



(51) International Patent Classification:  
C12N 15/10 (2006.01) G01N 33/52 (2006.01)  
C12Q 1/68 (2006.01)

(74) Agents: MULLEN, James, J. et al.; Morrison & Foerster Llp, 12531 High Bluff Drive, Suite 100, San Diego, CA 92130-2040 (US).

(21) International Application Number:  
PCT/US2016/047631

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

(22) International Filing Date:  
18 August 2016 (18.08.2016)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
62/207,359 19 August 2015 (19.08.2015) US

(71) Applicant: ARC BIO, LLC [US/US]; 16 Cherry Street, Cambridge, MA 02139 (US).

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

(72) Inventors: GOURGUECHON, Stephane, B.; C/o Arc Bio, Llc, 16 Cherry Street, Cambridge, MA 02139 (US). HARNESS, Eric; Arc Bio, LLC, 16 Cherry Street, Cambridge, MA 02139 (US). SINCLAIR, David; C/o Arc Bio, Llc, 16 Cherry Street, Cambridge, MA 02139 (US). CARPENTER, Meredith, L.; C/o Arc Bio, Llc, 16 Cherry Street, Cambridge, MA 02139 (US).

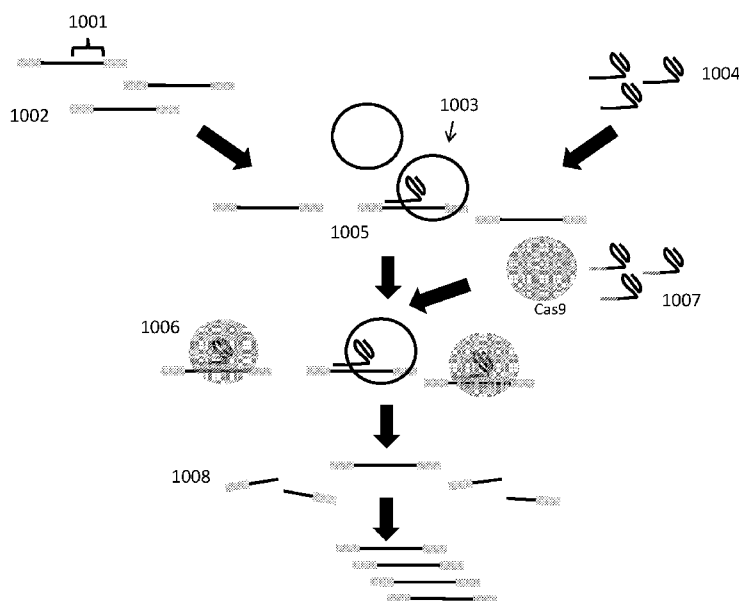
**Published:**

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

[Continued on next page]

(54) Title: CAPTURE OF NUCLEIC ACIDS USING A NUCLEIC ACID-GUIDED NUCLEASE-BASED SYSTEM

FIG. 10



(57) Abstract: Provided herein are methods and compositions for the capture of nucleic acids, for example by using a nucleic acid-guided nuclease-based system.



- 
- *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

## **CAPTURE OF NUCLEIC ACIDS USING A NUCLEIC ACID-GUIDED NUCLEASE-BASED SYSTEM**

### **CROSS REFERENCE TO RELATED APPLICATIONS**

[0001] This application claims priority to U.S. Provisional Application Serial No. 62/207,359, filed on August 19, 2015, which is incorporated herein by reference in its entirety.

### **BACKGROUND OF THE INVENTION**

[0002] Targeted sequencing of specific regions of the genome continues to be of interest to researchers, particularly in a clinical setting. Clinical diagnosis of genetic disease, cancer, and many research projects rely on targeted sequencing to enable high coverage sequencing of targeted sites while reducing sequencing costs. Currently, the primary methods used for this purpose are 1) hybridization-based enrichment, and 2) multiplex PCR. In the former approach, biotin-labeled short oligonucleotide probes are used to “pull out” sequences of interest from a library. This process can be time-consuming, expensive, and require many hands-on steps. Furthermore, often some “off-target” sequences can remain in the resulting product. The multiplex PCR-based approach can be faster, but the number of targets can be limited and the cost can be high. Needed are methods for the efficient capture of nucleic acid regions of interest that are easy, specific, rapid, and inexpensive. Provided herein are methods and compositions that address this need.

[0003] All patents, patent applications, publications, documents, web links, and articles cited herein are incorporated herein by reference in their entireties.

### **BRIEF SUMMARY OF THE INVENTION**

[0004] Provided herein are methods and compositions that allow for the selective capture of nucleic acid sequences of interest. The nucleic acids may contain DNA or RNA. The methods and compositions provided herein are particularly useful for working with complex nucleic acid samples.

[0005] In one aspect, the invention provides a method of capturing target nucleic acid sequences comprising: (a) providing a sample comprising a plurality of adapter-ligated nucleic acids, wherein the nucleic acids are ligated to a first adapter at one end and ligated to a second adapter at the other end; (b) contacting the sample with a plurality of nucleic acid-guided nuclease-gNA

complexes, wherein the gNAs are complementary to targeted sites of interest contained in a subset of the nucleic acids, thereby generating a plurality of nucleic acid fragments ligated to a first or second adapter at one end and no adapter at the other end; and (c) contacting the plurality of nucleic acid fragments with third adapters, thereby generating a plurality of nucleic acid fragments ligated to either the first or second adapter at one end and the third adapter at the other end. In one embodiment, the nucleic acid-guided nuclease is a CRISPR/Cas system protein. In one embodiment, the nucleic acid-guided nuclease is a non-CRISPR/Cas system protein. In one embodiment, the nucleic acid-guided nuclease is selected from the group consisting of CAS Class I Type I, CAS Class I Type III, CAS Class I Type IV, CAS Class II Type II, and CAS Class II Type V. In one embodiment, the nucleic acid-guided nuclease is selected from the group consisting of Cas9, Cpf1, Cas3, Cas8a-c, Cas10, Cse1, Csy1, Csn2, Cas4, Csm2, Cm5, Csf1, C2c2, and NgAgo. In one embodiment, the gNAs are gRNAs. In one embodiment, the gNAs are gDNAs. In one embodiment, the contacting with a plurality of nucleic acid-guided nuclease-gNA complexes cleaves the targeted sites of interest contained in a subset of the nucleic acids, thereby generating a plurality of nucleic acid fragments comprising a first or second adapter at one end and no adapter at the other end. In one embodiment the method further comprises amplifying the product of step (c) using first or second and third adapter-specific PCR. In one embodiment the nucleic acids are selected from the group consisting of single stranded DNA, double stranded DNA, single stranded RNA, double stranded RNA, and a DNA/RNA hybrid. In one embodiment the nucleic acids are double stranded DNA. In one embodiment the nucleic acids are from genomic DNA. In one embodiment the genomic DNA is human. In one embodiment the nucleic acids adapter-ligated ends are from 20 bp to 5000 bp in length. In one embodiment the targeted sites of interest are single nucleotide polymorphisms (SNPs), short tandem repeats (STRs), cancer genes, inserts, deletions, structural variations, exons, genetic mutations, or regulatory regions. In one embodiment amplified products are used for cloning, sequencing, or genotyping. In one embodiment the adapters are from 20bp to 100 bp in length. In one embodiment the adapters comprise primer binding sites. In one embodiment the adapters comprise sequencing adapters or restriction sites. In one embodiment the targeted sites of interest represent less than 50% of the total nucleic acid in the sample. In one embodiment the sample is obtained from a biological sample a clinical sample, a forensic sample, or an environmental sample. In one embodiment the first and second adapters are identical. In one embodiment the first and second adapters are different. In one embodiment the sample comprises a sequencing library.

[0006] In another aspect, the invention provides a method of introducing labeled nucleotides at targeted sites of interest comprising: (a) providing a sample comprising a plurality of nucleic acid fragments; (b) contacting the sample with a plurality of nucleic acid-guided nuclease nickase-gNA complexes wherein the gNAs are complementary to targeted sites of interest in the nucleic acid fragments, thereby generating a plurality of nicked nucleic acid fragments at the targeted sites of interest; and (c) contacting the plurality of nicked nucleic acid fragments with an enzyme capable of initiating nucleic acid synthesis at a nicked site, and labeled nucleotides, thereby generating a plurality of nucleic acid fragments comprising labeled nucleotides in the targeted sites of interest. In one embodiment, the nucleic acid-guided nuclease nickase is selected from the group consisting of CAS Class I Type I nickase, CAS Class I Type III nickase, CAS Class I Type IV nickase, CAS Class II Type II nickase, and CAS Class II Type V nickase. In one embodiment, the nucleic acid-guided nuclease nickase is selected from the group consisting of Cas9 nickase, Cpf1 nickase, Cas3 nickase, Cas8a-c nickase, Cas10 nickase, Cse1 nickase, Csy1 nickase, Csn2 nickase, Cas4 nickase, Csm2 nickase, Cm5 nickase, Csf1 nickase, C2C2 nickase, and NgAgo nickase. In one embodiment, the gNAs are gRNAs. In one embodiment, the gNAs are gDNAs. In one embodiment the nucleic acid fragments are selected from the group consisting of single stranded DNA fragments, double stranded DNA fragments, single stranded RNA fragments, double stranded RNA fragments, and a DNA/RNA hybrid fragments. In one embodiment the nucleic fragments are double stranded DNA fragments. In one embodiment double stranded DNA fragments are from genomic DNA. In one embodiment the genomic DNA is human. In one embodiment the nucleic acid fragments are from 20bp to 5000 bp in length. In one embodiment the targeted sites of interest are single nucleotide polymorphisms (SNPs), short tandem repeats (STRs), cancer genes, inserts, deletions, structural variations, exons, genetic mutations, or regulatory regions. In one embodiment the targeted sites of interest represent less than 50% of the total nucleic acid in the sample. In one embodiment the sample is obtained from a biological sample, a clinical sample, a forensic sample, or an environmental sample. In one embodiment the labeled nucleotides are biotinylated nucleotides. In one embodiment, the labeled nucleotides are part of an antibody conjugate pair. In one embodiment the method further comprises contacting the nucleic acid fragments comprising biotinylated nucleotides with avidin or streptavidin, thereby capturing the targeted nucleic acid sites of interest. In one embodiment the enzyme capable of initiating nucleic acid synthesis at a nicked site is DNA Polymerase I, a Klenow fragment, a TAQ polymerase, or a Bst DNA Polymerase. In one embodiment the Cas9 nickase nicks the 5' end of the nucleic acid fragments.

In one embodiment the nucleic acid fragments are from 20bp to 5000bp in length. In one embodiment the sample is obtained from a biological sample a clinical sample, a forensic sample, or an environmental sample.

[0007] In another aspect, the invention provides a method of capturing target nucleic acid sequences of interest comprising: (a) providing a sample comprising a plurality of adapter-ligated nucleic acids, wherein the nucleic acids are ligated to a first adapter at one end and are ligated to a second adapter at the other end; and (b) contacting the sample with a plurality of catalytically dead nucleic acid-guided nuclease-gRNA complexes, wherein the catalytically dead nucleic acid-guided nuclease is fused to a transposase, wherein the gRNAs are complementary to targeted sites of interest contained in a subset of the nucleic acids, and wherein the complexes are loaded with a plurality of third adapters, to generate a plurality of nucleic acids fragments comprising either a first or second adapter at one end, and a third adapter at the other end. In one embodiment the method further comprises amplifying the product of step (b) using first or second adapter and third adapter-specific PCR. In one embodiment, the catalytically dead nucleic acid-guided nuclease is derived from a CRISPR/Cas system protein. In one embodiment, the catalytically dead nucleic acid-guided nuclease is derived from a non-CRISPR/Cas system protein. In one embodiment, the catalytically-dead nucleic acid-guided nuclease is selected from the group consisting of dead CAS Class I Type I, dead CAS Class I Type III, dead CAS Class I Type IV, dead CAS Class II Type II, and dead CAS Class II Type V. In one embodiment, the catalytically dead nucleic acid-guided nuclease is selected from the group consisting of dCas9, dCpf1, dCas3, dCas8a-c, dCas10, dCse1, dCsy1, dCsn2, dCas4, dCsm2, dCm5, dCsf1, dC2C2, and dNgAgo. In one embodiment, the gNAs are gRNAs. In one embodiment, the gNAs are gDNAs. In one embodiment the nucleic acids sequences are from genomic DNA. In one embodiment the genomic DNA is human. In one embodiment the catalytically dead nucleic acid-guided nuclease is fused to the N-terminus of the transposase. In one embodiment the catalytically dead nucleic acid-guided nuclease is fused to the C-terminus of the transposase. In one embodiment the adapter-ligated nucleic acids are 20bp-5000bp. In one embodiment the contacting of step (b) allows for the insertion of the second adapter into the targeted nucleic acid sequences. In one embodiment the targeted sites of interest are single nucleotide polymorphisms (SNPs), short tandem repeats (STRs), cancer genes, inserts, deletions, structural variations, exons, genetic mutations, or regulatory regions. In one embodiment the amplified products are used for cloning, sequencing, or genotyping. In one embodiment the adapters are from 20 bp to 100bp in length. In one embodiment the adapters comprise primer binding sites. In one

embodiment the adapters comprise sequencing adapters or restriction sites. In one embodiment the targeted sites of interest represent less than 50% of the total nucleic acid in the sample. In one embodiment the sample is obtained from a biological sample a clinical sample, a forensic sample, or an environmental sample.

**[0008]** In another aspect, the invention provides a method of capturing target nucleic acid sequences of interest comprising: (a) providing a sample comprising a plurality of adapter-ligated nucleic acids, wherein the nucleic acids are ligated to the adapter at the 5' end and 3' ends; (b) contacting the sample with a plurality of catalytically dead nucleic acid-guided nuclease-gNA complexes, wherein the gNAs are complementary to targeted sites of interest contained in a subset of the nucleic acids, thereby generating a plurality of nucleic acids adapter-ligated at the 5' and 3' ends, bound to a catalytically dead nucleic acid-guided nuclease-gNA complex; and (c) contacting the sample with a plurality of catalytically dead nucleic acid-guided nuclease-gNA complexes, wherein the gNAs are complementary to both targeted sites of interest and targeted sites not of interest in the nucleic acids, thereby generating a plurality of nucleic acid fragments comprising nucleic acid sequences not of interest, adapter ligated at only one of the 5' or 3' ends. In one embodiment the contacting of step (c) does not displace the plurality of nucleic acids adapter-ligated at the 5' and 3' ends, bound to a dCAS9-gRNA complex of step (b). In one embodiment the contacting in step (d) cleaves the targeted sites not of interest contained in a subset of the nucleic acids, thereby generating a plurality of nucleic acid fragments comprising nucleic acid sequences not of interest, adapter ligated at only one of the 5' or 3' ends. In one embodiment the method further comprises removing the bound dCAS9-gRNA complex and amplifying the product of step (b) using adapter-specific PCR. In one embodiment the catalytically dead nucleic acid-guided nuclease is a CRISPR/Cas system protein. In one embodiment the catalytically dead nucleic acid-guided nuclease is a non-CRISPR/Cas system protein. In one embodiment, the catalytically-dead nucleic acid-guided nuclease is selected from the group consisting of dead CAS Class I Type I, dead CAS Class I Type III, dead CAS Class I Type IV, dead CAS Class II Type II, and dead CAS Class II Type V. In one embodiment the catalytically dead nucleic acid-guided nuclease is selected from the group consisting of dCas9, dCpf1, dCas3, dCas8a-c, dCas10, dCse1, dCsy1, dCsn2, dCas4, dCsm2, dCm5, dCsf1, dC2C2, and dNgAgo. In one embodiment the gNAs are gRNAs. In one embodiment the gNAs are gDNAs. In one embodiment the nucleic acids are selected from the group consisting of single stranded DNA, double stranded DNA, single stranded RNA, double stranded RNA, and a DNA/RNA hybrid. In one embodiment the nucleic acids are double stranded DNA. In one

embodiment the nucleic acids are from genomic DNA. In one embodiment the genomic DNA is human. In one embodiment the nucleic acids adapter-ligated at the 5' ends and the 3' ends are from 20 bp to 5000 bp in length. In one embodiment the targeted sites of interest are single nucleotide polymorphisms (SNPs), short tandem repeats (STRs), cancer genes, inserts, deletions, structural variations, exons, genetic mutations, or regulatory regions. In one embodiment the amplified products are used for cloning, sequencing, or genotyping. In one embodiment the adapters are from 20 bp to 100bp in length. In one embodiment the adapters comprise primer binding sites. In one embodiment the adapters comprise sequencing adapters or restriction sites. In one embodiment the targeted sites of interest represent less than 50% of the total nucleic acid in the sample. In one embodiment the sample is obtained from a biological sample, a clinical sample, a forensic sample, or an environmental sample.

**[0009]** In another aspect, the invention provides a method of capturing target DNA sequences of interest comprising: (a) providing a sample comprising a plurality of nucleic acid sequences, wherein the nucleic acid sequences comprise methylated nucleotides, and wherein the nucleic acid sequences are adapter ligated on the 5' and 3' ends; (b) contacting the sample with a plurality of nucleic acid-guided nuclease nickase-gRNA complexes, wherein the gRNAs are complementary to targeted sites of interest in a subset of the nucleic acid sequences, thereby generating a plurality of nicked sites of interest in the subset of the nucleic acid sequences, and wherein the target nucleic acid sequences are adapter ligated on the 5' and 3' ends; (c) contacting the sample with an enzyme capable of initiating DNA synthesis at a nicked site, and unmethylated nucleotides, thereby generating a plurality of nucleic acid sequences comprising unmethylated nucleotides in the targeted sites of interest and wherein the nucleic acid sequences are adapter ligated on the 5' and 3' ends; and (d) contacting the sample with an enzyme capable of cutting methylated nucleic acids, thereby generating a plurality of nucleic acid fragments comprising methylated nucleic acids, wherein the plurality of nucleic acid fragments comprising methylated nucleic acids that are adapter ligated on at most one of the 5' and 3' ends. In one embodiment, the nucleic acid-guided nuclease nickase is selected from the group consisting of CAS Class I Type I nickase, CAS Class I Type III nickase, CAS Class I Type IV nickase, CAS Class II Type II nickase, and CAS Class II Type V nickase. In one embodiment the nucleic acid-guided nuclease nickase is selected from the group consisting of Cas9 nickase, Cpf1 nickase, Cas3 nickase, Cas8a-c nickase, Cas10 nickase, Cse1 nickase, Csy1 nickase, Csn2 nickase, Cas4 nickase, Csm2 nickase, Cm5 nickase, Csf1 nickase, C2c2 nickase, and NgAgo nickase. In one embodiment the gRNAs are gRNAs. In one embodiment the gRNAs are gDNAs. In one



embodiment the DNA is double stranded DNA. In one embodiment the double stranded DNA is from genomic DNA. In one embodiment the genomic DNA is human. In one embodiment the DNA sequences are from 20 bp to 5000 bp in length. In one embodiment the targeted sites of interest are single nucleotide polymorphisms (SNPs), short tandem repeats (STRs), cancer genes, inserts, deletions, structural variations, exons, genetic mutations, or regulatory regions. In one embodiment the targeted sites of interest represent less than 50% of the total DNA in the sample. In one embodiment the sample is obtained from a biological sample, a clinical sample, a forensic sample, or an environmental sample. In one embodiment the enzyme capable of initiating nucleic acid synthesis at a nicked site is DNA Polymerase I, Klenow fragment a TAQ polymerase, or a Bst DNA Polymerase. In one embodiment the nucleic acid-guided nuclease nickase nicks the 5' end of the DNA sequences. In one embodiment the enzyme capable of cutting methylated DNA is DpnI.

**[0010]** In another aspect, the invention provides a method of capturing target DNA sequences of interest comprising: (a) contacting the sample with a plurality of nucleic acid-guided nuclease nickase -gNA complexes, wherein the gNAs are complementary to targeted sites flanking a region of interest in a subset of the DNA sequences, thereby generating a plurality of nicked DNA at sites adjacent to the regions of interest (b) heating to 65°C to cause nicks in close proximity to generate a double stranded break (c) contacting these double stranded breaks with a thermostable ligase thereby allowing ligation of adapter sequences at these sites only and (d), repeating these three steps to place a second adapter on the other side of the region of interest, thus allowing enrichment of the region of interest. In one embodiment, the gNAs are gRNAs. In one embodiment, the gNAs are gDNAs. In one embodiment, the nucleic acid-guided nuclease nickase is selected from the group consisting of CAS Class I Type I nickase, CAS Class I Type III nickase, CAS Class I Type IV nickase, CAS Class II Type II nickase, and CAS Class II Type V nickase. In one embodiment the nucleic acid-guided nuclease nickase is selected from the group consisting of Cas9 nickase, Cpf1 nickase, Cas3 nickase, Cas8a-c nickase, Cas10 nickase, Cse1 nickase, Csy1 nickase, Csn2 nickase, Cas4 nickase, Csm2 nickase, Cm5 nickase, Csf1 nickase, C2C2 nickase, and NgAgo nickase. In one embodiment the DNA is double stranded DNA. In one embodiment the double stranded DNA is from genomic DNA. In one embodiment the genomic DNA is human. In one embodiment the DNA sequences are from 20 bp to 5000 bp in length. In one embodiment the targeted sites of interest are single nucleotide polymorphisms (SNPs), short tandem repeats (STRs), cancer genes, inserts, deletions, structural variations, exons, genetic mutations, or regulatory regions. In one embodiment the targeted sites of interest

represent less than 50% of the total DNA in the sample. In one embodiment the sample is obtained from a biological sample, a clinical sample, a forensic sample, or an environmental sample. In one embodiment the ligase capable of contacting the double stranded break is Thermostable 5' App DNA/RNA ligase, or T4 RNA ligase. In one embodiment the nucleic acid-guided nuclease nicks the 5' end of the DNA sequences. In one embodiment the enzyme capable of cutting methylated DNA is DpnI.

[0011] In another aspect, the invention provides a method of enriching a sample for sequences of interest, comprising: (a) providing a sample comprising sequences of interest and targeted sequences for depletion, wherein the sequences of interest comprise less than 50% of the sample; and (b) contacting the sample with a plurality of either nucleic acid-guided RNA endonuclease-gRNA complexes or a plurality of nucleic acid-guided DNA endonuclease-gDNA complexes, wherein the gRNAs and gDNAs are complementary to the targeted sequences, and whereby the targeted sequences are cleaved. In one embodiment, the method further comprises extracting the sequences of interest and the targeted sequences for depletion from the sample. In one embodiment, the method further comprises fragmenting the extracted sequences. In one embodiment, the cleaved targeted sequences are removed by size- exclusion. In one embodiment, the sample is any one of a biological sample, a clinical sample, a forensic sample or an environmental sample. In one embodiment, the sample comprises host nucleic acid sequences targeted for depletion and non-host nucleic acid sequences of interest. In one embodiment, the non-host nucleic acid sequences comprise microbial nucleic acid sequences. In one embodiment, the microbial nucleic acid sequences are bacterial, viral or eukaryotic parasitic nucleic acid sequences. In one embodiment, the gRNAs and gDNAs are complementary to ribosomal RNA sequences, spliced transcripts, unspliced transcripts, introns, exons, or noncoding RNAs. In one embodiment, the extracted nucleic acids include single-stranded or double-stranded RNA. In one embodiment, the extracted nucleic acids include single-stranded or double-stranded DNA. In one embodiment, the sequences of interest comprise less than 10% of the extracted nucleic acids. In one embodiment, the nucleic acid-guided RNA endonuclease comprises C2c2. In one embodiment, the C2c2 is catalytically dead. In one embodiment, the nucleic acid-guided DNA endonuclease comprises NgAgo. In one embodiment, the NgAgo is catalytically dead. In one embodiment, the sample is selected from whole blood, plasma, serum, tears, saliva, mucous, cerebrospinal fluid, teeth, bone, fingernails, feces, urine, tissue, and a biopsy.

**[0012]** In another aspect, the invention provides a method of enriching a sample comprising: (a) providing a sample comprising host nucleic acids and non-host nucleic acids; (b) contacting the sample with a plurality of nucleic acid-guided RNA endonuclease -gRNA complexes or a plurality of nucleic acid-guided DNA endonuclease -gDNA complexes, wherein the gRNAs and gDNAs are complementary to targeted sites in the host nucleic acids, and (c) enriching the sample for non-host nucleic acids. In one embodiment, the nucleic acid-guided RNA endonuclease comprises C2c2. In one embodiment, the nucleic acid-guided RNA endonuclease comprises catalytically dead C2c2. In one embodiment, the nucleic acid-guided DNA endonuclease comprises NgAgo. In one embodiment, the nucleic acid-guided DNA endonuclease comprises catalytically dead NgAgo. In one embodiment, the host is selected from the group consisting of a human, cow, horse, sheep, pig, monkey, dog, cat, gerbil, bird, mouse, and rat. In one embodiment, the non-host is a prokaryotic organism. In one embodiment, the non-host is selected from the group consisting of a eukaryote, a virus, a bacterium, a fungus, and a protozoan. In one embodiment, the adapter-ligated host nucleic acids and non-host nucleic acids range from 50 bp to 1000 bp in length. In one embodiment, the non-host nucleic acids comprise less than 50% of the total nucleic acids in the sample. In one embodiment, the sample is any one of a biological sample, a clinical sample, a forensic sample or an environmental sample. In one embodiment, step (c) comprises reverse-transcribing the product of step (b) into cDNA. In one embodiment, step (c) comprises removing the host nucleic acids by size-exclusion. In one embodiment, step (c) comprises removing the host nucleic acids with the use of biotin. In one embodiment, the sample is selected from whole blood, plasma, serum, tears, saliva, mucous, cerebrospinal fluid, teeth, bone, fingernails, feces, urine, tissue, and a biopsy.

**[0013]** In another aspect, the invention provides a method of using a nucleic acid-guided RNA endonuclease to enrich for a target in an RNA sample using labeled, catalytically dead nucleic acid-guided RNA endonuclease protein. In some embodiments, the nucleic acid-guided RNA endonuclease protein is targeted to HIV RNA in a blood RNA sample, and host RNA is washed away. In some embodiments, nucleic acid-guided RNA endonuclease is C2c2.

**[0014]** In another aspect, the invention provides a composition comprising a nucleic acid fragment, a nucleic acid-guided nuclease nickase-gNA complex, and labeled nucleotides. In one embodiment, the nucleic acid comprises DNA. In one embodiment, the nucleic acid-guided nuclease nickase is selected from the group consisting of CAS Class I Type I nickase, CAS Class I Type III nickase, CAS Class I Type IV nickase, CAS Class II Type II nickase, and CAS

Class II Type V nickase. In one embodiment, the nucleic acid-guided nuclease nickase is selected from the group consisting of Cas9 nickase, Cpf1 nickase, Cas3 nickase, Cas8a-c nickase, Cas10 nickase, Cse1 nickase, Csy1 nickase, Csn2 nickase, Cas4 nickase, Csm2 nickase, Cm5 nickase, Csf1 nickase, C2C2 nickase, and NgAgo nickase. In one embodiment, the gNAs are gRNAs. In one embodiment, the gNAs are gDNAs. In one embodiment, the nucleic acid fragment comprises DNA. In one embodiment, the nucleic acid fragment comprises RNA. In one embodiment, the nucleotides are labeled with biotin. In one embodiment, the nucleotides are part of an antibody conjugate pair.

**[0015]** In another aspect, the invention provides a composition comprising a nucleic acid fragment and a catalytically dead nucleic acid-guided nuclease-gNA complex, wherein the catalytically dead nucleic acid-guided nuclease is fused to a transposase. In one embodiment, the catalytically-dead nucleic acid-guided nuclease is selected from the group consisting of dead CAS Class I Type I, dead CAS Class I Type III, dead CAS Class I Type IV, dead CAS Class II Type II, and dead CAS Class II Type V. In one embodiment, the catalytically dead nucleic acid-guided nuclease is selected from the group consisting of dCas9, dCpf1, dCas3, dCas8a-c, dCas10, dCse1, dCsy1, dCsn2, dCas4, dCsm2, dCm5, dCsf1, dC2C2, and dNgAgo. In one embodiment, the gNAs are gRNAs. In one embodiment, the gNAs are gDNAs. In one embodiment, the nucleic acid fragment comprises DNA. In one embodiment, the nucleic acid fragment comprises RNA. In one embodiment, the catalytically dead nucleic acid-guided nuclease is fused to the N-terminus of the transposase. In one embodiment, the catalytically dead nucleic acid-guided nuclease is fused to the C-terminus of the transposase. In one embodiment, the composition comprises a DNA fragment and a dCas9-gRNA complex, wherein the dCas9 is fused to a transposase.

**[0016]** In another aspect, the invention provides a composition comprising a nucleic acid fragment comprising methylated nucleotides, a nucleic acid-guided nuclease nickase-gNA complex, and unmethylated nucleotides. In one embodiment, the nucleic acid-guided nuclease nickase is selected from the group consisting of CAS Class I Type I nickase, CAS Class I Type III nickase, CAS Class I Type IV nickase, CAS Class II Type II nickase, and CAS Class II Type V nickase. In one embodiment, the nucleic acid-guided nuclease nickase is selected from the group consisting of Cas9 nickase, Cpf1 nickase, Cas3 nickase, Cas8a-c nickase, Cas10 nickase, Cse1 nickase, Csy1 nickase, Csn2 nickase, Cas4 nickase, Csm2 nickase, Cm5 nickase, Csf1 nickase, C2C2 nickase, and NgAgo nickase. In one embodiment, the gNAs are gRNAs. In one

embodiment, the gNAs are gDNAs. In one embodiment, the nucleic acid fragment comprises DNA. In one embodiment, the nucleic acid fragment comprises RNA. In one embodiment, the nucleotides are labeled with biotin. In one embodiment, the nucleotides are part of an antibody conjugate pair. In one embodiment, the composition comprises a DNA fragment comprising methylated nucleotides, a nickase Cas9-gRNA complex, and unmethylated nucleotides.

### **BRIEF DESCRIPTION OF THE DRAWINGS**

**[0017]** FIG. 1 illustrates a first protocol for the capture of target nucleic acids from a library of human genomic DNA.

**[0018]** FIG. 2 further illustrates the protocol for the capture of target nucleic acids (e.g. DNA) from a nucleic acid mixture. Target nucleic acid is cut with a nucleic acid-guided nuclease, following which adapters are ligated into the newly available blunt ends.

**[0019]** FIG. 3 illustrates that Cas9 cutting, followed by ligation of adapters, allows for specific amplification of target DNA.

**[0020]** FIG. 4 illustrates that upon sequencing the amplified DNA, ligation of adapters occurred only at the location specified by the guide RNA.

**[0021]** FIG. 5 illustrates that the method of FIG. 1 efficiently amplifies DNA that is under-represented in any given library.

**[0022]** FIG. 6 illustrates a second protocol for capture: the use of a nucleic acid-guided nuclease nickase to label target nucleic acids (e.g. DNA), allowing for further capture and purification.

**[0023]** FIG. 7 illustrates a proof of principle experiment using a restriction nickase as a substitute for the nucleic acid-guided nuclease nickase.

**[0024]** FIG. 8 illustrates that enrichment of test DNA by approximately 50-fold for the experiment illustrated in Fig 7 (using a Cas9-nickase).

**[0025]** FIG. 9 illustrates a third protocol for capture: the use of a catalytically dead nucleic acid-guided nuclease-transposase fusion to insert adapters in human genomic library, to allowing for enrichment of specific SNPs.

[0026] FIG. 10 illustrates a fourth protocol for capture: the use of dead nucleic acid-guided nuclease to protect targeted sites from subsequent fragmentation by a nucleic acid-guided nuclease, allowing for enrichment of regions of interest.

[0027] FIG. 11 illustrates a fifth protocol for capture: the use of a nucleic acid-guided nuclease nickase to protect and then enrich any targeted region, for example SNPs or STRs, from, for example, human genomic DNA, by replacing methylated DNA with unmethylated DNA.

[0028] FIG. 12 illustrates that the methylation of test DNA in the fifth protocol renders it susceptible to DpnI-mediated cleavage.

[0029] FIG. 13 illustrates a sixth protocol for capture: the use of a nucleic acid-guided nuclease nickase to introduce two double stranded breaks delineating a region of interest, allowing for 3' single stranded ligation of adapters and subsequent enrichment.

## DETAILED DESCRIPTION OF THE INVENTION

### Definitions

[0030] Unless defined otherwise herein, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, the preferred methods and materials are described.

[0031] The headings provided herein are not limitations of the various aspects or embodiments of the invention. Accordingly, the terms defined immediately below are more fully defined by reference to the specification as a whole.

[0032] Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Singleton, et al., *DICTIONARY OF MICROBIOLOGY AND MOLECULAR BIOLOGY*, 2D ED., John Wiley and Sons, New York (1994), and Hale & Markham, *THE HARPER COLLINS DICTIONARY OF BIOLOGY*, Harper Perennial, N.Y. (1991) provide one of skill with the general meaning of many of the terms used herein. Still, certain terms are defined below for the sake of clarity and ease of reference.

[0033] Numeric ranges are inclusive of the numbers defining the range.

[0034] The term “sample” as used herein relates to a material or mixture of materials, typically, although not necessarily, in liquid form, containing one or more analytes of interest.

[0035] The term “nucleic acid sample,” as used herein denotes a sample containing nucleic acids. Nucleic acid samples used herein may be complex in that they contain multiple different molecules that contain sequences. Genomic DNA from a mammal is a type of a complex sample. Complex samples may have more than  $10^4$ ,  $10^5$ ,  $10^6$  or  $10^7$  different nucleic acid molecules. A DNA target may originate from any source such as genomic DNA, cDNA, or an artificial DNA construct. Any sample containing nucleic acid, e.g., genomic DNA made from tissue culture cells or a sample of tissue, may be employed herein.

[0036] The term “nucleotide” is intended to include those moieties that contain not only the known purine and pyrimidine bases, but also other heterocyclic bases that have been modified. Such modifications include methylated purines or pyrimidines, acylated purines or pyrimidines, alkylated riboses or other heterocycles. In addition, the term “nucleotide” includes those moieties that contain hapten or fluorescent labels and may contain not only conventional ribose and deoxyribose sugars, but other sugars as well. Modified nucleosides or nucleotides also include modifications on the sugar moiety, e.g., wherein one or more of the hydroxyl groups are replaced with halogen atoms or aliphatic groups, or are functionalized as ethers, amines, or the like.

[0037] The term “nucleic acids” and “polynucleotides” are used interchangeably herein. Polynucleotide is used to describe a nucleic acid polymer of any length, e.g., greater than about 2 bases, greater than about 10 bases, greater than about 100 bases, greater than about 500 bases, greater than 1000 bases, up to about 10,000 or more bases composed of nucleotides, e.g., deoxyribonucleotides or ribonucleotides, and may be produced enzymatically or synthetically (e.g., PNA as described in U.S. Patent No. 5,948,902 and the references cited therein) which can hybridize with naturally occurring nucleic acids in a sequence specific manner analogous to that of two naturally occurring nucleic acids, e.g., can participate in Watson-Crick base pairing interactions. Naturally-occurring nucleotides include guanine, cytosine, adenine and thymine (G, C, A and T, respectively). DNA and RNA have a deoxyribose and ribose sugar backbones, respectively, whereas PNA's backbone is composed of repeating N-(2-aminoethyl)-glycine units linked by peptide bonds. In PNA various purine and pyrimidine bases are linked to the backbone

by methylene carbonyl bonds. A locked nucleic acid (LNA), often referred to as inaccessible RNA, is a modified RNA nucleotide. The ribose moiety of an LNA nucleotide is modified with an extra bridge connecting the 2' oxygen and 4' carbon. The bridge "locks" the ribose in the 3'-endo (North) conformation, which is often found in the A-form duplexes. LNA nucleotides can be mixed with DNA or RNA residues in the oligonucleotide whenever desired. The term "unstructured nucleic acid," or "UNA," is a nucleic acid containing non-natural nucleotides that bind to each other with reduced stability. For example, an unstructured nucleic acid may contain a G' residue and a C' residue, where these residues correspond to non-naturally occurring forms, i.e., analogs, of G and C that base pair with each other with reduced stability, but retain an ability to base pair with naturally occurring C and G residues, respectively. Unstructured nucleic acid is described in US20050233340, which is incorporated by reference herein for disclosure of UNA.

**[0038]** The term "oligonucleotide" as used herein denotes a single-stranded multimer of nucleotides.

**[0039]** Unless otherwise indicated, nucleic acids are written left to right in 5' to 3' orientation; amino acid sequences are written left to right in amino to carboxy orientation, respectively.

**[0040]** The term "cleaving," as used herein, refers to a reaction that breaks the phosphodiester bonds between two adjacent nucleotides in both strands of a double-stranded DNA molecule, thereby resulting in a double-stranded break in the DNA molecule.

**[0041]** The term "cleavage site, as used herein, refers to the site at which a double-stranded DNA molecule has been cleaved.

**[0042]** The "nucleic acid-guided nuclease-gNA complex" refers to a complex comprising a nucleic acid-guided nuclease protein and a guide nucleic acid (gNA, for example a gRNA or a gDNA). For example the "Cas9-gRNA complex" refers to a complex comprising a Cas9 protein and a guide RNA (gRNA). The nucleic acid-guided nuclease may be any type of nucleic acid-guided nuclease, including but not limited to wild type nucleic acid-guided nuclease, a catalytically dead nucleic acid-guided nuclease, or a nucleic acid-guided nuclease-nickase.

**[0043]** The term "nucleic acid-guided nuclease-associated guide NA" refers to a guide nucleic acid (guide NA). The nucleic acid-guided nuclease-associated guide NA may exist as an



isolated nucleic acid, or as part of a nucleic acid-guided nuclease-gNA complex, for example a Cas9-gRNA complex.

**[0044]** The terms “capture” and “enrichment” are used interchangeably herein, and refer to the process of selectively isolating a nucleic acid region containing: sequences of interest, targeted sites of interest, sequences not of interest, or targeted sites not of interest.

**[0045]** The term “hybridization” refers to the process by which a strand of nucleic acid joins with a complementary strand through base pairing as known in the art. A nucleic acid is considered to be “selectively hybridizable” to a reference nucleic acid sequence if the two sequences specifically hybridize to one another under moderate to high stringency hybridization and wash conditions. Moderate and high stringency hybridization conditions are known (see, e.g., Ausubel, et al., *Short Protocols in Molecular Biology*, 3rd ed., Wiley & Sons 1995 and Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Third Edition, 2001 Cold Spring Harbor, N.Y.). One example of high stringency conditions includes hybridization at about 42 °C in 50% formamide, 5X SSC, 5X Denhardt’s solution, 0.5% SDS and 100 µg/ml denatured carrier DNA followed by washing two times in 2X SSC and 0.5% SDS at room temperature and two additional times in 0.1 X SSC and 0.5% SDS at 42 °C.

**[0046]** The term “duplex,” or “duplexed,” as used herein, describes two complementary polynucleotides that are base-paired, i.e., hybridized together.

**[0047]** The term “amplifying” as used herein refers to generating one or more copies of a target nucleic acid, using the target nucleic acid as a template.

**[0048]** The term “genomic region,” as used herein, refers to a region of a genome, e.g., an animal or plant genome such as the genome of a human, monkey, rat, fish or insect or plant. In certain cases, an oligonucleotide used in the method described herein may be designed using a reference genomic region, i.e., a genomic region of known nucleotide sequence, e.g., a chromosomal region whose sequence is deposited at NCBI’s Genbank database or other databases, for example.

**[0049]** The term “genomic sequence,” as used herein, refers to a sequence that occurs in a genome. Because RNAs are transcribed from a genome, this term encompasses sequence that exist in the nuclear genome of an organism, as well as sequences that are present in a cDNA copy of an RNA (e.g., an mRNA) transcribed from such a genome.

[0050] The term “genomic fragment,” as used herein, refers to a region of a genome, e.g., an animal or plant genome such as the genome of a human, monkey, rat, fish or insect or plant. A genomic fragment may be an entire chromosome, or a fragment of a chromosome. A genomic fragment may be adapter ligated (in which case it has an adapter ligated to one or both ends of the fragment, or to at least the 5' end of a molecule), or may not be adapter ligated.

[0051] In certain cases, an oligonucleotide used in the method described herein may be designed using a reference genomic region, i.e., a genomic region of known nucleotide sequence, e.g., a chromosomal region whose sequence is deposited at NCBI's Genbank database or other databases, for example. Such an oligonucleotide may be employed in an assay that uses a sample containing a test genome, where the test genome contains a binding site for the oligonucleotide.

[0052] The term “ligating,” as used herein, refers to the enzymatically catalyzed joining of the terminal nucleotide at the 5' end of a first DNA molecule to the terminal nucleotide at the 3' end of a second DNA molecule.

[0053] If two nucleic acids are “complementary,” each base of one of the nucleic acids base pairs with corresponding nucleotides in the other nucleic acid. The term “complementary” and “perfectly complementary” are used synonymously herein.

[0054] The term “separating,” as used herein, refers to physical separation of two elements (e.g., by size or affinity, etc.) as well as degradation of one element, leaving the other intact. For example, size exclusion can be employed to separate nucleic acids, including cleaved targeted sequences.

[0055] In a cell, DNA usually exists in a double-stranded form, and as such, has two complementary strands of nucleic acid referred to herein as the “top” and “bottom” strands. In certain cases, complementary strands of a chromosomal region may be referred to as “plus” and “minus” strands, the “first” and “second” strands, the “coding” and “noncoding” strands, the “Watson” and “Crick” strands or the “sense” and “antisense” strands. The assignment of a strand as being a top or bottom strand is arbitrary and does not imply any particular orientation, function or structure. Until they become covalently linked, the first and second strands are distinct molecules. For ease of description, the “top” and “bottom” strands of a double-stranded nucleic acid in which the top and bottom strands have been covalently linked will still be described as the “top” and “bottom” strands. In other words, for the purposes of this disclosure,

the top and bottom strands of a double-stranded DNA do not need to be separated molecules. The nucleotide sequences of the first strand of several exemplary mammalian chromosomal regions (e.g., BACs, assemblies, chromosomes, etc.) is known, and may be found in NCBI's Genbank database, for example.

**[0056]** The term “top strand,” as used herein, refers to either strand of a nucleic acid but not both strands of a nucleic acid. When an oligonucleotide or a primer binds or anneals “only to a top strand,” it binds to only one strand but not the other. The term “bottom strand,” as used herein, refers to the strand that is complementary to the “top strand.” When an oligonucleotide binds or anneals “only to one strand,” it binds to only one strand, e.g., the first or second strand, but not the other strand. If an oligonucleotide binds or anneals to both strands of a double-stranded DNA, the oligonucleotide may have two regions, a first region that hybridizes with the top strand of the double-stranded DNA, and a second region that hybridizes with the bottom strand of the double-stranded DNA.

**[0057]** The term “double-stranded DNA molecule” refers to both double-stranded DNA molecules in which the top and bottom strands are not covalently linked, as well as double-stranded DNA molecules in which the top and bottom strands are covalently linked. The top and bottom strands of a double-stranded DNA are base paired with one other by Watson-Crick interactions.

**[0058]** The term “denaturing,” as used herein, refers to the separation of at least a portion of the base pairs of a nucleic acid duplex by placing the duplex in suitable denaturing conditions. Denaturing conditions are well known in the art. In one embodiment, in order to denature a nucleic acid duplex, the duplex may be exposed to a temperature that is above the  $T_m$  of the duplex, thereby releasing one strand of the duplex from the other. In certain embodiments, a nucleic acid may be denatured by exposing it to a temperature of at least 90 °C for a suitable amount of time (e.g., at least 30 seconds, up to 30 mins). In certain embodiments, fully denaturing conditions may be used to completely separate the base pairs of the duplex. In other embodiments, partially denaturing conditions (e.g., with a lower temperature than fully denaturing conditions) may be used to separate the base pairs of certain parts of the duplex (e.g., regions enriched for A-T base pairs may separate while regions enriched for G-C base pairs may remain paired). Nucleic acid may also be denatured chemically (e.g., using urea or NaOH).

[0059] The term “genotyping,” as used herein, refers to any type of analysis of a nucleic acid sequence, and includes sequencing, polymorphism (SNP) analysis, and analysis to identify rearrangements.

[0060] The term “sequencing,” as used herein, refers to a method by which the identity of consecutive nucleotides of a polynucleotide are obtained.

[0061] The term “next-generation sequencing” refers to the so-called parallelized sequencing-by-synthesis or sequencing-by-ligation platforms, for example, those currently employed by Illumina, Life Technologies, and Roche, etc. Next-generation sequencing methods may also include nanopore sequencing methods or electronic-detection based methods such as Ion Torrent technology commercialized by Life Technologies.

[0062] The term “complementary DNA” or cDNA refers to a double-stranded DNA sample that was produced from an RNA sample by reverse transcription of RNA (using primers such as random hexamers or oligo-dT primers) followed by second-strand synthesis by digestion of the RNA with RNaseH and synthesis by DNA polymerase.

[0063] The term “RNA promoter adapter” is an adapter that contains a promoter for a bacteriophage RNA polymerase, e.g., the RNA polymerase from bacteriophage T3, T7, SP6 or the like.

[0064] Other definitions of terms may appear throughout the specification.

#### Exemplary Methods of the Invention

[0065] As described herein, the invention provides exemplary protocols for the capture of nucleic acids and compositions for use in these protocols. Exemplary protocols are illustrated in FIGS. 1, 6, 9, 10, 11, and 13, respectively, and in the Examples section. Various uses are contemplated throughout. Specific terms referred to in this section are described in greater detail in subsequent sections.

[0066] In one embodiment, the invention provides a capture method (depicted as Protocol 1) as provided in FIGS. 1-5. In this embodiment, the method is used to capture target nucleic acid sequences. Referring to FIG. 1, the method comprises providing a sample or a library **100**, subject to extraction protocols **101** (e.g. DNA extraction protocols), resulting in a sample **102** comprising >99% non-target sequence and <1% target sequence. The sample is subjected to

library construction protocols **103**, resulting in a nucleic acid library comprising sequencing indexed adapters **104**, resulting in a plurality of adapter-ligated nucleic acids, wherein the nucleic acids are ligated to a first adapter at one end, and a second adapter at the other end. To produce a library of target-specific guide NAs (e.g. gRNAs), a library of target specific gNA precursors **110**, each comprising a RNA polymerase promoter **111**, a specific-base pair region **112** (e.g. a 20 base pair region), and a stem-loop binding site for a nucleic acid-guided nuclease **113**, was subjected to *in vitro* transcription **114**, yielding a library of target-specific guide RNAs **115**. The library of target-specific guide NAs was then combined with nucleic acid-guided nuclease proteins **116** to yield a library of nucleic acid-guided nuclease-gNA complexes. The nucleic acid-guided nuclease-gNA complexes were then combined with the nucleic acid library such that the nucleic acid-guided nuclease cleaved matching target nucleic acid sequences and left other nucleic acids uncleaved **117**. Second adapters **118** were added and allowed to ligate specifically to the 5' phosphorylated blunt ends of cleaved nucleic acids **119**. This allows for downstream applications, for example amplifying the nucleic acid fragments comprising a first or second adapter at one end and a third adapter at the other end, using adapter-specific PCR **120**.

[0067] In an exemplary depiction of Protocol 1, referring to FIG. 1, the method is used to capture target nucleic acid sequences. The method comprises providing a sample or a library comprising a plurality of adapter-ligated nucleic acids, wherein the nucleic acids are ligated to a first adapter at one end, and a second adapter at the other end. This is followed by contacting the sample with a plurality of Cas9-gRNA complexes, wherein the gRNAs are complementary to targeted sites of interest contained in a subset of the nucleic acids. The contacting cleaves the targeted sites of interest, thereby generating a plurality of nucleic acid fragments ligated to a first or second adapter at one end and no adapter at the other end. This step is followed by ligating the plurality of resulting nucleic acid fragments with third adapters, thereby generating a plurality of nucleic acid fragments ligated to a first or second adapter at one end and a third adapter at the other end. This allows for downstream applications, for example amplifying the nucleic acid fragments comprising a first or second adapter at one end and a third adapter at the other end, using adapter-specific PCR.

[0068] In one embodiment, the invention provides a capture method (depicted as Protocol 2) as provided in FIGS. 6-8. In this embodiment, the method is used to introduce labeled nucleotides at targeted sites of interest. The method comprises providing a sample comprising a plurality of

double stranded nucleic acid fragments **601** (e.g. double stranded DNA); contacting the sample with a plurality of nucleic acid-guided nuclease nickase-gNA complexes. The nickase nucleic acid-guided nuclease **603** is guided by target-specific guide NAs **604**, wherein the gNAs are complementary to targeted sites of interest in the nucleic acid fragments, thereby generating a plurality of nicked nucleic acid fragments at the targeted sites of interest. Nickase is used to nick target sequences **605**. The nickase-nucleic acid-guided nuclease cleaves at target sequence. Single strand cuts (nicks) are a substrate for DNA polymerase I which can be used to replace DNA downstream of the nick with biotin labeled DNA **606**.

[0069] In an exemplary depiction of Protocol 2, referring to FIG. 6, the method is used to introduce labeled nucleotides at targeted sites of interest. The method comprises providing a sample comprising a plurality of nucleic acid fragments; contacting the sample with a plurality of Cas9 nickase-gRNA complexes, wherein the gRNAs are complementary to targeted sites of interest in the nucleic acid fragments, thereby generating a plurality of nicked nucleic acid fragments at the targeted sites of interest; and then is followed by contacting the plurality of nicked nucleic acid fragments with an enzyme capable of initiating nucleic acid synthesis at a nicked site, and labeled nucleotides, thereby generating a plurality of nucleic acid fragments comprising labeled nucleotides in the targeted sites of interest.

[0070] In one embodiment, the invention provides a capture method (depicted as Protocol 3) as provided in FIG. 9. In this embodiment, the method is used to capture target nucleic acid sequences of interest **901**. The method first involves providing a sample comprising a plurality of adapter-ligated nucleic acids **902**, wherein the nucleic acids are ligated to a first adapter at one end and are ligated to a second adapter at the other end. This is then followed by contacting the sample with a plurality of catalytically dead nucleic acid-guided nuclease-gNA complexes, wherein the catalytically dead nucleic acid-guided nuclease is fused to a transposase **903**, wherein the gNAs **904** are complementary to targeted sites of interest contained in a subset of the nucleic acids, and wherein the catalytically dead nucleic acid-guided nuclease-gNA transposase complexes are loaded with a plurality of third adapters **905**, to generate a plurality of nucleic acids fragments **906** comprising either a first or second adapter at one end and a third adapter at the other end. These fragments can then be amplified **907** using the adapter sequences and then sequenced **908**.

[0071] In an exemplary depiction of Protocol 3, referring to FIG. 9, the method is used to capture target nucleic acid sequences of interest. The method first involves providing a sample comprising a plurality of adapter-ligated nucleic acids, wherein the nucleic acids are ligated to a first adapter at one end and are ligated to a second adapter at the other end. This is then followed by contacting the sample with a plurality of dCas9-gRNA complexes, wherein the dCas9 is fused to a transposase, wherein the gRNAs are complementary to targeted sites of interest contained in a subset of the nucleic acids, and wherein the dCas9-gRNA transposase complexes are loaded with a plurality of third adapters, to generate a plurality of nucleic acids fragments comprising either a first or second adapter at one end and a third adapter at the other end.

[0072] In one embodiment, the invention provides a capture method (depicted as Protocol 4) as provided for example in FIG. 10. In this embodiment, the method is used to capture target nucleic acid sequences of interest **1001**. The method comprises first providing a sample comprising a plurality of adapter-ligated nucleic acids **1002**, wherein the nucleic acids are ligated to the adapter at the 5' end and 3' ends. The method then involves contacting the sample with a plurality of catalytically dead nucleic acid-guided nuclease-gNA complexes **1003**, wherein the gNAs **1004** are complementary to targeted sites of interest contained in a subset of the nucleic acids, thereby generating a plurality of nucleic acids adapter-ligated at the 5' and 3' ends, bound to a catalytically dead nucleic acid-guided nuclease-gNA complex **1005**. This is followed by contacting the sample with a plurality of nucleic acid-guided nuclease-gNA complexes **1006**, wherein the gNAs **1007** are complementary to targeted sites of interest and targeted sites not of interest in the nucleic acids, thereby generating a plurality of nucleic acid fragments **1008** comprising nucleic acid sequences not of interest, adapter ligated at only one of the 5' or 3' ends. In this method wherein the second contacting step, contacting with a plurality of nucleic acid-guided nuclease-gNA complexes, does not displace the plurality of nucleic acids adapter-ligated at the 5' and 3' ends, bound to a catalytically dead nucleic acid-guided nuclease-gNA complex of step (b).

[0073] In an exemplary depiction of Protocol 4, referring to FIG. 10, the method is used to capture target nucleic acid sequences of interest. The method comprises first providing a sample comprising a plurality of adapter-ligated nucleic acids, wherein the nucleic acids are ligated to the adapter at the 5' end and 3' ends. The method then involves contacting the sample with a plurality of dead nucleic acid-guided nuclease-gNA complexes (e.g., dCas9-gRNA) complexes,

wherein the gNAs are complementary to targeted sites of interest contained in a subset of the nucleic acids, thereby generating a plurality of nucleic acids adapter-ligated at the 5' and 3' ends, bound to a dead nucleic acid-guided nuclease-gNA complex (e.g., dCAS9-gRNA complex). This is followed by contacting the sample with a plurality of nucleic acid-guided nuclease-gNA complexes (e.g., Cas9-gRNA complexes), wherein the gNAs are complementary to targeted sites of interest and targeted sites not of interest in the nucleic acids, thereby generating a plurality of nucleic acid fragments comprising nucleic acid sequences not of interest, adapter ligated at only one of the 5' or 3' ends. In this method wherein the second contacting step, contacting with a plurality of nucleic acid-guided nuclease-gNA complexes (e.g., Cas9-gRNA complexes), does not displace the plurality of nucleic acids adapter-ligated at the 5' and 3' ends, bound to a dead nucleic acid-guided nuclease-gNA complex (e.g., dCAS9-gRNA complex) of step (b).

**[0074]** In one embodiment, the invention provides a capture method (depicted as Protocol 5) as provided, for example, in FIG. 11 and FIG. 12. In this embodiment, the method is used to capture target nucleic sequences of interest **1101**. The method involves first providing a sample comprising a plurality of sequences **1102**, wherein the sequences comprise methylated nucleotides (e.g., treated with Dam methyltransferase), and wherein the sequences are adapter ligated on the 5' and 3' ends. The method then involves first contacting the sample with a plurality of nucleic acid-guided nuclease nickase-gNA complexes **1103**, wherein the gNAs **1104** are complementary to targeted sites of interest in a subset of the sequences, thereby generating a plurality of nicked nucleic acid sequences **1105** at the targeted sites of interest, and wherein the nucleic acid sequences are adapter ligated on the 5' and 3' ends. Where the nucleic acid is DNA, single strand cuts (nicks) can be, for example, a substrate for DNA polymerase I, which replaces the DNA downstream of the nick with unmethylated DNA. This can then be followed by then contacting the sample with an enzyme capable of initiating DNA synthesis at a nicked site, and unmethylated nucleotides, thereby generating a plurality of DNA comprising unmethylated nucleotides **1106** in the targeted sites of interest and wherein the DNA sequences are adapter ligated on the 5' and 3' ends. This is then followed by contacting the sample with an enzyme capable of cutting methylated DNA (e.g., DpnI) **1107**, thereby generating a plurality of DNA fragments comprising methylated DNA, wherein the plurality of DNA fragments comprising methylated DNA are adapter ligated only one of the 5' and 3' ends. Remaining intact nucleic acid can be amplified and sequenced **1108**. **FIG. 12** shows, for example, results from an experiment conducted according to this protocol. The first column on the gel shows a 1



kb ladder, the second column shows test DNA treated with Dam methyltransferase, then digested with DpnI, and the third column shows test DNA digested with DpnI. The second column shows a band corresponding to DpnI digested DNA, while the third column shows a band corresponding to uncut test DNA.

**[0075]** In an exemplary depiction of Protocol 5, referring to FIGS. 11-12, the method is used to capture target DNA sequences of interest. The method involves first providing a sample comprising a plurality of DNA sequences, wherein the DNA sequences comprise methylated nucleotides, and wherein the DNA sequences are adapter ligated on the 5' and 3' ends. The method then involves first contacting the sample with a plurality of Cas9 nickase-gRNA complexes, wherein the gRNAs are complementary to targeted sites of interest in a subset of the DNA sequences, thereby generating a plurality of nicked DNA at the targeted sites of interest, and wherein the DNA are adapter ligated on the 5' and 3' ends. This is then followed by then contacting the sample with an enzyme capable of initiating DNA synthesis at a nicked site, and unmethylated nucleotides, thereby generating a plurality of DNA comprising unmethylated nucleotides in the targeted sites of interest and wherein the DNA sequences are adapter ligated on the 5' and 3' ends. This is then followed by contacting the sample with an enzyme capable of cutting methylated DNA, thereby generating a plurality of DNA fragments comprising methylated DNA, wherein the plurality of DNA fragments comprising methylated DNA are adapter ligated only one of the 5' and 3' ends.

**[0076]** In one embodiment, the invention provides a capture method (depicted as Protocol 6) as provided in FIG. 13. The objective of this method is to enrich a region of a nucleic acid **1301**, from any source (e.g., library **1302**, genomic, or PCR), as depicted for example in FIG. 13. The nucleic acid-guided nuclease – nickase **1303** can be targeted to proximal sites using two guide NAs **1304** and **1305**, resulting in nicking of nucleic acid at each location **1306**. Alternatively, an adapter can be ligated on only one side, then filled in, then an adapter can be ligated on the other side. The two nicks can be close to each other (e.g., within 10 to 15 bp). Single nicks may be generated in non-target molecules. Because of the proximity of the two nicking sites, a double stranded break can be created when the reaction is heated **1307**, e.g. to 65 °C, resulting in long (e.g., 10-15 bp) 3' overhangs. These overhangs can be recognized by a thermostable single stranded DNA/RNA ligase, to allow for site-specific ligation **1308** of single stranded adapters. The ligase can, for example, only recognize long 3' overhangs, thus ensuring that adapters will not be ligated at other sites. This process can be repeated using nucleic acid-guided nuclease

nickase and guide NAs targeting on the other side of the region of interest, followed by ligation as above using a second single stranded adapter. Once two adapters have been ligated on either side of the region of interest, the region can be amplified or sequenced directly **1309**.

**[0077]** In an exemplary depiction of Protocol 6, referring to FIG. 13, the method is to enrich a region of DNA, from any DNA source (e.g., library, genomic, or PCR). A Cas9 Nickase can be targeted to proximal sites using two guide RNAs and, resulting in nicking of DNA at each location. Alternatively, an adapter can be ligated on only one side, then filled in, then an adapter can be ligated on the other side. The two nicks can be close to each other (e.g., within 10 to 15 bp). Single nicks may be generated in non-target molecules. Because of the proximity of the two nicking sites, a double stranded break can be created when the reaction is heated, e.g. to 65 °C, resulting in long (e.g., 10-15 bp) 3' overhangs. These overhangs can be recognized by a thermostable single stranded DNA/RNA ligase, such as a Thermostable 5' App DNA/RNA ligase to allow for site-specific ligation of single stranded adapters. The ligase can, for example, only recognize long 3' overhangs, thus ensuring that adapters will not be ligated at other sites. This process can be repeated using Cas9 Nickase and guide RNA targeting on the other side of the region of interest, followed by ligation as above using a second single stranded adapter. Once two adapters have been ligated on either side of the region of interest, the region can be amplified or sequenced directly.

**[0078]** In one embodiment, provided herein is a method of enriching a sample for sequences of interest, comprising: (a) providing a sample comprising sequences of interest and targeted sequences for depletion, wherein the sequences of interest comprise less than 50% of the sample; and (b) contacting the sample with a plurality of nucleic acid-guided RNA endonuclease-gRNA complexes or a plurality of nucleic acid-guided DNA endonuclease-gDNA complexes, wherein the gRNAs and gDNAs are complementary to the targeted sequences. In some embodiments, the targeted sequences are thereby cleaved. In one embodiment, the nucleic acid-guided RNA endonuclease is C2c2. In one embodiment the C2c2 is catalytically dead. In one embodiment, the nucleic acid-guided DNA endonuclease is NgAgo (Argonaute from *Natronobacterium gregoryi*). In one embodiment the NgAgo is catalytically dead.

**[0079]** In one embodiment, provided herein is a method of enriching a sample comprising: (a) providing a sample comprising host nucleic acids and non-host nucleic acids; (b) contacting the sample with a plurality of nucleic acid-guided RNA endonuclease-gRNA complexes or plurality

of nucleic acid-guided DNA endonuclease-gDNA complexes, wherein the gNAs are complementary to targeted sites in the host nucleic acids, and (c) enriching the sample for non-host nucleic acids. In one embodiment, the nucleic acid-guided RNA endonuclease is C2c2. In one embodiment the C2c2 is catalytically dead. In one embodiment, the nucleic acid-guided DNA endonuclease is NgAgo. In one embodiment the NgAgo is catalytically dead.

#### Nucleic Acids, Samples

**[0080]** Nucleic acids of the invention (targeted for capture) can be any DNA, any RNA, single stranded DNA, single stranded RNA, double stranded DNA, double stranded RNA, artificial DNA, artificial RNA, synthetic DNA, synthetic RNA, and RNA/DNA hybrids.

**[0081]** The nucleic acids of the invention can be a genomic fragment, comprising a region of the genome, or the whole genome itself. In one embodiment, the genome is a DNA genome. In another embodiment, the genome is a RNA genome.

**[0082]** Nucleic acids of the invention can be obtained from a eukaryotic or prokaryotic organism; from a mammalian organism or a non-mammalian organism; from an animal or a plant; from a bacteria or virus; from an animal parasite; or from a pathogen.

**[0083]** Nucleic acids of the invention can be obtained from any mammalian organism. In one embodiment the mammal is a human. In another embodiment the mammal is a livestock animal, for example a horse, a sheep, a cow, a pig, or a donkey. In another embodiment, a mammalian organism is a domestic pet, for example a cat, a dog, a gerbil, a mouse, a rat. In another embodiment the mammal is a type of a monkey.

**[0084]** Nucleic acids of the invention can be obtained from any bird or avian organism. An avian organism includes but is not limited to chicken, turkey, duck and goose.

**[0085]** Nucleic acids of the invention can be obtained from a plant. In one embodiment, the plant is rice, maize, wheat, rose, grape, coffee, fruit, tomato, potato, or cotton.

**[0086]** In some embodiments, nucleic acids of the invention are obtained from a species of bacteria. In one embodiment, the bacteria are tuberculosis-causing bacteria.

**[0087]** In some embodiments, nucleic acids of the invention are obtained from a virus.

[0088] In some embodiments, nucleic acids of the invention are obtained from a species of fungi.

[0089] In some embodiments, nucleic acids of the invention are obtained from a species of algae.

[0090] In some embodiments, nucleic acids of the invention are obtained from any mammalian parasite.

[0091] In some embodiments, nucleic acids of the invention are obtained from any mammalian parasite. In one embodiment, the parasite is a worm. In another embodiment, the parasite is a malaria-causing parasite. In another embodiment, the parasite is a Leishmaniasis-causing parasite. In another embodiment, the parasite is an amoeba.

[0092] In some embodiments the pathogen is a non-mammalian pathogen (is pathogenic in non-mammalian organisms).

[0093] In one embodiment, the nucleic acids of the invention include nucleic acids that are targets of gNAs and nucleic acids that are not the targets of gNAs, in the same sample.

[0094] In one embodiment, the nucleic acids of the invention include nucleic acids that are targets of gRNAs and nucleic acids that are not the targets of gRNAs, in the same sample.

[0095] In one embodiment, the nucleic acids of the invention include nucleic acids that are targets of gDNAs and nucleic acids that are not the targets of gDNAs, in the same sample.

[0096] In one embodiment, the nucleic acids of the invention include target nucleic acids (targets of gNAs) and nucleic acids of interest (not targeted by gNAs) from a sample.

[0097] In one embodiment, the nucleic acids of the invention include target nucleic acids (targets of gRNAs) and nucleic acids of interest (not targeted by gRNAs) from a sample.

[0098] In one embodiment, the nucleic acids of the invention include target nucleic acids (targets of gDNAs) and nucleic acids of interest (not targeted by gDNAs) from a sample.

[0099] In one embodiment, the target DNA (target of the gNAs, gRNAs, gDNAs) may be human non-mitochondrial DNA (e.g. genomic DNA), and the DNA of interest (for capture) may

be the human mitochondrial DNA, and the human mitochondrial DNA is enriched by targeting the non-mitochondrial human DNA.

**[0100]** In one embodiment, the nucleic acids to be captured may be a non-mapable region of a genome; and the nucleic acids to be retained for further analysis/sequencing/cloning may be mapable regions of a genome. In one embodiment, the nucleic acids to be captured out may be a mapable region of a genome; and the nucleic acids to be retained for further analysis/sequencing/cloning may be non-mapable regions of a genome. Examples of non mapable regions include telomeres, centromeres, or other genomic regions that contain features harder to map.

**[0101]** In one embodiment, the nucleic acids of the invention are obtained from a biological sample. The biological sample from which the nucleic acids are obtained include but are not limited to whole blood, plasma, serum, tears, saliva, mucous, cerebrospinal fluid, teeth, bone, fingernails, feces, urine, tissue, and biopsy. The biological sample may include forensic samples such as teeth, bone, fingernails or the like. The biological sample may include tissue, a tissue biopsy, for example a resected lung tissue. The biological sample may include a clinical sample, which refers to a sample obtained in a clinical setting, such as in a hospital, or clinic.

**[0102]** In one embodiment, the nucleic acids of the invention are obtained from an environmental sample, for example from water, soil, air, or rock.

**[0103]** In one embodiment, the nucleic acids of the invention are obtained from a forensic sample, for example, a sample obtained from an individual at a crime scene, from a piece of evidence, post-mortem, as a part of an ongoing investigation or the like.

**[0104]** In one embodiment, the nucleic acids of the invention are provided in a library.

**[0105]** The nucleic acids of the invention can be either provided or extracted from a sample. Extraction can extract substantially all the nucleic acid sequences from a specimen.

**[0106]** The methods of the invention may produce nucleic acids to be captured: nucleic acids to not be captured at a ratio of anywhere between 99.999:0.001 to 0.001:99.999. The methods of the invention may produce targeted nucleic acids and nucleic acids of interest at a ratio of anywhere between 99.999:0.001 to 0.001:99.999. The methods of the invention may produce nucleic acids to be captured to nucleic acids to be retained/analyzed/sequenced at a ratio of

anywhere between 99.999:0.001 to 0.001:99.999. In these embodiments, the ratios can be equal to or fall anywhere in between 99.999:0.001 to 0.001:99.999, for example the ratio can be 99:1, 95:5, 90:10, 85:15, 80:20, 75:25, 70:30, 65:35, 60:40, 55:45, 50:50, 45:55, 40:60, 35:65, 30:70, 25:75, 20:80, 15:85, 10:90, 5:95, and 1:99.

[0107] After the capture, the captured or retained nucleic sequences can be fragmented to reduce the lengths of each extracted nucleic acids to a more manageable length for amplifying, sequencing or the like.

[0108] As provided herein, at least 10%, 15%, 20%, 25%, 30%, 40%, 50%, 60%, 70%, 80%, 90% of the starting nucleic acid material can be captured. This capture can be achieved in no greater than 10 minutes, 15 minutes, 20 minutes, 30 minutes, 45 minutes, 60 minutes, 75 minutes, 90 minutes, 105 minutes, 120 minutes, 150 minutes, 180 minutes, or 240 minutes.

[0109] In some cases, the targeted sites of interest represent less than 90%, less than 80%, less than 70%, less than 60%, less than 50%, less than 40%, less than 30%, less than 20%, or less than 10% of the total DNA in the sample.

#### Adapters

[0110] As provided herein, the nucleic acids of the invention (referred to interchangeably as nucleic acids or nucleic acid fragments) are adapter-ligated, to aid in carrying out the methods provided herein.

[0111] Nucleic acids of the invention to be adapter-ligated can range from 20 bp in size to 5000 bp in size. For example, the nucleic acid to be adapter-ligated may be at least 20, 25, 50, 75, 100, 125, 150, 175, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1500, 2000, 2500, 3000, 3500, 4000, 4500, or 5000 bp. In one specific embodiment, the nucleic acid to be adapter ligated is 100 bp. In one specific embodiment, the nucleic acid to be adapter ligated is 200bp. In one specific embodiment, the nucleic acid to be adapter ligated is 300 bp. In one specific embodiment, the nucleic acid to be adapter ligated is 400 bp. In one specific embodiment, the nucleic acid to be adapter ligated is 500 bp.

[0112] An adapter can be ligated to each end of each of the nucleic acids, or nucleic acid fragments, at the 5' and 3' ends. In other embodiments an adapter may be ligated to only one end of each of the fragments or in other instances adapters may be ligated in a later step. In

one example the adapter is a nucleic acid that is ligatable to both strands of a double-stranded DNA molecule. In various embodiments the adapter may be a hairpin adapter e.g., one molecule that base pairs with itself to form a structure that has a double-stranded stem and a loop, where the 3' and 5' ends of the molecule ligate to the 5' and 3' ends of the double-stranded DNA molecule of the fragment, respectively. Alternately, the adapter may be a Y-adapter ligated to one end or to both ends of a fragment, also called a universal adapter. Alternately, the adapter may itself be composed of two distinct oligonucleotide molecules that are base paired with one another. Additionally, a ligatable end of the adapter may be designed to be compatible with overhangs made by cleavage by a restriction enzyme, or it may have blunt ends or a 5' T overhang. Generally, the adapter may include double-stranded as well as single-stranded molecules. Thus the adapter can be DNA or RNA, or a mixture of the two. Adapters containing RNA may be cleavable by RNase treatment or by alkaline hydrolysis.

**[0113]** Adapters can be 10 to 100 bp in length although adapters outside of this range are usable without deviating from the present invention. In specific embodiments, the adapter is at least 10 bp, at least 15 bp, at least 20 bp, at least 25 bp, at least 30 bp, at least 35 bp, at least 40 bp, at least 45 bp, at least 50 bp, at least 55 bp, at least 60 bp, at least 65 bp, at least 70 bp, at least 75 bp, at least 80 bp, at least 85 bp, at least 90 bp, or at least 95 bp in length.

**[0114]** In further examples the captured nucleic acid sequences may be derived from one or more DNA sequencing libraries. An adapter may be configured for a next generation sequencing platform, for example for use on an Illumina sequencing platform or for use on an IonTorrent platform.

**[0115]** An adapter may contain a restriction site of interest or a primer binding site.

**[0116]** Exemplary adapters include P5 and P7 adapters.

#### Guide Nucleic Acids (gNAs)

**[0117]** Provided herein are guide nucleic acids (gNAs), wherein the gNAs are complementary to (selective for, can hybridize with) targeted sites or sequences of interest, or sequences not of interest in the nucleic acids, for example in genomic DNA from a host. The gNAs guide nucleic acid-guided nucleases to specific sites on a nucleic acid.

[0118] In some embodiments, the gNAs are guide RNAs (gRNAs); in other embodiments, the gNAs are guide DNAs (gDNAs). In some embodiments the gNAs comprise a mixture of gRNAs and gDNAs.

[0119] The host to which the gNAs are directed to can be an animal, for example a human, cow, horse, sheep, pig, monkey, dog, cat, gerbil, bird, mouse, or rat. The host can be a plant. The non-host can be a prokaryotic organism, a eukaryote, a virus, a bacterium, a fungus, and a protozoan.

[0120] In one embodiment, the present invention provides a guide nucleic acid (gNA) library which comprises a collection of gNAs, configured to hybridize with a nucleic acid sequence targeted for capture. In another embodiment, the present invention provides a guide NA library which comprises a collection of gNAs, configured to hybridize with a nucleic acid sequence that is not targeted for capture.

[0121] In one embodiment, the present invention provides a guide RNA library which comprises a collection of gRNAs, configured to hybridize with a nucleic acid sequence targeted for capture. In another embodiment, the present invention provides a guide RNA library which comprises a collection of gRNAs, configured to hybridize with a nucleic acid sequence that is not targeted for capture.

[0122] In one embodiment, the present invention provides a guide DNA library which comprises a collection of gDNAs, configured to hybridize with a nucleic acid sequence targeted for capture. In another embodiment, the present invention provides a guide DNA library which comprises a collection of gDNAs, configured to hybridize with a nucleic acid sequence that is not targeted for capture.

[0123] In one embodiment, the gNAs are selective for target nucleic acids in a sample, but are not selective for sequences of interest from the sample.

[0124] In one embodiment, the gNAs are used to serially capture nucleic acid sequences.

[0125] In some embodiments, the gNAs are selective for a target nucleic acid sequences which are followed by Protospacer Adjacent Motif (PAM) sequences that can be bound by a nucleic acid-guided nuclease. In some embodiments, the sequence of the gNAs is determined by the nucleic acid-guided nuclease type. In various embodiments the gNAs may be tailored to



different nucleic acid-guided nuclease types as the PAM sequence can vary by the species of the organism from which the nucleic acid-guided nuclease is derived.

**[0126]** The gNAs (gRNAs or gDNAs) of the invention can range in size from 50-200 base pairs. For example, a gNA of the invention can be at least 50bp, 55bp, 60bp, 65bp, 70bp, 75bp, 80bp, 85bp, 90bp, 95bp, 100bp, 110bp, 120bp, 125bp, 130bp, 140bp, 150bp, 160bp, 170bp, 175bp, 180bp, 190bp, or 195 bp. In specific embodiments, the gNA is 80bp, 90bp, 100bp, or 110bp. In some embodiments, a target-specific gNA comprises a base pair sequence can be complementary to a pre-defined site in a target nucleic acid that is followed by a Protospacer Adjacent Motif or (PAM) sequence that can be bound by a nucleic acid-guided nuclease protein (e.g. Cas9) derived from a bacterial species. In specific embodiments, the base pair sequence of the gNA that is complementary to a pre-defined site in a target nucleic acid is 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, or 50 base pairs.

**[0127]** The present invention also provides for gNA libraries (e.g. gRNA libraries or gDNA libraries). A gNA library can comprise a number of different species- specific guide NA (e.g. gRNA or gDNA) elements each, configured to hybridize with (be selective for) a nucleic acid sequence being targeted capture, a nucleic acid sequence of interest, or a nucleic acid sequence not of interest. Each gNA includes a target-specific guide sequence and a stem loop binding site that is formed to bind with a nucleic acid-guided nuclease protein. In some embodiments, the library can comprise a plurality of different guide NAs, each having a different 15-30 base pair sequence that is complementary to a different pre-defined site in the nucleic acid being targeted, that is followed by an appropriate PAM sequence that can be bound by a nucleic acid-guided nuclease protein. For each guide NA the PAM sequence is present in the pre-defined DNA or RNA target sequence of the nucleic acid of interest but is not present in the corresponding target specific guide sequence.

**[0128]** Generally according to the present invention, any nucleic acid sequence in a genome of interest, with a pre-defined target sequence followed by the appropriate PAM sequence can be hybridized by a corresponding guide RNA provided in the guide NA library and bound by a nucleic acid-guided nuclease. In various embodiments the gNA library may be tailored to different nucleic acid-guided nuclease types since the PAM sequence can vary by the species of the bacteria from which nucleic acid-guided nuclease is derived. But in some variations, the pre-defined target sequence is not followed by a PAM sequence.

[0129] Different target-specific sequences in the gNAs can be generated. This can be done by using a promoter for a bacteriophage RNA polymerase, e.g., the RNA polymerase from bacteriophage T3, T7, SP6 or the like. Accordingly, each different T7 RNA polymerase promoter provides a different target specific sequence suitable for hybridizing to a different target nucleic acid sequence. A non-limiting exemplary set of forward primers usable for both annealing and subsequent PCR reactions is listed in Table 1 provided below.

[0130] A gNA library (e.g. a gRNA or gDNA library) can be amplified to include a large number of copies of each different guide NA element as well as a large number of different guide NA elements as may be suitable to for the desired capture results. The number of unique guide NA elements in a given guide NA library may range from 1 unique guide NA element to as many as 300,000,000 unique guide NA elements, or approximately 1 unique guide NA sequence for every 10 base pairs in the human genome. The number of unique gNAs (e.g., gRNAs or gDNAs) can be at least about 101, 102, 103, 104, 105, 106, 107, or 108 unique gNAs. The number of unique gNAs can result in that number of unique nucleic acid-guided nuclease – gNA complexes (e.g. CRISPR/Cas system protein-gRNA complexes).

[0131] Without being limited to theory, the distance between gNAs to arrive at >95% cleavage of the target nucleic acid can be computed, if the gNAs display ~100% efficacy: this can be computed by measuring the distribution of library size and determining the mean, N and the standard deviation SD;  $N - 2SD$  = minimum size for >95% of the library, ensuring that there is one guide NA per fragment of this size to ensure >95% capture. This can also be described as the Maximum distance between guide NAs = Mean of library size -  $2 \times$  (standard deviation of library size).

[0132] In various embodiments of the invention, the gNAs can be specific for various targeted sites of interest, including, but not limited to, single nucleotide polymorphisms (SNPs), short tandem repeats (STRs), cancer genes, inserts, deletions, structural variations, exons, genetic mutations, and regulatory regions.

#### Nucleic Acid-Guided Nucleases

[0133] Provided herein compositions and methods for the capture of nucleic acids from a sample. These compositions and methods utilize nucleic acid-guided nucleases. As used herein, a “nucleic acid-guided nuclease” is any endonuclease that cleaves DNA, RNA or

DNA/RNA hybrids, and which uses one or more nucleic acid guide nucleic acids (gNAs) to confer specificity. Nucleic acid-guided nucleases include CRISPR/Cas system proteins as well as non-CRISPR/Cas system proteins.

[0134] The nucleic acid-guided nucleases provided herein can be DNA guided DNA endonucleases; DNA guided RNA endonucleases; RNA guided DNA endonucleases; or RNA guided RNA endonucleases.

[0135] In one embodiment, the nucleic acid-guided nuclease is a nucleic acid-guided-DNA endonuclease.

[0136] In one embodiment, the nucleic acid-guided nuclease is a nucleic acid-guided-RNA endonuclease.

#### CRISPR/Cas System Nucleic Acid-Guided Nucleases

[0137] In some embodiments, CRISPR/Cas system proteins are used in the embodiments provided herein. In some embodiments, CRISPR/Cas system proteins include proteins from CRISPR Type I systems, CRISPR Type II systems, and CRISPR Type III systems.

[0138] In some embodiments, CRISPR/Cas system proteins can be from any bacterial or archaeal species.

[0139] In some embodiments, the CRISPR/Cas system protein is isolated, recombinantly produced, or synthetic.

[0140] In some embodiments, the CRIPR/Cas system proteins are from, or are derived from CRISPR/Cas system proteins from *Streptococcus pyogenes*, *Staphylococcus aureus*, *Neisseria meningitidis*, *Streptococcus thermophiles*, *Treponema denticola*, *Francisella tularensis*, *Pasteurella multocida*, *Campylobacter jejuni*, *Campylobacter lari*, *Mycoplasma gallisepticum*, *Nitratifactor salsuginis*, *Parvibaculum lavamentivorans*, *Roseburia intestinalis*, *Neisseria cinerea*, *Gluconacetobacter diazotrophicus*, *Azospirillum*, *Sphaerochaeta globus*, *Flavobacterium columnare*, *Fluviicola taffensis*, *Bacteroides coprophilus*, *Mycoplasma mobile*, *Lactobacillus farciminis*, *Streptococcus pasteurianus*, *Lactobacillus johnsonii*, *Staphylococcus pseudintermedius*, *Filifactor alocis*, *Legionella pneumophila*, *Suterella wadsworthensis*, or *Corynebacter diphtheria*.

[0141] In some embodiments, examples of CRISPR/Cas system proteins can be naturally occurring or engineered versions.

[0142] In some embodiments, naturally occurring CRISPR/Cas system proteins can belong to CAS Class I Type I, III, or IV, or CAS Class II Type II or V, and can include Cas9, Cas3, Cas8a-c, Cas10, Cse1, Csy1, Csn2, Cas4, Csm2, Cmr5, Csf1, C2c2, and Cpf1.

[0143] In an exemplary embodiment, the CRISPR/Cas system protein comprises Cas9.

[0144] A "CRISPR/Cas system protein-gNA complex" refers to a complex comprising a CRISPR/Cas system protein and a guide NA (e.g. a gRNA or a gDNA). Where the gNA is a gRNA, the gRNA may be composed of two molecules, i.e., one RNA ("crRNA") which hybridizes to a target and provides sequence specificity, and one RNA, the "tracrRNA", which is capable of hybridizing to the crRNA. Alternatively, the guide RNA may be a single molecule (i.e., a gRNA) that contains crRNA and tracrRNA sequences.

[0145] A CRISPR/Cas system protein may be at least 60% identical (e.g., at least 70%, at least 80%, or 90% identical, at least 95% identical or at least 98% identical or at least 99% identical) to a wild type CRISPR/Cas system protein. The CRISPR/Cas system protein may have all the functions of a wild type CRISPR/Cas system protein, or only one or some of the functions, including binding activity, nuclease activity, and nuclease activity.

[0146] The term "CRISPR/Cas system protein-associated guide NA" refers to a guide NA. The CRISPR/Cas system protein -associated guide NA may exist as isolated NA, or as part of a CRISPR/Cas system protein-gNA complex.

Cas9

[0147] In some embodiments the CRISPR/Cas System protein nucleic acid-guided nuclease is or comprises Cas9. The Cas9 of the present invention can be isolated, recombinantly produced, or synthetic.

[0148] Examples of Cas9 proteins that can be used in the embodiments herein can be found in F.A. Ran, L. Cong, W.X. Yan, D. A. Scott, J.S. Gootenberg, A.J. Kriz, B. Zetsche, O. Shalem, X. Wu, K.S. Makarova, E.V. Koonin, P.A. Sharp, and F. Zhang; "In vivo genome editing using *Staphylococcus aureus* Cas9," *Nature* 520, 186–191 (09 April 2015) doi:10.1038/nature14299, which is incorporated herein by reference.

[0149] In some embodiments, the Cas9 is a Type II CRISPR system derived from *Streptococcus pyogenes*, *Staphylococcus aureus*, *Neisseria meningitidis*, *Streptococcus thermophilus*, *Treponema denticola*, *Francisella tularensis*, *Pasteurella multocida*, *Campylobacter jejuni*, *Campylobacter lari*, *Mycoplasma gallisepticum*, *Nitratifractor salsuginis*, *Parvibaculum lavamentivorans*, *Roseburia intestinalis*, *Neisseria cinerea*, *Gluconacetobacter diazotrophicus*, *Azospirillum*, *Sphaerochaeta globus*, *Flavobacterium columnare*, *Fluviicola taffensis*, *Bacteroides coprophilus*, *Mycoplasma mobile*, *Lactobacillus farciminis*, *Streptococcus pasteurianus*, *Lactobacillus johnsonii*, *Staphylococcus pseudintermedius*, *Filifactor alocis*, *Legionella pneumophila*, *Suterella wadsworthensis*, or *Corynebacter diphtheria*.

[0150] In some embodiments, the Cas9 is a Type II CRISPR system derived from *S. pyogenes* and the PAM sequence is NGG located on the immediate 3' end of the target specific guide sequence. The PAM sequences of Type II CRISPR systems from exemplary bacterial species can also include: *Streptococcus pyogenes* (NGG), *Staph aureus* (NNGRRT), *Neisseria meningitidis* (NNNNGA TT), *Streptococcus thermophilus* (NNAGAA) and *Treponema denticola* (NAAAAC) which are all usable without deviating from the present invention.

[0151] In one exemplary embodiment, Cas9 sequence can be obtained, for example, from the pX330 plasmid (available from Addgene), re-amplified by PCR then cloned into pET30 (from EMD biosciences) to express in bacteria and purify the recombinant 6His tagged protein.

[0152] A “Cas9-gNA complex” refers to a complex comprising a Cas9 protein and a guide NA. A Cas9 protein may be at least 60% identical (e.g., at least 70%, at least 80%, or 90% identical, at least 95% identical or at least 98% identical or at least 99% identical) to a wild type Cas9 protein, e.g., to the *Streptococcus pyogenes* Cas9 protein. The Cas9 protein may have all the functions of a wild type Cas9 protein, or only one or some of the functions, including binding activity, nuclease activity, and nuclease activity.

[0153] The term “Cas9-associated guide NA” refers to a guide NA as described above. The Cas9-associated guide NA may exist isolated, or as part of a Cas9-gNA complex.

#### Non-CRISPR/Cas System Nucleic Acid-Guided Nucleases

[0154] In some embodiments, non-CRISPR/Cas system proteins are used in the embodiments provided herein.

[0155] In some embodiments, the non-CRISPR/Cas system proteins can be from any bacterial or archaeal species.

[0156] In some embodiments, the non-CRISPR /Cas system protein is isolated, recombinantly produced, or synthetic.

[0157] In some embodiments, the non-CRISPR /Cas system proteins are from, or are derived from *Aquifex aeolicus*, *Thermus thermophilus*, *Streptococcus pyogenes*, *Staphylococcus aureus*, *Neisseria meningitidis*, *Streptococcus thermophilus*, *Treponema denticola*, *Francisella tularensis*, *Pasteurella multocida*, *Campylobacter jejuni*, *Campylobacter lari*, *Mycoplasma gallisepticum*, *Nitratifractor salsuginis*, *Parvibaculum lavamentivorans*, *Roseburia intestinalis*, *Neisseria cinerea*, *Gluconacetobacter diazotrophicus*, *Azospirillum*, *Sphaerochaeta globus*, *Flavobacterium columnare*, *Fluviicola taffensis*, *Bacteroides coprophilus*, *Mycoplasma mobile*, *Lactobacillus farciminis*, *Streptococcus pasteurianus*, *Lactobacillus johnsonii*, *Staphylococcus pseudintermedius*, *Filifactor alocis*, *Legionella pneumophila*, *Suterella wadsworthensis*, *Natronobacterium gregoryi*, or *Corynebacter diphtheria*.

[0158] In some embodiments, the non-CRISPR /Cas system proteins can be naturally occurring or engineered versions.

[0159] In some embodiments, a naturally occurring non-CRISPR /Cas system protein is NgAgo (Argonaute from *Natronobacterium gregoryi*).

[0160] A "non-CRISPR /Cas system protein-gNA complex" refers to a complex comprising a non-CRISPR /Cas system protein and a guide NA (e.g. a gRNA or a gDNA). Where the gNA is a gRNA, the gRNA may be composed of two molecules, i.e., one RNA ("crRNA") which hybridizes to a target and provides sequence specificity, and one RNA, the "tracrRNA", which is capable of hybridizing to the crRNA. Alternatively, the guide RNA may be a single molecule (i.e., a gRNA) that contains crRNA and tracrRNA sequences.

[0161] A non-CRISPR /Cas system protein may be at least 60% identical (e.g., at least 70%, at least 80%, or 90% identical, at least 95% identical or at least 98% identical or at least 99% identical) to a wild type non-CRISPR /Cas system protein. The non-CRISPR /Cas system protein may have all the functions of a wild type non-CRISPR /Cas system protein, or only one or some of the functions, including binding activity, nuclease activity, and nuclease activity.

[0162] The term “non-CRISPR /Cas system protein-associated guide NA” refers to a guide NA. The non-CRISPR /Cas system protein -associated guide NA may exist as isolated NA, or as part of a non-CRISPR /Cas system protein-gNA complex.

#### Catalytically Dead Nucleic Acid-Guided Nucleases

[0163] In some embodiments, engineered examples of nucleic acid-guided nucleases include catalytically dead nucleic acid-guided nucleases (CRISPR/Cas system nucleic acid-guided nucleases or non-CRISPR/Cas system nucleic acid-guided nucleases). The term “catalytically dead” generally refers to a nucleic acid-guided nuclease that has inactivated nucleases, for example inactivated HNH and RuvC nucleases. Such a protein can bind to a target site in any nucleic acid (where the target site is determined by the guide NA), but the protein is unable to cleave or nick the nucleic acid.

[0164] Accordingly, the catalytically dead nucleic acid-guided nuclease allows separation of the mixture into unbound nucleic acids and catalytically dead nucleic acid-guided nuclease-bound fragments. Use of a dead nucleic acid-guided nuclease is depicted, for example, in FIG. 9 and FIG. 10, in Protocols 3 and 4, respectively. In one exemplary embodiment, a dCas9/gRNA complex binds to the targets determined by the gRNA sequence. The dCas9 bound can prevent cutting by Cas9 while other manipulations proceed, as pictured in FIG. 10.

[0165] In another embodiment, the catalytically dead nucleic acid-guided nuclease can be fused to another enzyme, such as a transposase, to target that enzyme’s activity to a specific site.

[0166] In some embodiments, the catalytically dead nucleic acid-guided nuclease is dCas9, dCpf1, dCas3, dCas8a-c, dCas10, dCse1, dCsy1, dCsn2, dCas4, dCsm2, dCm5, dCsf1, dC2C2, or dNgAgo.

[0167] In one exemplary embodiment the catalytically dead nucleic acid-guided nuclease protein is a dCas9.

#### Nucleic Acid-Guided Nuclease Nickases

[0168] In some embodiments, engineered examples of nucleic acid-guided nucleases include nucleic acid-guided nuclease nickases (referred to interchangeably as nickase nucleic acid-guided nucleases).

[0169] In some embodiments, engineered examples of nucleic acid-guided nucleases include CRISPR/Cas system nickases or non-CRISPR/Cas system nickases, containing a single inactive catalytic domain.

[0170] In some embodiments, the nucleic acid-guided nuclease nickase is a Cas9 nickase, Cpf1 nickase, Cas3 nickase, Cas8a-c nickase, Cas10 nickase, Cse1 nickase, Csy1 nickase, Csn2 nickase, Cas4 nickase, Csm2 nickase, Csm5 nickase, Csf1 nickase, C2C2 nickase, or a NgAgo nickase.

[0171] In one embodiment, the nucleic acid-guided nuclease nickase is a Cas9 nickase.

[0172] In some embodiments, a nucleic acid-guided nuclease nickase can be used to bind to target sequence. With only one active nuclease domain, the nucleic acid-guided nuclease nickase cuts only one strand of a target DNA, creating a single-strand break or "nick". Depending on which mutant is used, the guide NA-hybridized strand or the non-hybridized strand may be cleaved. nucleic acid-guided nuclease nickases bound to 2 gNAs that target opposite strands can create a double-strand break in the nucleic acid. This "dual nickase" strategy increases the specificity of cutting because it requires that both nucleic acid-guided nuclease /gNA complexes be specifically bound at a site before a double-strand break is formed.

[0173] In exemplary embodiments, a Cas9 nickase can be used to bind to target sequence. The term "Cas9 nickase" refers to a modified version of the Cas9 protein, containing a single inactive catalytic domain, i.e., either the RuvC- or the HNH-domain. With only one active nuclease domain, the Cas9 nickase cuts only one strand of the target DNA, creating a single-strand break or "nick". Depending on which mutant is used, the guide RNA-hybridized strand or the non-hybridized strand may be cleaved. Cas9 nickases bound to 2 gRNAs that target opposite strands will create a double-strand break in the DNA. This "dual nickase" strategy can increase the specificity of cutting because it requires that both Cas9/gRNA complexes be specifically bound at a site before a double-strand break is formed.

[0174] Capture of DNA can be carried out using a nucleic acid-guided nuclease nickase. This is illustrated in FIG.6 and FIG. 11, Protocols 2 and 5, respectively. In one exemplary embodiment, as pictured in FIGS. 6 and 11, a nucleic acid-guided nuclease nickase cuts a



single strand of double stranded nucleic acid, wherein the double stranded region comprises methylated nucleotides.

#### Dissociable and Thermostable Nucleic Acid-Guided Nucleases

**[0175]** In some embodiments thermostable nucleic acid-guided nucleases are used in the methods provided herein (thermostable CRISPR/Cas system nucleic acid-guided nucleases or thermostable non-CRISPR/Cas system nucleic acid-guided nucleases). In such embodiments, the reaction temperature is elevated, inducing dissociation of the protein; the reaction temperature is lowered, allowing for the generation of additional cleaved target sequences. In some embodiments, thermostable nucleic acid-guided nucleases maintain at least 50% activity, at least 55% activity, at least 60% activity, at least 65% activity, at least 70% activity, at least 75% activity, at least 80% activity, at least 85% activity, at least 90% activity, at least 95% activity, at least 96% activity, at least 97% activity, at least 98% activity, at least 99% activity, or 100% activity, when maintained for at least 75°C for at least 1 minute. In some embodiments, thermostable nucleic acid-guided nucleases maintain at least 50% activity, when maintained for at least 1 minute at least at 75°C, at least at 80°C, at least at 85°C, at least at 90°C, at least at 91°C, at least at 92°C, at least at 93°C, at least at 94°C, at least at 95°C, 96°C, at least at 97°C, at least at 98°C, at least at 99°C, or at least at 100°C. In some embodiments, thermostable nucleic acid-guided nucleases maintain at least 50% activity, when maintained at least at 75°C for at least 1 minute, 2 minutes, 3 minutes, 4 minutes, or 5 minutes. In some embodiments, a thermostable nucleic acid-guided nucleases maintains at least 50% activity when the temperature is elevated, lowered to 25°C-50°C. In some embodiments, the temperature is lowered to 25°C, to 30°C, to 35°C, to 40°C, to 45°C, or to 50°C In one exemplary embodiment, a thermostable enzyme retains at least 90% activity after 1 min at 95°C.

**[0176]** In some embodiments, the thermostable nucleic acid-guided nuclease is thermostable Cas9, thermostable Cpf1, thermostable Cas3, thermostable Cas8a-c, thermostable Cas10, thermostable Cse1, thermostable Csy1, thermostable Csn2, thermostable Cas4, thermostable Csm2, thermostable Cm5, thermostable Csf1, thermostable C2C2, or thermostable NgAgo.

**[0177]** In some embodiments the thermostable CRISPR/Cas system protein is thermostable Cas9.

[0178] Thermostable nucleic acid-guided nucleases can be isolated, for example, identified by sequence homology in the genome of thermophilic bacteria *Streptococcus thermophilus* and *Pyrococcus furiosus*. Nucleic acid-guided nuclease genes can then be cloned into an expression vector. In one exemplary embodiment, a thermostable Cas9 protein is isolated.

[0179] In another embodiment, a thermostable nucleic acid-guided nuclease can be obtained by in vitro evolution of a non-thermostable nucleic acid-guided nuclease. The sequence of a nucleic acid-guided nuclease can be mutagenized to improve its thermostability.

#### Exemplary Compositions of the Invention

[0180] In one embodiment, provided herein is a composition comprising a nucleic acid fragment, a nickase nucleic acid-guided nuclease-gRNA complex, and labeled nucleotides. In one exemplary embodiment, provided herein is a composition comprising a nucleic acid fragment, a nickase Cas9-gRNA complex, and labeled nucleotides. In such embodiments, the nucleic acid may comprise DNA. The nucleotides can be labeled, for example with biotin. The nucleotides can be part of an antibody-conjugate pair.

[0181] In one embodiment, provided herein is a composition comprising a nucleic acid fragment and a catalytically dead nucleic acid-guided nuclease-gRNA complex, wherein the catalytically dead nucleic acid-guided nuclease is fused to a transposase. In one exemplary embodiment, provided herein is a composition comprising a DNA fragment and a dCas9-gRNA complex, wherein the dCas9 is fused to a transposase.

[0182] In one embodiment, provided herein is a composition comprising a nucleic acid fragment comprising methylated nucleotides, a nickase nucleic acid-guided nuclease-gRNA complex, and unmethylated nucleotides. In an exemplary embodiment, provided herein is a composition comprising a DNA fragment comprising methylated nucleotides, a nickase Cas9-gRNA complex, and unmethylated nucleotides.

[0183] In one embodiment, provided herein is a gDNA complexed with a nucleic acid-guided -DNA endonuclease. In an exemplary embodiment, the nucleic acid-guided -DNA endonuclease is NgAgo.

[0184] In one embodiment, provided herein is a gDNA complexed with a nucleic acid-guided -RNA endonuclease.

[0185] In one embodiment, provided herein is a gRNA complexed with a nucleic acid-guided -DNA endonuclease.

[0186] In one embodiment, provided herein is a gRNA complexed with a nucleic acid-guided -RNA endonuclease. In one embodiment the nucleic acid-guided -RNA endonuclease comprises C2c2.

#### Kits and Articles of Manufacture

[0187] The present application provides kits comprising any one or more of the compositions described herein, including, but not limited to, adapters, gNAs, gDNAs, gRNAs, gNA libraries, gRNA libraries, gDNA libraries, a nucleic acid-guided nuclease, a catalytically dead nucleic acid-guided nuclease, a nickase nucleic acid-guided nuclease, a CRISPR/Cas system protein, a nickase CRISPR/Cas system protein, a catalytically dead CRISPR/Cas system protein, Cas9, dCas9, Cas9 nickase, methylated nucleotides, labeled nucleotides, biotinylated nucleotides, avidin, streptavidin, an enzyme capable of initiating nucleic acid synthesis at a nicked site, DNA Polymerase I, TAQ polymerase, bst DNA Polymerase, an enzyme capable of cleaving methylated nucleotides, a Dpn1 enzyme, and an enzyme capable of methylating DNA, for example a Dam/Dcm1 methyltransferase).

[0188] In one embodiment, the kit comprises a collection or library of gNAs wherein the gNAs are targeted to human genomic DNA sequences, for example particular genes of interest (e.g. cancer genes) SNPs, STRs. In another exemplary embodiment, the kit comprises a collection or library of gNAs wherein the gNAs are targeted to non-human mammalian DNA sequences. In another exemplary embodiment, the kit comprises a collection or library of gNAs wherein the gNAs are targeted to human ribosomal RNA sequences. In another exemplary embodiment, the kit comprises a collection or library of gNAs wherein the gNAs are targeted to human mitochondrial DNA sequences.

[0189] In one exemplary embodiment, the kit comprises a collection or library of gRNAs wherein the gRNAs are targeted to human genomic DNA sequences, for example particular genes of interest (e.g. cancer genes) SNPs, STRs. In another exemplary embodiment, the kit comprises a collection or library of gRNAs wherein the gRNAs are targeted to non-human mammalian DNA sequences. In another exemplary embodiment, the kit comprises a collection or library of gRNAs wherein the gRNAs are targeted to human ribosomal RNA sequences. In

another exemplary embodiment, the kit comprises a collection or library of gRNAs wherein the gRNAs are targeted to human mitochondrial DNA sequences.

[0190] The present application also provides articles of manufacture comprising any one of the kits described herein. Examples of an article of manufacture include vials (including sealed vials).

[0191] The following examples are included for illustrative purposes and are not intended to limit the scope of the invention.

### EXAMPLES

Example 1: Capture of mitochondrial DNA from total human genomic DNA (Protocol 1 for Capture of DNA)

#### Overview

[0192] The objective of this method was to capture mitochondrial DNA from a library of human genomic DNA, as depicted in FIG. 1. A human tissue specimen was subjected to DNA extraction protocols, resulting in a DNA sample comprising >99% human DNA and <1% target sequence. The DNA sample was subjected to sequencing library construction protocols, resulting in a nucleic acid library comprising sequencing indexed adapters. To produce a library of target-specific guide RNAs (gRNAs), a library of target specific gRNA precursors, each comprising a T7 RNA polymerase promoter, a human-specific 20-base pair region, and a stem-loop binding site for Cas9, was subjected to in vitro transcription, yielding a library of target-specific guide RNAs. The library of target-specific guide RNAs was then combined with Cas9 proteins to yield a library of Cas9-gRNA complexes. The Cas9-gRNA complexes were then combined with the nucleic acid library such that the Cas9 cleaved matching target DNA sequences and left other DNA uncleaved. Second adapters were added and allowed to ligate specifically to the 5' phosphorylated blunt ends of cleaved DNA. PCR was then used to amplify specifically using the first and second adapters.

[0193] Mitochondrial DNA makes up approximately 0.1-0.2% of total human genomic DNA. To test the precise site-specific cutting of DNA with Cas9, followed by ligation of adapters, referring to FIG. 2, a test DNA (e.g., a plasmid) 201 was cut with Cas9 at a first location 202 or a second location 203, yielding either first products 204 or second products 205 (see, e.g.,

FIG. 2). Adapters were ligated into the newly available blunt DNA ends. PCR amplification was performed for AUK-F/P7 or MB1OriR/P7, yielding two separate products per reaction. If the first location was cut, the products were 212 and 1.9k in size; if the second location was cut, the products were 359 and 1.75k in size. Verification was performed by sequencing.

[0194] Results showed that Cas9 cutting followed by ligation of adapters allowed for specific amplification of target DNA (see, e.g., FIG. 3); sequencing of the amplified DNA showed that ligation of adapters occurred only at the location specified by the guide RNA (see, e.g., FIG. 4).

[0195] Mitochondrial DNA was then enriched from a mixture containing predominantly human nuclear DNA. 25 guide RNAs specific for mitochondrial DNA were used, then cut with Cas9 and ligated with adapters, followed by amplification. As seen from FIG. 5, two separate reactions allowed for amplification of mitochondrial DNA whereas a reaction without guide RNAs did not amplify any DNA, thus showing that the method efficiently amplifies DNA that is under-represented in any given library.

#### Preparation of DNA libraries

[0196] Human genomic DNA libraries were generated by end repairing 500 ng of fragmented human genomic DNA (ds DNA fragmentase, NEB, treated for 1 hour at 37°C) using the blunt end repair kit (NEB) for 20 minutes at 25°C. Reactions were then heat inactivated at 75°C for 20 minutes, cooled to 25°C then ligated to 15 pmoles of P5/Myc adapters for two hours at 25°C using T4 DNA ligase (NEB). Adapter dimers were removed using the NGS cleanup kit (Life Technologies).

#### Expression of Cas9

[0197] Cas9 (from *S. pyogenes*) was cloned into the pET30 expression vector (EMD biosciences) to insert the hexahistidine tag immediately upstream of the Cas9 start codon. The resulting plasmid was transformed into the Rosetta (DE3) BL21 bacterial strain (EMD biosciences) and grown in 1L of LB media with vigorous aeration until optical density of the culture (OD at 600 nm) reached 0.4. The temperature was lowered to 25°C, 0.2 mM IPTG was added and the culture grown for another four hours. Cells were then harvested by centrifugation (1,000xg for 20 min at 4°C), resuspended in 10 ml binding buffer (20 mM Tris

pH8, 0.5 M NaCl, 5 mM Imidazole, 0.05% NP40) and lysed by sonication (7x10 second bursts at 30% power, Sonifier 250, Branson). Insoluble cell debris were removed by centrifugation at 10,000xg for 20 min; supernatant containing soluble protein was then mixed with 0.4 ml of NTA beads (Qiagen) and loaded onto a column. Beads were washed three times with 4 ml binding buffer, then eluted with 3 x 0.5 ml of binding buffer supplemented with 250 mM Imidazole. Eluted fractions were then concentrated and buffer exchanged with storage buffer (10 mM Tris pH8, 0.3 M NaCl, 0.1 mM EDTA, 1 mM DTT, 50% glycerol) using a 30,000 MWCO protein concentrator (Life Technologies), verified by SDS PAGE followed by Colloidal Blue staining (Life Technologies), quantified, then stored at -20°C for later use.

**[0198]** A mutant Cas9 nickase, a D10A mutant of *S. pyogenes* Cas9, can be produced and purified using the same procedures used to produce Cas9 as above.

#### Preparation of gRNA1 and gRNA2

**[0199]** Three oligonucleotides T7-guideRNA1 and 2 (sequences, in 5' to 3' direction GCCTCGAGCTAATACGACTCACTATAGGGATTATACAGCACTTTAA, and GCCTCGAGCTAATACGACTCACTATAGGGTCTTTTTGGTCCTCGAAG) and stlR (sequence, GT TTT AGA GCT AGA AAT AGC AAG TTA AAA TAA GGC TAG TCC GTT ATC AAC TTG AAA AAG TGG CAC CGA GTC GGT GCT TTT TTT GGA TCC GAT GC) were ordered and synthesized (IDT). The stlR oligonucleotide (300 pmol) was sequentially 5' phosphorylated using T4 PNK (New England Biolabs) and then 5' adenylated using the 5'adenylation kit (New England Biolabs), according to the manufacturer's instructions. T7-guide RNAs oligonucleotides (5 pmol) and the 5'adenylated stlR (10 pmol) were then ligated using thermostable 5'App DNA/RNA ligase (New England Biolabs) at 65°C for one hour. Ligation reactions were heat inactivated at 90°C for 5 min, then amplified by PCR (using OneTaq, New England Biolabs, 30 cycles of 95°C 30 secs, 57°C 20 secs, 72°C, 20 secs) with primers ForT7 (sequence GCC TCG AGC TAA TAC GAC TCA C) and gRU (sequence AAAAAAAGCACCGACTCGGTG). PCR products were purified using PCR cleanup kit (Life Technologies) and verified by agarose gel electrophoresis and sequencing. Verified products were then used as templates for in vitro transcription.

#### Preparation of guide RNA libraries

[0200] T7-guideRNA oligonucleotides (Table 1) and a separate oligonucleotide, stlgR (sequence, GT TTT AGA GCT AGA AAT AGC AAG TTA AAA TAA GGC TAG TCC GTT ATC AAC TTG AAA AAG TGG CAC CGA GTC GGT GCT TTT TTT GGA TCC GAT GC) were ordered and synthesized (IDT).

[0201] The stlgR oligonucleotide (300 pmol) was sequentially 5' phosphorylated using T4 PNK (New England Biolabs) and then 5' adenylated using the 5'adenylation kit (New England Biolabs), according to the manufacturer's instructions. T7-guide RNAs oligonucleotides (5 pmol) and the 5'adenylated stlgR (10 pmol) were then ligated using thermostable 5' App DNA/RNA ligase (New England Biolabs) at 65°C for one hour. Ligation reactions were heat inactivated at 90°C for 5 min, then amplified by PCR (using OneTaq, New England Biolabs, 30 cycles of 95°C 30 secs, 57°C 20 secs, 72°C, 20 secs) with primers ForT7 (sequence GCC TCG AGC TAA TAC GAC TCA C) and gRU (sequence AAAAAAAGCACCGACTCGGTG). PCR products were purified using PCR cleanup kit (Life Technologies) and verified by agarose gel electrophoresis and sequencing. Verified products were then used as templates for in vitro transcription.

Table 1: Mito/T7 primers for ligation reactions	
Series 1 T7-1-F	GCC TCG AGC TAA TAC GAC TCA CTA TAG GCT TGG ATT AGC GTT TAG AA
T7-13-F	GCC TCG AGC TAA TAC GAC TCA CTA TAG GCT CTT AAA ACT AGG CGG CTA
T7-39-F	GCC TCG AGC TAA TAC GAC TCA CTA TAG ATT TAC ACT CAC AAC ACC CT
T7-41-F	GCC TCG AGC TAA TAC GAC TCA CTA TAG AAC AGC TAT CCA TTG GTC TT
T7-43-F	GCC TCG AGC TAA TAC GAC TCA CTA TAG GCA GCC GGA AGC CTA TTC GC
T7-61-F	GCC TCG AGC TAA TAC GAC TCA CTA TAG GTA ATG AGG ATG TAA GCC

	CG
T7-63-F	GCC TCG AGC TAA TAC GAC TCA CTA TAG ATA TTT ACA AGA GGA AAA CC
T7-65-F	GCC TCG AGC TAA TAC GAC TCA CTA TAG GTT TGA AGC TTA GGG AGA GCT
T7-67-F	GCC TCG AGC TAA TAC GAC TCA CTA TAG GTA TGG CTT TGA AGA AGG CG
Series 2 T7mtgRNA3	GCC TCG AGC TAA TAC GAC TCA CTA TAG TAG ATG ACG GGT TGG GCC AG
T7mtgRNA7	GCC TCG AGC TAA TAC GAC TCA CTA TAG AGC TTT ACA GTG GGC TCT AG
T7mtgRNA11	GCC TCG AGC TAA TAC GAC TCA CTA TAG ATG GCA GCT TCT GTG GAA CG
T7mtgRNA15	GCC TCG AGC TAA TAC GAC TCA CTA TAG GTG GTA AGG GCG ATG AGT GT
T7mtgRNA31	GCC TCG AGC TAA TAC GAC TCA CTA TAG TCC ATA ACG CTC CTC ATA CT
T7mtgRNA33	GCC TCG AGC TAA TAC GAC TCA CTA TAG TCT CCC TTC ACC ATT TCC GA

#### In vitro transcription

**[0202]** Verified products were then used as templates for in vitro transcription reactions using the HiScribe T7 transcription kit (New England Biolabs). 500-1000 ng of template was incubated overnight at 37°C according to the manufacturer's instruction. To transcribe the guide libraries into guide RNA, we assembled the following in vitro transcription reaction



mixture: 10 µl purified library (~500 ng), 6.5 µl of H<sub>2</sub>O, of 2.25 µl of ATP, 2.25 µl of CTP, 2.25 µl of GTP, 2.25 µl of UTP, 2.25 µl 10x reaction buffer (NEB) and 2.25 µl of T7 RNA polymerase mix. The reaction was incubated at 37°C for 24 hours, then purified using the RNA cleanup kit (Life Technologies), eluted into 100 µl of RNase-free water, quantified and stored at -20°C until use.

#### DNA-specific Cas9-mediated fragmentation

**[0203]** For cutting the test DNA test guide RNA 1 and 2 were used separately. For cutting and enriching the mitochondrial DNA from a human genomic DNA library, guide RNA series 1 and 2 were used. Diluted guide RNA (1 µl, equivalent to 2 pmol) was combined with 3 µl 10x Cas9 reaction buffer (NEB), 20 µl H<sub>2</sub>O and 1 µl of recombinant Cas9 enzyme (NEB, 1 pmol/µl). A control reaction using a control guide RNA targeting the following sequence (5'-GGATTTATACAGCACTTTAA-3') was performed separately, using the same parameters. This sequence is absent from either the human chromosomal or mitochondrial DNA. Reactions were incubated for 15 min at 37°C, then supplemented with 5 µl diluted DNA library (50 pg/µl) and incubation at 37°C continued for 90 min. The reactions were terminated by adding RNase A (Thermo Fisher Scientific) at a 1:100 dilution, then purifying the DNA using a PCR cleanup kit (LifeTechnologies) and eluting in 30 µl 10 mM Tris-Cl pH 8. Reactions were then stored at -20°C until use.

#### Ligation of adapters and PCR analysis

**[0204]** For the test DNA, reactions after Cas9 digestion were incubated with 15 pmoles of P5/P7 adapters and T4 DNA ligase (NEB) for one hour at 25°C. Ligations were then used as templates for PCR using the TestDNA-F primer (sequence ATGCCGCAGCACTTGG) and P5 primer (sequence AATGATACGGCGACCACCGA). Successful PCR products were confirmed by agarose gel electrophoresis sequenced using the TestDNA-F primer (ElimBio), to show cutting and ligation occurred at the target DNA sequence.

#### Elimination of adapters ligated at old ends

**[0205]** For the enrichment of mitochondrial DNA, after Cas9 digestion, reactions were incubated with 15 pmoles of Flag5/P7 adapters and T4 DNA ligase (NEB) for one hour at 25°C. Multiple molecular biology methods can be employed to eliminate ligation of adapters at the ends of the old P5 and P7 adapters from the original library, for example with enzyme

treatment. Reactions were then used as templates for PCR using the P7 (sequence CAAGCAGAAGACGGCATACGA) and P5 primers (sequence AATGATACGGCGACCACCGA). Successful PCR products were confirmed by agarose gel electrophoresis.

Example 2: Using Cas9 Nickase to label, then purify test DNA from a DNA mixture (Protocol 2 for Capture of DNA)

#### Overview

**[0206]** The purpose of this method was to capture a region of interest (e.g., SNP, STR, etc.) from a library of human genomic DNA, as depicted for example in FIG. 6. To apply this protocol, a test DNA containing a site for a nicking enzyme guided by target-specific guide RNAs (NtAlwI (NEB)) was mixed at 1% or 5% into another pool of DNA that does not contain this site. Nickase Cas9 cleaves only at target sequence and only cuts one strand of DNA. Nickase was used to nick target sequences. Single strand cuts (nicks) are a substrate for DNA polymerase I which can be used to replace the DNA downstream of the nick with biotin labeled DNA. Biotinylated DNA of interest was then purified, amplified, and sequenced.

**[0207]** The mixture of DNA with 701 and without 702 target sites for nicking (e.g., GGATC) 703 was nicked 704 using NtAlwI (see, e.g., FIG. 7). DNA that does not have nickase sites was present in 100X excess compared to DNA of interest with the nickase sites. Single strand cuts (nicks) are a substrate for DNA polymerase I which are used to replace 705 the DNA downstream of the nick with biotin labeled DNA. Biotin-labeled DNA of interest was isolated by streptavidin binding and washing 706. PCR amplification was performed with specific primers 707, yielding amplified regions of interest 708 and amplified unlabeled sequences 709 present, for example, due to non-specific labeling or capture.

**[0208]** Two specific PCR reactions (one for the test DNA with regions of interest, the other for other DNA without regions of interest) showed that the test DNA had been enriched approximately 50-fold (see, e.g., FIG. 8).

#### Expression and purification of Cas9 Nickase

**[0209]** A mutant Cas9 nickase, a D10A mutant of *S. pyogenes* Cas9, can be produced and purified using the same procedures used to produce Cas9 as above.

#### Sequence specific Cas9 Nickase-mediated nicking

[0210] Diluted guide RNA, targeting the following sequence 5'-GGATTATACAGCACTTTAA-3' (1 ul, equivalent to 2 pmol) was combined with 3 ul 10x Cas9 reaction buffer (NEB), 20 ul H<sub>2</sub>O and 1 ul of recombinant Cas9 Nickase enzyme (10 pmol/ul). This target sequence is only present on the target DNA (that makes either 5% or 1% of the total DNA). Reactions were incubated for 15 min at 37°C, then supplemented with 5 µl diluted DNA (100 ng total) and incubation at 37°C continued for 90 min. The reactions were terminated by adding RNase A (Thermo Fisher Scientific) at a 1:100 dilution, then purifying the DNA using a PCR cleanup kit (LifeTechnologies) and eluting in 30 ul 10 mM Tris-Cl pH 8. Reactions were then stored at -20°C until use.

#### Biotin nick translation

[0211] Nicked DNA was incubated with E. coli DNA polymerase I which has 5'→3' exonuclease activity, and is capable of initiating DNA synthesis at nicks, thus allowing it to replace nucleotides downstream of a nick with labeled nucleotides (in the case of this procedure, biotin-labeled nucleotides). Nick labeling reactions were performed in 20 ul of DNA polymerase buffer (NEB) with 1 unit of E. coli DNA polymerase I (NEB) and 0.02 mM each of dCTP, dGTP and dTTP, 0.01 mM of dATP and 0.01 mM biotin-C14 labeled dATP (LifeTechnologies) for 30 min at 25°C. Reactions were terminated by adding 1 mM EDTA.

#### Enrichment of biotin labeled DNA

[0212] Streptavidin C1 beads (5 ul per reaction, LifeTechnologies) were resuspended in 1 ml binding buffer (50 mM Tris-Cl pH 8, 1 mM EDTA, 0.1 % Tween20), bound to a magnetic rack, and washed twice with binding buffer. Beads were then resuspended in 30 ul of binding buffer and mixed with the nick translation reaction, then incubated at 25°C for 30 minutes. Beads were captured using the magnetic rack and washed four times with 0.5 ml binding buffer, then three times with 0.5 ml of 10 mM Tris-Cl pH 8. Beads were then resuspended in 20 ul 10 mM Tris-Cl pH 8, then used as templates for PCR to determine the proportion of test DNA and other DNA.

Example 3: Use of a catalytically dead nucleic acid-guided nuclease -transposase fusion to insert adaptors in human genomic library followed by enriching for specific SNPs (Protocol 3 for Capture of DNA)

[0213] In this example, catalytically dead nucleic acid-guided nuclease-transposase fusion protein (e.g. a dCas9-transposase fusion protein) is expressed and purified from *E. coli* as described for the Cas9 purification. The fusion protein is complexed with adapters (Nextera) then with guide NAs (e.g. gRNAs) targeting the regions of interest (e.g. human SNPs). Then the complex is added to human genomic DNA; regions of interest are targeted by the catalytically dead nucleic acid-guided nuclease, bringing the transposase-adapter complex in close proximity to the regions of interest, allowing insertion of the adapters. Human SNPs can then be amplified by PCR and then sequenced using MiSeq; thus human SNPs can be enriched from human genomic DNA.

Example 4: Using catalytically dead nucleic acid-guided nuclease and nucleic acid-guided nuclease to protect mitochondrial DNA and digest remaining human nuclear DNA from a human genomic DNA library (Protocol 4 for Capture of DNA)

[0214] A human genomic DNA library with P5/P7 adapters from a clinical, forensic or environmental sample is obtained. To enrich for certain regions (e.g. SNPs) guide RNAs targeting these regions are made and incubated with catalytically dead nucleic acid-guided nuclease (e.g. dCas9) then added to the human genomic DNA library for 20 minutes at 37C. Then, a library of guide NAs covering the human genome complexed with active nucleic acid-guided nuclease (e.g. Cas9) is added. The catalytically dead nucleic acid-guided nuclease will remain bound at the target locations and protect the regions of interests (e.g. SNPs) from being cleaved and becoming non PCR-amplifiable and sequence-able. Thus, the DNA of interest remains intact while all other DNA will be cleaved and eliminated. DNA of interest is recovered from the reactions using the PCR cleanup kit, PCR amplified and sequenced using MiSeq.

[0215] For proof of concept testing, mitochondrial DNA is enriched from a total human genomic DNA library. Mitochondrial specific guide RNAs are added to dCas9, then added to the library. Then, random guide RNAs complexed with Cas9 degrade any sequence, except for those inaccessible because already protected by bound dCas9. This allows enrichment of the mitochondrial DNA to levels far higher than the original 0.1-0.2%.

Example 5: Using a nucleic acid-guided nuclease Nickase (e.g. Cas9 nickase) to protect and then enrich SNPs from human genomic DNA by replacing methylated DNA with unmethylated DNA (Protocol 5 for Capture of DNA)

## Overview

[0216] The objective of this method is to capture a region of interest from a library of human genomic DNA, as depicted in FIG. 11. As proof of principle, a test DNA containing a site for a nicking enzyme, NtBbvCI (NEB) was mixed at 1% or 5% into another pool of DNA that does not contain this site. The entire mixture was then treated with Dam methyltransferase to add methyl groups to all GATC sequences. Both test DNA and other DNA contain GATC motifs in their sequence. The mixture was then nicked using NtBbvCI, then incubated with DNA polymerase I and unlabeled nucleotides (dATP, dCTP, dGTP, dTTP), then heat inactivated at 75°C for 20 minutes. The mixture is then digested with DpnI, which only digests methylated DNA; unmethylated or hemimethylated DNA will not be digested. FIG. 12 shows that methylation of test DNA renders it susceptible to DpnI mediated cleavage.

## Expression and purification of Cas9 Nickase

[0217] A mutant Cas9 nickase, a D10A mutant of *S. pyogenes* Cas9, can be produced and purified using the same procedures used to produce Cas9 as above.

## Sequence specific Cas9 Nickase-mediated nicking

[0218] Diluted guide RNA, targeting the following sequence 5'-GGATTTATACAGCACTTTAA-3' (1 ul, equivalent to 2 pmol) is combined with 3 ul 10x Cas9 reaction buffer (NEB), 20 ul H<sub>2</sub>O and 1 ul of recombinant Cas9 Nickase enzyme (10 pmol/ul). This target sequence is only present on the target DNA (that makes either 5% or 1% of the total DNA). Reactions are incubated for 15 min at 37°C, then supplemented with 5 µl diluted DNA (100 ng total) and incubation at 37°C continued for 90 min. The reactions are terminated by adding RNase A (Thermo Fisher Scientific) at a 1:100 dilution, then purifying the DNA using a PCR cleanup kit (LifeTechnologies) and eluting in 30 ul 10 mM Tris-Cl pH 8. Reactions are then stored at -20°C until use.

## Unlabeled DNA nick translation

[0219] Nicked DNA is incubated with *E. coli* DNA polymerase I which has 5'→3' exonuclease activity, and is capable of initiating DNA synthesis at nicks, thus allowing it to replace nucleotides downstream of a nick with labeled nucleotides (in the case of this procedure, biotin-labeled nucleotides). Nick labeling reactions are performed in 20 ul of DNA polymerase

buffer (NEB) with 1 unit of *E. coli* DNA polymerase I (NEB) and 0.02 mM each of dATP, dCTP, dGTP and dTTP for 30 min at 25°C. Reactions are terminated by heat inactivation at 75°C for 20 minutes.

#### Digestion with DpnI

[0220] Reactions are then incubated with DpnI (NEB) for one hour at 37°C. DNA is recovered using the PCR cleanup kit (LifeTechnologies) and then used as template for PCR using test DNA specific primers.

Example 6: Using a nucleic acid-guided nuclease nickase (e.g. Cas9 nickase) to generate long 3' overhangs flanking regions of interests, then ligating adapters to these overhangs to enrich target DNA (Protocol 6 for capture of DNA)

[0221] The objective of this method can be used to enrich a region of DNA, from any DNA source (e.g., library, genomic, or PCR), as depicted for example in FIG. 13. The nucleic acid-guided nuclease nickase (e.g. Cas9 Nickase) can be targeted to proximal sites using two guide NAs (e.g. gRNAs) and, resulting in nicking of DNA at each location. Alternatively, an adapter can be ligated on only one side, then filled in, then an adapter can be ligated on the other side. The two nicks can be close to each other (e.g., within 10 to 15 bp). Single nicks may be generated in non-target molecules. Because of the proximity of the two nicking sites, a double stranded break can be created when the reaction is heated, e.g. to 65 °C, resulting in long (e.g., 10-15 bp) 3' overhangs. These overhangs can be recognized by a thermostable single stranded DNA/RNA ligase, such as a Thermostable 5' App DNA/RNA ligase to allow for site-specific ligation of single stranded adapters. The ligase can, for example, only recognize long 3' overhangs, thus ensuring that adapters will not be ligated at other sites. This process can be repeated using the nucleic acid-guided nuclease nickase and guide NA targeting on the other side of the region of interest, followed by ligation as above using a second single stranded adapter. Once two adapters have been ligated on either side of the region of interest, the region can be amplified or sequenced directly.

## CLAIMS

What is claimed is:

1. A method of capturing target nucleic acid sequences comprising:
  - (a) providing a sample comprising a plurality of adapter-ligated nucleic acids, wherein the nucleic acids are ligated to a first adapter at one end and ligated to a second adapter at the other end;
  - (b) contacting the sample with a plurality of nucleic acid-guided nuclease-gNA complexes, wherein the gNAs are complementary to targeted sites of interest contained in a subset of the nucleic acids, thereby generating a plurality of nucleic acid fragments ligated to a first or second adapter at one end and no adapter at the other end; and
  - (c) contacting the plurality of nucleic acid fragments with third adapters, thereby generating a plurality of nucleic acid fragments ligated to either the first or second adapter at one end and the third adapter at the other end.
2. The method of claim 1, wherein the nucleic acid-guided nuclease is a CRISPR/Cas system protein.
3. The method of claim 1, wherein the nucleic acid-guided nuclease is a non-CRISPR/Cas system protein.
4. The method of claim 1, wherein the nucleic acid-guided nuclease is selected from the group consisting of CAS Class I Type I, CAS Class I Type III, CAS Class I Type IV, CAS Class II Type II, and CAS Class II Type V.
5. The method of claim 1, wherein the nucleic acid-guided nuclease is selected from the group consisting of Cas9, Cpf1, Cas3, Cas8a-c, Cas10, Cse1, Csy1, Csn2, Cas4, Csm2, Cmr5, Csf1, C2c2, and NgAgo.
6. The method of any one of claims 1 to 5, wherein the gNAs are gRNAs.
7. The method of any one of claims 1 to 5, wherein the gNAs are gDNAs.
8. The method of any one of claims 1 to 7, wherein the contacting with the plurality of nucleic acid-guided nuclease-gNA complexes cleaves the targeted sites of interest contained in a subset of the nucleic acids, thereby generating a plurality of nucleic acid fragments comprising a first or second adapter at one end and no adapter at the other end.

9. The method of any one of claims 1 to 8, wherein the method further comprises amplifying the product of step (c) using first or second and third adapter-specific PCR.
10. The method of any one of claims 1 to 9, wherein the nucleic acids are selected from the group consisting of single stranded DNA, double stranded DNA, single stranded RNA, double stranded RNA, and a DNA/RNA hybrid.
11. The method of claim 10, wherein the nucleic acids are double stranded DNA.
12. The method of any one of claims 1 to 11, wherein nucleic acids are from genomic DNA.
13. The method of claim 12, wherein the genomic DNA is human.
14. The method of any one of claims 1 to 13, wherein the nucleic acids which are adapter-ligated are from 20 bp to 5000 bp in length.
15. The method of any one of claims 1 to 14, wherein the targeted sites of interest are single nucleotide polymorphisms (SNPs), short tandem repeats (STRs), cancer genes, inserts, deletions, structural variations, exons, genetic mutations, or regulatory regions.
16. The method of claim 9, wherein amplified products are used for cloning, sequencing, or genotyping.
17. The method of any one of claims 1 to 16, wherein the adapters are from 20 bp to 100 bp in length.
18. The method of any one of claims 1 to 17, wherein the adapters comprise primer binding sites.
19. The method of any one of claims 1 to 18, wherein the adapters comprise sequencing adapters or restriction sites.
20. The method of any one of claims 1 to 19, wherein the targeted sites of interest represent less than 50% of the total nucleic acid in the sample.
21. The method of any one of claims 1 to 20, wherein the sample is obtained from a biological sample a clinical sample, a forensic sample, or an environmental sample.



22. The method of any one of claims 1 to 21, wherein the first and second adapters are identical.
23. The method of any one of claims 1 to 21, wherein the first and second adapters are different.
24. The method of any one of claims 1 to 23, wherein the sample comprises a sequencing library.
25. A method of introducing labeled nucleotides at targeted sites of interest comprising:
- (a) providing a sample comprising a plurality of nucleic acid fragments;
  - (b) contacting the sample with a plurality of nucleic acid-guided nuclease nickase-gNA complexes wherein the gNAs are complementary to targeted sites of interest in the nucleic acid fragments, thereby generating a plurality of nicked nucleic acid fragments at the targeted sites of interest; and
  - (c) contacting the plurality of nicked nucleic acid fragments with an enzyme capable of initiating nucleic acid synthesis at a nicked site, and labeled nucleotides, thereby generating a plurality of nucleic acid fragments comprising labeled nucleotides in the targeted sites of interest.
26. The method of claim 25, wherein the nucleic acid-guided nuclease nickase is selected from the group consisting of CAS Class I Type I nickase, CAS Class I Type III nickase, CAS Class I Type IV nickase, CAS Class II Type II nickase, and CAS Class II Type V nickase.
27. The method of claim 25, wherein the nucleic acid-guided nuclease nickase is selected from the group consisting of Cas9 nickase, Cpf1 nickase, Cas3 nickase, Cas8a-c nickase, Cas10 nickase, Cse1 nickase, Csy1 nickase, Csn2 nickase, Cas4 nickase, Csm2 nickase, Cm5 nickase, Csf1 nickase, C2C2 nickase, CPF1 nickase, and NgAgo nickase.
28. The method of any one of claims 25 to 27, wherein the gNAs are gRNAs.
29. The method of any one of claims 25 to 27, wherein the gNAs are gDNAs.
30. The method of any one of claims 25 to 29, wherein the nucleic acid fragments are selected from the group consisting of single stranded DNA fragments, double stranded DNA fragments, single stranded RNA fragments, double stranded RNA fragments, and a DNA/RNA hybrid fragments.

31. The method of claim 30, wherein the nucleic acid fragments are double stranded DNA fragments.

32. The method of claim 31, wherein double stranded DNA fragments are from genomic DNA.

33. The method of claim 32, wherein the genomic DNA is human.

34. The method of any one of claims 25 to 33, wherein the nucleic acid fragments are from 20bp to 5000 bp in length.

35. The method of any one of claims 25 to 34, wherein the targeted sites of interest are single nucleotide polymorphisms (SNPs), short tandem repeats (STRs), cancer genes, inserts, deletions, structural variations, exons, genetic mutations, or regulatory regions.

36. The method of any one of claims 25 to 35, wherein the targeted sites of interest represent less than 50% of the total nucleic acid in the sample.

37. The method of any one of claims 25 to 36, wherein the sample is obtained from a biological sample a clinical sample, a forensic sample, or an environmental sample.

38. The method of any one of claims 25 to 37, wherein the labeled nucleotides are biotinylated nucleotides.

39. The method of any one of claims 25 to 37, wherein the labeled nucleotides are part of an antibody conjugate pair.

40. The method of claim 38, further comprising contacting the nucleic acid fragments comprising biotinylated nucleotides with avidin or streptavidin, thereby capturing the targeted sites of interest.

41. The method of any one of claims 25 to 40, wherein the enzyme capable of initiating nucleic acid synthesis at a nicked site is DNA Polymerase I, a Klenow fragment, a TAQ polymerase, or a Bst DNA Polymerase.

42. The method of any one of claims 25 to 41, wherein the nucleic acid-guided nuclease nickase nicks the 5' end of the nucleic acid fragments.

43. The method of any one of claims 25 to 42, wherein the nucleic acid fragments are from 20 bp to 5000 bp in length.

44. The method of any one of claims 25 to 43, wherein the sample is obtained from a biological sample, a clinical sample, a forensic sample, or an environmental sample.

45. A method of capturing target nucleic acid sequences of interest comprising:

(a) providing a sample comprising a plurality of adapter-ligated nucleic acids, wherein the nucleic acids are ligated to a first adapter at one end and are ligated to a second adapter at the other end; and

(b) contacting the sample with a plurality of catalytically dead nucleic acid-guided nuclease-gRNA complexes, wherein the catalytically dead nucleic acid-guided nuclease is fused to a transposase, wherein the gRNAs are complementary to targeted sites of interest contained in a subset of the nucleic acids, and wherein the complexes are loaded with a plurality of third adapters, to generate a plurality of nucleic acid fragments comprising either a first or second adapter at one end, and a third adapter at the other end.

46. The method of claim 45, wherein the method further comprises amplifying the product of step (b) using first or second adapter and third adapter-specific PCR.

47. The method of claim 45, wherein the catalytically dead nucleic acid-guided nuclease is derived from a CRISPR/Cas system protein.

48. The method of claim 45, wherein the catalytically dead nucleic acid-guided nuclease is derived from a non-CRISPR/Cas system protein.

49. The method of claim 45, wherein the catalytically dead nucleic acid-guided nuclease is selected from the group consisting of dead CAS Class I Type I, dead CAS Class I Type III, dead CAS Class I Type IV, dead CAS Class II Type II, and dead CAS Class II Type V.

50. The method of claim 45, wherein the catalytically dead nucleic acid-guided nuclease is selected from the group consisting of dCas9, dCpf1, dCas3, dCas8a-c, dCas10, dCse1, dCsy1, dCsn2, dCas4, dCsm2, dCmr5, dCsf1, dC2C2, and dNgAgo.

51. The method of any one of claims 45 to 50, wherein the gNAs are gRNAs.

52. The method of any one of claims 45 to 50, wherein the gNAs are gDNAs.

53. The method of any one of claims 45 to 52, wherein nucleic acid sequences are from genomic DNA.

54. The method of claim 53, wherein the genomic DNA is human.
55. The method of any one of claims 45 to 54, wherein the catalytically dead nucleic acid-guided nuclease is fused to the N-terminus of the transposase.
56. The method of any one of claims 45 to 54, wherein the catalytically dead nucleic acid-guided nuclease is fused to the C-terminus of the transposase.
57. The method of any one of claims 45 to 56, wherein the nucleic acids, which are adapter ligated, are from 20 bp to 5000 bp in length.
58. The method of any one of claims 45 to 57, wherein the contacting of step (b) allows for the insertion of the second adapter into the targeted nucleic acid sequences.
59. The method of any one of claims 45 to 58, wherein the targeted sites of interest are single nucleotide polymorphisms (SNPs), short tandem repeats (STRs), cancer genes, inserts, deletions, structural variations, exons, genetic mutations, or regulatory regions.
60. The method of claim 46, wherein amplified products are used for cloning, sequencing, or genotyping.
61. The method of any one of claims 45 to 59, wherein the adapters are from 20 bp to 100 bp in length.
62. The method of any one of claims 45 to 61, wherein the adapters comprise primer binding sites.
63. The method of any one of claims 45 to 62, wherein the adapters comprise sequencing adapters or restriction sites.
64. The method of any one of claims 45 to 63, wherein the targeted sites of interest represent less than 50% of the total nucleic acid in the sample.
65. The method of any one of claims 45 to 64, wherein the sample is obtained from a biological sample, a clinical sample, a forensic sample, or an environmental sample.
66. A method of capturing target nucleic acid sequences of interest comprising:
- (a) providing a sample comprising a plurality of adapter-ligated nucleic acids, wherein the nucleic acids are ligated to the adapter at the 5' end and 3' ends;

(b) contacting the sample with a plurality of catalytically dead nucleic acid-guided nuclease-gNA complexes, wherein the gNAs are complementary to targeted sites of interest contained in a subset of the nucleic acids, thereby generating a plurality of nucleic acids adapter-ligated at the 5' and 3' ends, bound to a catalytically dead nucleic acid-guided nuclease-gNA complex; and

(c) contacting the sample with a plurality of catalytically dead nucleic acid-guided nuclease-gNA complexes, wherein the gNAs are complementary to both targeted sites of interest and targeted sites not of interest in the nucleic acids, thereby generating a plurality of nucleic acid fragments comprising nucleic acid sequences not of interest, adapter ligated at only one of the 5' or 3' ends.

67. The method of claim 66, wherein the catalytically dead nucleic acid-guided nuclease is a CRISPR/Cas system protein.

68. The method of claim 66, wherein the catalytically dead nucleic acid-guided nuclease is a non-CRISPR/Cas system protein.

69. The method of claim 66, wherein the catalytically-dead nucleic acid-guided nuclease is selected from the group consisting of dead CAS Class I Type I, dead CAS Class I Type III, dead CAS Class I Type IV, dead CAS Class II Type II, and dead CAS Class II Type V.

70. The method of claim 66, wherein the catalytically dead nucleic acid-guided nuclease is selected from the group consisting of dCas9, dCpf1, dCas3, dCas8a-c, dCas10, dCse1, dCsy1, dCsn2, dCas4, dCsm2, dCm5, dCsf1, dC2C2, dCPF1, and dNgAgo.

71. The method of any one of claims 66 to 70, wherein the gNAs are gRNAs.

72. The method of any one of claims 66 to 70, wherein the gNAs are gDNAs.

73. The method of any one of claims 66 to 72, wherein the contacting of step (c) does not displace the plurality of nucleic acids adapter-ligated at the 5' and 3' ends, bound to a catalytically dead nucleic acid-guided nuclease-gNA complex of step (b).

74. The method of any one of claims 66 to 73, wherein the contacting in step (d) cleaves the targeted sites not of interest contained in a subset of the nucleic acids, thereby generating a plurality of nucleic acid fragments comprising nucleic acid sequences not of interest, adapter ligated at only one of the 5' or 3' ends.

75. The method of any one of claims 66 to 74, wherein the method further comprises removing the bound catalytically dead nucleic acid-guided nuclease-gNA complex and amplifying the product of step (b) using adapter-specific PCR.

76. The method of any one of claims 66 to 75, wherein the nucleic acids are selected from the group consisting of single stranded DNA, double stranded DNA, single stranded RNA, double stranded RNA, and a DNA/RNA hybrid.

77. The method of claim 76, wherein the nucleic acids are double stranded DNA.

78. The method of any one of claims 66 to 77, wherein nucleic acids are from genomic DNA.

79. The method of claim 78, wherein the genomic DNA is human.

80. The method of any one of claims 66 to 79, wherein the nucleic acids adapter-ligated at the 5'ends and the 3'ends are from 20 bp to 5000 bp.

81. The method of any one of claims 66 to 80, wherein the targeted sites of interest are single nucleotide polymorphisms (SNPs), short tandem repeats (STRs), cancer genes, inserts, deletions, structural variations, exons, genetic mutations, or regulatory regions.

82. The method of claim 75, wherein amplified products are used for cloning, sequencing, or genotyping.

83. The method of any one of claims 66 to 82, wherein the adapters are from 20 bp to 100 bp in length.

84. The method of any one of claims 66 to 82, wherein the adapters comprise primer binding sites.

85. The method of any one of claims 66 to 82, wherein the adapters comprise sequencing adapters or restriction sites.

86. The method of any one of claims 66 to 82, wherein the targeted sites of interest represent less than 50% of the total nucleic acid in the sample.

87. The method of any one of claims 66 to 82, wherein the sample is obtained from a biological sample a clinical sample, a forensic sample, or an environmental sample.

88. A method of capturing target nucleic acid sequences of interest comprising:

- (a) providing a sample comprising a plurality of nucleic acid sequences, wherein the nucleic acid sequences comprise methylated nucleotides, and wherein the nucleic acid sequences are adapter ligated on the 5' and 3' ends;
- (b) contacting the sample with a plurality of nucleic acid-guided nuclease nickase-gNA complexes, wherein the gNAs are complementary to targeted sites of interest in a subset of the nucleic acid sequences, thereby generating a plurality of nicked sites of interest in the subset of the nucleic acid sequences, and wherein the target nucleic acid sequences are adapter ligated on the 5' and 3' ends;
- (c) contacting the sample with an enzyme capable of initiating DNA synthesis at a nicked site, and unmethylated nucleotides, thereby generating a plurality of nucleic acid sequences comprising unmethylated nucleotides in the targeted sites of interest and wherein the nucleic acid sequences are adapter ligated on the 5' and 3' ends; and
- (d) contacting the sample with an enzyme capable of cutting methylated nucleic acids, thereby generating a plurality of nucleic acid fragments comprising methylated nucleic acids, wherein the plurality of nucleic acid fragments comprising methylated nucleic acids that are adapter ligated on at most one of the 5' and 3' ends.

89. The method of claim 88, wherein the nucleic acid-guided nuclease nickase is selected from the group consisting of CAS Class I Type I nickase, CAS Class I Type III nickase, CAS Class I Type IV nickase, CAS Class II Type II nickase, and CAS Class II Type V nickase.

90. The method of claim 88, wherein the nucleic acid-guided nuclease nickase is selected from the group consisting of Cas9 nickase, Cpf1 nickase, Cas3 nickase, Cas8a-c nickase, Cas10 nickase, Cse1 nickase, Csy1 nickase, Csn2 nickase, Cas4 nickase, Csm2 nickase, Cmr5 nickase, Csf1 nickase, C2C2 nickase, and NgAgo nickase.

91. The method of claim 88, wherein the gNAs are gRNAs.

92. The method of claim 88, wherein the gNAs are gDNAs.

93. The method of any one of claims 88 to 92, wherein the DNA is double stranded DNA.

94. The method of claim 93, wherein the double stranded DNA is from genomic DNA.

95. The method of claim 94, wherein the genomic DNA is human.
96. The method of any one of claims 88 to 95, wherein the DNA sequences are from 20 bp to 5000 bp in length.
97. The method of any one of claims 88 to 96, wherein the targeted sites of interest are single nucleotide polymorphisms (SNPs), short tandem repeats (STRs), cancer genes, inserts, deletions, structural variations, exons, genetic mutations, or regulatory regions.
98. The method of any one of claims 88 to 97, wherein the targeted sites of interest represent less than 50% of the total DNA in the sample.
99. The method of any one of claims 88 to 98, wherein the sample is obtained from a biological sample a clinical sample, a forensic sample, or an environmental sample.
100. The method of any one of claims 88 to 99, wherein the enzyme capable of initiating nucleic acid synthesis at a nicked site is a DNA Polymerase I, Klenow fragment a TAQ polymerase, or a Bst DNA Polymerase.
101. The method of any one of claims 88 to 100, wherein the nucleic acid-guided nuclease nickase nicks the 5' end of the DNA sequences.
102. The method of any one of claims 88 to 101, wherein the enzyme capable of cutting methylated DNA is DpnI.
103. A method of capturing target DNA sequences of interest comprising:
- (a) contacting the sample with a plurality of nucleic acid-guided nuclease nickase-gNA complexes, wherein the gNAs are complementary to targeted sites of interest flanking a region of interest in a subset of the DNA sequences, thereby generating a plurality of nicked DNA at sites adjacent to the regions of interest;
  - (b) heating the sample to 65°C thereby causing cause nicks in close proximity to generate a double stranded breaks;
  - (c) contacting the double stranded breaks with a thermostable ligase thereby allowing ligation of adapter sequences at these sites only; and
  - (d) repeating steps a-c to place a second adapter on the other side of the region of interest, thus allowing enrichment of the region of interest.
104. The method of claim 103, wherein the gNAs are gRNAs.



105. The method of claim 103, wherein the gNAs are gDNAs.
106. The method of any one of claims 103 to 105, wherein the the nucleic acid-guided nuclease nickase is selected from the group consisting of CAS Class I Type I nickase, CAS Class I Type III nickase, CAS Class I Type IV nickase, CAS Class II Type II nickase, and CAS Class II Type V nickase.
107. The method of any one of claims 103 to 105, wherein the nucleic acid-guided nuclease nickase is selected from the group consisting of Cas9 nickase, Cpf1 nickase, Cas3 nickase, Cas8a-c nickase, Cas10 nickase, Cse1 nickase, Csy1 nickase, Csn2 nickase, Cas4 nickase, Csm2 nickase, Cmr5 nickase, Csf1 nickase, C2C2 nickase, and NgAgo nickase.
108. The method of claim 103, wherein the DNA is double stranded DNA.
109. The method of claim 108, wherein the double stranded DNA is from genomic DNA.
110. The method of claim 109, wherein the genomic DNA is human.
111. The method of any one of claims 103 to 110, wherein the DNA sequences are from 20 bp to 5000 bp in length.
112. The method of any one of claims 103 to 111, wherein the targeted sites of interest are single nucleotide polymorphisms (SNPs), short tandem repeats (STRs), cancer genes, inserts, deletions, structural variations, exons, genetic mutations, or regulatory regions.
113. The method of any one of claims 103 to 112, wherein the targeted sites of interest represent less than 50% of the total DNA in the sample.
114. The method of any one of claims 103 to 113, wherein the sample is obtained from a biological sample, a clinical sample, a forensic sample, or an environmental sample.
115. The method of any one of claims 103 to 113, wherein the thermostable ligase capable of contacting the double stranded break is thermostable 5' App DNA/RNA ligase or T4 RNA ligase.
116. The method of any one of claims 103 to 115, wherein the nucleic acid-guided nuclease nickase nicks the 5' end of the DNA sequences.
117. A method of enriching a sample for sequences of interest, comprising:

- (a) providing a sample comprising sequences of interest and targeted sequences for depletion, wherein the sequences of interest comprise less than 50% of the sample; and
- (b) contacting the sample with a plurality of either nucleic acid-guided RNA endonuclease-gRNA complexes or a plurality of nucleic acid-guided DNA endonuclease-gDNA complexes, wherein the gRNAs and gDNAs are complementary to the targeted sequences, and whereby the targeted sequences are cleaved.

118. The method of claim 117, further comprising extracting the sequences of interest and the targeted sequences for depletion from the sample.

119. The method of claim 118, further comprising fragmenting the extracted sequences.

120. The method of any one of claims 117 to 119, wherein the cleaved targeted sequences are removed by size- exclusion.

121. The method of any one of claims 117 to 120, wherein the sample is any one of a biological sample, a clinical sample, a forensic sample or an environmental sample.

122. The method of any one of claims 117 to 121, wherein the sample comprises host nucleic acid sequences targeted for depletion and non-host nucleic acid sequences of interest.

123. The method of claim 122, wherein the non-host nucleic acid sequences comprise microbial nucleic acid sequences.

124. The method of claim 123, wherein the microbial nucleic acid sequences are bacterial, viral or eukaryotic parasitic nucleic acid sequences.

125. The method of any one of claims 117 to 124, wherein the gRNAs and gDNAs are complementary to ribosomal RNA sequences, spliced transcripts, unspliced transcripts, introns, exons, or noncoding RNAs.

126. The method of claims 118, wherein the extracted nucleic acids include single-stranded or double-stranded RNA.

127. The method of claim 118, wherein the extracted nucleic acids include single-stranded or double-stranded DNA.

128. The method of any one of claims 117 to 127, wherein the sequences of interest comprise less than 10% of the extracted nucleic acids.

129. The method of any one of claims 117 to 128, wherein the nucleic acid-guided RNA endonuclease comprises C2c2.

130. The method of claim 129, wherein the C2c2 is catalytically dead. □

131. The method of any one of claims 117 to 128, wherein the nucleic acid-guided DNA endonuclease comprises NgAgo.

132. The method of claim 131, wherein the NgAgo is catalytically dead.

133. The method of any one of claims 117 to 132, wherein the sample is selected from whole blood, plasma, serum, tears, saliva, mucous, cerebrospinal fluid, teeth, bone, fingernails, feces, urine, tissue, and a biopsy.

134. A method of enriching a sample comprising:

- (a) providing a sample comprising host nucleic acids and non-host nucleic acids;
- (b) contacting the sample with a plurality of nucleic acid-guided RNA endonuclease -gRNA complexes or a plurality of nucleic acid-guided DNA endonuclease -gDNA complexes, wherein the gRNAs and gDNAs are complementary to targeted sites in the host nucleic acids, and
- (c) enriching the sample for non-host nucleic acids.

135. The method of claim 134, wherein the nucleic acid-guided RNA endonuclease comprises C2c2.

136. The method of claim 134, wherein the nucleic acid-guided RNA endonuclease comprises catalytically dead C2c2.

137. The method of claim 134, wherein the nucleic acid-guided DNA endonuclease comprises NgAgo.

138. The method of claim 134, wherein the nucleic acid-guided DNA endonuclease comprises catalytically dead NgAgo.

139. The method of anyone of claims 134 to 138, wherein the host is selected from the group consisting of a human, cow, horse, sheep, pig, monkey, dog, cat, gerbil, bird, mouse, and rat.

140. The method of anyone of claims 134 to 138, wherein the non-host is a prokaryotic organism.

141. The method of anyone of claims 134 to 138, wherein the non-host is selected from the group consisting of a eukaryote, a virus, a bacterium, a fungus, and a protozoan.

142. The method of anyone of claims 134 to 141, wherein the adapter-ligated host nucleic acids and non-host nucleic acids range from 50 bp to 1000 bp.

143. The method of anyone of claims 134 to 142, wherein the non-host nucleic acids comprise less than 50% of the total nucleic acids in the sample.

144. The method of anyone of claims 134 to 143, wherein the sample is any one of a biological sample, a clinical sample, a forensic sample or an environmental sample.

145. The method of anyone of claims 134 to 144, wherein step (c) comprises reverse-transcribing the product of step (b) into cDNA.

146. The method of anyone of claims 134 to 145, wherein step (c) comprises removing the host nucleic acids by size-exclusion.

147. The method of anyone of claims 134 to 145, wherein step (c) comprises removing the host nucleic acids with the use of biotin.

148. The method of anyone of claims 134 to 147, wherein the sample is selected from whole blood, plasma, serum, tears, saliva, mucous, cerebrospinal fluid, teeth, bone, fingernails, feces, urine, tissue, and a biopsy.

149. A composition comprising a nucleic acid fragment, a nucleic acid-guided nuclease nickase-gNA complex, and labeled nucleotides.

150. The composition of claim 149, wherein the nucleic acid-guided nuclease nickase is selected from the group consisting of CAS Class I Type I nickase, CAS Class I Type III nickase, CAS Class I Type IV nickase, CAS Class II Type II nickase, and CAS Class II Type V nickase.

151. The composition of claim 149, wherein the nucleic acid-guided nuclease nickase is selected from the group consisting of Cas9 nickase, Cpf1 nickase, Cas3 nickase, Cas8a-c nickase,

Cas10 nickase, Cse1 nickase, Csy1 nickase, Csn2 nickase, Cas4 nickase, Csm2 nickase, Cmr5 nickase, Csf1 nickase, C2C2 nickase, and NgAgo nickase.

152. The composition of any one of claims 149 to 151, wherein the gNAs are gRNAs.
153. The composition of any one of claims 149 to 151, wherein the gNAs are gDNAs.
154. The composition of any one of claims 149 to 153, wherein the nucleic acid fragment comprises DNA.
155. The composition of any one of claims 149 to 153, wherein the nucleic acid fragment comprises RNA.
156. The composition of any one of claims 149 to 155, wherein the nucleotides are labeled with biotin.
157. The composition of any one of claims 149 to 156, wherein the nucleotides are part of an antibody conjugate pair.
158. A composition comprising a nucleic acid fragment and a catalytically dead nucleic acid-guided nuclease-gNA complex, wherein the catalytically dead nucleic acid-guided nuclease is fused to a transposase.
159. The composition of claim 158, wherein the catalytically-dead nucleic acid-guided nuclease is selected from the group consisting of dead CAS Class I Type I, dead CAS Class I Type III, dead CAS Class I Type IV, dead CAS Class II Type II, and dead CAS Class II Type V.
160. The composition of claim 158, wherein the catalytically dead nucleic acid-guided nuclease is selected from the group consisting of dCas9, dCpf1, dCas3, dCas8a-c, dCas10, dCse1, dCsy1, dCsn2, dCas4, dCsm2, dCmr5, dCsf1, dC2C2, and dNgAgo.
161. The composition of any one of claims 158, 159, or 160, wherein the gNAs are gRNAs.
162. The composition of any one of claims 158 to 161, wherein the gNAs are gDNAs.
163. The composition of any one of claims 158 to 162, wherein the nucleic acid fragment comprises DNA.

164. The composition of any one of claims 158 to 163, wherein the nucleic acid fragment comprises RNA.

165. The composition of any one of claims 158 to 164, wherein the catalytically dead nucleic acid-guided nuclease is fused to the N-terminus of the transposase.

166. The composition of any one of claims 158 to 165, wherein the catalytically dead nucleic acid-guided nuclease is fused to the C-terminus of the transposase.

167. A composition comprising a nucleic acid fragment comprising methylated nucleotides, a nucleic acid-guided nuclease nickase-gNA complex, and unmethylated nucleotides.

168. The composition of claim 167, wherein the nucleic acid-guided nuclease nickase is selected from the group consisting of CAS Class I Type I nickase, CAS Class I Type III nickase, CAS Class I Type IV nickase, CAS Class II Type II nickase, and CAS Class II Type V nickase.

169. The composition of claim 167, wherein the nucleic acid-guided nuclease nickase is selected from the group consisting of Cas9 nickase, Cpf1 nickase, Cas3 nickase, Cas8a-c nickase, Cas10 nickase, Cse1 nickase, Csy1 nickase, Csn2 nickase, Cas4 nickase, Csm2 nickase, Cmr5 nickase, Csf1 nickase, C2C2 nickase, and NgAgo nickase.

170. The composition of any one of claims 167, 168, or 169, wherein the gNAs are gRNAs.

171. The composition of any one of claims 167 to 170, wherein the gNAs are gDNAs.

172. The composition of any one of claims 167 to 171, wherein the nucleic acid fragment comprises DNA.

173. The composition of any one of claims 167 to 171, wherein the nucleic acid fragment comprises RNA.

174. The composition of any one of claims 167 to 173, wherein the nucleotides are labeled with biotin.

175. The composition of any one of claims 167 to 174, wherein the nucleotides are part of an antibody conjugate pair.

FIG. 1

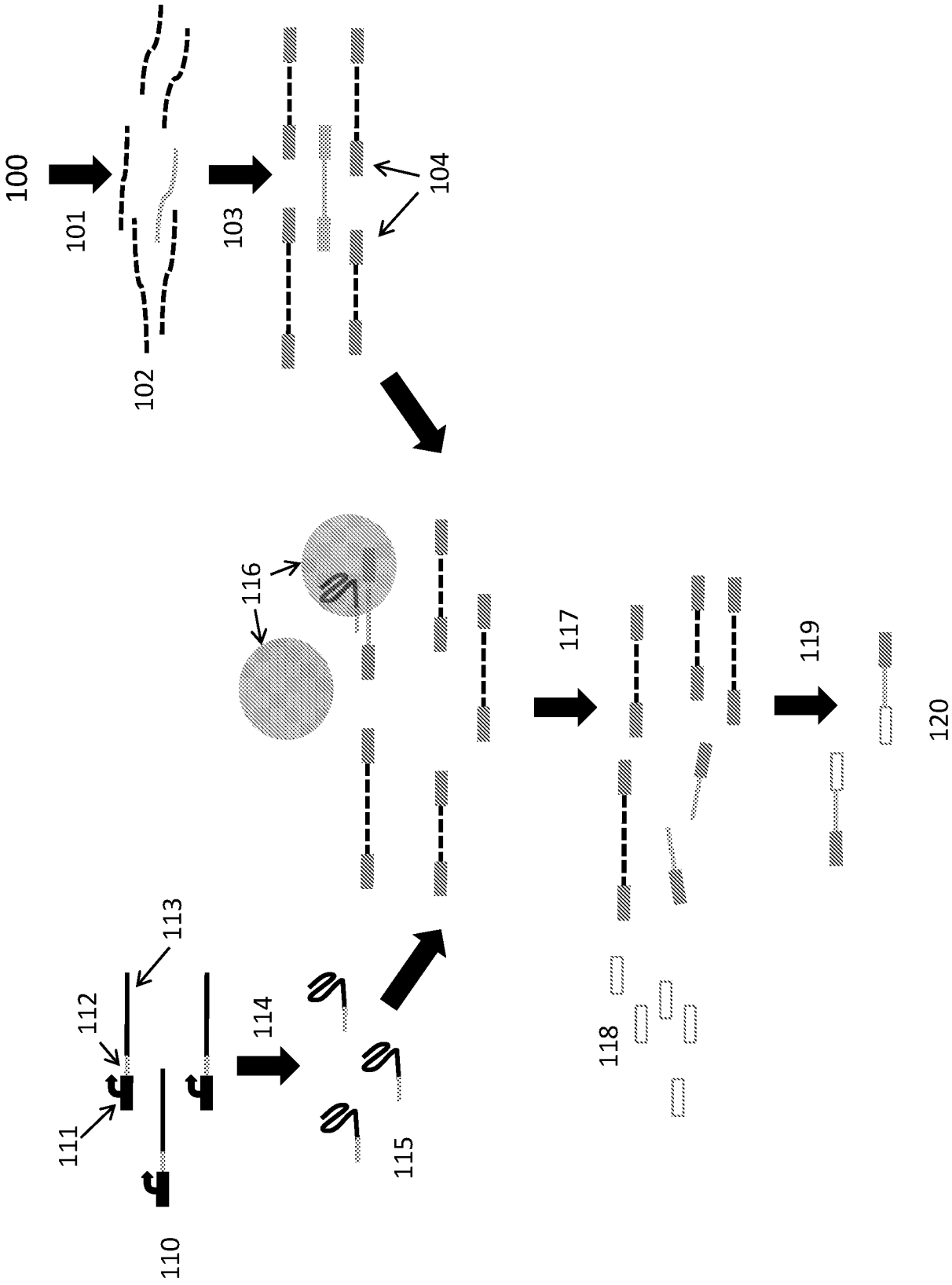


FIG. 2

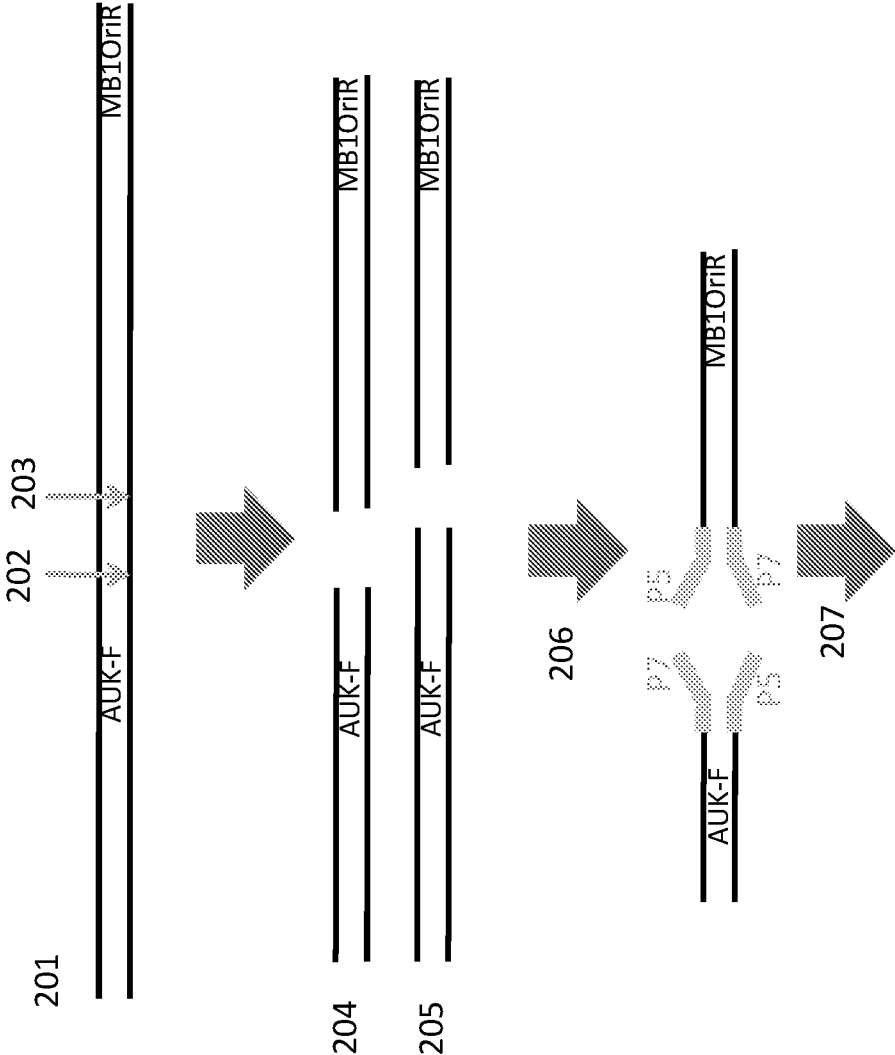




FIG. 3

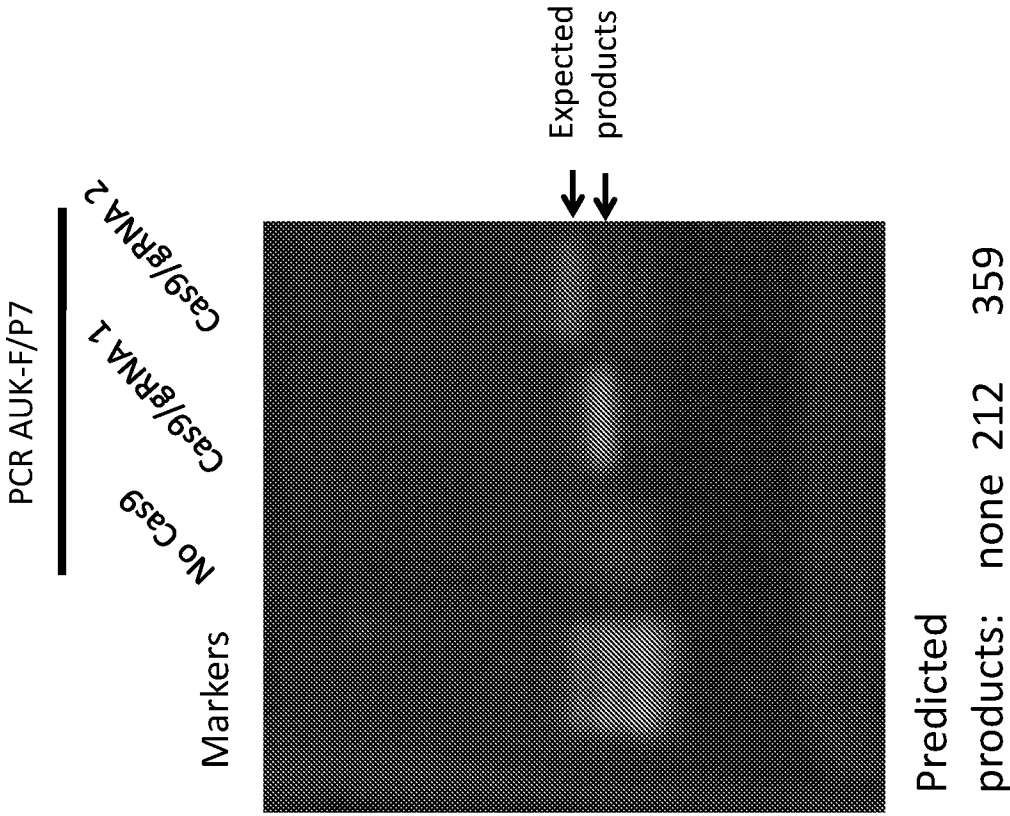


FIG. 4

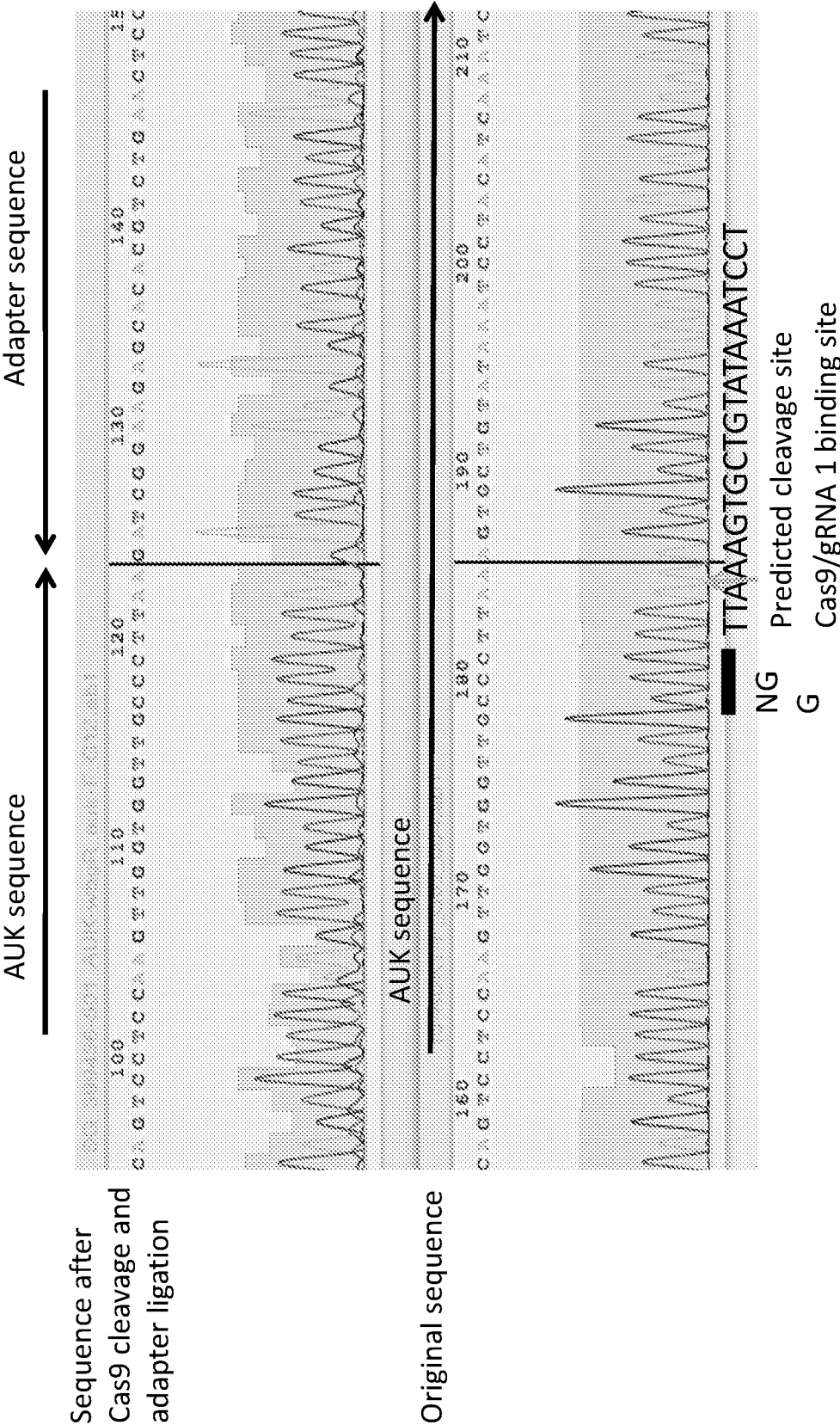


FIG. 5

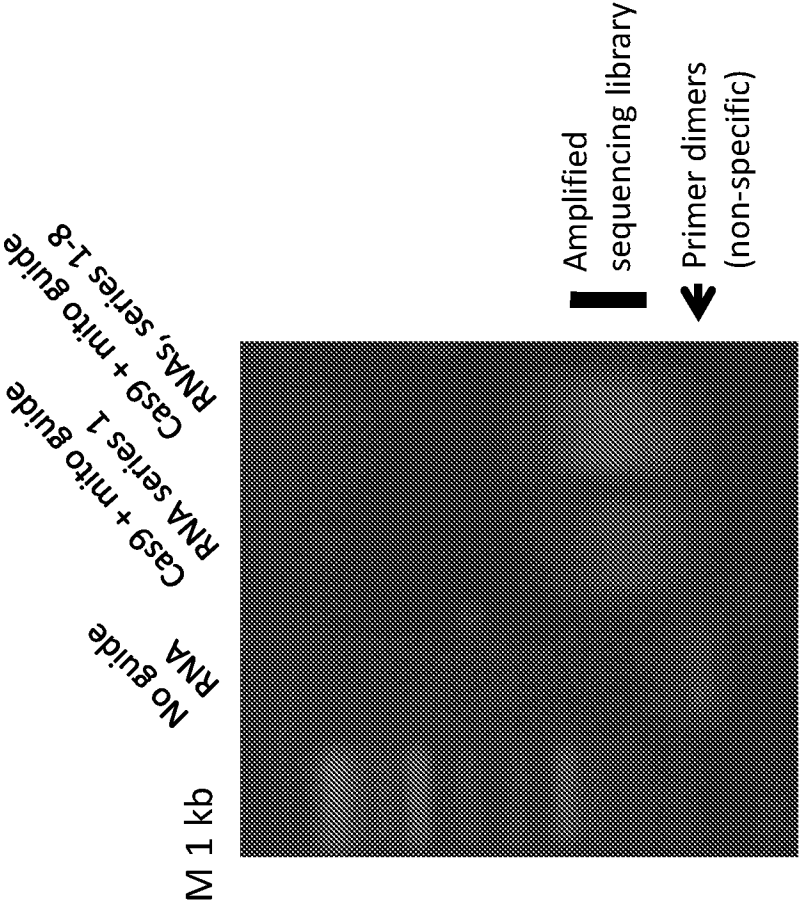


FIG. 6

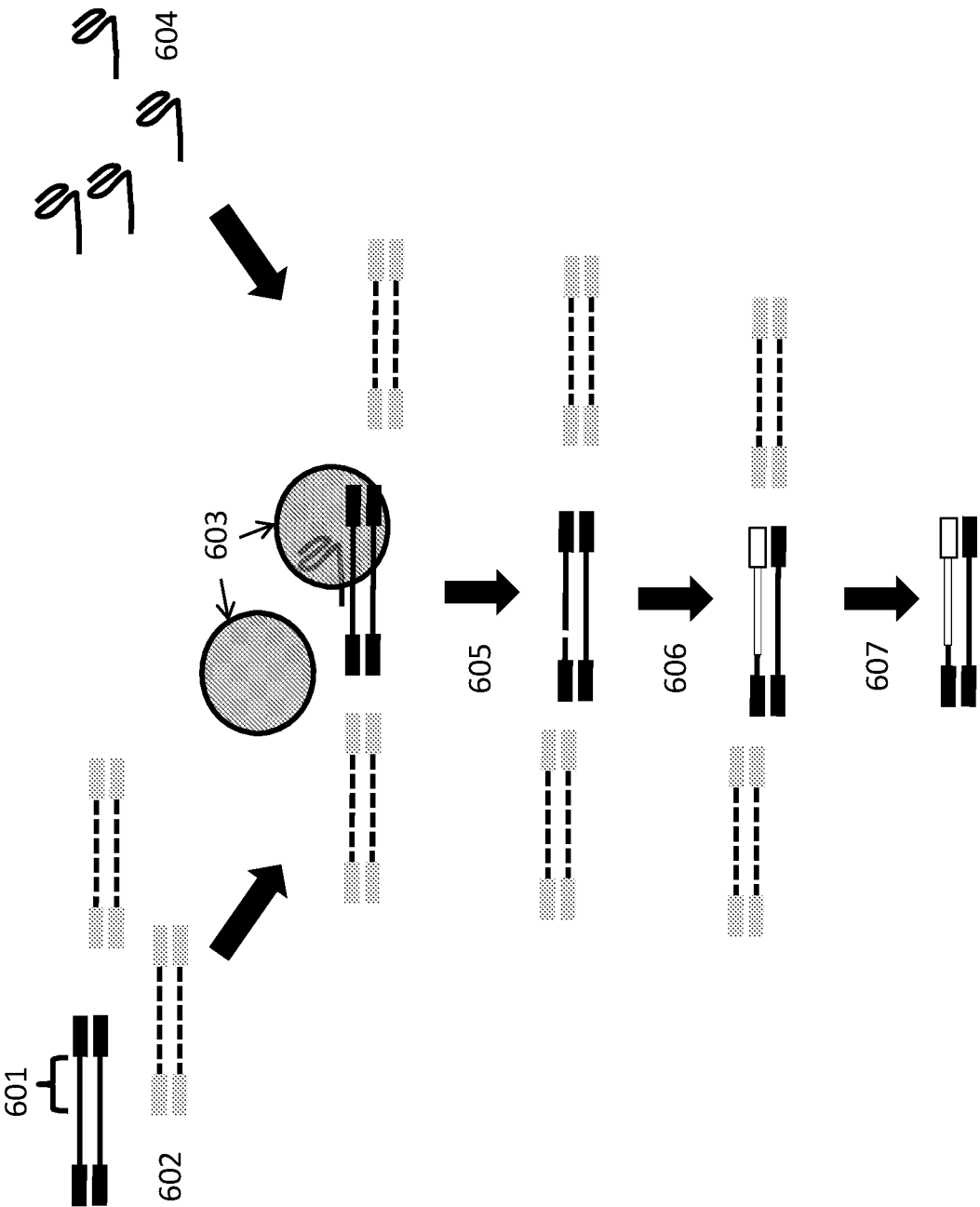
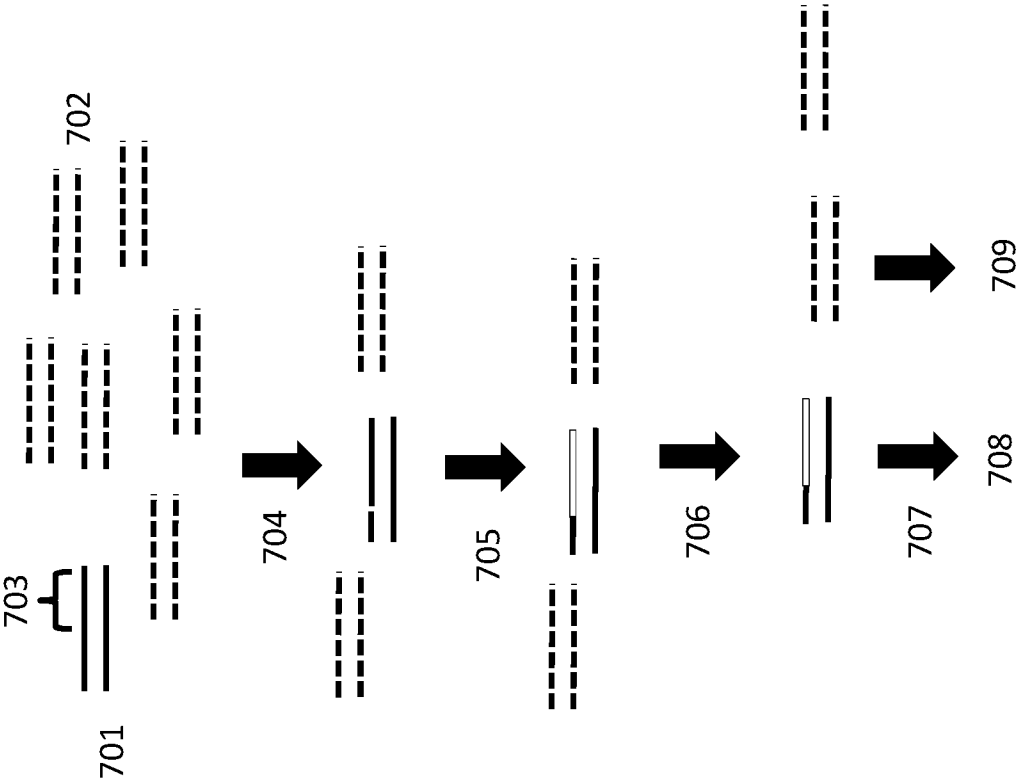


FIG. 7



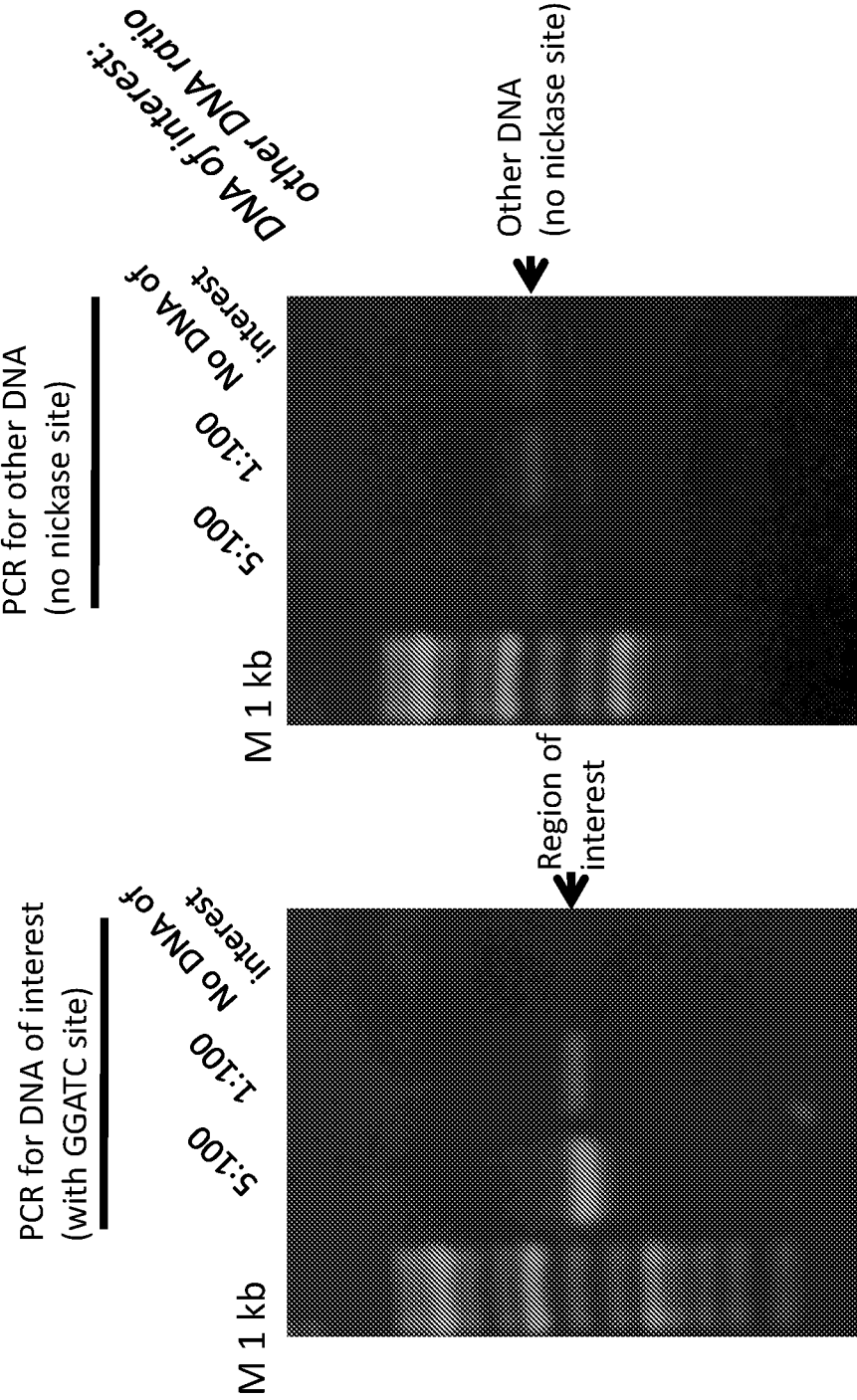


FIG. 8

FIG. 9

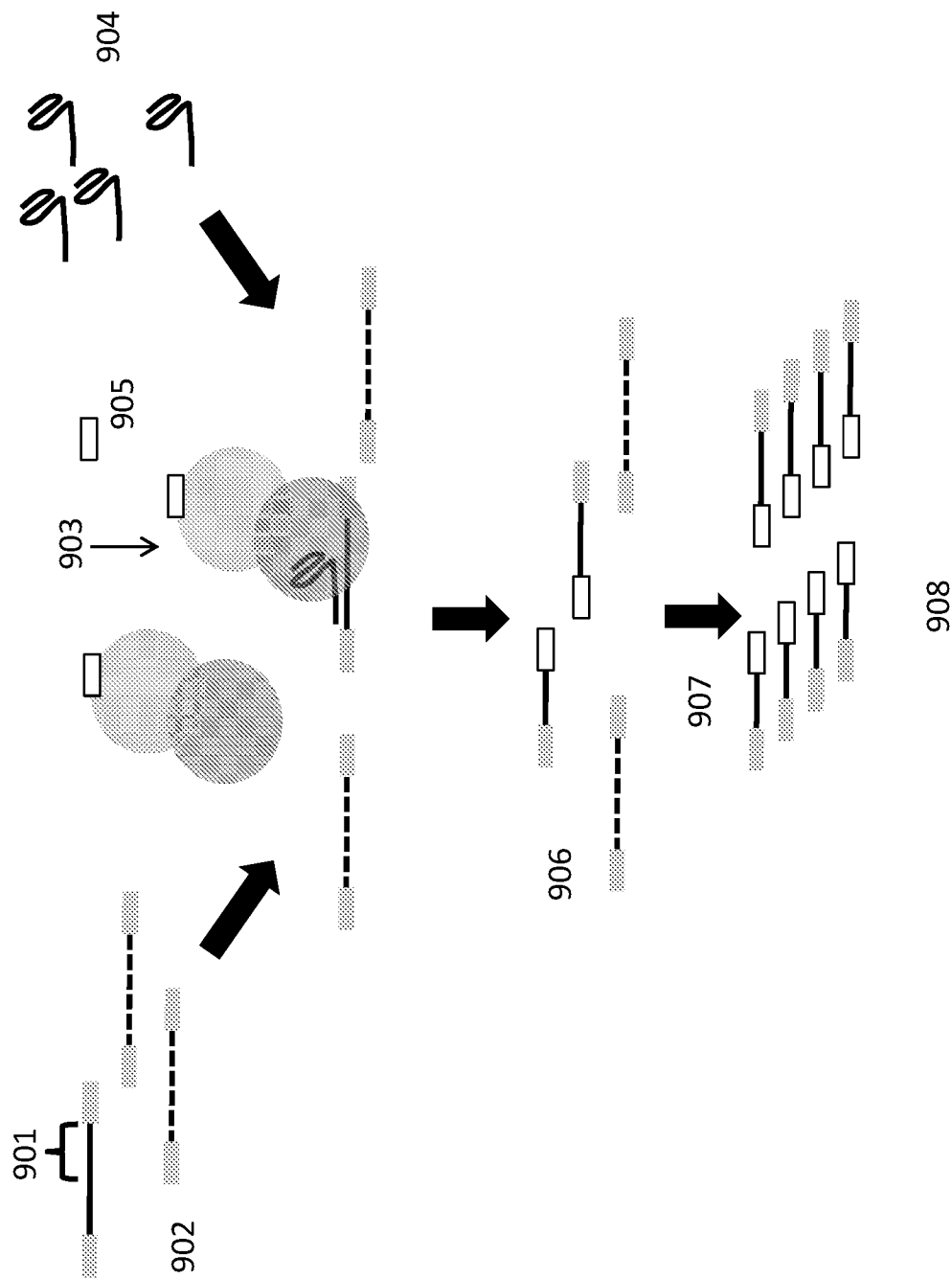


FIG. 10

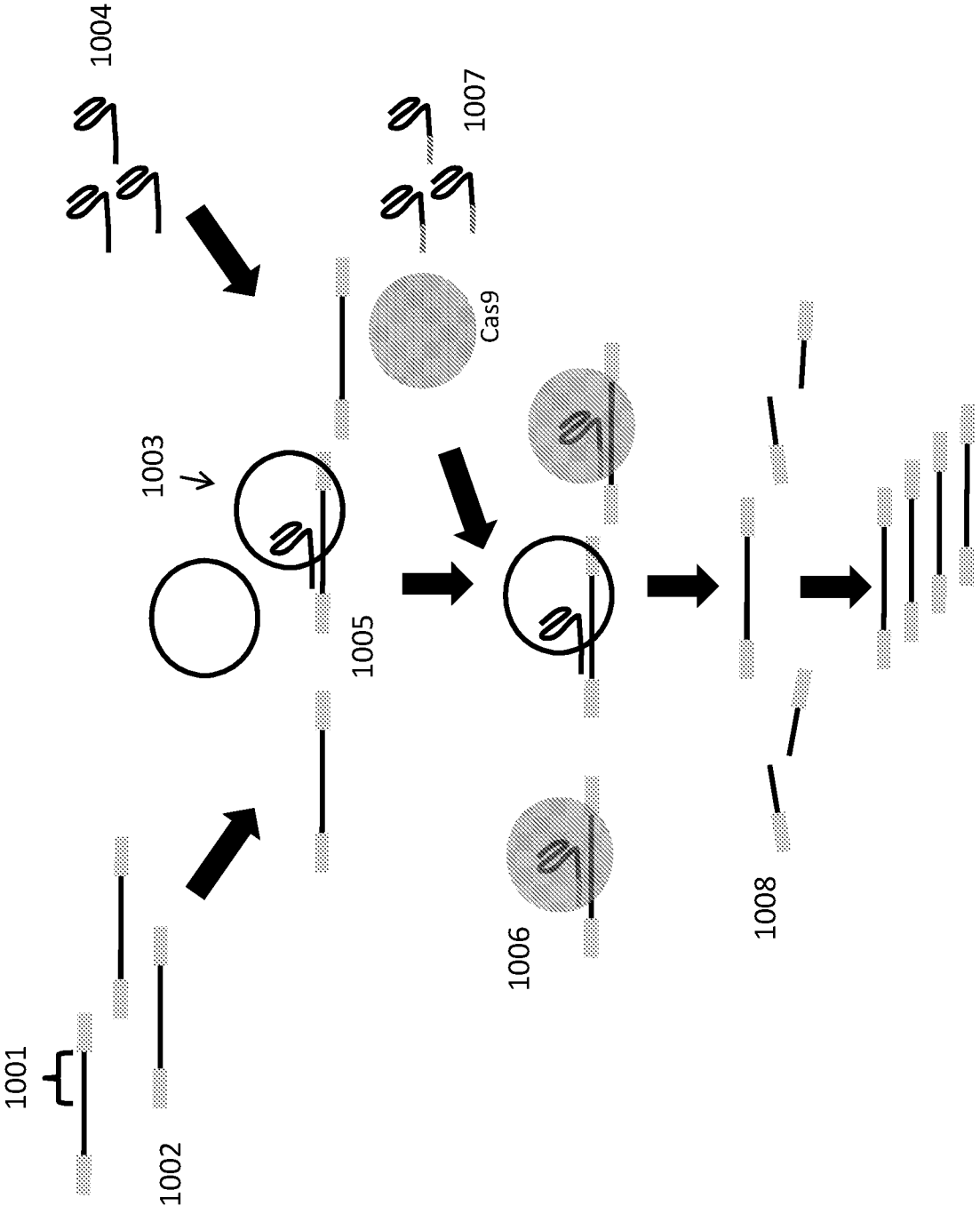




FIG. 11

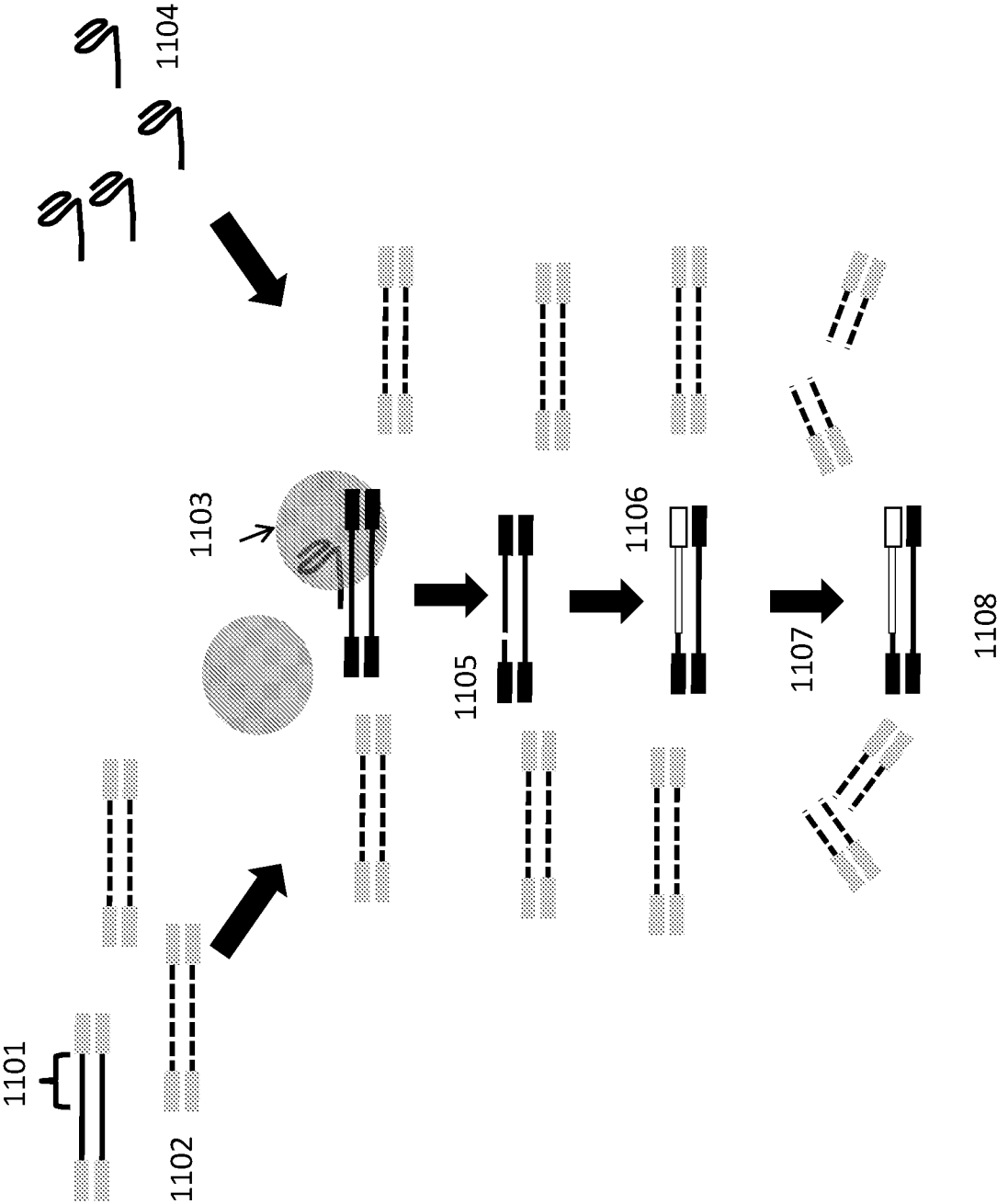


FIG. 12

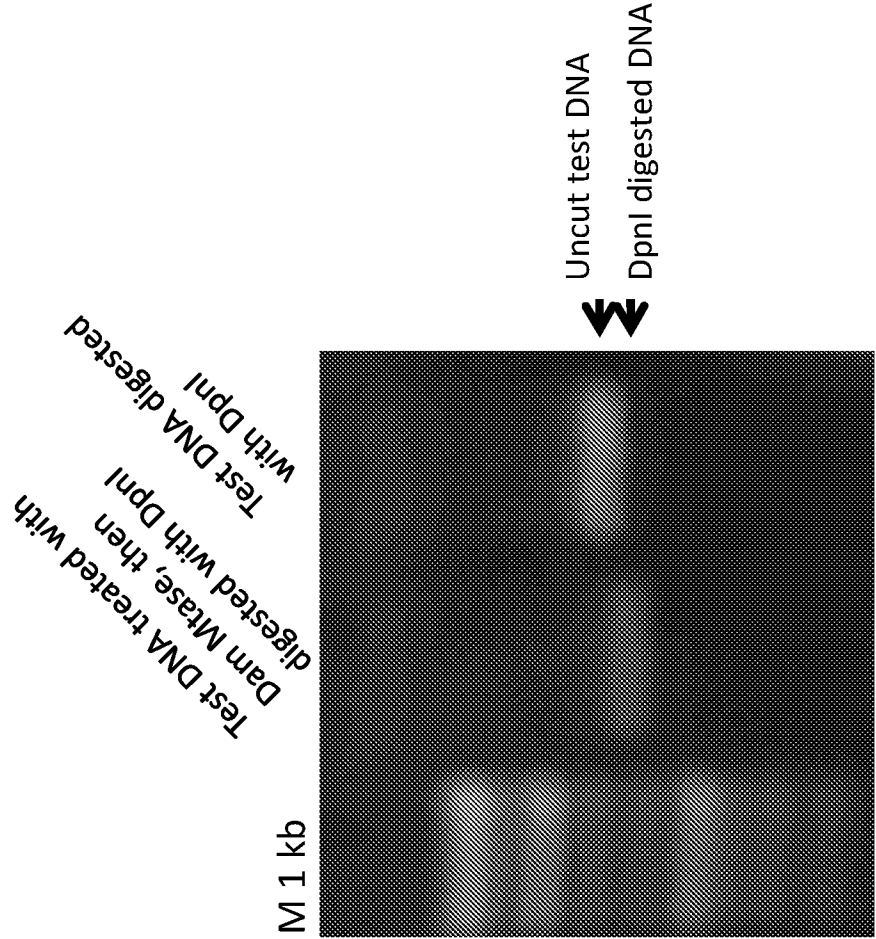
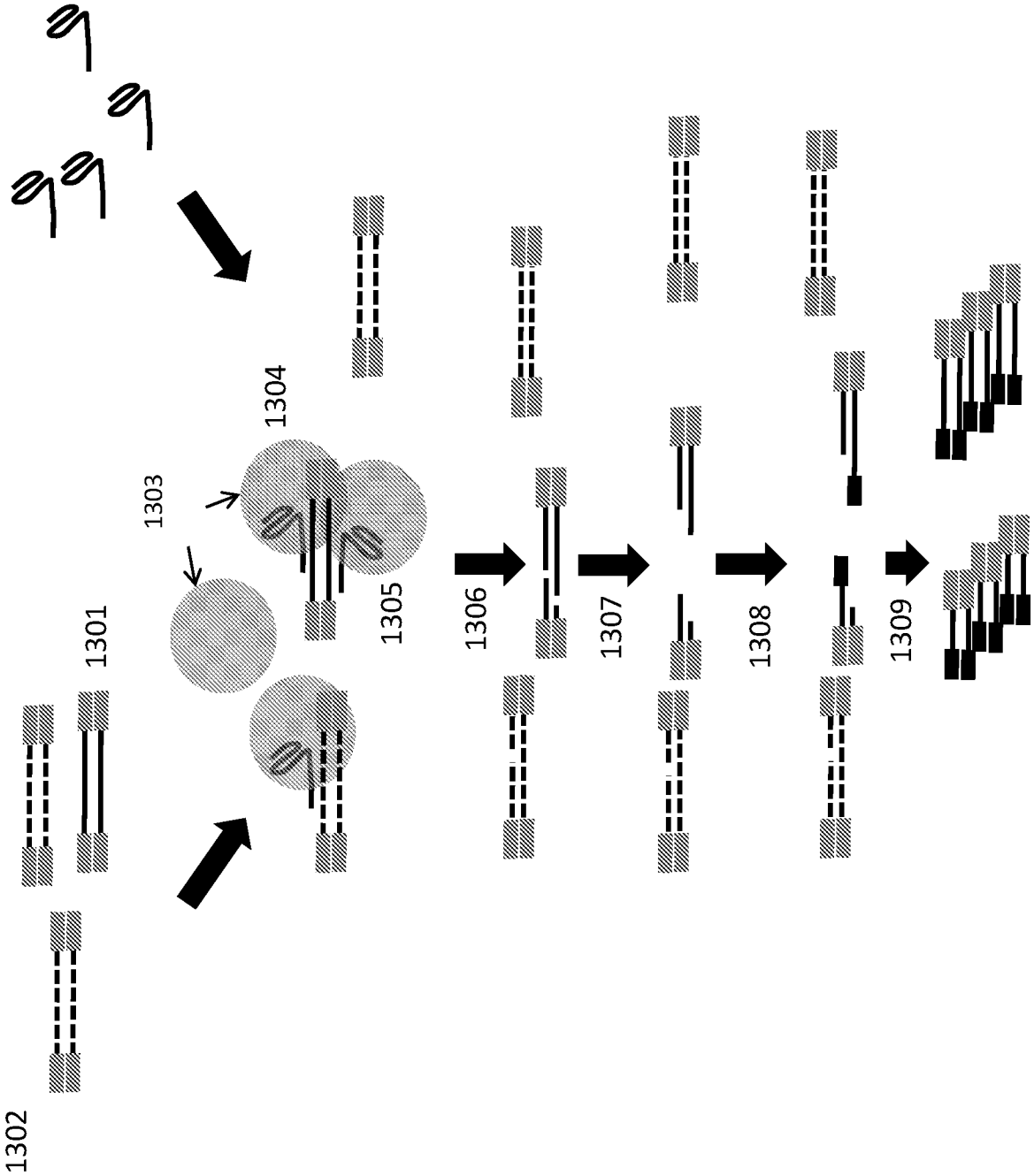


FIG. 13



## INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2016/047631

**Box No. I** Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:
- a. ☐ forming part of the international application as filed:  
☐ in the form of an Annex C/ST.25 text file.  
☐ on paper or in the form of an image file.
- b. ☐ furnished together with the international application under PCT Rule 13*ter*. 1(a) for the purposes of international search only in the form of an Annex C/ST.25 text file.
- c. ☒ furnished subsequent to the international filing date for the purposes of international search only:  
☒ in the form of an Annex C/ST.25 text file (Rule 13*ter*. 1(a)).  
☐ on paper or in the form of an image file (Rule 13*ter*. 1(b) and Administrative Instructions, Section 713).
2. ☒ In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.
3. Additional comments:

# INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2016/047631

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

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

1. ☐ Claims Nos.:  
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claims Nos.:  
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. ☒ Claims Nos.: 8-24, 30-44, 53-59, 61-65, 73-87, 96-102, 111-116, 121-125, 128-133, 142-148, 154-157, 162-166, 171-175  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

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

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

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

### Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
- ☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
- ☐ No protest accompanied the payment of additional search fees.

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2016/047631

<b>A. CLASSIFICATION OF SUBJECT MATTER</b> IPC(8) - C12N 15/10; C12Q 1/68; G01N 33/52 (2016.01) CPC - C12N 15/1082; C12N 15/66; C12Q 1/6844 (2016.11) According to International Patent Classification (IPC) or to both national classification and IPC		
<b>B. FIELDS SEARCHED</b> Minimum documentation searched (classification system followed by classification symbols) IPC - C12N 15/10; C12Q 1/68; G01N 33/52 CPC - C12N 15/1082; C12N 15/66; C12Q 1/6844 Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched USPC - 435/473; 435/6.12; 435/91.52; 506/11 (keyword delimited) Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) PatBase, Google Patents, Google Scholar, Google, PubMed Search terms used: guide, gNA, gRNA, gDNA; CRISPR, Cas argonaute; C2C2; NgAgo; adapter, adaptor; methylation; ligation; DpnI		
<b>C. DOCUMENTS CONSIDERED TO BE RELEVANT</b>		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2014/0357523 A1 (AGILENT TECHNOLOGIES, INC.) 04 December 2014 (04.12.2014) entire document	1, 2, 4-6, 45-47, 49-51, 60, 158-161
--		
Y		3, 7, 48, 52, 66-72, 134, 139-141
X	US 2015/0211058 A1 (AGILENT TECHNOLOGIES, INC.) 30 July 2015 (30.07.2015) entire document	25-28, 149-152
--		
Y		29, 153
X	US 2014/0356867 A1 (AGILENT TECHNOLOGIES, INC.) 04 December 2014 (04.12.2014) entire document	117-120, 126, 127
--		
Y		66-72, 134, 139-141
X	WO 2015/070083 A1 (EDITAS MEDICINE, INC.) 14 May 2015 (14.05.2015) entire document	167-170
Y	SWARTS et al. "DNA-guided DNA interference by a prokaryotic Argonaute," Nature, 13 March 2014 (13.03.2014), Vol. 507, Pgs. 258-261. entire document	3, 7, 29, 48, 52, 68, 72, 153
A	US 2015/0031089 A1 (AGILENT TECHNOLOGIES, INC.) 29 January 2015 (29.01.2015) entire document	1-7, 25-29, 45-52, 60, 66-72, 88-95, 103-110, 117-120, 126, 127, 134-141, 149-153, 158-161, 167-170
<input checked="" type="checkbox"/> Further documents are listed in the continuation of Box C. <input type="checkbox"/> See patent family annex.		
* Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier application or patent but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family		
Date of the actual completion of the international search 21 November 2016		Date of mailing of the international search report 20 DEC 2016
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, VA 22313-1450 Facsimile No. 571-273-8300		Authorized officer Blaine R. Copenheaver PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2016/047631

## C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2014/0186843 A1 (MASSACHUSETTS INSTITUTE OF TECHNOLOGY) 03 July 2014 (03.07.2014) entire document	1-7, 25-29, 45-52, 60, 66-72, 88-95, 103-110, 117-120, 126, 127, 134-141, 149-153, 158-161, 167-170
A	WU et al. "OsLLO1, a C2C2-type zinc finger protein, interacts with OsbZIP58 to promote seed germination through the modulation of gibberellin biosynthesis in <i>Oryza sativa</i> ," The Plant Journal, 29 October 2014 (29.10.2014), Vol. 80, Pgs. 1118-1130. entire document	1-7, 25-29, 45-52, 60, 66-72, 88-95, 103-110, 117-120, 126, 127, 134-141, 149-153, 158-161, 167-170