INTERNATIONAL SEARCH REPORT

International application No.
PCT/US14/24581

A. CLASSIFICATION OF SUBJECT MATTER
IPC(8) - C40B 40/06; C12P 19/34 (2014.01)
USPC - 506/16, 15, 13, 23
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC(8): C40B 40/06; C12P 19/34 (2014.01)
USPC: 435/91.2, 91.1, 89, 85, 84, 72, 41; 506/16, 15, 13, 7, 23; 514/44R, 43, 42, 23.1

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

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
MicroPatent (US-G, US-A, EP-A, EP-B, WO, JP-bib, DE-C,B, DE-A, DE-T, DE-U, GB-A, FR-A); Google Scholar; Pubmed; ScienceDirect; overhang, '3',' '5',' 'directional library,' 'adapter molecule,' polynucleotide

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 2011/0224105 A1 (KURN, N et al.) September 15, 2011; figure 7; paragraphs [0006], [0023], [0053], [0079], [0087], [0104]-[0107], [0129], [0130], [0134]	1-41
Y	BORODINA, T et al. A Strand-Specific Library Preparation Protocol For RNA Sequencing. Methods Enzymol. 2011, Vol. 500; pages 79-98; abstract; page 82, third paragraph; page 83, first paragraph. DOI: 10.1016/B978-0-12-385118-5.00005-0.	1-41
Y	US 2012/0283145 A1 (WANG, Y) November 8, 2012; paragraphs [0013], [0064]	13/12/9/1-13/12/9/4

☐ Further documents are listed in the continuation of Box C. ☐

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
08 July 2014 (08.07.2014)	29 JUL 2014
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-3201	Authorized officer: Shane Thomas PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

Form PCT/ISA/210 (second sheet) (July 2009)

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
3 September 2015 (03.09.2015)

WIPO | PCT

(10) International Publication Number
WO 2015/131107 A1

(51) International Patent Classification:
C12P 19/34 (2006.01)

(21) International Application Number:
PCT/US2015/018112

(22) International Filing Date:
27 February 2015 (27.02.2015)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/946,617 28 February 2014 (28.02.2014) US
61/968,982 21 March 2014 (21.03.2014) US

(71) Applicant: **NUGEN TECHNOLOGIES, INC.** [US/US]; 201 Industrial Road, Suite 310, San Carlos, CA 94070 (US).

(72) Inventor: **SCHROEDER, Benjamin, G.**; 1964 Shoreview Avenue, San Mateo, CA 94401 (US).

(74) Agents: **MCEVOY, Michael, T.** et al.; Wilson Sonsini Goodrich & Rosati, 650 Page Mill Road, Palo Alto, CA 94304-1050 (US).

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

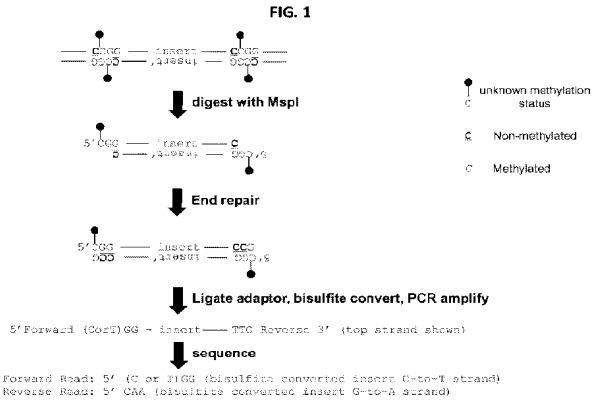
(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR,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

Published:

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

(54) **Title**: REDUCED REPRESENTATION BISULFITE SEQUENCING WITH DIVERSITY ADAPTORS

FIG. 1



(57) **Abstract**: Described herein are methods, compositions and kits for the generation of bisulfite-converted libraries useful for conducting reduced representation bisulfite sequencing (RRBS). The methods described herein can be employed to generate RRBS libraries in a manner that is easier and more cost-efficient than conventional RRBS methods, and can be efficiently sequenced with next generation sequencing (NGS) techniques without the need for genomic, higher diversity sequencing controls such as PhiX spike-ins.

WO 2015/131107 A1

REDUCED REPRESENTATION BISULFITE SEQUENCING WITH DIVERSITY ADAPTORS

CROSS-REFERENCE

[0001] This application claims the benefit of U.S. Provisional Application No. 61/946,617, filed on February 28, 2014, and U.S. Provisional Application No. 61/968,982, filed on March 21, 2014, each of which is herein incorporated by reference in its entirety.

BACKGROUND

[0002] DNA methylation is an epigenetic modification that can occur at the C5 position of cytosine residues. In mammals, 5-methylcytosine can appear in the CpG dinucleotide context (Ramsahoye *et al.*, *Proc Natl Acad Sci USA* 97:5237–5242, 2000). Recent data suggests that approximately 25 % of all cytosine methylation identified in stem cells can occur in non-CpG context (*see* Ziller *et al.*, *PLoS Genet.* 7(12):e1002389, 2011). Although CpG dinucleotides can be underrepresented in the genome, stretches of sequences known as CpG islands can exist that are rich in CpG dinucleotides. These CpG islands can be associated with promoter regions and span several hundred nucleotides or more.

[0003] Epigenomics, i.e., the study of the complete set of epigenetic modifications on the genetic material of a cell, has revealed that epigenetic modifications (e.g., DNA methylation) can play a role in mammalian development and disease. For example, DNA methylation is implicated in embryonic development, genomic imprinting and X-chromosome inactivation through regulation of transcriptional activity, chromatin structure and chromatin stability (Robertson, *Nat Review Genet* 6:597-610, 2005). Increased DNA methylation (hypermethylation) at promoter regions of genes can be associated with transcriptional silencing, whereas decreased methylation (hypomethylation) at promoter regions can be associated with increased gene activity. Aberrant methylation patterns can be associated with various human pathologies, including tumor formation and progression (Feinberg and Fogelstein, *Nature* 301:89-92, 1983; Esteller, *Nat Review Genet* 8: 286-298, 2007; and Jones and Paylin, *Cell* 128:683-692, 2007). Therefore, analysis of DNA methylation status across the human genome can be of interest.

[0004] Methods for measuring DNA methylation at specific genomic loci include, for example, immunoprecipitation of methylated DNA, methyl-binding protein enrichment of methylated fragments, digestion with methylation-insensitive restriction enzymes, and bisulfite conversion followed by Sanger sequencing (reviewed in Laird, *Nat Review Genet* 11: 191-203,

IDT Exhibit 1017.1551

2010). Bisulfite treatment can convert unmethylated cytosine residues into uracils (the readout of which can be thymine after amplification with a polymerase). Methylcytosines can be protected from conversion by bisulfite treatment to uracils. Following bisulfite treatment, methylation status of a given cytosine residue can be inferred by comparing the sequence to an unmodified reference sequence.

[0005] Techniques have been developed for profiling methylation status of the whole genome, i.e. the methylome, at a single-base resolution using high throughput sequencing technologies. Bisulfite conversion of genomic DNA combined with next generation sequencing (NGS), or BS-seq, is one strategy. Because of the high cost still associated with genome-wide methylation sequencing, variations of BS-seq technology that enable genome partitioning to enrich for regions of interest can be used. One such variation is reduced representation BS-seq (RRBS), which can involve digestion of a DNA sample with a methylation-insensitive restriction endonuclease that has CpG dinucleotide as a part of its recognition site, followed by bisulfite sequencing of the selected fragments (Meissner *et al.*, *Nucleic Acids Res.* 33(18):5868-5877, 2005).RRBS provides less coverage than whole genome bisulfite sequencing, but allows the researcher to obtain quantitative DNA methylation information across many features with approximately 50-fold fewer sequencing reads, resulting in substantial sequencing cost reduction. Since the RRBS technique was first described (Meissner, et al. (2005) *Nucleic Acids Res* 33(18):5868), there have been many enhancements and improvements (see for example, Boyle, et al. (2012) *Genome Biol* 13:R92). The current approach utilizes the methylation insensitive restriction enzyme MspI, which recognizes CCGG. As a result of partial fragmentation during bisulfite conversion, PCR, and efficiency of cluster generation, only a subset of these fragments, typically under 300 bp in length, can be sequenced. However these smaller fragments are derived from genomic DNA that has a high frequency of MspI sites and therefore a high frequency of potential CpG methylation sites. This is how RRBS can achieve reduced representation. One feature of RRBS can be that all forward reads start with either TGG or CGG. This feature can create a lack of color balance that can be problematic for some NGS sequencers. For example, the complete lack of color balance at bases 2 and 3 can make these samples very challenging for both Illumina MiSeq and HiSeq instruments. The methods, compositions and kits provided herein can allow RRBS libraries to be generated easily and sequenced more efficiently by increasing the diversity of nucleotide bases sequenced during the initial cycles of sequencing. The increased diversity can increase the efficiency and accuracy of cluster identification by sequencers during certain types of next generation sequencing applications (e.g., Illumina based sequencing by synthesis).

IDT Exhibit 1017.1552

SUMMARY

[0006] In one aspect, provided herein is a method for increasing diversity of sequences at ends of a plurality of polynucleotides in a sample, the method comprising introducing a pool of oligonucleotide sequences to 3' ends of the plurality of polynucleotides, a. wherein each oligonucleotide sequence in the pool of oligonucleotide sequences comprises a sequence complementary to a sequencing primer, b. wherein the pool of oligonucleotide sequences comprises a first sequence, wherein a terminal 3' nucleotide base of the sequencing primer is capable of annealing to a terminal 5' nucleotide base of the first sequence, wherein following introduction the 5' end of the first sequence is adjacent to a 3' end of a polynucleotide from the plurality of polynucleotides, c. wherein the pool of oligonucleotide sequences comprises a second sequence, wherein the second sequence terminates at its 5' end with an additional base relative to the 5' end of the first sequence, wherein following introduction the 5' end of the second sequence is adjacent to a 3' end of a polynucleotide from the plurality of polynucleotides; d. wherein the pool of oligonucleotide sequences comprises a third sequence, wherein the third sequence terminates at its 5' end with two additional bases relative to the 5' end of the first sequence, wherein following introduction the 5' end of the third sequence is adjacent to a 3' end of a polynucleotide from the plurality of polynucleotides; e. wherein the pool of oligonucleotide sequences comprises a fourth sequence, wherein the fourth sequence terminates at its 5' end with three additional bases relative to the 5' end of the first sequence, wherein following introduction the 5' end of the fourth sequence is adjacent to a 3' end of a polynucleotide from the plurality of polynucleotides; and f. wherein the introducing generates a plurality of oligonucleotide-polynucleotide complexes, thereby increasing diversity of sequences at ends of the plurality of polynucleotides. In some cases, the sample is a whole genome sample. In some cases, the sample comprises a polynucleotide library. In some cases, the method further comprises fragmenting the polynucleotides prior to introducing the pool of oligonucleotide sequences to the 3' ends of the plurality of polynucleotides. In some cases, the fragmenting is performed by contacting the polynucleotides with an enzyme. In some cases, the enzyme is a restriction enzyme. In some cases, the restriction enzyme is a methylation insensitive restriction enzyme. In some cases, the methylation insensitive restriction enzyme is MspI. In some cases, the additional base of the second sequence is an adenine, thymine, or cytosine base. In some cases, each of the two additional bases of the third sequence is an adenine, thymine, or cytosine base. In some cases, a 3' most base of the three additional bases of the fourth sequence comprises a thymine or cytosine base, and each of the two other bases of the three additional bases of the fourth sequence comprises an adenine, thymine, or cytosine base. In some cases, each oligonucleotide sequence

IDT Exhibit 1017.1553

in the pool of oligonucleotide sequences is within a strand of a duplexed adaptor. In some cases, the duplexed adaptor is a partial duplexed adaptor comprising a short strand and a long strand. In some cases, the duplexed adaptor is a partial duplexed adaptor comprising two strands of equal length. In some cases, the short strand is a non-ligation strand of the duplexed adaptor, and the long strand is a ligation strand of the duplexed adaptor. In some cases, the non-ligation strand comprises a 5' overhang. In some cases, the 5' overhang comprises two bases. In some cases, the two bases are 3'-GC-5'. In some cases, the introducing comprises a ligation reaction. In some cases, the ligation reaction comprises cohesive end ligation. In some cases, the ligation reaction comprises blunt end ligation. In some cases, the plurality of polynucleotides are strands of double-stranded nucleic acids, wherein each 5' end of the double-stranded nucleic acids comprises a ligation end and each 3' end comprises a non-ligation end, wherein each of the non-ligation ends of the double-stranded nucleic acid comprises part of the polynucleotide sequence. In some cases, the 3'-GC-5' of the non-ligation strand of the duplexed adaptor anneals to a 5'-CG-3' sequence of the ligation end of the double-stranded sequence comprising the polynucleotide sequence. In some cases, the 3'-GC-5' of the non-ligation strand of the duplexed adaptor is 5' to the terminal 5' nucleotide base of the first sequence. In some cases, the 3'-GC-5' of the non-ligation strand of the duplexed adaptor is 5' to the additional base of the second sequence relative to the 5' end of the first sequence. In some cases, the 3'-GC-5' of the non-ligation strand of the duplexed adaptor is 5' to the two additional bases of the third sequence relative to the 5' end of the first sequence. In some cases, the 3'-GC-5' of the non-ligation strand of the duplexed adaptor is 5' to the three additional bases of the fourth sequence relative to the 5' end of the first sequence. In some cases, the ligation strand is capable of ligating to a 5' end of the plurality of polynucleotides, and wherein a nick is formed between the non-ligation strand and the end of the plurality of polynucleotides. In some cases, the method further comprises an extending step, wherein the non-ligation end of the plurality of polynucleotides is extended using a polymerase in the presence of a mixture of dNTPs using the ligation strand of the adaptor as template, thereby introducing the pool of oligonucleotide sequences to the 3' ends of the plurality of polynucleotides. In some cases, the ligation strand of each adaptor duplex in the pool comprise cytosine analogs resistant to bisulfite treatment and the mixture of dNTPs comprises dATP, dTTP, dGTP, and dCTP, whereby the extending generates the plurality of oligonucleotide sequence-polynucleotide complexes, wherein each of the plurality of oligonucleotide sequence-polynucleotide complexes comprises complementary ends, wherein one end comprises cytosine bases sensitive to bisulfite treatment, while the opposite end comprises cytosine analogs resistant to bisulfite treatment. In some cases, the ligation strand of each adaptor duplex in the pool of

IDT Exhibit 1017.1554

adaptor duplexes comprise cytosine bases sensitive to bisulfite treatment and the mixture of dNTPs comprises dATP, dTTP, dGTP, and a dCTP analog resistant to bisulfite treatment, whereby the extending generates the plurality of oligonucleotide sequence-polynucleotide complexes, wherein each of the plurality of oligonucleotide sequence-polynucleotide complexes comprises complementary ends, wherein one end comprises cytosine bases sensitive to bisulfite treatment, while the opposite end comprises cytosine analogs resistant to bisulfite treatment. In some cases, each oligonucleotide sequence in the pool of oligonucleotide sequences is a primer, wherein the pool of oligonucleotide sequences comprises the primer and a complement thereof. In some cases, the introducing comprises an amplification reaction. In some cases, the amplification reaction comprises polymerase chain reaction. In some cases, each of the primers in the pool comprises cytosine bases sensitive to bisulfite treatment, wherein the amplification is conducted in the presence of a mixture of dNTPs comprising dATP, dTTP, dGTP, and a dCTP analog resistant to bisulfite treatment, thereby generating the plurality of oligonucleotide sequence-polynucleotide complexes, wherein each of the plurality of oligonucleotide sequence-polynucleotide complexes comprises complementary ends, wherein one end comprises cytosine bases sensitive to bisulfite treatment, while the opposite end comprises cytosine analogs resistant to bisulfite treatment. In some cases, each of the primers in the pool of primers comprises cytosine analogs resistant to bisulfite treatment, wherein the amplification is conducted in the presence of a mixture of dNTPs comprising dATP, dTTP, dGTP, and dCTP, thereby generating the plurality of oligonucleotide sequence-polynucleotide complexes, wherein each of the plurality of oligonucleotide sequence-polynucleotide complexes comprises complementary ends, wherein one end comprises cytosine bases sensitive to bisulfite treatment, while the opposite end comprises cytosine analogs resistant to bisulfite treatment. In some cases, the cytosine analog resistant to bisulfite treatment is 5-methylcytosine, 5-hydroxymethylcytosine, or 5-propynylcytosine. In some cases, the cytosine analog resistant to bisulfite treatment is 5-methylcytosine. In some cases, the method further comprises amplifying the plurality of oligonucleotide sequence-polynucleotide complexes, wherein the amplifying comprises solid-phase nucleic acid amplification to generate a plurality of clusters, wherein each cluster represents a plurality of copies of each of the plurality of oligonucleotide sequence-polynucleotide complexes. In some cases, the method further comprises treating the plurality of oligonucleotide sequence-polynucleotides complexes with bisulfite prior to the amplification, wherein the treating generates a plurality of oligonucleotide sequence-polynucleotide complexes comprising non-complementary ends. In some cases, the treating with bisulfite prior to amplification generates the sequence complementary to a sequencing primer in each

IDT Exhibit 1017.1555

oligonucleotide sequence in the pool of oligonucleotide sequences comprises. In some cases, the method further comprises sequencing each of the plurality of clusters, wherein the sequencing comprises hybridizing the sequencing primer to the sequence complementary to a sequencing primer, and conducting sequencing by synthesis. In some cases, the introducing of the pool of oligonucleotide sequences to the polynucleotides improves identification of each of the plurality of clusters generated from the plurality of oligonucleotide sequence-polynucleotide complexes relative to the identification of each cluster from a plurality of clusters generated from the plurality of polynucleotides not introduced to the pool of oligonucleotide sequences. In some cases, the sequencing comprises paired-end sequencing.

[0007] In one aspect, provided herein is a method for generating a bisulfite converted library, the method comprising: a. introducing a pool of oligonucleotide sequences to each end of a plurality of polynucleotides from a sample, i. wherein each oligonucleotide sequence in the pool of oligonucleotide sequences comprises a sequence complementary to a sequencing primer and a sequence element adjacent to the sequence complementary to a sequencing primer, ii. wherein the pool of oligonucleotide sequences comprises a first sequence, wherein the first sequence terminates at its 3' end with the sequence element, wherein following ligation the 3' end of the sequence element is adjacent to a 5' end of a polynucleotide from the plurality of polynucleotides, iii. wherein the pool of oligonucleotide sequences comprises a second sequence, wherein the second oligonucleotide sequence terminates at its 3' end with an additional base relative to the 3' end of the first sequence, wherein following ligation the 3' end of the second sequence is adjacent to a 5' end of a polynucleotide from the plurality of polynucleotides; iv. wherein the pool of oligonucleotide sequences comprises a third sequence, wherein the third sequence terminates at its 3' end with two additional bases relative to the 3' end of the first sequence, wherein following ligation the 3' end of the third sequence is adjacent to a 5' end of a polynucleotide from the plurality of polynucleotides; v. wherein the pool of oligonucleotide sequences comprises a fourth sequence, wherein the fourth sequence terminates at its 3' end with three additional bases relative to the 3' end of the first sequence, wherein following ligation the 3' end of the fourth sequence is adjacent to a 5' end of a polynucleotide from the plurality of polynucleotides; vi. wherein the introducing generates a plurality of oligonucleotide sequence-polynucleotide complexes; and b. treating the plurality of oligonucleotide sequence-polynucleotide complexes with bisulfite, thereby generating a bisulfite converted polynucleotide library. In some cases, the sample is a whole genome sample. In some cases, the sample comprises a polynucleotide library. In some cases, the method further comprises fragmenting the polynucleotides prior to introducing the pool of oligonucleotide sequences to the 3' ends of the

IDT Exhibit 1017.1556

plurality of polynucleotides. In some cases, the fragmenting is performed by contacting the polynucleotides with an enzyme. In some cases, the enzyme is a restriction enzyme. In some cases, the restriction enzyme is a methylation insensitive restriction enzyme. In some cases, the methylation insensitive restriction enzyme is MspI. In some cases, the additional base of the second sequence is an adenine, thymine, or guanine base. In some cases, each of the two additional bases of the third sequence is an adenine, thymine, or guanine base. In some cases, a 5' most base of the three additional bases of the fourth sequence comprises an adenine or guanine base, and each of the two other bases of the three additional bases of the fourth sequence comprises an adenine, thymine, or guanine base. In some cases, each oligonucleotide sequence in the pool of oligonucleotide sequences is within a strand of a duplexed adaptor. In some cases, the duplexed adaptor is a partial duplexed adaptor. In some cases, the strand is a ligation strand of the duplexed adaptor. In some cases, the non-ligation strand comprises a 5' overhang. In some cases, the 5' overhang comprises two bases. In some cases, the two bases are 3'-GC-5'. In some cases, the introducing comprises a ligation reaction. In some cases, the ligation reaction comprises cohesive end ligation. In some cases, the ligation reaction comprises blunt end ligation. In some cases, the plurality of polynucleotides are double-stranded nucleic acids, wherein each 5' end of the double-stranded nucleic acids comprises a ligation end and each 3' end comprises a non-ligation end. In some cases, each duplexed adaptor in the pool comprises a ligation strand and a non-ligation strand, wherein the ligation strand is capable of ligating to the ligation end of the plurality of polynucleotides, and wherein a nick is formed between the non-ligation strand and the non-ligation end of the plurality of polynucleotides. In some cases, the method further comprises an extending step, wherein the non-ligation end of the plurality of polynucleotides is extended using a polymerase in the presence of a mixture of dNTPs using the ligation strand of the adaptor as template, thereby introducing the pool of oligonucleotide sequences to the 3' ends of the plurality of polynucleotides. In some cases, the ligation strand of each adaptor duplex in the pool comprise cytosine analogs resistant to bisulfite treatment and the mixture of dNTPs comprises dATP, dTTP, dGTP, and dCTP, whereby the extending generates the plurality of oligonucleotide sequence-polynucleotide complexes, wherein each of the plurality of oligonucleotide sequence-polynucleotide complexes comprises complementary ends, wherein one end comprises cytosine bases sensitive to bisulfite treatment, while the opposite end comprises cytosine analogs resistant to bisulfite treatment. In some cases, the ligation strand of each adaptor duplex in the pool of adaptor duplexes comprise cytosine bases sensitive to bisulfite treatment and the mixture of dNTPs comprises dATP, dTTP, dGTP, and a dCTP analog resistant to bisulfite treatment, whereby the extending generates the plurality of oligonucleotide sequence-

IDT Exhibit 1017.1557

polynucleotide complexes, wherein each of the plurality of oligonucleotide sequence-polynucleotide complexes comprises complementary ends, wherein one end comprises cytosine bases sensitive to bisulfite treatment, while the opposite end comprises cytosine analogs resistant to bisulfite treatment. In some cases, each oligonucleotide sequence in the pool of oligonucleotide sequences is within a primer, wherein the pool of oligonucleotide sequences comprises the primer and a complement thereof. In some cases, the introducing comprises an amplification reaction. In some cases, the amplification reaction comprises polymerase chain reaction. In some cases, each of the primers in the pool comprises cytosine bases sensitive to bisulfite treatment, wherein the amplification is conducted in the presence of a mixture of dNTPs comprising dATP, dTTP, dGTP, and a dCTP analog resistant to bisulfite treatment, thereby generating the plurality of oligonucleotide sequence -polynucleotide complexes, wherein each of the plurality of oligonucleotide sequence -polynucleotide complexes comprises complementary ends, wherein one end comprises cytosine bases sensitive to bisulfite treatment, while the opposite end comprises cytosine analogs resistant to bisulfite treatment. In some cases, each of the primers in the pool of primers comprises cytosine analogs resistant to bisulfite treatment, wherein the amplification is conducted in the presence of a mixture of dNTPs comprising dATP, dTTP, dGTP, and dCTP, thereby generating the plurality of oligonucleotide sequence-polynucleotide complexes, wherein each of the plurality of oligonucleotide sequence-polynucleotide complexes comprises complementary ends, wherein one end comprises cytosine bases sensitive to bisulfite treatment, while the opposite end comprises cytosine analogs resistant to bisulfite treatment. In some cases, the cytosine analog resistant to bisulfite treatment is 5-methylcytosine, 5-hydroxymethylcytosine, or 5-propynylcytosine. In some cases, the cytosine analog resistant to bisulfite treatment is 5-methylcytosine. In some cases, the method further comprises amplifying the plurality of amplifying the plurality of oligonucleotide sequence-polynucleotide complexes. In some cases, the amplifying comprises solid-phase nucleic acid amplification to generate a plurality of clusters, wherein each cluster represents a plurality of copies of each of the plurality of oligonucleotide sequence-polynucleotide complexes. In some cases, the method further comprises treating the plurality of oligonucleotide sequence-polynucleotides complexes with bisulfite prior to the amplification, wherein the treating generates a plurality of oligonucleotide sequence-polynucleotide complexes comprising non-complementary ends. In some cases, the sequence complementary to a sequencing primer in one end of the plurality of oligonucleotide sequence-polynucleotide complexes comprising non-complementary ends is complementary to a reverse read sequencing primer, and wherein the sequence complementary to a sequencing primer in an opposite end of the non-complementary

IDT Exhibit 1017.1558

ends in the plurality of oligonucleotide sequence-polynucleotide complexes comprising non-complementary ends is complementary to a forward read sequencing primer. In some cases, the method further comprises sequencing each of the plurality of clusters, wherein the sequencing comprises hybridizing the sequencing primer to the sequence complementary to a sequencing primer, and conducting sequencing by synthesis, wherein the introducing of the pool of oligonucleotide sequences to the polynucleotides improves identification of each of the plurality of clusters generated from the plurality of oligonucleotide sequence-polynucleotide complexes relative to the identification of each cluster from a plurality of clusters generated from the plurality of polynucleotides not introduced to the pool of oligonucleotide sequences. In some cases, the sequence element comprises a barcode, a universal sequence, a linker sequence, or a random sequence.

[0008] In one aspect, provided herein is a kit comprising a pool of oligonucleotide sequences, a. wherein each oligonucleotide sequence in the pool of oligonucleotide sequences comprises a sequence complementary to a sequencing primer, b. wherein the pool of oligonucleotide sequences comprises a first oligonucleotide sequence, wherein a terminal 3' nucleotide base of the sequencing primer is capable of annealing to a terminal 5' nucleotide base of the first oligonucleotide sequence, c. wherein the pool of oligonucleotide sequences comprises a second oligonucleotide sequence, wherein the second oligonucleotide sequence terminates at a 5' end with an additional base relative to the 5' end of the first oligonucleotide sequence; d. wherein the pool of oligonucleotide sequences comprises a third oligonucleotide sequence, wherein the third oligonucleotide sequence terminates at a 5' end with two additional bases relative to the 5' end of the first oligonucleotide sequence; and e. wherein the pool of oligonucleotide sequences comprises a fourth oligonucleotide sequence, wherein the fourth oligonucleotide sequence terminates at a 5' end with three additional bases relative to the 5' end of the first oligonucleotide sequence. In some cases, each oligonucleotide sequence in the pool of oligonucleotide sequences is within a primer, wherein the pool of oligonucleotide sequences comprises the primer and a complement thereof. In some cases, the primer comprises cytosine analogs resistant to bisulfite conversion. In some cases, the primer comprises cytosine bases sensitive to bisulfite conversion. In some cases, each oligonucleotide sequence in the pool of oligonucleotide sequences is within a strand of a duplexed adaptor. In some cases, each duplexed adaptor in the pool comprises a ligation strand and a non-ligation strand. In some cases, the ligation strand and the non-ligation strand of each of duplexed adaptor in the pool comprise cytosine bases sensitive to bisulfite treatment. In some cases, the kit further comprises a mixture of dNTPs comprising dATP, dTTP, dGTP, and a dCTP analog resistant to bisulfite treatment. In

IDT Exhibit 1017.1559

some cases, the ligation strand and the non-ligation strand of each of duplexed adaptor in the pool comprise cytosine analogs resistant to bisulfite treatment. In some cases, the kit further comprises a mixture of dNTPs comprising dATP, dTTP, dGTP, and a dCTP analog sensitive to bisulfite treatment. In some cases, the kit further comprises a restriction enzyme. In some cases, the restriction enzyme is MspI. In some cases, the kit further comprises a DNA polymerase. In some cases, the DNA polymerase comprises exonuclease activity. In some cases, the DNA polymerase comprises strand displacement activity. In some cases, the non-ligation strand comprises a block at a 5' and 3' end, whereby the non-ligation strand is enzymatically unreactive. In some cases, the block at the 5' end comprises a biotin moiety. In some cases, the block at the 3' end comprises a dideoxy residue. In some cases, the cytosine analog resistant to bisulfite treatment is 5-methylcytosine, 5-hydroxymethylcytosine, or 5-propynylcytosine. In some cases, the cytosine analog resistant to bisulfite treatment is 5-methylcytosine.

[0009] In one aspect, provided herein is a method for analyzing sequence reads, the method comprising: a. generating a plurality of sequence reads, wherein the generating comprises conducting sequencing by synthesis on a plurality of clusters, wherein each cluster of the plurality of clusters comprises copies of one oligonucleotide sequence-polynucleotide from a plurality of oligonucleotide sequence-polynucleotides, wherein the sequencing by synthesis comprises hybridizing a sequencing primer to sequence complementary to the sequencing primer present in each cluster of the plurality of clusters, wherein the plurality of clusters are produced from solid-phase nucleic acid amplification of the plurality of oligonucleotide sequence-polynucleotides, wherein each of the plurality of oligonucleotide sequence-polynucleotides comprises a single nucleotide base extender sequence, a dinucleotide base extender sequence, a trinucleotide base extender sequence or no nucleotide base extender sequence between the sequence complementary to the sequencing primer of an oligonucleotide sequence and an end of a polynucleotide; and b. analyzing each of the plurality of sequence reads, wherein the analyzing comprises trimming or removing the extender sequence *in silico* from each of the plurality of sequence reads. In some cases, the plurality of oligonucleotide sequence-polynucloetides are generated by introducing a pool of oligonucleotide sequences to 3' ends of the plurality of polynucleotides from a sample, a. wherein each oligonucleotide sequence in the pool of oligonucleotide sequences comprises a sequence complementary to a sequencing primer, b. wherein the pool of oligonucleotide sequences comprises a first sequence, wherein a terminal 3' nucleotide base of the sequencing primer is capable of annealing to a terminal 5' nucleotide base of the first sequence, wherein following introduction the 5' end of the first sequence is adjacent to a 3' end of a polynucleotide from the plurality of polynucleotides, c. wherein the pool of

IDT Exhibit 1017.1560

oligonucleotide sequences comprises a second sequence, wherein the second sequence terminates at its 5' end with an additional base relative to the 5' end of the first sequence comprising the single nucleotide base extender, wherein following introduction the 5' end of the second sequence is adjacent to a 3' end of a polynucleotide from the plurality of polynucleotides; d. wherein the pool of oligonucleotide sequences comprises a third sequence, wherein the third sequence terminates at its 5' end with two additional bases relative to the 5' end of the first sequence comprising the dinucleotide base extender, wherein following introduction the 5' end of the third sequence is adjacent to a 3' end of a polynucleotide from the plurality of polynucleotides; e. wherein the pool of oligonucleotide sequences comprises a fourth sequence, wherein the fourth sequence terminates at its 5' end with three additional bases relative to the 5' end of the first sequence comprising the trinucleotide base extender, wherein following introduction the 5' end of the fourth sequence is adjacent to a 3' end of a polynucleotide from the plurality of polynucleotides; and f. wherein the introducing generates a plurality of oligonucleotide-polynucleotide complexes, thereby increasing diversity of sequences at ends of the plurality of polynucleotides. In some cases, the sample is a whole genome sample. In some cases, the sample comprises a polynucleotide library. In some cases, the method further comprises fragmenting the polynucleotides prior to introducing the pool of oligonucleotide sequences to the 3' ends of the plurality of polynucleotides. In some cases, the fragmenting is performed by contacting the polynucleotides with an enzyme. In some cases, the enzyme is a restriction enzyme. In some cases, the restriction enzyme is a methylation insensitive restriction enzyme. In some cases, the methylation insensitive restriction enzyme is MspI. In some cases, the additional base of the second sequence is an adenine, thymine, or cytosine base. In some cases, each of the two additional bases of the third sequence is an adenine, thymine, or cytosine base. In some cases, a 3' most base of the three additional bases of the fourth sequence comprises a thymine or cytosine base, and each of the two other bases of the three additional bases of the fourth sequence comprises an adenine, thymine, or cytosine base. In some cases, each oligonucleotide sequence in the pool of oligonucleotide sequences is within a strand of a duplexed adaptor. In some cases, the duplexed adaptor is a partial duplexed adaptor comprising a long strand and a short strand. In some cases, the short strand is a non-ligation strand of the duplexed adaptor and the long strand is a ligation strand. In some cases, the non-ligation strand comprises a 5' overhang. In some cases, the 5' overhang comprises two bases. In some cases, the two bases are 3'-GC-5'. In some cases, the introducing comprises a ligation reaction. In some cases, the ligation reaction comprises cohesive end ligation. In some cases, the ligation reaction comprises blunt end ligation. In some cases, the plurality of polynucleotides are strands of

double-stranded nucleic acids, wherein each 5' end of the double-stranded nucleic acids comprises a ligation end and each 3' end comprises a non-ligation end, wherein each of the non-ligation endsof the double-stranded nucleic acid comprises part of the polynucleotide sequence. In some cases, the 3'-GC-5' of the non-ligation strand of the duplexed adaptor anneals to a 5'-CG-3' sequence of the ligation end of the double-stranded sequence comprising the polynucleotide sequence. In some cases, the 3'-GC-5' of the non-ligation strand of the duplexed adaptor is 5' to the terminal 5' nucleotide base of the first sequence. In some cases, the 3'-GC-5' of the non-ligation strand of the duplexed adaptor is 5' to the additional base of the second sequence relative to the 5' end of the first sequence. In some cases, the 3'-GC-5' of the non-ligation strand of the duplexed adaptor is 5' to the two additional bases of the third sequence relative to the 5' end of the first sequence. In some cases, the 3'-GC-5' of the non-ligation strand of the duplexed adaptor is 5' to the three additional bases of the fourth sequence relative to the 5' end of the first sequence. In some cases, each duplexed adaptor in the pool comprises a ligation strand and a non-ligation strand, wherein the ligation strand is capable of ligating to a 5' end of the plurality of polynucleotides, and wherein a nick is formed between the non-ligation strand and the 3' end of the plurality of polynucleotides. In some cases, the method further comprises an extending step, wherein the 3' end of the plurality of polynucleotides is extended using a polymerase in the presence of a mixture of dNTPs using the ligation strand of the adaptor as template, thereby introducing the pool of oligonucleotide sequences to the 3' ends of the plurality of polynucleotides. In some cases, the ligation strand of each adaptor duplex in the pool comprise cytosine analogs resistant to bisulfite treatment and the mixture of dNTPs comprises dATP, dTTP, dGTP, and dCTP, whereby the extending generates the plurality of oligonucleotide-polynucleotide complexes, wherein each of the plurality of oligonucleotide-polynucleotide complexes comprises complementary ends, wherein one end comprises cytosine bases sensitive to bisulfite treatment, while the opposite end comprises cytosine analogs resistant to bisulfite treatment. In some cases, the ligation strand of each adaptor duplex in the pool of adaptor duplexes comprise cytosine bases sensitive to bisulfite treatment and the mixture of dNTPs comprises dATP, dTTP, dGTP, and a dCTP analog resistant to bisulfite treatment, whereby the extending generates the plurality of oligonucleotide-polynucleotide complexes, wherein each of the plurality of oligonucleotide-polynucleotide complexes comprises complementary ends, wherein one end comprises cytosine bases sensitive to bisulfite treatment, while the opposite end comprises cytosine analogs resistant to bisulfite treatment. In some cases, each oligonucleotide in the pool of oligonucleotides is a primer, wherein the pool of oligonucleotides comprises the primer and a complement thereof. In some cases, the introducing

IDT Exhibit 1017.1562

comprises an amplification reaction. In some cases, the amplification reaction comprises polymerase chain reaction. In some cases, each of the primers in the pool comprises cytosine bases sensitive to bisulfite treatment, wherein the amplification is conducted in the presence of a mixture of dNTPs comprising dATP, dTTP, dGTP, and a dCTP analog resistant to bisulfite treatment, thereby generating the plurality of oligonucleotide sequence-polynucleotide complexes, wherein each of the plurality of oligonucleotide sequence-polynucleotide complexes comprises complementary ends, wherein one end comprises cytosine bases sensitive to bisulfite treatment, while the opposite end comprises cytosine analogs resistant to bisulfite treatment. In some cases, each of the primers in the pool of primers comprises cytosine analogs resistant to bisulfite treatment, wherein the amplification is conducted in the presence of a mixture of dNTPs comprising dATP, dTTP, dGTP, and dCTP, thereby generating the plurality of oligonucleotide sequence-polynucleotide complexes, wherein each of the plurality of oligonucleotide sequence-polynucleotide complexes comprises complementary ends, wherein one end comprises cytosine bases sensitive to bisulfite treatment, while the opposite end comprises cytosine analogs resistant to bisulfite treatment. In some cases, the cytosine analog resistant to bisulfite treatment is 5-methylcytosine, 5-hydroxymethylcytosine, or 5-propynylcytosine. In some cases, the method further comprises treating the plurality of oligonucleotide sequence-polynucleotides complexes with bisulfite prior to the amplification, wherein the treating generates a plurality of oligonucleotide sequence-polynucleotide complexes comprising non-complementary ends. In some cases, the treating with bisulfite prior to amplification generates the sequence complementary to a sequencing primer in each oligonucleotide sequence in the pool of oligonucleotide sequences comprises. In some cases, the generating the plurality of sequence reads comprises paired-end sequencing.

[0010] In one aspect, provided herein is a method for increasing diversity of sequences at ends of a plurality of polynucleotide fragments in a library, the method comprising: a. fragmenting polynucleotides to generate polynucleotide fragments; b. ligating a pool of adaptor sequences to 3' ends of the plurality of polynucleotide fragments, wherein each adaptor sequence in the pool of adaptor sequences comprises a sequence complementary to a sequencing primer, wherein a 3' end of the sequencing primer anneals at least four bases from the 5' end of the adaptor sequence, and wherein each of four different types of nucleotide bases are present in each of the at least three terminal nucleotides at the 5' ends of the adaptors, thereby increase diversity of sequences at ends of the plurality of polynucleotide fragments.

IDT Exhibit 1017.1563

INCORPORATION BY REFERENCE

[0011] 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

[0012] The novel features described herein are set forth with particularity in the appended claims. A better understanding of the features and advantages described herein will be obtained by reference to the following detailed description that sets forth illustrative embodiments, in which the principles described herein are utilized, and the accompanying drawings of which:

[0013] **FIG. 1** illustrates the basic principle of RRBS. Nucleic acid Library inserts (or genomic DNA) is digested with MspI, bisulfite converted (a chemical reaction that converts non-methylated C to U, but leaves 5-mC unreacted), and sequenced. C in the sequence indicates the presence of 5-mC at that position. Note how the first base of every sequence contains a methylation measurement.

[0014] **FIG. 2A-B** illustrates a method for generating the diversity adapters used in the methods provided herein. **FIG. 2A** illustrates a long strand of nucleotides (R01446) and 4 different short strands (i.e., Ben383, Ben384, Ben385, and Ben341) such that each short strand comprises a stretch of 12 nucleotides that comprise bases complementary to the long strand (R01446). **FIG. 2B** illustrates the method used to generate a diversity adaptor using the long strand (R01446) and a representative short strand (Ben383) from **FIG. 2A.**

[0015] **FIG. 3** illustrates an alternative method for generating an RRBS library.

[0016] **FIG. 4A-B** illustrates the sequencing results of forward reads (FIG. 4A) and reverse reads (FIG. 4B) for sequencing an RRBS library generated using diversity-adapters comprising 3 "DDD" bases.

[0017] **FIG. 5A-B** illustrates the sequencing results of forward reads (FIG. 5A) and reverse reads (FIG. 5B) for sequencing an RRBS library generated using diversity-adapters comprising 2 "DD" bases.

[0018] **FIG. 6A-B** illustrates the sequencing results of forward reads (FIG. 6A) and reverse reads (FIG. 6B) for sequencing an RRBS library generated using diversity-adapters comprising a single "D" base.

[0019] **FIG. 7A-B** illustrates the sequencing results of forward reads (FIG. 7A) and reverse reads (FIG. 7B) for sequencing an RRBS library generated using diversity-adapters comprising no "D" bases.

[0020] **FIG. 8** illustrates expected sequencing reads (forward and reverse reads) for sequencing an RRBS library with a pool of diversity-adapters comprising a variable number of "D" bases.

IDT Exhibit 1017.1564

The expected forward or reverse reads can be from N to NNNNNN bases in the first 6 sequencing cycles, wherein an N base can be an H, D, Y or B base.

[0021] FIG. 9A-B illustrates the sequencing results of forward reads (FIG. 9A) and reverse reads (FIG. 9B) for sequencing an RRBS library with a pool of diversity-adapters comprising a variable number of "D" bases.

[0022] FIG. 10 illustrates the sequencing results for sequencing an RRBS library with a 5-plex pool of diversity-adapters comprising a variable number of "D" bases, wherein the sequencing was performed without using a PhiX control as described in Example 2.

[0023] FIG. 11 illustrates the % base for each cycle for the sequencing reactions performed in **FIG. 10**.

[0024] FIG. 12 illustrates the % >Q20 score for each cycle for the sequencing reactions performed in **FIG. 10**.

[0025] FIG. 13 illustrates a Q-score heat map for the sequencing reactions performed in **FIG. 10**.

[0026] FIG. 14 illustrates the sequencing results for sequencing a blood sample using a 16 plex sequencing system (e.g., Ovation® Blood Seq Library System) as described in Example 2.

[0027] FIG. 15 illustrates a Q-score heat map for the sequencing reactions performed in **FIG. 14.**

[0028] FIG. 16 illustrates a run summary for the sequencing reactions performed in **FIG. 10.**

[0029] FIG. 17 illustrates a run summary for the sequencing reactions performed in **FIG. 14.**

[0030] FIG. 18A-C illustrates proposed rules for trimming diversity "D" bases *in silico* from sequence reads from nucleic acid inserts generated with mixed diversity-Adaptor sequences comprising 0 to 3 "D" bases (D0-D3). **FIG. 18A** illustrates a first rule. **FIG. 18B** illustrates a second rule. **FIG. 18C** illustrates a third rule.

[0031] FIG. 19 illustrates the total theoretical diversity possible for introducing to the ends of a polynucleotide upon introduction of a pool of diversity adaptors comprising diversity-adapters comprising 0 to 3 "D" bases (D0-D3).

[0032] FIG. 20 illustrates total CpGs analyzed using the modified RRBS methods provided herein in comparison to Whole Genome Bisulfite Sequencing (WGBS).

[0033] FIG. 21 illustrates a comparison of the average CpG per read and average CpG per uniquely aligned read between the modified RRBS method as provided herein vs. WGBS.

[0034] FIG. 22 illustrates the feature coverage using the modified RRBS as provided herein.

IDT Exhibit 1017.1565

[0035] FIG. 23 illustrates bioanalyzer traces of MspI-digested human genomic DNA (left), and the resulting RRBS sequencing library (top right) including an example of the nucleotide distribution (Forward Reads) obtained using said method (bottom right).

[0036] FIG. 24 illustrates a comparison between RRBS and other DNA methylation profiling technologies.

[0037] FIG. 25 illustrates a comparison of the *in silico* assignment of CpGs into classes for RRBS vs. whole genome bisulfite sequencing.

[0038] FIG. 26 illustrates an example of a dark sequencing approach used that can be used separately or in conjunction with RRBS.

[0039] FIG. 27A illustrates a polynucleotide insert from an RRBS generated using diversity adapters as provided herein. **FIG. 27B** illustrates the increased color balance in the initial sequence reads of an RRBS library generated using diversity adapters as provided herein and comprising polynucleotide inserts as shown in **FIG. 27A**.

[0040] FIG. 28 illustrates library size distribution on Bioanalyzer High Sensitivity DNA Chip.

[0041] FIG. 29 illustrates determination of the number of PCR cycles to perform during the library amplification.

[0042] FIG. 30 illustrates results from selecting Plot > deltaRn vs. Cycle

[0043] FIG. 31 illustrates cluster identification requires sequence diversity. The Illumina software identifies clusters over the first several cycles of sequencing. During sequencing of normal, high diversity clusters (top), overlapping clusters can be distinguished because they are different colors. If overlapping clusters contain the same sequence (bottom), theycan be mistaken as a single cluster.

[0044] FIG. 32 illustrates random N6 sequence adjacent to the barcode that can be used to mark PCR duplicates for *in silico* removal.

[0045] FIG. 33 illustrates high coverage overlap between three NuGEN RRBS technical replicates, each made from 25 ng of IMR90 gDNA. CpG's covered at 10X or greater are shown, after PCR duplicate removal.

[0046] FIG. 34 illustrates concordance in methylation levels. Concordance for CpG's covered at 20X or greater depth between NuGEN RRBS technical replicates (left) or between NuGEN RRBS (30 M reads) and published whole genome bisulfite sequencing (WGBS) from Lister et al., *Nature*, 2009, 462:315–322 (1180 M reads), which is herein incorporated by reference in its entirety.

[0047] FIG. 35 illustrates types of sequence diversity added to the beginning of forward reads in RRBS libraries generated using diversity adaptors with 0, 1, 2, or 3 diversity "D" bases.

IDT Exhibit 1017.1566

[0048] FIG. 36 illustrates the process of generating an RRBS library using MspI digestion and diversity adaptors lacking diversity bases.

[0049] FIG. 37 illustrates a schematic of a nucleic acid insert from an RRBS library generated using the library preparation process provided herein

[0050] FIG. 38 illustrates an overview of the RRBS library preparation process described in Example 6.

DETAILED DESCRIPTION

I. Overview

[0051] Provided herein are methods, compositions, and kits for introducing sequence diversity into genomic regions comprising low sequence diversity. The genomic regions comprising low sequence diversity can be highly conserved genomic regions between organisms (e.g., 16s rRNA sequences between microbial species). The genomic regions comprising low sequence diversity can be repetitive sequences within a genome (e.g., CpG islands within mammalian genomes). The methods described herein can introduce sequence diversity into genomic regions comprising low sequence diversity by introducing oligonucleotide sequence to one or both ends of a polynucleotide. The introduction can be through ligation (e.g., blunt-end or cohesive-end) using adaptors such that the adaptors comprise the oligonucleotide sequence introduced to one or both ends of the polynucleotide. The introduction can through amplification of a target sequence in a polynucleotide (e.g., genomic region) using primers such that the primers comprise the oligonucleotide sequence introduced to one or both ends of the polynucleotide. The primers can comprise sequence complementary to a target sequence in a nucleic acid as well as the oligonucleotide sequence. The oligonucleotide sequence can be present in the primer as a tail such that it is non-complementary to the target sequence. In general, the oligonucleotide sequence can comprise sequence complementary to a sequencing primer, one or more barcodes, universal sequence, random sequence, and/or sequence useful for introducing sequence diversity at one or both ends of a polynucleotide (e.g., extender sequence as provided herein). The oligonucleotide sequence can be configured such that sequence complementary to a sequencing primer is immediately adjacent to the end of the polynucleotide upon introduction (i.e., through amplification or ligation) of the oligonucleotide sequence complementary to a sequencing primer. In some cases, the oligonucleotide sequence further comprises an extender sequence such that the sequence complementary to a sequencing primer precedes the extender sequence such that, upon introduction, the extender sequence is immediately adjacent to the end of the polynucleotide.

IDT Exhibit 1017.1567

[0052] In the methods, compositions and kits provided herein, the extender sequence in an oligonucleotide sequence as described herein can be a variable number of nucleotides in length. The extender sequence can be less than, more than, or about 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 nucleotides in length. In some cases, the extender sequence is a single nucleotide in length. In some cases, the extender sequence is two nucleotides in length. In some cases, the extender sequence is three nucleotides in length. In some cases, an additional sequence element is present in the oligonucleotide sequence such that the additional sequence is located between the sequence complementary to a sequencing primer and the extender sequence such that upon introduction the extender sequence is immediately adjacent to the end of the polynucleotide and is followed by the additional sequence element which is followed by the sequence complementary to a sequencing primer. In some cases, the oligonucleotide sequence can comprise one or more additional sequence elements such that the one or more additional sequence elements are not located between the sequence complementary to a sequencing primer and an extender sequence such that the one or more additional sequence elements precedes the sequence complementary to a sequencing primer. In this case, upon introduction of the oligonucleotide sequence to one or both ends of a polynucleotide, the extender sequence is immediately adjacent to the end of the polynucleotide and is followed by the sequence complementary to a sequencing primer which is followed by the one or more additional sequence elements. In some cases, the oligonucleotide sequence can comprise a plurality of additional sequence elements such that one of the plurality of additional sequence elements is located between the sequence complementary to a sequencing primer and the extender sequence while one or more additional sequence elements of the plurality of additional sequence elements are not located between the sequence complementary to a sequencing primer and an extender sequence. The one or more additional sequence elements of the plurality of additional sequence elements precede the sequence complementary to a sequencing primer. In this case, upon introduction of the oligonucleotide sequence to one or both ends of a polynucleotide, the extender sequence is immediately adjacent to the end of the polynucleotide and is followed in succession by one of the plurality of additional sequence elements, the sequence complementary to a sequencing primer and the one or more additional sequence elements of the plurality of additional sequence elements.

[0053] An additional sequence element can be one or more barcodes, universal adaptor sequence, linker sequence, and/or random sequence. The sequencing primer can be a custom sequencing primer or can be a sequencing primer compatible with a commercially available NGS sequencing system.

IDT Exhibit 1017.1568

[0054] The polynucleotides can be in a sample. The sample can be any sample as provided herein. The polynucleotides can be fragmented prior to the introduction of an oligonucleotide sequence as described herein. Fragmentation can be mechanical (e.g., sonication) or chemical (e.g., enzymatic) as provided herein.

[0055] Following introduction of the oligonucleotide sequence to the ends of a polynucleotide, the resulting oligonucleotide sequence-polynucleotide complex can be amplified. In some cases, oligonucleotide sequence is introduced to both ends of a polynucleotide, and subsequent amplification is performed using primers directed to complementary sequence in the oligonucleotide sequences introduced to the ends of the polynucleotide. The primers used for amplification can introduce further sequence elements to the oligonucleotide sequence-polynucleotide complex formed after introduction of the oligonucleotide sequence. The sequence elements introduced during amplification can be barcode(s) or sequence complementary to a capture probe on a solid surface. The solid surface can be a bead or planar surface (e.g., flow cell). The amplification can be a solid-state amplification (e.g., cluster amplification) that generates a cluster comprising a plurality of copies of the oligonucleotide sequence-polynucleotide complex. The methods provided herein can further comprise sequencing the amplified oligonucleotide-polynucleotide complexes. An extender sequence can be a barcode, universal sequence, linker sequence, or random sequence.

[0056] In the methods, compositions and kits provided herein, the extender sequence in an oligonucleotide sequence as described herein can comprise any of the four canonical nucleotide bases (i.e., thymine, adenine, cytosine, or guanine) found in DNA. The nucleotide base composition of a nucleotide in an extender sequence can be selected such that the nucleotide base composition can differ from the nucleotide base composition at an end of a polynucleotide to which the extender sequence can be introduced, whereby the presence of one or more nucleotides from the extender sequence at the end of the polynucleotide can serve to increase sequence read length as well as nucleotide base diversity. The increase based diversity can aid identification of individual clusters as well as maintain accurate label detection in NGS sequencing systems that use cluster based sequencing by synthesis (e.g., Illumina).

[0057] The methods, compositions, and kits provided herein can comprise a pool of oligonucleotide sequences as described herein. The pool of oligonucleotide sequences can be a pool of adaptors comprising the oligonucleotide sequence as described herein. The pool of adaptors comprising the oligonucleotide sequence can be introduced to one or both ends of a plurality of polynucleotides through ligation (e.g., blunt end or cohesive end). The pool of oligonucleotide sequences can be a pool of primers comprising the oligonucleotide sequence as

IDT Exhibit 1017.1569

described herein. The pool of primers comprising the oligonucleotide sequence can be introduced to one or both ends of a plurality of polynucleotides through amplification (e.g., isothermal or PCR). The pool of oligonucleotide sequences can comprise sets of oligonucleotide sequences such that each set comprises either no extender sequence or an extender sequence of a length that differs from the length of an extender sequence in a separate set of oligonucleotide sequences within the pool of oligonucleotide sequences. The length of an extender sequence in a set of oligonucleotide sequences in a pool of oligonucleotide sequences can be 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more nucleotides.

[0058] In some cases, a pool of oligonucleotide sequences comprises a first set of oligonucleotide sequences that do not comprise an extender sequence, a second set of oligonucleotide sequences that comprise an extender sequence comprising a single nucleotide, a third set of oligonucleotide sequences that comprise an extender sequence comprising two nucleotides, and a fourth set of oligonucleotide sequences that comprise an extender sequence comprising three nucleotides. Each oligonucleotide sequence within the pool of oligonucleotide sequences can comprise the same sequence complementary to a sequencing primer as described herein. The nucleotide base composition of each nucleotide in an extender sequence comprising more than one nucleotides can be identical or different. The nucleotide base composition of each nucleotide in an extender sequence comprising more than one nucleotide can be known or randomly assigned as provided herein.

[0059] In some cases, the methods, compositions, and kits provided herein are useful for generating bisulfite converted nucleic acid libraries with increased sequence diversity at the ends of nucleic acid inserts within the library. The bisulfite converted libraries can be reduced representation bisulfite sequencing (RRBS) libraries. Such libraries can be useful, for example, for determining the methylation status across a genome, or alternatively, for determining the methylation status at given genomic loci. The RRBS libraries generated by the methods, compositions, and kits provided herein can be useful for determining the methylation status of CpG islands in polynucleotides derived from whole genome samples. The whole genome samples can be from eukaryotic cells (e.g., mammalian cells). The methods can produce RRBS libraries that can be more efficiently sequenced by current NGS sequencers (e.g., Illumina sequencers), and in some embodiments, without the need for spiking in a genomic, higher-diversity sample nucleic acids (e.g. PhiX control library). The method for generating an RRBS library can comprise enzymatically fragmenting a nucleic acid (polynucleotide) such that a plurality of nucleic acids fragments comprising substantially similar nucleotide compositions at the ends or termini of each fragment are generated. In some cases, the method comprises

IDT Exhibit 1017.1570

fragmenting the nucleic acid with a methylation sensitive restriction enzyme such as, for example, MspI. In some cases, a pool of adaptors is ligated to the 5' ends of each nucleic acid fragment in the plurality of nucleic acid fragments such that each nucleic acid fragment comprises double stranded sequence and the ligation generates a plurality of adaptor-ligated nucleic acid fragments. The pool of adaptors can comprise adaptors that share a sequence complementary to a sequencing primer as well as a first set of adaptors comprising no extender sequence, a second set comprising a single nucleotide extender sequence, a third set comprising a dinucleotide extender sequence, and a fourth set comprising a trinucleotide extender sequence. In some cases, the nucleotide base composition of the extender sequences in a pool of adaptors comprising extender sequence are chosen such that the base composition differs for the nucleotide base composition at the end of the nucleic fragment to which the extender sequence is appended. In some cases, the single nucleotide in first set of adaptors can comprises a thymine, guanine, or adenine base, while the each nucleotide in the dinucleotide of the second set can comprise a thymine, guanine, or adenine base and two of the three nucleotides in the trinucleotide of the third set can be a thymine, guanine, or adenine base, while the third nucleotide can be guanine or adenine. As provided herein, the extender sequence can be used to increase sequence diversity between the sequencing primer and the end of a library nucleic acid insert. In some cases, the adaptor-ligated nucleic acid fragments are subjected to an extension reaction as provided herein followed by denaturation and bisulfite treatment, whereby the bisulfite treatment generates adaptor-ligated nucleic acid fragment comprising non-complementary adaptor sequence on opposite ends. The non-complementary adaptor sequences can subsequently be utilized to amplify and/or sequence each adaptor-ligated nucleic acid fragment. The amplification can be solution phase and/or solid-state amplification. In some cases, the plurality of adaptor-ligated nucleic acid fragments are PCR amplified followed by solid-state amplification to generate clusters such that each cluster comprises thousands of copies of an adaptor-ligated nucleic acid fragment form the plurality of adaptor-ligated nucleic acid fragments. The presence of extender sequence in the plurality of adaptor-ligated nucleic acid fragments can increase the ability of NGS sequencers (e.g., Illumina) to successfully identify and discern between neighboring clusters formed by solid-state amplification during sequencing of the clusters (e.g., as illustrated in **FIG. 31**).

[0060] Provided herein are methods for generating adaptors comprising extender sequences for use in the generation of bisulfite converted nucleic acid libraries (e.g., RRBS libraries). Adaptors comprising extender sequences can also be referred to as diversity adaptors.

IDT Exhibit 1017.1571

[0061] Provided herein are *in silico* methods for filtering and trimming the extender sequence present in sequences reads produced from sequencing nucleic acids or polynucleotides comprising extender sequence as generated through the methods provided herein. The *in silico* methods can comprise trimming away any oligonucleotide sequence (e.g., adaptor sequence) that may be present on the 3' ends of reads using publically available trimming algorithms (e.g., Trim Galore). In some cases, the trimming script or algorithm can remove adaptor sequence such as adaptor sequence associated with a specific sequencing platform (e.g., Illumina adaptor sequence) and/or any nucleotide bases whose quality score is below a set quality score (Q-score) threshold. In some cases, the Q-score threshold is set at 30, 20, or 10 such that any Q-score below said thresdold is trimmed or removed from analysis by the script or algorithm. Once the oligonucleotide sequence (e.g., adaptor sequence) is trimmed away, a custom python script can remove any reads that do not contain the known end sequence of the library insert. In RRBS libraries generated through fragmentation via MspI digestion, the python script can remove any reads that do not contain an MspI site signature, YGG (Y= cytosine of thymine bases), at the 5' end. For paired end data an MspI site signature can be required at 5' ends of both sequences. Following filtering and trimming, all reads should begin with YGG, where Y is C or T, 5 bases can be trimmed from every read (6 bases are trimmed for paired-end to prevent alignment issues). If YGG is not found in the first 6 bases, the read is discarded.

[0062] Provided herein are methods for identifying and removing duplicate sequence reads following NGS sequencing. In some cases, an oligonucleotide sequence introduced to one or both ends of a polynucleotide comprises additional sequence elements upstream of a sequence complementary to a sequencing primer relative to an end of a polynucleotide to which the oligonucleotide sequence has been appended. In some cases, the additional sequence elements comprise a barcode and a stretch of random nucleotides immediately adjacent thereto. In some cases, the stretch of random nucleotides is 6 nucleotides in length. The methods provided herein can utilize an *in silico* analysis to discriminate between true duplicates produced during amplification (e.g., PCR). The analysis can utilize *in silico* algorithms in order to utilize information provided by the unique stretch of random nucleotides to discriminate between true PCR duplicates and independent adaptor ligated-nucleic acid fragments. In some cases, the algorithm for removing duplicate sequence reads uses sequence reads recovered from an index read. The index read can provide sequence data or reads comprising the barcode(s) and the random sequence located adjacent to the barcode(s). In order to read the inserted random sequence adjacent to the barcode(s) during an index read, the cycle number of the index read can be increased by the number of nucleotide bases present in the inserted random sequence. For

IDT Exhibit 1017.1572

example, the barcode can be 6 nucleotides and the random sequence can be 6 nucleotides and the index read can be a total of 12 cycles.

II. Adaptors

[0063] In one aspect provided herein is a method for increasing sequence diversity at the ends of polynucleotides that comprise conserved sequence at one or both ends with other polynucleotides using adaptors comprising extender sequence as described herein. The adaptor comprising an extender sequence can be used in the methods described herein to increase the sequence diversity at one or both ends of a polynucleotide by appending the adaptor comprising the extender sequence to one or both ends of the polynucleotide. The increased sequence diversity can aid an NGS sequencer in determining the sequence of each polynucleotide in a group of polynucleotides that share identical or substantially similar nucleotide base compositions at their end or ends when conducting NGS sequencing using NGS methods and systems that generate and identify clusters of individual polynucleotides during sequencing (e.g., Illumina, SOLiD). As a result, an adaptor comprising extender sequence as described herein can be referred to as a diversity adaptor and each nucleotide and its corresponding nucleotide base can be referred to as a diversity base.

[0064] **FIG. 35** illustrates the DNA sequence derived from a bisulfite converted library constructed using diversity adaptors such that following introduction of the diversity adaptors to the ends of DNA polynucleotides, the library comprises polynucleotides with either no diversity bases (i.e., 0 diversity bases added) derived from the diversity adaptor, one diversity base (i.e., 1 diversity base added) derived from the diversity adaptor, two diversity bases (i.e., 2 diversity bases added) derived from the diversity adaptor, or three diversity bases (i.e., 3 diversity "3" bases added) derived from the diversity adaptor. The number of diversity bases added can be variable and can be tailored to the type of library being constructed (e.g., 16S rRNA libraries, RRBS libraries, etc.). The diversity bases in a diversity adaptor can be "D" bases such that a "D" base can be an adenine, thymine, or guanine such as in FIGs. 2A-B, 3, 4A-B, 5A-B, 6A-B, 7A-B, 8, 9A-B, 18A-B and 35. The diversity bases in a diversity adaptor can be "R" bases such that an "R" base can be an adenine or guanine such as in FIG. 18B and 35. The complement of a "D" or an "R" base can be an "H" base. In some cases, an "H" base is a complement of a "D" base such that the "H" base can be a thymine when the "D" base is an adenine, an adenine when the "D" base is a thymine, or a cytosine when the "D" base is guanine. In some cases, an "H" base is a complement of an "R" base such that the "H" base can be a thymine when the "R" base is an adenine or a cytosine when the "R" base is guanine. In some cases, each diversity base in a diversity adaptor is a "D" base. In some cases, each base in a diversity adaptor is an "R" base.

IDT Exhibit 1017.1573

Each diversity base in a diversity adaptor comprising more than one diversity base can be either a "D" base or an "R" base. In some cases, a pool of diversity adaptors is used in a method herein, wherein the pool comprises diversity adaptors comprising no diversity bases, diversity adaptors comprising 1 "D" base, diversity adaptors comprising 2 "D" bases, and diversity adaptors comprising 1 R base and 2 D bases. The diversity adaptor with 1 R base and two D bases can have the R base at the 5', middle, or 3' position.

[0065] In some cases, the diversity bases can be added to the ends of target polynucleotides as shown in **FIG. 35**, using primers comprising said diversity bases or complements thereof instead of adaptors, whereby the introduction is performed using polymerase extension reactions (e.g., PCR). The primers can be designed to target specific sequences in a polynucleotide such as specific genomic regions (e.g., 16S rRNA sequences in microbial species) such that the primers comprise tails that comprise the diversity bases (i.e., extender sequences) or complements thereof. The adaptors comprising diversity base(s) (i.e., extender sequence) can be designed such that the diversity base or bases are immediately adjacent to sequence derived from a polynucleotide or polynucleotide insert from a library. In other words, the diversity base or bases are the first non-target sequence polynucleotide base or bases present at the junction of a complex comprising a target sequence from a polynucleotide and non-target sequence derived sequence. Also as provided herein, the adaptor comprising diversity bases can comprise a sequence or complement thereof that is capable of hybridizing a sequencing primer and this sequence or complement thereof can be located adjacent to the diversity base or bases such that the diversity base or bases are the bases initially sequenced in a sequencing reaction. In some cases, there can also be additional sequence elements located between the sequence or complement thereof that is capable of hybridizing a sequencing primer and the diversity base or bases such as barcodes (e.g., inline barcodes), universal sequence, or random sequence. In some cases, the additional sequence elements (e.g., barcode(s), universal sequence and/or random sequence) can be located after the sequence or complement thereof that is capable of hybridizing a sequencing primer and distal to the diversity base or bases. In some cases, the additional sequence elements (e.g., barcode(s), universal sequence and/or random sequence) can be located after the sequence or complement thereof that is capable of hybridizing a sequencing primer and distal to the diversity base or bases as well as between the sequence or complement thereof that is capable of hybridizing a sequencing primer and the diversity base or bases such as barcodes (e.g., inline barcodes), universal sequence, or random sequence. An example of a polynucleotide library insert generated to possess one or more additional sequence elements after the sequence that is capable of hybridizing a sequencing primer and therefore distal to the end of the

IDT Exhibit 1017.1574

polynucleotide library insert is shown in **FIGs. 32** and **37**. The insert generated in **FIGs. 32** and **37** can be generated by introducing an adaptor or primer as provided herein.

[0066] As provided herein, the polynucleotide to which the adaptor attached can be a nucleic acid fragment generated using any fragmentation method known in the art. The fragmentation method can be one described herein such as enzymatic fragmentation. The nucleic acid fragment can be double stranded and the adaptor can be attached to both ends of the double stranded nucleic acid fragment. The term "adaptor", as used herein, can refer to an oligonucleotide of known sequence or an oligonucleotide comprising a portion or portions 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 can be envisioned. Various ligation processes and reagents are known in the art and can be useful for carrying out the methods described herein. For example, blunt end ligation can be employed. Similarly, a single dA nucleotide can be added to the 3′-end of a double-stranded, blunted DNA product, by a polymerase lacking 3′-exonuclease activity and can anneal to an adaptor 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 known in the art and kits and reagents for carrying out efficient ligation reactions are commercially available (e.g, from New England Biolabs, Roche).

[0067] In some embodiments, an adaptor comprises a duplex adaptor. In some cases, an adapter comprises a partial duplex adaptor. The partial duplex adaptor can be a forked adaptor. The forked adaptor can be a forked adaptor available from a commercial source and comprise known sequence. The forked adaptor can be compatible with specific NGS sequencing systems that utilize cluster amplification (e.g, Illumina forked adaptors for Illumina based sequencing by synthesis). The forked adaptors can comprise a duplex region where a first strand and a second strand share complementary sequence. The duplex region can be of varying length. The duplex region can be 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30 or more nucleotides in length. Outside of the duplex region, the forked adaptor can have single stranded non-complementary regions such that the first strand comprises a 5' single stranded region that is non-complementary to a 3' single-stranded region of the second strand. In some cases, one or more cytosine bases in the forked adaptor can be substituted with a cytosine analog resistant to bisulfite treatment. These types of forked adaptors can be useful in the generation of RRBS libraries using the methods provided herein. In some cases, the cytosine analog is 5-methylcytosine. In some cases, the cytosine analog is 5-hydroxymethylcytosine. In some cases, the cytosine analog is 5-

IDT Exhibit 1017.1575

propynylcytosine. The 3' and 5' ends of the duplex region of a forked adaptor used in the methods, compositions, and kits herein can comprise no or one or more bases of extender sequence (i.e., diversity bases) as described herein or complements thereof. In some cases, ligation of a forked adaptor as provided herein to both ends of a double stranded polynucleotide can generate forked adaptor-ligated polynucleotide complexes comprising non-complementary adaptor sequence on the 5' and 3' ends of each strand of the double stranded polynucleotide. The non-complementary sequence can be useful for further manipulation and downstream processing of the ligated polynucleotide complexes (e.g., PCR amplification, solid-state amplification, and/or sequencing).

[0068] In some cases, the partial duplex adaptor useful in the methods, compositions and kits provided herein can comprise a long strand and a short strand wherein the short strand comprises sequence complementary to the long strand. The short strand can hybridize completely with the long strand. The short strand can hybridize to a portion of the long strand, and thereby comprise either a 5' or 3' overhang. In some cases, an adaptor for use in the methods provided herein comprises a partial duplex adaptor as shown in **FIGs 2A-B** or **3**. In some cases, an adapter for use in the methods described herein comprises a partial duplex adapter wherein the long strand comprises adaptor sequence comprising one or more sequence elements, while the short strand comprises sequence complementary to a portion of the sequence of the long strand as well as an overhang complementary to sequence in an overhang found in a nucleic acid following cleavage of said nucleic acid.

[0069] The nucleic acid can be cleaved by a restriction endonuclease (RE). The RE can be any restriction endonuclease described herein. In some cases, the RE is MspI.

[0070] In some cases, an adaptor for use in the methods provided herein comprises a ligation strand and a non-ligation strand. In some cases, an adaptor for use in the methods provided herein comprises a partial duplex wherein the long strand comprises a ligation strand, and a short strand comprises a non-ligation strand.

[0071] In some cases, an adaptor (e.g., diversity adaptor) for use in the methods provided herein comprises one or more identifier sequences. The one or more identifier sequences can be present in a long strand and/or a short strand of a partial duplex diversity adaptor as provided herein. The one or more identifier sequences can be present in a ligation strand and/or a non-ligation strand of a partial duplex diversity adapter as provided herein. The identifier sequence can be a barcode sequence, a flow cell sequence, and/or an index sequence. In some cases, the index sequence is a Truseq primer sequence compatible with the NGS platform produced by Illumina. In some cases, the 3' and/or 5' ends of the non-ligation strand comprise a blocking group or

IDT Exhibit 1017.1576

moiety and are enzymatically unreactive. The blocking group or moiety can be a dideoxynucleotide (ddCMP, ddAMP, ddTMP, or ddGMP), various modified nucleotides (e.g. phosphorothioate-modified nucleotides), or non-nucleotide chemical moieties. In some cases, the blocking group comprises a nucleotide analog that comprises a blocking moiety. The blocking moiety can mean a part of the nucleotide analog that inhibits or prevents the nucleotide analog from forming a covalent linkage to a second nucleotide or nucleotide analog. For example, in the case of nucleotide analogs having a pentose moiety, a reversible blocking moiety can prevent formation of a phosphodiester bond between the 3′ oxygen of the nucleotide and the 5′ phosphate of the second nucleotide. Reversible blocking moieties can include phosphates, phosphodiesters, phosphotriesters, phosphorothioate esters, and carbon esters. In some cases, a blocking moiety can be attached to the 3′ position or 2′ position of a pentose moiety of a nucleotide analog. A reversible blocking moiety can be removed with a deblocking agent. The 3' end of the non-ligation strand can be modified to comprise a blocking group, for example, a dideoxynucleotide (ddCMP, ddAMP, ddTMP, or ddGMP) to prevent polymerase extension. The blocking group at the 3' end of the non-ligation strand can be a nucleotide terminator. In some cases, the block at the 3' end of the non-ligation strand comprises a terminal dideoxycytosine. The 5' end of the non-ligation strand can be modified to comprise a blocking group. The blocking group at the 5' end of the non-ligation strand can be a spacer (C3 phosphoramidite, triethylene glycol (TEG), photo-cleavable, hexa-ethyleneglycol), inverted dideoxy-T, biotin, thiol, dithiol, hexanediol, digoxigenin, an azide, alkynes, or an amino modifier. The biotin blocking group can be photocleavable biotin, biotin-triethylene glycol (TEG), biotin-dT, desthiobiotin-TEG, biotin-azide, or dual biotin. In some cases, the block at the 5' end of the non-ligation strand comprises a biotin moiety. In some cases, the 5' end of the non-ligation strand does not comprise a 5' phosphate. The 5' end can be removed by treatment with an enzyme. The enzyme can be a phosphatase. In some cases, the 5' end of the non-ligation strand is dephosphorylated by treatment with alkaline phosphatase. In some cases, the 5' end of the non-ligation strand does comprise a 5' phosphate, wherein the 3' end of the polynucleotide lacks a free 3' hydroxyl. In some cases, the non-ligation strand comprises a block at the 3' end comprising terminal dideoxycytosine and a block at the 5' end comprising a biotin moiety. In some cases, one or more cytosine bases in the ligation strand of a duplex adaptor can be substituted with a cytosine analog resistant to bisulfite treatment. The ligation strand comprising the one or more cytosine analogs resistant to bisulfite treatment can be the long strand in a partial duplex adaptor as provided herein. In some cases, one or more cytosine bases in the non-ligation strand of a duplex adaptor can be substituted with a cytosine analog resistant to bisulfite

IDT Exhibit 1017.1577

treatment. The non-ligation strand comprising the one or more cytosine analogs resistant to bisulfite treatment can be the short strand in a partial duplex adaptor as provided herein. These types of duplex or partial duplex adaptors can be useful in the generation of RRBS libraries using the methods provided herein. In some cases, the cytosine analog is 5-methylcytosine. In some cases, the cytosine analog is 5-hydroxymethylcytosine. In some cases, the cytosine analog is 5-propynylcytosine.

[0072] Provided herein is a method for generating the diversity adaptors useful for generating RRBS nucleic acid libraries comprising increased sequence diversity as provided herein. **FIG. 2A-B** illustrates a method for generating the diversity adaptors useful in the methods, compositions, and kits proved herein. **FIG. 2A** shows a long strand of nucleotides (R01446) and 4 different short strands (i.e., Ben383, Ben384, Ben385, and Ben341) such that each short strand comprises stretch of 12 nucleotides that comprise bases complementary to the 3' end of the long strand (R01446). Each of the four short strands differ in their 5' end sequences such that Ben341 comprises a 5'-CG-3' dinucleotide, while Ben385 comprises an additional H base 3' to the 5'-CG-3' dinucleotide, Ben384 an additional two H bases 3' to the 5'-CG-3' dinucleotide, and Ben383 an additional three H bases 3' to the 5'-CG-3' dinucleotide. The H bases in the short strands can be either and adenine (A), cytosine (C), or thymine (T). **FIG. 2B** illustrates the method used to generate a diversity adaptor using the long strand (R01446) and a representative short strand (Ben383) from **FIG. 2A**, although it should be noted that any of the four short strands can be substituted for the Ben383 short strand in **FIG. 2B.** Each of the short strands (Ben383, 384, 385, or 341) in **FIG. 2A** comprise a 5' terminal GC sequence with a Biosg label on the terminal cytosine residue, while the 3' terminus comprises a terminal 3' dideoxy cytosine residue.

[0073] In **FIG. 2B**, the long strand of nucleotides (R01446) can be annealed to a short strand of nucleotides (Ben383) such that a partial duplex is formed. The partial duplex can comprise a duplex region and a 5' overhang on both ends of the duplex, wherein both the long strand (R01446) and the short strand (Ben383) contribute sequence to the 5' overhang on the opposite strand. The 5' overhang comprising sequence from the short strand of the partial duplex can comprise a 5' terminal GC sequence with a Biosg label on the terminal cytosine residue, while the 3' terminus comprises a terminal 3' dideoxy cytosine residue. Additionally, the 5' overhang from the short strand comprises three "H" bases. If Ben384 were used in the method illustrated in **FIG. 2B**, the 5' overhang would further comprise two H bases, Ben385 would comprise one H base, and Ben341 would comprise no H bases. The "H" bases can be either an adenosine, cytosine or thymine residue, but cannot be a guanine residue. As shown in **FIG. 2B**, following

IDT Exhibit 1017.1578

annealing, the partial duplex adapter can be subjected to an extension reaction using a dNTP mix containing only dATP, dGTP, and dTTP (i.e., no dCTP), whereby a partial duplex adapter comprising a 5' overhang comprising at least one "H" (e.g., R01446 hybridized to Ben 383, 384, or 385) is extended to produce a partial duplex adapter containing a two base 5' overhang with the sequence GC at one end. In some cases, the diversity adaptors generated in **FIG. 2B** are ligated to polynucleotides fragmented with MspI. The diversity adaptors generated in **FIG. 2B** can be mixed into a single pool thereby comprising a mixture of diversity adaptors comprising 0, 1, 2, or 3 D bases (D0-D3) and ligated to the MspI polynucleotide fragments. In some cases, the MspI polynucleotide fragments were separated into different pools, wherein a first pool was ligated to diversity adaptors with no D bases (D0), a second pool was ligated to diversity adapters with one D base (D1), a third pool was ligated diversity adapters comprising two D bases (D2), a fourth pool was ligated to adapters comprising three D bases (D3), and a fifth pool was ligated to a mixture of D0, D1, D2, and D3 diversity adapters. The methods illustrated in **FIG. 2A-B** can be used to generate diversity adaptors useful for ligating to polynucleotide fragments generated by digestion with any restriction enzyme. In these cases, the nucleotide base compositions of the non-H base overhang can be adjusted to be complementary to overhang sequence produced by enzymatic digestion, while the base identity of the H bases can be altered to so as to be distinguishable from the recognition sequence of the restriction enzyme used to generate the site upon sequencing of the fragments. In some cases, diversity adaptors or pools of diversity adaptors can be ligated to polynucleotide ends using blunt end ligation. The ends of the polynucleotide can be blunt due to digestion with an enzyme that generates blunt ends, or can be blunted using an end repair reaction as described herein and known in the art. In some cases, the diversity adaptors are configured such that each strand is capable of ligation to an end of a polynucleotide. In some cases, the diversity adaptors where both strands are ligation capable differ in that one strand comprises cytosine analogs resistant to bisulfite treatment, while the complementary strand comprises cytosine bases sensitive to bisulfite treatment such that bisulfite treatment of polynucleotides comprising said diversity adaptors ligated thereto generate polynucleotides comprising non-complementary termini.

III. Polynucleotides, oligonucleotides, primers, samples, and nucleotides

[0074] A bisulfite converted polynucleotide library (e.g., reduced representation bisulfite library) can be generated from a polynucleotide obtained from a source of polynucleotides. The polynucleotides can be single-stranded or double stranded. In some cases, the polynucleotide is DNA. The DNA can be obtained and purified using standard techniques in the art and include DNA in purified or unpurified form. The DNA can be mitochondrial DNA, cell-free DNA,

IDT Exhibit 1017.1579

complementary DNA (cDNA), or genomic DNA. In some cases, the polynucleotide is genomic DNA. The DNA can be plasmid DNA, cosmid DNA, bacterial artificial chromosome (BAC), or yeast artificial chromosome (YAC). The DNA can be derived from one or more chromosomes. For example, if the DNA is from a human, the DNA can derived from one or more of chromosome 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, X, or Y. In some cases, the DNA is double-stranded DNA. In some cases, the double-stranded DNA is genomic DNA. In some cases, the DNA is cDNA. In some cases, the cDNA is double-stranded cDNA. In some cases, the cDNA is derived from RNA, wherein the RNA is subjected to first strand synthesis followed by second strand synthesis. The RNA can be obtained and purified using standard techniques in the art and include RNAs in purified or unpurified form, which include, but are not limited to, mRNAs, tRNAs, snRNAs, rRNAs, retroviruses, small non-coding RNAs, microRNAs, polysomal RNAs, pre-mRNAs, intronic RNA, viral RNA, cell free RNA and fragments thereof. The non-coding RNA or ncRNA can include snoRNAs, microRNAs, siRNAs, piRNAs and long nc RNAs. First strand synthesis can be performed using any number of RNA dependent DNA polymerases known in the art.

[0075] The source of polynucleotides for use in the methods described herein can be a sample comprising the polynucleotides. The polynucleotides can be isolated from the sample and purified by any of the methods known in the art for purifying the nucleic acid from the sample. The sample can be derived from a non-cellular entity comprising polynucleotides (e.g., a virus) or from a cell-based organism (e.g., member of archaea, bacteria, or eukarya domains). In some cases, the sample is obtained from a swab of a surface, such as a door or bench top.

[0076] The sample can from a subject, e.g., a plant, fungi, eubacteria, archeabacteria, protest, or animal. The subject can be an organism, either a single-celled or multi-cellular organism. The subject can be cultured cells, which can be primary cells or cells from an established cell line, among others. The sample can be isolated initially from a multi-cellular organism in any suitable form. The animal can be a fish, e.g., a zebrafish. The animal can be a mammal. The mammal can be, e.g., a dog, cat, horse, cow, mouse, rat, or pig. The mammal can be a primate, e.g., a human, chimpanzee, orangutan, or gorilla. The human can be a male or female. The sample can be from a human embryo or human fetus. The human can be an infant, child, teenager, adult, or elderly person. The female can be pregnant, suspected of being pregnant, or planning to become pregnant.

[0077] The sample can be from a subject (e.g., human subject) who is healthy. In some cases, the sample is taken from a subject (e.g., an expectant mother) at at least 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, or 26 weeks of gestation. In some cases, the

IDT Exhibit 1017.1580

subject is affected by a genetic disease, a carrier for a genetic disease or at risk for developing or passing down a genetic disease, where a genetic disease is any disease that can be linked to a genetic variation such as mutations, insertions, additions, deletions, translocation, point mutation, trinucleotide repeat disorders and/or single nucleotide polymorphisms (SNPs).

[0078] The sample can be from a subject who has a specific disease, disorder, or condition, or is suspected of having (or at risk of having) a specific disease, disorder or condition. For example, the sample can be from a cancer patient, a patient suspected of having cancer or a patient at risk of having cancer. The cancer can be, e.g., acute lymphoblastic leukemia (ALL), acute myeloid leukemia (AML), adrenocortical carcinoma, Kaposi Sarcoma, anal cancer, basal cell carcinoma, bile duct cancer, bladder cancer, bone cancer, osteosarcoma, malignant fibrous histiocytoma, brain stem glioma, brain cancer, craniopharyngioma, ependymoblastoma, ependymoma, medulloblastoma, medulloeptithelioma, pineal parenchymal tumor, breast cancer, bronchial tumor, Burkitt lymphoma, Non-Hodgkin lymphoma, carcinoid tumor, cervical cancer, chordoma, chronic lymphocytic leukemia (CLL), chronic myelogenous leukemia (CML), colon cancer, colorectal cancer, cutaneous T-cell lymphoma, ductal carcinoma in situ, endometrial cancer, esophageal cancer, Ewing Sarcoma, eye cancer, intraocular melanoma, retinoblastoma, fibrous histiocytoma, gallbladder cancer, gastric cancer, glioma, hairy cell leukemia, head and neck cancer, heart cancer, hepatocellular (liver) cancer, Hodgkin lymphoma, hypopharyngeal cancer, kidney cancer, laryngeal cancer, lip cancer, oral cavity cancer, lung cancer, non-small cell carcinoma, small cell carcinoma, melanoma, mouth cancer, myelodysplastic syndromes, multiple myeloma, medulloblastoma, nasal cavity cancer, paranasal sinus cancer, neuroblastoma, nasopharyngeal cancer, oral cancer, oropharyngeal cancer, osteosarcoma, ovarian cancer, pancreatic cancer, papillomatosis, paraganglioma, parathyroid cancer, penile cancer, pharyngeal cancer, pituitary tumor, plasma cell neoplasm, prostate cancer, rectal cancer, renal cell cancer, rhabdomyosarcoma, salivary gland cancer, Sezary syndrome, skin cancer, nonmelanoma, small intestine cancer, soft tissue sarcoma, squamous cell carcinoma, testicular cancer, throat cancer, thymoma, thyroid cancer, urethral cancer, uterine cancer, uterine sarcoma, vaginal cancer, vulvar cancer, Waldenstrom Macroglobulinemia, or Wilms Tumor. The sample can be from the cancer and/or normal tissue from the cancer patient.

[0079] The sample can be from a subject who is known or suspected to have a genetic disease, disorder or condition. The sample can be from a subject who is known or suspected to have gene silencing, imprinting, X-chromosome inactivation, phenotypic variation or genetic disease susceptibility. In some cases, the subject is known to be wild-type or mutant for a gene, or portion of a gene, e.g., CFTR, Factor VIII (F8 gene), beta globin, hemachromatosis, G6PD,

IDT Exhibit 1017.1581

neurofibromatosis, GAPDH, beta amyloid, or pyruvate kinase gene. In some cases, the status of the subject is either known or not known, and the subject is tested for the presence of a mutation or genetic variation of a gene, e.g., CFTR, Factor VIII (F8 gene), beta globin, hemachromatosis, G6PD, neurofibromatosis, GAPDH, beta amyloid, or pyruvate kinase gene.

[0080] The sample can be aqueous humour, vitreous humour, bile, whole blood, blood serum, blood plasma, breast milk, cerebrospinal fluid, cerumen, enolymph, perilymph, gastric juice, mucus, peritoneal fluid, saliva, sebum, semen, sweat, tears, vaginal secretion, vomit, feces, or urine. The sample can be obtained from a hospital, laboratory, clinical or medical laboratory. The sample can be taken from a subject.

[0081] The sample can comprise nucleic acid. The sample can comprise a whole genome, and thus be considered a whole genome sample. The sample can be a nucleic acid or polynucleotide library. The nucleic acid can be, e.g., mitochondrial DNA, genomic DNA, mRNA, siRNA, miRNA, cRNA, single-stranded DNA, double-stranded DNA, single-stranded RNA, double-stranded RNA, tRNA, rRNA, or cDNA. The sample can comprise cell-free nucleic acid. The sample can be a cell line, genomic DNA, cell-free plasma, formalin fixed paraffin embedded (FFPE) sample, or flash frozen sample. A formalin fixed paraffin embedded sample can be deparaffinized before nucleic acid is extracted. The sample can be from an organ, e.g., heart, skin, liver, lung, breast, stomach, pancreas, bladder, colon, gall bladder, brain, etc. Nucleic acids can be extracted from a sample by means available to one of ordinary skill in the art.

[0082] The sample can be processed to render it competent for fragmentation, ligation, denaturation, and/or amplification. Exemplary sample processing can include lysing cells of the sample to release nucleic acid, purifying the sample (e.g., to isolate nucleic acid from other sample components, which can inhibit enzymatic reactions), diluting/concentrating the sample, and/or combining the sample with reagents for further nucleic acid processing. In some examples, the sample can be combined with a restriction enzyme, reverse transcriptase, or any other enzyme of nucleic acid processing.

[0083] The methods described herein can be used for analyzing or detecting one or more target polynucleotides. The term polynucleotide, or grammatical equivalents, can refer to at least two nucleotides covalently linked together. A polynucleotide described herein can contain phosphodiester bonds, although in some cases, as outlined below (for example in the construction of primers and probes such as label probes), nucleic acid analogs are included that can have alternate backbones, comprising, for example, phosphoramide (Beaucage et al., Tetrahedron 49(10):1925 (1993) and references therein; Letsinger, J. Org. Chem. 35:3800 (1970); Sprinzl et al., Eur. J. Biochem. 81:579 (1977); Letsinger et al., Nucl. Acids Res. 14:3487

IDT Exhibit 1017.1582

(1986); Sawai et al, Chem. Lett. 805 (1984), Letsinger et al., J. Am. Chem. Soc. 110:4470 (1988); and Pauwels et al., Chemica Scripta 26:141 91986)), phosphorothioate (Mag et al., Nucleic Acids Res. 19:1437 (1991); and U.S. Pat. No. 5,644,048), phosphorodithioate (Briu et al., J. Am. Chem. Soc. 111:2321 (1989), O-methylphosphoroamidite linkages (see Eckstein, Oligonucleotides and Analogues: A Practical Approach, Oxford University Press), and peptide nucleic acid (also referred to herein as "PNA") backbones and linkages (see Egholm, J. Am. Chem. Soc. 114:1895 (1992); Meier et al., Chem. Int. Ed. Engl. 31:1008 (1992); Nielsen, Nature, 365:566 (1993); Carlsson et al., Nature 380:207 (1996), all of which are incorporated by reference). Other analog nucleic acids include those with bicyclic structures including locked nucleic acids (also referred to herein as "LNA"), Koshkin et al., J. Am. Chem. Soc. 120.13252 3 (1998); positive backbones (Denpcy et al., Proc. Natl. Acad. Sci. USA 92:6097 (1995); non-ionic backbones (U.S. Pat. Nos. 5,386,023, 5,637,684, 5,602,240, 5,216,141 and 4,469,863; Kiedrowshi et al., Angew. Chem. Intl. Ed. English 30:423 (1991); Letsinger et al., J. Am. Chem. Soc. 110:4470 (1988); Letsinger et al., Nucleoside & Nucleotide 13:1597 (1994); Chapters 2 and 3, ASC Symposium Series 580, "Carbohydrate Modifications in Antisense Research", Ed. Y. S. Sanghui and P. Dan Cook; Mesmaeker et al., Bioorganic & Medicinal Chem. Lett. 4:395 (1994); Jeffs et al., J. Biomolecular NMR 34:17 (1994); Tetrahedron Lett. 37:743 (1996)) and non-ribose backbones, including those described in U.S. Pat. Nos. 5,235,033 and 5,034,506, and Chapters 6 and 7, ASC Symposium Series 580, "Carbohydrate Modifications in Antisense Research", Ed. Y. S. Sanghui and P. Dan Cook. Nucleic acids containing one or more carbocyclic sugars are also included within the definition of nucleic acids (see Jenkins et al., Chem. Soc. Rev. (1995) pp 169 176). Several nucleic acid analogs are described in Rawls, C & E News Jun. 2, 1997 page 35. "Locked nucleic acids" are also included within the definition of nucleic acid analogs. LNAs are a class of nucleic acid analogues in which the ribose ring is "locked" by a methylene bridge connecting the 2'-O atom with the 4'-C atom. All of these references are hereby expressly incorporated by reference. These modifications of the ribose-phosphate backbone can be done to increase the stability and half-life of such molecules in physiological environments. For example, PNA:DNA and LNA-DNA hybrids can exhibit higher stability and thus can be used in some cases. The polynucleotides can be single stranded or double stranded, as specified, or contain portions of both double stranded or single stranded sequence. Depending on the application, the nucleic acids can be DNA (including, e.g., genomic DNA, mitochondrial DNA, and cDNA), RNA (including, e.g., mRNA and rRNA) or a hybrid, where the nucleic acid contains any combination of deoxyribo- and ribo-nucleotides, and any combination of bases,

IDT Exhibit 1017.1583

including uracil, adenine, thymine, cytosine, guanine, inosine, xathanine hypoxathanine, isocytosine, isoguanine, etc.

[0084] "Unmodified nucleotide" or "unmodified dNTP" can refer to the four deoxyribonucleotide triphosphates dATP (deoxyadenosine triphosphate), dCTP (deoxycytidine triphosphate), dGTP (deoxyguanosine triphosphate) and dTTP (deoxythymidine triphosphate) that can normally be used as building blocks in the synthesis of DNA.

[0085] "Modified nucleotide," "modified dNTP," or "nucleotide analog," can refer to any molecule suitable for substituting one corresponding unmodified nucleotide. The modified nucleotide or dNTP render the polynucleotide more or less susceptible to degradation or alteration by a suitable degrading or altering agent. In some cases, the modified nucleotide substitutes for cytosine, which in its unmodified state undergoes conversion to uracil when subjected to bisulfite treatment. In some cases, the modified nucleotide substituting for cytosine is 5-methylcytosine. In some cases, the modified nucleotide substituting for cytosine is 5-hydroxymethylcytosine. In some cases, the modified nucleotide is 5-propynylcytosine.

[0086] As used herein, "oligonucleotide" can refer to a polynucleotide chain, typically less than 200 residues long, most typically between 15 and 100 nucleotides long, but also intended to encompass longer polynucleotide chains. Oligonucleotides may be single-or double-stranded. The terms "primer", and "oligonucleotide primer", as used herein, can refer to an oligonucleotide capable of hybridizing to a complementary nucleotide sequence.

[0087] As used herein, "hybridization"/ "hybridizing" and "annealing" can be used interchangeably and refer to the pairing of complementary nucleic acids or sequences within nucleic acids. "Primer", as used herein, can refer to an oligonucleotide, generally with a free 3' hydroxyl group, that is capable of hybridizing with a template (such as a target polynucleotide or target sequence within a polynucleotide, target DNA, target RNA or a primer extension product) and is also capable of promoting polymerization of a polynucleotide complementary to the template. 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. In some cases, a first and/or second primer used to amplify a nucleic acid or polynucleotide fragment comprising diversity sequences on the 5' and/or 3' ends generated by the methods as provided herein further comprise one or more identifier sequences. In some cases, the identifier sequences comprise a non-hybridizable tail on the first and/or second primer. The identifier sequence can be a barcode sequence, a flow cell sequence, and/or an index sequence. In some cases, the index sequence is a Truseq primer sequence compatible with the next generation sequencing (NGS) platform produced by Illumina. In some cases, the

IDT Exhibit 1017.1584

first and/or second primer can bind to a solid surface. The solid surface can be a planar surface or a bead. The planar surface can be the surface of a chip, microarray, well, or flow cell. In some cases, the first and/or second primer comprises one or more sequence elements wherein products of the amplification reaction (i.e. amplification products) bind to a solid surface, whereby the one or more sequences elements are complementary to one or more capture probes attached to a solid surface.

[0088] The primers described herein can generally be oligonucleotides that are employed in an extension reaction by a polymerase along a template (such as a target polynucleotide or target sequence within a polynucleotide, target DNA, target RNA or a primer extension product), such as in PCR or cDNA synthesis, for example. An 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 a target polynucleotide. Normally, the 3′ region of the primer that hybridizes with the target nucleic acid has at least 80%, 90%, 95%, 100%, complementarity to a sequence or primer binding site.

[0089] In some cases, one or more primers can be used in the methods provided herein to generate nucleic acid libraries comprising increased diversity relative to the same nucleic acid libraries being generated without the increased diversity. The primers used to generate the nucleic acid libraries comprising increased diversity can comprise extender sequence diversity sequence as provided herein. The primers can comprise sequence complementary to a target sequence in a polynucleotide as well as sequence that is non-complementary to the target sequence. The non-complementary sequence can be a tail and can comprise one or more diversity bases or extender sequence as described herein. The primers comprising one or more diversity bases can be introduced to one or more ends of the target sequence from a polynucleotide using a polymerase extension reaction (e.g., PCR) as provided herein. The primers comprising diversity base(s) (i.e., extender sequence) can be designed such that the diversity base or bases are immediately adjacent to the target sequence derived from a polynucleotide following introduction. In other words, the diversity base or bases are the first non-target sequence polynucleotide base or bases present at the junction of a complex comprising a target sequence from a polynucleotide and non-target sequence derived sequence. Also as provided herein, the primers comprising diversity bases can comprise a sequence or complement thereof that is capable of hybridizing a sequencing primer and this sequence or complement thereof can be located adjacent to the diversity base or bases such that the diversity base or bases are the bases initially sequenced in a sequencing reaction. In some cases, there can also be additional sequence elements located between the sequence or complement thereof that is

IDT Exhibit 1017.1585

capable of hybridizing a sequencing primer and the diversity base or bases such as barcodes (e.g., inline barcodes), universal sequence, index sequence, or random sequence. In some cases, the additional sequence elements (e.g., barcode(s), universal sequence and/or random sequence) can be located after the sequence or complement thereof that is capable of hybridizing a sequencing primer and distal to the diversity base or bases. In some cases, the additional sequence elements (e.g., barcode(s), universal sequence and/or random sequence) can be located after the sequence or complement thereof that is capable of hybridizing a sequencing primer and distal to the diversity base or bases as well as between the sequence or complement thereof that is capable of hybridizing a sequencing primer and the diversity base or bases such as barcodes (e.g., inline barcodes), universal sequence, or random sequence. In some cases, the primers comprising diversity bases used in the methods provided herein can be used to generate low diversity libraries such as 16S rRNA libraries or RRBS libraries such that the low diversity libraries comprise increased diversity versus low diversity libraries generated without the use of primers comprising diversity bases. The primers comprising diversity bases can be mixed into a single pool thereby comprising a mixture of primers comprising diversity bases of varying lengths, wherein some of the primers in the pool comprise no diversity bases, while some comprise 1 diversity bases, some comprise 2 diversity bases, some comprise 3 diversity bases, some comprise 4 diversity bases, some comprise 5 diversity bases, some comprise 6 diversity bases, some comprise 7 diversity bases, some comprise 8 diversity bases, some comprise 9 diversity bases, some comprise 10 diversity bases, some comprise 11 diversity bases, and some comprise 12 diversity bases. Each primer in the pool of primers can comprise sequence complementary to the same target sequence. Each primer in the pool of primers can further comprise sequence that can be hybridized by a sequencing primer or a primer that can introduce a common sequencing primer target sequence. The single pool of primers comprising diversity bases of varying lengths can be introduced to a target sequence (e.g., 16S rRNA) present in polynucleotides from multiple samples. As described herein, the diversity bases in a primer comprising diversity bases can be "D" bases, "R" bases, complements of "D" bases, or complements of "R" bases or combinations thereof.

[0090] "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 can be such that stringency conditions used to hybridize the oligonucleotide primer can prevent excessive random non-specific hybridization. In some cases, the number of nucleotides in the hybridizing portion of the oligonucleotide primer can be at least as great as the defined sequence on the target polynucleotide that the oligonucleotide primer

IDT Exhibit 1017.1586

hybridizes to, namely, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 11, at least 12, at least 13, at least 14, at least 15, at least about 20, and generally from about 6 to about 10 or 6 to about 12 of 12 to about 200 nucleotides, usually about 10 to about 50 nucleotides. In general, the target polynucleotide can be larger than the oligonucleotide primer or primers as described previously.

[0091] 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 below. 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, 30,000, 40,000, 50,000, 60,000, 70,000, 80,000, 90,000, 100,000 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.

[0092] "Barcode" can refer to a known polynucleotide sequence that allows some feature of a polynucleotide with which the barcode is associated to be identified. In some cases, the feature of the polynucleotide to be identified is the sample from which the polynucleotide is derived. In some cases, barcodes are at least 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, or more nucleotides in length. In some cases, barcodes are shorter than 10, 9, 8, 7, 6, 5, or 4 nucleotides in length. A oligonucleotide (e.g., primer or adaptor) can comprise about, more than, less than, or at least 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 different barcodes. In some cases, barcodes associated with some polynucleotides are of different length than barcodes associated with other polynucleotides. Barcodes can be of sufficient length and comprise sequences that can be sufficiently different to allow the identification of samples based on barcodes with which they are associated. In some cases, a barcode, and the sample source with which it is associated, can be identified accurately after the mutation, insertion, or deletion of one or more nucleotides in the barcode sequence, such as the mutation, insertion, or deletion of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more nucleotides. In some cases, each barcode in a plurality of barcodes differ from every other barcode in the plurality at at least three nucleotide positions, such as at least 3, 4, 5, 6, 7, 8, 9, 10, or more

IDT Exhibit 1017.1587

positions. In some cases, an oligonucleotide (e.g., adaptor or primer) comprises at least one of a plurality of barcode sequences. In some cases, barcodes for a second oligonucleotide (e.g., adaptor or primer) are selected independently from barcodes for a first oligonucleotide (e.g., adaptor or primer). In some cases, first oligonucleotides (e.g., adaptor or primer) and second oligonucleotides (e.g., adaptor or primer) having barcodes are paired, such that oligonucleotides (e.g., adaptors or primers) of the pair comprise the same or different one or more barcodes. In some cases, the methods described herein further comprise identifying the sample from which a target polynucleotide is derived based on a barcode sequence to which the target polynucleotide is joined. A barcode can comprise a oligonucleotide sequence that when joined to a target polynucleotide serves as an identifier of the sample from which the target polynucleotide was derived.

[0093] "Extender sequence" can refer to a stretch of nucleotides that can vary in length and base composition and can serve to increase the distance between the 3' end of a sequencing primer hybridized to a sequence complementary to the sequencing primer and the first nucleotide of a target sequence in a polynucleotide to be read during a nucleotide sequencing reaction. "Extender sequence" can be used interchangeably with "diversity sequence" or "diversity base."

[0094] As provided herein, the sequence complementary to the sequencing primer and the extender sequence can be appended or introduced to the first nucleotide of a target sequence in a polynucleotide through ligation or amplification such that the extender sequence is immediately upstream of the first base with respect to the direction of sequencing and thus will be sequenced prior to the first base. The extender sequence can be 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 or more nucleotides in length. In some cases, the extender sequence and the sequence complementary to a sequencing primer are present in the same oligonucleotide sequence. As provided herein, the oligonucleotide sequence can be present in a primer or adaptor. The extender sequence and the sequence complementary to a sequencing primer can be immediately adjacent to each other or can be separated by a stretch of nucleotides. The stretch of nucleotides can be an additional sequence element. The additional sequence element can be one or more barcodes as provided herein, universal sequence, linker sequence, or random sequence. The nucleotide base composition of each nucleotide in an extender sequence comprising more than one nucleotide can be identical or different. The nucleotide base composition of each nucleotide in an extender sequence can be randomly assigned. The nucleotide base composition of each nucleotide in an extender sequence can be semi-randomly assigned such that each nucleotide in an extender sequence (or diversity sequence) can only comprise 2 or 3 of the four possible canonical bases (i.e., thymine, guanine, cytosine, or adenine) such that one or two

IDT Exhibit 1017.1588

specific bases can be excluded. The nucleotide base composition of each nucleotide in an extender sequence can have a 25% chance of being one of the four unmodified bases (i.e., thymine, guanine, cytosine, or adenine). The nucleotide base composition of each nucleotide in an extender sequence can be the same nucleotide base. The nucleotide base composition of an extender sequence comprising more than one nucleotide can be constructed such that one or more nucleotide positions can have a fixed base composition, while the remaining nucleotides can have a randomly or semi-randomly assigned base composition as described herein.

[0095] RNA-dependent DNA polymerases for use in the methods and compositions provided herein can be capable of effecting extension of a primer according to the methods provided herein. Accordingly, an RNA-dependent DNA polymerase can 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, compositions, and kits provided herein 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 can 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.

[0096] DNA-dependent DNA polymerases for use in the methods and compositions provided herein can be capable of effecting extension of a primer according to the methods provided herein. Accordingly, a DNA-dependent DNA polymerase can be one that is capable of

IDT Exhibit 1017.1589

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 provided herein 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 can be performed using a polymerase comprising strong strand displacement activity such as for example Bst polymerase. In other cases, the primer extension can be performed using a polymerase comprising weak or no strand displacement activity. One skilled in the art can recognize the advantages and disadvantages of the use of strand displacement activity during the primer extension step, and which polymerases can be expected to provide strand displacement activity (see e.g., New England Biolabs Polymerases). For example, strand displacement activity can be useful in ensuring whole transcriptome coverage during the random priming and extension step. Strand displacement activity can 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 can be useful in the generation of single stranded nucleic acid products during primer hybridization and extension that can be hybridized to the template nucleic acid.

[0097] In some cases, the double stranded products generated by the methods described herein can be end repaired to produce blunt ends for the adapter ligation applications described herein. Generation of the blunt ends on the double stranded products can be generated by the use of a single strand specific DNA exonuclease such as for example exonuclease 1, exonuclease 7 or a combination thereof to degrade overhanging single stranded ends of the double stranded products. Alternatively, the double stranded products can be blunt ended by the use of a single stranded specific DNA endonuclease for example but not limited to mung bean endonuclease or S1 endonuclease. Alternatively, the double stranded products can be blunt ended by the use of a polymerase that comprises single stranded exonuclease activity such as for example T4 DNA polymerase, any other polymerase comprising single stranded exonuclease activity or a combination thereof to degrade the overhanging single stranded ends of the double stranded products. In some cases, the polymerase comprising single stranded exonuclease activity can be incubated in a reaction mixture that does or does not comprise one or more dNTPs. In other cases, a combination of single stranded nucleic acid specific exonucleases and one or more polymerases can be used to blunt end the double stranded products of the primer extension

IDT Exhibit 1017.1590

reaction. In still other cases, the products of the extension reaction can be made blunt ended by filling in the overhanging single stranded ends of the double stranded products. For example, the fragments can 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 can be made blunt by a combination of a single stranded overhang degradation reaction using exonucleases and/or polymerases, and a fill-in reaction using one or more polymerases in the presence of one or more dNTPs.

[0098] In another embodiment, the adapter ligation applications described herein can leave a nick between a non-ligation strand of the adapters and a strand of the double stranded product. In these instances, a gap repair or fill-in reaction can be used to append the double stranded product with the sequence complementary to the ligation strand of the adapter. Gap repair can be performed with any number of DNA dependent DNA polymerase described herein. In some cases, gap repair can be performed with a DNA dependent DNA polymerase with strand displacement activity. In some cases, gap repair can be performed using a DNA dependent DNA polymerase with weak or no strand displacement activity. In some cases, the ligation strand of the adapter can serve as the template for the gap repair or fill-in reaction. In some cases, gap repair can be performed using *Taq* DNA polymerase.

[0099] Various ligation processes and reagents are known in the art and can be useful for carrying out the methods provided herein. For example, the NuGEN Ovation ® Ultralow Methyl-Seq Library System can be used. In some cases, blunt end ligation is 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 adaptor (e.g., diversity adaptor as provided herein) comprising a dT overhang (or the reverse). This design can allow the hybridized components to be subsequently ligated (e.g., by T4 DNA ligase). Other ligation strategies and the corresponding reagents known in the art can be employed and kits and reagents for carrying out efficient ligation reactions are commercially available (e.g, from New England Biolabs, Roche) and can be used for the methods provided herein.

[00100] "Bisulfite" as used herein can encompass all types of bisulfites, such as sodium bisulfite that are capable of chemically converting a cytosine (C) to a uracil (U) without chemically modifying a methylated cytosine and therefore can be used to differentially modify a DNA sequence based on the methylation status of the DNA.

[00101] "Low diversity" can refer to a polynucleotide. For example, the methods, compositions, and kits described herein can be used with pools of polynucleotides that have low diversity at

IDT Exhibit 1017.1591

one or more positions at one or both ends of polynucleotides in the pool. A "position" can be a single nucleotide, or 2 or more nucleotides. A "position" can be 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 nucleotides in length. A "position" can be a nucleotide sequence repeat such as CpG or a conserved region of a gene.

[00102] A polynucleotide can have low diversity at a position among polynucleotides if the position among polynucleotides has only one, two, or three nucleotide bases out of the four canonical DNA nucleotide bases. In some cases, a polynucleotide can have low diversity at a position among polynucleotides if it lacks one of the types of nucleotide bases present in the polynucleotide. For example, a pool of polynucleotides can one type of nucleotide base, e.g., only A, only T, only C, or only G at a position in the polynucleotides in the pool. In some cases, a pool of polynucleotides has only two types of polynucleotides at a position, e.g., only A or C, only A or G, only A or T, only C or T, only G or T, or only C or G. In some cases, a pool of polynucleotides has only three types of polynucleotides at a position, e.g., only A, C, and G; only A, C, and T; only C, T, and G.

[00103] In some cases, a pool of polynucleotides can have low diversity at a position if all types of nucleotide bases (e.g., A, T, C, and G) are at the position among polynucleotides in the pool, but the percentage of polynucleotides with one, two, or three types of bases at the position is low. For example, a pool of polynucleotides can have all four canonical bases found in DNA at the position, but one of the types of nucleotide bases is present at that position in less than 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2%, or 1% of the polynucleotides. For example, a pool polynucleotides can have A, T, C, and G at a position, where 50% of polynucleotides have an A at the position, 20% have a T at the position, 20% have a C at the position, and only 10% have a G at the position.

[00104] In some cases, a pool of polynucleotides can have all four canonical bases found in DNA at a position among the polynucleotides in the pool, but each of two of the types of nucleotide bases is present at that position in less than 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2%, or 1% of the polynucleotides. For example, a pool polynucleotides can have A, T, C, and G at a position, where 50% of polynucleotides have an A at the position, 40% have a T at the position, 5% have a C at the position, and 5% have a G at the position.

[00105] A pool of polynucleotides can have low diversity at a position if all four canonical bases found in DNA are at the position among the polynucleotides in the pool, but each of three of the types of nucleotide bases is present at that position in less than 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2%, or 1% of the polynucleotides. For example, a pool polynucleotides can have A, T, C,

IDT Exhibit 1017.1592

and G at a position, where 94% of polynucleotides have an A at the position, 2% have a T at the position, 2% have a C at the position, and 2% have a G at the position.

[00106] A pool of polynucleotides can have any of the characteristics of a position described herein at 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more positions. A pool of polynucleotides can have any of the characteristics of a position described herein at 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more positions immediately 5' to base that anneals to the 3' end of a primer, e.g., a sequencing primer. A pool of polynucleotides can have any of the characteristics of a position described herein at 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more positions immediately 5' to base that anneals to the 3' end of a primer, e.g., a sequencing primer.

[00107] In some cases, a pool of polynucleotides with low diversity at one or both ends is generated using restriction enzyme. In some cases the restriction enzyme has a palindromic recognition sequence. In some cases, a recognition site comprises a methylation site. In some cases, a palindromic recognition site comprises a methylation site. For example, a pool of polynucleotides can be generated by using the restriction enzyme MspI, which can have the recognition sequence CCGG, where CG can be a methylation site.

[00108] In some cases, a pool of polynucleotides with low sequence diversity is generated using an amplification based method, for example, polymerase chain reaction (PCR).

IV. Methods of Fragmentation

[00109] The methods provided herein can comprise fragmenting a polynucleotide to produce polynucleotide fragments. In some cases, the polynucleotides are double-stranded and fragmentation of the polynucleotides generates double stranded polynucleotide fragments. In some cases, fragmentation can be achieved through methods known in the art. Fragmentation can be through physical fragmentation methods and/or enzymatic fragmentation methods. Physical fragmentation methods can include nebulization, sonication, and/or hydrodynamic shearing. In some cases, the fragmentation can be accomplished mechanically comprising subjecting the nucleic acid to acoustic sonication. In some cases, the fragmentation comprises treating the nucleic acid with one or more enzymes under conditions suitable for the one or more enzymes to generate breaks in the double-stranded nucleic acid. Examples of enzymes useful in the generation of nucleic acid fragments include sequence specific and non-sequence specific nucleases. Non-limiting examples of nucleases include DNase I, Fragmentase, restriction endonucleases, variants thereof, and combinations thereof. Reagents for carrying out enzymatic fragmentation reactions are commercially available (e.g, from New England Biolabs). For example, digestion with DNase I can induce random double-stranded breaks in DNA in the absence of Mg^{++} and in the presence of Mn^{++}. In some cases, fragmentation comprises treating

IDT Exhibit 1017.1593

DNA with one or more restriction endonucleases. Fragmentation can produce fragments having 5' overhangs, 3' overhangs, blunt ends, or a combination thereof. In some cases, such as when fragmentation comprises the use of one or more restriction endonucleases, cleavage of the DNA leaves overhangs having a predictable sequence. In some cases, the methods provided herein are used on to generate nucleic acid libraries from polynucleotides fragmented by digestion with restriction enzymes. In some cases, the method includes the step of size selecting the fragments via standard methods known in the art such as column purification or isolation from an agarose gel.

[00110] In some cases, the methods, compositions, and kits provided herein are used to generate a reduced representation bisulfite sequencing (RRBS) library. An initial step in the generation of an RRBS library (see **FIG. 1**, **FIG. 3,** or **FIG. 36**) can comprise fragmenting a polynucleotide or a sample comprising polynucleotides by treating the polynucleotide or sample comprising polynucleotides with a methylation insensitive restriction enzyme. The methylation insensitive restriction enzyme can be AatII, AccB7I, AccIII, Acc65I, ApaI, AvaI, BanII, BbuI, BclI, BamHI, BalI, BglII, Bsp12861, BssHII, BstEII, BstOI, BstXI, ClaI, CaspI, Csp45I, DdeI, Eco47III, EcoRI, HaeIII, HhaI, HincII, HindIII, KpnI, MboII, MluI, MspI, NarI, NdeII, NruI, PstI, PvuII, SacI, ScaI, SinI, SnaBI, SphI, StuI, TaqI, XbaI, XhoII, or XmaI. In some cases, fragmenting a polynucleotide or a sample comprising polynucleotides comprises treating the polynucleotide or the sample comprising polynucleotides with MspI, thereby generating polynucleotide fragments comprising a 5'-CG-3" overhang at each end. The polynucleotides treated with MspI can be from mammalian sources, whereby the CCGG recognition site of MspI is not methylated at the first cytosine residue.

[00111] In some cases, the polynucleotide, for example DNA, can be fragmented into a population of fragmented polynucleotides of one or more specific size range(s). In some cases, the fragments can have an average length from about 10 to about 10,000 nucleotides or base pairs. In some cases, the fragments have an average length from about 50 to about 2,000 nucleotides or base pairs. In some cases, the fragments have an average length from about 100 to about 2,500, about 10 to about1000, about 10 to about 800, about 10 to about 500, about 50 to about 500, about 50 to about 250, or about 50 to about 150 nucleotides or base pairs. In some cases, the fragments have an average length less than 10,000 nucleotides or bp, less than 7,500 nucleotides or bp, less than 5,000 nucleotides or bp, less than 2,500 nucleotides or bp, less than 2,000 nucleotides or bp, less than 1,500 nucleotides or bp, less than 1,000 nucleotides or bp, less than 500 nucleotides or bp, less than 400 nucleotides or bp, less than 300 nucleotides or bp, less than 200 nucleotides or bp, or less than 150 nucleotides or bp. In some cases, the polynucleotide

IDT Exhibit 1017.1594

fragments have an average length of about, more than, less than, or at least 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 125, 150, 175, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000, 2100, 2200, 2300, 2400, 2500, 2600, 2700, 2800, 2900, 3000, 3500, 4000, 4500, 5000, 5500, 6000, 6500, 7000, 7500, 8000, 8500, 9000, 9500, or 10,000 nucleotides or base pairs.

[00112] In some cases, polynucleotide fragments generated by fragmentation are subjected to end repair. 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 nucleic acid product by a polymerase lacking 3′-exonuclease activity. In some cases, end repair is performed on the double stranded polynucleotide fragments to produce blunt ends wherein the ends of the polynucleotide fragments contain 5' phosphates and 3' hydroxyls. End repair can be performed using any number of enzymes and/or methods known in the art. An overhang can comprise about, more than, less than, or at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 nucleotides.

[00113] In some cases, the polynucleotide fragments are captured using a binding agent directed against an epigenetic modification within the sequence of the polynucleotide fragments. The epigenetic modification can be methylation. In some cases, the double stranded polynucleotide fragments are captured using a binding agent directed against 5-methylcytosine residues in the double-stranded polynucleotide fragments. The binding agent can be an antibody, or the binding domain of a protein directed against 5-methylcytosine residues. The protein can be a methyl-CpG-binding domain (MBD) protein. The MBD protein can be methyl-CpG-binding domain protein 1, 2, 4, or MECP2. In some cases, the double stranded polynucleotide fragments are captured using the binding domain of MBD2. In some cases, the double stranded polynucleotide fragments are captured using the binding domain of MECP2.

V. Methods of Amplification

[00114] 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. As provided herein, the method, compositions and kits can be used to generate nucleic acid library (e.g., bisulfite converted libraries) such that each nucleic acid insert in the library comprises non-complementary sequence in opposing termini. Primer pairs comprising primers designed to target the non-complementary sequence can be used to amplify the nucleic acid inserts of a nucleic acid library prepared using the methods provided herein using the methods of amplification as provided herein. Methods of amplification are well known in the art. In some

IDT Exhibit 1017.1595

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. Examples of PCR techniques that can be used include, but are not limited to, quantitative PCR, quantitative fluorescent PCR (QF-PCR), multiplex fluorescent PCR (MF-PCR), real time PCR(RT-PCR), single cell PCR, restriction fragment length polymorphism PCR(PCR-RFLP), PCR-RFLP/RT-PCR-RFLP, hot start PCR, nested PCR, in situ polony PCR, in situ rolling circle amplification (RCA), bridge PCR, picotiter PCR, digital PCR, droplet digital PCR, and emulsion PCR. Other suitable amplification methods include the ligase chain reaction (LCR), transcription amplification, molecular inversion probe (MIP) PCR, self-sustained sequence replication, selective amplification of target polynucleotide sequences, consensus sequence primed polymerase chain reaction (CP-PCR), arbitrarily primed polymerase chain reaction (AP-PCR), degenerate oligonucleotide-primed PCR (DOP-PCR) and nucleic acid based sequence amplification (NABSA), single primer isothermal amplification (SPIA, see e.g. U.S. Pat. No. 6,251,639), Ribo-SPIA, or a combination thereof. Other amplification methods that can be used herein include those described in U.S. Pat. Nos. 5,242,794; 5,494,810; 4,988,617; and 6,582,938. Amplification of target nucleic acids can occur on a bead. In other embodiments, amplification does not occur on a bead. Amplification can be by isothermal amplification, e.g., isothermal linear amplification. A hot start PCR can be performed wherein the reaction is heated to 95°C. for two minutes prior to addition of the polymerase or the polymerase can be kept inactive until the first heating step in cycle 1. Hot start PCR can be used to minimize nonspecific amplification. Other strategies for and aspects of amplification are described in U.S. Patent Application Publication No. 2010/0173394 A1, published Jul. 8, 2010, which is incorporated herein by reference. In some cases, the amplification methods can 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.

[00115] Techniques for amplification of target and reference sequences are known in the art and include the methods described in U.S. Pat. No. 7,048,481. Briefly, the techniques can include methods and compositions that separate samples into small droplets, in some instances with each containing on average less than about 5, 4, 3, 2, or one target nucleic acid molecule (polynucleotide) per droplet, amplifying the nucleic acid sequence in each droplet and detecting the presence of a target nucleic acid sequence. In some cases, the sequence that is amplified is

present on a probe to the genomic DNA, rather than the genomic DNA itself. In some cases, at least 200, 175, 150, 125, 100, 90, 80, 70, 60, 50, 40, 30, 20, 10, or 0 droplets have zero copies of a target nucleic acid.

[00116] PCR can involve in vitro amplification based on repeated cycles of denaturation, oligonucleotide primer annealing, and primer extension by thermophilic template dependent polynucleotide polymerase, which can result in the exponential increase in copies of the desired sequence of the polynucleotide analyte flanked by the primers. In some cases, two different PCR primers, which anneal to opposite strands of the DNA, can be 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.

[00117] LCR uses a ligase enzyme to join pairs of preformed nucleic acid probes. The probes can hybridize with each complementary strand of the nucleic acid analyte, if present, and ligase can be 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.

[00118] SDA (Westin et al 2000, Nature Biotechnology, 18, 199-202; Walker et al 1992, Nucleic Acids Research, 20, 7, 1691-1696), can involve isothermal amplification 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.

[00119] Some aspects of the methods described herein can utilize linear amplification of nucleic acids or polynucleotides. Linear amplification can refer 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.

IDT Exhibit 1017.1597

[00120] In some cases, 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.

VI. Generation of Bisulfite Converted Libraries

[00121] In one aspect, a method is provided for generating a directional, bisuflite-converted nucleic acid library using duplex-forming adaptors comprising diversity bases or extender sequence as described herein. The nucleic acid library generated using the duplex-forming adapters comprising diversity bases can maintain directional (strandedness) information of the original nucleic acid sample. In some cases, the original nucleic acid is DNA. In some cases, the DNA is double-stranded DNA. In some cases, the double-stranded DNA is genomic DNA. In some cases, the nucleic acid or polynucleotide is from a whole genome sample. In some cases, the methods provided herein comprise generating a plurality of bisulfite converted libraries. One of the plurality of bisulfite converted libraries can be generated using a duplex-forming adaptor lacking diversity bases, while others of the plurality of bisulfite converted libraries can be generated using a duplex-forming adaptor comprising a variable number of diversity bases. One of the others of the plurality of bisulfite converted libraries can be generated using a duplex-forming adaptor comprising 1 diversity base. One of the others of the plurality of bisulfite converted libraries can be generated using a duplex-forming adaptor comprising 2 diversity bases. One of the others of the plurality of bisulfite converted libraries can be generated using a duplex-forming adaptor comprising 3 diversity bases.

[00122] As provided herein, the diversity base or bases can be introduced to adjacent to the first nucleotide of sequence derived from the polynucleotide to which it is attached such that the diversity base is sequenced immediately prior to the sequencing of the polynucleotide to which it is attached. In some cases, each of the plurality of bisulfite converted libraries can comprises a polynucleotide comprising the same or substantially similar sequence across at least a portion of each polynucleotide. Each of the plurality of bisulfite converted libraries comprising the polynucleotide comprising the same or substantially similar sequence across at least a portion of each polynucleotide sequence can be pooled and subsequently processed (e.g., amplified and/or solid-state amplified) and sequenced in parallel. In some cases, the bisulfite converted libraries can be generated using the method described herein by introducing (e.g., ligating) a pool of adaptors comprising adaptors comprising no diversity bases as well as adaptors comprising a varying number of diversity bases as described herein to a sample comprising polynucleotides comprising the same or substantially similar sequence across at least a portion of each

IDT Exhibit 1017.1598

polynucleotide sequence. The sequencing can be performed using a commercially available NGS system such as Illumina. The presence of diversity bases can aid in cluster determination in sequencing systems such as Illumina sequencing which can suffer from poor cluster identification in sequencing samples with low sequence diversity in their initial nucleotide bases as shown in **FIG. 31**.

[00123] The method can comprise fragmenting a double stranded polynucleotide to produce double stranded polynucleotide fragments. In some cases, fragmentation can be achieved through methods known in the art. In some cases, fragmentation can be achieved through methods known in the art and provided herein. In some cases, the fragmentation comprises treating the nucleic acid with one or more enzymes under conditions suitable for the one or more enzymes to generate breaks in the double-stranded nucleic acid. In some cases, fragmentation comprises treating DNA with one or more restriction endonucleases. Fragmentation can produce fragments having 5' overhangs, 3' overhangs, blunt ends, or a combination thereof. In some cases, such as when fragmentation comprises the use of one or more restriction endonucleases, cleavage of the DNA leaves overhangs having a predictable sequence. In some cases, the method includes the step of size selecting the fragments via standard methods known in the art such as column purification or isolation from an agarose gel. In some cases, the methods provided herein are used to generate RRBS libraries such that a sample comprising polynucleotides is fragmented using a methylation insensitive enzyme such as MspI in order to generate fragments with a predictable 5'-CG-3' overhang and subsequently introduced (e.g., through ligation) to an adaptor as shown in **FIGs 1, 3** and **36**.

[00124] In some cases, polynucleotide fragments generated by fragmentation are subjected to end repair. 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 nucleic acid product by a polymerase lacking 3′-exonuclease activity. In some cases, end repair is performed on the double stranded nucleic acid fragments to produce blunt ends wherein the ends of the polynucleotide fragments contain 5' phosphates and 3' hydroxyls. End repair can be performed using any number of enzymes and/or methods known in the art. An overhang can comprise about, more than, less than, or at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 nucleotides.

[00125] In some cases, double stranded polynucleotide fragments are captured using a binding agent directed against an epigenetic modification within the sequence of the polynucleotide fragments. The epigenetic modification can be methylation. In some cases, the double stranded polynucleotide fragments are captured using a binding agent directed against 5-methylcytosine

IDT Exhibit 1017.1599

residues in the double-stranded polynucleotide fragments. The binding agent can be an antibody, or the binding domain of a protein directed against 5-methylcytosine residues. The protein can be a methyl-CpG-binding domain (MBD) protein. The MBD protein can be methyl-CpG-binding domain protein 1, 2, 4, or MECP2. In some cases, the double stranded polynucleotide fragments are captured using the binding domain of MBD2. In some cases, the double stranded polynucleotide fragments are captured using the binding domain of MECP2.

[00126] The method can further comprise ligating an adapter comprising either no diversity bases or a variable number of diversity bases as provided herein to the double-stranded polynucleotide fragments. Ligation can be blunt end ligation or sticky or cohesive end ligation. In some cases, the ligation is performed using cohesive end ligation as shown in **FIG. 3**. The ligation can be performed with any of the enzymes known in the art for performing ligation (e.g. T4 DNA ligase). The adapter can be any type of adapter known in the art including, but not limited to, a conventional duplex or double stranded adapter. The adapter can comprise DNA, RNA, or a combination thereof. The adapters can be about, less than about, or more than about 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 90, 100, or 200 nucleotides in length. The adapters can be a duplex adaptor, partial duplex adaptor, a forked adaptor or single stranded adaptor. In some cases, the adapter is a duplex adapter. In some cases, the duplex adapters comprises about, less than about, or more than about 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 90, 100, or 200 nucleotides in length. In some cases, the adapter is a partial duplex adapter, wherein the adapter comprises a long strand and a short strand. In some cases, a partial duplex adapter has overhangs of about, more than, less than, or at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 nucleotides. In some cases, the overhang is a 5' overhang. In some cases, the overhang is a 3' overhang. In some cases, the partial duplex adapter comprises a 5' and 3' overhang. In some cases, the adapter comprises duplexed sequence. In some cases, the adapters comprise about, more than, less than, or at least 5, 6, 7, 8, 9, 10, 12, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 90, 100, 200, or more of base paired or duplexed sequence. In some cases, the adapter comprises a single stranded adapter. In some cases, a single-stranded adapter comprises about, more than, less than, or at least 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 90, 100, or 200 nucleotides in length. In some cases, the single-stranded adapter forms a stem-loop or hairpin structure. In some cases, the stem of the hairpin adapter is about, less than about, or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50, 75, 100, or more nucleotides in length. In some cases, the loop sequence of a hairpin adapter is about, less than about, or more than about 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, or more nucleotides in

IDT Exhibit 1017.1600

length. The adapter can further comprise known or universal sequence and, thus, allow generation and/or use of sequence specific primers for the universal or known sequence. In some cases, an adapter comprises one or more barcodes. In some cases, the one or more barcodes are in a stem and/or a loop.

[00127] In some cases, an adapter is marked via incorporation of at least one modified dNTP. In some cases, the modified dNTP comprises a nucleotide analog resistant to conversion by treatment with a converting agent. The nucleotide analog can be a cytosine analog. The converting agent can be any biological, biochemical, and/or chemical agent capable of altering the base composition of a dNTP. In some cases, the converting agent is a chemical. In some cases, the converting agent is the chemical compound bisulfite or sodium bisulfite. In some cases, the adapter comprises a cytosine analog resistant to conversion by bisulfite treatment. In some cases, the long strand of a partial duplex adapter comprises cytosine analog residues in place of cytosine residues, which are protected from bisulfite conversion, while the short strand of the partial duplex adapter does not comprise cytosine analog residues in place of cytosine residues. In some cases, the short strand of a partial duplex adapter comprises cytosine analog residues in place of cytosine residues, which are protected from bisulfite conversion, while the long strand of the partial duplex adapter does not comprise cytosine analog residues in place of cytosine residues. In some cases, both the long and short strand of a partial duplex adapter comprises cytosine analog residues in place of cytosine residues. In some cases, the cytosine analog is 5-methylcytosine. In some cases, the cytosine analog is 5-hydroxymethylcytosine. In some cases, the cytosine analog is 5-propynylcytosine. A strand can comprise a modified cytosine at about, more than, less than, or at least every 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50, 65, 75, 85, 90, 95, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 225, or 250 nucleotides. In some cases, ligation of an adapter to a double stranded polynucleotide is by blunt end ligation. In some cases, ligation of an adapter to a double stranded polynucleotide is by cohesive or sticky end ligation, wherein an overhang in the adapter hybridizes to an overhang in the double stranded polynucleotide comprising complementary sequence such as shown in **FIGs. 3**. In some cases, an adapter comprising a modified dNTP (e.g. a cytosine analog resistant to bisulfite treatment) comprises a ligation strand or first strand capable of ligation to a 5'end of the polynucleotide fragments and a non-ligation strand or second strand incapable of ligation to either end of the polynucleotide fragments. In some cases, the duplex adapter is a partial duplex adapter, wherein the adapter comprises a long strand and a short strand, and wherein the long strand is the ligation strand or first strand, while the short strand is the non-ligation strand or second strand. In some cases, the partial duplex has strands of unequal length. In some cases,

IDT Exhibit 1017.1601

the partial duplex comprises an overhang at one end of the adapter and a blunt end at another end of the adapter. The overhang can be at the 3' end or the 5' end. In some cases, the partial duplex comprises an overhang at each end of the adapter. The overhang can be of equal length or unequal length. In some cases, the 5' end of the ligation strand does not comprise a 5' phosphate group. In some cases, the 5' end of the ligation strand does comprise a 5' phosphate, wherein the 3' end of the polynucleotide lacks a free 3' hydroxyl.

[00128] In some cases, the 3' and/or 5' ends of the non-ligation strand comprise a blocking group and are enzymatically unreactive. The blocking group or moiety on the 3' and/or 5' end can be any blocking group or moiety as described herein and/or known in the art. In some cases, the block at the 5' end of the non-ligation strand comprises a biotin moiety. In some cases, the 5' end of the non-ligation strand does not comprise a 5' phosphate. The 5' end can be removed by treatment with an enzyme. The enzyme can be a phosphatase. In some cases, the 5' end of the non-ligation strand is dephosphorylated by treatment with alkaline phosphatase. In some cases, the 5' end of the non-ligation strand does comprise a 5' phosphate, wherein the 3' end of the polynucleotide lacks a free 3' hydroxyl. In some cases, the non-ligation strand comprises a block at the 3' end comprising terminal dideoxycytosine and a block at the 5' end comprising a biotin moiety. In some cases, distinct adapters as described herein are ligated to a 5' end of a double strand polynucleotide.

[00129] In some cases, the adapter is a hairpin adapter comprising a stem-loop, wherein both strands of the stem comprise a modified dNTP (e.g. a cytosine analog resistant to bisulfite treatment). In some cases, the stem-loop adapter comprises a ligation or first strand and a non-ligation or second strand as described herein. In some cases, the 3' end of the stem comprises the ligation strand, while the 5' end of the stem comprises the non-ligation strand. In some cases, the 5' end of the stem does not comprise a 5' phosphate. In some cases, the 5' end of the stem comprises a 5' phosphate, while the 3' ends of the double strand polynucleotide lacks a free 3' hydroxyl. In some cases, the 5' end of the stem comprises a blocking group. The blocking group can be any of the blocking groups described herein. In some cases, the stem comprises an overhang. The overhang can be a 5' overhang or a 3' overhang and can comprise DNA, RNA, or both. A stem-loop adapter can be ligated to a double stranded polynucleotide by the methods described herein. In some cases, a stem loop adapter comprises a replication block. The replication block can be a non-replicable base or region in the loop or in a region of the stem adjacent to the loop comprising abasic sites. The replication block can comprise an inverted repeat. Abasic sites can be generated in the stem-loop by any of the methods known in the art, which can include, but is not limited to, incorporation of dUTP during generating of the adapter

IDT Exhibit 1017.1602

followed by treatment with dU-glycosylase (which is also referred to as Uracyl-DNA Glycosylase or UDG). In some cases, the replication block is removable or cleavable.

[00130] In some cases, the adapter comprises a ligation or first strand as described herein, and a non-ligation or second strand, wherein the non-ligation or second strand comprises RNA residues. In some cases, the adapter comprises a ligation or first strand as described herein, and a non-ligation or second strand, wherein the ligation or first strand comprises RNA residues.

[00131] In some cases, the ligation of an adapter to a first strand of a double stranded polynucleotide fragments creates a nick or break in the backbone between the non-ligation strand of the adapter and the 3' end of the second strand of the double-stranded polynucleotide fragments, wherein the non-ligation strand is not joined to the 3' end of the second strand of the polynucleotide fragments. In this case, the 5' end of the ligation strand does not comprise a 5' phosphate group. Further to this case, ligation of an adapter to the polynucleotide fragment can generate a polynucleotide fragment comprising the ligation strand comprising a cytosine analog joined to a first and second 5' end of the polynucleotide fragments. In some cases, the 5' end of the ligation strand comprises a 5' phosphate group, and the 3' ends of the polynucleotide fragment lacks a free 3' hydroxyl. Further to this case, ligation of an adapter to the polynucleotide fragment can generate a polynucleotide fragment comprising the ligation strand comprising a cytosine analog joined to a first and second 5' end of the polynucleotide fragments.

[00132] The method can further comprise performing an extension reaction. The extension 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 strand displacement activity or a DNA polymerase with exonuclease activity and all four dNTPs (i.e. dATP, dTTP, dCTP, and dGTP), wherein the dNTPs are unmodified. In some cases, adaptors (e.g., duplex or partial duplex) as described herein comprising cytosine analogs resistant to bisulfite treatment in a ligation strand as well as no or one or more diversity bases are ligated to a polynucleotide and further subject to an extension reaction that is performed with a DNA polymerase and unmodified dNTPs (i.e. dATP, dTTP, dCTP, and dGTP). In some cases, adaptors (e.g., duplex or partial duplex) as described herein comprising cytosine bases sensitive to bisulfite treatment in a ligation strand as well as no or one or more diversity bases are ligated to a polynucleotide and further subject to an extension reaction that is performed with a DNA polymerase and a dNTP mixture comprising dATP, dTTP, dGTP, and a dCTP analog resistant to bisulfite treatment. In some cases, the extension reaction extends the 3' ends of the polynucleotide fragments, whereby a non-ligation strand of an adapter is removed. The non-ligation strand can be removed by being displaced, degraded, or denatured. In some cases, the non-ligation strand of the joined adapter is

IDT Exhibit 1017.1603

removed by heat denaturation, and the 3' ends of the polynucleotide fragment are extended with a polymerase without strand displacement activity. In some cases, the melting temperature of the non-ligation strand bound to the ligation strand can be lower than the melting temperature of the two strands of the polynucleotide fragment to which the ligation strand of the adapter is joined. In some cases, the non-ligation strand is displaced by a polymerase comprising strand displacement activity during extension of the 3' ends of the double stranded polynucleotide fragment. In some cases, the adapter is a hairpin adapter and the extension reaction displaces the non-ligation strand of the stem. In some cases, the displaced strand of the stem adapter remains connected to the ligation strand of the stem via the loop. In some cases, the loop comprises a cleavage site for an enzyme (i.e. restriction endonuclease). In some cases, the cleavage site is within a replication block. In some cases, the cleavage site is cleaved, thereby removing the non-ligation strand of the stem. In these cases, the ligation strand of the stem can comprise the modified nucleotide (i.e. nucleotide with cytosine analog resistant to bisulfite treatment). In some cases, the ligation strand serves as the template, wherein the extension reaction generates sequence complementary to the ligation strand. In some cases a single adapter is ligated to the 5' ends of the double stranded polynucleotide fragment, whereby extension of the 3' ends of the polynucleotide fragment generates polynucleotide fragments comprising complementary adapter sequences at the 3' and 5' ends. In some cases, the adapter ligated to the polynucleotide fragments comprises a non-ligation strand comprising RNA thereby forming a DNA/RNA heteroduplex with the ligation strand, wherein the extension reaction extends the 3' ends of the polynucleotide fragments following degradation of the RNA in the non-ligation strand using an agent capable of degrading RNA in a DNA/RNA heteroduplex. The agent can be an enzyme. The enzyme can be RNase H. In this embodiment, the ligation or first strand serves as the template, wherein the extension reaction generates sequence complementary to the ligation or first strand, thereby generating polynucleotide fragments comprising complementary adapter sequences at the 3' and 5' ends.

[00133] In some cases, the duplex adapter is a partial duplex adapter, wherein the adapter comprises a long strand and a short strand, wherein both the long strand and the short strand are capable of ligation. In some cases, the long strand comprises a modified dNTP (e.g. a cytosine analog resistant to bisulfite treatment). In some cases, the short strand comprises a modified dNTP (e.g. a cytosine analog resistant to bisulfite treatment). In these cases, the partial duplex adapter comprises a 5' overhang and a blunt end, or both a 5' and 3' overhang. In order to reduce the formation of primer dimers, the 3' end of the short arm of the adapter can comprise a blocking group and can be enzymatically unreactive. The blocking group can be any of the

IDT Exhibit 1017.1604

blocking groups described herein. In some cases, the short arm of the adapter comprises a reversible blocking group, wherein the reversible blocking group can be removed following ligation of the adapter to the double stranded polynucleotide. In some cases, unligated adapter is removed by washing and/or degradation following ligation and prior to removal of the reversible blocking group. In some cases, the method can further comprise performing an extension reaction. The extension 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 strand displacement activity and all four dNTPs (i.e. dATP, dTTP, dCTP, and dGTP), wherein the dNTPs are unmodified. In some cases, the extension reaction is performed with a DNA polymerase and unmodified dNTPs (i.e. dATP, dTTP, dCTP, and dGTP). In some cases, the extension reaction extends the 3' ends of short strand of the adapters ligated to the ends of the double stranded polynucleotide fragments, thereby generating polynucleotide fragments comprising complementary adapter sequences at the 3' and 5' ends. The cytosine analog resistant to bisulfite treatment can be 5-methylcytosine, 5-hydroxymethylcytosine or 5-propynylcytosine.

[00134] In some cases, double stranded polynucleotide fragments comprising adapter sequence at the 3' and 5' ends are captured prior to treatment with a converting agent. In some cases, the double stranded polynucleotide fragments are captured using a binding agent directed against modified dNTPs in the double-stranded polynucleotide fragments with adapters. The modified dNTP can be a modified dCTP comprising a cytosine analog. The cytosine analog can be 5-methylcytosine, 5-hydroxymethylcytosine or 5-propynylcytosine. The binding agent can be an antibody, or the binding domain of a protein directed against a cytosine analog. In some cases, the binding domain is directed against 5-methylcytosine residues. The binding domain can be from a methyl-CpG-binding domain (MBD) protein. The MBD protein can be methyl-CpG-binding domain protein 1, 2, 4, or MECP2. In some cases, the double stranded polynucleotide fragments are captured using the binding domain of MBD2. In some cases, the double stranded polynucleotide fragments are captured using the binding domain of MECP2. In some cases, one or both strands of the adapter sequence on the end(s) of the double stranded polynucleotide fragments comprise a cytosine analog other than 5-methylcytosine, wherein the double stranded polynucleotide fragments are captured using the binding domain of a methyl-CpG-binding domain (i.e. MBD2 or MECP2). The cytosine analog other than 5-methylcytosine can be 5-hydroxymethylcytosine or 5-propynylcytosine.

[00135] In some cases, the method further comprises a denaturing step, wherein the polynucleotide fragments comprising adaptor sequences at the 3' and 5' ends are denatured.

IDT Exhibit 1017.1605

Denaturation can be achieved using any of the methods known in the art which can include, but are not limited to, heat denaturation, and/or chemical denaturation. Heat denaturation can be performed by raising the temperature of the reaction mixture to be above the melting temperature of the polynucleotide fragments comprising adapter sequence at both ends. The melting temperature can be about, more than, less than, or at least 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, or 95 degrees C. The temperature can be raised above the melting temperature by about, more than, less than, or at least 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 degrees C. Chemical denaturation can be performed using bases (i.e. NaOH), and/or competitive denaturants (i.e. urea, or formaldehyde). In some cases, denaturation generates single stranded polynucleotide fragments comprising complementary adaptor sequence at the 3' and 5' ends such that the juncture between the fragments and the adaptors comprise one or more diversity bases. In some cases, denaturation generates single stranded polynucleotides fragments comprising distinct adapter sequence at the 3' and 5' ends such that the juncture between the fragments and the adaptors comprise one or more diversity bases.

[00136] In some cases, single stranded polynucleotide fragments comprising adapter sequence at the 3' and 5' ends are captured prior to treatment with a converting agent. In some cases, the polynucleotide fragments are captured by a binding agent directed against one or more modified dNTPs present in the adapter sequence. In some cases, the modified dNTP is a nucleotide base analog. In some cases, the binding agent is a binding protein. In some cases, the binding protein is an antibody directed against the modified dNTP. In some cases, the binding protein is an antibody directed against the modified dNTP, wherein the modified dNTP is a nucleotide analog. In some cases, the single stranded polynucleotide fragments comprising adapter sequence at the 3' and 5' ends are captured prior to treatment with a bisulfite. In some cases, the nucleic acid fragments (polynucleotides) are captured by a binding agent directed against one or more elements present in the adapter sequence. In some cases, the one or more elements comprise a cytosine analog. In some cases, the cytosine analog is 5-methylcytosine. In some cases, the binding agent is a 5-methylcytosine binding protein. In some cases, the binding protein is an anti-5-methylcytosine antibody. In some cases, 5-methylcytosine capture is performed prior to bisulfite treatment, wherein the cytosine analog resistant to bisulfite treatment is a cytosine analog other than 5-methylcytosine. In some cases, the cytosine analog can be 5-hydroxymethylcytosine or 5-propynylcyotsine. The one or more elements can be introduced during the extension reaction. In some cases, a modified nucleotide can be incorporated during

IDT Exhibit 1017.1606

the extension reaction, wherein the modified nucleotide contains a tag. The tag can be a biotin moiety. In some cases, the binding agent is avidin, streptavidin, or an anti-biotin antibody.

[00137] Following denaturation, and optional capture by a binding agent, the single-stranded polynucleotide fragments comprising adapter sequence at the 3' and 5' ends can be treated with a converting agent. In some cases, treatment of the single-stranded polynucleotide fragments with a converting agent alters the sequence of the complement of the ligation strand as well as the first and second strands of the double stranded polynucleotide fragment, while leaving the sequence of the ligation or first strand unchanged. In some cases, a single adapter is ligated to the 5' ends of the polynucleotide fragments, whereby treatment with a converting agent generates single stranded polynucleotide fragments comprising non-complementary sequence at the 5' and 3' ends. In some cases, distinct adapters are ligated to the 5' ends of the polynucleotide fragments, whereby treatment with a converting agent generates single stranded polynucleotide fragments wherein the non-ligation strands of the distinct adapters is altered to be non-complementary to the ligation strands of the distinct adapters. In some embodiments, the sequence of the ligation or first strand of the adapter marks the 5' end of the polynucleotide fragments, thereby maintaining the strandedness of the polynucleotide fragment and thus providing information on directionality.

[00138] In some cases, the single-stranded nucleic acid fragments are treated with a converting agent wherein the converting agent is bisulfite. In some cases, treatment of the single-stranded polynucleotide fragments converts cytosine residues in the polynucleotide fragment and the complement of the ligation or first strand to uracil residues while the cytosine analogs in the ligation or first strand are resistant to conversion. In some cases, treatment of the single stranded polynucleotide fragments with bisulfite generates single stranded polynucleotide fragments comprising non-complementary adapter sequence at the 5' and 3' ends. In some cases, the sequence of the ligation strand of the adapter unaltered by bisulfite treatment marks the 5' end of the polynucleotide fragments, thereby maintaining the strandedness of the polynucleotide fragment and thus providing information on directionality. In some cases, distinct adapters are ligated to the 5' ends of the polynucleotide fragments, whereby treatment with a bisulfite generates single stranded polynucleotide fragments wherein cytosine residues in the non-ligation strands of the distinct adapters are converted to uracil residues, whereby the sequence of the non-ligation strand is no longer complementary to the ligation strands of the distinct adapters.

[00139] In some cases, the method further comprises amplifying the single-stranded polynucleotide fragments comprising adapter sequences at the 3' and 5' ends. In some cases, amplification of the single-stranded polynucleotide fragments comprising adapter sequence at

IDT Exhibit 1017.1607

the 3' and 5' ends generates directional polynucleotide libraries. In some cases, one end of the polynucleotide fragment marks the orientation of the original polynucleotide strand to which it is appended due to its resistance to conversion by the converting agent, whereby the sequence in said end is resistant to conversion to a different sequence by treatment with the converting agent. In some cases, amplification of the single-stranded polynucleotide fragments comprising adapter sequence at the 3' and 5' ends generates directional polynucleotide libraries wherein one end of the polynucleotide fragments marks the orientation of the original polynucleotide strand to which it is appended due to its resistance to conversion by bisulfite treatment. In some cases, the cytosine residues present in said end are resistant to conversion to uracil residues by bisulfite treatment.

[00140] In some cases, amplifying the single stranded polynucleotide fragments comprising adapter sequence at the 3' and 5' ends comprises the use of a first primer and a second primer. In some cases, the first primer is directed against sequence complementary to the ligation or first strand of an adapter altered following treatment with a converting agent. In some cases, the second primer is directed against sequence complementary to the ligation or first strand of an adapter, wherein the ligation or first strand to which said complementary sequence is complementary is not altered by treatment with the converting agent. In some cases, the converting agent is bisulfite, whereby treatment with bisulfite converts cytosine residues in the sequence complementary to the ligation or first strand to uracil residues. In some cases, the first primer is directed against sequence complementary to the ligation or first strand of the adapter comprising uracil residues following bisulfite treatment. In some cases, the second primer is directed against sequence complementary to the ligation or first strand of the adapter, wherein the ligation or first strand to which said complementary sequence is complementary to does not contain uracil residues following bisulfite treatment. The single stranded polynucleotide fragments comprising adapter sequence at the 3' and 5' ends can represent a first strand of a double stranded polynucleotide fragment or a second strand of a double stranded polynucleotide fragment.

[00141] In some cases, a single adapter is ligated to the 5' ends of the polynucleotide, whereby the first and second strands can comprise non-complementary sequence following treatment with the converting agent (e.g., bisulfite treatment). In some cases, distinct adapters are ligated to the 5' ends of the polynucleotide fragments, whereby treatment with bisulfite generates single stranded polynucleotide fragments from a first strand of a double stranded polynucleotide fragment or a second strand of a double stranded polynucleotide fragment, wherein cytosine residues in the non-ligation strands of the distinct adapters are converted to uracil residues. In

IDT Exhibit 1017.1608

these cases, the sequence of the non-ligation strand is no longer complementary to the ligation strands of the distinct adapters. Amplifying the single stranded polynucleotide fragments comprising adapter sequence at the 3' and 5' ends can produce amplification products from either or both of the first and second strand of the double stranded polynucleotide fragment following treatment with the converting agent (i.e. bisulfite). In some cases, the first and/or second primer further comprises one or more identifier sequences. In some cases, the identifier sequences comprise a non-hybridizable tail on the first and/or second primer. The identifier sequence can be a barcode sequence, a flow cell sequence, and/or an index sequence. In some cases, the index sequence is a Truseq primer sequence compatible with the next generation sequencing platform produced by Illumina. In some cases, the first and/or second primer can bind to a solid surface. The solid surface can be a planar surface or a bead. The planar surface can be the surface of a chip, microarray, well, or flow cell. In some cases, the first and/or second primer comprises one or more sequence elements products of the amplification reaction (i.e. amplification products) to a solid surface, wherein the one or more sequences are complementary to one or more capture probes attached to a solid surface.

[00142] In some cases, methods for generating a polynucleotide library using modified duplex-forming adapters described herein further comprise determining the methylation status of the input double stranded polynucleotide. In some cases, the input polynucleotide is genomic DNA and the amplification of single-stranded polynucleotide fragments comprising non-complementary sequence at the 3' and 5' ends is followed by sequencing. The non-complementary sequence at the 3' ends of the single-stranded polynucleotide fragments can be complementary to reverse read sequencing primers, while the 5' ends of the single-stranded polynucleotide fragments can be complementary to forward read sequencing primers. This differentially complementarity to forward and reverse sequencing primers can result from the bisulfite treatment differentially affecting the sequence of the 3' and 5' ends based on the presence or absence of cytosine analogs resistant to bisulfite treatment in the 3' and 5' ends as described herein. Further to this embodiment, the methylation status of the genomic DNA can be determined by comparing the sequence obtained from the sequencing of the single-stranded polynucleotide fragments comprising non-complementary sequence at the 3' and 5' ends representing either or both of the first and second strand of the double stranded polynucleotide following treatment with converting agent (i.e. bisulfite treatment) generated by the methods described herein against a reference sequence. The reference sequence can be the sequence of the genomic DNA (either or both strands) not subjected to alteration by treatment with the converting agent. The comparing can be performed on a computer. The comparing can be done

IDT Exhibit 1017.1609

on a computer using a sequence alignment tool or program. The sequence alignment tool or program can map bisulfite treated sequencing reads to a genome of interest and perform methylation calls. The bisulfite sequencing mapping tool can be the Bismark program. In some cases, the comparing comprises performing a nucleotide alignment between the sequence obtained from the sequencing of the single-stranded DNA fragments comprising non-complementary sequence at the 3' and 5' ends generated by the methods described herein with a reference sequence on a computer using any of the nucleotide alignment programs known in the art (e.g. Bismark). In some cases, the methods described herein can be used to determine the methylation status of a specific locus or region of genomic DNA or the entire genome (i.e. the methylome). In some cases, following bisulfite treatment, the methylation status of a given cytosine residue is inferred by comparing the sequence to an unmodified reference sequence.

[00143] Sequencing can be any method of sequencing, including any of the next generation sequencing (NGS) methods described herein. In some cases, the NGS method comprises sequencing by synthesis. In some embodiments, sequencing is performed with primers directed against known or universal sequence introduced into the nucleic acid fragments by the adapter ligated to the nucleic acid fragments. In some cases, the primers used for sequencing are directed against adapter sequence unaltered by treatment with a converting agent. In some cases, primers used for sequencing are directed against adapter sequence altered by treatment with a converting agent. The converting agent can be bisulfite, wherein bisulfite treatment converts cytosine residues to uracil residues. In some cases, the sequencing primers are directed against adapter sequence comprising thymine residues following bisulfite treatment and amplification. In some cases, the sequencing primers are directed against adapter sequence wherein the adapter sequence is resistant to conversion by bisulfite treatment. In this embodiment, the adapter sequence to which the sequencing primers are directed does not comprise thymine residues following bisulfite treatment and amplification. In some cases, sequencing is performed with primers directed against identifier sequence introduced into the polynucleotide fragments by the first and/or second primer used to amplify single-stranded polynucleotide fragments comprising non-complementary sequence at the 3' and 5' ends. The identifier sequence can be a barcode sequence, a flow cell sequence, and/or index sequence. In some cases, the index sequence is a Truseq primer sequence compatible with the next generation sequencing platform produced by Illumina.

[00144] As described herein, the methods, compositions and kits provided herein can be used to generate reduced representation bisulfite (RRBS) libraries. As shown in **FIG. 23**, generation of an RRBS library can result in representation of about 1% of the genome that is enriched for 40-

IDT Exhibit 1017.1610

300 bp fragments comprising microsatellite bands and aid *in silico* assignment of CpG sites into genomic classes using far less sequencing than can be required using whole genome bisulfite sequencing approach as shown in **FIG. 25**. A comparison of RRBS library generation with other DNA methylation profiling technologies (**FIG. 24**), as reproduced from Gu et al. *Nature Protocols*. 6(4): 468-481(2011), which is herein incorporated by reference in its entirety, reveals that RRBS provides about 9% actual genomic coverage at a resolution of 1 bp with an input of only 0.01-0.3 μg. However, as described herein, RRBS can sometimes create a lack of color balance that can be an issue for some sequencing platforms. The methods, compositions, and kits provided herein can be used to eliminate or substantially reduce these issues.

[00145] A schematic exemplary of an embodiment of the methods described herein for generating an RRBS library using partial duplex-forming adapters comprising diversity bases is shown in **FIG. 3**. The method can comprise digesting a double stranded polynucleotide (e.g., genomic DNA) at one or more recognition sites for a methylation insensitive restriction endonuclease (RE). In some cases, the one or more recognition sites comprise the sequence "CCGG", and the methylation insensitive RE is MspI. The methylation insensitive restriction enzyme can be DpnII, HpaII, Sall-HF®, ScrFI, and the recognition site can be any recognition site recognized by said restriction endonucleases. Cleavage with the RE produces one or more polynucleotide fragments with or without overhangs. As shown in **FIG. 3**, cleavage with MspI generates a double-stranded polynucleotide fragment comprising a 5' overhang dinucleotide comprising 5'-CG-3'. A pool of diversity adaptors as shown in **FIG. 2A-B** can be appended to the double stranded polynucleotide generated by cleavage with MspI. The appending can be through ligation. Ligation can be performed with any ligation method known in the art. The ligation can be cohesive end ligation such as shown in **FIG. 3**. For illustrative purposes, ligation of a single diversity adaptor to both ends of the MspI cleaved double stranded polynucleotide is shown. The cleaved polynucleotide can be ligated on each end by a diversity adaptor comprising a different number of diversity bases than on an opposite end of the same cleaved polynucleotide. The diversity adapter in **FIG. 3** comprises a short strand whose 5' end comprises a terminal "CG" dinucleotide which immediately follows three 'H' nucleotides, which can be either a adenine (A), cytosine (C), or thymine (T) base, but cannot be a guanine (G) base. The long strand of the diversity adaptor terminates with a "DDD" trinucleotide comprising all "D" bases, wherein the "DDD" is the complement of the HHH triplet on the short strand and, therefore can be an A, G, or T base, but cannot be a C base. The long strand of the diversity adapter comprises non-methylated cytosine residues. In some cases, the long strand of the diversity adaptor can comprise cytosine analogs resistant to bisulfite treatment. In some cases,

IDT Exhibit 1017.1611

the cytosine analog resistant to bisulfite treatment is 5-methyl-cytosine. As can be seen in **FIG. 2A-B**, the short strand can comprise 0, 1, 2, or 3 "H" bases, and thus, the long strand can comprise either 0, 1, 2, or 3 "D" bases. **FIG 2A-B** also illustrates that the short strand comprises both a 5' (e.g., Biosg), and 3' (e.g., ddC) group. One or both groups can serve to block ligation and/or extension of the short strand. Following ligation of the diversity adapters in **FIG. 3**, the 3' ends of the polynucleotide fragments can be extended in the presence of a nucleotide mixture comprising dTTP, dGTP, dATP, and 5-methyl dCTP using the long strand of the diversity adapter as template. Extension can produce polynucleotide fragments comprising the long strand diversity adapter sequence comprising non-methylated cytosine residues at the 5' end and sequence complementary to the long strand diversity adapter sequence comprising methylated cytosine residues at the 3' end. In some cases, the long strand of the diversity adaptor comprises 5-methylcytosine and following ligation of the diversity adapters in **FIG. 3**, the 3' ends of the polynucleotide fragments can be extended in the presence of a nucleotide mixture comprising dTTP, dGTP, dATP, and dCTP using the long strand of the diversity adapter as template. Thus, extension can produce polynucleotide fragments comprising the long strand diversity adapter sequence comprising methylated cytosine residues at the 5' end and sequence complementary to the long strand diversity adapter sequence comprising non-methylated cytosine residues at the 3' end. In some cases, both the long strand and the short strand of the diversity adaptor comprises 5-methylcytosines and can ligate to each end of the MspI digested polynucleotide and following ligation of the diversity adapters , the 3' ends of the short strand of the D-adaptor can be extended in the presence of a nucleotide mixture comprising dTTP, dGTP, dATP, and dCTP using the long strand of the diversity adapter as template. Thus, extension can produce polynucleotide fragments comprising the long strand diversity adapter sequence comprising methylated cytosine residues at the 5' end and sequence complementary to the long strand diversity adapter sequence comprising non-methylated cytosine residues at the 3' end. In some cases, both the long strand and the short strand of the diversity adaptor comprises non-methylated cytosines and can ligate to each end of the MspI digested polynucleotide and following ligation of the diversity adapters, the 3' ends of the short strand of the diversity adaptor can be extended in the presence of a nucleotide mixture comprising dTTP, dGTP, dATP, and 5-methyl dCTP using the long strand of the diversity adapter as template. Thus, extension can produce polynucleotide fragments comprising the long strand diversity adapter sequence comprising non-methylated cytosine residues at the 5' end and sequence complementary to the long strand diversity adapter sequence comprising methylated cytosine residues at the 3' end. 5-methyl cytosine in the diversity adaptor and/or the reaction mixture used for extension can be

IDT Exhibit 1017.1612

substituted with any cytosine analog resistant to bisulfite treatment known in the art. The cytosine analog can be 5-hydroxymethylcytosine. The cytosine analog can be 5-propynlcytosine. The polynucleotide fragment resulting from digestion with MspI and ligation with one or more diversity adaptors to each end can be denatured and then subjected to bisulfite treatment as shown in **FIG. 3**. In some cases, fragments of a desired size are selected prior to bisulfite treatment and PCR ampilfication by running the fragments comprising the adapters in a gel and size selecting. Bisulfite treatment can generate single stranded polynucleotide fragments comprising non-complementary diversity adaptor sequence at each terminus. As shown in **FIG. 3**, bisulfite treatment can generates single-stranded polynucleotide fragments comprising long strand diversity adapter sequence at the 5' end wherein each cytosine residue has been converted to a uracil, and sequence which remains complementary to the long strand of the diversity adapter sequence present in the fragment prior to treatment at the 3' end. Any non-methylated cytosine residues present in the polynucleotide fragment between the adapter sequence can also be converted to uracil residues. Following bisulfite treatment, the fragments can be subjected to PCR using primers directed to the adapter sequences following conversion. PCR amplification can generate polynucleotide fragments with distinct adapter sequences on each end. The "D," "R," and "H" bases present in the diversity adapters can retain their original base identity following bisulfite treatment and subsequent amplification. The resultant bisulfite converted library can be further subjected to sequencing. Sequencing can be performed by any of the known next generation sequencing platforms. As can be seen in **FIG. 3**, sequencing the RRBS library generated by the method depicted in **FIG3** can generate predictable forward and reverse reads in the initial cycles. The long strand of the adapter can further comprise additional sequence elements such as barcodes and/or random sequence. Additional sequence elements can also be introduced to RRBS nucleic acid inserts through amplification of the inserts with primers comprising the additonal sequence elements. The additional sequence elements can be one or more barcodes, random sequence, and/or flow cell sequences that can be used to hybridize and thus capture an insert on a solid surface for subsequent solid-phase amplification. **FIG. 27A** illustrates a polynucleotide insert from an RRBS generated using diversity adapters as provided herein. **FIG. 37** illustrates a polynucleotide insert from an RRBS generated using diversity adapters as provided herein that is subsequently attached to a solid surface. The solid surface can be a flow cell surface or a bead. As shown in **FIG. 37**, a capture probe on a flow cell surface can be used to capture (via hybridization) a library insert such as the one shown in **FIG. 27A**. The "N" in **FIG. 37** or **FIG. 27A** can be a D base or a mixture of D and H bases as described herein. **FIG. 27B** illustrates the increased color balance in the initial sequence reads of an RRBS

IDT Exhibit 1017.1613

library generated using diversity adapters as provided herein and comprising polynucleotide inserts as shown in **FIG. 27A**. The increased sequence diversity (as shown in **FIG. 27B** by the increase in nucleotide base composition in the initial sequencing cycles) present in one or more polynucleotide inserts from an RRBS library generated using diversity adaptors can serve to substantially lower levels of genomic, higher diversity control libraries (e.g., PhiX control libraries) typically employed when sequencing RRBS libraries. The amount of higher diversity control libraries can be reduced by about, more than, or less than 1%, 5%, 10 %, 20%, 30%, 40%, 50%, 60%, 70%, 80% , 90%, 95%, or 99%. The amount of higher diversity control libraries can be reduced by about 1-5%, 5-10%, 10-20%, 30-40%, 40-50%, 50-60%, 60-70%, 70-80%, 80-90%, 90-95% or 95-99%. The amount of higher diversity control libraries can be reduced by between 1-5%, 5-10%, 10-20%, 30-40%, 40-50%, 50-60%, 60-70%, 70-80%, 80-90%, 90-95% or 95-99%. **FIG. 19** illustrates the total theoretical diversity possible for introducing to the ends of a polynucleotide upon introduction of a pool of diversity adaptors comprising diversity adapters comprising 0 to 3 "D" bases (D0-D3).

[00146] In some cases, no genomic, higher diversity control libraries (e.g., PhiX control libraries) are required to be added or spiked into sequencing reactions from the RRBS libraries generated by the methods provided herein.

[00147] In some cases, the methods depicted in **FIG. 3** for generating RRBS libraries can be combined wtih a dark sequencing approach as illustrated in **FIG. 26** (see Boyle et al., *Genome Biology*. 13(R92):1-10 (2012), which is herein incorporated by reference in its entirety), which can serve to further reduce a lack of color balance. In some cases, the methods depicted in **FIG. 3** for generating RRBS libraries can be combined with spiking in a high diversity control library (e.g., Phi X) during sequencing, which can serve to further reduce a lack of color balance.

[00148] The method depicted in **FIG. 3** and described herein can also be used in the context of generating whole genome bisulfite sequencing libraries from samples spanning a 100-fold range of inputs, down to as little as 1ng of genomic DNA. Library performance metrics such as mapping, complexity, and 5-methylC content can be altered using the methods provided herein. In some cases, the methods provided herein improve Library performance metrics. The Library performance metrics can be mapping, complexity, and 5-methylC content.

VII. Downstream Applications

[00149] The methods and compositions disclosed herein can be efficiently and cost-effectively utilized for downstream analyses, such as next generation sequencing or hybridization platforms, with minimal loss of biological material of interest. The methods described herein can be useful

IDT Exhibit 1017.1614

for generating high throughput reduced representation sequencing libraries from bisulfite-converted DNA, for methylation analysis at specific regions across the genome or methylome.

[00150] For example, the methods described herein can be useful for sequencing by the method commercialized by Illumina, as described U.S. Pat. Nos. 5,750,341; 6,306,597; and 5,969,119. Directional (strand-specific) cDNA libraries are prepared using the methods described herein, and the selected single-stranded nucleic acid is amplified, for example, by PCR. The resulting nucleic acid can then be denatured and the single-stranded amplified polynucleotides can be randomly attached to the inside surface of flow-cell channels. Unlabeled nucleotides can be 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 can be added. After laser excitation, fluorescence from each cluster on the flow cell can be imaged. The identity of the first base for each cluster can then be recorded. Cycles of sequencing are performed to determine the fragment sequence one base at a time. The identity of the bases of the first few cycles (4-5) can then be used to discern separate clusters and facilitate sequencing of individual clusters. Additionally, clusters that share identical or substantially similar sequences at one or more ends as other clusters can experience problems with focusing of the laser and subsequent detection of fluorescence when sequencing such clusters in parallel due to the presence of the same base at each cycle. In some cases, the introduction of one or more diversity bases allows clear identification of individual clusters. For library inserts comprising low diversity portions at one or more ends of the inserts, the introduction of diversity bases at the low diversity ends can allow identification of individual clusters and allow different nucleotide bases to be incorporated in each cycle between inserts with identical or substantially similar sequences at their ends. The incorporation in each cycle of different bases between inserts with identical or substantially similar sequences at their ends can facilitate enhanced laser focusing during each sequencing cycle and thus accurate sequencing of individual clusters.

[00151] In some embodiments, the methods described herein are useful for preparing target polynucleotides for sequencing by the sequencing by ligation methods commercialized by Applied Biosystems (e.g., SOLiD sequencing). In other embodiments, the methods are useful for preparing target polynucleotides for sequencing by synthesis using the methods commercialized by 454/Roche Life Sciences, including but not limited to the methods and apparatus described in Margulies et al., *Nature*(2005) 437:376-380 (2005); and U.S. Pat. Nos. 7,244,559; 7,335,762; 7,211,390; 7,244,567; 7,264,929; and 7,323,305. In other embodiments, the methods are useful for preparing target polynucleotide(s) for sequencing by the methods commercialized by Helicos

IDT Exhibit 1017.1615

BioSciences Corporation (Cambridge, Mass.) as described in U.S. application Ser. No. 11/167,046, and U.S. Pat. Nos. 7,501,245; 7,491,498; 7,276,720; and in U.S. Patent Application Publication Nos. US20090061439; US20080087826; US20060286566; US20060024711; US20060024678; US20080213770; and US20080103058. In other embodiments, the methods are useful for preparing target polynucleotide(s) for sequencing by the methods commercialized by Pacific Biosciences as described in U.S. Pat. Nos. 7,462,452; 7,476,504; 7,405,281; 7,170,050; 7,462,468; 7,476,503; 7,315,019; 7,302,146; 7,313,308; and US Application Publication Nos. US20090029385; US20090068655; US20090024331; and US20080206764.

[00152] Another example of a sequencing technique that can be used in the methods described hereinis 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 can be 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.

[00153] Another example of a sequencing technique that can be used in the methods described herein 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.

IDT Exhibit 1017.1616

[00154] Another example of a sequencing technique that can be used in the methods described herein is DNA nanoball sequencing (as performed, e.g., by Complete Genomics; see e.g., Drmanac et al. (2010) Science 327: 78-81). DNA can be isolated, fragmented, and size selected. For example, DNA can be fragmented (e.g., by sonication) to a mean length of about 500 bp. Adapters (Adl) can be attached to the ends of the fragments. For example, DNA can be fragmented with MspI and size selected to a mean length of about 500 bp. Adapters (Adl) can be attached to the ends of the fragments. The adapters can be used to hybridize to anchors for sequencing reactions. DNA with adapters bound to each end can be PCR amplified. The adapter sequences can be modified so that complementary single strand ends bind to each other forming circular DNA. The adapter sequences can be further modified so that they comprise diversity bases. In some cases, the adaptors comprise a pool of adaptors such that some of the adaptors in the pool comprise no diversity bases, some adaptors comprise 1 diversity base, some adaptors comprise 2 diversity bases, and some comprise 3 diversity bases. The DNA can be methylated to protect it from cleavage by a type IIS restriction enzyme used in a subsequent step. An adapter (e.g., the right adapter) can have a restriction recognition site, and the restriction recognition site can remain non-methylated. The non-methylated restriction recognition site in the adapter can be recognized by a restriction enzyme (e.g., AcuI), and the DNA can be cleaved by AcuI 13 bp to the right of the right adapter to form linear double stranded DNA. A second round of right and left adapters (Ad2) can be ligated onto either end of the linear DNA, and all DNA with both adapters bound can be PCR amplified (e.g., by PCR). Ad2 sequences can be modified to allow them to bind each other and form circular DNA. The DNA can be methylated, but a restriction enzyme recognition site can remain non-methylated on the left Adl adapter. A restriction enzyme (e.g., AcuI) can be applied, and the DNA can be cleaved 13 bp to the left of the Adl to form a linear DNA fragment. A third round of right and left adapter (Ad3) can be ligated to the right and left flank of the linear DNA, and the resulting fragment can be PCR amplified. The adapters can be modified so that they can bind to each other and form circular DNA. A type III restriction enzyme (e.g., EcoP15) can be added; EcoP15 can cleave the DNA 26 bp to the left of Ad3 and 26 bp to the right of Ad2. This cleavage can remove a large segment of DNA and linearize the DNA once again. A fourth round of right and left adapters (Ad4) can be ligated to the DNA, the DNA can be amplified (e.g., by PCR), and modified so that they bind each other and form the completed circular DNA template. Rolling circle replication (e.g., using Phi 29 DNA polymerase) can be used to amplify small fragments of DNA. The four adapter sequences can contain palindromic sequences that can hybridize and a single strand can fold onto itself to form a DNA nanoball (DNB(TM)) which can be approximately 200-300 nanometers in diameter on

IDT Exhibit 1017.1617

average. A DNA nanoball can be attached (e.g., by adsorption) to a microarray (sequencing flowcell). The flow cell can be a silicon wafer coated with silicon dioxide, titanium and hexamehtyldisilazane (HMDS) and a photoresist material. Sequencing can be performed by unchained sequencing by ligating fluorescent probes to the DNA. The color of the fluorescence of an interrogated position can be visualized by a high resolution camera. The identity of nucleotide sequences between adapter sequences can be determined.

[00155] In some cases, the sequencing technique can comprise paired-end sequencing in which both the forward and reverse template strand can be sequenced. In some cases, the sequencing technique can comprise mate pair library sequencing. In mate pair library sequencing, DNA can be fragments, and 2-5 kb fragments can be end-repaired (e.g., with biotin labeled dNTPs). The DNA fragments can be circularized, and non-circularized DNA can be removed by digestion. Circular DNA can be fragmented and purified (e.g., using the biotin labels). Purified fragments can be end-repaired and ligated to sequencing adapters.

[00156] In some cases, a sequence read is about, more than about, less than about, or at least about 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 223, 224, 225, 226, 227, 228, 229, 230, 231, 232, 233, 234, 235, 236, 237, 238, 239, 240, 241, 242, 243, 244, 245, 246, 247, 248, 249, 250, 251, 252, 253, 254, 255, 256, 257, 258, 259, 260, 261, 262, 263, 264, 265, 266, 267, 268, 269, 270, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, 285, 286, 287, 288, 289, 290, 291, 292, 293, 294, 295, 296, 297, 298, 299, 300, 301, 302, 303, 304, 305, 306, 307, 308, 309, 310, 311, 312, 313, 314, 315, 316, 317, 318, 319, 320, 321, 322, 323, 324, 325, 326, 327, 328, 329, 330, 331, 332, 333, 334, 335, 336, 337, 338, 339, 340, 341, 342, 343, 344, 345, 346, 347, 348, 349, 350, 351, 352, 353, 354, 355, 356, 357, 358, 359, 360, 361, 362, 363, 364, 365, 366, 367, 368, 369, 370, 371, 372, 373, 374, 375, 376, 377, 378, 379, 380, 381, 382, 383, 384, 385, 386, 387, 388, 389, 390, 391, 392, 393, 394, 395, 396, 397, 398, 399, 400, 401, 402, 403, 404, 405, 406, 407, 408, 409, 410, 411, 412,

IDT Exhibit 1017.1618

413, 414, 415, 416, 417, 418, 419, 420, 421, 422, 423, 424, 425, 426, 427, 428, 429, 430, 431, 432, 433, 434, 435, 436, 437, 438, 439, 440, 441, 442, 443, 444, 445, 446, 447, 448, 449, 450, 451, 452, 453, 454, 455, 456, 457, 458, 459, 460, 461, 462, 463, 464, 465, 466, 467, 468, 469, 470, 471, 472, 473, 474, 475, 476, 477, 478, 479, 480, 481, 482, 483, 484, 485, 486, 487, 488, 489, 490, 491, 492, 493, 494, 495, 496, 497, 498, 499, 500, 525, 550, 575, 600, 625, 650, 675, 700, 725, 750, 775, 800, 825, 850, 875, 900, 925, 950, 975, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000, 2100, 2200, 2300, 2400, 2500, 2600, 2700, 2800, 2900, or 3000 bases. In some cases, a sequence read is about 10 to about 50 bases, about 10 to about 100 bases, about 10 to about 200 bases, about 10 to about 300 bases, about 10 to about 400 bases, about 10 to about 500 bases, about 10 to about 600 bases, about 10 to about 700 bases, about 10 to about 800 bases, about 10 to about 900 bases, about 10 to about 1000 bases, about 10 to about 1500 bases, about 10 to about 2000 bases, about 50 to about 100 bases, about 50 to about 150 bases, about 50 to about 200 bases, about 50 to about 500 bases, about 50 to about 1000 bases, about 100 to about 200 bases, about 100 to about 300 bases, about 100 to about 400 bases, about 100 to about 500 bases, about 100 to about 600 bases, about 100 to about 700 bases, about 100 to about 800 bases, about 100 to about 900 bases, or about 100 to about 1000 bases.

[00157] The number of sequence reads from a sample can be about, more than about, less than about, or at least about 100, 1000, 5,000, 10,000, 20,000, 30,000, 40,000, 50,000, 60,000, 70,000, 80,000, 90,000, 100,000, 200,000, 300,000, 400,000, 500,000, 600,000, 700,000, 800,000, 900,000,
1,000,000, 2,000,000, 3,000,000, 4,000,000, 5,000,000, 6,000,000, 7,000,000, 8,000,000, 9,000,000, or 10,000,000.

[00158] The depth of sequencing of a sample can be about, more than about, less than about, or at least about 1x, 2x, 3x, 4x, 5x, 6x, 7x, 8x, 9x, 10x, 11x, 12x, 13x, 14x, 15x, 16x, 17x, 18x, 19x, 20x, 21x, 22x, 23x, 24x, 25x, 26x, 27x, 28x, 29x, 30x, 31x, 32x, 33x, 34x, 35x, 36x, 37x, 38x, 39x, 40x, 41x, 42x, 43x, 44x, 45x, 46x, 47x, 48x, 49x, 50x, 51x, 52x, 53x, 54x, 55x, 56x, 57x, 58x, 59x, 60x, 61x, 62x, 63x, 64x, 65x, 66x, 67x, 68x, 69x, 70x, 71x, 72x, 73x, 74x, 75x, 76x, 77x, 78x, 79x, 80x, 81x, 82x, 83x, 84x, 85x, 86x, 87x, 88x, 89x, 90x, 91x, 92x, 93x, 94x, 95x, 96x, 97x, 98x, 99x, 100x, 110x, 120x, 130x, 140x, 150x, 160x, 170x, 180x, 190x, 200x, 300x, 400x, 500x, 600x, 700x, 800x, 900x, 1000x, 1500x, 2000x, 2500x, 3000x, 3500x, 4000x, 4500x, 5000x, 5500x, 6000x, 6500x, 7000x, 7500x, 8000x, 8500x, 9000x, 9500x, or 10,000x. The depth of sequencing of a sample can about 1x to about 5x, about 1x to about 10x, about 1x to about 20x, about 5x to about 10x, about 5x to about 20x, about 5x to about 30x, about 10x to about 20x, about 10x to about 25x, about 10x to about 30x, about 10x to about 40x, about 30x to

about 100x, about 100x to about 200x, about 100x to about 500x, about 500x to about 1000x, about 1000x, to about 2000x, about 1000x to about 5000x, or about 5000x to about 10,000x. Depth of sequencing can be the number of times a sequence (e.g., a genome) is sequenced. In some cases, the Lander/Waterman equation is used for computing coverage. The general equation can be: C = LN/G, where C = coverage; G = haploid genome length; L = read length; and N = number of reads.

[00159] In some cases, the methods described herein can be useful for enhancing sequence diversity in a sample. For example, the sample can be a sequencing library comprising polynucleotides sharing the same sequencing terminal 1, 2, or 3 bases before sequencing inserts. The polynucleotides can be fragments digested by MspI. 0, 1, 2, 3 diversity bases can be added before the shared sequencing terminal bases. For example, the diversity bases can be added between a sequencing primer and a sequencing insert. In some cases, the sequencing terminal bases of the polynucleotides are the diversity bases, not the shared bases. Thus, sequence diversity can be enhanced in the library. The library can be used in next generation sequencing. For example, when used for Illumina sequencing, the library with enhanced diversity can allow better cluster identification, phasing and color matrix calculation by the sequencer.

[00160] In some cases, the methods described herein can be useful for *in silico* analysis of sequence reads. In one aspect, computer-implemented software can be used to remove certain bases (e.g., bases added from an adaptor) from a sequence read. In some cases, Trim Galore (www.bioinformatics.babraham.ac.uk/projects/trim_galore) can be used to trim an adaptor sequence from a sequence read. In another case, python scriptscan be used to trim (i.e., remove) diversity bases from a sequence read. The computer implemented software can follow rules for removing certain bases from sequencing reads such as the rules shown in **FIG. 18A-C**. As shown in **FIG. 18A**, the rules can be used on library inserts that comprise adaptors (e.g., diversity adaptors or adaptors comprising extender sequence) on each end. The adaptors on each end can be different lengths (i.e., different numbers of nucleotides). As shown in **FIG. 18A**, the computer implemented software can be designed to look from a YGG triplet of bases at the 5' end of the sequence and trim or remove 0-2 D bases and/or an R base before the YGG triplet. The Y in the YGG triplet can be a cytosine (C) or thymine base (T). Within the 6 nucleotides at the 3' end of the forward read, the computer implemented software can look for a TCG triplet and trim it as well as the 0-3 "H" bases after it. Outside of the D, R, and H bases, the computer implemented software can be further designed to remove any additional adaptor sequence from or a plurality of sequence reads. In some cases, the MspI fragment can be larger than the read length and thus the software can look for TC at the 3' end and remove it. If TC is not found, then

IDT Exhibit 1017.1620

the software can look for a "T" and remove it. As shown in **FIG. 18B**, with regards to a reverse read, the computer-implemented software can look for a CGA triplet within the 6 nucleotides at the 5' end and trim the 0-3 D bases as well as the CGA triplet since this contains no genomic info other than the fact that it is an MspI site. Within the 6 nucleotides at the 3' end of the reverse reads, the computer implemented software can look for CCR and trim the 0-3 H bases after it. Again, the computer implemented software can be further designed to remove any additional adaptor sequence from or a plurality of sequence reads. With regards to the reverse read, if the software cannot find a CCR, the MspI fragment may be larger than the read length. In general, the number of D (or D and R base combination) removed from the 5' end of forward reads can be the same as the number of H bases removed from 3' end of the reverse read (if the read spans across the entire MspI fragment). Additionally, the number of H bases removed from the 3' end of a forward read (if the read spans across the entire MspI fragment) can be the same as the number of D bases (or D and R base combination) removed from the 3' end of the reverse read. In some cases, following trimming, a library insert can align with respect to MspI sites in a genome from which the insert was derived as shown in **FIG. 18C**.

[00161] In another aspect, the methods provided herein can be useful for removing duplicated sequence reads. In these cases, a stretch of nucleotides whose base compositions is selected at random is inserted within one or both ends of a nucleic acid library insert generated using the method provided herein. This stretch of nucleotides whose base compositions is selected at random can be inserted adjacent to another sequence element (e.g., a barcode) and can be located proximally to a sequence that is capable of binding a sequencing primer. An example of a library insert generated using the methods provided herein for generating an RRBS library that comprises this stretch of nucleotides whose base compositions is selected at random is shown in **FIG. 32** and **FIG. 37**. This stretch of nucleotides whose base compositions is selected at random can be present in an adaptor or pool of adaptors comprising no or varying numbers of diversity bases (or extender sequence) and can be introduced to the end or ends of a polynucleotide upon introduction of the adaptor or pool of adaptors to the end or ends of the polynucleotide using the methods provided herein. This stretch of nucleotides whose base compositions is selected at random can be present in a primer or pool of primers comprising no or varying numbers of diversity bases (or extender sequence) and can be introduced to the end or ends of a polynucleotide upon introduction of the primer or pool of primers to the end or ends of the polynucleotide using the methods provided herein. This stretch of nucleotides whose base compositions is selected at random can be considered an additional sequence element as described herein. In some cases, this stretch of nucleotides whose base compositions is selected

IDT Exhibit 1017.1621

at random is introduced to the end or ends of a nucleic acid library insert during amplification of the nucleic acid library inserts using a primer or set of primers comprising the stretch of nucleotides whose base compositions is selected at random. The length of the stretch of nucleotides whose base compositions is selected at random can comprise 1, 2, 3, 4, 5, 6, 7, 8, 9, 9, 10, 11, or 12 bases. This random sequence can be common to each adaptor in a pool of adaptors or can be different between each adaptor within a pool of adaptors. This random sequence can be different between pools of adaptors. This random sequence can be common to each primer in a pool of primers or can be different between each primer within a pool of primers. This random sequence can be different between pools of primers. The presence of this random sequence can be utilized by computer-implemented software during the analysis of sequence reads. The computer-implemented software can be used to analyze the random sequence associated with all sequence reads that map to the same genomic location. Any reads that both map to the same genomic location and contain the same random sequence are considered duplicates, and all but one copy are removed.

[00162] In some cases, different barcodes can be added to polynucleotides in different samples (e.g., by using primers and/or adapters), and the different samples can be pooled and analyzed in a multiplexed assay. The barcode can allow the determination of the sample from which a polynucleotide originated.

[00163] The compositions, kits, and methods provided herein can be used to treat, prevent, diagnose, and/or prognose a variety of methylation related diseases. Such methylation related diseases can be cancer, mental retardation, neurodegenerative disorders, imprinting disorders, and syndromes involving chromosomal abnormalities. Such methylation related diseases can be Immunodeficiency-centromeric instability-facial anomalies syndrome (ICF), Rett syndrome, Beckwith-Wiedemann Syndrome (BWS), ATRX-linked mental retardation, fragile X syndrome. The cancer can be breast, ovarian, lung, head and neck, testicular, colon, or brain cancer. The cancer can be medulloblastoma, hepatoblastoma, uterine leiomyosarcomata, cervical carcinoma, renal cell carcinoma, rhadbomyosarcoma, gliomas, colorectal cancer, Wilm's tumour, Burkitt's lymphoma, or leukemia. In some cases, the methods described herein are used to determine the status of one or more genes associated with methylation related disorders. The status can include the presence or absence of a nucleic acid modification (i.e. methylation) at one or more bases in a nucleic acid sequence. In some cases, the methods disclosed herein are used to determine or recommend a course of treatment or administration of a therapy based on the status of one or more genes. The therapy can reduce one or more signs or symptoms of a methylation related disease. The therapy can prevent one or more signs or symptoms of any methylation related

IDT Exhibit 1017.1622

diseases. In some cases, the methods disclosed herein are used to determine the outcome or progress of a course of treatment or administration of a therapy based on the status of one or more genes. Genes associated with methylation related diseases can be, but are not limited to *Socs1, Cdkn1c, Slc22a11, Bmp3b, Wit1, Rassf1a, Brca1, p16, Dapk, Mgmt, D4z4, Nbl2, H19, Igf2, G6pd, Rasgrf1 Sybl1, Ar, Pgk1, Dyz2, or Fmr1*. In some cases, the status of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 of any of the genes associated with methylation related diseases are analyzed.

[00164] The methods, kits, and compositions described herein can be used to prevent the development of one or more signs and/or symptoms of methylation related diseases or reduce the severity of one or more signs and/or symptoms of methylation related diseases. The severity of the sign and/or symptom can be reduced by about, or more than about, or at least about, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, or 100 percent. The severity of the sign or symptom can be decreased by about 1 percent to about 10 percent, about 1 percent to about 20 percent, about 1 percent to about 30 percent, about 1 percent to about 50 percent, about 1 percent to about 90 percent, about 1 percent to about 99 percent, about 10 percent to about 20 percent, about 10 percent to about 30 percent, about 10 percent to about 50 percent, about 50 percent to about 75 percent, about 75 percent to about 90 percent, about 75 percent to about 99 percent. The severity of the sign and/or symptom can be reduced by about, more than about, or at least about 2-fold, 3-fold, 4-fold, 5-fold, 6-fold, 7-fold, 8-fold, 9-fold, 10-fold, 11-fold, 12-fold, 13-fold, 14-fold, 15-fold, 16-fold, 17-fold, 18-fold, 19-fold, 20-fold, 21-fold, 22-fold, 23-fold, 24-fold, 25-fold, 30-fold, 35-fold, 40-fold, 45-fold, 50-fold, 55-fold, 60-fold, 65-fold, 70-fold, 75-fold, 80-fold, 85-fold, 90-fold, 95-fold, 100-fold, 200-fold, 300-fold, 400-fold, 500-fold, 600-fold, 700-fold, 800-fold, 900-fold, or 1000-fold. The severity of the sign and/or symptom can be reduced by about 2-fold to 10-fold, about 2-fold to about 50-fold, about 2-fold to about 100-fold, about 10-fold to about 20-fold, about 10-fold to about 50-fold, about 10-fold to about 75-fold, about 10-fold to about 100-fold, about 50-fold to about 75-fold, about 50-fold to about 100-fold, about 100-fold to about 500-fold, about 100-fold to about 1000-fold, or about 500-fold to about 1000-fold.

[00165] The methods, kits, and compositions described herein can be used to decrease the likelihood that a subject will develop one or more signs and/or symptoms of methylation related diseases. The decrease in likelihood can be about, or more than about, or at least about 1, 2, 3, 4,

IDT Exhibit 1017.1623

5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, or 100 percent. The decrease in likelihood can be about 1 percent to about 10 percent, about 1 percent to about 20 percent, about 1 percent to about 30 percent, about 1 percent to about 50 percent, about 1 percent to about 90 percent, about 1 percent to about 99 percent, about 10 percent to about 20 percent, about 10 percent to about 30 percent, about 10 percent to about 50 percent, about 50 percent to about 75 percent, about 75 percent to about 90 percent, about 75 percent to about 99 percent. The decrease in likelihood can be about, more than about, or at least about 1-fold, 2-fold, 3-fold, 4-fold, 5-fold, 6-fold, 7-fold, 8-fold, 9-fold, 10-fold, 11-fold, 12-fold, 13-fold, 14-fold, 15-fold, 16-fold, 17-fold, 18-fold, 19-fold, 20-fold, 21-fold, 22-fold, 23-fold, 24-fold, 25-fold, 30-fold, 35-fold, 40-fold, 45-fold, 50-fold, 55-fold, 60-fold, 65-fold, 70-fold, 75-fold, 80-fold, 85-fold, 90-fold, 95-fold, 100-fold, 200-fold, 300-fold, 400-fold, 500-fold, 600-fold, 700-fold, 800-fold, 900-fold, or 1000-fold. The decrease in likelihood can be about 2-fold to 10-fold, about 2-fold to about 50-fold, about 2-fold to about 100-fold, about 10-fold to about 20-fold, about 10-fold to about 50-fold, about 10-fold to about 75-fold, about 10-fold to about 100-fold, about 50-fold to about 75-fold, about 50-fold to about 100-fold, about 100-fold to about 500-fold, about 100-fold to about 1000-fold, or about 500-fold to about 1000-fold.

[00166] A diagnosis and/or prognosis of a methylation associated neurological in a subject can be made by a health care provider, e.g., a developmental-behavioral pediatrician, a neurologist, a pediatric psychologist, or a psychiatrist. A diagnosis and/or prognosis of a neurological condition can be made or supported by a genetic test performed by a diagnostic laboratory. In some cases, a neurological assessment is administered to a subject by an individual trained and certified to administer a neurological assessment.

[00167] In some cases, a procedure can be performed to diagnose a methylation associated neurological condition in a subject, e.g., angiography, biopsy, a brain scan (e.g., computed tomography (CT), magnetic resonance imaging (MRI), positron emission tomography (PET)), cerebrospinal fluid analysis (by, e.g., lumbar puncture or spinal tap), discography, intrathecal contrast-enhanced CT scan (cisternograhpy), electronencephalography (EEG), electromyography (EMG), nerve conduction velocity (NCV) test, electronystagmography (ENG), evoked potentials (evoked response; e.g., auditory evoked potentials, visual evoked potentials, somatosensory evoked potentials), myelography, polysomnogram, single photon emission computed tomography (SPECT), thermography, or ultrasound imaging (e.g., neurosonography, transcranial

IDT Exhibit 1017.1624

Doppler ultrasound). One or more procedures that can diagnose a neurological condition can be performed on a subject.

[00168] Instruments that can be used in neurological examination can include, e.g., a tuning fork, flashlight, reflex hammer, ophthalmoscope, X-ray, fluoroscope, or a needle.

[00169] The methods, kits, and compositions provided herein can be used to treat, prevent, diagnose, and/or prognose a methylation associate disease or condition in a subject. The subject can be a male or female. The subject can have, or be suspected of having, a methylation associated disease. The subject can have a relative (e.g., a brother, sister, monozygotic twin, dizygotic twin, father, mother, cousin, aunt, uncle, grandfather, grandmother) that was diagnosed with a methylation associate disease. The subject can be, for example, a newborn (birth to about 1 month old), an infant (about 1 to 12 months old), a child (about 1 year old to 12 years old), a teenager (about 13 years old to 19 years old), an adult (about 20 years old to about 64 years old), or an elderly person (about 65 years old and older). The subject can be, for example, about 1 day to about120 years old, about 1 day to about 110 years old, about 1 day to about 100 years old, about 1 day to about 90 years old, about 1 day to about 80 years old, about 1 day to about 70 years old, about 1 day to about 60 years old, about 1 day to about 50 years old, about 1 day to about 40 years old, about 1 day to about 30 years old, about 1 day to about 20 years old, about 1 day to about 15 years old, about 1 day to about 10 years old, about 1 day to about 9 years old, about 1 day to about 8 years old, about 1 day to about 7 years old, about 1 day to about 6 years old, about 1 day to about 5 years old, about 1 day to about 4 years old, about 1 day to about 3 years old, about 1 year to about 2 years old, about 3 years to about 15 years old, about 3 years to about 10 years old, about 3 years to about 7 years old, or about 3 years to about 5 years old. The subject can be about, more than about, at least about, or less than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, or 120 years old.

[00170] The methods for generating polynucleotide libraries as described herein can be used for detecting the presence of fetal DNA in a maternal sample. In some cases, the method comprises: (a) generating bisulfite treated DNA libraries as described herein using a sample obtained from a pregnant woman comprising maternal and fetal DNA; (b) detecting the methylation status of DNA sequence of one or more genes from the sample comprising maternal and fetal DNA; and (c) comparing the methylation status the one or more genes from the sample comprising

maternal and fetal DNA to a reference maternal DNA sample comprising only maternal DNA. In some cases, step (b) of the method comprises an amplification process. In some cases, the amplification process is a polymerase chain reaction (PCR), such as real-time PCR. In other embodiments, step (b) determines the quantity of the DNA sequence. In some cases, the methods provided herein can be used to determine the Rhesus D (RhD) blood group compatibility between a pregnant woman and a fetus. In some cases, the methods for generating directional polynucleotide libraries as described herein can be used for diagnosing, monitoring, or risk assessment of a number of prenatal conditions. For example, the prenatal conditions can include, but are not limited to, beta-thalassemia, cystic fibrosis, congenital adrenal hyperplasia, chromosomal aneuploidies, preeclampsia, preterm labor, and intrauterine growth retardation (IUGR). In some cases, the method comprises (a) generating directional, bisulfite treated DNA libraries as described herein using a sample obtained from a pregnant woman comprising maternal and fetal DNA; (b) detecting the amount of DNA sequence of one or more genes from the sample comprising maternal and fetal DNA; and (c) comparing the amount of the DNA sequence with a standard control, wherein an increase from the control indicates the presence of or an increased risk for developing the pregnancy-associated condition. In some cases, step (b) of the method comprises an amplification process, which can be accomplished by various means, including polymerase chain reaction (PCR), such as real-time PCR. The one or more genes can be *RASSF1A, APC, CASP8, RARB, SCGB3A1, DAB2IP, PTPN6, THY1, TMEFF2, or PYCARD.* The sample can be whole blood, plasma, serum, urine, or saliva. The DNA can be cell-free DNA and/or DNA derived from maternal and fetal cells present in the sample from the pregnant woman. "Standard control value" as used herein refers to a predetermined amount of a genomic sequence that is originated from a fetus and is present in an established sample. The standard control value is suitable for the use of a method described herein, in order for comparing the amount of a gene of interest (or a non-coding sequence) that is present in a test sample. The standard control can provide an average amount of a fetal gene of interest that is typical for a defined time (e.g., first trimester) during pregnancy in the blood of an average, healthy pregnant woman carrying a normal fetus, both of whom are not at risk of developing any pregnancy-associated disorders or complications. A standard control value can vary depending on the genomic sequence of interest and the nature of the sample.

[00171] The methods for generating polynucleotide libraries as described herein can be combined with one or more methods for measuring DNA methylation at specific genomic loci. For example, the methods for measuring DNA methylation can include, but are not limited to,

IDT Exhibit 1017.1626

immunoprecipitation of methylated DNA, methyl-binding protein enrichment of methylated fragments, and/or digestion with methylation-sensitive restriction enzymes.

[00172] The methods for generating polynucleotide libraries as described herein can be combined with one or more methods for profiling methylation status of the whole genome, i.e. the methylome. For example, the methods provided herein can be combined with reduced representation bisuflite sequencing (RRBS). RRBS involves digestion of a DNA sample with a methylation-insensitive restriction endonuclease that has CpG dinucleotide as a part of its recognition site, followed by bisulfite sequencing of the selected fragments (Meissner *et al.*, *Nucleic Acids Res.* 33(18):5868-5877, 2005).

VIII. Compositions and Reaction mixtures

[00173] The present methods further provide one or more compositions or reaction mixtures. In some cases, the reaction mixture comprises: (a) a pool of duplex adaptors comprising duplex adaptors comprising 0, 1, 2, or 3 diversity bases in a ligation strand, wherein the ligation strand comprises cytosine analogs resistant to bisulfite treatment and a non-ligation strand wherein the non-ligation strand is blocked at the 3' and 5' ends and is enzymatically unreactive; (b) a strand displacing polymerase; (c) unmodified dNTPs; and (d) bisulfite. The duplex adaptors in the pool can be partial duplex adaptors comprising a long strand and a short strand as provided herein, wherein the long strand comprises the ligation strand, while the short strand comprises the non-ligation strand. In some cases, the strand displacing polymerase in the reaction mixture is substituted with a polymerase comprising exonuclease activity (e.g., Taq polymerase). In some cases, the reaction mixture further comprises (e) amplification primers directed to unique priming sites created at each end of the DNA fragments following bisulfite treatment. In some cases, at least one of the amplification primers is directed against adaptor sequence following bisulfite treatment, whereby cytosine residues have been converted to uracil residues. In some cases, the reaction mixture further comprises (f) sequencing primers directed against sequences present in the adaptor sequence. In some cases, at least one of the sequencing primers is directed against adaptor sequence following bisulfite treatment, whereby cytosine residues have been converted to uracil residues and subsequently replaced with thymine residues following amplification.

[00174] In some cases, the reaction mixture comprises: (a) a pool of duplex adaptors comprising duplex adaptors comprising 0, 1, 2, or 3 diversity bases in a ligation strand and a non-ligation strand wherein the non-ligation strand is blocked at the 3' and 5' ends and is enzymatically unreactive; (b) a strand displacing polymerase; (c) modified dCTP (i.e. 5-methyl-dCTP, 5-hydroxymethyl-dCTP, or 5-propynyl-dCTP); (d) dATP, dGTP, and dTTP; and (e) bisulfite. The

IDT Exhibit 1017.1627

duplex adaptors in the pool can be partial duplex adaptors comprising a long strand and a short strand as provided herein, wherein the long strand comprises the ligation strand, while the short strand comprises the non-ligation strand. In some cases, the strand displacing polymerase in the reaction mixture is substituted with a polymerase comprising exonuclease activity (e.g., Taq polymerase). In some cases, the reaction mixture further comprises (f) amplification primers directed to unique priming sites created at each end of the DNA fragments following bisulfite treatment. In some cases, at least one of the amplification primers is directed against adaptor sequence following bisulfite treatment, whereby cytosine residues have been converted to uracil residues. In some cases, the reaction mixture further comprises (g) sequencing primers directed against sequences present in the adaptor sequence. In some cases, at least one of the sequencing primers is directed against adaptor sequence following bisulfite treatment, whereby cytosine residues have been converted to uracil residues and subsequently replaced with thymine residues following amplification.

IX. Kits

[00175] Any of the compositions described herein may be comprised in a kit. In a non-limiting example, the kit, in a suitable container, comprises: an adaptor or several adaptors, one or more of oligonucleotide primers and reagents for ligation, primer extension and amplification. In some cases, the adaptors are diversity adaptors. The diversity adaptors can be any adapter has generated by the method depicted in **FIG. 2B**. The diversity adaptors can be any adaptor that comprises a partial duplex as depicted in **FIG. 3**. The adapters can be the adapters depicted in **FIG. 3**. The kit may also comprise means for purification, such as a bead suspension, and nucleic acid modifying enzymes.

[00176] The containers of the kits can 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 can contain a second, third or other additional container into which the additional components may be separately placed. Various combinations of components may be comprised in a container.

[00177] When the components of the kit are provided in one or more liquid solutions, the liquid solution can be an aqueous solution. 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.

[00178] The present methods provide kits containing one or more compositions described herein and other suitable reagents suitable for carrying out the methods described herein. The methods described herein provide, e.g., diagnostic kits for clinical or criminal laboratories, or nucleic acid

IDT Exhibit 1017.1628

amplification or analysis kits for general laboratory use. The present methods thus include kits which include some or all of the reagents to carry out the methods described herein, 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 cases, the kit comprises a binding molecule, wherein the binding molecule is a nucleotide analog binding protein. In some cases, the nucleotide analog binding protein comprises a methylcytosine binding protein. In some cases, the methylcyotsine binding protein comprises an anti-5-methylcytosine antibody. In some cases, the kit contains a modified nucleotide. Suitable modified nucleotides include any nucleotides provided herein including but not limited to a nucleotide analog. In some cases, the nucleotide analog can be a cytosine analog. In some cases, the cytosine analogs can be 5-methyl dCTP, 5-hydroxymethyl dCTP, and/or 5-propynl dCTP. In some cases, the kit comprises a converting agent. In some cases, the converting agent is bisulfite or its equivalent.

[00179] In some cases, the kit can 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 can 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 can include any of the compositions provided herein including but not limited to buffers, salts, divalent cations, azeotropes, chaotropes, dNTPs, labeled nucleotides, modified nucleotides, dyes, fluorophores, biotin, enzymes (such as endonucleases, exonucleases, glycosylases), or any combination thereof.

[00180] In some cases, the kit can contain one or more oligonucleotide primers, such as the oligonucleotide primers described herein. The oligonucleotide primers can be one or more primers comprising diversity bases or extender sequence as described herein. In some cases, the kit comprises a pool of primers, wherein some of the primers comprise no diversity bases, while the remaining primers in the pool of primers comprise a variable number of diversity bases. The primers can comprise 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 or more diversity bases. In some cases, the pool of primers is used to generate RRBS libraries such that the base composition of the diversity base is chosen so as not to recreate the YGG (Y is cytosine or thymine) motif at the 5' end of a nucleic insert in the RRBS library. The kit can contain one or more oligonucleotide primers comprising sequence directed against the ligation strand of an adapter or its complement and/or sequence directed against the ligation strand of an adapter or its complement whose sequence is altered by treatment with a converting agent. In some cases, the

IDT Exhibit 1017.1629

converting agent is bisulfite. In some cases the kit can 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 can contain chimeric primers comprising an RNA portion and a DNA portion. In some cases, the 5' portion of the tailed primers comprises one or more barcodes or other identifier sequences. In some cases, the identifier sequences comprise flow cell sequences, TruSeq primer sequence, and/or second read barcode sequences. In some cases, the identifier sequence is a stretch of random sequences such that the random sequences are read during an index or barcode read and are used as mark to facilitate identification and removal of PCR duplicates during subsequent *in silico* analysis of sequencing reads.

[00181] In some cases, the kit can contain one or more polymerases or mixtures thereof. In some cases, the one or more polymerases or mixtures thereof can comprise strand displacement activity and/or exonuclease activity. Suitable polymerases include any of the polymerases provided herein. The kit can further contain one or more polymerase substrates such as for example dNTPs, non-canonical or modified nucleotides, or nucleotide analogs.

[00182] In some cases, the kit can 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.

[00183] In some cases, the kit can contain one or more reagents for producing blunt ends. For example, the kit can 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 can 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 can comprise dNTPs.

[00184] In some cases, the kit can contain one or more reagents for preparing the double stranded products for ligation to adapter molecules. For example, the kit can contain dATP, dCTP, dGTP, dTTP, or any mixture thereof. In some cases, the kit can contain a polynucleotide kinase, such as for example T4 polynucleotide kinase. Additionally, the kit can contain a polymerase suitable for producing a 3' extension from the blunt ended double stranded DNA fragments. Suitable polymerases can be included, for example, exo-Klenow polymerase.

IDT Exhibit 1017.1630

[00185] In some cases, the kit can contain one or more adaptor molecules such as any of the adapter molecules provided herein. Suitable adapter molecules include single or double stranded nucleic acid (DNA or RNA) molecules or derivatives thereof, stem-loop nucleic acid molecules, forked 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 adapter molecules known in the art that can be covalently or non-covalently attached, such as for example by ligation, to the double stranded DNA fragments. In some cases, the adapters can be duplex adapters wherein one strand comprises nucleotide analogs resistant to conversion by a converting agent, while the other strand comprises a 5' and 3' block. In some cases, the adapters can be duplex adapters wherein both strands comprise nucleotide analogs resistant to conversion by a converting agent. In some cases, the adapters can be duplex adapters wherein both strands comprise nucleotides sensitive to conversion by a converting agent. In a further embodiment, the duplex adapter is a partial duplex adapter. In some cases, the partial duplex adapter comprises a long strand comprising nucleotide analogs resistant to conversion by a converting agent, and a short strand comprising a 5' and 3' block. In some cases, the partial duplex adapter comprises a long strand comprising nucleotides sensitive to conversion by a converting agent, and a short strand comprising a 5' and 3' block. In some cases, the nucleotide analog is a cytosine analog. In some cases, the nucleotides sensitive to a converting agent is dCTP. In some cases, the cytosine analogs present in the adapter can be 5-methylcytosine, 5-hydroxymethylcytosine, and/or 5-propynlcytosine. In some cases, the 5' block comprises a biotin moiety. In some cases, the 3' block is blocked with a terminal dideoxycytosine. In some cases, one or more adaptors in the kit comprise one or more diversity bases or extender sequence as described herein.

[00186] In some cases, the kit can contain one or more reagents for performing nick or fill-in repair on the ligation complex formed between the adaptors and the double stranded products of the methods described herein. The kit can contain a polymerase suitable for performing nick repair. Suitable polymerases can be included, for example, *Taq* DNA polymerase.

[00187] The kit can further contain instructions for the use of the kit. For example, the kit can contain instructions for generating RRBS libraries, directional cDNA libraries or directional cDNA libraries representing the methylome or the methylation status of a specific genomic region or locus useful for large scale analysis of 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

IDT Exhibit 1017.1631

methods described herein, or any combination thereof. The kit can further contain instructions for mixing the one or more reaction mixture components to generate one or more reaction mixtures suitable for the methods described herein. The kit can further contain instructions for hybridizing the one or more oligonucleotide primers to a nucleic acid template. The kit can further contain instructions for extending the one or more oligonucleotide primers with for example a polymerase and/or nucleotide analogs. The kit can further contain instructions for treating the DNA products with a converting agent. In some cases, the converting agent is bisulfite. The kit can further contain instructions for purification of any of the products provided by any of the steps of the methods provided herein. The kit can 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 can further contain instructions for phosphorylating the 5' ends of the double stranded DNA fragments produced by the methods described herein. The kit can further contain instructions for ligating one or more adapter molecules to the double stranded DNA fragments. The kit can further contain instructions for analysing sequence reads obtained from libraries generated using the methods provided herein and subsequently sequenced. The instructions can describe how to trim diversity bases present in sequence reads *in silico*. The instructions can describe how to remove duplicates present in sequence reads *in silico*.

[00188] A kit will can include instructions for employing, the kit components as well the use of any other reagent not included in the kit. Instructions can include variations that can be implemented.

[00189] 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.

EXAMPLES

Example 1: Generation of a RRBS library using Diversity Adapters:

[00190] In this example, an RRBS library was generated using the diversity adapters shown in **FIG. 2A-B** and the sequencing results were subsequently analyzed. As shown in **FIG.**

IDT Exhibit 1017.1632

2A-B, the diversity adapters can be partial duplex adapters comprising a long strand (e.g., R01446 in **FIG. 2A**), hybridized with a short strand (e.g., Ben383, 384, 385, or 381 in **FIG. 2A**) in such a manner that both ends of the partial duplex comprise a 5' overhang. The 5' overhang comprising sequence from the short strand of the partial duplex comprises a 5' terminal GC sequence with a Biosg label on the terminal cytosine residue, while the 3' terminus comprises a terminal 3' dideoxy cytosine residue. As can be seen in **FIG. 2A-B**, the 5' overhang comprising short strand sequence contains either no (D0, e.g., Ben341 in **FIG. 2A**), 1 (D1, e.g., Ben385 in **FIG. 2A**), 2 (D2, e.g., Ben384 in **FIG. 2A**), or 3 ((D3, e.g., Ben383 in **FIG. 2A**) "H" bases. The "H" bases can be either an adenosine, cytosine or thymine residue, but cannot be a guanine residue. As shown in **FIG. 2A-B**, for use in the methods provided herein as well as this Example, the partial duplex adapter is subjected to an extension reaction using a dNTP mix containing only dATP, dGTP, and dTTP (i.e., no dCTP), whereby a partial duplex adapter comprising a 5' overhang comprising at least one "H" (e.g., R01446 hybridized to Ben 383, 384, or 385) is extended to produce a partial duplex adapter containing a two base 5' overhang with the sequence GC at one end. In this Example, 5 ng of human gDNA was digested with MspI and the resulting fragments were separated into different diversity adaptor pools as shown in FIG. 2A-B, wherein a first pool was ligated to the D0 diversity adaptors, a second pool was ligated to the D1 diversity adapters, a third pool was ligated the D2 diversity adapters, a fourth pool was ligated to the D3 adapters, and a fifth pool was ligated to a mixture of the D0, D1, D2, and D3 diversity adapters. The ligations and all subsequent steps were performed according to the protocol for the NuGEN Ovation ® Ultralow Methyl-Seq Library System, which is herein incorporated by reference in its entirety. Briefly, following digestion of the gDNA with MspI to produce DNA fragments, the DNA fragments are combined with a master mix containing the appropriate diversity adapters comprising 0, 1, 2, or 3 diveristy "D" bases (i.e., D3 ,D2, D1, D0, or a mix of D0-D3 in **FIGs. 4A-B, 5A-B, 6A-B, 7A-B, and 9A-B, respectively**), ligation buffer, water and ligation enzyme mix, placed in a thermal cycler and run through a program on the thermal cycler that comprised of a single cycle comprising incubations at 25 C for 30 minutes, 70 C for 10 minutes, and then holding the mixture at 4 C. Following ligation, the ligation complexes were subjected to a post-ligation purification followed by a final repair reaction entailing incubation of the ligation complexes with a DNA repair enzyme in a thermal cycler programmed for a single cycle of 60 C incubation for 10 minutes. Following final repair, the samples were subjected to bisulfite conversion, wherein 20 uL of the products from the final repair reactions were inputted directly into a Qiagen EpiTect Fast DNA Bisulfite Kit according to manufacturer protocols and subjected to two cycles of 95 C for 5 minutes, and 60 C for 20

IDT Exhibit 1017.1633

minutes on a thermal cycler followed by elution of the purified, bisulfite converted DNA libraries.. The bisulfite converted DNA libraries were then PCR amplified and purified. Following library amplification and purification, the libraries were pooled and sequenced on an Illumina MiSeq sequencer according to manufacturer instructions with a 40 nucleotide forward read, 6 nucleotide index read, and 25 nucleotide reverse read. **FIGs. 4A-B, 5A-B, 6A-B, 7A-B, 8, and 9A-B** show a high correlation between the expected forward and reverse sequencing reads and the actual nucleotide distributions at read positions 1-40 in the forward reads (FIGs. 4A, 5A, 6A, 7A, and 9A) and 1-25 in the reverse reads (FIGs. 4B, 5B, 6B, 7B, and 9B)for nucleic acid fragments appended with diversity adapters with three diversity "D" bases (**FIG. 4A-B**), diversity adapters with 2 diversity "D" bases (**FIG. 5A-B**), diversity adapters with 1 diversity "D" base (**FIG. 6A-B**), diversity adapters with no diversity bases (**FIG. 7A-B**), and the mix of diversity adapters comprising zero, 1, 2, or 3 diversity "D" bases (**FIGs. 8 & 9A-B**). In some cases, the diversity base in the 5' position of a diversity adaptor comprising three diversity bases can be an "R" base while the remaining 2 diversity bases can be "D" bases as shown in **FIG. 35**. The "R" base can be either an adenine or a guanine nucleotide base, while the "D" bases can be either an adenine, guanine, or thymine nucleotide base as shown in **FIG. 35** and as described herein. In some cases, an "R" base is substituted for a "D" base in a pool of diversity adaptors such that library inserts generated using the methods described herein for generating an RRBS library following MspI digestion using the pool of diversity adaptors with an "R" base comprise an "R" base as the first base read in a forward sequence read (e.g., **FIG. 35**).

Example 2: Sequencing of RRBS library generated using Diversity Adapters without Phi X control:

[00191] In this example, RRBS libraries were generated using diversity adapters as described in Example 1; however, sequencing of the RRBS libraries generated using diversity adapters with a 40 nucleotide forward read, 6 nucleotide index read, and 25 nucleotide reverse read was conducted without running a dedicated control high diversity Phi X control lane. As can be seen in **FIGs. 10-12**, the RRBS library sequenced without a PhiX control showed a Q-score distribution whereby > 95% of the Q-scores were > 30 (**FIG. 10**), an anticipated distribution of bases (**FIGs. 10** and **11**) whereby a diversity of bases existed at the start of the sequencing read while roughly equal numbers of "A" and "G" bases were reported further into the read, and the percentage of > Q20 scores being above 95% at PCR amplification cycles of less than 40 (**FIG. 12**) , which indicated a low probability of an incorrect base calling. **FIG. 13** shows a Q-score heatmap over 71 sequencing cycles that include 40 forward read cycles, 6 index read cycles, and 25 reverse read sequencing cycles. **FIG. 16** shows a summary of the run for this

IDT Exhibit 1017.1634

Example showing a high percentage (>95%) of total reads with a Q score above 30, with a forward read (read 1) having a Q 30 score of 95%, an index read (read 2) having a Q 30 score of >90%, anda reverse read (read 3) having a Q30 score of >90%. In summary, the use of the diversity adapters in generating RRBS libraries with diversity adaptors comprising D bases as described in Example 1 produced high quality sequencing reads with or without the use of PhiX control libraries.

[00192] As a comparison, inserts from a Human Blood RNA sequencing library generated for ~ 100 ng of total RNA from adult human whole blood according to the manufacturer's instructions for the NuGEN Ovation® Human Blood RNA-Seq Library System, which is herein incorporated by reference in its entirety, were sequenced with a 40 nucleotide forward read, 6 nucleotide index read, and 25 nucleotide reverse read. Briefly, ~100 ng of total RNA from adult human whole blood was subjected to first strand synthesis using random primers, followed by second strand synthesis in the presence of a non-canonical nucleotide (i.e., dUTP), fragmented using the Covaris sonication, end repaired, ligated with adapters marked with a non-canonical nucleotide (i.e., dUTP) and comprising barcode sequences, strand selected, adapter cleavage to remove unwanted sequences (e.g., rRNA and Globin RNA), and finally, library amplification using PCR with primers directed against the adapter sequences, thereby allowing for 16-plex sequencing capability. **FIGs. 14** and **15** show intensity by cycle, cluster density, Q-score distribution and heat-map data from the 40 cycle forward read by 6 cycle index read by 25 cycle reverse read run performed on the Blood-Seq library that was similar to the data obtained from the 40 cycle forward read by 6 cycle index read by 25 cycle reverse read run performed on the RRBS library generated without Phi X controls (see **FIGs. 10 and 13**). **FIG. 17** shows the run summary for the 40 cycle forward read by 6 cycle index read by 25 cycle reverse read run performed on the Blood-Seq library. A comparison between **FIGs 16** and **17** showed that sequencing of the RRBS library generated with diversity adapters as described in Example 1 produced high quality sequencing data without requiring a Phi X high diversity control.

Example 3: Comparison of the CpGs analyzed by sequencing a RRBS library generated using Diversity Adapters vs. Whole Genome Bisulfite Sequencing (WGBS):

[00193] In this example, a comparison of the CpGs analyzed by sequencing a RRBS library generated by the methods provided herein vs. WGBS was conducted. The RRBS library was generated from MspI fragmented genomic DNA (gDNA) from IMR90 human cell line was ligated to a mixture of DO-D3 diversity adapters as described in Example 1 using either 400, 100, 25, 6.25, or 1.56 ng of gDNA as input for generation of the RRBS library. The library used for WGBS was generated by first subjecting 50 ng of human gDNA to Covaris sonication to an

IDT Exhibit 1017.1635

average size of 200 bp, and then generating a library for WGBS using the fragmented gDNA as input into the NuGEN Ovation ® Ultralow Methyl-Seq Library System, wherein the library for WGBS was generated as directed by the manufacturer. The RRBS and WGBS libraries were each separately sequenced on an Illumina GAIIx sequencer with a 40 nucleotide single end (40nt SE) and 6 nucleotide index read. For the RRBS sequences, the D1-D3 adaptor sequences were trimmed off, and the resulting reads, which started with "YGG", were analyzed with the Bismark program for mapping bisulfite treated sequencing reads. Mapped reads were further analyzed by grouping into categories of CpG islands, CpG shores, and promoters. **FIG. 20** shows that the total CpGs analyzed using as little as 1.56 ng of starting gDNA was considerably greater than the total CpGs analyzed by WGBS. Moreover, the average CpG per read and average CpG per uniquely aligned read were at least 6 X and 7X greater than WGBS, respectively (**FIG. 21**). **FIG. 22** shows that the feature coverage (i.e., CpG Islands and Shores, as well as Promoters) using RRBS libraries generated using the methods provided herein are very close to the theoretical maximum predicted by an *in silico* analysis for a 40 nt SE read for a RRBS library (**FIGs. 22 and 25**). Overall, RRBS libraries generated using diversity adapters as provided herein can be easily sequenced (e.g., with commercially available NGS sequencers such Illumina systems), and provide CpG-dense data.

Example 4: Kit Components

[00194] A kit can be used for RRBS library formation.

[00195] A. Reagents

[00196] The kit can comprise the components in Table 1:

Table 1. Reagents

MspI Buffer Mix
MspI Enzyme Mix
Ligation Buffer Mix
Ligation Adaptor Mixes
Ligation Enzyme Mix
Final Repair Buffer Mix
Final Repair Enzyme Mix
Amplification Primer Mix
Amplification Enzyme Mix
25 μM Sequencing Primer

IDT Exhibit 1017.1636

Nuclease-free Water
Agencourt RNAClean XP Beads
DNA Resuspension Buffer

[00197] B. Additional Equipment, Reagents and Labware

[00198] Additional materials that can be used include the following Equipment, Reagents, and Labware.

[00199] Equipment includes: Agilent 2100 Bioanalyzer, High Sensitivity DNA Kit or materials andequipment for electrophoretic analysis of nucleic acids; Qubit® 2.0 Fluorometer and dsDNA HS Assay Kit (Life Technologies); microcentrifuge for individual 1.5 mL and 0.5 mL tubes; 0.5–10 µL pipette, 2–20 µL pipette, 20–200 µL pipette, 200–1000 µL pipette; vortexer; thermal cycler with 0.2 mL tube heat block, heated lid, and 100 µL reaction capacity.

[00200] Reagents include: EpiTect Fast DNA Bisulfite Kit (QIAGEN Cat. #59824 for 50 preps or Cat. #59826 for 200 preps); Ethanol (Sigma-Aldrich, Cat. #E7023), for purification steps; OPTIONAL: EvaGreen® Dye, 20X in water (Biotium, Cat. #31000).

[00201] Supplies and Labware include: Nuclease-free pipette tips; 1.5 mL and 0.5 mL RNase-free microcentrifuge tubes; 0.2 mL individual thin-wall PCR tubes or 8 X 0.2 mL strip PCR tubes or 0.2 mL thin-wall PCR plates; magnetic separation device for the Bead Purification of the Amplified Material protocol; disposable gloves; kimwipes; ice bucket; cleaning solutions such as DNA-OFF™ (MP Biomedicals, Cat. #QD0500); OPTIONAL: PhiX Control (Illumina, Cat. Cat. #FC-110-3001); OPTIONAL: Real-time PCR system.

Example 5: Planning an experiment

[00202] A. Input DNA

[00203] Methods, compositions, and kits provided herein can work with inputs of 100 ng of intact human genomic DNA. Quantitation of gDNA by a dsDNA assay, such as PicoGreen® or the Qubit System, can be performed. In some cases, OD260 readings for quantitation of input material are not relied upon. DNA samples can be free of contaminating proteins, RNA, organic solvents (including phenol and ethanol) and salts. A commercially available system for gDNA isolation can be used. The A260:A280 ratio for DNA samples can be in excess of 1.8. Using DNA samples with lower ratios can compromise results. RRBS libraries can be generated from less than 100 ng of gDNA, or from degraded gDNA, such as DNA extracted from formalin fixed, paraffin embedded (FFPE) specimens. If such samples are used, library amplification can

IDT Exhibit 1017.1637

require additional cycles of PCR. One additional cycle of PCR can be used for each two fold decrease in starting material. To accurately determine the number of PCR cycles to perform, the optional real-time PCR protocol in Example 7. An increase in adaptor artifacts can be observed, as well as decreased library complexity, in libraries made from less than 100 ng or degraded samples, and additional cycles of PCR amplification can be performed. Adaptor artifacts can appear as a peak of approximately 145 bp. These artifacts can be reduced by performing a second bead purification. One volume of beads can be added to the purified library.

[00204] Signs of reduced complexity can be monitored by sequencing unique 6 random bases appended to the index on each adaptor molecule, then a Duplicate Marking tool can be used to identify true PCR duplicates. The Duplicate Marking tool utilizes information provided by the unique 6 random bases sequence to discriminate between true PCR duplicates and independent adaptor ligation events to fragments with the same start site resulting in the recovery of more usable data. The Duplicate Marking tool and instructions for use can be obtained by contacting NuGEN Technical Support (techserv@nugen.com). After removing true PCR duplicates, a reliable measure of library complexity can be made. Sequencing the additional 6 bases after the 6-base barcode can be useful when attempting RRBS on rare or degraded samples. See Example 7 for information on index structure and sequencing.

[00205] B. Using compositions and kits described herein on Illumina NGS Platforms

[00206] Methods described herein can use the same approach to multiplexing found in the standard Illumina method. These libraries can be sequenced using the Illumina protocol for multiplex sequencing. The dedicated read (DR) barcode sequences, found in Table 2, can be entered into the Illumina software prior to analysis.

[00207] Barcode sequences and multiplex guidelines for adaptors used herein can be found in Table 2. These 6-nucleotide barcode sequences can be input into the Illumina Sequencing System prior to parsing of the data. You may combine anywhere from 2–16 barcoded libraries to allow for a range of multiplex sequencing. The barcodes can be chosen for their ability to parse properly and for color balancing and can have strict pairing requirements when performing 2-plex multiplexing. Users wishing to perform greater than a 2-plex multiplexing can choose a Duplex Set (as defined in Table 9), combined with any of the remaining barcoded libraries. The barcode sequences can be separated by an edit distance of three. For further details on the barcode design strategy, please refer to Faircloth BC and Glenn TC (2012). *PLoS ONE* 7(8): e42543. doi:10.1371/journal.pone.0042543, which is herein incorporated by reference in its entirety. NuGEN RRBS libraries can be mixed with other libraries if the barcodes are compatible (i.e. can be parsed). One mismatch allowed can be parsed with if the spike in library barcodes are

IDT Exhibit 1017.1638

an edit distance of 3 or greater from the RRBS barcodes used in that lane. The PhiX library from Illumina does not contain an index. As a result, the sequencer canproduce a low-quality index read from PhiX clusters. To remove PhiX reads prior to parsing, filter by index read quality and remove reads with quality less than 20.

[00208] Barcode sequences for dedicated read (DR) adaptors used in herein can be found in Table 2.

Table 2 Barcode sequences for dedicated read (DR) adaptors

6 NT BARCODE SEQUENCE AS READ BY THE SEQUENCER	BARCODE PAIRING (2-PLEX)
AACCAG	Duplex Set 1
TGGTGA	
AGTGAG	Duplex Set 2
GCACTA	
ACCTCA	Duplex Set 3
GTGCTT	
AAGCCT	Duplex Set 4
GTCGTA	
AAGAGG	Duplex Set 5
GGAGAA	
AGCATG	Duplex Set 6
GAGTCA	
CGTAGA	Duplex Set 7
TCAGAG	
CACAGT	Duplex Set 8
TTGGCA	

One of the duplex sets can be used in combination with any of the other remaining 14 individual barcodes.

[00209] The barcode sequences used can be chosen for their ability to parse properly and for color balancing.

[00210] Libraries generated using methods, compositions, and kits described herein can be sequenced using a custom Read 1 sequencing primer, e.g., MetSeq Primer 1, which can be included in a kit at a concentration of 25 μM. The Standard Read 1 Primer can be used when

IDT Exhibit 1017.1639

multiplexing with PhiX or other libraries. The standard primers provided in the Illumina sequencing kit can be sufficient for sequencing the barcodes and, if desired, the reverse read.

[00211] Methods, compositions, and kits provided herein can be used to produce directional bisulfite-converted libraries. The Read 1 sequencing primer can sequence the C-to-T converted strand and the overall nucleotide balance for the Read 1 sequencing primer can show a low proportion of C bases. Illumina can recommend how to obtain high-quality base calls from libraries containing unbalanced nucleotide ratios. These recommendations can differ by instrument. Illumina software can produce higher-quality reads from low-diversity samples such as RRBS. High quality results can be made if the sequencing instrument is running the following versions or later:

HiSeq- HCS v2.2.38 (includes RTA v1.18.61)

MiSeq- RTA v1.17.28

[00212] These software upgrades, in combination with the sequence diversity incorporated in the adaptors described herein, can produce high-quality RRBS reads without the need to spike in balanced library such as PhiX. However, spiking in 5% PhiX or another, previously characterized high-quality library can be useful for troubleshooting purposes in the event of a failed run, but can reduce the data output by 5%. Such a control can be used to tell if the sequencing run failed in general, or if there is a problem specific to the particular RRBS library being sequenced.

[00213] The first base of the read RRBS method contains a CpG methylation measurement. In principle, reads need only be long enough to accurately map them to the genome. However, mapping rates can be directly affected by sequence read length, and use of longer reads can potentially lead to more uniquely mapping reads and coverage of a greater number of CpG loci. Table 3 lists the percent of MspI fragments 40–300 nt in length that are uniquely or non-uniquely mappable to the hg19 reference genome. In this theoretical analysis, the 1,545,560 MspI fragments of hg19 that are 40–300 nt in length are bisulfite-converted in silico (assuming 0% methylation and 100% bisulfite conversion), trimmed to the indicated read length, then mapped using Bismark (www.bioinformatics.bbsrc.ac.uk/projects/bismark/).

[00214] Table 3. Percent of MspI fragments 40–300 nt in length that are uniquely or nonuniquely mappable to the hg19 reference genome

SINGLE END READ LENGTH	40 nt	50 nt	100 nt
Sequences mapping uniquely	65.2%	69.6%	83.9%
Sequences mapping non-uniquely	34.8%	30.4%	16.1%

IDT Exhibit 1017.1640

[00215] In addition to mappability, how read length affects CpG loci coverage can be considered. Many MspI fragments contain internal CpG's, so longer reads can sequence more CpGs. However, many MspI fragments can be smaller than 100 bp, and even smaller than 50 bp. For these fragments, long sequencing reads, or paired end reads, can provide no additional CpG data. Table 4 lists the number of CpG loci that can be expected to be covered by various sequencing protocols.

[00216] Table 4. Number of CpG loci expected to be covered by various sequencing protocols

SINGLE END READ LENGTH	40 nt	50 nt	100 nt
CpG loci covered	2,649,855	3,281,511	5,916,268

[00217] C. Amplified Library Storage

[00218] Amplified libraries can be stored at –20°C.

[00219] D. Data Analysis and Parsing Multiplex Libraries

[00220] The sequences of the barcodes can be entered prior to parsing (the sequences are listed in Table 3). With bisulfite-converted libraries, a slightly higher rate of unmatched barcodes relative to non- bisulfite-converted libraries can be observed. The methods can use edit distance 3 barcodes, which can allow one mismatch during parsing. This can significantly reduce the fraction of unmatched barcode reads.

[00221] The Duplicate Marking feature mentioned above can built into the NuGEN RRBS system, by adding an additional 6 nt to the index read. See Example 7 for more information on this feature.

[00222] Once the data have been parsed according to sample index, the reads can be trimmed before attempting alignment. Trimming can be done in two steps. First, any low-quality bases and adaptor sequences can be removed, then the sequence diversity provided by adaptors described herein can be removed. Example 8 contains detailed recommendations on trimming. At this point reads can be ready for downstream analysis, such as mapping to the genome and determining methylation status.

Example 6: Protocol

[00223] A. Overview

IDT Exhibit 1017.1641

[00224] As shown in **FIG. 38**, an RRBS library preparation process can be performed in five stages: 1. DNA digestion with MspI: 1.0 hour; 2. Adaptor ligation: 0.75 hours; 3. final repair: 0.25 hours; 4. bisulfite conversion: 3.0 hours; 5. amplification and purification: 1.5 hours. Total time to prepare amplified library: 6.5 hours.

[00225] Components in a kit can be color coded, with each color linked to a specific stage of the process. Performing each stage can involve making a master mix, then adding it to the reaction, followed by incubation. Master mixes can be prepared by mixing components provided for that stage.

[00226] Each library can be produced independently when performing multiplex sequencing. In some cases, adaptors are not mixed during the actual library construction protocol. Multiplexing can be achieved by mixing the amplified libraries prior to cluster formation. The barcode sequences in can be carefully chosen for their ability to parse properly and for color balancing.

[00227] B. Protocol Notes

[00228] Water provided with a kit or an alternate source of nuclease-free water can be used. In some cases, DEPC-treated water is used; in some cases, DEPC-treated water is not used. A minimum of four reactions can be set up at a time to avoid pipetting very small volumes. Components used in each step can be thawed and immediately placed on ice. In some cases, all reagents at not thawed once. Thawed reagents and reaction tubes can be kept on ice. After thawing and mixing buffer mixes, if any precipitate is observed, it can be re-dissolved completely prior to use. The buffer mix can be gently warmed for 2 minutes at room temperature followed by brief vortexing. In some cases, any enzyme, primer, or adaptor mixes is not warmed.

[00229] Small amounts of reagents can be placed into a reaction mix, pipetted up and down several times to ensure complete transfer from the pipet tip into the reaction mix. Pipetting a mix can involve gently aspirating and dispensing a volume that is at least half of the total volume of the reaction mix.

[00230] The thermal cycler can be allowed to reach the initial incubation temperature prior to placing the tubes or plates in the block. Fresh ethanol stocks can be used make 70% ethanol for the purification protocols. The ethanol mixes can be made fresh, carefully measuring both the ethanol and water with pipettes. Lower concentrations of ethanol in wash solutions can result in loss of yield as the higher aqueous content will dissolve the DNA and wash it off the beads or column.

[00231] C. Agencourt® RNAClean® XP Purification Beads

IDT Exhibit 1017.1642

[00232] The bead purification processes can include 1. binding of DNA to RNAClean XP beads; 2. magnetic separation of beads from supernatant; 3. ethanol wash of bound beads to remove salts, etc.; 4. elution of bound DNA from beads. Beads can be removed from 4°C and left at room temperature for at least 15 minutes before use. Beads can completely reach room temperature before using. Cold beads can reduce recovery. Beads can be fully resuspended by vortexing before adding to sample. The ratio of Agencourt RNAClean XP bead volume to sample volume can vary among different steps in the protocol. Furthermore, the bead:sample ratios used can differ from the standard Agencourt protocol. The beads can separate on the magnet for a full 5 minutes. Removing binding buffer before the beads have completely separated can impact DNA yields. After completing the binding step, bead loss can be minimized when removing the binding buffer. With the samples placed on the magnet, the amount of the binding buffer specified at each of the individual purification steps can be removed. Some liquid can remain at the bottom of the tube, helping to minimize bead loss at this step. Bead loss can be minimized throughout the procedure to maximize DNA yield. An ethanol wash can be freshly prepared from fresh ethanol stocks at the indicated concentration. Lower percent ethanol mixes can reduce recovery. During the ethanol washes, the samples can be kept on the magnet. Dispersion of the beads can be avoided. The magnet can keep the beads on the walls of sample wells or tubes. All residual ethanol can be removed prior to continuing with the next step. When removing the final ethanol wash, most of the ethanol can be removed first, then excess ethanol is collected at the bottom of the tube before removing the remaining ethanol. This process can reduce the required bead air drying time. After drying the beads for the time specified in the protocol, each tube can be carefully inspected to determine whether all the ethanol has evaporated before proceeding.

[00233] Strip tubes or partial plates can be firmly placed within the magnetic plate. Individual tubes can be difficult to position stably on the magnetic plates.

[00234] D. Programming the Thermal Cycler

[00235] A thermal cycler with a heat block designed for 0.2 mL tubes can be used, equipped with a heated lid, and with a capacity of 100 µL reaction volume. The programs can be prepared as shown in Table 5, following the operating instructions provided by the thermal cycler manufacturer. For thermal cyclers with an adjustable heated lid, the lid temperature can be set to 100°C when sample temperature reaches above 30°C. For thermal cyclers with a fixed temperature heated lid (e.g., ABI GeneAmp® PCR 9600 and 9700 models), default settings can be used (e.g., 100°C to 105°C).

IDT Exhibit 1017.1643

[00236] Table 5. Thermal cycler programming

Program 1 (MspI Digestion)	37°C – 60 min, hold at 4°C
Program 2 (Ligation)	25°C – 30 min, 70°C – 10 min, hold at 4°C
Program 3 (Final Repair)	60°C – 10 min, 70°C – 10 min, hold at 4°C
Program 4 (Bisulfite Conversion)	2 cycles of (95°C – 5 min, 60°C – 20 min) hold at 20°C
Program 5 (Library Amplification)	95°C – 2 min, 12 cycles of (95°C – 15 sec, 60°C – 1 min, 72°C – 30 sec), hold at 10°C

[00237] If less than 100 ng of intact gDNA is used, or if DNA is severely degraded (such as with FFPE), more than 12 cycles of PCR may be performed. As an estimate, one additional cycle of PCR for each two-fold reduction of input can be performed. For example, for 25 ng of intact gDNA, 14 cycles of PCR can be used.

[00238] E. MspI Digestion

[00239] The MspI Buffer Mix, MspI Enzyme Mix, and Nuclease-free Water can be removed from –20°C storage. Contents of MspI Enzyme Mix can be spun down and placed on ice. MspI buffer mix and nuclease-free water can be thawed at room temperature. MspI buffer mix can be mixed by vortexing, spun and placed on ice. Each 100 ng gDNA sample can be adjusted to 8.5 µL with nuclease-free water. A master mix can be prepared by combining MspI buffer mix and MspI enzyme mix in a 0.5 mL capped tube, according to the volumes shown in Table 6.

[00240] Table 6. MspI Master Mix (volumes listed are for a single reaction)

MspI Buffer Mix	MspI Enzyme Mix
1.0 µL	0.5 µL

[00241] 1.5 µL of the MspI master mix can be added to each sample tube. Mixing can be by pipetting; tubes can be capped, spun and placed on ice. Tubes can be placed in a thermal cycler programmed to run Program 1 (MspI Digestion; see Table 5): 37°C – 60 min, hold at 4°C. Tubes can be removed from the thermal cycler, spin to collect condensation and place on ice.

[00242] F. Ligation

[00243] A Ligation Buffer Mix, Ligation Adaptor Mixes, and Ligation Enzyme Mix can be removed from –20°C storage. Ligation Enzyme mix can be spun down and placed on ice.

IDT Exhibit 1017.1644

Ligation Buffer Mix (L1) and Ligation Adaptor Mixes (L2) can be thawed at room temperature. The solution can be mixed by vortexing, spun and placed on ice. The Ligation Buffer Mix (L1) can be extremely viscous. Care can be taken to ensure it is well mixed after thawing (it can be helpful to alternately vortex the tube right-side up and upside down). In some cases, Ligation Adaptor mixes are not warmed above room temperature. Heating can degrade performance.

[00244] Three µL of appropriate L2 Ligation Adaptor Mix can be added to each sample and mixed thoroughly by pipetting. Prior to use, a master mix can be made by combining nuclease-free water, L1, Ligation Enzyme Mix in a 0.5 mL capped tube according to the volumes in Table 7. The solutions can be mixed by pipetting slowly, without introducing bubbles, spun and placed on ice. The mix can be used immediately.

[00245] The L1 Ligation Buffer Mix can be very viscous. This reagent can be pipetted slowly, and care can be taken to ensure that the Ligation Master Mix and the ligation reactions are well mixed (visually observe that the solutions become homogeneous).

[00246] Table 7. Ligation Master Mix (volumes listed are for a single reaction)

WATER	2.0 µL
LIGATION BUFFER MIX	4.0 µL
LIGATION ENZYME MIX	1.0 µL

[00247] 7 µL Ligation Master Mix can be added to each reaction tube. The solution can be mixed thoroughly by pipetting slowly, gently spun and placed on ice. Incubation can be preceded immediately. Tubes can be placed in a pre-warmed thermal cycler programmed to run Program 2 (Ligation; see Table 5): 25°C – 30 min, 70°C – 10 min, hold at 4°C. The tubes can be removed from the thermal cycler, spun to collect condensation, and placed on ice.

[00248] **G. Final Repair**

[00249] The Final Repair Buffer Mix (FR1) and Final Repair Enzyme Mix (FR2) can be removed from –20°C storage. Contents of FR2 can be spun down and placed on ice. FR1 can be thawed at room temperature. The solutions can be mixed by vortexing, spun and placedon ice. A master mix can be prepared by combining FR1 and FR2 in a 0.5 mL capped tube, according to the volumes shown in Table 8.

[00250] Table 8. Final Repair Master Mix (volumes listed are for a single reaction)

FINAL REPAIR BUFFER	FINAL REPAIR ENZYME MIX	WATER

IDT Exhibit 1017.1645

MIX		
6.0 µL	0.5 µL	13.5µL

[00251] The solutions can be mixed by pipetting slowly, without introducing bubbles, spun for 2 seconds and place on ice. The mix can be used immediately.20 µL of the Final Repair Master Mix can be added to each of the 20 µL ligation reactions. The solution can be mixed by pipetting. The tubes can be capped, and spun, and placed on ice. The tubes can then be placed in a thermal cycler pre-heated to 60°C and programmed to run Program 3 (Final Repair; see Table 5): 60°C – 10 min, 70°C – 10 min, hold at 4°C. The tubes can be removed from the thermal cycler, spun to collect condensation and placed on ice.

[00252] H. Bisulfite Conversion

[00253] 1. QIAGEN EpiTect Fast DNA Bisulfite Kit (QIAGEN Cat. #59824 for 50 preps or Cat. #59826 for 200 preps) can be used for bisulfite conversion. Other commercial bisulfite conversion kits can also be used. The 40 µL product of the Final Repair reaction can be input directly into the bisulfite conversion kit. The input can be performed using the QIAGEN EpiTect Fast protocol "Bisulfite Conversion of Unmethylated Cytosines in DNA", on page 19, following the guidelines for 1–500 ng in a maximum volume of 40 µL (low concentration). 40 µL of input and 15 µL of DNA Protect Buffer can be used. The following thermal cycler conditions can be used: 95°C – 5 min, 60°C – 20 min, 95°C – 5 min, 60°C – 20 min, hold at 20°C. After thermal cycling, a blue ring at the meniscus can be observed. This can be normal and may not affect the results. The rest of the QIAGEN protocol can then be followed. Carrier RNA in the BL Buffer can be used. The purified, bisulfite-converted DNA can be eluted in 23 µL of EB. This can yield 20 µL of bisulfite-converted DNA ready for amplification. If necessary, the final eluted volume of bisulfite-converted DNA can be adjusted to 20 µL with nuclease-free water.

[00254] I. Library Amplification

[00255] The real-time protocol in Example 7 can be consulted to determine the number of cycles of PCR to perform (such as when starting with degraded DNA or significantly less than 100 ng of high-quality DNA). The Amplification Primer Mix (P2) and Amplification Enzyme Mix (P3)m can be removed from –20°C storage. P3 can be spun and placed on ice. P2 can be thawed at room temperature. The solutions can be mixed by vortexing, spun and placed on ice. The 20 µL product from the Bisulfite Conversion protocol can be obtained. A master mix can be made by combining P2 and P3 in an appropriately sized capped tube according to the volumes shown in Table 9. The solutions can be mixed by pipetting (care can be taken to avoid bubbles), spun and placed on ice.

IDT Exhibit 1017.1646

[00256] Table 9. Amplification Master Mix (volumes listed are for a single reaction)

AMP PRIMER MIX (P2)	AMP ENZYME MIX (P3)
5 µL	25 µL

[00257] On ice, 30 µL of the Amplification Master Mix can be added to each sample. Tubes can be placed in a pre-warmed thermal cycler programmed to run Program 5 (Library Amplification; see Table 5): 95°C – 2 min, 12 cycles of (95°C – 15 sec, 60°C – 1 min, 72°C – 30 sec), hold at 10°C. If less than 100 ng of intact gDNA is used, or if the DNA is severely degraded (such as with FFPE), more than 12 cycles of PCR can be performed. Methods provided in Example 8 can be used in such situations. Tubes can be removed from the thermal cycler, spun to collect condensation and placed on ice.

[00258] J. Amplified Library Purification

[00259] The Agencourt RNAClean XP Beads and 70% ethanol set aside previously can be used if they are still at room temperature. The beads can be resuspended by vortexing the tube and added to the sample. In some cases, the beads are not spun after resuspending. At room temperature, 50 µL (1 volume) of the bead suspension can be added to each reaction. The bead suspension can be quite viscous; therefore slow pipetting can be performed to enhance accuracy. The solutions can be mixed thoroughly by pipetting 10 times. A multichannel pipettor can be used to make the incubation times uniform. The mixture can be incubated at room temperature for 10 minutes. The tubes can be transferred to the magnet and let stand 5 minutes to clear the solution of beads. 90 µL of the binding buffer can be carefully removed and discarded. Some of the volume can be left behind to reduce bead loss at this step. In some cases, the beads are not disperse; instead, they can stay on the walls of the tubes. Significant loss of beads at this stage will impact the amount of purified DNA, so in some cases, beads are not removed with the binding buffer or the wash. The 70% ethanol wash in the library purification step can involve removing samples from the magnet, fully resuspending the beads in the 70% ethanol wash, and re-magnetizing until the beads have cleared from the solution. The beads can be washed in following steps: the plate can be removed from the magnet, 150 µL of freshly prepared 70% ethanol can be added and the mixture can be pipetted to resuspend the beads; the beads can be cleared by re-magnetizing and a 70% ethanol wash can be performed using a pipette. The washing steps can be repeated twice. In some cases, the ethanol can be removed as much as possible with the final wash. At least 2 pipetting steps can be performed to allow excess ethanol to collect at the bottom of the tubes after removing most of the ethanol in the first pipetting step.

IDT Exhibit 1017.1647

The beads can be air dried on the magnet for a minimum of 10 minutes. In some cases each tube can be carefully to ensure that all the ethanol has evaporated. All residual ethanol can be removed prior to continuing. The tubes can then be removed from the magnet. 20 µL DNA Resuspension Buffer can be added to the dried beads. The beads can be resuspended by thorough mixing. The tubes can be transferred to the magnet and let stand for 5 minutes. 18 µL of the eluate can be carefully removed, ensuring as few beads as possible are carried over, and transferred to a fresh set of tubes. When pipetting any portion of this eluted library downstream, the library can be let stand briefly on a magnet to reduce bead carryover. The eluate can then be used to quantitatively and qualitatively assess the library. After quantitation, barcoded libraries that will be run in the same flow cell can be mixed in equimolar ratios prior to processing on the flow cell, cBot or Cluster Station.

[00260] K. Quantitative and Qualitative Assessment of the Library

[00261] Samples can be diluted 1:3 with water. 1 µL of the diluted sample can be loaded on the Bioanalyzer High Sensitivity DNA Chip. In some cases, fragment distribution can be as shown in **FIG. 28**. The three peaks at 200 bp, 265 bp, and 330 bp are due to MspI-containing micro-satellite repeats, and are characteristic of RRBS libraries made from human DNA. For qPCR for quantification, 250 bp can be used as library size for calculations.

Example 7 - PCR amplification when using degraded DNA or less than 100 ng of high-quality DNA

[00262] Standard RRBS libraries can be generated using 100 ng of high-quality, MspI-digested genomic DNA. Library Amplification can be performed by preparing an Amplification Master Mix and adding 30 µL of this to 20 µL of bisulfite-converted sample for a total PCR volume of 50 µL. In some cases, the following two options can be used for degraded DNA or less than 100 ng of high-quality DNA: A) If high-quality DNA is used and the amount of DNA is known, the number of PCR cycles can be adjusted to obtain sufficient material for analysis and sequencing. One additional cycle of PCR can be performed for each two-fold reduction of input. For example, for 25 ng of intact gDNA, 14 cycles of PCR can be used. B) If degraded DNA is used or if the amount and/or quality of the DNA is unknown, a 1/5 scale real-time PCR with an aliquot of your sample can be performed as follows: A master mix can be made by combining P2 and P3 in an appropriately sized capped tube according to the volumes shown in Table 10. P3 Enzyme Mix can be added at the last moment and mixed well by pipetting taking care to avoid bubbles, spun and placed on ice. EvaGreen (Biotium, Cat. #31000) is less inhibitory to PCR than SYBR Green®, however SYBR Green can also be used at a 1X final concentration in the real-

IDT Exhibit 1017.1648

time PCR reaction. For each library, 6 µL of the Amplification Master Mix can be aliquoted into a well of on a real-time PCR plate. 4 µL of sample can be added to each well (after bisulfite conversion but before PCR amplification) for a total real-time PCR volume of 10 µL. Real-time PCR can be performed with the following cycling conditions: 95°C – 2 min, 30 cycles of (95°C – 15 sec, 60°C – 1 min, 72°C – 30 sec). The SYBR Green channel can be monitored. In some cases, ROX normalization is not used. After thermal cycling is complete, the amplification plot can be examined for each sample to determine the number of PCR cycles to perform during the Library Amplification step. A cycle in the late exponential phase can be selected. For example, in the amplification plot shown in **FIG. 29**, cycle 9 was chosen. This provides sufficient amplification without entering the plateau phase.

[00263] Table 10. Amplification Master Mix (volumes listed are for a single reaction)

AMP PRIMER MIX (P2)	20X EVAGREEN	AMP ENZYME MIX (P3)
1.0 µL	0.5 µL	4.5 µL

[00264] If using an Applied Biosystems real-time PCR instrument, Plot > Rn vs. Cycle, not Plot > deltaRn vs. Cycle is selected. Selecting Plot > deltaRn vs. Cycle may give unexpected results similar to those shown in **FIG. 30**.

[00265] After determining the optimal number of PCR cycles to perform, the remaining 16 µL of bisulfite-converted library can be amplified as follows: A master mix can be made by combining P2 and P3 in an appropriately sized capped tube according to the volumes shown in Table 11. P3 Enzyme Mix can be added at the last moment and mix well by pipetting taking care to avoid bubbles, spun and placed on ice. For each library, 24 µL of the Amplification Master Mix can be aliquoted into a well of on a real-time PCR plate. 16 µL of sample can be added to each well. PCR can be performed with the following cycling conditions, where N = number of cycles determined from the above real-time PCR assay: 95°C – 2 min, N cycles of (95°C – 15 sec, 60°C – 1 min, 72°C – 30 sec, hold at 10°C. Proceed with the Amplified Library Purification protocol in Example 6, Section J, by adding 1 volume (40 µL) of Agencourt RNAClean XP Beads to each amplified sample and following the remaining steps of the purification protocol.

[00266] Table 11. Amplification Master Mix (volumes listed are for a single reaction)

AMP PRIMER MIX (P2)	AMP ENZYME MIX (P3)
4.0 µL	20 µL

IDT Exhibit 1017.1649

Example 8 – Index structure and index read recommendations

[00267] The 6-base barcodes can be used for sample multiplexing. In addition to the 6-base barcode, the adaptor can contain 6 random bases immediately following the 6-base barcode, for a total of 12 bases. The additional 6 bases can be used for duplicate read determination using the Duplicate Marking Tool. The libraries can be sequenced using 12 cycles for the index read. The 6-base barcode sequences can be found in Table 2. In some cases, the MiSeq and HiSeq instruments do not provide a simple way to obtain the sequence information contained in the 12-base pair index read including the 6 random bases that can be necessary for duplicate read determination. Several methods can be used to generate the necessary index fastq file are provided below.

[00268] MiSeq Instrument

[00269] Parsing multiplex runs using the MiSeq built-in Illumina software can replace the barcode sequence from each library with a numerical substitute, which removes the duplicate information provided by the N6 sequence present after the barcode. To retrieve this information using the MiSeq instrument, one of the options can be used: Option 1: contact Illumina Technical Support and request a modification of the MiSeq config file to allow generation of an index fastq file during data analysis. Option 2: Modify the MiSeq config file to allow generation of an index fastq file during data analysis: 1. Stop the MiSeq Reporter process. 2. Locate the "MiSeq Reporter.exe.config" file located in C:/Illumina/MiSeq Reporter. 3. Open config file and search for a line that reads:

<add key="CreateFastqForIndexReads" value="0"/>

1. If this line is present, change the value from "0" to "1".
2. If this line is not present, add the line to the config file using the add keys function under the app Settings tab with the value set as "1".

[00270] 4. Restart the MiSeq reporter process. 5. Re-queue the run for data analysis if required. The 6-base barcodes followed by NNNNNN can be entered into the sample sheet to enable proper multiplex library parsing.

[00271] HiSeq Instrument

[00272] When setting up the HiSeq run, 12 bases of index sequencing (no sample sheet is required) can be specified. However, if a sample sheet to be included, the 6 bases of the actual barcode can be specified. In some cases, the N6 in your sample sheet is not included. The method described below can be used to parse and generate the N6 index fastq files for HiSeq using CASAVA: 1. Browse to the location of the run folder (called "RunFolder" in this

IDT Exhibit 1017.1650

example). 2. Open the "Data" folder in your run (/RunFolder/Data). 3. Run CASAVA. Use the "--use-bases-mask Y*, I6Y*" option to generate an Index fastq file along with the forward read (for paired end reads use "--use-bases-mask Y*,I6Y*, Y*"). For example:

"/illumina/pipeline/CASAVA-1.8.2/bin/configureBclToFastq.pl --input-dir /data/Runs/RunFolder/Data/Intensities/Basecalls/ --fastq-cluster-count 0 --use-basesmask
Y*,I6Y* --ignore-missing-control".

[00273] To generate the read and index fastq files without parsing, the –usebases- mask option to "--use-bases-mask Y*,Y*" can be modified. The generated fastq files can then be parsed using alternative software. 4. Open /RunFolder/Unaligned. 5. Type "nohup make -j 12" (this command will produce the actual fastq files, the 12 indicates how many threads to use). 6. The fastq files will be located in /RunFolder/Unaligned/Project_UniqueID

Example 9 -- Diversity trimming

[00274] Following sequencing and parsing, adaptors can be trimmed and the additional sequence added by the diversity adaptors to the 5' and 3' ends of the insert can be removed before downstream analysis.

[00275] Removing Adaptor Sequences

[00276] To accurately identify the diversity sequence and MspI (C^CGG) site, adaptor sequences present on the 3' end of your reads can be trimmed. Trim Galore (www.bioinformatics.babraham.ac.uk/projects/trim_galore/) can be used for this purpose. Other programs can also be used. Trim Galore may also trim some or all of a read due to low quality. In some cases, the program is run with default parameters, and the --RRBS option is not used. Single end reads can be trimmed with the following command:

trim_galore sampleR1.fq

[00277] Paired end reads can be trimmed with the following command:

trim_galore --adaptor2 AAATCAAAAAAAC sampleR1.fq sampleR2.fq

[00278] Diversity Trimming and Filtering

[00279] Additional filtering and trimming can be performed by a custom python script provided by NuGEN. The script can be obtained from NuGEN Technical Support at techserv@nugen.com. The script removes reads that do not contain an MspI site signature (YGG) at the 5' end. For paired end data an MspI site signature can be required at the 5' end of both sequences. The script will generate new file(s) with "_trimmed.fq" appended to the filename. In it, all reads have been trimmed at the 5' end to remove the diversity sequence (0–

3bases), and all reads can begin with YGG, where Y is C or T. On the 3' end, 5 bases are trimmed from every read (6 bases are trimmed for paired-end to prevent alignment issues).

[00280] Examples of Trimming the 5' Ends of Forward Reads

[00281] Underlined bases denote sequence derived from the adaptor. In this example, the fragment is derived from the genomic sequence, starting and ending with MspI sites:

5' CCGGAGTT...AAGGGCCGG 3'
3' GGCCTCAA...TTCCCGGCC 5'

[00282] After MspI digestion:

5' CGGAGTT...AAGGGC 3'
3' CTCAA...TTCCCGGC 5'

[00283] After ligation to adaptors, both with three bases of diversity:

5' RDDCGGAGTT...AAGGGCCGHHY 3'
3' YHHGCCTCAA...TTCCCGGCDDR 5'

[00284] After bisulfite conversion and PCR amplification of the top strand:

5' RDDYGGAGTT...AAGGGTCGHHY 3'
3' YHHRCCTCAA...TTCCCAGCDDR 5'

[00285] Assuming the insert is smaller than the read length, the forward read after Trim Galore is used to trim the adaptor from the 3' end will be:

5' RDDYGGAGTT...AAGGGTCGHHY 3'

[00286] The result of the NuGEN diversity trim of the forward read (if it's a single-end read) will be:

5' YGGAGTT...AAGGGT 3'

[00287] The reverse read after Trim Galore is used will be:

5' RDDCGACCCTT...AACTCCRHHY 3'

[00288] The result of the NuGEN diversity trim of the reverse read:

5' ACCCTT...AACT 3'

[00289] The adaptor can contain between 0 and 3 bases of diversity.

[00290] Tables 12–15 show how the script trims all types of adaptor variation.

IDT Exhibit 1017.1652

[00291] Table 12. Trimming of the 5' ends of forward reads

READ BEFORE TRIMMING	READ AFTER TRIMMING
5' YGGAGTT...	5' YGGAGTT...
5' DYGGAGTT...	5' YGGAGTT...
5' DDYGGAGTT...	5' YGGAGTT...
5' RDDYGGAGTT...	5' YGGAGTT...

Note: If YGG is not found in the first 6 bases, the read is discarded.

[00292] Table 13. Trimming of the 3' ends of forward reads

READ BEFORE TRIMMING	READ AFTER TRIMMING
5' ...AAGGGTCG	5' ...AAG
5' ...AAGGGTCGH	5' ...AAGG
5' ...AAGGGTCGHH	5' ...AAGGG
5' ...AAGGGTCGHHY	5' ...AAGGGT

[00293] Table 14. Trimming of the 5' ends of reverse reads

READ BEFORE TRIMMING	READ AFTER TRIMMING
5' CGACCCTT…	5' ACCCTT…
5' DCGACCCTT…	5' ACCCTT…
5' DDCGACCCTT…	5' ACCCTT…
5' RDDCGACCCTT…	5' ACCCTT…

Note: If CGR is not found in the first 6 bases, the read is discarded.

[00294] Table 15. Trimming of the 3' ends of reverse reads

READ BEFORE TRIMMING	READ AFTER TRIMMING
5' …AACTCCRHHY	5' ...AACTC
5' …AACTCCRHH	5' ...AACT
5' …AACTCCRH	5' ...AAC
5' …AACTCCR	5' ...AA

[00295] The script accepts as input one or two fastq file strings, given either as complete filenames or as a pattern in quotes. When a pattern is given, the script will find all the filenames matching a specified pattern according to the rules used by the Unix shell (*,?).

[00296] Example usage for single end reads with a complete filename:

python trimRRBSdiversityAdaptCustomers.py -1 sample_R1.fq

[00297] with a pattern:

python trimRRBSdiversityAdaptCustomers.py -1 "s_R1_1_edit3_6bp_BC??.fq"

[00298] Example usage for paired end reads with a complete filename:

python trimRRBSdiversityAdaptCustomers.py -1 sample_R1.fq -2 sample_R2.fq

[00299] with a pattern:

python trimRRBSdiversityAdaptCustomers.py -1 "s_R1*.fq" -2 "s_R2*.fq"

[00300] The help option of this script can be accessed for more details (-h). The script outputs a trimmed fastq file that can be used for downstream analysis including Bismark.

[00301] Following trimming, an additional algorithm can be used to remove sequencing duplicates. The Duplicate Marking tool can utilize information provided by the unique N6 sequence to discriminate between true PCR duplicates and independent adaptor ligation events to fragments with the same start site resulting in the recovery of more usable data. Using the index sequence reads, the Duplicate Marking tool can utilize the N6 random sequence found adjacent to the barcode sequence (as shown in **FIGs. 32** and **37**) in the sequence appended to nucleic acid inserts generated using the methods outlined in Example 4-8 to unambiguously identify other PCR duplicates comprising the same N6–barcode sequence and removing all but one of these reads.

Example 10 -- A simplified workflow for reduced representation bisulfite sequencing

[00302] RRBS can be used to obtain quantitative DNA methylation information across many features of the genome with approximately 50-fold fewer sequencing reads than can be needed with whole genome bisulfite sequencing which can result in substantial sequencing cost reduction. The methods described herein can be used for generating RRBS libraries in less than one day, without gel purification. The methods can entail digesting samples, ligating adaptors, performing a final repair, bisulfite converting, and PCR amplifying to generate libraries ready for NGS sequencing platforms such as Illumina. Additional benefits of the methods provided herein can include improved sequencing performance without the need for PhiX spike in, and the ability to mark PCR duplicates.

IDT Exhibit 1017.1654

[00303] DNA methylation can be a central component of epigenetic regulation, and can play a role in development, environmental exposure, and diseases including cancer. In humans, the predominant form of DNA methylation can be 5-methyl cytosine. When present, 5-mC can occur in the context of the CpG sequence motif. RRBS can utilize the restriction enzyme MspI, which can recognize and cleave the sequence CCGG regardless of the methylation state of the central CpG.

[00304] Reducing Representation to ~2%

[00305] After digestion with MspI, ~2% of the genome can be represented in fragments under 500 bp. However, many of these fragments can be from CpG-rich regions. By focusing RRBS sequencing on small fragments, the methylation status of these regions can be determined with much less sequencing than would be required using a whole genome bisulfite sequencing approach.

[00306] Challenges with Sequencing RRBS Libraries

[00307] On sequencers such as provided by Illumina, cluster identification as well as phasing and color matrix calculations can be negatively impacted by a lack of sequence complexity in the sample. RRBS libraries all begin with YGG (Y = C or T), which can make these libraries challenging to sequence. These issues can be mitigated by reducing cluster density and/or mixing in a high diversity library, such as a PhiX control library. This approach can reduce the number of useful reads as PhiX levels of 40% or higher may be needed to ensure a successful run.

[00308] Adding Diversity to RRBS Reads

[00309] The challenges described above can be overcome by adding diversity in the form of 0 to 3 bases of sequence between the sequencing primer and the insert. This can ensure all 4 bases are present during the first few cycles of sequencing. In addition, the YGG signature can be de-phased so that no cycle contains only G across all clusters.

[00310] The methods described herein can be multiplexed up to 16 samples per lane, can require no gel purification steps, and can have many advantages, including: simplified workflow that can be completed in a single day; adaptor features (i.e., sequence diversity) which can provide enhanced sequence quality without PhiX spike in, the ability to identify and remove PCR duplicates; and highly reproducible data that can displays excellent concordance with published methylation profiles such as shown in **FIG. 34**.

Example 11 -- Generation of and analysis of reduced representation bisulfite sequencing library from IMR-90 human genomic DNA using diversity adaptors

IDT Exhibit 1017.1655

[00311] In this example, an RRBS nucleic acid library from 25 ng of genomic DNA (gDNA) from the IMR-90 human cell line was generated using the kit, protocol, considerations, and analytical software outlined and described in Examples 4-9. Briefly, following digestion of 25 ng of gDNA from IMR-90 human cells with MspI to produce DNA fragments, the DNA fragments are combined with a master mix containing the appropriate diversity adapters (i.e., D0 ,D1, D2, D3, or a mix of D0-D3), ligation buffer, water and ligation enzyme mix, placed in a thermal cycler and run through a program on the thermal cycler that comprised of a single cycle comprising incubations at 25 C for 30 minutes, 70 C for 10 minutes, and then holding the mixture at 4 C. Following ligation, the ligation complexes were subjected to a post-ligation purification followed by a final repair reaction entailing incubation of the ligation complexes with a DNA repair enzyme in a thermal cycler programmed for a single cycle of 60 C incubation for 10 minutes. Following final repair, the samples were subjected to bisulfite conversion, wherein 20 uL of the products from the final repair reactions were inputted directly into a Qiagen EpiTect Fast DNA Bisulfite Kit according to manufacturer protocols and subjected to two cycles of 95 C for 5 minutes, and 60 C for 20 minutes on a thermal cycler followed by elution of the purified, bisulfite converted DNA libraries.. The bisulfite converted DNA libraries were then PCR amplified and purified. Following library amplification and purification, the libraries were pooled and sequenced on an Illumina MiSeq sequencer according to manufacturer instructions with a 40 nucleotide forward read, 6 nucleotide index read, and 25 nucleotide reverse read. The sequence reads were analyzed using the *in silico* diversity trimming and duplicate marking computer programs described in Example 9. This procedure was repeated two more times to generate technical replicates. The sequencing analysis was then compared between the replicates.

[00312] Following sequencing and *in silico* analysis of sequence reads obtained from the sequencing, a high coverage overlap can be seen between the three RRBS technical replicates as shown in **FIG. 33**. Moreover, as shown in Table 16, each of the technical replicates produced similar high levels of total mapping, as well as other measures.

[00313] Table 16. Replicate RRBS libraries from 25 ng IMR90 gDNA

	Replicate 1	Replicate 2	Replicate 3
Uniquely Mapping	67.60%	67.47%	66.98%
Non-Uniquely Mapping	23.12%	22.70%	23.93%

IDT Exhibit 1017.1656

Total Mapping	90.72%	90.18%	90.91%
Methyl CG	44.05%	44.14%	43.17%
Methyl CHG	0.37%	0.37%	0.36%
Methyl CHH	0.31%	0.31%	0.30%

[00314] Additionally, **FIG. 34** shows a concordance in methylation levels between IMR-90 RRBS replicates (left) and IMR-90 whole genome bisulfite sequencing (WGBS) libraries (right).

[00315] While preferred embodiments are shown and described herein, it will be obvious to those skilled in the art that such embodiments are provided by way of example only. Numerous variations, changes, and substitutions will now occur to those skilled in the art without departing from the instant disclosure. It should be understood that various alternatives to the embodiments described herein may be employed in practicing the disclosure. It is intended that the following claims define the scope of the invention and that methods and structures within the scope of these claims and their equivalents be covered thereby.

IDT Exhibit 1017.1657

CLAIMS

WHAT IS CLAIMED IS:

1. A method for increasing diversity of sequences at ends of a plurality of polynucleotides in a sample, the method comprising introducing a pool of oligonucleotide sequences to 3' ends of the plurality of polynucleotides,
   a. wherein each oligonucleotide sequence in the pool of oligonucleotide sequences comprises a sequence complementary to a sequencing primer,
   b. wherein the pool of oligonucleotide sequences comprises a first sequence, wherein a terminal 3' nucleotide base of the sequencing primer is capable of annealing to a terminal 5' nucleotide base of the first sequence, wherein following introduction the 5' end of the first sequence is adjacent to a 3' end of a polynucleotide from the plurality of polynucleotides,
   c. wherein the pool of oligonucleotide sequences comprises a second sequence, wherein the second sequence terminates at its 5' end with an additional base relative to the 5' end of the first sequence, wherein following introduction the 5' end of the second sequence is adjacent to a 3' end of a polynucleotide from the plurality of polynucleotides;
   d. wherein the pool of oligonucleotide sequences comprises a third sequence, wherein the third sequence terminates at its 5' end with two additional bases relative to the 5' end of the first sequence, wherein following introduction the 5' end of the third sequence is adjacent to a 3' end of a polynucleotide from the plurality of polynucleotides;
   e. wherein the pool of oligonucleotide sequences comprises a fourth sequence, wherein the fourth sequence terminates at its 5' end with three additional bases relative to the 5' end of the first sequence, wherein following introduction the 5' end of the fourth sequence is adjacent to a 3' end of a polynucleotide from the plurality of polynucleotides; and
   f. wherein the introducing generates a plurality of oligonucleotide-polynucleotide complexes, thereby increasing diversity of sequences at ends of the plurality of polynucleotides.
2. The method of claim 1, wherein the sample is a whole genome sample.
3. The method of claim 1, wherein the sample comprises a polynucleotide library.

IDT Exhibit 1017.1658

4. The method of any of claims 1-3, further comprising fragmenting the polynucleotides prior to introducing the pool of oligonucleotide sequences to the 3' ends of the plurality of polynucleotides.
5. The method of claim 4, wherein the fragmenting is performed by contacting the polynucleotides with an enzyme.
6. The method of claim 5, wherein the enzyme is a restriction enzyme.
7. The method of claim 6, wherein the restriction enzyme is a methylation insensitive restriction enzyme.
8. The method of claim 7, wherein the methylation insensitive restriction enzyme is MspI.
9. The method of claim 8, wherein the additional base of the second sequence is an adenine, thymine, or cytosine base.
10. The method of claim 8 or 9, wherein each of the two additional bases of the third sequence is an adenine, thymine, or cytosine base.
11. The method of any of claims 8-10 wherein a 3' most base of the three additional bases of the fourth sequence comprises a thymine or cytosine base, and each of the two other bases of the three additional bases of the fourth sequence comprises an adenine, thymine, or cytosine base.
12. The method of any of claims 1-11, wherein each oligonucleotide sequence in the pool of oligonucleotide sequences is within a strand of a duplexed adaptor.
13. The method of claim 12, wherein the duplexed adaptor is a partial duplexed adaptor comprising a short strand and a long strand.
14. The method of claim 12, wherein the duplexed adaptor is a partial duplexed adaptor comprising two strands of equal length.
15. The method of claim 13, wherein the short strand is a non-ligation strand of the duplexed adaptor, and the long strand is a ligation strand of the duplexed adaptor.
16. The method of claim 15, wherein the non-ligation strand comprises a 5' overhang.
17. The method of claim 16, wherein the 5' overhang comprises two bases.
18. The method of claim 17, wherein the two bases are 3'-GC-5'.
19. The method of any of the above claims, wherein the introducing comprises a ligation reaction.
20. The method of claim 19, wherein the ligation reaction comprises cohesive end ligation.

IDT Exhibit 1017.1659

21. The method of claim 19, wherein the ligation reaction comprises blunt end ligation.

22. The method of claim 20, wherein the plurality of polynucleotides are strands of double-stranded nucleic acids, wherein each 5' end of the double-stranded nucleic acids comprises a ligation end and each 3' end comprises a non-ligation end, wherein each of the non-ligation ends of the double-stranded nucleic acid comprises part of the polynucleotide sequence.

23. The method of claim 22, wherein the 3'-GC-5' of the non-ligation strand of the duplexed adaptor anneals to a 5'-CG-3' sequence of the ligation end of the double-stranded sequence comprising the polynucleotide sequence.

24. The method of claim 23, wherein the 3'-GC-5' of the non-ligation strand of the duplexed adaptor is 5' to the terminal 5' nucleotide base of the first sequence.

25. The method of claim 23 or 24, wherein the 3'-GC-5' of the non-ligation strand of the duplexed adaptor is 5' to the additional base of the second sequence relative to the 5' end of the first sequence.

26. The method of any of claims 23-25, wherein the 3'-GC-5' of the non-ligation strand of the duplexed adaptor is 5' to the two additional bases of the third sequence relative to the 5' end of the first sequence.

27. The method of any of claims 23-26, wherein the 3'-GC-5' of the non-ligation strand of the duplexed adaptor is 5' to the three additional bases of the fourth sequence relative to the 5' end of the first sequence.

28. The method of any of claims 16-20, wherein the ligation strand is capable of ligating to a 5' end of the plurality of polynucleotides, and wherein a nick is formed between the non-ligation strand and the end of the plurality of polynucleotides.

29. The method of claim 28, further comprising an extending step, wherein the non-ligation end of the plurality of polynucleotides is extended using a polymerase in the presence of a mixture of dNTPs using the ligation strand of the adaptor as template, thereby introducing the pool of oligonucleotide sequences to the 3' ends of the plurality of polynucleotides.

30. The method of claim 29, wherein the ligation strand of each adaptor duplex in the pool comprise cytosine analogs resistant to bisulfite treatment and the mixture of dNTPs comprises dATP, dTTP, dGTP, and dCTP, whereby the extending generates the plurality of oligonucleotide sequence-polynucleotide complexes, wherein each of the plurality of oligonucleotide sequence-polynucleotide complexes comprises complementary ends,

IDT Exhibit 1017.1660

wherein one end comprises cytosine bases sensitive to bisulfite treatment, while the opposite end comprises cytosine analogs resistant to bisulfite treatment.

31. The method of claim 29, wherein the ligation strand of each adaptor duplex in the pool of adaptor duplexes comprise cytosine bases sensitive to bisulfite treatment and the mixture of dNTPs comprises dATP, dTTP, dGTP, and a dCTP analog resistant to bisulfite treatment, whereby the extending generates the plurality of oligonucleotide sequence-polynucleotide complexes, wherein each of the plurality of oligonucleotide sequence-polynucleotide complexes comprises complementary ends, wherein one end comprises cytosine bases sensitive to bisulfite treatment, while the opposite end comprises cytosine analogs resistant to bisulfite treatment.

32. The method of any of claims 1-3, wherein each oligonucleotide sequence in the pool of oligonucleotide sequences is a primer, wherein the pool of oligonucleotide sequences comprises the primer and a complement thereof.

33. The method of claim 32, wherein the introducing comprises an amplification reaction.

34. The method of claim 33, wherein the amplification reaction comprises polymerase chain reaction.

35. The method of claim 33 or 34, wherein each of the primers in the pool comprises cytosine bases sensitive to bisulfite treatment, wherein the amplification is conducted in the presence of a mixture of dNTPs comprising dATP, dTTP, dGTP, and a dCTP analog resistant to bisulfite treatment, thereby generating the plurality of oligonucleotide sequence-polynucleotide complexes, wherein each of the plurality of oligonucleotide sequence-polynucleotide complexes comprises complementary ends, wherein one end comprises cytosine bases sensitive to bisulfite treatment, while the opposite end comprises cytosine analogs resistant to bisulfite treatment.

36. The method of claim 33 or 34, wherein each of the primers in the pool of primers comprises cytosine analogs resistant to bisulfite treatment, wherein the amplification is conducted in the presence of a mixture of dNTPs comprising dATP, dTTP, dGTP, and dCTP, thereby generating the plurality of oligonucleotide sequence-polynucleotide complexes, wherein each of the plurality of oligonucleotide sequence-polynucleotide complexes comprises complementary ends, wherein one end comprises cytosine bases sensitive to bisulfite treatment, while the opposite end comprises cytosine analogs resistant to bisulfite treatment.

IDT Exhibit 1017.1661

37. The method of claims 30, 31, 35, or 36, wherein the cytosine analog resistant to bisulfite treatment is 5-methylcytosine, 5-hydroxymethylcytosine, or 5-propynylcytosine.

38. The method of claims 30, 31, 35, or 36, wherein the cytosine analog resistant to bisulfite treatment is 5-methylcytosine.

39. The method of any of the above claims, further comprising amplifying the plurality of oligonucleotide sequence-polynucleotide complexes, wherein the amplifying comprises solid-phase nucleic acid amplification to generate a plurality of clusters, wherein each cluster represents a plurality of copies of each of the plurality of oligonucleotide sequence-polynucleotide complexes.

40. The method of claim 39, further comprising treating the plurality of oligonucleotide sequence-polynucleotides complexes with bisulfite prior to the amplification, wherein the treating generates a plurality of oligonucleotide sequence-polynucleotide complexes comprising non-complementary ends.

41. The method of claim 40, wherein the treating with bisulfite prior to amplification generates the sequence complementary to a sequencing primer in each oligonucleotide sequence in the pool of oligonucleotide sequences comprises.

42. The method of any of claims 39-41, further comprising sequencing each of the plurality of clusters, wherein the sequencing comprises hybridizing the sequencing primer to the sequence complementary to a sequencing primer, and conducting sequencing by synthesis.

43. The method of claim 42, wherein the introducing of the pool of oligonucleotide sequences to the polynucleotides improves identification of each of the plurality of clusters generated from the plurality of oligonucleotide sequence-polynucleotide complexes relative to the identification of each cluster from a plurality of clusters generated from the plurality of polynucleotides not introduced to the pool of oligonucleotide sequences.

44. The method of claim 42, wherein the sequencing comprises paired-end sequencing.

45. A method for generating a bisulfite converted library, the method comprising:

a.) introducing a pool of oligonucleotide sequences to each end of a plurality of polynucleotides from a sample,

i. wherein each oligonucleotide sequence in the pool of oligonucleotide sequences comprises a sequence complementary to a sequencing primer and a

IDT Exhibit 1017.1662

sequence element adjacent to the sequence complementary to a sequencing primer,

ii. wherein the pool of oligonucleotide sequences comprises a first sequence, wherein the first sequence terminates at its 3' end with the sequence element, wherein following ligation the 3' end of the sequence element is adjacent to a 5' end of a polynucleotide from the plurality of polynucleotides,

iii. wherein the pool of oligonucleotide sequences comprises a second sequence, wherein the second oligonucleotide sequence terminates at its 3' end with an additional base relative to the 3' end of the first sequence, wherein following ligation the 3' end of the second sequence is adjacent to a 5' end of a polynucleotide from the plurality of polynucleotides;

iv. wherein the pool of oligonucleotide sequences comprises a third sequence, wherein the third sequence terminates at its 3' end with two additional bases relative to the 3' end of the first sequence, wherein following ligation the 3' end of the third sequence is adjacent to a 5' end of a polynucleotide from the plurality of polynucleotides;

v. wherein the pool of oligonucleotide sequences comprises a fourth sequence, wherein the fourth sequence terminates at its 3' end with three additional bases relative to the 3' end of the first sequence, wherein following ligation the 3' end of the fourth sequence is adjacent to a 5' end of a polynucleotide from the plurality of polynucleotides;

vi. wherein the introducing generates a plurality of oligonucleotide sequence-polynucleotide complexes; and

b.) treating the plurality of oligonucleotide sequence-polynucleotide complexes with bisulfite, thereby generating a bisulfite converted polynucleotide library.

46. The method of claim 45, wherein the sample is a whole genome sample.

47. The method of claim 45, wherein the sample comprises a polynucleotide library.

48. The method of any of claims 45-47, further comprising fragmenting the polynucleotides prior to introducing the pool of oligonucleotide sequences to the 3' ends of the plurality of polynucleotides.

49. The method of claim 48, wherein the fragmenting is performed by contacting the polynucleotides with an enzyme.

50. The method of claim 49, wherein the enzyme is a restriction enzyme.

IDT Exhibit 1017.1663

51. The method of claim 50, wherein the restriction enzyme is a methylation insensitive restriction enzyme.

52. The method of claim 51, wherein the methylation insensitive restriction enzyme is MspI.

53. The method of claim 52, wherein the additional base of the second sequence is an adenine, thymine, or guanine base,

54. The method of claim 52 or 53, wherein each of the two additional bases of the third sequence is an adenine, thymine, or guanine base.

55. The method of any of claims 52-54 wherein a 5' most base of the three additional bases of the fourth sequence comprises an adenine or guanine base, and each of the two other bases of the three additional bases of the fourth sequence comprises an adenine, thymine, or guanine base.

56. The method of any of claims 45-55, wherein each oligonucleotide sequence in the pool of oligonucleotide sequences is within a strand of a duplexed adaptor.

57. The method of claim 56, wherein the duplexed adaptor is a partial duplexed adaptor.

58. The method of claim 57, wherein the strand is a ligation strand of the duplexed adaptor.

59. The method of claim 58, wherein the non-ligation strand comprises a 5' overhang.

60. The method of claim 59, wherein the 5' overhang comprises two bases.

61. The method of claim 60, wherein the two bases are 3'-GC-5'.

62. The method of any of claims 45-61, wherein the introducing comprises a ligation reaction.

63. The method of claim 62, wherein the ligation reaction comprises cohesive end ligation.

64. The method of claim 62, wherein the ligation reaction comprises blunt end ligation.

65. The method of claim 63, wherein the plurality of polynucleotides are double-stranded nucleic acids, wherein each 5' end of the double-stranded nucleic acids comprises a ligation end and each 3' end comprises a non-ligation end.

66. The method of any of claims 63-65, wherein each duplexed adaptor in the pool comprises a ligation strand and a non-ligation strand, wherein the ligation strand is capable of ligating to the ligation end of the plurality of polynucleotides, and wherein a

IDT Exhibit 1017.1664

nick is formed between the non-ligation strand and the non-ligation end of the plurality of polynucleotides.

67. The method of claim 66, further comprising an extending step, wherein the non-ligation end of the plurality of polynucleotides is extended using a polymerase in the presence of a mixture of dNTPs using the ligation strand of the adaptor as template, thereby introducing the pool of oligonucleotide sequences to the 3' ends of the plurality of polynucleotides.

68. The method of claim 67, wherein the ligation strand of each adaptor duplex in the pool comprise cytosine analogs resistant to bisulfite treatment and the mixture of dNTPs comprises dATP, dTTP, dGTP, and dCTP, whereby the extending generates the plurality of oligonucleotide sequence-polynucleotide complexes, wherein each of the plurality of oligonucleotide sequence-polynucleotide complexes comprises complementary ends, wherein one end comprises cytosine bases sensitive to bisulfite treatment, while the opposite end comprises cytosine analogs resistant to bisulfite treatment.

69. The method of claim 68, wherein the ligation strand of each adaptor duplex in the pool of adaptor duplexes comprise cytosine bases sensitive to bisulfite treatment and the mixture of dNTPs comprises dATP, dTTP, dGTP, and a dCTP analog resistant to bisulfite treatment, whereby the extending generates the plurality of oligonucleotide sequence-polynucleotide complexes, wherein each of the plurality of oligonucleotide sequence-polynucleotide complexes comprises complementary ends, wherein one end comprises cytosine bases sensitive to bisulfite treatment, while the opposite end comprises cytosine analogs resistant to bisulfite treatment.

70. The method of any of claims 45-47, wherein each oligonucleotide sequence in the pool of oligonucleotide sequences is within a primer, wherein the pool of oligonucleotide sequences comprises the primer and a complement thereof.

71. The method of claim 70, wherein the introducing comprises an amplification reaction.

72. The method of claim 71, wherein the amplification reaction comprises polymerase chain reaction.

73. The method of claim 71 or 72, wherein each of the primers in the pool comprises cytosine bases sensitive to bisulfite treatment, wherein the amplification is conducted in the presence of a mixture of dNTPs comprising dATP, dTTP, dGTP, and a dCTP analog resistant to bisulfite treatment, thereby generating the plurality of oligonucleotide sequence -polynucleotide complexes, wherein each of the plurality of oligonucleotide

IDT Exhibit 1017.1665

sequence -polynucleotide complexes comprises complementary ends, wherein one end comprises cytosine bases sensitive to bisulfite treatment, while the opposite end comprises cytosine analogs resistant to bisulfite treatment.

74. The method of claim 71 or 72, wherein each of the primers in the pool of primers comprises cytosine analogs resistant to bisulfite treatment, wherein the amplification is conducted in the presence of a mixture of dNTPs comprising dATP, dTTP, dGTP, and dCTP, thereby generating the plurality of oligonucleotide sequence-polynucleotide complexes, wherein each of the plurality of oligonucleotide sequence-polynucleotide complexes comprises complementary ends, wherein one end comprises cytosine bases sensitive to bisulfite treatment, while the opposite end comprises cytosine analogs resistant to bisulfite treatment.

75. The method of claims 68, 69, 73, or 74, wherein the cytosine analog resistant to bisulfite treatment is 5-methylcytosine, 5-hydroxymethylcytosine, or 5-propynylcytosine.

76. The method of claims 68, 69, 73, or 74, wherein the cytosine analog resistant to bisulfite treatment is 5-methylcytosine.

77. The method of any of claims 45-76, further comprising amplifying the plurality of amplifying the plurality of oligonucleotide sequence-polynucleotide complexes.

78. The method of claim 77, wherein the amplifying comprises solid-phase nucleic acid amplification to generate a plurality of clusters, wherein each cluster represents a plurality of copies of each of the plurality of oligonucleotide sequence-polynucleotide complexes.

79. The method of claim 77, further comprising treating the plurality of oligonucleotide sequence-polynucleotides complexes with bisulfite prior to the amplification, wherein the treating generates a plurality of oligonucleotide sequence-polynucleotide complexes comprising non-complementary ends.

80. The method of claim 79, wherein the sequence complementary to a sequencing primer in one end of the plurality of oligonucleotide sequence-polynucleotide complexes comprising non-complementary ends is complementary to a reverse read sequencing primer, and wherein the sequence complementary to a sequencing primer in an opposite end of the non-complementary ends in the plurality of oligonucleotide sequence-polynucleotide complexes comprising non-complementary ends is complementary to a forward read sequencing primer.

81. The method of any of claims 78-80, further comprising sequencing each of the plurality of clusters, wherein the sequencing comprises hybridizing the sequencing

IDT Exhibit 1017.1666

primer to the sequence complementary to a sequencing primer, and conducting sequencing by synthesis, wherein the introducing of the pool of oligonucleotide sequences to the polynucleotides improves identification of each of the plurality of clusters generated from the plurality of oligonucleotide sequence-polynucleotide complexes relative to the identification of each cluster from a plurality of clusters generated from the plurality of polynucleotides not introduced to the pool of oligonucleotide sequences.

82. The method of any of claims 45-81, wherein the sequence element comprises a barcode, a universal sequence, a linker sequence, or a random sequence.

83. A kit comprising a pool of oligonucleotide sequences,

a. wherein each oligonucleotide sequence in the pool of oligonucleotide sequences comprises a sequence complementary to a sequencing primer,

b. wherein the pool of oligonucleotide sequences comprises a first oligonucleotide sequence, wherein a terminal 3' nucleotide base of the sequencing primer is capable of annealing to a terminal 5' nucleotide base of the first oligonucleotide sequence,

c. wherein the pool of oligonucleotide sequences comprises a second oligonucleotide sequence, wherein the second oligonucleotide sequence terminates at a 5' end with an additional base relative to the 5' end of the first oligonucleotide sequence;

d. wherein the pool of oligonucleotide sequences comprises a third oligonucleotide sequence, wherein the third oligonucleotide sequence terminates at a 5' end with two additional bases relative to the 5' end of the first oligonucleotide sequence; and

e. wherein the pool of oligonucleotide sequences comprises a fourth oligonucleotide sequence, wherein the fourth oligonucleotide sequence terminates at a 5' end with three additional bases relative to the 5' end of the first oligonucleotide sequence.

84. The kit of claim 83, wherein each oligonucleotide sequence in the pool of oligonucleotide sequences is within a primer, wherein the pool of oligonucleotide sequences comprises the primer and a complement thereof

85. The kit of claim 84, wherein the primer comprises cytosine analogs resistant to bisulfite conversion.

86. The kit of claim 84, wherein the primer comprises cytosine bases sensitive to bisulfite conversion.

IDT Exhibit 1017.1667

87. The kit of claim 83, wherein each oligonucleotide sequence in the pool of oligonucleotide sequences is within a strand of a duplexed adaptor.

88. The kit of claim 87, wherein each duplexed adaptor in the pool comprises a ligation strand and a non-ligation strand.

89. The kit of claim 88, wherein the ligation strand and the non-ligation strand of each of duplexed adaptor in the pool comprise cytosine bases sensitive to bisulfite treatment.

90. The kit of claim 89, further comprising a mixture of dNTPs comprising dATP, dTTP, dGTP, and a dCTP analog resistant to bisulfite treatment.

91. The kit of claim 88, wherein the ligation strand and the non-ligation strand of each of duplexed adaptor in the pool comprise cytosine analogs resistant to bisulfite treatment.

92. The kit of claim 91, further comprising a mixture of dNTPs comprising dATP, dTTP, dGTP, and a dCTP analog sensitive to bisulfite treatment.

93. The kit of any of claims 83-92, further comprising a restriction enzyme.

94. The kit of claim 93, wherein the restriction enzyme is MspI.

95. The kit of any of claims 83-94, further comprising a DNA polymerase.

96. The kit of claim 95, wherein the DNA polymerase comprises exonuclease activity.

97. The kit of claim 95, wherein the DNA polymerase comprises strand displacement activity.

98. The kit of any of claims 88, 89, or 91, wherein the non-ligation strand comprises a block at a 5' and 3' end, whereby the non-ligation strand is enzymatically unreactive.

99. The kit of claim 98, wherein the block at the 5' end comprises a biotin moiety.

100. The kit of claim 98 or 99, wherein the block at the 3' end comprises a dideoxy residue.

101. The kit of claims 85, 90, or 91, wherein the cytosine analog resistant to bisulfite treatment is 5-methylcytosine, 5-hydroxymethylcytosine, or 5-propynylcytosine.

102. The kit of claims 85, 90, or 91, wherein the cytosine analog resistant to bisulfite treatment is 5-methylcytosine.

103. A method for analyzing sequence reads, the method comprising:

a. generating a plurality of sequence reads, wherein the generating comprises conducting sequencing by synthesis on a plurality of clusters, wherein each cluster of the plurality of clusters comprises copies of one oligonucleotide sequence-polynucleotide from a

IDT Exhibit 1017.1668

plurality of oligonucleotide sequence-polynucleotides, wherein the sequencing by synthesis comprises hybridizing a sequencing primer to sequence complementary to the sequencing primer present in each cluster of the plurality of clusters, wherein the plurality of clusters are produced from solid-phase nucleic acid amplification of the plurality of oligonucleotide sequence-polynucleotides, wherein each of the plurality of oligonucleotide sequence-polynucleotides comprises a single nucleotide base extender sequence, a dinucleotide base extender sequence, a trinucleotide base extender sequence or no nucleotide base extender sequence between the sequence complementary to the sequencing primer of an oligonucleotide sequence and an end of a polynucleotide; and

b. analyzing each of the plurality of sequence reads, wherein the analyzing comprises trimming or removing the extender sequence *in silico* from each of the plurality of sequence reads.

104. The method of claim 103, wherein the plurality of oligonucleotide sequence-polynucloetides are generated by introducing a pool of oligonucleotide sequences to 3' ends of the plurality of polynucleotides from a sample,

a. wherein each oligonucleotide sequence in the pool of oligonucleotide sequences comprises a sequence complementary to a sequencing primer,

b. wherein the pool of oligonucleotide sequences comprises a first sequence, wherein a terminal 3' nucleotide base of the sequencing primer is capable of annealing to a terminal 5' nucleotide base of the first sequence, wherein following introduction the 5' end of the first sequence is adjacent to a 3' end of a polynucleotide from the plurality of polynucleotides,

c. wherein the pool of oligonucleotide sequences comprises a second sequence, wherein the second sequence terminates at its 5' end with an additional base relative to the 5' end of the first sequence comprising the single nucleotide base extender, wherein following introduction the 5' end of the second sequence is adjacent to a 3' end of a polynucleotide from the plurality of polynucleotides;

d. wherein the pool of oligonucleotide sequences comprises a third sequence, wherein the third sequence terminates at its 5' end with two additional bases relative to the 5' end of the first sequence comprising the dinucleotide base extender, wherein following introduction the 5' end of the third sequence is adjacent to a 3' end of a polynucleotide from the plurality of polynucleotides;

e. wherein the pool of oligonucleotide sequences comprises a fourth sequence, wherein the fourth sequence terminates at its 5' end with three additional bases

IDT Exhibit 1017.1669

relative to the 5' end of the first sequence comprising the trinucleotide base extender, wherein following introduction the 5' end of the fourth sequence is adjacent to a 3' end of a polynucleotide from the plurality of polynucleotides; and

f. wherein the introducing generates a plurality of oligonucleotide-polynucleotide complexes, thereby increasing diversity of sequences at ends of the plurality of polynucleotides.

105. The method of claim 104, wherein the sample is a whole genome sample.

106. The method of claim 104, wherein the sample comprises a polynucleotide library.

107. The method of any of claims 104-106, further comprising fragmenting the polynucleotides prior to introducing the pool of oligonucleotide sequences to the 3' ends of the plurality of polynucleotides.

108. The method of claim 107, wherein the fragmenting is performed by contacting the polynucleotides with an enzyme.

109. The method of claim 108, wherein the enzyme is a restriction enzyme.

110. The method of claim 109, wherein the restriction enzyme is a methylation insensitive restriction enzyme.

111. The method of claim 110, wherein the methylation insensitive restriction enzyme is MspI.

112. The method of claim 111, wherein the additional base of the second sequence is an adenine, thymine, or cytosine base.

113. The method of claim 111 or 112, wherein each of the two additional bases of the third sequence is an adenine, thymine, or cytosine base.

114. The method of any of claims 111-113 wherein a 3' most base of the three additional bases of the fourth sequence comprises a thymine or cytosine base, and each of the two other bases of the three additional bases of the fourth sequence comprises an adenine, thymine, or cytosine base.

115. The method of any of claims 104-114, wherein each oligonucleotide sequence in the pool of oligonucleotide sequences is within a strand of a duplexed adaptor.

116. The method of claim 115, wherein the duplexed adaptor is a partial duplexed adaptor comprising a long strand and a short strand.

117. The method of claim 116, wherein the short strand is a non-ligation strand of the duplexed adaptor and the long strand is a ligation strand.

118. The method of claim 117, wherein the non-ligation strand comprises a 5' overhang.

IDT Exhibit 1017.1670

119. The method of claim 118, wherein the 5' overhang comprises two bases.

120. The method of claim 119, wherein the two bases are 3'-GC-5'.

121. The method of any of claims 104-120, wherein the introducing comprises a ligation reaction.

122. The method of claim 121, wherein the ligation reaction comprises cohesive end ligation.

123. The method of claim 121, wherein the ligation reaction comprises blunt end ligation.

124. The method of claim 122, wherein the plurality of polynucleotides are strands of double-stranded nucleic acids, wherein each 5' end of the double-stranded nucleic acids comprises a ligation end and each 3' end comprises a non-ligation end, wherein each of the non-ligation ends of the double-stranded nucleic acid comprises part of the polynucleotide sequence.

125. The method of claim 124, wherein the 3'-GC-5' of the non-ligation strand of the duplexed adaptor anneals to a 5'-CG-3' sequence of the ligation end of the double-stranded sequence comprising the polynucleotide sequence.

126. The method of claim 125, wherein the 3'-GC-5' of the non-ligation strand of the duplexed adaptor is 5' to the terminal 5' nucleotide base of the first sequence,

127. The method of claim 125 or 126, wherein the 3'-GC-5' of the non-ligation strand of the duplexed adaptor is 5' to the additional base of the second sequence relative to the 5' end of the first sequence.

128. The method of any of claims 125-127, wherein the 3'-GC-5' of the non-ligation strand of the duplexed adaptor is 5' to the two additional bases of the third sequence relative to the 5' end of the first sequence.

129. The method of any of claims 125-128, wherein the 3'-GC-5' of the non-ligation strand of the duplexed adaptor is 5' to the three additional bases of the fourth sequence relative to the 5' end of the first sequence.

130. The method of any of claims 118-122, wherein each duplexed adaptor in the pool comprises a ligation strand and a non-ligation strand, wherein the ligation strand is capable of ligating to a 5' end of the plurality of polynucleotides, and wherein a nick is formed between the non-ligation strand and the 3' end of the plurality of polynucleotides.

131. The method of claim 130, further comprising an extending step, wherein the 3' end of the plurality of polynucleotides is extended using a polymerase in the presence of a mixture of dNTPs using the ligation strand of the adaptor as template, thereby

IDT Exhibit 1017.1671

introducing the pool of oligonucleotide sequences to the 3' ends of the plurality of polynucleotides.

132. The method of claim 131, wherein the ligation strand of each adaptor duplex in the pool comprise cytosine analogs resistant to bisulfite treatment and the mixture of dNTPs comprises dATP, dTTP, dGTP, and dCTP, whereby the extending generates the plurality of oligonucleotide-polynucleotide complexes, wherein each of the plurality of oligonucleotide-polynucleotide complexes comprises complementary ends, wherein one end comprises cytosine bases sensitive to bisulfite treatment, while the opposite end comprises cytosine analogs resistant to bisulfite treatment.

133. The method of claim 131, wherein the ligation strand of each adaptor duplex in the pool of adaptor duplexes comprise cytosine bases sensitive to bisulfite treatment and the mixture of dNTPs comprises dATP, dTTP, dGTP, and a dCTP analog resistant to bisulfite treatment, whereby the extending generates the plurality of oligonucleotide-polynucleotide complexes, wherein each of the plurality of oligonucleotide-polynucleotide complexes comprises complementary ends, wherein one end comprises cytosine bases sensitive to bisulfite treatment, while the opposite end comprises cytosine analogs resistant to bisulfite treatment.

134. The method of any of claims 104-106, wherein each oligonucleotide in the pool of oligonucleotides is a primer, wherein the pool of oligonucleotides comprises the primer and a complement thereof.

135. The method of claim 134, wherein the introducing comprises an amplification reaction.

136. The method of claim 135, wherein the amplification reaction comprises polymerase chain reaction.

137. The method of claim 135 or 136, wherein each of the primers in the pool comprises cytosine bases sensitive to bisulfite treatment, wherein the amplification is conducted in the presence of a mixture of dNTPs comprising dATP, dTTP, dGTP, and a dCTP analog resistant to bisulfite treatment, thereby generating the plurality of oligonucleotide sequence-polynucleotide complexes, wherein each of the plurality of oligonucleotide sequence-polynucleotide complexes comprises complementary ends, wherein one end comprises cytosine bases sensitive to bisulfite treatment, while the opposite end comprises cytosine analogs resistant to bisulfite treatment.

138. The method of claim 135 or 136, wherein each of the primers in the pool of primers comprises cytosine analogs resistant to bisulfite treatment, wherein the

IDT Exhibit 1017.1672

amplification is conducted in the presence of a mixture of dNTPs comprising dATP, dTTP, dGTP, and dCTP, thereby generating the plurality of oligonucleotide sequence-polynucleotide complexes, wherein each of the plurality of oligonucleotide sequence-polynucleotide complexes comprises complementary ends, wherein one end comprises cytosine bases sensitive to bisulfite treatment, while the opposite end comprises cytosine analogs resistant to bisulfite treatment.

139. The method of claims 132, 133, 137, or 138, wherein the cytosine analog resistant to bisulfite treatment is 5-methylcytosine, 5-hydroxymethylcytosine, or 5-propynylcytosine.

140. The method of claim 137-139, further comprising treating the plurality of oligonucleotide sequence-polynucleotides complexes with bisulfite prior to the amplification, wherein the treating generates a plurality of oligonucleotide sequence-polynucleotide complexes comprising non-complementary ends.

141. The method of claim 140, wherein the treating with bisulfite prior to amplification generates the sequence complementary to a sequencing primer in each oligonucleotide sequence in the pool of oligonucleotide sequences comprises.

142. The method of any of claims 103-141, wherein the generating the plurality of sequence reads comprises paired-end sequencing.

143. A method for increasing diversity of sequences at ends of a plurality of polynucleotide fragments in a library, the method comprising:

a. fragmenting polynucleotides to generate polynucleotide fragments;
b. ligating a pool of adaptor sequences to 3' ends of the plurality of polynucleotide fragments, wherein each adaptor sequence in the pool of adaptor sequences comprises a sequence complementary to a sequencing primer, wherein a 3' end of the sequencing primer anneals at least four bases from the 5' end of the adaptor sequence, and wherein each of four different types of nucleotide bases are present in each of the at least three terminal nucleotides at the 5' ends of the adaptors, thereby increase diversity of sequences at ends of the plurality of polynucleotide fragments.

IDT Exhibit 1017.1673

FIG. 1

unknown methylation status
Non-methylated
Methylated
digest with MspI
End repair
Ligate adaptor, bisulfite convert, PCR amplify
sequence
5'Forward (CorT)GG - insert TTG Reverse 3' (top strand shown)
Forward Read: 5' (C or T)GG (bisulfite converted insert C-to-T strand)
Reverse Read: 5' CAA (bisulfite converted insert G-to-A strand)

FIG. 2A

```
Ben383   /5Biosg/CGHHHAGATCGGAAGAG/3ddC/     (H = A,C, or T, (no G))
Ben384   /5Biosg/CGHHAGATCGGAAGAG/3ddC/
Ben385    /5Biosg/CGHAGATCGGAAGAG/3ddC/
Ben341    /5Biosg/CGAGATCGGAAGAG/3ddC/

R01446   CAAGCAGAAGACGGCATACGTGATCTGGTTGTGACTGGAGTTCTGACGTGTGCTCTTCCGATCT
```

FIG. 2B

1. Anneal

```
                    R01446
                      ↓
5'-CAAGCAGAAGACGGCATACGTGATCTGGTTGTGACTGGAGTTCTGACGTGTGCTCTTCCGATCT
                                                  3dddCGHHHAGATCGGAAGAG/5Biosg
                                                                    ↑
                                                                  Ben383
```

2. Extend with D mix (dATP, dGTP, dTTP (no dCTP))

```
                    R01446
                      ↓
5'-CAAGCAGAAGACGGCATACGTGATCTGGTTGTGACTGGAGTTCTGACGTGTGCTCTTCCGATCTDDD
                                                  3dddCGHHHAGATCGGAAGAG/5Biosg
                                                                    ↑
                                                                  Ben383
```

FIG. 3

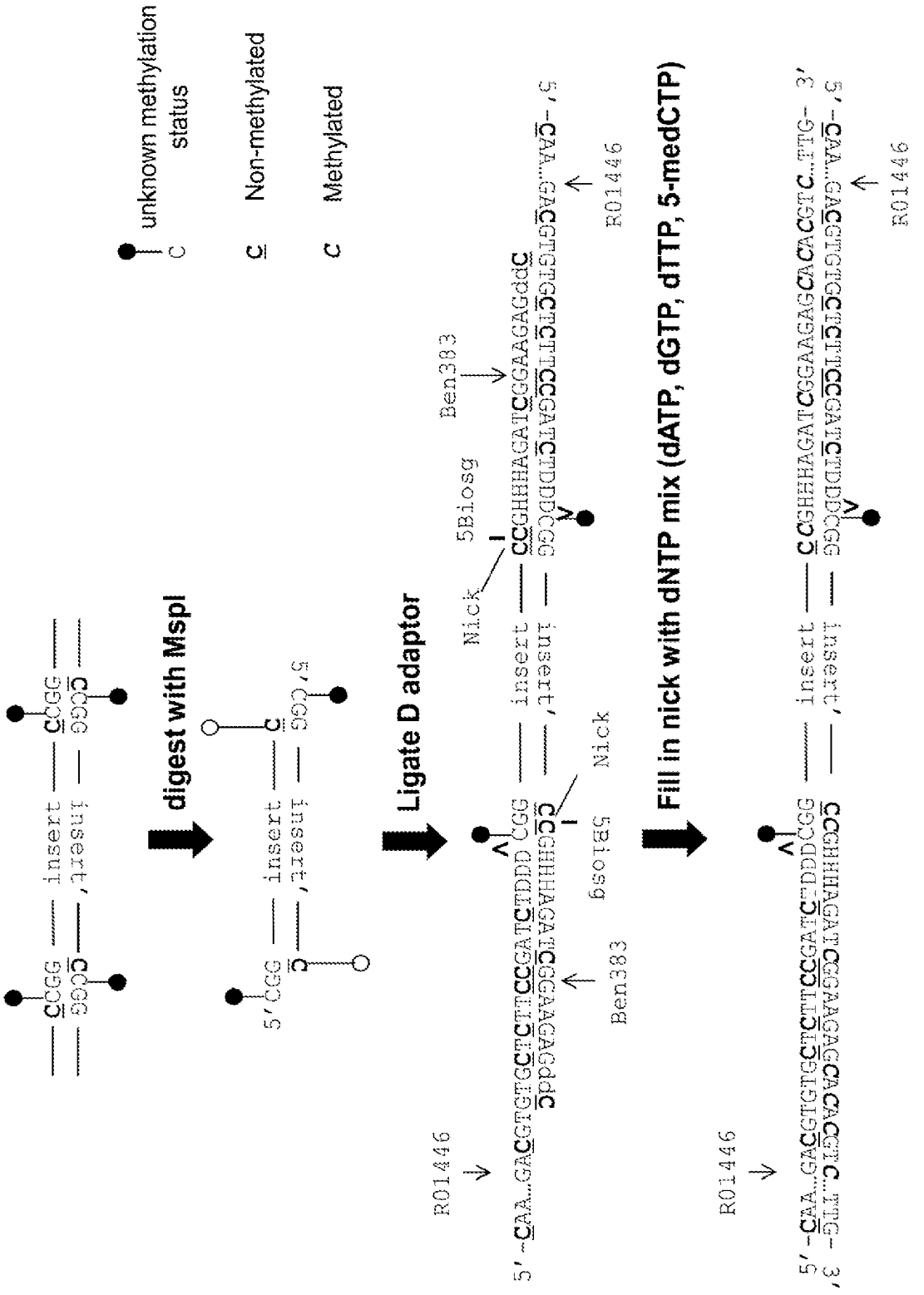

unknown methylation status
Non-methylated
Methylated
digest with MspI
Ligate D adaptor
Fill in nick with dNTP mix (dATP, dGTP, dTTP, 5-medCTP)
insert
Nick
5Biosg
Ben383
R01446

FIG. 3 (continued)

FIG. 4A

DR Rev Helper Style Ed3 BC

R01446 Ben383 D3 1

Expected Forward Read:

DDDYGG…

H = A,C, or T

D = A,G, or T

Y = C or T

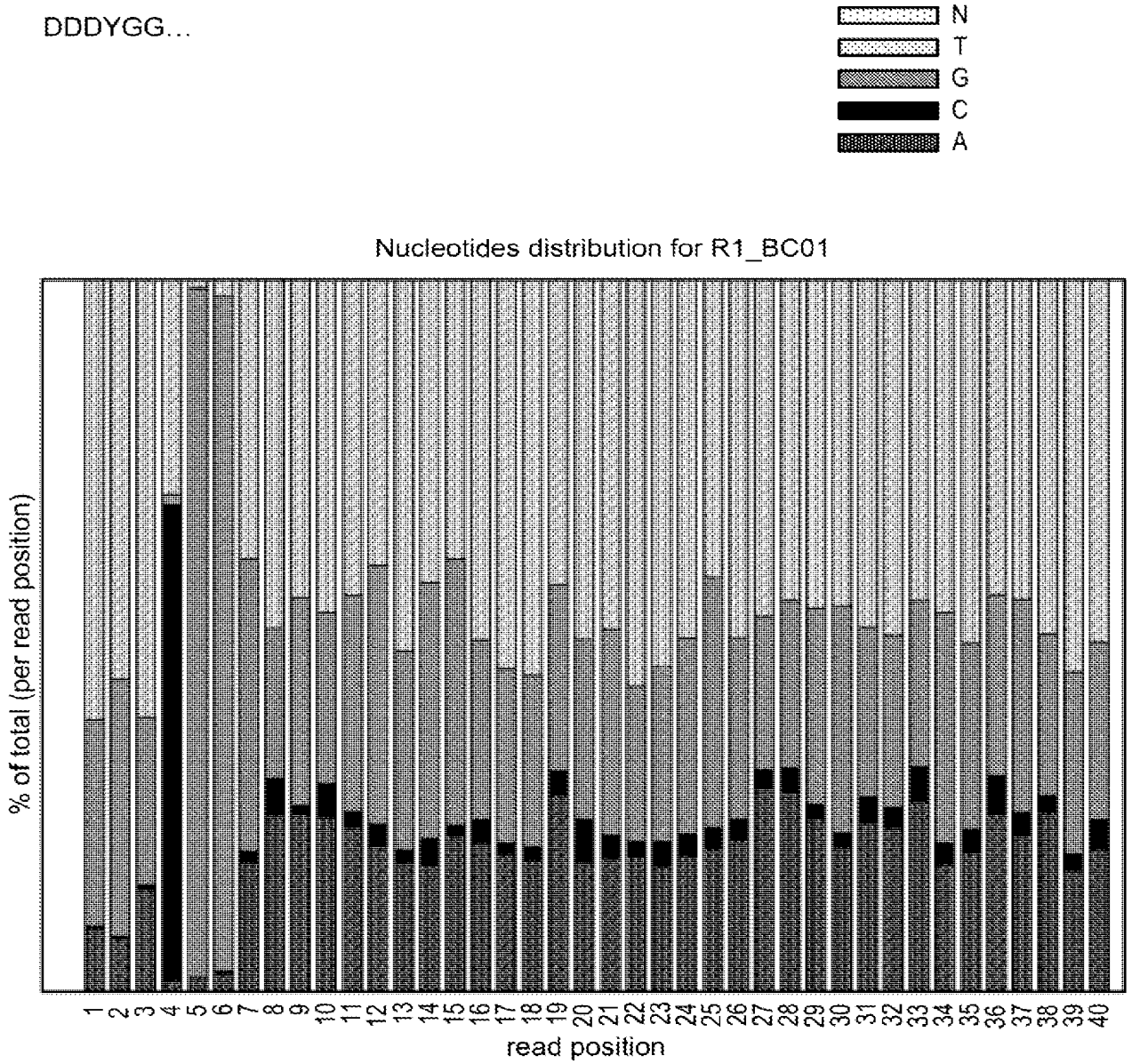


FIG. 4B

DR Rev Helper Style Ed3 BC

R01446 Ben383 D3 1

Expected Reverse Read:

DDDCGA…

H = A,C, or T

D = A,G, or T

Y = C or T

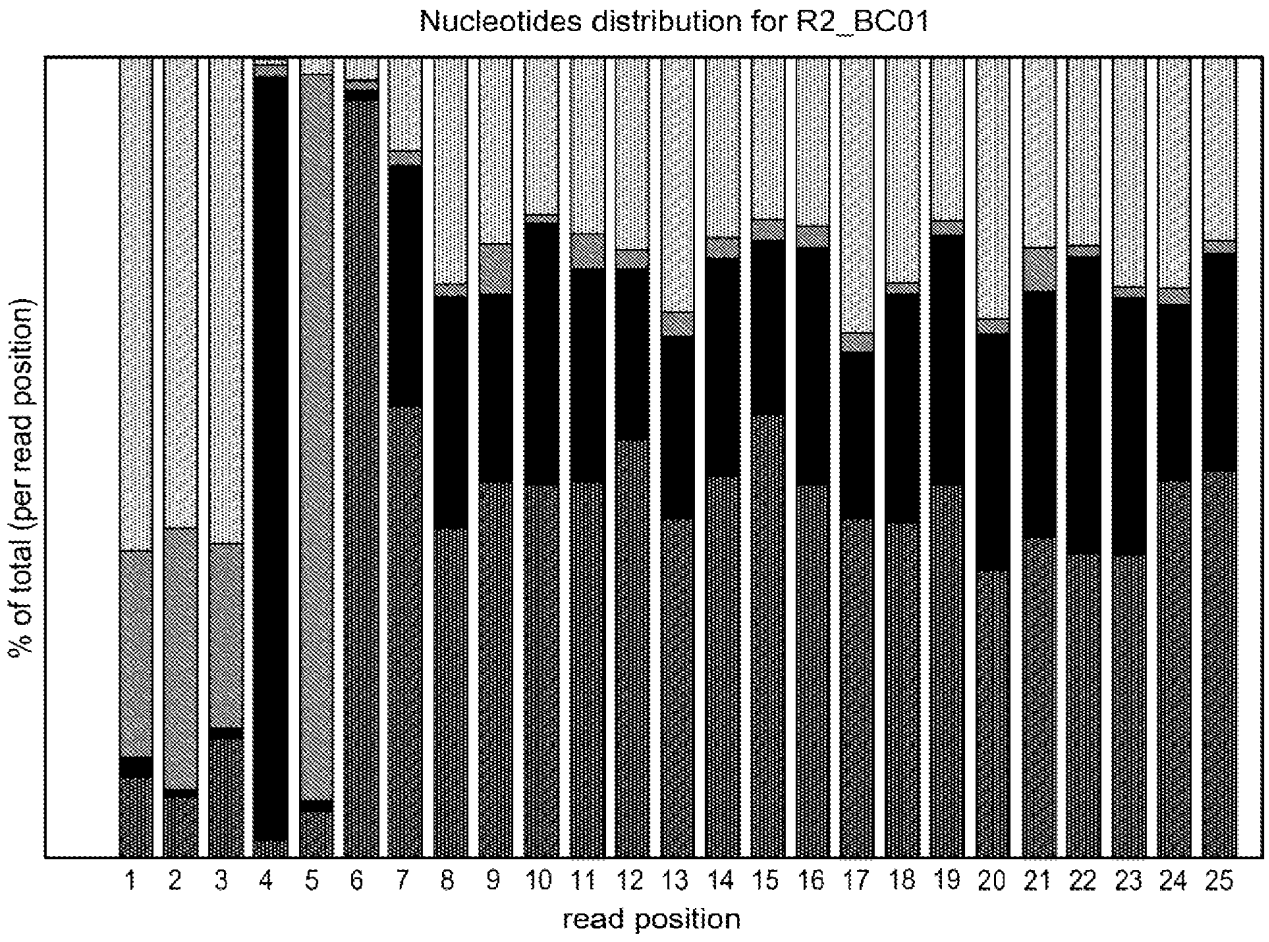


FIG. 5A

DR Rev Helper Style Ed3 BC

R01447 Ben384 D2 2

Expected Forward Read:

DDYGG...

H = A,C, or T

D = A,G, or T

Y = C or T

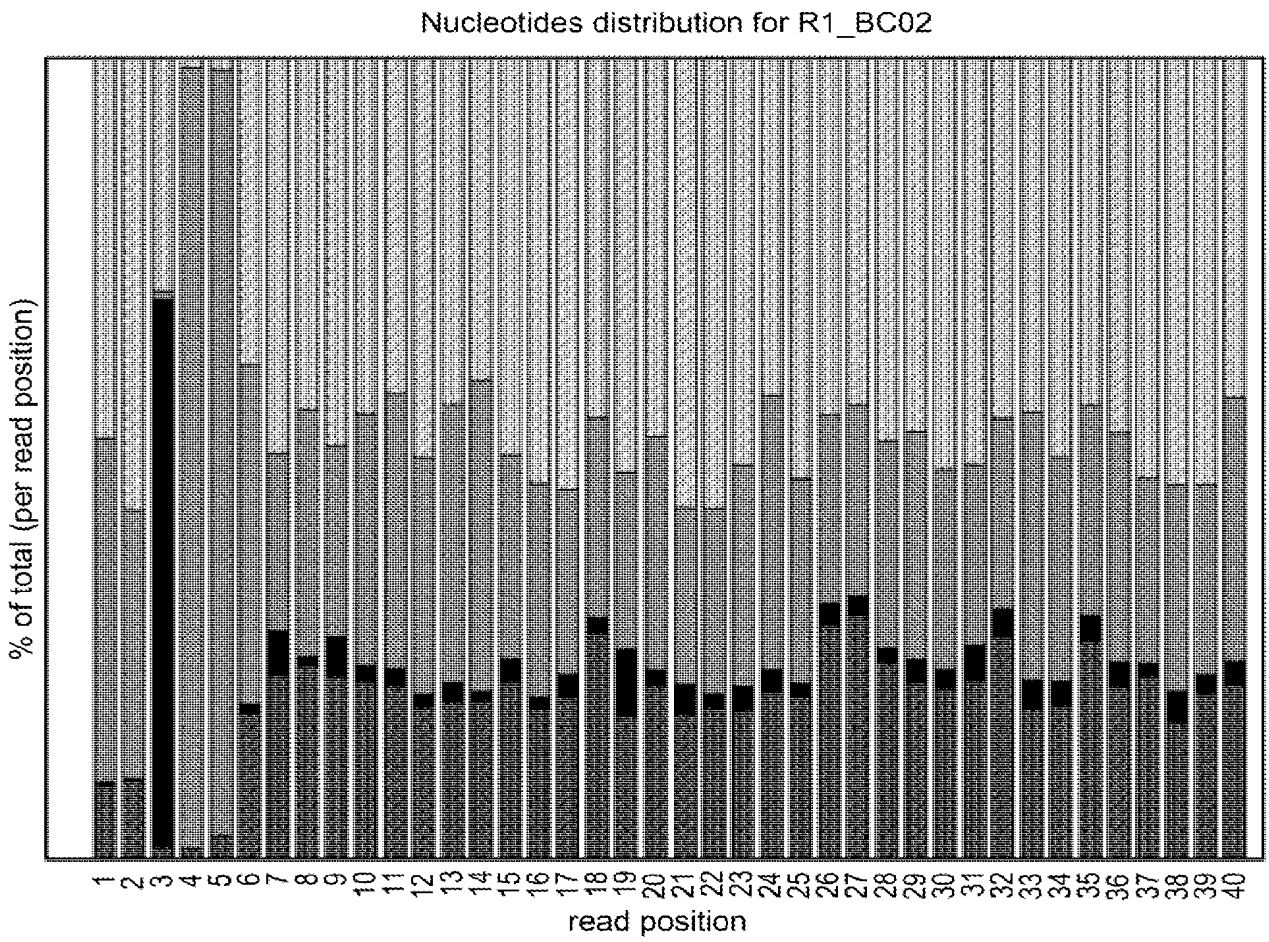


FIG. 5B

DR Rev Helper Style Ed3 BC

R01447 Ben384 D2 2

Expected Reverse Read:

DDCGA...

H = A,C, or T

D = A,G, or T

Y = C or T

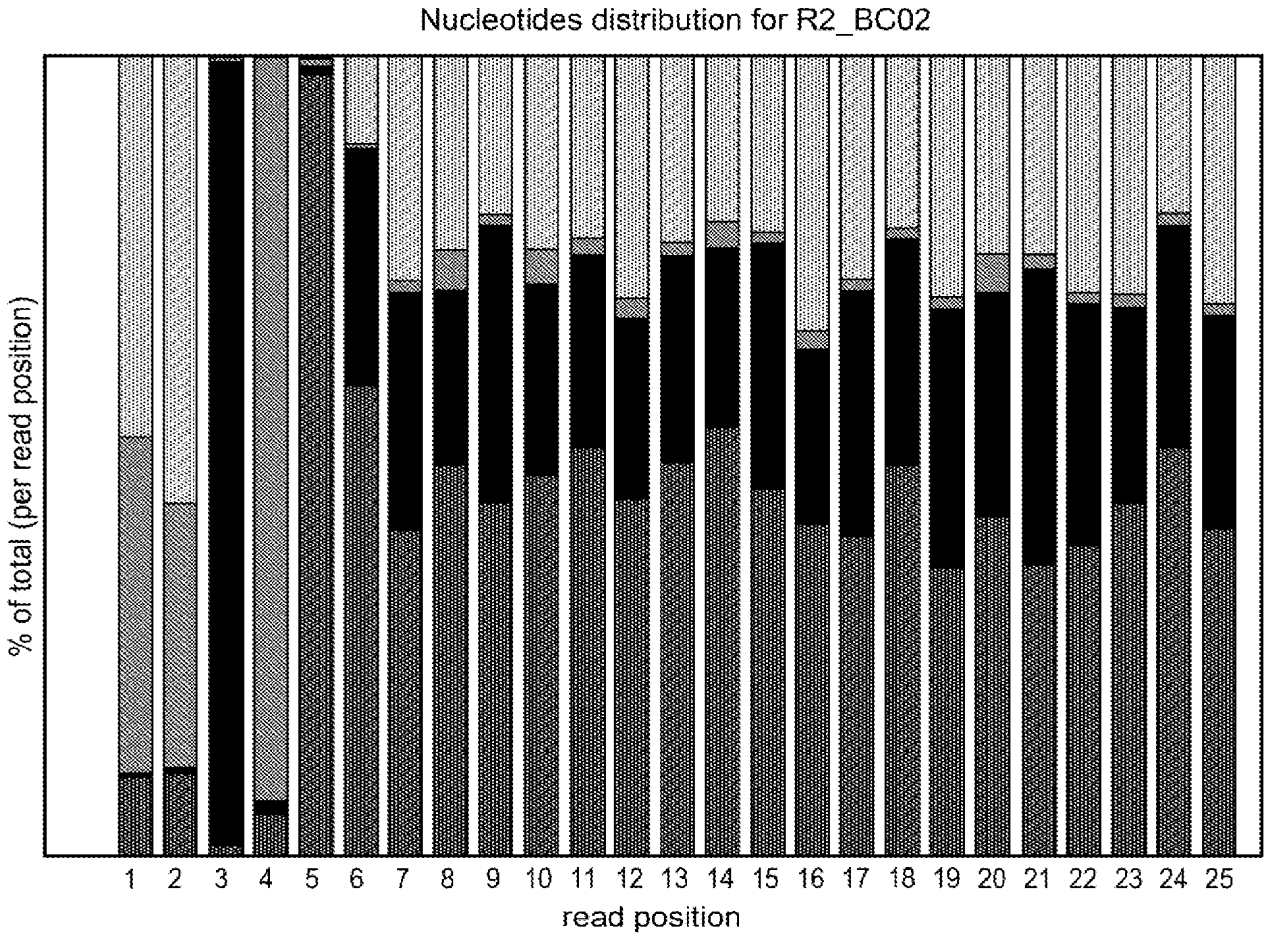


FIG. 6A

DR Rev Helper Style Ed3 BC

R01448 Ben385 D1 3

Expected Forward Read:

DYGG...

H = A,C, or T

D = A,G, or T

Y = C or T

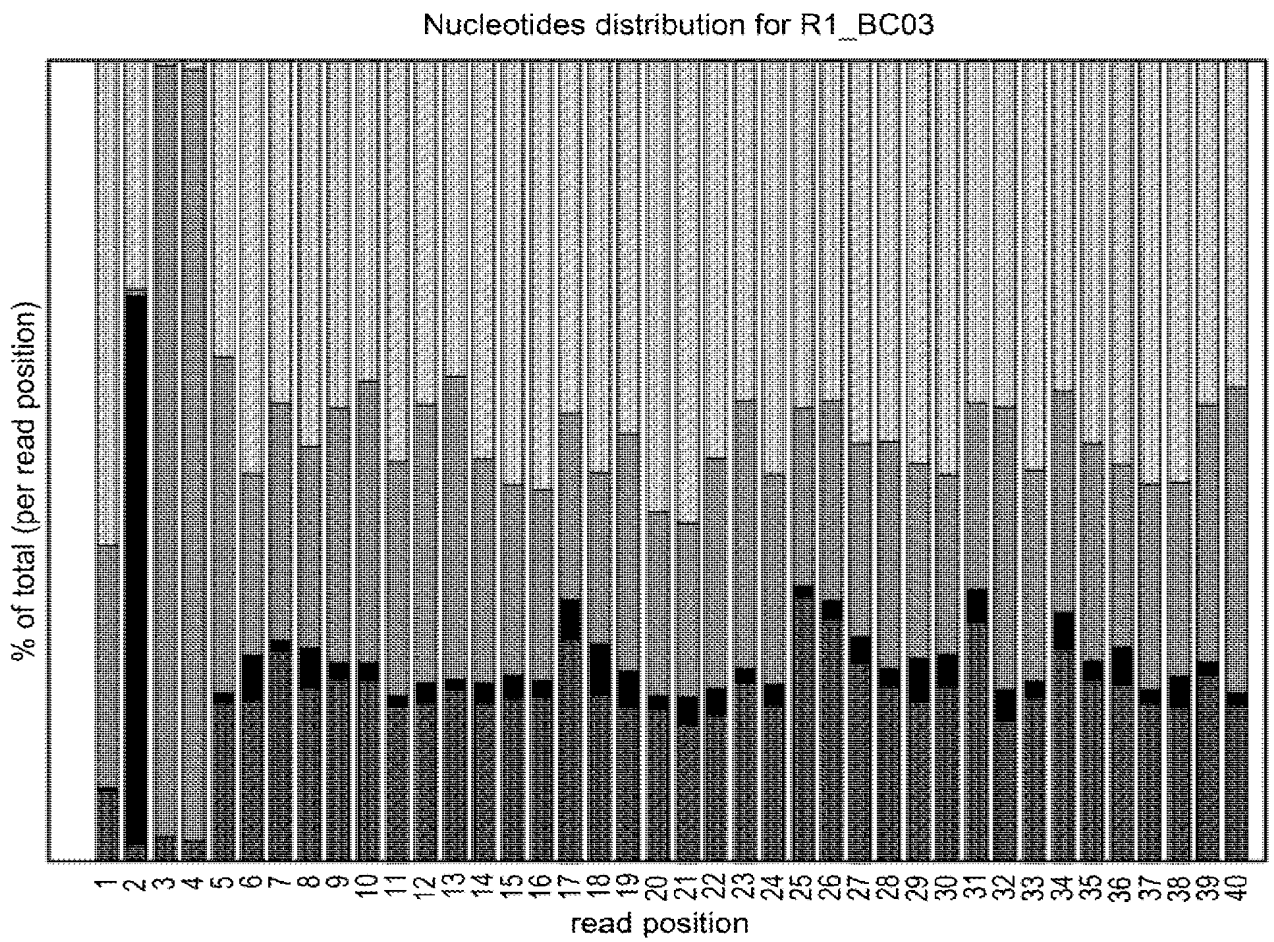


FIG. 6B

DR Rev Helper Style Ed3 BC

R01448 Ben385 D1 3

Expected Reverse Read:

DCGA...

H = A,C, or T

D = A,G, or T

Y = C or T

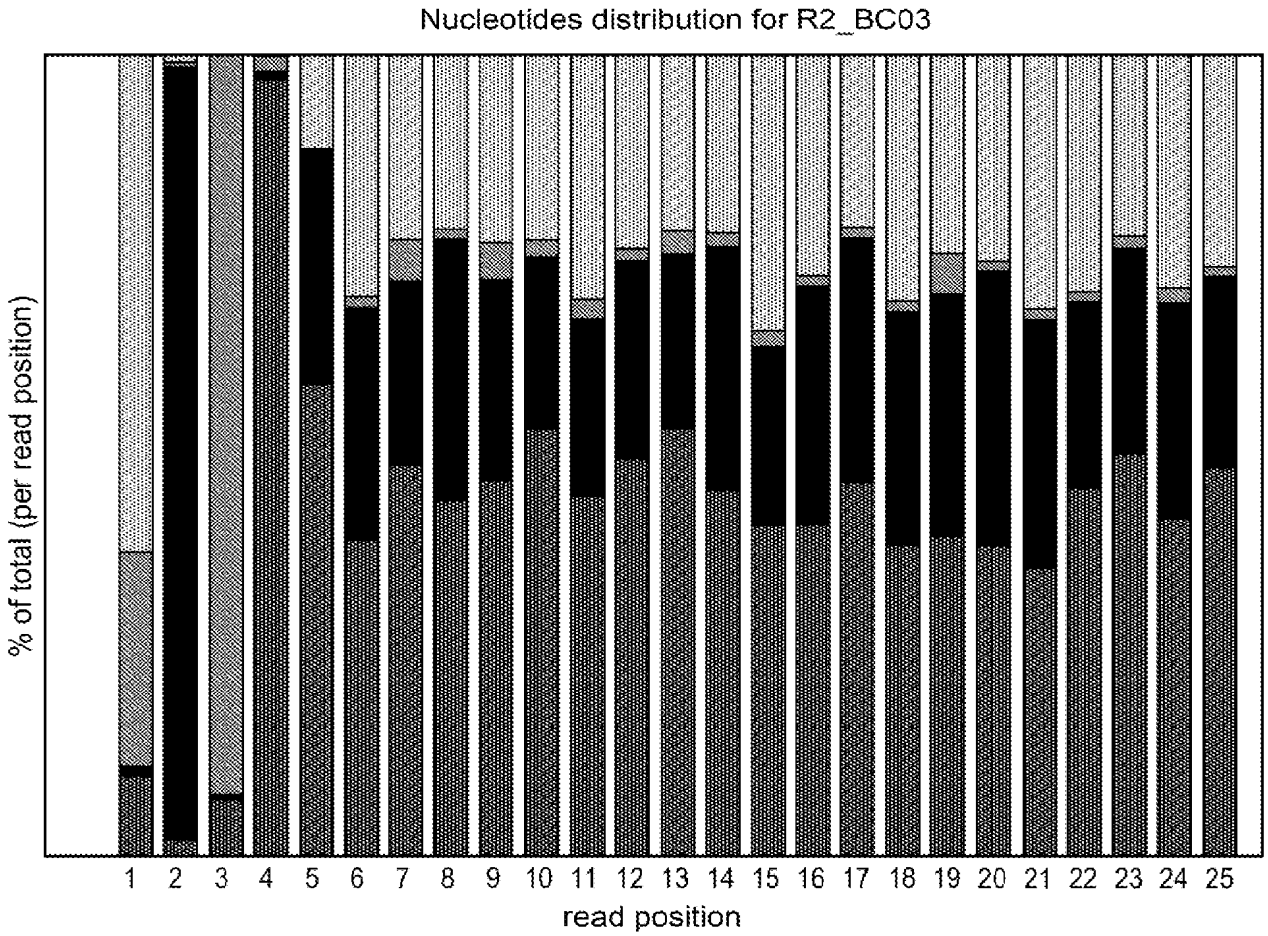


FIG. 7A

DR Rev Helper Style Ed3 BC

R01449 Ben341 D0 1

Expected Forward Read:

YGG...

H = A,C, or T

D = A,G, or T

Y = C or T

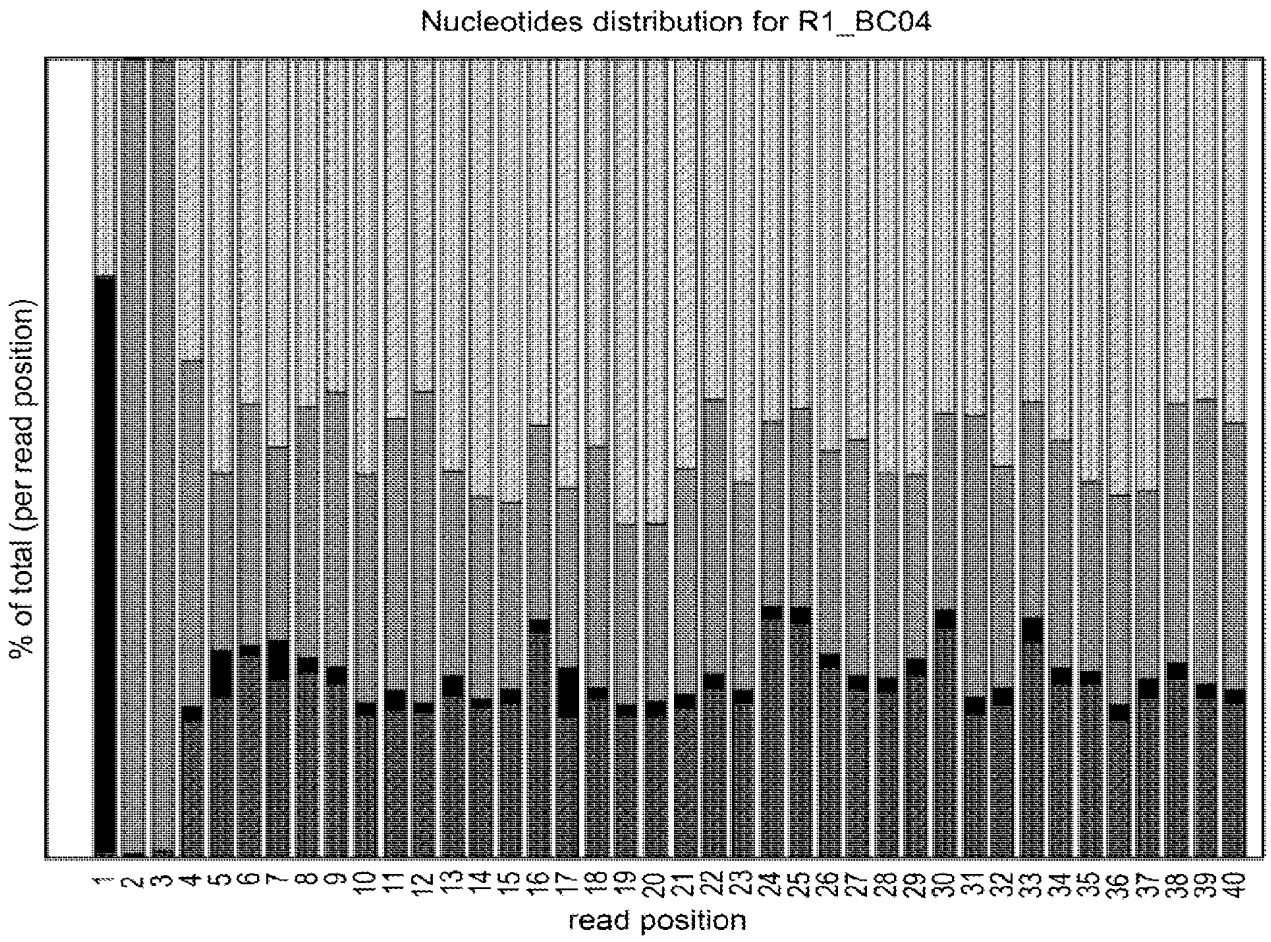


FIG. 7B

DR Rev Helper Style Ed3 BC

R01449 Ben341 D0 4

Expected Reverse Read:

CGA...

H = A,C, or T

D = A,G, or T

Y = C or T

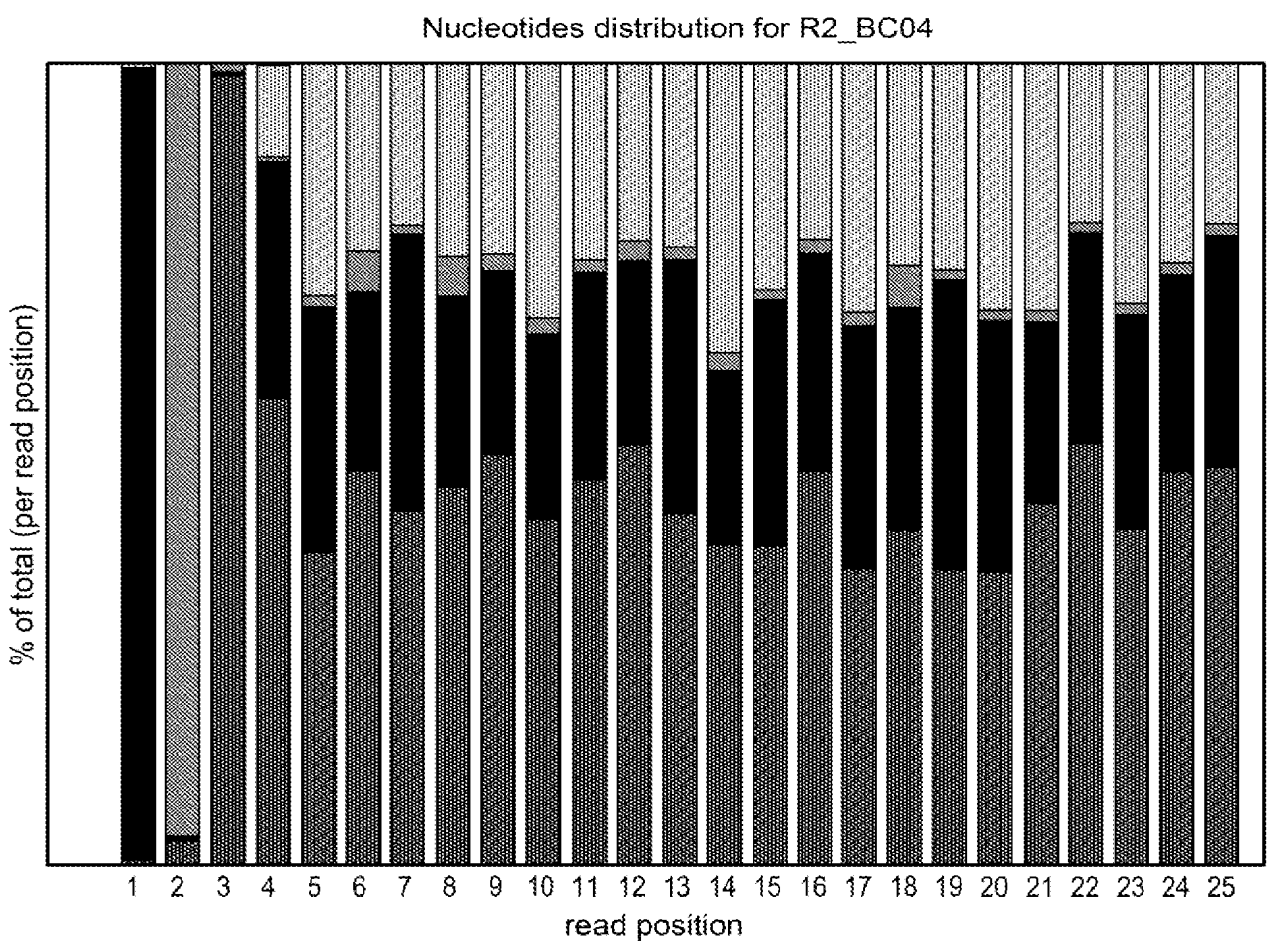


FIG. 8

DR Rev Helper Style Ed3 BC
R01450 mix4 MIX 5

H = A,C, or T
D = A,G, or T
Y = C or T
B = C,G, or T

Expected Forward Read
sum of:
YGG...
DYGG...
DDYGG...
DDDYGG...
NNNNNN...

Expected Reverse Read
sum of:
CGA...
DCGA...
DDCGA...
DDDCGA...
NNNNNN...

Expected Forward Read
sum of:
YGG...
DYGG...
DDYGG... (if don't have DDDYGG)
NNBNN... (missing A at position 3)

FIG. 9A

DR Rev Helper Style Ed3 BC

R01450 mix4 MIX 5

Expected Forward Read:

NNNNNN...

H = A,C, or T

D = A,G, or T

Y = C or T

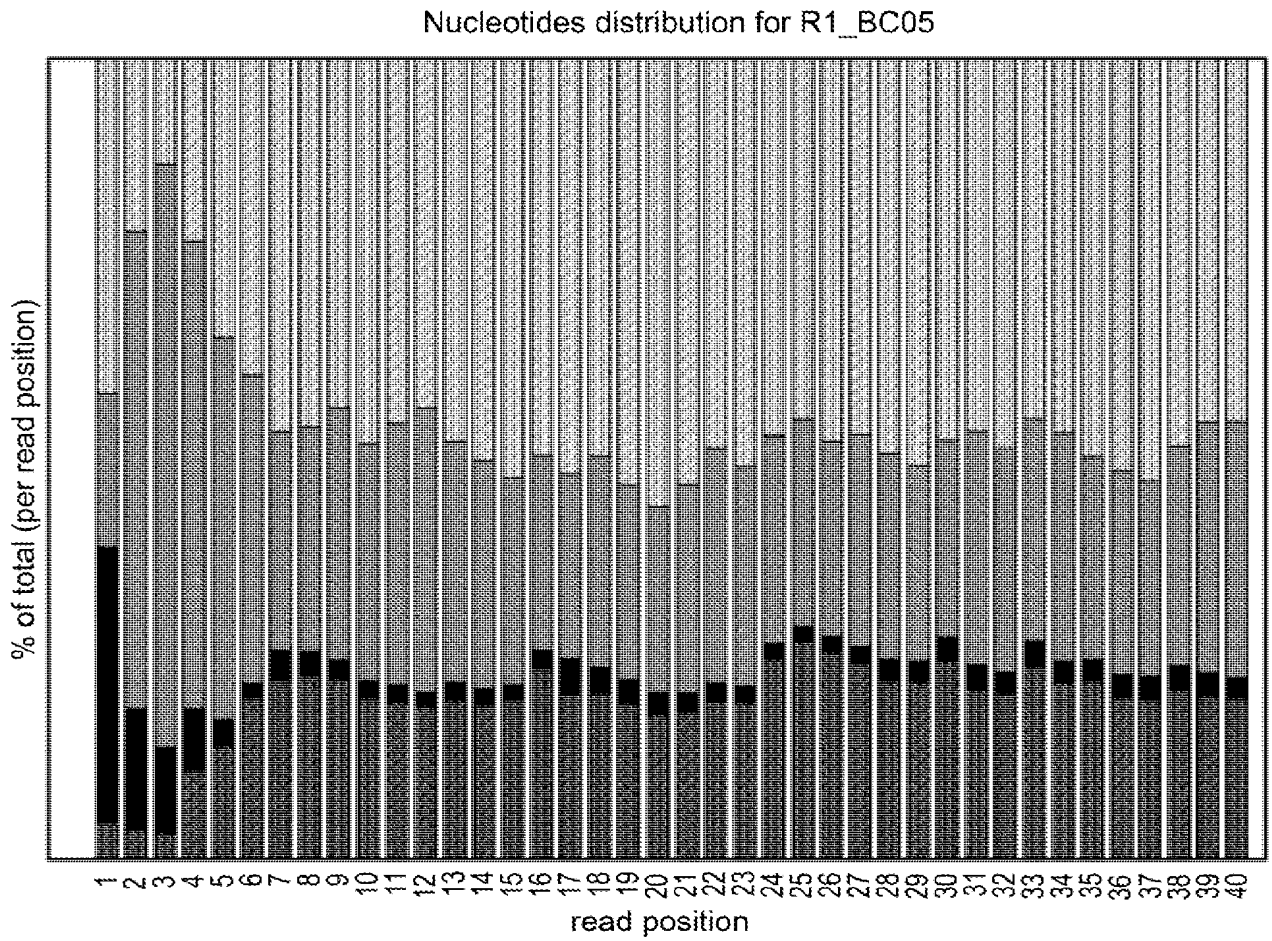


FIG. 9B

DR Rev Helper Style Ed3 BC

R01450 MIX4 5

Expected Reverse Read

NNNNNN...

H = A,C, or T

D = A,G, or T

Y = C or T

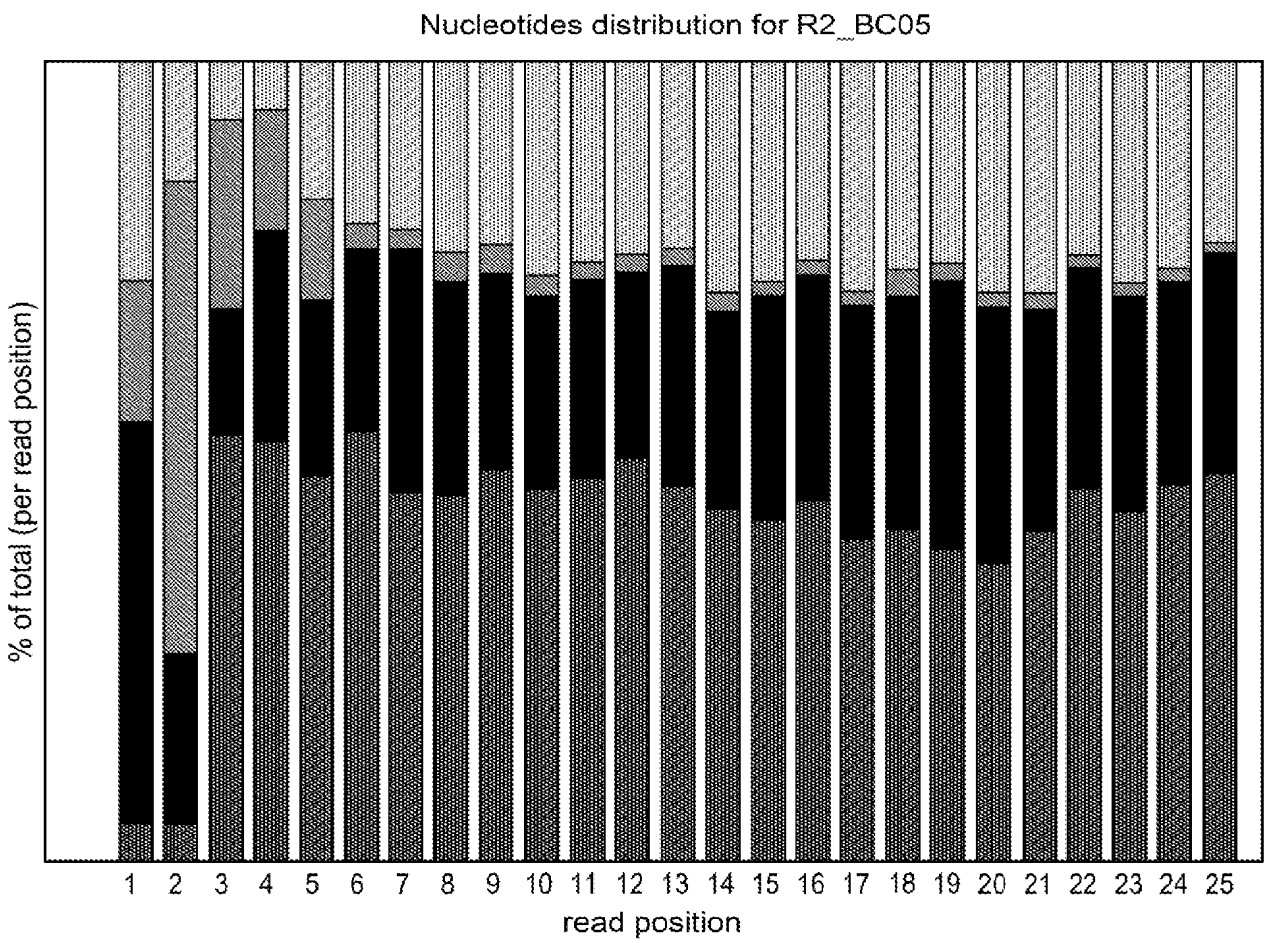


FIG. 10

40+6+25 run

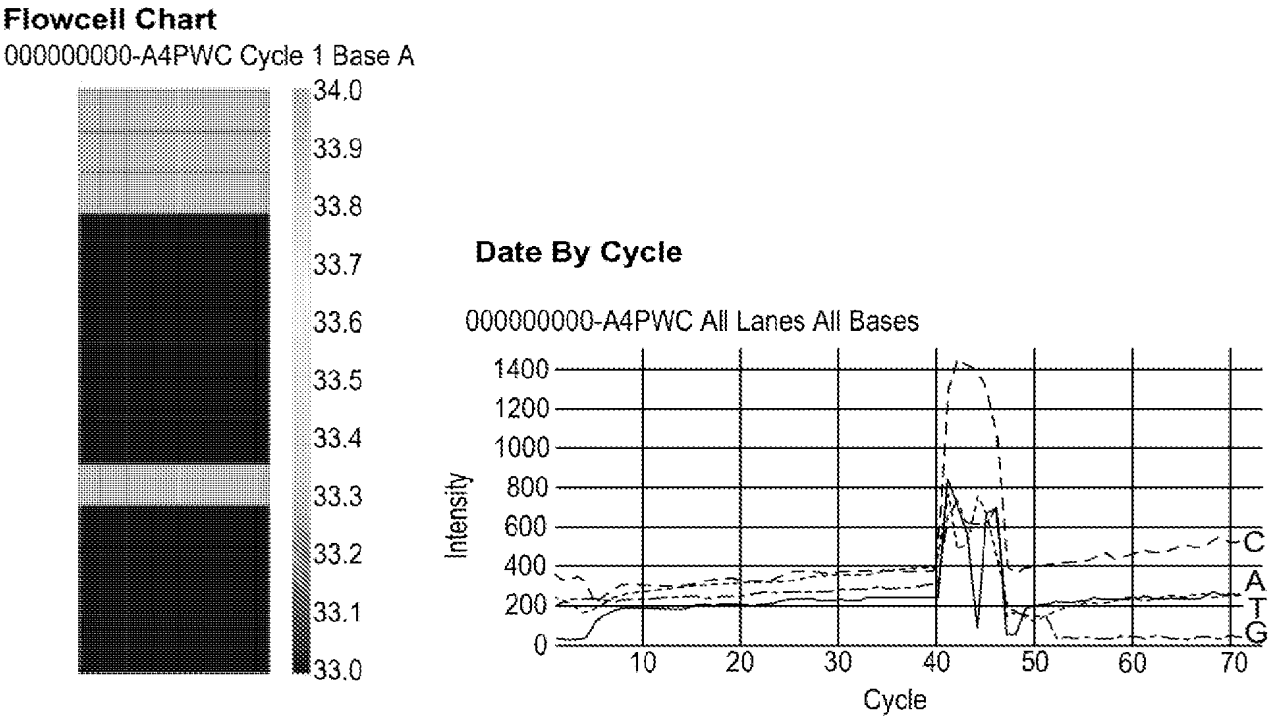
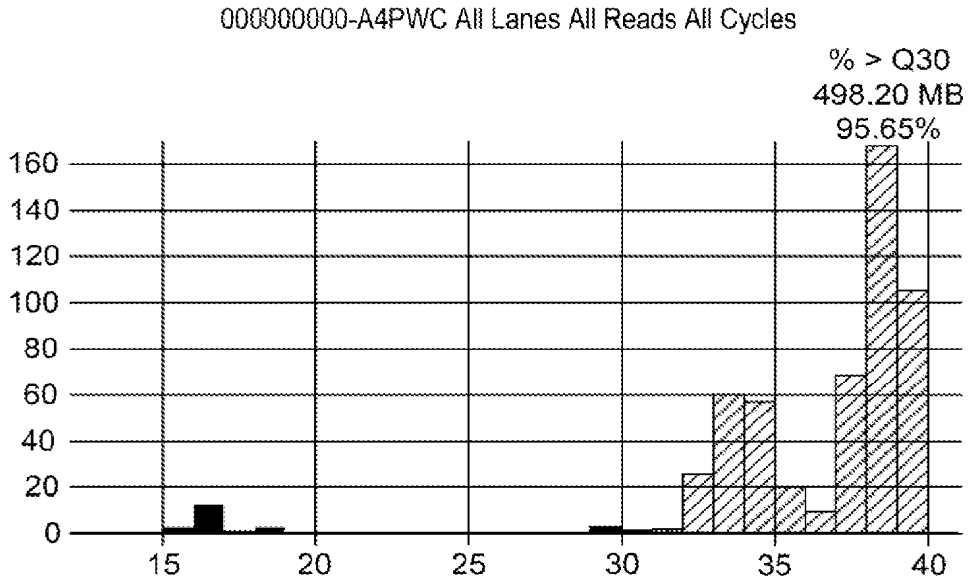


FIG. 10
(Continued)

QScore Distribution



QScore Heatmap

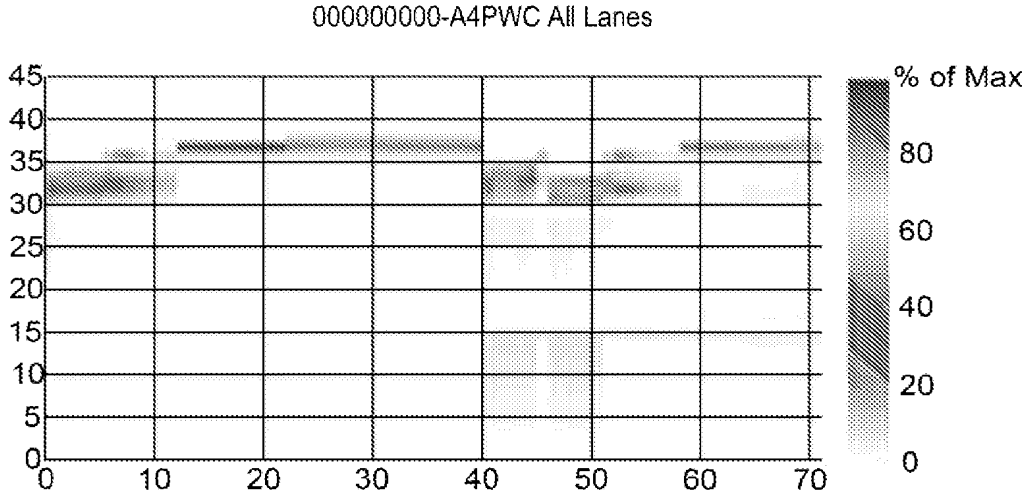


FIG. 11

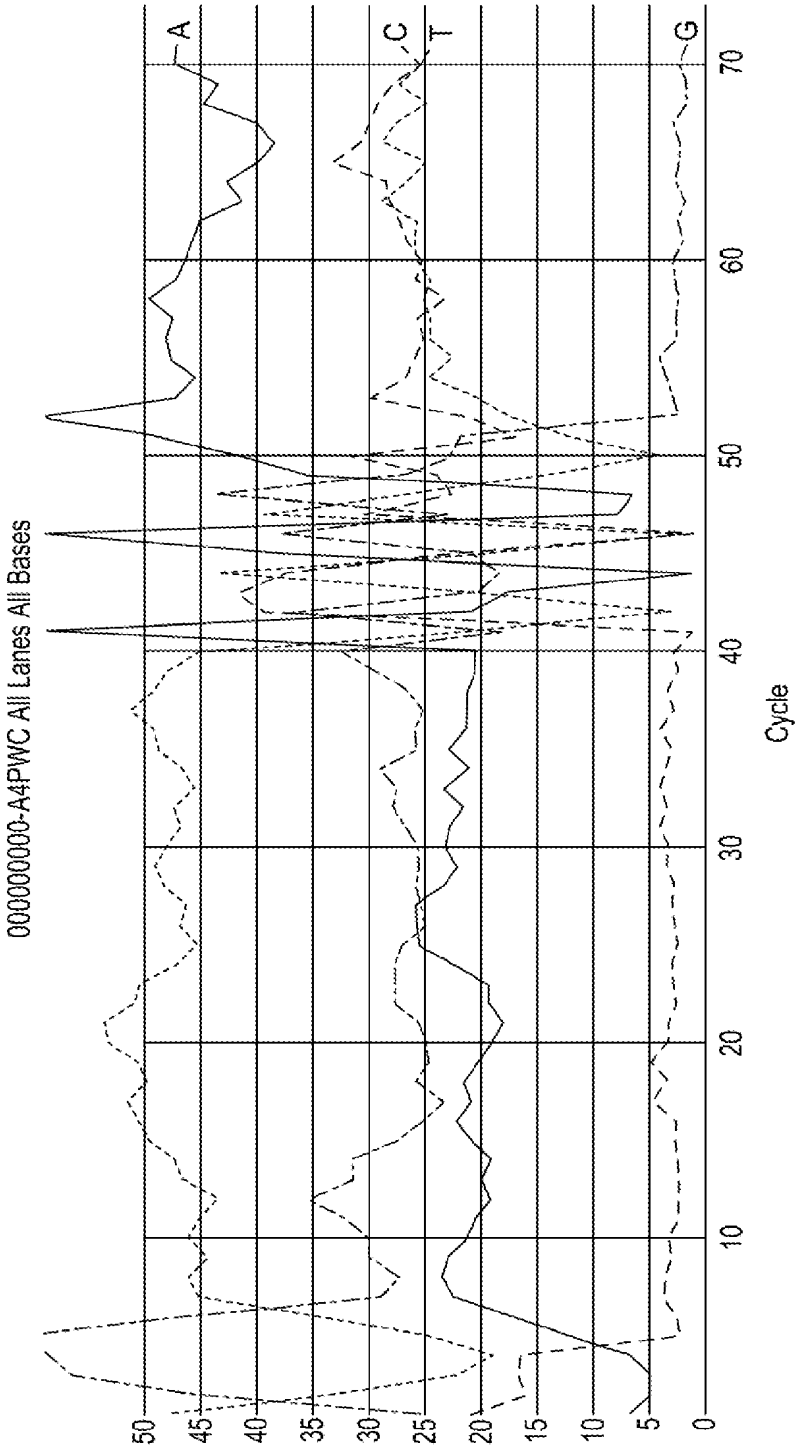

000000000-A4PWC All Lanes All Bases
A
C
T
G
50
45
40
35
30
25
20
15
10
5
0
10
20
30
40
50
60
70
Cycle

FIG. 12

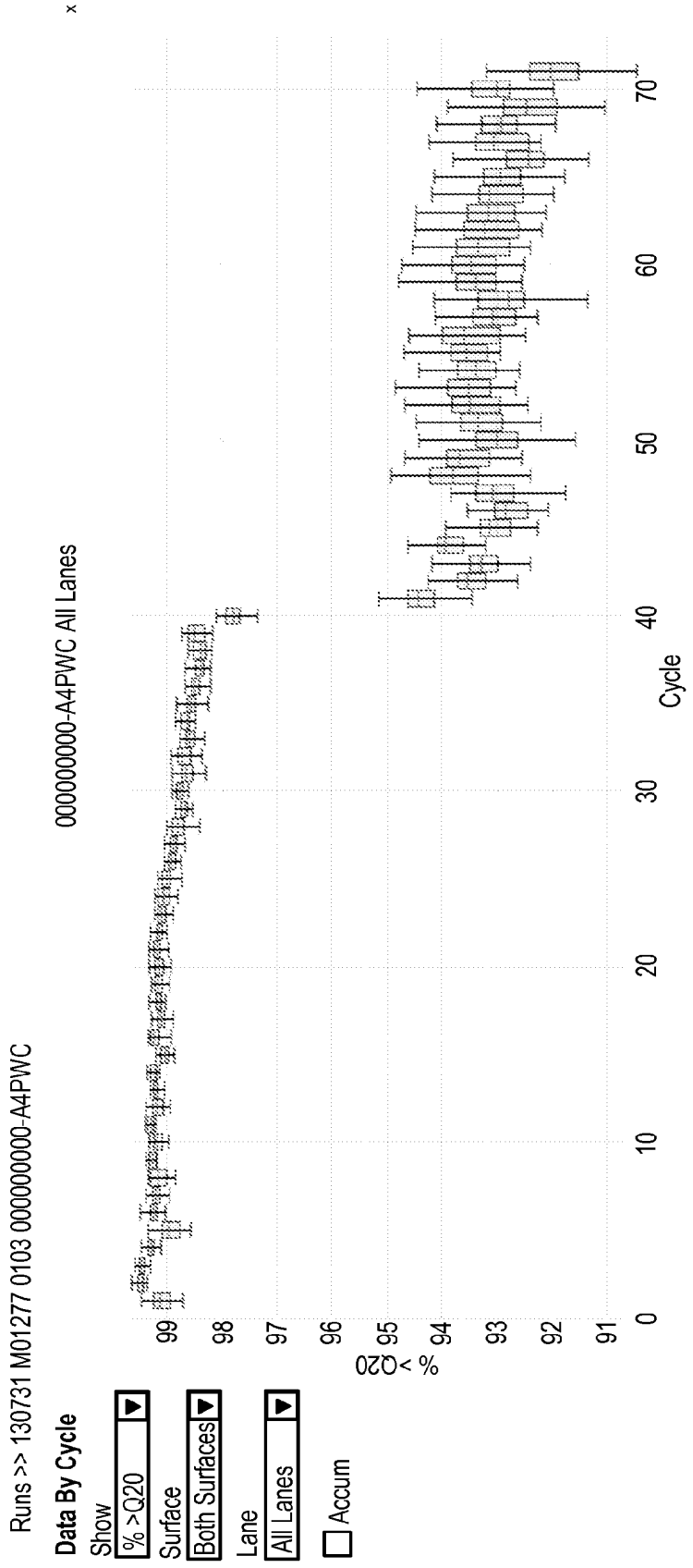

Runs >> 130731 M01277 0103 000000000-A4PWC
Data By Cycle
Show
% >Q20
Surface
Both Surfaces
Lane
All Lanes
Accum
000000000-A4PWC All Lanes
% >Q20
99
98
97
96
95
94
93
92
91
0
10
20
30
40
50
60
70
Cycle

FIG. 13

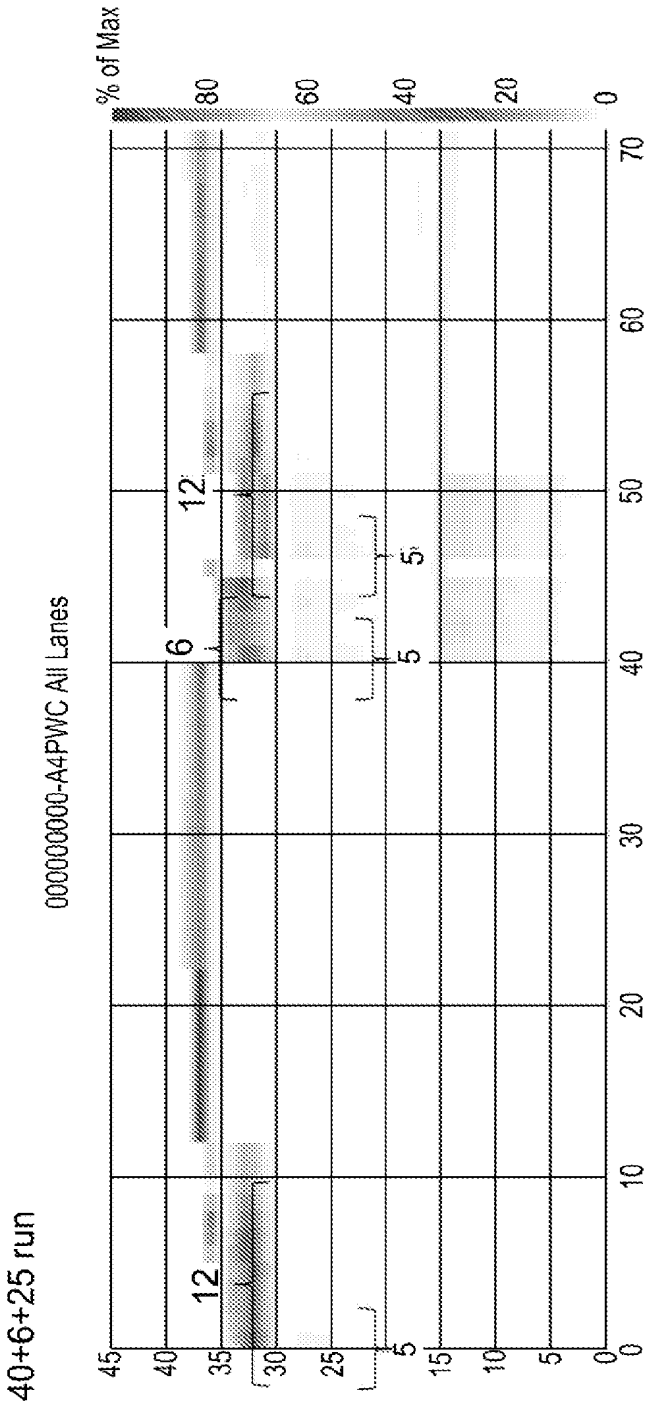

40+6+25 run
000000000-A4PWC All Lanes
% of Max
0
20
40
60
80
0
5
10
15
20
25
30
35
40
45
0
10
20
30
40
50
60
70
12
6
5
12
5
5

FIG. 14

40+6+25 run

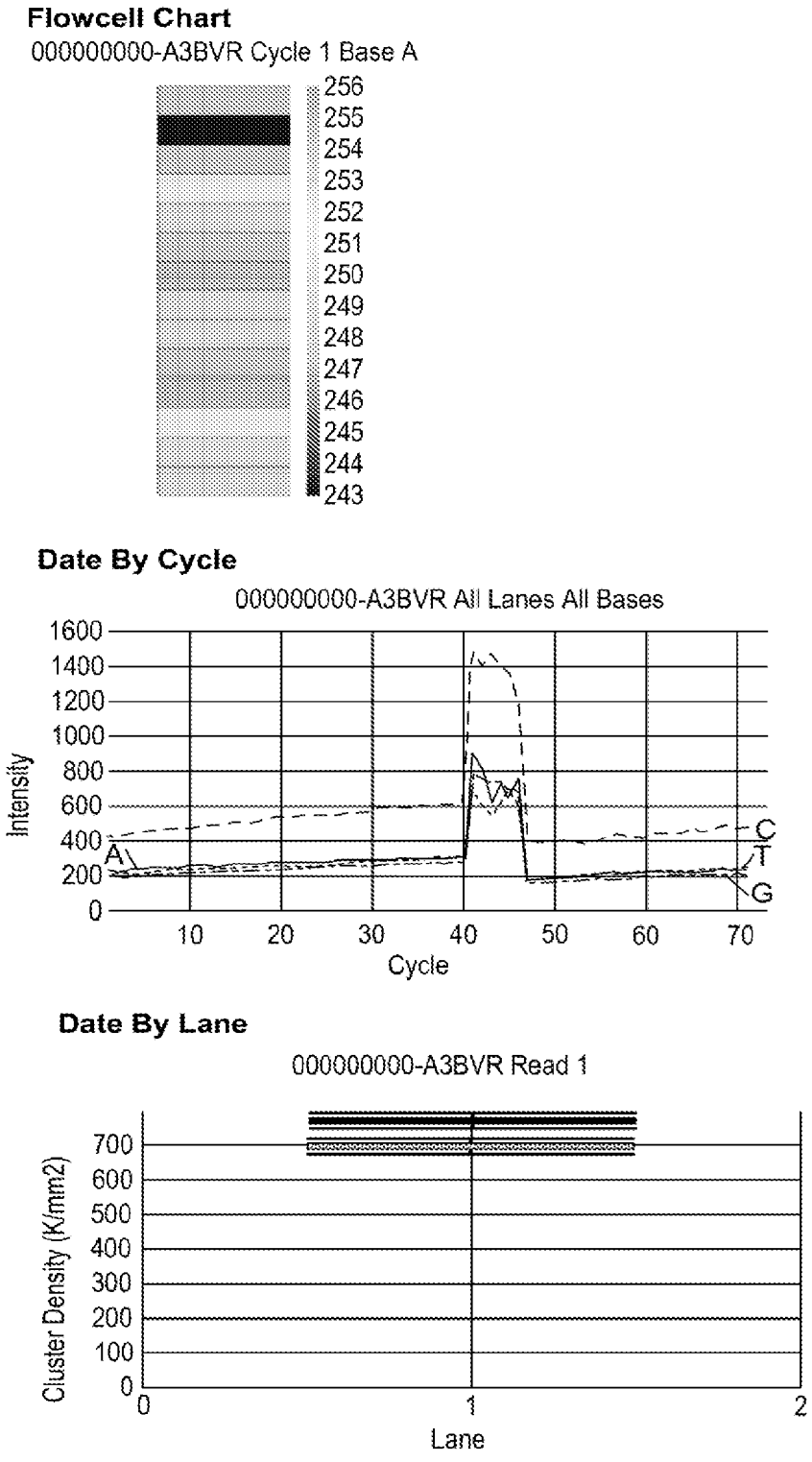
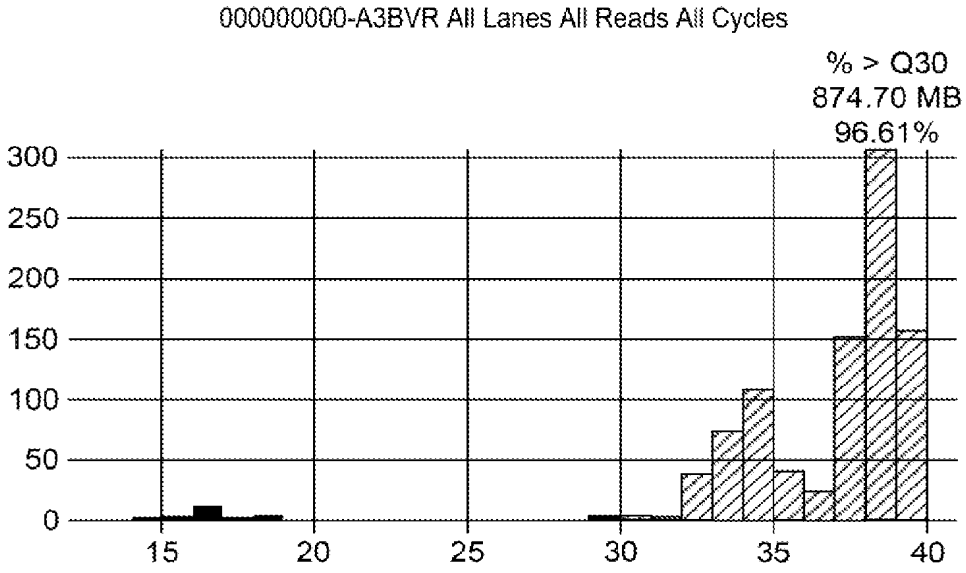


FIG. 14
(Continued)

QScore Distribution



QScore Heatmap

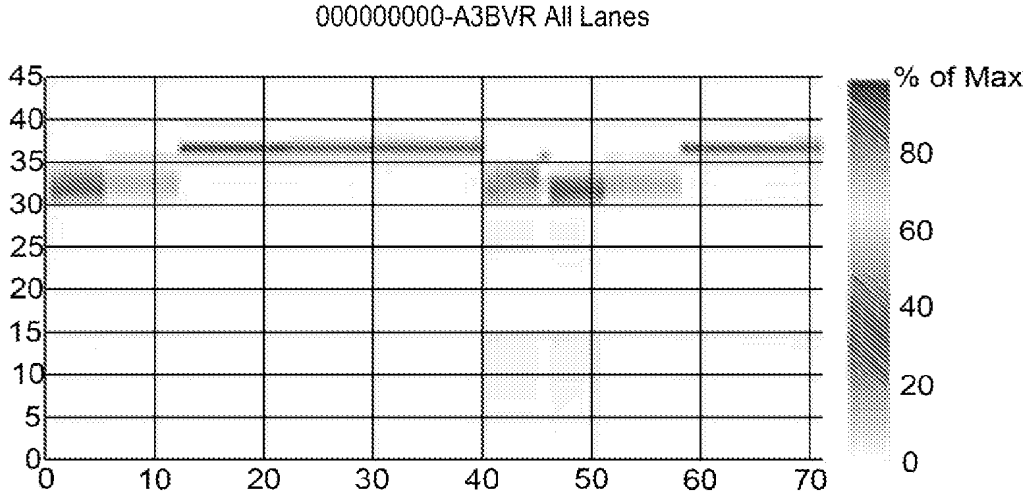


FIG. 15

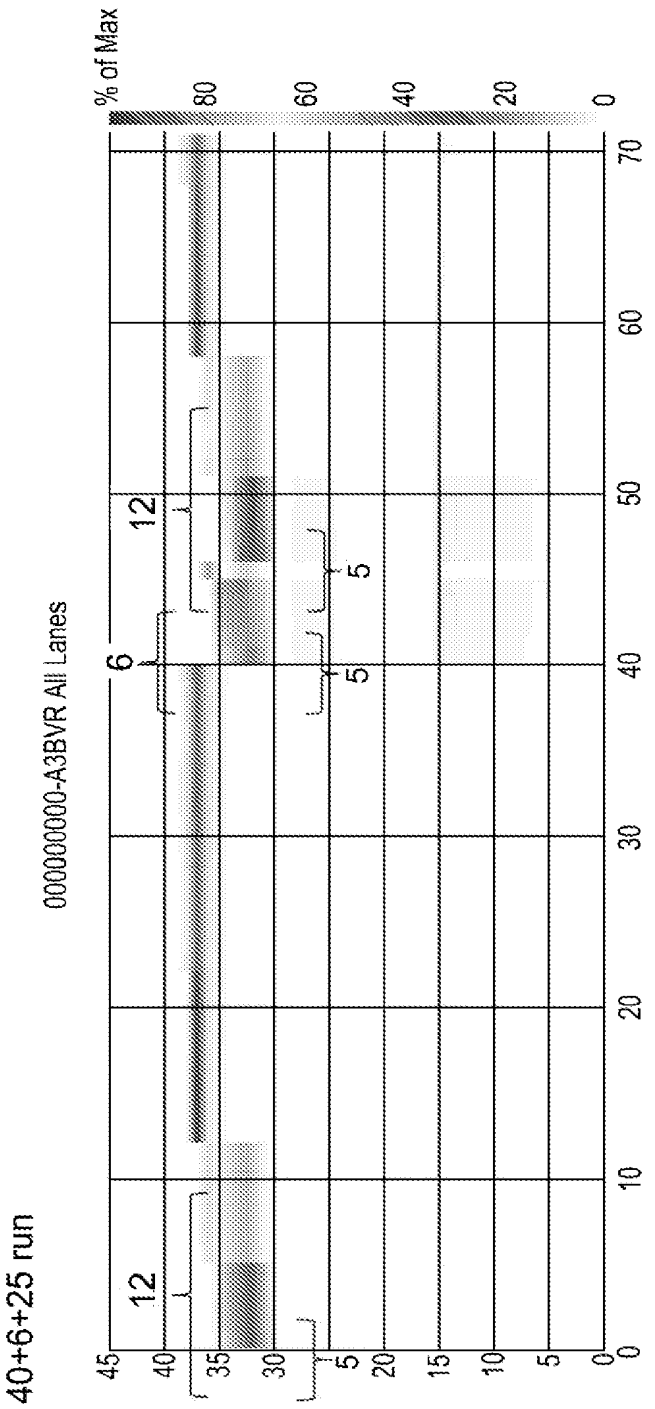

40+6+25 run
00000000-A3BVR All Lanes
% of Max
12
6
12
5
5
45
40
35
30
5
20
15
10
5
0
0
10
20
30
40
50
60
70
80
60
40
20
0

FIG. 16

Runs >> 130731 MC1277 0103 00000000-A4PWC

Run Summary

Level	Cycles	Yield Total	Projected Total Yield	Yield Perfect	Yield <= 3 errors	Aligned (%)	% Perfect [Non Usable Cycles]	% <= 3 errors [Num Usable Cycles]	Error Rate (%)	Intensity Cycle 1	% Intensity Cycle 20	% >= Q30
Read 1	40	300.0 M	300.0 M	0.0	0.0	0.00	0.0 [39]	0.0 [39]	0.00	31	632.9	98.5
Read 2 (I)	6	38.5 M	38.5 M	0.0	0.0	0.00	0.0 [5]	0.0 [5]	0.00	850	0.0	90.9
Read 3	25	184.6 M	184.6 M	0.0	0.0	0.00	0.0 [24]	0.0 [24]	0.00	53	447.1	92.2
Total	71	523.1 M	523.1 M	0.0	0.0	0.00	0.0	0.0	0.00	311	540.0	95.7

Lane	Tiles	Density (K/ mm2)	Cluster PF (%)	Phas/ Prephas (%)	Reads (M)	Reads PF (M)	% > Q30	Yield	Cycles Err Rated	Aligned (%)	Error Rate (%)	Error Rate 35 cycle (%)	Error Rate 75 cycle (%)	Error Rate 100 cycle (%)	Intensity Cycle 1	% Intensity Cycle 20
Read 1																
1	28	446 +/- 7	91.95 +/- 0.41	0.202 / 0.227	8.36	7.69	98.5	300.0 M	0	0.0 +/- 0.0	0.00 +/- 0.00	0.00 +/- 0.00	0.00 +/- 0.00	0.00 +/- 0.00	31 +/- 2	632.9 +/- 24.3
Read 2 (I)																
1	28	446 +/- 7	91.95 +/- 0.41	0.653 / 0.129	8.36	7.69	90.9	38.5 M	0	0.0 +/- 0.0	0.00 +/- 0.00	0.00 +/- 0.00	0.00 +/- 0.00	0.00 +/- 0.00	850 +/- 77	0.0 +/- 0.0
Read 3																
1	28	446 +/- 7	91.95 +/- 0.41	0.259 / 0.245	8.36	7.69	92.2	184.6 M	0	0.0 +/- 0.0	0.00 +/- 0.00	0.00 +/- 0.00	0.00 +/- 0.00	0.00 +/- 0.00	53 +/- 4	447 +/- 22.6

IDT Exhibit 1017.1697

FIG. 17

Runs >> 130314 MO1277 0048 00000000-A3BVR

Run Summary

Level	Cycles	Yield Total	Projected Total Yield	Yield Perfect	Yield <= 3 errors	Aligned (%)	% Perfect [Non Usable Cycles]	% <= 3 errors [Num Usable Cycles]	Error Rate (%)	Intensity Cycle 1	% Intensity Cycle 20	% >= Q30
Read 1	40	521.5 M	521.5 M	2.7 M	2.9 M	0.56	92.8 [39]	99.9 [39]	0.21	228	117.6	97.9
Read 2 (I)	6	66.9 M	66.9 M	0.0	0.0	0.00	0.0 [5]	0.0 [5]	0.00	894	0.0	93.9
Read 3	25	320.9 M	320.9 M	1.7 M	1.8 M	0.55	95.0 [24]	100.0 [24]	0.21	185	119.3	95.2
Total	71	909.3 M	909.3 M	4.4 M	4.7 M	0.55	83.9	100.0	0.21	436	118.4	96.6

Lane	Tiles	Density (K/ mm2)	Cluster PF (%)	Phas/ Prephas (%)	Reads (M)	Reads PF (M)	% > Q30	Yield	Cycles Err Rated	Aligned (%)	Error Rate (%)	Error Rate 35 cycle (%)	Error Rate 75 cycle (%)	Error Rate 100 cycle (%)	Intensity Cycle 1	% Intensity Cycle 20
Read 1																
1	28	770 +/- 14	91.02 +/- 0.27	0.118 / 0.249	14.69	13.37	97.9	521.5 M	39	0.6 +/- 0.0	0.21 +/- 0.02	0.20 +/- 0.02	0.00 +/- 0.0	0.00 +/- 0.0	228 +/- 22	117.6 +/- 2.6
Read 2 (I)																
1	28	770 +/- 14	91.02 +/- 0.27	0.000 / 0.000	14.69	13.37	93.9	66.9 M	0	0.0 +/- 0.0	0.00 +/- 0.0	0.00 +/- 0.0	0.00 +/- 0.0	0.00 +/- 0.0	894 +/- 88	0.0 +/- 0.0
Read 3																
1	28	770 +/- 14	91.02 +/- 0.27	0.126 / 0.266	14.69	13.37	95.2	320.9 M	24	0.6 +/- 0.0	0.21 +/- 0.02	0.00 +/- 0.0	0.00 +/- 0.0	0.00 +/- 0.0	185 +/- 17	119.3 +/- 3.1

IDT Exhibit 1017.1698

FIG. 18A

Adaptors at each end can be different lengths.

H = A,C, or T
D = A,G, or T
Y = C or T

FC-FWD/MetSeq 5′-…GAtGTGTGtTtTTttGATtTDDDYGG —— insert —— tCGHHHAGATCGGAAGAGCACACACGTC…TTG-3′
MetSeq 5′…GGAGTTTAGATGTGTGTTTTTTTGATTT

Forward Read
Within the 6nt at 5′ end, look for YGG and trim the 0-3 D's before it:
YGG…insert (what if YGG not found?)
DYGG…insert (what to do with non-D bases?)
DDYGG…insert
DDDYGG…insert (but in this case, DDD can = TGG, so how to tell if adaptor or real insert?)

Within the 6nt at 3′ end (after adaptor trimming is done), look for TCG and trim it and the 0-3 H's after it:
insert…TCGHHH
insert…TCGHH
insert…TCGH
insert…TCG (what to do with non-H bases?)
It's OK if TCG not found, the MspI fragment can be larger than the read length.
If TCG not found, look for TC at 3′ end and remove it, and if TC not found, look for T at 3′ end and remove it.

properly trimmed Forward read:
5′ YGG…insert…

C Methylated

IDT Exhibit 1017.1699

FIG. 18B

Adaptors at each end can be different lengths.

H = A,C, or T
D = A,G, or T
Y = C or T
R = A or G

FC-FWD/MetSeq 5′-…GAtGTGTGtTtTTttGATtTDDDYGG —— insert —— t***C***GHHHAGAT***C***GGAAGAG***C***A***C***A***C***GT***C***…TTG-3′

TS Rev 5′,_GACGTGTGCTCTTCCGATCT

Reverse Read
Within the 6nt at 5′ end, look for CGA and trim the 0-3 D's and the CGA (since this contains no genomic info other than it's an MspI site) :
CGA…insert
DCGA…insert
DDCGA…insert
DDDCGA…insert

Within the 6nt at 3′ end (after adaptor trimming is done), look for CCR and trim the 0-3 H's after it:
insert…CCRHHH
insert…CCRHH
insert…CCRH
insert…CCR
It's OK if CCR is not found, the MspI fragment can be larger than the read length.

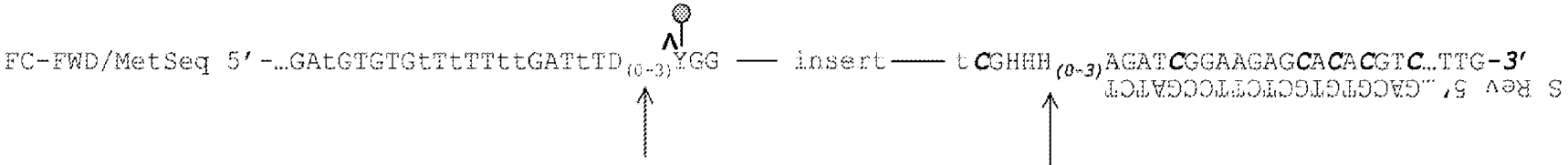
C Methylated

properly trimmed Reverse read:
5′ insert…(may or may not end in CCR, depending on the MspI fragment size and readlength)

FIG. 18C

Adaptors at each end can be different lengths.

C Methylated



i) Number of D's removed from 5' end of Forward read
can be the same as the number of H's removed from 3' end of Reverse read (if the read spans across the entire MspI fragment)

ii) Number of H's removed from 3' end of Forward read (if the read spans across the entire MspI fragment)
can be the same as the number of D's removed from 3' end of Reverse read

iii) Insert can align with respect to MspI sites in genome as shown:

(part shown in BOLD may or may not be present in the read depending on the MspI fragment length and read length)

```
Top strand fragment case
genome reference:              5′ …CCGG… - insert - CCGG…
properly trimmed Forward read: 5′ YGG… - insert -
properly trimmed Reverse read: 3′ RCC… - insert -

Bottom strand fragment case
genome reference:              5′ …CCGG… - insert - CCGG…
properly trimmed Forward read:           - insert - …GGY 5′
properly trimmed Reverse read:           - insert - …CCR 3′
```

IDT Exhibit 1017.1701

FIG. 19

D0 = 1
D1 = 3
D2 = 3x3 = 9
D3 = 3x3x3 = 27

total diversity = 40

FIG. 20

Sample	ng	Sequences analysed in total:	Mapping efficiency:	No alignments:	Did not map uniquely:	CT/CT:	CT/GA:	mCG	mCHG	mCHH	Total CpGs analysed
NA19238	400	28,467,191	59.58%	7.52%	32.91%	8,464,806	8,495,775	28.6%	0.5%	0.3%	**47,208,201**
NA19238	100	27,239,653	58.42%	6.94%	34.64%	7,944,860	7,968,954	32.7%	0.5%	0.3%	**42,071,026**
NA19238	25	26,732,815	58.28%	6.81%	34.92%	7,774,534	7,804,757	34.9%	0.5%	0.4%	**39,632,176**
NA19238	6.25	28,820,811	57.86%	7.09%	35.05%	8,324,948	8,351,549	34.9%	0.6%	0.4%	**42,654,227**
NA19238	1.56	25,889,539	57.46%	7.46%	35.08%	7,425,771	7,451,201	34.5%	0.6%	0.5%	**38,504,765**
IMR90	100	27,431,639	59.94%	6.59%	33.47%	8,213,675	8,228,851	30.5%	0.5%	0.3%	**44,343,781**
IMR90	100	29,671,768	59.59%	6.75%	33.67%	8,833,296	8,846,892	31.0%	0.5%	0.3%	**47,371,339**
WGBS	50	21,577,677	75.66%	6.5%	17.8%	8,161,875	8,163,270	76.7%	1.2%	1.2%	**5,495,435**

FIG. 21

	Sample	ng	Ave CpG per Read Total	Ave CpG per Uniquely Aligned Read
RRBS	NA19238	400	1.66	2.78
	NA19238	100	1.54	2.64
	NA19238	25	1.48	2.54
	NA19238	6.25	1.48	2.56
	NA19238	1.56	1.49	2.59
	IMR90	100	1.62	2.70
	IMR90	100	1.60	2.68
	WGBS	50	0.25	0.34

FIG. 22

		CpG Islands			CpG Shores		Promoters	
Sample	ng	1x depth	5x depth	20x depth	1x depth	5x depth	1x depth	5x depth	20x depth
NA19238	400	1,504,475	1,266,783	667,786	656,096	461,332	170,336	1,664,663	1,324,874	660,439
NA19238	100	1,476,388	1,260,019	432,222	659,484	482,788	110,350	1,647,854	1,334,295	430,005
NA19238	25	1,450,314	1,236,153	306,773	654,881	486,022	89,159	1,624,751	1,318,058	312,095
NA19238	6.25	1,436,633	1,228,801	399,082	644,323	479,246	125,823	1,603,990	1,306,739	408,945
NA19238	1.56	1,374,221	1,149,018	345,608	596,378	436,296	107,778	1,516,399	1,215,205	353,982
IMR90	100	1,463,177	1,257,887	552,607	653,059	472,682	145,786	1,625,626	1,323,880	546,676
IMR90	100	1,469,576	1,272,816	619,104	658,688	483,909	170,471	1,636,068	1,343,959	615,871
insilico 40nt SE 40-300bp		1,213,017			507,061			1,305,452		
insilico 40nt SE all sizes		1,665,942			852,847			2,025,019		

FIG. 23

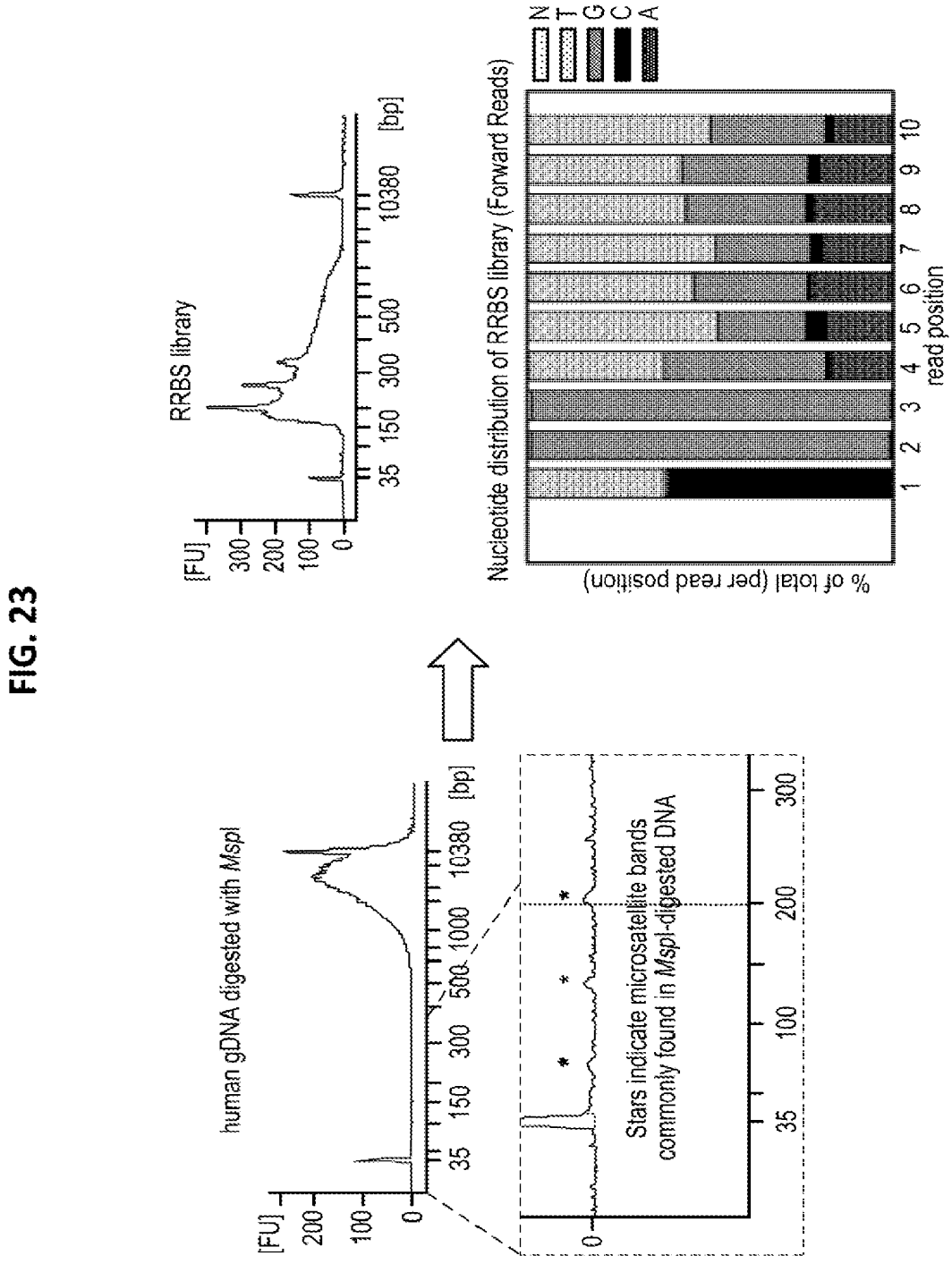

human gDNA digested with MspI
[FU]
[bp]
Stars indicate microsatellite bands commonly found in MspI-digested DNA
RRBS library
Nucleotide distribution of RRBS library (Forward Reads)
% of total (per read position)
read position
N
T
G
C
A

FIG. 24

TABLE 1 | Comparison of genome-wide DNA methylation profiling technologies[a].

	RRBS	MeDIP-seq	MethylCap-seq/MBD-seq	WGBS	Infinium 27K
Method for enriching CpGs	CCGG/MspI	Anti-5C antibody	5C-binding protein	N/A	Predefined set
Genomic DNA input	0.01–0.3 μg	0.3–5 μg	1–3 μg	5 μg	0.5–1 μg
Maximum resolution (bp)	1	>100	>100	1	1
Theoretical genomic coverage	10%	100%	100%	100%	0.1%
Actual coverage (≥10 reads[b])	9%	9%	14%	76%	0.1%
Illumina reads per sample	~10 million	~50 million	~50 million	>500 million[c]	Microarray based

RRBS is appropriate for analysis of small amounts of DNA, e.g., from FACS-purified cell populations or embryonic tissue; analysis of clinical samples in FFPE format or laser-microdissected tissue material; studies with a focus on CpG islands and gene promoters (while also sampling a sizable percentage of other genomic elements); studies investigating small effects (e.g., for non-cancer diseases or environmental effects) that require high sequencing depth and many samples to achieve sufficient sensitivity; and multicenter studies that require robustness and reproducibility across several labs. 5C, 5-methylcytosine. [a]Adapted from reference 31. [b]Quantitative measurement of DNA methylation requires a minimum number of overlapping reads. [c]Very deep sequencing is necessary to achieve the same read coverage per CpG as RRBS (which is the key determinant of method sensitivity and accuracy).

FIG. 25

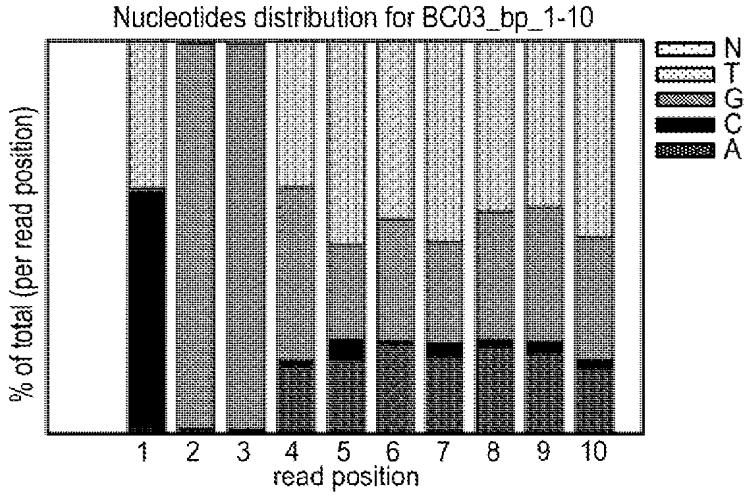
	Class	CpG island	CpG shores	Promoters	Enhancers	Coding Exons	Introns
Genome-wide	# regions in Class	27,718	45,494	82,145	1,339	377,744	404,405
	Average Size	764.1	2,225.6	2,334.9	1,522.2	165.6	6,422.8
	Total # CpG	2,089,538	2,287,717	4,148,299	22,160	1,951,024	26,571,158
	CpG per Kbp	97.3	20.9	20.3	11.4	28.7	14.0
	Stranded CpG	4,179,076	4,575,434	8,296,598	44,320	3,902,048	53,142,316
RRBS	100nt SE	2,187,671	1,003,815	2,431,738	3,860	984,468	5,022,090
	50nt SE	1,452,825	606,959	1,564,822	2,578	629,482	2,786,556
	40nt SE	1,213,017	507,061	1,305,452	2,270	528,990	2,277,500

RRBS analysis considers 40-300bp *Msp I* fragments and assumes 0% methyl-C and 100% bisulfite conversion.

1 CpG = 2 "Stranded CpG" measurements

FIG. 26

RRBS



DARK SEQUENCING

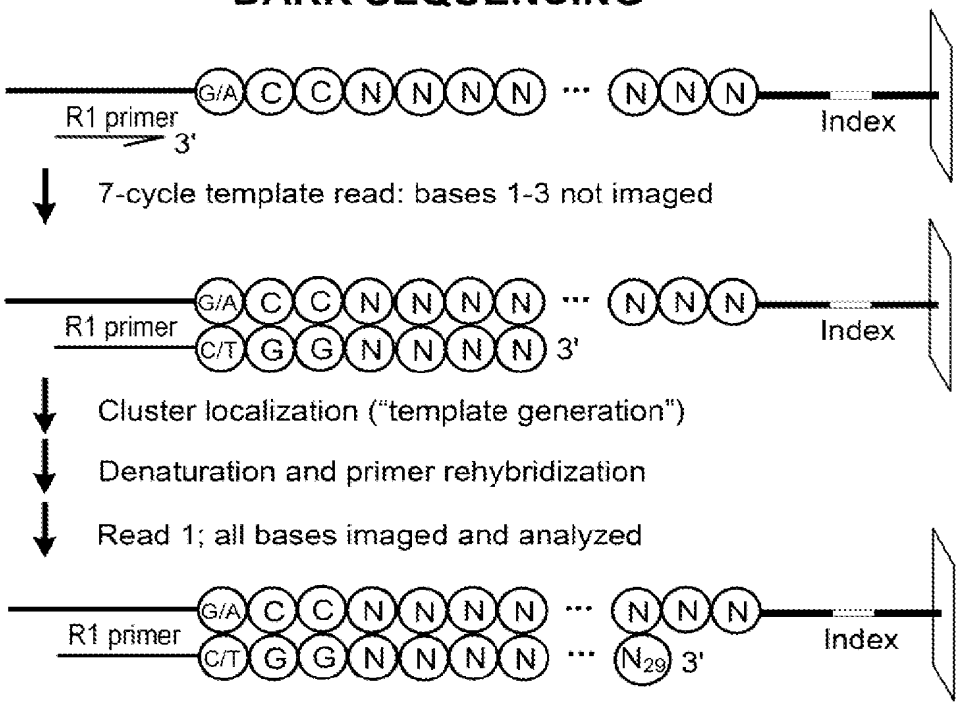


FIG. 27A

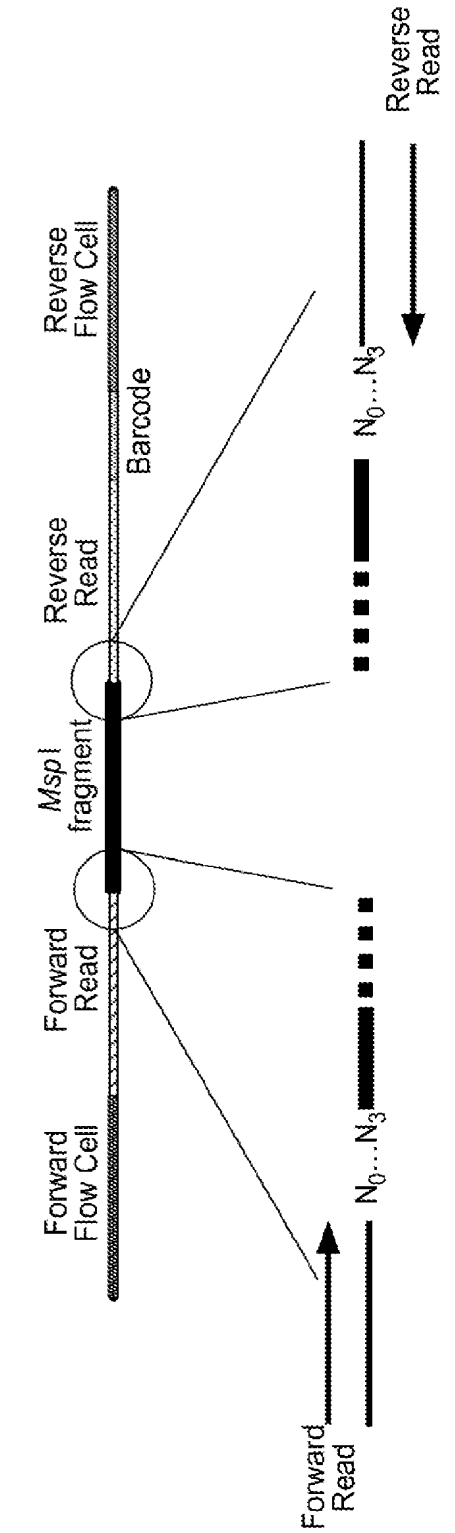


FIG. 27B

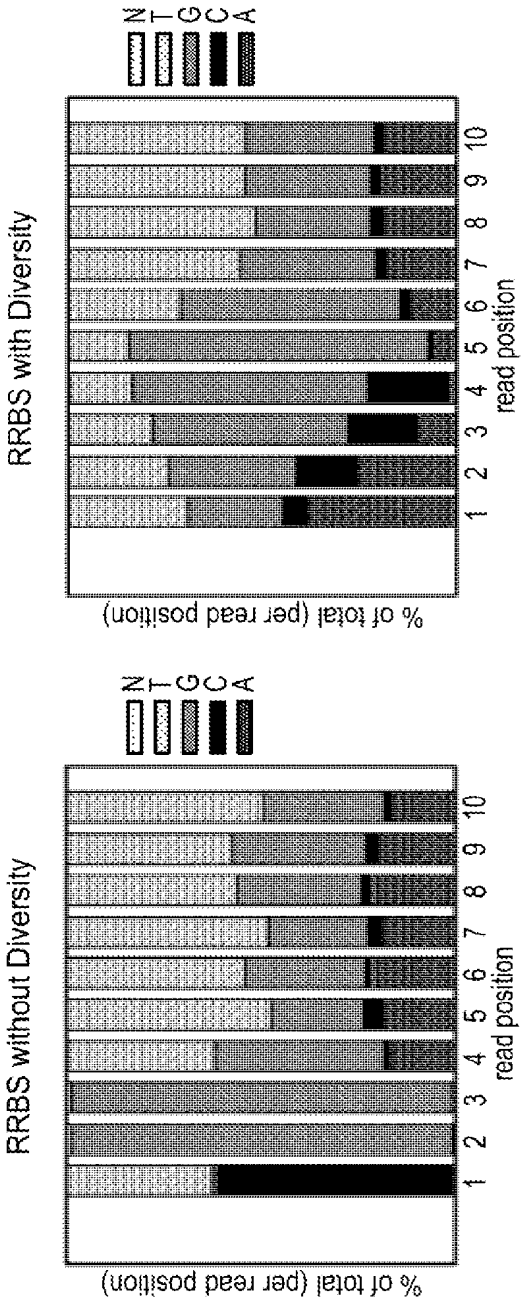
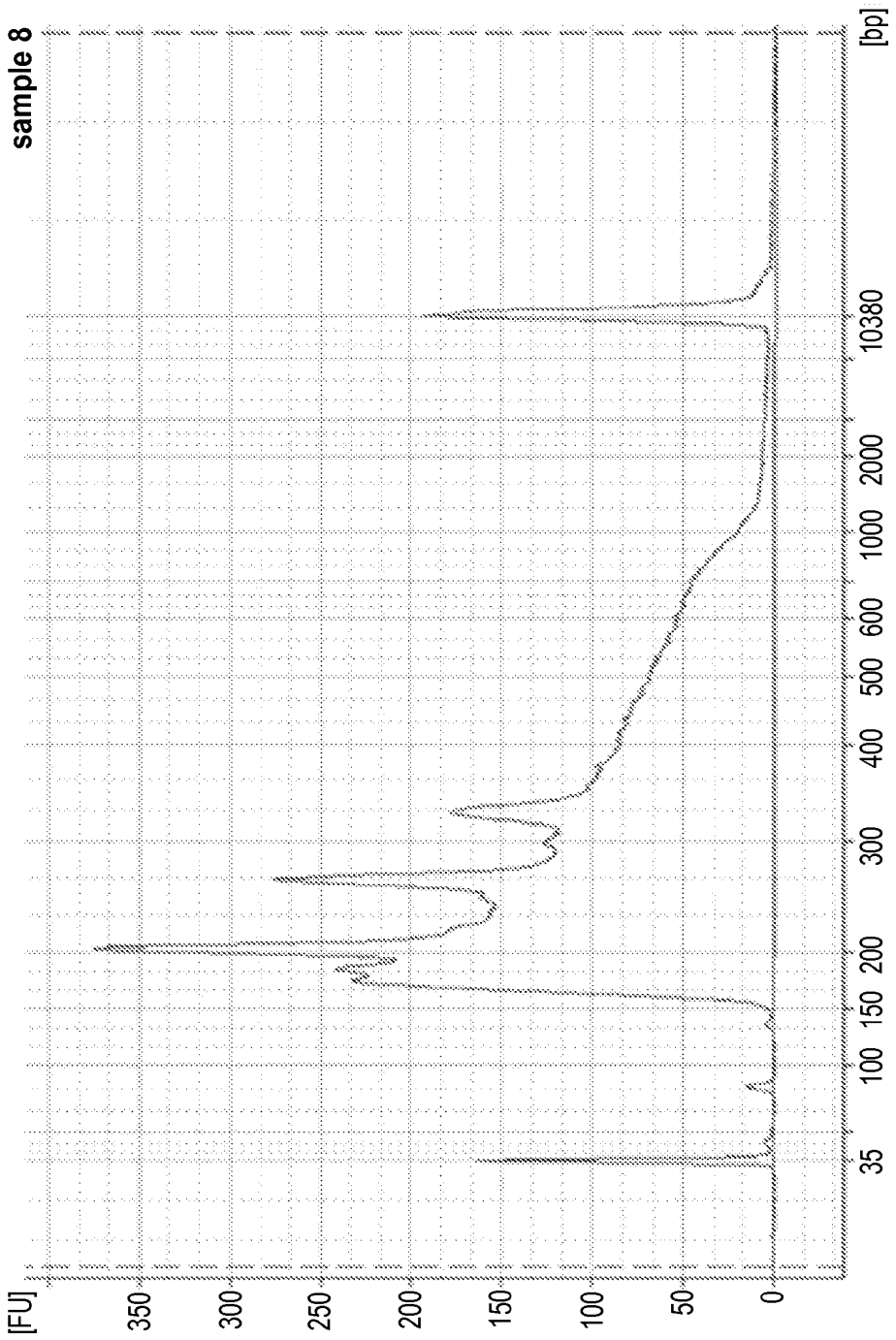
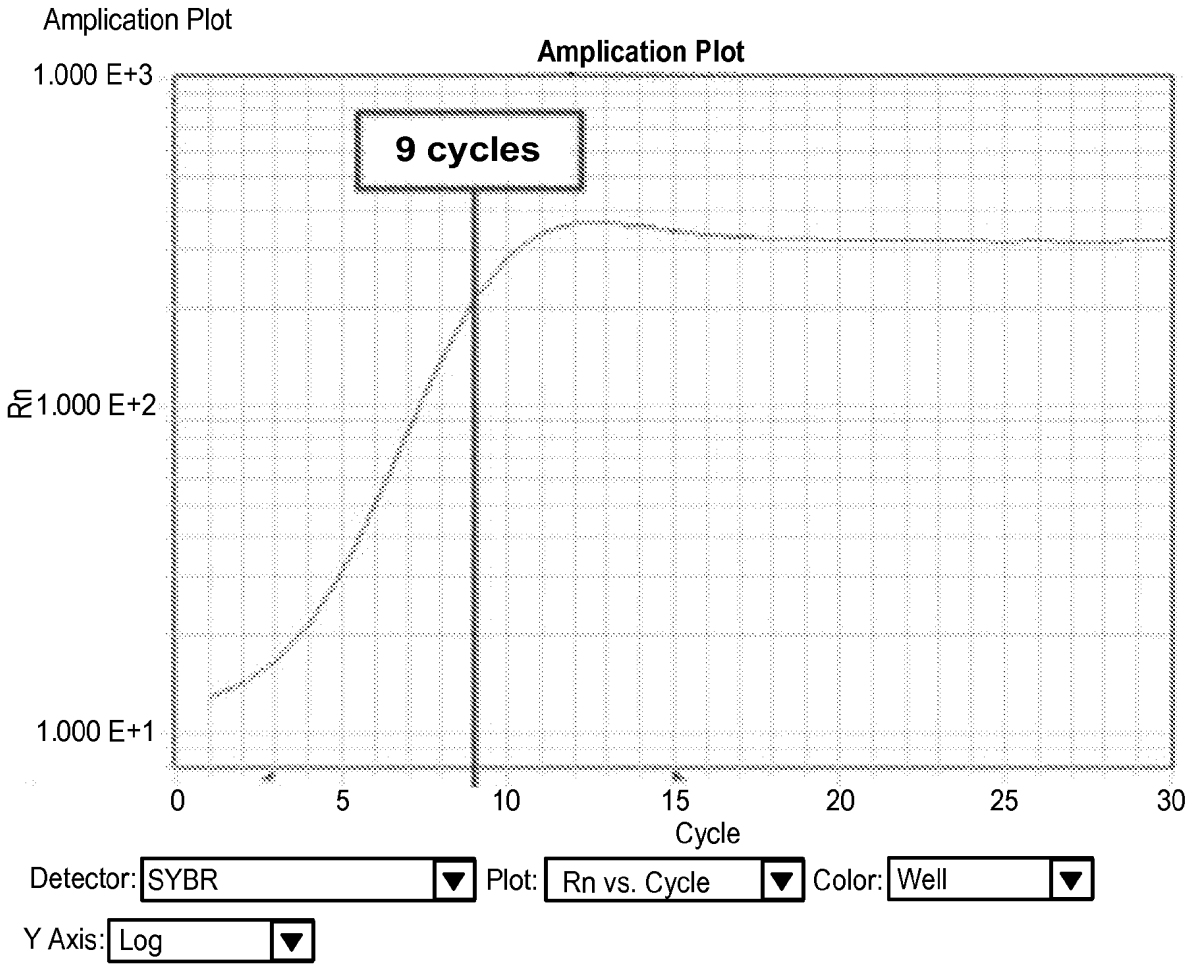


FIG. 28



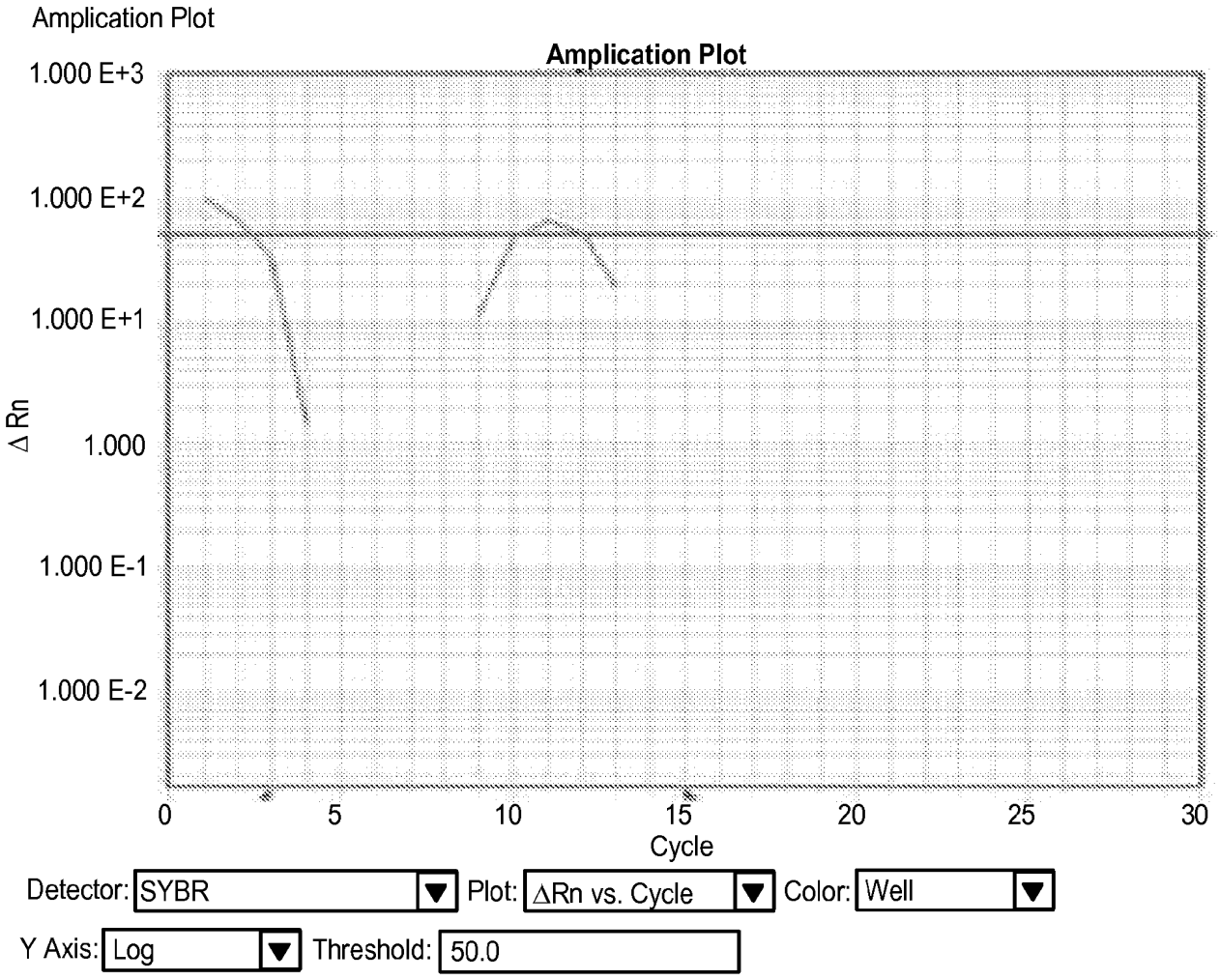
IDT Exhibit 1017.1711

FIG. 29



IDT Exhibit 1017.1712

FIG. 30



IDT Exhibit 1017.1713

FIG. 31

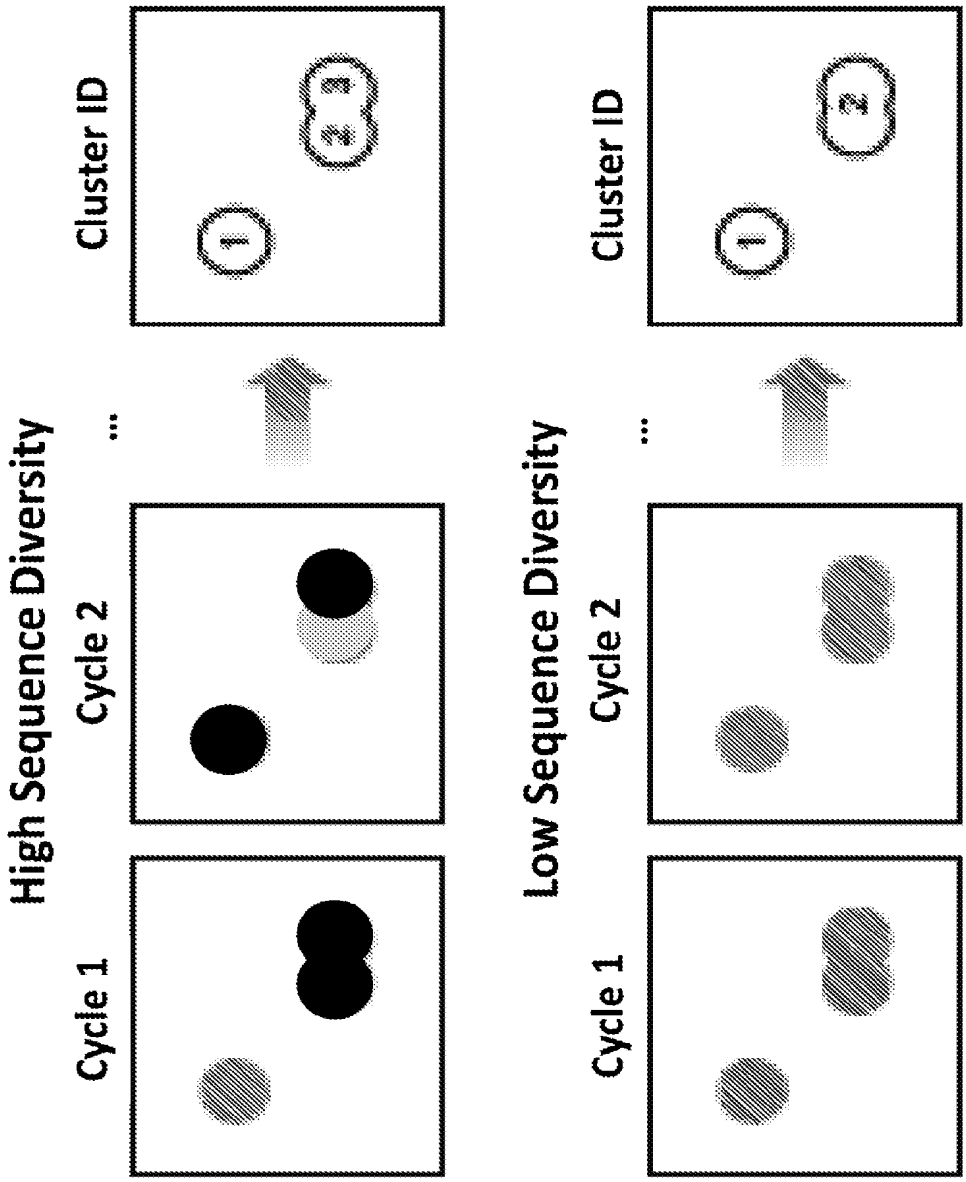

High Sequence Diversity
Cycle 1
Cycle 2
...
Cluster ID
1
2 3
Low Sequence Diversity
Cycle 1
Cycle 2
...
Cluster ID
1
2

FIG. 32

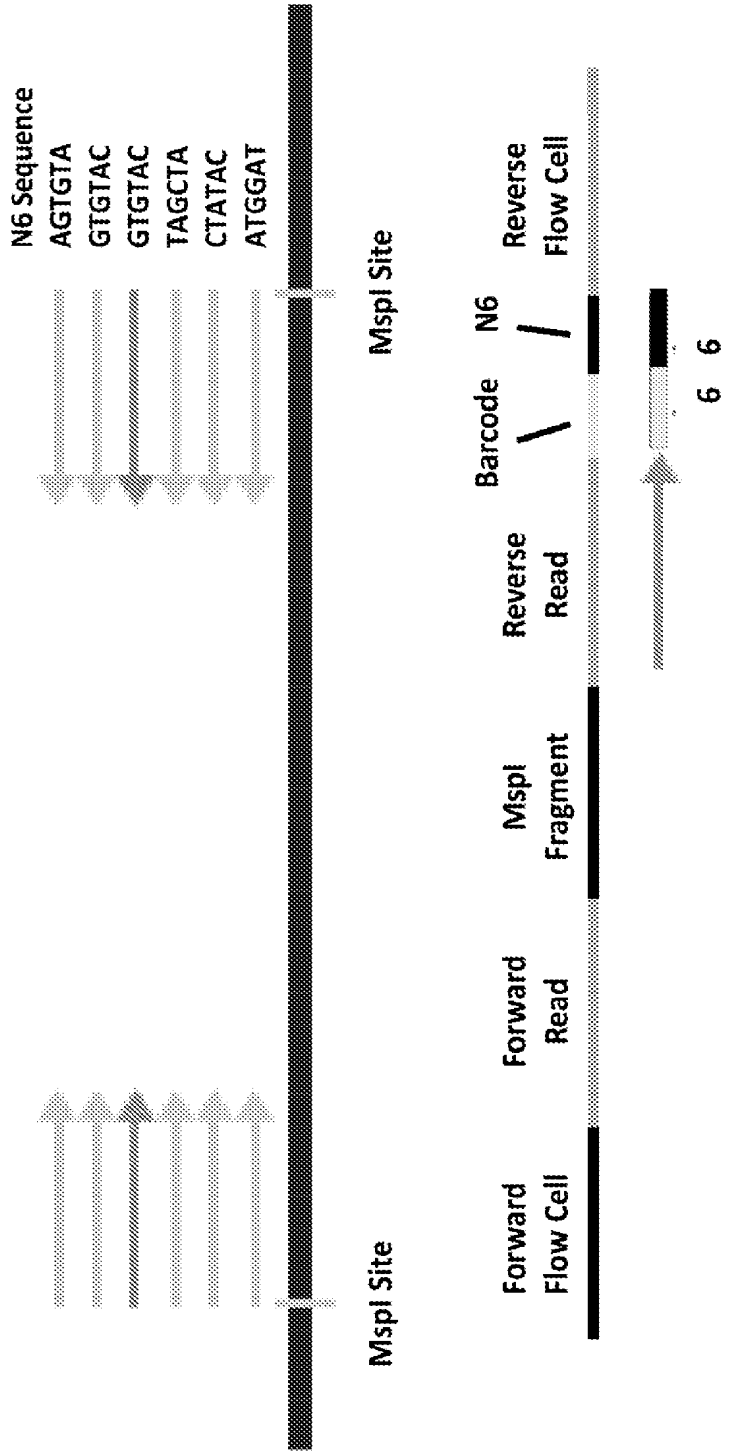

N6 Sequence
AGTGTA
GTGTAC
GTGTAC
TAGCTA
CTATAC
ATGGAT
MspI Site
MspI Site
Forward Flow Cell
Forward Read
MspI Fragment
Reverse Read
Barcode
N6
Reverse Flow Cell
6 6

FIG. 33

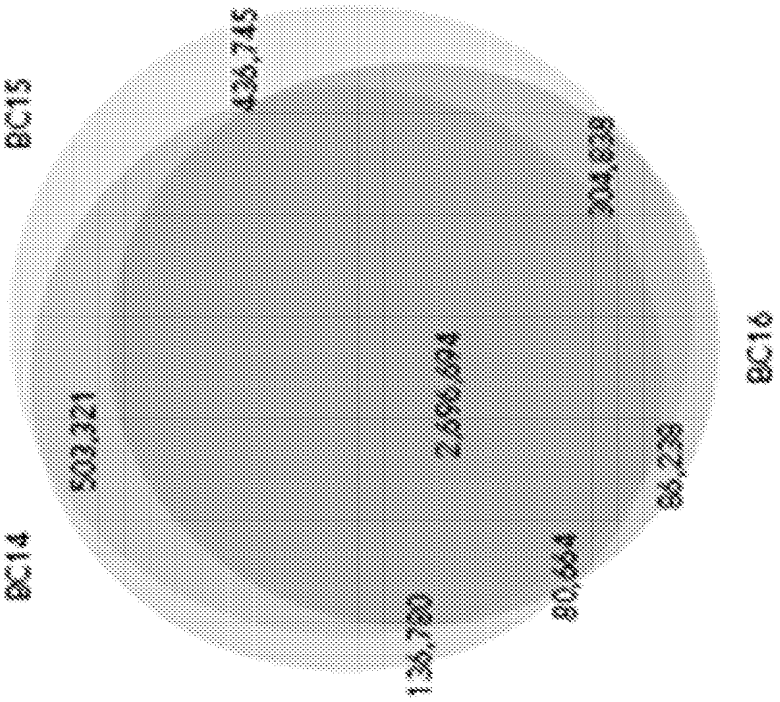

BC15
426,745
204,628
BC16
86,228
80,664
136,780
BC14
503,221
2,696,694

FIG. 34

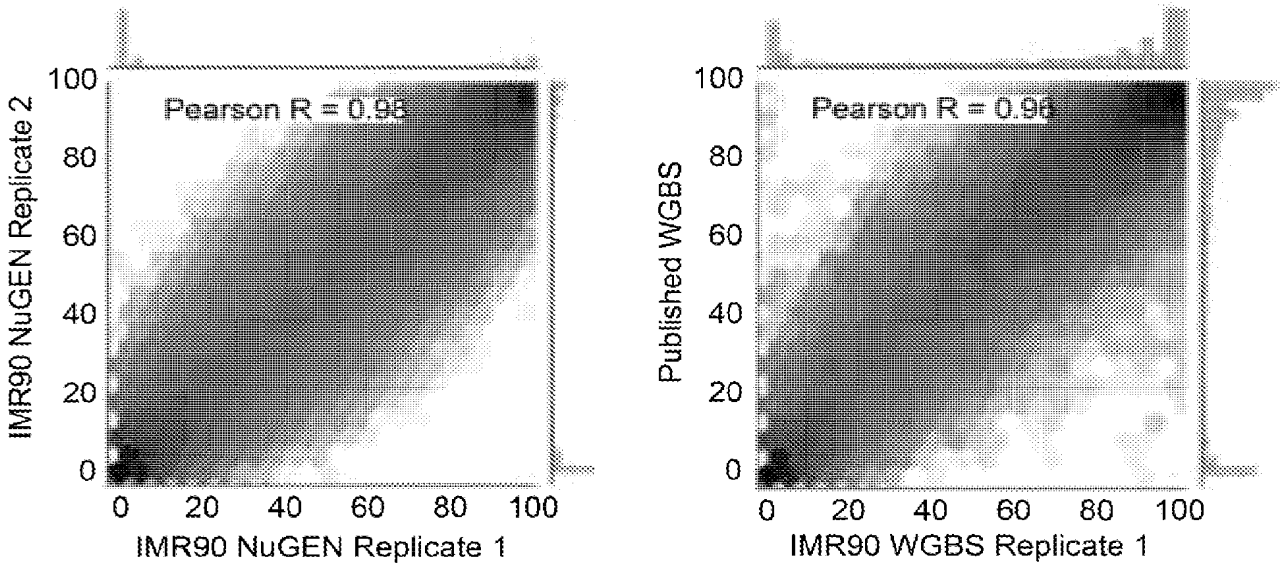

Pearson R = 0.98
IMR90 NuGEN Replicate 2
IMR90 NuGEN Replicate 1
0
20
40
60
80
100
Pearson R = 0.96
Published WGBS
IMR90 WGBS Replicate 1

FIG. 35

*D and R bases comprise DNA sequence derived from library construction

0 diversity bases added	5'	YGG(BCO)	BCO = Bisulfite-Converted Original genomic strand
1 diversity base added	5'	DYGG(BCO)	Y = C or T, depending on methylation status
2 diversity bases added	5'	DDYGG(BCO)	D = A, G, or T
3 diversity bases added	5'	RDDYGG(BCO)	R = A or G

FIG. 36

5' ...CCGG(OT)CCGG... 3'
3' ...GGCC(OB)GGCC... 5'
OT = Original Top genomic strand
OB = Original Bottom genomic strand
CG motifs that may contain 5-methyl C are underlined.
MspI digestion
5' pCGG(OT)C 3'
3' C(OB)GGCp 5'
Adaptor ligation
5' (Ad)CGG(OT)CCG(Ad) 3'
3' (Ad)GCC(OB)GGC(Ad) 5'
DNA sequence derived from library construction is shown in bold. Ad= Adaptor
Bisulfite conversion
5' (Ad)YGG(BCOT)CCG(Ad) 3'
and
3' (Ad)GCT(BCOB)GGY(Ad) 5'
BCOT = Bisulfite Converted Original Top genomic strand
BCOB = Bisulfite Converted Original Bottom genomic strand
Y = C or T, depending on methylation status

FIG. 37

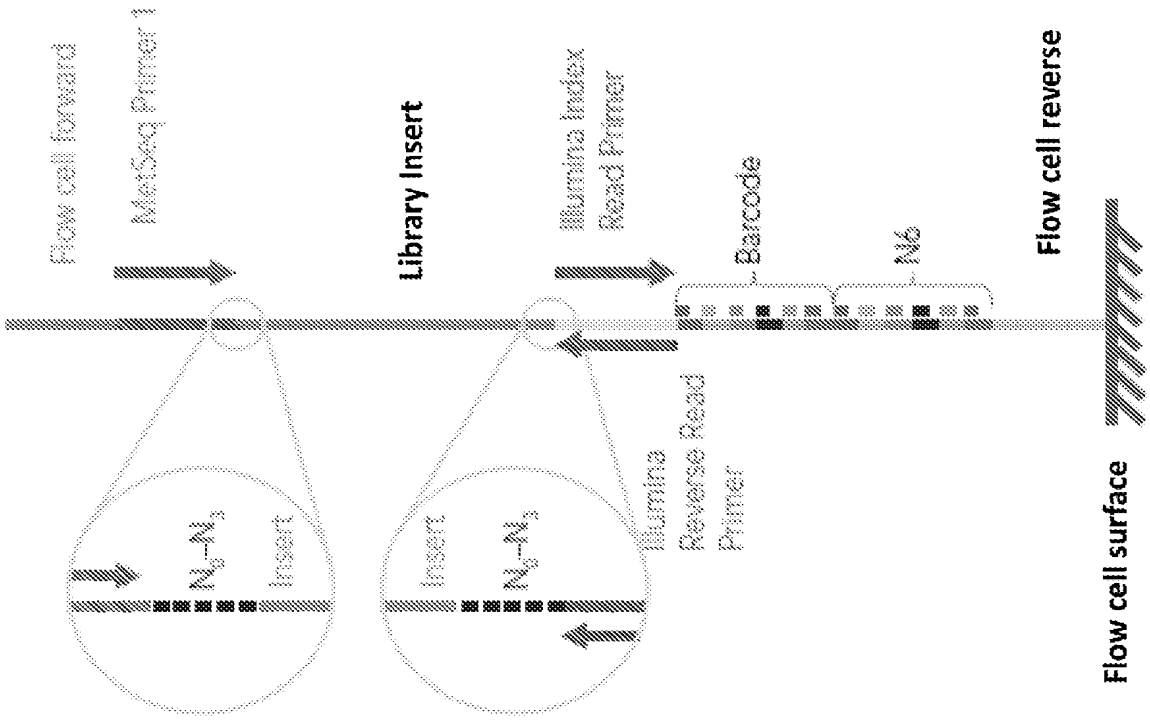
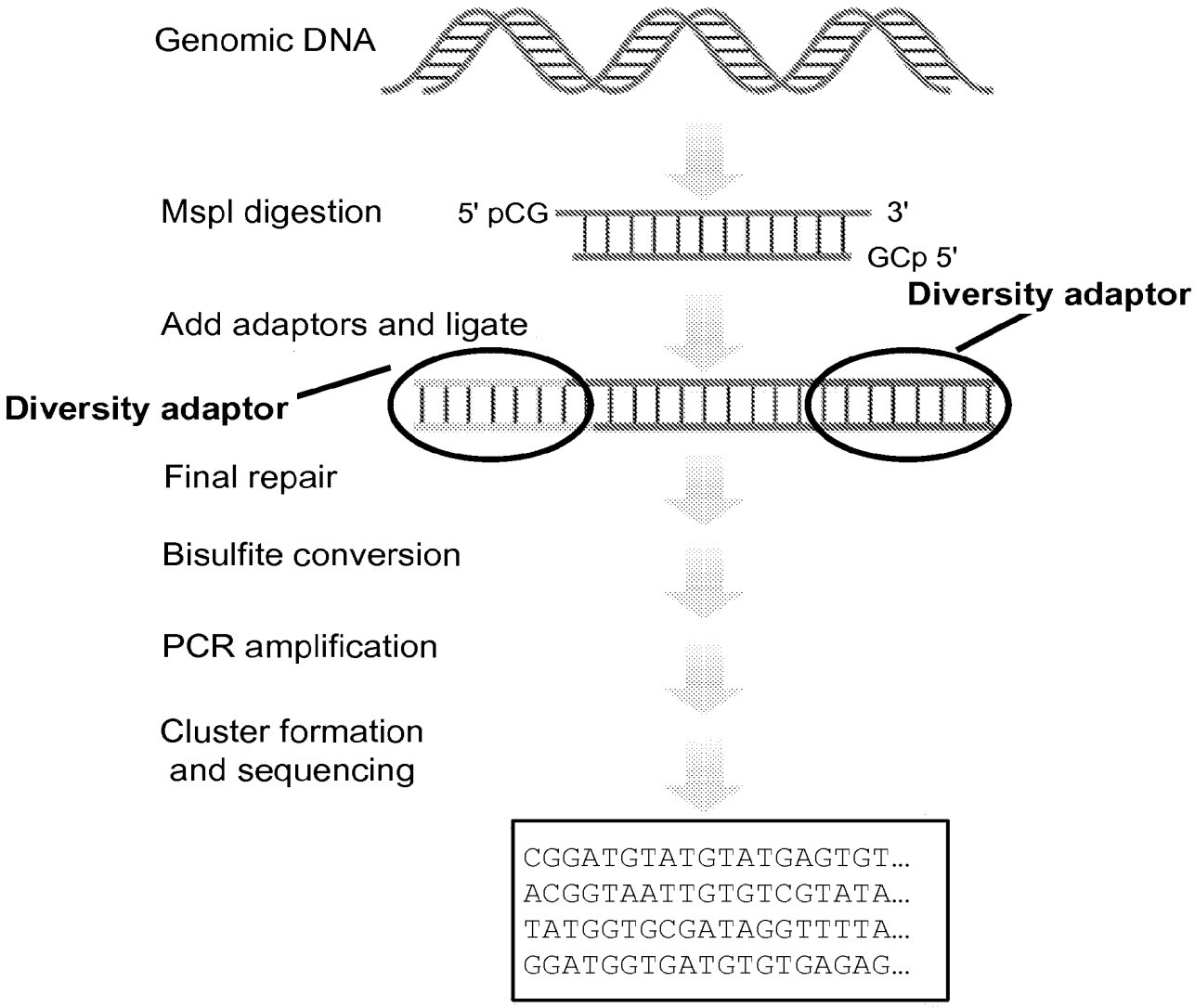

Flow cell forward
MetSeq Primer 1
Library Insert
Illumina Index Read Primer
Barcode
N6
Flow cell reverse
Flow cell surface
Illumina Reverse Read Primer
Insert
N_1-N_5
Insert
N_1-N_5

FIG. 38



IDT Exhibit 1017.1721

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2015/018112

A. CLASSIFICATION OF SUBJECT MATTER
IPC(8) - C12P 19/34 (2015.01)
CPC - C12Q 1/6874 (2015.04)
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC(8) - C12N 15/10; C12P 19/34; C40B 50/06; G06F 19/10 (2015.01)
CPC - C12Q 1/6874, 1/6806, 1/6837 (2015.04)

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
CPC - C12Q 1/6874, 1/6806, 1/6837 (2015.04) (keyword delimited)
US Classes - 435/6.11, 91.2; 506/16

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

PatBase, Google Patents, Google Scholar, Google, PubMed
Search terms used: bisulfite sequencing, methylation, epigenetic, diversity, nucleotide, oligonucleotide, polynucleotide, end, terminus, termini, terminal, sequence, extender, overhang, surface, cluster, in silico

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 2010/063711 A1 (ISTITUTO DI RICHERCHE DI BIOLOGIA MOLECOLARE P. ANGELETTI S. P. A. et al) 10 June 2010 (10.06.2010) entire document	1-9, 32-34, 45-53, 70-72, 83-94, 98, 99, 101-112, 134-136, 143
A	US 2012/0157322 A1 (MYLLYKANGAS et al) 21 June 2012 (21.06.2012) entire document	1-9, 32-34, 45-53, 70-72, 83-94, 98, 99, 101-112, 134-136, 143
A	WO 2013/177220 A1 (THE SCRIPPS RESEARCH INSTITUTE et al) 28 November 2013 (28.11.2013) entire document	1-9, 32-34, 45-53, 70-72, 83-94, 98, 99, 101-112, 134-136, 143
P, X	US 2014/0378345 A1 (10X TECHNOLOGIES INC et al) 25 December 2014 (25.12.2014) entire document	1-9, 32-34, 45-53, 83-94, 98, 99, 101-102
P, X	CHEN et al. "BisQC: an Operational Pipeline for Multiplexed Bisulfite Sequencing," BMC Genomics, 30 April 2014 (30.04.2014), Vol. 15, No. 290, Pgs. 1-22. entire document	1, 45, 83, 103, 143

☐ Further documents are listed in the continuation of Box C. ☐

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
08 May 2015	18 JUN 2015
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-8300	Authorized officer: Blaine R. Copenheaver PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

Form PCT/ISA/210 (second sheet) (July 2009)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2015/018112

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☒ Claims Nos.: 10-31, 35-44, 54-69, 73-82, 95-97, 100, 113-133, 137-142
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

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

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

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

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

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

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

Remark on Protest

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

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

☐ No protest accompanied the payment of additional search fees.

IDT Exhibit 1017.1723

Approved for use through 11/30/2020. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Substitute for form 1449/PTO

INFORMATION DISCLOSURE STATEMENT BY APPLICANT

(Use as many sheets as necessary)

Sheet 1 of 13

Complete if Known	
Application Number	16017340
Filing Date	06-25-2018
First Named Inventor	AMORESE; Doug
Art Unit	Not Yet Assigned
Examiner Name	Not Yet Assigned
Attorney Docket Number	25115-755.302

U. S. PATENT DOCUMENTS

Examiner Initials*	Cite No.[1]	Document Number Number-Kind Code[2] (if known)	Publication Date MM-DD-YYYY	Name of Patentee or Applicant of Cited Document	Pages, Columns, Lines, Where Relevant Passages or Relevant Figures Appear
	001	US-4458066	07-03-1984	CARUTHERS; Marvin H [Us] et al.	
	002	US-5525471	06-11-1996	ZENG; Jin	
	003	US-6197501	03-06-2001	CREMER; Thomas et al.	
	004	US-6913884	07-05-2005	STUELPNAGEL; John R. et al.	
	005	US-7056716	06-06-2006	POTTER; Robert Jason et al.	
	006	US-7115400	10-03-2006	ADESSI; Celine et al.	
	007	US-7232656	06-19-2007	BALASUBRAMANIAN; Shankar et al.	
	008	US-7264929	09-04-2007	ROTHBERG; Jonathan M. et al.	
	009	US-7294461	11-13-2007	KURN; Nurith	
	010	US-7351557	04-01-2008	KURN; Nurith	
	011	US-7354717	04-08-2008	KURN; Nurith	
	012	US-7771934	08-10-2010	KURN; Nurith	
	013	US-7771946	08-10-2010	KURN; Nurith	
	014	US-7846666	12-07-2010	KURN; Nurith	
	015	US-7846733	12-07-2010	KURN; Nurith	
	016	US-7867703	01-11-2011	SAMPSON; Jeffrey R. et al.	
	017	US-7939258	05-10-2011	KURN; Nurith et al.	
	018	US-7948015	05-24-2011	ROTHBERG; Jonathan M. et al.	
	019	US-7985565	07-26-2011	KAWASHIMA; Eric H. et al.	
	020	US-8034568	10-11-2011	KURN; Nurith et al.	
	021	US-8071311	12-06-2011	KURN; Nurith	
	022	US-8334116	12-18-2012	KURN; Nurith	

Examiner Signature		Date Considered	

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant. [1] Applicant's unique citation designation number (optional). [2] See Kinds Codes of USPTO Patent Documents at www.uspto.gov or MPEP 901.04. [3] Enter Office that issued the document, by the two-letter code (WIPO Standard ST.3). [4] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [5] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. [6] Applicant is to place a check mark here if English language Translation is attached. This collection of information is required by 37 CFR 1.97 and 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 2 hours to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 (1-800-786-9199) and select option 2.

Approved for use through 11/30/2020. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Substitute for form 1449/PTO

INFORMATION DISCLOSURE STATEMENT BY APPLICANT

(Use as many sheets as necessary)

Sheet 2 of 13

Complete if Known	
Application Number	16017340
Filing Date	06-25-2018
First Named Inventor	AMORESE; Doug
Art Unit	Not Yet Assigned
Examiner Name	Not Yet Assigned
Attorney Docket Number	25115-755.302

U. S. PATENT DOCUMENTS

Examiner Initials*	Cite No.[1]	Document Number Number-Kind Code[2] (if known)	Publication Date MM-DD-YYYY	Name of Patentee or Applicant of Cited Document	Pages, Columns, Lines, Where Relevant Passages or Relevant Figures Appear
	023	US-8465950	06-18-2013	KURN; Nurith et al.	
	024	US-8492095	07-23-2013	KURN; Nurith	
	025	US-8512956	08-20-2013	KURN; Nurith	
	026	US-8759036	06-24-2014	WANG; Yuker et al.	
	027	US-8852867	10-07-2014	KURN; Nurith et al.	
	028	US-8999677	04-07-2015	SOLDATOV; Aleksey et al.	
	029	US-9175325	11-03-2015	KURN; Nurith et al.	
	030	US-9175336	11-03-2015	SOLDATOV; Aleksey et al.	
	031	US-9181582	11-10-2015	KURN; Nurith	
	032	US-9206418	12-08-2015	ARMOUR; Chris et al.	
	033	US-9650628	05-16-2017	AMORESE; Doug [Us] et al.	
	034	US-9745614	08-29-2017	SCHROEDER; Benjamin G [Us] et al.	
	035	US-9822408	11-21-2017	AMORESE; Doug [Us] et al.	
	036	US-9957549	05-01-2018	ARMOUR; Christopher [Us] et al.	
	037	US-10036012	07-31-2018	AMORESE; Doug [Us] et al.	
	038	US-20060068415	03-30-2006	JONES; Keith W. et al.	
	039	US-20060134633	06-22-2006	CHEN; Yi-Ju et al.	
	040	US-20070141604	06-21-2007	GORMLEY; Niall A. et al.	
	041	US-20080160580	07-03-2008	ADESSI; Celine et al.	
	042	US-20080241831	10-02-2008	FAN; Jian-Bing et al.	
	043	US-20090036663	02-05-2009	KURN; Nurith	
	044	US-20090105081	04-23-2009	RODESCH; Matthew et al.	

Examiner Signature		Date Considered	

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant. [1] Applicant's unique citation designation number (optional). [2] See Kinds Codes of USPTO Patent Documents at www.uspto.gov or MPEP 901.04. [3] Enter Office that issued the document, by the two-letter code (WIPO Standard ST.3). [4] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [5] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. [6] Applicant is to place a check mark here if English language Translation is attached. This collection of information is required by 37 CFR 1.97 and 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 2 hours to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 (1-800-786-9199) and select option 2.

Approved for use through 11/30/2020. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Substitute for form 1449/PTO

INFORMATION DISCLOSURE STATEMENT BY APPLICANT

(Use as many sheets as necessary)

Sheet 3 of 13

Complete if Known	
Application Number	16017340
Filing Date	06-25-2018
First Named Inventor	AMORESE; Doug
Art Unit	Not Yet Assigned
Examiner Name	Not Yet Assigned
Attorney Docket Number	25115-755.302

U. S. PATENT DOCUMENTS

Examiner Initials*	Cite No.[1]	Document Number Number-Kind Code[2] (if known)	Publication Date MM-DD-YYYY	Name of Patentee or Applicant of Cited Document	Pages, Columns, Lines, Where Relevant Passages or Relevant Figures Appear
	045	US-20100081174	04-01-2010	DUNN; John J [Us] et al.	
	046	US-20100105049	04-29-2010	EHRICH; Mathias et al.	
	047	US-20100159559	06-24-2010	KURN; Nurith et al.	
	048	US-20100203597	08-12-2010	CHEN; Liangjing et al.	
	049	US-20100311066	12-09-2010	KURN; Nurith	
	050	US-20110091882	04-21-2011	GRANADOS; Edward N. et al.	
	051	US-20110269194	11-03-2011	MAKAROV; Vladimir	
	052	US-20110294132	12-01-2011	KURN; Nurith	
	053	US-20120003657	01-05-2012	MYLLYKANGAS; Samuel et al.	
	054	US-20120028310	02-02-2012	KURN; Nurith et al.	
	055	US-20120149068	06-14-2012	KURN; Nurith	
	056	US-20120156728	06-21-2012	LI; Bin et al.	
	057	US-20120190587	07-26-2012	KURN; Nurith et al.	
	058	US-20120238738	09-20-2012	HENDRICKSON; Cynthia	
	059	US-20120289426	11-15-2012	ROOS; Fredrik Johansson Henrik Isaksson Magnus Nilsson Mats Ericsson Olle et al.	
	060	US-20140038188	02-06-2014	KURN; Nurith	
	061	US-20140038236	02-06-2014	KURN; Nurith et al.	
	062	US-20140303000	10-09-2014	ARMOUR; Chris et al.	
	063	US-20150225773	08-13-2015	FARMER; Andrew Alan et al.	
	064	US-20150284769	10-08-2015	SCHROEDER; Benjamin G. et al.	

Examiner Signature		Date Considered	

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant. [1] Applicant's unique citation designation number (optional). [2] See Kinds Codes of USPTO Patent Documents at www.uspto.gov or MPEP 901.04. [3] Enter Office that issued the document, by the two-letter code (WIPO Standard ST.3). [4] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [5] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. [6] Applicant is to place a check mark here if English language Translation is attached. This collection of information is required by 37 CFR 1.97 and 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 2 hours to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 (1-800-786-9199) and select option 2.

Approved for use through 11/30/2020. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Substitute for form 1449/PTO

INFORMATION DISCLOSURE STATEMENT BY APPLICANT

(Use as many sheets as necessary)

Sheet 4 of 13

Complete if Known	
Application Number	16017340
Filing Date	06-25-2018
First Named Inventor	AMORESE; Doug
Art Unit	Not Yet Assigned
Examiner Name	Not Yet Assigned
Attorney Docket Number	25115-755.302

U. S. PATENT DOCUMENTS

Examiner Initials*	Cite No.[1]	Document Number Number-Kind Code[2] (if known)	Publication Date MM-DD-YYYY	Name of Patentee or Applicant of Cited Document	Pages, Columns, Lines, Where Relevant Passages or Relevant Figures Appear
	065	US-20160122756	05-05-2016	ARMOUR; Christopher	
	066	US-20160153039	06-02-2016	AMORESE; Doug et al.	
	067	US-20160251711	09-01-2016	AMORESE; Doug et al.	
	068	US-20160251712-A1	09-01-2016	AMORESE; Doug et al.	
	069	US-20160265042-A1	09-15-2016	SCHROEDER; Benjamin G. et al.	
	070	US-20160275240	09-22-2016	HUELGA; Stephanie et al.	
	071	US-20160275240-A1	09-22-2016	HUELGA; Stephanie et al.	
	072	US-20170298345-A1	10-19-2017	AMORESE; Doug et al.	

FOREIGN PATENT DOCUMENTS

Examiner Initials*	Cite No.[1]	Foreign Patent Document Country Code[3]-Number[4]-Kind Code[5] (if known)	Publication Date MM-DD-YYYY	Name of Patentee or Applicant of Cited Document	Pages, Columns, Lines, Where Relevant Passages Or Relevant Figures Appear	T[6]
	001	EP-2272976-A1	01-12-2011	MAX PLANCK GESELLSCHAFT [DE]		☐
	002	EP-2451973-A1	05-16-2012	MAX PLANCK GESELLSCHAFT [DE]		☐
	003	WO-0070039-A1	11-23-2000	UNIV MCGILL [CA], et al.		☐
	004	WO-2004092418-A2	10-28-2004	NUGEN TECHNOLOGIES INC [US], et al.		☐

Examiner Signature		Date Considered	

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant. [1] Applicant's unique citation designation number (optional). [2] See Kinds Codes of USPTO Patent Documents at www.uspto.gov or MPEP 901.04. [3] Enter Office that issued the document, by the two-letter code (WIPO Standard ST.3). [4] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [5] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. [6] Applicant is to place a check mark here if English language Translation is attached. This collection of information is required by 37 CFR 1.97 and 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 2 hours to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 (1-800-786-9199) and select option 2.

Approved for use through 11/30/2020. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Substitute for form 1449/PTO

INFORMATION DISCLOSURE STATEMENT BY APPLICANT

(Use as many sheets as necessary)

Sheet 5 of 13

Complete if Known	
Application Number	16017340
Filing Date	06-25-2018
First Named Inventor	AMORESE; Doug
Art Unit	Not Yet Assigned
Examiner Name	Not Yet Assigned
Attorney Docket Number	25115-755.302

FOREIGN PATENT DOCUMENTS

Examiner Initials*	Cite No.[1]	Foreign Patent Document Country Code[3]-Number[4]-Kind Code[5] (if known)	Publication Date MM-DD-YYYY	Name of Patentee or Applicant of Cited Document	Pages, Columns, Lines, Where Relevant Passages Or Relevant Figures Appear	T[6]
	005	WO-2006086668-A2	08-17-2006	EPICT TECHNOLOGIES [US], et al.		☐
	006	WO-2007019444-A2	02-15-2007	EUCLID DIAGNOSTICS LLC [US], et al.		☐
	007	WO-2009102878-A2	08-20-2009	NUGEN TECHNOLOGIES INC [US], et al.		☐
	008	WO-2009102896-A2	08-20-2009	NUGEN TECHNOLOGIES INC [US], et al.		☐
	009	WO-2009117698-A2	09-24-2009	NUGEN TECHNOLOGIES INC [US], et al.		☐
	010	WO-2010085715-A1	07-29-2010	QUANTA BIOSCIENCES [US], et al.		☐
	011	WO-2011019964-A1	02-17-2011	NUGEN TECHNOLOGIES INC [US], et al.		☐
	012	WO-2011053987-A1	05-05-2011	NUGEN TECHNOLOGIES INC [US], et al.		☐
	013	WO-2011151777-A1	12-08-2011	CONSIGLIO NAZIONALE RICERCHE [IT], et al.		☐
	014	WO-2012013932-A1	02-02-2012	UNIV ST ANDREWS [GB], et al.		☐

Examiner Signature		Date Considered	

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant. [1] Applicant's unique citation designation number (optional). [2] See Kinds Codes of USPTO Patent Documents at www.uspto.gov or MPEP 901.04. [3] Enter Office that issued the document, by the two-letter code (WIPO Standard ST.3). [4] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [5] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. [6] Applicant is to place a check mark here if English language Translation is attached. This collection of information is required by 37 CFR 1.97 and 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 2 hours to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 (1-800-786-9199) and select option 2.

Approved for use through 11/30/2020. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Substitute for form 1449/PTO

INFORMATION DISCLOSURE STATEMENT BY APPLICANT

(Use as many sheets as necessary)

Sheet 6 of 13

Complete if Known	
Application Number	16017340
Filing Date	06-25-2018
First Named Inventor	AMORESE; Doug
Art Unit	Not Yet Assigned
Examiner Name	Not Yet Assigned
Attorney Docket Number	25115-755.302

FOREIGN PATENT DOCUMENTS

Examiner Initials*	Cite No.[1]	Foreign Patent Document Country Code[3]–Number[4]–Kind Code[5] (if known)	Publication Date MM-DD-YYYY	Name of Patentee or Applicant of Cited Document	Pages, Columns, Lines, Where Relevant Passages Or Relevant Figures Appear	T[6]
	015	WO-2013059740-A1	04-25-2013	FOUNDATION MEDICINE INC [US]		☐
	016	WO-2013112923-A1	08-01-2013	NUGEN TECHNOLOGIES INC [US], et al.		☐
	017	WO-2014144092-A1	09-18-2014	NUGEN TECHNOLOGIES INC [US]		☐
	018	WO-2014150931-A1	09-25-2014	NUGEN TECHNOLOGIES INC [US]		☐
	019	WO-2015122967-A1	08-20-2015	CLONTECH LAB INC [US]		☐
	020	WO-2015131107-A1	09-03-2015	NUGEN TECHNOLOGIES INC [US]		☐
	021	WO-2016033251-A2	03-03-2016	NUGEN TECHNOLOGIES INC [US]		☐
	022	WO-2016100955-A2	06-23-2016	IDENTIFYGENOMICS LLC [US]		☐

Examiner Signature		Date Considered	

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant. [1] Applicant's unique citation designation number (optional). [2] See Kinds Codes of USPTO Patent Documents at www.uspto.gov or MPEP 901.04. [3] Enter Office that issued the document, by the two-letter code (WIPO Standard ST.3). [4] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [5] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. [6] Applicant is to place a check mark here if English language Translation is attached. This collection of information is required by 37 CFR 1.97 and 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 2 hours to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 (1-800-786-9199) and select option 2.

Approved for use through 11/30/2020. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Substitute for form 1449/PTO

INFORMATION DISCLOSURE STATEMENT BY APPLICANT

(Use as many sheets as necessary)

Sheet 7 of 13

Complete if Known	
Application Number	16017340
Filing Date	06-25-2018
First Named Inventor	AMORESE; Doug
Art Unit	Not Yet Assigned
Examiner Name	Not Yet Assigned
Attorney Docket Number	25115-755.302

NON-PATENT LITERATURE DOCUMENTS

Examiner Initials*	Cite No.[1]	Include name of the author(in CAPITAL LETTERS),title of the article(when appropriate), title of the item (book,magazine,journal,serial,symposium,catalog,etc.),date,page(s),volume-issue number(s),publisher, city and/or country where published.	T[2]
	001	BALDWIN et al. Multilocus sequence typing of Cronobacter sakazakii and Cronobacter malonaticus reveals stable clonal structures with clinical significance which do not correlate with biotypes. BMC Microbiology 9:223. 9 pages. (October 23, 2009). DOI: https://doi.org/10.1186/1471-2180-9-223.	☐
	002	BEAUCAGE et al. Deoxynucleoside Phosphoramidites-A New Class of Key Intermediates for Deoxypolynucleotide Synthesis. Tetrahedron Letters. 1981;22(20):1859-1862.	☐
	003	BigDye Terminator v3.1 Cycle Sequencing Kit. Applied Biosystems, pp. 1-72 (2002).	☐
	004	BODI, et al. Comparison of Commercially Available Target Enrichment Methods for Next-Generation Sequencing. J Biomol Tech. 2013 Jul; 24(2): 73-86.	☐
	005	BROWN, et al. Chemical synthesis and cloning of a tyrosine tRNA gene. Methods Enzymol. 1979;68:109-51.	☐
	006	KURN, et al., United States Patent Application No. US13/980987, filed on 07-22-2013.	☐
	007	AMORESE, et al., United States Patent Application No. US14/836936, filed on 08-26-2015.	☐
	008	European search report and opinion dated 01/29/2016 for EP Application No. 13806978.	☐
	009	FROUSSARD. A random-PCR method (rPCR) to construct whole cDNA library from low amounts of RNA. Nucleic Acids Res. 1992 Jun 11;20(11):2900.	☐
	010	GUDMUNDSSON, et al. Genome-wide association and replication studies identify four variants associated with prostate cancer susceptibility. Nat Genet. 2009 Oct;41(10):1122-6. doi: 10.1038/ng.448. Epub 2009 Sep 20.	☐

Examiner Signature		Date Considered	

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant. 1 Applicant's unique citation designation number (optional). 2 Applicant is to place a check mark here if English language Translation is attached. This collection of information is required by 37 CFR 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 2 hours to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 (1-800-786-9199) and select option 2.

Approved for use through 11/30/2020. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Substitute for form 1449/PTO

INFORMATION DISCLOSURE STATEMENT BY APPLICANT

(Use as many sheets as necessary)

Sheet 8 of 13

Complete if Known	
Application Number	16017340
Filing Date	06-25-2018
First Named Inventor	AMORESE; Doug
Art Unit	Not Yet Assigned
Examiner Name	Not Yet Assigned
Attorney Docket Number	25115-755.302

NON-PATENT LITERATURE DOCUMENTS

Examiner Initials*	Cite No.[1]	Include name of the author(in CAPITAL LETTERS),title of the article(when appropriate), title of the item (book,magazine,journal,serial,symposium,catalog,etc.),date,page(s),volume-issue number(s),publisher, city and/or country where published.	T[2]
	011	HUANG et al., The behavior of 5-hydroxymethylcytosine in Bisulfite sequencing. PLOS one, January 2010; 5(1):e8888. 9 Pages.	☐
	012	International search report and written opinion dated 02/05/2016 for PCT/US2015/047053.	☐
	013	JANG, et al., Spectra of BRCA1 and BRCA2 mutations in Korean patients with breast cancer: the importance of whole-gene sequencing. Journal of Human Genetics. Vol. 57. Pages 212-215. January 5, 2012.	☐
	014	KURN, Nurith, United States Patent Application No. 14/877075, filed 10/07/2015.	☐
	015	LEVESQUE-SERGERIE, et al. Detection limits of several commercial reverse transcriptase enzymes: impact on the low- and high-abundance transcript levels assessed by quantitative RT-PCR. BMC Mol Biol. 2007 Oct 22;8:93.	☐
	016	MULLIGAN et al., Differential binding of Escherichia coli McrA protein to DNA sequences that contain the dinucleotide m5CpG. Nucleic Acids Research, 2010; 38(6):1997-2005.	☐
	017	Multi-Site Study Validates End-to-End RNA-Seq Sample Prep Workflow for Human Whole Blood Clinical Samples. Nugen Technologies Press Release. San Carlos. July 15, 2015. 3 pages. http://www.businesswire.com/news/home/20150715005371/en#.VaZOCPlWLKE.	☐
	018	NARANG, et al. Improved phosphotriester method for the synthesis of gene fragments. Methods Enzymol. 1979;68:90-8.	☐
	019	Notice of allowance dated 01/05/2017 for US Application No. 13/750,768.	☐
	020	Notice of Allowance dated 04/17/2017 for US Application No. 14/634,326.	☐
	021	Notice of Allowance dated 06/21/2017 for US Application No. 14/634,326.	☐

Examiner Signature		Date Considered	

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant. 1 Applicant's unique citation designation number (optional). 2 Applicant is to place a check mark here if English language Translation is attached. This collection of information is required by 37 CFR 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 2 hours to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 (1-800-786-9199) and select option 2.

Approved for use through 11/30/2020. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Substitute for form 1449/PTO

INFORMATION DISCLOSURE STATEMENT BY APPLICANT

(Use as many sheets as necessary)

Sheet 9 of 13

Complete if Known	
Application Number	16017340
Filing Date	06-25-2018
First Named Inventor	AMORESE; Doug
Art Unit	Not Yet Assigned
Examiner Name	Not Yet Assigned
Attorney Docket Number	25115-755.302

NON-PATENT LITERATURE DOCUMENTS

Examiner Initials*	Cite No.[1]	Include name of the author(in CAPITAL LETTERS),title of the article(when appropriate), title of the item (book,magazine,journal,serial,symposium,catalog,etc.),date,page(s),volume-issue number(s),publisher, city and/or country where published.	T[2]
	022	Notice of Allowance dated 07/25/2017 for US Application No. 14/778,564. 9 Pages.	☐
	023	Notice of Allowance dated 12/26/2017 for U.S. Patent Application No. 14/390,012. 7 Pages.	☐
	024	NuGEN Launches Enhanced Target Enrichment Technology Allowing More Comprehensive and Efficient NGS Analysis of Cancer Samples. San Carlos, CA. August 03, 2015. 2 pages.	☐
	025	NuGEN scientists screen 400+ genes for fusion events in single assay; detect new fusions and those previously associated with cancers. Nugen Technologies. San Carlos. July 1, 2015.	☐
	026	Office action dated 02/05/2014 for US Application No. 13/750,768.	☐
	027	Office action dated 04/04/2016 for US Application No. 14/995,882.	☐
	028	Office action dated 04/07/2016 for US Application No. 14/390,012.	☐
	029	Office action dated 06/02/2016 for US Application No. 13/750,768.	☐
	030	Office action dated 07/15/2015 for US Application No. 13/750,768.	☐
	031	Office Action dated 08/07/2017 for US Application No. 14/920,254. 9 Pages.	☐
	032	Office Action dated 08/18/2017 for U.S. Patent Application No. 15/471,785.	☐
	033	Office Action dated 08/22/2017 for U.S. Patent Application No. 14/836,936. 20 Pages.	☐
	034	Office Action dated 08/30/2017 for U.S. Patent Application No. 14/990,339. 9 Pages.	☐
	035	Office action dated 09/25/2014 for US Application No. 13/750,768.	☐

Examiner Signature		Date Considered	

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant. 1 Applicant's unique citation designation number (optional). 2 Applicant is to place a check mark here if English language Translation is attached. This collection of information is required by 37 CFR 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 2 hours to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 (1-800-786-9199) and select option 2.

Approved for use through 11/30/2020. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Substitute for form 1449/PTO

INFORMATION DISCLOSURE STATEMENT BY APPLICANT

(Use as many sheets as necessary)

Sheet | 10 | of | 13

Complete if Known	
Application Number	16017340
Filing Date	06-25-2018
First Named Inventor	AMORESE; Doug
Art Unit	Not Yet Assigned
Examiner Name	Not Yet Assigned
Attorney Docket Number	25115-755.302

NON-PATENT LITERATURE DOCUMENTS

Examiner Initials*	Cite No.[1]	Include name of the author(in CAPITAL LETTERS),title of the article(when appropriate), title of the item (book,magazine,journal,serial,symposium,catalog,etc.),date,page(s),volume-issue number(s),publisher, city and/or country where published.	T[2]
	036	Office action dated 10/30/2015 for US Application No. 13/750,768.	☐
	037	Office Action dated 12/15/2017 for U.S. Patent Application No. 14/991,340, 14 Pages.	☐
	038	Office Action dated 7/19/2017 for US Application No. 14/390,012. 12 Pages.	☐
	039	OUT, et al. Deep sequencing to reveal new variants in pooled DNA samples. Hum Mutat. 2009 Dec; 30(12): 1703-12. doi: 10.1002/humu.21122.	☐
	040	Ovation® Target Enrichment System. User guide. Nugen. 2016. 45 pages.	☐
	041	SCOLNICK, et al. An Efficient Method for Identifying Gene Fusions by Targeted RNA Sequencing from Fresh Frozen and FFPE Samples. PLoS ONE 10(7): e0128916. doi:10.1371/journal.pone.0128916.	☐
	042	U.S. Serial No. 14/990,339 Office Action dated August 29, 2018.	☐
	043	U.S. Serial No. 14/991,340 Office Action dated September 28, 2018.	☐
	044	U.S. Serial No. 14/990,339 Office Action dated January 3, 2018.	☐
	045	TURNER, et al. Massively parallel exon capture and library-free resequencing across 16 genomes. Nat Methods. 2009 May; 6(5): 315-6. doi: 10.1038/nmeth.f.248. Epub 2009 Apr 6.	☐
	046	U.S. 15/471,785 Notice of Allowance dated April 10, 2018.	☐
	047	U.S. 15/471,785 Notice of Allowance dated June 21, 2018.	☐
	048	U.S. Serial 14/920,254 Office Action dated March 15, 2018, 13 pages.	☐

Examiner Signature		Date Considered	

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant. 1 Applicant's unique citation designation number (optional). 2 Applicant is to place a check mark here if English language Translation is attached. This collection of information is required by 37 CFR 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 2 hours to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 (1-800-786-9199) and select option 2.

Approved for use through 11/30/2020. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Substitute for form 1449/PTO

INFORMATION DISCLOSURE STATEMENT BY APPLICANT

(Use as many sheets as necessary)

Sheet 11 of 13

Complete if Known	
Application Number	16017340
Filing Date	06-25-2018
First Named Inventor	AMORESE; Doug
Art Unit	Not Yet Assigned
Examiner Name	Not Yet Assigned
Attorney Docket Number	25115-755.302

NON-PATENT LITERATURE DOCUMENTS

Examiner Initials*	Cite No.[1]	Include name of the author(in CAPITAL LETTERS),title of the article(when appropriate), title of the item (book,magazine,journal,serial,symposium,catalog,etc.),date,page(s),volume-issue number(s),publisher, city and/or country where published.	T[2]
	049	VATER, et al. Short bioactive Spiegelmers to migraine-associated calcitonin gene-related peptide rapidly identified by a novel approach: tailored-SELEX. Nucleic Acids Res. 2003 Nov 1;31(21):e130.	☐
	050	VOELKERDING, et al. Next generation sequencing: from basic research to diagnostics. Clin. Chem. 2009; 55(4):641-658.	☐
	051	WATSON, et al. Cloning and assembly of PCR products using modified primers and DNA repair enzymes. Biotechniques. 1997 Nov;23(5):858-62, 864.	☐
	052	WEBER et al. Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. Nat. Genet. 37:853-862 (2005).	☐
	053	ZHANG, et al. Copy number variation in human health, disease, and evolution. Annu Rev Genomics Hum Genet. 2009;10:451-81. doi: 10.1146/annurev.genom.9.081307.164217.	☐
	054	ZHONG, et al. High-throughput illumina strand-specific RNA sequencing library preparation. Cold Spring Harb. Protoc.; 2011; 940-949. doi:10.1101/pdb.prot5652.	☐

Examiner Signature		Date Considered	

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant. 1 Applicant's unique citation designation number (optional). 2 Applicant is to place a check mark here if English language Translation is attached. This collection of information is required by 37 CFR 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 2 hours to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 (1-800-786-9199) and select option 2.

Approved for use through 11/30/2020. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Substitute for form 1449/PTO

INFORMATION DISCLOSURE STATEMENT BY APPLICANT

(Use as many sheets as necessary)

Sheet 12 of 13

Complete if Known	
Application Number	16017340
Filing Date	06-25-2018
First Named Inventor	AMORESE; Doug
Art Unit	Not Yet Assigned
Examiner Name	Not Yet Assigned
Attorney Docket Number	25115-755.302

CERTIFICATION STATEMENT

Please see 37 CFR 1.97 and 1.98 to make the appropriate selection(s):

☐ That each item of information contained in the information disclosure statement was first cited in any communication from a foreign patent office in a counterpart foreign application not more than three months prior to the filing of the information disclosure statement. See 37 CFR 1.97(e)(1).

OR

☐ That no item of information contained in the information disclosure statement was cited in a communication from a foreign patent office in a counterpart foreign application, and, to the knowledge of the person signing the certification after making reasonable inquiry, no item of information contained in the information disclosure statement was known to any individual designated in 37 CFR 1.56(c) more than three months prior to the filing of the information disclosure statement. See 37 CFR 1.97(e)(2).

☐ See attached certification statement.

☐ Fee set forth in 37 CFR 1.17 (p) has been submitted herewith.

☒ A certification statement is not submitted herewith.

SIGNATURE

A signature of the applicant or representative is required in accordance with CFR 1.33, 10.18. Please see CFR 1.4(d) for the form of the signature.

Signature	/Jeffrey Seidel/	Date (YYYY-MM-DD)	2018-12-04
Name/Print	Jeffrey Seidel	Registration Number	61,063

This collection of information is required by 37 CFR 1.97 and 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR1.14. This collection is estimated to take 1 hour to complete, including gathering, preparing and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

Privacy Act Statement

The Privacy Act of 1974 (P.L. 93-579) requires that you be given certain information in connection with your submission of the attached form related to a patent application or patent. Accordingly, pursuant to the requirements of the Act, please be advised that: (1) the general authority for the collection of this information is 35 U.S.C. 2(b)(2); (2) furnishing of the information solicited is voluntary; and (3) the principal purpose for which the information is used by the U.S. Patent and Trademark Office is to process and/or examine your submission related to a patent application or patent. If you do not furnish the requested information, the U.S. Patent and Trademark Office may not be able to process and/or examine your submission, which may result in termination of proceedings or abandonment of the application or expiration of the patent.

The information provided by you in this form will be subject to the following routine uses:

1. The information on this form will be treated confidentially to the extent allowed under the Freedom of Information Act (5 U.S.C. 552) and the Privacy Act (5 U.S.C 552a). Records from this system of records may be disclosed to the Department of Justice to determine whether disclosure of these records is required by the Freedom of Information Act.
2. A record from this system of records may be disclosed, as a routine use, in the course of presenting evidence to a court, magistrate, or administrative tribunal, including disclosures to opposing counsel in the course of settlement negotiations.
3. A record in this system of records may be disclosed, as a routine use, to a Member of Congress submitting a request involving an individual, to whom the record pertains, when the individual has requested assistance from the Member with respect to the subject matter of the record.
4. A record in this system of records may be disclosed, as a routine use, to a contractor of the Agency having need for the information in order to perform a contract. Recipients of information shall be required to comply with the requirements of the Privacy Act of 1974, as amended, pursuant to 5 U.S.C. 552a(m).
5. A record related to an International Application filed under the Patent Cooperation Treaty in this system of records may be disclosed, as a routine use, to the International Bureau of the World Intellectual Property Organization, pursuant to the Patent Cooperation Treaty.
6. A record in this system of records may be disclosed, as a routine use, to another federal agency for purposes of National Security review (35 U.S.C. 181) and for review pursuant to the Atomic Energy Act (42 U.S.C. 218(c)).
7. A record from this system of records may be disclosed, as a routine use, to the Administrator, General Services, or his/her designee, during an inspection of records conducted by GSA as part of that agency's responsibility to recommend improvements in records management practices and programs, under authority of 44 U.S.C. 2904 and 2906. Such disclosure shall be made in accordance with the GSA regulations governing inspection of records for this purpose, and any other relevant (i.e., GSA or Commerce) directive. Such disclosure shall not be used to make determinations about individuals.
8. A record from this system of records may be disclosed, as a routine use, to the public after either publication of the application pursuant to 35 U.S.C. 122(b) or issuance of a patent pursuant to 35 U.S.C. 151. Further, a record may be disclosed, subject to the limitations of 37 CFR 1.14, as a routine use, to the public if the record was filed in an application which became abandoned or in which the proceedings were terminated and which application is referenced by either a published application, an application open to public inspection or an issued patent.
9. A record from this system of records may be disclosed, as a routine use, to a Federal, State, or local law enforcement agency, if the USPTO becomes aware of a violation or potential violation of law or regulation.

(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 2 272 976 A1

(12) EUROPEAN PATENT APPLICATION

(43) Date of publication:
12.01.2011 Bulletin 2011/02

(51) Int Cl.:
C12Q 1/68 *(2006.01)*

(21) Application number: **09008808.9**

(22) Date of filing: **06.07.2009**

(84) Designated Contracting States:
AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV MC MK MT NL NO PL PT RO SE SI SK SM TR
Designated Extension States:
AL BA RS

(71) Applicant: **Max-Planck-Gesellschaft zur Förderung der Wissenschaften e.V.**
80539 München (DE)

(72) Inventors:
- **Soldatov, Aleksey**
  14195 Berlin (DE)
- **Borodina, Tatiana**
  14195 Berlin (DE)
- **Lehrach, Hans**
  14129 Berlin (DE)

(74) Representative: **Arth, Hans-Lothar**
ABK Patent Attorneys
Jasminweg 9
14052 Berlin (DE)

(54) Method for differentiation of polynucleotide strands

(57) Objective of the present invention is to provide a method for keeping of directional information in double-stranded DNA. We suggest to convert polynucleotide into a hybrid double-stranded DNA. One particular strand of this hybrid double-stranded DNA should be synthesised using at least one modified nucleotide. Thus, this particular strand would contain modified nucleotides along the whole length. Density of directional markers would not depend on the length of polynucleotides. Any internal fragments of the hybrid double-stranded DNA would have directional information. When it is necessary the modified strand may be easily degraded or separated from the other strand. It was found that such hybrid double-stranded DNA may be easily generated in a number of molecular biology tasks and may be used for molecular cloning, library preparation and strand separation.

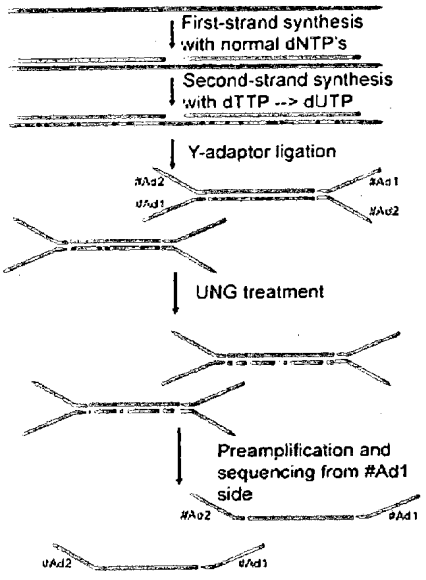


Figure 4

EP 2 272 976 A1

IDT Exhibit 1017.1737

Description

[0001] Cloning of single-stranded RNA and DNA molecules is much easier if they are transformed into double-stranded DNA. The invention relates to a novel method of preserving information about the direction of the original single-stranded molecules after being transformed into a double-stranded form.

[0002] Single-stranded RNA and DNA molecules play an important role in molecular biology. A number of viruses have single-stranded genomes. Most RNAs are present in single-stranded form. Recent studies have demonstrated an unexpected complexity of eukaryotic transcriptomes. A transcriptome is the set of all RNA transcripts produced in one or a population of cells. In addition to classical mRNAs (messenger ribonucleic acids) which cover approximately 1.5 percent of the genome in most higher eukaryotes, a large number of non-coding RNAs (ribonucleic acids) with widely varying expression levels have been identified. The biological functions of these novel transcripts are largely unknown and represent a new research area requiring high throughput transcriptome studies.

[0003] The polarity of the transcript is important for correct annotation of novel genes because it provides essential information about the possible function of a gene both on the RNA (structure and hybridization to other nucleic acid molecules) and the protein level. In addition, many genomic regions give rise to transcripts from both strands simultaneously. Antisense transcription is characteristic for eukaryotic genes and is thought to play an important regulatory role. Overlapping genes are common for compact genomes of prokaryotes and lower eukaryotes. Knowledge of transcript orientation helps to resolve the problem of colliding transcripts and to determine correctly gene expression levels in the presence of antisense transcription.

Background of the invention

[0004] From the art several methods are known that preserve directional information using distinguishable adapters for different ends of double-stranded DNA.

[0005] Adapters may be ligated directly to single-stranded RNA molecules.

[0006] Special composite primers may be used for first-strand cDNA synthesis: Herein the 5'- part of the primer provides a directional information, the 3'-part (oligo(dT)) or random nucleotides) is able to form a duplex with the RNA. Recently, first-strand cDNA synthesis was performed from a tagged random hexamer primer and second strand synthesis from a DNA-RNA template-switching primer.

[0007] Special homopolymeric adapters may be synthesised at the 3'-end of DNA molecules by Terminal Nucleotide Transferase.

[0008] DNA-RNA template-switching is used to attach a specific adapter to 3'-end of the DNA copy of the original single-stranded RNA molecule.

[0009] The main disadvantage of the adaptor-based approach it that only the ends of the clones preserve directional information. The problems are:

(i) strand-specific operations (directional immobilization on a solid phase; production of a single-stranded DNA for hybridization, etc.) should act on the adaptor sequence. For long clones the efficiency of such operations may be low. For example, strand separation where one of the strands contains a biotin molecule at the 5'-end is getting complicated if the length of the duplex is higher than 1 kb.
(ii) after fragmentation internal parts of the clones lose the directional information.

[0010] A special approach which is able to keep directional information in any part of the sequence was suggested recently: All cytidine residues in the RNA were converted to undines by bisulfite treatment prior to cDNA synthesis. The problem is that this approach is laborious and also leads to the loss of about 30% of uniquely matched sequencing reads because a part of the genome complexity is lost during the transformation from the four-bases into the three-bases code.

[0011] A lot of biologically important nucleic acids function in a single-stranded form where the polarity is extremely important. But for basic molecular cloning procedures single-stranded polynucleotides should be converted into double-stranded DNA molecules, because

* most of restriction enzymes work only on double-stranded DNA;
* ligation of double-stranded molecules is much simpler in comparison with the ligation of single-stranded molecules;
* the result of PCR is a double-stranded DNA.

[0012] At the moment most of the protocols keep directional information in the form of distinguishable adaptors on the ends of double-stranded molecules (Fig. 1B).

IDT Exhibit 1017.1738

Description of the invention

[0013] The objective of the present invention is to provide a method for keeping the directional information in double-stranded DNA. Surprisingly, it was found that this objective can be reached by the inventive method as disclosed by the description, the figures and the examples. A polynucleotide strand is converted into a hybrid double-stranded DNA. One particular strand of this hybrid double-stranded DNA should be synthesised using at least one modified nucleotide. Thus, this particular strand would contain modified nucleotides along the whole length of the strand. The density of directional markers would not depend on the length of hybrid double-stranded DNA. Any internal fragment of the hybrid would have directional information. If necessary, the modified strand may be easily degraded or separated from the other strand. It was found that such hybrid double-stranded DNA may be easily generated in a number of molecular biology tasks and may be used for molecular cloning, library preparation and strand separation.

[0014] The directional library construction method and the method of synthesis of single-stranded DNA according to the invention are easy to perform, inexpensive and highly reliable and thus overcome the shortcomings of the previously mentioned approaches.

[0015] The inventive method for preserving information about the direction of single-stranded nucleic acid molecules during molecular cloning operations with double-stranded DNA is provided. One of the strands of the double-stranded DNA is synthesized by using the other strand as a template and contains at least one modified dNTP. Thus, one strand of double-stranded DNA would contain unmodified nucleotides and the other strand at least one modified nucleotide along the whole length (Fig. 1C).

[0016] The new method of keeping of direction of polynucleotides is fully compatible with methods known from the art. Double-stranded DNA molecules may have simultaneously distinguishable primers and distinguishable strands (Fig. 1D). Fig. 2A shows how molecules with distinguishable strands may be converted into molecules with distinguishable primers. In an opposite operation the conversion of molecules with distinguishable primers into molecules with distinguishable strands is shown in Fig. 2B.

[0017] The main advantages compared to the adaptor-based method are:

(i) the density of modified nucleotides (directional markers) does not depend on the length of the double-stranded molecule. This is important for strand separation or strand-specific digestion of long DNA molecules.
(ii) any internal fragment of the clone carries directional information. After fragmentation internal parts of the clones do not lose the directional information.

[0018] Two examples demonstrate the convenience of the inventive method for differentiation of polynucleotide strands.

[0019] The method of synthesis of a single-stranded DNA is described in Fig. 3.

[0020] There is a number of methods of synthesis of single-stranded DNA relying on the difference between flanked adaptors:

* asymmetric PCR may be performed with the different amounts of PCR primers;
* the 5' end of one of the strands may contain thionucleotides which protect this strand from 5' to 3' exonucleases;
* protective thionucleotides may be introduced into the 3'-end part of only one strand to protect it from 3' to 5' exonucleases;
* the 5'-part of one of the strands may bear biotin for attachment of this strand to streptavidine-coated solid particles.

[0021] The problem is that all these methods proved to be unstable and show low efficiency for molecules larger than 1 kb.

[0022] Fig. 3 shows two possible ways for generating single-stranded DNA using the inventive method for differentiation of polynucleotide strands.

[0023] In a first approach (Fig. 3A) one strand of PCR product is substituted by a strand where dTTPs are substituted by dUTPs. After UNG treatment and APE 1 digestion (or alkalic hydrolisis) this strand is converted into a number of short single-stranded fragments which may be easily separated from the long unmodified strand by electrophoresis, anion-exchange chromatography, silica-based purification or gel-filtration.

[0024] In a second approach (Fig. 3B) one strand of the PCR product is substituted by a strand where a significant percentage of the dTTPs is substituted by biotin-associated dUTPs, respectively biotin-associated dTTPs. The concentration of biotin in a modified strand does not depend on the length of the double-stranded DNA, so the efficiency of the purification on streptavidine-coated particles should not depend on the length of the fragment.

[0025] The key steps of the method according to the invention include the generation of a cDNA strand including at least one modified nucleotide that in the course of the method is recognized by a selection factor which allows for a differentiation of the strand containing the at least one modified nucleotide from the strand containing the corresponding

IDT Exhibit 1017.1739

unmodified nucleotide or nucleotides.

[0026] The key steps of the inventive method are to be read in the following order:

I) Providing at least one polynucleotide template, primers, buffer solution, at least one enzyme with polymerase activity, sufficient amounts of the four dNTPs dATP, dCTP, dGTP and dTTP or analogues thereof, wherein at least one of the four dNTPs is replaced by a corresponding modified dNTP used for strand-specific differentiation;
II) Performing specific strand synthesis on said at least one polynucleotide; and
III) Performing strand-specific differentiation by using a selection factor for differentiating strands containing the at least one modified nucleotide from strands containing the corresponding unmodified nucleotide.

[0027] If the polynucleotide template is a single-stranded RNA a second cDNA strand has to be synthesized. For this purpose the at least one modified dNTP has to be removed from the solution and has to be replaced by the corresponding unmodified dNTP or dNTPs or analogues thereof. By the use of all four unmodified dNTPs a second cDNA strand is generated.

[0028] In this case the key steps according to the method read as follows:

I) Providing at least one polynucleotide template, primers, buffer solution, at least one enzyme with polymerase activity, sufficient amounts of the four dNTPs dATP, dCTP, dGTP and dTTP or analogues thereof, wherein at least one of the four dNTPs is replaced by a corresponding modified dNTP used for strand-specific differentiation;
II) Performing specific strand synthesis on said at least one polynucleotide;
III) Performing second-strand synthesis with all four unmodified dNTPs or analogues thereof after removing the at least one modified dNTP; and
IV) Performing strand-specific differentiation by using a selection factor for differentiating strands containing the at least one modified nucleotide from strands containing the corresponding unmodified nucleotide.

[0029] If the resulting cDNA shall undergo cloning, for example by an ensuing amplification step by PCR, it may be advantageous to carry out one or more preparative modifications on the strands to be cloned, such as shearing, restriction, digestion, end-polishing, A-tailing, G-tailing and adaptor ligation. These modifications may lead to a higher efficacy and unambiguous results in cloning.

[0030] With these additional modifications the key steps of the invention read as follows:

I) Providing at least one polynucleotide template, primers, buffer solution, at least one enzyme with polymerase activity, sufficient amounts of the four dNTPs dATP, dCTP, dGTP and dTTP or analogues thereof, wherein at least one of the four dNTPs is replaced by a corresponding modified dNTP used for strand-specific differentiation;
II) Performing specific strand synthesis on said at least one polynucleotide;
III) Performing one or more molecular cloning operations on the double-stranded DNA, wherein the molecular cloning operations are selected from shearing, restriction, digestion, end-polishing, A-tailing, G-tailing and adaptor ligation; and
IV) Performing strand-specific differentiation by using a selection factor for differentiating strands containing the at least one modified nucleotide from strands containing the corresponding unmodified nucleotide.

[0031] As said before, if the polynucleotide template is a single-stranded RNA a second-strand cDNA has to be synthesized after the exchange of the at least one modified dNTP by the corresponding unmodified dNTP or dNTPs. If this double-stranded cDNA shall undergo cloning it will be advantageous to perform one or more of the modifications for the molecular cloning operation as described before. By combining these conditions the key steps of the invention read as follows:

I) Providing at least one polynucleotide template, primers, buffer solution, at least one enzyme with polymerase activity, sufficient amounts of the four dNTPs dATP, dCTP, dGTP and dTTP or analogues thereof, wherein at least one of the four dNTPs is replaced by a corresponding modified dNTP used for strand-specific differentiation;
II) Performing specific strand synthesis on said at least one polynucleotide;
III) Performing second-strand synthesis with all four unmodified dNTPs or analogues thereof after removing the at least one modified dNTP;
IV) Performing one or more molecular cloning operations on the double-stranded DNA, wherein the molecular cloning operations are selected from shearing, restriction, digestion, end-polishing, A-tailing, G-tailing and adaptor ligation; and
V) Performing strand-specific differentiation by using a selection factor for differentiating strands containing the at least one modified nucleotide from strands containing the corresponding unmodified nucleotide.

IDT Exhibit 1017.1740

[0032] The initially provided at least one polynucleotide template may be a RNA such as mRNA, tRNA, rRNA, hnRNA, snRNA, snoRNA, scaRNA, asRNA, micro RNA, siRNA, short transcripts, synthetic polynucleotides and semisynthetic polynucleotides as well as physiological or externally induced modifications thereof, wherein mRNA is preferred.

[0033] mRNAs are protein-coding RNA. tRNAs are transfer RNAs responsible for providing correct amino acids in peptide (protein) synthesis in ribosomes. rRNAs are ribosomal RNAs, being a constitutive element of ribosomes. hnRNAs (heterogeneous nuclear RNA) is a precursor mRNA in the nucleus of eukaryotic cells. It is often named also pre-mRNA or precursor mRNA. snRNA (small nuclear RNA) is responsible in the eukaryotic nucleus for the processing of hnRNAs (spliceosomes). snoRNA (small nucleolar RNA) occurs in the nucleolus of eukaryotic cells. scaRNA (small Cajal body-specific RNA) is strongly related to snoRNA and occurs in Cajal bodies. asRNA (antisense RNA) plays an important role in regulating gene expression. microRNAs are involved in the regulation of cellular processes such as apoptosis and cell death. siRNA (small interfering RNA) is processed from cellular RNA into fragments with a length of ca. 22 base pairs by the enzyme Dicer and integrated into the enzyme complex RISC (RNA-induced silencing complex). RISC can thus regulate the expression of DNA segments or RNAs by switching off the respective coding region through analogous base pairing (silencing). This method by also used in gene therapeutic approaches. Retroviruses use RNA as the carrier of their genetic information.

[0034] Short transcripts refer to small RNA fragments with up to 20-40 base pairs.

[0035] Alternatively, the initially provided at least one polynucleotide template may be a DNA such as siDNA, DNA copies from retroviral or hepadnaviral RNA, single-stranded DNA, synthetic polynucleotides and semisynthetic polynucleotides as well as physiological or externally induced modifications thereof, wherein single-stranded DNA is preferred.

[0036] It shall be pointed out that in this invention the term *single-stranded DNA* can also refer to the first-strand cDNA derived from a RNA as template.

[0037] siDNA (small interfering DNA) is the corresponding DNA of siRNAs and mainly fulfils the same tasks in the cell. Single-stranded DNA is a single-stranded type of satellite DNA, occurring among others in certain soil-dwelling bacteria species.

[0038] According to the invention also double-stranded DNA can be the polynucleotide template. Specific primers and optionally adjuvants are needed to allow the generation of a cDNA strand against a particular strand of the double-stranded DNA (see below and Fig. 2B). Herein, the double-stranded DNA is destabilized in the region of hybridization of said specific primer for allowing annealing of the specific primer. Such a destabilization can be achieved for example by Uracil-N-Glycosylase digestion of deoxyuridines.

[0039] In a specific embodiment the double-stranded DNA template is a PCR product.

[0040] Sources for RNAs and DNAs can be eukaryotes (animals including humans and other mammals, plants, fungi), prokaryotes (bacteria, archaea), viruses, synthetic or semisynthetic constructs wherein these constructs or semisynthetic constructs may mirror at least partially a naturally occurring nucleic acids or may have a widely artificial sequence.

[0041] Said RNAs and DNAs can be isolated by well-known methods (Molecular Cloning: A Laboratory Manual, Third Edition. J. Sambrook, D. Russell).

[0042] The term *classic dNTPs,* respectively *unmodified dNTPS* refers to dATP (deoxyadenosine triphosphate), dCTP (deoxycytidine triphosphate), dGTP (deoxyguanosine triphosphate) and dTTP (deoxythymidine triphosphate). These four dNTPs (deoxynucleoside triphosphates) are usually used in nature for the generation of DNA strands.

[0043] It shall be pointed out that the plural form *classic dNTPs* refers to the total of dATP, dCTP, dGTP and dTTP. A referral in a quantifying sense to one or more of these classic dNTPs is not intended.

[0044] The term *analogues of dNTPs,* respectively *dNTPs and analogues thereof* refers to any molecule suitable for replacing a dNTP in presence and behavioural characteristics in a respective polynucleotide-containing solution and particularly in its matching characteristics to be able to undergo a base pair matching identical or similar to the classic dNTP it replaces. In particular, such analogues of dNTPs should be able to build the same or similar selective hydrogen bonds that allow for a selective pairing of base pairs (2 hydrogen bonds between A/T, T/A, A/U or U/A, and 3 hydrogen bonds between G/C and C/G), thus allowing for a correct translation of a polynucleotide sequence into its antisense sequence. Furthermore, it should be satirically similar to its corresponding classic dNTP so that it allows integration into the base pair sequence. Not only the nucleotide can be varied in such an analogue, but also the saccharide (usually a ribose or deoxyribose) or the phosphate backbone. For example, the phosphate backbone may be partially or entirely substituted by a sulphonated backbone. Also combinations of such variations are subsumed under this term.

[0045] The quantitative term *sufficient amount* herein refers to any amount, respectively concentration of an agent which allows for a qualitative and quantitative implementation of the method according to the invention in respect of the expected amounts which shall be needed for this purpose. Usually, the respective agents, in particular dNTPs, are added in exceeding amounts in order to ensure that all possible copies can be effected qualitatively and quantitatively.

[0046] The term *template* refers in molecular biology to a DNA or RNA strand which sets the base pair sequence to which a corresponding strand with an antisense sequence shall by polymerized. In a broader sense it may also refer to a RNA encoding a sequence for translation of this sequence into an amino acid sequence such as a peptide or a protein in ribosomes.

IDT Exhibit 1017.1741

[0047] The term *cDNA* stands for complementary deoxyribonucleic acid. It refers to a DNA which is generated as a copy of a RNA (usually by means of Reverse Transcriptase) or from a single DNA strand (usually by means of a DNA polymerase). Subsequently, a second cDNA strand may be generated by using the first cDNA strand as a template (usually by means of a DNA polymerase).

[0048] Enzymes with reverse transcriptase (RNA-dependent DNA polymerase) activity are able to transcribe a RNA sequence into a corresponding cDNA sequence. This group of enzymes includes among others HIV-1 reverse transcriptase from human immunodeficiency virus type 1 (PDB 1HMV), M-MLV reverse transcriptase from the Moloney murine leukemia virus, AMV reverse transcriptase from avian myeloblastosis virus and telomerase reverse transcriptase fulfilling a role in the maintenance of telomeres of eukaryotic chromosomes as well as commercially available forms such as "M-MLV reverse transcriptase RNase H Minus", "SuperScript® II", "SuperScript® III" or Tth polymerase. Among this group M-MLV reverse transcriptase RNase H Minus, SuperScript® II and SuperScript® III polymerases are preferred.

[0049] The terms *primers* and *random primers* refer to oligonucleotides which shall serve as starting point for DNA-replicating enzymes such as DNA polymerases. They can consist of a few base pairs, but in selected cases can reach up to 2.000 base pairs. Preferred are 5 to 50 base pairs. DNA polymerases need a hydroxy group for starting transcription. This hydroxy group is provided by the primer with its 3' hydroxy end. Often, the DNA sequence to be investigated is known. Hence suitable primers can be synthesized precisely according to the antisense sequence of aforesaid polynucleotide sequence. If this polynucleotide sequence is not or only incompletely known so-called random primers can be used. In such a method a large number of primers with a random base pair sequence are provided. One or more of these random primers statistically anneal to the polynucleotide sequence that shall be transcribed. Thus corresponding transcripts of different length can be obtained, depending where the random primer anneals to the polynucleotide sequence of interest. Alternatively, also composite primers with a random 3' part can be used. A further variation corresponding to the inventive method is the use of primers which are complementary to a homo-oligomeric tail synthesized by terminal transferase on the polynucleotide which serves as a template.

[0050] If the polynucleotide template is a RNA with a polyA tail it is preferred that said primers are oligo(dT), oligo(dU) or oligo(U) primers or composite primers with oligo(dT), oligo(dU) or oligo(U) region on 3' part. Alternatively, specific primers complementary to said single-stranded RNA template can be used.

[0051] If a double-stranded DNA template is used it is likewise possible to mark one particular strand of the double-stranded DNA with at least one modified nucleotide. Herein specific primers complementary to a particular strand of said double-stranded DNA can be used. In this case RecA protein may facilitate the hybridization of said primers to the double-stranded DNA template. Specific primers may appear as a result of introducing a nick in a particular strand of said double-stranded DNA template. Such nicks can for example result by the use of nicking endonuclease. Alternatively, such nicks may result from the digestion of the at least one modified nucleotide in a particular strand of said double-stranded DNA template.

[0052] In a specific embodiment the double-stranded DNA template is a PCR product. In this case, the recognition site of a nicking exonuclease or at least one modified nucleotide is located in one of the PCR primers. Herein it is preferred that said modified nucleotide is deoxyuridine.

[0053] If a double-stranded DNA template is used it is preferred that said polymerase is a DNA-dependent DNA polymerase with 5' to 3' exonuclease activity and said strand-specific synthesis on such a double-stranded DNA template is effected by means of a nick-translation.

[0054] In analogy, the polynucleotide template may be also a DNA::RNA hybrid. Hybrid herein means that a DNA strand is matching with a RNA strand having a corresponding antisense sequence. For example, but not limiting, it is possible that such a DNA::RNA hybrid is the product of a first-strand synthesis with unmodified nucleotides generated on a single-strand RNA as a template. For generating a cDNA of the DNA strand specific primers complementary to a particular DNA region of said double-stranded DNA::RNA hybrid have to be used. It is preferred that said primers are complementary to a homo-oligomeric tail synthesized by terminal transferase. In another preferred embodiment this specific primer is a template-switch primer. Such template-switch primers bind to the extra random nucleotides attached to the 3' end of the newly synthesized cDNA strand by using SuperScript® II enzyme (Invitrogen). This allows for full-length reverse transcription of the RNA strand. If the cDNA strand shall be generated against the RNA strand of the DNA::RNA hybrid said polymerase is DNA-dependent DNA polymerase with 5' to 3' exonuclease activity.

[0055] If the cDNA strand shall be generated against the DNA strand of the DNA::RNA hybrid the specific primers may appear as a result of introducing a nick into the RNA strand of said DNA::RNA hybrid. Such a nick may be set by using the RNA-specific enzyme RNase H. In a preferred embodiment RNase H, DNA polymerase I and E.coli ligase are used simultaneously. DNA ligases are enzymes able to ligate DNA strands by building an ester bond between the phosphate residue and the saccharide deoxyribose in the backbone of the DNA strands.

[0056] For second-strand cDNA synthesis using the first strand cDNA as a template an enzyme with DNA polymerase activity is needed. Also if the original polynucleotide is a single-stranded DNA and only one cDNA strand is generated such an enzyme with DNA polymerase activity is needed. DNA-dependent DNA polymerases are usually enzymes which generate a DNA copy strand from a DNA template. This group includes DNA polymerase I, II, III holoenzyme and

IDT Exhibit 1017.1742

IV, Taq polymerase, Tth polymerase, Pfu polymerase, wherein DNA polymerase I and Taq polymerase are preferred. This enzyme with polymerase activity does not have to be the same as used optionally for amplifying the remaining intact cDNA.

[0057] If the cDNA strand shall be generated against the RNA strand of the DNA::RNA hybrid the procedure is similar to that for generating a cDNA strand against one strand of a single-stranded RNA.

[0058] The term *modified dNTP* refers to the any molecule suitable for substituting exactly one corresponding classic dNTP in presence and behavioural characteristics in a respective polynucleotide-containing solution and particularly in its matching characteristics to be able to undergo a base pair matching identical or similar to the classic dNTP it replaces. Additionally, this modified dNTP must be suitable either for a specific degradation in which selectively this modified dNTP is degraded by a suitable degrading agent, thus rendering the DNA strand containing at least one of this degraded dNTP essentially unfit for amplification and/or hybridization, or it must render the DNA strand containing this modified dNTP eligible for specific removal of this DNA strand from the solution, or at least allowing a local and thus functional separation of polynucleotide strands. Such a removal respectively separation can be achieved by molecules or particles binding selectively to such exchange dNTPs, thus selectively only to one polynucleotide strand. Removal can occur by purification, adsorption to suitable structures or surfaces, binding to suitable selective structures such as antibodies, by selective degradation through certain chemicals, heat or radiation, or by centrifugation, magnetic or electrophoretic methods, isoelectric focusing. Also any kind of immobilization can be perceived.

[0059] Different purification methods may be used for removal of modified and unmodified dNTPs from the solution: electrophoretic separation, salt-alcohol precipitation, *anion*-exchange chromatography, silica purification, gel filtration. Purification on silica and gel-filtration are preferred.

[0060] Degradation respectively degradable herein means that a deoxynucleotide is specifically modified by addition of at least one enzymatic, chemical, heat, radiation (X-ray, UV, microwave) or optical methods such as lasers or combinations thereof in such a way that the resulting deoxynucleotide does not serve anymore for hybridization, cloning or as a suitable template in an amplification system such as PCR.

[0061] Analogously, the plural form *modified dNTPs* refers to the total of the respective modified dNTPs. A referral in a quantifying sense to one or more of these modified dNTPs is not intended.

[0062] The term *at least one modified dNTP* refers to the option that not only one dNTP is exchanged. According to the invention either one dNTP such as dATP, dCTP, dGTP or dTTP, or two dNTPs can be exchanged in the combinations dATP + dCTP, dATP + dGTP, dATP + dTTP, dCTP + dGTP, dCTP + dTTP or dGTP + dTTP, or three dNTPs can be exchanged in the combinations dATP + dCTP + dGTP, dATP + dCTP + dTTP, dATP + dGTP + dTTP or dCTP + dGTP + dTTP, or all four dNTPs are exchanged in the combination dATP + dCTP + dGTP + dTTP. It is understood that also analogues of these dNTPs as defined before shall be subsumed under the respective dNTPs. Also possible quantitative combinations thereof shall be subsumed under this combinational scheme.

[0063] There are several principles by which a selection factor can differentiate between a strand marked with at least one modified nucleotide and a strand with the corresponding unmodified nucleotide or nucleotides. For example, the at least one modified nucleotide has a chemically active group that can be modified after strand-specific synthesis. Another possibility is that the at least one nucleotide carries an affinity group and the selection factor binds selectively to this affinity group. Such am affinity group may for example be recognized by an antibody. A further possibility is that the at least one modified nucleotide carries a group which allows for selective recognition and binding or attachment by another agent which then enables selective removal of the such recognized strand. A further selective factor may be a different sensitivity to specific nucleases. Also the incorporation of ribonucleotides (NTPs) instead of deoxyribonucleotides (dNTPs) is suitable for enabling a differentiation between two strands.

[0064] The at least one modified nucleotide may display a different sensitivity to a nuclease, in comparison with the corresponding unmodified nucleotide or nucleotides. Thus such nuclease degrades specifically the at least one modified nucleotide while the unmodified nucleotides are not or only insignificantly affected by such nuclease action. Alternatively, such nuclease degrades specifically the unmodified nucleotides. In both cases the strand containing the such degraded nucleotide or nucleotides has become unfit for amplification or hybridization. Thus both strands can be differentiated.

[0065] In a preferred embodiment the differentiation of strands of polynucleotides is carried out with the modified dNTP being dUTP (deoxyuridine triphosphate). Herein the modified dNTP is dUTP (deoxyuridine triphosphate) as a substitute for dTTP. dUTP displays the same matching characteristics as dTTP, i.e. it is able to match with an opposed adenosine, but not with opposed guanosines, cytosines, thymidines or uridines. dUTP can be specifically degraded by the nuclease Uracil-N-Glycosylase (UNG). UNG is neutral towards dTTP and the other dNTPs or analogues thereof. This enzyme degrades uracil specifically by cleaving the N-glycosylic bond resulting in an abasic site. The such degraded DNA strand is unfit for undergoing appropriate amplification. At temperatures of over 50°C in slightly alkali conditions nicks occur in the DNA strand at the abasic positions. Additionally, Phusion DNA polymerase preferentially used for amplification (see below) is strongly suppressed by deoxyuridine-containing templates (Hogrefe, H.H., Hansen, C.J., Scott, B.R. and Nielsen K.B. 2002. Archaeal dUTPase enhances PCR amplifications with archaeal DNA polymerases by preventing dUTP incorporation. Proc Natl Acad Sci USA 99: 596-601). Consequently, only the DNA strand that does not contain

IDT Exhibit 1017.1743

dUTP can be amplified properly.

[0066] The combination of dUTP as exchange dNTP and UNG as the degrading agent of the exchange dNTP is a preferred embodiment. It is also preferred to replace only dTTP.

[0067] In another preferred embodiment the at least one modified nucleotide is thio_modified and said nuclease is an exonuclease that degrades thio_bonds with low efficiency. Such exonucleases sensitive to thio_bonds can be lambda-exonuclease, Exo III nuclease, T4 polymerase and T7 exonuclease.

[0068] Other preferred dNTP-specific degradation systems include the use of specific biotinylated dNTPs. Herein NHS esters of biotin are reacted with free amino groups of the respective nucleotide. These biotinylated nucleotides can be recognized by polymerase enzymes as eligible for integration into the actually formed strand. The matching properties to their counter nucleotide on the other strand are comparable. The biotinylated strand might be removed using streptavidin- or avidin-coated magnetic particles. Such specifically attached magnetic particles can be removed by applying a suitable magnetic field.

[0069] In another preferred embodiment the at least one modified polynucleotide is an Aminoallyl nucleotide (5-(3-aminoallyl)-nucleotide, aa-dNTP). Herein the nucleotide has an aminoallyl group attached to carbon 5 of the pyrimidine ring of uracil and cytosine. This aminoallyl group reacts with dyes from the cyanine series such as HiLyte Fluor® or Alexa Fluor® dyes that also been modified to be amine-reactive. As these dyes use to react also with standard buffer mediums a carbonate buffer is typically used. Herein it is preferred that the modified nucleotide is Aminoallyl-dUTP.

[0070] In a further preferred embodiment the modified nucleotide can be selectively recognized by an antibody. The antibody binds specifically to this modified nucleotide and the strands are differentiated by binding or immobilizing said antibody. It is further preferred that such antibody recognizes specifically the modified nucleotide 5-bromodeoxyuridine. It is also further preferred that such antibody or antibodies recognize specifically at least one modified nucleotide that has been modified by biotinylation.

[0071] If the strand carrying at least one modified nucleotide which carries a group which interacts which an affinity factor it is preferred that such strand is immobilized on a solid phase.

[0072] In a further preferred embodiment the at least one modified nucleotide is a ribonucleotide. Then it is preferred that said polymerase is a DNA polymerase displaying enhanced ability to incorporate modified substrates such as ribonucleotides (for example, Therminator™ II DNA Polymerase). The such incorporated at least one ribonucleotide is then specifically degraded by a specific nuclease, in this case a RNase.

[0073] If at least one classic dNTP is replaced by the corresponding modified dNTP(s) to generate a modified DNA strand the ratio of the at least one modified dNTP shall amount to at least 5% of the corresponding classic dNTP in the solution. Preferred are 20% ratio, more preferred 30%, more preferred 40%, more preferred 50%, more preferred 60%, more preferred 70%, more preferred 80% and particularly preferred more then 90%. Theoretically, only one degraded nucleotide is enough to render the degraded strand unfit for appropriate amplification. This requires that the DNA is long enough that at least one modified dNTP is integrated into the DNA strand. This is a matter of probability (when an even distribution of all four nucleotides is ideally assumed for the original polynucleotides) according to the distribution of the corresponding classic dNTPs and exchange dNTPs. With a length of at least 60 base pairs the probability is close to nil that not at least one modified dNTP has to be integrated into the second-strand cDNA. In order to ensure that also cDNAs with a length less than 60 base pairs always include at least one degradable nucleotide it is preferred that a sufficiently large percentage of the respective nucleoside triphosphate is modified.

[0074] If an amplification step is carried out in any of the embodiments of this invention an enzyme with polymerase activity is needed which is able to use the remaining intact cDNA strand as a template. In theory, any such enzyme with polymerase activity is suitable for this step. However, those enzymes that are used in conventional PCR protocols are preferred. They should grant a high reliability and a high throughput under conventional PCR conditions.

[0075] The term PCR refers to polymerase chain reaction. This is a general term comprising all protocols, conditions, automatized and half-automatized devices in which at least one DNA template is amplified by default by means of a enzyme with polymerase activity. In general, such molecules with polymerase activity are enzymes such as Taq DNA polymerase, Pfu DNA polymerase, Phusion DNA polymerase, wherein Taq and Phusion DNA Polymerases are_preferred.

[0076] It is particularly preferred that the at least one modified nucleotide is dUTP and said amplification is a Pfu DNA polymerase-dependent amplification.

[0077] If an amplification step is carried out in any of the embodiments of this invention a useful PCR (polymerase chain reaction) protocol shall be used by means of which the remaining intact DNA strand shall be amplified. Suitable PCR protocols of the include the known protocols for Standard PCR, Allele-Specific PCR, Alu PCR, Assembly PCR, Asymmetric PCR, Helicase Dependent Amplification HDA, Hot Start PCR, Inverse PCR, Intersequence-Specific PCR (ISSR PCR), Late-PCR, Ligation-mediated PCR, Long PCR, Methylation-Specific PCR, Miniprimer PCR, Multiplex PCR, Nested PCR, Overlap-Extension PCR, Quantitative PCR, Real Time PCR, Solid Phase PCR, Thermal Assymetric Interlaced PCR (TAIL-PCR), Touchdown PCR or combinations thereof. Preferred is the use of Standard PCR. Particularly preferred is the use of dUTP as the at least one modified nucleotide for amplification of the remaining intact strand with

IDT Exhibit 1017.1744

Standard PCR. Equally particularly preferred is the use of ribonucleotides as the least one modified nucleotide for amplification of the remaining intact strand with Standard PCR.

[0078] Herein the term "at least one intact DNA strand" means that the number of the intact DNA strands to be amplified can vary from one to many billions of different polynucleotides.

[0079] In a further optional step the resulting sequences can be aligned with already known sequence patterns of the corresponding DNA, if applicable. Such results can be plotted in a way such as shown in Fig. 6.

[0080] Optionally, the method according to the invention comprises the step of contig construction from the obtained sequences. Herein the sequences of different overlapping (contiguous) DNA fragments derived from the same source can be used for (re-)constructing the original respectively complete DNA sequence of this source.

[0081] Alternatively to the amplification step the at least one intact DNA strand can be hybridized to a nucleic acid structure displaying an antisense sequence of the at least one intact DNA strand. Suitable hybridization methods include *in situ* hybridization (hybridization of a DNA to a complementary sequence in isolated cells and/or tissues), Southern Blot (a method for verifying the presence of a DNA sequence in a DNA sample by separating the DNAs in the sample by size using gel electrophoresis and bringing said sample to a filter membrane for probe hybridization), Northern Blot (the equivalent of Southern Blot with RNA samples) and Southwestern Blot (identifying the presence of a DNA sequence in a DNA sample prepared according to the Southern blot method by hybridizing the sample on a nitrocellulose gel with characterizing DNA-binding proteins) microarray hybridization and correspondent "reverse" hybridizations were intact DNA's are loaded on a membrane and hybridised with particular probes.

[0082] When an amplification is intended the key steps of the inventive method are to be read in the following order:

I) Providing at least one polynucleotide template, primers, buffer solution, at least one enzyme with polymerase activity, sufficient amounts of the four dNTPs dATP, dCTP, dGTP and dTTP or analogues thereof, wherein at least one of the four dNTPs is replaced by a corresponding modified dNTP used for strand-specific differentiation;
II) Performing specific strand synthesis on said at least one polynucleotide;
III) Performing one or more molecular cloning operations on the double-stranded DNA, wherein the molecular cloning operations are selected from shearing, restriction, digestion, end-polishing, A-tailing, G-tailing and adaptor ligation; and
IV) Performing strand-specific differentiation by using a selection factor for differentiating strands containing the at least one modified nucleotide from strands containing the corresponding unmodified nucleotide.
V) Amplifying the at least one remaining intact cDNA strand by performing PCR;

[0083] In the particularly preferred embodiment 1 of the method according to the invention a strand specific cloning of RNAs such as a whole transcriptome of a cell or a tissue shall be carried out. Therefore, at least one single strand RNA is provided, for example by isolation from the respective cell or tissue. The four dNTPs dATP, dCTP, dGTP and dTTP are added in sufficient amounts to the solution containing the at least one RNA strand. Then, a first cDNA strand is generated by using random primers (or oligo(dT), oligo (dU), oligo (U) primers) for annealing to said at least one RNA strand and an enzyme with Reverse Transcriptase activity. In the following at least one of the aforesaid dNTPs is replaced by at least one corresponding modified dNTP. For example, the modified dNTP may be a dNTP which can be specifically degraded by adding a degrading agent in order that the strand containing the such degraded at least one modified dNTP is not suitable anymore for amplification, or the modified dNTP may expose the such marked cDNA strand for being removed by interaction with specific agents interacting with the additional functional group or groups on the at least one modified dNTP. With the four dNTPs including the at least one modified dNTP second strand cDNA synthesis is performed by using the first cDNA strand as a template and a suitable enzyme with polymerase activity. In the next step polarity specific adaptors are ligated to the ends of cDNA strands, allowing a polymerase to dock to the adaptor structure for PCR amplification. After the ligation the second cDNA strand containing the at least one modified dNTP is for example degraded by degrading the at least one modified dNTP with a suitable degrading agent, or the at least one second cDNA strand containing the at least one modified dNTP is separated from the at least one first cDNA strand by special molecules or particles attaching to characteristic structures on the modified dNTP. Finally, the remaining intact first cDNA strand is amplified by PCR using a suitable enzyme with polymerase activity.

[0084] Thus the steps of this embodiment of the inventive method can be summarized in the following schematic order:

1) Providing at least one single-stranded RNA molecule;
2) Adding sufficient amounts of the four dNTPs dATP, dCTP, dGTP and dTTP or corresponding analogues thereof;
3) Generating a first cDNA strand;
4) Replacing at least one dNTP with at least one corresponding modified dNTP;
5) Performing second-strand synthesis on the first cDNA strand;
6) Ligating polarity-specific adaptors to one or both ends of the cDNA double strand;
7) Specifically modifying or removing the second cDNA strand containing the at least one modified dNTP; and

IDT Exhibit 1017.1745

8) Amplifying the at least one remaining intact cDNA strand by performing PCR.

[0085] In embodiment 2 strand specific cloning of RNA can be achieved also by varying the steps of the inventive method. In this embodiment the first cDNA strand shall contain the at least one modified dNTP. In the replacement step this at least one modified dNTP is substituted by the corresponding classic dNTP. Thus in this embodiment the at least one first cDNA strand will be either degraded or removed. The amplification step is then performed on the remaining intact second cDNA strand. All other steps are adjusted correspondingly.

[0086] The steps of this embodiment of the inventive method can be summarized in the following schematic order:

1) Providing at least one single-stranded RNA molecule;
2) Adding sufficient amounts of the four dNTPs dATP, dCTP, dGTP and dTTP or corresponding analogues thereof, wherein at least one of the four dNTPs is a modified dNTP;
3) Generating a first cDNA strand;
4) Replacing the at least one modified dNTP with the corresponding dNTPs or analogues of the corresponding dNTPs;
5) Performing second-strand synthesis on the first cDNA strand;
6) Ligating polarity-specific adaptors to one or both ends of the cDNA double strand;
7) Specifically modifying or removing the first cDNA strand containing the at least one modified dNTP; and
8) Amplifying the at least one remaining intact cDNA strand by performing PCR.

[0087] In embodiment 3 strand specific differentiation according to the invention can also be achieved using single-stranded DNA as a template. Here at least one DNA single strand is provided. A cDNA strand is synthesized wherein at least one modified dNTP is integrated. In this embodiment only one cDNA strand is synthesized since a second cDNA strand would be identical to the original DNA single strand. Finally, the original single-stranded DNA shall be amplified. Thus all transcripts resulting from the PCR will bear the same reading direction of the sequence as the original single-stranded DNA. The other steps are adjusted accordingly.

[0088] This leads to the following schematic order:

1) Providing at least one single-stranded DNA molecule;
2) Adding sufficient amounts of the four dNTPs dATP, dCTP, dGTP and dTTP or corresponding analogues thereof, wherein at least one of the four dNTPs is a modified dNTP used for the strand specific differentiation;
3) Generating a cDNA strand;
4) Ligating polarity-specific adaptors to one or both ends of the resulting DNA double strand;
5) Specifically modifying or removing the cDNA strand containing the at least one modified dNTP; and
6) Amplifying the at least one intact DNA strand by performing PCR.

[0089] In embodiment 4 the purpose of the inventive method is not an amplification but a hybridization. This hybridization can occur against any structure carrying a corresponding antisense sequence to the at least one intact first cDNA strand (see below). For a hybridization without amplification no adaptors have to be ligated. Consequently, this step is omitted in this embodiment and the amplification step is replaced by a hybridization step. Thus the order of the steps of the inventive method reads as follows:

1) Providing at least one single-stranded RNA molecule;
2) Adding sufficient amounts of the four dNTPs dATP, dCTP, dGTP and dTTP or corresponding analogues thereof;
3) Generating a first cDNA strand;
4) Replacing at least one dNTP with at least one corresponding modified dNTP used for the strand specific differentiation;
5) Performing second-strand synthesis on the first cDNA strand;
6) Specifically modifying or removing the second cDNA strand containing the at least one modified dNTP; and
7) Hybridizing the at least one remaining intact cDNA strand to a suitable antisense DNA sequence.

[0090] Embodiment 5 varies embodiment 4 in so far that herein the first cDNA strand contains the at least one modified dNTP. Thus the remaining intact second cDNA strand is hybridized. Consequently, the sequence against which the remaining intact second cDNA strand is hybridized must display an antisense sequence of the second strand. The steps of the inventive method are adjusted according to embodiments 2 and 4 and read as follows:

1) Providing at least one single-stranded RNA molecule;
2) Adding sufficient amounts of the four dNTPs dATP, dCTP, dGTP and dTTP or corresponding analogues thereof, wherein at least one of the four dNTPs is a modified dNTP used for the strand specific differentiation;

IDT Exhibit 1017.1746

3) Generating a first cDNA strand;
4) Replacing the at least one modified dNTP with the corresponding dNTPs or analogues of the corresponding dNTPs;
5) Performing second-strand synthesis on the first cDNA strand;
6) Specifically modifying or removing the first cDNA strand containing the at least one modified dNTP; and
7) Hybridizing the at least one remaining intact cDNA strand to a suitable antisense DNA sequence.

[0091] In embodiment 6 the hybridization step follows after an amplification step. In various hybridization modes it may be advantageous to hybridize not only one or a few copies of the respective intact cDNA strand, but a large number as resulting from an amplification. Thus the steps of ligating a polarity-specific adaptor and of an amplification have to be integrated again. The order of the steps reads like embodiment 1 and a subsequent hybridization step:

1) Providing at least one single-stranded RNA molecule;
2) Adding sufficient amounts of the four dNTPs dATP, dCTP, dGTP and dTTP or corresponding analogues thereof;
3) Generating a first cDNA strand;
4) Replacing at least one dNTP with at least one corresponding modified dNTP;
5) Performing second-strand synthesis on the first cDNA strand;
6) Ligating polarity-specific adaptors to one or both ends of the cDNA double strand;
7) Specifically modifying or removing the second cDNA strand containing the at least one modified dNTP; and
8) Amplifying the at least one remaining intact cDNA strand by performing PCR.
9) Hybridizing the amplified cDNA to a suitable antisense DNA sequence.

[0092] Correspondingly, in embodiment 7 the hybridization step can ensue after performing an amplification of the intact second-strand cDNA, as in embodiment 2:

1) Providing at least one single-stranded RNA molecule;
2) Adding sufficient amounts of the four dNTPs dATP, dCTP, dGTP and dTTP or corresponding analogues thereof, wherein at least one of the four dNTPs is a modified dNTP;
3) Generating a first cDNA strand;
4) Replacing the at least one modified dNTP with the corresponding dNTPs or analogues of the corresponding dNTPs;
5) Performing second-strand synthesis on the first cDNA strand;
6) Ligating polarity-specific adaptors to one or both ends of the cDNA double strand;
7) Specifically modifying or removing the first cDNA strand containing the at least one modified dNTP; and
8) Amplifying the at least one remaining intact cDNA strand by performing PCR.
9) Hybridizing the amplified cDNA to a suitable antisense DNA sequence.

[0093] Also in the case of single-stranded DNA it may be advantageous to combine an amplification and a hybridization step. Thus in embodiment 8 the order of the steps reads as follows:

1) Providing at least one single-stranded DNA molecule;
2) Adding sufficient amounts of the four dNTPs dATP, dCTP, dGTP and dTTP or corresponding analogues thereof, wherein at least one of the four dNTPs is a modified dNTP used for the strand specific differentiation;
3) Generating a cDNA strand;
4) Ligating polarity-specific adaptors to one or both ends of the resulting DNA double strand;
5) Specifically modifying or removing the cDNA strand containing the at least one modified dNTP; and
6) Amplifying the at least one intact DNA strand by performing PCR.
7) Hybridizing the amplified DNA to a suitable antisense DNA sequence.

[0094] Embodiment 9: If the sequence of the original RNA is not known it may be desirable to perform sequencing. Thus such a step is added to the step order of embodiment 1:

1) Providing at least one single-stranded RNA molecule;
2) Adding sufficient amounts of the four dNTPs dATP, dCTP, dGTP and dTTP or corresponding analogues thereof;
3) Generating a first cDNA strand;
4) Replacing at least one dNTP with at least one corresponding modified dNTP;
5) Performing second-strand synthesis on the first cDNA strand;
6) Ligating polarity-specific adaptors to one or both ends of the cDNA double strand;
7) Specifically modifying or removing the second cDNA strand containing the at least one modified dNTP; and
8) Optionally, amplifying the at least one remaining intact cDNA strand by performing PCR.

IDT Exhibit 1017.1747

9) Sequencing the resulting copies of the at least one intact cDNA strand.

[0095] Embodiment 10: Consequently, it is also possible to include a sequencing step if the first-stranded included the at least one modified dNTP and the intact second-strand cDNA is amplified, as in embodiment 2:

1) Providing at least one single-stranded RNA molecule;
2) Adding sufficient amounts of the four dNTPs dATP, dCTP, dGTP and dTTP or corresponding analogues thereof, wherein at least one of the four dNTPs is a modified dNTP;
3) Generating a first cDNA strand;
4) Replacing the at least one modified dNTP with the corresponding dNTPs or analogues of the corresponding dNTPs;
5) Performing second-strand synthesis on the first cDNA strand;
6) Ligating polarity-specific adaptors to one or both ends of the cDNA double strand;
7) Specifically modifying or removing the first cDNA strand containing the at least one modified dNTP; and
8) Optionally, amplifying the at least one remaining intact cDNA strand by performing PCR.
9) Sequencing the resulting copies of the at least one intact cDNA strand.

[0096] Embodiment 11: After sequencing it may be desirable to determine the polarity of the resulting sequence(s). This may serve for several analytic or diagnostic purposes (see below) or for aligning RNA transcripts to gene sequences known from DNA sequencing projects. If the remaining intact first strand cDNA was sequenced the order of the steps of the inventive method read as follows:

1) Providing at least one single-stranded RNA molecule;
2) Adding sufficient amounts of the four dNTPs dATP, dCTP, dGTP and dTTP or corresponding analogues thereof;
3) Generating a first cDNA strand;
4) Replacing at least one dNTP with at least one corresponding modified dNTP;
5) Performing second-strand synthesis on the first cDNA strand;
6) Ligating polarity-specific adaptors to one or both ends of the cDNA double strand;
7) Specifically modifying or removing the second cDNA strand containing the at least one modified dNTP; and
8) Optionally, amplifying the at least one remaining intact cDNA strand by performing PCR.
9) Sequencing the resulting copies of the at least one intact cDNA strand.
10) Aligning of sequencing reads to the reference genome or reconstructing of the RNA sequence from individual sequencing reads.
11) Determining the polarity of the at least one RNA transcript.

[0097] Embodiment 12: Consequently, such a determination of the polarity and optionally an alignment to a gene structure known from a sequencing project can also be performed if the remaining intact second strand cDNA was amplified and sequenced. As in this case the antisense sequence is resulting the sequence has to be inverted, preferably by means of a computer programme, into the correct reading sequence.

[0098] Thus the order of the steps of the inventive method is:

1) Providing at least one single-stranded RNA molecule;
2) Adding sufficient amounts of the four dNTPs dATP, dCTP, dGTP and dTTP or corresponding analogues thereof, wherein at least one of the four dNTPs is a modified dNTP;
3) Generating a first cDNA strand;
4) Replacing the at least one modified dNTP with the corresponding dNTPs or analogues of the corresponding dNTPs;
5) Performing second-strand synthesis on the first cDNA strand;
6) Ligating polarity-specific adaptors to one or both ends of the cDNA double strand;
7) Specifically modifying or removing the first cDNA strand containing the at least one modified dNTP; and
8) Optionally, amplifying the at least one remaining intact cDNA strand by performing PCR;
9) Sequencing the resulting copies of the at least one intact cDNA strand;
10) Aligning of sequencing reads to the reference genome or reconstructing of the RNA sequence from individual sequencing reads;
11) Determining the polarity of the at least one RNA transcript

[0099] Embodiment 13: In analogy, it may be also desirable to sequence the amplified copies of a single-stranded DNA if the sequence isn't known yet. Thus a sequencing step ensues to the order of the steps of the inventive method of embodiment 3:

IDT Exhibit 1017.1748

1) Providing at least one single-stranded DNA molecule;
2) Adding sufficient amounts of the four dNTPs dATP, dCTP, dGTP and dTTP or corresponding analogues thereof, wherein at least one of the four dNTPs is a modified dNTP used for the strand specific differentiation;
3) Generating a cDNA strand;
4) Ligating polarity-specific adaptors to one or both ends of the resulting DNA double strand;
5) Specifically modifying or removing the cDNA strand containing the at least one modified dNTP; and
6) Optionally, amplifying the at least one intact DNA strand by performing PCR.
7) Sequencing the resulting copies of the at least one intact DNA strand.

[0100] Embodiment 14: Also for single-stranded DNA it may be desirable to determine the polarity of the resulting copies after amplification and sequencing. As for single-stranded DNA by needs only one cDNA strand is generated including the at least one modified dNTP the resulting sequence must be inverted, as in embodiment 12.

1) Providing at least one single-stranded DNA molecule;
2) Adding sufficient amounts of the four dNTPs dATP, dCTP, dGTP and dTTP or corresponding analogues thereof, wherein at least one of the four dNTPs is a modified dNTP used for the strand specific differentiation;
3) Generating a cDNA strand;
4) Ligating polarity-specific adaptors to one or both ends of the resulting DNA double strand;
5) Specifically modifying or removing the cDNA strand containing the at least one modified dNTP; and
6) Optionally, amplifying the at least one intact DNA strand by performing PCR;
7) Sequencing the resulting copies of the at least one intact DNA strand;
8) Aligning of sequencing reads to the reference genome or reconstructing of the DNA sequence from individual sequencing reads;
9) Determining the polarity of the at least one DNA molecule.

[0101] In another preferred embodiment 15 the original polynucleotide strand is a double-stranded DNA. Herein it may be necessary to set a nick into the particular strand of the double-stranded DNA which shall be substituted. A possible sequence of the steps according to this embodiment is:

1) Providing at least one double-stranded DNA;
2) Setting a nick in one particular strand of the double-stranded DNA by using a nicking endonuclease;
3) Adding sufficient amounts of the four dNTPs dATP, dCTP, dGTP and dTTP or corresponding analogues thereof, wherein at least one of the four dNTPs is a modified dNTP used for the strand specific differentiation;
4) Performing nick-translation of the nicked DNA strand by means of a DNA polymerase with 5' to 3' exonuclease activity;
5) Specifically modifying or removing the cDNA strand containing the at least one modified dNTP.

[0102] In a further preferred embodiment 16 the original polynucleotide strand is a double-stranded DNA obtained as a result of PCR reaction. A special modified nucleotide is introduced into one PCR primer to set a nick into the particular strand of the double-stranded DNA which shall be substituted. A possible sequence of the steps according to this embodiment is:

1) Providing at least one double-stranded DNA obtained in PCR reaction where one of PCR-primers contains dUTP;
2) Setting a nick in one particular strand of the double-stranded DNA by UNG treatment and APE 1 digestion or UNG treatment and alkalic hydrolisis;
3) Adding sufficient amounts of the four dNTPs dATP, dCTP, dGTP and dTTP or corresponding analogues thereof, wherein at least one of the four dNTPs is a modified dNTP used for the strand specific differentiation;
4) Performing nick-translation of the nicked DNA strand by means of a DNA polymerase with 5' to 3' exonuclease activity;
5) Specifically modifying or removing the cDNA strand containing the at least one modified dNTP.

[0103] It is obvious to a person skilled in the art that the different steps mentioned before in the respective embodiments can be combined for solving similar or different tasks provided the order and the selection of the different steps results in a meaningful order of the steps. All these possible combinations shall fall under the scope and the spirit of the invention.

[0104] By way of example, the method according to the invention was applied on the yeast transcriptome (Example 1). In a similar fashion, it was applied on the transcriptome of the mouse brain (Example 2). These two examples were chosen because comparative data from the Applicants' laboratory and other laboratories are available. Also complete genomes of these organisms are published. By way of comparing the results from conventional RNA-Seq and the

IDT Exhibit 1017.1749

inventive method according to the invention the advantages of the inventive method become apparent. As shown e.g. in Example 1 and Fig. 6, the inventive method allows a differentiation in sense and antisense transcripts. In this plot sense transcripts are marked in blue and antisense transcripts in red. Thus the method according to the invention can use the maintained information of the polarity of a transcript to generate more precise gene expression maps. Practical applications are discussed further below.

[0105] The reliability of the method according to the invention was validated in several control experiments. First, it was tested to which degree the inventive method can reproduce the results of conventional RNA sequencing. Herein the amplified sense and antisense transcripts according to the invention were pooled for allowing a correlation analysis. The results such as shown in Example 3 and Fig. 7A show a high correlation between the two methods. A high correlation corresponds to a distribution close to unity. A low correlation would be represented by a widely scattered distribution pattern. Thus the inventive method does not disturb the overall expression and transcriptome pattern in a sample. So the results of both methods can be compared to one another.

[0106] Next, two different strand-specific expression profilings were performed according to the invention on the same sample. Herein, only sense transcripts were included. The results were plotted again as described before. An excellent reproducibility of the inventive method is given, as can be seen in Example 4 and Fig. 7B.

[0107] In a further control experiment, the combination of sense and antisense transcripts obtained were compared to the results from another laboratory. There is still a very high correlation between the results, although with a somewhat broader scatter pattern (see Example 4 and Fig. 7C). Such a variation nearly always occurs when comparing the results from different laboratories. This experiment shows that this method does not work only under the specific conditions of the Applicants' laboratory, but can be universally used.

[0108] In order to test whether sense and antisense transcripts obtained according to the inventive method can be appropriately differentiated sense and antisense transcripts were compared and depicted in such an aforesaid scatter plot. As can be seen in Fig. 7D there is nearly no correlation. Thus sense and antisense transcripts obtained according to the inventive method are unequivocally different. This is a proof that the inventive method is highly suitable for differentiating accurately between sense and antisense transcripts.

[0109] The stability of the inventive method in respect of a complete exchange of at least one dNTP by at least one corresponding degradable analogue was tested. For this purpose in one reaction dTTP was exchanged by 100% dUTP. In another reaction dTTP was exchanged only by 75% dUTP, with 25% dTTP remaining in the reaction batch. The results from both experiments were compared for the sense transcripts and antisense transcripts (see Example 5 and Fig. 8). There is no significant deterioration in correlation when the exchange of the at least one dNTP was incomplete. When extrapolating these results it appears that only a part of the at least one dNTP must be replaced by the at least one corresponding degradable analogue to keep the method going with acceptable results, as shown in the correlation coefficient.

[0110] Actinomycin D is a substance which selectively inhibits DNA synthesis from DNA templates, but not from RNA templates. In a further experiment it could be shown that the presence of actinomycin D does not deteriorate significantly the results from the inventive protocol, neither for sense nor for antisense transcripts (see Example 6 and Fig. 9).

[0111] Practical applications of the inventive method include but are not limited to the provision of suitable kits for performing the inventive method according to the different protocols (embodiments) disclosed above.

[0112] Fig. 10 presents examples where information about transcription orientation helps to resolve overlaps of annotated and new genes in both yeast and mouse transcriptomes.

[0113] A kit in molecular biology is a package which includes all necessary ingredients for performing a certain method or singular step. Standard chemicals as present in any standard molecular biology laboratory are normally not included. Nevertheless some of these standard chemicals may be indispensable to carry out the inventive method properly. It is understood that all ingredients are provided in quantities that allow for a proper execution of the inventive method for the majority of scientific, diagnostic and industrial applications.

[0114] Often, but not always, these ingredients are provided in already prepared solutions ready- or close to ready-for-use. There may be also combinations of different ingredients already added together. A further advantage is that such kits use to be verified. Therefore the operator doesn't have to prove again the viability of the method and can save on at least some control experiments. Therefore kits are a very popular tool in molecular biology laboratories in research, diagnostics and industry.

[0115] For carrying out the key steps of the method according to the invention such a kit shall include the following components:

A) All four dNTPs dATP, dCTP, dGTP and dTTP or corresponding analogues thereof, additionally at least one modified dNTP used for the strand specific differentiation;
B) Primers
C) One or more DNA polymerases;
D) One or more degrading agents corresponding to the modified dNTPs and/or one or more removing agents apt

IDT Exhibit 1017.1750

for removing the strand containing the modified dNTPs.

[0116] If the protocol of the inventive method that shall be carried out includes an amplification step another embodiment of the kit should include:

A) All four dNTPs dATP, dCTP, dGTP and dTTP or corresponding analogues thereof, additionally at least one modified dNTP used for the strand specific differentiation;
B) Primers
C) One or more DNA polymerases;
D) One or more degrading agents corresponding to the modified dNTPs and/or one or more removing agents apt for removing the strand containing the modified dNTPs.
E) One or more polarity-specific adaptor.

[0117] If the protocol of the inventive method that shall be carried out includes an amplification step and the original polynucleotide is a RNA another embodiment of the kit should include:

A) All four dNTPs dATP, dCTP, dGTP and dTTP or corresponding analogues thereof, additionally at least one modified dNTP used for the strand specific differentiation;
B) Primers
C) One or more DNA polymerases;
D) One or more degrading agents corresponding to the modified dNTPs and/or one or more removing agents apt for removing the strand containing the modified dNTPs.
E) One or more polarity-specific adaptor;
F) One or more Reverse transcriptases.

[0118] In any of the aforegoing kits according to the invention it may be preferable to include additionally components for removal of unmodified nucleotides by gel-filtration or purification on silica-column. Thus the kits should include the following:

A) All four dNTPs dATP, dCTP, dGTP and dTTP or corresponding analogues thereof, additionally at least one modified dNTP used for the strand specific differentiation;
B) Primers
C) One or more DNA polymerases;
D) One or more degrading agents corresponding to the modified dNTPs and/or one or more removing agents apt for removing the strand containing the modified dNTPs;
E) Components for specific removal of at least one modified or at least one unmodified nucleotide by gel-filtration or purification on silica-column.

or

A) All four dNTPs dATP, dCTP, dGTP and dTTP or corresponding analogues thereof, additionally at least one modified dNTP used for the strand specific differentiation;
B) Primers
C) One or more DNA polymerases;
D) One or more degrading agents corresponding to the modified dNTPs and/or one or more removing agents apt for removing the strand containing the modified dNTPs.
E) One or more polarity-specific adaptor.
F) Components for specific removal of at least one modified or at least one unmodified nucleotide by gel-filtration or purification on silica-column.

or

A) All four dNTPs dATP, dCTP, dGTP and dTTP or corresponding analogues thereof, additionally at least one modified dNTP used for the strand specific differentiation;
B) Primers
C) One or more DNA polymerases;
D) One or more degrading agents corresponding to the modified dNTPs and/or one or more removing agents apt for removing the strand containing the modified dNTPs.

IDT Exhibit 1017.1751

E) One or more polarity-specific adaptor;
F) One or more Reverse transcriptases.
G) Components for specific removal of at least one modified or at least one unmodified nucleotide by gel-filtration or purification on silica-column.

[0119] Another application is the use of the results of this method for improved quantitative and qualitative diagnosis tests in selected diseases, as a part of the treatment or for prophylactic purposes. In general every disease is eligible which includes a temporary or permanent change in RNA expression levels. It is also possible to obtain a better monitoring of the effects of pharmaceutical agents used for therapy and/or prophylaxis of diseases provided these pharmaceutical agents produce a characteristic expression pattern different from that in untreated patients (humans or animals). Also improved gene expression patterns of plant, fungi, bacteria and viruses can be obtained by this method when the polarity of transcripts is available.

[0120] A further application is the production of improved and more selective microarrays. When the polarity of transcripts is known only the transcripts (respectively the antisense sequence) with a correct reading direction are meaningfully included in the microarray so that the limited space on the microarray is more economically used. Respectively, microarrays used for the generation of characteristic patterns instead of individual results may obtain a higher resolution by omitting non-characteristic transcripts of the incorrect reading direction.

[0121] As shown before, the advantages of the procedure according to the invention are providing a simple, inexpensive and reliable method for strand-specific cDNA cloning. This method is a novel and easy approach for determining the polarity of transcripts and using this information for purposes as outlined above.

[0122] Moreover, the method according to the invention has a higher accuracy in measuring expression levels of genes compared to RNA-Seq. Ignoring orientation of transcripts leads for example to overestimation of the expression level for about 10% of the mouse and about 15% of the yeast genes with noticeable expression level.

Figures

[0123]

Fig. 1: Directional double-stranded DNA
(A) original single-stranded polynucleotide; (B) directional double- stranded DNA with distinguishable adaptors (state of the art); (C) directional double-stranded DNA with distinguishable strands; (D) directional double-stranded DNA with distinguishable strands and distinguishable adapters.

Fig. 2A: Conversion of distinguishable strands into distinguishable adaptors
(A1) Y-shaped adaptors are ligated to double-stranded DNA with distinguishable strands; modified strands are separated from unmodified ones; after synthesis of complementary strands or PCR- amplification double-stranded DNA with distinguishable adapters is generated.
(A2) directional adaptors are ligated to double-stranded DNA with distinguishable strands; modified strands are separated from unmodified ones; after PCR-amplification double-stranded DNA with distinguishable adapters is generated.

Fig. 2B: Conversion of distinguishable adaptors into distinguishable strands
(B1) nick introduced into one of the adaptors using nicking endonuclease;
(B2) dUTP in one of the adaptors is converted into a nick: Uracil-N- Glycosylase converts dUTP into an abasic site and APE 1 nuclease converts the abasic site into a nick;
(B3) double-stranded DNA is destabilized in one of the Uracil-N- Glycosylase; new primer is annealed in the destabilized region;
(B4) position for annealing of the primer is organized by action 5' to 3' exonuclease; another adapter is protected from action of 5' to 3' exonuclease by thio-nucleotides; nick-translation is performed by DNA polymerase I (or Taq polymerase) in presence of modified nucleotides.

Fig. 3A: Preparation of single-stranded DNA according to the inventive method using dUTP as a modified nucleotide
* double-stranded DNA is a product of PCR reaction where one of the PCR-primers has dUTP at 3'-part;
* dUTP in the PCR primer is converted into a nick: Uracil-N-Glycosylase converts dUTP into an abasic site and APE 1 nuclease converts the abasic site into a nick;
* nick-translation is performed by DNA polymerase I (or Taq polymerase) in presence of dUTP;
* dUTP-containing strand is destroyed by Uracil-N-Glycosylase digestion and alkali hydrolysis;
* single-stranded DNA is purified by silica-based purification or gel- filtration or electrophoretic purification.

IDT Exhibit 1017.1752

Fig. 3B: Preparation of single-stranded DNA according to the inventive method using biotinilated nucleotides
* the double-stranded DNA is a product of PCR reaction where one of the PCR-primers has dUTP at 3'-part;
* dUTP in the PCR primer is converted into a nick: Uracil-N-Glycosylase converts dUTP into an abasic site and APE 1 nuclease converts the abasic site into a nick;
* nick-translation is performed by DNA polymerase I (or Taq polymerase) in presence of biotin-associated dTTPs and/or biotin- associated dUTPs;
* strands are separated using streptavidine-coated particles;

Fig. 4: Flowchart of the inventive method
RNA is shown in red colour, DNA in green. Arrows are in 5' to 3' direction.

Fig. 5: Schematic view of the inventive method on the level of base pairs

Fig. 6: Expression profile of a yeast gene
(A) A conventional transcriptional profile of the yeast YGR203W gene without orientation is shown. Vertical lines mark the boundaries of the YGR203W gene as they were determined previously
(B) A transcriptional profile of the yeast YGR203W gene with orientation according to the inventive method is depicted. Reads mapped in the forward direction are shown in blue, in the reverse direction in red.

Fig. 7: Scatter plots comparing mouse mRNA expression data (number of reads in annotated genes)
(A) The same mouse liver sample, strand-specific (according to the method, X-axis) and strand-unspecific (RNA-Seq, Y-axis) protocols (Pearson correlation coefficient (cc) = 0.999). A distribution close to unity would corresponds to a high correlation in these scatter plots.
(B) results according to the method for two biological replicas (mouse whole brain mRNA); cc=0.990.
(C) Mouse whole brain expression data according to the method (X- axis) and data from Mortazavi et al. 2008 (Y-axis); cc=0.817.
(D) Sense (X-axis) and antisense (Y-axis) expression in mouse brain. Thus the inventive method allows for a good differentiation of sense and antisense transcripts.

Fig. 8: Comparison of ssRNA-Seq results for protocols using 75%/25% dUTP/dTTP mixture and 100% dUTP
(A) Scatter plot of levels of sense transcription.
(B) Scatter plot of levels of antisense transcription.
Still a good correlation can be reached. Thus the inventive method does not depend on a complete exchange of at least one dNTP by a degradable analogue.

Fig. 9: Comparison of ssRNA-Seq results for protocols with and without actinomycin D added to the reverse transcription reaction
(A) Scatter plot of levels of sense transcription.
(B) Scatter plot of levels of antisense transcription.
X-Axe: with actinomycin D; Y-Axe: without actinomycin D. The use of actinomycin D does not bias significantly the correlation when using sense strands.

Fig. 10: Examples of overlapping genes and novel gene-candidates
A. *S. cerevisiae* transcriptome. Overlapping tails of two genes:
B. *S. cerevisiae* transcriptome. Overlapping of three genes:
C. *S. cerevisiae* transcriptome. Novel gene-candidate between
D. Mouse whole brain transcriptome. Overlapping genes: Ncaph2 and Ecgf1.
E. Mouse whole brain transcriptome. genes: Mrpl24 and BC023814.
F. Mouse whole brain transcriptome. One unannotated and one incorrectly annotated exon of Cdc42bpa mouse gene.
G. Mouse whole brain transcriptome. Novel exon of Chd3 gene.
H. Mouse whole brain transcriptome. Novel gene-candidate.

ssRNA-Seq stands for single-strand RNA sequencing carried out according to the method.

IDT Exhibit 1017.1753

Examples

Example 1: Transcriptome of Saccharomyces cerevisiae

[0124] In this example an application of the method according to the invention in a preferred embodiment is presented. Strand-specific sequencing of the yeast transcriptome is performed. Subsequently, the sequence of each clone is determined and matched with already existing DNA sequence of the yeast genome.

RNA isolation:

[0125] Yeast strain BY4741 (MATa; *his3Δ1; leu2Δ0; met15Δ0; ura3Δ0*) was grown in rich medium (YPD; BD Company) at 30°C overnight, diluted to an OD_{600} of 0.15 and grown until reaching an OD_{600} of 0.87. The cells were harvested by centrifugation at room temperature, washed once with 1xPBS, and frozen in liquid nitrogen. Total RNA was extracted using the RiboPure™-Yeast kit (Ambion) and analyzed by an Agilent 2100 bioanalyzer (Agilent Technologies).

polyA+ RNA purification:

[0126] polyA+ RNA was purified with the Dynabeads mRNA purification kit (Invitrogen) following manufacturer's instructions and treated for 30 min at 37°C with 0.2 u TURBO™ DNase (Ambion) per 1 μg of RNA.

First strand synthesis (FSS):

[0127] A typical FSS reaction was prepared by mixing 0.5 μg of polyA+ RNA, 40 ng of $(dN)_6$ primers (Invitrogen) and 25 pmol of oligo(dT) primer (Invitrogen) in 8.5 μl of 1x reverse transcription buffer (Invitrogen), 0.5 mM dNTP's, 5 mM $MgCl_2$ and 10 mM DTT. The mixture was incubated at 98°C for 1 min to melt RNA secondary structures, then at 70°C for 5 min and was ramped to 15°C at 0.1°C/sec. Slow temperature ramp was used to make secondary structures of RNA and primers annealing as reproducible as possible. At 15°C 0.5 μl of actinomycin D solution (120 ng/μl), 0.5 μl of RNase OUT (40 u/μl, Invitrogen) and 0.5 μl of Superscript III polymerase (200 u/μl, Invitrogen) were added to the reaction. Reverse transcription reaction temperature was increased gradually as a compromise between survival of the enzyme, stability of the primers and denaturation of RNA secondary structures: ramp from 15 to 25°C at 0.1 °C/sec; incubation at 25°C for 10 min; ramp from 25 to 42°C at 0.1 °C/sec; incubation at 42°C for 45 min; ramp from 42 to 50°C at 0.1°C/sec; incubation at 50°C for 25 min. Superscript III polymerase was finally inactivated at 75°C for 15 min.

Removal of dNTPs:

[0128] 20 μl of EB (10 mM TrisCl, pH 8.5, Qiagen) was added to the reaction. dNTPs were removed by purification of the first strand mixture on a self-made 200 μl G-50 gel-filtration spin-column equilibrated with 1 mM TrisCl, pH 7.0.

Second strand synthesis (SSS):

[0129] The Invitrogen kit for second strand synthesis was used. Therefore the first strand synthesis buffer was restored after gel-filtration column. It is possible to prepare self-made buffer and reduce the number of buffers. Water was added to the purified FSS reaction to bring the final volume to 52.5 μl. The mixture was cooled on ice. Then 22.5 μl of the "second strand mixture" (1 μl of 10 x reverse transcription buffer (Invitrogen); 0.5 μl of 100 mM $MgCl_2$; 1 μl of 0.1 M DTT; 2 μl of 10 mM mixture of each: dATP, dGTP, dCTP, dUTP; 15 μl of 5x SSS buffer (Invitrogen); 0.5 μl of E.coli ligase (10 u/μl, NEB); 2 μl of DNA polymerase I (10 u/μl, NEB) and 0.5 μl RNase H (2u/μl, Invitrogen)) were added. SSS reaction was incubated at 16°C for 2 hours. ds cDNA was purified on QIAquick columns (Qiagen) according to the manufacturer's instructions.

DNA fragmentation:

[0130] About 250 ng of double-stranded cDNA was fragmented by sonication on a UTR200 (Hielscher Ultrasonics GmbH, Germany) under the following conditions: 1 hour, 50% pulse, 100% power, and continuous cooling by 0°C water flow-through.

Preparation of libraries for Illumina sequencing platform:

[0131] Libraries were prepared using the DNA sample kit (#FC-102-1002, Illumina), as described in (Sultan et al.

IDT Exhibit 1017.1754

2008), but with the following modifications: just before library amplification uridine digestion was performed at 37°C for 15 min in 5 μl of 1 x TE buffer, pH 7.5 with 1 u UNG (Applied Biosystems).

[0132] Procedure of paired-end sequencing library preparation was the same except that other ligation adapters and PCR primers were used (#PE-102-1002, Illumina).

Sequencing:

[0133] Amplified material was loaded onto flowcell at 4 pM concentration. Sequencing was carried out on the Illumina 1 G Genome Analyser by running 36 cycles according to the manufacturer's instructions.

Data analysis:

[0134] Images deconvolution, quality values calculation, mapping of exonic reads and exon junctions were performed as described in the art. Sequencing reads were aligned to *S. cerevisiae* (UCSC sacCerl) genomes using modification of Eland software (Gerald module v.1.27, Illumina). The mapping criteria of Eland are the following: sequencing read should be uniquely match to the genome allowing up to two mismatches, without insertions or deletions. The following recursive modification of Eland procedure was applied: first 32 bp of reads (trimming the last 4 bp of 36 bp reads due to Eland limitations) were aligned, then reads which do not match on reference by Eland criteria were trimmed to 31 bp, and aligned again. This 3'-end trimming of unmatched reads was done recursively down to 25 bp length. The modified procedure increases gain of uniquely aligned fragments typically by 20-50%, because the sequencing errors, which prevent successful alignment by Eland criteria, are located mostly on reads ends, which are gradually trimmed. Under these conditions, ~60% of the reads obtained here are matched to unique locations of the reference genome, whereas ~25% of the reads map to more than one genomic position and about 15% do not map.

Mapping end tags:

[0135] Unmapped sequencing reads with 1-11 nt long leading oligo(dT) stretches were used for mapping of the 3'-gene boundaries. Leading oligo(dT) stretches were removed, and if the remaining fragment length was longer or equal to 25 bp, it was aligned on reference genome.

Search for novel genes:

[0136] The search for novel genes in yeast using RNA-Seq was previously performed. Such an algorithm was not used herein, since it was not accurately described in the art. To estimate roughly how the knowledge of RNA orientation influences the number of novel gene-candidates the following procedure was used. A search algorithm presented below gives 377 novel gene candidates without taking into account RNA orientation. It is about the same number as in the art. To use information about RNA orientation the search was performed twice (for reads mapped in the forward direction and then for reads mapped in the reverse direction), yielding in 549 novel gene candidates in forward orientation and 512 - in reverse. The whole genome was split into 50 bp windows (non-sliding). A "new gene" was defined as a joined group of more than 2 consecutive windows, with at least 2 sequence reads (of the same direction) mapped per window. The gap between "new genes" should be at least 50 bp, the gap between "new gene" and an annotated gene (with the same transcription direction as the "new gene") - at least 100 bp.

[0137] With the RNA-Seq procedure it is possible to determine the 3'-boundaries of genes using those sequencing reads, which overlap with 3'-borders of genes. These reads may be mapped to the reference genome only after removal of oligo(dT) tail. The inventive method has the advantage of reduced noise in comparison with the RNA-Seq protocol because only one orientation of the homopolymeric stretch is allowed.

Example 2: Transcriptome of the brain of *Mus musculus*

[0138] In this example a further application of the method according to the invention is presented. Strand-specific sequencing of the whole mouse brain transcriptome is performed. Subsequently, the sequence of each clone is determined and matched with already existing DNA sequence of the mouse genome. Methods are similar to those used in Example 1. Therefore only steps which differ from Example 1 are indicated.

RNA isolation:

[0139] Two eleven weeks old female mice (C57B1/6J) were dissected and the whole brain was taken for RNA preparation. Total RNA was extracted using the Trizol method.

IDT Exhibit 1017.1755

Data analysis:

[0140] The resulting sequencing reads were aligned to the *M. musculus* (UCSC mm9) genomes using modification of Eland software (Gerald module v.1.27, Illumina).

Example 3: Comparison of strand-specific and non-strand-specific amplification of polynucleotides

[0141] To demonstrate that second-strand synthesis with deoxyuridine does not disturb the transcriptional landscape both strand-specific (according to the inventive method) and non-strand-specific transcriptome (RNA-Seq) sequencing for the same RNA sample was performed (mouse liver transcriptome). The resulting scatter plot (Fig. 7A) shows that both RNA-Seq and the inventive method produce identical transcription patterns.

Example 4: Reproducibiliity of the inventive method and comparison with previously published results from another laboratory

[0142] In order to test the reproducibility of the inventive method independent analyses of whole brain mouse (Fig. 7B) transcriptomes were carried out under identical conditions. A high reproducibility can be seen in the respective scatter plots. A comparison of the data with previously published results also demonstrates good correlation (Fig. 7C).

Example 5: Stability of the inventive method

[0143] The inventive method results in a high degree of certainty in identifying transcript polarity. In theory, one uridine base in a molecule of a sequencing library is enough to prevent the appearance of a sequencing read from the false strand. Even if UNG occasionally does not remove the uridine base, the molecule still would not be amplified, since uridine-containing template strongly suppresses Phusion DNA polymerase used for library amplification. Apparently, most second cDNA strands in the library contain more than one uridine base, because a 75%/25% dUTP/dTTP mixture in the second strand synthesis reaction gives essentially the same results as 100% dUTP (Fig. 8A and 8B).

Example 6: Influence of actinomycin D

[0144] To prevent spurious second-strand cDNA synthesis, which was shown to be the major source of artifactual antisense transcripts. Antisense artifacts in transcriptome microarray experiments are resolved by actinomycin D. Actinomycin D was included in the reverse transcription reaction. Actinomycin D specifically inhibits DNA-dependent, but not RNA-dependent, DNA synthesis. The presence of actinomycin D in the reaction practically did not influence the level of antisense transcription (Fig. 9A and 9B). Actinomycin D is included in the protocol to make the reaction more reproducible.

IDT Exhibit 1017.1756

SEQUENCE LISTING

<110> Max-Planck-Gesellschaft zur Förderung der Wissenschaften e.V.

<120> Method for differentiation of polynucleotide strands

<130> GAR-P03013WO

<140> 000000
<141> 2010-07-06

<160> 39

<170> PatentIn version 3.3

<210> 1
<211> 113
<212> RNA
<213> Artificial

<220>
<223> first-strand synthesis: primers annealing

<220>
<221> misc_feature
<222> (1)..(6)
<223> n is a, c, g, or u

<220>
<221> misc_feature
<222> (8)..(13)
<223> n is a, c, g, or u

<220>
<221> misc_feature
<222> (16)..(20)
<223> n is a, c, g, or u

<220>
<221> misc_feature
<222> (22)..(24)
<223> n is a, c, g, or u

<220>
<221> misc_feature
<222> (26)..(31)
<223> n is a, c, g, or u

<220>
<221> misc_feature
<222> (34)..(44)
<223> n is a, c, g, or u

<220>
<221> misc_feature
<222> (46)..(49)
<223> n is a, c, g, or u

<220>
<221> misc_feature
<222> (51)..(57)
<223> n is a, c, g, or u

<220>
<221> misc_feature
<222> (59)..(66)

IDT Exhibit 1017.1757

<223> n is a, c, g, or u

<220>
<221> misc_feature
<222> (70)..(73)
<223> n is a, c, g, or u

<220>
<221> misc_feature
<222> (75)..(76)
<223> n is a, c, g, or u

<220>
<221> misc_feature
<222> (78)..(79)
<223> n is a, c, g, or u

<400> 1
nnnnnnunnn nnnuunnnnn unnnunnnnn nuunnnnnnn nnnnunnnnu nnnnnnnunn 60

nnnnnnuuun nnnunnunna aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaa 113

<210> 2
<211> 6
<212> DNA
<213> Artificial

<220>
<223> Primer first-strand synthesis

<220>
<221> misc_feature
<222> (1)..(6)
<223> n is a, c, g, or t

<400> 2
nnnnnn 6

<210> 3
<211> 20
<212> DNA
<213> Artificial

<220>
<223> oligo (dT) first-strand synthesis

<400> 3
tttttttttt tttttttttt 20

<210> 4
<211> 42
<212> DNA
<213> Artificial

<220>
<223> first-strand synthesis: reverse transcription; cDNA starting from primer

<220>
<221> misc_feature
<222> (1)..(11)
<223> n is a, c, g, or t

IDT Exhibit 1017.1758

<220>
<221> misc_feature
<222> (14)..(19)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (21)..(23)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (25)..(29)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (32)..(37)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (39)..(42)
<223> n is a, c, g, or t

<400> 4
nnnnnnnnnn naannnnnna nnnannnnna annnnnnann nn 42

<210> 5
<211> 50
<212> DNA
<213> Artificial

<220>
<223> first-strand synthesis: reverse transcription; cDNA starting from oligo (dT)

<220>
<221> misc_feature
<222> (24)..(25)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (27)..(28)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (30)..(33)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (37)..(44)
<223> n is a, c, g, or t

<400> 5
ttttttttt ttttttttt tttnnannan nnnaaannnn nnnnaaaaaa 50

<210> 6
<211> 102
<212> DNA
<213> Artificial

IDT Exhibit 1017.1759

<220>
<223> Second-strand synthesis: Second Strand

<220>
<221> misc_feature
<222> (1)..(6)
<223> n is a, c, g, t or u

<220>
<221> misc_feature
<222> (8)..(13)
<223> n is a, c, g, t or u

<220>
<221> misc_feature
<222> (16)..(20)
<223> n is a, c, g, t or u

<220>
<221> misc_feature
<222> (22)..(24)
<223> n is a, c, g, t or u

<220>
<221> misc_feature
<222> (26)..(31)
<223> n is a, c, g, t or u

<220>
<221> misc_feature
<222> (34)..(44)
<223> n is a, c, g, t or u

<220>
<221> misc_feature
<222> (46)..(49)
<223> n is a, c, g, t or u

<220>
<221> misc_feature
<222> (51)..(57)
<223> n is a, c, g, t or u

<220>
<221> misc_feature
<222> (59)..(66)
<223> n is a, c, g, t or u

<220>
<221> misc_feature
<222> (70)..(73)
<223> n is a, c, g, t or u

<220>
<221> misc_feature
<222> (75)..(76)
<223> n is a, c, g, t or u

<220>
<221> misc_feature
<222> (78)..(79)
<223> n is a, c, g, t or u

<400> 6
nnnnnnunnn nnnuunnnnn unnnunnnnn nuunnnnnnn nnnnunnnnu nnnnnnnunn 60

nnnnnnuuun nnnunnunna aaaaaaaaaa aaaaaaaaaa aa 102

IDT Exhibit 1017.1760

<210> 7
<211> 36
<212> DNA
<213> Artificial

<220>
<223> Fragmentation (ultrasound digestion): Fragment A

<220>
<221> misc_feature
<222> (1)..(1)
<223> n is a, c, g, t or u

<220>
<221> misc_feature
<222> (3)..(8)
<223> n is a, c, g, t or u

<220>
<221> misc_feature
<222> (11)..(15)
<223> n is a, c, g, t or u

<220>
<221> misc_feature
<222> (17)..(19)
<223> n is a, c, g, t or u

<220>
<221> misc_feature
<222> (21)..(26)
<223> n is a, c, g, t or u

<220>
<221> misc_feature
<222> (29)..(36)
<223> n is a, c, g, t or u

<400> 7
nunnnnnnuu nnnnnunnnu nnnnnnuunn nnnnnn 36

<210> 8
<211> 47
<212> DNA
<213> Artificial

<220>
<223> Fragmentation (ultrasound digestion): Fragment B (second strand)

<220>
<221> misc_feature
<222> (1)..(2)
<223> n is a, c, g, t or u

<220>
<221> misc_feature
<222> (4)..(11)
<223> n is a, c, g, t or u

<220>
<221> misc_feature
<222> (15)..(18)
<223> n is a, c, g, t or u

IDT Exhibit 1017.1761

<220>
<221> misc_feature
<222> (20)..(21)
<223> n is a, c, g, t or u

<220>
<221> misc_feature
<222> (23)..(24)
<223> n is a, c, g, t or u

<400> 8
nnunnnnnnn nuuunnnnun nunnaaaaaa aaaaaaaaaa aaaaaaa 47

<210> 9
<211> 42
<212> DNA
<213> Artificial

<220>
<223> Fragmentation (ultrasound digestion) Fragment B (first strand)

<220>
<221> misc_feature
<222> (21)..(22)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (24)..(25)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (27)..(30)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (34)..(41)
<223> n is a, c, g, or t

<400> 9
ttttttttt ttttttttt nnannannnn aaannnnnnn na 42

<210> 10
<211> 39
<212> DNA
<213> Artificial

<220>
<223> End-repair Fragment A (second strand))

<220>
<221> misc_feature
<222> (1)..(1)
<223> n is a, c, g, t or u

<220>
<221> misc_feature
<222> (3)..(8)
<223> n is a, c, g, t or u

<220>

IDT Exhibit 1017.1762

<221> misc_feature
<222> (11)..(15)
<223> n is a, c, g, t or u

<220>
<221> misc_feature
<222> (17)..(19)
<223> n is a, c, g, t or u

<220>
<221> misc_feature
<222> (21)..(26)
<223> n is a, c, g, t or u

<220>
<221> misc_feature
<222> (29)..(39)
<223> n is a, c, g, t or u

<400> 10
nunnnnnnuu nnnnnunnnu nnnnnnuunn nnnnnnnnn 39

<210> 11
<211> 39
<212> DNA
<213> Artificial

<220>
<223> End-repair Fragment A first strand

<220>
<221> misc_feature
<222> (1)..(11)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (14)..(19)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (21)..(23)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (25)..(29)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (32)..(37)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (39)..(39)
<223> n is a, c, g, or t

<400> 11
nnnnnnnnnn naannnnnna nnnannnnna annnnnnan 39

<210> 12
<211> 44

IDT Exhibit 1017.1763

<212> DNA
<213> Artificial

<220>
<223> End-repair Fragment B second strand

<220>
<221> misc_feature
<222> (1)..(2)
<223> n is a, c, g, t or u

<220>
<221> misc_feature
<222> (4)..(11)
<223> n is a, c, g, t or u

<220>
<221> misc_feature
<222> (15)..(18)
<223> n is a, c, g, t or u

<220>
<221> misc_feature
<222> (20)..(21)
<223> n is a, c, g, t or u

<220>
<221> misc_feature
<222> (23)..(24)
<223> n is a, c, g, t or u

<400> 12
nnunnnnnnn nuuunnnnun nunnaaaaaa aaaaaaaaaa aaaa 44

<210> 13
<211> 44
<212> DNA
<213> Artificial

<220>
<223> End-repair Fragment B first strand

<220>
<221> misc_feature
<222> (21)..(22)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (24)..(25)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (27)..(30)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (34)..(41)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (43)..(44)

IDT Exhibit 1017.1764

<223> n is a, c, g, or t

<400> 13
ttttttttttt ttttttttttt nnannannnn aaannnnnnn nann 44

<210> 14
<211> 40
<212> DNA
<213> Artificial

<220>
<223> A-tailing Fragment A second strand

<220>
<221> misc_feature
<222> (1)..(1)
<223> n is a, c, g, t or u

<220>
<221> misc_feature
<222> (3)..(8)
<223> n is a, c, g, t or u

<220>
<221> misc_feature
<222> (11)..(15)
<223> n is a, c, g, t or u

<220>
<221> misc_feature
<222> (17)..(19)
<223> n is a, c, g, t or u

<220>
<221> misc_feature
<222> (21)..(26)
<223> n is a, c, g, t or u

<220>
<221> misc_feature
<222> (29)..(39)
<223> n is a, c, g, t or u

<400> 14
nunnnnnnuu nnnnnunnnu nnnnnnuunn nnnnnnnnna 40

<210> 15
<211> 40
<212> DNA
<213> Artificial

<220>
<223> A-tailing Fragment A first strand

<220>
<221> misc_feature
<222> (1)..(11)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (14)..(19)
<223> n is a, c, g, or t

IDT Exhibit 1017.1765

<220>
<221> misc_feature
<222> (21)..(23)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (25)..(29)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (32)..(37)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (39)..(39)
<223> n is a, c, g, or t

<400> 15
nnnnnnnnnn naannnnnna nnnannnnna annnnnnana 40

<210> 16
<211> 45
<212> DNA
<213> Artificial

<220>
<223> A-tailing Fragment B second strand

<220>
<221> misc_feature
<222> (1)..(2)
<223> n is a, c, g, t or u

<220>
<221> misc_feature
<222> (4)..(11)
<223> n is a, c, g, t or u

<220>
<221> misc_feature
<222> (15)..(18)
<223> n is a, c, g, t or u

<220>
<221> misc_feature
<222> (20)..(21)
<223> n is a, c, g, t or u

<220>
<221> misc_feature
<222> (23)..(24)
<223> n is a, c, g, t or u

<400> 16
nnunnnnnnn nuuunnnnun nunnaaaaaa aaaaaaaaaa aaaaa 45

<210> 17
<211> 45
<212> DNA
<213> Artificial

<220>

IDT Exhibit 1017.1766

<223> A-tailing Fragment B first strand

<220>
<221> misc_feature
<222> (21)..(22)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (24)..(25)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (27)..(30)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (34)..(41)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (43)..(44)
<223> n is a, c, g, or t

<400> 17
ttttttttttt ttttttttttt nnannannnn aaannnnnnn nanna 45

<210> 18
<211> 107
<212> DNA
<213> Artificial

<220>
<223> Adaptor ligation Fragment A second strand

<220>
<221> misc_feature
<222> (35)..(35)
<223> n is a, c, g, t or u

<220>
<221> misc_feature
<222> (37)..(42)
<223> n is a, c, g, t or u

<220>
<221> misc_feature
<222> (45)..(49)
<223> n is a, c, g, t or u

<220>
<221> misc_feature
<222> (51)..(53)
<223> n is a, c, g, t or u

<220>
<221> misc_feature
<222> (55)..(60)
<223> n is a, c, g, t or u

<220>
<221> misc_feature

IDT Exhibit 1017.1767

<222> (63)..(68)
<223> n is a, c, g, t or u

<220>
<221> misc_feature
<222> (70)..(73)
<223> n is a, c, g, t or u

<400> 18
tacactcttt ccctacacga cgctcttccg atctnunnnn nnuunnnnnu nnnunnnnnn 60

uunnnnnnun nnnagatcgg aagagctcgt atgccgtctt ctgcttg 107

<210> 19
<211> 106
<212> DNA
<213> Artificial

<220>
<223> Adaptor ligation Fragment A first strand

<220>
<221> misc_feature
<222> (35)..(45)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (48)..(53)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (55)..(56)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (58)..(62)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (65)..(70)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (72)..(72)
<223> n is a, c, g, or t

<400> 19
tacactcttt ccctacacga cgctcttccg atctnnnnnn nnnnnaannn nnnannannn 60

nnaannnnnn anagatcgga agagctcgta tgccgtcttc tgcttg 106

<210> 20
<211> 112
<212> DNA
<213> Artificial

<220>
<223> Adaptor ligation Fragment B second strand

IDT Exhibit 1017.1768

<220>
<221> misc_feature
<222> (35)..(36)
<223> n is a, c, g, t or u

<220>
<221> misc_feature
<222> (38)..(45)
<223> n is a, c, g, t or u

<220>
<221> misc_feature
<222> (49)..(52)
<223> n is a, c, g, t or u

<220>
<221> misc_feature
<222> (54)..(55)
<223> n is a, c, g, t or u

<220>
<221> misc_feature
<222> (57)..(58)
<223> n is a, c, g, t or u

<400> 20
tacactcttt ccctacacga cgctcttccg atctnnunnn nnnnnuuunn nnunnunnaa 60

aaaaaaaaaa aaaaaaaaag atcggaagag ctcgtatgcc gtcttctgct tg 112

<210> 21
<211> 112
<212> DNA
<213> Artificial

<220>
<223> Adaptor ligation Fragment B first strand

<220>
<221> misc_feature
<222> (55)..(56)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (58)..(59)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (61)..(64)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (68)..(75)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (77)..(78)
<223> n is a, c, g, or t

<400> 21
tacactcttt ccctacacga cgctcttccg atctttttttt ttttttttttt ttttnnanna 60

IDT Exhibit 1017.1769

nnnnaaannn nnnnnannag atcggaagag ctcgtatgcc gtcttctgct tg 112

<210> 22
<211> 35
<212> DNA
<213> Artificial

<220>
<223> UDGase treatment Fragment A second strand piece 1

<220>
<221> misc_feature
<222> (35)..(35)
<223> n is a, c, g, or t

<400> 22
tacactcttt ccctacacga cgctcttccg atctn 35

<210> 23
<211> 6
<212> DNA
<213> Artificial

<220>
<223> UDGase treatment fragment A second sttrand piece 2

<220>
<221> misc_feature
<222> (1)..(6)
<223> n is a, c, g, or t

<400> 23
nnnnnn 6

<210> 24
<211> 5
<212> DNA
<213> Artificial

<220>
<223> UDGase treatment fragment A second strand piece 3

<220>
<221> misc_feature
<222> (1)..(5)
<223> n is a, c, g, or t

<400> 24
nnnnn 5

<210> 25
<211> 3
<212> DNA
<213> Artificial

<220>
<223> UDGase treatment fragment A second strand piece 4

<220>
<221> misc_feature

IDT Exhibit 1017.1770

<222> (1)..(3)
<223> n is a, c, g, or t

<400> 25
nnn 3

<210> 26
<211> 6
<212> DNA
<213> Artificial

<220>
<223> UDGase treatment fragment A second strand piece 5

<220>
<221> misc_feature
<222> (1)..(6)
<223> n is a, c, g, or t

<400> 26
nnnnnn 6

<210> 27
<211> 44
<212> DNA
<213> Artificial

<220>
<223> UDGase treatment Fragment A second strand piece 6

<220>
<221> misc_feature
<222> (1)..(10)
<223> n is a, c, g, or t

<400> 27
nnnnnnnnnn agatcggaag agctcgtatg ccgtcttctg cttg 44

<210> 28
<211> 36
<212> DNA
<213> Artificial

<220>
<223> UDGase treatment fragment B second strand piece 1

<220>
<221> misc_feature
<222> (35)..(36)
<223> n is a, c, g, or t

<400> 28
tacactcttt ccctacacga cgctcttccg atctnn 36

<210> 29
<211> 8
<212> DNA
<213> Artificial

<220>
<223> UDGase treatment fragment B second strand piece 2

IDT Exhibit 1017.1771

<220>
<221> misc_feature
<222> (1)..(8)
<223> n is a, c, g, or t

<400> 29
nnnnnnnn 8

<210> 30
<211> 4
<212> DNA
<213> Artificial

<220>
<223> UDGase treatment fragment B second strand piece 3

<220>
<221> misc_feature
<222> (1)..(4)
<223> n is a, c, g, or t

<400> 30
nnnn 4

<210> 31
<211> 2
<212> DNA
<213> Artificial

<220>
<223> UDGase treatment fragment B second strand piece 4

<220>
<221> misc_feature
<222> (1)..(2)
<223> n is a, c, g, or t

<400> 31
nn 2

<210> 32
<211> 56
<212> DNA
<213> Artificial

<220>
<223> UDGase treatment fragment B second strand piece 5

<220>
<221> misc_feature
<222> (1)..(2)
<223> n is a, c, g, or t

<400> 32
nnaaaaaaaa aaaaaaaaaa aaagatcgga agagctcgta tgccgtcttc tgcttg 56

<210> 33
<211> 34
<212> DNA

IDT Exhibit 1017.1772

<213> Artificial

<220>
<223> Preamplification fragment A second strand part 1

<400> 33
caagcagaag acggcatacg agctcttccg atct 34

<210> 34
<211> 58
<212> DNA
<213> Artificial

<220>
<223> Preamplification fragment A second strand part 2

<400> 34
tctagccttc tcgcagcaca tccctttctc acatctagag ccaccagcgg catagtaa 58

<210> 35
<211> 131
<212> DNA
<213> Artificial

<220>
<223> Amplification fragment A superior

<220>
<221> misc_feature
<222> (35)..(35)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (37)..(42)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (45)..(49)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (51)..(53)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (55)..(60)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (63)..(73)
<223> n is a, c, g, or t

<400> 35
caagcagaag acggcatacg agctcttccg atctntnnnn nnttnnnnnt nnntnnnnnn 60

ttnnnnnnnn nnnagatcgg aagagcgtcg tgtagggaaa gagtgtagat ctcggtggtc 120

gccgtatcat t 131

IDT Exhibit 1017.1773

<210> 36
<211> 131
<212> DNA
<213> Artificial

<220>
<223> Amplification fragment A inferior

<220>
<221> misc_feature
<222> (59)..(69)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (72)..(77)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (79)..(81)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (83)..(87)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (90)..(95)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (97)..(97)
<223> n is a, c, g, or t

<400> 36
aatgatacgg cgaccaccga gatctacact ctttccctac acgacgctct tccgatctnn 60

nnnnnnnnna annnnnnann nannnnnaan nnnnnanaga tcggaagagc tcgtatgccg 120

tcttctgctt g 131

<210> 37
<211> 136
<212> DNA
<213> Artificial

<220>
<223> Amplification fragment B superior

<220>
<221> misc_feature
<222> (35)..(36)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (38)..(45)
<223> n is a, c, g, or t

<220>
<221> misc_feature

IDT Exhibit 1017.1774

<222> (49)..(52)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (54)..(55)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (57)..(58)
<223> n is a, c, g, or t

<400> 37
caagcagaag acggcatacg agctcttccg atctnntnnn nnnnntttnn nntnntnnaa 60

aaaaaaaaaa aaaaaaaaag atcggaagag cgtcgtgtag ggaaagagtg tagatctcgg 120

tggtcgccgt atcatt 136

<210> 38
<211> 136
<212> DNA
<213> Artificial

<220>
<223> Amplification fragment B inferior

<220>
<221> misc_feature
<222> (79)..(80)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (82)..(83)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (85)..(88)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (92)..(99)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (101)..(102)
<223> n is a, c, g, or t

<400> 38
aatgatacgg cgaccaccga gatctacact ctttccctac acgacgctct tccgatcttt 60

ttttttttt ttttttttnn annannnnaa annnnnnnna nnagatcgga agagctcgta 120

tgccgtcttc tgcttg 136

<210> 39
<211> 33
<212> DNA
<213> Artificial

IDT Exhibit 1017.1775

<220>
<223> Sequencing primer annealing: sequencing primer

<400> 39
acactctttc cctacacgac gctcttccga tct 33

IDT Exhibit 1017.1776

SEQUENCE LISTING

<110> Max-Planck-Gesellschaft zur Förderung der Wissenschaften e.V.

<120> Method for differentiation of polynucleotide strands

<130> GAR-P03013EP

<140> EP09008808.9
<141> 2009-07-06

<160> 39

<170> PatentIn version 3.3

<210> 1
<211> 113
<212> RNA
<213> Artificial

<220>
<223> first-strand synthesis: primers annealing

<220>
<221> misc_feature
<222> (1)..(6)
<223> n is a, c, g, or u

<220>
<221> misc_feature
<222> (8)..(13)
<223> n is a, c, g, or u

<220>
<221> misc_feature
<222> (16)..(20)
<223> n is a, c, g, or u

<220>
<221> misc_feature
<222> (22)..(24)
<223> n is a, c, g, or u

<220>
<221> misc_feature
<222> (26)..(31)
<223> n is a, c, g, or u

<220>
<221> misc_feature
<222> (34)..(44)
<223> n is a, c, g, or u

<220>
<221> misc_feature
<222> (46)..(49)
<223> n is a, c, g, or u

<220>
<221> misc_feature
<222> (51)..(57)
<223> n is a, c, g, or u

<220>
<221> misc_feature
<222> (59)..(66)

IDT Exhibit 1017.1777

<223> n is a, c, g, or u

<220>
<221> misc_feature
<222> (70)..(73)
<223> n is a, c, g, or u

<220>
<221> misc_feature
<222> (75)..(76)
<223> n is a, c, g, or u

<220>
<221> misc_feature
<222> (78)..(79)
<223> n is a, c, g, or u

<400> 1
nnnnnnunnn nnnuunnnnn unnnunnnnn nuunnnnnnn nnnnunnnnu nnnnnnnunn 60

nnnnnnuuun nnnunnunna aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaa 113

<210> 2
<211> 6
<212> DNA
<213> Artificial

<220>
<223> Primer first-strand synthesis

<220>
<221> misc_feature
<222> (1)..(6)
<223> n is a, c, g, or t

<400> 2
nnnnnn 6

<210> 3
<211> 20
<212> DNA
<213> Artificial

<220>
<223> oligo (dT) first-strand synthesis

<400> 3
tttttttttt tttttttttt 20

<210> 4
<211> 42
<212> DNA
<213> Artificial

<220>
<223> first-strand synthesis: reverse transcription; cDNA starting from primer

<220>
<221> misc_feature
<222> (1)..(11)
<223> n is a, c, g, or t

IDT Exhibit 1017.1778

<220>
<221> misc_feature
<222> (14)..(19)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (21)..(23)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (25)..(29)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (32)..(37)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (39)..(42)
<223> n is a, c, g, or t

<400> 4
nnnnnnnnnn naannnnnna nnnannnnna annnnnnann nn 42

<210> 5
<211> 50
<212> DNA
<213> Artificial

<220>
<223> first-strand synthesis: reverse transcription; cDNA starting from oligo (dT)

<220>
<221> misc_feature
<222> (24)..(25)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (27)..(28)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (30)..(33)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (37)..(44)
<223> n is a, c, g, or t

<400> 5
ttttttttt ttttttttt tttnnannan nnnaaannnn nnnnaaaaaa 50

<210> 6
<211> 102
<212> DNA
<213> Artificial

IDT Exhibit 1017.1779

<220>
<223> Second-strand synthesis: Second Strand

<220>
<221> misc_feature
<222> (1)..(6)
<223> n is a, c, g, t or u

<220>
<221> misc_feature
<222> (8)..(13)
<223> n is a, c, g, t or u

<220>
<221> misc_feature
<222> (16)..(20)
<223> n is a, c, g, t or u

<220>
<221> misc_feature
<222> (22)..(24)
<223> n is a, c, g, t or u

<220>
<221> misc_feature
<222> (26)..(31)
<223> n is a, c, g, t or u

<220>
<221> misc_feature
<222> (34)..(44)
<223> n is a, c, g, t or u

<220>
<221> misc_feature
<222> (46)..(49)
<223> n is a, c, g, t or u

<220>
<221> misc_feature
<222> (51)..(57)
<223> n is a, c, g, t or u

<220>
<221> misc_feature
<222> (59)..(66)
<223> n is a, c, g, t or u

<220>
<221> misc_feature
<222> (70)..(73)
<223> n is a, c, g, t or u

<220>
<221> misc_feature
<222> (75)..(76)
<223> n is a, c, g, t or u

<220>
<221> misc_feature
<222> (78)..(79)
<223> n is a, c, g, t or u

<400> 6
nnnnnnunnn nnnuunnnnn unnnunnnnn nuunnnnnnn nnnnunnnnu nnnnnnnunn 60

nnnnnnuuun nnnunnunna aaaaaaaaaa aaaaaaaaaa aa 102

IDT Exhibit 1017.1780

<210> 7
<211> 36
<212> DNA
<213> Artificial

<220>
<223> Fragmentation (ultrasound digestion): Fragment A

<220>
<221> misc_feature
<222> (1)..(1)
<223> n is a, c, g, t or u

<220>
<221> misc_feature
<222> (3)..(8)
<223> n is a, c, g, t or u

<220>
<221> misc_feature
<222> (11)..(15)
<223> n is a, c, g, t or u

<220>
<221> misc_feature
<222> (17)..(19)
<223> n is a, c, g, t or u

<220>
<221> misc_feature
<222> (21)..(26)
<223> n is a, c, g, t or u

<220>
<221> misc_feature
<222> (29)..(36)
<223> n is a, c, g, t or u

<400> 7
nunnnnnnuu nnnnnunnnu nnnnnnuunn nnnnnn 36

<210> 8
<211> 47
<212> DNA
<213> Artificial

<220>
<223> Fragmentation (ultrasound digestion): Fragment B (second strand)

<220>
<221> misc_feature
<222> (1)..(2)
<223> n is a, c, g, t or u

<220>
<221> misc_feature
<222> (4)..(11)
<223> n is a, c, g, t or u

<220>
<221> misc_feature
<222> (15)..(18)
<223> n is a, c, g, t or u

IDT Exhibit 1017.1781

<220>
<221> misc_feature
<222> (20)..(21)
<223> n is a, c, g, t or u

<220>
<221> misc_feature
<222> (23)..(24)
<223> n is a, c, g, t or u

<400> 8
nnunnnnnnn nuuunnnnun nunnaaaaaa aaaaaaaaaa aaaaaaa 47

<210> 9
<211> 42
<212> DNA
<213> Artificial

<220>
<223> Fragmentation (ultrasound digestion) Fragment B (first strand)

<220>
<221> misc_feature
<222> (21)..(22)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (24)..(25)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (27)..(30)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (34)..(41)
<223> n is a, c, g, or t

<400> 9
ttttttttt ttttttttt nnannannnn aaannnnnnn na 42

<210> 10
<211> 39
<212> DNA
<213> Artificial

<220>
<223> End-repair Fragment A (second strand))

<220>
<221> misc_feature
<222> (1)..(1)
<223> n is a, c, g, t or u

<220>
<221> misc_feature
<222> (3)..(8)
<223> n is a, c, g, t or u

<220>

IDT Exhibit 1017.1782

<221> misc_feature
<222> (11)..(15)
<223> n is a, c, g, t or u

<220>
<221> misc_feature
<222> (17)..(19)
<223> n is a, c, g, t or u

<220>
<221> misc_feature
<222> (21)..(26)
<223> n is a, c, g, t or u

<220>
<221> misc_feature
<222> (29)..(39)
<223> n is a, c, g, t or u

<400> 10
nunnnnnnuu nnnnnunnnu nnnnnnuunn nnnnnnnnn 39

<210> 11
<211> 39
<212> DNA
<213> Artificial

<220>
<223> End-repair Fragment A first strand

<220>
<221> misc_feature
<222> (1)..(11)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (14)..(19)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (21)..(23)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (25)..(29)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (32)..(37)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (39)..(39)
<223> n is a, c, g, or t

<400> 11
nnnnnnnnnn naannnnnna nnnannnnna annnnnnan 39

<210> 12
<211> 44

IDT Exhibit 1017.1783

<212> DNA
<213> Artificial

<220>
<223> End-repair Fragment B second strand

<220>
<221> misc_feature
<222> (1)..(2)
<223> n is a, c, g, t or u

<220>
<221> misc_feature
<222> (4)..(11)
<223> n is a, c, g, t or u

<220>
<221> misc_feature
<222> (15)..(18)
<223> n is a, c, g, t or u

<220>
<221> misc_feature
<222> (20)..(21)
<223> n is a, c, g, t or u

<220>
<221> misc_feature
<222> (23)..(24)
<223> n is a, c, g, t or u

<400> 12
nnunnnnnnn nuuunnnnun nunnaaaaaa aaaaaaaaaa aaaa 44

<210> 13
<211> 44
<212> DNA
<213> Artificial

<220>
<223> End-repair Fragment B first strand

<220>
<221> misc_feature
<222> (21)..(22)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (24)..(25)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (27)..(30)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (34)..(41)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (43)..(44)

IDT Exhibit 1017.1784

<223> n is a, c, g, or t

<400> 13
ttttttttttt ttttttttttt nnannannnn aaannnnnnn nann 44

<210> 14
<211> 40
<212> DNA
<213> Artificial

<220>
<223> A-tailing Fragment A second strand

<220>
<221> misc_feature
<222> (1)..(1)
<223> n is a, c, g, t or u

<220>
<221> misc_feature
<222> (3)..(8)
<223> n is a, c, g, t or u

<220>
<221> misc_feature
<222> (11)..(15)
<223> n is a, c, g, t or u

<220>
<221> misc_feature
<222> (17)..(19)
<223> n is a, c, g, t or u

<220>
<221> misc_feature
<222> (21)..(26)
<223> n is a, c, g, t or u

<220>
<221> misc_feature
<222> (29)..(39)
<223> n is a, c, g, t or u

<400> 14
nunnnnnnnuu nnnnnunnnu nnnnnnuunn nnnnnnnnna 40

<210> 15
<211> 40
<212> DNA
<213> Artificial

<220>
<223> A-tailing Fragment A first strand

<220>
<221> misc_feature
<222> (1)..(11)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (14)..(19)
<223> n is a, c, g, or t

IDT Exhibit 1017.1785

<220>
<221> misc_feature
<222> (21)..(23)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (25)..(29)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (32)..(37)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (39)..(39)
<223> n is a, c, g, or t

<400> 15
nnnnnnnnnn naannnnnna nnnannnnna annnnnnana 40

<210> 16
<211> 45
<212> DNA
<213> Artificial

<220>
<223> A-tailing Fragment B second strand

<220>
<221> misc_feature
<222> (1)..(2)
<223> n is a, c, g, t or u

<220>
<221> misc_feature
<222> (4)..(11)
<223> n is a, c, g, t or u

<220>
<221> misc_feature
<222> (15)..(18)
<223> n is a, c, g, t or u

<220>
<221> misc_feature
<222> (20)..(21)
<223> n is a, c, g, t or u

<220>
<221> misc_feature
<222> (23)..(24)
<223> n is a, c, g, t or u

<400> 16
nnunnnnnnn nuuunnnnun nunnaaaaaa aaaaaaaaaa aaaaa 45

<210> 17
<211> 45
<212> DNA
<213> Artificial

<220>

IDT Exhibit 1017.1786

<223> A-tailing Fragment B first strand

<220>
<221> misc_feature
<222> (21)..(22)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (24)..(25)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (27)..(30)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (34)..(41)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (43)..(44)
<223> n is a, c, g, or t

<400> 17
ttttttttttt ttttttttttt nnannannnn aaannnnnnn nanna 45

<210> 18
<211> 107
<212> DNA
<213> Artificial

<220>
<223> Adaptor ligation Fragment A second strand

<220>
<221> misc_feature
<222> (35)..(35)
<223> n is a, c, g, t or u

<220>
<221> misc_feature
<222> (37)..(42)
<223> n is a, c, g, t or u

<220>
<221> misc_feature
<222> (45)..(49)
<223> n is a, c, g, t or u

<220>
<221> misc_feature
<222> (51)..(53)
<223> n is a, c, g, t or u

<220>
<221> misc_feature
<222> (55)..(60)
<223> n is a, c, g, t or u

<220>
<221> misc_feature

IDT Exhibit 1017.1787

<222> (63)..(68)
<223> n is a, c, g, t or u

<220>
<221> misc_feature
<222> (70)..(73)
<223> n is a, c, g, t or u

<400> 18
tacactcttt ccctacacga cgctcttccg atctnunnnn nnuunnnnnu nnnunnnnnn 60

uunnnnnnun nnnagatcgg aagagctcgt atgccgtctt ctgcttg 107

<210> 19
<211> 106
<212> DNA
<213> Artificial

<220>
<223> Adaptor ligation Fragment A first strand

<220>
<221> misc_feature
<222> (35)..(45)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (48)..(53)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (55)..(56)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (58)..(62)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (65)..(70)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (72)..(72)
<223> n is a, c, g, or t

<400> 19
tacactcttt ccctacacga cgctcttccg atctnnnnnn nnnnnaannn nnnannannn 60

nnaannnnnn anagatcgga agagctcgta tgccgtcttc tgcttg 106

<210> 20
<211> 112
<212> DNA
<213> Artificial

<220>
<223> Adaptor ligation Fragment B second strand

IDT Exhibit 1017.1788

<220>
<221> misc_feature
<222> (35)..(36)
<223> n is a, c, g, t or u

<220>
<221> misc_feature
<222> (38)..(45)
<223> n is a, c, g, t or u

<220>
<221> misc_feature
<222> (49)..(52)
<223> n is a, c, g, t or u

<220>
<221> misc_feature
<222> (54)..(55)
<223> n is a, c, g, t or u

<220>
<221> misc_feature
<222> (57)..(58)
<223> n is a, c, g, t or u

<400> 20
tacactcttt ccctacacga cgctcttccg atctnnunnn nnnnnuuunn nnunnunnaa 60

aaaaaaaaaa aaaaaaaaag atcggaagag ctcgtatgcc gtcttctgct tg 112

<210> 21
<211> 112
<212> DNA
<213> Artificial

<220>
<223> Adaptor ligation Fragment B first strand

<220>
<221> misc_feature
<222> (55)..(56)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (58)..(59)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (61)..(64)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (68)..(75)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (77)..(78)
<223> n is a, c, g, or t

<400> 21
tacactcttt ccctacacga cgctcttccg atctttttt ttttttttt ttttnnanna 60

IDT Exhibit 1017.1789

nnnnaaannn nnnnnannag atcggaagag ctcgtatgcc gtcttctgct tg 112

<210> 22
<211> 35
<212> DNA
<213> Artificial

<220>
<223> UDGase treatment Fragment A second strand piece 1

<220>
<221> misc_feature
<222> (35)..(35)
<223> n is a, c, g, or t

<400> 22
tacactcttt ccctacacga cgctcttccg atctn 35

<210> 23
<211> 6
<212> DNA
<213> Artificial

<220>
<223> UDGase treatment fragment A second sttrand piece 2

<220>
<221> misc_feature
<222> (1)..(6)
<223> n is a, c, g, or t

<400> 23
nnnnnn 6

<210> 24
<211> 5
<212> DNA
<213> Artificial

<220>
<223> UDGase treatment fragment A second strand piece 3

<220>
<221> misc_feature
<222> (1)..(5)
<223> n is a, c, g, or t

<400> 24
nnnnn 5

<210> 25
<211> 3
<212> DNA
<213> Artificial

<220>
<223> UDGase treatment fragment A second strand piece 4

<220>
<221> misc_feature

IDT Exhibit 1017.1790

<222> (1)..(3)
<223> n is a, c, g, or t

<400> 25
nnn 3

<210> 26
<211> 6
<212> DNA
<213> Artificial

<220>
<223> UDGase treatment fragment A second strand piece 5

<220>
<221> misc_feature
<222> (1)..(6)
<223> n is a, c, g, or t

<400> 26
nnnnnn 6

<210> 27
<211> 44
<212> DNA
<213> Artificial

<220>
<223> UDGase treatment Fragment A second strand piece 6

<220>
<221> misc_feature
<222> (1)..(10)
<223> n is a, c, g, or t

<400> 27
nnnnnnnnnn agatcggaag agctcgtatg ccgtcttctg cttg 44

<210> 28
<211> 36
<212> DNA
<213> Artificial

<220>
<223> UDGase treatment fragment B second strand piece 1

<220>
<221> misc_feature
<222> (35)..(36)
<223> n is a, c, g, or t

<400> 28
tacactcttt ccctacacga cgctcttccg atctnn 36

<210> 29
<211> 8
<212> DNA
<213> Artificial

<220>
<223> UDGase treatment fragment B second strand piece 2

IDT Exhibit 1017.1791

<220>
<221> misc_feature
<222> (1)..(8)
<223> n is a, c, g, or t

<400> 29
nnnnnnnn 8

<210> 30
<211> 4
<212> DNA
<213> Artificial

<220>
<223> UDGase treatment fragment B second strand piece 3

<220>
<221> misc_feature
<222> (1)..(4)
<223> n is a, c, g, or t

<400> 30
nnnn 4

<210> 31
<211> 2
<212> DNA
<213> Artificial

<220>
<223> UDGase treatment fragment B second strand piece 4

<220>
<221> misc_feature
<222> (1)..(2)
<223> n is a, c, g, or t

<400> 31
nn 2

<210> 32
<211> 56
<212> DNA
<213> Artificial

<220>
<223> UDGase treatment fragment B second strand piece 5

<220>
<221> misc_feature
<222> (1)..(2)
<223> n is a, c, g, or t

<400> 32
nnaaaaaaaa aaaaaaaaaa aaagatcgga agagctcgta tgccgtcttc tgcttg 56

<210> 33
<211> 34
<212> DNA

IDT Exhibit 1017.1792

<213> Artificial

<220>
<223> Preamplification fragment A second strand part 1

<400> 33
caagcagaag acggcatacg agctcttccg atct 34

<210> 34
<211> 58
<212> DNA
<213> Artificial

<220>
<223> Preamplification fragment A second strand part 2

<400> 34
tctagccttc tcgcagcaca tccctttctc acatctagag ccaccagcgg catagtaa 58

<210> 35
<211> 131
<212> DNA
<213> Artificial

<220>
<223> Amplification fragment A superior

<220>
<221> misc_feature
<222> (35)..(35)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (37)..(42)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (45)..(49)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (51)..(53)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (55)..(60)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (63)..(73)
<223> n is a, c, g, or t

<400> 35
caagcagaag acggcatacg agctcttccg atctntnnnn nnttnnnnnt nnntnnnnnn 60

ttnnnnnnnn nnnagatcgg aagagcgtcg tgtagggaaa gagtgtagat ctcggtggtc 120

gccgtatcat t 131

IDT Exhibit 1017.1793

<210> 36
<211> 131
<212> DNA
<213> Artificial

<220>
<223> Amplification fragment A inferior

<220>
<221> misc_feature
<222> (59)..(69)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (72)..(77)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (79)..(81)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (83)..(87)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (90)..(95)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (97)..(97)
<223> n is a, c, g, or t

<400> 36
aatgatacgg cgaccaccga gatctacact ctttccctac acgacgctct tccgatctnn 60

nnnnnnnnna annnnnnann nannnnnaan nnnnnanaga tcggaagagc tcgtatgccg 120

tcttctgctt g 131

<210> 37
<211> 136
<212> DNA
<213> Artificial

<220>
<223> Amplification fragment B superior

<220>
<221> misc_feature
<222> (35)..(36)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (38)..(45)
<223> n is a, c, g, or t

<220>
<221> misc_feature

IDT Exhibit 1017.1794

<222> (49)..(52)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (54)..(55)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (57)..(58)
<223> n is a, c, g, or t

<400> 37
caagcagaag acggcatacg agctcttccg atctnntnnn nnnnntttnn nntnntnnaa 60

aaaaaaaaaa aaaaaaaaag atcggaagag cgtcgtgtag ggaaagagtg tagatctcgg 120

tggtcgccgt atcatt 136

<210> 38
<211> 136
<212> DNA
<213> Artificial

<220>
<223> Amplification fragment B inferior

<220>
<221> misc_feature
<222> (79)..(80)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (82)..(83)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (85)..(88)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (92)..(99)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (101)..(102)
<223> n is a, c, g, or t

<400> 38
aatgatacgg cgaccaccga gatctacact ctttccctac acgacgctct tccgatcttt 60

ttttttttt ttttttttnn annannnnaa annnnnnnna nnagatcgga agagctcgta 120

tgccgtcttc tgcttg 136

<210> 39
<211> 33
<212> DNA
<213> Artificial

IDT Exhibit 1017.1795

<220>
<223> Sequencing primer annealing: sequencing primer

<400> 39
acactctttc cctacacgac gctcttccga tct 33

Claims

1. A method for differentiation of polynucleotide strands comprising the following steps:

   I) Performing strand-specific synthesis on a polynucleotide template by providing polynucleotide template, primers, correspondent buffer, polymerase and sufficient amounts of the four dNTPs: dATP, dCTP, dGTP and dTTP or corresponding analogues thereof, wherein at least one of the four dNTPs is a modified nucleotide used for strand specific differentiation;
   II) Optionally, second-strand synthesis with unmodified dNTPs or corresponding analogues thereof, if said polynucleotide template on step I) is a single-stranded RNA;
   III) Optionally, performing molecular cloning operations with double-stranded DNA; and
   IV) Strand specific differentiation using different interaction of modified and unmodified nucleotides with some selection factor.

2. A method for differentiation of polynucleotide strands according to claim 1 comprising the following steps:

   I) Providing at least one polynucleotide template, primers, buffer solution, at least one enzyme with polymerase activity, sufficient amounts of the four dNTPs dATP, dCTP, dGTP and dTTP or analogues thereof, wherein at least one of the four dNTPs is replaced by a corresponding modified dNTP used for strand-specific differentiation;
   II) Performing specific strand synthesis on said at least one polynucleotide;
   and
   IIIa) Performing strand-specific differentiation by using a selection factor for differentiating strands containing the at least one modified nucleotide from strands containing the corresponding unmodified nucleotide;
   or if said polynucleotide template is a single-stranded RNA,
   IIIb) Performing second-strand synthesis with all four unmodified dNTPs or analogues thereof after removing the at least one modified dNTP; and
   IVb) Performing strand-specific differentiation by using a selection factor for differentiating strands containing the at least one modified nucleotide from strands containing the corresponding unmodified nucleotide;
   or
   IIIc) Performing one or more molecular cloning operations on the double-stranded DNA, wherein the molecular cloning operations are selected from shearing, restriction, digestion, end-polishing, A-tailing, G-tailing and adaptor ligation; and
   IVc) Performing strand-specific differentiation by using a selection factor for differentiating strands containing the at least one modified nucleotide from strands containing the corresponding unmodified nucleotide;
   or if said polynucleotide template is a single-stranded RNA,
   IIId) Performing second-strand synthesis with all four unmodified dNTPs or analogues thereof after removing the at least one modified dNTP;
   IVd) Performing one or more molecular cloning operations on the double-stranded DNA, wherein the molecular cloning operations are selected from shearing, restriction, digestion, end-polishing, A-tailing, G-tailing and adaptor ligation; and
   Vd) Performing strand-specific differentiation by using a selection factor for differentiating strands containing the at least one modified nucleotide from strands containing the corresponding unmodified nucleotide.

3. The method for differentiation of polynucleotide strands according to claim 1 or 2, wherein said polynucleotide template is selected from single-stranded RNA, single-stranded DNA, double-stranded DNA and DNA/RNA hybrids.

4. The method for differentiation of polynucleotide strands according to claim 1 or 2, wherein said selection factor for differentiating strands is at least one chemical agent modifying selectively the at least one modified nucleotide or one or more chemical groups of the at least one modified nucleotide.

IDT Exhibit 1017.1796

5. The method for differentiation of polynucleotide strands according to claim 1 or 2, wherein said selection factor for differentiating strands is at least one factor displaying selective affinity to an affinity group of the at least one modified nucleotide.

6. The method for differentiation of polynucleotide strands according to claim 1 or 2, wherein said selection factor for differentiating strands is one or more antibodies binding selectively to the at least one modified nucleotide.

7. The method for differentiation of polynucleotide strands according to claim 1 or 2, wherein said selection factor for differentiating strands is a solid phase at which the strand containing the at least one modified nucleotide is immobilized.

8. The method for differentiation of polynucleotide strands according to claim 1 or 2, wherein said selection factor for differentiating strands is a different sensitivity to nucleases of the modified and the unmodified nucleotide.

9. The method for differentiation of polynucleotide strands according to claim 8, wherein said nuclease is Uracil-N-Glycosylase and said modified nucleotide is dUTP.

10. The method for differentiation of polynucleotide strands according to claim 5, wherein said factor displaying selective affinity to an affinity group of the at least one modified nucleotide is streptavidin or avidin and the at least one modified nucleotide is a biotinylated nucleotide.

11. The method for differentiation of polynucleotide strands according to claim 4, wherein the at least one modified nucleotide with said chemical group is Aminoallyl-d UTP.

12. The method for differentiation of polynucleotide strands according to claim 6, wherein the at least one modified nucleotide recognized selectively by an antibody is 5-bromodeoxyuridine.

13. The method for differentiation of polynucleotide strands according to claim 1 or 2, wherein the ratio of the at least one modified dNTP towards the corresponding unmodified dNTP or analogues thereof is at least 1:20.

14. The method for differentiation of polynucleotide strands according to claim 2, wherein if the polynucleotide strand is a single RNA strand the RNA is selected from mRNA, tRNA, rRNA, hnRNA, snRNA, snoRNA, scaRNA, asRNA, micro RNA, siRNA, short transcripts, synthetic RNA polynucleotides and semisynthetic RNA polynucleotides as well as physiological or externally induced modifications thereof, or if the polynucleotide strand is a DNA the DNA is selected from siDNA, single-stranded DNA copies from retroviral or hepadnaviral RNA and single-stranded DNA, single-stranded synthetic DNA polynucleotides and single-stranded semisynthetic DNA polynucleotides as well as physiological or externally induced modifications thereof.

15. A kit for performing differentiation of polynucleotide strands comprising:

A) All four dNTPs dATP, dCTP, dGTP and dTTP and corresponding modified deoxynucleotide;
B) Primers;
C) One or more DNA polymerases; and
D) One or more selection factors corresponding to the modified deoxynucleotide and/or one or more selective media for removing the strand containing the modified nucleotide;
or additionally comprising
E) One or more DNA polymerases with reverse transcriptase activity; or additionally comprising; and
F) Components for specific removal of at least one modified or at least one unmodified nucleotide by gel-filtration or purification on silica-column.

IDT Exhibit 1017.1797

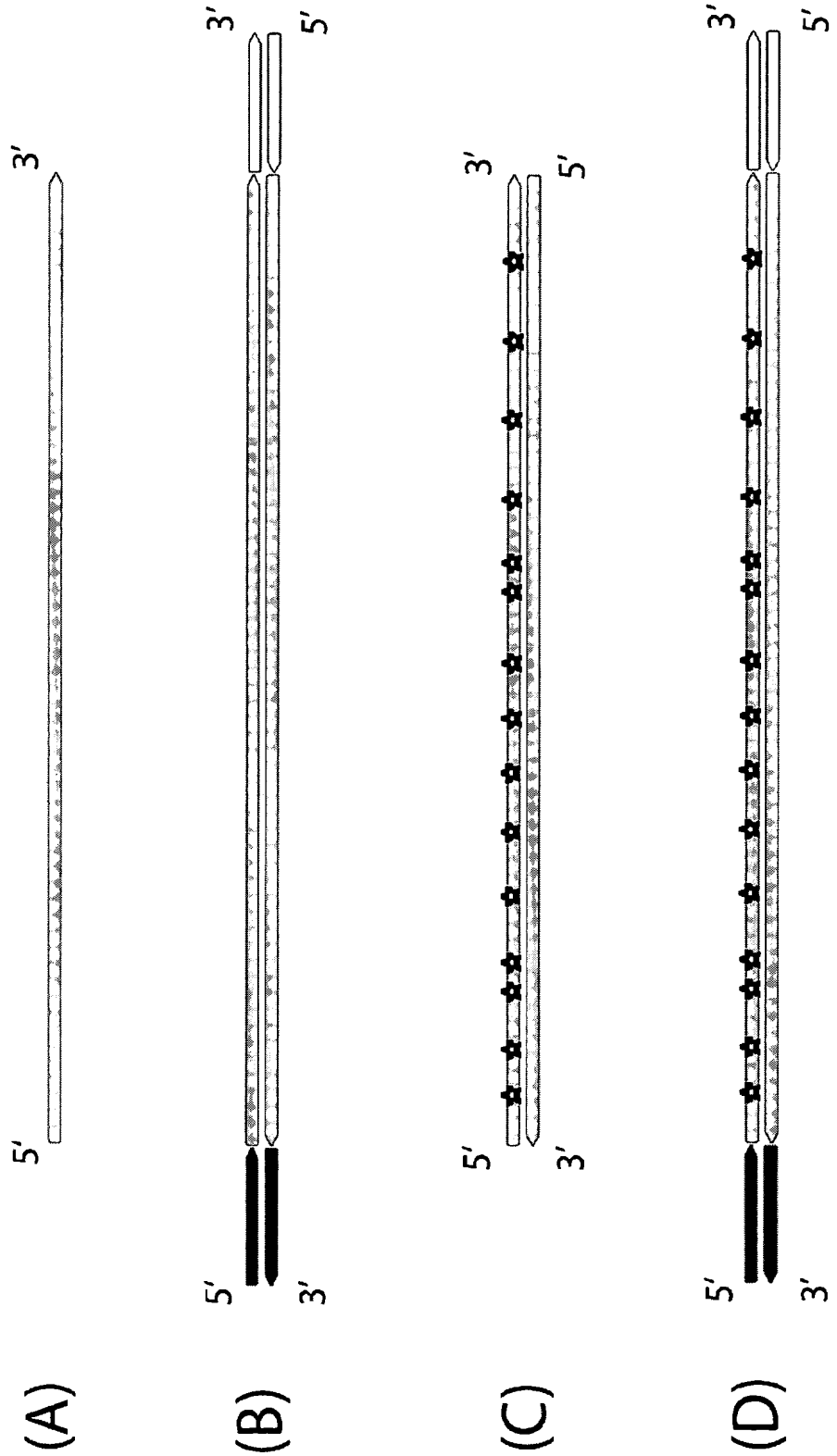


Figure 1

IDT Exhibit 1017.1798

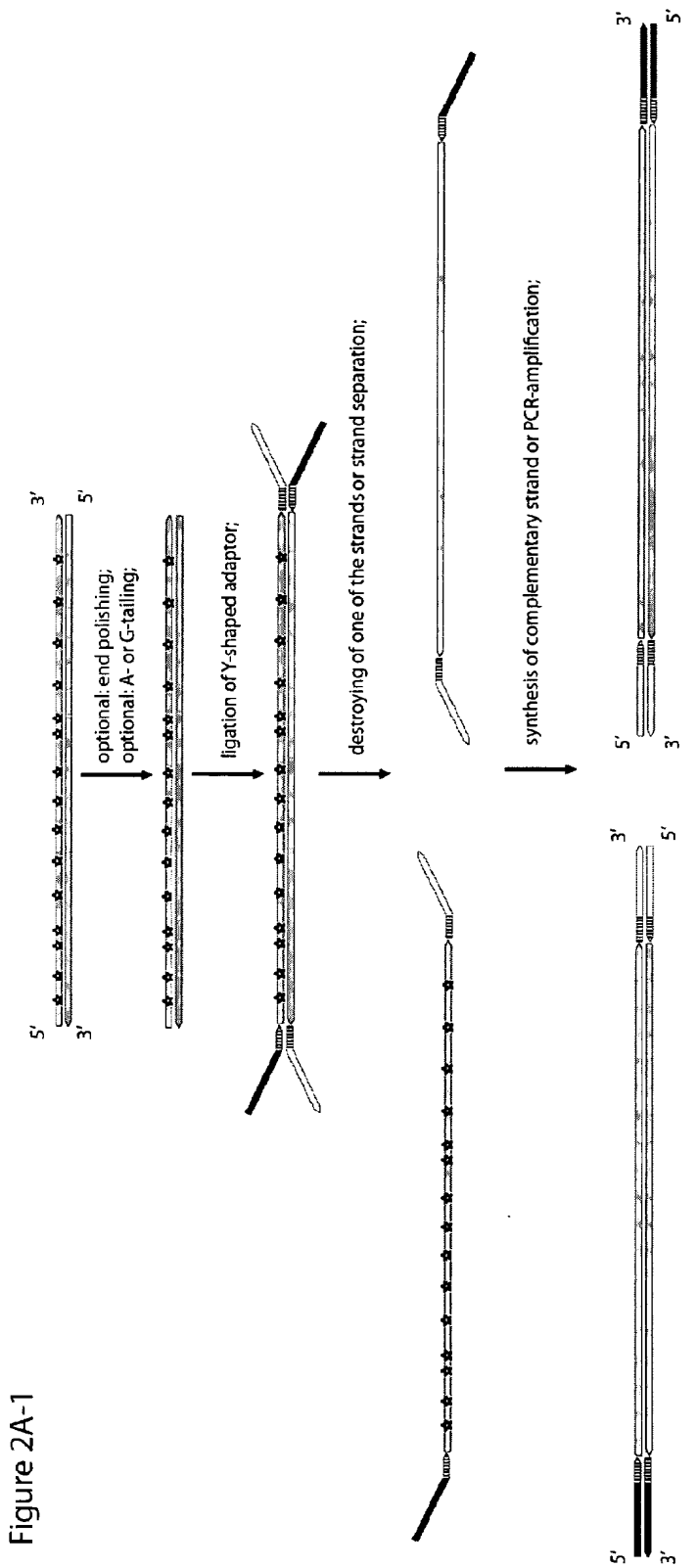


Figure 2A-1

IDT Exhibit 1017.1799

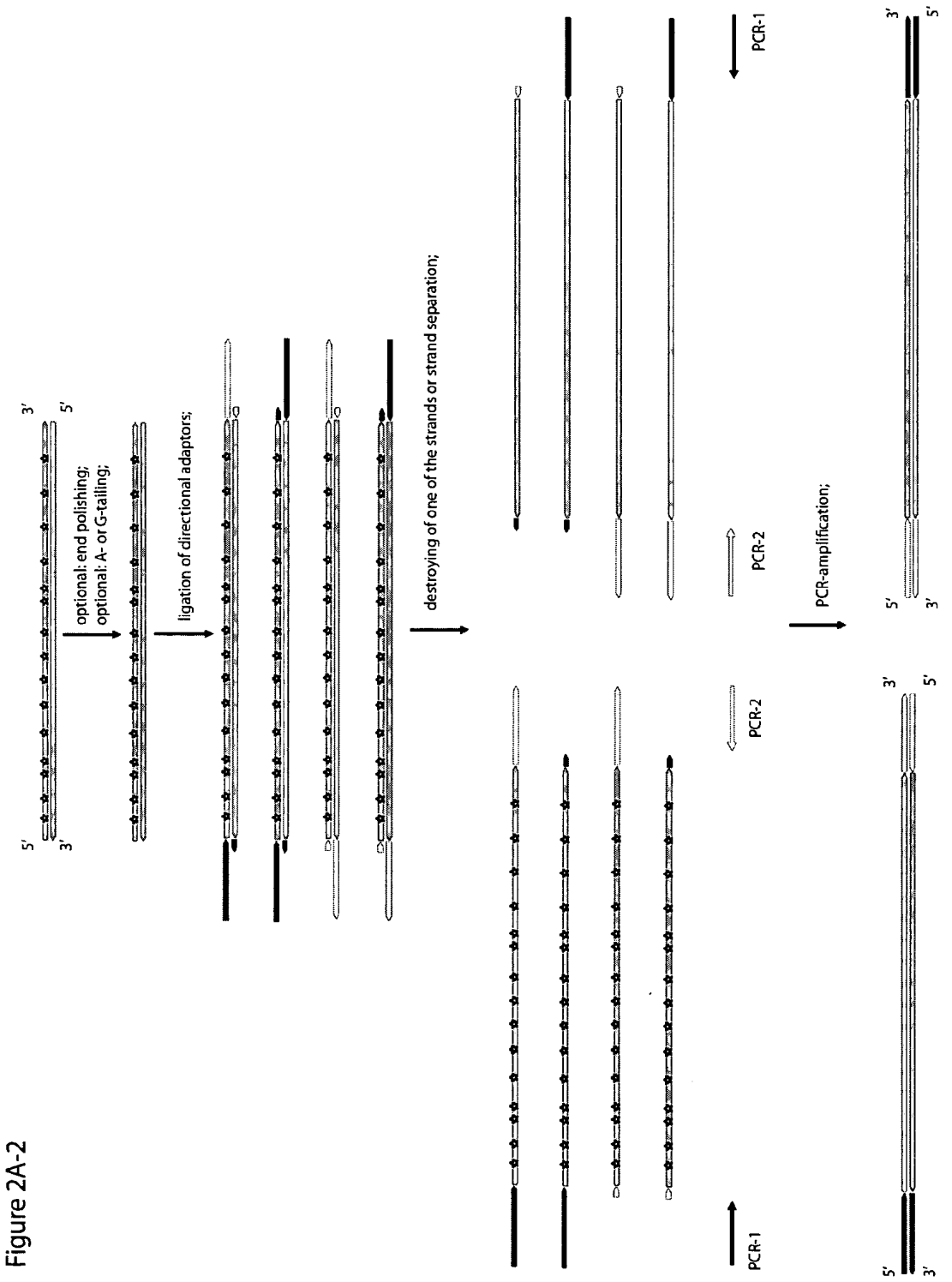


Figure 2A-2

IDT Exhibit 1017.1800

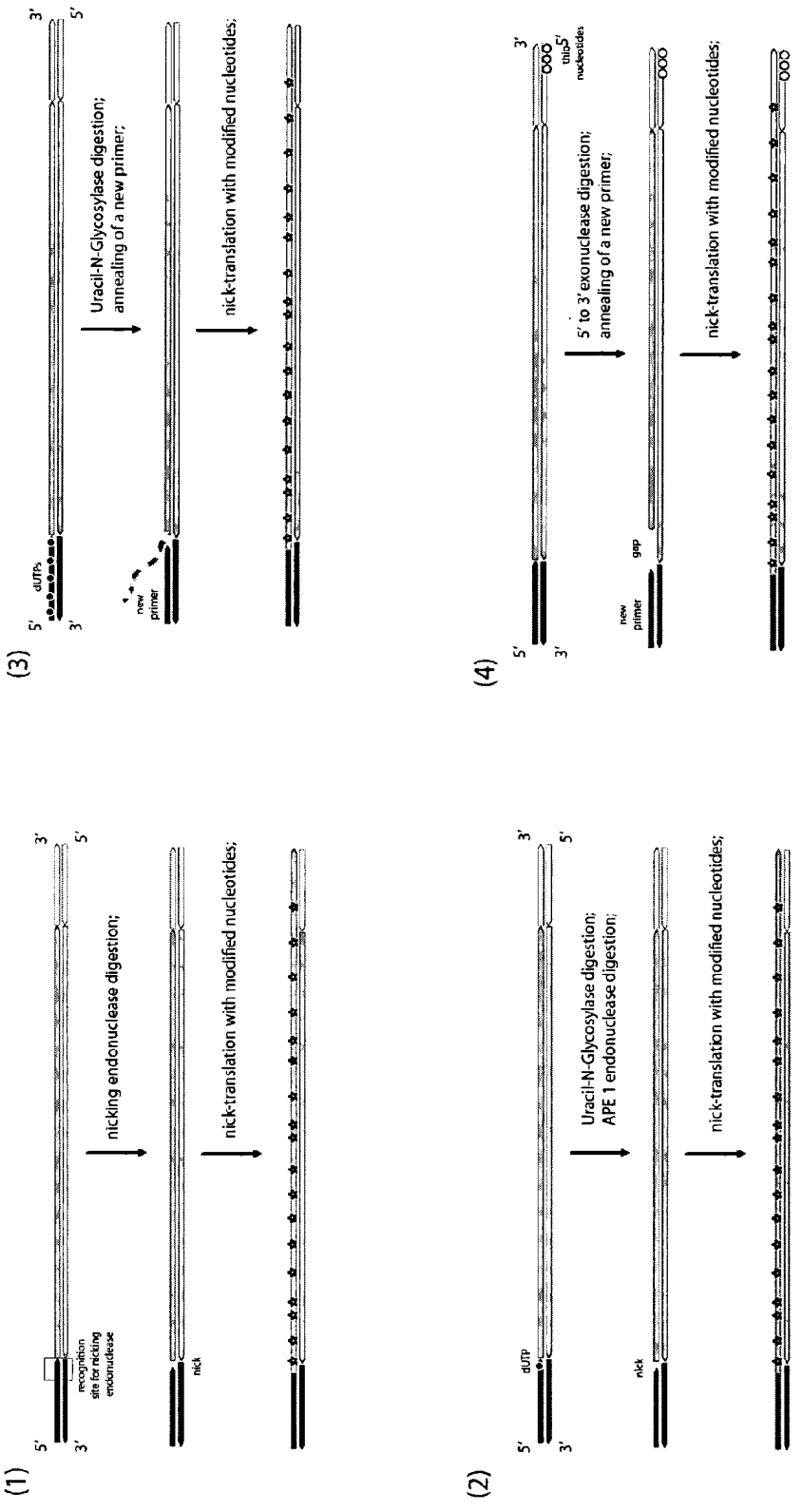


Figure 2B

IDT Exhibit 1017.1801

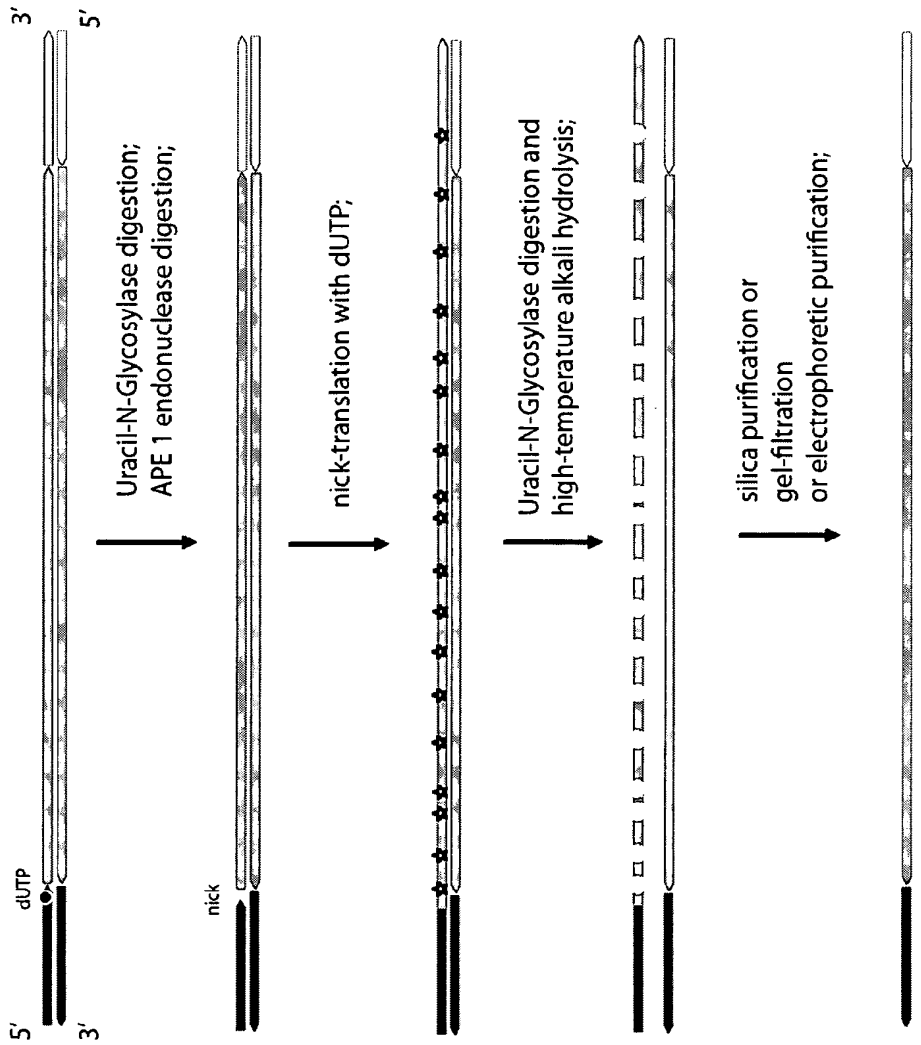


Figure 3A

IDT Exhibit 1017.1802

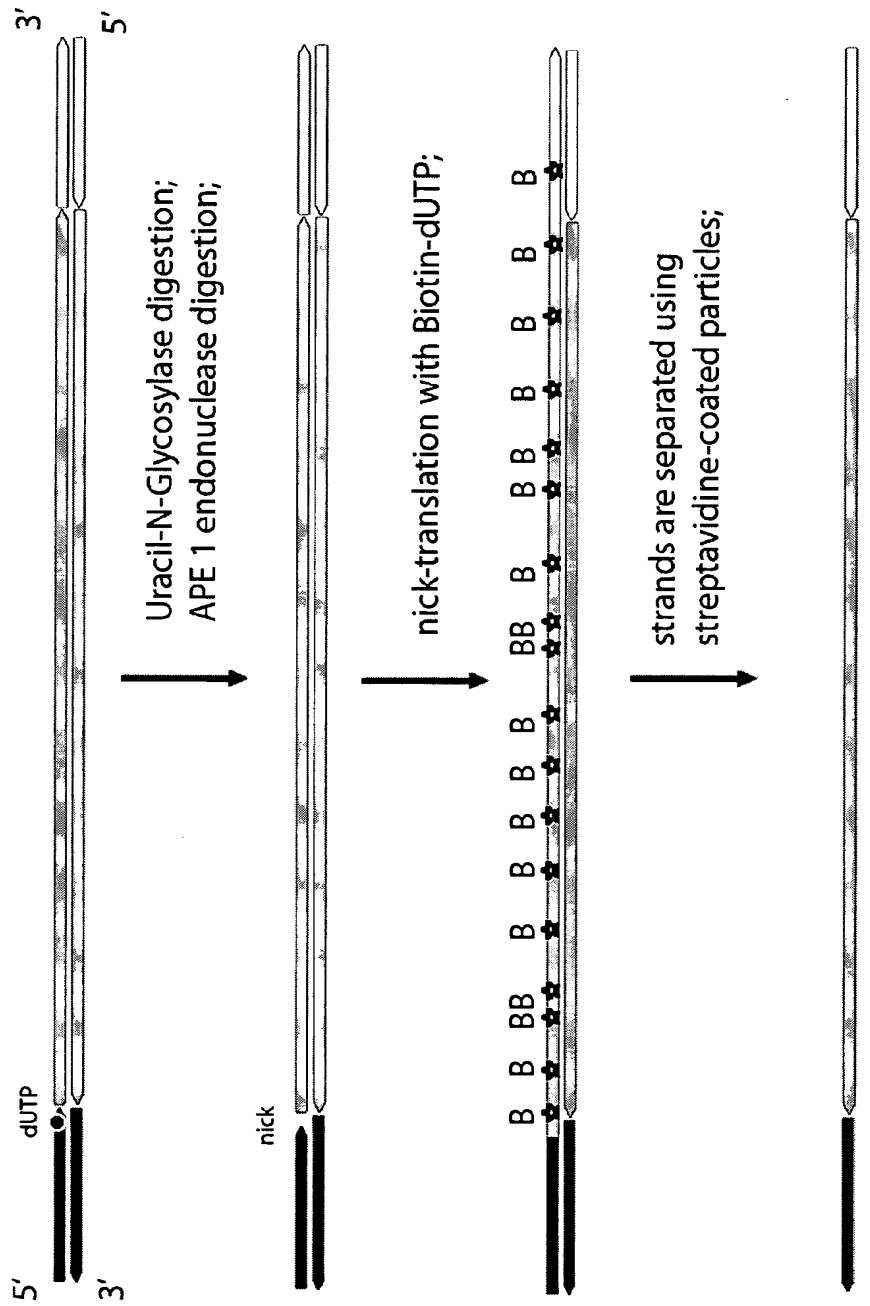


Figure 3B

IDT Exhibit 1017.1803

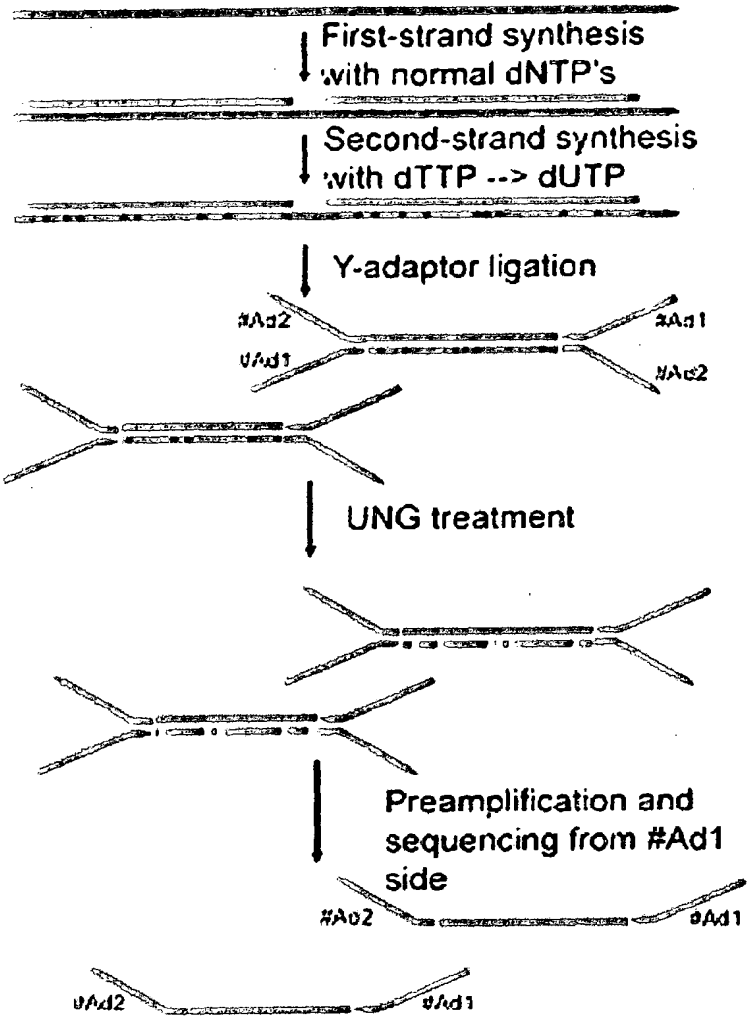
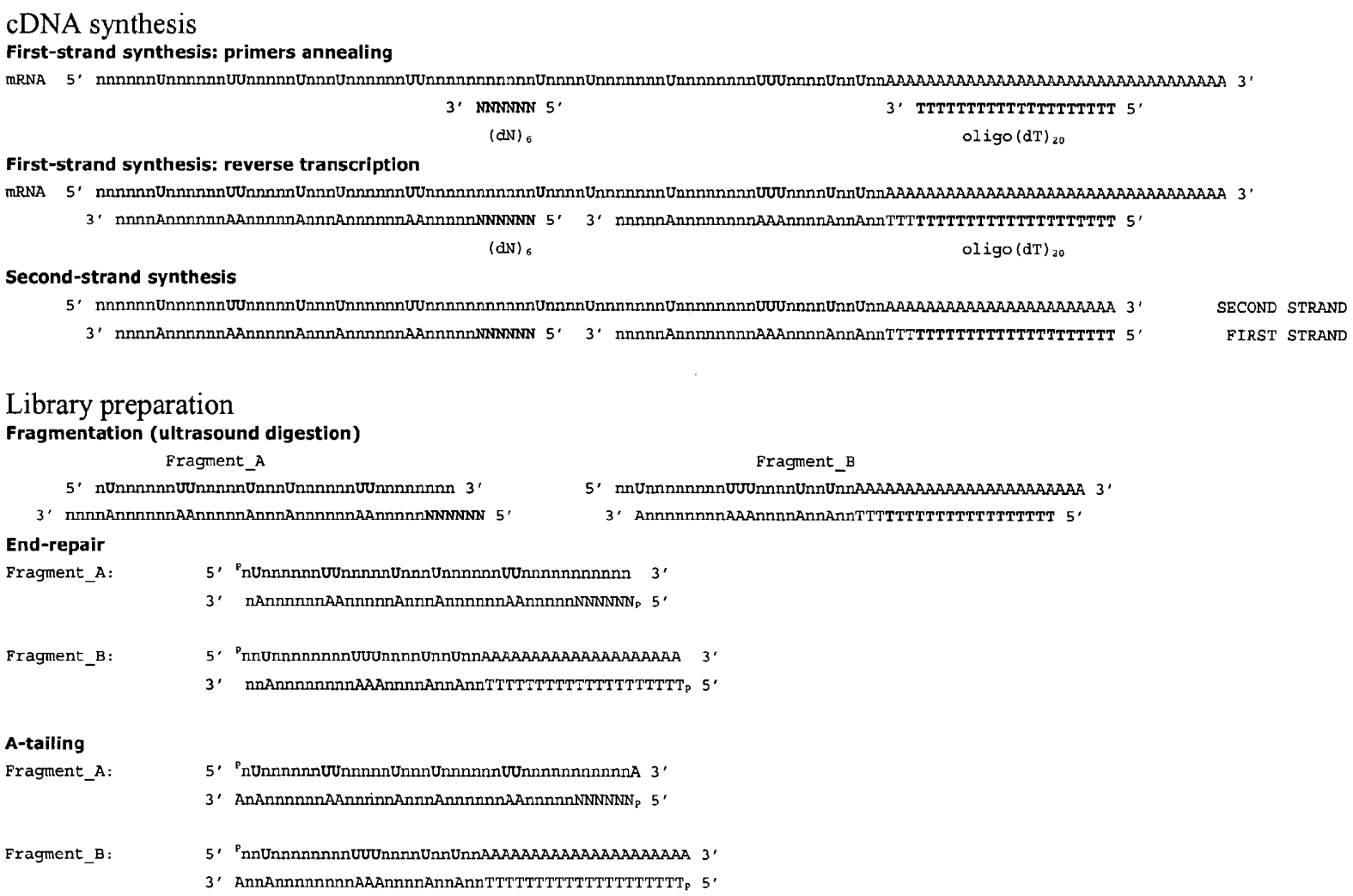


Figure 4

IDT Exhibit 1017.1804

Figure 5. Schematic view of the inventive method on the level of base pairs



EP 2 272 976 A1

IDT Exhibit 1017.1805

Adaptor ligation

```
Fragment_A:
                                                                                #Ad_2
        #Ad_1                                                                  TCGTATGCCGTCTTCTGCTTG 3'
5' TACACTCTTTCCCTACACGAC
                  GCTCTTCCGATCT nUnnnnnnUUnnnnnnUnnnUnnnnnnUUnnnnnnUnnnnnA GATCGGAAGAGC
                  CGAGAAGGCTAG AnAnnnnnnAAnnnnnnAnnnAnnnnnnAAnnnnnnNNNNNN TCTAGCCTTCTCG
                                                                                CAGCACATCCCTTTCTCACAT 5'
                                                                                      #Ad_1
Fragment_B:
        #Ad_1                                                                       #Ad_2
5' TACACTCTTTCCCTACACGAC                                                       TCGTATGCCGTCTTCTGCTTG 3'
                  GCTCTTCCGATCT nnUnnnnnnnnnUUUnnnnnUnnUnnAAAAAAAAAAAAAAAAAAAA GATCGGAAGAGC
                  CGAGAAGGCTAG AnnAnnnnnnnnnAAAnnnnnAnnnAnnTTTTTTTTTTTTTTTTTT TCTAGCCTTCTCG
                                                                                CAGCACATCCCTTTCTCACAT 5'
                                                                                      #Ad_1
        #Ad_2
3' GTTCGTCTTCTGCCGTATGCT
```

UDGase treatment

```
Fragment_A:
        #Ad_1                                                                  #Ad_2
5' TACACTCTTTCCCTACACGAC                                                       TCGTATGCCGTCTTCTGCTTG 3'
                  GCTCTTCCGATCT n·nnnnnn··nnnnn·nnn·nnnnnn··nnnnnnnnnnnnA GATCGGAAGAGC
                  CGAGAAGGCTAG AnAnnnnnnAAnnnnnnAnnnAnnnnnnAAnnnnnnNNNNNN TCTAGCCTTCTCG
                                                                                CAGCACATCCCTTTCTCACAT 5'
                                                                                      #Ad_1
        #Ad_2
Fragment_B:
        #Ad_1                                                                       #Ad_2
5' TACACTCTTTCCCTACACGAC                                                       TCGTATGCCGTCTTCTGCTTG 3'
                  GCTCTTCCGATCT nn·nnnnnnnn···nnnn·nn·nnAAAAAAAAAAAAAAAAAAAA GATCGGAAGAGC
                  CGAGAAGGCTAG AnnAnnnnnnnnnAAAnnnnnAnnnAnnTTTTTTTTTTTTTTTTTT TCTAGCCTTCTCG
                                                                                CAGCACATCCCTTTCTCACAT 5'
        #Ad_2                                                                       #Ad_1
3' GTTCGTCTTCTGCCGTATGCT
```


IDT Exhibit 1017.1806

Preamplification

Fragment_A:

#PCR_2

5' CAAGCACAAGACGGCATACGAGCTCTTCCGATCT 3'

|<------------ 34bp ------------>|< 50-150bp

#PCR 1

3' TCTAGCCTTCTCGCAGCACATCCCTTTCTCACATCTAGAGCCACCAGCGGCATAGTAA 5'

>|<------------------------ 58bp ------------------------>|

Fragment_A:

3' GTTCGTCTTCTGCCGTATGCTCGAGAAGGCTAGAnAnnnnnnAAnnnnnAnnnAnnnnnnAAnnnnnNNNNNNTCTAGCCTTCTCGCAGCACATCCCTTTCTCACAT 5'

#Ad_2 #Ad_1

: ragment_B:

3' GTTCGTCTTCTGCCGTATGCTCGAGAAGGCTAGAnnAnnnnnnnnnAAAnnnnAnnAnnTTTTTTTTTTTTTTTTTTTTCTAGCCTTCTCGCAGCACATCCCTTTCTCACAT 5'

#Ad_2 #Ad_1

FlowCell preparation

Amplification

Fragment_A:

5' CAAGCAGAAGACGGCATACGAGCTCTTCCGATCTnTnnnnnnTTnnnnnTnnnTnnnnnnTTnnnnnnnnnnnAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTATCATT 3'

3' GTTCGTCTTCTGCCGTATGCTCGAGAAGGCTAGAnAnnnnnnAAnnnnnAnnnAnnnnnnAAnnnnnNNNNNNTCTAGCCTTCTCGCAGCACATCCCTTTCTCACATCTAGAGCCACCAGCGGCATAGTAA 5'

Fragment_B:

5' CAAGCAGAAGACGGCATACGAGCTCTTCCGATCTnnTnnnnnnnnnTTTnnnnTnnTnnAAAAAAAAAAAAAAAAAAAAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTATCATT 3'

3' GTTCGTCTTCTGCCGTATGCTCGAGAAGGCTAGAnnAnnnnnnnnnAAAnnnnAnnAnnTTTTTTTTTTTTTTTTTTTTCTAGCCTTCTCGCAGCACATCCCTTTCTCACATCTAGAGCCACCAGCGGCATAGTAA 5'

Linearisation

Fragment_A:

5' CAAGCAGAAGACGGCATACGAGCTCTTCCGATCTnTnnnnnnTTnnnnnTnnnTnnnnnnTTnnnnnnnnnnnAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTATCATT 3'

Fragment_B:

5' CAAGCAGAAGACGGCATACGAGCTCTTCCGATCTnnTnnnnnnnnnTTTnnnnTnnTnnAAAAAAAAAAAAAAAAAAAAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTATCATT 3'

IDT Exhibit 1017.1807

Sequencing primer annealing

```
Fragment_A:

                                    sequencing primer
                                <---- 3' TCTAGCCTTCTCGCAGCACATCCCTTTCTCACA 5'
                                      ||||||||||||||||||||||||||||||||||
5' CAAGCAGAAGACGGCATACGAGCTCTTCCGATCTnTnnnnnnnTnnnnnnnTnnnTnnnnnnnnnnnnnnnAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTATCATT 3'
Fragment_B:

                                                    <---- 3' TCTAGCCTTCTCGCAGCACATCCCTTTCTCACA 5'
                                                          ||||||||||||||||||||||||||||||||||
5' CAAGCAGAAGACGGCATACGAGCTCTTCCGATCTnnTnnnnnnnnnTTTnnnnnTnnTnnTnnAAAAAAAAAAAAAAAAAAAAAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTATCATT
```


IDT Exhibit 1017.1808

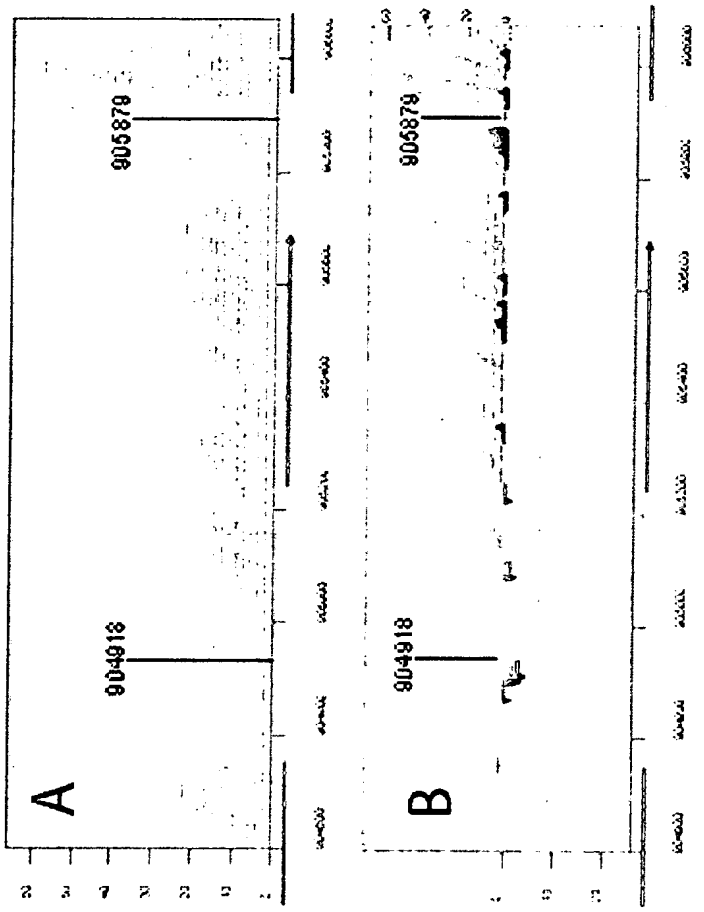


Figure 6

IDT Exhibit 1017.1809

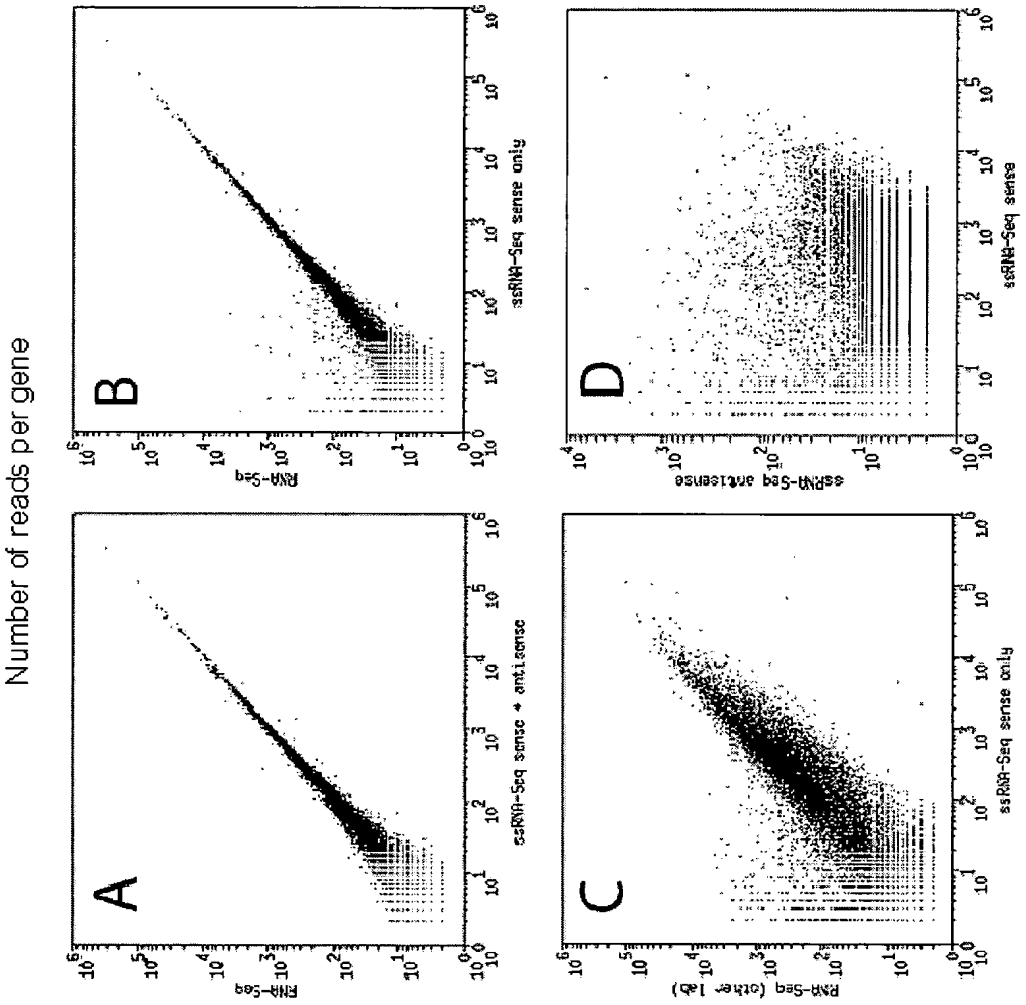
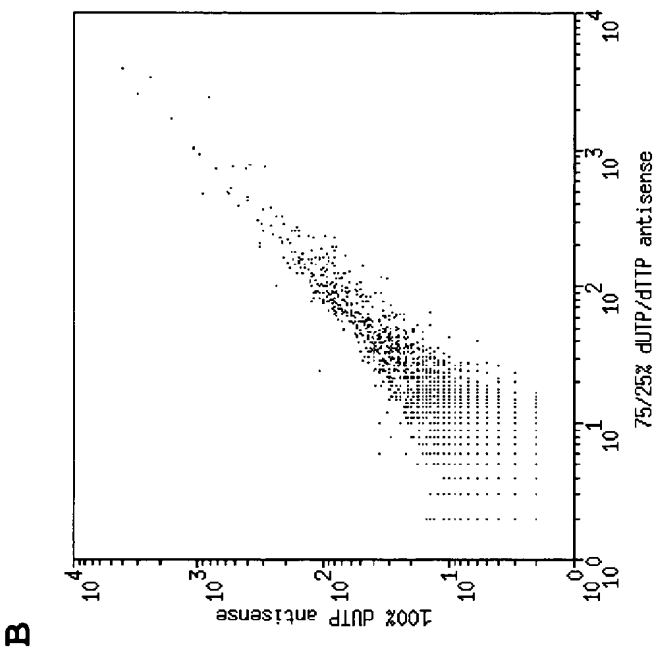


Figure 7

IDT Exhibit 1017.1810



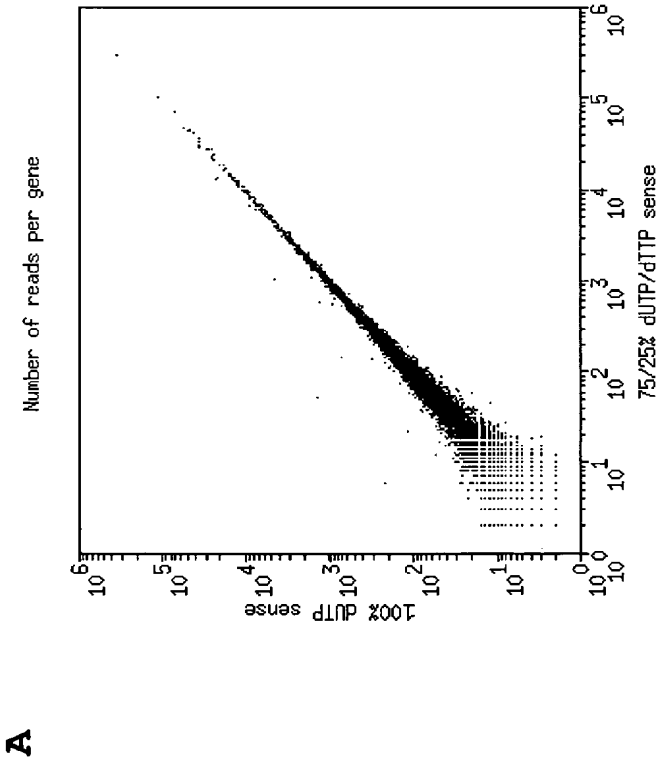
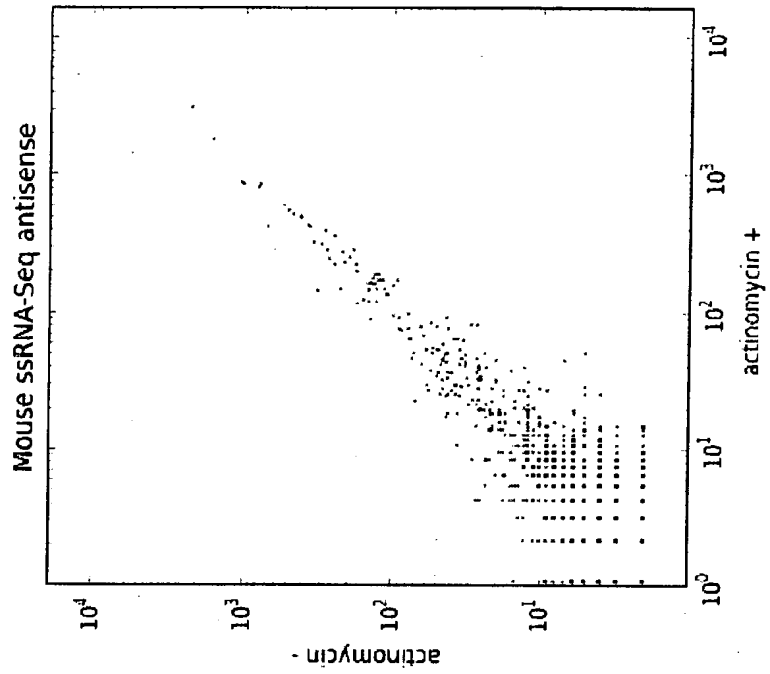


Figure 8

IDT Exhibit 1017.1811



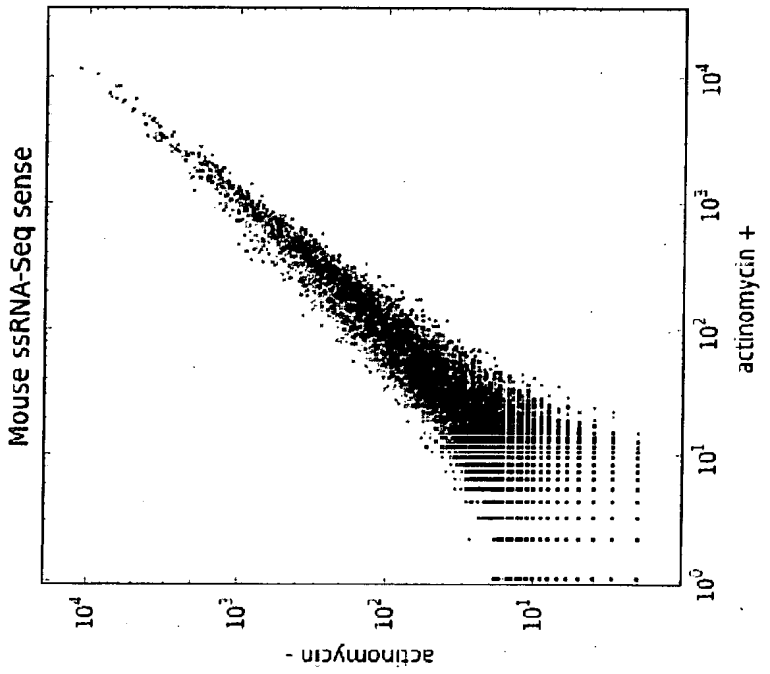


Figure 9

IDT Exhibit 1017.1812

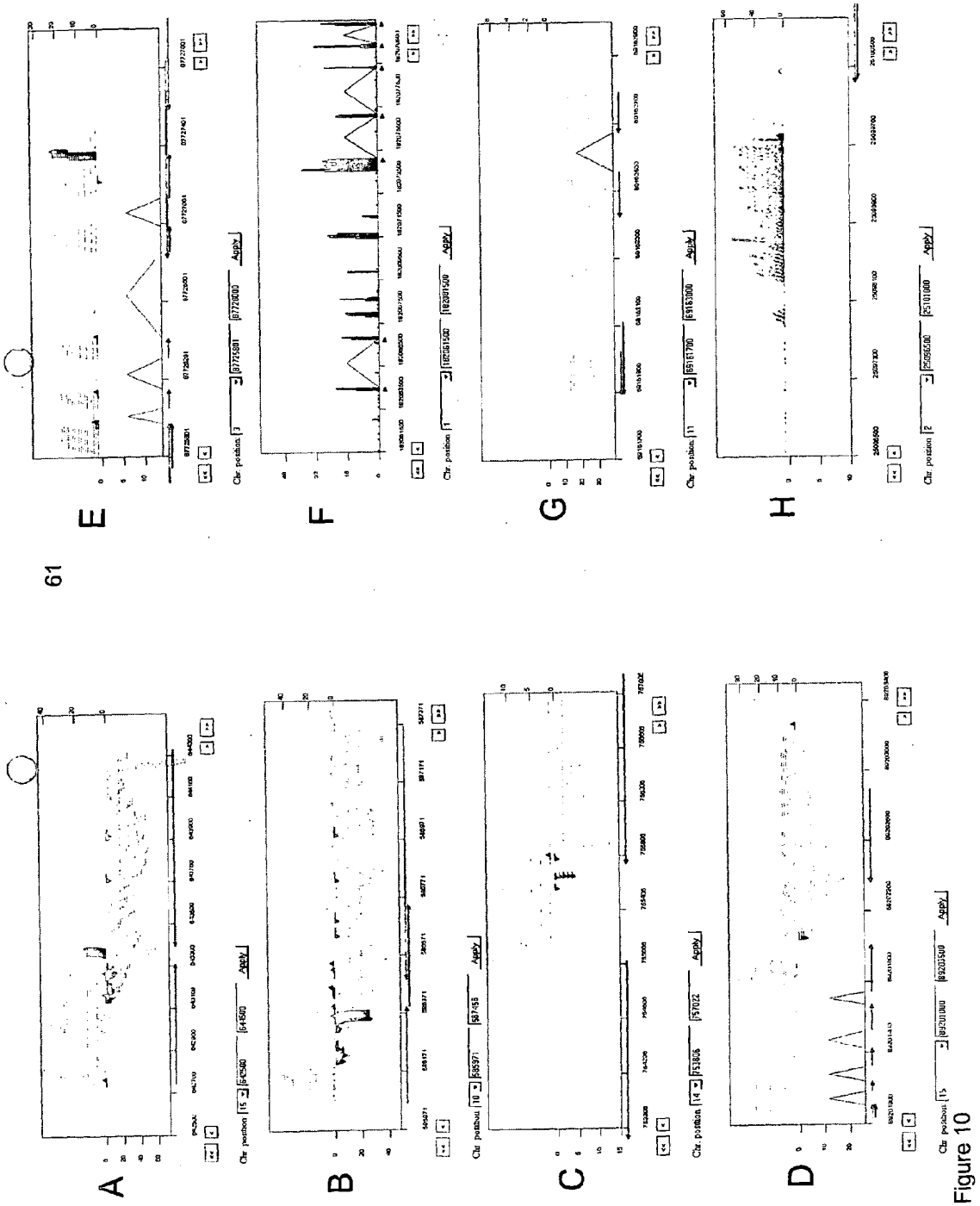


Figure 10

IDT Exhibit 1017.1813

Europäisches Patentamt
European Patent Office
Office européen des brevets

EUROPEAN SEARCH REPORT

Application Number
EP 09 00 8808

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (IPC)
X	US 2007/281863 A1 (RABBANI ELAZAR [US] ET AL) 6 December 2007 (2007-12-06) * abstract; examples 1,6 * * page 3, column 1, paragraph 4 - column 2, paragraph 1 * * page 6, column 2, paragraph 3; figure 2 *	1-14	INV. C12Q1/68
X	WO 97/12061 A1 (EPICENTRE TECHNOLOGIES CORP [US]) 3 April 1997 (1997-04-03) * claim 25 *	15	
A	WO 03/008538 A2 (MOLECULAR STAGING INC [US]) 30 January 2003 (2003-01-30) * abstract; claims 26-29; figure 5; example 9 *	6,12	
			TECHNICAL FIELDS SEARCHED (IPC) C12Q

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
Munich	22 December 2009	Tilkorn, A

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons

& : member of the same patent family, corresponding document

EPO FORM 1503 03.82 (P04C01)

IDT Exhibit 1017.1814

ANNEX TO THE EUROPEAN SEARCH REPORT ON EUROPEAN PATENT APPLICATION NO. EP 09 00 8808

This annex lists the patent family members relating to the patent documents cited in the above-mentioned European search report. The members are as contained in the European Patent Office EDP file on
The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

22-12-2009

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2007281863 A1	06-12-2007	NONE	
WO 9712061 A1	03-04-1997	AU 704625 B2	29-04-1999
		AU 7118396 A	17-04-1997
		CA 2233079 A1	03-04-1997
		EP 0854936 A1	29-07-1998
		JP 11512607 T	02-11-1999
		JP 3561523 B2	02-09-2004
		US 6190865 B1	20-02-2001
WO 03008538 A2	30-01-2003	AU 2002259196 A1	03-03-2003
		US 2004091857 A1	13-05-2004

EPO FORM P0459

For more details about this annex : see Official Journal of the European Patent Office, No. 12/82

IDT Exhibit 1017.1815

REFERENCES CITED IN THE DESCRIPTION

This list of references cited by the applicant is for the reader's convenience only. It does not form part of the European patent document. Even though great care has been taken in compiling the references, errors or omissions cannot be excluded and the EPO disclaims all liability in this regard.

Non-patent literature cited in the description

- **J. Sambrook ; D. Russell.** Molecular Cloning: A Laboratory Manual **[0041]**
- **Hogrefe, H.H. ; Hansen, C.J. ; Scott, B.R. ; Nielsen K.B.** Archaeal dUTPase enhances PCR amplifications with archaeal DNA polymerases by preventing dUTP incorporation. *Proc Natl Acad Sci USA,* 2002, vol. 99, 596-601 **[0065]**

IDT Exhibit 1017.1816

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
1 August 2013 (01.08.2013)

WIPO | PCT

(10) International Publication Number
WO 2013/112923 A1

(51) International Patent Classification:
C40B 50/06 (2006.01)

(21) International Application Number:
PCT/US2013/023278

(22) International Filing Date:
25 January 2013 (25.01.2013)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/591,241 26 January 2012 (26.01.2012) US

(71) Applicant *(for all designated States except US)*: **NUGEN TECHNOLOGIES, INC.** [US/US]; 821 Industrial Road, San Carlos, CA 95050 (US).

(72) Inventors; and
(71) Applicants *(for US only)*: **AMORESE, Doug** [US/US]; 357 S. El Monte Ave., Los Altos, CA 94022 (US). **ARMOUR, Chris** [US/US]; 13531 104th Place, Ne., Kirkland, WA 98034 (US). **KURN, Nurith** [IL/US]; 2876 Ramona Street, Palo Alto, CA 94306 (US).

(74) Agents: **ANDRES JR, Charles** et al.; Wilson Sonsini Goodrich & Rosati, 650 Page Mill Road, Palo Alto, CA 94304-1050 (US).

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

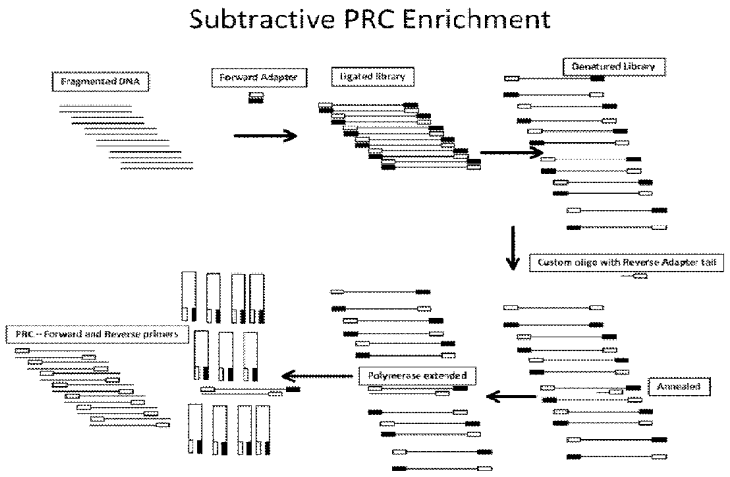
(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU,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

Published:

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

(54) Title: COMPOSITIONS AND METHODS FOR TARGETED NUCLEIC ACID SEQUENCE ENRICHMENT AND HIGH EFFICIENCY LIBRARY GENERATION

FIG. 1



(57) Abstract: The present invention provides methods, compositions and kits for targeted nucleic acid sequence enrichment in a nucleic acid sample and for high efficiency nucleic acid library generation for next generation sequencing (NGS). Specifically, the methods, compositions and kits provided herein are useful for the production and capture of amplification-ready, target-specific and strand-specific regions of interest from nucleic acid samples containing complex DNA.

WO 2013/112923 A1

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

CROSS-REFERENCE

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

BACKGROUND OF THE INVENTION

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

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

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

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

IDT Exhibit 1017.1818

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

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

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

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

IDT Exhibit 1017.1819

SUMMARY OF THE INVENTION

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

IDT Exhibit 1017.1820

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

IDT Exhibit 1017.1821

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

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

IDT Exhibit 1017.1822

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

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

IDT Exhibit 1017.1823

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

INCORPORATION BY REFERENCE

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

BRIEF DESCRIPTION OF THE DRAWINGS

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

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

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

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

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

IDT Exhibit 1017.1824

DETAILED DESCRIPTION OF THE INVENTION

General

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

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

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

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

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

IDT Exhibit 1017.1825

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

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

IDT Exhibit 1017.1826

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

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

IDT Exhibit 1017.1827

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

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

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

IDT Exhibit 1017.1828

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

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

IDT Exhibit 1017.1829

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

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

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

IDT Exhibit 1017.1830

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

IDT Exhibit 1017.1831

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

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

IDT Exhibit 1017.1832

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

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

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

IDT Exhibit 1017.1833

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

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

IDT Exhibit 1017.1834

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

IDT Exhibit 1017.1835

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

IDT Exhibit 1017.1836

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

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

IDT Exhibit 1017.1837

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

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

IDT Exhibit 1017.1838

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

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

IDT Exhibit 1017.1839

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

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

IDT Exhibit 1017.1840

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

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

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

IDT Exhibit 1017.1841

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

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

IDT Exhibit 1017.1842

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

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

IDT Exhibit 1017.1843

Input Nucleic Acid

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

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

IDT Exhibit 1017.1844

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

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

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

Oligonucleotides of the invention

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

IDT Exhibit 1017.1845

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

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

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

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

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

IDT Exhibit 1017.1846

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

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

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

IDT Exhibit 1017.1847

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

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

IDT Exhibit 1017.1848

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

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

IDT Exhibit 1017.1849

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

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

IDT Exhibit 1017.1850

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

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

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

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

IDT Exhibit 1017.1851

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

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

IDT Exhibit 1017.1852

Nucleic acid modifying enzymes

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

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

DNA-Dependent DNA Polymerases

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

IDT Exhibit 1017.1853

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

Methods of Amplification

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

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

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

IDT Exhibit 1017.1854

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

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

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

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

Downstream Applications

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

IDT Exhibit 1017.1855

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

Sequencing

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

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

IDT Exhibit 1017.1856

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

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

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

IDT Exhibit 1017.1857

Genetic Analysis

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

Kits

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

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

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

IDT Exhibit 1017.1858

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

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

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

IDT Exhibit 1017.1859

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

Products based on the Methods of the Invention

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

EXAMPLES

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

[0084] Sample nucleic acid

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

IDT Exhibit 1017.1860

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

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

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

[0087] Amplification

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

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

IDT Exhibit 1017.1861

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

[0089] Sequencing and data analysis

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

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

[0091] Sample nucleic acid

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

[0092] Generation of DNA fragments with ligated forward adapters

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

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

[0094] Primer extension

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

IDT Exhibit 1017.1862

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

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

[0096] Amplification

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

[0097] Sequencing and data analysis

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

IDT Exhibit 1017.1863

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

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

[0099] Sample nucleic acid

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

[00100] Generation of fragments with ligated forward adapters

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

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

[00102] Primer extension

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

IDT Exhibit 1017.1864

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

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

[00104] Amplification

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

[00105] Sequencing and data analysis

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

IDT Exhibit 1017.1865

CLAIMS

WHAT IS CLAIMED IS:

1. A method for enriching for target nucleic acid sequences of interest in a sample comprising nucleic acids, the method comprising:
    a. fragmenting the nucleic acids, thereby generating nucleic acid fragments;
    b. appending a first adaptor to a 5' end of each nucleic acid fragment;
    c. annealing one or more oligonucleotides to the nucleic acid fragments, whereby each of the one or more oligonucleotides comprise a 3' portion that is complementary to a target nucleic acid sequence of interest present in one or more of the nucleic acid fragments, and a 5' portion comprising a second adapter sequence;
    d. extending the one or more oligonucleotides with a polymerase thereby generating one or more oligonucleotide extension products with the first adaptor at a first end and the second adaptor sequence at a second end; and
    e. amplifying the one or more oligonucleotide extension product using a first primer that is complementary to the first adaptor and a second primer that is complementary to the second adaptor sequence to enrich for nucleic acid fragments containing the first adaptor and the second adaptor sequence at each end.
2. The method of claim 1, further comprising an additional step of sequencing the one or more oligonucleotide extension product following amplification.
3. The method of claim 1, wherein the target nucleic acid sequences of interest comprise genomic DNA, RNA, or cDNA.
4. The method of claim 1, wherein the target nucleic acid sequences of interest comprise genomic DNA.
5. The method of claim 1, wherein the target nucleic acid sequences of interest comprise cDNA.
6. The method of claim 1, further comprising denaturing the nucleic acid fragments prior to step c, thereby generating single-stranded nucleic acid fragments with the first adaptor sequence at the 5'end.

IDT Exhibit 1017.1866

7. The method of claim 1, wherein the first adaptor is common to each nucleic acid fragment.

8. The method of claim 1, wherein the second adaptor sequence is common to the one or more oligonucleotides.

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

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

11. The method of claim 1, wherein step b is performed by ligation.

12. The method of claim 1, further comprising an additional step of performing gap repair following ligation of the first adapter to create nucleic acid fragments with complementary termini.

13. A composition comprising the target nucleic acid sequences of interest enriched by the method of claim 1.

14. The method of claim 1, wherein the polymerase is a DNA polymerase.

15. A method for enriching for target nucleic acid sequences of interest in a sample comprising nucleic acids, the method comprising:
    a. fragmenting the nucleic acids, thereby generating nucleic acid fragments;
    b. appending a first adapter to the nucleic acid fragments wherein the first adaptor comprises a partial duplex with a short strand and a long strand wherein a 3' end of the short strand of the partial duplex adaptor comprises a blocking group, and a 5' end of the long strand of the partial duplex adaptor comprises a restriction and/or cleavage site for a nucleic acid modifying enzyme;
    c. denaturing the nucleic acid fragments, thereby creating single-stranded nucleic acid fragments;
    d. annealing one or more oligonucleotides to the single stranded nucleic acid fragments, whereby each of the one or more oligonucleotides comprise sequence that is complementary to a target nucleic acid sequence of interest present in one or more of the single-stranded nucleic acid fragments, extending the one or more oligonucleotides with a polymerase

IDT Exhibit 1017.1867

to produce one or more double stranded nucleic acid complexes comprising the target nucleic acid sequences and their complements, a first end with a double stranded restriction and/or cleavage site for the nucleic acid modifying enzyme , and a second end with a 3' overhang comprising the short strand of the first adaptor;

e. cleaving the double stranded restriction and/or cleavage site with the nucleic acid modifying enzyme, thereby generating a cleavage site;

f. ligating a second adaptor to the cleavage site, wherein the second adaptor comprises a duplex with two strands;

g. denaturing the one or more double-stranded nucleic acid complexes, thereby generating one or more single stranded nucleic acid fragments comprising the target nucleic acid sequences, a strand from the second adapter at the first end, and the short strand of the first adaptor at the second end; and

h. amplifying the one or more single stranded nucleic acid fragments comprising the one or more target nucleic acid sequences with a first primer comprising sequence complementary to the strand from the second adapter and a second primer comprising sequence complementary to the short strand of the first adaptor, thereby enriching for the one or more target nucleic acid sequences.

16. The method of claim 15, further comprising an additional step of sequencing the one or more single stranded nucleic acid fragments from step h following amplification.

17. The method of claim 15, wherein the target nucleic acid sequences of interest comprise genomic DNA, RNA, or cDNA.

18. The method of claim 15, wherein the target nucleic acid sequences of interest comprise genomic DNA

19. The method of claim 15, wherein the target nucleic acid sequences of interest comprise cDNA.

20. The method of claim 15, wherein the first adaptor and the second adaptor are common to each of the nucleic acid fragments.

IDT Exhibit 1017.1868

21. The method of claim 15, wherein the first adaptor and the second adaptor are distinct from each other.

22. The method of claim 15, wherein the first adaptor and/or the second adaptor further comprise barcode sequence.

23. The method of claim 15, wherein the double stranded restriction and/or cleavage site for the nucleic acid modifying enzyme from step e comprises the 5' end of the long strand of the partial duplex of the first adaptor and sequence complementary to the 5' end of the long strand of the partial duplex of the first adaptor generated from extension of the one or more oligonucleotides;

24. The method of claim 15, wherein the nucleic acid modifying enzyme comprises a restriction enzyme.

25. The method of claim 15, wherein step b is performed by ligation.

26. A composition comprising the target nucleic sequences of interest enriched by the method of claim 15.

27. The method of claim 15, wherein the polymerase is a DNA polymerase.

28. A method for generating a library of nucleic acid sequences, the method comprising:
    a. fragmenting a sample comprising nucleic acids, thereby generating nucleic acid fragments;
    b. appending a first adapter to each of the nucleic acid fragments;
    c. denaturing the nucleic acid fragments, thereby generating a library of single-stranded nucleic acid fragments;
    d. annealing one or more oligonucleotides to the single-stranded nucleic acid fragments wherein each of the one or more oligonucleotides comprises a 3' portion complementary to sequence in one or more of the single-stranded nucleic acid fragments and a 5' portion comprising a second adaptor sequence;
    e. extending the one or more oligonucleotides with a polymerase thereby generating one or more oligonucleotide extension products comprising the first adaptor at a first end and the second adapter sequence at a second end; and

IDT Exhibit 1017.1869

f. amplifying the one or more oligonucleotide extension products with a set of primers specific to the first adaptor and the second adaptor sequence to generate a library of nucleic acid fragments comprising the first adaptor and second adaptor sequence at each end.

29. The method of claim 28, further comprising an additional step of performing a gap repair reaction following ligation of the forward adapter to create nucleic acid fragments with complementary termini.

30. The method of claim 28, further comprising an additional step of sequencing the amplified one or more oligonucleotide extension product from step f.

31. The method of claim 28, wherein the nucleic acid sequence comprises genomic DNA.

32. The method of claim 28, wherein the nucleic acid sequence comprises cDNA.

33. The method of claim 28, without step c, wherein the nucleic acid fragments are double-stranded.

34. The method of claim 28, wherein the 3' portion of the one or more oligonucleotides of step d comprises random sequence.

35. The method of claim 28, wherein step b is performed by ligation.

36. The method of claim 28, wherein the first adaptor and the second adaptor sequence are common to each nucleic acid fragment.

37. The method of claim 28, wherein the first adaptor and the second adaptor sequence are distinct from each other.

38. The method of claim 28, wherein the first adaptor and/or the second adaptor sequence further comprise barcode sequence.

39. A composition comprising the library of nucleic sequences generated by the method of claim 28.

40. The method of claim 28, wherein the polymerase is a DNA polymerase.

41. A method for enriching for target nucleic acid sequences of interest from a library comprising nucleic acid inserts with a first adaptor on a first end and a second adaptor on a second end, the method comprising:

IDT Exhibit 1017.1870

a. denaturing the nucleic acid inserts, thereby generating a library of single stranded nucleic acid inserts;
b. annealing one or more oligonucleotides to the single stranded nucleic acid inserts, wherein each of the one or more oligonucleotides comprises a 3' portion that is complementary to a target nucleic acid sequence of interest present in one or more of the nucleic acid inserts, and a 5' portion comprising a third adaptor sequence;
c. extending the one or more oligonucleotides with a polymerase thereby generating one or more oligonucleotide extension products with the first adaptor at the first end and the third adaptor sequence at the second end; and
d. amplifying the one or more oligonucleotide extension products using a first primer that is complementary to the first adaptor and a second primer that is complementary to the third adaptor sequence to enrich for nucleic acid fragments containing the first adaptor and the third adaptor sequence at each end.

42. The method of claim 41, further comprising an additional step of sequencing the amplified one or more oligonucleotide extension products from step d.

43. The method of claim 41, wherein the target nucleic acid sequences of interest comprises genomic DNA.

44. The method of claim 41, wherein the target nucleic acid sequences of interest comprises cDNA.

45. The method of claim 41, without step a, wherein the nucleic acid fragments are double-stranded.

46. The method of claim 41, wherein the first adaptor and the second adaptor are common to each nucleic acid fragment.

47. The method of claim 41, wherein the third adaptor sequence is common to the one or more oligonucleotides.

48. The method of claim 41, wherein the first adaptor and the second adaptor are distinct from each other.

IDT Exhibit 1017.1871

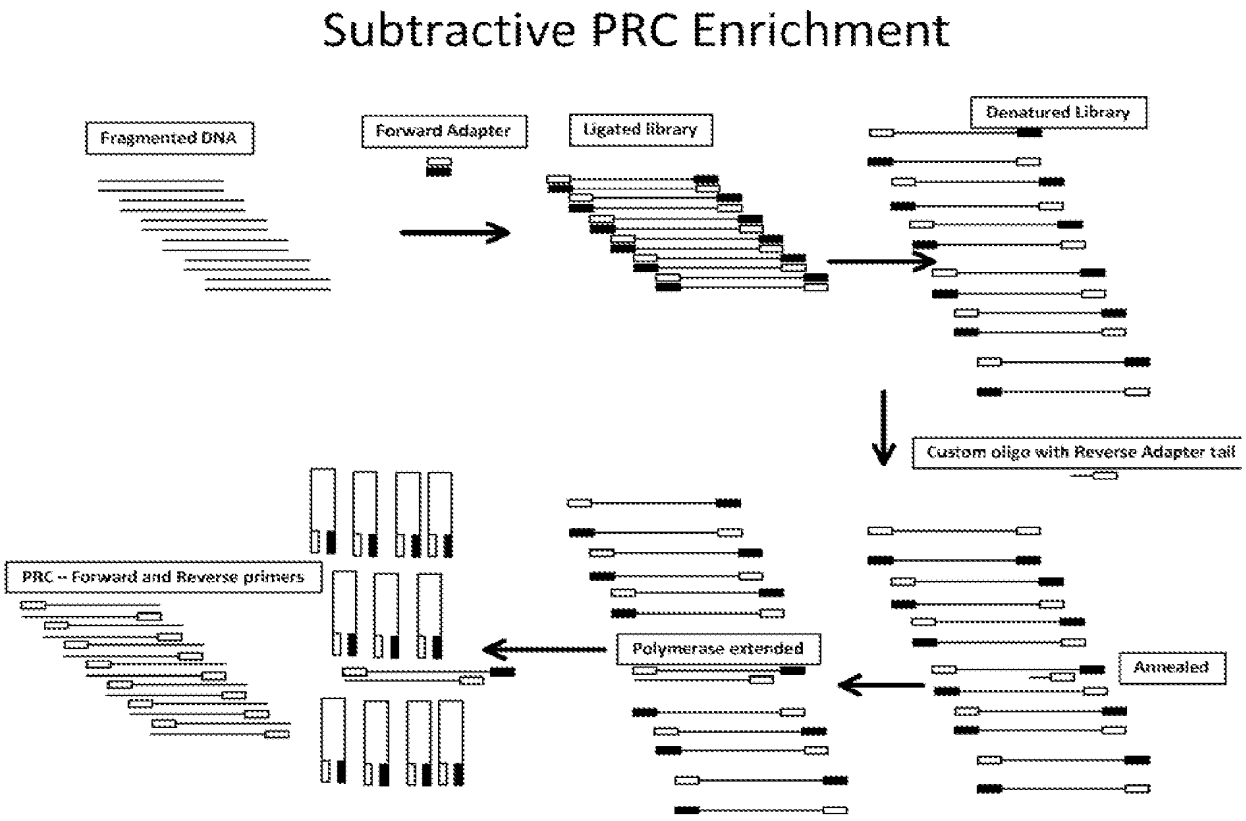
49. The method of claim 41, wherein the first adaptor and the second adaptor are the same.

50. The method of claim 41, wherein the first adaptor, the second adaptor and/or the third adaptor sequence further comprise barcode sequence.

51. A composition comprising the target nucleic acid sequences of interest enriched by the method of claim 41.

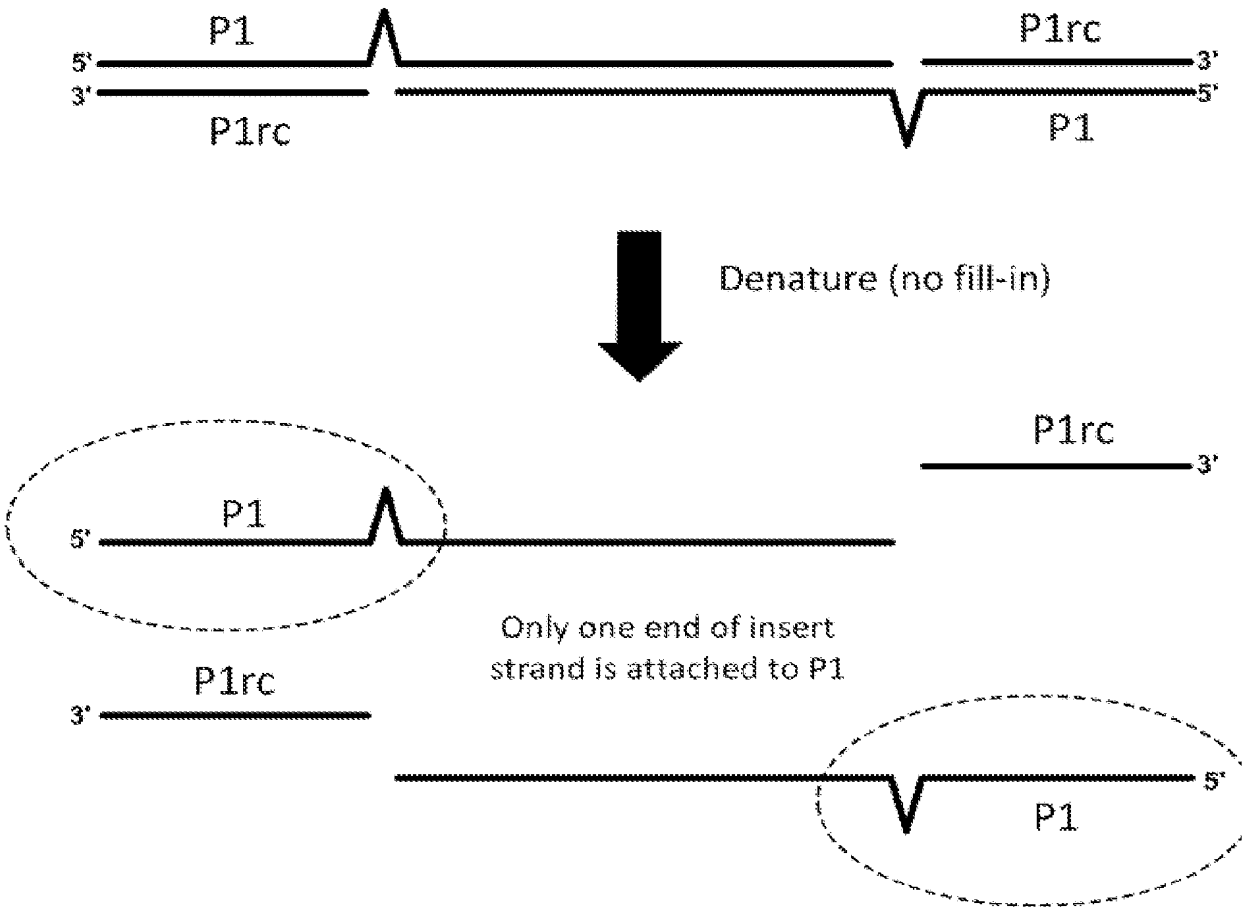
IDT Exhibit 1017.1872

FIG. 1



IDT Exhibit 1017.1873

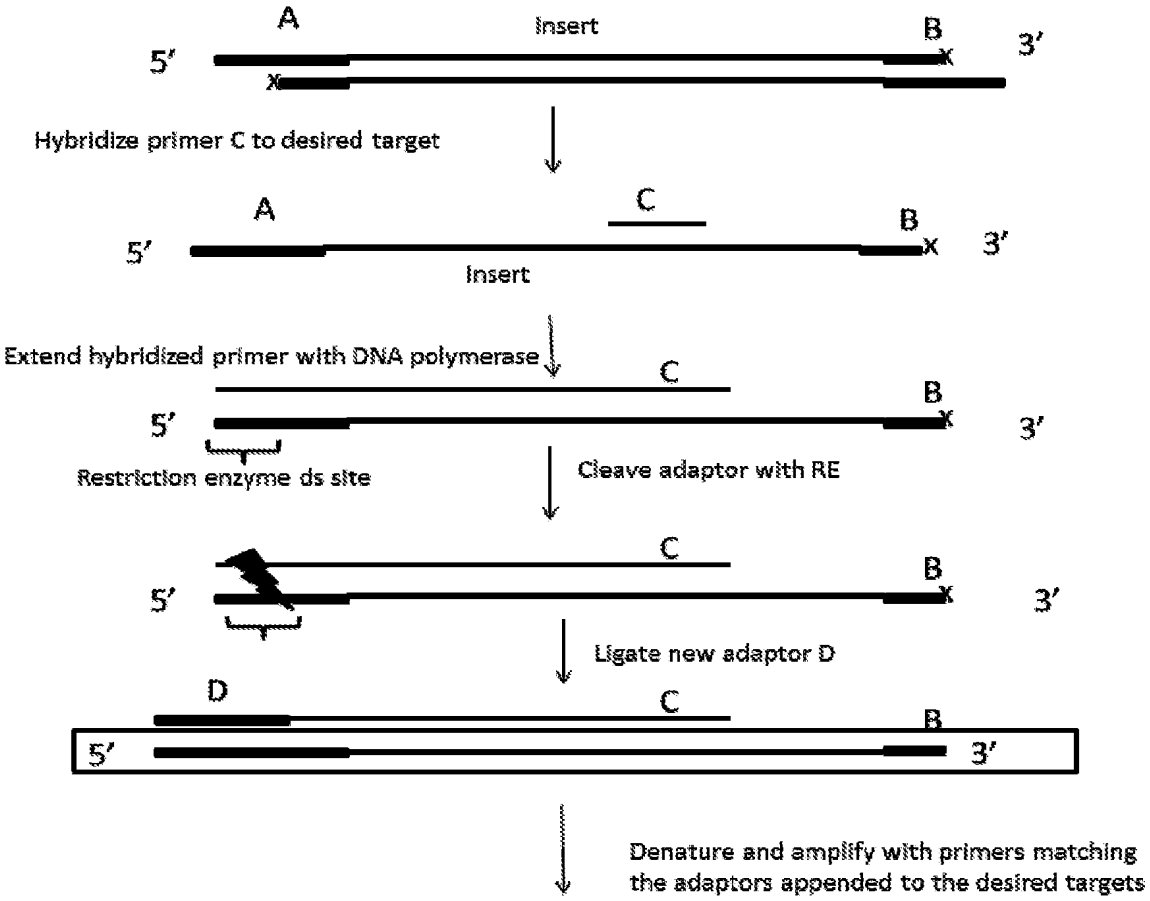
FIG. 2



IDT Exhibit 1017.1874

FIG. 3

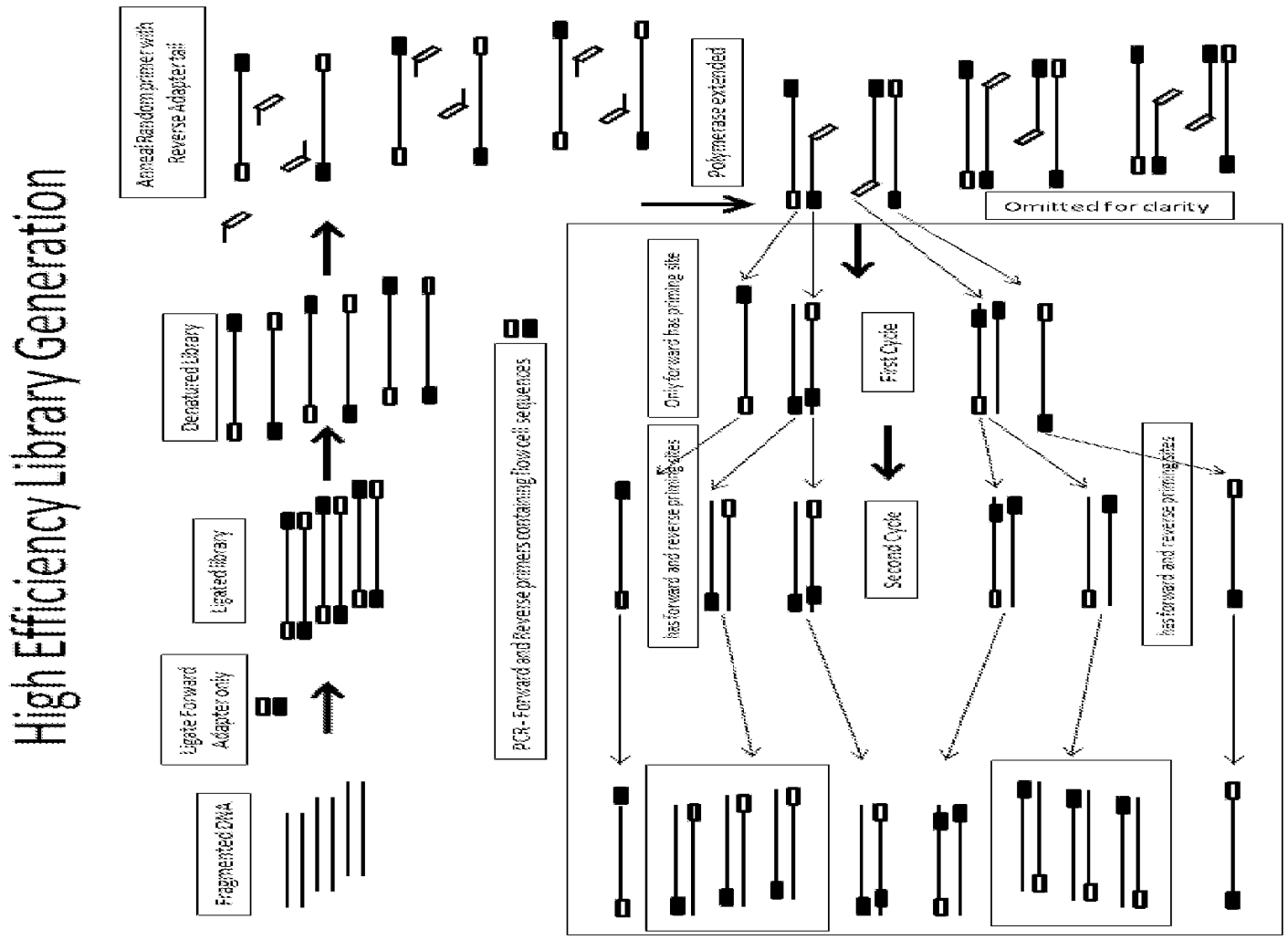
Primer Directed Targeted Enrichment Of Sequences Of Interest In A Complex Library



IDT Exhibit 1017.1875

IDT Exhibit 1017.1876

FIG. 4



INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2013/023278

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

B. FIELDS SEARCHED

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

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
CPC - C12N 15/10; C12P 19/34; C12Q 1/6806, 1/6844, 1/6853, 1/6855 (2013.01)

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

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	BORODINA et al. 'A Strand-Specific Library Preparation Protocol for RNA Sequencing Methods in Enzymology,' Methods in Enzymology, Volume 500 Pages 79-98 (2011). entire document	1-9, 11-14, 28-37, 39-48, 51
Y		10, 15-27, 38, 50
Y	US 2011/0104785 A1 (VAIDYANATHAN et al) 05 May 2011 (05.05.2011) entire document	10, 15-27, 38, 50
Y	US 2007/0231823 A1 (MCKERNAN et al) 04 October 2007 (04.10.2007) entire document	15-27
A	LEVIN et al. 'Comprehensive comparative analysis of strand-specific RNA sequencing methods nature methods,' Nature Methods, VOL.7 NO.9 SEPTEMBER 2010 Pages. 709-718. entire document	1-51

☐ Further documents are listed in the continuation of Box C. ☐

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
26 March 2013	16 APR 2013

Name and mailing address of the ISA/US	Authorized officer:
Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-3201	Blaine R. Copenheaver PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

Form PCT/ISA/210 (second sheet) (July 2009)

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
20 August 2015 (20.08.2015)

WIPO | PCT

(10) International Publication Number
WO 2015/122967 A1

(51) International Patent Classification:
C12N 15/10 (2006.01) *G06F 19/10* (2011.01)
C12Q 1/68 (2006.01)

(21) International Application Number:
PCT/US2014/072293

(22) International Filing Date:
23 December 2014 (23.12.2014)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/939,658 13 February 2014 (13.02.2014) US
62/040,804 22 August 2014 (22.08.2014) US

(71) Applicant: **CLONTECH LABORATORIES, INC.** [US/US]; 1290 Terra Bella Avenue, Mountain View, California 94043 (US).

(72) Inventors: **FARMER, Andrew Alan**; 343 Marich Way, Los Altos, California 94022 (US). **BETTS, Craig**; 1290 Terra Bella Avenue, Mountain View, California 94043 (US).

(74) Agent: **KUNG, Viola T.**; Perkins Coie LLP, 3150 Porter Drive, Palo Alto, CA 94304-1212 (US).

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

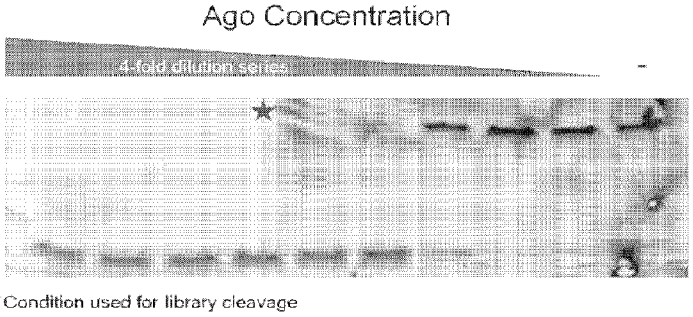
(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR,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

Published:

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

(54) Title: METHODS OF DEPLETING A TARGET MOLECULE FROM AN INITIAL COLLECTION OF NUCLEIC ACIDS, AND COMPOSITIONS AND KITS FOR PRACTICING THE SAME

FIG. 11



(57) Abstract: Provided are methods of depleting a target nucleic acid from an initial collection of nucleic acids. Aspects of the methods include contacting the initial collection with a nucleic acid guided nuclease specific for the target nucleic acid in a manner sufficient to deplete the target nucleic acid from the initial collection. Depending on a given application, depletion of a target nucleic acid may vary, e.g., where depleting may include cleaving a target nucleic acid in, or selectively separating a target nucleic acid from, the initial collection of nucleic acids. Also provided are compositions and kits for practicing embodiments of the methods.

WO 2015/122967 A1

IDT Exhibit 1017.1878

METHODS OF DEPLETING A TARGET MOLECULE FROM AN INITIAL COLLECTION OF NUCLEIC ACIDS, AND COMPOSITIONS AND KITS FOR PRACTICING THE SAME

CROSS REFERENCE TO RELATED APPLICATIONS

Pursuant to 35 U.S.C. § 119 (e), this application claims priority to the filing date of United States Provisional Patent Application Serial No. 61/939,658, filed February 13, 2014 and United States Provisional Patent Application Serial No. 62/040,804, filed August 22, 2014; the disclosure of which applications are herein incorporated by reference.

INTRODUCTION

Applications in biomedical research often involve the analysis of specific subsets of nucleic acids present in a complex mixture of other sequences – for example, analysis of gene expression by array hybridization, qPCR or massively parallel sequencing. If the sequences of nucleic acids of interest are known, PCR with specific primer sequences can be used to amplify the desired sequences out of the mixture. In some cases, however, it may be desired to analyze multiple different sequences, perhaps where sequence information is not fully known. Messenger RNAs in eukaryotic systems, for example, may be collectively amplified and analyzed using an oligo-dT primer to initiate first strand cDNA synthesis by priming on the poly A tail, thereby reducing or avoiding contamination by unwanted nucleic acids – such as ribosomal RNAs, mitochondrial RNAs and genomic DNA. A requirement for this approach, however, is that the RNA is intact and not degraded, e.g., the poly A tails are not lost or disconnected from the body of the RNA message. Unfortunately, many otherwise useful and interesting biological specimens – such as biopsied material retained as formalin-fixed and paraffin embedded tissue samples (FFPE samples) often suffer from such degradation making oligo-dT priming impractical for such samples. Further, many interesting RNA sequences do not have poly A tails – e.g., non-coding RNAs and non-eukaryotic RNAs. In such cases, random priming can be used to generally amplify all nucleotide species in the sample. However, random priming will also result in the amplification of potentially unwanted sequences – such as genomic DNA or ribosomal RNA.

IDT Exhibit 1017.1879

SUMMARY

Provided are methods of depleting a target nucleic acid from an initial collection of nucleic acids. Aspects of the methods include contacting the initial collection with a nucleic acid guided nuclease specific for the target nucleic acid in a manner sufficient to deplete the target nucleic acid from the initial collection. Depending on a given application, depletion of a target nucleic acid may vary, e.g., where depleting may include cleaving a target nucleic acid in, or selectively separating a target nucleic acid from, the initial collection of nucleic acids. Also provided are compositions and kits for practicing embodiments of the methods.

BRIEF DESCRIPTION OF THE FIGURES

FIG. 1 schematically illustrates a nucleic acid guided nuclease that finds use in certain embodiments of the present disclosure.

FIG. 2 schematically illustrates a method for producing a nucleic acid guide component according to one embodiment of the present disclosure.

FIG. 3 shows three example oligonucleotides that find use in the method schematically illustrated in FIG. 2.

FIG. 4 shows an image of template nucleic acids visualized by gel electrophoresis. The template nucleic acids find use in producing nucleic acid guide components by *in vitro* transcription according to one embodiment of the present disclosure.

FIG. 5 shows an image of nucleic acids visualized by gel electrophoresis. The image demonstrates the cleavage of a target nucleic acid using a nucleic acid guided nuclease according to one embodiment of the present disclosure.

FIG. 6 shows an image of nucleic acids visualized by gel electrophoresis. The image demonstrates the simultaneous cleavage of two different target nucleic acids using nucleic acid guided nucleases according to one embodiment of the present disclosure.

FIG. 7 provides data demonstrating the depletion of a target nucleic acid (18S rRNA in this example) in a nucleic acid library for next generation sequencing using a nucleic acid guided nuclease according to one embodiment of the present disclosure. Panel A shows a bar graph indicating the amount of depletion of a target nucleic using one or two example nucleic acid guided nucleases. Panel B shows the amount of depletion of the target nucleic acid using increasing amounts of a nuclease according to one embodiment of the present disclosure. In this example, a pool of nucleic acid guided nucleases is employed, in which the pool includes a

IDT Exhibit 1017.1880

single type of nuclease component and various species of nucleic acid guide components having different nucleotide sequences.

FIG. 8 provides data demonstrating the depletion of a target nucleic acid using a nucleic acid guided nuclease according to one embodiment of the present disclosure. In this example, a pool of nucleic acid guided nucleases is employed, in which the pool includes a single type of nuclease component and various species of nucleic acid guide components having different nucleotide sequences.

FIG. 9 is a bar graph showing sequencing results and demonstrating the effects of depleting a target nucleic acid from a sequencing library using a method according to one embodiment of the present disclosure.

FIG. 10 provides data demonstrating the cleavage of a target nucleic acid using a nucleic acid guided nuclease according to one embodiment of the present disclosure. Panel A shows a gel image demonstrating the purification of an example nuclease component according to one embodiment of the present disclosure. Panel B shows a gel image demonstrating the cleavage of a target nucleic acid using a nucleic acid guided nuclease that includes the nuclease component shown in Panel A.

FIG. 11 shows a gel image demonstrating cleavage of a target nucleic acid using various amounts of a nucleic acid guided nuclease according to one embodiment of the present disclosure.

FIG. 12 provides data demonstrating the expression and purification of a 6xHN tagged D10A/H840A mutant of Cas9 (Panel A), and proof-of-principle of the use of the mutant in combination with a pool of nucleic acid guide components to remove target nucleic acids from an initial collection of nucleic acids and produce a nucleic acid sample enriched for the target nucleic acids (Panel B).

DETAILED DESCRIPTION

Provided are methods of depleting a target nucleic acid from an initial collection of nucleic acids. Aspects of the methods include contacting the initial collection with a nucleic acid guided nuclease specific for the target nucleic acid in a manner sufficient to deplete the target nucleic acid from the initial collection. Depending on a given application, depletion of a target nucleic acid may vary, e.g., where depleting may include cleaving a target nucleic acid in, or selectively separating a target nucleic acid from, the initial collection of nucleic acids. Also provided are compositions and kits for practicing embodiments of the methods.

IDT Exhibit 1017.1881

Before the methods and kits of the present disclosure are described in greater detail, it is to be understood that the methods and kits are not limited to particular embodiments described, as such may, of course, vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to be limiting, since the scope of the methods and kits will be limited only by the appended claims.

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

Certain ranges are presented herein with numerical values being preceded by the term "about." The term "about" is used herein to provide literal support for the exact number that it precedes, as well as a number that is near to or approximately the number that the term precedes. In determining whether a number is near to or approximately a specifically recited number, the near or approximating unrecited number may be a number which, in the context in which it is presented, provides the substantial equivalent of the specifically recited number.

Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which the methods belong. Although any methods and kits similar or equivalent to those described herein can also be used in the practice or testing of the methods and kits, representative illustrative methods and materials are now described.

All publications and patents cited in this specification are herein incorporated by reference as if each individual publication or patent were specifically and individually indicated to be incorporated by reference and are incorporated herein by reference to disclose and describe the methods, kits and/or materials in connection with which the publications are cited. The citation of any publication is for its disclosure prior to the filing date and should not be construed as an admission that the present methods and kits are not entitled to antedate such publication by virtue of prior invention. Further, the dates of publication provided may be different from the actual publication dates which may need to be independently confirmed.

It is noted that, as used herein and in the appended claims, the singular forms "a", "an", and "the" include plural referents unless the context clearly dictates otherwise. It is further noted

IDT Exhibit 1017.1882

that the claims may be drafted to exclude any optional element. As such, this statement is intended to serve as antecedent basis for use of such exclusive terminology as "solely," "only" and the like in connection with the recitation of claim elements, or use of a "negative" limitation.

It is appreciated that certain features of the methods and kits, which are, for clarity, described in the context of separate embodiments, may also be provided in combination in a single embodiment. Conversely, various features of the methods and kits, which are, for brevity, described in the context of a single embodiment, may also be provided separately or in any suitable sub-combination. All combinations of the embodiments are specifically embraced by the present invention and are disclosed herein just as if each and every combination was individually and explicitly disclosed, to the extent that such combinations embrace operable processes and/or devices/systems/kits. In addition, all sub-combinations listed in the embodiments describing such variables are also specifically embraced by the present methods and kits and are disclosed herein just as if each and every such sub-combination was individually and explicitly disclosed herein.

As will be apparent to those of skill in the art upon reading this disclosure, each of the individual embodiments described and illustrated herein has discrete components and features which may be readily separated from or combined with the features of any of the other several embodiments without departing from the scope or spirit of the present methods and kits. Any recited method can be carried out in the order of events recited or in any other order which is logically possible.

In further describing embodiments of the invention, aspects of embodiments of the subject methods will be described first in greater detail. Thereafter, aspects of embodiments of the kits for practicing the subject methods are described in greater detail.

METHODS

Aspects of the invention include methods of selectively depleting a target nucleic acid from an initial collection of nucleic acids. By "depleting" a target nucleic acid is meant reducing the amount of the target nucleic acid in the initial collection of nucleic acids. For example, a target nucleic acid may be depleted by cleavage of the target nucleic acid at one or more locations within the target nucleic acid by one or more nucleic acid guided nuclease(s) in which the nuclease component(s) have nuclease activity. The non-depleted nucleic acids present in the initial collection may then be used in a downstream application of interest, such as nucleic acid amplification, nucleic acid sequencing, gene expression analysis (e.g., by array

IDT Exhibit 1017.1883

hybridization, quantitative RT-PCR, massively parallel sequencing, etc.), or any other downstream application of interest.

Alternatively, a target nucleic acid may be depleted by removal (and optionally, recovery) of the target nucleic acid from the initial collection of nucleic acids. As described in more detail below, in certain aspects, removal of a target nucleic acid from the initial collection is effected using a nucleic acid guided nuclease that includes a cleavage-deficient nuclease, which nuclease may include a heterologous component (e.g., a tag) that facilitates removal of the nuclease (and accordingly, the target nucleic acid to which the nuclease is bound) from the initial collection of nucleic acids. By removing/recovering the target nucleic acids from the initial collection, a subsequent collection of nucleic acids enriched for the target sequences may be obtained. This enriched sample (e.g., an exome-enriched sample, a sample enriched for a panel of genes of interest, etc.) may then be used in a downstream application of interest, such as nucleic acid amplification, nucleic acid sequencing, gene expression analysis (e.g., by array hybridization, quantitative RT-PCR, massively parallel sequencing, etc.), or any other downstream application of interest.

According to certain embodiments, depleting target nucleic acids present in the initial collection includes depleting certain species of target nucleic acids by cleavage of such target nucleic acids, and depleting certain other species of target nucleic acids by removal (and optionally, recovery) of such target nucleic acids from the initial collection of nucleic acids.

A target nucleic acid may vary. By "nucleic acid" is meant a polymer of any length, e.g., 10 bases or longer, 20 bases or longer, 50 bases or longer, 100 bases or longer, 500 bases or longer, 1000 bases or longer, 2000 bases or longer, 3000 bases or longer, 4000 bases or longer, 5000 bases or longer, 10,000 bases or longer, 50,000 or more bases composed of nucleotides, e.g., ribonucleotides or deoxyribonucleotides. In some instances, the length of the nucleic acids is 100,000 bases or less, e.g., 75,000 bases or less, including 50,000 bases or less, e.g., 25,000 bases or less, such as 10,000 bases or less, 5,000 bases or less, 2,000 bases or less, 1,000 bases or less, or 500 bases or less.

Depleting a target nucleic acid from the initial collection partially reduces, if not completely eliminates, the presence of the target nucleic acid in the collection. In some instances, the copy number of a given target nucleic acid in the initial collection of nucleic acids is reduced by 5% or more, such as 10% or more, e.g., 25% or more, including 50%, 75%, 90% or more, including embodiments where the presence of the target nucleic acid is completely eliminated. Depleting a target nucleic acid reduces the percentage of the target nucleic acid in a sample with respect to the total nucleic acid in the sample. In certain aspects, after depletion of

IDT Exhibit 1017.1884

the target nucleic acid, the percent remaining of the target nucleic acid as compared to the initial amount of target nucleic acid in the sample is 50%, 40%, 30%, 25%, 20%, 19%, 18%, 17%, 16%, 15%, 14%, 13%, 12%, 11%, 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2%, 1%, or less, including 0.5%, 0.1%, 0.01% or less. By depleting a target nucleic acid in the initial collection, a type of nucleic acid (e.g., a desirable nucleic acid) may be enriched in the collection. According to certain embodiments, in a sample in which a target nucleic acid has been depleted, a type of nucleic acid (e.g., DNA, mRNA, microRNA (miRNA), and/or the like) is enriched in the sample such that the percentage of the type of nucleic acid remaining in the sample relative to the total is 5% or more, such as 10% or more, 25% or more, 30% or more, 40% or more, 50% or more, 60% or more, 70% or more, 75% or more, including 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, and 99.5% or more.

The initial collection of nucleic acids is contacted with a nucleic acid guided nuclease. As used herein, a "nucleic acid guided nuclease" is an association (e.g., a complex) that includes a nuclease component and a nucleic acid guide component. The nucleic acid guided nuclease may have nuclease/cleavage activity, e.g., catalyzes the hydrolysis of a target nucleic acid (e.g., a target DNA, a target RNA, etc.) into two or more products thereby depleting the target nucleic acid. Cleavage products may be removed from the sample if desired (e.g., to purify the remaining collection of nucleic acids). In certain embodiments, when it is desirable to deplete certain target nucleic acids by cleavage, the number of distinct nucleic acid guided nucleases used and/or the location(s) of the target nucleic acid cleaved by the nucleic acid guided nuclease(s) may be selected such that all or nearly all of the target nucleic acid fragments are sufficiently small to be removed by nucleic acid purification steps such as ethanol or isopropanol precipitation, spin column purification (e.g., using NucleoSpin® Clean-Up columns by Clontech Laboratories, Inc. (Mountain View, CA)), Solid Phase Reversible Immobilization (SPRI) beads, or the like. When the nuclease component has nuclease activity, the nuclease may generate double-stranded breaks in the target nucleic acid, or the nuclease may be a nuclease that introduces a break in a single strand of a double-stranded target nucleic acid (e.g., the nuclease component may be a nickase).

In certain aspects, the nuclease is a modified nuclease that does not have nuclease activity (e.g., is cleavage deficient) as a result of the modification. Such nucleases may be employed to deplete the target nucleic acid, e.g., upon removal of the target nucleic acid present in a complex formed between the nucleic acid guided nuclease and the target nucleic acid, from the initial collection of nucleic acids, which in certain aspects is facilitated by a tag (e.g., an epitope tag) provided on the nuclease.

IDT Exhibit 1017.1885

Any suitable nuclease component may be employed by a practitioner of the subject methods. The nuclease component may be a wild-type enzyme that exhibits nuclease activity, or a modified variant thereof that retains its nuclease activity. In other aspects, the nuclease component may be a non-nuclease protein operatively linked to a heterologous nuclease (or "cleavage") domain, such that the protein is capable of cleaving the target nucleic acid by virtue of being linked to the nuclease domain. Suitable cleavage domains are known and include, e.g., the DNA cleavage domain of the FokI restriction endonuclease. For example, in certain aspects, the nuclease component of a nucleic acid guided nuclease may be a Cas9 (e.g., a wild-type Cas9 or cleavage deficient Cas9) or other nuclease operably linked to a cleavage domain, such as a FokI cleavage domain. According to certain embodiments, the nuclease is a mutant that is cleavage deficient – e.g., Sp, a Cas9 D10A mutant, a Cas9 H840A mutant, a Cas9 D10A/H840A mutant (see, e.g., Sander & Joung (2014) *Nature Biotechnology* 32:347–355 doi:10.1038/nbt.2842), or any other suitable cleavage deficient mutant. Such a strategy has been successfully employed to confer nuclease activity upon zinc finger and transcription-activator-like effector (TALE) proteins to generate zinc finger nucleases and TALENs, respectively, for genomic engineering purposes (see, e.g., Kim et al. (1996) *PNAS* 93(3):1156-1160, and US Patent Application Publication Numbers US2003/0232410, US2005/0208489, US2006/0188987, US2006/0063231, and US2011/0301073).

According to certain embodiments, the nuclease domain is derived from an endonuclease. Endonucleases from which a nuclease/cleavage domain can be derived include, but are not limited to: a Cas nuclease (e.g., a Cas9 nuclease), an Argonaute nuclease (e.g., Tth Ago, mammalian Ago2, etc.), S1 Nuclease; mung bean nuclease; pancreatic DNase I; micrococcal nuclease; yeast HO endonuclease; a restriction endonuclease; a homing endonuclease; and the like; see also Mishra (Nucleases: Molecular Biology and Applications (2002) ISBN-10: 0471394610). In certain aspects, the nuclease component of the nucleic acid guided nuclease is a Cas9 nuclease of *Francisella novicida* (or any suitable variant thereof), which uses a scaRNA to target RNA for degradation (see Sampson et al. (2013) *Nature* 497:254-257).

As described above, according to certain embodiments, the nucleic acid guided nuclease includes a CRISPR-associated (or "Cas") nuclease. The CRISPR/Cas system is an RNA-mediated genome defense pathway in archaea and many bacteria having similarities to the eukaryotic RNA interference (RNAi) pathway. The pathway arises from two evolutionarily (and often physically) linked gene loci: the CRISPR (clustered regularly interspaced short

IDT Exhibit 1017.1886

palindromic repeats) locus, which encodes RNA components of the system; and the Cas (CRISPR-associated) locus, which encodes proteins.

There are three types of CRISPR/Cas systems which all incorporate RNAs and Cas proteins. The Type II CRISPR system carries out double-strand breaks in target DNA in four sequential steps. First, two non-coding RNAs (the pre-crRNA array and tracrRNA), are transcribed from the CRISPR locus. Second, tracrRNA hybridizes to the repeat regions of the pre-crRNA and mediates the processing of pre-crRNA into mature crRNAs containing individual spacer sequences. Third, the mature crRNA:tracrRNA complex directs Cas9 to the target DNA via Watson-Crick base-pairing between the spacer on the crRNA and the protospacer on the target DNA next to the protospacer adjacent motif (PAM), an additional requirement for target recognition. Finally, Cas9 mediates cleavage of target DNA to create a double-stranded break within the protospacer.

CRISPR systems Types I and III both have Cas endonucleases that process the pre-crRNAs, that, when fully processed into crRNAs, assemble a multi-Cas protein complex that is capable of cleaving nucleic acids that are complementary to the crRNA. In type II CRISPR/Cas systems, crRNAs are produced by a mechanism in which a trans-activating RNA (tracrRNA) complementary to repeat sequences in the pre-crRNA, triggers processing by a double strand-specific RNase III in the presence of the Cas9 protein. Cas9 is then able to cleave a target DNA that is complementary to the mature crRNA in a manner dependent upon base-pairing between the crRNA and the target DNA, and the presence of a short motif in the crRNA referred to as the PAM sequence (protospacer adjacent motif).

The requirement of a crRNA-tracrRNA complex can be avoided by use of an engineered fusion of crRNA and tracrRNA to form a "single-guide RNA" (sgRNA) that comprises the hairpin normally formed by the annealing of the crRNA and the tracrRNA. See, e.g., Jinek et al. (2012) *Science* 337:816-821; Mali et al. (2013) *Science* 339:823-826; and Jiang et al. (2013) *Nature Biotechnology* 31:233-239. The sgRNA guides Cas9 to cleave target DNA when a double-stranded RNA:DNA heterodimer forms between the Cas-associated RNAs and the target DNA. This system, including the Cas9 protein and an engineered sgRNA containing a PAM sequence, has been used for RNA guided genome editing with editing efficiencies similar to ZFNs and TALENs. See, e.g., Hwang et al. (2013) *Nature Biotechnology* 31 (3):227.

According to certain embodiments, the nuclease component of the nucleic acid guided nuclease is a CRISPR-associated protein, such as a Cas protein. Non-limiting examples of Cas proteins include Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas9 (also known as Csn1 and Csx12), Cas10, Csy1, Csy2, Csy3, Cse1, Cse2, Csc1, Csc2, Csa5, Csn2, Csm2,

IDT Exhibit 1017.1887

Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, Csf4, homologues thereof, or modified versions thereof. In certain aspects, the nuclease component of the nucleic acid guided nuclease is Cas9. The Cas9 may be from any organism of interest, including but not limited to, *Streptococcus pyogenes* ("spCas9", Uniprot Q99ZW2) having a PAM sequence of NGG; *Neisseria meningitidis* ("nmCas9", Uniprot C6S593) having a PAM sequence of NNNNGATT; *streptococcus thermophilus* ("stCas9", Uniprot Q5M542) having a PAM sequence of NNAGAA, and *Treponema denticols* ("tdCas9", Uniprot M2B9U0) having a PAM sequence of NAAAAC. An example nucleic acid guided nuclease that includes a Cas9 nuclease and an sgRNA guide component, in which the sgRNA guide component is aligned with a complementary region of a generalized target nucleic acid, is schematically illustrated in FIG. 1.

In certain aspects, the nuclease component of the nucleic acid guided nuclease is an Argonaute (Ago) nuclease. Ago proteins are a family of evolutionarily conserved proteins central to the RNA interference (RNAi) platform and microRNA (miRNA) function and biogenesis. They are best known as core components of the RNA-induced silencing complex (RISC) required for small RNA-mediated gene regulatory mechanisms. In post-transcriptional gene silencing, Ago guided by a small RNA (e.g., siRNA, miRNA, piRNA, etc.) binds to the complementary transcripts via base-pairing and serve as platforms for recruiting proteins to facilitate gene silencing.

Mammals have eight Argonaute proteins, which are divided into two subfamilies: the Piwi clade and the Ago clade. Of the wild-type Ago proteins (Ago1-4, or EIF2C1–4), only Ago2 has endonuclease activity. The crystal structure of full-length human Ago2 (Uniprot Q9UKV8) has been solved. See, e.g., Elkayam et al. (2012) *Cell* 150(1):100-110. Similar to the bacteria counterpart, human Ago2 is a bilobular structure comprising the N-terminal (N), PAZ, MID, and PIWI domains. The PAZ domain anchors the 3′end of the small RNAs and is dispensable for the catalytic activity of Ago2. However, PAZ domain deletion disrupts the ability of the non-catalytic Agos to unwind small RNA duplex and to form functional RISC.

When the nuclease component of the nucleic acid guided nuclease is an Ago nuclease, the nuclease may be an Ago nuclease that cleaves DNA duplexes, RNA duplexes, or DNA-RNA duplexes. The Ago nuclease may be derived from any suitable organism, such as a prokaryotic or eukaryotic organism. In certain aspects, the Ago is a prokaryotic Ago. Prokaryotic Agos of interest include, but are not limited to, *Thermus thermophiles* Ago ("Tth Ago"), such as the Tth Ago nucleases described in Wang et al. (2008) *Nature* 456(7224):921-926; and Wang et al. (2009) *Nature* 461(7265):754-761. DNA-guided DNA interference *in vivo* using Tth Ago and 5′-

IDT Exhibit 1017.1888

phosphorylated DNA guides of from 13-25 nucleotides in length was recently described by Swarts et al. (2014) *Nature* 507:258-261.

The nucleic acid guided nuclease may include a nuclease having nuclease activity (e.g., catalyzes the hydrolysis of a target nucleic acid (e.g., a target DNA, a target RNA, etc.)), or may be a modified nuclease that does not have nuclease activity (e.g., is cleavage deficient) as a result of the modification. In some instances, the nuclease component (e.g., a Cas nuclease component) is a cleavage deficient mutant and the method results in the production of a product composition comprising target nucleic acid/nucleic acid guided nuclease complexes. When part of the resultant complex, the target nucleic acid is no longer free in the collection of nucleic acids, and therefore has been depleted from the initial collection of nucleic acids. In other aspects, the nuclease component of the complex is a cleavage competent nuclease, but the nucleic acid guided nuclease remains bound to a fragment of the target nucleic acid subsequent to cleavage of the target nucleic acid. In some instances, the method further includes separating (e.g., removing) target nucleic acid/nucleic acid guided nuclease complexes (e.g., including a cleavage competent nuclease component and/or a cleavage deficient nuclease component such as a D10A Cas9 mutant, a H840A Cas9 mutant, a D10A/H840A Cas9 mutant, and/or the like) from other constituents of the product composition. Where desired, the nuclease component may include a tag, e.g., an epitope tag, FLAG tag, HA tag, His tag, Myc tag, S-tag, SBP tag, Softag, GST tag, GFP tag, biotin, streptavidin, 6-His tag, etc., e.g., to facilitate separation of the complexes (e.g., by affinity purification) from the other components of the initial collection.

According to certain embodiments, when the method involves the formation of target nucleic acid/nucleic acid guided nuclease complexes, the method further includes recovering the target nucleic acids from the complexes. Any suitable strategy for recovering the target nucleic acids may be employed. Such strategies may include separating the complexes from other constituents of the composition, and then disassociating the target nucleic acids from the nucleic acid guided nucleases. In certain aspects, the nuclease component of the nucleic acid guided nuclease includes a tag (e.g., an epitope tag), and the complexes may be separated from other constituents by affinity purification. For example, the complexes may be immobilized on the surface of a solid phase (e.g., a column, a plate, beads (e.g., agarose or magnetic beads), and/or the like) that includes a binding partner of the tag (e.g., an antibody or other suitable binding partner that binds the tag), and then washed to remove any residual constituents of the composition. The target nucleic acids may then be recovered from the nucleic acid guided nucleases using a suitable elution buffer (e.g., a buffer that includes a

IDT Exhibit 1017.1889

protein denaturation agent, such as sodium dodecyl sulfate (SDS)), using a buffer that includes a reagent that digests the nuclease component (e.g., proteinase K), using heat denaturation, and/or the like, to disrupt the interactions between the target nucleic acids and the nucleic acid guided nuclease. Approaches for affinity purification and recovering nucleic acids from protein complexes are described, e.g., in *Methods for Affinity-Based Separations of Enzymes and Proteins* (Munishwar Nath Gupta, ed., Birkhäuser Verlag, Basel-Boston-Berlin, 2002); *Chromatin Immunoprecipitation Assays: Methods and Protocols* (Collas, ed., 2009); and *The Protein Protocols Handbook* (Walker, ed., 2002). If desired, the separated target nucleic acids may be further purified by alcohol precipitation, column purification, or any other convenient nucleic acid purification strategy.

As summarized above, in certain aspects, the nucleic acid guided nuclease includes a nucleic acid guide component. Any suitable nucleic acid guide component capable of guiding the nuclease component to the target nucleic acid may be employed. The nucleic acid guide component may be single-stranded or double-stranded as appropriate for the particular nuclease component employed.

The nucleic acid guide component may be one or more nucleic acid polymers of any suitable length. In certain aspects, the nucleic acid guide component is a nucleic acid polymer (e.g., a single- or double-stranded RNA or DNA) of from 10 to 200 nucleotides in length, such as from 10 to 150 nucleotides in length, including from 10 to 100, from 10 to 90, from 10 to 80, from 10 to 70, from 10 to 60, from 10 to 50, from 10 to 40, from 10 to 30, from 10 to 25, from 10 to 20, or from 10 to 15 nucleotides in length.

At least a portion of the nucleic acid guide component is complementary (e.g., 100% complementary or less than 100% complementary) to at least a portion of a target nucleic acid of interest. The sequence of all or a portion of the nucleic acid guide component may be selected by a practitioner of the subject methods to be sufficiently complementary to a target nucleic acid of interest to specifically guide the nuclease component to the target nucleic acid. The nucleic acid sequences of target nucleic acids of interest are readily available from resources such as the nucleic acid sequence databases of the National Center for Biotechnology Information (NCBI), the European Molecular Biology Laboratory-European Bioinformatics Institute (EMBL-EBI), and the like. By way of example, when the target nucleic acid(s) of interest is one or both of human 18S rRNA (1.9 kb) and/or human 28S rRNA, the nucleotide sequences of the 18S rRNA are readily available as those of NCBI reference sequences NR_003286.2 and NR_003287.2, respectively.

IDT Exhibit 1017.1890

Once a target nucleic acid is selected, and based on the available sequence information for the target nucleic acid, a nucleic acid guide component may be designed such that all or a portion of the nucleic acid guide component is sufficiently complementary to a target region of the target nucleic acid to specifically guide the nucleic acid guided nuclease under hybridization conditions to the target region of the target nucleic acid, e.g., for cleavage at the target region by the nuclease to deplete the target nucleic acid.

"Hybridization conditions" may include conditions in which the nucleic acid guide component specifically hybridizes to a target region of the target nucleic acid, interactions between the target nucleic acid and nuclease component, or both. Whether a nucleic acid guide component specifically hybridizes to a target nucleic acid is determined by such factors as the degree and length of complementarity between the nucleic acid guide component and the target nucleic acid, and the temperature at which the hybridization/contacting occurs, which may be informed by the melting temperature (T_M) of the region of the nucleic acid guide component that is complementary to the target region of the target nucleic acid. The melting temperature refers to the temperature at which half of the nucleic acid guide component-target nucleic acid duplexes remain hybridized and half of the duplexes dissociate into single strands. The T_m of a duplex may be experimentally determined or predicted using the following formula $T_m = 81.5 + 16.6(\log_{10}[Na^+]) + 0.41$ (fraction G+C) – (60/N), where N is the chain length and $[Na^+]$ is less than 1 M. *See* Sambrook and Russell (2001; Molecular Cloning: A Laboratory Manual, 3rd ed., Cold Spring Harbor Press, Cold Spring Harbor N.Y., Ch. 10). Other more advanced models that depend on various parameters may also be used to predict T_m of nucleic acid guide component-target nucleic acid duplexes depending on various hybridization conditions. Approaches for achieving specific nucleic acid hybridization may be found in, e.g., Tijssen, Laboratory Techniques in Biochemistry and Molecular Biology-Hybridization with Nucleic Acid Probes, Part I, chapter 2, "Overview of principles of hybridization and the strategy of nucleic acid probe assays," Elsevier (1993).

According to certain embodiments, the nucleic acid guide component is an RNA guide component (or "guide RNA"). The RNA guide component may include one or more RNA molecules. For example, the RNA guide component may include two separately transcribed RNAs (e.g., a crRNA and a tracrRNA) which form a duplex that guides the nuclease component (e.g., Cas9) to the target nucleic acid. In other aspects, the RNA guide component is a single RNA molecule, which may correspond to a wild-type single guide RNA, or alternatively, may be an engineered single guide RNA. According to certain embodiments, the nucleic acid guide

IDT Exhibit 1017.1891

component is an engineered single guide RNA that includes a crRNA portion fused to a tracrRNA portion, which single guide RNA is capable of guiding a nuclease (e.g., Cas9) to the target nucleic acid.

In certain aspects, the nucleic acid guide component is a DNA guide component, e.g., a single-stranded or double-stranded guide DNA. According to certain embodiments, the guide DNA is phosphorylated at one or both ends. For example, the guide DNA may be a 5'-phosphorylated guide DNA oligonucleotide of any suitable length (e.g., any of the lengths set forth above, including for example, from 10 to 30 nucleotides in length). The present inventors have demonstrated that nucleic acid guided nucleases that include such phosphorylated guide DNA oligonucleotides and Tth Ago efficiently deplete a target nucleic acid of interest from an initial collection of nucleic acids based on complementarity between the guide DNA oligonucleotide and the target nucleic acid of interest (see, e.g., the Examples section herein).

As summarized above, the methods of the present disclosure include contacting an initial collection of nucleic acids with a nucleic acid guided nuclease specific for the target nucleic acid of interest in a manner sufficient to deplete the target nucleic acid from the initial collection. In certain aspects, contacting the initial collection of nucleic acids with a nucleic acid guided nuclease includes combining in a reaction mixture the initial collection of nucleic acids, a nucleic acid guide component, and a nuclease component. The nucleic acid guide component and the nuclease component may be stably associated (e.g., as a complex) prior to being added to the reaction mixture, or these components may be added separately for subsequent association with each other and targeting/depletion of the target nucleic acid. In certain aspects, at least a portion of the contacting step occurs under conditions in which the nuclease component is active and able to cleave the target nucleic acid.

The conditions under which the initial collection of nucleic acids is contacted with the nucleic acid guided nuclease may vary. For example, the conditions may include a temperature at which the nucleic acid guide component specifically hybridizes to the target nucleic acid, such as from 0°C to 10°C (e.g., 4°C), from 10°C to 20°C (e.g., 16°C), from 20°C to 30°C (e.g., 25°C), from 30°C to 40°C (e.g., 37°C), from 40°C to 50°C, from 50°C to 60°C, from 60°C to 70°C, or from 70°C to 80°C. Factors and approaches for achieving specific hybridization between the nucleic acid guide component and the target nucleic acid are described hereinabove. In certain aspects, nucleic acids of the initial collection of nucleic acids are denatured (e.g., heat-denatured) to generate single-stranded nucleic acids prior to the contacting step to facilitate hybridization of the nucleic acid guide component to the target nucleic acid.

IDT Exhibit 1017.1892

According to embodiments in which the nuclease component cleaves the target nucleic acid, the contacting conditions may include a temperature at which the particular nuclease employed is active, e.g., has nuclease activity. Such temperatures may vary, and in certain aspects include temperatures from 0°C to 10°C (e.g., 4°C), from 10°C to 20°C (e.g., 16°C), from 20°C to 30°C (e.g., 25°C), from 30°C to 40°C (e.g., 37°C), from 40°C to 50°C, from 50°C to 60°C, from 60°C to 70°C, or from 70°C to 80°C.

The nuclease activity of certain nucleases depends on the presence of one or more cofactors. When such a nuclease component is employed to practice the subject methods, the contacting conditions may include providing any necessary cofactors to the reaction mixture. In certain aspects, the cofactor(s) is one or more divalent cations, such as Mg^{2+}, Mn^{2+}, Ca^{2+}, and/or the like.

The reaction mixture may include one or more buffers (e.g., a Tris buffer, a PBS buffer, or the like) to ensure that the contacting occurs a suitable pH, e.g., at which the nuclease exhibits nuclease activity. For example, the contacting conditions may include the pH of the reaction mixture being from pH 4.5 to 8.5, such as from 4.5 to 5.5, from 5.5 to 6.5, from 6.5 to 7.5, or from 7.5 to 8.5.

The contacting step may be performed such that the final concentrations of the initial collection of nucleic acids, the nucleic acid guide component, and the nuclease component are suitable to deplete the target nucleic acid. For example, the final concentration of the initial collection of nucleic acids may be from 0.1 pg/µl to 10 µg/µl, the final concentration of the nucleic acid guide component may be from 25 pM to 50 µM, and the final concentration of the nuclease component may be from 25 pM to 50 µM.

Aspects of the invention include methods of making a nucleic acid guided nuclease, e.g., any of the nucleic acid guided nucleases described elsewhere herein. Approaches for making the nucleic acid guided nuclease may vary. In certain aspects, the methods include expressing a nucleic acid guide component and a nuclease component from the same or different expression plasmids. Plasmids and associated protocols for expressing a nucleic acid guide component and/or a nuclease component are commercially available and include, e.g., the GeneArt® CRISPR nuclease vectors (Life Technologies, Carlsbad, CA).

According to certain embodiments, the present disclosure provides PCR-based methods of producing a nucleic acid guide component specific for a target nucleic acid of interest. One embodiment of such methods is schematically illustrated in FIG. 2. As shown, a user may design an oligonucleotide primer (shown here as the forward "F" primer) that includes: a sequence complementary to a promoter sequence of interest (e.g., a T7, U6, T3 or other

IDT Exhibit 1017.1893

promoter); a sequence complementary to an sgRNA guide sequence specific for a target nucleic acid of interest (e.g., specific for a target cDNA transcribed from an rRNA); and a sequence complementary to at least a portion of an sgRNA scaffold sequence. This forward primer may be used in conjunction with an oligonucleotide primer (shown here as the reverse "R" primer) having a sequence complementary to at least a portion of the sgRNA scaffold sequence and any other useful sequences (e.g., a poly dT tract) for producing an sgRNA. PCR amplification using these primers and a template that includes the scaffold and any other desirable elements produces a template (designated in FIG. 2 as the "DNA template of sgRNA") from which a particular sgRNA may be transcribed by *in vitro* transcription and employed in conjunction with a nuclease for, e.g., selective depletion of a target nucleic acid according to the methods of the present disclosure. Shown in FIG. 3 are example forward (T7-T1-AcGFP and T7-T2-AcGFP) and reverse (T7-Rev) oligonucleotides for producing templates from which sgRNAs may be produced by *in vitro* transcription, e.g., according to the embodiment shown in FIG. 2. The sequence outlined with dashed rectangles is the T7 promoter sequence. Underlined sequences are 20 bp crRNA sequences. A polyA sequence is outlined with a solid rectangle.

The initial collection of nucleic acids may vary. Examples of initial collection of nucleic acids of interest include collections of double-stranded nucleic acids (e.g., double-stranded DNA), collections of single stranded nucleic acids (e.g., single-stranded RNA or DNA), mixed collections of double and single stranded nucleic acids, etc. The complexity of the initial collection may also vary, where in some instances the collection includes 5 or more, 10 or more, 25 or more, 50 or more, 100 or more, 250 or more, 500 or more, 1000 or more, 5,000 or more, including 10,000, 100,000 or more, 500,000 or more, 1 million or more, 100 million or more, or 1 billion or more distinct nucleic acids of differing sequence. The initial collection of nucleic acids may include deoxyribonucleic acids, ribonucleic acids, or mixtures thereof.

In certain aspects, the initial collection of nucleic acids of interest is a collection of nucleic acids (e.g., undesired and desired nucleic acids) isolated from a nucleic acid source of interest, including but not limited to, a nucleic acid sample isolated from a single cell, a plurality of cells (e.g., cultured cells), a tissue, an organ, or an organism, e.g., bacteria, yeast, or a collection of organisms (such as a metagenomic sample, e.g., sea water containing multiple organisms, a fecal sample containing many distinct bacterial species, a buccal swab, etc.), or the like. The term "sample", as used herein, relates to a material or mixture of materials, typically, although not necessarily, in liquid form, containing nucleic acids and/or proteins which one desires to deplete from an initial collection (e.g., by cleavage or removal as described elsewhere herein). In certain aspects, the nucleic acid sample is isolated from a cell(s), tissue,

IDT Exhibit 1017.1894

organ, and/or the like of a mammal (e.g., a human, a rodent (e.g., a mouse), or any other mammal of interest). In other aspects, the nucleic acid sample is isolated from a source other than a mammal, such as bacteria, yeast, insects (e.g., drosophila), amphibians (e.g., frogs (e.g., *Xenopus*)), viruses, plants, or any other non-mammalian nucleic acid sample source. According to certain embodiments, the initial collection of nucleic acids of interest is not genomic DNA.

Any convenient protocol for isolating nucleic acids from such sources, as well as reagents and kits designed for isolating nucleic acids from such sources, may be employed. For example, kits for isolating nucleic acids from a source of interest – such as the NucleoSpin®, NucleoMag® and NucleoBond® genomic DNA or RNA isolation kits by Clontech Laboratories, Inc. (Mountain View, CA) – are commercially available. In certain aspects, the nucleic acid is isolated from a fixed biological sample, e.g., formalin-fixed, paraffin-embedded (FFPE) tissue. Nucleic acids from FFPE tissue may be isolated using commercially available kits – such as the NucleoSpin® FFPE DNA or RNA isolation kits by Clontech Laboratories, Inc. (Mountain View, CA).

According to certain embodiments, the initial collection of nucleic acids of interest is produced from a precursor collection of nucleic acids of interest. For example, the initial collection of nucleic acids of interest may be a collection of DNAs (e.g., cDNAs) transcribed from a precursor collection of nucleic acids of interest (e.g., RNAs). In certain aspects, the initial collection of nucleic acids of interest is a collection of cDNAs transcribed from a precursor collection of RNAs of interest, where the precursor collection of RNAs of interest include mRNAs, miRNAs, rRNAs, and/or the like, and the target nucleic acid to be depleted is cDNA transcribed from rRNAs present in the precursor collection of nucleic acids.

Generating a collection of cDNAs of interest from a precursor collection of RNAs of interest may include carrying out a reverse transcription reaction by combining the precursor collection of RNAs of interest with a suitable polymerase, dNTPs, buffer components that establish an appropriate pH, one or more salts (e.g., KCl), one or more metal cofactors (e.g., Mg^{2+} or Mn^{2+}), and the like, under conditions suitable for a polymerase-mediated extension reaction to occur. Other components may be included, such as one or more nuclease inhibitors (e.g., an RNase inhibitor and/or a DNase inhibitor), one or more additives for facilitating amplification/replication of GC rich sequences (e.g., GC-Melt™ reagent (Clontech Laboratories, Inc. (Mountain View, CA)), betaine, single-stranded binding proteins (e.g., T4 Gene 32, cold shock protein A (CspA), and/or the like), DMSO, ethylene glycol, 1,2-propanediol, or combinations thereof), one or more molecular crowding agents (e.g., polyethylene glycol, or the like), one or more enzyme-stabilizing components (e.g., DTT present at a final concentration

IDT Exhibit 1017.1895

ranging from 1 to 10 mM (e.g., 5 mM)), and/or any other reaction mixture components useful for facilitating polymerase-mediated extension reactions.

Polymerases that find use in generating a collection of cDNAs of interest from a precursor collection of RNAs of interest include, but are not limited to, reverse transcriptases, such as a retroviral reverse transcriptase, retrotransposon reverse transcriptase, retroplasmid reverse transcriptases, retron reverse transcriptases, bacterial reverse transcriptases, group II intron-derived reverse transcriptase, and mutants, variants derivatives, or functional fragments thereof. For example, the reverse transcriptase may be a Moloney Murine Leukemia Virus reverse transcriptase (MMLV RT) or a Bombyx mori reverse transcriptase (e.g., Bombyx mori R2 non-LTR element reverse transcriptase). In certain aspect, the polymerase is capable of template switching. Template switching polymerases are commercially available and include SMARTScribe™ reverse transcriptase and PrimeScript™ reverse transcriptase available from Clontech Laboratories, Inc. (Mountain View, CA). In certain aspects, a mix of two or more different polymerases is added to the reaction mixture, e.g., for improved processivity, proof-reading, and/or the like. In certain aspects, the polymerase (e.g., a reverse transcriptase such as an MMLV RT or a Bombyx mori RT) is present in the reaction mixture at a final concentration of from 0.1 to 200 units/μL (U/μL), such as from 0.5 to 100 U/μL, such as from 1 to 50 U/μL, including from 5 to 25 U/μL, e.g., 20 U/μL.

In certain aspects, the initial collection of nucleic acids of interest is produced from a precursor collection of nucleic acids of interest by shearing/fragmenting the precursor collection of nucleic acids of interest, e.g., when it is desirable to control the size of the nucleic acids in the initial collection of nucleic acids of interest. Shearing/fragmentation strategies include, but are not limited to, passing a precursor collection of nucleic acids of interest one or more times through a micropipette tip or fine-gauge needle, nebulizing the sample, sonicating the sample (e.g., using a focused-ultrasonicator by Covaris, Inc. (Woburn, MA)), bead-mediated shearing, enzymatic shearing (e.g., using one or more DNA- or RNA-shearing enzymes), chemical based fragmentation, e.g., using divalent cations (e.g., Mg^{2+}, Mn^{2+}, and/or Zn^{2+}), fragmentation buffer (e.g., a high pH buffer), and/or heat, or any other suitable approach for shearing/fragmenting a precursor collection of nucleic acids of interest to generate a shorter initial collection of nucleic acids of interest. In certain aspects, the initial collection of nucleic acids of interest generated by shearing/fragmentation has a length of from 50 to 10,000 nucleotides, from 100 to 5000 nucleotides, from 150 to 2500 nucleotides, from 200 to 1000 nucleotides, e.g., from 250 to 500 nucleotides in length, for example.

IDT Exhibit 1017.1896

According to certain embodiments, the initial collection of nucleic acids of interest includes nucleic acids (e.g., double-stranded DNA, such as double-stranded cDNA) having one or more sequencing adapter constructs at one or both ends of the nucleic acids. By "sequencing platform adapter construct" is meant a nucleic acid construct that includes at least a portion of a nucleic acid domain (e.g., a sequencing platform adapter nucleic acid sequence) or complement thereof utilized by a sequencing platform of interest, such as a sequencing platform provided by Illumina® (e.g., the HiSeq™, MiSeq™ and/or Genome Analyzer™ sequencing systems); Ion Torrent™ (e.g., the Ion PGM™ and/or Ion Proton™ sequencing systems); Pacific Biosciences (e.g., the PACBIO RS II sequencing system); Life Technologies™ (e.g., a SOLiD sequencing system); Roche (e.g., the 454 GS FLX+ and/or GS Junior sequencing systems); or any other sequencing platform of interest. Such an initial collection of nucleic acids finds use, e.g., when it is desirable to determine the sequence(s) of nucleic acids present in the initial collection of nucleic acids using a sequencing platform.

When the initial collection of nucleic acids (e.g., cDNAs) includes sequencing platform adapter constructs, the methods of the present disclosure find use in depleting one or more subpopulations of target nucleic acids in the initial collection (e.g., undesirable sequences such as cDNAs transcribed from rRNAs or particular subtypes thereof; or desirable sequences which are depleted by removal from the initial collection, recovered to produce a sample enriched for the desirable nucleic acids), followed by sequencing the desirable nucleic acids. On sequencing platforms that utilize adapter sequences at both ends of a nucleic acid to be sequenced (e.g., an Illumina®- Ion Torrent™-based platform), a single cleavage event by a nucleic acid guided nuclease in a target nucleic acid (e.g., a cDNA transcribed from an rRNA) renders the fragmented target nucleic acid invisible to the sequencer.

According to certain embodiments, the sequencing platform adapter construct includes a nucleic acid domain selected from: a domain (e.g., a "capture site" or "capture sequence") that specifically binds to a surface-attached sequencing platform oligonucleotide (e.g., the P5 or P7 oligonucleotides attached to the surface of a flow cell in an Illumina® sequencing system); a sequencing primer binding domain (e.g., a domain to which the Read 1 or Read 2 primers of the Illumina® platform may bind); a barcode domain (e.g., a domain that uniquely identifies the sample source of the nucleic acid being sequenced to enable sample multiplexing by marking every molecule from a given sample with a specific barcode or "tag"); a barcode sequencing primer binding domain (a domain to which a primer used for sequencing a barcode binds); a molecular identification domain (e.g., a molecular index tag, such as a randomized tag of 4, 6, or other number of nucleotides) for uniquely marking molecules of interest to determine

IDT Exhibit 1017.1897

expression levels based on the number of instances a unique tag is sequenced; a complement of any such domains; or any combination thereof. In certain aspects, a barcode domain (e.g., sample index tag) and a molecular identification domain (e.g., a molecular index tag) may be included in the same nucleic acid.

In some instances, the subject methods include contacting the initial collection of nucleic acids with a plurality (e.g., a "pool" or "library") of two or more distinct nucleic acid guided nucleases. The nucleic acid guided nucleases may be distinct in any desired respects. For example, the methods may employ a pool of nucleic acid guided nucleases in which the pool includes a single type of nuclease component (e.g., a nuclease which may have nuclease activity or, alternatively, be cleavage deficient) and two or more species of nucleic acid guide components having different nucleotide sequences. The two or more species of nucleic acid guide components may be designed such that the resulting different nucleic acid guided nucleases target multiple regions of a single target nucleic acid, target multiple different target nucleic acids, target multiple regions of multiple different target nucleic acids, etc.

Alternatively, or additionally, the plurality of two or more distinct nucleic acid guided nucleases may include two or more types of nuclease components. For example, the methods may employ a pool/library of any desired combination of nucleases that differ from one another with respect to the origin of the nuclease (e.g., nucleases from different prokaryotic and/or eukaryotic species), nucleases that differ from one another with respect to nuclease activity (e.g., the pool/library may include one or more nucleases that have nuclease activity, one or more nucleases that are cleavage deficient, one or more nickases, or any combination thereof), PAM sequence (e.g., the pool/library may include nucleases that utilize guide nucleic acids having different PAM sequences), and any other combination of nuclease components suitable for depleting one or more target nucleic acids from the initial collection. As set forth above, a pool of different nuclease components may be used in conjunction with a pool of different nucleic acid guide components to achieve a desired level of depletion of a desired number of target nucleic acids. The depletion may include cleaving target nucleic acids present in the initial collection (e.g., one or more ribosomal and/or mitochondrial RNAs), removing target nucleic acids from the initial collection (e.g., to produce a nucleic acid sample enriched for the target nucleic acids removed from the initial collection), or both.

As such, aspects of the present disclosure include methods of selectively depleting a subpopulation of nucleic acids (e.g., rRNA- and/or mtRNA-derived nucleic acids) from an initial collection of nucleic acids. In such embodiments, the methods may include contacting the initial collection of nucleic acids with a library of nucleic acid guided nucleases in a manner sufficient

IDT Exhibit 1017.1898

to deplete the subpopulation from the initial collection, where the library includes two or more distinct nucleic acid guided nucleases specific for two or more members and/or multiple regions of a single member of the subpopulation of nucleic acids. In these embodiments, the size of the library may vary. For example, the library may include 2 or more, 3 or more, 4 or more, 5 or more, 10 or more, 25 or more, 100 or more, 500 or more, 1000 or more, 10000 or more, 50000 or more, 100000 or more, 500000 or more, or 1 million or more distinct nucleic acid guided nucleases, etc. In certain aspects, the library includes two or more, but 1 million or less, 500000 or less, 100000 or less, 50000 or less, 10000 or less, 1000 or less, 500 or less, 100 or less, 25 or less, 10 or less, 5, 4, or 3 distinct nucleic acid guided nucleases.

Aspects of the present disclosure further include methods of making a library of nucleic acid guided nucleases. Such methods may include producing a plurality of distinct nucleic acid guide components using the PCR-based approach shown in FIG. 2 and described above. In certain embodiments, the methods include combining a plurality of distinct guide nucleic acids with a one or more distinct nucleases in a manner sufficient to produce the library of nucleic acid guided nucleases. The nuclease may vary, and in some instances is a Cas nuclease (e.g., Cas9) or Ago nuclease, which independently may or may not have cleavage activity. The size of the produced library may vary, and in some instances includes 2 or more, 3 or more, 4 or more, 5 or more, 10 or more, 25 or more, 100 or more, 500 or more, 1000 or more, 10,000 or more, 50,000 or more, 100,000 or more, 500,000 or more, or 1 million or more distinct nucleic acid guided nucleases. In certain aspects, the produced library includes two or more, but 1 million or less, 500,000 or less, 100,000 or less, 50,000 or less, 10,000 or less, 1000 or less, 500 or less, 100 or less, 25 or less, 10 or less, 5, 4, or 3 distinct nucleic acid guided nucleases. In some instances, the nucleic acid guides include separate crRNA and tracrRNA, or a component that includes functional elements thereof, e.g., an sgRNA. In some instances, the methods include producing the nucleic acid guides, e.g., by expressing the nucleic acid guides from plasmids encoding the nucleic acid guides, or by PCR and *in vitro* transcription, such as described in greater detail above and shown in FIG. 2. According to certain embodiments, the nucleic acid guides are produced by solid phase synthesis (e.g., as in standard oligonucleotide synthesis).

In some instances, the nucleic acid guided nucleases of the produced library each include a nucleic acid guide component (e.g., a RNA or DNA guide component) and a nuclease component, e.g., a Cas nuclease component (such as Cas9), an Argonaute nuclease component (e.g., Tth Ago, Ago2, or the like). The nuclease component may exhibit cleavage activity or be a cleavage deficient mutant. Where desired, the nuclease component may further include a tag, such as described above.

IDT Exhibit 1017.1899

Also provided by the present disclosure are compositions. In certain aspects, the compositions include a plurality of distinct nucleic acid guided nucleases. The plurality of distinct nucleic acid guided nucleases may include 2 or more, 3 or more, 4 or more, 5 or more, 10 or more, 25 or more, 100 or more, 500 or more, 1000 or more, 10,000 or more, 50,000 or more, 100,000 or more, 500,000 or more, or 1 million or more distinct nucleic acid guided nucleases, etc. In certain aspects, the plurality of distinct nucleic acid guided nucleases includes two or more, but 1 million or less, 500,000 or less, 100,000 or less, 50,000 or less, 10,000 or less, 1000 or less, 500 or less, 100 or less, 25 or less, 10 or less, 5, 4, or 3 distinct nucleic acid guided nucleases.

According to certain embodiments, the distinct nucleic acid guided nucleases are distinct based on: the nucleic acid guided nucleases having differing nuclease components; and/or the nucleic acid guided nucleases having nucleic acid guide components of differing nucleotide sequence. For example, the distinct nucleic acid guided nucleases may target different regions of the same target nucleic acid based on the guide components having differing sequences complementary to different regions of the same target nucleic acid, and/or the distinct nucleic acid guided nucleases may target different target nucleic acids based on the guide components having differing sequences complementary to the different target nucleic acids and/or different species of nucleases (e.g., different Cas9 species) with different PAM sequence requirements so as to broaden the array of target sequences.

The subject compositions may be present in any suitable environment. According to one embodiment, the composition is present in a reaction tube (e.g., a 0.2 mL tube, a 0.6 mL tube, a 1.5 mL tube, or the like) or a well. In certain aspects, the composition is present in two or more (e.g., a plurality of) reaction tubes or wells (e.g., a plate, such as a 96-well plate). The tubes and/or plates may be made of any suitable material, e.g., polypropylene, or the like. In certain aspects, the tubes and/or plates in which the composition is present provide for efficient heat transfer to the composition (e.g., when placed in a heat block, water bath, thermocycler, and/or the like), so that the temperature of the composition may be altered within a short period of time, e.g., as necessary for a particular enzymatic reaction to occur. According to certain embodiments, the composition is present in a thin-walled polypropylene tube, or a plate having thin-walled polypropylene wells. In certain embodiments it may be convenient for the reaction to take place on a solid surface or a bead, in such case, the initial collection of nucleic acids or the nucleic acid guided nuclease(s) may be attached to the solid support or bead by methods known in the art – such as biotin linkage or by covalent linkage) and reaction allowed to proceed on the support.

IDT Exhibit 1017.1900

Other suitable environments for the subject compositions include, e.g., a microfluidic chip (e.g., a "lab-on-a-chip device"). The composition may be present in an instrument configured to bring the composition to a desired temperature, e.g., a temperature-controlled water bath, heat block, or the like. The instrument configured to bring the composition to a desired temperature may be configured to bring the composition to a series of different desired temperatures, each for a suitable period of time (e.g., the instrument may be a thermocycler).

The nucleic acid targeted for depletion can be any target nucleic acid selected by a practitioner of the subject methods. According to one embodiment, the target nucleic acid is an initial RNA (e.g., an rRNA or mtRNA, and not a reverse transcription product of an RNA). In certain aspects, the target nucleic acid is a reverse (DNA) transcription product of an initial RNA (e.g., an rRNA or mtRNA). The RNA (e.g., the initial transcribed RNA) may be any type of RNA (or sub-type thereof) including, but not limited to, a ribosomal RNA (rRNA), a mitochondrial RNA (mtRNA), a microRNA (miRNA), a messenger RNA (mRNA), transfer RNA (tRNA), a small nucleolar RNA (snoRNA), a small nuclear RNA (snRNA), a long non-coding RNA (lncRNA), a non-coding RNA (ncRNA), a small interfering RNA (siRNA), a transacting small interfering RNA (ta-siRNA), a natural small interfering RNA (nat-siRNA), a transfer-messenger RNA (tmRNA), a precursor messenger RNA (pre-mRNA), a small Cajal body-specific RNA (scaRNA), a piwi-interacting RNA (piRNA), an endoribonuclease-prepared siRNA (esiRNA), a small temporal RNA (stRNA), a signal recognition RNA, a telomere RNA, a ribozyme, and any combination of RNA types thereof or subtypes thereof. When the target nucleic acid is a transcription product of an initial RNA, the methods may include depleting all types of such transcription products in the sample (e.g., ribosomal RNA, transfer RNA, microRNA, and the like), or one or more particular types of such transcription products. In certain aspects, the target nucleic acid is a transcription product of a ribosomal RNA (rRNA) template. The rRNA template in such instances may be a eukaryotic 28S, 26S, 25S, 18S, 5.8S, 5S rRNA, or any combination thereof. In other aspects, the rRNA template may be a prokaryotic 23S, 16S, 5S rRNA, or any combination thereof. The subject methods find use in depleting RNA transcription products other than those produced from ribosomal RNAs. For example, the target nucleic acid may be a transcription product of a messenger RNA (mRNA), e.g., a highly expressed but clinically irrelevant mRNA from a pool of total RNA or mRNA (e.g., a globulin mRNA in a sample of total or polyA$^+$ blood RNA). Other types of RNA transcription products may be targeted for depletion, including a mitochondrial RNA (mtRNA), a precursor messenger RNA (pre-mRNA), a micro RNA (miRNA), a transfer RNA (tRNA), and any combination thereof. The target transcription

IDT Exhibit 1017.1901

product may be a product of RNA from a particular organism, such as bacterial RNA or yeast RNA. According to certain embodiments, the target nucleic acid is not genomic DNA.

In certain aspects, the target molecule is a target nucleic acid, and the target nucleic acid is a deoxyribonucleic acid (DNA), e.g., intronic or inter-geneic DNA when it is desired to enrich a sample for exonic DNA (e.g., to enrich a sample for nucleic acids corresponding to the exome of a species of interest) by cleaving intronic or inter-geneic DNA present in the initial collection, or by capturing the exonic sequences directly. In certain aspects, DNA-based plasmids/vectors such as those used for *in vitro* transcription may be targeted for depletion by cleavage, e.g., after completion of an *in vitro* transcription reaction to enrich a nucleic acid sample for newly transcribed RNA.

When practicing the methods of the present disclosure, the nucleic acid guided nuclease(s) may be designed such that the frequency of cleavage and resulting fragment sizes of a particular target nucleic acid is selected by a practitioner of the subject methods. For example, as described above, two or more nucleic acid guided nucleases that target different sequences within a target nucleic acid may be used in a multiplex fashion when it is desirable to cleave the target nucleic acid into 3 or more fragments. The targeted sequences of the target nucleic acid may be chosen to produce fragments of a desired size, e.g., fragments which are small enough to be removed from the sample using a spin column, alcohol precipitation, and/or the like.

UTILITY

The subject methods find use in a variety of different applications, e.g., where it is desirable to deplete irrelevant and/or undesired molecules from a sample of interest; where it is desirable to deplete nucleic acids of interest by removing the nucleic acids of interest from the sample for subsequent recovery (thereby producing a sample enriched for the nucleic acids of interest; and/or the like. By depleting the irrelevant and/or undesired molecules, or removing desirable molecules to produce an enriched sample, the complexity of the sample is reduced and the sample is enriched for molecules of interest. When the molecules of interest are nucleic acids, reduced complexity and enrichment of nucleic acids of interest may facilitate and/or improve the results of downstream applications such as nucleic acid amplification, nucleic acid sequencing, gene expression analysis (e.g., by array hybridization, quantitative RT-PCR, massively parallel sequencing, etc.), the preparation of pharmaceutical compositions in which a therapeutic nucleic acid of interest is to be included, and any other applications in which reduced sample complexity and enrichment of nucleic acids of interest is beneficial.

IDT Exhibit 1017.1902

By way of example, certain embodiments of the subject methods include depleting nucleic acids from a sequencing library. For example, the initial collection may be a collection of nucleic acids to be sequenced on a sequencing platform of interest (e.g., a high-throughput or "next generation" sequencing platform such as an Illumina®- or Ion Torrent®-based sequencing platform), but the collection includes irrelevant/undesirable nucleic acids which may complicate or interfere with obtaining the sequences of nucleic acids of interest (e.g., research or clinical interest) in the collection. The irrelevant/undesirable nucleic acids may be reduced or eliminated using the methods of the present disclosure, e.g., by nucleic acid guided nuclease-mediated cleavage, or by producing a sequencing sample that is enriched for sequences of interest by removal of such sequences from the initial collection using the methods of the present disclosure.

In certain aspects, depletion of a target nucleic acid (e.g., an irrelevant/undesirable nucleic acid, such as a nucleic acid derived from an rRNA, an mtRNA, or the like) renders the target nucleic acid invisible to the sequencing platform. That is, the depleted (e.g., cleaved) target nucleic acid is no longer suitable for sequencing on the sequencing platform of interest. For example, Illumina®- and Ion Torrent®-based sequencing platforms require nucleic acids having adapters at each end of the nucleic acids. According to certain embodiments, the nucleic acids of the initial collection of nucleic acids include sequencing adapters at each end, where selective depletion of the target nucleic acid (e.g., cDNAs transcribed from rRNA) includes cleaving the target nucleic acid into at least two fragments, none of which will include sequencing adapters at each end as required for sequencing on an Illumina®- or Ion Torrent®-based sequencing platform. As such, the target nucleic acid is rendered invisible to the sequencing platform, thereby reducing the "load" on the sequencing platform and the complexity of the sequencing results.

In certain aspects, when the sequencing platform of interest only requires the nucleic acid to include an adapter at a single end of a nucleic acid, the target nucleic acid can be rendered unsuitable for sequencing on the platform by, e.g., cleaving the target nucleic acid such that the length of most or all of the resulting fragments are too short to be sequenced on the sequencing platform. This may be accomplished by cleaving the target nucleic acid at a single location, or at two or more locations within the target nucleic acid, to generate fragments of insufficient length to be sequenced on the sequencing platform, e.g., because the platform requires nucleic acids of greater length, or the fragments are lost during a purification procedure (e.g., bead purification, such as SPRI bead-based purification) prior to sequencing. The location(s) to be cleaved and the distance(s) between cleavage sites may be selected by a

IDT Exhibit 1017.1903

practitioner of the subject methods to generate cleavage fragments of the desired length, e.g., using available nucleic acid sequence information for a target nucleic acid of interest and designing nucleic acid guide component(s) with sequence complementarity to the selected cleavage site(s). In this way, the target nucleic acid is rendered invisible to the sequencing platform and the load on the sequencing platform and complexity of the sequencing results is reduced.

As described above, the methods of the present disclosure also find use in selectively recovering one or more types of nucleic acids of interest from an initial collection of nucleic acids. For example, in certain aspects, depleting one or more types of target nucleic acids from the initial collection of nucleic acids includes capturing the target nucleic acid via formation of target nucleic acid/nucleic acid guided nuclease complexes, and then recovering the target nucleic acids from the complexes, e.g., for downstream analysis (e.g., quantitative analysis, sequence analysis, and/or the like). Approaches for target nucleic acid capture and recovery include, e.g., affinity-based approaches (which in certain aspects is facilitated by the nuclease component including an epitope/affinity tag), as described hereinabove. Accordingly, the subject methods find use in selectively obtaining one or more target nucleic acids of interest from collections of nucleic acids, which in certain aspects are complex collections of nucleic acids (e.g., a collection of cDNAs produced by reverse transcription of a total RNA sample, exons from a genomic library, or any other complex nucleic acid collections of interest).

KITS

Also provided by the present disclosure are kits useful for practicing the subject methods. The kits may include one or more of any of the components described above in relation to the subject methods and compositions. For example, the kits may include a primer and a template for generating PCR amplification products from which a nucleic acid guide component may be produced by *in vitro* transcription. Such reagents may include any reagents useful in practicing the method for producing a nucleic acid guide component (e.g., an sgRNA) shown in FIG. 2 and described hereinabove. According to certain aspects, the kit includes a reverse primer for use in conjunction with a forward primer provided by a user of the kit, where at least a portion of the forward primer includes a nucleic acid sequence complementary to a target nucleic acid selected for depletion by the user.

In some instances, the kits include: a vector that includes a sgRNA scaffold template domain; a reverse primer configured for use with the vector in a PCR reaction; and an RNA polymerase. In some instances, the reverse primer comprises a polyA domain and an sgRNA

IDT Exhibit 1017.1904

scaffold domain. While the RNA polymerase may vary, in some embodiments the RNA polymerase is a T7 polymerase. Where desired, the kit further includes one or more of a DNA polymerase; a PCR buffer; a nuclease (e.g., a Cas, Ago, or other nuclease) which may or may not have cleavage activity; a control nucleic acid; etc.

Components of the subject kits may be present in separate containers, or multiple components may be present in a single container. For example, when the kit includes a primer for generating a template DNA from which an RNA guide component is produced by *in vitro* transcription, the primer may be provided in a separate container, or in a container that includes a second component of the kit (e.g., a buffer or the like).

In addition to the above-mentioned components, the subject kit may further include instructions for using the components of the kit to practice the subject methods. The instructions for practicing the subject methods may be recorded on a suitable recording medium. For example, the instructions may be printed on a substrate, such as paper or plastic, etc. As such, the instructions may be present in the kits as a package insert, in the labeling of the container of the kit or components thereof (i.e., associated with the packaging or subpackaging) etc. In other embodiments, the instructions are present as an electronic storage data file present on a suitable computer readable storage medium, e.g., portable flash drive, DVD, CD-ROM, diskette, etc. In yet other embodiments, the actual instructions are not present in the kit, but means for obtaining the instructions from a remote source, e.g. via the internet, are provided. An example of this embodiment is a kit that includes a web address where the instructions can be viewed and/or from which the instructions can be downloaded. As with the instructions, the means for obtaining the instructions is recorded on a suitable substrate.

The following examples are offered by way of illustration and not by way of limitation.

EXPERIMENAL

Example 1: Depletion of a Target PCR Fragment

In this example, a nucleic acid guided nuclease was produced and, as proof of concept, used to deplete a PCR product having a target sequence corresponding to an rRNA sequence. Using the method outlined in FIG. 2 and described above, two distinct sgDNA templates (CMBO640 and CMBO641) were generated by PCR using the Advantage® HD polymerase (Clontech Laboratories, Mountain View, CA). The PCR products are seen as the main (lower) bands in the gel image shown in FIG. 4. Between 100-120 ng of the templates were used in an

IDT Exhibit 1017.1905

in vitro transcription reaction in accordance with the Takara T7 RNA polymerase manual to produce the corresponding sgRNAs.

The *in vitro* transcribed CMBO640 (72 ng/µl) and CMBO641 (90 ng/µl) sgRNAs were then tested for their ability, when combined with Cas9 (500 ng), to cleave a PCR product (260 ng) having a sequence corresponding to an rRNA sequence. The results are provided in FIG. 5. As shown, combining Cas9, the COMBO640 and/or COMBO641 sgRNAs, and the PCR product results in cleavage of the PCR product.

Example 2: Multiplex depletion of DNAs corresponding to full-length 18S rRNA

In this example, it is shown that two different sgRNAs having different target sequence specificities can be used in conjunction (i.e., in a multiplex fashion) to deplete a target DNA having a sequence corresponding to full-length 18S rRNA. As shown in the gel image provided in FIG. 6, a nucleic acid guided nuclease that includes Cas9 and a first sgRNA ("sgRNA1", fourth lane from the left) and a nucleic acid guided nuclease that includes Cas9 and a second sgRNA ("sgRNA2", fifth lane from the left) are separately able to cleave at different locations a target DNA having a sequence corresponding to full-length 18S rRNA. When the two nucleic acid guided nucleases are both combined with the target DNA (sixth lane from the left), the target DNA is cleaved at both of the respective cleavage sites, indicating that multiplexed target depletion occurred.

Example 3: Depletion of Undesirable Sequences from Next Generation Sequencing Libraries

In this example, the ability of nucleic acid guided nucleases to deplete unwanted sequences from next generation sequencing libraries was tested. A next generation sequencing library of cDNAs transcribed from human brain total RNA was treated with a sgRNA/Cas9 nucleic acid guided nuclease to deplete cDNAs corresponding to 18S rRNA by cleaving at position 1022-1041 of the cDNAs corresponding to 18S rRNA prior to sequencing the library. The resulting sequences were mapped to the rRNA and the number of reads overlapping 1022-1041 was divided by the total number of reads mapping to rRNA. The ratio was plotted and normalized to control.

As shown in Panel A of FIG. 7, depletion of the target sequence only occurred in the library treated with both Cas9 and the specific guide RNA.

In a separate experiment, a target cDNA corresponding to 18S rRNA was degraded using a pool of sgRNAs targeting the target cDNA along its length about every 50 bp. The pool was generated by PCR amplifying 35 sgDNA templates (using the approach shown in FIG. 2),

IDT Exhibit 1017.1906

and then *in vitro* transcribing the corresponding 35 sgRNAs in a pooled *in vitro* transcription reaction. This is one of the key advantages of the method of the present disclosure shown in FIG. 2, as large quantities of a complex collection of different sgRNA sequences can be produced in a single reaction for depleting a single target (by tiling across the target with sgRNAs having different target-specific sequences) or depleting multiple targets using the collection of sgRNAs.

In this experiment, a model target was used that included a target corresponding to the full-length 18S RNA (25 ng) and 55 ng of a 5800 bp plasmid iPCR product (used as a surrogate for the 28S rRNA). This ratio approximates the ratio (in mass) of 18S and 28S rRNA in a 20 nM RNA-Seq library with no depletion. As shown in Panel B of FIG. 7, the 18S fragment was thoroughly depleted/degraded by treatment with the sgRNA pool, while the iPCR fragment was unaffected. Depletion of the target nucleic acid results in failure of the target to cluster on the sequencer (e.g., an Illumina® sequencer) because, at most, the target only has a sequencing adapter at one of its ends following cleavage. Alternatively, the depleted target fragments are sufficiently small that they are lost following purification (e.g., SPRI bead purification) and thus not available for sequencing.

Example 4: Depletion of 18s rRNA using a pool of nucleic acid guided nucleases

In this experiment, sequencing libraries were generated from 100ng Human Brain Total RNA (Clontech) using the SMARTer Stranded RNA-Seq Kit (Clontech). 70ng of library was incubated with 0, 2, or 5µg of recombinantly purified Cas9 and 0 or 191ng of the 35 sgRNA pool described in the previous experiment in 20µl 1X NEB3.1 buffer for one hour at 37°C. The Cas9 was then heat inactivated for 10 minutes at 70°C. The libraries were then pooled and sequenced on a MiSeq instrument. The resulting sequences were mapped against the human genome, hg19, and rRNA transcripts simultaneously using the STAR aligner. FIG. 8 illustrates the reduction in sequence coverage of the 18S rRNA transcript resulting only from treatment with both Cas9 and the sgRNA pool. The number of sequencing reads mapped to the 18S rRNA were normalized to the total number of reads sequenced and plotted in FIG. 9, demonstrating an ~75% reduction in sequences mapping to the 18S transcript.

Examples 1-4: Sequence Information

The sgRNA described in examples 1-4 contained the sgRNA scaffold shown in Figure 1 with the following target-specific sequences:

sgRNA1 used in examples 1 and 2: UUAUCAGAUCAAAACCAACC (SEQ ID NO:01)

IDT Exhibit 1017.1907

sgRNA2 used in examples 1, 2 and 3: UAAUCAAGAACGAAAGUCGG (SEQ ID NO:02)

pool of 35 sgRNA used in examples 3 and 4: (SEQ ID NO:03 to 37)

GACAAGCAUAUGCUACUGGC	CGGCGCAAUACGAAUGCCCC
CGGUACAGUGAAACUGCGAA	CGCUCUGGUCCGUCUUGCGC
GGAGAGGAGCGAGCGACCAA	UAAUCAAGAACGAAAGUCGG
UAGAGCUAAUACAUGCCGAC	CGGUCGGCAUCGUUUAUGGU
UUAUCAGAUCAAAACCAACC	GUUUCCCGGAAGCUGCCCGG
GGGGCGGGCGCCGGCGGCUU	CUGAAACUUAAAGGAAUUGA
CGAUCGCACGCCCCCCGUGG	GGCUUAAUUUGACUCAACAC
GGUAGUCGCCGUGCCUACCA	CUGUCAAUCCUGUCCGUGUC
UCAGGGUUCGAUUCCGGAGA	GCAUGGCCGUUCUUAGUUGG
UGCGCGCCUGCUGCCUUCCU	GCCAGAGUCUCGUUCGUUAU
AACAAUACAGGACUCUUUCG	GCGUCCCCCAACUUCUUAGA
CCUCGUUAAAGGAUUUAAAG	UGUUAUUGCUCAAUCUCGGG
UAUUGGAGCUGGAAUUACCG	AGCGUGUGCCUACCCUACGC
AAAGCUCGUAGUUGGAUCUU	CCGUUGAACCCCAUUCGUGA
CAAGGGGCGGGGACGGGCGG	UACUGGGAAUUCCUCGUUCA
UCUUAGCUGAGUGUCCCGCG	GGCGGUGUGUACAAAGGGCA
CAAAGCAGGCCCGAGCCGCC	GGCCCUCGGAUCGGCCCCGC
GGACCGCGGUUCUAUUUUGU	

Example 5: Depletion of Target Nucleic Acids by Argonaute (Ago)

In this example, the ability of the Argonaute (Ago) protein to deplete a target nucleic acid was assessed. His-tagged Tth Ago was expressed and purified (FIG. 10, Panel A). Depletion of a 5' FAM labelled ssDNA representative target was carried out by combining the 500nM single-stranded DNA, Tth Ago, and 100nM 5' phosphorylated targeting oligonucleotide complementary to the target single-stranded DNA and incubated at 75°C for one hour. As shown in FIG. 10, Panel B, the single stranded target was not cleaved in the absence of Ago (far right lane), but was cleaved in the presence of Ago and the targeting oligonucleotide at various Ago concentrations. FIG. 11 demonstrates a similar experiment, but with 20nM ssDNA target and 50nM guide DNA to be representative of depleting common library concentrations.

In Example 5, the guide DNA oligo was:

/5'Phos/TGAGGTAGTAGGTTGTATAGT (SEQ ID NO:38); and the targeting oligo was:

/5'6-FAM/AGGTGATAAGACTATACAACCTACTACCTCGAATGTCCGT (SEQ ID NO:39)

IDT Exhibit 1017.1908

Example 6: Depletion of Target Nucleic Acids from a Collection by Argonaute (Ago)

As demonstrated in Example 5, Argonaute is capable of depleting target ssDNA molecules in solution. If the desired target is double stranded, the double stranded material may be converted to a single strands by denaturation, such as with heat or high pH, or by degrading the non-target strand by exonuclease treatment. For example, the mixture may be amplified by a common pair of PCR primers, only one of which is 5' phosphorylated. The amplified product could then be treated with λ-exonuclease to yield a ssDNA product. This ssDNA product could then be used as a substrate for cleavage by Ago and a targeting oligonucleotide.

Example 7: Recovery of Targeted Sequences

In this Example the 6xHN tagged D10A/H840A mutant of Cas9 (referred to herein as dCas9), was expressed and purified in E. coli (FIG. 12, Panel A). RNA-Seq libraries were generated from 100ng Human Brain PolyA-plus RNA (Clontech) and the SMARTer Stranded RNA-Seq kit (Clontech). 10ng of library was combined with ~1.25µg dCas9 and 76ng of a pool of 190 sgRNA designed every ~50bp on the human 5S, 5.8S, 18S, 28S, mt12S, and mt16S sequences in 10µl 1X NEB3.1. The mixture was incubated at 37°C for one hour. Then 10µg of salmon sperm DNA (Life Technologies) was added and the reaction was allowed to proceed for an additional 30 minutes at 37°C. 5µl of TALON magnetic beads (Clontech) equilibrated in 10µl NEB3 was added to the tubes and incubated at 25°C with rotation for 30 minutes. The beads were washed twice with 200µl NEB. 50µl of SeqAmp PCR Mastermix (1X SeqAmp Buffer, 1µl SeqAmp DNA polymerase, and 250nM Illumina P5 and P7 PCR primers) was added to the washed beads and amplified with 15 cycles of PCR (94°C 1 minute, 15X (98°C -15s, 55°C-15s, 68°C-30s)). The amplified product was purified with 50µl AMPure beads (Beckman) according to the manufacturer's instructions and eluted in 20µl 10mM Tris-HCl pH8.5 0.1% Tween-20. The library was then sequenced on a MiSeq instrument. The resulting sequences were mapped against the human genome, hg19, and rRNA transcripts simultaneously using the STAR aligner. Reads mapping to rRNA were identified by Picard RNA-Seq Metrics, and the relative amounts of sequences mapping to rRNA are plotted in FIG. 12, Panel B, showing a ~340% enrichment verses an untreated library. This example demonstrates that nucleic acid guided nucleases may be used to deplete target nucleic acids by removal of the target nucleic acids from an initial collection of nucleic acids, for subsequent recovery and production of a sample enriched for the target nucleic acids.

IDT Exhibit 1017.1909

Notwithstanding the appended clauses, the disclosure is also defined by the following clauses:

1. A method of selectively depleting a target nucleic acid from an initial collection of nucleic acids, the method comprising:

contacting the collection of nucleic acids with a nucleic acid guided nuclease specific for the target nucleic acid in a manner sufficient to deplete the target nucleic acid from the initial collection.

2. The method according to Claim 1, wherein the target nucleic acid is a deoxyribonucleic acid (DNA) or a ribonucleic acid (RNA).

3. The method according to Claims 1 or 2, wherein the target nucleic acid is a double-stranded nucleic acid.

4. The method according to Claims 1 or 2, wherein the target nucleic acid is a single-stranded nucleic acid.

5. The method according to any of Claims 1 to 4, wherein the nucleic acid guided nuclease comprises a RNA guide component and a nuclease component.

6. The method according to Claim 5, wherein the nuclease component comprises a Cas nuclease component.

7. The method according to Claim 6, wherein the Cas nuclease component exhibits cleavage activity and the method results in cleavage of the target nucleic acid to deplete the initial collection of the target nucleic acid.

8. The method according to Claim 6, wherein the Cas nuclease component is a cleavage deficient mutant and the method results in the production of a product composition comprising target nucleic acid/nucleic acid guided nuclease complexes.

9. The method according to Claim 8, wherein the method further comprises separating the complexes from other constituents of the product composition.

10. The method according to Claim 9, wherein the Cas nuclease component further comprises a tag.

11. The method according to Claim 10, wherein the method further comprises recovering target nucleic acids from the complexes.

12. The method according to any of Claims 1 to 4, wherein the nucleic acid guided nuclease comprises a DNA guide component and a nuclease component.

13. The method according to Claim 12, wherein the nuclease component comprises a Tth Ago nuclease component.

IDT Exhibit 1017.1910

14. The method according to any of the preceding claims, wherein the method comprises contacting the initial collection of nucleic acids with a plurality of two or more distinct nucleic acid guided nucleases.

15. A method of selectively depleting a subpopulation of nucleic acids from an initial collection of nucleic acids, the method comprising:

contacting the initial collection of nucleic acids with a library of nucleic acid guided nucleases in a manner sufficient to deplete the subpopulation from the initial collection, wherein the library includes two or more distinct nucleic acid guided nucleases specific for one or more members of the subpopulation of nucleic acids.

16. The method according to Claim 15, wherein the subpopulation comprises deoxyribonucleic acids.

17. The method according to Claims 15 or 16, wherein the subpopulation comprises double-stranded nucleic acids.

18. The method according to Claims 15 or 16, wherein the subpopulation comprises single-stranded nucleic acids.

19. The method according to Claims 15 or 16, wherein the library comprises five or more distinct nucleic acid guided nucleases.

20. The method according to Claim 19, wherein the five or more distinct nucleic acid guided nucleases have nucleic acid guide components of differing sequence.

21. The method according to any of Claims 16 to 20, wherein the nucleic acid guided nucleases each comprise an RNA guide component and a nuclease component.

22. The method according to Claim 21, wherein the nuclease component comprises a Cas nuclease component.

23. The method according to Claim 22, wherein the Cas nuclease component exhibits cleavage activity and the method results in cleavage of the subpopulation to deplete the initial collection of the subpopulation.

24. The method according to Claim 22, wherein the Cas nuclease component is a cleavage deficient mutant and the method results in the production of a product composition comprising subpopulation nucleic acid/nucleic acid guided nuclease complexes.

25. The method according to Claim 24, wherein the method further comprises separating the complexes from other constituents of the product composition.

26. The method according to Claim 25, wherein the Cas nuclease component further comprises a tag.

IDT Exhibit 1017.1911

27. The method according to Claim 26, wherein the method further comprises recovering subpopulation nucleic acids from the complexes.

28. The method according to any of Claims 16 to 20, wherein the nucleic acid guided nucleases each comprise a DNA guide component and a Tth Ago nuclease component.

29. The method according to any of Claims 15 to 27, wherein the initial collection of nucleic acids comprises a next generation sequencing (NGS) nucleic acid collection.

30. The method according to Claim 29, wherein the subpopulation comprises sequences that are not desired to be sequenced.

31. The method according to Claim 30, wherein the method further comprises sequencing the subpopulation depleted NGS nucleic acid collection.

32. A composition comprising plurality of distinct nucleic acid guided nucleases.

33. The composition according to Claim 32, wherein the composition comprises 5 or more distinct nucleic acid guided nucleases.

34. The composition according to Claim 33, wherein the composition comprises 25 or more distinct nucleic acid guided nucleases.

35. The composition according to Claim 34, wherein the composition comprises 100 or more distinct nucleic acid guided nucleases.

36. The composition according to Claim 35, wherein the composition comprises 500 or more distinct nucleic acid guided nucleases.

37. The composition according to Claim 36, wherein the composition comprises 1000 or more distinct nucleic acid guided nucleases.

38. The composition according to any of Claims 32 to 37, wherein the nucleic acid guided nucleases have nucleic acid guide components of differing sequence.

39. The composition according to any of Claims 32 to 38, wherein the nucleic acid guided nucleases each comprise an RNA guide component and a nuclease component.

40. The composition according to Claim 39, wherein the nuclease component comprises a Cas nuclease component.

41. The composition according to Claim 40, wherein the Cas nuclease component exhibits cleavage activity.

42. The composition according to Claim 40, wherein the Cas nuclease component is a cleavage deficient mutant.

43. The composition according to Claim 42, wherein the Cas nuclease component further comprises a tag.

IDT Exhibit 1017.1912

44. The composition according to any of Claims 32 to 38, wherein the nucleic acid guided nucleases each comprise a DNA guide component and a nuclease component.

45. The composition according to Claim 44, wherein the nuclease component comprises a Tth Ago nuclease component.

46. A method of making a library of nucleic acid guided nucleases, the method comprising:

combining a plurality of distinct guide nucleic acids with a nuclease in a manner sufficient to produce the library of nucleic acid guided nucleases.

47. The method according to Claim 46, wherein the nuclease is a Cas nuclease.

48. The method according to Claim 47, wherein the Cas nuclease comprises cleavage activity.

49. The method according to Claim 47, wherein the Cas nuclease is a cleavage deficient mutant.

50. The method according to any of Claims 46 to 49, wherein the library comprises 5 or more distinct nucleic acid guided nucleases.

51. The method according to Claim 50, wherein the library comprises 25 or more distinct nucleic acid guided nucleases.

52. The method according to Claim 51, wherein the library comprises 100 or more distinct nucleic acid guided nucleases.

53. The method according to Claim 52, wherein the library comprises 500 or more distinct nucleic acid guided nucleases.

54. The method according to Claim 53, wherein the library comprises 1000 or more distinct nucleic acid guided nucleases.

55. The method according to any of Claims 46 to 54, wherein the nucleic acid guides comprise crRNA and tracrRNA.

56. The method according to any of Claims 46 to 55, wherein the nucleic acid guides comprises sgRNA.

57. The method according to any of the preceding claims, wherein the method further comprises producing the nucleic acid guides.

58. The method according to Claim 57, wherein the nucleic acid guides are produced by expressing the nucleic acid guides from vectors encoding the nucleic acid guides.

59. The method according to Claim 57, wherein the nucleic acid guides are produced using an in vitro transcription protocol.

60. A kit comprising:

a vector comprising a sgRNA scaffold template domain;

IDT Exhibit 1017.1913

a reverse primer configured for use with the vector in a PCR reaction; and

a RNA polymerase.

61. The kit according to Claim 60, wherein the reverse primer comprises a polyA domain and a sgRNA scaffold domain.

62. The kit according to any of Claims 60 and 61, wherein the RNA polymerase is a T7 polymerase.

63. The kit according to any of Claims 60 to 62, wherein the kit further comprises a DNA polymerase.

64. The kit according to any of Claims 60 to 63, wherein the kit further comprises a PCR buffer.

65. The kit according to any of Claims 60 to 64, wherein the kit further comprises a nuclease.

66. The kit according to Claim 65, wherein the nuclease is a Cas nuclease.

67. The kit according to Claim 66, wherein the Cas nuclease comprises cleavage activity.

68. The kit according to Claim 66, wherein the Cas nuclease is a cleavage deficient mutant.

Although the foregoing invention has been described in some detail by way of illustration and example for purposes of clarity of understanding, it is readily apparent to those of ordinary skill in the art in light of the teachings of this invention that certain changes and modifications may be made thereto without departing from the spirit or scope of the appended claims.

Accordingly, the preceding merely illustrates the principles of the invention. It will be appreciated that those skilled in the art will be able to devise various arrangements which, although not explicitly described or shown herein, embody the principles of the invention and are included within its spirit and scope. Furthermore, all examples and conditional language recited herein are principally intended to aid the reader in understanding the principles of the invention and the concepts contributed by the inventors to furthering the art, and are to be construed as being without limitation to such specifically recited examples and conditions. Moreover, all statements herein reciting principles, aspects, and embodiments of the invention as well as specific examples thereof, are intended to encompass both structural and functional equivalents thereof. Additionally, it is intended that such equivalents include both currently known equivalents and equivalents developed in the future, i.e., any elements developed that perform the same function, regardless of structure. The scope of the present invention, therefore, is not intended to be limited to the exemplary embodiments shown and described herein. Rather, the scope and spirit of present invention is embodied by the appended claims.

IDT Exhibit 1017.1914

WHAT IS CLAIMED IS:

1. A method of selectively depleting a target nucleic acid from an initial collection of nucleic acids, the method comprising:

contacting the collection of nucleic acids with a nucleic acid guided nuclease specific for the target nucleic acid in a manner sufficient to deplete the target nucleic acid from the initial collection.

2. The method according to Claim 1, wherein the target nucleic acid is a deoxyribonucleic acid (DNA) or a ribonucleic acid (RNA).

3. The method according to Claims 1 or 2, wherein the target nucleic acid is a double-stranded nucleic acid.

4. The method according to Claims 1 or 2, wherein the target nucleic acid is a single-stranded nucleic acid.

5. The method according to any of Claims 1 to 4, wherein the nucleic acid guided nuclease comprises a RNA guide component and a nuclease component.

6. The method according to any of Claims 1 to 4, wherein the nucleic acid guided nuclease comprises a DNA guide component and a nuclease component.

7. The method according to any of the preceding claims, wherein the method further comprises separating the complexes from other constituents of the product composition.

8. The method according to Claim 7, wherein the nuclease component further comprises a tag.

IDT Exhibit 1017.1915

9. The method according to Claim 8, wherein the method further comprises recovering target nucleic acids from the complexes.

10. The method according to any of the preceding claims, wherein the method comprises contacting the initial collection of nucleic acids with a plurality of two or more distinct nucleic acid guided nucleases.

11. A method of selectively depleting a subpopulation of nucleic acids from an initial collection of nucleic acids, the method comprising:

contacting the initial collection of nucleic acids with a library of nucleic acid guided nucleases in a manner sufficient to deplete the subpopulation from the initial collection, wherein the library includes two or more distinct nucleic acid guided nucleases specific for one or more members of the subpopulation of nucleic acids.

12. The method according to Claim 11, wherein the initial collection of nucleic acids comprises a next generation sequencing (NGS) nucleic acid collection.

13. A composition comprising plurality of distinct nucleic acid guided nucleases.

14. A method of making a library of nucleic acid guided nucleases, the method comprising:

combining a plurality of distinct guide nucleic acids with a nuclease in a manner sufficient to produce the library of nucleic acid guided nucleases.

15. A kit comprising:

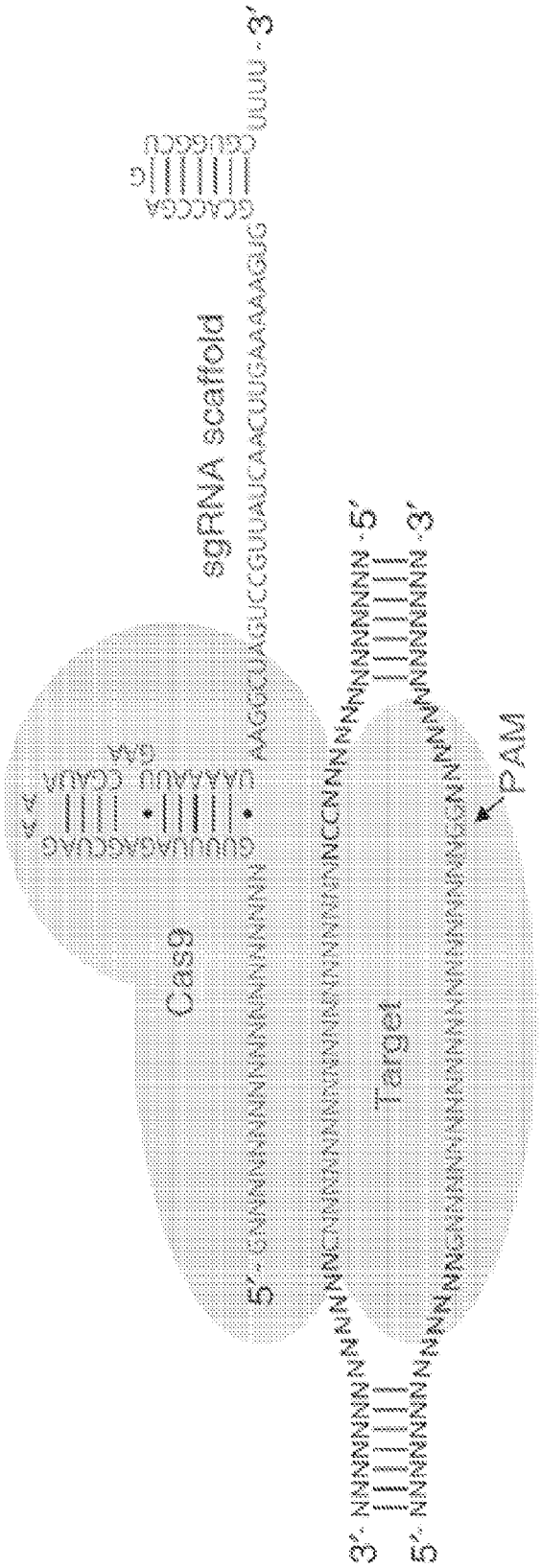
a vector comprising a sgRNA scaffold template domain;

a reverse primer configured for use with the vector in a PCR reaction; and

a RNA polymerase.

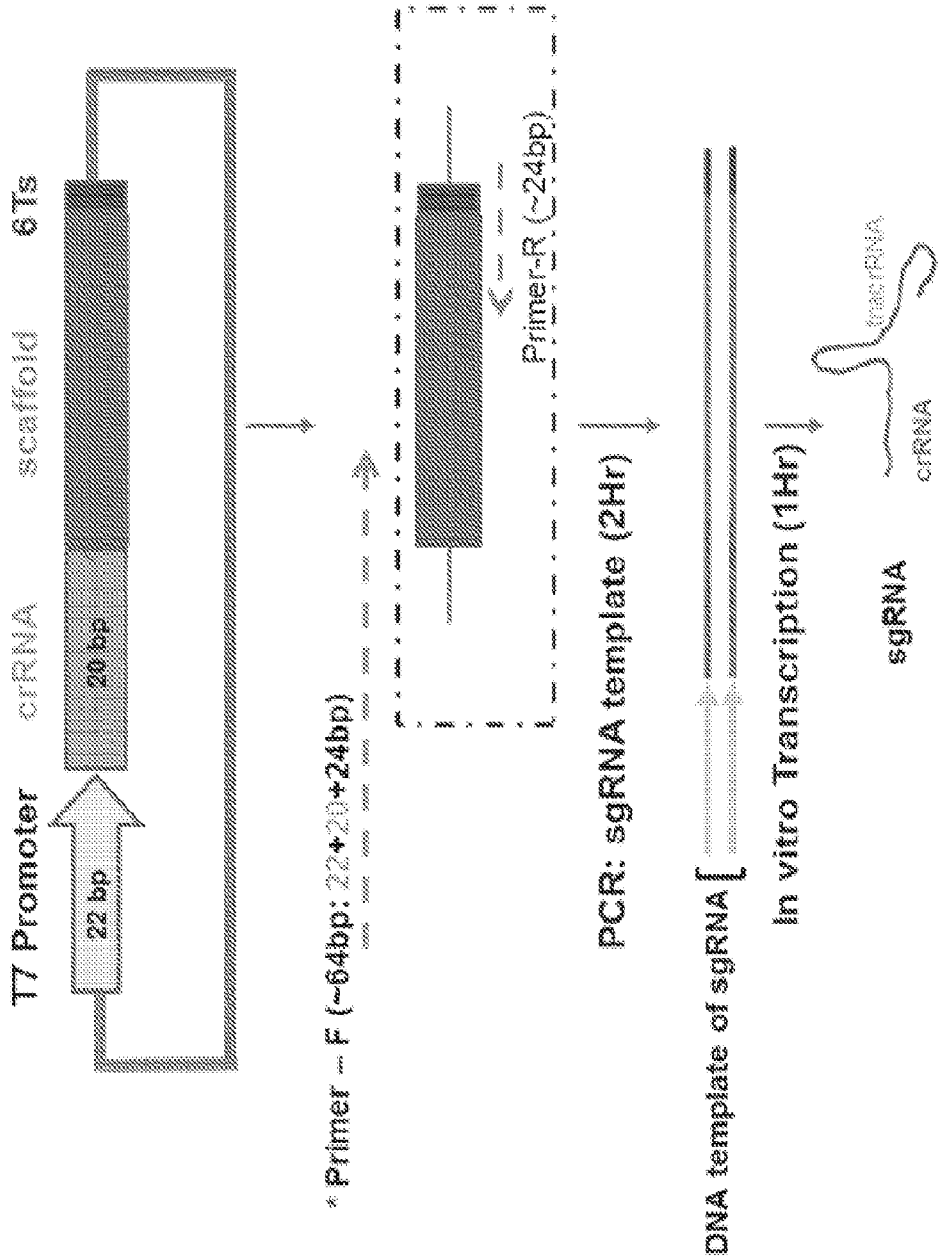
IDT Exhibit 1017.1916

FIG. 1



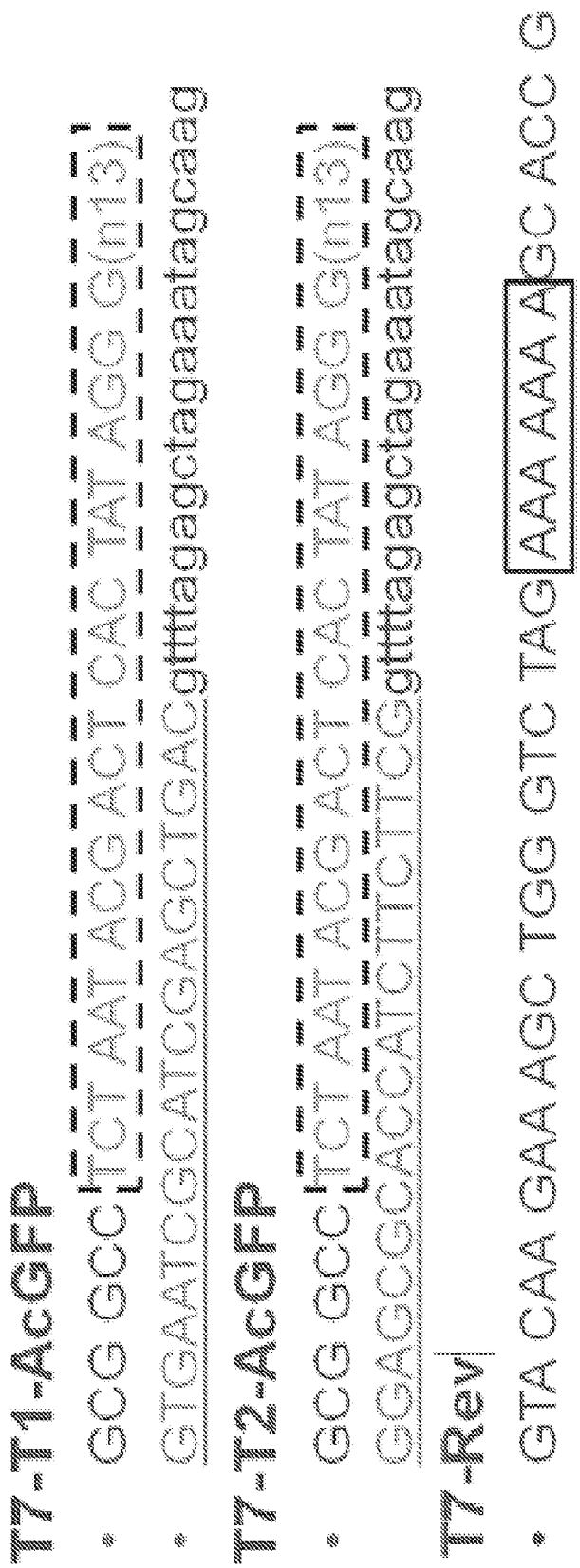
IDT Exhibit 1017.1917

FIG. 2



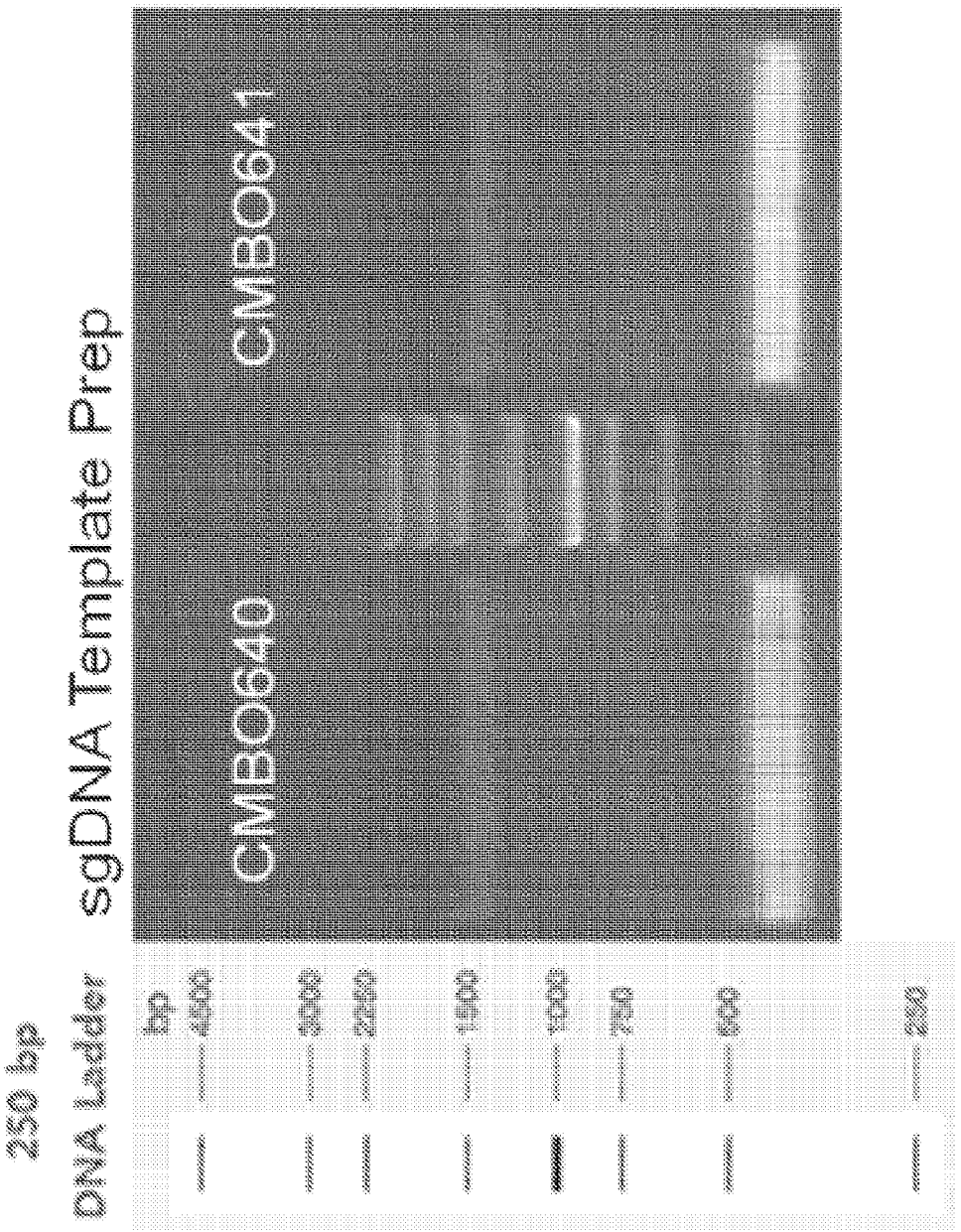
IDT Exhibit 1017.1918

FIG. 3



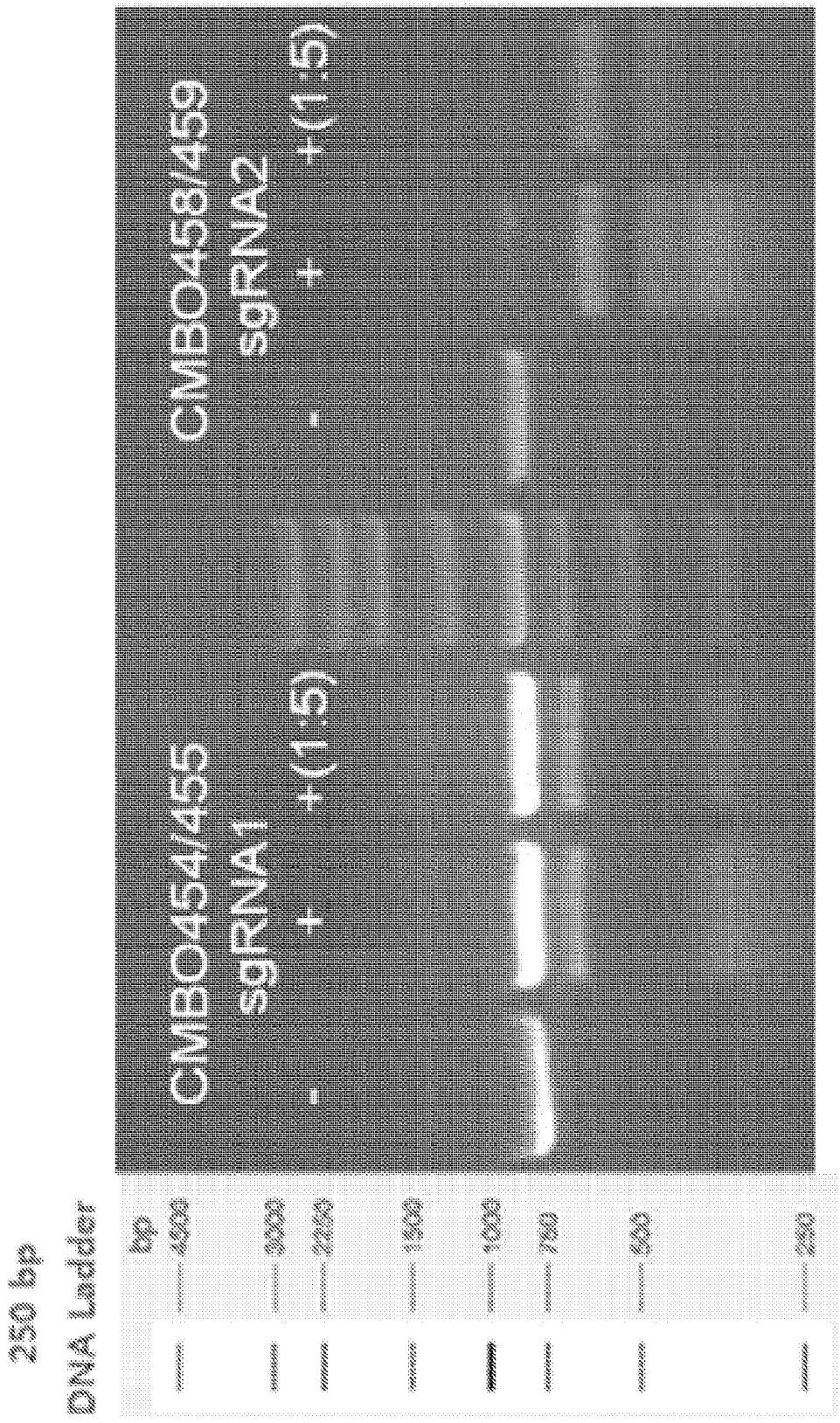
IDT Exhibit 1017.1919

FIG. 4



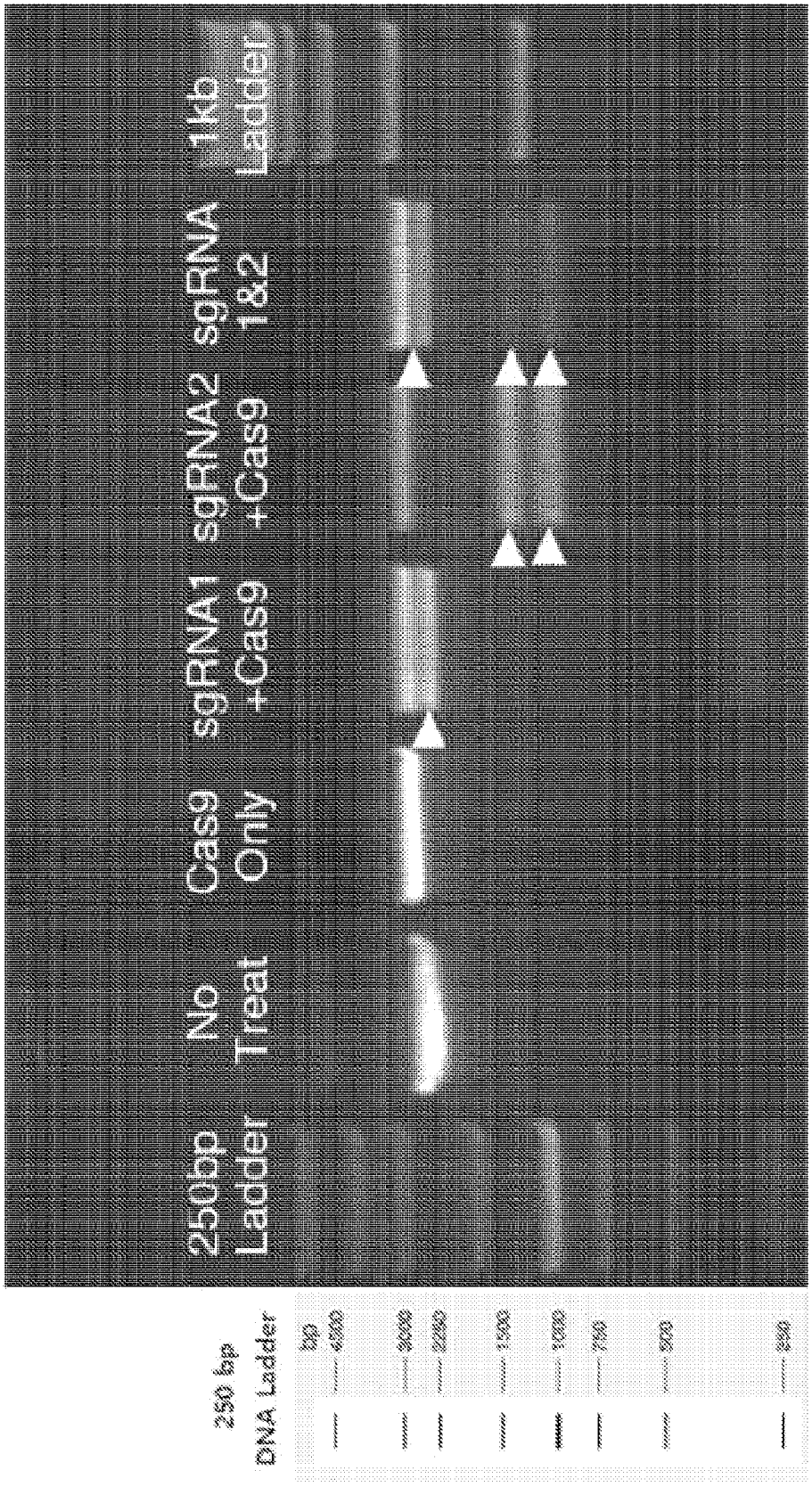
IDT Exhibit 1017.1920

FIG. 5



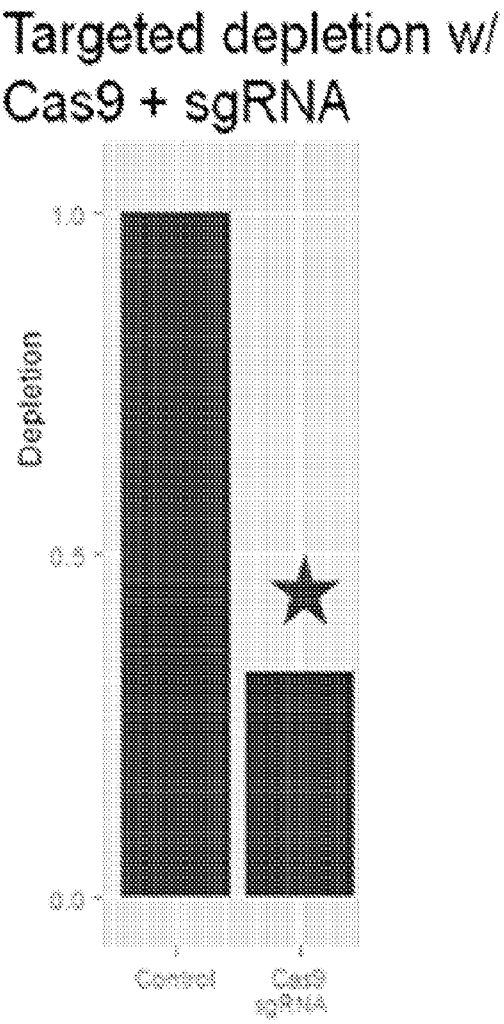
IDT Exhibit 1017.1921

FIG. 6



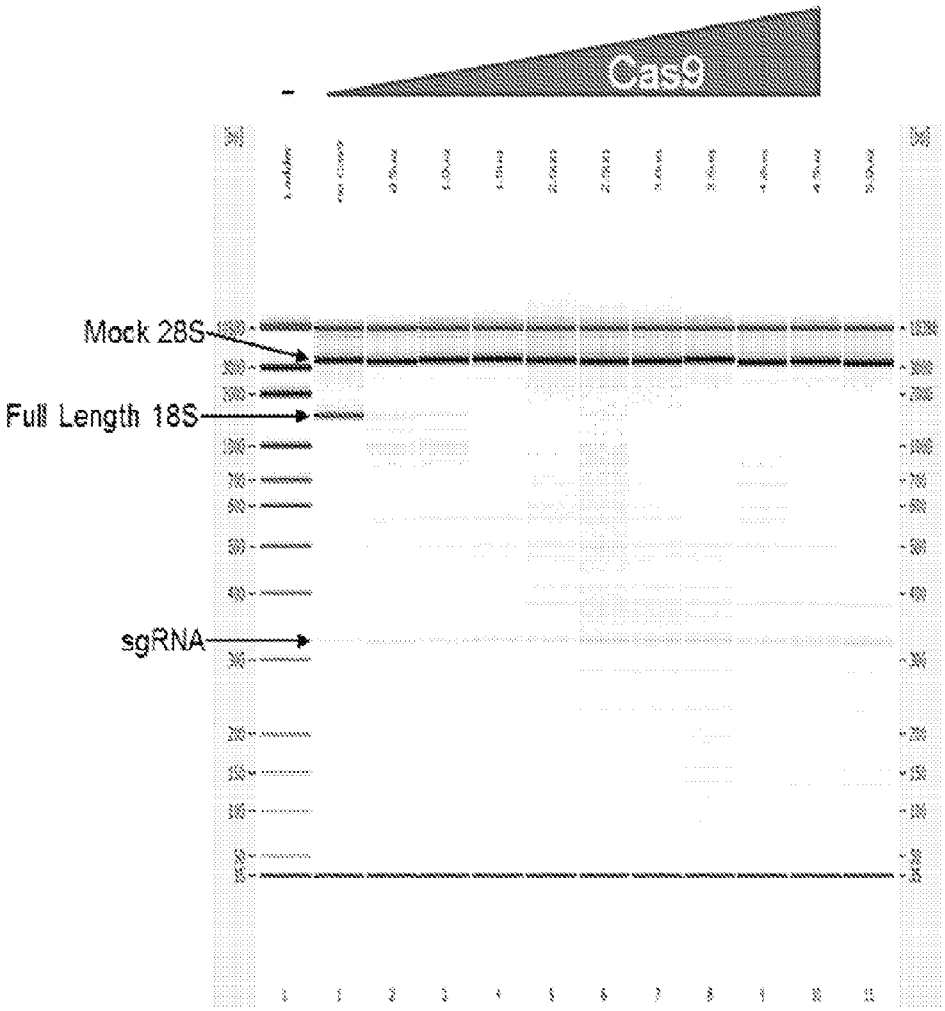
IDT Exhibit 1017.1922

FIG. 7A



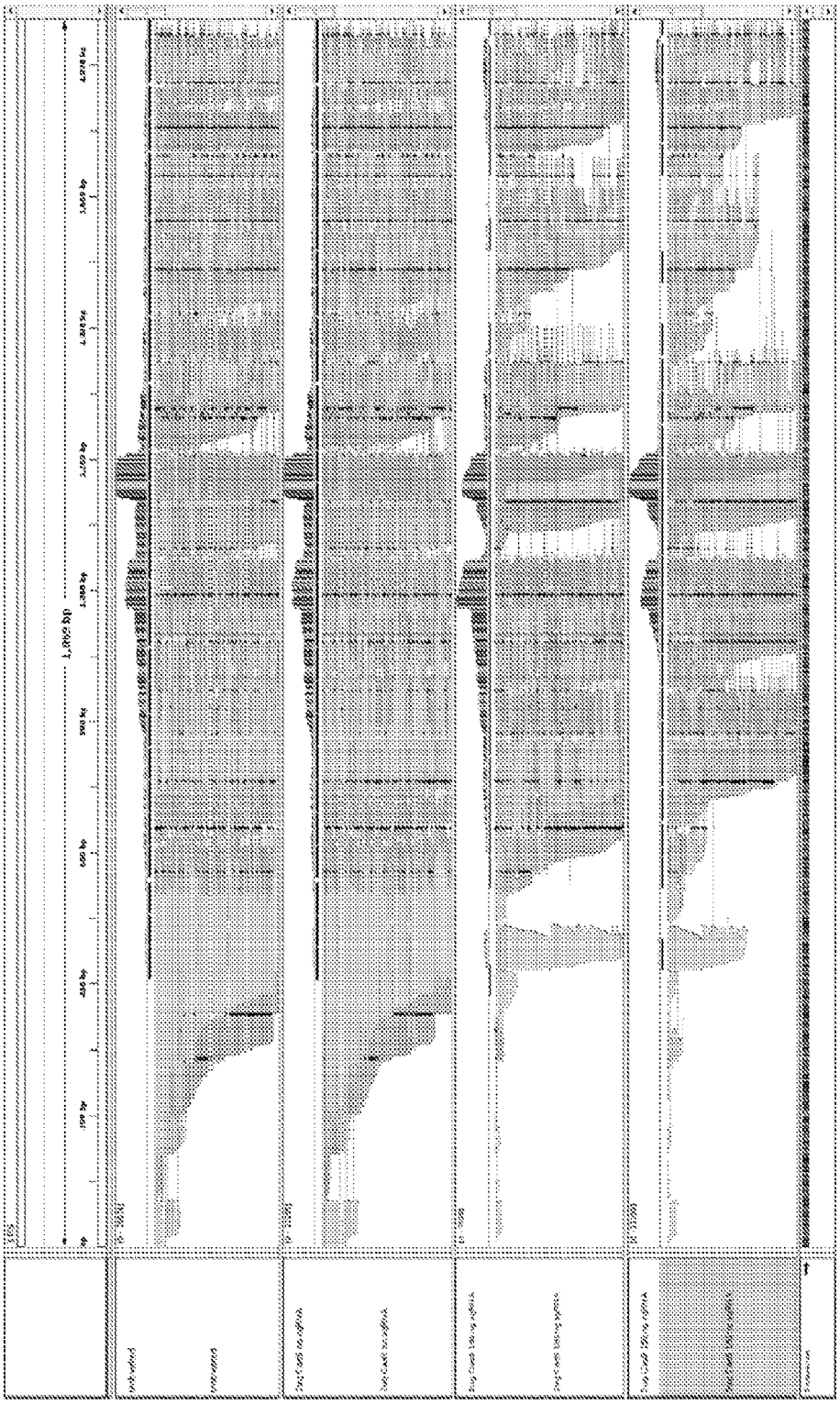
IDT Exhibit 1017.1923

FIG. 7B



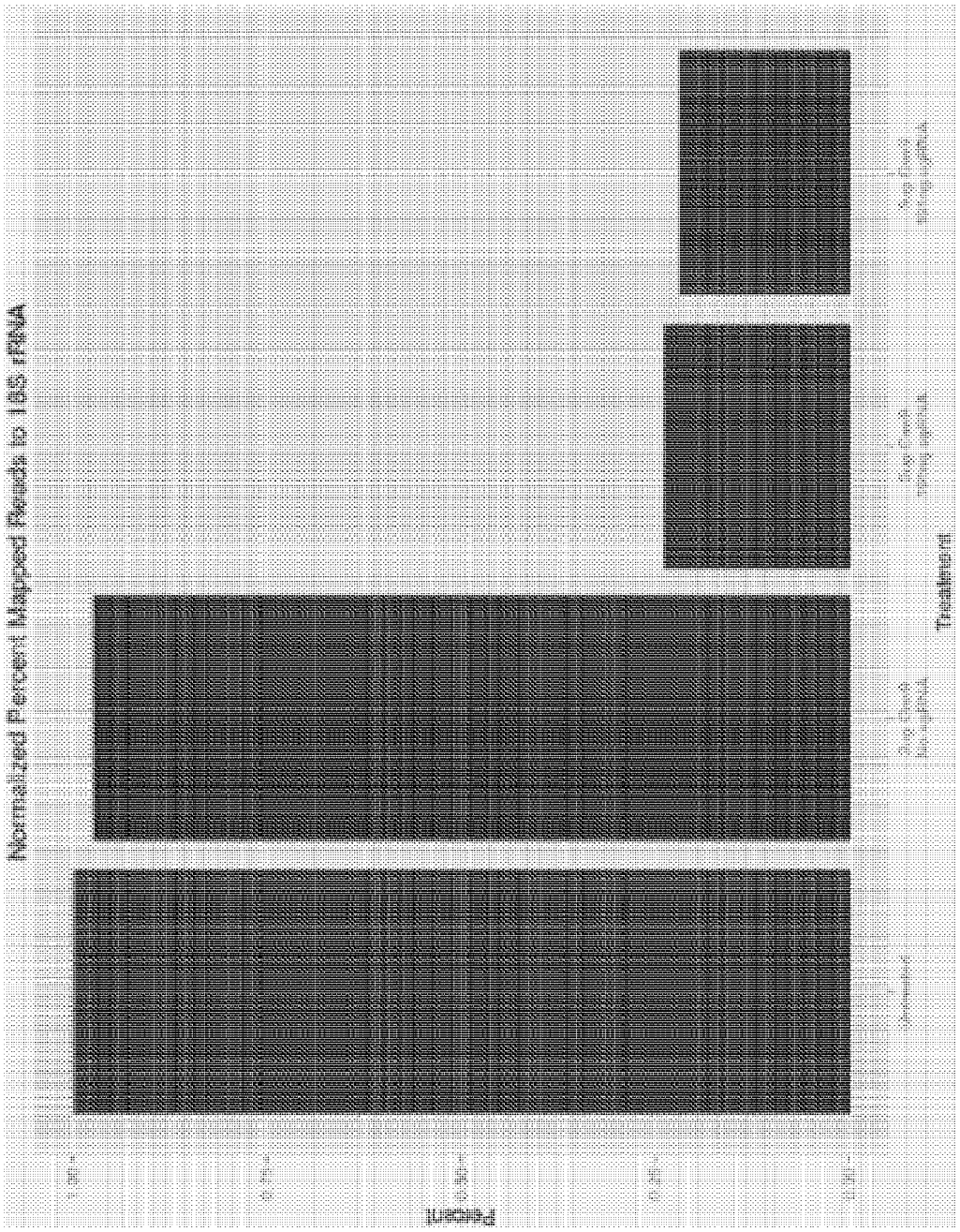
IDT Exhibit 1017.1924

FIG. 8



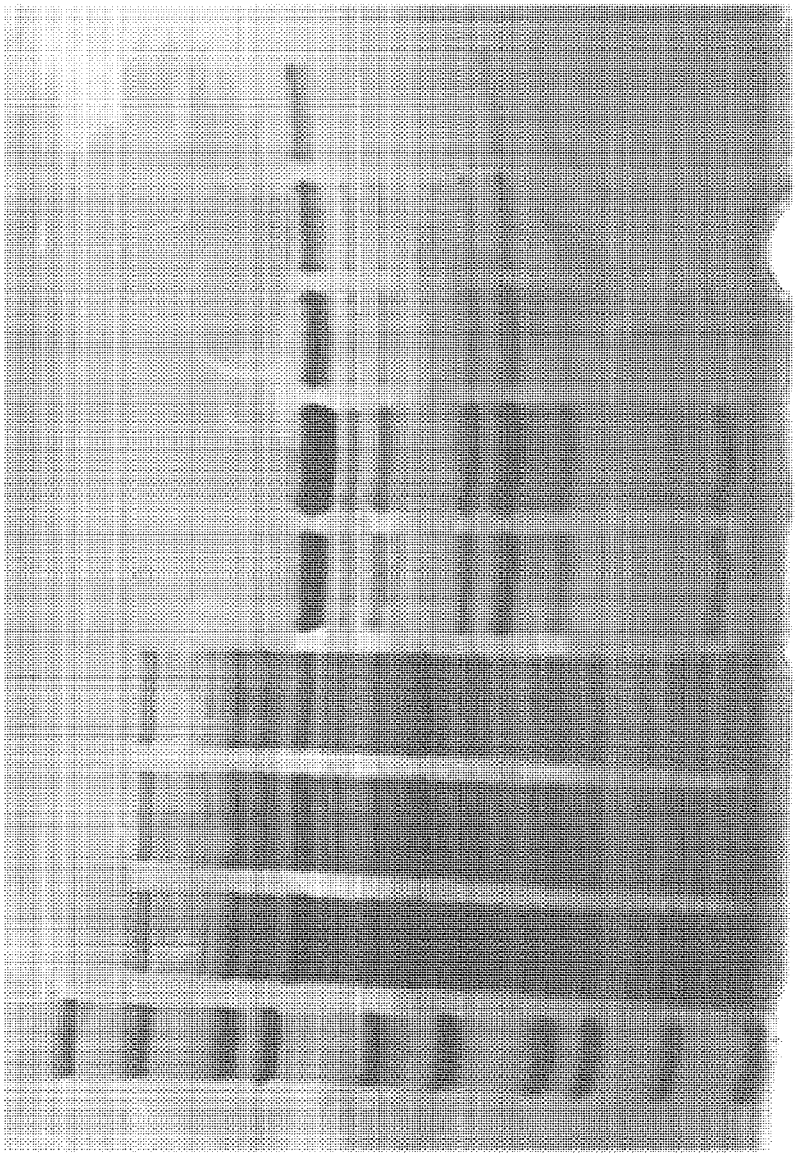
IDT Exhibit 1017.1925

FIG. 9



IDT Exhibit 1017.1926

FIG. 10A



IDT Exhibit 1017.1927

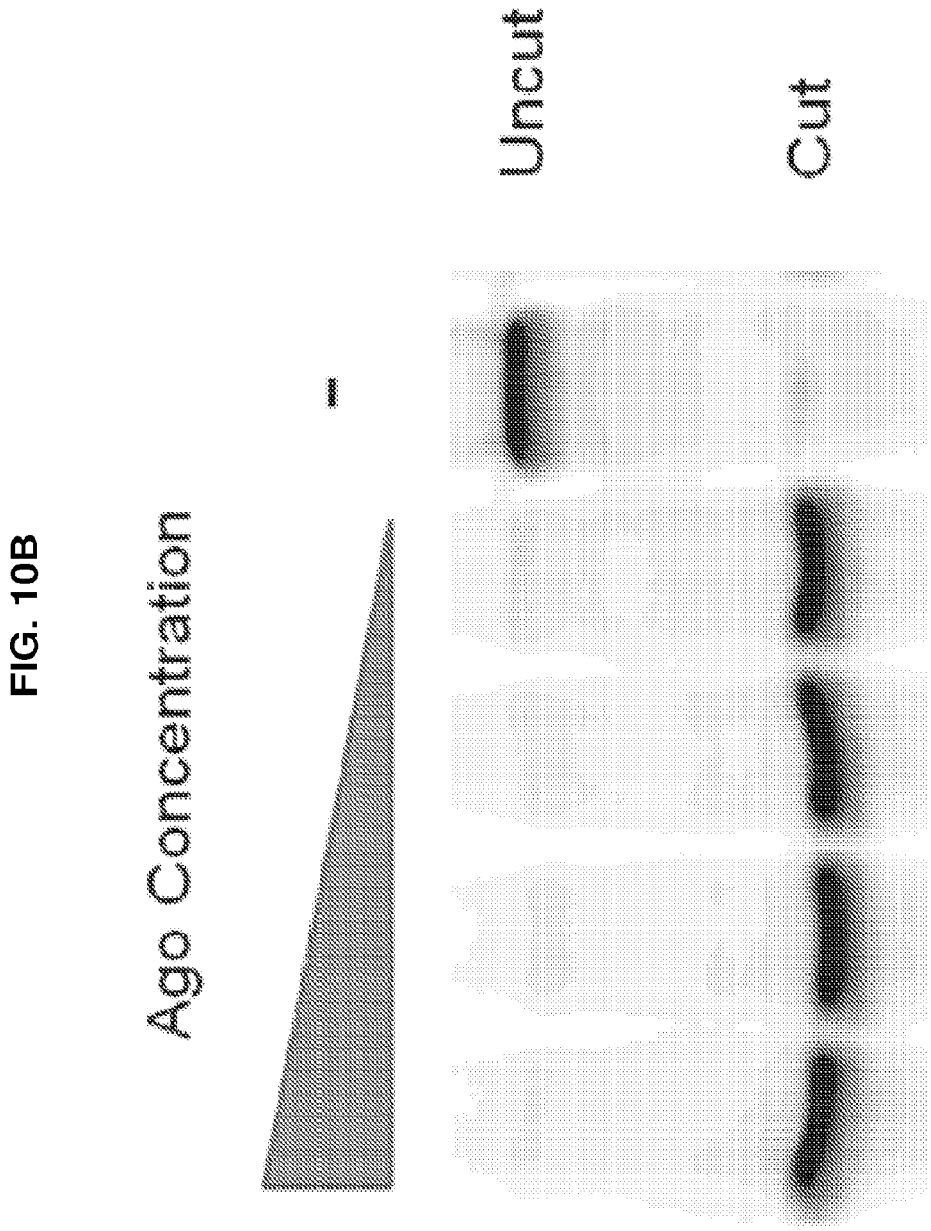
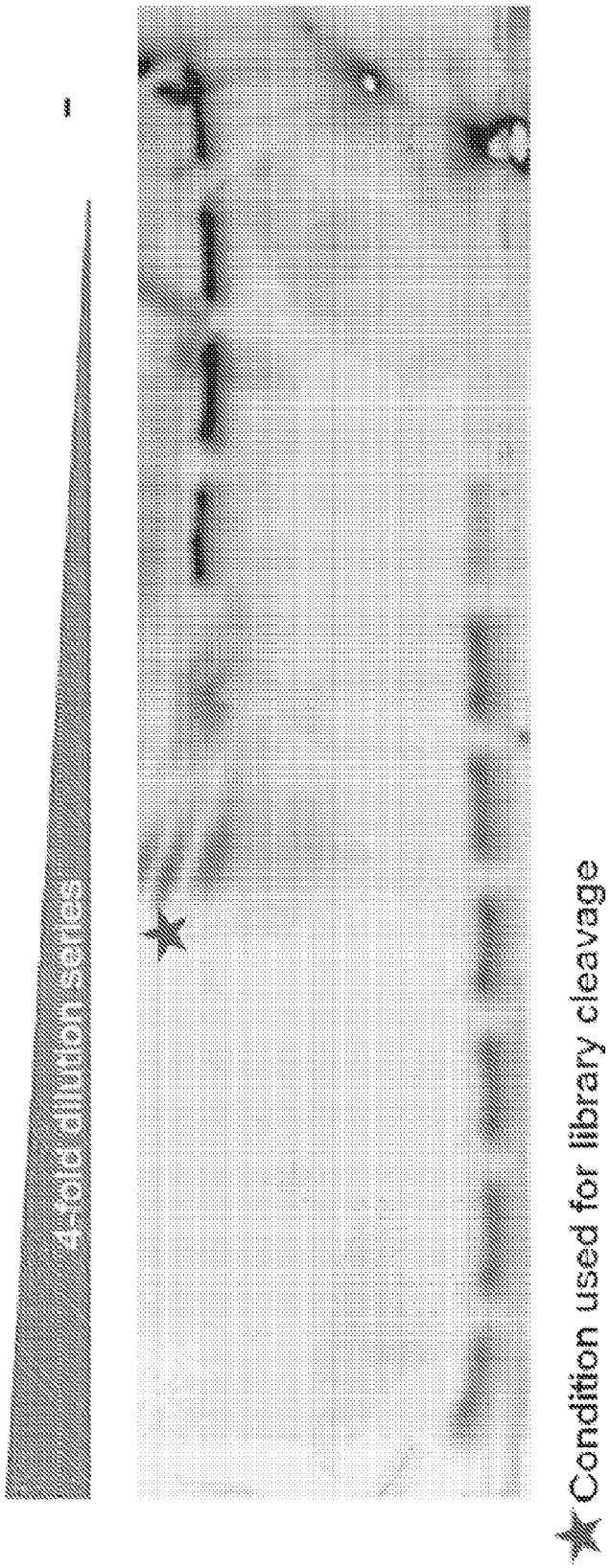


FIG. 10B

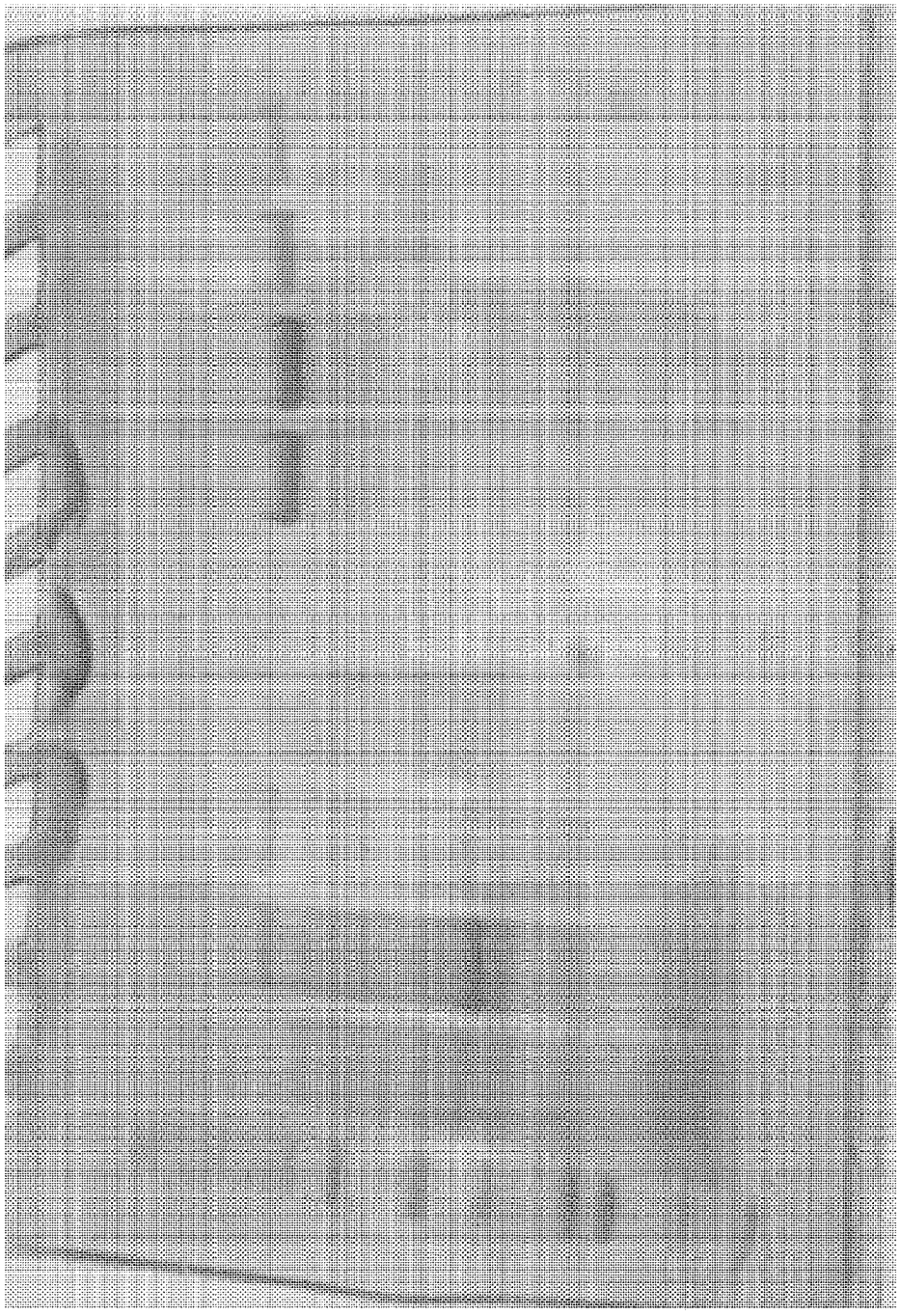
IDT Exhibit 1017.1928

FIG. 11



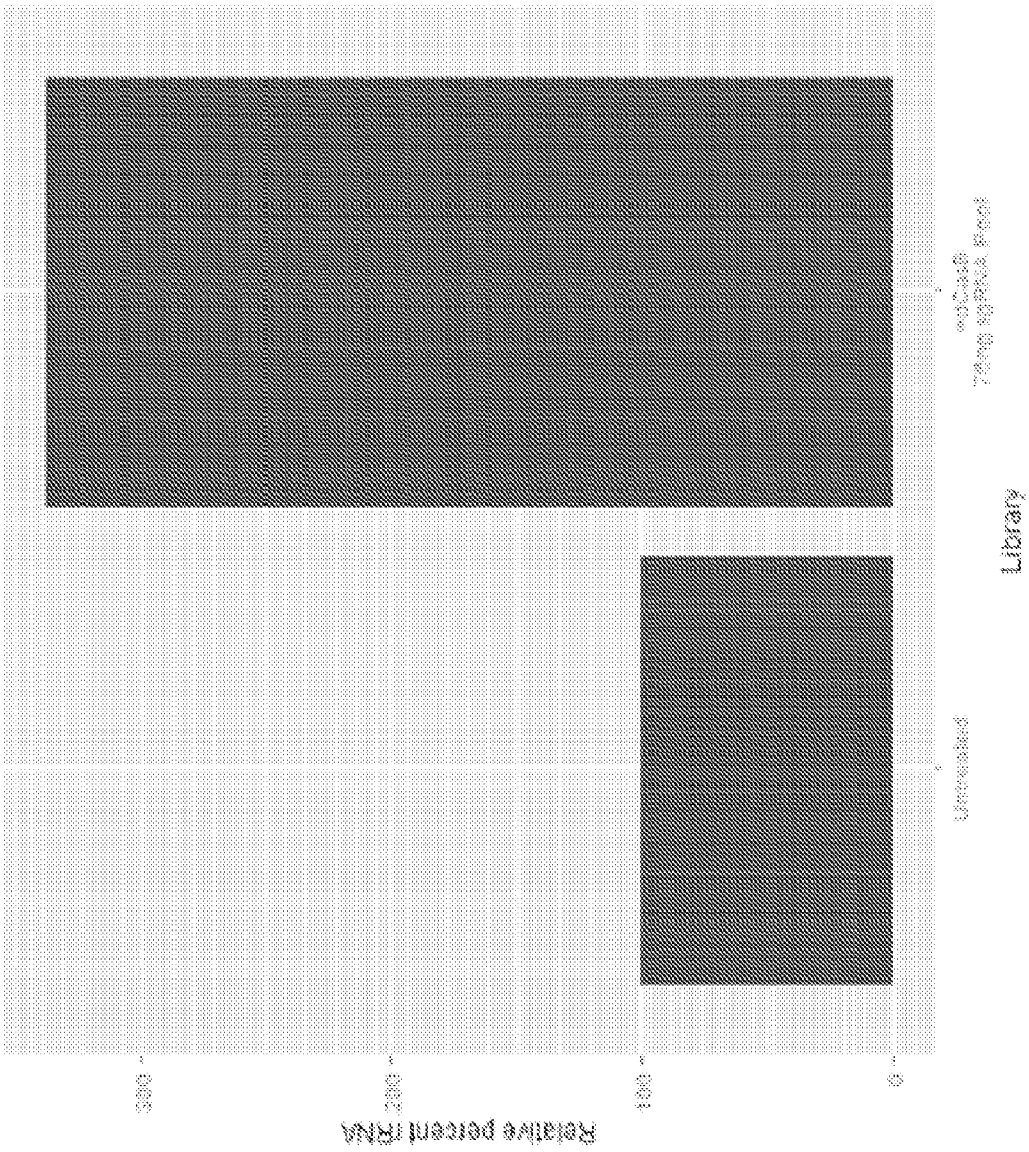
IDT Exhibit 1017.1929

FIG. 12A



IDT Exhibit 1017.1930

FIG. 12B



IDT Exhibit 1017.1931

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2014/072293

A. CLASSIFICATION OF SUBJECT MATTER

C12N 15/10(2006.01)i, C12Q 1/68(2006.01)i, G06F 19/10(2011.01)i

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C12N 15/10; C12N 5/06; C12N 5/08; C12P 21/02; C12N 15/87; C12Q 1/68; G06F 19/10

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
Korean utility models and applications for utility models
Japanese utility models and applications for utility models

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
eKOMPASS(KIPO internal) & keywords: nuclease, nucleic acid guide, guide RNA, depletion, target, sgRNA scaffold

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2013-141680 A1 (VILNIUS UNIVERSITY) 26 September 2013 See abstract; page 11, lines 1-3; claims 1-5, 23-28 and 35-41; figures 1-4 and 14.	1-4,11-14
A		15
X	HSU, P. D. et al., `DNA targeting specificity of RNA-guided Cas9 nucleases`, Nature Biotechnology, September 2013(Epub. 21 July 2013), Vol. 31, No. 9, pp. 827 832(NIH Public Access Author Manuscript Version internal pages 1 14). See abstract; page 9, lines 3-9; figures 1-2.	1-4,11-13,15
A		14
X	MALI, P. et al., `RNA-guided human genome engineering via Cas9`, Science, 15 February 2013(Epub. 3 January 2013), Vol. 339, No. 6121, pp. 823-826. See abstract; page 823; figures 1-2.	1-4,11-14
A		15
A	US 2006-0188987 A1 (GUSCHIN, D. et al.) 24 August 2006 See abstract; claims 1-11.	1-4,11-15
A	US 2003-0232410 A1 (LILJEDAHL, M. et al.) 18 December 2003 See abstract; claim 1.	1-4,11-15

☐ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents,such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
24 April 2015 (24.04.2015)	**27 April 2015 (27.04.2015)**

Name and mailing address of the ISA/KR
International Application Division
Korean Intellectual Property Office
189 Cheongsa-ro, Seo-gu, Daejeon Metropolitan City, 302-701, Republic of Korea
Facsimile No. ++82 42 472 7140

Authorized officer
HEO, Joo Hyung
Telephone No. +82-42-481-8150

Form PCT/ISA/210 (second sheet) (January 2015)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2014/072293

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☒ Claims Nos.: 7-10
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
Claims 8-9 are unclear since they are referring to the multiple dependent claims which do not comply with PCT Rule 6.4(a).
Claim 7 is worded in reference to the complexes and the product composition of any of the preceding claims. However, the complexes and the product composition have not been worded in any of the preceding claims. Therefore, claim 7 and claims 8-10, substantially depending on claim 7, are not clear and concise contrary to PCT Article 6.

3. ☒ Claims Nos.: 5-7, I0
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

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

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

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

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

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

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

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
☐ No protest accompanied the payment of additional search fees.

IDT Exhibit 1017.1933

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.
PCT/US2014/072293

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2013-141680 A1	26/09/2013	CA 2867849 A1	26/09/2013
		WO 2013-142578 A1	26/09/2013
US 2006-0188987 A1	24/08/2006	AU 2004-263865 A1	17/02/2005
		AU 2004-263865 B2	17/05/2007
		AU 2005-220148 A1	15/09/2005
		AU 2005-220148 B2	10/12/2009
		AU 2007-267887 A1	06/12/2007
		CA 2534296 A1	17/02/2005
		CA 2534296 C	26/03/2013
		CA 2554966 A1	15/09/2005
		CA 2554966 C	19/11/2013
		CA 2651494 A1	06/12/2007
		EP 1651660 A2	03/05/2006
		EP 1720995 A2	15/11/2006
		EP 2027262 A2	25/02/2009
		EP 2027262 B1	31/03/2010
		EP 2213731 A1	04/08/2010
		EP 2213731 B1	04/12/2013
		JP 04555292 B2	29/09/2010
		JP 05102274 B2	19/12/2012
		JP 05266210 B2	21/08/2013
		JP 2007-501626 A	01/02/2007
		JP 2009-538138 A	05/11/2009
		JP 2010-046099 A	04/03/2010
		KR 10-2006-0039019 A	04/05/2006
		US 2005-0064474 A1	24/03/2005
		US 2007-0218528 A1	20/09/2007
		US 2008-0131962 A1	05/06/2008
		US 2009-0305346 A1	10/12/2009
		US 2009-0311787 A1	17/12/2009
		US 2011-0030076 A1	03/02/2011
		US 2014-0065667 A1	06/03/2014
		US 7888121 B2	15/02/2011
		US 7914796 B2	29/03/2011
		US 7972854 B2	05/07/2011
		US 8034598 B2	11/10/2011
		US 8409861 B2	02/04/2013
		US 8524500 B2	03/09/2013
		WO 2005-014791 A2	17/02/2005
		WO 2005-014791 A3	16/02/2006
		WO 2005-084190 A2	15/09/2005
		WO 2005-084190 A3	10/08/2006
		WO 2007-139898 A2	06/12/2007
		WO 2007-139898 A3	03/04/2008
US 2003-0232410 A1	18/12/2003	AT 531796 T	15/11/2011
		AU 2003-218382 A1	08/10/2003
		AU 2003-218382 B2	13/12/2007

IDT Exhibit 1017.1934

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

PCT/US2014/072293

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
		CA 2479858 A1	02/10/2003
		EP 1504092 A2	09/02/2005
		EP 1504092 A4	08/08/2007
		EP 1504092 B1	02/11/2011
		EP 1504092 B2	25/06/2014
		EP 2368982 A2	28/09/2011
		EP 2368982 A3	12/10/2011
		HK 1073331 A1	08/06/2012
		US 2008-0209587 A1	28/08/2008
		US 2009-0305402 A1	10/12/2009
		WO 03-080809 A2	02/10/2003
		WO 03-080809 A3	16/12/2004

IDT Exhibit 1017.1935

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
17 August 2006 (17.08.2006)

PCT

(10) International Publication Number
WO 2006/086668 A2

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

(21) International Application Number:
PCT/US2006/004798

(22) International Filing Date: 9 February 2006 (09.02.2006)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/651,409 9 February 2005 (09.02.2005) US
60/685,367 27 May 2005 (27.05.2005) US

(71) Applicant *(for all designated States except US)*: **EPICENTRE TECHNOLOGIES** [US/US]; 726 Post Road, Madison, WI 53713 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **JENDRISAK, Jerome** [US/US]; 2721 Mason Street, Madison, WI 53705 (US). **MEIS, Judith** [US/US]; 2728 Tami Trail, Fitchburg, WI 53129 (US). **MEIS, Ronald** [US/US]; 2728 Tami Trail, Fitchburg, WI 53129 (US). **RADEK, Agnes** [HU/US]; 6510 Piedmont Road, Madison, WI 53711 (US). **DAHL, Gary** [US/US]; 2131 Kendall Avenue, Madison, WI 53726 (US).

(74) Agents: **CASIMIR, David, A.** et al.; Medlen & Carroll, LLP, 101 Howard Street, Suite 350, San Francisco, CA 94105 (US).

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

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 2006/086668 A2

(54) Title: COMPOSITIONS AND METHODS EMPLOYING 5'-PHOSPHATE-DEPENDENT NUCLEIC ACID EXONUCLEASES

(57) Abstract: The present invention relates to compositions and methods employing 5'- phosphate-dependent nucleic acid exonucleases. In particular, the present invention provides kits and methods employing 5 '-phosphate-dependent nucleic acid exonucleases for selective enrichment, isolation and amplification of a particular set of desired nucleic acid molecules from samples that also contain undesired nucleic acid molecules for a variety of uses. In preferred embodiments, the desired nucleic acid molecules comprise prokaryotic and/or eukaryotic mRNA.

IDT Exhibit 1017.1936

COMPOSITIONS AND METHODS EMPLOYING 5'-PHOSPHATE-DEPENDENT NUCLEIC ACID EXONUCLEASES

The present invention claims priority to U.S. Provisional Patent Application Serial Numbers 60/651,409, filed February 9, 2005 and 60/685,367 filed May 27, 2005, each of which is herein incorporated by reference in its entirety.

FIELD OF THE INVENTION

The present invention relates to compositions and methods employing 5'-phosphate-dependent nucleic acid exonucleases. In particular, the present invention provides kits and methods for employing 5'-phosphate-dependent nucleic acid exonucleases for selective enrichment, isolation and amplification of a particular set of desired nucleic acid molecules from samples that also contain undesired nucleic acid molecules for a variety of uses. In preferred embodiments, the desired nucleic acid molecules comprise prokaryotic and/or eukaryotic mRNA.

BACKGROUND OF THE INVENTION

With the enormous increase in the amount of bacterial genome sequence information over the last several years and the complete sequencing of a number of microbial genomes, the molecular biology field has entered a post-genomic era (Pang et al., Microbiol. Immunol., 48:91, 2004). The investigation of gene expression at the transcriptional level is of basic interest in microbial ecology and pathogenicity studies. This investigation will bridge the gap between structural and functional diversity. Expression analysis is being used to identify gene functions and metabolic pathways in many organisms, including humans, yeast, *Drosophila*, mice, and bacteria.

A major challenge in prokaryotic expression analysis is the preparation and analysis of prokaryotic mRNA. Oligo(dT) selection for poly(A) tails has long been used for isolating mRNAs from eukaryotic sources. However, the lack of relatively stable poly(A) tails, and their short half-lives and low quantity for bacteria make isolation of bacterial mRNA difficult (Adel et al., Nature. Biotechnol. 18:679, 2000) and Coller et al., Proc. Natl. Acad. Sci. U.S.A. 97:3260, 2000). Thus, isolation of mRNA from bacteria has been vitally important, but difficult. There has been an urgent need to develop rapid

IDT Exhibit 1017.1937

and simple methods to purify mRNA from bacteria. Several attempts are described below. However, none have provided sufficient solutions to the problem.

Wendisch et al. (Anal. Biochem. 290:205, 2001); U.S. Pat. No. 6,242,189, herein incorporated by reference in their entireties) developed a method to purify the whole population of cellular mRNAs by polyadenylation with *E. coli* poly(A) polymerase in crude cell extracts obtained by mechanical lysis. However, this method does not select for only mRNA, so other RNA molecules can be polyadenylated in addition to mRNA.

Affymetrix Inc. (Rosenow et al., Nucleic Acids Res. 29(22):e112, 2001) obtained enriched mRNA from *E. coli* with a series of enzymatic steps that specifically eliminate the 16S and 23S rRNA species in the total RNA. Reverse transcriptase and primers specific for 16S and 23S rRNA are used to synthesize complementary DNAs. Then rRNA is removed enzymatically by treatment with RNase H, which specifically digests RNA within an RNA:DNA hybrid. The cDNA molecules are then removed by DNase I digestion and the enriched mRNA is purified on QIAGEN RNeasy columns. The enriched mRNA will contain mRNAs, tRNAs, 5S rRNA, and other small RNAs. This method is complex and difficult to handle, and introduces the risk of mRNA losses due to mispriming of mRNA with rRNA primers.

Ambion Company developed the MICROBE EXRESS Bacterial mRNA Purification Kit that employs a modified capture hybridization approach, to remove abundant 16S and 23S ribosomal RNAs (rRNA) from purified total RNA and enrich bacterial mRNA. Briefly, purified RNA is incubated with the Capture Oligonucleotide Mix in Binding Buffer. Magnetic beads, derivatized with an oligonucleotide that hybridizes to the capture oligonucleotide, are then added to the mixture and allowed to hybridize. The magnetic beads, with 16S and 23S rRNAs attached, are pulled to the side of the tube with a magnet. The enriched RNA in the supernatant is removed and precipitated with ethanol. The enriched mRNA will contain mRNAs, tRNAs, 5S rRNA, and other small RNAs. The process is somewhat tedious and cumbersome, and cannot be applied to all species of bacteria. U.S. Pat. Appln. No. 2003/0175709 likewise describes a method for depleting unwanted RNA species using capture probes that hybridize to bridging oligonucleotides configured to bind to capture beads and target sequences.

IDT Exhibit 1017.1938

Pang et al., *supra*, demonstrate that magnetic capture-hybridization methods may be used for the purification of bacterial mRNAs, using biotin-labeled oligonucleotides as capture probes specific for 5S, 16S and 23S rRNA of bacteria. Ribosomal RNAs hybridize to biotin-labeled oligonucleotide capture probes that are fixed to streptavidin-coated paramagnetic beads. The mRNA remains in the supernatant and is recovered by ethanol precipitation. While this method enriched mRNA, improvements in efficiency are needed.

Thus, the art is in need of improved compositions and methods that provide high efficiency and ease of use for the enrichment and purification of bacterial mRNA for a variety of applications. What is needed are rapid and simple methods for isolation and purification of bacterial mRNA that avoid the tedious and cumbersome use of magnetic beads or purification columns, and/or the synthesis of ribosomal cDNA and subsequent RNase H digestion.

What is further needed in the art are compositions and methods for preparation of bacterial mRNA, whereby said compositions and methods remove 3' mRNA fragments that do not have the 5'-end of primary mRNA transcripts (such as 3' mRNA fragments resulting from exposure of primary mRNA transcripts to a ribonuclease or to physical forces such as shearing).

What is further needed are compositions and methods to prepare cDNA from bacterial mRNA and to amplify bacterial mRNA for gene expression studies such as, but not limited to, for microarray analyses. Moreover, what is needed are compositions and methods for preparation of cDNA from bacterial mRNA, whereby the cDNA is synthesized from full-length mRNA, meaning from mRNA sequences that are not truncated at the 5'-end or the 3'-end.

What is further needed are compositions and methods to isolate bacterial mRNA, to prepare cDNA from bacterial mRNA, and to amplify bacterial mRNA for gene expression studies from samples in which the bacteria are associated with cells of one or more other prokaryotic and/or eukaryotic organisms. By way of example, but not of limitation, what is needed are compositions and methods for purifying bacterial mRNA, making bacterial cDNA, and amplifying bacterial mRNA from Rhizobium nitrogen-fixing bacteria in the roots of legumes, from biofilms such as in the human mouth, or

IDT Exhibit 1017.1939

from pathogenic bacteria or mycoplasma in association with a plant, animal, human, or fungal host. Moreover, what is needed are improved compositions and methods for isolating mRNA, making cDNA, and amplifying mRNA from both pathogen (or symbiont) and host cells at the same time in order to simultaneously study gene expression in both organisms that are associated in pathogen-host (or symbiotic) interactions (or even to simultaneously study gene expression in multiple bacterial species associated in biofilms of any type, or in multiple cells of a single organism or of multiple organisms in any type of association).

What is also needed are compositions and methods that enable synthesis of cDNA and amplification of mRNA from bacteria that are difficult to grow in cell culture.

What is also needed in the art are improved compositions and methods to purify eukaryotic mRNA. What is needed are rapid and simple methods for isolation and purification of eukaryotic mRNA that avoid the tedious and cumbersome use of magnetic beads, purification columns, or membranes that bind polyadenylated nucleic acids (e.g., with oligo(dT)). Moreover, what is needed are compositions and methods for preparation of eukaryotic mRNA, whereby said compositions and methods remove mRNA fragments that do not have the 5'-end of a full-length mRNA transcript (such as 3' mRNA fragments resulting from exposure of the mRNA to a ribonuclease or to physical forces such as shearing and/or heat).

What is further needed are improved compositions and methods to prepare cDNA from eukaryotic mRNA and to amplify eukaryotic mRNA for gene expression studies such as, but not limited to, for microarray analyses. Moreover, what is needed are compositions and methods for preparation of cDNA from eukaryotic mRNA, whereby the cDNA is synthesized from full-length mRNA, meaning from mRNA sequences that are not truncated at the 5'-end or the 3'-end. What is further needed are improved compositions and methods for isolating mRNA, making cDNA, and amplifying mRNA from eukaryotes in a high-throughput manner.

What is also needed are improved kits and methods for analyzing gene expression in biological samples containing eukaryotic mRNA or both eukaryotic and prokaryotic mRNA. For example, it can be very difficult to obtain mRNA, make cDNA, and amplify mRNA in biological samples that contain degraded RNA, such as, but not limited to

IDT Exhibit 1017.1940

slides of formalin-fixed paraffin-embedded ("FFPE") tissue sections. Samples of cells from which it is desired to profile gene expression can be obtained from such sections by various methods known in the art. One method that can be used to obtain sample of cells from such tissue sections is laser capture, such as but not limited to, laser capture microdissection, using methods known in the art. Instruments for laser capture are available commercially from Arcturus, PALM, and other sources. Methods are also known in the art for isolating total RNA from samples from FFPE tissue sections. One such method that can be used is described by Acturus in the product literature for their Paradise™ Kit. Other methods are also known in the art that can be used.

However, once total RNA is obtained, it can be difficult to obtain good quality mRNA for gene expression analysis by methods known in the art, such as but not limited to, analysis by microarrays. In part, this is true because many methods rely on use of oligo(dT) resins or membranes to isolate poly(A) containing RNA, which is a common method for isolating polyadenylated mRNA. The problem is that much of the mRNA may be degraded into fragments that do not have a poly(A) tail and will not be isolated using such oligo(dT) columns or membranes. Also, the mRNA fragments that have a poly(A) tail and therefore are obtained using the oligo(dT) binding method may not contain the sequences that are complementary to many of the oligos on an array or microarray. Therefore, the gene expression analysis can be difficult to interpret. It would be desirable to obtain mRNA samples more easily, particularly for making cDNA, for RNA amplification and other methods of amplification, and especially, for obtaining better and more informative gene expression data than current methods.

SUMMARY OF THE INVENTION

The present invention relates to compositions and methods employing 5'-phosphate-dependent nucleic acid exonucleases. In particular, the present invention provides kits and methods employing 5'-phosphate-dependent nucleic acid exonucleases for selective enrichment, isolation and amplification of a particular set of desired nucleic acid molecules from samples that also contain undesired nucleic acid molecules for a variety of uses. In preferred embodiments, the desired nucleic acid molecules comprise prokaryotic and/or eukaryotic mRNA.

IDT Exhibit 1017.1941

The present invention comprises methods for using a 5' exoribonuclease for a number of applications related to biological research on gene expression, diagnostics, and human and animal therapeutics.

One preferred method of the invention comprises a method for enriching for an RNA having a 5'-triphosphate or a 5'-cap in a biological sample comprising prokaryotic RNA, eukaryotic RNA or both prokaryotic and eukaryotic RNA and at least one undesired nucleic acid, the method comprising treating the sample with purified 5' exoribonuclease under conditions in which the 5' exoribonuclease is active and for sufficient time so that the undesired nucleic acid is digested and the sample is enriched for RNA having a 5'-triphosphate or a 5'-cap.

In different embodiments of this method, the RNA having a 5'-triphosphate or a 5'-cap is selected from the group consisting of: (i) prokaryotic mRNA; (ii) eukaryotic mRNA, including polyadenylated and non-polyadenylated eukaryotic mRNA; (iii) a mixture of both prokaryotic and eukaryotic mRNA; (iv) eukaryotic snRNA; (v) eukaryotic pre-micro RNA; and (vi) prokaryotic or eukaryotic primary RNA transcripts of unknown function.

Another preferred embodiment of this method comprises treating the biological sample with a polynucleotide kinase to phosphorylate the RNA having 5'-hydroxyl groups prior to treating the sample with purified 5' exoribonuclease.

In one embodiment of the method, the RNA having a 5'-triphosphate or a 5'-cap comprises eukaryotic mRNA or both prokaryotic and eukaryotic mRNA, and the method additionally comprises binding the RNA to an oligo(dT) or oligo(dU) resin, membrane or other surface to which oligo(dT) or oligo(dU) is attached and eluting the bound RNA so as to obtain a solution containing polyadenylated eukaryotic mRNA.

In another embodiment of the method, the RNA having a 5'-triphosphate or a 5'-cap comprises prokaryotic mRNA, eukaryotic mRNA or both prokaryotic and eukaryotic mRNA and the undesired nucleic acid comprises prokaryotic rRNA, eukaryotic rRNA or both prokaryotic and eukaryotic rRNA of a size having a svedburg unit greater than about 10S. In specific embodiments of this method the rRNA is selected from the group consisting of prokaryotic rRNA having a svedburg unit of about 16S or 23S and eukaryotic rRNA having a svedburg unit of about 18S, 26S or 28S.

IDT Exhibit 1017.1942

Another embodiment of the invention is a method wherein the RNA having a 5'-triphosphate or a 5'-cap comprises prokaryotic mRNA, eukaryotic mRNA, or both prokaryotic and eukaryotic mRNA, and wherein the method additionally comprises precipitation of said mRNA having a 5'-triphosphate or a 5'-cap with a solution of LiCl and ethanol at concentrations and under conditions in which tRNA and 5S rRNA are not precipitated. In some embodiments of this method, the RNA having a 5'-triphosphate or a 5'-cap comprises prokaryotic mRNA, eukaryotic mRNA, or both prokaryotic and eukaryotic mRNA and the undesired nucleic acid comprises prokaryotic rRNA, eukaryotic rRNA or both prokaryotic and eukaryotic rRNA of a size having a svedburg unit greater than about 10S. In specific embodiments, the rRNA is selected from the group consisting of prokaryotic rRNA having a svedburg unit of about 16S or 23S and eukaryotic rRNA having a svedburg unit of about 18S, 26S or 28S.

Still another method of the invention that uses RNA obtained following 5' exoribonuclease treatment comprises: (i) contacting the sample enriched for RNA having a 5'-triphosphate or a 5'-cap with a poly(A) polymerase under conditions so that polyadenylated RNA is obtained; and (ii) contacting the polyadenylated RNA with an oligo(dT)-containing primer and an RNA-dependent DNA polymerase under conditions that cDNA complementary to the RNA is obtained. In some specific embodiments, the oligo(dT)-containing primer comprises a T7-type RNA polymerase promoter selected from the group consisting of T7 RNA polymerase, T3 RNA polymerase, and SP6 RNA polymerase. In one embodiment of this method, the biological sample is treated with a polynucleotide kinase to phosphorylate the RNA having 5'-hydroxyl groups prior to treating the sample with purified 5' exoribonuclease. In some other specific embodiments of this method, the RNA having a 5'-triphosphate or a 5'-cap comprises prokaryotic mRNA, eukaryotic mRNA, or both prokaryotic and eukaryotic mRNA and the undesired nucleic acid comprises prokaryotic rRNA, eukaryotic rRNA or both prokaryotic and eukaryotic rRNA of a size having a svedburg unit greater than about 10S. In specific embodiments, the rRNA is selected from the group consisting of prokaryotic rRNA having a svedburg unit of about 16S or 23S and eukaryotic rRNA having a svedburg unit of about 18S, 26S or 28S.

IDT Exhibit 1017.1943

In still another method, wherein the biological sample comprises eukaryotic RNA, and wherein the method additionally comprises contacting the sample enriched for RNA having a 5'-triphosphate or a 5'-cap with an oligo(dT)-containing primer and an RNA-dependent DNA polymerase under conditions that cDNA complementary to the RNA is obtained. In some specific embodiments, the oligo(dT)-containing primer comprises a T7-type RNA polymerase promoter selected from the group consisting of T7 RNA polymerase, T3 RNA polymerase, and SP6 RNA polymerase. In some of these embodiments, the method additionally comprises contacting the cDNA complementary to the RNA with one or more DNA or RNA primers and a DNA polymerase and incubating under conditions so as to obtain double-stranded cDNA having a functional promoter for the T7-type RNA polymerase and synthesizing RNA therefrom.

In still other embodiments, the method additionally comprises contacting the cDNA complementary to the RNA with an oligonucleotide sequence tag template and a DNA polymerase and incubating under conditions so as to obtain cDNA having a sequence tag on it the 3'-end. In some embodiments, the method additionally comprises contacting the cDNA having the sequence tag on its 3'-end with a DNA polymerase and an oligonucleotide primer, wherein the either the sequence tag or the oligonucleotide primer, or the combination of both the sequence tag and the oligonucleotide primer encodes the complete sequence of a T7-type RNA polymerase promoter, and incubating under conditions wherein double-stranded cDNA having a functional T7-type RNA polymerase promoter is obtained. In those cases, the method additionally comprises contacting the cDNA with the T7-type RNA polymerase under conditions so as to synthesize RNA using said T7-type RNA polymerase promoter. In some other specific embodiments of these methods, the biological sample is treated with a polynucleotide kinase to phosphorylate the RNA having 5'-hydroxyl groups prior to treating the sample with purified 5' exoribonuclease.

In preferred embodiments, the compositions, kits, and methods of the present invention employ Xrn1p/5' Exoribonuclease I and functional variants, homologues, and equivalents. Xrn1p/5' exoribonuclease I is a magnesium-dependent, 5'-to-3', processive exoribonuclease that acts preferentially on RNA substrates with a 5' phosphate (Stevens, Biochem. Biophys. Res. Commun. 81:656, 1978); Stevens, Biochem. Biophys. Res.

IDT Exhibit 1017.1944

Commun. 86:1126, 1979), Stevens, J. Biol. Chem. 255:3080, 1980); and Stevens and Maupin, Nucleic Acids Res. 15:695, 1987). Nucleic acid molecules with a cap, a triphosphate group, or an hydroxyl group on their 5'-termini (or 5'-ends) are not substantially digested by the enzyme. The exoribonuclease I is not inhibited by proteinaceous RNase inhibitors such as RNASIN (Promega, Madison, WI) or PRIME RNase Inhibitor (Eppendorf, Brinkmann, Westbury, NY). In some embodiments of the present invention, the 5'-phosphate-dependent nucleic acid exonuclease is obtained as described by Stevens (J. Biol. Chem. 255:3080, 1980). In some other embodiments, the 5'-phosphate-dependent nucleic acid exonuclease is obtained by cloning a gene for an exonuclease, such as but not limited to a Saccharomyces Xrn I gene, in a vector, expressing the gene in a host cell, and purifying the expressed protein from cultures of said host cells. For example, but without limitation, in some embodiments, the 5'-phosphate-dependent nucleic acid exonuclease is obtained from recombinant yeast by following the protocol of Johnson and Kolodner (J. Biol. Chem. 266:14046, 1991). In other embodiments, the 5'-phosphate-dependent nucleic acid exonuclease is obtained by cloning the gene for an exonuclease in a vector, expressing the gene in a heterologous host cell, such as but not limited to an *E. coli* host cell, and purifying the expressed protein from cultures of said host cells. Such a composition from recombinant source finds use for more efficient and less expensive production, as well as a higher purity and more consistent quality of 5'-phosphate-dependent nucleic acid exonuclease for use in the methods and kits of the present invention. The expression vector may further express other sequences that assist in the purification, expression, detection, or use of the 5'-phosphate-dependent nucleic acid exonucleases.

The exonucleases of the present invention find use in a number of applications including, but not limited to: analysis of the 5' ends of RNA molecules; removal of ribosomal RNA from RNA preparations; enrichment of prokaryotic and/or eukaryotic mRNA; enrichment of mixtures of prokaryotic and/or eukaryotic mRNA; preparation of prokaryotic and/or eukaryotic mRNA for cDNA synthesis; synthesis of amplified antisense or sense RNA by preparing prokaryotic mRNA using exoribonuclease, then polyadenylating the RNA with poly(A) polymerase, then making cDNA joined to an RNA polymerase promoter, and transcribing the cDNA from the promoter *in vitro*;

IDT Exhibit 1017.1945

enrichment of 5'-capped or 5'-triphosphorylated small nuclear RNAs (snRNA); removal of 5'-phosphorylated splice template oligos (or template switch oligos) and other oligos (e.g., ligation splints) from *in vitro* transcription reactions, including antisense or sense RNA amplification reactions, in order to reduce nonspecific background transcription in RNA amplification and other reactions; removal of linear RNA from circular RNA preparations; removal of linear first-strand cDNA molecules from circular first-strand cDNA molecules, diagnostic applications; and the like.

For example, in some embodiments, the present invention provides a method for isolating a nucleic acid molecule of interest in a mixture of nucleic acids, comprising the steps of: a) providing a 5'-to-3' exoribonuclease and a biological sample comprising a mixture of nucleic acids, said mixture of nucleic acids comprising a nucleic acid molecule of interest (i.e., a desired nucleic acid) and an undesired nucleic acid having a 5'-phosphate; and b) exposing the 5'-to-3' exoribonuclease to the biological sample under conditions such that the exoribonuclease digests the undesired nucleic acid so as to enrich the sample for the nucleic acid molecule of interest. The present invention is not limited by the nature of the 5'-phosphate-dependent nucleic acid exonuclease. In preferred embodiments, the 5'-phosphate-dependent nucleic acid exonuclease includes, but is not limited to, Xrn1p/5' exoribonuclease 1, fragments of Xrn1p/5' exoribonuclease 1 (e.g., that have exoribonuclease activity), and Xrn1p/5' exoribonuclease 1 variants (e.g., containing conserved amino acid changes, truncations, chimeras, etc.). A 5' exonuclease of the present invention comprises any 5'-exonuclease that has greater than 20-fold more 5'-to-3' exonuclease activity for a single-stranded RNA substrate that has a 5'-monophosphorylated terminus than for the same RNA substrate that has a 5'-triphosphorylated or 5'-capped terminus. That is, a 5' exonuclease of the present invention comprises any 5'-to'3' exonuclease that has a relative activity in digesting a particular defined-sequence single-stranded RNA substrate with a 5'-triphosphate or a 5'-cap that is less than 5% of the activity of an RNA substrate having the same sequence but with a 5'-monophosphate. Enzyme activity of a 5' exonuclease of the invention can be measured using a number of different methods. Without limitation, suitable methods that can be used for assaying activity and determining relative activity using RNA substrates with a

IDT Exhibit 1017.1946

5'-triphosphate, a 5'-cap, or a 5'-monophosphate are described by Stevens and Poole (J. Biol. Chem. 270: 16063, 1995).

The present invention is also not limited by the nature of the sample. Samples include, but are not limited to, cell lysates, provided that nucleases that can affect nucleic acids of interest are removed or inhibited in said lysates, mixtures of nucleic acids (e.g., unpurified, partially purified, etc.), environmental samples, etc. Samples can comprise nucleic acids from a 1-cell organism, or if an organism comprises multiple cells, from one or more cells. Samples comprising nucleic acids from multiple cells can comprise cells from one or more types of cells, including cells from different organisms, and/or different cells from the same organism. The present invention finds use with prokaryotic nucleic acid, with eukaryotic nucleic acid, or with a mixture of both eukaryotic and prokaryotic nucleic acid.

In some preferred embodiments of the present invention, the sample comprises (or contains) RNA. A sample of the invention that comprises or contains RNA can contain total RNA from any source, which RNA can be obtained using, for example, a MasterPure RNA Purification Kit, an ArrayPure Nano-Scale RNA Purification Kit, or a MasterPure Yeast RNA Purification Kit (all of which are from EPICENTRE), or using a kit from another commercial source, or using a "home-brew" method known in the art. Alternatively, in some embodiments a sample of the invention that comprises or contains RNA can contain a subfraction of total RNA obtained by any method known in the art, such as, but without limitation, a subfraction based on size (e.g., by purification on an agarose or polyacrylamide gel, or by column purification, including by HPLC), or a subfraction obtained by salt precipitation (e.g., using precipitation with 0.5-2.5 M LiCl (Barlow, JJ et al., Biochem. Biophys. Res. Comm. 13: 61, 1963); Cathala, G et al, DNA 2: 329, 1983) or 2.5 M ammonium acetate). In some embodiments, a sample comprising RNA can also contain DNA. In some preferred embodiments, the nucleic acid molecules of interest comprise mRNA having a 5'-triphosphate or a 5'-cap. However, the nucleic acid molecule of interest (or desired nucleic acid) can comprise any nucleic acid molecule that is resistant to digestion by the 5'-phosphate-dependent nucleic acid exonuclease used, including, but not limited to nucleic acid molecules having a 5'-triphosphate or a 5'-cap, a 5'-hydroxyl group, or circularized nucleic acid molecules.

IDT Exhibit 1017.1947

The undesired nucleic acid can comprise any nucleic acid molecule that is digested by the 5'-phosphate-dependent nucleic acid exoribonuclease used, including, but not limited to nucleic acid molecules having a 5'-monophosphate or a 5'-diphosphate group, which are substrates for the enzyme. In some preferred embodiments, the undesired nucleic acid comprises rRNA that is greater than approximately 200 nucleotides (e.g., without limitation, containing each of 16S and 23S prokaryotic rRNA or 18S and 28S human, mouse, drosophila, or Xenopus rRNA or 18S and 26S yeast rRNA).

In some preferred embodiments, little or no detectable level of the desired nucleic acid molecule of interest is digested by the 5'-phosphate-dependent nucleic acid exonuclease. However, the invention is not limited by the amount of the desired nucleic acid molecule of interest that is digested so long as a substantial amount of the undesired nucleic acid is digested and sufficient amount of the desired nucleic acid remains for the intended purpose, which can vary for different purposes. For example, less of the desired RNA may be needed from some samples and/or applications (e.g., for RNA amplification of mRNA from total RNA in large samples) than are needed from other samples and/or applications (e.g., preparing a cDNA library from only a small number of cells). In preferred embodiments, a substantial amount of the undesired nucleic acid is digested by the 5' exonuclease of the invention, which means that at least 50% (e.g., 60%, 70%, 80%, 90%, 95%, 98%, 99%) of the starting amount (i.e., the amount present prior to treatment with the exonuclease) of the undesired nucleic acid is digested.

In some preferred embodiments, the present invention provides a method for isolating prokaryotic mRNA (e.g., to conduct gene expression analysis) comprising the steps of: a) providing a sample comprising RNA from a prokaryotic organism, said sample containing desired mRNA having a 5'-triphosphate group and at least one undesired nucleic acid; and a 5'-phosphate-dependent nucleic acid exonuclease; b) treating the sample with the 5'-phosphate-dependent nucleic acid exonuclease under conditions such that said undesired nucleic acid is substantially digested and said desired mRNA is not digested; and c) recovering the mRNA (e.g., for additional analysis or use). In some embodiments, the 5'-phosphate-dependent nucleic acid exonuclease used in the method is provided in a kit with buffers, instructions, and optionally, with appropriate control samples, and the like.

IDT Exhibit 1017.1948

In some preferred embodiments, the present invention provides a method for isolating prokaryotic mRNA molecules that have the 5'-ends of primary mRNA transcripts (e.g., to conduct gene expression analysis) comprising the steps of: a) providing a sample comprising RNA from a prokaryotic organism, said sample containing desired mRNA having a 5'-triphosphate group and at least one undesired nucleic acid; a polynucleotide kinase; and a 5'-phosphate-dependent nucleic acid exonuclease; b) treating the sample with the polynucleotide kinase in the presence of ATP and under conditions such that nucleic acid molecules having an hydroxyl group at their 5'-termini are phosphorylated; c) treating the sample with the 5'-phosphate-dependent nucleic acid exonuclease under conditions such that RNA molecules having a 5'-monophosphate group are substantially digested; and d) recovering the mRNA (e.g., for additional analysis or use). In some embodiments, the 5'-phosphate-dependent nucleic acid exonuclease used in the method is provided in a kit with polynucleotide kinase (e.g., T4 polynucleotide kinase), ATP, buffers, instructions, and optionally, with appropriate control samples, and the like.

In some embodiments, the present invention provides methods and kits for producing cDNA copies of prokaryotic mRNA (e.g., for making cDNA libraries). For example, the present invention provides a method for preparing prokaryotic cDNA, comprising the steps of a) providing a sample comprising RNA from a prokaryotic organism, said sample containing desired mRNA having a 5'-triphosphate group and at least one undesired nucleic acid; and a 5'-phosphate-dependent nucleic acid exonuclease; b) contacting the sample with the 5'-phosphate-dependent nucleic acid exonuclease under conditions such that undesired nucleic acids in the sample are substantially digested but mRNA molecules in the sample having a 5'-triphosphate are not digested; and c) producing cDNA copies of the mRNA. The present invention is not limited by the manner in which the cDNA copies are produced.

In some preferred embodiments, the method of making cDNA copies comprises: i) contacting the mRNA molecules with a poly(A) polymerase to produce mRNA molecules having a 3'-poly(A) tail; and ii) contacting the mRNA molecules having a 3'-poly(A) tail with a primer that anneals to said tail (e.g., an oligo(dT)-containing primer) and extending the primer with an RNA-dependent DNA polymerase under conditions

IDT Exhibit 1017.1949

whereby cDNA is obtained. Kits of the invention for conducting such methods may have any one or more reagents (e.g., primers), control samples, buffers, etc. useful in the method. For example, preferred kits comprise: a) a 5'-phosphate-dependent nucleic acid exonuclease; b) a poly(A) polymerase (e.g., without limitation, an *E. coli* poly(A) polymerase encoded by the *pcn*B gene); and c) an RNA-dependent DNA polymerase (e.g., an AMV reverse transcriptase; an MMLV reverse transcriptase; SuperScript I, SuperScript II, SuperScript III, or AMV ThermoScript reverse transcriptase (INVITROGEN); or MonsterScript reverse transcriptase (EPICENTRE).

In some other embodiments, the method of making cDNA copies comprises adding a sequence tag template to the 3'-end of the mRNA molecules using a method as described in U.S. Patent Application No. 2005/0153333, incorporated herein by reference. Thus, one embodiment for making cDNA copies of mRNA molecules comprises: i) contacting the mRNA molecules with a mixture of tagging oligonucleotides, each tagging oligonucleotide having a 5'-portion comprising the same sequence tag template, a 3'-portion comprising a random sequence and a blocked 3' terminus, under conditions in which one of the tagging oligonucleotides anneals with each mRNA molecule and the mRNA molecules are extended with a nucleic acid polymerase to form mRNA molecules having a sequence complementary to the sequence tag template on their 3'-termini; and ii) contacting the mRNA molecules having a sequence complementary to the sequence tag template on their 3'-termini with a primer that anneals to said sequence that is complementary to the sequence tag template and extending the primer with an RNA-dependent DNA polymerase under conditions whereby cDNA is obtained. In some aspects of this embodiment, the nucleic acid polymerase used to extend the mRNA molecules using the sequence tag template as a template in step (i) is the same enzyme as the RNA-dependent DNA polymerase used to obtain cDNA in step (ii), wherein, in other aspects of this embodiment, the nucleic acid polymerase used to extend the mRNA molecules using the sequence tag template as a template in step (i) is different than the RNA-dependent DNA polymerase used to obtain cDNA in step (ii). Kits for synthesizing cDNA according to the embodiment of the invention comprise: a) tagging oligonucleotides; b) a primer that anneals to the sequence that is complementary to the sequence tag template; and c) an RNA-dependent DNA

IDT Exhibit 1017.1950

polymerase (e.g., an AMV reverse transcriptase; an MMLV reverse transcriptase; SuperScript I, SuperScript II, SuperScript III, or AMV ThermoScript reverse transcriptase (INVITROGEN); or MonsterScript reverse transcriptase (EPICENTRE)); and in aspects of this embodiment in which the nucleic acid polymerase that extends the 3'-termini of the mRNA molecules using the sequence that is complementary to the sequence tag template of the tagging oligonucleotide as a template is different from the RNA-dependent DNA polymerase used to obtain cDNA, the kit additionally comprises said nucleic acid polymerase for extending the 3'-termini of the mRNA molecules.

In some preferred embodiments, the present invention provides methods and kits for producing cDNA copies of full-length prokaryotic mRNA (e.g., for making full-length cDNA libraries). For example, the present invention provides a method for preparing prokaryotic cDNA, comprising the steps of a) providing a sample comprising RNA from a prokaryotic organism; a polynucleotide kinase; and a 5'-phosphate-dependent nucleic acid exonuclease; b) contacting the sample with the polynucleotide kinase in the presence of ATP and under conditions such that nucleic acid molecules having an hydroxyl group at their 5'-termini are phosphorylated; c) contacting the sample treated with the polynucleotide kinase with the 5'-phosphate-dependent nucleic acid exonuclease under conditions such that undesired nucleic acids in the sample are digested but mRNA molecules in the sample having a 5'-triphosphate are not digested; and d) producing cDNA copies of the mRNA. The present invention is not limited by the manner in which the cDNA copies are produced.

In some preferred embodiments, the method of making cDNA copies comprises: i) contacting the mRNA molecules with a poly(A) polymerase to produce mRNA molecules having a 3'-poly(A) tail; and ii) contacting the mRNA molecules having a 3'-poly(A) tail with a primer that anneals to said tail and extending the primer with an RNA-dependent DNA polymerase under conditions whereby cDNA is obtained. Kits for conducting such methods may have any one or more reagents (e.g., oligo(dT) primer and/or other primers), control samples, ATP, buffers, etc. useful in the method. For example, preferred kits comprise: a) a 5'-phosphate-dependent nucleic acid exonuclease; b) a polynucleotide kinase (e.g., T4 polynucleotide kinase); c) a poly(A) polymerase; and d) an RNA-dependent DNA polymerase. In some preferred embodiments, the desired

IDT Exhibit 1017.1951

nucleic acid molecules comprise both prokaryotic and eukaryotic mRNA. In some preferred kits in which the RNA in the sample is of sufficient quality for the intended purpose, meaning that a sufficient amount of the RNA comprises the desired mRNA with a 5'-triphosphate or 5'-cap for the intended purpose (which amount can vary for different purposes), the step comprising contacting the sample with the polynucleotide kinase in the presence of ATP and under conditions such that nucleic acid molecules having an hydroxyl group at their 5'-termini are phosphorylated can be omitted and a polynucleotide kinase is not needed in a preferred kit for this method, or if it is included in the kit, the polynucleotide kinase step can be omitted. The invention also envisions that this step can be omitted when a sample is rare or limited in amount and it is desirable to obtain the maximum amount of mRNA and/or cDNA, including degraded portions of mRNA and/or cDNA molecules, for the intended purposes.

Similar methods and kits are provided that are useful for amplifying prokaryotic mRNA. For example, the present invention provides a method for amplifying prokaryotic mRNA, comprising the steps of a) providing a sample comprising RNA from a prokaryotic organism; and a 5'-phosphate-dependent nucleic acid exonuclease; b) contacting the sample with the 5'-phosphate-dependent nucleic acid exonuclease under conditions such that undesired nucleic acids in the sample are substantially digested but mRNA molecules in the sample having a 5'-triphosphate are not digested; and c) amplifying the mRNA. Similar methods and kits are provided that are useful for amplifying prokaryotic mRNA, whereby the methods and kits result in enrichment of sequences corresponding to the 5' portions of mRNA. For example, the present invention provides a method for amplifying prokaryotic mRNA, comprising the steps of a) providing a sample comprising RNA from a prokaryotic organism; a polynucleotide kinase; and a 5'-phosphate-dependent nucleic acid exonuclease; b) contacting the sample with the polynucleotide kinase in the presence of ATP and under conditions such that nucleic acid molecules having an hydroxyl group at their 5'-termini are phosphorylated; c) contacting the sample with the 5'-phosphate-dependent nucleic acid exonuclease under conditions such that undesired nucleic acids in the sample are substantially digested but mRNA molecules in the sample having a 5'-triphosphate are not digested; and d) amplifying the mRNA. The present invention is not limited by the manner in which the

IDT Exhibit 1017.1952

mRNA is amplified. In some embodiments, the mRNA amplified by a method comprising: i) contacting said mRNA molecules with a poly(A) polymerase to produce mRNA molecules having a 3'-poly(A) tail; ii) contacting said mRNA molecules having a 3'-poly(A) tail with a primer that anneals to said tail and extending said primer with an RNA-dependent DNA polymerase under conditions whereby cDNA that is joined to an RNA polymerase promoter sequence is obtained (using any number of methods known in the art; methods are known for joining the promoter to either the 5'-end of the cDNA or the 3'-end of the cDNA, resulting in synthesis of amplified antisense or sense RNA, respectively, by subsequent *in vitro* transcription); iii) producing double stranded cDNA; and iv) contacting said double-stranded cDNA comprising said promoter sequence with an RNA polymerase that binds to said promoter and synthesizes RNA therefrom under conditions whereby RNA corresponding to mRNA in the biological sample is synthesized. Some other embodiments of the present invention use a single-stranded first-strand cDNA template that is functionally joined at either the 3'-end or the 5'-end to a single-stranded promoter that is recognized by an RNA polymerase, such as MiniV RNA polymerase, that can synthesize RNA *in vitro* therefrom; in these embodiments, the method does not comprise a step for synthesis of double-stranded cDNA. Kits for conducting such methods may have any one or more reagents (e.g., primers), control samples, buffers, etc. useful in the method. For example, preferred kits comprise: a) a 5'-phosphate-dependent nucleic acid exonuclease; c) an RNA-dependent DNA polymerase and c) an RNA polymerase (e.g., T7, T3, and SP6 RNA polymerases). Other kits that use single-stranded templates and promoters use an RNA polymerase that synthesize RNA *in vitro* therefrom, such as MiniV RNA polymerase. Other preferred kits further comprise: a) a polynucleotide kinase, such as but not limited to T4 polynucleotide kinase, and ATP; and b) a poly(A) polymerase, such as but not limited to a poly(A) polymerase encoded by an *E. coli pcn*B gene.

In other preferred embodiments, the present invention provides a method for enriching for or isolating mRNA from eukaryotes, comprising the steps of a) providing a 5'-phosphate-dependent nucleic acid exonuclease; and a biological sample comprising a mixture of nucleic acids, said mixture of nucleic acids comprising mRNA of interest and an undesired nucleic acid having a 5'-phosphate; and b) contacting the 5'-phosphate-

IDT Exhibit 1017.1953

dependent nucleic acid exonuclease to the biological sample under conditions such that the exonuclease substantially digests the undesired nucleic acid so as to enrich said sample for the mRNA of interest. Thus, the present invention provides alternative kits and methods for isolating eukaryotic mRNA that do not require binding to oligo(dT) cellulose or other oligo(dT) or oligo(dU) resins or membranes. The method is particularly suitable to high-throughput methods—much more so than methods that employ binding to a resin. However, it will also be understood, based on the description of the present invention, that a 5'-phosphate-dependent exonuclease can be used to obtain a mixture of both prokaryotic and eukaryotic mRNA, and that the eukaryotic mRNA can be separated from prokaryotic mRNA in a mixture by binding polyadenylated eukaryotic mRNA to an oligo(dT) resin, as is known in the art. Still further, although most eukaryotic mRNA is believe to be polyadenylated, some eukaryotic mRNA transcripts may not have a poly(A) tail. The methods disclosed herein can be used to study eukaryotic transcripts that lack a poly(A) tail, which transcripts would not have been detected in many studies that have used oligo(dT) resins and membranes for isolation of the eukaryotic mRNA.

In some preferred embodiments, the present invention provides a method for isolating eukaryotic mRNA comprising the steps of: a) providing a sample from a eukaryotic organism comprising a mixture of nucleic acids, said mixture of nucleic acids comprising desired mRNA of interest and at least one undesired nucleic acid; and a 5'-phosphate-dependent nucleic acid exonuclease; b) treating the sample with the 5'-phosphate-dependent nucleic acid exonuclease under conditions such that the undesired nucleic acid is substantially digested and mRNA having a 5'-cap or a 5'-triphosphate group is not digested; and c) recovering the mRNA (e.g., for additional analysis or use). In some embodiments, the 5'-phosphate-dependent nucleic acid exonuclease used in the method is provided in a kit with buffers, instructions, and optionally, with appropriate control samples, and the like.

One problem in the art that complicates analyses involving mRNA, including analyses involving eukaryotic mRNA, relates to the fact that samples containing the mRNA can include mRNA molecules that are not full-length because they have been degraded: 1) by ribonucleases in the sample, which ribonucleases can be derived either

IDT Exhibit 1017.1954

from the cells from which the mRNA is derived or from contamination of the sample, such as by contamination with a ribonuclease A-type nuclease from human skin or other sources; or 2) by physical forces, such as by shearing forces or by forces due to contact of the mRNA with heat and/or metal ions, such as but not limited to Mn 2+ cations. Many common ribonucleases, such as but not limited to RNase A-type ribonucleases (e.g., including those on human skin) and *E. coli* RNase I, are endoribonucleases that digest RNA to yield products having 5'-hydroxyl and 3'-phosphate groups. Without being bound by theory, it is believed that physical forces can yield a mixture of products having either a 5'-hydroxyl or a 5'-phosphate group. In any case, it is likely that many, if not most, of the degraded mRNA molecules in a sample will have a 5'-hydroxyl group, which means that they will be resistant to digestion with a 5'-exoribonuclease of the present invention. However, since most 5'-exoribonuclease-resistant eukaryotic mRNA molecules will have a 3'-poly(A) tail, these degraded molecules will be substrates for synthesis of oligo(dT)-primed cDNA synthesis and will also be substrates for many antisense RNA amplification methods or sense RNA amplification methods. Thus, these degraded, but 5'-exonuclease-resistant, mRNA molecules can result in a high representation of mRNA sequences corresponding to the 3'-end of the mRNA relative to mRNA sequences corresponding to the 5'-end of the mRNA, which can complicate interpretations of gene expression analyses, and the like. One embodiment of the present invention is a method for assessing the degree of RNA degradation, the method comprising: a) providing a first sample comprising a known quantity of RNA of unknown quality (i.e., of unknown degree of fragmentation or degradation) from any organism or mixture of organisms, the sample containing mRNA; a second control sample comprising a known quantity of substantially undegraded mRNA, preferably, but not necessarily, from the same organism or mixture of organisms; a polynucleotide kinase; and a 5'-phosphate-dependent exonuclease; b) treating one portion of each sample with the polynucleotide kinase in the presence of ATP and under conditions such that nucleic acid molecules having an hydroxyl group at their 5'-termini are phosphorylated; c) contacting the portion of each sample treated with the polynucleotide kinase and a second portion (of the same starting RNA quantity) of each sample that was not treated with the polynucleotide kinase with the 5'-phosphate-dependent nucleic acid exonuclease under conditions such that

IDT Exhibit 1017.1955

undesired nucleic acids in the sample having a 5'-phosphate are substantially digested but mRNA molecules in the sample having a 5'-triphosphate are not digested; and d) producing labeled cDNA copies or labeled amplified RNA of the mRNA in each portion of each sample using an oligo(dT)-containing primer; e), obtaining a ratio for each sample, the ratio having a numerator comprising the amount of labeled cDNA (or amplified RNA) obtained from the polynucleotide kinase-untreated portion and a denominator comprising the amount of labeled cDNA (or amplified RNA) obtained from the polynucleotide kinase-treated portion; and f) comparing the ratio obtained for the first sample with the ratio obtained for the second sample, whereby a higher ratio for the first sample than the ratio for the second sample indicates a higher relative degree of degradation of the RNA in the first sample compared to the second sample. This embodiment of the invention finds utility as an independent method of the invention, or more preferably, it is used as part of a method of the invention for obtaining cDNA or amplified sense or antisense RNA.

As indicated from the above discussion, it is usually desirable to increase the proportion of cDNA molecules and/or amplified sense or antisense RNA molecules that contain sequences corresponding to the 5'-end of the mRNA molecules in the sample (i.e., by removing degraded 3' mRNA fragments) so that a higher percentage of the cDNA and/or the amplified sense or antisense RNA is made from full-length mRNA molecules. Thus, the present invention also provides a method to obtain mRNA from which the degraded mRNA molecules with a 5'-hydroxyl group have been removed, which can improve various analyses involving mRNA, including analyses that involve eukarytic mRNA, prokaryotic mRNA, or both eukaryotic and prokaryotic mRNA. One method of the present invention uses a polynucleotide kinase, such as but not limited to T4 polynucleotide kinase, in order to phosphorylate the 5'-hydroxyl groups of degraded mRNA molecules, thereby making them substrates for digestion by a 5'-exonuclease. Following digestion of the 5'-phosphorylated mRNA with a 5'-exonuclease of the invention, the remaining undigested mRNA molecules will be enriched for full-length mRNA, which will in turn result in cDNA molecules (from cDNA synthesis reactions using an oligo(dT)-containing primer) and amplified sense or antisense RNA molecules (from RNA amplification reactions) that have better representations of sequences

IDT Exhibit 1017.1956

corresponding to the 5'- and 3'-ends of mRNA in the sample. Thus, in some preferred embodiments, the present invention provides a method for isolating desired mRNA of interest comprising the steps of: a) providing a sample containing RNA from an organism, said sample comprising mRNA and at least one undesired nucleic acid; a polynucleotide kinase; and a 5'-phosphate-dependent nucleic acid exonuclease; b) treating the sample with the polynucleotide kinase in the presence of ATP and under conditions such that nucleic acid molecules having an hydroxyl group at their 5'-termini are phosphorylated; c) treating the sample with the 5'-phosphate-dependent nucleic acid exonuclease under conditions such that RNA molecules having a 5'-monophosphate group are substantially digested; and d) recovering the mRNA (e.g., for additional analysis or use). In some embodiments, the 5'-phosphate-dependent nucleic acid exonuclease used in the method is provided in a kit with polynucleotide kinase (e.g., T4 polynucleotide kinase), ATP, buffers, instructions, and optionally, with appropriate control samples, and the like.

In some embodiments, the present invention provides methods and kits for producing cDNA copies of eukaryotic mRNA (e.g., for making cDNA libraries). For example, the present invention provides a method for preparing eukaryotic cDNA, comprising the steps of a) providing a sample comprising RNA from a eukaryotic organism; and a 5'-phosphate-dependent nucleic acid exonuclease; b) contacting the sample with the 5'-phosphate-dependent nucleic acid exonuclease under conditions such that undesired nucleic acids in the sample are substantially digested but mRNA molecules in the sample having a 5'-triphosphate or a 5'-cap are not digested; and c) producing cDNA copies of the mRNA. In some preferred embodiments, the present invention provides methods and kits for producing cDNA copies of full-length eukaryotic mRNA (e.g., for making full-length cDNA libraries). For example, the present invention provides a method for preparing eukaryotic cDNA copies of full-length eukaryotic mRNA, comprising the steps of a) providing a sample comprising RNA from a eukaryotic organism; a polynucleotide kinase; and a 5'-phosphate-dependent nucleic acid exonuclease; b) contacting the sample with the polynucleotide kinase in the presence of ATP and under conditions such that nucleic acid molecules having an hydroxyl group at their 5'-termini are phosphorylated; c) contacting the sample treated with the

IDT Exhibit 1017.1957

polynucleotide kinase with the 5'-phosphate-dependent nucleic acid exonuclease under conditions such that undesired nucleic acids in the sample are substantially digested but desired mRNA molecules in the sample having a 5'-cap or a 5'-triphosphate are not digested; and d) producing cDNA copies of the mRNA. The present invention is not limited by the manner in which the cDNA copies are produced. In some preferred embodiments, the method of making cDNA copies comprises contacting the mRNA molecules with a primer having an oligo(dT) sequence in its 3'-end portion under conditions such that the primer anneals with the mRNA, and extending the primer with an RNA-dependent DNA polymerase under conditions whereby cDNA is obtained. In preferred embodiments, an RNA-dependent DNA polymerase enzyme and suitable reaction conditions are used so that the amount of full-length cDNA is maximized. Suitable enzymes and reaction conditions for this purpose are known in the art. For example, but without limitation, U.S. Patent Nos. 5,962,271; and 5,962,272 describe one method to obtain increased amounts of full-length cDNA, which can be used together with the methods of the present invention. However, the invention is not limited with respect to enzymes or reaction conditions, and any RNA-dependent DNA polymerase and reaction conditions that accomplish the intended purpose can be used in a method or kit of the present invention. Kits for conducting such methods of the present invention may have any one or more reagents (e.g., an oligo(dT)-containing primer and/or other primers), control samples, ATP, buffers, etc. useful in the method. For example, preferred kits comprise: a) a 5'-phosphate-dependent nucleic acid exonuclease; b) a polynucleotide kinase (e.g., T4 polynucleotide kinase); and c) an RNA-dependent DNA polymerase. In some preferred kits in which the mRNA in the sample is known to be of sufficient quality for the intended purpose, meaning that approximately <90%, <75%, <50%, <40%, <30%, <20%, <15%, <10%, <5%, <4%, <3%, <2%, or <1% of the mRNA transcripts are degraded into fragments that lack a 5'-cap or a 5'-triphosphate group, the step comprising contacting the sample with the polynucleotide kinase in the presence of ATP and under conditions such that nucleic acid molecules having an hydroxyl group at their 5'-termini are phosphorylated can be omitted and a polynucleotide kinase is not needed in a preferred kit, or if included, its use can be omitted.

IDT Exhibit 1017.1958

Similar methods and kits are provided that are useful for amplifying eukaryotic mRNA. For example, the present invention provides a method for amplifying eukaryotic mRNA, comprising the steps of a) providing a sample comprising RNA from a eukaryotic organism; and a 5'-phosphate-dependent nucleic acid exonuclease; b) contacting the sample with the 5'-phosphate-dependent nucleic acid exonuclease under conditions such that undesired nucleic acids in the sample are substantially digested but mRNA molecules in the sample having a 5'-triphosphate or a 5'-cap are not digested; and c) amplifying the mRNA. In preferred embodiments, the present invention provides a method for amplifying eukaryotic mRNA, comprising the steps of a) providing a sample comprising RNA from a eukaryotic organism; a polynucleotide kinase; and a 5'-phosphate-dependent nucleic acid exonuclease; b) contacting the sample with the polynucleotide kinase in the presence of ATP and under conditions such that nucleic acid molecules having an hydroxyl group at their 5'-termini are phosphorylated; c) contacting the sample with the 5'-phosphate-dependent nucleic acid exonuclease under conditions such that undesired nucleic acids in the sample are substantially digested but mRNA molecules in the sample having a 5'-triphosphate or a 5'-cap are not digested; and d) amplifying the mRNA. The present invention is not limited by the manner in which the mRNA is amplified. In some embodiments, the mRNA is amplified by a method comprising: i) contacting said mRNA molecules with an oligo(dT)-containing primer that anneals to a poly(A) tail of said mRNA and extending said primer with an RNA-dependent DNA polymerase under conditions whereby cDNA that is joined to an RNA polymerase promoter sequence is obtained (using any number of methods known in the art; methods are known for joining the promoter to either the 5'-end of the cDNA or the 3'-end of the cDNA, resulting in synthesis of amplified antisense or sense RNA, respectively, by subsequent *in vitro* transcription); ii) producing double stranded cDNA; and iii) contacting said double-stranded cDNA comprising said promoter sequence with an RNA polymerase that binds to said promoter and synthesizes RNA therefrom under conditions whereby RNA corresponding to mRNA in the biological sample is synthesized. Some other embodiments of the present invention use a single-stranded first-strand cDNA template that is functionally joined at either the 3'-end or the 5'-end to a single-stranded promoter that is recognized by an RNA polymerase, such as MiniV

IDT Exhibit 1017.1959

RNA polymerase, that can synthesize RNA *in vitro* therefrom; in these embodiments, the method does not comprise a step for synthesis of double-stranded cDNA. Still further, the invention also comprises methods for amplifying mRNA enriched or isolated using a 5'-phosphate-dependent exonuclease of the invention using methods, such as but not limited to, methods described in U.S. Pat. Appln. No. 2004/0180361 of Dahl et al.; PCT Patent Publication Nos. WO 02/16639; WO 00/56877; and AU 00/29742 of Takara Shuzo Company; and in U.S. Pat. No. 6,251,639 and U.S. Pat. Appln. Nos. 2001/0034048; 2003/0017591; 2003/0087251; and 2003/0186234 of Kurn (all incorporated herein by reference), even if said methods comprise synthesis of DNA products rather than RNA products corresponding to mRNA in the sample. Thus, another embodiment of the invention comprises a method for amplifying mRNA enriched or isolated using a 5'-phosphate-dependent exonuclease, the method comprising: i) contacting said mRNA molecules with a first primer that anneals to said mRNA and extending said first primer with an RNA-dependent DNA polymerase under conditions whereby first-strand cDNA is obtained; ii) contacting said first-strand cDNA with a second primer and extending said second primer with a strand-displacing DNA-dependent DNA polymerase under conditions whereby double-stranded cDNA that is joined to a ribonucleotide-containing tail sequence at the 3'-end of at least one cDNA strand is obtained; and iii) contacting said double-stranded cDNA with a ribonuclease H and a ribonucleotide-containing primer that anneals to the tail sequence and extending said ribonucleotide-containing primer with the strand-displacing DNA-dependent DNA polymerase under conditions whereby at least one strand of said cDNA is amplified. Kits for conducting such methods may have any one or more reagents (e.g., primers), control samples, buffers, etc. useful in the method. For example, preferred kits comprise: a) a 5'-phosphate-dependent nucleic acid exonuclease; b); c) an RNA-dependent DNA polymerase; and c) an RNA polymerase (e.g., T7, T3, or SP6 RNA polymerases). Other kits that use single-stranded templates and promoters use an RNA polymerase that synthesize RNA *in vitro* therefrom, such as MiniV RNA polymerase. Other preferred kits further comprise: a) a polynucleotide kinase, such as but not limited to T4 polynucleotide kinase, and ATP. Still other kits also include a ribonucleotide-containing primer, a ribonuclease H, and a strand-displacing DNA-dependent DNA polymerase, such as but not limited to phi29 DNA

IDT Exhibit 1017.1960

polymerase or rBST DNA polymerase large fragment (both available from EPICENTRE).

In some cases in which a sample cannot be replaced (e.g, a forensics sample from a crime scene), it may not be desirable to remove mRNA fragments that do not have a 5'-triphosphate or a 5'-cap prior to synthesis of cDNA and/or RNA amplification. Thus, some embodiments of the present invention comprise a method for preparing prokaryotic mRNA, eukaryotic mRNA or a mixture of both prokaryotic and eukaryotic mRNA that is degraded for synthesis of cDNA and/or RNA amplification, the method comprising: a) providing a sample comprising degraded RNA from a prokaryotic and/or eukaryotic organism; a poly(A) polymerase; and a 5'-phosphate-dependent nucleic acid exonuclease; b) contacting the sample with the 5' exonuclease under conditions such that undesired nucleic acids in the sample are substantially digested but desired RNA molecules are not digested; and c) contacting the sample with a poly(A) polymerase in the presence of ATP and under conditions in which a 3'-poly(A) tail is added to the desired RNA molecules. The invention further comprises a method for preparing cDNA from said 3'-polyadenylated RNA, the method comprising: a) providing a sample comprising 3'-polyadenylated RNA; an RNA-dependent DNA polymerase; and an oligo(dT)-containing primer that anneals to said tail; and b) contacting the 3'-polyadenylated RNA with the oligo(dT)-containing primer and extending the primer with the RNA-dependent DNA polymerase under conditions whereby cDNA is obtained. Other embodiments further comprise methods for amplifying the polyadenylated RNA, the method comprising a) contacting the polyadenylated RNA with a primer that anneals to said tail and extending said primer with an RNA-dependent DNA polymerase under conditions whereby cDNA that is joined to an RNA polymerase promoter sequence is obtained (using any number of methods known in the art; methods are known for joining the promoter to either the 5'-end of the cDNA or the 3'-end of the cDNA, resulting in synthesis of amplified antisense or sense RNA, respectively, by subsequent *in vitro* transcription); b) producing double stranded cDNA; and c) contacting said double-stranded cDNA comprising said promoter sequence with an RNA polymerase that binds to said promoter and synthesizes RNA therefrom under conditions whereby amplified RNA corresponding to polyadenylated RNA in the sample is synthesized. Kits for conducting such methods may have any one

IDT Exhibit 1017.1961

or more reagents (e.g., primers, such as an oligo(dT) primer, or other oligonucleotides), control samples, buffers, etc. useful in the method. For example, preferred kits comprise: a) a 5'-phosphate-dependent nucleic acid exonuclease; and b) a poly(A) polymerase. Other preferred kits for preparing cDNA additionally comprise an RNA-dependent DNA polymerase. Still other preferred kits for methods comprising sense or antisense RNA amplification comprise, in addition to an RNA-dependent DNA polymerase, an RNA polymerase (e.g., without limitation, T7, T3, or SP6 RNA polymerase).

The present invention also provides a method for using a 5'-phosphate-dependent exonuclease of the invention to obtain 5'-ends of degraded mRNA and other primary transcripts that have a 5'-cap or a 5'-triphosphate group.

For example, one method that can be used to isolate the 5'-ends of degraded mRNA and other primary transcripts that have a 5'-cap or a 5'-triphosphate group comprises the steps of:

a) obtaining a sample of total RNA comprising degraded RNA;

b) treating the degraded RNA with a polynucleotide kinase (e.g., T4 polynucleotide kinase) under conditions that result in phosphorylation of 5'-hydroxyl groups of said degraded RNA;

c) treating the RNA from step (b) with a 5'-phosphate-dependent exonuclease of the invention under conditions that result in digestion of RNA having a 5'-phosphate group; and

d) isolating the RNA that is not digested in step (c).

Degraded RNA, such as but not limited to RNA degraded by RNase A-type nucleases, can have 5'-hydroxylated termini. The polynucleotide kinase phosphorylates the 5'-ends of degrade RNA molecules with a hydroxyl group, making these molecules subtrates which can be digested by the 5'-to-3' exoribonuclease. The RNA obtained using this procedure will comprise 5'-end fragments of RNA that have a 5'-cap or a 5'-triphosphate. Without limiting the invention, if sufficient quantities of these fragments are available, the RNA fragments can be directly labeled [e.g., using ULS™ (Kreatech) or other chemical dye labeling compounds] and used for gene expression studies with arrays or microarrays of oligonucleotide sequences that are complementary to the 5'-ends of all known sequences or desired known sequences corresponding to mRNA transcripts

IDT Exhibit 1017.1962

for the organism from which the total RNA was obtained. If sufficient quantities of the RNA fragments are not available from this method (e.g., if the RNA fragments are obtained from only a single cell or a small number of cells), then the 5'-end fragments of RNA having a 5'-cap or a 5'-triphosphate obtained can be amplified by one of several methods. By way of example but not of limitation, the 5'-end fragments can be polyadenylated using poly(A) polymerase using methods known in the art, and then amplified using an Eberwine-type RNA amplification method (e.g., see Van Gelder et al., Proc. Nat. Acad. Sci. USA 87: 1663, 1990), which uses an oligo(dT) promoter primer and generates anti-sense RNA. A 1-round or 2-round TargetAmp™ aRNA or aminoallyl-aRNA Amplification Kit (EPICENTRE Technologies, Madison, WI, USA) or another kit can also be used for this purpose. If the amplified and labeled aRNA is analyzed using a microarray, the oligos on the microarray must be complementary to the aRNA corresponding to the 5'-ends of the mRNA fragments rather than to the sense mRNA sequence. Alternatively, a sense mRNA amplification method, such as available in kits from BD-Clontech and from Genisphere can be use to amplify sense RNA corresponding to the 5'-ends of the mRNA fragments. A method comprising an oligonucleotide sequence tag template can be used to add a sequence tag to the 3'-end of a nucleic acid, including an RNA or DNA molecule, which can then be amplified following addition of a promoter, as described in U.S. Patent Application No. 2005/015333, which is incorporated herein by reference. After labeling, the amplified sense RNA can be annealed to a microarray of antisense oligos complementary to all known sequences or desired known sequences corresponding to the 5'-ends of mRNA transcripts for the organism from which the total RNA was obtained. Still further, the 5'-ends of the mRNA fragments can be amplified without polyadenylation by using a promoter primer with a random sequence for priming first-strand cDNA synthesis (e.g., see Ziman and Davis, WO 2002/044399). Still other methods for amplifying the mRNA 5'-end fragments are described herein which can be used as part of the present invention.

In other embodiments of the invention, a 5'-phosphate-dependent exonuclease is used to obtain 5'-end fragments from intact, undegraded mRNA or other primary transcripts having a 5'-cap or a 5'-triphosphate in a sample which contains other undesired nucleic acids. One way that the total RNA can be analyzed to determine that it is not

IDT Exhibit 1017.1963

significantly degraded is to determine the profile of large rRNA in the sample using an Agilent bioanalyzer and to verify that the scan of the rRNA indicates that the large rRNA is largely intact. If desired, small RNA, such as 5S rRNA and tRNA, can be substantially removed from the total RNA by precipitation of other RNA in the sample using LiCl and ethanol, leaving 5S rRNA and tRNA in solution, prior to performing additional steps to isolate 5'-end fragments of mRNA. Thus, one embodiment is a method for isolating the 5'-ends of mRNA and other primary transcripts that have a 5'-cap or a 5'-triphosphate group (i.e., the nucleic acid of interest) from a sample which also contains undesired nucleic acid (e.g., large rRNA in a sample of total RNA), wherein the nucleic acid of interest is not substantially degraded, the method comprising:

a) obtaining a sample of total RNA that contains mRNA;

b) optionally, obtaining an RNA fraction by that is reduced in 5S rRNA and tRNA LiCl and ethanol precipitation;

c) treating the RNA so as to bring about controlled degradation of the RNA into smaller fragments by physical or chemical means or by degradation with a nuclease;

d) treating the degraded RNA from step (c) with a polynucleotide kinase (e.g., T4 polynucleotide kinase) under conditions that result in phosphorylation of 5'-hydroxyl groups of said degraded RNA;

e) treating the RNA from step (d) with a 5'-to-3' exoribonuclease under conditions that result in digestion of RNA having a 5'-phosphate group; and

f) isolating the RNA that is not digested in step (e).

This procedure will give a solution of fragments of 5'-capped or 5'-triphosphorylated mRNA of the desired size range. These mRNA 5'-end fragments can then be used for any desired application, including for the applications and methods described with respect to the 5'-end fragments of mRNA obtained from FFPE degraded RNA discussed herein. For example, but without limitation, the fragments can be polyadenylated using poly(A) polymerase and amplified using an RNA amplification method that yields labeled aRNA or sense RNA as described herein. In other embodiments of the invention, the 5'-end fragments of mRNA can be amplified and analyzed using rolling circle replication or rolling circle transcription as also described herein.

IDT Exhibit 1017.1964

Some methods that can be used for treating the RNA so as to bring about controlled degradation of the RNA into smaller fragments include chemical treatments, such exposing the RNA to cations, including, but not limited to Mg^{2+} under elevated temperatures, and controlled degradation with a ribonuclease (after optimizing the enzyme, enzyme amount, time and other reactions conditions that yield RNA fragments in the desired size range). One RNase that can be used to fragment the RNA is RNase I, which cleaves the RNA after every base and which can be inactivated by heat after the treatment. Still another embodiment of a method of the invention for controlled degradation of RNA with a nuclease comprises treating the RNA with an RNase H in the presence of one or more DNA oligonucleotides complementary to specific sequences of the RNA of interest, under conditions wherein the DNA oligonucleotides can hybridize to the complementary mRNA and the RNase H is active.

Still another embodiment of the invention is a method for method for amplifying circularized cDNA, which cDNA is prepared from 5'-end fragments of 5'-capped or 5'-triphosphorylated RNA, by rolling circle replication. Thus, one embodiment of the invention is a method for amplifying the 5'-end fragments of mRNA and other primary transcripts, the method comprising the steps of:

a) providing a sample comprising 5'-end fragments of mRNA or other primary transcripts of a desired size range (e.g., without limitation, $\leq$ 100 bases);

b) contacting the 5'-end fragments with poly(A) polymerase under conditions that result in addition of a poly(A) tail of sufficient length to permit annealing of a primer for reverse transcription;

c) contacting the poly(A)-tailed 5'-end fragments with a 5'-phosphorylated primer that anneals to the poly(A) tail and an RNA-dependent DNA polymerase or reverse transcriptase under conditions that result in extension of the primer and synthesis of first-strand cDNA that is complementary to the poly(A)-tailed 5'-end fragments;

d) treating the complex between the first-strand cDNA and the poly(A)-tailed 5'-end fragments of mRNA from step (c) so as to remove the RNA from the first-strand cDNA;

e) contacting the first-strand cDNA from step (d) with a ligase that can ligate single-stranded DNA (ssDNA) in the absence of a ligation splint oligo under conditions that

IDT Exhibit 1017.1965

result in intramolecular ligation (i.e., circularization) of the linear first-strand cDNA so as to obtain circular first-strand cDNA; and

f) contacting the circular first-strand cDNA obtained from step (e) with a strand-displacing DNA polymerase (e.g., without limitation, phi29 DNA polymerase or rBst DNA polymerase large fragment) and a primer that anneals to the circular first-strand cDNA under conditions that result in synthesis of concatemeric second-strand cDNA that is complementary to the circular first-strand cDNA (e.g., by rolling circle replication).

The primer used to prime first-strand cDNA synthesis in step (c) can be an oligo$(dT)_n$ or an oligo$(dT)_nX$ anchored primer, wherein X is a mixture of dAMP, dCMP and dGMP. In step (c) above, the primer has a 5'-phosphate. In other embodiments, the primer has a 5'-hydroxyl group and is phosphorylated using a polynucleotide kinase (e.g., T4 polynucleotide kinase) under kinasing conditions after step (d) and prior to step (e). An RNA-dependent DNA polymerase or reverse transcriptase that does not have RNase H activity, such as but not limited to, an enzyme is selected from among an RNase H-minus mutant of MMLV or AMV reverse transcriptase (e.g., SuperScript™ I, II, or III from Invitrogen) can be used and is preferred in step (c), but an RNA-dependent DNA polymerase that has RNase H activity can be used in some embodiments. Any ligase that ligates ssDNA in the absence of a ligation splint oligo can be used, such as but not limited to, a ligase selected from the group consisting of a ligase derived from phage TS2126 that infect Thermus scotoductus, CircLigase™ ssDNA ligase (EPICENTRE), and ThermoPhage™ ssDNA Ligase (PROKARIA, Reykjavik, Iceland). In some embodiments, the primer used for rolling circle replication in step (f) is an oligo(dA) primer that anneals to the oligo(dT) sequence from the primer that was used to prime synthesis of first-strand cDNA. In other embodiments, a multiplicity of primers is used for rolling circle replication. For example, in one embodiment, a solution comprising all possible combinations of short primers such as hexamers (i.e., the hexamer is synthesized to contain any of the four canonical nucleotide bases at each of the six nucleotide positions) is used. In some aspects of this embodiment, the concatemeric rolling circle replication product is labeled by incorporation of a labeled nucleotide, but any other suitable method known in the art can also be used to detect the rolling circle replication product. In another embodiment in which a multiplicity of primers is used for rolling

IDT Exhibit 1017.1966

circle replication, the primers comprise oligonucleotides that have a sequence that is the same as at least one sequence within a certain defined distance of the 5'-end of the known mRNA transcripts for the organism for which transcription is analyzed. The defined distance from the 5'-end of the mRNA transcripts within which the complementary sequence for the priming oligos is chosen is similar to the average length of the 5'-end fragments provided in step (a). In a preferred aspect of this embodiment, the multiplicity of primers having the same sequence as at least one sequence within a certain defined distance of the 5'-end of the known mRNA transcripts for the organism for which transcription is analyzed are attached to a surface (e.g., in an array or microarray) so that their 3'-ends have an hydroxyl group and the sequences are at sufficient distance from the surface so that the sequences can prime rolling circle replication by primer extension of the tethered primer using a circular first-strand cDNA from step (e), to which the primer is complementary, as a template. In this way, the primer that is attached to the surface is primer extended by rolling circle replication using a strand-displacing DNA polymerase under polymerization conditions. In some aspects of this embodiment, the concatemeric rolling circle replication product is labeled by incorporation of a labeled nucleotide, but any other suitable method known in the art can also be used to detect the rolling circle replication product.

Yet another method that can be used to amplify 5'-end fragments of mRNA and other primary transcripts comprises the steps of:

a) providing a sample comprising 5'-end fragments of mRNA and other primary transcripts of a desired size range (e.g., without limitation, $\leq$ 100 bases);

b) contacting the 5'-end fragments with a 5'-phosphorylated random hexamer primer and an RNA-dependent DNA polymerase or reverse transcriptase under conditions that result in extension of the primer and synthesis of first-strand cDNA that is complementary to the 5'-end fragments;

c) treating the complex between the first-strand cDNA and the 5'-end fragments of mRNA from step (b) so as to remove the RNA from the first-strand cDNA;

d) contacting the first-strand cDNA from step (c) with a ligase that can ligate single-stranded DNA (ssDNA) in the absence of a ligation splint oligo under conditions that

IDT Exhibit 1017.1967

result in intramolecular ligation (i.e., circularization) of the linear first-strand cDNA so as to obtain circular first-strand cDNA; and

e) contacting the circular first-strand cDNA obtained from step (d) with a strand-displacing DNA polymerase (e.g., without limitation, phi29 DNA polymerase or rBst DNA polymerase large fragment) and a primer that anneals to the circular first-strand cDNA under conditions that result in synthesis of concatemeric second-strand cDNA that is complementary to the circular first-strand cDNA (e.g., by rolling circle replication).

In preferred embodiments this method, the RNA-dependent DNA polymerase or reverse transcriptase is an enzyme that does not have RNase H activity. By way of example, but not of limitation, the enzyme is selected from among an RNase H-minus mutant of MMLV or AMV reverse transcriptase (e.g., SuperScript™ I, II, or III from Invitrogen). A ligase that ligates ssDNA in the absence of a ligation splint oligo includes, but is not limited to a ligase selected from the group consisting of a ligase derived from phage TS2126 that infect Thermus scotoductus, CircLigase™ ssDNA ligase (EPICENTRE), and ThermoPhage™ ssDNA Ligase (PROKARIA, Reykjavik, Iceland). In some embodiments, a multiplicity of primers is used for step (e). For example, in one embodiment, a solution comprising all possible hexamer primers (i.e., the hexamer is synthesized to contain any of the four canonical nucleotide bases at each of the six positions). In some aspects of this embodiment, the concatemeric rolling circle replication product is labeled by incorporation of a labeled nucleotide, but any other suitable method known in the art can also be used to detect the rolling circle replication product. In other embodiments in which a multiplicity of primers is used for rolling circle replication, the primers comprise oligonucleotides that have a sequence that is the same as at least one sequence within a certain defined distance of the 5'-end of the known mRNA transcripts for the organism for which transcription is analyzed. The defined distance from the 5'-ends of the mRNA transcripts is chosen to be similar to the average length of the 5'-end fragments provided in step (a). In a preferred aspect of this embodiment, the multiplicity of primers having the same sequence as at least one sequence within a certain defined distance of the 5'-end of the known mRNA transcripts for the organism for which transcription is analyzed are attached to a surface (e.g., in an array or microarray) so that their 3'-ends have an hydroxyl group and the sequences are at

IDT Exhibit 1017.1968

sufficient distance from the surface so that the sequences can prime rolling circle replication by primer extension of the tethered primer using a circular first-strand cDNA from step (d), to which the primer is complementary, as a template. In this way, the primer that is attached to the surface is primer extended by rolling circle replication using a strand-displacing DNA polymerase under polymerization conditions. In some aspects of this embodiment, the concatemeric rolling circle replication product is labeled by incorporation of a labeled nucleotide, but any other suitable method known in the art can also be used to detect the rolling circle replication product.

Another method for amplifying circularized cDNA prepared from 5'-end fragments of 5'-capped or 5'-triphosphorylated RNA is by rolling circle transcription. Thus, one method that can be used to amplify the 5'-end fragments of mRNA and other primary transcripts comprises the steps of:

a) providing a sample comprising 5'-end fragments of mRNA and other primary transcripts of a desired size range (e.g., without limitation, ≤ 100 bases);

b) contacting the 5'-end fragments with poly(A) polymerase under conditions that result in addition of a poly(A) tail of sufficient length to permit annealing of a reverse transcription primer;

c) contacting the poly(A)-tailed 5'-end fragments with a 5'-phosphorylated primer that anneals to the poly(A) tail and an RNA-dependent DNA polymerase or reverse transcriptase under conditions that result in extension of the primer and synthesis of first-strand cDNA that is complementary to the poly(A)-tailed 5'-end fragments;

d) treating the complex between the first-strand cDNA and the poly(A)-tailed 5'-end fragments of mRNA from step (c) so as to remove the RNA from the first-strand cDNA;

e) contacting the first-strand cDNA from step (d) with a ligase that can ligate single-stranded DNA (ssDNA) in the absence of a ligation splint oligo under conditions that result in intramolecular ligation (i.e., circularization) of the linear first-strand cDNA so as to obtain circular first-strand cDNA; and

f) contacting the circular first-strand cDNA obtained from step (e) with an RNA polymerase under conditions that result in synthesis of concatemeric RNA complementary to the first-strand cDNA by rolling circle transcription.

IDT Exhibit 1017.1969

An oligo$(dT)_n$ or an oligo$(dT)_nX$ anchored primer, wherein X=a mixture of dAMP, dCMP and dGMP, can be used to prime first-strand cDNA synthesis in step (c). In step (c) above, the primer has a 5'-phosphate. In other embodiments, the primer has a 5'-hydroxyl group and is phosphorylated using a polynucleotide kinase (e.g., T4 polynucleotide kinase) under kinasing conditions after step (d) and prior to step (e). RNA-dependent DNA polymerases or reverse transcriptases that can be used include, but are not limited to enzymes that lack RNase H activity. By way of example, but not of limitation, the enzyme can be selected from among an RNase H-minus mutant of MMLV or AMV reverse transcriptase (e.g., SuperScript™ I, II, or III from Invitrogen). However, a reverse transcriptase with RNase H activity can also be used. Ligases that can be used to ligate ssDNA include, but are not limited to, ligases selected from the group consisting of a ligase derived from phage TS2126 that infect Thermus scotoductus, CircLigase™ ssDNA ligase (EPICENTRE), and ThermoPhage™ ssDNA Ligase (PROKARIA, Reykjavik, Iceland). The RNA polymerase can be any RNA polymerase that transcribes a single-stranded circular DNA in the absence of a primer and in the absence of a promoter sequence, although a promoter sequence can be present in some embodiments of the invention. RNA polymerases that can be used include, but are not limited to, *E. coli* RNA polymerase (including core enzyme); a T7-type RNA polymerase, including T7 RNAP, T3 RNAP, and SP6 RNAP; or phage N4 mini-vRNAP.

Still another method that can be used to amplify the 5'-end fragments of mRNA and other primary transcripts comprises the steps of:

a) providing a sample comprising 5'-end fragments of mRNA and other primary transcripts of a desired size range (e.g., without limitation, ≤ 100 bases);

b) contacting the 5'-end fragments with a 5'-phosphorylated random hexamer primer and an RNA-dependent DNA polymerase or reverse transcriptase under conditions that result in extension of the primer and synthesis of first-strand cDNA that is complementary to the 5'-end fragments;

c) treating the complex between the first-strand cDNA and the 5'-end fragments of mRNA from step (b) so as to remove the RNA from the first-strand cDNA;

d) contacting the first-strand cDNA from step (c) with a ligase that can ligate single-stranded DNA (ssDNA) in the absence of a ligation splint oligo under conditions that

IDT Exhibit 1017.1970

result in intramolecular ligation (i.e., circularization) of the linear first-strand cDNA so as to obtain circular first-strand cDNA; and

e) contacting the circular first-strand cDNA obtained from step (d) with an RNA polymerase under conditions that result in synthesis of concatemeric RNA complementary to the first-strand cDNA by rolling circle transcription.

RNA-dependent DNA polymerases or reverse transcriptases that can be used include, but are not limited to enzymes that lack RNase H activity. By way of example, but not of limitation, the enzyme can be selected from among an RNase H-minus mutant of MMLV or AMV reverse transcriptase (e.g., SuperScript™ I, II, or III from Invitrogen). However, a reverse transcriptase with RNase H activity can also be used. Ligases that can be used to ligate ssDNA include, but are not limited to, ligases selected from the group consisting of a ligase derived from phage TS2126 that infect Thermus scotoductus, CircLigase™ ssDNA ligase (EPICENTRE), and ThermoPhage™ ssDNA Ligase (PROKARIA, Reykjavik, Iceland). The RNA polymerase can be any RNA polymerase that transcribes a single-stranded circular DNA in the absence of a primer and in the absence of a promoter sequence, although a promoter sequence can be present in some embodiments of the invention. RNA polymerases that can be used include, but are not limited to, *E. coli* RNA polymerase (including core enzyme); a T7-type RNA polymerase, including T7 RNAP, T3 RNAP, and SP6 RNAP; or phage N4 mini-vRNAP.

The present invention also comprises embodiments comprising amplification of circular first-strand cDNA by both rolling circle replication and rolling circle transcription using any combination of the methods described herein.

The present invention further comprises methods methods, compositions, and kits for amplification of linear first-strand cDNA prepared from 5'-capped or 5'-triphosphorylated mRNA using a random hexamer primer and a strand-displacing DNA polymerase. Thus, one method that can be used to amplify the 5'-end fragments of mRNA and other primary transcripts comprises the steps of:

a) providing a sample comprising 5'-end fragments of mRNA and other primary transcripts of a desired size range (e.g., without limitation, approximately 100 - 500 bases);

IDT Exhibit 1017.1971

b) contacting the 5'-end fragments with poly(A) polymerase under conditions that result in addition of a poly(A) tail of sufficient length to permit annealing of a reverse transcription primer;

c) contacting the poly(A)-tailed 5'-end fragments with a primer that anneals to the poly(A) tail and an RNA-dependent DNA polymerase or reverse transcriptase under conditions that result in extension of the primer and synthesis of first-strand cDNA that is complementary to the poly(A)-tailed 5'-end fragments;

d) treating the complex between the first-strand cDNA and the poly(A)-tailed 5'-end fragments of mRNA from step (c) so as to remove the RNA from the first-strand cDNA; and

e) contacting the first-strand cDNA obtained from step (d) with a strand-displacing DNA polymerase (e.g., phi29 DNA polymerase) and a random hexamer primer that anneals to the first-strand cDNA under conditions that result in amplification of the cDNA.

An oligo$(dT)_n$ or an oligo$(dT)_nX$ anchored primer, wherein X=a mixture of dAMP, dCMP and dGMP, can be used to prime first-strand cDNA synthesis in step 4.A.(c). In preferred embodiments of step 4.A. (c), the RNA-dependent DNA polymerase or reverse transcriptase is an enzyme that does not have RNase H activity. By way of example, but not of limitation, the enzyme is selected from among an RNase H-minus mutant of MMLV or AMV reverse transcriptase (e.g., SuperScript™ I, II, or III from Invitrogen). However, reverse transcriptases with RNase H activity can be used in some embodiments. In some embodiments of step 4.A (e), phi29 DNA polymerase is used as the strand-displacing DNA polymerase and the random hexamer primer used comprises either a random hexamer DNA primer having alpha-thiophosphate internucleoside linkages, which are resistant to many exonucleases, or a random hexamer RNA primer.

Still another embodiment of the invention comprises a method for amplifying the 5'-end fragments of mRNA and other primary transcripts, the method comprising the steps of:

a) providing a sample comprising 5'-end fragments of mRNA and other primary transcripts of a desired size range (e.g., without limitation, approximately 100 - 500 bases);

IDT Exhibit 1017.1972

b) contacting the 5'-end fragments with a random primer that anneals to 5'-end fragments of mRNA and other primary transcripts and an RNA-dependent DNA polymerase or reverse transcriptase under conditions that result in extension of the primer and synthesis of first-strand cDNA that is complementary to the 5'-end fragments;

c) treating the complex between the first-strand cDNA and the 5'-end fragments of mRNA from step (b) so as to remove the RNA from the first-strand cDNA; and

d) contacting the first-strand cDNA obtained from step (c) with a strand-displacing DNA polymerase (e.g., phi29 DNA polymerase) and a random hexamer primer that anneals to the first-strand cDNA under conditions that result in amplification of the cDNA.

The random primer used in step 4.B. (b) can comprise a random hexamer DNA primer having alpha-thiophosphate internucleoside linkages. In preferred embodiments of step 4.A. (b), the RNA-dependent DNA polymerase or reverse transcriptase is an enzyme that does not have RNase H activity. By way of example, but not of limitation, the enzyme is selected from among an RNase H-minus mutant of MMLV or AMV reverse transcriptase (e.g., SuperScript™ I, II, or III from Invitrogen). However, reverse transcriptases with RNase H activity can be used in some embodiments. The primer in step 4.B.(d) can comprise either a random hexamer DNA primer having alpha-thiophosphate internucleoside linkages or a random hexamer RNA primer. One DNA polymerase that can be used in this step is phi29 DNA polymerase.

The present invention further provides kits and methods for selectively modifying nucleic acids that are not substrates for a 5' exonuclease of the present invention so that said nucleic acids will be substrates for said 5' exonuclease and thereby, can be removed from desired nucleic acid molecules. For example, but without limitation, an RNA molecule that is not a substrate for a 5'-phosphate-dependent exoribonuclease because it has a triphosphate group or an hydroxyl group on its 5'-terminus can be hybridized to a DNA oligo that anneals near to its 5'-end and the RNA:oligo complex can be contacted with a ribonuclease H, such as but not limited to *E. coli* RNase H or a *Thermus* RNase H (e.g., Hybridase RNase H from EPICENTRE). The RNase H digests the RNA within the portion that is annealed to the oligo, yielding an RNA molecule that has a 5'-phosphorylated RNA that is a substrate for the 5'-phosphate-dependent exoribonuclease that can be substantially digested in the presence of desired nucleic acids. Thus, in some

IDT Exhibit 1017.1973

preferred embodiments, the present invention provides a method for enriching for a desired RNA in the presence of an undesired RNA comprising the steps of: a) providing a 5'-phosphate-dependent nucleic acid exoribonuclease; a sample comprising RNA, said sample containing desired RNA and an undesired RNA that is not a substrate for the exoribonuclease; a DNA oligo that is complementary to a sequence near the 5'-terminus of the undesired RNA; and an RNase H; b) contacting the sample with the DNA oligo and the RNase H under conditions under which the oligo anneals to the undesired RNA and the RNase H is active; c) treating the sample with the exoribonuclease under conditions such that the undesired RNA is substantially digested and the desired RNA is not digested; and d) recovering the desired RNA (e.g., for additional analysis or use). In some embodiments, the 5'-phosphate-dependent nucleic acid exoribonuclease used in the method is provided in a kit that contains the DNA oligo, RNase H, buffers, instructions, and optionally, appropriate control samples, and the like.

The present invention is not limited with respect to kits and methods for modifying nucleic acids to be either desired or undesired nucleic acids that are, respectively, either resistant to digestion or susceptible to digestion by a 5' exonuclease of the present invention. Any appropriate modification method can be used to achieve a particular purpose in a method of the invention. For example, but without limitation, a 5'-exonuclease-resistant nucleic acid with a 5'-hydroxyl can be modified to a 5'-phosphorylated nucleic acid that is a substrate for digestion by treatment with a polynucleotide kinase, such as T4 polynucleotide kinase, and ATP under kinase reaction conditions; or a 5'-phosphorylated nucleic acid that is a substrate for digestion by a 5'-exonuclease can be dephosphorylated to a 5'-exonuclease-resistant nucleic acid with a 5'-hydroxyl by treatment with a DNA phosphatase, such as, but not limited to APex Alkaline Phosphatase (EPICENTRE) or shrimp alkaline phosphatase; or a 5'-exonuclease-resistant 5'-triphosphorylated or 5'-capped mRNA can be modified to a 5'-exonuclease substrate that has a 5'-phosphate by treatment with a pyrophosphatase, such as but not limited to, tobacco acid pyrophosphatase. Those with skill in the art will know or know how to identify other methods for modifying undesired nucleic acids so that they will be substrates for a 5' exonuclease of the invention or for modifying desired nucleic

IDT Exhibit 1017.1974

acids so they will not be substrates for the 5' exonuclease, all of which modification methods are included within the scope of the present invention.

The present invention further provides methods and kits for removing undesired 5'-phosphorylated oligonucleotides from nucleic acid reactions (e.g., without limitation, aRNA and sense RNA amplification reactions), comprising the steps of: a) providing a sample that contains desired nucleic acid molecules that have undergone or are undergoing enzymatic manipulation and undesired 5'-phosphorylated oligonucleotides that are involved in said manipulation of said desired nucleic acid molecules; and b) exposing the sample to a 5'-phosphate-dependent nucleic acid exonuclease under conditions such that the undesired 5'-phosphorylated oligonucleotides are substantially digested and the desired nucleic acid molecules are not digested. The method finds use in a wide variety of applications. By way of example, but not of limitation, U.S. Patent Appln. Nos. 2004/0171041 and 2004/0197802, which are incorporated herein by reference in their entireties, disclose methods for which this aspect of the present invention can be used. For example, the present method can be used for methods disclosed therein that use a splice template oligo (or template switch oligo), such as a promoter splice template oligo to synthesize first-strand cDNA comprising a target sequence in order to reduce subsequent background transcription due to the presence of the splice template oligo if it is not removed. The splice template oligo can have a 5'-phosphate group or it can be phosphorylated during the method using a polynucleotide kinase. The present invention is not limited to removal of splice template oligos. The invention also includes removal of other oligos that are undesired nucleic acids. (e.g., removal of linear promoter primers from circular first-strand cDNA synthesized by primer extension of the promoter primer and subsequent ligation using a ligase in the presence of a ligation splint oligo or using a ligase that ligates ssDNA in the absence of a ligation splint oligo (e.g., CircLigase ssDNA Ligase (EPICENTRE Biotechnologies)), and the like, and from RNA amplification reactions). In some embodiments, the undesired 5'-phosphorylated oligonucleotide includes, but is not limited to, primers, ligation splint oligos, etc. In some embodiments, the enzymatic manipulation comprises a ligation reaction, a nucleic acid synthesis reaction (e.g., *in vitro* transcription reaction, cDNA synthesis reaction, Eberwine-type aRNA amplification reaction, sense RNA

IDT Exhibit 1017.1975

amplification reaction, etc.). Kits for conducting such methods comprise a 5'-phosphate-dependent exonuclease and may have any one or more other reagents (e.g., primers), control samples (e.g., 5'-phosphorylated oligonucleotides), buffers, etc. useful in the method.

The compositions and methods of the present invention provided herein can be used in diagnostic methods. For example, the present invention provides a method for detecting the presence of a target nucleic acid in a sample, comprising the steps of: a) providing a sample comprising a mixture of nucleic acids, said mixture of nucleic acids suspected of containing the target nucleic acid; b) exposing the sample to a 5'-phosphate-dependent nucleic acid exonuclease; and c) detecting the presence of undigested nucleic acid molecules that are not substrates for the exonuclease or detecting the products and/or process of digestion of nucleic acid molecules that are substrates for the exonuclease in order to detect the presence of the target nucleic acid in the sample. The method is not limited by the nature of the target nucleic acid. In some embodiments, the target nucleic acid, if present in the sample, is modified as discussed herein above to make it susceptible or resistant to cleavage by the 5'-phosphate-dependent nucleic acid exonuclease in a target-specific manner. The target nucleic acid may also be differentiated from non-target sequence by its different behavior in response to particular modification treatments (e.g., susceptibility to RNase H digestion in the presence of a DNA oligo or restriction enzyme digestion due to the presence of a particular sequence). The present invention also provides kits, mixtures, and compositions for carrying out such methods. For example, the present invention provides a kit for isolating a nucleic acid molecule of interest in a mixture of nucleic acids comprising one or more of: a) a 5'-phosphate-dependent nucleic acid exonuclease; b) a positive control sample comprising nucleic acid having a 5' phosphate (e.g., rRNA); and c) a negative control sample comprising nucleic acid that is resistant to degradation by the 5'-phosphate-dependent nucleic acid exonuclease (e.g., mRNA).

The present invention further provides a method for isolating 5'-capped or 5'-triphosphorylated or 5'-hydroxylated small RNA molecules in samples containing eukaryotic RNA (e.g., small nuclear RNA (snRNA) or pre-microRNA (pre-miRNA) to study the their functions in gene processing and control of gene expression, etc. in health

IDT Exhibit 1017.1976

and disease) or 5'-capped or 5'-triphosphorylated transcripts of other small RNA molecules, such as but not limited to, primary, the method comprising the steps of: a) providing a sample comprising nuclear RNA from a eukaryotic organism (e.g., obtained from isolated nuclei or from total RNA preparations using methods known in the art), said sample containing desired small RNA having a 5'-cap or 5'-triphosphate or 5'-hydroxyl group and at least one undesired nucleic acid; and a 5'-to-3' exoribonuclease; b) treating the sample with the 5'- to-3' exoribonuclease under conditions such that said undesired nucleic acid is substantially digested and the desired 5'-capped or 5'-triphosphorylated or 5'-hydroxylated small RNA is not digested; and c) recovering the capped or triphosphorylated or hydroxylated small RNA. In some embodiments, the 5'-phosphate-dependent nucleic acid exonuclease used in the method is provided in a kit for isolation of capped or triphosphorylated or hydroxylated small RNA, with buffers, instructions, and optionally, with appropriate control samples, and the like. Pre-miRNA, which has a 5'-triphosphate, is further processed in the cell to micro RNA (miRNA) having a 5'-monophosphate, which is implicated in control of gene expression of many. In some embodiments of methods of the invention, a 5'-group is modified in order to achieve a purpose. For example, without limitation, in one embodiment of the method, small RNA with a 5'-phosphate group can be removed using a nucleic acid phosphatase, such as APex Phosphatase (EPICENTRE), to form a 5'-hydoxyl group, which is resistant to digestion with the 5'-to-3 exoribonuclease, thereby permitting enrichment for the small RNA (such as miRNA). In another embodiment, the small RNA with a 5'-triphosphate or 5'-cap can be treated with tobacco acid pyrophosphatase (EPICENTRE) to form a 5'-phosphate group, thereby permitting digestion of the small RNA, such as pre-miRNA, with the 5'-to-3 exoribonuclease.

The present invention further comprises methods to prepare cDNA from said isolated capped or triphosphorylated small RNA, the method comprising a) providing a sample comprising capped and/or triphosphorylated small RNA; a poly(A) polymerase; and an RNA-dependent DNA polymerase; b) contacting the sample with the poly(A) polymerase in the presence of ATP and under conditions in which a 3'-poly(A) tail is added to the desired small RNA molecules; and c) contacting the 3'-polyadenylated RNA with the oligo(dT)-containing primer and extending the primer with the RNA-dependent

IDT Exhibit 1017.1977

DNA polymerase under conditions whereby cDNA is obtained. Other embodiments comprise methods for amplifying the polyadenylated small RNA, the methods comprising a) contacting the polyadenylated small RNA with a primer that anneals to said tail and extending said primer with an RNA-dependent DNA polymerase under conditions whereby cDNA that is joined to an RNA polymerase promoter sequence is obtained (using any number of methods known in the art; methods are known for joining the promoter to either the 5'-end of the cDNA or the 3'-end of the cDNA, resulting in synthesis of amplified antisense or sense RNA, respectively, by subsequent *in vitro* transcription); b) producing double stranded cDNA; and c) contacting said double-stranded cDNA comprising said promoter sequence with an RNA polymerase that binds to said promoter and synthesizes RNA therefrom under conditions whereby amplified RNA corresponding to polyadenylated RNA in the sample is synthesized. Kits for conducting such methods may have any one or more reagents (e.g., primers), control samples, buffers, etc. useful in the method. For example, preferred kits comprise: a) a 5'-phosphate-dependent nucleic acid exonuclease; b) a poly(A) polymerase; c) an RNA-dependent DNA polymerase; and, for methods comprising RNA amplification, d) an RNA polymerase (e.g., T7, T3, and SP6 RNA polymerases).

One preferred embodiment of the invention is a kit for enrichment of mRNA from a sample comprising mRNA and at least one undesired nucleic acid, the kit comprising a 5'-to-3' exoribonuclease that has an activity in digesting an RNA having a 5'-triphosphate or a 5'-cap that is less than 5% of its activity in digesting said RNA having a 5'-monophosphate.

Another preferred embodiment is a kit for enrichment of mRNA from a sample comprising mRNA and at least one undesired nucleic acid, the kit comprising: a) a 5'-to-3' exoribonuclease, wherein the activity in digesting an RNA having a 5'-triphosphate or a 5'-cap is less than 5% of its activity in digesting said RNA having a 5'-monophosphate; and b) a negative control comprising an RNA having a 5'-triphosphate or a 5'-cap.

Another preferred embodiment is a kit for enrichment of mRNA from a sample comprising mRNA and at least one undesired nucleic acid, the kit comprising: a) a 5'-to-3' exoribonuclease, wherein the activity in digesting an RNA having a 5'-triphosphate or

IDT Exhibit 1017.1978

a 5'-cap is less than 5% of its activity in digesting said RNA having a 5'-monophosphate; and b) a positive control comprising an RNA having a 5'-monophosphate.

Still another preferred embodiment is a kit for enrichment of mRNA from a sample comprising mRNA and at least one undesired nucleic acid, the kit comprising: a) a 5'-to-3' exoribonuclease, wherein the activity in digesting an RNA having a 5'-triphosphate or a 5'-cap is less than 5% of its activity in digesting said RNA having a 5'-monophosphate; b) a negative control comprising an RNA having a 5'-triphosphate or a 5'-cap; and c) a positive control comprising an RNA having a 5'-monophosphate.

Another preferred embodiment of the invention is a kit for isolation of mRNA from a sample comprising mRNA and at least one undesired nucleic acid, the kit comprising: a) a 5'-to-3' exoribonuclease, wherein the activity in digesting an RNA having a 5'-triphosphate or a 5'-cap is less than 5% of its activity in digesting said RNA having a 5'-monophosphate; and b) a solution of LiCl for precipitation of low molecular RNA.

Another preferred embodiment is a kit for isolation of mRNA from a sample comprising mRNA and at least one undesired nucleic acid, the kit comprising: a) a 5'-to-3' exoribonuclease, wherein the activity in digesting an RNA having a 5'-triphosphate or a 5'-cap is less than 5% of its activity in digesting said RNA having a 5'-monophosphate; b) a solution of LiCl for precipitation of low molecular RNA; and c) a negative control comprising an RNA having a 5'-triphosphate or a 5'-cap.

Another preferred embodiment is a kit for isolation of mRNA from a sample comprising mRNA and at least one undesired nucleic acid, the kit comprising: a) a 5'-to-3' exoribonuclease, wherein the activity in digesting an RNA having a 5'-triphosphate or a 5'-cap is less than 5% of its activity in digesting said RNA having a 5'-monophosphate; b) a solution of LiCl for precipitation of low molecular RNA; and c) a positive control comprising an RNA having a 5'-monophosphate.

Still another preferred embodiment is a kit for isolation of mRNA from a sample comprising mRNA and at least one undesired nucleic acid, the kit comprising: a) a 5'-to-3' exoribonuclease, wherein the activity in digesting an RNA having a 5'-triphosphate or a 5'-cap is less than 5% of its activity in digesting said RNA having a 5'-monophosphate; b) a solution of LiCl for precipitation of low molecular RNA; c) a negative control

IDT Exhibit 1017.1979

comprising an RNA having a 5'-triphosphate or a 5'-cap; and d) a positive control comprising an RNA having a 5'-monophosphate.

The present invention further comprises a kit for enrichment of mRNA having a 5'-triphosphate or a 5'-cap from a sample comprising mRNA and at least one undesired nucleic acid, the kit comprising: a) a 5'-to-3' exoribonuclease, wherein the activity in digesting an RNA having a 5'-triphosphate or a 5'-cap is less than 5% of its activity in digesting said RNA having a 5'-monophosphate; and b) a polynucleotide kinase.

Another embodiment of the invention is a kit for isolation of mRNA having a 5'-triphosphate or a 5'-cap from a sample comprising mRNA and at least one undesired nucleic acid, the kit comprising: a) a 5'-to-3' exoribonuclease, wherein the an activity in digesting an RNA having a 5'-triphosphate or a 5'-cap is less than 5% of its activity in digesting said RNA having a 5'-monophosphate; b) a polynucleotide kinase; and c) a solution of LiCl for precipitation of low molecular RNA.

The present invention further comprises a kit for synthesis of first-strand cDNA complementary to mRNA from a sample comprising mRNA and at least one undesired nucleic acid, the kit comprising: a) a 5'-to-3' exoribonuclease, wherein the activity in digesting an RNA having a 5'-triphosphate or a 5'-cap is less than 5% of its activity in digesting said RNA having a 5'-monophosphate; b) at least one primer that anneals to said mRNA in said sample; and c) an RNA-dependent DNA polymerase. One kit according to this embodiment is a kit for synthesis of first-strand cDNA complementary to mRNA from a prokaryotic organism. Another kit according to this embodiment is a kit for synthesis of first-strand cDNA complementary to mRNA from a eukaryotic organism. Still another kit according to this embodiment is a kit for synthesis of first-strand cDNA complementary to mRNA from both a prokaryotic organism and a eukaryotic organism.

Still further, another preferred embodiment of the invention is a kit for synthesis of first-strand cDNA complementary to mRNA from a prokaryotic organism, the kit comprising: a) a 5'-to-3' exoribonuclease, wherein the activity in digesting an RNA having a 5'-triphosphate or a 5'-cap is less than 5% of its activity in digesting said RNA having a 5'-monophosphate; b) a poly(A) polymerase; c) an oligo(dT)-containing primer that is capable of annealing to a poly(A) tail; and d) an RNA-dependent DNA polymerase. Another kit according to this embodiment is a kit for synthesis of first-strand

IDT Exhibit 1017.1980

cDNA complementary to mRNA from both a prokaryotic organism and a eukaryotic organism in a sample comprising mRNA from both the prokaryotic and the eukaryotic organisms.

Still another preferred embodiment of the invention is a kit for synthesis of first-strand cDNA complementary to mRNA from a prokaryotic organism, the kit comprising: a) a 5'-to-3' exoribonuclease, wherein the activity in digesting an RNA having a 5'-triphosphate or a 5'-cap is less than 5% of its activity in digesting said RNA having a 5'-monophosphate; b) a poly(A) polymerase; c) a polynucleotide kinase; d) an oligo(dT)-containing primer that is capable of annealing to a poly(A) tail; and e) an RNA-dependent DNA polymerase. Another kit according to this embodiment is a kit for synthesis of first-strand cDNA complementary both to mRNA from a prokaryotic organism and to full-length mRNA from a eukaryotic organism in a sample comprising mRNA from both the prokaryotic and the eukaryotic organisms.

Yet another preferred embodiment of the present invention is a kit for synthesis of first-strand cDNA complementary to full-length mRNA from a eukaryotic organism, the kit comprising: a) a 5'-to-3' exoribonuclease, wherein the activity in digesting an RNA having a 5'-triphosphate or a 5'-cap is less than 5% of its activity in digesting said RNA having a 5'-monophosphate; b) a polynucleotide kinase; c) an oligo(dT)-containing primer that is capable of annealing to a poly(A) tail; and d) an RNA-dependent DNA polymerase.

Still further, the invention also comprises a kit for RNA amplification of mRNA from a sample comprising mRNA and at least one undesired nucleic acid, the kit comprising: a) a 5'-to-3' exoribonuclease, wherein the activity in digesting an RNA having a 5'-triphosphate or a 5'-cap is less than 5% of its activity in digesting said RNA having a 5'-monophosphate; b) at least one primer that anneals to said mRNA in said sample; c) an oligonucleotide that encodes an RNA polymerase promoter; d) an RNA-dependent DNA polymerase; and e) an RNA polymerase that binds to the RNA polymerase promoter encoded by the oligonucleotide. One kit according to this embodiment is a kit for RNA amplification of mRNA from a prokaryotic organism. Another kit according to this embodiment is a kit for RNA amplification of mRNA from

IDT Exhibit 1017.1981

a eukaryotic organism. Still another kit according to this embodiment is a kit for RNA amplification of mRNA from both a prokaryotic organism and a eukaryotic organism.

One preferred embodiment of the invention is a kit for RNA amplification of mRNA from a prokaryotic organism, the kit comprising: a) a 5'-to-3' exoribonuclease, wherein the activity in digesting an RNA having a 5'-triphosphate or a 5'-cap is less than 5% of its activity in digesting said RNA having a 5'-monophosphate; b) a poly(A) polymerase; c) an oligo(dT)-containing primer that is capable of annealing to a poly(A) tail; d) an oligonucleotide that encodes an RNA polymerase promoter; e) an RNA-dependent DNA polymerase; and f) an RNA polymerase that binds to the RNA polymerase promoter encoded by the oligonucleotide. One kit according to this embodiment of the invention is a kit for RNA amplification of mRNA from both a prokaryotic organism and a eukaryotic organism in a sample comprising mRNA from both the prokaryotic and the eukaryotic organisms.

Still another preferred embodiment of the invention is a kit for RNA amplification of mRNA from a prokaryotic organism, the kit comprising: a) a 5'-to-3' exoribonuclease, wherein the activity in digesting an RNA having a 5'-triphosphate or a 5'-cap is less than 5% of its activity in digesting said RNA having a 5'-monophosphate; b) a poly(A) polymerase; c) a polynucleotide kinase; d) an oligo(dT)-containing primer that is capable of annealing to a poly(A) tail; e) an oligonucleotide that encodes an RNA polymerase promoter; e) an RNA-dependent DNA polymerase; and f) an RNA polymerase that binds to the RNA polymerase promoter encoded by the oligonucleotide. One kit according to this embodiment is a kit for RNA amplification of both mRNA from a prokaryotic organism and full-length mRNA from a eukaryotic organism in a sample comprising mRNA from both the prokaryotic and the eukaryotic organisms.

The present invention further comprises a kit for RNA amplification of mRNA from a eukaryotic organism, the kit comprising: a) a 5'-to-3' exoribonuclease, wherein the activity in digesting an RNA having a 5'-triphosphate or a 5'-cap is less than 5% of its activity in digesting said RNA having a 5'-monophosphate; b) an oligo(dT)-containing primer that is capable of annealing to a poly(A) tail; c) an oligonucleotide that encodes an RNA polymerase promoter; d) an RNA-dependent DNA polymerase; and e) an RNA polymerase that binds to the RNA polymerase promoter encoded by the oligonucleotide.

IDT Exhibit 1017.1982

A preferred embodiment of the invention comprises a kit for RNA amplification of full-length mRNA from a eukaryotic organism, the kit comprising: a) a 5'-to-3' exoribonuclease, wherein the activity in digesting an RNA having a 5'-triphosphate or a 5'-cap is less than 5% of its activity in digesting said RNA having a 5'-monophosphate; b) a polynucleotide kinase; c) an oligo(dT)-containing primer that is capable of annealing to a poly(A) tail; d) an oligonucleotide that encodes an RNA polymerase promoter; e) an RNA-dependent DNA polymerase; and f) an RNA polymerase that binds to the RNA polymerase promoter encoded by the oligonucleotide.

The present invention also comprises a kit for removing ribosomal RNA from a sample, the kit comprising: a) a 5'-to-3' exoribonuclease, wherein the activity in digesting an RNA having a 5'-triphosphate or a 5'-cap is less than 5% of its activity in digesting said RNA having a 5'-monophosphate; and b) a positive control comprising rRNA selected from among a) 16S and 23S prokaryotic rRNA; and b) 18S and 26S or 28S eukaryotic rRNA. A preferred kit of this embodiment comprises a kit for removing ribosomal RNA >300 nucleotides from a sample. Other kits according to these embodiments comprise kits that additionally comprise a negative control comprising an RNA having a 5'-triphosphate, a 5'-cap, or a 5'-hydroxyl.

One preferred embodiment of the invention is a kit for enriching for mRNA having a 5'-triphosphate or a 5'-cap in a biological sample comprising prokaryotic mRNA, eukaryotic mRNA or both prokaryotic and eukaryotic mRNA and at least one undesired nucleic acid, the kit comprising: (i) a solution of purified and stabilized 5'-to-3' exoribonuclease; and (ii) a concentrated solution of a reaction buffer for providing a 1X reaction buffer in which the exoribonuclease is active. One suitable 1X concentration of a reaction buffer that can be used for the invention comprises: 50 mM Tris-HCl (pH 8.0), 2.0 mM $MgCl_2$, and 100 mM NaCl.

In another embodiment, a kit for enriching for mRNA additionally comprises one or more components selected from the group consisting of: (i) a negative control comprising an RNA having a 5'-triphosphate or a 5'-cap; (ii) a positive control comprising an RNA having a 5'-monophosphate; (iii) a solution of LiCl; (iv) a polynucleotide kinase; and (v) a poly(A) polymerase.

IDT Exhibit 1017.1983

In still another embodiment, the kit additionally comprises an RNA-dependent DNA polymerase (reverse transcriptase).

In still another embodiment, the kit additionally comprises one or more components selected from the group consisting of: (i) a poly(A) polymerase; (ii) an oligo(dT)-containing primer; and (iii) an RNA-dependent DNA polymerase.

In yet another embodiment, the kit additionally comprises one or more components selected from the group consisting of: (i) a polynucleotide kinase; (ii) an oligo(dT)-containing primer; and (iii) an RNA-dependent DNA polymerase for obtaining cDNA complementary to full-length mRNA from a eukaryotic organism.

In yet another embodiment, the kit additionally comprises one or more primers selected from the group consisting of (i) an oligo(dT)-containing primer; and (ii) a primer that is complementary to a specific nucleic acid sequence. In some aspects of these embodiments, the primer additionally comprises a sense or an antisense sequence of a double-stranded promoter for a T7-type RNA polymerase selected from the group consisting of T7 RNA polymerase, T3 RNA polymerase, and SP6 RNA polymerase, and wherein the kit additionally comprises the RNA polymerase that can transcribe RNA using said promoter.

In some embodiments the kit additionally comprises: (i) a oligonucleotide sequence tag template; and (ii) a DNA polymerase selected from the group consisting of a DNA-dependent DNA polymerase and an RNA-dependent DNA polymerase that can extend the 3'-end of a nucleic acid that is annealed to the oligonucleotide sequence tag template. In some of these embodiments, the kit additionally comprises: (i) an oligonucleotide that is complementary to the oligonucleotide sequence tag template; and (ii) a T7-type RNA polymerase selected from the group consisting of T7 RNA polymerase, T3 RNA polymerase, and SP6 RNA polymerase; wherein the oligonucleotide sequence tag template and the oligonucleotide that is complementary to the oligonucleotide sequence tag template together encode the complete sequence of a promoter that can be used for transcription by said T7-type RNA polymerase.

In many cases, a kit of the present invention will contain one or more components selected from the group consisting of: (i) a negative control comprising an RNA having

IDT Exhibit 1017.1984

a 5'-triphosphate or a 5'-cap; (ii) a negative control comprising an RNA or a DNA having a 5'-hydroxyl group; (iii) a positive control comprising an RNA having a 5'-monophosphate; (iv) a solution of LiCl; (v) a ribonuclease H enzyme; and (vi) a polynucleotide kinase. The ribonuclease H (RNase H)can be used, together with a DNA oligonucleotide that anneals to a specific sequence of an mRNA molecule, to cleave the mRNA in a particular region prior to adding sequence tag by extension of the 3'-end of the RNase H-cleaved mRNA having a 5'-triphosphate or a 5'cap using an oligonucleotide sequence tag template as a template. The polynucleotide kinase can be used to phosphorylate the 5'-end of degraded RNA molecules having a 5'-hydroxyl group in order to generate 5'-monophosphorylated RNA substrates that can digested using a 5'-to-3' exoribonuclease of the invention.

Some kits of the invention will also comprise one or more components selected from the group consisting of: (i) a poly(A) polymerase; (ii) an oligo(dT)-containing primer; and (iii) an RNA-dependent DNA polymerase, such as, but not limited to, kits for enriching for prokaryotic mRNA. Some of such kits may also contain an RNA-dependent DNA polymerase and one or more primers selected from the group consisting of (i) a random primer, such as a random hexamer primer; (ii) an oligo(dT)-containing primer; and (iii) a primer that is complementary to a specific nucleic acid sequence. Some of such kits may also comprise: (i) a oligonucleotide sequence tag template for adding a tag sequence to the 3'-end of mRNA having a 5'-triphosphate or a 5'-cap; and (ii) a DNA polymerase selected from the group consisting of a DNA-dependent DNA polymerase and an RNA-dependent DNA polymerase that can extend the 3'-end of the mRNA using the oligonucleotide sequence tag template annealed to the mRNA as a template.

The kit of the invention can also comprise: (i) an oligonucleotide sequence tag template for adding a tag sequence to the 3'-end of mRNA having a 5'-triphosphate or a 5'-cap; and (ii) a DNA polymerase selected from the group consisting of a DNA-dependent DNA polymerase and an RNA-dependent DNA polymerase that can extend the 3'-end of the mRNA using the oligonucleotide sequence tag template annealed to the mRNA as a template. The DNA polymerase can be an RNA-dependent DNA polymerase, or, if a DNA-dependent DNA polymerase is used for adding the tag sequence to the 3'-end of the mRNA, the kit additionally contains an RNA-dependent

IDT Exhibit 1017.1985

DNA polymerase that can synthesize cDNA from a primer using the mRNA as a template.

In some embodiments of the invention, the oligonucleotide sequence tag template can additionally encode either a sense or an antisense strand of a double-stranded promoter for a T7-type RNA polymerase selected from the group consisting of T7 RNA polymerase, T3 RNA polymerase, and SP6 RNA polymerase, and in those embodiments, the kit can additionally comprise the RNA polymerase that can transcribe RNA using said promoter.

Still further, a kit of the invention can also comprise a poly(A) polymerase for tailing of prokaryotic mRNA, and an oligo(dT)-containing primer and an RNA-dependent DNA polymerase for synthesis of cDNA. The kit can also contain an oligonucleotide sequence tag template (for adding a sequence tag to the 3'-end of first-strand cDNA); an oligonucleotide that is complementary to the oligonucleotide sequence tag template; wherein the oligonucleotide sequence tag template and the oligonucleotide complementary to the oligonucleotide sequence tag template together can encode the complete sequence of a promoter for a T7-type RNA polymerase. Thus, in some such embodiments, the kit can also contain a T7-type RNA polymerase selected from the group consisting of a T7 RNA polymerase, a T3 RNA polymerase, and an SP6 RNA polymerase, wherein the T7-type RNA polymerase can use the promoter encoded by the combination of the oligonucleotide sequence tag template and the oligonucleotide complementary to the oligonucleotide sequence tag template.

Still other kits of the invention are useful for removing intact prokaryotic or eukaryotic ribosomal RNA (rRNA) of a size having a svedburg unit greater than about 10S from a biological sample. These kits can comprise: (i) a solution of purified and stabilized 5'-to-3' exoribonuclease that is free of contaminating *S. cerevisiae* protein; (ii) a concentrated solution of a reaction buffer for providing a 1X reaction buffer in which the exoribonuclease is active; and (iii) a positive control rRNA, selected from the group consisting of a 16S and a 23S prokaryotic rRNA, and an 18S, a 26S and a 28S eukaryotic rRNA. In a preferred embodiment of this aspect of the invention, the 5'-to-3' exoribonuclease is a wild-type or recombinant *Saccharomyces cerevisiae* Xrn1p/5' exoribonuclease 1.

IDT Exhibit 1017.1986

In addition to the kits described above, it should be understood that the present invention provides a wide variety of kits for use in an equally wide variety of applications. In some preferred embodiments, the kits comprise a 5'-phosphate-dependent nucleic acid exonuclease (e.g., in a form, concentration, etc. suitable for use in the methods described herein). In other preferred embodiments, the kits comprise or consist of a 5'-phosphate-dependent nucleic acid exonuclease and one or more other components with which the 5'-phosphate-dependent nucleic acid exonuclease might be used directly or indirectly, either alone together or with yet other components. The one or more components may be any component involved in any of the methods described herein or in any of the methods described in the references or patents cited herein. The one or more components include, but are not limited to, buffers, salts, enzymes, proteins, antibodies, nucleic acid molecules (e.g., primers, probes, antisense oligonucleotides, RNA, DNA, etc.), nucleotides, control reagents, instructions, solid surfaces (e.g., beads, microarray components, etc.), detection reagents (e.g., fluorescent, luminescent, colorimetric, etc.), general laboratory equipment (tubes, multiwell plates, etc.), software, vectors or vector components, cell culture reagents and materials, and the like.

DESCRIPTION OF THE FIGURE

Figure 1 shows the amino acid (SEQ ID NO:1) and nucleic acid (SEQ ID NO:2) sequences of Saccharomyces cerevisiae Xrn1p.

DEFINITIONS:

The present invention will be understood and interpreted based on the definitions of terms as defined below.

A "5'-to-3' exoribonuclease" or a "5' exoribonuclease" or a "5' Xrn1p exoribonuclease" or a "5' exonuclease" or a "5'-phosphate-dependent exonuclease" or a "5'-phosphate-dependent nucleic acid exonuclease" or a "5 nuclease" of the present invention comprises an 5'-exonuclease that has greater than 20-fold more 5'-to-3' exonuclease activity for a single-stranded RNA substrate that has a 5'-monophosphorylated terminus than for the same RNA substrate that has a 5'-triphosphorylated or 5'-capped terminus, or, any 5'-to-3' exonuclease that has a relative

IDT Exhibit 1017.1987

activity in digesting a particular defined-sequence single-stranded RNA substrate with a 5'-triphosphate or a 5'-cap that is less than 5% of the activity for an RNA substrate having the same sequence but with a 5'-monophosphate. Enzyme activity of a 5' exonuclease of the invention can be measured using a number of different methods. Without limitation, suitable methods that can be used for assaying activity and determining relative activity using RNA substrates with a 5'-triphosphate, a 5'-cap, or a 5'-monophosphate are described by Stevens and Poole (J. Biol. Chem., 270: 16063, 1995).

Preferred embodiments of compositions, kits and methods of the invention employ Xrn1p/5' exoribonuclease 1, including *Saccharomyces cerevisiae* Xrn1p/5' exoribonuclease 1, and/or functional variants and homologues thereof. Xrn1p was first identified and characterized in *Saccharomyces cerevisiae*, but the activity has been reported in numerous organisms, including humans. In preferred embodiments of the present invention, the 5'-to-3' exoribonuclease comprises a solution of the enzyme that has been purified so as to be free of contaminating enzymes with activity on nucleic acids and *Saccharomyces cerevisiae* proteins. In some preferred embodiments, the 5'-to-3' exoribonuclease is obtained by expression of the *Saccharomyces cerevisiae* XRN1 gene that has been cloned in a plasmid, and then replicated and expressed in *Esherichia coli* cells, since the Xrn1p/5' exoribonuclease I obtained from such a recombinant source is of a higher purity, free from contaminating enzymatic activities, and generally at a higher enzyme concentration than is obtained from non-recombinant sources. In preferred embodiments of the invention, the 5'-to-3' exoribonuclease is "stabilized", by which we mean that the 5'-to-3' exoribonuclease is sufficiently pure of proteases and other contaminants which contribute to degradation and loss of enzyme activity and that the 5'-to-3' exoribonuclease is provided in a formulation of enzyme storage buffer in which there is no significant loss of activity during storage at -20 degrees C for six months. One suitable enzyme storage buffer for providing a stabilized 5'-to-3' exoribonuclease comprises a 50% glycerol solution containing 50 mM Tris-HCL (pH 7.5), 100 mM NaCl, 100 mM EDTA, 1 mM DTT and 0.1% of the non-ionic detergent Triton X-100. The *Saccharomyces cerevisiae* purified yeast enzyme was initially found to be a 160-kDa protein but, based on its gene sequence, the actual size is closer to 175 kDa. The amino acid (SEQ ID NO:1) and nucleic acid (SEQ ID NO:2) sequences of Saccharomyces

IDT Exhibit 1017.1988

cerevisiae Xrn1p are shown in Figure 1. "Xrn1p" generally refers to the protein, whereas "XRN1" generally refers to the gene; however, the terms "Xrn1p" and "Xrn1p/5' exoribonuclease I" are used interchangeably herein and the term "Xrn1p", as used herein, refers to both the protein and gene unless indicated otherwise. The enzyme is a processive exoribonuclease, requiring a divalent cation, and generally is stimulated by monovalent cations. The enzyme acts on a variety of substrates, including homopolymers, rRNA, and oligoadenylates. The enzyme also has some DNase activity. Single-stranded molecules are the preferred substrate. The XRN1 gene of Saccharomyces cerevisiae is identical to a gene (DST2, SEP1) encoding a DNA transferase and to genes involved in nuclease fusion, KEM1, and plasmid stability, RAR5 (see e.g., Larimer et al., Gene 120:51, 1992). Another name for this gene is Stpβ (see e.g., Heyer et al., Mol. Cell Biol. 15:2728, 1995).

Without being bound by theory, sequence comparison of several Mg2+-dependent 5'-3' exonucleases from bacteriophages, prokaryotes, and eukaryotes (Solinger et al., Mol. Cell. Biol. 19:5930, 1995) revealed some highly conserved amino acid residues. In the crystal structure of phage T4 RNase H, these residues are clustered in the proposed active site. In particular, some conserved aspartic (D) and glutamic (E) acid residues coordinate two Mg2+ ions in the reactive center of the exonuclease and are believed to play a crucial role in catalysis. Mutation of D206 and D208 of S. cerevisiae Xrn1p abolished exonuclease activity in the mutants.

Moreover, variant forms of Xrn1p are also contemplated as being equivalent to those peptides and DNA molecules that are set forth in more detail herein. For example, it is contemplated that isolated replacement of a leucine with an isoleucine or valine, an aspartate with a glutamate, a threonine with a serine, or a similar replacement of an amino acid with a structurally related amino acid (*i.e.*, conservative mutations) will not have a major effect on the biological activity of the resulting molecule. Accordingly, some embodiments of the present invention provide variants of Xrn1p disclosed herein containing conservative replacements. Conservative replacements are those that take place within a family of amino acids that are related in their side chains. Genetically encoded amino acids can be divided into four families: (1) acidic (aspartate, glutamate); (2) basic (lysine, arginine, histidine); (3) nonpolar (alanine, valine, leucine, isoleucine,

IDT Exhibit 1017.1989

proline, phenylalanine, methionine, tryptophan); and (4) uncharged polar (glycine, asparagine, glutamine, cysteine, serine, threonine, tyrosine). Phenylalanine, tryptophan, and tyrosine are sometimes classified jointly as aromatic amino acids. In similar fashion, the amino acid repertoire can be grouped as (1) acidic (aspartate, glutamate); (2) basic (lysine, arginine, histidine), (3) aliphatic (glycine, alanine, valine, leucine, isoleucine, serine, threonine), with serine and threonine optionally be grouped separately as aliphatic-hydroxyl; (4) aromatic (phenylalanine, tyrosine, tryptophan); (5) amide (asparagine, glutamine); and (6) sulfur -containing (cysteine and methionine) (*e.g.*, Stryer ed., *Biochemistry*, pg. 17-21, 2nd ed, WH Freeman and Co., 1981). It can be readily determined whether a change in the amino acid sequence of a peptide results in a functional polypeptide by assessing the ability of the variant peptide to function in a fashion similar to the wild-type protein. Peptides having more than one replacement can readily be tested in the same manner.

More rarely, a variant includes "nonconservative" changes (*e.g.*, replacement of a glycine with a tryptophan). Analogous minor variations can also include amino acid deletions or insertions, or both. Guidance in determining which amino acid residues can be substituted, inserted, or deleted without abolishing biological activity can be found using computer programs (*e.g.*, LASERGENE software, DNASTAR Inc., Madison, Wis.).

Variants may be produced by methods such as directed evolution or other techniques for producing combinatorial libraries of variants, described in more detail below. In still other embodiments of the present invention, the nucleotide sequences of the present invention may be engineered in order to alter an Xrn1p coding sequence including, but not limited to, alterations that modify the cloning, processing, localization, secretion, and/or expression of the gene product. For example, mutations may be introduced using techniques that are well known in the art (*e.g.*, site-directed mutagenesis to insert new restriction sites, alter glycosylation patterns, or change codon preference, etc.).

Still other embodiments of the present invention provide mutant or variant forms of Xrn1p. It is possible to modify the structure of a peptide having an activity of Xrn1p for such purposes as enhancing activity, or stability (*e.g.*, *ex vivo* shelf life, and/or

IDT Exhibit 1017.1990

resistance to proteolytic degradation *in vivo*). Such modified peptides are considered functional equivalents of peptides having an activity of the subject Xrn1p proteins as defined herein. A modified peptide can be produced in which the amino acid sequence has been altered, such as by amino acid substitution, deletion, or addition.

Moreover, as described above, variant forms (*e.g.*, mutants) of the subject Xrn1p proteins are also contemplated as being equivalent to those peptides and DNA molecules that are set forth in more detail. For example, as described above, the present invention encompasses mutant and variant proteins that contain conservative or non-conservative amino acid substitutions.

This invention further contemplates a method of generating sets of combinatorial mutants of the present Xrn1p proteins, as well as truncation mutants, and is especially useful for identifying potential variant sequences (*i.e.*, mutants) that are functional in exoribonuclease activity. The purpose of screening such combinatorial libraries is to generate, for example, novel Xrn1p variants that have improved or altered exoribonuclease activity.

Therefore, in some embodiments of the present invention, Xrn1p variants are engineered by the present method to provide altered (*e.g.*, increased or decreased) exoribonuclease activity. In other embodiments, Xrn1p variants are engineered to provide heat-stable (i.e., "thermostable") or heat-labile exoribonuclease activity for particular applications. In other embodiments of the present invention, combinatorially-derived variants are generated which have substrate variability different than that of a naturally occurring Xrn1p. Such proteins, when expressed from recombinant DNA constructs, find use in the methods described herein.

Still other embodiments of the present invention provide Xrn1p variants that have intracellular half-lives dramatically different than the corresponding wild-type protein. For example, the altered protein can be rendered either more stable or less stable to proteolytic degradation or other cellular process that result in destruction of, or otherwise inactivate Xrn1p. Such variants, and the genes which encode them, can be utilized to alter the location of Xrn1p expression by modulating the half-life of the protein. For instance, a short half-life can give rise to more transient Xrn1p biological effects and,

IDT Exhibit 1017.1991

when part of an inducible expression system, can allow tighter control of Xrn1p levels within the cell.

In still other embodiments of the present invention, Xrn1p variants are generated by the combinatorial approach to act as antagonists, in that they are able to interfere with the ability of the corresponding wild-type protein to regulate cell function.

In some embodiments of the combinatorial mutagenesis approach of the present invention, the amino acid sequences for a population of Xrn1p homologs, variants or other related proteins are aligned, preferably to promote the highest homology possible. Such a population of variants can include, for example, Xrn1p homologs from one or more species, or Xrn1p variants from the same species but which differ due to mutation or polymorphisms. Amino acids that appear at each position of the aligned sequences are selected to create a degenerate set of combinatorial sequences.

In a preferred embodiment of the present invention, the combinatorial Xrn1p library is produced by way of a degenerate library of genes encoding a library of polypeptides which each include at least a portion of potential Xrn1p protein sequences. For example, a mixture of synthetic oligonucleotides can be enzymatically ligated into gene sequences such that the degenerate set of potential Xrn1p sequences are expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (*e.g.*, for phage display) containing the set of Xrn1p sequences therein.

There are many ways by which the library of potential Xrn1p homologs and variants can be generated from a degenerate oligonucleotide sequence. In some embodiments, chemical synthesis of a degenerate gene sequence is carried out in an automatic DNA synthesizer, and the synthetic genes are ligated into an appropriate gene for expression. The purpose of a degenerate set of genes is to provide, in one mixture, all of the sequences encoding the desired set of potential Xrn1p sequences. The synthesis of degenerate oligonucleotides is well known in the art (*See e.g.*, Narang, Tetrahedron Lett., 39:39, 1983; Itakura *et al.*, Recombinant DNA, *in* Walton (ed.), *Proceedings of the 3rd Cleveland Symposium on Macromolecules,* Elsevier, Amsterdam, pp 273-289, 1981; Itakura *et al.*, Annu. Rev. Biochem., 53:323, 1984; Itakura *et al.*, Science 198:1056, 1984; Ike *et al.*, Nucl. Acid Res., 11:477, 1983). Such techniques have been employed in the directed evolution of other proteins (*See e.g.*, Scott *et al.*, Science 249:386 [1980];

IDT Exhibit 1017.1992

Roberts *et al.*, Proc. Natl. Acad. Sci. USA 89:2429 [1992]; Devlin *et al.*, Science 249: 404 [1990]; Cwirla *et al.*, Proc. Natl. Acad. Sci. USA 87: 6378 [1990]; as well as U.S. Pat. Nos. 5,223,409, 5,198,346, and 5,096,815; each of which is incorporated herein by reference).

It is contemplated that the Xrn1p nucleic acids (*e.g.*, SEQ ID NO:2, and fragments and variants thereof) can be utilized as starting nucleic acids for directed evolution. These techniques can be utilized to develop Xrn1p variants having desirable properties such as increased, decreased, or altered exoribonuclease activity.

In some embodiments, artificial evolution is performed by random mutagenesis (*e.g.*, by utilizing error-prone PCR to introduce random mutations into a given coding sequence). This method requires that the frequency of mutation be finely tuned. As a general rule, beneficial mutations are rare, while deleterious mutations are common. This is because the combination of a deleterious mutation and a beneficial mutation often results in an inactive enzyme. The ideal number of base substitutions for targeted gene is usually between 1.5 and 5 (Moore and Arnold, Nat. Biotech., 14, 458, 1996; Eckert and Kunkel, PCR Methods Appl., 1:17-24, 1991; Caldwell and Joyce, PCR Methods Appl., 2:28, 1992; and Zhao and Arnold, Nuc. Acids Res. 25:1307, 1997). After mutagenesis, the resulting clones are selected for desirable activity (*e.g.*, screened for Xrn1p activity). Successive rounds of mutagenesis and selection are often necessary to develop enzymes with desirable properties. It should be noted that only the useful mutations are carried over to the next round of mutagenesis.

In other embodiments of the present invention, the polynucleotides of the present invention are used in gene shuffling or sexual PCR procedures (*e.g.*, Smith, Nature, 370:324, 1994; U.S. Pat. Nos. 5,837,458; 5,830,721; 5,811,238; 5,733,731; all of which are herein incorporated by reference). Gene shuffling involves random fragmentation of several mutant DNAs followed by their reassembly by PCR into full length molecules. Examples of various gene shuffling procedures include, but are not limited to, assembly following DNase treatment, the staggered extension process (STEP), and random priming in vitro recombination. In the DNase mediated method, DNA segments isolated from a pool of positive mutants are cleaved into random fragments with DNaseI and subjected to multiple rounds of PCR with no added primer. The lengths of random fragments

IDT Exhibit 1017.1993

approach that of the uncleaved segment as the PCR cycles proceed, resulting in mutations present in different clones becoming mixed and accumulating in some of the resulting sequences. Multiple cycles of selection and shuffling have led to the functional enhancement of several enzymes (Stemmer, Nature, 370:398, 1994; Stemmer, Proc. Natl. Acad. Sci. USA, 91:10747, 1994; Crameri *et al.*, Nat. Biotech., 14:315, 1996; Zhang *et al.*, Proc. Natl. Acad. Sci. USA, 94:4504, 1997; and Crameri *et al.*, Nat. Biotech., 15:436, 1997).

A wide range of techniques are known in the art for screening gene products of combinatorial libraries made by point mutations, and for screening cDNA libraries for gene products having a certain property. Such techniques will be generally adaptable for rapid screening of the gene libraries generated by the combinatorial mutagenesis or recombination of Xrn1p homologs or variants. The most widely used techniques for screening large gene libraries typically comprise cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates relatively easy isolation of the vector encoding the gene whose product was detected.

Homologues and similar proteins that find use in the compositions and methods of the present invention include proteins and extracts containing proteins from a variety of organisms. For example, Rat1p/5' exonuclease 2 and homologues DHp1p and Dhm1p in mouse and humans may be used in the compositions and methods of the present invention. Work with crude extracts shows a homologue in Xenopus (see e.g., Furuichi et al., Nature 266:235, 1977) and wheat germ (see e.g., Shimotohno et al., Proc. Natl. Acad. Sci. USA 74:2734, 1977). Stevens et al. showed purification of a 5' exoribonuclease from human placental nuclei (Nucl. Acids Res., 15:695, 1987). Heyer et al., show that antibodies against the yeast Xrn1 protein also interact with the homologue protein in Drosophila, Xenopus, and mouse (Mol. Cell. Biol., 15:2728, 1995).

Fragments of the nucleic acids and proteins of the present invention may also be used, so long as the fragments encode or possess the desired enzymatic activity. Page et al. (Nucl. Acids Res., 26:3707, 1998), herein incorporated by reference in its entirety,

IDT Exhibit 1017.1994

provide deletion analysis demonstrating that the C-terminus is dispensable for exonuclease function.

A "unit" of exoribonuclease is the amount of enzyme required to convert 1 μg of *E. coli* ribosomal RNA (i.e., 16S and/or 23S rRNA) to acid-soluble form in sixty minutes at 30 °C under standard reaction assay conditions (e.g., 50 mM Tris-HCl, pH 8.0, 2 mM $MgCl_2$, and 0.1 M NaCl).

The symbol "S" with respect to a ribosomal RNA, such as "16S" or "23S" rRNA, or another macromolecule, designates a "svedberg unit" (named for Theodor Svedberg), which is a unit used to measure the sedimentation rate of a molecule in a colloidal suspension and which indicates the molecular weight of the molecule, with a svedberg unit being equal to 10^{-13} second.

The term "gene" as used herein, refers to a DNA sequence that comprises control and coding sequences necessary for the production of a polypeptide or protein precursor or other encoded molecule (e.g., rRNA, tRNA). The polypeptide can be encoded by a full-length coding sequence or by any portion of the coding sequence, as long as the desired protein activity is retained.

"Nucleoside", as used herein, refers to a compound consisting of a purine (guanine (G) or adenine (A)) or pyrimidine (thymine (T), uridine (U), or cytidine (C)) base covalently linked to a pentose sugar, whereas "nucleotide" refers to a nucleoside phosphorylated at one of the hydroxyl groups of the pentose sugar.

A "cap" is a modified guanine nucleotide that is joined to the 5'-end of the 5' nucleotide of a primary eukaryotic mRNA molecule. A common cap (sometimes referred to as "standard cap") has a structure designated as m^7G[5']ppp[5']N, in which "p" represents a phosphate group, "G" represents a guanosine nucleoside, "m^7" represents a methyl group on the 7-position of guanine (in this example, on the guanine base of the cap nucleotide), and "[5']" indicates the position at which the "p" is joined to the ribose of each respective nucleoside (in this example, indicating that the 5'-carbon of the ribose of the cap is joined via a triphosphate group to the 5'-carbon of the ribose of the first nucleoside ("N") of the mRNA. In addition to this "standard" cap, a variety of other naturally-occurring and synthetic cap analogs are known in the art.

IDT Exhibit 1017.1995

A "nucleic acid" or a "polynucleotide", as used herein, is a covalently linked sequence of nucleotides in which the 3' position of the sugar moiety of one nucleotide is joined by a phosphodiester group to the 5' position of the sugar moiety of the next nucleotide, and in which the nucleotide residues (bases) are linked in specific sequence; i.e., a linear order of nucleotides. An "oligonucleotide", as used herein, is a short polynucleotide or a portion of a polynucleotide. An oligonucleotide typically contains a sequence of about two to about one hundred bases, although longer molecules may also be encompassed by this term. The word "oligo" is sometimes used in place of the word "oligonucleotide".

An "oligonucleotide sequence tag template" or a "splice template oligo" or a "template switch oligo" is an oligonucleotide that complexes with a single-stranded target nucleic acid and is used as a template to extend the 3'-terminus of one or more target sequences in order to add a specific sequence or a "sequence tag". The 3'-portion of an oligonucleotide sequence tag template or a splice template is sufficiently complementary to the 3'-terminus of the target sequence that is to be extended to anneal thereto. A DNA- or RNA-dependent DNA polymerase is then used to extend the target nucleic acid molecule using the sequence in the 5'-portion of the oligonucleotide sequence tag template or splice template oligo as a template. The extension product of the primer-extended molecule has the specific sequence at its 3'-terminus that is complementary to the sequence in the 5'-portion of the oligonucleotide sequence tag template or splice template oligo. With respect to the present invention, an oligonucleotide sequence tag template can be used to add a sequence tag to the 3'-end of a nucleic acid, including an RNA or DNA molecule, as described in U.S. Patent Application No. 2005/015333 of Roy R. Sooknanan, which is incorporated herein by reference.

One embodiment of an oligonucleotide sequence tag template or splice template oligo of the present invention is a "promoter sequence tag template" or "promoter splice template oligo." A promoter sequence tag template or promoter splice template oligo comprises a sequence in its 3'-portion that is sufficiently complementary to the 3'-end of the target sequence to anneal thereto and a sequence in it 5'-portion that is complementary to a sequence comprising a single-stranded transcription promoter of the invention. Thus, a promoter sequence tag template or promoter splice template oligo can provide a

IDT Exhibit 1017.1996

template for synthesis of a sequence comprising a transcription promoter at the 3'-end of first-strand cDNA obtained either by RNA-dependent DNA polymerase (or "reverse transcriptase") primer extension of a target sequence comprising an RNA target nucleic acid, such as, but not limited to, an mRNA target, or by DNA polymerase primer extension of a target sequence comprising DNA or RNA.

A "ligation splint oligo" or "ligation splint" is an oligo that is used to provide an annealing site or a "ligation template" for joining two ends of one nucleic acid (i.e., "intramolecular joining") or two ends of two nucleic acids (i.e., "intermolecular joining") using a ligase or another enzyme with ligase activity. The ligation splint holds the ends adjacent to each other and "creates a ligation junction" between the 5'-phosphorylated and a 3'-hydroxylated ends that are to be ligated.

Nucleic acid molecules are said to have a "5'-terminus" (5' end) and a "3'-terminus" (3' end) because nucleic acid phosphodiester linkages occur at the 5' carbon and 3' carbon of the sugar moieties of the substituent mononucleotides. The end of a polynucleotide at which a new linkage would be to a 5' carbon is its 5' terminal nucleotide. The end of a polynucleotide at which a new linkage would be to a 3' carbon is its 3' terminal nucleotide. A terminal nucleotide, as used herein, is the nucleotide at the end position of the 3'- or 5'-terminus.

Nucleic acid molecules are said to have "5' ends" and "3' ends" because mononucleotides are joined in one direction via a phosphodiester linkage to make oligonucleotides, in a manner such that a phosphate on the 5'-carbon of one mononucleotide sugar moiety is joined to an oxygen on the 3'-carbon of the sugar moiety of its neighboring mononucleotide. Therefore, an end of an oligonucleotide referred to as the "5' end" if its 5' phosphate is not linked to the oxygen of the 3'-carbon of a mononucleotide sugar moiety and as the "3' end" if its 3' oxygen is not linked to a 5' phosphate of the sugar moiety of a subsequent mononucleotide.

Polypeptide molecules are said to have an "amino terminus" (N-terminus) and a "carboxy terminus" (C-terminus) because peptide linkages occur between the backbone amino group of a first amino acid residue and the backbone carboxyl group of a second amino acid residue.

IDT Exhibit 1017.1997

As used herein, the terms "complementary" or "complementarity" are used in reference to a sequence of nucleotides related by the base-pairing rules. For example, the sequence 5'-A-G-T-3', is complementary to the sequence 3'-T-C-A-5'. Complementarity may be "partial," in which only some of the nucleic acids' bases are matched according to the base pairing rules. Or, there may be "complete" or "total" complementarity between the nucleic acids. The degree of complementarity between nucleic acid strands has significant effects on the efficiency and strength of hybridization between nucleic acid strands. This is of particular importance in amplification reactions, as well as detection methods that depend upon hybridization of nucleic acids.

As used herein, the term "probe" refers to an oligonucleotide (i.e., a sequence of nucleotides), whether occurring naturally (e.g., as in a purified restriction digest) or produced synthetically, recombinantly or by PCR amplification, which is capable of hybridizing to another oligonucleotide of interest. A probe may be single-stranded or double-stranded (e.g., and rendered single-stranded or partially single-stranded in use). Probes are useful in the detection, identification and isolation of particular gene sequences. It is contemplated that a probe used in the present invention can be labeled with any "reporter molecule," so that it is detectable in a detection system, including, but not limited to enzyme (i.e., ELISA, as well as enzyme-based histochemical assays), visible, fluorescent, radioactive, and luminescent systems. It is not intended that the present invention be limited to any particular detection system or label. The terms "reporter molecule" and "label" are used herein interchangeably. In addition to probes, primers and deoxyribonucleoside triphosphates may contain labels; these labels may comprise, but are not limited to, ^{32}P, ^{33}P, ^{35}S, enzymes, or visible, luminescent, or fluorescent molecules (e.g., fluorescent dyes).

The term "recombinant protein" or "recombinant polypeptide" as used herein refers to a protein molecule expressed from a recombinant DNA molecule. In contrast, the term "native protein" is used herein to indicate a protein isolated from a naturally occurring (i.e., a nonrecombinant) source. Molecular biological techniques may be used to produce a recombinant form of a protein with identical or similar properties as compared to the native form of the protein. Variants of the native sequence may also be

IDT Exhibit 1017.1998

made to, for example, improve expression, purification, or other desired properties of the polypeptide.

As used herein in reference to an amino acid sequence or a protein, the term "portion" (as in "a portion of an amino acid sequence") refers to fragments of that protein. The fragments may range in size from four amino acid residues to the entire amino acid sequence minus one amino acid. Particularly preferred fragments retain one or more of the enzymatic activities associated with a whole protein (e.g., 5' exoribonuclease activity).

As used herein, the term "fusion protein" refers to a chimeric protein containing the protein of interest (e.g., 5' exoribonuclease 1 and fragments thereof) joined to an exogenous protein fragment (e.g., the fusion partner which contains a non-xrn-1 protein). The fusion partner may enhance the solubility of xrn-1 protein as expressed in a host cell, may provide an affinity tag to allow purification of the recombinant fusion protein from the host cell or culture supernatant, or both. If desired, the fusion protein may be removed from the protein of interest (e.g., xrn-1 or fragments thereof) by a variety of enzymatic or chemical means know to the art.

The term "5' to 3' exonuclease activity" refers to the presence of an activity in a protein that is capable of removing nucleotides from the 5' end of an oligonucleotide. 5' to 3' exonuclease activity may be measured using any of the assays provided herein or known in the art. Some enzymes contain 5' to 3' "exoribonuclease" activity, the ability to remove nucleotides from the 5' end of ribonucleic acid molecules. Some enzymes that are traditionally called "exoribonucleases" may also have 5'-exonuclease activity against DNA substrates. For example, enzymes of the present invention, while historically named "exoribonucleases", have 5'-exonuclease activity on DNA substrates that have a 5'-phosphate, meaning a phosphate group on the 5'-position of the deoxyribose sugar of the nucleotide at the 5'-end of the DNA. Thus, it is understood that such exoribonucleases (e.g., Xrn1 5' exoribonuclease 1), unless specified otherwise, have exonuclease activity for both RNA and DNA substrates.

The terms "cell," "cell line," "host cell," as used herein, are used interchangeably, and all such designations include progeny or potential progeny of these designations. The words "transformants" or "transformed cells" include the primary transformed cells

IDT Exhibit 1017.1999

derived from that cell without regard to the number of transfers. All progeny may not be precisely identical in DNA content, due to deliberate or inadvertent mutations. Nonetheless, mutant progeny that have the same functionality as screened for in the originally transformed cell are included in the definition of transformants.

Nucleic acids are known to contain different types of mutations. A "point" mutation refers to an alteration of a base nucleotide at a single nucleotide position in the sequence compared to the wild type sequence. Mutations may also refer to insertion or deletion of one or more bases, so that the nucleic acid sequence differs from the wild-type sequence.

The term "homology" refers to a degree of complementarity of one nucleic acid sequence with another nucleic acid sequence. There may be partial homology or complete homology (i.e., complementarity). A partially complementary sequence is one that at least partially inhibits a completely complementary sequence from hybridizing to a target nucleic acid and is referred to using the functional term "substantially homologous." The inhibition of hybridization of the completely complementary sequence to the target sequence may be examined using a hybridization assay (Southern or Northern blot, solution hybridization and the like) under conditions of low stringency. A substantially homologous sequence or probe will compete for and inhibit the binding (i.e., the hybridization) of a completely homologous sequence to a target under conditions of low stringency. This is not to say that conditions of low stringency are such that non-specific binding is permitted; low stringency conditions require that the binding of two sequences to one another be a specific (i.e., selective) interaction. The absence of non-specific binding may be tested by the use of a second target that lacks complementarity or that has only a low degree of complementarity (e.g., less than about 30% complementarity). In the case in which specific binding is low or non-existent, the probe will not hybridize to a nucleic acid target.

When used in reference to a double-stranded nucleic acid sequence such as a cDNA or a genomic clone, the term "substantially homologous" refers to any probe which can hybridize to either or both strands of the double-stranded nucleic acid sequence under conditions of low stringency as described herein.

IDT Exhibit 1017.2000

As used herein, the terms "hybridization" or "annealing" are used in reference to the pairing of complementary nucleic acid strands. Hybridization and the strength of hybridization (i.e., the strength of the association between nucleic acid strands) is impacted by many factors well known in the art including the degree of complementarity between the nucleic acids, stringency of the conditions involved affected by such conditions as the concentration of salts, the T_m (melting temperature) of the formed hybrid, the presence of other components (e.g., the presence or absence of polyethylene glycol or betaine), the molarity of the hybridizing strands and the G:C content of the nucleic acid strands.

The term "isolated" when used in relation to a nucleic acid, as in "isolated polynucleotide " or "isolated oligonucleotide " refers to a nucleic acid sequence that is identified and separated from at least one contaminant with which it is ordinarily associated in its source. Thus, an isolated nucleic acid is present in a form or setting that is different from that in which it is found in nature. In contrast, non-isolated nucleic acids (e.g., DNA and RNA) are found in the state they exist in nature. For example, a given DNA sequence (e.g., a gene) is found on the host cell chromosome in proximity to neighboring genes; and a specific RNA sequences (e.g., a specific mRNA sequence encoding a specific protein), is found in the cell as a mixture with numerous other mRNAs that encode a multitude of proteins. The isolated polynucleotide or nucleic acid or oligonucleotide may be present in single-stranded or double-stranded form. When an isolated polynucleotide or nucleic acid is to be utilized to express a protein, the polynucleotide contains at a minimum, the sense or coding strand (i.e., the polynucleotide may be single-stranded), but may contain both the sense and anti-sense strands (i.e., the polynucleotide may be double-stranded).

As used herein, the term "purified" or "to purify" means the result of any process that removes some of a contaminant from the component of interest, such as a protein or nucleic acid. The percent of a purified component is thereby increased in the sample.

As used herein, the term "enzyme" refers to molecules or molecule aggregates that are responsible for catalyzing chemical and biological reactions. Such molecules are typically proteins, but can also comprise short peptides, RNAs, ribozymes, antibodies,

IDT Exhibit 1017.2001

and other molecules. A molecule that catalyzes chemical and biological reactions is referred to as "having enzyme activity" or "having catalytic activity."

In addition to 5'-phosphate-dependent exonucleases, a variety of other enzymes are used in the methods of the present invention. In general, the invention comprises use of any enzyme from any source that has an activity that is equivalent to an enzyme referred to in a method so long as the enzyme functions in the same way in the particular method. By way of example, but not of limitation, an RNA-dependent DNA polymerase can comprise an AMV reverse transcriptase; an MMLV reverse transcriptase; SuperScript I, SuperScript II, SuperScript III, or AMV ThermoScript reverse transcriptase (INVITROGEN); or MonsterScript reverse transcriptase (EPICENTRE), or it can comprise another enzyme that has similar activity in the method; a polynucleotide kinase can comprise T4 polynucleotide kinase or it can comprise another enzyme that has similar activity in the method; a poly(A) polymerase can comprise an *E. coli* poly(A) polymerase encoded by the *pcn*B gene or it can comprise another enzyme that has similar activity in the method; a ribonuclease H can comprise *E. coli* RNase H or a *Thermus* RNase H (e.g., Hybridase RNase H from EPICENTRE) or it can comprise another enzyme that has similar activity in the method; a pyrophosphatase can be tobacco acid pyrophosphatase or it can comprise another enzyme that has similar activity in the method; and a DNA phosphatase can comprise APex Alkaline Phosphatase (EPICENTRE) or shrimp alkaline phosphatase or it can comprise another enzyme that has similar activity in the method. Methods and reaction conditions for use of these and other enzymes that can be used are well known in the art.

An "RNA amplification method" according to the present invention is a method that synthesizes an RNA product and that results in an increase in the number of copies of an RNA sequence or of its complementary sequence compared to the number of copies of the sequence present in a sample. A number of RNA amplification methods are known in the art that comprise steps for joining an RNA polymerase promoter to either the 5'-end of the cDNA or the 3'-end of cDNA that is made by reverse transcription of mRNA using an RNA-dependent DNA polymerase to extend a primer, such as, but not limited to an oligo(dT)-containing primer, which methods result in synthesis of amplified antisense RNA or amplified sense RNA, respectively, by subsequent *in vitro* transcription (e.g.,

IDT Exhibit 1017.2002

using T7, T3, or SP6 RNA polymerase). Kits and methods of the present invention that comprise RNA amplification can comprise any compositions or method, including without limitation, any compositions or methods known in the art. By way of example, but not of limitation, a method that uses an oligo(dT) promoter primer to synthesize an antisense RNA (aRNA), such as methods described by Van Gelder, R.N., et al. (Proc. Natl. Acad. Sci. USA 87: 1663, 1990) can be used; such methods are sometimes referred to as "Eberwine-type" amplification methods herein. Kits for this purpose are commercially available and can be used, including 1-round and 2-round amplification kits such as various 1-round and 2-round TargetAmp™ Aminoallyl-aRNA Amplification Kits or TargetAmp™ aRNA Amplification Kits available from EPICENTRE, or MessageAmp kits available from Ambion (Austin, TX). Still further, a sense RNA amplification method that attaches a promoter to the 3'-end of first-strand cDNA for synthesis of amplified sense RNA, such as methods described by Che, S. and Ginsberg, S.D (Laboratory Investigation 84: 131, 2004) and by Ginsberg and Che (Neurochemical Res. 27: 981, 2002; and PCT Internatl. Publication No. WO 02/065093), can be used in a method or kit of the present invention. The BD SMART mRNA amplification kit (BD Biosciences Clontech, Palo Alto, CA) is an example of a commercial kit of a type that can be used for sense RNA amplification in methods and kits of the present invention. Still another RNA amplification method that can be used in the methods of the present invention uses a oligonucleotide sequence tag template to add a tag sequence to the 3'-end of a target sequence comprising mRNA or first-strand cDNA, which tag sequence can be further used to add a promoter for synthesis of RNA corresponding to the target sequence, as described in U.S. Patent Application No. 2005/015333 of Roy R. Sooknanan, which is incorporated herein by reference. Examples of other methods for RNA amplification that can be used are described in: Murakawa et al., DNA 7:287-295, 1988; Phillips and Eberwine, Methods in Enzymol. Suppl. 10:283-288, 1996; Ginsberg et al., Ann. Neurol. 45:174-181, 1999; Ginsberg et al., Ann. Neurol. 48:77-87, 2000; VanGelder et al., Proc. Natl. Acad. Sci. USA 87:1663-1667, 1990; Eberwine et al., Proc. Natl. Acad. Sci. USA 89:3010-3014, 1992; U.S. Patent Nos. 5,021,335; 5,168,038; 5,545,522; 5,514,545; 5,716,785; 5,891,636; 5,958,688; 6,291,170; and PCT Patent

IDT Exhibit 1017.2003

Applications WO 00/75356 and WO 02/065093, which is incorporated herein by reference.

The present invention is also not limited to RNA amplification methods that require synthesis of double-stranded cDNA. By way of example, but without limitation, the present invention also comprises RNA amplification methods and compositions as described in U.S. Patent Appln. No. 2004/0171041 that use an RNA polymerase that can synthesize RNA using single-stranded templates that are functionally joined to a single-stranded promoter, such as, but not limited to methods that use MiniV RNA polymerase (available from EPICENTRE in the MiniV™ *In Vitro* Transcription Kit); in these embodiments, a single-stranded promoter is joined to either the 5'-end of the cDNA or the 3'-end of cDNA that is made by reverse transcription of mRNA using an RNA-dependent DNA polymerase to extend a primer, resulting in synthesis of amplified antisense RNA or amplified sense RNA, respectively, by subsequent *in vitro* transcription of single-stranded DNA templates (e.g., using MiniV RNA polymerase).

A "T7-type RNA polymerase" as defined herein is a wild-type or mutant form of an RNA polymerase derived from a T7-type bacteriophage, including both phage-encoded enzymes and enzymes obtained by cloning the RNA polymerase gene in a DNA vector and expressing it in a bacterial or other cell. This is based on the fact that the genetic organization of all T7-type bacteriophage that have been examined has been found to be essentially the same as that of T7. Examples of T7-type bacteriophages according to the invention include, but are not limited to Escherichia coli phages T3, phi I, phi II, W31, H, Y, A1, 122, cro, C21, C22, and C23; Pseudomonas putida phage gh-1; Salmonella typhimurium phage SP6; Serratia marcescens phages IV; Citrobacter phage ViIII; and Klebsiella phage No. 11 (Hausmann, Current Topics in Microbiology and Immunology 75:77-109, 1976; Korsten et al., J. Gen. Virol. 43:57-73, 1975; Dunn, et al., Nature New Biology 230:94-96, 1971; Towle, et al., J. Biol. Chem. 250:1723-1733, 1975; Butler and Chamberlin, J. Biol. Chem. 257:5772-5778, 1982). Mutant RNAPs (Sousa et al., U.S. Patent No. 5,849,546; Padilla, R and Sousa, R, Nucleic Acids Res., 15: e138, 2002; Sousa, R and Mukherjee, S, Prog Nucleic Acid Res Mol Biol., 73: 1-41, 2003), such as, but not limited to, T7 RNAP Y639F mutant enzyme, T3 RNAP Y573F mutant enzyme, SP6 RNAP Y631F mutant enzyme, T7 RNAP having altered amino

IDT Exhibit 1017.2004

acids at both positions 639 and 784, T3 RNAP having altered amino acids at both positions 573 and 785, or SP6 RNAP having altered amino acids at both positions 631 and 779 can also be used in some embodiments of methods or assays of the invention. In particular, such mutant enzymes can corporate dNTPs and 2'-F-dNTPs, in addition to ddNTPs and certain other substrates, which are advantageous for synthesis of RNA molecules with specific properties and uses. Phage N4 mini-vRNAP, which has certain domains in common with T7-type RNAPs, and which is also an RNA polymerase of the present invention, is a transcriptionally active 1,106-amino acid domain of the N4 vRNAP, which corresponds to amino acids 998-2103 of N4 vRNAP (Kazmierczak, K.M., et al., EMBO J., 21: 5815-5823, 2002; U.S. Patent Application No. 20030096349, incorporated herein by reference) can be used. Alternatively, an N4 mini-vRNAP Y678F mutant enzyme (U.S. Patent Application No. 20030096349) can incorporate non-canonical nucleotides such as 2'-F-dNTPs. In order to carry out transcription, a RNA polymerase recognizes and binds to a DNA sequence of approximately 25 nucleotides in length called an "RNA polymerase promoter," a "transcription promoter" or simply a "promoter," and initiates transcription therefrom. In most cases, the promoter sequence is double-stranded. As used herein, the strand of a double-stranded promoter that is covalently joined to the template strand for synthesis of RNA is defined as the "sense strand" or "sense promoter sequence" and its complement is defined as the "anti-sense strand" or the "anti-sense promoter sequence."

Still further, the term "RNA amplification method" according to the present invention also includes other methods, such as but not limited to methods described in U.S. Pat. Appln. No. 2004/0180361 of Dahl et al.; PCT Patent Publication Nos. WO 02/16639; WO 00/56877; and AU 00/29742 of Takara Shuzo Company; and in U.S. Pat. No. 6,251,639 and U.S. Pat. Appln. Nos. 2001/0034048; 2003/0017591; 2003/0087251; and 2003/0186234 of Kurn (all incorporated herein by reference), even if said methods comprise synthesis DNA products rather than RNA products corresponding to mRNA in the sample. Thus, the present invention also comprises embodiments in which a nucleic acid of interest (or a desired nucleic acid) obtained following treatment with a 5'-phosphate-dependent exonuclease of the present invention is amplified using any of said

IDT Exhibit 1017.2005

RNA amplification methods, including those that comprise synthesis of DNA amplification products.

A nucleic acid of interest (or a desired nucleic acid) obtained following treatment with a 5'-phosphate-dependent exonuclease of the present invention can also be amplified by PCR or reverse transcription-PCR (i.e., RT-PCR), including without limitation real-time PCR or real-time RT-PCR. Amplification of an mRNA of interest or a cDNA of interest by RT-PCR or PCR, respectively, is useful for validation of results of gene expression analysis using other methods, such as but not limited to, gene expression analysis results obtained by hybridization of labeled RNA or DNA corresponding to mRNA from a sample to nucleic acid arrays or microarrays.

"Rolling Circle Replication" and "Rolling Circle Transcription" are two other methods that can be used to amplify a nucleic acid of interest in a method of the present invention, which methods are defined and described in a number of patents and publications below, which are incorporated herein by reference. In PCT Patent Application No. WO 92/01813, Ruth and Driver disclosed a process for synthesizing circular single-stranded nucleic acids by hybridizing a linear polynucleotide to a complementary oligonucleotide and then ligating the linear polynucleotide. They further disclosed a process for generating multiple linear complements of the circular single-stranded nucleic acid template by extending a primer more than once around the circular template using a DNA polymerase. Japanese Patent Nos. JP4304900 and JP4262799 of Aono Toshiya et al. disclose detection of a target sequence by ligation of a linear single-stranded probe having target-complementary 3'- and 5'-end sequences which are adjacent when the linear probe is annealed to a target sequence in the sample, followed by either rolling circle replication or *in vitro* transcription of the circular single-stranded template. The inventors disclose that in vitro transcription is performed by first annealing to the circular single-stranded template a complementary nucleotide primer having an anti-promoter sequence in order to form a double-stranded promoter, and then transcribing the circular single-stranded template having the annealed anti-promoter primer with an RNA polymerase that has helicase-like activity, such as T7, T3 or SP6 RNA polymerase. In U.S. Patent Nos. 6,344,329; 6,210,884; 6,183,960; 5,854,033; 6,329,150; 6,143,495; 6,316,229; and 6,287,824, Paul M. Lizardi also used rolling circle replication to amplify

IDT Exhibit 1017.2006

and detect nucleic acid sequences. In a series of articles and patents, Eric Kool and co-workers disclosed synthesis of DNA or RNA multimers, meaning multiple copies of an oligomer or oligonucleotide joined end to end (i.e., in tandem) by rolling circle replication or rolling circle transcription, respectively, of a circular DNA template molecule. Rolling circle replication uses a primer and a strand-displacing DNA polymerase, such as phi29 DNA polymerase. With respect to rolling circle transcription, it was shown that circular single-stranded DNA (ssDNA) molecules can be efficiently transcribed by phage and bacterial RNA polymerases (Prakash, G. and Kool, E., J. Am. Chem. Soc. 114: 3523-3527, 1992; Daubendiek, S.L. et al., J. Am. Chem. Soc. 117: 7818-7819, 1995; Liu, D. et al., J. Am. Chem. Soc. 118: 1587-1594, 1996; Daubendiek, S.L. and Kool, E.T., Nature Biotechnol., 15: 273-277, 1997; Diegelman, A.M. and Kool, E.T., Nucleic Acids Res., 26: 3235-3241, 1998; Diegelman, A.M. and Kool, E.T., Chem. Biol., 6: 569-576, 1999; Diegelman, A.M. et al., BioTechniques 25: 754-758, 1998; Frieden, M. et al., Angew. Chem. Int. Ed. Engl. 38: 3654-3657, 1999; Kool, E.T., Acc. Chem. Res., 31: 502-510, 1998; U.S. Patent Nos. 5,426,180; 5,674,683; 5,714,320; 5,683,874; 5,872,105; 6,077,668; 6,096,880; and 6,368,802). Rolling circle transcription of these circular ssDNAs occurs in the absence of primers, in the absence of a canonical promoter sequence, and in the absence of any duplex DNA structure, and results in synthesis of linear multimeric complementary copies of the circle sequence up to thousands of nucleotides in length. Transcription of the linear precursor of the circular ssDNA template yielded only a small amount of RNA transcript product that was shorter than the template. Fire and Xu (U.S. Patent No. 5,648,245; Fire, A. and Xu, S-Q, Proc. Natl. Acad. Sci. USA, 92: 4641-4645, 1995) also disclose methods for using rolling circle replication of small DNA circles to construct oligomer concatamers. Other researchers, including, but not limited to, Mahtani (U.S. Patent No. 6,221,603), Rothberg et al. (U.S. Patent No. 6,274,320), Dean et al. (Genome Res., 11: 1095-1099, 2001), Lasken et al. (U.S. Patent No. 6,323,009), and Nilsson et al. (Nucleic Acids Res., 30(14): e66, 2002) disclose other methods and applications of rolling circle amplification. Pickering et al. (Nucleic Acids Res., 30(12): e60, 2002) discloses a ligation and rolling circle amplification method for homogeneous end-point detection of single nucleotide polymorphisms (SNPs).

IDT Exhibit 1017.2007

As used herein, the terms "buffer" or "buffering agents" refer to materials that when added to a solution, cause the solution to resist changes in pH.

As used herein, the terms "reducing agent" and "electron donor" refer to a material that donates electrons to a second material to reduce the oxidation state of one or more of the second material's atoms.

The term "monovalent salt" refers to any salt in which the metal (e.g., Na, K, or Li) has a net 1+ charge in solution (i.e., one more proton than electron).

As used herein, the term "divalent salt" refers to any salt in which a metal (e.g., Mg, Mn, Ca, or Sr) has a net 2+ charge in solution.

As used herein, the terms "chelator" or "chelating agent" refer to any materials having more than one atom with a lone pair of electrons that are available to bond to a metal ion.

As used herein, the term "solution" refers to an aqueous or non-aqueous mixture.

As used herein, the term "buffering solution" refers to a solution containing a buffering agent.

As used herein, the term "reaction buffer" refers to a buffering solution in which an enzymatic reaction is performed.

As used herein, the term "storage buffer" refers to a buffering solution in which an enzyme is stored.

IDT Exhibit 1017.2008

In keeping with standard polypeptide nomenclature, J. Biol. Chem., 243:3557-3559, 1969, abbreviations for amino acid residues are as shown in the following Table of Correspondence.

TABLE OF CORRESPONDENCE

1-Letter	3-Letter	AMINO ACID
Y	Tyr	tyrosine
G	Gly	glycine
F	Phe	phenylalanine
M	Met	methionine
A	Ala	alanine
S	Ser	serine
I	Ile	isoleucine
L	Leu	leucine
T	Thr	threonine
V	Val	valine
P	Pro	proline
K	Lys	lysine
H	His	histidine
Q	Gln	glutamine
E	Glu	glutamic acid
W	Trp	tryptophan
R	Arg	arginine
D	Asp	aspartic acid
N	Asn	asparagine
C	Cys	cysteine

The phrase "mixed population" or "complex population" refers to any sample containing both desired and undesired nucleic acids. As a non-limiting example, a complex population of nucleic acids may be total genomic DNA, total cellular RNA or a combination thereof. Moreover, a complex population of nucleic acids may have been enriched for a given population, but include other undesirable populations. For example, a complex population of nucleic acids may be a sample that has been enriched for desired messenger RNA (mRNA) sequences but still includes some undesired ribosomal RNA sequences (rRNA).

The term "undetectable levels" of exonuclease activity with respect to a particular nucleic acid substrate refers to levels of digestion of a nucleic acid substrate that are not detectable within the assay or method in which an exonuclease is employed. Undetectable levels include the complete absence of digestion. Although the level of digestion of a nucleic acid of interest (i.e., a desired nucleic acid) can be undetectable in some embodiments, the invention is not limited to embodiments in which digestion of the

IDT Exhibit 1017.2009

nucleic acid of interest by a 5'-phosphate-dependent exonuclease of the invention is undetectable so long as a substantial amount of the undesired nucleic acid is digested and sufficient amount of the desired nucleic acid remains for the intended purpose, which can vary for different purposes. In preferred embodiments of the invention, a substantial amount of the undesired nucleic acid is digested by the 5' exonuclease of the invention, which means that at least 50% (e.g., 60%, 70%, 80%, 90%, 95%, 98%, 99%) of the starting amount (i.e., the amount present prior to treatment with the exonuclease) of the undesired nucleic acid is digested.

The terms "sample" and "specimen" are used in their broadest sense and encompass samples or specimens obtained from any source including biological and environmental sources. As used herein, the term "sample" when used to refer to biological samples obtained from organisms, includes, but it not limited to fluids, solids, tissues, and gases. In preferred embodiments of this invention, biological samples include bodily fluids, isolated cells, fixed cells, cell lysates and the like. However, these examples are not to be construed as limiting the types of samples that find use with the present invention.

A "target nucleic acid" comprises at least one nucleic acid molecule or portion of at least one nucleic acid molecule, whether the molecule or molecules is or are DNA, RNA, or both DNA and RNA, and wherein each the molecule has, at least in part, a defined nucleotide sequence, which is referred to as the "target sequence." A goal of an assay or method of the invention is to detect whether or not the target nucleic acid is present in a sample by means of methods that can detect the target sequence. This is frequently, but without limitation, achieved using a "nucleic acid probe" that is complementary to the target sequence (e.g., a probe that is present on an array or microarray together with other probes that are complementary to other nucleic acid sequences). The target nucleic acid may also have at least partial complementarity with other molecules used in an assay, such as, but not limited to, primers, splice template oligos, ligation splint oligos, capture probes or detection probes. The target nucleic acid may be single- or double-stranded and may be of any length. However, it must comprise a polynucleotide sequence of sufficient sequence specificity and length so as to be useful for its intended purpose. By way of example, but not of limitation, a target nucleic acid

IDT Exhibit 1017.2010

that is to be detected using a sequence-complementary detection probe must have a sequence of sufficient sequence specificity and length so as remain hybridized by the detection probe under assay hybridization conditions wherein sequences that are not target nucleic acids are not hybridized. A target nucleic acid having sufficient sequence specificity and length for an assay of the present invention may be identified, using methods known to those skilled in the art, by comparison and analysis of nucleic acid sequences known for a target and for other sequences which may be present in the sample. For example, sequences for nucleic acids of many viruses, bacteria, humans (e.g., for genes and messenger RNA), and many other biological organisms can be searched using public or private databases, and sequence comparisons, folded structures, and hybridization melting temperatures (i.e., T_m's) may be obtained using computer software known to those knowledgeable in the art. The term "source of target nucleic acid" refers to any sample that contains a naturally occurring target nucleic acid, RNA or DNA.

DETAILED DESCRIPTION OF THE INVENTION

The present invention relates to compositions and methods employing 5'-phosphate-dependent nucleic acid exonucleases. In particular, the present invention provides kits and methods employing 5'-phosphate-dependent nucleic acid exonucleases for a variety of uses where it is desired to digest nucleic acid molecules having 5'-phosphate groups.

As defined herein, any exoribonuclease that selectively digests nucleic acid molecules having 5'-phosphate groups may be used in the methods and kits of the present invention. A 5'-to-3' exoribonuclease of the present invention comprises any 5'-to-3' exonuclease that has greater than 20-fold more 5'-to-3' exonuclease activity for a single-stranded RNA substrate that has a 5'-monophosphorylated terminus than for the same RNA substrate that has a 5'-triphosphorylated or 5'-capped terminus. That is, a 5' exonuclease of the present invention comprises any 5'-to'3' exonuclease that has a relative activity in digesting a particular defined-sequence single-stranded RNA substrate with a 5'-triphosphate or a 5'-cap that is less than 5% of the activity of an RNA substrate having the same sequence but with a 5'-monophosphate. Without limitation, suitable methods that can be used for assaying activity and determining relative activity using RNA

IDT Exhibit 1017.2011

substrates with a 5'-triphosphate, a 5'-cap, or a 5'-monophosphate are described by Stevens and Poole (J. Biol. Chem., 270: 16063, 1995). Preferred embodiments employ Saccharomyces cerevisiae Xrn1p/5' exoribonuclease 1 and/or functional variants and homologues thereof.

In some embodiments, the present invention provides a method of preparing a nucleic acid sample for analysis. It is often desirable to isolate, enrich, or increase the relative percentage of a particular population of sequences within a much larger population of sequences in order to limit analysis to sequences of interest and to reduce interference and unnecessary work that may be caused by the presence of undesirable sequences. The present invention provides a novel method wherein a complex sample or a mixed population of nucleic acids (e.g., containing both nucleic acids of interest and nucleic acids not of interest) is substantially depleted of undesired sequences (e.g., nucleic acids not of interest) and is thus enriched for a population of interest. For example, in preferred embodiments, the present invention provides a method of enriching a nucleic acid population of interest present in a complex sample by treating the sample with a 5'-phosphate-dependent nucleic acid exonuclease, thereby increasing the relative percentage of the nucleic acid population of interest in a given sample for further analysis. In preferred embodiments, the 5'-phosphate-dependent nucleic acid exonuclease digests greater than 50% (e.g., 60%, 70%, 80%, 90%, 95%, 97%, 99%, 99.5%, 99.9%) of the undesired sequence or sequences. Also, in preferred embodiments, the 5'-phosphate-dependent nucleic acid exonuclease digests less than 50%, (e.g., 40%, 30%, 20%, 15%, 10%, 5%, 4%, 3%, 2%, 1%) of the desired sequence or sequences, and most preferably, undetectable amounts of the desired sequence or sequences, but the invention also comprises embodiments in which more of the desired sequence or sequences is/are digested so long as the amount remaining can be used for the intended purpose and a substantial amount of the undesired sequence or sequences is/are digested.

For example, in some embodiments, methods of the present invention enrich for a nucleic acid population of interest (e.g., mRNA sequences) within a mixed population of nucleic acid sequences by targeting undesired sequences (e.g., 16S and 23S prokaryotic rRNA or 18S and 26S or 28S eukaryotic rRNA sequences comprising a 5' phosphate) for digestion by a 5'-phosphate-dependent exonuclease and eliminating them from the mixed

IDT Exhibit 1017.2012

population. In some embodiments, the undesired sequences are degraded by the enzymatic activity of xrn-1 exoribonuclease (See, e.g., Example 1, Table 1).

The mixed population of nucleic acids may be any nucleic acid sample comprising both desired and undesired sequences. The population may include different DNA and/or RNA molecules. In some preferred embodiments, the mixed population is an RNA sample. In further preferred embodiments, the nucleic acid sample is RNA derived from a prokaryotic organism. In other preferred embodiments, the nucleic acid sample comprises RNA from a eukaryotic organism. In still other embodiments, the sample comprises RNA from multiple prokaryotic organisms, or from both prokaryotic and eukaryotic organisms, or from multiple eukaryotic organisms. The mixed population may be derived from a wide variety of sources including for example, tissue samples, blood, isolated cells or environmental samples such as water or soil. The mixed population may be derived from any organism including both eukaryotes and prokaryotes such as human, rat, mouse, Escherichia coli (E. coli), Bacillus subtilis (B. subtilis), Pseudomonas aeruginosa, etc. Methods of deriving nucleic acid samples from eukaryotic and prokaryotic organisms are well known to those of skill in the art (See, e.g., Chapter 4, "Current Protocols in Molecular Biology," Ausubel et al., eds (1997 supplement) Johan Wilen & Sons, Inc. and Chapter 7, Sambrook, Fritsch, Maniatis "Molecular Cloning: A Laboratory Manual" (1989) Cold Spring Harbor Press, etc.).

The population of interest may comprise a subset of the mixed population. The population of interest may include RNA and/or DNA. The population of interest may, for example, be a particular type of RNA. In a preferred embodiment the population of interest is mRNA. The population of interest may comprise any sequence or mixture of sequences and the sequence or mixture of sequences need not be known. The population of interest may be chosen on a variety of bases, including by sequence, function (e.g. messenger RNA (mRNA), ribosomal RNA (rRNA), transfer RNA (tRNA), etc.), or a combination thereof. In some embodiments, the population of interest is provided in a sample that has been depleted of 5S rRNA and/or tRNA. Such a mixture may be generated by salt precipitation using LiCl. The larger rRNAs and mRNAs are precipitated, but the tRNAs and 5S rRNAs remain in solution and can be removed. In some embodiments, the tRNAs and 5S rRNAs are removed by LiCl precipitation after a

IDT Exhibit 1017.2013

sample is treated with exoribonuclease. Kits for conducting such reactions are provided by the present invention (e.g., kits containing LiCl and exoribonuclease). Other methods for isolating subsets of molecules (e.g., PEG precipitation, etc.) may also be employed in the kits and methods of the present invention.

The sequences that are targeted for digestion may comprise undesired sequences in the mixed population. The sequences that are targeted for digestion may comprise any sequence so long as they are distinguishable from the population of interest (e.g., the sequences that are targeted for digestion possesses an accessible 5' phosphate whereas the population of interest does not). Targeted double-stranded DNA (dsDNA) or RNA (dsRNA) or single-stranded DNA (ssDNA) or RNA (ssRNA) that self-anneals to form double-stranded regions can be denatured by various treatments known in the art, such as, but not limited to, treatment with heat or addition of compositions that affect the melting temperature of the targeted polynucleotide, such as, but not limited to betaine. In preferred embodiments, the sequences that are targeted for digestion are RNAs including rRNA (e.g., 16S and 23S prokaryotic rRNA, or 18S and 26S or 28S eukaryotic rRNA). In some embodiments, it may not be necessary to remove substantially all the undesired sequences from the mixed population. In these embodiments it is acceptable to remove only enough of the undesired sequences such that the undesired sequences do not interfere with analysis of the population of interest.

In preferred embodiments, any non-targeted undesirable sequences represent only a small proportion of the mixed population or are readily removed using other methods known in the art. These non-targeted undesirable sequences may include a variety of other nucleic acids such as DNAs, rRNAs, mRNAs or tRNAs. For the sake of simplicity, the presence of non-targeted RNAs will not be discussed throughout the remainder of the application; however, the possibility of their presence is contemplated by the scope of the presently claimed invention, unless described otherwise.

A particular example of the presently claimed invention provides a method of isolating or enriching for mRNAs within a mixed population of RNAs by specifically removing rRNAs >300 nucleotides that possess an accessible 5' phosphate. A mixed population of RNAs may include, for example, mRNAs, tRNAs and rRNAs.

IDT Exhibit 1017.2014

In some embodiments, methods of the present invention are used in combination with methods that expose a 5' phosphate on an undesired nucleic acid sequence (e.g., a method that removes a 5'-triphosphate or a 5' cap on an undesired nucleic acid sequence using an agent, such as, but not limited to, tobacco acid pyrophosphatase).

In other embodiments, methods of the present invention are used in combination with methods that result in addition of a 5'-phosphate to a nucleic acid molecule of interest (e.g., a method for 5'-phosphorylation using a polynucleotide kinase, such as, but not limited to, T4 polynucleotide kinase).

In other embodiments, methods of the present invention are used in combination with methods that result in addition of a cap to the 5'-terminus of a nucleic acid sequence (e.g., a method for 5'-capping of an mRNA molecule having a 5'-triphosphate using a guanyltransferase, such as, but not limited to, *Vaccinia* virus capping enzyme).

In some preferred embodiments, the sample containing the nucleic acids of interest (e.g., mRNA sequences) and the undesired nucleic acids (e.g., rRNA sequences) is partially purified prior to treatment using a composition or method of the present invention. By way of example, but not of limitation, total RNA can be purified from a sample using methods known in the art, including, for example, commercially available purification kits such as the MasterPure™ RNA Purification Kit (EPICENTRE Technologies, Madison, WI) or the RNeasy Kit (Qiagen, Valencia, CA) prior to treatment with an exoribonuclease according to the present invention.

In preferred embodiments, once the undesired nucleic acids (e.g., rRNA sequences) are eliminated, the nucleic acids of interest (e.g., mRNA sequences) are further purified using methods known in the art, including, for example, commercially available purification kits such as the MasterPure™ Complete DNA and RNA Purification Kit (EPICENTRE Technologies, Madison, WI) or the RNeasy Kit (Qiagen, Valencia, CA).

In some embodiments, the presently claimed invention provides a method of differentiating between nucleic acid sequences. For example, the present invention provides a method of distinguishing nucleic acid sequences that comprise an accessible 5' phosphate (e.g., that are targeted for digestion by a 5'-phosphate-dependent exonuclease) from nucleic acid sequences that do not possess an accessible 5' phosphate (e.g., that are

IDT Exhibit 1017.2015

not targeted for digestion by xrn-1). Thus, the present invention provides diagnostic methods. A wide variety of diagnostic methods are provided by the invention. Any method that utilizes an exoribonuclease to distinguish between nucleic acid molecules having and lacking 5' phosphate groups is encompassed by the present invention. For example, the exoribonucleases may be used to identify nucleic acid molecules that have been digested by restriction enzyme—generating a 5' phosphate—as compared to undigested nucleic acid molecules lacking a 5' phosphate.

In some embodiments, the method of differentiating between nucleic acids is used in combination with known diagnostic methods. For example, in some embodiments, methods of using exoribunucleases of the present invention are combined with restriction enzyme digestion used in restriction fragment length polymorphism (RFLP) analysis or other diagnostic methods.

In some embodiments, the enzymatic activity of the exoribonuclease of the present invention is controlled by altering the concentration of a divalent cation (e.g., Mg^{2+}) present in a reaction solution. In some embodiments, the enzymatic activity of exoribonuclease of the present invention is controlled by addition of a chelating agent.

EXAMPLES

The following examples serve to illustrate certain preferred embodiments and aspects of the present invention and are not to be construed as limiting the scope thereof.

Example 1

Activity of Exoribonuclease I

This Example describes the use of Exoribonuclease I in enriching for specific RNA substrates in the presence of other undesired nucleic acids.

A. Materials and Methods

Enzyme

Exoribonuclease I (xrn-1) (Stevens, Biochem. Biophys. Res. Commun. 81:656, 1978); Stevens, Biochem. Biophys. Res. Commun. 86:1126, 1979); Stevens, J. Biol.

IDT Exhibit 1017.2016

Chem 255:3080, 1980); and Stevens and Maupin, Nucleic Acids Res. 15:695, 1987) was purified from a recombinant source using methods similar to those described in the art. The enzyme was stored in 50% (v/v) glycerol containing, 0.05 M Tris-HCl (pH 7.5), 100 mM NaCl, 0.1 mM EDTA, 1 mM DTT, and 0.1% Triton X-100.

One unit of exoribonuclease I is the amount of enzyme that converts 1 μg of *E. coli* ribosomal RNA (comprising 16S and 23S rRNA) to acid-soluble form in 60 minutes at 30°C under standard reaction assay conditions. Exoribonuclease I used in the experiments described herein is free of other detectable contaminating ribonucleases, as well as free from detectable contaminating exo- and endodeoxyribonucleases (although the enzyme itself has DNase activity).

Reaction Conditions

All experiments were performed in a reaction buffer containing 50 mM Tris-HCl (pH 8.0), 2.0 mM $MgCl_2$, and 100 mM NaCl.

B. Results

Capped RNA is resistant to xrn-1 degradation

A 75-nucleotide, capped transcript was made with AmpliCap™ SP6 High Yield Message Maker Kit from EPICENTRE (Madison, WI). Samples containing this capped RNA were treated with tobacco acid pyrophosphatase (TAP; EPICENTRE) to remove the cap (or "de-cap") the RNA and/or with xrn-I, as indicated in Table 1, and analyzed by polyacrylamide gel electrophoresis. The results, shown in Table 1, demonstrated that the 5'-capped RNA was resistant to digestion (or "degradation") by xrn-1. De-capping the RNA with TAP resulted in degradation of the RNA by xrn-1 exonuclease.

IDT Exhibit 1017.2017

Table 1		
TAP Treated	Xrn-1 Treated	Degradation
-	-	-
-	+	-
+	-	-
+	+	+

Degradation of ribosomal RNA

A preparation of increasing amounts of tobacco RNA isolated from tobacco plants was treated with xrn-1 exonuclease and analyzed by agarose gel electrophoresis. Minimal RNA was visible in the lanes containing RNA treated with xrn-1. These results showed that the xrn-1 exonuclease was able to digest abundant ribosomal RNA.

The experiment was next repeated with a mixture of the tobacco RNA and a 1.4-kB capped mRNA transcript prepared using an AmpliCap™ T7 High Yield Message Maker Kit (EPICENTRE), which was added to the preparation of tobacco RNA, and with the 1.4-kB transcript alone. Results of the reactions were analyzed by agarose gel electrophoresis. The capped mRNA was not degraded by xrn-1. In the lanes containing both the tobacco RNA and the capped mRNA treated with xrn-1, the ribosomal RNA bands corresponding to the tobacco RNA were degraded, but those corresponding to the capped mRNA were not degraded. The present invention is not limited to a particular mechanism. Indeed, an understanding of the mechanism is not necessary to practice the present invention. Nonetheless, it is contemplated that the tobacco mRNA remains undegraded (see above) but is not visible due to its low concentration.

The xrn-1-treated tobacco RNA was concentrated approximately 50X by ethanol precipitation and analyzed on an agarose gel. The results indicated that the bulk of the RNA was degraded, but that a small amount of xrn-1-resistant RNA was present that appeared as a heterodisperse size range that could be observed on an agarose gel following ethanol concentration.

IDT Exhibit 1017.2018

5' pppG RNAs are resistant to degradation

A T7 control DNA template was used to synthesize a sample comprising a 1.4-kB RNA transcript using either: a) standard rNTPS without a cap; b) a capping rNTP mixture containing rNTPs and m^7GpppG; or c) a non-capping non-pppG 5' end mix containing rNTPs and an rGMP (which can be incorporated at the 5'-end of a transcript during *in vitro* transcription instead of a 5'-triphosphorylated G). Each transcript preparation was treated with xrn-1, and then xrn-1-treated and xrn-1-untreated samples were analyzed by agarose gel electrophoresis. The 5'pppG RNA was resistant to degradation by xrn-1. The RNA transcript that lacked a 5'-triphosphate or a cap structure was degraded by xrn-1. The RNA with rGMP in the transcription reaction was partially sensitive to degradation by xrn-1, presumably because some of the transcripts had a 5'-monophosphate due to incorporation of the GMP during *in vitro* transcription. These results show that xrn-1 can be used to enrich for uncapped but 5'-triphosphorylated RNA (e.g., bacterial mRNA).

xrn-1 degradation of E. coli RNA

E. coli total RNA was treated with xrn-1 exonuclease. Following xrn-1 treatment, a portion of the RNA preparation was concentrated approximately 20X by ethanol precipitation and analyzed by agarose gel electrophoresis. The majority of the RNA, including the *E. coli* 16S and 23S rRNA, was degraded. Following ethanol precipitation to concentrate the RNA remaining in the xrn-1-treated solution, a heterodisperse size range of undegraded RNA was observed on the agarose gel. Specific mRNAs for different genes were detected by RT-PCR of the xrn-1-treated material.

IDT Exhibit 1017.2019

Example 2

Enrichment of Eukaryotic mRNA from Eukaryotic Total RNA for Use in cDNA Synthesis, RNA Amplification and Preparation of Labeled Target RNA for Gene Expression Analysis Using DNA Microarrays.

Sample source and nucleic acid samples

Cells from a mouse stem cell line that had been transformed with a gene linked to an inducible promoter were treated with the inducing substance and then grown in culture for either 6 days or 12 days following induction. Total RNA was then isolated from each 6-day and 12-day culture using methods known in the art, resulting in samples that were designated as the "6-day sample" and the "12-day sample," respectively. Each RNA sample had prominent 18S and 28S rRNA peaks when they were analyzed using an Agilent 2100 bioanalyzer.

xrn-1 treatment of 6-day and 12-day RNA samples

The 6-day and 12-day RNA samples were each divided into two aliquots. The first aliquot was treated with xrn-1 exonuclease and the second aliquot was not treated with xrn-1 exonuclease. No 18S or 28S rRNA peaks are detectable using the Agilent 2100 bioanalyzer after the total RNA is treated with the xrn-1 exonuclease using the same standard protocol.

cDNA synthesis and RNA amplification of xrn-1-treated and xrn-1-untreated 6-day and 12-day RNA samples

Two hundred picograms of total RNA from the xrn-1-untreated aliquots of 6-day and 12-day RNA samples, or a volume of the xrn-1-treated aliquots equivalent to the volume of 200 pg of total RNA in the xrn-1-untreated aliquots, were each used for cDNA and aminoallyl-antisense RNA ("AA-aRNA") synthesis using the TargetAmp™ 2-Round Aminoallyl-aRNA Amplification Kit 1.0 (EPICENTRE) and SuperScript™_ III and SuperScript™ II reverse trancriptases (INVITROGEN) for cDNA synthesis in the first and second rounds of amplification, respectively, according to the instructions provided in the TargetAmp 1.0 kit; the TargetAmp 1.0 kit makes AA-aRNA corresponding to

IDT Exhibit 1017.2020

mRNA in the sample using an Eberwine-type aRNA amplification process (Van Gelder, R.N., et al., Proc. Natl. Acad. Sci. USA, 87: 1663, 1990). Sufficient amounts of AA-aRNA were obtained from amplification reactions for all four types of samples for use in preparing labeled target RNA for hybridization to Affymetrix mouse whole genome GeneChip®s.

Microarray analysis of amplified aRNA from mRNA in xrn-1-treated and xrn-1-untreated 6-day and 12-day RNA samples

Prior to hybridization to an Affymetrix GeneChip, each AA-aRNA was biotinylated using Biotin-X-X-NHS (EPICENTRE) according to the protocol of the supplier. The biotinylated a-RNA was then fragmented, and hybridized to mouse genome GeneChip arrays (MG 430 2.0, Affymetrix) in triplicates. Hybridization, staining, and washing were conducted as specified by Affymetrix. Data were analyzed using GeneSpring software (Agilent Technologies).

Microarray results using labeled target aRNA prepared from xrn-1-treated and xrn-1-untreated 6-day and 12-day RNA samples

The "% present calls" were 49-51% from all GeneChips hybridized to labeled target aRNA prepared from xrn-1-treated or xrn-1-untreated 12-day RNA samples, and ~43-48% for both 6-day samples (excluding one xrn-1-untreated replicate that had a "% present call" of only 29.2%). The concordance of present and absent calls between the xrn-1-treated and xrn-1-untreated 12-day samples was 73.6%. This concordance was close to the average concordance (82.6%) obtained from pairwise comparisons of three replicate slides hybridized with biotinylated aRNA samples obtained from untreated RNA. Thus, the high concordance indicates good preservation of messenger RNA in the xrn-1 treated samples.

IDT Exhibit 1017.2021

Example 3

Preservation of Relative mRNA Abundance Levels after xrn-1 exonuclease Treatment.

Two micrograms of human reference RNA and human skeletal muscle RNA were treated with 1U of xrn-1 exonuclease for 1 hour at 30°C. The RNA was extracted and concentrated by ethanol precipitation and the entire sample was used in a 40μl reverse transcription reaction using MMLV reverse transcriptase (EPICENTRE) with random nonamer primers at 37°C for 1 hour. qPCR was performed using TAQURATE GREEN Real-time PCR MasterMix (EPICENTRE) and optimized concentrations of target-specific primers. The abundance of beta-2-microglobulin (B2M) in the samples was used for expression level normalization. The expression levels of six different target messages were analyzed. Duplicate cDNA and qPCR reactions were preformed, averaged, and normalized for each comparison. Simultaneous analysis was performed with normalization and test primers, and non-template controls were included. The difference in threshold cycles between a target message in one RNA sample (in relation to the normalizer) and that of the other RNA sample (in relation to the normalizer) was determined. The results showed that xrn-1 exonuclease maintained mRNA abundance levels.

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

IDT Exhibit 1017.2022

CLAIMS

We claim:

1. A method for enriching for an RNA having a 5'-triphosphate or a 5'-cap in a biological sample comprising prokaryotic RNA, eukaryotic RNA or both prokaryotic and eukaryotic RNA and at least one undesired nucleic acid, the method comprising treating the sample with purified 5' exoribonuclease under conditions in which the 5' exoribonuclease is active and for sufficient time so that the undesired nucleic acid is digested and the sample is enriched for RNA having a 5'-triphosphate or a 5'-cap.

2. The method of claim 1, wherein the RNA having a 5'-triphosphate or a 5'-cap is selected from the group consisting of: (i) prokaryotic mRNA; (ii) eukaryotic mRNA, including polyadenylated and non-polyadenylated eukaryotic mRNA; (iii) a mixture of both prokaryotic and eukaryotic mRNA; (iv) eukaryotic snRNA; (v) eukaryotic pre-micro RNA; and (vi) prokaryotic or eukaryotic primary RNA transcripts.

3. The method of claim 1, wherein the RNA having a 5'-triphosphate or a 5'-cap comprises eukaryotic mRNA or both prokaryotic and eukaryotic mRNA, and wherein the method additionally comprises binding the RNA to an oligo(dT) or oligo(dU) resin, membrane or other surface to which oligo(dT) or oligo(dU) is attached and eluting the bound RNA so as to obtain a solution containing polyadenylated eukaryotic mRNA.

4. The method of claim 1 wherein the RNA having a 5'-triphosphate or a 5'-cap comprises prokaryotic mRNA, eukaryotic mRNA or both prokaryotic and eukaryotic mRNA and the undesired nucleic acid comprises prokaryotic rRNA, eukaryotic rRNA or both prokaryotic and eukaryotic rRNA of a size having a svedburg unit greater than about 10S.

IDT Exhibit 1017.2023

5. The method of claim 4 wherein the rRNA is selected from the group consisting of prokaryotic rRNA having a svedburg unit of about 16S or 23S and eukaryotic rRNA having a svedburg unit of about 18S, 26S or 28S.

6. The method of claim 1 wherein the RNA having a 5'-triphosphate or a 5'-cap comprises prokaryotic mRNA, eukaryotic mRNA, or both prokaryotic and eukaryotic mRNA, and wherein the method additionally comprises precipitation of said mRNA having a 5'-triphosphate or a 5'-cap with a solution of LiCl and ethanol at concentrations and under conditions in which tRNA and 5S rRNA are not precipitated.

7. The method of claim 6 wherein the RNA having a 5'-triphosphate or a 5'-cap comprises prokaryotic mRNA, eukaryotic mRNA, or both prokaryotic and eukaryotic mRNA and the undesired nucleic acid comprises prokaryotic rRNA, eukaryotic rRNA or both prokaryotic and eukaryotic rRNA of a size having a svedburg unit greater than about 10S.

8. The method of claim 7 wherein the rRNA is selected from the group consisting of prokaryotic rRNA having a svedburg unit of about 16S or 23S and eukaryotic rRNA having a svedburg unit of about 18S, 26S or 28S.

9. The method of claim 1, wherein the biological sample is treated with a polynucleotide kinase to phosphorylate the RNA having 5'-hydroxyl groups prior to treating the sample with purified 5' exoribonuclease.

10. The method of claim 1, wherein the method additionally comprises: (i) contacting the sample enriched for RNA having a 5'-triphosphate or a 5'-cap with a poly(A) polymerase under conditions so that polyadenylated RNA is obtained; and (ii) contacting the polyadenylated RNA with an oligo(dT)-containing primer and an RNA-dependent DNA polymerase under conditions that cDNA complementary to the RNA is obtained.

IDT Exhibit 1017.2024

11. The method of claim 9, wherein the biological sample is treated with a polynucleotide kinase to phosphorylate the RNA having 5'-hydroxyl groups prior to treating the sample with purified 5' exoribonuclease.

12. The method of claim 10, wherein the oligo(dT)-containing primer comprises a T7-type RNA polymerase promoter selected from the group consisting of T7 RNA polymerase, T3 RNA polymerase, and SP6 RNA polymerase.

13. The method of claim 1, wherein the biological sample comprises eukaryotic RNA, and wherein the method additionally comprises contacting the sample enriched for RNA having a 5'-triphosphate or a 5'-cap with an oligo(dT)-containing primer and an RNA-dependent DNA polymerase under conditions that cDNA complementary to the RNA is obtained.

14. The method of claim 13, wherein the biological sample is treated with a polynucleotide kinase to phosphorylate the RNA having 5'-hydroxyl groups prior to treating the sample with purified 5' exoribonuclease.

15. The method of claim 13, wherein the oligo(dT)-containing primer comprises a T7-type RNA polymerase promoter selected from the group consisting of T7 RNA polymerase, T3 RNA polymerase, and SP6 RNA polymerase.

16. The method of claim 15, wherein the method additionally comprises contacting the cDNA complementary to the RNA with one or more DNA or RNA primers and a DNA polymerase and incubating under conditions so as to obtain double-stranded cDNA having a functional promoter for the T7-type RNA polymerase and synthesizing RNA therefrom.

17. The method of claim 13, wherein the method additionally comprises contacting the cDNA complementary to the RNA with an oligonucleotide sequence tag

IDT Exhibit 1017.2025

template and a DNA polymerase and incubating under conditions so as to obtain cDNA having a sequence tag on it the 3'-end.

18. The method of claim 17, additionally comprising contacting the cDNA having the sequence tag on its 3'-end with a DNA polymerase and an oligonucleotide primer, wherein the either the sequence tag or the oligonucleotide primer, or the combination of both the sequence tag and the oligonucleotide primer encodes the complete sequence of a T7-type RNA polymerase promoter, and incubating under conditions wherein double-stranded cDNA having a functional T7-type RNA polymerase promoter is obtained.

19. The method of claim 18, wherein the method additionally comprises contacting the cDNA with the T7-type RNA polymerase under conditions so as to synthesize RNA using said T7-type RNA polymerase promoter.

20. The method of claim 1, wherein the 5' exoribonuclease is a wild-type or recombinant *Saccharomyces cerevisiae* Xrn1p/5' exoribonuclease 1.

21. A kit for enriching for mRNA having a 5'-triphosphate or a 5'-cap in a biological sample comprising prokaryotic mRNA, eukaryotic mRNA or both prokaryotic and eukaryotic mRNA and at least one undesired nucleic acid, the kit comprising: (i) a solution of purified and stabilized 5' exoribonuclease; and (ii) a concentrated solution of a reaction buffer for providing a 1X reaction buffer in which the 5' exoribonuclease is active.

22. The kit of claim 21, additionally comprising one or more components selected from the group consisting of: (i) a negative control comprising an RNA having a 5'-triphosphate or a 5'-cap; (ii) a positive control comprising an RNA having a 5'-monophosphate; (iii) a solution of LiCl; (iv) a polynucleotide kinase; and (v) a poly(A) polymerase.

IDT Exhibit 1017.2026

23. The kit of claim 21, additionally comprising an RNA-dependent DNA polymerase (reverse transcriptase).

24. The kit of claim 21, additionally comprising one or more components selected from the group consisting of: (i) a poly(A) polymerase; (ii) an oligo(dT)-containing primer; and (iii) an RNA-dependent DNA polymerase.

25. The kit of claim 21, additionally comprising one or more components selected from the group consisting of: (i) a polynucleotide kinase; (ii) an oligo(dT)-containing primer; and (iii) an RNA-dependent DNA polymerase.

26. The kit of claim 21, additionally comprising one or more primers selected from the group consisting of (i) an oligo(dT)-containing primer; and (ii) a primer that is complementary to a specific nucleic acid sequence.

27. The kit of claim 26, wherein said primer additionally comprises a sense or an antisense sequence of a double-stranded promoter for a T7-type RNA polymerase selected from the group consisting of T7 RNA polymerase, T3 RNA polymerase, and SP6 RNA polymerase, and wherein the kit additionally comprises the RNA polymerase that can transcribe RNA using said promoter.

28. The kit of claim 26, additionally comprising: (i) a oligonucleotide sequence tag template; and (ii) a DNA polymerase selected from the group consisting of a DNA-dependent DNA polymerase and an RNA-dependent DNA polymerase that can extend the 3'-end of a nucleic acid that is annealed to the oligonucleotide sequence tag template.

29. The kit of claim 28, additionally comprising: (i) an oligonucleotide that is complementary to the oligonucleotide sequence tag template; and (ii) a T7-type RNA polymerase selected from the group consisting of T7 RNA polymerase, T3 RNA polymerase, and SP6 RNA polymerase; wherein the oligonucleotide sequence tag

IDT Exhibit 1017.2027

template and the oligonucleotide that is complementary to the oligonucleotide sequence tag template together encode the complete sequence of a promoter that can be used for transcription by said T7-type RNA polymerase.

30. The kit of claim 21, additionally comprising one or more components selected from the group consisting of: (i) a negative control comprising an RNA having a 5'-triphosphate or a 5'-cap; (ii) a negative control comprising an RNA or a DNA having a 5'-hydroxyl group; (iii) a positive control comprising an RNA having a 5'-monophosphate; (iv) a solution of LiCl; (v) a ribonuclease H enzyme; and (vi) a polynucleotide kinase.

31. The kit of claim 30, additionally comprising one or more components selected from the group consisting of: (i) a poly(A) polymerase; (ii) an oligo(dT)-containing primer; and (iii) an RNA-dependent DNA polymerase.
(for prokaryotic mRNA)

32. The kit of claim 30, additionally comprising an RNA-dependent DNA polymerase and one or more primers selected from the group consisting of (i) a random primer, such as a random hexamer primer; (ii) an oligo(dT)-containing primer; and (iii) a primer that is complementary to a specific nucleic acid sequence.

33. The kit of claim 30, additionally comprising: (i) an oligonucleotide sequence tag template for adding a tag sequence to the 3'-end of mRNA having a 5'-triphosphate or a 5'-cap; and (ii) a DNA polymerase selected from the group consisting of a DNA-dependent DNA polymerase and an RNA-dependent DNA polymerase that can extend the 3'-end of the mRNA using the oligonucleotide sequence tag template annealed to the mRNA as a template.

34. The kit of claim 21, additionally comprising: (i) an oligonucleotide sequence tag template for adding a tag sequence to the 3'-end of mRNA having a 5'-triphosphate or a 5'-cap; and (ii) a DNA polymerase selected from the group consisting of

IDT Exhibit 1017.2028

a DNA-dependent DNA polymerase and an RNA-dependent DNA polymerase that can extend the 3'-end of the mRNA using the oligonucleotide sequence tag template annealed to the mRNA as a template.

35. The kit of claim 34, wherein the DNA polymerase is an RNA-dependent DNA polymerase, or if a DNA-dependent DNA polymerase is used for adding the tag sequence to the 3'-end of the mRNA, wherein the kit additionally contains an RNA-dependent DNA polymerase that can synthesize cDNA from a primer using the mRNA as a template.

36. The kit of claim 34, wherein said oligonucleotide sequence tag template additionally encodes either a sense or an antisense strand of a double-stranded promoter for a T7-type RNA polymerase selected from the group consisting of T7 RNA polymerase, T3 RNA polymerase, and SP6 RNA polymerase, and wherein the kit additionally comprises the RNA polymerase that can transcribe RNA using said promoter.

37. The kit of claim 21, additionally comprising: (i) a poly(A) polymerase; (ii) an oligo(dT)-containing primer; and (iii) an RNA-dependent DNA polymerase.

38. The kit of claim 37, additionally comprising: (i) an oligonucleotide sequence tag template; (ii) an oligonucleotide that is complementary to the oligonucleotide sequence tag template; wherein the oligonucleotide sequence tag template and the oligonucleotide complementary to the oligonucleotide sequence tag template together encode the complete sequence of a promoter for a T7-type RNA polymerase.

39. The kit of claim 37, additionally comprising a T7-type RNA polymerase selected from the group consisting of a T7 RNA polymerase, a T3 RNA polymerase, and an SP6 RNA polymerase, wherein the T7-type RNA polymerase can use the promoter

IDT Exhibit 1017.2029

encoded by the oligonucleotide sequence tag template and the oligonucleotide complementary to the oligonucleotide sequence tag template.

40. A kit for removing intact prokaryotic or eukaryotic ribosomal RNA (rRNA) of a size having a svedburg unit greater than about 10S from a biological sample, the kit comprising: (i) a solution of purified and stabilized 5' exoribonuclease; (ii) a concentrated solution of a reaction buffer for providing a 1X reaction buffer in which the exoribonuclease is active; and (iii) a positive control rRNA, selected from the group consisting of a 16S and a 23S prokaryotic rRNA, and an 18S, a 26S and a 28S eukaryotic rRNA.

41. The kit of claim 40, wherein the 5' exoribonuclease is a wild-type or recombinant *Saccharomyces cerevisiae* Xrn1p/5' exoribonuclease 1.

IDT Exhibit 1017.2030

Figure 1

SEQ ID NO:1

MGIPKFFRYISERWPMILQLIEGTQIPEFDNLYLDMNSILHNCT
HGNDDDVTKRLTEEEVFAKICTYIDHLFQTIKPKKIFYMAIDGVAPRAKMNQQRARRF
RTAMDAEKALKKAIENGDEIPKGEPFDSNSITPGTEFMAKLTKNLQYFIHDKISNDSK
WREVQIIFSGHEVPGEGEHKIMNFIRHLKSQKDFNQNTRHCIYGLDADLIMLGLSTHG
PHFALLREEVTFGRRNSEKKSLEHQNFYLLHLSLLREYMELEFKEIADEMQFEYNFER
ILDDFILVMFVIGNDFLPNLPDLHLNKGAFPVLLQTFKEALLHTDGYINEHGKINLKR
LGVWLNYLSQFELLNFEKDDIDVEWFNKQLENISLEGERKRQRVGKKLLVKQQKKLIG
SIKPWLMEQLQEKLSPDLPDEEIPTLELPKDLDMKDHLEFLKEFAFDLGLFITHSKSK
GSYSLKMDLDSINPDETEEEFQNRVNSIRKTIKKYQNAIIVEDKEELETEKTIYNERF
ERWKHEYYHDKLKFTTDSEEKVRDLAKDYVEGLQWVLYYYYRGCPSWSWYYPHHYAPR
ISDLAKGLDQDIEFDLSKPFTPFQQLMAVLPERSKNLIPPAFRPLMYDEQSPIHDFYP
AEVQLDKNGKTADWEAVVLISFVDEKRLIEAMQPYLRKLSPEEKTRNQFGKDLIYSFN
PQVDNLYKSPLGGIFSDIEHNHCVEKEYITIPLDSSEIRYGLLPNAKLGAEMLAGFPT
LLSLPFTSSLEYNETMVFQQPSKQQSMVLQITDIYKTNNVTLEDFSKRHLNKVIYTRW
PYLRESKLVSLTDGKTIYEYQESNDKKKFGFITKPAETQDKKLFNSLKNSMLRMYAKQ
KAVKIGPMEAIATVFPVTGLVRDSDGGYIKTFSPTPDYYPLQLVVESVVNEDERYKER
GPIPIEEEFPLNSKVIFLGDYAYGGETTIDGYSSDRRLKITVEKKFLDSEPTIGKERL
QMDHQAVKYYPSYIVSKNMHLHPLFLSKITSKFMITDATGKHINVGIPVKFEARHQKV
LGYARRNPRGWEYSNLTLNLLKEYRQTFPDFFFRLSKVGNDIPVLEDLFPDTSTKDAM
NLLDGIKQWLKYVSSKFIAVSLESDSLTKTSIAAVEDHIMKYAANIEGHERKQLAKVP
REAVLNPRSSFALLRSQKFDLGDRVVYIQDSGKVPIFSKGTVVGYTTLSSSLSIQVLF
DHEIVAGNNFGGRLRTNRGLGLDASFLLNITNRQFIYHSKASKKALEKKKQSNNRNNN
TKTAHKTPSKQQSEEKLRKERAHDLLNFIKKDTNEKNSESVDNKSMGSQKDSKPAKKV
LLKRPAQKSSENVQVDLANFEKAPLDNPTVAGSIFNAVANQYSDGIGSNLNIPTPPHP
MNVVGGPIPGANDVADVGLPYNIPPGFMTHPNGLHPLHPHQMPYPNMNGMSIPPPAPH
GFGQPISFPPPPPMTNVSDQGSRIVVNEKESQDLKKFINGKQHSNGSTIGGETKNSRK
GEIKPSSGTNSTECQSPKSQSNAADRDNKKDEST

SEQ ID NO:2

atgggtattc caaaattttt caggtacatc tcagaaagat ggcccatgat tttacagctt
attgagggaa cacagattcc tgagtttgat aacttatacc tggatatgaa ttcgatttta
cataattgta cgcatggtaa cgacgatgat gtaaccaagc gattaactga agaagaggtt
tttgcaaaaa tctgtacgta tatcgatcac ctttttcaaa caatcaagcc caagaagatt
ttctacatgg ctattgatgg tgtggcccct cgtgccaaga tgaatcaaca aagagctcgt
agattcagaa ccgctatgga tgcagaaaaa gccttgaaga aggctattga gaatggtgac
gagattccta agggtgagcc atttgattcg aattctatta ctccaggtac ggagtttatg
gccaaattga ccaaaaactt acaatatttt attcacgaca agatttctaa cgattccaaa
tggagggaag tgcaaatcat attttctggc catgaagttc caggtgaagg tgaacacaag
atcatgaact ttataaggca tttaaaatcc caaaaggatt tcaaccagaa tacgagacat
tgtatttacg gtcttgacgc agatttgatt atgctgggtt tgtctactca tgggccacat
tttgcgttat tgagagaaga agtgacattt ggtagaagaa atagtgaaaa aaaatcgctt
gaacatcaaa atttctactt attacatctt tctttattaa gagaatacat ggagttggaa
ttcaaagaaa ttgccgatga aatgcaattt gaatacaatt ttgaacgtat tttggatgat
tttattcttg tcatgttcgt cattggtaat gatttcttgc ccaatttgcc agatttgcac
cttaacaaag gagcatttcc cgttttgtta caaacgttca aagaagctct tttacatact
gatggctaca ttaatgaaca tggtaaaata aatttaaaga gattaggtgt ctggttaaat
tatctgtctc aatttgagtt attaaatttc gaaaaggatg atatagacgt tgagtggttc
aacaagcaat tagagaatat ttctttggag ggtgaaagga aaagacagag ggttggtaaa
aaattactgg taaaacaaca gaagaaatta attggaagta taaaaccatg gttgatggag
caattacagg aaaaattatc gcctgattta ccagatgaag aaattccaac tttagagtta
cctaaggact tagacatgaa agatcattta gaatttttaa aagaattcgc ttttgatttg

IDT Exhibit 1017.2031

ggtcttttta taacgcattc caaatccaaa ggtagttatt cgctaaaaat ggatcttgat
tctattaatc ctgatgaaac agaagaagaa tttcaaaatc gtgttaattc tatcaggaaa
acaataaaaa aatatcaaaa tgctatcatc gtggaggaca aagaagaatt ggaaactgaa
aaaacgattt ataatgaaag gtttgaacgt tggaagcatg agtattatca cgacaagtta
aaatttacga cagacagtga agaaaaagtg agagatcttg ctaaagacta cgttgaaggt
ttacaatggg ttctatatta ttattataga ggatgtccat cttggtcgtg gtactatccg
caccattatg caccaagaat ctccgactta gccaagggtt tagatcaaga cattgaattt
gatttgagca aaccatttac tccattccag caactaatgg cagttttacc ggaaaggtcc
aaaaatctga tacctcccgc ctttaggcca ttaatgtacg atgaacagtc gccaatccac
gatttctatc ccgctgaggt tcaacttgat aaaaacggca agacagctga ttgggaagct
gtggttttga tatcgtttgt agatgaaaaa aggttgattg aggctatgca accttatttg
cgcaagttat cacctgaaga aaaaacgaga aatcaatttg gcaaggactt gatatattcc
tttaatcctc aagttgataa cctttataag agtccgttgg gcggcatttt ttctgatatt
gaacacaatc attgtgtcga aaaagagtac atcaccatcc cattggacag ctccgagatt
cggtatggtt tattacctaa tgctaaactc ggtgccgaaa tgctggcggg tttccccacg
ttattgtctt taccatttac tagttcactg gagtacaatg agacaatggt tttccaacaa
ccttctaaac aacaatcaat ggtcttacaa ataactgaca tatacaaaac gaataatgtt
actttggagg acttttccaa gaggcattta aacaaagtga tttatacaag atggccatat
ttaagagaat ccaaattggt ctctttaacg gatggtaaga ctatctatga atatcaggag
tccaatgata agaaaaagtt cggattcata acgaagcctg cggaaaccca ggacaaaaaa
cttttcaata gtttgaagaa ttcaatgcta aggatgtatg ctaaacagaa agctgttaaa
ataggaccta tggaagccat tgctaccgtc tttccagtga ctggtttggt gagagactct
gatggtggtt atattaagac ctttagccct accccagatt actatccatt gcaactggtt
gttgaatctg ttgtcaacga ggatgaaaga tataaagaaa gaggacccat tcctattgaa
gaggaatttc cattgaattc aaaagttatt ttcttaggtg attatgccta tggtggtgaa
actactattg acggttacag cagtgaccgc agactaaaaa ttactgtaga aaagaagttt
ttggatagtg agcccaccat cggcaaagaa aggttacaaa tggatcatca agccgttaaa
tattatccgt cttatattgt gtccaagaac atgcacttac acccccttgtt tttgtctaag
attacttcca agttcatgat tactgacgct actgggaagc atatcaatgt tggtatcccg
gttaagttcg aagctagaca ccaaaaggtt ttaggttacg cgaggaggaa ccctaggggc
tgggaatact caaatttaac tctaaattta ctgaaagagt atagacaaac tttcccagat
ttttttttca ggttgtccaa ggtaggtaat gatatcccag ttttggaaga tcttttcccc
gatacttcca ctaaggatgc catgaattta ttagatggta tcaaacaatg gctaaagtat
gtctcatcga agtttatcgc ggtatctttg gagtctgact ccttaactaa gacatcgatt
gctgccgtgg aagatcatat catgaaatac gcagctaaca tcgaaggtca tgaaagaaaa
cagttagcca aagttcctcg tgaggctgtt ttgaatccaa gatcatcatt tgcactctta
cgtagtcaaa agttcgattt gggtgaccgt gttgtttata tccaagattc tggtaaggta
ccaatttttct caaagggtac agttgttggc tatactactc tcagttcatc attatcaatt
caggtcttat ttgatcatga aatcgtggcc ggtaataatt ttggcggcag gttgcgcacg
aatagaggct tagggcttga tgcctctttc ttattgaata ttactaacag gcagttcatt
tatcactcca aggcttccaa aaaggctttg gaaaagaaaa agcaatctaa caataggaac
aataatacca aaactgctca caagactcct tcaaagcaac aatctgaaga aaaactgaga
aaagaaaggg cacatgattt attgaatttt atcaaaaagg ataccaatga aaagaattct
gaaagtgtag acaacaagag catgggatcg caaaaagatt ccaaacccgc aaagaaagtt
ttgttaaaaa gaccagctca gaaaagcagt gaaaacgtgc aagttgattt ggccaatttt
gaaaaagcac cgcttgataa tccaactgtt gctggatcta ttttcaatgc cgttgcaaat
caatattctg atggtatagg cagtaatttg aatatcccaa ctccacctca cccaatgaat
gtggttgggg gtcctattcc tggagcgaat gatgttgcag atgttggttt gccgtacaat
attccccccag gttttatgac gcatcctaat ggtcttcacc cattacaccc tcaccagatg
ccttacccta atatgaatgg aatgtctatt ccgccaccag caccacatgg gtttggacaa
ccgatttcct tcccacctcc acctcctatg acaaatgttt cagatcaagg aagtcgtatt
gttgtcaatg aaaaggaaag ccaagatttg aaaaaattca ttaatggtaa acagcacagc
aatggttcaa ctattggggg agaaacaaag aacagtagga aaggcgagat taaaccttct
tctggcacaa actctactga atgtcaatcg ccaaagtcac aaagcaatgc tgctgaccgt
gataataaaa aagacgaatc

IDT Exhibit 1017.2032

Electronic Acknowledgement Receipt

EFS ID:	34484601
Application Number:	16017340
International Application Number:	
Confirmation Number:	2635
Title of Invention:	COMPOSITIONS AND METHODS FOR TARGETED NUCLEIC ACID SEQUENCE ENRICHMENT AND HIGH EFFICIENCY LIBRARY GENERATION
First Named Inventor/Applicant Name:	Doug AMORESE
Customer Number:	12754
Filer:	Na Kyung Lee/Keith Gneshin
Filer Authorized By:	Na Kyung Lee
Attorney Docket Number:	25115-755.302
Receipt Date:	04-DEC-2018
Filing Date:	25-JUN-2018
Time Stamp:	18:08:47
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	no

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Non Patent Literature	Z-BROWN-CHEMICAL-SYNTHESIS-109.PDF	2764248 f3baccc37d050f002971903321dbf93b82b66cd7	no	43

Warnings:

Information:					
2	Non Patent Literature	Z-FROUSSARD-A-RANDOM-2900.PDF	321286 59dc7cf0bbf77342fe4b8e2838869b50042b8f7a	no	1
Warnings:					
Information:					
3	Non Patent Literature	Z-GUDMUNDSSON-GENOME-WIDE-1122.PDF	1390801 28bcf79510f3a29ee1117b0610b7c3fcba3288c3	no	15
Warnings:					
Information:					
4	Non Patent Literature	ZHANG_COPY_NUMBER_VARIATION_451_EFS.PDF	3843552 5e850504ce0cf11305489eaeac1ce161242dc59e	no	33
Warnings:					
Information:					
5	Non Patent Literature	ZHONG-HIGH-THROUGHPUT-940.PDF	1135891 91d3cf63c651bcc6a0b0151fdf2df37e4c9ee43d	no	10
Warnings:					
Information:					
6	Non Patent Literature	Z-LEVESQUE-SERGERIE-DETECTION-LIMITS-93.PDF	875724 5588c405092b80d6441fa07c0c385bb2cfe516cb	no	18
Warnings:					
Information:					
7	Non Patent Literature	Z-MULTI-SITE-NUGEN-2015.PDF	68366 0fb2598e7c0edb643f81175fb85884b269366dcd	no	3
Warnings:					
Information:					
8	Non Patent Literature	Z-NARANG-IMPROVED-90.PDF	433255 d61350042592d4fa52959ebba1549f48236e0dae	no	9
Warnings:					
Information:					

9	Non Patent Literature	Z-NOA-01-05-2017-US13750768.PDF	458724 f041305b80e00d4c7828bbf6585b8b0a7518bd02	no	9
Warnings:					
Information:					
10	Non Patent Literature	Z-NOA-04-17-2017-US14634326.PDF	472234 616c89dbfccba66b797248fa4aee981bce005098	no	9
Warnings:					
Information:					
11	Non Patent Literature	Z-NOA-06-21-2017-US14634326.PDF	78871 0ce810d2190c411aa9f44f2d22f7c2fae832cf64	no	2
Warnings:					
Information:					
12	Non Patent Literature	Z-NOA-07-25-2017-US14778564.PDF	437633 ba45c137b7002d24788fff69531a1ad00df1e60c	no	9
Warnings:					
Information:					
13	Non Patent Literature	Z-NUGEN-LAUNCHES-ENHANCED-2015.PDF	65059 2f3410e19cd5ec4a1fc9f6f57c5bb6ddc46bf6aa	no	2
Warnings:					
Information:					
14	Non Patent Literature	Z-NUGEN-SCIENTISTS-SCREEN-2015.PDF	68983 1d85ace704d4c484dc8fd27de9582bd60cf52133	no	2
Warnings:					
Information:					
15	Non Patent Literature	Z-OA-06-02-16-USAPP-13-750768.PDF	1122701 b1a654cdfd285e9ca46365f73dfb230b308c077a	no	29
Warnings:					
Information:					

16	Non Patent Literature	Z-OA-07-19-2017-US14390012.PDF	397493 5a4e5c3727ec0d4f392fb51e6ea4c24728924e1d	no	12
Warnings:					
Information:					
17	Non Patent Literature	Z-OA-08-07-2017-US14920254.PDF	372188 cb2eb1796fa0759464113819bfc9f5bc1257fb5e	no	9
Warnings:					
Information:					
18	Non Patent Literature	Z-VOELKERDING-NEXT-GENERATION-641.PDF	2879816 483c81f9d952d262970a3cec5cb75740ee6283c4	no	24
Warnings:					
Information:					
		Total Files Size (in bytes):	17186825		

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111

If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371

If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office

If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

Electronic Acknowledgement Receipt

EFS ID:	34484161
Application Number:	16017340
International Application Number:	
Confirmation Number:	2635
Title of Invention:	COMPOSITIONS AND METHODS FOR TARGETED NUCLEIC ACID SEQUENCE ENRICHMENT AND HIGH EFFICIENCY LIBRARY GENERATION
First Named Inventor/Applicant Name:	Doug AMORESE
Customer Number:	12754
Filer:	Na Kyung Lee/Keith Gneshin
Filer Authorized By:	Na Kyung Lee
Attorney Docket Number:	25115-755.302
Receipt Date:	04-DEC-2018
Filing Date:	25-JUN-2018
Time Stamp:	17:43:23
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	no

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1		1-IDS-25115-755-302-12-04-18.pdf	1041845 ca846e3d51ce1b39adc85f0916eeff8ac95c8ecc	yes	65

	Multipart Description/PDF files in .zip description		
	Document Description	**Start**	**End**
	Transmittal Letter	1	4
	Information Disclosure Statement (IDS) Form (SB08)	5	65
Warnings:			
Information:			
	Total Files Size (in bytes):	1041845	

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111

If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371

If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office

If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Inventor:	AMORESE; Doug et al.	Group Art Unit:	Not Yet Assigned
Serial Number:	16/017,340	Examiner:	Not Yet Assigned
Filing or 371 (c) Date:	2018-06-25	**CONFIRMATION NO:**	**2635**
Title:	COMPOSITIONS AND METHODS FOR TARGETED NUCLEIC ACID SEQUENCE ENRICHMENT AND HIGH EFFICIENCY LIBRARY REGENERATION		

FILED ELECTRONICALLY ON: December 4, 2018

Commissioner for Patents
P.O. Box 1450
Alexandria VA 22313-1450

INFORMATION DISCLOSURE STATEMENT UNDER 37 CFR § 1.97

Commissioner for Patents:

An Information Disclosure Statement along with attached PTO/SB/08 is hereby submitted. The information provided in the attached 61-page PTO/SB/08 was previously considered by the Office in a parent application and is therefore deemed to have been considered in the present continuing application (see MPEP § 609.02). Applicant desires the information considered in the parent application to be printed on any patent issuing from the present continuing application and respectfully requests that the Examiner initial and return the attached PTO/SB/08 in accordance with MPEP § 609.

The Examiner is requested to review the information provided and to make the information of record in the above-identified application.

The right to establish the patentability of the claimed invention over any of the information provided herewith, and/or to prove that this information may not be prior art, and/or to prove that this information may not be enabling for the teachings purportedly offered, is hereby reserved.

IDT Exhibit 1017.2039

This statement is not intended to represent that a search has been made or that the information cited in the statement is, or is considered to be, prior art or material to patentability as defined in § 1.56.

A. ☒ *37 CFR § 1.97 (b).* This Information Disclosure Statement should be considered by the Office because:

☐ (1) It is being filed within 3 months of the filing date of a national application and is other than a continued prosecution application under § 1.53 (d);

-- OR --

☐ (2) It is being filed within 3 months of entry of the national stage as set forth in § 1.491 in an international application;

-- OR --

☒ (3) It is being filed before the mailing of a first Office action on the merits;

-- OR --

☐ (4) It is being filed before the mailing of a first Office action after the filing of a request for continued examination under § 1.114.

B. ☐ *37 CFR § 1.97(c).* Although this Information Disclosure Statement is being filed after the period specified in *37 CFR § 1.97(b),* above, it is filed before the mailing date of the earlier of (1) a final office action under § 1.113, (2) a notice of allowance under § 1.311, or (3) an action that otherwise closes prosecution in the application, this Information Disclosure Statement should be considered because it is accompanied by one of:

☐ a statement as specified in §1.97 (e) provided concurrently herewith;

-- OR --

☐ a fee of $240.00 as set forth in § 1.17 (p) authorized below, enclosed, or included with the payment of other papers filed together with this statement.

C. ☐ *37 CFR § 1.97 (d).* Although this Information Disclosure Statement is being filed after the mailing date of the earlier of (1) a final office action under § 1.113, (2) a notice of allowance under § 1.311, or (3) an action that otherwise closes prosecution in the application, it is being filed before payment of the issue fee and should be considered because it is accompanied by:

i. a statement as specified in § 1.97 (e);

-- AND --

ii. a fee of $240.00 as set forth in §1.17(p) is authorized below, enclosed, or included with the payment of other papers filed together with this Statement.

D. ☐ *37 CFR §1.97 (e).* Statement.

☐ A statement is provided herewith to satisfy the requirement under 37 CFR §§ 1.97 (c);

-- AND/OR --

☐ A statement is provided herewith to satisfy the requirement under 37 CFR §§ 1.97 (d);

-- AND/OR --

☐ A copy of a dated communication from a foreign patent office clearly showing that the information disclosure statement is being submitted within 3 months of the filing date on

IDT Exhibit 1017.2040

the communication is provided in lieu of a statement under 37 C.F.R. § 1.97(e) (1) as provided for under MPEP 609.04(b) V.

E. ☐ *Statement Under 37 C.F.R. §1.704(d).* Each item of information contained in the information disclosure statement was first cited in any communication from a patent office in a counterpart foreign or international application or from the Office or is a communication that was issued by a patent office in a counterpart foreign or international application or by the Office that was received by an individual designated in § 1.56(c) not more than thirty (30) days prior to the filing of this information disclosure statement. This statement is made pursuant to the requirements of 37 C.F.R. §1.704(d) to avoid reduction of the period of adjustment of the patent term for Applicant(s) delay.

F. ☒ *37 CFR §1.98 (a) (2).* The content of the Information Disclosure Statement is as follows:

☐ Copies of each of the references listed on the attached Form PTO/SB/08 are enclosed herewith.

-- OR --

☒ Copies of U.S. Patent Documents (issued patents and patent publications) listed on the attached Form PTO/SB/08 are not enclosed.

-- AND/OR --

☐ Copies of Foreign Patent Documents and/or Non Patent Literature Documents listed on the attached Form PTO/SB/08 are enclosed in accordance with 37 CFR §1.98 (a)(2).

-- AND/OR --

☐ Copies of pending unpublished U.S. patent applications are enclosed in accordance with 37 CFR §1.98 (a) (2) (iii).

G. ☐ *37 CFR §1.98(a)(3).* The Information Disclosure Statement includes non-English patents and/or references.

☐ Pursuant to 37 CFR §1.98(a)(3)(i), a concise explanation of the relevance of each patent, publication or other information provided that is not in English is provided herewith.

☐ Pursuant to MPEP 609(B), an English language copy of a foreign search report is submitted herewith to satisfy the requirement for a concise explanation where non-English language information is cited in the search report.

-- OR --

☐ A concise explanation of the relevance of each patent, publication or other information provided that is not in English is as follows: _____

☐ Pursuant to 37 CFR §1.98(a) (3) (ii), a copy of a translation, or a portion thereof, of the non-English language reference(s) is provided herewith.

H. ☒ *37 CFR §1.98(d).* Copies of patents, publications and pending U.S. patent applications, or other information specified in 37 C.F.R. § 1.98(a) are not provided herewith because:

☒ Pursuant to 37 CFR §1.98(d)(1) the information was previously submitted in an Information Disclosure Statement, or cited by examiner for another application under which this application claims priority for an earlier effective filing date under 35 U.S.C. 120.

Application in which the information was submitted: U.S. Patent Application No. 13/750,768 filed on 01/25/2013.

IDT Exhibit 1017.2041

Information Disclosure Statement(s) and/or PTO-892 Form(s) filed on: 04/16/2013, 07/16/2013, 09/11/2013, 09/24/2013, 01/30/2014, 02/05/2014, 06/23/2014, 08/29/2014, 09/25/2014, 10/01/2014, 11/21/2014, 02/05/2015, 02/18/2015, 03/02/2015, 03/10/2015, 03/18/2015, 04/10/2015, 04/14/2015, 04/16/2015, 06/03/2015, 06/26/2015, 07/15/2015, 07/22/2015, 10/16/2015, 10/30/2015, 11/09/2015, 02/01/2016, 05/06/2016, 05/11/2016, 05/23/2016, 06/02/2016, 12/02/2016, and 03/29/2017.

AND

☒ The information disclosure statement submitted in the earlier application complied with paragraphs (a) through (c) of 37 CFR §1.98.

I. ☒ *Fee Authorization*. The Commissioner is hereby authorized to charge the above-referenced fees of $0.00 and charge any additional fees or credit any overpayment associated with this communication to Deposit Account No. 23-2415 (Docket No.25115-755.302).

Respectfully submitted,

WILSON SONSINI GOODRICH & ROSATI

Dated: December 4, 2018

By: /Jeffrey Seidel/
Jeffrey Seidel
Reg. No. 61,063

650 Page Mill Road
Palo Alto, CA 94304-1050
(650) 493-9300
Customer No. 12754

IDT Exhibit 1017.2042

Approved for use through 11/30/2020. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Substitute for form 1449/PTO

INFORMATION DISCLOSURE STATEMENT BY APPLICANT

(Use as many sheets as necessary)

Sheet | 1 | of | 61

Complete if Known	
Application Number	16017340
Filing Date	06-25-2018
First Named Inventor	AMORESE; Doug
Art Unit	Not Yet Assigned
Examiner Name	Not Yet Assigned
Attorney Docket Number	25115-755.302

U. S. PATENT DOCUMENTS

Examiner Initials*	Cite No.[1]	Document Number Number-Kind Code[2] (if known)	Publication Date MM-DD-YYYY	Name of Patentee or Applicant of Cited Document	Pages, Columns, Lines, Where Relevant Passages or Relevant Figures Appear
	001	US-4362867	12-07-1982	PADDOCK; Gary V.	
	002	US-4469863	09-04-1984	TS'O; Paul O. P. et al.	
	003	US-4582877	04-15-1986	FAIRCHOK; William J. et al.	
	004	US-4876187	10-24-1989	DUCK; Peter et al.	
	005	US-4935357	06-19-1990	SZYBALSKI; Waclaw	
	006	US-4942124	07-17-1990	CHURCH; George M [Us] et al.	
	007	US-4988617	01-29-1991	LANDEGREN; Ulf et al.	
	008	US-4996143	02-26-1991	HELLER; Michael J. et al.	
	009	US-5011769	04-30-1991	DUCK; Peter [Ca] et al.	
	010	US-5034506	07-23-1991	SUMMERTON; James E. et al.	
	011	US-5035996	07-30-1991	HARTLEY; James L.	
	012	US-5043272	08-27-1991	HARTLEY; James L.	
	013	US-5082830	01-21-1992	BRAKEL; Christine L. et al.	
	014	US-5090591	02-25-1992	LONG; John D.	
	015	US-5130238	07-14-1992	MALEK; Lawrence T. et al.	
	016	US-5169766	12-08-1992	SCHUSTER; David M. et al.	
	017	US-5171534	12-15-1992	SMITH; Lloyd M. et al.	
	018	US-5194370	03-16-1993	BERNINGER; Mark S. et al.	
	019	US-5216141	06-01-1993	BENNER; Steven A.	
	020	US-5234809	08-10-1993	BOOM; Willem R. et al.	
	021	US-5235033	08-10-1993	SUMMERTON; James E. et al.	
	022	US-5242794	09-07-1993	WHITELEY; Norman M. et al.	

Examiner Signature		Date Considered	

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant. [1] Applicant's unique citation designation number (optional). [2] See Kinds Codes of USPTO Patent Documents at www.uspto.gov or MPEP 901.04. [3] Enter Office that issued the document, by the two-letter code (WIPO Standard ST.3). [4] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [5] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. [6] Applicant is to place a check mark here if English language Translation is attached. This collection of information is required by 37 CFR 1.97 and 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 2 hours to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 (1-800-786-9199) and select option 2.

Approved for use through 11/30/2020. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Substitute for form 1449/PTO

INFORMATION DISCLOSURE STATEMENT BY APPLICANT

(Use as many sheets as necessary)

Sheet 2 of 61

Complete if Known	
Application Number	16017340
Filing Date	06-25-2018
First Named Inventor	AMORESE; Doug
Art Unit	Not Yet Assigned
Examiner Name	Not Yet Assigned
Attorney Docket Number	25115-755.302

U. S. PATENT DOCUMENTS

Examiner Initials*	Cite No.[1]	Document Number Number-Kind Code[2] (if known)	Publication Date MM-DD-YYYY	Name of Patentee or Applicant of Cited Document	Pages, Columns, Lines, Where Relevant Passages or Relevant Figures Appear
	023	US-5386023	01-31-1995	SANGHVI; Yogesh S. et al.	
	024	US-5399491	03-21-1995	KACIAN; Daniel L. et al.	
	025	US-5409818	04-25-1995	DAVEY; Cheryl et al.	
	026	US-5418149	05-23-1995	GELFAND; David H. et al.	
	027	US-5427929	06-27-1995	RICHARDS; Rodney M. et al.	
	028	US-5480784	01-02-1996	KACIAN; Daniel L. et al.	
	029	US-5494810	02-27-1996	BARANY; Francis et al.	
	030	US-5508169	04-16-1996	DEUGAU; Kenneth V. et al.	
	031	US-5508178	04-16-1996	ROSE; Samuel et al.	
	032	US-5510270	04-23-1996	FODOR; Stephen P. A. et al.	
	033	US-5545522	08-13-1996	VAN; Gelder Russell N. et al.	
	034	US-5554516	09-10-1996	KACIAN; Daniel L. et al.	
	035	US-5554517	09-10-1996	DAVEY; Cheryl et al.	
	036	US-5556752	09-17-1996	LOCKHART; David J. et al.	
	037	US-5565340	10-15-1996	CHENCHIK; Alex et al.	
	038	US-5573913	11-12-1996	ROSEMEYER; Viola et al.	
	039	US-5578832	11-26-1996	TRULSON; Mark et al.	
	040	US-5589339	12-31-1996	HAMPSON; Ian N. et al.	
	041	US-5602240	02-11-1997	DE; Mesmaeker Alain et al.	
	042	US-5637684	06-10-1997	COOK; Phillip D. et al.	
	043	US-5641658	06-24-1997	ADAMS; Christopher P. et al.	
	044	US-5644048	07-01-1997	YAU; Eric K.	

Examiner Signature		Date Considered	

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant. [1] Applicant's unique citation designation number (optional). [2] See Kinds Codes of USPTO Patent Documents at www.uspto.gov or MPEP 901.04. [3] Enter Office that issued the document, by the two-letter code (WIPO Standard ST.3). [4] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [5] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. [6] Applicant is to place a check mark here if English language Translation is attached. This collection of information is required by 37 CFR 1.97 and 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 2 hours to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 (1-800-786-9199) and select option 2.

Approved for use through 11/30/2020. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Substitute for form 1449/PTO

INFORMATION DISCLOSURE STATEMENT BY APPLICANT

(Use as many sheets as necessary)

Sheet 3 of 61

Complete if Known	
Application Number	16017340
Filing Date	06-25-2018
First Named Inventor	AMORESE; Doug
Art Unit	Not Yet Assigned
Examiner Name	Not Yet Assigned
Attorney Docket Number	25115-755.302

U. S. PATENT DOCUMENTS

Examiner Initials*	Cite No.[1]	Document Number Number-Kind Code[2] (if known)	Publication Date MM-DD-YYYY	Name of Patentee or Applicant of Cited Document	Pages, Columns, Lines, Where Relevant Passages or Relevant Figures Appear
	045	US-5665549	09-09-1997	PINKEL; Daniel et al.	
	046	US-5667976	09-16-1997	VAN; Ness Jeffrey et al.	
	047	US-5667979	09-16-1997	BERRENS; Lubertus	
	048	US-5679512	10-21-1997	LANEY; Maureen et al.	
	049	US-5681726	10-28-1997	HUSE; William David et al.	
	050	US-5683879	11-04-1997	LANEY; Maureen et al.	
	051	US-5688648	11-18-1997	MATHIES; Richard A. et al.	
	052	US-5705628	01-06-1998	HAWKINS; Trevor	
	053	US-5708154	01-13-1998	SMITH; Steven S. et al.	
	054	US-5710028	01-20-1998	EYAL; Nurit et al.	
	055	US-5712126	01-27-1998	WEISSMAN; Sherman M. et al.	
	056	US-5716785	02-10-1998	VAN; Gelder Russell N. et al.	
	057	US-5726329	03-10-1998	JONES; David S. et al.	
	058	US-5750341	05-12-1998	MACEVICZ; Stephen C.	
	059	US-5759822	06-02-1998	CHENCHIK; Alex et al.	
	060	US-5763178	06-09-1998	CHIRIKJIAN; Jack G. et al.	
	061	US-5789206	08-04-1998	TAVTIGIAN; Sean V. et al.	
	062	US-5824517	10-20-1998	CLEUZIAT; Philippe et al.	
	063	US-5824518	10-20-1998	KACIAN; Daniel Louis et al.	
	064	US-5837832	11-17-1998	CHEE; Mark et al.	
	065	US-5876976	03-02-1999	RICHARDS; Rodney M. et al.	
	066	US-5882867	03-16-1999	ULLMAN; Edwin F. et al.	

Examiner Signature		Date Considered	

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant. [1] Applicant's unique citation designation number (optional). [2] See Kinds Codes of USPTO Patent Documents at www.uspto.gov or MPEP 901.04. [3] Enter Office that issued the document, by the two-letter code (WIPO Standard ST.3). [4] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [5] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. [6] Applicant is to place a check mark here if English language Translation is attached. This collection of information is required by 37 CFR 1.97 and 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 2 hours to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 (1-800-786-9199) and select option 2.

Approved for use through 11/30/2020. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Substitute for form 1449/PTO

INFORMATION DISCLOSURE STATEMENT BY APPLICANT

(Use as many sheets as necessary)

Sheet 4 of 61

Complete if Known	
Application Number	16017340
Filing Date	06-25-2018
First Named Inventor	AMORESE; Doug
Art Unit	Not Yet Assigned
Examiner Name	Not Yet Assigned
Attorney Docket Number	25115-755.302

U. S. PATENT DOCUMENTS

Examiner Initials*	Cite No.[1]	Document Number Number-Kind Code[2] (if known)	Publication Date MM-DD-YYYY	Name of Patentee or Applicant of Cited Document	Pages, Columns, Lines, Where Relevant Passages or Relevant Figures Appear
	067	US-5888779	03-30-1999	KACIAN; Daniel Louis et al.	
	068	US-5888819	03-30-1999	GOELET; Philip et al.	
	069	US-5945313	08-31-1999	HARTLEY; James L. et al.	
	070	US-5952176	09-14-1999	MCCARTHY; Thomas Valentine et al.	
	071	US-5958681	09-28-1999	WETMUR; James G. et al.	
	072	US-5965409	10-12-1999	PARDEE; Arthur B. et al.	
	073	US-5969119	10-19-1999	MACEVICZ; Stephen C.	
	074	US-5972618	10-26-1999	BLOCH; Will	
	075	US-6004744	12-21-1999	GOELET; Philip et al.	
	076	US-6004745	12-21-1999	ARNOLD Jr.; Lyle J. et al.	
	077	US-6027889	02-22-2000	BARANY; Francis et al.	
	078	US-6027923	02-22-2000	WALLACE; Robert Bruce	
	079	US-6030774	02-29-2000	LANEY; Maureen et al.	
	080	US-6037152	03-14-2000	RICHARDS; Rodney M. et al.	
	081	US-6056661	05-02-2000	SCHMIDT; Michael Roland	
	082	US-6077674	06-20-2000	SCHLEIFER; Arthur et al.	
	083	US-6087103	07-11-2000	BURMER; Glenna C.	
	084	US-6090553	07-18-2000	MATSON; Robert S.	
	085	US-6090591	07-18-2000	BURG; James Lawrence et al.	
	086	US-6107023	08-22-2000	REYES; Gregory R. et al.	
	087	US-6110709	08-29-2000	AUSUBEL; Frederick M. et al.	

Examiner Signature		Date Considered	

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant. [1] Applicant's unique citation designation number (optional). [2] See Kinds Codes of USPTO Patent Documents at www.uspto.gov or MPEP 901.04. [3] Enter Office that issued the document, by the two-letter code (WIPO Standard ST.3). [4] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [5] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. [6] Applicant is to place a check mark here if English language Translation is attached. This collection of information is required by 37 CFR 1.97 and 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 2 hours to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 (1-800-786-9199) and select option 2.

Approved for use through 11/30/2020. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Substitute for form 1449/PTO

INFORMATION DISCLOSURE STATEMENT BY APPLICANT

(Use as many sheets as necessary)

Sheet 5 of 61

Complete if Known	
Application Number	16017340
Filing Date	06-25-2018
First Named Inventor	AMORESE; Doug
Art Unit	Not Yet Assigned
Examiner Name	Not Yet Assigned
Attorney Docket Number	25115-755.302

U. S. PATENT DOCUMENTS

Examiner Initials*	Cite No.[1]	Document Number Number-Kind Code[2] (if known)	Publication Date MM-DD-YYYY	Name of Patentee or Applicant of Cited Document	Pages, Columns, Lines, Where Relevant Passages or Relevant Figures Appear
	088	US-6150112	11-21-2000	WEISSMAN; Sherman et al.	
	089	US-6159685	12-12-2000	PINKEL; Daniel et al.	
	090	US-6160105	12-12-2000	CUNNINGHAM; Mary Jane et al.	
	091	US-6169194	01-02-2001	THOMPSON; Michael et al.	
	092	US-6172208	01-09-2001	COOK; Alan Frederick	
	093	US-6174680	01-16-2001	MAKRIGIORGOS; Gerassimos M.	
	094	US-6190865	02-20-2001	JENDRISAK; Jerome J. et al.	
	095	US-6194211	02-27-2001	RICHARDS; Cynthia Ann et al.	
	096	US-6197557	03-06-2001	MAKAROV; Vladimir L. et al.	
	097	US-6210891	04-03-2001	NYREN; Paal et al.	
	098	US-6225109	05-01-2001	JUNCOSA; Robert D. et al.	
	099	US-6225451	05-01-2001	BALLINGER; Dennis G. et al.	
	100	US-6232104	05-15-2001	LISHANSKI; Alla et al.	
	101	US-6251639	06-26-2001	KURN; Nurith	
	102	US-6262490	07-17-2001	HSU; Kao-Yu et al.	
	103	US-6270961	08-07-2001	DRMANAC; Radoje	
	104	US-6280935	08-28-2001	MACEVICZ; Stephen C.	
	105	US-6287766	09-11-2001	NOLAN; John P. et al.	
	106	US-6287825	09-11-2001	WEISSMAN; Sherman et al.	
	107	US-6291170	09-18-2001	VAN; Gelder Russell N. et al.	
	108	US-6306365	10-23-2001	RUOSLAHTI; Erkki et al.	

Examiner Signature		Date Considered	

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant. [1] Applicant's unique citation designation number (optional). [2] See Kinds Codes of USPTO Patent Documents at www.uspto.gov or MPEP 901.04. [3] Enter Office that issued the document, by the two-letter code (WIPO Standard ST.3). [4] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [5] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. [6] Applicant is to place a check mark here if English language Translation is attached. This collection of information is required by 37 CFR 1.97 and 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 2 hours to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 (1-800-786-9199) and select option 2.

Approved for use through 11/30/2020. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Substitute for form 1449/PTO

INFORMATION DISCLOSURE STATEMENT BY APPLICANT

(Use as many sheets as necessary)

Sheet 6 of 61

Complete if Known	
Application Number	16017340
Filing Date	06-25-2018
First Named Inventor	AMORESE; Doug
Art Unit	Not Yet Assigned
Examiner Name	Not Yet Assigned
Attorney Docket Number	25115-755.302

U. S. PATENT DOCUMENTS

Examiner Initials*	Cite No.[1]	Document Number Number-Kind Code[2] (if known)	Publication Date MM-DD-YYYY	Name of Patentee or Applicant of Cited Document	Pages, Columns, Lines, Where Relevant Passages or Relevant Figures Appear
	109	US-6306597	10-23-2001	MACEVICZ; Stephen C.	
	110	US-6309843	10-30-2001	TIMMS; Kathy Lynn	
	111	US-6326142	12-04-2001	ROYER; Catherine A.	
	112	US-6335167	01-01-2002	PINKEL; Daniel et al.	
	113	US-6339147	01-15-2002	LUKHTANOV; Eugeny Alexander et al.	
	114	US-6440705	08-27-2002	STANTON Jr.; Vincent P. et al.	
	115	US-6449562	09-10-2002	CHANDLER; Van S. et al.	
	116	US-6582938	06-24-2003	SU; Xing et al.	
	117	US-6670461	12-30-2003	WENGEL; Jesper et al.	
	118	US-6686156	02-03-2004	KURN; Nurith	
	119	US-6692918	02-17-2004	KURN; Nurith	
	120	US-6770748	08-03-2004	IMANISHI; Takeshi et al.	
	121	US-6777180	08-17-2004	FISHER; Paul B. et al.	
	122	US-6794499	09-21-2004	WENGEL; Jesper et al.	
	123	US-6815164	11-09-2004	KURN; Nurith	
	124	US-6815167	11-09-2004	CROTHERS; Donald M. et al.	
	125	US-6825011	11-30-2004	ROMANTCHIKOV; Yuri	
	126	US-6833246	12-21-2004	BALASUBRAMANIAN; Shankar	
	127	US-6849404	02-01-2005	PARK; Han Oh [Kr] et al.	
	128	US-6858413	02-22-2005	KURN; Nurith	
	129	US-6917726	07-12-2005	LEVENE; Michael J. et al.	

Examiner Signature		Date Considered	

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant. [1] Applicant's unique citation designation number (optional). [2] See Kinds Codes of USPTO Patent Documents at www.uspto.gov or MPEP 901.04. [3] Enter Office that issued the document, by the two-letter code (WIPO Standard ST.3). [4] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [5] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. [6] Applicant is to place a check mark here if English language Translation is attached. This collection of information is required by 37 CFR 1.97 and 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 2 hours to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 (1-800-786-9199) and select option 2.

Approved for use through 11/30/2020. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Substitute for form 1449/PTO

INFORMATION DISCLOSURE STATEMENT BY APPLICANT

(Use as many sheets as necessary)

Sheet 7 of 61

Complete if Known	
Application Number	16017340
Filing Date	06-25-2018
First Named Inventor	AMORESE; Doug
Art Unit	Not Yet Assigned
Examiner Name	Not Yet Assigned
Attorney Docket Number	25115-755.302

U. S. PATENT DOCUMENTS

Examiner Initials*	Cite No.[1]	Document Number Number-Kind Code[2] (if known)	Publication Date MM-DD-YYYY	Name of Patentee or Applicant of Cited Document	Pages, Columns, Lines, Where Relevant Passages or Relevant Figures Appear
	130	US-6924104	08-02-2005	WEISMANN; Sherman et al.	
	131	US-6946251	09-20-2005	KURN; Nurith	
	132	US-7001724	02-21-2006	GREENFIELD; I. Lawrence	
	133	US-7033764	04-25-2006	KORLACH; Jonas et al.	
	134	US-7048481	05-23-2006	SUGATA; Shinsuke et al.	
	135	US-7052847	05-30-2006	KORLACH; Jonas et al.	
	136	US-7056676	06-06-2006	KORLACH; Jonas et al.	
	137	US-7060441	06-13-2006	BOURGET; Cecile et al.	
	138	US-7094536	08-22-2006	KURN; Nurith	
	139	US-7170050	01-30-2007	TURNER; Stephen et al.	
	140	US-7175982	02-13-2007	MCCARTHY; Thomas Valentine et al.	
	141	US-7176025	02-13-2007	KURN; Nurith et al.	
	142	US-7189512	03-13-2007	PORAT; Noga et al.	
	143	US-7211390	05-01-2007	ROTHBERG; Jonathan M. et al.	
	144	US-7244559	07-17-2007	ROTHBERG; Jonathan M. et al.	
	145	US-7244567	07-17-2007	CHEN; Yi-Ju et al.	
	146	US-7273730	09-25-2007	DU; Breuil Lastrucci Rusla M.	
	147	US-7276720	10-02-2007	ULMER; Kevin	
	148	US-7300755	11-27-2007	PETERSDORF; Effie W. et al.	
	149	US-7302146	11-27-2007	TURNER; Stephen et al.	
	150	US-7313308	12-25-2007	TURNER; Stephen et al.	

Examiner Signature		Date Considered	

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant. [1] Applicant's unique citation designation number (optional). [2] See Kinds Codes of USPTO Patent Documents at www.uspto.gov or MPEP 901.04. [3] Enter Office that issued the document, by the two-letter code (WIPO Standard ST.3). [4] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [5] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. [6] Applicant is to place a check mark here if English language Translation is attached. This collection of information is required by 37 CFR 1.97 and 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 2 hours to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 (1-800-786-9199) and select option 2.

Approved for use through 11/30/2020. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Substitute for form 1449/PTO

INFORMATION DISCLOSURE STATEMENT BY APPLICANT

(Use as many sheets as necessary)

Sheet 8 of 61

Complete if Known	
Application Number	16017340
Filing Date	06-25-2018
First Named Inventor	AMORESE; Doug
Art Unit	Not Yet Assigned
Examiner Name	Not Yet Assigned
Attorney Docket Number	25115-755.302

U. S. PATENT DOCUMENTS

Examiner Initials*	Cite No.[1]	Document Number Number-Kind Code[2] (if known)	Publication Date MM-DD-YYYY	Name of Patentee or Applicant of Cited Document	Pages, Columns, Lines, Where Relevant Passages or Relevant Figures Appear
	151	US-7315019	01-01-2008	TURNER; Stephen et al.	
	152	US-7323305	01-29-2008	LEAMON; John H. et al.	
	153	US-7335762	02-26-2008	ROTHBERG; Jonathan M. et al.	
	154	US-7361466	04-22-2008	KORLACH; Jonas et al.	
	155	US-7361468	04-22-2008	LIU; Guoying et al.	
	156	US-7402386	07-22-2008	KURN; Nurith et al.	
	157	US-7405281	07-29-2008	XU; Yue et al.	
	158	US-7414117	08-19-2008	SAITO; Isao et al.	
	159	US-7416844	08-26-2008	KORLACH; Jonas et al.	
	160	US-7462452	12-09-2008	WILLIAMS; John G. et al.	
	161	US-7462468	12-09-2008	WILLIAMS; John G. K. et al.	
	162	US-7476503	01-13-2009	TURNER; Stephen et al.	
	163	US-7476504	01-13-2009	TURNER; Stephen	
	164	US-7491498	02-17-2009	LAPIDUS; Stanley et al.	
	165	US-7501245	03-10-2009	QUAKE; Stephen et al.	
	166	US-7579153	08-25-2009	BRENNER; Sydney et al.	
	167	US-7741463	06-22-2010	GORMLEY; Niall A. et al.	
	168	US-7803550	09-28-2010	MAKAROV; Vladimir L. et al.	
	169	US-8017335	09-13-2011	SMITH; Geoffrey Paul	
	170	US-8053192	11-08-2011	BIGNELL; Helen et al.	
	171	US-8143001	03-27-2012	KURN; Nurith et al.	
	172	US-8551709	10-08-2013	KURN; Nurith et al.	

Examiner Signature		Date Considered	

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant. [1] Applicant's unique citation designation number (optional). [2] See Kinds Codes of USPTO Patent Documents at www.uspto.gov or MPEP 901.04. [3] Enter Office that issued the document, by the two-letter code (WIPO Standard ST.3). [4] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [5] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. [6] Applicant is to place a check mark here if English language Translation is attached. This collection of information is required by 37 CFR 1.97 and 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 2 hours to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 (1-800-786-9199) and select option 2.

Approved for use through 11/30/2020. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Substitute for form 1449/PTO

INFORMATION DISCLOSURE STATEMENT BY APPLICANT

(Use as many sheets as necessary)

Sheet 9 of 61

Complete if Known	
Application Number	16017340
Filing Date	06-25-2018
First Named Inventor	AMORESE; Doug
Art Unit	Not Yet Assigned
Examiner Name	Not Yet Assigned
Attorney Docket Number	25115-755.302

U. S. PATENT DOCUMENTS

Examiner Initials*	Cite No.[1]	Document Number Number-Kind Code[2] (if known)	Publication Date MM-DD-YYYY	Name of Patentee or Applicant of Cited Document	Pages, Columns, Lines, Where Relevant Passages or Relevant Figures Appear
	173	US-9284602	03-15-2016	ZHANG; David Yu et al.	
	174	US-20010000077	03-29-2001	ENGELHARDT; Dean L. et al.	
	175	US-20010031739	10-18-2001	DARE; Akintade Oyedele	
	176	US-20010034048	10-25-2001	KURN; Nurith	
	177	US-20010041334	11-15-2001	RASHTCHIAN; Ayoub et al.	
	178	US-20020028447	03-07-2002	LI; Wu Bo et al.	
	179	US-20020058270	05-16-2002	KURN; Nurith	
	180	US-20020115088	08-22-2002	KURN; Nurith	
	181	US-20020150919	10-17-2002	WEISMANN; Sherman et al.	
	182	US-20020155451	10-24-2002	MAKRIGIORGOS; Gerassimos M.	
	183	US-20020164628	11-07-2002	KURN; Nurith	
	184	US-20020164634	11-07-2002	PATIL; Nila et al.	
	185	US-20020197639	12-26-2002	SHIA; Michael A. et al.	
	186	US-20030017591	01-23-2003	KURN; Nurith	
	187	US-20030022207	01-30-2003	BALASUBRAMANIAN; Shankar et al.	
	188	US-20030082543	05-01-2003	SU; Xing et al.	
	189	US-20030087251	05-08-2003	KURN; Nurith	
	190	US-20030119150	06-26-2003	ANKENBAUER; Waltraud et al.	
	191	US-20030143555	07-31-2003	BOURGET; Cecile et al.	
	192	US-20030175780	09-18-2003	JONES; Douglas H.	
	193	US-20030180779	09-25-2003	LOFTON-DAY CATHY E . et al.	

Examiner Signature		Date Considered	

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant. [1] Applicant's unique citation designation number (optional). [2] See Kinds Codes of USPTO Patent Documents at www.uspto.gov or MPEP 901.04. [3] Enter Office that issued the document, by the two-letter code (WIPO Standard ST.3). [4] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [5] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. [6] Applicant is to place a check mark here if English language Translation is attached. This collection of information is required by 37 CFR 1.97 and 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 2 hours to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 (1-800-786-9199) and select option 2.

Approved for use through 11/30/2020. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Substitute for form 1449/PTO

INFORMATION DISCLOSURE STATEMENT BY APPLICANT

(Use as many sheets as necessary)

Sheet 10 of 61

Complete if Known	
Application Number	16017340
Filing Date	06-25-2018
First Named Inventor	AMORESE; Doug
Art Unit	Not Yet Assigned
Examiner Name	Not Yet Assigned
Attorney Docket Number	25115-755.302

U. S. PATENT DOCUMENTS

Examiner Initials*	Cite No.[1]	Document Number Number-Kind Code[2] (if known)	Publication Date MM-DD-YYYY	Name of Patentee or Applicant of Cited Document	Pages, Columns, Lines, Where Relevant Passages or Relevant Figures Appear
	194	US-20030186234	10-02-2003	KURN; Nurith	
	195	US-20030207279	11-06-2003	CROTHERS; Donald M. et al.	
	196	US-20030215926	11-20-2003	KURN; Nurith et al.	
	197	US-20030224439	12-04-2003	LAFFERTY; Mike et al.	
	198	US-20030232348	12-18-2003	JONES; Keith W. et al.	
	199	US-20040002371	01-01-2004	PAQUIN; Claude et al.	
	200	US-20040005614	01-08-2004	KURN; Nurith et al.	
	201	US-20040023271	02-05-2004	KURN; Nurith et al.	
	202	US-20040115815	06-17-2004	LI; Henry et al.	
	203	US-20040137456	07-15-2004	YOKOTA; Hiroki et al.	
	204	US-20040161742	08-19-2004	DEAN; Frank B. et al.	
	205	US-20040203019	10-14-2004	KURN; Nurith	
	206	US-20040203025	10-14-2004	KURN; Nurith	
	207	US-20040248153	12-09-2004	DEAR; Paul H. et al.	
	208	US-20050003441	01-06-2005	KURN; Nurith	
	209	US-20050014192	01-20-2005	KURN; Nurith	
	210	US-20050019793	01-27-2005	KURN; Nurith et al.	
	211	US-20050059048	03-17-2005	GUNDERSON; Kevin et al.	
	212	US-20050064456	03-24-2005	KURN; Nurith	
	213	US-20050123956	06-09-2005	BLUME; John E. et al.	
	214	US-20050136417	06-23-2005	COLE; Kyle B. et al.	
	215	US-20050142577	06-30-2005	JONES; Keith W. et al.	

Examiner Signature		Date Considered	

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant. [1] Applicant's unique citation designation number (optional). [2] See Kinds Codes of USPTO Patent Documents at www.uspto.gov or MPEP 901.04. [3] Enter Office that issued the document, by the two-letter code (WIPO Standard ST.3). [4] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [5] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. [6] Applicant is to place a check mark here if English language Translation is attached. This collection of information is required by 37 CFR 1.97 and 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 2 hours to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 (1-800-786-9199) and select option 2.

Approved for use through 11/30/2020. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Substitute for form 1449/PTO

INFORMATION DISCLOSURE STATEMENT BY APPLICANT

(Use as many sheets as necessary)

Sheet | 11 | of | 61

Complete if Known	
Application Number	16017340
Filing Date	06-25-2018
First Named Inventor	AMORESE; Doug
Art Unit	Not Yet Assigned
Examiner Name	Not Yet Assigned
Attorney Docket Number	25115-755.302

U. S. PATENT DOCUMENTS

Examiner Initials*	Cite No.[1]	Document Number Number-Kind Code[2] (if known)	Publication Date MM-DD-YYYY	Name of Patentee or Applicant of Cited Document	Pages, Columns, Lines, Where Relevant Passages or Relevant Figures Appear
	216	US-20050191656	09-01-2005	DRMANAC; Radoje et al.	
	217	US-20050191682	09-01-2005	BARONE; Anthony D. et al.	
	218	US-20050208538	09-22-2005	KURN; Nurith et al.	
	219	US-20060008824	01-12-2006	RONAGHI; Mostafa et al.	
	220	US-20060014182	01-19-2006	KURN; Nurith	
	221	US-20060024678	02-02-2006	BUZBY; Philip R.	
	222	US-20060024711	02-02-2006	LAPIDUS; Stanley N. et al.	
	223	US-20060035274	02-16-2006	DONG; Shoulian	
	224	US-20060046251	03-02-2006	SAMPSON; Jeffrey R. et al.	
	225	US-20060051789	03-09-2006	KAZAKOV; Sergei A. et al.	
	226	US-20060216724	09-28-2006	CHRISTIANS; Frederick C. et al.	
	227	US-20060263789	11-23-2006	KINCAID; Robert	
	228	US-20060281082	12-14-2006	ZHU; Jiahui	
	229	US-20060286566	12-21-2006	LAPIDUS; Stanley N. et al.	
	230	US-20060292597	12-28-2006	SHAPERO; Michael H. et al.	
	231	US-20070031857	02-08-2007	MAKAROV; Vladimir L. et al.	
	232	US-20070134128	06-14-2007	KORLACH; Jonas	
	233	US-20070224613	09-27-2007	STRATHMANN; Michael P.	
	234	US-20070231823	10-04-2007	MCKERNAN; Kevin J. et al.	
	235	US-20070238122	10-11-2007	ALLBRITTON; Nancy et al.	
	236	US-20080038727	02-14-2008	SPIER; Eugene G.	
	237	US-20080087826	04-17-2008	HARRIS; Timothy D. et al.	

Examiner Signature		Date Considered	

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant. [1] Applicant's unique citation designation number (optional). [2] See Kinds Codes of USPTO Patent Documents at www.uspto.gov or MPEP 901.04. [3] Enter Office that issued the document, by the two-letter code (WIPO Standard ST.3). [4] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [5] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. [6] Applicant is to place a check mark here if English language Translation is attached. This collection of information is required by 37 CFR 1.97 and 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 2 hours to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 (1-800-786-9199) and select option 2.

Approved for use through 11/30/2020. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Substitute for form 1449/PTO

INFORMATION DISCLOSURE STATEMENT BY APPLICANT

(Use as many sheets as necessary)

Sheet | 12 | of | 61

Complete if Known	
Application Number	16017340
Filing Date	06-25-2018
First Named Inventor	AMORESE; Doug
Art Unit	Not Yet Assigned
Examiner Name	Not Yet Assigned
Attorney Docket Number	25115-755.302

U. S. PATENT DOCUMENTS

Examiner Initials*	Cite No.[1]	Document Number Number-Kind Code[2] (if known)	Publication Date MM-DD-YYYY	Name of Patentee or Applicant of Cited Document	Pages, Columns, Lines, Where Relevant Passages or Relevant Figures Appear
	238	US-20080103058	05-01-2008	SIDDIQI; Suhaib	
	239	US-20080176311	07-24-2008	KURN; Nurith	
	240	US-20080182300	07-31-2008	KURN; Nurith	
	241	US-20080194413	08-14-2008	ALBERT; Thomas J.	
	242	US-20080194416	08-14-2008	CHEN; Fuqiang	
	243	US-20080206764	08-28-2008	WILLIAMS; John G.K. et al.	
	244	US-20080213770	09-04-2008	WILLIAMS; Peter et al.	
	245	US-20080242560	10-02-2008	GUNDERSON; Kevin L. et al.	
	246	US-20080286795	11-20-2008	KAWASHIMA; Eric H. et al.	
	247	US-20090011959	01-08-2009	COSTA; Gina L. et al.	
	248	US-20090024331	01-22-2009	TOMANEY; Austin B. et al.	
	249	US-20090026082	01-29-2009	ROTHBERG; Jonathan M. et al.	
	250	US-20090029385	01-29-2009	CHRISTIANS; Fred et al.	
	251	US-20090061425	03-05-2009	LO; Yuk Ming Dennis et al.	
	252	US-20090061439	03-05-2009	BUZBY; Philip Richard	
	253	US-20090068645	03-12-2009	SIBSON; Ross	
	254	US-20090068655	03-12-2009	WILLIAMS; John G.K. et al.	
	255	US-20090068709	03-12-2009	KURN; Nurith et al.	
	256	US-20090117573	05-07-2009	FU; Glenn K. et al.	
	257	US-20090117621	05-07-2009	BOUTELL; Jonathan Mark et al.	
	258	US-20090124514	05-14-2009	FU; Glenn et al.	
	259	US-20090127589	05-21-2009	ROTHBERG; Jonathan M. et al.	

Examiner Signature		Date Considered	

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant. [1] Applicant's unique citation designation number (optional). [2] See Kinds Codes of USPTO Patent Documents at www.uspto.gov or MPEP 901.04. [3] Enter Office that issued the document, by the two-letter code (WIPO Standard ST.3). [4] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [5] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. [6] Applicant is to place a check mark here if English language Translation is attached. This collection of information is required by 37 CFR 1.97 and 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 2 hours to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 (1-800-786-9199) and select option 2.

Approved for use through 11/30/2020. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Substitute for form 1449/PTO

INFORMATION DISCLOSURE STATEMENT BY APPLICANT

(Use as many sheets as necessary)

Sheet | 13 | of | 61

Complete if Known	
Application Number	16017340
Filing Date	06-25-2018
First Named Inventor	AMORESE; Doug
Art Unit	Not Yet Assigned
Examiner Name	Not Yet Assigned
Attorney Docket Number	25115-755.302

U. S. PATENT DOCUMENTS

Examiner Initials*	Cite No.[1]	Document Number Number-Kind Code[2] (if known)	Publication Date MM-DD-YYYY	Name of Patentee or Applicant of Cited Document	Pages, Columns, Lines, Where Relevant Passages or Relevant Figures Appear
	260	US-20090130721	05-21-2009	KURN; Nurith et al.	
	261	US-20090203085	08-13-2009	KURN; Nurith et al.	
	262	US-20090203531	08-13-2009	KURN; Nurith	
	263	US-20090233804	09-17-2009	KURN; Nurith et al.	
	264	US-20090239232	09-24-2009	KURN; Nurith	
	265	US-20090275486	11-05-2009	KURN; Nurith et al.	
	266	US-20090280538	11-12-2009	PATEL; Pranav et al.	
	267	US-20090298075	12-03-2009	TRAVERS; Kevin [Us] et al.	
	268	US-20100015666	01-21-2010	BRENNER; Sydney et al.	
	269	US-20100021973	01-28-2010	MAKAROV; Vladimir L. et al.	
	270	US-20100022403	01-28-2010	KURN; Nurith et al.	
	271	US-20100029511	02-04-2010	RAYMOND; Christopher K. et al.	
	272	US-20100105052	04-29-2010	DRMANAC; Radoje et al.	
	273	US-20100113296	05-06-2010	MYERSON; Joel	
	274	US-20100129879	05-27-2010	ACH; Robert A. et al.	
	275	US-20100137143	06-03-2010	ROTHBERG; Jonathan M. et al.	
	276	US-20100167954	07-01-2010	EARNSHAW; David James et al.	
	277	US-20100173394	07-08-2010	COLSTON Jr.; Billy Wayne et al.	
	278	US-20100273219	10-28-2010	MAY; Andrew et al.	
	279	US-20100323348	12-23-2010	HAMADY; Micah L [Us] et al.	
	280	US-20110015096	01-20-2011	CHIU; Kuo Ping	
	281	US-20110039732	02-17-2011	RAYMOND; Christopher et al.	

Examiner Signature		Date Considered	

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant. [1] Applicant's unique citation designation number (optional). [2] See Kinds Codes of USPTO Patent Documents at www.uspto.gov or MPEP 901.04. [3] Enter Office that issued the document, by the two-letter code (WIPO Standard ST.3). [4] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [5] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. [6] Applicant is to place a check mark here if English language Translation is attached. This collection of information is required by 37 CFR 1.97 and 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 2 hours to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 (1-800-786-9199) and select option 2.

Approved for use through 11/30/2020. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Substitute for form 1449/PTO

INFORMATION DISCLOSURE STATEMENT BY APPLICANT

(Use as many sheets as necessary)

Sheet | 14 | of | 61

Complete if Known	
Application Number	16017340
Filing Date	06-25-2018
First Named Inventor	AMORESE; Doug
Art Unit	Not Yet Assigned
Examiner Name	Not Yet Assigned
Attorney Docket Number	25115-755.302

U. S. PATENT DOCUMENTS

Examiner Initials*	Cite No.[1]	Document Number Number-Kind Code[2] (if known)	Publication Date MM-DD-YYYY	Name of Patentee or Applicant of Cited Document	Pages, Columns, Lines, Where Relevant Passages or Relevant Figures Appear
	282	US-20110104785	05-05-2011	VAIDYANATHAN; Ramesh et al.	
	283	US-20110105364	05-05-2011	KURN; Nurith	
	284	US-20110129827	06-02-2011	CAUSEY; Marie et al.	
	285	US-20110189679	08-04-2011	KURN; Nurith et al.	
	286	US-20110224105	09-15-2011	KURN; Nurith et al.	
	287	US-20110288780	11-24-2011	RABINOWITZ; Matthew et al.	
	288	US-20110319290	12-29-2011	RAYMOND; Christopher et al.	
	289	US-20120045797	02-23-2012	KURN; Nurith et al.	
	290	US-20120071331	03-22-2012	CASBON; James et al.	
	291	US-20120074925	03-29-2012	OLIVER; John S.	
	292	US-20120102054	04-26-2012	POPESCU; Liviu et al.	
	293	US-20120122701	05-17-2012	RYAN; Allison et al.	
	294	US-20120157322	06-21-2012	MYLLYKANGAS; Samuel et al.	
	295	US-20120220483	08-30-2012	KURN; Nurith et al.	
	296	US-20120237943	09-20-2012	SOLDATOV; Aleksey et al.	
	297	US-20120245041	09-27-2012	BRENNER; Sydney et al.	
	298	US-20120270212	10-25-2012	RABINOWITZ; Matthew [Us] et al.	
	299	US-20120283145	11-08-2012	WANG; Yan	
	300	US-20120309002	12-06-2012	LINK; Darren	
	301	US-20130059738	03-07-2013	LEAMON; John et al.	
	302	US-20140065692	03-06-2014	KURN; Nurith et al.	

Examiner Signature		Date Considered	

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant. [1] Applicant's unique citation designation number (optional). [2] See Kinds Codes of USPTO Patent Documents at www.uspto.gov or MPEP 901.04. [3] Enter Office that issued the document, by the two-letter code (WIPO Standard ST.3). [4] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [5] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. [6] Applicant is to place a check mark here if English language Translation is attached. This collection of information is required by 37 CFR 1.97 and 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 2 hours to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 (1-800-786-9199) and select option 2.

Approved for use through 11/30/2020. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Substitute for form 1449/PTO

INFORMATION DISCLOSURE STATEMENT BY APPLICANT

(Use as many sheets as necessary)

Sheet | 15 | of | 61

Complete if Known	
Application Number	16017340
Filing Date	06-25-2018
First Named Inventor	AMORESE; Doug
Art Unit	Not Yet Assigned
Examiner Name	Not Yet Assigned
Attorney Docket Number	25115-755.302

U. S. PATENT DOCUMENTS

Examiner Initials*	Cite No.[1]	Document Number Number-Kind Code[2] (if known)	Publication Date MM-DD-YYYY	Name of Patentee or Applicant of Cited Document	Pages, Columns, Lines, Where Relevant Passages or Relevant Figures Appear
	303	US-20140274729	09-18-2014	KURN; Nurith et al.	
	304	US-20140274731	09-18-2014	RAYMOND; Christopher K [Us] et al.	
	305	US-20140274738	09-18-2014	AMORESE; Doug et al.	
	306	US-20140378345	12-25-2014	HINDSON; Benjamin et al.	
	307	US-20150011396	01-08-2015	SCHROEDER; Benjamin G. et al.	
	308	US-20150017635	01-15-2015	MYLLYKANGAS; Samuel et al.	
	309	US-20150284769	10-08-2015	SCHROEDER; Benjamin G. et al.	
	310	US-20150299767	10-22-2015	ARMOUR; Christopher et al.	
	311	US-20160040215	02-11-2016	HENN; Matthew R. et al.	
	312	US-20160130576	05-12-2016	ARMOUR; Chris	
	313	US-20160251712	09-01-2016	AMORESE; Doug et al.	
	314	US-20160265042	09-15-2016	SCHROEDER; Benjamin G. et al.	
	315	US-20160362680-A1	12-15-2016	CHRISTOPHER; Armour et al.	

FOREIGN PATENT DOCUMENTS

Examiner Initials*	Cite No.[1]	Foreign Patent Document Country Code[3]-Number[4]-Kind Code[5] (if known)	Publication Date MM-DD-YYYY	Name of Patentee or Applicant of Cited Document	Pages, Columns, Lines, Where Relevant Passages Or Relevant Figures Appear	T[6]
	001	CA-2444926-A1	11-14-2002	CENTRE NAT RECH SCIENT [FR], et al.		☒

Examiner Signature		Date Considered	

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant. [1] Applicant's unique citation designation number (optional). [2] See Kinds Codes of USPTO Patent Documents at www.uspto.gov or MPEP 901.04. [3] Enter Office that issued the document, by the two-letter code (WIPO Standard ST.3). [4] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [5] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. [6] Applicant is to place a check mark here if English language Translation is attached. This collection of information is required by 37 CFR 1.97 and 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 2 hours to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 (1-800-786-9199) and select option 2.

Approved for use through 11/30/2020. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Substitute for form 1449/PTO

INFORMATION DISCLOSURE STATEMENT BY APPLICANT

(Use as many sheets as necessary)

Sheet | 16 | of | 61

Complete if Known	
Application Number	16017340
Filing Date	06-25-2018
First Named Inventor	AMORESE; Doug
Art Unit	Not Yet Assigned
Examiner Name	Not Yet Assigned
Attorney Docket Number	25115-755.302

FOREIGN PATENT DOCUMENTS

Examiner Initials*	Cite No.[1]	Foreign Patent Document Country Code[3]-Number[4]-Kind Code[5] (if known)	Publication Date MM-DD-YYYY	Name of Patentee or Applicant of Cited Document	Pages, Columns, Lines, Where Relevant Passages Or Relevant Figures Appear	T[6]
	002	EP-0329822-B1	06-08-1994	CANGENE CORP [CA]		☐
	003	EP-0365627-B1	12-22-1993	UNIV IOWA RES FOUND [US]		☐
	004	EP-0667393-A2	08-16-1995	ENZO DIAGNOSTICS INC [US]		☐
	005	EP-0667393-A3	11-15-1995	ENZO DIAGNOSTICS INC [US]		☐
	006	EP-0843735-B1	07-17-2002	FORFAS TRADING AS BIORES IRELA [IE], et al.		☐
	007	EP-1071811-B1	03-06-2002	ENTPR IE TRD AS BIORESEARCH IE [IE], et al.		☐
	008	EP-2322612-A1	05-18-2011	HITACHI HIGH TECH CORP [JP], et al.		☐
	009	WO-0008208-A2	02-17-2000	MEDICAL RES COUNCIL [GB], et al.		☐
	010	WO-0008208-A3	05-11-2000	MEDICAL RES COUNCIL [GB], et al.		☐
	011	WO-0009756-A1	02-24-2000	PERKIN ELMER CORP [US]		☐

Examiner Signature		Date Considered	

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant. [1] Applicant's unique citation designation number (optional). [2] See Kinds Codes of USPTO Patent Documents at www.uspto.gov or MPEP 901.04. [3] Enter Office that issued the document, by the two-letter code (WIPO Standard ST.3). [4] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [5] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. [6] Applicant is to place a check mark here if English language Translation is attached. This collection of information is required by 37 CFR 1.97 and 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 2 hours to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 (1-800-786-9199) and select option 2.

Approved for use through 11/30/2020. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Substitute for form 1449/PTO

INFORMATION DISCLOSURE STATEMENT BY APPLICANT

(Use as many sheets as necessary)

Sheet 17 of 61

Complete if Known	
Application Number	16017340
Filing Date	06-25-2018
First Named Inventor	AMORESE; Doug
Art Unit	Not Yet Assigned
Examiner Name	Not Yet Assigned
Attorney Docket Number	25115-755.302

FOREIGN PATENT DOCUMENTS

Examiner Initials*	Cite No.[1]	Foreign Patent Document Country Code[3]-Number[4]-Kind Code[5] (if known)	Publication Date MM-DD-YYYY	Name of Patentee or Applicant of Cited Document	Pages, Columns, Lines, Where Relevant Passages Or Relevant Figures Appear	T[6]
	012	WO-0018957-A1	04-06-2000	APPLIED RESEARCH SYSTEMS [NL], et al.		☐
	013	WO-0039345-A1	07-06-2000	DANA FARBER CANCER INST INC [US], et al.		☐
	014	WO-0052191-A1	09-08-2000	LIFE TECHNOLOGIES INC [US]		☐
	015	WO-0055364-A2	09-21-2000	AMERSHAM PHARM BIOTECH UK LTD [GB], et al.		☐
	016	WO-0055364-A3	10-11-2001	AMERSHAM PHARM BIOTECH UK LTD [GB], et al.		☐
	017	WO-0120035-A2	03-22-2001	NUGEN TECHNOLOGIES INC [US], et al.		☐
	018	WO-0120035-A3	12-06-2001	NUGEN TECHNOLOGIES INC [US], et al.		☐
	019	WO-0123613-A1	04-05-2001	ISIS PHARMACEUTICALS INC [US], et al.		☐
	020	WO-0146464-A1	06-28-2001	LION BIOSCIENCE AG [DE], et al.		☐
	021	WO-0157248-A2	08-09-2001	SOLEXA LTD [GB]		☐

Examiner Signature		Date Considered	

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant. [1] Applicant's unique citation designation number (optional). [2] See Kinds Codes of USPTO Patent Documents at www.uspto.gov or MPEP 901.04. [3] Enter Office that issued the document, by the two-letter code (WIPO Standard ST.3). [4] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [5] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. [6] Applicant is to place a check mark here if English language Translation is attached. This collection of information is required by 37 CFR 1.97 and 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 2 hours to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 (1-800-786-9199) and select option 2.

Approved for use through 11/30/2020. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Substitute for form 1449/PTO

INFORMATION DISCLOSURE STATEMENT BY APPLICANT

(Use as many sheets as necessary)

Sheet | 18 | of | 61

Complete if Known	
Application Number	16017340
Filing Date	06-25-2018
First Named Inventor	AMORESE; Doug
Art Unit	Not Yet Assigned
Examiner Name	Not Yet Assigned
Attorney Docket Number	25115-755.302

FOREIGN PATENT DOCUMENTS

Examiner Initials*	Cite No.[1]	Foreign Patent Document Country Code[3]–Number[4]–Kind Code[5] (if known)	Publication Date MM-DD-YYYY	Name of Patentee or Applicant of Cited Document	Pages, Columns, Lines, Where Relevant Passages Or Relevant Figures Appear	T[6]
	022	WO-0157248-A3	02-14-2002	SOLEXA LTD [GB]		☐
	023	WO-0164952-A2	09-07-2001	MOLECULAR STAGING INC [US]		☐
	024	WO-0164952-A3	12-27-2002	MOLECULAR STAGING INC [US]		☐
	025	WO-0200938-A2	01-03-2002	NUGEN TGECHNOLOGIES INC [US], et al.		☐
	026	WO-0200938-A3	08-28-2003	NUGEN TECHNOLOGIES INC [US], et al.		☐
	027	WO-02060318-A2	08-08-2002	UNIV OHIO STATE RES FOUND [US], et al.		☐
	028	WO-02060318-A3	10-09-2003	UNIV OHIO STATE RES FOUND [US], et al.		☐
	029	WO-02072772-A2	09-19-2002	NUGEN TECHNOLOGIES INC [US], et al.		☐
	030	WO-02072772-A3	09-12-2003	NUGEN TECHNOLOGIES INC [US], et al.		☐
	031	WO-02072773-A2	09-19-2002	NUGEN TECHNOLOGIES INC [US], et al.		☐

Examiner Signature		Date Considered	

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant. [1] Applicant's unique citation designation number (optional). [2] See Kinds Codes of USPTO Patent Documents at www.uspto.gov or MPEP 901.04. [3] Enter Office that issued the document, by the two-letter code (WIPO Standard ST.3). [4] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [5] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. [6] Applicant is to place a check mark here if English language Translation is attached. This collection of information is required by 37 CFR 1.97 and 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 2 hours to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 (1-800-786-9199) and select option 2.

Approved for use through 11/30/2020. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Substitute for form 1449/PTO

INFORMATION DISCLOSURE STATEMENT BY APPLICANT

(Use as many sheets as necessary)

Sheet | 19 | of | 61

Complete if Known	
Application Number	16017340
Filing Date	06-25-2018
First Named Inventor	AMORESE; Doug
Art Unit	Not Yet Assigned
Examiner Name	Not Yet Assigned
Attorney Docket Number	25115-755.302

FOREIGN PATENT DOCUMENTS

Examiner Initials*	Cite No.[1]	Foreign Patent Document Country Code[3]–Number[4]–Kind Code[5] (if known)	Publication Date MM-DD-YYYY	Name of Patentee or Applicant of Cited Document	Pages, Columns, Lines, Where Relevant Passages Or Relevant Figures Appear	T[6]
	032	WO-02072773-A3	12-18-2003	NUGEN TECHNOLOGIES INC [US], et al.		☐
	033	WO-02081753-A1	10-17-2002	ADVANCED RES & TECH INST [US], et al.		☐
	034	WO-02090584-A2	11-14-2002	BIO MERIEUX [FR], et al.		☒
	035	WO-02090584-A3	09-25-2003	BIO MERIEUX [FR], et al.		☒
	036	WO-0228876-A2	04-11-2002	RIKEN [JP], et al.		☐
	037	WO-0228876-A3	08-01-2002	RIKEN [JP], et al.		☐
	038	WO-0229117-A2	04-11-2002	NUGEN TECHNOLOGIES INC [US], et al.		☐
	039	WO-0229117-A3	08-14-2003	NUGEN TECHNOLOGIES INC [US], et al.		☐
	040	WO-0236821-A2	05-10-2002	UNIV COLLEGE CORK NAT UNIV IE [IE], et al.		☐
	041	WO-0236821-A3	03-27-2003	UNIV COLLEGE CORK NAT UNIV IE [IE], et al.		☐

Examiner Signature		Date Considered	

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant. [1] Applicant's unique citation designation number (optional). [2] See Kinds Codes of USPTO Patent Documents at www.uspto.gov or MPEP 901.04. [3] Enter Office that issued the document, by the two-letter code (WIPO Standard ST.3). [4] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [5] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. [6] Applicant is to place a check mark here if English language Translation is attached. This collection of information is required by 37 CFR 1.97 and 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 2 hours to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 (1-800-786-9199) and select option 2.

Approved for use through 11/30/2020. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Substitute for form 1449/PTO

INFORMATION DISCLOSURE STATEMENT BY APPLICANT

(Use as many sheets as necessary)

Sheet | 20 | of | 61

Complete if Known	
Application Number	16017340
Filing Date	06-25-2018
First Named Inventor	AMORESE; Doug
Art Unit	Not Yet Assigned
Examiner Name	Not Yet Assigned
Attorney Docket Number	25115-755.302

FOREIGN PATENT DOCUMENTS

Examiner Initials*	Cite No.[1]	Foreign Patent Document Country Code[3]–Number[4]–Kind Code[5] (if known)	Publication Date MM-DD-YYYY	Name of Patentee or Applicant of Cited Document	Pages, Columns, Lines, Where Relevant Passages Or Relevant Figures Appear	T[6]
	042	WO-0248402-A2	06-20-2002	NUGEN TECHNOLOGIES INC [US], et al.		☐
	043	WO-0248402-A3	04-08-2004	NUGEN TECHNOLOGIES INC [US], et al.		☐
	044	WO-03002736-A2	01-09-2003	ROCHE DIAGNOSTICS GMBH [DE], et al.		☐
	045	WO-03012118-A1	02-13-2003	AFFYMETRIX INC [US], et al.		☐
	046	WO-03027259-A2	04-03-2003	EPIGENX PHARMACEUTICALS INC [US], et al.		☐
	047	WO-03027259-A3	12-11-2003	EPIGENX PHARMACEUTICALS INC [US], et al.		☐
	048	WO-03078645-A2	09-25-2003	NUGEN TECHNOLOGIES INC [US], et al.		☐
	049	WO-03078645-A3	03-18-2004	NUGEN TECHNOLOGIES INC [US], et al.		☐
	050	WO-03083435-A2	10-09-2003	NUGEN TECHNOLOGIES INC [US], et al.		☐
	051	WO-03083435-A3	02-26-2004	NUGEN TECHNOLOGIES INC [US], et al.		☐

Examiner Signature		Date Considered	

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant. [1] Applicant's unique citation designation number (optional). [2] See Kinds Codes of USPTO Patent Documents at www.uspto.gov or MPEP 901.04. [3] Enter Office that issued the document, by the two-letter code (WIPO Standard ST.3). [4] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [5] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. [6] Applicant is to place a check mark here if English language Translation is attached. This collection of information is required by 37 CFR 1.97 and 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 2 hours to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 (1-800-786-9199) and select option 2.

Approved for use through 11/30/2020. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Substitute for form 1449/PTO

INFORMATION DISCLOSURE STATEMENT BY APPLICANT

(Use as many sheets as necessary)

Sheet | 21 | of | 61

Complete if Known	
Application Number	16017340
Filing Date	06-25-2018
First Named Inventor	AMORESE; Doug
Art Unit	Not Yet Assigned
Examiner Name	Not Yet Assigned
Attorney Docket Number	25115-755.302

FOREIGN PATENT DOCUMENTS

Examiner Initials*	Cite No.[1]	Foreign Patent Document Country Code[3]–Number[4]–Kind Code[5] (if known)	Publication Date MM-DD-YYYY	Name of Patentee or Applicant of Cited Document	Pages, Columns, Lines, Where Relevant Passages Or Relevant Figures Appear	T[6]
	052	WO-03106642-A2	12-24-2003	AFFYMETRIX INC [US], et al.		☐
	053	WO-03106642-A3	11-04-2004	AFFYMETRIX INC [US], et al.		☐
	054	WO-2004011665-A2	02-05-2004	NUGEN TECHNOLOGIES INC [US], et al.		☐
	055	WO-2004011665-A3	07-21-2005	NUGEN TECHNOLOGIES INC [US], et al.		☐
	056	WO-2005065321-A2	07-21-2005	NUGEN TECHNOLOGIES INC [US], et al.		☐
	057	WO-2005065321-A3	05-28-2009	NUGEN TECHNOLOGIES INC [US], et al.		☐
	058	WO-2006081222-A2	08-03-2006	COMPASS GENETICS LLC [US], et al.		☐
	059	WO-2006081222-A3	02-22-2007	COMPASS GENETICS LLC [US], et al.		☐
	060	WO-2007018601-A1	02-15-2007	RUBICON GENOMICS INC [US], et al.		☐
	061	WO-2007030759-A2	03-15-2007	NUGEN TECHNOLOGIES INC [US], et al.		☐

Examiner Signature		Date Considered	

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant. [1] Applicant's unique citation designation number (optional). [2] See Kinds Codes of USPTO Patent Documents at www.uspto.gov or MPEP 901.04. [3] Enter Office that issued the document, by the two-letter code (WIPO Standard ST.3). [4] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [5] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. [6] Applicant is to place a check mark here if English language Translation is attached. This collection of information is required by 37 CFR 1.97 and 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 2 hours to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 (1-800-786-9199) and select option 2.

Approved for use through 11/30/2020. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Substitute for form 1449/PTO

INFORMATION DISCLOSURE STATEMENT BY APPLICANT

(Use as many sheets as necessary)

Sheet | 22 | of | 61

Complete if Known	
Application Number	16017340
Filing Date	06-25-2018
First Named Inventor	AMORESE; Doug
Art Unit	Not Yet Assigned
Examiner Name	Not Yet Assigned
Attorney Docket Number	25115-755.302

FOREIGN PATENT DOCUMENTS

Examiner Initials*	Cite No.[1]	Foreign Patent Document Country Code[3]-Number[4]-Kind Code[5] (if known)	Publication Date MM-DD-YYYY	Name of Patentee or Applicant of Cited Document	Pages, Columns, Lines, Where Relevant Passages Or Relevant Figures Appear	T[6]
	062	WO-2007030759-A3	06-21-2007	NUGEN TECHNOLOGIES INC [US], et al.		☐
	063	WO-2007052006-A1	05-10-2007	SOLEXA LTD [GB], et al.		☐
	064	WO-2007057652-A1	05-24-2007	SOLEXA LTD [GB], et al.		☐
	065	WO-2007136717-A1	11-29-2007	NUGEN TECHNOLOGIES INC [US], et al.		☐
	066	WO-2008005459-A2	01-10-2008	NUGEN TECHNOLOGIES INC [US], et al.		☐
	067	WO-2008005459-A3	02-21-2008	NUGEN TECHNOLOGIES INC [US], et al.		☐
	068	WO-2008015396-A2	02-07-2008	SOLEXA LTD [GB], et al.		☐
	069	WO-2008033442-A2	03-20-2008	UNIV LELAND STANFORD JUNIOR [US], et al.		☐
	070	WO-2008033442-A3	10-16-2008	UNIV LELAND STANFORD JUNIOR [US], et al.		☐
	071	WO-2008115185-A2	09-25-2008	NIMBLEGEN SYSTEMS INC [US], et al.		☐

Examiner Signature		Date Considered	

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant. [1] Applicant's unique citation designation number (optional). [2] See Kinds Codes of USPTO Patent Documents at www.uspto.gov or MPEP 901.04. [3] Enter Office that issued the document, by the two-letter code (WIPO Standard ST.3). [4] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [5] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. [6] Applicant is to place a check mark here if English language Translation is attached. This collection of information is required by 37 CFR 1.97 and 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 2 hours to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 (1-800-786-9199) and select option 2.