(19) DANMARK

(10) DK/EP 3455372 T3

(12) Oversættelse af europæisk patentskrift

Patent- og Varemærkestyrelsen

(51) Int.Cl.: *C 12 Q 1/6844 (2018.01)*

(45) Oversættelsen bekendtgjort den: **2020-01-06**

(80) Dato for Den Europæiske Patentmyndigheds bekendtgørelse om meddelelse af patentet: **2019-10-30**

(86) Europæisk ansøgning nr.: **17725052.9**

(86) Europæisk indleveringsdag: **2017-05-10**

(87) Den europæiske ansøgnings publiceringsdag: **2019-03-20**

(86) International ansøgning nr.: **US2017032021**

(87) Internationalt publikationsnr.: **WO2017197027**

(30) Prioritet: **2016-05-11 US 201662335012 P**

(84) Designerede stater: **AL 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 RS SE SI SK SM TR**

(73) Patenthaver: **Illumina, Inc., 5200 Ilumina Way, San Diego, CA 92122, USA**

(72) Opfinder: **MANDELL, Jeffrey, G., 5200 Illumina Way, San Diego, CA 92122, USA**
HE, Molly, 5200 Illumina Way, San Diego, CA 92122, USA

(74) Fuldmægtig i Danmark: **MURGITROYD & COMPANY, Scotland House, 165-169 Scotland Street, Glasgow G5 8PL, Storbritannien**

(54) Benævnelse: **Polynukleotidberigelse og -amplifikation ved anvendelse af argonautsystemer**

(56) Fremdragne publikationer:
WO-A1-2016/077350
US-A1- 2015 211 058

DESCRIPTION

[0001] The present disclosure relates generally to methods for enriching or amplifying polynucleotides, and more specifically to methods for enriching or amplifying polynucleotides using CRISPR-Cas or Argonaute systems and applications thereof.

BACKGROUND

[0002] There are a variety of methods and applications for which it is desirable to enrich or amplify a target polynucleotide among a population of polynucleotides, *e.g.,* among whole genome.

[0003] Many of the methods currently used for sequence-specific DNA enrichment involve multiple steps and require relatively large amounts of sample nucleic acids, and usually are difficult, tedious, laborious, time-consuming, inefficient, and costly. In addition, methods currently used for targeted enrichment of double-stranded DNA require creating a single-stranded DNA prior to the sequence specific targeting. They also require longer time for hybridizing probes to target DNA. Thus, there exists a need for new methods that enable rapid and efficient sequence-specific polynucleotide enrichment.

[0004] Nucleic acid amplification is a key step of many nucleic acid based processes such as nucleic acid sequencing. Most currently used nucleic acid amplification methods, *e.g.,* those used in cluster generation in the next generation sequencing, require both temperature cycling and fluid exchanges. Isothermal amplification, on the other hand, can be time and energy efficient by eliminating temperature ramp and equilibration times. Several isothermal amplification methods have been developed, *e.g.,* recombinase polymerase amplification (RPA) based isothermal amplification. These isothermal amplification systems usually lack the desired speed and efficiency ideally suitable for some applications. Furthermore, some systems require additional enzymes and reagents including ATP. Thus, there remains a need in the art for convenient, rapid, and efficient isothermal nucleic acid amplification methods.

[0005] CRISPR-Cas and related proteins are discussed in PCT Publ. Nos. WO2016/014409 and WO2016/077350 which relates to methods for amplifying polynucleotides, and more specifically to methods for amplifying polynucleotides using CRISPR-Cas systems and applications thereof. US2015/0211058 relates to an isothermal amplification of target nucleic acid sequences. The method takes advantages of the CRISPR/CAS system and uses related nuclease (such as CAS9) and guide RNAs to target two separate sequences in the target nucleic acid strand.

[0006] The present disclosure addresses these and other needs by providing methods for enriching or amplifying nucleic acid using CRISPR-Cas or Argonaute systems. Wilson, R. et al. (2013) and Ender C., and Meister, G., (2010) both relate to Argonaute proteins. Related

advantages are provided as well.

SUMMARY

[0007] The invention is defined by the appended claims.

[0008] The present disclosure provides methods for enriching or amplifying polynucleotides, and more specifically to methods for enriching or amplifying a target DNA sequence using endonuclease systems, *e.g.,* CRISPR-Cas systems or Argonaute systems, and applications thereof.

[0009] In one aspect, provided herein are methods for amplifying a target double-stranded nucleic acid using CRISPR-Cas systems.

[0010] In certain embodiments, provided herein is a method for amplifying a target double-stranded nucleic acid including: (a) providing a system having: a clustered regularly interspaced short palindromic repeats (CRISPR) RNA (crRNA) or a derivative thereof, and a CRISPR-associated (Cas) protein or a variant thereof, wherein the crRNA or the derivative thereof contains a target-specific nucleotide region complementary to a region of a first strand of the target double-stranded nucleic acid; (b) contacting the target double-stranded nucleic acid with the system to form a complex; (c) hybridizing a primer to a second strand of the target double-stranded nucleic acid, the primer containing a sequence complementary to a region of the second strand of the target double-stranded nucleic acid, and (d) extending a nucleic acid complementary to the second strand of the target double-stranded nucleic acid from the primer using a polymerase.

[0011] In some embodiments, the method provided herein further including repeating step (c) to step (d) for one or more times, *e.g.,* until a desired degree of amplification is reached. In some embodiments, the method provided herein further including repeating step (a) to step (d) until a desired degree of amplification is reached.

[0012] In some embodiments, the target nucleic acid provided herein is a double-stranded DNA (dsDNA). In some embodiments, the target nucleic acid is a double-stranded RNA (dsRNA).

[0013] In some embodiments, the system is a Type I CRISPR-Cas system or a derivative thereof. In some embodiments, the system is a Type II CRISPR-Cas system or a derivative thereof. In some embodiments, the system is a Type III CRISPR-Cas system or a derivative thereof.

[0014] In some embodiments, the system further comprises a trans-activating crRNA (tracrRNA) or a derivative thereof. In some embodiments, the crRNA or the derivative thereof is a polynucleotide comprising a crRNA polynucleotide fused to a tracrRNA polynucleotide.

[0015] In some embodiments, the first strand of the target double-stranded nucleic acid contains a sequence complementary to a 5'-NGG protospacer-adjacent motif (PAM).

[0016] In some embodiments, the first strand of the target double-stranded nucleic acid contains a universal sequence, and wherein the crRNA or the derivative thereof contains a sequence complementary to a region of the universal sequence. In some embodiments, the primer contains a sequence of a region of the universal sequence.

[0017] In some embodiments, the universal sequence has a sequence of SEQ ID No. 3. In some embodiments, the crRNA contains a sequence of SEQ ID No.7. In some embodiments, the universal sequence has a sequence of SEQ ID No. 5.

[0018] In some embodiments, the universal sequence has a sequence of SEQ ID No. 4. In some embodiments, the crRNA contains a sequence of SEQ ID No.8. In some embodiments, the universal sequence has a sequence of SEQ ID No. 6.

[0019] In some embodiments, the Cas protein or the variant thereof is a Cas9 protein or a variant thereof. In some embodiments, the Cas9 protein contains two inactivated nuclease domains. In some embodiments, the two inactivated nuclease domains comprise a first mutation in the domain that cleaves the strand complementary to the crRNA and a second mutation in the domain that cleaves the strand non-complementary to the crRNA. In some embodiments, the first mutation is D10A and the second mutation is H840A. In some embodiments, the Cas protein or the variant thereof is a Cascade protein or a variant thereof. In some embodiments, the Cas protein or the variant thereof is a Cas3 protein or a variant thereof.

[0020] In some embodiments, the polymerase is a strand-displacing polymerase. In some embodiments, the polymerase is selected from a group consisting of Bst, Bsu, and Phi29.

[0021] In some embodiments, the method provided herein further comprises: applying at least one transposase and at least one transposon end composition containing a transferred strand to a sample containing a target nucleic acid under conditions where the target nucleic acid and the transposon end composition undergo a transposition reaction to generate a mixture, wherein the target nucleic acid is fragmented to generate a plurality of target nucleic acid fragments, and incorporating a universal primer sequence into each of the plurality of target nucleic acid fragments, wherein the crRNA or the derivative thereof contains a target-specific nucleotide region complementary to a region of the universal primer.

[0022] In some embodiments, the universal primer is incorporated into the plurality of target nucleic acid fragments by a PCR reaction. In some embodiments, the universal primer has sequence complementary to a 5'-NGG protospacer-adjacent motif (PAM).

[0023] In some embodiments, the universal sequence has a sequence of SEQ ID No. 3. In

some embodiments, the crRNA contains a sequence of SEQ ID No.7. In some embodiments, the universal sequence has a sequence of SEQ ID No. 5.

[0024] In some embodiments, the universal primer has a sequence of SEQ ID No. 4. In some embodiments, the crRNA contains a sequence of SEQ ID No.8. In some embodiments, the universal sequence has a sequence of SEQ ID No. 6.

[0025] In some embodiments, two universal primers are incorporated into two ends of each of the plurality of target nucleic acid fragments. In some embodiments, the two universal primers have sequences of SEQ ID No.3 and SEQ ID No.4. In some embodiments, the two universal primers have sequences of SEQ ID No.5 and SEQ ID No.6.

[0026] In some embodiments, the target double-stranded nucleic acid is linearly amplified. In some embodiments, the target double-stranded nucleic acid is exponentially amplified.

[0027] In some embodiments, provided herein is a method for amplifying a target double-stranded nucleic acid comprising: (a) providing a first system having: a first clustered regularly interspaced short palindromic repeats (CRISPR) RNA (crRNA) or a derivative thereof, and a first CRISPR-associated (Cas) protein or a variant thereof, wherein the first crRNA or the derivative thereof contains a target-specific nucleotide region complementary to a region of a first strand of the target double-stranded nucleic acid; (b) providing second system having: a second clustered regularly interspaced short palindromic repeats (CRISPR) RNA (crRNA) or a derivative thereof, and a second CRISPR-associated (Cas) protein or a variant thereof, wherein the second crRNA or the derivative thereof contains a target-specific nucleotide region complementary to a region of a second strand of the target double-stranded nucleic acid; (c) contacting the target double-stranded nucleic acid with the first system and the second system; (d) hybridizing a first primer to a second strand of the target double-stranded nucleic acid, the first primer containing a sequence complementary to a region of the second strand of the target double-stranded nucleic acid, and hybridizing a second primer to a first strand of the target double-stranded nucleic acid, the second primer containing a sequence complementary to a region of the first strand of the target double-stranded nucleic acid, and (e) extending the 3' end of the first primer and the second primer with one or more polymerases to generate a first and a second double stranded target nucleic acid. In some embodiments, the method provided herein further includes repeating step (a) and step (e) for one or more times, e.g., until a desired degree of amplification is reached.

[0028] In some embodiments, the target nucleic acid is a double-stranded DNA (dsDNA). In some embodiments, the target nucleic acid is a double-stranded RNA (dsRNA).

[0029] In some embodiments, the first system or the second system is a Type I CRISPR-Cas system or a derivative thereof. In some embodiments, the first system or the second system is a Type II CRISPR-Cas system or a derivative thereof. In some embodiments, the first system or a second system is a Type III CRISPR-Cas system or a derivative thereof.

[0030] In some embodiments, the first system or the second system further comprises a trans-activating crRNA (tracrRNA) or a derivative thereof. In some embodiments, the crRNA or the derivative thereof of the first system or the second system is a polynucleotide comprising a crRNA polynucleotide fused to a tracrRNA polynucleotide. In some embodiments, the first strand and the second strand of the target double-stranded nucleic acid contain a sequence complementary to a 5'-NGG protospacer-adjacent motif (PAM).

[0031] In some embodiments, the first strand of the target double-stranded nucleic acid contains a first universal sequence, and wherein the crRNA or the derivative thereof of the first system contains a sequence complementary to a region of the first universal sequence, and the second strand of the target double-stranded nucleic acid contains a second universal sequence, and wherein the crRNA or the derivative thereof of the second system contains a sequence complementary to a region of the second universal sequence.

[0032] In some embodiments, the first primer contains a sequence of a region of the first universal sequence, and the second primer contains a sequence of a region of the second universal sequence. In some embodiments, the first universal sequence (which contains a first primer) has a sequence of SEQ ID No. 3, the crRNA or the derivative thereof of the first system contains a sequence of SEQ ID No. 7, and the first primer contains a sequence of SEQ ID No. 5, and the second universal sequence (which contains a second primer) has a sequence of SEQ ID No. 4, the crRNA or derivative thereof of the second system contains a sequence of SEQ ID No.8, and the second primer contains a sequence of SEQ ID No. 6.

[0033] In some embodiments, the Cas protein or the variant thereof of the first system or the second system is a Cas9 protein or a variant thereof. In some embodiments, the Cas9 protein contains two inactivated nuclease domains. In some embodiments, the two inactivated nuclease domains comprise a first mutation in the domain that cleaves the strand complementary to the crRNA and a second mutation in the domain that cleaves the strand non-complementary to the crRNA. In some embodiments, the first mutation is D10A and the second mutation is H840A. In some embodiments, the Cas protein or the variant thereof of the first system or the second system is a Cascade protein or a variant thereof. In some embodiments, the Cas protein or the variant thereof of the first system or the second system is a Cas3 protein or a variant thereof.

[0034] In some embodiments, the polymerase is a strand-displacing polymerase. In some embodiments, the polymerase is selected from a group consisting of Bst, Bsu, and Phi29.

[0035] In some embodiments, the target nucleic acid is genomic DNA. In some embodiments, the target nucleic acid contains chromosomal DNA or a fragment thereof. In some embodiments, the target nucleic acid comprises a genome or a partial genome.

[0036] In some embodiments, the method provided herein further includes sequencing the target nucleic acid or target nucleic acid fragments. In some embodiments, the sequencing comprises use of one or more of sequencing by synthesis, bridge PCR, chain termination

sequencing, sequencing by hybridization, nanopore sequencing, and sequencing by ligation.

[0037] In another aspect, provided herein are methods for amplifying a target double-stranded nucleic acid using Argonaute systems.

[0038] In certain embodiments, provided herein is a method for amplifying a target double-stranded nucleic acid including: (a) providing a system having: a 5' phosphorylated single-stranded nucleic acid or a derivative thereof, and an Argonaute protein or a variant thereof, wherein the 5' phosphorylated single-stranded nucleic acid or the derivative thereof contains a target-specific nucleotide region complementary to a region of a first strand of the target double-stranded nucleic acid; (b) contacting the target double-stranded nucleic acid with the system to form a complex; (c) hybridizing a primer to a second strand of the target double-stranded nucleic acid, the primer containing a sequence complementary to a region of the second strand of the target double-stranded nucleic acid, and (d) extending a nucleic acid complementary to the second strand of the target double-stranded nucleic acid from the primer using a polymerase.

[0039] In some embodiments, the method provided herein further including repeating step (c) to step (d) for one or more times, *e.g.,* until a desired degree of amplification is reached. In some embodiments, the method provided herein further including repeating step (a) to step (d) until a desired degree of amplification is reached.

[0040] In some embodiments, the target nucleic acid provided herein is a double-stranded DNA (dsDNA). In some embodiments, the target nucleic acid is a double-stranded RNA (dsRNA).

[0041] In some embodiments, the 5' phosphorylated single-stranded nucleic acid is a single-stranded DNA (guide DNA or gDNA). In some embodiments, the 5' phosphorylated single-stranded nucleic acid is a single-stranded DNA of less than 100 nucleotides. In some embodiments, the 5' phosphorylated single-stranded nucleic acid is a single-stranded DNA of less than 50 nucleotides. In some embodiments, the 5' phosphorylated single-stranded nucleic acid is a single-stranded DNA of less than 25 nucleotides. Example lengths of the single-stranded DNA include 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, and 50 nucleotides.

[0042] In some embodiments, the Argonaute protein is Natronobacterium gregorgi Argonaute (Ng Argonaute). In some embodiments, the Argonaute protein or a variant thereof is engineered to remove the endonuclease activity so that the Argonaute system does not generate double-stranded or single-stranded DNA breaks. In other embodiments, the Argonaute protein or a variant thereof is engineered to be able to generate single-stranded DNA breaks.

[0043] In some embodiments, the first strand of the target double-stranded nucleic acid

contains a universal sequence, and wherein the gDNA or the derivative thereof contains a sequence complementary to a region of the universal sequence. In some embodiments, the primer contains a sequence of a region of the universal sequence.

[0044] In some embodiments, the polymerase is a strand-displacing polymerase. In some embodiments, the polymerase is selected from a group consisting of Bst, Bsu, and Phi29.

[0045] In some embodiments, the method provided herein further comprises: applying at least one transposase and at least one transposon end composition containing a transferred strand to a sample containing a target nucleic acid under conditions where the target nucleic acid and the transposon end composition undergo a transposition reaction to generate a mixture, wherein the target nucleic acid is fragmented to generate a plurality of target nucleic acid fragments.

[0046] In some embodiments, the target double-stranded nucleic acid is linearly amplified. In some embodiments, the target double-stranded nucleic acid is exponentially amplified.

[0047] In some embodiments, provided herein is a method for amplifying a target double-stranded nucleic acid comprising: (a) providing a first system having: a 5' phosphorylated single-stranded nucleic acid or a derivative thereof, and a first Argonaute protein or a variant thereof, wherein the first 5' phosphorylated single-stranded nucleic acid or the derivative thereof contains a target-specific nucleotide region complementary to a region of a first strand of the target double-stranded nucleic acid; (b) providing second system having: a second 5' phosphorylated single-stranded nucleic acid or a derivative thereof, and a second Argonaute protein or a variant thereof, wherein the second 5' phosphorylated single-stranded nucleic acid or the derivative thereof contains a target-specific nucleotide region complementary to a region of a second strand of the target double-stranded nucleic acid; (c) contacting the target double-stranded nucleic acid with the first system and the second system; (d) hybridizing a first primer to a second strand of the target double-stranded nucleic acid, the first primer containing a sequence complementary to a region of the second strand of the target double-stranded nucleic acid, and hybridizing a second primer to a first strand of the target double-stranded nucleic acid, the second primer containing a sequence complementary to a region of the first strand of the target double-stranded nucleic acid, and (e) extending the 3' end of the first primer and the second primer with one or more polymerases to generate a first and a second double stranded target nucleic acid. In some embodiments, the method provided herein further includes repeating step (a) and step (e) for one or more times, *e.g.,* until a desired degree of amplification is reached.

[0048] In some embodiments, the target nucleic acid is a double-stranded DNA (dsDNA). In some embodiments, the target nucleic acid is a double-stranded RNA (dsRNA).

[0049] In some embodiments, the 5' phosphorylated single-stranded nucleic acid is a single-stranded DNA (guide DNA or gDNA). In some embodiments, the 5' phosphorylated single-stranded nucleic acid is a single-stranded DNA of less than 100 nucleotides. In some

embodiments, the 5' phosphorylated single-stranded nucleic acid is a single-stranded DNA of less than 50 nucleotides. In some embodiments, the 5' phosphorylated single-stranded nucleic acid is a single-stranded DNA of less than 25 nucleotides. Example lengths of the single-stranded DNA include 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, and 50 nucleotides.

[0050] In some embodiments, the Argonaute protein is Natronobacterium gregorgi Argonaute (Ng Argonaute). In some embodiments, the Argonaute protein or a variant thereof is engineered to remove the endonuclease activity so that the Argonaute system does not generate double-stranded or single-stranded DNA breaks. In other embodiments, the Argonaute protein or a variant thereof is engineered to be able to generate single-stranded DNA breaks.

[0051] In some embodiments, the first strand of the target double-stranded nucleic acid contains a first universal sequence, and wherein the gDNA or the derivative thereof of the first system contains a sequence complementary to a region of the first universal sequence, and the second strand of the target double-stranded nucleic acid contains a second universal sequence, and wherein the gDNA or the derivative thereof of the second system contains a sequence complementary to a region of the second universal sequence.

[0052] In some embodiments, the first primer contains a sequence of a region of the first universal sequence, and the second primer contains a sequence of a region of the second universal sequence.

[0053] In some embodiments, the polymerase is a strand-displacing polymerase. In some embodiments, the polymerase is selected from a group consisting of Bst, Bsu, and Phi29.

[0054] In some embodiments, the target nucleic acid is genomic DNA. In some embodiments, the target nucleic acid contains chromosomal DNA or a fragment thereof. In some embodiments, the target nucleic acid comprises a genome or a partial genome.

[0055] In some embodiments, the method provided herein further includes sequencing the target nucleic acid or target nucleic acid fragments. In some embodiments, the sequencing comprises use of one or more of sequencing by synthesis, bridge PCR, chain termination sequencing, sequencing by hybridization, nanopore sequencing, and sequencing by ligation.

[0056] In yet another aspect, provided herein are methods for enriching polynucleotides using CRISPR-Cas or Argonaute systems. In certain embodiments, provided herein is a method for enriching a target nucleic acid including providing a system having: a 5' phosphorylated single-stranded nucleic acid or a derivative thereof, and an Argonaute protein or a variant thereof, wherein the 5' phosphorylated single-stranded nucleic acid or the derivative thereof contains a target-specific nucleotide region complementary to a region of the target nucleic acid; contacting the target nucleic acid with the endonuclease system to form a complex, and

separating the complex and thereby enriching for the target nucleic acid.

[0057] In some embodiments, the method further includes separating the target nucleic acid from the complex. In some embodiments, the method further includes amplifying the targeted nucleic acid.

[0058] In some embodiments, the 5' phosphorylated single-stranded nucleic acid is a single-stranded DNA (guide DNA or gDNA). In some embodiments, the 5' phosphorylated single-stranded nucleic acid is a single-stranded DNA of less than 100 nucleotides. In some embodiments, the 5' phosphorylated single-stranded nucleic acid is a single-stranded DNA of less than 50 nucleotides. In some embodiments, the 5' phosphorylated single-stranded nucleic acid is a single-stranded DNA of less than 25 nucleotides. Example lengths of the single-stranded DNA include 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, and 50 nucleotides.

[0059] In some embodiments, the target nucleic acid is a double-stranded DNA (dsDNA).

[0060] In some embodiments, the system is labeled. In some embodiments, the gDNA is labeled with biotin. In some embodiments, the method provided herein further includes adding streptavidin and thereby separating the complex. In some embodiments, the Argonaute protein or the derivative thereof is labeled with a capture tag.

[0061] In certain embodiments, provided herein is a method for enriching a target double-stranded nucleic acid including: providing an endonuclease system having: a 5' phosphorylated single-stranded nucleic acid or a derivative thereof, and an Argonaute protein or a variant thereof or a variant thereof, wherein the 5' phosphorylated single-stranded nucleic acid or the derivative thereof contains a target-specific nucleotide region complementary to a region of a first strand of the target double-stranded nucleic acid; contacting the target double-stranded nucleic acid with the endonuclease system to form a first complex; hybridizing a labeled nucleic acid to a second strand of the target double-stranded nucleic acid to form a second complex, the second strand of the target double-stranded nucleic acid being non-complementary to the 5' phosphorylated single-stranded nucleic acid or the derivative thereof, and separating the second complex and thereby enriching for the target nucleic acid.

[0062] In some embodiments, the method further includes separating the target nucleic acid from the complex. In some embodiments, the method further includes amplifying the targeted nucleic acid.

[0063] In some embodiments, the 5' phosphorylated single-stranded nucleic acids is a single-stranded DNA (guide DNA or gDNA). In some embodiments, the 5' phosphorylated single-stranded nucleic acids is a single-stranded DNA of less than 100 nucleotides. In some embodiments, the 5' phosphorylated single-stranded nucleic acids is a single-stranded DNA of less than 50 nucleotides. In some embodiments, the 5' phosphorylated single-stranded nucleic

acids is a single-stranded DNA of less than 25 nucleotides. Example lengths of the single-stranded DNA include 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, and 50 nucleotides.

[0064] In some embodiments, the target nucleic acid is a double-stranded DNA (dsDNA).

[0065] In some embodiments, the endonuclease system is labeled. In some embodiments, the 5' phosphorylated single-stranded nucleic acid is labeled with biotin. In some embodiments, the method provided herein further comprises adding streptavidin and thereby separating the complex.

[0066] In some embodiments, the Argonaute protein or the derivative thereof is labeled with a capture tag.

[0067] In certain embodiments, provided herein is a method for enriching a target nucleic acid including: obtaining a population of cell free DNA (cfDNA) from a subject's plasma or serum, the population of cell free DNA containing the target nucleic acid; providing a system having: a 5' phosphorylated single-stranded nucleic acid or a derivative thereof, and an Argonaute protein or a variant thereof, wherein the 5' phosphorylated single-stranded nucleic acid or the derivative thereof contains a target-specific nucleotide region complementary to a region of the target nucleic acid; contacting the target nucleic acid with the endonuclease system to form a complex, and separating the complex and thereby enriching for the target nucleic acid.

[0068] In some embodiments, the 5' phosphorylated single-stranded nucleic acids is a single-stranded DNA (guide DNA or gDNA). In some embodiments, the 5' phosphorylated single-stranded nucleic acids is a single-stranded DNA of less than 100 nucleotides. In some embodiments, the 5' phosphorylated single-stranded nucleic acids is a single-stranded DNA of less than 50 nucleotides. In some embodiments, the 5' phosphorylated single-stranded nucleic acids is a single-stranded DNA of less than 25 nucleotides. Example lengths of the single-stranded DNA include 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, and 50 nucleotides.

[0069] In some embodiments, the target nucleic acid contains a single nucleotide variant (SNV). In some embodiments, the method further includes separating the target nucleic acid from the complex. In some embodiments, the method further includes amplifying the targeted nucleic acid. In some embodiments, the target nucleic acid is a double-stranded DNA (dsDNA).

[0070] In some embodiments, the system is labeled. In some embodiments, the gDNA is labeled with biotin. In some embodiments, the method provided herein further includes adding streptavidin and thereby separating the complex. In some embodiments, the Argonaute protein or the derivative thereof is labeled with a capture tag.

[0071] In some embodiments, the target nucleic acid is in a fetal cell faction of the cell free DNA, and wherein the cell free DNA is from maternal plasma. In some embodiments, the subject is a cancer patient.

[0072] In certain embodiments, provided herein is a method for detecting single nucleotide variant (SNV) including: obtaining a population of cell free DNA from a subject's plasma or serum; providing a first system having: a first 5' phosphorylated single-stranded nucleic acid or a derivative thereof, and a first Argonaute protein or a variant thereof, wherein the first 5' phosphorylated single-stranded nucleic acid or the derivative thereof contains a first target-specific nucleotide region complementary to a region of a first target nucleic acid, and wherein the first Argonaute protein has nuclease activity; cleaving the first target nucleic acid using the first endonuclease system, and amplifying a second target nucleic acid using Polymerase Chain Reaction (PCR), wherein the second target nucleic acid contains a single nucleotide variant version of the first target nucleic acid.

[0073] In some embodiments, the 5' phosphorylated single-stranded nucleic acid is a single-stranded DNA (guide DNA or gDNA). In some embodiments, the 5' phosphorylated single-stranded nucleic acid is a single-stranded DNA of less than 100 nucleotides. In some embodiments, the 5' phosphorylated single-stranded nucleic acid is a single-stranded DNA of less than 50 nucleotides. In some embodiments, the 5' phosphorylated single-stranded nucleic acid is a single-stranded DNA of less than 25 nucleotides. Example lengths of the single-stranded DNA include 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, and 50 nucleotides.

[0074] In some embodiments, the method provided herein further includes: providing a second system having: a second 5' phosphorylated single-stranded nucleic acid or a derivative thereof, and a second Argonaute protein or a variant thereof, wherein the second 5' phosphorylated single-stranded nucleic acid or the derivative thereof contains a second target-specific nucleotide region complementary to a region of the second target nucleic acid; contacting the second target nucleic acid with the second endonuclease system to form a complex, and separating the complex and thereby enriching for the second target nucleic acid.

[0075] In some embodiments, the method provided herein further includes separating the second target nucleic acid from the complex. In some embodiments, the second target nucleic acid is a double-stranded DNA (dsDNA).

[0076] In some embodiments, the second system is labeled. In some embodiments, the second 5' phosphorylated single-stranded nucleic acid is labeled with biotin. In some embodiments, the method provided herein further includes adding streptavidin and thereby separating the complex. In some embodiments, the second Argonaute protein or the derivative thereof is labeled with a capture tag.

[0077] In some embodiments, the target nucleic acid is in a fetal cell faction of the cell free

DNA, and wherein the cell free DNA is from maternal plasma. In some embodiments, the subject is a cancer patient.

[0078] In certain embodiments, provided herein is a method for labeling a target nucleic including providing a first system having: a 5' phosphorylated single-stranded nucleic acid or a derivative thereof, and a first Argonaute protein or a variant thereof, wherein the first 5' phosphorylated single-stranded nucleic acid or the derivative thereof contains a first target-specific nucleotide region complementary to a first region of the target nucleic acid, and wherein the first Argonaute protein is capable of generating a single-stranded nick; contacting a double-stranded nucleic acid containing the target nucleic acid with the first nuclease system to generate a first single-stranded nick at the first region of the target nucleic acid, and labeling the target nucleic acid.

[0079] In some embodiments, the method provided herein further includes separating the target nucleic acid through the labeling and thereby enriching the target nucleic acid. In some embodiments, the method provided herein further includes amplifying the target nucleic acid.

[0080] In some embodiments, the target nucleic acid is a double-stranded DNA (dsDNA).

[0081] In some embodiments, the method provided herein further includes performing a nick translation. In some embodiments, the nick translation is performed by using a nick translation polymerase selected from a group consisting of DNA Pol 1, Bst, and Taq. In some embodiments, the nick translation is performed in a reaction mixture containing biotinylated dNTPs. In some embodiments, the biotinylated dNTPs are biotinylated dUTPs. In some embodiments, the method provided herein further includes adding magnetic streptavidin beads to enrich biotinylated target nucleic acid.

[0082] In some embodiments, the method provided herein further includes providing a second system having: a second 5' phosphorylated single-stranded nucleic acid or a derivative thereof, and a second Argonaute protein or a variant thereof, wherein the second 5' phosphorylated single-stranded nucleic acid or the derivative thereof contains a second target-specific nucleotide region complementary to a second region of the target nucleic acid, and wherein the second Argonaute protein is capable of generating a single-stranded nick, and contacting the double-stranded nucleic acid containing the target nucleic acid with the second nuclease system to generate a second single-stranded nick at the second region of the target nucleic acid, wherein the first region of the target nucleic acid is different from the second region of the target nucleic acid.

[0083] In some embodiments, the first single-stranded nick and the second single-stranded nick are on the same strand of the target nucleic acid. In some embodiments, the space between the first single-stranded nick and the second single-stranded nick on the same strand of the target nucleic acid is 1 bp to 20 bp. In some embodiments, the method further includes performing a nick translation. In some embodiments, the nick translation is performed by using a nick translation polymerase Phi29.

[0084] In some embodiments, the first region of the target nucleic acid and the second region of the target nucleic acid are on different strands of the target nucleic acid. In some embodiments, the space between the first single-stranded nick and the second single-stranded nick is from 20 bp to 500 bp.

[0085] In some embodiments, the method provided herein further includes adding a capture probe; and exchanging a single-stranded nucleic acid product between the first single-stranded nick and the second single-stranded nick with the capture probe, wherein the capture probe is able to hybridize to a nucleic acid complementary to the single-stranded nucleic acid product.

[0086] In some embodiments, the sequence of the capture probe is 10% to 100% identical to the sequence of the single-stranded nucleic acid product. In some embodiments, the capture probe is a biotinylated probe. In some embodiments, the method provided herein further includes adding magnetic streptavidin beads to enrich the target nucleic acid. In some embodiments, the capture probe contains an overhang nucleotide sequence, the overhang nucleotide sequence is complementary to an oligonucleotide immobilized on a surface.

[0087] In some embodiments, the first single-stranded nick and the second single-stranded nick are on opposite strands of the target nucleic acid, thereby generating a first double-stranded nucleic acid break end.

[0088] In some embodiments, the method provided herein further includes ligating an adapter to the first double-stranded DNA break end. In some embodiments, the adapter is biotinylated. In some embodiments, the method provided herein further includes adding magnetic streptavidin beads to enrich the target nucleic acid.

[0089] In some embodiments, the method provided herein further includes providing a third system having: a third 5' phosphorylated single-stranded nucleic acid or a derivative thereof, and a third Argonaute protein or a variant thereof, wherein the third 5' phosphorylated single-stranded nucleic acid or the derivative thereof contains a third target-specific nucleotide region substantially complementary to a third region of the target nucleic acid, and wherein the third Argonaute protein is capable of generating a single-stranded nick; providing a fourth system having: a fourth 5' phosphorylated single-stranded nucleic acid or a derivative thereof, and a fourth Argonaute protein or a variant thereof, wherein the fourth 5' phosphorylated single-stranded nucleic acid or the derivative thereof contains a fourth target-specific nucleotide region substantially complementary to a fourth region of the target nucleic acid, and wherein the fourth Argonaute protein is capable of generating a single-stranded nick; and contacting the double-stranded nucleic acid containing the target nucleic acid with the third and fourth systems to generate a third single-stranded nick at the third region of the target nucleic acid and a fourth single-stranded nick at the fourth region of the target nucleic acid, wherein in the third single-stranded nick and the fourth single-stranded nick are on opposite strands of the target nucleic acid, thereby generating a second double-stranded nucleic acid break end, the second double-stranded nucleic acid break end being different from the first double-stranded

nucleic acid break end. In some embodiments, the method further includes ligating an adapter to the second double-stranded nucleic acid break end.

[0090] In certain embodiments, provided herein is a method for enriching a target nucleic acid including: providing a population of Argonaute proteins programmed with a set of 5' phosphorylated single-stranded nucleic acids, wherein the set of 5' phosphorylated single-stranded nucleic acids contains 5' phosphorylated single-stranded nucleic acids complementary to a series of different regions of the target nucleic acid; contacting the target nucleic acid with the population of Argonaute proteins programmed with the set of 5' phosphorylated single-stranded nucleic acids to generate a series of nucleic acid fragments, and ligating adaptors to at least one of nucleic acid fragments, wherein the Argonaute proteins are capable of generating double-stranded DNA breaks.

[0091] In some embodiments, the set of 5' phosphorylated single-stranded nucleic acids contains 5' phosphorylated single-stranded nucleic acids complementary to two different regions of the target nucleic acid. In some embodiments, the target nucleic acid is a double-stranded DNA. In some embodiments, the target nucleic acid is a genomic DNA, a chromosomal DNA, a genome, or a partial genome.

[0092] In another aspect, provided herein is a method for sequencing a target nucleic acid including: providing a population of Argonaute proteins programmed with a set of 5' phosphorylated single-stranded nucleic acids, wherein the set of 5' phosphorylated single-stranded nucleic acids contains 5' phosphorylated single-stranded nucleic acids complementary to a series of different regions across the target nucleic acid; contacting the target nucleic acid with the population of Argonaute proteins programmed with the set of 5' phosphorylated single-stranded nucleic acids to generate a series of nucleic acid fragments, and sequencing the series of nucleic acid fragments.

[0093] In some embodiments, provided herein is a method for sequencing a target nucleic acids including: providing a plurality of populations of Argonaute proteins, each population of Argonaute proteins being programmed with a different set of 5' phosphorylated single-stranded nucleic acids, wherein each set of 5' phosphorylated single-stranded nucleic acids contains 5' phosphorylated single-stranded nucleic acids complementary to a different series of regions across the target nucleic acid, contacting the target nucleic acid with each of the plurality of populations of Argonaute proteins in a separate reaction to generate a different series of nucleic acid fragments, and sequencing the nucleic acid fragments.

[0094] In some embodiments, the plurality of populations of Argonaute proteins comprises three populations of Argonaute proteins, and wherein the nucleic acid fragments generated by each of the three populations of Argonaute proteins contain overlapping sequences with the nucleic acid fragments generated by at least another of the three populations of Argonaute proteins. In some embodiments, the Argonaute proteins are capable of generating double-stranded DNA breaks. In some embodiments, the target nucleic acid is a double-stranded DNA. In some embodiments, the target nucleic acid is a genomic DNA, a chromosomal DNA, a

genome, or a partial genome. In some embodiments, the method further includes ligating an adapter to the nucleic acid fragments. In some embodiments, the method provided herein further includes diluting a DNA sample containing the target DNA to haploid content. In some embodiments, the sequencing the nucleic acid fragments comprises use of one or more of sequencing by synthesis, bridge PCR, chain termination sequencing, sequencing by hybridization, nanopore sequencing, and sequencing by ligation.

BRIEF DESCRIPTION OF THE DRAWINGS

[0095] Figure 1A shows the primers designed according to the present methods for amplifying DNA fragments containing the entirety or portions of Illumina universal sequencing primer adaptors P5 and P7. The primers can be added using the Nextera (Illumina, Inc.) library preparation method. Figure 1B illustrates one round of Cas9-mediated linear amplification using a crRNA targeting a modified P5 primer. 2B(I) depicts the DNA to be amplified which possesses appropriate primer sequences P5 and P7. The guide RNA targeting P5, bound to Cas9, is also shown as P5'. 1B(II) shows the R loop created by Cas9 after the guide RNA has hybridized to the first strand. 1B(III) illustrates the immobilized P5 primer hybridizing to the displaced second strand, followed by polymerase extension. 1B(IV) shows the resulting extended primer P5. The resulting amplicon can be re-targeted by Cas9+crRNA as shown in step 1B(II). Figure 1C illustrates one round of Cas9-mediated linear amplification using a crRNA targeting a modified P7 primer. 1C(I) depicts the DNA to be amplified which possesses appropriate primer sequences P5 and P7. The guide RNA targeting P7, bound to Cas9, is also shown as P7'. 1C(II) shows the R loop created by Cas9 after the guide RNA has hybridized to the first strand. 1C(III) illustrates the immobilized P7 primer hybridizing to the displaced second strand, followed by polymerase extension. 1C(IV) shows the resulting extended primer P7. The resulting amplicon can be re-targeted by Cas9+crRNA as shown in step 1C(II). With both sides of the amplicon undergoing amplification (Figures. 1B and 1C), exponential amplification can be achieved.

DETAILED DESCRIPTION

[0096] The present disclosure provides methods for rapid and efficient enrichment or amplification of target nucleic acid using CRISPR-Cas or Argonaute systems.

[0097] Current target-specific enrichment protocols require that single-stranded nucleic acid be made prior to the target specific hybridization with probes. Among various advantages provided by the present disclosure, the present disclosure provides enrichment methods that can skip this step of generating single-stranded nucleic acid in the first place, and enable direct targeting to double-stranded nucleic acid, *e.g.,* double-stranded DNA (dsDNA). Methods targeting directly to double-stranded DNA (either partly or completely double-stranded) have unique advantages over single-stranded enrichment strategies. For example, non-specific

hybridization of single-stranded genomic DNA to targeted regions reduces specificity and often requires extensive stringency washing or other time-consuming steps; and single-stranded enrichment schemes often utilizes Cot-1 or other blocking DNA to reduce non-specific hybridization. These additives are not required from double-stranded DNA enrichment schemes, reducing both cost and number of required reagents. In addition, it is easier to make sequencing libraries from double-stranded DNA than from single-stranded DNA. As such, enrichment of double-stranded DNA allows library preparation (*e.g.*, tagmentation) to occur after enrichment. For another example, since specificity (tree-like structures and non-specific hybridization is less of an issue with double-stranded DNA enrichment, potentially larger DNA fragments can be better specifically enriched compared to single-stranded DNA enrichment schemes. This is a particularly important advantage if one considers targeted sequencing in the context of haplotyping and assembly. Also, since longer DNA fragments can potentially be enriched, we have greater flexibility to where the target probe is designed. For example, we can avoid high polymorphic regions but still capture these regions. Also, fewer probes need to be used to capture large regions, reducing both capture probe cost and design.

[0098] In addition, the current protocols of target specific hybridization have slow kinetics and usually require high temperature. The present disclosure provides enzyme-driven sequence targeting methods that offer faster kinetics and easier workflow for enrichment. Because the hybridization to the target nucleic acid is enzyme driven in the present methods, the process can take place isothermally. In some embodiments, the method herein provides isothermal targeting of DNA at 20-37°C. Furthermore, the guide RNA, *e.g.,* crRNA, in the system herein provides for sequence specificity as well as flexible programming that enables multiplex targeted enrichment (*e.g.*, targeting multiple targeted regions with more probes made in various ways including IVT from oligo pool).

[0099] Nucleic acid amplification is a step of many nucleic acid based processes such as next generation sequencing. Currently the polymerase chain reaction (PCR) is the most widely used method for DNA amplification for, *e.g.,* detection and identification of infectious diseases, genetic disorders and other research purposes. A PCR reaction typically uses two oligonucleotide primers, which are hybridized to the 3' ends of a duplex target nucleic acid sequence, and a DNA polymerase, which can extend the annealed primers by adding on deoxyribonucleoside triphosphates (dNTPs) to generate double-stranded nucleic acid products. Gill and Ghaemi, Nucleosides, Nucleotides, and Nucleic Acids, 2008, 27: 224-243. However, a PCR reaction requires thermocycling to separate the two DNA strands. Similarly, many currently used nucleic acid amplification methods, *e.g.*, those used in cluster generation in the next generation sequencing, require both temperature cycling and fluid exchanges.

[0100] Several isothermal amplification methods have been developed in order to eliminate temperature ramp and equilibration times, such as recombinase polymerase amplification (RPA) based isothermal amplification, transcription mediated amplification, nucleic acid sequence-based amplification, signal mediated amplification of RNA, strand displacement amplification, rolling circle amplification, loop-mediated amplification of DNA, isothermal multiple displacement amplification, helicase-dependent amplification, single primer isothermal

amplification, and circular helicase-dependent amplification, as described in Gill and Ghaemi, Nucleosides, Nucleotides, and Nucleic Acids, 2008, 27: 224-243.

[0101] For example, in a transcription mediated amplification (TMA), an RNA polymerase is used to make RNA from a promoter engineered in the primer region, and then a reverse transcriptase synthesizes cDNA from the primer. A third enzyme, *e.g.,* Rnase H can then be used to degrade the RNA target from cDNA without the heat-denatured step. This amplification technique is very similar to Self-Sustained Sequence Replication (3SR) and Nucleic Acid Sequence Based Amplification (NASBA), but varies in the enzymes employed. *Id.* For another example, helicase-dependent amplification (HDA) utilizes a thermostable helicase (Tte-UvrD) rather than heat to unwind dsDNA to create single-strands that are then available for hybridization and extension of primers by polymerase. Reaction times have been shown to be over 1 hour to amplify products 70-120 base pairs in length. *Id.* For yet another example, a loop mediated amplification (LAMP) employs a thermostable polymerase with strand displacement capabilities and a set of four or more specific designed primers. Each primer is designed to have hairpin ends that, once displaced, snap into a hairpin to facilitate self-priming and further polymerase extension. In a LAMP reaction, though the reaction proceeds under isothermal conditions, an initial heat denaturation step is required for double-stranded targets. In addition, amplification yields a ladder pattern of various length products. *Id.* For yet another example, a strand displacement amplification (SDA) combines the ability of a restriction endonuclease to nick the unmodified strand of its target DNA and an exonulease-deficient DNA polymerase to extend the 3' end at the nick and displace the downstream DNA strand. *Id.* Other exemplary isothermal amplification methods include, but not limited to, those described in Craw and Balachandran, Lab Chip, 2012, 12: 2469-2486.

[0102] However, these currently developed isothermal amplification systems usually lack the desired speed and efficiency ideally suitable for some applications. Furthermore, some systems require additional enzymes and reagents including ATP. Thus, there remains a need in the art for convenient, rapid, and efficient isothermal nucleic acid amplification methods.

[0103] The present disclosure addresses this and other needs by providing methods for enriching or amplifying nucleic acid using CRISPR-Cas or Argonautes systems. For example, one advantage provided by the present methods is that the system used herein, *e.g.,* Cas protein, recognizes target nucleic acid without consuming ATP or energy investment, and thus the methods herein provide for cost and time efficient isothermal amplification methods.

[0104] The present disclosure also provides methods for enriching and/or detecting polynucleotide variants with higher sensitivity and specificity. Furthermore, the present invention is defined by the appended claims.

DEFINITIONS

[0105] As used herein, the terms "includes," "including," "includes," "including," "contains,"

"containing," "have," "having," and any variations thereof, are intended to cover a non-exclusive inclusion, such that a process, method, product-by-process, or composition of matter that includes, includes, or contains an element or list of elements does not include only those elements but can include other elements not expressly listed or inherent to such process, method, product-by-process, or composition of matter.

[0106] As used herein, the singular forms "a", "an" and "the" include plural referents unless the content clearly dictates otherwise. Thus, for example, reference to "a protein" includes a mixture of two or more proteins, and the like.

[0107] As used herein, the term "about" or "approximately" means within 5% of a given value or range.

[0108] As used herein, the term "nucleic acid" means single-stranded and double-stranded polymers of nucleotide monomers, including 2'-deoxyribonucleotides (DNA) and ribonucleotides (RNA) linked by internucleotide phosphodiester bond linkages, or internucleotide analogs, and associated counter ions, *e.g.*, H^+, NH^{4+}, trialkylammonium, tetraalkylammonium, Mg^{2+}, Na^+ and the like. A nucleic acid can be a polynucleotide or an oligonucleotide. A nucleic acid may be composed entirely of deoxyribonucleotides, entirely of ribonucleotides, or chimeric mixtures thereof. The nucleotide monomer units may comprise any of the nucleotides described herein, including, but not limited to, naturally occurring nucleotides and nucleotides analogs. Nucleic acid typically ranges in size from a few monomeric units, *e.g.,* 5-40, to several thousands of monomeric nucleotide units. Nucleic acids include, but are not limited to, genomic DNA, cDNA, hnRNA, mRNA, rRNA, tRNA, fragmented nucleic acid, nucleic acid obtained from sub-cellular organelles such as mitochondria or chloroplasts, and nucleic acid obtained from microorganisms or DNA or RNA viruses that may be present on or in a biological sample.

[0109] As used herein, the term "target nucleic acid" is intended to mean a nucleic acid that is the object of an analysis or action. The analysis or action includes subjecting the nucleic acid to copying, amplification, sequencing and/or other procedure for nucleic acid interrogation. A target nucleic acid can include nucleotide sequences additional to the target sequence to be analyzed. For example, a target nucleic acid can include one or more adapters, including an adapter that functions as a primer binding site, that flank(s) a target nucleic acid sequence that is to be analyzed. A target nucleic acid hybridized to a capture oligonucleotide or capture primer can contain nucleotides that extend beyond the 5' or 3' end of the capture oligonucleotide in such a way that not all of the target nucleic acid is amenable to extension.

[0110] As used herein, the term "target specific" when used in reference to a guide RNA or DNA, a crRNA or a derivative thereof, a 5' phosphorylated single-stranded nucleic acid, or other nucleotide is intended to mean a polynucleotide that includes a nucleotide sequence specific to a target polynucleotide sequence, namely a sequence of nucleotides capable of selectively annealing to an identifying region of a target polynucleotide, *e.g.,* a target DNA. Target specific nucleotide can have a single species of oligonucleotide, or it can include two or

more species with different sequences. Thus, the target specific nucleotide can be two or more sequences, including 3, 4, 5, 6, 7, 8, 9 or 10 or more different sequences. In one embodiment, a crRNA or the derivative thereof contains a target-specific nucleotide region complementary to a region of the target DNA sequence. In one embodiment, a crRNA or the derivative thereof may contain other nucleotide sequences besides a target-specific nucleotide region. In one embodiment, the other nucleotide sequences may be from a tracrRNA sequence. In one embodiment, a 5' phosphorylated single-stranded nucleic acid or the derivative thereof contains a target-specific nucleotide region complementary to a region of the target DNA sequence. In one embodiment, a 5' phosphorylated single-stranded nucleic acid or the derivative thereof may contain other nucleotide sequences besides a target-specific nucleotide region.

[0111] As used herein, the term "complementary" when used in reference to a polynucleotide is intended to mean a polynucleotide that includes a nucleotide sequence capable of selectively annealing to an identifying region of a target polynucleotide under certain conditions. As used herein, the term "substantially complementary" and grammatical equivalents is intended to mean a polynucleotide that includes a nucleotide sequence capable of specifically annealing to an identifying region of a target polynucleotide under certain conditions. Annealing refers to the nucleotide base-pairing interaction of one nucleic acid with another nucleic acid that results in the formation of a duplex, triplex, or other higher-ordered structure. The primary interaction is typically nucleotide base specific, *e.g.,* A:T, A:U, and G:C, by Watson-Crick and Hoogsteen-type hydrogen bonding. In certain embodiments, base-stacking and hydrophobic interactions can also contribute to duplex stability. Conditions under which a polynucleotide anneals to complementary or substantially complementary regions of target nucleic acids are well known in the art, *e.g.,* as described in Nucleic Acid Hybridization, A Practical Approach, Hames and Higgins, eds., IRL Press, Washington, D.C. (1985) and Wetmur and Davidson, Mol. Biol. 31:349 (1968). Annealing conditions will depend upon the particular application, and can be routinely determined by persons skilled in the art, without undue experimentation.

[0112] As used herein, the term "hybridization" refers to the process in which two single-stranded polynucleotides bind non-covalently to form a stable double-stranded polynucleotide. A resulting double-stranded polynucleotide is a "hybrid" or "duplex." Hybridization conditions will typically include salt concentrations of less than about 1 M, more usually less than about 500 mM and may be less than about 200 mM. A hybridization buffer includes a buffered salt solution such as 5% SSPE, or other such buffers known in the art. Hybridization temperatures can be as low as 5° C, but are typically greater than 22° C, and more typically greater than about 30° C, and typically in excess of 37° C. Hybridizations are usually performed under stringent conditions, *i.e.,* conditions under which a probe will hybridize to its target subsequence but will not hybridize to the other, uncomplimentary sequences. Stringent conditions are sequence-dependent and are different in different circumstances, and may be determined routinely by those skilled in the art.

[0113] In the context of "polynucleotides," the terms "variant" and "derivative" as used herein

refer to a polynucleotide that comprises a nucleotide sequence of a polynucleotide or a fragment of a polynucleotide, which has been altered by the introduction of nucleotide substitutions, deletions or additions. A variant or a derivative of a polynucleotide can be a fusion polynucleotide which contains part of the nucleotide sequence of a polynucleotide. The term "variant" or "derivative" as used herein also refers to a polynucleotide or a fragment thereof, which has been chemically modified, *e.g.,* by the covalent attachment of any type of molecule to the polynucleotide. For example, but not by way of limitation, a polynucleotide or a fragment thereof can be chemically modified, *e.g.,* by acetylation, phosphorylation, methylation, etc. The variants or derivatives are modified in a manner that is different from naturally occurring or starting nucleotide or polynucleotide, either in the type or location of the molecules attached. Variants or derivatives further include deletion of one or more chemical groups which are naturally present on the nucleotide or polynucleotide. A variant or a derivative of a polynucleotide or a fragment of a polynucleotide can be chemically modified by chemical modifications using techniques known to those of skill in the art, including, but not limited to specific chemical cleavage, acetylation, formulation, etc. Further, a variant or a derivative of a polynucleotide or a fragment of a polynucleotide can contain one or more dNTPs or nucleotide analogs. A polynucleotide variant or derivative may possess a similar or identical function as a polynucleotide or a fragment of a polynucleotide described herein. A polynucleotide variant or derivative may possess an additional or different function compared with a polynucleotide or a fragment of a polynucleotide described herein.

[0114] As used herein, the term "dNTP" refers to deoxynucleoside triphosphates. NTP refers to ribonucleotide triphosphates such as those used to synthesize crRNA or tracrRNA. The purine bases (Pu) include adenine (A), guanine (G) and derivatives and analogs thereof. The pyrimidine bases (Py) include cytosine (C), thymine (T), uracil (U) and derivatives and analogs thereof. Examples of such derivatives or analogs, by way of illustration and not limitation, are those which are modified with a reporter group, biotinylated, amine modified, radiolabeled, alkylated, and the like and also include phosphorothioate, phosphite, ring atom modified derivatives, and the like. The reporter group can be a fluorescent group such as fluorescein, a chemiluminescent group such as luminol, a terbium chelator such as N-(hydroxyethyl) ethylenediaminetriacetic acid that is capable of detection by delayed fluorescence, and the like.

[0115] As used herein, the term "nucleotide analogs" refers to synthetic analogs having modified nucleotide base portions, modified pentose portions, and/or modified phosphate portions, and, in the case of polynucleotides, modified internucleotide linkages, as generally described elsewhere (*e.g.,* Scheit, Nucleotide Analogs, John Wiley, New York, 1980; Englisch, Angew. Chem. Int. Ed. Engl. 30:613-29, 1991; Agarwal, Protocols for Polynucleotides and Analogs, Humana Press, 1994; and S. Verma and F. Eckstein, Ann. Rev. Biochem. 67:99-134, 1998). Exemplary phosphate analogs include but are not limited to phosphorothioate, phosphorodithioate, phosphoroselenoate, phosphorodiselenoate, phosphoroanilothioate, phosphoranilidate, phosphoramidate, boronophosphates, including associated counterions, *e.g.,* H^+, NH_4^+, Na^+, if such counterions are present. Exemplary modified nucleotide base portions include but are not limited to 5-methylcytosine (5mC); C-5-propynyl analogs, including but not limited to, C-5 propynyl-C and C-5 propynyl-U; 2,6-diaminopurine, also known as 2-

amino adenine or 2-amino-dA); hypoxanthine, pseudouridine, 2-thiopyrimidine, isocytosine (isoC), 5-methyl isoC, and isoguanine (isoG; *see, e.g.,* U.S. Pat. No. 5,432,272). Exemplary modified pentose portions include but are not limited to, locked nucleic acid (LNA) analogs including without limitation Bz-A-LNA, 5-Me-Bz-C-LNA, dmf-G-LNA, and T-LNA (*see, e.g.,* The Glen Report, 16(2):5, 2003; Koshkin et al., Tetrahedron 54:3607-30, 1998), and 2'-or 3'-modifications where the 2'-or 3'-position is hydrogen, hydroxy, alkoxy (*e.g.*, methoxy, ethoxy, allyloxy, isopropoxy, butoxy, isobutoxy and phenoxy), azido, amino, alkylamino, fluoro, chloro, or bromo. Modified intemucleotide linkages include phosphate analogs, analogs having achiral and uncharged intersubunit linkages (*e.g.,* Sterchak, E. P. et al., Organic Chern., 52:4202, 1987), and uncharged morpholino-based polymers having achiral intersubunit linkages (*see, e.g.,* U.S. Pat. No. 5,034,506). Some intemucleotide linkage analogs include morpholidate, acetal, and polyamide-linked heterocycles.

[0116] The terms "polymerase chain reaction," or "PCR," as used herein, refers to a procedure wherein small amounts of a nucleic acid, *e.g.,* RNA and/or DNA, are amplified as described, for example, in U.S. Pat. No. 4,683,195 to Mullis. Generally, sequence information from the ends of the region of interest or beyond needs to be available, such that oligonucleotide primers can be designed; these primers will be identical or similar in sequence to opposite strands of the template to be amplified. The 5' terminal nucleotides of the two primers may coincide with the ends of the amplified material. PCR can be used to amplify specific RNA sequences, specific DNA sequences from total genomic DNA, and cDNA transcribed from total cellular RNA, bacteriophage or plasmid sequences, etc. *See generally* Mullis et al., Cold Spring Harbor Symp. Quant. Biol., 51: 263 (1987); Erlich, ed., PCR Technology, (Stockton Press, NY, 1989).

[0117] As used herein, the terms "ligation," "ligating," and grammatical equivalents thereof are intended to mean to form a covalent bond or linkage between the termini of two or more nucleic acids, *e.g.,* oligonucleotides and/or polynucleotides, typically in a template-driven reaction. The nature of the bond or linkage may vary widely and the ligation may be carried out enzymatically or chemically. As used herein, ligations are usually carried out enzymatically to form a phosphodiester linkage between a 5' carbon terminal nucleotide of one oligonucleotide with a 3' carbon of another nucleotide. Template driven ligation reactions are described in the following references: U.S. Patent Nos. 4,883,750; 5,476,930; 5,593,826; and 5,871,921. The term "ligation" also encompasses non-enzymatic formation of phosphodiester bonds, as well as the formation of non-phosphodiester covalent bonds between the ends of oligonucleotides, such as phosphorothioate bonds, disulfide bonds, and the like.

[0118] As used herein, the term "adapter" is a single-stranded or a double-stranded nucleic acid molecule that can be linked to the end of other nucleic acids. In one embodiment, an adapter is a short, chemically synthesized, double-stranded nucleic acid molecule which can be used to link the ends of two other nucleic acid molecules. In one embodiment, an adaptor is a double-stranded nucleic acid (*e.g.*, oligonucleotides) that comprises single-stranded nucleotide overhangs at the 5' and/or 3' ends. In some embodiments, the single-stranded overhangs are 1, 2 or more nucleotides. In some embodiments, adaptors comprise additional nucleic acid sequence for cloning or analysis of "inserts." In some embodiments, adaptors

comprise labels or affinity tags for analysis or purification of "inserts." The term "insert" refers to a nucleic acid sequence of interest. In some embodiments, inserts are double-stranded DNAs that comprise single stranded nucleotide overhangs at the 5 ' and/or 3' ends. In some embodiments, the single stranded overhangs are 1, 2 or more nucleotides.

[0119] As used herein, the term "CRISPR-Cas system" refers to an enzyme system including a guide RNA sequence that contains a nucleotide sequence complementary or substantially complementary to a region of a target polynucleotide, and a protein with nuclease activity. CRISPR-Cas systems include Type I CRISPR-Cas system, Type II CRISPR-Cas system, Type III CRISPR-Cas system, and derivatives thereof. CRISPR-Cas systems include engineered and/or programmed nuclease systems derived from naturally accruing CRISPR-Cas systems. CRISPR-Cas systems may contain engineered and/or mutated Cas proteins. CRISPR-Cas systems may contain engineered and/or programmed guide RNA.

[0120] As used herein, the term "Argonaute system" refers to an enzyme system including a guide nucleic acid, *e.g.,* a guide DNA, that contains a nucleotide sequence complementary or substantially complementary to a region of a target polynucleotide, and an Argonaute family protein or a variant thereof. Argonaute systems include engineered and/or programmed nuclease systems derived from naturally accruing Argonaute systems. Argonaute systems may contain engineered and/or mutated Argonaute proteins. Argonaute systems may contain engineered and/or programmed guide DNA.

[0121] As used herein, the term "guide RNA" refers to a RNA containing a sequence that is complementary or substantially complementary to a region of a target DNA sequence. A guide RNA may contain nucleotide sequences other than the region complementary or substantially complementary to a region of a target DNA sequence. A guide RNA may be a crRNA or a derivative thereof, *e.g.,* a crRNA: tracrRNA chimera.

[0122] As used herein, the term "guide DNA" refers to a DNA containing a sequence that is complementary or substantially complementary to a region of a target DNA sequence. A guide DNA may contain nucleotide sequences other than the region complementary or substantially complementary to a region of a target DNA sequence.

[0123] As used herein, the term "nuclease" refers to an enzyme capable of cleaving the phosphodiester bonds between the nucleotide subunits of nucleic acids; the term "endonuclease" refers to an enzyme capable of cleaving the phosphodiester bond within a polynucleotide chain; and the term "nickase" refers to an endonuclease which cleaves only a single strand of a DNA duplex. The term "Cas9 nickase" refers to a nickase derived from a Cas9 protein, typically by inactivating one nuclease domain of Cas9 protein.

[0124] In the context of a polypeptide, the terms "variant" and "derivative" as used herein refer to a polypeptide that comprises an amino acid sequence of a polypeptide or a fragment of a polypeptide, which has been altered by the introduction of amino acid residue substitutions, deletions or additions. A variant or a derivative of a polypeptide can be a fusion protein which

contains part of the amino acid sequence of a polypeptide. The term "variant" or "derivative" as used herein also refers to a polypeptide or a fragment of a polypeptide, which has been chemically modified, *e.g.,* by the covalent attachment of any type of molecule to the polypeptide. For example, but not by way of limitation, a polypeptide or a fragment of a polypeptide can be chemically modified, *e.g.,* by glycosylation, acetylation, pegylation, phosphorylation, amidation, derivatization by known protecting/blocking groups, proteolytic cleavage, linkage to a cellular ligand or other protein, etc. The variants or derivatives are modified in a manner that is different from naturally occurring or starting peptide or polypeptides, either in the type or location of the molecules attached. Variants or derivatives further include deletion of one or more chemical groups which are naturally present on the peptide or polypeptide. A variant or a derivative of a polypeptide or a fragment of a polypeptide can be chemically modified by chemical modifications using techniques known to those of skill in the art, including, but not limited to specific chemical cleavage, acetylation, formulation, metabolic synthesis of tunicamycin, etc. Further, a variant or a derivative of a polypeptide or a fragment of a polypeptide can contain one or more non-classical amino acids. A polypeptide variant or derivative may possess a similar or identical function as a polypeptide or a fragment of a polypeptide described herein. A polypeptide variant or derivative may possess an additional or different function compared with a polypeptide or a fragment of a polypeptide described herein.

[0125] As used herein, the term "detecting" a nucleic acid molecule or fragment thereof refers to determining the presence of the nucleic acid molecule, typically when the nucleic acid molecule or fragment thereof has been fully or partially separated from other components of a sample or composition, and also can include determining the charge-to-mass ratio, the mass, the amount, the absorbance, the fluorescence, or other property of the nucleic acid molecule or fragment thereof.

[0126] As used herein, the term "primer" refers to an oligonucleotide primer, whether natural or synthetic, which is capable of acting as a point of initiation of nucleic acid synthesis when placed under conditions in which primer extension (not limited in number of extended bases) is initiated. A primer can be a single-stranded oligodeoxyribonucleotide. The length of a primer can range from about 10 to about 50 nucleotides, for example 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, or about 50 nucleotides. A primer need not reflect the exact sequence of the template but must be sufficiently complementary to hybridize with a template for primer elongation to occur. A primer can be labeled, if desired, by incorporating a label that is detectable by, for example, spectroscopic, photochemical, biochemical, immunochemical, or chemical means. Exemplary labels include, but are not limited to biotin, amine, radiolabels (*e.g.*, 32P), fluorescent dyes, electron-dense reagents, enzymes (as commonly used in ELISAS), or biotin.

[0127] As used herein, the term "linear amplification" refers to an amplification process that uses multiple cycles of primer extension reactions to amplify a target nucleic acid. With linear amplification, the abundance of a transcript increases proportionately with the number of

cycles and scales linearly. An example of a linear amplification procedure is LCR, the aRNA method of Phillips and Eberwine, supra, and the linear amplification method described herein. Unlike exponential amplification, the amount of the amplification products does not grow exponentially. For example, in an ideal 4 hour linear amplification reaction whose copying rate is 2000 copies per minute, 2000 copies of template DNA will yield 960,000,000 copies.

[0128] As used herein, the term "exponential amplification" refers to an amplification procedure where the product (i.e., amplicon) doubles with every reaction cycle. "Exponential amplification" is a non-linear amplification that results in exponential growth in the number of nucleic acid copies present. For example, exponential amplification can occur when primer extension initiates from both ends of an amplicon in one amplificaiotn cycle. For example, PCR is an exponential amplification procedure. For example, in an ideal PCR reaction with 30 cycles, 2 copies of template DNA will yield 2^30 or 1,073,741,824 copies.

[0129] As used herein, in the context of enriching a target polynucleotide, the term "enrich," "enriching", or "enrichment" refers to a process which results in a higher percentage of the target polynucleotide in a polynucleotide population. In one embodiment, the percentage increases about 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% or 100%. In one embodiment, the percentage increases about 2 fold, 5 fold, 10 fold, 50 fold, or 100 fold. In one embodiment, the target polynucleotide is substantially isolated from the polynucleotide population.

[0130] As used herein, the term "polymerase" refers to a protein that is able to catalyze the specific incorporation of nucleotides to extend a 3' hydroxyl terminus of a primer molecule, such as, for example, the template oligonucleotide, against a nucleic acid target sequence. The polymerase can be, for example, thermophilic so that it is active at an elevated reaction temperature. It can also, for example, have strand displacement capabilities.

[0131] As used herein, the term "single nucleotide polymorphism (SNP)" refers to a DNA sequence variation occurring when a single nucleotide -A, T, C, or G- in the genome (or other shared sequence) differs between members of a species (or between paired chromosomes in an individual).

[0132] As used herein, the term "single nucleotide variant (SNV)" refers to one kind of genotype or polynucleotide including a single nucleotide polymorphism (SNP) or point mutation site.

[0133] As used herein, the terms "subject" and "patient" are used interchangeably. As used herein, a subject is preferably a mammal such as a non-primate (*e.g.*, cows, pigs, horses, cats, dogs, rats, etc.) or a primate (*e.g.*, monkey and human). In specific embodiments, the subject is a human. In one embodiment, the subject is a mammal (*e.g.*, a human) having a cancer.

[0134] As used herein, the terms "haplotype," "haploid genotype," and other grammatical equivalents herein refer to a set of nucleotide sequence polymorphisms or alleles present on a

single maternal or paternal chromosome, usually inherited as a unit.

[0135] As used herein, the terms "phased sequencing," "haplotype sequencing," and other grammatical equivalents when used in context of a genome or a chromosome refer to determining the nucleic acid sequence of a single genome or single chromosome, respectively, where the nucleic acid sequence is obtained from the sequencing of a single genome or a single chromosome. The terms "phased sequencing," "haplotype sequencing," and other grammatical equivalents when used in context of a chromosomal fragment refer to determining the nucleic acid sequence of a single chromosomal fragment where the nucleic acid sequence is obtained from the sequencing of a single chromosomal fragment.

METHODS FOR AMPLIFYING POLYNUCLEOTIDES

[0136] Provided herein are methods for nuclease system mediated amplification, such as CRISPR-Cas mediated amplification and Argonaute mediated amplification.

[0137] In one aspect, the present disclosure provides a method for CRISPR-Cas system mediated amplification. In another aspect, the present disclosure provides a method for Argonaute medicated amplification. The methods provided herein is partially based on that binding of a guide nucleic acid such as a guide RNA or a guide DNA to a region of a target double-stranded nucleic acid disrupts the interaction between the two strands of the target nucleic acid, and thereby creates a loop structure (also called "R-loop") exposing the strand non-complementary to the guide nucleic acid. This exposed strand can be subjected to hybridization with primer for extension by an appropriate polymerase, *e.g.,* in a nucleic acid amplification process. As illustrated in Example 1, this loop structure created by CRISPR-Cas systems, *e.g.,* systems containing Cas9 or Cascade proteins, can be accessible to other enzymes. Thus, the loop structure can be further utilized as a template to initiate primer hybridization and interact with polymerase enzymes for amplification.

[0138] In some embodiments, the present disclosure provides a method for amplifying a target double-stranded nucleic acid including providing a system having a clustered regularly interspaced short palindromic repeats (CRISPR) RNA (crRNA) or a derivative thereof, and a CRISPR-associated (Cas) protein or a variant thereof, wherein the crRNA or the derivative thereof contains a target-specific nucleotide region complementary to a region of a first strand of the target double-stranded nucleic acid; contacting the target double-stranded nucleic acid with the system to form a complex; hybridizing a primer to a second strand of the target double-stranded nucleic acid, the primer containing a sequence complementary to a region of the second strand of the target double-nucleic acid, and extending a nucleic acid complementary to the second strand of the target double-stranded nucleic acid from the primer using a polymerase.

[0139] In other embodiments, the present disclosure provides a method for amplifying a target double-stranded nucleic acid including providing a system having: a 5' phosphorylated single-

stranded nucleic acid or a derivative thereof, and an Argonaute protein or a variant thereof, wherein the 5' phosphorylated single-stranded nucleic acid or the derivative thereof contains a target-specific nucleotide region complementary to a region of a first strand of the target double-stranded nucleic acid; contacting the target double-stranded nucleic acid with the system to form a complex; hybridizing a primer to a second strand of the target double-stranded nucleic acid, the primer containing a sequence complementary to a region of the second strand of the target double-stranded nucleic acid, and extending a nucleic acid complementary to the second strand of the target double-stranded nucleic acid from the primer using a polymerase.

[0140] The methods provided herein can be used in various amplification methods, including but not limited to, linear nucleic acid amplification and exponential nucleic acid amplification.

[0141] In some embodiments, the target nucleic acid is amplified linearly according to the methods provided herein. For example, in some embodiments, the method provided herein further includes repeating hybridizing a primer to the second strand of the target double-stranded nucleic acid, and extending a nucleic acid complementary to the second strand of the target double-stranded nucleic acid from the primer using a polymerase for one or more times, e.g., until a desired amount of amplification is achieved.

[0142] In other embodiments, the target nucleic acid is amplified exponentially. In exemplary exponential nucleic acid amplification, the product (i.e., amplicon) doubles with every reaction cycle. "Exponential amplification" is a non-linear amplification that results in exponential growth in the number of nucleic acid copies present. Typically, in an exponential amplification, primer extension (or copying) occurs from both ends of an amplicon. To ensure that the newly created strand has a primer binding sites at both ends, in an exponential amplification, the 3' end of the new synthesized nucleic acid contains the reverse complement of a primer, and the ends of the template are usually copied in each amplification cycle.

[0143] The cycle number of the reactions according to the present methods depends on the applications and the desired amount of the amplification products. In some embodiments, cycle number is 5 to 100. In some embodiments, the cycle number is 10 to 90. In some embodiments, the cycle number is 20 to 80. In some embodiments, the cycle number is 30 to 70. In some embodiments, the cycle number is 40 to 60. Exemplary cycle number includes 10, 15, 20, 25, 30, 35, 40, 45, and 50. Cycles need not be synchronized between amplicons, as they are in a PCR reaction where the start of each cycle is controlled by changing the temperature. Cycles, as used herein, thus refers to the average number of rounds of amplification an amplicon undergoes.

[0144] In some embodiments, the initialization step includes provides multiple CRISPR-Cas systems or multiple Argonaute systems to the target nucleic acid to open the double-stranded nucleic acid structure at two or more targeted sequences and to form two or more R-Loop structures. The initialization step does not require heating the reaction, as required in a PCR reaction, because this initialization step is enzyme driven. The next step according to the

present methods involves providing and annealing primers targeted to the R-Loop regions to form relatively stable nucleic acid-nucleic acid interaction, e.g., DNA-DNA hybrids. Typically, stable nucleic acid-nucleic acid hybrids are formed when the primer sequence has substantial complementarity the template sequence. Then one or more polymerases bind to the primer-template hybrid and begin nucleic acid synthesis in an extension/elongation step. In some embodiments, the temperature at this extension/elgation step depends on the polymerase used. At this step the polymerase synthesizes a new nucleic acid strand complementary to the template strand by adding dNTPs that are complementary to the template. In some embodiments, when DNA polymerase is used, the reaction condenses the 5'-phosphate group of the dNTPs with the 3'-hydroxyl group at the end of the nascent (extending) DNA strand. The extension time depends both on the polymerase used and on the length of the nucleic acid fragment to be amplified. One or more of these steps can be repeated for one or more times, e.g., until a desired amount of amplification is achieved.

[0145] To permit exponential growth of the amplified products, it is benefical that the newly created nucleic acid product in each cycle contains primer binding sites at both ends. In some embodiments, the 3' end of the newly synthesized molecule is the reverse complement of a primer. In some embodiments, two primers, each targeting to one end of a target nucleic acid, can be used in an exponential amplification. In some embodiments, it is also beneficial that the primer is designed in such way that it can be targeted by the CRISPR-Cas systems provided herein-that is the primer can be targeted by the guide RNA, *e.g.*, crRNA, of the CRISPR-Cas systems so that the CRISPR-Cas systems can be repeatedly used to bind to the target nucleic acid to initiate a new round of amplification. Similarly, in some embodiments, it is also beneficial that the primer is designed in such way that it can be targeted by the Argonaute systems provided herein-that is the primer can be targeted by the 5' phosphorylated single-stranded nucleic acid (guide nucleic acid), *e.g.,* gDNA, of the Argonaute systems so that the Argonaute systems can be repeatedly used to bind to the target nucleic acid to initiate a new round of amplification.

[0146] Thus, in some embodiments, two or more CRISPR-Cas systems or Argonaute systems are used to initiate primer binding at both ends of a target nucleic acid.

[0147] In some embodiments, provided herein is a method for amplifying a target double-stranded nucleic acid including: (a) providing a first system having: a first clustered regularly interspaced short palindromic repeats (CRISPR) RNA (crRNA) or a derivative thereof, and a first CRISPR-associated (Cas) protein or a variant thereof, wherein the first crRNA or the derivative thereof contains a target-specific nucleotide region complementary to a region of a first strand of the target double-stranded nucleic acid; (b) providing second system having: a second clustered regularly interspaced short palindromic repeats (CRISPR) RNA (crRNA) or a derivative thereof, and a second CRISPR-associated (Cas) protein or a variant thereof, wherein the second crRNA or the derivative thereof contains a target-specific nucleotide region complementary to a region of a second strand of the target double-stranded nucleic acid; (c) contacting the target double-stranded nucleic acid with the first system and the second system; (d) hybridizing a first primer to a second strand of the target double-stranded nucleic acid, the

first primer containing a sequence complementary to a region of the second strand of the target double-stranded nucleic acid, and hybridizing a second primer to a first strand of the target double-stranded nucleic acid, the second primer containing a sequence complementary to a region of the first strand of the target double-stranded nucleic acid, and (e) extending the 3' end of the first primer and the second primer with one or more polymerases to generate a first and a second double stranded target nucleic acid. In some embodiments, primer hybridization and extension by polymerase steps are repeated for one or more times, *e.g.,* until a desired degree of amplification is reached.

[0148] In other embodiments, provided herein is a method for amplifying a target double-stranded nucleic acid comprising: (a) providing a first system having: a 5' phosphorylated single-stranded nucleic acid or a derivative thereof, and a first Argonaute protein or a variant thereof, wherein the first 5' phosphorylated single-stranded nucleic acid or the derivative thereof contains a target-specific nucleotide region complementary to a region of a first strand of the target double-stranded nucleic acid; (b) providing second system having: a second 5' phosphorylated single-stranded nucleic acid or a derivative thereof, and a second Argonaute protein or a variant thereof, wherein the second 5' phosphorylated single-stranded nucleic acid or the derivative thereof contains a target-specific nucleotide region complementary to a region of a second strand of the target double-stranded nucleic acid; (c) contacting the target double-stranded nucleic acid with the first system and the second system; (d) hybridizing a first primer to a second strand of the target double-stranded nucleic acid, the first primer containing a sequence complementary to a region of the second strand of the target double-stranded nucleic acid, and hybridizing a second primer to a first strand of the target double-stranded nucleic acid, the second primer containing a sequence complementary to a region of the first strand of the target double-stranded nucleic acid, and (e) extending the 3' end of the first primer and the second primer with one or more polymerases to generate a first and a second double stranded target nucleic acid. In some embodiments, the method provided herein further includes repeating step (a) and step (e) for one or more times, *e.g.,* until a desired degree of amplification is reached.

[0149] In other embodiments, the method provided herein is used for multiplex amplification. As used herein, the term "multiplex amplification" refers to the amplification of more than one nucleic acid of interest, *e.g.,* an amplification of multiple sequences from the same sample. The term "multiplex amplification" also refers to the amplification of one or more sequences present in multiple samples either simultaneously or in step-wise fashion. Thus, in some embodiments, two or more target nucleic acid sequences are being amplified in an amplification reaction, and the amplification reaction comprises the appropriate templates and enzymes to amplify at least two target nucleic acid sequences. One application of the multiplex amplification provided herein is to detect two or more target sequences in a sample because multiplex amplification is cable of amplifying two or more target sequences. When only one of the target sequences is actually present in the sample being tested, the result of the multiplex amplification can be amplification of the only one sequence that is present. Multiplex amplification may utilize the same primer pairs to amplify a one or more intervening amplicon sequences. Alternatively, multiplex amplification may utitilized one or more primer pairs.

[0150] In some embodiments, the double-stranded nucleic acid provided herein is a double-stranded DNA. In other embodiments, the double-stranded nucleic acid provided herein is a double-stranded RNA. In some embodiments, the target nucleic acid is genomic DNA. In other embodiments, the target nucleic acid contains chromosomal DNA or a fragment thereof. In yet other embodiments, the target nucleic acid comprises a genome or a partial genome.

[0151] In some embodiments, the systems provided herein are derived from CRISPR-Cas systems. CRISPR-Cas systems can generally be categorized into three major types (Type I-III), which are further subdivided into ten subtypes, based on core element content and sequences (Makarova et al., 2011, Nat Rev Microbiol 9:467-77). The two key elements of these CRISPR-Cas systems are Cas proteins and CRISPR RNA (crRNA). crRNA consists of short repeat sequences interspersed with spacer sequences derived from invader DNA. Cas proteins have various activities, *e.g.,* nuclease activity. Thus, CRISPR-Cas systems provide mechanisms for targeting a specific sequence as well as certain enzyme activities upon the sequence.

[0152] A typical Type I CRISPR-Cas system contains Cas3 protein with separate helicase and DNase activities. For example, in the Type 1-E system, crRNAs are incorporated into a multisubunit effector complex called Cascade (CRISPR-associated complex for antiviral defense) (Brouns et al., 2008, Science 321: 960- 4), which binds to the target DNA and triggers degradation by the Cas3 protein (Sinkunas et al.,2011, EMBO J 30:1335-1342; Beloglazova et al., 2011, EMBO J 30:616-627).

[0153] Type II CRISPR-Cas systems include the signature Cas9 protein, a single protein (about 160KDa), capable of generating crRNA and cleaving the target DNA. The Cas9 protein typically contains two nuclease domains, a RuvC-like nuclease domain near the amino terminus and the HNH (or McrA-like) nuclease domain near the middle of the protein. Each nuclease domain of the Cas9 protein is specialized for cutting one strand of the double helix (Jinek et al., 2012, Science 337 (6096): 816-821).

[0154] Type III CRISPR-Cas systems contain polymerase and RAMP modules. Type III systems can be further divided into sub-types III-A and III-B. Type III-A CRISPR-Cas systems have been shown to target plasmids, and the polymerase-like proteins of Type III-A systems are involved in the cleavage of target DNA (Marraffini and Sontheimer, 2008, Science 322:1843-1845). Type III-B CRISPR-Cas systems have also been shown to target RNA (Hale et al., 2009, Cell 139:945-956).

[0155] Thus, in some embodiments, the system is a Type I CRISPR-Cas system or a derivative thereof. In other embodiments, the system is a Type II CRISPR-Cas system or a derivative thereof. In yet other embodiments, the system is a Type III CRISPR-Cas system or a derivative thereof.

[0156] The key elements of a CRISPR-Cas system include a guide RNA, *e.g.,* a crRNA, and a

Cas protein. The crRNA or the derivative thereof contains a target specific nucleotide region complementary or substantially complementary to a region of the target nucleic acid. In some embodiments, the crRNA or the derivative thereof contains a user-selectable RNA sequence that permits specific targeting of the enzyme to a complementary double-stranded DNA. In some embodiments, the user-selectable RNA sequence contains 20-50 nucleotides complementary or substantially complementary to a region of the target DNA sequence. In some embodiments, the user-selectable RNA sequence contains less than 20 nucleotides complementary or substantially complementary to a region of the target DNA sequence. Exemplary user-selectable RNA sequence contains 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, or 19 nucelotides complementary or substantially complementary to a region of the target DNA sequence. In some embodiments, the target specific nucleotide region of the crRNA has 100% base pair matching with the region of the target nucleic acid. In some embodiments, the target specific nucleotide region of the crRNA has 90%-100%, 80%-100%, or 70%-100% base pair matching with the region of the target nucleic acid. In some embodiments, there is one base pair mismatch between the target specific nucleotide region of the crRNA and the region of the target nucleic acid. In some embodiments, there are two base pair mismatches between the target specific nucleotide region of the crRNA and the region of the target nucleic acid. In some embodiments, there are three base pair mismatches between the target specific nucleotide region of the crRNA and the region of the target nucleic acid. In some embodiments, there are four base pair mismatches between the target specific nucleotide region of the crRNA and the region of the target nucleic acid. In some embodiments, there are five base pair mismatches between the target specific nucleotide region of the crRNA and the region of the target nucleic acid.

[0157] In some embodiments, the system provided herein further includes a trans-activating crRNA (tracrRNA) or a derivative thereof.

[0158] The CRISPR-Cas systems provided herein include engineered and/or programmed nuclease systems derived from naturally occurring CRISPR-Cas systems. CRISPR-Cas systems may include contain engineered and/or mutated Cas proteins. CRISPR-Cas systems may also contain engineered and/or programmed guide RNA. In some embodiments, the crRNA or the derivative thereof provided herein is a polynucleotide having a crRNA polynucleotide fused to a tracrRNA polynucleotide. A chimeric single-guided RNA (sgRNA) is described in Jinek et al., 2012, Science 337, 816-821, which is incorporated herein in its entirety. In one embodiment, the Cas protein or the variant thereof provided herein can be directed by a chimeric sgRNA to any genomic locus followed by a 5'-NGG protospacer-adjacent motif (PAM). For example, in some embodiments, crRNA and tracrRNA are synthesized by *in vitro* transcription, using a synthetic double stranded DNA template containing the T7 promoter. The tracrRNA has a fixed sequence, whereas the target sequence dictates part of crRNA's sequence. Equal molarities of crRNA and tracrRNA are mixed and heated at 55°C for 30 seconds. Cas9 is added at the same molarity at 37°C and incubated for 10 minutes with the RNA mix. 10-20 fold molar excess of Cas9 complex is then added to the target DNA. The binding reaction can occur within 15 minutes.

[0159] In some embodiments, the Cas protein or the variant thereof is a Cas9 protein or a variant thereof. Isolated Cas9-crRNA complex from the *S. thermophilus* CRISPR-Cas system as well as complex assembled *in vitro* from separate components demonstrate that it binds to both synthetic oligodeoxynucleotide and plasmid DNA bearing a nucleotide sequence complementary to the crRNA. It has been shown that Cas9 has two nuclease domains-RuvC- and HNH- active sites/nuclease domains, and these two nuclease domains are responsible for the cleavage of opposite DNA strands. In some embodiments, the Cas9 protein is derived from Cas9 protein of *S. thermophilus* CRISPR-Cas system. In some embodiments, the Cas9 protein is a multi-domain protein having about 1,409 amino acids residues.

[0160] In some embodiments, the Cas9 protein or the variant thereof is a nuclease-null variant of the Cas9 protein, in which both RuvC- and HNH- active sites/nuclease domains are mutated. A nuclease-null variant of the Cas9 protein binds to double-stranded DNA, but not cleave the DNA, and thus it can be used for target specific DNA enrichment too. In some embodiments, the Cas9 protein has two inactivated nuclease domains with a first mutation in the domain that cleaves the strand complementary to the crRNA and a second mutation in the domain that cleaves the strand non-complementary to the crRNA. In some embodiments, the Cas9 protein has a first mutation D10A and a second mutation H840A.

[0161] In some embodiments, the Cas protein or the variant thereof is a Cascade protein or a variant thereof. Cascade complex in *E. coli* recognizes double-stranded DNA (dsDNA) targets in a sequence-specific manner. *E. coli* Cascade complex is a 405-kDa complex comprising five functionally essential CRISPR-associated (Cas) proteins (CasA1B2C6D1E1, also called Cascade protein) and a 61-nucleotide crRNA. The crRNA guides Cascade complex to dsDNA target sequences by forming base pairs with the complementary DNA strand while displacing the noncomplementary strand to form an R-loop. Cascade recognizes target DNA without consuming ATP, which suggests that continuous invader DNA surveillance takes place without energy investment. Matthijs et al., Nature Structural & Molecular Biology, 2011, 18,529-536.

[0162] In some embodiments, the Cas protein or the variant thereof is a Cas3 protein or a variant thereof. *E. coli* Cas3 can catalyse ATP-independent annealing of RNA with DNA forming R-loops, and hybrid of RNA base-paired into duplex DNA. Cas3 protein can use gRNA that is longer than that for Cas9. Howard et al., Biochem J., 2011, 439(1):85-95. Such longer RNA can permit easier access of other elements to the target DNA, *e.g.*, access of a primer to be extended by polymerase. Another advantage provided by Cas3 protein is that Cas3 protein does not require a PAM sequence as Cas9, thus provides more flexibility for targeting desired sequence. R-loop formation by Cas3 may require magnesium as a co-factor. Howard et al., Biochem J., 2011, 439(1):85-95. Thus, in some embodiments, the system provided herein further comprises magnesium.

[0163] It should be appreciated that any CRISPR-Cas systems capable of disrupting the double stranded nucleic acid and creating a loop structure can be used in the present methods. For example, the Cas proteins provided herein may include, but not limited to, the Cas proteins described in Haft et al., PLoS Comput Biol., 2005, 1(6): e60, and Zhang et al.,

Nucl. Acids Res., 2013, 10.1093/nar/gkt1262. Some these CRISPR-Cas systems require that a specific sequence be present for these CRISPR-Cas systems to recognize and bind to the target sequence. For instance, Cas9 requires the presence of a 5'-NGG protospacer-adjacent motif (PAM). Thus, in some embodiments, a PAM sequence or a sequence complementary to a PAM sequence is engineered into the target nucleic acid for initiating the binding of the CRISPR-Cas systems to the target nucleic acid.

[0164] In other embodiments, the systems provided herein are derived from Argonaute systems. Argonaute protein family was first discovered genetically, and was first identified in plants. Members are defined by the presence of PAZ (Piwi-Argonaute-Zwille) and PIWI domains. Argonaute proteins are highly conserved between species and many organisms encode multiple members of the family. Numbers of Argonaute genes range from 1 in the fission yeast Schizosaccharomyces pombe to 27 in the nematode worm Caenorhabditis elegans. In mammals there are eight Argonaute genes. The Argonaute protein family can be divided into the Ago subfamily and the Piwi subfamily. In most organisms investigated so far, which include Drosophila, the zebrafish and the mouse, the expression of Piwi proteins is restricted to the germ line, where they bind Piwi-interacting proteins (piRNAs). In contrast, Ago proteins are ubiquitously expressed in many organisms. Human Ago1, Ago3 and Ago4 genes are clustered on chromosome 1, whereas the Ago2 gene is located on chromosome 8. The human Piwi subfamily comprises HIWI1, HIWI2, HIWI3 and HILI; they are encoded by genes on chromosomes 12, 11, 22 and 8, respectively. It was shown that members of the Argonaute protein family are key players in gene-silencing pathways guided by small RNAs. Small RNAs such as short interfering RNAs (siRNAs), microRNAs (miRNAs) or Piwi-interacting RNAs (piRNAs) are anchored into specific binding pockets and guide Argonaute proteins to target mRNA molecules for silencing or destruction. Hock and Meister, Genome Biol,. 2008, 9(2): 210. Recently, it was also shown that certain Argonaute proteins are guided by small DNA for performing DNA cleavage. *See, e.g.,* Gao et al., nature biotechnology, 2016, doi: 10.1038/nbt.3547. Thus, Argonaute systems provide mechanisms for targeting a specific sequence as well as certain enzyme activities upon the sequence.

[0165] Thus, in some embodiments, the system is an Ago subfamily system or a derivative thereof. In other embodiments, the system is a Piwi subfamily system or a derivative thereof. In some specific embodiments, the Argonaute protein is Natronobacterium gregorgi Argonaute (Ng Argonaute) or a variant thereof.

[0166] The key elements of an Argonaute system include a guide nucleic acid, *e.g.,* 5' phosphorylated single-stranded nucleic acid. The 5' phosphorylated single-stranded nucleic acid or the derivative thereof contains a target specific nucleotide region complementary or substantially complementary to a region of the target nucleic acid. In some embodiments, the 5' phosphorylated single-stranded nucleic acid or the derivative thereof contains a user-selectable DNA sequence that permits specific targeting of the enzyme to a complementary double-stranded DNA. In some embodiments, the user-selectable DNA sequence contains 20-50 nucleotides complementary or substantially complementary to a region of the target DNA sequence. In some embodiments, the 5' phosphorylated single-stranded nucleic acid is user-

selectable single-stranded DNA sequence of less than 100 nucleotides. In some embodiments, the 5' phosphorylated single-stranded nucleic acid is a single-stranded DNA of less than 50 nucleotides. In some embodiments, the 5' phosphorylated single-stranded nucleic acid is a single-stranded DNA of less than 25 nucleotides. Example lengths of the single-stranded DNA include 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, and 50 nucleotides. In some embodiments, the user-selectable DNA sequence contains less than 20 nucleotides complementary or substantially complementary to a region of the target DNA sequence. Exemplary user-selectable DNA sequence contains 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, or 19 nucelotides complementary or substantially complementary to a region of the target DNA sequence.

[0167] In some embodiments, the target specific nucleotide region of the 5' phosphorylated single-stranded nucleic acid has 100% base pair matching with the region of the target nucleic acid. In some embodiments, the target specific nucleotide region of the 5' phosphorylated single-stranded nucleic acid has 90%-100%, 80%-100%, or 70%-100% base pair matching with the region of the target nucleic acid. In some embodiments, there is one base pair mismatch between the target specific nucleotide region of the crRNA and the region of the target nucleic acid. In some embodiments, there are two base pair mismatches between the target specific nucleotide region of the 5' phosphorylated single-stranded nucleic acid and the region of the target nucleic acid. In some embodiments, there are three base pair mismatches between the target specific nucleotide region of the 5' phosphorylated single-stranded nucleic acid and the region of the target nucleic acid. In some embodiments, there are four base pair mismatches between the target specific nucleotide region of the 5' phosphorylated single-stranded nucleic acid and the region of the target nucleic acid. In some embodiments, there are five base pair mismatches between the target specific nucleotide region of the 5' phosphorylated single-stranded nucleic acid and the region of the target nucleic acid.

[0168] The Argonaute systems provided herein include engineered and/or programmed nuclease systems derived from naturally occurring Argonaute systems. Argonaute systems may include contain engineered and/or mutated Argonaute proteins. Argonaute systems may also contain engineered and/or programmed guide nucleic acid. In some embodiments, the Argonaute protein or a variant thereof is engineered to remove the endonuclease activity so that the Argonaute system does not generate double-stranded or single-stranded DNA breaks. In other embodiments, the Argonaute protein or a variant thereof is engineered to be able to generate single-stranded DNA breaks.

[0169] It should be appreciated that any Argonaute systems capable of binding to a nucleic acid through a guide nucleic acid, or disrupting the double stranded nucleic acid and creating a loop structure can be used in the present methods.

[0170] In some embodiments, the primer provided herein is a single-stranded oligodeoxyribonucleotide. In some embodiments, the length of a primer ranges from about 10 to about 50 nucleotides. Exemplary length of the primer provided herein is 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, or 50 nucleotides.

[0171] In some embodiments, the primer provided herein has 100% base pair matching with a region of the target nucleic acid. In some embodiments, the primer provided herein has 90%-100%, 80%-100%, or 70%-100% base pair matching with the region of the target nucleic acid. In some embodiments, there is one base pair mismatch between the primer and the region of the target nucleic acid. In some embodiments, there are two base pair mismatches between the primer and the region of the target nucleic acid. In some embodiments, there are three base pair mismatches between the primer and the region of the target nucleic acid. In some embodiments, there are four base pair mismatches between the primer and the region of the target nucleic acid. In some embodiments, there are five base pair mismatches between the primer and the region of the target nucleic acid.

[0172] In some embodiments, the primer provided herein is labeled, *e.g.,* for enrichment or detection of the amplified products. In some embodiments, the primer provided herein is labeled for detection by spectroscopic, photochemical, biochemical, immunochemical, or chemical means. In some embodiments, the primer is labeled with biotin, amine, radiolabels (e.g., 32P), fluorescent dyes, electron-dense reagents, enzymes (as commonly used in ELISAS), or biotin.

[0173] In some embodiments, the polymerase provided herein catalyzes incorporation of nucleotides to extend a 3' hydroxyl terminus of a primer molecule against a nucleic acid target sequence. In some embodiments, the reaction by the polymerase provided herein is performed in isothermal conditions. In some embodiments, the polymerase is a strand-displacing polymerase. Exemplary polymerases include but not limited to Bst DNA polymerase, 9°Nm DNA polymerase, Phi29 DNA polymerase, DNA polymerase I (E. coli), DNA polymerase I (Large)m, (Klenow) fragment, Klenow fragment (3'-5' exo-), T4 DNA polymerase, T7 DNA polymerase, Deep VentR™ (exo-) DNA polymerase, Deep VentR™ DNA polymerase, DyNAzyme™ EXT DNA, DyNAzyme™ II Hot Start DNA Polymerase, Phusion™ High-Fidelity DNA Polymerase, Therminator™ DNA Polymerase, Therminator™ II DNA Polymerase, VentR® DNA Polymerase, VentR® (exo-) DNA Polymerase, RepliPHI™ Phi29 DNA Polymerase, rBst DNA Polymerase, rBst DNA Polymerase (Large), Fragment (IsoTherm™ DNA Polymerase), MasterAmp™ AmpliTherm™, DNA Polymerase, Taq DNA polymerase, Tth DNA polymerase, Tfl DNA polymerase, Tgo DNA polymerase, SP6 DNA polymerase, Tbr DNA polymerase, DNA polymerase Beta, and ThermoPhi DNA polymerase. In some embodiments, the polymerase is selected from a group consisting of Bst, Bsu, and Phi29. As the polymerase extends the hybridized strand, it can be beneficial to include single-stranded binding protein (SSB). SSB may stabilize the displaced (non-template) strand. Thus, in some embodiments, the method provided herein can further include SSB protein.

[0174] As discussed, one advantage provided by the linear or exponential amplifications provided herein is that the amplifications can be performed under constant temperature or isothermal conditions. As used herein, the term "constant temperature," "isothermal

conditions," or "isothermally" refers to a set of reaction conditions where the temperature of the reaction is kept essentially constant during the course of the amplification reaction. However, it is not necessary that the temperature be maintained at precisely one temperature. If the equipment used to maintain an elevated temperature allows the temperature of the reaction mixture to vary by a few degrees, this is not detrimental to the amplification reaction, and can still be considered to be an isothermal reaction. In some embodiments, the amplification reaction is run at a constant temperature between 20°C to 70°C. In some embodiments, the amplification reaction is run at a constant temperature between 25°C to 60°C. In some embodiments, the amplification reaction is run at a constant temperature between 40°C to 55°C. In some embodiments, the amplification reaction is run at a constant temperature between 30°C to 40°C. Exemplary amplification reaction temperatures include 30°C, 31°C, 32°C, 33 °C, 34 °C, 35 °C, 36 °C, 37 °C, 38 °C, 39 °C, 40 °C, 41 °C, 42 °C, 43 °C, 44 °C, 45 °C, 46 °C, 47 °C, 48 °C, 49 °C, 50 °C, 51 °C, 52 °C, 53 °C, 54 °C, 55 °C, 56 °C, 57 °C, 58 °C, 59 °C, and 60 °C. In one embodiment, the amplification reation temperature is about 37 °C.

[0175] In some embodiments, the time that the amplification reaction is run may vary from, for example, 1 minute to several hours until the desired amount of the amplification is achieved. In some embodiments, the time that the amplification reaction is 5 minutes to 1 hour. Exemplary times of the amplification reaction include 5 minutes, 10 minutes, 15 minutes, 20 minutes, 30 minutes, 40 minutes, 50 minutes, and 60 minutes.

[0176] In some embodiments, the present methods can be used in certain applications wherein detection and/or quantification of the amplified nucleic acid products are desired. The amplified target sequence can be detected by any method known to one of ordinary skill in the art.

[0177] In some embodiments, dyes that specifically stain double-stranded DNA can be used to detect or quantify the amplified products. In some embodiments, intercalating dyes exhibit enhanced fluorescence upon binding to DNA or RNA. In some embodiments, dyes can be, for example, DNA or RNA intercalating fluorophores. Exemplary DNA or RNA intercalating fluorophores include but are not limited to Acridine orange, ethidium bromide, Hoechst dyes, PicoGreen, propidium iodide, SYBR I (an asymmetrical cyanine dye), SYBR II, TOTO (a thiaxole orange dimer) and YOYO (an oxazole yellow dimer).

[0178] In some embodiments, amplified products with specific sizes can be detected by gel electrophoresis. In some embodiments, the nucleotides used in the amplification reaction, can be labeled, e.g., with biotin. Biotin-labeled amplified sequences can be captured using avidin bound to a signal generating enzyme, for example, peroxidase.

[0179] In some embodiments, labeled nucleotides can be incorporated directly into the target sequence or into primers containing complementary sequences to the target of interested. Such labels can be radioactive and/or fluorescent in nature and can be resolved in any of the manners discussed herein.

[0180] Methods of detecting and/or continuously monitoring the amplification of nucleic acid products are also well known to those skilled in the art. For example, in some embodiments, the production or presence of target nucleic acids and nucleic acid sequences may be detected and monitored by Molecular Beacons. Molecular Beacons are hairpin shaped oligonucleotides containing a fluorophore on one end and a quenching dye on the opposite end. The loop of the hair-pin contains a probe sequence that is complementary to a target sequence and the stem is formed by annealing of complementary arm sequences located on either side of the probe sequence. When the molecular beacon encounters a target molecule, hybridization occurs; the loop structure is converted to a stable more rigid conformation causing separation of the fluorophore and quencher molecules leading to fluorescence. Tyagi et al., Nature Biotechnology, 1996, 303-308. Thus, the generation of fluorescence indicates the synthesis of the intended amplified product. For another example, in some embodiments, the production or presence of target nucleic acids and nucleic acid sequences can be detected and monitored by Fluorescence resonance energy transfer (FRET). FRET is a useful tool to quantify molecular dynamics, for example, in DNA-DNA interactions. For monitoring the production of a specific product a probe can be labeled with a donor molecule on one end and an acceptor molecule on the other. Probe-target hybridization brings a change in the distance or orientation of the donor and acceptor and FRET change is observed. Lakowicz, "Principles of Fluorescence Spectroscopy," Plenum Publishing Corporation, 2nd edition (Jul. 1, 1999). Other exemplary methods for detecting and/or continuously monitoring the amplification of nucleic acid products include but not limited to Mass Spectrometry, capillary gel electrophoresis, sequencing, and various surface capture methods as known to those skilled in the art. In some embodiments, running the amplification reaction with asymmetric amounts of primers, *i.e.,* with one primer present in higher concentration than the other, would permit amplification with a bias towards one strand. An excess of a single strand may facilitate detection by such modalities as Molecular Beacons, which detect ssDNA.

[0181] The amplification methods provided herein can be used in various applications. For example, in some embodiments, the method provided herein can be used to isolate DNA fragments from genomic DNA by selective amplification of a specific region of DNA.

[0182] For another example, in some embodiments, the amplification methods provided herein can be used for diagnosing various diseases based on the production or presence of a target nucleic acid linked to a specific disease. In some embodiments, the methods provided herein can be used for early diagnosis of malignant diseases, *e.g.,* leukemia and lymphomas. In other embodiments, methods provided herein can be used directly on genomic DNA samples to amplify and detect translocation-specific malignant cells.

[0183] In other embodiments, the methods provided herein can be used for diagnosis of infectious diseases, including those caused by bacteria or viruses. For example, the present methods can be used to detect infectious agents and distinguish non-pathogenic from pathogenic strains by virtue of specific nucleic acid sequence. In some embodiments, the amplification methods provided herein can amplify and identify non-cultivatable or slow-growing microorganisms, *e.g.,* mycobacteria, anaerobic bacteria, or viruses from tissue culture

assays and animal models. In other embodiments, the amplification methods provided herein can amplify and identify viral DNA in a sample from an individual.

[0184] In yet other embodiments, the present amplification methods can be used to generate nucleic acid materials to augment other procedures. For example, in some embodiments, the present amplification methods can be used for generating hybridization probes for Southern or northern hybridization and DNA cloning, which require relatively larger amounts of DNA, representing a specific DNA region. For another example, the methods provided herein can be used for nucleic acid sequencing.

[0185] In a specific embodiment, the CRISPR-Cas or Argonaute mediated nucleic acid amplification methods provided herein can be used for amplifying a library of nucleic acid fragments, *e.g.,* generated using library preparation methods and/or kits available from Illumina, Inc. (San Diego, CA).

[0186] In some embodiments, the methods provided herein can be used to amplify a library of nucleic acid fragments generated from genomic DNA. In a specific embodiment, the library of nucleic acid fragments is generated using tagmentation. In a specific embodiment, the library of nucleic acid fragments is generated from genomic DNA using Illumina's Nextera library preparation methods and kits (available from Illumina, Inc, San Diego, CA).

[0187] Thus, in some embodiments, the method provided herein further includes applying at least one transposase and at least one transposon end composition containing a transferred strand to a sample containing a target nucleic acid under conditions where the target nucleic acid and the transposon end composition undergo a transposition reaction to generate a mixture, wherein the target nucleic acid is fragmented to generate a plurality of target nucleic acid fragments, and thus incorporates a universal primer sequence into each of the plurality of target nucleic acid fragments, wherein the guide nucleic acid in the nuclease system, such a crRNA or the derivative thereof or a 5' phosphorylated single-stranded nucleic acid, contains a target-specific nucleotide region complementary to a region of the universal primer.

[0188] In some embodiments, the target nucleic acid is subjected to a transposase mediated tagmentation that results in fragmentation of the target nucleic acid and ligation of adaptors to the 5' end of both strands of double-stranded DNA fragments. Optionally, the target nucleic acid can be fragmented and adaptors can be added to the 5' and 3' ends using tagmentation or transposition as described in U.S. Publication No. 2010/0120098. Briefly, a transposition reaction is a reaction wherein one or more transposons are inserted into target nucleic acids at random sites. Essential components in a transposition reaction are a transposase and DNA oligonucleotides that exhibit the nucleotide sequences of a transposon, including the transferred transposon sequence and its complement (the non- transferred transposon end sequence) as well as other components needed to form a functional transposition or transposome complex. The DNA oligonucleotides can further include additional sequences (*e.g.*, adaptor or primer sequences) as needed or desired. Exemplary transposition complexes, suitable for use in the methods provided herein, include, but are not limited to, those formed by

a hyperactive Tn5 transposase and a Tn5-type transposon end or by a MuA transposase and a Mu transposon end comprising R1 and R2 end sequences (*see, e.g.,* Goryshin and Reznikoff, J. Biol. Chem. 273: 7367, 1998; and Mizuuchi, Cell 35: 785, 1983; Savilahti et al., EMBO J. 14: 4893, 1995;). However, any transposition system that is capable of inserting a transposon end with sufficient efficiency to tag target nucleic acids for its intended purpose can be used in the provided methods. Other examples of known transposition systems that could be used in the provided methods include, but are not limited to, Staphylococcus aureus Tn552, Ty1, Transposon Tn7, Tn/O and IS10, Mariner transposase, Tel, P Element, Tn3, bacterial insertion sequences, retroviruses, and retrotransposon of yeast (*see, e.g.,* Colegio et al., 2001, J. Bacteriol. 183: 2384-8; kirby et al., 2002, Mol. Microbiol. 43: 173-86; Devine and Boeke, 1994, Nucleic Acids Res., 22: 3765-72; International Patent Application No. WO 95/23875; Craig, 1996, Science 271 : 1512; Craig, 1996, Review in: Curr Top Microbiol Immunol. 204: 27-48; Kleckner et al., 1996, Curr Top Microbiol Immunol. 204: 49-82; Lampe et al., 1996, EMBO J. 15: 5470-9; Plasterk, 1996, Curr Top Microbiol Immunol 204: 125-43; Gloor, 2004, Methods Mol. Biol. 260: 97-114; Ichikawa and Ohtsubo, 1990, J Biol. Chem. 265: 18829-32; Ohtsubo and Sekine, 1996, Curr. Top. Microbiol. Immunol. 204: 1-26; Brown et al., 1989, Proc Natl Acad Sci USA 86: 2525-9; Boeke and Corces, 1989, Annu Rev Microbiol. 43: 403-34;). In some embodiments, the method of the present disclosure further comprises removing the transposase enzyme and adding to the ends of the adapted DNA fragments by PCR.

[0189] The term "tagmentation," "tagment," or "tagmenting," as used herein, refers to transforming a nucleic acid, *e.g.,* a DNA, into adaptor-modified templates in solution ready for cluster formation and sequencing by the use of transposase mediated fragmentation and tagging. This process often involves the modification of the nucleic acid by a transposome complex comprising transposase enzyme complexed with adaptors comprising transposon end sequence. Tagmentation results in the simultaneous fragmentation of the nucleic acid and ligation of the adaptors to the 5' ends of both strands of duplex fragments. Following a purification step to remove the transposase enzyme, additional sequences are added to the ends of the adapted fragments by PCR.

[0190] The term "transposome complex," as used herein, refers to a transposase enzyme non-covalently bound to a double stranded nucleic acid. For example, the complex can be a transposase enzyme preincubated with double-stranded transposon DNA under conditions that support non-covalent complex formation. Double-stranded transposon DNA can include, without limitation, Tn5 DNA, a portion of Tn5 DNA, a transposon end composition, a mixture of transposon end compositions or other double-stranded DNAs capable of interacting with a transposase such as the hyperactive Tn5 transposase.

[0191] A "transposase" means an enzyme that is capable of forming a functional complex with a transposon end-containing composition (*e.g.*, transposons, transposon ends, transposon end compositions) and catalyzing insertion or transposition of the transposon end-containing composition into the double-stranded target nucleic acid with which it is incubated, for example, in an in vitro transposition reaction. A transposase as presented herein can also include integrases from retrotransposons and retroviruses. Transposases, transposomes and

transposome complexes are generally known to those of skill in the art, as exemplified by the disclosure of US 2010/0120098. Although many embodiments described herein refer to Tn5 transposase and/or hyperactive Tn5 transposase, it will be appreciated that any transposition system that is capable of inserting a transposon end with sufficient efficiency to 5'-tag and fragment a target nucleic acid for its intended purpose can be used in the present invention. In particular embodiments, a preferred transposition system is capable of inserting the transposon end in a random or in an almost random manner to 5'-tag and fragment the target nucleic acid.

[0192] As used herein, the term "transposition reaction" refers to a reaction wherein one or more transposons are inserted into target nucleic acids, *e.g.,* at random sites or almost random sites. Essential components in a transposition reaction are a transposase and DNA oligonucleotides that exhibit the nucleotide sequences of a transposon, including the transferred transposon sequence and its complement (the non- transferred transposon end sequence) as well as other components needed to form a functional transposition or transposome complex. The DNA oligonucleotides can further comprise additional sequences (*e.g.*, adaptor or primer sequences) as needed or desired. In some embodiments, the method provided herein is exemplified by employing a transposition complex formed by a hyperactive Tn5 transposase and a Tn5-type transposon end (Goryshin and Reznikoff, 1998, J. Biol. Chem., 273: 7367) or by a MuA transposase and a Mu transposon end comprising R1 and R2 end sequences (Mizuuchi, 1983, Cell, 35: 785; Savilahti et al., 1995, EMBO J., 14: 4893). However, any transposition system that is capable of inserting a transposon end in a random or in an almost random manner with sufficient efficiency to 5'- tag and fragment a target DNA for its intended purpose can be used in the present invention. Examples of transposition systems known in the art which can be used for the present methods include but are not limited to Staphylococcus aureus Tn552 (Colegio et al., 2001, J Bacterid., 183: 2384-8; Kirby et al., 2002, Mol Microbiol, 43: 173-86), Tyl (Devine and Boeke, 1994, Nucleic Acids Res., 22: 3765-72 and International Patent Application No. WO 95/23875), Transposon Tn7 (Craig, 1996, Science. 271 : 1512; Craig, 1996, Review in: Curr Top Microbiol Immunol, 204: 27-48), TnIO and ISIO (Kleckner et al., 1996, Curr Top Microbiol Immunol, 204: 49-82), Mariner transposase (Lampe et al., 1996, EMBO J., 15: 5470-9), Tci (Plasterk, 1996, Curr Top Microbiol Immunol, 204: 125-43), P Element (Gloor, 2004, Methods Mol Biol, 260: 97-114), TnJ (Ichikawa and Ohtsubo, 1990, J Biol Chem. 265: 18829-32), bacterial insertion sequences (Ohtsubo and Sekine, 1996, Curr. Top. Microbiol. Immunol. 204:1-26), retroviruses (Brown et al., 1989, Proc Natl Acad Sci USA, 86: 2525-9), and retrotransposon of yeast (Boeke and Corces, 1989, Annu Rev Microbiol. 43: 403-34). The method for inserting a transposon end into a target sequence can be carried out *in vitro* using any suitable transposon system for which a suitable in vitro transposition system is available or that can be developed based on knowledge in the art. In general, a suitable in vitro transposition system for use in the methods provided herein requires, at a minimum, a transposase enzyme of sufficient purity, sufficient concentration, and sufficient *in vitro* transposition activity and a transposon end with which the transposase forms a functional complex with the respective transposase that is capable of catalyzing the transposition reaction. Suitable transposase transposon end sequences that can be used in the invention include but are not limited to wild-type, derivative or mutant

transposon end sequences that form a complex with a transposase chosen from among a wild-type, derivative or mutant form of the transposase.

[0193] The term "transposon end" (TE) refers to a double-stranded nucleic acid, *e.g.,* a double-stranded DNA, that exhibits only the nucleotide sequences (the "transposon end sequences") that are necessary to form the complex with the transposase or integrase enzyme that is functional in an in vitro transposition reaction. In some embodiments, a transposon end is capable of forming a functional complex with the transposase in a transposition reaction. As non-limiting examples, transposon ends can include the 19-bp outer end ("OE") transposon end, inner end ("IE") transposon end, or "mosaic end" ("ME") transposon end recognized by a wild-type or mutant Tn5 transposase, or the R1 and R2 transposon end as set forth in the disclosure of US 2010/0120098. Transposon ends can include any nucleic acid or nucleic acid analogue suitable for forming a functional complex with the transposase or integrase enzyme in an in vitro transposition reaction. For example, the transposon end can include DNA, RNA, modified bases, non-natural bases, modified backbone, and can include nicks in one or both strands. Although the term "DNA" is sometimes used in the present disclosure in connection with the composition of transposon ends, it should be understood that any suitable nucleic acid or nucleic acid analogue can be utilized in a transposon end.

[0194] Through an *in vitro* transposition reaction, target nucleic acid fragments are tagged at the 5' end. In some embodiments, the method provided herein further includes steps to incorporate a 3' end tag to the 5' tagged nucleic acid fragments to make a library of di-tagged nucleic acid fragments. Adding 3' end tag can be performed through various methods, *e.g.*, by using DNA polymerase, terminal transferase, and/or ligase as described in WO 2010/048605.

[0195] In some embodiments, di-tagged nucleic acid fragments are generated by using a polymerase, *e.g.,* a DNA polymerase, with strand-displacement or 5' nuclease activity. In some embodiments, the method provided herein includes incubating the population of annealed 5'-tagged nucleic acid fragments with a DNA polymerase that has strand-displacement or 5' nuclease activity under conditions without thermocycling and wherein the annealed 5 '-tagged nucleic acid fragments are not denatured, wherein the DNA polymerase extends the 3 '-end of each strand of the annealed 5 '-tagged nucleic acid fragments using the complementary strand as a template and displaces or digests the non-transferred strand, thereby generating the library of di-tagged double-stranded DNA fragments.

[0196] In other embodiments, the 5'-tagged nucleic acid fragments are incubated with a DNA polymerase consisting of a terminal transferase and at least one substrate for the terminal transferase under conditions and for sufficient time wherein the terminal transferase joins the second tag to the 3' end of the 5'-tagged nucleic acid fragments, thereby generating a library of di-tagged nucleic acid fragments. In some embodiments, the 3 '-end of the non-transferred transposon end that composes the transposon end composition is blocked (*e.g.,* by using a non-transferred transposon end that has a dideoxy nucleotide or a 3'-O-methyl-nucleotide as the 3'-terminal nucleotide).

[0197] In yet other embodiments, di-tagged nucleic acid fragments are generated by using a template-dependent ligase and a ligation tagging oligonucleotide. In some embodiments, the 5'-tagged nucleic acid fragments are incubated with a template-dependent DNA ligase and a ligation tagging oligodeoxynucleotide having a 3 '-portion and a 5'- portion, wherein the 3'-portion exhibits a second tag that exhibits any sequence that is desired to be joined to the 3'-end of the 5'-tagged DNA fragments and the 5'-portion has a 5 '-monophosphate group and exhibits a random sequence, under conditions and for sufficient time wherein the second tag is joined to the annealed 5'-tagged DNA fragments, thereby generating a library of DNA fragments comprising annealed di-tagged DNA fragments.

[0198] In some embodiments, the nucleic acid fragments generated contain universal sequences at the two ends of the nucleic acid fragments. The universal sequences at the two ends of the nucleic acid fragments, e.g., in a library, can be targeted to by the CRISPR-Cas system or the Argonaute system according the methods provided herein, and as such the fragments can be amplified, *e.g.,* in a cluster formation reaction.

[0199] Such universal sequences can be introduced into the two ends of the nucleic acid fragments in a library by a PCR or Nextera transposon reaction. In some embodiments, after a library of tagged nucleic acid fragments is generated, the tagged nucleic acid fragments can be amplified, e.g., using limited-cycle polymerase chain reaction (PCR), to introduce other end sequences or adaptors, e.g., index, universal primers and other sequences required for cluster formation and sequencing. In a specific embodiment, a limited-cycle PCR amplification is performed to add index 1 (P7) and index 2 (P5) (available from Illumina, Inc, San Diego, CA) to the two ends of the nucleic acid fragments.

[0200] In some embodiments, such amplification is performed to a library of 5' tagged nucleic acid fragments. In some embodiments, such amplification is performed to a library of di-tagged nucleic acid fragments. Exemplary amplification methods include polymerase chain reaction (PCR), strand-displacement amplification reaction, rolling circle amplification reaction, ligase chain reaction, transcription-mediated amplification reaction, and loop-mediated amplification reaction.

[0201] In some embodiments, the method provided herein includes amplifying the library of di-tagged nucleic acid fragments using a PCR. In some embodiments, provided herein includes amplifying the library of di-tagged nucleic acid fragments using the Cas9 or Argonaute mediated amplification method provided herein. In some embodiments, the method provided herein uses single-primer PCR amplification of a library of di-tagged DNA fragments. In some embodiments, the step of amplifying di-tagged DNA fragments includes using a DNA polymerase and at least one primer that is complementary to the second tag. In some embodiments, the step of amplifying the library of di-tagged DNA fragments includes amplifying the library of tagged DNA fragments by PCR using only one oligodeoxyribonucleotide that exhibits the sequence of at least a portion of the transferred strand as a PCR primer and the di-tagged DNA fragments as templates. In some embodiments, the primer contains a 5' portion that contains additional sequence, *e.g.,* an adaptor sequence.

[0202] In some embodiments, two different PCR primers are used, each of which PCR primers exhibits the sequence of at least a portion of the transferred transposon end that composes the transposon end composition. In some embodiments, each PCR primer includes a 3'-portion and a 5 '-portion, wherein the 3 '-portion exhibits the respective transferred transposon end sequence and the 5 '-portion exhibits the sequence of a respective tag domain or an adaptor for a particular purpose (*e.g.,* a sequencing tag domain/adaptor or an amplification tag domain/adaptor, and optionally an address tag domain/adaptor for next-generation sequencing or amplification). For example, when a single transposon end composition is used in the *in vitro* transposition reaction to generate the library of di-tagged DNA fragments using a DNA polymerase that has strand-displacement or 5' nuclease activity, the di-tagged DNA fragments can be amplified by PCR using two different PCR primers. Each PCR primer contains a 3'-portion and a 5 '-portion, wherein the 3 '-portion exhibits the respective transferred transposon end sequence and the 5 '-portion exhibits the sequence of a respective tag domain/adaptor for a particular purpose (*e.g.,* a sequencing tag domain/adaptor or an amplification tag domain/adaptor, and optionally an address tag domain/adaptor for next-generation sequencing or amplification). In some embodiments, the 5' portion of each PCR primer is different from that of the other primer, and as such the sequences of the two ends of the PCR product are different. For example, one end contains one index and/or universal primer sequence, and the other end contains a different index and/or universal primer sequence.

[0203] In some embodiments, the two ends of di-tagged nucleic acid fragments originate from two different transferred strand sequences. For example, in some embodiments, two different transposomes can be used in the *in vitro* transposition reaction, and each of the two transposomes contains the same transposase but a different transposon end composition. In some embodiments, two different transposomes are used, and the two different transposomes each contains the same transposase and the transposon end compositions contain different transferred strands. In some embodiments, two different transposomes are used, and each of the two transposomes includes different transposase enzymes and different transposon end compositions, each of which forms a functional complex with the respective transposase. In some embodiments, wherein two different transposon end compositions are used in the *in vitro* transposition reaction, and the library of di-tagged single stranded nucleic acid fragments is generated using a DNA polymerase that has strand- displacement or 5' nuclease activity, the first tag exhibits the sequence of the transferred strand of one transposon end composition and the second tag exhibits the sequence of the non-transferred strand of the other transposon end composition.

[0204] In the above mentioned embodiments and other embodiments wherein two different transferred strands are linked to the 5' end of each opposite strands of the double stranded nucleic acid, the method provided herein can further include the step of amplifying the di-tagged nucleic acid fragments by PCR using two different PCR primers. One of the PCR primers exhibits the sequence of at least a portion of one transferred strand that compose one transposon end composition, and the other of PCR primers exhibits the sequence of at least a portion of the other transferred strand that composes the other transposon end composition.

[0205] In some embodiments wherein two primers are used, each PCR primer contains a 3 '-portion and a 5 '-portion, wherein the 3 '-portion exhibits the respective transferred transposon end sequence and the 5 '-portion exhibits the sequence of a respective tag domain/adaptor for a particular purpose (*e.g.*, a sequencing tag domain or an amplification tag domain, and optionally an address tag domain for next-generation sequencing or amplification). In some embodiments, the 5' portion of each PCR primer is different from that of the other primer, and as such to introduce different sequences to the two ends of the PCR product. In some embodiments, the 5' portion of the first PCR primer or the 5' portion of the second PCR primer, or the 5' portions of both the first and the second PCR primers contain first or second sequencing tags/adaptors, respectively, for generation of templates for next-generation sequencing for a particular sequencing platform (*e.g.*, sequencing tags for an Illumina Nextera sequencing platform). In some embodiments, the 5' portion of the first PCR primer or the 5' portion of the second PCR primer additionally contains an address tag domain/adaptor or another tag domain/adaptor for a particular purpose.

[0206] A wide variety of enzymes and kits are available for performing the amplification reaction by PCR as known by those skilled in the art. For example, in some embodiments, the PCR amplification is performed using either the FAILSAFE™ PCR System or the MASTERAMP™ Extra-Long PCR System from EPICENTRE Biotechnologies, Madison, WI, as described by the manufacturer. However, the present disclosure is not limited to the use of those products or conditions for the amplification reaction and any suitable thermostable DNA polymerase and reaction mixture that permits amplification of the sequence between the primer that anneals to the target sequence and the primer that anneals to the transposon can be used.

[0207] The method provide herein is not limited to the use of PCR to amplify the library of tagged nucleic acid fragments. Any suitable amplification method (*e.g.*, rolling circle amplification, riboprimer amplification (*e.g.*, U.S. Patent No. 7,413,857), ICAN, UCAN, ribospia, terminal tagging (U.S. Patent Application No. 20050153333), Eberwine-type aRNA amplification or strand-displacement amplification) that amplifies the same sequence, and generates a suitable composition and amount of amplification product for the intended purpose can be used in embodiments of the present invention. For example, some strand displacement methods that can be used are described in PCT Patent Publication Nos. WO 02/16639; WO 00/56877; and AU 00/29742; of Takara Shuzo Company, Kyoto, Japan; U.S. Patent Nos. 5,523,204; 5,536,649; 5,624,825; 5,631,147; 5,648,211; 5,733,752; 5,744,311; 5,756,702; and 5,916,779 of Becton Dickinson and Company; U.S. Patent Nos. 6,238,868; 6,309,833; and 6,326,173 of Nanogen/ Becton Dickinson Partnership; U.S. Patent Nos. 5,849,547; 5,874,260; and 6,218,151 of Bio Merieux; U.S. Patent Nos. 5,786,183; 6,087,133; and 6,214,587 of Gen-Probe, Inc.; U.S. Patent No. 6,063,604 of Wick et al; U.S. Patent No. 6,251,639 of Kurn; U.S. Patent No. 6,410,278; and PCT Publication No. WO 00/28082 of Eiken Kagaku Kabushiki Kaishi, Tokyo, Japan; U.S. Patent Nos. 5,591,609; 5,614,389; 5,773,733; 5,834,202; and 6,448,017 of Auerbach; and U.S. Patent Nos. 6,124,120; and 6,280,949 of Lizardi. In some embodiments, Cas mediated amplification provided herein is used to amplify the library of the target nucleic acid fragments.

[0208] In some embodiments, when the target nucleic acid has a universal primer sequence, *e.g.*, as generated using the methods provided above, a guide RNA, *e.g.*, a crRNA, can target to the universal primer sequence of the nucleic acid fragments in the library so that to amplify the library of nucleic acid fragments isothermally.

[0209] Figure 1 illustrates a method to amplify nucleic acid fragments generated by Nextera library preparation (available from Illumina, Inc., San Diego, CA) according to the present disclosure. In some embodiments, the Cas proteins provided herein require a PAM sequence in order to recognize a target. For example, Cas9 protein requires a sequence of the motif NGG residing adjacent to the target sequence. A sequence complementary to the PAM sequence can be incorporated into the flowcell primers (*e.g.*, P5 and P7) that are added to library inserts to result in PAM modified P5 and P7 primers capable of targeting PAM-containing sequences. The sequences of standard universal primers P5 and P7 are shown in Figure 1A (SEQ ID No.1 and SEQ ID No. 2). The sequences of PAM-modified primers P5 and P7 are also shown in Figure 1A (SEQ ID No.3 and SEQ ID No. 4). Thus, in some embodiments, PAM-modified primers P5 and P7 are added to nucleic acid fragments in the library. In a specific embodiment, PAM-modified primers P5 and P7 are added to nucleic acid fragments using a limited-cycle PCR.

[0210] As shown in Figure 1B-C, DNA fragments (*e.g.*, Nextera amplicons in solution) contain P5 and P7 primer sequences (SEQ ID No. 1 and SEQ ID No.2) at the ends of the amplicon. A CRISPR-Cas9 system containing a guide RNA of SEQ ID No. 7 targeting a region of P5 primer sequence (see Figure 1B), and a CRISPR-Cas9 system containing a guide RNA of SEQ ID No. 8 targeting a region of P7 primer sequence (see Figure 1C) are added. The CRISPR-Cas9 systems open some regions of double-stranded DNA to create R-loop structures near the two ends of the DNA fragment by binding crRNA to the primer sequences. Then truncated PAM-modified P5 and P7 primers (SEQ ID No. 5 and SEQ ID No.6) can be used to amplify the DNA fragment.

[0211] Thus, in some embodiments, the first strand of the target double-stranded nucleic acid contains a universal sequence, and wherein the crRNA or the derivative thereof contains a sequence complementary to a region of the universal sequence. In some embodiments, the primer contains a sequence of a region of the universal sequence.

[0212] In some embodiments, the universal primer has a sequence of SEQ ID No. 3. In some embodiments, the crRNA contains a sequence of SEQ ID No.7. In some embodiments, the primer contains a sequence of SEQ ID No. 5.

[0213] In some embodiments, the universal primer sequence has a sequence of SEQ ID No. 4. In some embodiments, the crRNA contains a sequence of SEQ ID No.8. In some embodiments, the primer contains a sequence of SEQ ID No. 6.

[0214] In some embodiments, the first strand of the target double-stranded nucleic acid

contains a first universal sequence, and wherein the crRNA or the derivative thereof of the first system contains a sequence complementary to a region of the first universal sequence, and the second strand of the target double-stranded nucleic acid contains a second universal sequence, and wherein the crRNA or the derivative thereof of the second system contains a sequence complementary to a region of the second universal sequence. In some embodiments, the first primer contains a sequence of a region of the first universal sequence, and the second primer contains a sequence of a region of the second universal sequence. In a specific embodiment, the first universal sequence has a sequence of SEQ ID No. 3, the crRNA or the derivative thereof of the first system contains a sequence of SEQ ID No. 7, and the first primer contains a sequence of SEQ ID NO.5, and the second universal sequence has a sequence of SEQ ID No. 4, the crRNA or derivative thereof of the second system contains a sequence of SEQ ID No. 8, and the second primer contains a sequence of SEQ ID NO. 6.

[0215] The methods provided herein can be used in isothermal amplification for sequencing, *e.g.*, in a cluster amplification developed by Illumina, Inc. (San Diego, CA). In some embodiments, the target nucleic acid of the present methods can be immobilized on a surface for amplification. For example, in some embodiments, immobilized nucleic acid fragments are amplified using cluster amplification methodologies as exemplified by the disclosures of US Patent Nos. 7,985,565 and 7,115,400. The incorporated materials of US Patent Nos. 7,985,565 and 7,115,400 describe methods of solid-phase nucleic acid amplification which allow amplification products to be immobilized on a solid support in order to form arrays comprised of clusters or "colonies" of immobilized nucleic acid molecules. Each cluster or colony on such an array is formed from a plurality of identical immobilized polynucleotide strands and a plurality of identical immobilized complementary polynucleotide strands. The arrays so-formed are generally referred to herein as "clustered arrays". The products of solid-phase amplification reactions such as those described in US Patent Nos. 7,985,565 and 7,115,400 are so-called "bridged" structures formed by annealing of pairs of immobilized polynucleotide strands and immobilized complementary strands, both strands being immobilized on the solid support at the 5' end, preferably via a covalent attachment. Cluster amplification methodologies are examples of methods wherein an immobilized nucleic acid template is used to produce immobilized amplicons. Other suitable methodologies can also be used to produce immobilized amplicons from immobilized nucleic acid fragments produced according to the methods provided herein. For example one or more clusters or colonies can be formed via solid-phase PCR whether one or both primers of each pair of amplification primers are immobilized.

[0216] As used herein, the terms "solid surface," "solid support" and other grammatical equivalents herein refer to any material that is appropriate for or can be modified to be appropriate for the attachment of a polynucleotide. Possible substrates include, but are not limited to, glass and modified or functionalized glass, plastics (including acrylics, polystyrene and copolymers of styrene and other materials, polypropylene, polyethylene, polybutylene, polyurethanes, Teflon™, *etc.*), polysaccharides, nylon or nitrocellulose, ceramics, resins, silica or silica-based materials including silicon and modified silicon, carbon, metals, inorganic glasses, plastics, optical fiber bundles, and a variety of other polymers. In some embodiments, solid supports and solid surfaces are located within a flow cell apparatus. In some

embodiments, the solid support comprises a patterned surface suitable for immobilization of molecules in an ordered pattern. A "patterned surface" refers to an arrangement of different regions in or on an exposed layer of a solid support. In some embodiments, the solid support comprises an array of wells or depressions in a surface. The composition and geometry of the solid support can vary with its use. In some embodiments, the solid support is a planar structure such as a slide, chip, microchip and/or array. As such, the surface of a substrate can be in the form of a planar layer. In some embodiments, the solid support comprises one or more surfaces of a flowcell. The term "flowcell" as used herein refers to a chamber comprising a solid surface across which one or more fluid reagents can be flowed. Examples of flowcells and related fluidic systems and detection platforms that can be readily used in the methods of the present disclosure are described, for example, in Bentley et al., Nature 456:53-59 (2008), WO 04/018497; US 7,057,026; WO 91/06678; WO 07/123744; US 7,329,492; US 7,211,414; US 7,315,019; US 7,405,281, and US 2008/0108082. In some embodiments, the solid support or its surface is non-planar, such as the inner or outer surface of a tube or vessel. In some embodiments, the solid support comprises microspheres or beads. "Microspheres," "beads," "particles," or grammatical equivalents herein are intended to mean small discrete particles made of various material including, but are not limited to, plastics, ceramics, glass, and polystyrene. In certain embodiments, the microspheres are magnetic microspheres or beads. Alternatively or additionally, the beads may be porous. The bead sizes range from nanometers, *e.g.* 100 nm, to millimeters, *e.g.* 1 mm.

[0217] In other embodiments, the immobilized nucleic acid fragments are amplified in solution. For example, in some embodiments, the immobilized nucleic acid fragments are cleaved or otherwise liberated from the solid support and amplification primers are then hybridized in solution to the liberated molecules. In other embodiments, amplification primers are hybridized to the immobilized nucleic acid fragments for one or more initial amplification steps, followed by subsequent amplification steps in solution. Thus, in some embodiments an immobilized nucleic acid template can be used to produce solution-phase amplicons. It will be appreciated that any of the amplification methodologies described herein can be utilized with universal or target-specific primers to amplify immobilized nucleic acid fragments.

[0218] The nucleic acid amplified according to the method provided herein can be sequenced according to any suitable sequencing methodology, such as direct sequencing, including sequencing by synthesis, sequencing by ligation, sequencing by hybridization, nanopore sequencing and the like. In some embodiments, the immobilized DNA fragments are sequenced on a solid support. In some embodiments, the solid support for sequencing is the same solid support upon which the amplification occurs.

[0219] In some embodiments, the sequencing methodology used in the method provided herein is sequencing-by-synthesis (SBS). In SBS, extension of a nucleic acid primer along a nucleic acid template (*e.g.* a target nucleic acid or amplicon thereof) is monitored to determine the sequence of nucleotides in the template. The underlying chemical process can be polymerization (*e.g.* as catalyzed by a polymerase enzyme). In a particular polymerase-based SBS embodiment, fluorescently labeled nucleotides are added to a primer (thereby extending

the primer) in a template dependent fashion such that detection of the order and type of nucleotides added to the primer can be used to determine the sequence of the template.

[0220] Other sequencing procedures that use cyclic reactions can be used, such as pyrosequencing. Pyrosequencing detects the release of inorganic pyrophosphate (PPi) as particular nucleotides are incorporated into a nascent nucleic acid strand (Ronaghi, et al., 1996, Analytical Biochemistry 242(1), 84-9; Ronaghi, 2001, Genome Res. 11(1), 3-11; Ronaghi et al., 1998, Science 281(5375), 363; US 6,210,891; US 6,258,568 and US. 6,274,320). In pyrosequencing, released PPi can be detected by being immediately converted to adenosine triphosphate (ATP) by ATP sulfurylase, and the level of ATP generated can be detected via luciferase-produced photons. Thus, the sequencing reaction can be monitored via a luminescence detection system. Excitation radiation sources used for fluorescence based detection systems are not necessary for pyrosequencing procedures. Useful fluidic systems, detectors and procedures that can be adapted for application of pyrosequencing to amplicons produced according to the present disclosure are described, for example, in WIPO Pat. App. Ser. No. PCT/US11/57111, US 2005/0191698 A1, US 7,595,883, and US 7,244,559.

[0221] Some embodiments can utilize methods involving the real-time monitoring of DNA polymerase activity. For example, nucleotide incorporations can be detected through fluorescence resonance energy transfer (FRET) interactions between a fluorophore-bearing polymerase and γ-phosphate-labeled nucleotides, or with zeromode waveguides (ZMWs). Techniques and reagents for FRET-based sequencing are described, for example, in Levene et al., 2003, Science 299, 682-686; Lundquist et al., 2008, Opt. Lett. 33, 1026-1028; Korlach et al., 2008, Proc. Natl. Acad. Sci. USA 105, 1176-1181.

[0222] Some SBS embodiments include detection of a proton released upon incorporation of a nucleotide into an extension product. For example, sequencing based on detection of released protons can use an electrical detector and associated techniques that are commercially available from Ion Torrent (Guilford, CT, a Life Technologies subsidiary) or sequencing methods and systems described in US 2009/0026082 A1; US 2009/0127589 A1; US 2010/0137143 A1; or US 2010/0282617 A1. Methods set forth herein for amplifying target nucleic acids using kinetic exclusion can be readily applied to substrates used for detecting protons. More specifically, methods set forth herein can be used to produce clonal populations of amplicons that are used to detect protons.

[0223] Another useful sequencing technique is nanopore sequencing (*see*, for example, Deamer et al., 2000, Trends Biotechnol., 18, 147-151; Deamer et al., 2002, Acc. Chem. Res. 35:817-825; Li et al., 2003, Nat. Mater. 2:611-615),). In some nanopore embodiments, the target nucleic acid or individual nucleotides removed from a target nucleic acid pass through a nanopore. As the nucleic acid or nucleotide passes through the nanopore, each nucleotide type can be identified by measuring fluctuations in the electrical conductance of the pore. (U.S. Patent No. 7,001,792; Soni et al., 2007, Clin. Chem., 53, 1996-200; Healy, 2007, Nanomed. 2, 459-481; Cockroft et al., 2008, J. Am. Chem. Soc., 130, 818-820,). In other nanopore sequencing embodiments, the DNA fragment to be sequenced creates a unique nanopore

current signature using gamma phosphate modified nucleotides.

METHODS FOR ENRICHING POLYNUCLEOTIDES

[0224] The present disclosure relates, in part, to utilizing CRISPR-Cas or Argonaute systems and derivatives thereof for target-specific enrichment.

[0225] Current target-specific enrichment protocols require that single-stranded nucleic acid be made prior to the target specific hybridization with probes. Among various advantages provided by the present disclosure, the present disclosure provides enrichment methods that can skip this step of generating single-stranded nucleic acid in the first place, and enable direct targeting to double-stranded nucleic acid, *e.g.*, double-stranded DNA (dsDNA). Methods targeting directly to double-stranded DNA (either partly or completely double-stranded) have unique advantages over single-stranded enrichment strategies. For example, non-specific hybridization of single-stranded genomic DNA to targeted regions reduces specificity and often requires extensive stringency washing or other time-consuming steps; and single-stranded enrichment schemes often utilizes Cot-1 or other blocking DNA to reduce non-specific hybridization. These additives are not required from double-stranded DNA enrichment schemes, reducing both cost and number of required reagents. In addition, it is easier to make sequencing libraries from double-stranded DNA than from single-stranded DNA. As such, enrichment of double-stranded DNA allows library preparation (*e.g.*, tagmentation) to occur after enrichment. For another example, since specificity (tree-like structures and non-specific hybridization is less of an issue with double-stranded DNA enrichment, potentially larger DNA fragments can be better specifically enriched compared to single-stranded DNA enrichment schemes. This is a particularly important advantage if one considers targeted sequencing in the context of haplotyping and assembly. Also, since longer DNA fragments can potentially be enriched, we have greater flexibility to where the target probe is designed. For example, we can avoid high polymorphic regions but still capture these regions. Also, fewer probes need to be used to capture large regions, reducing both capture probe cost and design.

[0226] In addition, the current protocols of target specific hybridization have slow kinetics and usually require high temperature. The present disclosure provides enzyme-driven sequence targeting methods that offer faster kinetics and easier workflow for enrichment. Because the hybridization to the target nucleic acid is enzyme driven in the present methods, the process can take place isothermally. In some embodiments, the method herein provides isothermal targeting of DNA at 20-37°C. Furthermore, the guide nucleic acid in the system herein provides for sequence specificity as well as flexible programming that enables multiplex targeted enrichment (*e.g.*, targeting multiple targeted regions with more probes made in various ways including IVT from oligo pool). The present disclosure also provides methods for enriching and/or detecting polynucleotide variants with higher sensitivity and specificity. Furthermore, the present disclosure also provides methods for targeted sequencing using CRISPR-Cas or Argonaute systems.

[0227] In one aspect, the present disclosure provides a method for enriching a target nucleic acid using a nuclease system derived from a CRISPR-Cas or Argonaute system. The present disclosure is based, in part, on the capability of CRISPR-Cas or Argonaute system to specifically bind with a target nucleic acid. Such target specific binding by the CRISPR-Cas or Argonaute system provides methods for efficiently enriching target nucleic acid, *e.g.*, by pulling down an element of CRISPR-Cas or Argonaute system that is associated with the target nucleic acid. CRISPR-Cas or Argonaute mediated nucleic acid enrichment bypasses traditionally required step of generating single-stranded nucleic acid prior to target specific binding, and enables directly targeting double-stranded nucleic acid, *e.g.*, double-stranded DNA (dsDNA). In addition, CRISPR-Cas or Argonaute mediated nucleic acid binding is enzyme-driven, and thus it can offer faster kinetics and easier workflows for enrichment with lower temperature and/or isothermal reaction conditions.

[0228] In some embodiments, the present disclosure provides a method for enriching a target nucleic acid including: providing an endonuclease system having a clustered regularly interspaced short palindromic repeats (CRISPR) RNA (crRNA) or a derivative thereof, and a CRISPR-associated (Cas) protein or a variant thereof, wherein the crRNA or the derivative thereof contains a target-specific nucleotide region complementary to a region of the target nucleic acid; contacting the target nucleic acid with the endonuclease system to form a complex, and separating the complex and thereby enriching for the target nucleic acid.

[0229] In other embodiments, provided herein is a method for enriching a target nucleic acid including: providing an endonuclease system having: a 5' phosphorylated single-stranded nucleic acid or a derivative thereof, and an Argonaute protein or a variant thereof, wherein the 5' phosphorylated single-stranded nucleic acid or the derivative thereof contains a target-specific nucleotide region complementary to a region of the target nucleic acid; contacting the target nucleic acid with the endonuclease system to form a complex, and separating the complex and thereby enriching for the target nucleic acid.

[0230] In some embodiments, the method provided herein further includes separating the target nucleic acid from the complex. In one embodiment, the CRISPR-Cas system or Argonaute system can be bound to a surface, *e.g.*, in plate once it has found the targeted region. This can prevent dissociation of the complex pre-maturely, and thus improve efficiency of capture. In some embodiments, the method provided herein further includes amplifying the target nucleic acid sequence.

[0231] In some embodiments, the target nucleic acid provided herein is a double-stranded DNA (dsDNA).

[0232] In some embodiments, the endonuclease system provided herein is derived from a CRISPR-Cas system. In some embodiments, the endonuclease system provided herein is a Type I CRISPR-Cas system or a derivative thereof. In some embodiments, the endonuclease system provided herein is a Type II CRISPR-Cas system. In some embodiments, the endonuclease system provided herein is a Type III CRISPR-Cas system or a derivative thereof.

The CRISPR-Cas systems provided herein include engineered and/or programmed nuclease systems derived from naturally occuring CRISPR-Cas systems. CRISPR-Cas systems may include contain engineered and/or mutated Cas proteins. CRISPR-Cas systems may also contain engineered and/or programmed guide RNA.

[0233] The crRNA or the derivative thereof contains a target specific nucleotide region complementary or substantially complementary to a region of the target nucleic acid. In some embodiments, the crRNA or the derivative thereof contains a user-selectable RNA sequence that permits specific targeting of the enzyme to a complementary double-stranded DNA. In some embodiments, the user-selectable RNA sequence contains 20-50 nucleotides complementary or substantially complementary to a region of the target DNA sequence. In some embodiments, the target specific nucleotide region of the crRNA has 100% base pair matching with the region of the target nucleic acid. In some embodiments, the target specific nucleotide region of the crRNA has 90%-100%, 80%-100%, or 70%-100% base pair matching with the region of the target nucleic acid. In some embodiments, there is one base pair mismatch between the target specific nucleotide region of the crRNA and the region of the target nucleic acid. In some embodiments, there are two base pair mismatches between the target specific nucleotide region of the crRNA and the region of the target nucleic acid. In some embodiments, there are three base pair mismatches between the target specific nucleotide region of the crRNA and the region of the target nucleic acid. In some embodiments, there are four base pair mismatches between the target specific nucleotide region of the crRNA and the region of the target nucleic acid. In some embodiments, there are five base pair mismatches between the target specific nucleotide region of the crRNA and the region of the target nucleic acid.

[0234] In some embodiments, the endonuclease system provided herein further includes a trans-activating crRNA (tracrRNA) or a derivative thereof. In some embodiments, the crRNA or the derivative thereof provided herein is a polynucleotide having a crRNA polynucleotide fused to a tracrRNA polynucleotide. A chimeric single-guided RNA (sgRNA) is described in Jinek et al., 2012, Science 337, 816-821, which is incorporated herein in its entirety. In one embodiment, the Cas protein or the variant thereof provided herein can be directed by a chimeric sgRNA to any genomic locus followed by a 5'-NGG protospacer-adjacent motif (PAM).

[0235] In some embodiments, the Cas protein or the variant thereof is a Cas9 protein or a variant thereof. Isolated Cas9-crRNA complex from the *S. thermophilus* CRISPR-Cas system as well as complex assembled *in vitro* from separate components demonstrate that it binds to both synthetic oligodeoxynucleotide and plasmid DNA bearing a nucleotide sequence complementary to the crRNA. It has been shown that Cas9 has two nuclease domains-RuvC- and HNH- active sites/nuclease domains, and these two nuclease domains are responsible for the cleavage of opposite DNA strands. In some embodiments, the Cas9 protein is derived from Cas9 protein of *S. thermophilus* CRISPR-Cas system. In some embodiments, the Cas9 protein is a multi-domain protein having about 1,409 amino acids residues.

[0236] In some embodiments, the Cas9 protein or the variant thereof retains the two nuclease

domains and is able to cleave opposite DNA strands and produce a double-stranded DNA break. The present method is partially based on a surprising discovery that wild-type Cas9 protein that retains the two nuclease domains can remain at the binding site following DNA cleavage with sufficient strength and length, so that to enable pulling down the DNA-endonuclease system complex through the endonuclease system. As illustrated in Figure 2A-2B, the CRISPR-Cas system containing a wild type Cas9 protein is added to a solution containing a target Braf sequence. The system is labeled with biotinylated dUTP, and streptavidin beads are added to pull down the system with its associated DNA fragments. As shown in the right panel of Figure 2A-2B, the cleaved DNA fragments are detected from the bead elution, indicating the association between the enzyme system and the DNA after the cleavage.

[0237] In other embodiments, the Cas9 protein or the variant thereof is a Cas9 nickase and is able to produce a single-stranded nucleic acid nick, *e.g.*, a single-stranded DNA nick. A nickase variant of Cas9 protein stays with the target nucleic acid after creating a nick, and thus it can be used for target specific enrichment. In some embodiment, only RuvC- nuclease domain is mutated and inactivated. In some embodiments, only HNH- nuclease domain is mutated and inactivated. In some embodiments, the Cas9 protein contains one inactivated nuclease domain having a mutation in the domain that cleaves a target nucleic acid strand that is complementary to the crRNA. In one embodiment, the mutation is D10A. In some embodiments, the Cas9 protein contains one inactivated nuclease domain having a mutation in the domain that cleaves a target nucleic acid strand that is non-complementary to the crRNA. In one embodiment, the mutation is mutation is H840A.

[0238] In other embodiments, the systems provided herein are derived from Argonaute systems. In some embodiments, the system is an Ago subfamily system or a derivative thereof. In other embodiments, the system is a Piwi subfamily system or a derivative thereof. In some specific embodiments, the Argonaute protein is Natronobacterium gregorgi Argonaute (Ng Argonaute) or a variant thereof.

[0239] In some embodiments, the 5' phosphorylated single-stranded nucleic acid or the derivative thereof contains a target specific nucleotide region complementary or substantially complementary to a region of the target nucleic acid. In some embodiments, the 5' phosphorylated single-stranded nucleic acid or the derivative thereof contains a user-selectable DNA sequence that permits specific targeting of the enzyme to a complementary double-stranded DNA. In some embodiments, the user-selectable DNA sequence contains 20-50 nucleotides complementary or substantially complementary to a region of the target DNA sequence. In some embodiments, the 5' phosphorylated single-stranded nucleic acid is user-selectable single-stranded DNA sequence of less than 100 nucleotides. In some embodiments, the 5' phosphorylated single-stranded nucleic acid is a single-stranded DNA of less than 50 nucleotides. In some embodiments, the 5' phosphorylated single-stranded nucleic acid is a single-stranded DNA of less than 25 nucleotides. Example lengths of the single-stranded DNA include 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, and 50 nucleotides. In some

embodiments, the user-selectable DNA sequence contains less than 20 nucleotides complementary or substantially complementary to a region of the target DNA sequence. Exemplary user-selectable DNA sequence contains 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, or 19 nucelotides complementary or substantially complementary to a region of the target DNA sequence.

[0240] In some embodiments, the target specific nucleotide region of the 5' phosphorylated single-stranded nucleic acid has 100% base pair matching with the region of the target nucleic acid. In some embodiments, the target specific nucleotide region of the 5' phosphorylated single-stranded nucleic acid has 90%-100%, 80%-100%, or 70%-100% base pair matching with the region of the target nucleic acid. In some embodiments, there is one base pair mismatch between the target specific nucleotide region of the crRNA and the region of the target nucleic acid. In some embodiments, there are two base pair mismatches between the target specific nucleotide region of the 5' phosphorylated single-stranded nucleic acid and the region of the target nucleic acid. In some embodiments, there are three base pair mismatches between the target specific nucleotide region of the 5' phosphorylated single-stranded nucleic acid and the region of the target nucleic acid. In some embodiments, there are four base pair mismatches between the target specific nucleotide region of the 5' phosphorylated single-stranded nucleic acid and the region of the target nucleic acid. In some embodiments, there are five base pair mismatches between the target specific nucleotide region of the 5' phosphorylated single-stranded nucleic acid and the region of the target nucleic acid.

[0241] The Argonaute systems provided herein include engineered and/or programmed nuclease systems derived from naturally occurring Argonaute systems. Argonaute systems may include contain engineered and/or mutated Argonuate proteins. Argonaute systems may also contain engineered and/or programmed guide nucleic acid. In some embodiments, the Argonaute protein or a variant thereof is engineered to remove the endonuclease activity so that the Argonaute system does not generate double-stranded or single-stranded DNA breaks. In other embodiments, the Argonaute protein or a variant thereof is engineered to be able to generate single-stranded DNA breaks.

[0242] In some embodiments, the present method can be used to enrich a target nucleic acid fragment in a library of nucleic acid fragments, *e.g.*, prepared using Illumina's Nextera library preparation.

[0243] In some embodiments, the Cas9 protein or the variant thereof or Argonaute protein or the variant thereof is a nuclease-null variant of the Cas9 or Argonaute protein. For example, in some embodiments, both RuvC- and HNH- active sites/nuclease domains of a Cas9 protein are mutated. A nuclease-null variant of the Cas9 protein or Argonaute protein binds to double-stranded DNA, but not cleave the DNA, and thus it can be used for target specific DNA enrichment too.

[0244] A target nucleic acid can be separated by pulling down its associated CRISPR-Cas or Argonaute system or elements thereof. In some embodiments, the endonuclease system is

labeled, and the enzyme-nucleic acid complex is pulled down through the label. In some embodiments, the guide nucleic acid is labeled. In one embodiment, the guide nucleic acid, *e.g.*, crRNA or 5' phosphorylated single-stranded nucleic acid, is labeled with biotin. In other embodiments, the Cas protein or the Argonaute, or the variant thereof, is labeled with a capture tag. The protein capture tag includes, but not limited to, GST, Myc, hemagglutinin (HA), Green fluorescent protein (GFP), flag, His tag, TAP tag, and Fc tag. Other protein capture tags, *e.g.*, affinity tags, recognized in the art can also be used in the present methods. Those skilled in the art will recognize that a protocol chosen for the purification step will be specific to the tag used. In some embodiments, antibodies or fragments thereof, *e.g.*, anti-Cas9 antibodies or anti-Argonaute antibodies, can also be used to separate the complex.

[0245] In another aspect, binding of a guide nucleic acid to a region of a target double-stranded nucleic acid disrupts the interaction between the two strands of the target nucleic acid, and thereby creates a loop structure exposing the strand non-complementary to the guide nucleic acid. This exposed strand can be subjected to hybridization with another nucleotide probe as provided herein. One advantage provided by the method herein is double specificity for the enrichment-one from the guide nucleic acid and the other from the probe.

[0246] In one embodiment, the present disclosure provides a method for enriching a target double-stranded nucleic acid including providing an endonuclease system having a clustered regularly interspaced short palindromic repeats (CRISPR) RNA (crRNA) or a derivative thereof, and a CRISPR-associated (Cas) protein or a variant thereof, wherein the crRNA or the derivative thereof contains a target-specific nucleotide region complementary to a region of a first strand of the target double-stranded nucleic acid; contacting the target double-stranded nucleic acid with the endonuclease system to form a first complex; hybridizing a labeled nucleic acid to a second strand of the target double-stranded nucleic acid to form a second complex, the second strand of the target double-stranded nucleic acid being non-complementary to the crRNA or the derivative thereof, and separating the second complex and thereby enriching for the target nucleic acid.

[0247] In another embodiment, the present disclosure provides a method for enriching a target double-stranded nucleic acid including providing an endonuclease system having: a 5' phosphorylated single-stranded nucleic acid or a derivative thereof, and an Argonaute protein or a variant thereof or a variant thereof, wherein the 5' phosphorylated single-stranded nucleic acid or the derivative thereof contains a target-specific nucleotide region complementary to a region of a first strand of the target double-stranded nucleic acid; contacting the target double-stranded nucleic acid with the endonuclease system to form a first complex; hybridizing a labeled nucleic acid to a second strand of the target double-stranded nucleic acid to form a second complex, the second strand of the target double-stranded nucleic acid being non-complementary to the 5' phosphorylated single-stranded nucleic acid or the derivative thereof, and separating the second complex and thereby enriching for the target nucleic acid.

[0248] In some embodiments, the method of the present disclosure further includes separating the target double-stranded DNA sequence from the second complex. In some embodiments,

the method the present application further includes amplifying the targeted double-stranded DNA sequence.

[0249] In some embodiments, the target nucleic acid provided herein is a double-stranded DNA (dsDNA).

[0250] In some embodiments, the endonuclease system provided herein is a Type I CRISPR-Cas system or a derivative thereof. In some embodiments, the endonuclease system provided herein is a Type II CRISPR-Cas system. In some embodiments, the endonuclease system provided herein is a Type III CRISPR-Cas system or a derivative thereof. The CRISPR-Cas systems provided herein include engineered and/or programmed nuclease systems derived from naturally accruing CRISPR-Cas systems. CRISPR-Cas systems may include contain engineered and/or mutated Cas proteins. CRISPR-Cas systems may also contain engineered and/or programmed guide RNA.

[0251] In some embodiments, the crRNA or the derivative thereof contains a user-selectable RNA sequence that permits specific targeting of the enzyme to a complementary double-stranded DNA. In some embodiment, the user-selectable RNA sequence contains 20-50 nucleotides complementary or substantially complementary to a region of the target DNA sequence. In some embodiments, the target specific nucleotide region of the crRNA has 100% base pair matching with the region of the target nucleic acid. In some embodiments, the target specific nucleotide region of the crRNA has 90%-100%, 80%-100%, or 70%-100% base pair matching with the region of the target nucleic acid. In some embodiments, there is one base pair mismatch between the target specific nucleotide region of the crRNA and the region of the target nucleic acid. In some embodiments, there are two, three, four, or five base pair mismatches between the target specific nucleotide region of the crRNA and the region of the target nucleic acid.

[0252] In some embodiments, the endonuclease system provided herein further includes a trans-activating crRNA (tracrRNA) or a derivative thereof. In some embodiments, the crRNA or the derivative thereof provided herein is a polynucleotide having a crRNA polynucleotide fused to a tracrRNA polynucleotide. In one embodiment, the Cas protein or the variant thereof provided herein can be directed by a chimeric sgRNA to any genomic locus followed by a 5'-NGG protospacer-adjacent motif (PAM).

[0253] In some embodiments, the Cas protein or the variant thereof is a Cas9 protein or a variant thereof. In some embodiment, the Cas9 protein is derived from Cas9 protein of S. *thermophilus* CRISPR-Cas system. In some embodiment, the Cas9 protein is a multi-domain protein of about 1,409 amino acids residues.

[0254] In some embodiments, the Cas9 protein or the variant thereof retains the two nuclease domains and is able to cleave opposite DNA strands and produce a double-stranded DNA break. In other embodiments, the Cas9 protein or the variant thereof is a Cas9 nickase and is able to produce a single-stranded nucleic acid nick, *e.g.*, a single-stranded DNA nick. In some

embodiment, only RuvC- nuclease domain is mutated and inactivated. In some embodiments, only HNH- nuclease domain is mutated and inactivated. In some embodiments, the Cas9 protein contains one inactivated nuclease domain having a mutation in the domain that cleaves a target nucleic acid strand that is complementary to the crRNA. In one embodiment, the mutation is D10A. In some embodiments, the Cas9 protein contains one inactivated nuclease domain having a mutation in the domain that cleaves a target nucleic acid strand that is non-complementary to the crRNA. In one embodiment, the mutation is mutation is H840A. In yet other embodiments, the Cas9 protein or the variant thereof is a nuclease-null variant of the Cas9 protein, in which both RuvC- and HNH- active sites/nuclease domains are mutated. A nuclease-null variant of the Cas9 protein binds to double-stranded DNA, but not cleave the DNA, and thus it can be used for target specific DNA enrichment too. In some embodiments, the Cas9 protein has two inactivated nuclease domains with a first mutation in the domain that cleaves the strand complementary to the crRNA and a second mutation in the domain that cleaves the strand non-complementary to the crRNA. In some embodiments, the Cas9 protein has a first mutation D10A and a second mutation H840A.

[0255] In other embodiments, the systems provided herein are derived from Argonaute systems. In some embodiments, the system is an Ago subfamily system or a derivative thereof. In other embodiments, the system is a Piwi subfamily system or a derivative thereof. In some specific embodiments, the Argonaute protein is Natronobacterium gregorgi Argonaute (Ng Argonaute) or a variant thereof.

[0256] In some embodiments, the 5' phosphorylated single-stranded nucleic acid or the derivative thereof contains a target specific nucleotide region complementary or substantially complementary to a region of the target nucleic acid. In some embodiments, the 5' phosphorylated single-stranded nucleic acid or the derivative thereof contains a user-selectable DNA sequence that permits specific targeting of the enzyme to a complementary double-stranded DNA. In some embodiments, the user-selectable DNA sequence contains 20-50 nucleotides complementary or substantially complementary to a region of the target DNA sequence. In some embodiments, the 5' phosphorylated single-stranded nucleic acid is user-selectable single-stranded DNA sequence of less than 100 nucleotides. In some embodiments, the 5' phosphorylated single-stranded nucleic acid is a single-stranded DNA of less than 50 nucleotides. In some embodiments, the 5' phosphorylated single-stranded nucleic acid is a single-stranded DNA of less than 25 nucleotides. Example lengths of the single-stranded DNA include 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, and 50 nucleotides. In some embodiments, the user-selectable DNA sequence contains less than 20 nucleotides complementary or substantially complementary to a region of the target DNA sequence. Exemplary user-selectable DNA sequence contains 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, or 19 nucelotides complementary or substantially complementary to a region of the target DNA sequence.

[0257] In some embodiments, the target specific nucleotide region of the 5' phosphorylated single-stranded nucleic acid has 100% base pair matching with the region of the target nucleic

acid. In some embodiments, the target specific nucleotide region of the 5' phosphorylated single-stranded nucleic acid has 90%-100%, 80%-100%, or 70%-100% base pair matching with the region of the target nucleic acid. In some embodiments, there is one base pair mismatch between the target specific nucleotide region of the crRNA and the region of the target nucleic acid. In some embodiments, there are two base pair mismatches between the target specific nucleotide region of the 5' phosphorylated single-stranded nucleic acid and the region of the target nucleic acid. In some embodiments, there are three base pair mismatches between the target specific nucleotide region of the 5' phosphorylated single-stranded nucleic acid and the region of the target nucleic acid. In some embodiments, there are four base pair mismatches between the target specific nucleotide region of the 5' phosphorylated single-stranded nucleic acid and the region of the target nucleic acid. In some embodiments, there are five base pair mismatches between the target specific nucleotide region of the 5' phosphorylated single-stranded nucleic acid and the region of the target nucleic acid.

[0258] The Argonaute systems provided herein include engineered and/or programmed nuclease systems derived from naturally occurring Argonaute systems. Argonaute systems may include contain engineered and/or mutated Argonuate proteins. Argonaute systems may also contain engineered and/or programmed guide nucleic acid. In some embodiments, the Argonaute protein or a variant thereof is engineered to remove the endonuclease activity so that the Argonaute system does not generate double-stranded or single-stranded DNA breaks. In other embodiments, the Argonaute protein or a variant thereof is engineered to be able to generate single-stranded DNA breaks.

[0259] In another aspect, the target nucleic acid can be fragmented and linked to an adaptor, preparing for other procedures such as sequencing. In some embodiments, the target nucleic acid is further subjected to a transposase mediated tagmentation that results in fragmentation of the target nucleic acid and ligation of adaptors to the 5' end of both strands of double-stranded DNA fragments. Optionally, the target nucleic acid can be fragmented and adaptors can be added to the 5' and 3' ends using tagmentation or transposition as described in U.S. Publication No. 2010/0120098. Briefly, a transposition reaction is a reaction wherein one or more transposons are inserted into target nucleic acids at random sites. Essential components in a transposition reaction are a transposase and DNA oligonucleotides that exhibit the nucleotide sequences of a transposon, including the transferred transposon sequence and its complement (the non- transferred transposon end sequence) as well as other components needed to form a functional transposition or transposome complex. The DNA oligonucleotides can further include additional sequences (*e.g.*, adaptor or primer sequences) as needed or desired. Exemplary transposition complexes, suitable for use in the methods provided herein, include, but are not limited to, those formed by a hyperactive Tn5 transposase and a Tn5-type transposon end or by a MuA transposase and a Mu transposon end comprising R1 and R2 end sequences (*see*, *e.g.*, Goryshin and Reznikoff, J. Biol. Chem. 273: 7367, 1998; and Mizuuchi, Cell 35: 785, 1983; Savilahti et al., EMBO J. 14: 4893, 1995;). However, any transposition system that is capable of inserting a transposon end with sufficient efficiency to tag target nucleic acids for its intended purpose can be used in the provided methods. Other examples of known transposition systems that could be used in the provided methods include,

but are not limited to, Staphylococcus aureus Tn552, Tyl, Transposon Tn7, Tn/O and IS10, Mariner transposase, Tel, P Element, Tn3, bacterial insertion sequences, retroviruses, and retrotransposon of yeast (*see*, *e.g.*, Colegio et al., 2001, J. Bacteriol. 183: 2384-8; kirby et al., 2002, Mol. Microbiol. 43: 173-86; Devine and Boeke, 1994, Nucleic Acids Res., 22: 3765-72; International Patent Application No. WO 95/23875; Craig, 1996, Science 271 : 1512; Craig, 1996, Review in: Curr Top Microbiol Immunol. 204: 27-48; Kleckner et al., 1996, Curr Top Microbiol Immunol. 204: 49-82; Lampe et al., 1996, EMBO J. 15: 5470-9; Plasterk, 1996, Curr Top Microbiol Immunol 204: 125-43; Gloor, 2004, Methods Mol. Biol. 260: 97-114; Ichikawa and Ohtsubo, 1990, J Biol. Chem. 265: 18829-32; Ohtsubo and Sekine, 1996, Curr. Top. Microbiol. Immunol. 204: 1-26; Brown et al., 1989, Proc Natl Acad Sci USA 86: 2525-9; Boeke and Corces, 1989, Annu Rev Microbiol. 43: 403-34;). In some embodiments, the method of the present disclosure further comprises removing the transposase enzyme and adding to the ends of the adapted DNA fragments by PCR.

[0260] In some embodiments, the tagmentation is performed after the target nucleic acid is enriched.

[0261] In some embodiments, the guide nucleic acid, *e.g.*, the RNA in the CRISPR-Cas system (*e.g.*, a crRNA or a derivative thereof, a sgRNA, and a tracrRNA or a derivative thereof), or the 5' phosphorylated single-stranded nucleic acid (e.g., a gDNA), contains a transposon end, and the method of the present disclosure further includes adding a transposase. The added transposase can assemble on the transposon end and the target DNA is thereby cleaved by the transposase. In some embodiments, the transposon end is a mosaic end (ME), and the transposase is a Tn5 transposase.

[0262] In some embodiments, the nuclease system provided herein further includes a transposase, and thus transposase is part of the endonuclease system, and the method of the present disclosure further includes adding transposon end to the target DNA sequence; and tagmenting the target DNA sequence by the transposase. In some embodiments, the transposase binds to a nucleotide sequence of the endonuclease system. In some embodiments, the transposase binds to a crRNA or a derivative thereof. In some embodiments, the transposase binds to a tracrRNA or a derivative thereof. In some embodiments, the transposase binds to a sgRNA or a chimeric polynucleotide having a crRNA polynucleotide and a tracrRNA polynucleotide. In some embodiments, the transposase binds to a 5' phosphorylated single-stranded nucleic acid or a derivative thereof. In some embodiments, the transposon end is a mosaic end (ME), and the transposase is a Tn5 transposase.

[0263] In another aspect, the present disclosure provides methods for enriching and/or detecting target nucleic acid in a population of cell free DNA using CRISPR-Cas or Argonaute systems. Cell free DNA in plasma or serum holds enormous potential as a non-invasive diagnostic tool in many areas of medicine. For example, cell free fetal DNA has been studied and even optimized for testing non-compatible RhD factors, sex determination for X-linked genetic disorders, testing for single gene disorders, identification of preeclampsia, and so on. For instance, sequencing the fetal cell fraction of cell free DNA in maternal plasma is a reliable

approach for detecting copy number changes associated with fetal chromosome anueploidy. For another instance, sequencing cell free DNA isolated from cancer patients (also called circulating tumor DNA) has been used to detect mutations in key genes that have relevance for treatment decisions. The present disclosure provides methods for improving enriching and/or detecting target DNA sequences in cell free DNA.

[0264] In some embodiments, the present disclosure provides a method for enriching a target nucleic acid including obtaining a population of cell free DNA (cfDNA) from a subject's plasma or serum, the population of cell free DNA containing the target nucleic acid; providing an endonuclease system having a clustered regularly interspaced short palindromic repeats (CRISPR) RNA (crRNA) or a derivative thereof, and a CRISPR-associated (Cas) protein or a variant thereof, wherein the crRNA or the derivative thereof contains a target-specific nucleotide region complementary to a region of the target nucleic acid; contacting the target nucleic acid with the endonuclease system to form a complex, and separating the complex and thereby enriching for the target nucleic acid.

[0265] In other embodiments, the present disclosure provides a method for enriching a target nucleic acid including obtaining a population of cell free DNA (cfDNA) from a subject's plasma or serum, the population of cell free DNA containing the target nucleic acid; providing an endonuclease system having: a 5' phosphorylated single-stranded nucleic acid or a derivative thereof, and an Argonaute protein or a variant thereof, wherein the 5' phosphorylated single-stranded nucleic acid or the derivative thereof contains a target-specific nucleotide region complementary to a region of the target nucleic acid; contacting the target nucleic acid with the endonuclease system to form a complex, and separating the complex and thereby enriching for the target nucleic acid.

[0266] In some embodiments, the method provided herein further includes separating the target DNA sequence from the complex. In some embodiments, the method provided herein further includes amplifying the targeted DNA sequence. In some embodiments, the target nucleic acid provided herein is a double-stranded DNA (dsDNA).

[0267] In some embodiments, the endonuclease system provided herein is a Type I CRISPR-Cas system or a derivative thereof. In some embodiments, the endonuclease system provided herein is a Type II CRISPR-Cas system. In some embodiments, the endonuclease system provided herein is a Type III CRISPR-Cas system or a derivative thereof. The CRISPR-Cas systems provided herein include engineered and/or programmed nuclease systems derived from naturally accruing CRISPR-Cas systems. CRISPR-Cas systems may include contain engineered and/or mutated Cas proteins. CRISPR-Cas systems may also contain engineered and/or programmed guide RNA.

[0268] In some embodiments, the crRNA or the derivative thereof contains a user-selectable RNA sequence that permits specific targeting of the enzyme to a complementary double-stranded DNA. In some embodiment, the user-selectable RNA sequence contains 20-50 nucleotides complementary or substantially complementary to a region of the target DNA

sequence. In some embodiments, the target specific nucleotide region of the crRNA has 100% base pair matching with the region of the target nucleic acid. In some embodiments, there is one base pair mismatch between the target specific nucleotide region of the crRNA and the region of the target nucleic acid. In some embodiments, there are two, three, four, or five base pair mismatches between the target specific nucleotide region of the crRNA and the region of the target nucleic acid.

[0269] In some embodiments, the endonuclease system provided herein further includes a trans-activating crRNA (tracrRNA) or a derivative thereof. In some embodiments, the crRNA or the derivative thereof provided herein is a polynucleotide having a crRNA polynucleotide fused to a tracrRNA polynucleotide. In one embodiment, the Cas protein or the variant thereof provided herein can be directed by a chimeric sgRNA to any genomic locus followed by a 5'-NGG protospacer-adjacent motif (PAM).

[0270] In some embodiments, the Cas protein or the variant thereof is a Cas9 protein or a variant thereof. In some embodiment, the Cas9 protein is derived from Cas9 protein of *S. thermophilus* CRISPR-Cas system. In some embodiment, the Cas9 protein is a multi-domain protein of about 1,409 amino acids residues.

[0271] In some embodiments, the Cas9 protein or the variant thereof retains the two nuclease domains and is able to cleave opposite DNA strands and produce a double-stranded DNA break. In other embodiments, the Cas9 protein or the variant thereof is a Cas9 nickase and is able to produce a single-stranded nucleic acid nick, *e.g.*, a single-stranded DNA nick. In some embodiment, only RuvC- nuclease domain is mutated and inactivated. In some embodiments, only HNH- nuclease domain is mutated and inactivated. In some embodiments, the Cas9 protein contains one inactivated nuclease domain having a mutation in the domain that cleaves a target nucleic acid strand that is complementary to the crRNA. In one embodiment, the mutation is D10A. In some embodiments, the Cas9 protein contains one inactivated nuclease domain having a mutation in the domain that cleaves a target nucleic acid strand that is non-complementary to the crRNA. In one embodiment, the mutation is mutation is H840A. In yet other embodiments, the Cas9 protein or the variant thereof is a nuclease-null variant of the Cas9 protein, in which both RuvC- and HNH- active sites/nuclease domains are mutated. A nuclease-null variant of the Cas9 protein binds to double-stranded DNA, but not cleave the DNA, and thus it can be used for target specific DNA enrichment too. In some embodiments, the Cas9 protein has two inactivated nuclease domains with a first mutation in the domain that cleaves the strand complementary to the crRNA and a second mutation in the domain that cleaves the strand non-complementary to the crRNA. In some embodiments, the Cas9 protein has a first mutation D10A and a second mutation H840A.

[0272] In other embodiments, the systems provided herein are derived from Argonaute systems. In some embodiments, the system is an Ago subfamily system or a derivative thereof. In other embodiments, the system is a Piwi subfamily system or a derivative thereof. In some specific embodiments, the Argonaute protein is Natronobacterium gregorgi Argonaute (Ng Argonaute) or a variant thereof.

[0273] In some embodiments, the 5' phosphorylated single-stranded nucleic acid or the derivative thereof contains a target specific nucleotide region complementary or substantially complementary to a region of the target nucleic acid. In some embodiments, the 5' phosphorylated single-stranded nucleic acid or the derivative thereof contains a user-selectable DNA sequence that permits specific targeting of the enzyme to a complementary double-stranded DNA. In some embodiments, the user-selectable DNA sequence contains 20-50 nucleotides complementary or substantially complementary to a region of the target DNA sequence. In some embodiments, the 5' phosphorylated single-stranded nucleic acid is user-selectable single-stranded DNA sequence of less than 100 nucleotides. In some embodiments, the 5' phosphorylated single-stranded nucleic acid is a single-stranded DNA of less than 50 nucleotides. In some embodiments, the 5' phosphorylated single-stranded nucleic acid is a single-stranded DNA of less than 25 nucleotides. Example lengths of the single-stranded DNA include 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, and 50 nucleotides. In some embodiments, the user-selectable DNA sequence contains less than 20 nucleotides complementary or substantially complementary to a region of the target DNA sequence. Exemplary user-selectable DNA sequence contains 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, or 19 nucelotides complementary or substantially complementary to a region of the target DNA sequence.

[0274] In some embodiments, the target specific nucleotide region of the 5' phosphorylated single-stranded nucleic acid has 100% base pair matching with the region of the target nucleic acid. In some embodiments, the target specific nucleotide region of the 5' phosphorylated single-stranded nucleic acid has 90%-100%, 80%-100%, or 70%-100% base pair matching with the region of the target nucleic acid. In some embodiments, there is one base pair mismatch between the target specific nucleotide region of the crRNA and the region of the target nucleic acid. In some embodiments, there are two base pair mismatches between the target specific nucleotide region of the 5' phosphorylated single-stranded nucleic acid and the region of the target nucleic acid. In some embodiments, there are three base pair mismatches between the target specific nucleotide region of the 5' phosphorylated single-stranded nucleic acid and the region of the target nucleic acid. In some embodiments, there are four base pair mismatches between the target specific nucleotide region of the 5' phosphorylated single-stranded nucleic acid and the region of the target nucleic acid. In some embodiments, there are five base pair mismatches between the target specific nucleotide region of the 5' phosphorylated single-stranded nucleic acid and the region of the target nucleic acid.

[0275] The Argonaute systems provided herein include engineered and/or programmed nuclease systems derived from naturally occurring Argonaute systems. Argonaute systems may include contain engineered and/or mutated Argonuate proteins. Argonaute systems may also contain engineered and/or programmed guide nucleic acid. In some embodiments, the Argonaute protein or a variant thereof is engineered to remove the endonuclease activity so that the Argonaute system does not generate double-stranded or single-stranded DNA breaks. In other embodiments, the Argonaute protein or a variant thereof is engineered to be able to

generate single-stranded DNA breaks.

[0276] In some embodiments, the target DNA is in a fetal cell faction of the cell free DNA, and the cell free DNA is from maternal plasma. Protocols for extracting cell free fetal DNA are known in the art (*see. e.g.*, Li et al., 2004, Clinical Chemistry 50 (6): 1002-1011; and Li et al., 2005, The Journal of the American Medical Association 293 (7): 843-849). Many protocols for extracting the fetal DNA from the maternal plasma use the size of the fetal DNA to distinguish it from the maternal DNA. Typical steps for isolation of plasma from maternal blood include centrifugation, followed by isolation and purification of cell-free DNA (*see*, *e.g.*, Chiu et al., 2001, Clinical Chemistry 47 (9): 1607-1613). Optionally, protocol developed by Legler *et al.* can be used for extracting cell free fetal DNA (*see* Legler et al. 2007, Prenatal Diagnosis 27 (9): 824-829). Optionally, formaldehyde can be added to maternal blood samples to increase the percentage of cell free fetal DNA. It has been shown that formaldehyde can stabilize intact cells, and inhibit further release of maternal DNA (*see*, *e.g.*, Dhallan et al. 2004, The Journal of the American Medical Association 291 (9): 1114-1119).

[0277] In some embodiments, the subject is a cancer patient. A tumor itself is usually the major source of tumor DNA. However, acquiring DNA through a biopsy is invasive and risky if possible at all. Cell-free circulating tumor DNA in the bloodstream released from dying tumor cells provides another useful tool for detecting somatic mutation present in the tumors. Cell free circulating tumor DNA with mutations has been identified in many types of cancers at both early stage and advanced stage. In addition, the amount of cell free circulating DNA has been shown to increase as the cancer advances. Accordingly, cell free circulating DNA can also be used as a way of monitoring tumor progression and testing whether a patient's tumor would respond to targeted drug treatments (*see*, *e.g.*, Bettegowda et al., 2014, Sci. Transl. Med, 6(224): 24). The present disclosure provides a method for enriching and/or detecting a target DNA sequence in the cell free circulating DNA from a cancer patient. In one embodiment, the cancer patient has pancreatic, ovarian, colorectal, bladder, gastroesophageal, breast, melanoma, hepatocellular, or head and neck cancer. In some embodiments, the cancer patient has brain, renal, prostate, or thyroid cancer. In some embodiments, the cancer patient has carcinoma. In some embodiments, the cancer patient has sarcoma. In some embodiments, the cancer patient has a lymphoma or leukemia. In some embodiments, the method provided herein is used to diagnose a cancer. In some embodiments, the method provided herein is used to monitor tumor progression and/or test a tumor patient's response to targeted drug treatments.

[0278] In some embodiments, the target nucleic acid contains a single nucleotide variant (SNV). In some embodiments, the SNV contains a single nucleotide polymorphism (SNP). In some embodiments, the SNV contains a point mutation. Single nucleotide polymorphism (SNP) is a common type of genetic variation which includes polymorphism in a DNA position at which two or more alternative bases occur at appreciable frequency in the people population (usually more than or equal to 1%). Point mutations are base variations with the frequency less than 1%. Single nucleotide polymorphism (SNP) and point mutations represent the largest source of diversity in the genome of a human. These single nucleotide polymorphisms (SNP) and point

mutations can serve as biological markers for locating a disease on the human genome map because they are usually located near a gene associated with a certain disease. Thus, detection of single nucleotide polymorphisms (SNPs), point mutations, and similar mutations are of great importance to clinical activities, human health, and control of genetic disease. Detection of fetal or cancer related SNV by sequencing cell free DNA can be difficult since these variants often are present at a very low percentage of total cell free DNA (typically 0.1% and below). One advantage provided by the present disclosure is a more sensitive method for detecting and/or enriching a DNA sequence having SNV. In one embodiment, the method of the present disclosure allows detection of SNV present in a cell free DNA sample in the 0.1% to 0.01% frequency range. In one embodiment, the method provided herein enriches and/or detects SNV present in a cell free DNA sample in the 0.01% to 0.05% frequency range. In some embodiments, the method provided herein enriches and/or detects SNV present in a cell free DNA sample at about 0.01%, 0.02%, 0.03%, 0.04%, 0.05%, 0.06%, 0.07%, 0.08%, 0.09%, or 0.1% frequency.

[0279] As an alternative to direct enrichment of the target nucleic acid sequence containing SNV, the present disclosure also provides a method for enriching nucleic acid sequence containing SNV by destroying other genotypes or polynucleotides that do not contain SNV using CRISPR-Cas or Argonaute systems.

[0280] In some embodiments, the present disclosure provides a method for detecting single nucleotide variant (SNV) including obtaining a population of cell free DNA from a subject's plasma or serum; providing a first endonuclease system having a first clustered regularly interspaced short palindromic repeats (CRISPR) RNA (crRNA) or a derivative thereof, and a first CRISPR-associated (Cas) protein or a variant thereof, wherein the first crRNA or the derivative thereof contains a first target-specific nucleotide region complementary to a region of a first target nucleic acid, and wherein the first Cas protein has nuclease activity; cleaving the first target nucleic acid using the endonuclease system, and amplifying a second target nucleic acid using Polymerase Chain Reaction (PCR), wherein the second target nucleic acid contains a single nucleotide variant version of the first target nucleic acid.

[0281] In other embodiments, the present disclosure provides a method for detecting single nucleotide variant (SNV) including: obtaining a population of cell free DNA from a subject's plasma or serum; providing a first endonuclease system having: a first 5' phosphorylated single-stranded nucleic acid or a derivative thereof, and a first Argonaute protein or a variant thereof, wherein the first 5' phosphorylated single-stranded nucleic acid or the derivative thereof contains a first target-specific nucleotide region complementary to a region of a first target nucleic acid, and wherein the first Argonaute protein has nuclease activity; cleaving the first target nucleic acid using the first endonuclease system, and amplifying a second target nucleic acid using Polymerase Chain Reaction (PCR), wherein the second target nucleic acid contains a single nucleotide variant version of the first target nucleic acid.

[0282] In some embodiments, the second target nucleic acid provided herein is a double-stranded DNA (dsDNA).

[0283] In some embodiments, the first endonuclease system provided herein is a Type I CRISPR-Cas system or a derivative thereof. In some embodiments, the endonuclease system provided herein is a Type II CRISPR-Cas system. In some embodiments, the endonuclease system provided herein is a Type III CRISPR-Cas system or a derivative thereof. The CRISPR-Cas systems provided herein include engineered and/or programmed nuclease systems derived from naturally accruing CRISPR-Cas systems. CRISPR-Cas systems may include contain engineered and/or mutated Cas proteins. CRISPR-Cas systems may also contain engineered and/or programmed guide RNA.

[0284] In some embodiments, the first crRNA or the derivative thereof contains a user-selectable RNA sequence that permits specific targeting of the enzyme to a complementary double-stranded nucleic acid. In some embodiment, the user-selectable RNA sequence contains 20-50 nucleotides complementary to a region of the first target DNA sequence. In some embodiments, the first target specific nucleotide region of the crRNA has 100% base pair matching with the region of first the target nucleic acid. In some embodiments, there is one base pair mismatch between the first target specific nucleotide region of the crRNA and the region of the first target nucleic acid. In some embodiments, there are two, three, four, or five base pair mismatches between the first target specific nucleotide region of the crRNA and the region of the first target nucleic acid.

[0285] In some embodiments, the first endonuclease system provided herein further includes a trans-activating crRNA (tracrRNA) or a derivative thereof. In some embodiments, the first crRNA or the derivative thereof provided herein is a polynucleotide having a crRNA polynucleotide fused to a tracrRNA polynucleotide. In one embodiment, the first Cas protein or the variant thereof provided herein can be directed by a chimeric sgRNA to any genomic locus followed by a 5'-NGG protospacer-adjacent motif (PAM).

[0286] In some embodiments, the first Cas protein or the variant thereof is a Cas9 protein or a variant thereof. In some embodiment, the Cas9 protein is derived from Cas9 protein of *S. thermophilus* CRISPR-Cas system. In some embodiment, the Cas9 protein is a multi-domain protein of about 1,409 amino acids residues. In some embodiments, the Cas9 protein or the variant thereof retains the two nuclease domains and is able to cleave opposite DNA strands and produce a double-stranded DNA break.

[0287] In other embodiments, the systems provided herein are derived from Argonaute systems. In some embodiments, the system is an Ago subfamily system or a derivative thereof. In other embodiments, the system is a Piwi subfamily system or a derivative thereof. In some specific embodiments, the Argonaute protein is Natronobacterium gregorgi Argonaute (Ng Argonaute) or a variant thereof.

[0288] In some embodiments, the 5' phosphorylated single-stranded nucleic acid or the derivative thereof contains a target specific nucleotide region complementary or substantially complementary to a region of the target nucleic acid. In some embodiments, the 5'

phosphorylated single-stranded nucleic acid or the derivative thereof contains a user-selectable DNA sequence that permits specific targeting of the enzyme to a complementary double-stranded DNA. In some embodiments, the user-selectable DNA sequence contains 20-50 nucleotides complementary or substantially complementary to a region of the target DNA sequence. In some embodiments, the 5' phosphorylated single-stranded nucleic acid is user-selectable single-stranded DNA sequence of less than 100 nucleotides. In some embodiments, the 5' phosphorylated single-stranded nucleic acid is a single-stranded DNA of less than 50 nucleotides. In some embodiments, the 5' phosphorylated single-stranded nucleic acid is a single-stranded DNA of less than 25 nucleotides. Example lengths of the single-stranded DNA include 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, and 50 nucleotides. In some embodiments, the user-selectable DNA sequence contains less than 20 nucleotides complementary or substantially complementary to a region of the target DNA sequence. Exemplary user-selectable DNA sequence contains 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, or 19 nucelotides complementary or substantially complementary to a region of the target DNA sequence.

[0289] In some embodiments, the target specific nucleotide region of the 5' phosphorylated single-stranded nucleic acid has 100% base pair matching with the region of the target nucleic acid. In some embodiments, the target specific nucleotide region of the 5' phosphorylated single-stranded nucleic acid has 90%-100%, 80%-100%, or 70%-100% base pair matching with the region of the target nucleic acid. In some embodiments, there is one base pair mismatch between the target specific nucleotide region of the crRNA and the region of the target nucleic acid. In some embodiments, there are two base pair mismatches between the target specific nucleotide region of the 5' phosphorylated single-stranded nucleic acid and the region of the target nucleic acid. In some embodiments, there are three base pair mismatches between the target specific nucleotide region of the 5' phosphorylated single-stranded nucleic acid and the region of the target nucleic acid. In some embodiments, there are four base pair mismatches between the target specific nucleotide region of the 5' phosphorylated single-stranded nucleic acid and the region of the target nucleic acid. In some embodiments, there are five base pair mismatches between the target specific nucleotide region of the 5' phosphorylated single-stranded nucleic acid and the region of the target nucleic acid.

[0290] The Argonaute systems provided herein include engineered and/or programmed nuclease systems derived from naturally occurring Argonaute systems. Argonaute systems may include contain engineered and/or mutated Argonuate proteins. Argonaute systems may also contain engineered and/or programmed guide nucleic acid. In some embodiments, the Argonaute protein or a variant thereof is engineered to remove the endonuclease activity so that the Argonaute system does not generate double-stranded or single-stranded DNA breaks. In other embodiments, the Argonaute protein or a variant thereof is engineered to be able to generate single-stranded DNA breaks.

[0291] In some embodiments, the second target nucleic acid contains a single nucleotide variant (SNV). In some embodiments, the SNV contains a single nucleotide polymorphism

(SNP). In some embodiments, the SNV contains a point mutation. In one embodiment, the method of the present disclosure allows detection of SNV present in a cell free DNA sample in the 0.1% to 0.01% frequency range. In one embodiment, the method provided herein enriches and/or detects SNV present in a cell free DNA sample in the 0.01% to 0.05% frequency range. In some embodiments, the method provided herein enriches and/or detects SNV present in a cell free DNA sample at about 0.01%, 0.02%, 0.03%, 0.04%, 0.05%, 0.06%, 0.07%, 0.08%, 0.09%, or 0.1% frequency.

[0292] Alternatively, two endonuclease systems can be provided: the first endonuclease system is used to digest the nucleic acid that does not contain SNV, and the second endonuclease system is used to pull down the nucleic acid with SNV

[0293] In some embodiments, the method herein further includes providing a second endonuclease system having a second clustered regularly interspaced short palindromic repeats (CRISPR) RNA (crRNA) or a derivative thereof, and a second CRISPR-associated (Cas) protein or a variant thereof, wherein the second crRNA or the derivative thereof contains a second target-specific nucleotide region complementary to a region of the second target nucleic acid; contacting the second target nucleic acid with the second endonuclease system to form a complex, and separating the complex and thereby enriching for the second target nucleic acid.

[0294] In some embodiments, the method herein further includes providing a second endonuclease system having a second 5' phosphorylated single-stranded nucleic acid or a derivative thereof, and a second Argonaute protein or a variant thereof, wherein the second 5' phosphorylated single-stranded nucleic acid or the derivative thereof contains a second target-specific nucleotide region complementary to a region of the second target nucleic acid; contacting the second target nucleic acid with the second endonuclease system to form a complex, and separating the complex and thereby enriching for the second target nucleic acid.

[0295] In some embodiments, the method provided herein further includes separating the second target nucleic acid from the complex. In some embodiments, the second target nucleic acid provided herein is a double-stranded DNA (dsDNA).

[0296] In some embodiments, the second endonuclease system provided herein is a Type I CRISPR-Cas system or a derivative thereof. In some embodiments, the second endonuclease system provided herein is a Type II CRISPR-Cas system. In some embodiments, the second endonuclease system provided herein is a Type III CRISPR-Cas system or a derivative thereof. The CRISPR-Cas systems provided herein include engineered and/or programmed nuclease systems derived from naturally accruing CRISPR-Cas systems. CRISPR-Cas systems may include contain engineered and/or mutated Cas proteins. CRISPR-Cas systems may also contain engineered and/or programmed guide RNA.

[0297] In some embodiments, the second crRNA or the derivative thereof contains a user-selectable RNA sequence that permits specific targeting of the enzyme to a complementary

double-stranded DNA. In some embodiment, the user-selectable RNA sequence contains 20-50 nucleotides complementary or substantially complementary to a region of the target DNA sequence. In some embodiments, the target specific nucleotide region of the crRNA has 100% base pair matching with the region of the target nucleic acid. In some embodiments, there is one base pair mismatch between the target specific nucleotide region of the crRNA and the region of the target nucleic acid. In some embodiments, there are two, three, four, or five base pair mismatches between the target specific nucleotide region of the crRNA and the region of the target nucleic acid.

[0298] In some embodiments, the second endonuclease system provided herein further includes a trans-activating crRNA (tracrRNA) or a derivative thereof. In some embodiments, the crRNA or the derivative thereof provided herein is a polynucleotide comprising a crRNA polynucleotide fused to a tracrRNA polynucleotide.

[0299] In some embodiments, the second Cas protein or the variant thereof is a Cas9 protein or a variant thereof. In some embodiment, the Cas9 protein is derived from Cas9 protein of S. *thermophilus* CRISPR-Cas system. In some embodiment, the Cas9 protein is a multi-domain protein of about 1,409 amino acids residues.

[0300] In some embodiments, the Cas9 protein or the variant thereof retains the two nuclease domains and is able to cleave opposite DNA strands and produce a double-stranded DNA break. In other embodiments, the Cas9 protein or the variant thereof is a Cas9 nickase and is able to produce a single-stranded nucleic acid nick, *e.g.*, a single-stranded DNA nick. In some embodiment, only RuvC- nuclease domain is mutated and inactivated. In some embodiments, only HNH- nuclease domain is mutated and inactivated. In some embodiments, the Cas9 protein contains one inactivated nuclease domain having a mutation in the domain that cleaves a target nucleic acid strand that is complementary to the crRNA. In one embodiment, the mutation is D10A. In some embodiments, the Cas9 protein contains one inactivated nuclease domain having a mutation in the domain that cleaves a target nucleic acid strand that is non-complementary to the crRNA. In one embodiment, the mutation is mutation is H840A. In yet other embodiments, the Cas9 protein or the variant thereof is a nuclease-null variant of a Cas9 protein, in which both RuvC- and HNH- active sites/nuclease domains are mutated. In some embodiments, the Cas9 protein has two inactivated nuclease domains with a first mutation in the domain that cleaves the strand complementary to the crRNA and a second mutation in the domain that cleaves the strand non-complementary to the crRNA. In some embodiments, the Cas9 protein has a first mutation D10A and a second mutation H840A.

[0301] In other embodiments, the systems provided herein are derived from Argonaute systems. In some embodiments, the system is an Ago subfamily system or a derivative thereof. In other embodiments, the system is a Piwi subfamily system or a derivative thereof. In some specific embodiments, the Argonaute protein is Natronobacterium gregorgi Argonaute (Ng Argonaute) or a variant thereof.

[0302] In some embodiments, the 5' phosphorylated single-stranded nucleic acid or the

derivative thereof contains a target specific nucleotide region complementary or substantially complementary to a region of the target nucleic acid. In some embodiments, the 5' phosphorylated single-stranded nucleic acid or the derivative thereof contains a user-selectable DNA sequence that permits specific targeting of the enzyme to a complementary double-stranded DNA. In some embodiments, the user-selectable DNA sequence contains 20-50 nucleotides complementary or substantially complementary to a region of the target DNA sequence. In some embodiments, the 5' phosphorylated single-stranded nucleic acid is user-selectable single-stranded DNA sequence of less than 100 nucleotides. In some embodiments, the 5' phosphorylated single-stranded nucleic acid is a single-stranded DNA of less than 50 nucleotides. In some embodiments, the 5' phosphorylated single-stranded nucleic acid is a single-stranded DNA of less than 25 nucleotides. Example lengths of the single-stranded DNA include 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, and 50 nucleotides. In some embodiments, the user-selectable DNA sequence contains less than 20 nucleotides complementary or substantially complementary to a region of the target DNA sequence. Exemplary user-selectable DNA sequence contains 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, or 19 nucelotides complementary or substantially complementary to a region of the target DNA sequence.

[0303] In some embodiments, the target specific nucleotide region of the 5' phosphorylated single-stranded nucleic acid has 100% base pair matching with the region of the target nucleic acid. In some embodiments, the target specific nucleotide region of the 5' phosphorylated single-stranded nucleic acid has 90%-100%, 80%-100%, or 70%-100% base pair matching with the region of the target nucleic acid. In some embodiments, there is one base pair mismatch between the target specific nucleotide region of the crRNA and the region of the target nucleic acid. In some embodiments, there are two base pair mismatches between the target specific nucleotide region of the 5' phosphorylated single-stranded nucleic acid and the region of the target nucleic acid. In some embodiments, there are three base pair mismatches between the target specific nucleotide region of the 5' phosphorylated single-stranded nucleic acid and the region of the target nucleic acid. In some embodiments, there are four base pair mismatches between the target specific nucleotide region of the 5' phosphorylated single-stranded nucleic acid and the region of the target nucleic acid. In some embodiments, there are five base pair mismatches between the target specific nucleotide region of the 5' phosphorylated single-stranded nucleic acid and the region of the target nucleic acid.

[0304] The Argonaute systems provided herein include engineered and/or programmed nuclease systems derived from naturally occurring Argonaute systems. Argonaute systems may include contain engineered and/or mutated Argonuate proteins. Argonaute systems may also contain engineered and/or programmed guide nucleic acid. In some embodiments, the Argonaute protein or a variant thereof is engineered to remove the endonuclease activity so that the Argonaute system does not generate double-stranded or single-stranded DNA breaks. In other embodiments, the Argonaute protein or a variant thereof is engineered to be able to generate single-stranded DNA breaks.

[0305] In some embodiments, the second target nucleic acid is in a fetal cell faction of the cell free DNA, and the cell free DNA is from maternal plasma. In some embodiments, the subject is a cancer patient. In one embodiment, the cancer patient has pancreatic, ovarian, colorectal, bladder, gastroesophageal, breast, melanoma, hepatocellular, or head and neck cancer. In some embodiments, the cancer patient has brain, renal, prostate, or thyroid cancer. In some embodiments, the cancer patient has carcinoma. In some embodiments, the cancer patient has sarcoma. In some embodiments, the cancer patient has a lymphoma or leukemia. In some embodiments, the method provided herein is used to diagnose a cancer. In some embodiments, the method provided herein is used to monitor tumor progression and/or test a tumor patient's response to targeted drug treatments.

[0306] In yet another aspect, the present disclosure provides a method for labeling a target nucleic acid sequence using CRISPR-Cas or Argonaute system containing a nickase. The nickase provided herein can introduce target specific nicks to the double-stranded nucleic acid. The nicks can be further used to insert capture tags, such as biotinylated dNTP, oligo probes, or double-stranded nucleic acid adapters, for enrichment strategies of the target nucleic acid. The current methods of a single-stranded nucleic acid enrichment schemes requires generating a "tree structure" of hybridized products, and such structure usually reduces specificity. The method provided herein directly targets to double-stranded nucleic acid and thus circumvents the need of creating such a "tree structure." In addition, the method provided here enables enrichment of long nucleic acid fragments.

[0307] In some embodiments, the method provided herein includes generating one single-stranded nick, and from this nick a nick translation is performed to introduce a capture label for recovering the target nucleic acid. In other embodiments, the method provided herein includes generating two consecutive single-stranded nicks on the same strand of the target nucleic acid. The single-stranded nucleic acid product between the two nicks can be replaced with a capture label for recovering the target nucleic acid. In yet other embodiments, the method provided herein includes generating two consecutive single-stranded nicks on the opposite strands of the target nucleic acid, and thus generate a double-stranded nucleic acid break that can be linked to an adapter for enrichment.

[0308] In some embodiment, the present disclosure provides a method for labeling a target nucleic acid including providing a first nuclease system having a first clustered regularly interspaced short palindromic repeats (CRISPR) RNA (crRNA) or a derivative thereof, and a first CRISPR-associated (Cas) protein or a variant thereof, wherein the first crRNA or the derivative thereof contains a first target-specific nucleotide region complementary to a first region of the target nucleic acid, and wherein the first Cas protein contains one inactivated nuclease domain; contacting a double-stranded nucleic acid containing the target nucleic acid with the first nuclease system to generate a first single-stranded nick at the first region of the target nucleic acid, and labeling the target nucleic acid.

[0309] In other embodiments, the present disclosure provides a method for labeling a target nucleic including providing a first nuclease system having a first 5' phosphorylated single-

stranded nucleic acid or a derivative thereof, and a first Argonaute protein or a variant thereof, wherein the first 5' phosphorylated single-stranded nucleic acid or the derivative thereof contains a first target-specific nucleotide region complementary to a first region of the target nucleic acid, and wherein the first Argonaute protein is capable of generating a single-stranded nick; contacting a double-stranded nucleic acid containing the target nucleic acid with the first nuclease system to generate a first single-stranded nick at the first region of the target nucleic acid, and labeling the target nucleic acid.

[0310] In some embodiments, the method herein further includes separating the target nucleic acid through the labeling and thereby enriching the target nucleic acid. In some embodiments, the method provided herein further includes amplifying the target nucleic acid.

[0311] In some embodiments, the first nuclease system provided herein further includes a trans-activating crRNA (tracrRNA). In some embodiments, the first crRNA or the derivative thereof provided herein is a polynucleotide having a crRNA polynucleotide fused to a tracrRNA polynucleotide. In some embodiments, the first nuclease system provided herein is a Type II CRISPR-Cas system or a derivative thereof. In some embodiments, the first Cas protein or the variant thereof is a Cas9 protein or a variant thereof. In some embodiments, the Cas9 protein or the variant thereof contains one inactivated nuclease domain with a mutation in the domain that cleaves a target nucleic acid strand that is complementary to the first crRNA. In some embodiments, the mutation is D10A. In some embodiments, the first Cas9 protein or the variant thereof contains one inactivated nuclease domain with a mutation in the domain that cleaves a target nucleic acid strand that is non-complementary to the first crRNA. In some embodiments, the mutation is H840A.

[0312] In some embodiments, the method of the present disclosure further includes performing a nick translation. In some embodiments, the nick translation provided herein is performed by using a nick translation polymerase selected from a group consisting of DNA Pol 1, Bst, and Taq. Other nick translation polymerases known in the art are also included in the method provided herein. In some embodiments, the nick translation provided herein is performed in a reaction mixture containing biotinylated dNTPs. In some embodiments, the biotinylated dNTPs provided herein are biotinylated dUTPs. In some embodiments, the method of the present disclosure further includes adding magnetic streptavidin beads to enrich biotinylated target DNA.

[0313] In some embodiments, the method of present disclosure further includes providing a second nuclease system having a second crRNA or a derivative thereof, and a second Cas protein or a variant thereof, wherein the second crRNA or the derivative thereof contains a second target-specific nucleotide region complementary to a second region of the target nucleic acid, and wherein the second Cas protein contains one inactivated nuclease domain, and contacting the double-stranded nucleic acid containing the target nucleic acid with the second nuclease system to generate a second single-stranded nick at the second region of the target nucleic acid, wherein the first region of the target nucleic acid is different from the second region of the target nucleic acid.

[0314] In some embodiments, the method of present disclosure further includes providing a second nuclease system having: a second 5' phosphorylated single-stranded nucleic acid or a derivative thereof, and a second Argonaute protein or a variant thereof, wherein the second 5' phosphorylated single-stranded nucleic acid or the derivative thereof contains a second target-specific nucleotide region complementary to a second region of the target nucleic acid, and wherein the second Argonaute protein is capable of generating a single-stranded nick, and contacting the double-stranded nucleic acid containing the target nucleic acid with the second nuclease system to generate a second single-stranded nick at the second region of the target nucleic acid, wherein the first region of the target nucleic acid is different from the second region of the target nucleic acid.

[0315] In some embodiments, the first single-stranded nick and the second single-stranded nick are on the same strand of the target nucleic acid.

[0316] For example, in some embodiments, the first region of the target nucleic acid and the second region of the target nucleic acid are on the same strand of the target nucleic acid, and the first Cas9 protein and the second Cas9 protein both contain a mutation in the domain that cleaves a target nucleic acid strand that is complementary to their respective crRNAs, so that the first single-stranded nick and the second single-stranded nick are on the same strand of the target nucleic acid. In some embodiments, the first Cas protein is a first Cas9 protein containing one inactivated nuclease domain having a first mutation in the domain that cleaves a target nucleic acid strand that is complementary to the first crRNA, and the second Cas protein is a second Cas9 protein containing one inactivated nuclease domain containing a second mutation in the domain that cleaves a target nucleic acid strand that is complementary to the second crRNA. In some embodiments, the first mutation and the second mutation are both D10A.

[0317] In some embodiments, the first region of the target nucleic acid and the second region of the target nucleic acid are on the same strand of the target nucleic acid, and the first Cas9 protein and the second Cas9 protein both contain a mutation in the domain that cleaves a target nucleic acid strand that is non-complementary to their respective crRNAs, so that the first single-stranded nick and the second single-stranded nick are on the same strand of the target nucleic acid. In some embodiments, the first Cas protein is a first Cas9 protein containing one inactivated nuclease domain having a first mutation in the domain that cleaves a target nucleic acid strand that is non-complementary to the first crRNA, and the second Cas protein is a second Cas9 protein containing one inactivated nuclease domain containing a second mutation in the domain that cleaves a target nucleic acid strand that is non-complementary to the second crRNA. In some embodiments, the first mutation and the second mutation are both H840A.

[0318] In some embodiments, the first region of the target nucleic acid and the second region of the target nucleic acid are on different strands of the target nucleic acid, and the two Cas9 proteins retain different nuclease domains, so that the first single-stranded nick and the second

single-stranded nick are on the same strand of the target nucleic acid. In some embodiments, the first Cas protein is a first Cas9 protein containing one inactivated nuclease domain with a first mutation in the domain that cleaves a target nucleic acid strand that is complementary to the first crRNA, and the second Cas protein is a second Cas9 protein containing one inactivated nuclease domain with a second mutation in the domain that cleaves a target nucleic acid strand that is non-complementary to the second crRNA. In some embodiments, the first mutation is D10A, and said second mutation is H840A.

[0319] In some embodiments, the space between the first single-stranded nick and the second single-stranded nick is from 20 base pairs (bp) to 10kp. In some embodiments, the space between the first single-stranded nick and the second single-stranded nick is from 20 base pairs (bp) to 5kp. In some embodiments, the space between the first single-stranded nick and the second single-stranded nick is from 20 base pairs (bp) to 1000 bp. In some embodiments, the space between the first single-stranded nick and the second single-stranded nick is from 20 base pairs (bp) to 500 bp. In some embodiments, the space between the first single-stranded nick and the second single-stranded nick is about 20 bp, 30 bp, 40 bp, 50 bp, 60 bp, 70 bp, 80 bp, 90 bp, 100 bp, 200 bp, 300 bp, 400 bp, or 500 bp.

[0320] In some embodiments, the method of the present disclosure further includes adding a capture probe; and exchanging a single-stranded nucleic acid product between the first single-stranded nick and the second single-stranded nick with the capture probe, wherein the capture probe is able to hybridize to a nucleic acid strand complementary to the single-stranded nucleic acid product. In some embodiments, the sequence of the capture probe is 10% to 100% identical to the sequence of the single-stranded nucleic acid product. In some embodiments, the sequence of the capture probe is about 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, or 90% identical to the sequence of the single-stranded nucleic acid product. In some embodiments, the capture probe provided herein is a biotinylated probe. In some embodiments, the method of the present disclosure further includes adding magnetic streptavidin beads to enrich the target nucleic acid.

[0321] In some embodiments, the nickases in the nuclease systems provide herein generate two consecutive single-stranded nicks on the same strand of the target DNA. Two enzymes systems are added with each targeting to a different region of the target DNA sequence, and thus two consecutive single-stranded nicks are generated on the same strand. The single-stranded DNA product between the two nicks is then replaced with a capture probe, *e.g.*, a biotinylated capture probe, for an enrichment step.

[0322] In some embodiments, the capture probe contains an overhang nucleotide sequence, the overhang nucleotide sequence is substantially complementary to an oligo immobilized on a surface. Therefore, the overhang can be used to pull down the target DNA by annealing the overhang to a complementary oligo immobilized on a surface. In one embodiment, the overhang contains or is complementary to the universal Illumina® capture primers P5 (available from Illunima, Inc, San Diego, CA). The surface can be an external part or external layer of a solid support. The solid support can be a rigid solid and optionally can be

impermeable to liquids or gases. The solid support can also be a semi-rigid solid, for example, being permeable to liquids or gases. The surface can be in contact with another material such as a gas, liquid, gel, second surface of a similar or different solid support, metal, or coat. The surface, or regions thereof, can be substantially flat. The surface can have surface features such as wells, pits, channels, ridges, raised regions, pegs, posts or the like. In some embodiments, a surface or region thereof can be located in a vessel such as a well, tube, channel, cuvette, Petri plate, bottle or the like. A useful vessel is a flow-cell. Exemplary flow-cells are those that are commercially available from Illumina, Inc (San Diego, CA). Another useful vessel is a well in a multiwell plate or microtiter plate. In some embodiments, the method provided herein further includes Nextera library preparation and clustering on the surface. In some embodiments, transposition can be performed prior to flow cell capture. Various embodiments have been described in context of a commercially available solid phase platform, e.g., available from Illumina Inc. (San Diego, CA), and those skilled in the art will understand that any of the various embodiments can be performed with various other solid phase configurations well known in the art. Such configurations essentially include solid phase and capture probe.

[0323] In other embodiments, the methods provided herein can be used to introduce specific gaps in repeat regions. In one embodiment, the capture probe has a "hairpin" or is a mismatched probe with 5' and 3' regions complementary to the target DNA. As a result, each repeat unit is replaced with a unique marker (or barcode) allowing the introduction of landmarks. The landmarks can be used for assembly of repeat regions or counting the exact number of repeats.

[0324] Certain polymerases *e.g.*, Phi29, can initiate a nick translation from a gap. Thus, in yet other embodiments, the space between the first single-stranded nick and the second single-stranded nick on the same strand of the target nucleic acid is 1 bp to 20 bp. In some embodiments, the method provided herein can further comprise performing a nick translation. In some embodiments, the nick translation is performed by using a nick translation polymerase Phi29.

[0325] In some embodiments, the first single-stranded nick and the second single-stranded nick are on opposite strands of the target DNA sequence, thereby generating a first double-stranded DNA break end.

[0326] In some embodiments, the first region of the target nucleic acid and the second region of the target nucleic acid are on the same strand of the target nucleic acid; the first Cas protein is a first Cas9 protein with one inactivated nuclease domain having a first mutation in the domain that cleaves a target nucleic acid strand that is complementary to the first crRNA, and the second Cas protein is a second Cas9 protein with one inactivated nuclease domain having a second mutation in the domain that cleaves a target nucleic acid strand that is non-complementary to the second crRNA. In some embodiments, the first mutation is D10A, and the second mutation is H840A.

[0327] In some embodiments, the first region of the target nucleic acid and the second region of the target nucleic acid are on opposite strands of the target nucleic acid; first Cas protein is a first Cas9 protein containing one inactivated nuclease domain having a first mutation in the domain that cleaves a target nucleic acid strand that is complementary to the first crRNA, and the second Cas protein is a second Cas9 protein containing one inactivated nuclease domain containing a second mutation in the domain that cleaves a target nucleic acid strand that is complementary to the second crRNA. In some embodiments, both the first mutation and the second mutation are D10A.

[0328] In some embodiments, the first region of the target nucleic acid and the second region of the target nucleic acid are on opposite strands of the target nucleic acid; the first Cas protein is a first Cas9 protein containing one inactivated nuclease domain having a first mutation in the domain that cleaves a target nucleic acid strand that is non-complementary to the first crRNA, and the second Cas protein is a second Cas9 protein containing one inactivated nuclease domain containing a second mutation in the domain that cleaves a target nucleic acid strand that is non-complementary to the second crRNA. In some embodiments, the first mutation and the second mutation are both H840A.

[0329] In some embodiments, nicks are made at relatively close nucleic acid positions, and a blunt ended break can be produced. In some embodiments, nicks are made at relatively far away from each other, and a sticky ended break with 5' or 3' overhangs can be produced. In some embodiments, the method of the present disclosure further includes ligating an adaptor to the double-stranded nucleic acid break end. In some embodiments, the adaptor of the present disclosure is biotinylated. In some embodiments, the method of the present disclosure includes adding magnetic streptavidin beads to enrich the target nucleic acid.

[0330] In some embodiments, the method provided herein further includes providing a third nuclease system having a third crRNA or a derivative thereof, and a third Cas protein or a variant thereof, wherein the third crRNA or the derivative thereof contains a third target-specific nucleotide region substantially complementary to a third region of the target nucleic acid, and wherein the third Cas protein contains one inactivated nuclease domain; providing a fourth nuclease system having a fourth crRNA or a derivative thereof, and a fourth Cas protein or a variant thereof, wherein the fourth crRNA or the derivative thereof contains a fourth target-specific nucleotide region substantially complementary to a fourth region of the target nucleic acid, and wherein the fourth Cas protein contains one inactivated nuclease domain; and contacting the double-stranded nucleic acid containing the target nucleic acid with the third and fourth nuclease systems to generate a third single-stranded nick at the third region of the target nucleic acid and a fourth single-stranded nick at the fourth region of the target nucleic acid, wherein in the third single-stranded nick and the fourth single-stranded nick are on opposite strands of the target nucleic acid, thereby generating a second double-stranded nucleic acid break end, the second double-stranded nucleic acid break end being different from the first double-stranded nucleic acid break end.

[0331] In other embodiments, the method provided herein further includes providing a third

nuclease system having: a third 5' phosphorylated single-stranded nucleic acid or a derivative thereof, and a third Argonaute protein or a variant thereof, wherein the third 5' phosphorylated single-stranded nucleic acid or the derivative thereof contains a third target-specific nucleotide region substantially complementary to a third region of the target nucleic acid, and wherein the third Argonaute protein is capable of generating a single-stranded nick; providing a fourth nuclease system having: a fourth 5' phosphorylated single-stranded nucleic acid or a derivative thereof, and a fourth Argonaute protein or a variant thereof, wherein the fourth 5' phosphorylated single-stranded nucleic acid or the derivative thereof contains a fourth target-specific nucleotide region substantially complementary to a fourth region of the target nucleic acid, and wherein the fourth Argonaute protein is capable of generating a single-stranded nick; and contacting the double-stranded nucleic acid containing the target nucleic acid with the third and fourth nuclease systems to generate a third single-stranded nick at the third region of the target nucleic acid and a fourth single-stranded nick at the fourth region of the target nucleic acid, wherein in the third single-stranded nick and the fourth single-stranded nick are on opposite strands of the target nucleic acid, thereby generating a second double-stranded nucleic acid break end, the second double-stranded nucleic acid break end being different from the first double-stranded nucleic acid break end.

[0332] In some embodiments, the nucleic acid fragment between the first and second double-stranded nucleic acid break ends can contain from 10 to multiple thousands of nucleotides. In some embodiments, capture probes, such as single-stranded oligos, DNA dumbbells, and double-stranded DNA adapters can be added to label the nucleic acid fragment. In some embodiments, the method provided herein further includes ligating an adapter to the second double-stranded nucleic acid break end.

[0333] For example, two pairs of CRISPR-Cas systems or Argonaute systems are provided. Each pair of enzymes contains two nickases, *e.g.*, Cas9 nickases or Argonaute nickases, and the two nickases can generate single-stranded DNA nicks on opposite strands of DNA. As such, each pair of enzymes generates a double-stranded DNA break end, and two double-stranded DNA break ends are generated surrounding or at the two ends of the target DNA sequence. In one embodiment, the DNA fragment between the two double-stranded DNA break ends is about 10kb. In some embodiments, the DNA fragment between the two double-stranded DNA break ends is about 1kb, 2kb, 3kb, 4kb, 5kb, 6kb, 7kb, 8kb, or 9kb. The DNA fragment can be further ligated to target-specific biotinylated PCR adapters through which the target DNA can be enriched.

[0334] The enriched double-stranded nucleic acid can be further subject to sequencing. In one embodiment, the enriched DNA is tagmented to smaller fragments and introduced to sequencing adapters. In some embodiments, the method provided herein further includes dilution prior to tagmentation. In one embodiment, the enriched DNA is diluted to haploid content prior to PCR and/or tagmentation.

[0335] In some embodiments, Nextera library preparation (available from Illumina, Inc, San Diego, CA) is performed to fragment input DNA and introduce sequencing primers, and then

the fragmented DNA is contacted with the CRISPR-Cas or Argonaute system provided herein to form a complex. The complex is pulled down and the target DNA can be released from the complex, *e.g.*, using EDTA, heat, SDS, and RNase. The sequencing can then be performed.

[0336] In another aspect, the present disclosure provides a method of enriching double-stranded DNA using multiple CRISPR-Cas or Argonaute systems which retain full endonuclease activity (*i.e.*, being capable of generating double-stranded breaks), *e.g.*, multiple CRISPR-Cas systems each of which having wild-type Cas9 containing two nuclease domains.

[0337] In some embodiments, provided herein is a method for enriching a target nucleic acid including: providing a population of Cas9 proteins programmed with a set of crRNAs, wherein the set of crRNAs contains crRNAs complementary to a series of different regions of the target nucleic acid; contacting the target nucleic acid with the population of Cas9 proteins programmed with the set of crRNAs to generate a series of nucleic acid fragments, and ligating adaptors to at least one of nucleic acid fragments, wherein the Cas9 protein retains two nuclease domains.

[0338] In other embodiments, provided herein is a method for enriching a target nucleic acid including: providing a population of Argonaute proteins programmed with a set of 5' phosphorylated single-stranded nucleic acids, wherein the set of 5' phosphorylated single-stranded nucleic acids contains 5' phosphorylated single-stranded nucleic acids complementary to a series of different regions of the target nucleic acid; contacting the target nucleic acid with the population of Argonaute proteins programmed with the set of 5' phosphorylated single-stranded nucleic acids to generate a series of nucleic acid fragments, and ligating adaptors to at least one of nucleic acid fragments, wherein the Argonaute proteins are capable of generating double-stranded DNA breaks.

[0339] In some embodiments, the set of crRNAs contains crRNAs complementary to two different regions of the target nucleic acid. In some embodiments, the set of 5' phosphorylated single-stranded nucleic acids contains 5' phosphorylated single-stranded nucleic acids complementary to two different regions of the target nucleic acid.

[0340] The method provided herein can be useful for enriching a long DNA fragment. In some embodiments, the space between the two different region is longer than 10 kb.

[0341] In some embodiments, the target nucleic acid is a double-stranded DNA. In some embodiments, the target nucleic acid is a genomic DNA, a chromosomal DNA, a genome, or a partial genome.

[0342] In another aspect, the present disclosure provides a method of Cas9 or Argonaute mediated nucleic acid fragmentation and targeted sequencing. The present disclosure provides a method for fragmenting DNA in a sequence specific manner in user defined regions, and generating nucleic acid fragments for subsequent sequencing, *e.g.*, DNA fragments amendable for incorporation into Illumina's sequencing libraries.

[0343] In some embodiments, the method for sequencing a target nucleic acid provided herein includes providing a population of Cas9 proteins programmed with a set of crRNAs, wherein the set of crRNAs contains crRNAs complementary to a series of different regions across the target nucleic acid; contacting the target nucleic acid with the population of Cas9 proteins programmed with the set of crRNAs to generate a series of nucleic acid fragments, and sequencing the series of nucleic acid fragments.

[0344] In other embodiments, the method for sequencing a target nucleic acid provided herein includes providing a population of Argonaute proteins programmed with a set of 5' phosphorylated single-stranded nucleic acids, wherein the set of 5' phosphorylated single-stranded nucleic acids contains 5' phosphorylated single-stranded nucleic acids complementary to a series of different regions across the target nucleic acid; contacting the target nucleic acid with the population of Argonaute proteins programmed with the set of 5' phosphorylated single-stranded nucleic acids to generate a series of nucleic acid fragments, and sequencing the series of nucleic acid fragments.

[0345] In some embodiments, targeted fragmentation of nucleic acid can be achieved by preparing a population of Cas9 or Argonaute proteins that are programmed with guide nucleic acids targeting regions tiled across the target nucleic acid. In some embodiments, the Cas9 or Argonaute proteins provided herein can generate double-stranded nucleic acid breaks and thus a series of nucleic acid fragments. These nucleic acid fragments can be further subjected to nucleic acid sequencing workflows.

[0346] The same nucleic acid sample can be treated separately with multiple populations of Cas9 or Argonaute proteins programmed with different sets of guide nucleic acid targeting regions tiled across the target nucleic acid. The nucleic acid fragments generated by each population overlap with nucleic acid fragments generated by another population. More reliable and comprehensive sequencing data can be achieved by sequencing nucleic acid fragments with overlapping sequences.

[0347] In some embodiments, the method for sequencing a target nucleic acid provided herein includes providing a plurality of populations of Cas9 proteins, each population of Cas9 proteins being programmed with a different set of crRNAs, wherein each set of crRNAs contains crRNAs complementary to a different series of regions across the target nucleic acid; contacting the target nucleic acid with each of the plurality of populations of Cas9 proteins in a separate reaction to generate a different series of nucleic acid fragments, and sequencing the nucleic acid fragments.

[0348] In other embodiments, the method for sequencing a target nucleic acid provided herein includes providing a plurality of populations of Argonaute proteins, each population of Argonaute proteins being programmed with a different set of 5' phosphorylated single-stranded nucleic acids, wherein each set of 5' phosphorylated single-stranded nucleic acids contains 5' phosphorylated single-stranded nucleic acids complementary to a different series of regions

across the target nucleic acid, contacting the target nucleic acid with each of the plurality of populations of Argonaute proteins in a separate reaction to generate a different series of nucleic acid fragments, and sequencing the nucleic acid fragments.

[0349] In some embodiments, the plurality of populations of Cas9 or Argonaute proteins includes three populations of Cas9 or Argonaute proteins, and wherein the nucleic acid fragments generated by each of the three populations of Cas9 or Argonaute proteins contain overlapping sequences with the nucleic acid fragments generated by at least another of the three populations of Cas9 proteins. For example, a 10kb target DNA can be treated with the Cas9 proteins programmed with three sets of crRNAs targeting regions with about 500 bp intervals across the target DNA sequence. Each set of crRNAs contains about 57 crRNAs. Cas9 proteins remain non-covalently associated with the ends of cleaved DNAs, cleaved target DNA can be released by treatment of the sample with protease or detergent. Cleavage products are then pooled and converted to sequencing libraries, *e.g.*, using Illumina's TruSeq Nano workflow. The cleavage can be carried out using a different set of crRNAs in a separate reaction. For instance, cleavage can be carried out in 3 tubes (Pot 1, Pot 2, and Pot 3) with three libraries of Cas9 complexes reconstituted with cRNAs that generate overlapping fragments about 500 bp in size. Such overlapping fragments can improve the sequencing accuracy.

[0350] In some embodiments, the present disclosure provides a method for targeted haplotype sequencing (phased sequencing). In some embodiments, the method provided herein further includes diluting a DNA sample containing the target DNA to haploid content. Phase or haplotype information, which refers to the unique content of the two homologous chromosomes in diploid organisms, provides a useful tool to better understand relationships between human DNA sequence and phenotype, including diseases. The present disclosure provides a method for haplotype sequencing using CRISPR-Cas or Argonaute systems. A haplotype sequencing workflow can take advantage of the ability of Cas9 or Argonuate proteins to hold onto ends of cleaved DNA. Since Cas9 or Argonaute proteins remain association with the ends of cleaved DNAs, this creates a haplotype block of DNA proportional in size to the number and distance between Cas9 or Argonuate target regions in a target sequence. In some embodiments, following cleave, reactions can be diluted in mcrotiter wells to subhaplotype levels, and then can be treated with protease to release joined fragments anc converted into a sequencing library, *e.g.*, using TruSeq Nano library preparation method available from Illumina, Inc. (San Diego, CA).

[0351] In some embodiments, the target nucleic acid provided herein is a double-stranded DNA. In some embodiments, the target nucleic acid provided herein is a genomic DNA, a chromosomal DNA, a genome, or a partial genome.

[0352] In some embodiments, the nucleic acid fragments can be amplified, *e.g.*, using limited-cycle polymerase chain reaction (PCR), to introduce other end sequences or adaptors, *e.g.,* index, universal primers and other sequences required for cluster formation and sequencing.

[0353] In some embodiments, the sequencing the nucleic acid fragments includes use of one or more of sequencing by synthesis, bridge PCR, chain termination sequencing, sequencing by hybridization, nanopore sequencing, and sequencing by ligation.

[0354] In some embodiments, the sequencing methodology used in the method provided herein is sequencing-by-synthesis (SBS). In SBS, extension of a nucleic acid primer along a nucleic acid template (*e.g.* a target nucleic acid or amplicon thereof) is monitored to determine the sequence of nucleotides in the template. The underlying chemical process can be polymerization (*e.g.* as catalyzed by a polymerase enzyme). In a particular polymerase-based SBS embodiment, fluorescently labeled nucleotides are added to a primer (thereby extending the primer) in a template dependent fashion such that detection of the order and type of nucleotides added to the primer can be used to determine the sequence of the template.

[0355] Other sequencing procedures that use cyclic reactions can be used, such as pyrosequencing. Pyrosequencing detects the release of inorganic pyrophosphate (PPi) as particular nucleotides are incorporated into a nascent nucleic acid strand (Ronaghi, et al., Analytical Biochemistry 242(1), 84-9 (1996); Ronaghi, Genome Res. 11(1), 3-11 (2001); Ronaghi et al. Science 281(5375), 363 (1998); US 6,210,891; US 6,258,568 and US. 6,274,320,). In pyrosequencing, released PPi can be detected by being immediately converted to adenosine triphosphate (ATP) by ATP sulfurylase, and the level of ATP generated can be detected via luciferase-produced photons. Thus, the sequencing reaction can be monitored via a luminescence detection system. Excitation radiation sources used for fluorescence based detection systems are not necessary for pyrosequencing procedures. Useful fluidic systems, detectors and procedures that can be adapted for application of pyrosequencing to amplicons produced according to the present disclosure are described, for example, in WIPO Pat. App. Ser. No. PCT/US11/57111, US 2005/0191698 A1, US 7,595,883, and US 7,244,559.

[0356] Some embodiments can utilize methods involving the real-time monitoring of DNA polymerase activity. For example, nucleotide incorporations can be detected through fluorescence resonance energy transfer (FRET) interactions between a fluorophore-bearing polymerase and γ-phosphate-labeled nucleotides, or with zeromode waveguides (ZMWs). Techniques and reagents for FRET-based sequencing are described, for example, in Levene et al. Science 299, 682-686 (2003); Lundquist et al. Opt. Lett. 33, 1026-1028 (2008); Korlach et al. Proc. Natl. Acad. Sci. USA 105, 1176-1181 (2008).

[0357] Some SBS embodiments include detection of a proton released upon incorporation of a nucleotide into an extension product. For example, sequencing based on detection of released protons can use an electrical detector and associated techniques that are commercially available from Ion Torrent (Guilford, CT, a Life Technologies subsidiary) or sequencing methods and systems described in US 2009/0026082 A1; US 2009/0127589 A1; US 2010/0137143 A1; or US 2010/0282617 A1. Methods set forth herein for amplifying target nucleic acids using kinetic exclusion can be readily applied to substrates used for detecting protons. More specifically, methods set forth herein can be used to produce clonal populations of amplicons that are used to detect protons.

[0358] Another useful sequencing technique is nanopore sequencing (*see*, for example, Deamer et al. Trends Biotechnol. 18, 147-151 (2000); Deamer et al. Acc. Chem. Res. 35:817-825 (2002); Li et al. Nat. Mater. 2:611-615 (2003),). In some nanopore embodiments, the target nucleic acid or individual nucleotides removed from a target nucleic acid pass through a nanopore. As the nucleic acid or nucleotide passes through the nanopore, each nucleotide type can be identified by measuring fluctuations in the electrical conductance of the pore. (U.S. Patent No. 7,001,792; Soni et al. Clin. Chem. 53, 1996-2001 (2007); Healy, Nanomed. 2, 459-481 (2007); Cockroft et al. J. Am. Chem. Soc. 130, 818-820 (2008),).

[0359] From the foregoing description, it will be apparent that variations and modifications can be made to the invention described herein to adopt it to various usages and conditions. Such embodiments are also within the scope of the following claims.

[0360] The recitation of a listing of elements in any definition of a variable herein includes definitions of that variable as any single element or combination (or subcombination) of listed elements. The recitation of an embodiment herein includes that embodiment as any single embodiment or in combination with any other embodiments or portions thereof.

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.

Patent documents cited in the description

- WO2016014409A **[0005]**
- WO2016077350A **[0005]**
- US20150211058A **[0005]**
- US5432272A **[0115]**
- US5034506A **[0115]**
- US4683195A **[0116]**
- US4883750A **[0117]**
- US5476930A **[0117]**
- US5593826A **[0117]**
- US5871921A **[0117]**
- US20100120098A **[0188]** **[0191]** **[0193]** **[0259]**
- WO9523875A **[0188]** **[0192]** **[0259]**
- WO2010048605A **[0194]**

- US7413857B [0207]
- US20050153333A [0207]
- WO0216639A [0207]
- WO0056877PCT [0207]
- AU0029742PCT [0207]
- US5523204A [0207]
- US5536649A [0207]
- US5624825A [0207]
- US5631147A [0207]
- US5648211A [0207]
- US5733752A [0207]
- US5744311A [0207]
- US5756702A [0207]
- US5916779A [0207]
- US6238868B [0207]
- US6309833B [0207]
- US6326173B [0207]
- US5849547A [0207]
- US5874260A [0207]
- US6218151B [0207]
- US5786183A [0207]
- US6087133A [0207]
- US6214587B [0207]
- US6063604A [0207]
- US6251639B [0207]
- US6410278B [0207]
- WO0028082A [0207]
- US5591609A [0207]
- US5614389A [0207]
- US5773733A [0207]
- US5834202A [0207]
- US6448017B [0207]
- US6124120A [0207]
- US6280949B [0207]
- US7985565B [0215] [0215] [0215]
- US7115400B [0215] [0215] [0215]
- WO04018497A [0216]
- US7057026B [0216]
- WO9106678A [0216]
- WO07123744A [0216]
- US7329492B [0216]
- US7211414B [0216]
- US7315019B [0216]
- US7405281B [0216]

- US20080108082A [0216]
- US6210891B [0220] [0355]
- US6258568B [0220] [0355]
- US6274320B [0220] [0355]
- US1157111W [0220] [0355]
- US20050191698A1 [0220] [0355]
- US7595883B [0220] [0355]
- US7244559B [0220] [0355]
- US20090026082A1 [0222] [0357]
- US20090127589A1 [0222] [0357]
- US20100137143A1 [0222] [0357]
- US20100282617A1 [0222] [0357]
- US7001792B [0223] [0358]

Non-patent literature cited in the description

- **ENDER C.MEISTER, G.**Argonaute proteins, 2010, [0006]
- **GILLGHAEMI**Nucleosides, Nucleotides, and Nucleic Acids, 2008, vol. 27, 224-243 [0099] [0100]
- **CRAWBALACHANDRAN**Lab Chip, 2012, vol. 12, 2469-2486 [0101]
- Nucleic Acid Hybridization, A Practical ApproachIRL Press19850000 [0111]
- **WETMURDAVIDSON**Mol. Biol., 1968, vol. 31, 349- [0111]
- **SCHEIT**Nucleotide AnalogsJohn Wiley19800000 [0115]
- **ENGLISCH**Angew. Chem. Int. Ed. Engl., 1991, vol. 30, 613-29 [0115]
- **AGARWAL**Protocols for Polynucleotides and AnalogsHumana Press19940000 [0115]
- **S. VERMAF. ECKSTEIN**Ann. Rev. Biochem., 1998, vol. 67, 99-134 [0115]
- The Glen Report, 2003, vol. 16, 25- [0115]
- **KOSHKIN et al.T**etrahedron, 1998, vol. 54, 3607-30 [0115]
- **STERCHAK, E. P. et al.**Organic Chern., 1987, vol. 52, 4202- [0115]
- **MULLIS et al.**Cold Spring Harbor Symp. Quant. Biol., 1987, vol. 51, 263- [0116]
- PCR TechnologyStockton Press19890000 [0116]
- **MAKAROVA et al.**Nat Rev Microbiol, 2011, vol. 9, 467-77 [0151]
- **BROUNS et al.**Science, 2008, vol. 321, 960-4 [0152]
- **SINKUNAS et al.**EMBO J, 2011, vol. 30, 1335-1342 [0152]
- **BELOGLAZOVA et al.**EMBO J, 2011, vol. 30, 616-627 [0152]
- **JINEK et al.**Science, 2012, vol. 337, 6096816-821 [0153]
- **MARRAFFINISONTHEIMER**Science, 2008, vol. 322, 1843-1845 [0154]
- **HALE et al.**Cell, 2009, vol. 139, 945-956 [0154]
- **JINEK et al.**Science, 2012, vol. 337, 816-821 [0158] [0234]
- **MATTHIJS et al.**Nature Structural & Molecular Biology, 2011, vol. 18, 529-536 [0161]

- **HOWARD et al.**Biochem J., 2011, vol. 439, 185-95 [0162] [0162]
- **HAFT et al.**PLoS Comput Biol., 2005, vol. 1, 6e60- [0163]
- **ZHANG et al.**Nucl. Acids Res., 2013, [0163]
- **HOCKMEISTER**Genome Biol, 2008, vol. 9, 2210- [0164]
- **GAO et al.**nature biotechnology, 2016, [0164]
- **TYAGI et al.**Nature Biotechnology, 1996, 303-308 [0180]
- **LAKOWICZ**Principles of Fluorescence SpectroscopyPlenum Publishing Corporation19990701 [0180]
- **GORYSHINREZNIKOFF**J. Biol. Chem., 1998, vol. 273, 7367- [0188] [0192] [0259]
- **MIZUUCHI**Cell, 1983, vol. 35, 785- [0188] [0192] [0259]
- **SAVILAHTI et al.**EMBO J, 1995, vol. 14, 4893- [0188]
- **COLEGIO et al.**J. Bacteriol., 2001, vol. 183, 2384-8 [0188] [0259]
- **KIRBY et al.**Mol. Microbiol., 2002, vol. 43, 173-86 [0188] [0259]
- **DEVINEBOEKE**Nucleic Acids Res., 1994, vol. 22, 3765-72 [0188] [0192]
- **CRAIG**Science, 1996, vol. 271, 1512- [0188] [0192] [0259]
- **CRAIG**Curr Top Microbiol Immunol., 1996, vol. 204, 27-48 [0188] [0259]
- **KLECKNER et al.**Curr **T**op Microbiol Immunol., 1996, vol. 204, 49-82 [0188] [0259]
- **LAMPE et al.**EMBO J., 1996, vol. 15, 5470-9 [0188] [0192] [0259]
- **PLASTERK**Curr **T**op Microbiol Immunol, 1996, vol. 204, 125-43 [0188] [0192] [0259]
- **GLOOR**Methods Mol. Biol., 2004, vol. 260, 97-114 [0188] [0259]
- **ICHIKAWAOHTSUBO**J Biol. Chem., 1990, vol. 265, 18829-32 [0188] [0259]
- **OHTSUBOSEKINE**Curr. **T**op. Microbiol. Immunol., 1996, vol. 204, 1-26 [0188] [0192] [0259]
- **BROWN et al.**Proc Natl Acad Sci USA, 1989, vol. 86, 2525-9 [0188] [0192] [0259]
- **BOEKECORCES**Annu Rev Microbiol., 1989, vol. 43, 403-34 [0188] [0192] [0259]
- **SAVILAHTI et al.**EMBO J., 1995, vol. 14, 4893- [0192] [0259]
- **COLEGIO et al.**J Bacterid., 2001, vol. 183, 2384-8 [0192]
- **KIRBY et al.**Mol Microbiol, 2002, vol. 43, 173-86 [0192]
- **CRAIG**Curr Top Microbiol Immunol, 1996, vol. 204, 27-48 [0192]
- **KLECKNER et al.**Curr **T**op Microbiol Immunol, 1996, vol. 204, 49-82 [0192]
- **GLOOR**Methods Mol Biol, 2004, vol. 260, 97-114 [0192]
- **ICHIKAWAOHTSUBO**J Biol Chem., 1990, vol. 265, 18829-32 [0192]
- **BENTLEY et al.**Nature, 2008, vol. 456, 53-59 [0216]
- **RONAGHI et al.**Analytical Biochemistry, 1996, vol. 242, 184-9 [0220] [0355]
- **RONAGHI**Genome Res., 2001, vol. 11, 13-11 [0220] [0355]
- **RONAGHI et al.**Science, 1998, vol. 281, 5375363- [0220] [0355]
- **LEVENE et al.**Science, 2003, vol. 299, 682-686 [0221] [0356]
- **LUNDQUIST et al.**Opt. Lett., 2008, vol. 33, 1026-1028 [0221] [0356]
- **KORLACH et al.**Proc. Natl. Acad. Sci. USA, 2008, vol. 105, 1176-1181 [0221] [0356]
- **DEAMER et al.**Trends Biotechnol., 2000, vol. 18, 147-151 [0223] [0358]
- **DEAMER et al.**Acc. Chem. Res., 2002, vol. 35, 817-825 [0223] [0358]
- **LI et al.**Nat. Mater., 2003, vol. 2, 611-615 [0223] [0358]
- **SONI et al.**Clin. Chem, 2007, vol. 53, 1996-200 [0223]
- **HEALY**Nanomed, 2007, vol. 2, 459-481 [0223] [0358]

- **COCKROFT et al.**J. Am. Chem. Soc, 2008, vol. 130, 818-820 [0223]
- **DEVINEBOEKE**Nucleic Acids Res, 1994, vol. 22, 3765-72 [0259]
- **LI et al.**Clinical Chemistry, 2004, vol. 50, 61002-1011 [0276]
- **LI et al.**The Journal of the American Medical Association, 2005, vol. 293, 7843-849 [0276]
- **CHIU et al.**Clinical Chemistry, 2001, vol. 47, 91607-1613 [0276]
- **LEGLER et al.**Prenatal Diagnosis, 2007, vol. 27, 9824-829 [0276]
- **DHALLAN et al.**The Journal of the American Medical Association, 2004, vol. 291, 91114-1119 [0276]
- **BETTEGOWDA et al.**Sci. Transl. Med, 2014, vol. 6, 22424- [0277]
- **SONI et al.**Clin. Chem., 2007, vol. 53, 1996-2001 [0358]
- **COCKROFT et al.**J. Am. Chem. Soc., 2008, vol. 130, 818-820 [0358]

PATENTKRAV:

1. Fremgangsmåde til amplificering af en dobbeltstrenget målnukleinsyre, som omfatter:
   (a) tilvejebringelse af et system, der har:
   en 5'-phosphoryleret enkeltstrenget nukleinsyre eller et derivat deraf og et argonautprotein eller en variant deraf,
   hvor den 5'-phosphorylerede enkeltstrengede nukleinsyre eller derivatet deraf indeholder en målspecifik nukleotidregion, der er komplementær til en region af en første streng af den dobbeltstrengede målnukleinsyre;
   (b) etablering af kontakt mellem den dobbeltstrengede målnukleinsyre og systemet til frembringelse af et kompleks;
   (c) hybridisering af en primer til en anden streng af den dobbeltstrengede målnukleinsyre, hvilken primer indeholder en sekvens, der er komplementær til en region af den anden streng af den dobbeltstrengede målnukleinsyre, og
   (d) forlængelse af en nukleinsyre, der er komplementær til den anden streng af den dobbeltstrengede målnukleinsyre, fra primeren ved anvendelse af en polymerase.

2. Fremgangsmåde ifølge krav 1, som yderligere omfatter gentagelse af trin (a) til trin (d) en eller flere gange.

3. Fremgangsmåde ifølge krav 1, hvor den dobbeltstrengede målnukleinsyre amplificeres lineært eller
   amplificeres eksponentielt.

4. Fremgangsmåde ifølge krav 1, hvor målnukleinsyren er et dobbeltstrenget DNA (dsDNA) eller
   et dobbeltstrenget RNA (dsRNA).

5. Fremgangsmåde ifølge krav 1, hvor argonautproteinet er *Natronobacterium gregorgyi*-argonaut (Ng-argonaut).

6. Fremgangsmåde ifølge krav 1, hvor den 5'-phosphorylerede enkeltstrengede nukleinsyre er et enkeltstrenget DNA eller
   et enkeltstrenget DNA på mindre end 25 nukleotider.

7. Fremgangsmåde ifølge krav 1, hvor den første streng af den dobbeltstrengede målnukleinsyre indeholder en universel sekvens, og hvor den 5'-phosphorylerede enkeltstrengede nukleinsyre eller derivatet deraf indeholder en sekvens, der er

komplementær til en region af den universelle sekvens, eventuelt
hvor primeren indeholder en sekvens af en region af den universelle sekvens.

8. Fremgangsmåde ifølge krav 1, hvor polymerasen er en strengforskydende polymerase, eventuelt
hvor polymerasen er valgt fra gruppen, der består af Bst, Bsu og Phi29.

9. Fremgangsmåde ifølge krav 1, som yderligere omfatter:
tilførsel af mindst én transposase og mindst én transposon-endesammensætning, der indeholder en overført streng, til en prøve, der indeholder en målnukleinsyre, under betingelser, hvor målnukleinsyren og transposon-endesammensætningen gennemgår en transpositionsreaktion til frembringelse af en blanding, hvor målnukleinsyren er fragmenteret til frembringelse af en flerhed af målnukleinsyrefragmenter, og
inkorporering af en universel primersekvens i hver af flerheden af målnukleinsyrefragmenter,
hvor den 5'-phosphorylerede enkeltstrengede nukleinsyre eller derivatet deraf indeholder en målspecifik nukleotidregion, der er komplementær til en region af den universelle primer.

10. Fremgangsmåde ifølge krav 9, hvor A) den universelle primer inkorporeres i flerheden af målnukleinsyrefragmenter ved hjælp af en PCR-reaktion, eller
B) to universelle primere inkorporeres i to ender af hver af flerheden af målnukleinsyrefragmenter.

11. Fremgangsmåde til amplificering af en dobbeltstrenget målnukleinsyre ifølge krav 1, hvor systemet omfatter et første og andet system, hvilken fremgangsmåde omfatter:

(a) tilvejebringelse af et første system, der har:
en første 5'-phosphoryleret enkeltstrenget nukleinsyre eller et derivat deraf og et første argonautprotein eller en variant deraf,
hvor den første 5'-phosphorylerede enkeltstrengede nukleinsyre eller derivatet deraf indeholder en målspecifik nukleotidregion, der er komplementær til en region af en første streng af den dobbeltstrengede målnukleinsyre;

(b) tilvejebringelse af et andet system, der har:
en anden 5'-phosphoryleret enkeltstrenget nukleinsyre eller et derivat deraf og et andet argonautprotein eller en variant deraf,
hvor den anden 5'-phosphorylerede enkeltstrengede nukleinsyre eller derivatet deraf indeholder en målspecifik nukleotidregion, der er komplementær til en region af en anden streng af den dobbeltstrengede målnukleinsyre;

(c) etablering af kontakt mellem den dobbeltstrengede målnukleinsyre og det første system og det andet system;

(d) hybridisering af en første primer til en anden streng af den dobbeltstrengede målnukleinsyre, hvilken første primer indeholder en sekvens, der er komplementær til en region af den anden streng af den dobbeltstrengede målnukleinsyre, og hybridisering af en anden primer til en første streng af den dobbeltstrengede målnukleinsyre, hvilken anden primer indeholder en sekvens, der er komplementær til en region af den første streng af den dobbeltstrengede målnukleinsyre, og

(e) forlængelse af 3'-enden af den første primer og den anden primer med en eller flere polymeraser til frembringelse af en første og en anden dobbeltstrenget målnukleinsyre.

12. Fremgangsmåde ifølge krav 11, i) som yderligere omfatter gentagelse af trin (a) til trin (e) en eller flere gange, eller

ii) hvor målnukleinsyren er et dobbeltstrenget DNA (dsDNA), eller

iii) hvor målnukleinsyren er et dobbeltstrenget RNA (dsRNA), eller

iv) hvor det første og det andet argonautprotein er *Natronobacterium gregorgyi*-argonauter (Ng-argonauter), eller

v) hvor den første og den anden 5'-phosphorylerede enkeltstrengede nukleinsyre er enkeltstrengede DNA'er, eller

vi) hvor den første og den anden 5'-phosphorylerede enkeltstrengede nukleinsyre er enkeltstrengede DNA'er på mindre end 25 nukleotider, eller

vii) hvor:
den første streng af den dobbeltstrengede målnukleinsyre indeholder en første universel sekvens, og hvor den første 5'-phosphorylerede enkeltstrengede nukleinsyre eller derivatet deraf i det første system indeholder en sekvens, der er komplementær til en region af den første universelle sekvens, og
den anden streng af den dobbeltstrengede målnukleinsyre indeholder en anden universel sekvens, og hvor den anden 5'-phosphorylerede enkeltstrengede nukleinsyre eller derivatet deraf i det andet system indeholder en sekvens, der er komplementær til en region af den anden universelle sekvens, eller

viii) hvor den første primer indeholder en sekvens af en region af den første universelle sekvens, og den anden primer indeholder en sekvens af en region af den anden universelle sekvens, eller

ix) hvor polymerasen er en strengforskydende polymerase, eventuelt hvor polymerasen er valgt fra gruppen, der består af Bst, Bsu og Phi29.

13. Fremgangsmåde ifølge et hvilket som helst af kravene 1-12, hvor målnukleinsyren er genomisk DNA.

14. Fremgangsmåde ifølge et hvilket som helst af kravene 1-12, hvor målnukleinsyren indeholder kromosomalt DNA eller et fragment deraf.

15. Fremgangsmåde ifølge et hvilket som helst af kravene 1-12, hvor målnukleinsyren omfatter et genom eller et partielt genom.

DRAWINGS

Standard P5 & P7:
P5 seq: 5' AATGATACGGCGACCACCGAGATCTACAC '3 (SEQ ID No. 1)
P7 seq: 5' CAAGCAGAAGACGGCATACGAGAT '3 (SEQ ID No. 2)

PAM-Modified P5 & P7:
P5 seq: 5' CCN AATGATACGGCGACCACCGAGATCTACAC '3 (SEQ ID No. 3)
P7 seq: 5' CCN CAAGCAGAAGACGGCATACGAGAT '3 (SEQ ID No. 4)

Truncated, PAM-Modified P5 & P7:
P5 seq: 5' CCN AATGATACGGCGACCACC~~GAGATCTACAC~~ '3 (SEQ ID No. 5)
P7 seq: 5' CCN CAAGCAGAAGACGGCATA~~CGAGAT~~ '3 (SEQ ID No. 6)

Guide RNAs:
P5 targeting: 5' UCGGUGGUCGCCGUAUCAUU 3' (SEQ ID No. 7)
P7 targeting: 5' CGUAUGCCGUCUUCUGCUUG 3' (SEQ ID No. 8)

Figure 1A

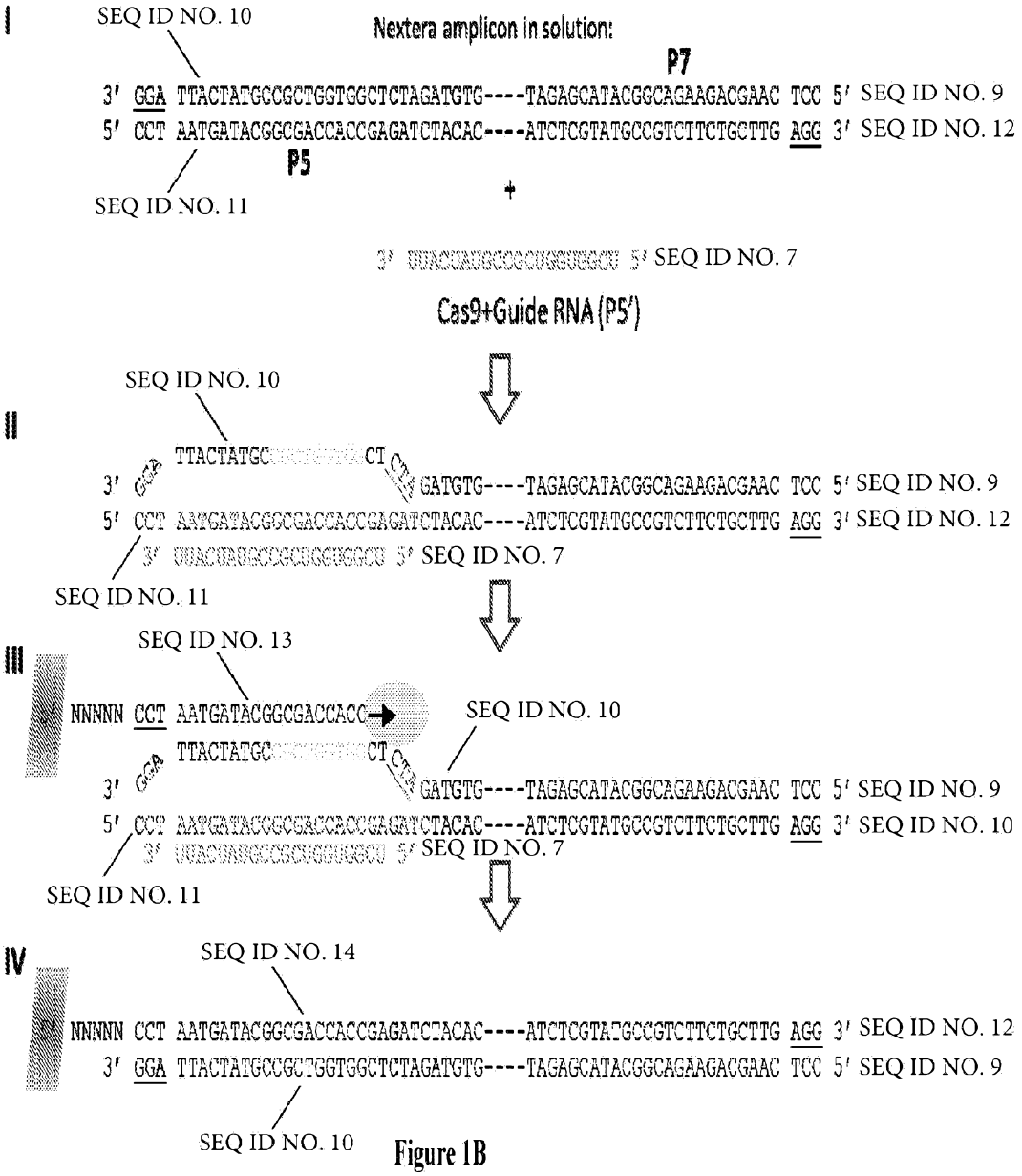
I
SEQ ID NO. 10
Nextera amplicon in solution:
P7
3' GGA TTACTATGCCGCTGGTGGCTCTAGATGTG----TAGAGCATACGGCAGAAGACGAAC TCC 5' SEQ ID NO. 9
5' CCT AATGATACGGCGACCACCGAGATCTACAC----ATCTCGTATGCCGTCTTCTGCTTG AGG 3' SEQ ID NO. 12
P5
SEQ ID NO. 11
+
3' UUACGAUGCCGCUGGUGGCU 5' SEQ ID NO. 7
Cas9+Guide RNA (P5')
II
SEQ ID NO. 10
TTACTATGC CT
3' GGA GATGTG----TAGAGCATACGGCAGAAGACGAAC TCC 5' SEQ ID NO. 9
5' CCT AATGATACGGCGACCACCGAGATCTACAC----ATCTCGTATGCCGTCTTCTGCTTG AGG 3' SEQ ID NO. 12
3' UUACUAUGCCGCUGGUGGCU 5' SEQ ID NO. 7
SEQ ID NO. 11
III
SEQ ID NO. 13
NNNNN CCT AATGATACGGCGACCACC
SEQ ID NO. 10
TTACTATGC CT
3' GGA GATGTG----TAGAGCATACGGCAGAAGACGAAC TCC 5' SEQ ID NO. 9
5' CCT AATGATACGGCGACCACCGAGATCTACAC----ATCTCGTATGCCGTCTTCTGCTTG AGG 3' SEQ ID NO. 10
3' UUACUAUGCCGCUGGUGGCU 5' SEQ ID NO. 7
SEQ ID NO. 11
IV
SEQ ID NO. 14
NNNNN CCT AATGATACGGCGACCACCGAGATCTACAC----ATCTCGTATGCCGTCTTCTGCTTG AGG 3' SEQ ID NO. 12
3' GGA TTACTATGCCGCTGGTGGCTCTAGATGTG----TAGAGCATACGGCAGAAGACGAAC TCC 5' SEQ ID NO. 9
SEQ ID NO. 10

Figure 1B

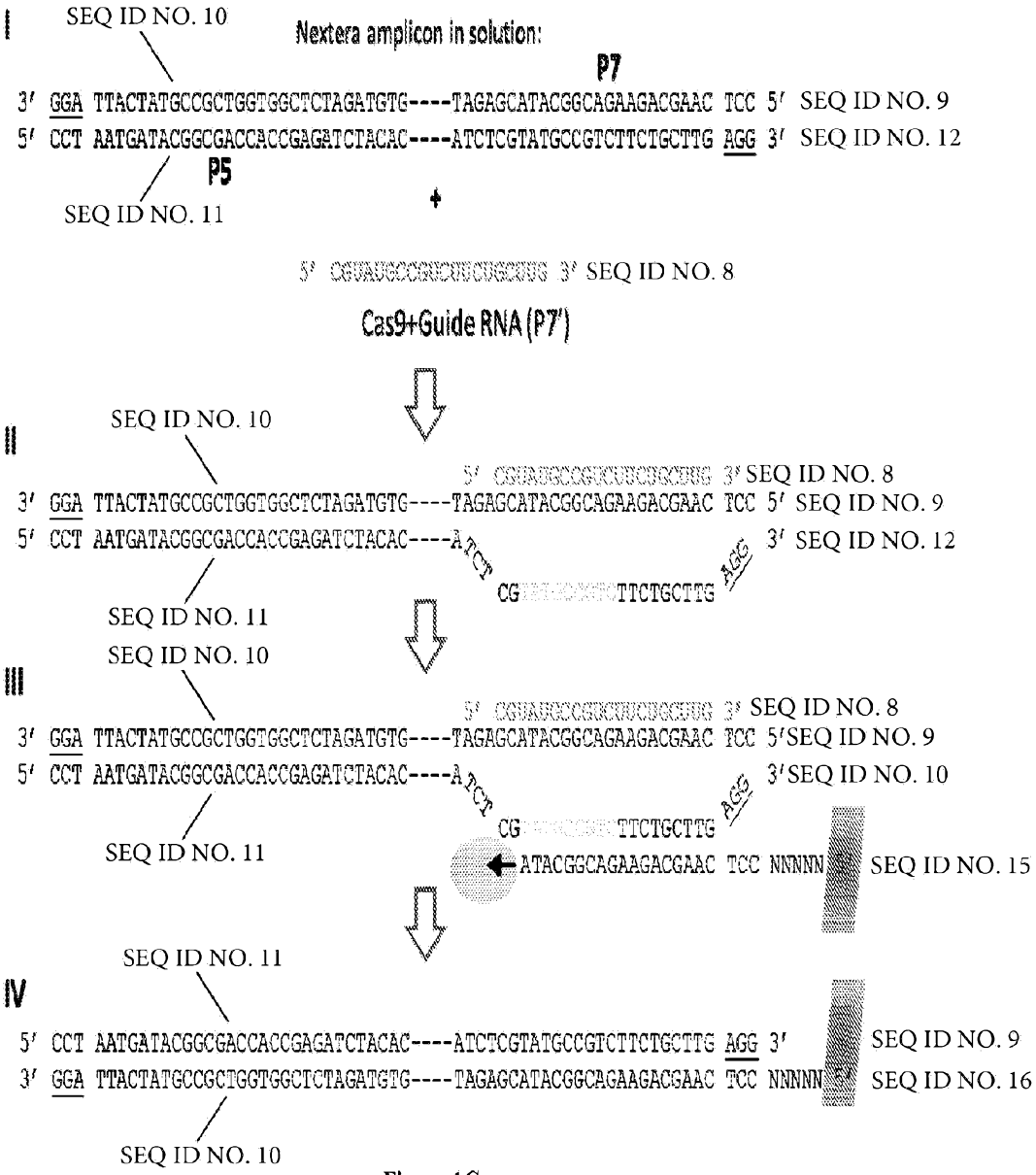
I
SEQ ID NO. 10
Nextera amplicon in solution:
P7
3′ GGA TTACTATGCCGCTGGTGGCTCTAGATGTG----TAGAGCATACGGCAGAAGACGAAC TCC 5′ SEQ ID NO. 9
5′ CCT AATGATACGGCGACCACCGAGATCTACAC----ATCTCGTATGCCGTCTTCTGCTTG AGG 3′ SEQ ID NO. 12
P5
SEQ ID NO. 11
5′ CGUAUGCCGUCUUCUGCUUG 3′ SEQ ID NO. 8
Cas9+Guide RNA (P7′)
II
SEQ ID NO. 10
5′ CGUAUGCCGUCUUCUGCUUG 3′ SEQ ID NO. 8
3′ GGA TTACTATGCCGCTGGTGGCTCTAGATGTG----TAGAGCATACGGCAGAAGACGAAC TCC 5′ SEQ ID NO. 9
5′ CCT AATGATACGGCGACCACCGAGATCTACAC----A
ATCT
CG
TTCTGCTTG
AGG
3′ SEQ ID NO. 12
SEQ ID NO. 11
III
SEQ ID NO. 10
5′ CGUAUGCCGUCUUCUGCUUG 3′ SEQ ID NO. 8
3′ GGA TTACTATGCCGCTGGTGGCTCTAGATGTG----TAGAGCATACGGCAGAAGACGAAC TCC 5′SEQ ID NO. 9
5′ CCT AATGATACGGCGACCACCGAGATCTACAC----A
ATCT
CG
TTCTGCTTG
AGG
3′SEQ ID NO. 10
SEQ ID NO. 11
ATACGGCAGAAGACGAAC TCC NNNNN
SEQ ID NO. 15
IV
SEQ ID NO. 11
5′ CCT AATGATACGGCGACCACCGAGATCTACAC----ATCTCGTATGCCGTCTTCTGCTTG AGG 3′
SEQ ID NO. 9
3′ GGA TTACTATGCCGCTGGTGGCTCTAGATGTG----TAGAGCATACGGCAGAAGACGAAC TCC NNNNN
SEQ ID NO. 16
SEQ ID NO. 10

Figure 1C