COMPOSITIONS AND METHODS FOR TARGETED NUCLEIC ACID SEQUENCE ENRICHMENT AND HIGH EFFICIENCY LIBRARY GENERATION

CROSS-REFERENCE

[0001] This application claims the benefit of U.S. Provisional Application No. 61/591,241, filed January 26, 2012, which application is incorporated herein by reference.

BACKGROUND OF THE INVENTION

[0002] With the rapid development of next generation sequencing (NGS) technologies and platforms, whole genome sequencing is becoming increasingly feasible. Researchers are driven to generate increasing amounts of data to achieve greater understanding of variance and biological trends, and to generate data from smaller sample sizes to avoid averaging across multiple cells within a tissue.

[0003] Although the cost of whole genome sequencing is decreasing and the throughput of the NGS platforms is increasing, it is nonetheless often more practical and cost-effective to select genomic regions of interest for sequencing and analysis. Target enrichment is a commonly employed strategy in genomic DNA sequencing in which genomic regions of interest are selectively captured from a DNA sample before sequencing. Focused target enrichment is an important tool especially in the fields of study where sequencing of a large number of samples is necessary (e.g. population-based studies of disease markers or SNPs), making whole genome sequencing cost-prohibitive. Similarly, improvements have been made that enable DNA libraries to be made from nucleic acid from fewer number of cells, but these are bound by the limitations of the efficiency of ligation reactions.

[0004] Several approaches to target enrichment have been developed which vary from one another in terms of sensitivity, specificity, reproducibility, uniformity, cost and ease of use. The target enrichment methods commonly employed today can be divided into three major categories, each with its distinct advantages and disadvantages: 1) PCR-based methods; 2) capture-by-hybridization, i.e. on-array or in-solution hybrid capture; and 3) capture-by-circularization, i.e. molecular inversion probe-based methods.

[0005] The PCR-based methods employ highly parallel PCR amplification, where each target sequence in the sample has a corresponding pair of unique, sequence-specific primers. The simultaneous use of numerous primer pairs makes multiplex PCR

impractical due to high level of non-specific amplification and primer-primer interactions. Recently developed microdroplet PCR technology (Tewhey *et al.*, 2009) in which each amplification reaction is physically separated into an individual droplet removes the constraints of multiplex PCR relating to non-specific amplification and primer-primer interactions. However, microdroplet PCR and other improved PCR-based methods require special instrumentation or platforms, are limited in their throughput, and, as with conventional multiplex PCR, require a large number of individual primer pairs when enriching for a multitude of regions on interest, thus making target enrichment costly.

[0006] Hybrid capture methods are based on the selective hybridization of the target genomic regions to user-designed oligonucleotides. The hybridization can be to oligonucleotides immobilized on high or low density microarrays (on-array capture), or solution-phase hybridization to oligonucleotides modified with a ligand (e.g. biotin) which can subsequently be immobilized to a solid surface, such as a bead (in-solution capture). The hybrid capture methods require complex pools of costly long oligonucleotides and long periods (typically 48 hours) of hybridization for efficient capture. For on-array hybrid capture, expensive instrumentation and hardware is also required. Because of the relatively low efficiency of the hybridization reaction, large quantities of input DNA are needed.

[0007] The molecular inversion probe (MIP) based method relies on construction of numerous single-stranded linear oligonucleotide probes, consisting of a common linker flanked by target-specific sequences. Upon annealing to a target sequence, the probe gap region is filled via polymerization and ligation, resulting in a circularized probe. The circularized probes are then released and amplified using primers directed at the common linker region. One of the main disadvantages of the MIP-based target enrichment is its relatively low capture uniformity, meaning there is large variability in sequence coverage across the target regions. As with PCR and hybrid capture, the MIP-based method requires a large number of target-specific oligonucleotides, which can be costly.

[0008] There is a need for improved methods for selective target enrichment that allow for low-cost, high throughput capture of genomic regions of interest without specialized instrumentation. Additionally, there is also a need for high efficiency nucleic acid library generation. The methods of the invention described herein fulfills these needs.

SUMMARY OF THE INVENTION

[0009] In one aspect, disclosed herein are methods for enriching for target nucleic acid sequences of interest in a sample comprising nucleic acids, the method comprising: (a) fragmenting the nucleic acids, thereby generating nucleic acid fragments; (b) appending a first adaptor to a 5' end of each nucleic acid fragment; (c) annealing one or more oligonucleotides to the nucleic acid fragments, whereby each of the one or more oligonucleotides comprise a 3' portion that is complementary to a target nucleic acid sequence of interest present in one or more of the nucleic acid fragments, and a 5' portion comprising a second adapter sequence; (d) extending the one or more oligonucleotides with a polymerase thereby generating one or more oligonucleotide extension products with the first adaptor at a first end and the second adaptor sequence at a second end; and (e) amplifying the one or more oligonucleotide extension product using a first primer that is complementary to the first adaptor and a second primer that is complementary to the second adaptor sequence to enrich for nucleic acid fragments containing the first adaptor and the second adaptor sequence at each end. In one embodiment, the method further comprises an additional step of sequencing the one or more oligonucleotide extension product following amplification. In one embodiment, the target nucleic acid sequences of interest comprise genomic DNA, RNA, or cDNA. In one embodiment, the target nucleic acid sequences of interest comprise genomic DNA. In one embodiment, the target nucleic acid sequences of interest comprise cDNA. In one embodiment, the method further comprises denaturing the nucleic acid fragments prior to step c, thereby generating single-stranded nucleic acid fragments with the first adaptor sequence at the 5'end. In one embodiment, the first adaptor can be common to each nucleic acid fragment. In one embodiment, the second adaptor sequence can be common to the one or more oligonucleotides. In one embodiment, the first adaptor and the second adaptor sequence can be distinct from each other. In one embodiment, the first adaptor and/or the second adaptor sequence further comprise barcode sequence. In one embodiment, step b can be performed by ligation. In one embodiment, the method further comprises an additional step of performing gap repair following ligation of the first adapter to create nucleic acid fragments with complementary termini. In one embodiment, a composition comprising enriched target nucleic acid sequences of interest can be generated by the methods disclosed herein. In one embodiment, the polymerase can be a DNA polymerase.

[0010] In another aspect, disclosed herein are methods for enriching for target nucleic acid sequences of interest in a sample comprising nucleic acids, the method comprising: (a) fragmenting the nucleic acids, thereby generating nucleic acid fragments; (b) appending a first adaptor to the nucleic acid fragments wherein the first adaptor comprises a partial duplex with a short strand and a long strand wherein a 3' end of the short strand of the partial duplex adaptor comprises a blocking group, and a 5' end of the long strand of the partial duplex adaptor comprises a restriction and/or cleavage site for a nucleic acid modifying enzyme; (c) denaturing the nucleic acid fragments, thereby creating single-stranded nucleic acid fragments; (d) annealing one or more oligonucleotides to the single stranded nucleic acid fragments, whereby each of the one or more oligonucleotides comprise sequence that is complementary to a target nucleic acid sequence of interest present in one or more of the single-stranded nucleic acid fragments, extending the one or more oligonucleotides with a polymerase to produce one or more double stranded nucleic acid complexes comprising the target nucleic acid sequences and their complements, a first end with a double stranded restriction and/or cleavage site for the nucleic acid modifying enzyme , and a second end with a 3' overhang comprising the short strand of the first adaptor; (e) cleaving the double stranded restriction and/or cleavage site with the nucleic acid modifying enzyme, thereby generating a cleavage site; (f) ligating a second adaptor to the cleavage site, wherein the second adaptor comprises a duplex with two strands; (g) denaturing the one or more double-stranded nucleic acid complexes, thereby generating one or more single stranded nucleic acid fragments comprising the target nucleic acid sequences, a strand from the second adapter at the first end, and the short strand of the first adaptor at the second end; and (h) amplifying the one or more single stranded nucleic acid fragments comprising the one or more target nucleic acid sequences with a first primer comprising sequence complementary to the strand from the second adapter and a second primer comprising sequence complementary to the short strand of the first adaptor, thereby enriching for the one or more target nucleic acid sequences. In one embodiment, the method further comprises an additional step of sequencing the one or more single stranded nucleic acid fragments from step h following amplification. In one embodiment, the target nucleic acid sequences of interest comprise genomic DNA, RNA, or cDNA. In one embodiment, the target nucleic acid sequences of interest comprise genomic DNA. In one embodiment, the target nucleic acid sequences of interest comprise cDNA. In one embodiment, the first adaptor and the second adaptor can

be common to each of the nucleic acid fragments. In one embodiment, the first adaptor and the second adaptor can be distinct from each other. In one embodiment, the first adaptor and/or the second adaptor further comprise barcode sequence. In one embodiment, the double stranded restriction and/or cleavage site for the nucleic acid modifying enzyme from step e comprises the 5' end of the long strand of the partial duplex of the first adaptor and sequence complementary to the 5' end of the long strand of the partial duplex of the first adaptor generated from extension of the one or more oligonucleotides. In one embodiment, the nucleic acid modifying enzyme comprises a restriction enzyme. In one embodiment, step b can be performed by ligation. In one embodiment, a composition comprising enriched target nucleic sequences of interest can be generated by the methods disclosed herein. In one embodiment, the polymerase can be a DNA polymerase.

[0011] In yet another aspect, disclosed herein are methods for generating a library of nucleic acid sequences, the method comprising: (a) fragmenting a sample comprising nucleic acids, thereby generating nucleic acid fragments; (b) appending a first adapter to each of the nucleic acid fragments; (c) denaturing the nucleic acid fragments, thereby generating a library of single-stranded nucleic acid fragments; (d) annealing one or more oligonucleotides to the single-stranded nucleic acid fragments wherein each of the one or more oligonucleotides comprises a 3' portion complementary to sequence in one or more of the single-stranded nucleic acid fragments and a 5' portion comprising a second adaptor sequence; (e) extending the one or more oligonucleotides with a polymerase thereby generating one or more oligonucleotide extension products comprising the first adaptor at a first end and the second adapter sequence at a second end; and (f) amplifying the one or more oligonucleotide extension products with a set of primers specific to the first adaptor and the second adaptor sequence to generate a library of nucleic acid fragments comprising the first adaptor and second adaptor sequence at each end. In one embodiment, the method further comprises an additional step of performing a gap repair reaction following ligation of the forward adapter to create nucleic acid fragments with complementary termini. In one embodiment, the method further comprises an additional step of sequencing the amplified one or more oligonucleotide extension product from step f. In one embodiment, the nucleic acid sequence comprises genomic DNA. In one embodiment, the nucleic acid sequence comprises cDNA. In one embodiment, step c can be omitted wherein the nucleic acid fragments are double-stranded. In one embodiment,

the 3' portion of the one or more oligonucleotides of step d comprises random sequence. In one embodiment, step b can be performed by ligation. In one embodiment, the first adaptor and the second adaptor sequence can be common to each nucleic acid fragment. In one embodiment, the first adaptor and the second adaptor sequence can be distinct from each other. In one embodiment, the first adaptor and/or the second adaptor sequence further comprise barcode sequence. In one embodiment, the methods disclosed herein can generate a composition comprising a library of nucleic sequences. In one embodiment, the polymerase can be a DNA polymerase.

[0012] In further aspect, disclosed herein are methods for enriching for target nucleic acid sequences of interest from a library comprising nucleic acid inserts with a first adaptor on a first end and a second adaptor on a second end, the method comprising: (a) denaturing the nucleic acid inserts, thereby generating a library of single stranded nucleic acid inserts; (b) annealing one or more oligonucleotides to the single stranded nucleic acid inserts, wherein each of the one or more oligonucleotides comprises a 3' portion that is complementary to a target nucleic acid sequence of interest present in one or more of the nucleic acid inserts, and a 5' portion comprising a third adaptor sequence; (c) extending the one or more oligonucleotides with a polymerase thereby generating one or more oligonucleotide extension products with the first adaptor at the first end and the third adaptor sequence at the second end; and (d) amplifying the one or more oligonucleotide extension products using a first primer that is complementary to the first adaptor and a second primer that is complementary to the third adaptor sequence to enrich for nucleic acid fragments containing the first adaptor and the third adaptor sequence at each end. In one embodiment, the method further comprises an additional step of sequencing the amplified one or more oligonucleotide extension products from step d. In one embodiment, the target nucleic acid sequences of interest comprise genomic DNA. In one embodiment, the target nucleic acid sequences of interest comprise cDNA. In one embodiment, step a can be omitted wherein the nucleic acid fragments can be double-stranded. In one embodiment, the first adaptor and the second adaptor can be common to each nucleic acid fragment. In one embodiment, the third adaptor sequence can be common to the one or more oligonucleotides. In one embodiment, the first adaptor and the second adaptor can be distinct from each other. In one embodiment, the first adaptor and the second adaptor can be the same. In one embodiment, the first adaptor, the second adaptor and/or the third adaptor sequence further comprise barcode sequence. In one

embodiment, the methods disclosed herein can generate a composition comprising enriched target nucleic acid sequences of interest.

INCORPORATION BY REFERENCE

[0013] All publications, patents, and patent applications mentioned in this specification are herein incorporated by reference to the same extent as if each individual publication, patent, or patent application was specifically and individually indicated to be incorporated by reference.

BRIEF DESCRIPTION OF THE DRAWINGS

[0014] The novel features of the invention are set forth with particularity in the appended claims. A better understanding of the features and advantages of the present invention will be obtained by reference to the following description that sets forth illustrative embodiments, in which the principles of the invention are utilized, and the accompanying drawings of which:

[0015] FIG. 1 depicts selective target enrichment using ligation of a single forward adaptor at the ends of the DNA fragments in the DNA library. The sequence-specific oligonucleotide that anneals to the target region of interest contains a common reverse adaptor sequence at its 5'end, and following sequence-specific oligonucleotide extension, PCR is performed using a set of primers specific to the forward and reverse adaptors.

[0016] FIG. 2 depicts an alternative ligation protocol wherein the DNA-fragment-adaptor complex is denatured following ligation without nick repair and adaptor fill-in, generating ligation products where non-complementary ends exist on each insert.

[0017] FIG. 3 depicts selective target enrichment using ligation of partial duplex adaptors. Cleavage of the 5' end of the long strand of the partial duplex adaptor (and the corresponding complementary sequence of the extended sequence-specific oligonucleotide) by a nucleic acid modifying enzyme specific for double-stranded DNA allows for ligation of a new adaptor pair , and consequently, amplification with primers corresponding to the new adaptors.

[0018] FIG. 4 depicts high efficiency NGS library generation using random priming. The oligonucleotide that anneals to the DNA fragment contains a common reverse adaptor sequence at its 5' end, and following primer extension, PCR is performed using a set of primers specific to the forward and reverse adaptors.

DETAILED DESCRIPTION OF THE INVENTION

General

[0019] The methods of the invention can be used for the selective enrichment of a plurality of defined target sequences from complex DNA with a set of common primers and adaptors, thus circumventing the need for multiplex PCR and multiple primer pairs. A multiplicity of target regions of interest are envisioned: for example, the regions of interest can represent all known coding regions, the entire exome, selected regions of coding genomic regions representing selected pathways, selected genomic regions known to comprise genomic variation related to altered phenotype, entire or selected regions of a specific chromosome, and the like. In another aspect, the methods of the invention can be used for high efficiency nucleic acid library production as well.

[0020] Altogether, the methods of the present invention create a simple, low cost, high throughput system for target enrichment and library preparation.

[0021] Reference will now be made in detail to exemplary embodiments of the invention. While the disclosed methods and compositions will be described in conjunction with the exemplary embodiments, it will be understood that these exemplary embodiments are not intended to limit the invention. On the contrary, the invention is intended to encompass alternatives, modifications and equivalents, which may be included in the spirit and scope of the invention.

[0022] In one embodiment, the present invention provides methods and compositions for the enrichment of specific target sequences of interest from a sample comprising nucleic acids. The methods described herein enrich target sequences using conventional duplex adaptors and/or partial duplex adaptors, sequence specific oligonucleotides, restriction enzymes and ligation. The methods further enable enrichment of target sequences from specific strands of template nucleic acids which can be further amplified using a variety of amplification methods. In another embodiment, the present invention provides methods for high efficiency generation of libraries comprising specific nucleic acid sequences of interest.

[0023] In one embodiment, the present invention provides methods and compositions for the enrichment of target nucleic acid sequences from a sample comprising nucleic acids. In one aspect, the method comprises fragmenting nucleic acids in an input sample to generate nucleic acid fragments. The nucleic acids can be DNA, or RNA. The nucleic acids can be single or double stranded. The DNA can be genomic DNA or cDNA or any

combination thereof. In one embodiment, the nucleic acids in an input sample are double stranded DNA. In one embodiment, fragmentation of the nucleic acids can be achieved through methods known in the art. Fragmentation can be through physical fragmentation methods and/or enzymatic fragmentation methods. Physical fragmentation methods can include nebulization, sonication, and/or hydrodynamic shearing. In some embodiments, the fragmentation can be accomplished mechanically comprising subjecting the nucleic acids in the input sample to acoustic sonication. In some embodiments, the fragmentation comprises treating the nucleic acids in the input sample with one or more enzymes under conditions suitable for the one or more enzymes to generate double-stranded nucleic acid breaks. Examples of enzymes useful in the generation of nucleic acid or polynucleotide fragments include sequence specific and non-sequence specific nucleases. Non-limiting examples of nucleases include DNase I, Fragmentase, restriction endonucleases, variants thereof, and combinations thereof. Reagents for carrying out enzymatic fragmentation reactions are commercially available (e.g, from New England Biolabs). For example, digestion with DNase I can induce random double-stranded breaks in DNA in the absence of Mg^{++} and in the presence of Mn^{++}. In some embodiments, fragmentation comprises treating the nucleic acids in the input sample with one or more restriction endonucleases. Fragmentation can produce fragments having 5' overhangs, 3' overhangs, blunt ends, or a combination thereof. In some embodiments, such as when fragmentation comprises the use of one or more restriction endonucleases, cleavage of sample polynucleotides leaves overhangs having a predictable sequence. In some embodiments, the method includes the step of size selecting the fragments via standard methods known in the art such as column purification or isolation from an agarose gel.

[0024] In some embodiments, the nucleic acids in the input sample can be fragmented into a population of fragmented nucleic acid molecules or polynucleotides of one or more specific size range(s). In some embodiments, the fragments have an average length from about 10 to about 10,000 nucleotides. In some embodiments, the fragments have an average length from about 50 to about 2,000 nucleotides. In some embodiments, the fragments have an average length from about 100-2,500, 10-1,000, 10-800, 10-500, 50-500, 50-250, or 50-150 nucleotides. In some embodiments, the fragments have an average length less than 10,000 nucleotide, such as less than 5,000 nucleotides, less than 2,500 nucleotides, less than 2,500 nucleotides, less than 1,000 nucleotides, less than 500

nucleotides, such as less than 400 nucleotides, less than 300 nucleotides, less than 200 nucleotides, or less than 150 nucleotides.

[0025] In one embodiment, fragmentation of the nucleic acids can be followed by end repair of the nucleic acid fragments. End repair can include the generation of blunt ends, non-blunt ends (i.e sticky or cohesive ends), or single base overhangs such as the addition of a single dA nucleotide to the 3′-end of the nucleic acid fragments, by a polymerase lacking 3′-exonuclease activity. End repair can be performed using any number of enzymes and/or methods known in the art including, but not limited to, commercially available kits such as the Encore ™ Ultra Low Input NGS Library System I. In a preferred embodiment, end repair can be performed on double stranded DNA fragments to produce blunt ends wherein the double stranded DNA fragments contain 5' phosphates and 3' hydroxyls. In some embodiments, the double-stranded DNA fragments can be blunt-end polished (or "end repaired") to produce DNA fragments having blunt ends, prior to being joined to adapters. Generation of the blunt ends on the double stranded fragments can be generated by the use of a single strand specific DNA exonuclease such as for example exonuclease 1, exonuclease 7 or a combination thereof to degrade overhanging single stranded ends of the double stranded products. Alternatively, the double stranded DNA fragments can be blunt ended by the use of a single stranded specific DNA endonuclease, for example, but not limited to, mung bean endonuclease or S1 endonuclease. Alternatively, the double stranded products can be blunt ended by the use of a polymerase that comprises single stranded exonuclease activity such as for example T4 DNA polymerase, or any other polymerase comprising single stranded exonuclease activity or a combination thereof to degrade the overhanging single stranded ends of the double stranded products. In some cases, the polymerase comprising single stranded exonuclease activity can be incubated in a reaction mixture that does or does not comprise one or more dNTPs. In other cases, a combination of single stranded nucleic acid specific exonucleases and one or more polymerases can be used to blunt end the double stranded fragments generated by fragmenting the sample comprising nucleic acids. In still other cases, the nucleic acid fragments can be made blunt ended by filling in the overhanging single stranded ends of the double stranded fragments. For example, the fragments may be incubated with a polymerase such as T4 DNA polymerase or Klenow polymerase or a combination thereof in the presence of one or more dNTPs to fill in the single stranded portions of the double stranded fragments. Alternatively, the double

stranded DNA fragments can be made blunt by a combination of a single stranded overhang degradation reaction using exonucleases and/or polymerases, and a fill-in reaction using one or more polymerases in the presence of one or more dNTPs.

[0026] In some embodiments, the 5' and/or 3' end nucleotide sequences of fragmented nucleic acids are not modified or end-repaired prior to ligation with the adapter oligonucleotides of the present invention. For example, fragmentation by a restriction endonuclease can be used to leave a predictable overhang, followed by ligation with one or more adapter oligonucleotides comprising an overhang complementary to the predictable overhang on a nucleic acid fragment. In another example, cleavage by an enzyme that leaves a predictable blunt end can be followed by ligation of blunt-ended nucleic acid fragments to adapter oligonucleotides comprising a blunt end. In some embodiments, end repair can be followed by an addition of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 or more nucleotides, such as one or more adenine, one or more thymine, one or more guanine, or one or more cytosine, to produce an overhang. Nucleic acid fragments having an overhang can be joined to one or more adapter oligonucleotides having a complementary overhang, such as in a ligation reaction. For example, a single adenine can be added to the 3' ends of end repaired DNA fragments using a template independent polymerase, followed by ligation to one or more adapters each having a thymine at a 3' end. In some embodiments, adapter oligonucleotides can be joined to blunt end double-stranded nucleic acid fragments which have been modified by extension of the 3' end with one or more nucleotides followed by 5' phosphorylation. In some cases, extension of the 3' end can be performed with a polymerase such as for example Klenow polymerase or any of the suitable polymerases provided herein, or by use of a terminal deoxynucleotide transferase, in the presence of one or more dNTPs in a suitable buffer containing magnesium. In some embodiments, nucleic acid fragments having blunt ends can be joined to one or more adapters comprising a blunt end. Phosphorylation of 5' ends of nucleic acid fragments can be performed for example with T4 polynucleotide kinase in a suitable buffer containing ATP and magnesium. The fragmented nucleic acid molecules may optionally be treated to dephosphorylate 5' ends or 3' ends, for example, by using enzymes known in the art, such as phosphatases.

[0027] The methods described herein for enriching for target nucleic acid sequences further comprise appending a first adaptor to the nucleic acid fragments generated by the methods described herein. In one embodiment, the first adaptor can be a forward adaptor.

Appending the first adaptor to the nucleic acid fragments generated by methods described herein can be achieved using a ligation reaction or a priming reaction. In one embodiment, appendage of a first adaptor to the nucleic acid fragments comprises ligation. In one embodiment, ligation of the first adaptor to the nucleic acid fragments can be following end repair of the nucleic acid fragments. In another embodiment, the ligation of the first adaptor to the nucleic acid fragments can be following generation of the nucleic acid fragments without end repair of the nucleic acid fragments. The first adaptor can be any type of adaptor known in the art including, but not limited to, conventional duplex or double stranded adaptors in which the adaptor comprises two complementary strands. In a preferred embodiment, the first adaptor can be a double stranded DNA adaptor. In one embodiment, the first adaptor can be an oligonucleotide of known sequence and, thus, allow generation and/or use of sequence specific primers for amplification and/or sequencing of any polynucleotides to which the first adaptor(s) is appended or attached. In one embodiment, the first adaptor can be a conventional duplex adaptor, wherein the first adaptor comprises sequence well known in the art. In a preferred embodiment, the first adaptor can be appended to the nucleic acid fragments generated by the methods described herein in multiple orientations. In a preferred embodiment, the methods described herein can involve the use of a first duplex adaptor comprising double stranded DNA of known sequence that is blunt ended and can bind to the double stranded nucleic acid fragments generated by the methods described herein in one of two orientations. In one embodiment, the first adaptor can be ligated to each of the nucleic acid fragments such that each of the nucleic acid fragments comprises the same first adaptor. In other words, each of the nucleic acid fragments comprises a common first adaptor. In another embodiment, a first adaptor can be appended or ligated to a library of nucleic acid fragments generated by the methods described herein such that each nucleic acid fragment in the library of nucleic acid fragments comprises the first adaptor ligated to one or both ends.

[0028] In one embodiment, the first adaptor can be ligated or appended to the 5' and/or 3' ends of the nucleic acid fragments generated by the methods described herein. The first adaptor can comprise two strands wherein each strand comprises a free 3' hydroxyl group but neither strand comprises a free 5' phosphate. In one embodiment, the free 3' hydroxyl group on each strand of the first adaptor can be ligated to a free 5' phosphate present on either end of the nucleic acid fragments of the present invention. In this

embodiment, the first adaptor comprises a ligation strand and a non-ligation strand whereby the ligation strand can be ligated to the 5'phosphate on either end of the nucleic acid fragment while a nick or gap can be present between the non-ligation strand of the first adaptor and the 3' hydroxyl on either end of the nucleic acid fragment. In one embodiment, the nick or gap can be filled in by performing a gap repair reaction. In one embodiment, the gap repair can be performed with a DNA dependent DNA polymerase with strand displacement activity. In one embodiment, the gap repair can be performed using a DNA dependent DNA polymerase with weak or no strand displacement activity. In one embodiment, the ligation strand of the first adaptor can serve as the template for the gap repair or fill-in reaction. In this embodiment, the gap repair or fill-in reaction comprises an extension reaction wherein the ligation strand of the first adaptor serves as a template and leads to the generation of nucleic acid fragments with complementary termini or ends as depicted, for example, in FIG. 1. In one embodiment, the gap repair can be performed using *Taq* DNA polymerase. In one embodiment, the ligation of the first adaptor to the nucleic acid fragments generated by the methods described herein may not be followed gap repair as depicted, for example, in FIG. 2. In this embodiment, the nucleic acid fragments comprise first adaptor sequence ligated only at the 5' end of each strand.

[0029] Ligation and, optionally gap repair, of the first adaptor to the nucleic acid fragments generates a first adaptor-nucleic acid fragment complex. In one embodiment, the first adaptor-nucleic acid fragment complex can be denatured. Denaturation can be achieved using any of the methods known in the art including, but not limited to, physical, thermal, and/ or chemical denaturation. In one embodiment, denaturation can be achieved using thermal or heat denaturation. In one embodiment, denaturation of the first adaptor-nucleic acid fragment complex generates single stranded nucleic acid fragments comprising first adaptor sequence at only the 5'end of the nucleic acid fragments as depicted, for example, in FIG. 2. In another embodiment, denaturation of the first adaptor-nucleic acid fragment complex generates single stranded nucleic acid fragments comprising first adaptor sequence at both the 5'end and 3'end of the nucleic acid fragments as depicted, for example, in FIG.1.

[0030] In one embodiment, the nucleic acid fragments comprising first adaptor sequence appended to either the 5' end or both the 5' and 3' end can be denatured to generate single stranded nucleic acid fragments comprising first adaptor sequence appended to either the

5' end or both the 5' and 3' end. In one embodiment, the methods of the present invention described herein can be used to generate a plurality of single stranded nucleic acid fragments comprising first adaptor sequence appended to either the 5' end or both the 5' and 3' end. In one embodiment, an oligonucleotide comprising at a first end sequence complementary to a target sequence of interest present in a single stranded nucleic acid fragment and at a second end sequence from a second adaptor , wherein the second adaptor sequence is not complementary to the target nucleic acid can be annealed to the single stranded nucleic acid fragments. In one embodiment, the second adaptor sequence can be sequence from a reverse adaptor. In one embodiment, the target nucleic acid sequence of interest can be present in one or more of the single stranded nucleic acid fragments. In one embodiment, different or distinct target nucleic acid sequences of interest can be present in one or more of the single stranded nucleic acid fragments. In one embodiment, one or more oligonucleotides can comprise sequence complementary to the same sequence of interest present in one or more single stranded nucleic acid fragments. In this embodiment, the one or more oligonucleotides can comprise sequence that is complementary to different parts or regions of the same sequence of interest. In one embodiment, the different regions can be adjacent to each other. In one embodiment, the different regions can be non-adjacent to each other. In a preferred embodiment, the one or more oligonucleotides that comprise sequence complementary to the same target nucleic acid sequence of interest further comprise the same second adaptor sequence. In another embodiment, one or more oligonucleotides can comprise sequence complementary to different or distinct sequences of interest which can be present in one or more single stranded nucleic acid fragments. In a preferred embodiment, the one or more oligonucleotides that comprise sequence complementary to different or distinct target nucleic acid sequences of interest further comprise the same second adaptor sequence. In one embodiment, the sequence complementary to the target sequence of interest can be at the 3'end of the oligonucleotide and the second adaptor sequence can be at the 5'end of the oligonucleotide. In a preferred embodiment, the second adaptor sequence is non-complementary to the target nucleic acid sequence of interest. In this manner, the second adaptor sequence serves as a tail. The second adaptor sequence can be a conventional adaptor sequence. In a preferred embodiment, the second adaptor sequence can be a conventional adaptor sequence that is different than or distinct from the sequence of the first adaptor appended to the single stranded nucleic acid fragment as

described above. In one embodiment, the second adaptor sequence can be of known sequence and, thus, allow generation and/or use of sequence specific primers for amplification and/or sequencing of any polynucleotides to which the second adaptor sequence is appended or attached. In a separate embodiment, the oligonucleotide can be annealed to the nucleic acid fragments comprising the first adaptor sequence appended to either the 5' end or both the 5' and 3' end without prior denaturation. In this embodiment, annealing of the oligonucleotide can be via formation of a triple helix or triplex between the oligonucleotide and a double stranded nucleic acid fragment comprising the first adaptor sequence appended to either the 5' end or both the 5' and 3' ends of the double stranded nucleic acid fragment. In this embodiment, the double stranded nucleic acid fragment comprises a sequence of interest and can be present amongst a plurality of double stranded nucleic acid fragments comprising first adaptor sequence appended to either the 5' end or both the 5' and 3' end. Further to this embodiment, the oligonucleotide comprises sequence complementary to the sequence of interest in the double stranded nucleic acid fragment. Overall, the use of the oligonucleotide comprising sequence complementary to a target sequence of interest present in a nucleic acid fragment amongst one or more or a plurality of nucleic acid fragments allows for selective binding and subsequent enrichment of said nucleic fragment using the methods described herein.

[0031] Following annealing of the oligonucleotide as described above, a polymerase can be used to extend the oligonucleotide. In one embodiment, the polymerase can be a DNA dependent DNA polymerase. In one embodiment, the DNA dependent DNA polymerase can be any of the DNA dependent DNA polymerases as described herein and extension of the oligonucleotide can be by any of the methods known in the art. In one embodiment, an oligonucleotide comprising the second adaptor sequence, wherein the second adaptor sequence is not complementary to the target nucleic acid, and sequence complementary to a target sequence of interest present in a nucleic acid fragment comprising a first adaptor appended to one and/or both ends can be annealed to the nucleic acid fragment and extended with a polymerase to generate an oligonucleotide extension product comprising the first adaptor sequence at a first end and the second adaptor sequence at a second end. In one embodiment, the nucleic acid fragment can be present amongst a plurality of nucleic acid fragments comprising first adaptor appended to one and/or both ends. In this

embodiment, the oligonucleotide extension product can only be generated for a nucleic acid fragment that contains the target sequence of interest.

[0032] In one embodiment, the oligonucleotide extension product generated by the methods described herein can be subjected to an amplification reaction. In one embodiment, the amplification reaction can be exponential, and may be carried out at various temperature cycles or isothermal. In one embodiment, the amplification can be polymerase chain reaction. In one embodiment, the amplification reaction can be isothermal. In one embodiment, the oligonucleotide extension product comprises first adaptor sequence on one end and second adaptor sequence on the other end as generated by the methods described herein. In a preferred embodiment, the oligonucleotide extension product can be separated from the template nucleic acid fragment in order to generate a single stranded oligonucleotide extension product with first adaptor sequence on the 5' end and second adaptor sequence on the 3' end. The single stranded oligonucleotide extension product can then be amplified using a first primer comprising sequence complementary to the first adaptor and a second primer comprising sequence complementary to the second adaptor sequence. In this manner only oligonucleotide extension products comprising both the first and the second adaptor sequence will be amplified and thus enriched. In one embodiment, the first adaptor and/or the second adaptor sequence can comprise an identifier sequence. In one embodiment, the identifier sequence can be barcode sequence. In one embodiment, the barcode sequence can be the same or different for the first adaptor and the second adaptor sequence. In one embodiment, the first adaptor and/or the second adaptor sequence can comprise sequence that can be used for downstream applications such as, for example, but not limited to, sequencing. In one embodiment, the first adaptor and/or the second adaptor sequence can comprise flow cell sequences which can be used for sequencing with the sequencing method developed by Illumina and described herein.

[0033] In an alternate embodiment, the methods of the present invention can be used to generate a library of nucleic acid fragments or inserts wherein each nucleic acid fragment comprises an adaptor at one or both ends. In one embodiment, the adaptors can be present at both ends and can be distinct from each other. In one embodiment, the adaptors can be present at both ends and can comprise the same adaptor sequence. The generation of the library comprising nucleic acid inserts with distinct adaptors at both ends can involve the methods for generating oligonucleotide extension products comprising first

adaptor sequence on one end and second adaptor sequence on the other end as described above with the exception that the oligonucleotide that binds to the nucleic acid fragments and can be extended comprises random sequence. In this embodiment, the oligonucleotide comprises random sequence at the 3' portion that is hybridizable to one or more nucleic acid fragments and further comprises second adaptor sequence at the 5'-portion. Extension of the oligonucleotide along the nucleic acid fragment and the corresponding first adaptor generates a product, or products, comprising the second adaptor at one end and a sequence complementary to the first adaptor at the other end, as illustrated in Figure 4. In one embodiment, the random sequence present in the oligonucleotide can bind to and be extended on one or more nucleic acid inserts. In one embodiment, one or more oligonucleotides comprising a 3' portion comprising random sequence and a 5' portion comprising second adaptor sequence can be annealed to a library comprising nucleic acid inserts comprising a first adaptor sequence on one or both ends of each of the nucleic acid inserts. In one embodiment, the first adaptor can be the same or common to each of the nucleic acid inserts. In one embodiment, the second adaptor sequence can be the same or common to each of the one or more oligonucleotides. In one embodiment, the methods described above can be used to generate a library of nucleic acid inserts comprising wherein each of the nucleic acid inserts comprises a common first adaptor on one end and a common second adaptor sequence on a second end. In one embodiment, the first adaptor and the second adaptor sequence can be distinct from each other. In one embodiment, the first adaptor and the second adaptor sequence can comprise the same adaptor sequence. Overall, the methods of the present invention as described above can be used for the high efficiency generation of a library of nucleic acid sequences.

[0034] In yet another alternate embodiment to the methods of the invention as described above, the first adaptor can be a double stranded DNA adaptor comprising a partial duplex, wherein the two strands of the adaptor can be different lengths with a complementary region and an overhanging region at the 5' end. In this embodiment, the 5' end of the long strand of the partial duplex adaptor can comprise a unique site for a nucleic acid modifying enzyme, such as a restriction enzyme, that is absent from the short strand of the duplex adaptor. In a further embodiment, the 3' end of the short strand adaptor can be modified by a replacement of the 3' OH-group by a blocking group, for example, a dideoxynucleotide (ddCMP, ddAMP, ddTMP, or ddGMP) to prevent

polymerase extension. In this embodiment, the first adaptor comprising the partial duplex can be ligated to nucleic acid fragments generated by the methods described herein. In one embodiment, ligation of the partial duplex first adaptor can be followed by a gap repair reaction as described above. In this embodiment, ligation of the partial duplex first adaptor is not followed by a gap repair reaction. In a preferred embodiment, the partial duplex first adaptor comprises a free 5' phosphate on the short strand and a free 3'hydroxyl on the long strand. In this embodiment, ligation of the partial duplex adaptor generates double stranded nucleic acid fragments wherein both ends of the double stranded nucleic acid fragment comprise the long strand and short strand of the partial duplex first adaptor. A double stranded partial duplex first adaptor-nucleic acid fragment complex can be generated by ligation. In one embodiment, the double stranded partial duplex first adaptor-nucleic acid fragment complex can be denatured to generate a single stranded nucleic acid fragment comprising the long strand of the first adaptor on a first end and the short strand of the first adaptor on a second end. In this embodiment, the first end is the 5' end and the second end is the 3' end. In one embodiment, the first adaptor can be appended to one or more nucleic acid fragments as generated by the methods described herein such that each of the nucleic acid fragments comprises the same first adaptor or, in other words, the first adaptor can be common to each of the nucleic acid inserts. An oligonucleotide or primer comprising sequence complementary to a sequence of interest in the single stranded nucleic acid fragment can be annealed to the single stranded nucleic acid fragment and extended using a polymerase. In one embodiment, the polymerase can be a DNA dependent DNA polymerase. In one embodiment, the DNA dependent DNA polymerase can be any of the DNA dependent DNA polymerases as described herein and extension of the oligonucleotide can be by any of the methods known in the art. Extension of the primer annealed to the single stranded nucleic acid fragment generates an oligonucleotide extension product comprising sequence complementary to the long strand of the first adaptor on one end. In one embodiment, the oligonucleotide extension product remains hybridized to the single stranded nucleic acid fragment such that the restriction and/or cleavage site specific for a nucleic acid modifying enzyme is made double stranded. The double stranded site can then be cleaved by the nucleic acid modifying enzyme specific for the double stranded restriction site. In one embodiment, the nucleic acid modifying enzyme can be a restriction enzyme. In one embodiment, the restriction enzyme can be specific for a double stranded restriction site.

In one embodiment, cleavage of the restriction site can generate a blunt end or non-blunt end. In one embodiment, end repair by any of the methods described herein can be performed on the end of the nucleic acid fragment following cleavage. Cleavage of the restriction and/or cleavage site generates a site to which a second adaptor can be ligated. Ligation of the second adaptor can be through any of the methods for ligation as described herein. In one embodiment, ligation generates a double stranded nucleic acid fragment comprising the second adaptor on a first end and a partial duplex on a second end, wherein the partial duplex comprises a 3' overhang comprising the sequence of the short strand of the first adaptor. The double stranded nucleic acid fragment can then be denatured using any of the methods for denaturation disclosed herein to generate a single stranded nucleic acid fragment comprising the second adaptor sequence on the first end and the sequence of the short strand of the first adaptor on the second end. In one embodiment, the first end and second end comprise the 5' end and 3' end, respectively. In one embodiment, the second adaptor can be appended to one or more nucleic acid fragments following cleavage of the double stranded restriction site such that each of the nucleic acid fragments comprises the same second adaptor or, in other words, the second adaptor can be common to each of the nucleic acid inserts. The single stranded nucleic acid fragment can then be amplified using a first primer specific for the second primer and a second primer specific for sequence present in the short strand of the first adaptor. In one embodiment, the amplification reaction can be exponential, and may be carried out at various temperature cycles or isothermal. In one embodiment, the amplification can be polymerase chain reaction. In one embodiment, the amplification reaction can be isothermal. Overall, only a fragment comprising the second adaptor and the short strand of the first adaptor will be amplified or enriched. In so far as the method provides for enrichment of targeted fragments of the library, and not enrichment of oligonucleotide extension products generated by the extension of the oligonucleotide comprising sequence complementary to a target sequence of interest, there is no distortion of the original DNA library, and the enrichment is independent of the insert length. Because the 3' end of the short strand of the partial duplex adaptor is 3' blocked, the method enables directional or asymmetric ligation. In one embodiment, the oligonucleotide that comprises sequence complementary to a sequence of interest in a nucleic acid fragment further comprises reverse adaptor sequence. In this embodiment, the sequence complementary to a sequence of interest in the nucleic acid fragment can be present in a 3' portion of the

oligonucleotide and the reverse adaptor sequence can be present at a 5' portion. Further to this embodiment, the reverse adaptor sequence can be a common or conventional adaptor sequence and can be different or distinct from the first and/or second adaptors. Further still to this embodiment, the methods described above can lead to the generation of a single stranded nucleic acid fragment comprising the second adaptor at one end and the reverse adaptor sequence at the other end. Subsequent to this embodiment, the single stranded nucleic acid fragment can be enriched through amplification using a first primer specific to the second adaptor and a second primer specific to the third adaptor sequence.

[0035] The methods of the inventions are further applicable to any enrichment of target nucleic acid sequences of interest from libraries comprising fragments of nucleic acid of a sample appended with adaptor sequence at one or both ends, wherein the libraries are generated using ligation of the adaptor or adaptor sequences to one or both ends as described herein or by ligation independent methods, such as for example Nextera, a transposome driven method. In one embodiment, the nucleic acid can be DNA such as genomic DNA or cDNA. In one embodiment, the nucleic acid can be double stranded. Enrichment of nucleic acid sequences of interest can be achieved using the methods described herein for target enrichment. In one embodiment, the method for enriching for target nucleic acid sequences of interest from a library comprising nucleic inserts with adaptors appended to one or both ends comprises denaturing the nucleic acid inserts to generate a library of single stranded nucleic acid inserts. In one embodiment, each of the nucleic acid inserts can comprise a first adaptor sequence on one end and a second adaptor sequence on an opposite end. In one embodiment, the first adaptor and the second adaptor can be distinct from each other. In one embodiment, the first adaptor and the second adaptor can comprise the same adaptor sequence. In one embodiment, each of the nucleic acid inserts can comprise a first adaptor sequence on one end and a second adaptor sequence on an opposite end such that denaturation generates a library of single stranded nucleic acid inserts comprising the first adaptor sequence on one end and the second adaptor sequence on an opposite end. Denaturation can be achieved using any of the methods described herein. Further to the embodiments described above, one or more oligonucleotides can be annealed to the single stranded nucleic acid inserts. In one embodiment, each of the one or more oligonucleotides comprises a 3' portion that is complementary to a target nucleic acid sequence of interest present in one or more of the nucleic acid inserts, and a 5' portion comprising a third adaptor sequence. In one

embodiment, the third adaptor sequence is distinct from either or both of the first adaptor and the second adaptor. The one or more oligonucleotides can be extended with a polymerase (i.e. a DNA polymerase) thereby generating one or more oligonucleotide extension products with the first or second adaptor at a first end and the third adaptor sequence at a second end. In one embodiment, the first end comprises the 5' end and the second end comprises the 3' end. The one or more oligonucleotide extension products can be amplified using a first primer that can be complementary to the first or second adaptor and a second primer that can be complementary to the third adaptor sequence to enrich for nucleic acid fragments comprising the first or second adaptor and the third adaptor sequence at each end. In one embodiment, the first and second adaptors can be common to each of the nucleic acid inserts in the library. In one embodiment, the third adaptor sequence can be common to each of the one or more oligonucleotides. Overall, the target enrichment methods as described above can be used to generate a composition comprising a library of nucleic acid inserts enriched for any target sequence of interest from a non-enriched library comprised of nucleic acid inserts with an adaptor ligated to one or both ends.

[0036] A schematic of a preferred embodiment of the methods described herein for enriching for target sequences of interest is illustrated in FIG. 1. Overall, FIG. 1 depicts a method for isolating or enriching for a nucleic acid fragment or insert comprising a target nucleic acid sequence from a library or plurality of nucleic acid fragments. The method in FIG. 1 involves generation of a ligated library of nucleic acid fragments or inserts wherein each fragment or insert of the ligated library comprises a common forward adaptor and a fragment or insert specific reverse adaptor distinct from the forward adaptor such that subtractive PCR using a primer directed against the common forward adaptor and a primer directed against the reverse adaptor enriches for a nucleic acid fragment or insert comprising a target nucleic acid sequence. The input for the method depicted in FIG.1 is fragmented DNA. The fragmented DNA is double stranded and comprises a plurality or library of DNA fragments. In one embodiment, the DNA fragments can be derived from complex DNA, such as double-stranded DNA, genomic DNA or mixed DNA from more than one organism. In one embodiment, the DNA fragments can be derived from RNA that has been converted to cDNA through a first strand synthesis reaction using any of the methods well known in the art for generating cDNA from an RNA template which can include, but is not limited to, combining the RNA with a primer

(i.e. random primer), and reverse transcribing the RNA template with an RNA-dependent DNA polymerase. In one embodiment, the DNA fragments can be derived from RNA that has been converted to double stranded cDNA through a first and second strand synthesis reaction using any of the methods well known in the art. Fragmentation of the DNA to produce the DNA fragments can be achieved through any of the methods described herein for fragmenting nucleic acids which can include, but are not limited to, physical (i.e. sonication), and/or chemical (i.e. restriction enzyme treatment) fragmentation reactions.

[0037] As depicted in FIG. 1, a single forward adaptor is ligated to the DNA fragments. In one embodiment, the single forward adaptor can comprise known sequence. In one embodiment, the single forward adaptor can be a common adaptor. In one embodiment, the DNA fragments can be subjected to an end repair reaction as described herein to produce blunt ends. In this embodiment, the single forward adaptor can also comprise blunt ends and ligation between the single forward adaptor and the DNA fragments can be through blunt end ligation as described herein. Ligation can be facilitated through the use of enzymes (i.e. T4 DNA ligase) and methods known in the art, including, but not limited to, commercially available kits such as the Encore ™ Ultra Low Input NGS Library System. In FIG. 1, the forward adaptor can contain a strand (the ligation strand) that ligates with the free 5'phosphate on a 5' end of the DNA fragments and a strand that does not ligate (non-ligation strand) to a 3' end of the DNA fragments. In one embodiment, the ligation reaction can lead to the generation of a nick or gap between the non-ligation strand of the single forward adaptor and the 3' end of the DNA fragments. In this embodiment, the nick or gap can be repaired or filled in through a gap repair or fill-in reaction wherein the 3' end of the DNA fragments can be extended with a polymerase (preferably with a DNA dependent DNA polymerase such as *Taq* DNA polymerase) wherein the ligation strand of the forward adaptor can serve as template. In this embodiment, the gap repair generates DNA fragments with complementary ends. As depicted in FIG. 1, the DNA fragments with complementary ends are denatured to generate a denatured library comprising single stranded DNA fragments with complementary ends. Denaturation can be achieved using any of the methods known in the art which can include, but are not limited to, heat denaturation, and/or chemical denaturation.

[0038] As depicted in FIG. 1, a custom oligonucleotide with a reverse adaptor tail is annealed to the single stranded DNA fragments with complementary ends. In one embodiment, the custom oligonucleotide with a reverse adaptor tail can comprise a 3' portion comprising sequence complementary to a target sequence of interest in one of the single-stranded DNA fragments and a 5' portion comprising reverse adaptor sequence that is not complementary to the single-stranded DNA fragments in the denatured library. In one embodiment, the reverse adaptor sequence can be known sequence. In a preferred embodiment, the reverse adaptor sequence can be distinct from the single forward adaptor as described herein. In one embodiment, a plurality of custom oligonucleotides with a reverse adaptor tail can be added to the denatured library wherein the plurality of custom oligonucleotides with a reverse adaptor tail comprise a 3' portion comprising sequence complementary to a target sequence of interest in one or more of the single-stranded DNA fragments of the denatured library and a 5' portion comprising a reverse adaptor sequence that is not complementary to the single-stranded DNA fragments in the denatured library. In one embodiment, the reverse adaptor tail comprises the same reverse adaptor sequence in each of the plurality of custom oligonucleotides with a reverse adaptor tail, and wherein the reverse adaptor sequence is distinct from the forward adaptor sequence. In another embodiment, the reverse adaptor tail comprises a different reverse adaptor sequence for each of the plurality of custom oligonucleotides with a reverse adaptor tail, and wherein each of the different reverse adaptor sequences is distinct from the forward adaptor sequence. In one embodiment, the 3' portion of the custom oligonucleotide with a reverse adaptor tail can be a specific sequence, wherein the custom oligonucleotide comprises a sequence complementary to the target sequence of interest and provides a means for targeted enrichment of sequence or sequences of interest using the methods of the invention. In another embodiment the 3' portion of the custom oligonucleotide with reverse adaptor tail can be a randomly generated sequence hybridizable to random sequences of the library of fragments with adaptor sequences on one or both ends, providing means for efficient, non-enriched library generation employing the methods of the invention.

[0039] Following annealing of the custom oligonucleotide with a reverse adaptor tail to a sequence of interest in a single-stranded DNA fragment of the denatured library, the custom oligonucleotide with a reverse adaptor tail is extended using any method known in the art, which can include but is not limited to, extension using a DNA dependent DNA

polymerase using the single stranded DNA fragment of the denatured library as a template. Extension of the custom oligonucleotide with a reverse adaptor tail generates an oligonucleotide extension product with forward adaptor sequence at one end and reverse adaptor sequence at the other end. In this embodiment, the custom oligonucleotide with a reverse adaptor tail can only anneal to and be extended on DNA fragments in the denatured library comprising the target sequence of interest for which the custom oligonucleotide with a reverse adaptor tail is directed. As illustrated in FIG. 1, a subtractive polymerase chain reaction (PCR) procedure is subsequently performed using a first primer directed against the forward adaptor sequence and a second primer directed toward the reverse adaptor sequence such that only the oligonucleotide extension product with the forward adaptor sequence at one end and the reverse adaptor sequence at the other end can be amplified and thus enriched.

[0040] FIG. 2 depicts another embodiment of subtractive PCR enrichment method as described for FIG. 1, wherein ligation of a duplex forward adaptor (P1) to a double stranded nucleic acid fragment is not subjected to gap repair. The duplex forward adaptor (P1) comprises a strand (the ligation strand; P1) that ligates with the free 5'phosphate on a 5' end of the nucleic acid fragment and a strand that does not ligate (non-ligation strand; P1rc) to a 3' end of the nucleic acid fragment. In this embodiment, the ligation strand is ligated to the 5' end of both strands of a double stranded nucleic fragment whereas a gap or nick is generated between the non-ligation strand and the 3'end of both strands of the double stranded nucleic acid fragment. As depicted in FIG. 2, the ligation of the P1 adaptor to the nucleic acid fragment is followed by denaturation without a gap repair or fill-in reaction thereby generating a single stranded nucleic acid fragment with non-complementary ends. In this embodiment, the single stranded nucleic acid fragment with non-complementary ends comprises a P1 forward adaptor sequence at a 5' end and fragment specific sequence at a 3' end. In a further embodiment, the single stranded nucleic fragment with non-complementary ends can be further processed as described above and illustrated in FIG. 1 to generate a single stranded nucleic fragment with the P1 forward adaptor sequence at the 5' end and a distinct reverse adaptor sequence on the 3' end.

[0041] FIG. 4 illustrates another embodiment of the present invention for the high efficiency generation of libraries comprising nucleic acid fragments with distinct adaptors on each end. In this embodiment, the methods for generating a denatured library

comprising single stranded DNA fragments with a single forward adaptor sequence on a 5' end and a reverse adaptor sequence on a 3' end is similar to the method described for FIG. 1. However, in FIG. 4, an oligonucleotide with a reverse adaptor tail is used whereby the oligonucleotide comprises a 3' portion comprising random sequence and a reverse adaptor tail wherein the reverse adaptor tail comprises reverse adaptor sequence that is distinct from the single forward adaptor sequence. As illustrated in FIG. 4, PCR can be carried out with a first primer directed against the single forward adaptor sequence and a second primer directed toward the reverse adaptor sequence wherein both the first and second primers further comprise flow cell sequences. In this manner, single stranded DNA fragments with a single forward adaptor sequence on a 5' end and a reverse adaptor sequence on a 3' end comprise flow cell sequences that can be used to adhere the amplified single stranded DNA fragments to flow cells for subsequent sequencing by the method commercialized by Illumina, as described U.S. Pat. Nos. 5,750,341; 6,306,597; and 5,969,119.

[0042] FIG. 3 illustrates a method for enrichment of a target nucleic acid sequence, or sequences of interest, contained in a double stranded nucleic acid insert from a complex library. In one embodiment, the complex library comprises nucleic acid inserts from a genomic DNA sample, In FIG. 3, the single forward adaptor comprises a partial duplex forward adaptor comprising a long strand, A, that forms a partial duplex with a short strand, B. Strand A of the partial duplex adaptor further comprises a restriction enzyme site while strand B does not contain the restriction enzyme site. Strand B further comprises a blocking group wherein the 3' end of strand B is modified by replacement of the 3' OH group with a blocking group that can prevent polymerase extension. In one embodiment, the partial duplex forward adaptor is ligated to the double stranded nucleic acid fragments such that a double stranded insert with the partial duplex forward adaptor appended to both ends is generated. In this embodiment, the 5' end of strand B of the partial duplex adaptor can contain a free 5' phosphate which can be ligated to a free 3' OH present on one or both strands of the double stranded insert. Subsequent denaturation generates a single stranded insert comprising sequence A on a 5'end and strand B on a 3' end. A primer, C, directed against a specific sequence of interest within the single stranded insert can be annealed to the specific sequence and extended with a DNA polymerase using the single stranded insert as a template. In one embodiment, primer C can be a sequence specific primer and is employed for enrichment of target, or targets, of

interest according to the methods of the invention. In one embodiment, primer C can be a random primer. Extension of primer C with DNA polymerase generates an extended primer C product that comprises sequence complementary to sequence A at its' 3' end in a double stranded complex with the template insert strand such that a double stranded restriction site has been generated between sequence A and its' complement. In one embodiment, the double stranded restriction enzyme recognition site can be cleaved by a restriction enzyme specific for the double stranded restriction site thereby generating a truncated, or cleaved, adaptor sequence at one end. A second forward adaptor comprising a common, or conventional, duplex adaptor D is then ligated to the cleavage site using any of the ligation methods described herein, thereby generating a double stranded complex comprising the second forward adaptor D at one end and a 3' overhang comprising strand B on the opposite end. The double stranded complex comprising the second forward adaptor D at one end and a 3' overhang comprising strand B on the opposite end is denatured and amplified using a first primer directed against the second forward adaptor D and a second primer directed against strand B. In this manner, the methods depicted in FIG. 3 can be used to enrich for specific sequences of interest from a complex library since the methods are designed such that the second forward adaptor D can only bind to the double stranded cleavage site generated by restriction enzyme digestion of the double stranded restriction site created between sequence A and its complement following extension of primer C. As described, primer C can be directed against a target sequence of interest present in a single, or multiple insert, or inserts, amongst a plurality of inserts. Moreover, the method can be made strand specific by designing primer C to bind to target sequence of interest present on one strand or the other of an insert in an amongst a plurality of inserts.

[0043] Unless otherwise specified, terms and symbols of genetics, molecular biology, biochemistry and nucleic acid used herein follow those of standard treatises and texts in the field, e.g. Kornberg and Baker, *DNA Replication*, Second Edition (W.H. Freeman, New York, 1992); Lehninger, *Biochemistry*, Second Edition (Worth Publishers, New York, 1975); Strachan and Read, Human *Molecular Genetics*, Second Edition (Wiley-Liss, New York, 1999); Eckstein, editor, *Oligonucleotides and Analogs: A Practical Approach* (Oxford University Press, New York, 1991); Gait, editor, *Oligonucleotide Synthesis: A Practical Approach* (IRL Press, Oxford, 1984); and the like.

Input Nucleic Acid

[0044] The input can be a nucleic acid. In one embodiment, the input can be DNA. In one embodiment, the input nucleic acid can be complex DNA, such as double-stranded DNA, genomic DNA or mixed DNA from more than one organism. In one embodiment, the input can be RNA. In one embodiment, the RNA can be obtained and purified using standard techniques in the art and include RNAs in purified or unpurified form, which include, but are not limited to, mRNAs, tRNAs, snRNAs, rRNAs, retroviruses, small non-coding RNAs, microRNAs, polysomal RNAs, pre-mRNAs, intronic RNA, viral RNA, cell free RNA and fragments thereof. The non-coding RNA, or ncRNA may include snoRNAs, microRNAs, siRNAs, piRNAs and long nc RNAs. In one embodiment, the input nucleic acid can be cDNA. The cDNA can be generated from RNA, e.g., mRNA. The cDNA can be single or double stranded. The input DNA can be of a specific species, for example, human, rat, mouse, other animals, specific plants, bacteria, algae, viruses, and the like. The input complex also can be from a mixture of genomes of different species such as host-pathogen, bacterial populations and the like. The input DNA can be cDNA made from a mixture of genomes of different species. Alternatively, the input nucleic acid can be from a synthetic source. The input DNA can be mitochondrial DNA. The input DNA can be cell-free DNA. The cell-free DNA can be obtained from, e.g., a serum or plasma sample. The input DNA can comprise one or more chromosomes. For example, if the input DNA is from a human, the DNA can comprise one or more of chromosome 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, X, or Y. The DNA can be from a linear or circular genome. The DNA can be plasmid DNA, cosmid DNA, bacterial artificial chromosome (BAC), or yeast artificial chromosome (YAC). The input DNA can be from more than one individual or organism. The input DNA can be double stranded or single stranded. The input DNA can be part of chromatin. The input DNA can be associated with histones.

[0045] In some embodiments, the oligonucleotides targeting the selected sequence regions of interest are designed to hybridize to single-stranded nucleic acid targets. In one embodiment, the oligonucleotides targeting the selected sequence regions of interest are designed to hybridize to single-stranded DNA targets. In the case where the input nucleic acid sample comprises genomic DNA or other double-stranded DNA, the input nucleic acid sample can be first denatured to render the target single stranded and enable hybridization of the oligonucleotides to the desired sequence regions of interest. In these

embodiments, the methods and compositions described herein can allow for region-specific enrichment and amplification of sequence regions of interest. In some embodiments, the other double-stranded DNA can be double-stranded cDNA generated by first and second strand synthesis of one or more target RNAs.

[0046] In other embodiments, the oligonucleotides targeting the selected sequence regions of interest are designed to hybridize to double-stranded nucleic acid targets, without denaturation of the double stranded nucleic acids. In other embodiments, the oligonucleotides targeting the selected sequence regions of interest are designed to hybridize to a double-stranded DNA target, without denaturation of the dsDNA. In these embodiments, the oligonucleotides targeting the selected sequence regions of interest are designed to form a triple helix (triplex) at the selected sequence regions of interest. The hybridization of the oligonucleotides to the double-stranded DNA sequence regions of interest can be carried out without prior denaturation of the double stranded nucleic acid sample. In such embodiments, the methods and compositions described herein can allow for region-specific enrichment as well as strand-specific enrichment and amplification of sequence regions of interest. This method can be useful for generation of copies of strand specific sequence regions of interest from complex nucleic acid without the need to denature the dsDNA input DNA, thus enabling enrichment and analysis of multiplicity of sequence regions of interest in the native complex nucleic acid sample. The method can find use for studies and analyses carried out in situ, enable studies and analysis of complex genomic DNA in single cells or collection of very small well defined cell population, as well as permit the analysis of complex genomic DNA without disruption of chromatin structures.

[0047] A "target nucleic acid sequence" or "target sequence" as used herein, is a polynucleotide sequence of interest, for which enrichment is desired. The target sequence may be known or not known, in terms of its actual sequence. Generally, a "template", as used herein, is a polynucleotide that contains the target nucleic acid sequence. The terms "target sequence," "target nucleic acid sequence," "target nucleotide sequence," "regions of interest," or "sequence of interest" and, variations thereof, are used interchangeably.

Oligonucleotides of the invention

[0048] As used within the invention, the term "oligonucleotide" refers to a polynucleotide chain, typically less than 200 residues long, most typically between 15 and

100 nucleotides long, but also intended to encompass longer polynucleotide chains. Oligonucleotides may be single-or double-stranded. As used in this invention, the term "oligonucleotide" may be used interchangeably with the terms "primer" and "adaptor".

[0049] As used herein, the terms "hybridization"/ "hybridizing" and "annealing" are used interchangeably and refer to the pairing of complementary nucleic acids.

[0050] The term "primer", as used herein, can refer to a nucleotide sequence, generally with a free 3' hydroxyl group, that is capable of hybridizing with a template (such as one or more target polynucleotides, one or more target DNAs, one or more target RNAs or a primer extension product) and is also capable of promoting polymerization of a polynucleotide complementary to the template. A primer can be, for example, an oligonucleotide. It can also be, for example, a sequence of the template (such as a primer extension product or a fragment of the template created following RNase [i.e. RNase H] cleavage of a template-DNA complex) that is hybridized to a sequence in the template itself (for example, as a hairpin loop), and that is capable of promoting nucleotide polymerization. Thus, a primer can be an exogenous (e.g., added) primer or an endogenous (e.g., template fragment) primer. A primer may contain a non-hybridizing sequence that constitutes a tail of the primer. A primer may still be hybridizing to a target even though its sequences are not fully complementary to the target.

[0051] The primers of the invention are generally oligonucleotides that are employed in an extension reaction by a polymerase along a polynucleotide template, such as in PCR, SPIA or cDNA synthesis, for example. The oligonucleotide primer can be a synthetic polynucleotide that is single stranded, containing a sequence at its 3′-end that is capable of hybridizing with a sequence of the target polynucleotide. Normally, the 3′ region of the primer that hybridizes with the target nucleic acid has at least 80%, preferably 90%, more preferably 95%, most preferably 100%, complementarity to a sequence or primer binding site.

[0052] "Complementary", as used herein, can refer to complementarity to all or only to a portion of a sequence. The number of nucleotides in the hybridizable sequence of a specific oligonucleotide primer should be such that stringency conditions used to hybridize the oligonucleotide primer will prevent excessive random non-specific hybridization. Usually, the number of nucleotides in the hybridizing portion of the oligonucleotide primer will be at least as great as the defined sequence on the target polynucleotide that the oligonucleotide primer hybridizes to, namely, at least 5, at least 6,

at least 7, at least 8, at least 9, at least 10, at least 11, at least 12, at least 13, at least 14, at least 15, at least about 20, and generally from about 6 to about 10 or 6 to about 12 or 12 to about 200 nucleotides, usually about 20 to about 50 nucleotides. In general, the target polynucleotide is larger than the oligonucleotide primer or primers as described previously.

[0053] In some cases, the identity of the investigated target polynucleotide sequence is known, and hybridizable sequence specific oligonucleotides or primers can be synthesized precisely according to the antisense sequence of the aforesaid target polynucleotide sequence. In some embodiments, multiple sequence-specific oligonucleotides or primers are employed to hybridize to a multiplicity of genomic regions of interest, allowing for selective enrichment of the regions of interest. In so far as the genomic regions may be very long, multiple oligonucleotides can be designed to hybridize to different sequence regions within the genomic regions of interest. In other embodiments, when the target polynucleotide sequence is unknown, the hybridizable sequence of an oligonucleotide or primer is a random sequence. Oligonucleotides or primers comprising random sequences may be referred to as "random primers", or "random oligonucleoties," as described herein. In one embodiment, an oligonucleotide or primer of the present invention hybridizable to a target sequence may comprise a mixture of primers or oilognucleotides designed to hybridize to a plurality (e.g. 2, 3, 4, about 6, 8, 10, 20, 40, 80, 100, 125, 150, 200, 250, 300, 400, 500, 600, 800, 1000, 1500, 2000, 2500, 3000, 4000, 5000, 6000, 7000, 8000, 10,000, 20,000, 25,000 or more) of target sequences. In some cases, the plurality of target sequences may comprise a group of related sequences, random sequences, a whole transcriptome or fraction (e.g. substantial fraction) thereof, or any group of sequences such as mRNA. In some embodiments, the primers can be directed to known sequences present in the adaptors used in the invention as described herein. In this embodiment, the primers can comprise groups of primers comprising one or more primers in each group, wherein each group of primers can be directed against distinct adaptors.

[0054] Tailed primers or oligonucleotides can be employed in certain embodiments of the invention. In general, a tailed primer comprises a 3' portion that is hybridizable to one or more target polynucleotides, and a 5' portion that is not hybridizable to the one or more target polynucleotides. In general, the non-hybridizable 5' portion does not hybridize to the one or more target polynucleotides under conditions in which the hybridizable 3'

portion of the tailed primer hybridizes to the one or more target polynucleotides. In some embodiments, the non-hybridizable 5' portion comprises an adaptor sequence. In some embodiments, the non-hybridizable 5' portion comprises a common or conventional adaptor sequence. In some embodiments, the non-hybridizable 5' portion comprises a common or conventional adaptor sequence that is distinct or different from the sequence of other adaptors used in the present invention. In some embodiments, the non-hybridizable 5' portion comprises a promoter-specific sequence. Generally, a promoter-specific sequence comprises a single-stranded DNA sequence region which, in double-stranded form is capable of mediating RNA transcription. Examples of promoter-specific sequences are known in the art, and include, without limitation, T7, T3, or SP6 RNA polymerase promoter sequences. When the tailed primer is extended with a DNA polymerase, a primer extension product with a 5' portion comprising a defined sequence can be created. This primer extension product can then have a second primer anneal to it, which can be extended with a DNA polymerase to create a double stranded product comprising a defined sequence at one end. In some embodiments, where the non-hybridizable 5' portion of one or more tailed primers comprises a promoter-specific sequence, creation of a double-stranded product comprising a defined sequence at one end generates a double-stranded promoter sequence that is capable of mediating RNA transcription. In some embodiments, a double-stranded promoter sequence can be generated by hybridizing to the promoter-specific sequence an oligonucleotide comprising a sequence complementary to the promoter-specific sequence. In some embodiments, formation of a double-stranded promoter can be followed by the generation of single-stranded RNA by RNA transcription of sequence downstream of the double-stranded promoter, generally in a reaction mixture comprising all necessary components, including but not limited to ribonucleoside triphosphates (rNTPs) and a DNA-dependent RNA polymerase. Tailed primers can comprise DNA, RNA, or both DNA and RNA. In some embodiments, the tailed primer consists of DNA.

[0055] Composite primers can be employed in certain embodiments of the invention. Composite primers are primers that are composed of RNA and DNA portions. In some aspects, the composite primer can be a tailed composite primer comprising, for example, a 3'-DNA portion and a 5'-RNA portion. In the tailed composite primer, a 3'-portion, all or a portion of which comprises DNA, is complementary to a polynucleotide; and a 5'-portion, all or a portion of which comprises RNA, is not complementary to the

polynucleotide and does not hybridize to the polynucleotide under conditions in which the 3'-portion of the tailed composite primer hybridizes to the polynucleotide target. When the tailed composite primer is extended with a DNA polymerase, a primer extension product with a 5'-RNA portion comprising a defined sequence can be created. This primer extension product can then have a second primer anneal to it, which can be extended with a DNA polymerase to create a double stranded product with an RNA/DNA heteroduplex comprising a defined sequence at one end. The RNA portion can be selectively cleaved from the partial heteroduplex to create a double-stranded DNA with a 3'-single-stranded overhang which can be useful for various aspects of the present invention including allowing for isothermal amplification using a composite amplification primer.

[0056] A "random primer," as used herein, can be a primer that generally comprises a sequence that is designed not necessarily based on a particular or specific sequence in a sample, but rather is based on a statistical expectation (or an empirical observation) that the sequence of the random primer is hybridizable (under a given set of conditions) to one or more sequences in the sample. A random primer will generally be an oligonucleotide or a population of oligonucleotides comprising a random sequence(s) in which the nucleotides at a given position on the oligonucleotide can be any of the four nucleotides, or any of a selected group of the four nucleotides (for example only three of the four nucleotides, or only two of the four nucleotides). In some cases all of the positions of the oligonucleotide or population of oligonucleotides can be any of two or more nucleotides. In other cases, only a portion of the oligonucleotide, for instance a particular region, will comprise positions which can be any of two or more bases. In some cases, the portion of the oligonucleotide which comprises positions which can be any of two or more bases is about 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, or about 15-20 nucleotides in length. In some cases, a random primer may comprise a tailed primer having a 3'-region that comprises a random sequence and a 5'-region that is a non-hybridizing sequence that comprises a specific, non-random sequence. The 3'-region may also comprise a random sequence in combination with a region that comprises poly-T sequences. The sequence of a random primer (or its complement) may or may not be naturally-occurring, or may or may not be present in a pool of sequences in a sample of interest. As is well understood in the art, a "random primer" can also refer to a primer that is a member of a population of primers (a plurality of random primers) which collectively are designed to hybridize to a desired

and/or a significant number of target sequences. A random primer may hybridize at a plurality of sites on a nucleic acid sequence. The use of random primers provides a method for generating primer extension products complementary to a target polynucleotide or target nucleic sequence which does not require prior knowledge of the exact sequence of the target. In some embodiments one portion of a primer is random, and another portion of the primer comprises a defined sequence. For example, in some embodiments, a 3'-portion of the primer will comprise a random sequence, while the 5'-portion of the primer comprises a defined sequence. In some embodiments a 3'-random portion of the primer will comprise DNA, and a 5'- defined portion of the primer will comprise RNA, in other embodiments, both the 3' and 5'-portions will comprise DNA. In some embodiments, the 5'-portion will contain a defined sequence and the 3'-portion will comprise a poly-dT sequence that is hybridizable to a multiplicity of RNAs in a sample (such as all mRNA). In some embodiments, a "random primer," or primer comprising a randomly generated sequence, comprises a collection of primers comprising one or more nucleotides selected at random from two or more different nucleotides, such that all possible sequence combinations of the nucleotides selected at random may be represented in the collection. In some embodiments, generation of one or more random primers does not include a step of excluding or selecting certain sequences or nucleotide combinations from the possible sequence combinations in the random portion of the one or more random primers.

[0057] In one embodiment, the oligonucleotides of the invention can be tailed oligonucleotides. In one embodiment, the 5'-tail can comprise RNA and is non hybridizable to the RNA in the sample. In one embodiment, the 5'-tail can comprise DNA and is non hybridizable to the DNA in the sample. In one embodiment, the 5'-tail can comprise an adaptor that is not hydridizable to the DNA and/or nucleic acid fragments derived from the sample comprising nucleic acid. In one embodiment, the 5'-tail can comprise an adaptor sequence that is not hydridizable to the DNA and/or nucleic acid fragments derived from the sample comprising nucleic acid. In some embodiments, the 5'-tail can comprise a common adaptor sequence that is not hydridizable to the DNA and is distinct from any other adaptor or adaptor sequence used in the methods of the invention described herein. In some embodiments, the 5'-tail can comprise an identifier sequence. In some embodiments, the identifier sequence can comprise a barcode

sequence. In some embodiments, the 5'-tail can comprise a common adaptor sequence that is not hydridizable to the DNA and a barcode sequence.

[0058] The term "adaptor", as used herein, refers to an oligonucleotide of known sequence, the ligation of which to a target polynucleotide or a target polynucleotide strand of interest enables the generation of amplification-ready products of the target polynucleotide or the target polynucleotide strand of interest. The target polynucleotide molecules may be fragmented or not prior to the addition of adaptors.

[0059] Various adaptor designs are envisioned which are suitable for generation of amplification-ready products of target sequence regions/strands of interest. For example, the two strands of the adaptor may be self-complementary, non-complementary or partially complementary. A common feature of the adaptors depicted in Figure 3 of the invention is the partial duplex design, wherein the two strands of the adaptor are different lengths with a complementary region and an overhanging region at the 5' end. The 5' end of the long strand of the partial duplex adaptor contains a unique site for a nucleic acid modifying enzyme, such as a restriction enzyme, that is absent from the short strand of the duplex adaptor. The 3' end of the short strand adaptor is modified by a replacement of the 3' OH-group is by a blocking group, for example, a dideoxynucleotide (ddCMP, ddAMP, ddTMP, or ddGMP) to prevent polymerase extension.

[0060] In some embodiments of the invention, the adaptors comprise an additional identifier sequence, i.e. a barcode sequence. As used herein, the term "barcode" refers to a known nucleic acid sequence that allows some feature of a polynucleotide with which the barcode is associated to be identified. In some embodiments, the feature of the polynucleotide to be identified is the sample from which the polynucleotide is derived. In some embodiments, barcodes are at least 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, or more nucleotides in length. In some embodiments, barcodes are shorter than 10, 9, 8, 7, 6, 5, or 4 nucleotides in length. In some embodiments, each barcode in a plurality of barcodes differ from every other barcode in the plurality at at least three nucleotide positions, such as at least 3, 4, 5, 6, 7, 8, 9, 10, or more positions. In some embodiments, barcodes associated with some polynucleotides are of different length than barcodes associated with other polynucleotides. In general, barcodes are of sufficient length and comprise sequences that are sufficiently different to allow the identification of samples based on barcodes with which they are associated. In some embodiments, both the forward and reverse adapter comprise at least one of a plurality of barcode sequences. In some

embodiments, the first, second, and/or third adaptor comprises at least one of a plurality of barcode sequences. In some embodiments, each reverse adapter comprises at least one of a plurality of barcode sequences, wherein each barcode sequence of the plurality of barcode sequences differs from every other barcode sequence in the plurality of barcode sequences. In some embodiments, both the first adapter and the second adapter comprise at least one of a plurality of barcode sequences. In some embodiments, barcodes for second adapter oligonucleotides are selected independently from barcodes for first adapter oligonucleotides. In some embodiments, first adapter oligonucleotides and second adapter oligonucleotides having barcodes are paired, such that adapters of the pair comprise the same or different one or more barcodes. In some embodiments, the methods of the invention further comprise identifying the sample from which a target polynucleotide is derived based on the barcode sequence to which the target polynucleotide is joined. In general, a barcode comprises a nucleic acid sequence that when joined to a target polynucleotide serves as an identifier of the sample from which the target polynucleotide was derived.

[0061] Recently, many improvements have been made in adaptor design that have reduced the occurrence of adapter dimer. These improvements include the use of nucleotide analogs and structured oligonucleotides, and have allowed for use of higher concentrations of oligonucleotides in ligation reactions. The higher concentrations of adapters in ligation reactions have enabled researchers to produce high quality libraries from as few as 150 copies of genome. Ligation of adaptors to the ends of DNA fragments, in particular those fragments containing the regions of interest is suitable for carrying out the methods of the invention. Various ligation modalities are envisioned, dependent on the choice of nucleic acid modifying enzymes and the resulting double-stranded DNA cleavage. For example, when a blunt end product comprising the target region/sequence of interest is generated, blunt end ligation can be suitable. Alternatively, where the cleavage is carried out using a restriction enzyme of known sequence specificity, leading to the generation of cleavage sites with known sequence overhangs, suitable ends of the adaptors can be designed to enable hybridization of the adaptor to the cleavage site of the sequence region of interest and subsequent ligation. Reagents and methods for efficient and rapid ligation of adaptors are commercially available and are known in the art.

Nucleic acid modifying enzymes

[0062] The nucleic acid (NA) -modifying enzyme can be DNA-specific modifying enzyme. The NA-modifying enzyme can be selected for specificity for double-stranded DNA. The enzyme can be a duplex-specific endonuclease, a blunt-end frequent cutter restriction enzyme, or other restriction enzyme. Examples of blunt-end cutters include DraI or SmaI. The NA-modifying enzyme can be an enzyme provided by New England Biolabs . The NA-modifying enzyme can be a homing endonuclease (a homing endonuclease can be an endonuclease that does not have a stringently-defined recognition sequence). The NA-modifying enzyme can be a nicking endonuclease (a nicking endonuclease can be an endonuclease that can cleave only one strand of DNA in a double-stranded DNA substrate). The NA-modifying enzyme can be a high fidelity endonuclease (a high fidelity endonuclease can be an engineered endonuclease that has less "star activity" than the wild-type version of the endonuclease).

[0063] In a preferred embodiment, the NA-modifying enzyme is a sequence and duplex-specific, DNA modifying enzyme.

DNA-Dependent DNA Polymerases

[0064] DNA-dependent DNA polymerases for use in the methods and compositions of the invention are capable of effecting extension of a primer or oligonucleotide according to the methods of the invention. In one embodiment, a preferred DNA-dependent DNA polymerase can be one that is capable of extending a nucleic acid primer in the presence of the DNA and/or cDNA template. Exemplary DNA dependent DNA polymerases suitable for the methods of the present invention include but are not limited to Klenow polymerase, with or without 3'-exonuclease, Bst DNA polymerase, Bca polymerase, .phi.29 DNA polymerase, Vent polymerase, Deep Vent polymerase, Taq polymerase, T4 polymerase, and E. coli DNA polymerase 1, derivatives thereof, or mixture of polymerases. In some cases, the polymerase does not comprise a 5'-exonuclease activity. In other cases, the polymerase comprises 5' exonuclease activity. In some cases, the primer or oligonucleotide extension of the present invention may be performed using a polymerase comprising strong strand displacement activity such as for example Bst polymerase. In other cases, the primer extension of the present invention may be performed using a polymerase comprising weak or no strand displacement activity. One skilled in the art may recognize the advantages and disadvantages of the use of strand

displacement activity during the primer extension step, and which polymerases may be expected to provide strand displacement activity (see e.g., New England Biolabs Polymerases).

Methods of Amplification

[0065] The methods, compositions and kits described herein can be useful to generate amplification-ready products for downstream applications such as massively parallel sequencing (i.e. next generation sequencing methods), generation of libraries with enriched population of sequence regions of interest, or hybridization platforms. Methods of amplification are well known in the art. Suitable amplification reactions can be exponential or isothermal and can include any DNA amplification reaction, including but not limited to polymerase chain reaction (PCR), strand displacement amplification (SDA), linear amplification, multiple displacement amplification (MDA), rolling circle amplification (RCA), single primer isothermal amplification (SPIA, see e.g. U.S. Pat. No. 6,251,639), Ribo-SPIA, or a combination thereof. In some cases, the amplification methods for providing the template nucleic acid may be performed under limiting conditions such that only a few rounds of amplification (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30 etc.), such as for example as is commonly done for cDNA generation. The number of rounds of amplification can be about 1-30, 1-20, 1-15, 1-10, 5-30, 10-30, 15-30, 20-30, 10-30, 15-30, 20-30, or 25-30.

[0066] PCR is an in vitro amplification procedure based on repeated cycles of denaturation, oligonucleotide primer annealing, and primer extension by thermophilic template dependent polynucleotide polymerase, resulting in the exponential increase in copies of the desired sequence of the polynucleotide analyte flanked by the primers. The two different PCR primers, which anneal to opposite strands of the DNA, are positioned so that the polymerase catalyzed extension product of one primer can serve as a template strand for the other, leading to the accumulation of a discrete double stranded fragment whose length is defined by the distance between the 5' ends of the oligonucleotide primers.

[0067] LCR uses a ligase enzyme to join pairs of preformed nucleic acid probes. The probes hybridize with each complementary strand of the nucleic acid analyte, if present,

and ligase is employed to bind each pair of probes together resulting in two templates that can serve in the next cycle to reiterate the particular nucleic acid sequence.

[0068] SDA (Westin et al 2000, Nature Biotechnology, 18, 199-202; Walker et al 1992, Nucleic Acids Research, 20, 7, 1691-1696), is an isothermal amplification technique based upon the ability of a restriction endonuclease such as HincII or BsoBI to nick the unmodified strand of a hemiphosphorothioate form of its recognition site, and the ability of an exonuclease deficient DNA polymerase such as Klenow exo minus polymerase, or Bst polymerase, to extend the 3'-end at the nick and displace the downstream DNA strand. Exponential amplification results from coupling sense and antisense reactions in which strands displaced from a sense reaction serve as targets for an antisense reaction and vice versa.

[0069] Some aspects of the invention utilize linear amplification of nucleic acids or polynucleotides. Linear amplification generally refers to a method that involves the formation of one or more copies of the complement of only one strand of a nucleic acid or polynucleotide molecule, usually a nucleic acid or polynucleotide analyte. Thus, the primary difference between linear amplification and exponential amplification is that in the latter process, the product serves as substrate for the formation of more product, whereas in the former process the starting sequence is the substrate for the formation of product but the product of the reaction, i.e. the replication of the starting template, is not a substrate for generation of products. In linear amplification the amount of product formed increases as a linear function of time as opposed to exponential amplification where the amount of product formed is an exponential function of time.

[0070] In some embodiments, the amplification is exponential, e.g. in the enzymatic amplification of specific double stranded sequences of DNA by a polymerase chain reaction (PCR). In other embodiments the amplification method is linear. In other embodiments the amplification method is isothermal.

Downstream Applications

[0071] An important aspect of the invention is that the methods and compositions disclosed herein can be efficiently and cost-effectively utilized for downstream analyses, such as next generation sequencing or hybridization platforms, with minimal loss of biological material of interest. The methods of the present invention can also be used in the analysis of genetic information of selective genomic regions of interest (e.g., analysis

of SNPs or other disease markers) as well as genomic regions which may interact with the selective region of interest.

Sequencing

[0072] For example, the methods of the invention are useful for sequencing by the method commercialized by Illumina, as described U.S. Pat. Nos. 5,750,341; 6,306,597; and 5,969,119. In general, double stranded fragment polynucleotides can be prepared by the methods of the present invention to produce amplified nucleic acid sequences tagged at one (e.g., (A)/(A′) or both ends (e.g., (A)/(A′) and (C)/(C′)). In some cases, single stranded nucleic acid tagged at one or both ends is amplified by the methods of the present invention (e.g., by SPIA or linear PCR).The resulting nucleic acid is then denatured and the single-stranded amplified polynucleotides are randomly attached to the inside surface of flow-cell channels. Unlabeled nucleotides are added to initiate solid-phase bridge amplification to produce dense clusters of double-stranded DNA. To initiate the first base sequencing cycle, four labeled reversible terminators, primers, and DNA polymerase are added. After laser excitation, fluorescence from each cluster on the flow cell is imaged. The identity of the first base for each cluster is then recorded. Cycles of sequencing are performed to determine the fragment sequence one base at a time.

[0073] In some embodiments, the methods of the invention are useful for preparing target polynucleotides for sequencing by the sequencing by ligation methods commercialized by Applied Biosystems (e.g., SOLiD sequencing). In other embodiments, the methods are useful for preparing target polynucleotides for sequencing by synthesis using the methods commercialized by 454/Roche Life Sciences, including but not limited to the methods and apparatus described in Margulies et al., *Nature*(2005) 437:376-380 (2005); and U.S. Pat. Nos. 7,244,559; 7,335,762; 7,211,390; 7,244,567; 7,264,929; and 7,323,305. In other embodiments, the methods are useful for preparing target polynucleotide(s) for sequencing by the methods commercialized by Helicos BioSciences Corporation (Cambridge, Mass.) as described in U.S. application Ser. No. 11/167,046, and U.S. Pat. Nos. 7,501,245; 7,491,498; 7,276,720; and in U.S. Patent Application Publication Nos. US20090061439; US20080087826; US20060286566; US20060024711; US20060024678; US20080213770; and US20080103058. In other embodiments, the methods are useful for preparing target polynucleotide(s) for sequencing by the methods commercialized by Pacific Biosciences as described in U.S. Pat. Nos. 7,462,452;

7,476,504; 7,405,281; 7,170,050; 7,462,468; 7,476,503; 7,315,019; 7,302,146; 7,313,308; and US Application Publication Nos. US20090029385; US20090068655; US20090024331; and US20080206764.

[0074] Another example of a sequencing technique that can be used in the methods of the provided invention is nanopore sequencing (see e.g. Soni G V and Meller A. (2007) *Clin Chem* 53: 1996-2001). A nanopore can be a small hole of the order of 1 nanometer in diameter. Immersion of a nanopore in a conducting fluid and application of a potential across it can result in a slight electrical current due to conduction of ions through the nanopore. The amount of current that flows is sensitive to the size of the nanopore. As a DNA molecule passes through a nanopore, each nucleotide on the DNA molecule obstructs the nanopore to a different degree. Thus, the change in the current passing through the nanopore as the DNA molecule passes through the nanopore can represent a reading of the DNA sequence.

[0075] Another example of a sequencing technique that can be used in the methods of the provided invention is semiconductor sequencing provided by Ion Torrent (e.g., using the Ion Personal Genome Machine (PGM)). Ion Torrent technology can use a semiconductor chip with multiple layers, e.g., a layer with micro-machined wells, an ion-sensitive layer, and an ion sensor layer. Nucleic acids can be introduced into the wells, e.g., a clonal population of single nucleic can be attached to a single bead, and the bead can be introduced into a well. To initiate sequencing of the nucleic acids on the beads, one type of deoxyribonucleotide (e.g., dATP, dCTP, dGTP, or dTTP) can be introduced into the wells. When one or more nucleotides are incorporated by DNA polymerase, protons (hydrogen ions) are released in the well, which can be detected by the ion sensor. The semiconductor chip can then be washed and the process can be repeated with a different deoxyribonucleotide. A plurality of nucleic acids can be sequenced in the wells of a semiconductor chip. The semiconductor chip can comprise chemical-sensitive field effect transistor (chemFET) arrays to sequence DNA (for example, as described in U.S. Patent Application Publication No. 20090026082). Incorporation of one or more triphosphates into a new nucleic acid strand at the 3′ end of the sequencing primer can be detected by a change in current by a chemFET. An array can have multiple chemFET sensors.

Genetic Analysis

[0076] The methods of the present invention can be used in the analysis of genetic information of selective genomic regions of interest as well as genomic regions which may interact with the selective region of interest. Amplification methods as disclosed herein can be used in the devices, kits, and methods known to the art for genetic analysis, such as, but not limited to those found in U.S. Pat. Nos. 6,449,562, 6,287,766, 7,361,468, 7,414,117, 6,225,109, and 6,110,709. In some cases, amplification methods of the present invention can be used to amplify target nucleic acid of interest for DNA hybridization studies to determine the presence or absence of polymorphisms. The polymorphisms, or alleles, can be associated with diseases or conditions such as genetic disease. In other cases the polymorphisms can be associated with susceptibility to diseases or conditions, for example, polymorphisms associated with addiction, degenerative and age related conditions, cancer, and the like. In other cases, the polymorphisms can be associated with beneficial traits such as increased coronary health, or resistance to diseases such as HIV or malaria, or resistance to degenerative diseases such as osteoporosis, Alzheimer's or dementia.

Kits

[0077] Any of the compositions described herein may be comprised in a kit. In a non-limiting example, the kit, in a suitable container, comprises: an adaptor or several adaptors, one or more of oligonucleotide primers and reagents for ligation, primer extension and amplification. The kit may also comprise means for purification, such as a bead suspension.

[0078] The containers of the kits will generally include at least one vial, test tube, flask, bottle, syringe or other containers, into which a component may be placed, and preferably, suitably aliquotted. Where there is more than one component in the kit, the kit also will generally contain a second, third or other additional container into which the additional components may be separately placed. However, various combinations of components may be comprised in a container.

[0079] When the components of the kit are provided in one or more liquid solutions, the liquid solution can be an aqueous solution. However, the components of the kit may be provided as dried powder(s). When reagents and/or components are provided as a dry powder, the powder can be reconstituted by the addition of a suitable solvent.

[0080] A kit will preferably include instructions for employing, the kit components as well the use of any other reagent not included in the kit. Instructions may include variations that can be implemented.

[0081] In one aspect, the invention provides kits containing any one or more of the elements disclosed in the above methods and compositions. In some embodiments, a kit comprises a composition of the invention, in one or more containers. In some embodiments, the invention provides kits comprising adapters, primers, and/or other oligonucleotides described herein. In some embodiments, the kit further comprises one or more of: (a) a DNA ligase, (b) a DNA-dependent DNA polymerase, (c) an RNA-dependent DNA polymerase, (d) a forward adapter (e) one or more oligonucleotides comprising reverse adaptor sequence and (f) one or more buffers suitable for one or more of the elements contained in said kit. The adapters, primers, other oligonucleotides, and reagents can be, without limitation, any of those described above. Elements of the kit can further be provided, without limitation, in any of the amounts and/or combinations (such as in the same kit or same container) described above. The kits may further comprise additional agents, such as those described above, for use according to the methods of the invention. For example, the kit can comprise a first forward adaptor that is a partial duplex adaptor as described herein, a second forward adapter, and a nucleic acid modifying enzyme specific for a restriction and/or cleavage site present in the first forward adaptor. The kit elements can be provided in any suitable container, including but not limited to test tubes, vials, flasks, bottles, ampules, syringes, or the like. The agents can be provided in a form that may be directly used in the methods of the invention, or in a form that requires preparation prior to use, such as in the reconstitution of lyophilized agents. Agents may be provided in aliquots for single-use or as stocks from which multiple uses, such as in a number of reaction, may be obtained.

[0082] In one embodiment, the kit comprises a plurality of forward adaptor oligonucleotides, wherein each of said forward adaptor oligonucleotides comprises at least one of a plurality of barcode sequences, wherein each barcode sequence of the plurality of barcode sequences differs from every other barcode sequence in said plurality of barcode sequences at at least three nucleotide positions, and instructions for using the same. Forward adapters comprising different barcode sequences can be supplied individually or in combination with one or more additional forward adapters having a different barcode sequence. In some embodiments, the kit can comprises a plurality of

first and second forward adapter oligonucleotides. Second forward adapter oligonucleotides can be supplied separately from or in combination with one or more first forward adapters, and/or one or more different second adapters. Combinations of first and second forward adapters can be supplied in accordance with combinations described above. In some embodiments, the kit can comprises a plurality of oligonucleotides comprising reverse adaptor sequence. In one embodiment, the kit can comprises a plurality of oligonucleotides comprising reverse adaptor sequence, wherein each of the plurality of oligonucleotides comprising reverse adaptor sequence further comprises sequence complementary to a specific target sequence of interest present in a nucleic acid. In one embodiment, the kit can comprises a plurality of oligonucleotides comprising reverse adaptor sequence, wherein each of the plurality of oligonucleotides comprising reverse adaptor sequence further comprises random sequence. In one embodiment, the kit comprises a plurality of oligonucleotides with reverse adaptor sequence, wherein each of said oligonucleotides with reverse adaptor sequence comprises at least one of a plurality of barcode sequences, wherein each barcode sequence of the plurality of barcode sequences differs from every other barcode sequence in said plurality of barcode sequences at at least three nucleotide positions, and instructions for using the same. Oligonucleotides with reverse adaptor sequence comprising different barcode sequences can be supplied individually or in combination with one or more additional oligonucleotides with reverse adaptor sequence having a different barcode sequence.

Products based on the Methods of the Invention

[0083] Products based on the methods of the invention may be commercialized by the Applicants under the trade name Ovation ®. Ovation ® is a trademark of NuGEN Technologies, Inc.

EXAMPLES

Example 1 – Characterization of the Human Oral Microbiome by Selective Enrichment of Bacterial 16S Ribosomal DNA Sequences.

[0084] Sample nucleic acid

Microbial genomic DNA is isolated from human saliva using the OMNIgene-DISCOVER sample collection kit (DNA Genotek) according to the manufacturer's instructions. Extracted DNA is then fragmented via sonication to an average length of 400 bp and purified using Agencourt AMPure XP beads (Beckman Coulter Genomics).

[0085] Generation of control and test libraries with ligated forward adaptors

The NuGEN Ovation Ultralow Library System (NuGEN Technologies) is used to generate two next generation sequencing libraries from 100 ng of the purified sample. The first library, an unenriched control, is made as recommended by the manufacturer. A second 'test' library, the input for downstream enrichment steps, is generated using the same library construction kit modified as follows. Briefly, DNA is blunted and prepared for ligation under the standard end-repair reaction conditions described in the kit. Fragments are then ligated to the forward adaptor only. As depicted in Figure 2, ligation attaches the forward adaptor to each end of each DNA fragment, leaving a single-strand nick on the opposite strand. Adaptor fill-in will be performed, thus generating ligation products where complementary ends exist on each insert.

[0086] Ligation products of at least 100 bp in length are purified by selective binding to Agencourt AMPure XP beads and taken forward into the enrichment process.

[0087] Amplification

Ribosomal DNA fragments from the test library are selectively amplified with two distinct steps: 1) gene-specific primer extension; and 2) PCR with universal adaptor sequences. The primer extension step is performed with oligonucleotides containing a 3' gene-specific region and a 5'common region that contains a portion of the Illumina reverse adaptor sequence. Consensus 16S sequences making up the gene-specific segment are selected by comparing the ribosomal operons from 40 diverse bacterial species using the ClustalW multiple sequence alignment program (European Bioinformatics Institute). Oligonucleotides representing each of the 18 highly conserved sequence blocks identified across the 16S genomic loci are synthesized and mixed in equimolar proportions.

[0088] The pool of primer extension probes is combined with the test DNA library (above) containing the forward adaptor and the HotStarTaq PCR mastermix (QIAGEN, USA) containing buffer, dNTPs, and a thermally-activated *Taq* DNA polymerase. This solution is placed in a thermal cycler, heated to 95°C for 15 minutes to activate the polymerase and cooled to 70°C for 5 minutes to allow the 16S primers to anneal to DNA inserts and extend into the forward adaptor site. Amplification primers that bind to the forward and reverse adaptor sites are added. Selection for fragments that contain both the forward (test library) and reverse (5' common region on 16S primers) adaptor, and the respective universal priming sites, is accomplished with PCR using a 3-step temperature routine (94°C for 30 seconds, 60°C for 30 seconds, 72° C for 1 minute) for 25 cycles.

PCR products are purified using AMPure XP beads and analyzed with a 2100 Bioanalyzer (Agilent Technologies).

[0089] Sequencing and data analysis

[0090] Single end sequencing reads of 100 nt length are obtained for both the control and enriched test libraries using a MiSeq System (Illumina). Raw sequencing data is processed using Illumina base calling software and mapped to a ribosomal RNA database. Sequences that do not align to bacterial rRNA are mapped to human and bacterial full genome reference sequences. Fold enrichment is determined by calculating the number of rRNA reads as a percentage of total mapped reads in the control and test samples.

Example 2 – Characterization of changes over time to the Human Oral Microbiome by Selective Enrichment of Bacterial 16S Ribosomal DNA Sequences.

[0091] Sample nucleic acid

Microbial genomic DNA is isolated from human saliva using the OMNIgene-DISCOVER sample collection kit (DNA Genotek) according to the manufacturer's instructions at 1 hour intervals for 16 hours following use of dental rinse. Extracted DNA is then fragmented via sonication to an average length of 400 bp and purified using Agencourt AMPure XP beads (Beckman Coulter Genomics).

[0092] Generation of DNA fragments with ligated forward adapters

Components from the NuGEN Ovation Ultralow Library System (NuGEN Technologies) are used to generate 16 independent next generation sequencing libraries from 100 ng of the purified sample. Briefly, DNA is blunted and prepared for ligation under the standard end-repair reaction conditions described in the kit. Fragments are then ligated to the forward adapter only. As depicted in Figure 2, ligation attaches the forward adapter to each end of each DNA fragment, leaving a single-strand nick on the opposite strand. Adapter fill-in will be performed, thus generating ligation products where complementary ends exist on each insert.

[0093] Ligation products of at least 100 bp in length are purified by selective binding to Agencourt AMPure XP beads and taken forward into the enrichment process.

[0094] Primer extension

Libraries containing ribosomal genes are generated by introducing the reverse adapter attached to the 5' end of oligonucleotides specific to conserved regions within these genes. There are two distinct steps: 1) annealing of the gene-specific primer; and 2) extension of that primer through the action of a DNA polymerase. The resulting product

is a functional library containing the forward adapter on one end and the reverse adapter on the other end. The gene-specific primer extension step is performed with oligonucleotides containing a 3' gene-specific region and a 5' region that contains a portion of the Illumina reverse adapter sequence. Embedded in the reverse adapter sequence is a variable region of 8 bases that differentiates this adapter from the 16 other adapters used with the other samples. Thus, 16 gene-specific libraries have been generated; one from each sample. Each library has a common forward adapter. Each library also contains a common sequence on the opposite end but within that common sequence there is a unique 8 nucleotide region. Consensus 16S sequences making up the gene-specific segment are selected by comparing the ribosomal operons from 40 diverse bacterial species using the ClustalW multiple sequence alignment program (European Bioinformatics Institute). Oligonucleotides representing each of the 18 highly conserved sequence blocks identified across the 16S genomic loci are synthesized and mixed in equimolar proportions.

[0095] Individual samples with forward adapters ligated onto each strand are combined with the primer extension probes (described above) in 16 independent reactions. These are mixed with HotStarTaq PCR mastermix (QIAGEN, USA) containing buffer, dNTPs, and a thermally-activatable *Taq* DNA polymerase. This solution is placed in a thermal cycler, heated to 95°C for 15 minutes to activate the polymerase and cooled to 70°C for 5 minutes to allow the 16S primers to anneal to DNA inserts and extend into the forward adapter site.

[0096] Amplification

The 16 individual primer extension products (above) are pooled, amplification primers that are complementary to the 5' ends of the forward and reverse adapter sites but also contain portions complementary to flow cell oligonucleotide sequences are added. Selection for fragments that contain both the forward and reverse (5' common region on 16S primers) adapter, and the respective universal priming sites, is accomplished with PCR using a 3-step temperature routine (94°C for 30 seconds, 60°C for 30 seconds, 72° C for 1 minute) for 25 cycles. PCR products are purified using AMPure XP beads and analyzed with a 2100 Bioanalyzer (Agilent Technologies).

[0097] Sequencing and data analysis

[0098] Single end sequencing reads of 100 nt length are obtained for both the control and enriched test libraries using a MiSeq System (Illumina). Raw sequencing data is

processed using Illumina base calling software. Samples from the various time points are binned based on their unique 8 base code and mapped to a ribosomal RNA database. Sequences that do not align to bacterial rRNA are mapped to human and bacterial full genome reference sequences. Changes in microbial populations are assessed by comparing 16S read counts from the different organisms in the samples over time.

Example 3 – Characterization of transcriptional activity of individual cells within a population.

[0099] Sample nucleic acid

Individual cells are isolated from whole blood using a FACS cell sorter. The cells are suspended in 10 µl of Prelude Lysis solution (a component of NuGEN Technologies, One Direct system), resulting in lysis of the cell membrane while the nuclear membrane remains intact. Sixteen of the single cell suspensions are selected for expression profiling. Briefly, kit reagents are used as described by the manufacturer to generate first and second strand cDNA from the total RNA present in the lysate. Double stranded cDNA products are purified using Agencourt AMPure XP beads (Beckman Coulter Genomics).

[00100] Generation of fragments with ligated forward adapters

Components from the NuGEN Ovation Ultralow Library System (NuGEN Technologies) are used to generate next generation sequencing libraries from each of the purified sample. Briefly, DNA is blunted and prepared for ligation under the standard end-repair reaction conditions described in the kit. Fragments are then ligated to the forward adapter only. As depicted in Figure 2, ligation attaches the forward adapter to each end of each DNA fragment, leaving a single-strand nick on the opposite strand. Adapter fill-in will be performed, thus generating ligation products where complementary ends exist on each insert.

[00101] Ligation products of at least 100 bp in length are purified by selective binding to Agencourt AMPure XP beads and taken forward into library generation.

[00102] Primer extension

Libraries are generated by introducing the reverse adapter attached to the 5' end of a random hexamer. There are two distinct steps: 1) annealing of the primer; and 2) extension of that primer through the action of a DNA polymerase. The resulting product is a functional library containing the forward adapter on one end and reverse adapter on the other end. The primer extension step is performed with oligonucleotides containing a

3' random region and a 5' region that contains a portion of the Illumina reverse adapter sequence. Embedded in the reverse adapter sequence is a variable region of 8 bases that differentiates this adapter from the 16 other adapters used with the other samples. Thus, 16 libraries have been generated; one from each sample. Each library has a common forward adapter. Each library also contains a common sequence on the opposite end but within that common sequence there is a unique 8 nucleotide region.

[00103] Individual samples with forward adapters ligated onto each strand are combined with the primer extension probes (described above) in 16 independent reactions. These are mixed with HotStarTaq PCR mastermix (QIAGEN, USA) containing buffer, dNTPs, and a thermally-activatable *Taq* DNA polymerase. This solution is placed in a thermal cycler, heated to 95°C for 15 minutes to activate the polymerase and cooled to 70°C for 5 minutes to allow the primers to anneal to DNA inserts and extend into the forward adapter site.

[00104] Amplification

Amplification primers that are complementary to the 5' ends of the forward and reverse adapter sites but also contain portions complementary to flow cell oligonucleotide sequences are added to the 16 individual primer extension products (above). Selection for fragments that contain both the forward and reverse adapter, and the respective universal priming sites, is accomplished with PCR using a 3-step temperature routine (94°C for 30 seconds, 60°C for 30 seconds, 72° C for 1 minute) for 25 cycles. PCR products are purified using AMPure XP beads and analyzed with a 2100 Bioanalyzer (Agilent Technologies).

[00105] Sequencing and data analysis

Equal masses of each of the amplified libraries (above) are pooled and diluted to working concentrations according to manufacturer's recommendations. Single end sequencing reads of 100 nt length are obtained for libraries using a MiSeq System (Illumina). Raw sequencing data is processed using Illumina base calling software. Samples from the various time points are binned based on their unique 8 base code and mapped to a reference database. Based on the mapping characteristics, individual samples or a new pool of samples can be rerun on the sequencer to obtain greater read depth. Samples with poor gene coverage will be eliminated from the pool.

CLAIMS

1. A method for enriching for a nucleic acid sequence of interest in a sample comprising nucleic acids, the method comprising:

a. fragmenting nucleic acids in the sample comprising the nucleic acids, thereby generating nucleic acid fragments, wherein the nucleic acid fragments comprise the nucleic acid sequence of interest, wherein the nucleic acid fragments are single-stranded or double-stranded;

b. ligating a first adaptor sequence to 5' ends of the nucleic acid fragments, thereby generating nucleic acid fragments comprising the ligated first adaptor sequence on the 5' ends of the nucleic acid fragments;

c. denaturing the nucleic acid fragments, thereby generating single-stranded nucleic acid fragments with the first adaptor sequence at the 5' ends, if the nucleic acid fragments comprising the ligated first adaptor sequence are double-stranded;

d. annealing one or more oligonucleotides to the nucleic acid sequence of interest in the nucleic acid fragments comprising the ligated first adaptor sequence, wherein the one or more oligonucleotides comprise a 3' portion with at least 10 bases that are complementary to the nucleic acid sequence of interest and a 5' tail portion comprising a second adaptor sequence that is non-complementary to the nucleic acid sequence of interest;

e. extending the one or more oligonucleotides annealed to the nucleic acid sequence of interest in the nucleic acid fragments comprising the ligated first adaptor sequence with a polymerase, thereby generating one or more oligonucleotide extension products comprising sequence complementary to the ligated first adaptor sequence at a first end, sequence complementary to the nucleic acid sequence of interest, and the second adaptor sequence at a second end; and

f. amplifying the one or more oligonucleotide extension products using a first primer that anneals to the complement of the first adaptor sequence and a second primer that anneals to the complement of the second adaptor sequence to enrich for the nucleic acid sequence of interest; and

g. sequencing the enriched nucleic acid sequence of interest by massively parallel sequencing.

2. The method of claim 1, further comprising adhering the amplified extension products generated in step f to a flow cell before step g.

3. The method of claim 1, wherein the nucleic acid sequence of interest comprises genomic DNA, RNA, or cDNA.

4. The method of claim 1, wherein the first adaptor sequence and the second adaptor sequence are distinct from each other.

5. The method of claim 1, wherein the first adaptor sequence and/or the second adaptor sequence comprise barcode sequence.

6. The method of claim 1, wherein the one or more oligonucleotides comprise a 3' portion that is complementary to a different nucleic acid sequence of interest in the nucleic acid fragments.

7. The method of claim 1, wherein the nucleic acid fragments are double stranded.

8. The method of claim 1, wherein the 3' portion of the one or more oligonucleotides comprises a random sequence.

9. The method of claim 1, wherein the first adaptor sequence is a partial duplex adaptor.

10. The method of claim 9, wherein the partial duplex adaptor comprises a long strand and wherein the partial duplex adaptor comprises a second strand, wherein the second strand is a short strand.

11. The method of claim 1, wherein the first adaptor sequence is in an adaptor, wherein the adaptor comprises a ligation strand and a non-ligation strand, and wherein the ligation strand ligates to the 5' ends of the nucleic acid fragments and the non-ligation strand does not ligate to a 3' end of the nucleic acid fragments.

12. The method of claim 1, wherein the 3' portion of each of the one or more oligonucleotides is designed to hybridize to the nucleic acid sequence of interest.

13. The method of claim 1, wherein the nucleic acid fragments are single-stranded.

14. The method of claim 1, wherein the sequencing comprises use of four labelled reversible terminators.

15. The method of claim 1, wherein the sequencing comprises semiconductor sequencing.