
10^4
10^5
10^6
10^7

FIG. 14B

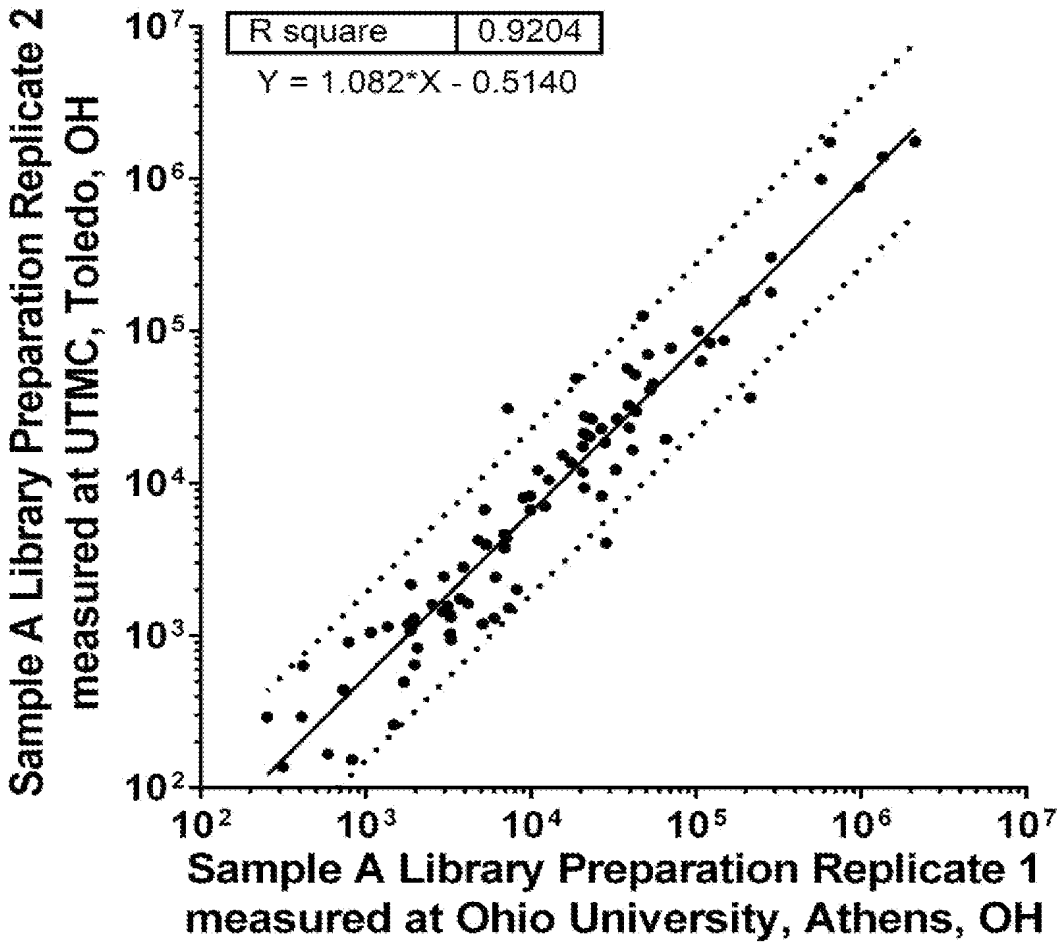

R square 0.9204
Y = 1.082*X - 0.5140
Sample A Library Preparation Replicate 2 measured at UTMC, Toledo, OH
Sample A Library Preparation Replicate 1 measured at Ohio University, Athens, OH
10^2
10^3
10^4
10^5
10^6
10^7

FIG. 15

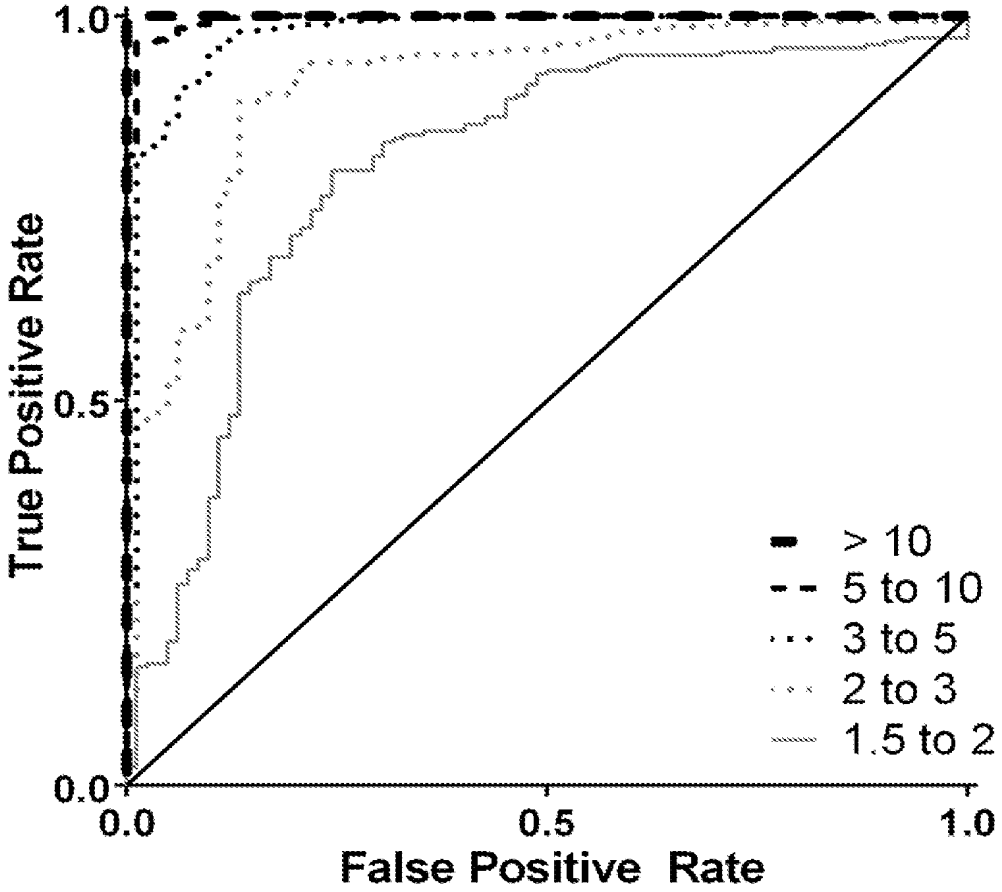


FIG. 16A

Detect Fold Change	Accuracy	95% CI
> 10	0.99	(0.99- 1.00)
5 to 10	0.99	(0.99- 1.00)
3 to 5	0.97	(0.95-0.99)
2 to 3	0.91	(0.87-0.94)
1.5 to 2	0.80	(0.74-0.86)

Fig. 4B

FIG. 16B

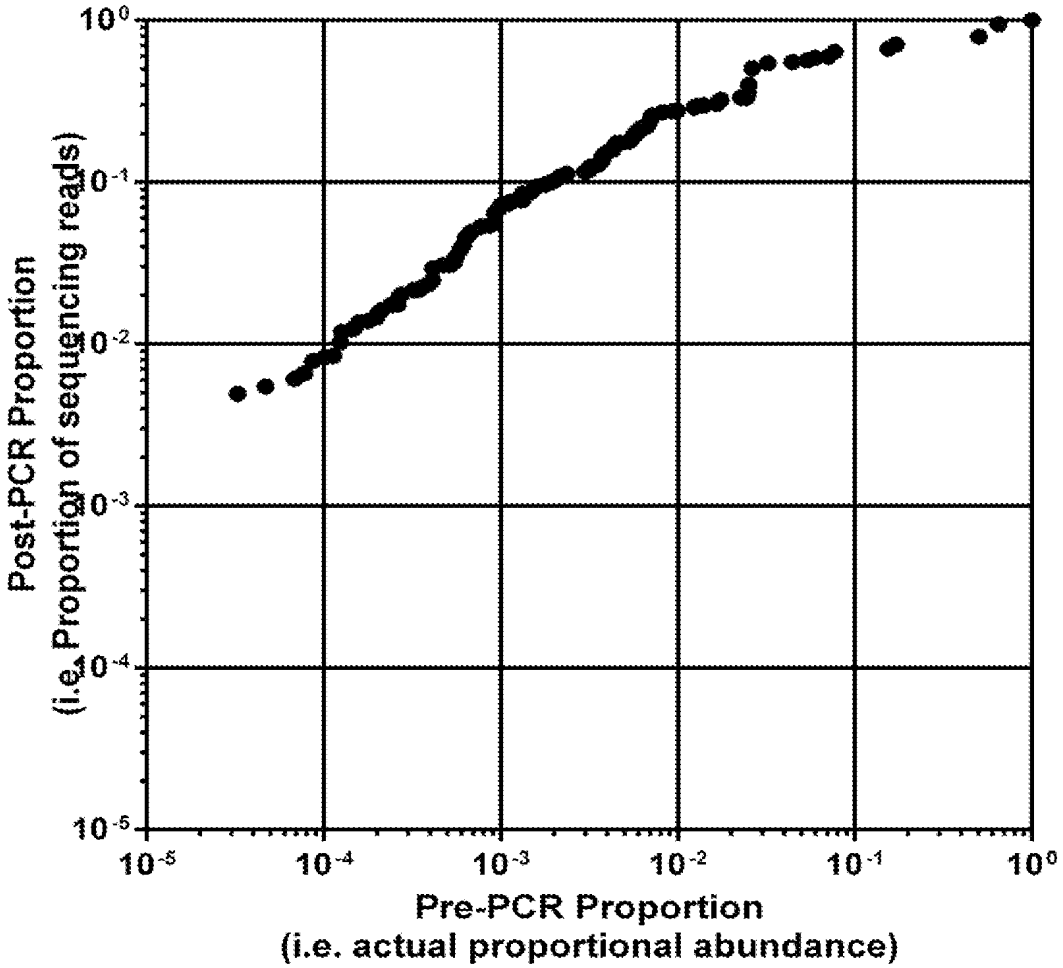

Post-PCR Proportion
(i.e. Proportion of sequencing reads)
10^0
10^-1
10^-2
10^-3
10^-4
10^-5
10^-5
10^-4
10^-3
10^-2
10^-1
10^0
Pre-PCR Proportion
(i.e. actual proportional abundance)

FIG. 17A

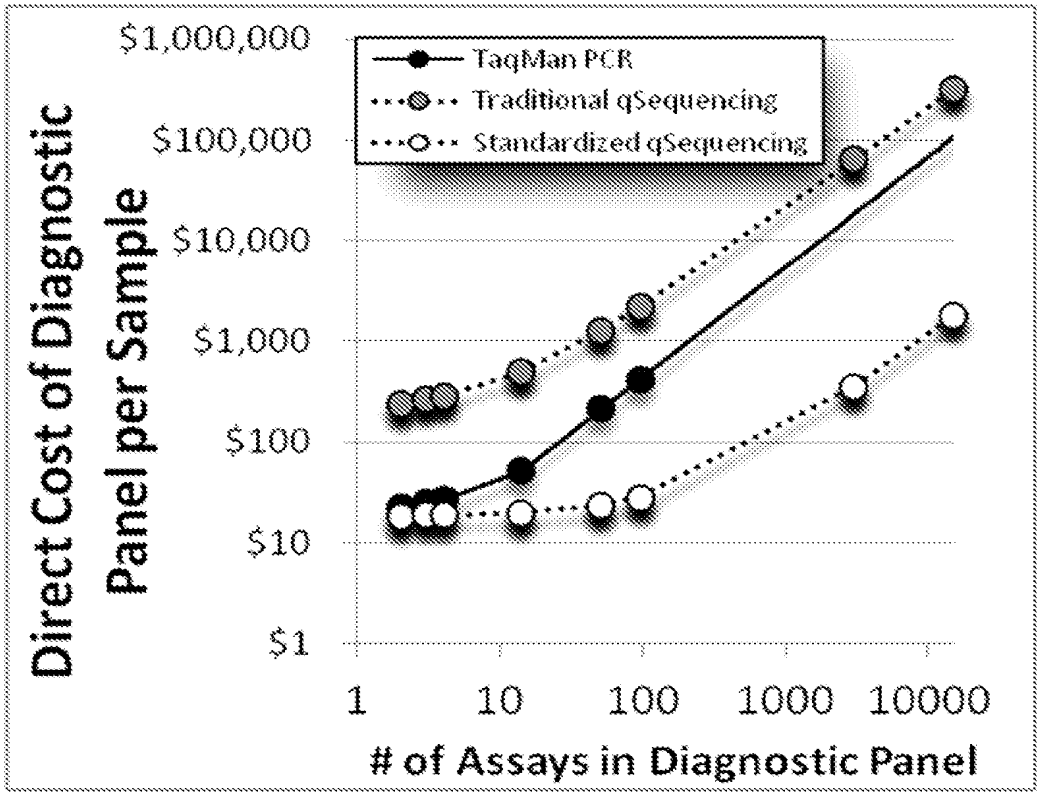

$1,000,000
$100,000
$10,000
$1,000
$100
$10
$1
Direct Cost of Diagnostic Panel per Sample
TaqMan PCR
Traditional qSequencing
Standardized qSequencing
1
10
100
1000
10000
of Assays in Diagnostic Panel

FIG. 17B

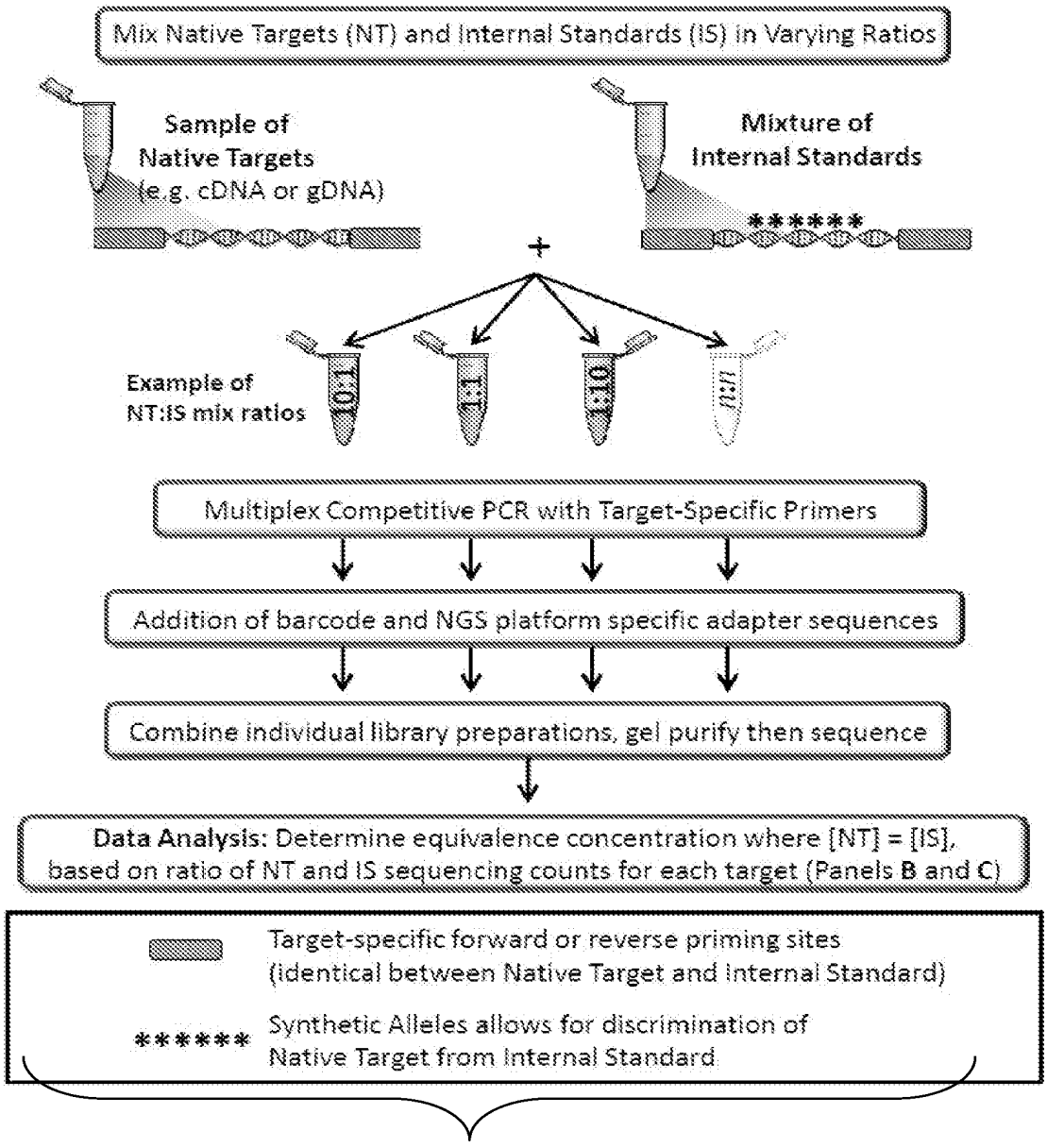

Mix Native Targets (NT) and Internal Standards (IS) in Varying Ratios
Sample of Native Targets (e.g. cDNA or gDNA)
Mixture of Internal Standards
******
+
Example of NT:IS mix ratios
10:1
1:1
1:10
n:n
Multiplex Competitive PCR with Target-Specific Primers
Addition of barcode and NGS platform specific adapter sequences
Combine individual library preparations, gel purify then sequence
Data Analysis: Determine equivalence concentration where [NT] = [IS], based on ratio of NT and IS sequencing counts for each target (Panels B and C)
Target-specific forward or reverse priming sites (identical between Native Target and Internal Standard)
****** Synthetic Alleles allows for discrimination of Native Target from Internal Standard

FIG. 18A

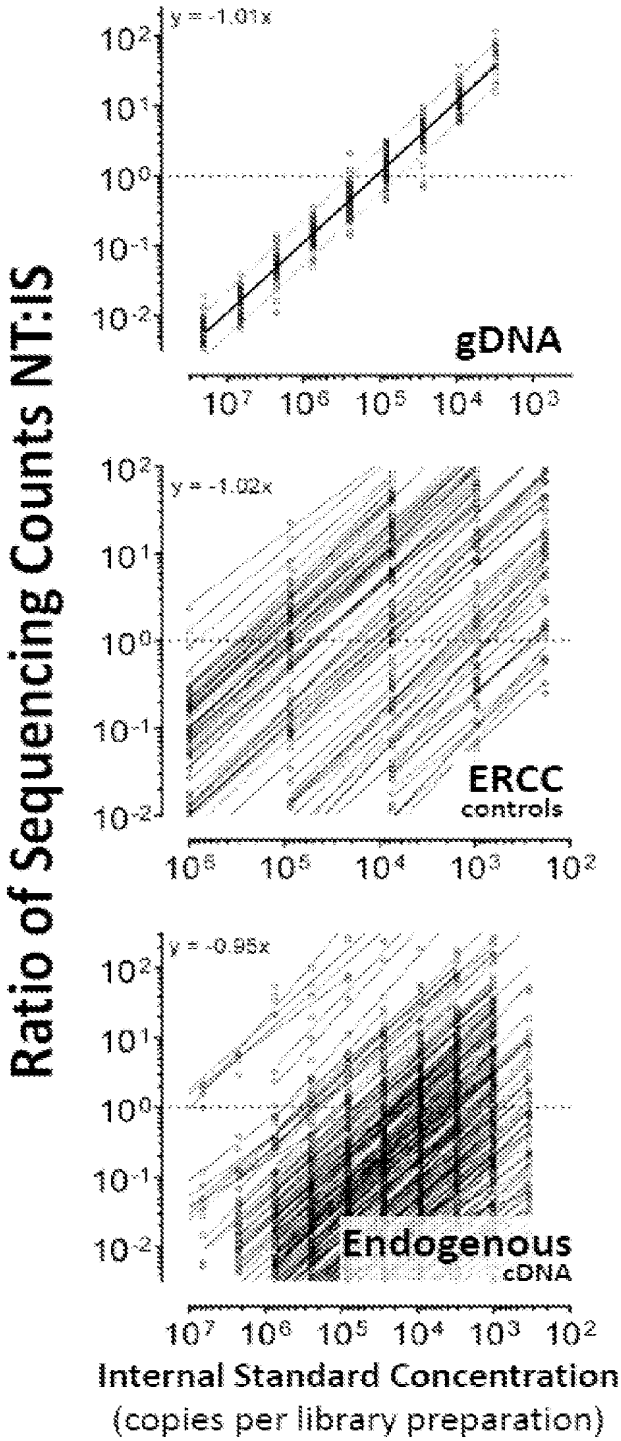

Ratio of Sequencing Counts NT:IS
y = -1.01x
gDNA
y = -1.02x
ERCC
controls
y = -0.95x
Endogenous
cDNA
Internal Standard Concentration
(copies per library preparation)

FIG. 18B

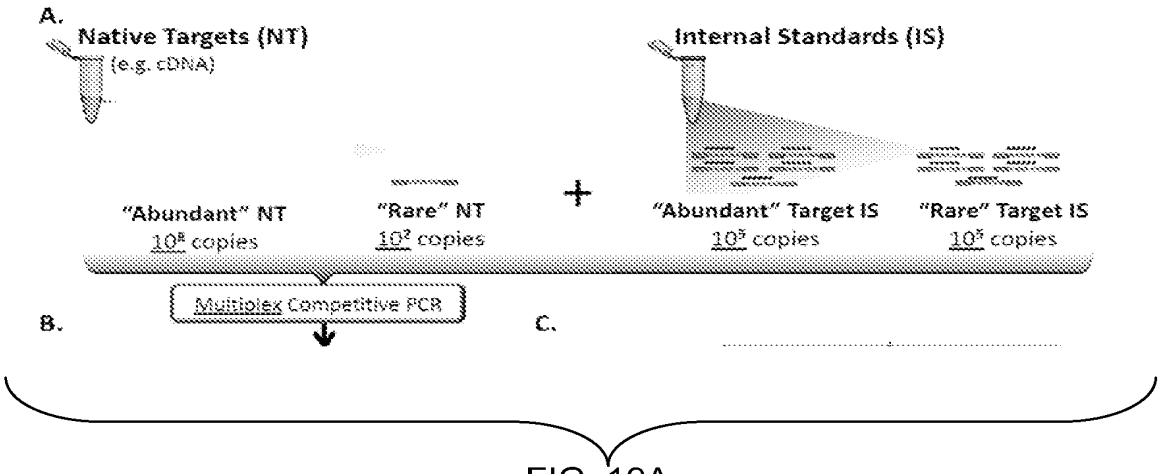


FIG. 19A

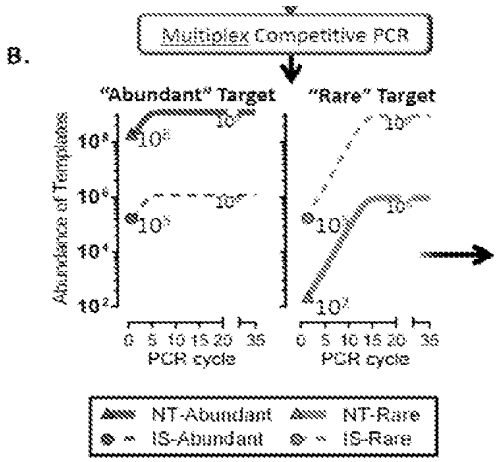


FIG. 19B

C.

	Before PCR		After PCR	
	Abundant	Rare	Abundant	Rare
Native Target (NT) Copies	10^8	10^2	10^8	10^2
Internal Standard (IS) Copies	10^5	10^5	10^5	10^5
NT:IS Ratio	1000:1	1:1000	1000:1	1:1000
Range of all templates	$(10^8/10^2) =$ **10^6**		$(10^8/10^5) =$ **10^3**	

FIG. 19C

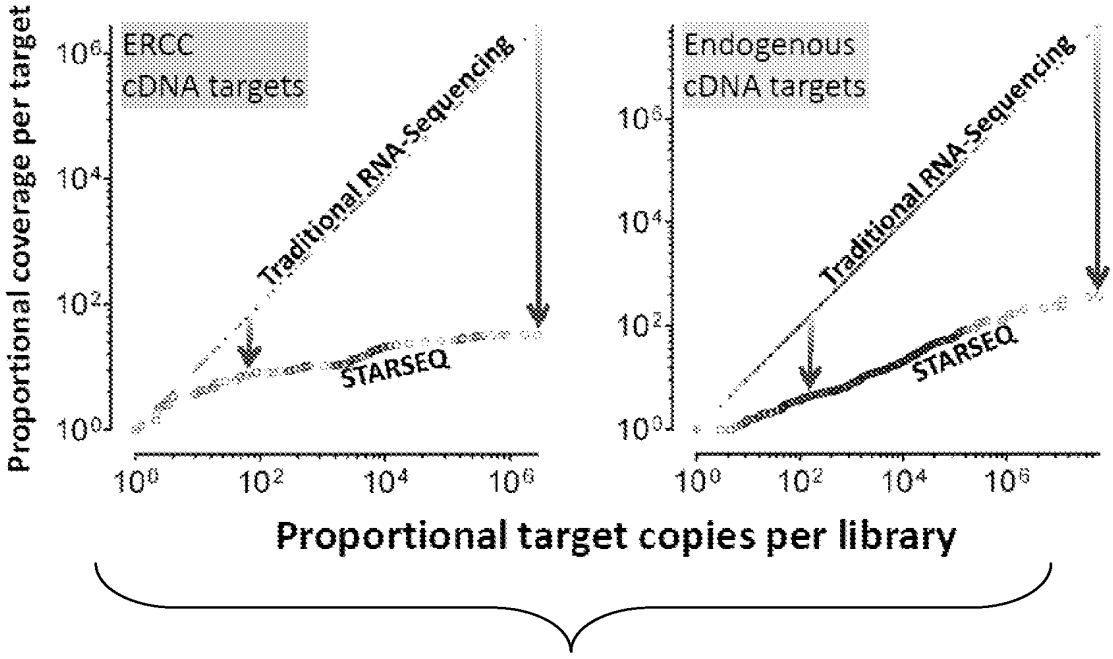


FIG. 20A

	Sequencing Reads	
	ERCC	Endogenous
Traditional RNA-Seq.	1.0×10^7	1.5×10^8
	÷	÷
STARSEQ	1.5×10^3	9.5×10^3
Fold Reduction	**6.9×10^3**	**1.6×10^4**

FIG. 20B

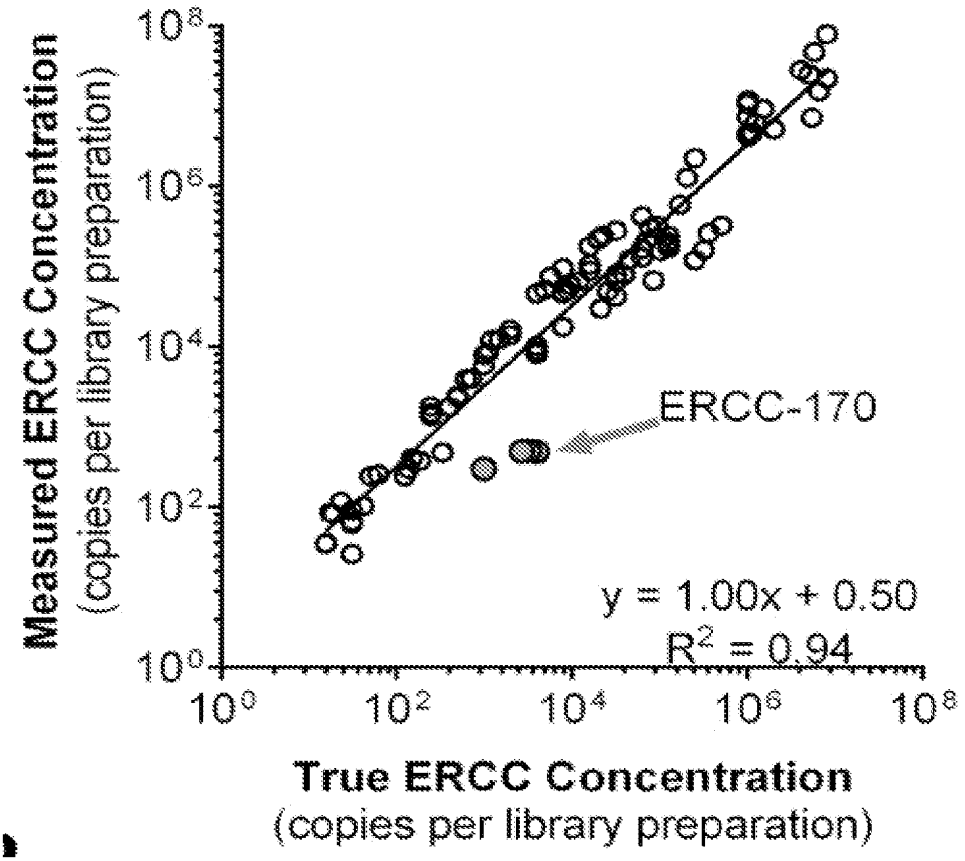

Measured ERCC Concentration
(copies per library preparation)
10^8
10^6
10^4
10^2
10^0
ERCC-170
y = 1.00x + 0.50
R^2 = 0.94
10^0
10^2
10^4
10^6
10^8
True ERCC Concentration
(copies per library preparation)

FIG. 21A

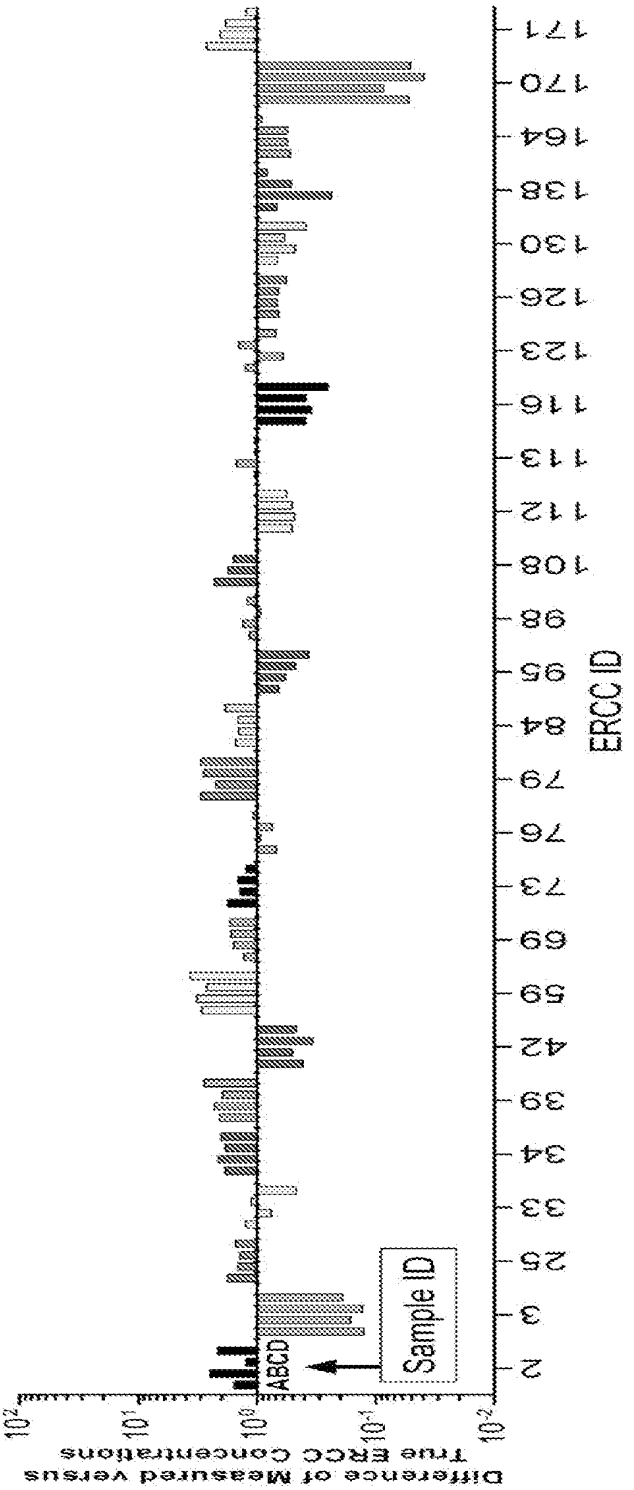

Difference of Measured versus True ERCC Concentrations
10^{2}
10^{1}
10^{0}
10^{-1}
10^{-2}
ABCD
Sample ID
2
3
25
33
34
39
42
59
69
73
76
79
84
95
98
108
112
113
116
123
126
130
138
164
170
171
ERCC ID

FIG. 21B

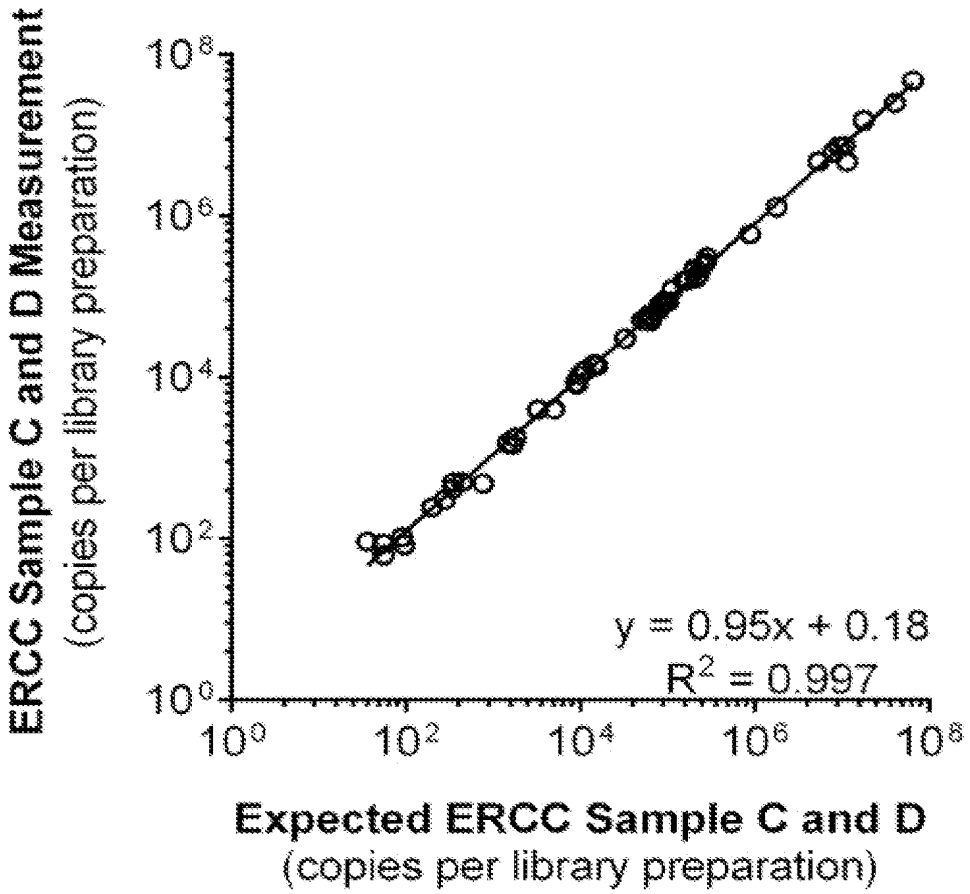

ERCC Sample C and D Measurement
(copies per library preparation)
10^8
10^6
10^4
10^2
10^0
y = 0.95x + 0.18
R^2 = 0.997
10^0
10^2
10^4
10^6
10^8
Expected ERCC Sample C and D
(copies per library preparation)

FIG. 21C

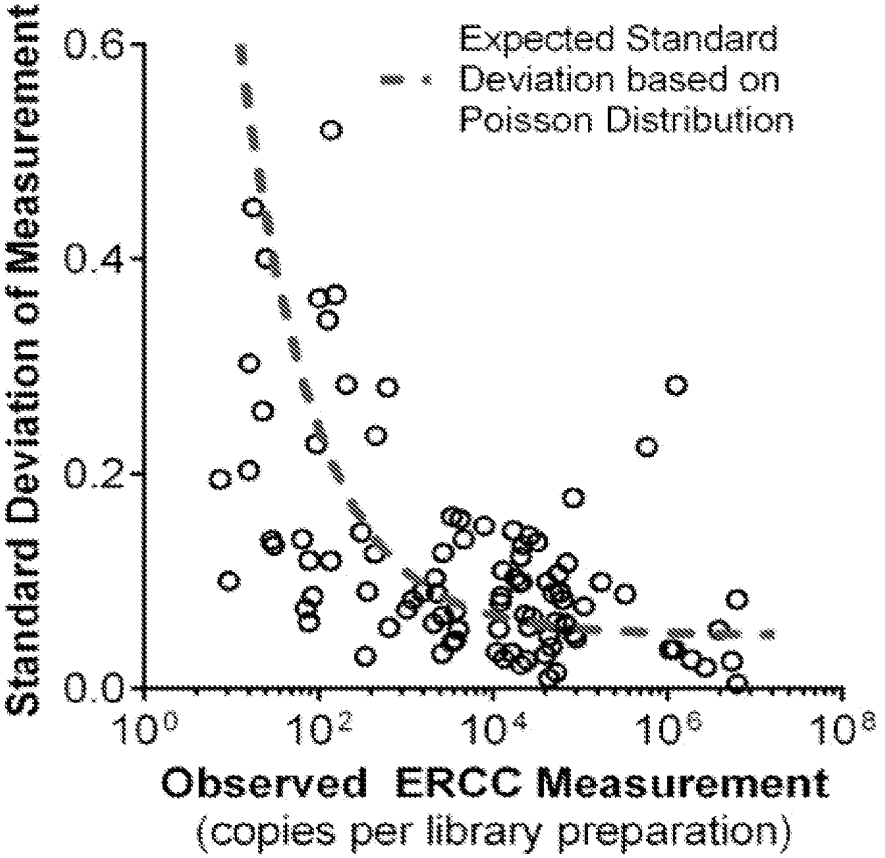

Standard Deviation of Measurement
0.6
0.4
0.2
0.0
Expected Standard Deviation based on Poisson Distribution
10^0
10^2
10^4
10^6
10^8
Observed ERCC Measurement
(copies per library preparation)

FIG. 21D

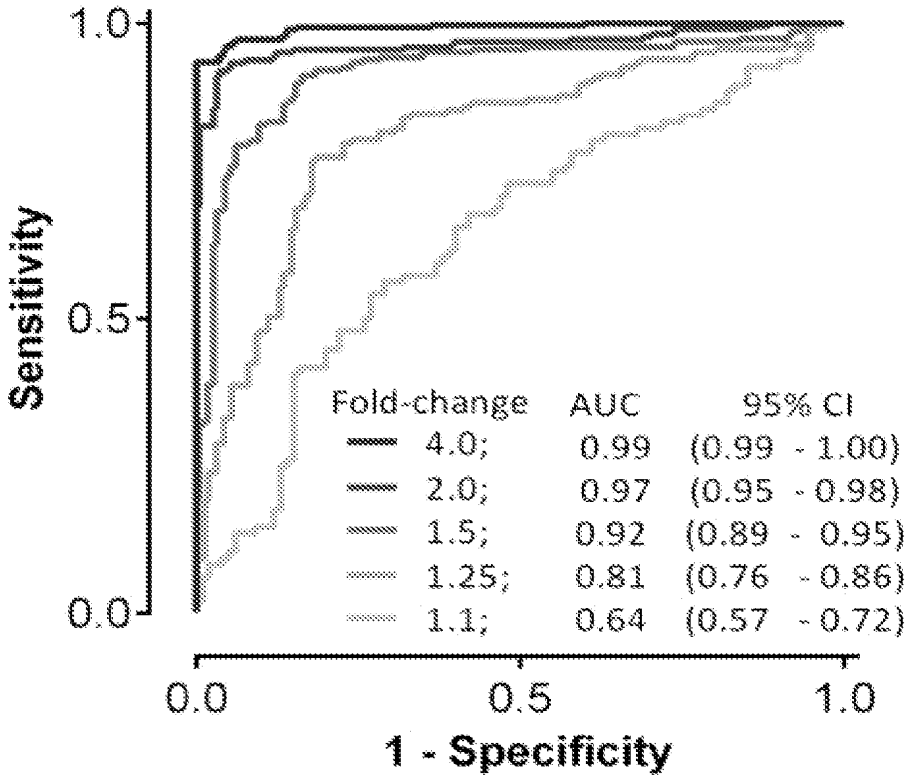

1.0
0.5
0.0
Sensitivity
Fold-change AUC 95% CI
4.0; 0.99 (0.99 - 1.00)
2.0; 0.97 (0.95 - 0.98)
1.5; 0.92 (0.89 - 0.95)
1.25; 0.81 (0.76 - 0.86)
1.1; 0.64 (0.57 - 0.72)
0.0
0.5
1.0
1 - Specificity

FIG. 21E

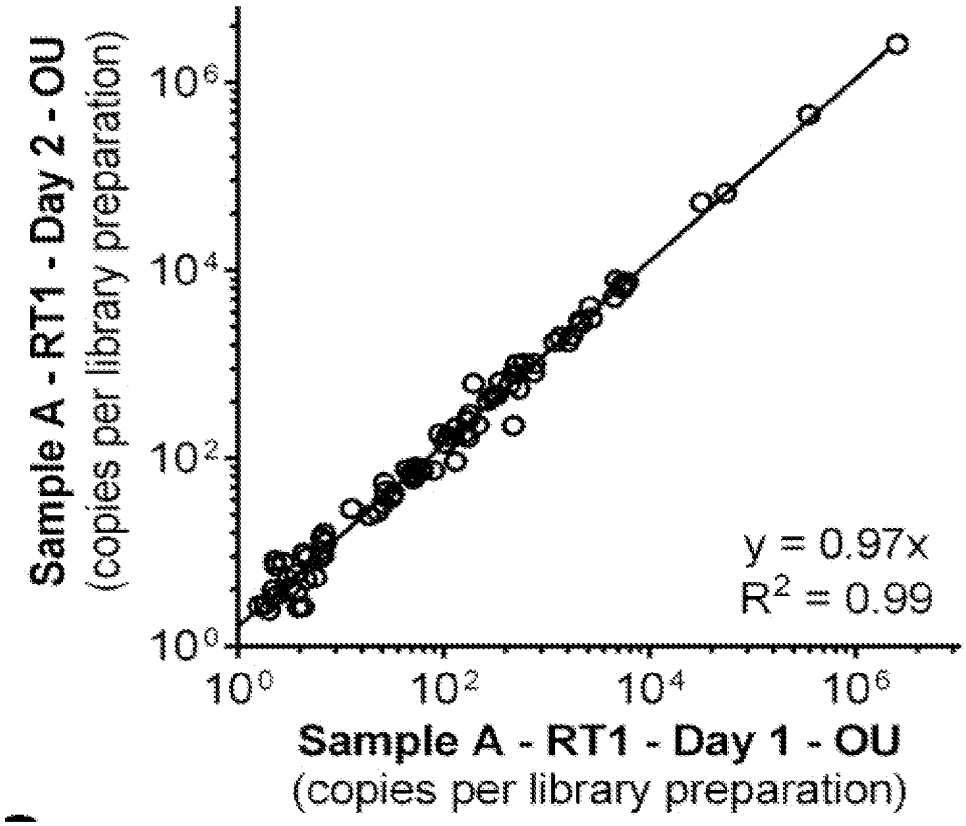

Sample A - RT1 - Day 2 - OU
(copies per library preparation)
10^6
10^4
10^2
10^0
y = 0.97x
R^2 = 0.99
10^0
10^2
10^4
10^6
Sample A - RT1 - Day 1 - OU
(copies per library preparation)

FIG. 22A

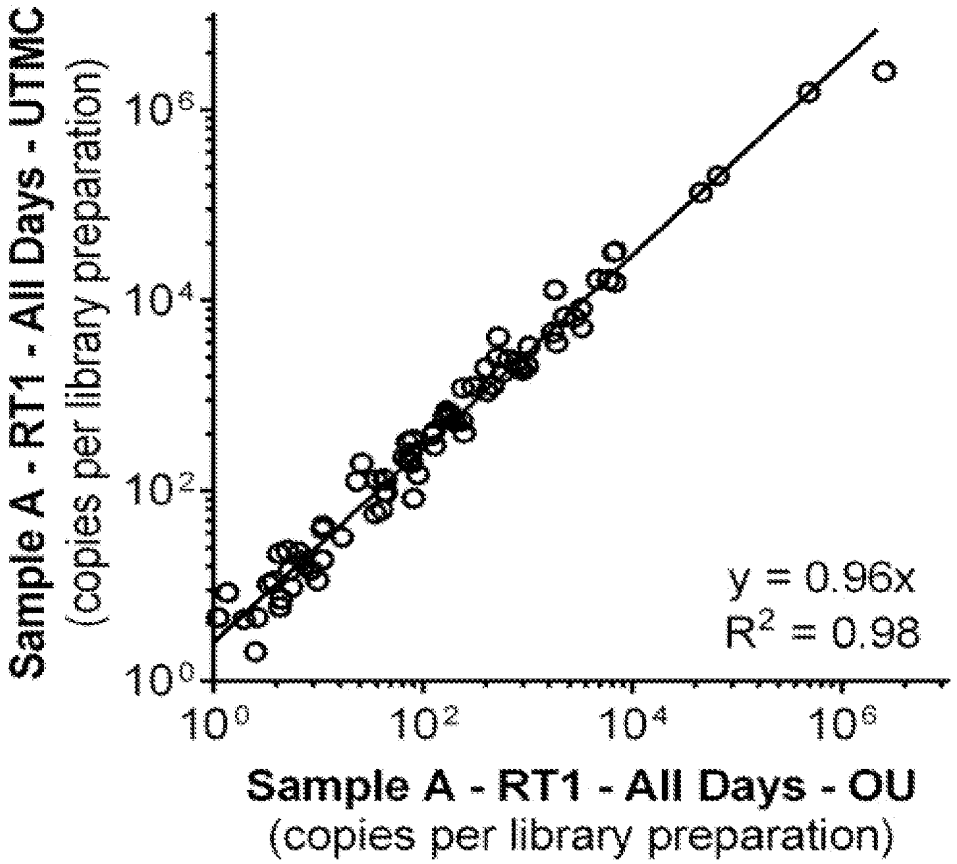

Sample A - RT1 - All Days - UTMC
(copies per library preparation)
10^6
10^4
10^2
10^0
y = 0.96x
R^2 = 0.98
10^0
10^2
10^4
10^6
Sample A - RT1 - All Days - OU
(copies per library preparation)

FIG. 22B

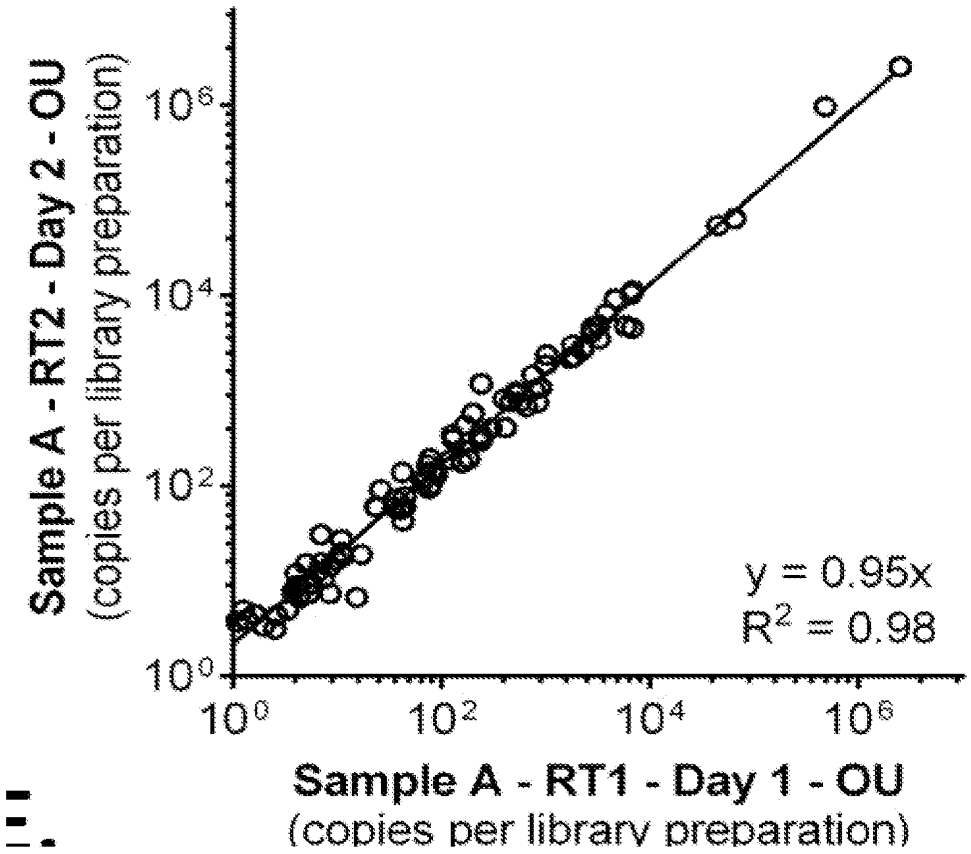

Sample A - RT2 - Day 2 - OU
(copies per library preparation)
10^6
10^4
10^2
10^0
y = 0.95x
R^2 = 0.98
10^0
10^2
10^4
10^6
Sample A - RT1 - Day 1 - OU
(copies per library preparation)

FIG. 22C

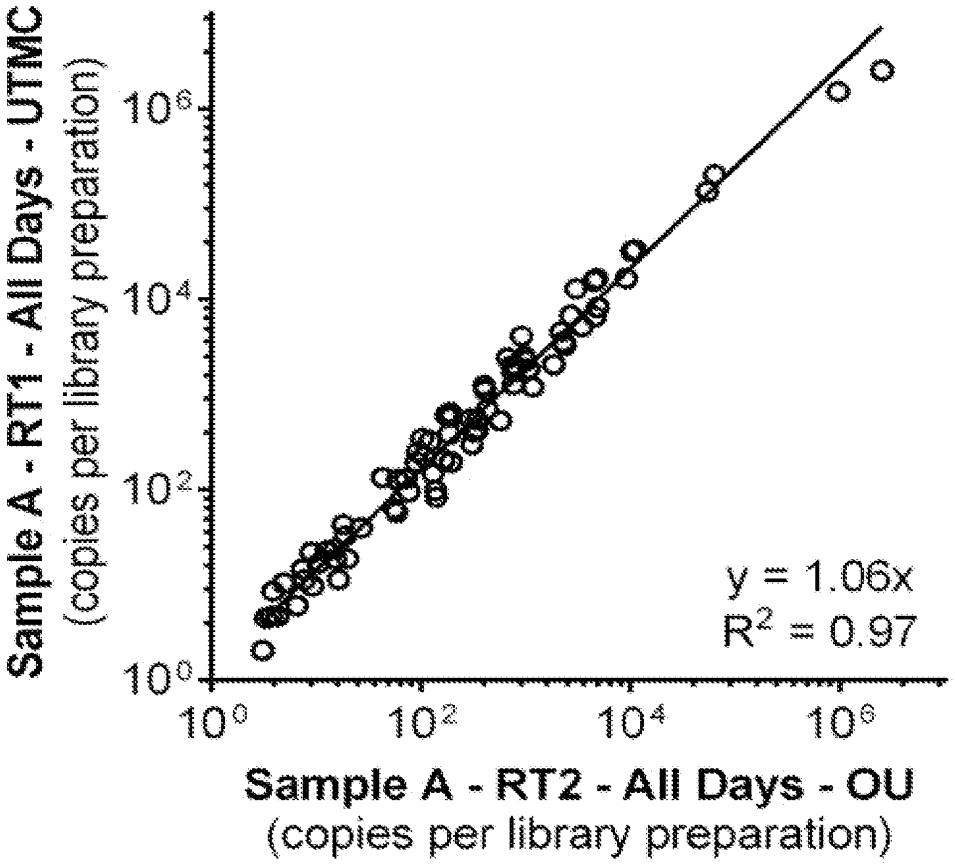

Sample A - RT1 - All Days - UTMC
(copies per library preparation)
10^6
10^4
10^2
10^0
y = 1.06x
R^2 = 0.97
10^0
10^2
10^4
10^6
Sample A - RT2 - All Days - OU
(copies per library preparation)

FIG. 22D

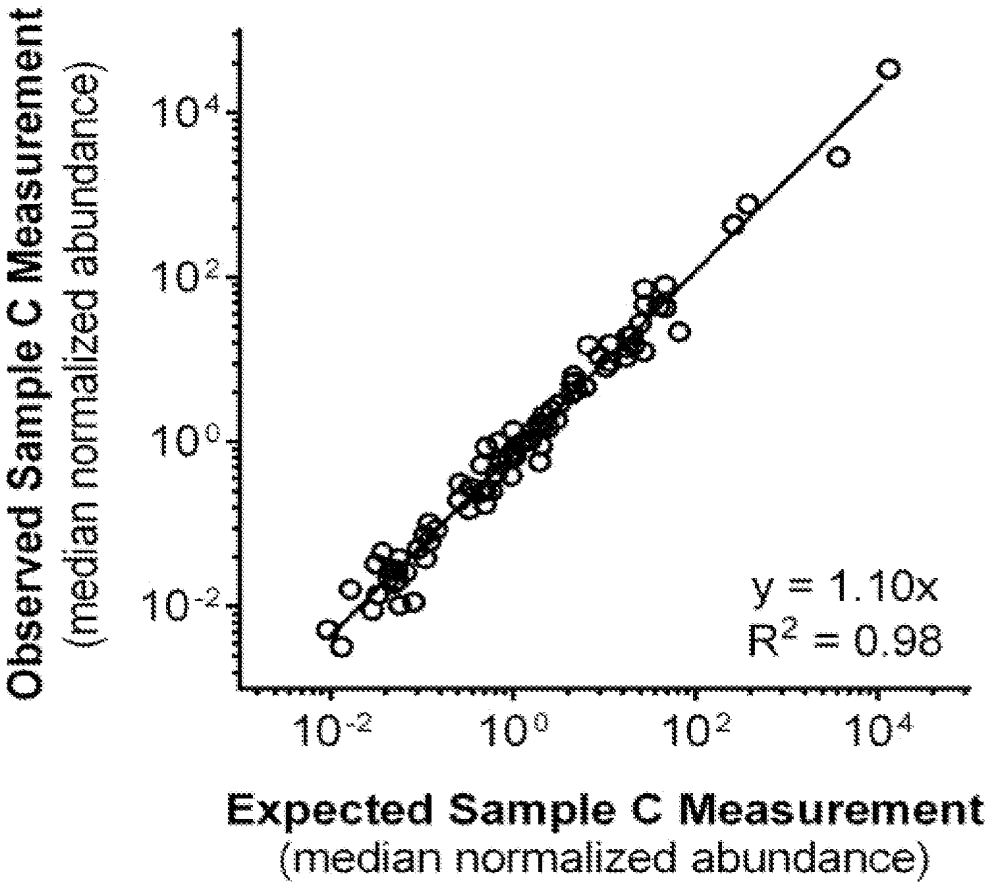

Observed Sample C Measurement
(median normalized abundance)
10^4
10^2
10^0
10^-2
y = 1.10x
R^2 = 0.98
10^-2
10^0
10^2
10^4
Expected Sample C Measurement
(median normalized abundance)

FIG. 22E

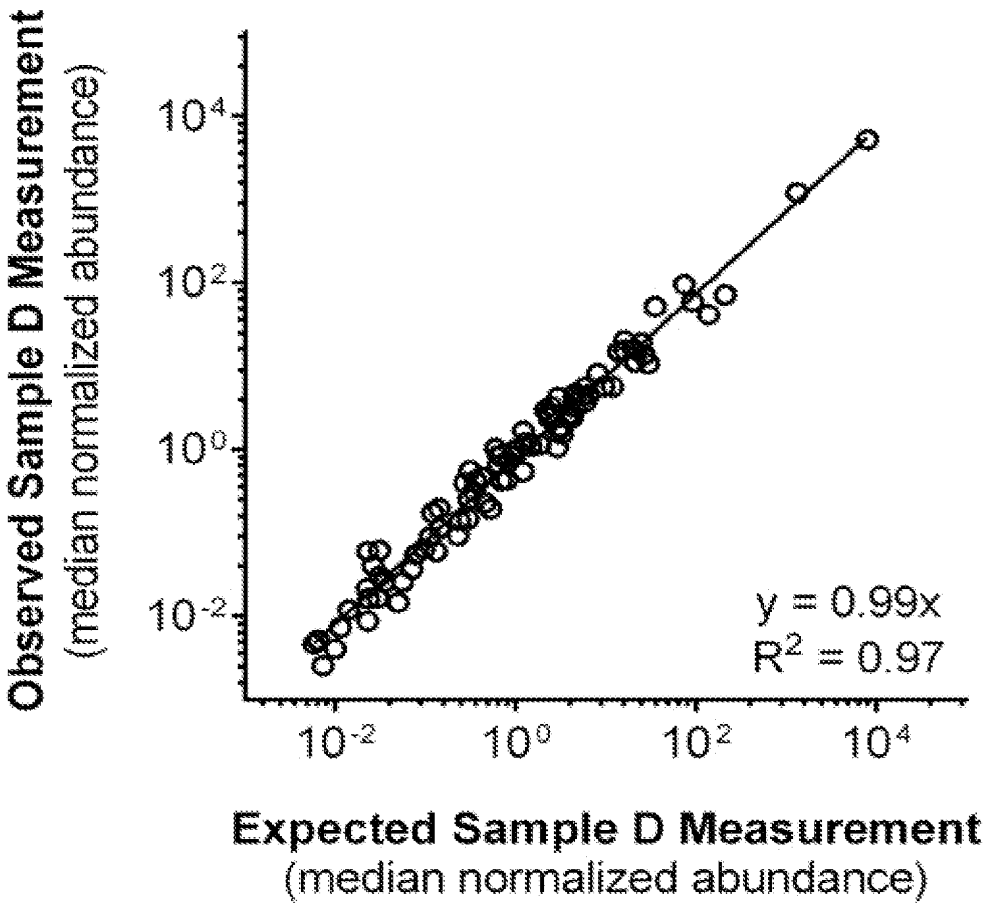

Observed Sample D Measurement
(median normalized abundance)
10^4
10^2
10^0
10^-2
y = 0.99x
R^2 = 0.97
10^-2
10^0
10^2
10^4
Expected Sample D Measurement
(median normalized abundance)

FIG. 22F

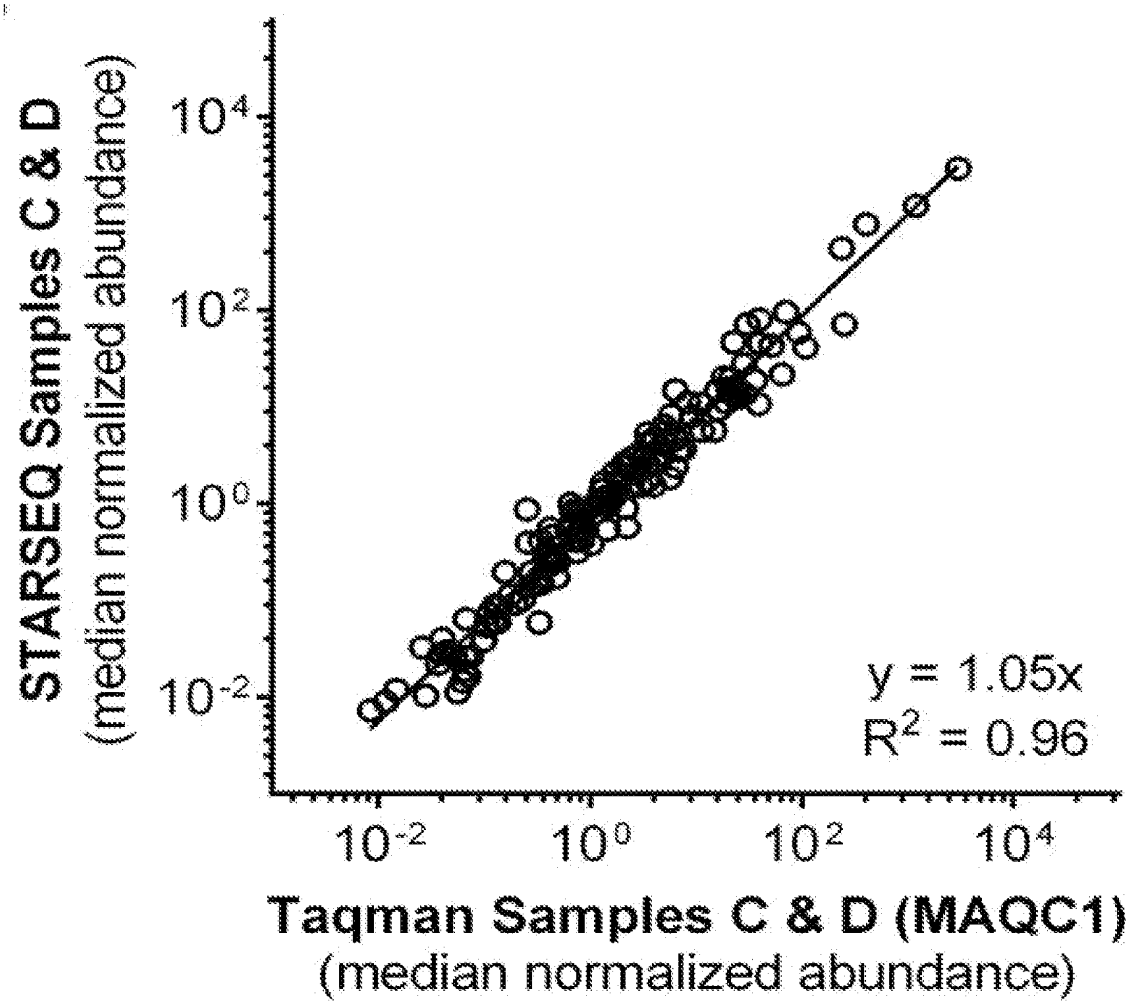

STARSEQ Samples C & D
(median normalized abundance)
10^4
10^2
10^0
10^{-2}
y = 1.05x
R^2 = 0.96
10^{-2}
10^0
10^2
10^4
Taqman Samples C & D (MAQC1)
(median normalized abundance)

FIG. 23A

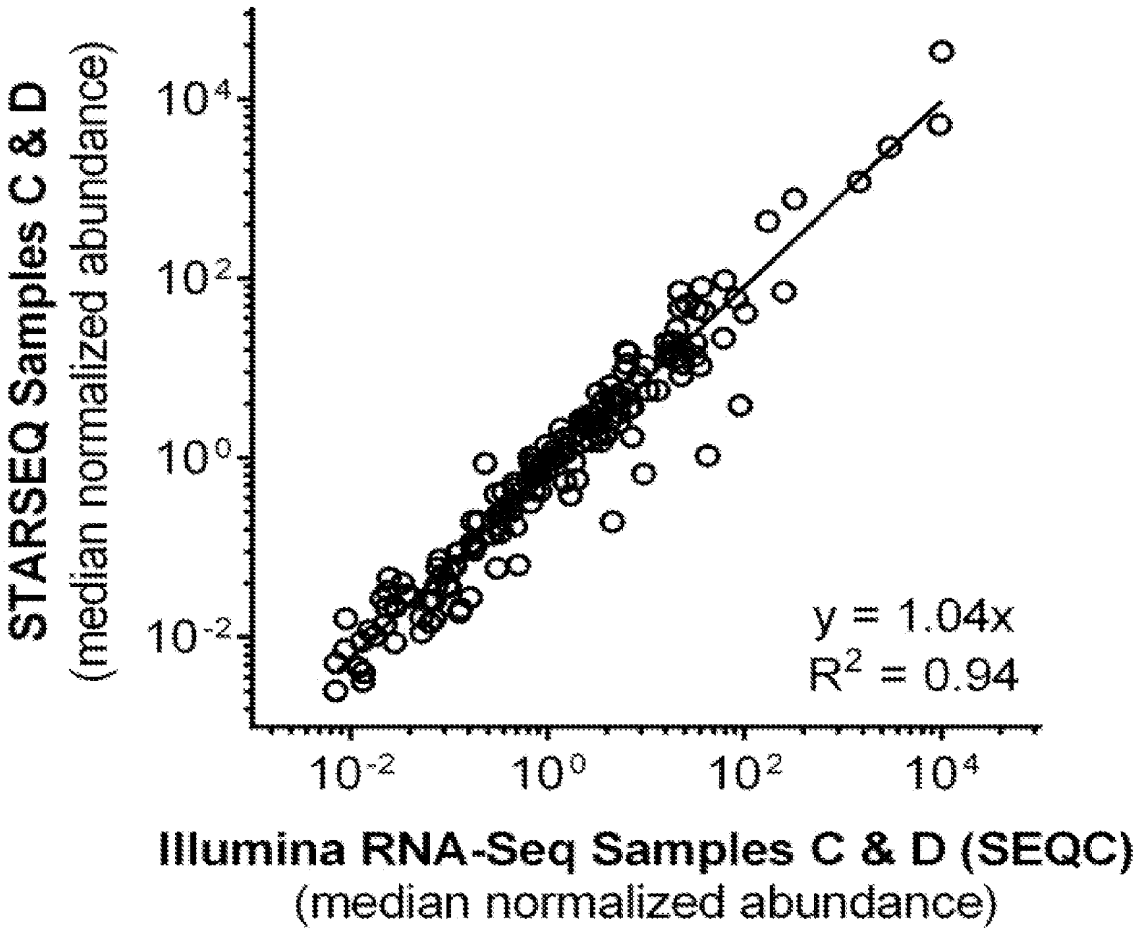

STARSEQ Samples C & D
(median normalized abundance)
10^4
10^2
10^0
10^{-2}
y = 1.04x
R^2 = 0.94
10^{-2}
10^0
10^2
10^4
Illumina RNA-Seq Samples C & D (SEQC)
(median normalized abundance)

FIG. 23B

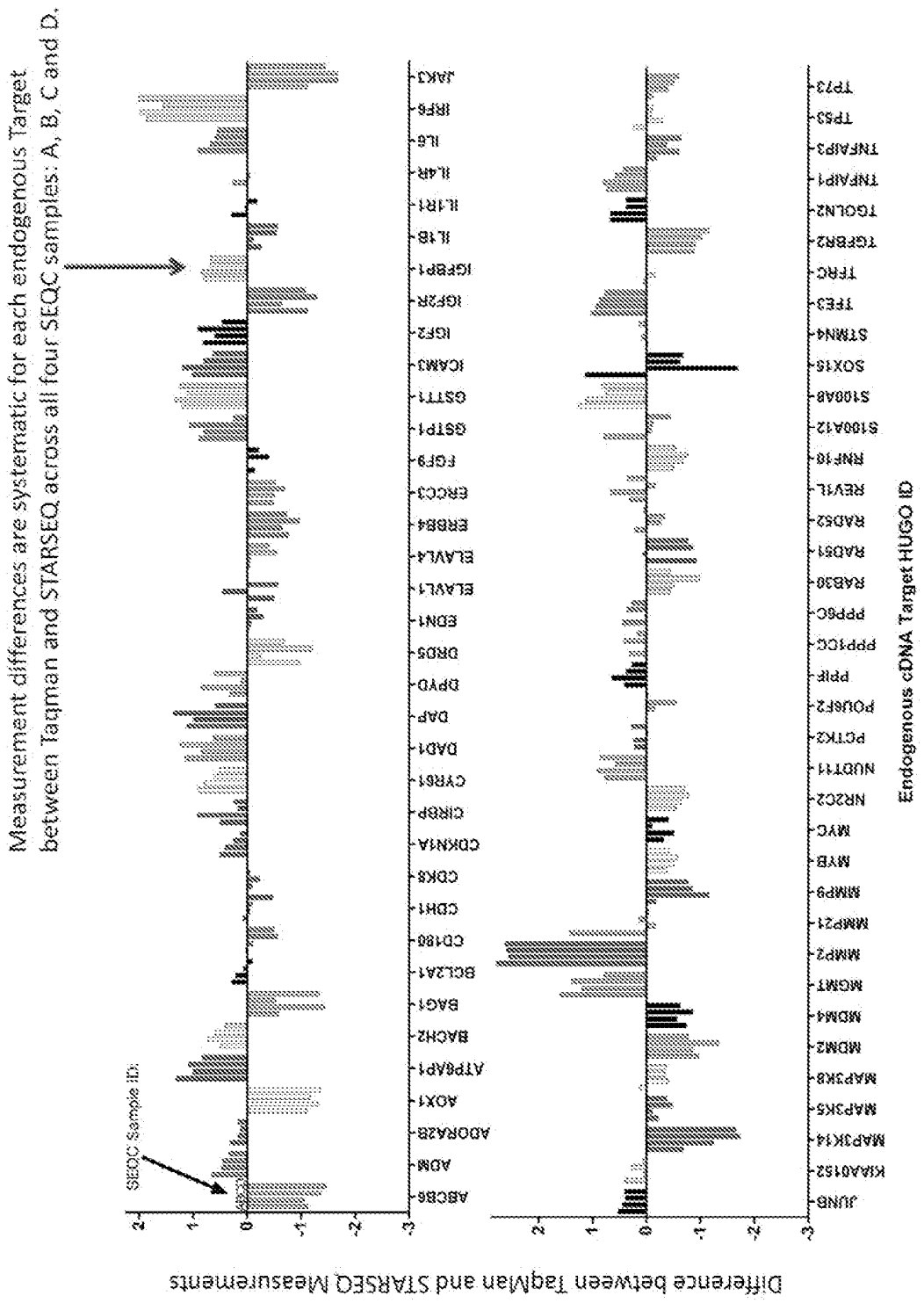

Measurement differences are systematic for each endogenous Target between Taqman and STARSEQ across all four SEQC samples: A, B, C and D.
SEQC Sample ID:
Difference between TaqMan and STARSEQ Measurements
Endogenous cDNA Target HUGO ID

FIG. 24

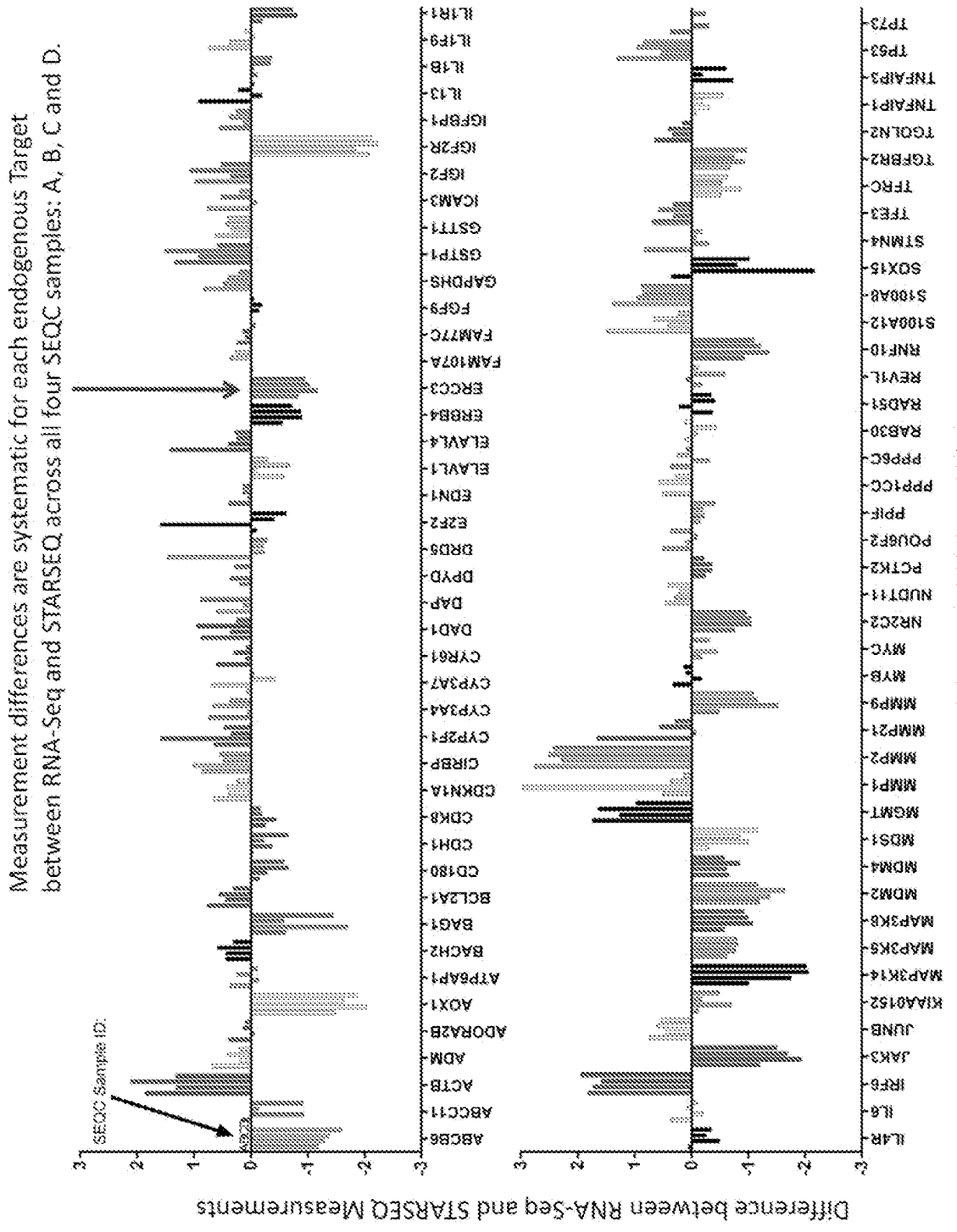

Measurement differences are systematic for each endogenous Target
between RNA-Seq and STARSEQ across all four SEQC samples: A, B, C and D.
SEQC Sample ID:
9223
Difference between RNA-Seq and STARSEQ Measurements
Endogenous cDNA Target HUGO ID

FIG. 25

ERCC Target Measurements	A	B	C	D
Valid	26	26	26	26
True negative	0	0	0	0
Failed	1	1	1	1

Endogenous Target Measurements	A	B	C	D	gDNA
Valid	99	99	96	107	123
True Negative	4	12	6	4	0
Failed	20	12	21	12	0

FIG. 26

HUGO and Sample ID	STARSEQ	Taqman	RNA-Seq
RPL3L Sample-D	< -2.4879	-3.24	-2.90
ANXA13 Sample-C	< -2.127	-3.20	-2.25
MMP7 Sample-C	< -1.7003	-3.27	-2.23
KRT24 Sample-D	< -1.6589	-3.00	-2.31
KIAA0101 Sample-B	< -1.3856	-2.24	-1.64
KRT24 Sample-B	< -1.2609	-2.83	-1.88
STMN2 Sample-A	< -1.0849	-1.47	-1.92
IL18R1 Sample-B	< -0.938	-2.28	-0.81
DPP4 Sample-B	< -0.733	-1.64	-0.61
IL18R1 Sample-D	< -0.7132	-2.37	-1.10
KIT Sample-C	< -0.0946	-0.80	-0.18
KIT Sample-A	< 0.1187	-1.21	-0.53
IL1F6 Sample-B	< -1.9726	ND	ND
POU1F1 Sample-C	< -1.9169	ND	-3.95
POU1F1 Sample-B	< -1.8552	ND	-2.62
RPL3L Sample-B	< -1.8183	ND	-3.15
KRT24 Sample-C	< -1.6485	ND	-2.72
POU1F1 Sample-A	< -1.5431	ND	-4.18
CYP2C9 Sample-D	< -1.1907	ND	-3.41
ABCB11 Sample-B	< -1.1123	ND	-2.38
SERPINB7 Sample-B	< -1.0772	ND	-3.15
DLG7 Sample-B	< -0.8081	ND	-2.31
FOXA1 Sample-B	< -0.4969	ND	-2.42
CYP2C9 Sample-C	< 0.37	ND	-3.20
CYP2C9 Sample-B	< 0.6952	ND	-2.92
CYP2C9 Sample-A	< 0.7906	ND	-3.16

FIG. 27

Intra-assay; Intra-sample SD	0.08
Intra-assay; Inter-sample SD	0.18
Inter-assay; Inter-sample SD	0.40

FIG. 28

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2013/071656

A. CLASSIFICATION OF SUBJECT MATTER
IPC(8) - C40B 20/04 (2014.01)
USPC - 506/17
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) - C12N 15/11; C12Q 1/68; C40B 20/04 (2014.01)
USPC - 435/91.5; 506/4, 17, 41

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
CPC - C12Q1/00, C12Q 1/6806, 1/686 (2013.01)

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
Orbit, Google Patents, Google Scholar, PubMed

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 2006/0188908 A1 (WILLEY et al) 24 August 2006 (24.08.2006) entire document	1, 2, 7, 10, 20, 21 --- 22
Y	SMITH et al., "Highly-multiplexed barcode sequencing: an efficient method for parallel analysis of pooled samples," Nucleic Acids Research, Vol. 38, No. 13, e142, Pgs. 1-7. 11 May 2010 (11.05.2010) entire document.	22
A	CROSBY et al., "Gene capture and random amplification for quantitative recovery of homologous genes," Mol Cell Probes. Vol. 21, No. 2, Pgs. 140-147. 30 September 2006 (30.09.2006) entire document	3-6, 8-10, 12, 13
A	HINDSON et al., "High-throughput droplet digital PCR system for absolute quantitation of DNA copy number," Anal Chem. Vol. 83, No. 22, Pgs. 8604-8610. 28 October 2011(28.10.2011) entire document	3-6, 8-10, 12, 13
A	US 2009/0221438 A1 (KITZMAN et al) 03 September 2009 (03.09.2009) entire document	3-6, 8-10, 12, 13
A	MAMANOVA et al., "Target-enrichment strategies for next-generation sequencing," Nat Methods, Vol. 7, No. 2, Pgs.111-118. February 2010. entire document	3-6, 8-10, 12, 13

☐ 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
23 January 2014	06 FEB 2014
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: Blaine R. Copenheaver 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/US2013/071656

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.: 11,14-19, 23-48
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

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

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

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

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

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

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

Remark on Protest

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

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

☐ No protest accompanied the payment of additional search fees.

IDT Exhibit 1020.5823

Espacenet

Bibliographic data: WO2009053039 (A1) — 2009-04-30

METHODS AND SYSTEMS FOR SOLUTION BASED SEQUENCE ENRICHMENT AND ANALYSIS OF GENOMIC REGIONS

Inventor(s): ALBERT THOMAS [US]; RODESCH MATTHEW [US] ± (ALBERT, THOMAS, ; RODESCH, MATTHEW)

Applicant(s): ROCHE DIAGNOSTICS GMBH [DE]; HOFFMANN LA ROCHE [CH]; ALBERT THOMAS [US]; RODESCH MATTHEW [US] ± (ROCHE DIAGNOSTICS GMBH, ; F. HOFFMANN-LA ROCHE AG, ; ALBERT, THOMAS, ; RODESCH, MATTHEW)

Classification: **- international:** ***C12Q1/68***
- cooperative: **C12Q1/6806**; **C12Q1/6813**; **C12Q1/6844**; C12N15/1006 more

Application number: WO2008EP08920 20081022 Global Dossier

Priority number(s): EP20070020660 20071023 ; US20080194574 20080820

Also published as: CA2699835 (A1) CA2699835 (C) CN101835907 (A) CN101835907 (B) EP2053132 (A1) more

Abstract of WO2009053039 (A1)

The present invention provides novel methods and systems for capturing and enrichment of target nucleic acids for reducing the complexity of target nucleic acids, preferably a genomic sample, for further analysis such as direct DNA sequencing, resequencing or SNP calling. In particular, the present invention provides for the enrichment of targeted sequences in a solution based format.

IDT Exhibit 1020.5824

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
30 April 2009 (30.04.2009)

PCT

(10) International Publication Number
WO 2009/053039 A1

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

(21) International Application Number:
PCT/EP2008/008920

(22) International Filing Date: 22 October 2008 (22.10.2008)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
07020660.2 23 October 2007 (23.10.2007) EP
12/194,574 20 August 2008 (20.08.2008) US

(71) Applicant *(for DE only)*: **ROCHE DIAGNOSTICS GMBH** [DE/DE]; Sandhofer Strasse 116, 68305 Mannheim (DE).

(71) Applicant *(for all designated States except DE, US)*: **F. HOFFMANN-LA ROCHE AG** [CH/CH]; Grenzacherstrasse 124, CH-4070 Basel (CH).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **ALBERT, Thomas** [US/US]; 2738 Tower Hill Dr., Fitchburg, WI 53711 (US). **RODESCH, Matthew** [US/US]; 205 W. Broadway St., Stoughton, WI 53589 (US).

(74) Agent: **BÖSL, Raphael**; Isenbruck Bösl Hörschler Wichmann Huhn LLP, Prinzregentenstrasse 68, 81675 München (DE).

(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, 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, 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, MT, NL, NO, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*
— *with sequence listing part of description published separately in electronic form and available upon request from the International Bureau*

WO 2009/053039 A1

(54) Title: METHODS AND SYSTEMS FOR SOLUTION BASED SEQUENCE ENRICHMENT AND ANALYSIS OF GENOMIC REGIONS

(57) Abstract: The present invention provides novel methods and systems for capturing and enrichment of target nucleic acids for reducing the complexity of target nucleic acids, preferably a genomic sample, for further analysis such as direct DNA sequencing, resequencing or SNP calling. In particular, the present invention provides for the enrichment of targeted sequences in a solution based format.

IDT Exhibit 1020.5825

METHODS AND SYSTEMS FOR SOLUTION BASED SEQUENCE ENRICHMENT AND ANALYSIS OF GENOMIC REGIONS

FIELD OF THE INVENTION

The present application relates to the field of enrichment and analysis of nucleic acid sequences by means of capturing said sequences onto a solid support. More precisely, the present invention provides a new method to capture specific genomic regions for subsequent further analysis, if the region of interest is too large to be amplified by only one or a few PCR reactions. In particular, the present invention provides for the enrichment of targeted sequences in a solution based format.

BACKGROUND OF THE INVENTION

The advent of nucleic acid microarray technology, for example DNA microarray technology, makes it possible to build an array of millions of nucleic acid sequences, for example DNA sequences, in a very small area, for example on a microscope slide (e.g., US Patent Nos. 6,375,903 and 5,143,854). Initially, such arrays were created by spotting pre-synthesized DNA sequences onto slides. However, the construction of maskless array synthesizers (MAS) in which light is used to direct synthesis of the DNA sequences, the light direction being performed using a digital micromirror device (DMD). as described in US Patent No. 6,375,903 now allows for the *in situ* synthesis of oligonucleotide sequences directly on the slide itself.

Using a MAS instrument, the selection of oligonucleotide sequences or DNA sequences to be constructed on the microarray is under software control such that it is now possible to create individually customized arrays based on the particular needs of an investigator. In general, MAS-based oligonucleotide or DNA microarray synthesis technology allows for the parallel synthesis of millions of unique oligonucleotide features in a very small area of a standard microscope slide. The microarrays are generally synthesized by using light to

direct which oligonucleotides are synthesized at specific locations on an array, these locations being called features

With the availability of the entire genomes of hundreds of organisms, for which a reference sequence has generally been deposited into a public database, microarrays have been used to perform sequence analysis on nucleic acids or DNA isolated from a myriad of organisms.

Nucleic acid or DNA microarray technology has been applied to many areas of research and diagnostics, such as gene expression and discovery, mutation detection, allelic and evolutionary sequence comparison, genome mapping, drug discovery, and more. Many applications require searching for genetic variants and mutations across the entire human genome that underly human diseases. In the case of complex diseases, these searches generally result in a single nucleotide polymorphism (SNP) or set of SNPs associated with diseases and/or disease risk. Identifying such SNPs has proved to be an arduous and frequently fruitless task because resequencing large regions of genomic DNA, usually greater than 100 kilobases (Kb), from affected individuals or tissue samples is required to find a single base change or to identify all sequence variants. Other applications involve the identification of gains and losses of chromosomal sequences which may also be associated with cancer, such as lymphoma (Martinez-Climent JA et al., 2003, Blood 101:3109-3117), gastric cancer (Weiss MM et al., 2004, Cell. Oncol. 26:307-317), breast cancer (Callagy G et al., 2005, J. Path. 205: 388-396) and prostate cancer (Paris, PL et al., 2004, Hum. Mol. Gen. 13:1303-1313). As such, microarray technology is a tremendously useful tool for scientific investigators and clinicians in their understanding of diseases and therapeutic regimen efficacy in treating diseases.

The genome is typically too complex to be studied as a whole, and techniques must be used to reduce the complexity of the genome. To address this problem, one solution is to reduce certain types of abundant sequences from a genomic nucleic acid or DNA sample, as found in US Patent 6,013,440. Alternatives employ methods and compositions for enriching genomic sequences as described, for example, in Albert et al. (2007, Nat. Meth., 4:903-5), Okou et al. (2007, Nat. Meth. 4:907-9), Olson M. (2007, Nat. Meth. 4:891-892), Hodges et al. (2007, Nat. Genet. 39:1522-1527) and as found in United States Patent Application Serial Nos. 11/638,004, 11/970,949, and 61/032,594. Albert et al. disclose an

IDT Exhibit 1020.5827

alternative that is both cost-effective and rapid in effectively reducing the complexity of a genomic sample in a user defined way to allow for further processing and analysis. Lovett et al. (1991, Proc. Natl. Acad. Sci. 88:9628-9632) also describes a method for genomic selection using a bacterial artificial chromosomes. However, existing methods are limited by, for example, their ease of use and inflexibility of materials and methods.

Prior art microarray technology, be it enrichment technology or otherwise, is typically a substrate associated technology with inherent variability, such as microarray slides, chips, and the like. Variability can take on many forms, for example variability in background, probe/hybridization kinetics, glass source, and the like. Variability plays a big part in experimental interpretation and can make or break an experiment.

As such, what are needed are methods, systems and compositions to provide enrichment of targeted sequences in a format that other than a typical substrate type of microarray format. The advent of new microarray formats will provide additional tools for researchers and clinicians in advancing their knowledge of diseases and disease states.

SUMMARY OF THE INVENTION

The present invention provides methods and systems for the capture and enrichment of target nucleic acids and analysis of the enriched target nucleic acids . In particular, the present invention provides for the enrichment of targeted sequences in a solution based format. Methods and systems of the present invention are useful in aiding investigators and clinicians in identifying, studying and following treatment regimens associated with disease and disease states.

The present invention is summarized as a novel method for reducing the complexity of a large nucleic acid sample, such as a genomic sample, cDNA library or mRNA or mRNA library to facilitate further processing and genetic analysis. Embodiments of the present invention comprise (pre-selected) immobilized or non immobilized nucleic acid probes to capture target nucleic acid sequences from, for example, a genomic sample by hybridizing the sample to probes, or probe derived amplicons, on a solid support or in solution. The captured target nucleic acids are preferably washed and eluted off of the probes. The eluted genomic sequences are more amenable to detailed genetic analysis than a sample that has

IDT Exhibit 1020.5828

not been subjected to the methods described herein. The present invention provides methods and systems for the capture and enrichment of target nucleic acids and analysis of the enriched target nucleic acids. In some embodiments, the present invention provides for the enrichment of targeted sequences in a solution based format. In some embodiments, the present invention provides methods and systems for solution based capture and enrichment of target nucleic acids (e.g., genomic DNA, RNA, cDNA, mRNA, etc.).

The disclosed methods provide a cost-effective, flexible and efficient approach for reducing the complexity of a genomic sample. Genomic samples are used herein for descriptive purposes, but it is understood that other non-genomic samples, also large, non-genomic samples, could be subjected to the same procedures. The methods and systems described herein provide for enrichment of target sequences in a solution based approach thereby providing an alternative to microarray substrate based methods for use in research and therapeutics associated with disease and disease states such as cancers (Durkin et al., 2008, Proc. Natl. Acad. Sci. 105:246-251; Natrajan et al., 2007, Genes, Chr. And Cancer 46:607-615; Kim et al., 2006, Cell 125:1269-1281; Stallings et al., 2006 Can. Res. 66:3673-3680), genetic disorders (Balciuniene et al., Am. J. Hum. Genet. In press), mental diseases (Walsh et al., 2008, Science 320:539-543; Roohi et al., 2008, J. Med. Genet. Epub 18 March 2008; Sharp et al., 2008, Nat. Genet. 40:322-328; Kumar et al., 2008, Hum. Mol. Genet. 17:628-638;) and evolutionary and basic research (Lee et al., 2008, Hum. Mol. Gen. 17:1127-1136; Jones et al., 2007, BMC Genomics 8:402; Egan et al., 2007, Nat. Genet. 39:1384-1389; Levy et al., 2007, PLoS Biol. 5:e254; Ballif et al., 2007, Nat. Genet. 39 :1071-1073 ; Scherer et al., 2007, Nat. Genet. S7-S15; Feuk et al., 2006, Nat. Rev. Genet. 7:85-97), to name a few.

In one embodiment, the present invention provides methods of reducing the genetic complexity of a population of nucleic acid molecules, the method comprising the steps of exposing fragmented, denatured nucleic acid molecules of said population to multiple, different oligonucleotide probes under hybridizing conditions followed by binding the complexes of hybridized molecules to a solid support to capture nucleic acid molecules that specifically hybridize to said probes, wherein said fragmented, denatured nucleic acid molecules have an average size of about 100 to about 1000 nucleotide residues, preferably about 250 to about 800 nucleotide residues and most preferably about 400 to about 600 nucleotide residues, separating unbound and non-specifically hybridized nucleic acids from

IDT Exhibit 1020.5829

the captured molecules, eluting the captured molecules, and optionally repeating the aforementioned processes for at least one further cycle with the eluted captured molecules.

In another embodiment of the present invention, solution based capture methods comprise probe derived amplicons wherein said probes for amplification are affixed to a solid support. The solid support comprises support-immobilized nucleic acid probes to capture specific nucleic acid sequences (e.g., target nucleic acids) from, for example, a genomic sample. Probe amplification provides probe amplicons in solution which are hybridized to target sequences. Following hybridization of probe amplicons to target sequences, target nucleic acid sequences present in the sample are enriched by capturing (e.g., via linker chemistry such as biotin, digoxigenin, etc.) and washing the probes and eluting the hybridized target nucleic acids from the captured probes (Figure 6). The target nucleic acid sequence(s) may be further amplified using, for example, non-specific ligation-mediated PCR (LM-PCR), resulting in an amplified pool of PCR products of reduced complexity compared to the original target sample.

In some embodiments, hybridization between the probes and target nucleic acids is performed under preferably stringent conditions sufficient to support hybridization between the solution based probe amplicons, wherein said probes comprise linker chemistry and complementary regions of the target nucleic acid sample to provide probe/target hybridization complexes. The complexes are subsequently captured via the linker chemistry and washed under conditions sufficient to remove non-specifically bound nucleic acids and the hybridized target nucleic acid sequences are eluted from the captured probe/target complexes.

In some embodiments, the multiple, different oligonucleotide probes comprise a chemical group or linker chemistry, for example a binding moiety such as biotin, digoxigenin, etc., which is able to bind to a solid support. The solid support for binding comprises the corresponding capture chemistry, for example steptavidin for biotin and anti-digoxigenin antibody for digoxigenin. A skilled artisan will recognize that the present invention is not limited by the linker chemistry used and alternative linker chemistries are equally amenable to methods and systems of the present invention.

IDT Exhibit 1020.5830

In embodiments of the present invention, the population or plurality of target nucleic acid molecules preferably contains the whole genome or at least one chromosome of an organism or at least one nucleic acid molecule with at least about 100 kb. In particular, the size(s) of the nucleic acid molecule(s) is/are at least about 200 kb, at least about 500 kb, at least about 1 Mb, at least about 2 Mb or at least about 5 Mb, especially a size between about 100 kb and about 5 Mb, between about 200 kb and about 5 Mb, between about 500 kb and about 5 Mb, between about 1 Mb and about 2 Mb or between about 2 Mb and about 5Mb.

In some embodiments, the target nucleic acid molecules are selected from an animal, a plant or a microorganism, in preferred embodiments the organism is a human. If only limited samples of nucleic acids (e.g. of the human genome) are available, the nucleic acids may be amplified, for example by whole genome amplification, prior to practicing the methods of the present invention. Prior amplification may be necessary for performing the inventive method(s), for example, for forensic purposes (e.g. in forensic medicine for genetic identity purposes).

In some embodiments, the population or plurality of target nucleic acid molecules is a population of genomic DNA molecules. The probes may be selected from a plurality of probes or sequences that, for example, define a plurality of exons, introns or regulatory sequences from a plurality of genetic loci, a plurality of probes that define the complete sequence of at least one single genetic locus, said locus having a size of at least 100 kb, preferably at least 1 Mb, or at least one of the sizes as specified above, a plurality of probes that define single nucleotide polymorphisms (SNPs), or a plurality of probes that define an array, for example a tiling array designed to capture the complete sequence of at least one complete chromosome.

In some embodiments, the solid support is either a nucleic acid microarray or a population of beads.

In some embodiments, the present invention comprises the step of ligating adaptor molecules to one or both, preferably both ends of the nucleic acid molecules prior to or after exposing fragmented nucleic samples to the probes for hybridization.

IDT Exhibit 1020.5831

In some embodiments, methods of the present invention further comprise the amplifying of the target nucleic acid molecules with at least one primer, said primer comprising a sequence which specifically hybridizes to the sequence of said adaptor molecule(s).

In some embodiments, the amplified target nucleic acid sequences may be sequenced, hybridized to a resequencing or SNP-calling array and the sequence or genotypes may be further analyzed.

In some embodiments, the present invention provides an enrichment method for target nucleic acid sequences in a genomic sample, such as exons or variants, preferably SNP sites. This can be accomplished by programming genomic probes specific for a region of the genome to be synthesized on a microarray or by synthesizing genomic probes specific for a region of the genome to capture complementary target nucleic acid sequences contained in a complex genomic sample.

In some embodiments, the present invention is directed to a method for determining nucleic acid sequence information of about at least one region of nucleic acid(s), in particular genomic nucleic acid(s), e.g. the whole genome or at least one chromosome, e.g. with a size as specified above, specifically in a sample, the method comprising the steps of performing the method(s) as described above and determining the nucleic acid sequence of the captured (and eluted) molecules, in particular by means of performing sequencing by synthesis reactions.

In some embodiments, the present invention is directed to a method for detecting coding region variation relative to a reference genome, in particular relative to a reference genome that comprises fragmented, denatured genomic nucleic acid molecules, the method as previously described further comprising determining the nucleic acid sequence of the captured (and eluted) target molecules, in particular by means of performing sequencing by synthesis reactions and comparing the determined sequence to a sequence in a database, in particular to a sequence in a database of polymorphisms in the reference genome to identify variants from the reference genome.

In some embodiments, the present invention is directed to a kit comprising a solid support and reagents for performing a method according to the present invention. Such a kit may

IDT Exhibit 1020.5832

comprise a double stranded adaptor molecule, and a solid support with multiple, different oligonucleotide probes, wherein the probes are selected from a plurality of probes that define a plurality of exons, introns or regulatory sequences from a plurality of genetic loci, a plurality of probes that define the complete sequence of at least one single genetic locus, said locus having a size of at least 100 kb, preferably at least 1 Mb, or at least one of the sizes as specified above, a plurality of probes that define sites known to contain SNPs, or a plurality of probes that define a tiling array designed to capture the complete sequence of at least one complete chromosome. Preferably, the kit comprises two different double stranded adaptor molecules (A and B). The solid support is again either a plurality of beads or a microarray. The kit may further comprise at least one or more other components selected from DNA polymerase, T4 polynucleotide kinase, T4 DNA ligase, an array hybridization solution, an array wash solution, and/or an array elution solution.

In other embodiments, the present invention is directed to a kit comprising compositions and reagents for performing a method according to the present invention. Such a kit may comprise, but is not limited to, a double stranded adaptor molecule, multiple, different oligonucleotide probes, a solid support for capturing said probes, wherein the probes are selected from a plurality of sequences that define a plurality of exons, introns or regulatory sequences from a plurality of genetic loci, a plurality of probes that define the complete sequence of at least one single genetic locus, said locus having a size of at least 100 kb, preferably at least 1 Mb, or at least one of the sizes as specified above, a plurality of probes that define sites known to contain SNPs, or a plurality of probes that define a tiling array designed to capture the complete sequence of at least one complete chromosome. In some embodiments, a kit comprises a plurality of beads or a microarray substrate (e.g., slide, chip, etc.). In some embodiments, a kit comprises two different double stranded adaptor molecules. A kit may further comprise at least one or more other components selected from DNA polymerase, T4 polynucleotide kinase, T4 DNA ligase, hybridization solution(s), wash solution(s), and/or elution solution(s).

In embodiments of the present invention, nucleic acid (pre-selected) capture probes are immobilized onto a solid support (e.g., slide, chip, bead, etc.) using any number of recognized methods (e.g., spotting, photolithography, *in situ* synthesis, etc.). In preferred embodiments, the probes are synthesized *in situ* by maskless array synthesis on a substrate and subsequently amplified by, for example, PCR resulting in probe derived amplicons in

IDT Exhibit 1020.5833

solution. In some embodiments, the probe sequences as synthesized comprise primer binding sites for amplification at one or both the 3' and 5' termini (e.g., at or near the ends) of the probes. In some embodiments, the sequence of the primer binding sites on the probes are the same at both the 3' and 5' prime ends or the probes, whereas in other embodiments the sequence of the primer binding sites is different at the 3' prime end than the sequence at the 5' prime end. In some embodiments, amplification primers for probe amplification further comprise a restriction endonuclease site, for example an MlyI site for easy removal of primer sequences from the final captured target, wherein one of the primers (e.g., forward or reverse primer) further comprises linker chemistry such as a binding moiety or sequence (e.g., biotin, digoxigenin, HIS tag, etc.) and are deposited onto the support with the immobilized probes along with reagents necessary for exponential PCR amplification (e.g., PCR procedures for exponential amplification of targets as known to a skilled artisan). PCR is performed thereby creating amplicons of probe capture sequences such that one of the strands comprises linker chemistry, such as a binding moiety or sequence. The amplicon containing solution is transferred to a vessel (e.g., tube, well of a 96 well plate, etc.) and, in some embodiments, purified from reaction components. An additional round of amplification is preferentially performed on the probe derived amplicons using asymmetric PCR, wherein the linker chemistry labeled primer is in abundance compared to the non-labeled primer to preferentially synthesize single stranded binding moiety/sequence labeled amplicons. The amplicons are purified away from reaction components and transferred to a vessel, denatured nucleic acid sample is added, and hybridization is allowed to occur.

Following hybridization, labeled amplicon/target nucleic acid complexes are captured. For example, when biotin is the binding moiety a streptavidin (SA) coated substrate such as SA coated beads (e.g., paramagnetic beads/particles) are used to capture the biotin labeled amplicon/target complex. The SA bound complex is washed and the hybridized target nucleic acids are eluted from the complex and utilized in downstream applications, such as sequencing applications.

In some embodiments, the present invention provides methods for isolating and reducing the complexity of a plurality of nucleic acid sequences comprising providing a solid support wherein said solid support comprises hybridization probes hybridizable to target nucleic acid sequences and providing a fragmented nucleic acid sample comprising target

IDT Exhibit 1020.5834

nucleic acid sequences, amplifying the hybridization probes wherein the amplification products comprise a binding moiety and wherein the amplification products are in solution, hybridizing the nucleic acid sample to the amplification products in solution under conditions such that hybridization between the amplification products and target nucleic acid sequences is allowed to occur, separating the hybridized target nucleic acid sequences/amplification product complexes from non-specifically hybridized nucleic acids by said binding moiety, and eluting the hybridized target nucleic acid sequences from the complex thereby isolation and reducing the complexity of a plurality of nucleic acid sequences. In some embodiments, the eluted target nucleic acid sequences are sequenced. In some embodiments, the solid support is a microarray slide. In some embodiments, the target nucleic acid sample is fragmented genomic DNA with or without adaptor molecules at one or both ends of the fragments. In some embodiments, the hybridization probes comprise a restriction endonuclease site, for example a MlyI site. In some embodiments, probe amplification comprises exponential polymerase chain reaction, and may further comprise asymmetric non-exponential amplification. In some embodiments, the binding moiety is biotin and the capture substrate, such as a bead for example a paramagnetic particle, is coated with streptavidin for separation of the target nucleic acid/ amplification product complex from non-specifically hybridized target nucleic acids. In some embodiments, the captured target nucleic acid/ amplification product complexes are washed prior to elution of the bound target nucleic acids.

In some embodiments, the present invention provides a kit comprising hybridization probe sequences comprising a binding moiety and a restriction enzyme site wherein said probe sequences are designed to hybridize to one or more target nucleic acid sequences and wherein said probe sequences are in solution, a substrate comprising a binding partner for binding said binding moiety, and instruction for performing methods of the present invention. In some embodiments, a kit further comprises one or more solutions such as hybridization, washing, and elution solution(s). In some embodiments, a kit comprises a magnet. In some embodiments, a kit comprises one or more enzymes and corresponding reagents, buffers, and the like, for example a restriction enzyme such as MlyI and buffers/reagents for performing restriction enzyme reactions using MlyI.

Other objects, advantages and features of the present invention will become apparent from the following specification taken in conjunction with the accompanying drawings.

IDT Exhibit 1020.5835

FIGURES

FIG. 1 is a general graphic depiction of flow diagram of a direct genomic selection process using a microarray.

FIG. 2 is another graphic depiction of flow diagram of a direct genomic selection process using a microarray.

FIG 3 (a-b) show the results of a direct genomic selection process using a microarray according to Example 2. (**a**) Sequence read map detail of ~190Kb of chromosome 16 from three microarray genomic selection replicates, indicating the reproducibility of targeted sequencing. Genomic DNA was from a Burkett lymphoma cell line was purified and fragmented. Tumor Sequencing Program exons (6726 genomic regions of 500bp in size), were captured using a NimbleGen oligonucleotide microarray and sequenced using a 454 sequencer. (1) Chromosome position, (2,3,4) read map of the highest BLAST score for 454 reads from three independent microarray selection and sequencing_experiments (5) regions targeted by microarray probes. (**b**) Sequence read map detail of ~2,000 bases of a chromosome 17 from a microarray selection of a 2Mb contiguous region that contains the BRCA1 gene. (1) Chromosome position, (2) microarray selection probes. Probes are spaced every 10pb and staggered along the y-axis. (3) Per-base fold sequence coverage. Coverage is from 0 to 100 fold. (4) Read map of the highest blast scores for 454 sequencing reads.

FIG. 4 (a-c) show the results of synthesizing probes on a microarray, releasing the probes from the microarray and immobilizing the probes on a support for use in a method for capturing target polynucleotides of interest. (a) Coverage depth comparison for 'Exonic' and 'Locus' selection and sequencing as disclosed in Example 2. Plot shows the fraction of bases of each aggregate target region and the corresponding cumulative depth of sequence coverage after one 454 FLX run. 'Exonic' sample represents 6,726 exon sized regions. The 2Mb BRCA1 region was targeted from positions 37,490,417 to 39,490,417 on human chromosome 17. Only the unique fraction was targeted by selection probes. (b) Histogram of per base sequence coverage depth for the Exonic exüero,emt as disclosed in Example 2. (c) Histogram of per base coverage depth for 2Mb Locus example according to Example 3.

IDT Exhibit 1020.5836

FIG. 5 illustrates a detail of the read mapping for a locus on chromosome 16 from three genomic samples. Data were generated by targeted sequencing of 6726 exons that were captured in solution. Capture oligonucleotides were cleaved and amplified from a microarray, using the protocol described in Example 4. The data presented represents an example gene map from chromosome 3. (1) chromosome position, (2) map of sequencing reads from one 454-FLX sequencing run, and (3) targeted regions. Analysis of the solution-phase capture data indicates that 83.8% of the reads map back to target regions, indicating similar performance to array-based capture protocols.

FIG. 6 exemplifies an embodiment of the present invention; a generalized flow diagram of an enrichment process wherein said enrichment methods are used to isolate and enrich for a plurality of nucleic acid sequences in an aqueous solution. Hybridization probes, affixed to a microarray substrate, are amplified *in situ* to produce probe derived amplicons in solution the amplicons of which comprise a binding moiety. Fragmented nucleic acids (e.g. labeled with a detection moiety) are hybridized in solution to the labeled probe amplicons, the complexes being subsequently captured (e.g., by paramagnetic capture particles). The captured and immobilized hybridized complexes are washed and the specifically bound targets are eluted from the bound immobilized probe amplicons. The eluted (e.g., isolated and enriched) target sequences are applied to downstream applications, such as sequencing.

FIG. 7 demonstrates resequencing conformance using the solution capture methods of the present invention. The resequencing assay is comprised of a subset of the targeted captured regions. The x-axis represents an arbitrary set of regions from the larger target region, which serves as a representation of the targeted capture region as a whole. The y-axis represents the percent sequence conformance with known target sequences.

DEFINITIONS

As used herein, the term "sample" is used in its broadest sense. In one sense, it is meant to include a specimen or culture obtained from any source, preferentially a biological source. Biological samples may be obtained from animals (including humans) and encompass fluids, solids, tissues, and gases. Biological samples include blood products, such as

IDT Exhibit 1020.5837

plasma, serum and the like. As such, a "sample of nucleic acids" or a "nucleic acid sample", a "target sample" comprises nucleic acids (e.g., DNA, RNA, cDNA, mRNA, tRNA, miRNA, etc.) from any source. In the present application, a nucleic acid sample preferably derives from a biological source, such as a human or non-human cell, tissue, and the like. The term "non-human" refers to all non-human animals and entities including, but are not limited to, vertebrates such as rodents, non-human primates, ovines, bovines, ruminants, lagomorphs, porcines, caprines, equines, canines, felines, aves, etc. Non-human also includes invertebrates and prokaryotic organisms such as bacteria, plants, yeast, viruses, and the like. As such, a nucleic acid sample used in methods and systems of the present invention is a nucleic acid sample derived from any organism, either eukaryotic or prokaryotic.

As used herein, the term "hybridization" is used in reference to the pairing of complementary nucleic acids. Hybridization and the strength of hybridization (e.g., the strength of the association between the nucleic acids) is affected by such factors as the degree of complementary between the nucleic acids, stringency of the conditions involved, the melting temperature (T_m) of the formed hybrid, and the G:C ratio of the nucleic acids. While the invention is not limited to a particular set of hybridization conditions, stringent hybridization conditions are preferably employed. Stringent hybridization conditions are sequence dependent and differ with varying environmental parameters (e.g., salt concentrations, presence of organics, etc.). Generally, "stringent" conditions are selected to be about 50°C to about 20°C lower than the T_m for the specific nucleic acid sequence at a defined ionic strength and pH. Preferably, stringent conditions are about 5°C to 10°C lower than the thermal melting point for a specific nucleic acid bound to a complementary nucleic acid. The T_m is the temperature (under defined ionic strength and pH) at which 50% of a nucleic acid (e.g., target nucleic acid) hybridizes to a perfectly matched probe.

"Stringent conditions" or "high stringency conditions," for example, can be hybridization in 50% formamide, 5x SSC (0.75 M NaCl, 0.075 M sodium citrate), 50 mM sodium phosphate (pH 6.8), 0.1% sodium pyrophosphate, 5x Denhardt's solution, sonicated salmon sperm DNA (50 mg/ml), 0.1% SDS, and 10% dextran sulfate at 42°C, with washes at 42°C in 0.2 % SSC (sodium chloride/sodium citrate) and 50% formamide at 55°C, followed by a wash with 0.lx SSC containing EDTA at 55°C. By way of example, but not limitation, it is contemplated that buffers containing 35% formamide, 5x SSC, and 0.1%

IDT Exhibit 1020.5838

(w/v) sodium dodecyl sulfate (SDS) are suitable for hybridizing under moderately non-stringent conditions at 45°C for 16-72 hours.

Furthermore, it is envisioned that the formamide concentration may be suitably adjusted between a range of 20-45% depending on the probe length and the level of stringency desired. Additional examples of hybridization conditions are provided in several sources, including Molecular Cloning: A Laboratory Manual, Eds. Sambrook et al., Cold Spring Harbour Press (incorporated herein by reference in its entirety).

Similarly, "stringent" wash conditions are ordinarily determined empirically for hybridization of a target to a probe, or in the present invention, a probe derived amplicon. The amplicon/target are hybridized (for example, under stringent hybridization conditions) and then washed with buffers containing successively lower concentrations of salts, or higher concentrations of detergents, or at increasing temperatures until the signal-to-noise ratio for specific to non-specific hybridization is high enough to facilitate detection of specific hybridization. Stringent temperature conditions will usually include temperatures in excess of about 30°C, more usually in excess of about 37°C, and occasionally in excess of about 45°C. Stringent salt conditions will ordinarily be less than about 1000 mM, usually less than about 500 mM, more usually less than about 150 mM (Wetmur et al., 1966, J. Mol. Biol., 31:349-370; Wetmur, 1991, Critical Reviews in Biochemistry and Molecular Biology, 26:227-259, incorporated by reference herein in their entireties).

As used herein, the term "primer" refers to an oligonucleotide, whether occurring naturally as in a purified restriction digest or produced synthetically, that is capable of acting as a point of initiation of synthesis when placed under conditions in which synthesis of a primer extension product that is complementary to a nucleic acid strand is induced, (e.g., in the presence of nucleotides and an inducing agent such as DNA polymerase and at a suitable temperature and pH). The primer is preferably single stranded for maximum efficiency in amplification. Preferably, the primer is an oligodeoxyribonucleotide. The primer must be sufficiently long to prime the synthesis of extension products in the presence of the inducing agent. The exact lengths of the primers will depend on many factors, including temperature, source of primer and the use of the method.

IDT Exhibit 1020.5839

As used herein, the term "probe" refers to an oligonucleotide (e.g., a sequence of nucleotides), whether occurring naturally as in a purified restriction digest or produced synthetically, recombinantly or by PCR amplification, that is capable of hybridizing to at least a portion of another oligonucleotide of interest, for example target nucleic acid sequences. A probe may be single-stranded or double-stranded. Probes are useful in the detection, identification and isolation of particular gene sequences.

As used herein, the term "target nucleic acid molecules" and "target nucleic acid sequences" are used interchangeably and refer to molecules or sequences from a target genomic region to be studied. The pre-selected probes determine the range of targeted nucleic acid molecules. Thus, the "target" is sought to be sorted out from other nucleic acid sequences. A "segment" is defined as a region of nucleic acid within the target sequence, as is a "fragment" or a "portion" of a nucleic acid sequence.

As used herein, the term "isolate" when used in relation to a nucleic acid, as in "isolating a nucleic acid" refers to a nucleic acid sequence that is identified and separated from at least one component or contaminant with which it is ordinarily associated in its natural source. Isolated nucleic acid is in a form or setting that is different from that in which it is found in nature. In contrast, non-isolated nucleic acids as nucleic acids such as DNA and RNA found in the state they exist in nature. The isolated nucleic acid, oligonucleotide, or polynucleotide may be present in single-stranded or double-stranded form.

DETAILED DESCRIPTION OF THE INVENTION

The present invention broadly relates to cost-effective, flexible and rapid methods for reducing nucleic acid sample complexity to enrich for target nucleic acids of interest and to facilitate further processing and analysis, such as sequencing, resequencing and SNP calling. The captured target nucleic acid sequences, which are of a more defined less complex genomic population are more amenable to detailed genetic analysis, for example genetic analysis of disease and disease states (e.g., cancers, genetic mutations, genetic diseases, etc). The present invention provides methods and systems useful, for example, in searching for genetic variants and mutations such a single nucleotide polymorphisms (SNPs), sets of SNPs, genomic insertions, deletions, etc. that underlie human diseases.

IDT Exhibit 1020.5840

Thus, the invention provides for method for enrichment of target nucleic acid in a complex nucleic acid sample.

In one embodiment, the present invention relates to a method of reducing the genetic complexity of a population of nucleic acid molecules, the method comprising the steps of either exposing fragmented, denatured nucleic acid molecules of said population to multiple, different oligonucleotide probes that are bound on a solid support under hybridizing conditions to capture nucleic acid molecules that specifically hybridize to said probes or exposing fragmented, denatured nucleic acid molecules of said population to multiple, different oligonucleotide probes under hybridizing conditions followed by binding the complexes of hybridized molecules on a solid support to capture nucleic acid molecules that specifically hybridize to said probes, wherein said fragmented, denatured nucleic acid molecules have an average size of about 100 to about 1000 nucleotide residues, preferably about 250 to about 800 nucleotide residues and most preferably about 400 to about 600 nucleotide residues, separating unbound and non-specifically hybridized nucleic acids from the captured molecules; eluting the captured molecules from the solid support, and optionally repeating steps (a) to (c) for at least one further cycle with the eluted captured molecules.

In a further embodiment, the present invention is directed to a kit comprising double stranded adaptor molecules, and a solid support with multiple, different oligonucleotid probes, wherein said probes are selected from: a plurality of probes that define a plurality of exons, introns or regulatory sequences from a plurality of genetic loci, a plurality of probes that define the complete sequence of at least one single genetic locus, said locus having a size of at least 100 kb, preferably at least 1 Mb, or at least one of the sizes as specified above, a plurality of probes that define sites known to contain SNPs, or a plurality of probes that define a tiling array designed to capture the complete sequence of at least one complete chromosome.

In one embodiment, a sample containing denatured (i.e., single-stranded) nucleic acid molecules, preferably genomic nucleic acid molecules, which can be fragmented molecules, is exposed under hybridizing conditions to a plurality of oligonucleotide probes, which are immobilized on a solid support prior or after hybridization with a plurality of oligonucleotide probes to capture from the sample target nucleic acid

IDT Exhibit 1020.5841

molecules that hybridize to the immobilized probes. Non-hybridizing regions of the genome or any other sample nucleic acid remain in solution.

The nucleic acids are typically deoxyribonucleic acids or ribonucleic acids, and include products synthesized *in vitro* by converting one nucleic acid molecule type (e.g. DNA, RNA and cDNA) to another as well as synthetic molecules containing nucleotide analogues, such as PNAs. Denatured genomic DNA molecules are in particular genome-derived molecules that are shorter than naturally occurring genomic nucleic acid molecules. The skilled person can produce molecules of random- or non-random size from larger molecules by chemical, physical or enzymatic fragmentation or cleavage using well known protocols. Chemical fragmentation can employ ferrous metals (e.g., Fe-EDTA). Physical methods can include sonication, hydrodynamic force or nebulization (see European patent application EP 0 552 290). Enzymatic protocols can employ nucleases such as micrococcal nuclease (Mnase) or exo-nucleases (such as ExoI or Bal31) or restriction endonucleases. The protocol by which fragments are generated should not affect the use of the fragments in the methods. It can be advantageous during enrichment to employ fragments in a size range compatible with the post-enrichment technology in which the enriched fragments will be used. A suitable fragment size can be in the range of between about 100 and about 1000 nucleotide residues or base pairs, or between about 250 and about 800 ~~1000~~ nucleotide residues or base pairs, and can be about 400 to about 600 nucleotide residues or base pairs, in particular about 500 nucleotide residues or base pairs.

The probes correspond in sequence to at least one region of the genome and can be provided on a solid support in parallel using maskless array synthesis technology. Alternatively, probes can be obtained serially using a standard DNA synthesizer and then applied to the solid support or can be obtained from an organism and then immobilized on the solid support. After the hybridization, nucleic acids that do not hybridize, or that hybridize non-specifically to the probes are separated from the support-bound probes by washing. The remaining nucleic acids, bound specifically to the probes, are eluted from the solid support in e.g. heated water or in a nucleic acid elution buffer containing e.g. TRIS puffer and/or EDTA to yield an eluate enriched for the target nucleic acid molecules.

In some embodiments, double-stranded linkers are provided at least at one of the termini of the (genomic) nucleic acid molecules before the fragments are denatured and hybridized to

IDT Exhibit 1020.5842

the immobilized probes. In such embodiments, target nucleic acid molecules can be amplified after elution to produce a pool of amplified products having reduced complexity relative to the original sample. The target nucleic acid molecules can be amplified using for example, non-specific LM-PCR through multiple rounds of thermal cycling. Optionally, the amplified products can be further enriched by a second selection against the probes. The products of the second selection can be amplified again prior to use as described. This approach is summarized graphically in Fig. 1 and in a flow chart in Fig. 2. The linkers can be provided in an arbitrary size and with an arbitrary nucleic acid sequence according to what is required for downstream analytical applications subsequent to the complexity reduction step. The linkers can range between about 12 and about 100 base pairs, including a range between about 18 and 100 base pairs, and preferably between about 20 and 24 base pairs.

Alternatively, nucleic acid probes for target molecules can be synthesized on a solid support, released from the solid support as a pool of probes and amplified as described. The amplified pool of released probes can be covalently- or non-covalently immobilized onto a support, such as glass, metal, ceramic or polymeric beads or other solid support. The probes can be designed for convenient release from the solid support by providing, e.g., at or near the support-proximal probe termini an acid- or alkali-labile nucleic acid sequence that releases the probes under conditions of low or high pH, respectively. Various cleavable linker chemistries are known in the art. The support can be provided, e.g., in a column having fluid inlet and outlet. The art is familiar with methods for immobilizing nucleic acids onto supports, for example by incorporating a biotinylated nucleotide into the probes and coating the support with streptavidin such that the coated support non-covalently attracts and immobilizes the probes in the pool. The sample or samples are passed across the probe-containing support under hybridizing conditions such that target nucleic molecules that hybridize to the immobilized support can be eluted for subsequent analysis or other use.

In one aspect, the invention enables capturing and enriching for target nucleic acid molecules or target genomic region(s) from a complex biological sample by direct genomic selection. The invention is also useful in searching for genetic variants and mutations, such as single nucleotide polymorphisms (SNP), or set of SNPs, that underlie human diseases. It is contemplated that capture and enrichment using microarray hybridization

IDT Exhibit 1020.5843

technology is much more flexible than other methods currently available in the field of genomic enrichment, such as use of BAC (bacterial artificial chromosome) for direct genomic selection (see Lovett et al., 1991).

The invention enables targeted array-based-, shotgun-, capillary-, or other sequencing methods known to the art. In general, strategies for shotgun sequencing of randomly generated fragments are cost-effective and readily integrated into a pipeline, but the invention enhances the efficiency of the shotgun approach by presenting only fragments from one or more genomic regions of interest for sequencing. The invention provides an ability to focus the sequencing strategies on specific genomic regions, such as individual chromosomes or exons for medical sequencing purposes.

Target nucleic acid molecules can be enriched from one or more samples that include nucleic acids from any source, in purified or unpurified form. The source need not contain a complete complement of genomic nucleic acid molecules from an organism. The sample, preferably from a biological source, includes, but is not limited to pooled isolates from individual patients, tissue samples, or cell culture. As used herein, the term "target nucleic acid molecules" refers to molecules from a target genomic region to be studied. The pre-selected probes determine the range of targeted nucleic acid molecules. The skilled person in possession of this disclosure will appreciate the complete range of possible targets and associated targets.

The target region can be one or more continuous blocks of several megabases (Mb), or several smaller contiguous or discontiguous regions such as all of the exons from one or more chromosomes, or sites known to contain SNPs. For example, the solid support can support a tiling array designed to capture one or more complete chromosomes, parts of one or more chromosomes, all exons, all exons from one or more chromosomes, selected exons, introns and exons for one or more genes, gene regulatory regions, and so on. Alternatively, to increase the likelihood that desired non-unique or difficult-to-capture targets are enriched, the probes can be directed to sequences associated with (e.g., on the same fragment as, but separate from) the actual target sequence, in which case genomic fragments containing both the desired target and associated sequences will be captured and enriched. The associated sequences can be adjacent or spaced apart from the target sequences, but the skilled person will appreciate that the closer the two portions are to one

IDT Exhibit 1020.5844

another, the more likely it will be that genomic fragments will contain both portions. Still further, to further reduce the limited impact of cross-hybridization by off-target molecules, thereby enhancing the integrity of the enrichment, sequential rounds of capture using distinct but related capture probe sets directed to the target region can be performed. Related probes are probes corresponding to regions in close proximity to one another in the genome that can, therefore, hybridize to the same genomic DNA fragment.

Microarray oligonucleotides are designed to target the target region or regions of the genome. The length of individual probes is typically between 50 and 200 bases. These probes may be either designed to be overlapping probes, meaning that the starting nucleotides of adjacent probes are less than the length of a probe, or non-overlapping probes, where the distance between adjacent probes are greater than the length of a probe. The distance between adjacent probes is generally overlapping, with spacing between the starting nucleotide of two probes varying between 1 and 100 bases. This distance can be varied to cause some genomic regions to be targeted by a larger number of probes than others. This variation can be used to modulate the capture efficiency of individual genomic regions, normalizing capture. Probes can be tested for uniqueness in the genome. To avoid non-specific binding of genomic elements to capture arrays, highly repetitive elements of the genome should be excluded from selection microarray designs using a new method that utilizes a strategy similar to the WindowMasker program developed by Morgolis (2006) to identify these regions and exclude them from probe selection. The process compared the set of probes against a pre-computed frequency histogram of all possible 15-mer probes in the human genome. For each probe, the frequencies of the 15-mers comprising the probe are then used to calculate the average 15-mer frequency of the probe. The higher the average 15-mer frequency, the more likely the probe is to lie within a repetitive region of the genome. Only probes with an average 15-mer frequency less than 100 should be used.

The nature and performance of the probes can be varied to advantageously normalize or adjust the distribution of the target molecules captured and enriched in accord with the methods. A goal of such normalization is to deliver one expressed gene per read (see Soares, et al., 1994) Normalization can be applied, for example, to populations of cDNA molecules before library construction, because the distribution of molecules in the population reflects the different expression levels of expressed genes from which the cDNA molecule populations are produced. For example, the number of sequencing

IDT Exhibit 1020.5845

reactions required to effectively analyze each target region can be reduced by normalizing the number of copies of each target sequence in the enriched population such that across the set of probes the capture performance of distinct probes are normalized, on the basis of a combination of fitness and other probe attributes. Fitness, characterized by a "capture metric," can be ascertained either informatically or empirically. In one approach, the ability of the target molecules to bind can be adjusted by providing so-called isothermal (Tm-balanced) oligonucleotide probes, as are described in U.S. Published Patent Application No. US-2005/0282209 (NimbleGen Systems, Madison, WI), that enable uniform probe performance, eliminate hybridization artifacts and/or bias and provide higher quality output. Probe lengths are adjusted (typically, about 20 to about 100 nucleotides, preferably about 40 to about 85 nucleotides, in particular about 45 to about 75 nucleotides, e.g. 45 nucleotides but optionally also more than 100 nucleotides until about 250 nucleotides) to equalize the melting temperature (e.g. Tm = 76°C, typically about 55°C to about 76°C, in particular about 72°C to about 76°C) across the entire set. Thus, probes are optimized to perform equivalently at a given stringency in the genomic regions of interest, including AT- and GC-rich regions. Relatedly, the sequence of individual probes can be adjusted, using natural bases or synthetic base analogs such as inositol, or a combination thereof to achieve a desired capture fitness of those probes. Similarly, locked nucleic acid probes, peptide nucleic acid probes or the like having structures that yield desired capture performance can be employed. The skilled artisan in possession of this disclosure will appreciate that probe length, melting temperature and sequence can be coordinately adjusted for any given probe to arrive at a desired capture performance for the probe. Conveniently, the melting temperature (Tm) of the probe can be calculated using the formula: Tm=5x(Gn+Cn)+1x(An+Tn), where n is the number of each specific base (A, T, G or C) present on the probe.

Capture performance can also be normalized by ascertaining the capture fitness of probes in the probe set, and then adjusting the quantity of individual probes on the solid support accordingly. For example, if a first probe captures twenty times as much nucleic acid as a second probe, then the capture performance of both probes can be equalized by providing twenty times as many copies of the second probe, for example by increasing by twenty-fold the number of features displaying the second probe. If the probes are prepared serially and applied to the solid support, the concentration of individual probes in the pool can be varied in the same way.

IDT Exhibit 1020.5846

Still further, another strategy for normalizing capture of target nucleic acids is to subject the eluted target molecules to a second round of hybridization against the probes under less stringent conditions than were used for the first hybridization round. Apart from the substantial enrichment in the first hybridization that reduces complexity relative to the original genomic nucleic acid, the second hybridization can be conducted under hybridization conditions that saturate all capture probes. Presuming that substantially equal amounts of the capture probes are provided on the solid support, saturation of the probes will ensure that substantially equal amounts of each target are eluted after the second hybridization and washing.

Another normalizing strategy follows the elution and amplification of captured target molecules from the solid support. Target molecules in the eluate are denatured using, for example, a chemical or thermal denaturing process, to a single-stranded state and are re-annealed. Kinetic considerations dictate that abundant species re-anneal before less abundant species. As such, by removing the initial fraction of re-annealed species, the remaining single-stranded species will be balanced relative to the initial population in the eluate. The timing required for optimal removal of abundant species is determined empirically.

Summarizing, an embodiment of the present invention provides a new method of reducing the genetic complexity of a population of nucleic acid molecules. This method comprises

(a) either exposing fragmented, denatured nucleic acid molecules of said population to multiple, different oligonucleotide probes that are bound on a solid support under hybridizing conditions to capture nucleic acid molecules that specifically hybridize to said probes,
or exposing fragmented, denatured nucleic acid molecules of said population to multiple, different oligonucleotide probes under hybridizing conditions followed by binding the complexes of hybridized molecules on a solid support to capture nucleic acid molecules that specifically hybridize to said probes,
wherein (in both cases) said fragmented, denatured nucleic acid molecules have an average size of about 100 to about 1000 nucleotide residues, preferably about 250 to about 800 nucleotide residues and most preferably about 400 to about 600 nucleotide residues,

IDT Exhibit 1020.5847

(b) separating unbound and non-specifically hybridized nucleic acids from the captured molecules;

(c) eluting the captured molecules from the solid support, preferably in an eluate pool having reduced genetic complexity relative to the original sample, and

(d) optionally repeating steps (a) to (c) for at least one further cycle with the eluted captured molecules.

In most cases, the population of nucleic molecules are molecules originated from a sample of genomic DNA (genomic nucleic acid molecules). However, it is also possible to start with a sample of cDNA or even RNA. Fragmentation can in principle be done by any method which is known in the art as already explained above. However, the fragmented denatured nucleic acid molecules should have an average size of about 100 to about 1000 nucleotide residues, preferably about 250 to about 800 nucleotide residues and most preferably about 400 to about 600 nucleotide residues. For example, this can be achieved by means of nebulization of genomic DNA (see e.g. the European patent application EP 0 552 290).

The parameters of genetic complexity reduction can be chosen almost arbitrarily, depending upon the user's desire for sequence selection, and are defined by the sequences of the multiple oligonucleotide probes. In one embodiment, said multiple probes define a plurality of exons, introns or regulatory sequences from a plurality of genetic loci. In another embodiment, said multiple probes define the complete sequence of at least one single genetic locus, said locus having a size of at least 100 kb and preferably at least 1 Mb or a size as specified above. In still another embodiment, said multiple probes define sites known to contain SNPs. In a further embodiment, said multiple probes define a tiling array. Such a tiling array in the context of the present invention is defined as being designed to capture the complete sequence of at least one complete chromosome. In this context, the term "define" is understood in such a way that the population of multiple probes comprises at least one probe for each target sequence that shall become enriched. Preferably, the population of multiple probes additionally comprises at least a second probe for each target sequence that shall become enriched, characterized in that said second probe has a sequence which is complementary to said first sequence.

IDT Exhibit 1020.5848

The solid support according to the present invention is either a nucleic acid microarray or a population of beads. Said beads may be glass, metal, ceramic and polymeric beads. If said solid support is a microarray, it is possible to synthesize the oligonucleotide capture probes in situ directly onto said solid support. For example, the probes may be synthesized on the microarray using a maskless array synthesizer (US 6,375,903). The lengths of the multiple oligonucleotide probes may vary, are dependent on the experimental design and are limited only by the possibility to synthesize such probes. Preferably, the average length of the population of multiple probes is about 20 to about 100 nucleotides, preferably about 40 to about 85 nucleotides, in particular about 45 to about 75 nucleotides, e.g. 45 nucleotides.

If the solid support is a population of beads, the capture probes may be initially synthesized on a microarray using a maskless array synthesizer, then released or cleaved off according to known standard methods, optionally amplified and then immobilized on said population of beads according to methods known in the art. The beads may be packed into a column so that a sample is loaded and passed through the column for reducing genetic complexity. Alternatively, in order to improve the hybridization kinetics, hybridization may take place in an aqueous solution comprising the beads with the immobilized multiple oligonucleotide molecules in suspension.

In one embodiment, the multiple different oligonucleotide probes each carry a chemical group or linker, i.e. a moiety which allows for immobilization onto a solid support, also named an immobilizable group. Then the step of exposing the fragmented, denatured nucleic acid molecules of the sample to the multiple, different oligonucleotide probes under hybridizing conditions is performed in an aqueous solution and immobilization onto an appropriate solid support takes place subsequently. For example, such a moiety may be biotin which can be used for immobilization on a streptavidin coated solid support. In another embodiment, such a moiety may be a hapten like digoxygenin, which can be used for immobilization on a solid support coated with a hapten recognizing antibody, e.g. a digoxygenin binding antibody.

In a specific embodiment, the plurality of immobilized probes is characterized by normalized capture performance. The normalized capture performance is generally achieved by methods as described above, typically comprising the steps of a) ascertaining the capture fitness of probes in the probe set; and b) adjusting the quantity of at least one

IDT Exhibit 1020.5849

probe on the solid support. Alternatively, the normalized capture performance is achieved by a method comprising the steps of a) ascertaining the capture fitness of probes in the probe set; and b) adjusting at least one of the sequence, the melting temperature and the probe length of at least one probe on the solid support. Still alternatively, the normalized capture performance is achieved by a method comprising the steps of a) exposing the captured molecules to the at least one immobilized probe on the solid support under less stringent conditions than in the first exposing step such that the at least one probe is saturated, b) washing unbound and non-specifically bound nucleic acids from the solid support; and c) eluting the bound target nucleic acids from the solid support. Still alternatively, the normalized capture performance is achieved by a method comprising the steps of a) denaturing the eluted captured molecules to a single-stranded state; b) re-annealing the single-stranded molecules until a portion of the molecules are double-stranded; and discarding the double-stranded molecules and c) retaining the single-stranded molecules.

Usually at least one immobilized probe hybridizes to a genomic region of interest on nucleic acid fragments in the sample. Alternatively, the at least one immobilized probe may hybridize to sequences on target nucleic acid fragments comprising a genomic region of interest, the hybridizing sequences being separate from the genomic region of interest. Furthermore, it is also within the scope of the present invention, that at least a second hybridization step using at least one oligonucleotide probe related to but distinct from the at least one probe used in the initial hybridization is performed.

In particular, the present invention is also directed to a method for determining nucleic acid sequence information of at least one region of genomic nucleic acid in a sample, the method comprising the steps of:

- reducing the genetic complexity of a population of nucleic acid molecules according to any method as disclosed herein, and
- determining the nucleic acid sequence of the captured molecules e.g. by means of performing a sequencing reaction. Preferably, such a sequencing reaction is a sequencing by synthesis reaction.

IDT Exhibit 1020.5850

According to this embodiment, the genomic DNA is preferably fragmented by mechanical stress. The desired average size of the DNA fragments shall be small (<= 1000 bp) and depends on the sequencing method to be applied.

Sequencing by synthesis according to the literature in the art (see e.g. Hyman, E. D., 1988) is defined as any sequencing method which monitors the generation of side products upon incorporation of a specific desoxynucleoside-triphosphate during the sequencing reaction (see e.g. Rhonaghi et al., 1998). One particular and most prominent embodiment of the sequencing by synthesis reaction is the pyrophosphate sequencing method. In this case, generation of pyrophosphate during nucleotide incorporation is monitored by means of an enzymatic cascade which finally results in the generation of a chemo-luminescent signal. For example, the 454 Genome Sequencer System (Roche Applied Science cat. No. 04 760 085 001) is based on the pyrophosphate sequencing technology. For sequencing on a 454 GS20 or 454 FLX instrument, the average genomic DNA fragment size should be in the range of 200 or 600 bp, respectively.

Alternatively, the sequencing by synthesis reaction is a terminator dye type sequencing reaction. In this case, the incorporated dNTP building blocks comprise a detectable label, which is preferably a fluorescent label that prevents further extension of the nascent DNA strand. The label is then removed and detected upon incorporation of the dNTP building block into the template/primer extension hybrid for example by means of using a DNA polymerase comprising a 3'-5' exonuclease or proofreading activity.

Advantageously, the inventive method of first reducing genomic complexity and then determining multiple sequences further comprises the step of ligating adaptor molecules to one or both, preferably both ends of the fragmented nucleic acid molecules. Adaptor molecules in the context of the present invention are preferably defined as blunt ended double stranded oligonucleotides. In addition, the inventive method may further comprise the step of amplification of said nucleic acid molecules with at least one primer, said primer comprising a sequence which corresponds to or specifically hybridizes with the sequence of said adaptor molecules.

In order to ligate adaptor molecules onto a double stranded target molecule, it is preferred that this target molecule itself is blunt ended. In order to achieve this, the double stranded

IDT Exhibit 1020.5851

target molecules are subjected to a fill-in reaction with a DNA Polymerase such as T4-DNA polymerase or Klenow polymerase in the presence of desoxynucleoside triphposphates, which results in blunt ended target molecules. In addition, e.g. T4 Polynucleotide kinase is added prior to the ligation in order to add phosphate groups to the 5' terminus for the subsequent ligation step. Subsequent ligation of the adaptors (short double stranded blunt end DNA oligonucleotides with about 3-20 base pairs) onto the polished target DNA may be performed according to any method which is known in the art, preferably by means of a T4-DNA ligase reaction.

Said ligation may be performed prior to or after the step of exposing a sample that comprises fragmented, denatured genomic nucleic acid molecules to multiple oligonucleotide probes under hybridizing conditions to capture target nucleic acid molecules that hybridize to said probes. In case ligation is performed subsequently, the enriched nucleic acids which are released from the solid support in single stranded form should be re-annealed first followed by a primer extension reaction and a fill-in reaction according to standard methods known in the art.

Ligation of said adaptor molecules allows for a step of subsequent amplification of the captured molecules. Independent from whether ligation takes place prior to or after the capturing step, there exist two alternative embodiments. In the first embodiment, one type of adaptor molecules is used. This results in population of fragments with identical terminal sequences at both ends of the fragment. As a consequence, it is sufficient to use only one primer in a potential subsequent amplification step. In an alternative embodiment, two types of adaptor molecules A and B are used. This results in a population of enriched molecules composed of three different types: (i) fragments having one adaptor (A) at one end and another adaptor (B) at the other end, (ii) fragments having adaptors A at both ends, and (iii) fragments having adaptors B at both ends.

Generation of enriched molecules according to type (i) is of outstanding advantage, if amplification and sequencing is e.g. performed with the 454 life science corporation GS20 and GSFLX instrument (see GS20 Library Prep Manual, Dec 2006, WO 2004/070007). If one of said adaptors, e.g. adaptor B carries a biotin modification, then molecules (i) and (iii) can e.g. be bound on streptavidin (SA) coated magnetic particles for further isolation and the products of (ii) washed away. In case the enriched and SA-immobilized DNA is

IDT Exhibit 1020.5852

single stranded following elution from the capture array/solid support, it is advantageous to make the DNA double-stranded. In this case primers complementary to adaptor A may be added to the washed SA pull down products. Since moieties that are B-B (iii above) do not have A or its complement available, only A-B adapted and SA captured products will be made double stranded following primer-extension from an A complement primer. Subsequently, the double stranded DNA molecules that have been bound to said magnetic particles are thermally or chemically (e.g. NaOH) denatured in such a way that the newly synthesized strand is released into solution. Due to the tight biotin/streptavidin bonding, for example, molecules with only two adaptors B will not be released into solution. The only strand available for release is the A-complement to B-complement primer-extension synthesized strand. Said solution comprising single stranded target molecules with an adaptor A at one end and an adaptor B at the other end can e.g. subsequently be bound on a further type of beads comprising a capture sequence which is sufficiently complementary to the adaptor A or B sequences for further processing.

In case of the Genome Sequencer workflow (Roche Applied Science Catalog No. 04 896 548 001), in a first step, (clonal) amplification is performed by means of emulsion PCR. Thus, it is also within the scope of the present invention, that the step of amplification is performed in the form of an emulsion PCR. The beads carrying the clonally amplified target nucleic acids may then become arbitrarily transferred into a picotiter plate according to the manufacturer's protocol and subjected to a pyrophosphate sequencing reaction for sequence determination.

Thus, the methods according to the present invention enable sequence determinations for a variety of different applications. For example, the present invention also provides a method for detecting coding region variation relative to a reference genome, preferably in a sample that comprises fragmented, denatured genomic nucleic acid molecules, the method comprising the steps of:

- performing the method(s) as described above,
- determining nucleic acid sequence of the captured molecules, and
- comparing the determined sequence to a database, in particular to a database of polymorphisms in the reference genome to identify variants from the reference genome.

IDT Exhibit 1020.5853

In a further major aspect, the present invention also provides a kit for performing a method or part of a method according to the present invention as disclosed herein. Thus, the present invention is also directed to a kit comprising

- a (first) double stranded adaptor molecule, and
- solid support with multiple probes, wherein the multiple probes are selected from:
- a plurality of probes that defines a plurality of exons, introns or regulatory sequences from a plurality of genetic loci
- a plurality of probes that defines the complete sequence of at least one single genetic locus, said locus having a size of at least 100 kb, preferably at least 1 Mb or a size as specified herein,
- a plurality of probes that defines sites known to contain SNPs, and
- a plurality of probes that defines an array, in particular a tiling array especially designed to capture the complete sequence of at least one complete chromosome.

Preferably, the kit contains two different double stranded adaptor molecules. The solid support can be either a plurality of beads or a microarray as disclosed herein.

In one embodiment, such a kit further comprises at least one or more compounds from a group consisting of DNA polymerase, T4 polynucleotide kinase, T4 DNA ligase, an array hybridization solution, e.g. as disclosed herein, an array wash solution, in particular a wash solution with SSC, DTT and optionally SDS, e.g. Wash Buffer I (0.2x SSC, 0.2% (v/v) SDS, 0.1mM DTT), Wash Buffer II (0.2x SSC, 0.1mM DTT) and/or Wash Buffer III (0.05x SSC, 0.1mM DTT), and/or an array elution solution, e. g water or a solution containing TRIS buffer and/or EDTA.

In a further specific embodiment, not mutually exclusive to the embodiment disclosed herein, the kit comprises a second adaptor molecule. At least one oligonucleotide strand of said first or second adaptor molecule may carry a modification, which allows for immobilization onto a solid support. For example, such a modification may be a Biotin label which can be used for immobilization on a streptavidin coated solid support. Alternatively, such a modification may be a hapten like digoxygenin, which can be used for immobilization on a solid support coated with a hapten recognizing antibody.

IDT Exhibit 1020.5854

As used herein, the term "hybridization" is used in reference to the pairing of complementary nucleic acids. Hybridization and the strength of hybridization (i.e., the strength of the association between the nucleic acids) is affected by such factors as the degree of complementary between the nucleic acids, stringency of the conditions involved, the T_m of the formed hybrid, and the G:C ratio of the nucleic acids. While the invention is not limited to a particular set of hybridization conditions, stringent hybridization conditions are preferably employed. Stringent hybridization conditions are sequence-dependent and will differ with varying environmental parameters (e.g., salt concentrations, and presence of organics). Generally, "stringent" conditions are selected to be about 5°C to 20°C lower than the thermal melting point (Tm) for the specific nucleic acid sequence at a defined ionic strength and pH. Preferably, stringent conditions are about 5°C to 10°C lower than the thermal melting point for a specific nucleic acid bound to a complementary nucleic acid. The Tm is the temperature (under defined ionic strength and pH) at which 50% of a nucleic acid (e.g., tag nucleic acid) hybridizes to a perfectly matched probe.

Similarly, "stringent" wash conditions are ordinarily determined empirically for hybridization of each set of tags to a corresponding probe array. The arrays are first hybridized (typically under stringent hybridization conditions) and then washed with buffers containing successively lower concentrations of salts, or higher concentrations of detergents, or at increasing temperatures until the signal-to-noise ratio for specific to non-specific hybridization is high enough to facilitate detection of specific hybridization. Stringent temperature conditions will usually include temperatures in excess of about 30°C, more usually in excess of about 37°C, and occasionally in excess of about 45°C. Stringent salt conditions will ordinarily be less than about 1000 mM, usually less than about 500 mM, more usually less than about 150 mM. For further information see e.g., Wetmur et al. (1966) and Wetmur (1991).

"Stringent conditions" or "high stringency conditions," as defined herein, can be hybridization in 50% formamide, 5x SSC (0.75 M NaCl, 0.075 M sodium citrate), 50 mM sodium phosphate (pH 6.8), 0.1% sodium pyrophosphate, 5x Denhardt's solution, sonicated salmon sperm DNA (50 mg/ml), 0.1% SDS, and 10% dextran sulfate at 42° C, with washes at 42°C in 0.2x SSC (sodium chloride/sodium citrate) and 50% formamide at 55°C, followed by a wash with 0.1x SSC containing EDTA at 55°C.

IDT Exhibit 1020.5855

By way of example, but not limitation, it is contemplated that buffers containing 35% formamide, 5x SSC, and 0.1% (w/v) sodium dodecyl sulfate are suitable for hybridizing under moderately non-stringent conditions at 45°C for 16-72 hours. Furthermore, it is envisioned that the formamide concentration may be suitably adjusted between a range of 20-45% depending on the probe length and the level of stringency desired. Also encompassed within the scope of the invention is that probe optimization can be obtained for longer probes (>>50mer), by increasing the hybridization temperature or the formamide concentration to compensate for a change in the probe length. Additional examples of hybridization conditions are provided in several sources, including: "Direct selection of cDNAs with large genomic DNA clones," in Molecular Cloning: A Laboratory Manual (2001).

In another embodiment, the present invention is directed to a method for isolating and reducing the complexity of a plurality of nucleic acid sequences comprising: providing a solid support wherein said solid support comprises hybridization probes hybridizable to target nucleic acid sequences, and a fragmented nucleic acid sample comprising target nucleic acid sequences, amplifying said hybridization probes wherein the amplification products comprise a binding moiety and wherein said amplification products are maintained in solution, hybridizing said nucleic acid sample to said amplification products in solution such that hybridization between said amplification products and target nucleic acid sequences is allowed to occur, separating the target nucleic acid/amplification product hybridization complexes from non-specifically hybridized nucleic acids by said binding moiety, and eluting the hybridized target nucleic acid sequences from the complex thereby isolating and reducing the complexity of a plurality of nucleic acid sequences.

In another embodiment, the present invention comprises the above method which further comprises sequencing the eluted target nucleic acid sequences.

In a further embodiment the invention is directed to a kit comprising: hybridization probe sequences comprising a binding moiety wherein said probe sequences are designed to hybridize to one or more target nucleic acid sequences and wherein said probe sequences are in solution, a substrate comprising a binding partner for binding said binding moiety, and instructions for performing the above methods.

IDT Exhibit 1020.5856

In embodiments of the present invention, a sample containing denatured (e.g., single-stranded) nucleic acid molecules, preferably genomic nucleic acid molecules, which can be fragmented molecules, is exposed under hybridizing conditions to a plurality of oligonucleotide probes, wherein the plurality of oligonucleotide probes or amplicons derived from said probes are in solution, to capture from the sample nucleic acid molecules target nucleic acid sequences and separating non-hybridizing regions of the genome or any other sample nucleic acids from the hybridized target sequences, wherein said separating comprises capturing via a binding moiety (e.g., associated with the probe or probe derived amplicon) the hybridization complexes that are in solution and washing the bound complexes thereby separating the hybridized target sequences from the non-specific non-target hybridized sequences (Figure 6).

The present invention provides methods and systems for isolating a plurality of nucleic acid sequences and reducing the complexity of a large nucleic acid sample, such as a genomic DNA or RNA sample, cDNA library or mRNA library to facilitate further processing and genetic analysis. In some embodiments of the present invention, methods and systems comprise *in situ* amplification of (pre-selected) immobilized nucleic acid probes wherein the probe derived amplicons comprise a binding moiety. The labeled amplicons capture, in solution, target nucleic acid sequences from a sample by hybridizing the sample to the amplicons in solution based methods. The labeled amplicon/target nucleic acid hybrid complex is captured via the binding moiety, preferably washed and the target nucleic acid eluted. The eluted genomic sequences are more amenable to detailed genetic analysis than a genomic sample that has not been subjected to this enrichment procedure. Accordingly, the disclosed methods provide a cost-effective, flexible and efficient approach for reducing the complexity of a genomic sample. Throughout the remainder of the description, genomic samples are used for descriptive purposes, but it is understood that other non-genomic samples could be subjected to the same procedures.

In some embodiments, the invention provides a method of isolating a plurality of nucleic acid sequences and reducing the complexity of a nucleic acid sample by hybridizing the sample against nucleic acid probe amplicons in solution under preferably stringent conditions sufficient to support hybridization between the probe amplicons and complementary regions of the nucleic acid sample. The probe amplicon/target nucleic acid

IDT Exhibit 1020.5857

complexes are washed under conditions sufficient to remove non-specifically bound nucleic acids. The hybridized target nucleic acid sequences are eluted from the probe derived amplicons and may optionally be further amplified (e.g., by LM-PCR), for example for downstream applications such as resequencing.

The present invention provides a method for isolating a plurality of nucleic acid sequences and reducing the genetic complexity of a population of nucleic acid molecules, the method comprising the steps of exposing fragmented, denatured nucleic acid molecules of a target population to multiple, different oligonucleotide probe derived amplicons wherein the amplicons are in solution and wherein the amplicons further comprise a binding moiety, under hybridizing conditions to capture nucleic acid molecules that specifically hybridize to the probe amplicons, binding or capturing the complexes of hybridized molecules by binding the binding moiety found on the probe amplicon to its binding partner (e.g., biotin/SA, digoxigenin/anti-digoxigenin, 6HIS/nickel, etc.), wherein the fragmented, denatured nucleic acid molecules have an average size of about 100 to about 1000 nucleotide residues, preferably about 250 to about 800 nucleotide residues and most preferably about 400 to about 600 nucleotide residues, separating unbound and non-specifically hybridized nucleic acids from the bound probe amplicons, eluting the hybridized target molecules from the amplicons, and optionally sequencing the target molecules.

As such, embodiments of the present invention provide solution based methods and systems for isolating a plurality of nucleic acid sequences and reducing the genetic complexity of a population of nucleic acid molecules. Methods and systems of the present invention comprise exposing fragmented, denatured nucleic acid sample sequences, which may or may not comprise one or more ligation adaptors at one or both ends of the fragmented nucleic acid sample prior to denaturation, to multiple, different hybridization probe amplicons in solution wherein said amplicons are derived from pre-designed multiple, different hybridization probes wherein said amplicons comprise a binding moiety or sequence and optionally a restriction endonuclease (RE) site, under hybridization conditions sufficient to hybridize the denatured nucleic acid target sequences to the probe derived amplicons (e.g., in solution), wherein the fragmented, denatured nucleic acid sequences have an average size of about 100 to about 1000 nucleotide residues, preferably about 250 to about 800 nucleotide residues and most preferably about 400 to about 600

IDT Exhibit 1020.5858

nucleotide residues, separating unbound and non-specifically hybridized nucleic acids from the probe derived amplicons by binding the amplicon/target complexes via the binding moiety and washing the bound complexes, eluting the target nucleic acid sequences from the bound complex wherein the target sequenced demonstrate reduced genetic complexity relative to the original sample and optionally repeating the hybridization, wash and elution steps using the initial eluted enriched target nucleic acid sequences to further enrich for target nucleic acid sequences.

In embodiments of the present invention, probes for capture of target nucleic acids are immobilized on a substrate by a variety of methods. In one embodiment, probes can be spotted onto slides (e.g., US Patent Nos. 6,375,903 and 5,143,854). In preferred embodiments, probes are synthesized *in situ* on a substrate by using maskless array synthesizers (MAS) as described in US Patent No. 6,375,903, 7,037,659, 7,083,975, 7,157, 229 that allows for the *in situ* synthesis of oligonucleotide sequences directly on the slide for subsequent *in situ* polymerase chain reaction (PCR) amplification.

In some embodiments, a solid support is a population of beads or particles. The capture probes are initially synthesized on a microarray slide using a maskless array synthesizer, amplified, released or cleaved off according to standard methods, optionally amplified and immobilized on said population of beads. The beads may be packed, for example, into a column so that a target sample is loaded and passed through the column and hybridization of probe/target sample takes place in the column, followed by washing and elution of target sample sequences for reducing genetic complexity. In some embodiments, a column has fluid inlet and outlet ports. In some embodiments, in order to enhance hybridization kinetics hybridization takes place in an aqueous solution comprising the beads with the immobilized multiple probes in suspension in an aqueous environment.

In some embodiments, nucleic acid probes for target molecules are synthesized on a solid support, released from the solid support as a pool of probes and amplified. The amplified pool of released probed are covalently or non-covalently immobilized onto a support (e.g., glass, metal, ceramic, polymeric beads, paramagnetic particles, etc.). The probes are designed for convenient release from the solid support by, for example, providing at or near the support-proximal probe termini an acid or alkali-labile nucleic acid sequence that releases the probes under conditions of low or high pH, respectively. The art is familiar

IDT Exhibit 1020.5859

with methods for immobilizing nucleic acids onto supports, for example by incorporating a biotinylated nucleotide into the probes and coating a support with streptavidin such that the coated support attracts and immobilizes the probes in the pool. The sample or samples pass across the probe containing support (e.g., slide, column, etc.) under hybridizing conditions such that target nucleic acid molecules that hybridize to the immobilized support can be eluted for subsequent analysis or other use.

In embodiments of the present invention, the initial hybridization probes designed for subsequent amplification for use in solution based capture methods as described herein are printed or deposited on a solid support such as a microarray slide, chip, microwell, column, tube, beads or particles. The substrates may be, for example, glass, metal, ceramic, polymeric beads, etc. In preferred embodiments, the solid support is a microarray (e.g., glass slide), wherein the probes are synthesized on the microarray using a maskless array synthesizer. The lengths of the multiple oligonucleotide probes may vary and are dependent on the experimental design and limited only by the possibility to synthesize such probes. In preferred embodiments, the average length of the population of multiple probes prior to *in situ* amplification is about 20 to about 100 nucleotides, preferably about 40 to about 85 nucleotides, in particular about 45 to about 75 nucleotides. The immobilized hybridization probes are subsequently utilized as the templates for *in situ* PCR amplification and optionally asymmetric PCR amplification thereby providing probe derived amplicons for solution based hybridization and enrichment of target nucleic acid molecules from a complex sample.

In embodiments of the present invention, hybridization probes correspond in sequence to at least one region of a genome and can be provided on a solid support in parallel using, for example, maskless array synthesis (MAS) technology. Alternatively, probes can be obtained serially using a standard DNA synthesizer and then applied to the solid support or can be obtained from an organism and then immobilized on the solid support. In embodiments of the present invention, it is contemplated that hybridization probes, regardless of the method of synthesis, comprise amplification primer sequences for use in amplification techniques. In embodiments of the present invention, amplification primer sequences incorporated into hybridization probe sequences further comprise restriction endonuclease (RE) sequences. In embodiments of the present invention, hybridization probes as found on a microarray substrate are amplified *in situ* using primers

IDT Exhibit 1020.5860

complementary to the primer sequences wherein one or both of the primers further comprise linker chemistry, such as a binding moiety (e.g., biotin, digoxigenin, etc.) and such that the hybridization probe derived PCR amplicons are in solution.

The solution comprising the probe derived amplicons is transferred to, for example, a tube, well, or other vessel and maintained in solution. It is contemplated that one or more additional rounds of amplification to boost the production of the amplicon strand that comprises the binding moiety, for example by asymmetric PCR, is additionally performed. A nucleic acid sample, preferably fragmented and denatured to yield fragmented single stranded target sequences, is added to the amplicons in solution and hybridization is allowed to occur between the probe derived amplicons and the fragmented single stranded target nucleic acid sample. After hybridization, nucleic acids that do not hybridize, or that hybridize non-specifically, are separated from the amplicon/target complex by capturing the amplicon/target complex via the binding moiety and washing the amplicon/target complex. For example, if the binding moiety is biotin, a streptavidin coated substrate is used to capture the complex. The bound complex is washed, for example with one or more washing solutions. The remaining nucleic acids (e.g., specifically bound to the amplicons) are eluted from the complex, for example, by using water or an elution buffer (e.g., comprising TRIS buffer and/or EDTA) to yield an eluate enriched for the target nucleic acid sequences.

Microarray based oligonucleotides provided for amplification in solution based capture methods and systems as described herein are designed to target a region or regions of a genome. The length of individual probes is typically between 50 and 200 bases. These probes may be either designed to be overlapping probes, meaning that the starting nucleotides of adjacent probes are separated in the genome by less than the length of a probe, or non-overlapping probes, where the distance between adjacent probes are greater than the length of a probe. The distance between adjacent probes is generally overlapping, with spacing between the starting nucleotide of two probes varying between 1 and 100 bases. The distance is varied to cause some genomic regions to be targeted by a larger number of probes than others. This variation is used, for example, to modulate the capture efficiency of individual genomic regions, normalizing capture. Probes can be tested for uniqueness in the genome. In preferred embodiments of the present invention, to avoid non-specific binding of genomic elements to probe derived amplicons, highly repetitive

IDT Exhibit 1020.5861

elements of the genome are excluded from selection probe designs using a method that utilizes a strategy similar to the WindowMasker program developed by, for example, Morgolis (2006, Bioinformatics 15:134-141, incorporated herein by reference in its entirety) to identify these regions and exclude them from probe design.

The nature and performance of the designed probes for amplification for solution based capture methods of the present invention can be varied to advantageously normalize or adjust the distribution of the target molecules captured and enriched in accord with the methods of the present invention. A goal of such normalization is to deliver one expressed gene per read (e.g., Soares, et al., 1994, Proc. Natl. Acad. Sci. 91:9228-9232). Normalization is applied, for example, to populations of cDNA molecules before library construction since generally the distribution of molecules in the population reflects the different expression levels of expressed genes from which the cDNA molecule populations are produced. For example, the number of sequencing reactions required to effectively analyze each target region is reduced by normalizing the number of copies of each target sequence in the enriched population such that across the set of probes the capture performance of distinct probes is normalized, on the basis of a combination of fitness and other probe attributes.

Fitness, characterized by a capture metric, is ascertained either informatically or empirically. In one approach, the ability of the target molecules to bind is adjusted by providing so-called isothermal (T_m-balanced) oligonucleotide probes, as described in U.S. Patent Publication No. 2005/10282209, that enable uniform probe performance, eliminate hybridization artifacts and/or bias and provide higher quality output. Probe lengths are adjusted (typically, about 20 to about 100 nucleotides, preferably about 40 to about 85 nucleotides, in particular about 45 to about 75 nucleotides, but optionally also more than 100 nucleotides until about 250 nucleotides) to equalize the melting temperature (e.g. T_m = 76°C, typically about 55°C to about 76°C, in particular about 72°C to about 76°C) across the entire set of probes prior to amplification. Thus, probes are optimized to perform equivalently at a given stringency in the genomic regions of interest, including AT- and GC-rich regions. The skilled artisan will appreciate that probe length, melting temperature and sequence can be coordinately adjusted for any given probe derived amplicon to arrive at a desired hybridization performance for the probe amplicon. For example, the melting temperature (T_m) of the probe derived amplicon can be calculated using the formula:

IDT Exhibit 1020.5862

$T_m=S_x(G_n+C_n)+lx(A_n+T_n)$, where n is the number of each specific base (A, T, G or C) present on the probe amplicon.

Capture performance can also be normalized by ascertaining the capture fitness of probe amplicons in the probe set, and then adjusting the quantity of individual probes on the solid support for amplification purposes accordingly. For example, if probe amplicons derived from a first probe is anticipated to capture twenty times as much nucleic acid as a second set of probe derived amplicons, then the capture performance of both probe amplicon sets can be equalized by providing twenty times as many copies of the second probe from amplification purposes, for example by increasing by twenty-fold the number of microarray probes displaying the second probe prior to amplification.

In other embodiments, an additional strategy for normalizing capture of target nucleic acids is to subject the eluted target molecules to a second round of solution based hybridization against the probe derived amplicons under less stringent conditions than were used for the first hybridization round. Apart from the substantial enrichment in the first hybridization that reduces complexity relative to the original genomic nucleic acid, the second hybridization can be conducted under hybridization conditions that saturate all capture probes. Presuming that substantially equal amounts of the probe derived amplicons are provided in solution, saturation of the amplicons will ensure that substantially equal amounts of each target are eluted after the second hybridization and washing.

In embodiments of the present invention, amplification primers utilized for *in situ* amplification of the hybridization probes for subsequent solution based capture and enrichment methods and systems described herein, comprise linker chemistry such as binding moieties. Binding moieties comprise any moiety that is attached or incorporated into the 5' end of an amplification primer useful in subsequent capture of the probe amplicon/target nucleic acid hybridization complex. A binding moiety is any sequence that is engineered into 5' of a primer sequence, such as a 6 histidine (6HIS) sequence that is capturable. For example, a primer that comprises a 6HIS sequence is captureable by nickel, for example in a tube, microwell, or purification column that is coated with nickel or contains nickel coated beads, particles, etc. wherein the beads are packed into a column and a sample is loaded and passed through the column for capture of the complex for reducing complexity (e.g., and subsequent target elution). Another example of a binding

IDT Exhibit 1020.5863

moiety useful in embodiments of the present invention includes a hapten, for example digoxigenin that is, for example, attached to the 5' end of an amplification primer. Digoxigenin is capturable by use of an antibody to digoxigenin, for example a substrate that is coated or contains an anti-digoxigenin antibody.

In preferred embodiments, an amplification primer used in methods and systems of the present invention contains a biotin moiety attached to the 5' end of the primer and subsequent probe derived amplicons. Biotin is capturable by streptavidin (SA), as such the biotin labeled amplicon can be captured on a substrate or column that is coated or contains SA. In preferred embodiments, streptavidin is coated onto paramagnetic particles that can be in turn captured magnetically for easy washing and elution of the target enriched nucleic acids. The present invention is not limited by the kind of linker chemistry used, and a skilled artisan will know of other options that are equally amendable to methods and systems of the present invention.

In embodiments of the present invention, the methods and systems comprise determining nucleic acid sequence information about at least one region of nucleic acid(s), in particular genomic nucleic acid(s), (the whole genome or at least one whole or partial chromosome) in a sample, the method comprising the steps of performing the methods as previously described followed by determining the nucleic acid sequence of the captured molecules, in particular by performing sequencing by synthesis reactions.

In embodiments of the present invention, target nucleic acids are typically deoxyribonucleic acids or ribonucleic acids, and include products synthesized *in vitro* by converting one nucleic acid molecule type (e.g., DNA, RNA and cDNA) to another as well as synthetic molecules containing nucleotide analogues. Denatured genomic DNA molecules are in particular molecules that are shorter than naturally occurring genomic nucleic acid molecules. A skilled person can produce molecules of random- or non-random size from larger molecules by chemical, physical or enzymatic fragmentation or cleavage using well known protocols. For example, chemical fragmentation can employ ferrous metals (e.g., Fe-EDTA), physical methods can include sonication, hydrodynamic force or nebulization (e.g., see European patent application EP 0 552 290) and enzymatic protocols can employ nucleases such as micrococcal nuclease (Mnase) or exo-nucleases (such as Exo1 or Bal31) or restriction endonucleases.

IDT Exhibit 1020.5864

The present invention is not limited to the method in which fragments are generated and any method useful in fragmenting nucleic acids is contemplated. In embodiments of the present invention, fragments in a size range compatible with the post-enrichment technology in which the enriched fragments are preferred. For example, embodiments of the present invention contemplate nucleic acid fragment sizes in the range of between about 100 and about 1000 nucleotide residues or base pairs, or between about 250 and about 800 nucleotide residues or base pairs, or about 400 to about 600 nucleotide residues or base pairs, in particular about 500 nucleotide residues or base pairs.

The population of nucleic acid molecules which may comprise the target nucleic acid sequences preferably contains the whole genome or at least one chromosome of an organism or at least one nucleic acid molecule with at least about 100 kb. In particular, the size(s) of the nucleic acid molecule(s) is/are at least about 200 kb, at least about 500 kb, at least about 1 Mb, at least about 2 Mb or at least about 5 Mb, especially a size between about 100 kb and about 5 Mb, between about 200 kb and about 5 Mb, between about 500 kb and about 5 Mb, between about 1 Mb and about 2 Mb or between about 2 Mb and about 5Mb. In some embodiments, the nucleic acid molecules are genomic DNA, while in other embodiments the nucleic acid molecules are cDNA, or RNA species (e.g., tRNA, mRNA, miRNA).

In embodiments of the present invention, the nucleic acid molecules which may or may not comprise the target nucleic acid sequences may be selected from an animal, a plant or a microorganism, in particular embodiments the nucleic acid molecules are from a primate, preferably a human. In some embodiments, if limited samples of nucleic acid molecules are available the nucleic acids are amplified (e.g., by whole genome amplification) prior to practicing the method of the present invention. For example, prior amplification may be necessary for performing embodiments of the present invention for forensic purposes (e.g., in forensic medicine, etc.).

It is contemplated that in preferred embodiments, the population of nucleic acid molecules is a population of genomic DNA molecules. The hybridization probes and subsequent amplicons may comprise one or more sequences that target a plurality of exons, introns or regulatory sequences from a plurality of genetic loci, the complete sequence of at least one

IDT Exhibit 1020.5865

single genetic locus, said locus having a size of at least 100 kb, preferably at least 1 Mb, or at least one of the sizes as specified above, sites known to contain SNPs, or sequences that define an array, in particular a tiling array, designed to capture the complete sequence of at least one complete chromosome.

It is contemplated that target nucleic acid sequences are enriched from one or more samples that include nucleic acids from any source, in purified or unpurified form. The source need not contain a complete complement of genomic nucleic acid molecules from an organism. The sample, preferably from a biological source, includes, but is not limited to, pooled isolates from individual patients, tissue samples, or cell culture. The target region can be one or more continuous blocks of several megabases, or several smaller contiguous or discontiguous regions, such as all of the exons from one or more chromosomes, or sites known to contain SNPs. For example, the hybridization probes and subsequent probe derived amplicons can support a tiling array designed to capture one or more complete chromosomes, parts of one or more chromosomes, all exons, all exons from one or more chromosomes, selected exons, introns and exons for one or more genes, gene regulatory regions, and so on.

Alternatively, to increase the likelihood that desired non-unique or difficult-to-capture targets are enriched, the probes can be directed to sequences associated with (e.g., on the same fragment as, but separate from) the actual target sequence, in which case genomic fragments containing both the desired target and associated sequences will be captured and enriched. The associated sequences can be adjacent or spaced apart from the target sequences, but a skilled person will appreciate that the closer the two portions are to one another, the more likely it will be that genomic fragments will contain both portions. To reduce the limited impact of cross-hybridization by off-target molecules, thereby enhancing the integrity of the enrichment, sequential rounds of capture using distinct but related capture probe sets, and thus probe derived amplicons, directed to the target region is performed. Related probes are probes corresponding to regions in close proximity to one another in the genome that hybridize to the same genomic DNA fragment.

In some embodiments of the present invention, the methods comprise the step of ligating adaptor or linker molecules to one or both ends of the nucleic acid molecules prior to denaturation and hybridization to the probe amplicons in solution.

IDT Exhibit 1020.5866

In some embodiments of the present invention the methods further comprise amplifying said adaptor modified nucleic acid molecules with at least one primer, said primer comprising a sequence which specifically hybridizes to the sequence of said adaptor molecule(s).

In some embodiments of the present invention, double-stranded linkers are provided at one or both ends of the fragmented nucleic acid molecules before sample denaturation and hybridization to the probe derived amplicons in solution. In such embodiments, target nucleic acid molecules are amplified after elution to produce a pool of amplified products having further reduced complexity relative to the original sample. The target nucleic acid molecules can be amplified using, for example, non-specific Ligation Mediated-PCR (LM-PCR) through multiple rounds of amplification and the products can be further enriched, if required, by one or more rounds of selection against the amplicon derived probes. The linkers or adaptors are provided, for example, in an arbitrary size and with an arbitrary nucleic acid sequence according to what is desired for downstream analytical applications subsequent to the complexity reduction step. The linkers can range between about 12 and about 100 base pairs, including a range between about 18 and 100 base pairs, and preferably between about 20 and 24 base pairs. Adaptor molecules in the context of the present invention are preferably defined as blunt-ended double-stranded oligonucleotides.

In order to ligate adaptor molecules onto a double stranded target molecule, it is preferred that this target molecule itself is blunt ended. In order to achieve this, the double stranded target molecules are subjected to, for example, a fill-in reaction with a DNA Polymerase such as T4 DNA polymerase or Klenow polymerase in the presence of dNTPs, which results in blunt ended target molecules. In addition, ends of the fragments are phosphorylated using T4 Polynucleotide kinase and methods known to skilled artisans (for example, see Molecular Cloning: A Laboratory Manual, Eds. Sambrook et al., Cold Spring Harbour Press; incorporated herein by reference in its entirety) to add phosphate groups to the 5' termini of the fragments prior to the ligation of the adaptors. Subsequent ligation of the adaptors (e.g., short double stranded blunt end DNA oligonucleotides with about 3-20 base pairs) onto the polished, phosphorylated target DNA may be performed according to any method which is known in the art, for example by T4 DNA ligase reaction.

IDT Exhibit 1020.5867

The ligation of the adaptors to the fragmented target nucleic acid molecules may be performed prior to, or after, exposing a sample that comprises fragmented, denatured genomic nucleic acid molecules to multiple oligonucleotide probes amplicons in solution under hybridizing conditions to capture target nucleic acid molecules. When ligation is performed after hybridization, the enriched nucleic acids which are released from the amplicons in single stranded form are initially re-annealed followed by a primer extension reaction and a fill-in reaction according to standard methods known in the art.

Ligation of adaptor molecules allows for a step of subsequent amplification of the captured molecules. Independent from whether ligation takes place prior to or after the capturing step, there exist several alternative embodiments. In one embodiment, one type of adaptor molecule (e.g., adaptor molecule A) is ligated that results in a population of fragments with identical terminal sequences at both ends of the fragment. As a consequence, it is sufficient to use only one primer in a potential subsequent amplification step. In an alternative embodiment, two types of adaptor molecules A and B are used. This results in a population of enriched molecules composed of three different types: (i) fragments having one adaptor (A) at one end and another adaptor (B) at the other end, (ii) fragments having adaptors A at both ends, and (iii) fragments having adaptors B at both ends. The generation of enriched molecules with adaptors is of outstanding advantage, if amplification and sequencing is to be performed, for example using the 454 Life Sciences Corporation GS20 and GSFLX instrument (e.g., see GS20 Library Prep Manual, Dec 2006, WO 2004/070007; incorporated herein by reference in their entireties).

The present invention is directed to a method for detecting coding region variation(s) of a test genome sample relative to a reference genome sample, in particular relative to a reference genome that comprises fragmented, denatured genomic nucleic acid molecules, the method comprising the steps as previously described on both a test and reference genome, further comparing the sequences to a sequence in a database, in particular to a sequence in a database of polymorphisms in a reference genome sample to identify variants from a test genome sample. The invention is, therefore, useful in searching for genetic variants and mutations, such as single nucleotide polymorphisms (SNP), or set of SNPs, genomic insertions and/or deletions, translocations, etc. that may underlie human diseases. It is contemplated that capture and enrichment using solution based hybridization

IDT Exhibit 1020.5868

technology as described herein is more flexible than other methods currently available in the field of genomic enrichment.

In some embodiments of the present invention, the eluted target nucleic acid sequences may be sequenced, hybridized to a resequencing or SNP-calling array and the sequence or genotypes may be further analyzed. Solution based enrichment as provided by embodiments of the present invention enables targeted array-based-, shotgun-, capillary-, or other sequencing methods known in the art. In general, strategies for shotgun sequencing of randomly generated fragments are cost-effective and readily integrated into a pipeline. The present invention enhances the efficiency of the shotgun approach by presenting only fragments from one or more genomic regions of interest for sequencing. The invention provides an ability to focus the sequencing strategies on specific genomic regions, such as individual chromosomes or exons for medical sequencing purposes. As such, a more focused approach to disease discovery is realized.

In embodiments of the present invention, the eluted target nucleic acid sequences resulting from solution based enrichment methods as described herein, are subsequently sequenced. Sequencing can be performed by a number of different methods, such as by employing sequencing by synthesis technology. Sequencing by synthesis according to the prior art is defined as any sequencing method which monitors the generation of side products upon incorporation of a specific deoxynucleoside-triphosphate during the sequencing reaction (Hyman, 1988, Anal. Biochem. 174:423-436; Rhonaghi et al., 1998, Science 281:363-365). One prominent embodiment of the sequencing by synthesis reaction is the pyrophosphate sequencing method. In this case, generation of pyrophosphate during nucleotide incorporation is monitored by an enzymatic cascade which results in the generation of a chemo-luminescent signal. The 454 Genome Sequencer System (Roche Applied Science cat. No. 04 760 085 001), an example of sequence by synthesis, is based on the pyrophosphate sequencing technology. For sequencing on a 454 GS20 or 454 FLX instrument, the average genomic DNA fragment size is in the range of 200 or 600 bp, respectively, as described in the product literature.

A sequencing by synthesis reaction can alternatively be based on a terminator dye type of sequencing reaction. In this case, the incorporated dye deoxynucleotriphosphates (ddNTPs) building blocks comprise a detectable label, which is preferably a fluorescent label that

IDT Exhibit 1020.5869

prevents further extension of the nascent DNA strand. The label is then removed and detected upon incorporation of the ddNTP building block into the template/primer extension hybrid for example by using a DNA polymerase comprising a 3'-5' exonuclease or proofreading activity.

In case of the Genome Sequencer workflow (Roche Applied Science Catalog No. 04 896 548 001), in a first step, (clonal) amplification is performed by emulsion PCR. Thus, it is also within the scope of the present invention, that the step of amplification is performed by emulsion PCR methods. The beads carrying the clonally amplified target nucleic acids may then become arbitrarily transferred into a picotiter plate according to the manufacturer's protocol and subjected to a pyrophosphate sequencing reaction for sequence determination.

In some embodiments, the present invention comprises a kit comprising reagents and materials for performing methods according to the present invention. Such a kit may include one or more of a microarray substrate upon which is immobilized a plurality of hybridization probes specific to one or more target nucleic acid sequences from one or more target genetic loci (e.g., specific to exons, introns, SNP sequences, etc.), a plurality of probes that define a tiling array designed to capture the complete sequence of at least one complete chromosome, amplification primers, reagents for performing polymerase chain reaction methods (e.g., salt solutions, polymerases, dNTPs, amplification buffers, etc.), reagents for performing ligation reactions (e.g., ligation adaptors, T4 polynucleotide kinase, ligase, buffers, etc.), substrates comprising a binding partner moiety, tubes, hybridization solutions, wash solutions, elution solutions, magnet(s), and tube holders.

In some embodiments, the present invention provides a system (e.g., kit) for performing a method or part of a method according to the present invention as disclosed herein. Thus, the present invention is a kit comprising a (first) double stranded adaptor molecule and multiple probe derived amplicons in solution, wherein the probe derived amplicons are amplified from a plurality of probes that defines a plurality of exons, introns and/or regulatory sequences from a plurality of genetic loci, and/or a plurality of probe derived amplicons in solution that defines the complete sequence of at least one single genetic locus, said locus having a size of at least 100 kb, preferably at least 1 Mb or a size as specified herein, and/or a plurality of probe derived amplicons that defines sites known to

IDT Exhibit 1020.5870

contain SNPs, and/or a plurality of probe derived amplicons that defines an array, in particular a tiling array especially designed to capture the complete sequence of at least one complete chromosome. In some embodiments, a kit further comprises two different double stranded adaptor molecules.

In some embodiments, a kit comprises one or more capture molecules or compounds. For example, at least one oligonucleotide probe comprises a modification which allows for immobilization onto a solid support. For example, a probe comprises a biotin moiety for immobilization onto a streptavidin coated paramagnetic particle. Another example is a hapten, such as digoxigenin, that is associated with a probe for immobilization on a solid support using a hapten recognizing antibody (e.g., anti-digoxigenin).

In some embodiments, a kit further comprises at least one or more compounds from a group consisting of DNA polymerase, T4 polynucleotide kinase, T4 DNA ligase, one or more array hybridization solutions, and/or one or more array wash solutions. In preferred embodiments, three wash solutions are included in a kit of the present invention, the wash solutions comprising SSC, DTT and optionally SDS. For example, kits of the present invention comprise Wash Buffer I (0.2% SSC, 0.2% (v/v) SDS, 0. 1 mM DTT), Wash Buffer II (0.2% SSC, 0.1mM, DTT) and/or Wash Buffer III (0.05% SSC, 0.1 mM DTT). In some embodiments, systems of the present invention further comprise an elution solution, for example water or a solution containing TRIS buffer and/or EDTA.

The following examples are provided as further non-limiting illustrations of particular embodiments of the invention.

EXPERIMENTATION

The following examples are provided in order to demonstrate and further illustrate certain preferred embodiments and aspects of the present invention and are not to be construed as limiting the scope thereof.

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 limits of that range is also specifically disclosed. Each smaller range

IDT Exhibit 1020.5871

between any stated value or intervening value in a stated 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 or excluded in the range, and each range where either, neither or both limits are included in the smaller ranges is 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.

Example 1 - Discovery of New Polymorphisms and Mutations in Large Genomic Regions

This generic example describes how to perform selection that allows for rapid and efficient discovery of new polymorphisms and mutations in large genomic regions. Microarrays having immobilized probes are used in one- or multiple rounds of hybridization selection with a target of total genomic DNA, and the selected sequences are amplified by LM-PCR (see Figs.1 and 2).

a) Preparation of the genomic DNA and double-stranded linkers

DNA is fragmented using sonication to an average size of ~500 base pairs. A reaction to polish the ends of the sonicated DNA fragments is set up:

DNA fragments	41 µl
T4 DNA Polymerase	20 µl
T4 DNA polymerase reaction mix	20 µl
Water	10 µl

The reaction is incubated at 11°C for 30 min. The reaction is then subjected to phenol/chloroform extraction procedures and the DNA is recovered by ethanol precipitation. The precipitated pellet is dissolved in 10 µl water (to give a final concentration of 2 µg/µl).

Two complementary oligonucleotides are annealed to create a double-stranded linker, by mixing the following:

IDT Exhibit 1020.5872

Oligonucleotide 1 (1μg/μl)	22.5 μl
(5'-CTCGAGAATTCTGGATCCTC-3') (SEQ ID NO:1)	
Oligonucleotide 2 (1μg/μl)	22.5 μl
(5'-GAGGATCCAGAATTCTCGAGTT-3') (SEQ ID NO:2)	
10x annealing buffer	5 μl
Water	to 50 μl

The reaction is heated at 65°C for 10 min; then allowed to cool at 15–25°C for 2 hours. The length of the 2 complementary oligonucleotides 1 and 2 is between 12 and 24 nucleotides, and the sequence is selected depending upon the functionality desired by the user. The double-stranded linker is then purified by column chromatography through a Sephadex G-50 spin column. The purified linker solution is then concentrated by lyophilization to a concentration of 2 μg/μl.

b) Ligation of linkers to genomic DNA fragments

The following reaction to ligate the linkers to genomic DNA fragments is set up. The reaction is incubated at 14°C overnight.

Annealed linkers from Step a) (20 μg)	10 μl
Genomic DNA from Step a) (10 μl)	5 μl
T4 DNA ligase	10 U
10x ligation buffer	2 μl
Water	to 20 μl

The reaction volume is adjusted to 500 μl with water and the ligated genomic DNA is purified using a QIAquick PCR purification kit. The purified DNA is stored at a concentration of 1 μg/μl.

c) Primary selection and capture of hybrids

To prepare the genomic DNA sample for hybridization to the microarray, linker modified genomic DNA (10μg) is resuspended in 3.5μl of nuclease-free water and combined with

IDT Exhibit 1020.5873

31.5µl NimbleGen Hybridization Buffer (Roche NimbleGen, Inc., Madison, WI), 9µl Hybridization Additive (Roche NimbleGen, Inc), in a final volume of 45µl. The samples are heat-denatured at 95°C for 5 minutes and transferred to a 42°C heat block.

To capture the target genomic DNA on the microarray, samples are hybridized to NimbleGen CGH arrays, manufactured as described in US6,375,903 (Roche NimbleGen, Inc.). Maskless fabrication of capture oligonucleotides on the microarrays is performed by light-directed oligonucleotide synthesis using a digital micromirror as described in Singh-Gasson et al. (1999, Nat. Biotech. 17:974-978, incorporated herein by reference in its entirety) as performed by a maskless array synthesizer. Gene expression analysis using oligonucleotide arrays produced by maskless photolithography is described in Nuwaysir et al. (2002, Genome Res. 12:1749-1755, incorporated herein by reference in its entirety). Hybridization is performed in a MAUI Hybridization System (BioMicro Systems, Inc., Salt Lake City, UT) according to manufacturer instructions for 16 hours at 42°C using mix mode B. Following hybridization, arrays are washed twice with Wash Buffer I (0.2x SSC, 0.2% (v/v) SDS, 0.1mM DTT, NimbleGen Systems) for a total of 2.5 minutes. Arrays are then washed for 1 minute in Wash Buffer II (0.2x SSC, 0.1mM DTT, NimbleGen Systems) followed by a 15 second wash in Wash Buffer III (0.05x SSC, 0.1mM DTT, Roche NimbleGen, Inc.).

To elute the genomic DNA hybridized to the microarray, the arrays are incubated twice for 5 minutes in 95°C water. The eluted DNA is dried down using vacuum centrifugation.

d) Amplification of the primary selected DNA

The primary selected genomic DNA is amplified as described below. Ten separate replicate amplification reactions are set up in 200µl PCR tubes. Only one oligonucleotide primer is required because each fragment has the same linker ligated to each end:

<u>Reaction reagents:</u>

Template: primary selection material	5 µl
Oligonucleotide 1 (200 ng/µl) (5'-CTCGAGAATTCTGGATCCTC-3') (SEQ ID NO:1)	1 µl

IDT Exhibit 1020.5874

dNTPs (25mM each)	0.4 µl
10x PfuUltra HF DNA polymerase	
Reaction buffer	5 µl
PfuUltra HF DNA polymerase	2.5 U
Water	to 50 µl

The reactions are amplified according to the following program:

Cycle number	Denaturation	Annealing	Polymerization
1	2 min at 95°C		
2-31	30 s at 95°C	30 s at 55°C	1 min at 72°C

The reaction products are analyzed by agarose gel electrophoresis. The amplification products are purified using a QIAquick PCR purification kit. The eluted samples are pooled and the concentration of amplified primary selected DNA is determined by spectrophotometry. A volume of DNA in the pool equivalent to 1 µg is reduced to 5 µl in a speed vacuum concentrator. One µl (at least 200 ng) of the primary selected material is set aside for comparison with the secondary selection products. As necessary, subsequent rounds of enrichment are performed by further rounds of array hybridization and amplification of the eluted sample.

e) Preparation of target oligonucleotide probes for release from microarray and immobilization on support

Probes are synthesized on a microarray, then are released using a base-labile Fmoc (9-fluorenylmethyloxycarbonyl) group. The probes are labelled with biotin and are then immobilized onto the surface of a streptavidin solid support using known methods for covalent or non-covalent attachment.

Optionally, prior to immobilization onto the solid support, the synthesized probes are amplified using LM-PCR, Phi29 or other amplification strategy to increase the amount of the synthesized probes by virtue of inserting sequences upon them that facilitate their amplification. This material can now be used for direct sequencing, array based resequencing, genotyping, or any other genetic analysis targeting the enriched region of the

IDT Exhibit 1020.5875

genome by employing solution phase hybridization and SA mediated capture of the hybridization products.

Example 2 - Array-Targeted Resequencing

A series of high-density oligonucleotide microarrays that capture short segments that correspond to 6,726 individual gene exon regions of at least 500 base pairs were chosen from 660 genes distributed about the human genome (sequence build HG17) (approximately 5Mb of total sequence) were synthesized according to standard Roche NimbleGen, Inc. microarray manufacturing protocols. Overlapping microarray probes of more than 60 bases each on the array spanned each target genome region, with a probe positioned each 10 bases for the forward strand of the genome.

Highly-repetitive genomic regions were excluded by design from the capture microarrays, to reduce the likelihood of non-specific binding between the microarrays and genomic nucleic acid molecules. The strategy for identifying and excluding highly-repetitive genomic regions was similar to that of the WindowMasker program (Morgulis et al.). The average 15-mer frequency of each probe was calculated by comparing the frequencies of all 15-mers present in the probe against a pre-computed frequency histogram of all possible 15-mer probes in the human genome. The likelihood that the probe represents a repetitive region of the genome increases as the average 15-mer frequency increases. Only probes having an average 15-mer frequency below 100 were included on the capture microarrays.

To test the reproducibility of the capture system, the exonic design was first used to capture fragmented genomic DNA from a human cell line (Burkitt's Lymphoma, NA04671 (Coriell)) using the method shown schematically in Fig.2. Briefly, genomic DNA (20 μg) was subjected to whole genome amplification (WGA; using Qiagen service (Hilden, Germany)). Twenty μg of the whole genome amplification (WGA) product was treated with Klenow fragment of DNA polymerase I (NEB, Beverly MA) to generate blunt-ends. The blunt-ended fragments were sonicated to generate fragments of about 500 base pairs and then 5′ phosphorylated with polynucleotide kinase (NEB). Oligonucleotide linkers 5′-Pi-GAGGATCCAGAATTCTCGAGTT-3′ (SEQ ID NO:2) and 5′-

IDT Exhibit 1020.5876

CTCGAGAATTCTGGATCCTC-3′ (SEQ ID NO:1) were annealed and ligated to the ends of the 5' phosphorylated fragments:

The linker-terminated fragments were denatured to produce single stranded products that were exposed to the capture microarrays under hybridization conditions in the presence of 1x hybridization buffer (Roche NimbleGen, Inc.) for approximately 65 hours at 42°C with active mixing using a MAUI hybridization station (Roche NimbleGen, Inc.). Single-stranded molecules that did not hybridize were washed from the microarrays under stringent washing conditions, 3 x 5 minutes with Stringent Wash Buffer (Roche NimbleGen, Inc.) and rinsed with Wash Buffers I, II, and III (Roche NimbleGen, Inc.). Fragments captured on the microarrays were immediately eluted with 2 x 250 µl of water at 95°C, dried and resuspended for amplification by LM-PCR using a primer complementary to the previously ligated linkers oligonucleotides.

To quantify enrichment of the exonic regions, eight random regions were selected for quantitative PCR (qPCR). These regions were amplified using the following primers:

Region 1 F: 5′-CTACCACGGCCCTTTCATAAAG-3′ (SEQ ID NO:3)
R: 5′-AGGGAGCATTCCAGGAGAGAA-3′ (SEQ ID NO:4)

Region 2 F: 5′-GGCCAGGGCTGTGTACAGTT-3′ (SEQ ID NO:5)
R: 5′-CCGTATAGAAGAGAAGACTCAATGGA-3′ (SEQ ID NO:6)

Region 3 F: 5′-TGCCCCACGGTAACAGATG-3′ (SEQ ID NO:7)
R: 5′-CCACGCTGGTGATGAAGATG-3′ (SEQ ID NO:8)

Region 4 F: 5′-TGCAGGGCCTGGGTTCT-3′ (SEQ ID NO:9)
R: 5′-GCGGAGGGAGAGCTCCTT-3′ (SEQ ID NO:10)

Region 5 F: 5′-GTCTCTTTCTCTCTCTTGTCCAGTTTT-3′ (SEQ ID NO:11)
R: 5′-CACTGTCTTCTCCCGGACATG-3′ (SEQ ID NO:12)

Region 6 F: 5′-AGCCAGAAGATGGAGGAAGCT-3′ (SEQ ID NO:13)
R: 5′-TTAAAGCGCTTGGCTTGGA-3′ (SEQ ID NO:14)

IDT Exhibit 1020.5877

Region 7 F: 5′-TCTTTTGAGAAGGTATAGGTGTGGAA-3′ (SEQ ID NO:15)
R: 5′-CAGGCCCAGGCCACACT-3′ (SEQ ID NO:16)

Region 8 F: 5′-CGAGGCCTGCACAGTATGC-3′ (SEQ ID NO:17)
R: 5′-GCGGGCTCAGCTTCTTAGTG-3′ (SEQ ID NO:18)

After a single round of microarray capture, the enriched, amplified samples and control genomic DNA, that was fragmented, linker-ligated and LM-PCR amplified, but not hybridized to a capture array, were compared using an ABI 7300 real time PCR system (Applied Biosystems, Foster City, CA) measuring SYBR green fluorescence according to manufacturer's protocols. An average of 378-fold enrichment was achieved for three replicate exonic capture products. The theoretical maximum enrichment level was 600 fold (3,000 Mb in the genome and 5 Mb of total sequence).

Samples eluted from the capture microarrays were ligated to 454-sequencing-compatible linkers, amplified using emulsion PCR on beads and sequenced using the 454 FLX sequencing instrument (454, Branford CT). Because each sequenced fragment also contained the 20bp LM-PCR linker used immediately after microarray elution, the majority of 454 sequencing reads contained that linker sequence. DNA sequencing of the three replicates on the 454 FLX instrument generated 63 Mb, 115 Mb, and 93 Mb of total sequence. Following *in silico* removal of the linker sequence, each sequencing read was compared to the entire appropriate version of the Human Genome using BLAST analysis (Altschul, et al., 1990, J. Mol. Biol. 215:403-410; incorporated herein by reference in its entirety) using a cutoff score of $e = 10^{-48}$, tuned to maximize the number of unique hits. Reads that did not uniquely map back to the genome (between 10 and 20%) were discarded. The rest were considered captured sequences. Captured sequences that, according to the original BLAST comparison, map uniquely back to regions within the target regions were considered sequencing hits. These were then used to calculate the % of reads that hit target regions, and the fold sequencing coverage for the entire target region. Data was visualized using SignalMap software (Roche NimbleGen, Inc.).

BLAST analysis showed that 91%, 89%, and 91% of reads, respectively, mapped back uniquely to the genome; 75%, 65%, and 77% were from targeted regions and 96%, 93%,

IDT Exhibit 1020.5878

and 95% of target sequences contained at least one sequence read (Table 1, upper three rows) representing an average enrichment of about 400 fold. The median per-base coverage for each sample was 5-, 7- and 7-fold coverage, respectively.

Table 1:

DNA Sample	qPCR Fold Enrichment	FLX – Yield (Mb)	Percentage of Reads Mapped Uniquely to the Genome	Percentage of Total Reads That Mapped to Selection Targets	Median Fold Coverage for Target Regions
NA04671	318	63.1	91%	75%	5
NA04671	399	115	89%	65%	7
NA04671	418	93.0	91%	76%	7
HapMap CEPH	217	77.6	88%	74%	7
HapMap JPT	153	96.7	84%	66%	8
HapMap CHB	240	52.8	83%	59%	4
HapMap YRI	363	81.3	53%	38%	4

Example 3 - Sequence Variation Captured by Genomic Enrichment and Resequencing

To ascertain the ability to discern variation in the human genome, genomic DNA samples from four cell types in the human HapMap collection (CEPH/ NA11839, CHB/NA18573, JPT/NA18942, YRI/NA18861, Coriell) were captured on the exon arrays of the prior examples, eluted and sequenced, as disclosed herein, except that the genomic DNAs were not whole genome amplified before capture. The capture results (shown in Table 1, rows 4-7) were similar to those above, except that sequence coverage was consistently more uniform than before, suggesting a bias introduced during WGA.

The sequence from the four HapMap samples was assembled and mutations were identified and compared to the HapMap SNP data for each sample (Tables 1 and 2). The total number of positions in the target regions that were genotyped in the HapMap project was 8103 (CEU), 8134 (CHB), 8134 (JPT), 8071 (YRI) for each of the four genomes. Of these, most (~6000) sites were homozygous for the reference genome allele. The number

IDT Exhibit 1020.5879

of known variant alleles (homozygous or heterozygous) is listed in the second row of Table 2. These positions were analyzed for coverage and to determine whether the allele(s) were found in the captured DNA.

Table 2:

Pop/Indiv	CEPH/ NA11839	CHB/NA18573	JPT/ NA18942	CEPH/ NA11839
# Known variant alleles	2235	2257	2206	2334
Stringency of at least one read per known variant HapMap allele				
Positions with ≥ 1 read	2176 (97.3%)	2104 (93.2%)	2168 (98.2%)	2133 (91.3%)
Variant alleles found in ≥ 1 read	2071 (92.6%)	1922 (85.1%)	2080 (94.2%)	1848 (79.1%)
False negative rate	7.4%	14.9%	5.8%	20.9%
Stringency of at least two reads per known variant HapMap allele				
Positions with ≥ 1 read	2176 (97.3%)	2104 (93.2%)	2168 (98.2%)	2133 (91.3%)
Variant alleles found in ≥ 2 reads	1907 (85.3%)	1569 (69.5%)	1939 (87.8%)	1469 (62.9%)
False negative rate	14.7%	30.5%	12.2%	37.1%

Between 94% and 79% of known variant positions among the HapMap samples were identified with at least one sequence read, which was expected, based upon the overall sequence coverage. There was no apparent biasing against alleles not present on the capture array when coverage of targets that contained 0, 1 or >1 known variants, (7.95, 8.48, and 8.82 fold coverage respectively) were compared.

There is considerable interest in the analysis of large contiguous genomic regions. Capture microarray series that target single long segments from 200kb – 5Mb surrounding the human BRCA1 gene were tested with the NA04671 DNA. For array series used to capture the BRCA1 gene locus, five genomic regions of increasing size (200 kb, 500kb, 1Mb, 2Mb, and 5Mb) surrounding the BRCA1 gene locus were chosen from the human genome

IDT Exhibit 1020.5880

sequence (build HG18). Attributes of the locus-capture arrays are shown in Table 3. The average probe tiling density is the average distance between the start of one probe and the start of the next probe.

Table 3:

BRCA1 Region Size	Average Selection Probe Tiling Density (base pairs)	Chromosome 17 coordinates (HG18)
200kb	1bp	38,390,417 – 38,590,417
500kb	1bp	38,240,417 – 38,740,417
1Mb	2bp	37,990,417 – 38,990,417
2Mb	3bp	37,490,417 – 39,490,417
5Mb	7bp	35,990,417 – 40,990,417

Table 4 shows that all capture targets performed well, with up to 140 Mb of raw sequence generated in a single sequencing machine run, generating ~18 fold coverage, from a 5 Mb capture region. Fig. 4b provides sequence read map details for the locus-specific capture and sequencing. Line 1 depicts the chromosome position of 2000 bases on human chromosome 17, line 2 shows the location of the probes, spaced every 10 base pairs and staggered along the Y axis, the chart at 3 shows the per-base fold sequence coverage, which ranges between 0 and 100 percent, and item 4 depicts the read map of the highest BLAST scores for 454 sequencing reads. Fig. 5 displays cumulative per-base sequence coverage (Fig. 5a) and a sequence coverage histogram (Fig. 4c) for the BRCA1 2Mb region. The percentage of reads that map to the target sequence increased with the size of the target region.

IDT Exhibit 1020.5881

Table 4:

Tiling Size (kb)	Average Selection Probe Tiling Density	FLX –Yield (Mb)	Percentage of Reads Mapped Uniquely to the Genome	Percentage of Total Reads That Mapped to Selection Targets	Median fold coverage of Unique Portion of Region
200	1bp	102	55%	14%	79
500	1bp	85.0	61%	36%	93
1,000	2bp	96.7	56%	35%	38
2000	3bp	112.6	81%	60%	37
5,000	7bp	140	81%	64%	18

These data illustrate the power of microarray-based direct selection methods for enriching targeted sequences. The inventor used a programmable high-density array platform with 385,000 probes that were readily able to capture at least 5Mb of total sequence. In addition to the specificity of the assay, the high yields of the downstream DNA sequencing steps are consistently superior to the routine average performance using non-captured DNA sources. This is attributed to the capture-enrichment process providing a useful purification of unique sequences away from repeats and other impurities that can confound, for example, the first emulsion PCR step of the 454 sequencing process.

Example 4 - Solution Phase Capture and Resequencing

The sample of Examples 2 and 3 was tested using capture probes synthesized upon, then liberated from, a solid support such that the enrichment was advantageously executed in solution phase. Standard microarray designs (e.g. the BRCA1 200K Tiling array and human exon capture arrays of the prior examples) were modified by adding terminal 15mer primer sequences containing an MlyI recognition site, which facilitates enzymatic primer removal while leaving the capture oligonucleotide sequence intact.

Arrays were synthesized by adding chemical phosphorylating reagent (Glen Research) after the initial T_5 linker and before the 3' primer sequence. Three individual couplings were performed to maximize subsequent cleavage of capture probes from the arrays.

IDT Exhibit 1020.5882

The array-immobilized capture probes were treated with 30% ammonium hydroxide (NH_4OH, Aldrich). After synthesis, arrays were placed in a humid chamber and approximately 700 µl of NH_4OH was applied to the synthesis area at ambient room temperature for 20 minutes to cleave the probes from the array. The NH_4OH remained largely within the confines of the synthesis area because of hydrophobicity differences between the reaction area and the surrounding glass. The solution was removed using a pipette and was retained. An additional 700 µl of fresh NH_4OH was applied to the surface. The process was repeated for a total of 3x (60 min and 2.1 ml total). Cleaved oligonucleotide capture probes were then dried by centrifugation under vacuum under standard conditions known in the art.

The cleaved capture probes were amplified under standard conditions. Dried probes were resuspended in 30 µl deionized water (diH_2O) and aliquoted into 30 individual PCR runs as follows:

Reaction reagents:

10x buffer	2.5 µl
25mM dNTPs	0.125 µl
20µM Primer 1a	1.25 µl
20µM Primer 1b (biotinylated)	1.25 µl
HotStart Taq	0.25 µl
MgCl	1 µl
Sample	1 µl
H_2O	17.625 µl
Total volume	25 µl

Primer 1a:

5'-TGCCGGAGTCAGCGT-3' (SEQ ID NO:19)

Primer 1b:

5'-Biotin-AGTCAGAGTCGCCAC-3' (SEQ ID NO:20)

IDT Exhibit 1020.5883

The reactions are amplified according to the following program:

Cycle number	Denaturation	Annealing	Polymerization
1	15 min at 95°C		
2-31	20 s at 95°C	45 s at 48°C	20 s at 72°

PCR products were purified away from reaction components using the QiaQuick Nucleotide Removal Kit (Qiagen), dried down, and resuspended in 20 µl diH_2O. Typical yield after purification was approximately 400-700ng/rxn by Nanodrop. Amplicons may be checked on a 3% agarose gel. Depending on quantity requirements of capture probes, additional PCR rounds were performed as above yielding approximately 200ng of sample per reaction. Amplicons were purified and characterized as above.

The final round of amplification of the capture probes was performed using asymmetric PCR. The protocol was as above, except that while the biotinylated primer concentration remained the same, the non-biotinylated primer concentration was reduced to 0.001x of the original concentration. The protocol was extended to 35 cycles to allow for non-exponential amplification. Amplicons were dried, resuspended in 20µl DIH_2O, and characterized.

The genomic DNA sample was prepared per standard protocol; 20 µg of WGA linkered sample was dried with 100 µg Cot-1 DNA and resuspended in 7.5 µl hybridization buffer and 3µl formamide. A 2 µg aliquot of capture probes was dried and resuspended in 4.5 µl diH_2O. The sample solution was mixed with the capture probe solution and incubated at 95°C for 10 minutes. The mixture was then transferred to a PCR tube and placed in a thermal cycler for 3 days at 42°C for hybridization to form duplexes.

After hybridization, the duplexes were bound to paramagnetic beads (Dynal). 25 µl of beads were washed three times in 2x BW buffer (10mM TrisHCl, 1mM EDTA, 2M NaCl), and the beads were resuspended in the hybridization mixture. Binding occurred over 45 minutes at 42°C with occasional gentle mixing.

Bound beads were isolated using a magnet and washed briefly with 40 µl Wash Buffer I, incubated for 2 x 5 minutes in 47°C stringent wash buffer, washed with Wash Buffer I for

IDT Exhibit 1020.5884

approximately 2 minutes at ambient room temperature, with Wash Buffer II for approximately 1 minute, and with Wash Buffer III for approximately 30 seconds.

To elute the captured fragments, the solution containing beads in Wash Buffer III was transferred to a 1.5 ml Eppendorf tube. The beads were isolated with a magnet. The wash buffer was removed and ~100ul of 95°C diH_2O is added. The solution was incubated at 95°C for 5 minutes, after which the beads were bound with a magnet and gently washed with 95°C diH_2O. The wash liquid was then removed and retained, and replaced with fresh 95°C diH_2O. Incubation and washing was repeated for a total of 3 times (15minutes, approximately 300 µl eluate). After the final wash, the Eppendorf tube containing eluate is placed on a magnetic stand for approximately 5 minutes to isolate any beads aspirated during elution. The solution was dried at high heat in a fresh Eppendorf tube. The eluted captured fragments were resuspended in 263µl diH_2O prior to standard LM-PCR.

Following LM-PCR, the captured fragments were subjected to standard ultra-deep sequencing using the 454 FLX platform, as above. Alternatively, LM-PCR can be avoided by ligating 454 sequencing adapter sequences to the pre-enrichment sample. In that case, the eluted enriched sequences can be piped directly into the emulsion PCR for ultra-deep sequencing.

Data indicated that 83.8% of the reads map back to target regions, which is comparable and indistinguishable from results obtained using array-based capture protocols.

Example 5-Solution phase capture using *in situ* amplification of capture probes

A standard microarray design was modified by addition of a terminal 15 mer primer sequence containing a MlyI (GAGTC(5/5)) recognition site. Incorporation of a MlyI site into the primer sequence facilitates enzymatic primer removal while leaving the capture oligonucleotide sequences intact. Arrays were synthesized via standard maskless array synthesis methods known to those skilled in the art.

Capture probes were amplified using *in situ* polymerase chain reaction (PCR) on an array in a thermal cycler using a sealing hybridization chamber (Grace Bio-Labs, Inc., Bend, OR) and Slide Griddle Adaptor (Bio-Rad Laboratories, Hercules, CA). PCR reaction

IDT Exhibit 1020.5885

constituents (25ul of 10X polymerase buffer, 1.25ul of 25mM dNTPs, 12.5ul each of 20uM primer 1a and 1b, 2.5ul of Hotstart Taq polymerase, 10ul 25mM $MgCl_2$ and 176.5ul diH_2O, total reaction volume of 250ul) were added to the microarray hybridization chambers and PCR was performed using the conditions; 100°C for 30s, 97°C for 15 min., 30 cycles of 100°C for 30s, 47.5°C for 45s, 78°C for 30s followed by cooling the reactions down to 1°C for 30s and 3.5°C to hold. Primer sequences were primer 1a 5'-TGCCGGAGTCAGCGT-3' (SEQ ID NO: 19) and primer 1b 5'-Biotin-AGTCAGAGTCGCCAC-3' (SEQ ID NO: 20), reflecting primer binding sites that were incorporated into the probe sequences.

Polymerase chain reaction capture probe amplicons were purified from the reaction components using the QIAquick® Nucleotide Removal Kit (Qiagen, Inc., Valencia, CA), dried down and resuspended in 20ul diH_2O. Amplification yield was roughly 5ug total as measured by NanoDrop® spectrophotometry (Thermo Fisher Scientific). Additional amplification rounds, following the protocol above, can be performed if additional amplicon quantity is needed (e.g., using the above protocol and 100ng sample per reaction).

The final round of amplification of the capture probes was performed using asymmetric PCR; 2.5ul 10X polymerase buffer, 0.125ul 25mM dNTPs, 0.0125ul of 20uM primer 1a, 1.25ul of 20uM primer 1b, 0.25ul Hotstart Taq, 1ul 25mM $MgCl_2$ and 18.86ul diH_2O (total reaction volume of 25ul). Amplicons were purified away from reaction components using the Qiagen MinElute™ columns and quantitated as previously described.

A genomic DNA sample was prepared per standard protocol. Twenty ug of the sample with linkers attached was dried with 100ug Cot-1 DNA and resuspended in 7.5 ul hybridization buffer (Roche NimbleGen, Madison, WI) and 3ul formamide. A 1ug aliquot of capture probes was dried and resuspended in 4.5ul diH_2O. The sample solution was incubated at 95°C for 10 min. to denature the DNA and added to the capture probe solution. The mixture was transferred to a PCR tube and placed in a thermal cycler at 42°C for 3 days to allow duplex formation to occur.

After hybridization, the duplexes were bound to streptavidin coated paramagnetic beads (Dynal®, Invitrogen, Carlsbad, CA). One hundred microliters of beads were washed three

IDT Exhibit 1020.5886

times with 2X BW buffer (10mM TrisHCl, 1mM EDTA, 2M NaCl) and resuspended in the hybridization duplex mixture. Binding between the beads and duplexes was allowed to occur over 45 min. at 42°C with occasional gentle mixing. Bound beads were isolated using a magnet and briefly washed in Wash Buffer I (0.2X SSC, 0.2% (v/v) SDS, 0.1mM DTT) at room temperature, followed by two washes (each wash for 5 min. at 47° C) in 200ul Stringent Wash Buffer (0.1M MES pH 6.65, 0.1M NaCl, 0.1% Tween 20), an additional wash in Wash Buffer I for 2 min. at room temperature, once with Wash Buffer II (0.2X SSC, 0.1mM DTT) for 1 min. at room temperature and finally for 30 sec. in Wash Buffer III (0.05X SSC, 0.1mM DTT) at room temperature.

Captured fragments were eluted from the beads. The washed bead solution in Wash Buffer III was transferred to a 1.5ml Eppendorf tube, the beads were isolated with a magnet, the wash buffer removed and replaced with 100ul 95°C diH_2O and the beads released from the magnet. The suspended beads were incubated at 95°C for 5 min. after which the beads were captured and gently washed with 95°C diH_2O to elute the captured fragments. The eluate was removed and the beads were washed again, for a total of three water washes; total of 10 min. with final volume of pooled eluate approximately 300 ul. After the final wash, residual magnetic beads were removed from the pooled eluate by additional magnetic capture and transfer of the eluate to a new tube. The solution was dried down and the captured, eluted fragments were resuspended in 263ul diH_2O in preparation for subsequent LM-PCR. Ligation was performed by established protocols known to those skilled in the art, using a ligator of sequence 5'-CTCGAGAATTCTGGATCC-3' (SEQ ID NO:21).

Following LM-PCR, the captured fragments were subjected to ultradeep sequencing using the 454 FLX platform (454 Life Sciences, Branford, CT). Alternatively, LM-PCR can be avoided by ligating 454 sequencing adaptor sequences to the pre-enrichment sample. In the latter case, the eluted enriched sequences can be added directly in the emulsion PCR of the 454 FLX platform workflow.

Figure 7 illustrates a resequencing experiment from fragments captured in solution using the methods as described above. The qPCR controls utilizing PCR control primer sequences indicate an average of 2600 fold enrichment across the four control loci.

qPCR control Primer Sequences:

IDT Exhibit 1020.5887

qPCR gSel-0210F	GACCCTCTTACCTTGGCATTCTC (SEQ ID NO: 22)
qPCR gSel-0210R	GCTGGTACCCATTGGCAACT (SEQ ID NO: 23)
qPCR gSel-0271F	GGAGTGAGTGGTTTTTCTTCATTTTT (SEQ ID NO: 24)
qPCR gSel-0271R	GCGCCACAAAGAGACATTCA (SEQ ID NO: 25)
qPCR gSel-0266F	AAGGCCATACTTGGGTGAACTG (SEQ ID NO: 26)
qPCR gSel-0266R	GCTCTGATTGGTGGCTTCGT (SEQ ID NO: 27)
qPCR gSel-0283F	TGCTTGCAGGTGTCTCTCAGA (SEQ ID NO: 28)
qPCR gSel-0283R	CAGTGAGATATTTGGTACCATGGTGTA (SEQ ID NO: 29)

Indeed, conformance wherein the percentage of what is expected upon resequencing to what is realized upon resequencing is approximately 100% for almost all regions resequenced.

All publications and patents mentioned in the present application are herein incorporated by reference. Various modification and variation of the described methods and compositions of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention that are obvious to those skilled in the relevant fields are intended to be within the scope of the following claims.

IDT Exhibit 1020.5888

List of References

Altschul, S.F. et al. (1990) J. Mol. Biol. 215,403-410

Hyman, E. D. (1988), Anal. Biochem. 174, 423-436

Lovett et al. (1991) PNAS USA, 88, 9628-9632

Morgulis, A. et al. (2006) Bioinformatics, 15, 134-41

Nuwaysir, E.F., et al., (2002) Genome Res. 12, 1749-1755

Rhonaghi et al. (1998), Science 281, 363-365

Soares, et al. (1994) PNAS, 91, 9228-9232

Singh-Gasson, S., et al. (1999) Nat. Biotechnol. 17, 974-978

Wetmur (1991) Critical Reviews in Biochemistry and Molecular Biology, 26(34):227-59

Wetmur et al. (1966) J. Mol. Biol., 31, 349-70

"Direct selection of cDNAs with large genomic DNA clones," in Molecular Cloning: A Laboratory Manual (eds. Sambrook, J. & Russell, D.W.) Chapter 11 Protocol 4, pages 11.98–11.106 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, USA, 2001)

EP 0 552 290

US 2005/0282209

US 5,143,854

US 6,013,440

US 6,375,903

WO 2004/070007

IDT Exhibit 1020.5889

CLAIMS

1. A method of reducing the genetic complexity of a population of nucleic acid molecules, the method comprising the steps of:

   (a) exposing fragmented, denatured nucleic acid molecules of said population to multiple, different oligonucleotide probes under hybridizing conditions followed by binding the complexes of hybridized molecules to a solid support to capture nucleic acid molecules that specifically hybridize to said probes,

   wherein said fragmented, denatured nucleic acid molecules have an average size of about 100 to about 1000 nucleotide residues, preferably about 250 to about 800 nucleotide residues and most preferably about 400 to about 600 nucleotide residues,

   (b) separating unbound and non-specifically hybridized nucleic acids from the captured molecules;

   (c) eluting the captured molecules from the solid support, and

   (d) optionally repeating steps (a) to (c) for at least one further cycle with the eluted captured molecules.

2. The method according to claim 1, wherein the multiple, different oligonucleotide probes each contain a chemical group or linker being able to bind to a solid support.

3. The method according to one of the claims 1 or 2, further comprising the step of ligating adaptor molecules to one or both, preferably both ends of the nucleic acid molecules prior or after the step (a).

4. The method according to claim 3, further comprising the step of amplification of said nucleic acid molecules with at least one primer, said primer comprising a sequence which specifically hybridizes to the sequence of said adaptor molecule(s).

IDT Exhibit 1020.5890

5. The method according to one of the claims 1-4, wherein said population of nucleic acid molecules is a population of genomic DNA molecules.

6. The method according to one of the claims 1-5, wherein said solid support is either a nucleic acid microarray or a population of beads.

7. A method for isolating and reducing the complexity of a plurality of nucleic acid sequences comprising:

a) providing:

i) a solid support wherein said solid support comprises hybridization probes hybridizable to target nucleic acid sequences,

ii) a nucleic acid sample comprising target nucleic acid sequences,

b) amplifying said hybridization probes wherein the amplification products comprise a binding moiety and wherein said amplification products are maintained in solution,

c) hybridizing said nucleic acid sample to said amplification products in solution such that hybridization between said amplification products and target nucleic acid sequences is allowed to occur,

d) separating the target nucleic acid/amplification product hybridization complexes from non-specifically hybridized nucleic acids by said binding moiety, and

e) eluting the hybridized target nucleic acid sequences from the complex thereby isolating and reducing the complexity of a plurality of nucleic acid sequences.

8. The method according to claim 7, further comprising sequencing the eluted target nucleic acid sequences.

IDT Exhibit 1020.5891

9. The method according to any of claims 7 or 8, wherein said solid support is a microarray slide.

10. The method according to any one of claims 7 to 9, wherein said nucleic acid sample is a fragmented genomic DNA sample.

11. The method according to claim 10, wherein said fragmented genomic DNA sample further comprises adaptor molecules at one or both ends of the fragments of genomic DNA sample.

12. The method according to any of claims 7 to 11, wherein said hybridization probes further comprise primer binding sequences at one or both ends of said probes.

13. The method according to claim 12, wherein said primer binding sequences when present at both ends of the probes are the same.

14. The method according to claim 13, wherein said primer binding sequences when present at both ends of the probes are different.

15. The method according to any of claims 7 to 14, wherein said amplifying comprises exponential polymerase chain reaction.

16. The method according claim 15, wherein said amplifying further comprises asymmetric polymerase chain reaction.

17. The method according to any of claims 7 to 16, wherein said binding moiety is a biotin binding moiety.

IDT Exhibit 1020.5892

18. The method according to claim 17, wherein said separating comprises binding said biotin binding moiety to a streptavidin coated substrate.

19. The method according to claim 18, wherein said streptavidin coated substrate is a streptavidin coated paramagnetic particle.

20. The method according to any of claims 7 to 19, further comprising washing said separated target nucleic acid/amplification product hybridization complexes prior to elution.

21. The method according to any one of the claims 1-20, wherein said population of nucleic acid molecules or said plurality of nucleic acid sequences contains the whole genome or at least one chromosome of an organism or at least one nucleic acid molecule with a size of at least about 200 kb, at least about 500 kb, at least about 1 Mb, at least about 2 Mb or at least about 5 Mb, especially a size between about 100 kb and about 5 Mb, between about 200 kb and about 5 Mb, between about 500 kb and about 5 Mb, between about 1 Mb and about 2 Mb or between 2 Mb and about 5 Mb.

22. The method according to any of claims 5, 6 or 10-21, wherein said probes are selected from:

- a plurality of probes that defines a plurality of exons, introns or regulatory sequences from a plurality of genetic loci,
- a plurality of probes that defines the complete sequence of at least one single genetic locus, said locus having a size of at least 100 kb, preferably at least 1 Mb, or at least one of the sizes as specified in claim 3,
- a plurality of probes that defines sites known to contain single nucleotide polymorphisms (SNPs), or
- a plurality of probes that defines an array, in particular a tiling array, designed to capture the complete sequence of at least one complete chromosome.

IDT Exhibit 1020.5893

23. A method for determining nucleic acid sequence information of about at least one region of nucleic acid(s), in particular genomic nucleic acid(s), the method comprising the steps of:

1. reducing the genetic complexity of a population of nucleic acid molecules according to a method according to one of the claims 1-22, and
2. determining the nucleic acid sequence of the captured molecules, in particular by means of performing sequencing by synthesis reactions.

24. A method for detecting coding region variation relative to a reference genome, the method comprising the steps of:

1. reducing the genetic complexity of a population of nucleic acid molecules according to a method according to one of the claims 1-22,
2. determining the nucleic acid sequence of the captured molecules, and
3. comparing the determined sequence to sequences in a database of the reference genome, in particular to sequences in a database of polymorphisms in the reference genome to identify variants from the reference genome.

25. A kit comprising

- double stranded adaptor molecules, and
- a solid support with multiple, different oligonucleotide probes, wherein said probes are selected from:
- a plurality of probes that define a plurality of exons, introns or regulatory sequences from a plurality of genetic loci,
- a plurality of probes that define the complete sequence of at least one single genetic locus, said locus having a size of at least 100 kb, preferably at least 1 Mb, or at least one of the sizes as specified in claim 21,
- a plurality of probes that define sites known to contain SNPs, or

IDT Exhibit 1020.5894

- a plurality of probes that define a tiling array designed to capture the complete sequence of at least one complete chromosome.

26. The kit according to claim 25, wherein the kit contains two different double stranded adaptor molecules.

27. The kit according to claim 25 or 26, wherein said solid support is either a plurality of beads or a microarray.

28. The kit according to one of the claims 25-27, further comprising at least one or more components selected from DNA polymerase, T4 polynucleotide kinase, T4 DNA ligase, an array hybridization solution, an array wash solution, and/or an array elution solution.

29. A kit comprising:

a) hybridization probe sequences comprising a binding moiety wherein said probe sequences are designed to hybridize to one or more target nucleic acid sequences and wherein said probe sequences are in solution,

b) a substrate comprising a binding partner for binding said binding moiety, and

c) instructions for performing the methods of claim 7.

30. The kit according to claim 29, further comprising one or more of a hybridization solution, a wash solution and an elution solution.

31. The kit according to claim 30, further comprising a magnet.

IDT Exhibit 1020.5895

Fig. 1

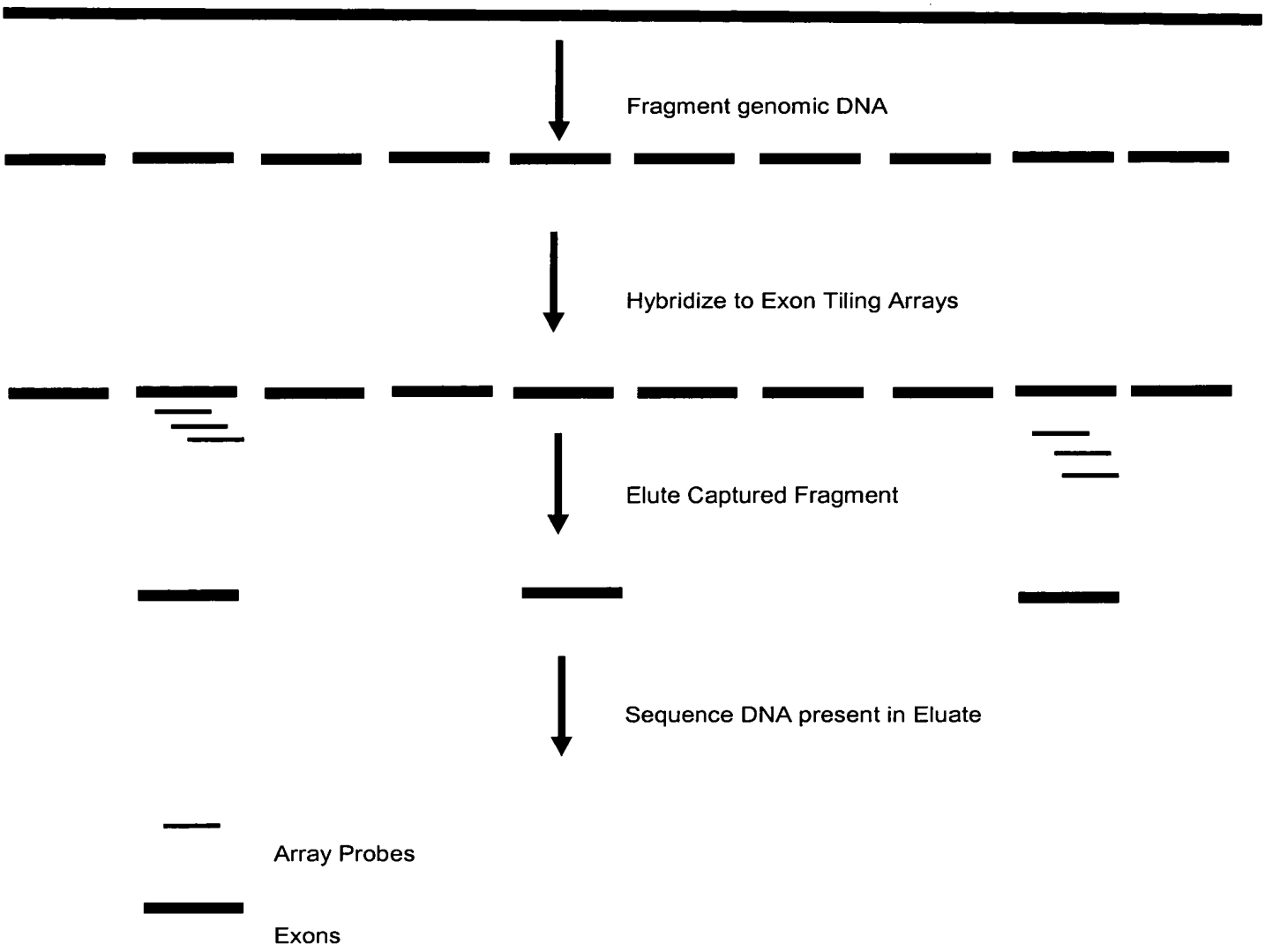

Fragment genomic DNA
Hybridize to Exon Tiling Arrays
Elute Captured Fragment
Sequence DNA present in Eluate
Array Probes
Exons

Fig. 2

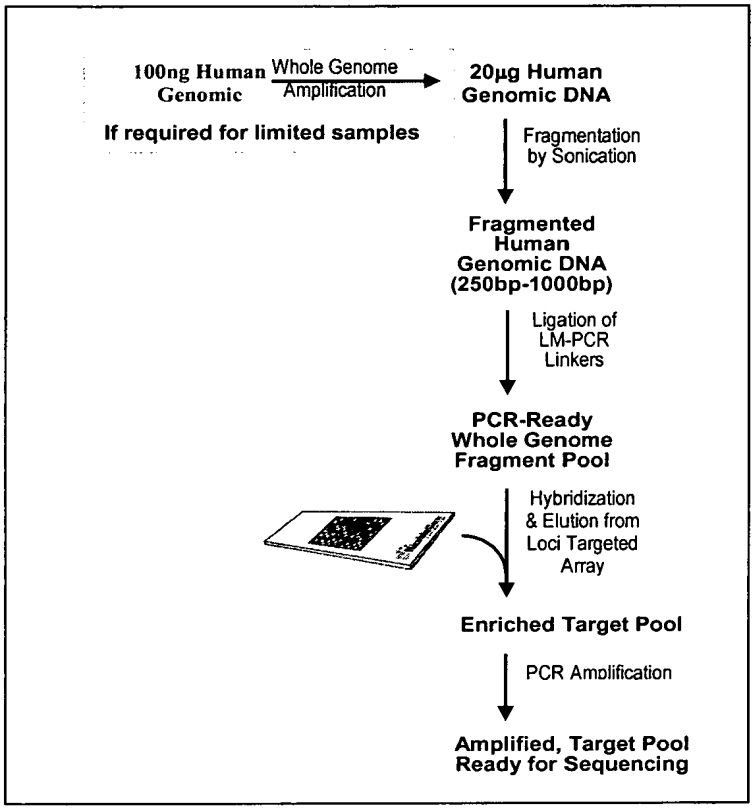
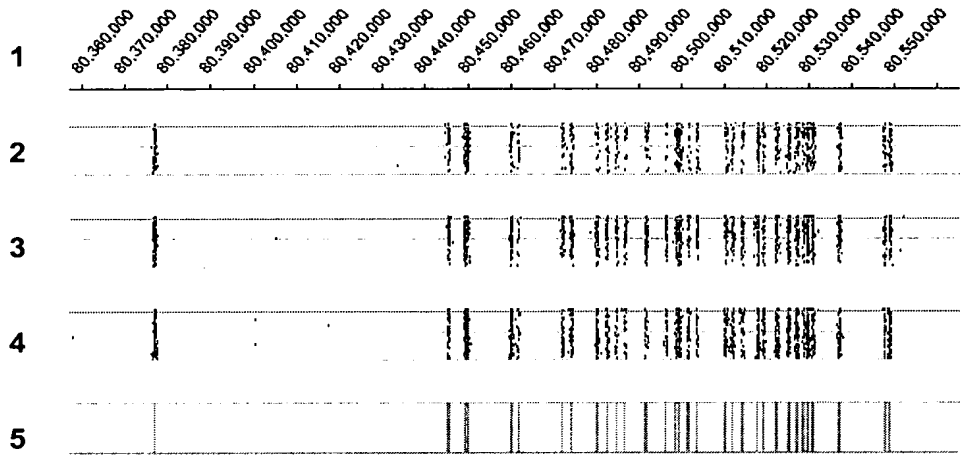

100ng Human Genomic
Whole Genome Amplification
20µg Human Genomic DNA
If required for limited samples
Fragmentation by Sonication
Fragmented Human Genomic DNA (250bp-1000bp)
Ligation of LM-PCR Linkers
PCR-Ready Whole Genome Fragment Pool
Hybridization & Elution from Loci Targeted Array
Enriched Target Pool
PCR Amplification
Amplified, Target Pool Ready for Sequencing

Fig. 3 (a-b)

a)


1
2
3
4
5
80,360,000
80,370,000
80,380,000
80,390,000
80,400,000
80,410,000
80,420,000
80,430,000
80,440,000
80,450,000
80,460,000
80,470,000
80,480,000
80,490,000
80,500,000
80,510,000
80,520,000
80,530,000
80,540,000
80,550,000

b)

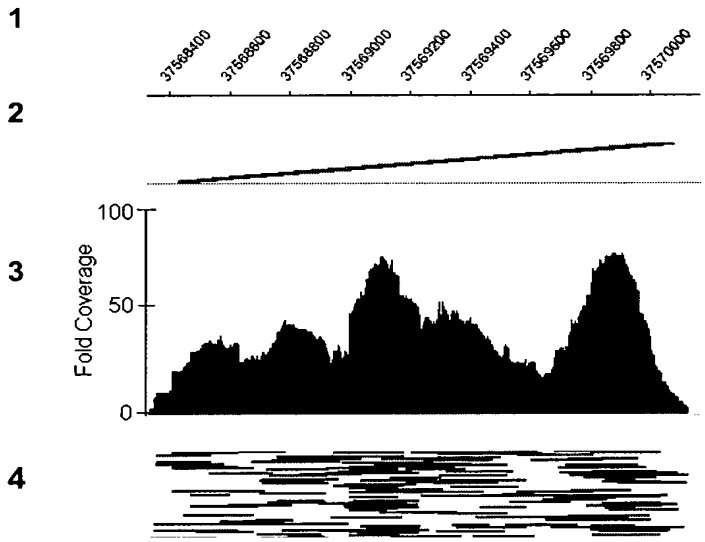
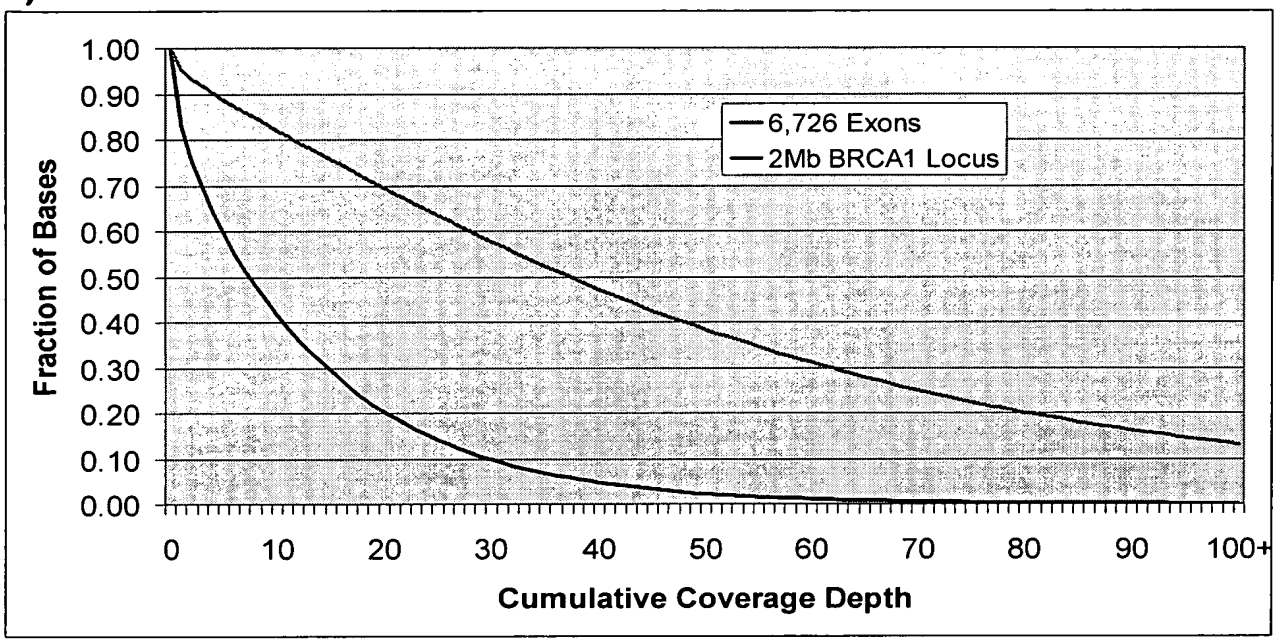
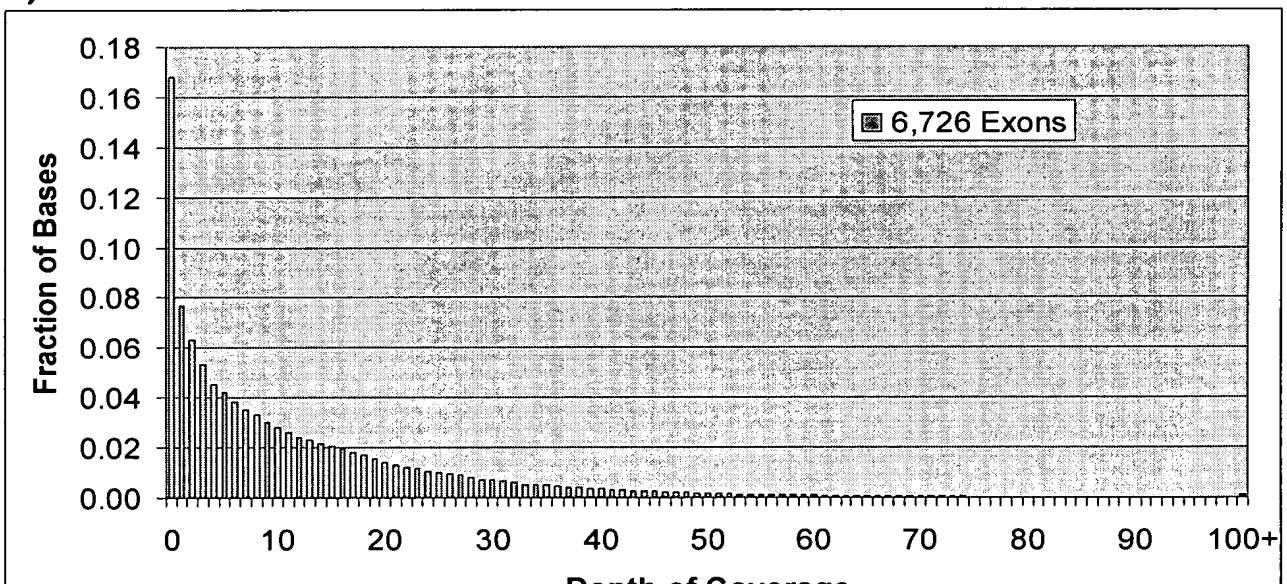

1
2
3
4
37568400
37568600
37568800
37569000
37569200
37569400
37569600
37569800
37570000
100
50
0
Fold Coverage

Fig. 4 (a-c)

a)


1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
Fraction of Bases
6,726 Exons
2Mb BRCA1 Locus
0
10
20
30
40
50
60
70
80
90
100+
Cumulative Coverage Depth

b)


0.18
0.16
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
Fraction of Bases
6,726 Exons
0
10
20
30
40
50
60
70
80
90
100+

c)

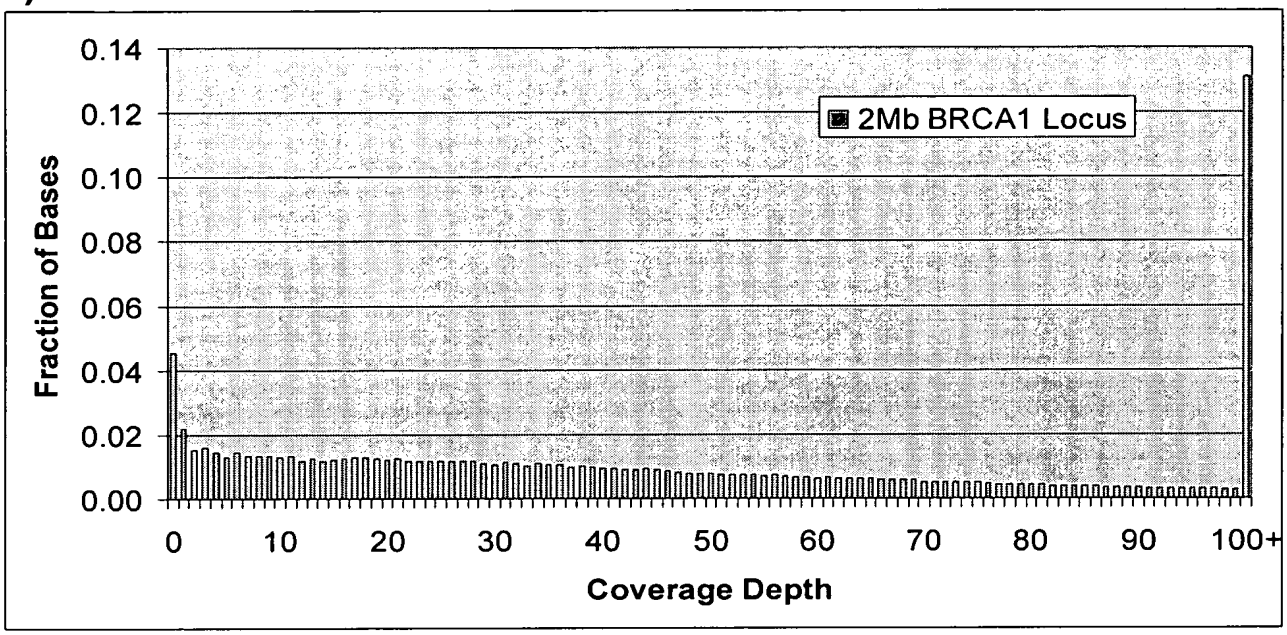

0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
Fraction of Bases
2Mb BRCA1 Locus
0
10
20
30
40
50
60
70
80
90
100+
Coverage Depth

Fig. 5:

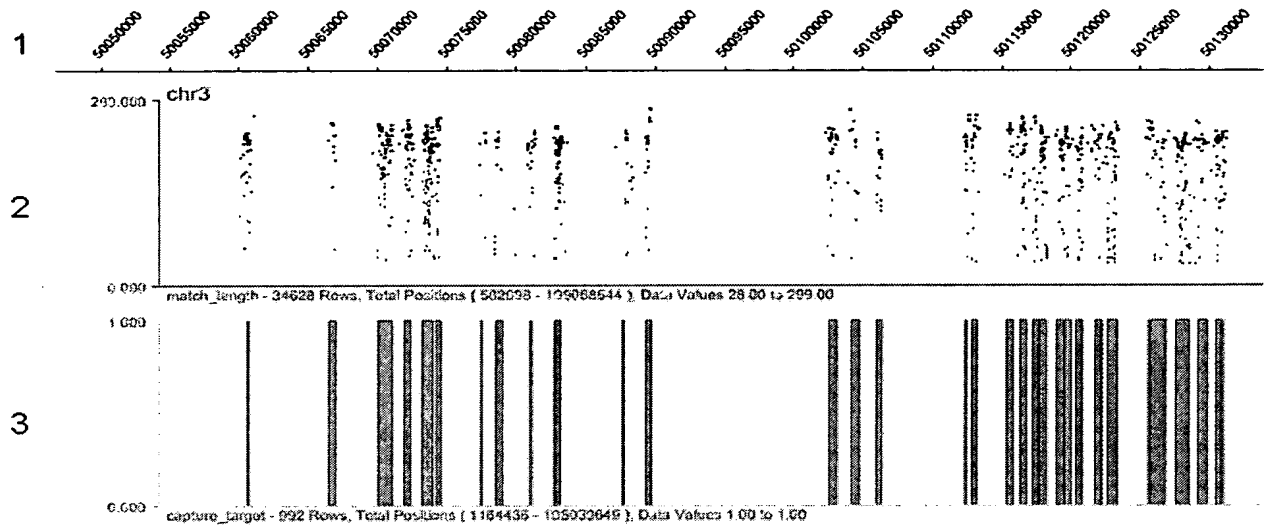

1
2
3
chr3
200.000
0.000
match_length - 34628 Rows, Total Positions (502098 - 105068544), Data Values 28.00 to 299.00
1.000
0.000
capture_target - 992 Rows, Total Positions (1164436 - 105032649), Data Values 1.00 to 1.00

FIG. 6

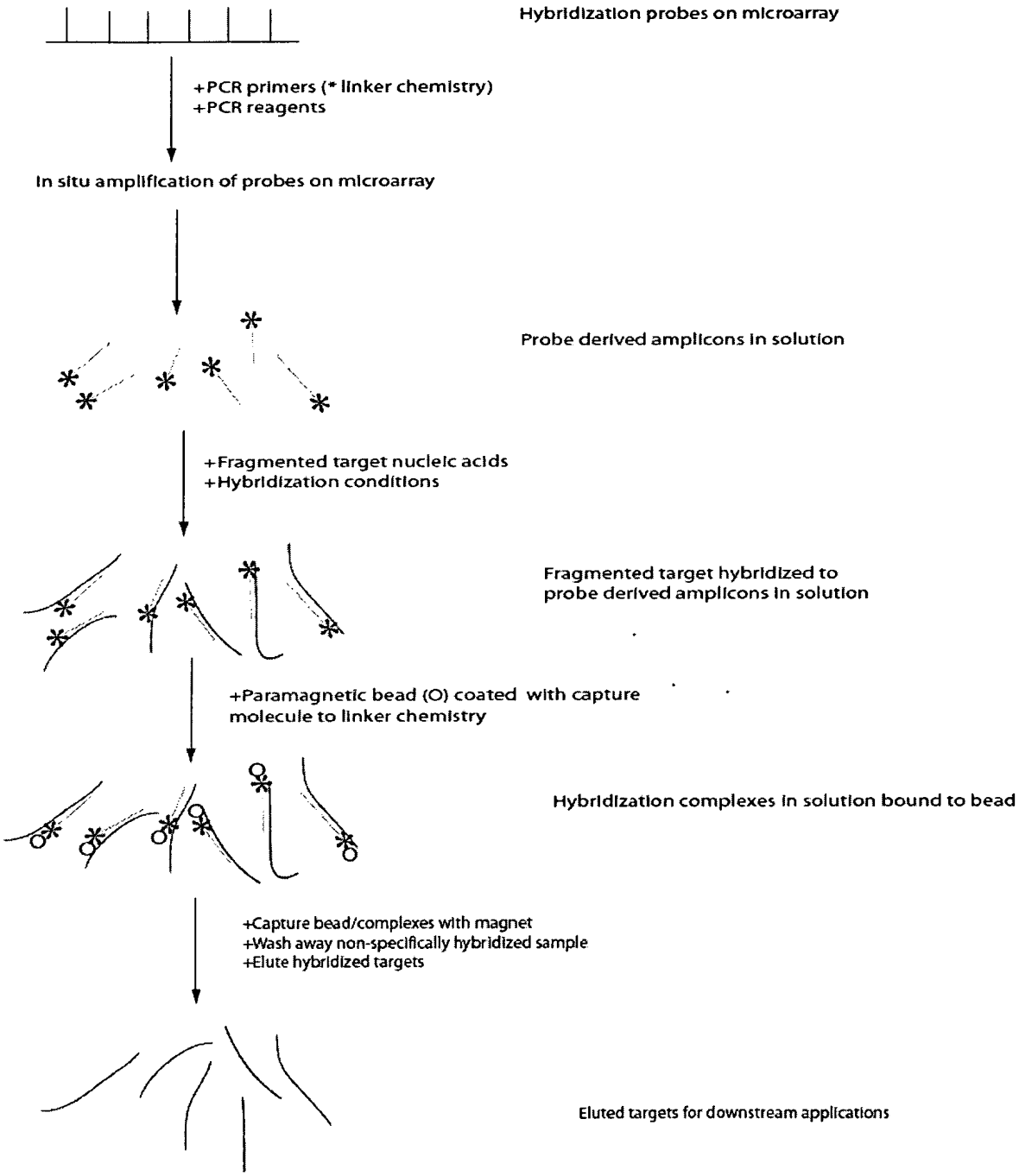
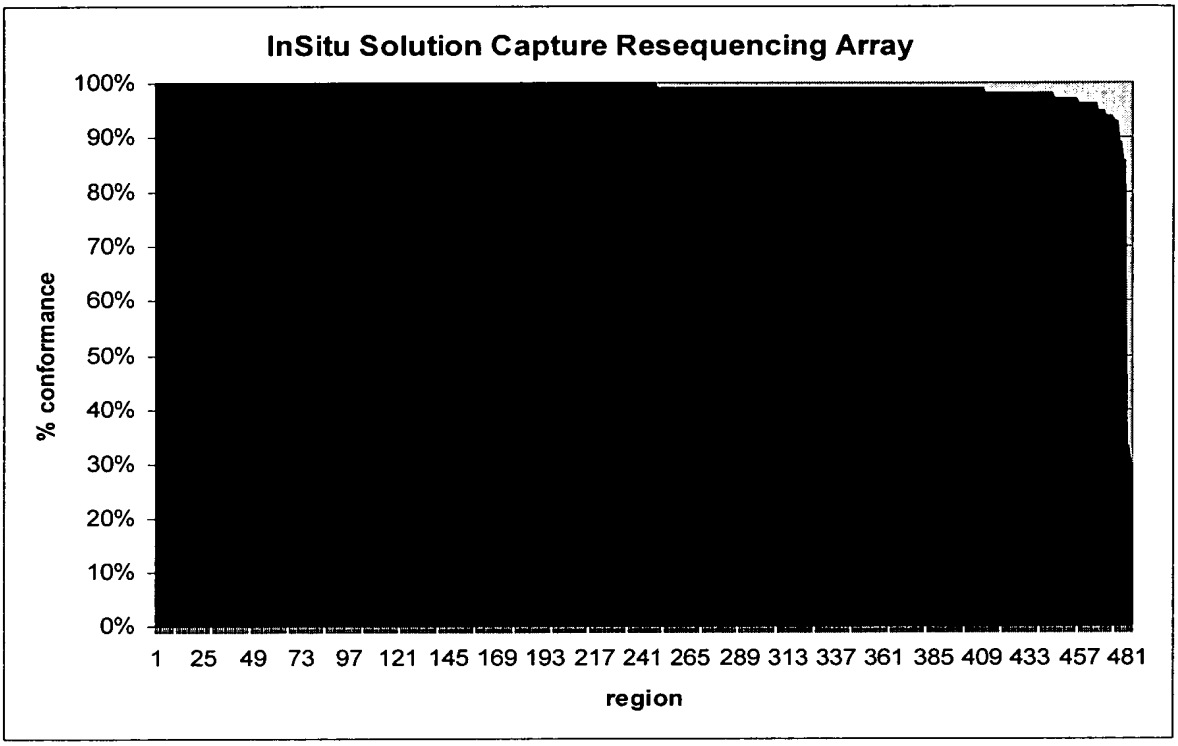

Hybridization probes on microarray
+PCR primers (* linker chemistry)
+PCR reagents
In situ amplification of probes on microarray
Probe derived amplicons in solution
+Fragmented target nucleic acids
+Hybridization conditions
Fragmented target hybridized to
probe derived amplicons in solution
+Paramagnetic bead (O) coated with capture
molecule to linker chemistry
Hybridization complexes in solution bound to bead
+Capture bead/complexes with magnet
+Wash away non-specifically hybridized sample
+Elute hybridized targets
Eluted targets for downstream applications

FIG. 7


InSitu Solution Capture Resequencing Array
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
% conformance
1 25 49 73 97 121 145 169 193 217 241 265 289 313 337 361 385 409 433 457 481
region

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2008/008920

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12Q1/68

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

B. FIELDS SEARCHED

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

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

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

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	EP 1 184 466 A (AFFYMETRIX INC [US]) 6 March 2002 (2002-03-06) figure 4 paragraphs [0022], [0029], [0031], [0033], [0037], [0044], [0049] - [0051], [0060]	1-31

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

☒ See patent family annex.

* Special categories of cited documents :

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

"E" earlier 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
20 January 2009	09/02/2009

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

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

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/EP2008/008920

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 1184466 A	06-03-2002	CA 2356396 A1	26-02-2002
		JP 2002330783 A	19-11-2002
		US 2003148273 A1	07-08-2003

IDT Exhibit 1020.5905

Espacenet

Bibliographic data: WO2011156529 (A2) — 2011-12-15

METHODS AND COMPOSITION FOR MULTIPLEX SEQUENCING

Inventor(s): RAYMOND CHRISTOPHER [US]; KURN NURITH [US]; MAGNUS JILL [US] + (RAYMOND, CHRISTOPHER, ; KURN, NURITH, ; MAGNUS, JILL)

Applicant(s): NUGEN TECHNOLOGIES INC [US]; RAYMOND CHRISTOPHER [US]; KURN NURITH [US]; MAGNUS JILL [US] + (NUGEN TECHNOLOGIES, INC, ; RAYMOND, CHRISTOPHER, ; KURN, NURITH, ; MAGNUS, JILL)

Classification: **- international:** ***C07H21/02; C40B30/00; C40B30/04; C40B40/06***
- cooperative: **C12Q1/6855**; **C12Q1/6869** more

Application number: WO2011US39683 20110608 Global Dossier

Priority number(s): US20100352801P 20100608

Also published as: WO2011156529 (A3) CN103119439 (A) EP2580378 (A2) EP2580378 (A4) US2011319290 (A1)

Abstract of WO2011156529 (A2)

Adapters are joined to target polynucleotides to create adapter-tagged polynucleotides. Adapter- tagged polynucleotides are sequenced simultaneously and sample sources are identified on the basis of barcode sequences.

IDT Exhibit 1020.5906

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
15 December 2011 (15.12.2011)

PCT

(10) International Publication Number
WO 2011/156529 A2

(51) International Patent Classification:
C40B 40/06 (2006.01) *C40B 30/04* (2006.01)
C40B 30/00 (2006.01) *C07H 21/02* (2006.01)

(21) International Application Number:
PCT/US2011/039683

(22) International Filing Date:
8 June 2011 (08.06.2011)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/352,801 8 June 2010 (08.06.2010) US

(71) Applicant *(for all designated States except US)*: **NUGEN TECHNOLOGIES, INC.** [US/US]; 201 Industrial Road, Suite 310, San Carlos, CA 94070 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **RAYMOND, Christopher** [US/US]; 2626 Ne 86th Street, Seattle, WA 98115 (US). **KURN, Nurith** [IL/US]; 2876 Ramona Street, Palo Alto, CA 94306 (US). **MAGNUS, Jill** [US/US]; 11528 39th Avenue Northeast, Seattle, WA 98125 (US).

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

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

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

Published:

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

— *with sequence listing part of description (Rule 5.2(a))*

WO 2011/156529 A2

(54) **Title**: METHODS AND COMPOSITION FOR MULTIPLEX SEQUENCING

(57) **Abstract**: Adapters are joined to target polynucleotides to create adapter-tagged polynucleotides. Adapter- tagged polynucleotides are sequenced simultaneously and sample sources are identified on the basis of barcode sequences.

IDT Exhibit 1020.5907

METHODS AND COMPOSITIONS FOR MULTIPLEX SEQUENCING

CROSS-REFERENCE

[0001] This application claims the benefit of U.S. Provisional Application No. 61/352,801, filed June 8, 2010, which application is incorporated herein by reference.

SEQUENCE LISTING

[0002] The instant application contains a Sequence Listing which has been submitted in ASCII format via EFS-Web and is hereby incorporated by reference in its entirety. Said ASCII copy, created on June 8, 2011, is named 25115-741-201.txt and is 21 Kilobytes in size.

BACKGROUND OF THE INVENTION

[0003] Large-scale sequence analysis of DNA can provide understanding of a wide range of biological phenomena related to states of health and disease, both in humans and in many economically important plants and animals, see e.g. Collins et al (2003), Nature, 422: 835-847; Service, Science, 311: 1544-1546 (2006); Hirschhorn et al (2005), Nature Reviews Genetics, 6: 95-108; National Cancer Institute, Report of Working Group on Biomedical Technology, "Recommendation for a Human Cancer Genome Project," (February, 2005); Tringe et al (2005), Nature Reviews Genetics, 6: 805-814. The need for low-cost high-throughput sequencing and re-sequencing has led to the development of several new approaches that employ parallel analysis of many target DNA fragments simultaneously, e.g. Margulies et al, Nature, 437: 376-380 (2005); Shendure et al (2005), Science, 309: 1728-1732; Metzker (2005), Genome Research, 15: 1767-1776; Shendure et al (2004), Nature Reviews Genetics, 5: 335-344; Lapidus et al, U.S. patent publication US 2006/0024711; Drmanac et al, U.S. patent publication US 2005/0191656; Brenner et al, Nature Biotechnology, 18: 630-634 (2000); and the like. Such approaches reflect a variety of solutions for increasing target polynucleotide density and for obtaining increasing amounts of sequence information within each cycle of a particular sequence detection chemistry.

[0004] Given the complexity of the mixture of sequences in a given reactions, sequencing is typically restricted to one sample per reaction chamber. However, the number of bases read in a given reaction using these next generation sequencing technologies can be far greater than that actually needed to acquire the sequence information of interest, which essentially amounts to wasted sequencing space. Coupled with increasing desires to sequence samples from multiple sources, the expense of utilizing these technologies can quickly become prohibitive. Sequencing runs are also often limited in the number of separate reactions that can be run in parallel, which places further restrictions on the efficiency with which large numbers of samples can be processed.

[0005] Some approaches to resolve these challenges have involved the incorporation of additional identifier sequences into each of the target fragments analyzed. Where different sequences are used for different samples, sequencing of pooled samples can be followed by resolution of sequences into subsets corresponding to sample sources based on the added sequences. However, addition of sequences to

IDT Exhibit 1020.5908

resolve sample sources faces two challenges. Firstly, random errors in sequencing can make it impossible to correctly identify an appended identifier sequence with its sample source when such errors occur within appended sequences that are either too short or insufficiently dissimilar from sequences corresponding to other samples. Secondly, the addition of longer sequences to allow for such sequencing error takes up valuable sequencing space from target reads that can be as short as 20 bases. In view of these limitations, there is a need to increase the efficiency of next generation sequencing technologies such that samples can be sequenced in greater numbers, with greater identification accuracy, while maximizing the available sequencing space.

SUMMARY OF THE INVENTION

[0006] In one aspect, the invention provides methods, compositions, and kits for multiplex sequencing. In one embodiment, the method comprises sequencing a plurality of target polynucleotides in a single reaction chamber, wherein said target polynucleotides are from two or more different samples; and identifying the sample from which each of said sequenced target polynucleotides is derived with an accuracy of at least 95% based on a single barcode contained in the sequence of said target polynucleotide. In some embodiments, the target polynucleotides comprise one or more sequences with which the sequencing reaction is calibrated. In some embodiments, each barcode differs from every other barcode at at least three nucleotide positions. In some embodiments, the identification of sample source is accurate after the mutation or deletion of a nucleotide in the barcode.

[0007] In another aspect, the invention provides methods, compositions, and kits for producing adapter-tagged target polynucleotides from a plurality of independent samples. In one embodiment, the method comprises: (a) providing a plurality of first adapter oligonucleotides, wherein each of said first adapter 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; and (b) joining at least one of said first adapter oligonucleotides to said target polynucleotides of each of said samples, such that no barcode sequence is joined to said target polynucleotides of more than one of said samples. In some embodiments, the method further comprises (c) joining at least one of a plurality of second adapter oligonucleotides to said target polynucleotides of each of said samples from step (b), such that at least some of said target polynucleotides comprise said first adapter oligonucleotide at one end and said second adapter oligonucleotide at the other end. One or more of the adapter oligonucleotides of the present invention can comprise SEQ ID NO: 1. One or more of the adapter oligonucleotides of the present invention can comprise SEQ ID NO: 2. One or more of the adapter oligonucleotides can comprise a hairpin structure. One or more of the adapter oligonucleotides can comprise an oligonucleotide duplex.

[0008] In some embodiments, the barcode sequences are at least 3 nucleotides in length. In some embodiments, the plurality of barcode sequences includes sequences selected from the group consisting of: AAA, TTT, CCC, and GGG. In some embodiments, the plurality of barcode sequences includes sequences selected from the group consisting of: AAAA, CTGC, GCTG, TGCT, ACCC, CGTA, GAGT,

IDT Exhibit 1020.5909

TTAG, AGGG, CCAT, GTCA, TATC, ATTT, CACG, GGAC, and TCGA. In some embodiments, the plurality of barcode sequences includes sequences selected from the group consisting of: AAAAA, AACCC, AAGGG, AATTT, ACACG, ACCAT, ACGTA, ACTGC, AGAGT, AGCTG, AGGAC, AGTCA, ATATC, ATCGA, ATGCT, ATTAG, CAACT, CACAG, CAGTC, CATGA, CCAAC, CCCCA, CCGGT, CCTTG, CGATA, CGCGC, CGGCG, CGTAT, CTAGG, CTCTT, CTGAA, CTTCC, GAAGC, GACTA, GAGAT, GATCG, GCATT, GCCGG, GCGCC, GCTAA, GGAAG, GGCCT, GGGGA, GGTTC, GTACA, GTCAC, GTGTG, GTTTT, TAATG, TACGT, TAGCA, TATAC, TCAGA, TCCTC, TCGAG, TCTCT, TGACC, TGCAA, TGGTT, TGTGG, TTAAT, TTCCG, TTGGC, and TTTTA.

[0009] In some embodiments, the method further comprises pooling the target polynucleotides from step (c). Target polynucleotides can be pooled based on the barcode sequences to which they are joined, such that all four bases are evenly represented at one or more positions along each barcode in the pool.

[0010] In some embodiments, target polynucleotides comprise fragmented sample polynucleotides. Fragmentation can comprise subjecting sample polynucleotides to acoustic sonication, and/or treating sample polynucleotides with one or more enzymes under conditions suitable for the one or more enzymes to generate random double-stranded nucleic acid breaks (which can include DNase I, Fragmentase, and variants thereof). In some embodiments, fragmentation comprises treating the sample polynucleotides with one or more restriction endonucleases. Fragments can have an average length of 10 to 10,000 nucleotides, such as an average length of 100-2,500 nucleotides, or 50-500 nucleotides. In some embodiments, samples comprise less than 500ng of nucleic acid. Target polynucleotides can comprise genomic DNA, DNA produced by a primer extension reaction, cDNA, mitochondrial DNA, chloroplast DNA, plasmid DNA, bacterial artificial chromosomes, yeast artificial chromosomes, or a combination thereof.

[0011] In some embodiments, the method further comprises performing the step of extending one or more 3' ends of the target polynucleotides, using the one or more joined adapter oligonucleotides as template. In some embodiments, the method further comprises amplifying the target polynucleotides after the extending step using a first primer and a second primer, wherein the first primer comprises a sequence that is hybridizable to at least a portion of the complement of one or more of the first adapter oligonucleotides, and further wherein the second primer comprises a sequence that is hybridizable to at least a portion of the complement of one or more of the second adapter oligonucleotides. One or more of the primers used in the amplification step can comprise SEQ ID NO: 1. One or more of the primers used in the amplification step can comprise SEQ ID NO: 2.

[0012] In some embodiments, each second adapter oligonucleotide 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 the plurality of barcode sequences at at least three nucleotide positions. Pairs of first and second adapter oligonucleotides can comprise the same or different barcode sequences.

IDT Exhibit 1020.5910

[0013] In some embodiments, the method further comprises sequencing one or more of the polynucleotides in a pool of target polynucleotides from independent samples. Sequencing can comprise extension of a sequencing primer comprising a sequence hybridizable to at least a portion of the complement of the first adapter oligonucleotide and/or the second adapter oligonucleotide. In some embodiments, the sequencing primer comprises SEQ ID NO: 1 or SEQ ID NO: 2. In some embodiments, sequencing comprises a calibration step, wherein calibration is based on each of the nucleotides at one or more nucleotide positions in the barcode sequences.

[0014] In some embodiments, the method further comprises identifying the sample from which a target polynucleotide is derived based on a barcode sequence to which it is joined.

[0015] In another aspect, the invention provides compositions for use in the described methods, comprising any one or more of the elements described herein. In one aspect, the invention provides a composition for multiplex sequencing. 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.

[0016] In another aspect, the invention provides a composition useful in the generation of adapter-tagged target polynucleotides, comprising any one or more of the elements described herein. In one embodiment, the composition comprises a plurality of first adapter oligonucleotides, wherein each of said first adapter 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. In some embodiments, the composition further comprises a plurality of second adapter oligonucleotides. In some embodiments, target polynucleotides are contained in a flow cell. First adapter oligonucleotides can be grouped in multiples of four such that all four bases are evenly represented at each position along each barcode. Where the second adapter oligonucleotide comprises a barcode, pairs of first and second adapter oligonucleotides can comprise the same or different barcode sequences. In some embodiments, the composition further comprises a first primer and a second primer, wherein said first primer comprises a sequence that is hybridizable to at least a portion of the complement of one or more of said first adapter oligonucleotides, and further wherein said second primers comprise a sequence that is hybridizable to at least a portion of the complement of one or more of said second adapter oligonucleotides. In some embodiments, the composition additionally comprises a sequencing primer comprising a sequence hybridizable to at least a portion of the complement of said first adapter oligonucleotide and/or said second adapter oligonucleotide.

[0017] In some embodiments, the composition comprises a plurality of first adapter oligonucleotides, wherein each of said first adapter oligonucleotides comprises a 5' end comprising sequence A and a 3'

IDT Exhibit 1020.5911

end comprising sequence A', and further wherein A is hybridizable to A', one of A or A' comprises DNA, and the other of A or A' comprises RNA and 5 or more terminal DNA nucleotides. In some embodiments, the composition further comprises a plurality of second adapter oligonucleotides, wherein each of said second adapter oligonucleotides comprises a 5' end comprising sequence B and a 3' end comprising sequence B', and further wherein B is hybridizable to B', one of B or B' comprises DNA, and the other of B or B' comprises RNA and 5 or more terminal DNA nucleotides.

[0018] In another aspect, the invention provides kits containing any one or more of the elements disclosed in the described methods and compositions. In one aspect, the invention provides a kit useful in the generation of adapter-tagged target polynucleotides. In one embodiment, the kit comprises a plurality of first adapter oligonucleotides, wherein each of said first adapter 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, and instructions for using the same. In some embodiments, the kit further comprises a plurality of second adapter oligonucleotides. In some embodiments, the kit further comprises a first primer and a second primer, wherein said first primer comprises a sequence that is hybridizable to at least a portion of the complement of one or more of said first adapter oligonucleotides, and further wherein said second primers comprise a sequence that is hybridizable to at least a portion of the complement of one or more of said second adapter oligonucleotides. In some embodiments, the kit additionally comprises a sequencing primer comprising a sequence hybridizable to at least a portion of the complement of said first adapter oligonucleotide and/or said second adapter oligonucleotide. In some embodiments, the kit further comprises one or more of: (a) a DNA ligase, (b) a DNA-dependent DNA polymerase, (c) an RNA-dependent DNA polymerase, (d) random primers, (e) primers comprising at least 4 thymidines at the 3' end, (f) a DNA endonuclease, (g) a DNA-dependent DNA polymerase having 3' to 5' exonuclease activity, (h) a plurality of primers, each primer having one of a plurality of selected sequences, (i) a DNA kinase, (j) a DNA exonuclease, (k) magnetic beads, (l) an enzyme comprising RNase H activity, (m) an RNA ligase, and (n) one or more buffers suitable for one or more of the elements contained in said kit.

[0019] In some embodiments, the kit comprises a plurality of first adapter oligonucleotides, wherein each of said first adapter oligonucleotides comprises a 5' end comprising sequence A and a 3' end comprising sequence A', and further wherein A is hybridizable to A', one of A or A' comprises DNA, and the other of A or A' comprises RNA and 5 or more terminal DNA nucleotides. In some embodiments, the kit further comprises a plurality of second adapter oligonucleotides, wherein each of said second adapter oligonucleotides comprises a 5' end comprising sequence B and a 3' end comprising sequence B', and further wherein B is hybridizable to B', one of B or B' comprises DNA, and the other of B or B' comprises RNA and 5 or more terminal DNA nucleotides.

[0020] In another aspect, the invention provides a method of producing adapter-tagged polynucleotides. In one embodiment, the method comprises: (a) providing a plurality of first adapter oligonucleotides,

IDT Exhibit 1020.5912

wherein each of said first adapter oligonucleotides comprises a 5' end comprising sequence A and a 3' end comprising sequence A', and further wherein A is hybridizable to A', one of A or A' comprises DNA, and the other of A or A' comprises RNA and 5 or more terminal DNA nucleotides; and, (b) joining at least one of said first adapter oligonucleotides to at least one of said target polynucleotides. Each of said first adapter oligonucleotides may comprise a barcode sequence. In some embodiments, the method further comprises the step of cleaving RNA with an enzyme that cleaves RNA from an RNA-DNA heteroduplex. In some embodiments, the method further comprises the step of extending one or more 3' ends of said target polynucleotides, using said one or more joined adapter oligonucleotides as template. In some embodiments, the method comprises joining at least one of a plurality of second adapter oligonucleotides to said target polynucleotides of each of said samples from step (b), such that at least one of said target polynucleotides comprises said first adapter oligonucleotide at one end and said second adapter oligonucleotide at the other end. In some embodiments, each of said second adapter oligonucleotides comprises a 5' end comprising sequence B and a 3' end comprising sequence B', and further wherein B is hybridizable to B', one of B or B' comprises DNA, and the other of B or B' comprises RNA and 5 or more terminal DNA nucleotides. In some embodiments, each of said second adapter oligonucleotides comprises a barcode sequence.

INCORPORATION BY REFERENCE

[0021] 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

[0022] 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:

[0023] Figure 1 shows a schematic drawing of one embodiment of the methods of the invention.

[0024] Figure 2A shows an example result of amplification products obtained for target polynucleotides joined to adapter oligonucleotides, also referred to as "adapters," according to methods of the invention.

[0025] Figure 2B shows a side by side comparison of selected lanes from Fig. 2A, along with details about elements contained in the ligation reaction.

[0026] Figure 3 shows a schematic drawing of one embodiment of the methods of the invention, with hairpin adapters comprising RNA at the 5' end.

[0027] Figure 4 shows a schematic drawing of one embodiment of the methods of the invention, with hairpin adapters comprising RNA at the 3' end.

IDT Exhibit 1020.5913

[0028] Figure 5 shows a schematic drawing of one embodiment of the methods of the invention, with hairpin adapters comprising RNA at the 3' end that are joined to a target polynucleotide, and further addition on non-hairpin adapters to ends of the target polynucleotide not joined to the hairpin adapter.

[0029] Figure 6 shows a schematic drawing of one embodiment of the methods of the invention.

[0030] Figure 7 shows various adapter designs, evaluated ligation efficiencies, and PCR amplified ligation products analyzed on an agarose gel.

[0031] Figure 8 shows an agarose gel containing target polynucleotides, adapter oligonucleotides, and ligation products.

[0032] Figure 9 shows an agarose gel containing PCR amplified ligation products.

[0033] Figure 10 shows a schematic drawing of one embodiment of the methods of the invention.

DEFINITIONS

[0034] The terms "polynucleotide", "nucleotide", "nucleotide sequence", "nucleic acid" and "oligonucleotide" are used interchangeably. They refer to a polymeric form of nucleotides of any length, either deoxyribonucleotides or ribonucleotides, or analogs thereof. Polynucleotides may have any three dimensional structure, and may perform any function, known or unknown. The following are non limiting examples of polynucleotides: coding or non-coding regions of a gene or gene fragment, intergenic DNA, loci (locus) defined from linkage analysis, exons, introns, messenger RNA (mRNA), transfer RNA, ribosomal RNA, short interfering RNA (siRNA), short-hairpin RNA (shRNA), micro-RNA (miRNA), small nucleolar RNA, ribozymes, cDNA, recombinant polynucleotides, branched polynucleotides, plasmids, vectors, isolated DNA of any sequence, isolated RNA of any sequence, nucleic acid probes, and primers. A polynucleotide may comprise modified nucleotides, such as methylated nucleotides and nucleotide analogs. If present, modifications to the nucleotide structure may be imparted before or after assembly of the polymer. The sequence of nucleotides may be interrupted by non nucleotide components. A polynucleotide may be further modified after polymerization, such as by conjugation with a labeling component. Polynucleotide sequences, when provided, are listed in the 5' to 3' direction, unless stated otherwise.

[0035] As used herein, the term "target polynucleotide" refers to a nucleic acid molecule or polynucleotide in a starting population of nucleic acid molecules having a target sequence whose presence, amount, and/or nucleotide sequence, or changes in these, are desired to be determined. In general, a target polynucleotide is a double-stranded nucleic acid molecule, and may be derived from any source of or process for generating double-stranded nucleic acid molecules.

[0036] As used herein, the term "target sequence" refers generally to a nucleic acid sequence on a single strand of nucleic acid. The target sequence may be a portion of a gene, a regulatory sequence, genomic DNA, cDNA, RNA including mRNA, miRNA, and rRNA, or others. The target sequence may be a target sequence from a sample or a secondary target such as a product of an amplification reaction.

[0037] A "nucleotide probe," "probe," or "tag oligonucleotide" refers to a polynucleotide used for detecting or identifying its corresponding target polynucleotide in a hybridization reaction. Thus, a tag

IDT Exhibit 1020.5914

oligonucleotide is hybridizable to one or more target polynucleotides. Tag oligonucleotides can be perfectly complementary to one or more target polynucleotides in a sample, or contain one or more nucleotides that are not complemented by a corresponding nucleotide in the one or more target polynucleotides in a sample.

[0038] "Hybridization" and "annealing" refer to a reaction in which one or more polynucleotides react to form a complex that is stabilized via hydrogen bonding between the bases of the nucleotide residues. The hydrogen bonding may occur by Watson Crick base pairing, Hoogstein binding, or in any other sequence specific manner. The complex may comprise two strands forming a duplex structure, three or more strands forming a multi stranded complex, a single self hybridizing strand, or any combination of these. A hybridization reaction may constitute a step in a more extensive process, such as the initiation of a PCR, or the enzymatic cleavage of a polynucleotide by a ribozyme. A first sequence that can be stabilized via hydrogen bonding with the bases of the nucleotide residues of a second sequence is said to be "hybridizable" to said second sequence. In such a case, the second sequence can also be said to be hybridizable to the first sequence.

[0039] In general, a "complement" of a given sequence is a sequence that is fully complementary to and hybridizable to the given sequence. In general, a first sequence that is hybridizable to a second sequence or set of second sequences is specifically or selectively hybridizable to the second sequence or set of second sequences, such that hybridization to the second sequence or set of second sequences is preferred (e.g. thermodynamically more stable under a given set of conditions, such as stringent conditions commonly used in the art) to hybridization with non-target sequences during a hybridization reaction. Typically, hybridizable sequences share a degree of sequence complementarity over all or a portion of their respective lengths, such as between 25%-100% complementarity, including at least about 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, and 100% sequence complementarity.

[0040] The term "hybridized" as applied to a polynucleotide refers to a polynucleotide in a complex that is stabilized via hydrogen bonding between the bases of the nucleotide residues. The hydrogen bonding may occur by Watson Crick base pairing, Hoogstein binding, or in any other sequence specific manner. The complex may comprise two strands forming a duplex structure, three or more strands forming a multi stranded complex, a single self hybridizing strand, or any combination of these. The hybridization reaction may constitute a step in a more extensive process, such as the initiation of a PCR reaction, or the enzymatic cleavage of a polynucleotide by a ribozyme. A sequence hybridized with a given sequence is referred to as the "complement" of the given sequence.

[0041] As used herein, "expression" refers to the process by which a polynucleotide is transcribed into mRNA and/or the process by which the transcribed mRNA (also referred to as "transcript") is subsequently being translated into peptides, polypeptides, or proteins. The transcripts and the encoded polypeptides are collectively referred to as "gene product." If the polynucleotide is derived from genomic DNA, expression may include splicing of the mRNA in a eukaryotic cell.

IDT Exhibit 1020.5915

DETAILED DESCRIPTION OF THE INVENTION

[0042] The practice of the present invention employs, unless otherwise indicated, conventional techniques of immunology, biochemistry, chemistry, molecular biology, microbiology, cell biology, genomics and recombinant DNA, which are within the skill of the art. See Sambrook, Fritsch and Maniatis, MOLECULAR CLONING: A LABORATORY MANUAL, 2nd edition (1989); CURRENT PROTOCOLS IN MOLECULAR BIOLOGY (F. M. Ausubel, et al. eds., (1987)); the series METHODS IN ENZYMOLOGY (Academic Press, Inc.): PCR 2: A PRACTICAL APPROACH (M.J. MacPherson, B.D. Hames and G.R. Taylor eds. (1995)), Harlow and Lane, eds. (1988) ANTIBODIES, A LABORATORY MANUAL, and ANIMAL CELL CULTURE (R.I. Freshney, ed. (1987)).

[0043] In one aspect, the present invention provides a method for multiplex sequencing. In one embodiment, the method comprises sequencing a plurality of target polynucleotides in a single reaction chamber, wherein said target polynucleotides are from two or more different samples; and identifying the sample from which each of said sequenced target polynucleotides is derived with an accuracy of at least 95% based on a single barcode contained in the sequence of said target polynucleotide. Reaction chambers can be any suitable chamber known in the art for containing a sequencing reaction, non-limiting examples of which include tubes of various dimensions, wells of multi-well plates, and channels of flow cells. In some embodiments, the target polynucleotides comprise one or more sequences with which the sequencing reaction is calibrated. In some embodiments, the one or more sequences with which the sequencing reaction is calibrated are joined to the target polynucleotides prior to sequencing.

[0044] In another aspect, the invention provides a method of producing adapter-tagged target polynucleotides from a plurality of independent samples. In one embodiment, the method comprises: (a) providing a plurality of first adapter oligonucleotides, wherein each of said first adapter 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; and (b) joining at least one of said first adapter oligonucleotides to said target polynucleotides of each of said samples, such that no barcode sequence is joined to said target polynucleotides of more than one of said samples. In some embodiments, the method further comprises (c) joining at least one of a plurality of second adapter oligonucleotides to said target polynucleotides of each of said samples from step (b), such that at least some of said target polynucleotides comprise said first adapter oligonucleotide at one end and said second adapter oligonucleotide at the other end. First and second adapter oligonucleotides can be the same or different, with different adapter oligonucleotides having different sequences and/or sequences of different lengths. A first adapter oligonucleotide can comprise one or more sequence regions that have the same sequence as one or more sequence regions of a second adapter oligonucleotide, and one or more sequence regions that have sequences that are different from one or more sequence regions of a second adapter oligonucleotide.

[0045] An adapter oligonucleotide includes any oligonucleotide having a sequence, at least a portion of which is known, that can be joined to a target polynucleotide. Adapter oligonucleotides can comprise

IDT Exhibit 1020.5916

DNA, RNA, nucleotide analogues, non-canonical nucleotides, labeled nucleotides, modified nucleotides, or combinations thereof. Adapter oligonucleotides can be single-stranded, double-stranded, or partial duplex. In general, a partial-duplex adapter comprises one or more single-stranded regions and one or more double-stranded regions. Double-stranded adapters can comprise two separate oligonucleotides hybridized to one another (also referred to as an "oligonucleotide duplex"), and hybridization may leave one or more blunt ends, one or more 3' overhangs, one or more 5' overhangs, one or more bulges resulting from mismatched and/or unpaired nucleotides, or any combination of these. In some embodiments, a single-stranded adapter comprises two or more sequences that are able to hybridize with one another. When two such hybridizable sequences are contained in a single-stranded adapter, hybridization yields a hairpin structure (hairpin adapter). When two hybridized regions of an adapter are separated from one another by a non-hybridized region, a "bubble" structure results. Adapters comprising a bubble structure can consist of a single adapter oligonucleotide comprising internal hybridizations, or may comprise two or more adapter oligonucleotides hybridized to one another. Internal sequence hybridization, such as between two hybridizable sequences in an adapter, can produce a double-stranded structure in a single-stranded adapter oligonucleotide. Adapters of different kinds can be used in combination, such as a hairpin adapter and a double-stranded adapter, or adapters of different sequences. Hybridizable sequences in a hairpin adapter may or may not include one or both ends of the oligonucleotide. When neither of the ends are included in the hybridizable sequences, both ends are "free" or "overhanging." When only one end is hybridizable to another sequence in the adapter, the other end forms an overhang, such as a 3' overhang or a 5' overhang. When both the 5'-terminal nucleotide and the 3'-terminal nucleotide are included in the hybridizable sequences, such that the 5'-terminal nucleotide and the 3'-terminal nucleotide are complementary and hybridize with one another, the end is referred to as "blunt." Different adapters can be joined to target polynucleotides in sequential reactions or simultaneously. For example, the first and second adapters can be added to the same reaction. Adapters can be manipulated prior to combining with target polynucleotides. For example, terminal phosphates can be added or removed.

[0046] In some embodiments, one of the hybridizable sequences in a single-stranded hairpin adapter comprises RNA. For example, an adapter can comprise a 5' end comprising sequence A and a 3' end comprising sequence A', where A is hybridizable to A', one of A or A' comprises DNA, and the other of A or A' comprises RNA. Similarly, an adapter can comprise a 5' end comprising sequence B and a 3' end comprising sequence B', where B is hybridizable to B', one of B or B' comprises DNA, and the other of B or B' comprises RNA. In some embodiments, one of A or A' consists entirely of DNA, and/or one of A or A' consists entirely of RNA. In some embodiment, one of B or B' consists entirely of DNA, and/or one of B or B' consists entirely of RNA. Sequence A can be the same as or different from sequence B and/or B'. Sequence A' can be the same as or different from sequence B and/or B'. In some embodiments, the end of a hairpin comprising RNA (e.g. A, A', B, or B') further comprises one or more terminal DNA residues (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, or more terminal DNA

IDT Exhibit 1020.5917

residues), such that the sequence comprising RNA is flanked by DNA residues at both ends (i.e. both the 5' end and the 3' end of the sequence comprising RNA). Hybridization of a sequence comprising RNA to a sequence comprising DNA creates an RNA-DNA heteroduplex. In some embodiments, RNA is cleaved by an enzyme that cleaves RNA from an RNA-DNA heteroduplex, such as enzymes comprising ribonuclease activity. Preferably, the enzyme comprising ribonuclease activity cleaves ribonucleotides in an RNA/DNA heteroduplex regardless of the identity and type of nucleotides adjacent to the ribonucleotide to be cleaved. It is preferred that the ribonuclease cleaves independent of sequence identity. Examples of suitable enzymes comprising ribonuclease activity for the methods and compositions of the invention are well known in the art, including ribonuclease H (RNase H) and enzymes comprising RNase H activity, e.g., Hybridase. In some embodiments, cleavage of RNA from an RNA-DNA heteroduplex removes all double-stranded character from a single-stranded hairpin adapter oligonucleotide, such that extension by a polymerase that uses the adapter as template requires no strand displacement step or strand displacement activity. In some embodiments, both ends of a hairpin adapter comprising one end comprising RNA are joined to a target polynucleotide, such that cleavage of the RNA from the RNA-DNA hetero duplex produces a 5' overhang or a 3' overhang. In some embodiments, an end comprising a 5' overhang produced by cleavage of RNA from an RNA-DNA heteroduplex is filled in by the extension of the produced 3' end using the 5' overhang as template.

[0047] In some embodiments, where hairpin adapters comprising 3' ends comprising RNA are joined to both 3' ends of a double-stranded target polynucleotide, cleavage of RNA from the RNA-DNA heteroduplex is followed by hybridization of oligonucleotides to the adapter sequences joined in the first step, and ligation of the hybridized oligonucleotides to the 5' ends of the double-stranded target polynucleotide to produce a target polynucleotide comprising non-complementary, single-stranded overhangs of both strands at both ends. Amplification of a double-stranded target polynucleotides comprising non-complementary, single-stranded overhangs on both strands at both ends can comprise the use of a first and second primer, wherein the first primer is hybridizable to one of the overhangs and the second primer is hybridizable to the complement of the overhang at the other end of the strand to which the first primer is hybridizable. Sequencing of double-stranded target polynucleotides comprising non-complementary, single-stranded overhangs on both strands at both ends can comprise the use of one or more sequencing primers hybridizable to one or more of the overhangs, or complements thereof. An illustrative example of the production of a double-stranded target polynucleotide comprising non-complementary, single-stranded overhangs on both strands at both ends is shown in Figure 5.

[0048] Adapters can contain one or more of a variety of sequence elements, including but not limited to, one or more amplification primer annealing sequences or complements thereof, one or more sequencing primer annealing sequences or complements thereof, one or more barcode sequences, one or more common sequences shared among multiple different adapters or subsets of different adapters, one or more restriction enzyme recognition sites, one or more overhangs complementary to one or more target polynucleotide overhangs, one or more probe binding sites (e.g. for attachment to a sequencing platform,

IDT Exhibit 1020.5918

such as a flow cell for massive parallel sequencing, such as developed by Illumina, Inc.), one or more random or near-random sequences (e.g. one or more nucleotides selected at random from a set of two or more different nucleotides at one or more positions, with each of the different nucleotides selected at one or more positions represented in a pool of adapters comprising the random sequence), and combinations thereof. Two or more sequence elements can be non-adjacent to one another (e.g. separated by one or more nucleotides), adjacent to one another, partially overlapping, or completely overlapping. For example, an amplification primer annealing sequence can also serve as a sequencing primer annealing sequence. Sequence elements can be located at or near the 3' end, at or near the 5' end, or in the interior of the adapter oligonucleotide. When an adapter oligonucleotide is capable of forming secondary structure, such as a hairpin, sequence elements can be located partially or completely outside the secondary structure, partially or completely inside the secondary structure, or in between sequences participating in the secondary structure. For example, when an adapter oligonucleotide comprises a hairpin structure, sequence elements can be located partially or completely inside or outside the hybridizable sequences (the "stem"), including in the sequence between the hybridizable sequences (the "loop"). In some embodiments, the first adapter oligonucleotides in a plurality of first adapter oligonucleotides having different barcode sequences comprise a sequence element common among all first adapter oligonucleotides in the plurality. In some embodiments, all second adapter oligonucleotides comprise a sequence element common among all second adapter oligonucleotides that is different from the common sequence element shared by the first adapter oligonucleotides. A difference in sequence elements can be any such that at least a portion of different adapters do not completely align, for example, due to changes in sequence length, deletion or insertion of one or more nucleotides, or a change in the nucleotide composition at one or more nucleotide positions (such as a base change or base modification). In some embodiments, an adapter oligonucleotide comprises a 5' overhang, a 3' overhang, or both that is complementary to one or more target polynucleotides. Complementary overhangs can be one or more nucleotides in length, including but not limited to 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, or more nucleotides in length. Complementary overhangs may comprise a fixed sequence. Complementary overhangs may comprise a random sequence of one or more nucleotides, such that one or more nucleotides are selected at random from a set of two or more different nucleotides at one or more positions, with each of the different nucleotides selected at one or more positions represented in a pool of adapters with complementary overhangs comprising the random sequence. In some embodiments, an adapter overhang is complementary to a target polynucleotide overhang produced by restriction endonuclease digestion. In some embodiments, an adapter overhang consists of an adenine or a thymine.

[0049] In some embodiments, one or more of the adapter oligonucleotides comprises SEQ ID NO: 1. In some embodiments, one or more of the adapter oligonucleotides comprises SEQ ID NO: 2. In some embodiments, the sequence element common among all first adapter oligonucleotides comprises SEQ ID NO: 1 or SEQ ID NO: 2. In some embodiments, the sequence element common among all second adapter oligonucleotides comprises SEQ ID NO: 1 or SEQ ID NO: 2. In some embodiments, one of SEQ

IDT Exhibit 1020.5919

ID NO: 1 or SEQ ID NO: 2 is common among all first adapter oligonucleotides and the other of SEQ ID NO: 1 or SEQ ID NO: 2 is common among all second adapter oligonucleotides. In some embodiments, one or more of the adapter oligonucleotides comprises SEQ ID NO: 3. In some embodiments, one or more of the adapter oligonucleotides comprises SEQ ID NO: 4. In some embodiments the 3'-most nucleotide of SEQ ID NO: 3 and/or SEQ ID NO: 4 is followed by one or more nucleotides of a barcode sequence.

[0050] In some embodiments, an adapter comprising an oligonucleotide duplex comprises an oligonucleotide comprising SEQ ID NO: 86 and/or an oligonucleotide comprising SEQ ID NO: 87. In some embodiments, an adapter comprising an oligonucleotide duplex comprises an oligonucleotide comprising SEQ ID NO: 88, and/or an oligonucleotide comprising SEQ ID NO: 89.

[0051] Adapter oligonucleotides can have any suitable length, at least sufficient to accommodate the one or more sequence elements of which they are comprised. In some embodiments, adapters are about, less than about, or more than about 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 90, 100, 200, or more nucleotides in length. In some embodiments, the stem of a hairpin adapter is about, less than about, or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50, 75, 100, or more nucleotides in length. Stems may be designed using a variety of different sequences that result in hybridization between the complementary regions on a hairpin adapter, resulting in a local region of double-stranded DNA. For example, stem sequences may be utilized that are from 15 to 18 nucleotides in length with equal representation of G:C and A:T base pairs. Such stem sequences are predicted to form stable dsDNA structures below their predicted melting temperatures of ~45°C. Sequences participating in the stem of the hairpin can be perfectly complementary, such that each base of one region in the stem hybridizes via hydrogen bonding with each base in the other region in the stem according to Watson-Crick base-pairing rules. Alternatively, sequences in the stem may deviate from perfect complementarity. For example, there can be mismatches and or bulges within the stem structure created by opposing bases that do not follow Watson-Crick base pairing rules, and/or one or more nucleotides in one region of the stem that do not have the one or more corresponding base positions in the other region participating in the stem. Mismatched sequences may be cleaved using enzymes that recognize mismatches. The stem of a hairpin can comprise DNA, RNA, or both DNA and RNA. In some embodiments, the stem and/or loop of a hairpin, or one or both of the hybridizable sequences forming the stem of a hairpin, comprise nucleotides, bonds, or sequences that are substrates for cleavage, such as by an enzyme, including but not limited to endonucleases and glycosylases. The composition of a stem may be such that only one of the hybridizable sequences forming the stem is cleaved. For example, one of the sequences forming the stem may comprise RNA while the other sequence forming the stem consists of DNA, such that cleavage by an enzyme that cleaves RNA in an RNA-DNA duplex, such as RNase H, cleaves only the sequence comprising RNA. The stem and/or loop of a hairpin can comprise non-canonical nucleotides (e.g. uracil), and/or methylated nucleotides. In some embodiments, one strand of a hairpin adapter stem comprises

IDT Exhibit 1020.5920

SEQ ID NO: 1 or SEQ ID NO: 2. In some embodiments, the loop sequence of a hairpin adapter is about, less than about, or more than about 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, or more nucleotides in length.

[0052] As used herein, the term "barcode" refers to a known nucleic acid sequence that allows some feature of a polynucleotide with which the barcode is associated to be identified. In some embodiments, the feature of the polynucleotide to be identified is the sample from which the polynucleotide is derived. In some embodiments, barcodes are at least 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, or more nucleotides in length. In some embodiments, barcodes are shorter than 10, 9, 8, 7, 6, 5, or 4 nucleotides in length. In some embodiments, barcodes associated with some polynucleotides are of different length than barcodes associated with other polynucleotides. In general, barcodes are of sufficient length and comprise sequences that are sufficiently different to allow the identification of samples based on barcodes with which they are associated. In some embodiments, a barcode, and the sample source with which it is associated, can be identified accurately after the mutation, insertion, or deletion of one or more nucleotides in the barcode sequence, such as the mutation, insertion, or deletion of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more nucleotides. In some embodiments, each barcode in a plurality of barcodes differ from every other barcode in the plurality at at least three nucleotide positions, such as at least 3, 4, 5, 6, 7, 8, 9, 10, or more positions. In some embodiments, both the first adapter and the second adapter comprise at least one of a plurality of barcode sequences. In some embodiments, barcodes for second adapter oligonucleotides are selected independently from barcodes for first adapter oligonucleotides. In some embodiments, first adapter oligonucleotides and second adapter oligonucleotides having barcodes are paired, such that adapters of the pair comprise the same or different one or more barcodes. In some embodiments, the methods of the invention further comprise identifying the sample from which a target polynucleotide is derived based on a barcode sequence to which the target polynucleotide is joined. In general, a barcode comprises a nucleic acid sequence that when joined to a target polynucleotide serves as an identifier of the sample from which the target polynucleotide was derived.

[0053] In some embodiments, the plurality of barcode sequences from which barcode sequences are selected includes sequences selected from the group consisting of: AAA, TTT, CCC, GGG. In some embodiments, the plurality of barcode sequences from which barcode sequences are selected includes sequences selected from the group consisting of: AAAA, CTGC, GCTG, TGCT, ACCC, CGTA, GAGT, TTAG, AGGG, CCAT, GTCA, TATC, ATTT, CACG, GGAC, and TCGA. In some embodiments, the plurality of barcode sequences from which barcode sequences are selected includes sequences selected from the group consisting of: AAAAA, AACCC, AAGGG, AATTT, ACACG, ACCAT, ACGTA, ACTGC, AGAGT, AGCTG, AGGAC, AGTCA, ATATC, ATCGA, ATGCT, ATTAG, CAACT, CACAG, CAGTC, CATGA, CCAAC, CCCCA, CCGGT, CCTTG, CGATA, CGCGC, CGGCG, CGTAT, CTAGG, CTCTT, CTGAA, CTTCC, GAAGC, GACTA, GAGAT, GATCG, GCATT, GCCGG, GCGCC, GCTAA, GGAAG, GGCCT, GGGGA, GGTTC, GTACA, GTCAC, GTGTG, GTTTT, TAATG, TACGT, TAGCA, TATAC, TCAGA, TCCTC, TCGAG, TCTCT, TGACC, TGCAA, TGGTT, TGTGG, TTAAT, TTCCG, TTGGC, and TTTTA.

IDT Exhibit 1020.5921

[0054] The terms "joining" and "ligation" as used herein, with respect to two polynucleotides, such as an adapter oligonucleotide and a target polynucleotide, refers to the covalent attachment of two separate polynucleotides to produce a single larger polynucleotide with a contiguous backbone. Methods for joining two polynucleotides are known in the art, and include without limitation, enzymatic and non-enzymatic (e.g. chemical) methods. Examples of ligation reactions that are non-enzymatic include the non-enzymatic ligation techniques described in U.S. Pat. Nos. 5,780,613 and 5,476,930, which are herein incorporated by reference. In some embodiments, an adapter oligonucleotide is joined to a target polynucleotide by a ligase, for example a DNA ligase or RNA ligase. Multiple ligases, each having characterized reaction conditions, are known in the art, and include, without limitation NAD^+-dependent ligases including tRNA ligase, Taq DNA ligase, *Thermus filiformis* DNA ligase, Escherichia coli DNA ligase, Tth DNA ligase, *Thermus scotoductus* DNA ligase (I and II), thermostable ligase, Ampligase thermostable DNA ligase, VanC-type ligase, 9° N DNA Ligase, Tsp DNA ligase, and novel ligases discovered by bioprospecting; ATP-dependent ligases including T4 RNA ligase, T4 DNA ligase, T3 DNA ligase, T7 DNA ligase, Pfu DNA ligase, DNA ligase 1, DNA ligase III, DNA ligase IV, and novel ligases discovered by bioprospecting; and wild-type, mutant isoforms, and genetically engineered variants thereof. Ligation can be between polynucleotides having hybridizable sequences, such as complementary overhangs. Ligation can also be between two blunt ends. Generally, a 5' phosphate is utilized in a ligation reaction. The 5' phosphate can be provided by the target polynucleotide, the adapter oligonucleotide, or both. 5' phosphates can be added to or removed from polynucleotides to be joined, as needed. Methods for the addition or removal of 5' phosphates are known in the art, and include without limitation enzymatic and chemical processes. Enzymes useful in the addition and/or removal of 5' phosphates include kinases, phosphatases, and polymerases. In some embodiments, both of the two ends joined in a ligation reaction (e.g. an adapter end and a target polynucleotide end) provide a 5' phosphate, such that two covalent linkages are made in joining the two ends. In some embodiments, only one of the two ends joined in a ligation reaction (e.g. only one of an adapter end and a target polynucleotide end) provides a 5' phosphate, such that only one covalent linkage is made in joining the two ends. In some embodiments, only one strand at one or both ends of a target polynucleotide is joined to an adapter oligonucleotide. In some embodiments, both strands at one or both ends of a target polynucleotide are joined to an adapter oligonucleotide. In some embodiments, 3' phosphates are removed prior to ligation. In some embodiments, an adapter oligonucleotide is added to both ends of a target polynucleotide, wherein one or both strands at each end are joined to one or more adapter oligonucleotides. When both strands at both ends are joined to an adapter oligonucleotide, joining can be followed by a cleavage reaction that leaves a 5' overhang that can serve as a template for the extension of the corresponding 3' end, which 3' end may or may not include one or more nucleotides derived from the adapter oligonucleotide. In some embodiments, a target polynucleotide is joined to a first adapter oligonucleotide on one end and a second adapter oligonucleotide on the other end. In some embodiments, the target polynucleotide and the adapter to which it is joined comprise blunt ends. In some embodiments, separate

IDT Exhibit 1020.5922

ligation reactions are carried out for each sample, using a different first adapter oligonucleotide comprising at least one barcode sequence for each sample, such that no barcode sequence is joined to the target polynucleotides of more than one sample. A target polynucleotide that has an adapter oligonucleotide joined to it is considered "tagged" by the joined adapter.

[0055] In some embodiments, joining of an adapter to a target polynucleotide produces a joined product polynucleotide having a 3' overhang comprising a nucleotide sequence derived from the adapter. 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.

[0056] In some embodiments, after joining at least one adapter oligonucleotide to a target polynucleotide, the 3' end of one or more target polynucleotides is extended using the one or more joined adapter oligonucleotides as template. For example, an adapter comprising two hybridized oligonucleotides that is joined to only the 5' end of a target polynucleotide allows for the extension of the unjoined 3'end of the target using the joined strand of the adapter as template, concurrently with or following displacement of the unjoined strand. If both strands of an adapter comprising two hybridized oligonucleotides are joined to a target polynucleotide such that the joined product has a 5' overhang, the complementary 3' end can be extended using the 5' overhang as template. As a further example, a hairpin adapter oligonucleotide can be joined to the 5' end of a target polynucleotide. While double-stranded in secondary structure, such a hairpin adapter remains single-stranded, and is thus a 5' overhang appended to the target polynucleotide (e.g. when the 5' end of the hairpin adapter is not joined to the target polynucleotide). Removal of the secondary structure, either prior to (e.g. thermal denaturing, or degradation) or concurrently with (e.g. strand displacement) the activity of a polymerase, provides a template for the extension of the 3' end of the complementary strand of the target polynucleotide. In some embodiments, the 3' end of the target polynucleotide that is extended comprises one or more nucleotides from an adapter oligonucleotide. For target polynucleotides to which adapters are joined on both ends, extension can be carried out for both 3' ends of a double-stranded target polynucleotide having 5' overhangs. This 3' end extension, or "fill-in" reaction, generates a complementary sequence, or "complement," to the adapter oligonucleotide template that is hybridized to the template, thus filling in the 5' overhang to produce a double-stranded sequence region. Where both ends of a double-stranded target polynucleotide have 5' overhangs that are filled in by extension of the complementary strands' 3' ends, the product is completely double-stranded. Extension can be carried out by any suitable polymerase known in the art, such as a DNA polymerase, many of which are commercially available. DNA polymerases can comprise DNA-dependent DNA polymerase activity, RNA-dependent DNA polymermase activity, or DNA-dependent and RNA-dependent DNA polymerase activity. DNA polymerases can be thermostable or non-thermostable. Examples of DNA polymerases include, but are

IDT Exhibit 1020.5923

not limited to, Taq polymerase, Tth polymerase, Tli polymerase, Pfu polymerase, Pfutubo polymerase, Pyrobest polymerase, Pwo polymerase, KOD polymerase, Bst polymerase, Sac polymerase, Sso polymerase, Poc polymerase, Pab polymerase, Mth polymerase, Pho polymerase, ES4 polymerase, VENT polymerase, DEEPVENT polymerase, EX-Taq polymerase, LA-Taq polymerase, Expand polymerases, Platinum Taq polymerases, Hi-Fi polymerase, Tbr polymerase, Tfl polymerase, Tru polymerase, Tac polymerase, Tne polymerase, Tma polymerase, Tih polymerase, Tfi polymerase, Klenow fragment, and variants, modified products and derivatives thereof. 3' end extension can be performed before or after pooling of target polynucleotides from independent samples.

[0057] In some embodiments, the fill-in reaction is followed by or performed as part of amplification of one or more target polynucleotides using a first primer and a second primer, wherein the first primer comprises a sequence that is hybridizable to at least a portion of the complement of one or more of the first adapter oligonucleotides, and further wherein the second primer comprises a sequence that is hybridizable to at least a portion of the complement of one or more of the second adapter oligonucleotides. Each of the first and second primers may be of any suitable length, such as about, less than about, or more than about 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 90, 100, or more nucleotides, any portion or all of which may be complementary to the corresponding target sequence (e.g. about, less than about, or more than about 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, or more nucleotides). "Amplification" refers to any process by which the copy number of a target sequence is increased. Methods for primer-directed amplification of target polynucleotides are known in the art, and include without limitation, methods based on the polymerase chain reaction (PCR). Conditions favorable to the amplification of target sequences by PCR are known in the art, can be optimized at a variety of steps in the process, and depend on characteristics of elements in the reaction, such as target type, target concentration, sequence length to be amplified, sequence of the target and/or one or more primers, primer length, primer concentration, polymerase used, reaction volume, ratio of one or more elements to one or more other elements, and others, some or all of which can be altered. In general, PCR involves the steps of denaturation of the target to be amplified (if double stranded), hybridization of one or more primers to the target, and extension of the primers by a DNA polymerase, with the steps repeated (or "cycled") in order to amplify the target sequence. Steps in this process can be optimized for various outcomes, such as to enhance yield, decrease the formation of spurious products, and/or increase or decrease specificity of primer annealing. Methods of optimization are well known in the art and include adjustments to the type or amount of elements in the amplification reaction and/or to the conditions of a given step in the process, such as temperature at a particular step, duration of a particular step, and/or number of cycles. In some embodiments, an amplification reaction comprises at least 5, 10, 15, 20, 25, 30, 35, 50, or more cycles. In some embodiments, an amplification reaction comprises no more than 5, 10, 15, 20, 25, 35, 50, or more cycles. Cycles can contain any number of steps, such as 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more steps. Steps can comprise any temperature or gradient of temperatures, suitable for achieving the purpose of the given step, including but not limited to, 3' end extension (e.g. adapter fill-in), primer annealing, primer

IDT Exhibit 1020.5924

extension, and strand denaturation. Steps can be of any duration, including but not limited to about, less than about, or more than about 1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 70, 80, 90, 100, 120, 180, 240, 300, 360, 420, 480, 540, 600, or more seconds, including indefinitely until manually interrupted. Cycles of any number comprising different steps can be combined in any order. In some embodiments, different cycles comprising different steps are combined such that the total number of cycles in the combination is about, less that about, or more than about 5, 10, 15, 20, 25, 30, 35, 50, or more cycles. In some embodiments, one or more of the primers comprises SEQ ID NO: 1. In some embodiments, one or more of the primers comprises SEQ ID NO: 2. In some embodiments, amplification is performed following the fill-in reaction. Amplification can be performed before or after pooling of target polynucleotides from independent samples.

[0058] In some embodiments, target polynucleotides from independent samples are pooled after the joining step. Pooling can be performed immediately after the joining step, or following one or more intervening steps between joining and pooling. Pools can comprise any fraction of the total target polynucleotides from a joining reaction, including the whole reaction volume. Samples can be pooled evenly or unevenly. Target polynucleotides can be further processed before or after pooling, for example to purify desired products or eliminate undesired products. Pools can comprise polynucleotides from any number of independent samples, such as at least 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 some embodiments, target polynucleotides are pooled based on the barcodes to which they are joined. In some embodiments, target polynucleotides from independent samples are pooled such that all four bases are evenly represented at one or more positions along the barcode, among barcodes included in the pool. In some embodiments, target polynucleotides from independent samples are pooled such that all four bases are evenly represented at every position along the barcode, among barcodes included in the pool. Where only one barcode is joined to polynucleotides of each sample, samples can be pooled in multiples of four in order to represent all four bases at one or more positions along the barcode evenly, for example 4, 8, 12, 16, 20, 24, 28, 32, 36, 40, 44, 48, 52, 56, 60, 64, 96, 128, 192, 256, 384, and so on. Where two barcodes are included in the joining reaction for each sample, such as two different first adapter oligonucleotides or one first adapter oligonucleotide and one second adapter oligonucleotide each having barcodes, samples can be pooled in multiples of two in order to evenly represent all four bases at one or more positions along the barcode, for example 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 48, 64, 96, 128, 256, 384, and so on. All combinations of the number of barcodes included in the joining reaction for target polynucleotides from each sample and the multiples in which samples are pooled in order to evenly represent all four nucleotides at one or more positions along the barcode are contemplated by the methods of the invention.

[0059] In some embodiments, pooling of target polynucleotides is followed by sequencing one or more polynucleotides in the pool. Sequencing processes are generally template dependent. Nucleic acid sequence analysis that employs template dependent synthesis identifies individual bases, or groups of

IDT Exhibit 1020.5925

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.

[0060] 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).

[0061] 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.

[0062] 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).

IDT Exhibit 1020.5926

[0063] In related processes, the primer/template/polymerase complex is immobilized upon a substrate and the complex is contacted with labeled nucleotides. The immobilization of the complex may be through the primer sequence, the template sequence and/or the polymerase enzyme, and may be covalent or noncovalent. For example, immobilization of the complex can be via a linkage between the polymerase or the primer and the substrate surface. A variety of types of linkages are useful for this attachment, including, e.g., provision of biotinylated surface components, using e.g., biotin-PEG-silane linkage chemistries, followed by biotinylation of the molecule to be immobilized, and subsequent linkage through, e.g., a streptavidin bridge. Other synthetic coupling chemistries, as well as non-specific protein adsorption can also be employed for immobilization. In alternate configurations, the nucleotides are provided with and without removable terminator groups. Upon incorporation, the label is coupled with the complex and is thus detectable. In the case of terminator bearing nucleotides, all four different nucleotides, bearing individually identifiable labels, are contacted with the complex. Incorporation of the labeled nucleotide arrests extension, by virtue of the presence of the terminator, and adds the label to the complex. The label and terminator are then removed from the incorporated nucleotide, and following appropriate washing steps, the process is repeated. In the case of non-terminated nucleotides, a single type of labeled nucleotide is added to the complex to determine whether it will be incorporated, as with pyrosequencing. Following removal of the label group on the nucleotide and appropriate washing steps, the various different nucleotides are cycled through the reaction mixture in the same process. See, e.g., U.S. Pat. No. 6,833,246, incorporated herein by reference in its entirety for all purposes). For example, the Illumina Genome Analyzer System is based on technology described in WO 98/44151, hereby incorporated by reference, wherein DNA molecules are bound to a sequencing platform (flow cell) via an anchor probe binding site (otherwise referred to as a flow cell binding site) and amplified in situ on a glass slide. The DNA molecules are then annealed to a sequencing primer and sequenced in parallel base-by-base using a reversible terminator approach. Typically, the Illumina Genome Analyzer System utilizes flow-cells with 8 channels, generating sequencing reads of 18 to 36 bases in length, generating >1.3 Gbp of high quality data per run (see www.illumina.com).

[0064] 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

IDT Exhibit 1020.5927

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).

[0065] 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.

[0066] 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 adapter 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 adapter 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, the sequencing primer comprises SEQ ID NO: 1 or SEQ ID NO:2. In some embodiments, the sequencing primer comprises SEQ ID NO: 5. In some embodiments, the sequencing primer comprises SEQ ID NO: 6. 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.

[0067] In some embodiments, accurate identification of the sample from which a target polynucleotide is derived is based on at least a portion of the sequence obtained for the target polynucleotide and is at least 90%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.8%, 99.85%, 99.9%, 99.95%, 99.99%, or more accurate. In some embodiments, the sample source of a target polynucleotide is identified based on a single barcode contained in the sequence. In some embodiments, accuracy can be increased by identifying the source of 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 adapter to which a target polynucleotide is joined, and/or by joining two or more adapters having one or more barcodes to a

IDT Exhibit 1020.5928

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.

[0068] The different samples from which the target polynucleotides are derived can comprise multiple samples from the same individual, samples from different individuals, or combinations thereof. In some embodiments, a sample comprises a plurality of polynucleotides from a single individual. In some embodiments, a sample comprises a plurality of polynucleotides from two or more individuals. An individual is any organism or portion thereof from which target polynucleotides can be derived, non-limiting examples of which include plants, animals, fungi, protists, monerans, viruses, mitochondria, and chloroplasts. Sample polynucleotides can be isolated from a subject, such as a cell sample, tissue sample, or organ sample derived therefrom, including, for example, cultured cell lines, biopsy, blood sample, or fluid sample containing a cell. The subject may be an animal, including but not limited to, an animal such as a cow, a pig, a mouse, a rat, a chicken, a cat, a dog, etc., and is usually a mammal, such as a human. Samples can also be artificially derived, such as by chemical synthesis. In some embodiments, the samples comprise DNA. In some embodiments, the samples comprise genomic DNA. In some embodiments, the samples comprise mitochondrial DNA, chloroplast DNA, plasmid DNA, bacterial artificial chromosomes, yeast artificial chromosomes, oligonucleotide tags, or combinations thereof. In some embodiments, the samples comprise DNA generated by primer extension reactions using any suitable combination of primers and a DNA polymerase, including but not limited to polymerase chain reaction (PCR), reverse transcription, and combinations thereof. Where the template for the primer extension reaction is RNA, the product of reverse transcription is referred to as complementary DNA (cDNA). Primers useful in primer extension reactions can comprise sequences specific to one or more targets, random sequences, partially random sequences, and combinations thereof. Reaction conditions suitable for primer extension reactions are known in the art. In general, sample polynucleotides comprise any polynucleotide present in a sample, which may or may not include target polynucleotides.

[0069] 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

IDT Exhibit 1020.5929

purification includes the use of magnetic particles to which nucleic acids can specifically or non-specifically bind, followed by isolation of the beads using a magnet, and washing and eluting the nucleic acids from the beads (see e.g. U.S. Pat. No. 5,705,628). In some embodiments, the above isolation methods may be preceded by an enzyme digestion step to help eliminate unwanted protein from the sample, e.g., digestion with proteinase K, or other like proteases. See, e.g., U.S. Pat. No. 7,001,724. If desired, RNase inhibitors may be added to the lysis buffer. For certain cell or sample types, it may be desirable to add a protein denaturation/digestion step to the protocol. Purification methods may be directed to isolate DNA, RNA, or both. When both DNA and RNA are isolated together during or subsequent to an extraction procedure, further steps may be employed to purify one or both separately from the other. Sub-fractions of extracted nucleic acids can also be generated, for example, purification by size, sequence, or other physical or chemical characteristic. In addition to an initial nucleic isolation step, purification of nucleic acids can be performed after any step in the methods of the invention, such as to remove excess or unwanted reagents, reactants, or products.

[0070] 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 subjection sample polynucleotides to acoustic sonication. In some embodiments, the fragmentation comprises treating the sample polynucleotides with one or more enzymes under conditions suitable for the one or more enzymes to generate double-stranded nucleic acid breaks. Examples of enzymes useful in the generation of polynucleotide fragments include sequence specific and non-sequence specific nucleases. Non-limiting examples of nucleases include DNase I, Fragmentase, restriction endonucleases, variants thereof, and combinations thereof. For example, digestion with DNase I can induce random double-stranded breaks in DNA in the absence of Mg^{++} and in the presence of Mn^{++}. In some embodiments, fragmentation comprises treating the sample polynucleotides with one or more restriction endonucleases. Fragmentation can produce fragments having 5' overhangs, 3' overhangs, blunt ends, or a combination thereof. In some embodiments, such as when fragmentation comprises the use of one or more restriction endonucleases, cleavage of sample polynucleotides leaves overhangs having a predictable sequence. In some embodiments, the method includes the step of size selecting the fragments via standard methods such as column purification or isolation from an agarose gel.

IDT Exhibit 1020.5930

[0071] In some embodiments, the 5' and/or 3' end nucleotide sequences of fragmented DNA are not modified prior to ligation with one or more adapter oligonucleotides. For example, fragmentation by a restriction endonuclease can be used to leave a predictable overhang, followed by ligation with one or more adapter 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 adapter 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 adapters. 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 adapter oligonucleotides having a complementary overhang, such as in a ligation reaction. For example, a single adenine can be added to the 3' ends of end repaired DNA fragments using a template independent polymerase, followed by ligation to one or more adapters each having a thymine at a 3' end. In some embodiments, adapter 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 adapters 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.

[0072] 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.

[0073] 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 polynucletides is derived can be identified in a combined

IDT Exhibit 1020.5931

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 adapter oligonucleotides, wherein each of said first adapter 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.

[0074] 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 adapters, primers, and/or other oligonucleotides described herein. In some embodiments, the kit further comprises one or more of: (a) a DNA ligase, (b) a DNA-dependent DNA polymerase, (c) an RNA-dependent DNA polymerase, (d) random primers, (e) primers comprising at least 4 thymidines at the 3' end, (f) a DNA endonuclease, (g) a DNA-dependent DNA polymerase having 3' to 5' exonuclease activity, (h) a plurality of primers, each primer having one of a plurality of selected sequences, (i) a DNA kinase, (j) a DNA exonuclease, (k) magnetic beads, (l) an enzyme comprising RNase H activity, (m) an RNA ligase, and (n) one or more buffers suitable for one or more of the elements contained in said kit. The adapters, 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 of the amounts and/or combinations (such as in the same kit or same container) described above. 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.

[0075] In one embodiment, the kit comprises a plurality of first adapter oligonucleotides, wherein each of said first adapter 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, and instructions for using the same. First adapters comprising different barcode sequences can be supplied individually or in combination with one or more additional first adapter having a different barcode sequence. In some embodiments, the kit further comprises a plurality of second adapter oligonucleotides. Second adapter oligonucleotides can be supplied separately from or in combination with one or more first adapters, and/or one or more different second adapters. Combinations of first and second adapters can be supplied in accordance with combinations described above.

IDT Exhibit 1020.5932

EXAMPLES

[0076] The following examples are given for the purpose of illustrating various embodiments of the invention and are not meant to limit the present invention in any fashion. The present examples, along with the methods described herein are presently representative of preferred embodiments, are exemplary, and are not intended as limitations on the scope of the invention. Changes therein and other uses which are encompassed within the spirit of the invention as defined by the scope of the claims will occur to those skilled in the art.

Example 1: Fragmentation and repair of sample nucleic acid

[0077] The sample comprising target polynucleotides ("sample") used in this example is human genomic DNA. In order to fragment the nucleic acid, 1ug to 5ug are diluted in 120uL of TE, and the dilution is subjected to mechanic fragmentation using a Covaris S-series sonication instrument (Covaris, Inc.) with the following setting: duty cycle = 10, intensity = 5, cycles/burst = 100, time = 10 minutes, and sample volume = 120uL. Fragmented nucleic acid is the purified with SPRI beads (Beckman Coulter, Inc.) at a ratio of 1:1.8 (sample:beads). DNA is eluted from the beads with 40uL of TE, and quantified, such as by using a Nanodrop, Quibit, or similar DNA quantitation instrument, or spectrophotometrically. Fragmentation products having 5' overhangs, 3' overhangs, non-phosphorylated 3' ends, and/or phosphorylated 3' ends are then end-repaired using a blend of enzymes that specifically eliminate overhangs and restore end residues to the appropriate 5' phosphorylation and 3' hydroxyl configuration. For end-repair using the Quick Blunting kit (New England Biolabs, Inc.), 100-200ng of fragmented DNA is combined with 1.25uL 10X quick blunt buffer, 1.25uL 1mM dNTP mix, and water to a final volume of 12uL. This combination is mixed well, spun down in the tube, and 0.5uL of quick blunt enzyme (a combination of T4 DNA polymerase and T4 polynucleotide kinase) is added, followed by incubation at room temperature for 30 minutes, and inactivation at 70°C for 10 minutes. Nucleic acid prepared according to the methods of this example can be stored at -20°C or used immediately is subsequent ligation reaction to join target polynucleotide fragments to adapters. A diagram of the steps in the procedure, including fragmentation, end-repair, adapter ligation, adapter fill-in, amplification, and sequencing is shown in Figure 1.

Example 2: Effects of target polynucleotide to adapter ratios on library construction

[0078] The present example examines the effects of varying target polynucleotide to adapter ratios on the construction of a collection of adapter tagged target polynucleotides (or "library"). The sample comprising target polynucleotides ("sample") used in this example is prepared as described in Example 1. The first adapter in this example consists of SEQ ID NO: 7. The second adapter consists of SEQ ID NO: 8. One of the primers used in the amplification step of this example consists of SEQ ID NO: 9, and the other primer in the pair consists of SEQ ID NO: 10. Ligation reactions are prepared such that each

IDT Exhibit 1020.5933

contains 10uL of 2X ligation buffer, 4uL of sample nucleic acid, 4uL of combined adapters, 1uL of water (5uL in reaction lacking sample or adapters), and 1uL ligase. In addition to buffer, water, and ligase, the tested reactions consist of: no sample (reactions 1 to 4), 20ng of sample (reaction 5 to 8), and 200ng of sample (reactions 9 to 12) in combination with (in reaction order) 1uM adapters, 0.2uM adapters, 0.04uM adapters, or 0.008uM adapters. Additional controls, in addition to buffer, water, and ligase, consist of the following, listed by reaction number: (13) 200ng of sample without adapters, (14) 200ng of sample and 1uM first adapter only, (15) 200ng of sample and 1uM second adapter only, (16) water only), (17) 1uM first adapter only, and (18) 1uM second adapter only. Ligation reactions are incubated at room temperature for 10 minutes. Ligation products are then subjected to an amplification step, with each amplification reaction containing 3uL water, 2uL 5X PCR buffer, 1uL 25mM $MgCl_2$, 1uL 10uM first primer, 1uL 10uM second primer, 0.5uL 10mM dNTPs, 0.5uM DMSO, 0.1uL Expand enzyme mix, 0.1uL Taq polymerase, and 1uL of one of the ligation reactions. Amplification reaction mixtures were then subjected to the following thermal cycling sequence: 1 cycle of 72°C for 2 minutes, 95°C for 2 minutes; 10 cycles of 95°C for 30 seconds, 60°C for 30 seconds, 72°C for 1 minute; 20 cycles of 95°C for 30 seconds, 60°C for 30 seconds, 72°C for 70°C seconds; 72°C for 7 minutes; hold at 10°C until proceeding. The first cycle in this process serves to extend the 3' ends of target polynucleotides using the adapters ligated to the 5' end as template (the "fill in" reaction), yielding a double-stranded DNA adapter tag. At the end of thermal cycling, 2uL of 6X loading dye is added to each reaction, and 5uL of the resulting mixture is loaded on a 2% agarose gel in TAE. Gels are imaged to illuminate the DNA products resulting from ligation and amplification.

[0079] A sample result is shown in Fig. 2A. The top half of Fig. 2A contains, in lanes from left to right, ladder, reactions 1 to 9, and ladder. The lower half of Fig. 2A contains, in lanes from left to right, ladder, reactions 10 to 18, and ladder. Lanes 1 to 4 and 13 to 18 show that both sample nucleic acid and both adapters are required for efficient amplification of target polynucleotides. Fig. 2B provides a side-by-side comparison of reactions 1 to 12, in order from left to right, in addition to lanes containing ladder. The results indicate that, under these conditions, amplified libraries can be obtained using a first and second hairpin adapter, that the higher sample amount decreases the formation of primer dimers, and amplification yield remains relatively constant with decreasing adapter input.

Example 3: Barcoded adapters and sample source identification

[0080] Nucleic acid is isolated from samples derived from 16 individuals using standard methods. Isolated polynucleotide samples are processed independently as in Example 1. Adapters are then ligated to target polynucleotides as in Example 2, with each sample being joined to a first adapter having a different barcode and a second adapter consisting of SEQ ID NO: 8. The first adapters are independently assigned to each of the samples, and have sequences provided by SEQ ID NOs: 11-26.

[0081] Target polynucleotides having 5' overhangs comprising the adapter sequences are then filled in by 3' end extension using the adapter sequences as template as in Example 2. Target polynucleotides are

IDT Exhibit 1020.5934

then also PCR amplified as in Example 2, using a pair of primers, one comprising SEQ ID NO: 84 and the other comprising SEQ ID NO: 85. The amplification products are then pooled and submitted for sequencing according to Illumina's Solexa sequencing platform (see e.g. www.illumina.com). Pooled sequencing data is then parsed on the basis of the barcodes contained in the sequencing reads to create 16 bins of sequencing data. The bins are then assembled as though each had been run independently, providing sorted and aligned sequencing data for 16 independent samples from a single pooled sequencing reaction.

Example 4: Use of hairpin adapters comprising heteroduplexes

[0082] The sample comprising target polynucleotides ("sample") used in this example is prepared as described in Example 1. First and second hairpin adapter oligonucleotides having stems involving both ends, forming blunt-ended structures, are ligated to target polynucleotides as in Example 2. With only target polynucleotides having 5' phosphates, only the 3' ends of the adapters are joined to the targets. As illustrated in Fig. 3, the hybridizable region of the 5' ends of the adapters comprise RNA, while the sequence to which the 5' ends hybridize comprises DNA. After ligation, RNase H cleaves the RNA of the RNA-DNA heteroduplexes, removing the secondary structure from the ligated adapters. DNA polymerase then extends the 3' ends of the target polynucleotides using the remaining sequence of the ligated adapters as template, a step that does not require any strand displacement. This step is performed as in Example 2, and can also be followed by an amplification step using primers that hybridize to sequences derived from the adapters. The resulting adapter-tagged oligonucleotides are then sequenced using sequencing primers that hybridize to sequences derived from the adapters. In both Fig. 3 and Fig. 4, S1 (half of stem 1) is hybridizable to S1' (the other half of stem 1), S2 (half of stem 2) is hybridizable to S2' (the other half of stem 2), L1 is the loop sequence of the first adapter oligonucleotide, and L2 is the loop sequence of the second adapter oligonucleotide. Similarly, in Fig. 5, S1 is hybridizable to S1', and L1 is the loop sequence of the adapter oligonucleotide. For the purpose of these illustrations, the sequences S1, S1', S2, and S2' correspond to sequences A, A', B, and B', respectively, as discussed above.

Example 5: Evaluation ligation efficiency of various hairpin adapter designs

[0083] In this example, hairpin adapter oligonucleotides having different nucleotide compositions were evaluated for efficiency of ligation to target polynucleotides. Each ligation reaction included target polynucleotides and a pair of adapters, with each member of the pair having a different sequence but sharing a specified characteristic. As illustrated in Figure 7, the various designs are, from left to right, blunt dU adapters, thymine-overhang adapters (ligated to blunt-ended target polynucleotides), thymine-overhang adapters (ligated to end repaired target polynucleotides modified to have 3' adenine single-base overhangs), duplex hairpin adapters, and blunt all-DNA adapters. Blunt dU adapters included a dinucleotide of deoxy uridine nucleotides at the 5' extreme end of the adapter loop (e.g. SEQ ID NO: 27

IDT Exhibit 1020.5935

and SEQ ID NO: 28). Treatment of ligated material with UDG + APE1 cleaves the U bases and opens the loop for a subsequent fill in reaction (the remaining stem dissociates at the 72°C temperature used in the fill-in reaction. Thymine-overhang adapters included an all-DNA sequence with a single thymine nucleotide 3' overhang (e.g. SEQ ID NO: 35 and SEQ ID NO 36). Duplex hairpin adapters included a first or second hairpin oligonucleotide having a stem and a 3' overhang (e.g. SEQ ID NO: 37 and SEQ ID NO: 38) hybridized to a short nucleotide (e.g. SEQ ID NO: 39), which hybridization included the 5' terminus of the short nucleotide and the 3' terminus of the hairpin oligonucleotide to form a stem effectively having a single-stranded break. Blunt all-DNA adapters consisted of DNA, with internal hybridization forming a blunt ended hairpin (e.g. SEQ ID NO: 40 and SEQ ID NO: 41). Example adapter sequences are provided by SEQ ID NOs: 27-43.

[0084] Human genomic DNA was fragmented as in Example 1. For end repair of fragmented genomic DNA, 52uL of 191ng/uL fragmented human genomic DNA was combined with 20uL 10X quick blunt buffer, 20uL 10X dNTPs, and 100uL water, which was mixed before further adding 8uL quick blunt enzyme mix. The end-repair reaction was incubated for 30 minutes at room temperature, and for 20 minutes at 75°C. For ligation to thymine-overhang adapters, 100uL of end-repaired DNA was modified to have single adenine nucleotide 3' overhangs ("tailed") by adding 2uL of 10mM dATP (final concentration of 0.2mM) and 8uL of Klenow (3'->5' exo minus), and incubating at 37°C for 30 minutes followed by 75°C for 20 minutes. Ligation reactions were prepared by combining 10uL of 2X ligation buffer, 4uL of end-repaired DNA or tailed DNA (approximately 200ng total), 0.2uL of each of a first and second adapter in a pair at concentrations of 10uM, and 5uL water, followed by mixing, adding 1uL of T4 DNA ligase, and incubating at room temperature for 10 minutes. For ligation reactions using blunt dU adapters, 1uL of a mixture of uracil DNA glycosylase (UDG) and apurinic endonuclease (APE) was added, followed by incubation at 37°C for 10 minutes. After ligation, and cleavage where indicated, duplicate reactions were prepared for fill-in of 5' overhangs by 3' end extension for each adapter type ligation reaction. One of each duplicate fill-in reaction was further amplified by PCR using a pair of amplification primers (SEQ ID NO: 42 and SEQ ID NO: 43), while the other of each duplicate was used in the determination of ligation efficiency. Each fill-in/amplification reaction contained 8uL of water, 2uL of 10X amplification buffer, 2uL of 25mM $MgCl_2$, 2uL of each amplification primer at 10uM concentrations, 2uL of one ligation reaction, 1uL of DMSO, 1uL of 10mM dNTPs, and 0.2uL of Taq polymerase enzyme. The fill-in/amplification reactions were incubated at 72°C for 2 minutes. Amplification included 20 cycles of 94°C for 30 seconds, 60°C for 30 seconds, and 72°C for 1 minute. Aliquots of amplification reactions were run on an agarose gel, the results of which are illustrated in Figure 7.

[0085] Ligation efficiency was measured by quantitative PCR (qPCR). Ligation efficiency is defined as the percent of target molecules that are added to the library construction as input that end up in the final, amplified library. It is determined by using an existing library of known composition and concentration as a standard. Dilutions of this library are used to create a standard curve in a qPCR reaction. For

IDT Exhibit 1020.5936

measurement of unknowns, a calculated portion of the target input is removed following end-repair, ligation, and fill in. The qPCR signal from this sample is plotted on the standard curve to establish the quantity of properly ligated molecules. The difference between the measured signal and the known input establishes the ligation efficiency. qPCR reaction mixes included12.5uL of 2X SYBR mix (Clontech Laboratories, Inc.), 0.5uL of each amplification primer at 10uM concentrations, 5uL of template (1/10 dilution of fill-in reaction, 1/100 dilution of fill-in reaction, library standard, or water for no template control), and 6.5uL of water. Amplification of the qPCR reactions was conducted using standard methods, and ligation efficiency for each adapter design is provide below the illustration of the respective design in Figure 7. In summary, for blunt dU adapters, thymine-overhang adapters (ligated to blunt-ended target polynucleotides), thymine-overhang adapters (ligated to end repaired target polynucleotides modified to have 3' adenine single-base overhangs), duplex hairpin adapters, and blunt all-DNA adapters, efficiencies were about 0.48%, 0.0035%, 0.20%, 0.22%, and 0.22%, respectively. All adapter pairs produced comparable PCR amplification products.

[0086] Examination of ligation products by agarose gel analysis indicated the presence of little or no adapter dimers. That amplification products contained target insert fragments of about the expected size was also confirmed. Figure 8 shows a gel of samples of various reactions, with lane contents from left to right as follows: end repaired human genomic DNA, blunt all-DNA adapters, end repaired and A-tailed DNA, thymine overhang adapters, ladder, ligated end repaired DNA without adapters, end repaired DNA ligated with blunt all-DNA adapters, end repaired and A-tailed DNA ligated without adapters, end repaired and A-tailed DNA ligated with thymine overhang adapters, and ladder.

[0087] In some embodiments, a first duplex adapter in a pair of duplex adapters includes a first hairpin oligonucleotide having a stem and a 3' overhang comprising a barcode hybridized to a short partner oligonucleotide comprising a sequence complementary to all or a portion of the 3' overhang, including the barcode. The duplex adapter comprising two oligonucleotides may have a 5' or 3' overhang, or a blunt end when both oligonucleotides in the duplex are hybridized. The first duplex adapter may be paired with a second duplex adapter that is the same or different from the first duplex adapter, and the second duplex adapter may or may not contain a barcode. In general, the second duplex adapter may include a hairpin oligonucleotide having a stem and a 3' overhang hybridized to a short nucleotide, such that the hybridized oligonucleotides together form an adapter having a 5' or 3' overhang, or a blunt end. Examples of first duplex adapters comprising a hairpin oligonucleotide having a barcode and paired with a short partner oligonucleotide includes the following sequence pairs: SEQ ID NO: 44 with SEQ ID NO: 45, SEQ ID NO: 46 with SEQ ID NO: 47, SEQ ID NO: 48 with SEQ ID NO: 49, SEQ ID NO: 50 with SEQ ID NO: 51, SEQ ID NO: 52 with SEQ ID NO: 53, SEQ ID NO: 54 with SEQ ID NO: 55, SEQ ID NO: 56 with SEQ ID NO: 57, SEQ ID NO: 58 with SEQ ID NO: 59, SEQ ID NO: 60 with SEQ ID NO: 61, SEQ ID NO: 62 with SEQ ID NO: 63, SEQ ID NO: 64 with SEQ ID NO: 65, SEQ ID NO: 66 with SEQ ID NO: 67, SEQ ID NO: 68 with SEQ ID NO: 69, SEQ ID NO: 70 with SEQ ID NO: 71, SEQ ID NO: 72 with SEQ ID NO: 73, and SEQ ID NO: 74 with SEQ ID NO: 75. In these sequences, barcodes

IDT Exhibit 1020.5937

are represented by the four bases at the 3' end of the hairpin oligonucleotide in each pair of oligonucleotides of the duplex adapter, and the complement of the barcode represented by the four bases at the 5' end of the short partner oligonucleotide in each pair of oligonucleotides of the duplex adapter. In general, each hairpin oligonucleotide of a pair is mixed with the corresponding short partner oligonucleotide in a ratio of 1:1.

Example 6: Evaluation ligation efficiency of hairpin adapters comprising RNA

[0088] In this example, hairpin adapter oligonucleotides having different nucleotide compositions were evaluated for efficiency of ligation to target polynucleotides, as in Example 5. Each ligation reaction included target polynucleotides and a pair of adapters, with each member of the pair having a different sequence but sharing a specified characteristic. The adapter pairs included blunt all-DNA adapters, and blunt RNA with DNA:DNA end adapters. Blunt all-DNA adapters consisted of DNA, with internal hybridization forming a blunt ended hairpin (SEQ ID NO: 76 and SEQ ID NO: 77). Blunt RNA with DNA:DNA end adapters included a stem having one strand with 10 RNA bases at the 5' end with five 5' terminal DNA bases, hybridized to a second strand of all DNA (SEQ ID NO: 80 and SEQ ID NO: 81). Amplification of ligation reactions using these adapters was performed using a pair of amplification primers (SEQ ID NO: 82 and SEQ ID NO: 83). Example adapter and amplification primer sequences are provided by SEQ ID NOs: 76-83.

[0089] Fragmented target polynucleotides were prepared as in Example 5. Fragmented DNA was end-repaired as in Example 1 by combining, for each reaction, 4.2uL of 47.5ng/uL fragmented genomic DNA, 1.25uL of 10X quick blunt buffer, 1.25uL of 1mM dNTP, 5.3uL water, mixing, and adding 0.5uL quick blunt enzyme. End-repair reactions were then incubated at room temperature (e.g. 20°C-27°C) for 30 minutes, and then incubated at 70°C for 10 minutes. Ligation reactions were prepared in duplicate, using a full 12.5uL end-repair reaction, combined with 12.5uL of 2X quick ligase buffer, 0.25uL of each adaptor in a pair at concentrations of 10uM, and 1.25uL of quick ligase. Ligation reactions were incubated at room temperature for 10 minutes prior to amplification. One ligation reaction in each duplicate was treated with RNase H in the amplification reaction mixture before beginning the amplification process. Both RNase H treated and untreated reactions were then subjected to 5' overhang fill-in and ligation product amplification. Samples not treated with RNase H included 59uL of water, 10uL of 10x PCR buffer, 3uL of 50mM $MgCl_2$, 5uL of each amplification primer at concentrations of 10uM, 5uL of DMSO, 2uL of 10mM dNTPs, 1uL of Taq polymerase, and 10uL of ligated template. Samples receiving RNase H treatment included 58uL of water, 10uL of 10x PCR buffer, 3uL of 50mM $MgCl_2$, 5uL of each amplification primer at concentrations of 10uM, 5uL of DMSO, 2uL of 10mM dNTPs, 1uL of Taq polymerase, 1ul of RNase H, and 10uL of ligated template. For samples receiving RNase H treatment, thermal cycling for amplification was preceded by incubation at 37°C for 10 minutes (non-amplified, RNase H treated samples used as a baseline for quantification included additional steps of 72°C for 2 minutes, and a 10°C hold step). Amplification reaction mixtures were then subjected to the

IDT Exhibit 1020.5938

following thermal cycling sequence for fill-in and amplification: 1 cycle of 72°C for 2 minutes; 20 cycles of 94°C for 45 seconds, 55°C for 30 seconds, and 72°C for 90 seconds; 1 cycle of 72°C for 7 minutes; and rest at 10°C. A 2% agarose gel containing 8uL samples of PCR amplification reactions is shown in Figure 9, with lanes from left to right corresponding to blunt RNA with DNA:DNA end adapter ligation products, blunt all-DNA adapter ligation products, and DNA ladder. A schematic diagram of the ligation between the 3' end of the adapter and 5' end of the target, RNase H treatment (where applicable), fill-in, and amplification reactions at one end of a target DNA is provided in Figure 10.

[0090] Ligation efficiency was measured for each pair of adapters with or without RNase H treatment as described in Example 5. Each qPCR reaction in this example comprised 5uL of 2X SYBR Green Mix, 0.4uL of each amplification primer, 2.2uL water, and 2uL of diluted ligation reaction, for a total of 10uL for each qPCR reaction. Ligation efficiencies for blunt all-DNA adapters with RNase H treatment, blunt all-DNA adapters without RNase H treatment, blunt RNA with DNA:DNA end adapters with RNase H treatment, and blunt RNA with DNA:DNA end adapters without RNase H treatment were 0.20%, 0.37%, 0.28%, and 0.13%, respectively. Successfully ligated and amplified fragments may be used as a next-generation sequence library.

[0091] 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 1020.5939

CLAIMS

WHAT IS CLAIMED IS:

1. A method for multiplex sequencing comprising sequencing a plurality of target polynucleotides in a single reaction chamber, wherein said target polynucleotides are from two or more different samples; and identifying the sample from which each of said sequenced target polynucleotides is derived with an accuracy of at least 95% based on a single barcode contained in the sequence of said target polynucleotide.

2. The method of claim 1, wherein said target polynucleotides comprise one or more sequences with which the sequencing reaction is calibrated.

3. The method of claim 1, wherein each barcode differs from every other barcode at at least three nucleotide positions.

4. The method of claim 1, wherein said identification is accurate after the mutation or deletion of a nucleotide in said barcode.

5. A method of producing adapter-tagged target polynucleotides from a plurality of independent samples, the method comprising:

a) providing a plurality of first adapter oligonucleotides, wherein each of said first adapter 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; and

b) joining at least one of said first adapter oligonucleotides to said target polynucleotides of each of said samples, such that no barcode sequence is joined to said target polynucleotides of more than one of said samples.

6. The method of claim 5, further comprising (c) joining at least one of a plurality of second adapter oligonucleotides to said target polynucleotides of each of said samples from step (b), such that at least some of said target polynucleotides comprise said first adapter oligonucleotide at one end and said second adapter oligonucleotide at the other end.

7. The method of claim 6, further comprising pooling the target polynucleotides from step (c).

8. The method of claim 7, further comprising sequencing one or more of said polynucleotides in said pool.

9. The method of claim 8, further comprising identifying the sample from which a target polynucleotide is derived based on a barcode sequence to which it is joined.

10. The method of claims 5 or 6, wherein one or more of said adapter oligonucleotides comprises SEQ ID NO: 1.

11. The method of claims 5 or 6, wherein one or more of said adapter oligonucleotides comprises SEQ ID NO: 2.

IDT Exhibit 1020.5940

12. The method of claims 5 or 6, wherein one or more of said adapter oligonucleotides comprise a hairpin structure.

13. The method of claims 5 or 6, wherein one or more of said adapter oligonucleotides comprise an oligonucleotide duplex.

14. The method of claims 1 or 5, wherein said barcode sequences are at least 3 nucleotides in length.

15. The method of claims 1 or 7, wherein said target polynucleotides are pooled based on said barcode sequences, such that all four bases are evenly represented at one or more positions along each barcode in said pool.

16. The method of claims 1 or 5, wherein said target polynucleotides comprise fragmented sample polynucleotides.

17. The method of claim 16, wherein said fragmentation comprises subjecting said sample polynucleotides to acoustic sonication.

18. The method of claim 16, wherein said fragmentation comprises treating said sample polynucleotides with one or more restriction endonucleases.

19. The method of claim 16, wherein said fragmentation comprises treating said sample polynucleotides with one or more enzymes under conditions suitable for said one or more enzymes to generate random double-stranded nucleic acid breaks.

20. The method of claim 19, wherein said one or more enzymes are selected from the group consisting of: DNase I, Fragmentase, and variants thereof.

21. The method of claim 16, wherein said fragments have an average length of 10 to 10000 nucleotides.

22. The method of claim 16, wherein said fragments have an average length of 100-2500 nucleotides.

23. The method of claim 16, wherein said fragments have an average length of 50-500 nucleotides.

24. The method of claims 12 or 13, further comprising performing the step of extending one or more 3' ends of said target polynucleotides, using said one or more joined adapter oligonucleotides as template.

25. The method of claim 24, further comprising amplifying said target polynucleotides after said extending step using a first primer and a second primer, wherein said first primer comprises a sequence that is hybridizable to at least a portion of the complement of one or more of said first adapter oligonucleotides, and further wherein said second primer comprises a sequence that is hybridizable to at least a portion of the complement of one or more of said second adapter oligonucleotides.

26. The method of claim 25, wherein one or more of said primers comprises SEQ ID NO: 1.

27. The method of claim 25, wherein one or more of said primers comprises SEQ ID NO: 2.

IDT Exhibit 1020.5941

28. The method of claim 6, wherein each of said second adapter 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.

29. The method of claim 28, wherein pairs of said first and second adapter oligonucleotides comprise different barcode sequences.

30. The method of claim 28, wherein pairs of said first and second adapter oligonucleotides comprise the same barcode sequence.

31. The method of claims 1or 5, wherein said target polynucleotides comprise genomic DNA.

32. The method of claims 1 or 5, wherein said target polynucleotides comprise mitochondrial DNA, chloroplast DNA, plasmid DNA, bacterial artificial chromosomes, yeast artificial chromosomes, or a combination thereof.

33. The methods of claims 1 or 5, wherein said target polynucleotides comprise cDNA.

34. The method of claims 1or 5, wherein said samples comprise target polynucleotides produced by a primer extension reaction.

35. The method of claim 8, wherein said sequencing comprises extension of a sequencing primer comprising a sequence hybridizable to at least a portion of the complement of said first adapter oligonucleotide and/or said second adapter oligonucleotide.

36. The method of claim 35, wherein said sequencing primer comprises SEQ ID NO: 1 or SEQ ID NO: 2.

37. The method of claims 1 or 8, wherein said sequencing comprises a calibration step, wherein said calibration is based on each of the nucleotides at one or more nucleotide positions in said barcode sequences.

38. The method of claims 1 or 5, wherein each of said samples comprises less than 500ng of nucleic acid.

39. The method of claims 1 or 5, wherein said plurality of barcode sequences includes sequences selected from the group consisting of: AAA, TTT, CCC, and GGG.

40. The method of claims 1 or 5, wherein said plurality of barcode sequences includes sequences selected from the group consisting of: AAAA, CTGC, GCTG, TGCT, ACCC, CGTA, GAGT, TTAG, AGGG, CCAT, GTCA, TATC, ATTT, CACG, GGAC, and TCGA.

41. The method of claims 1 or 5, wherein said plurality of barcode sequences includes sequences selected from the group consisting of: AAAAA, AACCC, AAGGG, AATTT, ACACG, ACCAT, ACGTA, ACTGC, AGAGT, AGCTG, AGGAC, AGTCA, ATATC, ATCGA, ATGCT, ATTAG, CAACT, CACAG, CAGTC, CATGA, CCAAC, CCCCA, CCGGT, CCTTG, CGATA, CGCGC, CGGCG, CGTAT, CTAGG, CTCTT, CTGAA, CTTCC, GAAGC, GACTA, GAGAT, GATCG, GCATT, GCCGG, GCGCC, GCTAA, GGAAG, GGCCT, GGGGA, GGTTC, GTACA,

IDT Exhibit 1020.5942

GTCAC, GTGTG, GTTTT, TAATG, TACGT, TAGCA, TATAC, TCAGA, TCCTC, TCGAG, TCTCT, TGACC, TGCAA, TGGTT, TGTGG, TTAAT, TTCCG, TTGGC, and TTTTA.

42. A composition configured for multiplex sequencing comprising: 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 polynucletides 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.

43. The composition of claim 42, 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.

44. A composition useful in the generation of adapter-tagged target polynucleotides, the composition comprising a plurality of first adapter oligonucleotides, wherein each of said first adapter 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.

45. The composition of claim 44, further comprising a plurality of second adapter oligonucleotides.

46. The composition of claims 42 or 44, wherein said target polynucleotides are contained in a flow cell.

47. The composition of claims 44 or 45, wherein one or more of said adapter oligonucleotides comprises SEQ ID NO: 1.

48. The composition of claims 44 or 45, wherein one or more of said adapter oligonucleotides comprises SEQ ID NO: 2.

49. The composition of claims 44 or 45, wherein one or more of said adapter oligonucleotides comprise a hairpin structure.

50. The composition of claims 44 or 45, wherein one or more of said adapter oligonucleotides comprise an oligonucleotide duplex.

51. The composition of claims 42or 44, wherein said barcode sequences are at least 3 nucleotides in length.

52. The composition of claim 44, wherein said first adapter oligonucleotides are grouped in multiples of four such that all four bases are evenly represented at each position along each barcode.

53. The composition of claim 45, wherein each of said second adapter 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 1020.5943

54. The composition of claim 53, wherein pairs of said first and second adapter oligonucleotides comprise the same barcode sequence.

55. The composition of claim 53, wherein pairs of said first and second adapter oligonucleotides comprise different barcode sequences.

56. The composition of claims 49 or 50, further comprising a first primer and a second primer, wherein said first primer comprises a sequence that is hybridizable to at least a portion of the complement of one or more of said first adapter oligonucleotides, and further wherein said second primers comprise a sequence that is hybridizable to at least a portion of the complement of one or more of said second adapter oligonucleotides.

57. The composition of claim 56, wherein one of said primers comprises SEQ ID NO: 1.

58. The composition of claim 56, wherein one of said primers comprises SEQ ID NO: 2.

59. The composition of claim 49 or 50, further comprising a sequencing primer comprising a sequence hybridizable to at least a portion of the complement of said first adapter oligonucleotide and/or said second adapter oligonucleotide.

60. The composition of claims 42or 44, wherein said plurality of barcode sequences includes sequences selected from the group consisting of: AAA, TTT, CCC, and GGG.

61. The composition of claims 42or 44, wherein said plurality of barcode sequences include sequences selected from the group consisting of: AAAA, CTGC, GCTG, TGCT, ACCC, CGTA, GAGT, TTAG, AGGG, CCAT, GTCA, TATC, ATTT, CACG, GGAC, and TCGA.

62. The composition of claims 42or 44, wherein said plurality of barcode sequences include sequences selected from the group consisting of: AAAAA, AACCC, AAGGG, AATTT, ACACG, ACCAT, ACGTA, ACTGC, AGAGT, AGCTG, AGGAC, AGTCA, ATATC, ATCGA, ATGCT, ATTAG, CAACT, CACAG, CAGTC, CATGA, CCAAC, CCCCA, CCGGT, CCTTG, CGATA, CGCGC, CGGCG, CGTAT, CTAGG, CTCTT, CTGAA, CTTCC, GAAGC, GACTA, GAGAT, GATCG, GCATT, GCCGG, GCGCC, GCTAA, GGAAG, GGCCT, GGGGA, GGTTC, GTACA, GTCAC, GTGTG, GTTTT, TAATG, TACGT, TAGCA, TATAC, TCAGA, TCCTC, TCGAG, TCTCT, TGACC, TGCAA, TGGTT, TGTGG, TTAAT, TTCCG, TTGGC, and TTTTA.

63. A kit useful in the generation of adapter-tagged target polynucleotides, the kit comprising a plurality of first adapter oligonucleotides, wherein each of said first adapter 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, and instructions for using the same.

64. The kit of claim 63, further comprising a plurality of second adapter oligonucleotides.

65. The kit of claims 63 or 64, wherein one or more of said adapter oligonucleotides comprises SEQ ID NO: 1.

66. The kit of claims 63 or 64, wherein one or more of said adapter oligonucleotides comprises SEQ ID NO: 2.

IDT Exhibit 1020.5944

67. The kit of claims 63 or 64, wherein one or more of said adapter oligonucleotides comprise a hairpin structure.

68. The kit of claims 63 or 64, wherein one or more of said adapter oligonucleotides comprise an oligonucleotide duplex.

69. The kit of claim 63, wherein said barcode sequences are at least 3 nucleotides in length.

70. The kit of claim 63, wherein said first adapter oligonucleotides are grouped in multiples of four such that all four bases are evenly represented at each position along each barcode.

71. The kit of claim 64, wherein each of said second adapter 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.

72. The kit of claim 71, wherein pairs of said first and second adapter oligonucleotides comprise the same barcode sequence.

73. The kit of claim 71, wherein pairs of said first and second adapter oligonucleotides comprise different barcode sequences.

74. The kit of claims 67 or 68, further comprising a first primer and a second primer, wherein said first primer comprises a sequence that is hybridizable to at least a portion of the complement of one or more of said first adapter oligonucleotides, and further wherein said second primers comprise a sequence that is hybridizable to at least a portion of the complement of one or more of said second adapter oligonucleotides.

75. The kit of claim 74, wherein one of said primers comprises SEQ ID NO: 1.

76. The kit of claim 74, wherein one of said primers comprises SEQ ID NO: 2.

77. The kit of claims 67 or 68, further comprising a sequencing primer comprising a sequence hybridizable to at least a portion of the complement of said first adapter oligonucleotide and/or said second adapter oligonucleotide.

78. The kit of claim 77, wherein said sequencing primer comprises SEQ ID NO:1 or SEQ ID NO: 2.

79. The kit of claim 63, further comprising one or more of: (a) a DNA ligase, (b) a DNA-dependent DNA polymerase, (c) an RNA-dependent DNA polymerase, (d) random primers, (e) primers comprising at least 4 thymidines at the 3' end, (f) a DNA endonuclease, (g) a DNA-dependent DNA polymerase having 3' to 5' exonuclease activity, (h) a plurality of primers, each primer having one of a plurality of selected sequences, (i) a DNA kinase, (j) a DNA exonuclease, (k) magnetic beads, (l) an enzyme comprising RNase H activity, (m) an RNA ligase, and (n) one or more buffers suitable for one or more of the elements contained in said kit.

80. The kit of claim 63, wherein said plurality of barcode sequences includes sequences selected from the group consisting of: AAA, TTT, CCC, and GGG.

IDT Exhibit 1020.5945

81. The kit of claim 63, wherein said plurality of barcode sequences include sequences selected from the group consisting of: AAAA, CTGC, GCTG, TGCT, ACCC, CGTA, GAGT, TTAG, AGGG, CCAT, GTCA, TATC, ATTT, CACG, GGAC, and TCGA.

82. The kit of claim 63, wherein said plurality of barcode sequences include sequences selected from the group consisting of: AAAAA, AACCC, AAGGG, AATTT, ACACG, ACCAT, ACGTA, ACTGC, AGAGT, AGCTG, AGGAC, AGTCA, ATATC, ATCGA, ATGCT, ATTAG, CAACT, CACAG, CAGTC, CATGA, CCAAC, CCCCA, CCGGT, CCTTG, CGATA, CGCGC, CGGCG, CGTAT, CTAGG, CTCTT, CTGAA, CTTCC, GAAGC, GACTA, GAGAT, GATCG, GCATT, GCCGG, GCGCC, GCTAA, GGAAG, GGCCT, GGGGA, GGTTC, GTACA, GTCAC, GTGTG, GTTTT, TAATG, TACGT, TAGCA, TATAC, TCAGA, TCCTC, TCGAG, TCTCT, TGACC, TGCAA, TGGTT, TGTGG, TTAAT, TTCCG, TTGGC, and TTTTA.

83. A method of producing adapter-tagged target polynucleotides, the method comprising:

a) providing a plurality of first adapter oligonucleotides, wherein each of said first adapter oligonucleotides comprises a 5' end comprising sequence A and a 3' end comprising sequence A', and further wherein A is hybridizable to A', one of A or A' comprises DNA, and the other of A or A' comprises RNA and 5 or more terminal DNA nucleotides; and,

b) joining at least one of said first adapter oligonucleotides to at least one of said target polynucleotides.

84. The method of claim 83, further comprising the step of cleaving RNA with an enzyme that cleaves RNA from an RNA-DNA heteroduplex.

85. The method of claim 84, further comprising performing the step of extending one or more 3' ends of said target polynucleotides, using said one or more joined adapter oligonucleotides as template.

86. The method of claim 83, further comprising joining at least one of a plurality of second adapter oligonucleotides to said target polynucleotides of each of said samples from step (b), such that at least one of said target polynucleotides comprises said first adapter oligonucleotide at one end and said second adapter oligonucleotide at the other end.

87. The method of claim 86, wherein each of said second adapter oligonucleotides comprises a 5' end comprising sequence B and a 3' end comprising sequence B', and further wherein B is hybridizable to B', one of B or B' comprises DNA, and the other of B or B' comprises RNA and 5 or more terminal DNA nucleotides.

88. The method of claim 83, wherein each of said first adapter oligonucleotides comprises a barcode sequence.

89. A composition useful in the generation of adapter-tagged target polynucleotides, the composition comprising a plurality of first adapter oligonucleotides, wherein each of said first adapter oligonucleotides comprises a 5' end comprising sequence A and a 3' end comprising sequence A', and

IDT Exhibit 1020.5946

further wherein A is hybridizable to A', one of A or A' comprises DNA, and the other of A or A' comprises RNA and 5 or more terminal DNA nucleotides.

90. The composition of claim 89, further comprising a plurality of second adapter oligonucleotides, wherein each of said second adapter oligonucleotides comprises a 5' end comprising sequence B and a 3' end comprising sequence B', and further wherein B is hybridizable to B', one of B or B' comprises DNA, and the other of B or B' comprises RNA and 5 or more terminal DNA nucleotides.

91. A kit useful in the generation of adapter-tagged target polynucleotides, the kit comprising a plurality of first adapter oligonucleotides, wherein each of said first adapter oligonucleotides comprises a 5' end comprising sequence A and a 3' end comprising sequence A', and further wherein A is hybridizable to A', one of A or A' comprises DNA, and the other of A or A' comprises RNA and 5 or more terminal DNA nucleotides.

92. The kit of claim 91, further comprising a plurality of second adapter oligonucleotides, wherein each of said second adapter oligonucleotides comprises a 5' end comprising sequence B and a 3' end comprising sequence B', and further wherein B is hybridizable to B', one of B or B' comprises DNA, and the other of B or B' comprises RNA and 5 or more terminal DNA nucleotides.

IDT Exhibit 1020.5947

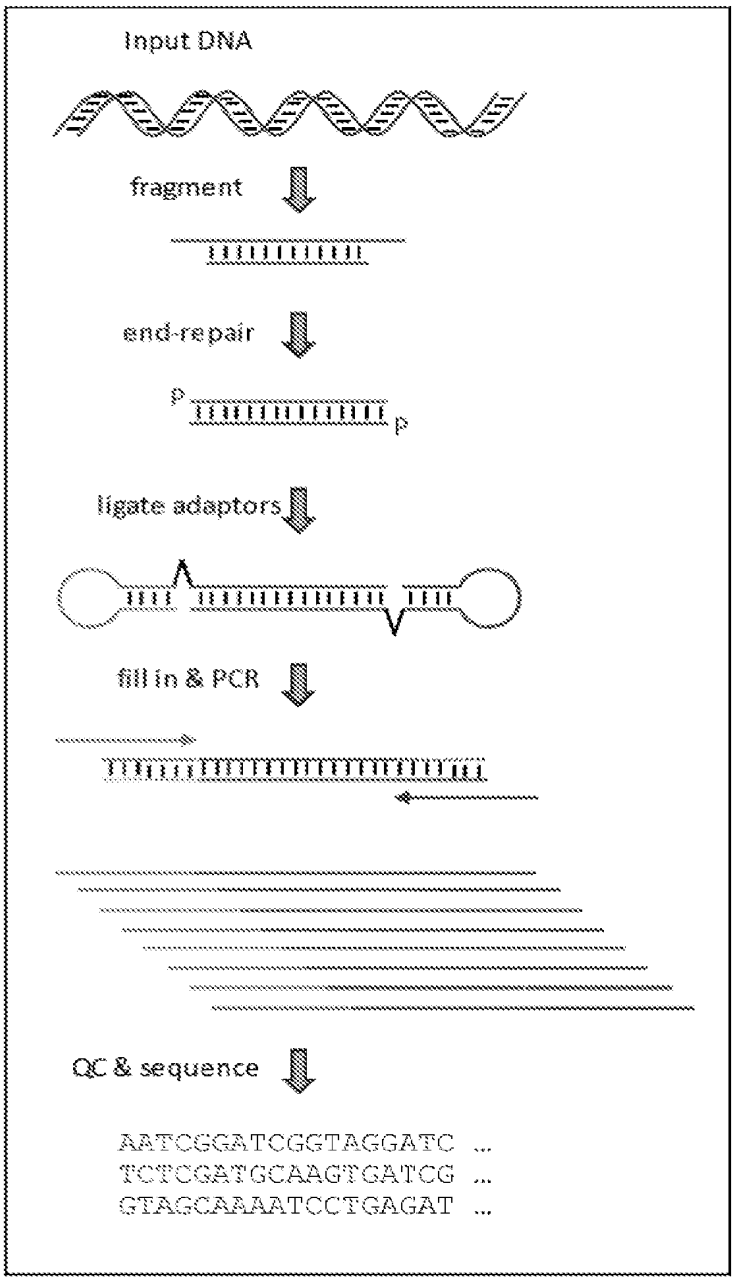

Input DNA
fragment
end-repair
P
P
ligate adaptors
fill in & PCR
QC & sequence
AATCGGATCGGTAGGATC ...
TCTCGATGCAAGTGATCG ...
GTAGCAAAATCCTGAGAT ...

Figure 1

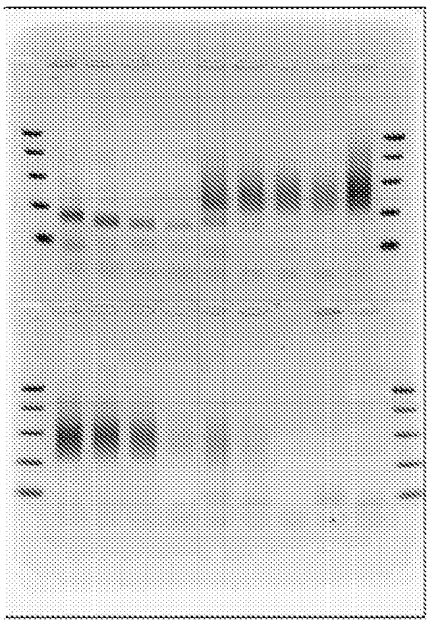
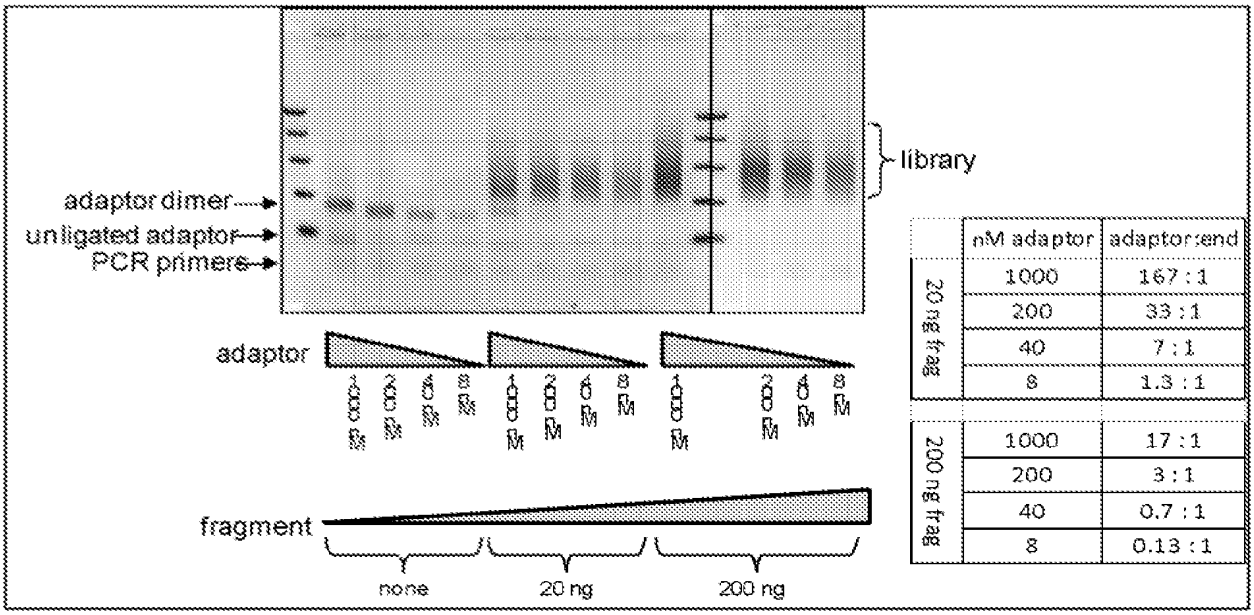


Figure 2A



	nM adaptor	adaptor:end
20 ng frag	1000	167 : 1
	200	33 : 1
	40	7 : 1
	8	1.3 : 1
200 ng frag	1000	17 : 1
	200	3 : 1
	40	0.7 : 1
	8	0.13 : 1

Figure 2B

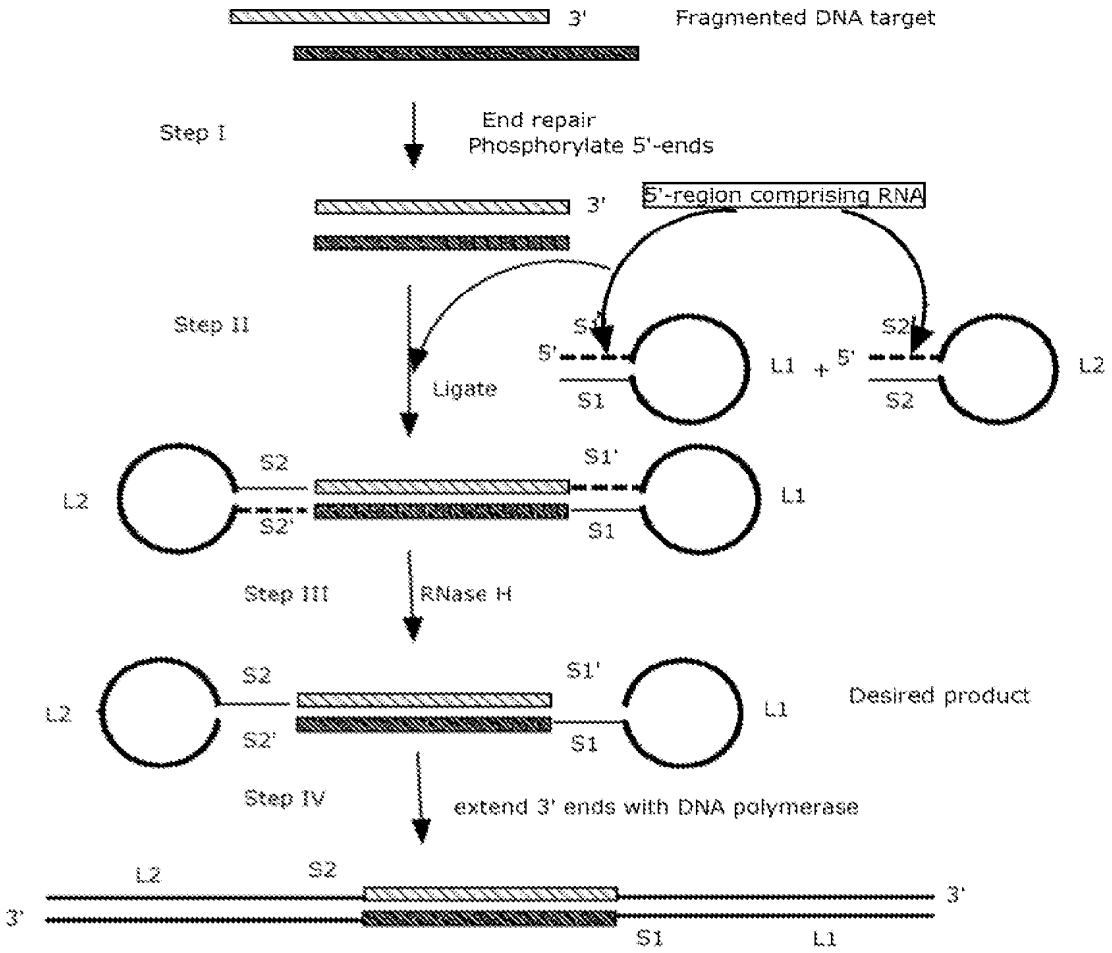

3'
Fragmented DNA target
Step I
End repair
Phosphorylate 5'-ends
3'
5'-region comprising RNA
Step II
S1'
S2'
5'
L1
+
5'
L2
Ligate
S1
S2
S2
S1'
L2
L1
S2'
S1
Step III
RNase H
S2
S1'
L2
L1
Desired product
S2'
S1
Step IV
extend 3' ends with DNA polymerase
L2
S2
3'
3'
S1
L1

Figure 3

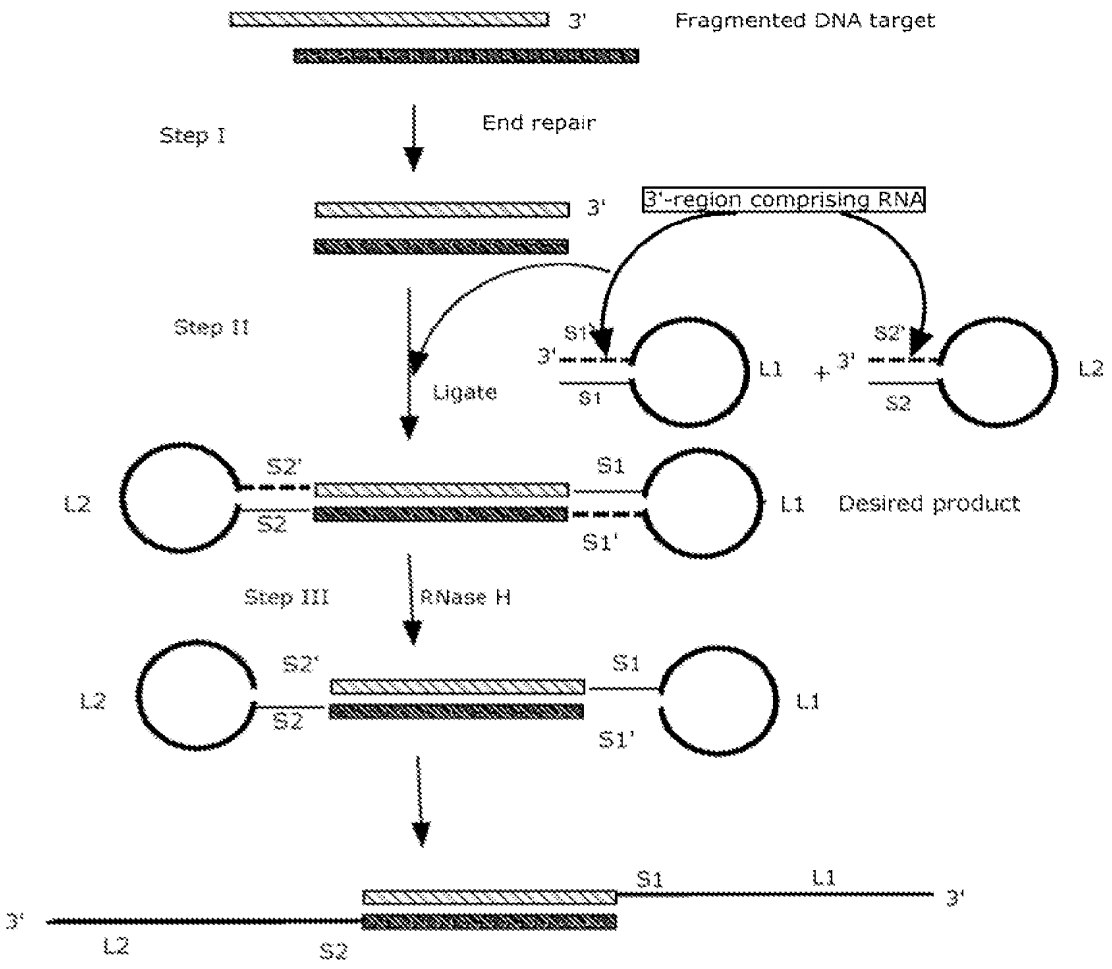

3'
Fragmented DNA target
Step I
End repair
3'
3'-region comprising RNA
Step II
S1'
3'
L1
+ 3'
S2'
L2
S1
S2
Ligate
S2'
S1
L2
L1
Desired product
S2
S1'
Step III
RNase H
S2'
S1
L2
L1
S2
S1'
S1
L1
3'
3'
L2
S2

Figure 4

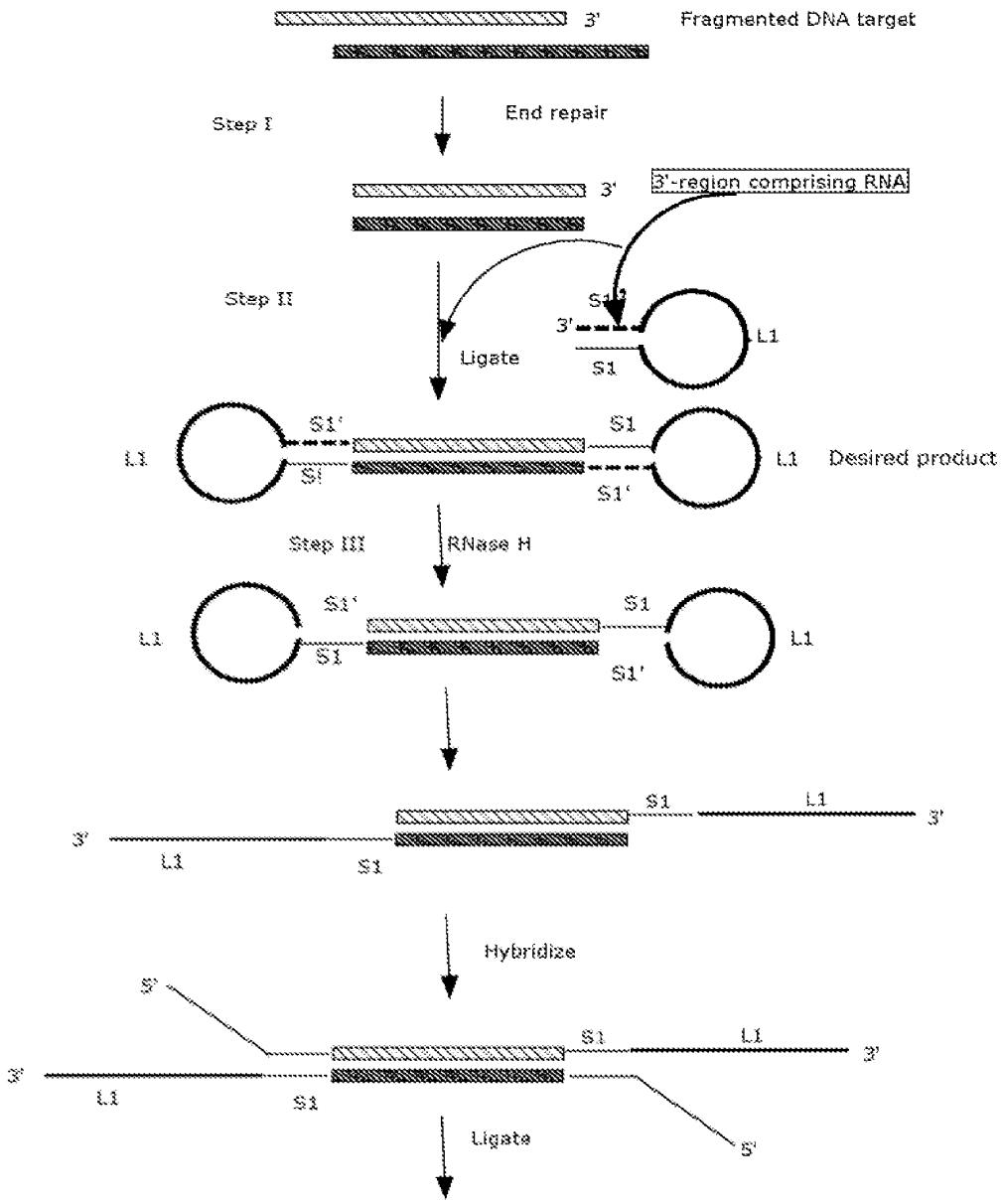

3'
Fragmented DNA target
Step I
End repair
3'
3'-region comprising RNA
Step II
S1'
3'
L1
S1
Ligate
S1'
S1
L1
L1
Desired product
S1'
Step III
RNase H
S1'
S1
L1
S1
L1
S1'
S1
L1
3'
3'
L1
S1
Hybridize
5'
S1
L1
3'
3'
L1
S1
Ligate
5'
Insert targets with unique non-complementary ssDNA 3'- and 5'- ends

Figure 5

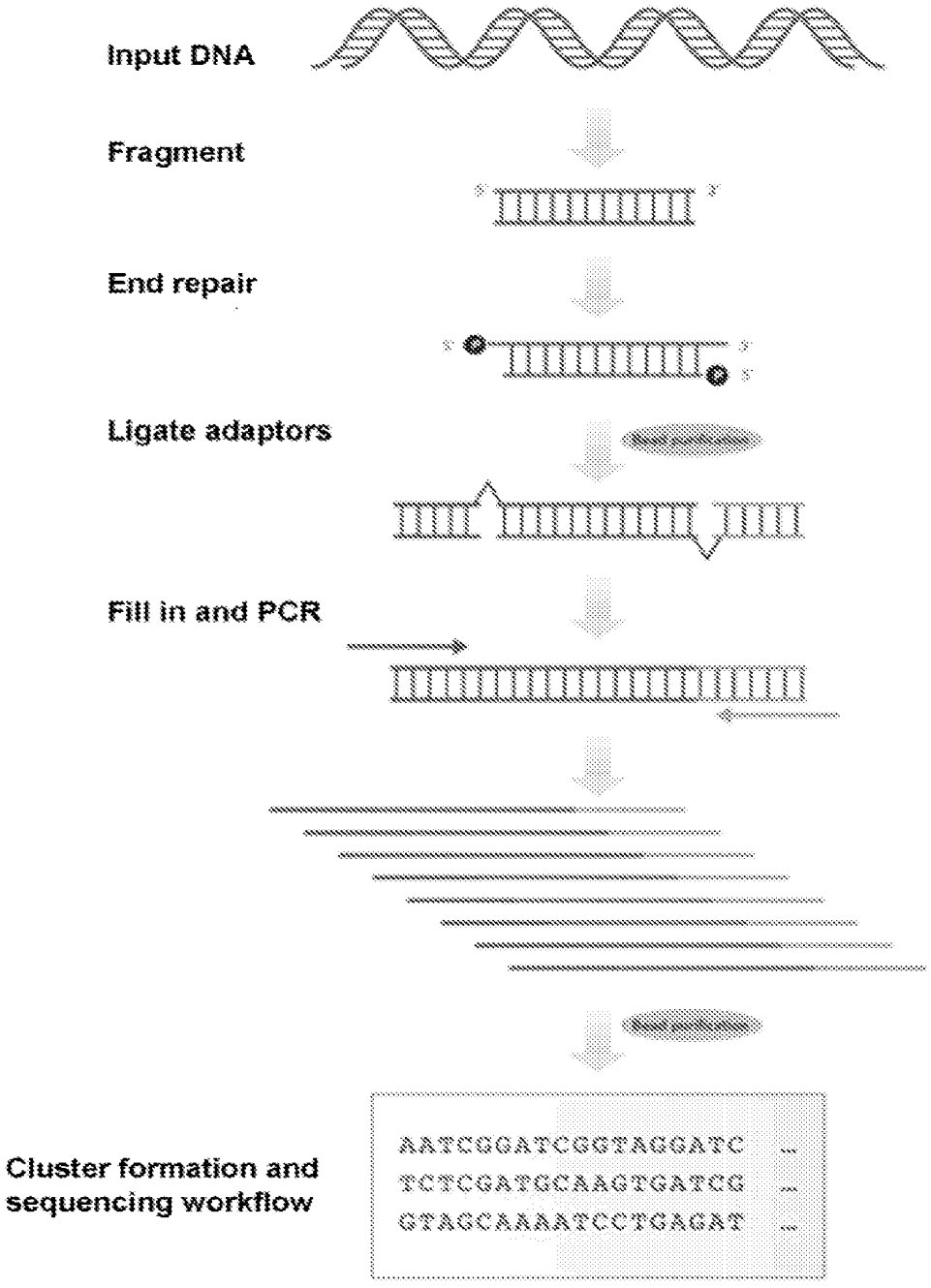
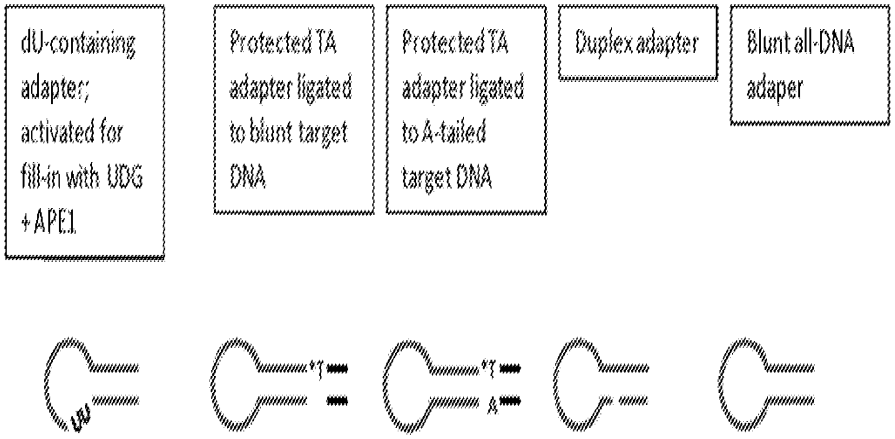

Input DNA
Fragment
End repair
Ligate adaptors
Fill in and PCR
Cluster formation and sequencing workflow
AATCGGATCGGTAGGATC ...
TCTCGATGCAAGTGATCG ...
GTAGCAAAATCCTGAGAT ...

Figure 6


dU-containing adapter; activated for fill-in with UDG + APE1
Protected TA adapter ligated to blunt target DNA
Protected TA adapter ligated to A-tailed target DNA
Duplex adapter
Blunt all-DNA adaper

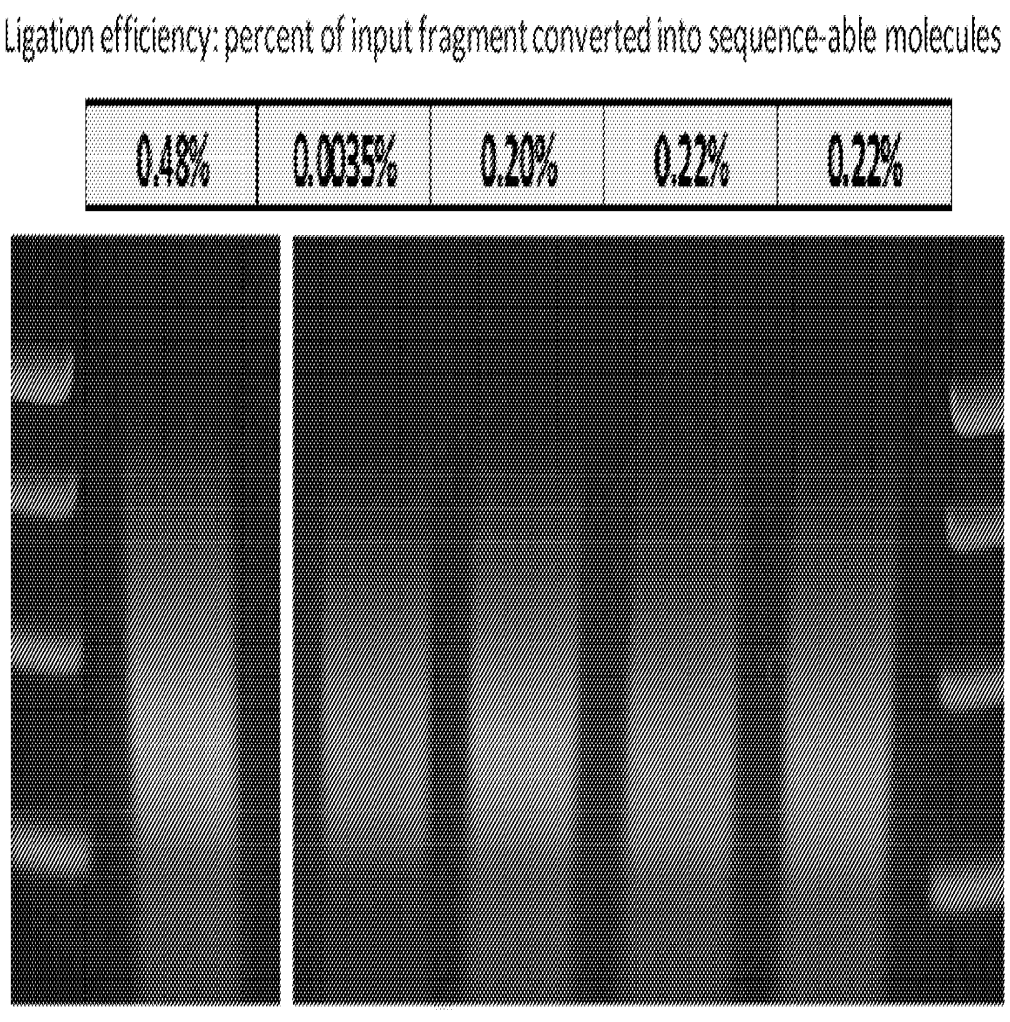

Ligation efficiency: percent of input fragment converted into sequence-able molecules
0.48%
0.0035%
0.20%
0.22%
0.22%

Figure 7

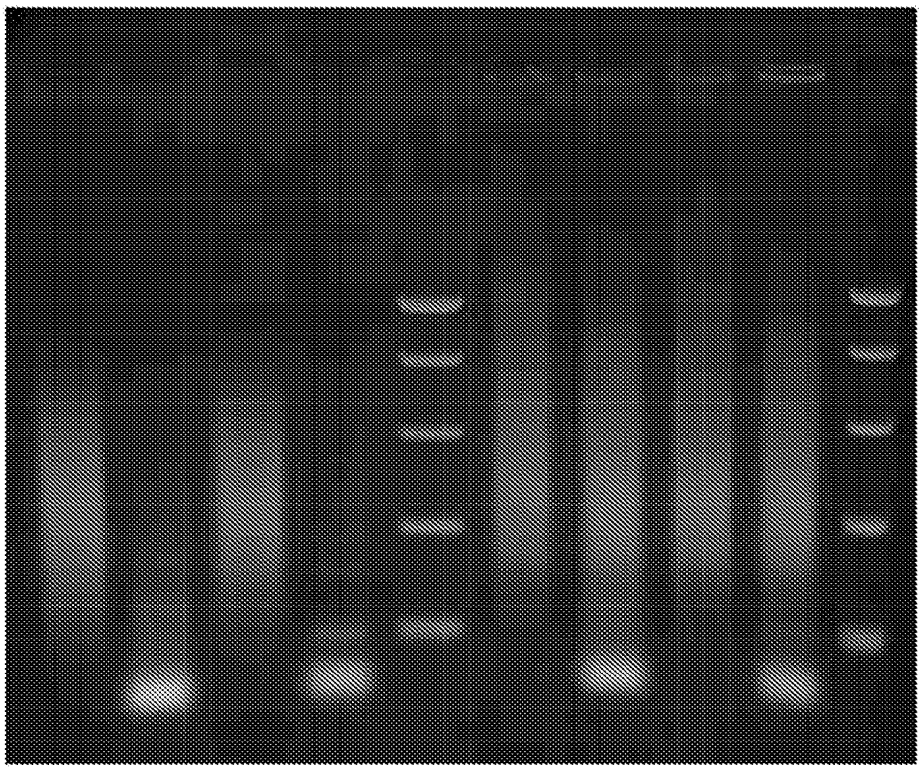


Figure 8

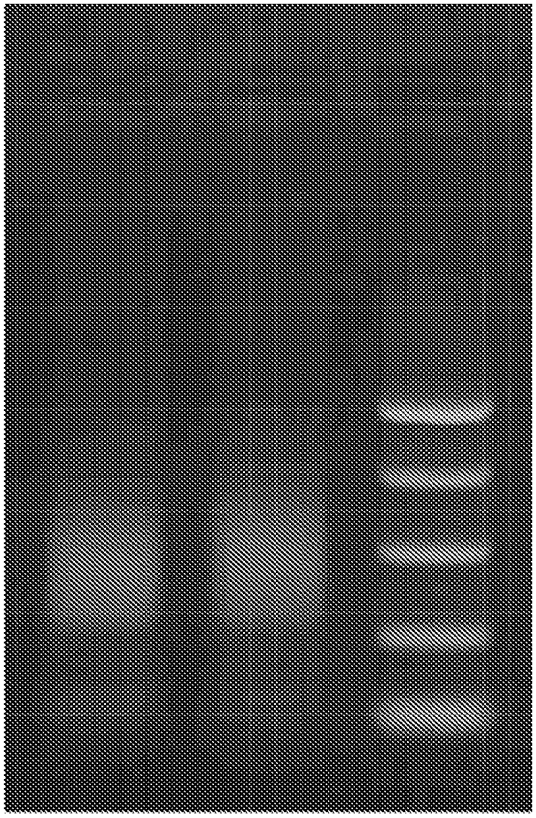


Figure 9

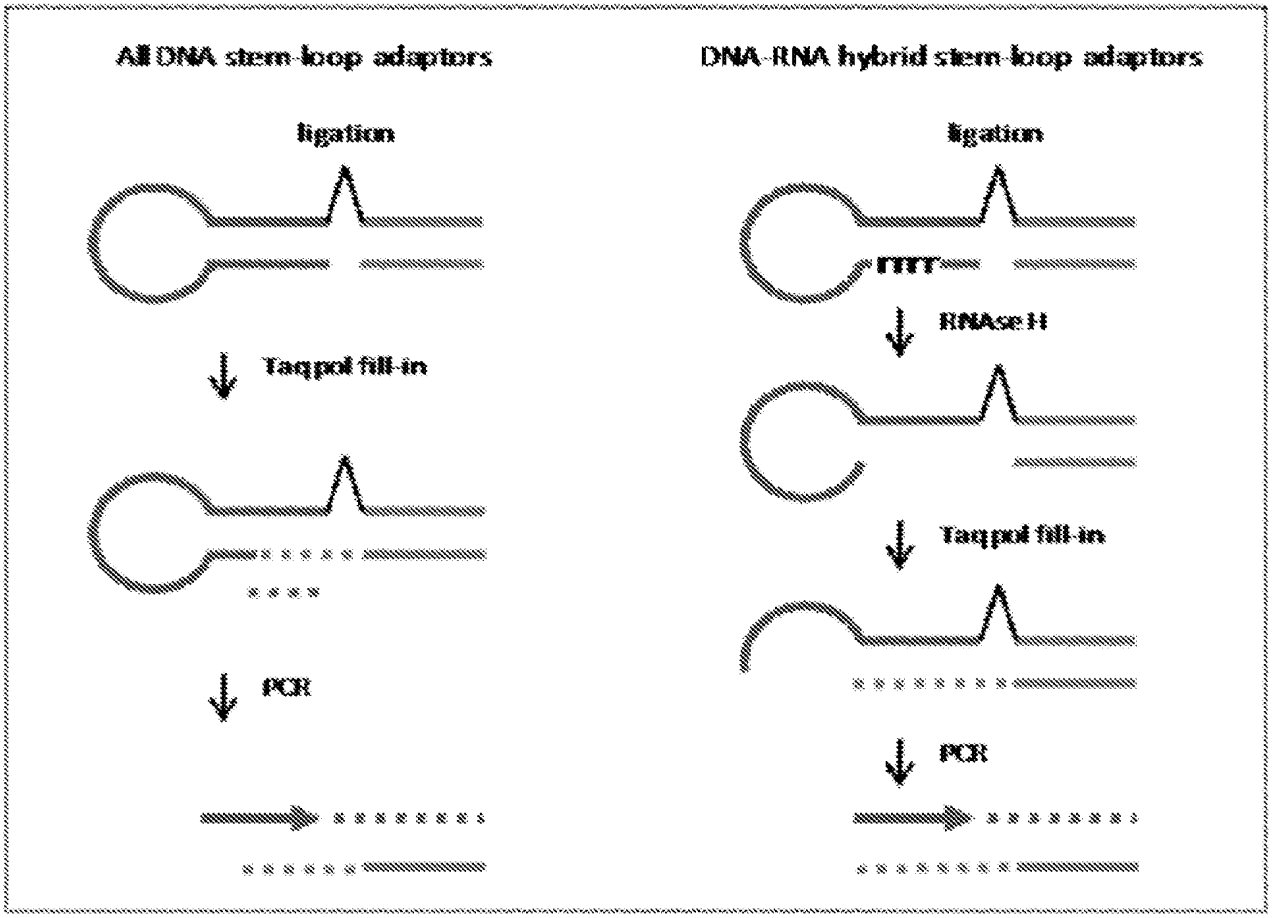

All DNA stem-loop adaptors
ligation
Taqpol fill-in
PCR
DNA-RNA hybrid stem-loop adaptors
ligation
RNAse H
Taqpol fill-in
PCR

Figure 10

Espacenet

Bibliographic data: WO2007057652 (A1) — 2007-05-24

METHOD OF TARGET ENRICHMENT

Inventor(s): GORMLEY NIALL ANTHONY [GB]; WEST JOHN STEPHEN [GB] ± (GORMLEY, NIALL ANTHONY, ; WEST, JOHN STEPHEN)

Applicant(s): SOLEXA LTD [GB]; GORMLEY NIALL ANTHONY [GB]; WEST JOHN STEPHEN [GB] ± (SOLEXA LIMITED, ; GORMLEY, NIALL ANTHONY, ; WEST, JOHN STEPHEN)

Classification: **- international:** ***C12N15/10; C12Q1/68***
- cooperative: **C12N15/1006**; **C12Q1/6806**; **C12Q1/6837**; C12Q1/6869 more

Application number: WO2006GB04244 20061114 Global Dossier

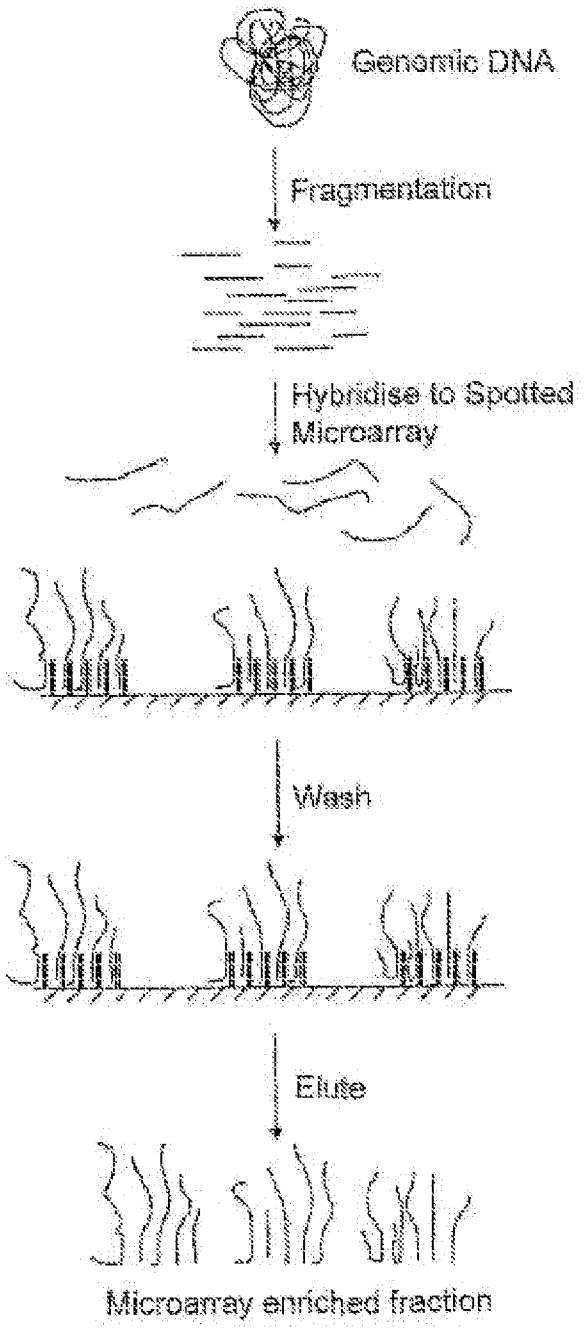
Priority number (s): US20050736785P 20051115 ; US20060837108P 20060810

Also published as: EP1957667 (A1) US2007141604 (A1)

Abstract of WO2007057652 (A1)

The invention relates to a method of enriching for specific target sequences using libraries of oligonucleotides, such as microarrays. In particular, the invention relates to a method of enriching for specific nucleic acid target sequences in a population of nucleic acid sequences in order to provide targets for sequencing.

IDT Exhibit 1020.5958



IDT Exhibit 1020.5959

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
24 May 2007 (24.05.2007)

PCT

(10) International Publication Number
WO 2007/057652 A1

(51) International Patent Classification:
C12Q 1/68 (2006.01) *C12N 15/10* (2006.01)

(21) International Application Number:
PCT/GB2006/004244

(22) International Filing Date:
14 November 2006 (14.11.2006)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/736,785 15 November 2005 (15.11.2005) US
60/837,108 10 August 2006 (10.08.2006) US

(71) Applicant *(for all designated States except US)*: **SOLEXA LIMITED** [GB/GB]; Chesterford Research Park, Little Chesterford, Saffron Walden Essex CB10 1XL (GB).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **GORMLEY, Niall Anthony** [IE/GB]; Solexa Limited, Chesterford Research Park, Little Chesterford, Saffron Walden Essex CB10 1XL (GB). **WEST, John Stephen** [US/GB]; Solexa Limited, Chesterford Research Park, Little Chesterford, Saffron Walden Essex CB10 1XL (GB).

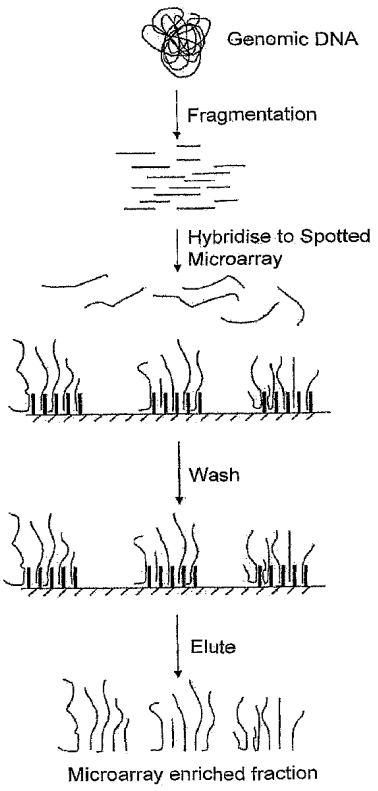
(74) Agents: **WHITE, Nina Louise** et al.; Boult Wade Tennant, Verulam Gardens, 70 Gray's Inn Road, London WC1X 8BT (GB).

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

(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),

[Continued on next page]

(54) Title: METHOD OF TARGET ENRICHMENT



(57) Abstract: The invention relates to a method of enriching for specific target sequences using libraries of oligonucleotides, such as microarrays. In particular, the invention relates to a method of enriching for specific nucleic acid target sequences in a population of nucleic acid sequences in order to provide targets for sequencing.

WO 2007/057652 A1

IDT Exhibit 1020.5960

European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report*

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

Method of Target Enrichment

Field of the invention

The present invention relates to a method of reducing the complexity of a nucleic acid sample in a reproducible manner by enriching for specific nucleic acid target sequences in the population. Specifically it relates to a method to enrich for specific target sequences using libraries of oligonucleotides such as micro-arrays, for example, for use in sequencing and particularly sequencing by synthesis.

Background to the invention

The draft sequence of the human genome was published in 2001 by the Human Genome Consortium (Nature Vol 409; issue 6822) and Celera genomics (Science, Vol 291, Issue 5507, 1304-1351), thus marking the beginning of the genetics chapter for society. Capitalizing on this investment and realizing the potential of the Human Genome Project requires a better understanding of genetic variation and its effect in disease.

It has been estimated that any two copies of the human genome differ from one another by as little as 0.1%, in other words a total of three million variants, or one variant every 1000 bases, over a total of three billion that make up the human genome. The consensus sequence of the human genome was based on information from just 12 genomes (six individuals; each person has two genomes, one from each parent), yet today there are six billion

individuals worldwide (Bennet, S., Current Drug Discovery, February 2004, p15-19).

Since such variation affects disease susceptibility and responses to drugs it is essential to identify the genetic factors which contribute to biological variation. DNA sequencing is a fundamental tool enabling the screening of genes for such genetic mutations associated with disease. High throughput, high accuracy sequencing methods are therefore required to screen the complete genome sequence of an animal in order to identify unique nucleic acid sequences which may indicate the presence of physiological or pathological conditions.

DNA sequencing of large and complex genomes is currently limited by cost. With a significant proportion of human genomic DNA comprising repetitive sequence, reducing the complexity of the sample reduces the amount of sequencing required. Furthermore, with prior genetic information, it is possible to correlate a phenotype, such as a predisposition to a disease, with the genetic variation of one or more regions of the genome, and what is desired is the application and advantages of high throughput sequencing methods specifically to these regions of interest among many individuals. Such studies are currently not feasible due to cost. In addition, in certain circumstances, it is desirable to generate a 'genome-wide' analysis of a particular genomic feature, such as exons, to correlate genetic diversity in the protein-coding regions across many individuals.

Consequently development of strategies that focus on targeted sequencing of gene rich regions provide an alternative approach to whole genome sequencing.

Weisburg et al, US6,534,273, describe a method for capturing a target polynucleotide present in a sample onto a solid support with an attached immobilised probe by using a capture probe and two different hybridisation conditions that control the order of hybridisation, where the first hybridisation condition allows hybridisation of the capture probe to the target polynucleotide, and the second hybridisation condition allows hybridisation of the capture probe to the immobilised probe. The method further includes amplifying the captured target polynucleotide by hybridising at least one primer oligonucleotide to the target polynucleotide and using nucleic acid amplification that initiates from the primer oligonucleotide. The method utilises two separate probes for use in diagnostic assays, for example in testing for the presence of bacteria in a biological sample.

Collins et al, US5,750,338, describe a method of assay for target polynucleotides which includes the steps of isolating target polynucleotides from extraneous non-target polynucleotides, debris, and impurities and amplifying the target polynucleotide. The method provides for detection of nucleic acid targets in clinical samples.

Urdea, US5,200,314, describes an analyte polynucleotide strand having an analyte sequence which is detected within a sample containing polynucleotides by contacting the analyte polynucleotide with a capture probe

under hybridization conditions, where the capture probe has a first binding partner specific for a solid-phase second binding partner.

The resulting duplex is then immobilized by specific binding between the binding partners, and non-bound polynucleotides are separated from the bound species. The analyte polynucleotide is optionally displaced from the solid phase, then amplified by PCR. The PCR primers each have a polynucleotide region capable of hybridizing to a region of the analyte polynucleotide, and at least one of the primers further has an additional binding partner capable of binding a solid-phase binding partner. The amplified product is then separated from the reaction mixture by specific binding between the binding partners, and the amplified product is detected.

Nisson et al, US6,268,133, discloses the use of amino acid denaturants for denaturing or separating double stranded nucleic acid molecules and more specifically, provides a method for the rapid isolation and recovery of a desired target DNA or RNA molecules from a mixture or library containing such molecules. The method involves the use of haptenylated probes and amino acid denaturants to select the desired molecules and eliminate the undesired library members from a sample. Their invention also provides a method in which larger or full-length nucleic acid molecules can be isolated from the subpopulation of desired molecules.

WO01/46470 in the name of Karolinska Innovations relates to a method for enrichment of specific nucleic acid

segments, such as a DNA, e.g. single nucleotide polymorphisms (SNPs), sequences that have been deleted, sequences that are identical between two complex genomes, etc. The disclosed method includes steps for providing a first sample A and a second sample B derived from different sources and digestion of both said samples; amplification of sample A with a suitable primer and dNTPs comprising one unconventional base and amplification of sample B with a labelled primer and all the conventional dNTPs, followed by combination of samples A and B; denaturation and hybridization; treatment with a nuclease specific for said unconventional base, such as uracil-DNA glycosylase (UDG), and isolation of the specific segment originally present in sample B by use of the primer label. Their invention further relates to a kit which comprises components suitable for working the above described method.

WO02/06528 in the name of Somalogic, Inc. relate to a method and apparatus for the automated generation of nucleic acid ligands. The disclosure includes a method and device for performing automated SELEX. The steps of the SELEX process are performed at one or more work stations on a work surface by a robotic manipulator controlled by a computer. The document also includes methods and reagents to obviate the need for size-fractionation of amplified candidate nucleic acids before beginning the next round of the SELEX process. SELEX or Systematic Evolution of Ligands by EXponential enrichment, is a procedure in which an initial pool of randomized polynucleotides (RNA or DNA, single stranded) is created, containing on the order of 10^15 molecules of a fixed length. The pool is then screened for some desired characteristic for example,

binding affinity for ATP. The molecules that are selected in this way are used as "parents" in the synthesis (with mutation) of a new pool of molecules, and the process repeats with more rounds of selection and amplification. The result of SELEX is a set of highly functional molecules of DNA or RNA that perform their selected function.

US6,013,440, Lipshutz et al, relates generally to matrices for conducting nucleic acid affinity chromatography. Specifically the invention relates to methods of preparing affinity chromatography matrices that bind a plurality of different pre-selected nucleic acids.

Su et al, US6,632,611, disclose methods and kits for amplifying a target sequence from within a nucleic acid population. The invention provides selection probes which are complementary to at least a portion of said target sequence and mechanisms for adding a probe sequence to the 3' end of a target sequence that is hybridized to a selection probe.

The added 3' probe sequence and a probe sequence added at the 5' end of the target by adaptor ligation allow for selective amplification of the target sequence.

Morgan et al (1992) disclose methods to direct cDNA selection allowing rapid and reproducible isolation of low abundance cDNA's encoded by large genomic clones.

Gill et al (2002) disclose a DNA microarray method for genome-wide monitoring of competitively grown transformants to identify genes whose overexpression confers a specific cellular phenotype.

These documents disclose methods of enriching a nucleic acid sample, usually by amplification, for the detection of specific target sequences. None of them provide a rapid, cost-effective method for reducing the complexity of a nucleic acid sample suitable for sequencing and particularly sequencing by synthesis. Furthermore, no methods have been described that utilise sample enrichment with high throughput sequencing methodology, such as, for example, reversible terminator chemistry described herein.

Summary of the invention

In a first aspect of the invention there is provided a method of enriching a complex nucleic acid sample for a population of target sequences for use in subsequent sequencing wherein each of said target sequences relates to a set of capture probes.

Said method comprises:

(a) fragmenting a first population of nucleic acid sequences;

(b) combining said first population of nucleic acid sequences with a set of probe sequences under conditions allowing for hybridisation of the probe sequences and said first population of nucleic acid sequences to form probe-target complexes; and

(c) purifying the probe-target complexes to discard the un-hybridised nucleic acid target sequences;

(d) sequencing the remaining probe selected population of target sequences.

In one embodiment of the invention, the method further comprises a ligation step wherein adaptors are ligated to the fragmented first population of nucleic acid sequences, either prior to or subsequent to the enrichment step.

In a further embodiment of the invention the method further comprises an amplification step whereby the fragmented first population of nucleic acid sequences are amplified using, for example, PCR. Preferably said amplification step is performed following ligation of adaptors to the fragmented first population of nucleic acid sequences. Yet more preferably said amplification step is performed on the first population of nucleic acid sequences as a whole and in contrast to the methods of the prior art is not intended to amplify only a subset of said first population.

In another embodiment of the invention the target sequences can be removed from the probe-target complex prior to sequencing, for example by elution. Removal by denaturation of the selected targets from the immobilised capture probes will generally give a solution of single stranded targets.

In another embodiment of the invention the method further comprises the step of ligating adaptors to the enriched target sequences after separation of said target sequences from the probe target complexes.

In another embodiment of the invention the target sequences remain bound to the probe(s) and are sequenced directly on the array using, for example, sequencing by synthesis (SBS).

In yet another embodiment of the invention the target sequences are removed from the array and are further amplified and/or immobilised to produce clustered arrays, or sequenced directly as single molecules.

In still yet another embodiment of the invention, enrichment of a first population of nucleic acid sequences and subsequent sequencing of target sequences takes place on a single surface i.e. a single array or 'chip'.

Brief description of the drawings

Figure 1a illustrates a simplified and schematised embodiment of the use of a microarray to enrich a fragmented complex nucleic acid sample for a population of target sequences.

Figure 1b illustrates a simplified and schematised embodiment of the use of a microarray to enrich a fragmented complex nucleic acid sample for a population of target sequences wherein adaptors are ligated to the fragmented genomic DNA prior to enrichment.

Figures 2a and 2b illustrate a simplified and schematised embodiment of the use of a microarray for 'one-chip' enrichment and sequencing. Genomic DNA is fragmented, adaptors are ligated to the ends of the fragments which are

then amplified. The capture probes hybridise to the target sequences which are then extended to produce a complimentary sequence bound to the capture probe. The capture probe can then be bound to the surface of an array, target sequence is removed and the complimentary sequence is sequenced.

Detailed description of the invention

It is an object of the present invention to provide a method for enrichment of a complex nucleic acid sample with capture probes, relating for example to genetic features of interest. More specifically the invention relates to the use of a predetermined panel of oligonucleotides, such as an array, designed to enrich said complex nucleic acid sample for use in sequencing and more particularly sequencing by synthesis.

Micro-arrays have been used primarily for gene expression analyses, although the strategy of using an ordered array of bio-molecules on such an array has also been extended to mutation detection, polymorphism analysis, mapping and evolutionary studies.

To date however there has been no disclosure on the use of such arrays to enrich a complex nucleic acid sample with specific nucleic acid sequences relating to genetic features of interest for use in sequencing by synthesis.

The use of capture probes allows the researcher to positively select for regions of the genome which are of interest whilst concomitantly negatively selecting for the

remainder of the genome. Such an approach has the advantage, for example, that highly repetitive DNA sequences which comprise 40% of genomic DNA can be removed quickly and efficiently from a complex population. As a direct result the complexity of sequence is reduced thus increasing throughput of subsequent sequencing. However, further enrichment for a target region or feature, such as exons, would further reduce the complexity of the sample. Usually, and in contrast to expression analysis studies for example, the nucleic acid sample applied to the array will not first be fluorescently labelled.

Since the method may be performed using micro-arrays there is the added benefit that the volume of input sample required is significantly reduced over methods of the prior art. The ability to use smaller quantities of input sample is a significant advantage over techniques in the prior art which often require complex strategies to increase the amount of nucleic acid by amplification. A further advantage is that the cost of carrying out both the enrichment and subsequent sequencing is also significantly reduced since less sequence data needs to be generated to produce meaningful results.

The term 'enrichment' refers to the process of increasing the relative abundance of particular nucleic acid sequences in a sample relative to the level of nucleic acid sequences as a whole initially present in said sample before treatment. Thus the enrichment step provides a percentage or fractional increase rather than directly increasing for example, the copy number of the nucleic acid sequences of interest as amplification methods, such as

PCR, would. The methods as described herein may be used to remove DNA strands that it is not desired to sequence, rather than to specifically amplify only the sequences of interest. At the level of the whole genome, removing 50% of the DNA sample gives a two fold reduction in the cost and time of sequencing the remaining regions of biological interest from the whole genome. The methods as described herein can also be used to select large regions of a genome (eg megabases) for resequencing of multiple individuals, or can select out all the exons in a genomic sample. The synthesis of one array, or pool of oligonucleotides, can be used to process multiple samples of interest, and thus the costs of the oligonucleotide synthesis can be amortised over many individual samples.

The complex nucleic acid sample or input sample is an initial sample of nucleotide sequences prior to enrichment, such as genomic DNA. As non-limiting examples, such a sample may consist of genomic DNA, cDNA, RNA, PCR products, pools or subsets thereof.

In a first embodiment there is provided a method of enriching a complex nucleic acid sample for a population of target sequences for use in subsequent sequencing wherein each of said target sequences relates to a set of capture probes.

Said method comprises:

(a) fragmenting a first population of nucleic acid sequences;

(b) combining said first population of nucleic acid sequences with a set of probe sequences under conditions allowing for hybridisation of the probe sequences and said first population of nucleic acid sequences to form probe-target complexes; and

(c) purifying the probe-target complexes to discard the un-hybridised nucleic acid target sequences;

(d) sequencing the remaining probe selected population of target sequences

Preferably said fragmented nucleic acid population comprises sequence fragments which are less than about 1000 base pairs in length, more preferably such sequences are in the range 100-1000 base pairs in length. Still more preferably such sequences are in the range of from 450-750 base pairs in length. It would be apparent to the skilled artisan that the following non-limiting fragmentation methods may be used: restriction endonucleases, other suitable enzymes, mechanical forms of fragmentation, such as nebulisation or sonication, or non-enzymatic chemical fragmentation.

In one embodiment, adaptors may be ligated to the fragmented first population of nucleic acid sequences, either prior to or subsequent to the enrichment step.

In a further embodiment the fragmented first population of nucleic acid sequences may be subjected to an amplification step using, for example, PCR. Preferably said amplification step is performed following ligation of adaptors to the fragmented first population of nucleic acid

sequences. Yet more preferably said amplification step is performed on the first population of nucleic acid sequences as a whole and in contrast to the methods of the prior art is not intended to amplify only a subset of said first population.

The capture probes are preferably nucleic acids, such as oligonucleotides, capable of binding to a target nucleic acid sequence through one or more types of chemical bonds, usually through complementary base pairing, usually through hydrogen bond formation. Such probes may include natural or modified bases and may be RNA or DNA. In addition the bases in probes may be joined by a linkage other than a phosphodiester bond so long as it does not interfere with hybridisation. Thus probes may also be peptide nucleic acids (PNA) in which the constituent bases are joined by peptide bonds rather than phosphodiester linkages.

Capture probes are reference populations of nucleic acid sequences. These have been selected such that said probes relate to, by way of non-limiting examples, a set of genes of interest, all of the exons of a genome, particular genetic regions of interest, disease or physiological states and the like. For example it can be envisaged that such reference populations will include commercially available populations available as micro-arrays or 'chips' more commonly used in expression profiling such as the Affymetrix® Exon Gene-Chip®. The capture probes can also be synthesised as oligonucleotides in solution, and can be used either in solution or immobilised on beads. The beads could contain multiple copies of individual sequences, such that each beads contains a single, different sequence, or

can just contain the whole pool of oligonucleotides immobilised on each bead such that each bead is the same mixture of sequences.

Capture probes may also be prepared from a sample of DNA from any source, for example bacterial artificial chromosomes (BACs), PCR fragments, whole chromosomes or cDNA libraries. Use of a suitably available nucleic acid sample that can be fragmented and enriched as described means that the same region can be re-sequenced from multiple individuals without the need for chemical synthesis of specific capture probes across that region.

Any available nucleic acid can be fragmented and undergo a ligation with an adaptor sequence to establish common known ends on each fragment. Such fragment libraries can be amplified using primers complementary to the known ends and modified with groups amenable to surface attachment, such as, for example, biotin. The fragment pools, once made single stranded, are attached to a suitably functionalised surface, such as, for example, streptavidin beads. If the bead pool is exposed to a single stranded target DNA sample, then the fragments of the target DNA sample complementary to the single stranded fragments immobilised on the beads will bind, and the non-complementary sequences will remain unbound in solution and can be easily separated from the immobilised fragments.

Thus removal of sequences that were not complementary to those fragments of the capture pool enriches the remaining, immobilised target DNA.

The hybridisation step may be performed either on the solid surface, such as on beads, to which the single stranded capture probes have been bound, or in solution. If the hybridisation is performed in solution, subsequent addition of beads results in binding of all the capture probes, either as duplexes with the target sample, or as single strands. The remainder of the target DNA which has not formed duplexes with one of the capture probes will not be able to bind to the beads. Unbound target sample can be removed from the beads by washing, for example, and the duplex sample can be treated to elute the hybridised target into solution.

The enriched sample may be eluted from the beads and can be attached to a surface and used for sequencing, either as arrays of single molecules, or amplified to form clustered arrays of clonal single molecules, for example as described in WO9844151. In an alternative embodiment the enriched sample may be amplified whilst still attached to the beads by, for example, emulsion phase PCR, or may be eluted from the beads and amplified in solution prior to surface attachment.

The terms 'target' or 'target sequence' refer to nucleic acid sequences of interest that is, those which hybridise to the capture probes. Thus the term includes those larger nucleic acid sequences, a sub-sequence of which binds to the probe and/or to the overall bound sequence. Since the target sequences are for use in sequencing methods, said target sequences do not need to have been previously defined to any extent, other than the bases complementary to the capture probes.

Capture probes hybridise to target sequences in the complex nucleic acid sample. It will be apparent to one skilled in the art that prior to hybridisation said complex nucleic acid sample will preferably comprise single stranded nucleic acid sequences. This can be achieved by a number of well known methods in the art such as, for example using heat to denature or separate complementary strands of double stranded nucleic acids, which on cooling can hybridise to the capture probes. It is also conceivable that said complex nucleic acid sample could comprise double stranded polynucleotides with a single stranded overhang ('sticky ends') which may hybridise to said capture probes.

To provide enrichment, the capture probes are preferably immobilised onto a support, either before or after hybridisation, such that sequences that do not hybridise to said capture probes can be removed for example, by washing.

In one embodiment the target sequences can be removed from the probe-target complex prior to sequencing for example by elution. Removal by denaturation of the selected targets from the immobilised capture probes will generally give a solution of single stranded targets.

In a further embodiment adaptors may be ligated to the enriched target sequences after removal of said target sequences from the probe target complexes. The target sequences may also be further fragmented after elution from the support used for enrichment. For example, it may be

advantageous to initially fragment the target sample to an average size of 10 kB, and thereby require fewer probe sequences to select out a specific megabase region. A 10 kB region can be selected, but not easily amplified, and therefore further fragmentation, to an average of a few hundred bases may be used after the enrichment step. If a second fragmentation step is used, then the universal adaptors will need to be ligated onto the enriched target sequences after the removal from the support and after the further fragmentation step.

The solid support may be any of the conventional supports used in arrays or 'DNA chips', beads, including magnetic beads or polystyrene latex microspheres, arrays of beads, or substrates such as membranes, slides and wafers made from cellulose, nitrocellulose, glass, plastics, silicon and the like.

Preferably the solid support is a flat planar surface or an array of beads. Still more preferably said solid support is an array and most preferably said array is a 'high density array' such as a micro-array.

Arrays are collections of biomolecular probes such as nucleic acids which are immobilised onto a solid support; as non-limiting examples, the biomolecular probes could be oligonucleotides of varying length (preferably 25 to 60mers), PCR products representing a cDNA clone library or BAC clones such as those used in comparative genome hybridisation. Multi-polynucleotide arrays or clustered arrays are 'high density arrays' of nucleic acid molecules which may be produced using techniques generally known in

the art. By way of example, WO98/44151 and WO00/18957 both describe methods of nucleic acid amplification which allow amplification products to be immobilised on a solid support in order to form arrays comprised of clusters or 'colonies' of immobilised nucleic acid molecules. An array of amplified molecules from a previously enriched, or otherwise obtained target may be used to select the same target regions from a new sample. The enriched DNA can be sequenced directly on the array, or removed from the array for subsequent sequencing by any desired sequencing process.

Preferably said array contains greater than 100 probes. More preferably said array contains greater than 1000 probes, still more preferably said array contains greater than 10,000 probes. Still yet more preferably said array contains greater than 100,000 probes.

Immobilisation of the probes may be by specific covalent or non-covalent interactions. If the molecule is a polynucleotide, immobilisation will preferably be at either the 5′ or 3′ position so that the polynucleotide is attached to the solid support at one end only. However, the polynucleotide may be attached to the solid support at any position along its length, the attachment acting to tether the polynucleotide to the solid support. The immobilised polynucleotide is then able to undergo interactions with other molecules or cognates at positions distant from the solid support. Typically the interaction will be such that it is possible to remove any molecules bound to the solid support through non-specific interactions, e.g. by washing.

In one embodiment the target sequences remain bound to the probe and can be sequenced directly on the array using, for example, sequencing by synthesis (SBS). The can either be on a chemically synthesised array, or as a lawn of primers deposited such that a single molecule array of selected templates can be formed. Single molecule arrays and their use in sequencing is described in WO0006770.

In another embodiment the target sequences are removed from the array and may be optionally amplified in solution prior to immobilisation. The target sequences, and their complementary copies can be immobilised on a solid support. The immobilised arrays can be further amplified to produce clustered arrays, or sequenced directly as single molecules.

Any suitable method for of sequencing may be used to determine a sequence read of the immobilised enriched targets. Suitable methods of sequencing include the use of sequencing by addition of nucleotide bases, for example sequencing by synthesis (SBS) using nucleoside triphosphates (as described in WO04018497) and DNA polymerases, or using oligonucleotide cassettes and ligases; as described in US6306597 or Science, 309:5741,1728-1732 (2005). The enriched targets may also be sequenced by pyrosequencing (Nature. 437:376-380 (2005)), or by MPSS where the strands are degraded rather than extended (Nat Biotechnol. 6:630-6344 (2000)).

In "sequencing by synthesis" or SBS a new polynucleotide strand based-paired to a template strand is

built up in the 5' to 3' direction by successive incorporation of individual nucleotides complementary to the template strand. In one embodiment of SBS the substrate nucleoside triphosphates used in the sequencing reaction are each labelled on the base with different labels permitting determination of the identity of the incorporated nucleotide as successive nucleotides are added. The labelled nucleoside triphosphates also have a 3′ blocking group which prevents further incorporation of complementary bases by the polymerase. The label of the incorporated base can then be determined and the blocking group removed to allow further polymerisation to occur.

There are known in the art methods of nucleic acid sequencing based on successive cycles of incorporation of fluorescently labelled nucleic acid analogues. In such "sequencing by synthesis" or "cycle sequencing" methods the identity of the added base is determined after each nucleotide addition by detecting the fluorescent label. In particular, US 5,302,509 describes a method for sequencing a polynucleotide template which involves performing multiple extension reactions using a DNA polymerase to successively incorporate labelled polynucleotides complementary to a template strand.

The present inventors have developed methods of sequencing multiple nucleic acid molecules in parallel based on the use of arrays, wherein multiple template molecules immobilised on the array are sequenced in parallel. Such arrays may be single molecule arrays or clustered arrays. The nucleotide (s) incorporated into the strand of nucleic acid complementary to the template

nucleic are each fluorescently labelled. The inclusion of a fluorescent label facilitates detection/identification of the base present in the incorporated nucleotide(s). Appropriate fluorophores are well known in the art.

The labels may be the same for each type of nucleotide, or each nucleotide type may carry a different label. This facilitates the identification of incorporation of a particular nucleotide. Thus, for example modified adenine, guanine, cytosine and thymine would all have attached a different fluorophore to allow them to be discriminated from one another readily. When sequencing on arrays, a mixture of labelled and unlabelled nucleotides may be used. Detectable labels such as fluorophores can be linked to nucleotides via the base using a suitable linker. The linker may be acid labile, photolabile or contain a disulfide linkage. Preferred labels and linkages include those disclosed in W003/048387. Other linkages, in particular phosphine-cleavable azide-containing linkers, may be employed in the invention as described in greater detail in W02004/018493. The contents of WO 03/048387 and WO 2004/018493 are incorporated herein in their entirety by reference.

The nucleotides described in W02004/018493 comprise a purine or pyrimidine base and a ribose or deoxyribose sugar moiety which has a removable blocking group covalently attached thereto, preferably at the 3'O position. 3' blocking groups are also described in W02004/018497, the contents of which are also incorporated herein in its entirety by reference. Use of such 3'-blocked nucleotides permits controlled incorporation of nucleotides in a

stepwise manner, since the presence of a blocking group at the 3'-OH position prevents incorporation of additional nucleotides. The detectable label may, if desirable, be incorporated into the blocking groups as is disclosed in WO2004/018497.

In further embodiments of SBS or cycle sequencing wherein the substrate nucleoside triphosphates used in the sequencing reaction are each labelled on the base with the same label and/or wherein the labelled nucleoside triphosphates do not have a 3′ blocking group to prevent further incorporation of complementary bases by the polymerase it will be apparent to the skilled person that in these cases the nucleotides can be supplied individually and serially and incorporation of a base can then be determined before applying the next nucleotide.

Methods for detecting fluorescently labeled nucleotides generally require use of incident light (e .g. laser light) of a wavelength specific for the fluorescent label, or the use of other suitable sources of illumination, to excite the fluorophore. Fluorescent light emitted from the fluorophore may then be detected at the appropriate wavelength using a suitable detection system such as for example a Charge-Coupled-Device (CCD) camera, which can optionally be coupled to a magnifying device, a fluorescent imager or a confocal microscope. If sequencing is carried out on an array, detection of an incorporated base may be carried out by using a confocal scanning microscope to scan the surface of the array with a laser, to image fluorescent labels attached to the incorporated nucleotide(s). Alternatively, a sensitive 2-D detector,

such as a charge-coupled detector (CCD), can be used to visualise the signals generated. This technique is particularly useful with single molecule arrays.

Other techniques such as scanning near-field optical microscopy (SNOM) are available and may be used when imaging dense arrays. For a description of scanning near-field optical microscopy, see Moyer et al., Laser Focus World **29**:10, 1993. An additional technique that may be used is surface-specific total internal reflection fluorescence microscopy (TIRFM) ; see, for example, Vale et al., Nature, (1996) **380**:451-453.

Suitable apparatus used for imaging polynucleotide arrays are known in the art and the technical set-up will be apparent to the skilled person. Detection buffers containing antioxidants, such as sodium ascorbate, show a clear improvement (over corresponding buffers absent such antioxidants) at preventing light-induced chemical artefacts in cycles of sequencing-by-synthesis based on detection of fluorescently labeled nucleotide analogues, as described in WO06064199. The inclusion of antioxidants prevents/reduces light-induced chemical reactions from damaging the integrity of the nucleic acid template and allows accurate determination of the identity of the incorporated base over at least 2, preferably at least 10 and more preferably at least 16 cycles of nucleotide incorporation.

Preferably from 10 to 50 and more preferably from 16 to 30 nucleotides are successively incorporated, and identified, in the sequencing reaction. The ability to

accurately sequence 10 or more, and preferably 16 or more, consecutive nucleotides in a sequencing reaction is a significant advantage in applications such as genome re-alignment. In the context of this invention the terms "sequencing reaction", "sequencing methodology" or "method of sequencing" generally refer to any polynucleotide "sequencing-by-synthesis" reaction which involves sequential addition of one or more nucleotides or oligonucleotides to a growing polynucleotide chain in the 5' to 3' direction using a polymerase or ligase in order to form an extended polynucleotide chain complementary to the template nucleic acid to be sequenced.

The identity of the base present in one or more of the added (oligo)nucleotide(s) is determined in a detection or "imaging" step. The identity of the added base is preferably determined after each nucleotide incorporation step. The sequence of the template may then be inferred using conventional Watson-Crick base-pairing rules.

The nucleic acid template to be sequenced in a sequencing reaction may be any polynucleotide that it is desired to sequence. The nucleic acid template for a sequencing reaction will typically comprise a double stranded region having a free 3' hydroxyl group which serves as a primer or initiation point for the addition of further nucleotides in the sequencing reaction. The region of the template to be sequenced will overhang this free 3' hydroxyl group on the complementary strand. The primer bearing the free 3' hydroxyl group may be added as a separate component (e .g. a short oligonucleotide) which hybridises to a region of the template to be sequenced.

Alternatively, the primer and the template strand to be sequenced may each form part of a partially self-complementary nucleic acid strand capable of forming an intramolecular duplex, such as for example a hairpin loop structure. Nucleotides are added successively to the free 3' hydroxyl group, resulting in synthesis of a polynucleotide chain in the 5' to 3' direction. After each nucleotide addition the nature of the base which has been added may be determined, thus providing sequence information for the nucleic acid template.

The term "incorporation" of a nucleotide into a nucleic acid strand (or polynucleotide) refers to joining of the nucleotide to the free 3' hydroxyl group of the nucleic acid strand via formation of a phosphodiester linkage with the 5' phosphate group of the nucleotide. The nucleic acid template to be sequenced may be DNA or RNA, or even a hybrid molecule comprised of deoxynucleotides and ribonucleotides. The nucleic acid may comprise naturally occurring and/or non-naturally occurring nucleotides and natural or non-natural backbone linkages.

Nucleic acid templates to be sequenced may be attached to a solid support via any suitable linkage method known in the art. Preferably linkage will be via covalent attachment. If the templates are "arrayed" on a solid support then the array may take any convenient form. Thus, the method of the invention is applicable to all types of "high density" arrays, including single-molecule arrays and clustered arrays.

The enrichment method of the invention may be carried out using essentially any type of array formed by immobilisation of nucleic acid molecules on a solid support, and more particularly any type of high-density array, including bead arrays. The sequencing aspect of the invention may be carried out using essentially any type of array formed by immobilisation of nucleic acid molecules on a solid support, and more particularly any type of high-density array, including single molecule, amplified single molecule (cluster) arrays, arrays of beads on which molecules have been amplified (for example in an emulsion PCR reaction), or arrays of beads on which amplified molecules have been hybridised.

In multi-polynucleotide or clustered arrays distinct regions on the array comprise multiple copies of single polynucleotide template molecules. Multi-polynucleotide or clustered arrays of nucleic acid molecules may be produced using techniques generally known in the art. By way of example, WO98/44151 and WO00/18957 both describe methods of nucleic acid amplification which allow amplification products to be immobilised on a solid support in order to form arrays comprised of clusters or "colonies" of immobilised nucleic acid molecules. The arrays are amplified such that both strands of a duplex are immobilised, but cleavage of one of the strands from the surface (for example using a chemical and/or a subsequent heat treatment to cleave and denature one of the amplification primers used to generate the copies of the immobilised single molecules), results in an array of single stranded templates suitable for sequencing.

The nucleic acid molecules present on the clustered arrays prepared according to these methods are suitable templates for sequencing using the method of the invention. Both WO98/44152 and WO00/18957, the contents of which are incorporated herein by reference describe methods of parallel sequencing of multiple templates located at distinct locations on a solid support, and in particular sequencing of "clustered" arrays. The single stranded arrays described above can be hybridised with a suitable sequencing primer, complementary to a region common to each of the amplified templates, to provide a free 3'-hydroxyl group suitable for sequencing against the unknown, variable region of the amplified templates.

Nevertheless, the method of the invention may also be used in the context of sequencing templates on single molecule arrays of nucleic acid templates. Single molecule arrays are generally formed by immobilisation of a single polynucleotide molecule at each discrete site that is detectable on the array. Single-molecule arrays comprised of nucleic acid molecules that are individually resolvable by optical means and the use of such arrays in sequencing are described, for example, in WO00/06770.

Single molecule arrays comprised of individually resolvable nucleic acid molecules including a hairpin loop structure are described in WO 01/57248. The method of the invention is suitable for sequencing template molecules on single molecule arrays prepared according to the disclosures of WO 00/06770 or WO 01/57248. The fluorescent moiety may be attached to a nucleic acid via any suitable covalent or non-covalent linkage. For example, the

fluorescent moiety may be attached to an oligonucleotide primer or probe which is hybridised to a target nucleic acid molecule.

In a preferred embodiment, enrichment of a first population of nucleic acid sequences and subsequent sequencing of target sequences takes place on a single surface i.e. a single array or 'chip'. More preferably the sequencing is by sequencing by synthesis.

Preferably a large number of target sequences are sequenced in parallel at the same time. More preferably greater than 100 target sequences are sequenced at a time. More preferably greater than 1000 target sequences are sequenced at one time, still more preferably greater than 10,000 target sequences are sequenced at one time. Still yet more preferably greater than 100,000 target sequences are sequenced at one time.

Experimental overview

The following experimental details describe the complete exposition of one embodiment of the invention as described above.

Example 1

Four probes can be designed that hybridise to the 5′ end of one exon from each of the four following genes present in the human BAC BCX98J21: PPP1R10, ABCF1,PRR3, GNL1. Each probe is 60 bases in length, contains a 5′ biotin group and hybridises uniquely at its intended sequence in the BAC. The sequences of the probes are as follows:

Probe #1 (PPP1R10)
TCGGTTAAGGAAGCTGTCCAGGCCCTTGAGAAGTTCTTTGGGGTCTATGGGACCCGAAC
C

Probe #2 (ABCF1)
GCCGTATCTGAGGAACAGCAGCCTGCACTCAAGGGCAAAAAGGGAAAGGAAGAGAAGTC
A

Probe #3 (PRR3)
CCGAAACGAAAGAAGCAGAATCATCACCAGCCACCGACACAGCAGCAGCCCCCGCTGCC
C

Probe #2 (GNL1)
CTCCCGTTTGTCCTGCAACTGCTTCTTCTTCTGCTTCACGCTGAATGGCTTCTTCCTCG
G

A solution is prepared containing a mixture of all four probes at a concentration of 1micromolar each in 5xSSC buffer and is added to a tube containing 1 microgram of BAC DNA that has been previously fragmented to less than 1000 base pairs using a nebulizer (Invitrogen® #K7025-05) in a total volume of 50 ul. The solution is heated to 97.5°C for 5 minutes, then cooled to room temperature to anneal the probes to their target sequences in the BAC fragments.

Wash 40 ul of streptavidin coated magnetic beads (Dynal® Biotech) twice with 100 ul of 2xB&W buffer (Dynal®), finally resuspending the beads in 50 ul of 2xB&W. Add to the solution of annealed DNA and incubate for 15 minutes at room temperature with gentle mixing on a roller mixer platform. Apply the tube of the DNA-bead mix to the

magnetic holder (Dynal®) and incubate at room temperature for 2 minutes. Wash the beads three times with 1xB&W buffer (Dynal®), each time discarding the supernatant, resuspending the beads in fresh 1XB&W buffer and reapplying to the magnets. Finally, add 50 ul of 100 mM NaOH to the DNA-bead mix and incubate at room temperature for 5 minutes. Reapply to the magnetic holder, then recover the supernatant after a 2 minute incubation, discarding the beads. The DNA recovered in the 100 mM NaOH solution may be neutralised by the addition of a titrated quantity of HCl. Alternatively, the DNA-NaOH solution can be exchanged for 5xSSC buffer by using a MicroSpin® S300 HR column (AmershamPharmacia® #27-5130-01).

Example 2

A set (500,000) of probes can be designed that hybridise to unique positions among the 10 regions of the human genome selected by the HapMap ENCODE resequencing and genotyping project. Each probe approximately 60 bases in length and contains a 5' phosphorothioate group.

A solution is prepared of a mixture of all probes at a total concentration of 10 micromolar in 100 mM potassium phosphate buffer pH7. The probe set is grafted onto the surface of an array chip by flowing the solution of probes over the functionalised array surface at 15 ul/min at 51°C. The chip is then washed by pumping consecutively across the surface of the array: 100 mM potassium phosphate buffer pH7, TE buffer (10 mM Tris pH8, 10 mM EDTA) and 5xSSC.

1ug of total human DNA can be fragmented to less than 1000 base pairs using a nebulizer (Invitrogen #K7025-05). The DNA is diluted to 10 nM in 5xSSC and then pumped onto the surface of the array. The array is heated to 97.5°C for 5 minutes, then cooled to room temperature to anneal the fragmented total human DNA to the probe sequences on the surface of the array. Non-hybridised DNA is then removed by washing the surface of the array consecutively with the following solutions: 5xSSC, 0.3x SSC, and 5xSSC. The DNA that has hybridised to the surface probes can be recovered by pumping TE (10 mM Tris pH8, 1 mM EDTA) onto the array and heating the array to 97.5°C for 5 minutes. Immediately thereafter, the contents of the array are pumped into a collecting tube at 97.5°C and cooled to 4°C.

Example 3

1ug of total human DNA can be fragmented to less than 1000 base pairs using a nebulizer (Invitrogen #K7025-05). The DNA is diluted to 1 nM in 5xSSC and then pumped onto the surface of an Affymetrixl® Genechip® Exon Array spotted microarray. The array is heated to 97.5°C for 5 minutes, then cooled to 45°C. The array is further incubated at 45°C for 16 hours to anneal the fragmented total human DNA to the primer oligonucleotides on the surface of the array. Non-hybridised DNA is then removed by washing the surface of the array with 3 cycles of the following consecutive wash solutions: 6x SSPE/0.01% Tween-20, 100 mM MES/0.01% Tween-20. The DNA that has hybridised to the surface probes can be recovered by pumping TE (10 mM Tris pH8, 1 mM EDTA) onto the array and heating the array to 97.5°C for 5 minutes. Immediately thereafter, the contents of the array

are pumped into a collecting tube at 97.5°C and cooled to 4°C.

Example 4

1ug of total human DNA can be fragmented to less than 1000 base pairs using a nebulizer (Invitrogen #K7025-05). The DNA is diluted to 1 nM in 5xSSC and then pumped onto the surface of an Affymetrix1® Genechip® Exon Array spotted microarray. The array is heated to 97.5°C for 5 minutes, then cooled to 45°C. The array is further incubated at 45°C for 16 hours to anneal the fragmented total human DNA to the primer oligonucleotides on the surface of the array. Non-hybridised DNA is then removed by washing the surface of the array with 3 cycles of the following consecutive wash solutions: 6x SSPE/0.01% Tween-20, 100 mM MES/0.01% Tween-20. The chip is then subject to multiple cycles of SBS sequencing.

Example 5

The following experimental details describe the complete exposition of one embodiment of the invention. The DNA source used is purified Human cell line DNA supplied by the Coriell Cell Repositories, Camden, NJ 08103 USA, catalog no. NA07055. The DNA is first prepared for the ligation reaction to a single adaptor by: fragmentation of the DNA by nebulisation, then polishing of the DNA ends to make them blunt-ended and phosphorylated. The ligation reaction is performed with the prepared fragmented DNA and an adaptor preformed by annealing 'Oligo A' and 'Oligo B' (sequences given below). Next, the product of the ligation reaction is subject to cycles of PCR with a single primer

'Oligo C' (sequence given below) to selectively amplify ligated product that contains adaptor at both ends of the fragments. The product of the PCR reaction is purified from unligated adaptor and primer 'Oligo C' by gel electrophoresis. These products are next denatured, then renatured in the presence of a set of probes. The set (500,000) of probes can be designed such that they hybridise to unique positions among the 10 regions of the human genome selected by the HapMap ENCODE resequencing and genotyping project. Each probe is approximately 80 bases in length and contains a common (universal) sequence 'Sequence D' at the 5' end as well as a terminal 5' phosphorothioate group. After the probes have hybridised, a polymerisation reaction is performed with klenow polymerase and dNTPs to extend the hybridised probes to the 5' end of the DNA fragments forming a duplex. The duplexes are next coupled to the surface of a DNA chip in conjunction with two amplification primers whose sequences are identical to the 5' end of 'Oligo A' and 'Sequence D'. The DNA coupled to the chip is denatured and washed to remove hybridised DNA. The chip can then be subjected to cluster amplification and sequencing by SBS.

Nebulization

Materials:

- Human DNA (1mg/ml) Corriell NMA07055
- Buffer (glycerol 53.1 ml, water 42.1 ml, 1 M TrisHCl pH7.5 3.7 ml, 0.5 M EDTA 1.1 ml)
- Nebulizer Invitrogen® (#K7025-05)
- Qiagen® columns PCR purification kit (#28104)

Mix: 25µl (5 micrograms) of DNA

725µl Buffer

Procedure:

Chill the DNA solution and fragment in a nebulizer on ice for 5 to 6 minutes under at least 32psi of pressure. Recover the solution by centrifugation (volume usually somewhere between 400 and 600µl), split into 3 aliquots and purify with a Qiagen® PCR-purification kit, but using only one column, and finally elute in 30µl of EB (Qiagen®).

End-repair

Materials:

- T4 DNA Polymerase NEB #M0203S
- 10xNEB 2 buffer NEB #M7002S
- 100x BSA NEB #M9001S
- dNTPs mix (10mM each) NEB #N0447S
- E. coli DNA Pol I large fragment (Klenow, NEB #M0210S)
- T4 polynucleotide kinase NEB #M0201S
- T4 PNK buffer NEB #M0201S
- 100 mM ATP
- Qiagen® columns PCR purification kit (#28104)

End repair mix assembled as follows:

DNA	30µl
Water	12µl
10xNEB2	5µl
100xBSA	0.5µl
10mM dNTPs	2µl
T4 DNA pol (3U/µl)	5µl
	50µl total

Incubate the reaction for 15 min at room temperature, then add 1µl of E. coli DNA Pol I large fragment (Klenow) and incubate the reaction for a further 15 min at room temperature. Purify the DNA from enzymes, buffer, etc by loading the reaction mix on a Qiagen® column, finally eluting in 30µl EB. The 5′ ends of the DNA are then phosphorylated using polynucleotide kinase as follows:

DNA	30µl
Water	9.5µl
10xPNK buffer	5µl
100mM ATP	0.5µl
T4 PNK (10U/µl)	5µl
	50µl total

Incubate the reaction for 30 min at 37°C, then heat inactivate at 65°C for 20 min. DNA is then purified from enzymes, buffer, etc by loading the reaction mix on a Qiagen® column, finally eluting in 30µl EB. Three separate tubes are pooled to give 90µl total.

Anneal adapter

Materials:

- Oligo:5′AAATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATC
- Oligo B:5′GATCGGAAGAGCGTCGTGTAG
- 50mM Tris/50mM NaCl pH7
- PCR machine

100µM Oligo A 20µl

100µm Oligo B	20µl	
Tris/NaCl	10µl	
	50µl at 40µM duplex in 10mM Tris/10mM NaCl pH7.5	

The adapter strands are annealed in a PCR machine programmed as follows:
Ramp at 0.5°C/sec to 97.5°C
Hold at 97.5°C for 150 sec
Then a step of 97.5°C for 2 sec with a temperature drop of 0.1°C/cycle for 775 cycles

Ligation reaction

Materials

- 15uM adaptor
- End-repaired fragmented DNA genomic DNA
- Quick Ligase NEB #M2200L
- Quick Ligase 2x buffer NEB #M2200L
- PCR machine
- Qiagen® columns x1 PCR purification kit (#28104)

DNA	10µl	
2x buffer	25µl	
15 uM adaptor	10µl	
Quick Ligase	5µl	
	~50µl total	

Incubate for 40 min @ RT
Clean up with a Qiagen® column. Elute in 30µl EB.

Pass down a S300 column to get rid of excess adaptor.

PCR amplification

Materials:

- Ligated DNA
- Oligo C: AATGATACGGCGACCACCGA
- 2x RedTaq™ PCR mix
- PCR machine
- Qiagen® MinElute columns Qiagen®(#28004)

The purified ligated DNA is diluted 25 fold, then a PCR reaction mix prepared as follows:

DNA	1µl
2x Red Taq™ mix	25µl
100 µM Oligo C	0.5µl
Water	23.5µl
	~50µl total

Thermocycling is carried out in a PCR machine under the following conditions:

- 2 min @ 70°C
- 2 min @ 94°C
- [45 sec@ 94°C, 45 sec @ 65°C, 2 min @ 70°C] 16 cycles
- 5 min @ 70°C
- Hold @ 4°C

PCR products are purified from enzymes, buffer, etc on a Qiagen® MinElute® column, eluting in 10µl EB.

Gel purification

Materials:

- Agarose Biorad® #161-3101
- 100 base pair ladder NEB #N3231L
- TAE
- Loading buffer (50 mM Tris pH8, 40 mM EDTA, 40% w/v sucrose)
- Ethidium bromide
- Gel trays and tank. Electrophoresis unit

The entire sample from the purified PCR amplification reaction is loaded into one lane of a 2% agarose gel containing ethidium bromide and run at 120V for 50 min. The gel is then viewed on a 'White-light®' box and fragments from above 300bp to at least 750bp excised and purified with a Qiagen® Gel purification kit, eluting in 30µl EB. For large gel slices two MinElute® columns are used, eluting each in 15µl EB and pool.

Annealing of the probe sets

The probes all consist of the following format:
5'PS- Sequence D-Probe sequence
where 'PS' represents a 5' phosphorothioate and 'Sequence D' is as follows:
CAAGCAGAAGACGGCATACGA

The gel-purified DNA is denatured by adding NaOH at a final concentration of 100 mM and incubating at room temperature for 5 minutes. This solution is neutralised by adding 100 microlitres of pre-warmed hybridisation solution (6xSSPE, 0.1% Tween 20) containing the probe set at a final concentration of 10 micromolar, and incubated at 37°C for 1 hour. The DNA is then purified on a Qiagen® MinElute® column, with a final elution in 10 microlitres of EB.

Probe extension

Materials:

1mM dNTPs mix

10x buffer (100mM Tris-HCl, pH 7.9, 100mM MgCl2, 10 mM DTT, 500mM NaCl)

Klenow fragment (3'->5' exo-) NEB #M0212S

Mix	
Probe-hybridised DNA	10 ul
10x buffer	5 ul
1mM dNTP mix	5 ul
H2O	29 ul
Klenow exo- (2Units/ul)	1 ul
	~50µl total

Incubate at 37°C for 1 hr, then purify on a Qiagen® DNA purification column, with a final elution in 30ul of EB buffer.

Covalent attachment to array chip

The coupling of DNA to the surface of a chip and sequencing of the template molecules on arrays may be carried out according to the disclosures of WO 00/06770, WO 01/57248 or WO06/064199 or PCT/GB2006/002687 the contents of which are herein incorporated by reference.

Claims:

1. A method of obtaining a population of target sequences for the purpose of sequencing, wherein each of said target sequences relates to a pre-determined nucleic acid sequence of interest comprising:

(a) fragmenting a first population of nucleic acid sequences;

(b) combining said first population of nucleic acid sequences with a set of probe sequences under conditions allowing for hybridisation of the probe sequences and said first population of nucleic acid sequences to form probe-target complexes; and

(c) purifying the probe-target complexes to discard the un-hybridised nucleic acid target sequences;

(d) sequencing the remaining probe selected population of target sequences.

2. A method of obtaining a population of target sequences for the purpose of sequencing, wherein each of said target sequences relates to a pre-determined nucleic acid sequence of interest comprising:

(a) fragmenting a first population of nucleic acid sequences;

(b) combining said first population of nucleic acid sequences with a set of probe sequences under conditions allowing for hybridisation of the probe sequences and said first population of

nucleic acid sequences to form probe-target complexes; and

(c) purifying the probe-target complexes to discard the un-hybridised nucleic acid target sequences;

(d) removing the bound sequences of the said first population of nucleic acid sequences from the probe- target complexes to form said population of target sequences

(e) immobilising said population of target sequences

(f) sequencing said immobilised population of target sequences.

3. A method of obtaining a selected population of target sequences for the purpose of sequencing, wherein each of said target sequences relates to a pre-determined nucleic acid sequence of interest comprising,

(a) fragmenting a first population of nucleic acid sequences;

(b) combining said fragmented nucleic acid population with a set of probe sequences under conditions allowing for hybridisation of the probe sequences and target sequences to form probe-target complexes;

(c) purifying the probe-target complexes from un-hybridised nucleic acid sequences by washing to leave probe-target complexes;

(d) Removing the selected targets from the purified probe-target complexes;

(e) amplifying the target sequences to produce multiple copies of the selected population of target sequences;

(f) Sequencing the multiple copies of the selected population of target sequences.

4. The method of claims 1 to 3 further comprising the step of ligating adaptors after fragmentation of nucleic acid sequences.

5. The method of claim 2 or 3 further comprising the step of ligating adaptors to the target sequences after removing said sequences from the probe-target complexes.

6. The method of claim 4 or 5 further comprising amplifying the nucleic acid sequences after ligating adaptor sequences.

7. The method of claims 1 to 6 wherein said pre-determined probe sequences are immobilised on a support prior to hybridisation with the fragmented nucleic acid population.

8. The method of claims 1 to 6 wherein said probe-target complexes are immobilised on a support subsequent to hybridisation of the pre-determined probe sequences with the target sequences.

9. The method of claims 7 or 8 wherein said support is a solid support.

10. The method of claim 7 wherein said support is an array.

11. The method of claim 10 wherein said array is a high density array.

12. The method of claim 9 wherein said solid support is magnetic particles.

13. The method of claim 9 wherein said solid support is beads.

14. The method of any preceding claim wherein said sequencing step comprises incorporation of one or more labelled nucleotide bases each having a reversible terminator attached thereto, and a suitable polymerase, and imaging to determine the identity of each incorporated base

15. Use of an array for the enrichment of a nucleic acid population wherein said use increases the relative abundance of specific target sequences in said nucleic acid population.

16. The use of claim 15 wherein said array is a high density array.

17. The use of claim 15 or 16 wherein the enriched nucleic acid population is for sequencing by synthesis.

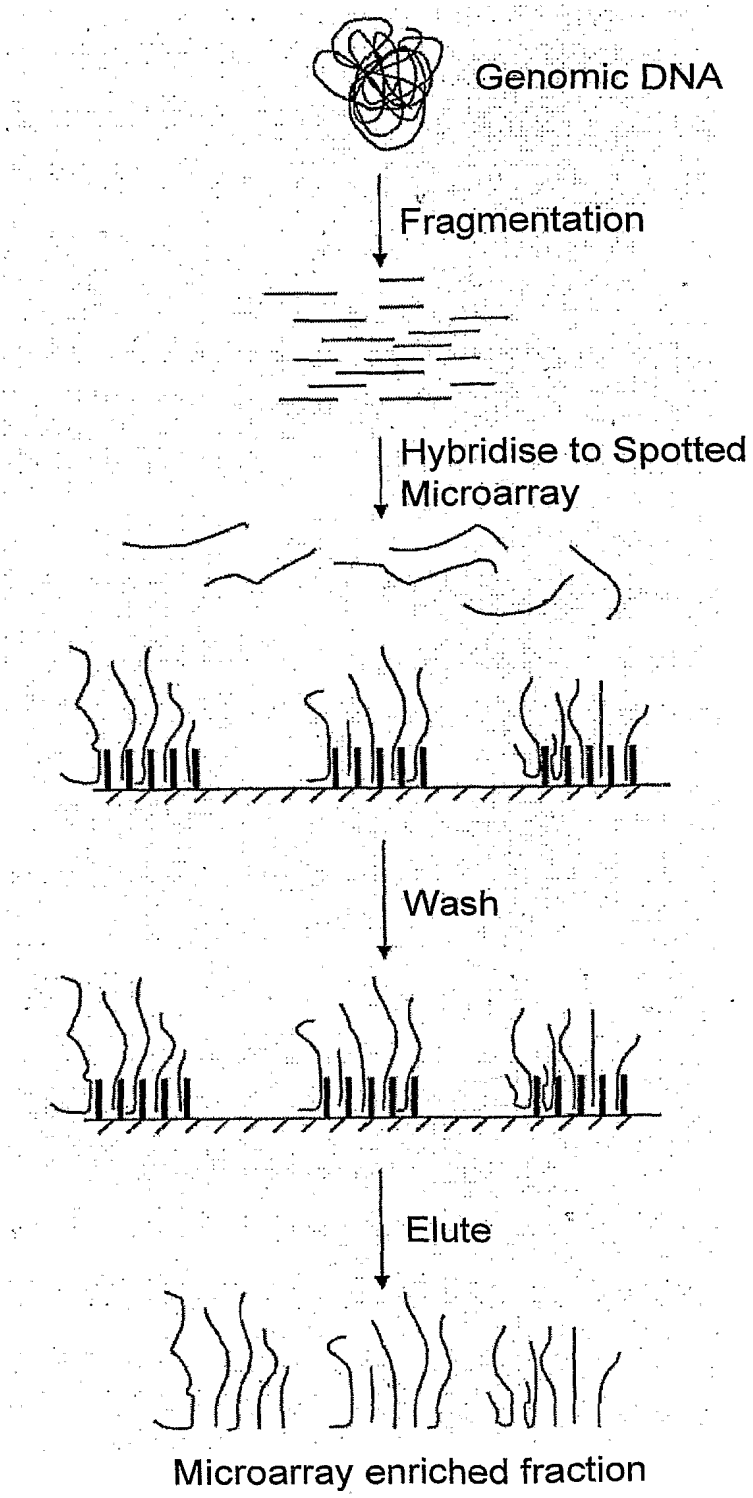

Genomic DNA
Fragmentation
Hybridise to Spotted Microarray
Wash
Elute
Microarray enriched fraction

Fig. 1a

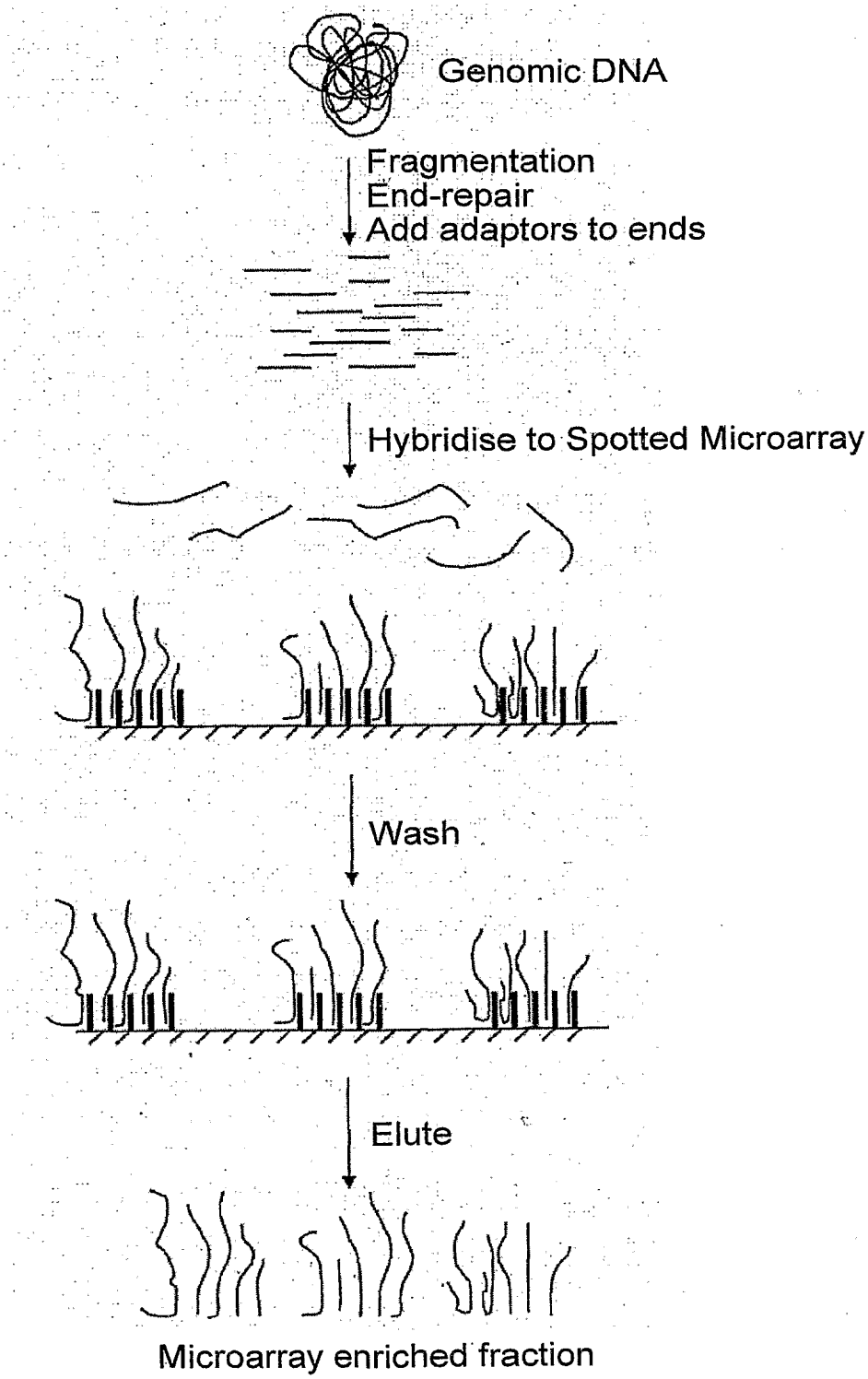

Genomic DNA
Fragmentation
End-repair
Add adaptors to ends
Hybridise to Spotted Microarray
Wash
Elute
Microarray enriched fraction

Fig. 1b

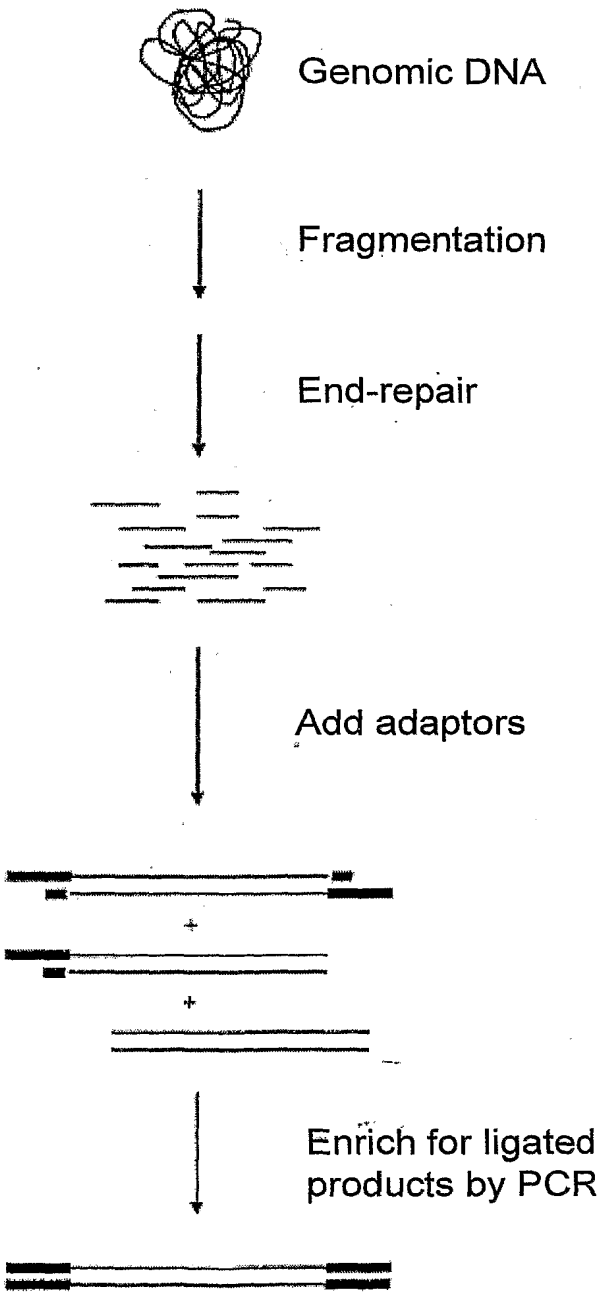

Genomic DNA
Fragmentation
End-repair
Add adaptors
Enrich for ligated products by PCR

Fig. 2a

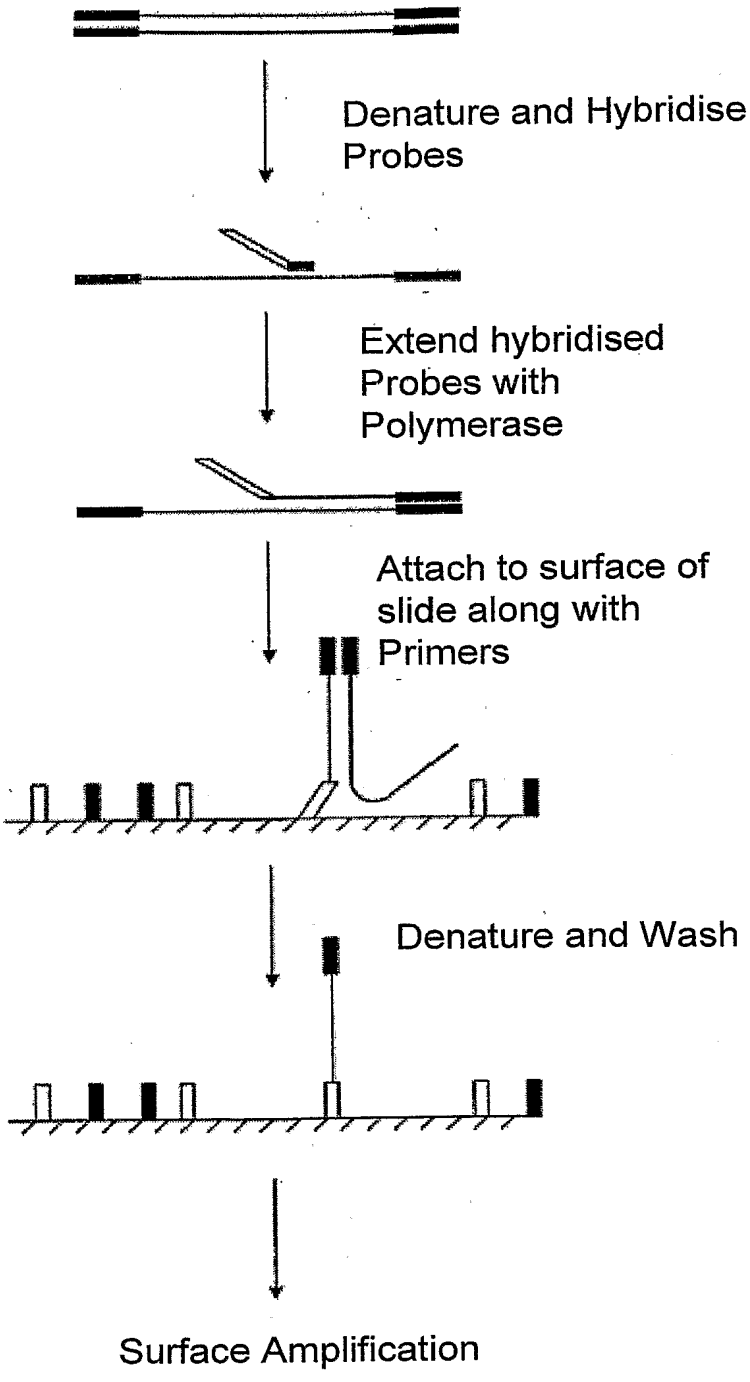

Denature and Hybridise Probes
Extend hybridised Probes with Polymerase
Attach to surface of slide along with Primers
Denature and Wash
Surface Amplification

Fig. 2b

INTERNATIONAL SEARCH REPORT

International application No
PCT/GB2006/004244

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12Q1/68 C12N15/10

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C12Q 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, MEDLINE, BIOSIS, EMBASE

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	EP 1 184 466 A2 (AFFYMETRIX INC [US]) 6 March 2002 (2002-03-06) see whole doc. esp. claims, par.5,51	1-17
X	WO 02/083943 A2 (EPIGENOMICS AG [DE]; DISTLER JUERGEN [DE]) 24 October 2002 (2002-10-24) see whole doc. esp. claims	1-17
A	US 2003/082543 A1 (SU XING [US] ET AL) 1 May 2003 (2003-05-01) see whole doc. esp. claims	4,5

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

☒ See patent family annex.

* Special categories of cited documents :

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

"E" earlier 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
14 February 2007	09/03/2007

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

2

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

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/GB2006/004244

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 1184466 A2	06-03-2002	CA 2356396 A1	26-02-2002
		JP 2002330783 A	19-11-2002
		US 2003148273 A1	07-08-2003
WO 02083943 A2	24-10-2002	DE 10119468 A1	24-10-2002
		EP 1379700 A2	14-01-2004
		US 2005009020 A1	13-01-2005
US 2003082543 A1	01-05-2003	US 2004132056 A1	08-07-2004

IDT Exhibit 1020.6011

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
16 January 2003 (16.01.2003)

PCT

(10) International Publication Number
WO 03/004690 A2

(51) International Patent Classification[7]: C12Q

(21) International Application Number: PCT/US02/21579

(22) International Filing Date: 8 July 2002 (08.07.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/303,576 6 July 2001 (06.07.2001) US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier application:
US 60/303,576 (CIP)
Filed on 6 July 2001 (06.07.2001)

(71) Applicant *(for all designated States except US)*: **454® CORPORATION** [US/US]; 20 Commercial Street, Branford, CT 06405 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **WEINER, Michael** [US/US]; 52 Leighton Trail, Guilford, CT 06437 (US). **ATTIYA, Said** [CA/US]; 57 Briarwood Lane, Apartment 13B, Branford, CT 06405 (US). **CRENSHAW, Hugh** [US/US]; 121 Pinecrest Road, Durham, NC 27705 (US). **ROTHBERG, Jonathan, M.** [US/US]; 1701 Moose Hill Road, Guilford, CT 06437 (US). **MATSON, Steve** [US/US]; 15 Withington Lane, Harvard, MA 01454 (US).

(74) Agent: **ELRIFI, Ivor, R.**; Mintz, Levin, Cohn, Ferris, Glovsky, and Popeo, P.C., One Financial Center, Boston, MA 02111 (US).

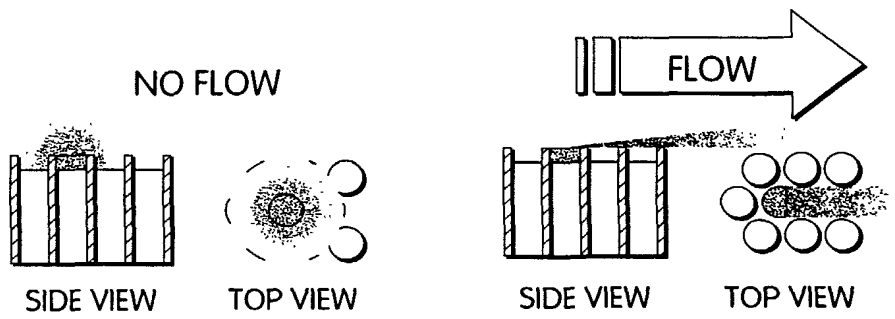
(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

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

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

(54) Title: METHOD FOR ISOLATION OF INDEPENDENT, PARALLEL CHEMICAL MICRO-REACTIONS USING A POROUS FILTER



(57) Abstract: The present invention relates to the field of fluid dynamics. More specifically, this invention relates to methods and apparatus for conducting densely packed, independent chemical reactions in parallel in a substantially two-dimensional array. Accordingly, this invention also focuses on the use of this array for applications such as DNA sequencing, most preferably pyrosequencing, and DNA amplification.

WO 03/004690 A2

METHOD FOR ISOLATION OF INDEPENDENT, PARALLEL CHEMICAL MICRO-REACTIONS USING A POROUS FILTER

FIELD OF THE INVENTION

The invention describes method and apparatus for conducting densely packed, independent chemical reactions in parallel in a substantially two-dimensional array comprising a porous filter.

BACKGROUND OF THE INVENTION

High throughput chemical synthesis and analysis are rapidly growing segments of technology for many areas of human endeavor, especially in the fields of material science, combinatorial chemistry, pharmaceuticals (drug synthesis and testing), and biotechnology (DNA sequencing, genotyping).

Increasing throughput in any such process requires either that individual steps of the process be performed more quickly, with emphasis placed on accelerating rate-limiting steps, or that larger numbers of independent steps be performed in parallel. Examples of approaches for conducting chemical reactions in a high-throughput manner include such techniques as:

- Performing a reaction or associated processing step more quickly:
  - Adding a catalyst
  - Performing the reaction at higher temperature
- Performing larger numbers of independent steps in parallel:
  - Conducting simultaneous, independent reactions with a multi-reactor system.

A common format for conducting parallel reactions at high throughput comprises two-dimensional (2-D) arrays of individual reactor vessels, such as the 96-well or 384-well microtiter plates widely used in molecular biology, cell biology, and other areas. Individual reagents, solvents, catalysts, and the like are added sequentially and/or in parallel to the appropriate wells in these arrays, and multiple reactions subsequently proceed in parallel. Individual wells may be further isolated from adjacent wells and/or from the environment by sealing means (e.g., a tight-fitting cover or adherent plastic sheet) or they may remain open. The base of the wells in such microtiter plates may or

may not be provided with filters of various pore sizes. The widespread application of robotics has greatly increased the speed and reliability of reagent addition, supplementary processing steps, and reaction monitoring – thus greatly increasing throughput.

Further increasing the number of microvessels or microreactors incorporated in such 2-D arrays has been a focus of much research, and this has been and is being accomplished by miniaturization. For instance, the numbers of wells that can be molded into plastic microtiter plates has steadily increased in recent years -- from 96, to 384, and now to 1536. Efforts to further increase the density of wells are ongoing (*e.g.* Matsuda and Chung, 1994; Michael *et al.*, 1998; Taylor and Walt, 1998).

Attempts to make arrays of microwells and microvessels for use as microreactors has also been a focal point for development in the areas of microelectromechanical and micromachined systems, applying and leveraging some of the microfabrication techniques originally developed for the microelectronics industry (see Matsuda and Chung, 1994; Rai-Choudhury, 1997; Madou, 1997; Cherukuri *et al.*, 1999; Kane *et al.* 1999; Anderson *et al.*, 2000; Dannoux *et al.*, 2000; Deng *et al.*, 2000; Zhu *et al.*, 2000; Ehrfeld *et al.*, 2000).

Yet another widely applied approach for conducting miniaturized and independent reactions in parallel involves spatially localizing or immobilizing at least some of the participants in a chemical reaction on a surface, thus creating large 2-D arrays of immobilized reagents. Reagents immobilized in such a manner include chemical reactants, catalysts, other reaction auxiliaries, and adsorbent molecules capable of selectively binding to complementary molecules. (For purposes of this patent specification, the selective binding of one molecule to another – whether reversible or irreversible – will be referred to as a reaction process, and molecules capable of binding in such a manner will be referred to as reactants.) Immobilization may be arranged to take place on any number of substrates, including planar surfaces and/or high-surface-area and sometimes porous support media such as beads or gels. Microarray techniques involving immobilization on planar surfaces have been commercialized for the hybridization of oligonucleotides (e.g. by Affymetrix Inc.) and for target drugs (*e.g.* by Graffinity AB).

A major obstacle to creating microscopic, discrete centers for localized reactions is that restricting unique reactants and products to a single, desired reaction center is frequently difficult. There are two dimensions to this problem. The first is that "unique" reagents – i.e., reactants and other reaction auxiliaries that are meant to differ from one

IDT Exhibit 1020.6014

reaction center to the next -- must be dispensed or otherwise deployed to particular reaction centers and not to their nearby neighbors. (Such "unique" reagents are to be distinguished from "common" reagents like solvents, which frequently are meant to be brought into contact with substantially all the reaction centers simultaneously and in parallel.) The second dimension of this problem has to do with restricting reaction products to the vicinity of the reaction center where they were created – i.e., preventing them from traveling to other reaction centers with attendant loss of reaction fidelity.

Focusing first on the problem of directed reagent addition, if the reaction center consists of a discrete microwell – with the microvessel walls (and cover, if provided) designed to prevent fluid contact with adjacent microwells -- then delivery of reagents to individual microwells can be difficult, particularly if the wells are especially small. For example, a reactor measuring 100 μm X 100 μm X 100 μm has a volume of only 1 nanoliter -- and this can be considered a relatively large reactor volume in many types of applications. Even so, reagent addition in this case requires that sub-nanoliter volumes be dispensed with a spatial resolution and precision of at least ±50 μm. Furthermore, addition of reagents to multiple wells must be made to take place in parallel, since sequential addition of reagents to at most a few reactors at a time would be prohibitively slow. Schemes for parallel addition of reagents with such fine precision exist, but they entail some added complexity and cost. For example, the use of inkjet print technologies to deliver sub-nanoliter-sized drops to surfaces has been widely explored and developed (Gamble *et al.*, 1999; Hughes *et al.*, 2001; Rosetta, Inc.; Agilent, Inc.) However, evaporation of such small samples remains a significant problem that requires careful humidity control.

If, on the other hand, the reaction centers are brought into contact with a common fluid -- *e.g.*, if the microwells all open out onto a common volume of fluid at some point during the reaction or subsequent processing steps -- then reaction products (and excess and/or unconverted reactants) originating in one reaction microwell or vessel can travel to and contaminate adjacent reaction microwells. Such cross-contamination of reaction centers can occur (i) via bulk convection of solution containing reactants and products from the vicinity of one well to another, (ii) by diffusion (especially over reasonably short distances) of reactant and/or product species, or (iii) by both processes occurring simultaneously. If the individual chemical compounds that are produced at the discrete reaction centers are themselves the desired objective of the process (e.g., as is the case in

IDT Exhibit 1020.6015

combinatorial chemistry), then the yield and ultimate chemical purity of this "library" of discrete compounds will suffer as a result of any reactant and/or product cross-contamination that may occur. If, on the other hand, the reaction process is conducted with the objective of obtaining information of some type – e.g., information as to the sequence or composition of DNA, RNA, or protein molecules -- then the integrity, fidelity, and signal-to-noise ratio of that information may be compromised by chemical "cross-talk" between adjacent or even distant microwells.

Consider the case of a two-dimensional, planar array of reaction sites in contact with a bulk fluid (e.g., a solution containing a common reagent or a wash solvent), and presume that at least one of the reactants and/or products involved in the chemistry at a particular reaction site is soluble in the bulk fluid. (Alternatively, consider the case of an array of microvessels that all open out onto a common fluid; the analyses are similar.) In the absence of convective flow of bulk fluid, transport of reaction participants (and cross-contamination or "cross-talk" between adjacent reaction sites or microvessels) can take place only by diffusion. If the reaction site is considered to be a point source on a 2-D surface, the chemical species of interest (e.g., a reaction product) will diffuse radially from the site of its production, creating a substantially hemispherical concentration field above the surface (see Fig. 1).

The distance that a chemical entity can diffuse in any given time t may be estimated in a crude manner by considering the mathematics of diffusion (Crank, 1975). The rate of diffusive transport in any given direction x (cm) is given by Fick's law as

$$j = -D\frac{\partial C}{\partial x} \qquad \text{Eq. 1}$$

where j is the flux per unit area (g-mol/cm^2-s) of a species with diffusion coefficient D (cm^2/s), and $\partial C/\partial x$ is the concentration gradient of that species. The mathematics of diffusion are such that a characteristic or "average" distance an entity can travel by diffusion alone scales with the one-half power of both the diffusion coefficient and the time allowed for diffusion to occur. Indeed, to order of magnitude, this characteristic diffusion distance can be estimated as the square root of the product of the diffusion coefficient and time – as adjusted by a numerical factor of order unity that takes into account the particulars of the system geometry and initial and/or boundary conditions imposed on the diffusion process.

It will be convenient to estimate this characteristic diffusion distance as the root-mean-square distance d_{rms} that a diffusing entity can travel in time t:

IDT Exhibit 1020.6016

$$d_{rms} = \sqrt{2Dt} \qquad \text{Eq. 2}$$

As stated above, the distance that a diffusing chemical typically travels varies with the square root of the time available for it to diffuse -- and inversely, the time required for a diffusing chemical to travel a given distance scales with the square of the distance to be traversed by diffusion. Thus, for a simple, low-molecular-weight biomolecule characterized by a diffusion coefficient D of order $1 \cdot 10^{-5}$ cm^2/s, the root-mean-square diffusion distances d_{rms} that can be traversed in time intervals of 0.1 s, 1.0 s, 2.0 s, and 10 s are estimated by means of Equation 2 as 14 μm, 45 μm, 63 μm, and 141 μm, respectively.

Such considerations place an upper limit on the surface density or number per unit area of microwells or reaction sites that can be placed on a 2-D surface if diffusion of chemicals from one microwell or reaction site to an adjacent well or site (and thus cross-contamination) is to be minimized. More particularly, given that the species diffusivity and the time available for diffusion are such that d_{rms} is the characteristic diffusion distance as estimated with Equation 2, it is evident that adjacent microwells or reaction sites can be spaced no more closely to one another than a fraction of this distance d_{rms} if diffusion of reaction participants between them is to be held to a minimum. This, then, restricts the numbers of reaction sites that can be placed on a 2-D surface. More precise calculations of the actual concentration of a diffusing species at an adjacent microwell or reaction site can be performed by solving – with either analytical or numerical methods -- the partial differential equations that describe unsteady-state diffusion subject to appropriate initial and boundary conditions (Crank, 1975). However, the order of magnitude analysis provided here suffices to illustrate the magnitude of the problem that must be solved if multiple, parallel reactions are to be conducted independently in a high-density format without risk of chemical cross-talk or contamination from nearby reaction centers.

The issue of contamination of a reaction center or well by chemical products being generated at nearby reaction centers or microwells becomes even more problematic when reaction sites are arrayed on a 2-D surface (or wells are arranged in an essentially two-dimensional microtiter plate) over which a fluid flows (again, see Fig. 1). In this case, compounds produced at a surface reaction site or within a well undergo diffusive transport up and away from the surface (or out of the reaction wells), where they are

IDT Exhibit 1020.6017

subsequently swept downstream by convective transport of fluid that is passing through a flow channel in fluid communication with the top surface of the array.

Several options exist for decreasing the spacing between (and thus increasing the number per unit area) of reaction sites. For example:

(1) Discrete reaction centers can be connected with microscopic tubes or channels in a "microfluidics" approach as described, e.g., by Cherukuri *et al.* (1999). However, this approach entails complex microfabrication, assembly of microcomponents, and control of fluid flow.

(2) The reaction centers can be placed at the bottom of microwells, such that d_{rms} is arranged to be small as compared to the sum of the depth of the microwell plus the spacing between adjacent microwells. Such microwells can be formed, for example, by microfabrication or microprinting (*e.g.* Aoki, 1992; Kane *et al.*, 1999; Dannoux *et al.*, 2000; Deng *et al.*, 2000; Zhu *et al.*, 2000), or by etching the ends of a fused fiber optic bundle (*e.g.* Taylor and Walt, 2000; Illumina Inc.; 454 Corporation -- see, e.g., US patent 6,274,320, incorporated fully herein by reference). In these etched wells, the distance from the top to the bottom of the microwells must be traversed not only by reaction products (the escape of which it is desired to minimize) but by reactants as well (whose access it is desired not to impede). That is, if a reaction is confined to the base of a microwell, reactants must traverse the distance from the top to the bottom of the microwells by diffusion, potentially reducing the rate of reactant supply and possibly limiting the rate of reaction.

(3) The space between reaction centers can be filled with a medium in which the diffusing chemical has low diffusivity, thus reducing the rate of transport of said compound to adjacent centers. Again, however, this adds complexity and may impede (i.e., slow) access of reactant to the reaction site.

(4) If the diffusing species is charged, it may be possible to establish an electric field so as to counter diffusion, as exemplified, *e.g.*, by Nanogen, Inc. Creation of the appropriate electrodes, however, again adds to the complexity of fabrication, and regulation of voltages at the electrodes adds complexity to the control system.

IDT Exhibit 1020.6018

SUMMARY OF THE INVENTION

An alternative technique for densely packing microreactors in a substantially 2-D arrangement is described here. This technique provides not only dense, two-dimensional packing of reaction sites, microvessels, and reaction wells – but also provides for efficient delivery of reagents and removal of products by convective flow rather than by diffusion alone. This latter feature permits much more rapid delivery of reagents and other reaction auxiliaries -- and it permits faster and more complete removal of reaction products and by-products -- than has heretofore been possible using methods and apparatus described in the prior art.

In one aspect, the invention includes a confined membrane reactor array (CMRA) comprising (a) a microreactor element comprising an array of open microchannels or open microwells, the longitudinal axes of said microchannels or microwells arranged in a substantially parallel manner; and (b) a porous filter element in contact with the microreactor element to form a bottom to the microchannels or microwells, thereby defining a series of reaction chambers, wherein the porous filter element comprises a permselective membrane that blocks the passage of nucleic acids, proteins and beads there across, but permits the passage of low molecular weight solutes, organic solvents and water there across. In a preferred embodiment, the microreactor element comprises a plate formed from a fused fiber optic bundle, wherein the microchannels extend from the top face of the plate through to the bottom face of the plate. In another embodiment, the CMRA further comprises an additional porous support between the microreactor element and the porous filter element. In one embodiment, the porous filter element comprises an ultrafilter. In a further embodiment, the CMRA further comprises at least one mobile solid support disposed in each of a plurality of the microchannels of the microreactor element. The mobile solid support can be a bead. In a preferred embodiment, the mobile solid support has an enzyme and/or a nucleic acid immobilized thereon. In another aspect, the invention provides a method of making the CMRA comprising attaching a microreactor element to a porous filter element.

In a further aspect, the invention includes an unconfined membrane reactor array (UMRA) comprising a porous filter element against which molecules are concentrated by concentration polarization wherein discrete reaction chambers are formed in discrete locations on the surface of or within the porous filter element by depositing reactant molecules at discrete sites on or within the porous filter element. In a preferred

IDT Exhibit 1020.6019

embodiment, the reaction chambers are formed by depositing mobile solid supports having said reactant molecules immobilized thereon, on the surface of, or within, the porous element. In one embodiment, the porous filter element comprises an ultrafilter. In another embodiment, the mobile solid support is a bead. In another embodiment, the mobile solid support has an enzyme and/or a nucleic acid immobilized thereon.

In another aspect, the invention includes a UMRA comprising (a) a porous membrane with discrete reaction sites formed by depositing mobile solid supports having said reactant molecules thereon, on the surface of, or within the porous membrane; (b) a nucleic acid template immobilized to a solid support; and (c) optionally, at least one immobilized enzyme. As used herein, the term discrete reaction sites refers to individual reaction centers for localized reactions whereby each site has unique reactants and products such that there is no cross-contamination between adjacent sites. In one embodiment, the mobile solid support is a bead. In another embodiment, the porous membrane is a nylon membrane. In another embodiment, the porous membrane is made of a woven fiber. In a preferred embodiment, the porous membrane pore size at least 0.02 μm. In another embodiment, the solid support is selected from the group consisting of a bead, glass surface, fiber optic or the porous membrane. In another embodiment, the immobilized enzyme is immobilized to a bead or the porous membrane. In one embodiment, the immobilized enzyme is selected from the group consisting of ATP sulfurylase, luciferase, hypoxanthine phosphoribosyltransferase, xanthine oxidase, uricase or peroxidase.

In another aspect, the invention includes an array comprising (a) a first porous membrane with a plurality of discrete reaction sites disposed thereon, and/or within, wherein each reaction site has immobilized template adhered to the surface; and (b) a second porous membrane with at least one enzyme located on the surface of, and/or within, the membrane, wherein the second porous membrane is in direct contact with the first porous membrane.

In another aspect, the invention provides a CMRA comprising an array of open microchannels or microwells attached to a porous filter or membrane. In one embodiment, the CMRA further comprises a mechanical support, wherein the mechanical support separates the microchannels from the porous membrane. In a preferred embodiment, the mechanical support is selected from the group consisting of plastic mesh, wire screening or molded or machined spacers.

IDT Exhibit 1020.6020

In another aspect, the invention includes an apparatus for determining the nucleic acid sequence in a template nucleic acid polymer, comprising: (a) a CMRA or UMRA; (b) nucleic acid delivery means for introducing template nucleic acid polymers to the discrete reaction sites; (c) nucleic acid delivery means to deliver reagents to the reaction sites to create a polymerization environment in which the nucleic acid polymers will act as template polymers for the synthesis of complementary nucleic acid polymers when nucleotides are added; (d) convective flow delivery means to immobilize reagents to the porous membrane; (e) detection means for detecting the formation of inorganic pyrophosphate enzymatically; and (f) data processing means to determine the identity of each nucleotide in the complementary polymers and thus the sequence of the template polymers. In one embodiment, the detection means is a CCD camera. In another embodiment, the data processing means is a computer.

In another aspect, the invention provides an apparatus for processing a plurality of analytes, the apparatus comprising: (a) a CMRA or an UMRA; (b) fluid means for delivering processing reagents from one or more reservoirs to the flow chamber so that the analytes disposed therein are exposed to the reagents; and (c) detection means for detecting a sequence of optical signals from each of the reaction sites, each optical signal of the sequence being indicative of an interaction between a processing reagent and the analyte disposed in the reaction site, wherein the detection means is in communication with the reaction site. In one embodiment, the convective flow delivery means is a peristaltic pump.

In another aspect, the invention includes an apparatus for determining the base sequence of a plurality of nucleotides on an array, the apparatus comprising: (a) a CMRA or UMRA; (b) reagent delivery means for adding an activated nucleotide 5'-triphosphate precursor of one known nitrogenous base to a reaction mixture to each reaction site, each reaction mixture comprising a template-directed nucleotide polymerase and a single-stranded polynucleotide template hybridized to a complementary oligonucleotide primer strand at least one nucleotide residue shorter than the templates to form at least one unpaired nucleotide residue in each template at the 3'-end of the primer strand, under reaction conditions which allow incorporation of the activated nucleoside 5'-triphosphate precursor onto the 3'-end of the primer strands, provided the nitrogenous base of the activated nucleoside 5'-triphosphate precursor is complementary to the nitrogenous base of the unpaired nucleotide residue of the templates; (c) detection means for detecting whether or not the nucleoside 5'-triphosphate precursor was incorporated into the primer

IDT Exhibit 1020.6021

strands in which incorporation of the nucleoside 5'-triphosphate precursor indicates that the unpaired nucleotide residue of the template has a nitrogenous base composition that is complementary to that of the incorporated nucleoside 5'-triphosphate precursor; (d) means for sequentially repeating steps (b) and (c), wherein each sequential repetition adds and detects the incorporation of one type of activated nucleoside 5'-triphosphate precursor of known nitrogenous base composition; and (e) data processing means for determining the base sequence of the unpaired nucleotide residues of the template in each reaction chamber from the sequence of incorporation of said nucleoside precursors.

In another aspect, the invention includes an apparatus for determining the nucleic acid sequence in a template nucleic acid polymer, comprising: (a) a CMRA or UMRA; (b) nucleic acid delivery means for introducing a template nucleic acid polymers onto the reaction sites; (c) nucleic acid delivery means to deliver reagents to the reaction chambers to create polymerization environment in which the nucleic acid polymers will act as a template polymers for the synthesis of complementary nucleic acid polymers when nucleotides are added; (d) reagent delivery means for successively providing to the polymerization environment a series of feedstocks, each feedstock comprising a nucleotide selected from among the nucleotides from which the complementary nucleic acid polymer will be formed, such that if the nucleotide in the feedstock is complementary to the next nucleotide in the template polymer to be sequenced said nucleotide will be incorporated into the complementary polymer and inorganic pyrophosphate will be released; (e) detection means for detecting the formation of inorganic pyrophosphate enzymatically; and (f) data processing means to determine the identity of each nucleotide in the complementary polymers and thus the sequence of the template polymers.

In one aspect, the invention includes a system for sequencing a nucleic acid comprising the following components: (a) a CMRA or UMRA; (b) at least one enzyme immobilized on a solid support; (c) means for flowing reagents over said porous membrane; (d) means for detection; and (e) means for determining the sequence of the nucleic acid.

In a further aspect, the invention includes a system for sequencing a nucleic acid comprising the following components: (a) a CMRA or UMRA; (b) at least one enzyme immobilized on a solid support; (c) means for flowing reagents over said porous membrane; (d) means for detection; and (e) means for determining the sequence of the nucleic acid.

IDT Exhibit 1020.6022

In another aspect, the invention provides a method for carrying out separate parallel independent reactions in an aqueous environment, comprising: (a) delivering a fluid containing at least one reagent to an array, using the CMRA of claim 1 or the UMRA of claim 9, wherein each of the reaction sites immersed in a substance such that when the fluid is delivered onto each reaction site, the fluid does not diffuse onto an adjacent site; (b) washing the fluid from the array in the time period after the starting material has reacted with the reagent to form a product in each reaction site; (c) sequentially repeating steps (a) and (b). In one embodiment, the product formed in any one reaction chamber is independent of the product formed in any other reaction chamber, but is generated using one or more common reagents. In another embodiment, the starting material is a nucleic acid sequence and at least one reagent in the fluid is a nucleotide or nucleotide analog. In a preferred embodiment, the fluid additionally comprises a polymerase capable of reacting the nucleic acid sequence and the nucleotide or nucleotide analog. In another embodiment, the method additionally comprises repeating steps (a) and (b) sequentially. In one embodiment, the substance is mineral oil. In a further embodiment, the reaction sites are defined by concentration polarization.

In one aspect, the invention includes a method of determining the base sequence of nucleotides in an array format, the method comprising the steps of: (a) adding an activated nucleoside 5'-triphopsphate precursor of one known nitrogenous base composition to a plurality of reaction sites localized on a CMRA or UMRA, wherein the reaction site is comprised of a template-directed nucleotide polymerase and a heterogenous population of single stranded templates hybridized to complementary oligonucleotide primer strands at least one nucleotide residue shorter than the templates to form at least one unpaired nucleotide residue in each template at the 3' end of the primer strand under reaction conditions which allow incorporation of the activated nucleoside 5'-triphosphate precursor onto the 3' end of the primer strand under reaction conditions which allow incorporation of the activated nucleoside 5'-triphosphate precursor onto the 3' end of the primer strands, provided the nitrogenous base of the activated nucleoside 5'-triphosphate precursor is complementary to the nitrogenous base of the unpaired nucleotide residue of the templates; (b) detecting whether or not the nucleoside 5'-triphosphate precursor was incorporated into the primer strands in which incorporation of the nucleoside 5'-triphosphate precursor indicates that the unpaired nucleotide residue of the template has a nitrogenous base composition that is complementary to that of the incorporated nucleoside 5'-triphosphate precursor; and (c) sequentially repeating steps (a)

IDT Exhibit 1020.6023

and (b), wherein each sequential repetition adds and detects the incorporation of one type of activated nucleoside 5'-triphosphate precursor of known nitrogenous base composition; (d) determining the base sequence of the unpaired nucleotide residues of the template from the sequence of incorporation of said nucleoside precursors.

In a preferred embodiment, the detection of the incorporation of the activated precursor is accomplished enzymatically. The enzyme utilized can be selected from the group consisting of ATP sulfurylase, luciferase, hypoxanthine phosphoribosyltransferase, xanthine oxidase, uricase or peroxidase. In one embodiment, the enzyme is immobilized to a solid support. In another embodiment, the solid support is selected from the group comprising a bead, glass surface, fiber optic or porous membrane.

In a further aspect, the invention includes a method of determining the base sequence of a plurality of nucleotides on an array, said method comprising: (a) providing a plurality of sample DNA's, each disposed within a plurality of reaction sites on a CMRA or UMRA; (b) detecting the light level emitted from a plurality of reaction sites on respective proportional of an optically sensitive device; (c) converting the light impinging upon each of said portions of said optically sensitive device into an electrical signal which is distinguishable from the signals from all of said other regions; (d) determining a light intensity for each of said discrete regions from the corresponding electrical signal; (e) recording the variations of said electrical signals with time.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1: Effect of flow across a 2-D array of microwells. On the left, there is no flow, and diffusion of compound (grey) creates a hemispherical chemical concentration gradient emanating from the microwell containing the reaction. On the right, flow carries compound downstream, creating a concentration plume. Cross-contamination of microwells resulting from diffusive and/or convective transport of compound to nearby and/or distant wells is minimized or avoided by the present invention.

FIG. 2: An integral or physical composite of a microchannel array and a porous membrane barrier forming a confined membrane reactor array (CMRA). The flow of fluid through the CMRA carries reaction participants along with it. Rejected macromolecules experience concentration polarization and so are concentrated and localized within the microchannels, without their being immobilized on a support.

IDT Exhibit 1020.6024

FIG. 3: Sequential addition of solutions containing different macromolecules permits their stratification within microchannels or microwells, thus forming the microscopic equivalent of a stacked column.

FIG. 4: Fluid flow atop and tangential to the CMRA surface may cause a pressure drop along the flow compartment that, in turn, may cause the pressure difference across the CMRA to decrease with distance along it. If this variation is significant, the flow rate through the CMRA will also vary with distance from the inlet, causing the delivery of molecules to the microchannels to be nonuniform.

FIG. 5: Radial or circumferential fluid inlets reduce the pressure variation over the CMRA, more nearly equalizing flow rates through the microchannels.

FIG. 6: Fluid flow restrictor, valve, or back-pressure regulator downstream of CMRA provides dominant flow resistance (i.e., large pressure drop) as compared to that associated with CMRA flow compartment – thus maintaining relatively uniform pressure difference across (and uniform fluid flow through) the CMRA. Shown here with optional fluid recirculation.

FIG. 7: An unconfined membrane reactor array (UMRA). Certain molecules may be concentrated adjacent to the porous filter by concentration polarization. Other molecules or particles are added in such a manner as to form a dispersed 2-D array of discrete reaction sites on or within the UMRA. Products formed at discrete reaction sites are carried by flow through the porous filter (e.g., a UF membrane), creating a plume of reaction products that extends downstream. Products are swept out of the UMRA before separate plumes merge, thus effectively minimizing or avoiding cross-contamination of independent reactions meant to occur at different reaction sites.

FIG. 8: An unconfined membrane reactor array (UMRA) comprised of an ultrafiltration membrane and a coarse filter, mesh, or other grossly porous matrix. The two filters are sandwiched together, with the coarser filter upstream. The latter can assist in stabilizing the layer of concentration-polarized molecules formed adjacent to the ultrafiltration membrane, and it can provide some resistance to lateral diffusion of

IDT Exhibit 1020.6025

reaction products. The coarser filter, mesh, or matrix may also provide mechanical support.

FIG. 9: A schematic of a pyrophosphate-based sequencing method with photon detection.

FIG. 10: Use of a CCD as a photodetector array to detect light production from a microchannel or microwell in a CMRA.

FIG. 11: Experimental setup for the convective flow embodiment described in Example 1.

FIG. 12: Effect of immobilization of the luciferase and ATP sulfurylase on the sepharose beads with Oligo seq 1. (A) Enzymes have not been immobilized on the beads. (B) Enzymes have been immobilized on the beads and the signal has been improved by factor of 3.5 times.

FIG. 13: An scanning electron micrograph (SEM) photo of the nylon weave filter that can be utilized for the CMRA and UMRA.

DETAILED DESCRIPTION OF THE INVENTION

Methods and apparati are described here for providing a dense array of discrete reaction sites, microreactor vessels, and/or microwells in a substantially two-dimensional configuration (see Fig. 2) and for charging such microreactors with reaction participants by effecting a convective flow of fluid normal to the plane of and through the array of reaction sites or microvessels. Reaction participants that may be charged to, concentrated, and contained within said reaction sites or microreactor vessels by methods of the present invention include high-molecular-weight reactants, catalysts, and other reagents and reaction auxiliaries. In the context of oligonucleotide sequencing and DNA/RNA analysis, such high-molecular-weight reactants include, for example, oligonucleotides, longer DNA/RNA fragments, and constructs thereof. These reactants may be free and unattached (if their molecular weight is sufficient to permit them to be

IDT Exhibit 1020.6026

contained by the method of the present invention), or they may be covalently bound to or otherwise associated with, e.g., high-molecular-weight polymers, high-surface-area beads or gels, or other supports known in the art. Examples of reaction catalysts that may be similarly delivered to and localized within said reaction sites or microvessels include enzymes, which may or may not be associated with or bound to solid-phase supports such as porous or non-porous beads.

The present invention also includes means for efficiently supplying relatively lower-molecular-weight reagents and reactants to said discrete reaction sites or microreactor vessels – as well as means for efficiently removing unconverted reactants and reaction products from said reaction sites or microvessels. More particularly, efficient reagent delivery and product removal are accomplished in the present invention by arranging for at least some convective flow of solution to take place in a direction normal to the plane of the substantially two-dimensional array of reaction sites or microreactor vessels – and thus past or through the discrete sites or microvessels, respectively, where chemical reaction takes place. In this instance, reactants and products will not necessarily be retained or concentrated at the reaction sites or within the reaction microvessels or microwells; indeed, it may be desired that certain reaction products be rapidly swept away from and/or out of said reaction sites or microvessels.

In addition to including means for providing a controlled convective flux of fluid normal to and across the substantially planar array of reaction sites or microreactors, the present invention also includes permselective, porous filter means capable of discriminating between large (i.e., high-molecular-weight) and small (i.e., low-molecular-weight) reaction participants. This filter means is capable of selectively capturing or retaining certain reaction participants while permitting others to be flushed through and/or out the bottom of the microreactor array. By the proper selection of porous filter and the judicious choice of convective flux rates, considerable control over the location, concentration, and fate of reaction participants can be realized.

CMRAs.

In a preferred embodiment, the apparatus of the present invention consists of an array of microreactor elements comprised of at least two functional elements that may take various physical or structural forms – namely, (i) a microreactor element comprised of an array of microchannels or microwells, and (ii) a porous filter element comprising, e.g., a porous film or membrane in the form of a sheet or thin layer (see, *e.g.*, Figure 13).

IDT Exhibit 1020.6027

These two elements are arranged next to and in close proximity or contact with one another, with the plane of the microchannel/microvessel element parallel to the plane of the porous filter element. For the sake of definiteness, the side of this composite structure containing the microchannel or microvessel array will be referred to hereinafter as the "top", while the side defined by the porous filter will be referred to as the "bottom" of the structure.

The microchannel or microvessel element consists of a collection of numerous microchannels, with the longitudinal axes of said microchannels being arranged in a substantially parallel manner, and with the downstream ends of said channels being in functional contact with a porous membrane or other filter element. The porous filter or membrane is chosen to be permselective – i.e., to block the passage of certain species like particles, beads, or macromolecules (e.g., proteins and DNA), while permitting the passage of relatively low-molecular-weight solutes, organic solvents, and water. The aspect ratio of the microchannels (i.e., their height- or length-to-diameter ratio) may be small or large, and their cross-section may take any of a number of shapes (e.g., circular, rectangular, hexagonal, etc.). As discussed further below, it is not at all essential that the microchannel walls be continuous or regular; indeed, highly porous, "spongy" matrices with interconnecting pores communicating between adjacent channels will also serve functionally as "microchannel" elements, despite the fact that they will not necessarily contain discrete or functionally confined or isolated microchannels *per se*. (This embodiment is described in more detail in discussion that follows and in Fig. 8).

In many cases it will be appropriate to consider the entire array assembly (i.e., the combination of microchannel/microvessel element plus porous filter element) as a single substantially two-dimensional structure comprised of either an integral or a physical composite, as described further below. For the sake of brevity, all such systems of the present invention that are comprised of composite structures containing at minimum a microchannel or microvessel element and a porous filter or membrane element will be referred to henceforth as "confining membrane reactor arrays" or CMRAs. The CMRAs of the present invention will be seen to possess some of the general structural features and functional attributes of commercially available microtiter filter plates of the sort commonly used in biology laboratories, wherein porous filter disks are molded or otherwise incorporated into the bottoms of plastic wells in 96-well plates. However, the CMRA is differentiated from these by the unparalleled high density of discrete reaction sites that it provides, by its unique construction, and by the novel and uniquely powerful

IDT Exhibit 1020.6028

way in which it can be operated to perform high-throughput chemistries – for example, of the sorts applied to the amplification and/or analysis of DNA.

The composite microreactor/filter structure – i.e., the CMRA of the present invention -- can take several physical forms; as alluded to above, two such forms are represented by physical composites and integral composites, respectively. As regards the former, the two functional elements of the structure – that is, the microchannel array and the porous filter – are provided as separate parts or components that are merely laid side-by-side, pressed together, or otherwise attached in the manner of a sandwich or laminate. This structural embodiment will be referred to hereinafter as a "physical composite". Additional porous supports (e.g., fine wire mesh or very coarse filters) and/or spacing layers may also be provided where warranted to provide mechanical support for the finely porous filter element and to ensure good contact between the microchannel and porous filter elements. Plastic mesh, wire screening, molded or machined spacers, or similar structures may be provided atop the CMRA to help define a compartment for tangential flow of fluid across the top of the CMRA, and similar structures may be provided beneath the CMRA to provide a pathway for egress of fluid that has permeated across the CMRA.

Alternatively, the two functional elements of the CMRA may be part and parcel of a single, one-piece composite structure – more particularly, an "integral composite". An integral composite has one surface that is comprised of a finely porous "skin" region that is permselective -- i.e., that permits solvent and low-molecular-weight solutes to permeate, but that retains or rejects high-molecular-weight solutes (e.g., proteins, DNA, etc.), colloids, and particles. However, the bulk of this structure's through-thickness will be comprised of microchannels and/or large voids or macropores that are incapable of exhibiting permselectivity by virtue of the very large size of the microchannels or voids contained therein.

Many synthetic membranes of the type employed in ultrafiltration processes and generally known as "ultrafilters" are known in the art and can be described as "integral composites" for present purposes (Kulkarni *et al.*, 1992; Eykamp, 1995). Ultrafiltration membranes are generally regarded as having effective pore sizes in the range of a nanometer or so up to at most a hundred nanometers. As a consequence, ultrafilters are capable of retaining species with molecular weights ranging from several hundred Daltons to several hundred thousand Daltons and up. Most UF membranes are described in terms of a nominal molecular-weight cut-off (MWCO). The MWCO can be defined in

IDT Exhibit 1020.6029

various ways, but commonly a membrane's MWCO corresponds to the molecular weight of the smallest species for which the membrane exhibits greater than 90% rejection.

Many ultrafilters are asymmetric. That is, they are characterized by having a thin (micron- or even submicron-thick) skin layer containing nanometer-size pores – said skin layer being integrally supported by and inseparable from a much thicker substrate region of order 100 μms and more in thickness, and with said substrate region having pore diameters that are generally at least an order of magnitude larger than those in the skin. Whereas the pores in the skin region of an ultrafiltration membrane that give rise to its permselectivity are typically in the range from a few to a few hundred nanometers, the voids in the substrate region of said ultrafilters might be as large as a few tenths of a micron to many microns in diameter. Most polymeric ultrafiltration membranes are generally prepared in a single membrane casting step, with both the ultraporous skin layer and the substrate region necessarily comprised of the same, continuous material.

An inorganic membrane filter with utility as a confined membrane reactor array (CMRA) of the integral composite type is exemplified by the Anopore™ and Anodisc™ families of ultrafiltration membranes sold, for example, by Whatman PLC (see, for instance, http://www.whatman.plc.uk/index2.html). These high-purity alumina membranes are prepared by an electrochemical oxidation process that gives rise to a rather unique membrane morphology (Furneaux *et al.*, 1989; Martin, 1994; Mardilovich *et al.*, 1995; Asoh *et al.*, 2001). In particular, such membranes provide both an array of parallel microchannels capable of housing independent reactions and a more finely porous permselective surface region capable of rejecting selected reaction participants. Commercially available alumina membranes (e.g., from Whatman) contain a densely packed array of regular, nearly hexagonal-shaped channels, nominally 0.2 μm in diameter, with no lateral crossovers between adjacent channels. The membranes have an overall thickness of about 60 μm, with almost the entire thickness being comprised of these 0.2-μm-diameter channels. However, on one surface is located a much more finely porous – even ultraporous – surface region of order 1 μm in thickness, said surface region containing pores characterized by pore diameters of about 0.02 μm or 20 nm – i.e., in the ultrafiltration range. These membranes have the additional interesting and useful optical property of being substantially transparent when wet with aqueous solutions – a feature that permits any light generated by chemical reaction within them to be readily detected.

IDT Exhibit 1020.6030

In conventional applications where Anopore™ or Anodisc™ membranes are used to concentrate and/or separate proteins by ultrafiltration, the higher fluid pressure will normally be applied to the side of the membrane characterized by the smaller pores. That is, fluid generally is made to flow first through the thin, permselective surface region of the UF membrane and only then through the much thicker substrate region with its larger, substantially parallel microchannels; in this event, the substrate region of the membrane serves merely to provide physical support and mechanical integrity. As explained in more detail below, however, the use of membranes of this type in CMRAs entails reversing the direction of convective flow through them, such that fluid flows first through the thick substrate region containing the parallel microchannels within which reaction occurs – and only then through the more finely porous and permselective surface region. In a sense, then, these ultrafiltration membranes are oriented "upside-down" (i.e., rejecting side "down" and opposite the higher-pressure side) when used as CMRAs – at least as compared to their more usual orientation in ordinary ultrafiltration applications.

Alternatively, ultrafiltration membranes may also be used as CMRA components with their more finely porous, rejecting side "up" – i.e., in contact with the higher fluid pressure – such that fluid flows first through the permselective skin region and then through the more grossly porous, spongy substrate region. However, in such instances it will generally be the case that the CMRA will be of the "physical composite" type – with a separate and distinct microchannel- or microvessel-containing element placed "above" and in intimate contact with the skin region of the ultrafilter. In this case, the order of fluid flow will be first through the microchannel-containing element, then across the thin skin of the permselective region of the integral composite UF membrane, and then finally across the substrate region of the UF membrane. Additional supporting layers may optionally be provided underneath the physical composite-type CMRA, and plastic mesh, wire screens, or like materials may be used atop the composite to help define a compartment for tangential flow of fluid across the top of the CMRA as before.

In contrast to the operation of many prior-art microreactor arrays, wherein diffusion of reactants into and products back out of an array of microvessels occurs solely by diffusion, the operation of the confined membrane reactor arrays of the present invention entails providing for a modest convective flux through the CMRA. In particular, a small pressure difference is applied from the top to the bottom surface of the CMRA sufficient to establish a controlled convective flux of fluid through the structure in a direction normal to the substantially planar surface of the structure. Fluid is thus made

IDT Exhibit 1020.6031

to flow first through the microchannel element and then subsequently across the porous filter element. This convective flow enables the loading and entrapment within the microchannel element of the CMRA of various high-molecular-weight reagents and reaction auxiliaries that are retained by the porous filter element of the CMRA. By the same token, this convective flow enables the rapid delivery to the site of reaction of low-molecular-weight reactants -- and the efficient and complete removal of low-molecular-weight reaction products from the site of their production. Particularly important is the fact that the convective flow serves to impede or substantially prevent the back-diffusion of reaction products out of the upstream ends of the microchannels, where otherwise they would be capable of contaminating adjacent or even distant microreactor vessels.

The relative importance of convection and diffusion in a transport process that involves both mechanisms occurring simultaneously can be gauged with the aid of a dimensionless number – namely, the Peclet number *Pe*. This Peclet number can be viewed as a ratio of two rates or velocities – namely, the rate of a convective flow divided by the rate of a diffusive "flow" or flux. More particularly, the Peclet number is a ratio of a characteristic flow velocity V (in cm/s) divided by a characteristic diffusion velocity D/L (also expressed in units of cm/s) – both taken in the same direction:

$$Pe = \frac{VL}{D} \qquad \text{Eq. 3}$$

In Equation 3, V is the average or characteristic speed of the convective flow, generally determined by dividing the volumetric flow rate Q (in cm^3/s) by the cross-sectional area A (cm^2) available for flow. The characteristic length L is a representative distance or system dimension measured in a direction parallel to the directions of flow and of diffusion (i.e., in the direction of the steepest concentration gradient) and selected to be representative of the typical or "average" distance over which diffusion occurs in the process. And finally D (cm^2/s) is the diffusion coefficient for the diffusing species in question. (An alternative but equivalent formulation of the Peclet number *Pe* views it as the ratio of two characteristic times – namely, of representative times for diffusion and convection. Equation 3 for the Peclet number can equally well be obtained by dividing the characteristic diffusion time L^2/D by the characteristic convection time L/V.)

The convective component of transport can be expected to dominate over the diffusive component in situations where the Peclet number *Pe* is large compared to unity. Conversely, the diffusive component of transport can be expected to dominate over the convective component in situations where the Peclet number *Pe* is small compared to

IDT Exhibit 1020.6032

unity. In extreme situations where the Peclet number is either very much larger or very much smaller than one, transport may be accurately presumed to occur either by convection or by diffusion alone, respectively. Finally, in situations where the estimated Peclet number is of order unity, then both convection and diffusion can be expected to play significant roles in the overall transport process.

The diffusion coefficient of a typical low-molecular-weight biomolecule will generally be of the order of 10^{-5} cm^2/s (e.g., $0.52 \cdot 10^{-5}$ cm/s for sucrose, and $1.06 \cdot 10^{-5}$ cm/s for glycine). Thus, for chemical reaction sites, microchannels, or microvessels separated by a distance of 100 μm (i.e., 0.01 cm), the Peclet number *Pe* for low-molecular-weight solutes such as these will exceed unity for flow velocities greater than about 10 μm/sec (0.001 cm/s). For sites or vessels separated by only 10 μm (i.e., 0.001 cm), the Peclet number *Pe* for low-molecular-weight solutes will exceed unity for flow velocities greater than about 100 μm/sec (0.01 cm/s). Convective transport is thus seen to dominate over diffusive transport for all but very slow flow rates and/or very short diffusion distances.

Where the molecular weight of a diffusible species is substantially larger -- for example as it is with large biomolecules like DNA/RNA, DNA fragments, oligonucleotides, proteins, and constructs of the former -- then the species diffusivity will be corresponding smaller, and convection will play an even more important role relative to diffusion in a transport process involving both mechanisms. For instance, the aqueous-phase diffusion coefficients of proteins fall in about a 10-fold range (Tanford, 1961). Protein diffusivities are bracketed by values of $1.19 \cdot 10^{-6}$ cm^2/s for ribonuclease (a small protein with a molecular weight of 13,683 Daltons) and $1.16 \cdot 10^{-7}$ cm^2/s for myosin (a large protein with a molecular weight of 493,000 Daltons). Still larger entities (e.g., tobacco mosaic virus or TMV at 40.6 million Daltons) are characterized by still lower diffusivities (in particular, $4.6 \cdot 10^{-8}$ cm^2/s for TMV) (Lehninger, 1975). The fluid velocity at which convection and diffusion contribute roughly equally to transport (i.e., *Pe* of order unity) scales in direct proportion to species diffusivity.

With the aid of the Peclet number formalism it is possible to gauge the impact of convection on reactant supply to -- and product removal from – microreactor vessels. On the one hand, it is clear that even modest convective flows can appreciably increase the speed at which reactants are delivered to the interior of the microchannels or microwells in a CMRA structure. In particular, suppose for the sake of simplicity that the criteria for

IDT Exhibit 1020.6033

roughly equal convective and diffusive flows is considered to be $Pe = 1$. One may then estimate that a convective flow velocity of the order of only 0.004 cm/s will suffice to carry reactant into a 25-μm-deep well at roughly the same rate as it could be supplied to the bottom of the well by diffusion alone, given an assumed value for reactant diffusivity of $1 \cdot 10^{-5}$ cm^2/s. The corresponding flow velocity required to match the rate of diffusion of such a species from the bottom to the top of a 2.5-μm-deep microwell is estimated to be of order 0.04 cm/s. Clearly, flow velocities through the CMRA much higher than this are possible, thereby illustrating the degree to which a modest convective flow can augment the diffusive supply of reactants to CMRA microchannels and wells.

By the same token, the Peclet number formalism assists in understanding how effective even a modest trans-CMRA convective flow component can be in impeding or substantially preventing the back-diffusion of excess unconverted reactants and/or reaction products and by-products into the flow compartment located above the "top" or upstream surface of the CMRA. Preventing such back-diffusion is critical since, once a compound escapes into this transverse flow compartment, it may readily diffuse and/or be swept along into the neighborhood of the mouths of adjacent microchannels, thus contributing to cross-contamination or cross-talk between reaction sites or microvessels. The magnitude of the "top-to-bottom" convective flow through the microvessels or microchannels of a CMRA that can be expected to have a significant effect in reducing the rate of diffusive compound loss out the top of the microchannels or microvessels can again be estimated to order of magnitude by setting the Peclet number equal to unity in Equation 3 -- with the understanding that, in this instance, the convective and diffusive flows will occur in opposite directions and oppose each other. The CMRA may be operated at microchannel Peclet numbers significantly greater than unity in situations where it is particularly critical that there be little or no escape of potential contaminant compounds from the top surface.

It may be noted that compounds which permeate across the porous filter element and out of the CMRA are made to flow straight-away out of the device, ideally in a direction substantially normal to the plane of the CMRA. By this and other fluid management strategies (e.g., the provision of thick, spongy pads underlying the CMRA), any potential for cross-contamination between nearby CMRA microchannels via the bottom surface of the structure may readily be avoided.

Thus far, it has been assumed that the freely diffusible reactants and products discussed in the paragraphs immediately above will experience only insignificant

IDT Exhibit 1020.6034

retention or rejection by the porous filter or membrane of the CMRA. The discussion turns now to focus on the fate of high-molecular-weight reaction participants and auxiliaries within the CMRA – in particular, of macromolecules including proteins, of oligo- and polynucleotides (and constructs thereof), and of otherwise low-molecular-weight reagents attached to high-molecular-weight polymers, nano- and micro-particles, or even beads. In these latter instances, said attachment facilitates the retention of reagents within the microvessels or microchannels of the CMRA.

A particularly useful feature of the present invention is its ability to permit the efficient and controlled loading of macromolecules and microparticles into said microvessels or microchannels by simple pressure-driven filtration across a suitably porous filter element. As discussed above, where the filter element has the ability to substantially reject and contain soluble macromolecules while permitting microsolutes to pass relatively freely, the filter is frequently referred to as an ultrafilter or ultrafiltration membrane – and the process is referred to as ultrafiltration.

Ultrafiltration (UF) is a process normally used to separate macromolecules from solutions according to the size and shape of the macromolecules relative to the pore size and morphology of the membrane. UF is a pressure-driven membrane permeation process, with flux of solvent (e.g., water) generally being proportional to an effective pressure difference that is equal to the applied hydraulic pressure difference ΔP (e.g., in atm) less any opposing osmotic pressure difference $\Delta\pi$ (in atm) that exists across the membrane by virtue of different solute concentrations in the feed or retentate relative to the permeate. The volumetric flux J_v across the membrane (expressed in units of cm/s) is the volume of permeated fluid per unit of time and membrane area; it is given by the expression

$$J_v = P \cdot (\Delta P - \Delta\pi) / \delta \qquad \text{Eq. 4}$$

where P is the membrane permeation coefficient or permeability (in cm^2/s-atm)and δ (cm) is the effective thickness of the membrane.

UF membranes have nominal pore sizes ranging from about 1 nm on the low end to about 0.02 μm to at most 0.1 μm (i.e., 20 to 100 nm) on the high end, so solutes with molecular weights of several hundred or less can readily flow convectively through the pores under an applied pressure differences; species larger than the nominal molecular-weight cut-off (MWCO) are rejected and retained to a greater or lesser extent. The extent to which a given UF membrane is effective in retaining a particular solute species "i" can be expressed in terms of a rejection coefficient R_i defined as

IDT Exhibit 1020.6035

$$R_i = 1 - (C_p / C_b) \qquad \text{Eq. 5}$$

where C_p is the solute concentration in the permeate and C_b is the solute concentration in the bulk solution (i.e., the feed or retentate in a separation application). (Equation 5 applies, strictly speaking, only in instances where boundary layer resistances are negligible and concentration polarization is insignificant. These considerations are discussed further below.) Low-molecular-weight solutes exhibit rejection coefficients close to zero, while macromolecular solutes with molecular weights well above the MWCO exhibit rejection coefficients approaching one. These concepts, normally applied to the ultrafiltrative separation and concentration of macromolecular solutes, are equally applicable to description of the structure and function of the porous membrane element in the contained membrane reactor arrays of the present invention.

When pressure-driven flux occurs through an ultrafiltration membrane, rejected macrosolutes accumulate at the high-pressure interface – normally at the high-pressure side of the rejecting skin layer of the UF membrane. As solutes accumulate, they are concentrated – not only within the bulk fluid but also within the thin fluid boundary later that normally resides at the high-pressure-side of the ultrafilter. The latter phenomenon is termed concentration polarization, and it is usually troublesome in conventional separation applications because it can reduce transmembrane flux (Lonsdale, 1982; Mulder, 1995; Cussler, 1997). However, in the CMRA applications of interest here, where the porous ultrafiltration element resides beneath a microchannel/microvessel element, concentration polarization can be a desirable phenomenon inasmuch as it provides a means for concentrating and effectively immobilizing high-molecular-weight reagents within the microchannels or microvessels of the CMRA. In effect, the microchannels or microvessels of a CMRA can be considered equivalent, structurally and functionally, to the stagnant film or fluid boundary layer that more typically resides atop the rejecting surface of an ultrafiltration membrane used to effect a separation.

The degree to which solutes are concentrated at the high-pressure interface (i.e., atop the permselective barrier that constitutes the skin region of an ultrafiltration membrane or functionally similar structure) can be estimated mathematically by considering the transport phenomena that give rise to concentration polarization. In ultrafiltration, solutes are continuously being carried to the surface of the membrane by the convective flow normal to the plane of the membrane, with solvent (typically, water) readily permeating the membrane. Those low-molecular-weight solutes that are not appreciably rejected (i.e., that have small R_i values) readily permeate the membrane as

IDT Exhibit 1020.6036

well and experience little concentration at the interface. However, those solutes that are highly rejected (i.e., with R_i values approaching unity) are blocked by the membrane and tend to diffuse away from its surface, back toward the main body or bulk of the fluid. Eventually a steady-state condition is reached, at which point the flow of solute towards the surface by convection is precisely balanced by the back-diffusion of solute to the bulk (and by solute leakage through the membrane to the extent that rejection is incomplete and $R_i < 1$). This balance is only established after the solute concentration gradient above the membrane (i.e., the driving force for diffusion) has become sufficiently steep. By writing the differential equation that describes this balance between convection and diffusion (and, in some cases, permeation) and then solving it subject to appropriate boundary conditions, an expression can be obtained for the ratio of the solute concentration at the membrane interface C_m relative to its concentration C_b in the bulk. For the particular case of a completely rejected solute (i.e., $R_i = 1$; $C_p = 0$) with diffusivity D, this expression takes the form:

$$C_m / C_b = exp\,[\,(J_v \cdot \delta) / D\,] \qquad \text{Eq. 6}$$

The grouping within the square brackets in Equation 6 will be recognized as the Peclet number Pe, i.e., the ratio of convective to diffusive fluxes, as discussed in considerable detail above. Substitution in Equation 6 leads to the simplified expression

$$C_m / C_b = exp\,(Pe) \qquad \text{Eq. 7}$$

In a typical separation application involving UF, δ generally refers to the effective thickness of the stagnant film or fluid boundary layer in contact with the surface of the membrane. In the present context, however, δ represents the effective thickness of the microchannel/microvessel element of the CMRA; that is, δ may be viewed either as the height of the microchannels or, alternatively, as the depth of the microwells in a confined membrane reactor array. Appropriate and straightforward modifications to the equations can be made where it is necessary to take into account the tortuosity and/or void volume of any structure (e.g., CMRA microchannels) that may reside atop the membrane surface. Inspection of the form of Equation 6 shows that the solute concentration within the microchannels or microwells of a polarized CMRA structure increases exponentially with distance as the surface of the porous membrane is approached.

The factor C_m / C_b is termed the "concentration polarization modulus," and it is readily calculated from the Peclet number Pe that quantifies the relative rates of the competing convective and diffusive transport processes. Typical values of the

IDT Exhibit 1020.6037

concentration polarization modulus are given in the following table for the limiting case of a completely rejected solute (i.e., $R_i = 1$):

Microchannel Peclet Number *Pe*	Concentration Polarization Modulus (C_m / C_b)
0.10	1.11
0.20	1.22
0.50	1.65
1.0	2.72
1.5	4.48
2.0	7.39
3.0	20.1
4.0	54.6
5.0	148.
10.0	22,030

For the sake of definiteness, consider a small, completely rejected protein (e.g., ribonuclease) with a diffusion coefficient D of about $1 \cdot 10^{-6}$ cm^2/s being swept by convection into a CMRA microchannel or microwell with a length or depth δ of 10 μm at a microchannel flow velocity (alternately, V or J_v) of 0.004 cm/s. The calculated Peclet number Pe corresponding to this situation is 4, and the resulting concentration polarization modulus C_m / C_b is estimated to be of order 55 – that is, the solute concentration at the membrane surface (i.e., at the bottom of the CMRA microchannel or well) will be 55 times higher than that in the bulk fluid at the top or mouth of the microchannel or well.

As noted previously, the solute concentration will vary exponentially with distance as measured from the top of a microchannel or mouth of a microwell, with the steepest gradient occurring at the base of the channel or well. More particularly, the concentration C_x of a polarized macrosolute at any point x along the length of a CMRA microchannel or microwell can be obtained by substituting the positional value x for the depth or thickness parameter δ in Equation 6:

$$C_x / C_b = exp\,[\,(J_v \cdot x) / D\,] \qquad \text{Eq. 8}$$

For the particular example of the small protein considered in the immediately preceding paragraph, the following local concentration factors C_x/C_b are obtained as a function of distance x from the top or mouth of the microchannel or well:

IDT Exhibit 1020.6038

Depth x (μm)	C_x/C_b
0.0	1.0 (at top or mouth of microvessel)
2.0	2.2
4.0	5.0
6.0	12.2
8.0	24.5
10.0	54.6 (at base of well or microchannel)

The average concentration of a concentration-polarized solute within a CMRA microchannel or microwell is readily calculated by integrating Equation 8 with respect to the position coordinate x over the interval from $x = 0$ to $x = \delta$ or L.

Lower-molecular-weight solutes that are incompletely rejected by the ultrafiltration membrane (i.e., $R_i < 1$) will also experience concentration polarization, albeit to a lesser extent. The concentration polarization moduli of such solutes will be smaller, as a consequence of their permeation or "leakage" across the ultrafilter. Although the mathematical equations that describe this situation are more tedious, they are known in the art and straightforward to solve and use.

Thus, while concentration polarization is normally considered a problem during conventional ultrafiltration because the increased concentration of retained molecules at the membrane surface increases the resistance to flow through it, the phenomenon is advantageous in the CMRA, where it is used specifically to create an elevated concentration of high-molecular-weight reagents inside the microchannels or microvessels. The increased concentration of macromolecules is then maintained by continued flow through the microchannels/microvessels and across the ultrafiltration membrane.

The maximum concentration of a molecule that is attainable in a CMRA – by concentration polarization or other means -- is set by the solubility limit for that molecule. This limit is generally met at lower concentrations, the larger the molecule at hand (at least when solubility is expressed on a molar basis). When the solubility limit is exceeded, molecules – and especially high-molecular-weight biomolecules -- will tend to come out of solution in the form of aggregates or gels, and where concentration polarization is the means by which high local concentrations are obtained, molecules will be deposited adjacent to the surface of the ultrafiltration membrane at a concentration corresponding to their solubility limit or gel concentration C_g. Formation of a gel is by no

IDT Exhibit 1020.6039

means required in the operation of a CMRA, but the process can be used to advantage, especially when particularly high local concentrations of molecules are desired. For many macromolecular solutions, gel concentrations C_g average around 25 wt % (with a range of from about 5 % to 50 %), whereas colloidal dispersions are characterized by C_g values that average about 65 % (with a range from 50 % to 75 %). Once a gel layer forms atop an ultrafiltration membrane, the hydraulic permeability of the gel layer itself (rather than the intrinsic hydraulic permeability P of the UF membrane) can control the transmembrane flux J_v.

Manipulation of the concentration of reaction participants by concentration polarization within a CMRA can be used to alter the phase or physical state of molecules in the system to advantage. For instance, if molecules in the gel state remain active (*e.g.*, if an enzyme retains its bioactivity when precipitated in the form of a gel), then gel formation provides a means for obtaining very high local concentrations of that molecule within the microvessels or microwells of a CMRA. Furthermore, molecules that have precipitated into a gel are less subject to diffusional motion; indeed, the processes of gel relaxation and resolubilization can be quite slow – even irreversible -- under certain circumstances. Thus, once macromolecules have been deposited in the form of gel layers by the application of convective flows sufficiently large as to cause severe concentration polarization and local concentrations that exceed the gel point concentration C_g, it is reasonable to expect that such molecules will tend to remain in the microchannels or microwells of a CMRA even when the convective flow rate through the microchannels or microwells is substantially reduced or even stopped. It may further be noted that molecules that form macromolecular complexes (*e.g.* multi-subunit proteins and certain polymers) can be added in bulk solution at concentrations that are too low for molecular association. Subsequently, however, they can be concentrated by the method of the present invention in the CMRA microchannels such that, *e.g.*, polymers polymerize or multi-subunit proteins assemble.

Macromolecules can thus be maintained at elevated concentrations inside CMRA microchannels without having to attach said macromolecules to the walls of the microchannel or to some other solid-phase support; that is, the macromolecules remain localized in the solution phase (or perhaps the gel phase) without the need for covalent attachment to a solid-phase support. This is advantageous because many enzymes lose activity or exhibit decreased activity when covalently bound or otherwise associated with a surface (Bickerstaff, 1997). An additional advantage is that the macromolecule

IDT Exhibit 1020.6040

"localization" (*cf.* immobilization) method of the present invention is generic -- *i.e.*, it functions in substantially the same manner for all macrosolutes -- rather than being macromolecule-specific, a drawback of many covalent immobilization protocols.

In certain reaction systems of interest (e.g., DNA analysis by pyrosequencing, as discussed in more detail below), it may be necessary to avoid covalently immobilizing certain macromolecular reagents altogether. The DNA polymerase used in pyrosequencing is a case in point. It is believed that DNA polymerase should retain at least a certain degree of mobility if it is to function optimally. As a consequence, this particular enzyme must normally be treated as a consumable reagent in pyrosequencing, since it is not desirable to covalently immobilize it and reuse it in subsequent pyrosequencing steps. However, the present invention provides means for localizing this macromolecular reagent within the microchannels or microvessels of a CMRA without having to covalently immobilize it.

Macromolecules (e.g., enzymes) can be added to CMRA microchannels or microvessels in a sequential manner, thereby creating a microscopic stacked column (see Fig. 3). This permits sequential processing of reactants/substrates and their products in the channel, with products produced upstream being made available as reactants for downstream processing steps.

Once macromolecules have been concentrated and deposited in the microchannels or microwells of a CMRA by methods disclosed above, other reaction participants (*e.g.* reactants) can be added. If the molecules are small enough to pass through the ultrafiltration membrane substantially unimpeded, then their local concentration within the microchannels or microwells will be unaltered by the filtration process. Considering for the moment only the two limiting cases of $R_i = 1$ and $R_i = 0$, it is seen that molecules that are swept into the microchannels or microwells of a CMRA will experience either one of two fates: either a molecule will be concentrated and localized by ultrafiltration, or it will pass through CMRA and emerge in the permeate or "ultrafiltrate". To simplify later descriptions, in what follows the word "pack" means to concentrate a molecule into the CMRA by concentration polarization, while the word "flow" means to generate bulk flow through the microchannel, carrying molecules into the CMRA and out in the ultrafiltrate without appreciable concentration within the CMRA.

As an aside, it should be noted that if the porous filter of the CMRA is capable of capturing particles in suspension (even if it is incapable of rejecting macromolecules in solution), then in a similar manner such particles will simply stack up next to the

IDT Exhibit 1020.6041

membrane in the form of a filter cake. In applications where certain of the reaction participants are immobilized on beads or other particulate supports, membranes other than ultrafilters – e.g., various microfiltration (MF) membranes known in the art -- may be suitable for the practice of the present invention (Eykamp, 1995). The fundamental requirement is that the effective pore size of the membrane be comparable to or smaller than the diameter of the particles that one desires to retain.

Clearly, molecules entrapped and concentrated in the microchannels or microwells of a CMRA may still undergo diffusional motion to a certain extent. While molecules that have precipitated into a gel layer will exhibit decreased mobility – and perhaps substantially decreased mobility -- it is still possible for them to return to solution and thus become free to diffuse, in the event that their concentration falls below the solubility or gel limit C_g. Referring to the stacked column configuration described above, it is evident that the order of the stack may be lost over time if the loaded macromolecules are subsequently able to diffuse at significant rates. To prevent this disordering, molecules can be tied together with polymers; for example, biotinylated macromolecules can be tied together with streptavidin-conjugated linear dextrans (*e.g.*, 2M Dalton linear dextran/streptavidin conjugate, product number F071100-1, Amdex A/S, Denmark) or with one of any number of chemical crosslinkers. The crosslinker can be added after the proteins have been deposited within the CMRA in order to prevent premature crosslinking and aggregation during the loading step. Alternatively, biotinylated molecules and biotinylated linear dextran can be added concurrently, and then tied together by the subsequent addition of avidin (molecular weight ~60 kD). Similarly, photoreactive crosslinkers can be added and then activated with light after other macromolecular species have been loaded into the microchannels or microwells of the CMRA.

As noted above, small molecules that would not typically be retained by ultrafilters can be attached to larger molecules (*e.g.* dextran or proteins such as albumin) or even to particles (*e.g.* polystyrene beads or colloidal gold, including porous beads such as those manufactured by Dynal, Inc.) to enhance their utility in connection with the present invention. Colloidal particles and microparticles will diffuse at much lower if not negligible rates as compared to microsolutes and macrosolutes. By attaching small molecules that would otherwise pass through the CMRA to larger macromolecules or particles, these smaller molecules can be retained in the microchannels or microwells of a CMRA.

IDT Exhibit 1020.6042

It should further be noted that the present invention permits one to selectively manipulate the degree to which different reaction participants experience concentration polarization within the microchannels or microwells of a CMRA. Smaller molecules have larger diffusion coefficients, and so they will be characterized by smaller Peclet numbers; thus, the relative significance of the diffusional component of transport of these smaller molecules vis-à-vis the convective component will be greater for smaller molecules than it will be for larger ones. This provides a degree of freedom in the design and operation of these systems.

For example, the convective flow rate can be modulated to manipulate the Peclet number of various species selectively. At any flow rate V or J_v across the CMRA, the smallest species ($R_i = 0$) will not be rejected at all by the porous membrane filter and will experience no concentration polarization. However, at a given flowrate small macromolecules with intermediate diffusivities may experience "intermediate" degrees of concentration polarization, while large macromolecules with small diffusivities will encounter "strong" polarization. If, then, one increases the flow rate and hence the Peclet number Pe for the smaller of two macromolecules, the extent of polarization of this molecule may be increased from "intermediate" to "strong." If instead the flow rate and Pe are decreased, the degree of polarization of this smaller macromolecule may be reduced from "intermediate" to "low." It should be noted, however, that at all three flow rate conditions, the smallest, unrejected solute will experience no polarization, while the larger of the two macromolecules will be strongly polarized.

As an example, reference to the above table shows that flow conditions could readily be chosen such that a small macromolecule characterized by a Peclet number Pe of 0.5 might exhibit a small concentration polarization modulus C_m / C_b of 1.65 at a first flow rate. However, increasing the flow rate 10-fold would result in a 10-fold increase in the Peclet number Pe to 5.0 – and a substantial, 90-fold increase in the polarization modulus C_m / C_b to a relatively large value of 148. All this time, however, a much larger macromolecule that was strongly polarized to begin with at the lower flowrate (and perhaps even deposited as a gel) would remain strongly polarized at the higher flowrate.

Alternatively, flow could be slowed to such a degree that smaller molecules with larger diffusion coefficients (but with R_i values greater than zero) would be permitted to diffuse throughout the microreactor volume – in the extreme, small molecules might even be permitted to diffuse upstream and out of the microreactor -- while at the same time larger molecules with smaller diffusivities would experience significant concentration

IDT Exhibit 1020.6043

polarization. Flow speed could then be increased to restore concentration polarization for both species, while smaller molecules, unrejected or poorly rejected by the porous filter or membrane, were swept past larger molecules. At the extreme, convective flow could be stopped altogether, causing all molecular transport to occur by diffusion.

Manipulation of the Peclet number in this manner thus permits sequential processing steps – e.g., macromolecule packing, reactants supply, chemical conversion, and product removal in the ultrafiltrate -- to be conducted in a highly controlled and advantageous manner. These steps can be performed in a constant-flow-rate system, or each step can be performed with different and time-varying flow speeds.

Controlling the pressure difference across the CMRA can, in principle, pose some minor problems if the fluid enters the flow compartment atop the CMRA via a plenum to one side of it. In this situation, there will be a pressure drop along the length of the fluid compartment atop and parallel to the CMRA due to viscous nature of the flows both parallel to and through the substantially 2-D confined membrane reactor array (Fig. 4). This pressure variation along the CMRA may, in extreme cases, cause the pressure difference across the CMRA (*i.e.* the pressure difference driving flow through it) to vary somewhat with distance, with the highest trans-CMRA pressure drop existing near the entrance plenum and the lowest pressure drop prevailing at the opposite end of the flow compartment. This effect can be reduced by introducing fluid via multiple inlets located along the sides or the circumference of the CMRA (see, for example, Fig. 5, where optional means to permit excess fluid to be withdrawn through a hole at the center of the CMRA may be provided). Alternatively, a back-pressure regulator, valve, or other flow restrictor may be introduced in the flow stream downstream of the CMRA, said regulator or valve introducing a controlling pressure drop that is arranged to be large as compared to the pressure drop within the CMRA flow compartment. In this manner, the fluid pressure within said flow compartment is increased -- and the end-to-end variation in pressure drop across the CMRA itself is minimized (Fig. 6). A fluid recirculation loop may optionally be provided. By these and other means, the potential variation in pressure drop parallel to the plane of the CMRA can be arranged to be only a small fraction of the pressure drop normal to and across the CMRA.

In a preferred embodiment of a CMRA, the microreactor element comprising an array of microchannels or microwells is defined as a fiberoptic reactor array plate similar to that described in US patent 6,274,320. In that patent the fiber optic reactor array is formed by etching one end a fused fiber optic bundle to forms wells. In the context of

IDT Exhibit 1020.6044