



- (51) **International Patent Classification:**
C12N 9/22 (2006.01)
- (21) **International Application Number:**
PCT/US2015/028514
- (22) **International Filing Date:**
30 April 2015 (30.04.2015)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**
61/986,849 30 April 2014 (30.04.2014) US
- (71) **Applicant:** MASSACHUSETTS INSTITUTE OF TECHNOLOGY [US/US]; 77 Massachusetts Avenue, Cambridge, MA 02139 (US).
- (72) **Inventors:** JACOBSON, Joseph, M.; 223 Grant Avenue, Newton, MA 02459 (US). JAKIMO, Noah; 409 Marlborough Street, Apt. 52, Boston, MA 02115 (US).
- (74) **Agents:** ZUCCHERO, Joseph C. et al.; Elmore Patent Law Group, 484 Groton Road, Westford, MA 01886 (US).
- (81) **Designated States** (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY,

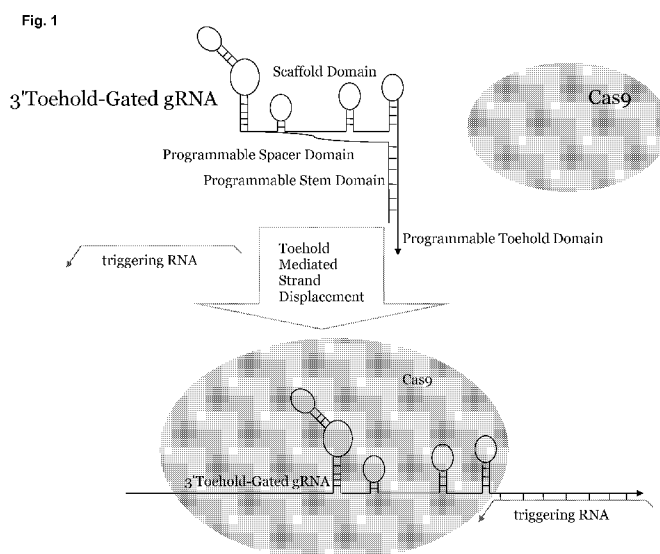
BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

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

Published:

- with international search report (Art. 21(3))
- before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))

(54) **Title:** TOEHOLD-GATED GUIDE RNA FOR PROGRAMMABLE CAS9 CIRCUITRY WITH RNA INPUT



(57) **Abstract:** The present invention relates to a toehold-gated guide RNA (thgRNA) as described herein and the use of thgRNA in cell based bimolecular circuitry to sense, report, and respond to the transcriptome state of the cellular environment. The present invention, referred to as toehold-gated gRNA (thgRNA) adds nucleotide sequences to the 3' and/or 5' ends of a gRNA, or within the scaffold of a gRNA, such that complementary sequences hybridize to form a stem domain in the secondary structure that constrains the distance between spacer and scaffold domains and sterically protects toehold-gated gRNA from activity with Cas9.

TOEHOLD-GATED GUIDE RNA FOR PROGRAMMABLE CAS9 CIRCUITRY WITH RNA INPUT

RELATED APPLICATIONS

- 5 **[0001]** This application claims the benefit of U.S. provisional patent application serial number 61/986,849, filed on April 30, 2014, the contents of which are incorporated herein by reference in their entirety.

BACKGROUND

- 10 **[0002]** Currently there is great interest in data sensing devices capable of measuring a signal and recording what is measured. A standard approach to creating such sensor devices is the use of semiconductor based electronics in which both the sensor and memory are comprised of transistor based integrated circuits. Although such devices can be made small relative to human dimensions, they are limited in the
15 ultimate size reduction which can be achieved by the size limits to which transistor logic can be scaled (current lithographic design rules are ~ 18 nm corresponding to transistor sizes or memory bit sizes of $\sim 100^3$ nm³) thus limiting them from a wide range of applications on the nanoscale such as recording information inside of a cell. Additional shortcomings of electronic circuits is that they generally operate far from
20 the physical limits of energy consumption and represent a manufactured artifact, limiting the ultimate cost which can be achieved of such devices and thus limiting their use in applications where large number of sensors are needed such as monitoring large areas of water or land (e.g. 100's of sq. km.) with high densities of sensors.

- 25 **[0003]** In distinction to electronics, the molecular elements of living cells, particularly DNA and proteins, possess a number of attributes which make them compelling for sensing and data recording. The 4 nucleotides of DNA, Guanine (G), Adenine (A), Thymine (T) and Cytosine (C) each represent 2 bits of information and in double stranded form have a volume of $\sim 0.34^3$ nm³ representing a volumetric bit density about 25 million greater than that of the current state of the art electronic
30 memory. In addition, proteins and protein switches serve as highly selective sensor elements and cells operate at close to physical limits for energy consumption.

- [0004]** The transcriptome of a biological cell enables classification of many important cellular properties, such as disease, metabolism, and cell type. Measuring

transcription levels of ribonucleic acid (RNA) is mainly achieved through cell lysate, which is applied to deoxyribonucleic acid (DNA) microarray or RNA-Seq. It is desirable to design biomolecular circuitry that works in the cell to sense, report, and respond to the transcriptome state. However, past approaches to such circuitry by other groups are not compatible with sensing both micro RNA and messenger RNA and are limited in the number of layers that circuitry can achieve.

[0005] It would be desirable to design and use a single or multi-layered cell based biomolecular circuitry for sensing and recording the transcriptome state of the cellular environment.

SUMMARY

[0006] The present invention, referred to as toehold-gated gRNA (thgRNA) adds nucleotide sequences to the 3' and/or 5' ends of a gRNA, or within the scaffold of a gRNA, such that complementary sequences hybridize to form a stem domain in the secondary structure that constrains the distance between spacer and scaffold domains and sterically protects toehold-gated gRNA from activity with Cas9. Additional sequence on either 3' or 5' end of this stem domain is unpaired, but provides a toehold for hybridization to RNA of interest, i.e., a "triggering RNA". The complementary sequences of a stem domain(s) are also complimentary to the triggering RNA at a region proximal or adjacent to where the toehold is complementary to the triggering RNA which initiates an entropy driven reaction of toehold-mediated strand displacement (Niranjan Srinivas, Thomas E. Ouldrige, Petr Šulc, Joseph M. Schaeffer, Bernard Yurke, Ard A. Louis, Jonathan P. K. Doye, and Erik Winfree, "On the biophysics and kinetics of toehold-mediated DNA strand displacement", Nucleic Acids Research, vol. 41, no. 22, pages 10641-10658, 2013, which is herein incorporated by reference in its entirety) to remove self-hybridization in the stem of toehold-gated gRNA. Complete removal of self-hybridization lifts the stem's constraint on the distance between spacer and scaffold domains, and permits the thgRNA to complex with a Cas9 protein. Further stem domains, each with a toehold domain, can be added to detect and/or implement logic for multiple triggering RNAs.

[0007] The present invention relates to the use of thgRNA in cell based biomolecular circuitry to sense, report and respond to the transcriptome state of the cellular environment.

BRIEF DESCRIPTION OF THE DRAWINGS

[0008] Other aspects, advantages and novel features of the invention will become more apparent from the following detailed description of the invention when considered in conjunction with the accompanying drawings wherein:

5 **[0009]** Fig. 1 is a schematic representation of the RNA triggered activity of toehold-gated guide RNA (thgRNA) with Cas9;

[0010] Fig. 2 depicts toehold-gated guide RNA (thgRNA) with inhibited activity with Cas9 and with toehold-gated gRNA (thgRNA) complexed with substrate;

10 **[0011]** Fig. 3 depicts toehold placement on the opposite terminal of the RNA molecule;

[0012] Fig. 4 depicts an exemplary stem and spacer sharing sequence;

[0013] Fig. 5 depicts layered expression of toehold-gated guide RNA;

[0014] Fig. 6 depicts the initial and intermediate structures of an exemplary
15 toehold-gated guide RNA hairpin that implements AND logic; and

[0015] Fig. 7 depicts an exemplary toehold-gated guide RNA (thgRNA) wherein the stem domain is inserted within the scaffold domain.

DETAILED DESCRIPTION

20 **[0016]** Previous research demonstrates that a non-natural chimeric guide RNA (gRNA) sequence can be used to localize the a Cas9 protein to a region of DNA with complementary sequence to a domain in the gRNA (Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E., "A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity", Science vol. 337, pages
25 816-821, 2012, which is herein incorporated by reference in its entirety). The "spacer" domain of gRNA is complementary to the targeted nucleic acid, e.g., the targeted sequence of DNA. The "scaffold" domain of gRNA has secondary structure that confers affinity to Cas9.

[0017] The present invention, referred to as toehold-gated gRNA (thgRNA) adds
30 nucleotide sequences to the 3' and/or 5' ends of a gRNA, or within the scaffold of a gRNA, such that complementary sequences hybridize to form a stem domain in the secondary structure that constrains the distance between spacer and scaffold domains and sterically protects toehold-gated gRNA from activity with Cas9. Additional

sequence on either 3' or 5' end of this stem domain is unpaired, but provides a toehold for hybridization to RNA of interest, i.e., a "triggering RNA". The complementary sequences of a stem domain(s) are also complimentary to the triggering RNA at a region proximal or adjacent to where the toehold is complementary to the triggering

- 5 RNA which initiates an entropy driven reaction of toehold-mediated strand displacement (Niranjan Srinivas, Thomas E. Ouldridge, Petr Šulc, Joseph M. Schaeffer, Bernard Yurke, Ard A. Louis, Jonathan P. K. Doye, and Erik Winfree, "On the biophysics and kinetics of toehold-mediated DNA strand displacement", Nucleic Acids Research, vol. 41, no. 22, pages 10641-10658, 2013, which is herein
10 incorporated by reference in its entirety) to remove self-hybridization in the stem of toehold-gated gRNA. Complete removal of self-hybridization lifts the stem's constraint on the distance between spacer and scaffold domains, and permits the thgRNA to complex with a Cas9 protein. Further stem domains, each with a toehold domain, can be added to detect and/or implement logic for multiple triggering RNAs.
15 **[0018]** The present invention relates to the use of thgRNA in cell based biomolecular circuitry to sense, report and respond to the transcriptome state of the cellular environment.

- [0019]** The terms "polynucleotide" and "nucleic acid," used interchangeably herein, refer to a polymeric form of nucleotides of any length, either ribonucleotides
20 or deoxyribonucleotides. Thus, this term includes, but is not limited to, single-, double-, or multi-stranded DNA or RNA, genomic DNA, cDNA, DNA-RNA hybrids, or a polymer comprising purine and pyrimidine bases or other natural, chemically or biochemically modified, non-natural, or derivatized nucleotide bases.

- "Oligonucleotide" generally refers to polynucleotides of between about 5 and about
25 100 nucleotides of single- or double-stranded DNA. However, for the purposes of this disclosure, there is no critical upper limit to the length of an oligonucleotide.

- Oligonucleotides are also known as "oligomers" or "oligos" and may be isolated from genes, or chemically synthesized by methods known in the art. The terms
"polynucleotide" and "nucleic acid" should be understood to include, as applicable to
30 the embodiments being described, single-stranded (such as sense or antisense) and double-stranded polynucleotides.

[0020] By "region of complementarity" it is meant that a nucleic acid comprises a sequence of nucleotides that enables it to non-covalently bind, i.e. form

Watson-Crick base pairs and/or G/U base pairs, "anneal", or "hybridize," to another nucleic acid in a sequence-specific, antiparallel, manner (i.e., a nucleic acid specifically binds to a complementary nucleic acid) under the appropriate *in vitro* and/or *in vivo* conditions of temperature and solution ionic strength. As is known in the art, standard Watson-Crick base-pairing includes: adenine (A) pairing with thymidine (T), adenine (A) pairing with uracil (U), and guanine (G) pairing with cytosine (C) [DNA, RNA]. In addition, it is also known in the art that for hybridization between two RNA molecules (e.g., dsRNA), guanine (G) base pairs with uracil (U). For example, G/U base-pairing is partially responsible for the degeneracy (i.e., redundancy) of the genetic code in the context of tRNA anti-codon base-pairing with codons in mRNA. In the context of this disclosure, a guanine (G) of a protein-binding segment (dsRNA duplex) of a thgRNA molecule is considered complementary to a uracil (U), and vice versa. As such, when a G/U base-pair can be made at a given nucleotide position a protein-binding segment (dsRNA duplex), the position is not considered to be non-complementary, but is instead considered to be complementary.

[0021] Hybridization and washing conditions are well known and exemplified in Sambrook, J., Fritsch, E. F. and Maniatis, T. *Molecular Cloning: A Laboratory Manual*, Second Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor (1989), particularly Chapter 11 and Table 11.1 therein; and Sambrook, J. and Russell, W., *Molecular Cloning: A Laboratory Manual*, Third Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor (2001). The conditions of temperature and ionic strength determine the "stringency" of the hybridization.

[0022] Hybridization requires that the two nucleic acids contain complementary sequences, although mismatches between bases are possible. The conditions appropriate for hybridization between two nucleic acids depend on the length of the nucleic acids and the degree of complementation, variables well known in the art. The greater the degree of complementation between two nucleotide sequences, the greater the value of the melting temperature (T_m) for hybrids of nucleic acids having those sequences. For hybridizations between nucleic acids with short stretches of complementarity (e.g. complementarity over 35 or less, 30 or less, 25 or less, 22 or less, 20 or less, or 18 or less nucleotides) the position of mismatches becomes important (see Sambrook et al., *supra*, 11.7-11.8). Typically, the length for a

hybridizable nucleic acid is at least about 10 nucleotides. Illustrative minimum lengths for a hybridizable nucleic acid are: at least about 15 nucleotides; at least about 20 nucleotides; at least about 22 nucleotides; at least about 25 nucleotides; and at least about 30 nucleotides). Furthermore, the skilled artisan will recognize that the temperature and wash solution salt concentration may be adjusted as necessary according to factors such as length of the region of complementation and the degree of complementation.

[0023] It is understood in the art that the sequence of polynucleotide need not be 100% complementary to that of its target nucleic acid to be specifically

hybridizable or hybridizable. Moreover, a polynucleotide may hybridize over one or more segments such that intervening or adjacent segments are not involved in the hybridization event (e.g., a stem structure, a loop structure or hairpin structure). A polynucleotide can comprise at least 70%, at least 80%, at least 90%, at least 95%, at least 99%, or 100% sequence complementarity to a target region within the target nucleic acid sequence to which they are targeted. For example, an antisense nucleic acid in which 18 of 20 nucleotides of the antisense compound are complementary to a target region, and would therefore specifically hybridize, would represent 90 percent complementarity. In this example, the remaining noncomplementary nucleotides may be clustered or interspersed with complementary nucleotides and need not be

contiguous to each other or to complementary nucleotides. Percent complementarity between particular stretches of nucleic acid sequences within nucleic acids can be determined routinely using BLAST programs (basic local alignment search tools) and PowerBLAST programs known in the art (Altschul et al., J. Mol. Biol., 1990, 215, 403-410; Zhang and Madden, Genome Res., 1997, 7, 649-656) or by using the Gap program (Wisconsin Sequence Analysis Package, Version 8 for Unix, Genetics Computer Group, University Research Park, Madison Wis.), using default settings, which uses the algorithm of Smith and Waterman (Adv. Appl. Math., 1981, 2, 482-489).

[0024] The terms "peptide," "polypeptide," and "protein" are used

interchangeably herein, and refer to a polymeric form of amino acids of any length, which can include coded and non-coded amino acids, chemically or biochemically modified or derivatized amino acids, and polypeptides having modified peptide backbones.

[0025] The phrase "capable of associating with" refers to a non-covalent interaction between macromolecules (e.g., between a Cas9 protein and a thgRNA). While in a state of non-covalent interaction, the macromolecules are said to be "associated" or "interacting" or "binding" or "co-localizing". Not all components of a binding interaction need be sequence-specific (e.g., contacts with phosphate residues in a DNA backbone), but some portions of a binding interaction may be sequence-specific. Binding interactions are generally characterized by a dissociation constant (Kd) of less than 10^{-6} M, less than 10^{-7} M, less than 10^{-8} M, less than 10^{-9} M, less than 10^{-10} M, less than 10^{-11} M, less than 10^{-12} M, less than 10^{-13} M, less than 10^{-14} M, or less than 10^{-15} M. "Affinity" refers to the strength of binding, increased binding affinity being correlated with a lower Kd.

[0026] The term "conservative amino acid substitution" refers to the interchangeability in proteins of amino acid residues having similar side chains. For example, a group of amino acids having aliphatic side chains consists of glycine, alanine, valine, leucine, and isoleucine; a group of amino acids having aliphatic - hydroxyl side chains consists of serine and threonine; a group of amino acids having amide containing side chains consisting of asparagine and glutamine; a group of amino acids having aromatic side chains consists of phenylalanine, tyrosine, and tryptophan; a group of amino acids having basic side chains consists of lysine, arginine, and histidine; a group of amino acids having acidic side chains consists of glutamate and aspartate; and a group of amino acids having sulfur containing side chains consists of cysteine and methionine. Exemplary conservative amino acid substitution groups are: valine -leucine - isoleucine, phenylalanine-tyrosine, lysine-arginine, alanine-valine, and asparagine-glutamine.

[0027] A polynucleotide or polypeptide has a certain percent "sequence identity" to another polynucleotide or polypeptide, meaning that, when aligned, that percentage of bases or amino acids are the same, and in the same relative position, when comparing the two sequences. Sequence identity can be determined in a number of different manners. To determine sequence identity, sequences can be aligned using various methods and computer programs (e.g., BLAST, T-COFFEE, MUSCLE, MAFFT, etc.), available over the world wide web at sites including ncbi.nlm.nih.gov/BLAST, ebi.ac.uk/Tools/msa/tcoffee/, ebi.ac.uk/Tools/msa/muscle/,

mafft.cbrc.jp/alignment/software/. See, e.g., Altschul et al. (1990), J. Mol. Bioi. 215:403-10.

[0028] A nucleic acid sequence that "encodes" a particular RNA is a DNA nucleic acid sequence that can be transcribed into RNA. A DNA polynucleotide may
5 encode an RNA (mRNA) that is translated into protein (e.g., a Cas9 protein), or a DNA polynucleotide may encode an RNA that is not translated into protein (e.g. a thgRNA according to the invention).

[0029] A "protein coding sequence" or a sequence that encodes a particular protein or polypeptide, is a nucleic acid sequence that is transcribed into mRNA (in
10 the case of DNA) and is translated (in the case of mRNA) into a polypeptide *in vitro* or *in vivo* when placed under the control of appropriate regulatory sequences. The boundaries of the coding sequence are determined by a start codon at the 5' terminus (N-terminus) and a translation stop nonsense codon at the 3' terminus (C -terminus). A coding sequence can include, but is not limited to, cDNA from prokaryotic or
15 eukaryotic mRNA, genomic DNA sequences from prokaryotic or eukaryotic DNA, and synthetic nucleic acids. A transcription termination sequence will usually be located 3' to the coding sequence.

[0030] As used herein, a "promoter sequence" is a DNA regulatory region capable of binding RNA polymerase and initiating transcription of a downstream (3'
20 direction) coding or non-coding sequence. For purposes of defining the present invention, the promoter sequence is bounded at its 3' terminus by the transcription initiation site and extends upstream (5' direction) to include the minimum number of bases or elements necessary to initiate transcription at levels detectable above background. Within the promoter sequence will be found a transcription initiation
25 site, as well as protein binding domains responsible for the binding of RNA polymerase. Eukaryotic promoters will often, but not always, contain "TATA" boxes and "CAT" boxes. Various promoters, including inducible promoters, may be used to drive the various vectors of the present invention.

[0031] The term "regulatory element" is intended to include promoters,
30 enhancers, internal ribosomal entry sites (IRES), and other expression control elements (e.g. transcription termination signals, such as polyadenylation signals and poly-U sequences), sequences, such as promoters, enhancers, polyadenylation signals, terminators, protein degradation signals, and the like, that provide for and/or regulate

transcription of a non-coding sequence (e.g. a guide RNA) or a coding sequence (e.g., Cas9 polypeptide) and/or regulate translation of an encoded polypeptide. In some embodiments, a regulatory element is operably linked to one or more elements of a CRISPR system so as to drive expression of the one or more elements of the CRISPR system.

[0032] The term "naturally-occurring" or "unmodified" as used herein as applied to a nucleic acid, a polypeptide, a cell, or an organism, refers to a nucleic acid, polypeptide, cell, or organism that is found in nature. For example, a polypeptide or polynucleotide sequence that is present in an organism (including viruses) that can be isolated from a source in nature and which has not been intentionally modified by a human in the laboratory is naturally occurring.

[0033] The term "chimeric" as used herein as applied to a nucleic acid or polypeptide refers to two components that are defined by structures derived from different sources. For example, where "chimeric" is used in the context of a chimeric polypeptide (e.g., a chimeric Cas9 fusion protein), the chimeric polypeptide includes amino acid sequences that are derived from different polypeptides. "Chimeric" in the context of a polynucleotide includes nucleotide sequences derived from different coding regions of the same or different genomes.

[0034] The term "chimeric polypeptide" refers to a polypeptide which is made by the combination (i.e., "fusion") of two otherwise separated segments of amino sequence, usually through human intervention. A polypeptide that comprises a chimeric amino acid sequence is a chimeric polypeptide.

[0035] "Heterologous," as used herein, means a nucleotide or polypeptide sequence that is not found in the native nucleic acid or protein, respectively.

[0036] "Recombinant," as used herein, means that a particular nucleic acid (DNA or RNA) is the product of various combinations of cloning, restriction, polymerase chain reaction (PCR) and/or ligation steps resulting in a construct having a structural coding or non-coding sequence distinguishable from endogenous nucleic acids found in natural systems. DNA sequences encoding polypeptides can be assembled from cDNA fragments or from a series of synthetic oligonucleotides, to provide a synthetic nucleic acid which is capable of being expressed from a recombinant transcriptional unit contained in a cell or in a cell-free transcription and translation system. Genomic DNA comprising the relevant sequences can also be used

in the formation of a recombinant gene or transcriptional unit. Sequences of non-translated DNA may be present 5' or 3' from the open reading frame, where such sequences do not interfere with manipulation or expression of the coding regions, and may indeed act to modulate production of a desired product by various mechanisms (see "DNA regulatory sequences", below). Alternatively, DNA sequences encoding RNA that is not translated may also be considered recombinant. Thus, e.g., the term "recombinant" nucleic acid refers to one which is not naturally occurring, e.g., is made by the artificial combination of two otherwise separated segments of sequence through human intervention. This artificial combination is often accomplished by either chemical synthesis means, or by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques. Such is usually done to replace a codon with a codon encoding the same amino acid, a conservative amino acid, or a non-conservative amino acid. Alternatively, it is performed to join together nucleic acid segments of desired functions to generate a desired combination of functions. This artificial combination is often accomplished by either chemical synthesis means, or by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques. When a recombinant polynucleotide encodes a polypeptide, the sequence of the encoded polypeptide can be naturally occurring ("wild type") or can be a variant (e.g., a mutant) of the naturally occurring sequence. Thus, the term "recombinant" polypeptide does not necessarily refer to a polypeptide whose sequence does not naturally occur. Instead, a "recombinant" polypeptide is encoded by a recombinant DNA sequence, but the sequence of the polypeptide can be naturally occurring ("wild type") or non-naturally occurring (e.g., a variant, a mutant, etc.). Thus, a "recombinant" polypeptide is the result of human intervention, but may be a naturally occurring amino acid sequence.

[0037] A "vector" or "expression vector" is a replicon, such as plasmid, phage, virus, or cosmid, to which another DNA segment, i.e. an "insert", may be attached so as to bring about the replication of the attached segment in a cell.

[0038] An "expression cassette" comprises a DNA coding sequence operably linked to a promoter.

[0039] "Operably linked" refers to a juxtaposition wherein the components so described are in a relationship permitting them to function in their intended manner.

For instance, a promoter is operably linked to a coding sequence if the promoter affects its transcription or expression.

[0040] The terms "recombinant expression vector," or "DNA construct" are used interchangeably herein to refer to a DNA molecule comprising a vector and at least one insert. Recombinant expression vectors are usually generated for the purpose of expressing and/or propagating the insert(s), or for the construction of other recombinant nucleotide sequences. The insert(s) may or may not be operably linked to a promoter sequence and may or may not be operably linked to DNA regulatory sequences.

[0041] A cell has been "genetically modified" or "transformed" or "transfected" by exogenous DNA, e.g. a recombinant expression vector, when such DNA has been introduced inside the cell. The presence of the exogenous DNA results in permanent or transient genetic change. The transforming DNA may or may not be integrated (covalently linked) into the genome of the cell. In prokaryotes, yeast, and mammalian cells for example, the transforming DNA may be maintained on an episomal element such as a plasmid. With respect to eukaryotic cells, a stably transformed cell is one in which the transforming DNA has become integrated into a chromosome so that it is inherited by daughter cells through chromosome replication. This stability is demonstrated by the ability of the eukaryotic cell to establish cell lines or clones that comprise a population of daughter cells containing the transforming DNA. A "clone" is a population of cells derived from a single cell or common ancestor by mitosis. A "cell line" is a clone of a primary cell that is capable of stable growth *in vitro* for many generations. Suitable methods of genetic modification (also referred to as "transformation") include e.g., viral or bacteriophage infection, transfection, conjugation, protoplast fusion, lipofection, electroporation, calcium phosphate precipitation, polyethyleneimine (PEI) -mediated transfection, DEAE-dextran mediated transfection, liposome-mediated transfection, particle gun technology, calcium phosphate precipitation, direct micro injection, nanoparticle-mediated nucleic acid delivery. The choice of method of genetic modification is generally dependent on the type of cell being transformed and the circumstances under which the transformation is taking place (e.g., *in vitro*, *ex vivo*, or *in vivo*).

[0042] By "cleavage" it is meant the breakage of the covalent backbone of a target nucleic acid molecule. Cleavage can be initiated by a variety of methods

including, but not limited to, enzymatic or chemical hydrolysis of a phosphodiester bond. Both single-stranded cleavage and double-stranded cleavage are possible, and double-stranded cleavage can occur as a result of two distinct single-stranded cleavage events. DNA cleavage can result in the production of either blunt ends or
5 staggered ends. In certain embodiments, a complex comprising a thgRNA and a Cas9 protein is used for targeted double-stranded DNA cleavage.

[0043] "Nuclease" and "endonuclease" are used interchangeably herein to mean an enzyme which possesses catalytic activity for DNA cleavage.

[0044] By "cleavage domain" or "active domain" or "nuclease domain" of a
10 nuclease such as a Cas9 protein, it is meant the polypeptide sequence or domain within the nuclease which possesses the catalytic activity for DNA cleavage. A cleavage domain can be contained in a single polypeptide chain or cleavage activity can result from the association of two (or more) polypeptides. A single nuclease domain may consist of more than one isolated stretch of amino acids within a given
15 polypeptide.

[0045] A "host cell," as used herein, denotes an *in vivo* or *in vitro* eukaryotic cell, a prokaryotic cell (e.g., bacterial or archaeal cell), or a cell from a multicellular organism (e.g., a cell line) cultured as a unicellular entity, which eukaryotic or prokaryotic cells can be, or have been, used as recipients for a nucleic acid, and
20 include the progeny of the original cell which has been transformed by the nucleic acid. It is understood that the progeny of a single cell may not necessarily be completely identical in morphology or in genomic or total DNA complement as the original parent, due to natural, accidental, or deliberate mutation. A "recombinant host cell" (also referred to as a "genetically modified host cell") is a host cell into which
25 has been introduced a heterologous nucleic acid, e.g., an expression vector. For example, a subject bacterial host cell is a genetically modified bacterial host cell by virtue of introduction into a suitable bacterial host cell of an exogenous nucleic acid (e.g., a plasmid or recombinant expression vector) and a subject eukaryotic host cell is a genetically modified eukaryotic host cell (e.g., a mammalian germ cell), by virtue of
30 introduction into a suitable eukaryotic host cell of an exogenous nucleic acid.

[0046] An "activity", a "biological activity", or a "functional activity" of a polypeptide or nucleic acid refers to an activity exerted by a polypeptide or nucleic acid molecule as determined *in vivo* or *in vitro*, according to standard techniques.

Such activities can be a direct activity, such as an association with or an enzymatic activity on a second protein, or an indirect activity, such as a cellular signaling activity mediated by interaction of a protein with a second protein.

- [0047]** “Biological system” is meant, material, in a purified or unpurified form, from biological sources, including but not limited to human, animal, plant, insect, microbial, viral or other sources, wherein the system comprises the components required for biologic activity (e.g., inhibition of gene expression). The term “biological system” includes, for example, a cell, a virus, a microbe, an organism, an animal, or a plant.
- 10 **[0048]** A “subject” as used herein, refers to an organism to which the nucleic acid molecules of the invention can be administered. A subject can be an animal or a plant, preferably a mammal, most preferably a human, who has been the object of treatment, observation or experiment, or any cell thereof.
- [0049]** As used herein, a “promoter” refers to a control region of a nucleic acid sequence at which initiation and rate of transcription of the remainder of a nucleic acid sequence are controlled. A promoter may also contain sub-regions at which regulatory proteins and molecules may bind, such as RNA polymerase and other transcription factors. Promoters may be constitutive, inducible, activatable, repressible, tissue-specific or any combination thereof.
- 15 **[0050]** A promoter drives expression or drives transcription of the nucleic acid sequence that it regulates. Herein, a promoter is considered to be “operably linked” when it is in a correct functional location and orientation in relation to a nucleic acid sequence it regulates to control (“drive”) transcriptional initiation and/or expression of that sequence.
- 20 **[0051]** A promoter may be classified as strong or weak according to its affinity for RNA polymerase (and/or sigma factor); this is related to how closely the promoter sequence resembles the ideal consensus sequence for the polymerase. The strength of a promoter may depend on whether initiation of transcription occurs at that promoter with high or low frequency. Different promoters with different strengths
- 25 may be used to construct biomolecular circuits with different levels of gene/protein expression (e.g., the level of expression initiated from a weak promoter is lower than the level of expression initiated from a strong promoter).
- 30

[0052] A promoter may be one naturally associated with a gene or sequence, as may be obtained by isolating the 5' non-coding sequences located upstream of the coding segment of a given gene or sequence. Such a promoter can be referred to as "endogenous." Similarly, an activator/enhancer may be one naturally associated with a nucleic acid sequence, located either within or downstream or upstream of that sequence.

[0053] In some embodiments, a coding nucleic acid segment may be positioned under the control of a recombinant or heterologous promoter, which refers to a promoter that is not normally associated with the encoded nucleic acid sequence in its natural environment. A recombinant or heterologous enhancer refers to an enhancer not normally associated with a nucleic acid sequence in its natural environment. Such promoters or enhancers may include promoters or enhancers of other genes; promoters or enhancers isolated from any other prokaryotic cell; and synthetic promoters or enhancers that are not "naturally occurring" such as, for example, those that contain different elements of different transcriptional regulatory regions and/or mutations that alter expression through methods of genetic engineering that are known in the art. In addition to producing nucleic acid sequences of promoters and enhancers synthetically, sequences may be produced using recombinant cloning and/or nucleic acid amplification technology, including polymerase chain reaction (PCR) (see U.S. Pat. No. 4,683,202 and U.S. Pat. No. 5,928,906).

[0054] Cells according to the present disclosure include any cell into which foreign nucleic acids can be introduced and expressed as described herein. It is to be understood that the basic concepts of the present disclosure described herein are not limited by cell type. Cells according to the present disclosure include eukaryotic cells, prokaryotic cells, animal cells, plant cells, fungal cells, archael cells, eubacterial cells and the like. Cells include eukaryotic cells such as yeast cells, plant cells, and animal cells. Particular cells include mammalian cells, for example, human cells. Further, cells may include those which overexpress or are deficient in expression of a particular protein or which have a different expression pattern leading to a disease or detrimental condition. Such diseases or detrimental conditions are readily known to those of skill in the art.

[0055] Target nucleic acids include any nucleic acid sequence to which a guide sequence is designed to have complementarity and to which a co-localization complex such as a Cas9 protein as described herein can be useful to either regulate or nick. Target nucleic acids include genes. For purposes of the present disclosure, DNA, such as double stranded DNA, can include the target nucleic acid and a co-localization complex can bind to or otherwise co-localize with the DNA at or adjacent or near the target nucleic acid and in a manner in which the co-localization complex may have a desired effect on the target nucleic acid. Such target nucleic acids can include endogenous (or naturally occurring) nucleic acids and exogenous (or foreign) nucleic acids. One of skill based on the present disclosure will readily be able to identify or design gRNA and Cas9 proteins which co-localize to a DNA including a target nucleic acid. One of skill will further be able to identify transcriptional regulator proteins or domains which likewise co-localize to a DNA including a target nucleic acid. DNA includes genomic DNA, mitochondrial DNA, viral DNA or exogenous DNA.

[0056] Triggering RNA (or “fuelRNA”) refers to any RNA of interest. The triggering RNA can be any naturally or non-naturally occurring RNA. The triggering RNA can include naturally occurring nucleosides, modified nucleosides, or mixtures thereof. As used herein, the term “modified nucleoside” is a nucleoside that includes a modified heterocyclic base, a modified sugar moiety, or a combination thereof. The nucleoside residues can be coupled to each other by any of the numerous known internucleoside linkages. Such internucleoside linkages include, without limitation, phosphodiester, phosphorothioate, phosphorodithioate, alkylphosphonate, alkylphosphonothioate, phosphotriester, phosphoramidate, siloxane, carbonate, carboalkoxy, acetamidate, carbamate, morpholino, borano, thioether, bridged phosphoramidate, bridged methylene phosphonate, bridged phosphorothioate, and sulfone internucleoside linkages. Examples of triggering RNA include, but are not limited to, an RNA that is translated into protein (mRNA), or an RNA that is not translated into protein (e.g. microRNA, tRNA, or rRNA; also called “non-coding” RNA or “ncRNA”).

[0057] As used herein, the term “proximal” refers to next to or nearest the first nucleobase of the toehold domain. In some embodiments, the hybridization of the at least one first stem domain and the at least one second stem domain is from 0 to 30

nucleobases from the toehold domain, preferably from 0 to 20 nucleobases, or from 0 to 10 nucleobases, or from 0 to 5 nucleobases from the toehold domain.

[0058] The practice of the present invention can employ, unless otherwise indicated, conventional techniques of immunology, biochemistry, chemistry, molecular biology, microbiology, cell biology, genomics and recombinant DNA, which are within the skill of the art. See Sambrook, Fritsch and Maniatis, MOLECULAR CLONING: A LABORATORY MANUAL, 2nd edition (1989); CURRENT PROTOCOLS IN MOLECULAR BIOLOGY (F. M. Ausubel, et al. eds., (1987)); the series METHODS IN ENZYMOLOGY (Academic Press, Inc.): PCR 2: A PRACTICAL APPROACH (M. J. MacPherson, B. D. Hames and G. R. Taylor eds. (1995)), Harlow and Lane, eds. (1988) ANTIBODIES, A LABORATORY MANUAL, and ANIMAL CELL CULTURE (R. I. Freshney, ed. (1987)).

[0059] In one aspect, the invention provides a toehold-gated guide RNA (thgRNA) comprising a scaffold domain, a spacer domain, at least one first stem domain and at least one second stem domain, wherein the at least one first stem domain is complementary to a triggering RNA, wherein the at least one second stem domain is complementary to a portion of the at least one first stem domain, and wherein at least a portion of the first stem domain remains unpaired thereby forming a toehold domain.

[0060] In some embodiments, the at least one first stem domain and at least one second stem domain are located at opposing ends of the thgRNA. In one embodiment, the at least one first stem domain is at the 5' end of the thgRNA and the at least one second stem domain is at the 3' end of the thgRNA. In another embodiment, the at least one second stem domain is at the 5' end of the thgRNA and the at least one first stem domain is at the 3' end of the thgRNA.

[0061] The complementary regions of the at least one first and the at least one second stem domain hybridize to form a stem in the secondary structure of the thgRNA. This stem should be sufficiently long as to constrain the distance between the spacer and scaffold domains. This constraint results in the spacer domain and/or the stem domain sterically hindering the association of the Cas protein to the scaffold domain until the toehold domain binds to the triggering RNA. In a preferred embodiment, the spacer domain hinders association of the Cas protein with the scaffold domain.

[0062] In one embodiment, the hybridization of the at least one first stem domain and the at least one second stem domain is proximal to the toehold domain. The binding of the triggering RNA to the toehold domain initiates an entropy driven reaction of toehold-mediated strand displacement which disrupts the complementarity of the stem domain. Complete removal of the hybridization of the at least one first and at least one second stem domains eliminates the steric hindrance between the spacer domain and/or the stem domain with the scaffold domain thereby allowing the thgRNA to complex with the Cas protein.

[0063] In one embodiment, the thgRNA can comprise at least two toehold domains. In one embodiment, the thgRNA comprises a plurality of first stem domains and a plurality of second stem domains which can be positioned serially or in parallel. The portions of the plurality first and second stem domains that are complementary can have the same complementary sequence or have different complementary sequences. The toehold domains of the plurality of first stem domains can be complementary to the same region of a triggering RNA, to different regions of a triggering RNA, or to different triggering RNAs.

[0064] When multiple first and second stem domains, and as such multiple toeholds, are positioned serially the terminal at the 3' or 5' terminal end is considered the "first" toehold. Binding of the first toehold to its triggering RNA (e.g., triggering RNA 1) displaces the stem domain to expose the next, second toehold domain. Binding of the second toehold domain to its triggering RNA (e.g., triggering RNA 2) results in displacement of the hybridization in the following stem domain. Therefore, triggering RNA2 can only perform strand displacement after triggering RNA1 does so.

[0065] In a preferred embodiment, the second toehold domain, and subsequent toehold domains, comprise an intervening segment that is not involved in the hybridization event with the triggering RNA so that one portion of the second, and subsequent, toehold(s) is weak relative to the other portion of the second, and subsequent, toehold(s). In this manner, only the dominant toehold segment of the second, and subsequent, toehold domains can initially bind to its triggering RNA. Binding of the weaker portion of the second, and subsequent, toehold domains is necessary to initiate strand displacement. Removal of all hybridization in the stem domains is only complete when a complex is formed between all the toehold domains

and their corresponding triggering RNAs. Hence the affinity of the toehold-gated guide RNA toehold-gated guide RNA to Cas9 effectively performs the AND logic operation on the presence of the two or more triggering RNAs.

[0066] In another embodiment, the at least one first stem domain and/or the at least one second stem domain is inserted within the scaffold domain. When inserting the at least one first and/or the at least one second stem domains within the scaffold the stem domains are positioned so that they does not interfere with the scaffold's ability to associate with a Cas protein when the thgRNA binds to the triggering RNA. In some embodiments, both the at least one first stem domain and the at least one second stem domain is inserted within the scaffold domain. In another embodiment, the at least one first stem domain is inserted within the scaffold domain whereas the at least one second stem domain is at the 5' or 3' end of the thgRNA. In another embodiment, the at least one second stem domain is inserted within the scaffold domain whereas the at least one first stem domain is at the 5' or 3' end of the thgRNA.

[0067] In a further embodiment, the spacer domain and the at least one first stem domain can form a single domain referred to as "spacerstem" domain. In this embodiment, the spacer domain is complementary to the triggering RNA. The second stem domain is complementary to at least a portion of the spacerstem domain. The spacerstem and the second stem domain hybridize to constrain the distance between the spacer and scaffold domains until the toehold domain binds to the triggering RNA. Alternatively, the spacer domain and the second stem domain can form a single, spacerstem domain. In this embodiment, the spacerstem domain is complementary to at least a portion of the first stem domain. The spacerstem and the first stem domain hybridize to constrain the distance between the spacer and scaffold domains until the toehold domain binds to the triggering RNA.

[0068] In a further aspect, the invention provides a nucleic acid encoding the thgRNA. In one embodiment, the nucleic acid encoding the thgRNA is inserted into an expression vector, for example, a recombinant expression vector. In another embodiment, the vector comprising the nucleic acid encoding the thgRNA is transfected into a cell. In some embodiments, the cell is a prokaryotic cell. In some embodiments, the cell is a eukaryotic cell such as, but not limited to, a human cell. In

a preferred embodiment, the thgRNA of the invention is expressed in the cell from the vector comprising the nucleic acid encoding the thgRNA.

[0069] In a further aspect, the invention provides a cell comprising the thgRNA. In some embodiments, the cell is a prokaryotic cell. In some embodiments, the cell is a eukaryotic cell such as, but not limited to, a human cell.

[0070] In some embodiments, the thgRNA can be used to detect the presence or absence of one or more triggering RNAs. In some embodiments, the thgRNA of the invention is useful in methods for, but not limited to, (i) activating the expression of a nucleic acid encoding a protein or a second, or subsequent, thgRNA; (ii) cutting or nicking a target nucleic acid; or (iii) inhibiting the expression of nucleic acid.

[0071] In a further aspect, the invention provides a biomolecular circuit comprising at least one thgRNA of the invention as described above and at least one Cas protein. In some embodiments, the biomolecular circuit is capable of detecting of the presence or absence of one or more triggering RNAs. In some embodiments, the biomolecular circuit is also useful in methods for, but not limited to, (i) activating the expression of a nucleic acid encoding a protein or a second, or subsequent, thgRNA; (ii) cutting or nicking a target nucleic acid; or (iii) inhibiting the expression of nucleic acid.

[0072] As used herein, a Cas protein refers to any naturally or non-naturally occurring RNA guided DNA binding protein. According to some embodiments, the Cas protein may have nuclease activity. Such Cas proteins are known to those skilled in the art and include, for example, Cas9 proteins. According to some embodiments, the Cas protein may be a nuclease-null DNA binding protein. The nuclease-null Cas protein may result from the alteration or modification of a DNA binding protein having nuclease activity.

[0073] Non-limiting examples of Cas proteins include Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas9 (also known as Csn1 and Csx12), Cas10, Csy1, Csy2, Csy3, Cse1, Cse2, Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, Csf4, homologs thereof, or modified versions thereof. In a preferred embodiment the Cas protein is Cas9, and may be Cas9 from *S. pyogenes* or *S. pneumoniae*.

[0074] Typically, in the context of an endogenous CRISPR system, formation of a CRISPR complex results in cleavage (e.g., a cutting or nicking) of one or both strands in or near (e.g. within 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 50, or more base pairs from) the target nucleic acid.

- 5 **[0075]** In some embodiments, the Cas protein directs cleavage of one or both strands at the location of a target nucleic acid, such as within the target nucleic acid and/or within the complement of the target nucleic acid.

- [0076]** In some embodiments, the Cas protein is mutated with respect to a corresponding wild-type enzyme such that the mutated Cas protein lacks the ability to
10 cleave one or both strands of a target nucleic acid. For example, an aspartate-to-alanine substitution (D10A) in the RuvC I catalytic domain of Cas9 from *S. pyogenes* converts Cas9 from a nuclease that cleaves both strands to a nickase (cleaves a single strand). Other examples of mutations that render Cas9a nickase include, without limitation, H840A, N854A, and N863A.

- 15 **[0077]** In some embodiments, the Cas protein is altered or otherwise modified to inactivate the nuclease activity. Modifications to inactivate the nuclease activity are known to one skilled in the art. The nuclease-null Cas protein retains the ability to bind to DNA even though the nuclease activity has been inactivated.

- [0078]** In some embodiments, the Cas protein is a Cas-regulator element
20 fusion protein, such as a Cas9-regulator element fusion protein. The Cas-regulator element fusion proteins are created by fusing a heterologous functional domain (e.g., a transcriptional activation domain, e.g., from VP64 or NF- κ B p65), to the N-terminus or C-terminus of a catalytically inactive Cas protein such as, but not limited to, an inactive Cas9 protein (dCas9). The dCas9 can be from any species including but not
25 limited to *S. pyogenes*. In some embodiments, the Cas9 contains mutations to render the nuclease portion of the protein catalytically inactive. The transcriptional activation domains can be fused on the N or C terminus of the Cas9. In addition, to transcriptional activation domains, other heterologous functional domains (e.g., transcriptional repressors (e.g., KRAB, ERD, SID, and others, e.g., amino acids 473-
30 530 of the ets2 repressor factor (ERF) repressor domain (ERD), amino acids 1-97 of the KRAB domain of KOX1, or amino acids 1-36 of the Mad mSIN3 interaction domain (SID); see Beerli et al., PNAS USA 95:14628-14633 (1998)) or silencers such as Heterochromatin Protein 1 (HP1, also known as swi6), e.g., HP1 α or HP1 β ;

proteins or peptides that could recruit long non-coding RNAs (lncRNAs) fused to a fixed RNA binding sequence such as those bound by the MS2 coat protein, endoribonuclease Csy4, or the lambda N protein; enzymes that modify the methylation state of DNA (e.g., DNA methyltransferase (DNMT) or TET proteins); or

5 enzymes that modify histone subunits (e.g., histone acetyltransferases (HAT), histone deacetylases (HDAC), histone methyltransferases (e.g., for methylation of lysine or arginine residues) or histone demethylases (e.g., for demethylation of lysine or arginine residues)) can also be used. A number of sequences for such domains are known in the art. In some embodiments, all or part of the full-length sequence of the

10 catalytic domain can be included. In some embodiments, the heterologous functional domain is a biological tether, and comprises all or part of (e.g., DNA binding domain from) the MS2 coat protein, endoribonuclease Csy4, or the lambda N protein. These proteins can be used to recruit RNA molecules containing a specific stem-loop structure to a locale specified by the dCas9 thgRNA targeting sequences. For

15 example, a dCas9 fused to MS2 coat protein, endoribonuclease Csy4, or lambda N can be used to recruit a long non-coding RNA (lncRNA) such as XIST or HOTAIR; see, e.g., Keryer-Bibens et al., Biol. Cell 100:125-138 (2008), that is linked to the Csy4, MS2 or lambda N binding sequence. Alternatively, the Csy4, MS2 or lambda N protein binding sequence can be linked to another protein, e.g., as described in

20 Keryer-Bibens et al., supra, and the protein can be targeted to the dCas9 binding site using the methods and compositions described herein. In some embodiments, the Csy4 is catalytically inactive.

[0079] In some embodiments, the fusion proteins include a linker between the dCas9 and the heterologous functional domains. Linkers that can be used in these

25 fusion proteins (or between fusion proteins in a concatenated structure) can include any sequence that does not interfere with the function of the fusion proteins. In preferred embodiments, the linkers are short, e.g., 2-20 amino acids, and are typically flexible (i.e., comprising amino acids with a high degree of freedom such as glycine, alanine, and serine).

30 **[0080]** In a further aspect, the invention provides a cell based biomolecular circuit comprising expressing in the cell, or introducing into the cell, at least one thgRNA of the invention as described above and at least one Cas protein. The biomolecular circuit is responsive to at least one triggering RNA originating from

outside or inside of the cell. Upon binding of the triggering RNA to the toehold, and the subsequent removal of the constraint on the thgRNA ability to complex with a Cas protein, the thgRNA binds to a target nucleic acid. Once bound to the target nucleic acid, the thgRNA can complex with the Cas protein. In one embodiment, the Cas protein cuts or nicks the target nucleic acid. The cut or nick in the target nucleic acid is then repaired wherein the repair introduces at least one error (also referred to herein as “indels”) such as, but not limited to, an insertion, deletion, or substitution of one or more nucleotides in the target nucleic acid. The presence of the at least one indel in the target nucleic acid provides a data record.

10 **[0081]** The cell based biomolecular circuit of the invention may be used in methods for detecting changes in transcriptome state of the cell. This process also is referred to herein as sensing and recording data from the cellular environment in a cell. The cell based biomolecular circuit of the invention may be used in methods for determining the presence of and/or the level of at least one triggering RNA of interest
15 originating outside or inside of a cell. In some embodiments, the methods comprise providing a recombinant cell-based biomolecular circuit to a subject, harvesting the cell or the cell’s progeny; isolating the DNA from the cell or the cell’s progeny; sequencing the target site(s) within the target nucleic acid(s); and comparing the sequenced target site with a reference sequence, wherein the introduction of an indel
20 into the target nucleic acid of the cell or its progeny is indicative of the presence of the triggering RNA(s).

[0082] In some embodiments, the cell based biomolecular circuit of the invention may be used in methods for monitoring the activation of all known signaling pathways. Additionally, since most of the cellular phenotypic changes are
25 communicated through transcriptional changes, the use of transcription as the counter input will facilitate detection of subtle cellular reactions to different treatments or experimental conditions. In addition to recording levels of transcriptional activation, the proposed synthetic biological circuit can provide a compact barcode for determining cell lineage of proliferating cells and serve as a platform for more
30 complex computational operations in the mammalian cell.

[0083] The repair of the cut or nicked target nucleic acid is done by, for example, native end joining mechanisms. Native end joining mechanisms include, but are not limited to, Non-Homologous End Joining (NHEJ) or Alternative End

Joining (AEJ). These repair mechanisms have a finite probability of introducing an error and modify a recognition sequence in a way that prevents future localization of the directed endonuclease. As such, these error(s) represent a record of the data and can serve as a counting currency the readout of which will be sequence based.

- 5 **[0084]** Any directed endonuclease known to those skilled in the art including, but are not limited to, Zinc Finger Nucleases (ZFNs), Transcription Activator Like Effector Nucleases (TALENs), and the Clustered Regularly Interspaced Palindromic Repeats (CRISPR) system. In general, "CRISPR system" refers collectively to transcripts and other elements involved in the expression of or directing the activity of
- 10 CRISPR-associated ("Cas") genes. Directed endonucleases can be designed to target specific recognition sites. This programmable specificity allows the library of available recognition sequences to grow exponentially with the length and/or number of the recognition sequence(s). As discussed further below, scalable memory follows from the large number of different triggering RNAs and/or nucleic acids that can be
- 15 targeted.

- [0085]** In order to use the toehold gated-guide RNAs described, it may be desirable to express them from a nucleic acid that encodes them. This can be performed in a variety of ways. For example, the nucleic acid encoding the guide RNA can be cloned into an intermediate vector for transformation into prokaryotic or
- 20 eukaryotic cells for replication and/or expression. Intermediate vectors are typically prokaryote vectors, e.g., plasmids, or shuttle vectors, or insect vectors, for storage or manipulation of the nucleic acid encoding the guide RNA for production of the guide RNA. The nucleic acid encoding the guide RNA can also be cloned into an expression vector, for administration to a plant cell, animal cell, preferably a mammalian cell or a
- 25 human cell, fungal cell, bacterial cell, or protozoan cell.

- [0086]** To obtain expression, a sequence encoding a guide RNA is typically subcloned into an expression vector that contains a promoter to direct transcription. Suitable bacterial and eukaryotic promoters are well known in the art and described, e.g., in Sambrook et al., *Molecular Cloning, A Laboratory Manual* (3d ed. 2001);
- 30 Kriegler, *Gene Transfer and Expression: A Laboratory Manual* (1990); and *Current Protocols in Molecular Biology* (Ausubel et al., eds., 2010). Bacterial expression systems for expressing the engineered protein are available in, e.g., *E. coli*, *Bacillus* sp., and *Salmonella* (Palva et al., 1983, *Gene* 22:229-235). Kits for such expression

systems are commercially available. Eukaryotic expression systems for mammalian cells, yeast, and insect cells are well known in the art and are also commercially available.

[0087] The promoter used to direct expression of a nucleic acid depends on the particular application. For example, a strong constitutive promoter is typically used for expression and purification of fusion proteins. In contrast, when the guide RNA is to be administered *in vivo* for gene regulation, either a constitutive or an inducible promoter can be used, depending on the particular use of the guide RNA. In addition, a preferred promoter for administration of the guide RNA can be a weak promoter, such as HSV TK or a promoter having similar activity. The promoter can also include elements that are responsive to transactivation, e.g., hypoxia response elements, Gal4 response elements, lac repressor response element, and small molecule control systems such as tetracycline-regulated systems and the RU-486 system (see, e.g., Gossen & Bujard, 1992, Proc. Natl. Acad. Sci. USA, 89:5547; Oligino et al., 1998, Gene Ther., 5:491-496; Wang et al., 1997, Gene Ther., 4:432-441; Neering et al., 1996, Blood, 88:1147-55; and Rendahl et al., 1998, Nat. Biotechnol., 16:757-761).

[0088] In addition to the promoter, the expression vector typically contains a transcription unit or expression cassette that contains all the additional elements required for the expression of the nucleic acid in host cells, either prokaryotic or eukaryotic. A typical expression cassette thus contains a promoter operably linked, e.g., to the nucleic acid sequence encoding the gRNA, and any signals required, e.g., for efficient polyadenylation of the transcript, transcriptional termination, ribosome binding sites, or translation termination. Additional elements of the cassette may include, e.g., enhancers, and heterologous spliced intronic signals.

[0089] The particular expression vector used to transport the genetic information into the cell is selected with regard to the intended use of the gRNA, e.g., expression in plants, animals, bacteria, fungus, protozoa, etc. Standard bacterial expression vectors include plasmids such as pBR322 based plasmids, pSKF, pET23D, and commercially available tag-fusion expression systems such as GST and LacZ.

[0090] Expression vectors containing regulatory elements from eukaryotic viruses are often used in eukaryotic expression vectors, e.g., SV40 vectors, papilloma virus vectors, and vectors derived from Epstein-Barr virus. Other exemplary eukaryotic vectors include pMSG, pAV009/A+, pMTO10/A+, pMAMneo-5,

baculovirus pDSVE, and any other vector allowing expression of proteins under the direction of the SV40 early promoter, SV40 late promoter, metallothionein promoter, murine mammary tumor virus promoter, Rous sarcoma virus promoter, polyhedrin promoter, or other promoters shown effective for expression in eukaryotic cells.

- 5 **[0091]** The vectors for expressing the guide RNAs can include RNA Pol III promoters to drive expression of the guide RNAs, e.g., the H1, U6 or 7SK promoters. These human promoters allow for expression of gRNAs in mammalian cells following plasmid transfection. Alternatively, a T7 promoter may be used, e.g., for in vitro transcription, and the RNA can be transcribed in vitro and purified. Vectors
10 suitable for the expression of short RNAs, e.g., siRNAs, shRNAs, or other small RNAs, can be used.

- [0092]** Some expression systems have markers for selection of stably transfected cell lines such as thymidine kinase, hygromycin B phosphotransferase, and dihydrofolate reductase. High yield expression systems are also suitable, such as
15 using a baculovirus vector in insect cells, with the gRNA encoding sequence under the direction of the polyhedrin promoter or other strong baculovirus promoters.

- [0093]** The elements that are typically included in expression vectors also include a replicon that functions in *E. coli*, a gene encoding antibiotic resistance to permit selection of bacteria that harbor recombinant plasmids, and unique restriction
20 sites in nonessential regions of the plasmid to allow insertion of recombinant sequences.

- [0094]** Standard transfection methods are used to produce bacterial, mammalian, yeast or insect cell lines that express large quantities of protein, which are then purified using standard techniques (see, e.g., Colley et al., 1989, *J. Biol.*
25 *Chem.*, 264:17619-22; Guide to Protein Purification, in *Methods in Enzymology*, vol. 182 (Deutscher, ed., 1990)). Transformation of eukaryotic and prokaryotic cells are performed according to standard techniques (see, e.g., Morrison, 1977, *J. Bacteriol.* 132:349-351; Clark-Curtiss & Curtiss, *Methods in Enzymology* 101:347-362 (Wu et al., eds, 1983)).

- 30 **[0095]** Any of the known procedures for introducing foreign nucleotide sequences into host cells may be used. These include the use of calcium phosphate transfection, polybrene, protoplast fusion, electroporation, nucleofection, liposomes, microinjection, naked DNA, plasmid vectors, viral vectors, both episomal and

integrative, and any of the other well-known methods for introducing cloned genomic DNA, cDNA, synthetic DNA or other foreign genetic material into a host cell (see, e.g., Sambrook et al., supra). It is only necessary that the particular genetic engineering procedure used be capable of successfully introducing at least one gene into the host cell capable of expressing.

[0096] Turning to the figures, domains are identified around the structure with numbers where 1 refers to the scaffold domain, 2 refers to the spacer domain, 3 refers to the at least one first stem domain, 4 refers to the toehold domain, 5 refers to the at least one second stem domain, 6 refers to the triggering RNA, and 7 refers to the spacerstem domain.

[0097] Fig. 1 depicts transition and Cas9 affinity of toehold-gated gRNA. In Fig. 1, above the hollow arrow illustrates a toehold-gated guide RNA beside Cas9, implying low affinity between the two. Following common representation of secondary structure for nucleic acid strands, an arrow represents the 3' end of the molecule and short connecting lines represent basepairing of complementary sequence. The scaffold domain is a fixed sequence that confers Cas9 affinity to chimeric gRNA and the spacer sequence guides Cas9 to a complementary region of a target nucleic acid. Addition of a stem reduces the distance between spacer and scaffold to reduce Cas9 affinity. Sequence for the toehold domain is selected to hybridize to RNA of interest, labeled "triggering RNA". As the toehold domain is on the 3' end, the toehold-gated guide RNA is referred to as 3'toehold-gated guide RNA. Sequence for the stem is selected such that hybridization is removed by adjacent complementary sequence in the triggering RNA through the established entropy driven reaction of toehold-mediated strand displacement. The completion of this reaction is illustrated beneath the hollow arrow, where Cas9 and 3'toehold-gated guide RNA are overlaid, implying greater affinity between the two as a result of removal of hybridization in the stem. Additional RNA processing domains such as ribozymes and endonuclease target sequences may be added between the stem domains and the scaffold or spacer domain to further increase Cas9 affinity [Lior Nissim, Samuel D Perli, Alexandra Fridkin, et al., "bioRxiv preprint server", [http:// biorxiv.org / content / early /2014/04/23/ 004432](http://biorxiv.org/content/early/2014/04/23/004432), 2014].

[0098] Fig. 2 is an illustration of secondary structure for 3'toehold-gated guide RNA and triggering RNA and secondary structure for the complex between 3'toehold-

gated guide RNA and triggering RNA. Removal of hybridization in the stem domain increases affinity to Cas9.

[0099] Fig. 3 shows that the toehold can be added to either the 3' or 5' end of toehold-gated guide RNA. The result of the latter is called 5'toehold-gated guide RNA. Fig. 3 is an illustration of secondary structure for 5'toehold-gated guide RNA and the complex between 5'toehold-gated guide RNA and triggering RNA. Removal of self-hybridization in the stem domains increases affinity to Cas9.

[00100] Fig. 4 depicts a spacer comprising sequence complementary to the triggering RNA sequence. In this embodiment, the spacer domain can be used to form the stem, which should further inhibit affinity to Cas9. This spacer domain combined with the at least one first, or the at least one second, stem domain is labeled "SpacerStem". Fig. 4 is an illustration of secondary structure for 3'toehold-gated guide RNA and the complex between 3'toehold-gated guide RNA and triggering RNA for a 3'toehold-gated guide RNA with a SpacerStem domain. Removal of self-hybridization in the SpacerStem domain increases affinity to Cas9.

[00101] Fig. 5 depicts layered toehold-gated gRNA. A complex between triggering RNA and toehold-gated guide RNA from a first layer can bind with Cas9 that has deactivated nuclease domains and fused to an activator protein domain. As shown, this RNA-protein complex localizes to the region upstream of the weak promoter controlling expression of the next layer of toehold-gated guide RNAs. Thus, fused activator on Cas9 induces transcription of the next toehold-gated guide RNAs, which can be multiplexed with RNA processing elements such as ribozymes and endonuclease target sequences.

[00102] Fig. 6 depicts the initial and intermediate states of Boolean logic within toehold-gated gRNA. Fig. 6 is an illustration of secondary structure for AND toehold-gated guide RNA. The AND thgRNA includes a plurality of first and second stem domains each contiguous or proximal to a toehold. The second, and subsequent, toehold(s) is designed to be interrupted by a stem domain so that one portion of the second, and subsequent, toehold(s) is weak relative to the other portion of the second, and subsequent, toehold(s) so only the latter can initiate strand displacement. However, the dominant portion of the second, and subsequent, toehold(s) is flanked by sequence containing hybridization until trigger RNA1 forms a complex with the first toehold domain (603). Therefore, trigger RNA2 can only perform strand

displacement after trigger RNA1 does so. Removal of all self-hybridization in the stem domains is only complete when a complex is formed with both trigger RNA molecules, increasing affinity to Cas9. Hence the affinity of the toehold-gated guide RNA toehold-gated guide RNA to Cas9 effectively performs the AND logic operation on the presence of the two trigger RNA molecules.

[00103] Fig. 7 is an illustration of secondary structure for toehold-gated guide RNA wherein the at least one first and/or the at least one second stem domains are inserted within the scaffold domain.

10 Example 1: *In vitro* experiment to test activity of the RNA with Cas9 in the presence or absence of corresponding triggering RNA.

[00104] For a given thgRNA design, DNA to transcribe thgRNA is synthesized (e.g., with services such as IDT gBlock Fragments) operably linked to a T7 promoter. Similarly, DNA to transcribe the corresponding triggering RNA is synthesized and

15 PCR amplified. Commercial *in vitro* transcription kits, such as New England BioLabs' HiScribe T7 High Yield RNA Synthesis Kit, and RNA extraction kits, such as Qiagen's miRNeasy Mini Kit, are used to separately produce and purify the thgRNA and its triggering RNA. DNA synthesis and/or PCR amplification is used to prepare, for example, a roughly 1 to 2 kb nucleic acid as the target for thgRNA at a position more than 200 bases from halfway along the length of the nucleic acid. In commercial reaction mix with Cas9, such as New England BioLabs' solution S. pyogenes Cas9 Nuclease, 1 microgram said target nucleic acid is incubated for 2 hours with 30 nM thgRNA and Cas9 in the presence or absence of ~100 nM triggering RNA. Each reaction and 1 microgram of the target nucleic acid are run in separate wells on a 1% agarose gel with a DNA/RNA stain to separate the molecules. After one hour the gel is imaged and the activity of thgRNA with Cas9 is quantified by the intensity of the two bands resulting from Cas9 cleavage of the target nucleic acid relative to the band resulting from the full length target nucleic acid.

25 **[00105]** While a preferred embodiment is disclosed, many other implementations will occur to one of ordinary skill in the art and are all within the scope of the invention. Each of the various embodiments described above may be combined with other described embodiments in order to provide multiple features. Furthermore, while the foregoing describes a number of separate embodiments of the

apparatus and method of the present invention, what has been described herein is merely illustrative of the application of the principles of the present invention. Other arrangements, methods, modifications, and substitutions by one of ordinary skill in the art are therefore also considered to be within the scope of the present invention,
5 which is not to be limited except by the claims.

CLAIMS

What is claimed is:

1. A toe hold-gated guide RNA (thgRNA) comprising a scaffold domain, a spacer domain, at least one first stem domain and at least one second stem domain,
5 wherein the at least one first stem domain is complementary to a triggering RNA and wherein the at least one second stem domain is complementary to a portion of the at least one first stem domain, and wherein at least a portion of the at least one first stem domain is unpaired thereby forming a toe hold domain
- 10 2. The thgRNA of claim 1, wherein the at least one first stem domain is positioned at the 5'-end of the thgRNA and the at least one second stem domain is positioned at the 3'-end of the thgRNA.
3. The thgRNA of claim 1, wherein the at least one second stem domain is
15 positioned at the 5'-end of the thgRNA and the at least one first stem domain is positioned at the 3'-end of the thgRNA.
4. The thgRNA of claim 1, wherein the at least one first stem domain and/or the at least one second stem domain is inserted within the scaffold domain.
20
5. The thgRNA of claim 1, wherein the at least one first stem domain and at least one second stem domain hybridize at a position proximal to the toehold domain.
6. The thgRNA of claim 1, comprising at least two toehold domains.
25
7. The thgRNA of claim 1, comprising at least two first stem domains and at least two second stem domains.
8. The thgRNA of claim 1, wherein the spacer domain hinders association of Cas
30 protein with the scaffold domain.
9. A nucleic acid encoding the thgRNA of claim 1.
10. A vector comprising the nucleic acid of claim 9.

11. A cell comprising the thgRNA of claim 1.
12. The cell of claim 11, wherein the cell is a prokaryotic cell or eukaryotic cell.
- 5 13. The cell of claim 12, wherein the cell is a human cell.
14. A cell comprising the vector of claim 10.
- 10 15. The cell of claim 14, wherein the cell is a prokaryotic cell or eukaryotic cell.
16. The cell of claim 15, wherein the cell is a human cell.
17. A biomolecular circuit comprising at least one thgRNA of claim 1 and at least
15 one Cas protein.
18. The method according to claim 17, wherein the Cas protein is a Cas9 protein.
19. A method of detecting of the presence or absence of one or more triggering
20 RNAs in a cell comprising the steps of:
expressing in the cell, or introducing into the cell, a biomolecular circuit of
claim 17; and
detecting the presence or absence of the one or more triggering RNAs.
- 25 20. The method according to claim 19, wherein the Cas protein is a Cas9 protein.
21. A kit comprising:
(i) at least one thgRNA of claim 1 or a nucleic acid encoding the same;
(ii) at least one Cas protein or nucleic acid encoding the same;
30 (iii) a reagent for reconstitution and/or dilution; and
(iv) instructions for using the kit
22. The kit of claim 21, further comprising a reagent selected is a buffer for
introducing the elements of (i) and (ii) into a cell.

23. The kit according to claim 21, wherein the Cas protein is a Cas9 protein.

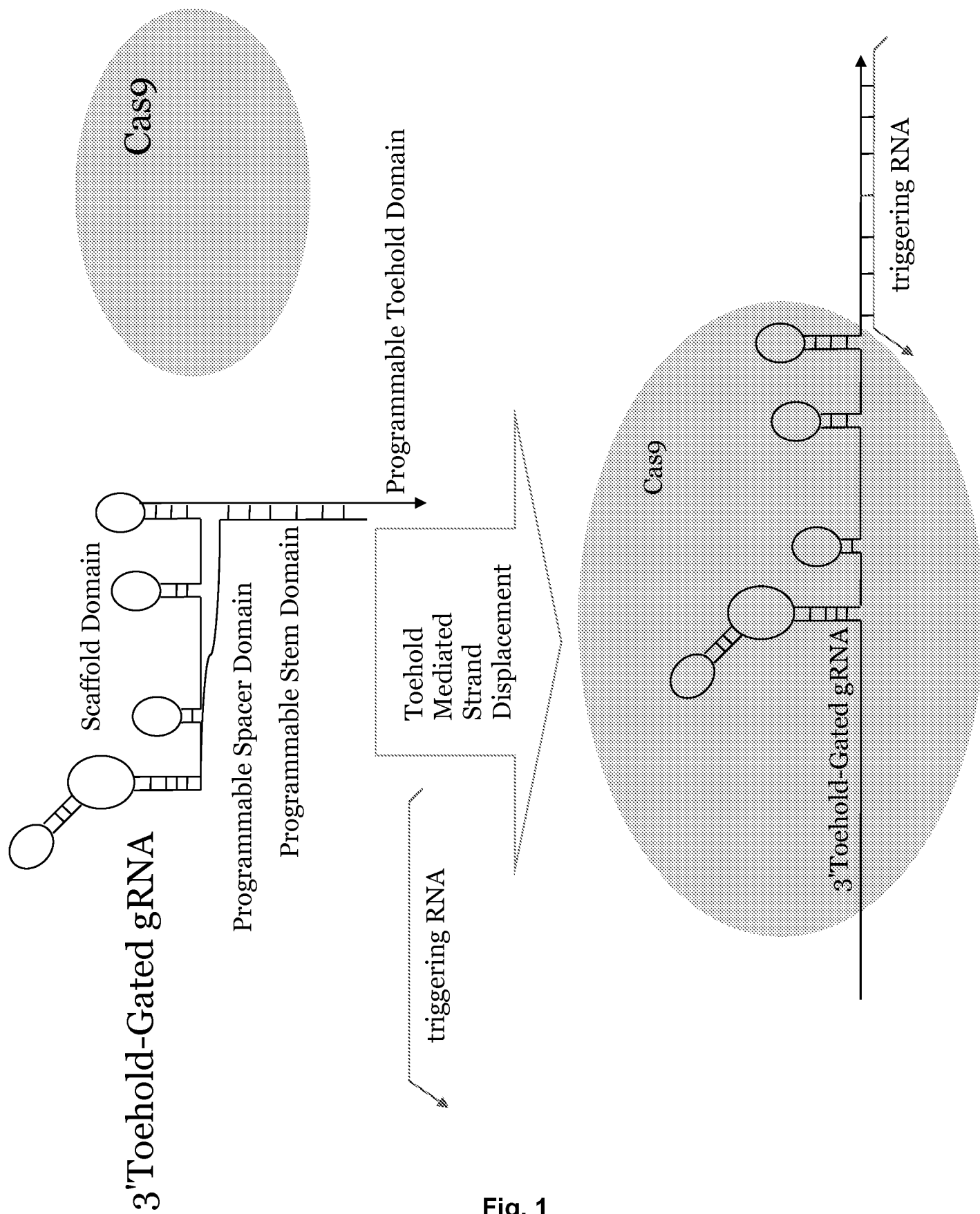
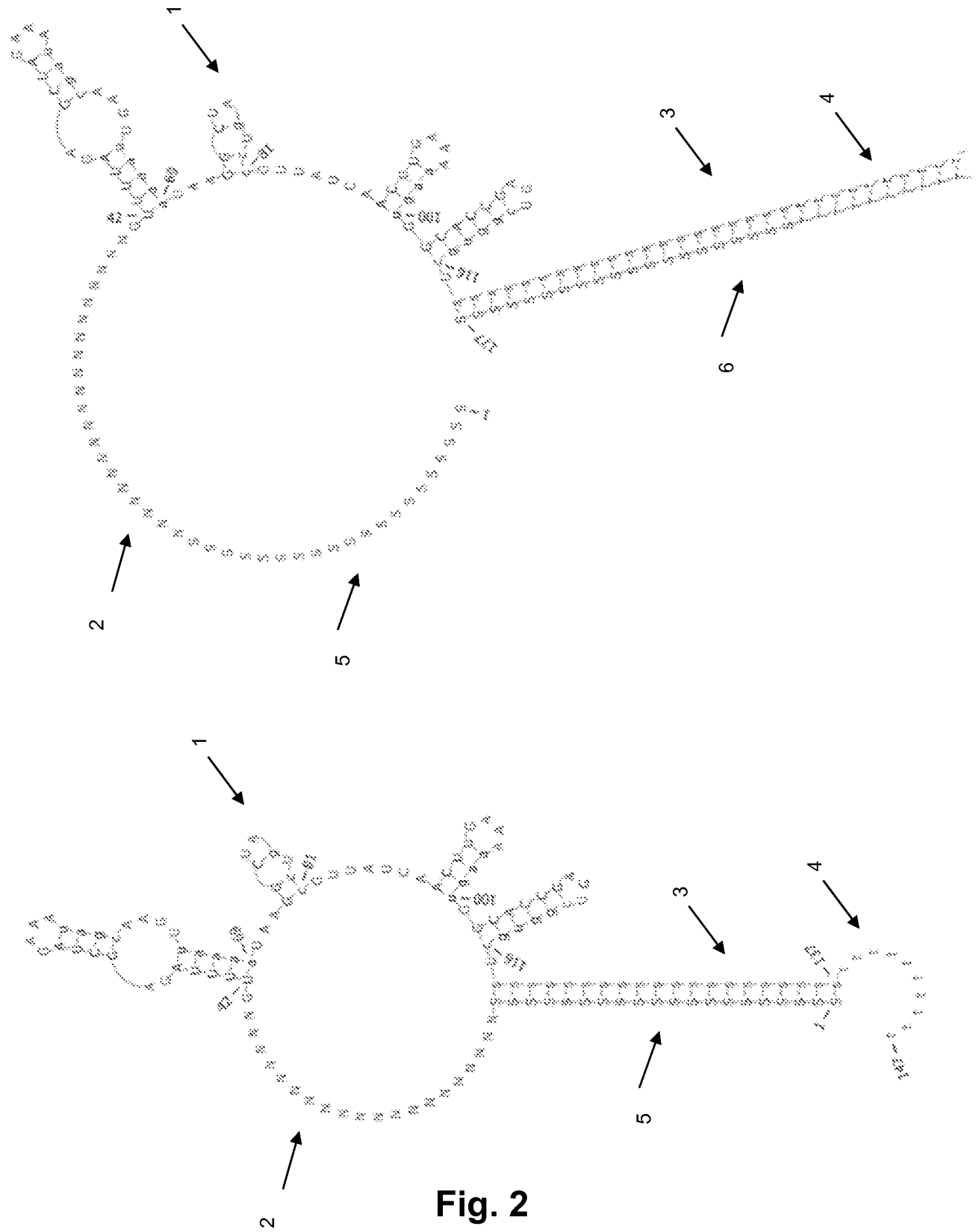


Fig. 1

3'Toehold-Gated gRNA

triggering RNA-3'Toehold-Gated gRNA Complex



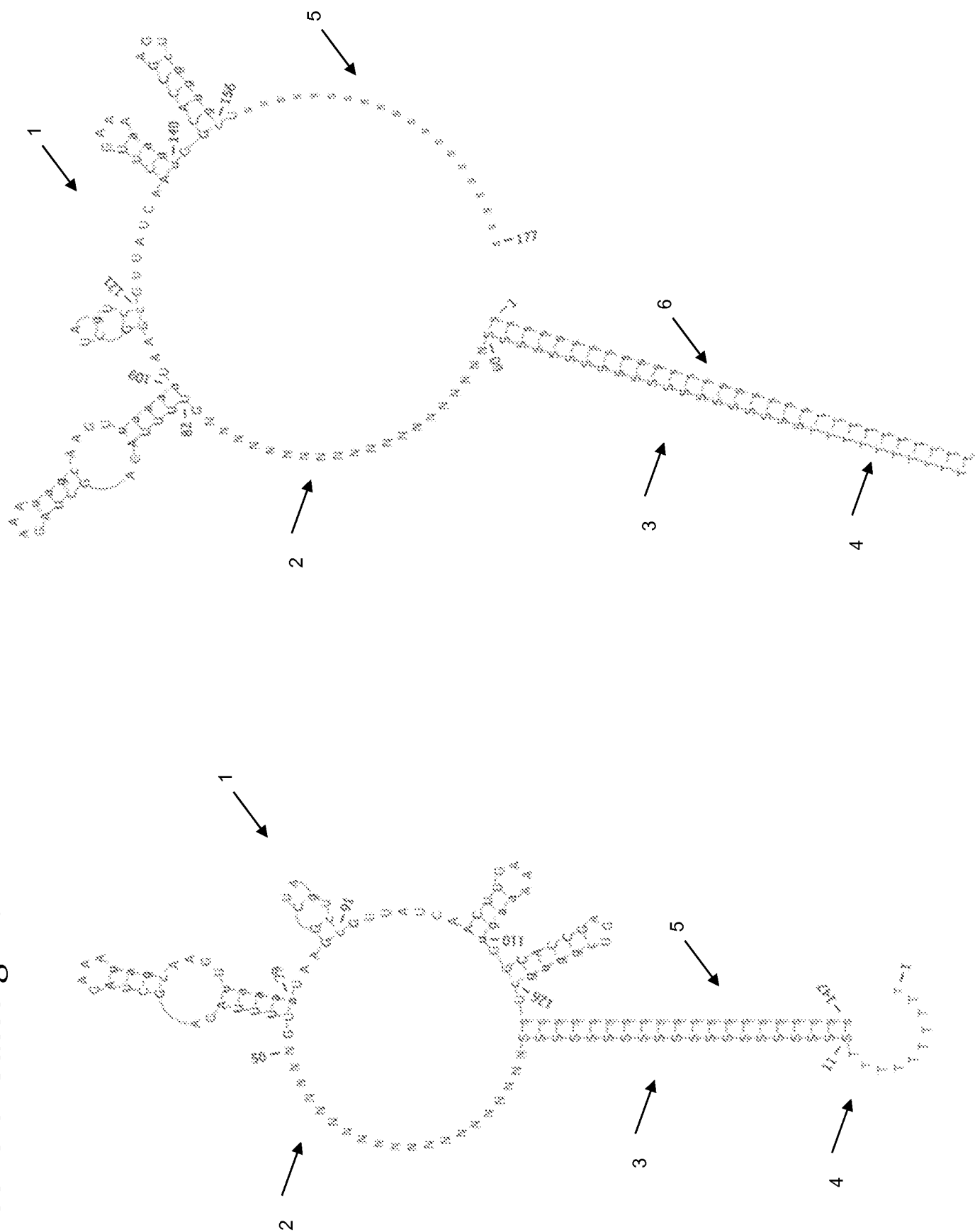


Fig. 3

triggering RNA-3' Toehold-Gated gRNA Complex

3' Toehold-Gated gRNA

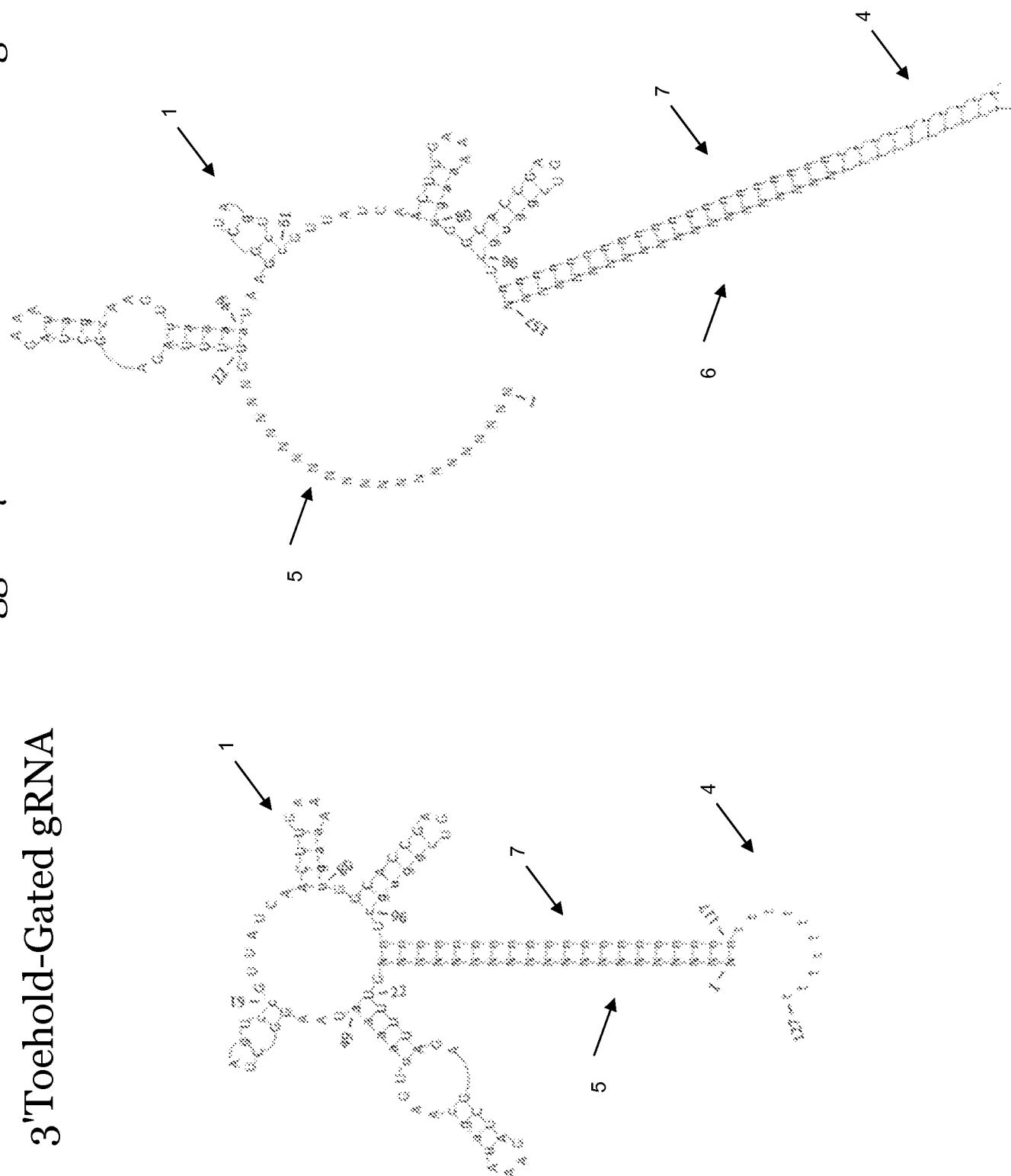


Fig. 4



AND Toehold-Gated gRNA

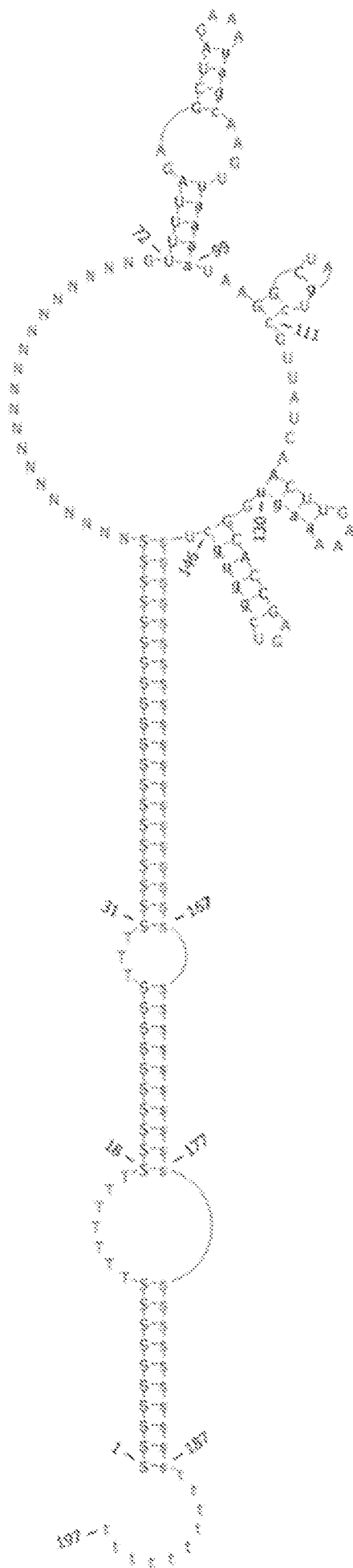


Fig. 6

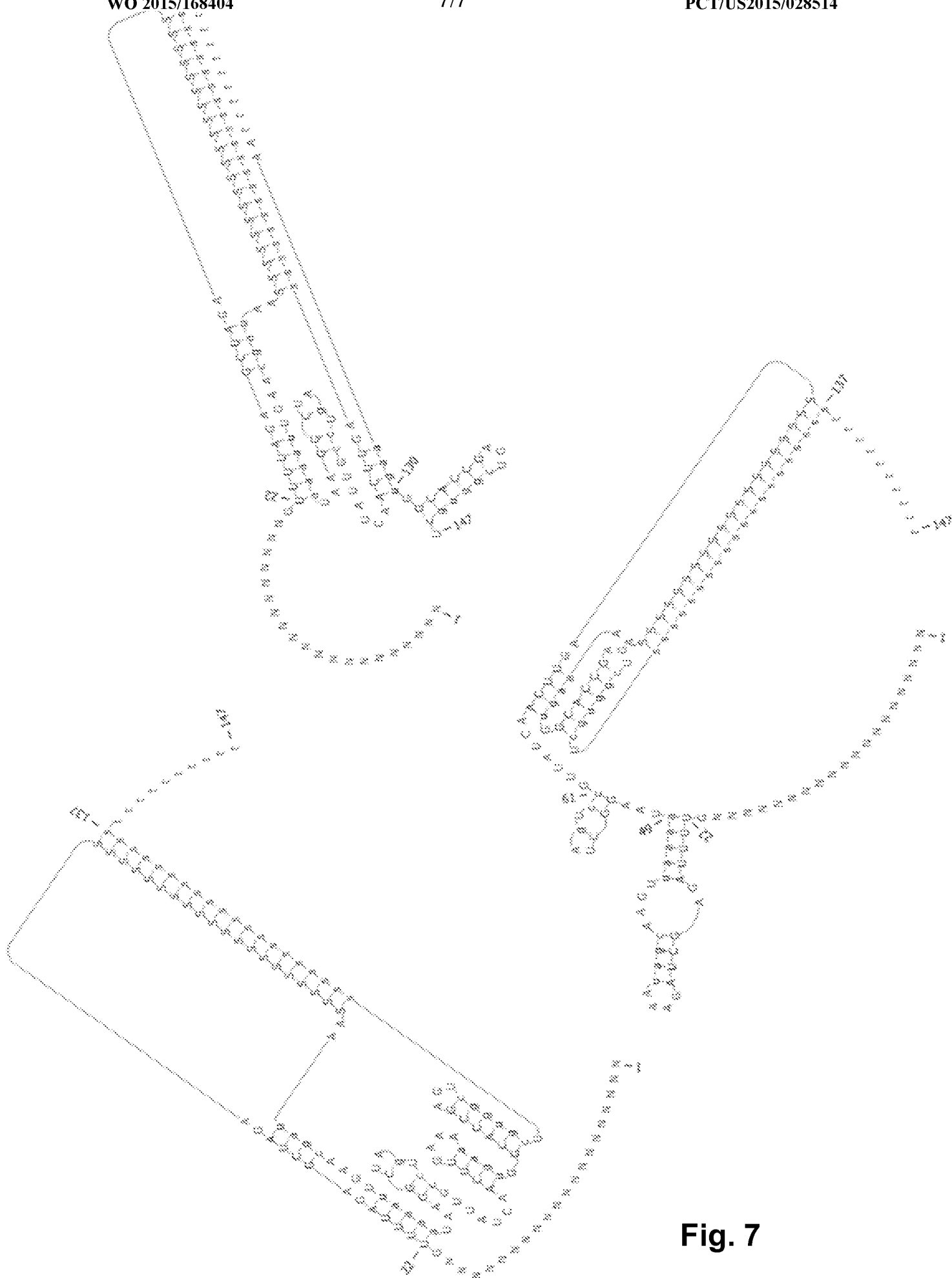


Fig. 7

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2015/028514

A. CLASSIFICATION OF SUBJECT MATTER

IPC(8) - C12N 9/22 (2015.01)

CPC - C12N 9/22 (2015.07)

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC(8) - C07H 21/02; C12N 9/22, 9/96, 15/85, 15/90, 15/113; C12Q 1/68; H01L 51/00 (2015.01)

CPC - C12N 9/22, 9/96, 15/85, 15/113, 15/907 (2015.07)

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

CPC - C12N 9/22, 9/96, 15/85, 15/113, 15/907 (2015.07) (keyword delimited)

USPC - 257/40; 435/6.1, 6.11, 188, 462; 536/24.5; 800/18; 977/704

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

Orbit, Google Patents, Google Scholar.

Search terms used: CAS OR CAS9 gated guide RNA thgRNA trigger circuit toehold toe hold in inventor:jacobson author:jakimo

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X -- Y	US 2007/0072215 A1 (SEELIG et al) 29 March 2007 (29.03.2007) entire document	1-7, 9-11, 14 ----- 8, 12, 13, 15-23
Y	JINEK et al. "Structures of Cas9 Endonucleases Reveal RNA-Mediated Conformational Activation," Science, 14 March 2014 (14.03.2014), Vol. 343, Pgs. 1247997-1247997. entire document	8
Y	US 2009/0234109 A1 (HAN et al) 17 September 2009 (17.09.2009) entire document	12, 13, 15, 16
Y	US 2014/0068797 A1 (UNIVERSITY OF VIENNA et al) 06 March 2014 (06.03.2014) entire document	17-23
A	US 8,697,359 B1 (ZHANG) 15 April 2014 (15.04.2014) entire document	1-23
A	GILBERT et al. "CRISPR-Mediated Modular RNA-Guided Regulation of Transcription in Eukaryotes," Cell, 18 July 2013 (18.07.2013), Vol. 154, Pgs. 442-451. entire document	1-23
A	KARVELIS et al. "crRNA and tracrRNA guide Cas9-mediated DNA interference in Streptococcus thermophilus," RNA Biology, 01 May 2013 (01.05.2013), Vol. 10, Pgs. 841-851. entire document	1-23

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

☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

14 August 2015

Date of mailing of the international search report

08 SEP 2015

Name and mailing address of the ISA/

Mail Stop PCT, Attn: ISA/US, Commissioner for Patents
P.O. Box 1450, Alexandria, Virginia 22313-1450

Facsimile No. 571-273-8300

Authorized officer

Blaine Copenheaver

PCT Helpdesk: 571-272-4300
PCT OSP: 571-272-7774