COMPOSITIONS AND METHODS FOR DIRECTIONAL NUCLEIC ACID AMPLIFICATION AND SEQUENCING

CROSS-REFERENCE

[0001] This application claims the benefit of U.S. Provisional Application No. 61/549,162, filed Oct. 19, 2011, which application is incorporated herein by reference.

BACKGROUND OF THE INVENTION

[0002] Recent advances in high-throughput, next generation sequencing technologies have enabled whole genome sequencing and new approaches to functional genomics, including comprehensive analysis of any transcriptome. One of these next generation sequencing methods involves direct sequencing of complementary DNA (cDNA) generated from messenger and structural RNAs (RNA-Seq). RNA-Seq provides several key advantages over traditional sequencing methods. First, it allows for high resolution study of all expressed transcripts, annotating the 5' and 3' ends and splice junctions of each transcript. Second, RNA-Seq allows for quantification of the relative number of transcripts in each cell. Third, RNA-Seq provides a way to measure and characterize RNA splicing by measuring the levels of each splice variant. Together, these advancements have provided new insights into individual cell function.

[0003] One drawback of performing standard RNA-Seq is the lack of information on the direction of transcription. Standard cDNA libraries constructed for RNA-Seq consist of randomly primed double-stranded cDNA. Non-directional ligation of adaptors containing universal priming sites prior to sequencing leads to a loss of information as to which strand was present in the original RNA template. Although strand information can be inferred in some cases by subsequent analysis, for example, by using open reading frame (ORF) information in transcripts that encode for a protein, or by assessing splice site information in eukaryotic genomes, direct information on the originating strand is highly desirable. For example, direct information on which strand was present in the original RNA sample is needed to assign the sense strand to a non-coding RNA, and when resolving overlapping transcripts.

[0004] Several methods have recently been developed for strand-specific RNA-Seq. These methods can be divided into two main classes. The first class utilizes distinct adaptors in a known orientation relative to the 5' and 3' end of the RNA transcript. The end result is a cDNA library where the 5' and 3' end of the original RNA are flanked by two distinct adaptors. A disadvantage of this method is that only the ends of the cloned molecules preserve directional information. This can be problematic for strand-specific manipulations of long clones, and can lead to loss of directional information when there is fragmentation.

[0005] The second class of strand-specific RNA-Seq methods marks one strand of either the original RNA (for example, by bisulfite treatment) or the transcribed cDNA (for example, by

incorporation of modified nucleotides), followed by degradation of the unmarked strand. Strand marking by bisulfite treatment of RNA is labor intensive and requires alignment of the sequencing reads to reference genomes that have all the cytosine bases converted to thymines on one of the two strands. The analysis is further complicated due to the fact that base conversion efficiency during bisulfite treatment is imperfect, i.e. less than 100%.

[0006] Strand marking by modification of the second strand of cDNA has become the preferred approach for directional cDNA cloning and sequencing (Levin *et al.,* 2010). However, cDNA second strand marking approaches, such as the one described in WO 2011/003630, are not sufficient to preserve directionality information when using conventional blunt-end ligation and cDNA library construction strategies with duplex adaptors, where two universal sequencing sites are introduced by two separate adaptors. The marking approach described in WO 2011/000360 utilizes a four-step process, consisting of 1) incorporation of a cleavable nucleotide into one strand of the cDNA insert, 2) end repair of the cDNA insert, 3) non-directional ligation of adaptors containing universal sequencing sites and 4) selective hydrolysis of library fragments with undesired adaptor orientation. To preserve directionality information, the method requires that the 5' and 3' ends of the strand selected for amplification are marked differentially, which can be achieved, for example, by ligation of directional (i.e. polarity-specific) adaptors, or by use of a specialized forked adaptor where each strand of a double-stranded polynucleotide is covalently attached to two distinct universal sequencing sites, one sequencing site at each end of the strand. Application of the methodology described in WO 2011/000360 does not result in directional sequencing libraries when using conventional duplex adaptors because the marked strand, i.e. the strand with incorporated cleavable nucleotides, is not differentially labeled at its 5' and 3' ends.

[0007] There is a need for improved methods for directional cDNA sequencing from cDNA libraries constructed with conventional duplex adaptors. The invention described herein fulfills this need.

SUMMARY OF THE INVENTION

[0008] The present invention provides novel methods, compositions, and kits for construction of directional cDNA libraries and directional cDNA sequencing. Specifically, an important aspect of this invention is the methods and compositions that allow for directional cDNA cloning and strand retention using duplex adaptors and blunt-end ligation, thereby generating ligation products with two adaptor orientations. In one aspect, the invention provides a method for cloning cDNA while retaining the directionality and strand information of the original RNA sample. In some embodiments, the method comprises: a) reverse transcribing a RNA sample to generate a first strand cDNA; b) generating a second strand cDNA from the first strand cDNA, wherein at least one of the four dNTPs dATP, dCTP, dGTP or dTTP is replaced by a modified dNTP during second strand synthesis and incorporated into the second strand, thereby generating a double-stranded cDNA; c) performing end repair on the double-stranded cDNA; d) ligating adaptors to the double-stranded cDNA, wherein only

one of the adaptors has the modified dNTP incorporated into a ligation strand of the adaptor; e) performing gap repair; and f) selectively cleaving the second strand and the ligation strand of the adaptor that has the modified dNTP by a suitable cleavage agent, thereby generating a directional cDNA library comprising the first strand cDNA In afurther aspect, the method optionally comprises fragmenting the double stranded cDNA prior to performing end repair on the double-stranded cDNA. In a specific embodiment, the method further comprises amplifying the remaining cDNA strand or the cDNA strand that does not comprise the modified nucleotide, thereby generating amplified products. In another specific embodiment, the method further comprises sequencing the remaining cDNA strand or the amplified products.

[0009] In another aspect, the invention provides for a method for selective removal of cDNA constructs in the undesired orientation.

[0010] In yet another aspect, the invention provides a method for whole transcriptome directional sequencing, comprising: a) reverse transcribing a RNA sample to generate a first strand cDNA; b) generating a second strand cDNA from the first strand cDNA, wherein at least one of the four dNTPs dATP, dCTP, dGTP or dTTP is replaced by a modified dNTP during second strand synthesis and incorporated into the second strand, thereby generating a double-stranded cDNA; c) performing end repair on the double-stranded cDNA; d) ligating adaptors to the double-stranded cDNA, wherein only one of the adaptors has the modified dNTP incorporated into a ligation strand of the adaptor; e) performing gap repair; f) selectively cleaving the second strand and the ligation strand of the adaptor that has the modified dNTP by a suitable cleavage agent, thereby generating a directional cDNA library comprising the first strand cDNA and h) amplification and/or sequencing of the directional cDNA library. In a further aspect, the method optionally comprises fragmenting the double stranded cDNA prior to performing end repair on the double-stranded cDNA.

[0011] In one aspect of any one of the foregoing aspects, the present invention provides for cleaving a base portion of the modified nucleotide thereby forming an abasic site. In a preferred embodiment, the modified nucleotide comprises dUTP. In some cases the cleavage agent comprises a glycosylase. In a preferred embodiment, the glycosylase comprises UNG. In some cases, the cleavage agent comprises a glycosylase and an endonuclease. In some cases, the endonuclease comprises an apurinic/apyrimidinic endonuclease (APE). In some cases, the cleavage agent comprises a glycosylase and a APE. In some cases, the cleavage agent comprises a UNG and a APE. In some cases, the cleavage agent comprises a glycosylase and a polyamine. In a preferred embodiment, the polyamine is N,N-dimethylethylenediamine (DMED). In some cases, the cleavage agent comprises a glycosylase and DMED. In some cases, the cleavage agent comprises a UNG and DMED.

[0012] In another aspect of any one of the foregoing aspects, the method further comprises creating nicks in a phosphodiester backbone at an abasic site with an enzyme, chemical agent, and/or heat following removal of a base portion of the modified nucleotide. In some cases, cleaving a phosphodiester backbone at an abasic site following removal of a base portion of the modified

nucleotide comprises using an enzyme. In a preferred embodiment, the enzyme is an endonuclease. In some cases, the endonuclease comprises an apurinic/apyrimidinic endonuclease (APE). In some cases, creating nicks at an abasic site following removal of a base portion of the modified nucleotide comprises using a chemical agent. In some cases, the chemical agent comprises a primary amine or a polyamine. In a preferred embodiment, the polyamine is N,N-dimethylethylenediamine (DMED).

[0013] In another aspect of any of the foregoing aspects, the method further comprises cleaving the RNA sample following reverse transcription of the RNA sample. In some cases, cleaving the RNA sample comprises exposing the RNA sample to an RNase. In a preferred embodiment, the RNase is RNase H. In some cases, cleaving the RNA sample comprises exposing the RNA sample to heat or chemical treatment or a combination thereof.

[0014] In another aspect of any of the foregoing aspects, the method further comprises reducing or depleting non-desired nucleic acid sequences. In some cases, the non-desired nucleic acid is ribosomal RNA (rRNA).

[0015] In another aspect of any of the foregoing aspects, the amplification of the remaining cDNA strand comprises polymerase chain reaction (PCR), strand displacement amplification (SDA), multiple displacement amplification (MDA), rolling circle amplification (RCA), single primer isothermal amplification (SPIA), or ligase chain reaction (LCR). In some cases, the amplification comprises PCR. In some case, the amplification comprises SPIA.

[0016] In another aspect of any of the foregoing aspects, the sequencing of the remaining cDNA strand or amplified products from the remaining cDNA strand comprises next generation sequencing.

[0017] Kits for performing any of the methods described herein are another feature of the invention. Such kits may include reagents, enzymes and platforms for amplification, cloning and sequencing of nucleic acids. In one embodiment, a kit is provided comprising: a) one or more primers; b) a reverse transcription enzyme, c) a glycosylase and d) an adaptor or several adaptors wherein one of the adaptors comprises at least one modified nucleotide in a ligation strand of said adaptor. In another embodiment, the kit further comprises reagents for amplification. In another embodiment, the kit further comprises a polyamine, an APE, or a combination thereof. In another embodiment, the kit further comprises at least one modified nucleotide or dNTP. In some cases, the modified nucleotide comprises dUTP. In some cases, the glycosylase comprises UNG. In yet another embodiment, the kit further comprises reagents for sequencing. A kit will preferably include instructions for employing the kit components as well as the use of any other reagent not included in the kit.

INCORPORATION BY REFERENCE

[0018] All publications, patents, and patent applications mentioned in this specification are herein incorporated by reference to the same extent as if each individual publication, patent, or patent application was specifically and individually indicated to be incorporated by reference.

BRIEF DESCRIPTION OF THE DRAWINGS

[0019] The novel features of the invention are set forth with particularity in the appended claims. A better understanding of the features and advantages of the present invention will be obtained by reference to the following description that sets forth illustrative embodiments, in which the principles of the invention are utilized, and the accompanying drawings of which:

[0020] Figure 1 depicts construction of a non-directional cDNA library with conventional duplex adaptors (i.e. where ligation products have two adaptor orientations) and strand marking.

[0021] Figure 2 depicts construction of directional (strand-specific) cDNA libraries with conventional duplex adaptors using the methods of the invention.

[0022] Figure 3 depicts a table summarizing strand retention efficiency data using the methods of the invention.

[0023] Figure 4 depicts a flow diagram illustrating the steps for generating a directional cDNA library using the methods of the invention.

DETAILED DESCRIPTION OF THE INVENTION

General

[0024] Reference will now be made in detail to exemplary embodiments of the invention. While the disclosed methods and compositions will be described in conjunction with the exemplary embodiments, it will be understood that these exemplary embodiments are not intended to limit the invention. On the contrary, the invention is intended to encompass alternatives, modifications and equivalents, which may be included in the spirit and scope of the invention.

[0025] In one embodiment, the present invention provides methods and compositions for construction of directional cDNA libraries. The methods described herein enable directional cDNA cloning and strand retention using conventional duplex adaptors and blunt-end ligation. The methods further enable generation of strand-specific cDNA which can be further amplified using a variety of amplification methods. In another embodiment, the present invention provides methods for whole transcriptome directional sequencing. In yet another embodiment, the present invention provides methods and compositions for generation of a directional, rRNA-depleted cDNA library.

[0026] One aspect involves a method for generation of a directional cDNA library. The first step in the method entails use of an RNA sample or RNA template from which a first strand cDNA can be generated through reverse transcription. The RNA sample can be derived from any number of sources known in the art including, but not limited to, messenger RNA (mRNA) or ribosomal RNA (rRNA) in either purified or unpurified forms and reverse transcription can be performed using any number of RNA dependent DNA polymerases known in the art. In one embodiment, the RNA template can be derived from DNA including, but not limited to, genomic DNA wherein the DNA is converted to RNA using methods known in the art including, but not limited to, transcription. In a preferred embodiment, as exemplified in FIG. 4, the RNA can be poly A+ RNA. Reverse transcription of the RNA sample can be performed using primers comprising sequence complementary to known

sequences or comprising random sequences. In one embodiment, the primers used in the methods described herein can be composite primers comprising both DNA and RNA. In a preferred embodiment, the RNA sample can be reverse transcribed using random hexamer primers.

[0027] The second step in the method described herein for the generation of a directional cDNA library entails generating a second strand cDNA from the first strand cDNA in order to form a double stranded cDNA. Second strand synthesis can be performed in the presence of a modified dNTP. In a preferred embodiment, second strand synthesis can be performed in the presence of dATP, dCTP, dGTP and dUTP in place of dTTP. Second strand synthesis in the presence of dUTP causes incorporation of at least one dUTP in the strand of the second strand cDNA. The dUTP in the second strand cDNA serves to mark the second strand since in this context dUTP is a modified or non-canonical dNTP. Second strand synthesis can be performed using any number of second strand synthesis protocols known in the art including, but not limited to, those that utilize RNase H mediated nick translation in combination with a DNA dependent DNA polymerase such as DNA polymerase I (not Klenow Fragment). Second strand synthesis can also be performed using commercially available kits such as New England Biolabs NEBNext Second Strand Synthesis Module The second strand cDNA product produced during second strand synthesis can also be referred to as the sense strand product since the sequence of the second strand cDNA comprises the sequence found in the template RNA, while the first strand cDNA can be also be referred to as the antisense strand product. In another embodiment, second strand synthesis can be performed following removal of the RNA template after first strand synthesis. Removal of the RNA template can be achieved using enzymes, heat denaturation, or chemical denaturation. Enzymatic mediated removal of the RNA template can be performed with an RNase, preferably RNase H, or a combination of enzymes, such as RNase H and RNase1 As a further aspect to this embodiment, second strand synthesis can be performed using a primer comprising sequence that is complementary to sequence present in the first strand product in conjunction with the use of a DNA dependent DNA polymerase.

[0028] In one embodiment, second strand synthesis can be followed by end repair of the double stranded cDNA generated following second strand synthesis. End repair can include the generation of blunt ends, non-blunt ends (i.e sticky or cohesive ends), or single base overhangs such as the addition of a single dA nucleotide to the 3′-end of the double-stranded DNA product, by a polymerase lacking 3′-exonuclease activity. In a preferred embodiment, end repair can be performed on the double stranded cDNA to produce blunt ends wherein the double stranded cDNA contains 5’ phosphates and 3’ hydroxyls. End repair can be performed using any number of enzymes and/or methods known in the art including, but not limited to, commercially available kits such as the Encore ^{TM} Ultra Low Input NGS Library System I.

[0029] In one embodiment, end repair can be performed after the double-stranded cDNA has been fragmented. Fragmentation of the double-stranded products can be achieved through methods known in the art. Fragmentation can be through physical fragmentation methods and/or enzymatic

fragmentation methods. Physical fragmentation methods can include nebulization, sonication, and/or hydrodynamic shearing. In a preferred embodiment, the fragmentation of the double-stranded cDNA is performed by sonication. Reagents for carrying out enzymatic fragmentation reactions are commercially available (e.g, from New England Biolabs).

[0030] Following end repair of the double stranded cDNA, the methods described herein for generating directional cDNA libraries involve ligating adaptors to the double-stranded cDNA. The adaptors can be any type of adaptor known in the art including, but not limited to, conventional duplex or double stranded adaptors. In a preferred embodiment, the adaptors can be double stranded DNA adaptors. In an embodiment, the adaptors can be oligonucleotides of known sequence and, thus, allow generation and/or use of sequence specific primers for amplification and/or sequencing of any polynucleotides to which the adaptor(s) is appended or attached. Preferably, the adaptors can be any adaptors that can be marked and selected for by methods known in the art. In a preferred embodiment, the adaptors can be marked via incorporation of at least one modified dNTP. In a preferred embodiment, one and only one of the adaptors comprises a modified dNTP while the other or any other adaptor(s) does not comprise the modified dNTP. In a preferred embodiment one and only one of the adaptors comprises a modified dNTP in a ligation strand of said adaptor, while the other or any other adaptor(s) does not comprise the modified dNTP in a ligation strand of said adaptor(s). In one embodiment, the modified dNTP is dUTP. In a preferred embodiment, the adaptors can be appended to the double-stranded product in multiple orientations. In a preferred embodiment, the methods described herein can involve the use of two conventional duplex adaptors comprising double stranded DNA of known sequence that are blunt ended and can bind to the double stranded product in one of two orientations, wherein one of the adaptors comprises a modified dNTP incorporated into the ligation strand while the other adaptor does not contain the modified dNTP in the ligation strand. In a preferred embodiment, the modified dNTP is dUTP.

[0031] According to the methods described herein, the adaptors can be ligated to the double-stranded cDNA by blunt end ligation in either of two orientations. One of the adaptors comprises a modified dNTP (preferably dUTP) incorporated into the ligation strand while the other adaptor does not comprise the modified dNTP (preferably dUTP) incorporated into the ligation strand. In one embodiment, the ligation of the adaptors to the double stranded cDNA creates a gap between the non-ligation strand of either of the adaptors and a strand of the double-stranded cDNA, whereby the non-ligation strand of the respective adaptor is not bound to a strand of the double stranded cDNA. As such, a gap repair or fill-in reaction can be performed using any number of methods known in the art including, but not limited to, the use of a DNA dependent DNA polymerase with weak or no strand displacement activity.

[0032] Following ligation of the adaptors and, optionally, gap repair, a double stranded cDNA/adaptor complex is generated. The complex can then be subjected to strand selection. Strand selection can entail base excision of the modified dNTP that is incorporated into the second strand of

the double stranded cDNA and the ligation strand of one and only one of the adaptors ligated to the double stranded cDNA. Base excision of the modified dNTP can be performed using an enzyme, chemical agent, and/or heat and creates an abasic site wherever the modified dNTP is incorporated in a nucleotide sequence. In addition to base excision, the methods of the present invention can also entail cleavage of the phoshodiester bond at the abasic site. The phosphodiester bond can also be referred to as the phosphodiester backbone or DNA backbone. Cleavage of the DNA backbone can be performed using any number of agents including anenzyme, chemical agent, heat, or any combination thereof.

[0033] Base excision and/or cleavage of the DNA backbone leads to the cleavage or removal of the marked second strand that comprises a modified dNTP as well as the ligation strand of the one adaptor that comprises the modified dNTP, while the unmarked strand and the ligation strand of the adaptor that does not contain the modified dNTP remain intact. In this instance, the remaining strand of the double strand cDNA following base excision and/or DNA backbone cleavage is the first strand cDNA or the antisense strand product. In addition, base excision and/or DNA backbone cleavage also cleaves or removes the ligation strand of the one adaptor that comprises the modified dNTP, regardless of which strand of the double stranded cDNA said ligation strand is ligated to. Amplification of the remaining strand can be preferentially performed using a primer whose sequence is complementary to the ligation strand of the adaptor that does not comprise the modified dNTP. Amplification using a primer whose sequence is complementary to the ligation strand of the adaptor that comprises the modified dNTP produces no product since that ligation strand has been cleaved and/or removed following base excision and/or cleavage of the modified dNTP. Amplification of the remaining strand can be performed using any number of amplification techniques known in the art including, but not limited to, polymerase chain reaction (PCR). Following amplification of the remaining strand, sequencing of the amplified products can be performed using primers complementary to sequences present in the ligation strand of the adaptor that does not comprise the modified dNTP, which ensures sequencing of only the unmarked first strand cDNA or the antisense strand product that remained after strand selection. Sequencing can be performed on the remaining cDNA directly and/or on the products resulting from amplification of the remaining strand. Sequencing can be performed using any of sequencing methods known in the art including, but not limited to, next generation sequencing methods. The methods of the present invention as described above result in the generation of directional cDNA libraries that comprises cDNAs of the antisense orientation or first strand cDNA due to the marking and cleavage or removal of the sense strand product (second strand cDNA).

[0034] In another embodiment, the cleaved or degraded marked strands can be removed prior to amplification and/or sequencing of the remaining strand. In a preferred embodiment, the cleaved second strand and cleaved ligation strand of the one adaptor that comprises the modified dNTP can be removed prior to amplification and/or sequencing of the remaining first strand cDNA or antisense

product. In yet another preferred embodiment, the remaining first strand cDNA (antisense product) can be purified prior to amplification and/or sequencing. Purification of the remaining strand can be performed using methods known in the art for purification of cDNA such as kits commercially available from Qiagen and/or Roche.

[0035] In an alternative embodiment, the methods described herein can be used to generate a directional cDNA library that comprises cDNAs in the sense orientation or second strand cDNA. In this embodiment, the methods described herein can be performed as described above with the exceptions that the first strand synthesis from the RNA template, as opposed to the second strand synthesis as described above, can be performed in the presence of a modified dNTP, and second strand synthesis is performed in the presence of unmodified dNTPs or classic dNTPs.In one aspect of this embodiment, the dNTPs including any and all modified dNTPs used during first strand synthesis can be removed, washed away, or replaced with unmodified dNTPs prior to second strand synthesis. As a further aspect of this embodiment, unmodified dNTPs can be used during second strand synthesis. In a preferred embodiment, the modified dNTP comprises dUTP. The antisense strand product (first strand cDNA) marked with a modified dNTP and ligated to the one and only one adaptor that comprises the modified dNTP in the ligation strand of said adaptor can be selectively cleaved or removed. As such, the remaining strand and ligated adaptor available for downstream amplification and/or sequencing comprises the sense strand product ligated to the adaptor that does not comprise the modified dNTP in the ligation strand of said adaptor.

[0036] A schematic of a preferred embodiment of the methods described herein for generating and sequencing a directional strand specific cDNA.library is illustrated in FIG. 2. Overall, the method illustrated in FIG. 2 allows determination of the strand orientation of a template RNA used to generate cDNA with improved efficiency over conventional methods as illustrated in FIG 1. The methods illustrated in FIGs. 1 and 2 both use strand marking of the cDNA and blunt end ligation of conventional duplex adaptors to the cDNA generated from template RNA as means for determining strand orientation. In both FIGs. 1 and 2, the method involves blunt end ligating double-stranded duplex adaptors (P1/P2 in FIGs. 1 and 2) to a double stranded cDNA complex formed from an RNA sample wherein the second strand product, which is also referred to as the sense strand product since it is complementary to and of the same strand orientation as the RNA template, is marked via incorporation of dUTP during second strand synthesis. In a preferred embodiment as illustrated in both FIGs. 1 and 2, the duplex adaptors do not contain free 5' phosphate groups. As such, both adaptors (P1/P2) contain a strand (the ligation strand) that ligates with the free 5'phosphate on the double-stranded cDNA and a strand that does not ligate (non-ligation strand) to the double-stranded cDNA.. Ligation can be facillitated through the use of enzymes (i.e. T4 DNA ligase) and methods known in the art, including, but not limited to, commercially available kits such as the Encore ^{TM} Ultra Low Input NGS Library System. In a preferred embodiment of the present invention as depicted in FIG. 2, the ligation strand of one and only one of the adaptors (P2) is marked via incorporation of

dUTP. As depicted in FIGs. 1 and 2, ligation of the duplex adaptors can occur in one of two orientations. In the schematic on the left side of FIGs. 1 and 2, the ligation strand of the P2 adaptor is ligated to the marked sense strand (second strand product). In the schematic on the right side of FIG. 2, the ligation strand of the P2 adaptor is ligated to the unmarked antisense strand (first strand product).

[0037] In the methods illustrated in both FIGs. 1 and 2, the duplex adaptors are unphosphorylated and thus do not contain free 5' phosphate groups. As such, both adaptors (P1/P2) contain a strand (the ligation strand) that will ligate with the free 5'phosphate on the double-stranded cDNA and a strand that does not ligate (non-ligation strand) to the double-stranded cDNA and thus leaves a gap. As such, in either orientation, the double-stranded cDNA containing the ligated adaptors is subjected to gap or fill-in repair (preferably with a DNA dependent DNA polymerase such as *Taq* DNA polymerase) in order to fill-in the gap through DNA dependent DNA polymerase mediated synthesis of the sequence of the non-ligation strand of the duplex adaptors using the respective ligation strand as template.

[0038] In both FIGs. 1 and 2, gap repair is followed by base excision via treatment with a cleavage agent, which can be an enzyme. In one embodiment as shown in FIGs. 1 and 2, base excision can be performed with an enzyme such as UNG. In a preferred embodiment, base excision can be followed by cleavage of the phosphodiester or DNA backbone using an enzyme, chemical agent, and/or heat at the site where the base was cleaved. In FIG. 1, base excision leads to the cleavage of the marked sense strand product, while both adaptors remain intact. In FIG. 2, base excision leads to the cleavage of both the marked sense strand product and the one adaptor that has dUTP incorporated into the ligation strand of said adaptor, while the adaptor that does not have dUTP incorporated into the ligation strand remains intact. In both FIGs. 1 and 2, the marked sense strand product or second strand product can be cleaved and thus only the antisense strand product or first strand product remains following base excision.

[0039] In contrast to FIG. 1 wherein both adaptor orientations remain intact following base excision, the schematic on the left side of FIG. 2 shows that the marked sense strand product can be cleaved along with the marked ligation strand of the P2 adaptor that is ligated to the sense strand product. As such, only the antisense strand product and the ligation strand of the adaptor ligated to the antisense strand (P1 in FIG. 2) remain intact and available for downstream processing. In a preferred embodiment downstream processing entails amplification of the remaining cDNA strand or antisense strand product. In contrast to FIG. 1 wherein amplification of the antisense strand product can be performed using primers complementary to sequence contained in either the ligation strand of the P1 adaptor (P1amp) or P2 adaptor (P2amp), FIG. 2 shows that amplification of the antisense strand product can only be performed using primers complementary to sequence contained in the ligation strand of the P1 adaptor (P1amp).. In a preferred embodiment, downstream processing can also entail sequencing of the antisense strand product (first strand product) and/or the amplified products. In FIG. 1, downstream sequencing using primers complementary to sequence in the ligation strand of the P1

adaptor will sequence either the sense or antisense strands relative to the RNA template. In FIG. 2, downstream sequencing using primers complementary to sequence in the ligation strand of the P1 adaptor will sequence only the antisense strand relative to the RNA template.

[0040] FIG. 4 illustrates a flow chart depicting one embodiment of the method for generating a directional strand specific cDNA library. The method involves the steps of generating first strand cDNA by performing random-primed reverse transcription on polyA+ RNA; generating 2nd strand cDNA using a DNA-dependent DNA polymerase using dATP, dCTP, dGTP, and dUTP in place of dTTP; fragmenting the double-stranded cDNA using sonication; end-repairing the purified fragmented double-stranded cDNA to generate blunt ends; ligating duplex adaptors wherein one of the ligation strands of one of the duplex adaptors is marked via incorporation of dUTP; nick repairing the ligation products to generate double-stranded cDNA containing the ligated adaptors; performing strand selection of the purified double-stranded cDNA containing the ligated adaptors using an enzyme and/or chemical agent; amplifying the remaining cDNA strand using PCR.

[0041] In an aspect to any of the embodiments above, the directional cDNA libraries created by the methods described herein can be depleted of non-desired nucleic acid sequences. In one embodiment, the non-desired nucleic acid comprises RNA. In a preferred embodiment, the non-desired nucleic acid comprises ribosomal or rRNA. Removal or depletion of rRNA from the directional cDNA libraries generated by the methods of the present invention can be performed by any of the methods known in the art including, but not limited to, removal of rRNA from the starting population, differential priming using oligo dT primers (i.e. priming polyadenylated transcripts only), and/or differential priming where primers complementary to rRNA sequences are specifically eliminated (or under-represented) in a primer pool (Not-So-Random or NSR primer approach).

[0042] In general, the methods described herein can be used to create nucleic acid libraries preferentially populated with nucleic acids of specific strand orientations relative to the nucleic acid template from which the library was generated. The nucleic acid libraries generated by the methods described herein can be used to ascertain the directionality and strand orientation of the nucleic acid template. In one embodiment, the nucleic acid template can be an RNA template and the nucleic acid library can be a cDNA library. In a preferred embodiment, the RNA template can be non-rRNA. In yet another preferred embodiment, the RNA template can be rRNA. In an aspect of the methods described herein, the cDNA library can be a directional cDNA library that retains the directionality and strand information pertaining to the original RNA template or sample, that is to say, the directional library of the methods of the invention represents products generated from first strand cDNA, or reverse transcription of the template RNA, or the second strand cDNA (a copy of the first strand cDNA). The methods of the invention provide means for exclusive retention of either first strand cDNA products or second strand cDNA products, thus enabling assigning the directionality of

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

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

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

RNA Sample

[0045] The RNA sample of the present invention can be double-stranded, partially double-stranded; and single-stranded nucleic acids from any source including, but not limited to, synthetic or semisynthetic nucleic acids in purified or unpurified form, which can be DNA (dsDNA and ssDNA) or RNA, including tRNA, mRNA, rRNA, mitochondrial DNA and RNA, chloroplast DNA and RNA, DNA-RNA hybrids, or mixtures thereof, genes, chromosomes, plasmids, the genomes of biological material such as microorganisms, e.g., bacteria, yeasts, viruses, viroids, molds, fungi, plants, animals,

humans, and fragments thereof. Exemplary starting material comprising DNA (including genomic DNA) can be transcribed into RNA form, which can be achieved using methods disclosed in Kurn, U.S. Pat. No. 6,251,639, and by other techniques, such as expression systems. RNA copies of genomic DNA would generally include untranscribed sequences generally not found in mRNA, such as introns, regulatory and control elements, etc. Exemplary RNA samples can be obtained and purified using standard techniques in the art and includes RNAs in purified or unpurified form, which include, but are not limited to, mRNAs, tRNAs, snRNAs, rRNAs, retroviruses, small non-coding RNAs, microRNAs, polysomal RNAs, pre-mRNAs, intronic RNA, viral RNA, cell free RNA and fragments thereof. In one embodiment, the RNA sample provided for the methods of the present invention includes a whole transcriptome which can include tRNA, mRNA, rRNA, and non-coding RNA. The non-coding RNA, or ncRNA may include snoRNAs, microRNAs, siRNAs, piRNAs and long nc RNAs. In a preferred embodiment, the RNA sample has the rRNA content reduced or removed using standard techniques in the art. In a most preferred embodiment, the RNA sample is mRNA.

Primers

[0046] The term "primer", as used herein, can refer to a nucleotide sequence, generally with a free 3' hydroxyl group, that is capable of hybridizing with a template (such as one or more target polynucleotides, one or more target DNAs, one or more target RNAs or a primer extension product) and is also capable of promoting polymerization of a polynucleotide complementary to the template. A primer can be, for example, an oligonucleotide. It can also be, for example, a sequence of the template (such as a primer extension product or a fragment of the template created following RNase [i.e. RNase H] cleavage of a template-DNA complex) that is hybridized to a sequence in the template itself (for example, as a hairpin loop), and that is capable of promoting nucleotide polymerization. Thus, a primer can be an exogenous (e.g., added) primer or an endogenous (e.g., template fragment) primer. A primer may contain a non-hybridizing sequence that constitutes a tail of the primer. A primer may still be hybridizing to a target even though its sequences are not fully complementary to the target.

[0047] The primers of the invention are generally oligonucleotides that are employed in an extension reaction by a polymerase along a polynucleotide template, such as in PCR, SPIA or cDNA synthesis, for example. The oligonucleotide primer can be a synthetic polynucleotide that is single stranded, containing a sequence at its 3′-end that is capable of hybridizing with a sequence of the target polynucleotide. Normally, the 3′ region of the primer that hybridizes with the target nucleic acid has at least 80%, preferably 90%, more preferably 95%, most preferably 100%, complementarity to a sequence or primer binding site.

[0048] "Complementary", as used herein, can refer to complementarity to all or only to a portion of a sequence. The number of nucleotides in the hybridizable sequence of a specific oligonucleotide primer should be such that stringency conditions used to hybridize the oligonucleotide primer will

prevent excessive random non-specific hybridization. Usually, the number of nucleotides in the hybridizing portion of the oligonucleotide primer will be at least as great as the defined sequence on the target polynucleotide that the oligonucleotide primer hybridizes to, namely, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 11, at least 12, at least 13, at least 14, at least 15, at least about 20, and generally from about 6 to about 10 or 6 to about 12 or 12 to about 200 nucleotides, usually about 20 to about 50 nucleotides. In general, the target polynucleotide is larger than the oligonucleotide primer or primers as described previously.

[0049] In some cases, the identity of the investigated target polynucleotide sequence is known, and hybridizable primers can be synthesized precisely according to the antisense sequence of the aforesaid target polynucleotide sequence. In other cases, when the target polynucleotide sequence is unknown, the hybridizable sequence of an oligonucleotide primer is a random sequence. Oligonucleotide primers comprising random sequences may be referred to as "random primers", as described herein. In yet other cases, an oligonucleotide primer such as a first primer or a second primer comprises a set of primers such as for example a set of first primers or a set of second primers. In some cases, the set of first or second primers may comprise a mixture of primers designed to hybridize to a plurality (e.g. 2, 3, 4, about 6, 8, 10, 20, 40, 80, 100, 125, 150, 200, 250, 300, 400, 500, 600, 800, 1000, 1500, 2000, 2500, 3000, 4000, 5000, 6000, 7000, 8000, 10,000, 20,000, 25,000 or more) of target sequences. In some cases, the plurality of target sequences may comprise a group of related sequences, random sequences, a whole transcriptome or fraction (e.g. substantial fraction) thereof, or any group of sequences such as mRNA.

[0050] Tailed primers can be employed in certain embodiments of the invention. In general, a tailed primer comprises a 3' portion that is hybridizable to one or more target polynucleotides, such as one or more target RNAs in an RNA sample, and a 5' portion that is not hybridizable to the one or more target polynucleotides. In general, the non-hybridizable 5' portion does not hybridize to the one or more target polynucleotides under conditions in which the hybridizable 3' portion of the tailed primer hybridizes to the one or more target polynucleotides. In some embodiments, the non-hybridizable 5' portion comprises a promoter-specific sequence. Generally, a promoter-specific sequence comprises a single-stranded DNA sequence region which, in double-stranded form is capable of mediating RNA transcription. Examples of promoter-specific sequences are known in the art, and include, without limitation, T7, T3, or SP6 RNA polymerase promoter sequences. When the tailed primer is extended with a DNA polymerase, a primer extension product with a 5' portion comprising a defined sequence can be created. This primer extension product can then have a second primer anneal to it, which can be extended with a DNA polymerase to create a double stranded product comprising a defined sequence at one end. In some embodiments, where the non-hybridizable 5' portion of one or more tailed primers comprises a promoter-specific sequence, creation of a double-stranded product comprising a defined sequence at one end generates a double-stranded promoter sequence that is capable of mediating RNA transcription. In some embodiments, a double-stranded promoter

sequence can be generated by hybridizing to the promoter-specific sequence an oligonucleotide comprising a sequence complementary to the promoter-specific sequence. In some embodiments, formation of a double-stranded promoter can be followed by the generation of single-stranded RNA by RNA transcription of sequence downstream of the double-stranded promoter, generally in a reaction mixture comprising all necessary components, including but not limited to ribonucleoside triphosphates (rNTPs) and a DNA-dependent RNA polymerase. Tailed primers can comprise DNA, RNA, or both DNA and RNA. In some embodiments, the tailed primer consists of DNA.

[0051] Composite primers can be employed in certain embodiments of the invention. Composite primers are primers that are composed of RNA and DNA portions. In some aspects, the composite primer can be a tailed composite primer comprising, for example, a 3'-DNA portion and a 5'-RNA portion. In the tailed composite primer, a 3'-portion, all or a portion of which comprises DNA, is complementary to a polynucleotide; and a 5'-portion, all or a portion of which comprises RNA, is not complementary to the polynucleotide and does not hybridize to the polynucleotide under conditions in which the 3'-portion of the tailed composite primer hybridizes to the polynucleotide target. When the tailed composite primer is extended with a DNA polymerase, a primer extension product with a 5'-RNA portion comprising a defined sequence can be created. This primer extension product can then have a second primer anneal to it, which can be extended with a DNA polymerase to create a double stranded product with an RNA/DNA heteroduplex comprising a defined sequence at one end. The RNA portion can be selectively cleaved from the partial heteroduplex to create a double-stranded DNA with a 3'-single-stranded overhang which can be useful for various aspects of the present invention including allowing for isothermal amplification using a composite amplification primer.

[0052] In other aspects, the composite primer can be an amplification composite primer (interchangeably called composite amplification primer). In the amplification composite primer, both the RNA and the DNA portions are generally complementary and hybridize to a sequence in the polynucleotide to be copied or amplified. In some embodiments, a 3'-portion of the amplification composite primer is DNA and a 5'-portion of the composite amplification primer is RNA. The composite amplification primer is designed such that the primer is extended from the 3'-DNA portion to create a primer extension product. The 5'-RNA portion of this primer extension product, in a RNA/DNA heteroduplex is susceptible to cleavage by RNase H, thus freeing a portion of the polynucleotide to the hybridization of an additional composite amplification primer. The extension of the amplification composite primer by a DNA polymerase with strand displacement activity releases the primer extension product from the original primer and creates another copy of the sequence of the polynucleotide. Repeated rounds of primer hybridization, primer extension with strand displacement DNA synthesis, and RNA cleavage create multiple copies of the sequence of the polynucleotide. Composite primers are described in more detail below.

[0053] A "random primer," as used herein, can be a primer that generally comprises a sequence that is designed not necessarily based on a particular or specific sequence in a sample, but rather is based

on a statistical expectation (or an empirical observation) that the sequence of the random primer is hybridizable (under a given set of conditions) to one or more sequences in the sample. A random primer will generally be an oligonucleotide or a population of oligonucleotides comprising a random sequence(s) in which the nucleotides at a given position on the oligonucleotide can be any of the four nucleotides, or any of a selected group of the four nucleotides (for example only three of the four nucleotides, or only two of the four nucleotides). In some cases all of the positions of the oligonucleotide or population of oligonucleotides can be any of two or more nucleotides. In other cases, only a portion of the oligonucleotide, for instance a particular region, will comprise positions which can be any of two or more bases. In some cases, the portion of the oligonucleotide which comprises positions which can be any of two or more bases is about 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, or about 15-20 nucleotides in length. In some cases, a random primer may comprise a tailed primer having a 3'-region that comprises a random sequence and a 5'-region that is a non-hybridizing sequence that comprises a specific, non-random sequence. The 3'-region may also comprise a random sequence in combination with a region that comprises poly-T sequences. The sequence of a random primer (or its complement) may or may not be naturally-occurring, or may or may not be present in a pool of sequences in a sample of interest. The amplification of a plurality of RNA species in a single reaction mixture can employ, but not necessarily employ, a multiplicity, preferably a large multiplicity, of random primers. As is well understood in the art, a "random primer" can also refer to a primer that is a member of a population of primers (a plurality of random primers) which collectively are designed to hybridize to a desired and/or a significant number of target sequences. A random primer may hybridize at a plurality of sites on a nucleic acid sequence. The use of random primers provides a method for generating primer extension products complementary to a target polynucleotide which does not require prior knowledge of the exact sequence of the target. In some embodiments one portion of a primer is random, and another portion of the primer comprises a defined sequence. For example, in some embodiments, a 3'-portion of the primer will comprise a random sequence, while the 5'-portion of the primer comprises a defined sequence. In some embodiments a 3'-random portion of the primer will comprise DNA, and a 5'- defined portion of the primer will comprise RNA, in other embodiments, both the 3' and 5'-portions will comprise DNA. In some embodiments, the 5'-portion will contain a defined sequence and the 3'-portion will comprise a poly-dT sequence that is hybridizable to a multiplicity of RNAs in a sample (such as all mRNA). In some embodiments, a “random primer,” or primer comprising a randomly generated sequence, comprises a collection of primers comprising one or more nucleotides selected at random from two or more different nucleotides, such that all possible sequence combinations of the nucleotides selected at random may be represented in the collection. In some embodiments, generation of one or more random primers does not include a step of excluding or selecting certain sequences or nucleotide combinations from the possible sequence combinations in the random portion of the one or more random primers.

[0054] In one embodiment, the primers of the invention can be tailed primers. In this embodiment, the 5'-tail can comprise RNA and is non hybridizable to the RNA in the sample. The 3'-end of the first primer(s) can be hybridizable to the RNA in the sample, comprise DNA and comprise a random sequence, enabling hybridization across the whole transcriptome. The first primer may also comprise a mixture of primers. The mixture of first primers may also include a first primer comprising a 3'-DNA sequence hybridizable to the 3'-poly A tail of mRNA, in addition to the first primers comprising a random sequence at the 3'-ends.

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

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

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

RNA-Dependent DNA Polymerases

[0058] RNA-dependent DNA polymerases for use in the methods and compositions of the invention are capable of effecting extension of a primer according to the methods of the invention. Accordingly, a preferred RNA-dependent DNA polymerasecan be one that is capable of extending a nucleic acid primer along a nucleic acid template that is comprised at least predominantly of ribonucleotides. Suitable RNA-dependent DNA polymerases for use in the methods and compositions of the invention include reverse transcriptases (RTs). RTs are well known in the art. Examples of RTs include, but are not limited to, Moloney murine leukemia virus (M- MLV) reverse transcriptase, human immunodeficiency virus (HIV) reverse transcriptase, rous sarcoma virus (RSV) reverse transcriptase, avian myeloblastosis virus (AMV) reverse transcriptase, rous associated virus (RAV) reverse transcriptase, and myeloblastosis associated virus (MAV) reverse transcriptase or other avian sarcoma-leukosis virus (ASLV) reverse transcriptases, and modified RTs derived

therefrom. See e.g. US7056716. Many reverse transcriptases, such as those from avian myeoloblastosis virus (AMV-RT), and Moloney murine leukemia virus (MMLV-RT) comprise more than one activity (for example, polymerase activity and ribonuclease activity) and can function in the formation of the double stranded cDNA molecules. However, in some instances, it is preferable to employ a RT which lacks or has substantially reduced RNase H activity. RTs devoid of RNase H activity are known in the art, including those comprising a mutation of the wild type reverse transcriptase where the mutation eliminates the RNase H activity. Examples of RTs having reduced RNase H activity are described in US20100203597. In these cases, the addition of an RNase H from other sources, such as that isolated from E. coli, can be employed for the degradation of the starting RNA sample and the formation of the double stranded cDNA. Combinations of RTs are also contemplated, including combinations of different non-mutant RTs, combinations of different mutant RTs, and combinations of one or more non-mutant RT with one or more mutant RT.

DNA-Dependent DNA Polymerases

[0059] DNA-dependent DNA polymerases for use in the methods and compositions of the invention are capable of effecting extension of a primer according to the methods of the invention. Accordingly, a preferred DNA-dependent DNA polymerase can be one that is capable of extending a nucleic acid primer along a first strand cDNA in the presence of the RNA template or after selective removal of the RNA template. Exemplary DNA dependent DNA polymerases suitable for the methods of the present invention include but are not limited to Klenow polymerase, with or without 3'-exonuclease, Bst DNA polymerase, Bca polymerase, .phi.29 DNA polymerase, Vent polymerase, Deep Vent polymerase, Taq polymerase, T4 polymerase, and E. coli DNA polymerase 1, derivatives thereof, or mixture of polymerases. In some cases, the polymerase does not comprise a 5'-exonuclease activity. In other cases, the polymerase comprises 5' exonuclease activity. In some cases, the primer extension of the present invention may be performed using a polymerase comprising strong strand displacement activity such as for example Bst polymerase. In other cases, the primer extension of the present invention may be performed using a polymerase comprising weak or no strand displacement activity. One skilled in the art may recognize the advantages and disadvantages of the use of strand displacement activity during the primer extension step, and which polymerases may be expected to provide strand displacement activity (see e.g., New England Biolabs Polymerases). For example, strand displacement activity may be useful in ensuring whole transcriptome coverage during the random priming and extension step. Strand displacement activity may further be useful in the generation of double stranded amplification products during the priming and extension step. Alternatively, a polymerase which comprises weak or no strand displacement activity may be useful in the generation of single stranded nucleic acid products during primer hybridization and extension that are hybridized to the template nucleic acid.

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

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

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

Methods of Strand-specific Selection

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

[0063] The term "strand specific" or "directional", as used herein, can refer to the ability to differentiate in a double-stranded polynucleotide between the original template strand and the strand that is complementary to the original template strand.

[0064] In some embodiments, the methods of the invention can be used to preserve information about the direction of single-stranded nucleic acid molecules while generating double-stranded polynucleotides more suitable for molecular cloning applications. One of the strands of the double-stranded polynucleotide can be synthesized so that it has at least one modified nucleotide incorporated into it along the entire length of the strand. In some embodiments, the incorporation of the modified nucleotide marks the strand for degradation or removal.

[0065] The term "first strand synthesis" can refer to the synthesis of the first strand using the original nucleic acid (RNA or DNA) as a starting template for the polymerase reaction. The nucleotide sequence of the first strand corresponds to the sequence of the complementary strand.

[0066] The term "second strand synthesis" can refer to the synthesis of the second strand that uses the first strand as a template for the polymerase reaction. The nucleotide sequence of the second strand corresponds to the sequence of the original nucleic acid template.

[0067] The term "unmodified dNTPs" or "classic dNTPs" can refer to the four deoxyribonucleotide triphosphates dATP (deoxyadenosine triphosphate), dCTP (deoxycytidine triphosphate), dGTP (deoxyguanosine triphosphate) and dTTP (deoxythymidine triphosphate) that are normally used as building blocks in the synthesis of DNA. Similarly, the term "canonical dNTP" or "canonical nucleotide" can be used to refer to the four deoxyribonucleotide triphosphates dATP, dCTP, dGTP and dTTP that are normally found in DNA.

[0068] The term "canonical", as used herein, can refer to the nucleic acid bases adenine, cytosine, guanine and thymine that are commonly found in DNA or their deoxyribonucleotide or deoxyribonucleoside analogs. The term "non-canonical" can refer to nucleic acid bases in DNA other than the four canonical bases in DNA, or their deoxyribonucleotide or deoxyribonucleoside analogs. Although uracil is a common nucleic acid base in RNA, uracil is a non-canonical base in DNA.

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

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

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

[0071] The selective removal or cleavage of a marked strand in the present invention can be achieved through the use of enzymatic treatment of the marked strand. Enzymes that can be used for selective removal or cleavage of the marked strand according to the methods of the present invention can include glycosylases such as Uracil-N-Glycosylase (UNG), which selectively degrades the base portion of dUTP from the DNA backbone. Additional glycosylases which can be used in the methods of the present invention and their non-canonical or modified nucleotide substrates include 5-methylcytosine DNA glycosylase (5-MCDG), which cleaves the base portion of 5-methylcytosine (5-MeC) from the DNA backbone (Wolffe et al., Proc. Nat. Acad. Sci. USA 96:5894-5896, 1999); 3-methyladenosine-DNA glycosylase I, which cleaves the base portion of 3-methyl adenosine from the DNA backbone (see, e.g. Hollis et al (2000) Mutation Res. 460: 201-210); and/or 3-methyladenosine DNA glycosylase II, which cleaves the base portion of 3-methyladenosine, 7-methylguanine, 7-methyladenosine, and/3-methylguanine from the DNA backbone. See McCarthy et al (1984) EMBO J. 3:545-550. Multifunctional and mono-functional forms of 5-MCDG have been described. See Zhu et al., Proc. Natl. Acad. Sci. USA 98:5031-6, 2001; Zhu et al., Nuc. Acid Res. 28:4157-4165, 2000; and Neddermann et al., J. B. C. 271:12767-74, 1996 (describing bifunctional 5-MCDG; Vairapandi & Duker, Oncogene 13:933-938, 1996; Vairapandi et al., J. Cell. Biochem. 79:249-260, 2000 (describing mono-functional enzyme comprising 5-MCDG activity). In some embodiments, 5-MCDG preferentially cleaves fully methylated polynucleotide sites (e.g., CpG dinucleotides), and in other embodiments, 5-MCDG preferentially cleaves a hemi-methylated polynucleotide. For example, mono-functional human 5-methylcytosine DNA glycosylase cleaves DNA specifically at fully methylated CpG sites, and is relatively inactive on hemimethylated DNA (Vairapandi & Duker, supra; Vairapandi et al., supra). By contrast, chick embryo 5-methylcytosine-DNA glycosylase has greater activity directed to hemimethylated methylation sites. In some embodiments, the activity of 5-MCDG is potentiated (increased or enhanced) with accessory factors, such as recombinant CpG-rich RNA, ATP, RNA helicase enzyme, and proliferating cell nuclear antigen (PCNA). See U.S. Patent Publication No. 20020197639 A1. One or more agents may be used. In some embodiments, the one or more agents cleave a base portion of the same methylated nucleotide. In other embodiments, the one or more agents cleave a base portion of different methylated nucleotides. Treatment with two or more agents may be sequential or simultaneous.

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

cleavage of the backbone at an abasic site include: heat treatment and/or chemical treatment (including basic conditions, acidic conditions, alkylating conditions, or amine mediated cleavage of abasic sites, (see e.g., McHugh and Knowland, Nucl. Acids Res. (1995) 23(10):1664-1670; Bioorgan. Med. Chem. (1991) 7:2351; Sugiyama, Chem. Res. Toxicol. (1994) 7: 673-83; Horn, Nucl. Acids. Res., (1988) 16:11559-71), and/ or the use of enzymes that catalyze cleavage of polynucleotides at abasic sites, For example AP endonucleases (also called "apurinic, apyrimidinic endonucleases") (e.g., E. coli Endonuclease IV, available from Epicentre Tech., Inc, Madison Wis.), E. coli endonuclease III or endonuclease IV, E. coli exonuclease III in the presence of calcium ions. See, e.g. Lindahl, PNAS (1974) 71(9):3649-3653; Jendrisak, U.S. Pat. No. 6,190,865 B1; Shida, Nucleic Acids Res. (1996) 24(22):4572-76; Srivastava, J. Biol. Chem. (1998) 273(13):21203-209; Carey, Biochem. (1999) 38:16553-60; Chem Res Toxicol (1994) 7:673-683. As used herein "agent" encompasses reaction conditions such as heat. In one embodiment, the AP endonuclease, E. coli endonuclease IV, is used to cleave the phosphodiester backbone or phosphodiester bond at an abasic site. In another embodiment, cleavage is with an amine, such as N,N'-dimethylethylenediamine (DMED). See, e.g., McHugh and Knowland, supra.

[0073] In some cases, the nucleic acid comprising one or more abasic sites may be treated with a nucleophile or a base. In some cases, the nucleophile is an amine such as a primary amine, a secondary amine, or a tertiary amine. For example, the abasic site may be treated with piperidine, moropholine, or a combination thereof. In some cases, hot piperidine (e.g., 1M at 90^0 C) may be used to cleave the nucleic acid comprising one or more abasic sites. In some cases, morpholine (e.g., 3M at 37^0 C or 65^0 C) may be used to cleave the nucleic acid comprising one or more abasic sites. Alternatively, a polyamine may be used to cleave the nucleic acid comprising one or more abasic sites. Suitable polyamines include for example spermine, spermidine, 1,4-diaminobutane, lysine, the tripeptide K--W--K, DMED, piperazine, 1,2-ethylenediamine, or any combination thereof. In some cases, the nucleic acid comprising one or more abasic sites may be treated with a reagent suitable for carrying out a beta elimination reaction, a delta elimination reaction, or a combination thereof. In some cases, the methods of the present invention provide for the use of an enzyme or combination of enzymes and a polyamine such as DMED under mild conditions in a single reaction mixture which does not affect the canonical or unmodified nucleotides and therefore may maintain the sequence integrity of the products of the method. Suitable mild conditions may include conditions at or near neutral pH. Other suitable conditions include pH of about 4.5 or higher, 5 or higher, 5.5 or higher, 6 or higher, 6.5 or higher, 7 or higher, 7.5 or higher, 8 or higher, 8.5 or higher, 9 or higher, 9.5 or higher, 10 or higher, or about 10.5 or higher. Still other suitable conditions include between about 4.5 and 10.5, between about 5 and 10.0, between about 5.5 and 9.5, between about 6 and 9, between about 6.5 and 8.5, between about 6.5 and 8.0, or between about 7 and 8.0. Suitable mild conditions also may include conditions at or near room temperature. Other suitable conditions include a temperature of about 10^0 C, 11^0 C, 12^0 C, 13^0 C, 14^0 C, 15^0 C, 16^0 C, 17^0 C, 18^0 C, 19^0 C, 20^0 C, 21^0 C, 22^0 C, 23^0 C,

24°C, 25°C, 26°C, 27°C, 28°C, 29°C, 30°C, 31°C, 32°C, 33°C, 34°C, 35°C, 36°C, 37°C, 38°C, 39 °C, 40°C, 41°C, 42°C, 43°C, 44°C, 45°C, 46°C, 47°C, 48°C, 49°C, 50°C, 51°C, 52°C, 53°C, 54° C, 55°C, 56°C, 57°C, 58°C, 59°C, 60°C, 61°C, 62°C, 63°C, 64°C, 65°C, 66°C, 67°C, 68°C, 69°C, or 70°C or higher. Still other suitable conditions include between about 10°C and about 70°C, between about 15°C and about 65°C, between about 20°C and about 60°C, between about 20°C and about 55°C, between about 20°C and about 50°C, between about 20°C and about 45°C, between about 20°C and about 40°C, between about 20°C and about 35°C, or between about 20°C and about 30°C. In some cases, the use of mild cleavage conditions may provide for less damage to the primer extension products produced by the methods of the present invention. In some cases, the fewer damaged bases, the more suitable the primer extension products may be for downstream analysis such as sequencing. In other cases, the use of mild cleavage conditions may increase final product yields, maintain sequence integrity, or render the methods of the present invention more suitable for automation.

[0074] In embodiments involving fragmentation, the backbone of the polynucleotide comprising the abasic site is cleaved at the abasic site, whereby two or more fragments of the polynucleotide are generated. At least one of the fragments comprises an abasic site, as described herein. Agents that cleave the phosphodiester backbone or phosphodiester bonds of a polynucleotide at an abasic site are provided herein. In some embodiments, the agent is an AP endonuclease such as E. coli AP endonuclease IV. In other embodiments, the agent is DMED. In other embodiments, the agent is heat, basic condition, acidic conditions, or an alkylating agent. In still other embodiments, the agent that cleaves the phosphodiester backbone at an abasic site is the same agent that cleaves the base portion of a nucleotide to form an abasic site. For example, glycosylases of the present invention may comprise both a glycosylase and a lyase activity, whereby the glycosylase activity cleaves the base portion of a nucleotide (e.g., a modified nucleotide) to form an abasic site and the lyase activity cleaves the phosphodiester backbone at the abasic site so formed. In some cases, the glycosylase comprises both a glycosylase activity and an AP endonuclease activity.

[0075] Appropriate reaction media and conditions for carrying out the cleavage of a base portion of a non-canonical or modified nucleotide according to the methods of the invention are those that permit cleavage of a base portion of a non-canonical or modified nucleotide. Such media and conditions are known to persons of skill in the art, and are described in various publications, such as Lindahl, PNAS (1974) 71(9):3649-3653; and Jendrisak, U.S. Pat. No. 6,190,865 B1; U.S. Pat. No. 5,035,996; and U.S. Pat. No. 5,418,149. In one embodiment, UDG (Epicentre Technologies, Madison Wis.) is added to a nucleic acid synthesis reaction mixture, and incubated at 37°C for 20 minutes. In one embodiment, the reaction conditions are the same for the synthesis of a polynucleotide comprising a non-canonical or modified nucleotide and the cleavage of a base portion of the non-canonical or modified nucleotide. In another embodiment, different reaction conditions are used for

these reactions. In some embodiments, a chelating regent (e.g. EDTA) is added before or concurrently with UNG in order to prevent a polymerase from extending the ends of the cleavage products.

[0076] In a one embodiment, the selection is done by incorporation of at least one modified nucleotide into one strand of a synthesized polynucleotide, and the selective removal is by treatment with an enzyme that displays a specific activity towards the at least one modified nucleotide. In a preferred embodiment, the modified nucleotide being incorporated into one strand of the synthesized polynucleotide is deoxyuridine triphosphate (dUTP), replacing dTTP in the dNTP mix, and the selective removal of the marked strand from downstream applications is carried by out by UNG. UNG selectively degrades dUTP while it is neutral towards other dNTPs and their analogs. Treatment with UNG results in the cleavage of the N-glycosylic bond and the removal of the base portion of dU residues, forming abasic sites. In a preferred embodiment, the UNG treatment is done in the presence of an apurinic/apyrimidinic endonuclease (APE) to create nicks at the abasic sites. Consequently, a polynucleotide strand with incorporated dUTP that is treated with UNG/APE is cleaved and unable to undergo amplification by a polymerase. In another embodiment, nick generation and cleavage is achieved by treatment with a polyamine, such as DMED, or by heat treatment. In a preferred embodiment, UNG treatment is conducted in a reaction buffer containing 32 mM DMED.

[0077] As used in this application, the term "at least one nucleotide" or "at least one modified nucleotide" refers to a plurality of dNTP molecules of the same kind or species. Thus, use of "one modified nucleotide" refers to the replacement in the dNTP mix of one of the classic dNTPs dATP, dCTP, dGTP or dTTP with a corresponding modified nucleotide species.In a preferred embodiment, the at least one modified nucleotide is dUTP, replacing dTTP in the dNTP mix. In another embodiment, the at least one modified nucleotide is a biotinylated dNTP. In another embodiment, the at least one modified nucleotide contains a thio group. In another embodiment, the at least one modified nucleotide in an aminoallyl dNTP. In yet another embodiment, the at least one modified nucleotide is inosine, replacing dGTP in the dNTP mix.In some embodiments, the methods of the invention are used for construction of directional cDNA libraries. Strand marking is necessary, but not sufficient for construction of directional cDNA libraries when using adaptors that are not polarity-specific, i.e. adaptors generating ligation products with two adaptor orientations. Construction of directional cDNA libraries according to the methods of invention requires strand marking of both the cDNA insert and one of the two adaptors at the ligation strand of the adaptor. A useful feature of the present invention is the ability to switch around the adaptor orientation. For example, in a duplex adaptor system where P1/P2 designates adaptor orientation resulting in sense strand selection and (optional) sequencing, and where the P2 adaptor has at least one modified nucleotide incorporated along the ligation strand of the adaptor, modification of the protocol such that the P1 adaptor (as opposed to P2 adaptor) has at least one modified nucleotide incorporated along the ligation strand allows for antisense strand selection and (optional) sequencing.

[0078] In an embodiment where the second strand and one of the adaptors contains at least one modified nucleotide, the second strand and the one of the adaptors may be synthesized so that each comprises a sufficient and predictable density of modified nucleotides to provide for sufficient and predictable fragmentation, and when used with one or more agents capable of cleaving at the modified nucleotides (e.g., a glycosylase, a glycosylase and an amine, a glycosylase and heat, or a glycosylase and an AP endonuclease) to further generate fragments of desirable size range. Generally, a modified base can be incorporated at about every 5, 10, 15, 20, 25, 30, 40, 50, 65, 75, 85, 100, 123, 150, 175, 200, 225, 250, 300, 350, 400, 450, 500, 550, 600, 650 or more nucleotides apart in the resulting polynucleotide comprising a modified nucleotide. In one embodiment, the modified nucleotide is incorporated about every 200 nucleotides, about every 100 nucleotide, or about every 50, 25, 20, 15, 10, 9, 8, 7, 6, 5, or fewer nucleotides. In another embodiment, the modified nucleotide is incorporated about every 50 to about 200 nucleotides. In some embodiments, a 1:1, 1:2, 1:3, 1:4, 1:5, 1:6, 1:10, 1:15, 1:20 or higher ratio of modified to non-modified nucleotide may be used in the reaction mixture. In some cases, a 1:1, 1:2, 1:3, 1:4, 1:5, 1:6, 1:10, 1:15, 1:20 or higher ratio of the modified nucleotide dUTP to non-modified nucleotide dTTP is used in the reaction mixture.

[0079] The term "adaptor", as used herein, refers to an oligonucleotide of known sequence, the ligation of which to a target polynucleotide or a target polynucleotide strand of interest enables the generation of amplification-ready products of the target polynucleotide or the target polynucleotide strand of interest. Various adaptor designs are envisioned. Suitable adaptor molecules include single or double stranded nucleic acid (DNA or RNA) molecules or derivatives thereof, stem-loop nucleic acid molecules, double stranded molecules comprising one or more single stranded overhangs of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 bases or longer, proteins, peptides, aptamers, organic molecules, small organic molecules, or any adaptor molecules known in the art that may be covalently or non-covalently attached, such as for example by ligation, to the double stranded DNA fragments. The adaptors can be designed to comprise a double-stranded portion which can be ligated to double-stranded DNA (or double-stranded DNA with overhang) products. Various ligation processes and reagents are known in the art and can be useful for carrying out the methods of the invention. For example, blunt ligation can be employed. Similarly, a single dA nucleotide can be added to the 3′-end of the double-stranded DNA product, by a polymerase lacking 3′-exonuclease activity and can anneal to an adapter comprising a dT overhang (or the reverse). This design allows the hybridized components to be subsequently ligated (e.g., by T4 DNA ligase). Other ligation strategies and the corresponding reagents are known in the art and kits and reagents for carrying out efficient ligation reactions are commercially available (e.g, from New England Biolabs, Roche). The double-stranded DNA portion of the adaptors can further comprise indexing or bar-coding sequences designed to mark either the samples or sequences of interest.

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

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

Methods of Amplification

[0081] The methods, compositions and kits described herein can be useful to generate amplification-ready products for downstream applications such as massively parallel sequencing (i.e. next generation sequencing methods) or hybridization platforms. Methods of amplification are well known in the art. Suitable amplification reactions can include any DNA amplification reaction, including but not limited to polymerase chain reaction (PCR), strand displacement amplification (SDA), linear amplification, multiple displacement amplification (MDA), rolling circle amplification (RCA), single primer isothermal amplification (SPIA, see e.g. U.S. Pat. No. 6,251,639), Ribo-SPIA, or a combination thereof. In some cases, the amplification methods for providing the template nucleic acid may be performed under limiting conditions such that only a few rounds of amplification (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30 etc.), such as for example as is commonly done for cDNA generation. The number of rounds of amplification can be about 1-30, 1-20, 1-15, 1-10, 5-30, 10-30, 15-30, 20-30, 10-30, 15-30, 20-30, or 25-30.

[0082] PCR is an in vitro amplification procedure based on repeated cycles of denaturation, oligonucleotide primer annealing, and primer extension by thermophilic template dependent polynucleotide polymerase, resulting in the exponential increase in copies of the desired sequence of the polynucleotide analyte flanked by the primers. The two different PCR primers, which anneal to opposite strands of the DNA, are positioned so that the polymerase catalyzed extension product of one primer can serve as a template strand for the other, leading to the accumulation of a discrete double stranded fragment whose length is defined by the distance between the 5' ends of the oligonucleotide primers.

[0083] LCR uses a ligase enzyme to join pairs of preformed nucleic acid probes. The probes hybridize with each complementary strand of the nucleic acid analyte, if present, and ligase is employed to bind each pair of probes together resulting in two templates that can serve in the next cycle to reiterate the particular nucleic acid sequence.

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

[0085] Some aspects of the invention utilize linear amplification of nucleic acids or polynucleotides. Linear amplification generally refers to a method that involves the formation of one or more copies of the complement of only one strand of a nucleic acid or polynucleotide molecule, usually a nucleic acid or polynucleotide analyte. Thus, the primary difference between linear amplification and exponential amplification is that in the latter process, the product serves as substrate for the formation of more product, whereas in the former process the starting sequence is the substrate for the formation of product but the product of the reaction, i.e. the replication of the starting template, is not a substrate for generation of products. In linear amplification the amount of product formed increases as a linear function of time as opposed to exponential amplification where the amount of product formed is an exponential function of time.

[0086] In some embodiments, the amplification is exponential, e.g. in the enzymatic amplification of specific double stranded sequences of DNA by a polymerase chain reaction (PCR). In other embodiments the amplification method is linear. In other embodiments the amplification method is isothermal.

Downstream Applications for Whole Transcriptome Analysis

[0087] An important aspect of the invention is that the methods and compositions disclosed herein can be efficiently and cost-effectively utilized for downstream analyses, such as next generation sequencing or hybridization platforms, with minimal loss of biological material of interest. Specifically, the methods of the invention are useful for sequencing a cDNA library or a whole transcriptome while retaining information on which strand was present in the original RNA sample. In one embodiment, the invention provides for a method for whole transcriptome sequencing comprising providing a RNA sample, providing one or more primers of known or unknown sequence, combining the one or more primers with a reverse transcriptase, reverse transcribing the sample, generating double-stranded cDNA from the reverse transcribed RNA sample, wherein at least one of the four dNTPs dATP, dCTP, dGTP or dTTP is replaced by a modified dNTP during second strand synthesis and incorporated into the second strand, performing end repair on the double-stranded cDNA, ligating adaptors to the double-stranded cDNA, wherein one of the adaptors has the modified dNTP incorporated into a ligation strand of the adaptor, performing gap repair, selectively removing or marking for removal the second strand by a suitable degrading agent, amplifying the RNA sample using one or more primers to produce amplified products, and performing sequencing on the products. In some embodiments, sequencing is performed on single-stranded cDNA as generated by the methods of the present invention without amplifying the RNA sample following selective removal of the marked second strand. In some embodiments, the starting amount of RNA is .01ng to 100mg. The primers used for reverse transcription and/or amplification can be tailed primers, chimeric primers, or tailed and chimeric primers.

[0088] In one embodiment, a collection of tailed primers, and a RT enzyme is provided, wherein the RT is used in combination with the tailed primers to reverse transcribe a whole transcriptome. In

one embodiment, a collection of chimeric primers, each comprising RNA and DNA, and a RT enzyme is provided, wherein the RT is used in combination with the chimeric primers to reverse transcribe a whole transcriptome. In some embodiments, no more than 5%, 10%, 15%, 20%, 25%, 30%, 40%, 50%, 60%, 70%, 75%, 80% of the resulting products are rRNA sequences.

[0089] The methods of the invention can be useful for sequencing by the method commercialized by Illumina, as described in U.S. Pat. Nos. 5,750,341; 6,306,597; and 5,969,119. Directional (strand-specific) cDNA libraries are prepared using the methods of the present invention, and the selected single-stranded nucleic acid is amplified, for example, by PCR. The resulting nucleic acid is then denatured and the single-stranded amplified polynucleotides are randomly attached to the inside surface of flow-cell channels. Unlabeled nucleotides are added to initiate solid-phase bridge amplification to produce dense clusters of double-stranded DNA. To initiate the first base sequencing cycle, four labeled reversible terminators, primers, and DNA polymerase are added. After laser excitation, fluorescence from each cluster on the flow cell is imaged. The identity of the first base for each cluster is then recorded. Cycles of sequencing are performed to determine the fragment sequence one base at a time.

[0090] In some embodiments, the methods of the invention can be useful for preparing target polynucleotides for sequencing by the sequencing by ligation methods commercialized by Applied Biosystems (e.g., SOLiD sequencing). In other embodiments, the methods are useful for preparing target polynucleotides for sequencing by synthesis using the methods commercialized by 454/Roche Life Sciences, including but not limited to the methods and apparatus described in Margulies et al., *Nature*(2005) 437:376-380 (2005); and U.S. Pat. Nos. 7,244,559; 7,335,762; 7,211,390; 7,244,567; 7,264,929; and 7,323,305. In other embodiments, the methods can be useful for preparing target polynucleotide(s) for sequencing by the methods commercialized by Helicos BioSciences Corporation (Cambridge, Mass.) as described in U.S. application Ser. No. 11/167,046, and U.S. Pat. Nos. 7,501,245; 7,491,498; 7,276,720; and in U.S. Patent Application Publication Nos. US20090061439; US20080087826; US20060286566; US20060024711; US20060024678; US20080213770; and US20080103058. In other embodiments, the methods can be useful for preparing target polynucleotide(s) for sequencing by the methods commercialized by Pacific Biosciences as described in U.S. Pat. Nos. 7,462,452; 7,476,504; 7,405,281; 7,170,050; 7,462,468; 7,476,503; 7,315,019; 7,302,146; 7,313,308; and US Application Publication Nos. US20090029385; US20090068655; US20090024331; and US20080206764.

[0091] Another example of a sequencing technique that can be used in the methods of the provided invention is nanopore sequencing (see e.g. Soni G V and Meller A. (2007) *Clin Chem* 53: 1996-2001). A nanopore can be a small hole of the order of 1 nanometer in diameter. Immersion of a nanopore in a conducting fluid and application of a potential across it can result in a slight electrical current due to conduction of ions through the nanopore. The amount of current that flows is sensitive to the size of the nanopore. As a DNA molecule passes through a nanopore, each nucleotide on the DNA molecule

obstructs the nanopore to a different degree. Thus, the change in the current passing through the nanopore as the DNA molecule passes through the nanopore can represent a reading of the DNA sequence.

[0092] Another example of a sequencing technique that can be used in the methods of the provided invention is semiconductor sequencing provided by Ion Torrent (e.g., using the Ion Personal Genome Machine (PGM)). Ion Torrent technology can use a semiconductor chip with multiple layers, e.g., a layer with micro-machined wells, an ion-sensitive layer, and an ion sensor layer. Nucleic acids can be introduced into the wells, e.g., a clonal population of single nucleic can be attached to a single bead, and the bead can be introduced into a well. To initiate sequencing of the nucleic acids on the beads, one type of deoxyribonucleotide (e.g., dATP, dCTP, dGTP, or dTTP) can be introduced into the wells. When one or more nucleotides are incorporated by DNA polymerase, protons (hydrogen ions) are released in the well, which can be detected by the ion sensor. The semiconductor chip can then be washed and the process can be repeated with a different deoxyribonucleotide. A plurality of nucleic acids can be sequenced in the wells of a semiconductor chip. The semiconductor chip can comprise chemical-sensitive field effect transistor (chemFET) arrays to sequence DNA (for example, as described in U.S. Patent Application Publication No. 20090026082). Incorporation of one or more triphosphates into a new nucleic acid strand at the 3′ end of the sequencing primer can be detected by a change in current by a chemFET. An array can have multiple chemFET sensors.

Kits

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

[0094] The containers of the kits can generally include at least one vial, test tube, flask, bottle, syringe or other containers, into which a component may be placed, and preferably, suitably aliquotted. Where there is more than one component in the kit, the kit also will generally contain a second, third or other additional container into which the additional components may be separately placed. However, various combinations of components may be comprised in a container.

[0095] When the components of the kit are provided in one or more liquid solutions, the liquid solution can be an aqueous solution. However, the components of the kit may be provided as dried powder(s). When reagents and/or components are provided as a dry powder, the powder can be reconstituted by the addition of a suitable solvent.

[0096] The present invention provides kits containing one or more compositions of the present invention and other suitable reagents suitable for carrying out the methods of the present invention. The invention provides, e.g., diagnostic kits for clinical or criminal laboratories, or nucleic acid amplification or analysis kits for general laboratory use. The present invention thus includes kits which include some or all of the reagents necessary to carry out the methods of the present invention, e.g., sample preparation reagents, oligonucleotides, binding molecules, stock solutions, nucleotides,

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

[0097] In some embodiments, the kit may contain one or more reaction mixture components, or one or more mixtures of reaction mixture components. In some cases, the reaction mixture components or mixtures thereof may be provided as concentrated stocks, such as 1.1x, 1.5 x, 2 x, 2.5 x, 3 x, 4 x, 5 x, 6 x, 7 x, 10 x, 15 x, 20 x, 25 x, 33 x, 50 x, 75 x, 100 x or higher concentrated stock. The reaction mixture components may include any of the compositions provided herein including but not limited to buffers, salts, divalent cations, azeotropes, chaotropes, dNTPs, labeled nucleotides, non-canonical or modified nucleotides, dyes, fluorophores, biotin, enzymes (such as endonucleases, exonucleases, glycosylases), or any combination thereof.

[0098] In some embodiments, the kit may contain one or more oligonucleotide primers, such as the oligonucleotide primers provided herein. For example, the kit may contain one or more oligonucleotide primers comprising random hybridizing portions. Alternatively, the kit may contain oligonucleotide primers comprising polyT hybridizing portions. In some cases, the kit may contain oligonucleotide primers that comprise random hybridizing portions and primers comprising polyT hybridizing portions. In still other cases, the kit may contain "not so random" primers that have been pre-selected to hybridize to desired nucleic acids, but not hybridize to undesired nucleic acids. In some cases the kit may contain tailed primers comprising a 3'-portion hybridizable to the target nucleic acid and a 5'-portion which is not hybridizable to the target nucleic acid. In some cases, the kit may contain chimeric primers comprising an RNA portion and a DNA portion. In some cases, the kit may contain primers comprising non-canonical or modified nucleotides.

[0099] In some embodiments, the kit of the present invention may contain one or more polymerases or mixtures thereof. In some cases, the one or more polymerases or mixtures thereof may comprise strand displacement activity. Suitable polymerases include any of the polymerases provided herein. The kit may further contain one or more polymerase substrates such as for example dNTPs, non-canonical or modified nucleotides.

[00100] In some embodiments, the kit of the present invention may contain one or more means for purification of the nucleic acid products, removing of the fragmented products from the desired products, or combination of the above. Suitable means for the purification of the nucleic acid products include but are not limited to single stranded specific exonucleases, affinity matrices, nucleic acid purification columns, spin columns, ultrafiltration or dialysis reagents, or electrophoresis reagents including but not limited acrylamide or agarose, or any combination thereof.

[00101] In some embodiments, the kit of the present invention may contain one or more agents capable of cleaving the base portion of a non-canonical nucleotide to generate an abasic site. In some

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

[00102] In some embodiments, the kit of the present invention may contain one or more agents capable of fragmenting a phosphodiester backbone at an abasic site to fragment the input nucleic acid template. In some cases, this agent may comprise one or more amines, primary amines, secondary amines, polyamines such as DMED, piperidine, AP endonucleases, or any combination thereof.

[00103] In some embodiments, the kit of the present invention may contain one or more reagents for producing blunt ends from the double stranded products generated by the extension reaction. For example, the kit may contain one or more of single stranded DNA specific exonucleases including but not limited to exonuclease 1 or exonuclease 7; a single stranded DNA specific endonucleases such as mung bean exonuclease or S1 exonuclease, one or more polymerases such as for example T4 DNA polymerase or Klenow polymerase, or any mixture thereof. Alternatively, the kit may contain one or more single stranded DNA specific exonucleases, endonucleases and one or more polymerases, wherein the reagents are not provided as a mixture. Additionally, the reagents for producing blunt ends may comprise dNTPs.

[00104] In some embodiments, the kit of the present invention may contain one or more reagents for preparing the double stranded products for ligation to adaptor molecules. For example, the kit may contain dATP, dCTP, dGTP, dTTP, or any mixture thereof. In some cases, the kit may contain a polynucleotide kinase, such as for example T4 polynucleotide kinase. Additionally, the kit may contain a polymerase suitable for producing a 3' extension from the blunt ended double stranded DNA fragments. Suitable polymerases are included, for example, exo-Klenow polymerase.

[00105] In some embodiments, the kit of the present invention may contain one or more adaptor molecules such as any of the adaptor molecules provided herein. Suitable adaptor molecules include single or double stranded nucleic acid (DNA or RNA) molecules or derivatives thereof, stem-loop nucleic acid molecules, double stranded molecules comprising one or more single stranded overhangs of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 bases or longer, proteins, peptides, aptamers, organic molecules, small organic molecules, or any adaptor molecules known in the art that may be covalently or non-covalently attached, such as for example by ligation, to the double stranded DNA fragments.

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

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

Q-PCR, and characterizing amplified nucleic acid products generated by the methods of the invention, or any combination thereof. In some cases, the kit may contain instructions for generating a second strand comprising one or more modified nucleotides. The kit may further contain instructions for mixing the one or more reaction mixture components to generate one or more reaction mixtures suitable for the methods of the present invention. The kit may further contain instructions for hybridizing the one or more oligonucleotide primers to a nucleic acid template. The kit may further contain instructions for extending the one or more oligonucleotide primers with for example a polymerase. The kit may further contain instructions for cleaving the base portion of a modified nucleotide to generate an abasic site, with, for example, a glycosylase. The kit may further contain instructions for fragmenting a phosphodiester backbone at an abasic site to fragment the input nucleic acid template, with, for example, any of the suitable agents provided herein such as a polyamine. The kit may further contain instructions for purification of any of the products provided by any of the steps of the methods provided herein. The kit may further contain instructions for producing blunt ended fragments, for example by removing single stranded overhangs or filling in single stranded overhangs, with for example single stranded DNA specific exonucleases, polymerases, or any combination thereof. The kit may further contain instructions for phosphorylating the 5' ends of the double stranded DNA fragments produced by the methods of the present invention. The kit may further contain instructions for ligating one or more adaptor molecules to the double stranded DNA fragments of the present invention.

[00108] A kit will preferably include instructions for employing, the kit components as well the use of any other reagent not included in the kit. Instructions may include variations that can be implemented.

Products based on the Methods of the Invention

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

EXAMPLES

Example 1 – Generation of a directional cDNA library.

[00110] This example describes the generation of a directional cDNA library using conventional blunt-end ligation with modified duplex adaptors and 50 ng of poly(A)+ selected messenger RNA as a starting material. An overview of the end-end work-flow for the generation of the directional cDNA library is shown in FIG. 4.

[00111] First strand synthesis

First strand cDNA was generated using random hexamer priming. The first strand synthesis reaction was conducted using the Invitrogen SuperScript III Reverse Transcriptase kit, with 10 μM of random

hexamers, 3.0 mM $MgCl_2$ and 1.0 mM dNTPs. The cDNA synthesis reaction was carried out in 10 µL volume, incubated at 40 degrees Celsius for 60 minutes and chilled to 4 degrees Celsius.

[00112] Second strand synthesis with dUTP incorporation

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

[00113] Fragmentation and purification of cDNA fragments

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

[00114] End repair

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

[00115] Ligation with dU marked adaptors

Duplex adaptors were ligated to blunt-ended cDNA fragments according to the Encore ™ Ultra Low Input NGS Library System I User Guide Instructions, with the exception that the Ligation Adaptor Mix contained one adaptor where the ligation strand of the adaptor had at least one dU incorporated into it.

[00116] Nick repair/adaptor fill-in

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

[00117] Strand selection with UDG/APE I treatment

Uridine digestion was performed with 1 unit of UNG and 1,000 units of APE I at 37° C for 20 minutes. Incorporation of dUTP into one strand of the cDNA insert and the ligation strand of one of the two adaptors allowed for selective removal of the products with the undesired adaptor orientation.

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

[00118] Library amplification

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

Example 2 – RNA Strand Retention Efficiency.

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

Example 3 – RNA Strand Retention Efficiency.

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

Example 4 – Generation of a directional cDNA library.

[00121] This example describes the generation of a directional cDNA library using conventional blunt-end ligation with modified duplex adaptors and 50 ng of poly(A)+ selected messenger RNA as a starting material.

[00122] First strand synthesis with dUTP incorporation

First strand cDNA was generated using random hexamer priming. The first strand synthesis reaction was conducted using the Invitrogen SuperScript III Reverse Transcriptase kit, with 10 μM of random hexamers, 3.0 mM $MgCl_2$ and supplemented with dNTP mix containing dATP, dCTP, dGTP, and dUTP in place of dTTP. The cDNA synthesis reaction was carried out in 10 μL volume, incubated at 40 degrees Celsius for 60 minutes and chilled to 4 degrees Celsius. After first strand synthesis, non-incorporated dNTPs were removed prior to second strand synthesis.

[00123] Second strand synthesis

Second strand synthesis was performed using the New England Biolabs NEBNext Second Strand Synthesis Module, where the Second Strand Synthesis (dNTP-free) Reaction Buffer was

supplemented with dNTP mix containing dATP, dCTP, dGTP, and dTTP. RNAse H-mediated nick translation was carried out by adding 65 µL of second strand synthesis master mix and incubating for one hour at 16 degrees Celsius. The reaction was stopped by adding 45 µL of 25 mM EDTA.

[00124] Fragmentation and purification of cDNA fragments

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

[00125] End repair

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

[00126] Ligation with dU marked adaptors

Duplex adaptors were ligated to blunt-ended cDNA fragments according to the Encore ™ Ultra Low Input NGS Library System I User Guide Instructions, with the exception that the Ligation Adaptor Mix contained one adaptor where the ligation strand of the adaptor had at least one dU incorporated into it.

[00127] Nick repair/adaptor fill-in

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

[00128] Strand selection with UDG/APE I treatment

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

[00129] Library amplification

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

CLAIMS

WHAT IS CLAIMED IS:

1. A method for construction of a directional cDNA library comprising:
    a. reverse transcribing a RNA sample to generate a first strand cDNA;
    b. generating a second strand cDNA from the first strand cDNA, wherein at least one of the four dNTPs dATP, dCTP, dGTP or dTTP is replaced by a modified dNTP during second strand synthesis and incorporated into the second strand, thereby generating a double-stranded cDNA;
    c. performing end repair on the double-stranded cDNA;
    d. ligating adaptors to the double-stranded cDNA, wherein only one of the adaptors has the modified dNTP incorporated into a ligation strand of the adaptor;
    e. performing gap repair; and
    f. selectively cleaving the second strand and the adaptor that has the modified dNTP by a suitable cleavage agent, thereby generating a directional cDNA library comprising the first strand cDNA.
2. The method of claim 1, wherein the modified dNTP comprises dUTP.
3. The method of claim 1, wherein the method, optionally, further comprises fragmenting the double-stranded cDNA prior to step (c).
4. The method of claim 1, wherein step (f) comprises cleaving a base portion of the modified nucleotide, thereby forming an abasic site.
5. The method of claim 4, further comprising cleaving a phosphodiester backbone at the abasic site.
6. The method of claim 4, wherein the cleaving the base portion of the modified nucleotide comprises use of an enzyme.
7. The method of claim 5, wherein the cleaving the phosphodiester backbone at the abasic site comprises use of an enzyme, chemical agent, and/or heat.
8. The method of claim 7, wherein the enzyme is an endonuclease.
9. The method of claim 8, wherein the endonuclease is an apurinic/apyrimidinic endonuclease (APE).
10. The method of claim 6, wherein the enzyme comprises a glycosylase.
11. The method of claim 10, wherein the glycosylase is uracil-N-glycosylase (UNG) or uracil DNA glycosylase (UDG).
12. The method of claim 7, wherein the chemical agent comprises a primary amine.
13. The method of claim 7, wherein the chemical agent comprises a polyamine.
14. The method of claim 13, wherein the polyamine comprises N,N-dimethylethylenediamine (DMED).

15. The method of claim 1, wherein the cleavage agent comprises a glycosylase and a polyamine.
16. The method of claim 1, wherein the cleavage agent comprises a glycosylase and a APE.
17. The method of claim 1, further comprising amplifying the directional cDNA library, thereby generating amplified products.
18. The method of claim 17, wherein the amplification of the directional cDNA library comprises polymerase chain reaction (PCR), strand displacement amplification (SDA), multiple displacement amplification (MDA), rolling circle amplification (RCA), single primer isothermal amplification (SPIA), or ligase chain reaction (LCR).
19. The method of claim 18, wherein the amplification comprises PCR.
20. The method of claim 18, wherein the amplification comprises SPIA.
21. The method of claim 17, further comprising an additional step of sequencing the amplified products.
22. The method of claim 21, wherein the sequencing comprises next generation sequencing.
23. The method of claim 1, further comprising cleaving the RNA sample following reverse transcription.
24. The method of claim 23, wherein cleaving the RNA sample comprises exposing the RNA sample to an RNase following reverse transcription.
25. The method of claim 24, wherein the RNase is RNase H.
26. The method of claim 23, wherein cleaving the RNA sample comprises exposing the RNA sample to heat or chemical treatment or a combination thereof.
27. The method of claim 1 further comprising reducing or depleting non-desired nucleic acid sequences from the directional cDNA library.
28. The method of claim 27, wherein the non-desired nucleic acid is ribosomal RNA (rRNA).
29. A method for whole transcriptome directional sequencing comprising:
    a. reverse transcribing a RNA sample to generate a first strand cDNA;
    b. generating a second strand cDNA from the first strand cDNA, wherein at least one of the four dNTPs dATP, dCTP, dGTP or dTTP is replaced by a modified dNTP during second strand synthesis and incorporated into the second strand, thereby generating a double-stranded cDNA;
    c. performing end repair on the double-stranded cDNA;
    d. ligating adaptors to the double-stranded cDNA, wherein only one of the adaptors has the modified dNTP incorporated into a ligation strand of the adaptor;
    e. performing gap repair;
    f. selectively cleaving the second strand and the adaptor that has the modified dNTP, by a suitable cleavage agent, thereby generating a directional cDNA library comprising the first strand cDNA; and

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

45. The method of claim 29, wherein the amplification of the remaining cDNA strand comprises polymerase chain reaction (PCR), strand displacement amplification (SDA), multiple displacement amplification (MDA), rolling circle amplification (RCA), single primer isothermal amplification (SPIA), or ligase chain reaction (LCR).

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

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

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

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

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

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