DE NOVO SYNTHESIZED NUCLEIC ACID LIBRARIES

CROSS-REFERENCE

[0001] This application claims the benefit of U.S. Provisional Application No. 62/419,881, filed on November 9, 2016; U.S. Provisional Application No. 62/411,388, filed on October 21, 2016; U.S. Provisional Application No. 62/393,948, filed on September 13, 2016; U.S. Provisional Application No. 62/382,191, filed on August 31, 2016; and U.S. Provisional Application No. 62/378,134, filed on August 22, 2016, each of which is incorporated herein by reference in its entirety.

SEQUENCE LISTING

[0002] The instant application contains a Sequence Listing which has been submitted electronically in ASCII format and is hereby incorporated by reference in its entirety. Said ASCII copy, created on April 4, 2017, is named 44854-727_601_SL.txt and is 13,567 bytes in size.

BACKGROUND

[0003] The cornerstone of synthetic biology is the design, build, and test process — an iterative process that requires DNA to be made accessible for rapid and affordable generation and optimization of these custom pathways and organisms. In the design phase, the A, C, T, and G nucleotides that constitute DNA are formulated into the various sequences that would comprise a region of interest, with each sequence variant representing a specific hypothesis that will be tested. These variant sequences represent subsets of sequence space, a concept that originated in evolutionary biology and pertains to the totality of sequences that make up genes, genomes, transcriptome, and proteome. In the context of targeted genome editing, there exists a need for rapid generation of highly accurate and uniform nucleic acid libraries for specifically directing enzymatic editing of a gene, a gene cluster, a pathway, or an entire genome.

INCORPORATION BY REFERENCE

[0004] All publications, patents, and patent applications mentioned in this specification are herein incorporated by reference to the same extent as if each individual publication, patent, or patent application was specifically and individually indicated to be incorporated by reference.

BRIEF SUMMARY

[0005] Provided herein are systems, methods, and compositions for the efficient de novo synthesis and screening of highly accurate nucleic acid libraries. Nucleic acid libraries as described

herein comprise nucleic acids for specifically targeting and editing a gene, a gene cluster, a biological pathway, or an entire genome.

[0006] Provided herein are nucleic acid libraries, wherein the nucleic acid library comprises at least 500 non-identical DNA molecules, wherein each non-identical DNA molecule encodes for a different gRNA sequence, and wherein at least about 80% of the at least 500 non-identical DNA molecules are each present in the nucleic acid library in an amount within 2x of a mean frequency for each of the non-identical DNA molecules in the library. Provided herein are nucleic acid libraries, wherein each non-identical DNA molecule has a GC base content of about 20% to about 85%. Provided herein are nucleic acid libraries, wherein each non-identical DNA molecule has a GC base content of about 30% to about 70%. Provided herein are nucleic acid libraries, wherein at least about 90% of the at least 500 non-identical DNA molecules are each present in the nucleic acid library in an amount within 2x of the mean frequency for each of the non-identical DNA molecules in the library. Provided herein are nucleic acid libraries, wherein at least 99% of the at least 500 non-identical DNA molecules are each present in the nucleic acid library in an amount within 2x of the mean frequency for each of the non-identical DNA molecules in the library. Provided herein are nucleic acid libraries, wherein the at least 500 non-identical DNA molecules comprises at least 2000 non-identical DNA molecules. Provided herein are nucleic acid libraries, wherein the at least 500 non-identical DNA molecules comprises at least 3500 non-identical DNA molecules. Provided herein are nucleic acid libraries, wherein the at least 500 non-identical DNA molecules comprises at least 100,000 non-identical DNA molecules. Provided herein are nucleic acid libraries, wherein each non-identical DNA molecule comprises up to 200 bases in length. Provided herein are nucleic acid libraries, wherein each non-identical DNA molecule comprises about 100 to about 200 bases in length. Provided herein are nucleic acid libraries, wherein the at least 500 non-identical DNA molecules comprises non-identical DNA molecules encoding for gRNA sequences targeting genes in a biological pathway. Provided herein are nucleic acid libraries, wherein the at least 500 non-identical DNA molecules comprises non-identical DNA molecules encoding for gRNA sequences targeting genes in an entire genome. Provided herein are nucleic acid libraries, wherein the gRNA is a single gRNA or a dual gRNA.

[0006a] Provided herein is a nucleic acid library, wherein the nucleic acid library comprises at least 500 non-identical DNA molecules, wherein each non-identical DNA molecule encodes for a different gRNA sequence, wherein each gRNA sequence comprises a targeting domain complementary to a mammalian gene, and wherein at least about 80% of the at least 500 non-identical DNA molecules are each present in the nucleic acid library in an amount within 2x of a mean frequency for each of the non-identical DNA molecules in the library.

[0006b] Provided herein is a nucleic acid library, wherein the nucleic acid library comprises at least 2000 non-identical nucleic acids, wherein each non-identical nucleic acid encodes for a different gRNA sequence, wherein each gRNA sequence comprises a targeting domain complementary to a eukaryotic gene, and wherein at least about 80% of the at least 2000 non-identical nucleic acids are present in the nucleic acid library in an amount within 2x of a mean frequency for each of the non-identical nucleic acids in the library.

[0006c] Provided herein is an amplicon library, wherein the amplicon library comprises a plurality of non-identical DNA molecules, wherein each non-identical DNA molecule is present in a population of amplification products, wherein each non-identical DNA molecule encodes for a different gRNA sequence, wherein each gRNA sequence comprises a targeting domain complementary to a eukaryotic gene, and wherein at least about 80% of the plurality of non-identical DNA molecules are each present in the amplicon library in an amount within 2x of a mean frequency for each of the non-identical DNA molecules in the library.

[0007] Provided herein are nucleic acid libraries, wherein the nucleic acid library comprises at least 2000 non-identical nucleic acids, wherein each non-identical nucleic acid encodes for a different sgRNA sequence, wherein each sgRNA sequence comprises a targeting domain complementary to a eukaryotic gene, and wherein at least about 80% of the at least 2000 non-identical nucleic acids are present in the nucleic acid library in an amount within 2x of a mean frequency for each of the non-identical nucleic acids in the library. Provided herein are nucleic

acid libraries, wherein each non-identical nucleic acid has a GC base content of about 20% to about 85%. Provided herein are nucleic acid libraries, wherein each non-identical nucleic acid has a GC base content of about 30% to about 70%. Provided herein are nucleic acid libraries, wherein at least about 90% of the at least 2000 non-identical nucleic acids are each present in the nucleic acid library in an amount within 2x of the mean frequency for each of the non-identical nucleic acids in the library. Provided herein are nucleic acid libraries, wherein at least 99% of the at least 2000 non-identical nucleic acids are each present in the nucleic acid library in an amount within 2x of the mean frequency for each of the non-identical nucleic acids in the library. Provided herein are nucleic acid libraries, wherein each non-identical nucleic acid comprises up to 200 bases in length. Provided herein are nucleic acid libraries, wherein each non-identical nucleic acid comprises about 100 to about 200 bases in length. Provided herein are nucleic acid libraries, wherein the at least 2000 non-identical nucleic acids comprise non-identical nucleic acids encoding for sgRNA sequences targeting genes in a biological pathway. Provided herein are nucleic acid libraries, wherein the at least 2000 non-identical nucleic acids comprise non-identical nucleic acids encoding for sgRNA sequences targeting genes in an entire genome. Provided herein are nucleic acid libraries, wherein each non-identical nucleic acid comprises DNA or RNA molecules.

[0008] Provided herein are amplicon libraries, wherein the amplicon library comprises a plurality of non-identical DNA molecules, wherein each non-identical DNA is present in a population of amplification products, wherein each non-identical DNA molecule encodes for a different gRNA sequence, and wherein at least about 80% of the plurality of non-identical DNA molecules are each present in the amplicon library in an amount within 2x of a mean frequency for each of the non-identical DNA molecules in the library. Provided herein are amplicon libraries, wherein each non-identical DNA molecule has a GC base content of about 30% to about 70%. Provided herein are amplicon libraries, wherein the gRNA is a single gRNA or a dual gRNA.

[0009] Provided herein are cell libraries, wherein the cell library comprises a plurality of cell populations, wherein each of the cell populations comprises a DNA molecule encoding for a different gRNA sequence, wherein each gRNA sequence comprises a targeting region for binding to a gene, and wherein at least 15% of the cell populations have at least 2-fold depletion in expression of the gene. Provided herein are cell libraries, wherein at least 45% of the cell populations have at least 2-fold depletion in expression of the gene. Provided herein are cell libraries, wherein the gRNA is a single gRNA or a dual gRNA. Provided herein are cell libraries, wherein the plurality of cell populations comprises DNA molecules encoding for at least 3 different gRNA sequences per a single gene. Provided herein are cell libraries, wherein the plurality of cell populations comprises DNA molecules encoding for at least 5 different gRNA sequences per a

single gene. Provided herein are cell libraries, wherein the plurality of cell populations comprises at least 2000 cell populations. Provided herein are cell libraries, wherein the plurality of cell populations comprises DNA molecules encoding for gRNA sequences in a biological pathway. Provided herein are cell libraries, wherein the plurality of cell populations comprises DNA molecules encoding for gRNA sequences in an entire genome. Provided herein are cell libraries, wherein the genome is *Arabidopsis thaliana, Caenorhabditis elegans, Canis lupus familiaris, Chlamydomonas reinhardtii, Danio rerio, Dictyostelium discoideum, Drosophila melanogaster, Escherichia coli, Homo sapiens, Macaca mulatta, Mus musculus, Oryctolagus cuniculus, Rattus norvegicus, Saccharomyces cerevisiae, or Sus scrofa*. Provided herein are cell libraries, wherein each of the cell populations comprises prokaryotic cells. Provided herein are cell libraries, wherein each of the cell populations comprises eukaryotic cells. Provided herein are cell libraries, wherein each of the cell populations comprises mammalian cells. Provided herein are cell libraries, wherein each of the cell populations further comprises an exogenous nuclease enzyme. Provided herein are cell libraries, wherein the DNA molecule further comprises a vector sequence.

[0010] Provided herein are cell libraries, wherein the cell library comprises a plurality of cell populations, wherein each of the cell populations comprises a DNA molecule encoding for a different gRNA sequence, wherein each gRNA sequence comprises a targeting region for binding to a gene, and wherein at most 20% of the cell populations have a zero or negative depletion in expression of the gene. Provided herein are cell libraries, wherein the gRNA is a single gRNA or a dual gRNA. Provided herein are cell libraries, wherein the plurality of cell populations comprises DNA molecules encoding for at least 3 different gRNA sequences per a single gene. Provided herein are cell libraries, wherein the plurality of cell populations comprises DNA molecules encoding for at least 5 different gRNA sequences per a single gene. Provided herein are cell libraries, wherein the plurality of cell populations comprises at least 2000 cell populations. Provided herein are cell libraries, wherein the plurality of cell populations comprises at least 10000 cell populations.

[0011] Provided herein are methods for synthesis of a gRNA library, comprising: providing predetermined sequences for at least 500 non-identical DNA molecules, wherein each non-identical DNA molecule encodes for a gRNA; synthesizing the at least 500 non-identical DNA molecules; and transcribing the at least 500 non-identical DNA molecules to generate a library of gRNAs, wherein at least about 75% of the gRNAs in the library of gRNAs are error free compared to the predetermined sequences for the at least 500 non-identical DNA molecules. Provided herein are methods for synthesis of a gRNA library, further comprising transferring the at least 500 non-identical DNA molecules into cells prior to the transcribing step. Provided herein are methods for

synthesis of a gRNA library, wherein at least 96% of the gRNAs encoded by the at least 500 non-identical DNA molecules are present in the library of gRNAs. Provided herein are methods for synthesis of a gRNA library, wherein at least 87% of the gRNAs in the library of gRNAs are error free compared to the predetermined sequences for the at least 500 non-identical DNA molecules. Provided herein are methods for synthesis of a gRNA library, further comprising inserting the at least 500 non-identical DNA molecules into vectors. Provided herein are methods for synthesis of a gRNA library, further comprising transferring the at least 500 non-identical DNA molecules to cells of an organism. Provided herein are methods for synthesis of a gRNA library, wherein the organism is *Arabidopsis thaliana, Caenorhabditis elegans, Canis lupus familiaris, Chlamydomonas reinhardtii, Danio rerio, Dictyostelium discoideum, Drosophila melanogaster, Escherichia coli, Homo sapiens, Macaca mulatta, Mus musculus, Oryctolagus cuniculus, Rattus norvegicus, Saccharomyces cerevisiae, or Sus scrofa*. Provided herein are methods for synthesis of a gRNA library, wherein each non-identical DNA molecule encodes for a single gRNA or a dual gRNA.

[0012] Provided herein are methods for synthesis of a gRNA library, comprising: providing predetermined sequences for a plurality of non-identical DNA molecules, wherein each non-identical DNA molecule encodes for a gRNA; providing a surface, wherein the surface comprises clusters of loci for nucleic acid extension reaction; synthesizing the plurality of non-identical DNA molecules, wherein each non-identical DNA molecule extends from the surface; and transferring the plurality of non-identical DNA molecules into cells. Provided herein are methods for synthesis of a gRNA library, wherein each cluster comprises about 50 to about 500 loci. Provided herein are methods for synthesis of a gRNA library, wherein each non-identical DNA molecule comprises up to about 200 bases in length. Provided herein are methods for synthesis of a gRNA library, wherein each non-identical DNA molecule encodes for a single gRNA or a dual gRNA. Provided herein are methods for synthesis of a gRNA library, wherein the cells are prokaryotic cells. Provided herein are methods for synthesis of a gRNA library, wherein the cells are eukaryotic cells. Provided herein are methods for synthesis of a gRNA library, wherein the eukaryotic are mammalian cells. Provided herein are methods for synthesis of a gRNA library, wherein each of the cells comprises an exogenous nuclease enzyme.

BRIEF DESCRIPTION OF THE DRAWINGS

[0013] **Figure 1A** illustrates a Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) complex which includes the following components: PAM, target sequence, CAS9 enzyme, Guide RNA (gRNA), and donor DNA.

[0014] **Figure 1B** illustrates a Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) complex which includes the following components: PAM, target sequence, CAS9

enzyme, Guide RNA (gRNA), and donor DNA for a non-homologous end joining repair (NHEJ) pathway.

[0015] **Figure 2** illustrates a gRNA library screening workflow, including design, synthesis, cloning, packaging, screening and analysis of a gRNA library.

[0016] **Figure 3** illustrates gRNA library screening workflow for building a library, including: synthesizing an oligonucleic acid library on an array, amplifying and transferring the oligonucleic acids into vectors, and forming an expression library for gRNA expression.

[0017] **Figures 4A-4C** are diagrams of various gRNAs. **Figure 4A** is diagram of a sgRNA sequence (SEQ ID NO: 40) having a base-pairing region, a dCas9 handle, and a *S. pyogenes* terminator region. **Figure 4B** is a diagram of a sgRNA alone. **Figure 4C** is a diagram of a dgRNA alone.

[0018] **Figure 5A** is a diagram of a sgRNA sequence in a template strand targeting arrangement.

[0019] **Figure 5B** is a diagram of a sgRNA sequence in a non-template strand targeting arrangement.

[0020] **Figure 6A** is a diagram of a gRNA sequence with a T7 promoter that, when transcribed, results in gRNA sequence that forms hairpin secondary structure.

[0021] **Figure 6B** is a diagram of a gRNA sequence with a T7 promoter that, when transcribed, results in gRNA sequence that does not form a hairpin secondary structure.

[0022] **Figure 7** depicts a workflow for *in vitro* Cas9 mediated cleavage of target DNA.

[0023] **Figure 8** illustrates an example of a computer system.

[0024] **Figure 9** is a block diagram illustrating an example architecture of a computer system.

[0025] **Figure 10** is a diagram demonstrating a network configured to incorporate a plurality of computer systems, a plurality of cell phones and personal data assistants, and Network Attached Storage (NAS).

[0026] **Figure 11** is a block diagram of a multiprocessor computer system using a shared virtual address memory space.

[0027] **Figure 12** depicts 4 sgRNA designs. **Figure 12** discloses SEQ ID NOS 20, 15, 21, 15, 22, 15, 41, 15, 42, and 43, respectively, in order of appearance.

[0028] **Figures 13A-13B** are plots from a BioAnalyzer reading, with nucleotide bases on the X axis and fluorescent units on the Y axis.

[0029] **Figures 14A-14J** are plots from a BioAnalyzer reading, with nucleotide bases on the X axis and fluorescent units on the Y axis.

[0030] **Figure 15** is an image of a 256 clusters, each cluster having 121 loci with oligonucleic acids extending therefrom.

[0031] **Figure 16A** is a plot of oligonucleic acid representation (oligonucleic acid frequency v. absorbance) across a plate from synthesis of 29,040 unique oligonucleic acids from 240 clusters, each cluster having 121 oligonucleic acids.

[0032] **Figure 16B** is a plot of measurement of oligonucleic acid frequency v. absorbance across each individual cluster, with control clusters identified by a box.

[0033] **Figure 17** is a plot of measurements of oligonucleic acid frequency v. absorbance across four individual clusters.

[0034] **Figure 18A** is a plot of on error rate v. frequency across a plate from synthesis of 29,040 unique oligonucleic acids from 240 clusters, each cluster having 121 oligonucleic acids.

[0035] **Figure 18B** is a plot of measurement of oligonucleic acid error rate v. frequency across each individual cluster, with control clusters identified by a box.

[0036] **Figure 19** is a plot of measurements of oligonucleic acid error rate v. frequency across four clusters.

[0037] **Figure 20** is a plot of GC content as a measure of percent per oligonucleic acid v. the number of oligonucleic acids.

[0038] **Figure 21** provides plots with results from PCR with two different polymerases. Each chart depicts "observed frequency" ("0 to 35" measured in counts per 100,000) v. number of oligonucleic acids (0 to 2000).

[0039] **Figure 22** provides a chart with quantification of oligonucleic acid population uniformity post amplification that was recorded.

[0040] **Figure 23** depicts a plot of impact of over amplification on sequence dropouts.

[0041] **Figures 24A-24B** depict results from sequencing recovered oligonucleic acids from a 10,000 sgRNA oligonucleic acid CRISPR library.

[0042] **Figure 25** depicts results from sequencing recovered oligonucleic acids from a 101,000 sgRNA oligonucleic acid CRISPR library.

[0043] **Figure 26A** depicts a graph of percentage of sgRNAs with at least 2-fold depletion.

[0044] **Figure 26B** depicts a graph of percentage of sgRNAs with zero or negative depletion.

DETAILED DESCRIPTION

[0045] Provided herein are systems, methods, and compositions for the efficient synthesis and screening of highly accurate guide RNA ("gRNA") libraries. De novo synthesis methods described herein provide for a rapid and highly accurate generation of large libraries of gRNA for incorporation into enzymatic systems for targeted gene editing.

[0046] Definitions

[0047] Throughout this disclosure, various embodiments are presented in a range format. It should be understood that the description in range format is merely for convenience and brevity and should not be construed as an inflexible limitation on the scope of any embodiments. Accordingly, the description of a range should be considered to have specifically disclosed all the possible subranges as well as individual numerical values within that range to the tenth of the unit of the lower limit unless the context clearly dictates otherwise. For example, description of a range such as from 1 to 6 should be considered to have specifically disclosed subranges such as from 1 to 3, from 1 to 4, from 1 to 5, from 2 to 4, from 2 to 6, from 3 to 6 etc., as well as individual values within that range, for example, 1.1, 2, 2.3, 5, and 5.9. This applies regardless of the breadth of the range. The upper and lower limits of these intervening ranges may independently be included in the smaller ranges, and are also encompassed within the invention, subject to any specifically excluded limit in the stated range. Where the stated range includes one or both of the limits, ranges excluding either or both of those included limits are also included in the invention, unless the context clearly dictates otherwise.

[0048] The terminology used herein is for the purpose of describing particular embodiments only and is not intended to be limiting of any embodiment. As used herein, the singular forms “a,” “an” and “the” are intended to include the plural forms as well, unless the context clearly indicates otherwise. It will be further understood that the terms “comprises” and/or “comprising,” when used in this specification, specify the presence of stated features, integers, steps, operations, elements, and/or components, but do not preclude the presence or addition of one or more other features, integers, steps, operations, elements, components, and/or groups thereof. As used herein, the term “and/or” includes any and all combinations of one or more of the associated listed items.

[0049] Unless specifically stated or obvious from context, as used herein, the term “about” in reference to a number or range of numbers is understood to mean the stated number and numbers +/- 10% thereof, or 10% below the lower listed limit and 10% above the higher listed limit for the values listed for a range.

[0050] The term “gRNA” as referred to herein refers to guide RNA sequence and encompasses both single and dual guide RNA sequence. Unless specifically stated or obvious from context, as used herein, the term “dgRNA” as referred to herein refers to dual guide RNA sequence: crRNA (spacer sequence comprising a seed region complementary to a target sequence) and a separate tracrRNA (trans-activating sequence), which are partially complementary RNAs. Unless specifically stated or obvious from context, as used herein, the term “sgRNA” as referred to herein refers to single guide RNA sequence, comprising both a fused crRNA and tracrRNA.

[0051] Unless specifically stated or obvious from context, as used herein, the terms "oligonucleic acid" and "nucleic acid" encompass double- or triple-stranded nucleic acids, as well as single-stranded molecules. In double- or triple-stranded nucleic acids, the nucleic acid strands need not be coextensive (i.e., a double-stranded nucleic acid need not be double-stranded along the entire length of both strands). Nucleic acid sequences, when provided, are listed in the 5' to 3' direction, unless stated otherwise. Methods described herein provide for the generation of isolated nucleic acids. Methods described herein additionally provide for the generation of isolated and purified nucleic acids. An "oligonucleic acid" and "nucleic acid" as referred to herein can comprise at least 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 125, 150, 175, 200, 225, 250, 275, 300, 325, 350, 375, 400, 425, 450, 475, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000, or more bases in length.

[0052] Unless specifically stated or obvious from context, as used herein, the term "amplicon" as used herein refers to an amplification reaction product.

[0053] The present disclosure employs, unless otherwise indicated, conventional molecular biology techniques, which are within the skill of the art. Unless defined otherwise, all technical and scientific terms used herein have the same meaning as is commonly understood by one of ordinary skill in the art to which these inventions belong.

[0054] ***gRNA Library Screening***

[0055] Provided herein are methods for designing, building, and screening a library of highly accurate gRNAs for incorporation in a Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-enzyme complex. *See, e.g.*, **FIGS. 1A-AB**. gRNA libraries generated using methods described herein include both sgRNA and dgRNA libraries. Provided herein are methods for highly uniform synthesis resulting in high representation of predetermined gRNAs in the resulting libraries. In the design phase, gRNAs are designed. *See* **FIG. 2**. Design strategies include, without limitation, design of gRNAs to span a gene. Depending on the desired workflow, the de novo synthesized nucleic acids are DNA or RNA bases.

[0056] In the case of de novo synthesized DNA, a library comprising nucleic acids is synthesized, wherein each nucleic acid synthesized is a DNA sequence that encodes for a gRNA (*e.g.*, sgRNA) sequence as a transcription product. In some instances, the synthesized nucleic acids are then inserted into expression vectors. In one exemplary workflow, the synthesized nucleic acids are inserted into viral vectors, and then packaged for transduction into cells, followed by screening and analysis. **FIG. 2**. Exemplary cells include without limitation, prokaryotic and eukaryotic cells. Exemplary eukaryotic cells include, without limitation, animal, plant, and fungal cells. Exemplary animal cells include, without limitation, insect, fish and mammalian cells.

Exemplary mammalian cells include mouse, human, and primate cells. Exemplary cellular functions tested include, without limitation, changes in cellular proliferation, migration/adhesion, metabolic, and cell-signaling activity. In the case of de novo synthesized RNA, the gRNA itself is synthesized and available for downstream applications, such as transfection into cells.

[0057] Oligonucleic acids may be synthesized within a cluster **303** of locations ("loci") for extension on an array **301**. *See* **FIG. 3**. Such an arrangement may provide for improved oligonucleic acid representation of products from amplification of the synthesized oligonucleic acids –termed "amplicons"–when compared to amplification products of oligonucleic acids synthesized across an entire plate without a clustered loci arrangement. In some instances, amplification **310** of oligonucleic acids synthesized within a single cluster counters negative effects on representation due to repeated synthesis of large oligonucleic acid populations having oligonucleic acids with heavy GC content, commonly termed "drift," due to underrepresentation of GC low or GC high amplicons in the amplification reaction product. In some instances, the single cluster described herein, comprises about 50-1000, 75-900, 100-800, 125-700, 150-600, 200-500, or 300-400 discrete loci. In some instances, the single cluster comprises 50-500 discrete loci. In some instances, a locus is a spot, well, microwell, channel, or post. In some instances, each cluster has at least 1X, 2X, 3X, 4X, 5X, 6X, 7X, 8X, 9X, 10X, or more redundancy of separate features supporting extension of oligonucleic acids having identical sequence.

[0058] Provided herein are gRNA libraries for insertion into expression vectors. Continuing the workflow in **FIG. 3**, an array **301** includes multiple clusters **303** of loci for oligonucleic acid synthesis and extension. De novo DNA is synthesized and removed from the plate to form a population of oligonucleic acids **305** (e.g., DNAs encoding for sgRNAs), which are subject to amplification **310** to form a library of amplified oligonucleic acids **320** for insertion into a vector **330** to form a library of vectors including the synthesized DNAs **335**. Once in the cells, the DNAs are transcribed into gRNAs (e.g., sgRNAs) and are available for binding with genomic editing regime (e.g., a Cas9-based system). The cells may have natural or ectopic expression of the editing enzyme (e.g., Cas9). The editing enzyme (e.g., Cas9) may have double DNA strand cleavage activity, or a modified activity, such as nicking, base swapping or sequence swapping activity. The synthesized DNA for insertion into a vector may comprise sgRNAs, dgRNAs, or fragments thereof.

[0059] Expression vectors for inserting nucleic acid libraries disclosed herein comprise eukaryotic or prokaryotic expression vectors. Exemplary expression vectors include, without limitation, mammalian expression vectors: pSF-CMV-NEO-NH2-PPT-3XFLAG, pSF-CMV-NEO-COOH-3XFLAG, pSF-CMV-PURO-NH2-GST-TEV, pSF-OXB20-COOH-TEV-FLAG(R)-6His, pCEP4 pDEST27, pSF-CMV-Ub-KrYFP, pSF-CMV-FMDV-daGFP, pEF1a-mCherry-N1 Vector,

pEF1a-tdTomato Vector, pSF-CMV-FMDV-Hygro, pSF-CMV-PGK-Puro, pMCP-tag(m), and pSF-CMV-PURO-NH2-CMYC; bacterial expression vectors: pSF-OXB20-BetaGal,pSF-OXB20-Fluc, pSF-OXB20, and pSF-Tac; plant expression vectors: pRI 101-AN DNA and pCambia2301; and yeast expression vectors: pTYB21 and pKLAC2, and insect vectors: pAc5.1/V5-His A and pDEST8.

[0060] De novo oligonucleic acid libraries synthesized by methods described herein may be expressed in cells. In some instances, the cells are associated with a disease state. For example, cells associated with a disease state include, but not limited to, cell lines, tissue samples, primary cells from a subject, cultured cells expanded from a subject, or cells in a model system. In some instances, the model system is a plant or animal system. In some instances, the de novo oligonucleic acid libraries are expressed in cells to assess for a change in cellular activity. Exemplary cellular activities include, without limitation, proliferation, cycle progression, cell death, adhesion, migration, reproduction, cell signaling, energy production, oxygen utilization, metabolic activity, aging, response to free radical damage, or any combination thereof.

[0061] Provided herein are methods for synthesizing a gRNA library (or a DNA library that when transcribed results in a gRNA library), wherein the gRNA library comprises a plurality of non-identical gRNAs per a gene. The gRNA may encode a sgRNA or a dgRNA. In some instances, the gRNA library comprises at least 3, 4, 5, 6, 7, 8, 9, 10, or more than 10 non-identical gRNAs per the gene. In some instances, the gRNA library targets one or more genes. In some instances, the gRNA library targets about 1, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, or more than 100 genes. In some instances, the gRNA library targets about 1-100, 2-95, 5-90, 10-85, 15-80, 20-75, 25-70, 30-65, 35-60, or 40-50 genes. In some instances, the gRNA library described herein targets genes in a pathway. Exemplary pathways include, without limitation a metabolic, cell death, cell cycle progression, immune cell activation, inflammatory response, angiogenesis, lymphogenesis, hypoxia and oxidative stress response, cell adhesion, and cell migration pathways.

[0062] Methods for synthesizing a gRNA library as described herein may provide for synthesis of non-identical gRNAs having a base-pairing region complementary to part of a genome, a genome target region. The genome target region may comprise exon, intron, coding, or non-coding sequence. In some instances, the gRNA library comprises non-identical gRNAs collectively having a base-pairing region complementary to at least or about 5% of the genes in an entire genome. In some instances, the gRNA library comprises non-identical gRNAs collectively having a base-pairing region complementary to at least or about 80% of the genes in an entire genome. In some instances, the gRNA library comprises non-identical gRNAs collectively having a base-pairing region complementary to at least or about 90% of the genes in an entire genome. In some

instances, the gRNA library comprises non-identical gRNAs collectively having a base-pairing region complementary to at least or about 95% of the genes in an entire genome. In some instances, the gRNA library comprises non-identical gRNAs collectively having a base-pairing region complementary to at least or about 100% of the genes in an entire genome.

[0063] Provided herein are gRNA libraries synthesized by methods described herein that result in gRNAs with at least 2X depletion of a gene across different cells. In some instances, the gRNA libraries comprise at least or about 10%, 12%, 15%, 16%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65% or more of gRNAs that provide for at least 2X depletion a gene when present in cells or in a plurality of cell populations. In some instances, the gene is an essential gene, i.e. a gene critical for cell survival. Exemplary essential genes include, without limitation, PCNA, PSMA7, RPP21, and SF3B3. In some instances, the gRNA libraries comprise gRNAs that provide for at least 2X, 3X, 4X, 5X, 6X, or more than 6X depletion of a gene when present in cells. In some instances, the gRNA libraries comprise at most 5%, 10%, 12%, 15%, or 20% of the gRNAs with zero or negative depletion of the gene when present in cells or in a plurality of cell populations. In some instances, the plurality of cell populations comprises at least or about 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000, 12000, 15000, 20000, 25000, 30000, or more than 30000 cell populations. In some instances, the gRNA libraries comprise gRNAs with at least 2X, 3X, 4X, 5X, 6X, or more than 6X depletion for the plurality of genes. In some instances, the gRNA libraries comprise an average of at least or about 40%, 45%, 50%, 60%, 65%, 70%, 75%, 80% or more than 90% of gRNAs providing at least 2X depletion for the plurality of genes. The gRNAs providing such gene deletion profiles can be sgRNAs or dgRNAs.

[0064] Provided herein are methods for synthesizing highly uniform libraries of oligonucleic acids. In some cases, more than 90% of synthesized oligonucleic acids (RNA or DNA) are represented within 4X of the mean for oligonucleic acid representation for a nucleic acid library. In some cases, more than 90% of oligonucleic acids are represented within 2X of the mean for oligonucleic acid representation for the library. In some cases, more than 90% of oligonucleic acids are represented within 1.5X of the mean for oligonucleic acid representation for the library. In some cases, more than 80% of oligonucleic acids are represented within 1.5X of the mean for oligonucleic acid representation for the library.

[0065] Oligonucleic acid libraries de novo synthesized by methods described herein comprise a high percentage of correct sequences compared to predetermined sequences. In some instances, de novo oligonucleic acids libraries disclosed herein have greater than 70% correct sequence compared to predetermined sequences for oligonucleic acids. In some instances, de novo oligonucleic acids libraries disclosed herein have greater than 75% correct sequence compared to

predetermined sequences for the oligonucleic acids. In some instances, de novo oligonucleic acids libraries disclosed herein have greater than 80% correct sequence compared to predetermined sequences for the oligonucleic acids. In some instances, de novo oligonucleic acids libraries disclosed herein have greater than 85% correct sequence compared to predetermined sequences for the oligonucleic acids. In some instances, de novo oligonucleic acids libraries disclosed herein have greater than 90% correct sequence compared to predetermined sequences for the oligonucleic acids. In some instances, de novo oligonucleic acids libraries disclosed herein have greater than 95% correct sequence compared to predetermined sequences for the oligonucleic acids. In some instances, de novo oligonucleic acids libraries disclosed herein have greater than 100% correct sequence compared to predetermined sequences for the oligonucleic acids.

[0066] In some instances, de novo synthesized oligonucleic acids libraries disclosed herein have greater than 70% correct sequence compared to predetermined sequences for the oligonucleic acids following an amplification reaction. In some instances, de novo synthesized oligonucleic acids libraries disclosed herein have greater than 75% correct sequence compared to predetermined sequences for the oligonucleic acids following an amplification reaction. In some instances, de novo synthesized oligonucleic acids libraries disclosed herein have greater than 80% correct sequence compared to predetermined sequences for the oligonucleic acids following an amplification reaction. In some instances, de novo synthesized oligonucleic acids libraries disclosed herein have greater than 85% correct sequence compared to predetermined sequences for the oligonucleic acids following an amplification reaction. In some instances, de novo synthesized oligonucleic acids libraries disclosed herein have greater than 90% correct sequence compared to predetermined sequences for the oligonucleic acids following an amplification reaction. In some instances, de novo synthesized oligonucleic acids libraries disclosed herein have greater than 95% correct sequence compared to predetermined sequences for the oligonucleic acids following an amplification reaction. In some instances, de novo synthesized oligonucleic acids libraries disclosed herein have 100% correct sequence compared to predetermined sequences for the oligonucleic acids following an amplification reaction.

[0067] In some instances, de novo synthesized oligonucleic acids libraries disclosed herein, when transferred into cells, results in greater than 80% correct sequence compared to predetermined sequences for the oligonucleic acids. In some instances, de novo synthesized oligonucleic acids libraries disclosed herein, when transferred into cells, results in greater than 85% correct sequence compared to predetermined sequences for the oligonucleic acids. In some instances, de novo synthesized oligonucleic acids libraries disclosed herein, when transferred into cells, results in greater than 90% correct sequence compared to predetermined sequences for the

oligonucleic acids. In some instances, de novo synthesized oligonucleic acids libraries disclosed herein, when transferred into cells, results in greater than 95% correct sequence compared to predetermined sequences for the oligonucleic acids. In some instances, de novo synthesized oligonucleic acids libraries disclosed herein, when transferred into cells, results in 100% correct sequence compared to predetermined sequences for the oligonucleic acids.

[0068] In some instances, de novo synthesized oligonucleic acids libraries disclosed herein, when transferred into cells, result in greater than 80% sequence representation. In some instances, de novo synthesized oligonucleic acids libraries disclosed herein, when transferred into cells, result in greater than 90% sequence representation. In some instances, de novo synthesized oligonucleic acids libraries disclosed herein, when transferred into cells, result in greater than 95% sequence representation. In some instances, de novo synthesized oligonucleic acids libraries disclosed herein, when transferred into cells, result in 100% sequence representation.

[0069] De novo oligonucleic acid libraries described herein may be subject to amplification reactions with the addition of a polymerase enzyme and amplification reagents (e.g., buffers, phosphates, and dNTPs). In some instances, the de novo oligonucleic acid libraries are amplified by PCR for at least or about 6, 8, 10, 15, 20, or more than 20 cycles. In some instances, the de novo oligonucleic acid libraries are amplified by PCR in a range of about 6 to 20, 7 to 18, 8 to 17, 9 to 16, or 10 to 15 cycles. In some instances, the de novo oligonucleic acid libraries are amplified by PCR for about 15 cycles.

[0070] In some instances, amplification of the de novo oligonucleic acid libraries provides for an amplicon library of DNA molecules. In some instances, the amplicon library comprises non-identical nucleic acids that encode for a gRNA sequence. In some instances, the gRNA sequence is a sgRNA or a dgRNA.

[0071] In some instances, the de novo oligonucleic acid libraries comprise non-identical nucleic acids, wherein each non-identical nucleic acid comprises DNA molecules. In some instances, the number of DNA molecules is about 500, 2000, 3500 or more molecules. In some instances, the number of DNA molecules is at least or about 250, 500, 1000, 1250, 2000, 2500, 3000, 3500, 4000, 4500, 5000, 5500, 6000, 6500, 7000, 7500, 8000, 8500, 9000, 9500, 10000, 15000, 20000, 50000, 100000, 250000, 500000, 750000, 1 million, or more than 1 million molecules. In some instances, the number of DNA molecules is at most 250, 500, 1000, 1250, 2000, 2500, 3000, 3500, 4000, 4500, 5000, 5500, 6000, 6500, 7000, 7500, 8000, 8500, 9000, 9500, 10000, 15000, 20000, 50000, 100000, 250000, 500000, 750000, 1 million, or more than 1 million molecules. In some instances, the DNA molecule encodes for a gRNA sequence. In some instances, the gRNA sequence is a sgRNA or a dgRNA.

[0072] In some instances, the de novo oligonucleic acid libraries comprise non-identical nucleic acids, wherein each non-identical nucleic acid comprises RNA molecules. In some instances, the number of RNA molecules is about 2000 molecules. In some instances, the number of RNA molecules is at least or about 250, 500, 1000, 1250, 2000, 2500, 3000, 3500, 4000, 4500, 5000, 5500, 6000, 6500, 7000, 7500, 8000, 8500, 9000, 9500, 10000, 15000, 20000, 50000, 100000, 250000, 500000, 750000, 1 million, or more than 1 million molecules. In some instances, the number of RNA molecules is at most 250, 500, 1000, 1250, 2000, 2500, 3000, 3500, 4000, 4500, 5000, 5500, 6000, 6500, 7000, 7500, 8000, 8500, 9000, 9500, 10000, 15000, 20000, 50000, 100000, 250000, 500000, 750000, 1 million, or more than 1 million molecules. In some instances, the RNA molecule encodes for a gRNA sequence. In some instances, the gRNA sequence is a sgRNA or a dgRNA.

[0073] Provided herein are de novo oligonucleic acid libraries having high uniformity following amplification. In some instances, more than 80% of oligonucleic acids in a de novo oligonucleic acid library described herein are represented within at least about 1.5X the mean representation for the entire library following amplification. In some instances, more than 90% of oligonucleic acids in a de novo oligonucleic acid library described herein are represented within at least about 1.5X the mean representation for the entire library following amplification. In some instances, more than 80% of oligonucleic acids in a de novo oligonucleic acid library described herein are represented within at least about 2X the mean representation for the entire library following amplification. In some instances, more than 80% of oligonucleic acids in a de novo oligonucleic acid library described herein are represented within at least about 2X the mean representation for the entire library following amplification.

[0074] An unamplified population of oligonucleic acids de novo synthesized using methods described herein can vary in a number of non-identical oligonucleic acid sequences. In some instances, the number of non-identical oligonucleic acid sequences is in a range of about 2000-1 million, 3000 to 900000, 4000-800000, 5000-700000, 6000-600000, 7000-500000, 8000-400000, 9000-300000, 10000-200000, 11000-100000, 12000-75000, 14000-60000, and 20000-50000 sequences. In some cases, the number of non-identical oligonucleic acid sequences is in the range of about 50-2000, 75-1800, 100-1700, 150-1600, 200-1500, 250-1400, 300-1300, 400-1200, 500-1100, 600-1000, 700-900 sequences. In some instances, the number of non-identical oligonucleic acid sequences is 2000 sequences. In some instances, the number of non-identical oligonucleic acid sequences is more than 1 million sequences. In some instances, the number of non-identical oligonucleic acid sequences is at least 50, 100, 200, 400, 600, 800, 1000, 1200, 1400, 1600, 1800, 2000, 3000, 5000, 7000, 10000, 20000, 30000, 50000, 100000, 500000, 700000, 1000000,

10000000, 1000000000, or more sequences. In some instances, the number of non-identical oligonucleic acids sequence is up to 50, 100, 200, 400, 600, 800, 1000, 1200, 1400, 1600, 1800, 2000, 3000, 5000, 7000, 10000, 20000, 30000, 50000, 100000, 500000, 700000, 1000000, or more sequences. In some instances, the number of non-identical oligonucleic acid sequences is at most 50, 100, 200, 400, 600, 800, 1000, 1200, 1400, 1600, 1800, 2000, 3000, 5000, 7000, 10000, 20000, 30000, 50000, 100000, 500000, 700000, and 1000000 sequences.

[0075] An oligonucleic acid of an unamplified population may be present in varying amounts. In some instances, an oligonucleic acid of an unamplified population is present in an amount of at least or about 0.25 femtomole. In some instances, an oligonucleic acid of an unamplified population is present in an amount of at least or about 1 femtomole. In some instances, an oligonucleic acid of an unamplified population is present in an amount of at least 0.25, 1, 10, 20, 30, 40, 50, 100, 250, 500, 750, 1000, or more than 1000 femtomoles. In some instances, an oligonucleic acid of an unamplified population is present in an amount of at most 0.25, 1, 10, 20, 30, 40, 50, 100, 250, 500, 750, and 1000 femtomoles.

[0076] Provided herein are methods for synthesizing libraries of non-identical oligonucleic acids, wherein a sequence length or average sequence length of the non-identical oligonucleic acids vary. In some cases, the sequence length or average sequence length of the non-identical oligonucleic acids is up to 150 bases. In some cases, the sequence length or average sequence length of the non-identical oligonucleic acids is in a range of about 100 to about 200 bases. In some instances, the sequence length or average sequence length of the non-identical oligonucleic acids is at least 30, 50, 100, 125, 150, 200, 250, 300, 350, 400, 450, 500, or more than 500 bases. In some instances, the sequence length or average sequence length of the non-identical oligonucleic acids is at most 150, 200, 250, 300, 350, 400, 450, or 500 bases. An exemplary sequence length of the non-identical oligonucleic acid is in a range of about 25 to about 150 or about 50 to about 200 bases. In some cases, the sequence length or average sequence length of the non-identical oligonucleic acids is in the range of about 125 to about 200 or about 150 to about 200 bases.

[0077] ***Guide RNA Sequences***

[0078] Provided herein are single guide RNA (sgRNA) sequences for directing a genomic sequence editing enzyme (e.g., Cas9) to a particular target nucleic acid sequence. An example sgRNA in complex with a Cas9 enzyme is illustrated in **FIG. 4A**, and an example alone in **FIG. 4B**. The gRNA may be a dual guide RNA, as illustrated in **FIG. 4C**. Guide sequences disclosed herein comprises a base-pairing region. The base-pairing region comprises a seed region for binding to a target sequence and, optionally, a spacer region. The base-pairing region may vary in length. For example, the base-pairing region may comprise about 1 to 10, 1 to 20, 20 to 25, or 1 to

30 bases in length. In some instance, the base-pairing region comprises at least 10, 15, 20, 25, 30 or more bases in length. In some instances, the base-pairing region comprises a seed region of at least 10 bases in length. The seed region may comprise about 8 to 20 bases in length. In some instances, the seed region is about 12 bases in length. In some instances, a base-pairing region described herein is designed to target a template strand during transcription, **FIG. 5A**. In some instances, a base-pairing region described herein is designed to target a non-template strand during transcription, **FIG. 5B**.

[0079] In some instances, 3' of the base-pairing region of a sgRNA is a Cas9 handle region for binding to Cas9. In some instances, the Cas9 handle region is a dCas9 handle region for binding to a dCas9 enzyme. The handle region may vary in length. For example, the handle region may comprise about 1 to 50, 20 to 45, or 15 to 60 bases in length. In some instance, the handle region comprises at least 35, 40, 45, 50 or more bases in length. The handle region may comprise about 42 bases in length.

[0080] In some instances, 3' of the handle region of the sgRNA is a terminator region. In some instances, the terminator region is a *S. pyogenes* terminator region. In some instances, the terminator region comprises at about 40 bases in length. In some instances, the terminator region comprises about 10 to 50, 20 to 60, or 30 to 55 bases in length.

[0081] Design schemes for gRNA sequences described herein may comprise inclusion of a DNA dependent RNA polymerase promoter region 5' upstream of DNA encoding for the gRNA sequence. Exemplary DNA dependent RNA polymerase promoter regions include, without limitation a T3 and a T7 RNA polymerase promoter sequence. For example, **FIG. 6A** illustrates an arrangement where a T7 promoter region is 5' upstream of a gRNA and the resultant gRNA transcribed is produced wherein the gRNA includes hairpins. In some arrangements, a gRNA is designed to lack a sequence that forms a hairpin secondary structure, **FIG. 6B**. The hairpin secondary structure may be lacking in the Cas9 handle and/or the terminator region.

[0082] Provided herein are dgRNAs for directing a genomic sequence editing enzyme (e.g., Cas9) to a particular target nucleic acid sequence. In some instances, libraries comprises oligonucleic acid sequences that encode sequences for dgRNAs. In some instances, the libraries comprise nucleic acids, wherein each nucleic acid synthesized is a DNA sequence that encodes for a dgRNA sequence as a transcription product. In some instances, the libraries comprise nucleic acids, wherein each nucleic acid synthesized is a RNA sequence and the dgRNA itself is synthesized. In some instances, libraries of dgRNAs comprise oligonucleic acid sequences for crRNA and tracrRNA that are synthesized as separate oligonucleic acids. In some instances, the oligonucleic acid nucleic acids encode for crRNA and tracrRNA separately. In some instances, the

oligonucleic acid nucleic acids encode for single sequence that when transcribed result in a separate crRNA sequence and a separate tracrRNA sequence. Exemplary sequences for crRNA and tracrRNA are seen in **Table 1**.

Table 1.

SEQ ID NO	Name	Sequence
5	crRNA-sp2	5'ATAACTCAATTTGTAAAAAAGTTTTAGAGCTATGCTGTTTTG3'
6	tracrRNA	5'GGAACCATTCAAAACAGCATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTT3'

[0083] gRNA libraries described herein may be used for *in vitro* screening and analysis. An illustration of such an arrangement is depicted in **FIG. 7**, where a target double-stranded DNA sequence is incubated with a gRNA sequence and Cas9 enzyme. The mixture results in a double strand DNA break. The DNA break may result in a measureable change in the function or expression of a genomic element. gRNAs described herein, or DNA encoding for gRNAs, may be added to cells via various methods known in the art, including, without limitation, transfection, transduction, or electroporation.

[0084] In some instances, gRNA libraries described herein are used for in vivo or ex vivo screening and analysis. Cells for screening include primary cells taken from living subjects or cell lines. Cells may be from prokaryotes (e.g., bacteria and fungi) or eukaryotes (e.g., animals and plants). Exemplary animal cells include, without limitation, those from a mouse, rabbit, primate, and insect. In some instances, gRNA libraries described herein may also be delivered to a multicellular organism. Exemplary multicellular organisms include, without limitation, a plant, a mouse, rabbit, primate, and insect.

[0085] ***Genome Engineering***

[0086] Provided herein are libraries comprising nucleic acids for nuclease targeting of a particular target nucleic acid sequence. In some instances, libraries described herein comprise synthesized nucleic acids, wherein the nucleic acids is DNA, RNA, any analogs, or derivatives thereof. In some instances, the target nucleic acid sequence comprises DNA, RNA, any analogs, or derivatives thereof. In some instances, the nuclease cleaves the target nucleic acid sequence. In some instances, the nuclease binds the target nucleic acid but does not cleave it. Types of nucleases include, but are not limited to, Transcription Activator-Like Effector Nuclease (TALEN), zinc finger nuclease (ZFN), meganuclease, Argonaute, and Clustered Regularly Interspaced Short

Palindromic Repeats (CRISPR)-associated (Cas) protein. In some instances, the nuclease is wild-type, genetically modified, or recombinant.

[0087] A model system for targeted gene editing comprises a Cas9-based approach. When expressed or transferred into cells alongside a gRNA, Cas9 allows for the targeted introduction or deletion of genetic information via a complex with a Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) sequence of mRNA. A Cas9 complex, as illustrated in **FIGS. 1A-1B**, includes a Cas9 protein, engaged with a guide RNA ("gRNA") transcript. The illustrated gRNA contains a target sequence region, a PAM region, and a hairpin region. In a CRISPR/Cas9 process, a gRNA shepherds the Cas9 enzyme to a specific stretch of DNA. While the gRNA depicted is a sgRNA (single stranded guide RNA), the complex may be formed with a dgRNA (dual stranded guide RNA). Cas9 then cleaves the DNA to disable or repair a gene. A non-limiting list of exemplary modifications to this process is described here. In a CRISPR/dCas9 process, a disabled or "dead" Cas9 ("dCas9") no longer has a splicing function but, with the addition of another enzymatic activity, performs a different target molecule modifying function. For example, tethering a cytidine deaminase to dCas9 converts a C-G DNA base pair into T-A base pair. In an alternative dCas9 process, a different enzyme tethered to the dCas9 results in changing the base C into a T, or a G to an A in a target DNA. Alternatively, the dCas9 process can be modified by fusion of transcription factors to block or activate RNA polymerase activity, resulting in turning off (CRISPRi) or turning on (CRISPRa) gene transcription and therefore regulate gene expression. For example, the dCas9 process is modified by fusion with a transcriptional repressor. In some instances, the dCas9 process is modified by fusion with a transcriptional activator. In some instances, the dCas9 process is modified by fusion with a plurality of transcriptional repressors or transcriptional activators. In alternative arrangements, a gRNA has multiple sites for cleavage, resulting in a gRNA having multiple regions for gene editing. In the case of Cas9n, or "nicking Cas9," either the RuvC or HNH cleavage domain is modified to be inactive. This inactivation leaves Cas9 only able to produce only a stranded break in the DNA (a nick), not a double stranded break. In some arrangements, two Cas9n enzymes, one for each strand, are used to produce the double stranded break. As they can recognize both the upstream and downstream regions of the cut site, off target effects are ablated. In the case of hfCas9, instead of using dual Cas9n proteins to generate the off-target effect-free Cas9 cut, a modified Cas9 enzyme has relaxed binding target specificity stringency to allow for less than perfect matches prior to enzymatic activity. In some instances, the dCas9 process is modified by fusion with a label or tag for detecting a target nucleic acid. For example, the label is a fluorescent marker (*e.g.*, GFP) for detecting the target nucleic

acid. In some instances, the dCas9 is fused to an epitope tag and is used for purification of the target nucleic acid specified by a gRNA.

[0088] Provided herein are libraries comprising nucleic acids for directing a nuclease to a particular target nucleic acid sequence. In some instances, the target nucleic acid sequence comprises DNA. In some instances, the target nucleic acid sequence comprises RNA. For example, libraries comprising nucleic acids for directing C2c2 are generated for targeting a RNA sequence. In some instances, the DNA or RNA is single stranded or double stranded.

[0089] Provided herein are libraries comprising nucleic acids for nuclease targeting of a particular target nucleic acid sequence, wherein the nuclease is from a species of, but not limited to, *Streptococcus, Campylobacter, Nitratifractor, Staphylococcus, Parvibaculum, Roseburia, Neisseria, Gluconacetobacter, Azospirillum, Sphaerochaeta, Lactobacillus, Eubacterium, Corynebacter, Carnobacterium, Rhodobacter, Listeria, Paludibacter, Clostridium, Lachnospiraceae, Clostridiaridium, Leptotrichia, Francisella, Legionella, Alicyclobacillus, Methanomethyophilus, Porphyromonas, Prevotella, Bacteroidetes, Helcococcus, Letospira, Desulfovibrio, Desulfonatronum, Desulfurococcus, Opitutaceae, Tuberibacillus, Bacillus, Brevibacilus, Methylobacterium, Natronobacterium, Flavobacterium, Saccharomyces, Chlamydomonas, Thermus, Pyrococcus*, *Mycoplasma,* or *Acidaminococcus.* Exemplary nucleases are listed in **Table 2A**. gRNAs described herein may bind to the terminator sequence of a nuclease from any of the species listed above, or nucleases from additional species where the enzyme allows for genome editing functions. Exemplary terminator sequences include, without limitation, those listed in **Table 2B**. Exemplary PAM sequences include, without limitation, those listed in **Table 2C**.

Table 2A.

Name	Accession Number	Species
Cas9	Q99ZW2	*Streptococcus pyogenes*
Cas9	J7RUA5	*Staphylococcus aureus*
Cas9	G3ECR1	*Streptococcus thermophilus*
C2c2	P0DOC6	*Leptotrichia shahii*
Cpf1	U2UMQ6	*Acidaminococcus sp.*
C2c1	T0D7A2	*Alicyclobacillus acidoterrestris*
FokI	P14870	*Flavobacterium okeanokoites*
AciI	A0A0C5GQT3	*Lactobacillus acidophilus*
I-SceI	P03882	*Saccharomyces cerevisiae*
I-CreI	P05725	*Chlamydomonas reinhardtii*
I-DmoI	1B24_A	*Desulfurococcus mobilis*
TtAgo	Q746M7	*Thermus thermophilus*
PfAgo	Q8U3D2	*Pyrococcus furiosus*
NgAgo	A0A172MAH6	*Natronobacterium gregoryi*

Table 2B

SEQ ID NO	Species	Sequence
7	*Streptococcus thermophilus*	5' TACTCAACTTGAAAAGGTGGCACCGATTCGGTGTTTTT 3'
8	*Streptococcus mutans*	5' TACACAACTTGAAAAAGTGCGCACCGATTCGGTGCTTTT 3'
9	*Listeria innocua*	5' TTATCAACTTTTAATTAAGTAGCGCTGTTTCGGCGCTTTT 3'
10	*Mycoplasma mobile*	5' TATGCCGTAACTACTACTTATTTTCAAAATAAGTAGTTTT 3'
11	*Campylobacter jejuni*	5' GACTCTGCGGGGTTACAATCCCCTAAAACCGCTTTT 3'

Table 2C

Species/Variant of Cas9	PAM Sequence
Streptococcus pyogenes/ SpCas9	NGG
Streptococcus pyogenes/SpCas9 D1135E variant	NGG
Streptococcus pyogenes/SpCas9 VRER variant	NGCG
Streptococcus pyogenes/SpCas9 EQR variant	NGAG
Streptococcus pyogenes/SpCas9 VQR variant	NGAN or NGNG
Staphylococcus aureus/SaCas9	NNGRRT or NNGRR(N)
Neisseria meningitidis	NNNNGATT
Streptococcus thermophilus	NNAGAAW
Treponema denticola	NAAAAC

[0090] Provided herein are libraries comprising nucleic acids for targeting one or more nuclease(s) to a particular nucleic acid sequence. In some instances, the nuclease is at least one of TALEN, ZFN, meganuclease, Argonaute, and Cas protein. For example, more than one nuclease can be multiplexed to generate large genomic deletions, modify multiple sequences at once, or be used in conjunction with other enzymes such as a nickase. In some instances, the number of nucleases is at least 2 nucleases for the target nucleic acid sequence. In some instances, the number of nucleases is in a range of about 2 to 3, 2 to 4, 2 to 5, 2 to 6, 2 to 7, 2 to 8, 2 to 9, or 2 to 10 nucleases for the target nucleic acid sequence.

[0091] Provided herein are libraries comprising synthesized nucleic acids for nickase targeting of a particular nucleic acid sequence. A nickase is an enzyme that generates a single stranded break in a nucleic acid sequence. In some instances, the synthesized nucleic acids are DNA, RNA, any analogs, or derivatives thereof. In some instances, the particular nucleic acid sequence comprises DNA, RNA, any analogs, or derivatives thereof. In some instances, the nickase cleaves the particular nucleic acid sequence. In some instances, the nickase binds the particular nucleic acid but does not cleave it. In some instances, the nickase is an altered nuclease, wherein the nuclease is TALEN, ZFN, meganuclease, Argonaute, or Cas protein. In some instances, the nickase is generated by altering a nuclease domain of TALEN, ZFN, meganuclease, Argonaute, or Cas protein. In some instances, the nickase is generated by altering the nuclease domain of Cas9.

[0092] In some instances, libraries comprise nucleic acids for one or more nickase(s) targeting of a particular nucleic acid sequence. In some instances, the number of nickases is at least 2 nickases for the particular nucleic acid sequence. In some instances, the number of nickases is in a range of about 2 to 3, 2 to 4, 2 to 5, 2 to 6, 2 to 7, 2 to 8, 2 to 9, or 2 to 10 nickases for the particular nucleic acid sequence. In some instances, libraries comprise nucleic acids for directing one or more nickase and one or more nuclease to the particular nucleic acid sequence.

[0093] Libraries comprising nucleic acids for targeting a nuclease to a particular nucleic acid sequence provided herein can result in cleavage of the particular nucleic acid sequence. In some instances, the nuclease is at least one of TALEN, ZFN, meganuclease, Argonaute, and Cas protein. In some instances, the nuclease is a chimeric nuclease that provides for a modification of the particular nucleic acid sequence other than cleavage. For example, the chimeric nuclease results in methylation, demethylation, polyadenylation, deadenylation, deamination, or polyuridinylation.

[0094] *Transcription Activator-Like Effector Nuclease*

[0095] Provided herein are methods for synthesizing nucleic acid libraries comprising nucleic acids for Transcription Activator-Like Effector Nuclease (TALEN) targeting of a particular nucleic acid sequence. TALENs are a class of engineered sequence-specific nucleases that can be used to induce double-strand breaks at specific target sequences. TALENs can be generated by fusing transcription activator- like (TAL) effector DNA-binding domain, or a functional part thereof, to the catalytic domain of a nuclease. The TAL effector DNA binding domain comprises a series of TAL repeats, which are generally highly conserved 33 or 34 amino acid sequence segments that each comprise a highly variable 12th and 13th amino acid known as the repeat variable diresidue (RVD). Each RVD can recognize and bind to a specific nucleotide. Thus, a TAL effector binding domain can be engineered to recognize a specific sequence of nucleotides by combining TAL repeats comprising the appropriate RVDs.

[0096] Provided herein are methods for synthesizing a nucleic acid library comprising non-identical nucleic acids that encode for a TAL effector DNA-binding domain. In some instances, the TAL effector DNA-binding domain is designed to recognize a particular target nucleic acid sequence and induce double-stranded breaks at a particular site. In some instances, the TAL effector DNA-binding domain comprises a number of TAL repeats that are designed to recognize and bind to a particular nucleic acid sequence. In some instances, the number of TAL repeats is at least or about 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, or more TAL repeats.

[0097] In some instances, a nucleic acid library comprising non-identical nucleic acids encoding for TAL effector DNA-binding domain are synthesized. In some instances, the nucleic acid library as described herein that when translated encodes for a protein library. In some instances, the nucleic acid library is expressed in cells and a protein library is generated. In some instances, the synthesized nucleic acids libraries are inserted into expression vectors. In some instances, the synthesized nucleic acids libraries are inserted into expression vectors and expressed in cells.

[0098] Nucleic acid libraries comprising nucleic acids that encode for a TAL effector DNA-binding domain generated by methods described herein can be used for generating a TALEN. In some instances, this is accomplished by mixing the TAL effector binding domain library that is cloned and expressed in vectors with a nuclease. Exemplary nucleases include, but are not limited to, AciI, AcuI, AlwI, BbvI, BccI, BceAI, BciVI, BfuAI, BmgBI, BmrI, BpmI, BpuEI, BsaI, BsmAI, BsmFI, BseRI, BspCNI, BsrI, BsgI, BsmI, BspMI, BsrBI, BsrDI, BtgZI, BtsI, BtsCI, EarI, EciI, FokI, HgaI, HphI, HpyAV, MboII, MlyI, MmeI, MnlI, NmeAIII, PleI, SfaNI, BbvCI, Bpu10I, BspQI, SapI, BaeI, BsaXI, or CspCI. In some instances, mixing occurs by ligation. Exemplary ligases, included, but are not limited to, *E. coli* ligase, T4 ligase, mammalian ligases (e.g., DNA ligase I, DNA ligase II, DNA ligase III, DNA ligase IV), thermostable ligases, and fast ligases. TALENs generated by methods described herein can be inserted into expression vectors. In some instances, TALENs are inserted into expression vectors and expressed in cells.

[0099] Provided herein are methods for synthesizing a TAL effector DNA-binding domain library comprising non-identical nucleic acid sequences for a gene in a genome of a prokaryotic or eukaryotic organism. In some instances, the TAL effector DNA-binding domain library comprises at least 3, 4, 5, 6, 7, 8, 9, 10, or more than 10 non-identical nucleic acid sequences for a gene for at least 5% of the genome. In some instances, the TAL effector DNA-binding domain library comprise non-identical nucleic acid sequences for one or more genes for at least 5% of the genome. In some instances, the TAL effector DNA-binding domain library comprises non-identical nucleic

acid sequence for about 1, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, or more than 100 genes for at least 5% of the genome. In some instances, the TAL effector DNA-binding domain library comprises non-identical nucleic acid sequence for about 1-100, 2-95, 5-90, 10-85, 15-80, 20-75, 25-70, 30-65, 35-60, or 40-50 genes for at least 5% of the genome.

[00100] *Zinc Finger Nucleases*

[00101] Provided herein are methods for synthesizing nucleic acid libraries comprising nucleic acids for Zinc Finger Nuclease (ZFN) targeting of a particular nucleic acid sequence. ZFNs can be generated by fusion of a nuclease with a DNA binding zinc finger domain (ZFD). The ZFD can bind to a target nucleic acid sequence through one or more zinc fingers. In some instances, the ZFD comprises at least 2, 3, 4, 5, 6, 7, 8, 9, 10, or more zinc fingers. In some instances, the ZFD comprises at most 2, 3, 4, 5, 6, 7, 8, 9, 10, or more zinc fingers. In some instances, the ZFD is designed to recognize a particular target nucleic acid sequence and induce double-stranded breaks at a particular site.

[00102] Provided herein are methods for synthesizing a nucleic acid library comprising nucleic acids that when transcribed and translated encode for a ZFD. In some instances, when the nucleic acid library is translated encode for a protein library. In some instances, the nucleic acid library is expressed in cells and a protein library is generated. In some instances, the synthesized nucleic acids libraries are inserted into expression vectors. In some instances, the synthesized nucleic acids libraries are inserted into expression vectors and expressed in cells.

[00103] Nucleic acid libraries comprising nucleic acids that encode for a ZFD generated by methods described herein can be used for generating a ZFN. In some instances, this is accomplished by mixing the ZFD that is cloned and expressed in vectors with a nuclease. Exemplary nucleases include, but are not limited to, AciI, AcuI, AlwI, BbvI, BccI, BceAI, BciVI, BfuAI, BmgBI, BmrI, BpmI, BpuEI, BsaI, BsmAI, BsmFI, BseRI, BspCNI, BsrI, BsgI, BsmI, BspMI, BsrBI, BsrDI, BtgZI, BtsI, BtsCI, EarI, EciI, FokI, HgaI, HphI, HpyAV, MboII, MlyI, MmeI, MnlI, NmeAIII, PleI, SfaNI, BbvCI, Bpu10I, BspQI, SapI, BaeI, BsaXI, or CspCI. In some instances, mixing occurs by ligation. Exemplary ligases, included, but are not limited to, *E. coli* ligase, T4 ligase, mammalian ligases (e.g., DNA ligase I, DNA ligase II, DNA ligase III, DNA ligase IV), thermostable ligases, and fast ligases. ZFNs generated by methods described herein can be inserted into expression vectors. In some instances, ZFNs are inserted into expression vectors and expressed in cells.

[00104] Provided herein are methods for synthesizing a ZFD library comprising non-identical nucleic acid sequences for a gene in a genome of a prokaryotic or eukaryotic organism. In some instances, the ZFD library comprises at least 3, 4, 5, 6, 7, 8, 9, 10, or more than 10 non-identical

nucleic acid sequences for a gene for at least 5% of the genome. In some instances, the ZFD library comprise non-identical nucleic acid sequences for one or more genes for at least 5% of the genome. In some instances, the ZFD library comprises non-identical nucleic acid sequence for about 1, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, or more than 100 genes for at least 5% of the genome. In some instances, the ZFD library comprises non-identical nucleic acid sequence for about 1-100, 2-95, 5-90, 10-85, 15-80, 20-75, 25-70, 30-65, 35-60, or 40-50 genes for at least 5% of the genome.

[00105] *Meganucleases*

[00106] Provided herein are methods for synthesizing nucleic acid libraries comprising nucleic acids for meganuclease targeting of a particular nucleic acid sequence. Meganucleases are enzymes that can recognize and cleave long base pair (*e.g.*, 12-40 base pairs) DNA targets. In some instances, meganucleases are engineered to comprise domains of other enzymes to confer specificity for a target nucleic acid sequence. For example, meganucleases are engineered to comprise a TAL effector DNA binding domain.

[00107] Provided herein are methods for synthesizing a nucleic acid library comprising nucleic acids that when transcribed and translated encode for a binding domain for use with a meganuclease. In some instances, when the nucleic acid library is translated encode for a protein library. In some instances, the nucleic acid library is expressed in cells and a protein library is generated. In some instances, the synthesized nucleic acids libraries are inserted into expression vectors. In some instances, the synthesized nucleic acids libraries are inserted into expression vectors and expressed in cells.

[00108] Nucleic acid libraries comprising nucleic acids that encode for a domain generated by methods described herein can be used for engineering a meganuclease for targeting a particular nucleic acid sequence. In some instances, this is accomplished by mixing a binding domain library such as a TAL effector binding domain library that is cloned and expressed in vectors with a meganuclease. Exemplary meganucleases for use with the methods provided herein include, but are not limited to, I-Scel, I- Scell, I-SceIII, I-SceIV, I-SceV, I-SceVI, I-SceVII, I-Ceul, I-CeuAIIP, I-Crel, I-CrepsbIP, I- CrepsbllP, I-CrepsbIIIP, I-CrepsbIVP, I-Tlil, I-Ppol, PI-PspI, F-Scel, F-Scell, F-Suvl, F- TevI, F-TevII, I-Amal, I-Anil, I-Chul, I-Cmoel, I-Cpal, I-CpaII, I-Csml, I-Cvul, I-CvuAIP, I-Ddil, I-DdiII, I-Dirl, I-Dmol, I-Hmul, I-HmuII, I-HsNIP, I-Llal, I-Msol, I-Naal, I-Nanl, I- NcIIP, I-NgrIP, I-Nitl, I-Njal, I-Nsp236IP, I-Pakl, I-PboIP, I-PcuIP, I-PcuAI, I-PcuVI, I- PgrlP, 1-PobIP, I-Porl, I-PorIIP, I-PbpIP, I-SpBetaIP, I-Scal, I-SexIP, 1-SneIP, I-Spoml, I- SpomCP, I-SpomIP, I-SpomIIP, I-SquIP, I-Ssp6803I, I-SthPhiJP, I-SthPhiST3P, I- SthPhiSTe3bP, I-TdeIP, I-Tevl, I-TevII, I-TevIII, I-UarAP, I-UarHGPAIP, I-UarHGPA13P, I-VinIP, 1-ZbiIP, PI-MtuI, PI-MtuHIP PI-MtuHIIP, PI-PfuI, PI-PfuII, PI-PkoI, Pl-PkoII, PI- Rma43812IP, PI-SpBetaIP, PI-SceI,

PI-TfuI, PI-TfuII, PI-Thyl, PI-Tlil, PI-TliII, or fragments thereof. In some instances, mixing occurs by ligation. Exemplary ligases, included, but are not limited to, *E. coli* ligase, T4 ligase, mammalian ligases (e.g., DNA ligase I, DNA ligase II, DNA ligase III, DNA ligase IV), thermostable ligases, and fast ligases. Engineered meganucleases generated by methods described herein can be inserted into expression vectors. In some instances, the engineered meganucleases are inserted into expression vectors and expressed in cells.

[00109] Provided herein are methods for synthesizing a binding domain library for use with a meganuclease comprising non-identical nucleic acid sequences for a gene in a genome of a prokaryotic or eukaryotic organism. In some instances, the domain library comprises at least 3, 4, 5, 6, 7, 8, 9, 10, or more than 10 non-identical nucleic acid sequences for a gene for at least 5% of the genome. In some instances, the domain library comprise non-identical nucleic acid sequences for one or more genes for at least 5% of the genome. In some instances, the domain library comprises non-identical nucleic acid sequence for about 1, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, or more than 100 genes for at least 5% of the genome. In some instances, the domain library comprises non-identical nucleic acid sequence for about 1-100, 2-95, 5-90, 10-85, 15-80, 20-75, 25-70, 30-65, 35-60, or 40-50 genes for at least 5% of the genome.

[00110] *Argonautes*

[00111] Provided herein are methods for synthesizing nucleic acid libraries comprising nucleic acids for Argonaute targeting of a particular nucleic acid sequence. Argonautes are a family of RNA or DNA guided nucleases. In some instances, Argonautes use a guide nucleic acid to identify a target nucleic acid. In some instances, the guide nucleic acid is a single guide RNA (sgRNA). In some instances, the guide nucleic acid is a guide DNA (gDNA). Exemplary Argonautes include, but are not limited to, TtAgo, PfAgo, and NgAgo. In some embodiments, the Argonaute is NgAgo.

[00112] Provided herein are methods for synthesizing a guide nucleic acid library comprising non-identical nucleic acid sequences for a gene in a genome of a prokaryotic or eukaryotic organism. In some instances, the guide nucleic acid library is a sgRNA library. In some instances, the guide nucleic acid library is a dgRNA library. In some instances, the guide nucleic acid library comprises at least 3, 4, 5, 6, 7, 8, 9, 10, or more than 10 non-identical nucleic acid sequences for a gene for at least 5% of the genome. In some instances, the guide nucleic acid library comprise non-identical nucleic acid sequences for one or more genes for at least 5% of the genome. In some instances, the guide nucleic acid library comprises non-identical nucleic acid sequence for about 1, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, or more than 100 genes for at least 5% of the genome. In some instances, the guide nucleic acid library comprises non-identical nucleic acid sequence for

about 1-100, 2-95, 5-90, 10-85, 15-80, 20-75, 25-70, 30-65, 35-60, or 40-50 genes for at least 5% of the genome.

[00113] *CRISPR-associated Proteins*

[00114] Provided herein are methods for synthesizing nucleic acid libraries comprising nucleic acids encoding for gRNAs for CRISPR-associated (Cas) protein targeting of a particular nucleic acid sequence. In some instances, the Cas protein is at least one of Cpf1, C2c1, C2c2, Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas9 (Csn1 or Csx12), Cas10, Csy1, Csy2, Csy3, Cse1, Cse2, Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, Csf4, homologues thereof, and modified versions thereof. In some instances, the Cas protein is Cas9.

[00115] Provided herein are methods for synthesizing a gRNA library comprising non-identical nucleic acid sequences for a gene in a genome of a prokaryotic or eukaryotic organism. In some instances, the gRNA library comprises at least 3, 4, 5, 6, 7, 8, 9, 10, or more than 10 non-identical nucleic acid sequences for a gene for at least 5% of the genome. In some instances, the gRNA library comprise non-identical nucleic acid sequences for one or more genes for at least 5% of the genome. In some instances, the gRNA library comprises non-identical nucleic acid sequence for about 1, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, or more than 100 genes for at least 5% of the genome. In some instances, the gRNA library comprises non-identical nucleic acid sequence for about 1-100, 2-95, 5-90, 10-85, 15-80, 20-75, 25-70, 30-65, 35-60, or 40-50 genes for at least 5% of the genome. The gRNA library may encode for sgRNA or dgRNAs.

[00116] **Variant Library Synthesis**

[00117] Provided herein are methods for synthesis of a variant nucleic acid library generated by combination of nucleic acids encoding complete or partial gene sequence with gRNAs and a nuclease, e.g., Cas9 enzyme or Cas9 variant enzyme. The fragments may collectively space the entire region of a gene. In some cases, the library encodes DNA or RNA. In some cases, the library encodes for a single gene or for up to an entire genome. For example, a gRNA library encoding for 5 gRNAs per a gene for a genome comprising about 20,000 genes would result in about 100,000 gRNAs. Such a library can be used to selectively silence or modify a single gene, a pathway of genes, or all genes in a single genome. In some arrangement, gRNAs lack a homology sequence and random end joining occurs. Such a process results in non-homologous end joining ("NHEJ"). In some instances, following NHEJ, an insertion, a deletion, a frameshift, or single base swapping occurs. *See* **FIG. 1B**.

[00118] Synthesized libraries described herein may be used for application in CRISPR-Cas9 functions, wherein the gRNA sequence generated is used to disrupt expression of or alter the expression product sequence of a target DNA sequence in a cell or in a mixture comprising a target DNA and Cas9 enzyme. In some embodiments, each variant encodes for a codon resulting in a different amino acid during translation. **Table 3** provides a listing of each codon possible (and the representative amino acid) for a variant site.

Table 3. List of codons and amino acids

Amino Acids	One letter code	Three letter code	Codons					
Alanine	A	Ala	GCA	GCC	GCG	GCT		
Cysteine	C	Cys	TGC	TGT				
Aspartic acid	D	Asp	GAC	GAT				
Glutamic acid	E	Glu	GAA	GAG				
Phenylalanine	F	Phe	TTC	TTT				
Glycine	G	Gly	GGA	GGC	GGG	GGT		
Histidine	H	His	CAC	CAT				
Isoleucine	I	Iso	ATA	ATC	ATT			
Lysine	K	Lys	AAA	AAG				
Leucine	L	Leu	TTA	TTG	CTA	CTC	CTG	CTT
Methionine	M	Met	ATG					
Asparagine	N	Asn	AAC	AAT				
Proline	P	Pro	CCA	CCC	CCG	CCT		
Glutamine	Q	Gln	CAA	CAG				
Arginine	R	Arg	AGA	AGG	CGA	CGC	CGG	CGT
Serine	S	Ser	AGC	AGT	TCA	TCC	TCG	TCT

Threonine	T	Thr	ACA	ACC	ACG	ACT
Valine	V	Val	GTA	GTC	GTG	GTT
Tryptophan	W	Trp	TGG			
Tyrosine	Y	Tyr	TAC	TAT		

[00119] Provided herein are methods for synthesis of a variant nucleic acid library generated by combination of nucleic acids encoding complete or partial gene sequence with a nuclease, wherein the nuclease is TALEN, ZFN, or an engineered meganuclease. In some instances, methods for synthesis of a variant nucleic acid library generated by combination of nucleic acids encoding complete or partial gene sequence with guide nucleic acids such as sgRNAs with a nuclease, wherein the nuclease is Argonaute or a Cas protein. Synthesized libraries described herein may be used for application in nuclease functions, wherein the nucleic acid sequence generated is used to disrupt expression of or alter the expression product sequence of a target DNA sequence in a cell or in a mixture comprising a target DNA and a nuclease. In some embodiments, each variant encodes for a codon resulting in a different amino acid during translation.

[00120] Variant nucleic acid libraries as described herein comprise sgRNAs or dgRNAs for varying a target nucleic acid sequence encoding in at least a single codon such that a plurality of different variants of a single residue in the subsequent protein encoded by the synthesized nucleic acid are generated by standard translation processes. In some instances, each variant encodes for a codon resulting in a different amino acid of a protein domain. For example, the protein domain is a conserved domain or catalytic domain. In some embodiments, the protein domain is, but not limited to, a kinase domain, an ATP-binding domain, a GTP-binding domain, a guanine nucleotide exchange factor (GEF) domain, a GTPase activating protein (GAP) domain, a hydrolase domain, an endonuclease domain, an exonuclease domain, a protease domain, a phosphatase domain, a phospholipase domain, a pleckstrin homology domain, a Src homology domain, and a ubiquitin-binding domain. In some instances, the variant nucleic acid libraries comprise sgRNAs or dgRNAs for targeting a nucleic acid sequence that encodes for variation in at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more than 10 protein domains.

[00121] In some embodiments, the variants encode for amino acids for a protein with particular activity. For example, the variants encode for a protein that comprises methyltransferase activity, demethylase activity, dismutase activity, alkylation activity, depurination activity, oxidation activity, pyrimidine dimer forming activity, integrase activity, transposase activity, recombinase

activity, polymerase activity, ligase activity, helicase activity, photolyase activity, glycosylase activity, acetyltransferase activity, deacetylase activity, kinase activity, phosphatase activity, ubiquitin ligase activity, deubiquitinating activity, adenylation activity, deadenylation activity, SUMOylating activity, deSUMOylating activity, ribosylation activity, deribosylation activity, myristoylation activity, remodelling activity, protease activity, oxidoreductase activity, transferase activity, hydrolase activity, lyase activity, isomerase activity, synthase activity, synthetase activity, or demyristoylation activity.

[00122] *Variation generated by homology-directed repair (HDR)*

[00123] In an exemplary process for variant nucleic acid library generation, Cas9 cleavage and homologous recombination are incorporated to generate variety in a target DNA library. First, a library of gRNA is synthesized (either by de novo synthesis of RNA or de novo synthesis of DNA followed by transcription (in vivo or in vitro) to generate gRNA), wherein the library comprises a plurality of gRNA molecules per a gene. For example, the gRNA library may comprise 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more gRNAs per a gene. The gRNA library is mixed with a Cas9 enzyme and a target DNA library, where the target DNA library comprises nucleic acid sequence encoding for at least one gene fragment or at least one gene. For example, the target DNA library may comprise 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more genes or gene fragments. In some instances, the target DNA library comprises sequence from multiple genes in a pathway or from all genes in an organism. Also added to the mixture are replacement sequences which comprise a homology sequence and a variant nucleic acid sequence such that variation is introduced into target DNA strands. The resultant target DNA library will comprise a plurality of variant DNA sequences. In some instances, variation introduces a deletion, frame shift, or insertion into target DNA sequence. In some instances, the variant DNA sequences result in variation for at least one codon per a gene or gene fragment. In some instances, a portion of a gene is inserted into the target DNA or, alternatively, a portion of a target DNA sequence (i.e. a fragment of a gene or an entire gene) is removed from the target DNA. In some instances, the variant DNA sequences result in variation for at least one transcription regulatory sequence, e.g., a promoter, UTR, or terminator sequence, associated with gene or gene fragment.

[00124] In some instances for variant nucleic acid library generation, nuclease cleavage and homologous recombination are incorporated to generate variety in a target DNA library, wherein the nuclease is a TALEN, a ZFN, a meganuclease, a Cas, or an Argonaute. In some instances, where the nuclease is TALEN, a library of TAL effector DNA-binding domains is synthesized (either by de novo synthesis of RNA or de novo synthesis of DNA followed by transcription and translation (in vivo or in vitro)), wherein the library comprises a plurality of TAL effector DNA-

binding domain molecules per a gene. For example, the TAL effector DNA-binding domain library may comprise 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more TAL effector DNA-binding domain molecules per a gene. The TAL effector DNA-binding domain library can then be mixed with a nuclease enzyme to generate a TALEN. In some instances, the TALEN is combined with a target DNA library, where the target DNA library comprises nucleic acid sequence encoding for at least one gene fragment or at least one gene. For example, the target DNA library may comprise 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more genes or gene fragments. In some instances, the target DNA library comprises sequence from multiple genes in a pathway or from all genes in an organism. In some instances, also added to the mixture are replacement sequences which comprise a homology sequence and a variant nucleic acid sequence such that variation is introduced into target DNA strands. The resultant target DNA library will comprise a plurality of variant DNA sequences. In some instances, variation introduces a deletion, frame shift, or insertion into target DNA sequence. In some instances, the variant DNA sequences result in variation for at least one codon per a gene or gene fragment. In some instances, a portion of a gene is inserted into the target DNA or, alternatively, a portion of a target DNA sequence (i.e. a fragment of a gene or an entire gene) is removed from the target DNA. In some instances, the variant DNA sequences result in variation for at least one transcription regulatory sequence, e.g., a promoter, UTR, or terminator sequence, associated with gene or gene fragment.

[00125] *Variation generated by modified Cas9 enzymes*

[00126] In a second exemplary process for variant nucleic acid library generation, modified Cas9 enzymes are incorporated to generate a variant target DNA library. First, a library of gRNA is synthesized (either by de novo synthesis of RNA or de novo synthesis of DNA followed by transcription to generate gRNA), wherein the library comprises a plurality of gRNA molecules per a gene. For example, the gRNA library may comprise 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more gRNAs per a gene. The gRNA library is mixed with a modified Cas9 enzyme and a target DNA library, where the target DNA library comprises nucleic acid sequence encoding for at least one gene fragment or at least one gene. For example, the target DNA library may comprise 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more genes or gene fragments. In some instances, the target DNA library comprises sequence from multiple genes in a pathway or from all genes in an organism. The modified Cas9 enzyme has tethered to it another enzyme with nucleic acid sequence modification capabilities. An exemplary modified Cas9 enzymes includes dCas9 process in which a disabled or “dead” Cas9 (“dCas9”) no longer has a splicing function but, with the addition of another enzymatic activity, performs a different target molecule modifying function. For example, tethering a cytidine deaminase to dCas9 converts a C-G DNA base pair into T-A base pair. In an alternative dCas9

process, a different enzyme tethered to the dCas9 results in changing the base C into a T, or a G to an A in a target DNA. The resultant target DNA library comprises a plurality of variant target DNA sequences. In some instances, variation introduces a deletion, frame shift, or insertion into target DNA sequence. In some instances, the variant DNA sequences result in variation for at least one codon per a gene or gene fragment. In some instances, the variant DNA sequences result in variation for at least one transcription regulatory sequence, e.g., a promoter, UTR, or terminator sequence, associated with gene or gene fragment.

[00127] *Variation generated by modified nucleases*

[00128] Provided herein are methods for variant nucleic acid library generation comprising a modified nuclease enzyme that is incorporated to generate a variant target DNA library. In some instances, the nuclease is TALEN. In some instances, a TAL effector DNA binding domain library is synthesized (either by de novo synthesis of RNA or de novo synthesis of DNA followed by transcription and translation to generate the TAL effector DNA binding domain library), wherein the library comprises a plurality of non-identical nucleic acid sequences per a gene. For example, the TAL effector DNA binding domain library may comprise 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more non-identical nucleic acid sequences per a gene. The TAL effector DNA-binding domain library can then be mixed with a nuclease enzyme to generate a TALEN. In some instances, the TALEN is then mixed with a target DNA library, where the target DNA library comprises nucleic acid sequence encoding for at least one gene fragment or at least one gene. For example, the target DNA library may comprise 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more genes or gene fragments. In some instances, the target DNA library comprises sequence from multiple genes in a pathway or from all genes in an organism.

[00129] In some instances, the nuclease is ZFN. In some instances, a ZFD library is synthesized (either by de novo synthesis of RNA or de novo synthesis of DNA followed by transcription and translation to generate the ZFD library), wherein the library comprises a plurality of non-identical nucleic acid sequences per a gene. For example, the ZFD library may comprise 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more non-identical nucleic acid sequences per a gene. The ZFD library can then be mixed with a nuclease enzyme to generate a ZFN. In some instances, the ZFN is then mixed with a target DNA library, where the target DNA library comprises nucleic acid sequence encoding for at least one gene fragment or at least one gene. For example, the target DNA library may comprise 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more genes or gene fragments. In some instances, the target DNA library comprises sequence from multiple genes in a pathway or from all genes in an organism.

[00130] In some instances, the nuclease is a meganuclease. In some instances, a binding domain library such as a TAL effector DNA binding domain library for targeting the meganuclease to a

particular nucleic acid sequence is synthesized (either by de novo synthesis of RNA or de novo synthesis of DNA followed by transcription to generate the binding domain library), wherein the binding domain library comprises a plurality of non-identical nucleic acid sequences per a gene. For example, the binding domain library may comprise 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more non-identical nucleic acid sequences per a gene. The binding domain library can then be mixed a meganuclease enzyme to generate an engineered meganuclease. In some instances, the engineered meganuclease is then mixed with a target DNA library, where the target DNA library comprises nucleic acid sequence encoding for at least one gene fragment or at least one gene. For example, the target DNA library may comprise 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more genes or gene fragments. In some instances, the target DNA library comprises sequence from multiple genes in a pathway or from all genes in an organism.

[00131] In some instances, the nuclease is Argonaute. In some instances, a guide nucleic acid library (gRNA or gDNA) is synthesized (either by de novo synthesis of RNA or de novo synthesis of DNA followed by transcription to generate the guide nucleic acid library), wherein the guide nucleic acid library comprises a plurality of non-identical nucleic acid sequences per a gene. For example, the guide nucleic acid library may comprise 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more non-identical nucleic acid sequences per a gene. The guide nucleic acid library is mixed with a modified Argonaute enzyme and a target DNA library, where the target DNA library comprises nucleic acid sequence encoding for at least one gene fragment or at least one gene. For example, the target DNA library may comprise 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more genes or gene fragments. In some instances, the target DNA library comprises sequence from multiple genes in a pathway or from all genes in an organism.

[00132] In some instances, the modified nuclease enzyme has tethered to it another enzyme with nucleic acid sequence modification capabilities. Exemplary modification capabilities include, but are not limited to, methylation, demethylation, polyadenylation, deadenylation, deamination, and polyuridinylation. In some instances, a target DNA library comprising a plurality of variant target DNA sequences results in variation. In some instances, variation introduces a deletion, frame shift, or insertion into target DNA sequence. In some instances, the variant DNA sequences result in variation for at least one codon per a gene or gene fragment. In some instances, the variant DNA sequences result in variation for at least one transcription regulatory sequence, e.g., a promoter, UTR, or terminator sequence, associated with gene or gene fragment.

[00133] *gRNA Library Synthesis for Targeting Genes of a Model System*

[00134] Provided herein are methods for screening model systems with a nucleic acid library described herein. In some instances, the nucleic acid library is a gRNA library described herein. In

some instances, the nucleic acid library is a DNA library described herein, that when transcribed results in transcription of gRNA sequences. A non-limiting exemplary list of model organisms is provided in **Table 4**.

Table 4. Organisms and Gene Number

Model System	**Protein Coding Genes***
Arabidopsis thaliana	27000
Caenorhabditis elegans	20000
Canis lupus familiaris	19000
Chlamydomonas reinhardtii	14000
Danio rerio	26000
Dictyostelium discoideum	13000
Drosophila melanogaster	14000
Escherichia coli	4300
Macaca mulatta	22000
Mus musculus	20000
Oryctolagus cuniculus	27000
Rattus norvegicus	22000
Saccharomyces cerevisiae	6600
Sus scrofa	21000
Homo sapiens	21000

* Numbers here reflect the number of protein coding genes and excludes tRNA and non-coding RNA. Ron Milo & Rob Phillips, Cell Biology by the Numbers 286 (2015).

[00135] A library of gRNAs is synthesized (either by de novo synthesis of RNA or de novo synthesis of DNA followed by transcription to generate gRNAs), wherein the library comprises a plurality of gRNA molecules per a gene. For example, a library described herein may comprises 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more gRNAs per a gene. In some instances, the nucleic acids within a de novo synthesized library encode sequences for at least or about 3 non-identical gRNAs per a single gene. In some instances, the nucleic acids encode sequences in a range of about 1 to about 10 non-identical gRNAs per a single gene. In some instances, the nucleic acids encode sequences for at least or about 1 non-identical gRNAs per a single gene. In some instances, the nucleic acids encode sequences for at most 10 non-identical gRNAs per a single gene. In some instances, the nucleic acids encode sequences for 1 to 2, 1 to 3, 1 to 4, 1 to 5, 1 to 6, 1 to 7, 1 to 8, 1 to 9, or 1 to 10 non-identical gRNAs per a single gene. In some instances, the gRNAs are sgRNAs. In some instances, the gRNAs are dgRNAs.

[00136] In some instances, a gRNA library described herein comprises one or more non-identical gRNAs per a gene of an organism. In some instances, the gRNA library comprises 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more non-identical gRNAs per a gene for the organism. Exemplary organisms include, without limitation, *Arabidopsis thaliana, Caenorhabditis elegans, Canis lupus familiaris, Chlamydomonas reinhardtii, Danio rerio, Dictyostelium discoideum, Drosophila melanogaster, Escherichia coli, Macaca mulatta, Mus musculus, Oryctolagus cuniculus, Rattus norvegicus, Saccharomyces cerevisiae, Sus scrofa,* and *Homo sapiens*. In some instances, the gRNAs are sgRNAs. In some instances, the gRNAs are dgRNAs. In some cases, the gRNA library comprises non-identical gRNAs for at least or about 5% of the entire genome of the organism. In some cases, the gRNA library comprises non-identical gRNAs for about 5% to about 100% of the entire genome of the organism. In some instances, the gRNA library comprises non-identical gRNAs for at least or about 80% of the entire genome of the organism. In some instances, the sgRNA library comprises non-identical gRNAs for at least or about 90% of the entire genome of the organism. In some instances, the gRNA library comprises non-identical gRNAs for at least or about 95% of the entire genome of the organism. In some cases, the gRNA library comprises non-identical gRNAs for at least or about 100% of the entire genome of the organism. In some cases, the gRNA library comprises non-identical gRNAs for about 5% to 10%, 5% to 20%, 5% to 30%, 5% to 40%, 5% to 50%, 5% to 60%, 5% to 70%, 5% to 80%, 5% to 90%, 5% to 95%, 5% to 100%, 10% to 20%, 10% to 30%, 10% to 40%, 10% to 50%, 10% to 60%, 10% to 70%, 10% to 80%, 10% to 90%, 10% to 95%, 10% to 100%, 20% to 30%, 20% to 40%, 20% to 50%, 20% to 60%, 20% to 70%, 20% to 80%, 20% to 90%, 20% to 95%, 20% to 100%, 30% to 40%, 30% to 50%, 30% to 60%, 30% to 70%, 30% to 80%, 30% to 90%, 30% to 95%, 30% to 100%, 40% to 50%, 40% to 60%, 40% to 70%, 40% to 80%, 40% to 90%, 40% to 95%, 40% to 100%, 50% to 60%, 50% to 70%, 50% to 80%, 50% to 90%, 50% to 95%, 50% to 100%, 60% to 70%, 60% to 80%, 60% to 90%, 60% to 95%, 60% to 100%, 70% to 80%, 70% to 90%, 70% to 95%, 70% to 100%, 80% to 90%, 80% to 95%, 80% to 100%, 90% to 95%, 90% to 100%, or 95% to 100% of the entire genome of the organism. In some instances, the gRNA library comprises sequences from multiple genes in a pathway or from all genes in an organism. The number of gRNAs may comprise at least 2X, 3X, 5X, or 10X per a gene in an organism listed in **Table 4**. In some instances, the gRNA library targets at least one of a gene, a group of genes (*e.g.*, 3-10 genes), a pathway (*e.g.*, 10-100 genes), or a chassis (*e.g.*, 100-1000 genes).

[00137] ***Highly Parallel De Novo Nucleic Acid Synthesis***

[00138] Described herein is a platform approach utilizing miniaturization, parallelization, and vertical integration of the end-to-end process from oligonucleic acid synthesis to gene assembly

within nanowells on silicon to create a revolutionary synthesis platform. Devices described herein provide, with the same footprint as a 96-well plate, a silicon synthesis platform is capable of increasing throughput by a factor of 100 to 1,000 compared to traditional synthesis methods, with production of up to approximately 1,000,000 oligonucleic acids in a single highly-parallelized run. In some instances, a single silicon plate described herein provides for synthesis of about 6100 non-identical oligonucleic acids. In some instances, each of the non-identical oligonucleic acids is located within a cluster. A cluster may comprise 50 to 500 non-identical oligonucleic acids.

[00139] In some instances, DNA libraries encoding for gRNA libraries described herein have an error rate of less than 1:500 when compared to predetermined sequences for the DNAs. In some instances, de novo oligonucleic acids libraries disclosed herein have an aggregated error rate of less than 1:500, 1:1000, 1:1500, 1:2000, 1:3000, 1:5000 or less when compared to predetermined sequences for the DNAs. In some instances, the aggregate error rate is less than 1:1000 when compared to predetermined sequences for the DNAs. The error rate may be an aggregate error rate or an average error rate.

[00140] In some instances, RNA libraries encoding for gRNA libraries described herein have an error rate of less than 1:500 when compared to predetermined sequences for the RNAs. In some instances, de novo oligonucleic acids libraries disclosed herein have an aggregated error rate of less than 1:500, 1:1000, 1:1500, 1:2000, 1:3000, 1:5000, 1:10,000 or less when compared to predetermined sequences for the RNAs. In some instances, the aggregate error rate is less than 1:1000 when compared to predetermined sequences for the RNAs.

[00141] Substrates

[00142] In some cases, described herein are substrates comprising a plurality of clusters, wherein each cluster comprises a plurality of loci that support the attachment and synthesis of oligonucleic acids. The term “locus” as used herein refers to a discrete region on a structure which provides support for oligonucleic acids encoding for a single predetermined sequence to extend from the surface. In some embodiments, a locus is on a two dimensional surface, *e.g.*, a substantially planar surface. In some embodiments, a locus is on a three-dimensional surface, *e.g.*, a well, microwell, channel, or post. In some embodiments, a surface of a locus comprises a material that is actively functionalized to attach to at least one nucleotide for oligonucleic acid synthesis, or preferably, a population of identical nucleotides for synthesis of a population of oligonucleic acids. In some embodiments, oligonucleic acid refers to a population of oligonucleic acids encoding for the same nucleic acid sequence. In some cases, a surface of a substrate is inclusive of one or a plurality of surfaces of a substrate. The average error rates for oligonucleic acids synthesized within a library

using the systems and methods provided are often less than 1 in 1000, less than about 1 in 2000, less than about 1 in 3000 or less often.

[00143] In some embodiments, a substrate comprises a surface that supports the synthesis of a plurality of oligonucleic acids having different predetermined sequences at addressable locations on a common support. In some embodiments, a substrate provides support for the synthesis of more than 50, 100, 200, 400, 600, 800, 1000, 1200, 1400, 1600, 1800, 2,000; 5,000; 10,000; 20,000; 50,000; 100,000; 200,000; 300,000; 400,000; 500,000; 600,000; 700,000; 800,000; 900,000; 1,000,000; 1,200,000; 1,400,000; 1,600,000; 1,800,000; 2,000,000; 2,500,000; 3,000,000; 3,500,000; 4,000,000; 4,500,000; 5,000,000; 10,000,000 or more non-identical oligonucleic acids. In some cases, the substrate provides support for the synthesis of more than 50, 100, 200, 400, 600, 800, 1000, 1200, 1400, 1600, 1800, 2,000; 5,000; 10,000; 20,000; 50,000; 100,000; 200,000; 300,000; 400,000; 500,000; 600,000; 700,000; 800,000; 900,000; 1,000,000; 1,200,000; 1,400,000; 1,600,000; 1,800,000; 2,000,000; 2,500,000; 3,000,000; 3,500,000; 4,000,000; 4,500,000; 5,000,000; 10,000,000 or more oligonucleic acids encoding for distinct sequences. In some embodiments, at least a portion of the oligonucleic acids have an identical sequence or are configured to be synthesized with an identical sequence. In some embodiments, the substrate provides a surface environment for the growth of oligonucleic acids having at least 80, 90, 100, 120, 150, 175, 200, 225, 250, 275, 300, 325, 350, 375, 400, 425, 450, 475, 500 bases or more.

[00144] In some embodiments, oligonucleic acids are synthesized on distinct loci of a substrate, wherein each locus supports the synthesis of a population of oligonucleic acids. In some cases, each locus supports the synthesis of a population of oligonucleic acids having a different sequence than a population of oligonucleic acids grown on another locus. In some embodiments, the loci of a substrate are located within a plurality of clusters. In some instances, a substrate comprises at least 10, 500, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000, 11000, 12000, 13000, 14000, 15000, 20000, 30000, 40000, 50000 or more clusters. In some embodiments, a substrate comprises more than 2,000; 5,000; 10,000; 100,000; 200,000; 300,000; 400,000; 500,000; 600,000; 700,000; 800,000; 900,000; 1,000,000; 1,100,000; 1,200,000; 1,300,000; 1,400,000; 1,500,000; 1,600,000; 1,700,000; 1,800,000; 1,900,000; 2,000,000; 300,000; 400,000; 500,000; 600,000; 700,000; 800,000; 900,000; 1,000,000; 1,200,000; 1,400,000; 1,600,000; 1,800,000; 2,000,000; 2,500,000; 3,000,000; 3,500,000; 4,000,000; 4,500,000; 5,000,000; or 10,000,000 or more distinct loci. In some embodiments, a substrate comprises about 10,000 distinct loci. The amount of loci within a single cluster is varied in different embodiments. In some cases, each cluster includes 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 120, 130, 150, 200, 300, 400, 500 or more loci. In some embodiments, each cluster includes about 50-500 loci. In some embodiments, each

cluster includes about 100-200 loci. In some embodiments, each cluster includes about 100-150 loci. In some embodiments, each cluster includes about 109, 121, 130 or 137 loci. In some embodiments, each cluster includes about 19, 20, 61, 64 or more loci.

[00145] Provided herein are methods for synthesizing non-identical oligonucleic acids on a silicon plate. In some instances, the silicon plate includes about 1-10, 1-50, or 50-500 clusters. In some instances, the silicon plate includes more than about 50, 100, 250, 500, 2500, 5000, 6000, 6150, 10000 or more clusters. In some instances, each cluster includes 121 loci. In some instances, each cluster includes about 50-500, 50-200, 100-150 loci. In some instances, each cluster includes at least about 50, 100, 150, 200, 500, 1000 or more loci. In some instances, a single plate includes 100, 500, 10000, 20000, 30000, 50000, 100000, 500000, 700000, 1000000 or more loci.

[00146] In some embodiments, the number of distinct oligonucleic acids synthesized on a substrate is dependent on the number of distinct loci available in the substrate. In some embodiments, the density of loci within a cluster of a substrate is at least or about 1 locus per mm^2, 10 loci per mm^2, 25 loci per mm^2, 50 loci per mm^2, 65 loci per mm^2, 75 loci per mm^2, 100 loci per mm^2, 130 loci per mm^2, 150 loci per mm^2, 175 loci per mm^2, 200 loci per mm^2, 300 loci per mm^2, 400 loci per mm^2, 500 loci per mm^2, 1,000 loci per mm^2 or more. In some cases, a substrate comprises from about 10 loci per mm^2 to about 500 mm^2, from about 25 loci per mm^2 to about 400 mm^2, from about 50 loci per mm^2 to about 500 mm^2, from about 100 loci per mm^2 to about 500 mm^2, from about 150 loci per mm^2 to about 500 mm^2, from about 10 loci per mm^2 to about 250 mm^2, from about 50 loci per mm^2 to about 250 mm^2, from about 10 loci per mm^2 to about 200 mm^2, or from about 50 loci per mm^2 to about 200 mm^2. In some embodiments, the distance between the centers of two adjacent loci within a cluster is from about 10 um to about 500 um, from about 10 um to about 200 um, or from about 10 um to about 100 um. In some cases, the distance between two centers of adjacent loci is greater than about 10 um, 20 um, 30 um, 40 um, 50 um, 60 um, 70 um, 80 um, 90 um or 100 um. In some cases, the distance between the centers of two adjacent loci is less than about 200 um, 150 um, 100 um, 80 um, 70 um, 60 um, 50 um, 40 um, 30 um, 20 um or 10 um. In some instances, each loci has a width of about 0.5 um, 1 um, 2 um, 3 um, 4 um, 5 um, 6 um, 7 um, 8 um, 9 um, 10 um, 20 um, 30 um, 40 um, 50 um, 60 um, 70 um, 80 um, 90 um or 100 um. In some cases, the each loci is has a width of about 0.5 um to 100 um, about 0.5 um to 50 um, about 10 um to 75 um, or about 0.5 um to 50 um.

[00147] In some embodiments, the density of clusters within a substrate is at least or about 1 cluster per 100 mm^2, 1 cluster per 10 mm^2, 1 cluster per 5 mm^2, 1 cluster per 4 mm^2, 1 cluster per 3 mm^2, 1 cluster per 2 mm^2, 1 cluster per 1 mm^2, 2 clusters per 1 mm^2, 3 clusters per 1 mm^2, 4 clusters per 1 mm^2, 5 clusters per 1 mm^2, 10 clusters per 1 mm^2, 50 clusters per 1 mm^2 or more. In

some embodiments, a substrate comprises from about 1 cluster per 10 mm^2 to about 10 clusters per 1 mm^2. In some embodiments, the distance between the centers of two adjacent clusters is less than about 50 um, 100 um, 200 um, 500 um, 1000 um, or 2000 um or 5000 um. In some cases, the distance between the centers of two adjacent clusters is between about 50 um and about 100 um, between about 50 um and about 200 um, between about 50 um and about 300 um, between about 50 um and about 500 um, and between about 100 um to about 2000 um. In some cases, the distance between the centers of two adjacent clusters is between about 0.05 mm to about 50 mm, between about 0.05 mm to about 10 mm, between about 0.05 mm and about 5 mm, between about 0.05 mm and about 4 mm, between about 0.05 mm and about 3 mm, between about 0.05 mm and about 2 mm, between about 0.1 mm and 10 mm, between about 0.2 mm and 10 mm, between about 0.3 mm and about 10 mm, between about 0.4 mm and about 10 mm, between about 0.5 mm and 10 mm, between about 0.5 mm and about 5 mm, or between about 0.5 mm and about 2 mm. In some cases, each cluster has a cross section of about 0.5 to 2 mm, about 0.5 to 1 mm, or about 1 to 2 mm. In some cases, each cluster has a cross section of about 0.5, 0.6, 0.7, 0.8, 0.9, 1, 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9 or 2 mm. In some cases, each cluster has an interior cross section of about 0.5, 0.6, 0.7, 0.8, 0.9, 1, 1.1, 1.15, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9 or 2 mm.

[00148] In some embodiments, a substrate is about the size of a standard 96 well plate, for example between about 100 and 200 mm by between about 50 and 150 mm. In some embodiments, a substrate has a diameter less than or equal to about 1000 mm, 500 mm, 450 mm, 400 mm, 300 mm, 250 nm, 200 mm, 150 mm, 100 mm or 50 mm. In some embodiments, the diameter of a substrate is between about 25 mm and 1000 mm, between about 25 mm and about 800 mm, between about 25 mm and about 600 mm, between about 25 mm and about 500 mm, between about 25 mm and about 400 mm, between about 25 mm and about 300 mm, or between about 25 mm and about 200. Non-limiting examples of substrate size include about 300 mm, 200 mm, 150 mm, 130 mm, 100 mm, 76 mm, 51 mm and 25 mm. In some embodiments, a substrate has a planar surface area of at least about 100 mm^2; 200 mm^2; 500 mm^2; 1,000 mm^2; 2,000 mm^2; 5,000 mm^2; 10,000 mm^2; 12,000 mm^2; 15,000 mm^2; 20,000 mm^2; 30,000 mm^2; 40,000 mm^2; 50,000 mm^2 or more. In some embodiments, the thickness of a substrate is between about 50 mm and about 2000 mm, between about 50 mm and about 1000 mm, between about 100 mm and about 1000 mm, between about 200 mm and about 1000 mm, or between about 250 mm and about 1000 mm. Non-limiting examples of substrate thickness include 275 mm, 375 mm, 525 mm, 625 mm, 675 mm, 725 mm, 775 mm and 925 mm. In some cases, the thickness of a substrate varies with diameter and depends on the composition of the substrate. For example, a substrate comprising materials other than silicon has a different thickness than a silicon substrate of the same diameter.

Substrate thickness may be determined by the mechanical strength of the material used and the substrate must be thick enough to support its own weight without cracking during handling.

[00149] Surface materials

[00150] Substrates, devices and reactors provided herein are fabricated from any variety of materials suitable for the methods and compositions described herein. In certain embodiments, substrate materials are fabricated to exhibit a low level of nucleotide binding. In some cases, substrate materials are modified to generate distinct surfaces that exhibit a high level of nucleotide binding. In some embodiments, substrate materials are transparent to visible and/or UV light. In some embodiments, substrate materials are sufficiently conductive, *e.g.*, are able to form uniform electric fields across all or a portion of a substrate. In some embodiments, conductive materials are connected to an electric ground. In some cases, the substrate is heat conductive or insulated. In some cases, the materials are chemical resistant and heat resistant to support chemical or biochemical reactions, for example oligonucleic acid synthesis reaction processes. In some embodiments, a substrate comprises flexible materials. Flexible materials include, without limitation, modified nylon, unmodified nylon, nitrocellulose, polypropylene, and the like. In some embodiments, a substrate comprises rigid materials. Rigid materials include, without limitation, glass, fuse silica, silicon, silicon dioxide, silicon nitride, plastics (for example, polytetraflouroethylene, polypropylene, polystyrene, polycarbonate, and blends thereof, and the like), and metals (for example, gold, platinum, and the like). In some embodiments, a substrate is fabricated from a material comprising silicon, polystyrene, agarose, dextran, cellulosic polymers, polyacrylamides, polydimethylsiloxane (PDMS), glass, or any combination thereof. In some cases, a substrate is manufactured with a combination of materials listed herein or any other suitable material known in the art.

[00151] Surface Architecture

[00152] In various embodiments, a substrate comprises raised and/or lowered features. One benefit of having such features is an increase in surface area to support oligonucleic acid synthesis. In some embodiments, a substrate having raised and/or lowered features is referred to as a three-dimensional substrate. In some cases, a three-dimensional substrate comprises one or more channels. In some cases, one or more loci comprise a channel. In some cases, the channels are accessible to reagent deposition via a deposition device such as an oligonucleic acid synthesizer. In some cases, reagents and/or fluids collect in a larger well in fluid communication one or more channels. For example, a substrate comprises a plurality of channels corresponding to a plurality of loci with a cluster, and the plurality of channels are in fluid communication with one well of the

cluster. In some methods, a library of oligonucleic acids is synthesized in a plurality of loci of a cluster.

[00153] In some embodiments, the structure is configured to allow for controlled flow and mass transfer paths for oligonucleic acid synthesis on a surface. In some embodiments, the configuration of a substrate allows for the controlled and even distribution of mass transfer paths, chemical exposure times, and/or wash efficacy during oligonucleic acid synthesis. In some embodiments, the configuration of a substrate allows for increased sweep efficiency, for example by providing sufficient volume for a growing an oligonucleic acid such that the excluded volume by the growing oligonucleic acid does not take up more than 50, 45, 40, 35, 30, 25, 20, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2, 1%, or less of the initially available volume that is available or suitable for growing the oligonucleic acid. In some embodiments, a three-dimensional structure allows for managed flow of fluid to allow for the rapid exchange of chemical exposure.

[00154] In some embodiments, segregation is achieved by physical structure. In some embodiments, segregation is achieved by differential functionalization of the surface generating active and passive regions for oligonucleic acid synthesis. Differential functionalization is also be achieved by alternating the hydrophobicity across the substrate surface, thereby creating water contact angle effects that cause beading or wetting of the deposited reagents. Employing larger structures can decrease splashing and cross-contamination of distinct oligonucleic acid synthesis locations with reagents of the neighboring spots. In some cases, a device, such as an oligonucleic acid synthesizer, is used to deposit reagents to distinct oligonucleic acid synthesis locations. Substrates having three-dimensional features are configured in a manner that allows for the synthesis of a large number of oligonucleic acids (*e.g.*, more than about 10,000) with a low error rate (*e.g.*, less than about 1:500, 1:1000, 1:1500, 1:2,000; 1:3,000; 1:5,000; or 1:10,000). In some cases, a substrate comprises features with a density of about or greater than about 1, 5, 10, 20, 30, 40, 50, 60, 70, 80, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 300, 400 or 500 features per mm^2.

[00155] A well of a substrate may have the same or different width, height, and/or volume as another well of the substrate. A channel of a substrate may have the same or different width, height, and/or volume as another channel of the substrate. In some embodiments, the diameter of a cluster or the diameter of a well comprising a cluster, or both, is between about 0.05 mm to about 50 mm, between about 0.05 mm to about 10 mm, between about 0.05 mm and about 5 mm, between about 0.05 mm and about 4 mm, between about 0.05 mm and about 3 mm, between about 0.05 mm and about 2 mm, between about 0.05 mm and about 1 mm, between about 0.05 mm and about 0.5 mm, between about 0.05 mm and about 0.1 mm, between about 0.1 mm and 10 mm,

between about 0.2 mm and 10 mm, between about 0.3 mm and about 10 mm, between about 0.4 mm and about 10 mm, between about 0.5 mm and 10 mm, between about 0.5 mm and about 5 mm, or between about 0.5 mm and about 2 mm. In some embodiments, the diameter of a cluster or well or both is less than or about 5 mm, 4 mm, 3 mm, 2 mm, 1 mm, 0.5 mm, 0.1 mm, 0.09 mm, 0.08 mm, 0.07 mm, 0.06 mm or 0.05 mm. In some embodiments, the diameter of a cluster or well or both is between about 1.0 and 1.3 mm. In some embodiments, the diameter of a cluster or well, or both is about 1.150 mm. In some embodiments, the diameter of a cluster or well, or both is about 0.08 mm. The diameter of a cluster refers to clusters within a two-dimensional or three-dimensional substrate.

[00156] In some embodiments, the height of a well is from about 20 um to about 1000 um, from about 50 um to about 1000 um, from about 100 um to about 1000 um, from about 200 um to about 1000 um, from about 300 um to about 1000 um, from about 400 um to about 1000 um, or from about 500 um to about 1000 um. In some cases, the height of a well is less than about 1000 um, less than about 900 um, less than about 800 um, less than about 700 um, or less than about 600 um.

[00157] In some embodiments, a substrate comprises a plurality of channels corresponding to a plurality of loci within a cluster, wherein the height or depth of a channel is from about 5 um to about 500 um, from about 5 um to about 400 um, from about 5 um to about 300 um, from about 5 um to about 200 um, from about 5 um to about 100 um, from about 5 um to about 50 um, or from about 10 um to about 50 um. In some cases, the height of a channel is less than 100 um, less than 80 um, less than 60 um, less than 40 um or less than 20 um.

[00158] In some embodiments, the diameter of a channel, locus (*e.g.*, in a substantially planar substrate) or both channel and locus (*e.g.*, in a three-dimensional substrate wherein a locus corresponds to a channel) is from about 1 um to about 1000 um, from about 1 um to about 500 um, from about 1 um to about 200 um, from about 1 um to about 100 um, from about 5 um to about 100 um, or from about 10 um to about 100 um, for example, about 90 um, 80 um, 70 um, 60 um, 50 um, 40 um, 30 um, 20 um or 10 um. In some embodiments, the diameter of a channel, locus, or both channel and locus is less than about 100 um, 90 um, 80 um, 70 um, 60 um, 50 um, 40 um, 30 um, 20 um or 10 um. In some embodiments, the distance between the center of two adjacent channels, loci, or channels and loci is from about 1 um to about 500 um, from about 1 um to about 200 um, from about 1 um to about 100 um, from about 5 um to about 200 um, from about 5 um to about 100 um, from about 5 um to about 50 um, or from about 5 um to about 30 um, for example, about 20 um.

[00159] Surface Modifications

[00160] In various embodiments, surface modifications are employed for the chemical and/or physical alteration of a surface by an additive or subtractive process to change one or more chemical and/or physical properties of a substrate surface or a selected site or region of a substrate surface. For example, surface modifications include, without limitation, (1) changing the wetting properties of a surface, (2) functionalizing a surface, i.e., providing, modifying or substituting surface functional groups, (3) defunctionalizing a surface, i.e., removing surface functional groups, (4) otherwise altering the chemical composition of a surface, *e.g.*, through etching, (5) increasing or decreasing surface roughness, (6) providing a coating on a surface, *e.g.*, a coating that exhibits wetting properties that are different from the wetting properties of the surface, and/or (7) depositing particulates on a surface.

[00161] In some cases, the addition of a chemical layer on top of a surface (referred to as adhesion promoter) facilitates structured patterning of loci on a surface of a substrate. Exemplary surfaces for application of adhesion promotion include, without limitation, glass, silicon, silicon dioxide and silicon nitride. In some cases, the adhesion promoter is a chemical with a high surface energy. In some embodiments, a second chemical layer is deposited on a surface of a substrate. In some cases, the second chemical layer has a low surface energy. In some cases, surface energy of a chemical layer coated on a surface supports localization of droplets on the surface. Depending on the patterning arrangement selected, the proximity of loci and/or area of fluid contact at the loci are alterable.

[00162] In some embodiments, a substrate surface, or resolved loci, onto which nucleic acids or other moieties are deposited, *e.g.*, for oligonucleic acid synthesis, are smooth or substantially planar (*e.g.*, two-dimensional) or have irregularities, such as raised or lowered features (*e.g.*, three-dimensional features). In some embodiments, a substrate surface is modified with one or more different layers of compounds. Such modification layers of interest include, without limitation, inorganic and organic layers such as metals, metal oxides, polymers, small organic molecules and the like. Non-limiting polymeric layers include peptides, proteins, nucleic acids or mimetics thereof (*e.g.*, peptide nucleic acids and the like), polysaccharides, phospholipids, polyurethanes, polyesters, polycarbonates, polyureas, polyamides, polyetheyleneamines, polyarylene sulfides, polysiloxanes, polyimides, polyacetates, and any other suitable compounds described herein or otherwise known in the art. In some cases, polymers are heteropolymeric. In some cases, polymers are homopolymeric. In some cases, polymers comprise functional moieties or are conjugated.

[00163] In some embodiments, resolved loci of a substrate are functionalized with one or more moieties that increase and/or decrease surface energy. In some cases, a moiety is chemically inert.

In some cases, a moiety is configured to support a desired chemical reaction, for example, one or more processes in an oligonucleic acid synthesis reaction. The surface energy, or hydrophobicity, of a surface is a factor for determining the affinity of a nucleotide to attach onto the surface. In some embodiments, a method for substrate functionalization comprises: (a) providing a substrate having a surface that comprises silicon dioxide; and (b) silanizing the surface using, a suitable silanizing agent described herein or otherwise known in the art, for example, an organofunctional alkoxysilane molecule. In some cases, the organofunctional alkoxysilane molecule comprises dimethylchloro-octodecyl-silane, methyldichloro-octodecyl-silane, trichloro-octodecyl-silane, trimethyl-octodecyl-silane, triethyl-octodecyl-silane, or any combination thereof. In some embodiments, a substrate surface comprises functionalized with polyethylene/polypropylene (functionalized by gamma irradiation or chromic acid oxidation, and reduction to hydroxyalkyl surface), highly crosslinked polystyrene-divinylbenzene (derivatized by chloromethylation, and aminated to benzylamine functional surface), nylon (the terminal aminohexyl groups are directly reactive), or etched with reduced polytetrafluoroethylene. Other methods and functionalizing agents are described in U.S. Patent No. 5474796, which is herein incorporated by reference in its entirety.

[00164] In some embodiments, a substrate surface is functionalized by contact with a derivatizing composition that contains a mixture of silanes, under reaction conditions effective to couple the silanes to the substrate surface, typically via reactive hydrophilic moieties present on the substrate surface. Silanization generally covers a surface through self-assembly with organofunctional alkoxysilane molecules. A variety of siloxane functionalizing reagents can further be used as currently known in the art, *e.g.*, for lowering or increasing surface energy. The organofunctional alkoxysilanes are classified according to their organic functions. Non-limiting examples of siloxane functionalizing reagents include hydroxyalkyl siloxanes (silylate surface, functionalizing with diborane and oxidizing the alcohol by hydrogen peroxide), diol (dihydroxyalkyl) siloxanes (silylate surface, and hydrolyzing to diol), aminoalkyl siloxanes (amines require no intermediate functionalizing step), glycidoxysilanes (3-glycidoxypropyl-dimethyl-ethoxysilane, glycidoxy-trimethoxysilane), mercaptosilanes (3-mercaptopropyl-trimethoxysilane, 3-4 epoxycyclohexyl-ethyltrimethoxysilane or 3-mercaptopropyl-methyl-dimethoxysilane), bicyclohepthenyl-trichlorosilane, butyl-aldehydr-trimethoxysilane, or dimeric secondary aminoalkyl siloxanes. Exemplary hydroxyalkyl siloxanes include allyl trichlorochlorosilane turning into 3-hydroxypropyl, or 7-oct-1-enyl trichlorochlorosilane turning into 8-hydroxyoctyl. The diol (dihydroxyalkyl) siloxanes include glycidyl trimethoxysilane-derived (2,3-dihydroxypropyloxy)propyl (GOPS). The aminoalkyl siloxanes include 3-aminopropyl

trimethoxysilane turning into 3-aminopropyl (3-aminopropyl-triethoxysilane, 3-aminopropyl-diethoxy-methylsilane, 3-aminopropyl-dimethyl-ethoxysilane, or 3-aminopropyl-trimethoxysilane). Exemplary dimeric secondary aminoalkyl siloxanes include bis (3-trimethoxysilylpropyl) amine turning into bis(silyloxylpropyl)amine. In some embodiments, the functionalizing agent comprises 11-acetoxyundecyltriethoxysilane, n-decyltriethoxysilane, (3-aminopropyl)trimethoxysilane, (3-aminopropyl)triethoxysilane, glycidyloxypropyl/trimethoxysilane and N-(3-triethoxysilylpropyl)-4-hydroxybutyramide.

[00165] Oligonucleic Acid Synthesis

[00166] Methods for oligonucleic acid synthesis, in various embodiments, include processes involving phosphoramidite chemistry. In some embodiments, oligonucleic acid synthesis comprises coupling a base with phosphoramidite. In some embodiments, oligonucleic acid synthesis comprises coupling a base by deposition of phosphoramidite under coupling conditions, wherein the same base is optionally deposited with phosphoramidite more than once, i.e., double coupling. In some embodiments, oligonucleic acid synthesis comprises capping of unreacted sites. In some cases, capping is optional. In some embodiments, oligonucleic acid synthesis comprises oxidation. In some embodiments, oligonucleic acid synthesis comprises deblocking or detritylation. In some embodiments, oligonucleic acid synthesis comprises sulfurization. In some cases, oligonucleic acid synthesis comprises either oxidation or sulfurization. In some embodiments, between one or each step during an oligonucleic acid synthesis reaction, the substrate is washed, for example, using tetrazole or acetonitrile. Time frames for any one step in a phosphoramidite synthesis method include less than about 2 min, 1 min, 50 sec, 40 sec, 30 sec, 20 sec and 10 sec.

[00167] Oligonucleic acid synthesis using a phosphoramidite method comprises the subsequent addition of a phosphoramidite building block (*e.g.*, nucleoside phosphoramidite) to a growing oligonucleic acid chain for the formation of a phosphite triester linkage. Phosphoramidite oligonucleic acid synthesis proceeds in the 3' to 5' direction. Phosphoramidite oligonucleic acid synthesis allows for the controlled addition of one nucleotide to a growing nucleic acid chain per synthesis cycle. In some embodiments, each synthesis cycle comprises a coupling step. Phosphoramidite coupling involves the formation of a phosphite triester linkage between an activated nucleoside phosphoramidite and a nucleoside bound to the substrate, for example, via a linker. In some embodiments, the nucleoside phosphoramidite is provided to the substrate activated. In some embodiments, the nucleoside phosphoramidite is provided to the substrate with an activator. In some embodiments, nucleoside phosphoramidites are provided to the substrate in a 1.5, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90,

100-fold excess or more over the substrate-bound nucleosides. In some embodiments, the addition of nucleoside phosphoramidite is performed in an anhydrous environment, for example, in anhydrous acetonitrile. Following addition of a nucleoside phosphoramidite, the substrate is optionally washed. In some embodiments, the coupling step is repeated one or more additional times, optionally with a wash step between nucleoside phosphoramidite additions to the substrate. In some embodiments, an oligonucleic acid synthesis method used herein comprises 1, 2, 3 or more sequential coupling steps. Prior to coupling, in many cases, the nucleoside bound to the substrate is de-protected by removal of a protecting group, where the protecting group functions to prevent polymerization. A common protecting group is 4,4'-dimethoxytrityl (DMT).

[00168] Following coupling, phosphoramidite oligonucleic acid synthesis methods optionally comprise a capping step. In a capping step, the growing oligonucleic acid is treated with a capping agent. A capping step is useful to block unreacted substrate-bound 5'-OH groups after coupling from further chain elongation, preventing the formation of oligonucleic acids with internal base deletions. Further, phosphoramidites activated with 1H-tetrazole may react, to a small extent, with the O6 position of guanosine. Without being bound by theory, upon oxidation with I_2 /water, this side product, possibly via O6-N7 migration, may undergo depurination. The apurinic sites may end up being cleaved in the course of the final deprotection of the oligonucleic acid thus reducing the yield of the full-length product. The O6 modifications may be removed by treatment with the capping reagent prior to oxidation with I_2/water. In some embodiments, inclusion of a capping step during oligonucleic acid synthesis decreases the error rate as compared to synthesis without capping. As an example, the capping step comprises treating the substrate-bound oligonucleic acid with a mixture of acetic anhydride and 1-methylimidazole. Following a capping step, the substrate is optionally washed.

[00169] In some embodiments, following addition of a nucleoside phosphoramidite, and optionally after capping and one or more wash steps, the substrate bound growing nucleic acid is oxidized. The oxidation step comprises the phosphite triester is oxidized into a tetracoordinated phosphate triester, a protected precursor of the naturally occurring phosphate diester internucleoside linkage. In some cases, oxidation of the growing oligonucleic acid is achieved by treatment with iodine and water, optionally in the presence of a weak base (*e.g.*, pyridine, lutidine, collidine). Oxidation may be carried out under anhydrous conditions using, *e.g.* tert-Butyl hydroperoxide or (1S)-(+)-(10-camphorsulfonyl)-oxaziridine (CSO). In some methods, a capping step is performed following oxidation. A second capping step allows for substrate drying, as residual water from oxidation that may persist can inhibit subsequent coupling. Following oxidation, the substrate and growing oligonucleic acid is optionally washed. In some embodiments,

the step of oxidation is substituted with a sulfurization step to obtain oligonucleotide phosphorothioates, wherein any capping steps can be performed after the sulfurization. Many reagents are capable of the efficient sulfur transfer, including but not limited to 3-(Dimethylaminomethylidene)amino)-3H-1,2,4-dithiazole-3-thione, DDTT, 3H-1,2-benzodithiol-3-one 1,1-dioxide, also known as Beaucage reagent, and N,N,N'N'-Tetraethylthiuram disulfide (TETD).

[00170] In order for a subsequent cycle of nucleoside incorporation to occur through coupling, the protected 5' end of the substrate bound growing oligonucleic acid is removed so that the primary hydroxyl group is reactive with a next nucleoside phosphoramidite. In some embodiments, the protecting group is DMT and deblocking occurs with trichloroacetic acid in dichloromethane. Conducting detritylation for an extended time or with stronger than recommended solutions of acids may lead to increased depurination of solid support-bound oligonucleotide and thus reduces the yield of the desired full-length product. Methods and compositions of the invention described herein provide for controlled deblocking conditions limiting undesired depurination reactions. In some cases, the substrate bound oligonucleic acid is washed after deblocking. In some cases, efficient washing after deblocking contributes to synthesized oligonucleic acids having a low error rate.

[00171] Methods for the synthesis of oligonucleic acids typically involve an iterating sequence of the following steps: application of a protected monomer to an actively functionalized surface (*e.g.*, locus) to link with either the activated surface, a linker or with a previously deprotected monomer; deprotection of the applied monomer so that it is reactive with a subsequently applied protected monomer; and application of another protected monomer for linking. One or more intermediate steps include oxidation or sulfurization. In some cases, one or more wash steps precede or follow one or all of the steps.

[00172] Methods for phosphoramidite based oligonucleic acid synthesis comprise a series of chemical steps. In some embodiments, one or more steps of a synthesis method involve reagent cycling, where one or more steps of the method comprise application to the substrate of a reagent useful for the step. For example, reagents are cycled by a series of liquid deposition and vacuum drying steps. For substrates comprising three-dimensional features such as wells, microwells, channels and the like, reagents are optionally passed through one or more regions of the substrate via the wells and/or channels.

[00173] Oligonucleic acids synthesized using the methods and/or substrates described herein comprise, in various embodiments, at least about 20, 30, 40, 50, 60, 70, 75, 80, 90, 100, 120, 150 or more bases. In some embodiments, at least about 1 pmol, 10 pmol, 20 pmol, 30 pmol, 40 pmol, 50

pmol, 60 pmol, 70 pmol, 80 pmol, 90 pmol, 100 pmol, 150 pmol, 200 pmol, 300 pmol, 400 pmol, 500 pmol, 600 pmol, 700 pmol, 800 pmol, 900 pmol, 1 nmol, 5 nmol, 10 nmol, 100 nmol or more of an oligonucleic acid is synthesized within a locus. Methods for oligonucleic acid synthesis on a surface provided herein allow for synthesis at a fast rate. As an example, at least 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, 55, 60, 70, 80, 90, 100, 125, 150, 175, 200 nucleotides per hour, or more are synthesized. Nucleotides include adenine, guanine, thymine, cytosine, uridine building blocks, or analogs/modified versions thereof. In some embodiments, libraries of oligonucleic acids are synthesized in parallel on substrate. For example, a substrate comprising about or at least about 100; 1,000; 10,000; 100,000; 1,000,000; 2,000,000; 3,000,000; 4,000,000; or 5,000,000 resolved loci is able to support the synthesis of at least the same number of distinct oligonucleic acids, wherein oligonucleic acid encoding a distinct sequence is synthesized on a resolved locus. In some embodiments, a library of oligonucleic acids are synthesized on a substrate with low error rates described herein in less than about three months, two months, one month, three weeks, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2 days, 24 hours or less. In some embodiments, larger nucleic acids assembled from an oligonucleic acid library synthesized with low error rate using the substrates and methods described herein are prepared in less than about three months, two months, one month, three weeks, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2 days, 24 hours or less.

[00174] Once large oligonucleic acids for generation are selected, a predetermined library of oligonucleic acids is designed for de novo synthesis. Various suitable methods are known for generating high density oligonucleic acid arrays. In the workflow example, a substrate surface layer is provided. In the example, chemistry of the surface is altered in order to improve the oligonucleic acid synthesis process. Areas of low surface energy are generated to repel liquid while areas of high surface energy are generated to attract liquids. The surface itself may be in the form of a planar surface or contain variations in shape, such as protrusions or microwells which increase surface area. In the workflow example, high surface energy molecules selected serve a dual function of supporting DNA chemistry, as disclosed in International Patent Application Publication WO/2015/021080, which is herein incorporated by reference in its entirety.

[00175] In situ preparation of oligonucleic acid arrays is generated on a solid support and utilizes single nucleotide extension process to extend multiple oligomers in parallel. A deposition device, such as an oligonucleic acid synthesizer, is designed to release reagents in a step wise fashion such that multiple oligonucleic acids extend, in parallel, one residue at a time to generate oligomers with a predetermined nucleic acid sequence. In some cases, oligonucleic acids are cleaved from the surface at this stage. Cleavage includes gas cleavage, *e.g.*, with ammonia or methylamine.

[00176] Computer systems

[00177] Any of the systems described herein, may be operably linked to a computer and may be automated through a computer either locally or remotely. In various embodiments, the methods and systems of the invention may further comprise software programs on computer systems and use thereof. Accordingly, computerized control for the synchronization of the dispense/vacuum/refill functions such as orchestrating and synchronizing the material deposition device movement, dispense action and vacuum actuation are within the bounds of the invention. The computer systems may be programmed to interface between the user specified base sequence and the position of a material deposition device to deliver the correct reagents to specified regions of the substrate.

[00178] The computer system **800** illustrated in **FIG. 8** may be understood as a logical apparatus that can read instructions from media **811** and/or a network port **805**, which can optionally be connected to server **809** having fixed media **812**. The system, such as shown in **FIG. 8** can include a CPU **801**, disk drives **803**, optional input devices such as keyboard **815** and/or mouse **816** and optional monitor **807**. Data communication can be achieved through the indicated communication medium to a server at a local or a remote location. The communication medium can include any means of transmitting and/or receiving data. For example, the communication medium can be a network connection, a wireless connection or an internet connection. Such a connection can provide for communication over the World Wide Web. It is envisioned that data relating to the present disclosure can be transmitted over such networks or connections for reception and/or review by a party **822** as illustrated in **FIG. 8**.

[00179] **FIG. 9** is a block diagram illustrating a first example architecture of a computer system **900** that can be used in connection with example embodiments of the present invention. As depicted in **FIG. 9**, the example computer system can include a processor **902** for processing instructions. Non-limiting examples of processors include: Intel XeonTM processor, AMD OpteronTM processor, Samsung 32-bit RISC ARM 1176JZ(F)-S v1.0TM processor, ARM Cortex-A8 Samsung S5PC100TM processor, ARM Cortex-A8 Apple A4TM processor, Marvell PXA 930TM processor, or a functionally-equivalent processor. Multiple threads of execution can be used for parallel processing. In some embodiments, multiple processors or processors with multiple cores can also be used, whether in a single computer system, in a cluster, or distributed across systems over a network comprising a plurality of computers, cell phones, and/or personal data assistant devices.

[00180] As illustrated in **FIG. 9**, a high speed cache **904** can be connected to, or incorporated in, the processor **902** to provide a high speed memory for instructions or data that have been recently, or are frequently, used by processor **902**. The processor **902** is connected to a north bridge **906** by a

processor bus **908**. The north bridge **906** is connected to random access memory (RAM) **910** by a memory bus **912** and manages access to the RAM **910** by the processor **902**. The north bridge **906** is also connected to a south bridge **914** by a chipset bus **916**. The south bridge **914** is, in turn, connected to a peripheral bus **918**. The peripheral bus can be, for example, PCI, PCI-X, PCI Express, or other peripheral bus. The north bridge and south bridge are often referred to as a processor chipset and manage data transfer between the processor, RAM, and peripheral components on the peripheral bus **918**. In some alternative architectures, the functionality of the north bridge can be incorporated into the processor instead of using a separate north bridge chip. In some embodiments, system **900** can include an accelerator card **922** attached to the peripheral bus **918**. The accelerator can include field programmable gate arrays (FPGAs) or other hardware for accelerating certain processing. For example, an accelerator can be used for adaptive data restructuring or to evaluate algebraic expressions used in extended set processing.

[00181] Software and data are stored in external storage **924** and can be loaded into RAM **910** and/or cache **904** for use by the processor. The system **900** includes an operating system for managing system resources; non-limiting examples of operating systems include: Linux, WindowsTM, MACOSTM, BlackBerry OSTM, iOSTM, and other functionally-equivalent operating systems, as well as application software running on top of the operating system for managing data storage and optimization in accordance with example embodiments of the present invention. In this example, system **900** also includes network interface cards (NICs) **920** and **921** connected to the peripheral bus for providing network interfaces to external storage, such as Network Attached Storage (NAS) and other computer systems that can be used for distributed parallel processing.

[00182] **FIG. 10** is a diagram showing a network **1000** with a plurality of computer systems **1002a**, and **1002b**, a plurality of cell phones and personal data assistants **1002c**, and Network Attached Storage (NAS) **1004a**, and **1004b**. In example embodiments, systems **1002a**, **1002b**, and **1002c** can manage data storage and optimize data access for data stored in Network Attached Storage (NAS) **1004a** and **1004b**. A mathematical model can be used for the data and be evaluated using distributed parallel processing across computer systems **1002a**, and **1002b**, and cell phone and personal data assistant systems **1002c**. Computer systems **1002a**, and **1002b**, and cell phone and personal data assistant systems **1002c** can also provide parallel processing for adaptive data restructuring of the data stored in Network Attached Storage (NAS) **1004a** and **1004b**. **FIG. 10** illustrates an example only, and a wide variety of other computer architectures and systems can be used in conjunction with the various embodiments of the present invention. For example, a blade server can be used to provide parallel processing. Processor blades can be connected through a

back plane to provide parallel processing. Storage can also be connected to the back plane or as Network Attached Storage (NAS) through a separate network interface. In some example embodiments, processors can maintain separate memory spaces and transmit data through network interfaces, back plane or other connectors for parallel processing by other processors. In other embodiments, some or all of the processors can use a shared virtual address memory space.

[00183] **FIG. 11** is a block diagram of a multiprocessor computer system using a shared virtual address memory space in accordance with an example embodiment. The system includes a plurality of processors **1102a-f** that can access a shared memory subsystem **1104**. The system incorporates a plurality of programmable hardware memory algorithm processors (MAPs) **1106a-f** in the memory subsystem **1104**. Each MAP **1106a-f** can comprise a memory **1108a-f** and one or more field programmable gate arrays (FPGAs) **1110a-f**. The MAP provides a configurable functional unit and particular algorithms or portions of algorithms can be provided to the FPGAs **1110a-f** for processing in close coordination with a respective processor. For example, the MAPs can be used to evaluate algebraic expressions regarding the data model and to perform adaptive data restructuring in example embodiments. In this example, each MAP is globally accessible by all of the processors for these purposes. In one configuration, each MAP can use Direct Memory Access (DMA) to access an associated memory **1108a-f**, allowing it to execute tasks independently of, and asynchronously from, the respective microprocessor **1102a-f**. In this configuration, a MAP can feed results directly to another MAP for pipelining and parallel execution of algorithms.

[00184] The above computer architectures and systems are examples only, and a wide variety of other computer, cell phone, and personal data assistant architectures and systems can be used in connection with example embodiments, including systems using any combination of general processors, co-processors, FPGAs and other programmable logic devices, system on chips (SOCs), application specific integrated circuits (ASICs), and other processing and logic elements. In some embodiments, all or part of the computer system can be implemented in software or hardware. Any variety of data storage media can be used in connection with example embodiments, including random access memory, hard drives, flash memory, tape drives, disk arrays, Network Attached Storage (NAS) and other local or distributed data storage devices and systems.

[00185] In example embodiments, the computer system can be implemented using software modules executing on any of the above or other computer architectures and systems. In other embodiments, the functions of the system can be implemented partially or completely in firmware, programmable logic devices such as field programmable gate arrays (FPGAs) as referenced in **FIG. 11**, system on chips (SOCs), application specific integrated circuits (ASICs), or other processing and logic elements. For example, the Set Processor and Optimizer can be implemented with

hardware acceleration through the use of a hardware accelerator card, such as accelerator card **922** illustrated in **FIG. 9**.

[00186] The following examples are set forth to illustrate more clearly the principle and practice of embodiments disclosed herein to those skilled in the art and are not to be construed as limiting the scope of any claimed embodiments. Unless otherwise stated, all parts and percentages are on a weight basis.

EXAMPLES

[00187] The following examples are given for the purpose of illustrating various embodiments of the invention and are not meant to limit the present invention in any fashion. The present examples, along with the methods described herein are presently representative of preferred embodiments, are exemplary, and are not intended as limitations on the scope of the invention. Changes therein and other uses which are encompassed within the spirit of the invention as defined by the scope of the claims will occur to those skilled in the art.

[00188] Example 1: Functionalization of a substrate surface

[00189] A substrate was functionalized to support the attachment and synthesis of a library of oligonucleic acids. The substrate surface was first wet cleaned using a piranha solution comprising 90% H_2SO_4 and 10% H_2O_2 for 20 minutes. The substrate was rinsed in several beakers with DI water, held under a DI water gooseneck faucet for 5 min, and dried with N_2. The substrate was subsequently soaked in NH_4OH (1:100; 3 mL:300 mL) for 5 min, rinsed with DI water using a handgun, soaked in three successive beakers with DI water for 1 min each, and then rinsed again with DI water using the handgun. The substrate was then plasma cleaned by exposing the substrate surface to O_2. A SAMCO PC-300 instrument was used to plasma etch O_2 at 250 watts for 1 min in downstream mode.

[00190] The cleaned substrate surface was actively functionalized with a solution comprising N-(3-triethoxysilylpropyl)-4-hydroxybutyramide using a YES-1224P vapor deposition oven system with the following parameters: 0.5 to 1 torr, 60 min, 70 °C, 135 °C vaporizer. The substrate surface was resist coated using a Brewer Science 200X spin coater. SPR™ 3612 photoresist was spin coated on the substrate at 2500 rpm for 40 sec. The substrate was pre-baked for 30 min at 90 °C on a Brewer hot plate. The substrate was subjected to photolithography using a Karl Suss MA6 mask aligner instrument. The substrate was exposed for 2.2 sec and developed for 1 min in MSF 26A. Remaining developer was rinsed with the handgun and the substrate soaked in water for 5 min. The substrate was baked for 30 min at 100 °C in the oven, followed by visual inspection for lithography defects using a Nikon L200. A cleaning process was used to remove residual resist using the SAMCO PC-300 instrument to O_2 plasma etch at 250 watts for 1 min.

[00191] The substrate surface was passively functionalized with a 100 μL solution of perfluorooctyltrichlorosilane mixed with 10 μL light mineral oil. The substrate was placed in a chamber, pumped for 10 min, and then the valve was closed to the pump and left to stand for 10 min. The chamber was vented to air. The substrate was resist stripped by performing two soaks for 5 min in 500 mL NMP at 70 °C with ultrasonication at maximum power (9 on Crest system). The substrate was then soaked for 5 min in 500 mL isopropanol at room temperature with ultrasonication at maximum power. The substrate was dipped in 300 mL of 200 proof ethanol and blown dry with N_2. The functionalized surface was activated to serve as a support for oligonucleic acid synthesis.

[00192] EXAMPLE 2: Synthesis of a 50-mer sequence on an oligonucleic acid synthesis device

[00193] A two dimensional oligonucleic acid synthesis device was assembled into a flowcell, which was connected to a flowcell (Applied Biosystems "ABI394 DNA Synthesizer"). The two-dimensional oligonucleic acid synthesis device was uniformly functionalized with N-(3-TRIETHOXYSILYLPROPYL)-4-HYDROXYBUTYRAMIDE (Gelest) was used to synthesize an exemplary oligonucleic acid of 50 bp ("50-mer oligonucleic acid") using oligonucleic acid synthesis methods described herein.

[00194] The sequence of the 50-mer was as described in SEQ ID NO.: 1. 5'AGACAATCAACCATTTGGGGTGGACAGCCTTGACCTCTAGACTTCGGCAT##TTTTTTTTTT3' (SEQ ID NO.: 1), where # denotes Thymidine-succinyl hexamide CED phosphoramidite (CLP-2244 from ChemGenes), which is a cleavable linker enabling the release of oligonucleic acids from the surface during deprotection.

[00195] The synthesis was done using standard DNA synthesis chemistry (coupling, capping, oxidation, and deblocking) according to the protocol in **Table 5** and an ABI synthesizer.

Table 5:

	Table 5	
General DNA Synthesis Process Name	**Process Step**	**Time (sec)**
WASH (Acetonitrile Wash Flow)	Acetonitrile System Flush	4
	Acetonitrile to Flowcell	23
	N2 System Flush	4
	Acetonitrile System Flush	4
DNA BASE ADDITION (Phosphoramidite + Activator Flow)	Activator Manifold Flush	2
	Activator to Flowcell	6
	Activator + Phosphoramidite to Flowcell	6
	Activator to Flowcell	0.5

	Table 5	
General DNA Synthesis Process Name	**Process Step**	**Time (sec)**
	Activator + Phosphoramidite to Flowcell	5
	Activator to Flowcell	0.5
	Activator + Phosphoramidite to Flowcell	5
	Activator to Flowcell	0.5
	Activator + Phosphoramidite to Flowcell	5
	Incubate for 25sec	25
WASH (Acetonitrile Wash Flow)	Acetonitrile System Flush	4
	Acetonitrile to Flowcell	15
	N2 System Flush	4
	Acetonitrile System Flush	4
DNA BASE ADDITION (Phosphoramidite + Activator Flow)	Activator Manifold Flush	2
	Activator to Flowcell	5
	Activator + Phosphoramidite to Flowcell	18
	Incubate for 25sec	25
WASH (Acetonitrile Wash Flow)	Acetonitrile System Flush	4
	Acetonitrile to Flowcell	15
	N2 System Flush	4
	Acetonitrile System Flush	4
CAPPING (CapA+B, 1:1, Flow)	CapA+B to Flowcell	15
WASH (Acetonitrile Wash Flow)	Acetonitrile System Flush	4
	Acetonitrile to Flowcell	15
	Acetonitrile System Flush	4
OXIDATION (Oxidizer Flow)	Oxidizer to Flowcell	18
WASH (Acetonitrile Wash Flow)	Acetonitrile System Flush	4
	N2 System Flush	4
	Acetonitrile System Flush	4
	Acetonitrile to Flowcell	15
	Acetonitrile System Flush	4
	Acetonitrile to Flowcell	15
	N2 System Flush	4
	Acetonitrile System Flush	4
	Acetonitrile to Flowcell	23
	N2 System Flush	4
	Acetonitrile System Flush	4
DEBLOCKING (Deblock Flow)	Deblock to Flowcell	36
WASH (Acetonitrile Wash	Acetonitrile System Flush	4

	Table 5	
General DNA Synthesis Process Name	**Process Step**	**Time (sec)**
Flow)	N2 System Flush	4
	Acetonitrile System Flush	4
	Acetonitrile to Flowcell	18
	N2 System Flush	4.13
	Acetonitrile System Flush	4.13
	Acetonitrile to Flowcell	15

[00196] The phosphoramidite/activator combination was delivered similar to the delivery of bulk reagents through the flowcell. No drying steps were performed as the environment stays “wet” with reagent the entire time.

[00197] The flow restrictor was removed from the ABI 394 synthesizer to enable faster flow. Without flow restrictor, flow rates for amidites (0.1 M in ACN), Activator, (0.25 M Benzoylthiotetrazole ("BTT"; 30-3070-xx from GlenResearch) in ACN), and Ox (0.02 M I2 in 20% pyridine, 10% water, and 70% THF) were roughly ~100uL/sec, for acetonitrile ("ACN") and capping reagents (1:1 mix of CapA and CapB, wherein CapA is acetic anhydride in THF/Pyridine and CapB is 16% 1-methylimidizole in THF), roughly ~200uL/sec, and for Deblock (3% dichloroacetic acid in toluene), roughly ~300uL/sec (compared to ~50uL/sec for all reagents with flow restrictor). The time to completely push out Oxidizer was observed, the timing for chemical flow times was adjusted accordingly and an extra ACN wash was introduced between different chemicals. After oligonucleic acid synthesis, the chip was deprotected in gaseous ammonia overnight at 75 psi. Five drops of water were applied to the surface to recover oligonucleic acids. The recovered oligonucleic acids were then analyzed on a BioAnalyzer small RNA chip (data not shown).

[00198] EXAMPLE 3: Synthesis of a 100-mer sequence on an oligonucleic acid synthesis device

[00199] The same process as described in Example 2 for the synthesis of the 50-mer sequence was used for the synthesis of a 100-mer oligonucleic acid (“100-mer oligonucleic acid”; 5' CGGGATCCTTATCGTCATCGTCGTACAGATCCCGACCCATTTGCTGTCCACCAGTCATG CTAGCCATACCATGATGATGATGATGATGAGAACCCCGCAT##TTTTTTTTTT3', where # denotes Thymidine-succinyl hexamide CED phosphoramidite (CLP-2244 from ChemGenes); SEQ ID NO.: 2) on two different silicon chips, the first one uniformly functionalized with N-(3-TRIETHOXYSILYLPROPYL)-4-HYDROXYBUTYRAMIDE and the second one functionalized with 5/95 mix of 11-acetoxyundecyltriethoxysilane and n-decyltriethoxysilane, and the oligonucleic acids extracted from the surface were analyzed on a BioAnalyzer instrument (data not shown).

[00200] All ten samples from the two chips were further PCR amplified using a forward (5'ATGCGGGGTTCTCATCATC3'; SEQ ID NO.: 3) and a reverse (5'CGGGATCCTTATCGTCATCG3'; SEQ ID NO.: 4) primer in a 50uL PCR mix (25uL NEB Q5 mastermix, 2.5 uL 10 uM Forward primer, 2.5 uL 10 uM Reverse primer, 1 uL oligonucleic acid extracted from the surface, and water up to 50 uL) using the following thermalcycling program:

98 °C, 30 sec

98 °C, 10 sec; 63 °C, 10 sec; 72 °C, 10 sec; repeat 12 cycles

72 °C, 2min

[00201] The PCR products were also run on a BioAnalyzer (data not shown), demonstrating sharp peaks at the 100-mer position. Next, the PCR amplified samples were cloned, and Sanger sequenced. **Table 6** summarizes the results from the Sanger sequencing for samples taken from spots 1-5 from chip 1 and for samples taken from spots 6-10 from chip 2.

Table 6:

Spot	Error rate	Cycle efficiency
1	1/763 bp	99.87%
2	1/824 bp	99.88%
3	1/780 bp	99.87%
4	1/429 bp	99.77%
5	1/1525 bp	99.93%
6	1/1615 bp	99.94%
7	1/531 bp	99.81%
8	1/1769 bp	99.94%
9	1/854 bp	99.88%
10	1/1451 bp	99.93%

[00202] Thus, the high quality and uniformity of the synthesized oligonucleic acids were repeated on two chips with different surface chemistries. Overall, 89%, corresponding to 233 out of 262 of the 100-mers that were sequenced were perfect sequences with no errors.

[00203] Finally, **Table 7** summarizes key error characteristics for the sequences obtained from the oligonucleic acid samples from spots 1-10.

Table 7:

Sample ID/Spot no.	OSA_0046/1	OSA_0047/2	OSA_0048/3	OSA_0049/4	OSA_0050/5	OSA_0051/6	OSA_0052/7	OSA_0053/8	OSA_0054/9	OSA_0055/10
Total Sequences	32	32	32	32	32	32	32	32	32	32
Sequencing Quality	25 of 28	27 of 27	26 of 30	21 of 23	25 of 26	29 of 30	27 of 31	29 of 31	28 of 29	25 of 28
Oligo Quality	23 of 25	25 of 27	22 of 26	18 of 21	24 of 25	25 of 29	22 of 27	28 of 29	26 of 28	20 of 25
ROI Match	2500	2698	2561	2122	2499	2666	2625	2899	2798	2348

Count										
ROI Mutation	2	2	1	3	1	0	2	1	2	1
ROI Multi Base Deletion	0	0	0	0	0	0	0	0	0	0
ROI Small Insertion	1	0	0	0	0	0	0	0	0	0
ROI Single Base Deletion	0	0	0	0	0	0	0	0	0	0
Large Deletion Count	0	0	1	0	0	1	1	0	0	0
Mutation: G>A	2	2	1	2	1	0	2	1	2	1
Mutation: T>C	0	0	0	1	0	0	0	0	0	0
ROI Error Count	3	2	2	3	1	1	3	1	2	1
ROI Error Rate	Err: ~1 in 834	Err: ~1 in 1350	Err: ~1 in 1282	Err: ~1 in 708	Err: ~1 in 2500	Err: ~1 in 2667	Err: ~1 in 876	Err: ~1 in 2900	Err: ~1 in 1400	Err: ~1 in 2349
ROI Minus Primer Error Rate	MP Err: ~1 in 763	MP Err: ~1 in 824	MP Err: ~1 in 780	MP Err: ~1 in 429	MP Err: ~1 in 1525	MP Err: ~1 in 1615	MP Err: ~1 in 531	MP Err: ~1 in 1769	MP Err: ~1 in 854	MP Err: ~1 in 1451

[00204] EXAMPLE 4: sgRNA Design

[00205] A chimera sgRNA sequence with a variable region at the 5' end was designed for direct sequence specific cleavage by the Cas9 protein. *See* **FIG. 4A**. The sgRNA sequence had a base-pairing region of 20 bases for specific DNA binding, which included a seed region of 12 bases. The 5' end of the base-pairing region was designed to be the transcription start site. 3' proximal to the base-pairing region was the dCas9 handled region for Cas9 binding, which was 42 bases in length. 3' proximal to the dCas9 handled region was the *S. pyogenes* terminator region which was 40 bases in length. The dCas9 handled region and the terminator region each were designed to include sequence that would result in a hairpin structure.

[00206] sgRNAs were also designed to target the template (T) or nontemplate (NT) DNA strands, **FIGS. 5A-5B**. sgRNAs designed for targeting the template DNA strand included a base-pairing region of the sgRNA having the same sequence identity as the transcribed sequence. sgRNAs designed for targeting the nontemplate DNA strand included the base-pairing region of the sgRNA that was a reverse-complement of the transcribed sequence.

[00207] In an additional arrangement, a T7 promoter was designed immediately upstream of variable base-pairing region. *See* **FIGS. 6A-6B**. The T7 promoter region was added to enable *in vitro* production of the sgRNA with T7 polymerase.

[00208] EXAMPLE 5: Synthesis of DNA encoding for sgRNA - Design and Polymerase Analysis

[00209] DNA oligonucleic acids were designed as fragments that, when joined, encode for an sgRNA sequence. **FIG. 12**. The sgRNAs were designed for inclusion of a T7 promoter immediately upstream of a variable sequence region **1233**. Following de novo synthesis of the DNA oligonucleic acids, an amplification reaction was performed to join and extend overlapping fragments.

[00210] Transcription of the DNA oligonucleic acids at **1201** resulted in *in vitro* production of the sgRNA with T7 polymerase from a DNA template.

[00211] A sequence for Design 1 **1220**, Design 2 **1222**, Design 3 **1224**, and Design 4 **1226** was designed as indicated in **Table 8**. The sequence for each Design 1, Design 2, Design 3, and Design 4 comprises a T7 promoter, a variable sequence portion, and a constant sequence region (the handle and terminator) (**Table 8**). Specifically, the constant sequence region as seen in **FIG. 12** comprises a Cas9 handle hairpin comprising base pairing regions **1211**, **1213**, **1215**, **1217**, **1223**, and **1225**, and a terminator hairpin comprising base pairing regions **1219** and **1221**.

Table 8.

SEQ ID NO	Name	Sequence
12	Design 1	TAATACGACTCACTATAGGGGATGCGCGCAGTTGTCCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTT
13	T7 Promoter of Design 1	TAATACGACTCACTATA
14	Variable Sequence of Design 1	GGATGCGCGCAGTTGTCC
15	Handle and Terminator of Design 1	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTT
44	Design 2	GAAATTAATACGACTCACTATAGGGGATGCGCGCAGTTGTCCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTT
13	T7 Promoter of Design 2	TAATACGACTCACTATA
14	Variable Sequence of Design 2	GGATGCGCGCAGTTGTCC
15	Handle and Terminator of Design 2	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTT
16	Design 3	GAGCTAATACGACTCACTATAGGGGATGCGCGCAGTTGTCCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTT
13	T7 Promoter of Design 3	TAATACGACTCACTATA
14	Variable Sequence of Design 3	GGATGCGCGCAGTTGTCC
15	Handle and Terminator of Design 3	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTT

17	Design 4	CGAGCTAATACGACTCACTATAGGGGATGCGCGCA GTTGTCCGTTTTAGAGCTAGAAATAGCAAGTTAAAA TAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCA CCGAGTCGGTGCTTTT
13	T7 Promoter of Design 4	TAATACGACTCACTATA
14	Variable Sequence of Design 4	GGATGCGCGCAGTTGTCC
15	Handle and Terminator of Design 4	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC TAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTC GGTGCTTTT

[00212] For *in vitro* analysis, it is noted that the T7 RNA polymerase promoter region should be double stranded for recognition by T7 RNA polymerase. An antisense oligonucleic acid was used for hybridization: 5'- TAATACGACTCACTATAGG- 3' (SEQ ID NO: 18). In addition, **Table 9** provides a list of primers that were used for analysis of 4 different sets of template and amplification oligonucleic acids. See **FIG. 12**.

Table 9.

SEQ ID NO	**Name (# bases - melting temp.)**	**Sequence**
19	SgR1-R1 (80bp)	AAAAGCACCGACTCGGTGCCACTTTTTCAAGTTG ATAACGGACTAGCCTTATTTTAACTTGCTATTTCT AGCTCTAAAAC
20	SgR1-F1 (58bp - 51ºC/54ºC)	TAATACGACTCACTATAGGGGATGCGCGCAGTTG TCCGTTTTAGAGCTAGAAATAGCA
21	SgR1-F2 (65bp - 56ºC/56ºC)	GAAATTAATACGACTCACTATAGGGGATGCGCGC AGTTGTCCGTTTTAGAGCTAGAAATAGCAAG
22	SgR1-F3 (66bp - 59ºC/58ºC)	GAGCTAATACGACTCACTATAGGGGATGCGCGCA GTTGTCCGTTTTAGAGCTAGAAATAGCAAGTT
23	SgR1-F4 (78bp – 62ºC/62ºC)	GAGCTAATACGACTCACTATAGGGGATGCGCGCA GTTGTCCGTTTTAGAGCTAGAAATAGCAAGTTAA AATAAGG
24	SgR1-AR1 (14bp - 52ºC)	AAAAGCACCGACTC
25	SgR1-AR2 (15bp - 56ºC)	AAAAGCACCGACTCG
26	SgR1-AR3 (16bp - 60ºC)	AAAAGCACCGACTCGG
27	SgR1-AR4 (17bp - 62ºC)	AAAAGCACCGACTCGGT
18	sgR1-AF1 (19bp - 51ºC)	TAATACGACTCACTATAGG
28	SgR1-AF2 (24bp - 56ºC)	GAAATTAATACGACTCACTATAGG
29	SgR1-AF3 (25bp - 59ºC)	GAGCTAATACGACTCACTATAGG

30	SgR1-AF4 (25bp - 62ºC)	GCGAGCTAATACGACTCACTATAGG

[00213] The 4 different sets of template and amplification oligonucleic acids were analyzed under a variety of condition to optimize purity and yield of the full length template. 10 ul PCR reactions were performed with the template oligonucleic acids (SgR1-R1 & SgR1-F1, SgR1-F2, SgR1-F3, SgR1-F4) at 100 fMol and the respective sets of amplification primers at Polymerase-1 PCR concentration of 600 nMol. Using the gradient on the Eppendorf Mastercycler, 3 annealing temps (50 ºC, 55 ºC, 60 ºC) were evaluated in a 25 cycle PCR using two high fidelity DNA polymerases (polymerase 1 and 3) and standard DNA polymerase (polymerase 2). **Table 10** provides a summary of reaction conditions and **Table 11** provides the amplification protocol.

Table 10:

Reagents (10ul rxns)		**Pol. 1**			**Pol. 2**			**Pol. 3**	
		1x	15x		1x	15x		1x	15x
Pol		0.1	1.5		0.1	1.5		0.1	1.5
Buffer		2	30		1	15		2	30
dNTP's		0.2	3		0.2	3		0.2	3
Amp Primers (10 uM)		0.6			0.6			0.6	
Template Oligos (100 nM)		1			1			1	
H2O		6.1	91.5		7.1	106.5		6.1	91.5
Total		10	150		10	150		10	150

Table 11.

Polymerase 1, 2				Polymerase 3		
98	30sec			95	3min	
98	10sec	25x		98	10sec	25x
50 / 55 / 60	15sec			50 / 55 / 60	15sec	
72	10sec			72	10sec	
72	30sec			72	30sec	
4	hold			4	hold	

[00214] Results from Polymerase-1 PCR reactions were run on a BioAnalyzer (data not shown) to estimate the yield, and are summarized in **Table 12**. DNA yield is presented in ng/ul (**Table 12**). Oligonucleic acid designs 3 and 4 each resulted in higher DNA yield than oligonucleic acid designs 1 and 2. Higher annealing temperatures resulted in increased yield as well, with 60 ºC having higher yields.

Table 12.

Polymerase 1	50 ºC	55 ºC	60 ºC
Oligo design 1	0	0	0
Oligo design 2	2.9	4.5	3.7
Oligo design 3	6.8	9.2	10
Oligo design 4	9.9	13	15.3

Yields listed in ng/ul.

[00215] Results from the Polymerase 2 PCR reactions were run on a BioAnalyzer (data not shown) to estimate the yield, and are summarized in **Table 13**. DNA yield is presented in ng/ul (**Table 13**). Again, oligonucleic acid designs 3 and 4 each resulted in higher DNA yield than oligonucleic acid designs 1 and 2. Higher annealing temperatures resulted in increased yield as well, with 60 ºC having higher yields.

Table 13.

Polymerase 2	50 ºC	55 ºC	60 ºC
Oligo design 1	0	0	0
Oligo design 2	7.6	5.9	6.9
Oligo design 3	6.1	8.5	10.5
Oligo design 4	7.4	11.1	19.4

Yields listed in ng/ul.

[00216] Results from the Polymerase 3 PCR reactions were run on a BioAnalyzer (data not shown) to estimate the yield, and are summarized in **Table 14**. DNA yield is presented in ng/ul (**Table 14**). Oligonucleic acid designs 3 and 4 each resulted in higher DNA yield than oligonucleic acid designs 1 and 2. Higher annealing temperatures resulted in increased yield as well, with 60 ºC having higher yields.

Table 14.

Polymerase 3	50 ºC	55 ºC	60 ºC
Oligo design 1	10	13	12.1
Oligo design 2	12.4	14.3	15.9
Oligo design 3	13.2	26.1	28.8
Oligo design 4	16.1	13.2	18.5

Yields listed in ng/ul.

[00217] In sum, oligonucleic acid designs 3 and 4 resulted in increased DNA yield with all three polymerases. In addition, the higher annealing temperature of 60 ºC resulted in increased DNA yield.

[00218] EXAMPLE 6: CRISPR sgRNA Synthesis - Temperature Analysis

[00219] Using oligonucleic acid primers from Example 5, the impact of increase annealing temperature conditions was analyzed after running a PCR reaction as described in Example 5. Amplification product was run on a BioAnalyzer (data not shown) to estimate the yield, and is summarized in **Table 15**. DNA yield is presented in ng/ul (**Table 15**). In sum, Polymerase 3 provides increased DNA yield and 60 ºC annealing temperature resulted in an increased DNA yield.

Table 15.

	Oligo Design 3				Oligo Design 4			
	Polymerase 3		Polymerase 1		Polymerase 3		Polymerase 1	
	60 ºC	65 ºC	60 ºC	65 ºC	60 ºC	65 ºC	60 ºC	65 ºC
25 cycles	27.3	15.2	11.6	3.9	28.4	29.5	13.8	10.2

Yields listed in ng/ul.

[00220] EXAMPLE 7: sgRNA Generation - Structure Free RNA

[00221] Two assembly oligonucleic acids were designed to generate a modified sgRNA template (120 bp) with T7 promoter sequence and terminator, but without the tracrRNA hairpin containing sequence. *See* **Table 16**.

Table 16.

SEQ ID NO	**Name**	**Sequence**
31	sgR2-Template	CGAGCTAATACGACTCACTATAGGGGCACAACGTGGAGGATGGCAGCGTGCAGCTGGCTGATCACTACCAGCAAAACACTCCAATCGGTGATGGTCCTGTTGCACCGAGTCGGTGCTTTT
32	sgR2-F	CGAGCTAATACGACTCACTATAGGGGCACAACGTGGAGGATGGCAGCGTGCAGCTGGCTGATCACTACCAG
33	sgR2-R	AAAGCACCGACTCGGTGCAACAGGACCATCACCGATTGGAGTGTTTTGCTGGTAGTGATCAGCCAGCTG

[00222] The assembly oligonucleic acids were amplified with same primer oligonucleic acids used to amplify the sgRNA in Example 5. The reaction conditions used are summarized in **Table 17**.

Table 17.

Ingredients	1x		PCR Condition	Time	Cycles
Pol 1	0.1		95°C	3 min	
Buffer	2		98 °C	10 sec	25 x
dNTP's	0.2		65 °C	15 sec	
Amp Primers (10um)	0.6		72 °C	10 sec	
Template Oligos (100nm)	1		72 °C	30 sec	
H2O	6.1		4 °C	hold	
Total	10				

[00223] Transcription with T7 RNA polymerase was expected to yield an RNA product of 80bp, devoid of secondary structure. Transcription of the amplification product was carried out with an in vitro transcription kit (NEB HiScribe). The reaction mixture was analyzed on a BioAnalyzer. See **FIGS. 13A-13B**. The modified sgRNA product was cleaner with the structure free design (**FIG. 13B**) than the sgRNA having the tracrRNA hairpin containing sequence (**FIG. 13A**).

[00224] EXAMPLE 8: sgRNA Directed Cas9 Cleavage

[00225] Three sgRNA sequences were designed with a T7 promoter region and each with a different recognition sequence for regions of a 720 bp GFP encoding sequence. Each of the sgRNA sequences was assembled from PCR of two oligonucleic acids. The sgRNA backbone and primers are provided in **Table 18**.

Table 18.

SEQ ID NO	Name	Sequence
34	sgRNA backbone	CGAGCTAATACGACTCACTATAggNNNNNNNNNNNNNNNNNNNGTTTTAGAGCTATGCTGAAAAGCATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTT
35	GFP	ATGcgtAAAggcGAAgagCTGttcACTggtGTCgtcCCTattCTGgtgGAActgGATggtGATgtcAACggtCATaagTTTtccGTGcgtGGCgagGGTgaaGGTgacGCAactAATggtAAActgACGctgAAGttcATCtgtACTactGGTaaaCTGccgGTAcctTGGccgACTctgGTAacgACGctgACTtatGGTgttCAGtgcTTTgctCGTtatCCGgacCATatgAAGcagCATgacTTCttcAAGtccGCCatgCCGgaaGGCtatGTGcagGAAcgcACGattTCCtttAAGgatGACggcACGtacAAAacgCGTgcgGAAgtgAAAtttGAAggcGATaccCTGgtaAACcgcATTgagCTGaaaGGCattGACtttAAAgaaGACggcAATatcCTGggcCATaagCTGgaaTACaatTTTaacAGCcacAATgttTACatcACCgccGATaaaCAAaaaAATggcATTaaaGCGaatTTTaaaAT

		TcgcCACaacGTGgagGATggcAGCgtgCAGctgGCTgatCACtacCAGcaaAACactCCAatcGGTgatGGTcctGTTctgCTGccaGACaatCACtatCTGagcACGcaaAGCgttCTGtctAAAgatCCGaacGAGaaaCGCgatCATatgGTTctgCTGgagTTCgtaACCgcaGCGggcATCacgCATggtATGgatGAActgTACaaaTGAtaa
36	sgR35-F	CGAGCTAATACGACTCACTATAGGAAcgcACGattTCCtttAGTTTTAGAGCTATGCTGAAAAGCATAGC
37	sgR36-F	CGAGCTAATACGACTCACTATAGGCattGACtttAAAgaaGAGTTTTAGAGCTATGCTGAAAAGCATAGC
38	sgR37-F	CGAGCTAATACGACTCACTATAGGagGATggcAGCgtgCAGcGTTTTAGAGCTATGCTGAAAAGCATAGC
39	sgR3-R	AAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATGCTTTTCAGCATAGCTCTAAAAC
30	SgR1-AF4	GCGAGCTAATACGACTCACTATAGG
27	SgR1-AR4	AAAAGCACCGACTCGGT

[00226] The assembly oligonucleic acids were amplified under reaction conditions summarized in **Table 19**.

Table 19.

Ingredients	**1x**		**PCR Condition**	**Time**	**Cycles**
Polymerase 3	0.1		95°C	3 min	
Buffer	2		98 °C	10 sec	25 x
dNTP's	0.2		95 °C	15 sec	
Amp Primers (10 um)	0.6		60 °C	15 sec	
Template Oligos (100 nm)	1		72 °C	10 sec	
H2O	6.1		72 °C	30 sec	
Total	10		4 °C	hold	

[00227] Samples from each sgRNA assembly reaction were analyzed on a BioAnalyzer (**FIGS. 14A-14C**). Transcription reactions using T7 RNA polymerase PCR amplification product were conducted. Samples from each reaction were analyzed on a BioAnalyzer (**FIGS. 14D-14F**).

[00228] Cas9 digests were prepared using GFP amplification product, Cas9 and the transcribed sgRNA. 2 peaks were observed for all three digests, compared to a single peak for the control. (**FIGS. 14G-14J**). Expected and resultant fragments from Cas9 cleavage using the 3 synthesized sgRNAs are listed **Table 20**.

Table 20.

sgRNA	Predicted Fragment 1	Resultant Fragment 1	Predicted Fragment 2	Resultant Fragment 2
sgR35	321	324	439	451
sgR36	342	350	418	430
sgR37	208	137	552	560

[00229] Cas9 digestion samples were purified and analyzed again on a BioAnalyzer (data not shown). Results from purified samples are summarized in **Table 21**.

Table 21.

sgRNA	Predicted Fragment 1	Resultant Fragment 1	Predicted Fragment 2	Resultant Fragment 2
sgR35	321	323	439	451
sgR36	342	353	418	427
sgR37	208	220	552	560

[00230] EXAMPLE 9: Parallel assembly of 29,040 unique oligonucleic acids

[00231] A structure comprising 256 clusters **1505** each comprising 121 loci on a flat silicon plate was manufactured as shown in **FIG. 15**. An expanded view of a cluster is shown in **1510** with 121 loci. Loci from 240 of the 256 clusters provided an attachment and support for the synthesis of oligonucleic acids having distinct sequences. Oligonucleic acid synthesis was performed by phosphoramidite chemistry using general methods from Example 3. Loci from 16 of the 256 clusters were control clusters. The global distribution of the 29,040 unique oligonucleic acids synthesized (240 non-control clusters x 121 oligonucleic acid populations per cluster) is shown in **FIG. 16A**. NGS sequencing confirmed 100% representation of designed oligonucleic acids selected for synthesis. Distribution was measured for each cluster, as shown in **FIG. 16B**. The distribution of unique oligonucleic acids synthesized in 4 representative clusters is shown in **FIG. 17**. On a global level, all oligonucleic acids the designed for synthesis were present and 99% of the oligonucleic acids had abundance that was within 2x of the mean, indicating high synthesis uniformity. This same observation was consistent on a per-cluster level.

[00232] The error rate for each oligonucleic acid was determined using an Illumina MiSeq gene sequencer. The error rate distribution for the 29,040 unique oligonucleic acids is shown in **FIG. 18A** and averages around 1 in 500 bases, with some error rates as low as 1 in 800 bases.

Distribution was measured for each cluster, as shown in **FIG. 18B**. The error rate distribution for unique oligonucleic acids in four representative clusters is shown in **FIG. 19**. The library of 29,040 unique oligonucleic acids was synthesized in less than 20 hours. Analysis of GC percentage v. oligonucleic acid representation across all of the 29,040 unique oligonucleic acids showed that synthesis was uniform despite GC content (roughly 20% to 85% GC per oligonucleic acid) , **FIG. 20**.

[00233] EXAMPLE 10: PCR amplification analysis of de novo synthesized DNA library encoding for sgRNAs

[00234] 9,996 oligonucleic acids 100 bases in length of randomized sequences with varying GC content, from 20-80% GC were designed and synthesized on a structure with a similar arrangement is described in Example 9. To determine the effect of PCR amplification on GC representation, the oligonucleic acid population was amplified for either 6 or 20 cycles with a high fidelity DNA polymerase (DNA polymerase 1). Alternatively, the oligonucleic acid population was amplified using two other high-fidelity PCR enzymes for 6, 8, 10, or 15 cycles, to determine whether polymerase selection had an effect on overall sequence representation post-amplification. Following PCR amplification, samples were prepped for next generation sequencing and sequenced on the Illumina MiSeq platform. 150 base pair SE reads were generated to an approximate read coverage of 100X. Raw FASTQ files were analyzed. Oligonucleic acid representation with either polymerase for 6, 10 or 15 cycles is depicted in **FIG. 21**. Oligonucleic acid uniformity measured by frequency of representation in sequencing reads was assessed for the various conditions and is summarized in **Table 22**.

Table 22.

	Cycles	**% within 1.5x**	**% within 2x**
Polymerase 1	6	72.1%	92.6%
	8	76.1%	90.3%
	10	70.9%	86.6%
	15	64.1%	82.7%
Polymerase 2	6	91.9%	98.9%
	8	89.9%	98.1%
	10	90.1%	98.4%
	15	89.2%	97.9%

[00235] The number of dropouts for each amplified oligonucleic acid population was quantified as shown in **FIG. 22**, amplification cycles v. fraction of population below a 10% of mean threshold. Polymerase 1 dropouts grew quickly whereas Polymerase 2 dropouts stayed relatively constant.

[00236] The impact of over amplification on GC distribution was assessed, **FIG. 23**. Generally, oligonucleic acids with a GC content 30% to 70% followed the trend line, Y=X, and increased in frequency with more cycles. Oligonucleic acids with a GC content greater than 70% were, generally, slightly more frequent after 20 cycles, while oligonucleic acids with a GC content lower than 30% were, generally, slightly more frequent after 6 cycles.

[00237] EXAMPLE 11: Human Epigenetic CRISPR Screen

[00238] A sgRNA screen was performed to introduce mutations into exons that encode functional domains using CRISPR-Cas9. About 10,000 DNA oligonucleic acids were de novo synthesized using methods similar to those described in Example 3 on a silicon chip as described in Example 1 on a structure with a similar arrangement is described in Example 9. Collectively, the oligonucleic acids had an aggregated error rate of about 1:500 or lower. Each oligonucleic acid was up to 200 bases in length, and at least 1 fmole per an oligonucleic acid species was generated. The oligonucleic acids were PCR amplified, cloned into vectors, and electroporated into cultured cells for sgRNA transcription. Nucleic acids were isolated from the cells and sequenced, using next generation sequencing.

[00239] Sequencing results showed highly accurate and uniform library synthesis with minimal bias and high fidelity production of sgRNAs. More reads per guide sequence with minimal sequencing 30% higher recovery of sgRNA with correct sequence for downstream screening compared to competitor pool. *See* **Table 23**. Pooled sequencing results showed more reads per guide sequence and a much tighter distribution of reads (4 logs) compared to 6 logs with the array based competitor pool. *See* **FIGS. 24A-24B**. Sequencing validation of clones showed 100% sgRNA recovery (**FIG. 24A**) and higher sequence accuracy compared to a commercially available array-based pool (**FIG. 24B**). Of the clones that were sequenced, significantly more were recovered with the correct sgRNA sequence. *See* **Table 23**. 100% of the predetermined sequences were represented in the oligonucleic acid population. NGS-based validation of sgRNA clones showed 100% sgRNA recovery and 13% higher accuracy of synthesis per clone compared to the competitor's population (data not shown).

Table 23.

	Synthesized oligo population	Commercially available oligo population
sgRNA oligos recovered	100%	>95.5%
Correct sequence rate (MiSeq)	about 87%	about 74%

Correct sequence (Sanger 10 clones)	about 100%	about 70%
Ave reads per sgRNA in cloned oligo population (100x normalized)	about 256	about 1024

[00240] EXAMPLE 12: Whole Genome sgRNA Library

[00241] A DNA library was designed to include DNAs encoding for sgRNAs for generating clones for 101,000 different oligonucleic acids (5 sgRNAs per 20200 gene targets). 101,000 oligonucleic acids were de novo synthesized using methods similar to those described in Example 3 on a silicon chip as described in Example 1 on a structure with a similar arrangement is described in Example 9. The synthesized oligonucleic acids were PCR amplified, digested and cloned into lentiviral vectors, and transformed into cells. Nucleic acids were isolated from the cells and sequenced, using next generation sequencing. Alternatively, the synthesized oligonucleic acids were PCR amplified to form an amplicon-based library and sequenced.

[00242] A plot of next generation sequencing reads v. number of sgRNAs recovered shows that as the oligonucleic acid pool size increases, the oligonucleic acid population maintained a more uniform tighter distribution of reads across the entire library, with a minimal tail compared to a commercially available array-based reference oligonucleic acid population. **FIG. 25**.

[00243] EXAMPLE 13. Design of sgRNA Libraries with Improved Targeting and Activity

[00244] sgRNA libraries were designed and de novo synthesized using methods similar to those described in Example 3 on a silicon chip as described in Example 1 on a structure with a similar arrangement is described in Example 9. The synthesized oligonucleic acids were PCR amplified, digested and cloned into vectors, and transferred into cells for use for downstream applications including screening and analysis.

[00245] Different sgRNA design parameters were compared including libraries characterized by a NAG PAM, a NGG PAM, high activity, low off-target, and filtered. The sgRNA library designed by methods described herein provided for a higher percentage of sgRNAs resulting in at least 2-fold depletion of gene expression, around 16% of sgRNAs, compared to other commercially available gRNA systems. **FIG. 26A**. The sgRNA libraries also provided for a lower percentage of sgRNAs resulting in zero or negative depletion of gene expression, around 17%, compared to other commercially available gRNA systems. **FIG. 26B**.

[00246] sgRNA-mediated depletion was assessed for essential gene expression levels as well, where the following genes were targeted by sgRNAs: PCNA, PSMA7, RPP21, and SF3B3. Analyzing the number of sgRNAs that exhibited at least 2-fold depletion, the sgRNA library had a higher percentage sgRNAs depleting essential genes as compared to Comparator 1, Comparator 2, and Comparator 3. *See* **Table 24**.

Table 24.

Gene Name	Comparator 1	Comparator 2	Comparator 3	CRISPR Library
PCNA	1/5	2/6	5/9	5/5
PSMA7	1/5	0/6	2/9	3/5
RPP21	3/5	1/6	4/9	2/5
SF3B3	0/5	0/6	4/9	3/5
Average (%)	**25%**	**12.5%**	**42%**	**65%**

[00247] EXAMPLE 14. sgRNA Library for MS2

[00248] A DNA library comprising non-identical DNA sequences encoding for sgRNAs was designed for sequence specific cleavage by the C2c2 protein. The library comprised all possible spacer sequences for C2c2 targeting of bacteriophage MS2 genome. Because mature crRNAs of C2c2 from *Leptotrichia shahii* comprises a maximum spacer length of 28 nucleotides, tiling all possible 28 nucleotide target sites in the bacteriophage genome resulted in a library of about 3500 spacer sequences.

[00249] About 3500 non-identical oligonucleic acids were de novo synthesized using methods similar to those described in Example 3 on a silicon chip as described in Example 1 on a structure with a similar arrangement is described in Example 9. The library of about 3500 sequences were inserted into vectors and transformed into *E. coli*. *E. coli* cells were infected with MS2 bacteriophage using three dilutions of MS2. The library was then screened for sequences that conferred *E. coli* resistance to MS2 infection.

[00250] A number of spacer sequences were found to confer resistance. Comparing spacer representation (crRNA frequencies), many spacer sequences exhibited more than 1.25 $\log_2$-fold enrichment in the three dilutions of MS2 infection whereas no non-targeting spacer sequences were found to be enriched.

[00251] EXAMPLE 15: sgRNA Library for Zebrafish

[00252] A DNA library is designed with sequences encoding for about 130,000 sgRNAs. On average, about 5 sgRNAs templates are designed for each zebrafish gene. The oligonucleic acids are de novo synthesized using methods similar to those described in Example 3 on a silicon chip as described in Example 1 on a structure with a similar arrangement as described in Example 9. De novo synthesis produces the 130,000 oligonucleic acids, each extending from a different locus on the surface of a silicon plate. The oligonucleic acids are removed from the plate, amplified by PCR, and cloned into expression vectors. Each template is subject to sequencing. The sgRNA library is injected into zebrafish embryos. Zebrafish are raised to adulthood. Sperm are then cryopreserved and screened by sequencing to identify the sequence of germline transmitted

insertions and deletions. Following the germline screen, sperm are genotyped by competitive allele-specific PCR.

[00253] EXAMPLE 16: gRNA Library for Mouse

[00254] A DNA library is designed with sequences encoding for about 100,000 sgRNAs. On average, about 5 sgRNAs templates are designed per mouse gene. The oligonucleic acids are de novo synthesized using methods similar to those described in Example 3 on a silicon chip as described in Example 1 on a structure with a similar arrangement as described in Example 9. A sgRNA library encoding for the sgRNA sequences is de novo synthesized to generate 100,000 oligonucleic acids. De novo synthesis produces the 100,000 oligonucleic acids, each extending from a different locus on the surface of a silicon plate. The oligonucleic acids are removed from the plate, amplified by PCR, and cloned into vectors. Each template is subject to sequencing. sgRNA on-target efficiency is verified by surveyor nuclease assay or sequencing. sgRNAs are then microinjected in mouse zygotes with a desired genetic background. Alternately, following verification of sgRNA efficiency, sgRNAs are packaged into viral vectors such as adeno-associated viruses (AAVs). sgRNAs are then stereotactically delivered into mice at a desired location. Expression levels for the preselected target genes are observed in tissue collected from mice.

[00255] EXAMPLE 17: gRNA library for a Receptor Tyrosine Kinases

[00256] A DNA oligonucleic acid library is designed with sequences encoding for 5 sgRNAs targeting genes for 58 human receptor tyrosine kinases listed in **Table 25**, totaling 290 different DNA oligonucleic acids. The oligonucleic acids are de novo synthesized using methods similar to those described in Example 3 on a silicon chip as described in Example 1 on a structure with a similar arrangement as described in Example 9. The oligonucleic acids are removed from the plate, amplified by PCR, cloned into vectors, and transferred into preselected populations of cells. Expression levels for the preselected genes listed in **Table 25** are compared in each preselected populations of cells against a control population of cells exposed to a control vector without the kinase-specific sgRNA.

Table 25.

Gene	Hs NT ACC #	Hs PROT ACC#
ALK	NM_004304	NP_004295
LTK	NM_002344	NP_002335
AXL	NM_001699	NP_001690
MER	NM_006343	NP_006334
TYRO3	NM_006293	NP_006284
DDR1	NM_013993	NP_001945
DDR2	NM_006182	NP_006173
EGFR	NM_005228	NP_005219

ERBB2	NM_004448	NP_004439
ERBB3	NM_001982	NP_001973
ERBB4	NM_005235	NP_005226
EPHA1	NM_005232	NP_005223
EPHA2	NM_004431	NP_004422
EPHA3	NM_005233	NP_005224
EPHA4	NM_004438	NP_004429
EPHA5	L36644	P54756
EPHA6	AL133666	
EPHA7	NM_004440	NP_004431
EPHA8	AB040892	CAB81612
EPHB1	NM_004441	NP_004432
EPHB2	AF025304	AAB94602
EPHB3	NM_004443	NP_004434
EPHB4	NM_004444	NP_004435
EPHB6	NM_004445	NP_004436
EPHX		
FGFR1	M34641	AAA35835
FGFR2	NM_000141	NP_000132
FGFR3	NM_000142	NP_000133
FGFR4	NM_002011	NP_002002
IGF1R	NM_000875	NP_000866
INSR	NM_000208	NP_000199
INSRR	J05046	AAC31759
MET	NM_000245	NP_000236
RON	NM_002447	NP_002438
MUSK	NM_005592	NP_005583
CSF1R	NM_005211	NP_005202
FLT3	NM_004119	NP_0041110
KIT	NM_000222	NP_000213
PDGFRA	NM_006206	NP_006197
PDGFRB	NM_002609	NP_002600
PTK7	NM_002821	NP_002812
RET	X12949	P07949
ROR1	NM_005012	NP_005003
ROR2	NM_004560	NP_004551
ROS1	NM_002944	NP_002935
RYK	S59184	AAB26341
TEK	NM_000459	NP_000450
TIE	NM_005424	NP_005415
NTRK1	NM_002529	NP_002520
NTRK2	NM_006180	NP_006171
NTRK3	NM_002530	NP_002521
VEGFR1	NM_002019	NP_002010
VEGFR2		AAB88005
VEGFR3	NM_002020	NP_002011
AATYK	NM_004920	NP_004911

AATYK2	NM_014916	NP_055731
AATYK3		
DKFZp761P1010	NM_018423	NP_060893

[00257] EXAMPLE 18: gRNA library for Human Kinome

[00258] A DNA oligonucleic acid library is designed with sequences encoding for 5 sgRNAs targeting genes for 518 human kinases, totaling 2,590 different DNA oligonucleic acids. The oligonucleic acids are removed from the plate, amplified by PCR, cloned into vectors, and transferred into preselected populations of cells. Expression levels for the preselected 518 genes are compared in each preselected populations of cells against a control population of cells exposed to a control vector without the kinase-specific sgRNA.

[00259] EXAMPLE 19: gRNA library for Human Phosphatome

[00260] A DNA oligonucleic acid library is designed with sequences encoding for 5 sgRNAs targeting genes for 200 human phosphatases, totaling 1000 different DNA oligonucleic acids. The oligonucleic acids are removed from the plate, amplified by PCR, cloned into vectors, and transferred into preselected populations of cells. Expression levels for the 200 preselected genes are compared in each preselected populations of cells against a control population of cells exposed to a control vector without the kinase-specific sgRNA.

[00261] While preferred embodiments of the present invention have been shown and described herein, it will be obvious to those skilled in the art that such embodiments are provided by way of example only. Numerous variations, changes, and substitutions will now occur to those skilled in the art without departing from the invention. It should be understood that various alternatives to the embodiments of the invention described herein may be employed in practicing the invention. It is intended that the following claims define the scope of the invention and that methods and structures within the scope of these claims and their equivalents be covered thereby.

CLAIMS

WHAT IS CLAIMED IS:

1. A nucleic acid library, wherein the nucleic acid library comprises at least 500 non-identical DNA molecules, wherein each non-identical DNA molecule encodes for a different gRNA sequence, wherein each gRNA sequence comprises a targeting domain complementary to a mammalian gene, and wherein at least about 80% of the at least 500 non-identical DNA molecules are each present in the nucleic acid library in an amount within 2x of a mean frequency for each of the non-identical DNA molecules in the library.
2. The nucleic acid library of claim 1, wherein each non-identical DNA molecule has a GC base content of about 20% to about 85%.
3. The nucleic acid library of claim 1, wherein each non-identical DNA molecule has a GC base content of about 30% to about 70%.
4. The nucleic acid library of claim 1, wherein at least about 90% of the at least 500 non-identical DNA molecules are each present in the nucleic acid library in an amount within 2x of the mean frequency for each of the non-identical DNA molecules in the library.
5. The nucleic acid library of claim 1, wherein at least 99% of the at least 500 non-identical DNA molecules are each present in the nucleic acid library in an amount within 2x of the mean frequency for each of the non-identical DNA molecules in the library.
6. The nucleic acid library of claim 1, wherein the at least 500 non-identical DNA molecules comprise at least 2000 non-identical DNA molecules.
7. The nucleic acid library of claim 1, wherein the at least 500 non-identical DNA molecules comprise at least 3500 non-identical DNA molecules.
8. The nucleic acid library of claim 1, wherein the at least 500 non-identical DNA molecules comprise at least 100,000 non-identical DNA molecules.
9. The nucleic acid library of claim 1, wherein each non-identical DNA molecule comprise up to 200 bases in length.
10. The nucleic acid library of claim 1, wherein each non-identical DNA molecule comprises about 100 to about 200 bases in length.
11. The nucleic acid library of claim 1, wherein the at least 500 non-identical DNA molecules comprises non-identical DNA molecules encoding for gRNA sequences targeting genes in a biological pathway.

12. The nucleic acid library of claim 1, wherein the at least 500 non-identical DNA molecules comprises non-identical DNA molecules encoding for gRNA sequences targeting genes in an entire genome.
13. The nucleic acid library of claim 1, wherein the gRNA is a single gRNA or a dual gRNA.
14. A nucleic acid library, wherein the nucleic acid library comprises at least 2000 non-identical nucleic acids, wherein each non-identical nucleic acid encodes for a different gRNA sequence, wherein each gRNA sequence comprises a targeting domain complementary to a eukaryotic gene, and wherein at least about 80% of the at least 2000 non-identical nucleic acids are present in the nucleic acid library in an amount within 2x of a mean frequency for each of the non-identical nucleic acids in the library.
15. The nucleic acid library of claim 14, wherein each non-identical nucleic acid has a GC base content of about 20% to about 85%.
16. The nucleic acid library of claim 14, wherein each non-identical nucleic acid has a GC base content of about 30% to about 70%.
17. The nucleic acid library of claim 14, wherein at least about 90% of the at least 2000 non-identical nucleic acids are each present in the nucleic acid library in an amount within 2x of the mean frequency for each of the non-identical nucleic acids in the library.
18. The nucleic acid library of claim 14, wherein at least 99% of the at least 2000 non-identical nucleic acids are each present in the nucleic acid library in an amount within 2x of the mean frequency for each of the non-identical nucleic acids in the library.
19. The nucleic acid library of claim 14, wherein each non-identical nucleic acid comprises up to 200 bases in length.
20. The nucleic acid library of claim 14, wherein each non-identical nucleic acid comprises about 100 to about 200 bases in length.
21. The nucleic acid library of claim 14, wherein the at least 2000 non-identical nucleic acids comprise non-identical nucleic acids encoding for gRNA sequences targeting genes in a biological pathway.
22. The nucleic acid library of claim 14, wherein the at least 2000 non-identical nucleic acids comprise non-identical nucleic acids encoding for gRNA sequences targeting genes in an entire genome.
23. The nucleic acid library of claim 14, wherein each non-identical nucleic acid comprises DNA or RNA molecules.

24. An amplicon library, wherein the amplicon library comprises a plurality of non-identical DNA molecules, wherein each non-identical DNA molecule is present in a population of amplification products, wherein each non-identical DNA molecule encodes for a different gRNA sequence, wherein each gRNA sequence comprises a targeting domain complementary to a eukaryotic gene, and wherein at least about 80% of the plurality of non-identical DNA molecules are each present in the amplicon library in an amount within 2x of a mean frequency for each of the non-identical DNA molecules in the library.

25. The amplicon library of claim 24, wherein each non-identical DNA molecule has a GC base content of about 30% to about 70%.

26. The amplicon library of claim 24, wherein the gRNA is a single gRNA or a dual gRNA.