ligation process may incorporate exogenous linking segments, e.g., not part of the original target sequence, while in other instances; portions of the original target nucleic acid may be used to form the linking oligonucleotides. In the case of internal sequences used as the linking oligonucleotides, such sequences may derive from single stranded overhang sequences, or may be derived from a double stranded portion of a blunt ended fragment.

[00108] In either event, covalent linkage of adjacent 3' and 5' ends of a double stranded nucleic acid segment, whether from the original target segment, or as a result of an appended or ligated exogenous linking oligonucleotide sequence, may be carried out using, e.g., a template independent, dsDNA end (TIDE) ligation process, using, for example a circligase enzyme system. Typically, this process will require the presence of a phosphate group at the 5' end of each segment in order to permit circligase action. Addition of the 5' phosphate may be accomplished upon the fragment enzymatically, e.g., using a T4 polynucleotide kinase, or the like. Alternatively, where the double stranded segment was synthesized, or created from another template, rather than from the fragmenting of a larger nucleic acid, the phosphorylated 5' terminus could be provided during the synthesis process, e.g., on a primer sequence used for amplification of the original target sequence, as an initial building block in solid phase synthesis, or the like.

[00109] Figure 8 schematically illustrates an exemplary process for coupling the 3' and 5' ends of a double stranded fragment. As shown, a double stranded fragment 800 of a target sequence is provided, that is comprised of strands 802 and 804. Overhang sequences 806 and 808 are provided on the double stranded segment 802 and 804, respectively. Such overhang sequences may be added to the fragments using by a variety of methods, e.g., using standard tailing techniques, such as treatment with terminal transferases to add poly-A tails, ligating adapter sequences to the target fragments that contain such overhang sequences, or the like. Also, although shown as sequences added to the double stranded fragment, it will be appreciated that such overhang sequences may be provided during the fragmentation process, e.g., as overhang sequences from, e.g., restriction endonuclease digestion of larger nucleic acids.

[00110] As shown, a 5' phosphate group 810 is coupled to each strand to permit TIDE ligation and closure of the two adjacent ends. Upon treatment with a ligase having appropriate closure activity , e.g., Circligase ssDNA ligase (Epicentre Biotechnologies, Madison WI), T4 RNA ligase, or the like each end of the double stranded target is closed to provide a completely contiguous template sequence 812 of the invention.

IDT Exhibit 1020.1185

[00111] Demonstration of the joining of a 5' phosphated nucleotide to a 3' hydroxyl of double stranded nucleic acid fragments employed a commercial Circligase enzyme system, but with additional modifications to the protocol (addition of 5'phosphate, presence of $MnCl_2$, ATP and a reaction temperature of 60°C for greater than 1 hour). The resulting molecule was resistant to exonuclease digestion (by both exonuclease I and exonuclease III) as monitored by PAGE, indicating that the resulting molecule was closed on both ends.

[00112] An alternative process for providing overhang sequences for the foregoing process employs blocked primer sets in an amplification process to generate double stranded nucleic acid segments that retain overhang sequences. In particular, amplification primer pairs are provided to amplify a segment of double stranded DNA, e.g., they prime opposing ends of complementary strands of the targeted segment for antiparallel amplification. The primer pairs are configured to be partially complementary to the target segment, and have within their sequence, one or more non-native nucleotides (referred to herein as a "5th base"). Inclusion of the fifth base within the primer sequence, for which no complements are provided in the amplification mix, will prevent the target sequence from extending the target along the primer sequence, and thus, retain the single stranded overhang sequence in the resulting double stranded product. Likewise, repeated cycles of amplification will result in the vast majority, and approaching substantially all, of the amplification product having the overhang sequence retained on both strands of the double stranded product. These double stranded segments may then be used in the template generation processes described herein.

[00113] A schematic illustration of an example of the foregoing process is provided in Figure 9. As shown, a double stranded target nucleic acid segment 900, is anti-parallel primed using primers 902 and 904 (Panel I). As shown, each primer includes a first portion 906, that is complementary to its strand of the target sequence 900, a second portion 908 that includes one or more non-native nucleotides or 5th bases, and a third portion 910, which, as shown, is non-complementary to the target segment 900. Although shown as being non-complementary, this is not required for operability of the process. In some cases, for example, use of complementary third portions 910, may provide higher affinity of the primers to the desired target segments by virtue of the primer having two segments that are complementary to the desired segment, and less likelihood of binding to non target regions, where the 5th base portions do not excessively interfere with hybridization of both the first and third portions to the target segment. The primers are extended in a standard amplification process, e.g., PCR, but in the absence of the complement of the 5th base.

IDT Exhibit 1020.1186

[00114] As shown by the arrows in panel II, primer extension against each amplification product will terminate at the same position, i.e., the position complementary to the 5th base in the complementary strand. Following multiple rounds of amplification (Panel III), the amplification product will be substantially made up of complementary strands having overhang sequences that contain the 5th base containing portion (second portion 908) and the third portion of the primer (910), that can be annealed to provide double stranded nucleic acids 912.

[00115] The double stranded nucleic acids 912, having overhang sequences on the 5' end of each segment, are then subjected to the ligation processes described above (and illustrated in Panel IV), to provide contiguous templates 914 of the invention.

[00116] As will be appreciated, the primer sequences may be separately synthesized and configured to include those functional groups necessary and/or desirable for the ligation process. For example, such primers may be synthesized to include the 5' phosphate group used in a TIDE ligation process. Additionally, they may be synthesized to include, e.g., within third portion 910, a sequencing priming site, e.g., different from the amplification priming site in portion 906, or other functional sequence, as set forth elsewhere herein. In addition, the presence of the 5th base portion, e.g., one or more non-native bases in the linking oligonucleotide portion of a resulting contiguous template construct, can provide yet another indicator and/or control sequence or sequence event outside of the double stranded segment of the target sequence. Alternatively, the region 910 could also be partially self-complimentary to form a stem-loop structure, placing the 5' end adjacent to the 3'-end of the blocked extension. This could then potentially be used as a substrate for a more standard T4 DNA ligase mediated method, e.g., as described above.

[00117] By way of example, a sequencing reaction may be initiated in the absence of the complement to the 5th base. Because this is a non-native base, it's absence will not impact the overall sequence determination of the target portion for the sequence. However, by starving the reaction for this complement, one can prohibit synthesis, and thus, the sequencing process, until the 5th base complement is added to the mixture, providing a hot start capability for the system. Additionally, as a non-native base, this portion of the overall template construct provides an internal check on sequencing process and progress that is configurable to not interfere with sequence analysis of the native bases in the template. For example, the 5th base complement in the sequence mixture may be provided with a wholly detectably different label than the complements to the four native bases in the sequence. The production of incorporation based signals associated with such labels then provides an indication that the process is about to start processing one strand of the

IDT Exhibit 1020.1187

target nucleic acid. Likewise, it can provide a clocking function for the number of times the process has proceeded around a completely contiguous template. Although described as the "5th base" it will be appreciated that this may comprise a set of non-natural bases that can provide multiple control elements within the template structure. For example, two different non-native or 5th bases could be included within the template structure, but at different points, to regulate procession of the sequencing process, e.g., allowing controlled initiation, and a controlled stop/start, later in the sequence, e.g., prior to sequencing the antisense strand. For example, one could add the complement to the first non-native base in order to initiate sequencing. Upon encountering the second non-native base, e.g., at the first hairpin turn, sequencing would stop in all reactions, until the complement to that second base was added to the reaction mixture. This would allow a resynchronization of the various sequencing reactions, and or an ability to control sequencing the opposing strand, providing a paired end sequencing configuration as discussed elsewhere herein.

[00118] A similar or related process for preparing either partially or completely contiguous template constructs is illustrated in Figure 10. In particular, as shown, a first amplification primer sequence 1000 is provided that includes first and second complementary segments 1002 and 1004, respectively, linked by a linking oligonucleotide 1006, e.g., as described elsewhere herein. In addition, a single stranded target priming segment 1008 is provided at the 3' end of the overall amplification primer. In some cases, the target priming segment 1008 may be specifically selected to prime adjacent to or within desired sequence locations. In other preferred aspects, the target priming segment will randomly prime within a given genome or other large DNA sequence to ensure optimal coverage in generating template libraries for sequencing. By way of example, in the case of random priming, the target priming segment 1008 may be comprised of relatively small oligonucleotides, e.g., hexamers, heptamers, octamers, or the like. For more specific priming, the target priming segment will typically comprise larger segments on the order of 16, 20 or more nucleotides within the target priming segment. Additionally, such segments will typically comprise a sequence complementary to a known sequence segment adjacent to or otherwise proximal to a desired target sequence.

[00119] As shown in Figure 10, the first amplification primer 1000 is denatured and allowed to hybridize to a target nucleic acid segment 1010. In some cases, the primer 1000 is configured to be denatured under conditions which still allow it to anneal to the target sequence 1010. In other cases, the structure of the primer will permit hybridization and priming of the target segment 1010, even when the primer is in its hairpin structure, e.g., without denaturation.

IDT Exhibit 1020.1188

[00120] By way of example, and as shown, at step A, the reaction mixture is heated to an appropriate denaturing temperature for the first amplification primer 1000, e.g., 37°C, and the primer is allowed to anneal to the target segment 1010. As shown, isothermal amplification of the target segment 1010 is then carried out (Steps A and B) to generate further amplifiable target segments having the hairpin structure of the first amplification primer appended to each end (Segment 1012 in Step C). Note that although illustrated as a single line in Figure 10, it will be appreciated that for purposes of discussion, the single illustrated line illustrating the target segment 1010 and amplifiable target segments 1012 represents either one or both of the sense and antisense strands of a complementary nucleic acid.

[00121] This segment is then subjected to geometric amplification, e.g., PCR, using second amplification primers 1014 against the initial amplification primer sequence 1000, e.g., complementary to one or more of segments 1002, 1004, 1006 and even 1008, or their complements, to yield amplification products, e.g., complementary template segments 1016 and 1018. Following the amplification of segment 1012 (Step E), renaturation of the original first amplification primer segments, or the partially overlapping isothermal or PCR amplification primer segments or their complements within the amplification product, e.g., 1016, results in the formation of hairpin structures at each end of the amplification products (Step E) to form partially double stranded partially contiguous nucleic acid segments 1020. These partially double stranded segments 1020 are then converted (Step F and G) to completely contiguous segments 1022, by subjecting the self priming partially double stranded segments to 3' extension, e.g., using non-strand displacing nucleic acid polymerases, e.g., Klenow fragment, followed by ligase treatment to couple the resulting 3' terminus to the 5'end. Following ligation, the amplification mixture is then subjected to exonuclease digestion to remove any nucleic acid segments that are not fully contiguous, e.g., were either not ligated or not fully extended.

[00122] Although the constructs of the invention are described primarily, and preferably, for use directly as templates for, e.g., sequencing applications, it will be appreciated that these structures may also serve as intermediate structures in the preparation of templates that provide for sequence redundancy in line with that provided by such constructs. For example, the structurally circular nucleic acid segments described herein, may be used as templates in a rolling circle replication process to produce concatamer molecules that include repeating copies of both the sense and antisense strands of the originating double stranded segment included within the circular nucleic acid. These replicated products may then be employed directly as template molecules in a

IDT Exhibit 1020.1189

template dependent sequencing process, as described elsewhere herein (See also U.S. Patent No. 7,476,503, which is incorporated herein by reference in its entirety for all purposes). Likewise, duplication processes may be employed to produce multiple copies of a prepared circular construct using methods previously described (See, e.g., U.S. Patent Application No. 61/072,160, previously incorporated herein by reference in its entirety for all purposes).

[00123] Generation of the double stranded nucleic acid segments that are used in the preparation of the template constructs of the invention may be accomplished by a number of means. For example, nucleic acids derived from samples to be analyzed may be fragmented into double stranded fragments through known fragmentation methods, e.g., as noted below. Alternatively, double stranded templates may be generated from targeted regions of the sample nucleic acid segments through bidirectional amplification of desired sequence segments within a larger sequence. In particular, one can employ sequence specific PCR primers on flanking portions of a desired sequence region to amplify the region bounded by the primers through antiparallel amplification, either alone, or in conjunction with an initial linear amplification process. The resulting amplified product includes the double stranded sequence region of interest, which is then subjected to additional processing to yield the template constructs of the invention.

[00124] As described above, the template nucleic acids of the invention that are provided by the methods described herein, e.g., for use in single molecule sequencing reactions, can be derived from a genomic DNA. Genomic DNA can be prepared from any source by three steps: cell lysis, deproteinization and recovery of DNA. These steps are adapted to the demands of the application, the requested yield, purity and molecular weight of the DNA, and the amount and history of the source. Further details regarding the isolation of genomic DNA can be found in Berger and Kimmel, Guide to Molecular Cloning Techniques, Methods in Enzymology volume 152 Academic Press, Inc., San Diego, CA (Berger); Sambrook et al., Molecular Cloning - A Laboratory Manual (3rd Ed.), Vol. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 2008 ("Sambrook"); Current Protocols in Molecular Biology, F.M. Ausubel et al., eds., Current Protocols, a joint venture between Greene Publishing Associates, Inc. and John Wiley & Sons, Inc ("Ausubel"); Kaufman et al. (2003) Handbook of Molecular and Cellular Methods in Biology and Medicine Second Edition Ceske (ed) CRC Press (Kaufman); and The Nucleic Acid Protocols Handbook Ralph Rapley (ed) (2000) Cold Spring Harbor, Humana Press Inc (Rapley). In addition, many kits are commercially available for the purification of genomic DNA from cells, including Wizard™ Genomic DNA Purification Kit, available from Promega; Aqua Pure™ Genomic DNA

Isolation Kit, available from BioRad; Easy-DNA™ Kit, available from Invitrogen; and DnEasy™ Tissue Kit, which is available from Qiagen. Alternatively, or additionally, target nucleoc acid segments may be obtained through targeted capture protocols where target nucleic acids are obtained initially as single stranded segments on microarrays or other capture techniques, followed by amplification of the captured material to generate double stranded sample materials. A variety of such capture protocols have been described in, e.g., Hodges E, *et al.* Nat. Genet. 2007 Nov 4, Olson M., Nature Methods 2007 Nov;4(11):891-2, Albert TJ, *et al.* Nature Methods 2007 Nov;4(11):903-5, and Okou DT, *et al.* Nature Methods 2007 Nov;4(11):907-9.

[00125] The nucleic acids that can be prepared by the methods described herein, e.g., for use with high-throughput sequencing systems can also be derived from a cDNA, e.g. cDNAs prepared from mRNA obtained from, e.g., a eukaryotic subject or a specific tissue derived from a eukaryotic subject. Data obtained from sequencing the nucleic acid templates derived from a cDNA library, e.g., using a high-throughput sequencing system, can be useful in identifying, e.g., novel splice variants of a gene of interest or in comparing the differential expression of, e.g., splice isoforms of a gene of interest, e.g., between different tissue types, between different treatments to the same tissue type or between different developmental stages of the same tissue type.

[00126] mRNA can typically be isolated from almost any source using protocols and methods described in, e.g., Sambrook and Ausubel. The yield and quality of the isolated mRNA can depend on, e.g., how a tissue is stored prior to RNA extraction, the means by which the tissue is disrupted during RNA extraction, or on the type of tissue from which the RNA is extracted. RNA isolation protocols can be optimized accordingly. Many mRNA isolation kits are commercially available, e.g., the mRNA-ONLY™ Prokaryotic mRNA Isolation Kit and the mRNA-ONLY™ Eukaryotic mRNA Isolation Kit (Epicentre Biotechnologies), the FastTrack 2.0 mRNA Isolation Kit (Invitrogen), and the Easy-mRNA Kit (BioChain). In addition, mRNA from various sources, e.g., bovine, mouse, and human, and tissues, e.g. brain, blood, and heart, is commercially available from, e.g., BioChain (Hayward, CA), Ambion (Austin, TX), and Clontech (Mountainview, CA).

[00127] Once the purified mRNA is recovered, reverse transcriptase is used to generate cDNAs from the mRNA templates. Methods and protocols for the production of cDNA from mRNAs, e.g., harvested from prokaryotes as well as eukaryotes, are elaborated in *cDNA Library Protocols*, I. G. Cowell, et al., eds., Humana Press, New Jersey, 1997, Sambrook and Ausubel. In addition, many kits are commercially available for the preparation of cDNA, including the Cells-to-cDNA™ II Kit (Ambion), the RETROscript™ Kit (Ambion), the CloneMiner™ cDNA Library

IDT Exhibit 1020.1191

Construction Kit (Invitrogen), and the Universal RiboClone® cDNA Synthesis System (Promega). Many companies, e.g., Agencourt Bioscience and Clontech, offer cDNA synthesis services.

[00128] In some embodiments of the invention described herein, nucleic acid fragments are generated from a genomic DNA or a cDNA. There exist a plethora of ways of generating nucleic acid fragments from a genomic DNA, a cDNA, or a DNA concatamer. These include, but are not limited to, mechanical methods, such as sonication, mechanical shearing, nebulization, hydroshearing, and the like; chemical methods, such as treatment with hydroxyl radicals, Cu(II):thiol combinations, diazonium salts, and the like; enzymatic methods, such as exonuclease digestion, restriction endonuclease digestion, and the like; and electrochemical cleavage. These methods are further explicated in Sambrook and Ausubel.

VIII. Kits and Systems

[00129] In addition to the template compositions described above, and methods of making and using such compositions, the present invention also provides applied embodiments of such methods and compositions.

[00130] For example, in certain embodiments, the present invention provides kits that are used in preparation and use of the template constructs of the invention. A first exemplary kit provides the materials and methods for preparation of the template constructs in accordance with the invention, as described elsewhere herein. As such, the kit will typically include those materials that are required to prepare template constructs as outlined herein, e.g., in accordance with the various template preparation processes outlined above. As will be appreciated, depending upon the nature of the template construct, and the method used, the kit contents can vary. For example, where one is employing hairpin adapters that are coupled to the ends of double stranded nucleic acid segments, the kits of the invention will typically include such hairpin adapters, along with appropriate ligation enzymes and protocols for attaching such adapters to the free ends of double stranded nucleic acids, as well as any processing enzymes that may be desirable for treating the ends of the double stranded segments prior to ligation, e.g., to provide overhangs or blunt end nucleic acids. As noted previously, such adapters may include overhang segments to facilitate coupling to complementary overhang ends on the double stranded segment. These overhang ends may be the result of a known added sequence to the double stranded segment, e.g., resulting from restriction enzyme cleavage, tailing processes, or the like, and the reagents for preparing templates

IDT Exhibit 1020.1192

having these characterisrtics may optionally be provided within the kits or they may be obtained from commercial sources.

[00131] In other embodiments, these kits may include primers or other sequence adapters or sequence extension reagents for providing overhang sequences that may be used to provide linking oligonucleotides between the two strands of the double stranded nucleic acid segment. In some cases, these kits may include enzyme systems for providing 5' phosphate groups to the linking oligonucleotides, or may provide amplification primers that have such 5' phosphate groups predisposed upon the primers. Such primers may additionally include the 5th base configurations set forth above, for controlling both the amplification process as well as the use of the resulting templates in sequencing applications.

[00132] A second exemplary kit provides materials and methods not just for the preparation of the template constructs of the invention, but also for the use of such templates in performing sequence analysis on target nucleic acid sequences. Thus, in addition to the materials and methods set forth above, such kits may additionally include reagents used in such sequencing processes, such as primer sequences for initiating the sequence process, polymerase enzymes, and in preferred cases, substrates that provide for optical confinement of nucleic acid synthesis complexes. In particularly preferred aspects, such substrates will typically include one or more arrays of zero mode waveguides. Such waveguide arrays may further include surface treatments that provide for enhanced localization of synthesis complexes within the illumination volumes of such zero mode waveguides, e.g., as described in Published International Patent Application No. WO 2007/123763, incorporated herein by reference in its entirety for all purposes. Additionally, such kits may optionally include nucleotide compositions for use in sequencing applications, including, for example labeled nucleotides that include fluorescent or otherwise detectable labeling groups coupled to the phosphate groups in a nucleoside polyphosphate construct at a phosphate group other than the alpha phosphate. A variety of other types of labeled and unlabeled nucleotides may be optionally includes within the kits and are generally known in the art.

[00133] The invention also provides systems that are used in conjunction with the template constructs of the invention in order to provide for analysis of target nucleic acid molecules. In particular, such systems typically include the reagent systems described herein, in conjunction with an analytical system for detecting sequence information from those reagent systems. For example, depending upon the nature of the sequencing process employed, the sequencing systems may include the system components provided with or sold for use with commercially available nucleic

IDT Exhibit 1020.1193

acid sequencing systems, such as the Genome Analyzer System available from Illumina, Inc., the GS FLX System, available from 454 Life Sciences, or the ABI 3730 System available from Life Technologies, Inc.

[00134] In preferred aspects, such systems include fluorescence microscopes capable of resolving fluorescent signals from individual sequencing complexes. In particularly preferred aspects, such systems include arrays of reaction regions, e.g, zero mode waveguide arrays, that are illuminated by the system, in order to detect fluorescent signals therefrom, that are in conjunction with sequencing reactions being carried out within each ZMW.

[00135] The systems of the invention also typically include information processors or computers operably coupled to the detection portions of the systems, in order to store the signal data generated by the system (e.g., the sequencing reactions incorporating labeled nucleotides which are illuminated in the system and thereby produce fluorescent signals indicative of such incorporation) obtained from the detector(s) on a computer readable medium, e.g., hard disk, CD, DVD or other optical medium, flash memory device, or the like. For purposes of this aspect of the invention, such operable connection provide for the electronic transfer of data from the detection system to the processor for subsequent analysis and conversion. Operable connections may be accomplished through any of a variety of well known computer networking or connecting methods, e.g., Firewire®, USB connections, wireless connections, WAN or LAN connections, or other connections that preferably include high data transfer rates. The computers also typically include software that analyzes the raw signal data, identifies signal pulses that are likely associated with incorporation events, and identifies bases incorporated during the sequencing reaction, in order to convert or transform the raw signal data into user interpretable sequence data (See, e.g., Published U.S. Patent Application No. 2009-0024331, the full disclosure of which is incorporated herein by reference in its entirety for all purposes).

[00136] Exemplary systems are described in detail in, e.g., U.S. Patent Application No. 11/901,273, filed September 14, 2007 and U.S. Patent Application No. 12/134,186, filed June 5, 2008, the full disclosures of which are incorporated herein by reference in their entirety for all purposes.

IX. Examples

[00137] The template constructs described herein were employed in various sequencing applications which are described in greater detail, below.

IDT Exhibit 1020.1194

Example 1: Template Construction and Sequencing

[00138] Template constructs in accordance with the invention were prepared for use in a sequencing by incorporation context. In particular, the template was subjected to single molecule real time (SMRT™) sequencing in which the polymerase mediated, template dependent extension of a primer sequence was monitored as labeled nucleotide analogs were being used for incorporation into the resulting nascent strand.

Template Construction:

[00139] A double-stranded fragment of DNA was amplified from a plasmid clone using primers 5' – GTACGGGTCTCACCCGGGGATCCTCTAGAATCGAT – 3' and 5' – CCTAAGGTCTCGGAAGCTACTAGTCCTCAGCAAGCTT – 3'. The resulting product was purified using a Zymo-25 PCR purification kit. Overhangs were generated on each end of the PCR product by incubating overnight in the presence of the restriction enzyme BsaI (NEB). The digested product was purified using a Qiagen PCR purification kit, and then ligated to the synthetic hairpin oligos: 5' – CGGGCTCGGAACGAAAGTTCCGAG – 3' and 5' – CTTCGGTCGCCAGATTAGAAAATCAGTCACGTCTAGATGCAGTCAGGTTCTTAAATCCT AGTTCCTTGGCGACC – 3'. Ligation was performed by incubating the digested PCR product with a 2-fold excess of each hairpin oligo in the presence of T4 DNA Ligase (NEB) at 23°C for one hour. Non-ligated products were removed by incubating the reaction in the presence of Exonuclease III for 1 hour at 37°C. The final product was purified using a Qiagen PCR Purification Kit, and annealed to an equimolar amount of sequencing primer: 5' – CTGACTGCATCTAGACGTGACTGA – 3'. The final template construct included a 244 nucleotide duplex segment and an overall length of 546 nucleotides (including linking/hairpin segments).

[00140] The SMRT™ sequencing reaction was carried out with 2 nM DNA polymerase; 100 nM template; 500 nM A647-dA6P, 500 nM A660-dC6P, 500 nM A568-dG6P, and 500 nM A555-dT6P; Trolox at 0.5 mM; PCA (protocatechuic acid) at 4 mM; and PCD (protocatechuate 3,4 dioxygenase) at 0.5X.

[00141] The sequencing reaction was carried out in a zero mode waveguide array having 3000 discrete cores of the ZMWs. The reaction was observed using a highly multiplexed confocal fluorescent microscope providing a targeted illumination profile, e.g., a separate spot for each core (See, e.g., U.S. Patent Application No. 12/151,979, filed May 9, 2008, and incorporated herein by reference in its entirety for all purposes). Fluorescent signals from the various ZMWs were

IDT Exhibit 1020.1195

detected on an EMCCD camera for 5 minutes, and were subjected to pulse recognition and base calling processes (See, e.g., Published U.S. Patent Application No. 2009-0024331, and incorporated herein by reference in its entirety for all purposes). Figure 11 shows a sequencing trace showing 700 bases of sequence data around the template construct providing 154 bases of consensus information.

Example 2: Single Molecular Consensus Sequencing

[00142] Single molecular consensus sequencing accuracy was assessed using the template constructs of the invention. A completely contiguous SMRTbell™ template was generated containing 162 nucleotides total. In generating the contiguous template, two hairpin sequences were used. The first contained a 54 nt loop closed with an 8 bp stem and a 4 nt overhang that is used to ligate to an overhang present in the insert oligonucleotides. The second consisted of a 4 nt loop closed with an 8 bp stem and a 4 nt overhang for ligation. When sequenced in a single molecule sequencing system, each strand of this template can be sequenced multiple times. The resulting template was subjected to SMRT™ Sequencing as described above, which iteratively sequenced the template in a single molecule configuration, multiple times. Individual iterative sequence reads of a single template molecule showed increasing accuracy over multiple reads, as was expected. In particular, the accuracy obtained from repeatedly sequencing the single template molecule increased with each iterative sequence read or "loop", and approaching the asymptotic maximum after only several complete loops around the template.

[00143] This molecular redundant sequencing was applied in the identification of a single nucleotide variation, e.g., a SNP. Two SMRTbell™ templates were generated that differ at only a single nucleotide (indicated as the 'T' allele and the 'A' allele. The two templates are illustrated in Figure 12, with the T allele shown in Panel A and marked at the variant base with the arrow, and the A allele shown in panel B. The two templates were mixed together at known ratios, ranging from 0% 'A':100% 'T' to 100% 'A':0% 'T'. Single molecule sequencing reactions were performed on each mixture. The resulting traces of sequencing data were filtered for those which contained 6 or greater reads of the insert sequence, and then used to generated consensus calls of the polymorphic position on individual molecules. Figure 13 shows the comparison of the called polymorphism ratios to the expected ratios.

IDT Exhibit 1020.1196

Example 3: Genomic E. coli Sequencing Using Contiguous Template Constructs

[00144] E. coli strain MG1655 was purchased from the ATCC and grown in LB media. DNA was harvested from the cell cultures using the High-molecular-weight genomic DNA purification kit from Qiagen. The genomic DNA was sheared using a nebulizer to generate fragments ranging from 500-1500 bp in size. Fragments were recovered using Qiagen PCR purification columns. The fragmented DNA was end repaired in a reaction containing T4 DNA polymerase, T4 polynucleotide kinase, and dNTPs. After purification, the end-repaired DNA was incubated in the presence of Taq DNA polymerase and dATP to add a single A nucleotide to each fragment. The tailed DNA was then ligated to hairpin oligos to generate the final SMRTbell™ templates. A single hairpin structure was used at both ends of the template. This hairpin contained a 54 nt single-stranded loop closed with an 8 bp stem, as well as a single T nucleotide at the 3' end to enable ligation to A-tailed insert fragments.

[00145] Fragments of sample material that failed to ligate with a hairpin at one or the other end, or that contained a nick due to incomplete ligation, were removed through the use of Exonuclease III and Exonuclease VII. The ligation products were concentrated by ethanol precipitation and then applied to a ChromSpin 1000 column to remove any templates that contained no insert or short inserts. The elution from the ChromaSpin column was purified using the Qiagen PCR purification columns and quantitated by absorbance at 260 nm. The templates were annealed to an equivalent amount of primer, and then subjected to sequencing.

[00146] Prior to immobilization on a ZMW array chip, 60nM SMRTbell™ DNA Library was incubated at 37°C for 1 hour with 10nM modified Phi29 DNA polymerase (N62D, E375Y, K512Y, T368F) (see, e.g., U.S. Patent Application No. 61/072,645, filed March 31, 2008, and incorporated herein by reference in its entirety for all purposes) bearing a biotinylation fusion protein tag, in the following buffer composition: 50mM MOPS, pH 7.5, 75 mM Potassium Acetate, 0.05% Tween-20, 5 mM DTT, 500nM ALEXA568-O-dG6P, 500nM ALEXA555-O-dT6P, 500nM ALEXA647-O-dA6P, 500nM Cy5.5-O-dC6P, 1mM Calcium Chloride. Just prior to immobilization, the mixture was diluted 10-fold in the same buffer composition and 8 μl was loaded onto the ZMW chip having surface immobilized streptavidin. The immobilization was carried out at room temperature for one hour. Prior to sequencing, the immobilization mixture was removed from the ZMW chip. The chip was washed 5 times with 8 μl of the following buffer: 50mM ACES pH 7.1, 120 mM Potassium Acetate, 0.1 mM Calcium Chloride, 120 mM DTT. After these wash steps, 2 additional washes were performed with the following composition: 50mM ACES pH 7.1, 120mM Potassium Acetate,

IDT Exhibit 1020.1197

0.1mM Calcium Chloride, 250nM ALEXA568-O-dG6P, 250nM ALEXA555-O-dT6P, 250nM ALEXA647-O-dA6P, 250nM Cy5.5-O-dC6P, and 120nM DTT. After the washes 4 µl of this nucleotide mix was left on the chip and the chip was placed in sequencing system as previously desribed. The reaction was initiated in real time as previously described by the addition of MnOAc to a final concentration of 0.5mM. Three 9 minute movies were taken for each ZMW chip for generating sequencing data as previously described. The sequenced fragments were aligned to the K12 MG1665 reference sequence.

[00147] Overall, the E. coli genome was sequenced to a depth of 38X coverage where 99.3% of the genome was unambiguously covered. Approximately 4.5 Mbp had relatively high coverage rates (i.e., greater than 20X coverage), giving approximately 99.99992% accuracy for a sequence accuracy score of Q61. For the entire genome, sequence accuracy was determined to be 99.9996%, equating to a quality score of approximately Q54.

[00148] Figure 14 illustrates the coverage map for the E. coli sequence. The plot was corrected for a known artifact related to reduced E. coli replication away from the origin of replication. As can be seen, the coverage level is highly uniform around the average level of 38X coverage. Plotted as a histogram of number of bases vs. level of coverage, the data show comparable distribution to the theoretical maximum coverage See Figure 15, Panel A). Further, when corrected for the variation in replication away from the origin of replication, one can see that the actual observed sequence coverage (Figure 15, Panel B, bars) begins to approach the theoretical maximum sequence coverage based upon Poisson statistics (shown by the dashed line).

Example 4: Large Insert Sequencing

[00149] Large repeated segments of genomic material are traditionally difficult to sequence and assemble into whole genomes since most sequencing platforms have readlength limitations (from tens of bases to hundreds of bases for pyrosequencing platforms and traditional capillary systems) that will not span these repeated regions, thus creating ambiguity in genomic assembly of sequence data, as to where a given sequence read may fit. Using SMRT™ sequencing and its substantially longer readlength, one is better able to span entire large repeat segments in individual reads, thus eliminating ambiguity faced by the shorter read systems.

[00150] A portion of the genome sequence of Escherichia coli strain MG1655 that contains a 2.56 kb exact repeat was targeted for sequencing using the single molecule processes described herein. Figure 16 provides a schematic illustration of the portion of the E. coli genome that

IDT Exhibit 1020.1198

includes this segment. The repeated segment is indicated as a grey bar over genome sequence. Primers targeting the sequence flanking this repeat region are indicated as arrows below the genome sequence. The resulting PCR product was phosphorylated using T4 DNA kinase, and then purified using Qiagen PCR purification columns. A single adenine nucleotide was added using Taq DNA polymerase in the presence of dATP, and then ligated to hairpin oligonucleotides using T4 DNA Ligase. Unligated material was removed using ExoIII and ExoVII. The SMRTbell™ template was concentrated by ethanol precipitation, passed through a ChromaSpin 1000 column, and then purified using Qiagen MinElute columns.

[00151] The template was then sequenced using a SMRT™ sequencing. Sequence reads were generated that spanned substantial portions, and in at least one case, the entire 3 kb segment (~3.2 kb read). By providing single sequence reads spanning the large insert, one significantly reduces any ambiguity that may be associated with such repeats.

[00152] Although described in some detail for purposes of illustration, it will be readily appreciated that a number of variations known or appreciated by those of skill in the art may be practiced within the scope of present invention. To the extent not already expressly incorporated herein, all published references and patent documents referred to in this disclosure are incorporated herein by reference in their entirety for all purposes.

IDT Exhibit 1020.1199

What is claimed is:

1. A method of determining a consensus sequence of nucleotides in a template nucleic acid segment, comprising:

providing a sense and an antisense strand of the template nucleic acid segment in a contiguous nucleic acid molecule;

sequencing both the sense and antisense strands in a polymerase mediated, template dependent sequencing process; and

determining a consensus sequence of the target nucleic acid segment from a sequence of the sense and antisense strands.

2. The method of claim 1, wherein the sense and antisense strands are disposed in a partially contiguous nucleic acid segment.

3. The method of claim 2, wherein the sense and antisense strands are linked at at least a first end by a first linking oligonucleotide.

4. The method of claim 1, wherein the sense and antisense strands are disposed in a completely contiguous nucleic acid segment.

5. The method of claim 4, wherein the sense and antisense strands are linked at each end by first and second linking oligonucleotides, respectively.

6. The method of claim 1, wherein the step of sequencing both the sense and antisense strands comprises sequencing at least one of the sense and antisense strands more than once.

7. The method of claim 4, wherein the sequencing step comprises contacting the contiguous nucleic acid molecule with a primer sequence that is complementary to at least a portion of the first linking oligonucleotide.

8. The method of claim 3, wherein the step of sequencing the sense and antisense strands comprises sequencing the completely contiguous nucleic acid segment more than once.

IDT Exhibit 1020.1200

9. The method of claim 4, wherein the step of sequencing the sense and antisense strands comprises sequencing the completely contiguous nucleic acid segment more than twice.

10. The method of claim 1, wherein the polymerase mediated, template dependent sequencing process comprises detecting incorporation of each nucleotide incorporated by the polymerase.

11. The method of claim 1, wherein the step of detecting incorporation comprises detecting incorporation of each nucleotide incorporated by an individual polymerase.

12. A method of sequencing a nucleic acid sequence, comprising:

providing a template nucleic acid that comprises a double stranded nucleic acid segment having a first and second end;

a first hairpin oligonucleotide connecting each strand of the template nucleic acid at the first end;

a second hairpin oligonucleotide connecting each strand of the template nucleic acid at the second end; and

determining a nucleotide sequence of at least one strand of the template nucleic acid using a template directed, polymerase mediated nucleic acid sequencing process.

13. The method of claim 12, wherein the sequencing process sequences each strand of the double stranded nucleic acid segment .

14. The method of claim 13, comprising the step of comparing sequence data from each strand of the double stranded nucleic acid segment to determine a consensus sequence for the double stranded segment.

15. The method of claim 12, wherein the sequencing process sequences at least one strand of the double stranded nucleic acid segment more than once.

16. The method of claim 15, comprising comparing sequence data obtained from sequencing the at least one strand more than once, to determine a consensus nucleotide sequence of the double stranded nucleic acid segment.

IDT Exhibit 1020.1201

17. The method of claim 13, wherein the sequencing process sequences each strand of the double stranded nucleic acid segment more than once.

18. The method of claim 17, comprising comparing sequence data obtained from sequencing each strand of the double stranded nucleic acid segment more than once to determine a consensus nucleotide sequence for the double stranded nucleic acid segment

19. The method of claim 12, wherein at least one of the first hairpin oligonucleotide and the second hairpin oligonucleotide comprises a primer recognition sequence.

20. The method of claim 12, wherein at least one of the first and second hairpin oligonucleotides comprises a registration sequence.

21. The method of claim 12, wherein the sequencing process is carried out under conditions that permit template directed, polymerase mediated nucleic acid synthesis through the double stranded nucleic acid segment of the template nucleic acid.

22. The method of claim 21, wherein the polymerase mediated nucleic acid synthesis comprises the use of a strand displacing polymerase enzyme.

23. A method of sequencing a nucleic acid sequence, comprising:

providing a template nucleic acid that comprises a double stranded segment having a first and second end;

a first hairpin oligonucleotide connecting each strand of the template nucleic acid at the first end; and

determining a nucleotide sequence of the template nucleic acid using a template directed, polymerase mediated nucleic acid sequencing process.

24. The method of claim 23, wherein the step of providing the template nucleic acid comprises anttiparallel amplification of a region of a sample nucleic acid.

IDT Exhibit 1020.1202

25. A method of sequencing a nucleic acid segment, comprising:

providing a template nucleic acid segment comprising:

first and second complementary nucleic acid strands;

a first connecting nucleic acid linking a 3' end of the first nucleic acid strand with a 5' end of the second nucleic acid strand; and

a second connecting nucleic acid linking the 5' end of the first nucleic acid strand to a 3' end of the second nucleic acid strand;

contacting the template nucleic acid segment with a primer sequence complementary to at least a portion of the template nucleic acid segment; and

monitoring template dependent primer extension to determine a sequence of nucleotides in the template nucleic acid segment.

26. A method of sequencing a nucleic acid, comprising:

providing a template nucleic acid comprising a double stranded segment comprising first and second complementary strands and at least a first single stranded oligonucleotide segment linking a 3' end of the first strand to the 5' end of the second segment; and

monitoring nucleotides incorporated in a template dependent synthesis reaction to identify a sequence of nucleotides in the first strand, the linking oligonucleotide segment and the second strand.

27. A method of preparing a template nucleic acid for nucleic acid sequence determination, comprising:

providing a double stranded target nucleic acid segment;

providing a first linking oligonucleotide that connects a 3' end of a first strand of the double stranded target nucleic acid segment to a 5' end of a second strand of the double stranded target nucleic acid segment, and providing a second linking oligonucleotide that connects a 3' end of the second strand to the 5' end of the first strand; and

providing a primer recognition sequence within at least one of the first and second linking oligonucleotides.

IDT Exhibit 1020.1203

28. The method of claim 27, wherein the first and second linking oligonucleotides comprise hairpin adapter oligonucleotides that are ligated to ends of the double stranded target nucleic acid segment.

29. The method of claim 27, wherein

the first and second linking oligonucleotides comprise a single stranded extension oligonucleotide segment at a 3' end of each of the first and second strands of the target nucleic acid segment; and

further comprising the step of ligating each of the single stranded extension oligonucleotide segments to a 5' end of the second and first strands of the double stranded target nucleic acid segment, respectively.

30. A method of determining a consensus nucleic acid sequence for a template nucleic acid segment, comprising:

providing a sense and an antisense strand of the template nucleic acid segment in a contiguous nucleic acid molecule;

contacting the contiguous nucleic acid molecule with a primer sequence complementary to at least a portion of the contiguous nucleic acid molecule and a polymerase enzyme to provide a nucleic acid synthesis complex, in the presence of a plurality of types of nucleotides or nucleotide analogs, wherein individual nucleic acid synthesis complexes are disposed upon a substrate such that individual complexes are optically resolvable;

monitoring nucleotides or nucleotide analogs incorporated into a nucleic acid synthesis reaction by the individual complexes in a template dependent manner to determine a nucleic acid sequence of the sense and antisense strands; and

comparing the nucleic acid sequence of the sense and antisense strands to determine a consensus nucleic acid sequence of the template nucleic acid.

31. A method of sequencing a nucleic acid sequence, comprising:

preparing a template nucleic acid that comprises a double stranded segment having a first and second end;

a first hairpin loop connecting each strand of the template nucleic acid at the first end; and

IDT Exhibit 1020.1204

a second hairpin loop connecting each strand of the template nucleic acid at the second end; and

determining a nucleotide sequence of the template nucleic acid using a template directed, polymerase mediated nucleic acid sequencing process.

32. A method of sequencing a nucleic acid sequence, comprising:

preparing a template nucleic acid sequence comprising:

first and second complementary nucleic acid strands;

a first connecting nucleic acid linking a 3' end of the first nucleic acid strand with a 5' end of the second nucleic acid strand; and

a second connecting nucleic acid linking the 5' end of the first nucleic acid strand to a 3' end of the second nucleic acid strand;

contacting the template nucleic acid sequence with a primer sequence complementary to at least a portion of the template nucleic acid; and

monitoring template dependent primer extension to determine a sequence of nucleotides in the template nucleic acid sequence.

33. A method of sequencing a nucleic acid, comprising:

preparing a template nucleic acid comprising a double stranded segment comprising first and second complementary strands and at least a first single stranded oligonucleotide segment linking a 3' end of the first strand to the 5' end of the second segment; and

monitoring nucleotides incorporated in a template dependent synthesis reaction to identify a sequence of nucleotides in the first strand, the linking oligonucleotide segment and the second strand; and

recording the sequence of nucleotides in the template nucleic acid in a computer readable medium.

34. A method of sequencing nucleic acid samples, comprising:

preparing template nucleic acid segments from each of a plurality of discrete nucleic acid samples, wherein the template nucleic acid segments comprise double stranded segments of the nucleic acid samples, a first strand of the double stranded segment being linked to a second strand

IDT Exhibit 1020.1205

of the double stranded segment by a linking oligonucleotide, wherein the linking oligonucleotide in each discrete nucleic acid sample comprises a unique, identifiable sequence characteristic;

pooling the template nucleic acid segments from the plurality of discrete nucleic acid samples;

sequencing the template nucleic acid segments pooled in the pooling step; and

identifying nucleic acid sequences as deriving from a discrete nucleic acid sample based at least in part on the unique identifiable sequence characteristic identified in the sequencing step.

35. A composition, comprising:

a template nucleic acid, comprising:

a double stranded nucleic acid segment comprising a first strand segment and a second strand segment substantially complementary to the first strand segment; and

at least a first linking oligonucleotide segment joining a 3' end of the first strand segment to the 5' end of the second strand segment;

a primer sequence capable of hybridizing to at least a portion of the template nucleic acid and initiating polymerase mediated nucleic acid synthesis; and

a polymerase enzyme.

36. The composition of claim 35, wherein the primer is capable of hybridizing to a portion of the first linking oligonucleotide segment.

37. The composition of claim 35, further comprising a second linking oligonucleotide segment joining a 5' end of the first strand with a 3' end of the second strand.

38. The composition of claim 35, wherein the template, primer and polymerase are present as a complex.

39. The composition of claim 38, wherein the complex is immobilized upon a substrate surface.

40. The composition of claim 39, wherein the complex is immobilized upon a substrate surface such that individual complexes are optically resolvable from each other.

IDT Exhibit 1020.1206

41. The composition of claim 40, wherein individual complexes are immobilized within discrete zero mode waveguides.

42. The composition of claim 39, wherein the complex is immobilized upon the substrate surface through a linkage between one of the primer sequence or the polymerase, and the surface of the substrate.

43. The composition of claim 39, wherein the substrate comprises one or more zero mode waveguides disposed thereon, and the complex is immobilized within a core of the one or more zero mode waveguides.

44. The composition of claim 35, further comprising a plurality of types of nucleotide analogs labeled with a detectable label component.

45. The composition of claim 44, wherein the detectable label component comprises one or more of a fluorescent moiety, an energy donor moiety, a fluorescence quencher moiety, a FRET donor, a FRET acceptor, and a fluorescent nanocrystal.

46. A template preparation kit, comprising:

a first linking oligonucleotide;

a primer sequence complementary to at least a portion of the first linking oligonucleotide; and

one or more ligation reagents for coupling the first linking oligonucleotide to a 3' end of a first strand of a double stranded template and a 5' end of a second strand of the double stranded template nucleic acid.

47. A system for sequencing a nucleic acid template, comprising:

a reaction mixture comprising:

a complex comprising a template nucleic acid, a polymerase enzyme and a primer sequence complementary to at least a portion of the template sequence, wherein the template sequence comprises a double stranded segment having a sense strand and an antisense strand, and at

IDT Exhibit 1020.1207

least a first linking oligonucleotide linking a 5' end of the sense strand to the 3' end of the antisense strand;

at least a first nucleotide or nucleotide analog bearing a detectable labeling group; and

a detection system configured to detect incorporation of the first nucleotide or nucleotide analog into a primer extension product by the polymerase.

48. The system of claim 47, further comprising a processor operably connected to the detection system, the processor including machine implemented instruction set for transforming signal data transmitted from the detection system to the processor to nucleic acid sequence information interpretable by a user.

IDT Exhibit 1020.1208

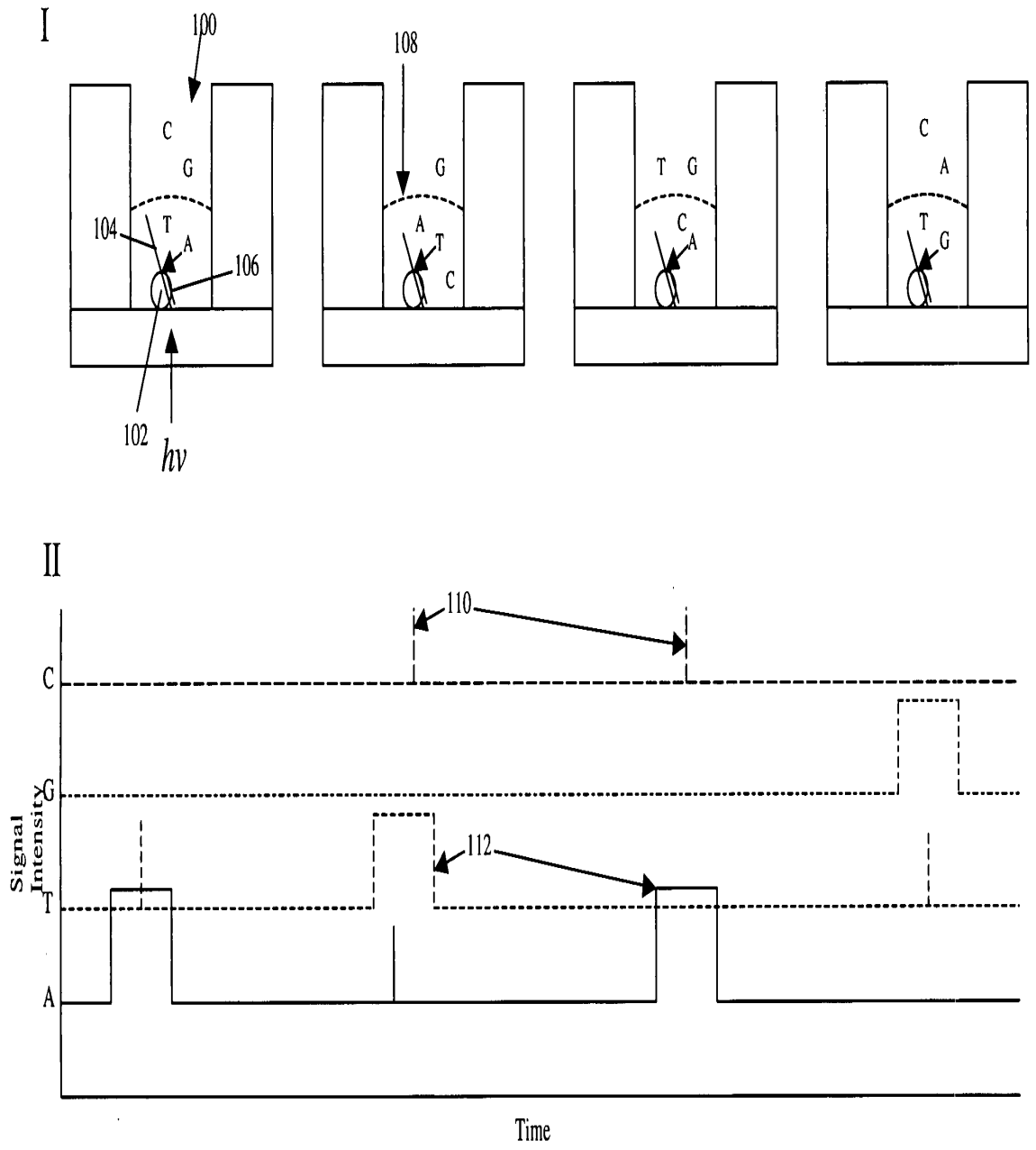

I
100
108
C
G
T
A
104
106
102
hv
G
A
T
C
T
G
C
A
C
A
T
G
II
110
C
G
T
A
112
Signal Intensity
Time

Fig. 1A

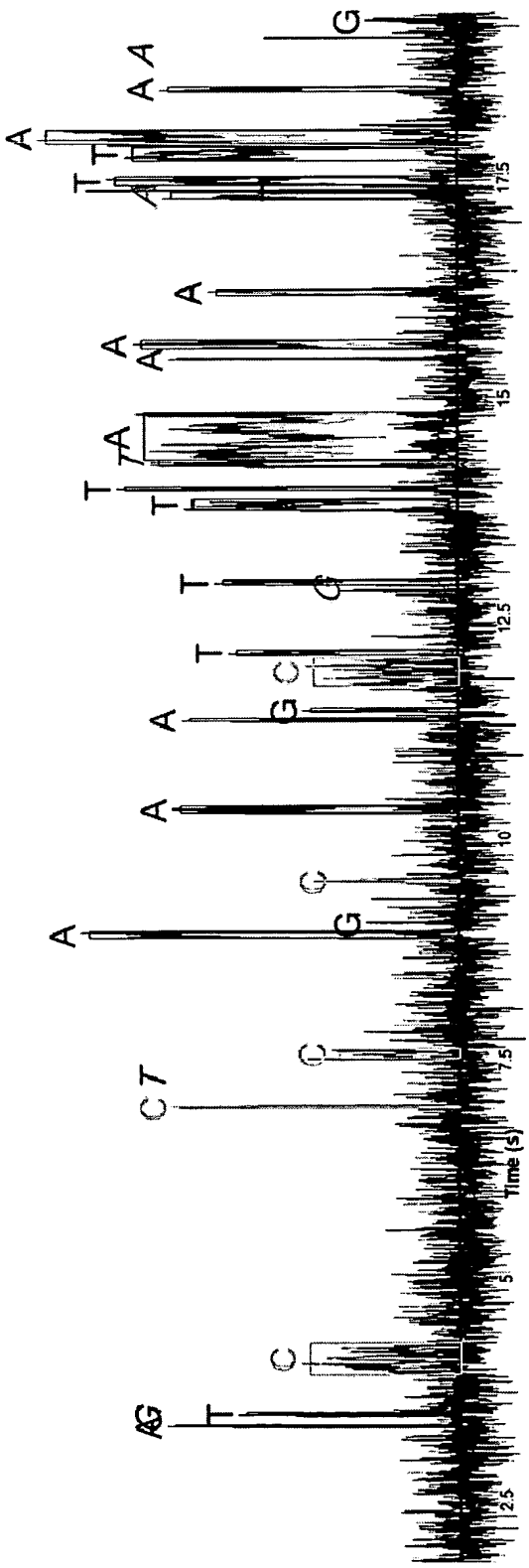


Fig. 1B

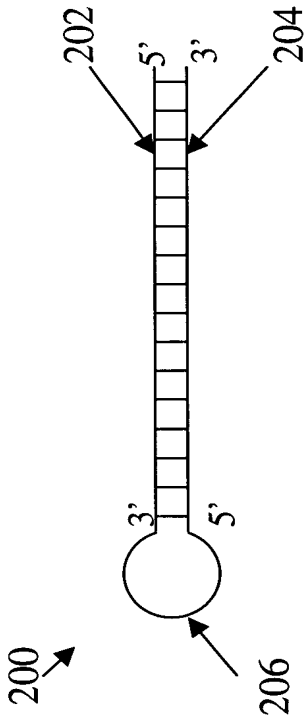


Fig. 2A

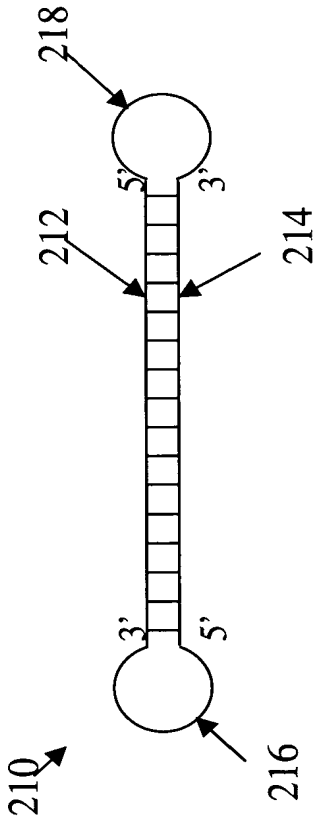


Fig. 2B

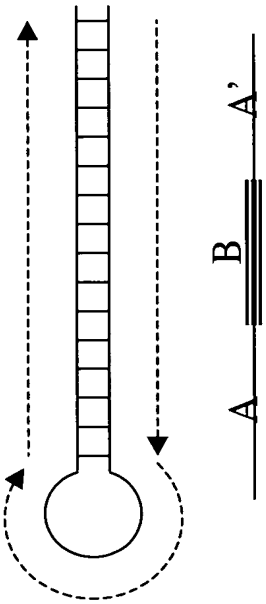


Fig. 3A

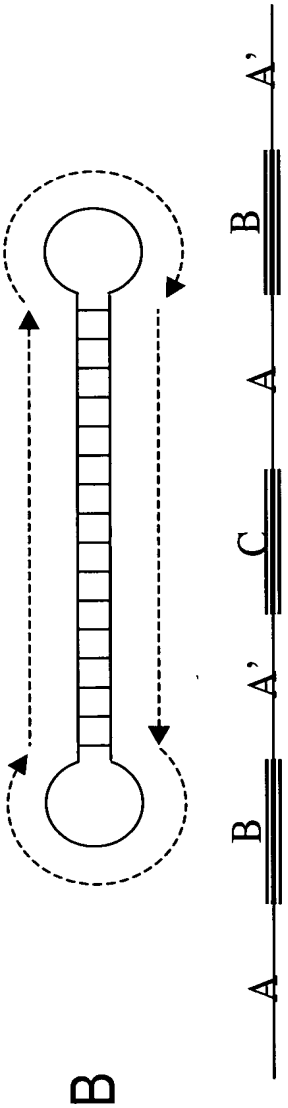


Fig. 3B

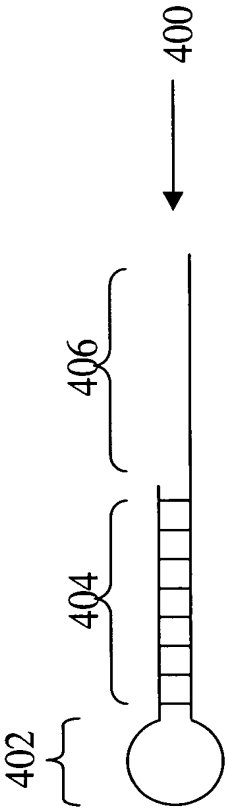


Fig. 4A

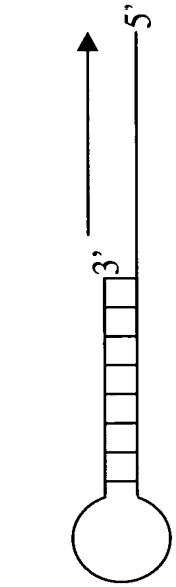


Fig. 4C

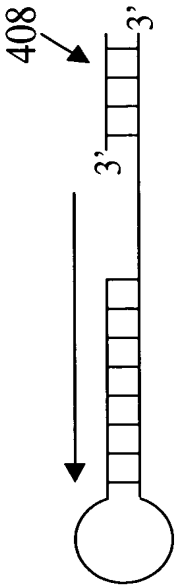


Fig. 4B

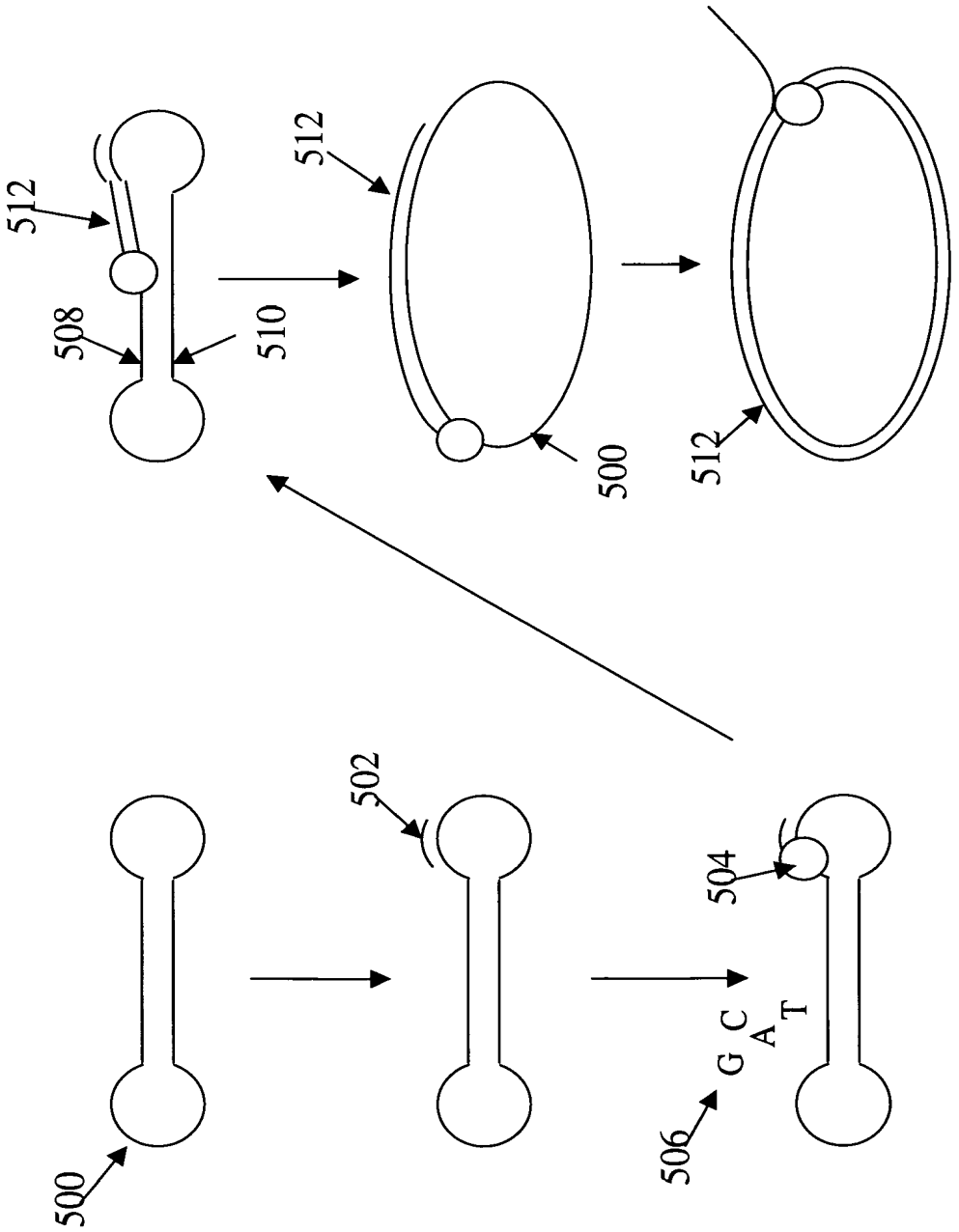


Fig. 5

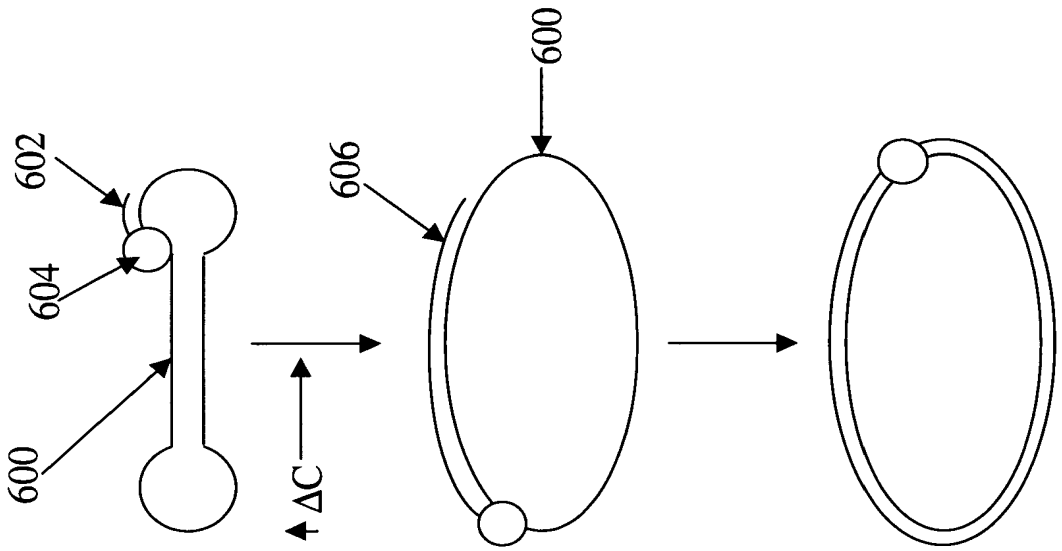


Fig. 6

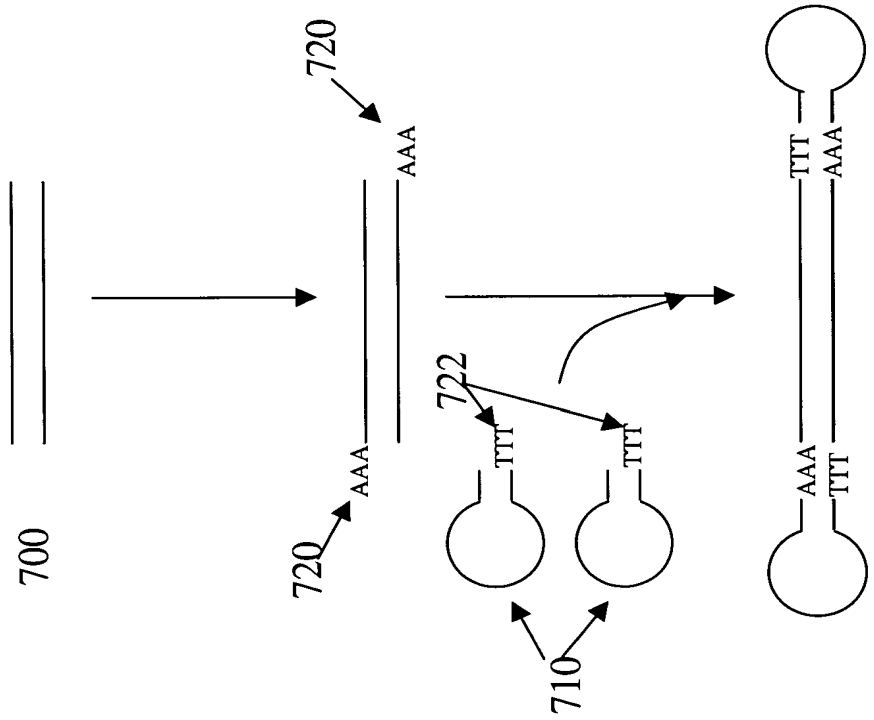


Fig. 7

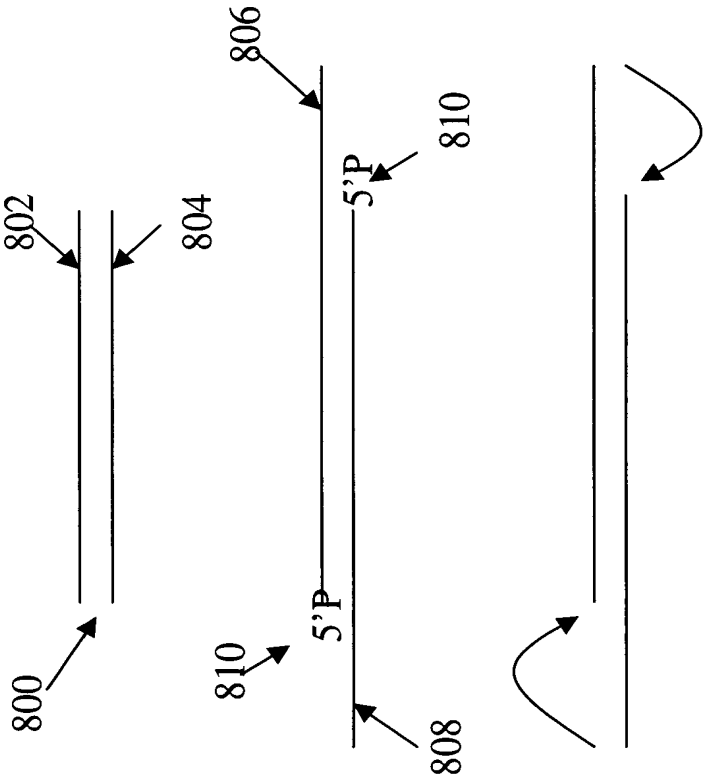


Fig. 8

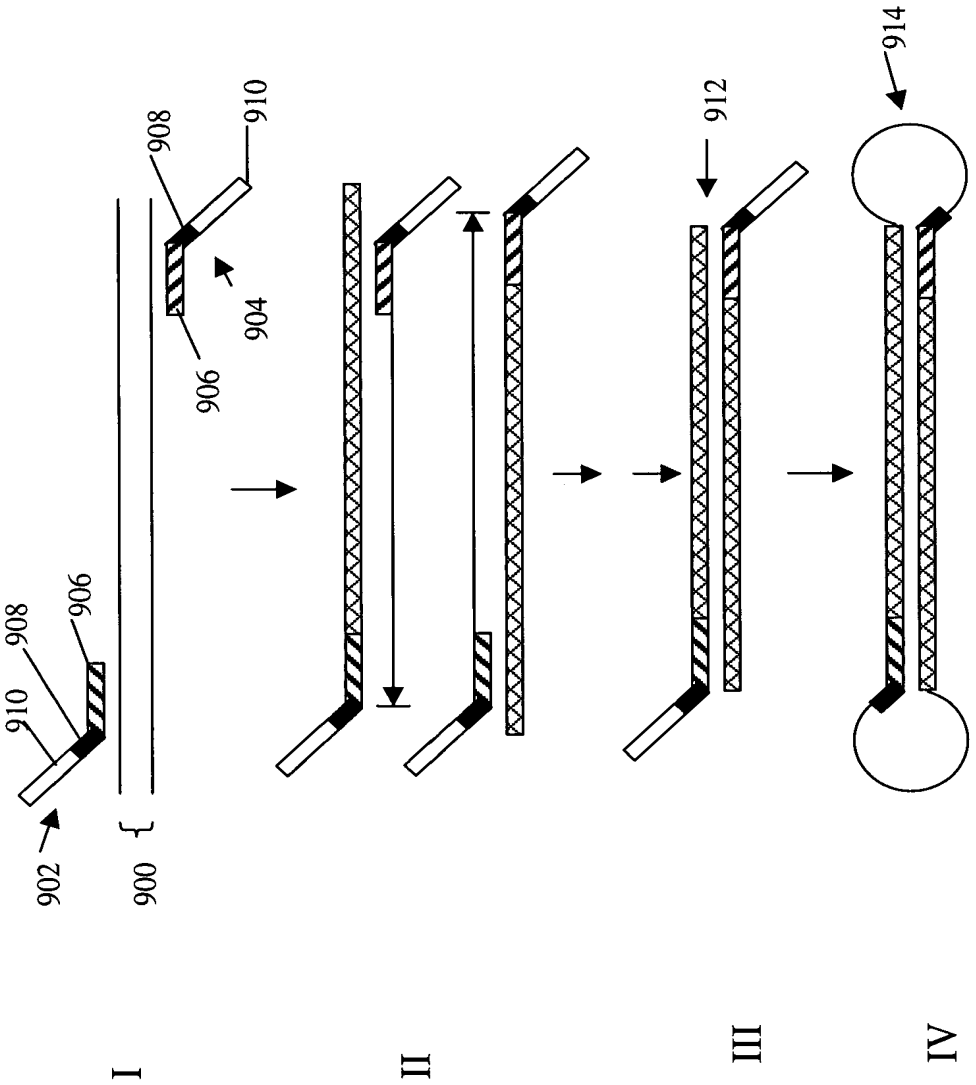


Fig. 9

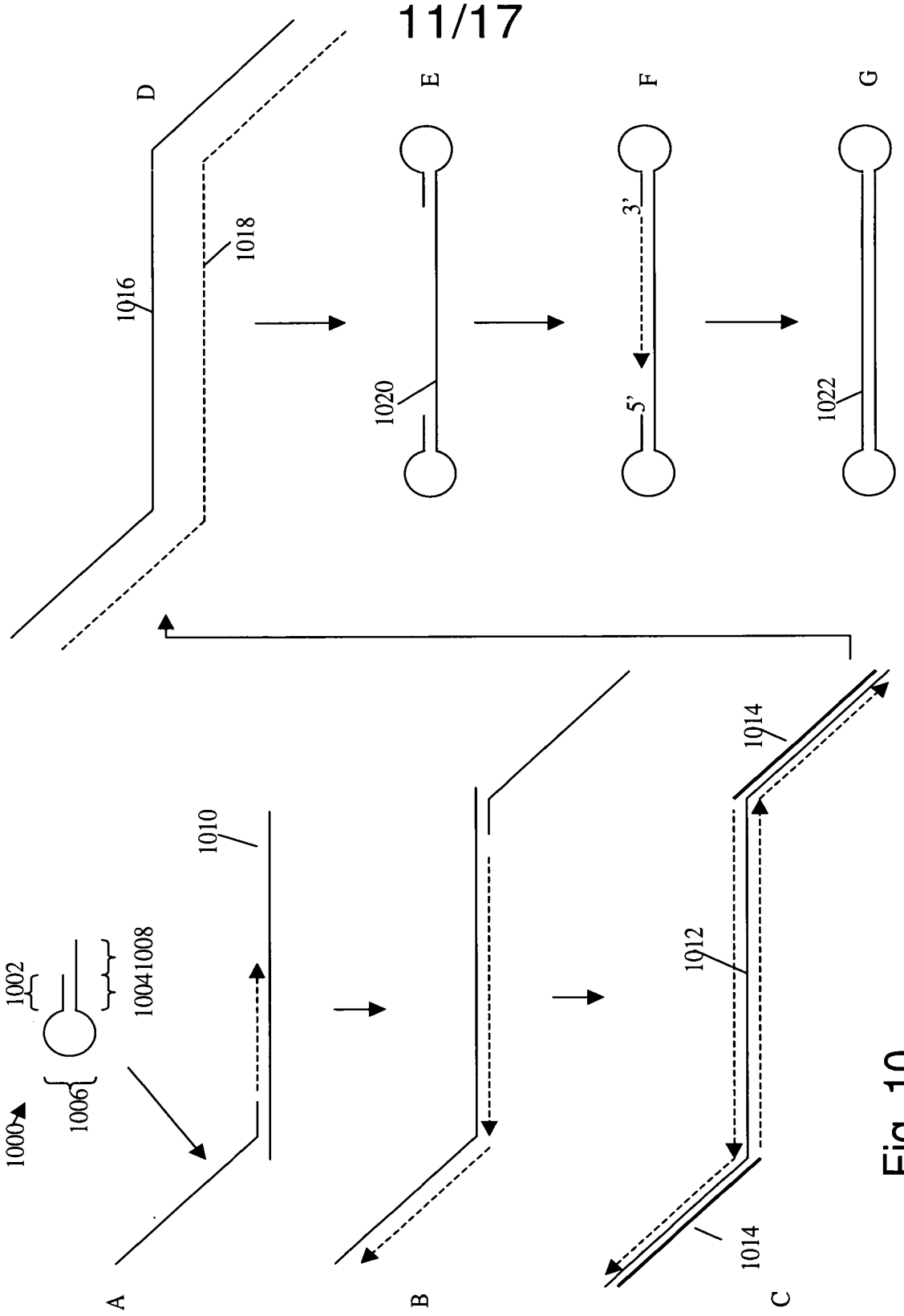

A
B
C
D
E
F
G
1000
1002
1004
1006
1008
1010
1012
1014
1014
1016
1018
1020
1022
3'
5'

Fig. 10

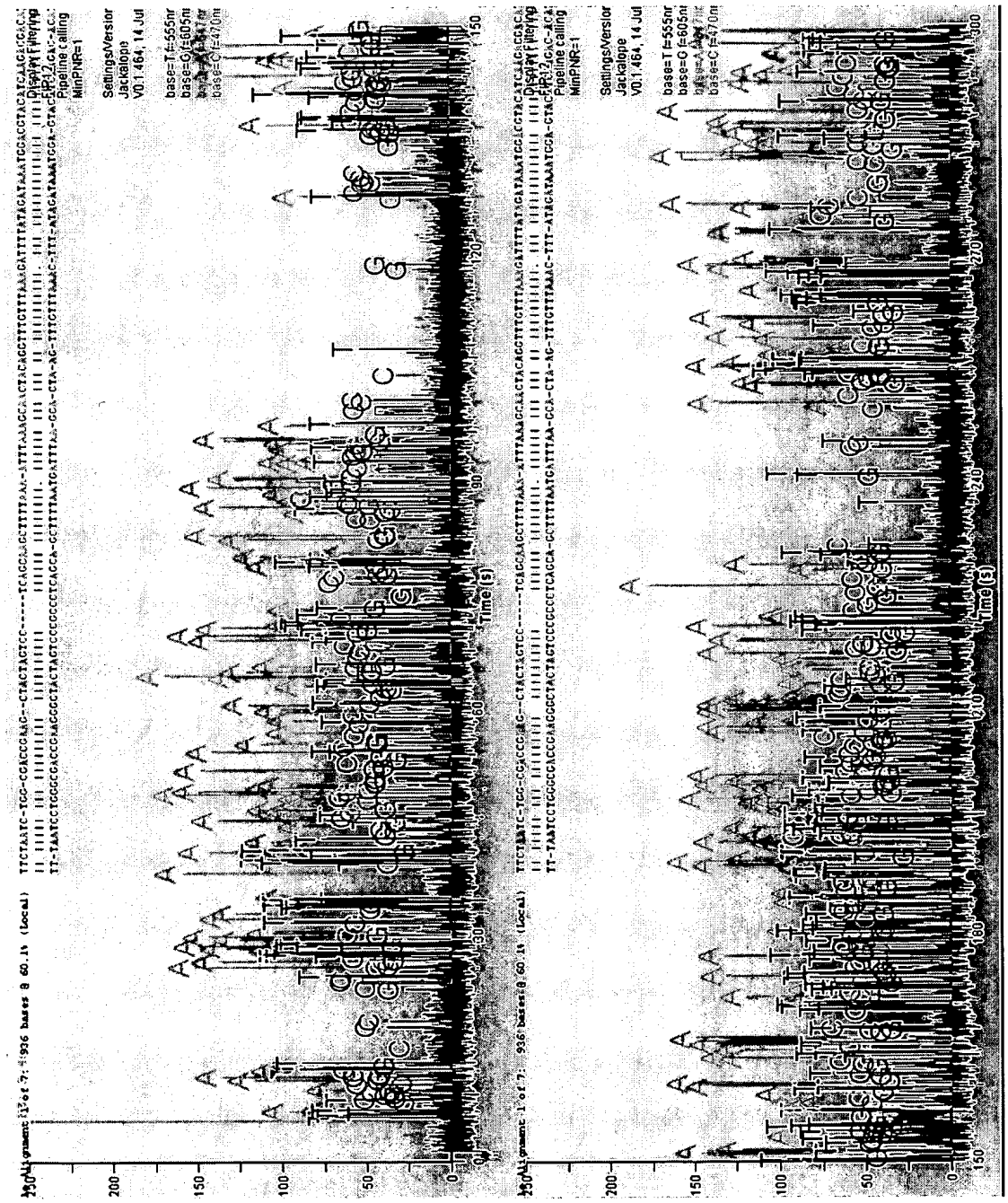
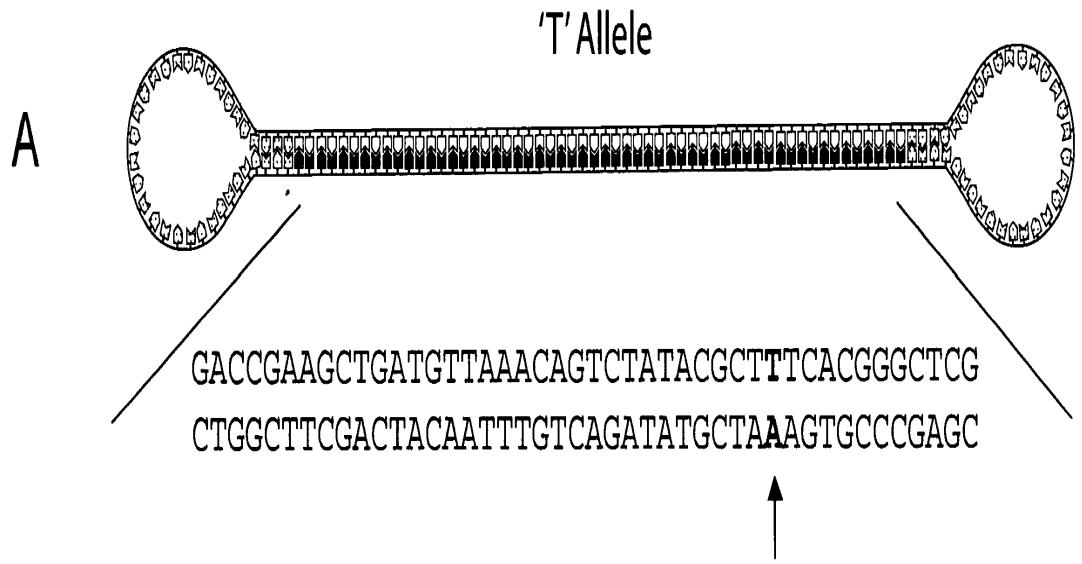


Fig. 11


A
'T' Allele
GACCGAAGCTGATGTTAAACAGTCTATACGCTTTCACGGGCTCG
CTGGCTTCGACTACAATTTGTCAGATATGCTAAAGTGCCCGAGC

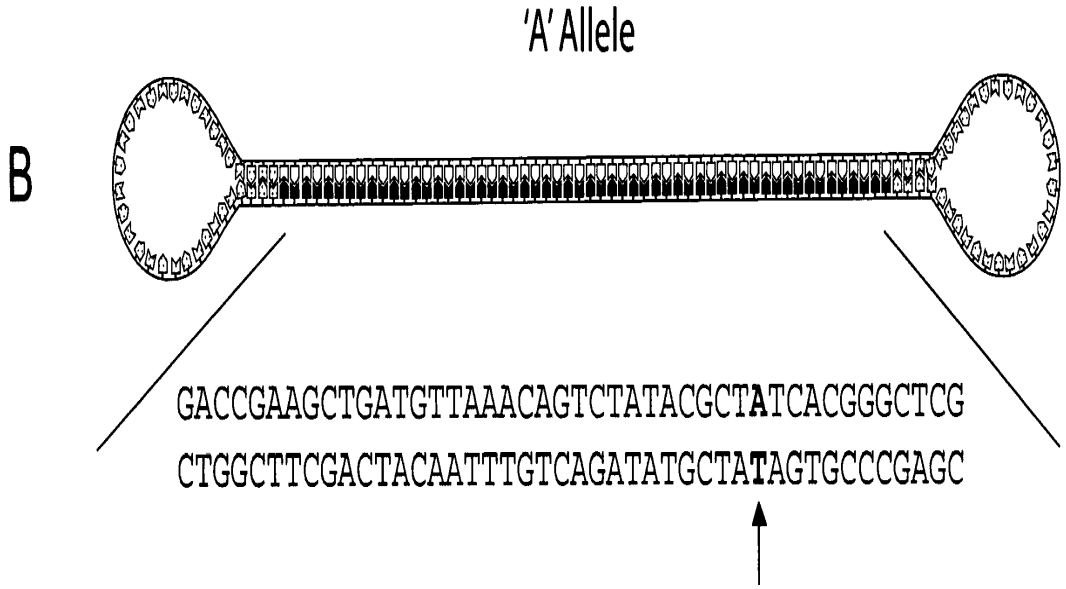

B
'A' Allele
GACCGAAGCTGATGTTAAACAGTCTATACGCTATCACGGGCTCG
CTGGCTTCGACTACAATTTGTCAGATATGCTATAGTGCCCGAGC

Fig. 12

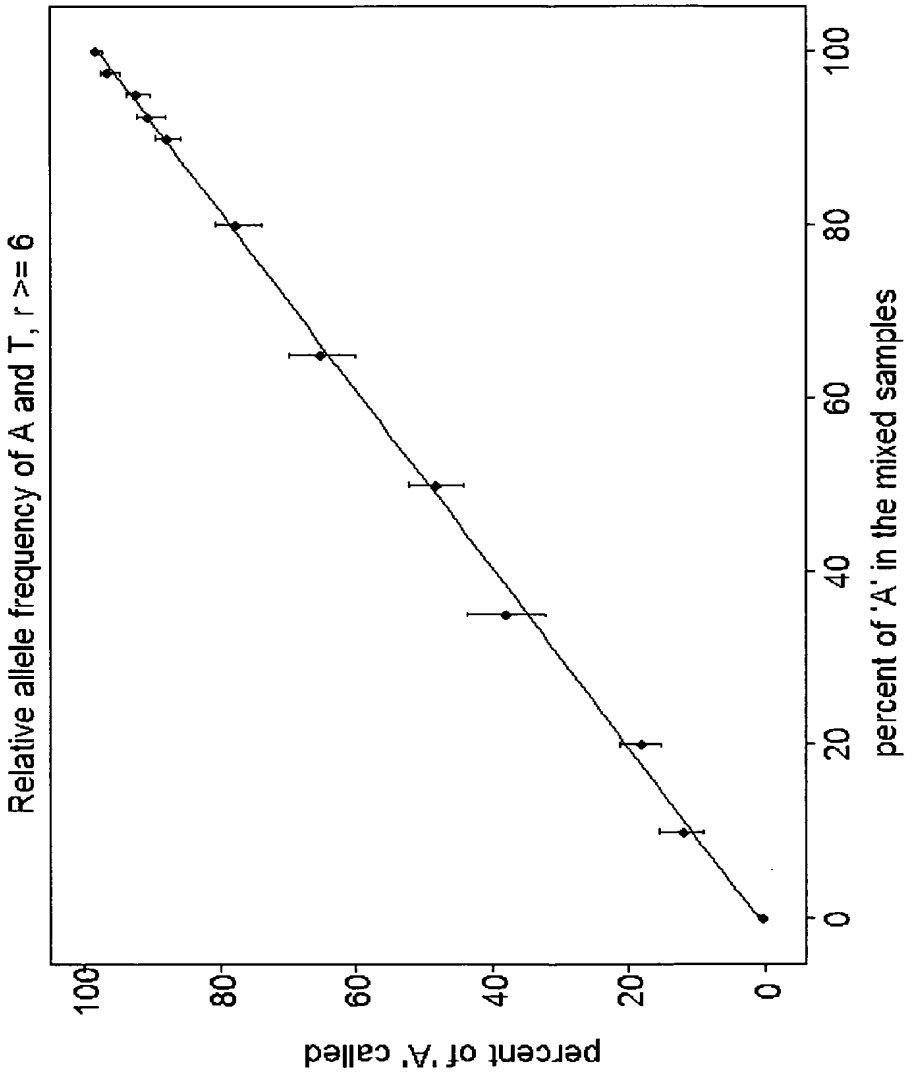

Relative allele frequency of A and T, r >= 6
percent of 'A' called
percent of 'A' in the mixed samples
0
20
40
60
80
100

Fig. 13

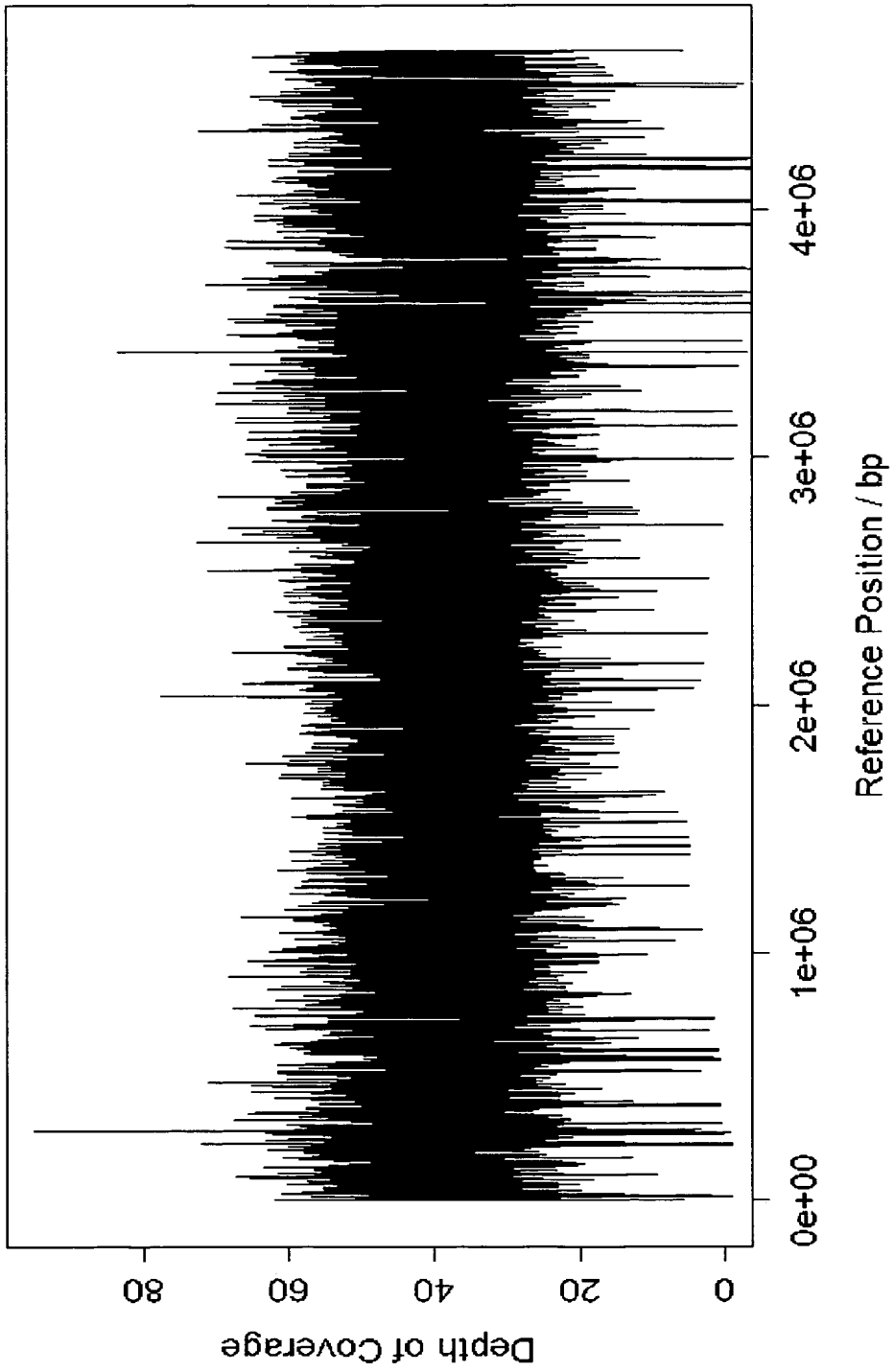

Depth of Coverage
0
20
40
60
80
Reference Position / bp
0e+00
1e+06
2e+06
3e+06
4e+06

Fig. 14

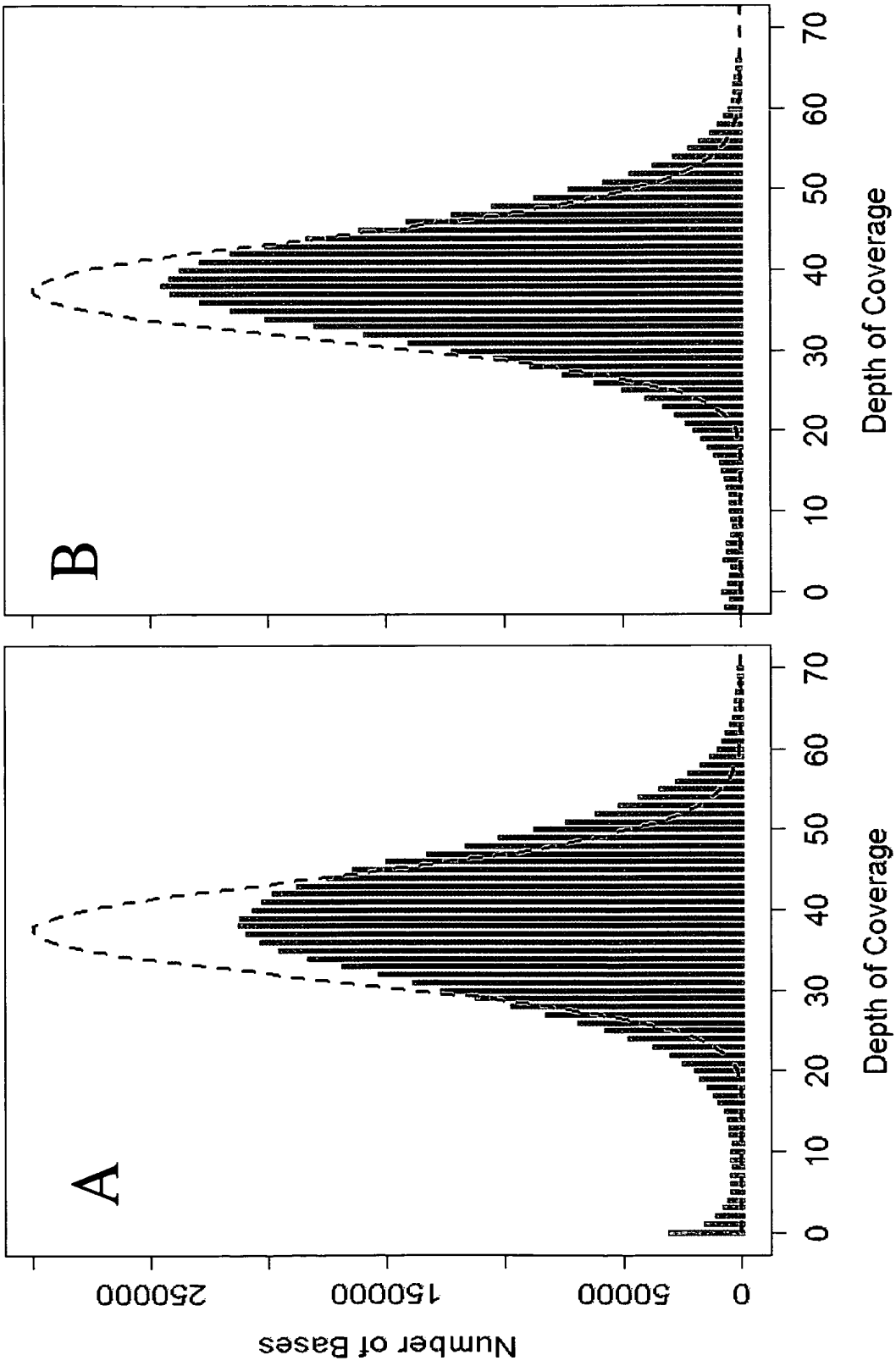

A
B
Number of Bases
0
50000
150000
250000
Depth of Coverage
0
10
20
30
40
50
60
70

Fig. 15

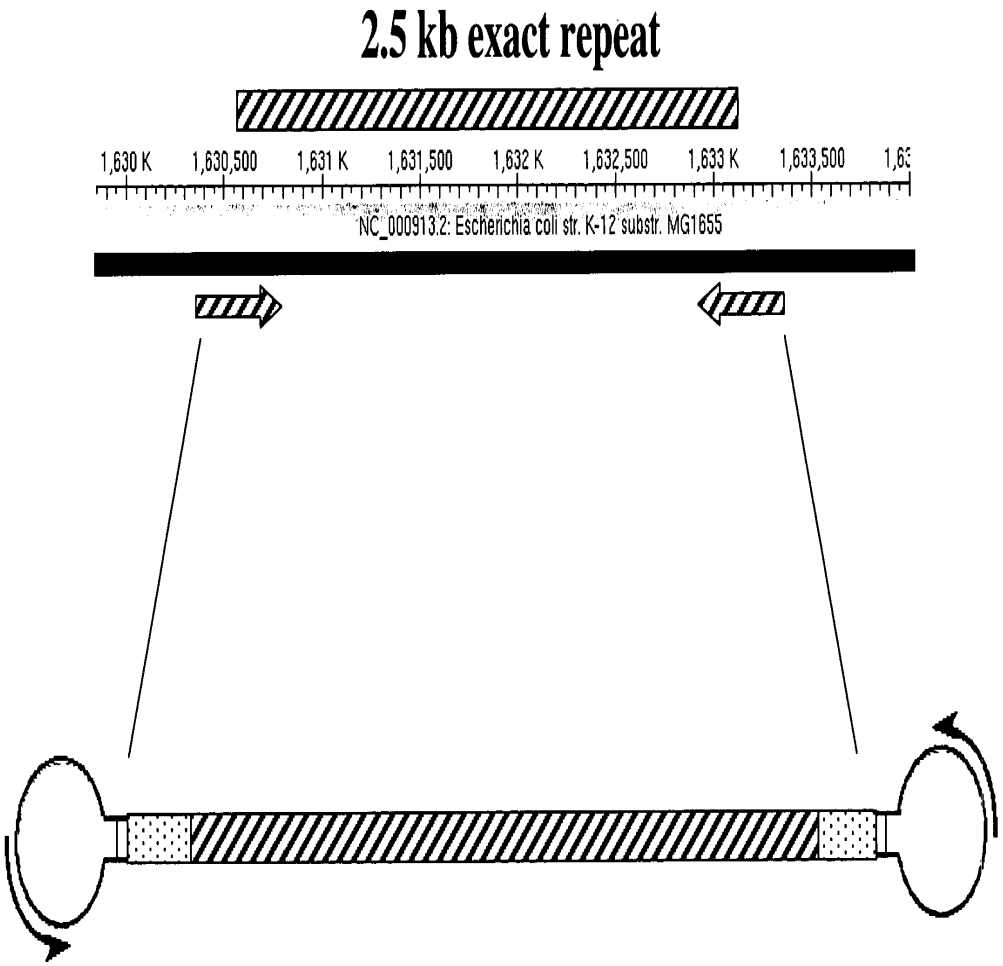

2.5 kb exact repeat
1,630 K
1,630,500
1,631 K
1,631,500
1,632 K
1,632,500
1,633 K
1,633,500
NC_000913.2: Escherichia coli str. K-12 substr. MG1655

Fig. 16

Espacenet

Bibliographic data: WO2011009941 (A1) — 2011-01-27

PROBES FOR SPECIFIC ANALYSIS OF NUCLEIC ACIDS

Inventor(s): ERICSSON OLOF [SE]; ISAKSSON MAGNUS [SE]; JOHANSSON HENRIK [SE]; LANDEGREN ULF [SE] ± (ERICSSON, OLOF, ; ISAKSSON, MAGNUS, ; JOHANSSON, HENRIK, ; LANDEGREN, ULF)

Applicant(s): OLINK GENOMICS AB [SE]; ERICSSON OLOF [SE]; ISAKSSON MAGNUS [SE]; JOHANSSON HENRIK [SE]; LANDEGREN ULF [SE] ± (OLINK GENOMICS AB, ; ERICSSON, OLOF, ; ISAKSSON, MAGNUS, ; JOHANSSON, HENRIK, ; LANDEGREN, ULF)

Classification: **- international:** ***C12Q1/68***
- cooperative: **C12Q1/6855**

Application number: WO2010EP60715 20100723 Global Dossier

Priority number(s): GB20090012909 20090723

Also published as: CN102575293 (A) CN102575293 (B) EP2456888 (A1) EP2456888 (B1) JP2012533314 (A) more

Abstract of WO2011009941 (A1)

The present invention provides a method for detecting or enriching for a target deoxyribonucleic acid (DNA) present in a nucleic acid sample, said method comprising: (a) fragmenting a nucleic acid sample to generate nucleic acid fragments including a target fragment containing said target DNA; (b) rendering said fragments, including said target fragment, at least partially single-stranded, wherein the single- stranded portion includes an end portion and wherein the length of said single- stranded portion is sufficient to allow hybridisation of at least part of the single- stranded portion of said target fragment to the probe of step (c); (c) contacting the at least partially single-stranded fragments of step (b) with oligonucleotides A and B of a single target-specific nucleic acid probe, wherein: (i) oligonucleotide A is a single- stranded oligonucleotide comprising at one end a first target-specific part comprising at least 10 nucleotides complementary in sequence to at least part of said single- stranded portion of said target fragment, and comprising at the other end a second non-target-specific part which comprises a nucleotide sequence complementary to at least a portion, including

IDT Exhibit 1020.1226

one end, of oligonucleotide B of the probe, and (ii) oligonucleotide B is a single-stranded oligonucleotide which may contain or carry at least one element for detection and/or enrichment of said target fragment, and of which at least a portion, including one end, is complementary in sequence to the second non-target-specific part of oligonucleotide A, such that said target fragment becomes annealed to said probe through hybridisation to the first target-specific part of oligonucleotide A resulting in only one target-specific probe-binding event per target fragment; (d) ligating oligonucleotide B of said probe to the part of the single- stranded portion of said target fragment which is hybridised to oligonucleotide A of said probe to produce a probe-target fragment hybrid; and (e) detecting or enriching for said probe-target fragment hybrid. Kits for use in the method of the invention are also provided.

IDT Exhibit 1020.1227

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
27 January 2011 (27.01.2011)

PCT

(10) International Publication Number
WO 2011/009941 A1

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

(21) International Application Number:
PCT/EP2010/060715

(22) International Filing Date:
23 July 2010 (23.07.2010)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
0912909.9 23 July 2009 (23.07.2009) GB

(71) Applicant *(for all designated States except US)*: **OLINK GENOMICS AB** [SE/SE]; Dag Hammarskjölds väg 36B, S-752 37 Uppsala (SE).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **ERICSSON, Olof** [SE/SE]; Vaksalagatan 38 A, S-753 31 Uppsala (SE). **ISAKSSON, Magnus** [SE/SE]; Salagatan 5 B, S-753 30 Uppsala (SE). **JOHANSSON, Henrik** [SE/SE]; Botvidsgatan 8C, S-753 29 Uppsala (SE). **LANDEGREN, Ulf** [SE/SE]; Norra Parkvägen 29, S-756 45 Uppsala (SE).

(74) Agent: **DZIEGLEWSKA, Hanna**; St Bride's House, 10 Salisbury Square, London Greater London EC4Y 8JD (GB).

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

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

Published:

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

(54) **Title**: PROBES FOR SPECIFIC ANALYSIS OF NUCLEIC ACIDS

(57) **Abstract**: The present invention provides a method for detecting or enriching for a target deoxyribonucleic acid (DNA) present in a nucleic acid sample, said method comprising: (a) fragmenting a nucleic acid sample to generate nucleic acid fragments including a target fragment containing said target DNA; (b) rendering said fragments, including said target fragment, at least partially single-stranded, wherein the single- stranded portion includes an end portion and wherein the length of said single- stranded portion is sufficient to allow hybridisation of at least part of the single- stranded portion of said target fragment to the probe of step (c); (c) contacting the at least partially single-stranded fragments of step (b) with oligonucleotides A and B of a single target-specific nucleic acid probe, wherein: (i) oligonucleotide A is a single- stranded oligonucleotide comprising at one end a first target-specific part comprising at least 10 nucleotides complementary in sequence to at least part of said single- stranded portion of said target fragment, and comprising at the other end a second non-target-specific part which comprises a nucleotide sequence complementary to at least a portion, including one end, of oligonucleotide B of the probe, and (ii) oligonucleotide B is a single-stranded oligonucleotide which may contain or carry at least one element for detection and/or enrichment of said target fragment, and of which at least a portion, including one end, is complementary in sequence to the second non-target-specific part of oligonucleotide A, such that said target fragment becomes annealed to said probe through hybridisation to the first target-specific part of oligonucleotide A resulting in only one target-specific probe-binding event per target fragment; (d) ligating oligonucleotide B of said probe to the part of the single- stranded portion of said target fragment which is hybridised to oligonucleotide A of said probe to produce a probe-target fragment hybrid; and (e) detecting or enriching for said probe-target fragment hybrid. Kits for use in the method of the invention are also provided.

WO 2011/009941 A1

Probes for specific analysis of nucleic acids

The present invention relates to a method for detecting or enriching target nucleic acids in a complex nucleic acid sample. In such a method the sample is fragmented and fragments containing the target nucleic acid are covalently attached, at or near one end by nucleic acid ligation ("single-sided ligation", i.e. one end of the target fragment is not involved in the ligation reaction and remains free), to a single target-specific probe per target fragment. Optionally, common nucleic acid adaptors may be non-target-specifically added to said free end of the fragments. The probes may contain elements facilitating enrichment of the target fragments, such as by purification (e.g., by immobilisation to a solid phase), or amplification (e.g., by utilising elements in the probes and optional adaptors). In this way, target nucleic acids may be obtained from complex samples for subsequent analysis by, e.g., nucleic acid sequencing, microarray, qPCR, visualisable hybridisation probes, *in situ* analyses, etc. The feature of the present invention of providing (and using) a ligation probe specific for only one end of a target nucleic acid (or target nucleic acid fragment) is advantageous, and facilitates particular applications not possible with known methods.

It is often desired to isolate large numbers of sub-genomic sequences to allow their further characterisation, e.g. the isolation of the genomic regions believed to be of relevance to a particular physiological or pathological condition. This is particularly so following the advent of parallel methods of high-throughput sequencing, which necessitate methods for rapidly isolating genomic sequences of interest for use in such sequencing methods.

One known method for amplifying (amplification constituting a means of "enriching" according to the present invention) in parallel a plurality of target nucleic acids from a sample of nucleic acids is disclosed in WO 2005/111236. In this method, partially double-stranded "Selector" nucleic acid molecules (either a single symmetrical molecule in which the longer strand overhangs at both ends, or two asymmetrical molecules each having a single-stranded overhang at only one end) are hybridised via their single-stranded overhangs in a target-specific manner to both ends of single-stranded (denatured) target fragments resulting from fragmentation of the nucleic acid sample (in the case of the single symmetrical Selector, the hybridised Selector-target fragment becomes circularised). In a particular embodiment of the method using the symmetrical Selector, only one end of the target fragment hybridises to an end of the Selector, the other end of the Selector

IDT Exhibit 1020.1229

hybridising internally of the target nucleic acid fragment and requiring a structure-specific endonuclease to resolve the resulting structure by cleaving off the portion of the target fragment "protruding" beyond the internal hybridised region. In all cases, therefore, the amplifiable portion of the target fragment is delineated by the regions (whether both end regions or one end and one internal region) of known sequence to which the Selector(s) has been designed to hybridise. Following hybridisation (and, where appropriate, resolution of the secondary structure) the Selector(s) and target nucleic acid fragment are joined by ligation to give (i) in the case of the symmetrical Selector, a circular nucleic acid molecule and (ii) in the case of the asymmetrical Selectors a linear molecule comprising the target fragment flanked by Selectors. The double-stranded region of the Selector(s) contains a primer pair motif which is common to the plurality of different target-specific Selectors used in a multiplex assay. Hence, amplification of multiple target fragments can be achieved simultaneously whilst avoiding amplification artefacts which can result from the use of multiple, different primer pairs.

In the method of WO 2005/111236, two target-specific hybridisation events are required for each target fragment; hybridisation of Selector(s) to both ends, or to one end and one internal sequence, of the target fragment. This inevitably requires knowledge of the sequence of the target nucleic acid at least at these two hybridising regions, in order for target-specific hybridisation probes to be designed.

However, it is often desired to isolate or enrich for genomic sequences or fragments based on knowledge of only one region of sequence, i.e. wherein the fragment (or region of sequence) it is desired to isolate is not bounded at both ends (delineated) by regions of known sequence. This will frequently be the case when the need to isolate the target sequence (e.g. fragment) arises from a desire to subject the sequence/fragment to sequencing, i.e. at least part of the sequence is not known. A particular example of this is the analysis of chromosomal translocation breakpoints, where it is desired to determine the precise sequence of the region of a chromosome to which a region of another chromosome, containing known sequence, has become fused. Other examples include the analysis of splicing patterns or VDJ recombination events. Hence, there is a need for a method of rapidly and efficiently detecting (or enriching for or capturing) one or a plurality of target nucleic acids which does not require knowledge of the sequence of regions at both ends of the corresponding target nucleic acid fragment.

It has now been found that the limitation of the above-described Selector(s) method may be overcome through the use of a single (i.e. a single species of) probe, that is a single (or single species of) probe per target nucleic acid, which has

IDT Exhibit 1020.1230

specificity to a region of known sequence in the nucleic acid fragment corresponding to the target nucleic acid. In other words, the present invention is based on the novel feature of using only one target-specific probe for each target nucleic acid, which binds in a target-specific manner only once (more particularly at only one site) in the target nucleic acid fragment. Thus, for each target nucleic acid fragment resulting from fragmentation of the nucleic acid sample, only one target-specific hybridisation and ligation ("binding") event occurs; the target-specific ligation is single-sided with respect to the target nucleic acid fragment and therefore only a single region of known sequence in the target nucleic acid fragment is required in order to design the probe. In other words, there need only be, and is only, one defined binding site for the probe in each target nucleic acid fragment, i.e. the sequence of the target fragment needs to be defined at only one site, e.g. at only one end. The method of the invention thereby enables and, advantageously, facilitates the detection or enrichment of genomic nucleic acids lacking known sequence at both ends (i.e. having known sequence at only one end).

The use of such a single target-specific probe offers further advantages beyond the ability to detect or enrich for target nucleic acids having unknown sequence at one end (i.e. the end other than that to which the target-specific probe hybridises). The ligation of the probe to the target nucleic acid fragment results in a covalent connection between these molecules. Thus, the actual genomic fragment, and not an amplification product thereof, becomes covalently connected to the probe. In certain embodiments of the invention the probe contains elements facilitating immobilisation of the probe to a solid phase. In such embodiments the covalent connection between the target nucleic acid fragment and the probe results in a stable capture of the former and allows the use of highly stringent washing steps to remove non-specifically hybridised (non-ligated) fragments, resulting in a high specificity. Washes of such high stringency cannot be used with known hybridisation- (as opposed to ligation-) based methods such as microarray capture and fluorescence *in situ* hybridisation (FISH), and the ability to use such washes in the present method is therefore advantageous.

The use of the single target-specific probes in the methods of the invention need not be limited only to methods using genomic DNA. In view of the favourable results obtained using human genomic DNA (see the Examples) it is envisaged that the methods of the invention would work equally well with any DNA sample that contains one or more target nucleic acids. Thus, for example, the methods of the invention could use one or more cDNA samples, e.g. cDNA obtained, derived or synthesised from the RNA of whole organisms, specific tissues, cell types etc or

IDT Exhibit 1020.1231

biological samples exposed to various conditions which may alter the gene expression of said sample resulting in an altered cDNA population.

Accordingly, the present invention provides a method for detecting or enriching for a target deoxyribonucleic acid (DNA) present in a nucleic acid sample, said method comprising:

(a) fragmenting a nucleic acid sample to generate nucleic acid fragments including a target fragment containing said target DNA;

(b) rendering said fragments, including said target fragment, at least partially single-stranded, wherein the single-stranded portion includes an end portion and wherein the length of said single-stranded portion is sufficient to allow hybridisation of at least part of the single-stranded portion of said target fragment to the probe of step (c);

(c) contacting the at least partially single-stranded fragments of step (b) with oligonucleotides A and B of a single target-specific nucleic acid probe, wherein:

(i) oligonucleotide A is a single-stranded oligonucleotide comprising at one end a first target-specific part comprising at least 10 nucleotides complementary in sequence to at least part of said single-stranded portion of said target fragment, and comprising at the other end a second non-target-specific part which comprises a nucleotide sequence complementary to at least a portion, including one end, of oligonucleotide B of the probe, and

(ii) oligonucleotide B is a single-stranded oligonucleotide which may contain or carry at least one element for detection and/or enrichment (in particular detection, amplification and/or capture) of said target fragment, and of which at least a portion, including one end, is complementary in sequence to the second non-target-specific part of oligonucleotide A,

such that said target fragment becomes annealed to said probe through hybridisation to the first target-specific part of oligonucleotide A resulting in only one target-specific probe-binding event per target fragment;

(d) directly or indirectly ligating oligonucleotide B of said probe to the part of the single-stranded portion of said target fragment which is hybridised to oligonucleotide A of said probe to produce a probe-target fragment hybrid; and

(e) detecting or enriching for said probe-target fragment hybrid.

As described in more detail below, the method of the invention may advantageously include an additional step, between steps (a) and (b) in which a common nucleic acid adaptor is non-target-specifically annealed to the ends of the fragments. Such a method involving the use of a common adaptor represents a preferred embodiment of the present invention. In particular, in such a step the

IDT Exhibit 1020.1232

annealed adaptor becomes ligated to the fragments only at the 3' ends or only at the 5' ends of the strands of the fragments. More particularly, the adaptor is ligated to one end of the strand of the target fragment to which, at the other end, oligonucleotide B of the probe is ligated in step (d). In other words, in the target fragment the adaptor is ligated at the opposite end of the strand to which the target-specific probe becomes ligated.

The method of the invention may be performed as described above ("simplex" format) to enrich for a single (i.e. a single species of, which will normally be present in many copies) target fragment or for a plurality of target fragments which are sufficiently similar in sequence so as to be possible to enrich for using the same probe. In this context it will be seen that the term "single" as used in relation to the target-specific probe in part (c) means single in the context of a particular target fragment, namely that one probe (or more particularly one type or species of probe) is used for each target fragment (i.e. a single probe per target fragment). Thus, where there is only one target fragment only one probe (in the sense of one species of probe) will be used. Accordingly, part (c) above may alternatively be expressed as:

"(c) contacting the at least partially single-stranded fragments of (b) with oligonucleotides A and B of a target-specific nucleic acid probe wherein for each target fragment a single probe is used, wherein"

It is clear from the above that "single" probe means single species of probe and does not imply any limitation on the actual number of probe molecules used.

Alternatively, a plurality (i.e. a plurality of species) of probes may be used in a "multiplex" format simultaneously to enrich for a plurality of target DNAs. Hence, in such a latter aspect the method as defined above is for enriching for a plurality of target DNAs, wherein in step (c) said target fragments are contacted with oligonucleotides A and B of a plurality of nucleic acid probes, each having an oligonucleotide A with a different first target-specific part, whereby a plurality of different target fragments may be annealed to said probes. In such a multiplex method, as stated above, for each target fragment of the plurality (i.e. each different type or species of target fragment) a single (i.e. in the sense of a single species of) probe will be used. Thus, a plurality of probes will be used, with a (different) single probe for each target fragment. The term "plurality" as used herein means 2 or more (or at least 2), more particularly 3 or more (or at least 3), or 4, 5, 6, 8, 10, 15, 20, 30, 50, 70 or 100 or more etc. In certain embodiments even higher numbers of probes may be used and very many different targets may be enriched or detected, e.g. 500, 1000, 2000, 5000 or 10,000 or more. For example, 10, 100, 1000 or 10000 different

IDT Exhibit 1020.1233

probes may simultaneously be used to detect or enrich for, respectively, 10, 100, 1000 or 10000 different target fragments.

The term "enriching" as used herein is used broadly and includes any means of selecting, isolating and/or capturing a DNA sequence of interest ("target DNA") from a nucleic acid sample which contains other nucleic acids, particularly other DNAs, in addition to (and including DNAs being part of the same DNA molecule as) the target DNA. Thus "enriching" encompasses any means of practically, if not necessarily physically, "separating" the target DNA from the other nucleic acids present in the sample, more particularly by means of covalently attaching a probe thereto in order that the target DNA may be subjected to analytical and/or preparative or synthetic, or other enrichment, techniques. "Enriching" for the target DNA as used herein may, for example, be amplification by one of the many known methods of nucleic acid amplification, or may be physical capture, for example by immobilisation to a solid phase, optionally followed by amplification.

Although as discussed further below, in certain embodiments the method may involve immobilisation on a solid phase, this is not a necessary feature of the method. Thus, the method may be carried out in solution, i.e. may be a homogeneous method.

The term "detecting" is also used broadly herein and includes any means of detecting or determining or assaying for the presence of the target DNA, or any means of analysing the target DNA. Direct analysis of the target DNA sequence (i.e. sequencing of all or any part of the target DNA) is encompassed by the term detecting. As described further below this may be accomplished by any means enabled by the probe ligated to the target molecule, such as the presence in the probe of a sequencing primer binding site Thus, the method of the invention allows a selected target fragment to be sequenced. More particularly, a target fragment selected by means of the target-specific probe according to the present invention may be directly sequenced. Of course, the invention also covers indirect analysis, e.g. sequencing, of the target DNA, for example analysis, e.g. sequencing, of an amplified target DNA, or of a captured target DNA fragment.

The step of detecting or enriching for the target DNA involves detecting or enriching for the probe-target fragment hybrid. Thus, detection or enrichment of the target DNA may be achieved by any means of detection or enrichment which is selective for the hybrid. As described in more detail below, this may be achieved by means of detection and/or enrichment (e.g. capture and/or amplification) elements which are present in the probe (and thereby incorporated into the hybrid, but not present on other, non-target, fragments present in the sample which have not

IDT Exhibit 1020.1234

become ligated to the probe) or more generally by any means which is dependent on the ligation of the probe to the target fragment or, put another way, dependent on formation of the probe-target fragment hybrid.

The term "target DNA" as used herein means a DNA of interest which it is desired to detect or for which it is desired to enrich. This will generally be a part (or a portion or segment) of a longer DNA molecule which may be present in the sample. Thus, it may be a region or stretch of a longer DNA present in the nucleic acid sample (or more particularly, DNA sample). The target DNA may be of any length, but in order to be detected or enriched by the method of the invention must comprise or be contained within a fragment (the "target fragment") produced by the step of fragmenting the nucleic acid sample. The sequence of the target DNA may not be known, providing that at least a portion of the target fragment is of known sequence in order to facilitate the design of the probe, which must be able to hybridise to a single region of the target fragment.

Whilst the target DNA may be of any length, advantageously the methods disclosed herein may be used for detecting or enriching large target fragments or, alternatively viewed, target fragment polynucleotides of a minimum length, by which is meant target fragments comprising at least 30 nucleotides. More preferably, the target fragments comprise at least 40, 50, 60, 70, 80, 90, 100, 125, 150, 200, 250 or 300 nucleotides. Where the target fragment comprises an overhang, e.g. as a result of cleavage (fragmentation), the minimum length of the target may include or exclude the overhang sequence.

The "nucleic acid sample" referred to above may be any sample which contains any amount of nucleic acid, from any source or of any origin, in or from which it is desired to detect or enrich a target DNA known or suspected to be comprised therein. More particularly, the sample may be any sample that contains DNA. The sample may be complex, e.g. whole genomic DNA, cDNA from a whole organism, tissue or cell population, or a fraction thereof. In this regard it may for example be a direct product of a nucleic acid isolation procedure, or of a cell lysis procedure, or it may be further be fractionated or purified in some way, e.g. it may contain nucleic acids which have been partially or fully separated in some way, or treated in any way, e.g. RNA treated to produce cDNA. The sample may be from any eukaryotic or prokaryotic or viral source, e.g. may be of microbial (for example bacterial or fungal), plant, or animal (e.g. vertebrate, mammalian or primate) origin. In a particular aspect, the sample may be of human origin e.g. human genomic DNA or cDNA. The sample may be from a single origin, or may be formed from a pooling of a plurality of samples from different origins. In the latter case, the fragmenting step (a),

IDT Exhibit 1020.1235

and optionally one or more further steps of the method, may be performed on the samples separately prior to said pooling. For example, said nucleic acid sample may represent the pooling of a number of patient samples, such that the method of the invention allows parallel enrichment of target DNAs from multiple patients.

Thus, the target DNA is preferably genomic DNA. It may represent total genomic DNA or it may be a sub-fraction thereof. Thus, the sample may comprise or may be derived from genomic DNA, e.g. by separating/isolating genomic DNA directly or copying (e.g. amplifying) genomic DNA.

Alternatively, the target DNA is preferably cDNA (complementary or copy DNA). The, the sample may comprise cDNA or a subfraction thereof, wherein cDNA is DNA that is complementary to messenger RNA or DNA that has been synthesized from messenger RNA by reverse transcriptase. Preferably, in the context of the present invention the cDNA has been treated to comprise double stranded DNA. Thus cDNA can be considered to be a copy of the RNA present in a cell at the time of extraction or isolation or a fraction thereof, i.e. it represents all or some of the genes that were expressed in said cell at the time of isolation. Hence, a cDNA sample may represent the genes expressed from a whole organism or a part thereof, e.g. a tissue or cell type or group or sub-group thereof, and may further represent genes expressed under specific conditions, e.g. at a particular time, in a specific environment, at a stage of development or in response to stimulus etc. The cDNA may represent a subfraction of the RNA isolated from any of the above sources, e.g. the RNA or cDNA may be fractionated, e.g. by size, to encompass only a proportion of the genes expressed in the source. The cDNA may be derived from a single source or multiple sources as described elsewhere herein. The sample may comprise cDNA derived directly from the mRNA or the cDNA may be derived indirectly, e.g. the cDNA may be amplified, e.g. via the production of a cDNA library.

In a first step (a) of the method the nucleic acid sample known or suspected to contain the target DNA is fragmented to produce fragments, amongst which will exist (if the target DNA is present in the sample, and if the method of fragmentation is selected appropriately) at least one (i.e. at least one species of) fragment containing the target DNA. The term "fragmenting" is used broadly herein to include any means by which the nucleic acid in the sample may be fragmented or cleaved (i.e. divided or "cut" into smaller pieces or fragments). Thus, fragmentation may be carried out enzymatically, e.g. using restriction or other endonucleases or nucleases such as DNase, and/or physically, e.g. by nebulisation or sonication or any shear-based methods. Such physical methods result in unpredictable, non-sequence-specific fragmentation, as do certain (non-restriction) endonucleases. Thus both random, and

IDT Exhibit 1020.1236

pre-determined (or site-specific) fragmentation is encompassed, but the latter is preferred. Accordingly, fragmentation using an enzyme which cleaves at a known or defined site is preferred, in other words enzymes which cleave sequence-specifically or structure-specifically, or, put another way, which cleave to generate ends of known (defined) sequence, for example restriction endonucleases and FLAP endonucleases. However, also encompassed by the reference to "fragmenting" in step (a) is fragmentation of a nucleic acid sample which inherently may occur as a result of the age of a sample, the conditions in which it is stored and any treatment of the sample (e.g. fixation, for example in a tissue sample), and the degradation to which these factors contribute. Any suitable class of restriction endonuclease may be used, including type II and type IIs enzymes. The use of type IIs endonucleases is particularly advantageous as these result in fragments which will not have the same end sequences, thereby improving the chances targeting the probe to the intended target fragment through the use of an oligonucleotide A of appropriate sequence. Alternatively, the cleaving (fragmenting) may be achieved using a flap endonuclease, wherein an added nucleic acid or oligonucleotide is only partially hybridisable, due to being partially double-stranded, to a sequence in the nucleic acid sample resulting in a protruding non-hybridised region of the nucleic acid sample adjacent to a hybridised region. This secondary structure is the substrate for a so-called structure-specific "flap endonuclease" enzyme which cleaves the nucleic acid sample at the junction of the hybridised and non-hybridised regions (Lyamichev V et al, Science. 1993 May 7;260(5109):778-83). The use of a flap endonuclease may be advantageous in the absence of a (known) restriction enzyme recognition sequence near to the target DNA within the nucleic acid sample, as it allows cleavage (fragmentation) to be targeted at any region of known sequence. Flexibility in positioning the site of cleavage is thereby afforded. When using a flap endonuclease, a preceding physical fragmentation step may be desirable.

Fragmenting means may be used in combination, e.g. the use together of two or more endonucleases, more particularly two or more restriction endonucleases, or the use together of an enzymatic and a physical means. Furthermore, the nucleic acid sample may be differently fragmented in separate aliquots, which aliquots are then pooled and together subjected to the remaining steps of the method of the invention.

Hence, the fragmenting may be achieved by separating the nucleic acid sample into a plurality of aliquots and fragmenting the respective aliquots with different means or different combinations of means, such means being for example restriction enzymes. The aliquots are then subjected to the remaining steps of the

IDT Exhibit 1020.1237

method and may be pooled at any point during the method, for example before step (b), before step (c), before step (d) or before or after step (e) to give the (single) nucleic acid sample referred to in the above method. Such an embodiment should be distinguished from the separate fragmenting and subsequent pooling discussed above, in which case the samples are of different origins (rather than aliquots of a single sample). However, in that aspect the samples of different origin may themselves respectively be fragmented in multiple separate aliquots and respectively pooled as mentioned above, before the pooling to give the (single) nucleic acid sample, referred to in the above method, for use in subsequent steps of the method.

Known heterozygous polymorphisms in the sample nucleic acid may be exploited by fragmenting the sample using, if possible, endonucleases which recognise a sequence inactivated in at least one instance by such a polymorphism. By designing probes targeted to the fragments produced in the presence and absence of cleaving at the polymorphic endonuclease recognition site said haplotype-specific fragments may independently be enriched and analysed.

After the fragmenting step the fragments of the nucleic acid sample, including the fragment which contains the target DNA, are rendered at least partially single-stranded (step (b)). Where the fragments are not made completely single-stranded, they will at least be single-stranded at an "end portion", which means that the part of such fragments which is single-stranded will include one end of the fragments, and the single-stranded portion will be of sufficient length to allow, in the case of the target fragment, hybridisation of at least part of said portion to the probe. Thus "rendering at least partially single-stranded" as used herein encompasses all means of causing a double-stranded DNA fragment to become single-stranded wholly or at least at an end-containing portion. Such means include denaturation, e.g. by heat or pH or through the use of chemicals, as is known in the art. Heat denaturation is particularly preferred. Rendering the fragments only partially single-stranded such that they remain largely or at least partially double-stranded is advantageous, especially for the enrichment of long genomic sequences, as this avoids to some extent the undesirable cross-reactivity between single-stranded nucleic acid fragments and thereby reduces the incidence of hybrids which must be distinguished between.

Alternatively, at least partial single-strandedness can be achieved by 3' or 5' exonucleolysis using an appropriate 3' or 5' exonuclease. Starting at a free double-stranded fragment end, such enzymes progressively degrade or digest one strand of a double-stranded nucleic acid, leaving the complementary strand and rendering the nucleic acid single-stranded along the length of the enzyme's action. The extent of

IDT Exhibit 1020.1238

exonucleolytic degradation (i.e. the length of the resulting single-stranded region) may be controlled by the duration of the reaction. The duration of the exonuclease reaction is chosen in order that an appropriate length of one end of the strands of the fragments is removed. The extent of digestion must be sufficient to allow a productive hybridisation (i.e. hybridisation capable of templating ligation) with the first target-specific part of oligonucleotide A of the probe, but not so much that the fragment loses all double-strandedness and becomes two single-stranded fragments. Suitable exonucleases are known in the art and include e.g. exonuclease III (3') and lambda exonuclease (5'). The use of exonuclease III is particularly preferred. As regards the use of exonucleases it should be noted that, if fragmentation is to be performed in step (a) by restriction endonucleolysis, it must be borne in mind when selecting the endonucleases whether they produce 5' or 3' overhanging, or blunt, ends, as certain types of restriction ends are poor substrates for certain exonucleases. The substrate preferences of exonucleases are known in the art. For example, exonuclease III does not favour 3' overhanging ends, and hence the use of endonucleases which leave 5' overhanging ends is in some aspects preferred. As mentioned above, the single-stranded end-containing portion of the target fragment must be able to hybridise, at least in part, to the probe. Thus, the single-stranded portions of the fragments must be of sufficient length to support, in step (c), stable base-pairing between at least part of the single-stranded portion of the target fragment and the first target-specific part of oligonucleotide A of the probe. "Sufficient to allow hybridisation" as used herein therefore means that the partially single-stranded target fragment must be single-stranded to the extent necessary to permit a productive hybridisation with the probe, i.e. a hybridisation capable of templating ligation of the hybridised portion of the fragment to oligonucleotide B of the probe in step (d). This does not require, but does include, 100% complementarity between the first target-specific part of oligonucleotide A of the probe and the corresponding region of the single-stranded portion of the target fragment. "Complementary", as used herein, means functionally complementary, i.e. a level of complementarity sufficient to mediate hybridisation, which encompasses degrees of complementarity less than 100%.

The reference in step (b) above to "at least part" of the single-stranded portion of the target fragment encompasses, for instances where said single-stranded portion is longer than the first target-specific part of oligonucleotide A of the probe to which at least part of said single-stranded portion is complementary, hybridisation of said part of oligonucleotide A to a portion of said single-stranded portion which includes the end of the target fragment as well as, in the alternative, to an internal non-end portion. Thus, the first target-specific part of oligonucleotide A

IDT Exhibit 1020.1239

has functional complementarity, and therefore hybridises, either to an end-containing portion or to an internal non-end-containing portion of the single-stranded portion which, as mentioned above, includes an end portion. Thus, in certain embodiments of the invention the first target-specific part of oligonucleotide A hybridises at (is functionally complementary to) one end of the single-stranded portion of the target fragment. As discussed in more detail below, in certain other embodiments of the invention the hybridisation and ligation steps (c) and (d) occur through resolution of a generated secondary structure using a flap endonuclease (as discussed above in the context of fragmenting step (a)), the substrate for which is generated by hybridisation of oligonucleotide A of the probe to such an internal non-end portion of the target fragment.

Once the fragments of the nucleic acid sample have been rendered at least partially single-stranded in step (b) they are contacted with oligonucleotides A and B of the single target-specific probe of step (c). As noted above, reference to a "single" probe is intended as a reference to a single species of probe; although in practice many copies of the probe (and other nucleic acid reagents) will be used, for a given particular target fragment only one probe (albeit many copies thereof), rather then two or more different probes, is contacted with the fragments of step (b). The probe comprises or consists of oligonucleotides A and B hybridised together. As oligonucleotide A has a first target-specific part which is not complementary to oligonucleotide B, the probe is partially single stranded. As is apparent from step (c), the probe may be provided "intact" (with oligonucleotides A and B pre-hybridised) or oligonucleotides A and B may be added un-hybridised in which case hybridisation to form an intact probe will occur during the step (c) of annealing the probe (oligonucleotide A) with the target fragment (as used herein, "annealed" means non-covalently joined or connected and therefore does not include ligation). Reference in step (c) to "oligonucleotides A and B of a single target-specific nucleic acid probe" also encompasses a probe wherein oligonucleotides A and B are connected together, or joined, directly or indirectly, to form a single contiguous oligonucleotide, oligonucleotides A and B representing a self-complementary portion thereof (wherein oligonucleotides A and B may be annealed (hybridised) together). Thus, in such a probe oligonucleotides A and B may represent the respective "parts" (more particularly "hybridising parts") of a single continuous partially self-complementary oligonucleotide. In such a case, the end of the probe other than the single-stranded target-specific part (of oligonucleotide A) is a hairpin or hairpin loop. Hence, as used herein, reference to "oligonucleotides" A and B is a reference to the respective parts of such an oligonucleotide. Alternatively viewed, oligonucleotides A and B may be

IDT Exhibit 1020.1240

seen as oligonucleotide sequences, which may be provided either as separate oligonucleotide molecules (i.e. separate moieties) or they may be part of a single oligonucleotide molecule. When oligonucleotides A and B are parts of such a single oligonucleotide, the oligonucleotide may be pre-hybridised or un-hybridised when contacted with the at least partially single-stranded fragments in step (c).

Advantageously, the fragments are contacted with un-hybridised oligonucleotides A and B of the probe, with a large amount of oligonucleotide A, and an even greater amount of oligonucleotide B (except in the case that oligonucleotides A and B are parts of a single oligonucleotide), being added in order that as high a proportion as possible of the target fragments are annealed thereto. The use of oligonucleotides A and B of the probe in excess, or large amounts of these probe oligonucleotides relative to the amounts of the fragments of the nucleic acid sample is, in addition to minimising the requirement for sample material, to try to ensure that a large fraction (at or close to 100%) of the available target fragments are bound to the probe, which is necessary for equal representation of the target fragments. Equal representation is important in, for example, sequencing applications, for efficient resequencing. Where oligonucleotides A and B are provided as part of a single probe, this helps to ensure that A and B oligonucleotides are hybridised together without having to add excess oligonucleotides.

The target-specific nucleic acid probe, and hence oligonucleotides A and B, are generally comprised of DNA. However, also included are probes made up of or including ribonucleotides or synthetic or modified nucleotide residues that are capable of participating in Watson-Crick type or analogous base-pair interactions. Thus, also included are probes comprised of DNA analogues or modified DNA, e.g. PNA or other derivatives including non-nucleotide backbones.

Thus, the partially double-stranded probe used in the method of the invention is comprised of two hybridised oligonucleotides, or two parts of a single, partially self-complementary oligonucleotide, A and B. Oligonucleotide A acts to template the ligation of the target fragment to oligonucleotide B and therefore comprises a first target-specific part and a second non-target-specific part. By "first target-specific part" is meant a part of oligonucleotide A of at least 10 nucleotides in length which is complementary in sequence to at least part of the single-stranded portion of the target fragment. The reference in step (c)(i) to "at one end" is meant in a general sense, i.e. that the first target-specific part and the second non-target-specific part are generally at opposite ends of oligonucleotide A, and does not require that the first target-specific part is at the very terminus of oligonucleotide A. "Complementary" as used here is as defined above, i.e. functionally complementary (capable of mediating

IDT Exhibit 1020.1241

hybridisation). Hence, "complementary" refers to said part as a whole rather than the individual nucleotides and may not necessarily mean 100% complementarity between the first target-specific part of oligonucleotide A and the part of the single-stranded portion of the target fragment to which it hybridises. However, the furthest-apart target fragment-complementary nucleotides of said part must delineate a stretch of at least 10 nucleotides. Thus, the first target-specific part may be 10 nucleotides long or of any length above 10 nucleotides such as, e.g., 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 50, 100 or any integer between or above these, providing that the length is sufficient to mediate a productive hybridisation, i.e. a hybridisation sufficient to template a ligation between the target fragment and oligonucleotide B. Such a productive hybridisation does not necessarily require that the terminal nucleotides of the relevant end of oligonucleotide A are comprised within the target fragment-complementary stretch of nucleotides (first target-specific part), and therefore oligonucleotide A may contain one or more terminal nucleotides at that end which are not complementary to the target fragment.

The non-target-specific part of oligonucleotide A will typically be of at least 20 nucleotides in length in order to facilitate a stable, productive hybridisation with oligonucleotide B. If the probe is a single, partially self-complementary oligonucleotide, the intramolecular hybridisation may permit a shorter non-target-complementary region of oligonucleotide A, though in such a case the ligase required in the subsequent ligation step will require a double-stranded region of about 6-10 nucleotides. The length of the non-target-specific part of oligonucleotide A will represent the difference between the length of the target-specific part and a maximum length of oligonucleotide A of about 200 nucleotides. Preferably, the first target-specific part and the second non-target-specific part of oligonucleotide A each consist of 20, 25 or 30 or more nucleotides. Oligonucleotide A may in total be, for example, 16 to 200 nucleotides long, more particularly, 18, 20, 30, 40, 50 or 60 to 100, 120, 150 or 200. A representative length range may yet more particularly be 16, 18 or 20 to 40, 50, 60, 70, 80 or 100, or at the higher end, 30, 40 or 50 to 60, 70, 80, 90, 100, 120, 150 or 200.

Oligonucleotide B must be at least of a length sufficient to maintain a productive hybridisation with at least part of the non-target-specific part of oligonucleotide A, but as discussed below may be longer than this and may overhang oligonucleotide A at the other end of the probe from the single-stranded target-specific part. Thus typically, oligonucleotide B will be at least 20 nucleotides long, but may be less, e.g. at least 6, 8 or 10 nucleotides long in the case of a single partially

IDT Exhibit 1020.1242

self-complementary probe. More particularly, a representative oligonucleotide B may thus be 6, 8, or 10 to 20, 30, 40, 50, 70, 80, 100, 120, 150, or 180 nucleotides long.

It will be noted that a stretch of at least 10 nucleotides (representing the first, target-specific, part of oligonucleotide A) is longer than the single-stranded "sticky" overhangs produced by restriction endonucleases. Thus, as used herein "target-specific" indicates a specificity for the target fragment which is based on a stretch of at least 10 nucleotides of the fragment. This means the annealing (hybridisation) of the probe (or more particularly oligonucleotide A of the probe) to the target fragment is dependent on (or contingent on or dictated by) the sequence of the target fragment. The sticky end hybridisation of restriction fragments is not encompassed. The probe is target-specific in the sense of being selective for the target fragment, i.e. capable of hybridising to the target fragment but not to other, non-target, fragments which do not contain a region (or part) of complementarity (in the single stranded portion) to oligonucleotide A of the probe. Thus the probe (or more particularly oligonucleotide A of the probe) discriminates or distinguishes between target and non-target fragments. The person skilled in the art will readily understand that the degree of specificity of the probe for the target fragment may be increased by increasing the length of said first target-specific part. By altering the length of said first target-specific part the "uniqueness" within the nucleic acid sample of the fragment captured by the probe may be varied, and in that way highly similar sequences such as family members or homologues of the target fragment may be included within or excluded from the fragments captured using the probe. Generally, it will be desired that in practice "target-specific" means that said first target-specific part of oligonucleotide A targets a DNA which is unique (albeit possibly present in many copies; a unique species) within the nucleic acid sample. However, as discussed above, it may be desired to use a single probe to detect or enrich for more than one DNA on the basis of shared sequence and in such cases "target-specific" will not mean that the target DNA is unique within the sample. On account of only a single target-specific nucleic acid probe being contacted in step (c) with the fragments of step (b), the method involves only one target-specific probe-binding event per target fragment. More particularly there is only one target-specific probe-binding (probe-hybridisation) event per target fragment before ligation step (d). Put another way, the method does not encompass the use of two or more target-specific probes per target fragment, nor the use of a probe which binds the target fragment more than once in a target-specific manner, in other words at more than one site, and in particular before ligation step (d).

IDT Exhibit 1020.1243

The second non-target-specific part of oligonucleotide A comprises a nucleotide sequence complementary to at least a portion, including one end, of oligonucleotide B of the probe. Hence, the end of oligonucleotide A other than the end containing the first target-specific part hybridises to at least one end of oligonucleotide B. If oligonucleotide B contains a portion non-complementary to oligonucleotide A, this is located at the distal end of oligonucleotide B relative to the first target-specific part of oligonucleotide A. The region of complementarity may be of any length and degree of complementarity capable of mediating a productive hybridisation, i.e. a hybridisation sufficient to template a ligation between oligonucleotide B and the target fragment.

As a result of the templating action of oligonucleotide A which hybridises to oligonucleotide B and the target fragment, oligonucleotide B may be ligated to the target fragment. Oligonucleotide B therefore comprises at least a portion which is complementary in sequence to the second non-target-specific part of oligonucleotide A such that together oligonucleotides A and B form a partially double-stranded nucleic acid probe comprising at one end a single-stranded target-specific part as well as a double-stranded non-target-specific part. In cases where the probe (or more particularly oligonucleotide A) hybridises to the end of the target fragment, the probe may be designed such that, upon hybridisation, oligonucleotide B is positioned immediately adjacent to the end nucleotide of the target fragment. However, as discussed further below this is not an absolute requirement. The end of the probe other than the single-stranded target-specific part may be single- or double-stranded (or a hairpin or hairpin loop, if oligonucleotides A and B are parts of a single oligonucleotide). Thus, reference in step (c)(i) and (ii) above to "at least a portion" of oligonucleotide B means that oligonucleotide B may be longer than the portion of oligonucleotide B complementary to oligonucleotide A, in which case the "overhanging" (protruding) portion non-complementary to oligonucleotide A is at the end of oligonucleotide B other than the end which becomes ligated to the target fragment. Hence, reference in step (c)(i) and (ii) to "including one end" means that the portion of oligonucleotide B which is complementary to the non-target-specific part of oligonucleotide A includes one or, if oligonucleotide B is entirely complementary to (no longer than) the non-target-specific part of oligonucleotide A, both ends of oligonucleotide B. The ligation of oligonucleotide B to a target fragment is the only target-specific probe-binding event that will occur (and more particularly, that will occur up to or before ligation step (d)) for a given target fragment in the method of the invention.

IDT Exhibit 1020.1244

If, as discussed above, oligonucleotide B is longer than the portion of oligonucleotide B complementary to oligonucleotide A, and the overhanging portion is at the 3' end of oligonucleotide B, said 3' end could serve as a primer for rolling circle amplification of a circularised or circularisable nucleic acid molecule hybridised to said 3' end. The nucleic acid may be pre-hybridised to oligonucleotide B or may be added separately, and may be pre-circularised or circularised by ligation as a result of a ligation reaction mediated by hybridisation to oligonucleotide B. Upon addition of DNA polymerisation reagents, an RCA product will be generated from the circularised template, said product being continuous with the probe-target fragment hybrid. Preferably, the polymerase of said polymerisation reagents is phi29 polymerase. Hence, in an *in situ* embodiment where the target nucleic acid fragment is immobilised, the generation of an RCA product may serve to detect the target nucleic acid. The RCA product could be visualised *in situ* by hybridisation of fluorescently-labelled oligonucleotides, for example.

Oligonucleotide B may contain or carry an element by which the target fragment may be detected or enriched. By "contains or carries" is meant that such an element may be contained within the nucleotide sequence of the oligonucleotide, e.g. a probe or primer binding site or other nucleic acid-based affinity-binding site (for example a binding site for a hybridisation probe or for a DNA binding protein etc., which binding site may be viewed as a capture or detection element depending on the nature of the probe or affinity binding element, or a binding site for a sequencing primer, which sequencing primer binding site may be accordingly be viewed as a detection element, or for an amplification primer, which amplification primer binding site may accordingly be viewed as an amplification element), or it may be attached or conjugated or in any way linked or coupled to or associated with oligonucleotide B. For example, it may be a functional moiety (e.g. a chemical group or a molecule) which is attached etc. to the oligonucleotide, such as an immobilisation moiety or a detection moiety (e.g. a reporter or a label). An immobilisation moiety may for example be an affinity binding moiety or group, e.g. one member of an affinity binding pair (i.e. an affinity ligand), which is attached or conjugated etc. to said oligonucleotide, and is capable of binding to the other member of the affinity binding pair (i.e. its cognate binding partner) for the purposes of capture or separation, e.g. when the cognate binding partner is attached to a solid phase.

Oligonucleotide B may contain one or a plurality of such detection or enrichment (e.g. amplification and/or capture) elements.

A detection element may as noted above be a binding site contained in the oligonucleotide sequence (e.g. a binding site for a detection probe or moiety or for a

IDT Exhibit 1020.1245

primer to be used in a detection reaction, e.g. a sequencing primer) or it may be a detection moiety which is carried in any way by the oligonucleotide, e.g. a reporter group or moiety or a label, which may be directly or indirectly signal-giving. For example it may a visualisable label, such as a coloured or fluorescent or particulate label, or a moiety which contributes to or takes part in a signal-giving reaction, e.g. an affinity binding partner or ligand or a substrate or co-factor for an enzyme.

An enrichment element may be any element for the amplification and/or capture of the target fragment, or indeed for enrichment of the target fragment by any means.

As can be seen from the discussion above an "amplification element" may be any feature of, or attached to, oligonucleotide B which may be used to amplify the target fragment of the probe-target fragment hybrid. Typically it will be an amplification primer binding site. Such an amplification primer binding site may be a binding site for a primer for single-sided amplification or polymerisation runoff, e.g. using T7 RNA polymerase primer to repeatedly prime transcription resulting in amplification (Van Gelder RN et al, Proc Natl Acad Sci U S A. 1990 Mar;87(5):1663-7) or a rolling circle amplification primer (see further below). It may also be a binding site for one of a number or set (e.g. pair) of amplification primers, for example to allow exponential amplification, e.g. a PCR primer or a primer for a PCR-based procedure. As described further below, if more than one amplification primer is used one or more further primer binding sites may be provided to the target fragment in a separate step. The primer binding site may also be used for the binding of a sequencing primer, or a sequencing primer binding site may be located elsewhere in oligonucleotide B.

A "capture element" is any moiety carried by (e.g. attached or conjugated to etc.) oligonucleotide B, or any feature of the sequence of oligonucleotide B (e.g. a binding site), which may potentially be used selectively to attach a probe-bound target fragment (probe-target fragment hybrid) to a solid phase or support, including for example a particle such as a bead. Hence, a capture element may be viewed as an "immobilisation element". Numerous examples of such elements are known in the art and include, e.g., an affinity binding partner, e.g. biotin or a hapten, capable of binding to its binding partner, i.e. a cognate binding partner, e.g. streptavidin or avidin, or an antibody, provided on the solid phase or support. Said interaction between oligonucleotide B and a solid phase may particularly be mediated by click chemistry (Kolb HC et al, Angew Chem Int Ed Engl. 2001 Jun 1;40(11):2004-2021).

The solid phase may be any of the well known supports or matrices which are currently widely used or proposed for immobilisation, separation etc. These may

IDT Exhibit 1020.1246

take the form of particles (e.g. beads which may be magnetic or non-magnetic), sheets, gels, filters, membranes, fibres, capillaries, or microtitre strips, tubes, plates or wells etc. The support may be made of glass, silica, latex or a polymeric material. Suitable are materials presenting a high surface area for binding of the analyte. Such supports may have an irregular surface and may be e.g. porous or particulate e.g. particles, fibres, webs, sinters or sieves. Particulate materials e.g. beads are useful due to their greater binding capacity, particularly polymeric beads. Conveniently, a particulate solid support used according to the invention will comprise spherical beads. The size of the beads is not critical, but they may e.g. be of the order of diameter of at least 1 and preferably at least 2 μm, and have a maximum diameter of preferably not more than 10, and e.g. not more than 6 μm. Monodisperse particles, that is those which are substantially uniform in size (e.g. size having a diameter standard deviation of less than 5%) have the advantage that they provide very uniform reproducibility of reaction. Representative monodisperse polymer particles may be produced by the technique described in US-A-4336173. However, to aid manipulation and separation, magnetic beads are advantageous. The term "magnetic" as used herein means that the support is capable of having a magnetic moment imparted to it when placed in a magnetic field, and thus is displaceable under the action of that field. In other words, a support comprising magnetic particles may readily be removed by magnetic aggregation, which provides a quick, simple and efficient way of separating the particles following the analyte binding steps. Particularly advantageous solid phases include very small particles which can efficiently contact a high proportion of the immobilisable oligonucleotide Bs. Such particles may further be useful by retarding the movement of particle-attached target fragments through a gel, allowing separation from free, non-particle-attached (non-target) fragments. Alternatively, also preferred is the use of a chromatographic matrix modified with groups that can be reacted covalently or non-covalently with groups on oligonucleotide B of the probe.

As noted above, enrichment and/or detection of the probe-target fragment hybrid can occur by any means selective for the hybrid. As described further below, in certain embodiments of the invention the probe-target fragment hybrid may be circularised. In such an embodiment enrichment for the hybrid may take place by enriching for circular molecules, e.g. by degrading any non-circular (i.e. linear) nucleic acid molecules using an exonuclease.

Oligonucleotide B may further contain a recognition sequence for a restriction endonuclease such that the partially double-stranded, intact probe formed from the hybridisation of oligonucleotides A and B is endonucleolytically cleavable. In

IDT Exhibit 1020.1247

particular, the recognition sequence may be for a rare, infrequently-cutting endonuclease. Such a cleavage site may be useful in releasing immobilised probe-target fragment hybrids from the solid phase.

Alternatively or additionally, oligonucleotide B may contain a "molecular tag", i.e. a feature which allows an oligonucleotide B used in the method of the invention as performed on a particular sample to be distinguished from an oligonucleotide B used in the method as performed on a different sample and which thereby allows identification of the sample from which a given fragment has been enriched or detected. The molecular tag may be contained in a portion of oligonucleotide B non-complementary to oligonucleotide A. The samples may, for example, correspond to patient samples. In such an embodiment a different oligonucleotide A would be needed for each target nucleic acid, whilst the oligonucleotides B would need to differ only between the method as performed on different samples.

As discussed above, the method of the invention may be performed in multiplex. In such an embodiment, the respective oligonucleotide Bs of the plurality of nucleic acid probes comprise a common sequence which is the same in each probe. Thus, in the context of oligonucleotide B of the probe, "common" refers to a sequence of oligonucleotide B, which may comprise all of oligonucleotide B, which is the same (generic) amongst the plurality of probes used in a multiplex reaction. In particular, the common sequence in each probe may comprise the detection and/or enrichment element (e.g. amplification and/or capture element), whereby the plurality of different target fragments may be amplified and/or captured together, and/or the restriction endonuclease recognition sequence. More particularly, the plurality of nucleic acid probes in a multiplex aspect of the method comprise the same oligonucleotide B.

Once the target fragment is hybridised at one end, and oligonucleotide B is hybridised at least at one end, to oligonucleotide A of the probe, the respective ends of the target fragment and oligonucleotide B may be ligated to produce a probe-target fragment hybrid. As mentioned above, and discussed further below, a target fragment "end" for ligation may be created by hybridising a probe (or oligonucleotide A of a probe) internally in the target fragment in such a manner that a substrate for a flap endonuclease is formed, and cleaving such a structure with a flap endonuclease. The probe-target fragment hybrid comprises the target fragment ligated to the probe.

Enzymes appropriate for the ligation step are known in the art and include, e.g. Tth DNA ligase, Taq DNA ligase, *Thermococcus* sp. (strain 9°N) DNA ligase (9°N™ DNA ligase, New England Biolabs), Ampligase™ (Epicentre Biotechnologies) and T4 DNA ligase. The ligation of oligonucleotide B and the target fragment may

IDT Exhibit 1020.1248

occur directly, i.e. the respective ends are immediately juxtaposed on hybridisation to oligonucleotide A, or indirectly, in which case the hybridised ends are separated by a gap. In the latter case, the gap may be filled by an added oligonucleotide complementary to the region of oligonucleotide A in the gap (i.e. a "gap" oligonucleotide), necessitating two ligation reactions, or alternatively the gap may be filled by polymerase extension of one of the ends until the extended end meets the other end, the immediately juxtaposed ends then being ligated together. Polymerases suitable to perform such a "gap-filling" reaction are known in the art.

In aspects of the method in which the fragments of the nucleic acid sample are rendered partially (incompletely) single-stranded, after the step of ligating the probe to the target fragment a polymerase optionally may be added to the reaction to repair (fill-in) any gap between the target fragment-hybridised end of oligonucleotide A of the probe and the exonucleolytically-degraded end of the corresponding strand of the target fragment. The polymerase may extend said end of oligonucleotide A or said end of the corresponding strand of the target fragment, whichever of these ends is a 3' end. The polymerase may be a strand-displacing or a non-strand-displacing polymerase. In the former case the polymerase will, after filling said gap, continue to the end of the template by displacing (and "replacing") oligonucleotide A or the corresponding strand of the target fragment as the case may be (i.e. depending on whether the polymerase is extending from the end of oligonucleotide A or from the end of the corresponding strand of the target fragment). Such strand-displacing polymerases are known in the art and include, e.g., phi29 DNA polymerase. Non-strand-displacing polymerases may also be used and are known to include, e.g., T7 DNA polymerase. Such a polymerase will stop once it has repaired the gap and encounters double-stranded nucleic acid. The extended end may then be ligated to the end of, as the case may be, oligonucleotide A of the probe or the corresponding strand of the target fragment. In this way the double-strandedness of the probe-target fragment hybrid is substantially restored. After the ligation, and optional gap-filling, steps, unbound probe may optionally be removed from the reaction by a size separation step.

As discussed above, the probe-target fragment hybrid may be detected or enriched for by any means selective for the hybrid, for example by degradation of nucleic acid other than the hybrid or by means of a detection and/or enrichment element in oligonucleotide B. As also discussed above, such an enrichment element may be a capture, or immobilisation, element, which may be any moiety attached to, or stretch of the nucleotides of, oligonucleotide B capable of interacting with an element on a solid phase or support.

IDT Exhibit 1020.1249

By means of the capture element the probe-target fragment hybrid may be separated (enriched) from the non-target fragments for, by example, washing or gel purification. As discussed above, the covalent nature of the ligation between the probe and the target fragment permits the use of highly stringent washes effective to remove fragments which are not ligated to the probe. Such washes are not possible in known methods in which the probe and target fragment are annealed through hybridisation, and therefore the ability to use such high stringency conditions in the present method is advantageous. Alternatively, if particles are used, another option is size selection such as gel purification.

The method of the invention may be performed with a probe having an oligonucleotide B which is immobilisable, i.e. carrying or containing an immobilisation (capture) element, or which is immobilised, i.e. wherein said immobilisation element-carrying oligonucleotide B is already bound to a solid phase. In such a latter embodiment, the step of enriching constitutes washing the immobilised probe-target fragment hybrid to remove non-target fragments.

A target fragment captured, or immobilised, as described above may directly be subjected to analytical techniques such as sequencing, or may instead firstly be amplified. Alternatively, the target fragment may not be captured/immobilised but may instead simply be amplified and/or sequenced. Thus, amplification of the target fragment may be by any suitable method for amplifying nucleic acids from a single amplification primer binding site (i.e. in oligonucleotide B), such as by the use of a T7 RNA polymerase primer, as discussed above. Sequencing reactions may be primed from such an amplification primer binding site, if present, or a separate sequencing primer binding site within oligonucleotide B. As used herein, reference to sequencing the target fragment means sequencing (i.e. determining the identity thereof) at least two consecutive nucleotides in the target fragment. As described further below, amplification may also be performed by methods which involve more than one amplification primer. A binding site for such a second or further primer may be provided to the target fragment in a separate step, as described below. Such a further amplification primer binding site will not be target fragment-specific, or will not be provided to the target fragment in a target-specific manner.

It may be desirable to release the immobilised, enriched fragments from the solid phase to facilitate their analysis. This may be achieved by cleaving in the probe of the probe-target fragment hybrid by, e.g., a restriction endonuclease having a recognition site in oligonucleotide B of the probe, as discussed above. Preferably, the restriction endonuclease is a rare-cutting endonuclease. If the immobilised probe-target fragment hybrid has become single-stranded, it may be necessary to add an

IDT Exhibit 1020.1250

oligonucleotide to render the appropriate part of the probe double-stranded prior to cleaving. Released target fragments may be then be subject to analytical techniques, such as those discussed above. Such techniques may be performed directly on the released fragments, or on amplification products thereof. In the latter case, the released fragments may be inter- or intramolecularly ligated to give linear concatemers or double-stranded circular molecules. This may require modification of the ends of the released fragments, e.g. by restriction digestion and optionally enzymatic blunting using e.g. Klenow fragment of DNA polymerase I that fills in 5' overhangs by polymerase activity and degrades 3' overhangs with exonucleolytic activity, thereby converting sticky ends to blunt ends. Circular molecules may be amplified by rolling circle amplification (RCA). For sequencing, sequences within oligonucleotide B of the probe, may be used to bind sequencing primers.

Numerous variations of the general method of the invention are contemplated, several of which are described below.

After step (a) above but before step (b) an optional further step may be performed in which a common nucleic acid adaptor is non-target-specifically annealed to the ends of the fragments. In this context, "common" means that the same adaptor is annealed to all fragments (or to a number of fragments - see further below - but in any case to more fragments than solely the target fragment(s)) within a given sample and is therefore common, or generic, to the fragments, or to a sub-population of fragments, of that sample, in contrast to the probe which anneals only to the target fragment. In cases where more than one enzyme, e.g. restriction endonucleases, have been used to fragment the sample, more than one "species" of adaptor may be used, wherein such adaptors are identical except for one "annealing end" in each case which differs to "match" the differing fragment ends produced by the enzymes.

The nucleic acid adaptors of the invention are substantially double-stranded molecules. However, by "double-stranded" in the present context does not mean (but does encompass) that said ends are "blunt", and "sticky" ends consisting of a single-stranded overhang of 1, 2, 3, 4, 5 or 6 etc. nucleotides generated by restriction enzyme digestion are included in this meaning. The adaptors may be of any suitable length.

By "non-target-specifically" is meant that said adaptor is annealed by a means which has no specificity for the target fragment in the sense of "target-specific" as defined elsewhere herein. In other words the adaptor is not selective for the target, or for non-target fragments, i.e. it does not discriminate or distinguish between target or non-target fragments. Thus, the binding (hybridisation) of the

IDT Exhibit 1020.1251

adaptor is not dependent or contingent upon a nucleotide sequence which is particular to (or specific for) the target fragment relative to the non-target fragments. As a result, such an adaptor is annealed to the ends of all of the fragments (including the target fragment). This is unlike the annealing of the probe of step (c) which is targeted to the target fragment by virtue of the complementary sequence of prescribed length present in oligonucleotide A of the probe. Such non-target specific annealing may be achieved by any method which facilitates the annealing of the adaptor to all of the fragments, such as through annealing with sticky restriction ends (if present) of the fragments, or blunt- or artificially-created sticky-end annealing. Hence, the adaptors must be designed with at least one end suitable for such annealing. For example, sticky ends of the fragments may be blunted by the use of a suitable polymerase, and single adenosines added as 3' overhangs, again through the action of an appropriate polymerase. Examples of suitable polymerases for either purpose are known in the art and include, in the former case, T4 DNA polymerase, and in the latter case Taq polymerase. If the adaptor is designed with a 5' thymidine overhang, the adenosine and thymidine overhangs will act as complementary sticky ends susceptible to ligation (known as "TA ligation"), promoting ligation of adaptors to fragments .The adaptors may be modified at one end to reduce or prevent ligation of adaptors with each other by any means which prevent ligation but do not otherwise interfere with the assay. Such modifications include the absence of 5' phosphate or 3'OH groups, addition of fluorophores or affinity groups, and the introduction of "blocking" groups such as amine groups.

Importantly, whilst said adaptor may become annealed to both ends of the fragments, it becomes ligated only at the 3' ends, or only at the 5' ends, of the strands of the fragments (thus, "annealed" is meant as defined above, i.e. it does not include ligation). This is in order that it can be ensured that, for the target fragment, an adaptor is ligated to the same strand to which oligonucleotide B will become ligated in step (d), but at the other end of that strand. Selective ligation of the adaptor to only the 3' ends, or only the 5' ends, of the strands of the fragments may be achieved by any suitable means. For example, ligation of the annealed adaptor to either the 3' or 5' ends of the strands may be controlled by dephosphorylation of the fragments or the adaptors, respectively. Suitable dephosphorylases (phosphatases) are known in the art and include, e.g., Antarctic phosphatase (New England Biolabs), shrimp alkaline phosphatase and calf intestinal phosphatase. Ligation of the fragments and adaptors may be achieved using suitable ligases as known in the art, such as T4 DNA ligase. The addition of single adenosines as 3' overhangs to the

IDT Exhibit 1020.1252

blunted fragments, and/or dephosphorylating said fragments, is also beneficial in reducing or avoiding inter-fragment ligation.

In one embodiment, therefore, the method of the invention further comprises, between steps (a) and (b), the step of non-target-specifically annealing to the ends of said fragments a common (or generic) nucleic acid adaptor, wherein the annealed adaptor becomes ligated to the fragments only at the 3' ends, or only at the 5' ends, of the strands of the fragments. As mentioned above, a single species of adaptor may be used, or one or more adaptors may be used, depending on the fragmentation procedure. Thus, the adaptor is annealed such that an adaptor is ligated to one end of the strand of the target fragment to which, at the other end, oligonucleotide B of the probe is ligated in step (d). In other words, in the target fragment the adaptor is ligated at the opposite end of the strand to which the target-specific probe becomes ligated.

In a further embodiment, the ends of the fragments are rendered sticky with respect to at least one end of the adaptor prior to ligation. In a still further aspect, dephosphorylation of the fragments or the adaptor is used to facilitate ligation of the adaptor to, respectively, either the 3' ends or the 5' ends of the strands of the fragments.

In embodiments of the method of the invention in which the adaptor-ligated target fragments are rendered completely single-stranded in step (b), the method must be designed so that the specificity of the subsequently-contacted probe is for a part of the target fragment which results in ligation of oligonucleotide B of the probe to the other end of the target fragment relative to the end to which the adaptor is ligated. Hence in such an aspect, the first target-specific part of oligonucleotide A comprises at least 10 nucleotides complementary in sequence to a portion at the end of said single-stranded target fragment other than the end to which the common nucleic acid adaptor is ligated.

The adaptor, which if used in the method of the invention will become comprised in the probe-target fragment hybrid, may contain an element (advantageously a sequence) useful for the enrichment and/or detection of the target nucleic acid. In particular, such an element or sequence will be used in conjunction with an element present in the probe. It will thus be appreciated that such an element may be a binding site for one of an amplification primer pair (or number or set of primers), the binding site(s) for the other(s) of which are provided in the probe. Accordingly, in one embodiment, the common nucleic acid adaptor comprises an element for amplification, preferably an amplification primer binding site. For example, the adaptor may comprise an amplification primer binding site orientated

IDT Exhibit 1020.1253

appropriately with respect to another such binding site in the probe such that an amplification product may be generated when the target fragment is ligated to both the probe and the adaptor. Such amplification may be by, e.g., the polymerase chain reaction (PCR), of which many modified versions (e.g. "real-time" or quantitative PCR) are well known in the art, or other amplification methods. The amplification product may be analysed by, e.g., massive parallel sequencing platforms (e.g. SOLiD (Applied Biosystems, Inc.), Illumina Genome Analyzer (Illumina, Inc.), Genome Sequencer (454 Life Sciences)) microarrays or hybridisation-based sequencing.

Said amplification primer binding site, or a separate sequence located elsewhere in the adaptor, may be used as a binding site for a sequencing primer for sequencing the enriched target nucleic acid.

Additionally or alternatively to the presence of an amplification element, the common nucleic acid adaptor may carry a molecular tag. By "molecular tag" is meant a feature which can be used to differentiate between otherwise identical adaptors. Put another way, such a tag is an element which may be used to generate distinguishable variants of a single adaptor. The tag may be a particular stretch of the adaptor's nucleotide sequence.

As discussed above, the nucleic acid sample on which the method of the invention is performed may be comprised of a plurality of pooled nucleic acid samples of different origins. Such pooling to give a single nucleic acid sample may occur after fragmenting step (a). In such an aspect, adaptors carrying different molecular tags may be ligated to the fragments of the respective nucleic acid samples prior to pooling, such that in the pooled nucleic acid sample subjected to the remaining steps of the method the fragments originating from different samples will be ligated to adaptors distinguishable by virtue of the different molecular tags. This is highly advantageous as it allows the simultaneous (parallel) analysis, in a simplex (one probe, for one or a plurality of sufficiently similar target nucleic acids) or multiplex (a plurality of probes, for a plurality of target nucleic acids) format, of multiple samples. For example, one or more target nucleic acids may be detected or enriched for in a single assay, from a plurality of differently-tagged patient samples. The use of molecularly-tagged common nucleic acid adaptors to identify the fragments of the respective samples, instead of incorporating tags in the target-specific probes, vastly reduces the number of different probes which would need to be synthesised in order to process multiple samples (containing a plurality of target nucleic acids) in parallel.

In one embodiment, therefore, in step (a) said fragmenting a nucleic acid sample comprises separately fragmenting a plurality of nucleic acid samples, ligating

IDT Exhibit 1020.1254

to the fragments of which samples, respectively, common (or generic) nucleic acid adaptors carrying different molecular tags, and pooling said adaptor-ligated fragmented nucleic acid samples prior to step (b).

The adaptor may further comprise a recognition site for an endonuclease.

In another aspect of the method of the invention, the hybridisation and ligation steps (c) and (d) occur through resolution of a generated secondary structure using a flap endonuclease (FEN). In order to generate the secondary structure which is the substrate for the flap endonuclease, the first target-specific part of oligonucleotide A of the probe must be complementary to a portion internal to the target fragment, rather than to an end portion. Hybridisation of the probe (comprising oligonucleotide A hybridised to oligonucleotide B) to the target fragment causes the non-hybridised 5' end of the target fragment to protrude. The resulting secondary structure is recognised by a flap endonuclease which cleaves off the protruding non-hybridised 5' end of the target fragment to reveal a new 5' end which is hybridised to oligonucleotide A and is ligatable to the oligonucleotide B. Thus, the part of the single-stranded portion of the target fragment which is ligated to oligonucleotide B of the probe (i.e. the part that is hybridised to oligonucleotide A of the probe as described in step (d)) is said ligatable 5' end, and the probe-target fragment hybrid will lack a portion at the 5' end of the original, pre-probe-bound target fragment corresponding to said protruding end.

In such an embodiment, therefore, the first target-specific part of oligonucleotide A comprises at least 10 nucleotides complementary in sequence to an internal non-end portion of said single-stranded target fragment, and said annealing of said probe to said target fragment is through hybridisation of said internal non-end portion to the first target-specific part of oligonucleotide A causing the 5' end of the target fragment to form a substrate for flap endonucleolytic cleavage and further comprising cleaving said flap endonucleolytic cleavage substrate to produce a ligatable 5' end of said target fragment which is hybridised to oligonucleotide A of said probe.

As such a use of a flap endonuclease results in internal cleavage of the target fragment to generate a 5' end for ligation to oligonucleotide B of the probe, if a common nucleic acid adaptor is used said adaptor must not be ligated only to the 5' ends of the strands of the fragments. Thus, the adaptor may be ligated to both the 5' and 3' ends of the strands of the fragments (in which case the 5' end adaptor will be cleaved off during the flap cleavage step) or may be ligated to only the 3' ends of the strands of the fragments (which can be achieved by, for example, dephosphorylation of the fragments as described above).

IDT Exhibit 1020.1255

Hence in a further embodiment, if a common (or generic) nucleic acid adaptor is non-target-specifically annealed to the ends of the fragments between steps (a) and (b), said annealed adaptor becomes ligated to the fragments only at the 3' ends, or at both the 3' and 5' ends, of the strands of the fragments, more particularly such that an adaptor is ligated to at least the 3' end of the strand of the target fragment to which, at the 5' end, oligonucleotide B of the probe is ligated in step (d).

In another aspect of the method of the invention, the probe-target fragment hybrid is circularised. In such a case enriching may involve increasing the amount of circular molecules relative to the amount of linear molecules. Thus, following the ligation step (d), an end of at least the target fragment, being the end to which the probe is not ligated (which end may in certain embodiments comprise a common nucleic acid adaptor) is rendered double-stranded in the event that said end is not double-stranded following the step (b) of rendering the fragments at least partially single-stranded (e.g. in aspects of the method wherein the fragments were made completely single-stranded). "Double-stranded" in the present context does not mean (but does encompass) that said ends are made "blunt", and "sticky" ends formed by a single-stranded overhang of 1, 2, 3, 4, 5 or 6 etc. nucleotides generated by restriction enzyme digestion are included in this meaning. Such rendering double-stranded may be achieved by any suitable method, e.g. by polymerisation from a hybridised probe 3' end or from added oligonucleotides (e.g. hexamers) annealed to the target fragment or, if present, adaptor. In embodiments in which an adaptor is ligated to the fragments, the end (comprising said adaptor) may be made double-stranded by adding an oligonucleotide complementary to the adaptor. In embodiments wherein the fragments are rendered only partially single-stranded, the ends will typically already be double-stranded and no such action is required.

Said double-stranded end of the target fragment is then non-target-specifically annealed intramolecularly with the free, non-target bound end of the probe at the other end of the probe-target fragment hybrid. As the remaining, non-target, fragments will lack ligated probe and therefore be single-stranded at the other end, such intramolecular annealing will not occur. Thus, the probe-target fragment hybrid, optionally also containing a common nucleic acid adaptor, will intramolecularly adopt a circular conformation. In this context, by "non-target-specifically" is meant that said annealing is by means which do not take advantage of or rely on (i.e. are not dependent upon) the particular nucleotide sequence of the target fragment in the sense that no knowledge of the sequence of the target fragment is required in order to effect such annealing. Hence, the means of

IDT Exhibit 1020.1256

annealing may differ according to whether or not the relevant end of the target fragment comprises an adaptor. If an adaptor is present, such annealing will include that between sticky restriction ends, as no sequence of the target fragment is required to exploit a restriction recognition sequence within the known sequence of the adaptor. However, if no adaptor is present, such annealing is limited to that between two blunt ends or two "artificial" sticky ends (such as used in TA-ligation, discussed above). In such embodiments the probe is designed to have, at the free, non-target fragment-bound end, a blunt or complementary sticky end, as appropriate, or contains a restriction endonuclease recognition sequence allowing the creation, through cleavage, of such an end. After the intramolecular-annealing of the probe-target fragment hybrid, oligonucleotide B of the probe is ligated with the corresponding strand of the double-stranded end resulting in circularisation of the hybrid.

Hence, in a further embodiment the method further comprises, between steps (d) and (e), rendering double-stranded an end of at least the target fragment, being the end to which the probe is not ligated, which end will comprise a common nucleic acid adaptor sequence in the case that such an adaptor is ligated to the fragments between steps (a) and (b); non-target-specifically annealing said double-stranded end intramolecularly with the free, non-target fragment-bound end of the probe; and ligating oligonucleotide B of the probe-target fragment hybrid with the corresponding strand of said double-stranded end to circularise said hybrid.

The selective circularisation of the target fragment in this aspect of the method can be exploited in the enriching and/or detecting step. In one embodiment, therefore, said enriching and/or detecting is by means that increases the ratio of circular to linear nucleic acids. Such enrichment or detection of circular molecules may be achieved by any suitable means, which include methods to selectively amplify circular nucleic acids, e.g., rolling circle replication (RCA), hyperbranched RCA, multiple strand displacement. Furthermore, enrichment of circular molecules may be achieved by methods to selectively remove non-circular nucleic acids, e.g. exonucleolysis. Suitable exonucleases are known, and include exonuclease I, exonuclease III, lambda exonuclease.

A number of representative preferred specific embodiments of the invention are enumerated below.

In a first such embodiment, the method of the invention comprises:

(a) fragmenting a nucleic acid sample to generate nucleic acid fragments including a target fragment containing said target nucleic acid;

IDT Exhibit 1020.1257

(b) non-target-specifically annealing to the ends of said fragments a common nucleic acid adaptor, wherein the annealed adaptor becomes ligated to the fragments only at the 3' ends, or only at the 5' ends, of the strands of the fragments;

(c) rendering said fragments, including said target fragment, single-stranded;

(d) contacting the single-stranded fragments of step (c) with oligonucleotides A and B of a single target-specific nucleic acid probe, wherein:

(i) oligonucleotide A is a single-stranded oligonucleotide comprising at one end a first target-specific part comprising at least 10 nucleotides complementary in sequence to a portion at the end of said single-stranded target fragment other than the end to which the common nucleic acid adaptor is ligated in step (b), and comprising at the other end a second non-target-specific part which comprises a nucleotide sequence complementary to at least a portion, including one end, of oligonucleotide B of the probe, and

(ii) oligonucleotide B is a single-stranded oligonucleotide which contains or carries an amplification primer binding site and optionally an element for immobilisation to a solid phase, and of which at least a portion, including one end, is complementary in sequence to the second non-target-specific part of oligonucleotide A,

such that said target fragment becomes annealed to said probe through hybridisation to the first target-specific part of oligonucleotide A resulting in only one target-specific probe-binding event per target fragment ;

(e) ligating oligonucleotide B of said probe to the end of said target fragment which is hybridised to oligonucleotide A of said probe to produce a probe-target fragment hybrid; and

(f) enriching for said probe-target fragment hybrid through amplification by means of the amplification primer binding site in oligonucleotide B of the probe and an amplification primer binding site in the common nucleic acid adaptor, and optionally immobilising said fragment to a solid phase by means of the immobilisation element in oligonucleotide B of the probe.

Preferably, the fragmenting in step (a) is by restriction endonuclease and/or in step (c) the fragments are rendered single-stranded by denaturation.

As discussed above, the method of the invention is advantageous in that the target fragment undergoes only a single target-specific binding event with the probe, necessitating knowledge of only a single, short region of the target fragment. This is an advantage over known methods such as PCT and the Selectors of WO 2005/111236 which require knowledge of the target fragment/nucleic acid sequence

IDT Exhibit 1020.1258

at two regions (which delineate the target nucleic acid). In particular, the method of the invention facilitates the analysis of unknown nucleic acid junctions such as DNA translocation breakpoints, recombination sites and splice junctions. The single-sided ligation of the probe with the target nucleic acid fragment, containing the target nucleic acid, can be thought of as the fragments being selectively "pulled out" of the nucleic acid sample by one end (the end containing known sequence, on the basis of which the probe is designed). Providing the sample has been fragmented appropriately, i.e. to give fragments of sufficient length, the junction of interest will be located in the pulled-out fragment and may be analysed and identified by, for example, nucleic acid sequencing. The method clearly has potential application in the field of medical diagnostics, particularly given the ease with which the method may be multiplexed through the use of target-specific probes having the same enrichment elements in oligonucleotide B, and optionally the use of a common adaptor. Whilst such an adaptor, where used, will be common in terms of all "functional" elements (such as amplification elements or restriction endonuclease recognition sites), as described above distinguishable variants may be generated by adding molecular tags. Such differently-tagged adaptors may respectively be ligated to different fragmented sub-samples of the single (pooled) nucleic acid sample which is subjected to the method. In addition to allowing the rapid preparation of hundreds or thousands of samples in a single probing and enrichment step, this facilitates ease of identification of target nucleic acids enriched from such samples in parallel, which in a medical diagnostic context may be patient samples. The use of such tagged adaptors is advantageous in dramatically reducing the number of individual probes which would need to be designed and synthesised in order for the sample provenance of each targeted nucleic acid to be indicated by use of "tagged" probes. Furthermore, the probe, and if used the adaptor, may be designed to comprise the primers used for sequencing, thereby forming an amplification product that can be sequenced directly with very little additional preparation.

The single-sided ligation feature of the method further enables other useful applications. For example, the fact that a target-specific probe-binding event occurs only at one "end" of the target fragment, with the other end binding non-target-specifically to a common adaptor (where used), means that the method may have use in analysis of degraded DNA which may be present in scarce amounts, such as in histological samples (e.g. formalin-fixed, paraffin-embedded samples). Since the adaptors are non-target-specifically ligated to the target fragments, every probe-bound fragment will also have an adaptor ligated thereto even if the non-probe-bound end has been generated by degradation or e.g. physical, "random" fragmentation. As

IDT Exhibit 1020.1259

a result, the method may facilitate enrichment of nucleic acids from samples not amenable to enrichment by other methods which require target-specific binding at both ends of the target nucleic acid (e.g. PCR or methods relying on intramolecular circularisation). Further, such known methods may require the target fragment to be of a minimum length, for efficient amplification, whereas fragment length is not critical in the method of the invention. Hence, the method could be used on degraded or fragmented nucleic acid with very short average fragments, representing a suitable approach for analysis of degraded DNA in scarce amounts.

As discussed above, a further advantage of the method of the invention is the covalent nature of the connection (ligation) between the target fragment and the probe. The probe becomes directly ligated to the target fragment of the sample (not to an amplification or polymerisation product thereof), allowing the use of stringent separation conditions to effectively remove free probe and non-target fragments and other reagents leading to enrichment relative to prior art methods such as microarray capture and FISH which rely on hybridisation between the probe and the target.

The method, which in most embodiments does not rely on circularisation of the target fragment, is for that reason particularly suitable for enriching for long genomic target fragments, such as fragments over 2kb in length which can be more difficult to circularise due to the lower proximity of fragment ends relative to shorter fragments. This makes the method very suitable for analysing, by e.g. sequencing and/or high resolution polymorphism analysis, genomic regions identified by whole genome association studies to be associated with disease states. The enrichment of long fragments is also useful in another application enabled by the method of the invention, namely haplotype analysis. If the sample is known to contain a heterozygous polymorphism, it may be possible to fragment the sample using one or more enzymes including one which recognises a site inactivated by the polymorphism. By designing probes respectively to target fragments generated by the presence and absence of cleavage at the polymorphic sites, fragments corresponding to each haplotype may separately be enriched allowing analysis/identification of the respective haplotypes. In such an aspect, the generation and enrichment of long fragments is desirable to maximise the amount of haplotype-specific nucleic acid available for analysis.

As the method generally represents a highly specific way to detect and amplify particular nucleic acids from a complex sample, it may be used in preference to known methods for analysing microbes and other exogenic pathogens in complex DNA samples. The probes of the invention are shorter and cheaper to make than padlock probes, known in the art. The foreign nucleic acid could be amplified and

IDT Exhibit 1020.1260

analysed by, e.g., sequencing, in order to characterize the infectious agent for, e.g., diagnostic or drug sensitivity screening purposes.

Unlike, e.g. PCR, the method of the invention is compatible with combinatorial oligonucleotide synthesis, e.g. on microarrays, enabling the production of large probe sets at low cost, and allows analysis single molecules.

In a second embodiment of the method of the invention, said method comprises:

(a) fragmenting a nucleic acid to generate nucleic acid fragments including a target fragment containing said target nucleic acid;

(b) rendering said fragments, including said target fragment, partially single-stranded by 3' or 5' exonuclease digestion, wherein the length of the resulting single-stranded end portion is sufficient to allow hybridisation of at least part of the single-stranded end portion of said target fragment to the probe of step (c);

(c) contacting the partially single-stranded fragments of step (b) with oligonucleotides A and B of a single target-specific nucleic acid probe, wherein:

(i) oligonucleotide A is a single-stranded oligonucleotide comprising at one end a first target-specific part comprising at least 10 nucleotides complementary in sequence to at least part of said single-stranded portion (or more particularly to a portion at the end of said single-stranded portion) of said target fragment, and comprising at the other end a second non-target-specific part which comprises a nucleotide sequence complementary to at least a portion, including one end, of oligonucleotide B of the probe, and

(ii) oligonucleotide B is a single-stranded oligonucleotide which contains or carries an element for immobilisation to a solid phase and optionally an amplification primer binding site, and of which at least a portion, including one end, is complementary in sequence to the second non-target-specific part of oligonucleotide A,

such that said target fragment becomes annealed to said probe through hybridisation to the first target-specific part of oligonucleotide A resulting in only one target-specific probe-binding event per target fragment ;

(d) ligating oligonucleotide B of said probe to the end of said target fragment which is hybridised to oligonucleotide A of said probe to produce a probe-target fragment hybrid; and

(e) enriching for said probe-target fragment hybrid through immobilisation to a solid phase by means of the immobilisation element in oligonucleotide B of the probe, and optionally by amplifying said target fragment by means of the amplification primer binding site in oligonucleotide B of the probe.

IDT Exhibit 1020.1261

Such an embodiment is particularly advantageous in that the fragments remain largely or at least partially double-stranded, i.e. are not rendered completely single-stranded. This avoids to some extent the undesirable cross-reactivity between single-stranded nucleic acid fragments and thereby reduces the incidence of hybrids which must be distinguished between. This is especially for the enrichment of long genomic sequences which are more prone to such cross-reactivity when single-stranded. Applications of the method of the invention involving the enrichment of long fragments are discussed above.

Preferably, the fragmenting in step (a) is by restriction endonuclease. In another preferred aspect there exists between steps (d) and (e) a step of contacting the probe-target fragment hybrid with a strand-displacing or a non-strand-displacing polymerase such that any gap between the target fragment-hybridised end of oligonucleotide A of the probe and the exonucleolytically-degraded end of the corresponding strand of the target fragment is filled, substantially restoring the double-strandedness of the probe-target fragment hybrid. In a further preferred aspect, the ligation step is followed by a size separation step to remove free non-target fragment-bound probes.

In a third representative specific embodiment of the method of the invention, said method comprises:

(a) fragmenting a nucleic acid sample to generate nucleic acid fragments including a target fragment containing said target nucleic acid;

(b) optionally, non-target-specifically annealing to the ends of said fragments a common nucleic acid adaptor, wherein the annealed adaptor becomes ligated to the fragments at the 3' ends or at the 3' and 5' ends of the strands of the fragments such that an adaptor is ligated to at least the 3' end of the strand of the target fragment to which, at the 5' end, oligonucleotide B of the probe is ligated in step (f);

(c) rendering said fragments, including said target fragment, single-stranded;

(d) contacting the single-stranded fragments of step (c) with oligonucleotides A and B of a single target-specific nucleic acid probe, wherein:

(i) oligonucleotide A is a single-stranded oligonucleotide comprising at one end a first target-specific part comprising at least 10 nucleotides complementary in sequence to an internal non-end portion of said single-stranded target fragment, and comprising at the other end a second non-target-specific part which comprises a nucleotide sequence complementary to at least a portion, including one end, of oligonucleotide B of the probe, and

IDT Exhibit 1020.1262

(ii) oligonucleotide B is a single-stranded oligonucleotide which contains or carries at least one element for detection and/or enrichment (or more particularly detection, amplification and/or capture) of said target fragment, and of which at least a portion, including one end, is complementary in sequence to the second non-target-specific part of oligonucleotide A,
such that said target fragment becomes annealed at said internal non-end portion to said probe through hybridisation to the first target-specific part of oligonucleotide A resulting in only one target-specific probe-binding event per target fragment, wherein said hybridisation causes the 5' end of the target fragment to form a substrate for flap endonucleolytic cleavage;

(e) cleaving said flap endonucleolytic cleavage substrate to produce a ligatable 5' end of said target fragment which is hybridised to oligonucleotide A of said probe;

(f) ligating oligonucleotide B of said probe to the part of the single-stranded portion of said target fragment which is hybridised to oligonucleotide A of said probe to produce a probe-target-fragment hybrid;

(g) enriching for said probe-target fragment hybrid by means of said detection and/or enrichment (or more particularly detection, capture and/or amplification) element.

In this embodiment, the use of a flap endonuclease is advantageous in that it avoids the need to identify an enzyme, e.g. a restriction endonuclease, recognition sequence near to the target nucleic acid in the nucleic acid sample. The sample may be fragmented by any one or more enzymes, even if such enzymes are not known to cut in the proximity of the target nucleic acid. Alternatively, non-sequence-specific fragmentation methods, which produce fragment ends of unpredictable sequence, such as physical methods, e.g. sonication or nebulisation, may be used. In such an embodiment the probe is designed to target a region of known sequence of the target nucleic acid which is internal to a fragment, i.e. is not situated at a fragment end as in other embodiments. Because of this, the sequence at the fragment ends is not important and does not need to be known. The flap endonuclease, recognising the secondary structure (protruding (non-hybridised) 5' region of the target fragment hybridised to the probe), cleaves the fragment to create a "new" 5' end which is hybridised to oligonucleotide A of the probe and is ligatable to oligonucleotide B. The method of the invention is therefore not limited the use of sequence-specific fragmentation of the sample, or to instances wherein the region of known sequence of the target nucleic acid contains an endonuclease recognition sequence.

IDT Exhibit 1020.1263

Preferably, said fragmenting in step (a) is by restriction endonucleolysis and/or in step (c) said fragments are rendered single-stranded by denaturation. Also preferred is that oligonucleotide B contains or carries an amplification primer binding site and optionally an element for immobilisation to a solid phase such that in step (g) said enriching is through amplification by means of the amplification primer binding site in oligonucleotide B of the probe or, if a common nucleic acid adaptor is non-target-specifically annealed to the ends of the fragments between in step (b), said amplification is by means of the amplification primer binding site in oligonucleotide B of the probe and an amplification primer binding site in the common nucleic acid adaptor. In a further preferred aspect, between steps (f) and (g) there exists a step of size-selecting the probe-target fragment hybrid by, e.g., gel purification.

In a fourth representative specific embodiment, the method of the invention comprises:

(a) fragmenting a nucleic acid sample to generate nucleic acid fragments including a target fragment containing said target nucleic acid;

(b) optionally, non-target-specifically annealing to the ends of said fragments a common nucleic acid adaptor, wherein the annealed adaptor becomes ligated to the fragments only at the 3' ends, or only at the 5' ends, of the strands of the fragments;

(c) rendering said fragments, including said target fragment, at least partially single-stranded, wherein the single-stranded portion includes an end portion and wherein the length of said single-stranded portion is sufficient to allow hybridisation of at least part of the single-stranded portion of said target fragment to the probe of step (d);

(d) contacting the at least partially single-stranded fragments of step (c) with oligonucleotides A and B of a single target-specific nucleic acid probe, wherein:

(i) oligonucleotide A is a single-stranded oligonucleotide comprising at one end a first target-specific part comprising at least 10 nucleotides complementary in sequence to at least part (or more particularly to a portion at the end) of said single-stranded portion of said target fragment, wherein if a common nucleic acid adaptor is ligated to the fragments according to optional step (b) said first target-specific part comprises at least 10 nucleotides complementary in sequence to a portion at the end of said at least partially single-stranded target fragment other than the end to which the common nucleic acid adaptor is ligated; and comprising at the other end a second non-target-specific part which comprises a nucleotide sequence complementary to at least a portion, including one end, of oligonucleotide B of the probe, and

IDT Exhibit 1020.1264

(ii) oligonucleotide B is a single-stranded oligonucleotide which contains or carries at least one element for detection and/or enrichment (or more particularly detection, amplification and/or capture) of said target fragment, and of which at least a portion, including one end, is complementary in sequence to the second non-target-specific part of oligonucleotide A,
such that said target fragment becomes annealed to said probe through hybridisation to the first target-specific part of oligonucleotide A resulting in only one target-specific probe-binding event per target fragment;

(e) ligating oligonucleotide B of said probe to the part of the single-stranded portion of said target fragment which is hybridised to oligonucleotide A of said probe to produce a probe-target-fragment hybrid;

(f) if necessary rendering double-stranded an end of at least the target fragment, being the end to which the probe is not ligated, which end will comprise a common nucleic acid adaptor sequence in the case that such an adaptor is ligated to the fragments according to optional step (b);

(g) non-target-specifically annealing the double-stranded end of step (f) intramolecularly with the free, non-target fragment-bound end of the probe;

(h) ligating oligonucleotide B of the probe-target fragment hybrid with the corresponding strand of the double-stranded end of step (f) to circularise said hybrid; and

(i) enriching for said probe-target fragment hybrid by means that increases the ratio of circular to linear molecules.

By circularising the probe-target fragment hybrid, in this embodiment enrichment or detection may advantageously be based on the circularised nature of said hybrids in contrast to the linear form of the non-target fragments.

The step (f) of rendering double-stranded an end of at least the target fragment will not be necessary in embodiments wherein the fragments are rendered only partially single-stranded in part (c) as in such cases the fragment end other than that to which the probe becomes ligated will remain double-stranded. In instances where such rendering double-stranded is necessary, this is preferably achieved by polymerisation from the target fragment-hybridised end of oligonucleotide A of the probe (in cases where this end is a 3' end) or, when a common nucleic acid adaptor is ligated to the fragments, by annealing of an oligonucleotide complementary thereto. The non-target-specific annealing of step (g) may be by blunt or artificially-created sticky end annealing, i.e. by annealing means which does not rely on knowledge of sequence of the target fragment. If a common nucleic acid adaptor is ligated to the target fragment, the annealing may be between sticky ends generated

IDT Exhibit 1020.1265

through sequence-specific digestion of the adaptor and a corresponding sticky end at the free end of the probe.

The enrichment step may involve selective removal of non-circular nucleic acids by e.g. exonuclease treatment. Preferably a mixture of exonucleases enabling degradation of linear double-stranded DNA, for example including exonuclease III, lambda exonuclease and exonuclease I, is used, though any of these or any other suitable exonucleases may be used alone. Alternatively, the circular:linear nucleic acid molecule ratio is increased by selectively amplifying the circularised probe-target fragment hybrid, such as by RCA, hyperbranched RCA or multiple strand displacement.

The invention provides kits for use in the above-described methods. Such a kit will comprise at least an oligonucleotide B and one or more oligonucleotides A, and one or more restriction endonucleases and/or a common nucleic acid adaptor or a set of such adaptors which differ from each other only by molecular tags and/or annealing ends. The oligonucleotide B of the kit may comprise at least one detection and/or enrichment element (more particularly at least one of a detection, capture and/or amplification element) as defined above, and optionally also a cleavage site such as a recognition sequence for an endonuclease. Where one or more restriction endonucleases are included in the kit, such endonucleases will be those which fragment a particular target in a way which renders useful the one or more oligonucleotides A of the kit. If the kit comprises one or more common nucleic acid adaptors, such adaptors will have an annealing end compatible with the sticky ends generated by the one or more restriction endonucleases which fragment a particular target in a way which renders useful the one or more oligonucleotides A of the kit, regardless of whether such one or more restriction endonucleases are included in the kit. The one or more adaptors may contain an amplification and/or sequencing primer binding site, and/or a cleavage site.

The kit may further comprise one or any combination of two or more of the following components. Thus, the kit may contain a ligase for ligating the end of oligonucleotide B of the probe to the target nucleic acid fragment, particularly a ligase as specifically mentioned above in connection with the step of ligating the target fragment to the probe. A solid phase as defined above, to which the oligonucleotide B of the kit is immobilisable, or immobilised, may be included. Alternatively or additionally, an amplification or sequencing primer specific for a primer binding site in oligonucleotide B (i.e. corresponding to sequences of the oligonucleotide B) may be included, and optionally also a polymerase and/or other reagents required for amplification and/or sequencing such as nucleotides, buffers, ions, etc. Means for

IDT Exhibit 1020.1266

cleaving the probe at the cleavage site in oligonucleotide B, such as a restriction endonuclease, may be included. For embodiments of the method of the invention employing an exonuclease for rendering the sample fragments single-stranded, the kit may include such an enzyme, particularly such an enzyme as specifically mentioned above in connection with the rendering single-stranded step. A polymerase may be included for filling any gap between oligonucleotide A and the partially-degraded strand of the probe-hybridised target fragment and/or for extending a 3' end of oligonucleotide A in order to render the non-probe end of the probe-target fragment hybrid double-stranded in certain embodiments. Optional other kit components include a dephosphorylase (phosphatase) such as discussed above, a kinase for adding 5' phosphates for selective ligation, a polymerase (such as T4 DNA polymerase) for blunting overhanging (sticky) restriction fragment ends, free adenosine, a polymerase for adding same to blunted fragment ends, random hexamers for priming polymerisation in order to render the non-probe end of the probe-target fragment hybrid double-stranded in certain embodiments or for multiple strand displacement amplification, qPCR reagents e.g. fluorescently-labelled nucleotides or Taqman® probes, a primer or primers for RCA or hyperbranched RCA, and a flap endonuclease for fragmenting the nucleic acid sample.

The kit may optionally also contain appropriate primers, amplification/sequencing reagents and/or cleaving means as mentioned above for use in connection with an amplification and/or sequencing primer binding site, and/or a cleavage site, in a common nucleic acid adaptor which may be employed in the methods of the invention. Also optionally included in such a kit are single-stranded oligonucleotides corresponding to either strand of the double-stranded adaptor, in order to render the non-probe end of the probe-target fragment hybrid double-stranded in certain embodiments. In certain embodiments hybridisation/ligation of the target fragment to the probe is mediated by a flap endonuclease and accordingly the kit may further include such an enzyme and optionally appropriate co-factors.

The kit components may be present in separate containers, or one or more of the components may be present in the same container, where the containers may be storage containers and/or containers that are employed during the assay for which the kit is designed.

In addition to the above components, the subject kits may further include instructions for practicing the subject methods. These instructions may be present in the subject kits in a variety of forms, one or more of which may be present in the kit. One form in which these instructions may be present is as printed information on a suitable medium or substrate, e.g., a piece or pieces of paper on which the

IDT Exhibit 1020.1267

information is printed, in the packaging of the kit, in a package insert, etc. Yet another means would be a computer readable medium, e.g., diskette, CD, etc., on which the information has been recorded. Yet another means that may be present is a website address which may be used via the internet to access the information at a remote site. Any convenient means may be present in the kits.

Accordingly, in a further embodiment of the invention is provided a kit for detecting or enriching for a target nucleic acid present in a nucleic acid sample, said kit comprising:

(a) an oligonucleotide B, being a single-stranded oligonucleotide which may contain or carry at least one element for detection and/or enrichment (or more particularly detection. amplification and/or capture) of a target fragment, and of which at least a portion, including one end, is complementary in sequence to the second non-target-specific part of an oligonucleotide A, wherein said target fragment and said second non-target-specific part are as defined above, and;

(b) one or more oligonucleotides A, being a single-stranded oligonucleotide comprising at one end a first target-specific part comprising at least 10 nucleotides complementary in sequence to at least part of the single-stranded portion of a target fragment, and comprising at the other end a second non-target-specific part which comprises a nucleotide sequence complementary to at least a portion, including one end, of oligonucleotide B, wherein said single-stranded portion and said target fragment are as defined above, and;

(c)

(i) one or more restriction endonucleases, and/or;

(ii) a common nucleic acid adaptor, or a plurality of variants of such an adaptor which differ from each other at a carried molecular tag and/or at the end by which the adaptor anneals to said fragments.

The invention will be further described with reference to the following non-limiting Examples with reference to the following drawings in which:

Figure 1 shows a particular embodiment of the method.

The nucleic acid sample is restriction enzyme digested;

A common double-stranded adaptor is designed to ligate selectively to either the 3' or 5' ends of the fragments. This is achieved by selective phosphorylation/dephosphorylation of the adaptor or the fragments (phosphorylated ends shown as "P");

The fragments are denatured;

The probe is annealed to the target fragment;

The probe is ligated to the target fragment;

IDT Exhibit 1020.1268

Optionally the probe is equipped with a function for immobilization to a solid phase;

The target fragment is amplified using common sequences in the adaptor and oligonucleotide B of the probe.

Figure 2 shows a further particular embodiment of the method.

The nucleic acid sample is restriction enzyme digested;

A portion of either the 5'- or 3'-ends of all fragments are made single-stranded by exonucleolysis, whilst a middle portion remains double-stranded;

The probe is designed to hybridize with one of the two single stranded ends of the target fragment so that a junction susceptible to ligation is created;

The probe is ligated to the target fragment;

The probe is equipped with a function such as a biotin molecule (shown as "B") for solid phase enrichment;

Optionally the target fragment is amplified using a common sequence in oligonucleotide B of the probe.

Figure 3 shows a further particular embodiment of the method.

The sample nucleic acid is fragmented using physical or enzymatic means;

Optionally, a common double-stranded adaptor is ligated specifically to the 3´ ends of the fragments by e.g. dephosphorylating the sample and using a phosphorylated adaptor (phosphorylated ends shown as "P");

The fragments are denatured;

The probe is annealed to the target fragment to form a substrate for a flap endonucleolytic cleavage of the target fragment 5' end;

The probe is ligated to the target fragment;

Optionally the probe is equipped with a function for solid phase immobilization and enrichment;

The probe is amplified and or detected using the common sequences in oligonucleotide B of the probe and/or optionally the adaptor.

Figure 4 shows a further particular embodiment of the method.

The target fragment is equipped with a probe according to any of the embodiments of the method of the invention;

The distal part of the target fragment with respect to the probe is made double-stranded where this is not already the case. This may, for example, be

IDT Exhibit 1020.1269

mediated by polymerisation initiated from a hybridised probe 3' end or by annealing of hexamers followed by polymerisation;

The probe is designed to allow intramolecular circularisation with the distal end of the probe-target fragment hybrid molecule, formed by ligation;

The intramolecular ligation may, e.g., be by sticky end ligation or blunt end ligation;

Circularised probe-target fragment hybrids are then enriched for using rolling circle amplification (RCA), hyperbranched RCA, multiple strand displacement, exonucleolysis or other methods that increase the ratio between circular and linear molecules.

Figure 5 shows the amplification products from the amplification reaction. Am3 = amine adaptor ligated to fragments. NoDNA = negative control where no genomic DNA was included in the hybridization reaction. All 17 fragments that were selected are between 200 and 450 nucleotides long.

Figure 6 shows qPCR readout of amine blocked selection. Ct values for each primer pair are shown left to right along the x-axis, corresponding to descending order of the primer pairs listed in the key.

EXAMPLES

Example 1

Restriction enzyme digested human genomic DNA was incubated with T4 DNA polymerase in order to generate blunt ends. Subsequently a 3' adenosine overhang was generated by addition of Taq polymerase and the product was ligated to an adaptor with 3' protruding thymidines in separate reactions using T4 DNA ligase. Seventeen *in silico* designed probe-oligonucleotide As together with a common probe-oligonucleotide B were then added to the reaction mix. The reaction was then heat denatured and the probes were allowed to hybridize to the single stranded ends of 17 specific genomic fragments in the restriction enzyme digested genomic DNA sample. After hybridization the probes were captured by a solid phase (beads) using a biotin on the probe. Following stringent washing probes were ligated on the beads and all fragments were amplified by PCR using one common primer pair located in the probe and adaptor respectively. The results were analyzed by gel electrophoresis and qPCR.

IDT Exhibit 1020.1270

The sequences of the 17 probe-oligonucleotide As, adaptor, probe-oligonucleotide B and primer pair are outlined in Table 1. The "amine" terminology used in connection with the adaptors indicates that one end of the adaptors (the end not intended to ligate with the nucleic acid fragments) is "protected" from such ligation by an amine (NH_2) group.

Table 1

Target ID	Probe-oligonucleotide A
ROI_1 (SEQ ID NO: 1)	ACAACGGGAATTCAAATCAAGATGGTGGCCACACCCCATGCGTACCAATGGATGCGGTCT
ROI_2 (SEQ ID NO: 2)	AAAAGAAGCAGGGAGTCAACTTCCTGCACCTTTAACTCTACGTACCAATGGATGCGGTCT
ROI_3 (SEQ ID NO: 3)	CACCTGATGATGGTCMTGGAGGCATTGTTCTGATTCTTTGCGTACCAATGGATGCGGTCT
ROI_4 (SEQ ID NO: 4)	CACACACCAGACTTGTCAGCCTCCCAAAGAGCCATGCTCCCGTACCAATGGATGCGGTCT
ROI_5 (SEQ ID NO: 5)	CACTGGCTGAGGTGCCAGATGGGTCTCCTGGCTATAGGGACGTACCAATGGATGCGGTCT
ROI_6 (SEQ ID NO: 6)	TCTTCAAAGGCAAGGTTTCAATTATGCAAAAACCACCTTACGTACCAATGGATGCGGTCT
ROI_7 (SEQ ID NO: 7)	AAYCTACGGCCATGGCAGGTAAGACGCTGCAGGGAAGCAGCGTACCAATGGATGCGGTCT
ROI_8 (SEQ ID NO: 8)	TGCTGTTTCCACATAGCAGAGGCTGTTAGGCTCAGAGCTACGTACCAATGGATGCGGTCT
ROI_9 (SEQ ID NO: 9)	GGGGTGCTTGAGAGCTTCGCTTGCACTTATTCGCCAACTACGTACCAATGGATGCGGTCT
ROI_10 (SEQ ID NO: 10)	TCTGTGGTAGTACCTAGAATAAGCTATGCAGCCTCTCTGACGTACCAATGGATGCGGTCT
ROI_11 (SEQ ID NO: 11)	GACAGCCGTGTATTACCCAATATCCCCAAAGGCAGCCTTACGTACCAATGGATGCGGTCT
ROI_12 (SEQ ID NO: 12)	GGTTCTGCTCCTCATCCGCTGTGGATCAGATGTGCTCTGACGTACCAATGGATGCGGTCT
ROI_13 (SEQ ID NO: 13)	AATCTTTTGGAACCCAATTGGTCATATTTCCTTCTACTTACGTACCAATGGATGCGGTCT
ROI_14 (SEQ ID NO: 14)	TGCTGGTTCTCAAATCCGAACGCTCTGTGATAAATTTCTGCGTACCAATGGATGCGGTCT
ROI_15 (SEQ ID NO: 15)	CTCACAGAAACCAAAGCGTTTCCCAACAGCACGTTTCCTGCGTACCAATGGATGCGGTCT
ROI_16 (SEQ ID NO: 16)	AATTTTAGAAATCCTTRAATTTTCCATGTCTACATTCATGCGTACCAATGGATGCGGTCT
ROI_17 (SEQ ID NO: 17)	GTTCTATGAAAAGAAAAAAAGAAACGATTAAGGTTTCATGCGTACCAATGGATGCGGTCT
Adaptor 5'-3'	
Amine plus (SEQ ID NO: 18)	CGTTATCAACCTGGGTCCGA
Amine minus (SEQ ID NO: 19)	TCGGACCCAGGTTGATAACGT
Probe-oligonucleotide B 5'-3'	
(SEQ ID NO: 20)	CTGGACCTTAATCGTGTGCGAGACCGCATCCATTGGTACG
Common primer pair 5'-3'	
Fwd primer (SEQ ID NO: 21)	AGACCGCATCCATTGGTACG
Rev primer (SEQ ID NO: 22)	TCGGACCCAGGTTGATAACG

Restriction enzyme digestion

Genomic DNA was restriction enzyme digested using restriction 10 different restriction enzymes combined in 5 separate reactions outlined in Table 2. 10 units of each restriction enzyme were used to digest 1 µg of DNA at 37 ˚C for 2 h.

IDT Exhibit 1020.1271

Table 2

Target ID	Target length (bp)	Restriction reaction	5' end	3' end	qPCR primer pair
ROI_1	314	XmnI/HpyCH4V	blunt	blunt	P01
ROI_2	305	BfaI/AlwI	5'o-h	5'o-h	P02
ROI_3	395	XmnI/HpyCH4V	blunt	blunt	P03
ROI_4	260	HpyAV/DraI	blunt	blunt	P04
ROI_5	386	NspI/MlyI	3'o-h	3'o-h	P05
ROI_6	288	DdeI/BspHI	5'o-h	5'o-h	P06
ROI_7	270	HpyAV/DraI	blunt	blunt	P07
ROI_8	264	BfaI/AlwI	5'o-h	5'o-h	P08
ROI_9	287	BfaI/AlwI	5'o-h	5'o-h	P09
ROI_10	205	DdeI/BspHI	5'o-h	5'o-h	P10
ROI_11	301	DdeI/BspHI	5'o-h	5'o-h	P11
ROI_12	226	DdeI/BspHI	5'o-h	5'o-h	P12
ROI_13	361	DdeI/BspHI|	5'o-h	5'o-h	P13
ROI_14	355	XmnI/HpyCH4V	blunt	blunt	P14
ROI_15	226	XmnI/HpyCH4V	blunt	blunt	P15
ROI_16	385	DdeI/BspHI	5'o-h	5'o-h	P16
ROI_17	365	DdeI/BspHI	5'o-h	5'o-h	P17

Table 3 - Primer pairs used

Primer pair	Target	Forward	Reverse
P1	ROI_1	TTCCCCACTGACAGCCTC (SEQ ID NO: 23)	GACCCAAAACCCAAAATGG (SEQ ID NO: 24)
P2	ROI_2	CCCCTGTGGACCTCAACC (SEQ ID NO: 25)	TGCTTGAAAAGCCAGTGC (SEQ ID NO: 26)
P3	ROI_3	TCACGGAGGCATTCTAAAGTC (SEQ ID NO: 27)	TTGATGCCCCCAAGAATC (SEQ ID NO: 28)
P4	ROI_4	CCAAGGGCATCCAGTTTG (SEQ ID NO: 29)	GGGGCCACACACATCTTC (SEQ ID NO: 30)
P5	ROI_5	ACCTTCCTTGCCCCTCTG (SEQ ID NO: 31)	TCCAGCCGTCAACTCCTC (SEQ ID NO: 32)
P6	ROI_6	TGAGCAAATCCAGTCAGGG (SEQ ID NO: 33)	ACTGTGTGGCAAACTGCG (SEQ ID NO: 34)
P7	ROI_7	TGAAAAGAGAACATGGGGG (SEQ ID NO: 35)	GAGAAGCCCTTTCCAGGC (SEQ ID NO: 36)
P8	ROI_8	CTCACCTTTGCGCCTCTG (SEQ ID NO: 37)	GAGGTGGAGAAACGCAGG (SEQ ID NO: 38)
P9	ROI_9	GCGAAGCTCTCAAGCACC (SEQ ID NO: 39)	CATTGAGTCTGGAGTGGAGC (SEQ ID NO: 40)
P10	ROI_10	CTCCTCTGTGCAGGTGGG (SEQ ID NO: 41)	GGATGTCCTCAAGCCGTG (SEQ ID NO: 42)
P11	ROI_11	GGAAACTCCCCTTACCCG (SEQ ID NO: 43)	TGTTGCCCATGTCAGCAC (SEQ ID NO: 44)
P12	ROI_12	GTCCCATGGTGCTTGCAG (SEQ ID NO: 45)	ATCTCTGGCTCCGTCGTG (SEQ ID NO: 46)
P13	ROI_13	TTTGACGGGCATCCTTTC (SEQ ID NO: 47)	ACAGGCGAAGGAGGTGTG (SEQ ID NO: 48)
P14	ROI_14	GTGACCTGCCACCTCCAG (SEQ ID NO: 49)	GCTGGCGTAAAGGTGAGG (SEQ ID NO: 50)
P15	ROI_15	AATGGCAACGACGGGTAG (SEQ ID NO: 51)	CAAACGCTCTGAGACAGCC (SEQ ID NO: 52)
P16	ROI_16	GCCTGGCAGAGCTGAATC (SEQ ID NO: 53)	CCAAGCACCTAACAGGCATC (SEQ ID NO: 54)
P17	ROI_17	CCGAAACATGGATTTGGC (SEQ ID NO: 55)	TCAACCGGCAAAGTCAGC (SEQ ID NO: 56)
G1	Not enriched	GACAGCTCCCCACACACC (SEQ ID NO: 57)	TTCCTGCCTGAGCTGACC (SEQ ID NO: 58)
G2	Not enriched	TGCCTCTCTTGCTCTGGG (SEQ ID NO: 59)	GTGGGCATGGGTCAGAAG (SEQ ID NO: 60)
G3	Not enriched	ACAGCTGCCCACTTCTGG (SEQ ID NO: 61)	GCGAGGACCAAACTCAGG (SEQ ID NO: 62)

IDT Exhibit 1020.1272

G4	Not enriched	ATTCAGGCGCTTTGCATC (SEQ ID NO: 63)	AGGCTGGTCACATGGGTG (SEQ ID NO: 64)
G5	Not enriched	CATGGTCTTGGACTGGGC (SEQ ID NO: 65)	AGCTCGATCTTCATGCGG (SEQ ID NO: 66)
Alu1	Alu repeat	GCGCGGTGGCTCACGCCTGT (SEQ ID NO: 67)	CCTCCCAAAGTGCTGGGATT (SEQ ID NO: 68)
Alu2	Alu repeat	CGCCACTGCACTCCAGCCTG (SEQ ID NO: 69)	CGATCTCCTGACCTCATGAT (SEQ ID NO: 70)

Blunt ending fragments using T4 DNA polymerase

Blunt ends were generated from 0.5 µg of digested genomic DNA using 3 Units of T4 polymerase and 0.1 mM dNTP in 1 × Special Buffer (50 mM KAc, 3 mM MgAc, 20 mM TrisAc, 1 mM DTT, 0.1 µg/µl BSA) in a total reaction volume of 15 µl. Two separate enrichment reactions were performed and one additional negative control reaction without addition of restriction enzyme digested genomic DNA. All reactions were incubated at 12 ˚C for 15 min and 65 ˚C for 20 min.

Shrimp alkaline phosphatase (SAP)

5' phosphates were removed by addition of 1 U of SAP in 5 µl 1 × Special Buffer. Reactions were incubated at 37 ˚C for 15 min and 65 ˚C for 15 min.

Addition of dATP 3' overhangs

Protruding adenosine 3' ends were generated by addition of 5 U Taq polymerase, 0.5 mM dATP in 5 µl 1 × Special Buffer. Reactions were incubated at 72 ˚C for 20 min.

TA-ligation to adaptors

Fifteen Units of T4 DNA Ligase was used to ligate 0.83 µM amine modified adaptor (5 × the concentration of all fragment ends) and 1 mM ATP to the genomic DNA. Total reaction volumes were 30 µl in 1 × Special Buffer. Reactions were incubated over night at 16 ˚C and heat inactivated at 65 ˚C for 20 min.

Enrichment

Probe hybridizations were performed using 12.5 µl of the 30 µl reactions, 0.1 nM of each probe-oligonucleotide A and 7.73 nM of the biotinylated probe-oligonucleotide B, 27.27 % formamide and 0.7 × Bind&Wash buffer in a total reaction volume of 27.5 µl (1 × B&W: 1 M NaCl, 5 mM EDTA, 0,1 % Tween-20, 10 mM Tris-HCl pH 7.5). Reactions were incubated at:

95 ˚C for 10 min, 75 ˚C for 30 min, 68 ˚C for 30 min, 62 ˚C for 30 min, 55 ˚C for 30 min, 46 ˚C for 10h, 10 ˚C overnight.

Probes were captured using 10 µg streptavidin-coated magnetic beads (Dynabeads M-280, Invitrogen) in a total volume of 200 µl 1.09 × B&W buffer and 3.8 % Formamide. Reactions were incubated at room temperature for 1 h with rotation.

IDT Exhibit 1020.1273

The supernatant was removed using a magnet rack and reactions were washed with 200 µl 1 × B&W buffer and 20 % formamide at 46 ˚C for 30 min with rotation.

The washing solution was removed using a magnet rack and beads washed with 200 µl 1 × B&W buffer at 46 ˚C for 10 min with rotation.

After removing the washing buffer using a magnet rack reactions were phosphorylated and ligated in the same reaction by addition of 5 U PNK, 1 mM ATP, 12.5 U Ampligase, 1 × Ampligase buffer, 1 mM NAD and 0.2 µg/µl BSA in a total volume of 50 µl. Reactions were incubated at 37 ˚C for 30 min and 55 ˚C for 1 h.

The ligation reactions were removed using a magnetic rack. PCR was performed using 1.5 U Taq Platinum polymerase, 0.1 µM primers, 0.2 mM of each dATP, dUTP, dCTP and dGTP, 2 mM $MgCl_2$ and 1 × PCR buffer in a final volume of 50 µl.

PCR reactions were incubated at 95 ˚C for 3 min followed by 40 cycles at 95 ˚C for 30 sec, 55 ˚C for 1 min and 72 ˚C for 1 min and a final step at 72 ˚C for 10 min.

Reactions were heated to 95 ˚C for 5 min and supernatants were transferred to new tubes using a magnet rack.

Five microlitres of the amplified products were analyzed using 1 % ethidium bromide-containing agarose gel electrophoresis and the results are presented in Figure 5.

qPCR was then performed on the enriched fragments. The qPCR analysis used 17 primer pairs, one targeting each of the 17 enriched genomic fragments (P1-17, Table 3) as well as 5 primer pairs amplifying regions that were not enriched (G1-5). Two PCR reactions using primers targeting Alu-repeats were also performed (Alu1-2). 10 µl 1:100 diluted PCR product was used in each reaction with 0.9 Units Taq Platinum polymerase, 0.17 µM primer, 0.2 mM dNTP, 1 × SYBR® Green I, 10 % DMSO and 2 mM MgCl2 in 1 × PCR buffer to a total volume of 30 µl.

Example 2

Genomic DNA from DNA samples of interest is digested with a rare-cutting restriction enzyme. It may be preferable to use a type IIs enzyme, because thereby restriction fragments will not share end sequences. The enzyme should furthermore leave 5' overhangs in order to be suitable for processing in the subsequent step.

Treat the digested DNA with exonuclease III for a predetermined time, adjusted to remove a suitable length of the 3' end sequence of all restriction fragments.

IDT Exhibit 1020.1274

Add a large amount of probe-oligonucleotide As whose 5' "halves" are complementary to (one of) the 5' ends of the restriction fragments that should be isolated, while the other "halves" of the oligonucleotide As contain a standard sequence. The aim is to reach all copies of the restriction fragments. Preferably each half of the oligonucleotide As encompasses 30 or so nucleotides.

Also add an even greater amount of probe-oligonucleotide Bs complementary to the standard sequence of the oligonucleotide As under step 3, and modified at the 5' ends for efficient capture on a solid phase either before or after being ligated to the genomic restriction fragments.

The digested genomic DNA (step 2) and probe oligonucleotides (steps 3 and 4) are incubated under suitably stringent conditions.

Add DNA ligase and cofactors so that probes now become covalently joined to the appropriate genomic DNA fragments.

If desired, add a strand-displacing polymerase and nucleotides to repair the ExoIII-digested ends of the restriction fragments, also copying the covalently joined oligonucleotides.

If ligation was performed in solution, then free probes can now optionally be removed by a simple size separation step.

If ligation was performed in solution, the probes that were joined to the restriction fragments by ligation are next immobilized on a solid-phase. An attractive mechanism for capture on solid support is through click chemistry, but many alternatives can be contemplated, including via biotin-streptavidin. The nature of the solid phase is also important. Possibly very small particles could be used to efficiently reach all oligonucleotides, and then prevent movement of the modified DNA through a gel. Alternatively, a chromatographic matrix could be extensively modified with groups that can be reacted covalently or non-covalently with ones on oligonucleotide B (step 4).

Separate the particles/matrix from DNA remaining in solution.

Release the captured fragments by digestion with an enzyme designed to cleave the immobilized oligonucleotides, e.g. using a rarely-cutting restriction site included in oligonucleotide B on the solid phase (step 4).

If required the released and fragmented DNA can now be ligated again to give rise to double-stranded DNA circles or linear concatemers that can be amplified by whole-genome amplification procedures. Alternatively the released DNA can be directly applied for sequencing.

IDT Exhibit 1020.1275

Claims

1. A method for detecting or enriching for a target deoxyribonucleic acid (DNA) present in a nucleic acid sample, said method comprising:

(a) fragmenting a nucleic acid sample to generate nucleic acid fragments including a target fragment containing said target DNA;

(b) non-target-specifically annealing to the ends of said fragments one, or a plurality of variants of a, common nucleic acid adaptor, wherein the annealed adaptor becomes ligated to the fragments at the 3' ends and/or the 5' ends of the strands of the fragments such that an adaptor is ligated only to the end of the strand of the target fragment to which, at the other end, oligonucleotide B of the probe is ligated in step (e), and wherein said variant adaptors differ from each other at a carried molecular tag and/or at the end by which the adaptor anneals to said fragments;

(c) rendering said fragments, including said target fragment, at least partially single-stranded, wherein the single-stranded portion includes an end portion and wherein the length of said single-stranded portion is sufficient to allow hybridisation of at least part of the single-stranded portion of said target fragment to the probe of step (d);

(d) contacting the at least partially single-stranded fragments of step (c) with oligonucleotides A and B of a single target-specific nucleic acid probe, wherein:

(i) oligonucleotide A is a single-stranded oligonucleotide comprising at one end a first target-specific part comprising at least 10 nucleotides complementary in sequence to at least part of said single-stranded portion of said target fragment, and comprising at the other end a second non-target-specific part which comprises a nucleotide sequence complementary to at least a portion, including one end, of oligonucleotide B of the probe, and

(ii) oligonucleotide B is a single-stranded oligonucleotide which may contain or carry at least one element for detection and/or enrichment of said target fragment, and of which at least a portion, including one end, is complementary in sequence to the second non-target-specific part of oligonucleotide A,

such that said target fragment becomes annealed to said probe through hybridisation to the first target-specific part of oligonucleotide A resulting in only one target-specific probe-binding event per target fragment;

(e) directly or indirectly ligating oligonucleotide B of said probe to the part of the single-stranded portion of said target fragment which is hybridised to oligonucleotide A of said probe to produce a probe-target fragment hybrid; and

IDT Exhibit 1020.1276

(f) detecting or enriching for said probe-target fragment hybrid.

2. The method of claim 1 for detecting or enriching for a plurality of target DNAs, wherein in step (d) said target fragments are contacted with oligonucleotides A and B of a plurality of nucleic acid probes, each having an oligonucleotide A with a different first target-specific part, whereby a plurality of different target fragments may be respectively annealed to said probes.

3. The method of claim 2, wherein in said plurality of nucleic acid probes oligonucleotide B comprises a common sequence which is the same in each probe.

4. The method of claim 3, wherein said common sequence in each probe comprises the detection and/or enrichment element.

5. The method of any one of claims 2 to 4, wherein the plurality of nucleic acid probes comprise the same oligonucleotide B.

6. The method of any one of claims 1 to 5, wherein said nucleic acid sample comprises genomic DNA or a subfraction thereof.

7. The method of any one of claims 1 to 5, wherein said nucleic acid sample comprises cDNA or a subfraction thereof.

8. The method of any one of claims 1 to 7, wherein said target fragment comprises at least 30 nucleotides.

9. The method of any one of claims 1 to 8, wherein in step (a) said fragmenting is: (i) enzymatic; and/or (ii) physical.

10. The method of claim 9, wherein said enzymatic fragmenting is by means of a restriction endonuclease.

11. The method of claim 9 or claim 10, wherein said enzymatic fragmenting is by means of an enzyme for which a recognition sequence close to the target DNA in the nucleic acid sample is polymorphic, whereby said detecting or enriching for said target DNA is haplotype-specific.

IDT Exhibit 1020.1277

12. The method of any one of claims 1 to 9, wherein said fragmenting is by subjecting aliquots of said nucleic acid sample to respectively different combinations of one or more enzymatic and/or physical methods, and pooling said aliquots of fragmented nucleic acid sample prior to step (c).

13. The method of any one of claims 1 to 11, wherein in step (c) said rendering at least partially single-stranded is by: (i) denaturation or (ii) exonucleolysis.

14. The method of any one of claims 1 to 12, wherein in step (d) said first target-specific part of oligonucleotide A comprises at least 10 nucleotides complementary in sequence to a single-stranded portion at the end of said at least partially single-stranded target fragment.

15. The method of any one of claims 1 to 14, wherein the length of the first target-specific part of oligonucleotide A in step (d) is at least 20 nucleotides.

16. The method of any one of claims 1 to 15, wherein in step (d) said enrichment element contained or carried by oligonucleotide B is an amplification and/or capture element.

17. The method of claim 16, wherein said amplification and capture elements are, respectively, an amplification primer binding site and an element for immobilisation to a solid phase.

18. The method of any one of claims 1 to 14, wherein in step (d) oligonucleotide B is immobilised to a solid phase.

19. The method of any one of claims 1 to 18, wherein in step (d) oligonucleotide B further carries or contains a molecular tag.

20. The method of any one of claims 1 to 19, wherein in step (e) said ligating of oligonucleotide B is to the end of said target fragment.

21. The method of claim 17, wherein in step (f) said target fragment in said probe-target fragment hybrid is amplified from a primer hybridised to said amplification primer binding site.

IDT Exhibit 1020.1278

22. The method of claim 17, wherein in step (f) said target fragment in said probe-target fragment hybrid is immobilised to a solid phase by means of said immobilisation element.

23. The method of any one of claims 1 to 22, wherein in step (f) said target fragment in said probe-target fragment hybrid, or an amplification product thereof, is sequenced.

24. The method of any one of claims 1 to 23, wherein said common nucleic acid adaptor comprises an element for amplification.

25. The method of claim 24, wherein in step (f) said target fragment in said probe-target fragment hybrid is amplified by means of the amplification elements in oligonucleotide B of the probe and in the adaptor.

26. The method of any one of claims 1 to 23, wherein said common nucleic acid adaptor carries a molecular tag.

27. The method of any one of claims 24 to 26, wherein said non-specific annealing is by rendering the ends of the fragments sticky with respect to at least one end of the adaptor, and wherein said ligating of the adaptor to the fragments at the 3' ends, or at the 5' ends, of the strands of the fragments is by selective dephosphorylation of the fragments or the adaptor, respectively, prior to ligation.

28. The method of any one of claims 24 to 27, wherein in step (a) said fragmenting a nucleic acid sample comprises separately fragmenting a plurality of nucleic acid samples, and step (b) comprises ligating to the fragments of which samples, respectively, variant adaptors carrying different molecular tags and pooling said adaptor-ligated fragmented nucleic acid samples.

29. The method of any one of claims 1 to 28, wherein in step (c) said fragments are rendered partially, but not completely, single-stranded.

30. The method of any one of claims 24 to 27, wherein in step (a) said fragmenting is as defined in claim 10, and step (b) comprises ligating to the fragments of said pooled aliquots variant adaptors which differ at the end by which the adaptor anneals to said fragments.

IDT Exhibit 1020.1279

31. The method of any one of claims 1 to 13 and 15 to 30, wherein in step (d) said first target-specific part of oligonucleotide A comprises at least 10 nucleotides complementary in sequence to a single-stranded internal non-end portion of said at least partially single-stranded target fragment, and wherein said annealing of said probe to said target fragment is through hybridisation of said internal non-end portion to said first target-specific part of oligonucleotide A, causing the 5' end of the target fragment to form a substrate for flap endonucleolytic cleavage and further comprising cleaving said flap endonucleolytic cleavage substrate to produce a ligatable 5' end of said target fragment which is hybridised to oligonucleotide A of said probe.

32. The method of claim 31, wherein in step (e) said ligating of oligonucleotide B is to said ligatable 5' end of said target fragment.

33. The method of any one of claims 1 to 10, 24 and 26 to 32, wherein in step (f) said detecting or enriching is by means that increases the ratio of circular to linear molecules, further comprising, betweens steps (e) and (f), rendering double-stranded an end of at least the target fragment, being the end to which the probe is not ligated, which end will comprise a common nucleic acid adaptor sequence ligated to the fragments in step (b); non-target-specifically annealing said double-stranded end intramolecularly with the free, non-target fragment-bound end of the probe; and ligating oligonucleotide B of the probe-target fragment hybrid with the corresponding strand of said double-stranded end to circularise said hybrid.

34. The method of claim 33, wherein in step (f) said means that increases the ratio of circular to linear molecules is by exonuclease treatment and/or rolling circle amplification.

35. A method for detecting or enriching for a target deoxyribonucleic acid (DNA) present in a nucleic acid sample, said method comprising:

(a) fragmenting a nucleic acid sample to generate nucleic acid fragments including a target fragment containing said target DNA;

(b) non-target-specifically annealing to the ends of said fragments a common nucleic acid adaptor, wherein the annealed adaptor becomes ligated to the fragments only at the 3' ends, or only at the 5' ends, of the strands of the fragments;

(c) rendering said fragments, including said target fragment, single-stranded;

IDT Exhibit 1020.1280

(d) contacting the single-stranded fragments of step (c) with oligonucleotides A and B of a single target-specific nucleic acid probe, wherein:

(i) oligonucleotide A is a single-stranded oligonucleotide comprising at one end a first target-specific part comprising at least 10 nucleotides complementary in sequence to a portion at the end of said single-stranded target fragment other than the end to which the common nucleic acid adaptor is ligated in step (b), and comprising at the other end a second non-target-specific part which comprises a nucleotide sequence complementary to at least a portion, including one end, of oligonucleotide B of the probe, and

(ii) oligonucleotide B is a single-stranded oligonucleotide which contains or carries an amplification primer binding site and optionally an element for immobilisation to a solid phase, and of which at least a portion, including one end, is complementary in sequence to the second non-target-specific part of oligonucleotide A,

such that said target fragment becomes annealed to said probe through hybridisation to the first target-specific part of oligonucleotide A resulting in only one target-specific probe-binding event per target fragment;

(e) ligating oligonucleotide B of said probe to the end of said target fragment which is hybridised to oligonucleotide A of said probe to produce a probe-target fragment hybrid; and

(f) enriching for said probe-target fragment hybrid through amplification by means of the amplification primer binding site in oligonucleotide B of the probe and an amplification primer binding site in the common nucleic acid adaptor, and optionally immobilising said fragment to a solid phase by means of the immobilisation element in oligonucleotide B of the probe.

36. A method for detecting or enriching for a target deoxyribonucleic acid (DNA) present in a nucleic acid sample, said method comprising:

(a) fragmenting a nucleic acid to generate nucleic acid fragments including a target fragment containing said target DNA;

(b) rendering said fragments, including said target fragment, partially single-stranded by 3' or 5' exonuclease digestion, wherein the length of the resulting single-stranded end portion is sufficient to allow hybridisation of at least part of the single-stranded end portion of said target fragment to the probe of step (c);

(c) contacting the partially single-stranded fragments of step (b) with oligonucleotides A and B of a single target-specific nucleic acid probe, wherein:

IDT Exhibit 1020.1281

(i) oligonucleotide A is a single-stranded oligonucleotide comprising at one end a first target-specific part comprising at least 10 nucleotides complementary in sequence to at least part of said single-stranded portion of said target fragment, and comprising at the other end a second non-target-specific part which comprises a nucleotide sequence complementary to at least a portion, including one end, of oligonucleotide B of the probe, and

(ii) oligonucleotide B is a single-stranded oligonucleotide which contains or carries an element for immobilisation to a solid phase and optionally an amplification primer binding site, and of which at least a portion, including one end, is complementary in sequence to the second non-target-specific part of oligonucleotide A,

such that said target fragment becomes annealed to said probe through hybridisation to the first target-specific part of oligonucleotide A resulting in only one target-specific probe-binding event per target fragment ;

(d) ligating oligonucleotide B of said probe to the end of said target fragment which is hybridised to oligonucleotide A of said probe to produce a probe-target fragment hybrid; and

(e) enriching for said probe-target fragment hybrid through immobilisation to a solid phase by means of the immobilisation element in oligonucleotide B of the probe, and optionally by amplifying said target fragment by means of the amplification primer binding site in oligonucleotide B of the probe.

37. A method for detecting or enriching for a target deoxyribonucleic acid (DNA) present in a nucleic acid sample, said method comprising:

(a) fragmenting a nucleic acid sample to generate nucleic acid fragments including a target fragment containing said target DNA;

(b) optionally, non-target-specifically annealing to the ends of said fragments a common nucleic acid adaptor, wherein the annealed adaptor becomes ligated to the fragments at the 3' ends or at the 3' and 5' ends of the strands of the fragments such that an adaptor is ligated to at least the 3' end of the strand of the target fragment to which, at the 5' end, oligonucleotide B of the probe is ligated in step (f);

(c) rendering said fragments, including said target fragment, single-stranded;

(d) contacting the single-stranded fragments of step (c) with oligonucleotides A and B of a single target-specific nucleic acid probe, wherein:

IDT Exhibit 1020.1282

(i) oligonucleotide A is a single-stranded oligonucleotide comprising at one end a first target-specific part comprising at least 10 nucleotides complementary in sequence to an internal non-end portion of said single-stranded target fragment, and comprising at the other end a second non-target-specific part which comprises a nucleotide sequence complementary to at least a portion, including one end, of oligonucleotide B of the probe, and

(ii) oligonucleotide B is a single-stranded oligonucleotide which contains or carries at least one element for detection and/or enrichment of said target fragment, and of which at least a portion, including one end, is complementary in sequence to the second non-target-specific part of oligonucleotide A,

such that said target fragment becomes annealed at said internal non-end portion to said probe through hybridisation to the first target-specific part of oligonucleotide A resulting in only one target-specific probe-binding event per target fragment, wherein said hybridisation causes the 5' end of the target fragment to form a substrate for flap endonucleolytic cleavage;

(e) cleaving said flap endonucleolytic cleavage substrate to produce a ligatable 5' end of said target fragment which is hybridised to oligonucleotide A of said probe;

(f) ligating oligonucleotide B of said probe to the part of the single-stranded portion of said target fragment which is hybridised to oligonucleotide A of said probe to produce a probe-target fragment hybrid;

(g) enriching for said probe-target fragment hybrid by means of said detection and/or enrichment element.

38. A method for detecting or enriching for a target deoxyribonucleic acid (DNA) present in a nucleic acid sample, said method comprising:

(a) fragmenting a nucleic acid sample to generate nucleic acid fragments including a target fragment containing said target DNA;

(b) optionally, non-target-specifically annealing to the ends of said fragments a common nucleic acid adaptor, wherein the annealed adaptor becomes ligated to the fragments at the 3' ends and/or at the 5' ends of the strands of the fragments such that an adaptor is ligated only to the end of the strand of the target fragment to which, at the other end, oligonucleotide B of the probe is ligated in step (e);

(c) rendering said fragments, including said target fragment, at least partially single-stranded, wherein the single-stranded portion includes an end portion and wherein the length of said single-stranded portion is sufficient to allow

IDT Exhibit 1020.1283

hybridisation of at least part of the single-stranded portion of said target fragment to the probe of step (d);

(d) contacting the at least partially single-stranded fragments of step (c) with oligonucleotides A and B of a single target-specific nucleic acid probe, wherein:

(i) oligonucleotide A is a single-stranded oligonucleotide comprising at one end a first target-specific part comprising at least 10 nucleotides complementary in sequence to at least part of said single-stranded portion of said target fragment, wherein if a common nucleic acid adaptor is ligated to the fragments according to optional step (b) said first target-specific part comprises at least 10 nucleotides complementary in sequence to a portion at the end of said at least partially single-stranded target fragment other than the end to which the common nucleic acid adaptor is ligated;
and comprising at the other end a second non-target-specific part which comprises a nucleotide sequence complementary to at least a portion, including one end, of oligonucleotide B of the probe, and

(ii) oligonucleotide B is a single-stranded oligonucleotide which contains or carries at least one element for detection and/or enrichment of said target fragment, and of which at least a portion, including one end, is complementary in sequence to the second non-target-specific part of oligonucleotide A,
such that said target fragment becomes annealed to said probe through hybridisation to the first target-specific part of oligonucleotide A resulting in only one target-specific probe-binding event per target fragment;

(e) ligating oligonucleotide B of said probe to the part of the single-stranded portion of said target fragment which is hybridised to oligonucleotide A of said probe to produce a probe-target fragment hybrid;

(f) if necessary, rendering double-stranded an end of at least the target fragment, being the end to which the probe is not ligated, which end will comprise a common nucleic acid adaptor sequence in the case that such an adaptor is ligated to the fragments according to optional step (b);

(g) non-target-specifically annealing the double-stranded end of step (f) intramolecularly with the free, non-target fragment-bound end of the probe;

(h) ligating oligonucleotide B of the probe-target fragment hybrid with the corresponding strand of the double-stranded end of step (f) to circularise said hybrid; and

(i) enriching for said probe-target fragment hybrid by means that increases the ratio of circular to linear molecules.

IDT Exhibit 1020.1284

39. A kit for detecting or enriching for a target DNA present in a nucleic acid sample, said kit comprising:

(a) an oligonucleotide B, being a single-stranded oligonucleotide which may contain or carry at least one element for detection and/or enrichment of a target fragment, and of which at least a portion, including one end, is complementary in sequence to the second non-target-specific part of an oligonucleotide A, wherein said target fragment and said second non-target-specific part are as defined in claim 1, and;

(b) one or more oligonucleotides A, being a single-stranded oligonucleotide comprising at one end a first target-specific part comprising at least 10 nucleotides complementary in sequence to at least part of the single-stranded portion of a target fragment, and comprising at the other end a second non-target-specific part which comprises a nucleotide sequence complementary to at least a portion, including one end, of oligonucleotide B, wherein said single-stranded portion and said target fragment are as defined in claim 1, and;

(c)

(i) one or more restriction endonucleases; and/or

(ii) a common nucleic acid adaptor, or a plurality of variants of such an adaptor which differ from each other at a carried molecular tag and/or at the end by which the adaptor anneals to said fragments.

IDT Exhibit 1020.1285

Figure 1

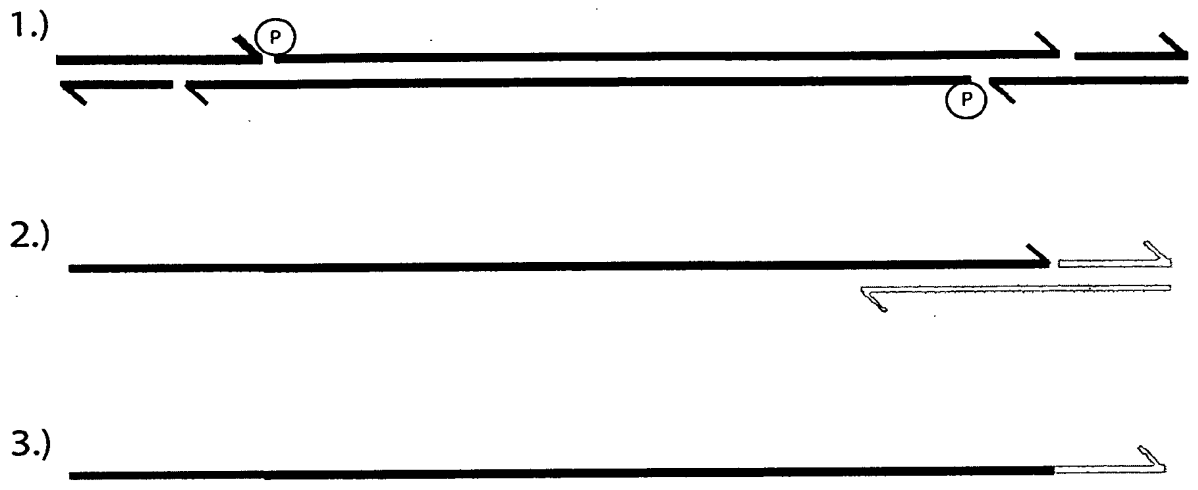
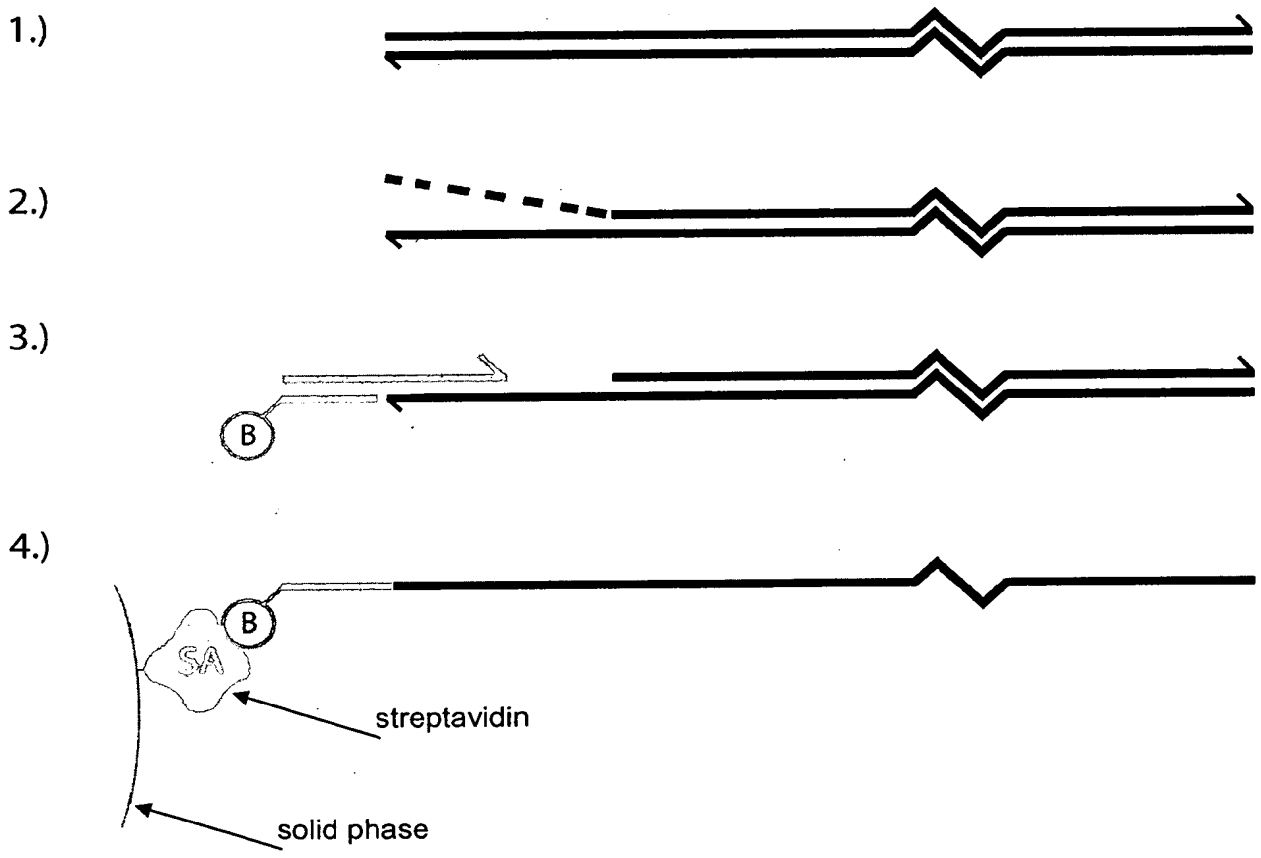


Figure 2



IDT Exhibit 1020.1286

Figure 3

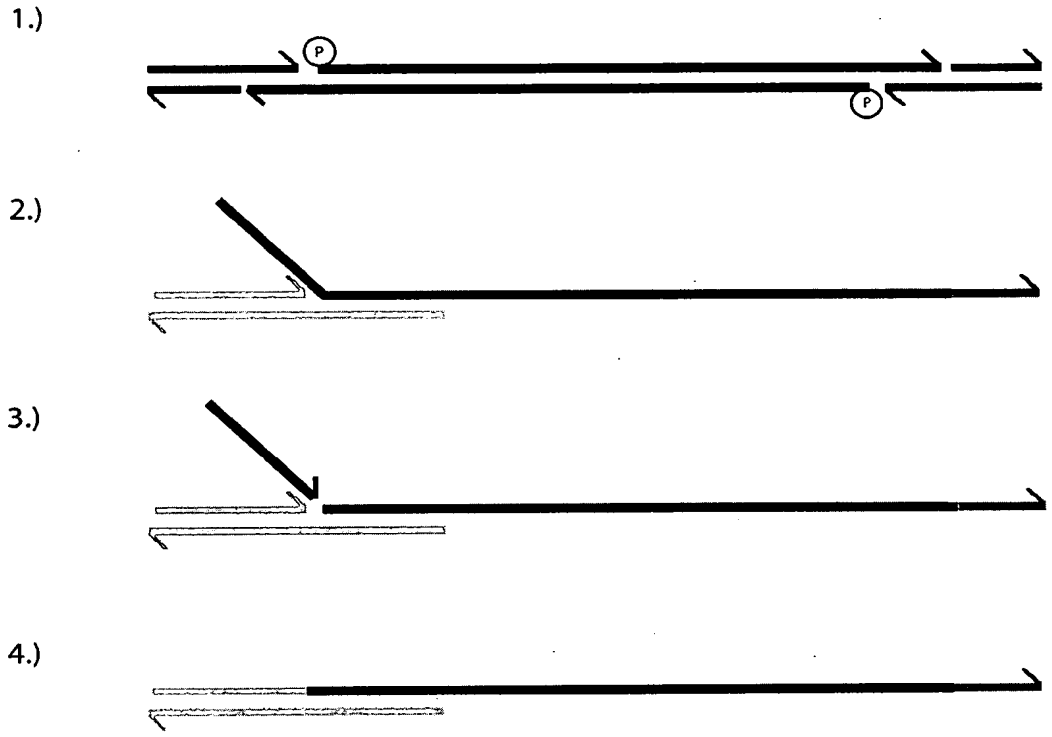
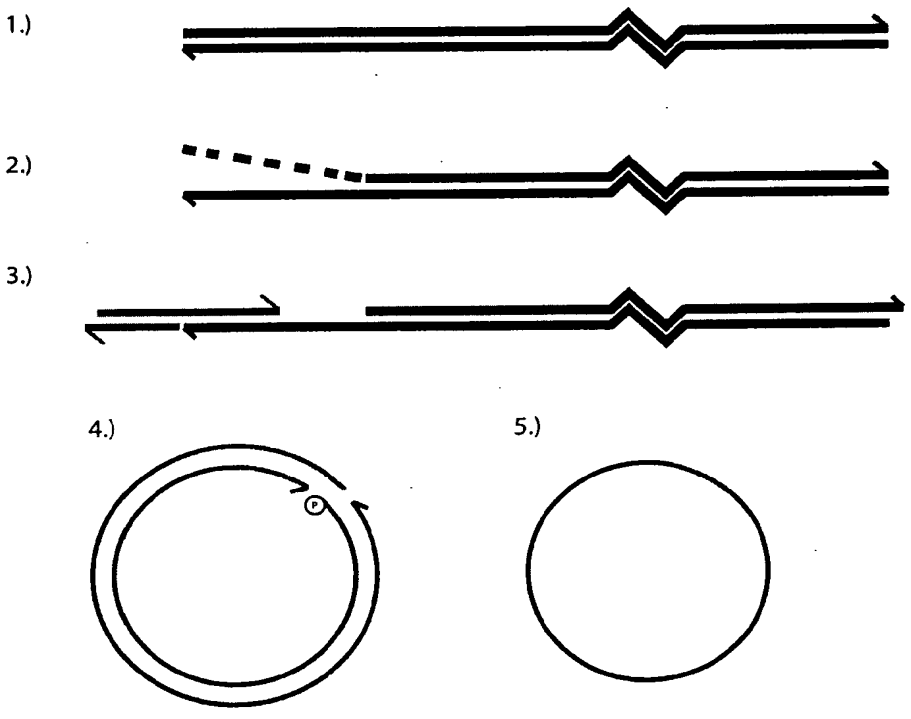


Figure 4



IDT Exhibit 1020.1287

Figure 5

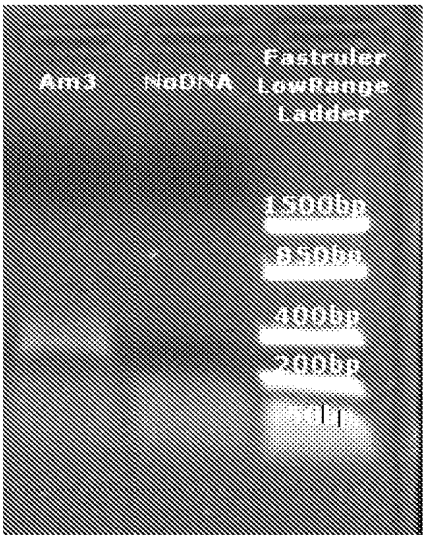
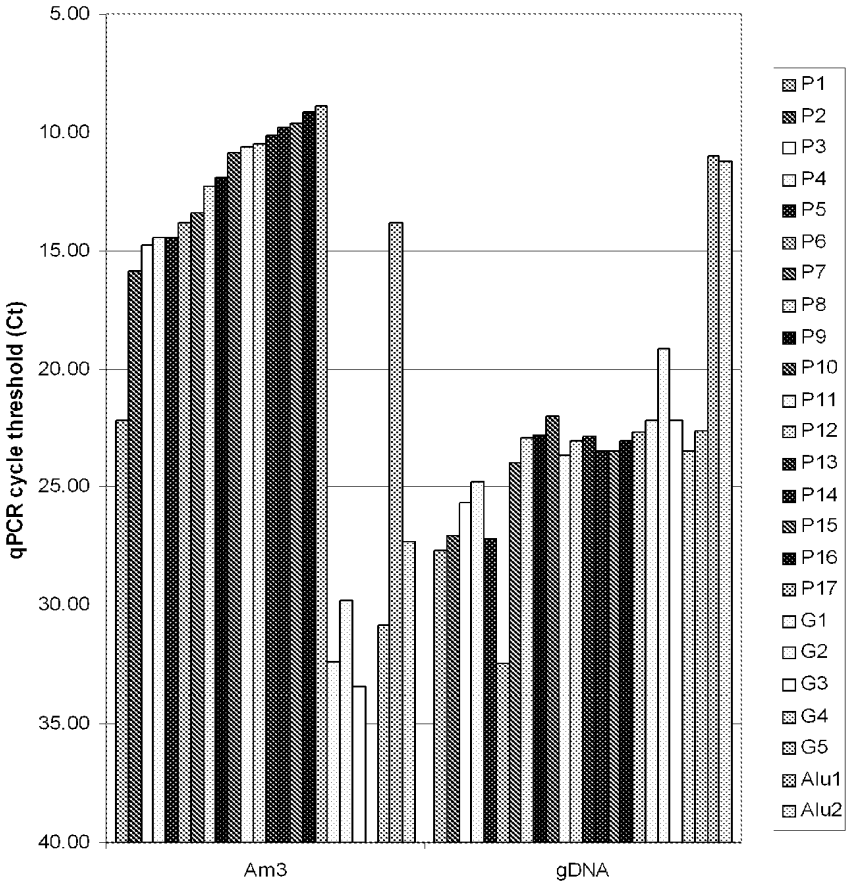


Figure 6



INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2010/060715

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12Q1/68
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, MEDLINE, BIOSIS, EMBASE, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2008/153492 A1 (JOHANSSON HENRIK [SE]; ERICSSON OLOF [SE]) 18 December 2008 (2008-12-18) the whole document	1-34
Y	WO 2005/111236 A1 (OLINK AB [SE]; DAHL FREDRIK [SE]) 24 November 2005 (2005-11-24) the whole document	1-39
Y	WO 98/40518 A2 (WISCONSIN ALUMNI RES FOUND [US]; GUILFOYLE RICHARD A [US]; GUO ZHEN [U) 17 September 1998 (1998-09-17) the whole document	1-39

-/--

[X] Further documents are listed in the continuation of Box C.

[X] See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
16 September 2010	22/09/2010

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Botz, Jürgen

Form PCT/ISA/210 (second sheet) (April 2005)

IDT Exhibit 1020.1289

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	EP 1 350 853 A1 (ID LELYSTAD INST DIERHOUDERIJ [NL]) 8 October 2003 (2003-10-08) the whole document	1-39
Y	US 2006/292597 A1 (SHAPERO MICHAEL H [US] ET AL) 28 December 2006 (2006-12-28) the whole document	1-38
Y	CALLOW M J ET AL: "Selective DNA amplification from complex genomes using universal double-sided adapters" NUCLEIC ACIDS RESEARCH, OXFORD UNIVERSITY PRESS, SURREY, GB LNKD-DOI:10.1093/NAR/GNH019, vol. 32, no. 2, 28 January 2004 (2004-01-28), pages 1-6, XP002991601 ISSN: 0305-1048 the whole document	1-39

IDT Exhibit 1020.1290

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/EP2010/060715

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2008153492 A1	18-12-2008	EP 2152901 A1	17-02-2010
WO 2005111236 A1	24-11-2005	EP 1747285 A1	31-01-2007
		JP 2007537751 T	27-12-2007
WO 9840518 A2	17-09-1998	AU 746620 B2	02-05-2002
		AU 7465198 A	29-09-1998
		CA 2284121 A1	17-09-1998
		EP 0970250 A1	12-01-2000
		US 6228999 B1	08-05-2001
		US 6280948 B1	28-08-2001
		US 5994068 A	30-11-1999
EP 1350853 A1	08-10-2003	AU 2003219619 A1	27-10-2003
		CA 2481417 A1	23-10-2003
		WO 03087409 A1	23-10-2003
		US 2005208513 A1	22-09-2005
US 2006292597 A1	28-12-2006	US 2009075345 A1	19-03-2009

IDT Exhibit 1020.1291

Espacenet

Bibliographic data: WO9318052 (A1) — 1993-09-16

COVALENTLY CROSS-LINKED OLIGONUCLEOTIDES

Inventor(s): COOK PHILLIP DAN [US]; MANOHARAN MUTHIAN [US]; BRUICE THOMAS [US] + (COOK, PHILLIP, DAN, ; MANOHARAN, MUTHIAN, ; BRUICE, THOMAS)

Applicant(s): ISIS PHARMACEUTICALS INC [US] + (ISIS PHARMACEUTICALS, INC)

Classification: **- international:** ***C07H21/00; C07H21/04; C12N15/09;*** (IPC1-7): C07H21/04
- cooperative: **C07H21/00**

Application number: WO1993US02059 19930305 Global Dossier

Priority number (s): US19920846376 19920305

Also published as: AT212998 (T) CA2131311 (A1) CA2131311 (C) DE69331543 (T2) EP0635023 (A1) more

Abstract of WO9318052 (A1)

Covalent cross-linkages for two oligonucleotide strands or for first and second regions of a single oligonucleotide strand connect sugar moieties of nucleotides on the respective strands or the regions of the single strand. The cross-linkages are connected to at least one strand or region via a space-spanning group. The cross-linkage also can be connected to the other strand or other region via a space-spanning group or via an abasic site located on the other strand or other region.

IDT Exhibit 1020.1292

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 : C07H 21/04 — A1

(11) International Publication Number: WO 93/18052

(43) International Publication Date: 16 September 1993 (16.09.93)

(21) International Application Number: PCT/US93/02059

(22) International Filing Date: 5 March 1993 (05.03.93)

(30) Priority data:
846,376 — 5 March 1992 (05.03.92) — US

(60) Parent Application or Grant
(63) Related by Continuation
US — 846,376 (CIP)
Filed on — 5 March 1992 (05.03.92)

(71) Applicant *(for all designated States except US):* ISIS PHARMACEUTICALS, INC. [US/US]; 2280 Faraday Avenue, Carlsbad, CA 92008 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only) :* COOK, Phillip, Dan [US/US]; 7340 Bolero Street, Carlsbad, CA 92009 (US). MANOHARAN, Muthian [US/US]; 2340 Hosp Way, #119, Carlsbad, CA 92008 (US). BRUICE, Thomas [US/US]; 6880 Shearwater Drive, Carlsbad, CA 92009 (US).

(74) Agents: CALDWELL, John, W. et al.; Woodcock Washburn Kurtz MacKiewicz & Norris, One Liberty Place - 46th Floor, Philadelphia, PA 19103 (US).

(81) Designated States: AU, BB, BG, BR, CA, CZ, FI, HU, JP, KP, KR, LK, MG, MN, MW, NO, NZ, PL, RO, RU, SD, SK, UA, US, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, SN, TD, TG).

Published
With international search report.
With amended claims and statement.

(54) Title: COVALENTLY CROSS-LINKED OLIGONUCLEOTIDES

(57) Abstract

Covalent cross-linkages for two oligonucleotide strands or for first and second regions of a single oligonucleotide strand connect sugar moieties of nucleotides on the respective strands or the regions of the single strand. The cross-linkages are connected to at least one strand or region via a space-spanning group. The cross-linkage also can be connected to the other strand or other region via a space-spanning group or via an abasic site located on the other strand or other region.

IDT Exhibit 1020.1293

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	**FR**	France	**MR**	Mauritania
AU	Australia	**GA**	Gabon	**MW**	Malawi
BB	Barbados	**GB**	United Kingdom	**NL**	Netherlands
BE	Belgium	**GN**	Guinea	**NO**	Norway
BF	Burkina Faso	**GR**	Greece	**NZ**	New Zealand
BG	Bulgaria	**HU**	Hungary	**PL**	Poland
BJ	Benin	**IE**	Ireland	**PT**	Portugal
BR	Brazil	**IT**	Italy	**RO**	Romania
CA	Canada	**JP**	Japan	**RU**	Russian Federation
CF	Central African Republic	**KP**	Democratic People's Republic of Korea	**SD**	Sudan
CG	Congo			**SE**	Sweden
CH	Switzerland	**KR**	Republic of Korea	**SK**	Slovak Republic
CI	Côte d'Ivoire	**KZ**	Kazakhstan	**SN**	Senegal
CM	Cameroon	**LI**	Liechtenstein	**SU**	Soviet Union
CS	Czechoslovakia	**LK**	Sri Lanka	**TD**	Chad
CZ	Czech Republic	**LU**	Luxembourg	**TG**	Togo
DE	Germany	**MC**	Monaco	**UA**	Ukraine
DK	Denmark	**MG**	Madagascar	**US**	United States of America
ES	Spain	**ML**	Mali	**VN**	Viet Nam
FI	Finland	**MN**	Mongolia		

IDT Exhibit 1020.1294

COVALENTLY CROSS-LINKED OLIGONUCLEOTIDES

FIELD OF THE INVENTION

This invention is directed to covalent cross-linkages for oligonucleotides. The cross-linkages connect either two oligonucleotide strands or different regions of a single oligonucleotide strand via a space-spanning group that extends from a sugar moiety on a nucleotide of the first strand or strand region to a nucleotide on the other strand or strand region.

BACKGROUND OF THE INVENTION

Several examples have been reported of cross-linking between complementary oligonucleotides. The purpose of such cross-linking often has been to disrupt the function of a nucleic acid, primarily by cleavage of the target. It is known to cross-link nucleic acids in a positionally uncontrolled manner with UV exposure. However, Grineva, and Karpova, *FEBS.*, **1973**, *32*, 351-355, appear to have been the first to covalently cross-link complementary strands of oligonucleotides at a specific site utilizing controlled chemistry. A nitrogen mustard was attached to the 3′ terminal ribose unit of an oligonucleotide via an acetal linkage or to the 5′ end of an oligonucleotide via a phosphoramide linkage. On hybridization, the reactive mustards covalently cross-linked to the complementary strand via alkylation of the ternary heteroaromatic nitrogen atom at the 7-position of guanine or adenine. However, the alkylated base thus formed is a quaternary charged species that is subject to rapid chemical degradation via

imidazole ring opening followed by cleavage of the targeted strand.

Enzymes are known to remove alkylated bases from nucleic acids. Indeed, such enzymatic removal is implicated in DNA repair mechanisms. The destruction of a targeted alkylated nucleic acid theoretically could be utilized for antimicrobial or antitumor therapy. However, cross-linking alkylation reactions and subsequent degradations are not desirable where host cell viability is necessary.

Summerton and Bartlett have proposed another type of alkylated oligonucleotide wherein an α-bromomethylketone attached to the 4-position of a cytidine nucleotide spans the major groove and alkylates the 7-position of a complementary guanine residue in a targeted strand. [*See, J. Mol. Biol.*, **1978**, *122*, 145-162; *J. Theor. Biology*, **1979**, *78*, 61-75; and United States Patent 4,123,610]. As with the nitrogen mustards noted above, this alkylation yields a charged species that effects cleavage of the targeted nucleic acid strand following alkylation.

A further example of a covalent cross-linkage to the 7-position nitrogen of guanine is described by Meyer *et. al.*, *J. Am. Chem. Soc.*, **1989**, *111*, 8517. The authors attached an iodoacetamidopropyl moiety to the 5-position of a thymidine nucleotide of DNA. The iodoacetamidopropyl moiety subsequently alkylated the 7-position of a guanine nucleotide at a position two base pairs down the complementary strand. Cleavage of the targeted strand was observed at various times and temperatures.

Matteucci and Webb have described a hybridization triggered cross-linking process where an N6,N6-ethano- adenine or N4,N4-ethanocytosine alkylates an appropriately positioned nucleophile in a complementary strand. [*See, Nucleic Acids Res.*, **1986**, *14*, 7661; *Tetrahedron Letters*, **1987**, *28*, 2469-2472.] This process has been designed to inactivate the normal function of the targeted DNA either by forming a stable adduct or by hydrolytic or enzymatic cleavage.

A cross-linkage similar to the above-noted N6,N6-ethano linkage was described by Ferentz, *et al.*, *J. Am. Chem.*

Soc., **1991**, *113*, 4000, who utilized either an ethyl disulfide or a propyl disulfide linkage to connect the N6-positions of adenine residues on a self-complementary oligonucleotide.

Lee, *et al.*, *Biochemistry*, **1988**, *27*, 3197-3203, described the interaction of psoralen-derivatized methyl phosphonate oligonucleotides with single-stranded DNA. According to this method, irradiation effects cross-linking which, in turn, inactivates DNA. Psoralen cross-linkages occur only between pyrimidine bases to join single-stranded or, more commonly, double-stranded structures.

The generation of an abasic site on a nucleic acid has been implicated in the above-noted enzymatic removal of alkylated bases from nucleic acids. Manoharan, *et al.*, *J. Am. Chem. Soc.*, **1987**, *109*, 7217 and Manoharan, *et al.*, *J. Am. Chem. Soc.*, **1988**, *110*, 2690, have characterized abasic sites in oligonucleotides utilizing ^{13}C NMR spectroscopy. Abasic sites also have been created on nucleic acids to attach the intercalator 9-aminoellipticine to nucleic acids. Vasseur, *et al.*, *Nucleosides & Nucleotides*, **1989**, *8*, 863-866, reported that the trimer Tp(Ap)Pt, where (Ap) is an apurinic site yielded a mixture of products when reacted with the intercalator 9-aminoellipticine. The reaction produced a desired Schiff's base adduct between the apurinic site and the 9-aminoellipticine, but also resulted in cleavage of the phosphate backbone. This work was extended by Bertrand, *et al.*, *Nucleic Acids Research*, **1989**, *17*, 10307, to the interaction of 9-aminoellipticine and a 13-mer oligonucleotide. The 13-mer oligonucleotide contained only a single purine nucleotide. Acid treatment resulted in loss of the purine base to yield an unstable abasic site. The 9-aminoellipticine was successfully linked to the abasic site in the oligonucleotide by reducing the Schiff's base adduct during the reaction. Using an oligonucleotide having a single purine site, the authors were able to study the coupling of the intercalator 9-aminoellipticine to the oligonucleotide. It is unlikely, however, that oligonucleotides which bind to sequences of biological significance contain only a single purinic site.

Other abasic site-containing oligonucleotides have been assembled by synthetically incorporating abasic nucleoside precursors into an oligonucleotide. For example, Philippe, *et al.*, *Tetrahedron Letters*, **1990**, *31*, 6347-6350, incorporated a 2-deoxy-D-ribose having a primary amide function attached via a pentamethylene chain to its 1'-position. 1,2-Dideoxy-D-ribofuranose and butane 1,3-diol likewise have been incorporated into oligonucleotides. [*See*, P.lyer, *et al.*, *Nucleic Acids Research*, **1990**, *18*, 2855.]

Groehke, *et al.*, *Helvetica Chimica Acta*, **1990**, *73*, 608, reported the preparation of a (*tert*-butyl)dimethylsilyl-protected deoxy-D-ribose and its incorporation into oligonucleotides utilizing solid state techniques. Reoc'h, *et al.*, *Tetrahedron Letters*, **1991**, *32*, 207, used 1-(*o*-nitrobenzyl)-2-deoxy-D-ribofuranose in the synthesis of other oligonucleotides. None of these chemically generated abasic sites have been utilized to effect cross-linking.

Application PCT/US91/01822, filed March 19, 1991 and entitled Reagents And Methods For Modulating Gene Expression Through RNA Mimicry, discloses effective use of oligonucleotides that mimic RNA structure. The entire disclosure Application PCT/US91/01822, which is assigned to the assignee of this application, is herein incorporated by reference. The oligonucleotides disclosed in PCT/US91/01822 are selected to mimic a portion of an RNA encoded by a gene. The RNA-mimicking oligonucleotides interfere with or modulate gene expression by interfering with protein-RNA interactions.

In order to mimic an RNA, an oligonucleotide must assume and retain the RNA's secondary and tertiary structure. Although the physical and chemical forces that normally retain the RNA in its secondary and tertiary structure are non-covalent in nature, one way to fix an RNA mimic in the necessary secondary and tertiary structure would be to covalently cross-link the mimic. However, presently-known techniques for cross-linking nucleic acid strands either lead to strand cleavage, disruption of structure or Watson/Crick

hydrogen bonding, or are only useful for limited, specific sequences or at specific locations in a sequence.

Cross-linkages between two oligonucleotide strands or between regions of a single strand would be undesirable where the cross-linkages could destroy the strands or strand or could unduly modify the conformational structure of the oligonucleotide. Cross-linkages also would be undesirable where they do not allow the cross-linked product to approximate the conformation of natural nucleic acids. Heretofore, there has be no suggestions in the art of cross-linkages or methods of cross-linking that do not destroy the strands, that allow suitable conformations, that are useful on various sequences and at various positions within the sequences, and that allow normal ranges of features such as the "tilt" and "propeller twist" features found in naturally occurring nucleic acid duplexes. Accordingly there remains a long-felt need for nucleic acid cross-linkages and methods of cross-linking nucleic acids.

OBJECTS OF THE INVENTION

It is a primary object of this invention to provide improved covalently cross-linked nucleic acids and methods of forming the same.

It is a further object of this invention to provide individual oligonucleotide strands that are covalently bonded to one another at one or more sites on the strands.

It is an even further object of this invention to provide a single oligonucleotide strand that is covalently bonded to itself at one or more sites on the strand.

It is an even further object of this invention to provide RNA mimics that are fixed in specific spatial conformations via cross-linking covalent bonds.

It is an additional object of this invention to provide covalently cross-linked oligonucleotides that have the same sequence as known nuclease-resistant mimics of binding receptors for nucleic acid-binding proteins.

These and other objects of this invention will become apparent to those skilled in the art from a review of this specification and the claims appended hereto.

BRIEF DESCRIPTION OF THE INVENTION

In accordance with this invention there are provided sequence-specific, covalently cross-linked nucleic acids comprising a first nucleotide located either on a first strand of complementary oligonucleotide strands or on a single oligonucleotide strand. The cross-linked nucleic acids further comprise a second nucleotide located on a further strand of the complementary strands or on the single strand at a site distal to the first nucleotide. A first bond means is located on a sugar moiety of the first nucleotide and a second bond means is located on a sugar moiety of the second nucleotide. A covalent cross-linkage connects the first and the second bond means and, in doing so, cross-links the strand or strands.

In one embodiment of the invention, the first and the second nucleotides are located within first and second nucleotide sequences, respectively, on a single oligonucleotide strand. In a first preferred embodiment, the second nucleotide sequence is separated from the first nucleotide sequence by a distance that allows the oligonucleotide strand to assume a conformation wherein the first and second nucleotide sequences are specifically hybridizable with one another. In a further preferred embodiment, the first and the second nucleotides are located in non-hybridizable positions but are in sufficient proximity to allow cross-linking between the first and second nucleotides.

In a preferred embodiment of the invention, the first bond means includes an abasic site and the second bond means includes a space-spanning group that, in turn, includes an active functional group that is capable of covalently bonding with the abasic site.

Further in accordance with the invention there are provided sequence-specific, cross-linked oligonucleotides that include a first nucleotide located on a first complementary

oligonucleotide strand, a second nucleotide located on a second complementary strand, a first bond precursor located on a sugar moiety of the first nucleotide, a second bond precursor located on a sugar moiety of the second nucleotide, and a covalent cross-linkage connecting the first and the second bond precursors to cross-link the first and the second strands. In a preferred embodiment of the invention, the first nucleotide is located in a first nucleotide sequence, the second nucleotide is located in a second nucleotide sequence, and the second nucleotide sequence is complementary to and specifically hybridizable with the first nucleotide sequence.

Additionally in accordance with this invention there are provided sequence-specific, cross-linked oligonucleotides that include a first nucleotide located on an oligonucleotide strand and a second nucleotide located on the same oligonucleotide strand. The first and second nucleotides are separated by a distance that allows the nucleotides to specifically hybridize with one another. A first bond precursor is located on a sugar moiety of the first nucleotide and a second bond precursor is located on a sugar moiety of the second nucleotide. A covalent cross-linkage connects the first and the second bond precursors, cross-linking the first and second nucleotides. In a preferred embodiment, the first nucleotide is located in a first nucleotide sequence, the second specific nucleotide is located in a second nucleotide sequence, and the second nucleotide sequence is complementary to and hybridizable with the first nucleotide sequence. In a further preferred embodiment, the first and second nucleotides are located in nonhybridizable positions but are in sufficient proximity that a cross-linkage can be formed between them.

Also in accordance with this invention there are provided cross-linked oligonucleotides comprising a first sequence region that includes an abasic site and a second sequence region that includes a space-spanning group. An active functional group is located on the space-spanning group and is covalently coupled with the abasic site. In a first preferred embodiment, the first sequence region and the second

sequence region are on the same oligonucleotide strand. In a further preferred embodiment, the first sequence region and the second sequence region are on different oligonucleotide strands. In these preferred embodiments, the first sequence region and the second sequence region are complementary to and specifically hybridizable with one another. Preferably, the active functional group is an amine, hydrazine, hydroxylamine, semicarbazide, thiosemicarbazide, hydrazide, thiol or alcohol, and the abasic site includes an aldehydic functionality.

Space-spanning groups according to the invention comprise elongated chains from about 1 to about 50 atoms in length. Space-spanning groups preferably comprise elongated chains from about 4 to about 20 atoms in length, more preferably chains from about 8 to about 13 atoms in length.

Further in accordance with this invention there are provided cross-linked oligonucleotides that include a first sequence region, a second sequence region, and a covalently bonded linker of the structure:

$$X_1-L_1-Y-L_2-X_2$$

wherein X_1 is covalently connected to the first sequence region and X_2 is covalently connected to the second sequence region and where:

(i) X_1 and X_2, independently, are O, S, NH, CH_2 or CO;

L_1 and L_2, independently, are C_1-C_{20} straight chain alkyl, C_1-C_{20} straight chain substituted alkyl, C_2-C_{50} branched chain alkyl, C_2-C_{50} branched chain substituted alkyl, C_2-C_{20} straight chain alkenyl, C_2-C_{20} straight chain substituted alkenyl, C_3-C_{50} branched chain alkenyl, C_3-C_{50} branched chain substituted alkenyl, C_2-C_{20} straight chain alkyne, C_2-C_{20} straight chain substituted alkyne, C_3-C_{50} branched chain alkyne, C_3-C_{50} branched chain substituted alkyne, polyamine,

polyamide, polyester, polyethylene glycol, polyether, aryl, aralkyl or heterocyclic; and

Y is an imine, amine, oxime, hydroxylamine, hydrazine, hydrazone, azine, hydrazide-hydrazone, amide, hydrazide, semicarbazide, semicarbazone, thiosemicarbazide, thiocarbazone, disulfide, hemiacetal, thiohemiacetal, α-keto-alkylthioalkyl or α-keto-alkylaminoalkyl; or

(ii) X_1 and X_2, independently, are O or NH; and

L_1-Y-L_2 together are NH-L_3-NH or NH-NH-L_3-NH-NH; and

L_3 is C_1-C_{20} straight chain alkyl, C_1-C_{20} straight chain substituted alkyl, C_2-C_{50} branched chain alkyl, C_2-C_{50} branched chain substituted alkyl, C_2-C_{20} straight chain alkenyl, C_2-C_{20} straight chain substituted alkenyl, C_3-C_{50} branched chain alkenyl, C_3-C_{50} branched chain substituted alkenyl, C_2-C_{20} straight chain alkyne, C_2-C_{20} straight chain substituted alkyne, C_3-C_{50} branched chain alkyne, C_3-C_{50} branched chain substituted alkyne, polyamine, polyamide, polyester, polyethylene glycol, polyether, aryl, aralkyl or heterocyclic; or

(iii) X_1 and X_2, independently, are O, S, NH, CH_2 or CO;

L_1 and L_2, independently, are C_1-C_{20} straight chain alkylamino, C_1-C_{20} straight chain substituted alkylamino, C_2-C_{50} branched chain alkylamino, C_2-C_{50} branched chain substituted alkylamino, C_2-C_{20} straight chain alkenylamino, C_2-C_{20} straight chain substituted alkenylamino, C_3-C_{50} branched chain alkenylamino, C_3-C_{50} branched chain substituted alkenylamino, C_2-C_{20} straight chain alkynylamino, C_2-C_{20} straight chain substituted alkynylamino, C_3-C_{50} branched chain alkynylamino, C_3-C_{50} branched chain substituted alkynylamino, C_1-C_{20} straight chain alkylthio, C_1-C_{20} straight chain substituted alkylthio, C_2-C_{50} branched chain alkylthio, C_2-C_{50} branched chain substituted

alkylthio, C_2-C_{20} straight chain alkenylthio, C_2-C_{20} straight chain substituted alkenylthio, C_3-C_{50} branched chain alkenylthio, C_3-C_{50} branched chain substituted alkenylthio, C_2-C_{20} straight chain alkynylthio, C_2-C_{20} straight chain substituted alkynylthio, C_3-C_{50} branched chain alkynylthio or C_3-C_{50} branched chain substituted alkynylthio; and

Y is a heterobifunctional or homobifunctional cross-linking reagent;

where said substituent of said substituted alkyl, substituted alkenyl and substituted alkyne moieties is an intercalator, a conjugate, polyamine, polyamide, polyethylene glycol, a group that enhances the pharmacodynamic properties of oligonucleotides or a group that enhances the pharmacokinetic properties of oligonucleotides.

Further in accordance with this invention there are provided cross-linked oligonucleotides that individually include a first sequence region, a second sequence region, a first space-spanning group connected to a 2'-position of a nucleotide located in the first sequence region, a second space-spanning group connected to a 2'-position of a nucleotide in the second sequence region, and a linking moiety covalently connecting the first and second space-spanning groups. In preferred embodiments the linking moiety is:

-N=CH-, -N(R)-, -N(R)-CH_2-N(R)-, -O-N=CH-, -O-N(R)-, -N(R)-N(R)-, -N(R)-N=CH-, -HC=N-N=CH-, -C(O)-N(R)-N=CH-, -C(O)N(R)-, -C(O)-N(R)-N(R)-, -N(R)-C(O)-N(R)-N(R)-, -N(R)-C(O)-N(R)-N(R)=CH-, -N(R)-C(S)-N(R)-N(R)-, -N(R)-C(S)-N(R)-N(R)=CH-, -CH(OH)-O-CH_2-, -CH(OH)-S-CH_2-, -CH_2-S-CH_2-C(O)-CH_2-, -CH_2-NH-CH_2-C(O)-CH_2- or -S-S-;

wherein R is H, alkyl, alkenyl, aralkyl, an intercalator, a conjugate, polyamine, polyamide, polyethylene glycol, a group that enhances the pharmacodynamic properties of oligonucleotides or a group that enhances the pharmacokinetic properties of oligonucleotides.

Further in accordance with this invention there are provided methods for fixing a sequence-specific oligonucleotide in a set configuration. Such methods include selecting a first site in the oligonucleotide, said first site having a first nucleotide sequence. The methods further include selecting a second site in the oligonucleotide, said second site having a second nucleotide sequence. A first bond precursor is attached to a sugar moiety of a first nucleotide of the first site, while a second bond precursor is attached to a sugar moiety of a second nucleotide of the second site. A covalent cross-linkage is formed between the first and the second bond precursors. In a preferred embodiment, the first bond precursor is attached to the first site by reacting a first nucleotide precursor with a reagent for introducing the first bond precursor on a sugar moiety of the first nucleotide precursor. The first nucleotide precursor is derivatized to a protected and activated form and is synthetically incorporated into the first specific nucleotide sequence. The second nucleotide sequence can be complementary to and hybridizable with the first nucleotide sequence.

In one embodiment of this method, the first bond precursor is an abasic nucleotide and the second bond precursor is attached to a 2'-position of the second nucleotide. In this embodiment, the second bond precursor is reacted with the abasic site to form an imine, oxime, hydrazone, hydrazide-hydrazone, semicarbazone, thiosemicarbazone, hemiacetal or thiohemiacetal linkage between the first nucleotide sequence and the second nucleotide sequence. Certain of these linkages can be reduced to amine, hydroxylamine, hydrazine, hydrazide, semicarbazide or thiosemicarbazide linkages.

In a further embodiment of the method, the first bond precursor is attached to a 2'-position of the first nucleotide and the second bond precursor is attached to a 2'-position of the second nucleotide. Preferred linkages for this embodiment include:

-N=CH-, -N(R)-, -N(R)-CH_2-N(R)-, -O-N=CH-,
-O-N(R)-, -N(R)-N(R)-, -N(R)-N=CH-, -HC=N-N=CH-,

-C(O)-N(R)-N=CH-, -C(O)N(R)-, -C(O)-N(R)-N(R)-, -N(R)-C(O)-N(R)-N(R)-, -N(R)-C(O)-N(R)-N(R)=CH-, -N(R)-C(S)-N(R)-N(R)-, -N(R)-C(S)-N(R)-N(R)=CH-, -CH(OH)-O-CH_2-, -CH(OH)-S-CH_2-, -CH_2-S-CH_2-C(O)-CH_2-, -CH_2-NH-CH_2-C(O)-CH_2- or -S-S-;

wherein R is H, alkyl, alkenyl, aralkyl, an intercalator, a conjugate, polyamine, polyamide, polyethylene glycol, a group that enhances the pharmacodynamic properties of oligonucleotides or a group that enhances the pharmacokinetic properties of oligonucleotides.

Further in accordance with this invention there are provided methods for cross-linking a single sequence-specific oligonucleotide strand. These methods include selecting first and second nucleotide sequences in the oligonucleotide, incorporating an abasic nucleotide within the first nucleotide sequence, attaching a bond precursor to one of the nucleotides of the second sequence, and forming a covalent cross-linkage between the abasic site and the bond precursor. In one embodiment of this method, the bond precursor is attached to a sugar moiety of the nucleotide of the second sequence, preferably at the 2'-position. In preferred embodiments, the abasic site includes an aldehydic functional group and the bond precursor includes a functional group capable of forming an oxime, imine, hydrazone, hydrazide-hydrazone, semicarbazone, thiosemicarbazone, hemiacetal or thiohemiacetal linkage with the aldehydic functional group. The second nucleotide sequence preferably is selected to be complementary to and specifically hybridizable with the first nucleotide sequence.

Further methods for cross-linking a single, sequence-specific oligonucleotide include selecting a first nucleotide sequence and a second nucleotide sequence in the oligonucleotide. A first bond precursor is attached to a sugar moiety of a nucleotide of the first nucleotide sequence and a second bond precursor is attached to a sugar moiety of a nucleotide of the second nucleotide sequence. A covalent cross-linkage is formed between the first and the second precursors. In preferred embodiments, the cross-linkage is an imine, amine, aminoalkyl-

amine, oxime, hydroxylamine, hydrazine, hydrazone, azine, hydrazide-hydrazone, amide, hydrazide, semicarbazide, semicarbazone, thiosemicarbazide, thiocarbazone, disulfide, hemiacetal, thiohemiacetal, α-keto-alkylthioalkyl or α-keto-alkylaminoalkyl cross-linkage. In other preferred embodiments the second nucleotide sequence is complementary to and hybridizable with the first nucleotide sequence.

Further in accordance with this invention there are provided methods for cross-linking first and second strands of oligonucleotides, comprising the steps of selecting a first nucleotide sequence in a first oligonucleotide strand and selecting a second nucleotide sequence in a second oligonucleotide strand. The methods further include incorporating an abasic nucleotide within the first nucleotide sequence, attaching a bond precursor to a nucleotide of the second sequence, and forming a covalent cross-linkage between the abasic site and the bond precursor. In preferred embodiments, the second nucleotide sequence is complementary to and hybridizable with the first nucleotide sequence.

Further, in accordance with this invention there are provided methods for cross-linking first and second strands of oligonucleotides comprising the steps of selecting a first nucleotide sequence in a first oligonucleotide strand, selecting a second nucleotide sequence in a second oligonucleotide strand, attaching a first bond precursor to a sugar moiety of a nucleotide of the first sequence, locating a second covalent precursor on a sugar moiety of a nucleotide of the second sequence, and forming a covalent cross-linkage between the first and the second bond precursors. In preferred embodiments, the second nucleotide sequence is complementary to and hybridizable with the first nucleotide sequence.

DETAILED DESCRIPTION OF THE INVENTION

In accordance with the objects of this invention novel, covalently cross-linked oligonucleotide strands are disclosed. Preferably such strands are "duplexed" or "complemen-

tary" along at least a portion of their length. In the context of this invention the terminology "duplex strands" or "complementary strands" or like terminology refers in a first instance to two separate strands wherein at least one nucleotide or nucleotide sequence on one of the strands is complementary to and specifically hybridizable with an opposing nucleotide or nucleotide sequence on the other strand. In a second instance, this terminology refers to two sections or sequence regions of a single strand that are spaced from one another by a distance sufficient for the strand to fold onto itself and permit hybridization of complementary nucleotides in the respective sequence regions. Folded, self-hybridizing, single strands form hairpin loop, stem loop, interior loop or bulge type structures as a result of such hybridization.

Further in the context of this invention, the term "oligonucleotide" refers to polynucleotides formed from a plurality of joined nucleotide units. In certain instances, the nucleotides are formed from naturally occurring bases and pentofuranosyl sugars groups and are joined together via native inter-nucleoside phosphodiester linkages. The term "oligonucleotide" thus effectively includes naturally occurring or synthetic species formed from naturally occurring subunits. The term "oligonucleotide" also includes structures formed from modified or non-naturally occurring nucleotide subunits. Modifications can occur on the base portion of a nucleotide, on a sugar portion of a nucleotide or on the linkage joining one nucleotide to the next. In addition, modifications can be made wherein nucleoside units are joined together with connecting groups that substitute for the inter-nucleoside phosphate linkages. Inter-nucleoside connecting groups, which are to be distinguished from the cross-linkages of the invention, include $-O-CH_2-CH_2-O-$ linkages and other novel linkages disclosed in U.S. Patent Application Serial No. 566,836, filed August 13, 1990, entitled Novel Nucleoside Analogs, and U.S. Patent Application Serial No. 703,619, filed May 21, 1991, entitled Backbone Modified Oligonucleotide Analogs, both assigned to the assignee of this invention. Other modifications can be made to

the sugar, to the base, or to the phosphate group of the nucleotide. Exemplary modifications are disclosed in the following U.S. Patent Applications: Serial No. 463,358, filed January 11, 1990, entitled Compositions And Methods For Detecting And Modulating RNA Activity; Serial No. 566,977, filed August 13, 1990, entitled Sugar Modified Oligonucleotides That Detect And Modulate Gene Expression; Serial No. 558,663, filed July 27, 1990, entitled Novel Polyamine Conjugated Oligonucleotides; Serial No. 558,806, filed July 27, 1991, entitled Nuclease Resistant Pyrimidine Modified Oligonucleotides That Detect And Modulate Gene Expression and serial number PCT/US91/00243, filed January 11, 1991, entitled Compositions and Methods For Detecting And Modulating RNA Activity, all assigned to the assignee of this invention. The disclosures of all of the above-noted patent applications are herein incorporated by reference.

The term oligonucleotides, as used in connection with this invention, refers to structures including modified portions such as modified sugar moieties, modified base moieties or modified sugar linking moieties. These modified portions function in a manner similar to natural bases, natural sugars and natural phosphodiester linkages. Thus, in the context of this invention, oligonucleotides may have altered base moieties, altered sugar moieties or altered inter-sugar linkages. Exemplary among these are: phosphorothioate, phosphorodithioate, methyl phosphonate, phosphotriester or phosphoramidate inter-nucleoside linkages used in place of phosphodiester inter-nucleoside linkages; deaza or aza purines and pyrimidines used in place of natural purine and pyrimidine bases; pyrimidine bases having substituent groups at the 5- or 6-positions; purine bases having altered or replacement substituent groups at the 2-, 6- or 8- positions; or sugars having substituent groups at their 2'-position, substitutions for one or more of the hydrogen atoms of the sugar, or carbo-cyclic or acyclic sugars. Oligonucleotides may also comprise other modifications consistent with the spirit of this invention. Such oligonucleotides are best described as being

functionally interchangeable with yet structurally distinct from natural oligonucleotides. All such oligonucleotides are comprehended by this invention so long as they effectively mimic the function of a desired RNA or DNA strand.

Further in the context of this invention, it will be recognized that the terms "cross-linkage," "cross-link" or "cross-linking" specifically exclude those portions of the oligonucleotide structure that linearly connect individual nucleoside subunits or nucleoside-like, modified subunits, *i.e.*, those portions that connect the nucleoside units in the sequence that defines the primary structure of the oligonucleotide. Thus, "cross-linkage," "cross-link" or "cross-linking" exclude native phosphodiester linkages that connect nucleosides, as well as analogous, non-native inter-nucleoside linkages such as phosphorothioate, phosphorodithioate methyl phosphonate, phosphotriester, phosphoramidate, and $-O-CH_2-CH_2-O-$ linkages used in place of phosphodiester inter-nucleoside linkages. The terms "cross-linkage," "cross-link" or "cross-linking" as used in the context of this invention refer to "additional" linkages created across a single oligonucleotide strand or several oligonucleotide strands to hold the strand or strands in secondary or other, higher-ordered structures. Primary, secondary, and higher-order structures are as defined in standard reference texts such as *Nucleic Acids In Chemistry And Biology*, edited by G. Michael Blackburn and Michael J. Gait, Oxford University Press, New York, 1991.

Cross-linking is employed in accordance with the present invention to fix separate oligonucleotide strands in duplex structures or to fix a single oligonucleotide strand in a hairpin loop, stem loop, interior loop, bulge or other similar higher-order structure. Such fixation is believed to confer nuclease resistance to oligonucleotides and to optimize structure-dependent function. Fixing a strand or strands in a duplex structure also can disrupt the normal function of single-strand nucleic acid-binding proteins by forming nuclease-resistant mimics of the proteins' binding receptors.

Fixing a strand or strands in set secondary structures also makes it possible to mimic the RNA secondary structures found in diseased cells, particularly cells infected with viruses and retroviruses. A detailed delineation of such RNA mimicry is disclosed in Application PCT/US91/01822, filed March 19, 1991, entitled Reagents and Methods For Modulating Gene Expression Through RNA Mimicry, assigned to the same assignee as this application. The entire contents of Application PCT/US91/01822 is herein incorporated by reference.

The "fixing" of multiple oligonucleotide strands or a single oligonucleotide strand in defined secondary or other higher-order structure is achieved in accordance with this invention by cross-linking the oligonucleotide strands or oligonucleotide. In one preferred embodiment, novel cross-linkages connect a nucleotide or nucleoside pentofuranosyl moiety on one strand with a nucleotide or nucleoside pentofuranosyl moiety on another strand. In another preferred embodiment, novel cross-linkages connect a nucleotide or nucleoside pentofuranosyl moiety on a first region of a single strand with a nucleotide or nucleoside pentofuranosyl moiety on a further region of the same strand. It is believed that such "sugar to sugar" cross-linkages were heretofore unknown. In a further embodiment of the invention, novel cross-linkages connect an abasic site on one oligonucleotide strand to a further site on another strand. Also, novel cross-linkages can connect an abasic site on a first region of a single strand with another site on another region of the single strand. It is believed that use of an abasic site as the attachment site for a oligonucleotide cross-linkage was also heretofore unknown.

In a further embodiment of the invention, novel cross-linkages of the invention connect a terminal sugar on one oligonucleotide strand or a first region of a single strand with a further site on a further strand or to a further region of a single strand.

One preferred cross-linkage of the invention connects the 2'-position of a ribose sugar on one oligonucleotide strand

(or a first region of a single strand) to the 2'-position of a ribose sugar on another strand (or a further region of a single strand). This cross-linkage can be considered a 2'-ribose to 2'-ribose cross-linkage. A second preferred cross-linkage connects a 2'-ribose position on one oligonucleotide strand (or a first region of a single strand) to a C1'-ribose position (an abasic site) on another strand (or a further region of a single strand). This linkage can be considered a 2'-ribose to C1'-ribose cross-linkage. A third preferred cross-linkage connects a 2'-ribose position on one strand (or a first region of a single strand) to a seco-ribose on the 3'-terminal end of another strand (or a further region of a single strand). This linkage can be considered a 2'-ribose to seco-ribose linkage. A fourth preferred cross-linkage connects a 2'-ribose position on one strand (or a first region of a single strand) to 2' and 3' hydroxyls of a nucleoside sugar on the 3'-terminal end of another strand (or a further region of a single strand). This linkage can be considered a 2'-ribose to 2',3'-OH-ribose linkage.

For 2'-ribose to 2'-ribose cross-linkages, presently preferred are: cross-linkages between the 2'-hydroxyl groups (or other 2' substituent groups) of the nucleotide sugar moieties of a Watson/Crick nucleotide base pair; a cross-linkage between a 2'-hydroxyl group (or other 2' substituent group) of a nucleotide sugar moiety and a 2'-hydroxyl (or other 2' substituent group) of a nucleotide sugar moiety one base removed (in the 5'-direction) of the opposite strand; a cross-linkage between nucleotides one helical loop removed; or cross-linkage between nucleotides that are separated from each other along the strand or strands but that are located proximal or adjacent to each other in the higher-ordered structure. Molecular models and computer modeling indicate that the 2'-hydroxyls of RNA-RNA duplexes are exposed in the minor groove or minor side. Thus, direct covalent cross-linkage of the 2'-hydroxyls with a linker of appropriate length should not disturb the natural conformation of the duplexes.

As per the teachings of this invention, RNA-RNA, DNA-DNA or DNA-RNA type duplexes can be covalently cross-linked via a space-spanning group attached to nucleotide units at desired attachment sites. The length of the space-spanning group on each strand or region of a single strand can be between 1 to about 25 atoms. Thus, when two space-spanning groups are cross-linked together, the total length of the space-spanning groups, the cross-linker, and any other connecting atoms should be less than about 50 atoms in length.

The space-spanning groups are composed of carbon atoms or of carbon atoms plus one or more heteroatoms such as nitrogen, sulfur or oxygen. It is preferred that the space-spanning group on an individual strand or on a sequence region of a single strand be about 20 atoms or less in length. More preferably, the cross-linking group is from 8 to 13 atoms in length.

For 2'-ribose to C1'-ribose cross-linkages, a cross-linkage is formed between a 2'-position of a nucleotide sugar moiety on one strand (or on a first region of a single strand) and an abasic 1'-site on another strand (or on a further region of a single strand). The cross-linking group preferably is from 8 to 13 atoms in length.

For 2'-ribose to seco-ribose cross-linkages, a cross-linkage is formed between a 3'-terminal ribofuranosyl nucleoside of a first strand and a reactive functionality on a space-spanning group attached to the sugar moiety of a nucleotide on another strand (or on a further region of the first strand). Preferably, such cross-linkage is made through 2' and 3' hydroxyl groups of the 3'-terminal nucleoside. The 2' and 3' hydroxyl groups are first converted to a 2',3'-dialdehyde structure with periodate. A cross-linkage is formed upon hybridization with a complementary strand that has a connecting or space-spanning group bearing a functional group capable of reacting with the aldehydic functionality. If a 2'-aminoalkoxy connecting or space-spanning group is provided on the 2'-ribose strand, reaction of the aldehydic functionality of the seco-ribose strand should form a Schiff's

base. Reduction of this Schiff's base with sodium cyanoborohydride affords a stable cross-linkage.

For 2'-ribose to 2',3'-OH-ribose cross-linkages, a cross-linkage is formed between 2' and 3' hydroxyls of a 3'-terminal ribofuranosyl nucleoside of a first strand and an aldehyde functionality on a space-spanning group attached to a nucleotide sugar moiety on another strand (or on a further region of the first strand). On hybridization of the strands (or strand), the aldehyde functionality on the space-spanning group reacts with the 2',3'-hydroxyls to form an acetal linkage.

Oligonucleotides in accordance with this invention preferably comprise from about 3 to about 100 subunits. It is preferred that such oligonucleotides comprise greater than about 6 subunits, more preferably from about 8 to about 60 subunits. Oligonucleotides having from about 10 to about 30 subunits are still more preferred. It will be recognized that a subunit comprises a base moiety linked to a sugar moiety. A subunit typically is bound to adjacent subunits through internucleoside bonds. As indicated in the above-referenced United States patent applications, oligonucleotides have diagnostic, therapeutic, and prophylactic applications and also can be used as research reagents. For therapeutic use, cross-linked oligonucleotide analogs can be administered to animals, including humans, suffering from a disease state that it is desirous to treat.

It is generally preferred to apply a therapeutic agent of the invention internally such as orally, intravenously or intramuscularly with or without pharmacologically acceptable carriers. Other forms of administration, such as transdermal, topical or intralesional administration, may also be useful. Cross-linked oligonucleotides also can be included in suppositories.

In certain embodiments, a space-spanning group bearing an active functional group is located on one strand of a pair of duplexed oligonucleotide strands. Preferably, the space-spanning group is attached to a 2'-position of a nucleotide on

the strand. The functional group is positioned on the space-spanning group at a point distal from the point at which the space-spanning group connects to the nucleotide. The space-spanning group is connected to the 2'-position via one or more connecting atoms. Preferably, the connecting atoms includes O, S, NH, CH_2 or CO.

The space-spanning group is preferably a linear or straight chain group. In one embodiment of the invention, the space-spanning group comprises a methylene chain from 1 to about 20 carbon atoms in length. Where the connecting atoms are methylene groups, the connecting group and the space-spanning group together form an alkyl chain. In other embodiments, one or more unsaturated sites are located in the space-spanning group, resulting in alkenyl or alkynyl chains. In these embodiments the alkenyl or alkynyl space-spanning groups have from 2 to about 20 carbon atoms.

In other embodiments the space-spanning group further includes branched chains and substituent groups that extend from a backbone or primary straight chain. Branched chains can be selected from branched alkyl, branched alkenyl or branched alkyne space-spanning groups. The branched chains, in addition to the carbon atoms of the main chain, have from 1 to about 30 carbon atoms. Further substituent atoms or substituent groups can also extend from the backbone or from any branched chains. Thus, the space-spanning group can include C_1-C_{20} straight chain alkyl, C_1-C_{20} straight chain substituted alkyl, C_2-C_{50} branched chain alkyl, C_2-C_{50} branched chain substituted alkyl, C_2-C_{20} straight chain alkenyl, C_2-C_{20} straight chain substituted alkenyl, C_3-C_{50} branched chain alkenyl, C_3-C_{50} branched chain substituted alkenyl, C_2-C_{20} straight chain alkyne, C_2-C_{20} straight chain substituted alkyne, C_3-C_{50} branched chain alkyne and C_3-C_{50} branched chain substituted alkyne groups. Other space-spanning groups include aryl, aralkyl and heterocyclic groups.

Representative examples of the space-spanning groups include but are not limited to methyl, ethyl, propyl, butyl, pentyl, hexyl, heptyl, octyl, nonyl, decyl, undecyl, dodecyl,

tridecyl, tetradecyl, pentadecyl, hexadecyl, heptadecyl, octadecyl, nonadecyl and eicosyl straight chained alkyl groups; 2-methylpropyl, 2-methyl-4-ethylbutyl, 2,4-diethylpropyl, 3-propylbutyl, 2,8-dibutyldecyl, 6,6-dimethyloctyl, 6-propyl-6-butyloctyl, 2-methylbutyl, 2-methylpentyl, 3-methylpentyl, 2-ethylhexyl, 3-penten-2-one, 3-methyl-2-butanol, 2-cyanooctyl, 3-methoxy-4-heptanal, 3-nitrobutyl, 4-isopropoxydodecyl, 4-azido-2-nitrodecyl, 5-mercaptononyl, 4-amino-1-pentenyl branched or substituted groups; allyl, crotyl, propargyl, 2-pentenyl unsaturated groups; 4-methylenebenzyl, 1,4-naphthyl, 2,7-anthracyl and 2,6-phenanthryl aryl or aralkyl groups.

Other space-spanning groups include polyamine, polyamide, polyester, polyethylene glycol and polyether, (polyalkoxyalkyl) groups. Polyamines include compounds of the structure [amine functionality-(linker)$_m$]$_n$. Normally the linker is a methylene chain. The amine functionalities can be primary amines, hydrazines, semicarbazides, thiosemicarbazides or similar nitrogenous species. The number of individual units of the polyamine (n) is selected in conjunction with the number of linker atoms (m) such that the total number of atoms is preferably 20 or less. Thus, if [-HN(CH$_2$)$_m$-]$_n$ is selected as the polyamine backbone, n could be from 1 to 10 depending on m. For example, if m is 1, n can range up to 10, and if m is 2, n can range up to 5. If larger nitrogenous species such as hydrazine or semicarbazide are utilized, n and m would be correspondingly smaller to reflect the increased number of atoms contributed by the nitrogenous species.

Polyamides, polyesters and polyethylene glycols according to the invention have structures analogous to the above-described polyamines, except that an amide, ester or alcohol functionality is substituted for the nitrogenous species of the polyamine. Polyether groups also have analogous structures, except that one or more ether oxygen atoms are interspersed in the carbon chains.

Aryl, aralkyl and heterocyclic space-spanning groups are similar to the above-described alkyl, alkenyl and alkynyl groups, except that appropriate aromatic, mixed alkyl-aromatic

or heterocyclic groups are selected. Such space-spanning groups preferably are less than about 20 atoms in length, excluding atoms that are not directly in the space-spanning group chain but are within the structure of the aromatic or heterocyclic ring. Aryl, aralkyl, and heterocyclic space-spanning groups can include up to about 50 atoms.

Substituent groups can be present on the above-described alkyl, alkenyl, alkyne, polyamine, polyamide, polyester, polyethylene glycol, polyether, aryl, aralkyl and heterocyclic space-spanning groups. Substituent groups include but are not limited to halogen, hydroxyl, keto, carboxy, nitrates, nitrites, nitro, nitroso, nitrile, trifluoromethyl, O-alkyl, S-alkyl, NH-alkyl, amino, azido, sulfoxide, sulfone, sulfide, silyl, intercalators, conjugates, polyamines, polyamides, polyethylene glycols, polyethers, groups that enhance the pharmacodynamic properties of oligonucleotides and groups that enhance the pharmacokinetic properties of oligonucleotides. Typical intercalators and conjugates include cholesterols, phospholipids, biotin, phenanthroline, phenazine, phenanthridine, anthraquinone, acridine, fluoresceins, rhodamines, coumarins, and dyes. Halogens include fluorine, chlorine, bromine and iodine. Groups that enhance pharmacodynamic properties, in the context of this invention, include groups that improve oligonucleotide uptake, enhance oligonucleotide resistance to degradation, and/or strengthen sequence-specific hybridization with RNA. Groups that enhance pharmacokinetic properties, in the context of this invention, include groups that improve oligonucleotide uptake, distribution, metabolism or excretion.

Nucleotides incorporating a connecting atom, a space-spanning group and a functional group on the end of the space-spanning group are generally synthesized as per the methods and syntheses disclosed in United States Patent Application Serial Nos. 463,358, 566,977 and PCT/US91/00243. To introduce amine-functionalized space-spanning groups on a nucleotide within desired oligonucleotide sequences, 5'-dimethoxytrityl-2'-O, 2'-S or 2'-NH-(space-spanning group-N-phthalimido) nucleoside

phosphoramidites are synthesized. The N-phthalimido amine protecting group is removed with concentrated NH_4OH. N-phthalimido compounds thus provide an amine-functionalized space-spanning group attached to the 2'-position of nucleotide components of an oligonucleotide.

The following compounds for forming compounds having amine-functionalized space-spanning groups are commercially available from Aldrich Chemical Co., Inc., Milwaukee, WI: N-(2-bromoethyl)phthalimide, N-(3-bromopropyl)phthalimide and N-(4-bromobutyl)phthalimide. Other phthalimide-protected amine compounds can be conveniently synthesized from appropriate alkyl, aralkyl or aryl halides and phthalimide. Representative compounds include N-(7-bromoheptyl)phthalimide; N-(8-bromooctyl)phthalimide; N-(9-bromononyl)phthalimide; N-(10-bromododecyl)phthalimide; N-(11-bromoundecyl)phthalimide; N-(12-bromodocecyl)phthalimide; N-(13-bromotridecyl) phthalimide; N-(14-bromotetradecyl)phthalimide; N-(15-bromopentadecyl)phthalimide; N-(16-bromohexadecyl)-phthalimide; N-(17-bromoheptadecyl)phthalimide;N-(18-bromooctadecyl)phthalimide; N-(19-bromononadecyl)phthalimide; N-(3-bromo-2-methylpropyl)-phthalimide; N-(4-bromo-2-methyl-3-ethylbutyl)phthalimide; N-(3-bromo-2,2-diethylpropyl)phthalimide; N-(4-bromo-3-propylbutyl)phthalimide; N-(10-bromo-2,8-dibutyldecyl)phthalimide; N-(8-bromo-6,6-dimethyloctyl)phthalimide;N-(8-bromo-6-propyl-6-butyloctyl)phthalimide; N-(4-bromo-2-methylbutyl)phthalimide; N-(5-bromo-2-methylpentyl)phthalimide; N-(5-bromo-3-methylpentyl)phthalimide; N-(6-bromo-2-ethylhexyl)phthalimide; N-(5-bromo-3-penten-2-one)phthalimide; N-(4-bromo-3-methyl-2-butanol)phthalimide; N-(8-bromo-3-amino-4-chloro-2-cyanooctyl)phthalimide; N-(7-bromo-3-methoxy-4-heptanal)phthalimide; N-(4-bromo-2-iodo-3-nitrobutyl)phthalimide; N-(12-bromo-4-isopropoxydodecyl)phthalimide; N-(10-bromo-4-azido-2-nitrodecyl)phthalimide;N-(9-bromo-5-mercaptononyl)phthal- imide;N-(5-bromo-4-aminopentenyl)phthalimide; N-(5-bromo-penten-2-yl)phthalimide; N-(3-bromoallyl)phthalimide; N-(4-bromocrotyl)phthalimide; N-(3-bromopropargyl)phthalimide; N-(1-bromonaphth-4-yl)phthalimide; N-(2-bromoanthrac-7-yl)-

phthalimide; and N-(2-bromophenanthr-6-yl)phthalimide. Such halide compounds are then reacted with an appropriate 2'-oxygen, 2'-sulfur or 2'-amine nucleoside.

In addition to amines ($-NH_2$), further preferred reactive moieties suitable for the functional groups of the invention include hydrazines (-NH-NH-), hydroxylamines (-NH-OH), semicarbazides (-NH-C(O)-NH-NH_2), thiosemicarbazides (-NH-C(S)-NH-NH_2), hydrazones (-N=NH), hydrazides (-C(O)-NH-NH_2), alcohols (-OH), thiols (-SH), and aldehydes (-CH=O).

To introduce a nucleotide having a hydrazine-functionalized space-spanning group at its 2'-position into desired oligonucleotide sequences, 5'-dimethoxytrityl-2'-O-, -2'-S- or -2'-NH-(space-spanning group-N-benzylcarbazide) nucleoside phosphoramidites are synthesized. Once an oligonucleotide incorporating these nucleotides is synthesized, the hydrazine functionality is generated by reduction of the carbazide with palladium on charcoal.

To introduce a nucleotide having a hydroxylamine-functionalized space-spanning group at its 2'-position within desired oligonucleotide sequences, 5'-dimethoxytrityl-2'-O-, -2'-S- or -2'-NH-(space-spanning group-O-phthalimide) nucleoside phosphoramidites are synthesized. Once an oligonucleotide incorporating these nucleotides is synthesized, the hydroxylamine functionality is generated by treatment of the oligonucleotide with methylhydrazine. Semicarbazide and thiosemicarbazide functionalities on space-spanning groups can be introduced directly without protective groups. 5'-Dimethoxytrityl-2'-O-, 2'-S- or 2'-NH-(alkylsemicarbazide)-nucleoside phosphoramidites are synthesized and incorporated into desired oligonucleotides to provide a semicarbazide functionalized space-spanning group attached to the 2'-position of a component nucleotide of an oligonucleotide.

Hydrazide functionalities are formed on space-spanning groups via reaction of hydrazine with an nucleotide bearing a 2'-O-alkylester. Alcohol functionalities can be formed on space-spanning groups via Ω-bromo alcohols. The alcohol hydroxyl group is protected with a tetrahydropyran (THP)

blocking group. The THP- protected compound is then reacted with an activated nucleoside to attach the THP-protected compound at the 2'-position of the nucleoside. After oligonucleotide synthesis, the THP blocking group is removed to expose the alcoholic functionality. Thiol functionalities can be directly formed via ring opening of a 2,2'-anhydro pyrimidine nucleoside using α,Ω-dithiol compounds. Alternatively, the thiol functionality of space-spanning compounds can be protected with S-trityl protecting groups. After oligonucleotide synthesis, the S-trityl protecting groups are removed with silver nitrate.

Aldehydes used as active functional groups in the formation of cross-linkages can be "masked" as protected aldehydes or aldehyde precursors during initial oligonucleotide synthesis. The aldehyde is generated prior to cross-linking from the protected aldehyde or aldehyde precursor and is utilized as one of the reactants for the formation of a covalent cross-linkage. Representative protected aldehydes are C-formyl, *o*-methylaminobenzenethio, aryl substituted imidazolidino moieties such as di-phenylimidazolidino, dithianes, bis(p-nitrophenoxy) and acetals. Particularly preferred acetals are di-aralkyl acetals such as bis(*o*-nitrobenzyl) acetals that are removed via catalytic hydrogenolysis. The diphenylimidazolidino group, which is removed via hydrazine hydrate/sodium cyanoborohydride treatment, is also particularly preferred. Aldehyde precursors include nitriles, alcohols and esters. Nitriles and esters are reduced to their corresponding aldehydes, while alcohols are oxidized to their corresponding aldehydes. Nitriles can be treated with hydrazine, followed by reduction with sodium cyanoborohydride. Esters can be reduced with lithium aluminum hydride. Alcohols can be oxidized under mild oxidation conditions such as Pfitzner-Moffat conditions. *See, J. Am. Chem. Soc.*, **1965**, *87*, 5661. Such conditions have been used by Hansske, *et al.*, *Tetrahedron*, **1984**, *40*, 125, to oxidize 2' or 3' hydroxyl groups of various nucleosides using both a chromic acid oxidation reagent as well as a DMSO/Ac_2O oxidation reagent.

Aldehyde precursors and protected aldehydes can be used both at the nucleoside/nucleotide level and at the oligonucleotide level. Individual nucleotides (or their nucleoside precursors) incorporating an aldehyde precursor or protected aldehyde can be directly incorporated into an oligonucleotide. Alternately, an aldehyde precursor can be converted to a protected aldehyde, as with C-formyl protected aldehydes, and then directly incorporated into an oligonucleotide. Upon synthesis of the oligonucleotide, the aldehyde is regenerated from the protected aldehyde or generated from the aldehyde precursor and the oligonucleotide strand is cross-linked with a further oligonucleotide strand or with a further region of the same oligonucleotide strand. Use of aldehyde precursors is suggested for certain oligonucleotides, such as mimics of pyrimidine tract sequences located upstream from gene splice junctions. One such pyrimidine tract sequence includes the sequence CCC CTC CCC. Since such tracts only contain pyrimidine nucleotides, acidic oxidation conditions used for conversion of alcohols to aldehydes will not result in acidic de-purination or acidic strand cleavage of the oligonucleotide.

As will be recognized, hydrazine, hydroxylamine, semicarbazide, thiosemicarbazide, hydrazone, hydrazide, alcohol or thiol functional groups on one space-spanning group can be reacted with an aldehydic group (a -CH=O group) on a further space-spanning group as, for example, in the above-described embodiment wherein a 2'-ribose to 2'-ribose linkage is formed. These functional groups also can be reacted with an aldehydic group at the C1'-position of an abasic site to form a 2'-ribose to C1'-ribose linkage. Alternatively, these functional groups can be reacted with one of the aldehydic groups of a 2',3'-dialdehyde 3'-terminal nucleoside to form a 2'-ribose to a seco-ribose linkage, or to form imine (-C=N-), hydrazone (-C=N-NH-), oxime (-O-N=CH-), semicarbazone (NH-C(O)-NH-N=C-), thiosemicarbazone (NH-C(S)-NH-N=C-), azine (-HC=N-N=CH-), hydrazide-hydrazone (-C(O)-NH-N=C-), hemiacetal

(-CH(OH)-O-CH_2-) or thiohemiacetal (-CH(OH)-S-CH_2-) cross-linkages.

When an aldehyde on a first strand reacts with an alcohol or a thioacetal on another strand or on a further region of the first strand, the resulting hemiacetal or thiohemiacetal theoretically can react with a further alcohol or thiol group. However, such reaction generally is not kinetically (entropically) favorable, as it usually occurs in a water solution or under other conditions in which the concentration of the further alcohol or thiol group is low.

In those embodiments of the invention wherein a cross-linkage is formed between a 2'-ribose on a first strand (or a first region of a single strand) and a 2'-ribose on another strand (or a further region of the single strand), space-spanning groups and reactive functionalities are attached to each of the strands (or regions of a single strand). The space-spanning groups and reactive functionalities are attached to the 2'-position of each of the strands (or regions of a single strand), as above, utilizing connecting atoms. The strand or strands are then cross-linked by reaction of the reactive functionalities. Accordingly, the reactive functionalities can each be considered "bond precursors" or "bond means" since they together form a covalent bond.

Depending upon the reactive functionality on each of the strands (or regions of a single strand), a covalent bond will be formed as an integral part of the cross-linkage. When a thiol group is the active functionality on each of the strands, oxidization results in a disulfide cross-linkage. When an aldehyde is the functional group on one strand (or a first region on a single strand) and an amine is the functional group on the further strand (or further region on a single strand), an imine (-N=CH-, Schiff's base) cross-linkage results. The imine cross-linkage can be reduced to an amine (-NH-CH_2-) using a reducing agent such as sodium cyanoborohydride.

Reaction of a hydrazine (-NH-NH-) with an aldehyde yields a hydrazone (-HC=N-NH-) cross-linkage, while reaction of

a hydroxylamine (HO-N-) with an aldehyde yields an oxime (-O-N=CH-) cross-linkage. Reaction of an aldehyde with a hydrazide (-C(O)-NH-NH_2) yields a hydrazide-hydrazone (-C(O)-NH-N=CH-) cross-linkage, reaction of an aldehyde with a semicarbazide (-NH-C(O)-NH-NH_2) yields a semicarbazone (-NH-C(O)-NH-N=C-) cross-linkage, and reaction of an aldehyde with a thiosemicarbazide (-NH-C(S)-NH-NH_2) yields a thiosemicarbazone (-NH-C(S)-NH-N=C-) cross-linkage. The hydrazone, oxime, hydrazide-hydrazone, semicarbazone and thiosemicarbazone linkages can be reduced to corresponding hydrazine, hydroxylamine, hydrazide-hydrazine, semicarbazide and thiosemicarbazide, respectively, if desired.

The active functional groups on the space-spanning moieties can be reacted with both heterobifunctional and homobifunctional groups. As will be recognized, formaldehyde is a homobifunctional group when it is reacted with two amines. Reaction of formaldehyde with amine functional groups that are attached to each of two strands (or regions of a single strand) yields an aminoalkylamine (-NH-CH_2-NH-) linkage.

Hydrazine also can be used as a bifunctional group to join strands that each have a space-spanning group terminated with an aldehyde. The hydrazine adds to a first aldehyde group to form an aldehyde-hydrazone that, in turn, reacts with the second aldehyde to yield an azine (-CH=N-N=CH-) cross-linkage.

Heterobifunctional groups allow the use of one active functionality on one strand (or a region of a single strand) and the use of a different active functionality on another strand (or another region of a single strand). Thus, one strand could have an amine functionality on an end of its space-spanning group and another strand could have a thiol functionality on the end of its space-spanning group.

Space-spanning groups on one of the strands (or region of a single strand) may not be necessary where heterobifunctional groups or certain of the above-noted active functionalities are employed. An attaching group, for example a 2'-amino or 2'-thio of a 2'-substituted-2'-deoxy-*erythro*-

pentofuranosyl moiety, could be used directly as an active functionality. The space-spanning group and its active functionality from the other strand could react directly with this 2'-positioned active functionality. For example a 2'-deoxy-2'-thio-*erythro*-pentofuranosyl sugar could be utilized as the sugar moiety of a selected nucleotide on one strand. An omega thiol group on the space-spanning group of a complementary duplex strand can then be coupled to the first 2'-thio group via a disulfide linkage. Other 2' groups suitable for this type of cross-linkage include 2'-amino groups. A 2'-keto group also could be reacted with a suitable hydrazine or other reactive functionality.

A variety of heterobifunctional coupling reagents are available from commercial sources including Pierce (Rockford, Il). These reagents include 1-ethyl-3-(3-dimethylaminopropyl)carbodimide (a water soluble DCC type coupling reagent), bifunctional imidoesters such as dimethyl adipimidate dihydrochloride, dimethyl pimelimidate dihydrochloride, and dimethyl 3,3'-dithiobispropionimidate dihydrochloride (for coupling amine groups), and 2-imino-thiolane (for coupling amines thiols). Coupling reagents also include compounds having N-hydroxysuccinimide moieties that react with amines, compounds having internal disulfide bonds that can be released to expose an active thiol moiety that can be further coupled via a disulfide bond with a further thiol functionality, and compounds having sulfosuccinimidyl groups (*e.g.*, N-hydroxysulfosuccinimide) that have different reactivity toward amines than N-hydroxysuccinimides. Coupling reagents further include pyridyl-disulfides that exchange with free thiol moieties, active halogen species such as ß-carbonyl alkylhalides, e.g. iodo acetamide, maleimide moieties and photoreactive aryl azides such as phenyl azide. Also, thioisocyanates can be used to couple with active functionalities on space-spanning groups.

The various functional groups of the bifunctional cross-linkers are joined together with suitable linking atoms. Representative, commercially available bifunctional compounds

include sulfosuccinimidyl 4-(N-maleimidoethyl)cyclohexane-1-carboxylate, N-succinimidyl 3-(2-pyridyldithio)propionate, N-sulfosuccinimidyl(4-iodoacetyl)aminobenzoate, sulfo-succinimidyl 4-(p-maleimidophenyl)butyrate, m-maleimido-benzoyl-N-hydroxysulfosuccinimide ester, succinimidyl 4-(N-maleimidomethyl)cyclohexane-1-carboxylate,bismaleimidohexane, disulfosuccinimidyl tartrate, ethylene glycol*bis*-(succinimidylsuccinate), dithio*bis*(succinimidylpropionate), disuccinimidyl suberate, N-hydroxysuccinimidyl-4-azidosalicylic acid, N-5-azido-2-nitrobenzoyloxysuccinimide and sulfonsuccinimidyl 2-(*m*-azido-*o*-nitrobenzamido)-ethyl-1,3'-dithiopropionate. Each of these bifunctional groups are reacted with an appropriate active functionality on the space-spanning group. Thus, oligonucleotide strands (or regions of a single strand) bearing the space-spanning groups are cross-linked through the bifunctional group.

The 6-bromo-5,5-dimethoxyhexanohydrazide and N-hydroxysuccinimidyl iodoacetate reagents disclosed by Summerton, *et al.* and Meyer, *et al.*, respectively, in the references cited above can also be used as heterobifunctional cross-linking reagents. The hydrazine moiety of the Summerton reagent and the succinimidyl moiety of the Meyer reagent can be linked to one strand of duplex strands and the other strand can be linked to the bromoketone or the iodoacetamido portions of these reagents. For instance, the hydrazine functionality of the Summerton reagent would first be reacted with an aldehydic functionality on one strand, the α-haloketal functionality of the reagent would be activated to an α-haloketone via pH adjustment, and the α-haloketone would then be further reacted with an amine or thiol functionality on the other strand. The active succinimide functionality of the Meyer reagent would be linked to a first amine functionality on a first strand and the resulting iodoacetamidoalkyl functionality would be linked to a further amine or thiol functionality on the other strand.

Since both the Summerton and Meyer reagents can be used to join two amine functionalities (two hydrazine

functionalities in the case of the Summerton reagent), they can also be considered homobifunctional linking reagents. Neither the Summerton reagent nor the Meyer reagent are used in accordance with the invention to couple duplex strands via the 7-position of a purine ring. Thus, no charged (*i.e.*, chemically unstable) species are formed but, rather uncharged, chemically stable cross-linkages of the structures $-CH_2-S-CH_2-C(O)-CH_2-$ and $-CH_2-NH-CH_2-C(O)-CH_2-$.

By terminating a first oligonucleotide strand at its 3′ end with a terminal ribofuranosyl nucleoside, cross-linking to a further oligonucleotide strands (or to a further region of a single strand) can be effected via one of two mechanisms. The first mechanism involves generation of a seco ribose by, for example, periodate oxidation of the 2′,3′ hydroxyl groups of the 3′-terminal ribose. [*See, e.g.*, Lemaitre, *Proc. Natl. Acad. Sci. USA*, **1986**, *84*, 648; Chang, *Biochem. J.*, **1981**, *199*, 281; and Easterbrook-Smith, *Eur. J. Biochem*, **1976**, *62*, 125.] Such oxidation yields a 2′,3′ di-aldehyde (a seco ribose), which is reacted with a second strand (or a further region of a single strand) that includes a space-spanning group bearing an active amine functionality. Such reaction yields a Schiff's base, which is reduced with, for example, sodium cyanoborohydride, to form an amine cross-linker having an amine group directly coupled to the seco ribose of the 3′-terminal nucleoside of the first oligonucleotide strand.

The second mechanism involves formation of an acetal cross-linkage from the 2′,3′-hydroxyl groups of the terminal ribose. To form the acetal linkage, the hydroxyl groups of the terminal ribose are reacted with an aldehyde group on a space-spanning group on the second strand (or further region of a single strand). Acetals of this type have been utilized to attach a nitrogen mustard to a terminal ribose, as described by Knorre, *et al.*, Nucleotide and Oligonucleotide Derivatives as Enzyme and Nucleic Acid Targeted Irreversible Inhibitors, Chemical Aspects, G. Weber (Ed.), Advances Press, Oxford, 1986, pp. 277-299. The nitrogen mustard was then further reacted with a complementary sequence of a short oligonucleotide.

Significantly, the cross-linkage was formed via the nitrogen mustard, not via the acetal. As noted above, cross-linkages formed via nitrogen mustards generate charged species that are chemically unstable. The present invention employs an acetal, which does not generate a charged species, to effect cross-linking.

Taken together, the space-spanning groups, the functional groups, and the connecting atoms comprise a moiety selected from the group consisting of: C_1-C_{20} straight chain alkylamine, C_1-C_{20} straight chain O-alkylamine, C_1-C_{20} straight chain S-alkylamine, C_1-C_{20} straight chain NH-alkylamine, C_1-C_{20} straight chain substituted alkylamine, C_1-C_{20} straight chain substituted O-alkylamine, C_1-C_{20} straight chain substituted S-alkylamine, C_1-C_{20} straight chain substituted NH-alkylamine, C_2-C_{50} branched chain alkylamine, C_2-C_{50} branched chain O-alkylamine, C_2-C_{50} branched chain S-alkylamine, C_2-C_{50} branched chain NH-alkylamine, C_2-C_{50} branched chain substituted alkylamine, C_2-C_{50} branched chain substituted O-alkylamine, C_2-C_{50} branched chain substituted S-alkylamine, C_2-C_{50} branched chain substituted NH-alkylamine, C_2-C_{20} straight chain alkenylamine, C_2-C_{20} straight chain O-alkenylamine, C_2-C_{20} straight chain S-alkenylamine, C_2-C_{20} straight chain NH-alkenylamine, C_2-C_{20} straight chain substituted alkenylamine, C_2-C_{20} straight chain substituted O-alkenylamine, C_2-C_{20} straight chain substituted S-alkenylamine, C_2-C_{20} straight chain substituted NH-alkenylamine, C_3-C_{50} branched chain alkenylamine, C_3-C_{50} branched chain O-alkenylamine, C_3-C_{50} branched chain S-alkenylamine, C_3-C_{50} branched chain NH-alkenylamine, C_3-C_{50} branched chain substituted alkenylamine, C_3-C_{50} branched chain substituted O-alkenylamine, C_3-C_{50} branched chain substituted S-alkenylamine, C_3-C_{50} branched chain substituted NH-alkenylamine, C_2-C_{20} straight chain alkynylamine, C_2-C_{20} straight chain O-alkynylamine, C_2-C_{20} straight chain S-alkynylamine, C_2-C_{20} straight chain NH-alkynylamine, C_2-C_{20} straight chain substituted alkynylamine, C_2-C_{20} straight chain substituted O-alkynylamine, C_2-C_{20} straight chain substituted S-alkynylamine, C_2-C_{20} straight chain substituted NH-alkynylamine,

C_3-C_{50} branched chain alkynylamine, C_2-C_{50} branched chain O-alkynylamine, C_3-C_{50} branched chain S-alkynylamine, C_3-C_{50} branched chain NH-alkynylamine, C_3-C_{50} branched chain substituted alkynylamine, C_3-C_{50} branched chain substituted O-alkynylamine, C_3-C_{50} branched chain substituted S-alkynylamine, C_3-C_{50} branched chain substituted NH-alkynylamine, C_1-C_{20} straight chain alkylhydrazine, C_1-C_{20} straight chain O-alkylhydrazine, C_1-C_{20} straight chain S-alkylhydrazine, C_1-C_{20} straight chain NH-alkylhydrazine, C_1-C_{20} straight chain substituted alkylhydrazine, C_1-C_{20} straight chain substituted O-alkylhydrazine, C_1-C_{20} straight chain substituted S-alkylhydrazine, C_1-C_{20} straight chain substituted NH-alkylhydrazine, C_2-C_{50} branched chain alkylhydrazine, C_2-C_{50} branched chain O-alkylhydrazine, C_2-C_{50} branched chain S-alkylhydrazine, C_2-C_{50} branched chain NH-alkylhydrazine, C_2-C_{50}, branched chain substituted alkylhydrazine, C_2-C_{50} branched chain substituted O-alkylhydrazine, C_2-C_{50} branched chain substituted S-alkylhydrazine, C_2-C_{50} branched chain substituted NH-alkylhydrazine, C_2-C_{20} straight chain alkenylhydrazine, C_2-C_{20} straight chain O-alkenylhydrazine, C_2-C_{20} straight chain S-alkenylhydrazine, C_2-C_{20} straight chain NH-alkenylhydrazine, C_2-C_{20} straight chain substituted alkenylhydrazine, C_2-C_{20} straight chain substituted O-alkenylhydrazine, C_2-C_{20} straight chain substituted S-alkenylhydrazine, C_2-C_{20} straight chain substituted NH-alkenylhydrazine, C_3-C_{50} branched chain alkenylhydrazine, C_3-C_{50} branched chain O-alkenylhydrazine, C_3-C_{50} branched chain S-alkenylhydrazine, C_3-C_{50} branched chain NH-alkenylhydrazine, C_3-C_{50} branched chain substituted alkenylhydrazine, C_3-C_{50} branched chain substituted O-alkenylhydrazine, C_3-C_{50} branched chain substituted S-alkenylhydrazine, C_3-C_{50} branched chain substituted NH-alkenylhydrazine, C_2-C_{20} straight chain alkynylhydrazine, C_2-C_{20} straight chain O-alkynylhydrazine, C_2-C_{20} straight chain S-alkynylhydrazine, C_2-C_{20} straight chain NH-alkynylhydrazine, C_2-C_{20} straight chain substituted alkynylhydrazine, C_2-C_{20} straight chain substituted O-alkynylhydrazine, C_2-C_{20} straight chain substituted S-alkynylhydrazine, C_2-C_{20} straight chain

substituted NH-alkynylhydrazine, C_3-C_{50} branched chain alkynylhydrazine, C_2-C_{50} branched chain O-alkynylhydrazine, C_3-C_{50} branched chain S-alkynylhydrazine, C_3-C_{50} branched chain NH-alkynylhydrazine, C_3-C_{50} branched chain substituted alkynylhydrazine, C_3-C_{50} branched chain substituted O-alkynylhydrazine, C_3-C_{50} branched chain substituted S-alkynylhydrazine, C_3-C_{50} branched chain substituted NH-alkynylhydrazine, C_1-C_{20} straight chain alkylhydroxylamine, C_1-C_{20} straight chain O-alkylhydroxylamine, C_1-C_{20} straight chain S-alkylhydroxylamine, C_1-C_{20} straight chain NH-alkylhydroxylamine, C_1-C_{20} straight chain substituted alkylhydroxylamine, C_1-C_{20} straight chain substituted O-alkylhydroxylamine, C_1-C_{20} straight chain substituted S-alkylhydroxylamine, C_1-C_{20} straight chain substituted NH-alkylhydroxylamine, C_2-C_{50} branched chain alkylhydroxylamine, C_2-C_{50} branched chain O-alkylhydroxylamine, C_2-C_{50} branched chain S-alkylhydroxylamine, C_2-C_{50} branched chain NH-alkylhydroxylamine, C_2-C_{50} branched chain substituted alkylhydroxylamine, C_2-C_{50} branched chain substituted O-alkylhydroxylamine, C_2-C_{50} branched chain substituted S-alkylhydroxylamine, C_2-C_{50} branched chain substituted NH-alkylhydroxylamine, C_2-C_{20} straight chain alkenylhydroxylamine, C_2-C_{20} straight chain O-alkenylhydroxylamine, C_2-C_{20} straight chain S-alkenylhydroxylamine, C_2-C_{20} straight chain NH-alkenylhydroxylamine, C_2-C_{20} straight chain substituted alkenyl-hydroxylamine, C_2-C_{20} straight chain substituted O-alkenylhydroxylamine, C_2-C_{20} straight chain substituted S-alkenylhydroxylamine, C_2-C_{20} straight chain substituted NH-alkenylhydroxylamine, C_3-C_{50} branched chain alkenyl-hydroxylamine, C_3-C_{50} branched chain O-alkenylhydroxylamine, C_3-C_{50} branched chain S-alkenylhydroxylamine, C_3-C_{50} branched chain NH-alkenylhydroxylamine, C_3-C_{50} branched chain substituted alkenylhydroxylamine, C_3-C_{50} branched chain substituted O-alkenylhydroxylamine, C_3-C_{50} branched chain substituted S-alkenylhydroxylamine, C_3-C_{50} branched chain substituted NH-alkenylhydroxylamine, C_2-C_{20} straight chain alkynylhydroxylamine, C_2-C_{20} straight chain O-alkynylhydroxylamine, C_2-C_{20} straight chain S-alkynyl-

hydroxylamine, C_2-C_{20} straight chain NH-alkynylhydroxylamine, C_2-C_{20} straight chain substituted alkynylhydroxylamine, C_2-C_{20} straight chain substituted O-alkynylhydroxylamine, C_2-C_{20} straight chain substituted S-alkynylhydroxylamine, C_2-C_{20} straight chain substituted NH-alkynylhydroxylamine, C_3-C_{50} branched chain alkynylhydroxylamine, C_2-C_{50} branched chain O-alkynylhydroxylamine, C_3-C_{50} branched chain S-alkynylhydroxylamine, C_3-C_{50} branched chain NH-alkynylhydroxylamine, C_3-C_{50} branched chain substituted alkynylhydroxylamine, C_3-C_{50} branched chain substituted O-alkynylhydroxylamine, C_3-C_{50} branched chain substituted S-alkynylhydroxylamine, C_3-C_{50} branched chain substituted NH-alkynylhydroxylamine, C_1-C_{20} straight chain alkylsemicarbazide, C_1-C_{20} straight chain O-alkylsemicarbazide, C_1-C_{20} straight chain S-alkylsemicarbazide, C_1-C_{20} straight chain NH-alkylsemicarbazide, C_1-C_{20} straight chain substituted alkylsemicarbazide, C_1-C_{20} straight chain substituted O-alkylsemicarbazide, C_1-C_{20} straight chain substituted S-alkylsemicarbazide, C_1-C_{20} straight chain substituted NH-alkylsemicarbazide, C_2-C_{50} branched chain alkylsemicarbazide, C_2-C_{50} branched chain O-alkylsemicarbazide, C_2-C_{50} branched chain S-alkylsemicarbazide, C_2-C_{50} branched chain NH-alkylsemicarbazide, C_2-C_{50} branched chain substituted alkylsemicarbazide, C_2-C_{50} branched chain substituted O-alkylsemicarbazide, C_2-C_{50} branched chain substituted S-alkylsemicarbazide, C_2-C_{50} branched chain substituted NH-alkylsemicarbazide, C_2-C_{20} straight chain alkenylsemicarbazide, C_2-C_{20} straight chain O-alkenylsemicarbazide, C_2-C_{20} straight chain S-alkenylsemicarbazide, C_2-C_{20} straight chain NH-alkenylsemicarbazide, C_2-C_{20} straight chain substituted alkenylsemicarbazide, C_2-C_{20} straight chain substituted O-alkenylsemicarbazide, C_2-C_{20} straight chain substituted S-alkenylsemicarbazide, C_2-C_{20} straight chain substituted NH-alkenylsemicarbazide, C_3-C_{50} branched chain alkenylsemicarbazide, C_3-C_{50} branched chain O-alkenylsemicarbazide, C_3-C_{50} branched chain S-alkenylsemicarbazide, C_3-C_{50} branched chain NH-alkenylsemicarbazide,

C_3-C_{50} branched chain substituted alkenylsemicarbazide, C_3-C_{50} branched chain substituted O-alkenylsemicarbazide, C_3-C_{50} branched chain substituted S-alkenylsemicarbazide, C_3-C_{50} branched chain substituted NH-alkenylsemicarbazide, C_2-C_{20} straight chain alkynylsemicarbazide, C_2-C_{20} straight chain O-alkynylsemicarbazide, C_2-C_{20} straight chain S-alkynylsemicarbazide, C_2-C_{20} straight chain NH-alkynylsemicarbazide, C_2-C_{20} straight chain substituted alkynylsemicarbazide, C_2-C_{20} straight chain substituted O-alkynylsemicarbazide, C_2-C_{20} straight chain substituted S-alkynylsemicarbazide, C_2-C_{20} straight chain substituted NH-alkynylsemicarbazide, C_3-C_{50} branched chain alkynylsemicarbazide, C_2-C_{50} branched chain O-alkynylsemicarbazide, C_3-C_{50} branched chain S-alkynylsemicarbazide, C_3-C_{50} branched chain NH-alkynylsemicarbazide, C_3-C_{50} branched chain substituted alkynylsemicarbazide, C_3-C_{50} branched chain substituted O-alkynylsemicarbazide, C_3-C_{50} branched chain substituted S-alkynylsemicarbazide, C_3-C_{50} branched chain substituted NH-alkynylsemicarbazide, C_1-C_{20} straight chain alkylthiosemicarbazide, C_1-C_{20} straight chain O-alkylthiosemicarbazide, C_1-C_{20} straight chain S-alkylthiosemicarbazide, C_1-C_{20} straight chain NH-alkylthiosemicarbazide, C_1-C_{20} straight chain substituted alkylthiosemicarbazide, C_1-C_{20} straight chain substituted O-alkylthiosemicarbazide, C_1-C_{20} straight chain substituted S-alkylthiosemicarbazide, C_1-C_{20} straight chain substituted NH-alkylthiosemicarbazide, C_2-C_{50} branched chain alkylthiosemicarbazide, C_2-C_{50} branched chain O-alkylthiosemicarbazide, C_2-C_{50} branched chain S-alkylthiosemicarbazide, C_2-C_{50} branched chain NH-alkylthiosemicarbazide, C_2-C_{50} branched chain substituted alkylthiosemicarbazide, C_2-C_{50} branched chain substituted O-alkylthiosemicarbazide, C_2-C_{50} branched chain substituted S-alkylthiosemicarbazide, C_2-C_{50} branched chain substituted NH-alkylthiosemicarbazide, C_2-C_{20} straight chain alkenylthiosemicarbazide, C_2-C_{20} straight chain O-alkenylthiosemicarbazide, C_2-C_{20} straight chain S-alkenylthiosemicarbazide, C_2-C_{20} straight chain NH-alkenylthiosemi-

carbazide, C_2-C_{20} straight chain substituted alkenylthiosemicarbazide, C_2-C_{20} straight chain substituted O-alkenylthiosemicarbazide, C_2-C_{20} straight chain substituted S-alkenylthiosemicarbazide, C_2-C_{20} straight chain substituted NH-alkenylthiosemicarbazide, C_3-C_{50} branched chain alkenylthiosemicarbazide, C_3-C_{50} branched chain O-alkenylthiosemicarbazide, C_3-C_{50} branched chain S-alkenylthiosemicarbazide, C_3-C_{50} branched chain NH-alkenylthiosemicarbazide, C_3-C_{50} branched chain substituted alkenylthiosemicarbazide, C_3-C_{50} branched chain substituted O-alkenylthiosemicarbazide, C_3-C_{50} branched chain substituted S-alkenylthiosemicarbazide, C_3-C_{50} branched chain substituted NH-alkenylthiosemicarbazide, C_2-C_{20} straight chain alkynylthiosemicarbazide, C_2-C_{20} straight chain O-alkynylthiosemicarbazide, C_2-C_{20} straight chain S-alkynylthiosemicarbazide, C_2-C_{20} straight chain NH-alkynylthiosemi- carbazide, C_2-C_{20} straight chain substituted alkynylthiosemicarbazide, C_2-C_{20} straight chain substituted O-alkynylthiosemicarbazide, C_2-C_{20} straight chain substituted S-alkynylthiosemicarbazide, C_2-C_{20} straight chain substituted NH-alkynylthiosemicarbazide, C_3-C_{50} branched chain alkynylthiosemicarbazide, C_2-C_{50} branched chain O-alkynylthiosemicarbazide, C_3-C_{50} branched chain S-alkynylthiosemicarbazide, C_3-C_{50} branched chain NH-alkynylthiosemicarbazide, C_3-C_{50} branched chain substituted alkynylthiosemicarbazide, C_3-C_{50} branched chain substituted O-alkynylthiosemicarbazide, C_3-C_{50} branched chain substituted S-alkynylthiosemicarbazide, C_3-C_{50} branched chain substituted NH-alkynylthiosemicarbazide, C_1-C_{20} straight chain alkylhydrazone, C_1-C_{20} straight chain O-alkylhydrazone, C_1-C_{20} straight chain S-alkylhydrazone, C_1-C_{20} straight chain NH-alkylhydrazone, C_1- C_{20} straight chain substituted alkylhydrazone, C_1-C_{20} straight chain substituted O-alkylhydrazone, C_1-C_{20} straight chain substituted S-alkylhydrazone, C_1-C_{20} straight chain substituted NH-alkylhydrazone, C_2-C_{50} branched chain alkylhydrazone, C_2-C_{50} branched chain O-alkylhydrazone, C_2-C_{50} branched chain S-alkylhydrazone, C_2-C_{50} branched chain NH-alkylhydrazone, C_2-C_{50} branched chain substituted alkylhydrazone, C_2-C_{50} branched chain substituted O-alkylhydrazone, C_2-C_{50} branched chain substituted

S-alkylhydrazone, C_2-C_{50} branched chain substituted NH-alkylhydrazone, C_2-C_{20} straight chain alkenylhydrazone, C_2-C_{20} straight chain O-alkenylhydrazone, C_2-C_{20} straight chain S-alkenylhydrazone, C_2-C_{20} straight chain NH-alkenylhydrazone, C_2-C_{20} straight chain substituted alkenylhydrazone, C_2-C_{20} straight chain substituted O-alkenylhydrazone, C_2-C_{20} straight chain substituted S-alkenylhydrazone, C_2-C_{20} straight chain substituted NH-alkenylhydrazone, C_3-C_{50} branched chain alkenylhydrazone, C_3-C_{50} branched chain O-alkenylhydrazone, C_3-C_{50} branched chain S-alkenylhydrazone, C_3-C_{50} branched chain NH-alkenylhydrazone, C_3-C_{50} branched chain substituted alkenylhydrazone, C_3-C_{50} branched chain substituted O-alkenylhydrazone, C_3-C_{50} branched chain substituted S-alkenylhydrazone, C_3-C_{50} branched chain substituted NH-alkenylhydrazone, C_2-C_{20} straight chain alkynylhydrazone, C_2-C_{20} straight chain O-alkynylhydrazone, C_2-C_{20} straight chain S-alkynylhydrazone, C_2-C_{20} straight chain NH-alkynylhydrazone, C_2-C_{20} straight chain substituted alkynylhydrazone, C_2-C_{20} straight chain substituted O-alkynylhydrazone, C_2-C_{20} straight chain substituted S-alkynylhydrazone, C_2-C_{20} straight chain substituted NH-alkynylhydrazone, C_3-C_{50} branched chain alkynylhydrazone, C_2-C_{50} branched chain O-alkynylhydrazone, C_3-C_{50} branched chain S-alkynylhydrazone, C_3-C_{50} branched chain NH-alkynylhydrazone, C_3-C_{50} branched chain substituted alkynylhydrazone, C_3-C_{50} branched chain substituted O-alkynylhydrazone, C_3-C_{50} branched chain substituted S-alkynylhydrazone, C_3-C_{50} branched chain substituted NH-alkynylhydrazone, C_1-C_{20} straight chain alkylhydrazide, C_1-C_{20} straight chain O-alkylhydrazide, C_1-C_{20} straight chain S-alkylhydrazide, C_1-C_{20} straight chain NH-alkylhydrazide, C_1-C_{20} straight chain substituted alkylhydrazide, C_1-C_{20} straight chain substituted O-alkylhydrazide, C_1-C_{20} straight chain substituted S-alkylhydrazide, C_1-C_{20} straight chain substituted NH-alkylhydrazide, C_2-C_{50} branched chain alkylhydrazide, C_2-C_{50} branched chain O-alkylhydrazide, C_2-C_{50} branched chain S-alkylhydrazide, C_2-C_{50} branched chain NH-alkylhydrazide, C_2-C_{50} branched chain substituted alkylhydrazide, C_2-C_{50} branched chain

substituted O-alkylhydrazide, C_2-C_{50} branched chain substituted S-alkylhydrazide, C_2-C_{50} branched chain substituted NH-alkylhydrazide, C_2-C_{20} straight chain alkenylhydrazide, C_2-C_{20} straight chain O-alkenylhydrazide, C_2-C_{20} straight chain S-alkenylhydrazide, C_2-C_{20} straight chain NH-alkenylhydrazide, C_2-C_{20} straight chain substituted alkenylhydrazide, C_2-C_{20} straight chain substituted O-alkenylhydrazide, C_2-C_{20} straight chain substituted S-alkenylhydrazide, C_2-C_{20} straight chain substituted NH-alkenylhydrazide, C_3-C_{50} branched chain alkenylhydrazide, C_3-C_{50} branched chain O-alkenylhydrazide, C_3-C_{50} branched chain S-alkenylhydrazide, C_3-C_{50} branched chain NH-alkenylhydrazide, C_3-C_{50} branched chain substituted alkenylhydrazide, C_3-C_{50} branched chain substituted O-alkenylhydrazide, C_3-C_{50} branched chain substituted S-alkenylhydrazide, C_3-C_{50} branched chain substituted NH-alkenylhydrazide, C_2-C_{20} straight chain alkynylhydrazide, C_2-C_{20} straight chain O-alkynylhydrazide, C_2-C_{20} straight chain S-alkynylhydrazide, C_2-C_{20} straight chain NH-alkynylhydrazide, C_2-C_{20} straight chain substituted alkynylhydrazide, C_2-C_{20} straight chain substituted O-alkynylhydrazide, C_2-C_{20} straight chain substituted S-alkynylhydrazide, C_2-C_{20} straight chain substituted NH-alkynylhydrazide, C_3-C_{50} branched chain alkynylhydrazide, C_2-C_{50} branched chain O-alkynylhydrazide, C_3-C_{50} branched chain S-alkynylhydrazide, C_3-C_{50} branched chain NH-alkynylhydrazide, C_3-C_{50} branched chain substituted alkynylhydrazide, C_3-C_{50} branched chain substituted O-alkynylhydrazide, C_3-C_{50} branched chain substituted S-alkynylhydrazide, C_3-C_{50} branched chain substituted NH-alkynylhydrazide, C_1-C_{20} straight chain alkanol, C_1-C_{20} straight chain O-alkanol, C_1-C_{20} straight chain S-alkanol, C_1-C_{20} straight chain NH-alkanol, C_1-C_{20} straight chain substituted alkanol, C_1-C_{20} straight chain substituted O-alkanol, C_1-C_{20} straight chain substituted S-alkanol, C_1-C_{20} straight chain substituted NH-alkanol, C_2-C_{50} branched chain alkanol, C_2-C_{50} branched chain O-alkanol, C_2-C_{50} branched chain S-alkanol, C_2-C_{50} branched chain NH-alkanol, C_2-C_{50} branched chain substituted alkanol, C_2-C_{50} branched chain substituted O-alkanol, C_2-C_{50} branched chain substituted S-

alkanol, C_2-C_{50} branched chain substituted NH-alkanol, C_2-C_{20} straight chain alkenol, C_2-C_{20} straight chain O-alkenol, C_2-C_{20} straight chain S-alkenol, C_2-C_{20} straight chain NH-alkenol, C_2-C_{20} straight chain substituted alkenol, C_2-C_{20} straight chain substituted O-alkenol, C_2-C_{20} straight chain substituted S-alkenol, C_2-C_{20} straight chain substituted NH-alkenol, C_3-C_{50} branched chain alkenol, C_3-C_{50} branched chain O-alkenol, C_3-C_{50} branched chain S-alkenol, C_3-C_{50} branched chain NH-alkenol, C_3-C_{50} branched chain substituted alkenol, C_3-C_{50} branched chain substituted O-alkenol, C_3-C_{50} branched chain substituted S-alkenol, C_3-C_{50} branched chain substituted NH-alkenol, C_2-C_{20} straight chain alkynol, C_2-C_{20} straight chain O-alkynol, C_2-C_{20} straight chain S-alkynol, C_2-C_{20} straight chain NH-alkynol, C_2-C_{20} straight chain substituted alkynol, C_2-C_{20} straight chain substituted O-alkynol, C_2-C_{20} straight chain substituted S-alkynol, C_2-C_{20} straight chain substituted NH-alkynol, C_3-C_{50} branched chain alkynol, C_2-C_{50} branched chain O-alkynol, C_3-C_{50} branched chain S-alkynol, C_3-C_{50} branched chain NH-alkynol, C_3-C_{50} branched chain substituted alkynol, C_3-C_{50} branched chain substituted O-alkynol, C_3-C_{50} branched chain substituted S-alkynol, C_3-C_{50} branched chain substituted NH-alkynol, C_1-C_{20} straight chain alkanthiol, C_1-C_{20} straight chain O-alkanthiol, C_1-C_{20} straight chain S-alkanthiol, C_1-C_{20} straight chain NH-alkanthiol, C_1-C_{20} straight chain substituted alkanthiol, C_1-C_{20} straight chain substituted O-alkanthiol, C_1-C_{20} straight chain substituted S-alkanthiol, C_1-C_{20} straight chain substituted NH-alkanthiol, C_2-C_{50} branched chain alkanthiol, C_2-C_{50} branched chain O-alkanthiol, C_2-C_{50} branched chain S-alkanthiol, C_2-C_{50} branched chain NH-alkanthiol, C_2-C_{50} branched chain substituted alkanthiol, C_2-C_{50} branched chain substituted O-alkanthiol, C_2-C_{50} branched chain substituted S-alkanthiol, C_2-C_{50} branched chain substituted NH-alkanthiol, C_2-C_{20} straight chain alkenthiol, C_2-C_{20} straight chain O-alkenthiol, C_2-C_{20} straight chain S-alkenthiol, C_2-C_{20} straight chain NH-alkenthiol, C_2-C_{20} straight chain substituted alkenthiol, C_2-C_{20} straight chain substituted O-alkenthiol, C_2-C_{20} straight chain substituted S-alkenthiol, C_2-C_{20} straight chain substituted NH-alkenthiol, C_3-

C_{50} branched chain alkenthiol, C_3-C_{50} branched chain O-alkenthiol, C_3-C_{50} branched chain S-alkenthiol, C_3-C_{50} branched chain NH-alkenthiol, C_3-C_{50} branched chain substituted alkenthiol, C_3-C_{50} branched chain substituted O-alkenthiol, C_3-C_{50} branched chain substituted S-alkenthiol, C_3-C_{50} branched chain substituted NH-alkenthiol, C_2-C_{20} straight chain alkynthiol, C_2-C_{20} straight chain O-alkynthiol, C_2-C_{20} straight chain S-alkynthiol, C_2-C_{20} straight chain NH-alkynthiol, C_2-C_{20} straight chain substituted alkynthiol, C_2-C_{20} straight chain substituted O-alkynthiol, C_2-C_{20} straight chain substituted S-alkynthiol, C_2-C_{20} straight chain substituted NH-alkynthiol, C_3-C_{50} branched chain alkynthiol, C_2-C_{50} branched chain O-alkynthiol, C_3-C_{50} branched chain S-alkynthiol, C_3-C_{50} branched chain NH-alkynthiol, C_3-C_{50} branched chain substituted alkynthiol, C_3-C_{50} branched chain substituted O-alkynthiol, C_3-C_{50} branched chain substituted S-alkynthiol and C_3-C_{50} branched chain substituted NH-alkynthiol.

The present invention also provides oligonucleotide cross-linkers for connecting two oligonucleotide strands or for connecting two separate regions of a single strand. In certain embodiments, the cross-linkers are of a structure:

$$X_1-L_1-Y-L_2-X_2$$

wherein X_1 is covalently connected to the first strand or to a first region of a single strand and X_2 is covalently connected to the second strand or to a further region of a single strand and where:

(i) X_1 and X_2, independently, are O, S, NH, CH_2 or CO;

L_1 and L_2, independently, are C_1-C_{20} straight chain alkyl, C_1-C_{20} straight chain substituted alkyl, C_2-C_{50} branched chain alkyl, C_2-C_{50} branched chain substituted alkyl, C_2-C_{20} straight chain alkenyl, C_2-C_{20} straight chain substituted alkenyl, C_3-C_{50} branched chain alkenyl, C_3-C_{50} branched chain substituted alkenyl, C_2-C_{20} straight chain alkyne, C_2-C_{20} straight

chain substituted alkyne, C_3-C_{50} branched chain alkyne, C_3-C_{50} branched chain substituted alkyne, polyamine, polyamide, polyester, polyethylene glycol, polyether, aryl, aralkyl or heterocyclic; and

Y is an imine, amine, oxime, hydroxylamine, hydrazine, hydrazone, azine, hydrazide-hydrazone, amide, hydrazide, semicarbazide, semicarbazone, thiosemicarbazide, thiocarbazone, disulfide, hemiacetal, thiohemiacetal, α-keto-alkylthioalkyl or α-keto-alkylaminoalkyl; or

(ii) X_1 and X_2, independently, are O or NH; and

L_1-Y-L_2 together are NH-L_3-NH or NH-NH-L_3-NH-NH; and

L_3 is C_1-C_{20} straight chain alkyl, C_1-C_{20} straight chain substituted alkyl, C_2-C_{50} branched chain alkyl, C_2-C_{50} branched chain substituted alkyl, C_2-C_{20} straight chain alkenyl, C_2-C_{20} straight chain substituted alkenyl, C_3-C_{50} branched chain alkenyl, C_3-C_{50} branched chain substituted alkenyl, C_2-C_{20} straight chain alkyne, C_2-C_{20} straight chain substituted alkyne, C_3-C_{50} branched chain alkyne, C_3-C_{50} branched chain substituted alkyne, polyamine, polyamide, polyester, polyethylene glycol, polyether, aryl, aralkyl or heterocyclic; or

(iii) X_1 and X_2, independently, are O, S, NH, CH_2 or CO;

L_1 and L_2, independently, are C_1-C_{20} straight chain alkylamino, C_1-C_{20} straight chain substituted alkylamino, C_2-C_{50} branched chain alkylamino, C_2-C_{50} branched chain substituted alkylamino, C_2-C_{20} straight chain alkenylamino, C_2-C_{20} straight chain substituted alkenylamino, C_3-C_{50} branched chain alkenylamino, C_3-C_{50} branched chain substituted alkenylamino, C_2-C_{20} straight chain alkynylamino, C_2-C_{20} straight chain substituted alkynylamino, C_3-C_{50} branched chain alkynylamino, C_3-C_{50} branched chain substituted alkynylamino, C_1-C_{20} straight chain alkylthio, C_1-C_{20}

straight chain substituted alkylthio, C_2-C_{50} branched chain alkylthio, C_2-C_{50}, branched chain substituted alkylthio, C_2-C_{20} straight chain alkenylthio, C_2-C_{20} straight chain substituted alkenylthio, C_3-C_{50} branched chain alkenylthio, C_3-C_{50} branched chain substituted alkenylthio, C_2-C_{20} straight chain alkynylthio, C_2-C_{20} straight chain substituted alkynylthio, C_3-C_{50} branched chain alkynylthio or C_3-C_{50} branched chain substituted alkynylthio; and

Y is a heterobifunctional or homobifunctional cross-linking reagent;

where said substituent of said substituted alkyl, substituted alkenyl and substituted alkyne moieties is an intercalator, a conjugate, polyamine, polyamide, polyethylene glycol, a group that enhances the pharmacodynamic properties of oligonucleotides or a group that enhances the pharmacokinetic properties of oligonucleotides.

As illustrated in the examples below, the invention further includes methods for fixing a sequence-specific oligonucleotide in a set configuration. These methods include selecting a first site in the oligonucleotide. The first site is selected to have a first specific sequence of nucleotides. A first bond precursor is attached to a sugar moiety of one of the nucleotides of the first site. The first bond precursor is for forming a covalent bond. The methods include selecting a second site in the oligonucleotide. The second site is selected to have a second specific nucleotide sequence. A second bond precursor is attached to a sugar moiety of a nucleotide of the second site. The second bond precursor also is for forming a covalent bond. The methods further include forming a covalent cross-linkage between the first and the second bond precursors. The covalent cross-linkage is independent of the phosphate inter-nucleoside linkages that define the primary structure of the oligonucleotide. In certain embodiments, the methods further include attaching the first bond precursor to a sugar moiety of a nucleotide of the first site by reacting a precursor of the nucleotide with a

suitable reagent, derivatizing the nucleotide precursor to a protected and activated form, and then synthetically incorporating the derivatized nucleotide precursor into the first nucleotide sequence. In still other embodiments, the methods further include attaching the second bond precursor to a sugar moiety of a nucleotide of the second site by reacting a precursor of the nucleotide with a suitable reagent, derivatizing the nucleotide precursor to a protected and activated form, and synthetically incorporating the derivatized precursor into the second nucleotide sequence.

The invention further includes methods for the intra-strand cross-linkage of an oligonucleotide. In certain embodiments, the methods include the steps of selecting a first nucleotide sequence and a second nucleotide sequence in the oligonucleotide, incorporating an abasic nucleotide within the first sequence, attaching a bond precursor to a nucleotide of the second sequence, and forming a covalent cross-linkage between the abasic site and the bond precursor.

In other embodiments, the methods include selecting a first nucleotide sequence and a second nucleotide sequence in the oligonucleotide, attaching a first bond precursor to a sugar moiety of a nucleotide of the first specific sequence, attaching a second bond precursor to a sugar moiety of a nucleotide of the second sequence, and forming covalent cross-linkage between the first and the second bond precursors. This covalent cross-linkage is independent of the phosphate inter-nucleotide linkages that define the primary structure of the oligonucleotide. Preferably, the first nucleotide sequence is selected to be complementary to and hybridizable with the second nucleotide sequence.

The invention further includes methods for cross-linking first and second oligonucleotide strands including the steps of selecting a first nucleotide sequence in a first oligonucleotide strand, selecting a second nucleotide sequence in a second oligonucleotide strand, incorporating an abasic nucleotide within the first nucleotide sequence, attaching a bond precursor to a nucleotide of the second sequence, and

forming a covalent cross-linkage between the abasic site and the bond precursor. These methods can further include selecting the first nucleotide sequence to be complementary to and hybridizable with the second nucleotide sequence, attaching the bond precursor at a 2'-position of a nucleotide of the second sequence, and forming an aldehydic functional group at the abasic site. The bond precursor should be capable of forming an oxime, imine, hydrazone, semicarbazone, thiosemicarbazone, hydrazide-hydrazone, hemiacetal or thiohemiacetal linking group with the abasic site aldehydic functional group.

The invention further includes methods for cross-linking first and second oligonucleotide strands including the steps of selecting a first nucleotide sequence in a first oligonucleotide strand, selecting a second nucleotide sequence in a second oligonucleotide strand, attaching a first bond precursor to a sugar moiety of a nucleotide of the first sequence, attaching a bond precursor on a sugar moiety of a nucleotide of the second sequence, and forming a covalent cross-linkage between the first and the second bond precursors. This covalent cross-linkage is independent of the phosphate inter-nucleotide linkages that define oligonucleotide primary structure.

The following examples are illustrative of the invention. It is understood that this invention is not limited by these illustrative examples but solely by the claims appended hereto.

EXAMPLE 1

Standard Oligonucleotide Synthesis

Oligonucleotide syntheses were performed on an Applied Biosystems 380 B or 394 DNA synthesizer following standard phosphoramidite protocols and cycles using reagents supplied by the manufacture. When modified phosphoramidites are used, a longer coupling time (10-15 min) was employed. The oligonucleotides were normally synthesized in either a 10 μmol scale or a 3 x 1 μmol scale in the "Trityl-On" mode. Standard

deprotection conditions (30% NH_4OH, 55°C, 16 hr) were employed. HPLC was performed on a Waters 600E instrument equipped with a model 991 detector. For analytical chromatography, the following reverse phase HPLC conditions were employed: Hamilton PRP-1 column (15 x 2.5 cm); solvent A: 50mm TEAA, pH 7.0; solvent B: 45mm TEAA with 80% CH_3CN; flow rate: 1.5ml/min; gradient: 5% B for the first 5 minutes, linear (1%) increase in B every minute thereafter. For preparative purposes, the following reverse phase HPLC conditions were employed: Waters Delta Pak Waters Delta-Pak C_4 15 μm, 300A, 25x100 mm column equipped with a guard column of the same material; column flow rate: 5ml/min; gradient: 5% B for the first 10 minutes, linear 1% increase for every minute thereafter. Following HPLC purification, oligonucleotides are detritylated and further purified by size exclusion using a Sephadex G-25 column. All oligonucleotide sequences are listed in a standard 5' to 3' order from left to right.

EXAMPLE 2

Preparation Of Activated and Protected Nucleotides Having A Space-spanning Group And Active Functional Group

A. Preparation of 5'-Dimethoxytrityl-2'-O-(Functionalized Space-spanning Group)Nucleoside Phosphoramidites

i. 5'-Dimethoxytrityl-2'-O-(Pentyl-N-phthalimido)-N6-Benzyladenosine Phosphoramidite

5'-Dimethoxytrityl-2'-O-(pentyl-N-phthalimido)-N6-benzyladenosine phosphoramidite was synthesized as per the procedures of United States Patent Applications US91/00243 and 463,358, herein incorporated by reference, starting from adenosine. Briefly, this procedure treats adenosine with NaH in DMF followed by treatment with N-(5-bromopentyl)phthalimide (Transworld Chemical Co. Inc., Rockville, MD). Further treatment with $(CH_3)_3SiCl$, Ph-C(O)-Cl and NH_4OH yielded N6-benzyl protected 2'-pentyl-N-phthalimido functionalized adenosine. Treatment with DIPA and CH_2Cl_2 added a DMT blocking group at the 5'-position. Finally phosphitylation gave the desired phosphoramidite compound.

ii. 5'-Dimethoxytrityl-2'-O-(2-Methyloctyl-N-phthalimido)-N4-Benzylcytidine Phosphoramidite

a. N-(8-bromo-2-methyloctyl)phthalimide

8-Bromo-2-methylaminooctane is treated with phthalic anhydride in toluene in the presence of triethylamine under reflux. The solvent is removed *in vacuo*. The residue, when triturated with water, acidified with hydrochloric acid and crystallized from alcohol will yield N-(8-bromo-2-methyloctyl)phthalimide.

b. Preparation of 5'-Dimethoxytrityl-2'-O-(2-Methyloctyl-N-Phthalimido)-N4-Benzylcytidine Phosphoramidite

Cytidine, when treated with NaH and N-(8-Bromo-2-methyloctyl)phthalimide followed by treatment with DIPA and CH_2Cl_2 and phosphitylation utilizing the procedure of Example 2-A-i, will give 5'-dimethoxytrityl-2'-O-(2-methyloctyl-N-phthalimido)-N4-benzylcytidine phosphoramidite.

c. 5'-Dimethoxytrityl-2'-O-(Eicosyl-N-Phthalimido)Adenosine Phosphoramidite

N-(20-bromoeicosyl)phthalimide is prepared as per the procedure of Example 2-A-ii-a from 1-amino-20-bromoeicosane. Adenosine is treated with NaH in DMF and 1-amino-20-bromoeicosane, benzylated, treated with DIPA and CH_2Cl_2 and phosphitylation as per the procedure of Example 2-A-ii will yield the title compound.

iii. 5'-Dimethoxytrityl-2'-O-(Eicosyl-N-Phthalimido)Uridine Phosphoramidite

N-(20-bromoeicosyl)phthalimide is prepared as per the procedure of Example 2-A-ii-a from 1-amino-20-bromoeicosane. Uridine is treated as per the procedure of Akiyama, *et al.*, *NOTES Bull. Chem. Soc. Jpn.*, **1990**, *63*, 3356, to give 3-(4-methoxyphenylmethyl)-3',5'-O-(tetraisopropyldisiloxane-1,3-diyl)uridine. Briefly this treatment first blocks the 3' and 5' sugar hydroxyls of uridine followed by blocking the N3 imide group of the pyrimidine heterocycle with a *p*-methoxybenzyl group. Treatment of the blocked uridine with Ag_2O and N-(20-bromoeicosyl)phthalimide will yield a 2'-alkylated blocked

intermediate compound. The tetraisopropyldisiloxane blocking group is removed by taking the uridine intermediate up in CH_3CN and treating with 60% HF. The *p*-methoxyphenylmethyl blocking group is removed by treatment with $AlCl_3$ in anisole under a N_2 atmosphere followed by Hcl. Both of these procedures are conducted as per the above procedure of Akiyama et al. Further treatment as per the procedure of Example 2-A-ii with DIPA and CH_2Cl_2 to add a DMT blocking group at the 5'-position followed by phosphitylation will yield the title compound.

iv. 5'-Dimethoxytrityl-2'-O-(2-Amylenyl-N-phthalimido)-N2-Isobutyrylguanosine Phosphoramidite

N-(5-bromo-2-amylenyl)phthalimide is prepared as per the procedure of Example 2-A-ii-a from 1-amino-5-bromoamyl-2-ene. 3',5'-O-(Tetraisopropyldisiloxane-1,3-diyl)-2-chloro-6-(2,6-dichlorophenoxy)purine riboside is prepared as described by Sproat, *Nucleic Acid Res.*, **1990**, *18*, 41. Treatment of the latter compound with the first followed by deblocking and enzymatic deaminization with adenosine deaminase utilizing the procedure of Sproat, *Nucleic Acids Research*, **1991**, *19*, 733, will yield the desired 2'-alkylated guanosine intermediate. Further treatment of this alkylated guanosine intermediate as per the procedure of Example 2-A-i with $(CH_3)_3SiCl$, isobutyryl chloride and NH_4OH yields N2-isobutyryl protected 2'-amylenyl-N-phthalimido functionalized guanosine. Treatment with DIPA and CH_2Cl_2 to add a DMT blocking group at the 5'-position followed by phosphitylation will give the desired phosphoramidite compound.

v. 5'-Dimethoxytrityl-2'-O-(Ethyl-N-Phthalimido)Thymidine Phosphoramidite

Thymidine is blocked and treated with Ag_2O and N-(2-bromoethyl)phthalimide (Aldrich Chemical Co., Inc., Milwaukee, WI) as per the procedure of Example 2-A-iii followed by further treatment as per the procedure of Example 2-A-i with DIPA and CH_2Cl_2 to add a DMT blocking group at the 5'-position and finally phosphitylation will yield the title compound.

vi. Preparation of Other 5'-Dimethoxytrityl-2'-O-(Phthalimido Blocked Amine Functionalized Space-spanning Group)-Nucleoside Phosphoramidites

a. Preparation of N-(Ω-bromoalkyl, Ω-bromoalkenyl and Ω-bromoalkynyl)phthalimides

In the manner of Example 2-A-ii-a the following Ω-bromo phthalimide blocked amine compounds, useable as amine functionalized space-spanning groups, can be prepared:

N-(7-bromoheptyl)phthalimide;N-(8-bromooctyl)phthalimide; N-(9-bromononyl)phthalimide; N-(10-bromododecyl)-phthalimide; N-(11-bromoundecyl)phthalimide; N-(12-bromodocecyl)phthalimide; N-(13-bromotridecyl)phthalimide; N-(14-bromotetradecyl)phthalimide; N-(15-bromopentadecyl)-phthalimide; N-(16-bromohexadecyl)phthalimide; N-(17-bromoheptadecyl)phthalimide;N-(18-bromooctadecyl)phthalimide; N-(19-bromononadecyl)phthalimide; N-(3-bromo-2-methyl-propyl)phthalimide; N-(4-bromo-2-methyl-3-ethyl-butyl)phthalimide; N-(3-bromo-2,2-diethylpropyl)phthalimide; N-(4-bromo-3-propylbutyl)phthalimide; N-(10-bromo-2,8-dibutyl-decyl)phthalimide; N-(8-bromo-6,6-dimethyloctyl)phthalimide; N-(8-bromo-6-propyl-6-butyloctyl)phthalimide; N-(4-bromo-2-methylbutyl)phthalimide; N-(5-bromo-2-methylpentyl)phthalimide; N-(5-bromo-3-methylpentyl)phthalimide; N-(6-bromo-2-ethyl-hexyl)phthalimide; N-(5-bromo-3-penten-2-one)phthalimide; N-(4-bromo-3-methyl-2-butanol)phthalimide; N-(8-bromo-3-amino-2-cyanooctyl)phthalimide; N-(7-bromo-3-methoxy-4-heptanal)phthalimide; N-(4-bromo-2-iodo-3-nitrobutyl)-phthalimide; N-(12-bromo-4-isopropoxydodecyl)phthalimide; N-(10-bromo-4-azido-2-nitrodecyl)phthalimide; N-(9-bromo-5-mercaptononyl)phthalimide; N-(5-bromo-4-aminopentenyl)phthalimide;N-(5-bromo-penten-2-yl)phthalimide; N-(3-bromoallyl)phthalimide; N-(4-bromocrotyl)phthalimide; N-(3-bromopropargyl)phthalimide; N-(1-bromonaphth-4-yl)phthalimide; N-(2-bromoanthrac-7-yl)phthalimide; and N-(2-bromophenanthr-6-yl)phthalimide.

b. Preparation of 5'-Dimethoxytrityl-2'-O-(N-Phthalimide Functionalized Space-spanning Group)Nucleoside Phosphoramidites

In the manner of Examples 2-A-i, 2-A-ii-b, 2-A-iii, 2-A-iv and 2-A-v, the following 5'-dimethoxytrityl-2'-O-(space-spanning group-N-phthalimide)nucleoside phosphoramidite can be prepared:

5'-dimethoxytrityl-2'-O-(dodecyl-N-phthalimido)-N4-benzylcytidine phosphoramidite; 5'-dimethoxytrityl-2'-O-(tetradecyl-N-phthalimido)-N6-benzyladenosine phosphoramidite; 5'-dimethoxytrityl-2'-O-(2-methylpropyl-N-phthalimido)uridine phosphoramidite; 5'-dimethoxytrityl-2'-O-(2,8-dibutyldecyl-N-phthalimido)-N2-isobutyrylguanosine phosphoramidite; 5'-dimethoxytrityl-2'-O-(3-methylpentyl-N-phthalimido)thymidine phosphoramidite; 5'-dimethoxytrityl-2'-O-(2-cyanooctyl-N-phthalimido)-N4-benzylcytidine phosphoramidite; 5'-dimethoxytrityl-2'-O-(3-methoxy-4-heptanal-N-phthalimido)uridine phosphoramidite; 5'-dimethoxytrityl-2'-O-(3-nitrobutyl-N-phthalimido)-N6-adenosine phosphoramidite; 5'-dimethoxytrityl-2'-O-(propargyl-N-phthalimido)-N2-isobutyrylguanosine phosphoramidite; and 5'-dimethoxytrityl-2'-O-(3-anthrac-7-yl-N-phthalimido)thymidine phosphoramidite.

B. Preparation of 5'-Dimethoxytrityl-2'-S-(Amine Functionalized Space-spanning Group)-2'-Deoxy-Nucleoside Phosphoramidites

i. Preparation of N-(1-Bromo-8-dodecyl)Phthalimide

1-Aminododecyl-12-ol is treated with N-ethoxycarbonylphthalimide utilizing the reaction conditions of McArthur, *et. al.*, *Synth. Commun.*, **1983**, *13*, 311, to yield N-(dodecyl-12-ol)phthalimide which in turn when treated with phosphorous tribromide and pyridine utilizing the reaction conditions of Hall, *et al.*, *Synthetic Procedures in Nucleic Acid Chemistry*, Zorbach and Tipson, Editors, Volume 1, John Wiley & Sons, 1968 will yield N-(12-bromododecyl)phthalimide.

ii. 5'-Dimethoxytrityl-2'-S-(Dodec-12-N-Phthalimido)-2'-Deoxyuridine Phosphoramidite

2'-Deoxy-2-thiouridine is prepared as per the procedure of Divakar, *J. Chem. Soc. Perkin Trans.*, **1990**, *1*, 969. 2'-Deoxy-2-thiouridine when treated with NaH in DMF followed by treatment with N-(12-bromododecyl)phthalimide will yield 2'-S-(dodecyl-N-phthalimido) functionalized uridine. Treatment with DIPA and CH_2Cl_2 to add a DMT blocking group at the 5'-position followed by phosphitylation will give the desired phosphoramidite compound.

C. Preparation of 5'-Dimethoxytrityl-2'-NH-(Alkyl-N-Phthalimido)-2'-Deoxy-Nucleoside Phosphoramidites

i. N-(Butyl-4-al)Phthalimide (an Ω-aldehydic terminating α-N-Phthalimide functionalized space-spanning group)

4-Aminobutyraldehyde diethyl acetal (Aldrich Chemical Co., Inc., Milwaukee, WI) when treated with phthalic anhydride in toluene in the presence of triethylamine as per the procedure of Example 2-A-ii-a will give an N-phthalimide blocked amine group at the 1-position of the butyl group and an diethyl acetal blocked aldehyde group at the 4-position. The diethyl acetal blocking group is then removed by acid catalyzed hydrolysis to give N-(butyl-4-al)phthalimide.

ii. 5'-Dimethoxytrityl-2'-NH-(Butyl-N-Phthalimido)-2'-Deoxy-N6-Benzyladenosine Phosphoramidite

2'-Amino-N6-benzyl-2'-deoxyadenosine, prepare as per the procedure of G. Burke, *et. al.*, *J. Carbohydrates, Nucleosides, Nucleotides*, **1988**, 7, 63, and N-(butyl-4-al)phthalimide are stirred in dry CH_2Cl_2. The solvent is removed and the residue taken up in NaOAc buffer (pH 5.02) and treated with $NaCNBH_3$ to reduce the intermediate imine to the desired amine linking functionality. Further treatment as per the procedure of Example 2-A-i with DIPA and CH_2Cl_2 to add a DMT blocking group at the 5'-position followed by phosphitylation will yield the title compound.

D. Preparation of 5'-Dimethoxytrityl-2'-(Alkyl-N-Phthalimido)-2'-Deoxy-Nucleoside Phosphoramidites

Introduction of a 2' directly alkylated nucleotide (i.e. a methylene linking group joining a space-spanning group to the 2'-position of a nucleotide sugar moiety) into an oligonucleotide is accomplished by free radical alkylation of the corresponding nucleoside followed by conversion to the appropriate 5'-dimethoxytrityl phosphoramidite nucleotides.

i. 5'-Dimethoxytrityl-2'-(Allyl-N-Phthalimido)-2'-Deoxy-N6-Benzyladenosine Phosphoramidite

2'-O-(Phenoxythioxomethyl)-2'-deoxyadenosine will be treated with N-(5-bromopenten-2-yl)phthalimide (from Example 2-A-vi) and tributyltin hydride utilizing conditions analogous to those described in Flandor, *et al.*, *Tetrahedron Letters*, **1990**, *31*, 597, to effect a free radical alkylation at the 2'-position of the nucleoside. The resulting alkylated compound is then further treated with $(CH_3)_3SiCl$, Ph-C(O)-Cl and NH_4OH to give the N6-benzyl protected 2'-[N-(penten-2-yl)phthalimido] functionalized adenosine. This is then followed by treatment with DIPA and CH_2Cl_2 to add a DMT blocking group at the 5'-position which in turn is followed by phosphitylation, all as per the procedure of Example 2-A-i, to give the desired phosphoramidite compound.

ii. 5'-Dimethoxytrityl-2'-(Allyl-N-Phthalimido)-2'-Deoxyuridine Phosphoramidite

2'-O-(Phenoxythioxomethyl)-2'-deoxyuridine will be treated with N-(5-bromopenten-2-yl)phthalimide (from Example 2-A-vi) and tributyltin hydride utilizing conditions analogous to those described in Flandor, *et al.*, *Tetrahedron Letters*, **1990**, *31*, 597 to effect a free radical alkylation at the 2'-position of the nucleoside. The resulting alkylated compound is then further treated with DIPA and CH_2Cl_2 to add a DMT blocking group at the 5'-position which in turn is followed by phosphitylation, all as per the procedure of Example 2-A-iii, to give the desired phosphoramidite compound.

E. Preparation of 5'-Dimethoxytrityl-2'-O-(Hydrazine Functionalized Space-spanning Group) Nucleoside Phosphoramidites

i. 8-Bromooctylbenzylcarbazide (an α-bromo carbazide protected Ω-hydrazine functionalized space-spanning group)

8-Bromooctanol is tosylated with tosyl chloride in pyridine. The solvent is removed and the 8-bromooctanol O-tosylate is treated with benzylcarbazide and activated molecular sieves in anhydrous dimethylacetamide overnight at 110°C. The solvent is removed and the residue purified by silica gel chromatography to yield bromooct-8-ylbenzylcarbazide.

ii. 5'-Dimethoxytrityl-2'-O-(Oct-8-ylbenzylcarbazide) uridine Phosphoramidite

Uridine will be treated with 8-bromooctylbenzylcarbazide utilizing the method of Example 2-A-iii to give the title compound.

iii. 5'-Dimethoxytrityl-2'-O-(Oct-8-ylbenzylcarbazide)-N4-Benzylcytidine Phosphoramidite

Cytidine will be treated with 8-bromooctylbenzylcarbazide utilizing the method of Example 2-A-ii-b to give the title compound.

iv. 5'-Dimethoxytrityl-2'-O-(Oct-8-ylbenzylcarbazide)-N2-Isobutyrylguanosine Phosphoramidite

Guanosine will be treated with 8-bromooctylbenzylcarbazide utilizing the method of Example 2-A-iv to give the title compound.

v. 5'-Dimethoxytrityl-2'-O-(Oct-8-ylbenzylcarbazide)-N6-Benzyladenosine Phosphoramidite

Adenosine will be treated with 8-bromooctylbenzylcarbazide utilizing the method of Example 2-A-i to give the title compound.

F. Preparation of 5'-Dimethoxytrityl-2'-O-(Hydroxylamine-Functionalized Space-spanning Group)Nucleoside Phosphoramidites

i. O-(8-Bromooctanol)phthalimide (an α-bromo O-phthalimide protected Ω-hydroxylamine functionalized space-spanning group)

1,8-Octanediol is treated with one equivalent of dihydropyrane in the presence of acid to give the 8-(tetrahydropyran-2-yl)octanol. 8-(Tetrahydropyran-2-yl)octanol, N-hydroxyphthalimide and triphenylphosphine in DMF are treated at 0°C with diisopropylazodicarboxylate utilizing Mitsunobu reaction conditions to give O-[8-(tetrahydropyran-2-yl)octanol)phthalimide. Acid treatment of O-[8-(tetrahydropyran-2-yl)octanol]phthalimide will yield O-(octanol)phthalimide which in turn when treated with phosphorous tribromide and pyridine utilizing the reaction conditions of Hall, *et al.*, *Synthetic Procedures in Nucleic Acid Chemistry*, Zorbach and Tipson, Editors, Volume 1, John Wiley & Sons, 1968 will yield O-(8-bromooctanol)phthalimide.

ii. 5'-Dimethoxytrityl-2'-O-(Octan-8-ol-O-Phthalimido)-N4-Benzylcytidine Phosphoramidite

Cytidine is treated with NaH in DMF followed by treatment with O-(8-bromooctanol)phthalimide from Example 2-F-i. Further treatment with $(CH_3)_3SiCl$, Ph-C(O)-Cl and NH_4OH will yield N4-benzyl protected 2'-octyl-O-phthalimido functionalized cytidine. Treatment with DIPA and CH_2Cl_2 will add a DMT blocking group at the 5'-position. Finally phosphitylation as per the procedure of Example 2-A-ii will give the desired phosphoramidite compound.

iii. 5'-Dimethoxytrityl-2'-O-(Octan-8-ol-O-Phthalimide)uridine Phosphoramidite

Uridine is blocked and then treated with Ag_2O and O-(8-bromooctanol)phthalimide from Example 2-F-i. Treatment with DIPA and CH_2Cl_2 will add a DMT blocking group at the 5'-position. Finally phosphitylation as per the procedure of Example 2-A-iii will give the desired phosphoramidite compound.

G. Preparation of 5'-Dimethoxytrityl-2'-O-(Semicarbazide Functionalized Space-spanning Group)Nucleoside Phosphoramidites

i. 5'-Dimethoxytrityl-2'-O-(pent-5-ylsemicarbazide)uridine Phosphoramidite

Uridine is blocked and then treated with Ag_2O and N-(8-bromopenyl)semicarbazide utilizing the method of Example 2-A-iii to give the title compound.

ii. 5'-Dimethoxytrityl-2'-O-(pent-5-ylsemicarbazide)-N6-Benzyladenosine Phosphoramidite

Adenosine will be treated with N-(8-bromopenyl)-semicarbazide utilizing the method of Example 2-A-i to give the title compound.

H. Preparation of 5'-Dimethoxytrityl-2'-O-(Hydrazide Functionalized Space-spanning Group)Nucleoside Phosphoramidites

i. α-Cloroethoxycarbonylethane

Chloroacetyl chloride (Air Products, Emmaus, PA) is added drop wise to stirring ethanol to esterify the acid chloride. The ester is then recovered by distillation.

ii. 5'-Dimethoxytrityl-2'-O-(ethoxycarbonylethyl)uridine Phosphoramidite

Uridine is blocked and then treated with Ag_2O and α-cloroethoxycarbonylethane as per the procedure of Example 2-A-iii to yield the title compound.

iii. 5'-Dimethoxytrityl-2'-O-(eth-2-ylhydrazide)uridine Phosphoramidite

5'-Dimethoxytrityl-2'-O-(ethoxycarbonylethyl)uridine phosphoramidite will be treated with hydrazine to give the corresponding 2'-O-(ethylhydrazide)uridine title compound.

I. Preparation of 5'-Dimethoxytrityl-2'-O-(Alcohol Functionalized Space-spanning Group)Nucleoside Phosphoramidites

i. 1-O-(Tetrahydropyran-2-yl)-7-bromoheptanol

7-Bromoheptanol will be treated with dihydropyrane in the presence of acid to give 1-O-(tetrahydropyran-2-yl)-7-bromoheptanol.

ii. 5'-Dimethoxytrityl-2'-O-[O-(Tetrahydropyran-2-yl)heptan-7-ol)]-N4-Benzylcytidine Phosphoramidite

Cytidine will be treated with 1-O-(tetrahydropyran-2-yl)-7-bromoheptanol as per the procedure of Example 2-A-ii to give the title compound.

iii. 5'-Dimethoxytrityl-2'-O-[O-(Tetrahydropyran-2-yl)heptan-7-ol)]uridine Phosphoramidite

Uridine is blocked and then treated with Ag_2O and 1-O-(tetrahydropyran-2-yl)-8-bromoheptanol as per the procedure of Example 2-A-iii to give the title compound.

J. Preparation of 5'-Dimethoxytrityl-2'-O or 2'-S (Thiol Functionalized Space-spanning Group)-Nucleoside Phosphoramidites

i. Via Ring Opening of 2,2'-Anhydro Pyrimidines

a. 2'-S-(Octan-8-Thiol)thymidine

2,2'-Anhydro thymidine will be ring opened with octane-1,8-dithiol to give 2'-S-(octan-8-thiol)thymidine.

b. 5'-Dimethoxytrityl-2'-S-[S-Trityl-(Octan-8-thiol)]-2'-Deoxy-Thymidine Phosphoramidite

2'-S-(octan-8-thiol)thymidine when treated with trityl chloride will give 2'-S-[S-trityl(octan-8-thiol)]thymidine which will then be further treated with DIPA and CH_2Cl_2 to add a DMT blocking group at the 5'-position and finally phosphitylation as per Example 2-A-v to give the title compound.

ii. Via O-Alkylation

a. 1-S-Trityl-8-Bromohexanethiol

8-Bromohexanethiol will be tritylated with trityl chloride to give 1-S-trityl-8-bromohexanethiol.

b. 5'-Dimethoxytrityl-2'-O-[S-Trityl-(Hexan-8-thiol)]-N6-Benzyladenosine Phosphoramidite

Adenosine is treated with NaH in DMF as per example 2-A-i. The reaction product will then be treated with 1-S-trityl-8-bromohexanethiol followed by treatment with $(CH_3)_3SiCl$, Ph-C(O)-Cl and NH_4OH to yield the N6-benzyl protected 2'-O-[S-trityl(hexyl-8-thiol)] functionalized adenosine. Treatment with DIPA and CH_2Cl_2 to add a DMT blocking group at the 5'-

position followed by phosphitylation, further as per the procedure of Example 2-A-i, will give the title compound.

c. 5′-Dimethoxytrityl-2′-O-[S-Trityl-(Hexan-8-thiol)]uridine Phosphoramidite

Uridine is blocked and then treated with Ag_2O and 1-S-trityl-8-bromohexanethiol as per Example 2-A-iii. The reaction product will then be treated with followed by treatment with DIPA and CH_2Cl_2 to add a DMT blocking group at the 5′-position and finally phosphitylation, further as per the procedure of Example 2-A-iii, to give the title compound.

d. 1-S-Trityl-8-Bromopentanethiol

8-Bromopentanethiol will be tritylated with trityl chloride to give 1-S-trityl-8-bromopentanethiol.

e. 5′-Dimethoxytrityl-2′-O-[S-Trityl-(Pentan-8-thiol)]uridine Phosphoramidite

Uridine is blocked and then treated with Ag_2O and 1-S-trityl-8-bromopentanethiol as per Example 2-A-iii followed by treatment with DIPA and CH_2Cl_2 to add a DMT blocking group at the 5′-position and finally phosphitylation, further as per the procedure of Example 2-A-iii, to give the title compound.

K. Preparation of 5′-Dimethoxytrityl-2′-O-(Protected Aldehyde Functionalized Space-spanning Group)Nucleoside Phosphoramidites

i. Preparation of Acetal Functionalization

a. N6-Benzyl-5′-Dimethoxytrityl-2′-O-[Propion-3-al Bis(o-Nitrobenzyl Acetal)adenosine Phosphoramidite

Adenosine is treated with NaH and 3-bromopropionaldehyde bis(*o*-nitrobenzyl) acetal as per the procedure of Example 2-A-i. Further treatment with $(CH_3)_3SiCl$, Ph-C(O)-Cl and NH_4OH to add a benzyl blocking group followed by DIPA and CH_2Cl_2 to add a DMT blocking group at the 5′-position and further followed by phosphitylation as per the procedure of Example 2-A-i, will give the desired compound.

b. 5'-Dimethoxytrityl-2'-O-[Propion-3-al Bis(*o*-Nitrobenzyl Acetal)uridine Phosphoramidite

Uridine is blocked followed by treatment with Ag_2O and 3-bromopropionaldehyde bis(*o*-nitrobenzyl) acetal as per the procedure of Example 2-A-iii. Further treatment with DIPA and CH_2Cl_2 to add a DMT blocking group at the 5'-position and further followed by phosphitylation as per the procedure of Example 2-A-i, will give the desired compound.

ii 6-(Diphenylimidazolidino)hexyl Bromide

6-Bromohexionyl acid chloride will be treated with ethylmercaptan. The resulting ethyl thiol-6-bromohexionate is then desulfurized with Raney nickel giving 6-bromohexenal. The 6-bromohexenal is then treated with 1,2-dianilinoethylene in benzene/DMSO to give the title 6-(diphenylimidazolidino)hexyl bromide.

iii. 5'-Dimethoxytrityl-2'-O-[6-(diphenylimidazolidino)hexyl]-N2-Isobutyrylguanosine Phosphoramidite

3',5'-O-(Tetraisopropyldisiloxane-1,3-diyl)-2-chloro-6-(2,6-dichlorophenoxy)purine riboside is reacted with 6-(diphenylimidazolidino)hexyl bromide as per Example 2-A-iv. Further treatment with $(CH_3)_3SiCl$, isobutyryl chloride and NH_4OH will yield N2-isobutyryl protected 2'-(diphenylimidazolidino)hexyl functionalized guanosine. Treatment with DIPA and CH_2Cl_2 to add a DMT blocking group at the 5'-position followed by phosphitylation, further as per the procedure of Example 2-A-iv, will give the title compound.

iv. 5'-Dimethoxytrityl-2'-O-[6-(diphenylimidazolidino)hexyl]-N6-Benzyladenosine Phosphoramidite

Adenosine is treated with NaH/DMF followed by treatment with 6-(diphenylimidazolidino)hexyl bromide as per Example 2-A-i. Further treatment with $(CH_3)_3SiCl$, benzyl chloride and NH_4OH will yield N6-benzyl protected 2'-(diphenylimidazolidino)hexyl functionalized adenosine. Treatment with DIPA and CH_2Cl_2 to add a DMT blocking group at the 5'-position followed by phosphitylation, further as per the procedure of Example 2-A-i, will give the title compound.

v. 5'-Dimethoxytrityl-2'-O-[6-(diphenylimidazolidino)hexyl)-N4-Benzylcytidine Phosphoramidite

Cytidine is treated with NaH/DMF followed by treatment with 6-(diphenylimidazolidino)hexyl bromide as per Example 2-A-ii. Further treatment with $(CH_3)_3SiCl$, benzyl chloride and NH_4OH will yield N4-benzyl protected 2'-(diphenylimidazolidino)hexyl functionalized cytidine. Treatment with DIPA and CH_2Cl_2 to add a DMT blocking group at the 5'-position followed by phosphitylation, further as per the procedure of Example 2-A-ii, will give the title compound.

vi. 5'-Dimethoxytrityl-2'-O-[6-(diphenylimidazolidino)hexyl]uridine Phosphoramidite

Uridine is blocked and then treated with Ag_2O and 6-(diphenylimidazolidino)hexyl bromide as per Example 2-A-iii. Treatment with DIPA and CH_2Cl_2 to add a DMT blocking group at the 5'-position followed by phosphitylation, further as per the procedure of Example 2-A-ii, will give the titled compound.

EXAMPLE 3

Preparation Of Oligonucleotides Having An Abasic Sites Located Thereon

A. Preparation of An Abasic Site Containing Oligonucleotide Via Enzymatic Reaction

i. Oligonucleotide Containing Single Uridine Site

An oligonucleotide of the sequence:

Oligomer 1 CGC AGU* CAG CC

wherein U* represents a 2'-deoxy uridine nucleotide, was prepared utilizing the procedure of Example 1. The deoxyuridine nucleotide in the middle of the sequence was added during the synthesis utilizing deoxyuridine phosphoramidite (Glen Research, Sterling, VA). The oligonucleotide was prepared utilizing standard synthesis cycles. It was deprotected by normal deprotection utilizing ammonium hydroxide, 30%, for 16 hours. The solvent was evaporated, and the residue was purified by HPLC and detritylated. Final purification was effected on Sephadex G-25.

ii. Enzyme Stock Solution Preparation

Uracil-DNA glycosylase was isolated from *E. coli* M5219 cells transformed with the expression plasmid pBD396 containing the *ung* gene. The enzyme was purified to electrophoretic homogeneity as described by Lindahl, *et al.*, *J. Biol. Chem.*, **1977**, *252*, 3286, and stored in 30 Mm HEPES-NaOH, pH 7.4, containing 5% glycerol, 2 mM DTT and 1 mM EDTA.

iii. Oligonucleotide Containing Single Abasic Site

An abasic oligonucleotide of the sequence:

Oligomer 2 CGC AGD* CAG CC

wherein D* represents an abasic site, was prepared by treating 237 O.D. units of Oligomer 1 of Example 3-A-i in 0.5 ml water with 200 μl of the stock solution of Example 3-A-ii (200 micrograms of uracil DNA-glycosylase) and incubating at room temperature overnight. HPLC analysis showed quantitative removal of uracil as indicated by a 1:10 ratio between uracil and the abasic dodecamer oligonucleotide. The uracil retention time was 2.43 mins and the abasic oligonucleotide was 21.68 mins. The solution was lyophilized and stored in the freezer until further use.

iv. Oligonucleotide Containing Multiple Uridine Sites

In the manner of Example 3-A-i the following oligonucleotide was prepared:

Oligomer 3 GAC AGA GGU* AGG AGA AGU* GA

wherein U* represents a 2'-deoxyuridine nucleotide. The oligonucleotide when treated as per the procedure of Example 3-A-iii will give an oligonucleotide of the sequence

Oligomer 4 GAC AGA GGD* AGG AGA AGD* GA

wherein D* represents an abasic site within the oligonucleotide.

B. Preparation of An Abasic Site Containing Oligonucleotide Via Abasic Sugar Precursor

i. 5-O-4,4'-Dimethoxytrityl-1,2-Dideoxy-D-Ribofuranose-3-O-(2-Cyanoethyl-N,N'-Diisopropyl) Phosphoramidite

5-O-4,4'-Dimethoxytrityl-1,2-dideoxy-D-ribofuranose-3-O-(2-cyanoethyl-N,N'-diisopropyl) phosphoramidite is prepared as per the procedure of Lyer, *et al.*, *Nucleic Acids Research*, **1990**, *18*, 2855, or as per the procedure of Didier Peoch *et al.*, *Tetrahedron Letters*, **1991**, *32*, 207.

ii. Oligonucleotide Containing Abasic Site

Oligomer 2 of Example 3-A-iii, i.e. the oligonucleotide of the sequence:

Oligomer 2 CGC AGD* CAG CC

wherein D* represents an abasic site, can also be prepared utilizing the synthetic procedures of the papers identified in Example 3-B-i. Utilizing those procedures, an *o*-Nitrobenzyldeoxyfuranose containing oligonucleotide is synthesized using the oligonucleotide synthetic methods of these papers. Photolysis, utilizing a high intensity Hg lamp, generates the corresponding abasic site containing oligonucleotide. Such abasic oligonucleotides are also described by Horn, *et al.*, *Nucleosides & Nucleotides*, **1991**, *10*, 299.

EXAMPLE 4

Preparation of Oligonucleotides Incorporating 2'-O-(Amine Functionalized Space-spanning Group)

A. Single Site

An oligonucleotide of the sequence:

Oligomer 5 GGC TGA* CTG CG

wherein A* indicates a 2'-O-(pentylamino)adenosine nucleotide was prepared as per the procedure of Example 1 utilizing an extend coupling time of 10 minutes during the coupling step of the modified nucleotide unit. 5'-Dimethoxytrityl-2'-O-(pentyl-N-phthalimido)-N6-benzyladenosine phosphoramidite (from Example 2-A-i was utilized in the DNA synthesizer as a 0.09M solution in anhydrous CH_3CN. Oligonucleotide synthesis was carried out

in either an ABI 390B or 394 synthesizer employing the standard synthesis cycles with an extended coupling time of 10 minutes during coupling of the 5'-dimethoxytrityl-2'-O-(pentyl-N-phthalimido)-N6-benzyladenosine phosphoramidite into the oligonucleotide sequence. Coupling efficiency of > 98% was observed for the 5'-dimethoxytrityl-2'-O-(pentyl-N-phthalimido)-N6-benzyladenosine phosphoramidite coupling. Removal of the N-phthalimido amine protecting group is effected simultaneously with NH_4OH deprotection of the benzyl and isobutyryl nucleotide base blocking groups.

B. Multiple Sites

An oligonucleotide of the sequence:

Oligomer 6 CTG TCT CCA* TCC TCT TCA* CT

wherein A* represents a nucleotide modified to incorporate a pentylamino functionality was prepared as per the procedures of Example 4-A. This oligonucleotide is an antisense compound to the E2 region of the bovine papilloma virus-1 (BPV-1).

EXAMPLE 5

A. Coupling of Oligonucleotide Strand Having An Abasic Site To An Oligonucleotide Having An Amine Functionalized Space-spanning Group

Two stranded, duplexed, cross-linked oligonucleotides of the structure:

Cross-linked Strands I

GGC AGD CAG CC
|
GGC TGA CTG CG

where " | " represents an $O-(CH_2)_5-N-CH_2-$ cross-linkage between the oligonucleotide strands is prepared by reacting 2.8 O.D. units of Oligomer 5, i.e. the GGC TGA* CTG CG oligonucleotide of Example 4-A, in 100 μl of 0.2 M NaOAc buffer (pH 5.02) with 2.37 O.D. units of Oligomer 2, i.e. the abasic oligonucleotide CGC AGD* CAG CC of Example 3-A-iii, dissolved in 10 μl of water. The combined solution was let stand for 2 hours. 10 mg of $NaCNBH_3$ was dissolved in 400 μl of 250 mM NaOAc (pH 5.02) and 25 μl of this solution was added to the reaction solution. After this addition, the final concentration of the cyanoboro-

hydride in the solution was nearly 80 mM. The reaction was followed at 1 hr intervals by HPLC. The amine linked oligonucleotide had a retention time of 21.65 mins and the abasic oligonucleotide 22.46 mins. A product peak was observed at 24.07 mins. A slow increase in the product peak was observed. After 12 hrs the product was separated from the reaction mixture by HPLC yielding 1.05 O.D. units, 20.3%. A small aliquot of the product was re-injected into the HPLC column and observed as a single peak.

B. Gel Analysis

A gel analysis was carried out on the above cross-linked 11 mer oligonucleotides, i.e. Cross-linked Strands I. The cross linked material moved slower than either of its component 11 mer oligonucleotides, i.e. Oligomers 2 or 5. Indeed it had the same electrophoretic mobility as that of a 21 mer oligonucleotide (11 + 11 -1 for the loss of a base).

C. Ion Exchange Chromatography

Cross-linking of the Cross-linked Strands I was further confirmed by using ion exchange chromatography. The 21 mer cross-linked strands, since they have the greatest amount of charge, were eluted at 0.25 M salt concentration whereas the 11 mer components oligonucleotides were eluted at 0.15M salt concentration.

D. Cross-Linking Affirmation

Cross-linking as opposed to hybridization duplex formation was confirmed independently by gel analysis and by re-injection of the cross-linked strands into HPLC. Gel analysis of a mixture the two component oligonucleotides mixed together without cross-linking did not yield the 21 mer mobility product. Re-injection of the cross-linked product into HPLC did not produce an equilibrium mixture of the two single oligonucleotide strands but came out as a single peak.

EXAMPLE 6

Preparation of Oligonucleotide Having Sulfur Linking Atom Attaching An Amine Functionalized Space-spanning Groups Thereon

An oligonucleotide of the sequence:

Oligomer 7 GGC T*GA CTG CG

wherein T* represents a nucleotide modified to incorporate a dodecylamino functionality joined to the thymidine nucleotide via a sulfur linking atom will be prepared as per the procedure of Example 4-A utilizing the oligonucleotide of Example 2-B-ii.

EXAMPLE 7

Preparation of Oligonucleotide Having Nitrogen Linking Atom Attaching An Amine Functionalized Space-spanning Groups Thereon

An oligonucleotide of the sequence:

Oligomer 8 GGC TGA* CTG CG

wherein A* represents a nucleotide modified to incorporate a butylamino functionality joined to the adenosine nucleotide via a nitrogen linking atom will be prepared as per the procedure of Example 4-A utilizing the oligonucleotide of Example 2-C-ii.

EXAMPLE 8

Preparation of Oligonucleotide Having Carbon Linking Atom Attaching An Amine Functionalized Space-spanning Groups Thereon

An oligonucleotide of the sequence:

Oligomer 9 GGC TGA* CTG CG

wherein A* represents a nucleotide modified to incorporate a allylamine functionality joined to the adenosine nucleotide via a methylene linking atom will be prepared as per the procedure of Example 4-A utilizing the oligonucleotide of Example 2-D-i.

EXAMPLE 9

A. Coupling of Oligonucleotide Strands Via Abasic Site and Amine Functionalized Space-spanning Group Connecting Through A Sulfur Linking Group

Two stranded, duplexed, cross-linked oligonucleotides of the structure:

Cross-linked Strands II

GGC AGD CAG CC
|
GGC TGA CTG CG

where " | " indicates represents an S-$(CH_2)_{12}$-N-CH_2- cross-linkage between the individual oligonucleotides of the structure will be prepared by reacting Oligomer 7 of Example 6 with the abasic Oligomer 2 of Example 3-A-iii utilizing the procedure of Example 5-A.

B. Coupling of Oligonucleotide Strands Via Abasic Site and Amine Functionalized Space-spanning Group Connecting Through A Nitrogen Linking Group

Two stranded, duplexed, cross-linked oligonucleotides of the structure:

Cross-linked Strands III

GGC AGD CAG CC
|
GGC TGA CTG CG

where " | " indicates represents an -NH-$(CH_2)_4$-N-CH_2- cross-linkage between the individual oligonucleotides of the structure will be prepared by reacting Oligomer 8 of Example 7 with the abasic Oligomer 2 of Example 3-A-iii utilizing the procedure of Example 5-A.

C. Coupling of Oligonucleotide Strands Via Abasic Site and Amine Functionalized Space-spanning Group Connecting Through A Methylene Linking Group

Two stranded, duplexed, cross-linked oligonucleotides of the structure:

Cross-linked Strands IV

GGC AGD CAG CC
|
GGC TGA CTG CG

where " | " indicates represents an -CH_2-CH=CH-N-CH_2- cross-linkage between the individual oligonucleotides of the structure will be prepared by reacting Oligomer 9 of Example 8 with the abasic Oligomer 2 of Example 3-A-iii utilizing the procedure of Example 5-A.

EXAMPLE 10

A. Preparation of Oligonucleotide Having A Hydrazine Functionalized Space-spanning Groups Thereon

i. Single Functionality

An oligonucleotide of the sequence:

Oligomer 10 GGC TGU* CTG CG

wherein U* represents a nucleotide modified to incorporate a 2'-O-(octylhydrazino) functionality joined to the uridine nucleotide will be prepared as per the procedure of Example 4-A utilizing the oligonucleotide of Example 2-E-ii followed by deblocking of the benzylcarbazide protecting group via hydrogenolysis in the presence of palladium on charcoal.

ii. Multiple Functionalities

An oligonucleotide of the structure:

Oligomer 11 AGC CAG AU*C U GA GCC UGG G**AG CUC UCU GGC U

wherein U* represents an uridine nucleotide having a 2'-O-(octylhydrazine) group thereon and G** represents a guanosine nucleotide having a 2'-O-(octylhydrazine) group thereon will be prepared as per the procedure of Example 4-A utilizing the nucleotides of Examples 2-E-ii and 2-E-iv, respectively, to introduce the hydrazine functionalized space-spanning group containing nucleotides in the proper sequence. Deblocking of the benzylcarbazide protecting group is effected via hydrogenolysis in the presence of palladium on charcoal.

B. Preparation of Oligonucleotide Having A Hydroxylamine Functionalized Space-spanning Groups Thereon

An oligonucleotide of the sequence:

Oligomer 12 GGC TGU* CTG CG

wherein U* represents a nucleotide modified to incorporate a octyl-hydroxylamine functionality joined to the cytidine nucleotide via an oxygen linking atom will be prepared as per the procedure of Example 4-A utilizing the oligonucleotide of Example 2-F-iii followed by treatment of the O-phthalimide group with methylhydrazine in DCM to generate the hydroxylamine functionality on the octyl space-spanning group.

C. Preparation of Oligonucleotide Having A Semicarbazide Functionalized Space-spanning Groups Thereon

An oligonucleotide of the sequence:

Oligomer 13 GGC TGU* CTG CG

wherein U* represents a nucleotide modified to incorporate a pentyl-N-semicarbazide functionality joined to the uridine nucleotide via an oxygen linking atom will be prepared as per the procedure of Example 4-A utilizing the oligonucleotide of Example 2-G-i.

D. Preparation of Oligonucleotide Having A Hydrazide Functionalized Space-spanning Groups Thereon

An oligonucleotide of the sequence:

Oligomer 14 GGC TGU* CTG CG

wherein U* represents a nucleotide modified to incorporate an ethyl hydrazide functionality joined to the uridine nucleotide via an oxygen linking atom will be prepared as per the procedure of Example 4-A utilizing the oligonucleotide of Example 2-H-iii.

E. Preparation of Oligonucleotide Having An Alcohol Functionalized Space-spanning Groups Thereon

An oligonucleotide of the sequence:

Oligomer 15 GGC TGU* CTG CG

wherein U* represents a nucleotide modified to incorporate a heptan-7-ol functionality joined to the uridine nucleotide will be prepared as per the procedure of Example 4-A utilizing the oligonucleotide of Example 2-I-iii.

EXAMPLE 11

A. Coupling of Oligonucleotide Strands Via Abasic Site and Hydrazine or Hydrazone Functionalized Space-spanning Group

i. Hydrazine Linked Strands

Two stranded, duplexed, cross-linked oligonucleotides of the structure:

Cross-linked Strands V

```
GGC AGD CAG CC
    |
GGC TGU CTG CG
```

where " | " indicates represents an O-$(CH_2)_8$-NH-NH-CH_2- cross-linkage between the individual oligonucleotides of the structure will be prepared by reacting Oligomer 10 of Example 10-A-i with the abasic Oligomer 2 of Example 3-A-iii followed by reduction of the resulting hydrazone linkage with $NaCNBH_3$ in NaOAc buffer as per the procedure of Example 5-A to form the hydrazine linkage.

ii. Hydrazone Linked Strands

The cross-linker strands of Example 13-A-i having a hydrazone cross-linkage can be formed by omitting the $NaCNBH_3$ reduction step of Example 13-A-i.

B. Coupling of Oligonucleotide Strands Via Abasic Site and Hydroxylamine or Oxime Functionalized Space-spanning Group

i. Hydroxylamine Linked Strands

Two stranded, duplexed, cross-linked oligonucleotides of the structure:

Cross-linked Strands VI

GGC AGD CAG CC
|
GGC TGU CTG CG

where " | " indicates represents an O-$(CH_2)_8$-NH-O-CH_2- cross-linkage between the individual oligonucleotides of the structure will be prepared by reacting Oligomer 12 of Example 10-B with the abasic Oligomer 2 of Example 3-A-iii followed by reduction of the resulting oxime linkage with $NaCNBH_3$ in NaOAc buffer as per the procedure of Example 5-A to form the hydroxylamine.

ii. Oxime Linked Strands

The cross-linker strands of Example 13-B-i having an oxime cross-linkage can be formed by omitting the $NaCNBH_3$ reduction step of Example 13-A-i.

C. Coupling of Oligonucleotide Strands Via Abasic Site and Semicarbazide or Semicarbazone Functionalized Space-spanning Group

i. Semicarbazide Linked Strands

Two stranded, duplexed, cross-linked oligonucleotides of the structure:

Cross-linked Strands VII

GGC AGD CAG CC
|
GGC TGU CTG CG

where " | " indicates represents an O-$(CH_2)_8$-NH-NH-CH_2- cross-linkage between the individual oligonucleotides of the structure will be prepared by reacting Oligomer 13 of Example 10-C with the abasic Oligomer 2 of Example 3-A-iii followed by reduction of the resulting semicarbazone linkage with $NaCNBH_3$ in NaOAc buffer as per the procedure of Example 5-A to form the semicarbazide linkage.

ii. Semicarbazone Linked Strands

The cross-linker strands of Example 13-A-i having a semicarbazone cross-linkage can be formed by omitting the $NaCNBH_3$ reduction step of Example 13-A-i.

D. Coupling of Oligonucleotide Strands Via Hemiacetal Linkage Across An Abasic Site and An Alcohol Functionalized Space-spanning Group

Two stranded, duplexed, cross-linked oligonucleotides of the structure:

Cross-linked Strands VIII

GGC AGD CAG CC
|
GGC TGU CTG CG

where " | " indicates represents an O-$(CH_2)_7$-O-CH(OH)- cross-linkage between the individual oligonucleotides of the structure will be prepared by reacting Oligomer 15 of Example 10-E with the abasic Oligomer 2 of Example 3-A-iii in the presence of trifluoroacetic acid utilizing a modification of the procedure of Knorre, *et al.*, "Nucleotide and Oligonucleotide Derivatives As Enzyme and Nucleic Acid Targeted Irreversible Inhibitors", Chemical Aspects, G. Weber (Ed.), Advance Press, Oxford, 1986, pp. 277-320.

EXAMPLE 12

Preparation of Oligonucleotide Having Multiple Space-spanning Groups That Include The Same Or Different Functionalities Thereon

A. Aldehyde and Amine Functionality Containing Oligonucleotide

A oligonucleotide of the structure:

Oligomer 16 AGC CAG AUC UGA* GCC UGG GAG CU**C UCU GGC U

wherein A* represents an adenosine nucleotide having a 2'-O-(pentylamino) group thereon and U** represents a uridine nucleotide having a 2'-O-[propion-4-al bis(*o*-nitrobenzyl) acetal] group thereon will be prepared as per the procedure of Example 4-A utilizing the nucleotides of Examples 2-A-i and 2-K-i-b, respectively, for the functionalized space-spanning group containing nucleotides.

B. Multiple Thiol Functionality Containing Oligonucleotide

A oligonucleotide of the structure:

Oligomer 17 A*GC CAG AUC U GA GCC UGG GAG CUC UCU GGC U**

wherein A* represents an adenosine nucleotide having a 2'-O-[S-trityl(hexyl-8-thiol)] group thereon and U** represents a uridine nucleotide having a 2'-O-[S-trityl(hexyl-8-thiol)] group thereon will be prepared as per the procedure of Example 4-A utilizing the nucleotides of Examples 2-J-ii-b and 2-J-ii-c, respectively, for the functionalized space-spanning group containing nucleotides.

C. Multiple Aldehydic Functionality Containing Oligonucleotide

An oligonucleotide of the structure:

Oligomer 18 GAC AGA GGU* AGG AGA AU* GA

wherein U* is a 2'-O-[propion-3-al bis(*o*-nitrobenzyl) acetal] group will be prepared as per the procedure of Example 4-A utilizing the nucleotide of Example 2-K-i-b for the aldehydic functionalized space-spanning group containing nucleotides.

D. Thiol and Amine Functionality Containing Oligonucleotide

A oligonucleotide of the structure:

Oligomer 19 AGC CAG* AUC U GA GCC UGG GAG CUC U**CU GGC U

wherein G* represents an guanosine nucleotide having a 2'-O-([2-amylenyl]-N-phthalimide) group thereon and U** represents a uridine nucleotide having a 2'-O-[S-trityl(pentan-8-thiol)] group thereon will be prepared as per the procedure of Example 4-A utilizing the nucleotides of Examples 2-A-iv and 2-J-ii-e, respectively, for the functionalized space-spanning group containing nucleotides.

EXAMPLE 13

Multi-Site Cross-Linking Of Dual Strands of Oligonucleotides

A. Cross-Linking Via Dual Abasic Sites

Two stranded duplexed, multi-site cross-linked oligonucleotides of the structure:

Cross-linked Strand IX

CTG TCT CCA TCC TCT TCA CT
| |
GAC AGA GGD AGG AGA AGD GA

wherein " | " represents a cross-linkage between the strands, are prepared from Oligomer 4 from Example 3-A-iv and Oligomer 6 from Example 4-B utilizing the cross-linking procedure of Example 5-A. The resulting cross-linked strands are held together not only by hybridization of the duplex structures but also by the dual cross-linkages extending between the individual oligonucleotide strands.

B. Via Dual Space-spanning Groups

Two stranded duplexed, multi-site cross-linked oligonucleotides of the structure:

Cross-linked Strands X

CTG TCT CCA TCC TCT TCA CT
| |
GAC AGA GGU AGG AGA AGU GA

wherein " | " represents a cross-linkage having a covalent bond formed via an oxime linkage between the strands, is prepared from Oligomer 18 from Example 12-C and Oligomer 6 from Example 4-B. A solution of Oligomer 18 in water will be treated with

hydrazine hydrate/ sodium cyanoborohydride to remove the aldehydic protecting groups.

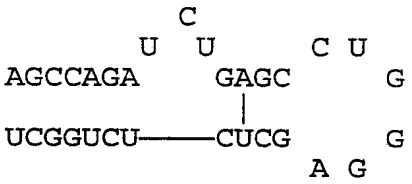
Oligomer 6 is then added in 0.1 M NaCl to hybridize the strands. Sodium acetate buffer is then added to effect cross-linking of the strands via formation of Schiff's bases between the amine and aldehyde groups. $NaCNBH_3$ is added to reduced the oxime linkage to an amine linkage whereby the resulting cross-linked strands will be held together not only by hybridization of the duplex structures but also by the multiple cross-links extending between the individual oligonucleotide strands.

EXAMPLE 14

Coupling of Two Sites on Single Strand

A single strand duplexed, cross-linked, hairpin loop oligonucleotide of the structure:

Cross-linked Strand XI



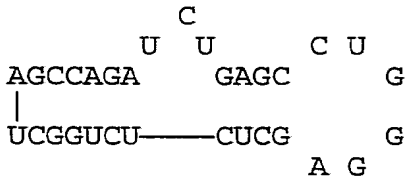
wherein " | " represents a cross-linkage having a covalent bond formed via an oxime linkage between the regions of the strand, is prepared from Oligomer 16 from Example 12-A. Oligomer 16 is taken up in 0.1 M NaCl solution to effect hybridization and then treated with hydrogen in the presence of palladium on charcoal at atmospheric pressure until two equivalents of hydrogen are taken up indicating removal of the aldehydic protecting groups. Sodium acetate buffer is then added to effect the formation of a Schiff's base linkage between the amine and aldehyde groups. $NaCNBH_3$ is added to reduced the oxime linkage to an amine linkage whereby the resulting cross-linked strand will be held together not only by hybridization of the duplex stem portion of the hairpin loop structure but also by the covalent cross-linkage extending across the stem nucleotides.

EXAMPLE 15

Coupling Via Disulfide Linkage

A single strand of duplexed, cross-linked, bulged, hairpin loop oligonucleotide of the structure:

Cross-linked Strand XII



wherein " | " represents a cross-linkage having a covalent bond formed via a disulfide linkage between the regions of the strand, is prepared from Oligomer 17 from Example 12-B.

Oligomer 17 will be solubilized in 0.1 M NaCl solution under nitrogen to effect hybridization and then treated with silver nitrate to remove the S-trityl protecting groups. A disulfide bond is formed between the two free thiol groups via mild oxidation by bubbling air through the solution to cross-linkage the A-U site of the oligonucleotide. The oligonucleotide will be held in the hairpin loop structure not only by hybridization of the duplexed stem portions of the hairpin loop structure but also by the covalent disulfide bond connecting the stem regions.

EXAMPLE 16

Coupling Via Dialdehyde

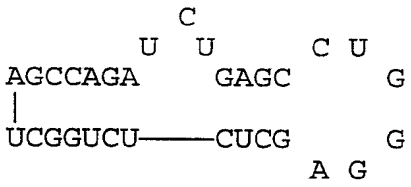
Utilizing the procedure of Example 4-A an oligonucleotide of the sequence:

Oligomer 20 A*GC CAG AUC U GA GCC UGG GAG CUC UCU GGC U

wherein A* is a 2'-O-(eicosylamino)adenosine nucleotide of Example 2-A-ii-c is prepared. The oligonucleotide is removed from the CPG synthetic column and de-salted on Sephadex G-25 Oligomer 20 will then be treated as per the peroxidation procedure of Lemairte et. al., *Proc. Natl. Acad. Sci. USA*, 84:648 (1986) to effect periodate-oxidation of the sugar moiety of the 3' terminus uridine nucleotide.

A single strand of duplexed, cross-linked, hairpin loop oligonucleotide of the structure:

Cross-linked Strand XIII



wherein " | " represents a covalent cross-linkage formed via the oxidized, ring opened sugar of the 3'-terminal nucleotide and the amine group on the opposing nucleotide of the duplexed stem portion of the strand. Such cross-linkage is formed from the periodate-oxidized oligonucleotide by solubilizing the oxidized oligonucleotide in 0.1 M NaCl to effect hybridization followed by the addition of sodium acetate buffer and $NaCNBH_3$. As the intermediate Schiff's base cross-linkage is formed, it is reduced to an amine linkage by the addition of $NaCNBH_3$ as per the procedure of Example 5-A.

EXAMPLE 17

Coupling Via Acetal Linkage

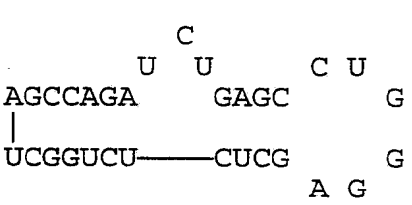
Utilizing the procedure of Example 4 an oligonucleotide of the sequence:

Oligomer 20 A*GC CAG AUC U GA GCC UGG GAG CUC UCU GGC U

wherein A* is a 2'-O-[propion-3-al bis(*o*-nitrobenzyl) acetal] functionalized adenosine nucleotide will be prepared as per the procedure of Example 4-A utilizing the nucleotide of Example 2-K-i-a to introduce the aldehydic functionalized space-spanning group containing nucleotide with the oligonucleotide sequence. The oligonucleotide is removed from the CPG synthetic column and de-salted on Sephadex G-25.

A single strand of duplexed, cross-linked, hairpin loop oligonucleotide of the structure:

Cross-linked Strand XIV



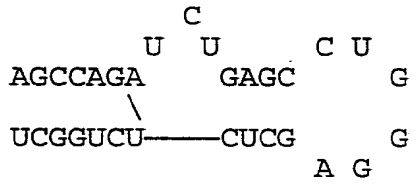
wherein " | " represents an acetal cross-linkage formed between the 2',3'-hydroxyl moieties of the sugar on the 3'-terminal U nucleotide and the aldehydic functionality on the opposing A nucleotide of the duplexed stem portion of the strand, is prepared by effecting cross-linking in 0.1 M NaCl in the presence of trifluoroacetic acid. This procedure is a modification of the procedure of Knorre et. al., "Nucleotide and Oligonucleotide Derivatives As Enzyme and Nucleic Acid Targeted Irreversible Inhibitors", Chemical Aspects, G. Weber (Ed.), Advance Press, Oxford, 1986, pp. 277-320. Upon reaction of the free aldehydic group on the A nucleotide, cross-linking to the opposing 2',3'-hydroxyl's of the 3'-terminus U nucleoside will yield a covalent cross-linkage between nucleotides of the stem portion of the hairpin loop bulged single stranded oligonucleotide.

EXAMPLE 18

Coupling Via Heterobifunctional Linker on Single Strand

A single strand of duplexed, cross-linked, hairpin loop oligonucleotide of the structure:

Cross-linked Strand XV



wherein " \ " represents a cross-linkage formed via a homobifunctional linker between the G and U nucleotides of duplexed regions of the strand across the G and U nucleotides of the two separate regions that are one base removed from Watson/Crick base pairs, is prepared from Oligomer 19 from Example 12-D. The Oligomer 19 will first be taken up in a 0.1M NaCl solution to effect hybridization of the oligonucleotide into a hairpin loop single strand duplexed structure. The reaction solution is then treated with silver nitrate in buffer to remove the S-trityl blocking group. A aliquot of N-

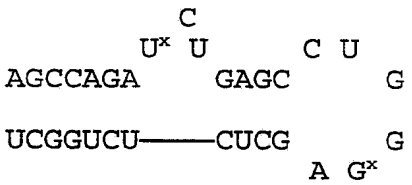
succinimidyl 3-(2-pyridyldithio)propionate (SPDP) (Pierce, Rockford, Il) in weak alkaline is added to effect coupling of the amine functionality to the linker. The pH is then adjusted to pH 4.5 to effect exchange of the 2-pyridyl-disulfide and the thiol functionality on the oligonucleotide and in doing so to effect cross-linking of the strand. The strand will thus be held together not only by hybridization of the duplex stem portion of the hairpin loop bulged structure but also by the heterobifunctional cross-linker extending across the stem nucleotides.

EXAMPLE 19

Coupling Via Homobifunctional Linker

A single strand of duplexed, cross-linked, bulged, hairpin loop oligonucleotide of the structure:

Cross-linked Strand XVI



wherein a cross-linkage is formed via a homobifunctional linker spanning between the U^x nucleotide on the UCU bulge and the G^x nucleotide of the AGGGUC loop regions of the strand, is prepared from Oligomer 11 from Example 10-A-ii. Molecular models of the proposed tertiary structure of Strand XVI indicate that the U^x and G^x nucleotides lie in close physical proximity to each other. To prepare Strand XVI the Oligomer 11 will first be taken up in a 0.1M NaCl solution to effect hybridization of the oligonucleotide into a hairpin loop bulged single strand duplexed structure. The reaction solution is then adjusted to be a 0.2 M sodium borate buffer. Disuccinimidyl suberate (DSS, Pierce, Rockford, Il) is added to effect cross-linking of the hydrazine reactive functionalities of the uridine-guanosine nucleotides. The U nucleotide of the bulge and G nucleotide of the loop are linked via hydrazide

linkages between the functionalities on the space-spanning groups and the homobifunctional cross-linker. In this structure, the U^x nucleotide of the interior loop and the G^x nucleotide of the hairpin loop are neither complementary to nor hybridizable with one another but they are located in 3-dimensional space in proximity with each other because of the higher order structure of the oligonucleotide. The cross-linkage spans this 3-dimensional space to fix the oligonucleotide in the higher order structure.

Example 20

Synthesis of Uridine-2'-hexyl Aminolinker

Preparation of 5'-dimethoxytrityl-2'-(O-hexyl-N-phthalimido) uridine phosphoramidite

2',3'-*O*-Dibutyl stannylene-uridine was synthesized according to the procedure of Wagner, *et al.*, *J. Org. Chem.*, **1974**, *39*, 24. This compound was dried over P_2O_5 under vacuum for 12 hrs. To a solution of this compound (29 g, 42.1 mmols) in 200 ml of anhydrous DMF were added (16.8 g, 50 mmols) of 6-bromohexyl phthalimide and 4.5 g of sodium iodide and the mixture was heated at 130°C for 16 hrs. under argon. The reaction mixture was evaporated, co-evaporated once with toluene and the gummy tar residue was applied on a silica column (500 g). The column was washed with 2L of EtOAc followed by eluting with 10% MeOH:90% EtOAc. The product, 2'- and 3'-isomers of O-hexyl-Ω-N-phthalimido uridine, eluted as an inseparable mixture (R_f=0.64 in 10% MeOH in EtOAc). By ^{13}C NMR, the isomeric ratio was ca. 55% of the 2' isomer and ca. 45% of the 3' isomer. The combined yield was 9.2 g (46.2%). This mixture was dried under vacuum and re-evaporated twice with pyridine. It was dissolved in 150 mL anhydrous pyridine and treated with 7.5 g of dimethyoxytrityl chloride (22.13 mmols) and 500 mg of DMAP. After 2 hrs. TLC (6:4 EtOAc:Hexane) indicated complete disappearance of the starting material and a good separation between 2' and 3' isomers (R_f=0.29 for the 2' isomer and 0.12 for the 3' isomer). The reaction mixture was quenched by the addition of 5ML of CH_3OH and evaporated under

reduced pressure. The residue was dissolved in 300mL CH_2Cl_2, washed successively with saturate NaHCO3 followed by saturated NaCl solution. It was dried ($MgSO_4$) and evaporated to give 15 g of a brown foam which was purified on a silica gel (500 g) to give 6.5 g of the 2'-isomer and 3.5 g of the 3' isomer.

The 2'-O-hexyl-Ω-N-phthalimido-5'-DMT-uridine (4 g, 5.2 mmole) was dissolved in 40 ML of anhydrous CH_2Cl_2. To this solution diisopropylaminetetrazolide (0.5 g, 2.9 mmol) and bis(isopropylamino)-ß-cyanoethoxyphosphine (2.5 mL, 7.0 mmol) were added and stirred overnight. TLC (1:1 EtoAC/hexane) showed complete disappearance of starring material. The reaction mixture was transferred with CH_2Cl_2 and washed with saturated $NaHCO_3$ (100 ML) followed by saturated NaCl solution. The organic layer was dried over anhydrous Na_2SO_4 and evaporated to yield 6.4 g of a crude product which was purified in a silica column (200 g) using 1:1 hexane EtOAc to give 4.6 g (4.7 mmol, 90%) of the desired phosphoramidite.

Example 21

Oligomer Synthesis and NMR characterization of an oligonucleotide containing U-2'-O-hexylamine linker

An oligonucleotide of the sequence:

Oligomer 22 TU*T

where U* represents the hexylamine modified uridine moiety, was synthesized on a 10 mmol synthesis in the DNA synthesizer. The phosphoramidite from Example 20 was used as a 0.12 M solution in CH_3CN. After deprotection of the oligonucleotide it was purified in a reverse phase HPLC column to get the pure trimer (any oligonucleotide containing undeprotected phthalimido group was separated from the trimer). The oligonucleotide was dried, desalted and dissolved in 600 uL D_2O (app. 50 O.D., A_{260}). Both the proton NMR and ^{31}P NMR were examined. ^{1}H NMR: 1.0-1.8 (m, CH2 protons from the hexyl chain); 5.75 (d); 5.9(d); 6.1 (t); 6.15 (t); 7.5 (s). ^{31}p NMR: -0.2; -0.4 ppm (two lines).

Example 22

Synthesis illustrating preparation of any nucleotide (A, U, C or G) placed opposite to an abasic site in an opposing strand

A. Oligonucleotide Synthesis

An oligonucleotide of the sequence:

Oligomer 23 GGC TGU* CTG CG (P=O)

wherein U* indicates a 2'-O-(hexylamino) uridine nucleotide was prepared using an extended coupling time of 10 minutes. 5'-Dimethyoxytrityl-2'-O-(hexyl-N-phthalimido) uridine phosphoramidite was used as a 0.12M solution in anhydrous CH3CN. The oligonucleotide was deprotected, purified in reverse phase HPLC (trityl-on), deprotected at the 5'-position and purified again to remove any undeprotected phthalimido oligonucleotide.

B. Cross linking reaction

The Oligomer 23 oligonucleotide was reacted with the abasic oligonucleotide CGC AGD* CAG CC (Oligomer 2) with a crosslinking efficient of >90%. The formation of the conjugate was confirmed by reverse phase HPLC, ion-exchange HPLC and finally by the observation of base paired crosslinked duplex imino protons in the NMR spectrum.

Example 23

Formation of a crosslink between a chimeric oligonucleotide and an abasic site containing oligonucleotide.

An oligonucleotide of the sequence:

Oligomer 24 $G_sG_sC_o$ $T_oG_oU^*_o$ $C_oT_oG_s$ C_sG

where s represents a phosphorothioate backbone, o represents a phosphodiester backbone and U* represents a 2'-O-(hexylamino) uridine, was synthesized with a mixed (phosphodiester/phosphorothioate) backbone using either a conventional iodine oxidation (P=O) or a Beaucage reagent oxidation (P=S) in the respective oxidation steps. The oligonucleotide was synthesized, purified and crosslinked with CGCAGD*CAGCC. The crosslinked duplex was characterized by ^{1}H NMR, Gel analysis, HPLC (reverse phase and ion-exchange).

Example 24

Cross linked nucleic acid duplex involving connectivity between two abasic sites:

Oligonucleotides with the following sequences were synthesized:

Oligomer 2 CGC AGD* CAG CC

Oligomer 26 GGC TGD* CTG CG.

Oligomer 26 was obtained from a precursor oligomer:

Oligomer 25 GGC TGU CTG CG

where U represents 2'-deoxyuridine, by the action of uracil-DNA glycosylase. Oligomer 2 is reacted with 1,6-hexanediamine in NaOAc/$NaCNBH_3$ medium and a product isolated that has a free amino group located on the linker. This modified oligonucleotide is mixed with oligomer 26 in NaOAc buffer followed by the addition of NaCNBH3 in NaOAC buffer. The resultant cross link has a connectivity between two abasic sites separated by the -NH-$(CH_2)_6$-NH-linker.

CGC AGD CGG CC
|
GGC TGD CTG CG

where "|" represents a -CH2-NH-$(CH_2)_6$-NH-CH2-crosslinker.

Example 25

T_m determination of the Cross-linked Strands I.

The melting behavior of the crosslinked duplex, Cross-linked Strand I, was determined at 100mM Nacl, 10mM sodium phosphate, 0.1 mM EDTA at pH 7.0 at 4mM each strand concentration.

	T_m
Oligomer 5 - Oligomer 1 duplex (wild type)	48°C
Oligomer 5 - Oligomer 2 duplex (abasic site)	<30°C
Oligomer 5 - Oligomer 2 Cross-linked Strands I	74°C

The higher T_m of the crosslinked duplex confirmed the covalent linkage between the two strands (Oligomer 5 and Oligomer 2).

EVALUATION

PROCEDURE 1 - Nuclease Resistance

A. Evaluation of the resistance of cross-linked oligonucleotides to serum and cytoplasmic nucleases.

Cross-linked oligonucleotides of the invention can be assessed for their resistance to serum nucleases by incubation of the cross-linked oligonucleotide in media containing various concentrations of fetal calf serum or adult human serum. Labeled cross-linked oligonucleotides are incubated for various times, treated with protease K and then analyzed by gel electrophoresis on 20% polyacrylamine-urea denaturing gels and subsequent autoradiography. Autoradiograms are quantitated by laser densitometry. Based upon the location of the modified linkage and the known length of the cross-linked oligonucleotide it is possible to determine the effect on nuclease degradation by the particular modification. For the cytoplasmic nucleases, an HL-60 cell line can be used. A post-mitochondrial supernatant is prepared by differential centrifugation and the labelled cross-linked oligonucleotides are incubated in this supernatant for various times. Following the incubation, the cross-linked oligonucleotides are assessed for degradation as outlined above for serum nucleolytic degradation. Autoradiography results are quantitated for comparison of the unmodified and the cross-linked oligonucleotides of the invention. It is expected that the cross-linked oligonucleotides will be completely resistant to serum and cytoplasmic nucleases.

B. Evaluation of the resistance of cross-linked oligonucleotides to specific endo- and exo-nucleases

Evaluation of the resistance of natural oligonucleotides and cross-linked oligonucleotides of the invention to specific nucleases (i.e., endonucleases, 3',5'-exo-, and 5',3'-exonucleases) can be done to determine the exact effect of the modified linkage on degradation. The cross-linked oligonucleotides are incubated in defined reaction buffers specific for various selected nucleases. Following treatment of the products with protease K, urea is added and analysis on

20% polyacrylamide gels containing urea is done. Gel products are visualized by staining with Stains All reagent (Sigma Chemical Co.). Laser densitometry is used to quantitate the extent of degradation. The effects of the modified linkage are determined for specific nucleases and compared with the results obtained from the serum and cytoplasmic systems. As with the serum and cytoplasmic nucleases, it is expected that the cross-linked oligonucleotides of the invention will be completely resistant to endo- and exo-nucleases.

PROCEDURE 2 - 5-Lipoxygenase Analysis and Assays

A. Therapeutics

For therapeutic use, an animal suspected of having a disease characterized by excessive or abnormal supply of 5-lipoxygenase is treated by administering cross-linked oligonucleotides in accordance with this invention. Persons of ordinary skill can easily determine optimum dosages, dosing methodologies and repetition rates. Such treatment is generally continued until either a cure is effected or a diminution in the diseased state is achieved. Long term treatment is likely for some diseases.

B. Research Reagents

The cross-linked oligonucleotides of this invention will also be useful as research reagents when used to cleave or otherwise modulate 5-lipoxygenase mRNA in crude cell lysates or in partially purified or wholly purified RNA preparations. This application of the invention is accomplished, for example, by lysing cells by standard methods, optimally extracting the RNA and then treating it with a composition at concentrations ranging, for instance, from about 100 to about 500 ng per 10 μg of total RNA in a buffer consisting, for example, of 50 mm phosphate, pH ranging from about 4-10 at a temperature from about 30° to about 50° C. The cleaved 5-lipoxygenase RNA can be analyzed by agarose gel electrophoresis and hybridization with radiolabeled DNA probes or by other standard methods.

C. Diagnostics

The cross-linked oligonucleotides of this invention will also be useful in diagnostic applications, particularly for the determination of the expression of specific mRNA species in various tissues or the expression of abnormal or mutant RNA species. In this example, the cross-linked oligonucleotides target a hypothetical abnormal mRNA by being designed complementary to the abnormal sequence, but would not hybridize to normal mRNA.

Tissue samples can be homogenized, and RNA extracted by standard methods. The crude homogenate or extract can be treated for example to effect cleavage of the target RNA. The product can then be hybridized to a solid support which contains a bound oligonucleotide complementary to a region on the 5′ side of the cleavage site. Both the normal and abnormal 5′ region of the mRNA would bind to the solid support. The 3′ region of the abnormal RNA, which is cleaved, would not be bound to the support and therefore would be separated from the normal mRNA.

Targeted mRNA species for modulation relates to 5-lipoxygenase; however, persons of ordinary skill in the art will appreciate that the present invention is not so limited and it is generally applicable. The inhibition or modulation of production of the enzyme 5-lipoxygenase is expected to have significant therapeutic benefits in the treatment of disease. In order to assess the effectiveness of the compositions, an assay or series of assays is required.

D. In Vitro Assays

The cellular assays for 5-lipoxygenase preferably use the human promyelocytic leukemia cell line HL-60. These cells can be induced to differentiate into either a monocyte-like cell or neutrophil-like cell by various known agents. Treatment of the cells with 1.3% dimethyl sulfoxide, DMSO, is known to promote differentiation of the cells into neutrophils. It has now been found that basal HL-60 cells do not synthesize detectable levels of 5-lipoxygenase protein or secrete leukotrienes (downstream products of 5-lipoxygenase).

Differentiation of the cells with DMSO causes an appearance of 5-lipoxygenase protein and leukotriene biosynthesis 48 hours after addition of DMSO. Thus induction of 5-lipoxygenase protein synthesis can be utilized as a test system for analysis of cross-linked oligonucleotides which interfere with 5-lipoxygenase synthesis in these cells.

A second test system for cross-linked oligonucleotides makes use of the fact that 5-lipoxygenase is a "suicide" enzyme in that it inactivates itself upon reacting with substrate. Treatment of differentiated HL-60 or other cells expressing 5-lipoxygenase, with 10 μM A23187, a calcium ionophore, promotes translocation of 5-lipoxygenase from the cytosol to the membrane with subsequent activation of the enzyme. Following activation and several rounds of catalysis, the enzyme becomes catalytically inactive. Thus, treatment of the cells with calcium ionophore inactivates endogenous 5-lipoxygenase. It takes the cells approximately 24 hours to recover from A23187 treatment as measured by their ability to synthesize leukotriene B_4. Cross-linked oligonucleotides directed against 5-lipoxygenase can be tested for activity in two HL-60 model systems using the following quantitative assays. The assays are described from the most direct measurement of inhibition of 5-lipoxygenase protein synthesis in intact cells to more downstream events such as measurement of 5-lipoxygenase activity in intact cells.

The most direct effect which cross-linked oligonucleotides can exert on intact cells and which can be easily quantitated is specific inhibition of 5-lipoxygenase protein synthesis. To perform this technique, cells can be labelled with ^{35}S-methionine (50 μCi/mL) for 2 hours at 37° C to label newly synthesized protein. Cells are extracted to solubilize total cellular proteins and 5-lipoxygenase is immunoprecipitated with 5-lipoxygenase antibody followed by elution from protein A-Sepharose beads. The immunoprecipitated proteins are resolved by SDS-polyacrylamide gel electrophoresis and exposed for autoradiography. The amount of

immunoprecipitated 5-lipoxygenase is quantitated by scanning densitometry.

A predicted result from these experiments would be as follows. The amount of 5-lipoxygenase protein immunoprecipitated from control cells would be normalized to 100%. Treatment of the cells with 1 μM, 10 μM, and 30 μM of effective cross-linked oligonucleotides for 48 hours would reduce immunoprecipitated 5-lipoxygenase by 5%, 25% and 75% of control, respectively.

Measurement of 5-lipoxygenase enzyme activity in cellular homogenates could also be used to quantitate the amount of enzyme present which is capable of synthesizing leukotrienes. A radiometric assay has now been developed for quantitating 5-lipoxygenase enzyme activity in cell homogenates using reverse phase HPLC. Cells are broken by sonication in a buffer containing protease inhibitors and EDTA. The cell homogenate is centrifuged at 10,000 x g for 30 min and the supernatants analyzed for 5-lipoxygenase activity. Cytosolic proteins are incubated with 10 μM ^{14}C-arachidonic acid, 2mM ATP, 50 μM free calcium, 100 μg/ml phosphatidylcholine, and 50 mM bis-Tris buffer, pH 7.0, for 5 min at 37° C. The reactions are quenched by the addition of an equal volume of acetone, and the fatty acids extracted with ethyl acetate. The substrate and reaction products are separated by reverse phase HPLC on a Novapak C18 column (Waters Inc., Millford, MA). Radioactive peaks are detected by a Beckman model 171 radiochromatography detector. The amount of arachidonic acid converted into di-HETE's and mono-HETE's is used as a measure of 5-lipoxygenase activity.

A predicted result for treatment of DMSO differentiated HL-60 cells for 72 hours with effective cross-linked oligonucleotides at 1 μM, 10 μM, and 30 μM would be as follows. Control cells oxidize 200 pmol arachidonic acid/ 5 min/ 10^6 cells. Cells treated with 1 μM, 10 μM, and 30 μM of an effective cross-linked oligonucleotides would oxidize 195 pmol, 140 pmol, and 60 pmol of arachidonic acid/ 5 min/ 10^6 cells respectively.

A quantitative competitive enzyme linked immunosorbant assay (ELISA) for the measurement of total 5-lipoxygenase protein in cells has been developed. Human 5-lipoxygenase expressed in *E. coli* and purified by extraction, Q-Sepharose, hydroxyapatite, and reverse phase HPLC is used as a standard and as the primary antigen to coat microtiter plates. 25 ng of purified 5-lipoxygenase is bound to the microtiter plates overnight at 4° C. The wells are blocked for 90 min with 5% goat serum diluted in 20 mM Tris-HCL buffer, pH 7.4, in the presence of 150 mM NaCl (TBS). Cell extracts (0.2% Triton X-100, 12,000 x g for 30 min.) or purified 5-lipoxygenase were incubated with a 1:4000 dilution of 5-lipoxygenase polyclonal antibody in a total volume of 100 μL in the microtiter wells for 90 min. The antibodies are prepared by immunizing rabbits with purified human recombinant 5-lipoxygenase. The wells are washed with TBS containing 0.05% Tween 20 (TBST), then incubated with 100 μL of a 1:1000 dilution of peroxidase conjugated goat anti-rabbit IgG (Cappel Laboratories, Malvern, PA) for 60 min at 25° C. The wells are washed with TBST and the amount of peroxidase labelled second antibody determined by development with tetramethylbenzidine.

Predicted results from such an assay using a 30-mer cross-linked oligonucleotide at 1 μM, 10 μM, and 30 μM would be 30 ng, 18 ng and 5 ng of 5-lipoxygenase per 10^6 cells, respectively, with untreated cells containing about 34 ng 5-lipoxygenase.

A net effect of inhibition of 5-lipoxygenase biosynthesis is a diminution in the quantities of leukotrienes released from stimulated cells. DMSO-differentiated HL-60 cells release leukotriene B4 (LTB4) upon stimulation with the calcium ionophore A23187. Leukotriene B4 released into the cell medium can be quantitated by radioimmunoassay using commercially available diagnostic kits (New England Nuclear, Boston, MA). Leukotriene B4 production can be detected in HL-60 cells 48 hours following addition of DMSO to differentiate the cells into neutrophil-like cells. Cells (2 x 10^5 cells/mL) will be treated with increasing concentrations of cross-linked

oligonucleotides for 48-72 hours in the presence of 1.3 % DMSO. The cells are washed and resuspended at a concentration of 2 x 10^6 cell/mL in Dulbecco's phosphate buffered saline containing 1% delipidated bovine serum albumin. Cells are stimulated with 10 μM calcium ionophore A23187 for 15 min and the quantity of LTB4 produced from 5 x 10^5 cells determined by radioimmunoassay as described by the manufacturer.

Using this assay the following results would likely be obtained with a cross-linked oligonucleotide directed to the 5-lipoxygenase mRNA. Cells will be treated for 72 hours with either 1 μM, 10 μM or 30 μM cross-linked oligonucleotide in the presence of 1.3% DMSO. The quantity of LTB_4 produced from 5 x 10^5 cells would be expected to be about 75 pg, 50 pg, and 35 pg, respectively with untreated differentiated cells producing 75 pg LTB_4.

E. In Vivo Assay

Inhibition of the production of 5-lipoxygenase in the mouse can be demonstrated in accordance with the following protocol. Topical application of arachidonic acid results in the rapid production of leukotriene B_4, leukotriene C_4 and prostaglandin E_2 in the skin followed by edema and cellular infiltration. Certain inhibitors of 5-lipoxygenase have been known to exhibit activity in this assay. For the assay, 2 mg of arachidonic acid is applied to a mouse ear with the contralateral ear serving as a control. The polymorphonuclear cell infiltrate is assayed by myeloperoxidase activity in homogenates taken from a biopsy 1 hour following the administration of arachidonic acid. The edematous response is quantitated by measurement of ear thickness and wet weight of a punch biopsy. Measurement of leukotriene B_4 produced in biopsy specimens is performed as a direct measurement of 5-lipoxygenase activity in the tissue. Cross-linked oligonucleotides will be applied topically to both ears 12 to 24 hours prior to administration of arachidonic acid to allow optimal activity of the compounds. Both ears are pretreated for 24 hours with either 0.1 μmol, 0.3 μmol, or 1.0 μmol of the oligonucleotide prior to challenge with arachidonic acid.

Values are expressed as the mean for three animals per concentration. Inhibition of polymorphonuclear cell infiltration for 0.1 μmol, 0.3 μmol, and 1 μmol is expected to be about 10 %, 75 % and 92 % of control activity, respectively. Inhibition of edema is expected to be about 3%, 58% and 90%, respectively while inhibition of leukotriene B_4 production would be expected to be about 15%, 79% and 99%, respectively.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: David Ecker

(ii) TITLE OF INVENTION: Covalently Cross-Linked Oligonucleotides

(iii) NUMBER OF SEQUENCES: 24

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & Norris

(B) STREET: One Liberty Place - 46th Floor

(C) CITY: Philadelphia

(D) STATE: PA

(E) COUNTRY: USA

(F) ZIP: 19103

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 Mb STORAGE

(B) COMPUTER: IBM PS/2

(C) OPERATING SYSTEM: PC-DOS

(D) SOFTWARE: WORDPERFECT 5.1

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER: n/a

(B) FILING DATE: Herewith

(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: 846,376

(B) FILING DATE: March 5, 1992

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: Joseph Lucci

(B) REGISTRATION NUMBER: 33,307

(C) REFERENCE/DOCKET NUMBER: ISIS-0980

(ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: (215) 568-3100

(B) TELEFAX: (215) 568-3439

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 11

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(iv) ANTI-SENSE: no

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

CGCAGUCAGC C 11

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 11

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(iv) ANTI-SENSE: no

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

CGCAGDCAGC C 11

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(iv) ANTI-SENSE: no

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

GACAGAGGUA GGAGAAGUGA 20

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(iv) ANTI-SENSE: no

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

GACAGAGGDA GGAGAAGDGA 20

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 11

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(iv) ANTI-SENSE: no

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

GGCTGACTGC G 11

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(iv) ANTI-SENSE: yes

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

CTGTCTCCAT CCTCTTCACT 20

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 11

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(iv) ANTI-SENSE: no

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

GGCTGACTGC G 11

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 11

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(iv) ANTI-SENSE: no

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

GGCTGACTGC G 11

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 11

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(iv) ANTI-SENSE: no

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

GGCTGACTGC G 11

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 11

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(iv) ANTI-SENSE: no

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

GGCTGUCTGC G 11

(2) INFORMATION FOR SEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 31

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(iv) ANTI-SENSE: no

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

AGCCAGAUCU GAGCCUGGGA GCUCUCUGGC U 31

(2) INFORMATION FOR SEQ ID NO: 12:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 11

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(iv) ANTI-SENSE: no

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

GGCTGUCTGC G 11

(2) INFORMATION FOR SEQ ID NO: 13:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 11

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(iv) ANTI-SENSE: no

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

GGCTGUCTGC G 11

(2) INFORMATION FOR SEQ ID NO: 14:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 11

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(iv) ANTI-SENSE: no

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

GGCTGUCTGC G 11

(2) INFORMATION FOR SEQ ID NO: 15:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 11

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(iv) ANTI-SENSE: no

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

GGCTGUCTGC G 11

(2) INFORMATION FOR SEQ ID NO: 16:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 31

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(iv) ANTI-SENSE: no

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

AGCCAGAUCU GAGCCUGGGA GCUCUCUGGC U 31

(2) INFORMATION FOR SEQ ID NO: 17:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 31

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(iv) ANTI-SENSE: no

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

AGCCAGAUCU GAGCCUGGGA GCUCUCUGGC U 31

(2) INFORMATION FOR SEQ ID NO: 18:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 19

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(iv) ANTI-SENSE: no

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

GACAGAGGUA GGAGAAUGA 19

(2) INFORMATION FOR SEQ ID NO: 19:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 31

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(iv) ANTI-SENSE: no

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

AGCCAGAUCU GAGCCUGGGA GCUCUCUGGC U 31

(2) INFORMATION FOR SEQ ID NO: 20:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 31

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(iv) ANTI-SENSE: no

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

AGCCAGAUCU GAGCCUGGGA GCUCUCUGGC U 31

(2) INFORMATION FOR SEQ ID NO: 21:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 11

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(iv) ANTI-SENSE: no

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

GGCTGUCTGC G 11

(2) INFORMATION FOR SEQ ID NO: 22:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 11

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(iv) ANTI-SENSE: no

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

GGCTGUCTGC G 11

(2) INFORMATION FOR SEQ ID NO: 23:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 11

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(iv) ANTI-SENSE: no

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23:

GGCTGDCTGC G 11

(2) INFORMATION FOR SEQ ID NO: 24:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 11

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(iv) ANTI-SENSE: no

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 24:

GGCTGUCTGC G 11

WHAT IS CLAIMED IS:

1. A cross-linked nucleic acid comprising:

a first nucleotide located on a first oligonucleotide strand;

a first bond means located on a sugar moiety of said first nucleotide;

a second nucleotide located on a further oligonucleotide strand;

a second bond means located on a sugar moiety of said second nucleotide; and

a covalent cross-linkage between said first and said second bond means.

2. A cross-linked nucleic acid comprising:

a first nucleotide located on an oligonucleotide strand;

a first bond means located on a sugar moiety of said first nucleotide;

a second nucleotide located on said oligonucleotide strand;

a second bond means located on a sugar moiety of said second nucleotide; and

a covalent cross-linkage between said first and said second bond means.

3. A cross-linked nucleic acid comprising:

a first nucleotide located on a 3′ end of a first oligonucleotide strand, said first nucleotide including a 2′,3′-dialdehyde structure;

a second nucleotide located on a further oligonucleotide strand or on said first oligonucleotide strand;

a space-spanning group located on a sugar moiety of said second nucleotide;

an active functional group located on said space-spanning group; and

a covalent cross-linkage between said dialdehyde structure and said active functional group.

4. A cross-linked nucleic acid comprising:

a first nucleotide located on a 3′ end of a first oligonucleotide strand, said first nucleotide including 2′ and 3′ hydroxyl groups;

a second nucleotide located on a further oligonucleotide strand or on said first oligonucleotide strand;

a space-spanning group located on a sugar moiety of said second specific nucleotide;

an active functional group located on said space-spanning group; and

a covalent acetal cross-linkage between said 2′ and 3′ hydroxyl groups and said active functional group.

5. A cross-linked nucleic acid comprising:

a first nucleotide located on a first oligonucleotide strand;

a first bond means located on a sugar moiety of said first nucleotide;

a second nucleotide located on a further oligonucleotide strand or on said first oligonucleotide strand;

a second bond means located on said second nucleotide; and

a covalent cross-linkage between said first and said second bond means.

6. The cross-linked nucleic acid of claim 5 wherein:

said first nucleotide is located in a first nucleotide sequence and said second nucleotide is located in a second nucleotide sequence;

said second nucleotide sequence is complementary to and hybridizable with said first nucleotide sequence; and

said first and said second nucleotide sequences are located on a single strand of said oligonucleotide with said second nucleotide sequence located on said oligonucleotide at a distance separated from said first nucleotide sequence sufficient to allow said oligonucleotide to assume a conformation wherein said first and second nucleotide sequences are mutually aligned with and specifically hybridized with one another.

7. The cross-linked nucleic acid of claim 5 wherein:

said first nucleotide is located in a first nucleotide sequence and said second nucleotide is located in a second nucleotide sequence;

said second nucleotide sequence is complementary to and specifically hybridizable with said first nucleotide sequence; and

said first and said second nucleotide sequences are located on different oligonucleotide strands.

8. The cross-linked nucleic acid of claim 5 wherein:

said first bond means includes an abasic site;

said second bond means includes a space-spanning group having an active functional group, and

said active functional group is covalently bonded with said abasic site.

9. The cross-linked nucleic acid of claim 8 wherein:

said abasic site includes an aldehyde; and

said active functional group is capable of reacting with said aldehyde.

10. The cross-linked nucleic acid of claim 8 wherein said active functional group is an amine, hydrazine,

hydroxylamine, semicarbazide, thiosemicarbazide, hydrazone, hydrazide, alcohol or thiol.

11. The cross-linked nucleic acid of claim 10 wherein said active functional group is an amine.

12. The cross-linked nucleic acid of claim 8 wherein said space-spanning group is attached to a 2'-position of said second nucleotide.

13. The cross-linked nucleic acid of claim 8 wherein said space-spanning group is C_1-C_{20} straight chain alkyl, C_1-C_{20} straight chain O-alkyl, C_1-C_{20} straight chain S-alkyl, C_1-C_{20} straight chain NH-alkyl, C_1-C_{20} straight chain substituted alkyl, C_1-C_{20} straight chain substituted O-alkyl, C_1-C_{20} straight chain substituted S-alkyl, C_1-C_{20} straight chain substituted NH-alkyl, C_2-C_{50} branched chain alkyl, C_2-C_{50} branched chain O-alkyl, C_2-C_{50} branched chain S-alkyl, C_2-C_{50} branched chain NH-alkyl, C_2-C_{50} branched chain substituted alkyl, C_2-C_{50} branched chain substituted O-alkyl, C_2-C_{50} branched chain substituted S-alkyl, C_2-C_{50} branched chain substituted NH-alkyl, C_2-C_{20} straight chain alkenyl, C_2-C_{20} straight chain O-alkenyl, C_2-C_{20} straight chain S-alkenyl, C_2-C_{20} straight chain NH-alkenyl, C_2-C_{20} straight chain substituted alkenyl, C_2-C_{20} straight chain substituted O-alkenyl, C_2-C_{20} straight chain substituted S-alkenyl, C_2-C_{20} straight chain substituted NH-alkenyl, C_3-C_{50} branched chain alkenyl, C_3-C_{50} branched chain O-alkenyl, C_3-C_{50} branched chain S-alkenyl, C_3-C_{50} branched chain NH-alkenyl, C_3-C_{50} branched chain substituted alkenyl, C_3-C_{50} branched chain substituted O-alkenyl, C_3-C_{50} branched chain substituted S-alkenyl, C_3-C_{50} branched chain substituted NH-alkenyl, C_2-C_{20} straight chain alkyne, C_2-C_{20} straight chain O-alkyne, C_2-C_{20} straight chain S-alkyne, C_2-C_{20} straight chain NH-alkyne, C_2-C_{20} straight chain substituted alkyne, C_2-C_{20} straight chain substituted O-alkyne, C_2-C_{20} straight chain substituted S-alkyne, C_2-C_{20} straight chain substituted NH-alkyne, C_3-C_{50} branched chain alkyne, C_2-C_{50}

branched chain O-alkyne, C_3-C_{50} branched chain S-alkyne, C_3-C_{50} branched chain NH-alkyne, C_3-C_{50} branched chain substituted alkyne, C_3-C_{50} branched chain substituted O-alkyne, C_3-C_{50} branched chain substituted S-alkyne, C_3-C_{50} branched chain substituted NH-alkyne, polyamine, polyamide, polyester, polyethylene glycol, aryl, aralkyl or heterocyclic.

14. The cross-linked nucleic acid of claim 13 wherein said space-spanning group is from about 8 to about 13 atoms in length.

15. The cross-linked nucleic acid of claim 5 wherein:

said first bond means includes a first space-spanning group having a first active functional group;

said second bond means includes a second space-spanning group having a second active functional group; and

said covalent cross-linkage is between said first and said second active functional groups.

16. The cross-linked nucleic acid of claim 15 wherein said first active functional group is an aldehyde, a protected aldehyde or an aldehyde precursor.

17. The cross-linked nucleic acid of claim 16 wherein said first active functional group is C-formyl, *o*-methylaminobenzenethiol, aryl substituted imidazolidino, dithiane, bis(*p*-nitrophenoxy), acetal, nitrile, ester or alcohol.

18. The cross-linked nucleic acid of claim 15 wherein said second active functional group is an amine, hydrazine, hydroxylamine, semicarbazide, thiosemicarbazide, hydrazide, alcohol or thiol.

19. The cross-linked nucleic acid of claim 15 wherein:

said first space-spanning group is attached to a 2'-position of said first nucleotide; and

said second space-spanning group is attached to a 2'-position of said second nucleotide.

20. The cross-linked nucleic acid of claim 15 wherein each of said first and second space-spanning groups are, independently, C_1-C_{20} straight chain alkyl, C_1-C_{20} straight chain O-alkyl, C_1-C_{20} straight chain S-alkyl, C_1-C_{20} straight chain NH-alkyl, C_1-C_{20} straight chain substituted alkyl, C_1-C_{20} straight chain substituted O-alkyl, C_1-C_{20} straight chain substituted S-alkyl, C_1-C_{20} straight chain substituted NH-alkyl, C_2-C_{50} branched chain alkyl, C_2-C_{50} branched chain O-alkyl, C_2-C_{50} branched chain S-alkyl, C_2-C_{50} branched chain NH-alkyl, C_2-C_{50} branched chain substituted alkyl, C_2-C_{50} branched chain substituted O-alkyl, C_2-C_{50} branched chain substituted S-alkyl, C_2-C_{50} branched chain substituted NH-alkyl, C_2-C_{20} straight chain alkenyl, C_2-C_{20} straight chain O-alkenyl, C_2-C_{20} straight chain S-alkenyl, C_2-C_{20} straight chain NH-alkenyl, C_2-C_{20} straight chain substituted alkenyl, C_2-C_{20} straight chain substituted O-alkenyl, C_2-C_{20} straight chain substituted S-alkenyl, C_2-C_{20} straight chain substituted NH-alkenyl, C_3-C_{50} branched chain alkenyl, C_3-C_{50} branched chain O-alkenyl, C_3-C_{50} branched chain S-alkenyl, C_3-C_{50} branched chain NH-alkenyl, C_3-C_{50} branched chain substituted alkenyl, C_3-C_{50} branched chain substituted O-alkenyl, C_3-C_{50} branched chain substituted S-alkenyl, C_3-C_{50} branched chain substituted NH-alkenyl, C_2-C_{20} straight chain alkyne, C_2-C_{20} straight chain O-alkyne, C_2-C_{20} straight chain S-alkyne, C_2-C_{20} straight chain NH-alkyne, C_2-C_{20} straight chain substituted alkyne, C_2-C_{20} straight chain substituted O-alkyne, C_2-C_{20} straight chain substituted S-alkyne, C_2-C_{20} straight chain substituted NH-alkyne, C_3-C_{50} branched chain alkyne, C_2-C_{50} branched chain O-alkyne, C_3-C_{50} branched chain S-alkyne, C_3-C_{50} branched chain NH-alkyne, C_3-C_{50} branched chain substituted alkyne, C_3-C_{50} branched chain sub-

stituted O-alkyne, C_3-C_{50} branched chain substituted S-alkyne, C_3-C_{50} branched chain substituted NH-alkyne, polyamine, polyamide, polyester, polyethylene glycol, aryl, aralkyl or heterocyclic.

21. The cross-linked nucleic acid of claim 15 wherein said first and second active functional groups are thiol groups and said covalent cross-linkage is a disulfide bond.

22. The cross-linked nucleic acid of claim 5 wherein:

said first nucleotide is located in a first nucleotide sequence and said second nucleotide is located in a second nucleotide sequence; and

said first and said second nucleotide sequences are located on a single strand of an oligonucleotide with said second nucleotide sequence located on said oligonucleotide at a non-hybridizable position with respect to said first nucleotide sequence and at a distance from said first nucleotide sequence sufficient to allow said oligonucleotide to assume a conformation wherein said first and second nucleotide sequences are located in spatial proximity with each other.

23. The cross-linked nucleic acid of claim 5 wherein:

said first bond means includes a 3'-terminal nucleoside having a 2',3'-dialdehyde structure;

said second bond means includes an active functional group located on a space-spanning group; and

said active functional group is capable of reacting with a dialdehyde.

24. The cross-linked nucleic acid of claim 5 wherein:

said first bond means includes a 3'-terminal nucleotide having 2' and 3'-sugar hydroxyl groups;

said second bond means includes an active functional group located on a space-spanning group; and

an acetal cross-linkage connects said 2' and 3' hydroxyl groups and said active functional group.

25. A cross-linked oligonucleotide comprising:

a first sequence region in an oligonucleotide;

a second sequence region in an oligonucleotide;

an abasic site located in said first sequence region;

a space-spanning group located on said second sequence region; and

an active functional group located on said space-spanning group, wherein said active functional group is covalently bonded with said abasic site.

26. The cross-linked oligonucleotide of claim 25 wherein said first sequence region and said second sequence region are on a single oligonucleotide strand.

27. The cross-linked oligonucleotide of claim 25 wherein said first sequence region and said second sequence region are on different oligonucleotide strands.

28. The cross-linked oligonucleotide of claim 25 wherein said first sequence region and said second sequence region are complementary to and specifically hybridizable with one another.

29. The cross-linked oligonucleotide of claim 25 wherein said space-spanning group is C_1-C_{20} straight chain alkyl, C_1-C_{20} straight chain O-alkyl, C_1-C_{20} straight chain S-alkyl, C_1-C_{20} straight chain NH-alkyl, C_1-C_{20} straight chain

substituted alkyl, C_1-C_{20} straight chain substituted O-alkyl, C_1-C_{20} straight chain substituted S-alkyl, C_1-C_{20} straight chain substituted NH-alkyl, C_2-C_{50} branched chain alkyl, C_2-C_{50} branched chain O-alkyl, C_2-C_{50} branched chain S-alkyl, C_2-C_{50} branched chain NH-alkyl, C_2-C_{50} branched chain substituted alkyl, C_2-C_{50} branched chain substituted O-alkyl, C_2-C_{50} branched chain substituted S-alkyl, C_2-C_{50} branched chain substituted NH-alkyl, C_2-C_{20} straight chain alkenyl, C_2-C_{20} straight chain O-alkenyl, C_2-C_{20} straight chain S-alkenyl, C_2-C_{20} straight chain NH-alkenyl, C_2-C_{20} straight chain substituted alkenyl, C_2-C_{20} straight chain substituted O-alkenyl, C_2-C_{20} straight chain substituted S-alkenyl, C_2-C_{20} straight chain substituted NH-alkenyl, C_3-C_{50} branched chain alkenyl, C_3-C_{50} branched chain O-alkenyl, C_3-C_{50} branched chain S-alkenyl, C_3-C_{50} branched chain NH-alkenyl, C_3-C_{50} branched chain substituted alkenyl, C_3-C_{50} branched chain substituted O-alkenyl, C_3-C_{50} branched chain substituted S-alkenyl, C_3-C_{50} branched chain substituted NH-alkenyl, C_2-C_{20} straight chain alkyne, C_2-C_{20} straight chain O-alkyne, C_2-C_{20} straight chain S-alkyne, C_2-C_{20} straight chain NH-alkyne, C_2-C_{20} straight chain substituted alkyne, C_2-C_{20} straight chain substituted O-alkyne, C_2-C_{20} straight chain substituted S-alkyne, C_2-C_{20} straight chain substituted NH-alkyne, C_3-C_{50} branched chain alkyne, C_2-C_{50} branched chain O-alkyne, C_3-C_{50} branched chain S-alkyne, C_3-C_{50} branched chain NH-alkyne, C_3-C_{50} branched chain substituted alkyne, C_3-C_{50} branched chain substituted O-alkyne, C_3-C_{50} branched chain substituted S-alkyne, C_3-C_{50} branched chain substituted NH-alkyne, polyamine, polyamide, polyester, polyethylene glycol, aryl, aralkyl or heterocyclic.

30. The cross-linked oligonucleotide of claim 25 wherein said active functional group is an amine, hydrazine, hydroxylamine, semicarbazide, hydrazide, alcohol or thiol.

31. The cross-linked oligonucleotide of claim 25 wherein said space-spanning group is CH_2-L_3, O-L_3, S-L_3, NH-L_3 or CO-L_3 where L_3 is C_1-C_{20} straight chain alkyl, C_1-C_{20} straight

chain substituted alkyl, C_2-C_{50} branched chain alkyl, C_2-C_{50} branched chain substituted alkyl, C_2-C_{20} straight chain alkenyl, C_2-C_{20} straight chain substituted alkenyl, C_3-C_{50} branched chain alkenyl, C_3-C_{50} branched chain substituted alkenyl, C_2-C_{20} straight chain alkyne, C_2-C_{20} straight chain substituted alkyne, C_3-C_{50} branched chain alkyne, C_3-C_{50} branched chain substituted alkyne, polyamine, polyamide, polyester, polyethylene glycol, aryl, aralkyl or heterocyclic.

32. The cross-linked oligonucleotide of claim 25 wherein said space-spanning group is located on a 2'-position of a nucleotide within said second sequence region.

33. A cross-linked oligonucleotide comprising:
a first sequence region in an oligonucleotide;
a second sequence region in an oligonucleotide; and
a covalently bonded linker of the structure:

$$X_1-L_1-Y-L_2-X_2$$

wherein X_1 is covalently connected to said first sequence region, X_2 is covalently connected to said second sequence region, and where:

(i) X_1 and X_2, independently, are O, S, NH, CH_2 or CO;

L_1 and L_2, independently, are C_1-C_{20} straight chain alkyl, C_1-C_{20} straight chain substituted alkyl, C_2-C_{50} branched chain alkyl, C_2-C_{50} branched chain substituted alkyl, C_2-C_{20} straight chain alkenyl, C_2-C_{20} straight chain substituted alkenyl, C_3-C_{50} branched chain alkenyl, C_3-C_{50} branched chain substituted alkenyl, C_2-C_{20} straight chain alkyne, C_2-C_{20} straight chain substituted alkyne, C_3-C_{50} branched chain alkyne, C_3-C_{50} branched chain substituted alkyne, polyamine, polyamide,

polyester, polyethylene glycol, polyether, aryl, aralkyl or heterocyclic; and

Y is an imine, amine, oxime, hydroxylamine, hydrazine, hydrazone, azine, hydrazide-hydrazone, amide, hydrazide, semicarbazide, semicarbazone, thiosemicarbazide, thiocarbazone, disulfide, hemiacetal, thiohemiacetal, α-keto-alkylthioalkyl or α-keto-alkylaminoalkyl; or

(ii) X_1 and X_2, independently, are O or NH; and

L_1-Y-L_2 together are NH-L_3-NH or NH-NH-L_3-NH-NH; and

L_3 is C_1-C_{20} straight chain alkyl, C_1-C_{20} straight chain substituted alkyl, C_2-C_{50} branched chain alkyl, C_2-C_{50} branched chain substituted alkyl, C_2-C_{20} straight chain alkenyl, C_2-C_{20} straight chain substituted alkenyl, C_3-C_{50} branched chain alkenyl, C_3-C_{50} branched chain substituted alkenyl, C_2-C_{20} straight chain alkyne, C_2-C_{20} straight chain substituted alkyne, C_3-C_{50} branched chain alkyne, C_3-C_{50} branched chain substituted alkyne, polyamine, polyamide, polyester, polyethylene glycol, polyether, aryl, aralkyl or heterocyclic; or

(iii) X_1 and X_2, independently, are O, S, NH, CH_2 or CO;

L_1 and L_2, independently, are C_1-C_{20} straight chain alkylamino, C_1-C_{20} straight chain substituted alkylamino, C_2-C_{50} branched chain alkylamino, C_2-C_{50} branched chain substituted alkylamino, C_2-C_{20} straight chain alkenylamino, C_2-C_{20} straight chain substituted alkenylamino, C_3-C_{50} branched chain alkenylamino, C_3-C_{50} branched chain substituted alkenylamino, C_1-C_{20} straight chain alkynylamino, C_2-C_{20} straight chain substituted alkynylamino, C_3-C_{50} branched chain alkynylamino, C_3-C_{50} branched chain substituted alkynylamino, C_1-C_{20} straight chain alkylthio, C_1-C_{20} straight chain substituted alkylthio, C_2-C_{50} branched chain alkylthio, C_2-C_{50},

branched chain substituted alkylthio, C_2-C_{20} straight chain alkenylthio, C_2-C_{20} straight chain substituted alkenylthio, C_3-C_{50} branched chain alkenylthio, C_3-C_{50} branched chain substituted alkenylthio, C_2-C_{20} straight chain alkynylthio, C_2-C_{20} straight chain substituted alkynylthio, C_3-C_{50} branched chain alkynylthio or C_3-C_{50} branched chain substituted alkynylthio; and

Y is a heterobifunctional or homobifunctional cross-linking reagent;

where said substituent of said substituted alkyl, substituted alkenyl and substituted alkyne moieties is an intercalator, a conjugate, polyamine, polyamide, polyethylene glycol, a group that enhances the pharmacodynamic properties of oligonucleotides or a group that enhances the pharmacokinetic properties of oligonucleotides.

34. The cross-linked oligonucleotide of claim 33 wherein:

X_1 and X_2, independently, are O, S, NH, CH_2 or CO;

L_1 and L_2, independently, are C_1-C_{20} straight chain alkyl, C_1-C_{20} straight chain substituted alkyl, C_2-C_{50} branched chain alkyl, C_2-C_{50}, branched chain substituted alkyl, C_2-C_{20} straight chain alkenyl, C_2-C_{20} straight chain substituted alkenyl, C_3-C_{50} branched chain alkenyl, C_3-C_{50} branched chain substituted alkenyl, C_2-C_{20} straight chain alkyne, C_2-C_{20} straight chain substituted alkyne, C_3-C_{50} branched chain alkyne, C_3-C_{50} branched chain substituted alkyne, polyamine, polyamide, polyester, polyethylene glycol, aryl, aralkyl or heterocyclic; and

Y is imine, amine, oxime, hydroxylamine, hydrazine, hydrazone, azine, hydrazide-hydrazone, amide, hydrazide, semicarbazide, semicarbazone, thiosemicarbazide, thiocarbazone, disulfide, hemiacetal, thiohemiacetal, α-keto-alkyl-thioalkyl or α-keto-alkylaminoalkyl.

35. The cross-linked oligonucleotide of claim 34 wherein X_1 and X_2, independently, are O, S or NH.

36. The cross-linked oligonucleotide of claim 34 wherein Y is imine, amine, oxime, hydrazine, hydrazone or hydroxylamine.

37. The cross-linked oligonucleotide of claim 34 wherein Y is disulfide.

38. The cross-linked oligonucleotide of claim 33 wherein:

X_1 and X_2, independently, are O or NH;

together L_1-Y-L_2 are HN-L_3-NH; and

L_3 is C_1-C_{20} straight chain alkyl, C_1-C_{20} straight chain substituted alkyl, C_2-C_{50} branched chain alkyl, C_2-C_{50} branched chain substituted alkyl, C_2-C_{20} straight chain alkenyl, C_2-C_{20} straight chain substituted alkenyl, C_3-C_{50} branched chain alkenyl, C_3-C_{50} branched chain substituted alkenyl, C_2-C_{20} straight chain alkyne, C_2-C_{20} straight chain substituted alkyne, C_3-C_{50} branched chain alkyne, C_3-C_{50} branched chain substituted alkyne, polyamine, polyamide, polyester, polyethylene glycol, aryl, aralkyl or heterocyclic.

39. The cross-linked oligonucleotide of claim 38 wherein X_1 and X_2 are O.

40. The cross-linked oligonucleotide of claim 38 wherein L_3 is polymethylene.

41. The cross-linked oligonucleotide of claim 33 wherein said first sequence region and said second sequence region are on a single oligonucleotide strand.

42. The cross-linked oligonucleotide of claim 33 wherein said first sequence region and said second sequence region are on different oligonucleotide strands.

43. The cross-linked oligonucleotide of claim 33 wherein:

X_1 and X_2, independently, are O, S, NH, CH_2 or CO;

L_1 and L_2, independently, are C_1-C_{20} straight chain alkylamino, C_1-C_{20} straight chain substituted alkylamino, C_2-C_{50} branched chain alkylamino, C_2-C_{50} branched chain substituted alkylamino, C_2-C_{20} straight chain alkenylamino, C_2-C_{20} straight chain substituted alkenylamino, C_3-C_{50} branched chain alkenylamino, C_3-C_{50} branched chain substituted alkenylamino, C_2-C_{20} straight chain alkynylamino, C_2-C_{20} straight chain substituted alkynylamino, C_3-C_{50} branched chain alkynylamino, C_3-C_{50} branched chain substituted alkynylamino, C_1-C_{20} straight chain alkylthio, C_1-C_{20} straight chain substituted alkylthio, C_2-C_{50} branched chain alkylthio, C_2-C_{50} branched chain substituted alkylthio, C_2-C_{20} straight chain alkenylthio, C_2-C_{20} straight chain substituted alkenylthio, C_3-C_{50} branched chain alkenylthio, C_3-C_{50} branched chain substituted alkenylthio, C_2-C_{20} straight chain alkynylthio, C_2-C_{20} straight chain substituted alkynylthio, C_3-C_{50} branched chain alkynylthio or C_3-C_{50} branched chain substituted alkynylthio; and

Y is a heterobifunctional or homobifunctional cross-linking reagent.

44. The cross-linked oligonucleotide of claim 43 wherein Y is a heterobifunctional cross-linking reagent comprising two reactive moieties independently selected from maleimide, N-hydroxysuccinimide, N-hydroxysulfosuccinimide, imidoester, pyridyl disulfide, ß-carbonyl alkylhalide or photoreactive aryl azide.

45. A cross-linked oligonucleotide comprising:

a first sequence region in an oligonucleotide;

a second sequence region in an oligonucleotide;

a first space-spanning group connected to a 2'-position of a nucleotide located in said first sequence region;

a second space-spanning group connected to a 2'-position of a nucleotide located on said second sequence region; and

a linking moiety covalently bonding said first and second space-spanning groups.

46. The cross-linked oligonucleotide of claim 45 wherein:

said linking moiety is -N=CH-, -N(R)-CH_2-N(R)-, -N(R)-, -O-N=CH-, -O-N(R)-, -N(R)-N(R)-, -HC=N-N(R)-, -HC=N-N=CH-, -C(O)-N(R)-N=CH-, -C(O)N(R)-, -C(O)-N(R)-N(R)-, -N(R)-C(O)-N(R)-N(R)-, -N(R)-C(O)-N(R)-N=CH-, -N(R)-C(S)-N(R)-N(R)-, -N(R)-C(S)-N(R)-N=CH-, -S-S-, -CH(OH)-O-CH_2-, -CH(OH)-S-CH_2-, -CH_2-S-CH_2-C(O)-CH_2- or -CH_2-NH-CH_2-C(O)-CH_2-; and

R is H, alkyl, alkenyl, aralkyl, an intercalator, a conjugate, polyamine, polyamide, polyethylene glycol, a group that enhances the pharmacodynamic properties of oligonucleotides or a group that enhances the pharmacokinetic properties of oligonucleotides.

47. The cross-linked oligonucleotide of claim 46 wherein R is H.

48. A method of fixing an oligonucleotide in a set configuration comprising:

selecting a first site in said oligonucleotide, said first site having a first nucleotide sequence;

attaching a first bond precursor to the sugar moiety of a first nucleotide of said first site;

selecting a second site in said oligonucleotide, said second site having a second nucleotide sequence;

attaching a second bond precursor to the sugar moiety of a second nucleotide of said second site; and

forming a covalent cross-linkage between said first and said second bond precursors.

49. The method of claim 48 including:

attaching said first bond precursor to said first nucleotide by reacting a precursor of said first nucleotide with a reagent for introducing said first bond precursor on the sugar moiety of said first nucleotide precursor;

derivatizing said first nucleotide precursor to a protected and activated form;

synthetically incorporating said derivatized first nucleotide precursor into said first nucleotide sequence;

attaching said second bond precursor to said second nucleotide by reacting a precursor of said second nucleotide with a reagent for introducing said second bond precursor on the sugar moiety of said second nucleotide precursor;

derivatizing said second nucleotide precursor to a protected and activated form; and

synthetically incorporating said derivatized second nucleotide precursor into said second nucleotide sequence.

50. The method of claim 48 including:

selecting an abasic nucleotide as said first bond precursor; and

attaching said second bond precursor to a 2'-position of said second nucleotide.

51. The method of claim 48 including:

attaching said first bond precursor to a 2'-position of said first nucleotide; and

attaching said second bond precursor to a 2'-position of said second nucleotide.

52. The method of claim 51 including selecting said cross-linkage from:

-N=CH-, -N(R)-, -N(R)-CH_2-NH-, -NH-CH_2-NH-, -O-N=CH-, -O-N(R)-, -N(R)-N(R)-, -HC=N-N(R)-, -HC=N-N=CH-, -C(O)-N(R)-N=CH-, -C(O)N(R)-, -C(O)-N(R)-N(R)-, -N(R)-C(O)-N(R)-N(R)-, -N(R)-C(O)-N(R)

-N=CH-, -N(R)-C(S)-N(R)-NH-, -N(R)-C(S)-N(R)-N=CH-, -S-S-, -CH(OH)-O-CH_2-, -CH(OH)-S-CH_2-, -CH_2-S-CH_2-C(O)-CH_2- or -CH_2-NH-CH_2-C(O)-CH_2-;

wherein R is H, alkyl, alkenyl, aralkyl, an intercalator, a conjugate, polyamine, polyamide, polyethylene glycol, a group that enhances the pharmacodynamic properties of cross-linked duplexes or a group that enhances the pharmacokinetic properties of cross-linked duplexes.

53. The method of claim 51 including selecting said cross-linkage from imine, amine, aminoalkylamine, oxime, hydroxylamine, hydrazine, hydrazone, azine, hydrazide-hydrazone, amide, hydrazide, semicarbazide, semicarbazone, thiosemicarbazide, thiocarbazone, disulfide, hemiacetal, thiohemiacetal, α-keto-alkylthioalkyl or α-keto-alkyl-aminoalkyl linkages.

54. The method of claim 48 including selecting said sequence of nucleotides of said first site to be complementary to and specifically hybridizable with said sequence of nucleotides of said second site.

55. A method for cross-linking a single oligonucleotide comprising:

selecting a first nucleotide sequence in an oligonucleotide;

selecting a second nucleotide sequence in said oligonucleotide;

incorporating an abasic nucleotide within said first nucleotide sequence;

attaching a bond precursor on a nucleotide of said second sequence; and

forming a covalent cross-linkage between said abasic site and said bond precursor.

56. The method of claim 55 including attaching said bond precursor on the sugar moiety of said second sequence nucleotide.

57. The method of claim 55 including attaching said bond precursor on a 2'-position of said second nucleotide.

58. The method of claim 55 including:

forming an aldehydic functional group at said abasic site; and

selecting said bond precursor to include a functional group capable of forming one of an oxime, imine, hydrazone, hydrazide-hydrazone, semicarbazone, thiosemicarbazone, hemiacetal or thio-hemiacetal covalent linkage with said abasic site aldehydic functional group.

59. The method of claim 55 including selecting said first nucleotide sequence to be complementary to and specifically hybridizable with said second nucleotide sequence.

60. A method of cross-linking a single oligonucleotide strand comprising:

selecting a first nucleotide sequence in an oligonucleotide;

selecting a second nucleotide sequence in said oligonucleotide;

attaching a first bond precursor to the sugar moiety of a first nucleotide of said first nucleotide sequence;

attaching a second bond precursor to the sugar moiety of a second nucleotide of said second nucleotide sequence; and

forming a covalent cross-linkage between said first and said second bond precursors.

61. The method of claim 60 including selecting said first nucleotide sequence to be complementary to and specifically hybridizable with said second nucleotide sequence.

62. The method of claim 60 including selecting said covalent cross-linkage from imine, amine, aminoalkylamine, oxime, hydroxylamine, hydrazine, hydrazone, azine, hydrazide-hydrazone, amide, hydrazide, semicarbazide, semicarbazone, thiosemicarbazide, thiocarbazone, disulfide, hemiacetal or thiohemiacetal linkages,

63. A method of cross-linking first and second strands of oligonucleotides comprising:

selecting a first nucleotide sequence in a first oligonucleotide strand;

selecting a second nucleotide sequence in a second oligonucleotide strand;

incorporating an abasic nucleotide within said first nucleotide sequence;

attaching a bond precursor to a nucleotide of said second nucleotide sequence; and

forming a covalent cross-linkage between said abasic site and said bond precursor.

64. The method of claim 63 including selecting said first nucleotide sequence to be complementary to and specifically hybridizable with said second nucleotide sequence.

65. The method of claim 63 including attaching said sequence bond precursor to the sugar moiety of said second sequence nucleotide.

66. The method of claim 65 including attaching said bond precursor to a 2'-position of said second nucleotide.

67. The method of claim 63 including:

forming an aldehydic functional group at said abasic site; and

selecting said bond precursor to include a functional group capable of forming one of an oxime, imine, hydrazone, semicarbazone, thiosemicarbazone, hydrazide-hydrazone, hemiacetal or thiohemiacetal linkage with said abasic site aldehydic functional group.

68. A method of cross-linking first and second strands of oligonucleotides comprising:

selecting a first nucleotide sequence in a first oligonucleotide strand;

selecting a second nucleotide sequence in a second oligonucleotide strand;

attaching a first bond precursor to the sugar moiety of a selected nucleotide of said first nucleotide sequence;

attaching a second bond precursor to the sugar moiety of a nucleotide of said second nucleotide sequence; and

forming a covalent cross-linkage between said first and said second bond precursors.

69. The method of claim 68 including selecting said first nucleotide sequence to be complementary to and specifically hybridizable with said second nucleotide sequence.

70. The method of claim 68 including choosing said covalent cross-linkage from imine, amine, aminoalkylamine, oxime, hydroxylamine, hydrazine, hydrazone, azine, hydrazide-hydrazone, amide, hydrazide, semicarbazide, semicarbazone, thiosemicarbazide, thiocarbazone, disulfide, hemiacetal or thiohemiacetal linkages.

AMENDED CLAIMS

[received by the International Bureau on 1 July 1993 (01.07.93); original claims 2-5, 8-11, 15-18, 21, 23, 24, 25, 30, 48, 60 and 68 amended; other claims unchanged (24 pages)]

1. A cross-linked nucleic acid comprising:

a first nucleotide located on a first oligonucleotide strand, said first nucleotide being covalently bound to at least two further nucleotides on said first oligonucleotide strand;

a first bond means located on a sugar moiety of said first nucleotide;

a second nucleotide located on a further oligonucleotide strand, said second nucleotide being covalently bound to at least two further nucleotides on said further oligonucleotide strand;

a second bond means located on a sugar moiety of said second nucleotide; and

a covalent cross-linkage between said first and said second bond means.

2. A cross-linked nucleic acid comprising:

a first nucleotide located on an oligonucleotide strand, said first nucleotide being covalently bound to at least two further nucleotides on said oligonucleotide strand;

a first bond means located on a sugar moiety of said first nucleotide;

a second nucleotide located on said oligonucleotide strand, said second nucleotide being covalently bound to at least two further nucleotides on said oligonucleotide strand different from said first nucleotide;

a second bond means located on a sugar moiety of said second nucleotide; and

a covalent cross-linkage between said first and said second bond means.

3. A cross-linked nucleic acid comprising:

a first nucleotide located on a 3′ end of a first oligonucleotide strand, said first nucleotide including a 2′,3′-dialdehyde structure;

a second nucleotide located on a further oligonucleotide strand or on said first oligonucleotide strand;

a space-spanning group located on a sugar moiety of said second nucleotide; and

a covalent cross-linkage between said dialdehyde structure and said space-spanning group.

4. A cross-linked nucleic acid comprising:

a first nucleotide located on a 3′ end of a first oligonucleotide strand, said first nucleotide including 2′ and 3′ hydroxyl groups;

a second nucleotide located on a further oligonucleotide strand or on said first oligonucleotide strand;

a space-spanning group located on a sugar moiety of said second specific nucleotide; and

a covalent acetal cross-linkage between said 2′ and 3′ hydroxyl groups and said space-spanning group.

5. A cross-linked nucleic acid comprising:

a first nucleotide located on a first oligonucleotide strand, said first nucleotide being covalently bound to at least two further nucleotides on said first oligonucleotide strand;

a first bond means located on a sugar moiety of said first nucleotide;

a second nucleotide located on a further oligonucleotide strand or on said first oligonucleotide strand, said second nucleotide being covalently bound to at least two further nucleotides on said further oligonucleotide strand or said first oligonucleotide strand;

a second bond means located on said second nucleotide; and

a covalent cross-linkage between said first and said second bond means.

6. The cross-linked nucleic acid of claim 5 wherein:

said first nucleotide is located in a first nucleotide sequence and said second nucleotide is located in a second nucleotide sequence;

said second nucleotide sequence is complementary to and hybridizable with said first nucleotide sequence; and

said first and said second nucleotide sequences are located on a single strand of said oligonucleotide with said second nucleotide sequence located on said oligonucleotide at a distance separated from said first nucleotide sequence sufficient to allow said oligonucleotide to assume a conformation wherein said first and second nucleotide sequences are mutually aligned with and specifically hybridized with one another.

7. The cross-linked nucleic acid of claim 5 wherein:

said first nucleotide is located in a first nucleotide sequence and said second nucleotide is located in a second nucleotide sequence;

said second nucleotide sequence is complementary to and specifically hybridizable with said first nucleotide sequence; and

said first and said second nucleotide sequences are located on different oligonucleotide strands.

8. The cross-linked nucleic acid of claim 5 wherein:

said first bond means includes an abasic site; and

said second bond means includes a space-spanning group covalently bonded with said abasic site.

9. The cross-linked nucleic acid of claim 8 wherein said abasic site includes an aldehyde.

10. The cross-linked nucleic acid of claim 8 wherein said space-spanning group includes an amine,

hydrazine, hydroxylamine, semicarbazide, thiosemicarbazide, hydrazone, hydrazide, alcohol or thiol.

11. The cross-linked nucleic acid of claim 10 wherein said space-spanning group includes an amine.

12. The cross-linked nucleic acid of claim 8 wherein said space-spanning group is attached to a 2'-position of said second nucleotide.

13. The cross-linked nucleic acid of claim 8 wherein said space-spanning group is C_1-C_{20} straight chain alkyl, C_1-C_{20} straight chain O-alkyl, C_1-C_{20} straight chain S-alkyl, C_1-C_{20} straight chain NH-alkyl, C_1-C_{20} straight chain substituted alkyl, C_1-C_{20} straight chain substituted O-alkyl, C_1-C_{20} straight chain substituted S-alkyl, C_1-C_{20} straight chain substituted NH-alkyl, C_2-C_{50} branched chain alkyl, C_2-C_{50} branched chain O-alkyl, C_2-C_{50} branched chain S-alkyl, C_2-C_{50} branched chain NH-alkyl, C_2-C_{50} branched chain substituted alkyl, C_2-C_{50} branched chain substituted O-alkyl, C_2-C_{50} branched chain substituted S-alkyl, C_2-C_{50} branched chain substituted NH-alkyl, C_2-C_{20} straight chain alkenyl, C_2-C_{20} straight chain O-alkenyl, C_2-C_{20} straight chain S-alkenyl, C_2-C_{20} straight chain NH-alkenyl, C_2-C_{20} straight chain substituted alkenyl, C_2-C_{20} straight chain substituted O-alkenyl, C_2-C_{20} straight chain substituted S-alkenyl, C_2-C_{20} straight chain substituted NH-alkenyl, C_3-C_{50} branched chain alkenyl, C_3-C_{50} branched chain O-alkenyl, C_3-C_{50} branched chain S-alkenyl, C_3-C_{50} branched chain NH-alkenyl, C_3-C_{50} branched chain substituted alkenyl, C_3-C_{50} branched chain substituted O-alkenyl, C_3-C_{50} branched chain substituted S-alkenyl, C_3-C_{50} branched chain substituted NH-alkenyl, C_2-C_{20} straight chain alkyne, C_2-C_{20} straight chain O-alkyne, C_2-C_{20} straight chain S-alkyne, C_2-C_{20} straight chain NH-alkyne, C_2-C_{20} straight chain substituted alkyne, C_2-C_{20} straight chain substituted O-alkyne, C_2-C_{20} straight chain substituted S-alkyne, C_2-C_{20} straight chain substituted NH-alkyne, C_3-C_{50} branched chain alkyne, C_2-C_{50}

branched chain O-alkyne, C_3-C_{50} branched chain S-alkyne, C_3-C_{50} branched chain NH-alkyne, C_3-C_{50} branched chain substituted alkyne, C_3-C_{50} branched chain substituted O-alkyne, C_3-C_{50} branched chain substituted S-alkyne, C_3-C_{50} branched chain substituted NH-alkyne, polyamine, polyamide, polyester, polyethylene glycol, aryl, aralkyl or heterocyclic.

14. The cross-linked nucleic acid of claim 13 wherein said space-spanning group is from about 8 to about 13 atoms in length.

15. The cross-linked nucleic acid of claim 5 wherein:

said first bond means includes a first space-spanning group;

said second bond means includes a second space-spanning group; and

said covalent cross-linkage is between said first and said second space-spanning groups.

16. The cross-linked nucleic acid of claim 15 wherein said first space-spanning group is an aldehyde, a protected aldehyde or an aldehyde precursor.

17. The cross-linked nucleic acid of claim 16 wherein said first space-spanning group is C-formyl, *o*-methylamino-benzenethiol, aryl substituted imidazolidino, dithiane, bis(*p*-nitrophenoxy), acetal, nitrile, ester or alcohol.

18. The cross-linked nucleic acid of claim 15 wherein said second space-spanning group is an amine, hydrazine, hydroxylamine, semicarbazide, thiosemicarbazide, hydrazide, alcohol or thiol.

19. The cross-linked nucleic acid of claim 15 wherein:

said first space-spanning group is attached to a 2'-position of said first nucleotide; and

said second space-spanning group is attached to a 2'-position of said second nucleotide.

20. The cross-linked nucleic acid of claim 15 wherein each of said first and second space-spanning groups are, independently, C_1-C_{20} straight chain alkyl, C_1-C_{20} straight chain O-alkyl, C_1-C_{20} straight chain S-alkyl, C_1-C_{20} straight chain NH-alkyl, C_1-C_{20} straight chain substituted alkyl, C_1-C_{20} straight chain substituted O-alkyl, C_1-C_{20} straight chain substituted S-alkyl, C_1-C_{20} straight chain substituted NH-alkyl, C_2-C_{50} branched chain alkyl, C_2-C_{50} branched chain O-alkyl, C_2-C_{50} branched chain S-alkyl, C_2-C_{50} branched chain NH-alkyl, C_2-C_{50} branched chain substituted alkyl, C_2-C_{50} branched chain substituted O-alkyl, C_2-C_{50} branched chain substituted S-alkyl, C_2-C_{50} branched chain substituted NH-alkyl, C_2-C_{20} straight chain alkenyl, C_2-C_{20} straight chain O-alkenyl, C_2-C_{20} straight chain S-alkenyl, C_2-C_{20} straight chain NH-alkenyl, C_2-C_{20} straight chain substituted alkenyl, C_2-C_{20} straight chain substituted O-alkenyl, C_2-C_{20} straight chain substituted S-alkenyl, C_2-C_{20} straight chain substituted NH-alkenyl, C_3-C_{50} branched chain alkenyl, C_3-C_{50} branched chain O-alkenyl, C_3-C_{50} branched chain S-alkenyl, C_3-C_{50} branched chain NH-alkenyl, C_3-C_{50} branched chain substituted alkenyl, C_3-C_{50} branched chain substituted O-alkenyl, C_3-C_{50} branched chain substituted S-alkenyl, C_3-C_{50} branched chain substituted NH-alkenyl, C_2-C_{20} straight chain alkyne, C_2-C_{20} straight chain O-alkyne, C_2-C_{20} straight chain S-alkyne, C_2-C_{20} straight chain NH-alkyne, C_2-C_{20} straight chain substituted alkyne, C_2-C_{20} straight chain substituted O-alkyne, C_2-C_{20} straight chain substituted S-alkyne, C_2-C_{20} straight chain substituted NH-alkyne, C_3-C_{50} branched chain alkyne, C_2-C_{50} branched chain O-alkyne, C_3-C_{50} branched chain S-alkyne, C_3-C_{50} branched chain NH-alkyne, C_3-C_{50} branched chain

substituted alkyne, C_3-C_{50} branched chain substituted O-alkyne, C_3-C_{50} branched chain substituted S-alkyne, C_3-C_{50} branched chain substituted NH-alkyne, polyamine, polyamide, polyester, polyethylene glycol, aryl, aralkyl or heterocyclic.

21. The cross-linked nucleic acid of claim 15 wherein said first and second space-spanning groups are thiol groups and said covalent cross-linkage is a disulfide bond.

22. The cross-linked nucleic acid of claim 5 wherein:

said first nucleotide is located in a first nucleotide sequence and said second nucleotide is located in a second nucleotide sequence; and

said first and said second nucleotide sequences are located on a single strand of an oligonucleotide with said second nucleotide sequence located on said oligonucleotide at a non-hybridizable position with respect to said first nucleotide sequence and at a distance from said first nucleotide sequence sufficient to allow said oligonucleotide to assume a conformation wherein said first and second nucleotide sequences are located in spatial proximity with each other.

23. The cross-linked nucleic acid of claim 5 wherein:

said first bond means includes a 3'-terminal nucleoside having a 2',3'-dialdehyde structure;

said second bond means includes a space-spanning group; and

said active functional group is capable of reacting with a dialdehyde.

24. The cross-linked nucleic acid of claim 5 wherein:

said first bond means includes a 3'-terminal nucleotide having 2' and 3'-sugar hydroxyl groups;

said second bond means includes a space-spanning group; and

an acetal cross-linkage connects said 2' and 3' hydroxyl groups and said space-spanning group.

25. A cross-linked oligonucleotide comprising:

a first sequence region in an oligonucleotide;

a second sequence region in an oligonucleotide;

an abasic site located in said first sequence region; and

a space-spanning group located on said second sequence region and covalently bonded with said abasic site.

26. The cross-linked oligonucleotide of claim 25 wherein said first sequence region and said second sequence region are on a single oligonucleotide strand.

27. The cross-linked oligonucleotide of claim 25 wherein said first sequence region and said second sequence region are on different oligonucleotide strands.

28. The cross-linked oligonucleotide of claim 25 wherein said first sequence region and said second sequence region are complementary to and specifically hybridizable with one another.

29. The cross-linked oligonucleotide of claim 25 wherein said space-spanning group is C_1-C_{20} straight chain alkyl, C_1-C_{20} straight chain O-alkyl, C_1-C_{20} straight chain S-alkyl, C_1-C_{20} straight chain NH-alkyl, C_1-C_{20} straight chain substituted alkyl, C_1-C_{20} straight chain substituted O-alkyl, C_1-C_{20} straight chain substituted S-alkyl, C_1-C_{20} straight chain

substituted NH-alkyl, C_2-C_{50} branched chain alkyl, C_2-C_{50} branched chain O-alkyl, C_2-C_{50} branched chain S-alkyl, C_2-C_{50} branched chain NH-alkyl, C_2-C_{50} branched chain substituted alkyl, C_2-C_{50} branched chain substituted O-alkyl, C_2-C_{50} branched chain substituted S-alkyl, C_2-C_{50} branched chain substituted NH-alkyl, C_2-C_{20} straight chain alkenyl, C_2-C_{20} straight chain O-alkenyl, C_2-C_{20} straight chain S-alkenyl, C_2-C_{20} straight chain NH-alkenyl, C_2-C_{20} straight chain substituted alkenyl, C_2-C_{20} straight chain substituted O-alkenyl, C_2-C_{20} straight chain substituted S-alkenyl, C_2-C_{20} straight chain substituted NH-alkenyl, C_3-C_{50} branched chain alkenyl, C_3-C_{50} branched chain O-alkenyl, C_3-C_{50} branched chain S-alkenyl, C_3-C_{50} branched chain NH-alkenyl, C_3-C_{50} branched chain substituted alkenyl, C_3-C_{50} branched chain substituted O-alkenyl, C_3-C_{50} branched chain substituted S-alkenyl, C_3-C_{50} branched chain substituted NH-alkenyl, C_2-C_{20} straight chain alkyne, C_2-C_{20} straight chain O-alkyne, C_2-C_{20} straight chain S-alkyne, C_2-C_{20} straight chain NH-alkyne, C_2-C_{20} straight chain substituted alkyne, C_2-C_{20} straight chain substituted O-alkyne, C_2-C_{20} straight chain substituted S-alkyne, C_2-C_{20} straight chain substituted NH-alkyne, C_3-C_{50} branched chain alkyne, C_2-C_{50} branched chain O-alkyne, C_3-C_{50} branched chain S-alkyne, C_3-C_{50} branched chain NH-alkyne, C_3-C_{50} branched chain substituted alkyne, C_3-C_{50} branched chain substituted O-alkyne, C_3-C_{50} branched chain substituted S-alkyne, C_3-C_{50} branched chain substituted NH-alkyne, polyamine, polyamide, polyester, polyethylene glycol, aryl, aralkyl or heterocyclic.

30. The cross-linked oligonucleotide of claim 25 wherein said space-spanning group includes an amine, hydrazine, hydroxylamine, semicarbazide, hydrazide, alcohol or thiol.

31. The cross-linked oligonucleotide of claim 25 wherein said space-spanning group is CH_2-L_3, O-L_3, S-L_3, NH-L_3 or CO-L_3 where L_3 is C_1-C_{20} straight chain alkyl, C_1-C_{20} straight

chain substituted alkyl, C_2-C_{50} branched chain alkyl, C_2-C_{50} branched chain substituted alkyl, C_2-C_{20} straight chain alkenyl, C_2-C_{20} straight chain substituted alkenyl, C_3-C_{50} branched chain alkenyl, C_3-C_{50} branched chain substituted alkenyl, C_2-C_{20} straight chain alkyne, C_2-C_{20} straight chain substituted alkyne, C_3-C_{50} branched chain alkyne, C_3-C_{50} branched chain substituted alkyne, polyamine, polyamide, polyester, polyethylene glycol, aryl, aralkyl or heterocyclic.

32. The cross-linked oligonucleotide of claim 25 wherein said space-spanning group is located on a 2'-position of a nucleotide within said second sequence region.

33. A cross-linked oligonucleotide comprising:
a first sequence region in an oligonucleotide;
a second sequence region in an oligonucleotide; and
a covalently bonded linker of the structure:

$$X_1-L_1-Y-L_2-X_2$$

wherein X_1 is covalently connected to said first sequence region, X_2 is covalently connected to said second sequence region, and where:

(i) X_1 and X_2, independently, are O, S, NH, CH_2 or CO;

L_1 and L_2, independently, are C_1-C_{20} straight chain alkyl, C_1-C_{20} straight chain substituted alkyl, C_2-C_{50} branched chain alkyl, C_2-C_{50} branched chain substituted alkyl, C_2-C_{20} straight chain alkenyl, C_2-C_{20} straight chain substituted alkenyl, C_3-C_{50} branched chain alkenyl, C_3-C_{50} branched chain substituted alkenyl, C_2-C_{20} straight chain alkyne, C_2-C_{20} straight chain substituted alkyne, C_3-C_{50} branched chain alkyne, C_3-C_{50} branched chain substituted alkyne, polyamine, polyamide, polyester, polyethylene glycol, polyether, aryl, aralkyl or heterocyclic; and

Y is an imine, amine, oxime, hydroxylamine, hydrazine, hydrazone, azine, hydrazide-hydrazone, amide, hydrazide, semicarbazide, semicarbazone,

thiosemicarbazide, thiocarbazone, disulfide, hemiacetal, thiohemiacetal, α-keto-alkylthioalkyl or α-keto-alkylaminoalkyl; or

(ii) X_1 and X_2, independently, are O or NH; and

L_1-Y-L_2 together are NH-L_3-NH or NH-NH-L_3-NH-NH; and

L_3 is C_1-C_{20} straight chain alkyl, C_1-C_{20} straight chain substituted alkyl, C_2-C_{50} branched chain alkyl, C_2-C_{50} branched chain substituted alkyl, C_2-C_{20} straight chain alkenyl, C_2-C_{20} straight chain substituted alkenyl, C_3-C_{50} branched chain alkenyl, C_3-C_{50} branched chain substituted alkenyl, C_2-C_{20} straight chain alkyne, C_2-C_{20} straight chain substituted alkyne, C_3-C_{50} branched chain alkyne, C_3-C_{50} branched chain substituted alkyne, polyamine, polyamide, polyester, polyethylene glycol, polyether, aryl, aralkyl or heterocyclic; or

(iii) X_1 and X_2, independently, are O, S, NH, CH_2 or CO;

L_1 and L_2, independently, are C_1-C_{20} straight chain alkylamino, C_1-C_{20} straight chain substituted alkylamino, C_2-C_{50} branched chain alkylamino, C_2-C_{50} branched chain substituted alkylamino, C_2-C_{20} straight chain alkenylamino, C_2-C_{20} straight chain substituted alkenylamino, C_3-C_{50} branched chain alkenylamino, C_3-C_{50} branched chain substituted alkenylamino, C_2-C_{20} straight chain alkynylamino, C_2-C_{20} straight chain substituted alkynylamino, C_3-C_{50} branched chain alkynylamino, C_3-C_{50} branched chain substituted alkynylamino, C_1-C_{20} straight chain alkylthio, C_1-C_{20} straight chain substituted alkylthio, C_2-C_{50} branched chain alkylthio, C_2-C_{50}, branched chain substituted alkylthio, C_2-C_{20} straight chain alkenylthio, C_2-C_{20} straight chain substituted alkenylthio, C_3-C_{50} branched chain alkenylthio, C_3-C_{50} branched chain substituted alkenylthio, C_2-C_{20} straight chain alkynylthio, C_2-C_{20} straight chain substituted alkynylthio, C_3-C_{50} branched chain alkynylthio or C_3-C_{50} branched chain substituted

alkynylthio; and

Y is a heterobifunctional or homobifunctional cross-linking reagent;

where said substituent of said substituted alkyl, substituted alkenyl and substituted alkyne moieties is an intercalator, a conjugate, polyamine, polyamide, polyethylene glycol, a group that enhances the pharmacodynamic properties of oligonucleotides or a group that enhances the pharmacokinetic properties of oligonucleotides.

34. The cross-linked oligonucleotide of claim 33 wherein:

X_1 and X_2, independently, are O, S, NH, CH_2 or CO;

L_1 and L_2, independently, are C_1-C_{20} straight chain alkyl, C_1-C_{20} straight chain substituted alkyl, C_2-C_{50} branched chain alkyl, C_2-C_{50}, branched chain substituted alkyl, C_2-C_{20} straight chain alkenyl, C_2-C_{20} straight chain substituted alkenyl, C_3-C_{50} branched chain alkenyl, C_3-C_{50} branched chain substituted alkenyl, C_2-C_{20} straight chain alkyne, C_2-C_{20} straight chain substituted alkyne, C_3-C_{50} branched chain alkyne, C_3-C_{50} branched chain substituted alkyne, polyamine, polyamide, polyester, polyethylene glycol, aryl, aralkyl or heterocyclic; and

Y is imine, amine, oxime, hydroxylamine, hydrazine, hydrazone, azine, hydrazide-hydrazone, amide, hydrazide, semicarbazide, semicarbazone, thiosemicarbazide, thiocarbazone, disulfide, hemiacetal, thiohemiacetal, α-keto-alkylthioalkyl or α-keto-alkylaminoalkyl.

35. The cross-linked oligonucleotide of claim 34 wherein X_1 and X_2, independently, are O, S or NH.

36. The cross-linked oligonucleotide of claim 34 wherein Y is imine, amine, oxime, hydrazine, hydrazone or hydroxylamine.

37. The cross-linked oligonucleotide of claim 34 wherein Y is disulfide.

38. The cross-linked oligonucleotide of claim 33 wherein:

X_1 and X_2, independently, are O or NH;

together L_1-Y-L_2 are HN-L_3-NH; and

L_3 is C_1-C_{20} straight chain alkyl, C_1-C_{20} straight chain substituted alkyl, C_2-C_{50} branched chain alkyl, C_2-C_{50} branched chain substituted alkyl, C_2-C_{20} straight chain alkenyl, C_2-C_{20} straight chain substituted alkenyl, C_3-C_{50} branched chain alkenyl, C_3-C_{50} branched chain substituted alkenyl, C_2-C_{20} straight chain alkyne, C_2-C_{20} straight chain substituted alkyne, C_3-C_{50} branched chain alkyne, C_3-C_{50} branched chain substituted alkyne, polyamine, polyamide, polyester, polyethylene glycol, aryl, aralkyl or heterocyclic.

39. The cross-linked oligonucleotide of claim 38 wherein X_1 and X_2 are O.

40. The cross-linked oligonucleotide of claim 38 wherein L_3 is polymethylene.

41. The cross-linked oligonucleotide of claim 33 wherein said first sequence region and said second sequence region are on a single oligonucleotide strand.

42. The cross-linked oligonucleotide of claim 33 wherein said first sequence region and said second sequence region are on different oligonucleotide strands.

43. The cross-linked oligonucleotide of claim 33 wherein:

X_1 and X_2, independently, are O, S, NH, CH_2 or CO;

L_1 and L_2, independently, are C_1-C_{20} straight chain alkylamino, C_1-C_{20} straight chain substituted alkylamino, C_2-C_{50} branched chain alkylamino, C_2-C_{50} branched chain substituted alkylamino, C_2-C_{20} straight chain alkenylamino, C_2-C_{20} straight chain substituted alkenylamino, C_3-C_{50} branched chain alkenylamino, C_3-C_{50} branched chain substituted alkenylamino, C_2-C_{20} straight chain alkynylamino, C_2-C_{20} straight chain substituted alkynylamino, C_3-C_{50} branched chain alkynylamino, C_3-C_{50} branched chain substituted alkynylamino, C_1-C_{20} straight chain alkylthio, C_1-C_{20} straight chain substituted alkylthio, C_2-C_{50} branched chain alkylthio, C_2-C_{50} branched chain substituted alkylthio, C_2-C_{20} straight chain alkenylthio, C_2-C_{20} straight chain substituted alkenylthio, C_3-C_{50} branched chain alkenylthio, C_3-C_{50} branched chain substituted alkenylthio, C_2-C_{20} straight chain alkynylthio, C_2-C_{20} straight chain substituted alkynylthio, C_3-C_{50} branched chain alkynylthio or C_3-C_{50} branched chain substituted alkynylthio; and

Y is a heterobifunctional or homobifunctional cross-linking reagent.

44. The cross-linked oligonucleotide of claim 43 wherein Y is a heterobifunctional cross-linking reagent comprising two reactive moieties independently selected from maleimide, N-hydroxysuccinimide, N-hydroxysulfosuccinimide, imidoester, pyridyl disulfide, ß-carbonyl alkylhalide or photoreactive aryl azide.

45. A cross-linked oligonucleotide comprising:

a first sequence region in an oligonucleotide;

a second sequence region in an oligonucleotide;

a first space-spanning group connected to a 2'-position of a nucleotide located in said first sequence region;

a second space-spanning group connected to a 2'-position of a nucleotide located on said second sequence region; and

a linking moiety covalently bonding said first and second space-spanning groups.

46. The cross-linked oligonucleotide of claim 45 wherein:

said linking moiety is -N=CH-, -N(R)-CH_2-N(R)-, -N(R)-, -O-N=CH-, -O-N(R)-, -N(R)-N(R)-, -HC=N-N(R)-, -HC=N-N=CH-, -C(O)-N(R)-N=CH-, -C(O)N(R)-, -C(O)-N(R)-N(R)-, -N(R)-C(O)-N(R)-N(R)-, -N(R)-C(O)-N(R)-N=CH-, -N(R)-C(S)-N(R)-N(R)-, -N(R)-C(S)-N(R)-N=CH-, -S-S-, -CH(OH)-O-CH_2-, -CH(OH)-S-CH_2-, -CH_2-S-CH_2-C(O)-CH_2- or -CH_2-NH-CH_2-C(O)-CH_2-; and

R is H, alkyl, alkenyl, aralkyl, an intercalator, a conjugate, polyamine, polyamide, polyethylene glycol, a group that enhances the pharmacodynamic properties of oligonucleotides or a group that enhances the pharmacokinetic properties of oligonucleotides.

47. The cross-linked oligonucleotide of claim 46 wherein R is H.

48. A method of fixing an oligonucleotide in a set configuration comprising:

selecting a first site in said oligonucleotide, said first site having a first nucleotide sequence;

attaching a first bond precursor to the sugar moiety of a first nucleotide of said first site, said first nucleotide being covalently bound to at least two further nucleotides in said first nucleotide sequence;

selecting a second site in said oligonucleotide, said second site having a second nucleotide sequence;

attaching a second bond precursor to the sugar moiety of a second nucleotide of said second site, said second nucleotide being covalently bound to at least two further nucleotides in said second nucleotide sequence; and

forming a covalent cross-linkage between said first and said second bond precursors.

49. The method of claim 48 including:

attaching said first bond precursor to said first nucleotide by reacting a precursor of said first nucleotide with a reagent for introducing said first bond precursor on the sugar moiety of said first nucleotide precursor;

derivatizing said first nucleotide precursor to a protected and activated form;

synthetically incorporating said derivatized first nucleotide precursor into said first nucleotide sequence;

attaching said second bond precursor to said second nucleotide by reacting a precursor of said second nucleotide with a reagent for introducing said second bond precursor on the sugar moiety of said second nucleotide precursor;

derivatizing said second nucleotide precursor to a protected and activated form; and

synthetically incorporating said derivatized second nucleotide precursor into said second nucleotide sequence.

50. The method of claim 48 including:

selecting an abasic nucleotide as said first bond precursor; and

attaching said second bond precursor to a 2'-position of said second nucleotide.

51. The method of claim 48 including:

attaching said first bond precursor to a 2'-position of said first nucleotide; and

attaching said second bond precursor to a 2'-position of said second nucleotide.

52. The method of claim 51 including selecting said cross-linkage from:

-N=CH-, -N(R)-, -N(R)-CH_2-NH-, -NH-CH_2-NH-, -O-N=CH-, -O-N(R)-, -N(R)-N(R)-, -HC=N-N(R)-, -HC=N-N=CH-, -C(O)-N(R)-N=CH-, -C(O)N(R)-, -C(O)-N(R)-N(R)-, -N(R)-C(O)-N(R)-N(R)-, -N(R)-C(O)-N(R)

-N=CH-, -N(R)-C(S)-N(R)-NH-, -N(R)-C(S)-N(R)-N=CH-, -S-S-, -CH(OH)-O-CH_2-, -CH(OH)-S-CH_2-, -CH_2-S-CH_2-C(O)-CH_2- or -CH_2-NH-CH_2-C(O)-CH_2-;

wherein R is H, alkyl, alkenyl, aralkyl, an intercalator, a conjugate, polyamine, polyamide, polyethylene glycol, a group that enhances the pharmacodynamic properties of cross-linked duplexes or a group that enhances the pharmacokinetic properties of cross-linked duplexes.

53. The method of claim 51 including selecting said cross-linkage from imine, amine, aminoalkylamine, oxime, hydroxylamine, hydrazine, hydrazone, azine, hydrazide-hydrazone, amide, hydrazide, semicarbazide, semicarbazone, thiosemicarbazide, thiocarbazone, disulfide, hemiacetal, thiohemiacetal, α-keto-alkylthioalkyl or α-keto-alkylaminoalkyl linkages.

54. The method of claim 48 including selecting said sequence of nucleotides of said first site to be complementary to and specifically hybridizable with said sequence of nucleotides of said second site.

55. A method for cross-linking a single oligonucleotide comprising:

selecting a first nucleotide sequence in an oligonucleotide;

selecting a second nucleotide sequence in said oligonucleotide;

incorporating an abasic nucleotide within said first nucleotide sequence;

attaching a bond precursor on a nucleotide of said second sequence; and

forming a covalent cross-linkage between said abasic site and said bond precursor.

56. The method of claim 55 including attaching said bond precursor on the sugar moiety of said second sequence nucleotide.

57. The method of claim 55 including attaching said bond precursor on a 2'-position of said second nucleotide.

58. The method of claim 55 including:

forming an aldehydic functional group at said abasic site; and

selecting said bond precursor to include a functional group capable of forming one of an oxime, imine, hydrazone, hydrazide-hydrazone, semicarbazone, thiosemicarbazone, hemi-acetal or thio-hemiacetal covalent linkage with said abasic site aldehydic functional group.

59. The method of claim 55 including selecting said first nucleotide sequence to be complementary to and specifically hybridizable with said second nucleotide sequence.

60. A method of cross-linking a single oligonucleotide strand comprising:

selecting a first nucleotide sequence in an oligo-nucleotide;

selecting a second nucleotide sequence in said oligonucleotide;

attaching a first bond precursor to the sugar moiety of a first nucleotide that is covalently bound to at least two further nucleotides in said first nucleotide sequence;

attaching a second bond precursor to the sugar moiety of a second nucleotide that is covalently bound to at least two further nucleotides in said second nucleotide sequence; and

forming a covalent cross-linkage between said first and said second bond precursors.

61. The method of claim 60 including selecting said first nucleotide sequence to be complementary to and specifically hybridizable with said second nucleotide sequence.

62. The method of claim 60 including selecting said covalent cross-linkage from imine, amine, aminoalkylamine, oxime, hydroxylamine, hydrazine, hydrazone, azine, hydrazide-hydrazone, amide, hydrazide, semicarbazide, semicarbazone, thiosemicarbazide, thiocarbazone, disulfide, hemiacetal or thiohemiacetal linkages,

63. A method of cross-linking first and second strands of oligonucleotides comprising:

selecting a first nucleotide sequence in a first oligonucleotide strand;

selecting a second nucleotide sequence in a second oligonucleotide strand;

incorporating an abasic nucleotide within said first nucleotide sequence;

attaching a bond precursor to a nucleotide of said second nucleotide sequence; and

forming a covalent cross-linkage between said abasic site and said bond precursor.

64. The method of claim 63 including selecting said first nucleotide sequence to be complementary to and specifically hybridizable with said second nucleotide sequence.

65. The method of claim 63 including attaching said sequence bond precursor to the sugar moiety of said second sequence nucleotide.

66. The method of claim 65 including attaching said bond precursor to a 2'-position of said second nucleotide.

67. The method of claim 63 including:

forming an aldehydic functional group at said abasic site; and

selecting said bond precursor to include a functional group capable of forming one of an oxime, imine, hydrazone, semicarbazone, thiosemicarbazone, hydrazide-hydrazone, hemiacetal or thiohemiacetal linkage with said abasic site aldehydic functional group.

68. A method of cross-linking first and second strands of oligonucleotides comprising:

selecting a first nucleotide sequence in a first oligonucleotide strand;

selecting a second nucleotide sequence in a second oligonucleotide strand;

attaching a first bond precursor to the sugar moiety of a first nucleotide that is covalently bound to at least two further nucleotides in said first nucleotide sequence;

attaching a second bond precursor to the sugar moiety of a second nucleotide that is covalently bound to at least two further nucleotides in said second nucleotide sequence; and

forming a covalent cross-linkage between said first and said second bond precursors.

69. The method of claim 68 including selecting said first nucleotide sequence to be complementary to and specifically hybridizable with said second nucleotide sequence.

70. The method of claim 68 including choosing said covalent cross-linkage from imine, amine, aminoalkylamine, oxime, hydroxylamine, hydrazine, hydrazone, azine, hydrazide-hydrazone, amide, hydrazide, semicarbazide, semicarbazone, thiosemicarbazide, thiocarbazone, disulfide, hemiacetal or thiohemiacetal linkages.

STATEMENT UNDER ARTICLE 19

Further to the International Search Report mailed April 28, 1993 in connection with this patent application, Applicants have amended the claims to specify nucleotides covalent bonding at the site of a cross-linkage. The claims also have been amended to more clearly define the invention with respect to the role of space-spanning groups in the formation of covalent cross-linkages. These amendments find support in the specification at, for example, page 6, lines 32-36, page 7, line 36 - page 8, line 14, and page 16, line 18 - page 17, line 3.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US93/02059

A. CLASSIFICATION OF SUBJECT MATTER

IPC(5) :C07H 21/04
US CL :536/23.1, 24.3, 25.3

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 536/23.1, 24.3, 25.3

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS, Medline, Biosis

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	J. Am. Chem. Soc., Vol. 111, issued 1989, Meyer et al., "Efficient, Specific Cross-Linking and Cleavage of DNA by Stable, Synthetic Complementary Oligodeoxynucleotides", pp. 8517-8519, see entire document.	1-70
Y	Biochemistry, Vol. 27, No. 13, issued 1988, Farahani et al., "Hydroxyl Radical Induced Cross-Linking Between Phenylalanine and 2-Deoxyribose", pp. 4695-4698, see entire document.	1-70

☐ 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 part of particular relevance

"E" earlier document 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

20 April 1993

Date of mailing of the international search report

28APR 1993

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231

Facsimile No. NOT APPLICABLE

Authorized officer

KENNETH R. HORLICK

Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)★

Espacenet

Bibliographic data: WO2008115185 (A2) — 2008-09-25

USE OF MICROARRAYS FOR GENOMIC REPRESENTATION SELECTION

Inventor(s): ALBERT THOMAS J [US] ± (ALBERT, THOMAS, J)

Applicant(s): NIMBLEGEN SYSTEMS INC [US]; ALBERT THOMAS J [US] ± (NIMBLEGEN SYSTEMS, INC, ; ALBERT, THOMAS, J)

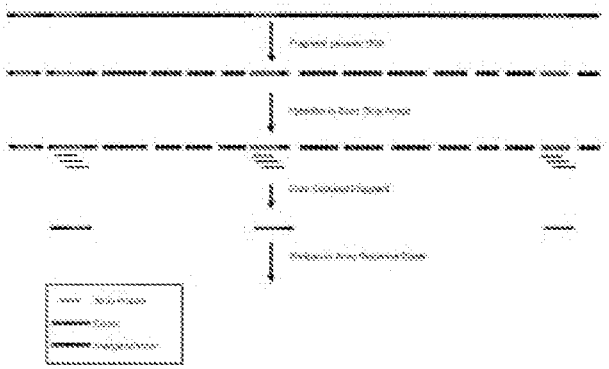
Classification: **- international:**
- cooperative: **C12N15/1093**; **C12Q1/6806** more

Application number: WO2007US10064 20070424 Global Dossier

Priority number (s): US20060794560P 20060424 ; US20060832719P 20060721

Also published as: WO2008115185 (A3) EP2010657 (A2) US2008194413 (A1) US2011184161 (A1)

Abstract of WO2008115185 (A2)



The present invention provides novel methods for reducing the complexity of a genomic sample for further analysis such as direct DNA sequencing, resequencing or SNP calling. The methods use pre-selected immobilized oligonucleotide probes to capture target nucleic acid molecules from a sample containing denatured, fragmented genomic nucleic acid. The disclosed method provides for cost-effective, flexible and rapid enrichment of target nucleic acid from complex biological samples.

IDT Exhibit 1020.1439

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
25 September 2008 (25.09.2008)

PCT

(10) International Publication Number
WO 2008/115185 A2

(51) International Patent Classification: Not classified

(21) International Application Number:
PCT/US2007/010064

(22) International Filing Date: 24 April 2007 (24.04.2007)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/794,560 24 April 2006 (24.04.2006) US
60/832,719 21 July 2006 (21.07.2006) US

(71) Applicant *(for all designated States except US)*: **NIMBLEGEN SYSTEMS, INC.** [US/US]; 1 Science Court, Madison, WI 53711 (US).

(72) Inventor; and

(75) Inventor/Applicant *(for US only)*: **ALBERT, Thomas, J.** [US/US]; 2738 Tower Hill Drive, Fitchburg, WI 53711 (US).

(74) Agent: **BERSON, Bennett, J.**; Quarles & Brady LLP, P. O. Box 2113, Madison, WI 53701-2113 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BH, 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, 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, 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, HU, IE, IS, IT, LT, LU, LV, MC, MT, 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:
— *without international search report and to be republished upon receipt of that report*

(54) **Title:** USE OF MICROARRAYS FOR GENOMIC REPRESENTATION SELECTION

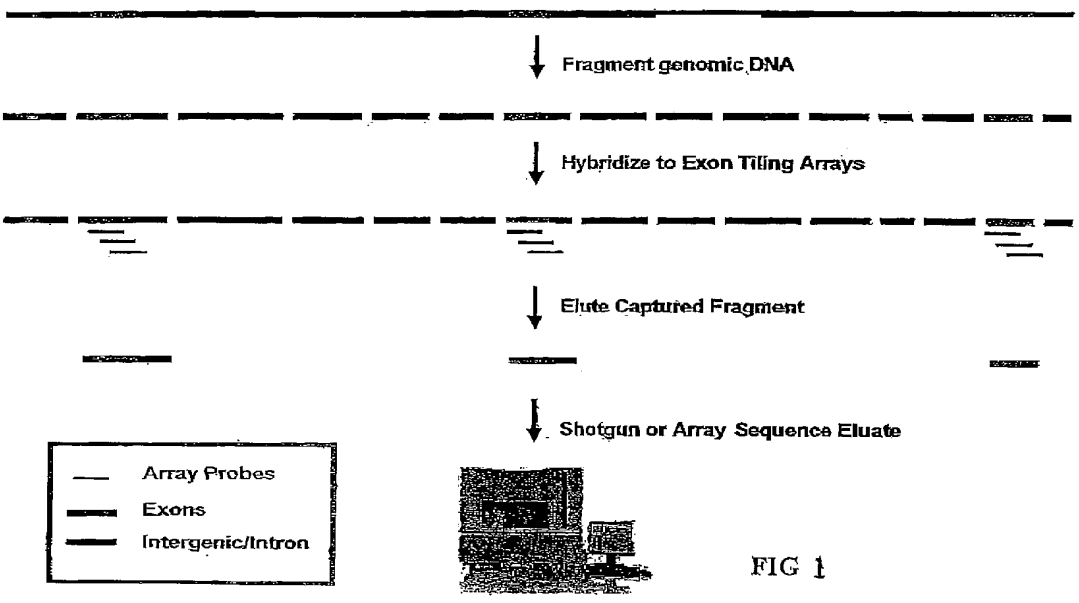


FIG 1

(57) **Abstract:** The present invention provides novel methods for reducing the complexity of a genomic sample for further analysis such as direct DNA sequencing, resequencing or SNP calling. The methods use pre-selected immobilized oligonucleotide probes to capture target nucleic acid molecules from a sample containing denatured, fragmented genomic nucleic acid. The disclosed method provides for cost-effective, flexible and rapid enrichment of target nucleic acid from complex biological samples.

WO 2008/115185 A2

USE OF MICROARRAYS FOR GENOMIC REPRESENTATION SELECTION

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims priority to co-pending U.S. Provisional Application No. 60/794,560, filed April 24, 2006, and U.S. Provisional Application No. 60/832,719, filed July 21, 2006, the disclosure of each of which is incorporated herein by reference in its entirety as if set forth herein.

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH OR DEVELOPMENT

[0002] Not Applicable

BACKGROUND OF THE INVENTION

[0003] The advent of DNA microarray technology makes it possible to build an array of millions of DNA sequences in a very small area, such as the size of a microscope slide. See, e.g., U. S. Patent No. 6, 375,903 and U.S. Pat. No. 5,143,854, each of which is hereby incorporated by reference in its entirety. The disclosure of U.S. Pat. No. 6,375,903 enables the construction of so-called maskless array synthesizer (MAS) instruments in which light is used to direct synthesis of the DNA sequences, the light direction being performed using a digital micromirror device (DMD). Using an MAS instrument, the selection of DNA sequences to be constructed in the microarray is under software control so that individually customized arrays can be built to order. In general, MAS-based DNA microarray synthesis technology allows for the parallel synthesis of over 4 million 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.

[0004] 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 DNA isolated from such organisms. DNA microarray technology has also been applied to many areas such as gene expression and discovery, mutation detection, allelic and evolutionary sequence comparison, genome mapping and more.

[0005] Many applications require searching for genetic variants and mutations across the entire human genome that underlie human diseases. In the case of complex diseases, these searches generally result in a single nucleotide polymorphism (SNP) or set of SNPs associated with 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

IDT Exhibit 1020.1441

(Kb) from affected individuals or tissue samples is frequently required to find a single base change or identify all sequence variants. Accordingly, the genome is typically too complex to be studied as a whole, and techniques must be used to reduce the complexity of the genome.

[0006] Therefore, alternative cost-effective and rapid methods for reducing the complexity of a genomic sample in a user defined way to allow for further processing and analysis would be a desirable contribution to the art.

BRIEF SUMMARY OF THE INVENTION

[0007] The present invention is summarized as a novel method for reducing the complexity of a genomic sample to facilitate further processing and genetic analysis. The method uses pre-selected immobilized nucleic acid probes to capture target nucleic acid sequences from a genomic sample by hybridizing the sample to the probes on a substrate. The captured target genomic nucleic acids are washed and then eluted off of the substrate. The eluted genomic sequences are more amenable to detailed genetic analysis than a genomic sample that has not been subjected to this procedure. Accordingly, the disclosed method provides a cost-effective, flexible and efficient approach for reducing the complexity of a genomic sample.

[0008] In one aspect, the method provides a microarray having pre-selected array-immobilized nucleic acid probes to capture target nucleic acid sequences from a genomic sample. This may be accomplished by hybridizing the genomic sample of target nucleic acid sequence(s) against a microarray having array-immobilized nucleic acid probes directed to a specific region of the genome. After hybridization, target nucleic acid sequences present in the sample may be enriched by washing the array and eluting the hybridized genomic nucleic acids from the array. The target nucleic acid sequence(s), preferably DNA may be amplified using for example, non-specific ligation mediated PCR, resulting in an amplified pool of PCR products of reduced complexity compared to the original genomic sample.

[0009] In a related aspect, the invention provides a method of reducing the complexity of a genomic sample by hybridizing the sample against a microarray having array-immobilized pre-selected target nucleic acid probes under preferably stringent conditions sufficient to support hybridization between the array-immobilized probes and complementary regions of the genomic sample. The microarray is subsequently washed under conditions sufficient to remove non-specifically bound nucleic acids and the hybridized target genomic nucleic acid sequences are eluted from the microarray. The eluted target sequences may optionally be amplified.

IDT Exhibit 1020.1442

[00010] In another aspect, 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.

[00011] In another aspect, the 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 to capture complementary target nucleic acid sequences contained in a complex genomic sample.

[00012] Other objects, advantages and features of the present invention will become apparent from the following specification taken in conjunction with the accompanying drawings.

BRIEF DESCRIPTION OF THE SEVERAL VIEWS OF THE DRAWINGS

[00013] FIG. 1 is a graphic depiction of flow diagram of the direct genomic selection process using a microarray.

[00014] FIG. 2 is a flow diagram of the direct genomic selection process using a microarray.

[00015] FIG. 3 diagrams a method for 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.

DETAILED DESCRIPTION OF THE INVENTION

[00016] The present invention broadly relates to cost-effective, flexible and rapid methods for reducing genomic 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. Thus, the invention provides for method for enrichment of target nucleic acid in a complex genomic sample.

[00017] In one embodiment, a sample containing fragmented, denatured genomic nucleic acid molecules is exposed under hybridizing conditions to at least one oligonucleotide probe, and more typically a plurality of oligonucleotide probes, immobilized on a substrate to capture from the sample target nucleic acid molecules that hybridize to the immobilized probes. Non-hybridizing regions of the genome remain in solution. The probes correspond in sequence to at least one region of the genome and can be provided on a substrate in parallel using maskless array synthesis technology. Alternatively, probes can be obtained serially using a standard DNA

IDT Exhibit 1020.1443

synthesizer and then applied to the substrate. After the hybridization, nucleic acids that do not hybridize, or that hybridize non-specifically to the probes are separated from the substrate-bound probes by washing. The remaining nucleic acids, bound specifically to the probes, are eluted from the substrate in heated water or in a nucleic acid elution buffer to yield an eluate enriched for the target nucleic acid molecules.

[00018] In some embodiments, double-stranded linkers are provided at the termini of the fragmented genomic nucleic acid before the fragments are denatured and hybridized to 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 ligation mediated 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.

[00019] Alternatively, as shown in Fig. 3, nucleic acid probes for target molecules can be synthesized on a substrate, released from the substrate 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 substrate. The probes can be designed for convenient release from the substrate by providing, e.g., at or near the substrate-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.

[00020] 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 technology is much more flexible than other methods currently available in the field of genomic enrichment, such as

IDT Exhibit 1020.1444

use of BAC (bacterial artificial chromosome) for direct genomic selection (see Lovett et al. (1991) *PNAS USA* 88, 9628-9632).

[00021] 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 such as individual chromosomes or exons for medical sequencing purposes.

[00022] 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.

[00023] 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 substrate 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 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.

IDT Exhibit 1020.1445

[00024] The nature and performance of the probes can be varied to advantageously adjust the distribution of the target molecules captured and enriched in accord with the methods. For example, the number of sequencing 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. Publication 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, 45mer - 85mer but optionally more than 100 nt) to equalize the melting temperature (Tm = 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.

[00025] 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 substrate 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 substrate, the concentration of individual probes in the pool can be varied in the same way.

[00026] 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

IDT Exhibit 1020.1446

are provided on the substrate, saturation of the probes will ensure that substantially equal amounts of each target are eluted after the second hybridization and washing.

[00027] Another normalizing strategy follows the elution and amplification of captured target molecules from the substrate. Target molecules in the eluate are denatured 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.

[00028] 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.

[00029] 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. However, the combination of parameters is more important than the measure of any single parameter. See, e.g., Wetmur et al., *J. Mol. Biol.* 31:349-70 (1966), and Wetmur, Critical Reviews in Biochemistry and *Molecular Biology* 26(34):227-59 (1991).

IDT Exhibit 1020.1447

[00030] "Stringent conditions" or "high stringency conditions," as defined herein, can be hybridization in 50% formamide, 5XSSC (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.

[00031] 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* (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).

[00032] The following examples are provided as further non-limiting illustrations of particular embodiments of the invention.

EXAMPLES

Example 1

[00033] This example describes modifications to direct selection that allow for rapid and efficient discovery of new polymorphisms and mutations in large genomic regions. Microarrays having immobilized probes were used in one- or multiple rounds of hybridization selection with a target of total genomic DNA, and the selected sequences were amplified by the polymerase chain reaction (PCR) (see Figs. 1 and 2).

[00034] **Preparation of the genomic DNA and double-stranded linkers**

[00035] DNA was fragmented using sonication to an average size of ~500 base pairs.

[00036] A reaction to polish the ends of the sonicated DNA fragments was set up:

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

IDT Exhibit 1020.1448

[00037] The reaction was incubated at 11°C for 30 min. The reaction was subjected to phenol/chloroform extraction procedures and the DNA was recovered by ethanol precipitation. The precipitated pellet was dissolved in 10 µl water (to give a final concentration of 2 µg/µl).

[00038] The oligonucleotides were annealed to create a double-stranded linker, by mixing the following:

Oligonucleotide 1 (1µg/µl)	22.5 µl
Oligonucleotide 2 (1µg/µl)	22.5 µl
10x annealing buffer	5 µl
Water	to 50 µl

[00039] The reaction was heated at 65°C for 10 min; then allowed to cool at 15–25°C for 2 h. The oligonucleotides in this protocol were designated 1 and 2.

[00040] The double-stranded linker was purified by column chromatography through a Sephadex G-50 spin column. The purified linker solution was concentrated by lyophilization to a concentration of 2 µg/µl.

IDT Exhibit 1020.1449

[00041] **Ligation of linkers to genomic DNA fragments**

[00042] The following reaction to ligate the linkers to genomic DNA fragments was set up. The reaction was incubated at 14°C overnight.

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

[00043] The reaction volume was adjusted to 500 μl with water and purify the ligated genomic DNA using a QIAquick PCR purification kit. The purified DNA was stored at a concentration of 1 μg/μl.

[00044] **Primary selection and capture of hybrids**

[00045] To prepare the genomic DNA sample for hybridization to the microarray, linkered genomic DNA (10μg) was resuspended in 3.5μl of nuclease-free water and combined with 31.5μl NimbleGen Hybridization Buffer (NimbleGen Systems Inc., Madison, WI), 9μl Hybridization Additive (NimbleGen Systems), in a final volume of 45μl. The samples were heat-denatured at 95°C for 5 minutes and transferred to a 42°C heat block.

[00046] To capture the target genomic DNA on the microarray, samples were hybridized to NimbleGen CGH arrays, manufactured as described, using 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. (Singh-Gasson, S., et al., (1999), Maskless fabrication of light-directed oligonucleotides microarrays using a digital micromirror array. Nat. Biotechnol. 17:974-978; and Nuwaysir, E.F., et al., (2000), Gene expression analysis using oligonucleotide arrays produced by maskless photolithography, Genome Res. 12:1749-1755, incorporated by reference herein in its entirety). Following hybridization, arrays were 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 were 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, NimbleGen Systems).

[00047] To elute the genomic DNA hybridized to the microarray, the arrays were incubated twice for 5 minutes in 95°C water. The eluted DNA was dried down using vacuum centrifugation.

IDT Exhibit 1020.1450

[00048] **Amplification of the primary selected DNA**

[00049] The primary selected genomic DNA was amplified as described below. Ten separate replicate amplification reactions were set up in 200µl PCR tubes:

Reagents	
Template: primary selection material	5 µl
Oligonucleotide 3 (200 ng/µl)	1 µl
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

[00050] The reactions were 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

[00051] The reaction products were analyzed by agarose gel electrophoresis. The amplification products were purified using a QIAquick PCR purification kit. The eluted samples were pooled and the concentration of amplified primary selected DNA was determined by spectrophotometry. A volume of DNA in the pool equivalent to 1 µg was reduced to 5 µl in a speed vacuum concentrator. 1 µl (at least 200 ng) of the primary selected material was set aside for comparison with the secondary selection products. As necessary, subsequent rounds of enrichment were performed by further rounds of array hybridization and amplification of the eluted sample.

IDT Exhibit 1020.1451

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

[00052] Probes were synthesized on a microarray, then were released using a base-labile Fmoc (9-fluorenylmethyloxycarbonyl) group. The probes were then immobilized onto the surface of a solid support using known methods for covalent or non-covalent attachment. Optionally, prior to immobilization onto the solid support, the synthesized probes were amplified using ligation mediated PCR, Phi29 or other amplification strategy to increase the amount of the synthesized probes. This material can now be used for direct sequencing, array based resequencing, genotyping, or any other genetic analysis targeting the enriched region of the genome.

[00053] It is understood that certain adaptations of the invention described in this disclosure are a matter of routine optimization for those skilled in the art, and can be implemented without departing from the spirit of the invention, or the scope of the appended claims.

IDT Exhibit 1020.1452

CLAIMS

WE CLAIM:

1. A method of reducing the genetic complexity of a population of genomic nucleic acid molecules, the method comprising the steps of:

exposing a sample that comprises fragmented, denatured genomic nucleic acid molecules to at least one oligonucleotide probe immobilized on a substrate under hybridizing conditions to capture target nucleic acid molecules that hybridize to the at least one probe;

separating unbound and non-specifically bound nucleic acids from the captured molecules; and

eluting the captured molecules from the substrate in an eluate pool having reduced genetic complexity relative to the sample.

2. The method of Claim 1 further comprising the step of amplifying the eluted target nucleic acids.

3. The method of Claim 1 wherein the probe is immobilized on the substrate by synthesizing the probe on the substrate.

4. The method of Claim 3 wherein the substrate is a microarray.

5. The method of Claim 4 wherein the probes are synthesized on the microarray using a maskless array synthesizer.

6. The method of Claim 1 wherein the probe is immobilized on the substrate by synthesizing the probe and then applying the probe to the substrate.

7. The method of Claim 6 wherein the substrate is a microarray.

8. The method of Claim 6 wherein the substrate is selected from the group consisting of glass, metal, ceramic and polymeric beads.

9. The method of Claim 8 wherein the probe is immobilized on the substrate by synthesizing the probe on a microarray, releasing the probe from the microarray and immobilizing the released probe on the substrate.

IDT Exhibit 1020.1453

10. The method of Claim 1 wherein the sample is exposed to a plurality of immobilized probes on the substrate.

11. The method of Claim 1 wherein the plurality of immobilized probes are characterized by normalized capture performance.

12. The method of Claim 11 wherein the normalized capture performance is achieved by a method comprising the steps of:

ascertaining the capture fitness of probes in the probe set; and

adjusting the quantity of at least one probe on the substrate.

13. The method of Claim 11 wherein the normalized capture performance is achieved by a method comprising the steps of:

ascertaining the capture fitness of probes in the probe set; and

adjusting at least one of the sequence, the melting temperature and the probe length of at least one probe on the substrate.

14. The method of Claim 11 wherein the normalized capture performance is achieved by a method comprising the steps of:

exposing the eluted target nucleic acids to the at least one immobilized probe on the substrate under less stringent conditions than in the first exposing step such that the at least one probe is saturated;

washing unbound and non-specifically bound nucleic acids from the substrate; and

eluting the bound target nucleic acids from the substrate.

15. The method of Claim 11 wherein the normalized capture performance is achieved by a method comprising the steps of:

denaturing the eluted target nucleic acids to a single-stranded state;

re-annealing the single-stranded target nucleic acids until a portion of the target nucleic acids are double-stranded; and

discarding the double-stranded target nucleic acids and retaining the single stranded target nucleic acids.

IDT Exhibit 1020.1454

16. The method of Claim 1 wherein the at least one immobilized probe hybridizes to a genomic region of interest on nucleic acid fragments in the sample.

17. The method of Claim 1 wherein the at least one immobilized probe hybridizes to sequences on target nucleic acid fragments comprising a genomic region of interest, the hybridizing sequences being separate from the genomic region of interest.

18. The method of Claim 1 further comprising the step of performing 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.

IDT Exhibit 1020.1455

Title: USE OF MICROARRAYS FOR GENOMIC REPRESENTATION SELECTION
Inventor: Thomas J. Albert
Application No.:
Docket Number: 700706.00027

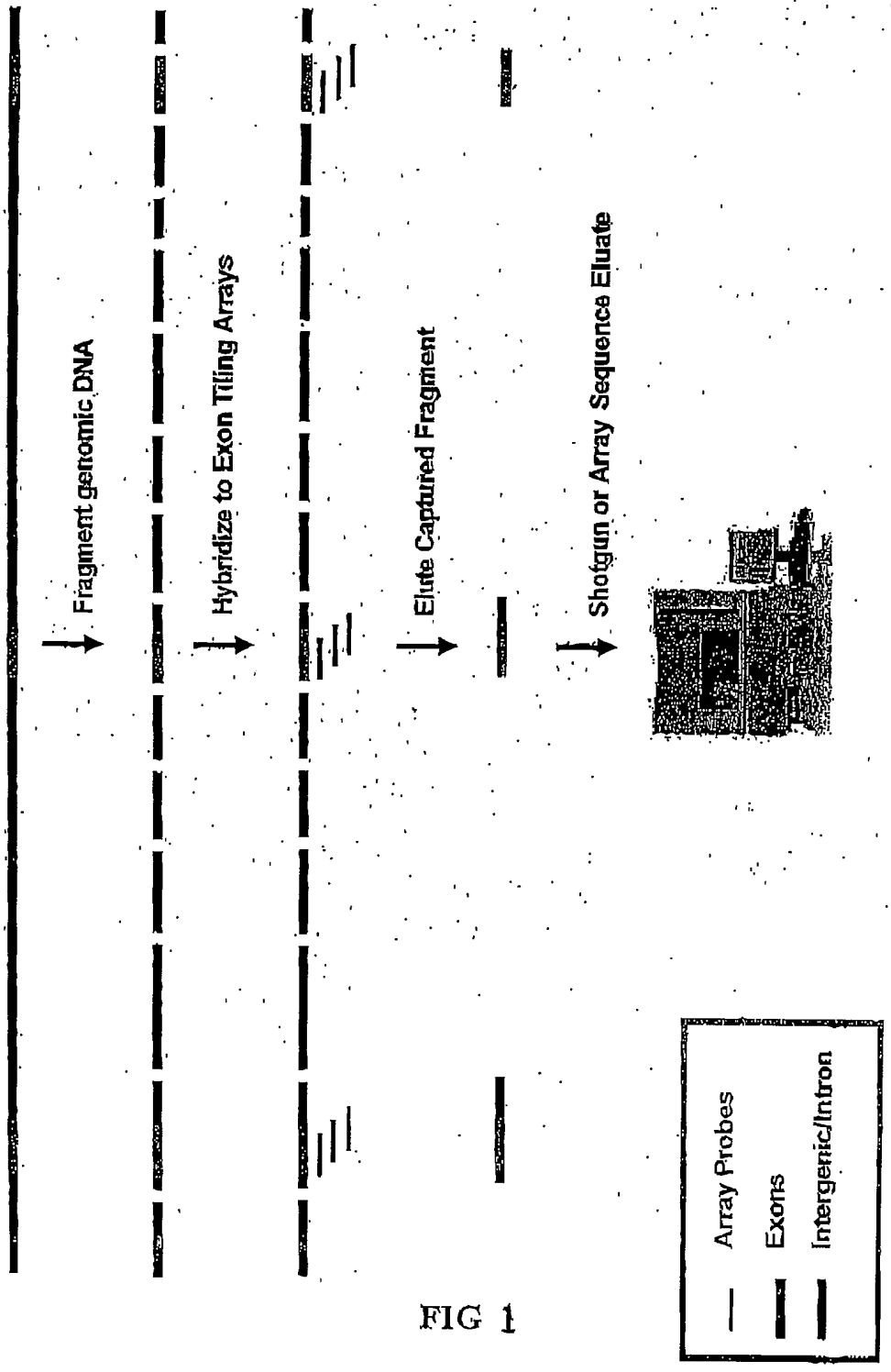


FIG 1

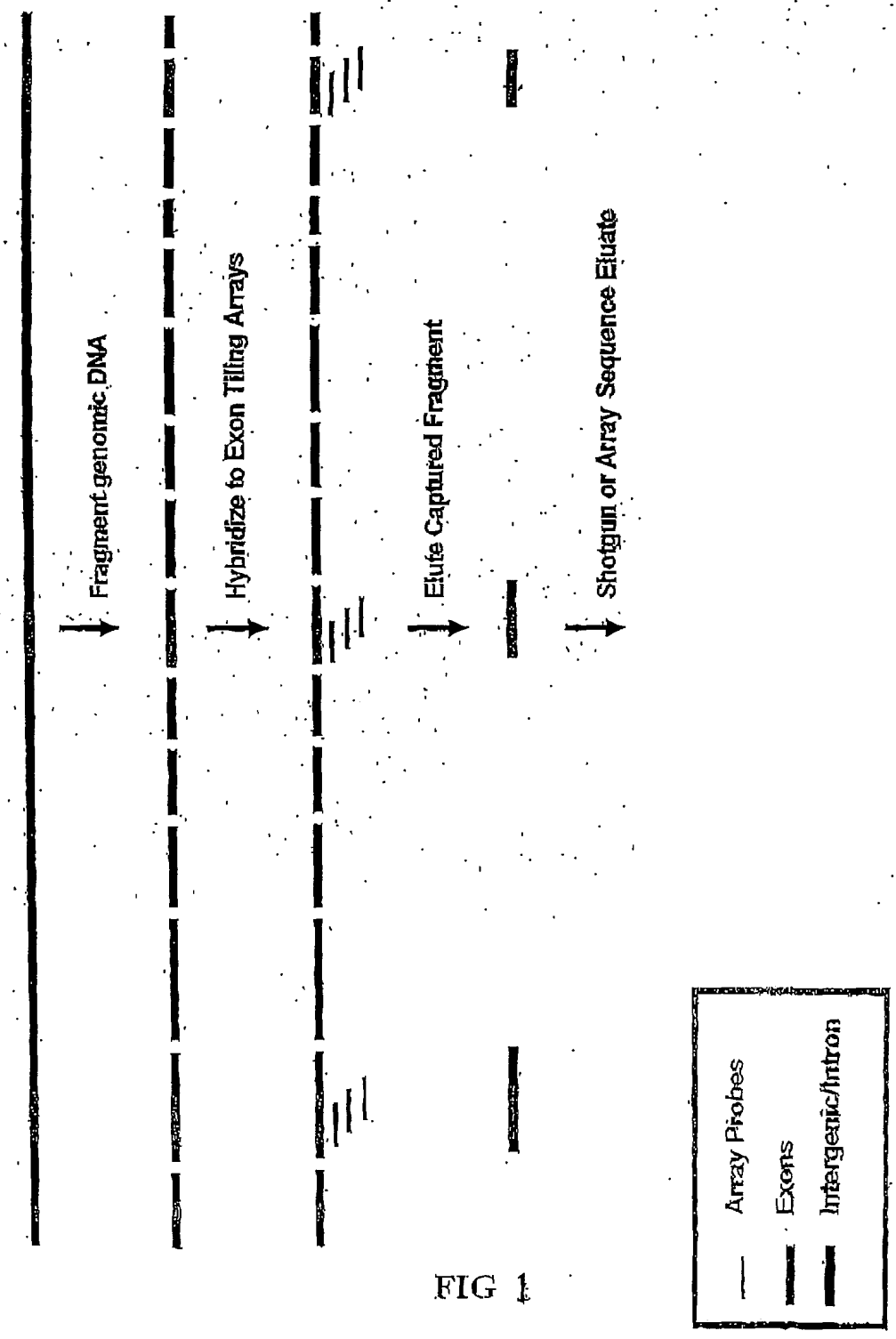

Fragment genomic DNA
Hybridize to Exon Tiling Arrays
Elute Captured Fragment
Shotgun or Array Sequence Eluate
Array Probes
Exons
Intergenic/Intron

FIG 1

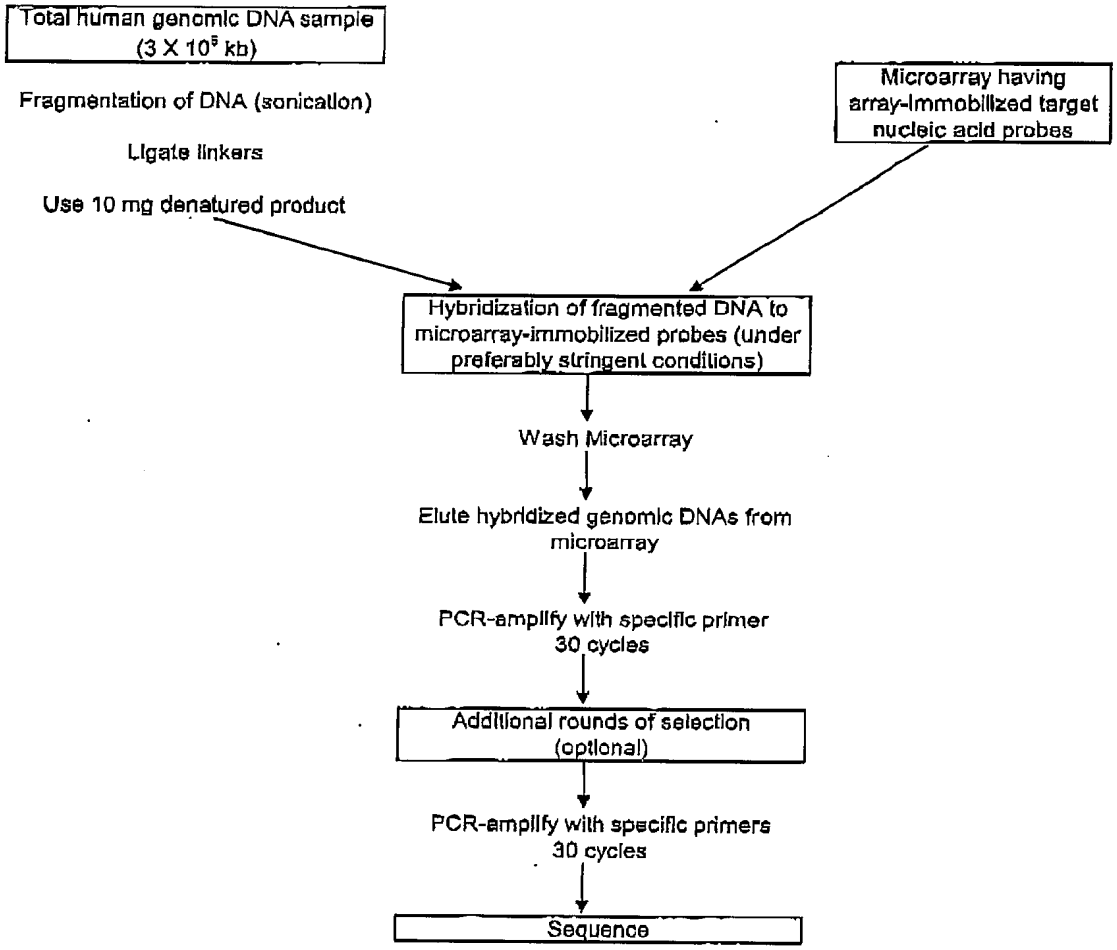

Total human genomic DNA sample (3 X 10⁵ kb)
Fragmentation of DNA (sonication)
Ligate linkers
Use 10 mg denatured product
Microarray having array-immobilized target nucleic acid probes
Hybridization of fragmented DNA to microarray-immobilized probes (under preferably stringent conditions)
Wash Microarray
Elute hybridized genomic DNAs from microarray
PCR-amplify with specific primer 30 cycles
Additional rounds of selection (optional)
PCR-amplify with specific primers 30 cycles
Sequence

FIG 2

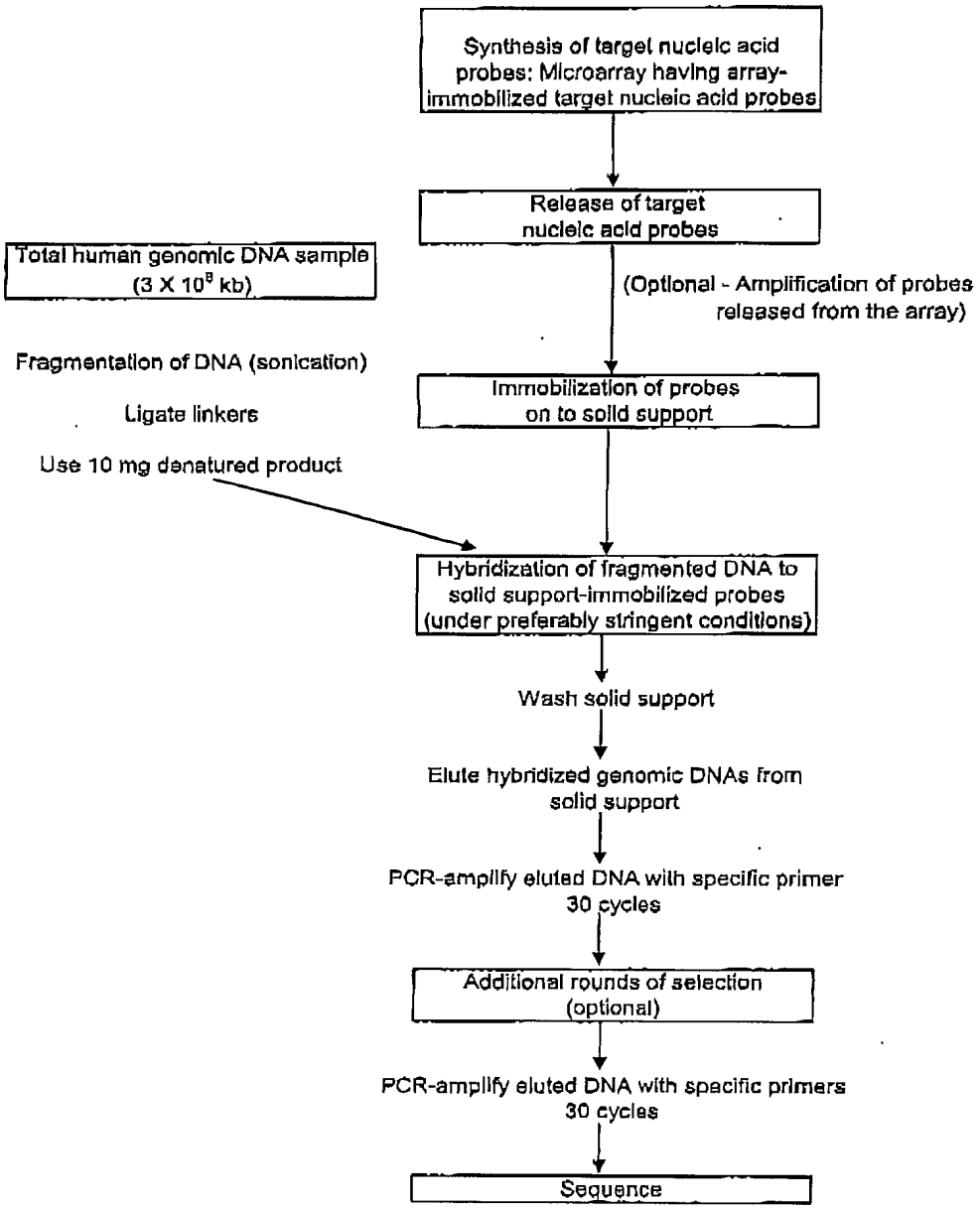

Synthesis of target nucleic acid probes: Microarray having array-immobilized target nucleic acid probes
Release of target nucleic acid probes
Total human genomic DNA sample (3 X 10^9 kb)
(Optional - Amplification of probes released from the array)
Fragmentation of DNA (sonication)
Immobilization of probes on to solid support
Ligate linkers
Use 10 mg denatured product
Hybridization of fragmented DNA to solid support-immobilized probes (under preferably stringent conditions)
Wash solid support
Elute hybridized genomic DNAs from solid support
PCR-amplify eluted DNA with specific primer 30 cycles
Additional rounds of selection (optional)
PCR-amplify eluted DNA with specific primers 30 cycles
Sequence

FIG 3

Espacenet

Bibliographic data: WO2009120374 (A2) — 2009-10-01

METHODS AND COMPOSITIONS FOR NUCLEIC ACID SAMPLE PREPARATION

Inventor(s): PATEL PRANAV [US]; BJORNSON KEITH [US]; TRAVERS KEVIN [US]; HEINER CHERYL [US] ± (PATEL, PRANAV, ; BJORNSON, KEITH, ; TRAVERS, KEVIN, ; HEINER, CHERYL)

Applicant(s): PACIFIC BIOSCIENCES CALIFORNIA [US]; PATEL PRANAV [US]; BJORNSON KEITH [US]; TRAVERS KEVIN [US]; HEINER CHERYL [US] ± (PACIFIC BIOSCIENCES OF CALIFORNIA, INC, ; PATEL, PRANAV, ; BJORNSON, KEITH, ; TRAVERS, KEVIN, ; HEINER, CHERYL)

Classification: **- international:** ***C07H21/00; C12N15/10; C12N15/11; C12Q1/68***
- cooperative: **C12N15/10**; **C12N15/65**; **C12N15/66**

Application number: WO2009US01930 20090327 Global Dossier

Priority number(s): US20080072160P 20080328

Also published as: WO2009120374 (A3) US2009280538 (A1) US2012309650 (A1) US8236499 (B2) US8535886 (B2)

Abstract of WO2009120374 (A2)

Provided are methods and compositions for the production of linear single- stranded nucleic acids, which can be used as templates in high-throughput sequencing systems. Also provided are methods and compositions for the production of closed single-stranded nucleic acid loops, which can be used as templates in high-throughput sequencing systems.

IDT Exhibit 1020.1460

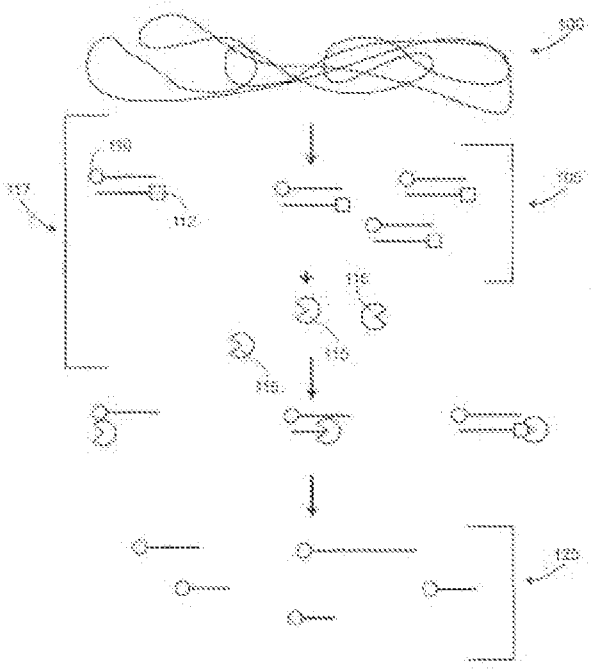


Fig. 1

IDT Exhibit 1020.1461

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
1 October 2009 (01.10.2009)

PCT

(10) International Publication Number
WO 2009/120374 A2

(51) International Patent Classification:
C12N 15/10 (2006.01) *C12N 15/11* (2006.01)
C12Q 1/68 (2006.01) *C07H 21/00* (2006.01)

(21) International Application Number:
PCT/US2009/001930

(22) International Filing Date:
27 March 2009 (27.03.2009)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/072,160 28 March 2008 (28.03.2008) US

(71) Applicant *(for all designated States except US)*: **PACIFIC BIOSCIENCES OF CALIFORNIA, INC.** [US/US]; 1505 Adams Drive, Menlo Park, CA 94025 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **PATEL, Pranav** [IN/US]; 34203 O'neil Terrace, Fremont, CA 94555 (US). **BJORNSON, Keith** [US/US]; 8098 Peachtree Avenue, Newark, CA 94560 (US). **TRAVERS, Kevin** [US/US]; 3385 Gonzaga Place, Santa Clara, CA 95051 (US). **HEINER, Cheryl** [US/US]; 1952 Ticonderoga Drive, San Mateo, CA 94402 (US).

(74) Agents: **QUINE, Jonathan Alan** et al.; Quine Intellectual Property, Law Group, P.c., P.o. Box 458, Alameda, CA 94501 (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, 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,

[Continued on next page]

(54) Title: METHODS AND COMPOSITIONS FOR NUCLEIC ACID SAMPLE PREPARATION

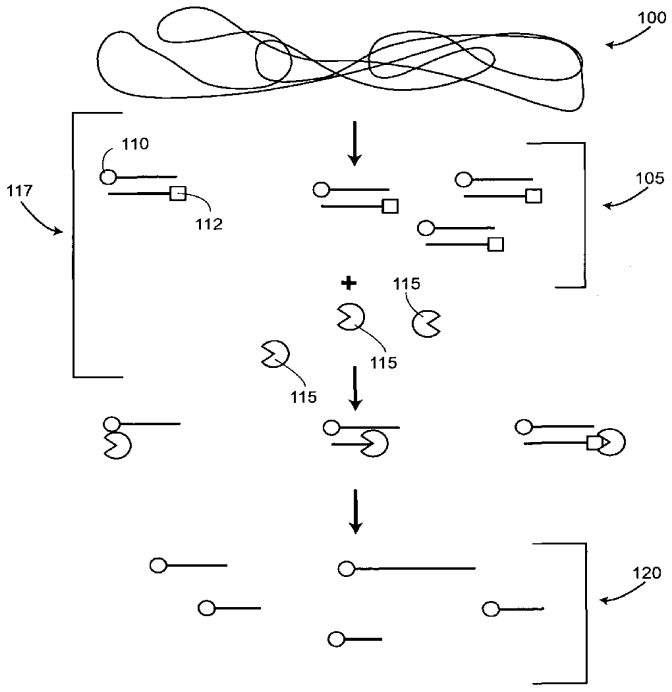


Fig. 1

(57) Abstract: Provided are methods and compositions for the production of linear single- stranded nucleic acids, which can be used as templates in high-throughput sequencing systems. Also provided are methods and compositions for the production of closed single- stranded nucleic acid loops, which can be used as templates in high-throughput sequencing systems.

WO 2009/120374 A2

IDT Exhibit 1020.1462

GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK, 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))*

METHODS AND COMPOSITIONS FOR NUCLEIC ACID SAMPLE PREPARATION

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims priority to and benefit of U.S. Provisional Patent Application 61/072,160, entitled, "Methods, Compositions, and Systems for Nucleic Acid Sample Preparation," by Patel, Bjornson, Travers, and Heiner, filed March 28, 2008, the disclosure of which is incorporated herein in its entirety for all purposes.

[0002] This application is also related to U.S. Patent Application No. [unassigned], filed March 27, 2009 (Attorney Docket No. 01-0005903) and U.S. Patent Application No. [unassigned], filed March 27, 2009 (Attorney Docket No. 01-0007701), all of which are incorporated herein by reference in their entireties for all purposes.

FIELD OF THE INVENTION

[0003] This invention is in the field of nucleic acid sequencing, particularly the preparation of templates for nucleic acid sequencing.

BACKGROUND OF THE INVENTION

[0004] Nucleic acid sequence data is valuable in myriad applications in biological research and molecular medicine, including determining the hereditary factors in disease, in developing new methods to detect disease and guide therapy (van de Vijver et al. (2002) "A gene-expression signature as a predictor of survival in breast cancer," *New England Journal of Medicine* **347:** 1999-2009), and in providing a rational basis for personalized medicine. Obtaining and verifying sequence data for use in such analyses has made it necessary for sequencing technologies to undergo advancements to expand throughput, lower reagent and labor costs and improve accuracy (*See*, e.g., Chan, et al. (2005) "Advances in Sequencing Technology" (Review) *Mutation Research* **573:** 13-40, Levene et al. (2003) "Zero Mode Waveguides for Single Molecule Analysis at High Concentrations," *Science* **299:** 682-686).

[0005] Methods of preparing templates for large-scale sequencing projects have typically entailed constructing shotgun libraries that comprise overlapping fragments of, e.g., a genomic DNA; transforming cells with the library; growing cells to amplify each library member; and isolating and purifying library DNA. For example, shotgun cloning was initially used to prepare nucleic acid templates for sequencing small genomes such as

IDT Exhibit 1020.1464

that of the cauliflower mosaic virus (CMV) (Gardner, et al. (1981) "The complete nucleotide sequence of an infectious clone of cauliflower mosaic virus by M13mp7 shotgun sequencing." *NAR* 9: 2871-2888). More recently, this template preparation strategy has been used to produce templates for the sequencing of complex genomes, including the mouse, *Drosophila*, and human genomes (Mural, et al. (2001) "A comparison of whole-genome shotgun-derived mouse chromosome 16 and the human genome." *Science* **296:** 1661-1671; Adams, et al. (2000) "The genome sequence of *Drosophila melanogaster.*" *Science* **287:** 2185-95; Venter, et al. (2001) "The Sequence of the Human Genome." *Science* **291:** 1304-1351).

[0006] However, the cloning and cell culture techniques used in shotgun library construction methods are time consuming, labor-intensive, costly, and not easily amenable to automation. Purification protocols that are used to isolate cloned nucleic acid templates from cells, e.g., bacterial cells, for sequencing do not reliably produce nucleic acid samples that are sufficiently free of sequencing reaction inhibitors such as salts, carbohydrates and/or proteins. Furthermore, these disadvantages are exacerbated when conventional template preparation methods are scaled to the quantities that would be useful for high throughput sequencing technologies, e.g., single-molecule real-time (SMRT) sequencing systems, such as those described in, e.g., Levene et al. (2003) "Zero Mode Waveguides for single Molecule Analysis at High Concentrations," *Science* **299:** 682-686; and Eid, et al. (2009) "Real-Time DNA Sequencing from Single Polymerase Molecules." *Science* **323:** 133-138.

[0007] Consequently, there is an increasing demand for efficient, low-cost methods for the preparation of high-quality nucleic acid templates for next generation sequencing technologies. The present invention provides methods and compositions that would be useful for supplying high throughput DNA sequencing systems with such templates.

SUMMARY

[0008] The present invention provides methods and compositions that can be useful for supplying high throughput DNA sequencing systems with nucleic acid templates. The methods circumvent the need for expensive, labor-intensive library construction and cell culture methods and can be scaled to accommodate template production for a variety of high-throughput sequencing applications, including, e.g., sequencing large genomes, gene expression profiling (Spinella, et al. (1999) "Tandem arrayed ligation of expressed sequence

IDT Exhibit 1020.1465

tags (TALEST): a new method for generating global gene expression profiles." *Nucleic Acids Res* **27:** e22, Velculescu, et al. (1995) "Serial analysis of gene expression." *Science* **270:** 484–487), genome-wide methylation analysis, and microbiome analysis. The methods and compositions provided by the invention can be used to produce either linear or single-stranded nucleic acid loops.

[0009] Thus in a first aspect, the invention provides methods of producing a population of single-stranded nucleic acids. In the first set of methods, a genomic DNA, a cDNA, or a DNA concatamer is provided, and double-stranded fragments that comprise first exonuclease-sensitive strands and second exonuclease-resistant strands are generated from the genomic DNA, the cDNA, or the concatamer. The methods include eliminating the exonuclease-sensitive strands of the double-stranded fragment to produce the population of single-stranded nucleic acids. The genomic DNA, the cDNA, or the concatamer provided for any of the methods described herein can be derived from any of a variety of sources, e.g., a eukaryote, a prokaryote, an archaea, a virus, a phage, etc.

[0010] In certain embodiments, generating double-stranded fragments that comprise first exonuclease-sensitive strands and second exonuclease-resistant strands can comprise cleaving the genomic DNA, cDNA, or concatamer, e.g., via enzymatic digestion, sonication, mechanical shearing, electrochemical cleavage, and/or nebulization, to produce linear double-stranded nucleic acids. In these embodiments, exonuclease-sensitive tags can be ligated to the 5' ends of the first strands of the double-stranded nucleic acids and exonuclease-resistant tags can be ligated to the 5' ends of the second strands of the double-stranded nucleic acids to produce double stranded fragments with first exonuclease-sensitive strands and second exonuclease-resistant strands. The tags can optionally comprise one or more fluorescent label, blocking group, phosphorylated nucleotide, phosphorothioated nucleotide, biotinylated nucleotide, methylated nucleotide, nucleotide analog, uracil, sequence capable of forming a secondary structure, oligonucleotide hybridization site, restriction site, DNA promoter, RNA promoter, sample or library identification sequence, and/or cis regulatory sequence.

[0011] Optionally, producing the double stranded fragments with first exonuclease-sensitive strands and second exonuclease-resistant strands can comprise annealing a population of primer pairs, which comprise a first primer that comprises an exonuclease-sensitive 5' end and a second primer that comprises an exonuclease-resistant 5' end, to

IDT Exhibit 1020.1466

subsequences of the genomic DNA, the cDNA, or the concatamer. The primers can be extended with a polymerase to produce the double stranded fragments with first exonuclease-sensitive strands and second exonuclease-resistant strands. The primers can optionally comprise any of the moieties and/or sequences that can be included in the tags.

[0012] Those of skill in the art will appreciate that double-stranded nucleic acid fragments that comprise one exonuclease-sensitive strand and one exonuclease-resistant strand can be prepared using a variety of techniques well known in the art and that the embodiments described above should not be taken as limiting.

[0013] Producing a population of linear single-stranded nucleic acids from the double-stranded nucleic acid fragments that comprise first exonuclease-sensitive strands and second exonuclease-resistant strands can optionally include eliminating the exonuclease sensitive strand from each double-stranded fragment, e.g., via exonuclease digestion.

[0014] Compositions provided by the invention, related to the methods described above, comprise an exonuclease and a population of double-stranded nucleic acid fragments that comprise first exonuclease-sensitive strands and second exonuclease-resistant strands. The nucleic acid fragments of the composition, which can be derived from any of the sources described above, can optionally comprise overlapping subsequences of a genomic DNA, a cDNA, or a DNA concatamer. In preferred embodiments of the compositions, the genomic DNA, cDNA, or DNA concatamer are derived from a eukaryote. The fragments in the compositions can optionally comprise any one or more of the moieties that can be included in the tags.

[0015] The invention also provides methods of preparing closed single-stranded nucleic acid loops. One set of methods for preparing closed single-stranded loops includes providing a genomic DNA, a cDNA or a DNA concatamer, and generating double-stranded fragments that comprise first strands, e.g., exonuclease-sensitive strands, and second strands, e.g., exonuclease-resistant strands, from the genomic DNA, the cDNA, or the concatamer. The methods include separating the first strands of the double-stranded fragments from the second strands to produce single-stranded fragments, and circularizing the single-stranded fragments to produce the closed single-stranded nucleic acid loops.

[0016] The genomic DNA, the cDNA, or the concatamer from which the closed single-stranded nucleic acid loops are produced can optionally be derived from any of the

IDT Exhibit 1020.1467

sources described above. The population of double-stranded fragments that comprise first strands (or exonuclease-sensitive strands) and second strands (or exonuclease-resistant strands) can optionally be produced using any of the strategies described above. Separating the first strands from the second strands of the double-stranded fragments can optionally comprise digesting the first strands, e.g., exonuclease sensitive strands, of the fragments with an exonuclease to produce the single-stranded fragments.

[0017] Circularizing the single-stranded fragments to produce the closed single-stranded nucleic acid loops can optionally comprise annealing single-stranded nucleic acid splints to the single-stranded fragments. A splint can optionally be about 18-40 nucleotides long, and the single-stranded nucleic acid that is to be circularized can be about 100 nucleotides long. The first ends of the splints can optionally comprise first nucleotide sequences complementary to second nucleotide sequences at the first ends of the single-stranded fragments, and the second ends of the splints can optionally comprise third nucleotide sequences complementary to fourth nucleotide sequences at the second ends of the single-stranded fragments. Thus, annealing the splints to the single-stranded nucleic acids can bring the first and the second ends of each single-stranded fragment into proximity with one another. The first and the second ends of the single-stranded fragments can then be ligated to one another, producing the closed single-stranded nucleic acid loops. The splint can optionally be removed, e.g., via exonuclease digestion, e.g., before using the single-stranded nucleic acid loops in sequencing reactions.

[0018] The invention also provides compositions related to these methods of producing single-stranded nucleic acids loops. These compositions comprise a plurality of single-stranded nucleic acid fragments derived from overlapping subsequences of a genomic DNA, a cDNA, or a DNA concatamer, and a population of single-stranded nucleic acid splints. In the compositions, first ends of the splints are annealed to first ends of the fragments and second ends of the splints are annealed to seconds end of the fragments, circularizing each fragment by bringing its ends within proximity of one another. In preferred embodiments, the genomic DNA, cDNA or concatamer is derived from a eukaryote. The single stranded nucleic acids can optionally comprise any one or more of the moieties that can be included in nucleic acid tags. Optionally, the compositions can include a ligase.

IDT Exhibit 1020.1468

[0019] The invention also provides additional methods of producing closed single-stranded nucleic acid loops. These methods include providing a genomic DNA, a cDNA, or a DNA concatamer, as described above, and generating double-stranded nucleic acid fragments with first strands that comprise a gap site from the genomic DNA, cDNA, or concatamer. These double-stranded fragments can be circularized to form double-stranded loops that comprise first non-contiguous strands and second contiguous strands. The first non-contiguous strands can be removed from the double-stranded loops to produce the population of closed single-stranded nucleic acid loops. Optionally, the methods include sequencing the single-stranded nucleic acid loops.

[0020] The genomic DNA or the cDNA from which fragments are generated can optionally be derived any of the sources described above. A gap site can be introduced into a nucleic acid strand using a variety of methods known in the art. The double-stranded fragments that comprise a gap site can optionally be generated by the strategies described previously. The gap site can optionally comprise an unphosphorylated 5' end or a sequence recognized by a nicking enzyme, e.g., a uracil-DNA glycosylase (UDG), a uracil-DNA N-glycosylase (UNG), or a site-specific restriction endonuclease engineered to cut one strand.

[0021] In embodiments wherein the gap site comprises an unphosphorylated 5' end, circularizing the double-stranded nucleic acid fragments to produce a population of double-stranded loops that comprise first non-contiguous strands and second contiguous strands can optionally comprise ligating first ends of the fragments to second ends of the fragments. Optionally, in embodiments wherein the gap site comprises a sequence recognized by a nicking enzyme, producing a population of double-stranded loops that comprise first non-contiguous strands and second contiguous strand can include ligating first ends of the fragments to second ends of the fragments and cleaving the strands that comprise the gap site with, e.g., a nicking enzyme that specifically recognizes the gap site, e.g., a UNG, a UDG, or a site-specific restriction endonuclease engineered to cut one strand. Removing first non-contiguous strands from the double-stranded loops can comprise digesting the non-contiguous strand with, e.g., an exonuclease, to produce the population of single-stranded nucleic acid loops.

[0022] Compositions related to these methods of producing single-stranded nucleic acid loops are also provided by the invention. The compositions comprise a plurality of double-stranded nucleic acid loops that comprise first non-contiguous strands and second

IDT Exhibit 1020.1469

contiguous strands. The loops of the composition comprise overlapping sequences of a genomic DNA, a cDNA, or DNA concatamer. The source of the genomic DNA, cDNA, or DNA concatamer is not limited. However, in preferred embodiments of the compositions, the genomic DNA, cDNA, or DNA concatamer is derived from a eukaryote. The double-stranded nucleic acid loops can optionally comprise any one or more of the moieties that can be included in the tags. Optionally, the compositions can comprise an exonuclease.

[0023] In other embodiments, single-stranded nucleic acid loops can also be produced by providing a genomic DNA, a cDNA, or a DNA concatamer, e.g., derived from any of the sources described above, and generating a population of double-stranded fragments that comprise first sacrificial strands with a rolling-circle replication (RCR) protein recognition sequence and second target strands. These methods include copying target strands of the double-stranded fragments to produce a population of single-stranded copies, and circularizing the single-stranded copies to produce the population of closed single-stranded nucleic acid loops.

[0024] The double-stranded fragments with first sacrificial strands with a rolling-circle replication (RCR) protein recognition sequence can optionally be generated as described previously. The RCR protein recognition sequence can optionally comprise a sequence that is bound by an RCR protein, optionally, an RCR protein that comprises a histidine-U-histidine-U-U-U amino acid motif, wherein U is a bulky hydrophobic amino acid residue, e.g., cisA.

[0025] Copying the target strands of the double-stranded fragments can comprise nicking sacrificial strands of the fragments, displacing the sacrificial strands, and replicating sequences of the target strands to produce the population of single-stranded copies. The sacrificial strands of the target strands can be nicked with an RCR protein, e.g., cisA, and displaced by a replisome, e.g., a replisome that comprises a single-stranded DNA binding protein (SSB), a helicase, a polymerase, and an RCR protein. Displacing the sacrificial strands can optionally comprise unwinding the sacrificial strands from the target strands of the double-stranded fragments with a replisome.

[0026] Replicating the target strand can comprise synthesizing nucleic acid strands that are complementary to the target strands to produce the population of single-stranded copies. Circularizing the single-stranded copies, e.g., the copied strands and the sacrificial

IDT Exhibit 1020.1470

strands, to produce the closed single-stranded nucleic acid loops can comprise ligating the ends of the copied strands and the sacrificial strands with an RCR initiation protein, e.g., cisA.

[0027] In a related aspect, compositions provided by the invention include a plurality of double-stranded nucleic acids, each of which comprise an overlapping subsequence of a genomic DNA, a cDNA, or a DNA concatamer. The double-stranded nucleic acids comprise a rolling-circle replication (RCR) protein recognition sequence, e.g., a sequence that is bound by an RCR protein, and optionally an RCR protein that comprises a histidine-U-histidine-U-U-U amino acid motif, wherein U is a bulky hydrophobic amino acid residue, e.g., cisA. The double-stranded nucleic acids of the composition can be closed loops or linear, and they can optionally comprise genomic DNA, a cDNA, or a DNA concatamer that is derived from a eukaryote. The nucleic acids of the compositions can optionally include any one or more of the moieties that can be included in tags.

[0028] The invention provides additional methods of producing a population of single-stranded linear nucleic acids that include providing a genomic DNA, a cDNA, or a DNA concatamer, and generating a population of double-stranded fragments from the genomic DNA, the cDNA, or the concatamer. The fragments generated in these methods have first strands with first ends that comprise a sequence that is recognized by a nicking enzyme. The methods include circularizing the double-stranded fragments to produce a set of closed double-stranded nucleic acid loops, cleaving sacrificial strands of the closed loops with the nicking enzyme to produce nicked loops, and copying target strands of the nicked loops to produce copied strands that comprise sequences that are recognized by the nicking enzyme. The copied strands are then nicked with the nicking enzyme to produce the single-stranded linear nucleic acids.

[0029] The genomic DNA, cDNA, or concatamer from which the double-stranded fragments are generated can optionally be derived from any of a variety of sources described previously, including, e. g., a eukaryote, a prokaryote, an archaea, a virus, a phage, etc. Methods of producing double-stranded fragments with first strands that comprise first ends with a sequence recognized by a nicking enzyme include, but are not limited to, any of the methods for generating double-stranded fragments that have been described previously.

IDT Exhibit 1020.1471

[0030] Circularizing the double-stranded fragments can optionally comprise ligating the ends the fragments to produce closed loops. The sacrificial strands of the closed double-stranded loops can be optionally cleaved by an enzyme that recognizes the nicking sequence, e.g., a UNG, a UDG, or a site-specific restriction endonuclease engineered to cleave only one strand. One of skill in the art will recognize that the nicking enzyme used in these methods need not be limited to those listed above. Copying the target strands can optionally comprise displacing the sacrificial strands and copying the target strands, i.e., synthesizing nucleic acid strands that are complementary to the target strands, with, e.g., a strand-displacing polymerase, to produce the copied strands. Strand displacing polymerases optionally include, e.g., a PolI, a BstI, a Phi29, or a Phi29-like polymerases, such as those described in United States Patent Application No. 11/645,223, entitled POLYMERASES FOR NUCLEOTIDE ANALOGUE INCORPORATION, published August 23, 2007 as Publication No. 2007-0196846. The copied strands can then be nicked with a nicking enzyme to produce the single-stranded linear nucleic acids. The nicking enzyme can optionally be any of the nicking enzymes described previously.

[0031] The invention also provides compositions that are related to these methods. Such compositions comprise a plurality of closed double-stranded nucleic acid loops that comprise overlapping fragments of a genomic DNA, a cDNA, or a DNA concatamer. Though the genomic DNA, the cDNA, or the DNA concatamer from which the loops have been generated can be derived from any source, in preferred embodiments, the loops comprise nucleic acids derived from a eukaryote. The double-stranded loops of the compositions each comprise a sequence that is specifically recognized by enzymes that can introduce a nick in the sacrificial strands, e.g., those described above. Optionally, the loops in the compositions can comprise any one or more of the moieties that can be included in tags. The compositions can optionally comprise a nicking enzyme.

[0032] Methods for generating closed single-stranded nucleic acid loops can also include providing a genomic DNA, a cDNA, or a DNA concatamer, producing a population of double-stranded nucleic acid fragments from the genomic DNA, the cDNA, or the concatamer, providing first hairpin oligonucleotides to first ends of the nucleic acid fragments and providing second hairpin oligonucleotides to second ends of the nucleic acid fragments to generate the set of single-stranded nucleic acid loops. The first and second hairpins that are provided to the double-stranded fragments can optionally comprise

IDT Exhibit 1020.1472

identical nucleotide sequences and/or be the same length. In other embodiments, the first and second hairpins can comprise two different sequences and/or be two different lengths.

[0033] The genomic DNA, the cDNA, or the concatamer from which the closed single-stranded nucleic acid loops are produced can be derived from any of the previously described sources. The population of double-stranded nucleic acid fragments can be optionally generated according to strategies elaborated above.

[0034] Providing the double-stranded nucleic acid fragments with hairpin oligonucleotide segments can comprise annealing the first hairpin oligonucleotides to the first ends of the fragments, annealing the second hairpin oligonucleotides to the second ends of the fragments, and ligating the hairpin oligonucleotides to the ends of the fragments to generate the set of closed single-stranded nucleic acid loops that comprise regions of internal complementarity. Attaching the hairpins to the fragments can optionally comprise linking 5' strands of the first hairpins to 3' strands of the first ends of the fragments, linking 3'strands the first hairpins to 5' strands of the first ends of the fragments, linking 5' strands of the second hairpins to 3' strands of the second ends of the fragments, and linking 3' strands of the second hairpins to 5' strands of the second ends of the fragments to form a population of closed single-stranded nucleic acid loops.

[0035] The first hairpins can optionally comprise first single-stranded terminal sequences that are complementary to second single-stranded terminal sequences at the first ends of the fragments, and the second hairpins can optionally comprise third single-stranded terminal sequences that are complementary to fourth single stranded terminal sequences at the second ends of the fragments. The first or second hairpins can optionally comprise or encode one or more ligand, fluorescent label, blocking group, phosphorylated nucleotide, phosphorothioated nucleotide, biotinylated nucleotide, methylated nucleotide, nucleotide analog, uracil, sequence capable of forming a secondary structure, oligonucleotide hybridization site, restriction site, DNA promoter, RNA promoter, sample or library identification sequence, and/or cis regulatory sequence.

[0036] Compositions comprising first hairpin oligonucleotides, second hairpin oligonucleotides, and a plurality of double-stranded nucleic acid fragments that comprise overlapping subsequences of a genomic DNA, a cDNA, or a DNA concatamer, are provided by the invention. The genomic DNA, cDNA, or DNA concatamer can optionally be derived

IDT Exhibit 1020.1473

from any of the sources described previously. However, in preferred embodiments, the fragments of the composition are derived from a eukaryote.

[0037] The first or second hairpins of the compositions can optionally comprise or encode any of the moieties described previously. The first and second hairpins can optionally comprise the same sequence of nucleotides. In some embodiments of these compositions, the first and second hairpins can comprise two different nucleotide sequences. The first hairpins of the compositions can optionally comprise first single-stranded terminal sequences complementary to second single-stranded terminal sequences at the first ends of the fragments, and the second hairpins of the composition can optionally comprise third single-stranded terminal sequences complementary to fourth single stranded terminal sequences at the second ends of the fragments. The compositions can optionally include a ligase.

[0038] Linear single-stranded nucleic acids can be produced by another set of methods provided by the invention. This set of methods includes providing a genomic DNA, a cDNA, or a DNA concatamer, generating a set of double-stranded nucleic acid fragments from the genomic DNA, the cDNA, or the concatamer, and nicking sacrificial strands of the double-stranded nucleic acid fragments to produce nicked fragments. In these methods, the sacrificial strands are displaced from the fragments, and sequences between nick sites and ends on un-nicked strands that are exposed by the displacement of the sacrificial strands are copied with a strand displacing polymerase to produce the linear single-stranded nucleic acids.

[0039] The genomic DNA, the cDNA, or the concatamer from which the fragments are generated can optionally be derived from a eukaryote or any of the other sources described above. The double-stranded fragments can optionally be generated using any of the methods described above. Nicking the sacrificial strands of the fragments can optionally comprise cleaving the sacrificial strands with a nicking enzyme, such as a UDG, a UNG, or a site-specific restriction endonuclease engineered to cleave only one strand, to produce the nicked fragments. Displacing the sacrificial strands can optionally comprise removing the sacrificial strands from the un-nicked strands with a strand-displacing polymerase, e.g., those described above. The exposed sequences can be copied by replicating sequences between the nick sites and the ends on the un-nicked strands with the strand-displacing polymerase. Replicating can optionally comprise synthesizing nucleic

IDT Exhibit 1020.1474

acid strands that are complementary to the un-nicked strands, to produce the linear single-stranded nucleic acids.

[0040] Any of the preceding methods of generating closed single stranded nucleic acid loops or single-stranded nucleic acid fragments can further include the step of sequencing the single-stranded nucleic acids, e.g., in a high-throughput sequencing system, such as an array of zero-mode waveguides (ZMWs). In addition, any of the compositions described herein can optionally include any compound or molecule useful for sequencing including, but not limited to, e.g., a DNA polymerase, a buffer solution and/or salt solution, including, e.g., divalent metal ions, i.e., Mg^{2+}, Mn^{2+}, Ca^{2+} and/or Fe^{2+}. Optionally, any of the compositions can be present in a ZMW.

[0041] Those of skill in the art will appreciate that the methods and compositions provided by the invention can be used alone or in combination. Systems that include modules for the production and/or sequencing of linear single-stranded nucleic acids and/or closed single-stranded nucleic acid loops are also a feature of the invention. Such systems can optionally include detectors, array readers, excitation light sources, one or more output devices, such as a printer and/or a monitor to display results, and the like.

[0042] Kits are also a feature of the invention. The present invention provides kits that incorporate the compositions of the invention, optionally with additional useful reagents such as one or more enzymes that are used in the methods, e.g., an\ nicking enzyme, a DNA polymerase, an RCR protein, etc., that can be unpackaged in a fashion to enable their use. Depending upon the desired application, the kits of the invention optionally include additional reagents, such as a control nucleic acids, buffer solutions and/or salt solutions, including, e.g., divalent metal ions, i.e., Mg^{2+}, Mn^{2+}, Ca^{2+} and/or Fe^{2+}, to prepare the single-stranded nucleic acids produced by the methods for sequencing, e.g., in a high-throughput sequencing system. Such kits also typically include a container to hold the kit components, instructions for use of the compositions, and other reagents in accordance with the desired application methods.

BRIEF DESCRIPTION OF THE DRAWINGS

[0043] **Figure 1** illustrates methods and compositions for generating single-stranded nucleic acid fragments from double-stranded fragments.

IDT Exhibit 1020.1475

[0044] **Figure 2** illustrates methods and compositions related to generating closed single-stranded nucleic acid loops from double-stranded fragments.

[0045] **Figures 3A and 3B** illustrate other methods and compositions related to producing closed single-stranded nucleic acid loops from double-stranded fragments.

[0046] **Figure 4** illustrates additional methods and compositions related to generating closed single-stranded nucleic acid loops from double-stranded nucleic acids.

[0047] **Figure 5** illustrates methods and compositions for producing closed double-stranded nucleic acid loops from which linear single-stranded nucleic acids are copied.

[0048] **Figure 6** illustrates methods and compositions for generating single-stranded nucleic acid loops from double-stranded fragments and hairpin oligonucleotides.

[0049] **Figure 7** illustrates methods and compositions for generating linear single-stranded nucleic acids from double-stranded fragments using strand-displacing polymerases.

DETAILED DESCRIPTION

[0050] Collecting reliable sequence data using high-throughput sequencing technologies depends in part on the availability of methods for the rapid and efficient production of high-quality nucleic acid templates. However, many of the methods of template preparation that are currently available entail constructing a library to clone, e.g., overlapping fragments of a genomic DNA, and growing cells to amplify each member of the library. These steps can be both time-consuming and expensive, and they can grow even more so when, e.g., a large genome is to be sequenced. The present invention provides methods and compositions that can be useful in supplying templates to such high throughput DNA sequencing systems as, e.g., single-molecule real-time (SMRT) systems and others. The methods circumvent the need for costly, labor-intensive cloning and cell culture methods, which can limit sample production from matching the capacities of modern sequencing systems (such systems are reviewed in, e.g., Chan, et al. (2005) "Advances in Sequencing Technology" *Mutation Research* **573:** 13-40, and described in Levene et al. (2003) "Zero Mode Waveguides for Single Molecule Analysis at High Concentrations," *Science* **299:** 682-686). Accordingly, a reduction in sequencing costs from current levels is a benefit of using the improved template preparation methods provided herein. Advantageously, the methods can be scaled to accommodate template production for a variety of sequencing applications, e.g., sequencing individuals' genomes, gene expression

IDT Exhibit 1020.1476

profiling (Spinella, et al. (1999) "Tandem arrayed ligation of expressed sequence tags (TALEST): a new method for generating global gene expression profiles." *Nucleic Acids Res* 27: e22, Velculescu, et al. (1995) "Serial analysis of gene expression." *Science* 270: 484–487), genome-wide methylation analysis, microbiome analysis, and others.

[0051] The methods and compositions provided by the invention can be used to produce either linear single-stranded nucleic acids or closed single-stranded nucleic acid loops. Single-stranded templates are typically preferable to double-stranded nucleic acids for sequencing because the thermodynamic stability of, e.g., homoduplex DNA, can promote the rapid reannealing of, e.g., double-stranded DNA that has been denatured in preparation for a sequencing reaction. Such reannealing reduces the efficiency with which a primer can hybridize to a template. Furthermore, the reannealing of double-stranded DNA can also impede polymerase-catalyzed extension of a sequencing reaction, decreasing the speed of the reaction and the accuracy of the results.

[0052] The invention provides a variety of methods and compositions related to the production of linear single-stranded nucleic acids or closed single-stranded nucleic acid loops, e.g., for sequencing. For example, in one embodiment, linear single stranded nucleic acids are produced from double stranded fragments that comprise an exonuclease-resistant strand and an exonuclease-sensitive strand. In a second embodiment, closed, single-stranded nucleic acid loops are produced from double stranded fragments using short nucleic acid splints. In a third embodiment, closed, single-stranded nucleic acid loops are produced from double stranded fragments that comprise one strand with a gap site. In another embodiment, replisomes comprising a rolling-circle replication protein are used to generate closed, single-stranded nucleic acid loops from double-stranded fragments that comprise one strand with a rolling circle replication (RCR) protein recognition sequence. In an another embodiment, strand-displacing polymerases are used to generate single-stranded linear nucleic acids from double-stranded fragments that comprise one strand with a sequence recognized by a nicking enzyme. In another embodiment provided by the invention, single-stranded nucleic acid loops are produced from self-annealing hairpin oligonucleotides that are attached to the ends of double-stranded nucleic acid fragments. Lastly, the invention provides methods of using a strand-displacing polymerase to generate single-stranded linear nucleic acids from double-stranded nucleic acid fragments that comprise one non-contiguous strand.

IDT Exhibit 1020.1477

[0053] The detailed description is organized to first elaborate the various methods and compositions provided by the invention for the preparation of templates for high-speed, high-capacity sequencing platforms. Next, details regarding sequencing reactions and high-throughput sequencing systems are described. Broadly applicable molecular biological techniques that can be used to perform any of the methods are described thereafter.

METHODS AND COMPOSITIONS FOR GENERATING SINGLE-STRANDED NUCLEIC ACID FRAGMENTS FROM DOUBLE-STRANDED NUCLEIC ACID FRAGMENTS

[0054] The methods and compositions described below are related to preparing linear single-stranded nucleic acids and can be used to supply high-throughput sequencing systems with templates in an efficient, timely, and cost-effective manner. Unlike shotgun cloning strategies for template preparation, the methods do not require library construction or cell culture, which can become impracticable if scaled to the degree necessary to meet the capacity of high-speed next generation sequencing platforms. Advantageously, the methods below produce single-stranded templates. Double-stranded templates can reanneal in a sequencing reaction, reducing primer annealing efficiency and impeding the polymerase-catalyzed extension of a sequencing reaction.

[0055] The methods entail providing a genomic DNA, a cDNA, or a DNA concatamer and producing double-stranded fragments that comprise one exonuclease-sensitive strand and one exonuclease-resistant strand. As used herein, a "DNA concatamer" refers to a long DNA molecule comprising a multiplicity of short sequence tags, e.g., derived from expressed transcripts, that have been linked in tandem. Such concatamers can be used in genome-wide expression profiling. The exonuclease-sensitive strands of each double-stranded fragment are then eliminated, producing a population of linear, single-stranded nucleic acid fragments.

[0056] These methods are schematically illustrated in **Figure 1**. In a first step, genomic DNA, cDNA or a DNA concatamer **100**, is provided. The genomic DNA, cDNA, or concatamer can be derived from any of a variety of sources, including prokaryotes, archaea, viruses, phage, eukaryotes, etc. In preferred embodiments of the methods, the genomic DNA, cDNA or concatamer is derived from a eukaryote, such as a human or other mammal with a complex genome.

IDT Exhibit 1020.1478

[0057] In the next step, population of double-stranded nucleic acid fragments **105** is generated from the genomic DNA, cDNA, or concatamer **100**. The double-stranded fragments, which are preferably between about 100 and 2000 base pairs long, comprise exonuclease-sensitive strands **112** and exonuclease-resistant strands **110**.

[0058] Population of fragments **105** can be generated using any of a variety of techniques well known in the art. For example, the genomic DNA, cDNA, or concatamer can be cleaved, e.g., via enzymatic digestion, sonication, mechanical shearing, electrochemical cleavage, and/or nebulization. Following the fragmentation of the genomic DNA, cDNA, or concatamer, exonuclease-sensitive tags **112** can be attached to the 5' ends of the first strands of the fragments and exonuclease-resistant tags **110** can be attached to the 5'ends of the second strands of the fragments, e.g., with a ligase, via primer extension, via chemical linkage, and the like. Optionally, the double-stranded fragments can be produced by annealing a population of nested primer pairs, e.g., which pairs comprise one primer that comprises an exonuclease-sensitive 5' end and a second primer that comprises an exonuclease-resistant 5' end, to the genomic DNA, the cDNA or the concatamer, and extending the primers with a polymerase. Those of skill in the art will recognize that the methods of generating double stranded nucleic acid fragments comprising one exonuclease sensitive strand and one exonuclease resistant strand that are described above should not be taken as limiting.

[0059] As used herein, a "tag" refers to a moiety linked to a nucleic acid of interest that can be used as a molecular recognition site to identify or distinguish the nucleic acid in a population, e.g., as a means to permit a protein, e.g. a DNA-binding protein, or an enzyme, e.g., an exonuclease, a restriction enzyme, a nicking enzyme, or the like, to recognize the nucleic acid and perform an activity, and/or as a means by which to separate the nucleic acid from the population. A tag can comprise one or more of a number of moieties, including labeled or modified nucleotides, e.g., fluorescently labeled nucleotides, nucleotide analogs, or the like. Tags can also comprise specific nucleotide sequences, e.g., restriction sites, cis regulatory elements, recognition sites for nucleic acid-binding proteins, sequences capable of forming secondary structures, or the like. The tags and/or primers used in generating the double-stranded fragments that comprise first exonuclease-sensitive strands and second exonuclease-resistant strands can comprise one or more ligand, fluorescent label, blocking group, phosphorylated nucleotide, phosphorothioated nucleotide,

IDT Exhibit 1020.1479

biotinylated nucleotide, methylated nucleotide, nucleotide analog, uracil, a sequence capable of forming a secondary structure, oligonucleotide hybridization site, restriction site, DNA promoter, RNA promoter, sample or library identification sequence, cis regulatory sequence, and/or the like. For example, an "exonuclease resistant tag" can include a 5' phosphorylated nucleotide, which prevents the nucleic acid to which it is attached from being digested by a 5' exonuclease. An "exonuclease sensitive strand" can include a 5' unphosphorylated nucleotide, which renders the nucleic acid to which it is attached susceptible to digestion by a 5' exonuclease.

[0060] Following the production of double-stranded fragments that comprise exonuclease-sensitive strands **112** and exonuclease-resistant strands **110**, the exonuclease-sensitive strands **112** are eliminated, e.g., via digestion with exonuclease **115**, thereby producing single-stranded linear nucleic acid fragments **120** that can subsequently be sequenced in high-throughput sequencing systems (described elsewhere herein). The invention provides related composition **117** that includes an exonuclease and overlapping double-stranded fragments of a genomic DNA, cDNA, or concatamer that each comprise one exonuclease-sensitive strand and one exonuclease-resistant strand. The fragments can include any one or more of the moieties described previously.

METHODS AND COMPOSITIONS FOR GENERATING SINGLE-STRANDED NUCLEIC ACID LOOPS FROM DOUBLE-STRANDED FRAGMENTS

[0061] The methods and compositions described below relate to preparing closed nucleic acid loops that can be used, e.g., in sequencing reactions in high-throughput sequencing systems. In contrast to shotgun cloning strategies for template preparation, these methods do not require library construction or cell culture, which are costly, time-consuming, and which can become impracticable if scaled to the degree necessary to meet the capacity of high-speed next generation sequencing platforms. Advantageously, the methods below produce single-stranded nucleic acid loops, which can be preferable in sequencing reactions. Double-stranded templates can reanneal, reducing primer annealing efficiency and impeding the polymerase-catalyzed extension of a sequencing reaction. In fact, loops can also be preferable to linear templates because a DNA polymerase can only copy a linear template, e.g., to which a primer has been annealed, once before it falls off the distal end of the template. In contrast, a strand-displacing polymerase can replicate a contiguous nucleic acid loop several times. The primer that is annealed to the loop is eventually displaced at its 5'-end upon completion of one revolution of the polymerase

IDT Exhibit 1020.1480

around the nucleic acid loop, and as polymerization and displacement continue, a linear, single-stranded product comprising several copies of the nucleic acid sequence of the loop is generated. Accordingly, using nucleic acid loops in sequencing can provide an internal sequencing control.

[0062] The methods for preparing closed, single-stranded nucleic acid loops include providing a genomic DNA, a cDNA, or a DNA concatamer and generating double-stranded fragments that each comprise a first strand (e.g., an exonuclease sensitive strand) and a second strand (e.g., an exonuclease resistant strand). In a following step, the two strands in each fragment are separated, and the resulting single-stranded fragments are circularized to produce closed single-stranded nucleic acid loops, which can then be used as templates in a high-throughput sequencing system, .

[0063] **Figure 2** depicts a schematic for these methods. First, population of double-stranded nucleic acid fragments **210**, which comprise first strands (or exonuclease-sensitive strands) **202** and second strands (or exonuclease-resistant strands) **201,** are generated from genomic DNA, cDNA, or a DNA concatamer **200**, using any of the strategies described previously. The genomic DNA, cDNA, or concatamer can be derived from any source known to those of skill in the art.

[0064] Next, strands **201** and **202** of the double-stranded fragments are separated from one another, e.g., via digestion with exonuclease **235** to produce set of single-stranded nucleic acid fragments **240** e.g., fragments that are about 100 to about 1000 base pairs long. Those of skill in the art will recognize that the strands of a double-stranded nucleic acid fragment can be separated using a variety of techniques that those described herein are not to be taken as limiting. Single-stranded fragments **240** can then be circularized to produce a population of single-stranded nucleic acid loops, e.g., for sequencing.

[0065] As used herein, "circularizing" a nucleic acid refers to the formation of a loop, e.g., a ring-like structure that does not intersect itself, from a linear nucleic acid fragment, e.g., a single-stranded fragment or double-stranded fragment. A linear nucleic acid fragment can be circularized by joining, e.g., ligating, the ends of the fragment to one another, e.g., to form a "closed loop" or "contiguous loop", wherein each nucleotide is covalently joined to the nucleotides preceding and following it, or by bringing the ends of

IDT Exhibit 1020.1481

the fragment into proximity with one another, e.g., to form a "non-contiguous loop", wherein at least two nucleotides of the fragment are not joined by a covalent bond.

[0066] Single-stranded fragments **240** can be circularized using any number of strategies. However, in a preferred embodiment, fragments **240** can be circularized using splints **245**. As used herein, a "splint" refers to a short, single-stranded nucleic acid oligomer about 18-40 nucleotides long that comprises 3' sequences and 5' sequences that form a perfectly matched homoduplex when hybridized to the 5' and 3' sequences, respectively, at the ends of a second, preferably longer single-stranded nucleic acid molecule. Following hybridization to the longer single-stranded nucleic acid, the splint functions as a molecular bridge that circularizes the second nucleic acid molecule by holding its ends in apposition.

[0067] For example, splints **245** comprise first ends that comprise sequences **220** that are complementary to sequences **215** at the first ends of single-stranded nucleic acid fragments **240**. Splints **245** also comprise second ends that comprise sequences **230** that are complementary to sequences **225** at the second ends of single-stranded fragments **240**. Annealing fragments **240** to the splints **245** brings the two ends of each fragment within close proximity of one another (*See*, e.g., composition **250)**, which increases the efficiency of ligation, wherein the two ends of each fragment **240** are covalently linked, e.g., via enzymatic ligation. Splints **245** can be removed, e.g., via digestion with exonuclease **235,** to produce single-stranded nucleic acid loops **255**. Composition **260** can then be provided to a high-throughput sequencing system.

[0068] Methods of generating single-stranded nucleic acid loops have been described in Kuhn, et al. (2002) "Rolling-circle amplification under topological constraints." *NAR* 30: 574-580 for use in determining the efficiency of rolling-circle amplification using templates with varying topologically linked DNA constructs. However, because the methods in Kuhn et al. only produced a homogenous population of single-stranded nucleic acid loops, e.g., loops that each comprise the same sequence, they are not suitable for use in sequencing reactions. In contrast, the methods provided by the invention include the steps of fragmenting a genomic DNA, a cDNA, or a DNA concatamer to produce a heterogeneous population of nucleic acid loops that comprise overlapping sequences of the genomic DNA, cDNA, or concatamer. Accordingly, the heterogeneous

IDT Exhibit 1020.1482

templates can be sequenced to generate data that can be assembled to determine the nucleotide sequence of, e.g., a complex mammalian genome.

[0069] Methods of generating single-stranded nucleic acid loops for use in nucleic acid sequencing have also been described in United States Patent Publication No. US 2008/0213771 METHODS AND COMPOSITIONS FOR LARGE SCALE ANALYSIS OF NUCLEIC ACIDS USING DNA DELETIONS, by Drmanac. These methods entail generating linear single-stranded nucleic acids from a fragmented, denatured genomic DNA and ligating poly dA tails to 3-prime ends of the nucleic acids. This is then followed by ligation of the free ends intramolecularly with the aid of a splint, which splint is complementary to the poly dA tail at one end and complementary to any sequence at the other end by virtue of a segment of degenerate nucleotides. Whereas in the methods of 2008/0213771, single stranded linear nucleic acids that are to be circularized using a splint are generated by fragmenting and denaturing a genomic DNA, the single stranded linear nucleic acids, e.g., that are to be circularized, are generated, e.g., using the methods herein, by fragmenting a genomic DNA and digesting one strand of each double-stranded fragment. The thermodynamic stability of, e.g., homoduplex DNA, can promote the rapid reannealing of, e.g., double-stranded DNA fragments that have been denatured in preparation for circularization.

[0070] The invention also provides Composition **250**, which is related to the present methods. Composition **250** comprises a population of single-stranded nucleic acid fragments **240**, which comprise overlapping subsequences of a genomic DNA, a cDNA, or a DNA concatamer, and a population of single-stranded nucleic acid splints **245**. In this composition, first ends **220** of the splints are annealed to first ends **215** of the fragments and second ends **230** of the splints are annealed to seconds ends **225** of the fragments, such that the two ends of each fragment are brought within proximity of one another. The genomic DNA, the cDNA, or the concatamer from which the single-stranded fragments in composition **250** can optionally be derived from, e.g., a eukaryote, a prokaryote, an archaebacterium, a phage, or a virus. The fragments in the composition can optionally comprise any one or more of the moieties described previously. Optionally, the composition can include a ligase.

[0071] In another embodiment, single-stranded nucleic acid loops are generated from double-stranded nucleic acid fragments, e.g., derived from a genomic DNA, a cDNA,

IDT Exhibit 1020.1483

or a DNA concatamer, that comprise a gap site. As described herein, a "gap site" is a nucleotide sequence or nucleotide-associated moiety, such as an unphosphorylated 5' C, that permits the introduction of a gap into contiguous nucleic acid strand. For example, a gap site in a nucleic acid fragment can comprise an unphosphorylated 5' end or a sequence that is recognized by a nicking enzyme. As used herein, a "nicking enzyme" refers to an enzyme that can cleave one strand of a double-stranded nucleic acid. A nicking enzyme can optionally nick a nucleic acid strand at any location, or it can recognize a specific sequence and cleave a nucleic acid strand only at that sequence. Examples of nicking enzymes include, e.g., a UDG, a UNG, or a restricition endonuclease engineered to cleave only one strand of a double-stranded nucleic acid. Those of skill in the art will recognize that there exist myriad enzymes that can cleave a single strand of a double-stranded nucleic acid and that the enzymes listed above are not to be taken as limiting.

[0072] Next, the double stranded fragments are circularized to produce a population of double-stranded loops that comprise first strands (or non-contiguous strands) and second strands (or contiguous strands). The first (e.g., non-contiguous) strands of each double-stranded loop are then removed to produce a population of closed, single-stranded nucleic acid loops useful, e.g., in a high-throughput sequencing system. This method is schematically depicted in **Figure 3**.

[0073] In **Figure 3**, double stranded nucleic acid fragments wherein one strand includes a gap site, e.g., populations **305** and **350**, can be generated as described previously or by methods well-known in the art. In one embodiment, double stranded fragments **350**, comprising 5' unphosphorylated first strands **336** and 5' phosphorylated second strands **335**, are produced from DNA, e.g., genomic DNA, cDNA, or concatamer **300**. In this embodiment, the 5' unphosphorylated end of each first strand **336** in each fragment comprises the gap site. In an alternate embodiment, set of double-stranded fragments **305**, which each comprise a first strand that encodes sequence **306**, are generated from genomic DNA, cDNA, or concatamer **300**. In this embodiment, sequence **306,** which is recognized by nicking enzyme **315**, comprises the gap site.

[0074] Fragments **350**, which comprise 5' unphosphorylated first strands **336** and 5' phosphorylated second strands **335**, can be circularized, e.g., with a ligase, to produce population of double-stranded loops **375** that each comprise first non-contiguous strands and second contiguous strands. (See **Figure 3B**.) Alternately, fragments **305** can be

IDT Exhibit 1020.1484

circularized, e.g., with a ligase, to produce closed double-stranded closed loops **311**. (See **Figure 3A**.) The first strand of each double-stranded loop that encodes sequence **306** can then be cleaved with nicking enzyme **315** to produce population **312**, which comprises double-stranded loops that comprise first non-contiguous strands **325** and second contiguous strands. The nicking enzyme used in this embodiment of the methods can be, e.g., a site-specific restriction endonucleases engineered to cleave only one strand. Alternately, the nicking can be uracil-DNA N-glycosylase (UNG), which cleaves uracil N-glycosylic bonds, or uracil DNA glycosylase (UDG), which catalyzes the release of free uracil from uracil-containing DNA. Those of skill in the art will appreciate that there exist a myriad of enzymes that can nick a single strand in double-stranded nucleic acid and that those described herein are not to be taken as limiting. As shown in **Figures 3A** and **3B**, the non-contiguous strands, e.g., the first strands, of each double-stranded loop in populations **312** and **375** can then be removed, e.g., by digestion with exonuclease **320**, to produce single-stranded nucleic acid loops **326** that can be used, e.g., in high-throughput sequencing systems.

[0075] Related composition **375** is provided by the invention. The composition comprises a set of double-stranded nucleic acid loops comprising overlapping fragments of a genomic DNA, a cDNA, or DNA concatamer derived from any of the sources described previously. The double-stranded nucleic acid loops of the composition each comprise one non-contiguous strand and one contiguous strand. The genomic DNA, cDNA, or concatamer present in the loops can optionally be derived from any of the previously described sources. The double-stranded nucleic acid loops can optionally comprise any one or more of the moieties described previously. Optionally, the composition can include an exonuclease.

METHODS AND COMPOSITIONS FOR PRODUCING DOUBLE-STRANDED LINEAR TEMPLATES FROM WHICH CLOSED SINGLE-STRANDED NUCLEIC ACID LOOPS ARE COPIED

[0076] The invention also provides methods and compositions related to generating a population of double-stranded fragments, copying the fragments, and producing closed single-stranded loops from the copied fragments. The closed, single-stranded nucleic acid loops produced by these methods can be used to provide templates in high-throughput sequencing systems in, e.g., a time-efficient and cost-effective manner. Unlike shotgun cloning strategies for template preparation, these methods do not require library

IDT Exhibit 1020.1485

construction or cell culture, which are laborious, expensive, and not easily scalable to meet the capacities of next generation sequencing systems. Advantageously, the methods below produce closed single-stranded nucleic acid loops, which are preferable in sequencing reactions. For example, denatured double-stranded templates can rehybridize, reducing primer annealing efficiency and impeding the polymerase-catalyzed extension of a sequencing reaction.

[0077] Double-stranded fragments that each comprise one strand that comprises a rolling-circle replication (RCR) protein recognition sequence are generated, e.g., using any of the strategies known in the art or described herein, e.g., from a genomic DNA, a cDNA, or a DNA concatamer derived from any of the sources described above. RCR protein recognition sequences are described in further detail in, e.g., Koonin, et al. (1993) "Computer-assisted dissection of rolling circle DNA replication." *BioSystems* **30:** 241-268; and Novick (1998) "Contrasting Lifestyles of rolling-circle phages and plasmids." *TIBS* **23:** 434- 438. Next, target strands of each double-stranded fragment are copied to produce a population of single-stranded nucleic acid copies. As used herein, "copying" refers to the process of replicating a nucleic acid molecule to generate a new nucleic acid that comprises a sequence complementary to that of the original. The nucleic acids that are synthesized via copying are the "copied strands," and the rate of nucleotide misincorporation during the synthesis of the copied strands is assumed to be approximately < 5%, or, more preferably, 0%. The copied strands are then circularized to produce a population of single-stranded nucleic acid loops, which can optionally be provided to a high-throughput sequencing system. This method is illustrated in **Figure 4**.

[0078] Double-stranded nucleic acid fragments **410** are generated from genomic DNA, cDNA, or a DNA concatamer **400**, as described above. Fragments **410** comprise target strands **420** and sacrificial strands **416**. Sacrificial strands **416** each comprise rolling-circle replication (RCR) protein recognition sequence **405.** As used herein, "target strands" refer to the strands in double strand nucleic acid molecules, e.g., linear nucleic acids or nucleic acid loops, that serve as templates from which new molecules are synthesized. "Sacrificial strands", as used herein for this method and other methods of the invention, refer to the strands in double-stranded nucleic acid molecules, e.g., linear nucleic acids or nucleic acid loops, that are cleaved and removed, e.g., via displacement or enzymatic

IDT Exhibit 1020.1486

digestion, to permit the copying of target strands to produce single-stranded nucleic acid molecules.

[0079] In a following step, target strands **420** of each of the double-stranded fragments in population **410** are repeatedly copied to produce single-stranded loops **435**. In preferred embodiments, copying target strands **420** to produce single-stranded nucleic acid loops **435** includes nicking sacrificial strands **416** of each double-stranded fragment, e.g., with RCR protein **415**. The RCR protein that nicks sacrificial strands **416** can optionally be an RCR protein that comprises a histidine-U-histidine-U-U-U amino acid motif, wherein U is a bulky hydrophobic amino acid, e.g., cisA.

[0080] Replisome **425**, which generally comprises a single-stranded DNA-binding protein (SSB), a helicase, a polymerase, and an RCR protein, then displaces sacrificial strands **416** and copies sequences of target strands **420** that are exposed by the displacement of sacrificial strands **416**. Further details regarding RCR replispomes can be found in, e.g., e.g., Koonin, et al. (1993) "Computer-assisted dissection of rolling circle DNA replication." *BioSystems* **30:** 241-268; and Novick (1998) "Contrasting Lifestyles of rolling-circle phages and plasmids." *TIBS* **23:** 434- 438. In copying target strands **420**, replisome **425** synthesizes single-stranded copies **430**. The RCR protein in replisome **425** comprises an activity which permits it to then circularize single-stranded copies **430** (and sacrificial strands **416**) to produce closed single-stranded nucleic acid loops **435**.

[0081] It will be apparent to one of skill in the art that these methods can optionally include the step of circularizing population of fragments **410** to produce a population of double-stranded loops. The remaining steps can be performed using the double-stranded loops to generate population **435**.

[0082] Composition **410**, which is related to the present methods, is also an embodiment of the invention. Composition **410** comprises a population of double stranded nucleic acids derived from a genomic DNA, a cDNA or a DNA concatamer. Each of the double-stranded nucleic acids in population **410** includes one strand that comprises an RCR protein recognition sequence. The nucleic acids of the composition **410** can optionally be closed loops or linear, and they can optionally be derived from a eukaryote. In some embodiments of this composition, the RCR protein recognition sequence comprises a sequence recognized by an RCR protein that comprises a histidine-U-histidine-U-U-U

IDT Exhibit 1020.1487

motif, wherein U is a bulky hydrophobic amino acid. Optionally, the RCR protein that recognizes the RCR protein recognition sequence can be a cisA protein. The double stranded nucleic acids in the composition can optionally comprise any one or more of the moieties described previously.

METHODS AND COMPOSITIONS FOR PRODUCING CLOSED DOUBLE-STRANDED LOOPS FROM WHICH LINEAR SINGLE-STRANDED NUCLEIC ACIDS ARE COPIED

[0083] Methods of preparing templates for large-scale sequencing projects have typically entailed constructing shotgun libraries that comprise overlapping fragments of, e.g., a genomic DNA; transforming cells with the library; growing cells to amplify each library member; and isolating and purifying library DNA. However, template production can be scaled to the volumes that can be accommodated by high-throughput sequencing platforms using novel methods provided herein.

[0084] The present methods can be used to produce linear single-stranded nucleic acids. These methods not only circumvent cloning and cell culture steps, they also beneficially produce single-stranded nucleic acid templates, which are preferable in sequencing reactions.

[0085] In one embodiment, the methods include generating a population of double stranded fragments that comprise a sacrificial strand that encodes a sequence recognized by a nicking enzyme. The fragments are then circularized to produce closed double-stranded loops. Sacrificial strands in each loop are cleaved by a nicking enzyme, and the target strands of each loop are copied to produce copied strands that encode a sequence recognized by the nicking enzyme. The copied strands are then nicked to produce single stranded nucleic acids that can be used in a high-throughput sequencing systems.

[0086] In this set of methods, which are illustrated in **Figure 5**, double-stranded nucleic acid fragments **510** are produced from genomic DNA, cDNA, or a DNA concatamer **500** using, e.g., any of the previously described strategies. Double-stranded fragments **510** each comprise a sacrificial strand that encodes sequence **520**, which is recognized by nicking enzyme **543**, e.g., a UNG, a UDG, or a site-specific restriction endonuclease engineered to cleave only one strand. The double-stranded fragments can be circularized, e.g., with a ligase, to produce a set of closed double-stranded loops **530**. Sacrificial strands **544** of each double-stranded loop are cleaved by nicking enzyme **543**, and displaced, e.g.,

IDT Exhibit 1020.1488

by strand-displacing polymerase **540**, which also copies target strands **546**, thereby generating copied strands **550**. Copied strands **550** are eventually displaced at their 5'-ends upon completion of one revolution of the target strand **546** by strand-displacing polymerase **540**. As polymerization and displacement continue, copied strands **550**, which each comprise the sequence that is recognized by nicking enzyme **543**, are cleaved by nicking enzyme **543** to produce single-stranded nucleic acids **545**.

[0087] The strand displacing polymerases that can be used in preferred embodiments of these methods include, e.g., PolI, BstI, Phi29, or Phi29-like polymerases, such as those described in United States Patent Application No. 11/645,223, entitled POLYMERASES FOR NUCLEOTIDE ANALOGUE INCORPORATION.

[0088] The invention provides composition **530**, which is related to the methods above. Composition **530** comprises closed double-stranded nucleic acid loops that comprise overlapping subsequences of a genomic DNA, a cDNA, or a DNA concatamer. The loops include sequence sacrificial strands **544** that encode sequence **520**, which is recognized by nicking enzyme **543**. In preferred embodiments of the compositions, the closed double-stranded nucleic acid loops in composition **530** are derived from a eukaryotic genomic DNA, a eukaryotic cDNA, or concatamer comprising eukaryotic DNA. The double stranded nucleic acid loops can optionally comprise any one or more of the moieties described previously. The composition can optionally include a nicking enzyme, e.g., nicking enzyme **543**.

METHODS AND COMPOSITIONS FOR GENERATING SINGLE-STRANDED NUCLEIC ACID LOOPS FROM DOUBLE-STRANDED FRAGMENTS AND HAIRPIN OLIGONUCLEOTIDES

[0089] There is an increasing demand for efficient, low-cost methods for the preparation of high-quality nucleic acid templates for next generation sequencing technologies. The methods and compositions described below can be useful for supplying high throughput DNA sequencing systems with such templates. The methods avoid labor-intensive, costly cloning and cell culture steps that are typically used in currently available strategies to generate nucleic acid templates for sequencing, e.g., shotgun cloning. Moreover, these methods beneficially produce single-stranded nucleic acids, which do not reduce primer annealing efficiency and impede the polymerase-catalyzed extension of a sequencing reaction due to reannealing of double-stranded templates.

IDT Exhibit 1020.1489

[0090] The methods include providing double-stranded nucleic acid fragments generated by any one of the strategies described previously. Next, hairpin oligonucleotides are provided to each end of each double-stranded fragment and attached to each end of each fragment, generating a set of single-stranded nucleic acid loops that comprise regions of internal complementarity. These methods are schematically depicted in **Figure 6**.

[0091] Double-stranded fragments **610** are generated from genomic DNA, cDNA, or a DNA concatamer **600**. Procedures for generating double-stranded nucleic acid fragments **610** from genomic DNA, cDNA, or concatamer **600** are described above and are well known in the art. Genomic DNA, cDNA or concatamer **600** can be derived from any of the sources described above.

[0092] Next, population of hairpin oligonucleotides **620** is provided to one end of each double-stranded fragment in population **610**, and population of hairpin oligonucleotides **625** is provided to the other end of each fragment, as shown in composition **630**. The hairpins in population **620** and **625** can be about 20 to about 100 nucleotides in length. In some embodiments, hairpin oligonucleotides **620** and **625** are the same length and, optionally, identical in sequence. In other embodiments, the two populations of hairpins, e.g., **620** and **625**, comprise unique sequences, and one population of hairpins can optionally be longer than the other. Either population of hairpins can comprise one or more ligand, fluorescent label, blocking group, phosphorylated nucleotide, phosphorothioated nucleotide, biotinylated nucleotide, methylated nucleotide, nucleotide analog, uracil, a sequence capable of forming a secondary structure, oligonucleotide hybridization site, restriction site, DNA promoter, RNA promoter, sample or library identification sequence, cis regulatory sequence, and/or the like.

[0093] In preferred embodiments of the methods, hairpins **620** and **625** anneal to the first and second ends of fragments **610**, respectively. For example, hairpins **620** can comprise first single-stranded terminal sequences complementary to second single-stranded terminal sequences at the first ends of fragments **610**, and hairpins **625** can comprise third single-stranded terminal sequences complementary to fourth single stranded terminal sequences at the second ends of fragments **610**. Attaching hairpins **620** and **625** to fragments **610** can include linking 5' strands of hairpins **620** to 3'strands at the first ends of the fragments **610**, linking 3' strands of hairpins **620** to 5' strands at the first ends of fragments **610**, linking 5' strands of hairpins **625** to 3'strands at the second ends of

IDT Exhibit 1020.1490

fragments **610**, and linking 3' strands of hairpins **625** to 5' strands at the second ends of the fragments **610**, e.g., via ligation or chemical linkage. Thus, the hairpins are attached to the fragments in a configuration that produces population **635**, which comprises closed single-stranded nucleic acid loops with regions of internal complementarity.

[0094] Methods of producing closed nucleic acid loops from hairpin for use as templates in rolling circle replication are described in USPN No. 6,498,023 B1, by Abarzua, entitled, "Generation of Single-Strand Circular DNA from Linear Self-Annealing Segments." However, the present invention uses self-annealing nucleic acid hairpins with double-stranded nucleic acid fragments, e.g., derived from a genomic DNA, a cDNA, or a DNA concatamer. Because the methods herein produce a heterogeneous population of nucleic acid loops that comprise overlapping sequences of a genomic DNA, cDNA, or DNA concatamer, they can be used as templates provide high-throughput sequencing systems. Accordingly, the heterogeneous templates can be sequenced to generate data that can be assembled to determine the nucleotide sequence of, e.g., a complex mammalian genome.

[0095] Composition **630** is provided by the invention. Composition **630** comprises a population of double stranded fragments **610**, which comprise overlapping sequences of genomic DNA, cDNA, or DNA concatamer **600**. The genomic DNA, cDNA or DNA concatamer **600** can optionally be derived from a eukaryote. Composition **630** includes population of hairpin oligonucleotides **620**, and population of hairpin oligonucleotides **625**. Hairpin oligonucleotides **620** and/or **625** in composition **630** can optionally comprise or encode any of the additional moieties or sequences described above. Hairpins **620** of the composition **630** can optionally comprise single-stranded terminal sequences that are complementary to second single-stranded terminal sequences at first ends of fragments **610**, and hairpins **625** of composition **630** can optionally comprise third single-stranded terminal sequences that are complementary to fourth single-stranded terminal sequences at the second ends of fragments **610**. In some embodiments of the compositions, hairpin oligonucleotides **620** and **625** are the same length and, optionally, identical in sequence. In other embodiments, the two populations hairpins, e.g., **620** and **625**, comprise unique sequences, and one population of hairpins can optionally be longer than the other. The composition can optionally include a ligase. Further details regarding methods and compositions related to this embodiment can be found in U.S. Patent Application No. [unassigned], filed March 27, 2009 (Attorney Docket No. 01- 0005903).

IDT Exhibit 1020.1491

METHODS AND COMPOSITIONS FOR GENERATING SINGLE-STRANDED NUCLEIC ACIDS FROM DOUBLE-STRANDED FRAGMENTS USING STRAND-DISPLACING POLYMERASES

[0096] The following methods and compositions for preparing linear nucleic acid single-stranded nucleic acids, can be used to supply high-throughput sequencing systems with templates in an efficient, timely, and cost-effective manner. Because the methods do not entail cell culture or library construction, they can be scaled to meet the capacity of high-speed next generation sequencing platforms. Advantageously, the methods below produce single-stranded templates, which, unlike double-stranded templates, do not reduce primer annealing efficiency or impede the polymerase-catalyzed extension of a sequencing reaction by reannealing

[0097] In this embodiment, double-stranded fragments are generated from a genomic DNA, a cDNA, or a DNA concatamer, e.g., derived from any one of the sources described previously. Each of the fragments comprises a sacrificial strand and a target strand. The sacrificial strands in each fragment are nicked, e.g., anywhere along the length of the fragment, and displaced from the target strands. Next, sequences on the target strands that are exposed by the displacement of the sacrificial strands are copied by a strand-displacing polymerase to produce linear single-stranded nucleic acids, which can optionally be sequenced, e.g., in a high-throughput sequencing system.

[0098] A schematic of these methods is provided in **Figure 7**. Double-stranded fragments **705** can be produced from genomic DNA, cDNA, or DNA concatamer **700** using any of the strategies that have been described above. An advantage to this embodiment is that it does not entail the addition of tags to the fragments. Sacrificial strands **703** of double-stranded fragments **705** are nicked by nicking enzyme **710** to produce population of nicked fragments **720**, e.g., double-stranded fragments that comprise one nicked strand. Nicking enzyme **710** can optionally comprise any of the nicking enzymes described previously. Those of skill in the art will appreciate that there exist a myriad of enzymes that can nick a single strand in double-stranded nucleic acid and that those described herein are not to be taken as limiting.

[0099] The sacrificial strands of nicked fragments **720** are displaced, e.g., by strand displacing polymerase **725**, which then copies the sequences on target strands **704** that are exposed by the displacement of sacrificial strands **703**. Strand-displacing polymerases **725**

IDT Exhibit 1020.1492

synthesize complementary copies of target strands **704**, thereby producing population of single-stranded nucleic acids **730**.

DETAILS REGARDING SEQUENCING REACTIONS AND HIGH-THROUGHPUT SEQUENCING SYSTEMS.

[0100] DNA sequencing refers to methods for determining the order of the nucleotide bases, e.g., adenine, guanine, cytosine, and thymine, in a molecule of DNA, such as a genomic DNA, a cDNA, or a DNA concatamer. Typically, a sequencing reaction mix includes a polymerase; adenine, guanine, cytosine, and thymine nucleotides; a template strand, an oligonucleotide primer that comprises a sequence complementary to a sequence in the template strand, and a divalent cation, e.g., Mn^{2+} or Mg^{2+}, which improves the polymerase's activity. In general, a sequencing reaction entails annealing the oligonucleotide primer to the single-stranded DNA template and extending the primer with the polymerase, which incorporates nucleotide bases into a nascent chain to synthesize a DNA molecule whose sequence is complementary to that of the template strand. If a double-stranded template is provided, it is denatured prior to the annealing and extension steps. During synthesis, the incorporation of each individual nucleotide is detected, permitting the determination of the pattern of adenines, guanines, cytosines, and thymines in the template strand.

[0101] One sequencing method that is routinely used is chain termination sequencing, in which modified nucleotides that terminate DNA strand elongation. In chain termination sequencing, a sequencing reaction is divided into four separate sequencing reactions, each containing all four of the standard deoxynucleotides, a radiolabeled nucleotide, a template strand, a divalent cation, and a DNA polymerase. To each of the four reactions, one of four dideoxynucleotides (ddATP, ddGTP, ddCTP, or ddTTP) are added. Dideoxynucleotides are chain-terminating nucleotides because they lack a 3'-OH group required for the formation of a phosphodiester bond between two nucleotides, thus terminating DNA strand extension and resulting in various DNA fragments of varying length.

[0102] The newly synthesized and labeled DNA fragments are heat denatured, and separated by size (with a resolution of just one nucleotide) by gel electrophoresis on a denaturing polyacrylamide-urea gel with each of the four reactions run in one of four

IDT Exhibit 1020.1493

individual lanes (lanes A, T, G, C); the DNA bands are then visualized by autoradiography or UV light, and the DNA sequence can be directly read off the X-ray film or gel image.

[0103] Dye-terminator sequencing is a variation of the chain termination methods in which each of the four chain terminator ddNTPs is labeled with a fluorescent dye that has a unique wavelengths of fluorescence and emission. This strategy circumvents the need for four separate reactions, since all four fluorescent signals can be run and read, e.g., in the same lane on a gel or in the same capillary in a capillary electrophoresis system.

[0104] The high demand for large-scale sequencing has driven the development of high-throughput sequencing technologies that parallelize the sequencing process, producing thousands or millions of sequences at once. High-throughput sequencing technologies can lower the cost of DNA sequencing beyond what is possible with standard dye-terminator or chain termination methods. Certain commercial high-throughput sequencing systems, e.g., those available from 454 Life Sciences, Illumina, and Pacific Biosciences, are based on multiplexed direct sequencing methods, e.g., "sequencing by synthesis" (SBS), in which each base position in a single-stranded DNA template is determined individually during the synthesis of a complementary strand.

[0105] For example, pyrosequencing is a bioluminometric DNA sequencing technique in which the real-time release of the inorganic pyrophosphate (PPi) that is produced upon each successful incorporation of a nucleotide into a DNA is monitored (Nyren (2007) "The History of Pyrosequencing." *Methods Mol Biol* **373:** 1-14; Ronaghi (2001) "Pyrosequencing sheds light on DNA sequencing." *Genome Res* **11:** 3-11; and Wheeler, et al. (2008) "The complete genome of an individual by massively parallel DNA sequencing." Nature 452: 872-876). In pyrosequencing, PPi release begins an enzymatic cascade in which PPi is immediately converted to ATP by ATP sulfurylase. The ATP then fuels the luciferase-catalyzed oxidation luciferin, in which photons are emitted.

[0106] 454 Sequencing, a technology available from 454 Life Sciences, is a massively-parallellized, multiplex pyrosequencing system that relies on fixing nebulized, adapter-ligated single-stranded DNA fragments, e.g., which can be prepared by the methods decribed above, to small DNA-capture beads. The single-stranded DNAs fixed to these beads are then amplified, e.g., via PCR. Each DNA-bound bead is placed into a well on a proprietary PicoTiterPlate™, to which a mix of enzymes, including, e.g., DNA polymerase,

IDT Exhibit 1020.1494

ATP sulfurylase, and luciferase, has also been added. The PicoTiterPlate™ is then placed into a sequencing module, where dideoxyribonucleotides, e.g., A, C, G, and T, are washed in series over the PicoTiterPlate™. During the nucleotide flow, the copies of DNA that are attached to the beads are sequenced in parallel. If a nucleotide complementary to a template strand is flowed into a well of the PicoTiterPlate™, the polymerase extends the existing DNA strand by adding the nucleotide, releasing PPi and generating a light signal. The presence or absence of PPi, and, therefore, the incorporation or non-incorporation of each nucleotide washed over the PicoTiterPlate™, is ultimately assessed on the basis of whether or not photons are detected. There is a minimal time lapse between these events, and the conditions of the reaction are such that iterative addition of nucleotides and PPi detection are possible.

[0107] Recently, 454 Sequencing technology was used to determine the complete sequence of an individual's genome at a cost of approximately $2,000,000 (Wheeler, et al. (2008) "The complete genome of an individual by massively parallel DNA sequencing." Nature 452: 872-876), a 5-fold reduction in costs compared to that of sequencing an individual's genome using Sanger dideoxy sequencing methods (Levy, et al., (2007) "The Diploid Genome Sequence of an Individual Human." *PLoS Biol* **5:** e254).

[0108] Currently, single-stranded DNAs are prepared for sequencing in the 454 system by nebulizing genomic DNA and performing Solid Phase Reversible Immobilization (SPRI) step to remove DNA fragments less than 300 base-pairs long prior to attaching the DNA fragments to capture beads (Wheeler, et al. (2008) "The complete genome of an individual by massively parallel DNA sequencing." Nature 452: 872-876, Supplementary Information). However, SPRI removes approximately 10% of the total DNA fragments. In contrast, the methods provided herein avoid the arbitrary loss of genomic sequences inherent in SPRI.

[0109] Single-stranded nucleic acid fragments, e.g., prepared using methods desribed above, can be sequenced using systems that include bridge amplification technologies, e.g., in which primers bound to a solid phase are used in the extension and amplification of solution phase target nucleic acid acids prior to SBS. (See, e.g., Mercier, et al. (2005) "Solid Phase DNA Amplification: A Brownian Dynamics Study of Crowding Effects." *Biophysical Journal* **89:** 32-42; Bing, et al. (1996) "Bridge Amplification: A Solid Phase PCR System for the Amplification and Detection of Allelic Differences in Single

IDT Exhibit 1020.1495

Copy Genes." Proceedings of the Seventh International Symposium on Human Identification, Promega Corporation Madison, WI.) Solexa sequencing, available from Illumina, is one such sequencing system.

[0110] Single-stranded nucleic acid fragments can be prepared for bridge amplification the following manner: First, unique adapter tags are attached to ends of single-stranded linear nucleic acids during sample preparation. Methods by which the tags are attached to the nucleic acids are not particularly limiting and can include the strategies used to produce tagged double-stranded nucleic acid fragments in the methods detailed above. The nucleic acids to which the adapters have been attached can then be amplified in a "bridged" amplification reaction on the surface of a flow cell. The flow cell surface is coated with single stranded oligonucleotides that correspond to the sequences of the adapters ligated to the linear single-stranded nucleic acids during sample preparation. The single-stranded, adapter-ligated nucleic acids are bound to the surface of the flow cell and exposed to reagents for polymerase-based extension. Priming occurs as the free/distal end of a ligated fragment "bridges" to a complementary oligonucleotide on the surface, and during the annealing step, the extension product from one bound primer forms a second bridge strand to the other bound primer. Repeated denaturation and extension results in localized amplification of single molecules in millions of unique locations, creating clonal "clusters" across the flow cell surface.

[0111] The flow cell is then placed in a fluidics cassette within a sequencing module, where primers, DNA polymerase, and fluorescently-labeled, reversibly terminated nucleotides, e.g., A, C, G, and T, are added to permit the incorporation of a single nucleotide into each clonal DNA in each cluster. Each incorporation step is followed by the high-resolution imaging of the entire flow cell to identify the nucleotides that were incorporated at each cluster location on the flow cell. After the imaging step, a chemical step is performed to deblock the 3' ends of the incorporated nucleotides to permit the subsequent incorporation of another nucleotide. Iterative cycles are performed to generate a series of images each representing a single base extension at a specific cluster. This system typically produces sequence reads of up to 20-50 nucleotides. Further details regarding this sequencing system are discussed in, e.g., Bennett, et al. (2005) "Toward the 1,000 dollars human genome." *Pharmacogenomics* **6:** 373–382; Bennett, S. (2004) "Solexa Ltd."

IDT Exhibit 1020.1496

Pharmacogenomics **5:** 433–438; and Bentley, D. R. (2006) "Whole genome re-sequencing." *Curr Opin Genet Dev* **16:** 545-52.

[0112] Single molecule real-time sequencing (SMRT) is another massively parallel sequencing technology that can be used to sequence single-stranded nucleic acid fragments or loops, e.g., produced by any of the methods described herein, in a high-throughput manner. Developed and commercialized by Pacific Biosciences, SMRT technology relies on arrays of multiplexed zero-mode waveguides (ZMWs) in which, e.g., thousands of sequencing reactions can take place simultaneously. The ZMW is a structure that creates an illuminated observation volume that is small enough to observe, e.g., the template-dependent synthesis of a single single-stranded DNA molecule by a single DNA polymerase (See, e.g., Levene, et al. (2003) "Zero Mode Waveguides for Single Molecule Analysis at High Concentrations," *Science* **299:** 682-686). When a DNA polymerase incorporates complementary, fluorescently labeled nucleotides into the DNA strand that is being synthesized, the enzyme holds each nucleotide within the detection volume for tens of milliseconds, e.g., orders of magnitude longer than the amount of time it takes an unincorporated nucleotide to diffuse in and out of the detection volume. During this time, the fluorophore emits fluorescent light whose color corresponds to the nucleotide base's identity. Then, as part of the nucleotide incorporation cycle, the polymerase cleaves the bond that previously held the fluorophore in place and the dye diffuses out of the detection volume. Following incorporation, the signal immediately returns to baseline and the process repeats. Additional descriptions of ZMWs and their application in single molecule analyses, such as SMRT sequencing can be found in, e.g., Published U.S. Patent Application No. 2003/0044781, and U.S. Patent No. 6,917,726, each of which is incorporated herein by reference in its entirety for all purposes. See also, Levene et al. (2003) "Zero Mode Waveguides for single Molecule Analysis at High Concentrations," Science 299:682-686 and Eid, et al. (2009) "Real-Time DNA Sequencing from Single Polymerase Molecules." *Science* **323:**133-138.

FURTHER DETAILS REGARDING BROADLY USED MOLECULAR BIOLOGY TECHNIQUES

Preparing Genomic DNA

[0113] Determining the nucleotide sequence of an organism's genome can be useful in a myriad of applications, including, e.g., guiding biological and medical research, identifying an organism's susceptibility to disease, predicting an organism's ecological

IDT Exhibit 1020.1497

niche, providing a rational basis for personalized medicine, and others. Because reference genome sequences for many organisms are now publicly available, cataloging sequence variations and understanding their biological consequences has become a major research goal. The invention provides novel methods of producing single-stranded nucleic acids, e.g., linear nucleic acids or nucleic acid loops, from a genomic DNA that can be used to provide templates to a sequencing reaction, e.g., in a high-throughput sequencing system.

[0114] Genomic DNA can be prepared from any source, e.g., eukaryotic, prokaryotic, archaeal, viral, etc., by three steps: cell lysis, deproteinization and recovery of DNA. These steps are adapted to the demands of the application, the requested yield, purity and molecular weight of the DNA, and the amount and history of the source. Further details regarding the isolation of genomic DNA can be found in Berger and Kimmel, Guide to Molecular Cloning Techniques, Methods in Enzymology volume 152 Academic Press, Inc., San Diego, CA (Berger); Sambrook et al., Molecular Cloning - A Laboratory Manual (3rd Ed.), Vol. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 2008 ("Sambrook"); Current Protocols in Molecular Biology, F.M. Ausubel et al., eds., Current Protocols, a joint venture between Greene Publishing Associates, Inc. and John Wiley & Sons, Inc ("Ausubel"); Kaufman et al. (2003) Handbook of Molecular and Cellular Methods in Biology and Medicine Second Edition Ceske (ed) CRC Press (Kaufman); and The Nucleic Acid Protocols Handbook Ralph Rapley (ed) (2000) Cold Spring Harbor, Humana Press Inc (Rapley). In addition, many kits are commercially available for the purification of genomic DNA from cells, including Wizard™ Genomic DNA Purification Kit, available from Promega; Aqua Pure™ Genomic DNA Isolation Kit, available from BioRad; Easy-DNA™ Kit, available from Invitrogen; and DnEasy™ Tissue Kit, which is available from Qiagen.

Preparing cDNA

[0115] Alternative splicing (AS) is a major source of protein diversity in higher eukaryotic organisms, and this process is frequently regulated in a developmental stage-specific or tissue-specific manner. Thus, an understanding of changes in splicing patterns can be critical to a comprehensive understanding of biological regulation and disease. Data obtained from sequencing cDNAs can be useful in identifying novel splice variants of a gene of interest and/or in comparing the differential expression of splice isoforms of a gene of interest, e.g., between different tissue types, between different treatments to the same

IDT Exhibit 1020.1498

tissue type or between different developmental stages of the same tissue type. The methods for preparing single-stranded nucleic acids, e.g., linear nucleic acids or nucleic acid loops, that are provided by the invention can be beneficially used to produce templates derived from cDNAs to high throughput sequencing systems.

[0116] cDNAs are prepared from mRNA. mRNA can typically be isolated from almost any source using protocols and methods described in, e.g., Sambrook and Ausubel. The yield and quality of the isolated mRNA can depend on, e.g., how a tissue is stored prior to RNA extraction, the means by which the tissue is disrupted during RNA extraction, or on the type of tissue from which the RNA is extracted. RNA isolation protocols can be optimized accordingly. Many mRNA isolation kits are commercially available, e.g., the mRNA-ONLY™ Prokaryotic mRNA Isolation Kit and the mRNA-ONLY™ Eukaryotic mRNA Isolation Kit (Epicentre Biotechnologies), the FastTrack 2.0 mRNA Isolation Kit (Invitrogen), and the Easy-mRNA Kit (BioChain). In addition, mRNA from various sources, e.g., bovine, mouse, and human, and tissues, e.g. brain, blood, and heart, is commercially available from, e.g., BioChain (Hayward, CA), Ambion (Austin, TX), and Clontech (Mountainview, CA).

[0117] Once the purified mRNA is recovered, reverse transcriptase is used to generate cDNAs from the mRNA templates. Methods and protocols for the production of cDNA from mRNAs, e.g., harvested from prokaryotes as well as eukaryotes, are elaborated in *cDNA Library Protocols*, I. G. Cowell, et al., eds., Humana Press, New Jersey, 1997, Sambrook and Ausubel. In addition, many kits are commercially available for the preparation of cDNA, including the Cells-to-cDNA™ II Kit (Ambion), the RETROscript™ Kit (Ambion), the CloneMiner™ cDNA Library Construction Kit (Invitrogen), and the Universal RiboClone® cDNA Synthesis System (Promega). Many companies, e.g., Agencourt Bioscience and Clontech, offer cDNA synthesis services.

Preparing DNA Concatamers

[0118] Short sequence tags can be linked together to from long serial molecules termed "concatamers" that can be prepared, e.g., using the methods described herein, for sequencing, e.g., using a high-throughput sequencing system, e.g., a ZMW. A short sequence tag, e.g., 10-14bp, can contain sufficient information to uniquely identify a transcript, provided that that the tag is obtained from a unique sequence within the transcript. Quantitation of the number of times a particular tag is observed provides the

IDT Exhibit 1020.1499

expression level of the corresponding transcript. Thus, sequencing the nucleic acid templates, e.g. prepared according to the methods provided by the invention, derived from concatenated short ESTs, e.g., using a high-throughput sequencing system, can be useful in analyzing global gene expression patterns of, e.g., a tissue at different developmental stages, tissues in different organs from a common genotype, common tissues of different genotypes, common tissues that have been exposed to different treatments, and the like. In addition, sequencing templates, e.g., produced using method described herein, derived from concatamers of short ESTs can eliminate the need for a practitioner to carry out laborious and time-consuming *in vivo* cloning and cell culturing techniques that are common for other EST-based systems for the analysis of global gene expression, e.g. SAGE (Velculescu, et al. (1995) "Serial analysis of gene expression." *Science* **270:** 484-487) and TALEST (Spinella, et al (1999) "Tandem arrays ligation of expressed sequence tags (TALEST): a new method for generating global gene expression profiles." *Nucl Acid Res* **27:** e22).

[0119] Preparing concatenated ESTs can comprise preparing a cDNA library, e.g., as described above. Typically, the prepared cDNA can then be digested with a restriction enzyme that would be expected to cleave most transcripts at least once, e.g., a restriction enzyme with a 4-base pair recognition site. The 3'-most cDNA fragments are then captured and ligated to adapter molecules that each contain a type-II restriction site, e.g., BsgI, and a second restriction site. Digestion of the adapter-ligated cDNAs, e.g., with BsgI, produces DNA fragments that consist of the adapter itself and an additional 10-12 nucleotides of unknown cDNA sequence separated from the adapter by the restriction site originally used to digest the cDNA. The fragments can then be ligated to a second adapter containing a second restriction site at one end and degenerate overhangs, e.g., which render the second adapter compatible with all possible cDNA sequences, e.g., produced by the BsgI digestion, at the other. The resulting double-tagged DNA molecules can be digested with enzymes that recognize the restriction sites on the adapters and ligated together to form concatamers that can then be prepared, e.g., using the methods described herein, for sequencing, e.g., using a high-throughput system. Additional information and methods describing the preparation of concatamers comprising short ESTs can be found in, e.g., Velculescu, et al. (1995) "Serial analysis of gene expression." *Science* **270:** 484-487; Spinella, et al (1999) "Tandem arrays ligation of expressed sequence tags (TALEST): a new method for generating global gene expression profiles." *Nucl Acid Res* **27:** e22; WIPO

IDT Exhibit 1020.1500

Patent Application Number WO/2004/024953; and Unneberg, et al. (2003) "Transcript identification by analysis of short sequence tags – influence of tag length, restriction site, and transcript database." *Nucl Acids Res* **31:** 2217-2226.

Generating Nucleic Acid Fragments

[0120] The methods of preparing single-stranded nucleic acids that are described herein entail generating double-stranded fragments from, e.g., a genomic DNA, a cDNA, or a DNA concatamer. There exist a plethora of ways of generating nucleic acid fragments from a genomic DNA, a cDNA, or a DNA concatamer. These include, but are not limited to, mechanical methods, such as sonication, mechanical shearing, nebulization, hydroshearing, and the like; enzymatic methods, such as exonuclease digestion, restriction endonuclease digestion, and the like; and electrochemical cleavage. These methods are further explicated in Sambrook and Ausubel.

Copying Nucleic Acids

[0121] In certain embodiments of the methods described herein, populations of double-stranded nucleic acid fragments are produced by copying subsequences, e.g. overlapping subsequences, of a genomic DNA, a cDNA, or a DNA concatamer. A variety of nucleic acid amplification and/or copying methods are known in the art and can be implemented to perform these steps.

[0122] The most widely used *in vitro* technique among these methods is polymerase chain reaction (PCR), which requires the addition of a template of interest, e.g., a DNA comprising the sequence that is to be amplified, nucleotides, oligonucleotide primers, buffer, and an appropriate polymerase to an amplification reaction mix. In PCR, the primers anneal to complementary sequences on denatured template DNA and are extended with a thermostable DNA polymerase to copy the sequence of interest. As a result, a nucleic acid that comprises a sequence complementary to that of the template strand (or "target strand") is synthesized. Repeated cycles of PCR can generate myriad copies. Primers ideally comprise sequences that are complementary to the template. However, they can also comprise sequences that are not complementary, but which comprise e.g., restriction sites, cis regulatory sites, oligonucleotide hybridization sites, protein binding sites, DNA promoters, RNA promoters, sample or library identification sequences, and the like. Primers can comprise modified nucleotides, such as methylated, biotinylated, or fluorinated nucleotides; and nucleotide analogs, such as dye-labeled nucleotides, non-hydrolysable nucleotides, and nucleotides comprising heavy atoms. Primers can be custom

IDT Exhibit 1020.1501

synthesized by commercial suppliers as described below. PCR can be a useful means by which to attach tags to fragments. Further details regarding PCR and its uses are described in *PCR Protocols A Guide to Methods and Applications* (Innis *et al.* eds) Academic Press Inc. San Diego, CA (1990) (Innis); Chen et al. (ed) *PCR Cloning Protocols, Second Edition* (Methods in Molecular Biology, volume 192) Humana Press; and in Viljoen et al. (2005) *Molecular Diagnostic PCR Handbook* Springer, ISBN 1402034032.

[0123] Additional methods that can be used to amplify, or copy, nucleic acids include strand displacement amplification (SDA), multiple-displacement amplification (MDA), rolling circle replication (RCR). Some methods use RCR to copy single-stranded nucleic acids, e.g., which will be used as templates in sequencing reactions, from double-stranded templates. In RCR, DNA replication is initiated by an initiator protein, e.g., cis A, which nicks one strand of the double-stranded, closed DNA loop at a specific nucleotide sequence called the double-strand origin, or DSO. The initiator protein remains bound to the 5' phosphate end of the nicked strand, and the free 3' hydroxyl end is released to serve as a primer for DNA synthesis by DNA polymerase III. Using the un-nicked strand as a template, replication proceeds around the DNA loop, displacing the nicked strand as single-stranded DNA. Displacement of the nicked strand is carried out by a replisome, e.g., a multiprotein complex that comprises a single-stranded DNA binding protein (SSB), a helicase, a polymerase, and an RCR initiation protein, e.g., cisA.

[0124] Further details regarding Rolling Circle Amplification can be found in Demidov, et al. (2002) "Rolling-circle amplification in DNA diagnostics: the power of simplicity," *Expert Rev Mol Diagn* **2:** 89-94; Demidov and Broude (eds) (2005) *DNA Amplification: Current Technologies and Applications.* Horizon Bioscience, Wymondham, UK; and Bakht et al. (2005) "Ligation-mediated rolling-circle amplification-based approaches to single nucleotide polymorphism detection" *Expert Rev Mol Diagn* **5:** 111-116; Koonin, et al. (1993) "Computer-assisted dissection of rolling circle DNA replication." *BioSystems* **30:** 241-268; and Novick (1998) "Contrasting Lifestyles of rolling-circle phages and plasmids." *TIBS* **23:** 434- 438.

[0125] Copying steps in the methods can be performed with a strand-displacing polymerase. The term "strand displacement" describes the ability of a polymerase to displace downstream DNA encountered during synthesis. Examples of strand-displacing polymerases that can be used with the methods include, e.g., a Phi29 polymerase, a PolI

IDT Exhibit 1020.1502

polymerase, a BstI polymerase, or a Phi29-like polymerases, such as those described in United States Patent Application No. 11/645,223, entitled POLYMERASES FOR NUCLEOTIDE ANALOGUE INCORPORATION.

Nucleic Acid Tags

[0126] In some methods of nucleic acid template preparation provided by this invention, tags, e.g., tags comprising phosphorylated 5' ends, can be added to ends of nucleic acid fragments in order to protect the fragments from, e.g., degradation by an exonuclease. In certain other methods, tags comprising specific sequences can be added to the ends of fragments so that the fragments can be, e.g., recognized by and nicked by, an e.g., an RCR protein. In other methods, tags comprising, e.g., complementary single stranded terminal sequences, can be added to the ends of fragments to, e.g., promote more efficient circularization of the fragments. The addition of tags, e.g., via ligation, chemical linkage, primer extension, etc. to the ends of nucleic acid fragments is a useful step in the methods for producing single-stranded nucleic acids that are provided by the invention.

[0127] Nucleic acid tags can comprise any of a plethora of ligands, such as high-affinity DNA-binding proteins; modified nucleotides, such as methylated, biotinylated, or fluorinated nucleotides; and nucleotide analogs, such as dye-labeled nucleotides, non-hydrolysable nucleotides, or nucleotides comprising heavy atoms. Such reagents are widely available from a variety of vendors, including Perkin Elmer, Jena Bioscience and Sigma-Aldrich. Nucleic acid tags can also include oligonucleotides that comprise specific sequences, such as restriction sites, cis regulatory sites, oligonucleotide hybridization sites, protein binding sites, and the like. Such oligonucleotide tags can be custom synthesized by commercial suppliers such as Operon (Huntsville, AL), IDT (Coralville, IA) and Bioneer (Alameda, CA). The methods that can be used to join tags to nucleic acids of interest include chemical linkage, ligation, and extension of a primer by a polymerase (described above). Further details regarding nucleic acid tags and the methods by which they are attached to nucleic acids of interest are elaborated in Sambrook and Ausubel.

KITS AND ARTICLES OF MANUFACTURE

[0128] Kits are also a feature of the invention. The present invention provides kits that incorporate the compositions of the invention, optionally with additional useful reagents such as one or more enzymes that are used in the methods, e.g., an nicking enzyme polymerase, a DNA polymerase, an RCR protein, etc., that can be unpackaged in a fashion

IDT Exhibit 1020.1503

to enable their use. Depending upon the desired application, the kits of the invention optionally include additional reagents, such as a control nucleic acids, buffer solutions and/or salt solutions, including, e.g., divalent metal ions, i.e., Mg^{2+}, Mn^{2+}, Ca^{2+} and/or Fe^{++}, to prepare the single-stranded nucleic acids produced by the methods for sequencing, e.g., in a high-throughput sequencing system. Such kits also typically include a container to hold the kit components, instructions for use of the compositions, and other reagents in accordance with the desired application methods.

[0129] While the foregoing invention has been described in some detail for purposes of clarity and understanding, it will be clear to one skilled in the art from a reading of this disclosure that various changes in form and detail can be made without departing from the true scope of the invention. For example, all the techniques and apparatus described above can be used in various combinations. All publications, patents, patent applications, and/or other documents cited in this application are incorporated by reference in their entirety for all purposes to the same extent as if each individual publication, patent, patent application, and/or other document were individually indicated to be incorporated by reference for all purposes.

IDT Exhibit 1020.1504

CLAIMS

WHAT IS CLAIMED IS:

1. A method of producing a population of linear single-stranded nucleic acids, the method comprising:

providing a genomic DNA, a cDNA, or a DNA concatamer,

generating double-stranded fragments from the genomic DNA, the cDNA, or the concatamer, wherein first strands of the fragments are exonuclease-sensitive and second strands of the fragments are exonuclease-resistant; and,

eliminating the first strands from the double-stranded fragments to produce the population of linear single-stranded nucleic acids.

2. The method of claim 1, wherein the genomic DNA, the cDNA, or the concatamer is derived from a eukaryote.

3. The method of claim 1, wherein generating the double-stranded fragments that comprise the first exonuclease-sensitive strands and the second exonuclease-resistant strands comprises:

cleaving the genomic DNA, the cDNA, or the concatamer to produce linear double-stranded nucleic acids, and;

ligating exonuclease-sensitive tags to the 5' ends of the first strands of the linear double-stranded nucleic acids and ligating exonuclease-resistant tags to the 5' ends of the second strands of the linear double-stranded nucleic acids, thereby producing the double-stranded fragments with the first exonuclease-sensitive strands and the second exonuclease-resistant strands.

4. The method of claim 3, wherein cleaving the genomic DNA, the cDNA, or the concatamer comprises one or more of: enzymatic digestion, sonication, mechanical shearing, electrochemical cleavage, or nebulization of the genomic DNA, the cDNA, or the concatamer to produce the linear double-stranded nucleic acids.

5. The method of claim 3, wherein the tags comprise one or more moieties selected from: a ligand, a fluorescent label, a blocking group, a phosphorylated nucleotide, a phosphorothioated nucleotide, a biotinylated nucleotide, a methylated nucleotide, a nucleotide analog, a uracil, a sequence capable of forming a secondary structure, an oligonucleotide hybridization site, a restriction site, a DNA promoter, an RNA promoter, a sample or library identification sequence, and a cis regulatory sequence.

IDT Exhibit 1020.1505

6. The method of claim 1, wherein generating the double-stranded fragments that comprise the first exonuclease-sensitive strands and the second exonuclease-resistant strands comprises:

annealing a population of primer pairs, which primer pairs comprise first primers with exonuclease-sensitive 5' ends and second primers with exonuclease-resistant 5' ends, to subsequences of the genomic DNA, the cDNA, or the concatamer; and,

extending the primers with a polymerase to produce the double stranded fragments with the first exonuclease-sensitive strands and the second exonuclease-resistant strands.

7. The method of claim 6, wherein the primers comprise one or more moieties selected from: a ligand, a fluorescent label, a blocking group, a phosphorylated nucleotide, a phosphorothioated nucleotide, a biotinylated nucleotide, a methylated nucleotide, a uracil, a nucleotide analog, a sequence capable of forming a secondary structure, an oligonucleotide hybridization site, a restriction site, a DNA promoter, an RNA promoter, a sample or library identification sequence, and a cis regulatory sequence.

8. The method of claim 1, wherein the first strands are eliminated by digestion with an exonuclease, thereby producing the population of linear single-stranded nucleic acids.

9. A composition comprising:

a population of double-stranded nucleic acid fragments, which fragments comprise overlapping subsequences of a genomic DNA, a cDNA, or a DNA concatamer, wherein the fragments comprise first exonuclease-sensitive strands and second exonuclease-resistant strands; and,

an exonuclease.

10. The composition of claim 9, wherein the nucleic acid fragments comprise tags, which tags include one or more moieties selected from: a ligand, a fluorescent label, a blocking group, a phosphorylated nucleotide, a phosphorothioated nucleotide, a biotinylated nucleotide, a methylated nucleotide, a uracil, a nucleotide analog, a sequence capable of forming a secondary structure, an oligonucleotide hybridization site, a restriction site, a DNA promoter, an RNA promoter, a sample or library identification sequence, and a cis regulatory sequence.

11. A method of preparing closed single-stranded nucleic acid loops, the method comprising:

providing a genomic DNA, a cDNA, or a DNA concatamer;

IDT Exhibit 1020.1506

generating double-stranded fragments from the genomic DNA, the cDNA, or the concatamer, wherein first strands of the fragments are exonuclease-sensitive and second strands of the fragments are exonuclease-resistant; and,

separating the first strands of the fragments from the second strands to produce single-stranded fragments; and,

circularizing the single-stranded fragments to produce the closed single-stranded nucleic acid loops.

12. The method of claim 11, wherein the genomic DNA, the cDNA, or the concatamer is derived from a eukaryote.

13. The method of claim 11, wherein generating the double-stranded fragments that comprise the first exonuclease-sensitive strands and the second exonuclease-resistant strands comprises:

cleaving the genomic DNA, the cDNA, or the concatamer to produce linear double-stranded nucleic acids, and;

ligating exonuclease-sensitive tags to 5' ends of first strands of the linear double-stranded nucleic acids and ligating exonuclease-resistant tags to 5' ends of second strands of the linear double-stranded nucleic acids, thereby producing the population of double-stranded fragments with the first strands that comprise the exonuclease-sensitive 5' ends and the second strands that comprise the exonuclease-resistant 5' ends.

14. The method of claim 13, wherein cleaving the genomic DNA, the cDNA, or the concatamer comprises one or more of: enzymatic digestion, sonication, mechanical shearing, electrochemical cleavage, or nebulization of the genomic DNA, the cDNA, or the concatamer to produce the linear double-stranded nucleic acids.

15. The method of claim 13, wherein the tags comprise one or more moieties selected from: a ligand, a fluorescent label, a blocking group, a phosphorylated nucleotide, a phosphorothioated nucleotide, a biotinylated nucleotide, a methylated nucleotide, a nucleotide analog, a uracil, a sequence capable of forming a secondary structure, an oligonucleotide hybridization site, a restriction site, a DNA promoter, an RNA promoter, a sample or library identification sequence, and a cis regulatory sequence.

16. The method of claim 11, wherein generating the double-stranded fragments that comprise the first exonuclease-sensitive strands and the second exonuclease-resistant strands comprises:

IDT Exhibit 1020.1507

annealing a population of primer pairs, which primer pairs comprise first primers that comprise exonuclease-sensitive 5' ends and second primers that comprise exonuclease-resistant 5' ends, to subsequences of the genomic DNA, the cDNA or the concatamer; and,

extending the primers with a polymerase to produce the double-stranded fragments with the first strands that comprise the exonuclease-sensitive 5' ends and the second strands that comprise the exonuclease-resistant 5' ends.

17. The method of claim 16, wherein the primers comprise one or more moieties selected from: a ligand, a fluorescent label, a blocking group, a phosphorylated nucleotide, a phosphorothioated nucleotide, a biotinylated nucleotide, a methylated nucleotide, a nucleotide analog, a uracil, a sequence capable of forming a secondary structure, an oligonucleotide hybridization site, a restriction site, a DNA promoter, an RNA promoter, a sample or library identification sequence, and a cis regulatory sequence.

18. The method of claim 11, wherein separating the first strands from the second strands comprises digesting the first strands of the double-stranded fragments with an exonuclease to produce the single-stranded fragments.

19. The method of claim 11, wherein circularizing the single-stranded fragments to produce the closed single-stranded nucleic acid loops comprises:

annealing single-stranded nucleic acid splints to the single-stranded fragments, wherein first ends of the splints comprise a first nucleotide sequence complementary to a second nucleotide sequence at the first ends of the single-stranded fragments, and wherein second ends of the splints comprise a third nucleotide sequence complementary to a fourth nucleotide sequence at the second ends of the single-stranded fragments, such that the first and the second ends of each single-stranded fragment are brought into proximity of one another;

ligating the first and the second ends of the single-stranded fragments to one another; and,

removing the splints, thereby producing the closed single-stranded nucleic acid loops.

20. The method of claim 19, wherein removing the splints comprises digesting the splints with an exonuclease.

21. A composition, comprising:

a plurality of single-stranded nucleic acid fragments, which fragments comprise overlapping subsequences of a genomic DNA, a cDNA, or a DNA concatamer; and,

IDT Exhibit 1020.1508

a population of single-stranded nucleic acid splints, wherein first ends of the splints are annealed to first ends of the fragments and second ends of the splints are annealed to seconds end of the fragments, such that the first and second ends of each fragment are brought within proximity of one another.

22. The composition of claim 21, wherein the nucleic acid fragments comprise tags, which tags include one or more moieties selected from: a ligand, a fluorescent label, a blocking group, a phosphorylated nucleotide, a phosphorothioated nucleotide, a biotinylated nucleotide, a methylated nucleotide, a uracil, a nucleotide analog, a sequence capable of forming a secondary structure, an oligonucleotide hybridization site, a restriction site, a DNA promoter, an RNA promoter, a sample or library identification sequence, and a cis regulatory sequence.

23. The composition of claim 21, wherein the composition comprises a ligase.

24. A method of producing a population of closed single-stranded nucleic acid loops, the method comprising:

providing a genomic DNA, a cDNA, or a DNA concatamer;

generating double-stranded nucleic acid fragments from the genomic DNA, the cDNA, or the concatamer, which fragments comprise a first strand that comprises a gap site;

circularizing the double-stranded nucleic acid fragments to produce nucleic acid loops that comprise first non-contiguous strands and second contiguous strands; and,

removing the non-contiguous strands from the double-stranded loops, thereby producing a population of closed single-stranded nucleic acid loops.

25. The method of claim 24, wherein the genomic DNA, the cDNA, or the concatamer is derived from a eukaryote.

26. The method of claim 24, wherein generating the double-stranded fragments that comprise the gap site comprises:

cleaving the genomic DNA, the cDNA, or the concatamer to produce linear double-stranded nucleic acids, and;

ligating tags, which tags comprise the gap site, to the ends of the linear double-stranded nucleic acids to produce the double-stranded fragments that comprise the gap site.

27. The method of claim 26, wherein cleaving the genomic DNA, the cDNA, or the concatamer comprises: enzymatic digestion, sonication, mechanical shearing,

IDT Exhibit 1020.1509

electrochemical cleavage, or nebulization of the genomic DNA, the cDNA, or the concatamer to produce the linear double-stranded nucleic acids.

28. The method of claim 26, wherein the tags comprise one or more moieties selected from: a ligand, a fluorescent label, a blocking group, a phosphorylated nucleotide, a phosphorothioated nucleotide, a biotinylated nucleotide, a methylated nucleotide, a nucleotide analog, a uracil, a sequence capable of forming a secondary structure, an oligonucleotide hybridization site, a restriction site, a DNA promoter, an RNA promoter, a sample or library identification sequence, and a cis regulatory sequence.

29. The method of claim 24, wherein generating the double-stranded fragments that comprise the gap site comprises:

annealing a population of primer pairs, which primers comprise the gap site, to subsequences of the genomic DNA, the cDNA, or the concatamer; and,

extending the primers with a polymerase, thereby producing the generating the double-stranded fragments that comprise the gap site.

30. The method of claim 29, wherein the primer pairs comprise one or more moieties selected from: a ligand, a fluorescent label, a blocking group, a phosphorylated nucleotide, a phosphorothioated nucleotide, a biotinylated nucleotide, a methylated nucleotide, a nucleotide analog, a uracil, a sequence capable of forming a secondary structure, an oligonucleotide hybridization site, a restriction site, a DNA promoter, an RNA promoter, a sample or library identification sequence, and a cis regulatory sequence.

31. The method of claim 24, wherein the gap site comprises an unphosphorylated 5'end

32. The method of claim 24, wherein the gap site comprises a sequence recognized by a nicking enzyme.

33. The method of claim 32, wherein the nicking enzyme is uracil-DNA N-glycosylase (UNG), uracil-DNA glycosylase (UDG) or a site-specific restriction endonuclease that has been engineered to cleave the gap site.

34. The method of claim 24, wherein the gap site comprises an unphosphorylated 5' and wherein circularizing the double-stranded nucleic acid fragments to produce double-stranded nucleic acid loops that comprise first non-contiguous strands and second contiguous strands comprises ligating first ends of the fragments to second ends of the fragments, thereby forming the population of nucleic acid loops that comprise first non-contiguous strands and second contiguous strands.

IDT Exhibit 1020.1510

35. The method of claim 24, wherein the gap site comprises a sequence recognized by a nicking enzyme and wherein circularizing the double-stranded nucleic acid fragments to produce double-stranded nucleic acid loops that comprise first non-contiguous strands and second contiguous strands comprises:

ligating first ends of the fragments to second ends of the fragments, thereby forming a population of double-stranded loops; and,

cleaving the strands of the loops that comprise the gap site with a nicking enzyme that specifically cleaves the gap site, thereby forming the population of nucleic acid loops that comprise first non-contiguous strands and second contiguous strands

36. The method of claim 24, wherein removing the non-contiguous strands from the double-stranded loops comprises digesting the non-contiguous strands with an exonuclease, thereby producing the population of closed single-stranded nucleic acid loops.

37. The method of claim 24, further comprising sequencing the single-stranded nucleic acid loops.

38. A composition, comprising a plurality of double-stranded nucleic acid loops that comprise overlapping sequences of a genomic DNA, a cDNA, or a DNA concatamer, wherein the loops comprise first non-contiguous strands and second contiguous strands.

39. The composition of claim 38, wherein the loops comprise tags, which tags include one or more moieties selected from: a ligand, a fluorescent label, a blocking group, a phosphorylated nucleotide, a phosphorothioated nucleotide, a biotinylated nucleotide, a methylated nucleotide, a uracil, a nucleotide analog, a sequence capable of forming a secondary structure, an oligonucleotide hybridization site, a restriction site, a DNA promoter, an RNA promoter, a sample or library identification sequence, and a cis regulatory sequence.

40. The composition of claim 38, wherein the composition comprises an exonuclease.

41. A method of producing a population of closed single-stranded nucleic acid loops, the method comprising:

providing a genomic DNA, a cDNA, or a DNA concatamer;

generating double-stranded fragments from the genomic DNA, the cDNA, or the concatamer, which fragments comprise first sacrificial strands with first ends that comprise a rolling-circle replication (RCR) protein recognition sequence and second target strands;

IDT Exhibit 1020.1511

copying target strands of the double-stranded fragments to produce a population of single-stranded copies, and;

circularizing the single-stranded copies to produce the population of closed single-stranded nucleic acid loops.

42. The method of claim 41, wherein the genomic DNA, the cDNA, or the concatamer is derived from a eukaryote.

43. The method of claim 41, wherein generating the population of double-stranded fragments with the ends that comprise the rolling-circle replication (RCR) protein recognition sequence comprises:

cleaving the genomic DNA, the cDNA, or the concatamer to produce linear double-stranded nucleic acids, and;

ligating tags that comprise the rolling-circle replication (RCR) protein recognition sequence to first ends of first strands of the linear double-stranded nucleic acids to produce the double-stranded fragments with the ends that comprise the rolling-circle replication (RCR) protein recognition sequence.

44. The method of claim 43, wherein cleaving the genomic DNA, the cDNA, or the concatamer comprises: enzymatic digestion, sonication, mechanical shearing, electrochemical cleavage, or nebulization of the genomic DNA, the cDNA, or the concatamer to produce the linear double-stranded nucleic acids.

45. The method of claim 43, wherein the tags comprise one or more moieties selected from: a ligand, a fluorescent label, a blocking group, a phosphorylated nucleotide, a phosphorothioated nucleotide, a biotinylated nucleotide, a methylated nucleotide, a nucleotide analog, a uracil, a sequence capable of forming a secondary structure, an oligonucleotide hybridization site, a restriction site, a DNA promoter, an RNA promoter, a sample or library identification sequence, and a cis regulatory sequence.

46. The method of claim 41, wherein the double stranded fragments with the ends that comprise the rolling-circle replication (RCR) protein recognition sequence are generated by:

annealing a population of primer pairs that comprise first primers that comprise the rolling-circle replication (RCR) protein recognition sequence to subsequences of the genomic DNA, the cDNA, or the concatamer; and,

extending the primers with a polymerase, thereby producing the fragments with the ends that comprise the rolling-circle replication (RCR) protein recognition sequence.

IDT Exhibit 1020.1512

47. The method of claim 46, wherein the primers comprise one or more moieties selected from: a ligand, a fluorescent label, a blocking group, a phosphorylated nucleotide, a phosphorothioated nucleotide, a biotinylated nucleotide, a methylated nucleotide, a nucleotide analog, a uracil, a sequence capable of forming a secondary structure, an oligonucleotide hybridization site, a restriction site, a DNA promoter, an RNA promoter, a sample or library identification sequence, and a cis regulatory sequence.

48. The method of claim 41, wherein the RCR protein recognition sequence comprises a sequence that is bound by an RCR protein that comprises a histidine-U-histidine-U-U-U amino acid motif, wherein U is a bulky hydrophobic amino acid residue.

49. The method of claim 48 wherein the RCR protein is cisA.

50. The method of claim 41, wherein copying the target strands comprises:

nicking sacrificial strands of the double-stranded fragments,

displacing the sacrificial strands, and

replicating sequences of the target strands to produce the population of single-stranded copies.

51. The method of claim 50, wherein nicking the sacrificial strands of the double-stranded fragments comprises cleaving the strands with an RCR protein.

52. The method of claim 50, wherein displacing the sacrificial strands comprises unwinding the sacrificial strands from the target strands of the double-stranded fragments with a replisome.

53. The method of claim 52, wherein the replisome comprises a single-stranded DNA binding protein (SSB), a helicase, a polymerase, and an RCR protein.

54. The method of claim 50, wherein the replicating comprises synthesizing nucleic acid strands that are complementary to the target strands to produce the population of single-stranded copies.

55. The method of claim 41, wherein circularizing the single-stranded copies comprises ligating first ends of the single-stranded copies to second ends of the single-stranded copies with an RCR protein to produce the closed single-stranded nucleic acid loops.

56. A composition comprising a plurality of double-stranded nucleic acids wherein the nucleic acids comprise overlapping subsequences of a genomic DNA, a cDNA, or a DNA concatamer, and wherein the nucleic acids comprise ends that comprise a rolling-circle replication (RCR) protein recognition sequence.

IDT Exhibit 1020.1513

57. The composition of claim 56, wherein the nucleic acids comprise tags, which tags include one or more moieties selected from: a ligand, a fluorescent label, a blocking group, a phosphorylated nucleotide, a phosphorothioated nucleotide, a biotinylated nucleotide, a methylated nucleotide, a uracil, a nucleotide analog, a sequence capable of forming a secondary structure, an oligonucleotide hybridization site, a restriction site, a DNA promoter, an RNA promoter, a sample or library identification sequence, and a cis regulatory sequence.

58. The composition of claim 56, wherein the nucleic acids are linear.

59. The composition of claim 56, wherein the nucleic acids are closed loops.

60. The composition of claim 56, wherein the rolling-circle replication (RCR) protein recognition sequence comprises a sequence that is bound by an RCR protein that comprises a histidine-U-histidine-U-U-U amino acid motif, wherein U is a bulky hydrophobic amino acid residue.

61. The composition of claim 60, wherein the RCR protein is cisA.

62. A method of producing a population of single-stranded linear nucleic acids, the method comprising:

providing a genomic DNA, a cDNA, or a DNA concatamer;

generating a population of double-stranded fragments from the genomic DNA, the cDNA, or the concatamer, which fragments comprise first strands with first ends that comprise a sequence recognized by a nicking enzyme;

circularizing the fragments to produce a set of closed double-stranded nucleic acid loops,

cleaving sacrificial strands of the closed loops with the nicking enzyme to produce nicked loops;

copying target strands of the nicked loops to produce copied strands that comprise sequences that are recognized by the nicking enzyme; and,

nicking the copied strands with the nicking enzyme to produce the single-stranded linear nucleic acids.

63. The method of claim 62, wherein the genomic DNA, the cDNA, or the concatamer is derived from a eukaryote.

IDT Exhibit 1020.1514

64. The method of claim 62, wherein generating the population of double-stranded fragments with the ends that comprise the sequence recognized by the nicking enzyme comprises:

cleaving the genomic DNA, the cDNA, or the concatamer to produce linear double-stranded nucleic acids, and;

ligating tags, which tags comprise the sequence recognized by the nicking enzyme, to the first ends of the first strands of the linear double-stranded nucleic acids to produce the double-stranded fragments that comprise the first strands with the first ends that comprise the sequence recognized by the nicking enzyme.

65. The method of claim 64, wherein cleaving the genomic DNA, the cDNA, or the concatamer comprises: enzymatic digestion, sonication, mechanical shearing, electrochemical cleavage, or nebulization of the genomic DNA, the cDNA, or the concatamer to produce the double-stranded fragments.

66. The method of claim 64, wherein the tags comprise one or more moieties selected from: a ligand, a fluorescent label, a blocking group, a phosphorylated nucleotide, a phosphorothioated nucleotide, a biotinylated nucleotide, a methylated nucleotide, a nucleotide analog, a uracil, a sequence capable of forming a secondary structure, an oligonucleotide hybridization site, a restriction site, a DNA promoter, an RNA promoter, a sample or library identification sequence, and a cis regulatory sequence.

67. The method of claim 62, wherein the double-stranded fragments with the ends that comprise the sequence recognized by the nicking enzyme are generated by:

annealing a population of primer pairs, which primer pairs comprise first primers with the sequence recognized by the nicking enzyme, to subsequences of the genomic DNA, the cDNA, or the concatamer; and,

extending the primers with a polymerase, thereby producing the double-stranded fragments with the first strands with the first ends that comprise the sequence recognized by the nicking enzyme.

68. The method of claim 67, wherein the primers comprise one or more moieties selected from: a ligand, a fluorescent label, a blocking group, a phosphorylated nucleotide, a phosphorothioated nucleotide, a biotinylated nucleotide, a methylated nucleotide, a nucleotide analog, a uracil, a sequence capable of forming a secondary structure, an oligonucleotide hybridization site, a restriction site, a DNA promoter, an RNA promoter, a sample or library identification sequence, and a cis regulatory sequence.

IDT Exhibit 1020.1515

69. The method of claim 62, wherein the circularizing comprises ligating first ends of the fragments to second ends of the fragments to produce the closed loops.

70. The method of claim 62, wherein the nicking enzyme is selected from the group consisting of: a uracil-DNA glycosylase, a uracil-DNA N-glycosylase, and a site-specific restriction endonuclease engineered to cut one strand

71. The method of claim 62, wherein the copying comprises:

displacing the sacrificial strands; and,

replicating the target strands, wherein the replicating comprises synthesizing nucleic acid strands that are complementary to the target strands, to produce the copied strands.

72. A composition comprising a plurality of closed double-stranded nucleic acid loops, wherein the loops comprise overlapping fragments of a genomic DNA, a cDNA, or a DNA concatamer and wherein the loops comprise a sequence that is specifically recognized by a nicking enzyme.

73. The composition of claim 72, wherein the loops comprise tags, which tags include one or more moieties selected from: a ligand, a fluorescent label, a blocking group, a phosphorylated nucleotide, a phosphorothioated nucleotide, a biotinylated nucleotide, a methylated nucleotide, a uracil, a nucleotide analog, a sequence capable of forming a secondary structure, an oligonucleotide hybridization site, a restriction site, a DNA promoter, an RNA promoter, a sample or library identification sequence, and a cis regulatory sequence.

74. The composition of claim 72, wherein the composition comprises a nicking enzyme.

75. A method of generating a set of single-stranded nucleic acid loops, the method comprising:

providing a genomic DNA, a cDNA, or a DNA concatamer;

producing a population of double-stranded nucleic acid fragments from the genomic DNA, the cDNA, or the concatamer;

providing first hairpin oligonucleotides to first ends of the double-stranded nucleic acid fragments and providing second hairpin oligonucleotides to second ends of the double-stranded nucleic acid fragments, thereby generating the set of single-stranded nucleic acid loops that comprise regions of internal complementarity.

IDT Exhibit 1020.1516

76. The method of claim 75, wherein the genomic DNA, the cDNA, or the concatamer is derived from a eukaryote.

77. The method of claim 75, wherein producing the population of double-stranded nucleic acid fragments comprises cleaving the genomic DNA, the cDNA, or the concatamer.

78. The method of claim 77, wherein cleaving the genomic DNA, the cDNA, or the concatamer comprises: enzymatic digestion, sonication, mechanical shearing, electrochemical cleavage, or nebulization of the genomic DNA, the cDNA, or the concatamer to produce the double-stranded fragments.

79. The method of claim 75, wherein producing the population of double-stranded nucleic acid fragments comprises:

annealing a population of primer pairs to subsequences of the genomic DNA, the cDNA, or the concatamer; and,

extending the primer pairs with a polymerase to produce the double-stranded fragments.

80. The method of claim 79, wherein the primers comprise one or more moieties selected from: a ligand, a fluorescent label, a blocking group, a phosphorylated nucleotide, a phosphorothioated nucleotide, a biotinylated nucleotide, a methylated nucleotide, a nucleotide analog, a uracil, a sequence capable of forming a secondary structure, an oligonucleotide hybridization site, a restriction site, a DNA promoter, an RNA promoter, a sample or library identification sequence, and a cis regulatory sequence.

81. The method of claim 75, wherein providing the hairpin oligonucleotides to the ends of the double-stranded nucleic acids fragments comprises:

annealing the first hairpin oligonucleotides to the first ends of the fragments;

annealing the second hairpin oligonucleotides to the second ends of the fragments; and,

ligating the hairpin oligonucleotides to the ends of the fragments to generate the set of single-stranded nucleic acid loops.

82. The method of claim75, wherein providing the hairpin oligonucleotides to the ends of the double-stranded nucleic acids fragments comprises:

linking 5'strands of the first hairpins to 3' strands of the first ends of the fragments;

linking 3'strands the first hairpins to 5' strands of the first ends of the fragments;

IDT Exhibit 1020.1517

linking 5' strands of the second hairpins to 3' strands of the second ends of the fragments; and,

linking 3' strands of the second hairpins to 5' strands of the second ends of the fragments to form a population of single-stranded nucleic acid loops.

83. The method of claim 75, wherein the first hairpins comprise first single-stranded terminal sequences complementary to second single-stranded terminal sequences at the first ends of the fragments and the second hairpins comprise third single-stranded terminal sequences complementary to fourth single stranded terminal sequences at the second ends of the fragments.

84. The method of claim 75, wherein the first or the second hairpins comprise one or more moieties selected from: a ligand, a fluorescent label, a blocking group, a phosphorylated nucleotide, a phosphorothioated nucleotide, a biotinylated nucleotide, a methylated nucleotide, a nucleotide analog, a uracil, a sequence capable of forming a secondary structure, an oligonucleotide hybridization site, a restriction site, a DNA promoter, an RNA promoter, a sample or library identification sequence, and a cis regulatory sequence.

85. The method of claim 75, wherein the first and second hairpins comprise identical nucleotide sequences.

86. The method of claim 75, wherein the first and second hairpins comprise first and second nucleotide sequences, wherein the first and second sequences are not identical.

87. A composition comprising first hairpin oligonucleotides, second hairpin oligonucleotides, and a plurality of double-stranded nucleic acid fragments, which fragments comprise overlapping subsequences of a genomic DNA, a cDNA, or a DNA concatamer.

88. The composition of claim 87, wherein the first and/or second hairpin oligonucleotides comprise one or more moieties selected from: a ligand, a fluorescent label, a blocking group, a phosphorylated nucleotide, a phosphorothioated nucleotide, a biotinylated nucleotide, a methylated nucleotide, a nucleotide analog, a uracil, a sequence capable of forming a secondary structure, an oligonucleotide hybridization site, a restriction site, a DNA promoter, an RNA promoter, a sample or library identification sequence, and a cis regulatory sequence.

89. The composition of claim 87, wherein the first hairpins comprise first single-stranded terminal sequences complementary to second single-stranded terminal sequences at

IDT Exhibit 1020.1518

the first ends of the fragments and the second hairpins comprise third single-stranded terminal sequences complementary to fourth single stranded terminal sequences at the second ends of the fragments.

90. The composition of claim 87, wherein the genomic DNA, the cDNA, or the concatamer is derived from a eukaryote.

91. The composition of claim 87, wherein the composition comprises a ligase.

92. A method of producing linear single-stranded nucleic acids, the method comprising:

providing a genomic DNA, a cDNA, or a DNA concatamer;

generating a set of double-stranded nucleic acid fragments from the genomic DNA, the cDNA, or the concatamer;

nicking sacrificial strands of the double-stranded nucleic acid fragments to produce nicked fragments;

displacing the sacrificial strands from the fragments; and,

copying sequences between nick sites and ends on un-nicked strands that are exposed by the displacement of the sacrificial strands with a strand displacing polymerase to produce the linear single-stranded nucleic acids.

93. The method of claim 92, wherein the genomic DNA, the cDNA, or the concatamer is derived from a eukaryote.

94. The method of claim 92, wherein producing the population of double-stranded nucleic acid fragments comprises cleaving the genomic DNA, the cDNA, or the concatamer.

95. The method of claim 94, wherein cleaving the genomic DNA, the cDNA, or the concatamer comprises one or more of: enzymatic digestion, sonication, mechanical shearing, electrochemical cleavage, or nebulization of the genomic DNA, the cDNA, or the concatamer to produce the double-stranded fragments.

96. The method of claim 92, wherein producing the population of double-stranded nucleic acid fragments comprises

annealing a population of primer pairs to subsequences of the genomic DNA, the cDNA, or the concatamer; and,

extending the primer pairs with a polymerase to produce the double-stranded fragments.

97. The method of claim 96, wherein the primers comprise one or more moieties selected from: a ligand, a fluorescent label, a blocking group, a phosphorylated nucleotide, a

IDT Exhibit 1020.1519

phosphorothioated nucleotide, a biotinylated nucleotide, a methylated nucleotide, a nucleotide analog, a uracil, a sequence capable of forming a secondary structure, an oligonucleotide hybridization site, a restriction site, a DNA promoter, an RNA promoter, a sample or library identification sequence, and a cis regulatory sequence.

98. The method of claim 92, wherein nicking the sacrificial strands comprises cleaving the sacrificial strands with an enzyme to produce the nicked fragments.

99. The method of claim 92, wherein displacing the sacrificial strands comprises removing the sacrificial strands from the un-nicked strands with the strand-displacing polymerase.

100. The method of claim 92, wherein the exposed sequences are copied by:

replicating sequences between the nick sites and the ends on the un-nicked strands with the strand-displacing polymerase, wherein replicating the sequences comprises synthesizing nucleic acid strands that are complementary to the un-nicked strands, to produce the linear single-stranded nucleic acids.

101. The method of claim 92, wherein the strand-displacing polymerase is a Phi29 polymerase, a Phi29-like polymerase, PolI polymerase, or a BstI polymerase.

IDT Exhibit 1020.1520

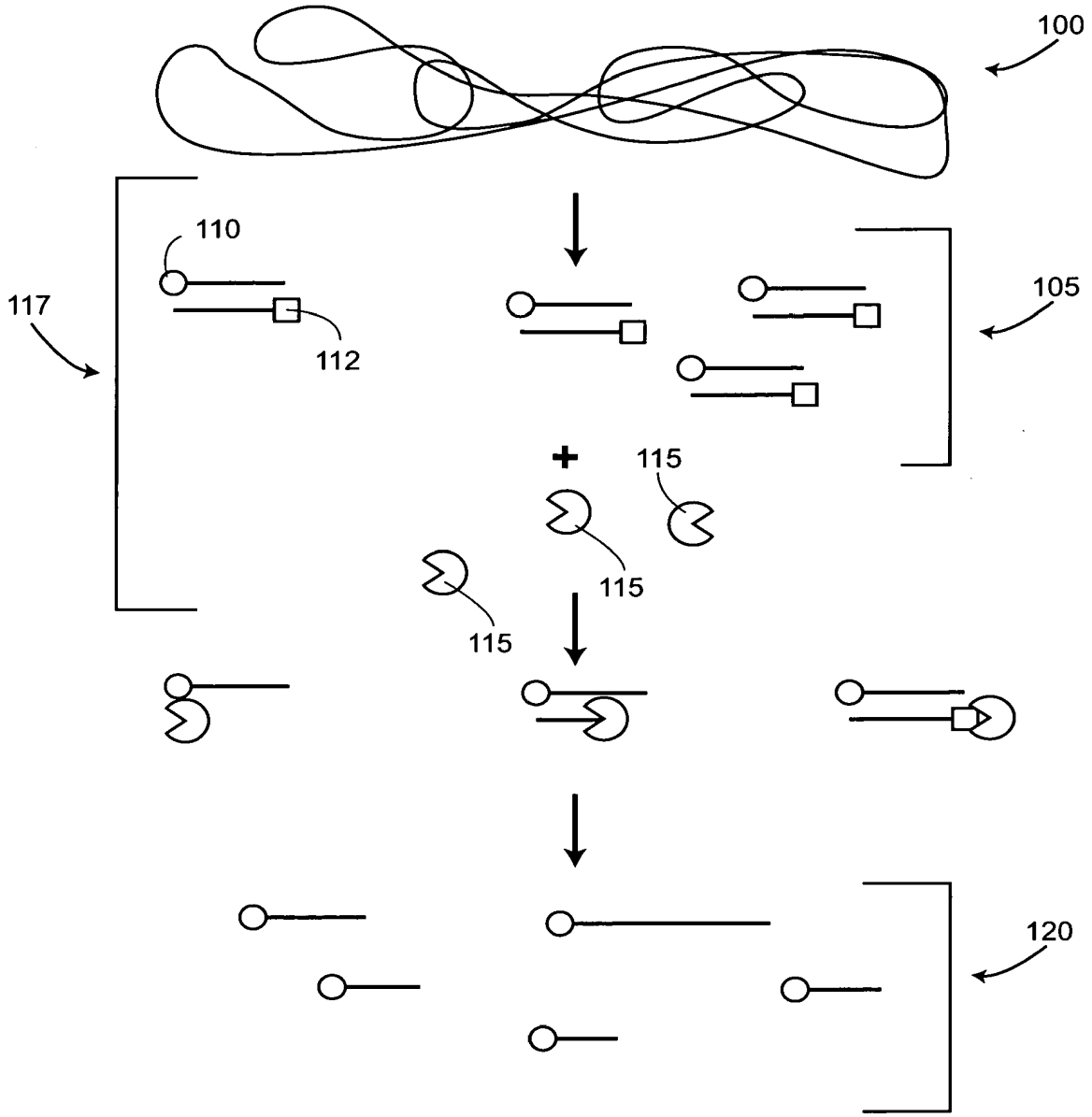


Fig. 1

IDT Exhibit 1020.1521

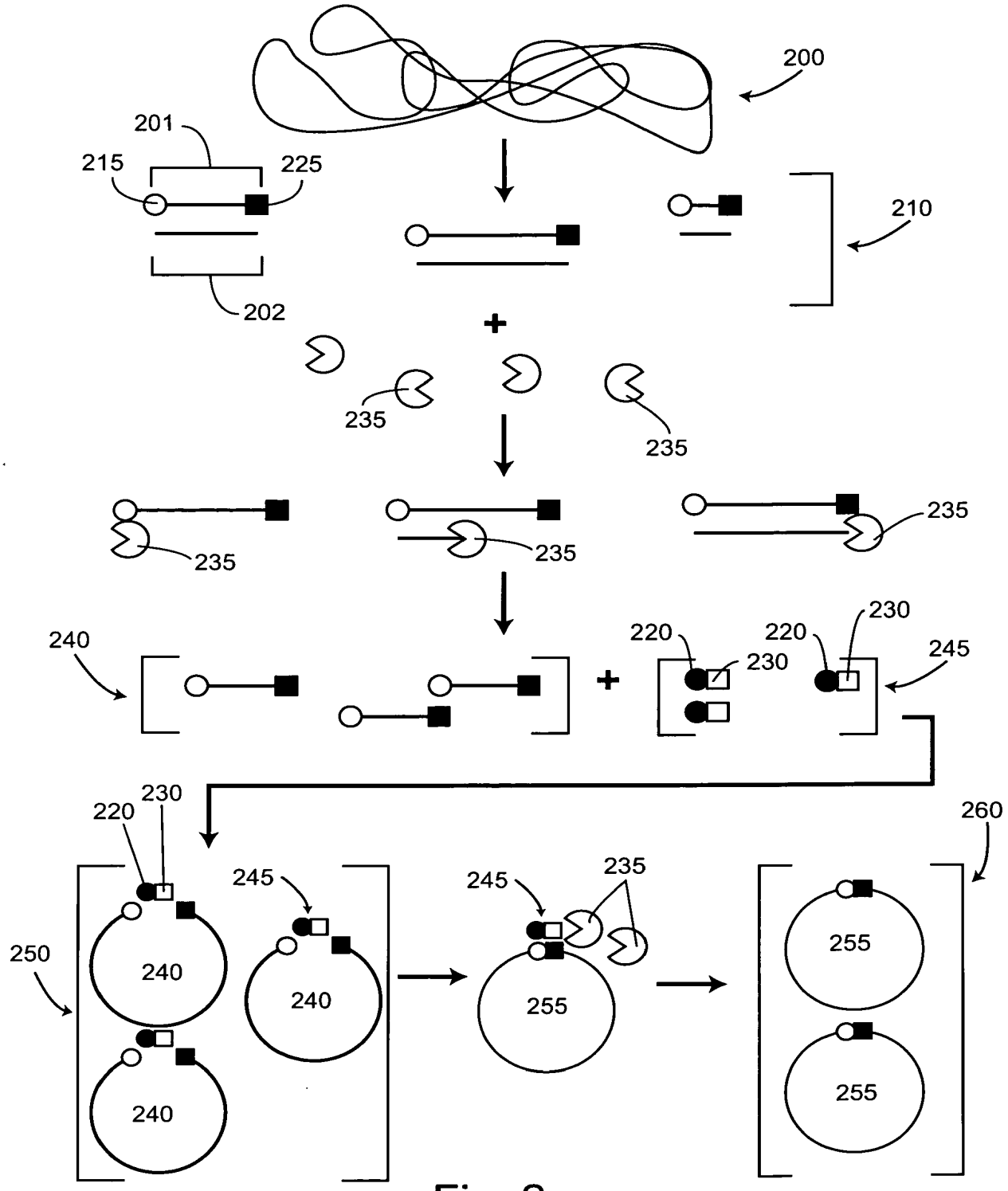


Fig. 2

IDT Exhibit 1020.1522

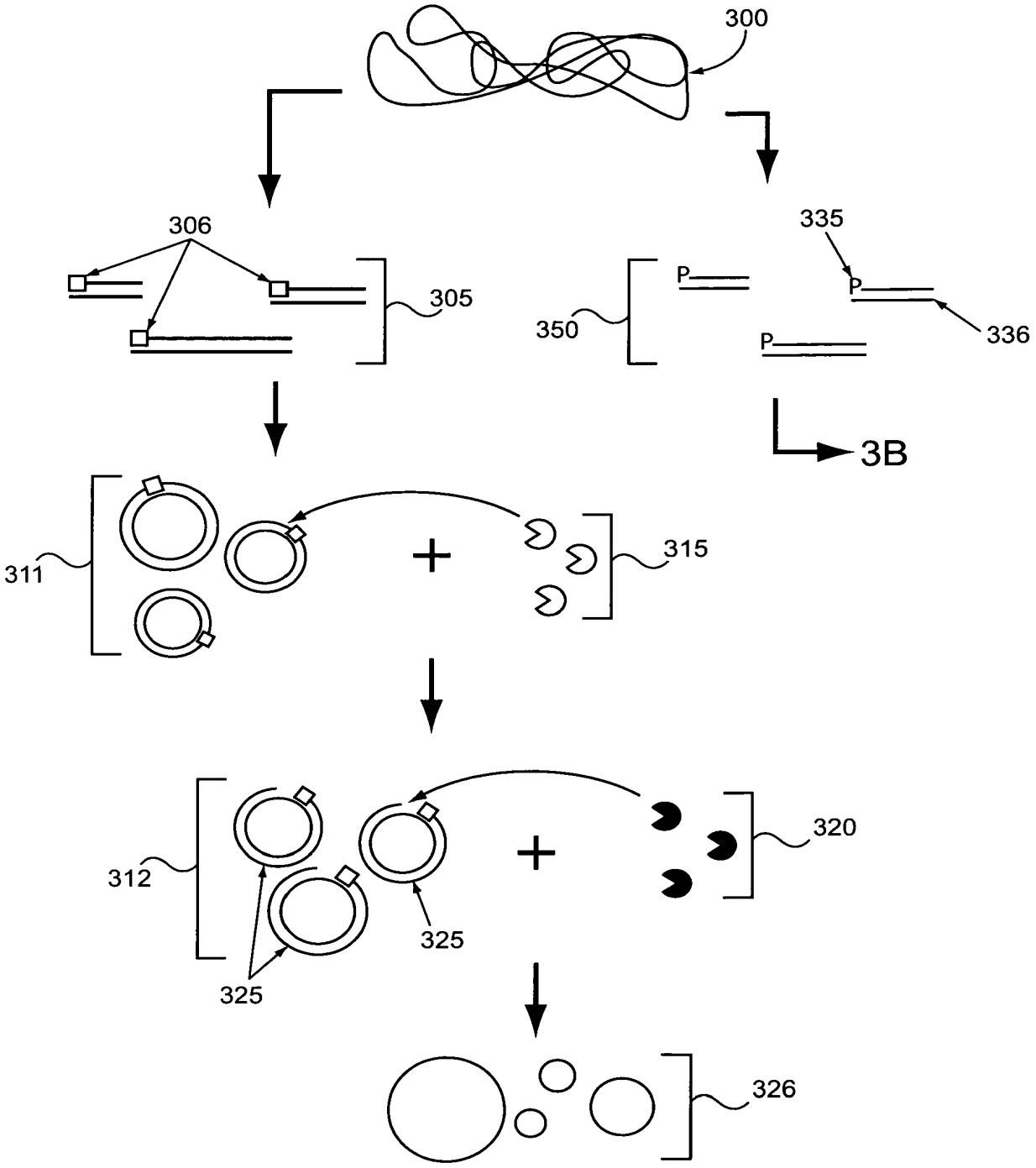


Fig. 3A

IDT Exhibit 1020.1523

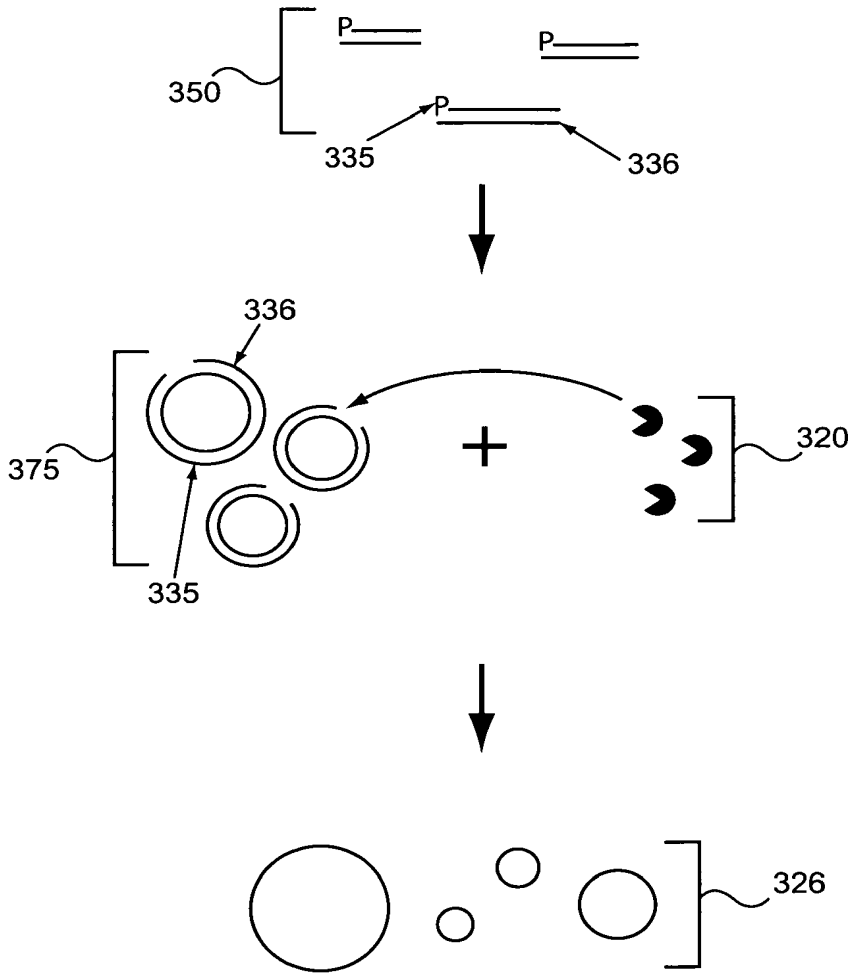


Fig. 3B

IDT Exhibit 1020.1524

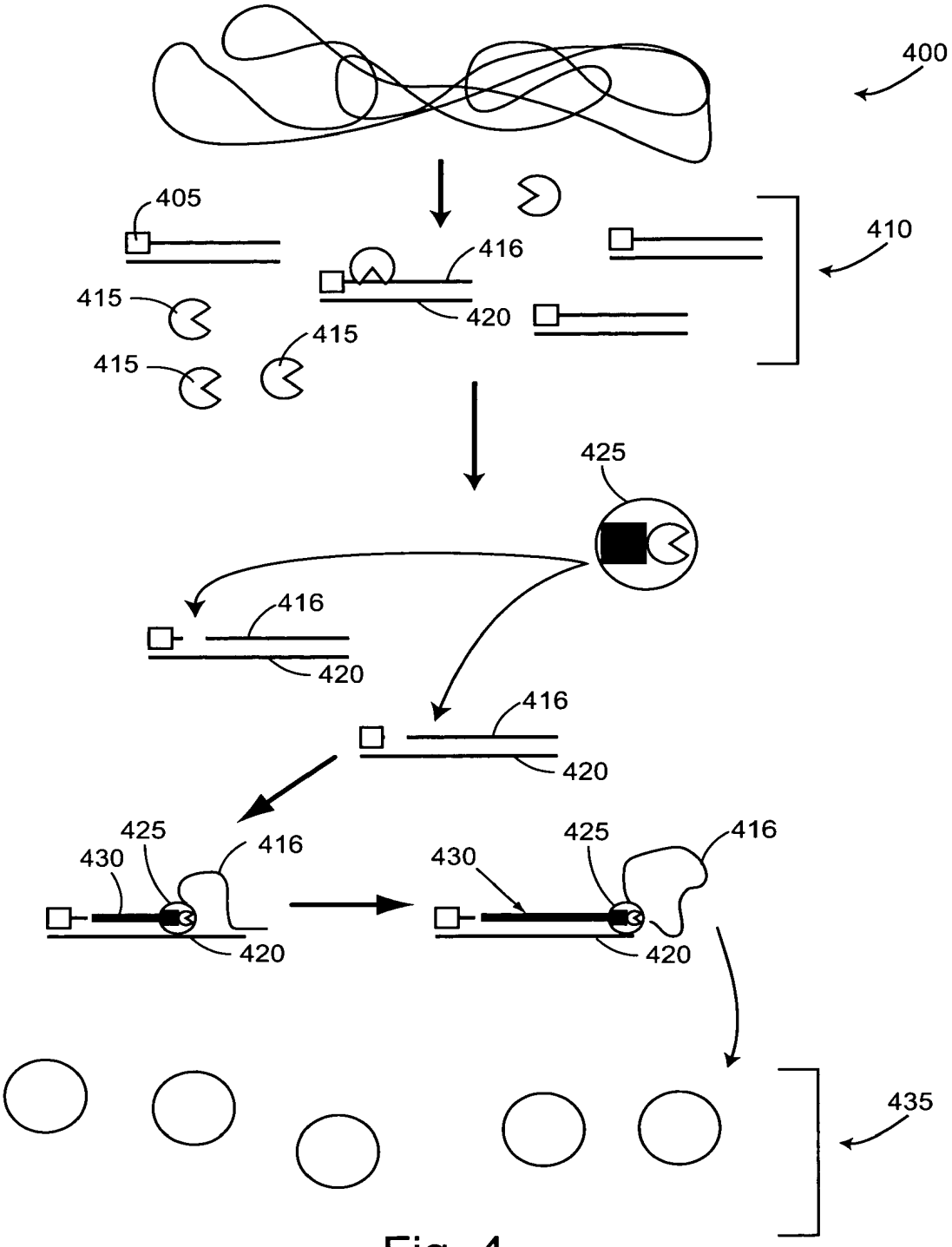


Fig. 4

IDT Exhibit 1020.1525

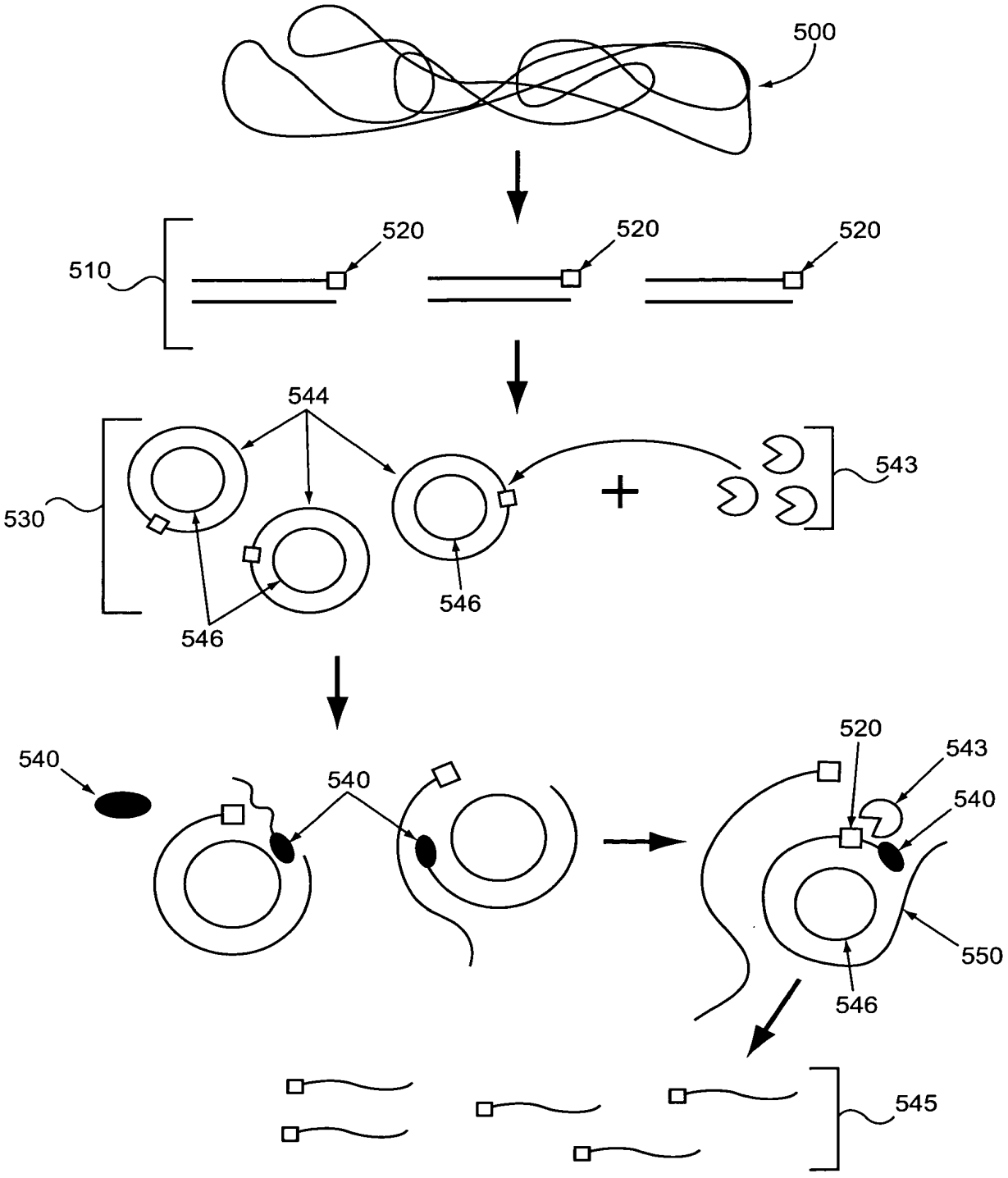


Fig. 5

IDT Exhibit 1020.1526

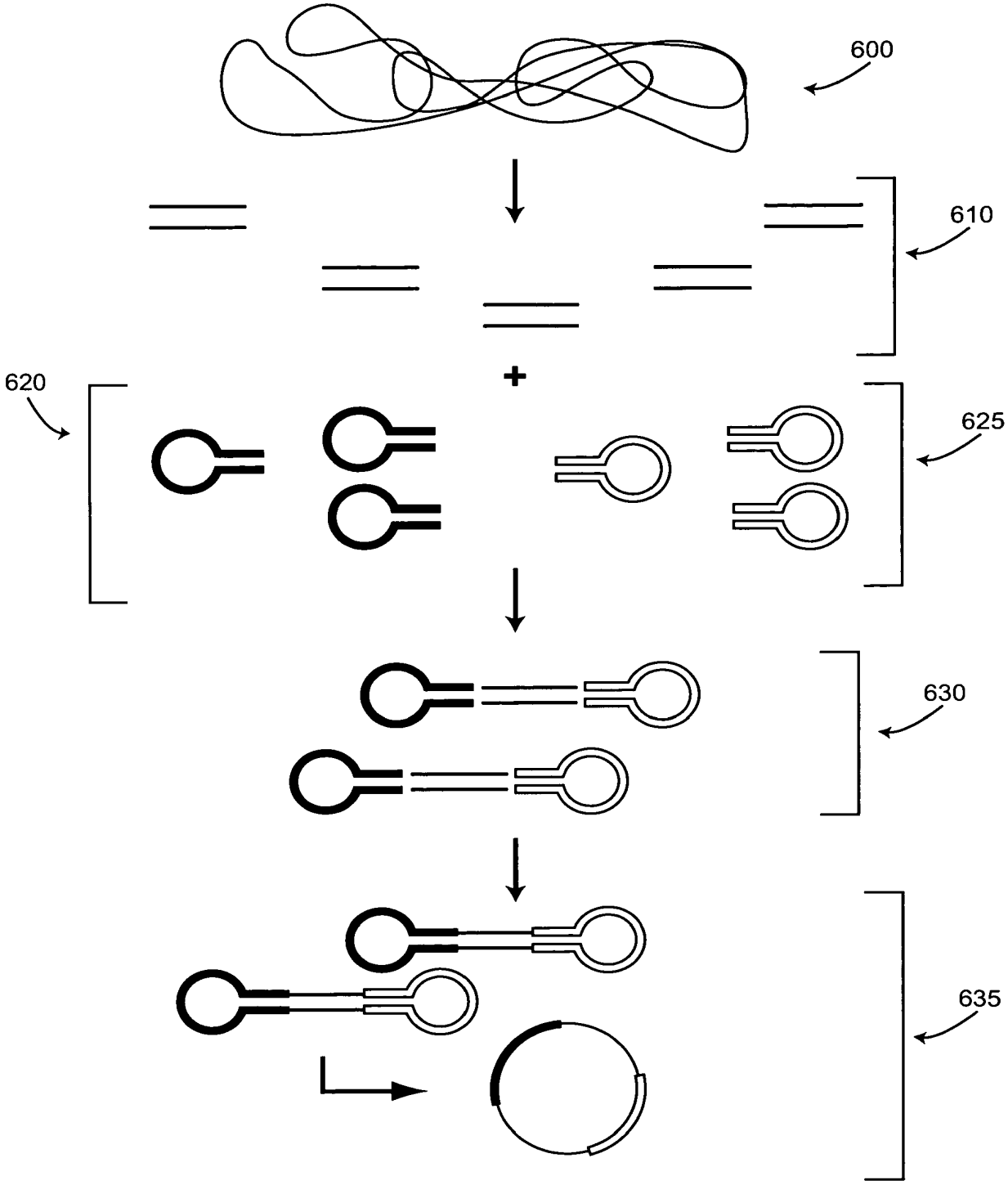


Fig. 6

IDT Exhibit 1020.1527

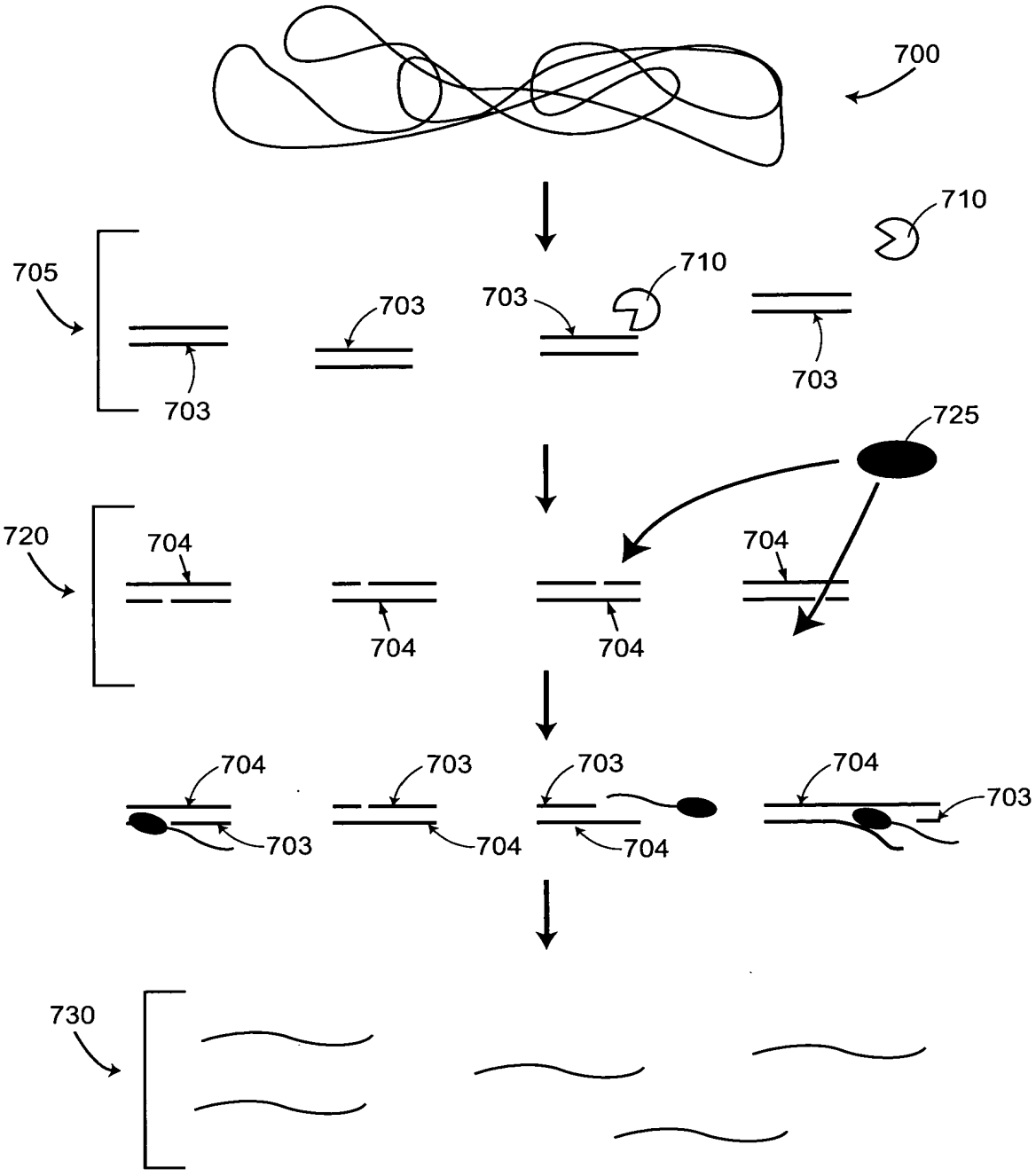


Fig. 7

IDT Exhibit 1020.1528

Espacenet

Bibliographic data: WO2010039991 (A2) — 2010-04-08

METHOD OF GENERATING INFORMATIVE DNA TEMPLATES FOR HIGH-THROUGHPUT SEQUENCING APPLICATIONS

Inventor(s): MULLET JOHN [US]; MORISHIGE DARYL [US] ± (MULLET, JOHN, ; MORISHIGE, DARYL)

Applicant(s): TEXAS A & M UNIV SYS [US]; MULLET JOHN [US]; MORISHIGE DARYL [US] ± (THE TEXAS A&M UNIVERSITY SYSTEM, ; MULLET, JOHN, ; MORISHIGE, DARYL)

Classification: **- international:** ***C07H21/00; C12N15/10; C12N15/11; C12Q1/68***
- cooperative: **C12Q1/6811** more

Application number: WO2009US59274 20091001 Global Dossier

Priority number (s): US20080102118P 20081002

Also published as: WO2010039991 (A3)

Abstract of WO2010039991 (A2)

Methods and compositions are disclosed for generating informative DNA templates for sequencing. The methods generally include fragmenting an individual's genomic DNA to generate templates with predetermined range of sequence lengths, ligating adapters; selecting informative DNA templates, which contain a polymorphism; and recovering enriched informative DNA templates. The DNA templates may then be sequenced using a high throughput sequencing process, and the resulting sequences compared to a reference set of sequences obtained from the genomes of other individuals, to identify DNA polymorphisms and to generate genotypes.

IDT Exhibit 1020.1529

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
8 April 2010 (08.04.2010)

PCT

(10) International Publication Number
WO 2010/039991 A2

(51) International Patent Classification:
C12N 15/10 (2006.01) *C12N 15/11* (2006.01)
C12Q 1/68 (2006.01) *C07H 21/00* (2006.01)

(21) International Application Number:
PCT/US2009/059274

(22) International Filing Date:
1 October 2009 (01.10.2009)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/102,118 2 October 2008 (02.10.2008) US

(71) Applicant *(for all designated States except US)*: **THE TEXAS A&M UNIVERSITY SYSTEM** [US/US]; 3369 Tamu, College Station, TX 77843-3369 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **MULLET, John** [US/US]; 809 Southern Hills Court, College Station, TX 77845 (US). **MORISHIGE, Daryl** [US/US]; 3902 E. 29th St., #B2, Bryan, TX 77802 (US).

(74) Agent: **HANSON, Robert, E.**; Sonnenschein Nath & Rosenthal, 7800 Sears Tower, 233 South Wacker Dr., Chicago, IL 60606 (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, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

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

Published:

— *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 2010/039991 A2

(54) **Title**: METHOD OF GENERATING INFORMATIVE DNA TEMPLATES FOR HIGH-THROUGHPUT SEQUENCING APPLICATIONS

(57) **Abstract**: Methods and compositions are disclosed for generating informative DNA templates for sequencing. The methods generally include fragmenting an individual's genomic DNA to generate templates with predetermined range of sequence lengths, ligating adapters; selecting informative DNA templates, which contain a polymorphism; and recovering enriched informative DNA templates. The DNA templates may then be sequenced using a high throughput sequencing process, and the resulting sequences compared to a reference set of sequences obtained from the genomes of other individuals, to identify DNA polymorphisms and to generate genotypes.

IDT Exhibit 1020.1530

TITLE OF INVENTION

METHOD OF GENERATING INFORMATIVE DNA TEMPLATES FOR HIGH-THROUGHPUT SEQUENCING APPLICATIONS

REFERENCE TO RELATED APPLICATIONS

This application claims priority to U.S. Provisional Application No. 61/102,118, filed on October 2, 2008, which is incorporated herein by reference in its entirety.

INCORPORATION OF SEQUENCE LISTING

The Sequence Listing, which is a part of the present disclosure, includes a computer readable 11 KB file created on September 25, 2009 entitled "TAMC011WO_ST25.txt" comprising nucleotide and/or amino acid sequences of the present invention submitted via EFS-Web. The subject matter of the Sequence Listing is incorporated herein by reference in its entirety.

TECHNICAL FIELD

The present invention generally relates to the generation of DNA templates from specific sites in genomes for high throughput sequence-based analysis, for such applications as genotyping, marker-assisted breeding, genetic mapping, haplotyping, physical map construction, and gene mapping. More particularly, the invention relates to methods for isolating and enriching the population of informative DNA templates for such high throughput sequencing applications.

BACKGROUND

A number of techniques have been described for discovering and detecting DNA polymorphisms. Most involve an initial DNA sequence polymorphism discovery step, which usually involves direct sequencing of cDNA or genes, or involves hybridization to oligonucleotide arrays, followed by the development of targeted single nucleotide polymorphism (SNP) or insertion/deletion (InDel) detection assays. Digestion of genomic DNA by restriction enzymes, and/or shearing, followed by adapter ligation, is a standard procedure for preparing amplifiable templates from genomes for a variety of uses, such as polymerase chain reaction (PCR)-based genotyping assays. One recent methodology known as amplified fragment length

IDT Exhibit 1020.1531

polymorphism (AFLP) employs PCR to amplify restriction fragments obtained from a complex mixture of DNA fragments that are prepared by the action of restriction endonucleases on genomic DNA. In some methodologies, the primers used for amplification of DNA, or to initiate sequence analysis of DNA, are not directed against a known genomic DNA sequence, but rather are designed such that they are complementary to sequences in adapters ligated to the ends of the restriction fragments. This strategy yields restriction site localized (RSL) DNA sequences, some of which include polymorphisms.

Bulk methods for isolating restriction fragments that contain polymorphisms recover the entire collection of restriction fragments generated by the selected restriction enzyme. However, not all of those DNA fragments can be used to generate useful genotyping information. First, the sequences adjacent to many of the restriction sites will be repetitive and therefore not useful for genotyping. Secondly, some restriction site localized (RSL) fragments will be too large or too small or contain secondary structures that make them poor templates for bridge amplification and sequencing. Thirdly, in genetic mapping applications, the only informative restriction fragments are those that contain DNA sequences spanning polymorphisms that distinguish parental genotypes. In addition to those challenges, secondary sources of DNA such as chloroplast, mitochondria, and contaminating bacteria or fungal genomes, may be present at relatively high copy number in some samples of a nuclear DNA selected for analysis. These non-target sources of DNA restriction fragments reduce the efficiency of bulk amplification and sequencing procedures.

There is continuing interest in developing ways to rapidly identify a large number of polymorphisms and mutations distributed over the entire expanse of large genomes or targeted to specific sites, to facilitate their use in a wide range of applications, including DNA fingerprinting, marker-assisted plant and animal breeding, genetic/quantitative trait locus (QTL) mapping, haplotyping, and other applications.

SUMMARY

Embodiments of the present methods comprise the sequential use of restriction enzymes and adapter ligation for template generation, followed by selection of informative templates by hybridization to oligonucleotides attached to a solid matrix, or by targeted PCR amplification, with subsequent direct sequence analysis. Embodiments of this specific combination of procedures provide a highly flexible, very

IDT Exhibit 1020.1532

low cost, and highly accurate way to obtain genotyping information. Compositions and kits for carrying out such methods are also provided in accordance with some embodiments of the invention.

In accordance with certain embodiments, a method of generating informative DNA templates for sequencing is provided that comprises: (a) obtaining a fragmented genomic DNA sequence from a first individual, to provide a mixture of DNA fragments, wherein the genomic DNA comprises a plurality of polymorphisms; (b) ligating at least a first adapter to the DNA fragments, to provide a mixture comprising adapter-modified informative DNA templates and adapter-modified non-informative DNA templates, wherein each said informative DNA template comprises a unique sequence in a location compatible with high throughput DNA sequencing of the template, wherein the unique sequence comprises a unique polymorphic site in the species genome sequence; and (c) selecting adapter-modified informative DNA templates (*e.g.,* by either hybridization-based selection or targeted PCR amplification of the adapter-modified informative templates), to obtain an enriched mixture of adapter-modified informative DNA templates. In some embodiments, a hybridization-based selection such as in step (c) involves forming hybridized complexes comprising the adapter-modified informative DNA templates, oligonucleotides and a solid matrix, and excluding the non-informative DNA templates. In some embodiments, the method further comprises (d) separating non-hybridized non-informative DNA templates from the hybridized complexes; and (e) releasing the informative DNA templates from the hybridized complexes, to obtain an enriched mixture of adapter-modified informative DNA templates. In some embodiments, forming the hybridized complexes such as in step (c) comprises hybridizing oligonucleotides to complementary sequences in the informative DNA templates.

In some embodiments, targeted PCR amplification of informative templates such as in step (c) involves a first primer complementary to the first adapter, and a set of second primer (or plurality of primers) complementary to one or more unique sequences in the informative DNA fragments, wherein each said second primer(s) is designed such that the resulting amplified DNA templates are of a predetermined sequence length (or range of lengths) and comprise informative DNA sequences.

In some embodiments, in step (b), ligating comprises ligating a second adapter to a terminus of each said DNA fragment opposite the first adapter, to provide the mixture of adapter-modified informative DNA templates and adapter-modified non-

IDT Exhibit 1020.1533

informative DNA templates. In the process, some templates are generated that comprise adapter A or adapter B ligated to both termini of some DNA fragments making them incompatible with some high throughput DNA sequencing technologies. Therefore, some embodiments further comprise the use of suppression PCR to amplify and enrich DNA templates comprising adapters A and B that are compatible with high throughput sequencing relative to templates containing only adapter A or adapter B that are not suitable for sequencing.

In some embodiments, a method according to the invention further comprises selecting adapter-modified informative DNA templates having a predetermined sequence length that is compatible DNA sequencing (*e.g.*, bridge amplification in the case of ILLUMINA® SGAII™ sequencing) and read length of a selected high throughput sequencing process.

In some embodiments, the method further comprises a step (d) for subjecting the enriched mixture of adapter-modified informative DNA templates to a high-throughput DNA sequencing procedure, to obtain the sequences of the informative DNA templates, and a step (e) comparing the sequences of the informative DNA templates to at least one set of reference genomic DNA sequences to identify the specific (*e.g.*, polymorphic) allele sequence obtained from each template or site in the genome. In certain embodiments, at least one set of reference genomic DNA sequences is obtained from at least one reference individual. In further embodiments, at least the first adapter comprises an indexing sequence that can be correlated to the first individual.

In accordance with still another embodiment of the invention, a reagent is provided for selecting informative DNA templates, comprising a solid matrix (*e.g.*, a plurality of beads) and a plurality of different oligonucleotides attached to the solid matrix. For example, each of the oligonucleotides may be in the range of about 17-60 nucleotides in length and is complementary to a unique sequence present in a respective informative DNA template. In certain aspects, each such informative DNA template comprises at least one polymorphism located within the read length of a selected high throughput sequencing process, wherein the location is measured from either terminus of the informative DNA template.

In accordance with another embodiment of the invention, a hybrid DNA complex is provided which comprises a reagent for selection of informative DNA templates as described herein and a plurality of adapter-modified informative DNA

IDT Exhibit 1020.1534

templates hybridized to the matrix-attached oligonucleotides. For instance, each such informative DNA template may comprise at least one polymorphism located within the read length of a selected high throughput sequencing process, wherein the location is measured from either terminus of the informative DNA template. In some aspects, the adapter-modified informative DNA templates comprise at least a first adapter ligated to the informative templates. In various embodiments, the informative DNA templates are obtained from a single individual, or are derived from a plurality of individuals, in which case the first adapter comprises a unique indexing sequence ligated to each of the informative DNA templates for matching it to the individual from which it was derived. These and other embodiments and potential advantages will be apparent with reference to the following description and drawings.

In still a further embodiment, the invention provides a method for marker-assisted selection. For instance, fragmented genomic DNA is obtained from a plurality individual plants or plant cells, to provide a plurality of genomic DNA fragments comprising a plurality of polymorphic sequences at least one of which is linked to a trait of interest. Fragmented genomic DNA is then ligated to at least a first adapter, to provide a plurality of adapter-modified informative DNA templates and adapter-modified non-informative DNA templates, wherein each of said informative DNA templates comprises a polymorphic sequence and wherein said first adapter comprises an index sequence that can be correlated to genomic DNA of an individual plant or plant cell. Adapter-modified informative DNA templates are selected by either hybridization-based selection or targeted PCR amplification of the adapter-modified informative DNA templates, to obtain an enriched mixture of adapter-modified informative DNA templates that can be sequenced. Based on the sequence an individual plant or plant cell is selected based on the presence of at least one polymorphism linked to a trait of interest. For example, selecting an individual plant or plant cell may comprise selecting a plant cell for regeneration of a plant and/or a plant comprising a trait of interest can be selected for commercial production or breeding. In certain aspects, a plant or plant cell for selection is a wheat, maize, rye, rice, oat, barley, turfgrass, sorghum, millet, sugarcane, tobacco, tomato, potato, soybean, cotton, canola, sunflower or alfalfa plant or plant cell. Thus, in certain aspects, the invention provides a method for marker-assisted selection of a genomic region or gene that regulates expression of a trait such as trait of agronomic interest in a plant (*e.g.,* a drought

IDT Exhibit 1020.1535

tolerance, enhanced yield, cold tolerance, pest resistance, insect resistance, salt tolerance or herbicide tolerance trait).

As used herein the specification, "a" or "an" may mean one or more. As used herein in the claim(s), when used in conjunction with the word "comprising", the words "a" or "an" may mean one or more than one.

The use of the term "or" in the claims is used to mean "and/or" unless explicitly indicated to refer to alternatives only or the alternatives are mutually exclusive, although the disclosure supports a definition that refers to only alternatives and "and/or." As used herein "another" may mean at least a second or more.

Throughout this application, the term "about" is used to indicate that a value includes the inherent variation of error for the device, the method being employed to determine the value, or the variation that exists among the study subjects.

Other objects, features and advantages of the present invention will become apparent from the following detailed description. It should be understood, however, that the detailed description and the specific examples, while indicating preferred embodiments of the invention, are given by way of illustration only, since various changes and modifications within the spirit and scope of the invention will become apparent to those skilled in the art from this detailed description.

BRIEF DESCRIPTION OF THE DRAWINGS

The following drawings form part of the present specification and are included to further demonstrate certain aspects of the present invention. The invention may be better understood by reference to one or more of these drawings in combination with the detailed description of specific embodiments presented herein.

FIG. 1 is a schematic flow diagram of a method of generating DNA polymorphism-enriched DNA templates for sequencing, in accordance with certain embodiments of the invention.

FIG. 2 is a box flow diagram summarizing template preparation, and illustrating the structure of representative adapter- and primer-modified DNA templates and sequencing complexes, in accordance with certain embodiments of the invention.

FIG. 3 is a box flow diagram of a method of generating enriched informative DNA templates, commencing with restriction enzyme digestion of genomic DNA and/or shearing, in accordance with certain embodiments of the invention.

IDT Exhibit 1020.1536

FIG. 4 is a box flow diagram of the steps following preparation and enrichment of informative templates that involve optional amplification, then sequencing and analysis of DNA sequences to identify polymorphisms and alleles.

FIG. 5 is a box flow diagram of a method of FIG. 1, including PCR amplification of informative DNA fragments, according to certain embodiments of the invention.

FIG. 6 illustrates an example of a specific 30 bp restriction site localized sequence from a sorghum genotype containing a unique polymorphism.

FIG. 7 illustrates an example of another specific 30 bp restriction site localized sequence from a sorghum genotype containing another unique polymorphism.

FIG. 8 illustrates an example of another specific 30 bp restriction site localized sequence from a sorghum genotype containing another unique polymorphism.

FIG. 9 illustrates an example of another specific 30 bp restriction site localized sequence from a sorghum genotype containing another unique polymorphism.

FIG. 10 illustrates an example of another specific 30 bp restriction site localized sequence from a sorghum genotype containing another unique polymorphism.

FIG. 11 illustrates an example of another specific 30 bp restriction site localized sequence from a sorghum genotype containing another unique polymorphism.

FIG. 12 illustrates an example of another specific 30 bp restriction site localized sequence from a sorghum genotype containing another unique polymorphism.

FIG. 13 illustrates an example of another specific 30 bp restriction site localized sequence from a sorghum genotype containing another unique polymorphism.

FIG. 14 illustrates an example of another specific 30 bp restriction site localized sequence from a sorghum genotype containing another unique polymorphism.

FIG. 15 illustrates an example of another specific 30 bp restriction site localized sequence from a sorghum genotype containing another unique polymorphism.

DETAILED DESCRIPTION

Restriction site localized (RSL)-sequences derived from nearly any genome will be a mixture of "unique" highly informative DNA sequences that occur once per genome, along with less informative DNA sequences that occur two or more times in the target genome. Many existing bulk methods for isolating RSL-templates recover the entire collection of RSL-templates generated by a specific targeting restriction enzyme. However, not all of those recovered RSL-templates can be used to generate

IDT Exhibit 1020.1537

useful genotyping information. First, the sequences adjacent to many of these restriction sites will be repetitive and therefore not useful for genotyping. Secondly, some RSL-templates will be too large or too small or contain DNA sequences with a propensity to form secondary structures that make them poor templates for bridge amplification, or for some other reason. Such size considerations are discussed in more detail below. Thirdly, in genetic mapping applications, the only informative sequences, and therefore informative RSL-templates, are those that correspond to DNA sequences spanning polymorphisms that distinguish parental genotypes. For example, if the parents used to create a population for genetic mapping purposes contain one sequence polymorphism per 1,000 bp, then only 1 specific template out of 40 RSL-templates sequenced at random will contain a polymorphic sequence, assuming 25 bp of sequence is obtained per template. In this instance, only the single DNA template containing the polymorphism is considered "informative," while the 39 other DNA templates in the mixture would be "uninformative," for the purposes of genetic mapping in the target population.

In certain embodiments described herein, direct hybridization (*e.g.*, to oligonucleotide-beads or microarrays) is utilized to select large numbers of RSL-templates that contain polymorphisms (*e.g.*, SNP/INDEL) for sequence-based genotyping. In other embodiments, a primer-targeted PCR-based enrichment is described that allows smaller numbers of polymorphism-containing RSL-templates to be amplified for sequence-based genotyping.

Embodiments of the present methods that provide targeted isolation of RSL-templates for sequencing, which contain informative polymorphic sequences, will potentially increase efficiency by approximately 100-fold or more (by removing repeats and monomorphic templates. This will potentially reduce the cost of genotyping considerably and expand the number and complexity of applications for this technology. Some applications require the collection of information from a large number of polymorphic sites per genome (*e.g.*, haplotype-based analysis which typically requires analysis of about 5,000 to 20,000 or more polymorphic sites per genome depending on linkage disequilibrium, whereas other applications, such as DNA fingerprinting, early phase genetic mapping or pedigree analysis require analysis of a relatively small number of polymorphic sites (approximately 500–2,000) per genome. Fine mapping genetic studies require higher density analysis (a large number of polymorphic sites analyzed per Mbp), but only in one or a few regions of the genome.

IDT Exhibit 1020.1538

Various embodiments of the methods described herein for sequence-based genotyping substantially enhance the output of these forms of genomic analysis.

Embodiments of the methods disclosed herein offer a flexible method to generate and perform a targeted analysis of subsets of RSL-templates generated by any specific set of restriction enzymes, allowing variation in the number analyzed per genome and their distribution across the genome (or based on their utility in larger population studies).

FIG. **1** is a box flow diagram of a method of generating a mixture or pool of informative DNA templates, and then sequencing those templates to identify DNA polymorphisms by comparison to a reference sequence, to generate genotypes and haplotypes. The method generally comprises the following stages or a subset thereof:

I. Fragment an individual genomic DNA to generate templates with predetermined range of sequence lengths.

II. Ligate adapters and select DNA template size.

III. Informative DNA template selection.

IV. Recovery of enriched informative DNA templates.

V. High throughput sequencing of the recovered DNA templates.

VI. Compare the resulting sequences to a reference set of sequences obtained from the genomes of one or more other individuals, to identify DNA polymorphisms and to generate a genotype.

Stage I. Fragmentation of the genomic DNA of an individual to generate templates of a predetermined sequence length.

The fragmentation stage is designed to produce DNA fragments having a specified complexity, sequence length and information content. Template size is dependent on two factors: (1) DNA fragment length suitable for processing by the high throughput sequencer used to collect sequence information *(e.g.,* an optimal size of about 50 to 250 bp, for uniform bridge-amplification on the ILLUMINA® SGAII™), and (2) DNA template length sufficient to utilize the sequence read length of the high throughput sequencer (*e.g.*, optimally, at least 35-50bp for the SGAII™ and at least 200-400bp for the Roche 454 GS-FLX™ Genome Analysis System). In general, this predetermined sequence length of the DNA fragments is less than that of a protein-coding region of a gene, and in most instances is also smaller than an average exon length. The predetermined template size may vary somewhat based on whether the

IDT Exhibit 1020.1539

fragments are generated by restriction enzymes or shearing, or by a combination of those techniques. For example, for restriction fragments, the distance of the DNA polymorphism away from the restriction enzyme cleavage site (where adapter A is added) is the primary determinant of the minimum fragment sequence length (as can be appreciated in the schematic illustration of Stage IIIa in FIG. 1). The "sequencer read length" of a given sequencing system refers to the number of bases of sequence information that can be acquired from each template sequenced.

The minimum sequence length of an informative DNA template is that which spans a DNA polymorphism and that can be uniquely located in a genome. A sequence as small as about 17 bp is sufficient to span a polymorphism and is typically large enough to often be a unique sequence in a genome, to permit identification of its location on a genetic map or within a genome sequence. Somewhat larger sequences (*e.g.*, 25-35) are employed in some embodiments, in that a relatively high percent of the sequences in this size range can be uniquely located in most genomes. When a high-throughput sequencer such as the ILLUMINA® SGAII™ or the ABI SOLID™ sequencer is used (*e.g.,* in Stage V), the sequence read length is approximately 35-50bp. Accordingly, the targeted DNA polymorphisms should be within about 35-50 bp of a restriction site where the adapter containing sequences used to initiate sequencing is located. In instances in which a high-throughput sequencer such as the Roche 454 sequencer is used as the sequencing system (*e.g.,* in Stage V), DNA polymorphisms within about 100–400 bp of restriction sites can be sequenced. If a DNA fragment is not within the read-length adjacent to the cleavage end obtained with one restriction enzyme, a second restriction enzyme with a different recognition sequence can be used in addition to the first enzyme, to generate template.

Alternatively, genomic DNA can be sheared to generate template with a random set of termini. Adapter A, that is later used to initiate sequencing reactions in Stage V, is ligated to DNA termini generated by shearing, and the informative templates in the resulting mixture of sheared fragments are then enriched as described below (hybridization-based or PCR-based methods). This approach to template preparation allows any region of any genome to be targeted for genotyping by direct sequence analysis.

In the resulting mixture of genomic fragments, only those that contain an approximately 17-400 bp stretch of DNA sequence located adjacent to a restriction enzyme recognition site or shearing site (*i.e.*, restriction site or cleavage site) must be

IDT Exhibit 1020.1540

unique in the target genome and span a polymorphism to be useful for genotyping. These restriction fragments or sheared fragments, once modified by the addition of suitable terminal sequences required for sequencing, are the "informative DNA templates."

In some embodiments, in which the DNA fragments are prepared entirely by shearing, the predetermined sequence length is in the range of about 35-50bp (*e.g.,* for the ILLUMINA® SGAII™ and ABI SOLID™ sequencers) or about 400bp (*e.g.,* for the Roche 454 sequencer, which has a comparatively longer sequence read length). In the case of DNA fragments generated by random shearing, the informative DNA templates are those that can be used to generate sequences unique in the target genome, that span DNA polymorphisms, and that have adapters ligated sufficiently close to the DNA polymorphism so that sequences spanning the target polymorphic sequence are acquired.

Informative DNA templates will in most cases be larger than the sequences ultimately derived from the templates by sequencing. For example, templates generated specifically for use with the SGAII™ sequencer may be 50-200bp in length, even though the SGAII™ only sequences about 35bp from one end of the template. Likewise, templates generated specifically for use with the Roche 454 sequencer may be about 400-800bp in length, even though only 200-400bp of sequence are acquired from one end of the template. In other instances, larger templates may be generated for sequencing from both ends of the template (using the SGAII™ sequencer, for instance). It should be understood that the templates and hybridization complexes illustrated in FIG. 1 are not drawn to scale, and the relative lengths of the oligonucleotides and the templates may vary greatly.

In some cases, the genomic DNA is fragmented in Stage I to a desired range of sequence lengths and complexity by digesting the DNA with one or more restriction enzymes (as schematically shown in Step 1 of FIG. 3). Complexity in this instance refers to the number of termini generated by digestion of a genome with a restriction enzyme and that correspond to informative sequences. This is accomplished, for example, with restriction enzymes having different recognition sequences (*e.g.*, 4 bp, 6 bp, 7 bp, 8 bp) to vary the number, complexity, repetitiveness, and GC content of the DNA templates generated.

In some variations of the Step 1 fragmentation protocol, the DNA is digested by one or more restriction enzyme that is sensitive to DNA methylation to prepare DNA

IDT Exhibit 1020.1541

templates enriched in templates derived from the gene rich regions of a genome. Enrichment for unique gene rich templates occurs because, in many organisms, repeat sequences are differentially methylated and therefore no template will be generated from these sites. In addition, by digesting DNA with a pair of restriction enzymes that recognize and digest the same DNA sequence (isoschizomers), where one of the pair of enzymes will cut methylated DNA and the second enzyme of the pair will not cut methylated DNA, DNA templates are provided for assaying differences in genome DNA methylation. More specifically, the DNA from two or more individuals is digested separately using each of the enzymes that recognizes and digests the same DNA sequence. One of the enzymes is sensitive to DNA methylation (*i.e.,* will not cut these sites), and the other enzyme is not sensitive to DNA methylation (*i.e.*, will cut all recognition sites, regardless of whether it is methylated or not). The individual DNAs are analyzed and tracked using different indexing sequences that can be used for identification. Differences in the complement of sequences derived from the two restriction enzymes correspond to differences in sites of methylation in the genome.

A subset of the polymorphisms that distinguish accessions or parental genotypes reside in the recognition sites of the restriction enzymes used for template preparation. Polymorphisms in a genotype that change the recognition sequence of the restriction enzyme used for template generation will prevent DNA digestion and as a consequence, no templates will be generated adjacent to that site. DNA polymorphisms of this type are identified by the presence of the two template sequences derived adjacent to a digestion site in one genotype, while the same two template sequences are absent in genotypes that contain polymorphisms in that specific restriction enzyme site.

In some embodiments, in order to reduce sequencing background "noise" that might arise due to the undesirable inclusion of mitochondrial or bacterial DNA in the individual's genomic DNA sample, the selected restriction enzymes will have the characteristic of cleaving organelle or bacterial DNA infrequently or not at all. This feature is potentially helpful during the initial phase of discovering the unique polymorphic templates useful for genotyping applications. However, for genotyping assays involving enrichment of informative templates by hybridization (in Stage IIIa) or PCR (in Stage IIIb of FIG. 1), the presence of organelle or non-target genome sequences is not a significant impediment and collection of this information may in some cases be informative.

IDT Exhibit 1020.1542

Stage II. Adapter ligation.

Following digestion of the DNA sample by one or more restriction enzymes, at least one adapter sequence is ligated to the fragments. A large excess of adapters is used relative to the number of RFs in the mixture, to ensure that most of the restriction fragments will receive an adapter, as illustrated schematically in FIG. 1. In some embodiments, the restriction fragments (RFs) obtained in Stage I are of the desired sequence length(s) appropriate for serving as templates in the selected genotyping by sequencing system. Accordingly, a first adapter ("adapter A") is ligated to the fragments. In most cases, a second adapter ("adapter B") is ligated to the opposite end of each fragment (PCR-based enrichment does not require this).

Adapters A and B are different, unique double stranded oligonucleotide sequences about 17 bp or longer in length. There is a practical upper limit of the sequence length of about 60–90bp, based on present costs, however even longer lengths may be used in some instances. The adapters need to be long enough to encode a sequence complementary to the primer used to initiate sequencing, the indexing sequence, and in some cases, sequences allowing binding to complementary termini generated by digestion of the target genome. The relationship of the adapters, DNA fragment, index sequences and restriction termini, in some embodiments, is illustrated in FIG. **2**. One or both adapters may contain an indexing sequence of 2–6 base pairs. The indexing sequence may be located in the adapter immediately adjacent to the nucleotide sequence (priming site) that will be used as the binding site for the sequencing primer that is used to initiate sequencing (in Stage V). The indexing sequence serves as an identification tag that allows the informative DNA templates to be sorted according to their respective sources following sequence analysis (Stage VI).

In an embodiment illustrated in boxes 2 and 3 of FIG. 2, a representative adapter A sequence in underlined (SEQ ID NOs: 1-2), and a representative adapter B sequence is italicized and underlined (SEQ ID NOs: 3-4). In box 2, adapter A is ligated to the restriction site end of the DNA fragment. The "indexing sequence" comprised in adapter A is denoted "XXX". The "restriction site overhang" comprised in adapter A is denoted "YYY." The remaining portion of adapter A corresponds to a DNA sequencing primer ("primer A"), which is described below. As shown in boxes 2 and 3, in some embodiments, adapter B is ligated to the blunt end of the DNA fragment, after shearing and end repair. A primer sequence ("primer B") (*e.g.*, ILLUMINA® primer A), ligated to the 5' end of adapter A, is a italicized sequence in boxes 3 and 4 of FIG. 2

IDT Exhibit 1020.1543

(SEQ ID NOs: 5-6). This primer B sequence may be added to the adapter-modified template through the process of PCR. The sequence of primer A (*e.g.*, ILLUMINA® primer B), is complementary to a "primer binding site" sequence within adapter A, exclusive of the indexing sequence and restriction overhang. Primers A and B are useful for bridge amplification during cluster generation in certain sequencing systems, such as the ILLUMINA® SGII sequencer. In some embodiments, the DNA sequencing primer, as shown in box 4 of FIG. 2 by dashed lines, is complementary to the primer binding sequence in adapter A. Sets of uniquely indexed adapters A and/or adapters B, and the primer sequences (*e.g.*, primer A and primer B) used to initiate sequencing are typically provided by the manufacturer of the selected sequencing platform for sequencing the informative DNA templates obtained in Stage IV. In some embodiments, the adapter sequences may also comprise specific sequences added to the ends to facilitate a primary amplification step and selection on magnetic beads, for example. It should be understood that any other suitable adapter and primer sequences capable of functioning in a similar manner may be used instead of the examples that are shown in FIG. 2.

By way of either or both adapter A and adapter B, an indexing sequence may be included at the end(s) of every DNA fragment derived from the individual's genome. Including a unique indexing sequence in each adapter-ligated set of DNA templates allows the user to pool the resulting index-modified adapter-ligated DNA templates derived from a large number of different individuals following ligation of an index-modified adapter, as discussed in more detail below in the section titled "*Multiplexing DNA Templates for Sequencing*." These samples can then be processed in bulk, sequenced, and the resulting sequences assigned to the individuals from which they were derived.

First and Second Restriction Enzyme Treatments of Genomic DNA. In some cases, the fragmenting of genomic DNA is accomplished by digesting the DNA molecules with a first restriction enzyme treatment, followed by ligation of a first adapter ("adapter A") to the ends of the resulting fragments. In some embodiments, these fragments are further digested by a second restriction enzyme treatment, to generate a pool of adapter-linked DNA templates with a predetermined range of shorter sequence lengths, as illustrated schematically in Steps 1–4 of FIG. 3). As mentioned above, in some embodiments, adapter A is a nucleotide sequence that includes an indexing sequence and an adjacent nucleotide sequence (*i.e.*, primer binding site) that

IDT Exhibit 1020.1544

can be used in a Stage IIIb PCR amplification step, shown in FIGs. 1 and 4) and/or to initiate sequencing in Stage V (Step 9, as shown in FIG. 4). Adapter B is a different nucleotide sequence, with or without the indexing sequence, that includes a primer binding sequence that can be used to amplify templates containing adapters A and B (optional PCR amplification step indicated by dashed box in FIG. 3), a step that enriches templates containing both adapter A and adapter B relative to templates containing only one type of adapter due to suppression PCR. The sequences of Adapters A and B are unique and non-complementary to each other. However, in some instances (when used in conjunction with the ILLUMINA® SGAII™, for instance), the adapter sequences may contain sequences that are complementary to oligonucleotides that are covalently attached to solid surfaces, to facilitate bridge amplification.

Restriction Enzyme and Shear Treatment of Genomic DNA. Referring to FIG. 3, in some embodiments, the genome is fragmented to a predetermined sequence length by a combination of digestion of DNA with restriction enzymes, and subsequent shearing of the restriction fragments. The shearing step may be carried out by sonication or high pressure hydrodynamic shearing. As illustrated by Steps 1–4, fragmenting of the genome DNA is accomplished by digesting the DNA molecules with a first restriction enzyme treatment, followed by ligation of a first adapter ("adapter A") to the ends of the resulting fragments, and then shearing the resulting restriction fragments to obtain a mixture or pool of informative and non-informative DNA templates. After end repair of the shear fragments using standard procedures, adapter B is ligated to the blunt ends of the shear fragments, opposite adapter A.

Digestion of genomic DNA with one restriction enzyme. In some embodiments, the genomic DNA is fragmented with a single restriction enzyme that generates template of an optimal size for a high throughput sequencer. For example, CspCI recognizes a specific 7bp sequence and digests DNA flanking that sequence at a distance such that DNA fragments of approximately 32bp are generated. In this instance, adapters A and B would be ligated to the DNA fragments at the same time, to generate DNA fragments containing adapter A or adapter B ligated to both termini of a given fragment, and DNA fragments that have adapter A ligated at one end and adapter B ligated to the other end of the fragment. PCR amplification of this mixture of adapter ligated fragments with primers complementary to sequences in adapter A and adapter B will preferentially amplify fragments containing both adapters due to suppression PCR.

IDT Exhibit 1020.1545

The optional step of limited amplification by suppression PCR is shown in the flow diagram of FIG. 3.

Shearing Genomic DNA. In some embodiments, the genomic DNA is fragmented to a predetermined sequence length by shearing instead of, or in addition to digestion with restriction enzyme(s), as shown in Step 1 of FIG. 3. In many cases, shearing treatments may be adjusted to generate the range of DNA fragment lengths of a predetermined range useful for downstream analysis on different DNA sequencers. Adapters A and B are ligated to the resulting DNA fragments (Step 2' of FIG. 3). Fragments containing Adapter A ligated to one end of the fragment and Adapter B ligated to the other end of the same fragment can be enriched relative to fragments having the same adapter ligated to both ends of the fragment, by suppression PCR amplification of the DNA templates using primers complementary to adapters A and B (as indicated in FIG. 3). In some embodiments, size selection (*i.e.*, enrichment of adapter-modified DNA fragments of specified sequence length), is performed before or after the limited amplification step. In some instances, shearing of the genomic DNA is less desirable than restriction enzyme digestion because (1) shearing produces templates with a distribution of different termini across the informative regions of the genome targeted for sequence analysis, (2) templates with high complexity are generated, (3) it is more labor intensive and less reproducible to shear and process a large number of different DNA samples compared to restriction digestion which can be done in 96-well or 384-well format, and (4) sheared DNA requires additional steps in template preparation compared to simple ligation of adapters to termini generated by restriction enzymes. Nevertheless, in some cases, a user may wish to employ shearing alone or in combination with a restriction enzyme digestion step, as outlined in FIG. 3, for ease of use or for specific applications. In some embodiments, the fragmentation protocol includes an initial shearing step followed by further reduction of fragment lengths by digestion with one or more restriction enzyme, as shown in Steps 1–4 of FIG. 3.

Biotin/Streptavidin Purification of Adapter-linked Target DNAs. In some applications, adapter A is biotinylated. In other applications, DNA templates are amplified using biotinylated primers that are complementary to sequences in adapters. The biotinylated adapters are ligated to termini generated by a restriction enzyme, allowing purification of DNA templates on streptavidin beads. Following release of the single strand template DNA (sstDNA) from the avidin beads, the sstDNA may be further processed as described below. Purification using biotinylated adapters

IDT Exhibit 1020.1546

potentially eliminates or reduces the content of DNA fragments that are not linked to adapters.

Selecting DNA Templates of Specified Sequence Length. Enrichment of templates of a given size enhances the efficiency of the procedure with respect to the selected high throughput DNA sequencing system to be used for sequencing the final templates. In some embodiments, template preparation also includes enriching the mixture of adapter-linked DNA templates for a specific range of sequence lengths. This may be done, for example, by electrophoresis of the adapter-linked DNA template on an agarose gel followed by extraction of the DNA templates of a predetermined size range, using standard techniques. Selecting only the DNA templates of a predetermine range of sequence lengths potentially increases the efficiency of the overall method by reducing or eliminating DNA templates that are larger or smaller than a predetermined sequence length (*e.g.*, longer than the sequence read length of a selected nucleotide sequencing system; *e.g.*, 30–50 bp). Size selection may be carried out after adapter ligation, or after PCR of adapter-ligated templates as shown in FIG. 3, or after enrichment of informative templates depending on the application.

Enriching for DNA Templates containing Adapter A and Adapter B. In some instances, an enrichment step may be utilized following ligation of adapters (Step II) and hybridization based template selection (Step IIIa). Referring to FIG. 3, in some applications, the adapter-linked DNA templates are amplified prior to performing Step 5, as indicated by the dashed box describing limited amplification and enrichment that involves suppression PCR. The adapter A- and adapter B-linked DNA fragments from Steps 2' or 2–4 of FIG. 3 may be subjected to PCR amplification using primers that are complementary to the non-indexing portion of adapters A and B. This step increases the relative abundance of templates containing adapters A and B relative to templates containing only adapter A or adapter B due to suppression PCR. Optional PCR amplification steps are indicated in FIG. 1 by dashed boxes labeled "PCR."

Reduction of Contaminating DNA Sequences. In some applications, undesirable secondary sources of DNA such as chloroplast, mitochondria and bacterial DNA may be present at relatively high copy number in a sample of nuclear DNA targeted for analysis. In such cases, it may be desirable to enhance the efficiency of template preparation and sequencing of the individual's DNA. In addition to taking preliminary steps to purify the target individual's nuclei to reduce organelle DNA content, utilization

IDT Exhibit 1020.1547

of tissues with relatively low organelle DNA copy number, purification of bacterial artificial chromosome (BAC) DNA in situations where BACs are being profiled, and selection of restriction enzymes that cleave the organelle or bacterial DNA infrequently or not at all (as described above), background "noise" arising from multiple copies of extraneous DNA sequences may be selectively reduced after fragmentation. The representation of those sequences during preparation of the individual's informative DNA templates may be reduced by including certain non-amplifiable primers in a PCR template preparation step (see FIG. 1) to suppress amplification of these amplicons. These non-amplifiable primers may extend into the adapter sequence and are complementary to the undesirable repeat sequences located adjacent to the restriction digestion site.

Stage III. Informative DNA Template Selection.

Prior to selecting informative DNA templates, a set of oligonucleotides is developed corresponding to the set of informative DNA templates in the target genome under investigation in a particular application. The complementary oligonucleotides used for enrichment may correspond to any unique genomic sequence in an informative DNA template. Therefore, in most instances, the adapter sequences and repetitive sequences are avoided, and, in some cases, polymorphic sequences *per se* are also avoided, thereby minimizing variable hybridization due to sequence miss-matches. The oligonucleotides are attached to the desired solid matrix (*e.g.*, magnetic beads, planar or curved surfaces, tubes). The chemistry and method of preparing the oligonucleotide-beads is carried out in accordance with the instructions of an oligonucleotide-bead manufacturer known to those in the art, or in accordance with other techniques for attaching oligonucleotides to solid substrates that are known and described in the literature. The resulting set of complementary oligonucleotides (or substrate-bound oligonucleotides) may then be assembled in any of a variety of combinations depending on the parental genotypes being assayed or the region of the genome being analyzed. Thus, in many cases, the technical burden and costs of primer design may be distributed across multiple experiments.

Stage IIIa: Hybridization-based Enrichment for Informative DNA Templates.

After the mixture of adapter-linked informative and non-informative DNA templates have been obtained (in Stage II), and, optionally, amplified and/or sized, as

IDT Exhibit 1020.1548

schematically shown in FIGs. 1 and 3, in some embodiments, the DNA templates are then enriched for informative DNA templates (*i.e.*, the templates corresponding to a site that contains a polymorphism), as illustrated in Step IIIa of FIG. 1, and Step 5 of FIG. 3. In these embodiments, enrichment of the informative DNA templates includes hybridizing only the informative DNA templates to oligonucleotides having complementary sequences. The mixed DNA templates are denatured to form single stranded DNAs, which are then hybridized to oligonucleotides that could range from about 17 to about 60 nucleotides in length. The oligonucleotides contain sequences complementary to genomic DNA sequences present in the informative DNA templates and are designed to enhance a uniform specificity of binding under a standard hybridization condition. In some embodiments of this protocol, the complementary oligonucleotides are attached to magnetic beads to facilitate the separation of hybridized informative DNA template from non-hybridized non-informative templates. In various embodiments, oligonucleotides may be attached to any other suitable solid surface or matrix. The oligonucleotide-substrate complexes must be different than the oligonucleotide-substrate complexes that are used for bridge amplification or are used in the DNA sequencing procedure. One reason for this requirement is that, if the oligonucleotide sequences used for template selection were the same as the sequences in adapters A or B (*i.e.*, sequences used for bridge amplification, etc.), then they would bind to all templates containing adapter A and/or B, which would prevent template enrichment. In certain embodiments, the hybridization-based selection procedure includes purifying the informative single stranded DNA templates by hybridization to a collection of oligonucleotides that are covalently attached to a solid matrix or surface (*e.g.*, magnetic beads).

One way in which a representative collection of oligonucleotides is obtained is described in Examples 5-14, below. The mixture of adapter-linked DNA templates (from Stage II), optionally amplified and enriched for adapter-linked DNA templates of a specified size, is mixed with the oligonucleotide-beads under hybridization promoting conditions. For some applications, the hybridization-based enrichment procedure is improved by using oligonucleotides that hybridize to their complements at similar melting temperatures (Tm), however small differences in DNA template selection efficiency will usually not affect the results significantly because the purpose of this step is to enrich a subset of the mixture or pool of informative templates, not to discriminate between polymorphic sequences. Moreover, a low level of off target

IDT Exhibit 1020.1549

template selection will only introduce a low level of off target sequencing rather than creating a source of error. In addition, it is the type of alleles at each site that is important (*e.g.,* AA, AB or BB in a simple case), and redundant sequencing of DNA templates will be carried out. For genotyping studies, in certain embodiments, a polymorphic marker is assayed within a given interval of the genetic map, however, any marker from the given region will provide similar information, and given a sufficient density of markers, missing data will not affect results. As illustrated in Stage IIIa of FIG. 1, in some embodiments the polymorphic site on the informative DNA template will be outside of the oligonucleotide:DNA template annealed region of the hybridized complex. In many embodiments this will be the preferred situation in order to avoid variation in selection of templates with perfect vs. non-perfect sequence complementarity. However, in some embodiments, the oligos used for selection will overlap a sequence containing a polymorphism. For example, this may be the case when approximately 35bp templates are generated with the restriction enzyme CspCI, as further described below.

The resulting hybridized complexes (*e.g.*, each containing adapter A, adapter B, informative DNA template, oligonucleotide, and magnetic bead or other substrate) are isolated by centrifugation, magnetic bead capture, filtering, or other suitable technique. The hybridized complexes are washed with aqueous media containing buffer (pH 7-8) and salts and at a temperature and under conditions that will not disrupt interaction between oligonucleotides and templates, to remove the non-hybridized DNA templates (Stage III of FIG. 1 and Steps 5–6 of FIG. 3). The exact hybridization and wash conditions and time (buffer, salts, temperature, etc.) will vary to some extent depending on the length and GC content of the oligonucleotides used for the selection step.

Enrichment of Informative DNA Templates Using Biotinylated Oligonucleotides and Avidin Beads. In some applications biotinylated oligonucleotides may be used to carry out selective enrichment of informative templates by hybridization (FIG. 1, Step IIIa). For some applications, the oligonucleotides complementary to informative templates are modified to contain biotin, allowing biotin/streptavidin bead-based capture to facilitate separating the hybridized informative DNA template from non-hybridized non-informative templates. More specifically, in some embodiments the oligonucleotide is modified allowing oligonucleotide:informative template hybrids to be separated from non-

IDT Exhibit 1020.1550

hybridized template. For example, if the oligonucleotide is modified by incorporation of a terminal biotin, then oligonucleotide:template hybrids may be purified by binding to a biotin:strepavidin bead or similarly modified surface. In this procedure, single-stranded adapter A- and B-modified informative DNA templates are hybridized to complementary biotinylated oligonucleotides, followed by binding of the resulting complex to streptavidin beads, washing of beads to remove non target template and other materials, and release of single strand template DNA (sstDNA) from the streptavidin beads for downstream sequencing (FIG. 1, Step V).

Stage IIIb: Enrichment of Informative DNA Templates by Targeted PCR.

An alternative to the above-described hybridization-based oligonucleotide:solid substrate enrichment procedure or biotinylated oligo-avidin bead enrichment procedure is shown schematically in FIGs. 1 and 4. In some embodiments, the enrichment of informative DNA templates is accomplished by targeted PCR amplification using a primer complementary to adapter A, and a primer complementary to a unique genomic sequence within each informative template targeted for analysis, designed so that the resulting amplified template is of the desired sequence length (compatible or optimal for the selected sequencing system) and the resulting amplified template spans informative sequences (*i.e.*, DNA polymorphisms). Embodiments of this alternative approach to selection of informative DNA templates are especially useful when the number of polymorphic sequences and informative templates targeted for analysis is less than about 2,000. The PCR-based enrichment route is shown along the lower portion of FIG. 1 (Stages I, IIb, IIIb, IVb, and V). Stage I, DNA fragment generation by digestion with restriction enzymes and/or shearing is carried out in a similar manner for both PCR-based and hybridization-based enrichment pathways. For PCR-based enrichment, step IIb involves ligation of a single adapter (A or B depending on embodiment) after which templates from different individuals may be pooled. FIG. 1 shows the instance in which adapter A is ligated to one of the termini of a DNA fragment in Stage IIb. In some instances, adapter A may be ligated to both ends of a DNA fragment. For example, when DNA fragments are generated by digestion with one enzyme or when shearing is the only method used to create DNA fragments, ligation with a single adapter will result in DNA fragments with the same adapter ligated to both termini. PCR amplification techniques are well known in the art and have been described in

IDT Exhibit 1020.1551

the literature. Briefly, PCR amplification includes selectively amplifying the single-stranded adapter A-modified informative templates using primers that hybridize to adapter A and a set of second informative template specific primers, each of which is complementary to a unique genomic DNA sequence flanking adapter A in an adapter A-modified informative DNA-template. The second primers that are unique to each informative template targeted for analysis are configured so that the amplified DNA contains a DNA sequence polymorphism. In this way, only the informative DNA templates are amplified, and the non-informative templates are not resulting in an enrichment of informative DNA templates. In some instances, primers that bind to adapter A and 10 or more different primers complementary to different informative templates are pooled (multiplexed) to streamline template preparation. In this way about 1,000–2,000 informative templates can be selectively amplified using a 96-well plate format for PCR, using pools of 10 or 20 different primers specific to informative templates per well.

In some embodiments, the PCR amplification primers used to target informative templates in Stage IIIb are designed to include two primary features: (1) unique "targeting" sequences are present at the 3'-end of each primer, usually 17–30bp in length, that are complementary to respective informative templates targeted for analysis, and (2) a "universal" amplification sequence, usually 17–30bp in length, that is not present in the target genome (as shown in FIG. 4). After several rounds of PCR amplification of informative templates with primers having this design, all templates may then be amplified with primers complementary to adapter A and a primer complementary to the "universal" sequence present in the mixture of informative templates. As a result of this primer design, the amplified templates are of a predetermined sequence length and contain informative sequences. In some embodiments, the "universal" sequence (*e.g.*, primer B in FIG. 4) is selected to be compatible with template preparation for high throughput sequencing on the ILLUMINA® SGAII™, ABI SOLID™, or the Roche 454. In some instances, one of the primers used to amplify the informative templates contains a biotin that allows purification of template prior to sequencing (Stage IVb).

Size Selection. As illustrated in FIGs. 1 and 3, and described above following Stage II, size selection may be performed either before or after the first PCR step, or, in some cases, later in the procedure. Size selection is desirable in many applications for increasing overall efficiency of a template generation and sequencing process. It is

IDT Exhibit 1020.1552

optional in some instances, however, such as when Stage IIIb PCR based enrichment is employed. In Stage IIIa, the DNA fragments have adapters A and B ligated to the fragments. In Stage IIIb (PCR-based enrichment route), however, only one adapter is ligated in Stage II (either A or B, depending on the application). In the PCR-enrichment route, the second priming site is incorporated during targeted amplification, as shown. The primers are indicated by arrows in Stage IIIb, and the adapter B linked to the second primer is designated by an open box. The PCR enrichment route (Stage IIIb) is also illustrated in FIG. 4. The asterisk (*) denotes the site of a DNA polymorphism in the genome (Stage I) and DNA templates (Stages II-IV) derived from the genome.

Stage IV. Recovery of Informative DNA Templates.

Following hybridization and separation from non-informative templates, the adapter-linked informative single stranded DNA templates are separated from the bead-bound complementary oligonucleotides by heat denaturation or treatment with alkali to release the selected templates (Stage IV of FIG. 1 and Step 7 of FIG. 3). The enriched set of DNA templates is then ready for an optional further amplification, and/or direct sequencing and analysis. The resulting single-stranded DNA templates are then sequenced (Step 9 of FIG. 4).

Enrichment for Informative DNA Templates. In some applications, the informative DNA templates recovered in Stage IVa (FIG. 1 and Step 7 of FIG. 3) are further enriched by PCR or any other suitable amplification technique. A PCR amplification procedure includes denaturing the hybridized DNA templates (as in Step 7) and PCR amplification of the resulting mixture of informative DNA templates using a primer that is complementary to adapter A and a second primer that is complementary to adapter B.

One suitable PCR amplification technique is the well known solid-phase bridge amplification used to create DNA clusters for DNA sequencing on the ILLUMINA® SGAII™ in which adapter-linked single stranded DNA templates attach to a surface containing a multiplicity of regularly spaced single-stranded primers having sequences that are complementary to the primer sequences contained in adapters A and B. The polymerase enzyme incorporates nucleotides to build double-stranded "bridges" between the spaced-apart primers on the solid surface. After

IDT Exhibit 1020.1553

amplification, the resulting double-stranded DNA sequences may be represented as shown in box 3 of FIG. 2.

Multiplexing DNA Templates for Sequencing. Due to the inclusion of a unique indexing sequence ("index") in one or both of adapter A and adapter B, the mixed DNA templates from Stage II, or the informative DNA templates from Stage IV of one individual may be pooled with similarly prepared, but differently indexed, adapter-linked DNA templates derived from other individuals. The pool of differently indexed adapter-linked informative DNA templates may then be further processed and sequenced together. An indexing sequence (XXX) is shown in box 2 of FIG. 2, flanking a polymorphism-containing DNA fragment having restriction termini denoted by "YYYyy," at the 5' end of the fragment.

Stage V. High Throughput Sequencing of Informative DNA Templates.

The recovered, and optionally amplified, adapter-linked informative DNA templates may then be directly sequenced using a high throughput DNA sequencing platform in which the optimal template sequence length for sequencing is in the range of about 50 bp to about 600 bp, in accordance with the instructions provided by the manufacturer of the chosen high throughput sequencing platform. This high throughput sequencing step may also be referred to as "resequencing" of the informative DNA templates in instances and in situations in which a previously sequenced genome is the target genome under investigation. In some instances, a reference sequence will have been previously obtained for the same or a similar individual, and the polymorphisms in the reference sequence will have been previously identified. For example, a reference sequence may have been previously obtained as part of the process of preparing an above-mentioned set of complementary nucleotide sequences for use in the Stage III hybridization-based template selection process.

In some embodiments, the informative DNA templates are sequenced using an ILLUMINA® SGAII™ sequencing system, in which the optimal template sequence length is about 50-250bp. Optimal template size in this instance is related to a template size that will uniformly be amplified in the sequencer by bridge amplification. In some embodiments, the informative DNA templates are sequenced using an ABI DNA SOLID™ System, in which the optimal template sequence length is about 50 bp. In some embodiments, the informative DNA templates are sequenced using a Roche 454

IDT Exhibit 1020.1554

GS-FLX™ Genome Analysis System, in which the optimal template sequence length is at least 400 bp in cases where the read length of the sequencer is 400 bp.

In these and similar sequencing systems, a number of "sequencing by synthesis" reactions are used to elucidate the identity of a plurality of bases at target positions within the target sequence. Sequencing by synthesis techniques are well known in the art, and have been described in the literature. All of these reactions rely on the use of a target sequence comprising at least two domains, including a first domain (*i.e.*, an adapter) to which a sequencing primer will hybridize, and an adjacent second domain, for which sequence information is desired (*i.e.*, an informative DNA template of initially undetermined sequence). Upon formation of the assay complex, enzymes are used to add dNTPs to the primer, and each addition of dNTP is "read" to determine the identity of the added dNTP. This may proceed for many cycles.

Sequencing primers specific to the adapters ligated to termini created by the "targeting" restriction enzyme are used to generate sequences adjacent to these sites. RSL-sequencing is a flexible technology that will allow the investigator to vary the number of sites sequenced and depth of sequencing depending on application. Another feature of sequence-based genotyping from RSL-templates is that sequencing always starts from the same sites in the genome, immediately adjacent to the targeting restriction enzyme recognition sequence. As a result, the location of a specific sequence polymorphism is always a specified number of bases away from the start site for DNA sequencing in each template selected for analysis. For many applications, this is a potentially valuable feature because the accuracy of sequencing decreases in a predictable way as a function of position in the sequence. Conversely, sequence accuracy is highest close to the sequencing primer. Therefore, in some cases a quality score can be assigned to sequence-based genotypes based in part on this information.

Varying the number of sites sequenced per genome. Depending on the application, a user may want to sequence a large number of different DNA sequences or a small number of sequences in a target genome. Accordingly, the depth and redundancy of RSL-sequencing may be modulated in any of several ways, which are briefly described as follows: (1) by selecting restriction enzymes that cleave the target genome with difference frequencies depending on the recognition site, (2) by digesting the target genome with two or more restriction enzymes that recognize different 8-base, 6- or 4-base sequences, for example, to increase the number of different DNA segments targeted for sequencing; (3) by amplifying and pooling DNA from two or more

IDT Exhibit 1020.1555

different genotypes, each containing a specific sequence identification tag (indexing), to track the origin of the DNA sequences; and (4) by using restriction enzymes that are sensitive to the methylation state of DNA within their recognition sequences; all of these variations are further described elsewhere herein.

Stage VI. Identification of Polymorphisms* and *Genotyping.

The flexible methods disclosed herein allow targeted analysis of subsets of RSL-templates generated by any specific set of restriction enzymes, by permitting variation in the number of informative DNA templates analyzed per genome and their distribution across the genome, or based on their utility in larger population studies. A global analysis of the sequences of all RSL-templates generated by a particular restriction enzyme or combination of restriction enzymes from the parental genotypes of interest results in data that, in many embodiments, will identify; (1) the subset of informative RSL-templates that can be successfully sequenced at a reasonable frequency. Differences in template utility are affected by size (*i.e.*, sequence length), repetitiveness, presence of polymorphisms, and other factors; (2) the subset of RSL-templates that contain unique sequences in cases where a reference genome sequence is available, and (3) the subset of unique sequences that contain polymorphisms that distinguish parental genotypes. In applications in which the user is working with a sequenced genome that is cross-referenced to a genetic map, the unique sequences may be mapped using bioinformatics. In applications in which the target genome has not been previously sequenced, then the polymorphic sequences will need to be mapped through normal segregation analysis in a population. Polymorphic RSL-sequences may be genetically mapped by analyzing mapping progeny followed by linkage analysis.

Once the sequences of the recovered informative DNA templates have been ascertained, they may then be compared to those derived from one or more different individuals or to a reference set of sequences, to identify specific DNA polymorphisms (alleles) and to generate genotyping/haplotyping information. Suitable software for analyzing the sequence data, and for aligning the sequences (BLAST, BLAT, etc.) to a reference in resequencing applications is available from well-known commercial sources.

For instance, an approximately 1,000 Mbp "genome" of random sequence is predicted to contain about 15,500 digestion sites for a restriction enzyme that recognizes a specific 8-base sequence. The collection of about 100 bp sequences

IDT Exhibit 1020.1556

flanking this set of restriction enzyme digestion sites constitutes a specific approximately 3.1 Mbp sub-sample of the target "genome." RSL-sequencing, as described above, allows the sequences flanking this set of restriction sites to be obtained from genomic DNA or a library of large insert bacterial artificial chromosome (BAC) clones prepared from that genome. The RSL-sequences derived from large insert clones may potentially be used to build physical maps, as the overlapping clones will contain common RSL-sequences. RSL-sequences mapped onto a BAC-based physical map spanning a genome may potentially be used to locate gene sequences and whole genome sequence assemblies on the genome map. In addition, RSL-sequences from different genotypes may be compared to identify DNA polymorphisms useful for the design of DNA marker assays and for diversity and haplotype analyses. Therefore, RSL-sequences obtained using any suitable high throughput-sequencing platform will potentially integrate genome map building, genome sequencing, and diversity analysis.

Definitions

The term "informative DNA template" refers to a DNA template that has three properties; (1) the template is compatible with high throughput sequencing, (2) the template contains one or more sequences that can be mapped to a unique location in a genome, and (3) the sequence or sequences derived from the template are polymorphic in the target species (*i.e.*, parental lines used for genetic mapping, lines being fingerprinted, individuals being analyzed as part of a diversity or haplotyping study). Thus, an informative DNA template contains a sequence in a location suitable for high throughput DNA sequencing, and contains a unique polymorphic site in the species genome sequence. Although an informative DNA template must span a sequence that is polymorphic in the target species, it should be understood that it may or may not be polymorphic in the particular individual analyzed.

DNA templates generated using at least one restriction enzyme are referred to herein as "restriction site localized templates" or "RSL-templates". For ease of reference, the DNA templates generated by shearing only, by digestion with restriction enzyme(s) only, or by a combination of shearing and restriction enzyme(s) digestion, are sometimes referred to herein as "RSL-templates." Accordingly, the term "RSL-templates" should be interpreted to include DNA templates generated by either restriction enzyme(s) digestion or shearing, or by a combination of those.

IDT Exhibit 1020.1557

A "nuclear genome" is all the DNA or genetic material in the chromosomes of a eukaryotic organism. Eukaryotic organisms such as plants will also contain organellar genomes in their mitochondria and chloroplasts.

An "allele" is a distinct DNA sequence or "spelling" of a chromosomal region. Customarily, the first identified allelic form is arbitrarily designated as the reference form and other allelic forms are designated as alternative or variant alleles. The allelic form occurring most frequently in a selected population is sometimes referred to as the wild type form. Diploid organisms may be homozygous or heterozygous for allelic forms. A diallelic polymorphism has two forms. A triallelic polymorphism has three forms. Most organisms have multiple alleles of gene sequences in their naturally occurring populations (germplasm in the case of plant species).

"Polymorphism" refers to the occurrence of two or more alternative sequences or alleles in a species or population. A "polymorphic site" is the locus or specific sequence location in a genome at which sequence divergence occurs (*i.e.*, the site of variation between allelic sequences). A polymorphism may comprise one or more base changes, a nucleotide insertion or deletion (INDEL), a nucleotide inversion, or variation in the size of a simple sequence repeat (SSR), relative to a reference allele. A single base pair polymorphism, termed a "single nucleotide polymorphism" (SNP), occurs at a polymorphic site occupied by a single nucleotide. A single nucleotide polymorphism usually arises due to substitution of one nucleotide for another at the polymorphic site. A transition is the replacement of one purine by another purine or one pyrimidine by another pyrimidine. A transversion is the replacement of a purine by a pyrimidine or vice versa.

A "genotype" is a collection of all the polymorphisms or alleles of an individual's genome.

A "haplotype" is a combination of alleles or polymorphisms at multiple loci that tend to be transmitted together. For example, neighboring polymorphisms that are inherited together on the same chromosome.

A "monomorphic template" does not contain DNA polymorphisms.

An "amplicon" refers to a fragment of DNA that can be amplified using specific priming sites located at each terminus, sequences usually added by ligation of adapters.

An "oligonucleotide" is a relatively short nucleic acid sequence, such as DNA or RNA, and may be single- or double-stranded. Oligonucleotides are typically prepared by synthetic means, however they may also be isolated from naturally

IDT Exhibit 1020.1558

occurring sources. For the purposes of this disclosure, oligonucleotides are usually in the range of about 17 to about 30 base pairs (bp) in length, and in some instances are about 30-90 bp long, for example.

The term "hybridization" refers to the non-covalent interaction or binding of two complementary single-stranded nucleic acid strands (*i.e.*, DNA and/or RNA) into a single double-stranded molecule. Two perfectly complementary strands will bind to each other readily (*i.e.*, anneal or "hybridize") because the nucleotides of the complementary strands bind to their complements under normal hybridization conditions. Hybridizations are usually performed under stringent conditions that are dependent on sequence length, GC content, temperature, salt, and other characteristics of the hybridizing media. For example, conditions of 5X SSPE (750 mM NaCl, 50 mM Na_2PO_4, 5 mM EDTA, pH 7.4) and a temperature of 25–30°C are suitable for SNP-specific oligonucleotide probe hybridizations. For stringent hybridization conditions, see, for example, Sambrook, Fritsche and Maniatis. "Molecular Cloning A laboratory Manual" 2nd Ed. Cold Spring Harbor Press (1989).

A "hybridization array" is an array comprising a solid support or matrix with attached oligonucleotide probes. Arrays typically comprise a plurality of different oligonucleotide probes that are coupled to a surface of a substrate in different locations. Substrates may be beads, planar or curved surfaces, fibers such as fiber optics, glass or any other suitable material or structure.

"High throughput sequencing" refers to instances where 400,000 to 70 million templates or more are sequenced in parallel in a single run generating up to 2 billion bases of sequence or more per run, in an automated nucleotide sequencing system.

References herein to "an individual" may apply to human beings, other mammals, plants, bacteria, or any other organism, as the context allows in this disclosure. As used herein, the term "individual" refers to a single member of any species. In most cases, the genotypes and templates derived from each individual of a species or of a different species will be different, but occasionally (*e.g.*, twins or clones) the genotypes and corresponding templates will be the same. For example, different individuals may have different genotypes (*e.g.*, Genbank accession numbers, ecotypes, germplasm accession numbers, etc.), with the exception of clones or genetically identical twins.

IDT Exhibit 1020.1559

EXAMPLES

Example 1. DNA template preparation using one restriction enzyme.

Digest genomic DNA of a first individual with the CspCI, a type II restriction enzyme, to generate DNA restriction fragments (RFs) approximately 34-36 bp in size (*i.e.*, sequence length), which is optimal for sequencing on the ILLUMINA® SGAII™ sequencer. CspCl recognizes and cuts the sequence:

5'….$_{10\text{-}11}$(N)CAA(N)$_5$GTGG(N)$_{12\text{-}13}$….3' (SEQ ID NO: 39)

3'….$_{12\text{-}13}$(N)GTT(N)$_5$CACC(N)$_{10\text{-}11}$….5' (SEQ ID NO: 40)

This enzyme typically digests a random genome about every 8,192 bp and thereby generates a large number of approximately 35 bp fragments

Ligate adapters A and B to the resulting RFs, using standard ligation techniques. Adapter A is a double stranded nucleotide sequence containing short terminal sequences that allow binding and ligation to DNA fragments generated by one or more restriction enzymes, selected 2-6 bp nucleotide sequence (index), a unique primer binding sequence useful for amplification or to initiate DNA sequencing ("primer A"), and adapter B contains a unique primer binding sequence. The sets of uniquely indexed adapters A and adapters B can be prepared by individual users of the technology but are typically provided by the manufacturer of the selected sequencing platform to be used for sequencing the informative DNA templates obtained in Step 1–5. If adapter A contains an indexing sequence, similarly prepared RFs from different individuals can be pooled to increase the efficiency of downstream processing

Amplify the adapter-linked RFs by PCR using primers specific to sequences in adapter A and adapter B to enrich for DNA templates containing both adapters

Enrich for DNA templates with sequence lengths optimal for the DNA sequencer selected for analysis (*i.e.*, in the range of about 50-200bp for the ILLUMINA® SGAII™, 400-600 for the Roche 454 sequencer).

Enrich for adapter-linked informative DNA templates (*i.e.*, those containing sequences corresponding to a polymorphism).

Sequence the resulting informative DNA templates with an ILLUMINA® SGAII™, ABI SOLID™, or Roche 454 sequencer.

Compare the sequences obtained to those of a set of reference sequences, to identify DNA sequence polymorphisms and genotypes.

IDT Exhibit 1020.1560

Example 2. DNA template preparation using two restriction enzymes.

Digest genomic DNA with the restriction enzyme, FseI, to generate RFs ranging from less than 100 bp to more than 100,000 bp in length with an average fragment size of 64,000 bp for a genome of random sequence.

Ligate adapter A to the resulting RFs. If adapter A contains an indexing sequence, similarly prepared RFs from different individuals can be pooled to increase the efficiency of downstream processing.

Digest the adapter A-linked RFs with a second restriction enzyme that has a 4-base recognition sequence to reduce their average sequence lengths to about 256 bp.

Ligate adapter B to the resulting Adapter A-linked DNA templates, opposite adapter A.

Enrich the adapter-linked DNA templates in the range of about 50-400bp for sequencing on the ILLUMINA® SGAII™ and to about 400-600bp for sequencing on the Roche 454 sequencer.

PCR amplify DNA templates that contain both adapter A and adapter B.

Enrich for informative DNA templates.

Sequence the informative DNA templates.

Compare the sequences obtained from each individual to each other and/or to a reference set of sequences to identify DNA polymorphisms and genotypes.

Example 3. DNA template preparation using restriction enzymes and shearing.

Digest genomic DNA with the restriction enzyme, FseI, to generate RFs ranging from less than 100bp to more than 100,000 bp in length with an average length of 64,000 bp for a genome of random sequence.

Ligate adapter A to the resulting RFs. If adapter A contains an indexing sequence, similarly prepared RFs from different individuals can be pooled to increase the efficiency of downstream processing.

Fragment the adapter A-linked RFs by shearing to generate smaller fragments in the range of about 50bp to about 200bp sequence length for sequencing on the ILLUMINA® SGAII™ or 400-600 for sequencing on the Roche 454 sequencer.

Ligate adapter B to the DNA fragments.

PCR amplify templates containing both adapter A and adapter B using primers specific to sequences contained in adapter A and adapter B.

Enrich DNA templates of a predetermined size.

IDT Exhibit 1020.1561