



- (51) International Patent Classification:
C12Q 1/68 (2006.01) *C12N 15/11* (2006.01)
- (21) International Application Number:
PCT/US2015/040460
- (22) International Filing Date:
14 July 2015 (14.07.2015)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
62/024,165 14 July 2014 (14.07.2014) US
- (71) Applicants: **PRESIDENT AND FELLOWS OF HARVARD COLLEGE** [US/US]; 17 Quincy Street, Cambridge, MA 02138 (US). **TRUSTEES OF BOSTON UNIVERSITY** [US/US]; One Silber Way, Boston, MA 02215 (US).
- (72) Inventors: **GREEN, Alexander, A.**; 7157 East Rancho Vista Drive, Suite 2010, Scottsdale, AZ 85251-1599 (US). **YIN, Peng**; 51 Winthrop Road, Brookline, MA 02445 (US). **COLLINS, James, J.**; 44 Cummington Street, Bo-

ston, MA 02215 (US). **KIM, Jongmin**; 24 Peabody Terrace, Apt 614, Cambridge, MA 02138 (US).

(74) Agent: **TREVISAN, Maria, A.**; Wolf Greenfield & Sacks, P.C., 600 Atlantic Avenue, Boston, MA 02210-2206 (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,

[Continued on next page]

(54) Title: COMPOSITIONS COMPRISING RIBOREGULATORS AND METHODS OF USE THEREOF

4-input OR with 2-input AND triggers (8-input circuit)

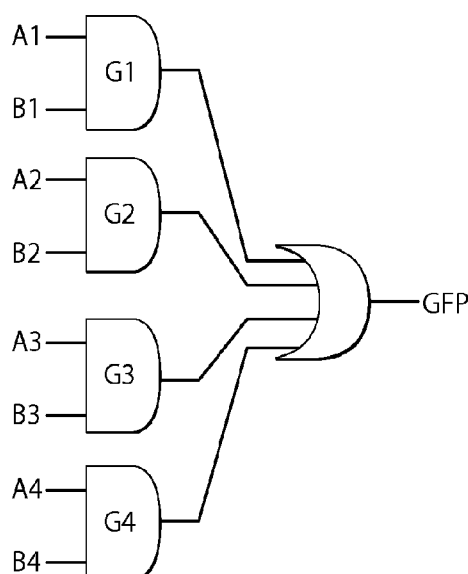


FIG. 37A

(57) Abstract: The invention provides novel and versatile classes of riboregulators, including inter alia activating and repressing riboregulators, switches, and trigger and sink RNA, and methods of their use for detecting RNAs in a sample such as a well and in modulating protein synthesis and expression.



SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ,
GW, KM, ML, MR, NE, SN, TD, TG).

— *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))*

Published:

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

— *with sequence listing part of description (Rule 5.2(a))*

- 1 -

COMPOSITIONS COMPRISING RIBOREGULATORS AND METHODS OF USE THEREOF

FEDERALLY SPONSORED RESEARCH

This invention was made with U.S. Government support under grant number 635-
5 67116-XXXX-167832-435548-0002.66625 awarded by DARPA and grant number
1DP2OD007292 awarded by the National Institutes of Health. The U.S. Government has
certain rights in the invention.

RELATED APPLICATIONS

10 This application claims the benefit of U.S. Provisional Application No. 62/024,165
filed July 14, 2014, the entire contents of each of which are incorporated by reference herein.

BACKGROUND OF INVENTION

15 Riboregulators are sequences of RNA that effect changes in cells in response to a
nucleic acid sequence. These RNA-based devices, which typically regulate protein translation
or trigger mRNA degradation, have been used for a number of applications in synthetic
biology, including sensitive control over gene expression, shunting of metabolic flux through
different metabolic pathways, and synthetic control over cell death.

In riboregulators that control gene expression, repression of protein translation has
20 relied on sequestration of the normally single-stranded ribosome binding site (RBS) within a
duplex RNA region that is upstream of a gene of interest (GOI). An mRNA in which the
RBS is sequestered within a hairpin upstream of the GOI is thus a cis-repressed RNA
(crRNA). A riboregulator based on an engineered crRNA can be constructed in which a
trans-activating RNA (taRNA) binds to the crRNA and unwinds the repressing RNA duplex
25 thereby exposing a now single-stranded RBS and activating translation of the downstream
gene. In riboregulators that decrease expression of the GOI, the RBS and initiation codon of
the GOI are both exposed in the absence of the trigger RNA. However, a trans-repressing
RNA (trRNA), which bears anti-sense sequence to the RBS and start codon, can bind to the
riboregulator and strongly suppress translation of the downstream gene.

30 Over the past decade, researchers have developed a number of different riboregulator
systems to control gene expression in prokaryotic cells. These previous systems have utilized
a number of recurring design motifs. The vast majority of riboregulators have employed loop-

- 2 -

linear interactions to drive the crRNA/trans-activating RNA hybridization reaction forward. In these interactions, a linear, single-stranded region in one of the strands binds to a loop established at the end of a duplex in the other strand. Essential in this interaction is the formation of a kissing loop structure in which the tertiary structure of the RNA sequence causes bases within the loop to flip outwards, facilitating binding with the linear RNA strand. Importantly, this kissing loop structure is only established with specific sequences inside the loop region, which can severely limit the number of possible crRNA designs.

All previous riboregulator systems have relied on sequestration of the RBS to impede translation of the GOI. This design choice has two crucial implications. First, much of the work in optimization of genetic circuits in synthetic biology relies on varying the strength of the RBS to finely tune protein levels inside the cell. Since the RBS sequence is a functional part of these riboregulators, one cannot simply replace the riboregulator RBS with variants and expect the output characteristics of the device to vary in a predictable manner following the strength of the new RBS. Furthermore, changes to the RBS will require corresponding modifications in the sequence of the trans-activating RNA. Second, for riboregulators that activate gene expression, riboregulators that sequester the RBS must be activated by taRNA sequences that are at least partially complementary to the crRNA RBS sequence. Consequently, unbound taRNAs can compete with de-repressed crRNA species for ribosome binding. Alternatively, RBS sequences within the taRNAs can also be sequestered within stem regions. This additional secondary structure can decrease the kinetics of binding with the crRNA and the dynamic range of the riboregulator.

SUMMARY OF INVENTION

The invention provides, in part, programmable riboregulators that can be activated by RNAs, including RNAs endogenous to a cell or sample of interest. The invention further provides programmable riboregulators, also referred to herein as toehold switches, that can be integrated into a genome, such as bacterial genome such as an *E. coli* genome, to regulate endogenous nucleic acids, such as genes, and to generate toehold switch sensors that respond to endogenous nucleic acids, such as RNAs. The invention further provides methods of use of the toehold switches, including for example methods of regulating a plurality (i.e., n number) of nucleic acids, such as genes, independently of each other using a plurality (e.g., n number) of toehold switches in a single cell. Such methods can be used in a synthetic biology application. In one exemplification, twelve such switches were used to regulate

- 3 -

twelve nucleic acids independently, in the same cell. The invention further provides methods for using the switches to generate a genetic circuit that evaluates 4-input AND logic.

Such programmable riboregulators have not been possible previously due in part to the structural constraints, including sequence constraints, outlined herein. The novel
5 riboregulators of the invention provide sufficient freedom in the sequence of the taRNA (trigger RNA) (and corresponding region of crRNA (e.g., switch RNA) to which the taRNA hybridizes) to allow for activation by, for example, RNAs such as but not limited to endogenous RNAs. When coupled to protein reporters such as fluorescent reporters, such riboregulators would act as sensors to probe RNA levels in real time in living cells or other
10 types of RNA-containing samples. The invention can be used to detect and quantitate endogenous RNA in real time without having to harvest the RNA from the cell (or sample). The method is sufficiently sensitive to detect RNA present at physiological copy numbers.

The riboregulators of the invention are less constrained in sequence than are those of the prior art, and accordingly a variety of riboregulators may be generated and importantly
15 used together in a single system such as a cell. Such orthogonality has not been possible heretofore using the riboregulators of the prior art. The riboregulators of the invention also do not depend upon the RBS for their structure. As a result, it is possible to modify the RBS without affecting the function of the riboregulator. The programmable nature of the riboregulators of the invention allow “plug and play” implementations of higher order
20 cellular logic.

The invention therefore provides methods for detecting (sensing) and measuring levels of one or more endogenous RNA, effecting sensitive control over one or more proteins simultaneously in a cell or sample (including translational control), performing complex logic operations in a cell or a sample, programming in a cell or sample, detecting single-nucleotide
25 polymorphisms (SNPs) in living systems, and detecting RNAs and SNP RNAs in in vitro translation systems, using the riboregulator (including the toehold switch RNA and/or the toehold repressors) and/or the taRNA (trigger RNA) and/or the sink RNA compositions of the invention.

The cis-repressing RNA (crRNA) and trans-activating RNA (taRNA) of the invention
30 may be comprised of RNA in whole or in part. They may comprise naturally occurring nucleotides and/or non-naturally occurring nucleotides. The crRNA may also be referred to herein as switch RNA. A crRNA intends an RNA that is typically repressed until bound to a taRNA (or trigger RNA), as such binding results in translation of a protein of interest from

- 4 -

the crRNA/switch RNA. Binding of the trigger RNA to the crRNA/switch typically occurs via a toehold domain, in some instances, and as described in greater detail herein.

The invention contemplates crRNA that may be modularly used via operable linkage to a coding domain. The invention further contemplates taRNA that may be modularly used to de-repress or activate crRNA.

Thus, in one aspect, this disclosure provides a system comprising a host cell having, integrated or encoded into its genome, a plurality of riboregulators, each riboregulator comprising an RNA comprising (i) a single-stranded toehold domain, (ii) a fully or partially double-stranded stem domain comprising an initiation codon, (iii) a loop domain comprising a ribosome binding site (RBS), and (iv) a coding domain.

In another aspect, this disclosure provides a system comprising a host cell having, integrated or encoded into its genome, a plurality of riboregulators, each riboregulator comprising an RNA comprising (i) a single-stranded toehold domain, (ii) a fully or partially double-stranded stem domain comprising an initiation codon, and (iii) a loop domain comprising a ribosome binding site, wherein each riboregulator is integrated upstream of an endogenous coding sequence, and wherein expression of the endogenous coding sequence is controlled by the riboregulator.

In certain embodiments, the host cell is a prokaryotic cell. In certain embodiments, the host cell is a bacterial cell. In certain embodiments, the host cell is an *E. coli* bacterium. In certain embodiments, the plurality is 5-15. In certain embodiments, the plurality is 10-15. In certain embodiments, the plurality is at least 10, including 10 to 20, or 10-30, or 10-40, or 10 to 50, or 10 to 100, or 10 to 500. In certain embodiments, the plurality is at least 12, and may range up to 20, 30, 40, 50, 100 or 500. In certain embodiments, riboregulators within a plurality are separated from each other by 0-30 nucleotides, or 9-15 nucleotides.

In certain embodiments, the riboregulator further comprises a spacer domain. In certain embodiments, the spacer domain encodes low molecular weight amino acids. In certain embodiments, the spacer domain is about 9-33 nucleotides in length. In certain embodiments, the spacer domain is about 21 nucleotides in length. In certain embodiments, the spacer domain is situated between the stem domain and the coding domain.

In certain embodiments, the stem domain comprises sequence upstream (5') and/or downstream (3') of the initiation codon. In certain embodiments, the sequence upstream of the initiation codon is about 6 nucleotides. In certain embodiments, the sequence downstream of the initiation codon is about 9 nucleotides. In certain embodiments, the

- 5 -

sequence downstream of the initiation codon does not encode a stop codon. In certain embodiments, the initiation codon is wholly or partially present in a 1-3 nucleotide bulge in the stem domain.

In certain embodiments, the coding domain encodes a reporter protein. In certain
5 embodiments, the reporter protein is green fluorescent protein (GFP).

In certain embodiments, the coding domain encodes a non-reporter protein.

In certain embodiments, the toehold domain is complementary in sequence to a naturally occurring RNA or a portion thereof. In certain embodiments, the toehold domain is complementary in sequence to a non-naturally occurring RNA. In certain embodiments, the
10 first domain is 100% complementary to the toehold domain.

In certain embodiments, the system further comprises a plurality of trans-activating RNA (taRNA), each comprising (i) a first domain that hybridizes to the toehold domain of one of the riboregulators in the plurality and that comprises no or minimal secondary structure, and (ii) a second domain that hybridizes to a sequence downstream (3') of the
15 toehold domain, wherein each taRNA has a cognate (or partner) riboregulator. In this context, a taRNA and a riboregulator are cognates if they are able to bind to each other and effect changes to the riboregulator structure, but not bind to other taRNA and riboregulators with the same structural (and functional) effect.

In another aspect, this disclosure provides a method of detecting presence of a
20 plurality of RNA in a sample, comprising combining a plurality of riboregulators with a sample, wherein each riboregulator (i) comprises a toehold domain that is complementary to an endogenous RNA and (ii) a coding domain that encodes a reporter protein, under conditions that allow translation of the coding domain in the presence of the endogenous RNA but not in the absence of the endogenous RNA, and detecting the reporter protein as an
25 indicator of the endogenous RNA, wherein each riboregulator detects a different endogenous RNA from all other riboregulators in the plurality, and each riboregulator encodes a different reporter protein from all other riboregulators in the plurality.

In another aspect, this disclosure provides a method of detecting presence of a
30 plurality of RNA in a cell, comprising introducing into the cell a plurality of riboregulators, wherein each riboregulator comprises (i) a toehold domain that is complementary to an endogenous RNA in the cell and (ii) a coding domain that encodes a reporter protein, and detecting the reporter protein as an indicator of the endogenous RNA, wherein each riboregulator detects a different endogenous RNA from all other riboregulators in the

- 6 -

plurality, and each riboregulator encodes a different reporter protein from all other riboregulators in the plurality. In this and other embodiments, the riboregulators may be introduced into the cell as an RNA or encoded in a DNA expression vector, for example.

In certain embodiments, the amount of reporter protein is an indicator of amount of endogenous RNA.

In another aspect, this disclosure provides a method of controlling gene and/or protein expression in a cell comprising integrating or encoding a plurality of riboregulators into the genome of the cell, each riboregulator integrated or encoded upstream of a target coding sequence, modulating expression of one or more of plurality of trans-activating RNA

(taRNA) in the cell, wherein expression of a taRNA in the cell results in increased expression of the target coding sequence, and wherein each riboregulator comprises an RNA comprising (i) a single-stranded toehold domain, (ii) a fully or partially double-stranded stem domain comprising an initiation codon, and (iii) a loop domain comprising a ribosome binding site, wherein each taRNA comprises (i) a first domain that hybridizes to the toehold domain of one of the riboregulators in the plurality and that comprises no or minimal secondary structure, and (ii) a second domain that hybridizes to a sequence downstream (3') of the toehold domain.

In certain embodiments, expression of a plurality of target coding sequences is controlled. In certain embodiments, the plurality of target coding sequences encode proteins that interact with each other directly or indirectly.

In certain embodiments, the plurality of taRNA are integrated or encoded in the host cell genome. In certain embodiments, each taRNA is operably linked to an inducible promoter that is different from all the other taRNA in the plurality. In certain embodiments, each taRNA has a cognate riboregulator in the cell. In certain embodiments, at least one taRNA activates two or more riboregulators in the cell.

It will be understood in the context of this and other embodiments provided herein examples of riboregulators include crRNAs, switch RNAs, toehold switches, toehold riboregulators, toehold repressors, beacon switches, beacon riboregulators, and the like. Similarly, in the context of this and other embodiments provided herein, the terms taRNA, input RNA, trigger RNA, input, trigger, and the like refer to the nucleic acid that binds to a repressor, in whole or in part, and/or which binds to other input or trigger nucleic acids thereby forming a nucleic acid complex that binds to a repressor and effects a change in the repressor structure and/or function. The latter category of inputs include those that contribute

- 7 -

to an AND gate. Thus, an AND gate involves two or more triggers that must hybridize to each other to form a complex that itself is capable of binding to the repressor and causing structural and functional changes to the repressor. Some but not all such AND gate triggers may comprise nucleotide sequence that is complementary and capable of hybridizing to a nucleotide sequence in the repressor.

In another aspect, this disclosure provides a method of controlling gene and/or protein expression in a cell comprising introducing a plurality of riboregulators into a cell, each riboregulator comprises an RNA comprising (i) a single-stranded toehold domain, (ii) a fully or partially double-stranded stem domain comprising an initiation codon, (iii) a loop domain comprising a ribosome binding site, and (iv) a coding domain for a reporter protein or a protein of interest, modulating expression of one or more of a plurality of trans-activating RNA (taRNA) in the cell, wherein expression of a taRNA in the cell results in increased expression of the coding domain, and wherein each taRNA comprises (i) a first domain that hybridizes to the toehold domain of one of the riboregulators in the plurality and that comprises no or minimal secondary structure, and (ii) a second domain that hybridizes to a sequence downstream (3') of the toehold domain.

In certain embodiments, one or more riboregulators comprise a coding domain for a transcription factor.

In certain embodiments, modulating expression of one or more of a plurality of taRNA comprises increasing expression of a subset of taRNA substantially simultaneously.

Also provided herein is a system comprising a plurality of riboregulators upstream of a coding domain, each riboregulator comprising (i) a single-stranded toehold domain, (ii) a fully or partially double-stranded stem domain comprising an initiation codon, and (iii) a loop domain comprising a ribosome binding site, wherein the riboregulators are separated from each other by a spacer of 9-15 nucleotides in length. The spacer between the last base of one riboregulator (the last 3' base at the stem of the riboregulator) and the first base of the adjacent riboregulator (the first 5' base of the toehold domain) may be 9, 10, 11, 12, 13, 14, or 15 nucleotides.

In certain embodiments, the system further comprises a plurality of trans-activating RNA (taRNA), wherein a different taRNA or a different subset of taRNAs is required to activate each of the riboregulators.

In certain embodiments, a different subset of taRNAs is required to activate each of the riboregulators, and the members of each subset of taRNAs hybridize to each other to form

- 8 -

a complex that is capable of hybridizing to the toehold domain of a riboregulator. In certain embodiments, a different subset of taRNAs is required to activate each of the riboregulators, and at least two members of each subset of taRNAs are partially complementary to a toehold domain and/or to the sequence downstream of the toehold domain in a single riboregulator.

5 In certain embodiments, the plurality of riboregulators is 5 or 6. In certain embodiments, the subset of taRNAs comprises 2 taRNAs.

Also provided herein is a toehold crRNA (toehold switch) riboregulator comprising a single-stranded toehold domain, a fully or partially double-stranded stem domain comprising an initiation codon, and a loop domain comprising a ribosome binding site. The toehold
10 crRNA/toehold switch may comprise an RBS sequence located in the loop domain.

Also provided herein is an RNA comprising more than one crRNA, optionally operably linked to a coding domain (as described below), wherein the multiple crRNA may be activated by the same or by different taRNA (trigger RNA). In some embodiments, a single taRNA may activate expression of a downstream coding sequence. In such
15 embodiments, the toehold crRNA riboregulator may be used to detect expression of a plurality of taRNA using a single readout.

Also provided herein is a toehold riboregulator system comprising (1) a crRNA riboregulator comprising a single-stranded toehold domain, a fully or partially double-stranded stem domain comprising an initiation codon, and a loop domain comprising a
20 ribosome binding site, and (2) a coding domain. In some embodiments, taRNAs that hybridize to complementary regions in the stem domain activate expression of a downstream coding sequence. In some embodiments, 2, 3, 4, 5, 6, or more or all of the taRNAs are required in order to activate expression of the downstream coding sequence. The terms system and device are used interchangeably herein to refer to a collection of riboregulator
25 components including but not limited to and in any combination crRNA (switch RNA), taRNA (trigger RNA), sink RNA, and the like.

In some embodiments, the riboregulator further comprises a spacer domain. In some embodiments, the spacer domain encodes low molecular weight amino acids. In some embodiments, the spacer domain is about 9-33 nucleotides in length. In some embodiments,
30 the spacer domain is about 21 nucleotides in length. In some embodiments, the spacer domain is situated between the stem domain and the coding domain. In some embodiments, the spacer domain is greater than 33 nucleotides in length and can contain single- and double-stranded regions, including other riboregulators.

- 9 -

In some embodiments, the stem domain comprises sequence upstream (5') and/or downstream (3') of the initiation codon. In some embodiments, the sequence upstream of the initiation codon is about 6 nucleotides. In some embodiments, the sequence downstream of the initiation codon is about 9 nucleotides. In some embodiments, the sequence downstream
5 of the initiation codon does not encode a stop codon.

In some embodiments, the coding domain encodes a reporter protein. In some embodiments, the reporter protein is green fluorescent protein (GFP). In some embodiments, the coding domain encodes a non-reporter protein. As used herein, a non-reporter protein is any protein that is used or that functions in a manner in addition to or instead of as a reporter
10 protein. A non-reporter protein may interact with another entity in the cell or sample, and may thereby effect a change in the cell or sample or in another moiety.

In some embodiments, the toehold domain is complementary in sequence to a naturally occurring RNA. A naturally occurring RNA may be an RNA that is capable of being expressed from the cell of interest (e.g., from an endogenous gene locus). In some
15 embodiments, the toehold domain is complementary in sequence to a non-naturally occurring RNA. A non-naturally occurring RNA may be an RNA that is not naturally expressed in a cell of interest (e.g., it is not expressed from an endogenous gene locus), and may instead be expressed from an exogenous nucleic acid introduced into the cell of interest.

Also provided herein is a trans-activating RNA (taRNA) comprising a first domain
20 that hybridizes to a toehold domain of any of the foregoing riboregulators and that comprises no or minimal secondary structure, and a second domain that hybridizes to a sequence downstream (3') of the toehold domain. In some embodiments, the first domain is 100% complementary to the toehold domain. In some embodiments, the second domain may be less than 100% complementary to the sequence downstream of the toehold domain.

The taRNA may consist of more than one strand of RNA, and such multiple RNAs in
25 combination provide the first and second domain for hybridization with the crRNA. In some embodiments, one or more RNAs may be used to bring multiple taRNAs into close proximity via hybridization to enable them to efficiently hybridize with the riboregulator. Examples of such embodiments are illustrated in FIGs. 9 and 10A-C.

Also provided herein is a system comprising one or more of any of the foregoing
30 crRNA riboregulators, and/or one or more of any of the foregoing trans-activating RNA (taRNA). The taRNA may all be naturally occurring RNA, or they may all be non-naturally

- 10 -

occurring RNA, or they may be a mixture of naturally occurring RNA and non-naturally occurring RNA.

The systems of the invention may include a plurality of riboregulators (e.g., a plurality of crRNA/switches, optionally together with cognate taRNA/trigger RNA) having minimal cross-talk amongst themselves. In some embodiments, the systems may comprise at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 or more toehold crRNA/switches, having minimal cross-talk (e.g., on the level of less than 20%, 19%, 18%, 17%, 16%, 15%, 14%, 13%, 12%, 11%, 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2% or less). In some embodiments, the toehold crRNA/switches have an average ON/OFF fluorescence ratio of more than 50, 100, 150, 200, 250, 300, 350, 400, or more. In some instances, the invention provides systems having a plurality of toehold crRNA/switches having an average ON/OFF fluorescence ratio in the range of about 200-665, including about 400. In some embodiments, the level of cross-talk amongst a plurality of toehold riboregulators in a system ranges from about 2% to less than 20%, or from about 2% to about 15%, or from about 5% to about 15%. Such systems may comprise 7 or more, including 8, 9, 10, etc. toehold riboregulators.

In some embodiments, the system is a cell. In some embodiments, the cell is a prokaryotic cell. The riboregulator system or components of the system may be introduced into the system, including encoded in nucleic acids that are introduced into the system.

In some embodiments, the system is a cell-free in vitro system.

In some embodiments, the crRNA riboregulator and the taRNA are hybridized to each other.

In some embodiments, the ratio of crRNA riboregulator to taRNA is less than 1, less than 0.5, or less than 0.1.

In some embodiments, the crRNA riboregulator or riboregulator system is comprised or encoded in a first nucleic acid and the taRNA is comprised or encoded in a second nucleic acid. In some embodiments, the first nucleic acid is a first plasmid and the second nucleic acid is a second plasmid. In some embodiments, the first plasmid comprises a medium copy origin of replication and the second plasmid comprises a high copy origin of replication.

The plasmids may be DNA plasmids or RNA plasmids. In the event the plasmids are DNA plasmids, the riboregulator and taRNA are encoded in the DNA plasmid. It will be understood that upon transcription of the DNA plasmid, as described and demonstrated in the Examples, the resultant RNA species will include the riboregulator and taRNA in RNA form.

- 11 -

It will be further understood that any given nucleic acid construct, whether DNA or RNA in nature, such as but not limited to a plasmid or an expression vector, may comprise or encode one or more riboregulators (including any of the toehold or beacon switches described herein) or one or more taRNAs (or other input or trigger RNAs described herein).

5 Also provided herein is a nucleic acid comprising any of the foregoing crRNA riboregulators or riboregulator systems or comprising sequences that encode any of the foregoing crRNA riboregulators or riboregulator systems. In another aspect, the invention provides a host cell comprising any of the foregoing nucleic acids including nucleic acids that encode any of the foregoing nucleic acids.

10 Also provided herein is a nucleic acid comprising any of the foregoing trans-activating RNA (taRNA) or comprising sequences that encode any of the foregoing taRNA. In another aspect, the invention provides a host cell comprising the nucleic acid.

Also provided herein is a method of detecting presence of an RNA in a sample, comprising combining any of the foregoing toehold crRNA riboregulator systems with a
15 sample, wherein the crRNA riboregulator comprises a toehold domain that is complementary to an endogenous RNA, and wherein the riboregulator system comprises a coding domain that encodes a reporter protein, under conditions that allow translation of the coding domain in the presence of the endogenous RNA but not in the absence of the endogenous RNA, and detecting the reporter protein as an indicator of the endogenous RNA. As used herein,
20 conditions that allow translation of the coding domain are conditions that include all the necessary machinery to produce a protein from an RNA such as but not limited to ribosomes, tRNAs, and the like.

Also provided herein is a method of detecting presence of an RNA in a cell, comprising introducing into the cell any of the foregoing toehold riboregulator systems,
25 wherein the crRNA riboregulator comprises a toehold domain that is complementary to an endogenous RNA in the cell, and wherein the riboregulator system comprises a coding domain that encodes a reporter protein, and detecting the reporter protein as an indicator of the endogenous RNA. In some embodiments, the reporter protein is green fluorescent protein (GFP). In some embodiments, amount of reporter protein is an indicator of amount of
30 endogenous RNA.

Also provided herein is a method of controlling protein translation, comprising combining any of the foregoing toehold riboregulator systems with any of the foregoing complementary taRNA, wherein the toehold crRNA riboregulator comprises a toehold

- 12 -

domain that is complementary to the taRNA, and wherein the toehold riboregulator system comprises a coding domain that encodes a non-reporter protein, under conditions that allow translation of the coding domain in the presence of the taRNA but not in the absence of the taRNA.

5 Also provided herein is a beacon riboregulator system comprising (1) a beacon crRNA riboregulator comprising a fully or partially double-stranded stem domain comprising a ribosome binding site, and a loop domain, (2) a coding domain, and (3) an initiation codon present between the stem domain and the coding domain.

In some embodiments, the stem domain comprises sequence upstream (5') of the
10 initiation codon. In some embodiments, the sequence upstream of the initiation codon is about 6 nucleotides.

In some embodiments, the coding domain encodes a reporter protein. In some
embodiments, the reporter protein is green fluorescent protein (GFP). In some embodiments, the coding domain encodes a non-reporter protein.

15 In some embodiments, the loop domain is complementary in sequence to a naturally occurring RNA. In some embodiments, the loop domain is complementary in sequence to a non-naturally occurring RNA. In some embodiments, the loop domain is about 21 nucleotides in length. In some embodiments, the loop domain ranges in length from about 15-30 nucleotides.

20 In some embodiments, the beacon crRNA riboregulator comprises a binding domain (i.e., a domain that hybridizes to its complementary taRNA) that includes but is not limited to the loop domain. The binding domain may comprise a region upstream (5') of the loop domain that may be about 9 nucleotides in length and which may exist in the stem domain.

The stem domain may be about 23 bps in length. The stem domain may range from
25 about 15 bp to about 30 bps.

Also provided herein is a trans-activating RNA (taRNA) comprising a first domain that hybridizes to a loop domain of any of the foregoing beacon riboregulators and that comprises no or minimal secondary structure, and a second domain that hybridizes to a
30 sequence upstream (5') of the loop domain and present in the stem domain. In some embodiments, the first domain is 100% complementary to the loop domain.

Also provided herein is a system comprising one or more of any of the foregoing beacon crRNA riboregulators, optionally operably linked to a coding domain, and any of the foregoing complementary trans-activating RNA (taRNA).

- 13 -

In some embodiments, the system is a cell. In some embodiments, the cell is a prokaryotic cell. In some embodiments, the system is a cell-free in vitro system.

In some embodiments, the beacon crRNA riboregulator and the taRNA are hybridized to each other.

5 In some embodiments, the ratio of beacon crRNA riboregulator to taRNA is less than 1, less than 0.5, or less than 0.1.

In some embodiments, the beacon crRNA riboregulator (or system) is comprised or encoded in a first nucleic acid and the taRNA is comprised or encoded in a second nucleic acid. In some embodiments, the first nucleic acid is a first plasmid and the second nucleic acid is a second plasmid. In some embodiments, the first plasmid comprises a medium copy origin of replication and the second plasmid comprises a high copy origin of replication. The plasmids may be DNA plasmids or RNA plasmids.

Also provided herein is a nucleic acid comprising any of the foregoing beacon crRNA riboregulators (or systems) or sequences that encode any of the foregoing beacon crRNA riboregulators (or systems). In another aspect, the invention provides a host cell comprising said nucleic acid.

Also provided herein is a nucleic acid comprising any of the foregoing trans-activating RNA (taRNA) or sequences that encode any of the foregoing taRNA. In another aspect, the invention provides a host cell comprising said nucleic acid.

20 Also provided herein is a method of detecting presence of an RNA in a sample, comprising combining a beacon riboregulator system with a sample, wherein the beacon crRNA riboregulator comprises a loop domain that is complementary to an endogenous RNA, and wherein the beacon riboregulator system comprises a coding domain that encodes a reporter protein, under conditions that allow translation of the coding domain in the presence of the endogenous RNA but not in the absence of the endogenous RNA, and detecting the reporter protein as an indicator of the endogenous RNA.

Also provided herein is a method of detecting presence of an RNA in a cell, comprising introducing into the cell a beacon riboregulator system, wherein the beacon crRNA riboregulator comprises a loop domain that is complementary to an endogenous RNA in the cell, and wherein the beacon riboregulator system comprises a coding domain that encodes a reporter protein, and detecting the reporter protein as an indicator of the endogenous RNA.

In some embodiments, the reporter protein is green fluorescent protein (GFP).

- 14 -

In some embodiments, amount of reporter protein is an indicator of amount of endogenous RNA.

Also provided herein is a method of controlling protein translation, comprising combining a beacon riboregulator system with a complementary taRNA, wherein the beacon crRNA riboregulator comprises a loop domain that is complementary to the taRNA, and wherein the beacon riboregulator system comprises a coding domain that encodes a non-reporter protein, under conditions that allow translation of the coding domain in the presence of the taRNA but not in the absence of the taRNA.

These and other aspects and embodiments of the invention will be described in greater detail herein.

BRIEF DESCRIPTION OF DRAWINGS

FIG. 1. Schematic of the toehold riboregulator crRNA base design. The corresponding taRNA has the sequence 5'-b-a-3' where domains a and b are the reverse complements of domains a* and b*, respectively.

FIG. 2. Characterization of the repression level of six inactivated toehold riboregulator crRNAs.

FIG. 3. On/off mode fluorescence ratio obtained for a high performance toehold riboregulator.

FIG. 4. On/off mode fluorescence ratios obtained for a set of 61 toehold riboregulators three hours after induction with IPTG.

FIG. 5. Beacon riboregulator base design. The taRNA has the sequence 5'-b-a-3'.

FIG. 6. On/off median fluorescence intensity obtained for a set of six beacon riboregulator devices. Dotted red line marks an on/off ratio of 10.

FIG. 7. Response of a beacon riboregulator targeted by the small RNA ryhB. The riboregulator sensor was induced using 1 mM IPTG and ryhB was induced using 0.5 mM 2,2'-dipyridyl. The riboregulator sensor responded to increased intracellular ryhB levels by increasing output of GFP by a factor ~5.

FIGs. 8A-B. Design schematics for other endogenous sensors based on the toehold (FIG. 8A) and beacon (FIG. 8B) riboregulators that are programmed to sense targets (taRNAs or triggers) with the sequence 5'-b-a-3'. Both designs employ strong RNA duplexes before and after the AUG start codon to repress protein translation. (FIG. 8A) Toehold riboregulator with an extended toehold (more than 21 nucleotides (nts) in some implementations) to

- 15 -

encourage strong binding of an RNA target with significant secondary structure. crRNA stem unwinding region is reduced in size but will allow trans-activation of translation since the stem nearest RBS is short (typically 6 base pairs (bp)) and likely to spontaneously unwind. (FIG. 8B) Beacon riboregulator possesses a larger loop (typically 32-nts) for target binding and the RBS is now in the loop to allow greater programmability.

FIGs. 9A-B illustrate a system in which two taRNAs work together and contribute to the 5'-a-b-3' sequence that hybridizes to a riboregulator crRNA. (FIG. 9A) Schematic illustration of a two-input AND gate system in which RNA strands A and B are inputs and strand C, a crRNA, functions as the gate. (FIG. 9B) On/off fluorescence ratios obtained for all combinations of RNA strands A, B, and C.

FIG. 10A illustrates a system in which two taRNAs each with part of the 5'-a-b-3' sequence are brought into close proximity by a third taRNA that does not contain any part of the 5'-a-b-3' sequence.

FIG. 10B illustrates a 3 input AND system and ON/OFF ratios in the presence of the various combinations of the 3 RNA inputs.

FIG. 10C illustrates a 4 input AND system and ON/OFF ratios in the presence of the various combinations of the 4 RNA inputs.

FIGs. 11A-C. Implementation of 2-input OR logic in vivo using riboregulators. (FIG. 11A) Three programmed RNA strands in the system. (FIG. 11B) Schematic of OR gate activation in vivo. (FIG. 11C) Flow cytometry measurements of on/off fluorescence from GFP upon transcription of different input RNAs to the system. In the off case, a taRNA that is non-cognate to the gate is expressed.

FIGs. 12A-B. Implementation of a 6-input OR gate in vivo. (FIG. 12A, top) The OR gate system is comprised of six crRNA arranged in series upstream of the GFP gene. (FIG. 12A, middle) The corresponding six taRNA inputs were all found to activate GFP expression from *E. coli* colonies induced on LB/IPTG plates. In contrast, four different non-cognate taRNAs did not elicit GFP production when co-expressed with OR gate construct. (FIG. 12B) Flow cytometry measurements of the On/Off mode GFP fluorescence ratio for the OR gate system. All six programmed input taRNAs exhibit greater than 10-fold higher GFP expression compared to the non-cognate taRNA with lowest GFP leakage levels (Y).

FIGs. 13A-B. (FIG. 13A) Schematic illustration of the six-input AND gate system. The gate consists of an extended hairpin containing sequences from validated toehold riboregulator crRNAs. The six input RNA triggers contain sequences from the corresponding

- 16 -

taRNAs and hybridization domains for binding to neighboring input strands. (FIG. 13B) Images of GFP fluorescence from *E. coli* colonies for the 6-bit AND gate exposed to the specified combinations of inputs A through F. Strong GFP expression is observed only when all six inputs are present, as shown in the far right column.

5 FIGs. 14A-B. In vivo demonstration of trigger RNA inactivation by a sink RNA. (FIG. 14A) Schematic showing the molecular interactions underlying the logic operations. The sink RNA is designed to outcompete the switch RNA for binding to the trigger. This preferential binding prevents the trigger from activating the switch whenever the sink is also present. (FIG. 14B) GFP fluorescence measured from the switch RNA with different
10 combinations of trigger and sink RNAs. Ninety per cent (90%) repression of fluorescence is observed when the sink is co-expressed with the trigger RNA compared to when the trigger alone is expressed.

 FIGs. 15A-B. Toehold repressor design and performance. (FIG. 15A) Schematic illustration showing the molecular interactions of a toehold repressor system. The trigger
15 RNA causes the switch RNA to refold into a configuration that prevents the ribosome from accessing binding elements on the RNA. (FIG. 15B) Repression levels measured from a library of 44 toehold repressors. Half of the systems provide greater than 90% repression. Dashed and dotted lines at 90% and 80% repression, respectively, are provided.

 FIG. 16. Time course measurements for a high performance toehold repressor.
20 Measurements were taken at hour time points following addition of the inducer IPTG.

 FIGs. 17A-D. Toehold switch design and output characteristics. (FIG. 17A) Conventional riboregulator systems repress translation by base pairing directly to the RBS region. RNA-RNA interactions are initiated via a loop-linear interaction at the YUNR-loop in an RNA hairpin. Interaction initiation region is denoted by thicker lines. (FIG. 17B)
25 Toehold switches repress translation through base pairs programmed before and after the start codon AUG, leaving the RBS and start codon regions completely unpaired. RNA-RNA interactions are initiated via linear-linear interaction domains called toeholds. The toehold domain (a*) binds to a complementary a domain on the trigger RNA. The ensuing branch migration de-represses the toehold switch mRNA to enable translation of the downstream
30 gene. (FIG. 17C) GFP mode fluorescence levels measured for switches in their on and off states as well as positive controls in which GFP with an identical sequence is expressed. Dashed black line marks the background fluorescence level obtained from IPTG-induced cells not bearing a GFP expressing plasmid. (FIG. 17D) On/off GFP fluorescence levels

- 17 -

obtained for a set of 168 toehold switches with 20 displaying on/off ≥ 100 . Inset: On/off GFP fluorescence measured for four toehold switches of varying performance levels at different time points following induction with IPTG.

FIGs. 18A-C. Comprehensive assessment of toehold switch orthogonality. (FIG. 18A) GFP fluorescence from colonies of *E. coli* expressing 676 pairwise combinations of switch mRNAs and trigger RNAs. GFP expressing colonies are visible along the diagonal in cells containing cognate switch and trigger strands. Off diagonal components have low fluorescence as a result of minimal interaction between non-cognate RNA components. (FIG. 18B) Crosstalk measured by flow cytometry for all trigger-switch combinations confirming strong overall system orthogonality. (FIG. 18C) Comparison of orthogonal library dynamic range (reciprocal of the threshold crosstalk level) and orthogonal library size for the toehold switches and a number of previous RNA-based regulators.

FIGs. 19A-D. Sequence analysis and forward engineering of toehold switches. (FIG. 19A) Regions and parameters critical to toehold switch output characteristics. (FIG. 19B) Evaluation of 168-member toehold switch library as a function of the number of G-C base pairs in the top and bottom three base pairs in the switch mRNA stem. Color of the background squares in the figure correspond to the mean on/off GFP fluorescence for the set of riboregulators that satisfy the specified GC base pairing constraints. Color of the circles within each square corresponds to the actual on/off ratio obtained for each of the components that satisfy the constraints. (FIG. 19C) On/off GFP fluorescence ratios obtained for the set of 13 forward engineered toehold switches. Dashed black line marks the mean on/off fluorescence level measured for the full set of 168 random sequence toehold switches. Inset. Time course measurements for forward engineered switches number 6 and number 9. (FIG. 19D) Percentage of random sequence and forward engineered library components that had on/off ratios that exceeded a specific value.

FIGs. 20A-D. Thermodynamic analysis of toehold switches. (FIG. 20A) Map of R2 values as a function of different thermodynamic parameters applied to subsets of on/off levels from the random sequence toehold switch library. The strongest correlation is found with the Δ GRBS-linker parameter (shown in red) for the subset of switches with a weak A-U base pair at the top of their stem. (FIG. 20B) Schematic illustrations showing position of the stem top base pair and the sequence range used to define Δ GRBS-linker. (FIG. 20C) Correlation between Δ GRBS-linker and on/off ratio measured for the 68 components in the toehold switch library with an A-U base pair at the top of the hairpin stem. (FIG. 20D) Strong

- 18 -

correlation between Δ GRBS-linker and on/off ratio measured for the set of forward engineered systems.

FIGs. 21A-E. Independent regulation and mRNA-based triggering using toehold switches. (FIG. 21A) Two orthogonal toehold switches triggered by RNAs A and B that independently regulate GFP and mCherry, respectively. (FIG. 21B) Two dimensional histograms of GFP and mCherry fluorescence for cells expressing all four input combinations of RNAs A and B confirm intended system behavior four hours after induction with IPTG. (FIG. 21C) mRNA-responsive toehold switches utilize an extended toehold domain denoted c* to bind to mRNA triggers with extensive secondary structure and activate expression of a GFP reporter. (FIG. 21D) On/off GFP fluorescence ratios for a series of toehold switches activated by the mCherry mRNA, and cat and aadA mRNAs conferring antibiotic resistance. (FIG. 21E) Mode GFP and mCherry fluorescence obtained from flow cytometry of mCherry sensors in their repressed and active states. Control expression levels were obtained from uninduced cells free of GFP-bearing plasmids and induced cells expressing either GFP or mCherry.

FIGs. 22A-B. Toehold switch activated by mRNA and endogenous small RNA triggers. (FIG. 22A) Endogenous and synthetic gene networks used for sensing the *ryhB* sRNA. (FIG. 22B) Transfer function for the *ryhB* sensor as a function of *ryhB* inducer concentration. Output of a constitutive GFP expression cassette is shown for comparison.

FIGs. 23A-D. Synthetic regulation of endogenous genes. (FIG. 23A) Integration of switch modules into the genome. A linear DNA fragment containing a kanamycin resistance marker and a switch sequence is inserted into the genome upstream of the targeted gene (gene B) using “lambda” Red recombination. The resistance marker is excised from the chromosome using FLP recombinase leaving a lone FRT site and the switch module at the 5' end of gene B. The switch-edited gene B is translationally repressed, but can be activated post-transcriptionally via the cognate trigger RNA. (FIG. 23B) Images of *uidA*::Switch A and *uidA*::Switch B spread onto X-Gluc plates with different trigger RNAs. *uidA* expression like the wild-type (top left) is only observed with cognate trigger RNAs as seen by blue/green color change. (FIG. 23C) Images of *lacZ*::Switch C with different combinations of IPTG and aTc chemical inducers. *lacZ*::Switch C only activates with aTc-induced expression of trigger C in conditions where *lacZ* is transcribed (in the presence of IPTG). The change in color to blue/green colonies only occurs when both IPTG and aTc are present. Wild-type *lacZ* (top left) is activated whenever IPTG is present. (FIG. 23D) Motility assays for *cheY*::Switch D

- 19 -

on soft agar plates. cheY::Switch D is only able to move away from the point of inoculation at the plate center when trigger D is induced with IPTG. In the absence of IPTG or with non-cognate trigger RNAs, motility is repressed.

FIGs. 24A-C. Simultaneous regulation of gene expression by twelve toehold switches. (FIG. 24A) Schematics of plasmids and ~3.4-kb polycistronic mRNAs used for multiplexing studies. A set of three compatible plasmids are used to each express four different fluorescent reporters. Each reporter has its own switch RNA that can be independently activated by its cognate trigger RNA. (FIG. 24B-C) Percentage of cells expressing each of the four reporters for a set of 24 different trigger RNA combinations. Gray and colored circles are used to identify the particular trigger RNA being expressed by the cell and the corresponding switch RNA. Successful operation is observed for all 12 single-input possibilities and all 2-, 3-, and 4-output color combinations. Output behavior for two non-cognate trigger RNAs is shown in graphs on lower right to demonstrate low system leakage levels.

FIGs. 25A-B. Layered 4-input AND circuit genetic program. (FIG. 25A) Design schematic for the 4-input AND circuit consisting of three 2-input AND gates formed by three toehold switches, two orthogonal transcription factors (ECF41_491 and ECF42_4454), and a GFP reporter. (FIG. 25B) Complete 16-element truth table for the 4-input AND system. GFP expression from the sole logical TRUE output case with all input RNAs are expressed (far right) is significantly higher than the logical FALSE output cases where one or more input RNAs is absent.

FIGs. 26A-B illustrate a system comprising a 4-input OR system controlling translation of the GOI (GFP) and expression data derived therefrom. FIG. 26A illustrates the 4 repressors (switches, hairpins, etc.) denoted G1, G2, G3 and G4. The orientation of these repressors relative to the GOI is G1 – G2 – G3 – G4 – GOI. Each G1, G3 and G4 is controlled by a 2-input AND gate. G2 is controlled by a 3-input AND gate. Thus, as illustrated, repressor G1 requires the presence of input RNAs A1 and B1, repressor G2 requires the presence of input RNAs A2, B2 and C2, repressor G3 requires the presence of input RNAs A3 and B3, and repressor G4 requires the presence of input RNAs A4 and B4. As denoted, the RNA inputs for G1-G4 AND gates are typically different from each other, or at a minimum the combined RNA inputs, including the resultant complex that such inputs form, are different for each of the repressors. FIG. 26B provides the ON/OFF ratio for the G1 repressor in the presence or absence of one or both of its RNA inputs (A1 and B1). The experiments were performed using the methodology described in Example 1.

- 20 -

FIGs. 27A-B illustrate the same system as shown in FIG. 26A and data corresponding to the activation of repressor G2. FIG. 27B provides the ON/OFF ratio for the G2 repressor in the presence or absence of its three RNA inputs (A2, B2 and C2). The experiments were performed using the methodology described in Example 1.

5 FIGs. 28A-B illustrate the same system as shown in FIG. 26A and data corresponding to the activation of repressor G3. FIG. 28B provides the ON/OFF ratio for the G3 repressor in the presence or absence of its two RNA inputs (A3 and B3). The experiments were performed using the methodology described in Example 1.

10 FIGs. 29A-B illustrate the same system as shown in FIG. 26A and data corresponding to the activation of repressor G4. FIG. 29B provides the ON/OFF ratio for the G4 repressor in the presence or absence of its two RNA inputs (A4 and B4). The experiments were performed using the methodology described in Example 1.

15 FIG. 30 is a bar graph showing the ON/OFF ratio for the same system as shown in FIGs. 26A. The bars denoted by arrows indicate the condition in which the necessary AND inputs were present (and thus they denote the highest ratio measured for each repressor), thereby activating the respective repressor and leading to translation of the GOI. The remainder of the bars indicate conditions in which not all of the necessary inputs were present. The experiments were performed using the methodology described in Example 1.

20 FIGs. 31A-B illustrate a 4-input OR system comprising 4 repressors each of which is activatable by a single input trigger and data corresponding to the activation of each repressor. In this case, each input may be referred to as being cognate to its repressor because it is able to hybridize to a sequence of the repressors (in the absence of another trigger, as is the case with AND gates). FIG. 31B provides the ON/OFF ratio in the presence of each of the input triggers, individually, as well as in the presence of non-cognate inputs (controls). The experiments were performed using the methodology described in Example 1.

25 FIGs. 32A-B illustrate a 5-input OR system controlling translation of the GOI (GFP), wherein the trigger for each of the 5 repressors is a complex of 2 or 3 inputs, and data corresponding to the activation of repressors G1 and G5. The 5 repressors (hairpins, switches, etc.) are denoted G1, G2, G3, G4 and G5. The orientation of these repressors relative to the GOI is G1 – G2 – G3 – G4 – G5 – GOI. G1, G3, G4 and G5 is each controlled by a 2-input AND gate. G2 is controlled by a 3-input AND gate. FIG. 32B provides the ON/OFF ratio for the G1 and G5 repressors in the presence or absence of one or both of its

respective AND inputs (A1 and B1 for G1 and A5 and B5 for G5). The experiments were performed using the methodology described in Example 1.

FIG. 33 is a bar graph showing the ON/OFF ratio for the same system as shown in FIG. 32A. The bars denoted by arrows indicate the condition in which the necessary AND inputs were present (and thus typically the highest measured ratio), thereby activating the repressor and leading to translation of the GOI. The remainder of the bars indicate conditions in which not all of the necessary inputs were present. The experiments were performed using the methodology described in Example 1.

FIGs. 34A-B illustrate a 5-input OR system comprising 5 repressors each of which can be activated by a single input trigger and data corresponding to the activation of each of the repressors in the presence of cognate triggers and lack of activation in the presence of non-cognate triggers. FIG. 34B provides the ON/OFF ratio in the individual presence of each of the cognate input triggers as well as in the presence of non-cognate inputs (controls). The experiments were performed using the methodology described in Example 1.

FIG. 35 is a bar graph showing the ON/OFF ratio for the a 6-input OR system. Each of the repressors is controlled by a 2-input AND gate. The orientation of the repressors relative to the GOI is as follows: G1 – G2 – G3 – G4 – G5 – G6 -- GOI. The bars denoted by arrows indicate the condition in which the necessary AND inputs were present (and thus the highest observed ratio per repressor), thereby activating the repressor and leading to translation of the GOI. The remainder of the bars indicate conditions in which not all of the necessary inputs were present. The experiments were performed using the methodology described in Example 1.

FIGs. 36A-B illustrate a 5-input OR system with 2-input AND triggers, resulting in a 10-input system. The 5-input OR repressors are denoted G1, G2, G3, G4 and G5. Each is activated by the presence of its specific 2-input AND triggers. FIG. 36B provides the ON/OFF ratio in the presence of specific AND trigger combinations. The experiments were performed using the methodology described in Example 1.

FIGs. 37A-B illustrate a 4-input OR system with 2-input AND triggers, resulting in an 8-input system. The 4-input OR repressors are denoted G1, G2, G3 and G4. Each is activated by the presence of its specific 2-input AND triggers. FIG. 37B provides the ON/OFF ratio in the presence of specific AND trigger combinations. The experiments were performed using the methodology described in Example 1.

- 22 -

FIGs. 38A-B illustrate a 2-input NAND gate with toehold repressor. FIG. 38A provides a design schematic for the NAND circuit. A two-input AND gate is formed by two triggers (also referred to as half-triggers) with complementary domains s and s^* , which together present the full-length trigger for the toehold repressor. The complex formed by hybridization of the half-triggers A1 and A2 to each other, hybridizes to the switch RNA and causes the switch RNA to refold into a configuration that prevents the ribosome from accessing binding elements on the RNA. FIG. 38B provides the GFP repression fold change for different combinations of trigger RNAs.

DETAILED DESCRIPTION OF INVENTION

The invention provides two general classes of riboregulators: toehold riboregulators and beacon riboregulators. Both can be used to activate protein production (or translation) in various systems including cells such as prokaryotic cells. Unlike previous engineered riboregulators of gene expression, these “devices” can be trans-activated using separate RNAs of virtually arbitrary sequence. The sequence of the activating RNA need not be related to a ribosome binding site (RBS) sequence.

The advantages of these new riboregulators are multifold. First, many riboregulators of the invention can be active in a single cell simultaneously, with each interacting only with its cognate (specific) targets or triggers. This allows simultaneous control over multiple cellular activities. This is illustrated herein in the context of an *E. coli* cell having twelve riboregulators, all of which are acting independently of each other. Second, riboregulators of the invention can be incorporated into complex nucleic acid circuits in vivo with low system cross-talk and high programmability. Third, riboregulators of the invention can trigger protein (e.g., reporter protein) production from endogenous RNAs. When riboregulator output is coupled to a fluorescent protein reporter, these riboregulators act as genetically encodable sensors and imaging probes for endogenous RNAs in cells. For other proteins, such as those involved in cellular metabolism, activation of gene expression using these riboregulators can facilitate the interaction between pathways endogenous to the cell and synthetic gene networks for new applications in biotechnology.

The invention therefore provides a variety of novel riboregulators and “devices” derived therefrom that offer greatly improved diversity, orthogonality, and functionality compared to previously described riboregulators. In contrast to prior art riboregulators that inhibit translation solely by disrupting binding of the ribosome to the RBS, certain

- 23 -

riboregulators of the invention allow ribosome docking (in some cases) but prevent translation initiation by blocking ribosome access to the initiation codon (in all cases) and usually extension from it. A benefit of this approach is that the RBS is no longer required to be part of the trans-RNA sequence enabling new riboregulators to be designed without any
5 dependence on the Shine-Dalgarno sequence and with only few overall sequence constraints. In addition, these new riboregulators do not rely on kissing-loop interactions to drive hybridization between the crRNA and the trans-RNA. Instead, they utilize linear-linear (or large-loop-linear) RNA interactions, whose strength can be rationally controlled simply by changing the number of nucleotides driving the initial RNA-RNA interaction and/or by
10 changing its base composition. In contrast, changes in base composition and/or sequence length in a kissing loop interaction can affect the tertiary structure of interacting domains and decrease the kinetics of the hybridization reaction.

Riboregulators generally

15 Riboregulators are RNA molecules that can be used to repress or activate translation of an open reading frame and thus production of a protein. Repression is achieved through the presence of a regulatory nucleic acid element (the cis-repressive RNA or crRNA) within the 5' untranslated region (5' UTR) of an mRNA molecule. The nucleic acid element forms a hairpin structure comprising a stem domain and a loop domain through complementary base
20 pairing. The hairpin structure blocks access to the mRNA transcript by the ribosome, thereby preventing translation. In some embodiments, including for example embodiments involving prokaryotic cells, the stem domain of the hairpin structure sequesters the ribosome binding site (RBS). In some embodiments, including for example embodiments involving eukaryotic cells, the stem domain of the hairpin structure is positioned upstream of the start (or
25 initiation) codon, within the 5' UTR of an mRNA. RNA expressed and acting in trans (and thus referred to as trans-activating RNA, or taRNA) interacts with the crRNA and alters the hairpin structure. This alteration allows the ribosome to gain access to the region of the transcript upstream of the start codon, thereby releasing the RNA from its repressed state and facilitating protein translation from the transcript. The crRNA are typically engineered RNA
30 molecules. The taRNA may be engineering molecules although in some instances, as described herein, they may be regions of endogenous, naturally occurring RNAs within a system such as a cell.

- 24 -

The invention generally provides nucleic acids, constructs, plasmids, host cells and methods for post-transcriptional regulation of protein expression using RNA molecules to modulate and thus control translation of an open reading frame.

It is to be understood that the invention contemplates modular crRNA encoding nucleic acids and modular taRNA encoding nucleic acids. Modular crRNA encoding nucleic acids as used herein refer to nucleic acid sequences that do not comprise an open reading frame (or coding domain for a gene of interest). Such modular crRNA may be toehold crRNA or beacon crRNA. Thus the invention contemplates riboregulators in their final form (e.g., comprising a coding domain for a gene of interest) or riboregulator components (e.g., a toehold crRNA or a beacon crRNA not operably linked to gene of interest).

The invention further provides oligonucleotides comprising a crRNA sequence and oligonucleotides comprising a taRNA sequence. In addition, the invention provides sets of two or more oligonucleotides. A first set of oligonucleotides includes two or more oligonucleotides whose sequences together comprise a crRNA sequence. The invention also provides a second set of oligonucleotides whose sequences together comprise a taRNA sequence. For ease of cloning, it may be preferable to employ two oligonucleotides each of which includes a single stem-forming portion, in different cloning steps, rather than a single oligonucleotide comprising two stem-forming portions, in order to avoid formation of a stem within the oligonucleotide, which may hinder cloning. The oligonucleotides may be provided in kits with any of the additional components mentioned herein. The oligonucleotides may include restriction sites at one or both ends.

Toehold riboregulators

In a toehold riboregulator system, the interaction between the crRNA and the trans-RNA species is mediated through a single-stranded RNA domain that is located to the 5' end of the crRNA stem. This domain, which is referred to as the toehold domain, provides the trans-RNA with sufficient binding affinity to enable it to unwind the crRNA stem. The degree of complementarity between the trans-RNA and the toehold domain may vary. In some embodiments, it is at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, or 100%. For optimal riboregulator kinetics, the trans-RNA should possess minimal secondary structure and full complementarity (i.e., 100%) to the toehold domain of the crRNA. As used herein, secondary structure refers to non-linear structures including for example hairpin structures, stem loop structures, and the like. Accordingly, it is preferable

- 25 -

that the trans-RNA consists of a sequence with little to no probability of forming secondary structure under the conditions of its use. Those of ordinary skill in the art are able to determine such sequences either manually or through the use of computer programs available in the art.

5 Toehold riboregulator crRNAs do not sequester the RBS within their stem domain. Instead, RBS are confined to the loop domain formed by the repressing stem domain. This allows the region immediately before (upstream or 5') and after (downstream or 3') the initiation codon to be sequestered within the stem domain, thus frustrating translation initiation. The respective lengths of the crRNA toehold, stem, and loop domains can be
10 changed to a large extent without affecting the performance of the toehold riboregulator as will be detailed below. In addition, the crRNA stem domain can retain its repression efficiency even if it contains a number of bulges or mispaired bases, which enables trans-RNAs that do not contain the start codon AUG sequence to trigger the riboregulator. In principle, the tolerance of bulges enables arbitrary tRNA sequences, including endogenous
15 RNAs, to act as input RNAs into the toehold riboregulator, although other criteria such as high secondary structure can affect the response of the regulator.

 An exemplary, non-limiting, class of toehold riboregulators has design parameters shown in FIG. 1. crRNAs of this class possess a toehold domain that is about 12-nucleotides (nts) long and a loop domain that is about 11-nts long and that contains, optionally at its 3'
20 end, an RBS sequence AGAGGAGA. Immediately adjacent to this loop domain is a stem domain comprising a 6-bp duplex spacer region and a 9-bp duplex region flanking a start codon (i.e., AUG). The 9-nts downstream (3') of the start codon were programmed to ensure they did not code for any stop codons since this would lead to early termination of translation. As will be understood based on this disclosure, the trigger RNA is responsible for
25 unwinding this region of the crRNA stem. In addition, the 3-nt region opposite the start codon triad was completely unpaired leading to a crRNA stem domain having a 3-nt long bulge. (This design precludes a trigger RNA from having an AUG sequence at positions programmed to hybridize to this bulge.) To reduce the likelihood that the 9-nt duplex region code for amino acids that affect folding of the gene of interest (GOI), a common 21-nt (7-
30 amino-acid) spacer domain containing a number of low molecular weight residues was inserted between the crRNA stem domain and the coding domain (e.g., the domain coding the GOI or the reporter protein. Thus, in some instances, the toehold switches add 11 residues to

- 26 -

the N-terminus of the regulated protein, which includes the 12-nt translated portion of the stem and the common 21-nt linker region immediately thereafter.

It is to be understood that the embodiment illustrated in FIG. 1 is non-limiting and that other riboregulators of differing lengths and functions are contemplated and

5 encompassed by the invention. Thus, the length of the toehold domain, the stem domain, the loop domain and the linker domain, as well as the duplex regions within the stem domain may differ in length from the embodiment shown in FIG. 1.

It is to be understood that the afore-mentioned conditions imposed on the trigger RNA and output protein can be avoided with a few modifications to the toehold switch design. The
10 sequence constraints on the trigger RNA are a byproduct of the base-pairing conditions specified for the switch RNA stem and the trigger-switch complex. However, these particular secondary structures are not strictly required for switch operation. We have tested multiple high performance switches that have less than a 3-nt bulge at the AUG position in the switch RNA or with an additional base pair at the base of the switch RNA stem. For instance,
15 forward-engineered switch number 5 has a 1-nt bulge in the stem. This switch still provides an ON/OFF value of 453 ± 119 even though the trigger RNA must disrupt two additional base pairs in order to activate the switch. Accordingly, similar design modifications that add and subtract base pairs from the switch RNA will still allow the toehold switches to modulate gene expression while simultaneously providing sufficient design flexibility to eliminate the
20 stop-codon- and AUG-bulge-related constraints on the trigger sequence.

Moreover, the toehold switches can also be modified to incorporate the coding sequence of the output protein directly into the switch RNA stem. Switches of this type would be compatible with any protein sensitive to N-terminal modifications. The specificity of toehold-mediated interactions, redistribution of bulges in the switch stem, and the use of
25 synonymous codons provide sufficient sequence space for these toehold switches to operate with high dynamic range and orthogonality.

Further toehold riboregulator system designs are described in Example 7.

As shown in FIG. 3, toehold riboregulators can display strong trans-activation using a target RNA at the taRNA species, with fluorescence increasing by a factor of over 200 only
30 two to three hours after induction. The same measurements were performed in vivo on an additional 60 toehold riboregulator designs and the on/off ratios are displayed in FIG. 4. Roughly one third of the riboregulators tested increase GFP output by a factor of 50 or more in the presence of their cognate taRNA.

- 27 -

Additional experimental testing has also enabled us to gain a better understanding of the crRNA secondary structure and domain lengths required for optimal toehold riboregulator operation. A toehold domain of at least 5 or 6 nts in length is preferable for taRNA initial binding. The toehold domain can therefore be 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, or more nucleotides in length. Moreover, it was also found that the taRNA need only unwind two-thirds of the crRNA stem in order to allow translation of the GOI. Based on these findings, the stem domain may be as small as 12 bps for adequate repression in the crRNA. The stem domain may however be longer than 12 bps, including 13, 14, 15, 16, 17, 18, 19, 20, or more base pairs in length. Furthermore, expanding the loop length to 12-nts and replacement of the RBS with a slightly stronger version with the canonical Shine-Dalgarno sequence did not decrease the degree of repression by the crRNA. Accordingly, the length of the loop domain may be 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 or more nucleotides. Variations of toehold riboregulators are shown in FIG. 8A and are described in greater detail in Example 7.

The invention further provides crRNA/switches having additional features. In some instances, the top three bases of the hairpin stem may be A-U base pairs. In some instances, the bottom three base pairs of the stem may comprise two strong G-C base pairs and one A-U base pair. In some instances, the length of the switch toehold may range from about 12- to about 15-nts. This latter feature may in some instances strengthen the initial binding between a trigger RNA and its switch RNA. In some instances, the size of the hairpin loop may range from about 11- to about 15-nts to enhance translation of the output protein upon switch activation. In some instances, the loop size is 15-nts. In yet other instances, the cognate trigger may be used that unwinds the first 15 of the 18 bases in the switch stem. In some instances, one or more, including all, of these features may be used simultaneously. The Examples demonstrate the results using such riboregulators.

The toehold riboregulators described herein may be used in logic gates that function through more than one trigger RNA or that sense more than one trigger RNA. FIGs. 12A and 13A illustrate these additional embodiments of the invention. FIG. 12A illustrates a toehold riboregulator comprising a plurality of hairpin structures (i.e., stem-loop structures, crRNAs) connected together in a linear manner, and a downstream GOI coding sequence. In the Figure, the riboregulator comprises 6 hairpin structures and a GOI is GFP. Each hairpin structure is connected to a toehold sequence that is complementary to an input RNA trigger (or taRNA). Each of the input RNA triggers (or taRNA) is capable of activating expression

- 28 -

of the downstream GOI. This riboregulator is referred to as an “OR” gate because it requires the presence of only one of the input RNA (and thus only a single input RNA) in order to observe expression of the GOI. This OR gate activates expression of GFP when any of the input RNA triggers (or taRNAs) is expressed and binds to its corresponding crRNA
5 sequence. The Figure further shows the on/off fluorescence ratio in the presence of individual input RNA triggers A-F or non-cognate RNA triggers W-Z. The on/off ratios are much greater in the presence of the input RNA triggers as compared to the non-cognate RNA triggers.

FIG. 13A illustrates an “AND” gate which comprises a single hairpin (crRNA)
10 structure with an extended stem region. The crRNA encodes a plurality of regions each acting as a binding domain for a taRNA. Input RNA (or taRNAs) hybridize with one another to form a structure that is able to bind to the hairpin, and can also unwind corresponding portions of the crRNA stem. This system only activates when all input RNA triggers are present, in order to form the complex, and thus to completely unwind the crRNA. It is
15 referred to as an AND gate because it requires all of the input RNA in order to observe expression of the GOI. The Figure further provides photographs showing GFP fluorescence in the presence of different combinations of 3, 4, 5 and 6 input RNA triggers.

As described herein, an n-OR system having n-number of switches or repressors
20 (hairpins), where n is greater than 1, is contemplated. Such a system may be referred to as a concatenated system. In general, an (n+1)-OR system has greater noise (or leakiness) than an n-OR system. That is, the greater the number of repressors or switches in the system, in some instances, the weaker the signal to noise ratio. This has been observed for example for a particular series of 4, 5, and 6-OR systems. Such systems can be optimized by first selecting
25 single AND gate configurations that operate well in isolation (e.g., show sufficiently high S/N ratios). These selected AND gates can then be combined to form an n-OR system. When combining such switches to form an OR system, the spacer between successive toehold switches should be of appropriate length, free of secondary structure(s), and should lack stop codons. Spacers can range from 0 to 30 nucleotides. In some embodiments, the OR-systems
30 comprise 9 to 15 nucleotide spacers between repressors. It is to be understood that such spacers are located between the base of one repressor and the initial nucleotide of the toehold domain of the adjacent downstream repressor. Furthermore, the toehold switch with the greatest S/N ratio (when tested individually) may be positioned closest to the GOI. This

- 29 -

should serve to counteract leaky expression in the system. To counteract the dampening of signal transmission, a toehold switch with the widest dynamic range may be positioned the farthest away from the GOI. These design considerations resulted in robust 8-input and 10-input circuits such as those provided herein, including for example those in FIGs. 36A-B and 37A-B.

The toehold switches exhibited low crosstalk with exogenous RNAs, including the coding sequence of the output protein, and endogenous RNAs, even in the absence of initial screening of devices *in silico* to avoid these interactions. If this type of crosstalk were common, a large fraction of the switches would be expected to display significant OFF state leakage. Variations in ON/OFF levels were generally dictated by changes in ON state expression. This insensitivity to non-cognate RNAs can be attributed to two main factors. First, most RNAs are expected to have substantial secondary structure *in vivo*, which reduces the kinetics of association with the switch RNA. Multiple new features can be incorporated into the switches to improve their ability to reliably detect mRNAs and endogenous RNAs. Second, switch RNA activation generally requires disruption of 12 or more base pairs in the switch stem. Such an event is unlikely in the absence of toehold binding to more than 6-nts. Thus, homology over more than 18-nts is required to activate a typical switch RNA. The combined requirements of significant homology and RNA accessibility make activation of the toehold switch by non-cognate RNAs unlikely. Nevertheless, the invention still contemplates in some instances that toehold switch RNA sequences can be screened against the host genome and other exogenous transcripts using BLAST to ensure that unintended interactions with the transcriptome do not occur.

In still other aspects, the invention recognizes that it is useful to prevent a trigger RNA from acting on its cognate switch RNA to prevent activation of a system or as a means of adding another layer of logic to an *in vivo* circuit. Provided herein is a method to reduce or eliminate the activity of a trigger RNA using an RNA referred to herein as a "sink RNA". The sink RNA is designed to outcompete the switch RNA for binding to its cognate trigger strand. In these systems, flanking sequences v^* and u^* are added to the 5' and 3' ends of the trigger RNA, respectively (FIG. 14A). The cognate sink RNA for the trigger is completely complementary to the central b^*-a^* region of the trigger and its flanking domains. Consequently, the thermodynamics of the sink-trigger RNA interaction are much stronger than the interaction between the trigger RNA and its cognate switch, which occurs through the shorter b^*-a^* sequence. This effect leads to preferential binding of the trigger to the sink,

- 30 -

and in the event a trigger RNA is bound to a switch, the v^* and u^* domains will behave as exposed toeholds that the sink RNA can use to complete a branch migration process to drive the trigger off the switch. To make sink-trigger hybridization still more likely, the sink RNA is expressed at a higher level than the trigger RNA. The lengths of the v^* and u^* domains can vary depending on the particular system. Either domain can be completely removed from the system and still retain the desired network behavior as long as the other domain is present. In other words, the sink RNA may comprise one or both flanking domains. The v^* and u^* domains may be 12 to 21 nts long, in some instances.

FIG. 14B displays the behavior of the sink-trigger-switch RNA system in *E. coli* using GFP as a model readout. It will be understood that the invention contemplates other systems in which GFP is replaced with a protein (or gene) or interest. When the switch RNA is expressed on its own, there is low output of the GFP reporter protein. When the trigger and switch are co-expressed, binding occurs and the switch activates strongly leading to an increase in GFP output. However, when all three RNAs are co-expressed GFP output drops ~90% from its fully activated level as a result of preferential sink-trigger RNA binding. Overall, the output protein is expressed only when the trigger RNA is present in the absence of the sink; otherwise, protein output from the device is low. As a result, this system carries out the logical operation A N-IMPLY B where the trigger RNA represents the A input and the sink RNA is the B input. The switch RNA in this case acts as the gate performing the A N-IMPLY B operation and the output is protein regulated by the switch RNA.

This approach can also be directly applied to the toehold repressors discussed below. When a trigger/sink combination is used with a repressor, the system turns off only when the trigger RNA is expressed in the absence of the sink RNA. This behavior is equivalent to an A IMPLY B operation where the trigger serves as the A input and the sink is the B input.

The sink RNA/trigger RNA system can be applied to thresholding circuits. The experiments shown in FIGs. 14A-B employed constant levels of each of the trigger, switch, and sink RNAs. A stoichiometric excess of the sink RNA was also expressed over the trigger RNA to ensure complete elimination of free trigger RNAs from the cell. However, if the levels of both the trigger and sink RNAs are allowed to vary, this system can provide thresholding behavior. For instance, if the expression of the sink RNA is held constant at a medium level but the expression of trigger RNA is varied from low to high levels, the switch RNA will be activated once the trigger RNA concentration exceeds that of the sink RNA concentration (or a particular percentage of the sink RNA concentration subject to variability

- 31 -

in RNA hybridization behavior in the cell or non-cellular environment). Alternatively, if the expression of the trigger RNA is held constant and the expression of the sink RNA is varied, the sink RNA acts as a modulator of trigger RNA activity, tuning protein output from the switch RNA up or down as a function of sink RNA concentration. These behaviors can be used for neural network type behavior (see for example Qian et al. Nature, 475:368-372, 2011), and for constructing majority and minority gates.

The invention therefore contemplates toehold riboregulator compositions (or systems or devices) comprising a switch RNA (comprising a coding sequence for a gene of interest), a trigger RNA, and a sink RNA. In some instances, the trigger RNA is an activating RNA (i.e., its presence, at a sufficient level, activates protein expression (or translation) from the switch RNA and thus of the coding sequence of interest). In some instances, the trigger RNA is a repressing RNA (i.e., its presence, at a sufficient level, represses protein expression (or translation) from the switch RNA and thus of the coding sequence of interest). The inter-related structural features of the switch RNA, trigger RNA and sink RNA are as described herein.

As discussed briefly herein, toehold riboregulators may also function as repressors of protein translation. In accordance with the invention, a new class of riboregulators is provided that can repress translation of a gene of interest in response to a trigger RNA by a novel strand reconfiguration mechanism. These switch RNA/trigger RNA riboregulator systems are referred to herein as toehold repressors as a result of their toehold-based interaction mechanism. The molecular implementation of these RNA devices is shown in FIG. 15A. The toehold repressors consist of two RNAs: a switch RNA that contains the coding sequence(s) of the gene of interest, and a trigger RNA that causes protein translation from the switch to stop. In the illustrated example, the switch RNA contains a 5'-toehold domain that is about 15-nts in length. This toehold is followed by a stem-loop region with a stem that is about 30-nts long and contains a 9-nt loop. The domains b and c that form the stem are about 18- and 12-nts, respectively. The stem contains bulges at three locations 8-, 16-, and 24-nts from the bottom of the stem. These bulges are incorporated to reduce the likelihood of transcriptional termination, but are not required for successful operation. The bulges can also be moved to other locations and increased in number without necessarily preventing successful switch operation. The size of the loop can also be changed without affecting operation. The stem region is followed by a single-stranded region that contains (in the 5' to 3' direction): a 4-nt spacer, the RBS sequence (8-nt in this implementation), a 6-nt

- 32 -

spacer, the start codon AUG, a 9-nt spacer, a 21-nt linker, and then the coding sequence for the gene of interest. As a result of the exposed RBS to start codon region in the switch RNA, expression is turned on in the absence of the trigger RNA.

The trigger RNA is a single-stranded RNA containing a sequence that is perfectly
5 complementary to the early region of the switch RNA as shown in FIG. 15A, and thus it has a total length of 45-nts. When the trigger and switch RNAs are co-expressed, the trigger RNA binds to the toehold domain of the switch RNA and completes a branch migration reaction with the switch stem. Displacement of the stem completely exposes 30-nts and the loop of the switch RNA. These newly exposed bases can rapidly refold. This strand reconfiguration
10 causes the downstream bases of the switch RNA to form a new hairpin domain. This hairpin sequesters the region surrounding the start codon of the gene, repressing in an identical manner to the switch RNA in toehold switch translational activator system. In addition, it is worth noting that the trigger-switch RNA complex formed by the toehold repressors yields a hairpin with an extended toehold that can in turn interact with an activating trigger RNA
15 having the sequence 5'-b*-a*-3' to reactivate translation of the gene/protein of interest. The behavior of this system with separate repressing and activating triggers is equivalent to an A IMPLY B gate, where A is the repressing trigger and B is the activating trigger.

Like the toehold activator switches, toehold repressors can adopt trigger RNAs with virtually arbitrary sequences. Consequently, it is possible to design large repressor libraries
20 with a high degree of orthogonality. In addition, they can be used to trigger translational repression in response to exogenous and endogenous RNAs.

The invention further contemplates and provides higher order logic circuitry based on toehold repressors. Given their similarities to the toehold activator switches, toehold repressor switches can be incorporated into complex logic systems in much the same way as
25 the translational activators.

Thus, some aspects provide NAND logic gates, which are repressor versions of the systems shown in FIGs. 9A, 10A-C and 13A. An example of a NAND logic gate is provided in FIG. 38A. N-bit NAND logic can be carried out using complexes formed by N-input RNA strands that produce a functional trigger RNA. For the simple 2-bit case, two input
30 RNAs are programmed to bind to one another in the same fashion as the tRNA used for the 2-bit AND system. Each of these input RNAs contains only part of the cognate trigger for the switch RNA and thus each is incapable on its own of carrying out the branch migration required to change the state of the switch. However, when both input RNAs bind, they form a

- 33 -

complete trigger RNA sequence and can bind to the switch toehold and unwind its stem to trigger repression of translation. This base concept can be extended to N-bit operation by dividing the complete trigger RNA sequence among multiple input RNAs that bind together in the proper order to provide the trigger sequence. In an alternative approach, two inputs can
 5 be used to each provide roughly half of the trigger sequence. These two inputs are then brought into close proximity through the assembly of N-2 programmed input RNAs.

Other aspects provide NOR logic gates, which are repressor versions of the systems shown in FIGs. 11A and 12A. N-bit NOR logic can be evaluated by using concatenated toehold repressor hairpins positioned upstream of the coding sequence for the protein of
 10 interest. For the simple two-bit NOR case, the NOR gate is composed of a pair of orthogonal toehold repressors upstream of the gene. In the absence of either trigger RNA, the RBS and start codon for both toehold repressors are exposed and available for translation. When only one of the trigger RNAs is expressed, one of the RBS-start codons regions remains free for translation and the ribosome has sufficient processivity to unwind strong hairpins along its
 15 path. Consequently, the 2-bit NOR gate can only turn OFF when both trigger RNAs are expressed and cause strand reconfiguration for both of the toehold repressor domains. These base concepts can be extended to N-bit NOR gate operation.

The riboregulators provided herein can be used in complex logic circuitry. As an example, toehold switches and toehold repressors can be incorporated into higher-order logic
 20 circuits for AND/NAND, OR/NOR, and IMPLY/N-IMPLY operations. The modularity of this computational approach enables even more complex calculations by combining all these operations in a single extended gate RNA containing concatenated toehold regulator hairpins along with a network of affiliated input trigger and sink RNAs. Importantly, the base set of computational elements provided herein enables evaluation of any logic operation by
 25 decomposing it into an expression in disjunctive normal form (i.e., an outer OR operation applied to nested NOT and AND expressions), such as:

(A AND B) OR (C AND D) OR (E AND F AND G),

or with the addition of sink RNAs:

NOT(A AND B) OR (C AND (NOT D)) OR (E AND F AND G).

30 Analogous expressions can be evaluated with the NAND and NOR gates incorporated as well. Computations using the toehold regulators operate in a single computational layer (i.e., they do not require the output from one operation to be used as an input for a later operation) and can readily integrate multiple input species, which increases their computation

- 34 -

speed and enables fewer gates to be used. This is in contrast to other molecular computation techniques such as those described by Qian et al. Science, 332:1196-1201, 2011 and Moon et al. Nature, 491:249-253, 2012.

Still further embodiments provide and apply multiple input XOR and XNOR logic.

5 As an example, N-bit XOR (XNOR) calculations can be performed using a combination of the OR (NOR) gates and trigger/sink RNAs. The main concepts behind this operation can be described using the simple 2-bit XOR case. The constitutively-expressed gate RNA for this operation is a 2-bit OR system containing a pair of concatenated orthogonal toehold switches upstream of the regulated gene. These switches accept cognate triggers A and B. Expression
10 of triggers A and B is controlled by two orthogonal chemical inducers indA and indB, respectively. Each of the triggers has a cognate sink RNA A* and B* that preferentially bind to their corresponding trigger to prevent activation of the switch hairpin in the gate. Importantly, these sink RNAs are expressed from a higher copy plasmid or using a stronger promoter than the trigger RNAs to ensure they reach higher concentrations when induced in
15 the cell. Furthermore, production of sink RNAs A* and B* is tied to indB and indA, respectively. Consequently, addition of indA to the growth media will cause expression of trigger A and sink B*, while addition indB will cause trigger B and sink A* to be produced.

When only one inducer is present, expression of the trigger RNA and a non-cognate sink RNA allows activation of one of the switch hairpins within the gate RNA. However,
20 when both inducers are present, the two trigger RNAs are expressed, but sink RNAs are also transcribed at higher levels. These sink RNAs outcompete the gate RNA for trigger molecules and prevent activation of protein translation. In the case where neither inducer is present, triggers are not expressed and the gate remains off. As a result, this synthetic gene network carries out 2-bit XOR logic.

25 This general approach can be extended to N-bit XOR logic in which each of the N inducers initiates expression of a single trigger RNA along with a complement of N-1 non-cognate sink RNAs. Lastly, N-bit XNOR is evaluated by replacing the N-bit OR gate formed from N concatenated toehold switches with a set of N concatenated toehold repressors.

30 ***Beacon riboregulators***

In a beacon riboregulator system, the crRNA comprise a stem domain of variable length that contains the RBS and, in some cases, the start codon (see FIG. 5 for an exemplary embodiment). The stem domain also includes a ~9 bp region upstream of the RBS containing

- 35 -

nucleotides complementary to the taRNA target. Binding of the taRNA target is initiated through a large (~21-nt) loop domain in the crRNA and proceeds into the 5' portion of the crRNA stem domain. Binding of the taRNA target through this big-loop-linear interaction results in a rigid duplex that provides mechanical force to encourage the rest of the crRNA to
5 unwind. After unwinding, both the RBS and start codon of the activated crRNA are exposed, enabling translation of the GOI. Since the target binding region of the crRNA is independent of both the RBS and start codon, the taRNA of the beacon riboregulator can, in principle, adopt arbitrary sequences. taRNAs having little secondary structure will offer the better reaction kinetics. In addition, the target taRNA must be sufficiently long to force unwinding
10 of the crRNA stem domain.

Beacon riboregulators were tested using identical conditions to those used for the toehold riboregulator devices. FIG. 6 shows the on/off median fluorescence intensity ratios obtained for six beacon riboregulators. Four of the devices show on/off ratios exceeding ten with one design exceeding a factor of 200.

15 Variations of toehold riboregulators are shown in FIG. 8B.

As described herein, the trans-activating RNA (taRNA) (also referred to herein as trigger RNA) may be small RNA molecules encompassing only those sequences that hybridize to the binding domains (first or second or first and second domains) of the toehold or beacon riboregulators, or they may be longer RNA molecules such as mRNA molecules
20 that hybridize to the binding domains of the toehold or beacon riboregulators using only part of their sequence. In still other instances, activation of the crRNA may require two or more RNA or other nucleic acid molecules that work in concert to unwind the hairpin structure of the crRNA. The taRNA may be of varied length. In some instances, the taRNA is about 30 nts in length. Such a taRNA may bind to a crRNA having a 12 nt toehold domain, as
25 described herein including in Example 7.

The crRNA of the invention comprise a hairpin structure that minimally comprises a stem domain and a loop domain. The crRNA and its hairpin typically comprise a single nucleic acid molecule or portion thereof that adopts secondary structure to form (a) a duplex (double helical, partially or fully double-stranded) region (referred to herein as the stem
30 domain) when complementary sequences within the molecule hybridize to each other via base pairing interactions and (b) a single-stranded loop domain at one end of the duplex. FIGs. 1, 5 and 8A show various stem-loop structures. In various embodiments of the invention the stem domain, while predominately double-stranded, may include one or more

- 36 -

mismatches, bulges, or inner loops. The length of a stem domain may be measured from the first pair of complementary nucleotides to the last pair of complementary bases and includes mismatched nucleotides (e.g., pairs other than AT, AU, GC), nucleotides that form a bulge, or nucleotides that form an inner loop.

5 It will be appreciated that although a hairpin is formed from a single nucleic acid molecule, the two regions or sequences of the molecule that form the stem domain may be referred to herein as "strands". Thus the stem may be referred to herein as being partially or fully double-stranded. Nucleic acid sequences within a single molecule that are complementary to each other and are capable of forming a stem domain are said to be "self-complementary" or to "self-hybridizing" or able to "self-hybridize". In general, the hairpin and stem domains described herein form at and are stable under physiological conditions, e.g., conditions present within a cell (e.g., conditions such as pH, temperature, and salt concentration that approximate physiological conditions). Such conditions include a pH between 6.8 and 7.6, more preferably approximately 7.4. Typical temperatures are approximately 37°C, although prokaryotes and some eukaryotic cells such as fungal cells can grow at a wider temperature range including at temperatures below or above 37°C.

 Various of the nucleic acids of the invention may be referred to herein as non-naturally occurring, artificial, engineered or synthetic. This means that the nucleic acid is not found naturally or in naturally occurring, unmanipulated, sources. A non-naturally occurring, artificial, engineered or synthetic nucleic acid may be similar in sequence to a naturally occurring nucleic acid but may contain at least one artificially created insertion, deletion, inversion, or substitution relative to the sequence found in its naturally occurring counterpart. A cell that contains an engineered nucleic acid may be referred to as an engineered cell.

 Various embodiments of the invention involve nucleic acid sequences that are complementary to each other. In some instances, the sequences are preferably fully complementary (i.e., 100% complementary). In other instances, however the sequences are only partially complementary. Partially complementary sequences may be at least 60%, at least 70%, at least 80%, at least 90%, or at least 95% complementary. Sequences that are only partially complementary, when hybridized to each other, will comprise double-stranded regions and single-stranded regions. The single-stranded regions may be single mismatches, loops (where for instances a series of consecutive nucleotides on one strand are unhybridized), bulges (where for instances a series of consecutive nucleotides on both strands, opposite to each other, are unhybridized). It will be appreciated that

- 37 -

complementarity may be determined with respect to the entire length of the two sequences or with respect to portions of the sequences.

Nucleic acids and/or other moieties of the invention may be isolated. As used herein, "isolated" means separate from at least some of the components with which it is usually
5 associated whether it be from a naturally occurring source or made synthetically.

Nucleic acids and/or other moieties of the invention may be purified. As used herein, purified means separate from the majority of other compounds or entities. A compound or moiety may be partially purified or substantially purified. Purity may be denoted by a weight by weight measure and may be determined using a variety of analytical techniques such as
10 but not limited to mass spectrometry, HPLC, etc.

Nucleic acids generally refer to polymers comprising nucleotides or nucleotide analogs joined together through backbone linkages such as but not limited to phosphodiester bonds. Nucleic acids include deoxyribonucleic acids (DNA) and ribonucleic acids (RNA) such as messenger RNA (mRNA), transfer RNA (tRNA), etc. Nucleic acids may be single-
15 stranded, double-stranded, and also tripled-stranded.

A naturally occurring nucleotide consists of a nucleoside, i.e., a nitrogenous base linked to a pentose sugar, and one or more phosphate groups which is usually esterified at the hydroxyl group attached to C-5 of the pentose sugar (indicated as 5') of the nucleoside. Such compounds are called nucleoside 5'-phosphates or 5'-nucleotides. In DNA the pentose sugar
20 is deoxyribose, whereas in RNA the pentose sugar is ribose. The nitrogenous base can be a purine such as adenine or guanine (found in DNA and RNA), or a pyrimidine such as cytosine (found in DNA and RNA), thymine (found in DNA) or uracil (found in RNA). Thus, the major nucleotides of DNA are deoxyadenosine 5'-triphosphate (dATP), deoxyguanosine 5'-triphosphate (dGTP), deoxycytidine 5'-triphosphate (dCTP), and deoxythymidine 5'-
25 triphosphate (dTTP). The major nucleotides of RNA are adenosine 5'-triphosphate (ATP), guanosine 5'-triphosphate (GTP), cytidine 5'-triphosphate (CTP) and uridine 5'-triphosphate (UTP). In general, stable base pairing interactions occur between adenine and thymine (AT), adenine and uracil (AU), and guanine and cytosine (GC). Thus adenine and thymidine, adenine and uracil, and guanine and cytosine (and the corresponding nucleosides and
30 nucleotides) are referred to as being complementary to each other.

In general, one end of a nucleic acid has a 5'-hydroxyl group and the other end of the nucleic acid has a 3'-hydroxyl group. As a result, the nucleic acid has polarity. The position or location of a sequence or moiety or domain in a nucleic acid may be denoted as being

- 38 -

upstream or 5' of a particular marker, intending that it is between the marker and the 5' end of the nucleic acid. Similarly, the position or location of a sequence or moiety or domain in a nucleic acid may be denoted as being downstream or 3' of a particular marker, intending that it is between the marker and the 3' end of the nucleic acid.

5 Nucleic acids may comprise nucleotide analogs including non-naturally occurring nucleotide analogs. Such analogs include nucleoside analogs (e.g., 2-aminoadenosine, 2-thiothymidine, inosine, 3-methyl adenosine, C5-propynylcytidine, C5-propynyluridine, C5-bromouridine, C5-fluorouridine, C5-iodouridine, C5-methylcytidine, 7-deazaadenosine, 7-deazaguanosine, 8-oxoadenosine, 8-oxoguanosine, O(6)-methylguanine, and 2-thiocytidine),
10 chemically modified bases, biologically modified bases (e.g., methylated bases), intercalated bases, modified sugars (e.g., 2'-fluororibose, ribose, 2'-deoxyribose, arabinose, and hexose), or modified phosphate groups (e.g., phosphorothioates and 5'-N-phosphoramidite linkages).

The nucleic acids of the invention, including the crRNA and taRNA, may be provided or present in a larger nucleic acid. The larger nucleic acid may be responsible for the
15 transcription and thus production of the crRNA and taRNA, as described in Example 1, for example. The larger nucleic acid may comprise a nucleotide sequence that is transcribed to produce the crRNA and taRNA of the invention. For convenience, the invention may refer to the larger nucleic acid as comprising the crRNA and/or taRNA although it is to be understood that in practice this intends that the larger nucleic acid comprises a sequence that encodes the
20 crRNA and/or taRNA. Such encoding sequences may be operably linked to other sequences in the larger nucleic acid such as but not limited to origins of replication. As used herein, "operably linked" refers to a relationship between two nucleic acid sequences wherein the production or expression of one of the nucleic acid sequences is controlled by, regulated by, modulated by, etc., the other nucleic acid sequence. For example, the transcription of a
25 nucleic acid sequence is directed by an operably linked promoter sequence; post-transcriptional processing of a nucleic acid is directed by an operably linked processing sequence; the translation of a nucleic acid sequence is directed by an operably linked translational regulatory sequence; the transport or localization of a nucleic acid or polypeptide is directed by an operably linked transport or localization sequence; and the post-
30 translational processing of a polypeptide is directed by an operably linked processing sequence. Preferably a nucleic acid sequence that is operably linked to a second nucleic acid sequence is covalently linked, either directly or indirectly, to such a sequence, although any effective association is acceptable.

- 39 -

As used herein, a regulatory sequence or element intends a region of nucleic acid sequence that directs, enhances, or inhibits the expression (e.g., transcription, translation, processing, etc.) of sequence(s) with which it is operatively linked. The term includes promoters, enhancers and other transcriptional and/or translational control elements. The crRNA and taRNA moieties of the invention may be considered to be regulatory sequences or elements to the extent they control translation of a gene of interest that is operably linked to the crRNA. The invention contemplates that the crRNA and taRNA of the invention may direct constitutive or inducible protein expression. Inducible protein expression may be controlled in a temporal or developmental manner.

The term vector refers to a nucleic acid capable of mediating entry of, e.g., transferring, transporting, etc., a second nucleic acid molecule into a cell. The transferred nucleic acid is generally linked to, e.g., inserted into, the vector nucleic acid. A vector may include sequences that direct autonomous replication, or may include sequences sufficient to allow integration into host cell DNA. Useful vectors include, for example, plasmids (typically DNA molecules although RNA plasmids are also known), cosmids, and viral vectors.

In the context of the invention, reporter proteins are typically used to visualize activation of the crRNA. Reporter proteins suitable for this purpose include but are not limited to fluorescent or chemiluminescent reporters (e.g., GFP variants, luciferase, e.g., luciferase derived from the firefly (*Photinus pyralis*) or the sea pansy (*Renilla reniformis*) and mutants thereof), enzymatic reporters (e.g., β -galactosidase, alkaline phosphatase, DHFR, CAT), etc. The eGFPs are a class of proteins that has various substitutions (e.g., Thr, Ala, Gly) of the serine at position 65 (Ser65). The blue fluorescent proteins (BFP) have a mutation at position 66 (Tyr to His mutation) which alters emission and excitation properties. This Y66H mutation in BFP causes the spectra to be blue-shifted compared to the wtGFP. Cyan fluorescent proteins (CFP) have a Y66W mutation with excitation and emission spectra wavelengths between those of BFP and eGFP. Sapphire is a mutant with the suppressed excitation peak at 495 nM but still retaining an excitation peak at 395 and the emission peak at 511 nM. Yellow FP (YFP) mutants have an aromatic amino acid (e.g. Phe, Tyr, etc.) at position 203 and have red-shifted emission and excitation spectra.

It is to be understood that although various embodiments of the invention are described in the context of RNA, the nucleic acids of the invention can be RNA or DNA. In general, RNA and DNA can be produced using in vitro systems, within cells, or by chemical synthesis using methods known in the art. It will be appreciated that insertion of crRNA

- 40 -

elements upstream of an open reading frame (ORF) can be accomplished by modifying a nucleic acid comprising the ORF.

The invention provides DNA templates for transcription of a crRNA or taRNA. The invention also provides DNA constructs and plasmids comprising such DNA templates. In certain embodiments, the invention provides a construct comprising the template for transcription of a crRNA or a taRNA operably linked to a promoter.

In certain embodiments, the invention provides a DNA construct comprising (i) a template for transcription of a crRNA; and (ii) a promoter located upstream of the template. In certain embodiments, a construct or plasmid of the invention includes a restriction site downstream of the 3' end of the portion of the construct that serves as a template for the crRNA, to allow insertion of an ORF of choice. The construct may include part or all of a polylinker or multiple cloning site downstream of the portion that serves as a template for the crRNA. The construct may also include an ORF downstream of the crRNA.

In certain embodiments, the invention provides a DNA construct comprising (i) a template for transcription of a taRNA; and (ii) a promoter located upstream of the template. The invention further provides a DNA construct comprising: (i) a template for transcription of a crRNA; (ii) a promoter located upstream of the template for transcription of the crRNA; (iii) a template for transcription of a taRNA; and (iv) a promoter located upstream of the template for transcription of the taRNA. The promoters may be the same or different.

The constructs may be incorporated into plasmids, e.g., plasmids capable of replicating in bacteria. In certain embodiments, the plasmid is a high copy number plasmid (e.g., a pUC-based or pBR322-based plasmid), while in other embodiments, the plasmid is a low or medium copy number plasmid, as these terms are understood and known in the art. The plasmid may include any of a variety of origins of replication, which may provide different copy numbers. For example, any of the following may be used (copy numbers are listed in parenthesis): ColEI (50-70 (high)), p15A (20-30 (medium)), pSC101 (10-12 (low)), pSOO1* (< 4 (lowest)). It may be desirable to use plasmids with different copy numbers for transcription of the crRNA and the taRNA in order to alter their relative amounts in a cell or system. In addition, in certain embodiments a tunable copy number plasmid is employed.

The invention further provides viruses and cells comprising the nucleic acids, constructs (such as DNA constructs), and plasmids described above. In various embodiments, the cell is a prokaryotic cell. In various embodiments, the cell is a eukaryotic cell (e.g., a fungal cell, mammalian cell, insect cell, plant cell, etc.). The nucleic acids or constructs may

- 41 -

be integrated into a viral genome using recombinant nucleic acid technology, and infectious virus particles comprising the nucleic acid molecules and/or templates for their transcription can be produced. The nucleic acid molecules, DNA constructs, plasmids, or viruses may be introduced into cells using any of a variety of methods known in the art, e.g., electroporation, calcium-phosphate mediated transfection, viral infection, etc.

As discussed herein, the nucleic acid constructs can be integrated into the genome of a cell. Such cells may be present in vitro (e.g., in culture) or in vivo (e.g., in an organism). The cells may be eukaryotic or prokaryotic cells, including but not limited to mammalian cells and bacterial cells. An example of a bacterial cell is an *E. coli* bacterium. An example of a mammalian cell is a human cell or a mouse cell. The invention further provides transgenic plants and non-human transgenic animals comprising the nucleic acids, DNA constructs, and/or plasmids of the invention. Methods for generating such transgenic organisms are known in the art.

The invention further provides a variety of kits. For example, the invention provides a kit comprising a plasmid, wherein a first plasmid comprises (i) a template for transcription of a crRNA, and (ii) a promoter located upstream of the template for transcription of the crRNA element, and optionally a second plasmid that comprises (i) a template for transcription of a cognate (complementary) tRNA element, and (ii) a promoter located upstream of the template for transcription of the tRNA element. The promoters may be the same or, preferably, different. One or more of the promoters may be inducible. The plasmids may have the same or different copy numbers. The invention further provides a kit comprising a single plasmid that comprises a template for transcription of a crRNA element and a promoter located upstream of the template for transcription of the crRNA element and further comprises a template for transcription of a cognate tRNA element and a promoter located upstream of the template for transcription of the cognate tRNA element. In certain embodiments, the plasmids comprise one or more restriction sites upstream or downstream of the template for transcription of the crRNA element. If downstream, the restriction sites may be used for insertion of an open reading frame of choice. The kits may further include one or more of the following components: (i) one or more inducers; (ii) host cells (e.g., prokaryotic or eukaryotic host cells); (iii) one or more buffers; (iv) one or more enzymes, e.g., a restriction enzyme; (v) nucleic acid isolation and/or purification reagents; (vi) a control plasmid lacking a crRNA or tRNA sequence; (vii) a control plasmid containing a crRNA or

- 42 -

taRNA sequence or both; (viii) sequencing primers; (ix) instructions for use. The control plasmids may comprise a reporter sequence.

The riboregulators of the invention in some instances comprise a consensus prokaryotic RBS. However, in various embodiments of the invention any of a variety of

5 alternative sequences may be used as the RBS. The sequences of a large number of bacterial ribosome binding sites have been determined, and the important features of these sequences are known. Preferred RBS sequences for high level translation contain a G-rich region at positions -6 to -11 with respect to the AUG and typically contain an A at position -3.

Exemplary RBS sequences for use in the present invention include, but are not limited to,

10 AGAGGAGA (or subsequences of this sequence, e.g., subsequences at least 6 nucleotides in length, such as AGGAGG). Shorter sequences are also acceptable, e.g., AGGA, AGGGAG, GAGGAG, etc. Numerous synthetic ribosome binding sites have been created, and their translation initiation activity has been tested. In various embodiments any naturally

occurring RBS may be used in the crRNA constructs. The activity of any candidate sequence
15 to function as an RBS may be tested using any suitable method. For example, expression may be measured as described in Example 1 of published PCT application WO 2004/046321, or as described in reference 53 of that published PCT application, e.g., by measuring the activity of a reporter protein encoded by an mRNA that contains the candidate RBS

appropriately positioned upstream of the AUG. Preferably an RBS sequence for use in the
20 invention supports translation at a level of at least 10% of the level at which the consensus RBS supports translation (e.g., as measured by the activity of a reporter protein). For

example, if the candidate RBS is inserted into a control plasmid in place of the consensus RBS, the measured fluorescence will be at least 10% of that measured using the consensus

RBS. In certain embodiments, an RBS that supports translation at a level of at least 1%, 5%,

25 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, or more relative to the level at which the consensus RBS supports translation is used. In certain embodiments of the invention an RBS that supports translation at higher levels than the consensus RBS is used.

Further general teachings relating to riboregulators are found in published PCT application WO 2004/046321, the entire contents of which are incorporated by reference
30 herein.

Advantages of toehold and beacon riboregulators

Riboregulators of the invention offer a number of benefits compared to existing techniques. For instance, quantitative real-time PCR (qRT-PCR) offers highly sensitive detection of RNA levels, northern blots exhibit high specificity, and microarrays enable simultaneous detection of thousands of targets. However, in all these techniques, cells must be sacrificed to obtain the RNA for quantitation and thus it is challenging to measure RNA levels in real time. Fluorescence in situ hybridization (FISH) and the use of fluorescent RNA aptamers enable visualization of RNA localization inside cells. FISH requires cells to be fixed for visualization and hybridization takes a number of hours using expensive probes. RNA aptamers can be used to image RNA in living cells; however, those aptamers with the highest fluorescence intensity still require copy numbers far exceeding those of endogenous RNAs in order to be detected in most optical microscopes. RNA levels can also be measured using a fluorescent reporter protein driven from the same promoter as the RNA target. The reporter in this method can reflect the level of RNA target, yet it cannot recapitulate regulatory behavior from chromosomal regions distant (e.g. multiple kilobases) from the promoter region. Furthermore, the presence of additional copies of the promoter can titrate RNA polymerase activity away from the target gene. Lastly, RNAs tagged with protein binding aptamers have also been used to measure localization and levels of RNAs inside cells using fusions of the binding protein with fluorescent protein reporters. This technique, however, requires chromosomal modifications to either tag or knockout the gene corresponding to the RNA to be visualized. The riboregulators of the invention are not encumbered by these various limitations of the prior art techniques.

EXAMPLES***Example 1. Toehold riboregulator and reporter protein/GOI***

An exemplary riboregulator of FIG. 1 was tested experimentally. The GOI was an EGFP variant GFPmut3b, which was tagged with an ASV degradation signal to set its half-life to approximately 110 minutes. taRNAs cognate to the crRNA were designed using the software package NUPACK to have minimal secondary structure and perfect complementarity to the 30-nt long target binding site of the crRNA.

The riboregulator was tested in *E. coli* BL21 DE3 star, an RNase E deficient strain that contained a lambda phage lysogen bearing T7 RNA polymerase under the control of the IPTG inducible lacUV5 promoter. crRNA and taRNA constructs were expressed from

- 44 -

separate plasmids to enable rapid characterization of the interaction of the crRNA with cognate and non-cognate taRNA sequences. For both the crRNA and the taRNA, transcription was initiated from an upstream T7 promoter and transcription terminated using a T7 RNA polymerase termination signal. The crRNA-GFP transcripts were generated from a plasmid with a medium copy number colA origin, while the taRNAs transcripts were generated from a higher copy number plasmid with a colE1 origin. These variations in plasmid copy number led to an estimated 7-fold excess of taRNA compared to crRNAs inside fully-induced cells. This ratio is similar to previous studies and typical copy number differences observed for anti-sense RNAs and their targets.

In vivo testing was performed in *E. coli* transformed with either a crRNA and its cognate taRNA target (ON state strains) or a crRNA and a non-cognate taRNA (OFF state strains) and grown overnight in 1 mL of selective LB media at 37°C in deep well 96-well plates covered with a gas permeable seal. Transformation of *E. coli* with two plasmids in both ON and OFF state riboregulator conditions ensured that both strains were subject to similar metabolic loads, at least with respect to the number of exogenous RNAs that were being transcribed. Overnight cultures were diluted 100-fold and grown up for 80 minutes at 37°C in the deep well plates. The early log phase cells were then induced with 0.1 mM of IPTG with aliquots taken at 1 hour time points for characterization via flow cytometry. For comparison of GFP fluorescence intensity between samples, the mode GFP intensity was calculated from fluorescence intensity histograms generate from flow cytometry data.

As a first measure of riboregulator performance, the fluorescence intensity of the crRNA-GFP constructs was compared to the fluorescence from a non-cis-repressed GFP construct induced at the same level of IPTG in the same BL21 DE3 star *E. coli* strain. These measurements demonstrated extremely high levels of translational repression with six tested riboregulator crRNAs reducing fluorescence output by 99.5% or more (see FIG. 2). To calculate the effectiveness of the trans-activation of the riboregulator, mode GFP fluorescence of the crRNA-GFP in the presence of its cognate taRNA was compared to the fluorescence of the crRNA-GFP in the presence of a non-cognate taRNA. By dividing these two numbers, the on/off ratio was calculated for all the riboregulators tested. FIG. 3 presents this on/off ratio taken at one hour time points for a high performance toehold riboregulator. Error bars are the standard deviation in the on/off ratio calculated from three biological replicates. From these data, it is clear that the toehold riboregulator can display strong trans-

activation by a target RNA, with fluorescence increasing by a factor of over 200 only two to three hours after induction.

The same measurements were performed in vivo on an additional 60 toehold riboregulator designs and the on/off ratios are displayed in FIG. 4. Roughly one third of the riboregulators tested increase GFP output by a factor of 50 or more in the presence of their cognate target.

Example 2. Beacon riboregulators

Beacon riboregulators, such as those having a structure shown in FIG. 5, were tested using identical conditions to those used for the toehold riboregulators. FIG. 6 shows the on/off median fluorescence intensity ratios obtained for six beacon riboregulators. Four of the devices show on/off ratios exceeding ten with one design exceeding a factor of 200.

Example 3. Endogenous RNA sensing

The novel riboregulators described herein can be used for the detection of endogenous RNAs. As a proof of concept, a beacon riboregulator was designed and generated that could be triggered by the small RNA ryhB in *E. coli*. RyhB is a 90-nt long non-coding RNA that is upregulated when iron levels are low in *E. coli*. This RNA can be induced through the addition of the iron chelator 2,2'-dipyridyl to the culture medium.

To test this endogenous sensor, a plasmid was constructed that contained the beacon riboregulator upstream of a GFP reporter. Expression of the crRNA transcript was controlled using the IPTG-inducible P_{llacO-1} promoter. MG1655 *E. coli* cells transformed with the riboregulator sensor plasmid were induced with 1 mM IPTG in early log phase. At the same time, ryhB expression was induced through the addition of the iron chelator. Flow cytometry measurements taken from cells harvested after 2 hours demonstrated a five-fold increase in GFP fluorescence intensity for the ryhB containing cells compared to a control population that was not induced with 2,2'-dipyridyl (FIG. 7). In addition, control cells containing a non-cis-repressed GFP reporter under the P_{llacO-1} promoter exhibited a decrease in fluorescence intensity when induced with both IPTG and 2,2'-dipyridyl compared to those induced with IPTG alone. This additional control demonstrates that GFP output from the sensor was not caused by an increase in transcription levels caused by the addition of the iron chelator.

Example 4. Genetic encoding of complex OR logic operations using riboregulators

We have used members of the riboregulator library to successfully carry out multiple logical OR operations in vivo. The simplest OR operation involves two inputs, A and B, that activate a logic gate if either of the inputs is present. We implemented this system in vivo simply by taking two high performance riboregulators and placing them one after the other along the same mRNA upstream of the coding sequence for GFP (FIG. 11A). The intended operation of this gate in vivo is shown in FIG. 11B. When either input RNA molecule is present in the cell, it will bind to its corresponding crRNA module and de-repress the module by unwinding its stem. Since a ribosome engaged in protein translation has strong RNA helicase activity, it can unwind downstream crRNAs in its path, continuing translation unimpeded. Flow cytometry of the 2-input OR gate revealed strong activation of GFP expression when either programmed taRNA input was transcribed (FIG. 11C). In addition, parallel experiments in which the positions of the crRNAs were interchanged showed similar system performance.

Motivated by the successful implementation of the 2-input gate, we pursued a 6-input OR logic system featuring six crRNA modules placed upstream of GFP (FIG. 12A). Since three of the parent crRNAs in the OR logic system contained stop codons, we modified their sequences to eliminate these unwanted codons and tested them individually to ensure the stop-codon-free variants retained the activities of their parents. Following these tests, the 474-bp six-crRNA construct was synthesized using gene assembly and transformed into *E. coli* along with plasmids expressing different taRNA elements. Cells expressing both the 6-input OR mRNA and one of the cognate taRNAs exhibit strong GFP fluorescence when measured on plates containing the inducer IPTG. A set of 4 non-cognate taRNAs, however, did not activate significant expression of GFP, highlighting the impressive orthogonality of our in vivo logic framework. Flow cytometry measurements from these transformants also confirmed successful OR gate operation, with all six inputs providing at least 5-fold higher GFP output compared to the set of 4 non-cognate taRNAs.

Example 5. Genetic encoding of complex AND logic operations using riboregulators

We have developed generalizable systems for carrying out AND logic operations using toehold riboregulators. FIG. 9A depicts a two input AND gate that features a crRNA sequence upstream of a GFP reporter sequence. The two inputs in the system are two RNA sequences A and B that contain one half of the cognate taRNA sequence of the crRNA gate

- 47 -

(FIG. 9A). The two input RNAs also possess a hybridization domain (u-u*) that enables both RNAs to bind to one another when they are present inside the cell. When this hybridization event occurs, the two halves of the taRNA are brought into close proximity providing a sequence capable of unwinding the gate crRNA to trigger translation of GFP. Each of the input RNAs when expressed on its own is unable to derepress that crRNA since they are either: (1) unable to unwind a long enough region of the crRNA stem, which is the case for input B, or (2) they are kinetically and thermodynamically disfavored from binding to the crRNA, which is the case for input A. Flow cytometry measurements for the 2-input AND logic system validate its operation in *E. coli* (FIG. 9B). GFP output is activated only if all three RNAs in the system are expressed inside the cell, while it is low in all other cases. FIGS. 10A-C illustrate 3-input AND systems (FIGS. 10A-B) and 4-input AND systems (FIG. 10C). 4-input systems (or larger systems) may involve constructs that carry more than one input. For example, in a 4-input AND system, constructs such as plasmids encoding 2 or more inputs may be used, rather than for example using a separate construct for each input.

AND gates based on toehold riboregulators have been successfully extended to 6-bit operation. As shown in FIG. 13A, the gate in this system consists of a hairpin with an extended stem consisting of the stem sequences of six validated toehold riboregulator crRNAs and a toehold sequence from the bottommost crRNA. The input RNAs thus contain the corresponding taRNA sequences and also possess hybridization sequences to their neighboring input strands. The hybridization sequence of a given input is complementary to the toehold binding domain of the next input RNA. For example, input A contains a 12- to 15-nt sequence to which input B binds, and this sequence is the toehold for the cognate crRNA of input B. Consequently, binding of input A to the gate unwinds the bottom bases of its stem, and also provides a new toehold for binding of input B. This stem unwinding/toehold presentation process repeats until all inputs are bound to the gate. Upon binding of all inputs, the RBS and start codon are de-repressed thereby triggering production of GFP or another protein of interest. We have validated this gate by expressing different combinations of the input RNAs in *E. coli* also expressing the gate mRNA. FIG. 13B shows GFP intensities measured from colonies induced on LB plates. Strong GFP fluorescence is only visible when all six inputs are expressed in the cell. GFP expression is low in the six other input combinations, including stringent tests where all but one of the input RNAs is expressed.

Example 6. Riboregulator toehold repressors

We constructed a library of 44 toehold repressors (devices/systems) and tested their function in *E. coli* BL21 Star DE3. We used flow cytometry to test the performance of the systems, calculating the mode GFP fluorescence from the switch in its OFF state (i.e., in the presence of its cognate trigger); and in its ON state (i.e., in the absence of its cognate trigger). We then calculated percent repression levels using the equation:

$$\% \text{ repression} = 1 - [\text{OFF state mode fluorescence} \div \text{ON state mode fluorescence}].$$

FIG. 15B displays the % repression levels obtained for the 44 repressors in the library. Repressors 40 to 44 have highly variable performance. We postulate that their behavior is due to instability of the folding of the switch RNA, which causes fluctuations between the ON and the OFF state configurations of the switch RNA even when the trigger RNA is not present. The rest of the devices/systems perform quite well on average. 73% of the total library has repression levels of at least 80%. Moreover, 50% of the library exhibits repression of greater than 90%. This impressive 90% repression level exceeds the performance of almost all previously reported translational repressors (see Mutalik et al., Nat. Chem. Biol. 8:447-454, 2012). Additional measurements also demonstrated that the highest performance toehold switches could achieve translational repression greater than 95% within 1 hour and increase the level to above 99% in subsequent time points (FIG. 16).

Example 7. Toehold switches

As described herein, the invention provides a new class of post-transcriptional riboregulators of gene expression in called toehold switches that have no known natural counterparts. Toehold switches activate expression of a regulated gene in response to a trans-acting trigger RNA. Their operation in living cells is facilitated by two novel mechanisms: toehold-based linear-linear RNA interactions pioneered in in vitro studies and efficient translational repression via base pairing in regions surrounding the initiation codon. We demonstrate that toehold switches routinely enable modulation of protein expression by over 100 fold, with the best switches rivaling the dynamic range of protein-based regulators. We validate large sets of orthogonal components, including a library of 18 toehold switches exhibiting system cross talk levels below 2%, which constitutes the largest and most stringent family of orthogonal regulatory elements, protein or RNA based, ever reported. We then

- 49 -

forward engineered a set of 13 toehold switches with an average on/off fluorescence ratio of 406. We further applied thermodynamic analyses to predict variations in system performance. Furthermore, we demonstrate a set of toehold switches that are capable of effective triggering from functional mRNA molecules. The high dynamic range, orthogonality, programmability, and versatility of these toehold switches suggest they will be powerful new tools for synthetic biology.

Methods

Strains, plasmids, and growth conditions. The following *E. coli* strains were used in this study: BL21 Star DE3 ($F^- ompT hsdS_B (r_B^- m_B^-) gal dcm rne131$ (DE3); Invitrogen), BL21 DE3 ($F^- ompT hsdS_B (r_B^- m_B^-) gal dcm$ (DE3); Invitrogen), MG1655Pro ($F^- \lambda^- ilvG^- rfb-50 rph-1 Sp^R lacR tetR$), and DH5 α ($endA1 recA1 gyrA96 thi-1 glnV44 relA1 hsdR17(r_K^- m_K^+) \lambda^-$). All strains were grown in LB medium with appropriate antibiotics. Antibiotics were used at the following concentrations: ampicillin ($50 \mu\text{g mL}^{-1}$), kanamycin ($30 \mu\text{g mL}^{-1}$), and chloramphenicol ($34 \mu\text{g mL}^{-1}$).

To characterize the toehold switches, chemically competent *E. coli* were transformed with the desired combination of toehold switch and trigger plasmids, and spread onto LB/agar plates containing the appropriate pair of antibiotics. For colony GFP fluorescence measurements, LB/agar plates were supplemented with 0.1 mM isopropyl β -D-1-thiogalactopyranoside (IPTG) to induce RNA expression. For flow cytometry measurements, LB medium containing antibiotics was inoculated with cells picked from individual colonies and incubated overnight with shaking at 37°C. Cells were then diluted 100-fold into fresh selective LB medium and returned to shaking at 37°C in 96-well plates. For T7 RNA polymerase driven expression in BL21 Star DE3 and BL21 DE3, cells were induced with 0.1 mM IPTG at 0.2-0.3 OD₆₀₀ after 80 minutes of growth. Unless otherwise noted, measurements on cell cultures were taken 3 hours after addition of IPTG. For expression using the constitutive PN25 promoter, overnight cultures were diluted 100-fold into selective LB media. The time of this dilution was defined as $t = 0$ for subsequent measurements.

Plasmid construction. All DNA oligonucleotides were purchased from Integrated DNA Technologies, Inc. Double-stranded trigger and switch DNA was produced from either single > 100-nt oligonucleotides amplified using universal primers or using gene assembly from short < 50-nt oligonucleotides segmented using gene2oligo (Rouillard et al., Nucleic Acids Res 32:W176-180, 2004). These PCR products were then inserted into vector

- 50 -

backbones using Gibson assembly with 30-bp overlap regions (Gibson et al., Nat. Methods 6:343-345, 2009). Vector backbones were PCR amplified using the universal backbone primers and digested prior to assembly using DpnI (New England Biolabs, Inc.). Backbones were generated from the T7-based expression plasmids pET15b, pCOLADuet, and pACYCDuet (EMD Millipore). pET15b, pCOLADuet, and pACYCDuet plasmids all contain a constitutively expressed lacI gene, a T7 RNA polymerase promoter and terminator pair, and the following respective resistance markers/replication origins: ampicillin/ColE1, kanamycin/ColA, and chloramphenicol/P15A. All trigger RNAs presented herein were expressed using pET15b backbones, and the switch mRNAs were expressed using either pCOLADuet or pACYCDuet backbones. Reverse primers for the backbones were designed to bind to the region upstream of the T7 promoter. Forward primers for trigger backbones amplified from the beginning of the T7 promoter. Forward primers for the switch backbones were designed to prime off the 5' end of either GFPmut3b-ASV or mCherry and add a 30-nt sequence containing the linker for Gibson assembly. Constructs were cloned inside DH5 α and sequenced to ensure all toehold switch components were synthesized correctly. All transformations were performed using established chemical transformation protocols (Inoue et al., Gene, 96:23-28, 1990).

Flow cytometry measurements and analysis. Flow cytometry was performed using a BD LSRFortessa cell analyzer equipped with a high throughput sampler. GFP fluorescence intensities were measured using 488 nm excitation laser and a 530/30 nm filter. mCherry fluorescence intensities were measured using a 561 nm laser and a 610/20 nm emission filter. In a typical experiment, cells were diluted by a factor of ~65 into phosphate buffered saline (PBS) and sampled from 96-well plates. Forward scatter (FSC) was used for trigger and ~30,000 individual cells analyzed.

Error levels for the fluorescence measurements of on state and off state cells were calculated from the standard deviation of measurements from at least three biological replicates. The relative error levels for the on/off fluorescence ratios were then determined by adding the relative errors of on and off state fluorescence in quadrature. For measurements of in vivo system cross talk, single colonies of each of the 676 strains of transformed cells were measured using flow cytometry. To estimate colony-to-colony variations in GFP output for these strains, we measured a randomly selected subset of 18 transformants and measured them in sextuplicate. The relative uncertainties for these measurements was 12% on average,

- 51 -

which is comparable to uncertainties obtained for flow cytometry experiments used for determining on/off fluorescence ratios for library components.

Colony fluorescence imaging. Images of fluorescence from *E. coli* colonies were obtained using a Typhoon FLA 9000 biomolecular imaging system. All images were

5 obtained using the same PMT voltage, an imaging resolution of 0.1 mm, 473 nm laser excitation, and an LPB (>510 nm long pass) filter for detection of GFP. Induced cells were imaged ~18 hours after they were plated. Since IPTG exhibits low-level fluorescence in the same channel as GFP, variations in the thickness of the LB/agar in the plates result in variations in background fluorescence levels. To compensate for this effect, the minimum
10 GFP intensity measured over each plate was subtracted from the intensity levels of the entire plate, thereby removing most background IPTG fluorescence.

Results

Provided herein is a new system of riboregulators that enable post-transcriptional
15 activation of protein translation. Unlike conventional riboregulators, the synthetic riboregulators of the invention take advantage of toehold-mediated linear-linear interactions to initiate RNA-RNA strand displacement interactions. Furthermore, they rely on sequestration of the region around the start codon to repress protein translation, eschewing any base pairing to the RBS or start codon itself to frustrate translation. As a result, these
20 riboregulators can be designed to activate protein translation in response to a trigger RNA with virtually arbitrary sequence, enabling substantial improvements in component orthogonality. The absence of binding to the RBS and use of thermodynamically favorable linear-linear interactions also enables facile tuning of translational efficiency via RBS engineering. Consequently, these systems routinely enable modulation of protein expression
25 over two orders of magnitude. Based on their interaction mechanism near-digital signal processing behavior, these riboregulator systems are referred to herein as toehold switches.

This disclosure further demonstrates the utility of toehold switches by validating dozens of translational activators in *E. coli* that increase protein production by more than 100-fold in response to a prescribed trigger RNA. Furthermore, we capitalize on the
30 expanded RNA sequence space afforded by the novel riboregulator design to construct libraries of components with unprecedented part orthogonality, including a set of 26 systems that exhibit less than 12% cross talk across the entire set, which exceeds the size of all previous orthogonal regulator libraries by a factor of more than 3. Sequence and

- 52 -

thermodynamic analyses of the toehold switches yield a set of design principles that can be used to forward engineer new riboregulators. These forward engineered parts on average exhibit on/off ratios exceeding 400, a dynamic range typically reserved for protein-based genetic networks using components constructed from a purely rational design framework.

5 *Toehold Switch Design.* The toehold switch systems are composed of two programmed RNA strands referred to as the switch and trigger (FIG. 17B). The switch mRNA contains the coding sequence of a gene being regulated. Upstream of this coding sequence is a hairpin-based processing module containing both a strong ribosome binding site and a start codon followed by a short linker sequence coding for amino acids added to the N-terminus of the gene of interest. A single-stranded toehold sequence at the 5' end of the
10 hairpin module provides the initial binding site for the trigger RNA strand. This trigger molecule is a single-stranded RNA that completes a branch migration process with the hairpin that exposes the RBS and initiation codon, thereby causing activation of translation of the gene of interest.

15 The hairpin processing unit functions as a repressor of translation in the absence of the trigger strand. Unlike previous riboregulators, the RBS sequence is left completely unpaired within the 11-nt loop of the hairpin. Instead, the bases immediately before and after the initiation codon are sequestered within RNA duplexes that are six and nine base pairs long, respectively. The start codon itself is left unpaired in the switches we tested, leaving a
20 3-nt bulge near the midpoint of the 18-nt hairpin stem. Since the repressing domain b (FIG. 17B) does not possess complementary bases to the start codon, the cognate trigger strand in turn does not need to contain corresponding start codon bases, thereby increasing the number of potential trigger sequences. The sequences of the hairpin sequence added after the start codon were also screened for the presence of stop codons, as they would prematurely
25 terminate translation of the gene of interest when the riboregulator was activated. Studies of the GFP expression from the repressed toehold switch mRNA revealed typical repression levels of 98% or higher compared to unrepressed GFP mRNAs. After confirming successful translational repression with this design, we employed a 12-nt toehold domain at the 5' end of the hairpin to initiate its interaction with the cognate trigger strand. The trigger strand bears a
30 30-nt single-stranded RNA sequence that is perfectly complementary to early bases in the switch mRNA.

From this base toehold switch design, we used the NUPACK nucleic acid sequence design package (Zadeh et al., J. Comput. Chem. 32:170-173, 2011) to generate a library of

- 53 -

translational activators. A common 21-nt sequence was used to link the hairpin module of the switch mRNAs to the coding sequence of the gene of interest. This linker sequence was programmed to encode low molecular weight amino acids to minimize its effect on folding of the gene of interest, which was selected in this case to be a GFP reporter. To reduce
5 computational load, only the first 29-nts of GFP were considered for secondary structure analysis. The complete trigger transcript, however, was simulated during the design process. This transcript included a GGG leader sequence to promote efficient transcription from the T7 RNA polymerase promoter, a 5' hairpin domain to increase RNA stability, and the 47-nt T7 RNA polymerase terminator at the 3' end of the transcript. NUPACK was used to
10 generate toehold switch designs satisfying the prescribed secondary structures and having the specified RBS and terminator sequences. Unspecified bases in the designs were random and thus allowed to become any of the four RNA bases, with some sequence constraints applied to NUPACK to preclude extended runs of the same bases. We initially designed a set of 24 toehold switches to gauge in vivo performance and constructed them as described in the
15 Methods section. After confirming that a number of these switches exhibited high dynamic range, we began to design an extended library of toehold switches containing elements selected for low crosstalk with the rest of the library.

To generate this library, a total of 672 toehold switch designs with randomized sequences were generated using NUPACK. Of the resulting designs, 25 were found to encode
20 stop codons in the hairpin region after the start codon. In the remaining systems, one duplicate design was found leaving 646 unique riboregulator designs in the library.

We next selected a subset of 144 of these toehold switch designs for testing in *E. coli* that exhibited the lowest levels of unintended riboregulator-trigger cross talk. In silico screening for cross talk served two purposes. First, the resulting library of orthogonal
25 regulators could provide a large set of components to independently regulate translation in vivo. Second, systems screened for orthogonality would necessarily span a large portion of the sequence space of possible toehold switches and inform future system designs. We simulated pairwise interactions between riboregulator and trigger strands for the complete set of 646 corresponding to 417,316 RNA-RNA interactions. These simulations determined the
30 concentration of any resulting riboregulator-trigger complexes and their secondary structures. The integrity of the toehold switch stem in these riboregulator-trigger complexes was used to determine the likelihood of unintended trigger activation, since the destruction of the duplex regions nearby the start codon would lead to translation of the gene of interest. Through this

- 54 -

stem integrity metric, we used a Monte Carlo algorithm to select 144 toehold switch designs with the predicted lowest net system cross talk. This resulted in a toehold switch library composed of 168 different components with random sequences subject to the same secondary structure constraints.

5 *Component validation.* The toehold switches were tested in *E. coli* BL21 Star DE3 with the switch mRNA expressed off a medium copy plasmid (ColA origin) and the trigger RNA expressed from a high copy plasmid (ColE1 origin). Expression of both strands was induced using IPTG, which triggered production of both RNA species through T7 RNA polymerase. To enable quantitative assessment of switch performance, we used an ASV-
10 tagged GFPmut3b with a reported half-life of 110 min (Andersen et al., Appl. Environ. Microbiol. 64:2240-2246, 1998) as a fluorescent reporter. In these experimental conditions, the copy number differences in the plasmids expressing switch and trigger RNAs led to a 6-8 fold excess of trigger compared to switch molecules as determined by fluorescence measurements of GFPmut3b-ASV expressed separately from each plasmid.

15 Flow cytometry was used to characterize the performance of the toehold switches. Cells were measured at one-hour intervals after induction with IPTG. ON fluorescence was measured for cells transformed with the riboregulator and its cognate trigger, while OFF fluorescence was determined from cells containing the riboregulator and a randomly selected non-cognate trigger. Fluorescence histograms from both activated and repressed toehold
20 switches are almost exclusively unimodal, highlighting their potential use in cellular digital logic (data not shown). The mode fluorescence value from the histograms was used to calculate the on/off ratios of each riboregulator design. FIG. 17C displays the mode GFP fluorescence measured from three toehold switches (numbers 2, 3 and 5) in their on and off states (switch only is first bar, switch and trigger is second bar, and positive control is third
25 bar). For comparison, unrepressed versions of each switch mRNA containing the same sequences for the GFP reporter were also evaluated as positive controls. The off state fluorescence of the switches is near the background fluorescence levels measured for induced cells not expressing GFP. On state fluorescence for the activated toehold switches was comparable to the positive controls, indicating that nearly all switch mRNAs were bound by
30 their trigger RNAs.

Activation of the systems was observed within one hour of induction and increased over time with accumulation of GFP (FIG. 17D, inset). FIG. 17D presents the on/off mode GFP fluorescence ratio determined three hours after induction for all 168 of the switches in

- 55 -

the random sequence library. Of the systems tested, 20 exhibit on/off ratios exceeding 100 and nearly two thirds display at least an on/off greater than 10. In comparison, we also characterized the widely used engineered riboregulators crRNA 10 and 12 (described by Isaacs et al., Nat. Biotechnol. 22:841-847, 2004) in identical conditions. These earlier
5 riboregulation systems exhibited significantly lower dynamic range with on/off values of 11 ± 2 and 13 ± 4 for crRNA systems 10 and 12, respectively.

Evaluation of toehold switch orthogonality. To evaluate the orthogonality of the translational activators, we selected the top 35 riboregulators from the 144 orthogonal component library and performed additional in silico screening to isolate a subset of 26 that
10 displayed extremely low levels of cross talk, both in terms of stem integrity and unwanted binding between non-cognate trigger and switch strands. The pairwise interactions between the 26 riboregulators were then assayed in *E. coli* by transforming cells with all 676 combinations of riboregulator and trigger plasmids. FIG. 18A displays images of GFP fluorescence from colonies of *E. coli* induced on LB plates. The set of orthogonal switches
15 are shown in order of decreasing on/off fluorescence ratio measured in FIG. 17D. Clearly visible is the strong emission from cognate switch and trigger pairs along the diagonal of the grid with the final switch at index 26 displaying lower fluorescence as a result of its low on/off ratio. In contrast, low fluorescence levels are observed for the off-diagonal elements featuring non-cognate trigger/switch RNA pairs.

20 To gain quantitative information, we used flow cytometry to measure the GFP output from all pairwise trigger-switch interactions. Crosstalk was calculated by dividing the GFP fluorescence obtained from a non-cognate trigger and a given switch mRNA by the fluorescence of the switch in its triggered state. The resulting matrix of crosstalk interactions is shown in FIG. 18B. By definition, crosstalk levels along the diagonal are 100%, while
25 those off the diagonal agree with the qualitative output levels from colony images. Based on these data, the toehold switches exhibit an unprecedented degree of orthogonality with the full set of 26 regulators tested displaying under 12% crosstalk. Since the number of regulators in an orthogonal set is defined by its threshold crosstalk level, we identified orthogonal subsets for a range of different crosstalk thresholds. For instance, a subset of 18 of the
30 toehold switches exhibits less than 2% subset-wide crosstalk.

When choosing toehold switches for a given application, a potentially more relevant metric for assessing their performance is the reciprocal of the threshold crosstalk level. For translational activators, this parameter represents the minimum fold change to expect between

- 56 -

when using the set of switches to regulate a protein with similar output characteristics to our GFPmut3b-ASV reporter. FIG. 18C plots this library dynamic range metric against the maximum orthogonal subset size for the toehold switches as a well as a number of other RNA-based regulators. The largest previously reported orthogonal riboregulator set consisted of seven transcriptional attenuators displaying 20% crosstalk (Takahashi et al., Nucleic Acids Res., 2013). This result was obtained using transcriptional attenuators composed of cognate sense and antisense transcripts. A crosstalk level of 20% means that the set of 42 off-target RNA sense-antisense interactions attenuated transcription by at most 20%. For that library, 20% crosstalk results in an upper bound in its overall dynamic range of 5 (FIG. 18C). Earlier orthogonal translational activators and repressors have been limited to sets of four (Callura et al., Proc. Natl. Acad. Sci. USA 109: 5850-5855, 2012) and five (Mutalik et al., Nat. Chem. Biol. 8: 447-454, 2012), respectively, at 20% crosstalk. For proteins, an engineered library of five orthogonal eukaryotic transcription factors crosstalk of ~30% was also reported (Khalil et al., Cell 150:647-658, 2012).

To our knowledge, the switches provided herein constitute the largest set of orthogonal regulatory elements, RNA- or protein-based, ever reported. Furthermore, subsets of orthogonal toehold switches of comparable size to previously reported libraries exhibit minimum dynamic ranges over an order of magnitude larger than previously reported systems. In comparison to previous attempts, cognate RNA interactions using a library of devices described herein reduced transcription by up to 83%.

Component analysis and forward engineering. Flow cytometry data from the toehold switches provided a substantial dataset with which to determine sequence-dependent variations in riboregulator performance. As a coarse screen for sequence-dependent effects, we began to investigate toehold switch output as a function of base pairing at the top and bottom in the stem of the riboregulator strand (FIG. 19A). We hypothesized that the strength of the base pairing in these regions would have a strong effect on the repression strength of the hairpin as they are essential to sequestering the start codon, and they could also affect the secondary structure of the RBS and mRNA region once the riboregulator is activated, which in turn influences translational efficiency (Kudla et al., Science 324:255-258, 2009). Analysis of the top and bottom three base pairs in the hairpin module revealed significant variations in the on/off ratio of riboregulators as a function of the G-C base pair content in these regions. FIG. 19B displays the average on/off fluorescence obtained for all 16 possible permutations of G-C content in the two stem regions, as well as the on/off values obtained for each toehold

- 57 -

switch that satisfied the specified G-C conditions. Based on the size of the library and secondary structure constraints imposed during in silico design, a number of G-C permutations had only one or two representative toehold switches. Toehold switches containing zero and two G-C base pairs at the top and bottom regions of the stem, respectively, displayed an average on/off fluorescence ratio of 154, over three times higher than the next highest permutation. Mean on/off levels also tended to steadily decrease as G-C combinations deviated further from this combination.

The bias toward low G-C content at the top of the riboregulator stem suggested potential interaction between the bound ribosome and the nearby RNA duplex in the activated riboregulator-trigger complex. In particular, weak base pairing at the end of the RNA duplex could allow the duplex to breathe open, spontaneously freeing bases upstream of the RBS to facilitate ribosome binding. To investigate this effect, we studied a series of riboregulators with different hairpin loop sizes to tune the size of the pre-RBS region (FIG. 19A), defined as the nucleotides between the RNA duplex and the start of the RBS sequence. Measurements of the loop variant riboregulators demonstrated steady increases in the on state fluorescence output as the pre-RBS region was increased from 3- to 19-nts in size through the addition of an A-rich sequence (data not shown). Notably, these increases in on state expression did not result in corresponding increases in system off state until the pre-RBS sequence was 13-nts in length, which corresponded to a loop of 21-nts. These observations are consistent with previous studies that demonstrated translational enhancement through A/U bases placed immediately upstream of the RBS (Vimberg et al., BMC Mol. Biol. 8:1-13, 2007). Furthermore, they provided a straightforward means of increasing toehold switch dynamic range by increasing the length of its hairpin loop. Systematic studies of toehold switch behavior as a function of trigger RNA length were also conducted. These studies revealed a strong positive correlation between system on/off ratios and the length of the toehold domain (data not shown) and demonstrated that switch output could be increased by only partially unwinding the stem of the switch (data not shown).

Previous riboregulators have been designed on a case-by-case basis (Isaacs et al., Nat. Biotechnol. 22:841-847, 2004; and Callura et al., Proc. Natl. Acad. Sci. USA 109:5850-5855, 2012) and those that have utilized computer-assisted design have not demonstrated consistently high on/off levels (Rodrigo et al., Proc. Natl. Acad. Sci. USA 109:15271-15276, 2012). In silico designed riboregulators forward engineered to exhibit high performance in vivo have the potential to significantly reduce the time required for generating new genetic

- 58 -

circuits, in turn enabling the realization of more complex cellular logic. Consequently, we integrated the above findings into designs for a set of toehold switches forward engineered for high dynamic range. Our forward engineered systems retain the same general secondary structure and interaction mechanisms of the library of 168 toehold switches, but adopt several of the insights described above to significantly improve their dynamic range. First, we incorporated the combination of switch mRNA sequence constraints revealed in FIG. 19B. Specifically, the top three bases of the hairpin stem were restricted to weak A-U base pairs. The bottom three base pairs of the stem were specified to contain two strong G-C base pairs and one A-U base pair. Second, we increased the length of the switch toehold from 12- to 15-nts. This change strengthened the initial binding between the trigger and the switch. Third, we increased the size of the hairpin loop from 11- to 15-nts to enhance translation of the output protein upon switch activation. We selected a fairly conservative loop size of 15-nts to ensure that leakage from the system in its off state remained low. Lastly, we exploited a cognate trigger that only unwound the first 15 of the 18 bases in the switch stem. This design change yielded a number of benefits. It enabled the trigger RNA to bypass binding to the top three bases in the hairpin stem, which were all specified to be A and U bases, thereby eliminating corresponding sequence constraints for the trigger and leaving its length unchanged at 30-nts. Furthermore, avoiding disruption of the top three weak base pairs of the stem allowed them to breathe open spontaneously after lower bases in the stem were unwound. This design change effectively increased the size of the pre-RBS region by adding a 3-nt A/U enhancer element without a concomitant increase in off state leakage.

We employed NUPACK to design 13 forward engineered toehold switches with the four system modifications detailed above. FIG. 19C presents the on/off mode fluorescence ratios for the forward engineered translational activators regulating GFP after 3 hours of induction. There is dramatic increase in on/off fluorescence for almost all the systems tested, with 12 out of 13 exhibiting a dynamic range that was comparable to or higher than the highest performance toehold switch from the initial library. These forward engineered systems exhibit an average on/off ratio of 406 compared to 43 for the initial toehold switch design. This mean on/off ratio rivals the dynamic range of protein-based regulation systems using a highly programmable system design and without requiring any evolution or large scale screening experiments. Furthermore, even the lowest performing optimized toehold switch displayed an on/off ratio of 33 ± 4 , which is still sufficient for many cellular decision making operations. Hourly time course measurements reveal activation of forward engineered

- 59 -

switches after 1 or 2 hours of induction (FIG. 19C, inset). Furthermore, on state fluorescence increased steadily over 4 hours, yielding on/off levels well over 600 for both switches.

We quantified the effectiveness of our forward engineering strategy by calculating the percentage of forward engineered designs with on/off ratios exceeding a given minimal level and comparing them to the same calculation performed on the library of 168 toehold switches with random sequences (FIG. 19D). The yield of high performance switches is higher for the forward engineered switches for all on/off ratios tested. For instance, 92% of the forward engineered designs had on/off GFP fluorescence of at least 287 compared to a single switch out of 168 for of the random sequence library.

Thermodynamic analysis of system performance. Our forward engineering resulted in riboregulators with 92% likelihood of high dynamic range. To develop a predictive model of riboregulator activity, on/off ratios of the 168 initial switches with random sequences were analyzed in terms of a number of thermodynamic parameters falling into six different categories (FIG. 20A). On/off ratios as opposed to fluorescence output in the on and off states alone were used for quantitative analysis since fluorescence off states varied relatively little over the library, leaving on/off ratios essentially a measure of on state fluorescence. Following the treatment by Salis et al., Nat. Biotechnol. 27: 946-950, 2009, the amount of expressed protein p can be related to thermodynamic free energies through the equation $p \propto \exp(-k\Delta G)$, where k is a fitting parameter. Consequently, relationships between thermodynamic parameters and riboregulator on/off values can be evaluated by the coefficient of determination R^2 of a linear regression applied to a semi-logarithmic plot of free energy versus on/off ratio. However, each of the thermodynamic parameters failed to demonstrate any significant correlation with riboregulator output characteristics when applied to the full component library.

Based on the sequence-dependent effects observed in FIG. 19B, we began to search for relationships between thermodynamic parameters and subsets of toehold switches sharing similar sequence characteristics (FIG. 20A, data not shown). By probing R^2 values for a number of switch subsets, we identified a single parameter $\Delta G_{\text{RBS-linker}}$ that displayed a clear correlation with system output. $\Delta G_{\text{RBS-linker}}$ is the free energy associated with the secondary structure of the region beginning immediately downstream of the RNA duplex of the riboregulator-trigger complex and running through to the end of the common 21-nt linker added after the hairpin module (FIG. 20B). It reflects the amount of energy required by the ribosome to unwind the RBS/early-mRNA region as it binds and begins translation of the

- 60 -

output gene. Variations in translational efficiency have previously been linked to secondary structure early in mRNAs and similar thermodynamic factors have been employed to calculate the strength of prokaryotic RBSs (Salis et al., Nat. Biotechnol. 27: 946-950, 2009). FIG. 20C provides an example of the relationship between Δ GRBS-linker and the on/off ratios for a subset of 68 riboregulators each containing a weak A-U base pair at the top of its stem. This set of riboregulators from the library was the largest for which a correlation with Δ GRBS-linker with $R^2 \geq 0.4$ was identified. In contrast, the complementary subset of 100 riboregulators containing a strong GC base pair at the top of their stems displayed no correlation with Δ GRBS-linker with $R^2 = 0.024$, possibly as a result of sequence dependent interactions with the ribosome at its standby site.

Having identified the importance of Δ GRBS-linker, we proceeded to investigate its relationship with on/off levels from the forward engineered systems. We found that Δ GRBS-linker exhibited a much stronger correlation with on/off levels, yielding $R^2 = 0.79$ (FIG. 20D). Most importantly, we found that this single thermodynamic term was sufficient to explain the single low performance forward engineered toehold switch. This particular toehold switch possessed relatively high secondary structure in the RBS-linker region that significantly decreased the translational efficiency of the activated switch mRNA.

Multiplexed Regulation. The orthogonality of the toehold switches can enable them to independently regulate multiple proteins simultaneously within the cell. To demonstrate this capability, we transformed cells with plasmids expressing two orthogonal toehold switch mRNAs expressing spectrally distinct fluorescent proteins GFP and mCherry, denoted A* and B*, respectively (FIG. 21A). The cognate trigger RNAs of these toehold switches were then expressed in all four possible combinations with reporter expression quantified using flow cytometry (FIG. 21B). Upon transcription of either the A or B trigger alone, GFP and mCherry fluorescence increases by over an order of magnitude, respectively, while fluorescence levels in the orthogonal channel are virtually unchanged. Co-expression of both A and B trigger RNAs yields strong increases in expression of both fluorophores, as expected for the two toehold switches.

Toehold switches triggered by functional mRNAs. The sequence space afforded by the toehold switch design enables them to be triggered by functional mRNAs (FIG. 21C). However, the fixed sequences of these mRNA triggers present significant challenges for effective system activation. Unlike synthetic trigger RNAs designed to be completely single-stranded, strong secondary structures abound within the mRNAs, frustrating toehold binding

- 61 -

and decreasing the thermodynamics driving the branch migration process. The toehold sequences defined by the trigger mRNA can also exhibit base pairing both internally and with sequences downstream of the hairpin module, and thus pose similar challenges to switch activation. In order to counter these effects, we increased the toehold domain length of the

mRNA-responsive switches from 12-nts to ≥ 24 -nts. This modification helped shift the importance of single-stranded regions for binding initiation from the trigger mRNA to the toehold switch itself, where only downstream sequences in the switch could hybridize with the binding region. In addition, we exploited a number of design features identified during detailed study of the highest performance toehold switch from the 168 system library.

Toehold switch number 1 had an on/off ratio of 290 ± 20 when paired with its complete 30-nt cognate trigger RNA. We found that on/off ratios increased sharply by using shortened trigger RNAs truncated from their 5' end. In particular, we observed that toehold switch number 1 could provide an 1900 ± 200 on/off fluorescence in response to a trigger RNA intended to only unwind the bottom five bases of its stem. Secondary structure and thermodynamic analyses of the toehold switch number 1 system indicated that this extreme dynamic range was due to two factors. First, the stem of switch number 1 contained a relatively high proportion of weak A-U base pairs, and G-C base pairs in the stem were concentrated toward the bottom of the stem. Consequently, when a trigger disrupted the bottom five base pairs, half of the G-C base pairs in the stem were eliminated leaving a weak stem containing predominantly A-U base pairs available for ribosome binding. Furthermore, trigger binding to only the lower bases in the stem increased the pre-RBS region of the activated switch and provided additional enhancement of translation. Second, bases freed from the stem upon trigger binding were shown to interact with downstream bases in the switch linker region forming a weak stem loop (data not shown). This refolding mechanism led to an additional base of the stem being disrupted, which further weakened its repression strength, and decreased the energetic barrier to trigger binding.

We incorporated all the design features discussed above to generate toehold switches that were responsive to mRNAs. The switch hairpin modules were derived from the toehold switch number 1 sequence. Specifically, the top 12-bases and loop of the switch number 1 stem were used in all mRNA sensors (FIG. 21C). In addition, the size of the sensor loop was increased from 11- to 18-nts to increase reporter expression. The toehold and the bottom 6 base pairs of the sensor stem had variable bases programmed to interact with the trigger mRNA. 24- and 30-nt toeholds were used for initial mRNA binding and the bottom 6 base

- 62 -

pairs were specified to be unwound by the trigger. To decrease the energetic barrier to stem unwinding by the trigger mRNA, we also explicitly encoded the downstream RNA refolding mechanism discussed above into the sensors. These RNA refolding elements induced the formation of a 6-bp stem loop after disruption of the bottom four base pairs of the switch stem and in turn forced the disruption of two additional bases in the switch stem. Using this
5 bases toehold switch mRNA sensor design, we simulated the secondary structures and thermodynamics of all possible sensors along the full length of the trigger mRNA. We then used in silico screening to identify toehold switches that offered the best combination of sensor secondary structure and mRNA binding site availability.

10 The resulting mRNA sensors were tested in the same manner as previous experiments, with the trigger mRNA expressed from a high copy ColE1 origin vector and toehold switches regulating GFP expressed from a medium copy ColA origin vector. We selected a trio of exogenous mRNA triggers, mCherry, chloramphenicol acetyltransferase (cat, conferring chloramphenicol resistance), and aadA (conferring spectinomycin resistance),
15 for sensing experiments to minimize the likelihood of switch activation by endogenous RNAs. The mCherry trigger RNA featured an RBS region to enable efficient translation, while the two antibiotic resistance conferring mRNAs lacked an RBS, as translation by the ribosome could interfere with recognition and binding of the toehold switch. FIG. 21D presents the on/off GFP fluorescence measured from five toehold switches. The three sensors
20 triggered by the translatable mCherry mRNA provide the strongest activation with design A displaying best on/off ratio of 57 ± 10 . The toehold switches triggered by the non-translatable mRNAs displayed more modest ~7-fold activation levels.

To establish the effect of toehold switch binding to translation from the trigger mRNA, we also performed experiments measuring mCherry output in the presence or
25 absence of the mCherry sensor. FIG. 21E contains the fluorescence of GFP measured for the three mCherry-responsive switches in their active and repressed states in addition to the mCherry fluorescence measured from the activated cells. For comparison, fluorescence measurements obtained from control experiments are also presented showing background GFP fluorescence measured from uninduced cells as well as fluorescence measured from
30 unregulated expression of GFP and mCherry from ColA and ColE1 origin vectors, respectively. Expression of mCherry is not strongly affected by transcription of the toehold switch RNA. This suggests that binding between the trigger and switch does not inhibit translation by the ribosome, although the molar excess of trigger RNA compared to switch

- 63 -

dampens the strength of this effect in our experiments. The GFP expression levels from the activated switches vary within a factor of only 2.5, while leakage from the repressed switches varies by about 5-fold. This variation in leakage is the determining factor explaining variations in on/off levels of the mCherry sensors and is due to the use of the highly sensitive parent toehold switch as the parent design for the mRNA sensors.

We also designed a toehold switch sensor to detect the endogenous *E. coli* small RNA (sRNA) *ryhB*. RyhB is a 90-nt transcript that down-regulates iron-associated genes in conditions where iron levels are low (Masse and Gottesman, PNAS 99: 4620-4625, 2002). To characterize the sensors, cells were transformed with plasmids constitutively expressing a *ryhB*-responsive toehold switch regulating GFP (FIG. 22A). The *ryhB* sRNA was induced by adding the iron-chelating compound 2,2'-bipyridyl to the growth media (FIG. 22A), which is known to rapidly stimulate expression of *ryhB* (Masse and Gottesman, PNAS 99: 4620-4625, 2002). We measured sensor output one hour after induction by the chelator using flow cytometry. FIG. 22B presents the GFP fluorescence from the sensor as a function of 2,2'-bipyridyl concentration. The sensor transfer function shows a steady increase in GFP expression as 2,2'-bipyridyl levels increase to 0.3 mM, beyond which levels plateau. For comparison, we also measured GFP output for a positive control construct using the same constitutive promoter as the sensor and having an exposed RBS with similar reporter output levels. The GFP positive control, in contrast, decreased in output as the concentration of the iron chelator increased, which demonstrates the response of the *ryhB* sensor was not the result of increased translation in response to the chelator. In addition, we performed control experiments using exogenously expressed *ryhB* sRNA and an off-target RNA (data not shown). The *ryhB* sensor activated GFP expression in response to the exogenously-expressed *ryhB*, but remained inactive for the off-target RNA confirming its specificity for the intended target RNA.

Discussion

Toehold switches represent a versatile and powerful new platform for regulating translation at the post-transcriptional level. They combine an unprecedented degree of component orthogonality with system dynamic range comparable to widely used protein-based regulatory elements²². Comprehensive evaluation of in vivo switch-trigger pairwise interactions resulted in a set of 26 toehold switches with sub-12% cross talk levels. To our knowledge, this represents the largest library of orthogonal regulatory elements ever reported

- 64 -

and exceeds previous libraries by a factor of over three in size (Takahashi et al., *Nucleic Acids Res.*, 2013). At this point, the ultimate size of the orthogonal sets of toehold switches is limited by the throughput of our cross talk assay, not design features intrinsic to the riboregulators. Furthermore, forward engineering of 13 toehold switch systems yielded a subset of 12 new high performance components that exhibited an average on/off fluorescence ratio of 406, with the performance of the complete set predicted by a two parameter thermodynamic model.

Crucial to these advances was the adoption of new mechanisms for translational repression and initiation of RNA-RNA interactions in vivo. Toehold switches strongly repress translation in their off state by sequestering the sequences nearby the initiation codon of the regulated gene within RNA duplexes, in contrast to previous riboregulators that repress by blocking access to the RBS and in some cases the start codon (Isaacs et al., *Nat. Biotechnol.* 22:841-847, 2004; Rodrigo et al., *Proc. Natl. Acad. Sci. USA* 109:15271-15276, 2012; and Mutalik et al., *Nat. Chem. Biol.* 8:447-454, 2012). While earlier riboregulators have relied on loop-linear (Isaacs et al., *Nat. Biotechnol.* 22:841-847, 2004; and Mutalik et al., *Nat. Chem. Biol.* 8:447-454, 2012) and loop-loop (Lucks et al., *Proc. Natl. Acad. Sci. USA* 108:8617-8622, 2011; Rodrigo et al., *Proc. Natl. Acad. Sci. USA* 109:15271-15276, 2012; and Takahashi et al., *Nucleic Acids Res.* 2013) interactions, toehold switches exploit toehold-mediated linear-linear RNA interactions to initiate binding between the riboregulator mRNA and trigger RNA. Taken together, these operating mechanisms enable the toehold switches to accept trigger RNAs with nearly arbitrary sequences, greatly expanding the sequence space for orthogonal operation, and they promote RNA-RNA interactions with high reaction kinetics by using extended toehold domains 12- to 15-nts in length. In contrast to earlier reports, thermodynamic analyses of toehold switch performance did not reveal significant correlations between riboregulator on/off levels and the free energy of the riboregulator-trigger interaction nor the free energy of toehold-trigger binding Mutalik et al., *Nat. Chem. Biol.* 8:447-454, 2012). These observations suggest that RNA-RNA interactions for the toehold switches are strongly thermodynamically and kinetically favoured.

We attribute the increased dynamic range of our toehold switches to three main factors. First, the increased kinetics and thermodynamic free energy driving trigger-switch interaction causes a higher percentage of the total switch mRNAs present in the cell to be triggered to produce the output produce. We found that the fraction of activated switch mRNAs was around 100% based on comparison measurements with unrepressed versions of

- 65 -

the switch mRNA (FIG. 17C). Second, the design of the toehold switches, in which no bases of the RBS are enclosed within a stem, provides a much better platform through which to engineer the RBS and its surrounding bases for optimal expression of the regulated gene. Experiments varying the loop size of the toehold switch mRNA demonstrated a very strong
5 dependence between the on state fluorescence of the switch and the presence of longer A-rich regions upstream of the RBS (data not shown). Importantly, this RBS enhancement required only additional bases to be added to the loop region and did not require corresponding changes in the sequence of the trigger RNA. In contrast, for many previous riboregulator systems, similar RBS engineering would require modifications to be made to both pairs of
10 riboregulator RNAs, complicating the design and requiring deconvolution of effects from RBS and the RNA-RNA interactions to properly interpret results. Lastly, the toehold switches were designed in silico to provide RBS and early mRNA regions with very little secondary structure to promote efficient translation of the regulated gene. Although this was accomplished by adding additional bases and a linker to the N-terminal of the output gene,
15 algorithms to select optimal codons with respect to mRNA secondary structure can be used to produce toehold switches without adding N-terminal bases. Synonymous codons should also enable the construction of large orthogonal sets of such N-terminal restricted of switches.

Example 8. Synthetic regulation of endogenous genes

20 A library of toehold switches was used to detect mRNAs and endogenous RNAs in vivo, and to regulate endogenous gene expression by integrating switches into the genome. We demonstrate their potential applications in synthetic biology by using toehold switches to regulate a dozen components in the cell at the same time, and by incorporating them into a genetic circuit to compute a 4-input AND expression.

25 Toehold switches can be integrated into the genome to provide synthetic regulation of endogenous genes. We used “lambda” Red recombination (Datsenko and Wanner, PNAS, 97:6640-6645, 2000) to insert toehold-switch hairpin modules upstream of targeted genes in the *E. coli* chromosome. Template genome-editing plasmids were constructed that contained a high-performance second-generation switch adjacent to a kanamycin resistance marker
30 flanked by a pair of FRT sites (FIG. 23A). Linear DNA fragments were amplified from these plasmids using primers with homology to the desired insertion site in the genome. Primers were designed to integrate the switch into the same reading frame of the targeted endogenous gene and to replace the endogenous RBS with the RBS of the toehold switch. Following

- 66 -

transformation of the linear DNA, cells successfully integrated with the insertion cassette were selected on kanamycin plates and the resistance marker subsequently excised by expressing FLP recombinase, which recognized its flanking FRT sites. The resulting *E. coli* strain retains a functional copy of the targeted gene in its chromosome; however, it is
5 deactivated as a result of the co-transcribed switch RNA module. This repressed gene can be activated post-transcriptionally by expression of a cognate trigger.

We validated this approach for regulating endogenous genes by inserting switches upstream of three genes *uidA*, *lacZ*, and *cheY*. The genes *uidA* and *lacZ* produce the enzymes beta-glucuronidase and beta-galactosidase, respectively. Cells expressing these enzymes can
10 be readily identified by their blue/green color on plates containing the corresponding substrates X-Gluc and X-Gal. We constructed two strains with synthetic *uidA* regulation by integrating switches A and B into the chromosome (*uidA*::Switch A and *uidA*::Switch B, respectively). FIG. 23B displays these strains upon expression of different trigger RNAs as well as a control strain with the wild-type *uidA* genotype. As expected, *uidA*::Switch A only
15 exhibits the blue/green wild-type phenotype upon expression of trigger A. Similarly, *uidA*::Switch B activates beta-glucuronidase only with cognate trigger B.

The edited strain *lacZ*::Switch C provides more complicated behavior since the *lac* operon is regulated at the transcriptional level by lactose or chemical analogs such as IPTG. Consequently, *lacZ*::Switch C requires both lactose/IPTG and trigger RNA C to turn on
20 expression of beta-galactosidase. This behavior results in a genetic AND circuit combining transcriptional and post-transcriptional regulation. We tested this AND circuit by expressing different trigger RNAs using inducible promoters responsive to anhydrotetracycline (aTc). FIG. 23C provides images of *lacZ*::Switch C transformed with different trigger plasmids for all four combinations of the two chemical inputs (i.e. IPTG and aTc) of the AND circuit. In
25 the absence of IPTG, none of the strains shows any change in color since expression of the *lac* operon is strongly repressed. For a plate containing IPTG and no aTc, the wild-type *lacZ* strain (upper left quadrant) becomes blue/green in color, while the *lacZ*::Switch C strains with different triggers do not change in color since the trigger RNAs are not being expressed. When the AND condition is satisfied by adding both IPTG and aTc to the plate, *lacZ*::Switch
30 C exhibits the blue/green color change with trigger RNA C as expected, while those expressing triggers A and B are unchanged.

Lastly, we conditionally regulated the *E. coli* chemotaxis gene *cheY* using a fourth toehold switch. The resulting strain *cheY*::Switch D was transformed with plasmids that

- 67 -

expressed triggers inducibly via IPTG. Regulation of cheY is readily observed through changes in the motility of cheY::Switch D on soft agar plates (FIG. 23D). Only cells expressing trigger RNA D with IPTG demonstrated significant motility, while those expressing non-cognate trigger RNA A or lacking the IPTG inducer were unable to move from the point of inoculation.

Example 9. Multiplexed regulation

To demonstrate the full multiplexing capabilities of toehold switches, we expressed twelve toehold switches in the same cell and independently confirmed their activity via flow cytometry. We used four different fluorescent proteins (GFP, venus, cerulean, and mCherry) as reporters and constructed three compatible plasmids to express each of the reporter proteins (FIG. 24A). Each of these proteins was regulated using a different high-performance switch from the second-generation library. The resulting constructs were expressed from a single T7 promoter as ~3.4-kb polycistronic mRNAs resulting in a total synthetic network size of over 10-kb. Trigger RNAs were then expressed in different combinations from a fourth compatible plasmid.

FIG. 24B presents the outcome of the multiplexing experiments. Cells were measured 6 hours after IPTG induction and the output of each reporter represented in terms of the percentage of cells expressing the reporter. A cell was determined to be expressing a given reporter if its fluorescence exceeded a threshold level held constant for all the plots in FIG. 24B. This threshold was set at 10 times the median fluorescence measured for cells expressing a trigger not cognate to any of the switches in the cell (data not shown). The first two rows of FIG. 24B display the output from all twelve of the switches activated separately from a single expressed trigger RNA. In all twelve cases, significant expression is only observed from the intended reporter with limited crosstalk in the three other channels.

We also tested activation of all two- and three-color combinations of reporter proteins. These also provided the intended output color combinations (FIG. 24B). We observed some unintended leakage of cerulean and GFP for trigger combinations activating GFP/venus and venus/cerulean, respectively. Since single trigger measurements of each of these triggers displayed low leakage, we attribute much of the observed leakage in these two-trigger experiments to imperfect compensation of the flow cytometry data caused by the spectral overlap between GFP, venus, and cerulean (data not shown). We also successfully activated all four output proteins by simultaneously expressing four trigger RNAs. Finally,

we observed low system leakage from each of two non-cognate trigger RNAs (FIG. 24B, two bottom-right panels).

Example 10. Implementation of a 4-input AND circuit

5 Toehold switches are readily integrated with existing biological components to build sophisticated genetic programs. We demonstrate this capability by constructing a layered 4-input AND gate consisting of three toehold switches coupled to two orthogonal transcription factors and a GFP reporter (FIG. 25A). In this circuit, toehold switch RNA pairs act as two independent input species that must both be present before a 2-input logical AND expression
10 evaluates as TRUE. The first computational layer in the circuit consists of two 2-input AND toehold switch gates, which each produce a transcription factor. We employed a pair of extracytoplasmic function (ECF) sigma factors as the transcription factors in the circuit since they had previously been reported to be highly orthogonal (Rhodius et al., Mol. Syst. Biol. 9, 2013). The sigma factors produced from the first layer then activate transcription of the
15 toehold switch RNAs in the second computational layer, which in turn lead to expression of a GFP reporter. Similar layered circuits have previously been constructed using transcription factors that required a second chaperone protein for full activity and made use of directed evolution to ensure component orthogonality (Moon et al., Nature, 491:249-253, 2012). Here, we directly use components from the forward-engineered toehold switch library as is without
20 any additional optimization.

 To validate the circuit, we constructed plasmids to express all 16 combinations of the four input trigger and switch RNAs. Input combinations in which a given trigger or switch RNA was missing, a logical FALSE value, were tested by replacing that RNA with a non-cognate trigger or switch RNA, such that the total RNA expressed by the cells remained the
25 same for all input possibilities. We induced expression of the input RNAs using IPTG and measured output from the circuits 8 hours after induction to provide sufficient time for signal propagation. The full truth table for the 4-input AND computation is shown in FIG. 25B with ON/OFF levels calculated by normalizing GFP fluorescence to that measured for the case in which none of the input bits are expressed. GFP output when all four input bits are present is
30 significantly higher than all other input permutations, as expected for a functional 4-input AND computation.

Example 11. Implementation of combined n-input AND m-input OR systems

A variety of AND/OR systems were designed, generated and tested using the methodology described in Example 1.

FIGs. 26A-30 provide the schematics and experimental data relating to a 4-input OR system, wherein each repressor is activated by a multiple input AND gate. Expression of the GOI can be induced by the disruption of any of the repressors, and thus the construct is referred to as an OR system. As illustrated, the system comprises 4-input OR repressors (G1, G2, G3 and G4), and each of G1, G3 and G4 were designed to be activated in the presence of their respective 2 inputs (and thus each is a 2-input AND gate), while G2 was designed to be activated in the presence of its 3 input (and it is a 3-input AND gate). In comparison, FIG. 31A provides the schematic and experimental data relating to a 4-input OR system, wherein each repressor is activated by a single input (rather than by 2-3 inputs together, as is the case with the system of FIGs. 26A-30).

Similarly FIGs. 32A-33 provide the schematics and experimental data relating to a 5-input OR system, wherein each repressor is activated by an multiple input AND gate. As illustrated, the system comprises 5-input OR repressors (G1, G2, G3, G4 and G5), and each of G1, G3, G4 and G5 were designed to be activated in the presence of their respective 2 inputs (and thus each is a 2-input AND gate), while G2 was designed to be activated in the presence of its 3 inputs (and it a 3-input AND gate). In comparison, FIGs. 34A-B provide the schematic and experimental data relating to a 5-input OR system, wherein each repressor is activated by a single input (rather than by 2-3 inputs together, as is the case with the system of FIGs. 32A-33).

FIG. 35 provides experimental data relating to a 6-input OR system, wherein each repressor is activated by a 2-input AND gate. The system comprises 6-input OR repressors (G1, G2, G3, G4, G5 and G6), each of which was designed to be activated in the presence of its respective 2 inputs. The orientation of the repressors relative to the GOI is as follows: G1 – G2 – G3 – G4 – G5 – GOI. This may be regarded as a 12-input system.

FIGs. 36A-B provides the schematic and experimental data relating to a 5-input OR system, wherein each repressor is activated by a 2-input AND gate. The system comprises 5-input OR repressors (G1, G2, G3, G4 and G5), each of which was designed to be activated in the presence of its respective 2 inputs. The orientation of the repressors relative to the GOI is as follows: G1 – G2 – G3 – G4 – G5 – GOI. This may be regarded as a 10-input system.

- 70 -

Comparison of FIGs. 35 and 36B indicate that the 5-input OR system provides a greater signal to noise ratio than does the 6-input OR system.

FIGs. 37A-B provides the schematic and experimental data relating to a 4-input OR system, wherein each repressor is activated by a 2-input AND gate. The system comprises 4-
5 input OR repressors (G1, G2, G3 and G4), each of which was designed to be activated in the presence of its respective 2 inputs. The orientation of the repressors relative to the GOI is as follows: G1 – G2 – G3 – G4 – GOI. This may be regarded as an 8-input system.

FIGs. 37A-B provides the schematic and experimental data relating to a 4-input OR system, wherein each repressor is activated by a 2-input AND gate. The system comprises 4-
10 input OR repressors (G1, G2, G3 and G4), each of which was designed to be activated in the presence of its respective 2 inputs. The orientation of the repressors relative to the GOI is as follows: G1 – G2 – G3 – G4 – GOI. This may be regarded as an 8-input system.

FIGs. 38A-B provide a schematic for a N-AND system. In the absence of triggers, the GOI is expressed. In the combined presence of triggers (or half-triggers) A1 and A2,
15 such triggers complex with each other and then hybridize to the switch RNA. This induces a structural change in the switch RNA that sequesters the RBS and start codon in the stem loop structure, thereby reducing expression of the GOI.

Sequences for various riboregulators and triggers described herein:

20 Table 1. Sequence and performance information for certain toehold switches and triggers from the initial set of 24 random sequence toehold switches.

Toehold switch number	Switch sequence	Trigger sequence	On/off
4	GGGAAUUGAUAUUGUGAUU AUGUGAUGAUUGUAAACAG AGGAGAUACAAUAUGCACA UAAUCAACCUGGCGGCAGCG CAAAAGAUGCGUAAA (SEQ ID NO:1)	GGGACAUACGGACUCACG UGUCCGUAUGUCAAUACA AUCAUCACAUAUUCACAA UAUCAAUUACU (SEQ ID NO:2)	264.6 ± 19.9
10	GGGAAUUGAUAUUGUUCGU UUCGUAUGAUCUAAGACAG AGGAGAUUAGAUUGACGA AACGAAACCUGGCGGCAGCG CAAAAGAUGCGUAAA (SEQ ID NO:3)	GGGACAUACGGACUCACG UGUCCGUAUGUAAACUAG AUCAUACGAAACGAACAA UAUCAAUUACU (SEQ ID NO:4)	192.4 ± 26.9

- 71 -

Toehold switch number	Switch sequence	Trigger sequence	On/off
14	GGGAAUUGAUUUGUAGUA UGUUGAAGUGAUUGAACAG AGGAGACAAUCAUGCAAC AUACUAACCUGGCGGCAGCG CAAAAGAUGCGUAAA (SEQ ID NO:5)	GGGACAUACGGACUCACG UGUCCGU AUGUCAGCAAU CACUUAACAUAUCUACAA UAUCAAUUACU (SEQ ID NO:6)	153.4 ± 10.1

5 Table 2. Sequence and performance information for certain toehold switches and triggers from the set of 144 orthogonal random sequence toehold switches.

Toehold switch number	Switch sequence	Trigger sequence	On/off
1	GGGUGAAUGAAUUGUAGGC UUGUUAUAGUUAUGAACAG AGGAGACAUACAUGAACA AGCCUAACCUGGCGGCAGCG CAAAAGAUGCGUAAA (SEQ ID NO:7)	GGGACCGUGGACCGCAUG AGGUCCACGGUAAACAU ACUAUAACAAGCCUACAA UUCAUCAAAC (SEQ ID NO:8)	292.0 ± 19.5
2	GGGUAAAGUAAAUCGCUU GCUGUAUGUCGUUAAACAG AGGAGAUAAAGAAUGACAG CAAGCAACCUGGCGGCAGCG CAAAAGAUGCGUAAA (SEQ ID NO:9)	GGGAUGCCCGUAGUUCAU UCUACGGGCAUGAAUAAAC GACAUACAGCAAGCGAUU UACUUAUACUA (SEQ ID NO:10)	279.6 ± 17.6
3	GGGUGAUGGAAUAAGGCUG UGUAUAUGAUGUAGACAG AGGAGAUACAUAUGAUAC ACAGCAACCUGGCGGCAGCG CAAAAGAUGCGUAAA (SEQ ID NO:11)	GGGUCAGUUCUGAGGUA CCAGGAACUGAAACUAAAC AUCAUAUACACAGCCUUA UCCAUCACAC (SEQ ID NO:12)	265.3 ± 28.2
5	GGGUAGAUUUGAAUGCUG CUGUUAUGUCGUUAAACAG AGGAGAUAAAGAAUGAACA GCAGCAACCUGGCGGCAGCG CAAAAGAUGCGUAAA (SEQ ID NO:13)	GGGACGCGAAUUGCUAUU CCAUUUCGCGUGACUAAAC GACAUAAACAGCAGCAUUC AAUAUCUAAAC (SEQ ID NO:14)	253.0 ± 12.5
6	GGGAUAAGUAGAUAAAGAUU GUUAGAUGGCUUCGAACAG AGGAGACGAAGCAUGCUGAA CAAUCAACCUGGCGGCAGCG CAAAAGAUGCGUAAA (SEQ ID NO:15)	GGGUCGUGCGCUCUGAGC CGAGCGCACGAGAACGAA GCCAUCUAAACAAUCUUAU CUACUUAUCAC (SEQ ID NO:16)	219.8 ± 16.5
7	GGGAUCACUUAUUGUCGUC UUUGUAUGUCUGUAAACAG AGGAGAUACAGAAUGACAA AGACGAACCUGGCGGCAGCG CAAAAGAUGCGUAAA (SEQ ID NO:17)	GGGACGUCGAUUCACCGU CGAAUCGACGUAGAUACA GACAUACAAAGACGACAA UAAGUGAUAGA (SEQ ID NO:18)	213.3 ± 28.3

- 72 -

Toehold switch number	Switch sequence	Trigger sequence	On/off
8	GGGACAAAGAUUGGUCGUU UCAUUAACGUUAGAAACAG AGGAGAUUCUAACAUGAAUG AAACGAACCUGGCGGCAGCG CAAAAGAUGCGUAAA (SEQ ID NO:19)	GGGACUCCAGGCGGAAUA ACGCCUGGAGUAAAUCUA ACGGUAAUGAAACGACCA AUCUUUGUAUG (SEQ ID NO:20)	200.8 ± 13.5
9	GGGCGAAAGUGUAUGGCUG AUAUGAUGUAGUUAACAG AGGAGAUAAACUAAUGCAUA UCAGCAACCUGGCGGCAGCG CAAAAGAUGCGUAAA (SEQ ID NO:21)	GGGUCGCACGGUUCGCC UAACCGUGCGAGAAUAAAC UACAUCAUAUCAGCCAUA CACUUUCGAAC (SEQ ID NO:22)	194.4 ± 31.3
11	GGGUAAGAUUUGAUGGCUA UUUGUACGUGUUCGAACAG AGGAGACGAACAAUGACAA AUAGCAACCUGGCGGCAGCG CAAAAGAUGCGUAAA (SEQ ID NO:23)	GGGACCAUUCGCCCACU UGGCGAAUGGUAAGCGAA CACGUACAAAUAGCCAUC AAAUCUUAACU (SEQ ID NO:24)	181.2 ± 13.0
12	GGGAAUUGGAUGAAGGCGG UAAGUAUGAUUGUAGACAG AGGAGAUACAAUAGACUU ACCGCAACCUGGCGGCAGCG CAAAAGAUGCGUAAA (SEQ ID NO:25)	GGGUCCAGUCUAGACUGA CCUAGACUGGACAAUACA AUCAUACUUACCGCCUUC AUCCAUUUACU (SEQ ID NO:26)	169.9 ± 16.7
13	GGGUAGAAUUGAUACUUG AUUUGAUGGCUUGAAACAG AGGAGAUCAAGCAUGCAAA UCAAGAACCUGGCGGCAGCG CAAAAGAUGCGUAAA (SEQ ID NO:27)	GGGUCUAAUGCGAGUAAA GUCGCAUUAGAAUAUCAA GCCAUCAAAUCAAGUAUC AAAUUCUAAAAG (SEQ ID NO:28)	155.3 ± 8.6
15	GGGCGUUAUACUUUGUCGU UCUGCGUGUCGUUAAACAG AGGAGAUAAACGAAUGGCAG AACGAAACCUGGCGGCAGCG CAAAAGAUGCGUAAA (SEQ ID NO:29)	GGGACUGAGGUACCUCGA AGUACCUCAGUAGAUAAAC GACACGCAGAACGACAAA GUUAACGAAA (SEQ ID NO:30)	124.7 ± 11.0
16	GGGUAAGAAUGAUAAAGGU AAGUAGUGAGUUGAAACAG AGGAGAUCAACUAUGUACU UACCUAACCUGGCGGCAGCG CAAAAGAUGCGUAAA (SEQ ID NO:31)	GGGCCAUGACUCCUAAUU CGGAGUCAUGGAAAUCAA CUCACUACUUACCUUUAU CAUUCUUACUU (SEQ ID NO:32)	123.2 ± 23.0
17	GGGCGAUAAAGACUGAGGC UGGGUAUGGUUAGAAACAG AGGAGAUUCUAACAUGACCCA GCCUAACCUGGCGGCAGCGC AAAAGAUGCGUAAA (SEQ ID NO:33)	GGGCCUCGACGUUCGUGA UAACGUCGAGGCAAUCUA ACCAUACCCAGCCUCAGUC UUUAUCGCAA (SEQ ID NO:34)	122.6 ± 0.8
18	GGGAUAGAUGAUUGUGCUU AGUUUAUGAUUCUGAACAG AGGAGACAGAAUUGAAAC UAAGCAACCUGGCGGCAGCG CAAAAGAUGCGUAAA (SEQ ID NO:35)	GGGACAUCCUAAGUUGAC UCUUAGGAUGUAAACAGA AUCUAAAACUAAAGCACA UCAUCUAUACA (SEQ ID NO:36)	115.1 ± 7.9

- 73 -

Toehold switch number	Switch sequence	Trigger sequence	On/off
19	GGGAUAAUGAUGAUGAGUA UGUUGAAGGUGUAAGACAG AGGAGAUUACACAUGCAAC AUACUAACCUGGCGGCAGCG CAAAAGAUGCGUAAA (SEQ ID NO:37)	GGGUCGUGUGAUGCUAUC UCAUCACACGACAAUUAC ACCUUCAACAUACUCAUC AUCAUUAUCAC (SEQ ID NO:38)	103.5 ± 10.5
20	GGGCGUUAUUCUCUGGCUU GCUUUAUGUCUGUAAACAG AGGAGAUACAGAAUGAAAG CAAGCAACCUGGCGGCAGCG CAAAAGAUGCGUAAA (SEQ ID NO:39)	GGGUCAUGACUGGGACAC GCCAGUCAUGAGAAUACA GACAUAAAGCAAGCCAGA GAUUAACGAAG (SEQ ID NO:40)	101.9 ± 8.4
21	GGGUAAAGAUGAAACGCGU GAAUGAUAGUAUUGAACAG AGGAGACAAUACAUGCAU CACGCAACCUGGCGGCAGCG CAAAAGAUGCGUAAA (SEQ ID NO:41)	GGGUCACCUCCAGGCACG ACUGGAGGUGAAUACAAU ACUAUCAUUCACGCGUUU CAUCUUUACUU (SEQ ID NO:42)	86.6 ± 14.5
22	GGGCAUUAAGAUUGUACUU GUAAGAUCGUGUCGAACAG AGGAGACGACACAUGCUUAC AAGUAACCUGGCGGCAGCGC AAAAGAUGCGUAAA (SEQ ID NO:43)	GGGACCUAGCGGAGCGCA GUCCGCUAGGUAAACGAC ACGAUCUUACAAGUACAA UCUUAUGAAA (SEQ ID NO:44)	83.1 ± 2.0
23	GGGUAUGAAUUGAUGUCGA UUGUUAUGUCUUGAGACAG AGGAGAUCAAGAAUGAACA AUCGAAACCUGGCGGCAGCG CAAAAGAUGCGUAAA (SEQ ID NO:45)	GGGACCCGACCGCGUCCUG GCGGUCGGGUAAAUCAAG ACAUAACAAUCGACAUCA AUUCAUACUA (SEQ ID NO:46)	75.2 ± 4.9
24	GGGUAAAGAUGAGAACGCU UGUGAAUGAUGUGAAACAG AGGAGAUACAUUUGUCAC AAGCGAACCUGGCGGCAGCG CAAAAGAUGCGUAAA (SEQ ID NO:47)	GGGUCGGGACACGGGCAU ACGUGUCCCGACAAUCAC AUCAUUCACAAGCGUUCU CAUCUUUACUU (SEQ ID NO:48)	74.8 ± 13.0
25	GGGUGAUAGAUGAAGGCAG GCGUUAUAGUUUAGAACAG AGGAGACUAAACAUGAACG CCUGCAACCUGGCGGCAGCG CAAAAGAUGCGUAAA (SEQ ID NO:49)	GGGUCGUGGGCCUGCCUA AAGGCCACGAGAACUAA ACUAUAACGCCUGCCUUC AUCUAUCAAAC (SEQ ID NO:50)	72.6 ± 3.7
26	GGGUAAGUAAUGAAGUCUA AGUGUAUCGUGUCGGACAG AGGAGACGACACAUGACACU UAGAAACCUGGCGGCAGCGC AAAAGAUGCGUAAA (SEQ ID NO:51)	GGGACGAACUUGAUCAAA UUCAAGUUCGUAAACGAC ACGAUACACUAGACUUC AUUACUUACAU (SEQ ID NO:52)	68.2 ± 10.3
27	GGGAUAGAAUUAGAAAUGA AAUAGAUGGUUACGAACAG AGGAGACGUAAACAUUCU UUCAUAACCUGGCGGCAGCG CAAAAGAUGCGUAAA (SEQ ID NO:53)	GGGUCGCUUGUACUCUUG CGUACAAGCGAAAGCGUA ACCAUCUAUUUCAUUUCU AAUUCUAUCUA (SEQ ID NO:54)	67.3 ± 10.7

- 74 -

Toehold switch number	Switch sequence	Trigger sequence	On/off
28	GGGAUAGAAAUUGAUCGUU AGUUUAUGUUGCCGGACAG AGGAGACGGCAAUUGAAAC UACGAACCUGGCGGCAGCG CAAAAGAUGCGUAAA (SEQ ID NO:55)	GGGUAUCUACUCGACUUC GCGAGUAGAUAAAGCGGC AACAUAAACUAAACGAUCA AUUUCUAUAAC (SEQ ID NO:56)	66.9 ± 8.0
29	GGGAGUUUGAAUAUGGCGA AAUGAAUGCUUUGAAACAG AGGAGAUCAAAGAUGUCAU UUCGCAACCUGGCGGCAGCG CAAAAGAUGCGUAAA (SEQ ID NO:57)	GGGACGAGAGCUCCUAGC AGAGCUCUCGUUAUAUCAA AGCAUUCAUUUCGCCAUUA UUCAAACUAAAC (SEQ ID NO:58)	63.1 ± 6.6
30	GGGUAAGAUGAUAAAGAUG UGAGUAAGGUAGUAAACAG AGGAGAUACUACAUGACUC ACAUCAACCUGGCGGCAGCG CAAAAGAUGCGUAAA (SEQ ID NO:59)	GGGCACGACUAGAAACCG AUCUAGUCGUGCAAUACU ACCUUACUCACAUCUUUAU CAUCUUUACUU (SEQ ID NO:60)	63.0 ± 5.6
31	GGGAUCUAAAUGUAUUCGU UCGUUAUGGUAUUGAACAG AGGAGACAAUACAUGAACG AACGAAACCUGGCGGCAGCG CAAAAGAUGCGUAAA (SEQ ID NO:61)	GGGCGUCGCGUGUUCGGGU AAACAGCGACGGAACAAU ACCAUAACGAACGAAUAC AUUUAGAUACU (SEQ ID NO:62)	62.0 ± 6.2
32	GGGUCAAUUUCGUUUUAGU AUGUUAUGGUUCUGAACAG AGGAGACAGAACAUGAACU UACUAAACCUGGCGGCAGCG CAAAAGAUGCGUAAA (SEQ ID NO:63)	GGGACUCGGUCUCUAUGU CGAGACCGAGUAAACAGA ACCAUAACAUACUAAUAC GAAAUUGAAGC (SEQ ID NO:64)	59.4 ± 7.6
33	GGGAUUUUGGAAGUAGAGU AGUAGAUAGUUAUGAACAG AGGAGACAUAAACUAGCUAC UACUCAACCUGGCGGCAGCG CAAAAGAUGCGUAAA (SEQ ID NO:65)	GGGCAGAGCGGUCUGUUU CGACCGCUCUGAAACAUUA ACUAUCUACUACUCUACU UCCAAAUUCUA (SEQ ID NO:66)	58.5 ± 9.6
34	GGGAGUAAUGAUGAUUAUAG UUUGAAUGUAGUGAAACAG AGGAGAUACUAAUGUCAU ACUAUAACCUGGCGGCAGCG CAAAAGAUGCGUAAA (SEQ ID NO:67)	GGGACCGUCACCCUACAUC GGGUGACGGUAAAUCACU ACAUUCAAACUAUAUCAU CAUUACUAAAC (SEQ ID NO:68)	57.1 ± 5.0
35	GGGAUAAUGGAGAUGGAGU AGGGUAUGAUUGUAGACAG AGGAGAUACAAUUGACCC UACUCAACCUGGCGGCAGCG CAAAAGAUGCGUAAA (SEQ ID NO:69)	GGGCCUAGAAACUGUACG AAGUUUCUAGGCAUUACA AUCAUACCCUACUCCAUCU CCAUAUUCUU (SEQ ID NO:70)	53.9 ± 4.5
36	GGGAUGAAUAUGGACAGUU GAGUAGUGAUGUGAAACAG AGGAGAUACAUUAUGUACU CAACUAACCUGGCGGCAGCG CAAAAGAUGCGUAAA (SEQ ID NO:71)	GGGAGUCAGGACCGCAUC AGGUCCUGACUAAAUCAC AUCACUACUACUGUCC AUUAUUAUCUU (SEQ ID NO:72)	53.4 ± 6.7

- 75 -

Toehold switch number	Switch sequence	Trigger sequence	On/off
37	GGGAUGGAGAUUGAUUAUG AUUGGAUGUGCUUAAACAG AGGAGAUAAAGCAAUGCCAA UCAUAAACCUGGCGGCAGCG CAAAAGAUGCGUAAA (SEQ ID NO:73)	GGGUCGACCGCUCCUGCUC GAGCGGUCGACAAUAAGC ACAUCCAAUCAUAAUCAA UCUCCAUAAC (SEQ ID NO:74)	52.7 ± 1.8
38	GGGAGUAAGAAUUGUGAUA AAGUAAUGUGCGUGAACAG AGGAGACACGCAAUGUACU UUAUCAACCUGGCGGCAGCG CAAAAGAUGCGUAAA (SEQ ID NO:75)	GGGUCACGGAGCGGAUUU GCGCUCCGUGAGAACACG CACAUAUACUUUAUCACAA UUCUUACUUA (SEQ ID NO:76)	51.0 ± 4.5

Table 3. Sequence and performance information for the set of forward engineered toehold switches.

Toehold switch number	Switch sequence	Trigger sequence	On/off
1	GGGUCUUAUCUUAUCUAUCUC GUUUAUCCUGCAUACAGAAA CAGAGGAGAUUAGCAAUGAUA AACGAGAACCUGGCGGCAGCG CAAAAGAUGCGUAAA (SEQ ID NO:77)	GGGACUGACUAUUCUGUG CAAUAGUCAGUAAAGCAG GGAUAAACGAGAUAGAU AGAUAAAGAUAG (SEQ ID NO:78)	665 ± 135
2	GGGAGUUUGAUUACAUUGUCG UUUAGUUUAGUGAUACAUA CAGAGGAGAUUACACAUGACU AAACGAAACCUGGCGGCAGCG CAAAAGAUGCGUAAA (SEQ ID NO:79)	GGGACAGAUCCACUGAGG CGUGGAUCUGUGAACACU AAACUAAACGACAAUGUA AUCAAACUAAAC (SEQ ID NO:80)	586 ± 92
3	GGGAUCUAUUACUACUUAACCA UUGUCUUGCUCUAUACAGAAA CAGAGGAGAUUAGAAUGAGA CAAUGGAACCUGGCGGCAGCG CAAAAGAUGCGUAAA (SEQ ID NO:81)	GGGUGAUGGGACAUUCCG AUGUCCCAUCAUAAGAG CAAGACAAUGGUAAGUAG UAAUAGAUAAAG (SEQ ID NO:82)	557 ± 68
4	GGGCGAUUAUGGAUUAGAGCU CCGUUACUGUCAUACAAGAA CAGAGGAGAUUAGACAUGAAA CGGAGCAACCUGGCGGCAGCG CAAAAGAUGCGUAAA (SEQ ID NO:83)	GGGACGAAUUCACCCUAA UGUGAAUUCGUAAAGACA GUAAACGGAGCUCUAAUC CAUAAUCGAAC (SEQ ID NO:84)	483 ± 40
5	GGGUAUGUAAUUGAUUUGGCU UCUGUUAGUUUCAUACAAGAA CAGAGGAGAUUAGAAAUGAAC AGAAGCAACCUGGCGGCAGCG CAAAAGAUGCGUAAA (SEQ ID NO:85)	GGGUCCAUUCUAGGUGAU ACUAGAAUGGAGCAGAAA CUAACAGAAGCCAAAUCA AUUACAUACUA (SEQ ID NO:86)	453 ± 119

- 76 -

Toehold switch number	Switch sequence	Trigger sequence	On/off
6	GGGCUUAAUCUUACCUUCGCU UGUUCUGUUCCGAUACAGAAA CAGAGGAGAUUUCGGAUGAGA ACAAGCAACCUGGCGGCAGCG CAAAAGAUGCGUAAA (SEQ ID NO:87)	GGGACAAUCGGGACGACA CUCCCGAUUGUGAACGGA ACAGAACAAAGCGAAGGUA AGAUUAAGGUA (SEQ ID NO:88)	409 ± 62
7	GGGUCACUAAUCAUUUGUCG UCGUUUCUAUCUAUACAAGAA CAGAGGAGAUUAGAAUGAAA CGACGAAACCUGGCGGCAGCG CAAAAGAUGCGUAAA (SEQ ID NO:89)	GGGUCGAGUAGACAGAGC UGUCUACUCGAAUAAGAU AGAAACGACGACAAAUGA UUAAGUGAGAA (SEQ ID NO:90)	403 ± 99
8	GGGACCUCUACUUACUCUCAC UCUUACUUCUGCAUAGUAGAA CAGAGGAGAUUUGCAAUGGUA AGAGUGAACCCUGGCGGCAGCG CAAAAGAUGCGUAAA (SEQ ID NO:91)	GGGACUGAGCUGCUAUCA CGCAGCUCAGUAGAGCAG AAGUAAGAGUGAGAGUAA GUAGAGGUAGA (SEQ ID NO:92)	393 ± 41
9	GGGCUUACUACUUUGACACCU GAUUCUGACACGAUACAGAA CAGAGGAGAUUUCGUUAGAGA AUCAGGAACCUGGCGGCAGCG CAAAAGAUGCGUAAA (SEQ ID NO:93)	GGGUCAAUUACCCGUGGU AGGGUAAUUGAAAGCGUG UCAGAAUCAGGUGUCAAA GUAGUAAGUAG (SEQ ID NO:94)	381 ± 32
10	GGGAAUGGAAUGAAUGAACUG CUUGUCUUAUGUAUACAGAAA CAGAGGAGAUUACAAUGGAC AAGCAGAACCUGGCGGCAGCG CAAAAGAUGCGUAAA (SEQ ID NO:95)	GGGCGAAGUGUCCGUUAG AGGACACUUCGACGACAU AAGACAAAGCAGUUCAUUC AUUCCAUUUAG (SEQ ID NO:96)	343 ± 64
11	GGGCGAAUAGAAAUGAAGGCU AGUGUCGUUGUCAUACAGAAA CAGAGGAGAUUAGACAUGGAC ACUAGCAACCUGGCGGCAGCG CAAAAGAUGCGUAAA (SEQ ID NO:97)	GGGACUCCACGCCGACCGU GGCGUGGAGUAAAGACAA CGACACUAGCCUUCAUUU CUAUUCGAUU (SEQ ID NO:98)	292 ± 40
12	GGGCAAUUUCGUUAUUGUUCG UCUUUGCUGUUCUAACAAGAA CAGAGGAGAUUAGAAAUGCAA AGACGAAACCUGGCGGCAGCG CAAAAGAUGCGUAAA (SEQ ID NO:99)	GGGCGACGAUGCCAGGUA UGGCAUCGUCGGACGAAC AGCAAAGACGAACAUUA CGAAAUUGAAA (SEQ ID NO:100)	288 ± 19
13	GGGAUGGAAUUGAGAUGGGCU UUCGCGAGAUUGAUACAGAAA CAGAGGAGAUUCAAUUGCGC GAAAGCAACCUGGCGGCAGCG CAAAAGAUGCGUAAA (SEQ ID NO:101)	GGGUGCGGUAGUAGGUUC CUACUACCGCAAUACAAU CUCGCGAAAGCCCAUCUCA AUUCCAUAUCU (SEQ ID NO:102)	33 ± 3

- 77 -

EQUIVALENTS

While several inventive embodiments have been described and illustrated herein, those of ordinary skill in the art will readily envision a variety of other means and/or structures for performing the function and/or obtaining the results and/or one or more of the advantages described herein, and each of such variations and/or modifications is deemed to be within the scope of the inventive embodiments described herein. More generally, those skilled in the art will readily appreciate that all parameters, dimensions, materials, and configurations described herein are meant to be exemplary and that the actual parameters, dimensions, materials, and/or configurations will depend upon the specific application or applications for which the inventive teachings is/are used. Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific inventive embodiments described herein. It is, therefore, to be understood that the foregoing embodiments are presented by way of example only and that, within the scope of the appended claims and equivalents thereto, inventive embodiments may be practiced otherwise than as specifically described and claimed. Inventive embodiments of the present disclosure are directed to each individual feature, system, article, material, kit, and/or method described herein. In addition, any combination of two or more such features, systems, articles, materials, kits, and/or methods, if such features, systems, articles, materials, kits, and/or methods are not mutually inconsistent, is included within the inventive scope of the present disclosure.

All definitions, as defined and used herein, should be understood to control over dictionary definitions, definitions in documents incorporated by reference, and/or ordinary meanings of the defined terms.

All references, patents and patent applications disclosed herein are incorporated by reference with respect to the subject matter for which each is cited, which in some cases may encompass the entirety of the document.

The indefinite articles “a” and “an,” as used herein in the specification and in the claims, unless clearly indicated to the contrary, should be understood to mean “at least one.”

The phrase “and/or,” as used herein in the specification and in the claims, should be understood to mean “either or both” of the elements so conjoined, i.e., elements that are conjunctively present in some cases and disjunctively present in other cases. Multiple elements listed with “and/or” should be construed in the same fashion, i.e., “one or more” of the elements so conjoined. Other elements may optionally be present other than the elements

- 78 -

specifically identified by the “and/or” clause, whether related or unrelated to those elements specifically identified. Thus, as a non-limiting example, a reference to “A and/or B”, when used in conjunction with open-ended language such as “comprising” can refer, in one embodiment, to A only (optionally including elements other than B); in another embodiment, to B only (optionally including elements other than A); in yet another embodiment, to both A and B (optionally including other elements); etc.

As used herein in the specification and in the claims, “or” should be understood to have the same meaning as “and/or” as defined above. For example, when separating items in a list, “or” or “and/or” shall be interpreted as being inclusive, i.e., the inclusion of at least one, but also including more than one, of a number or list of elements, and, optionally, additional unlisted items. Only terms clearly indicated to the contrary, such as “only one of” or “exactly one of,” or, when used in the claims, “consisting of,” will refer to the inclusion of exactly one element of a number or list of elements. In general, the term “or” as used herein shall only be interpreted as indicating exclusive alternatives (i.e. “one or the other but not both”) when preceded by terms of exclusivity, such as “either,” “one of,” “only one of,” or “exactly one of.” “Consisting essentially of,” when used in the claims, shall have its ordinary meaning as used in the field of patent law.

As used herein in the specification and in the claims, the phrase “at least one,” in reference to a list of one or more elements, should be understood to mean at least one element selected from any one or more of the elements in the list of elements, but not necessarily including at least one of each and every element specifically listed within the list of elements and not excluding any combinations of elements in the list of elements. This definition also allows that elements may optionally be present other than the elements specifically identified within the list of elements to which the phrase “at least one” refers, whether related or unrelated to those elements specifically identified. Thus, as a non-limiting example, “at least one of A and B” (or, equivalently, “at least one of A or B,” or, equivalently “at least one of A and/or B”) can refer, in one embodiment, to at least one, optionally including more than one, A, with no B present (and optionally including elements other than B); in another embodiment, to at least one, optionally including more than one, B, with no A present (and optionally including elements other than A); in yet another embodiment, to at least one, optionally including more than one, A, and at least one, optionally including more than one, B (and optionally including other elements); etc.

- 79 -

It should also be understood that, unless clearly indicated to the contrary, in any methods claimed herein that include more than one step or act, the order of the steps or acts of the method is not necessarily limited to the order in which the steps or acts of the method are recited.

5 In the claims, as well as in the specification above, all transitional phrases such as “comprising,” “including,” “carrying,” “having,” “containing,” “involving,” “holding,” “composed of,” and the like are to be understood to be open-ended, i.e., to mean including but not limited to. Only the transitional phrases “consisting of” and “consisting essentially of” shall be closed or semi-closed transitional phrases, respectively, as set forth in the United
10 States Patent Office Manual of Patent Examining Procedures, Section 2111.03.

- 80 -

What is claimed is:

CLAIMS

1. A system comprising
a host cell having, integrated or encoded into its genome, a plurality of riboregulators,
5 each riboregulator comprising an RNA comprising
 - (i) a single-stranded toehold domain,
 - (ii) a fully or partially double-stranded stem domain comprising an
initiation codon,
 - (iii) a loop domain comprising a ribosome binding site, and
 - 10 (iv) a coding domain.
2. A system comprising
a host cell having, integrated or encoded into its genome, a plurality of riboregulators,
each riboregulator comprising an RNA comprising
15
 - (i) a single-stranded toehold domain,
 - (ii) a fully or partially double-stranded stem domain comprising an
initiation codon, and
 - (iii) a loop domain comprising a ribosome binding site,
wherein each riboregulator is integrated upstream of an endogenous coding sequence,
20 and wherein expression of the endogenous coding sequence is controlled by the riboregulator.
3. The system of claim 1 or 2, wherein the host cell is a prokaryotic cell.
4. The system of claim 1 or 2, wherein the host cell is a bacterial cell.
- 25 5. The system of claim 1 or 2, wherein the host cell is an *E. coli* bacterium.
6. The system of any one of claims 1-5, wherein the plurality is 5-15 or 10-15.
- 30 7. The system of any one of claims 1-6, wherein riboregulators within the plurality are
separated from each other by 0-30 nucleotides, or 9-15 nucleotides.

- 81 -

8. The system of any one of claims 1-7, wherein the riboregulator further comprises a spacer domain.

9. The system of claim 8, wherein the spacer domain is encodes low molecular weight amino acids.

10. The system of claim 8 or 9, wherein the spacer domain is about 9-33 nucleotides in length, or about 21 nucleotides in length.

11. The system of any one of claims 8-10, wherein the spacer domain is situated between the stem domain and the coding domain.

12. The system of any one of claims 1-11, wherein the initiation codon is wholly or partially present in a 1-3 nucleotide bulge in the stem domain.

13. The system of any one of claims 1-12, wherein the stem domain comprises sequence upstream (5') and/or downstream (3') of the initiation codon.

14. The system of claim 12, wherein the sequence upstream of the initiation codon is about 6 nucleotides.

15. The system of claim 13 or 14, wherein the sequence downstream of the initiation codon is about 9 nucleotides.

16. The system of any one of claims 13-15, wherein the sequence downstream of the initiation codon does not encode a stop codon.

17. The system of any one of claims 1-16, wherein the coding domain encodes a reporter protein.

18. The system of claim 17, wherein the reporter protein is green fluorescent protein (GFP).

- 82 -

19. The system of any one of claims 1-16, wherein the coding domain encodes a non-reporter protein.

20. The system of any one of claims 1-19, wherein the toehold domain is complementary
5 in sequence to a naturally occurring RNA.

21. The system of any one of claims 1-19, wherein the toehold domain is complementary in sequence to a non-naturally occurring RNA.

10 22. The system of any one of claims 1-21, wherein the first domain is 100% complementary to the toehold domain.

23. The system of claim 1 or 2, further comprising
a plurality of trans-activating RNA (taRNA), each comprising

15 (i) a first domain that hybridizes to the toehold domain of one of the riboregulators in the plurality and that comprises no or minimal secondary structure, and

(ii) a second domain that hybridizes to a sequence downstream (3') of the toehold domain,

20 wherein each taRNA has a cognate riboregulator.

24. A method of detecting presence of a plurality of RNA in a sample, comprising combining a plurality of riboregulators with a sample, wherein each riboregulator (i) comprises a toehold domain that is complementary to an endogenous RNA and (ii) a coding
25 domain that encodes a reporter protein, under conditions that allow translation of the coding domain in the presence of the endogenous RNA but not in the absence of the endogenous RNA, and

detecting the reporter protein as an indicator of the endogenous RNA,

wherein each riboregulator detects a different endogenous RNA from all other

30 riboregulators in the plurality, and each riboregulator encodes a different reporter protein from all other riboregulators in the plurality.

- 83 -

25. A method of detecting presence of a plurality of RNA in a cell, comprising
introducing into the cell a plurality of riboregulators, wherein each riboregulator
comprises (i) a toehold domain that is complementary to an endogenous RNA in the cell and
(ii) a coding domain that encodes a reporter protein, and

5 detecting the reporter protein as an indicator of the endogenous RNA,
wherein each riboregulator detects a different endogenous RNA from all other
riboregulators in the plurality, and each riboregulator encodes a different reporter protein
from all other riboregulators in the plurality.

10 26. The method of claim 24 or 25, wherein amount of reporter protein is an indicator of
amount of endogenous RNA.

27. A method of controlling gene and/or protein expression in a cell comprising
integrating or encoding a plurality of riboregulators into the genome of the cell, each
15 riboregulator integrated upstream of a target coding sequence,
modulating expression of one or more of plurality of trans-activating RNA (taRNA)
in the cell,

wherein expression of a taRNA in the cell results in increased expression of the target
coding sequence, and

20 wherein each riboregulator comprises an RNA comprising

(i) a single-stranded toehold domain,

(ii) a fully or partially double-stranded stem domain comprising an
initiation codon, and

(iii) a loop domain comprising a ribosome binding site,

25 wherein each taRNA comprises

(i) a first domain that hybridizes to the toehold domain of one of the
riboregulators in the plurality and that comprises no or minimal secondary
structure, and

(ii) a second domain that hybridizes to a sequence downstream (3') of
30 the toehold domain.

28. The method of claim 27, wherein expression of a plurality of target coding sequences
is controlled.

- 84 -

29. The method of claim 28, wherein the plurality of target coding sequences encode proteins that interact with each other directly or indirectly.

30. The method of any one of claims 27-29, wherein the plurality of taRNA are integrated or encoded in the host cell genome.

31. The method of any one of claims 27-30, wherein each taRNA is operably linked to an inducible promoter that is different from all the other taRNA in the plurality.

32. The method of any one of claims 27-31, wherein each taRNA has a cognate riboregulator in the cell.

33. The method of any one of claims 27-31, wherein at least one taRNA activates two or more riboregulators in the cell.

34. A method of controlling gene and/or protein expression in a cell comprising introducing a plurality of riboregulators into a cell, each riboregulator comprises an RNA comprising

- (i) a single-stranded toehold domain,
- (ii) a fully or partially double-stranded stem domain comprising an initiation codon,
- (iii) a loop domain comprising a ribosome binding site, and
- (iv) a coding domain for a reporter protein or a protein of interest,

modulating expression of one or more of a plurality of trans-activating RNA (taRNA) in the cell,

wherein expression of a taRNA in the cell results in increased expression of the coding domain, and

wherein each taRNA comprises

- (i) a first domain that hybridizes to the toehold domain of one of the riboregulators in the plurality and that comprises no or minimal secondary structure, and

- 85 -

(ii) a second domain that hybridizes to a sequence downstream (3') of the toehold domain.

35. The method of claim 34, wherein one or more riboregulators comprise a coding domain for a transcription factor.

36. The method of claim 34 or 35, wherein modulating expression of one or more of a plurality of taRNA comprises increasing expression of a subset of taRNA substantially simultaneously.

37. A system comprising a plurality of riboregulators upstream of a coding domain, each riboregulator comprising

(i) a single-stranded toehold domain,

(ii) a fully or partially double-stranded stem domain comprising an initiation codon, and

(iii) a loop domain comprising a ribosome binding site,

wherein the riboregulators are separated from each other by a spacer of 9-15 nucleotides in length.

38. The system of claim 37, further comprising a plurality of trans-activating RNA (taRNA), wherein a different taRNA or a different subset of taRNAs is required to activate each of the riboregulators.

39. The system of claim 38, wherein a different subset of taRNAs is required to activate each of the riboregulators, and the members of each subset of taRNAs hybridize to each other to form a complex that is capable of hybridizing to the toehold domain of a riboregulator.

40. The system of claim 38, wherein a different subset of taRNAs is required to activate each of the riboregulators, and at least two members of each subset of taRNAs are partially complementary to a toehold domain and/or to the sequence downstream of the toehold domain in a single riboregulator.

- 86 -

41. The system of any one of claims 37-40, wherein the plurality of riboregulators is 5 or 6.

42. The system of any one of claims 38-41, wherein the subset of taRNAs comprises 2
5 taRNAs.

1/57

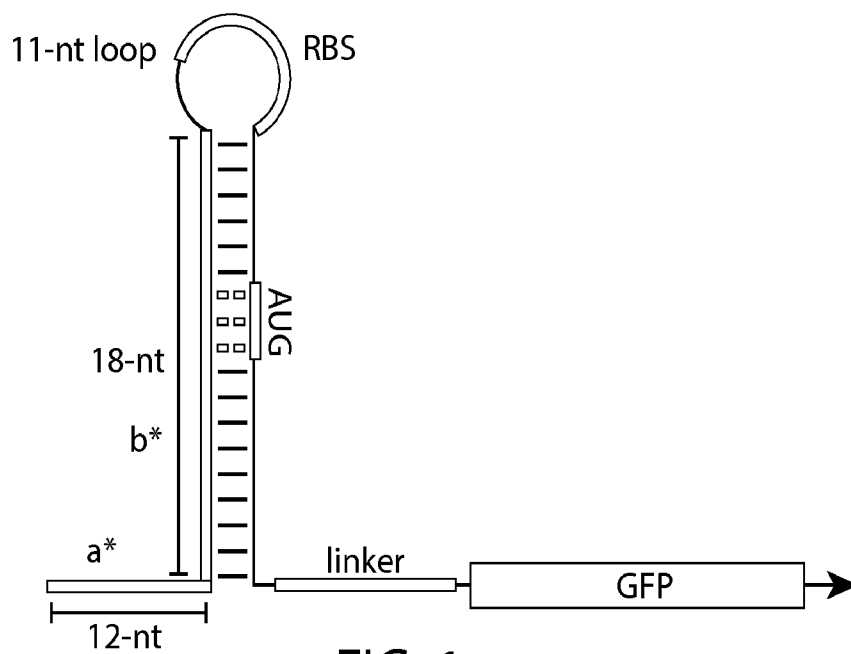


FIG. 1

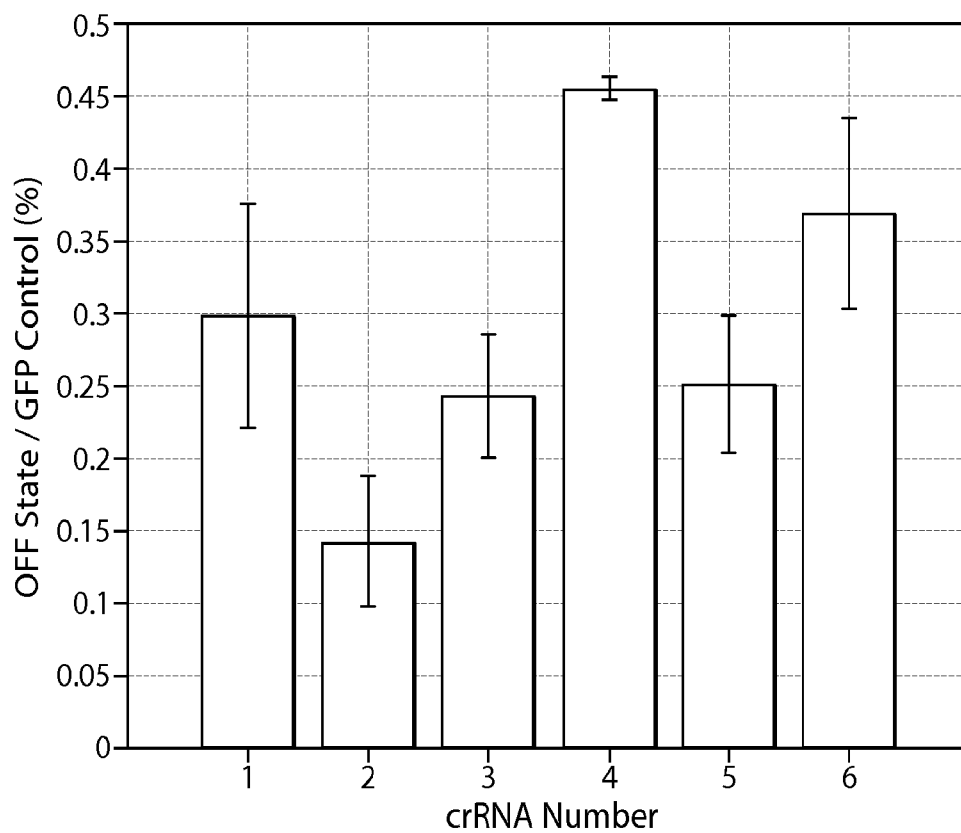


FIG. 2

2/57

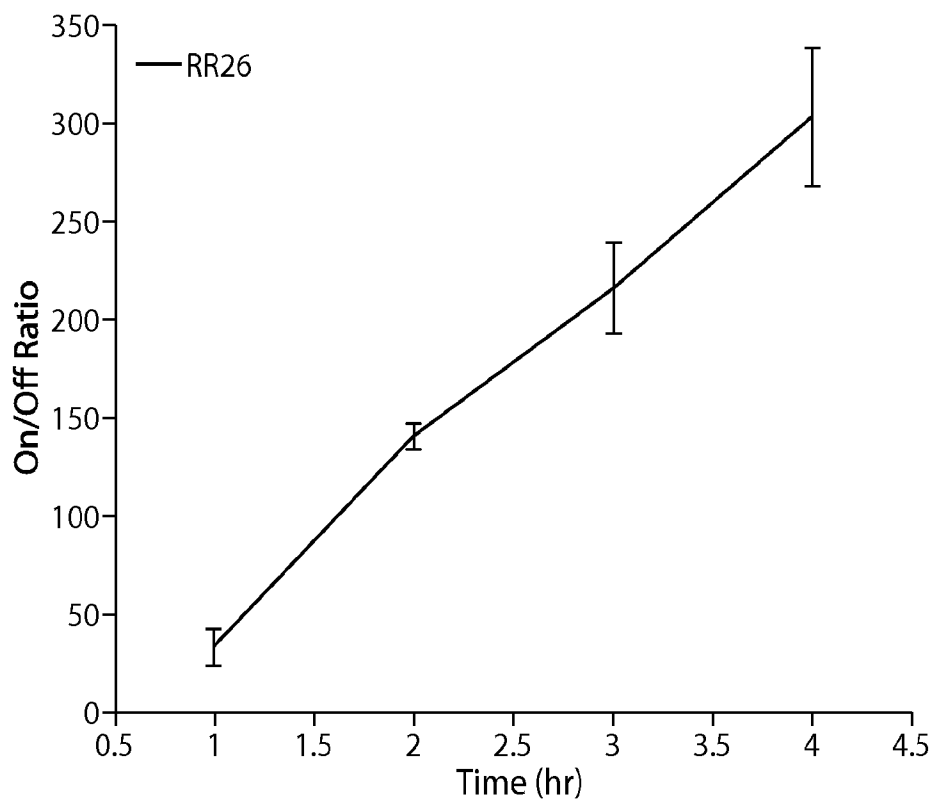


FIG. 3

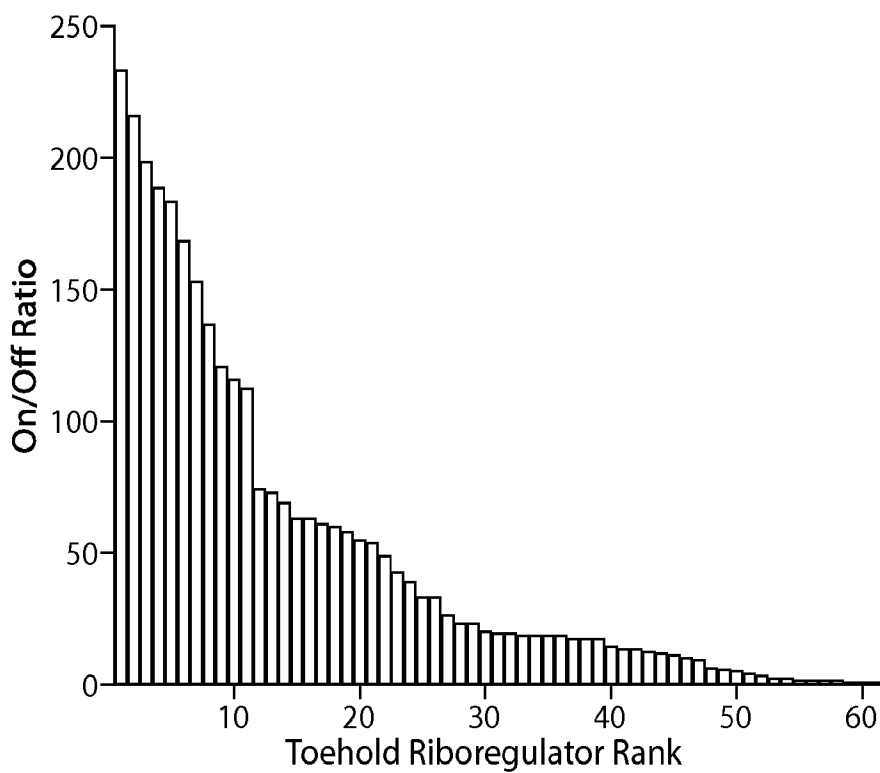


FIG. 4

3/57

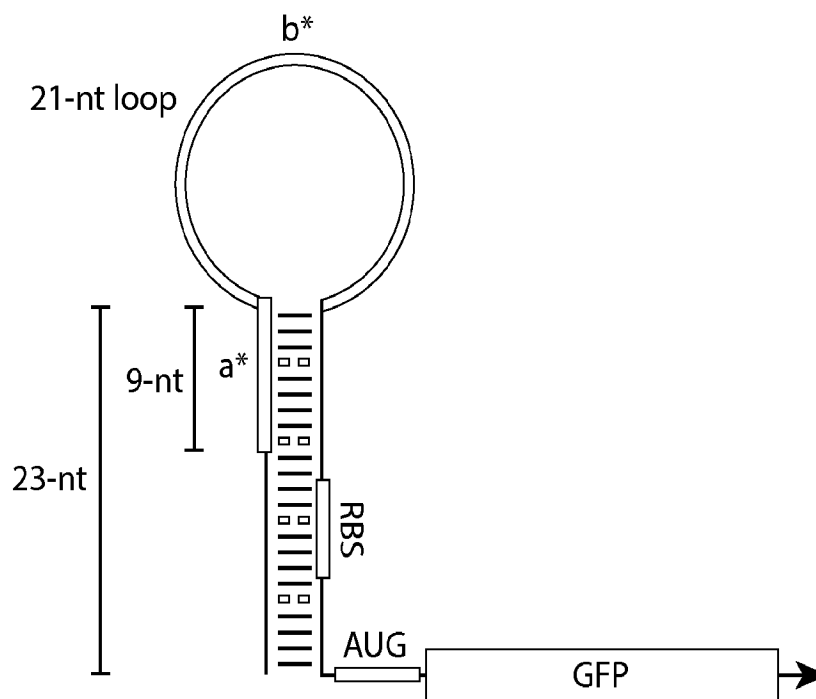


FIG. 5

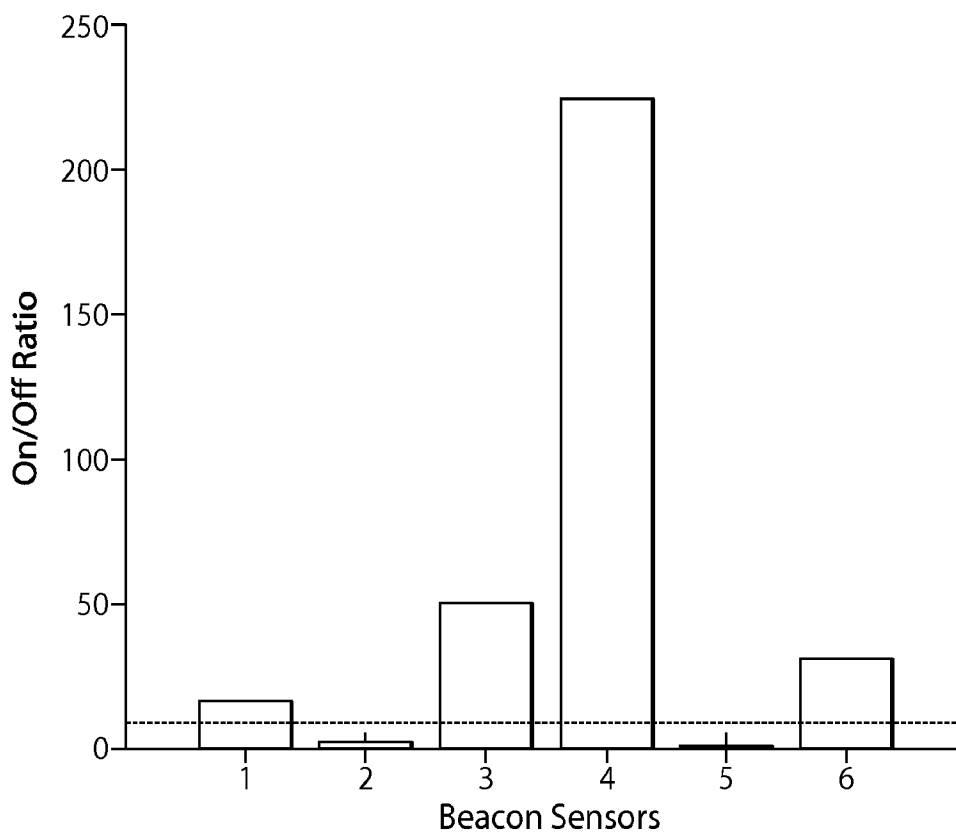


FIG. 6

4/57

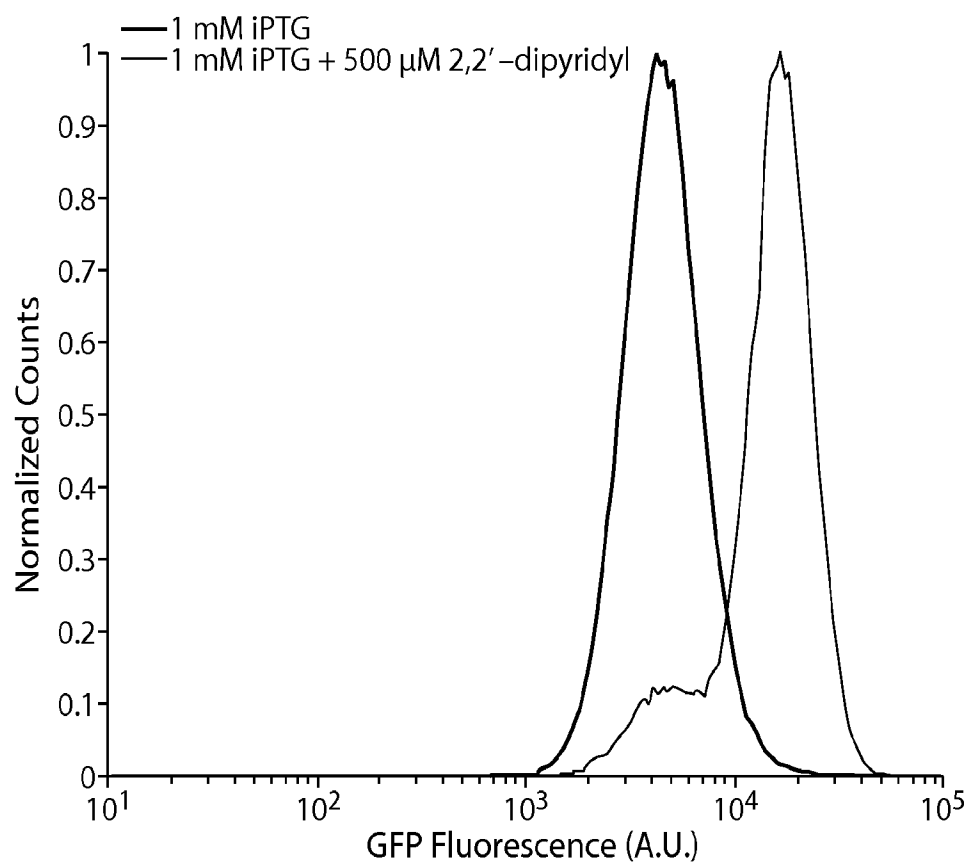


FIG. 7

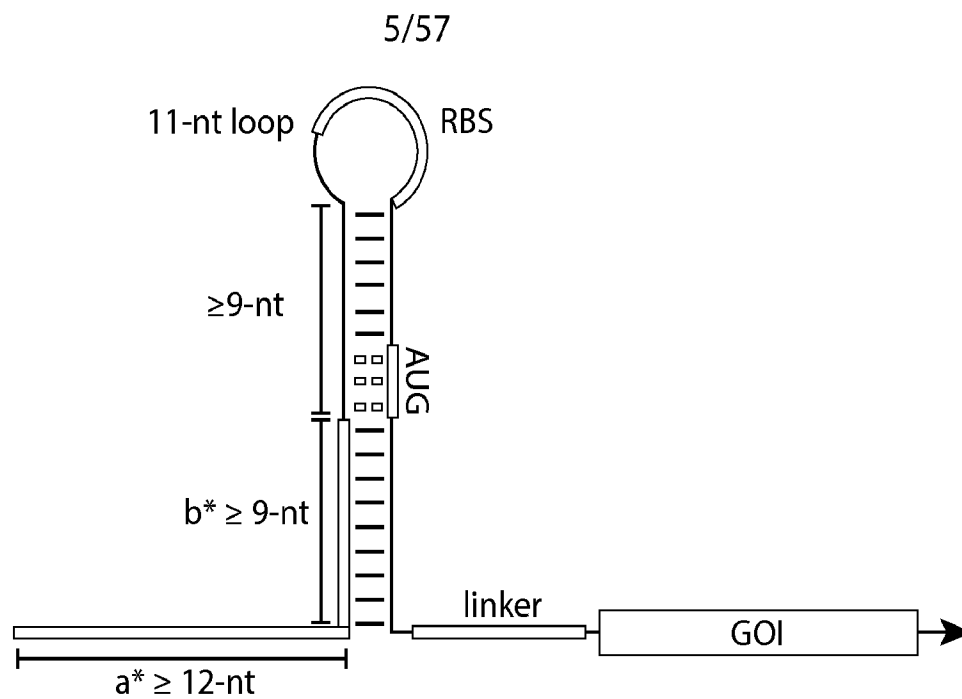


FIG. 8A

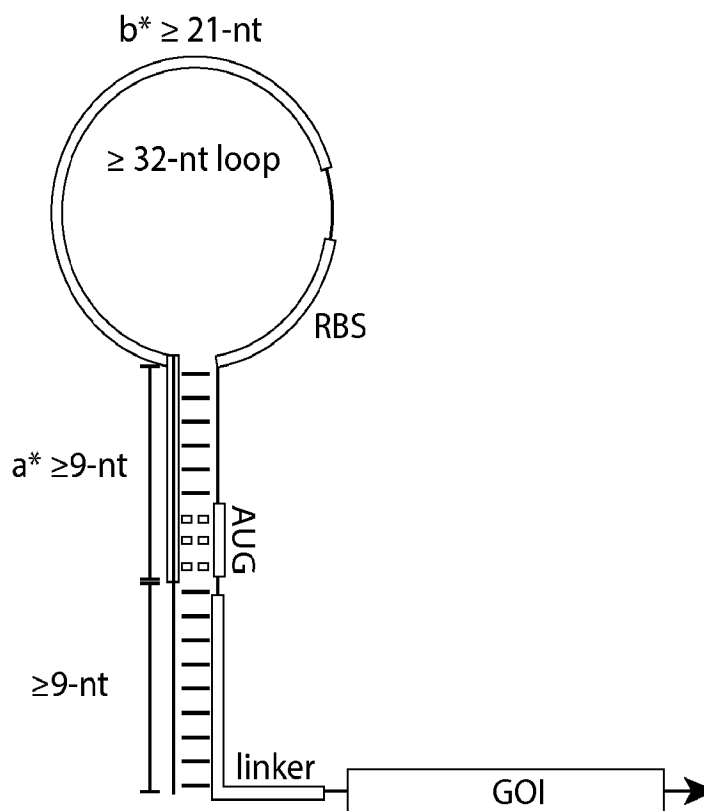


FIG. 8B

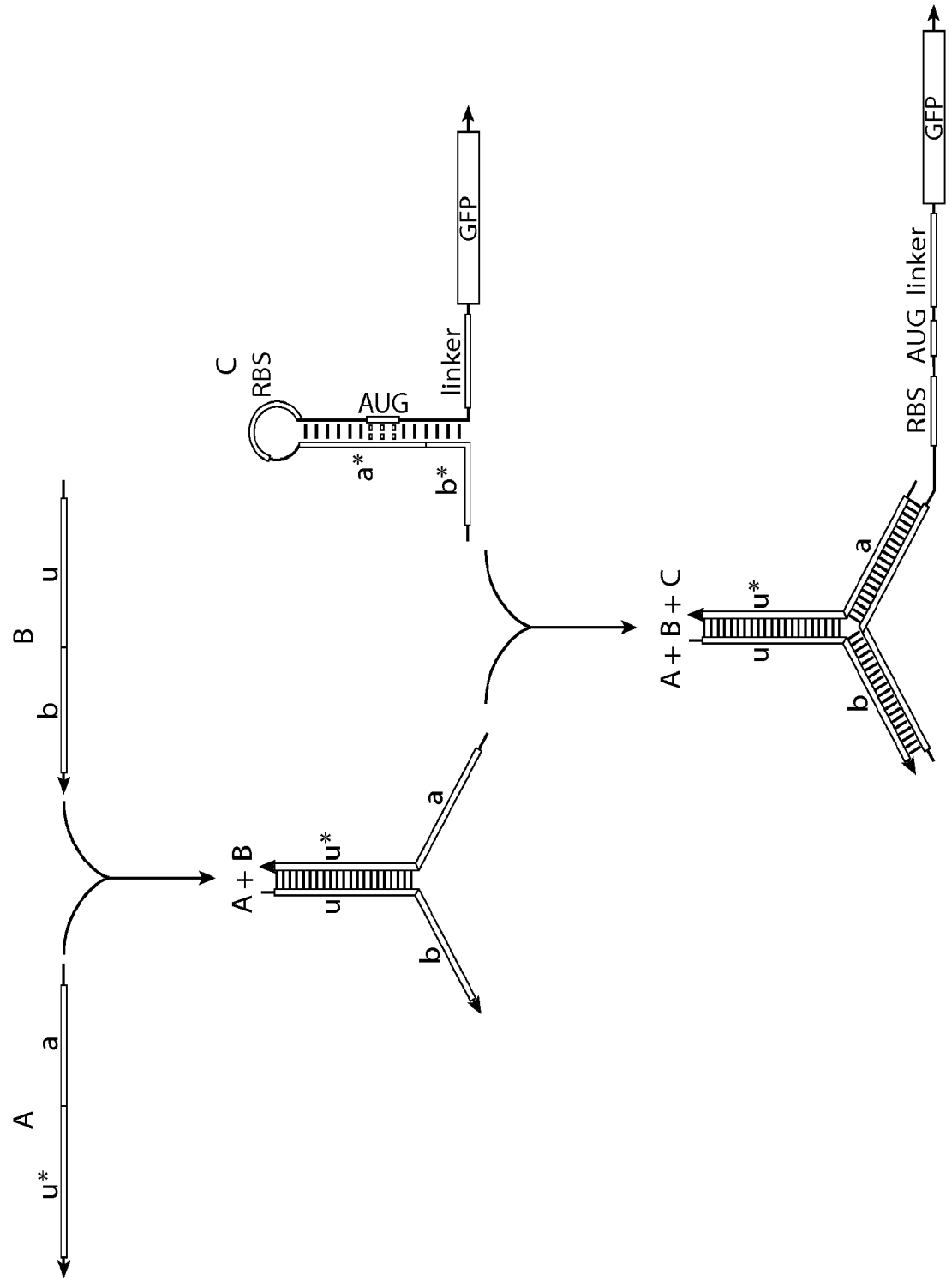


FIG. 9A

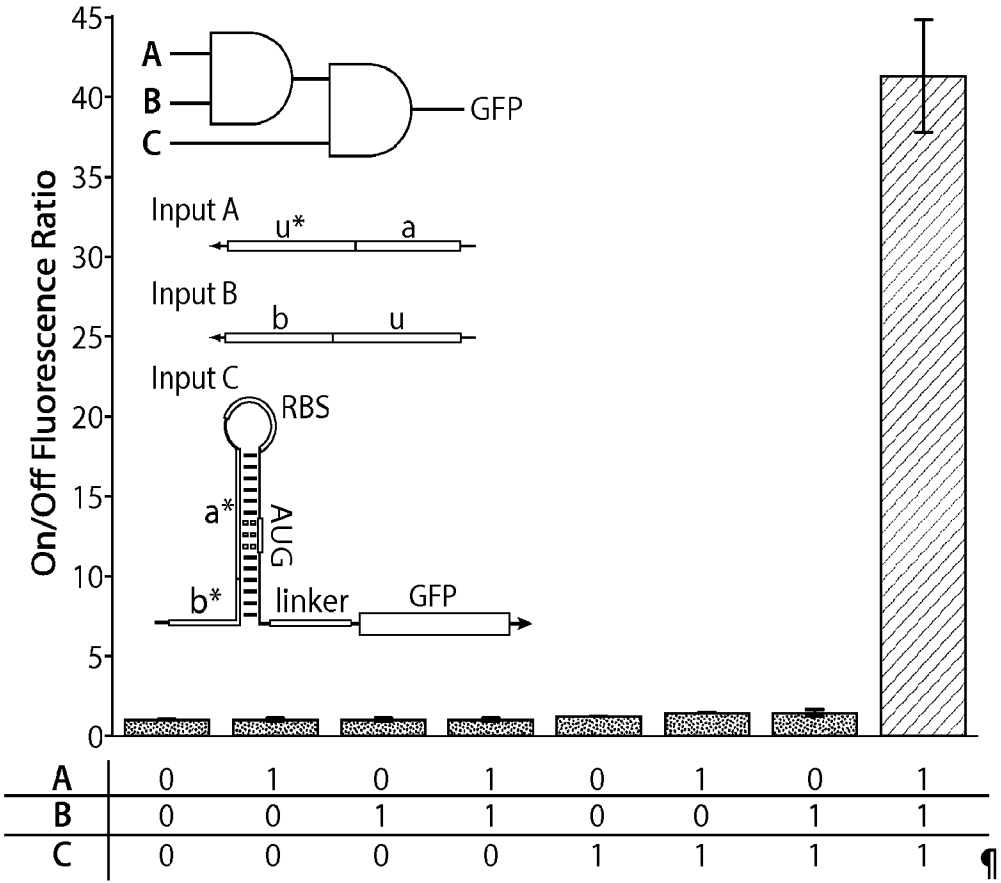


FIG. 9B

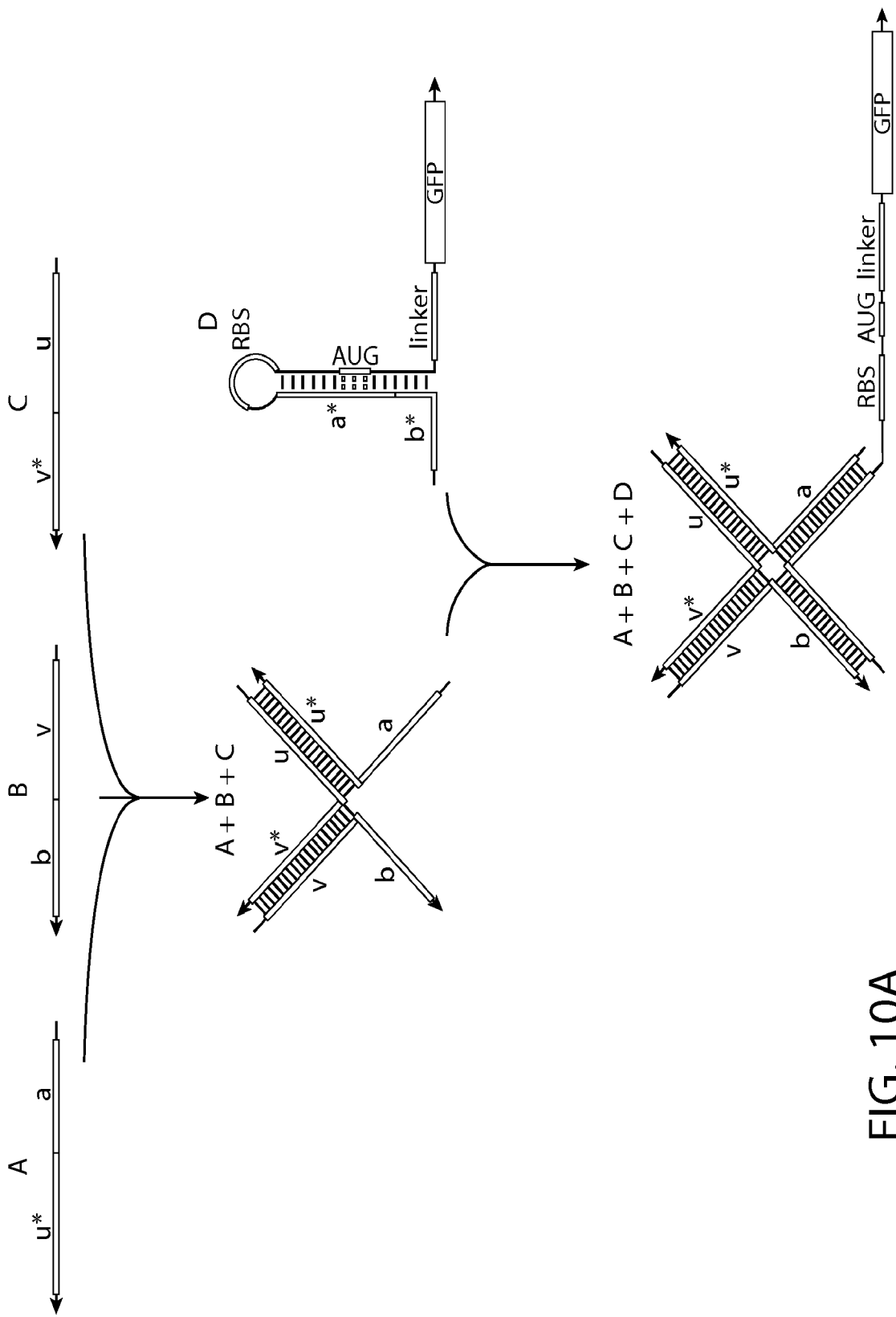


FIG. 10A

9/57

3-input AND

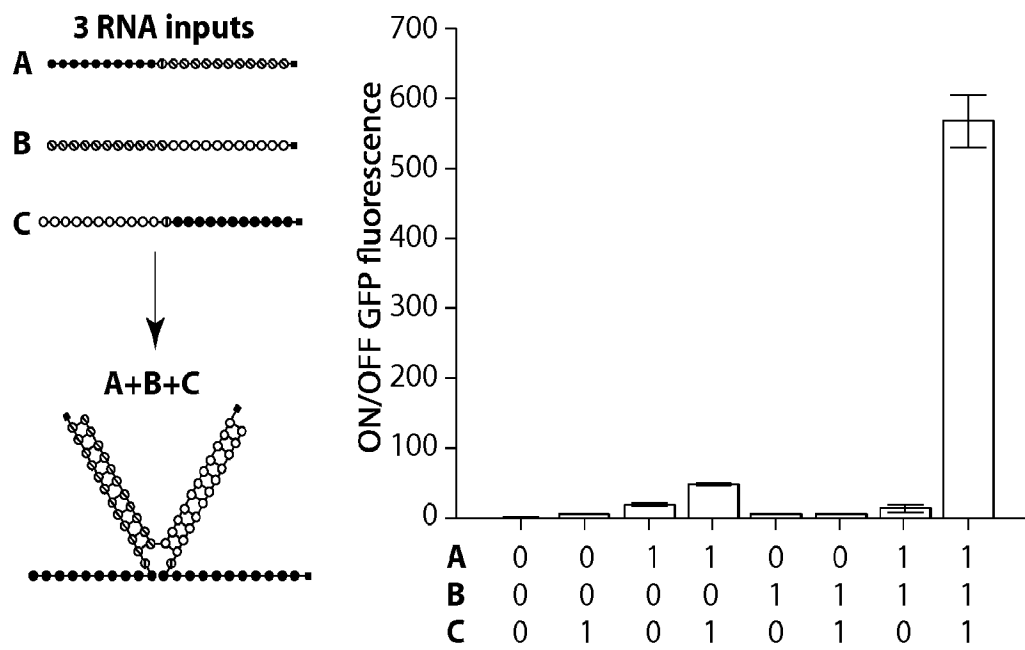


FIG. 10B

4-input AND

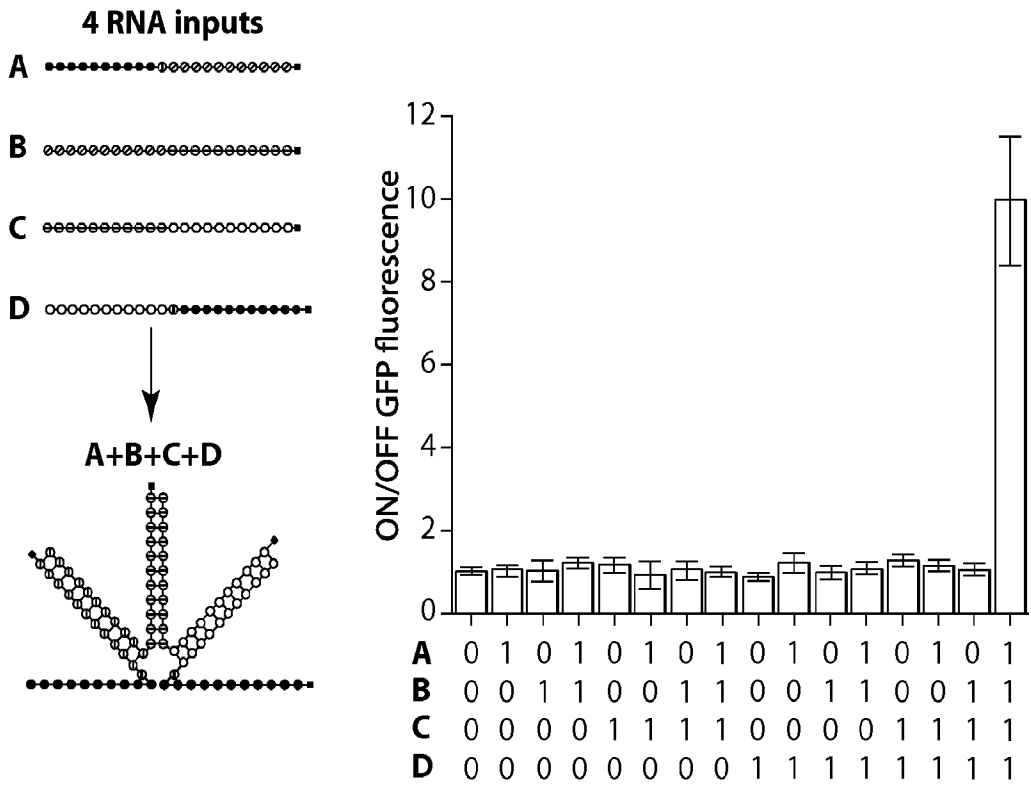


FIG. 10C

11/57

taRNA Input A



taRNA Input B



crRNA OR Gate C

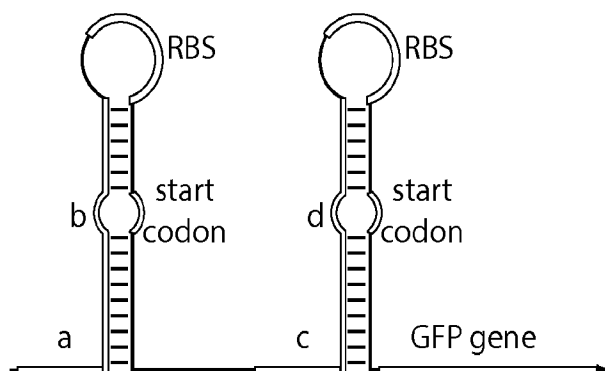


FIG. 11A

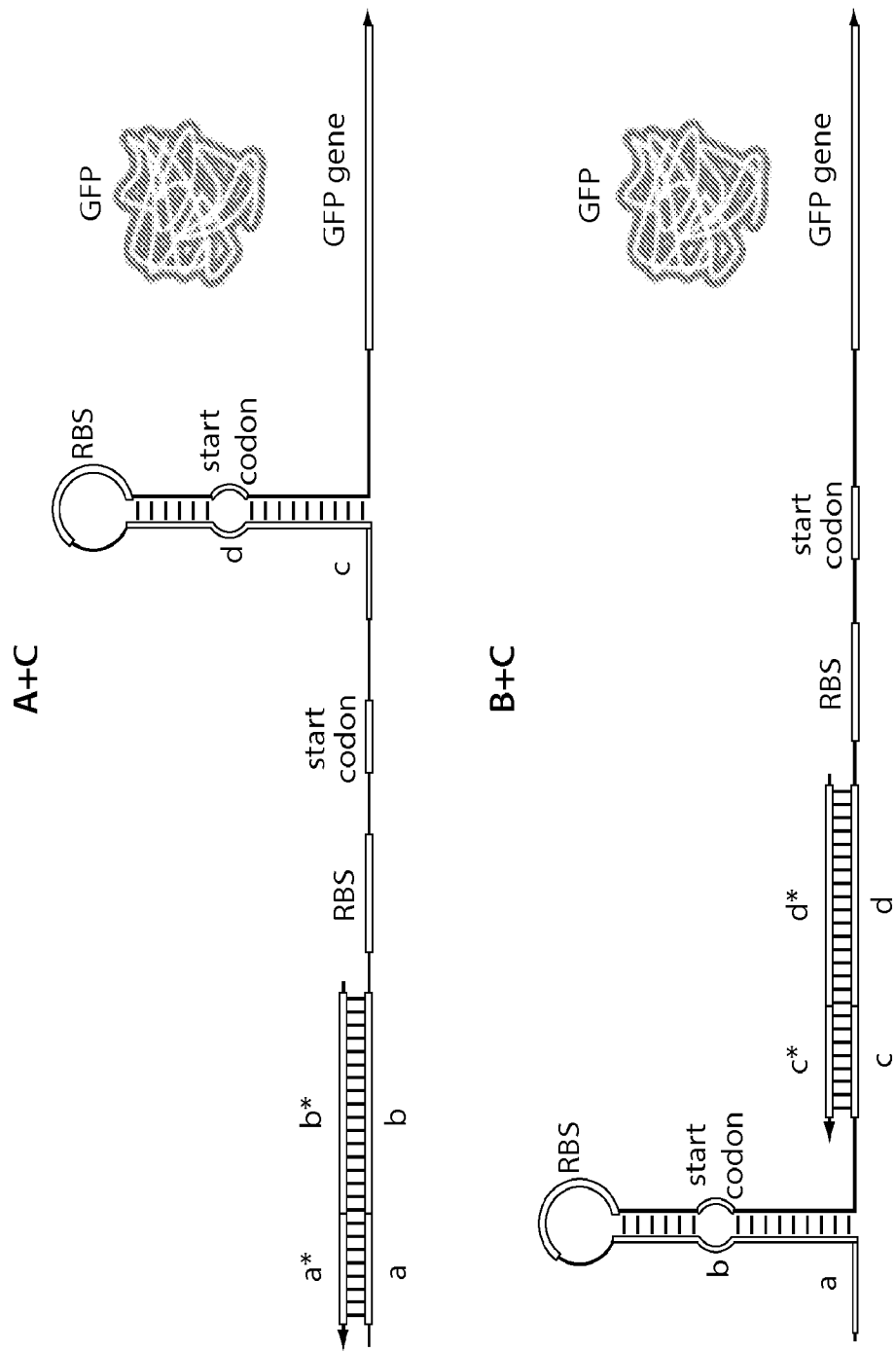


FIG. 11B

13/57

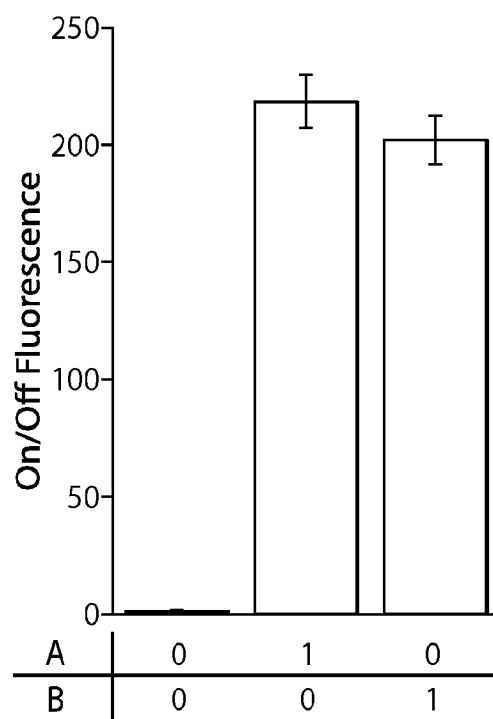


FIG. 11C

14/57

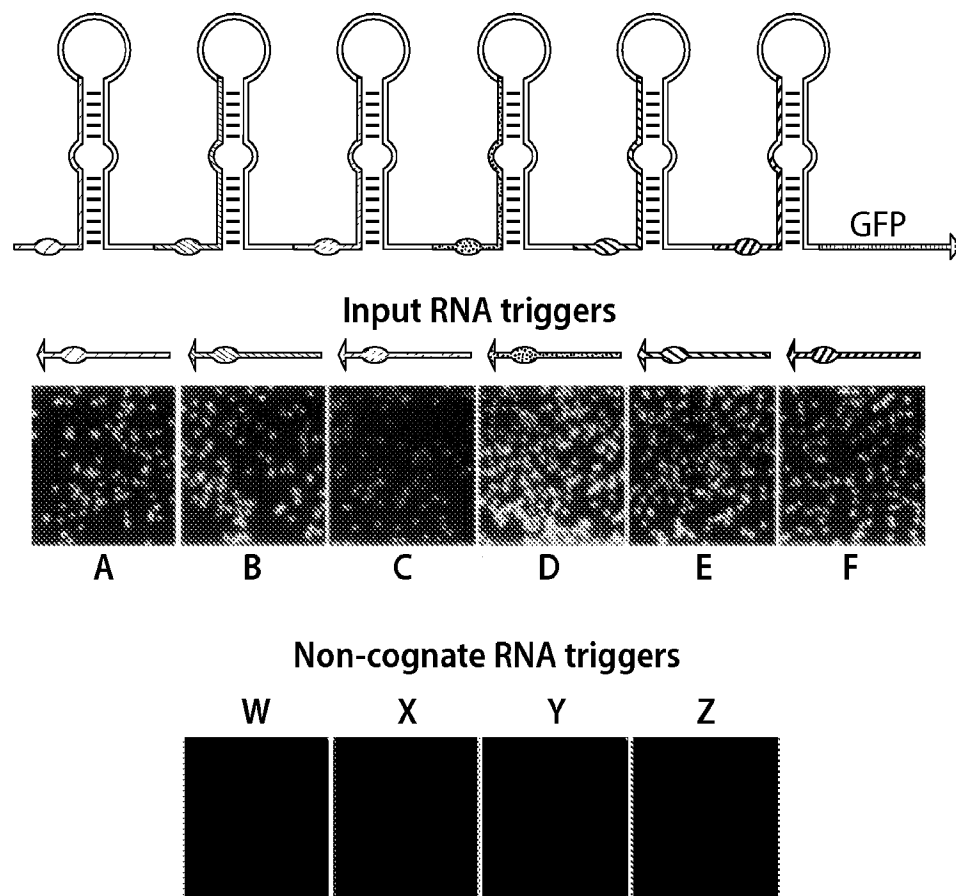


FIG. 12A

15/57

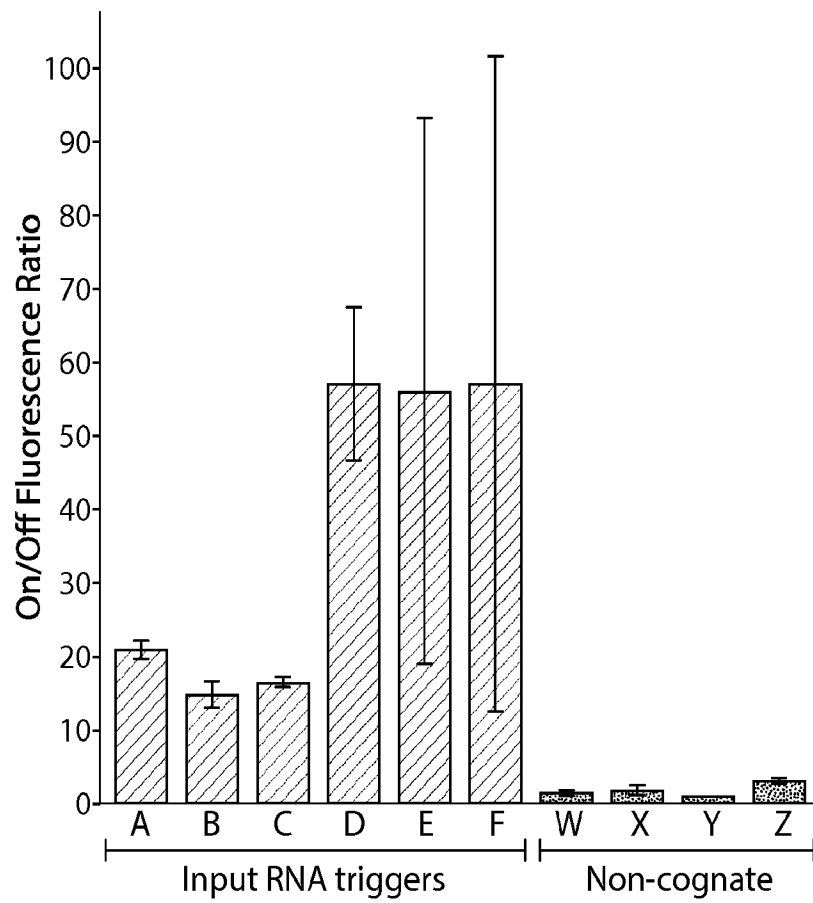


FIG. 12B

16/57

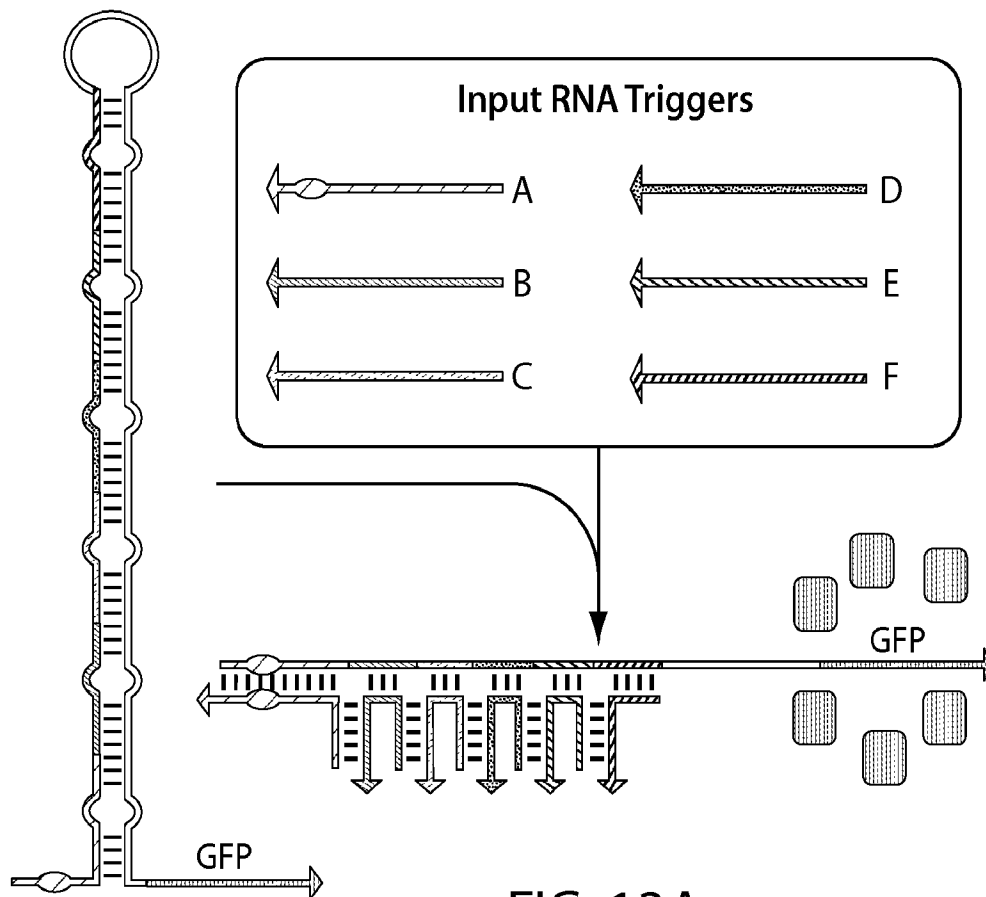
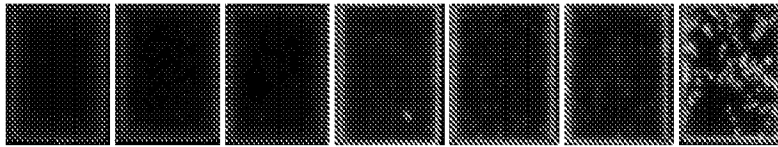


FIG. 13A



A	0	0	1	1	1	1	1
B	0	1	0	1	1	1	1
C	0	0	1	1	1	1	1
D	0	1	0	1	1	1	1
E	0	0	1	0	0	1	1
F	0	1	0	0	1	0	1

FIG. 13B

17/57

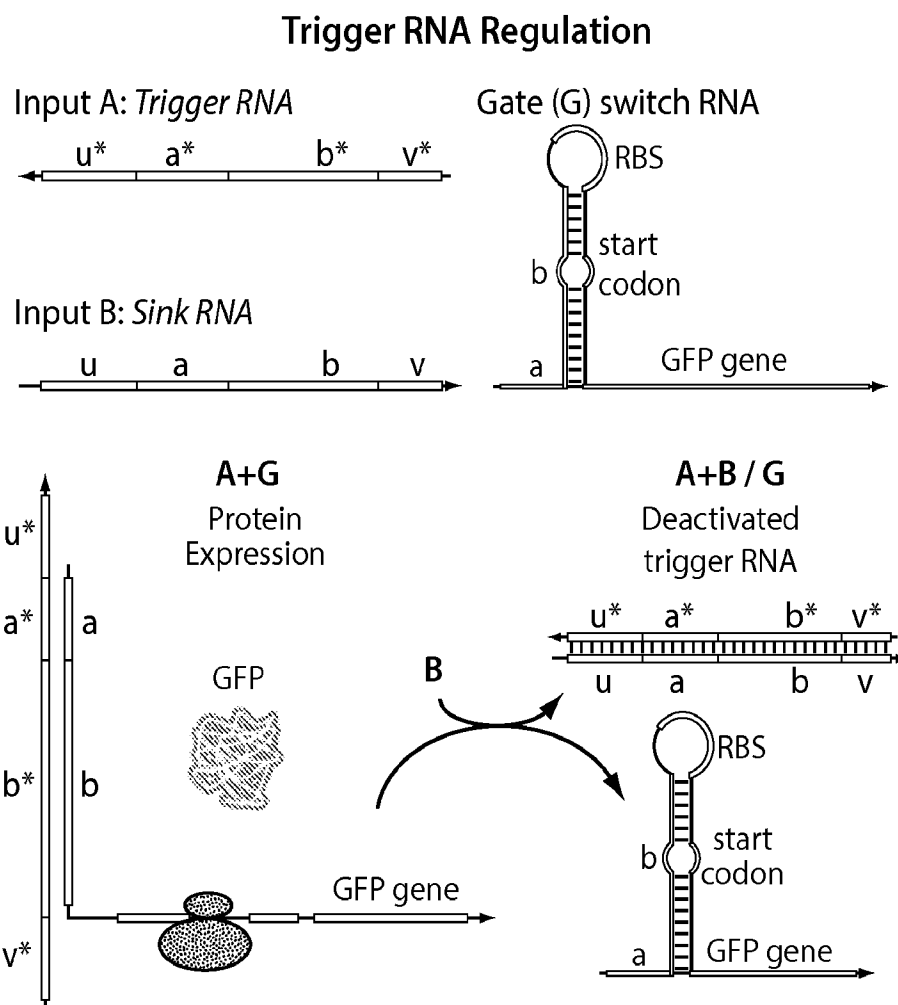


FIG. 14A

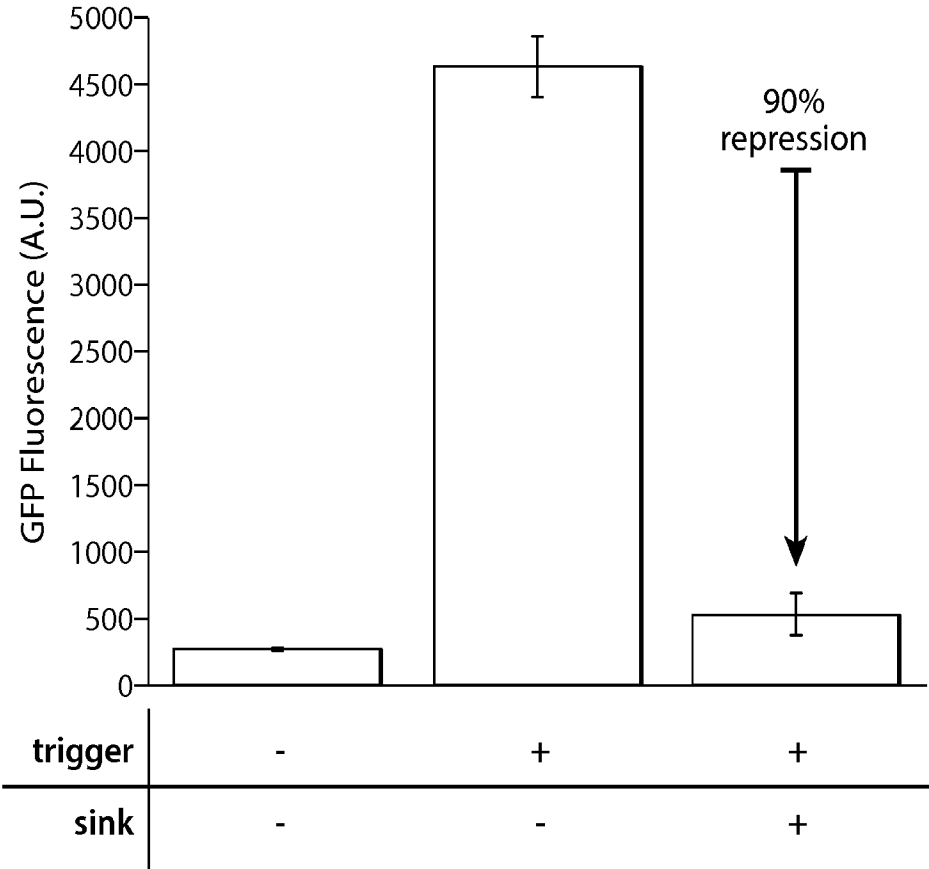


FIG. 14B

19/57

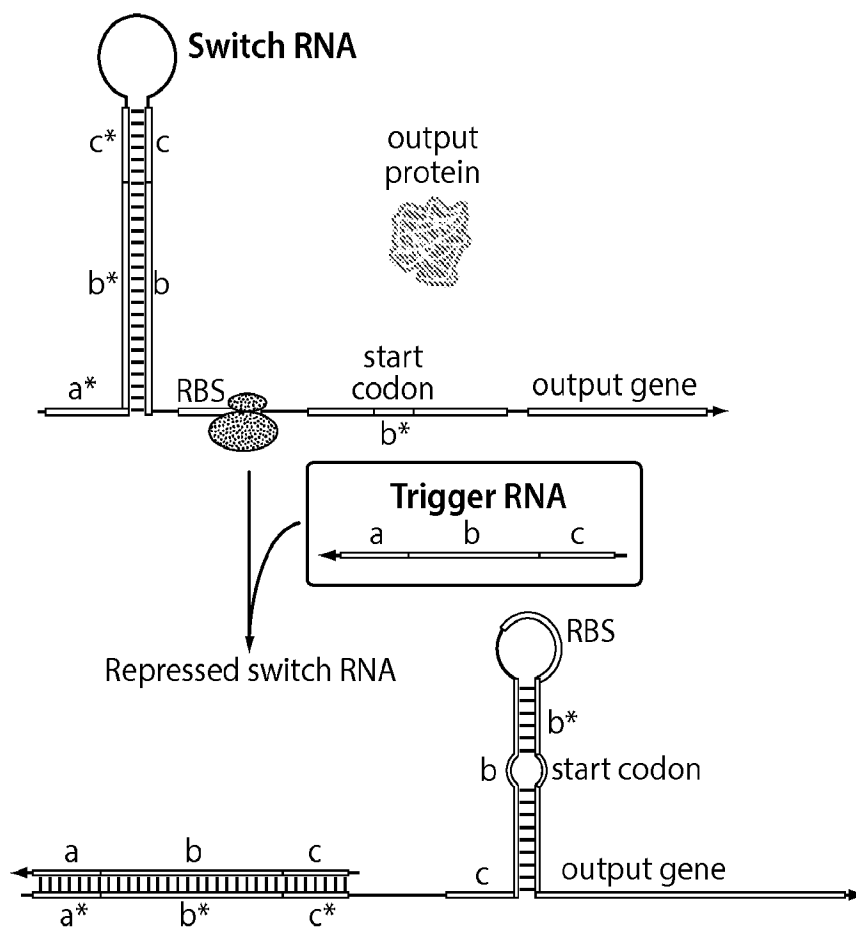


FIG. 15A

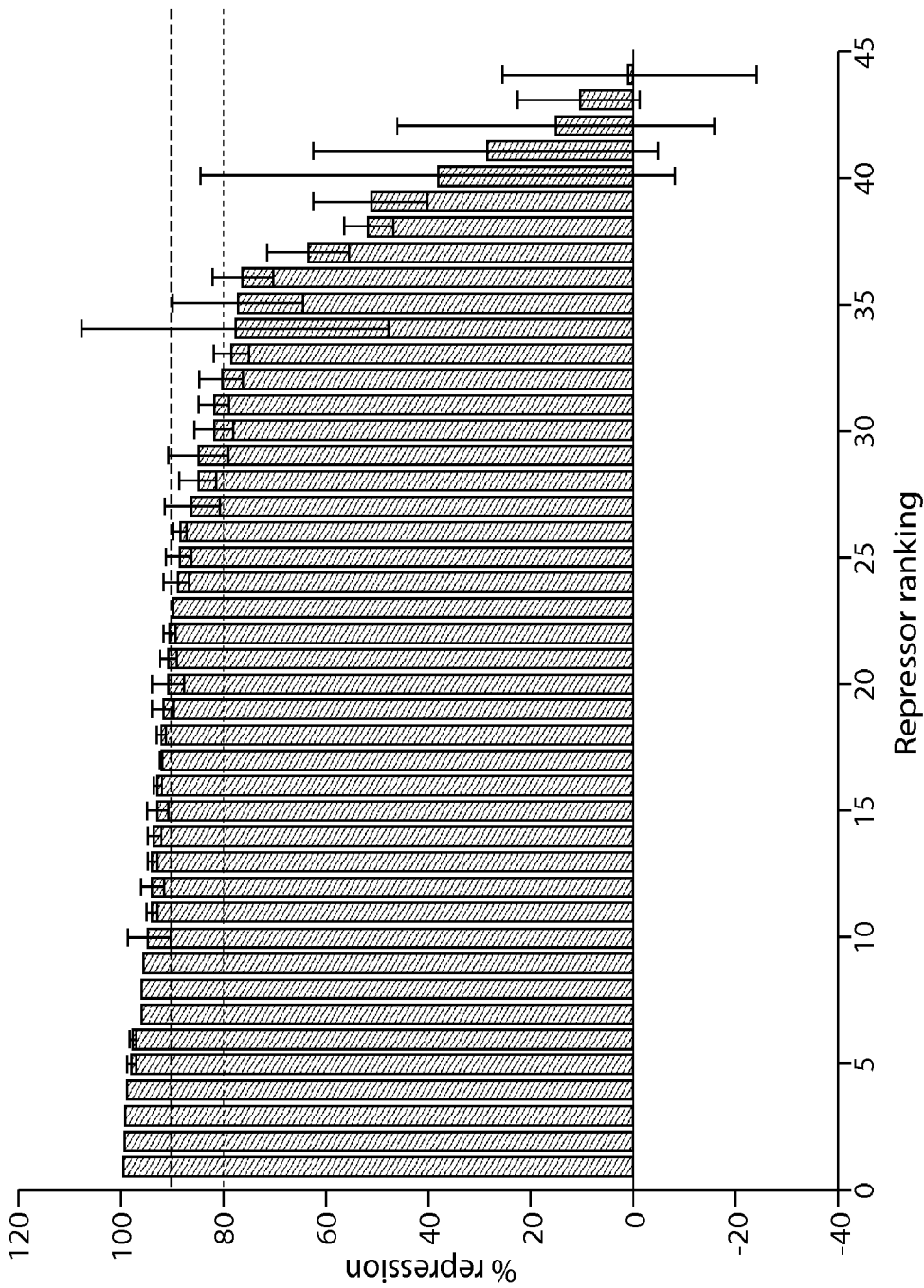


FIG. 15B

21/57

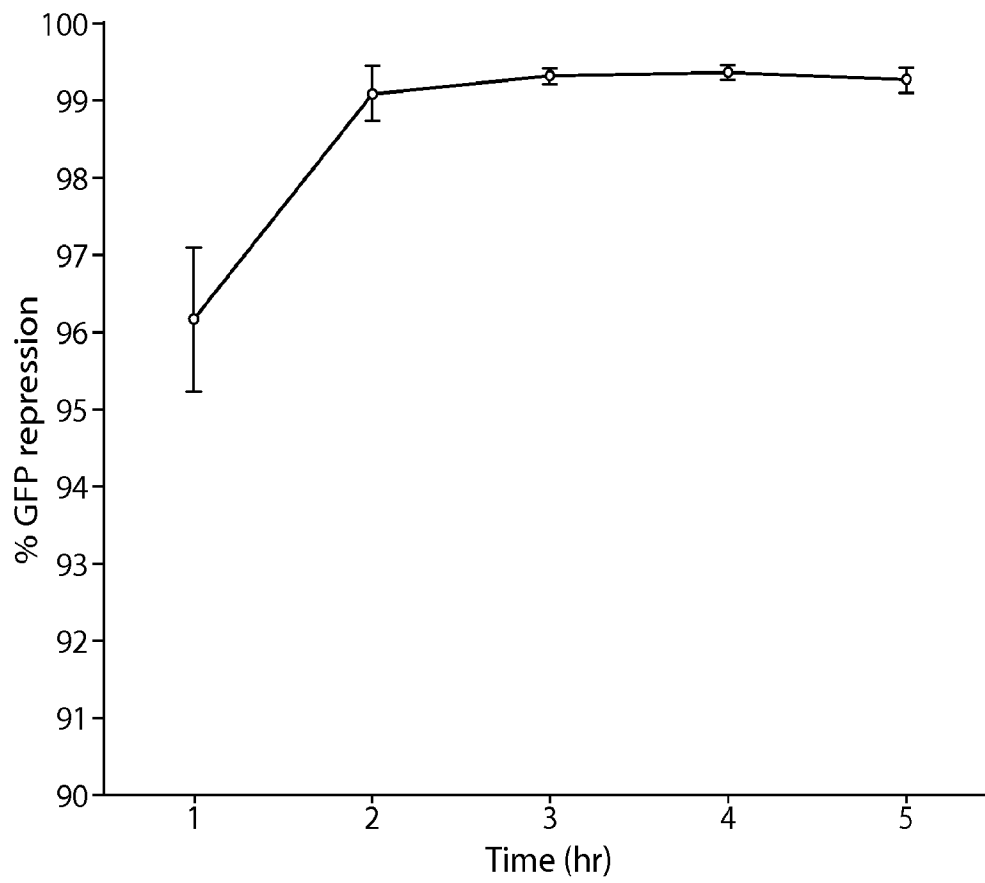


FIG. 16

22/57

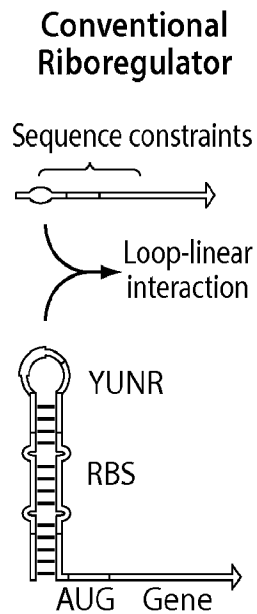


FIG. 17A

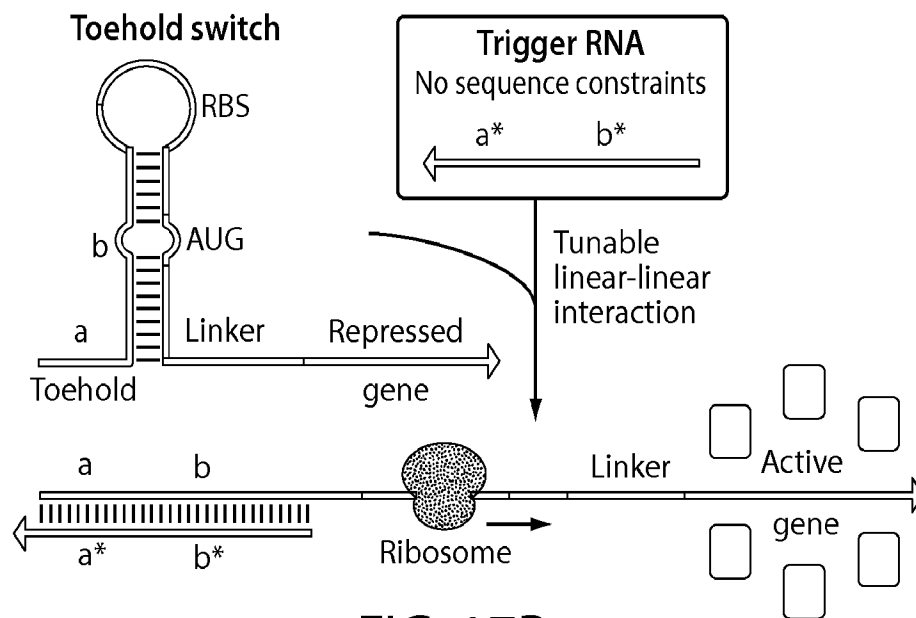


FIG. 17B

23/57

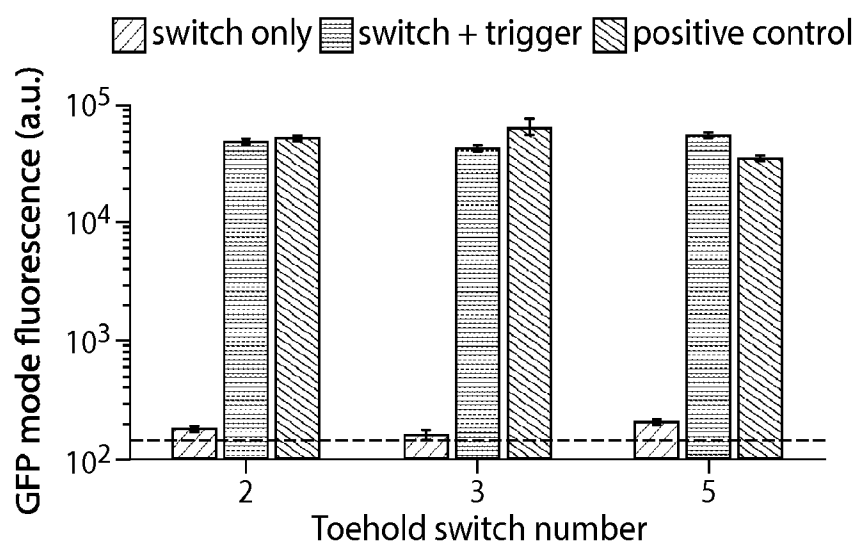


FIG. 17C

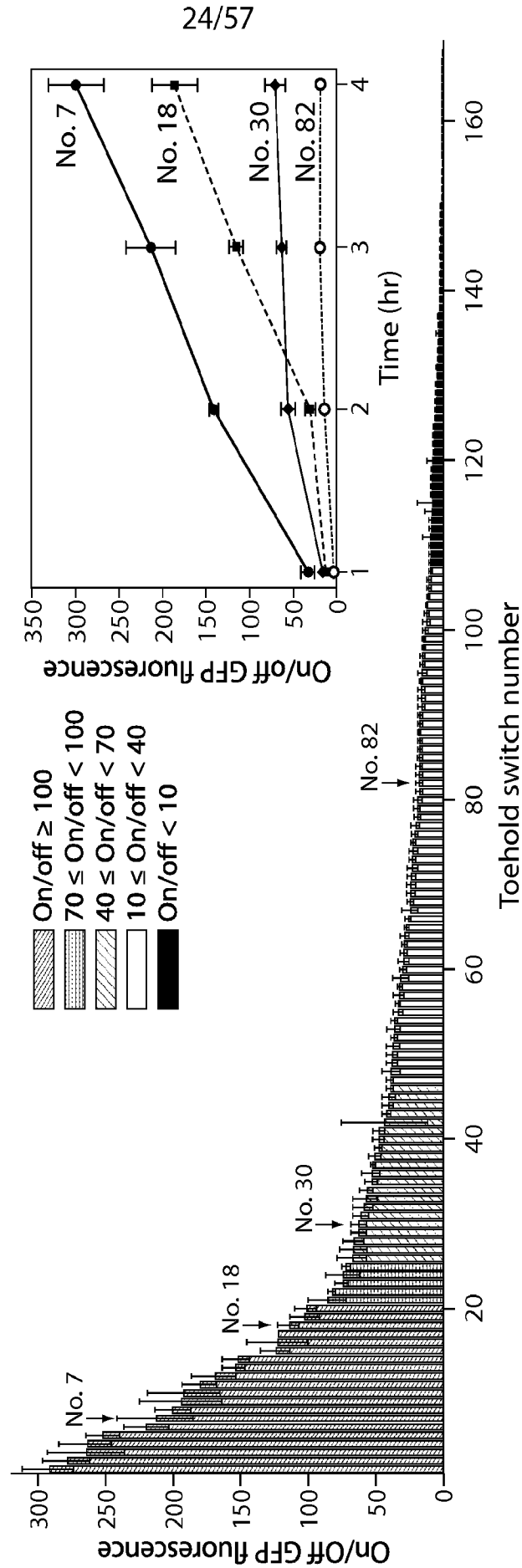


FIG. 17D

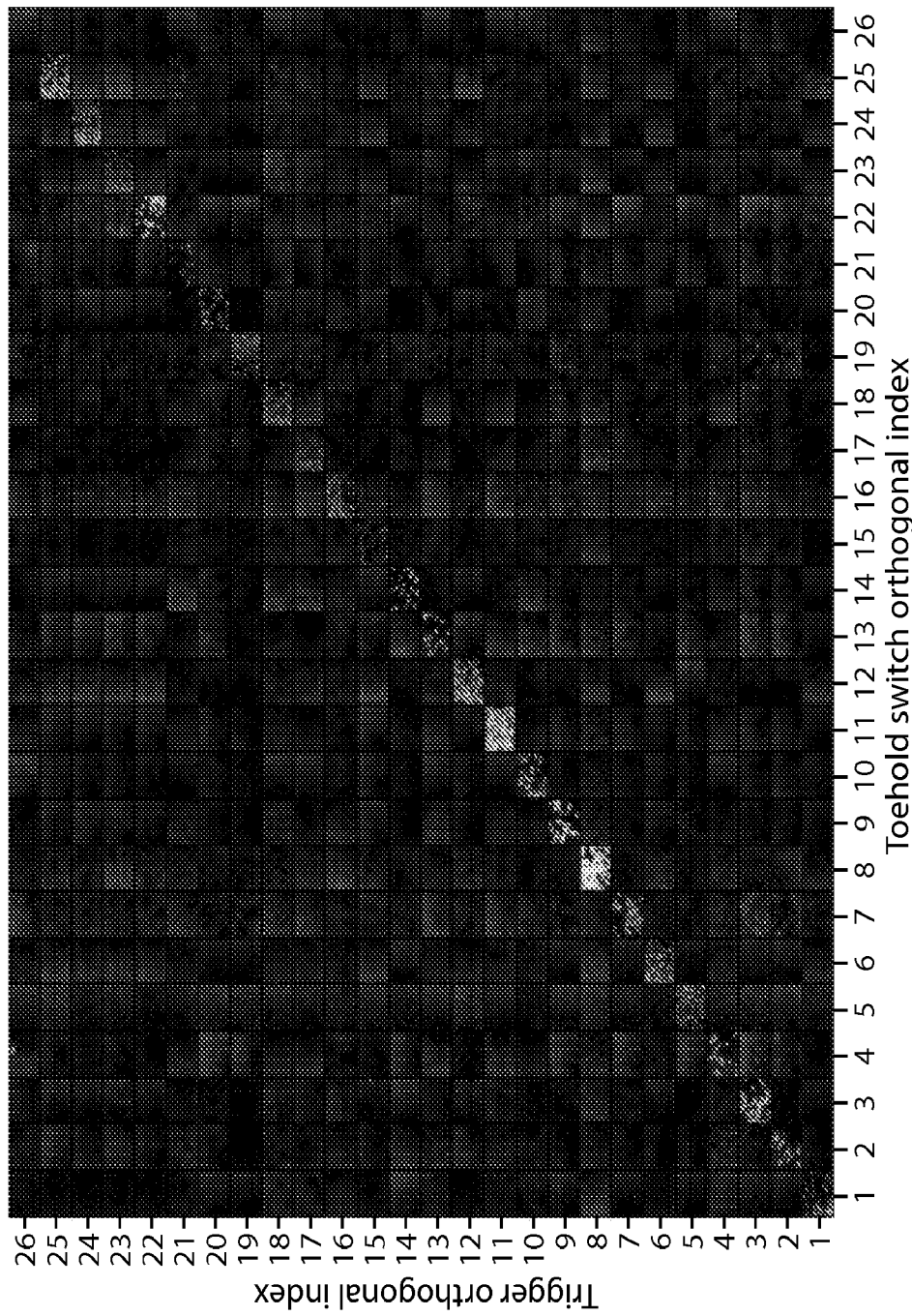


FIG. 18A

26/57

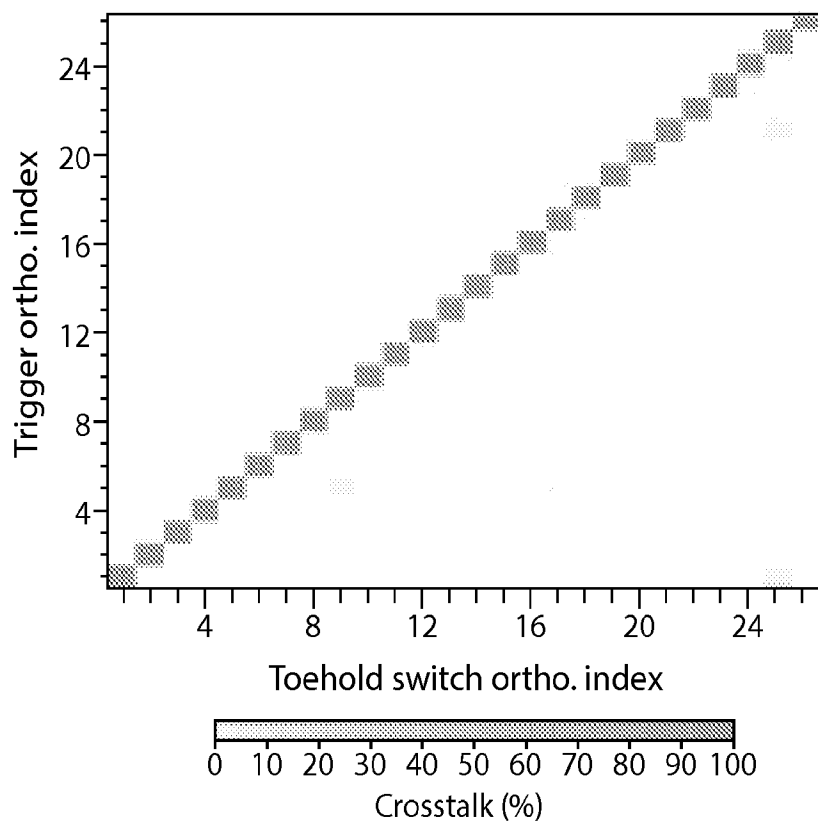


FIG. 18B

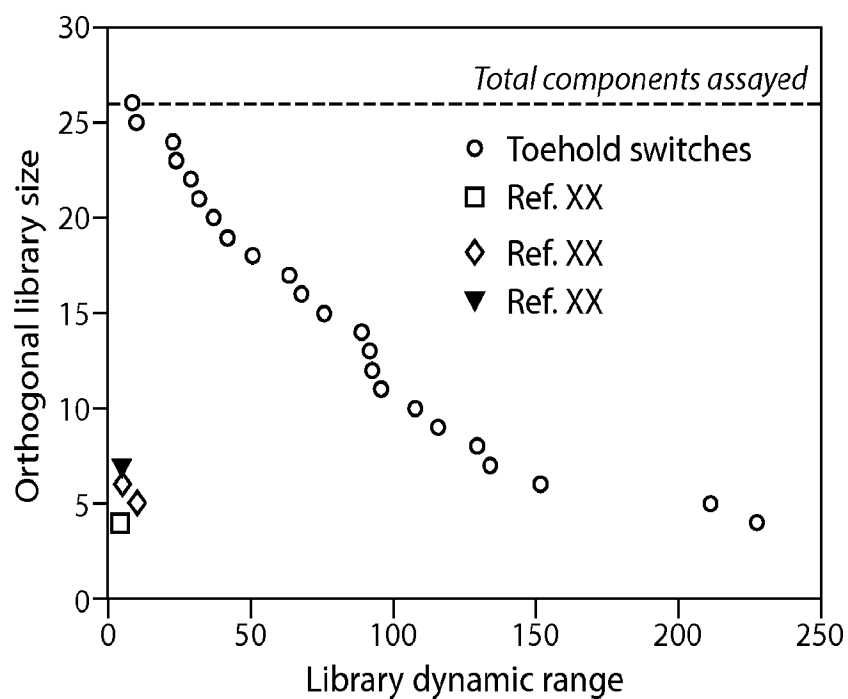


FIG. 18C

27/57

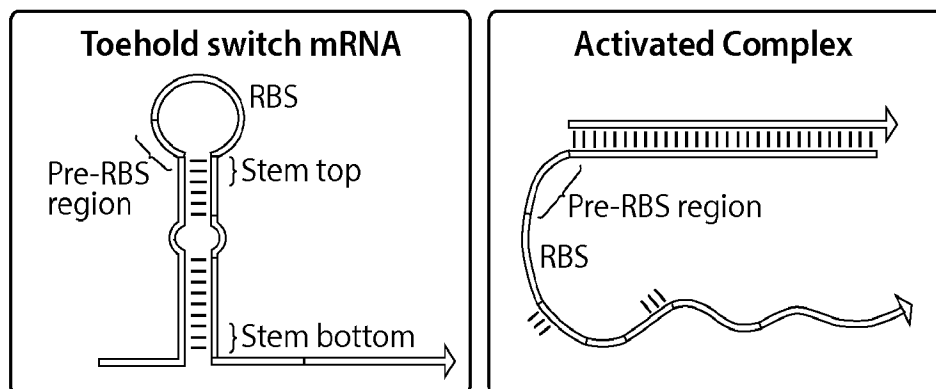


FIG. 19A

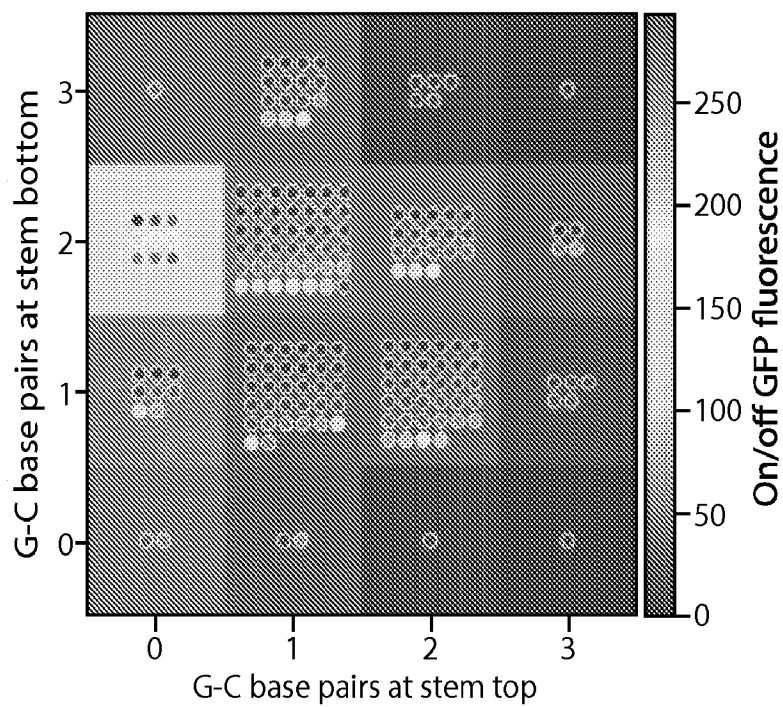


FIG. 19B

28/57

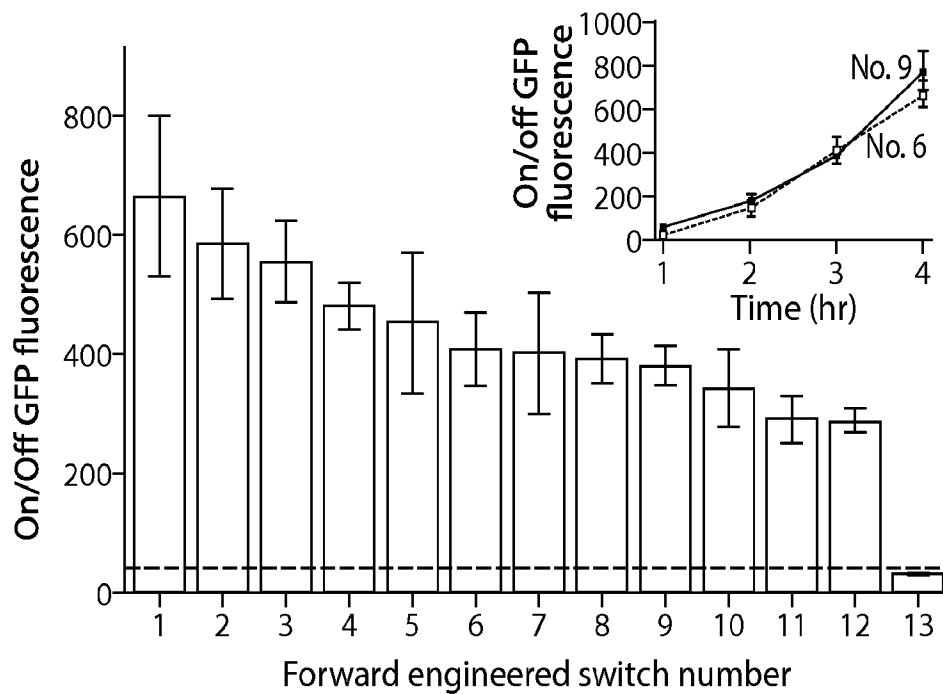


FIG. 19C

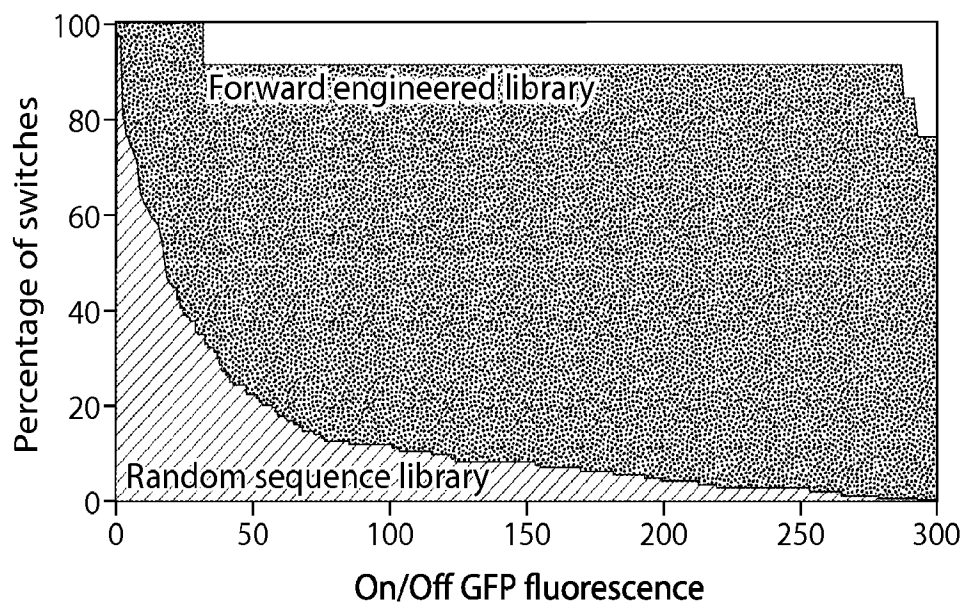


FIG. 19D

29/57

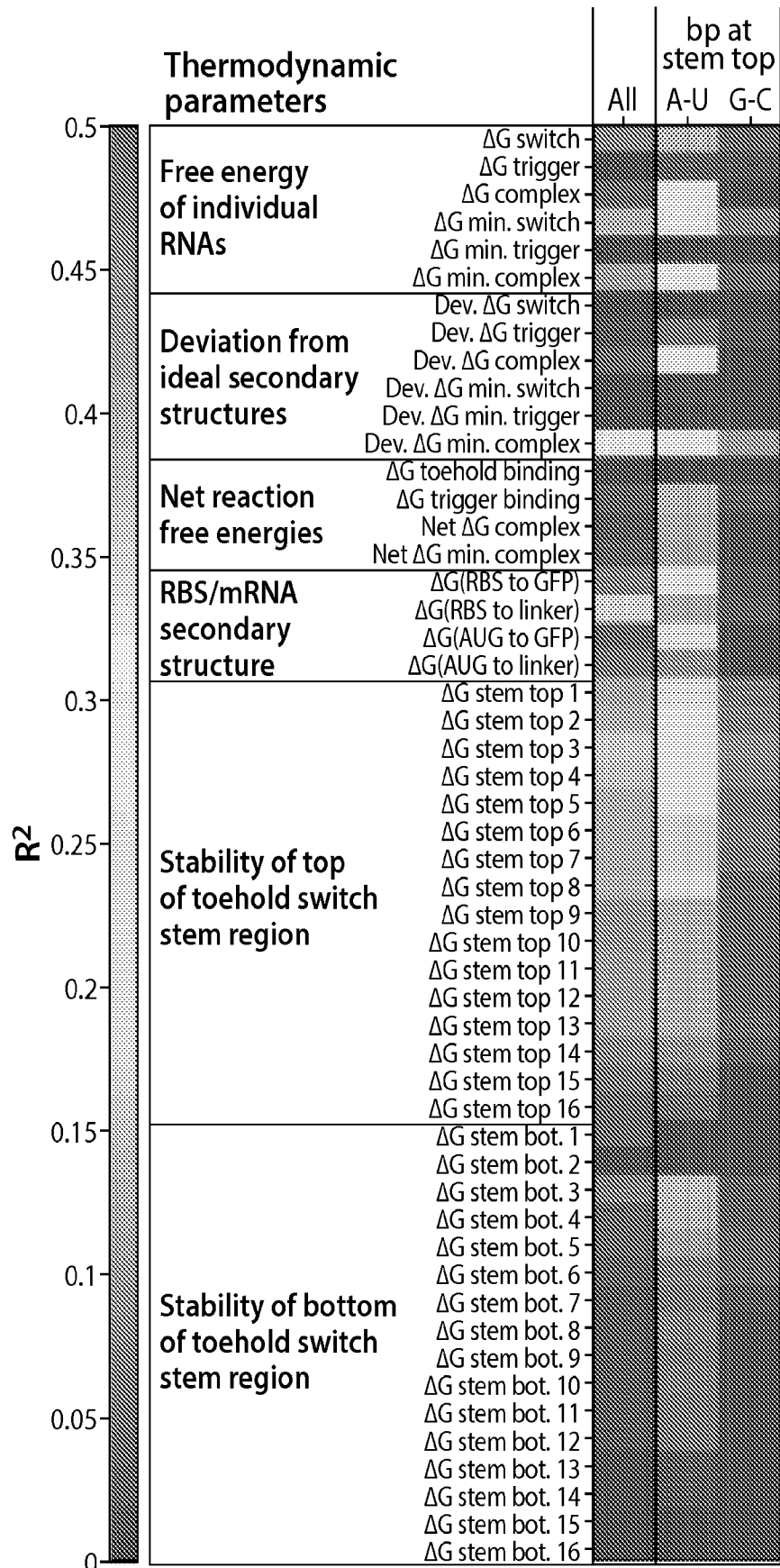


FIG. 20A

30/57

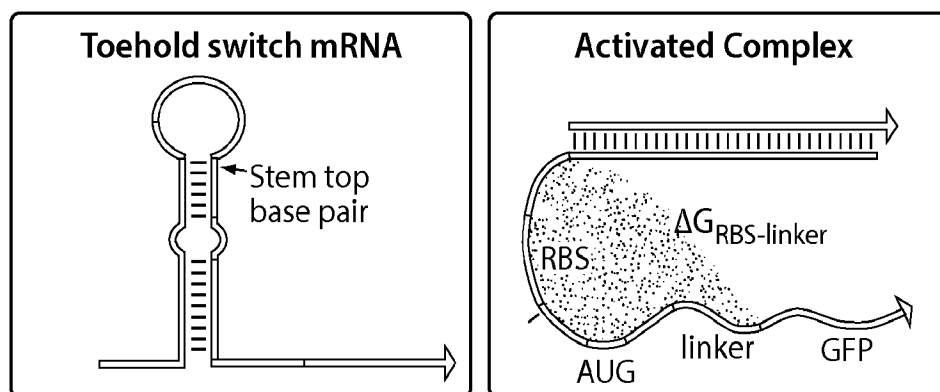


FIG. 20B

31/57

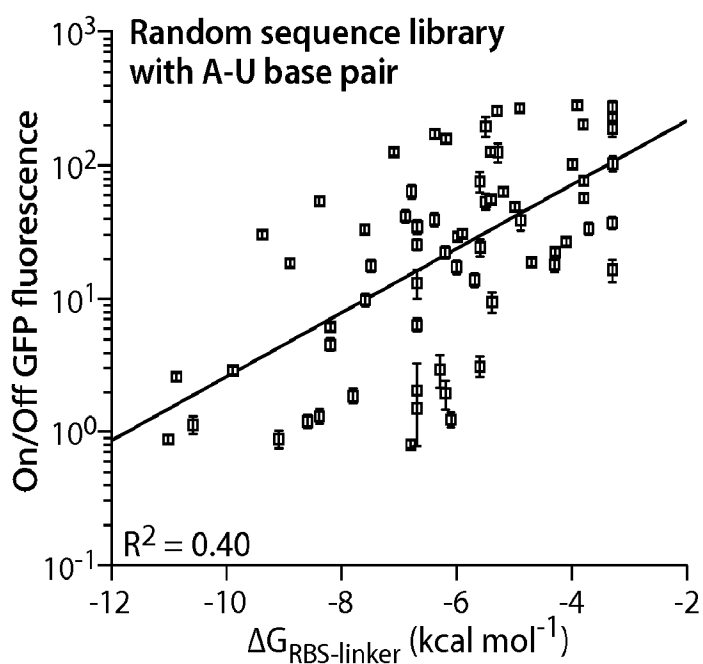


FIG. 20C

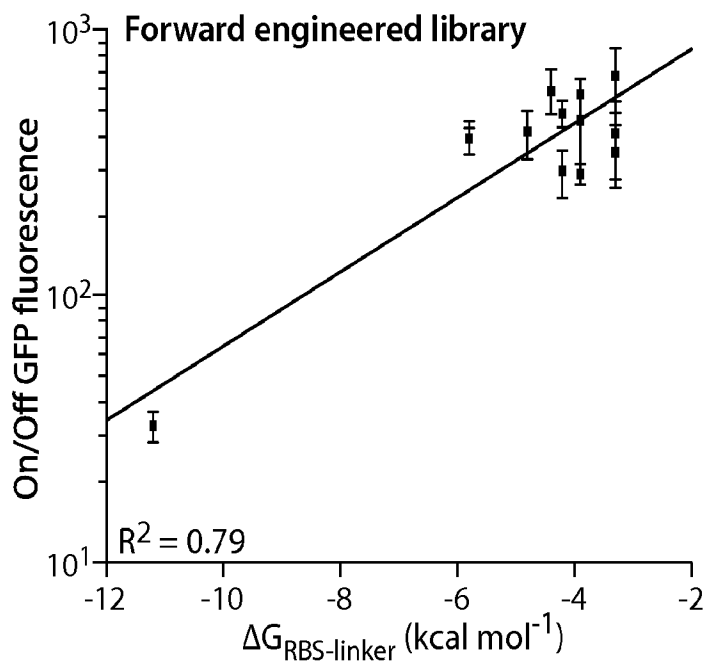


FIG. 20D

32/57

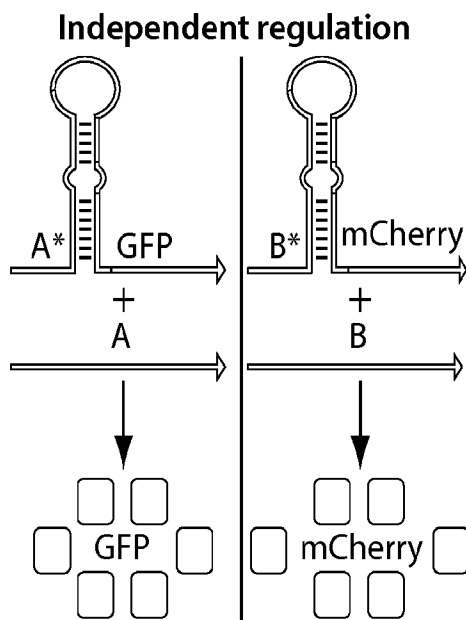


FIG. 21A

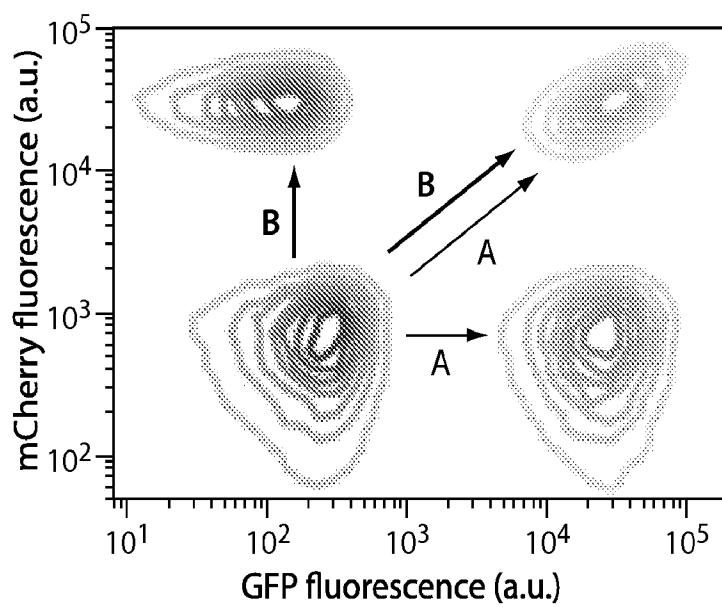


FIG. 21B

33/57

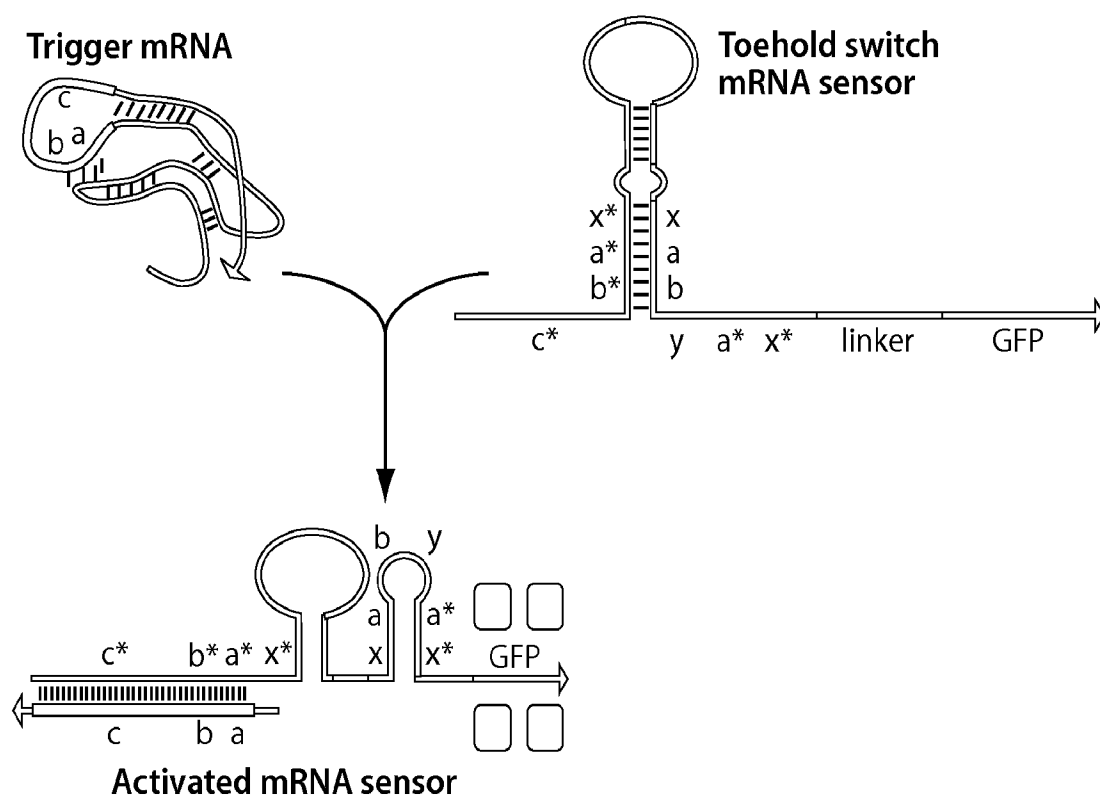


FIG. 21C

34/57

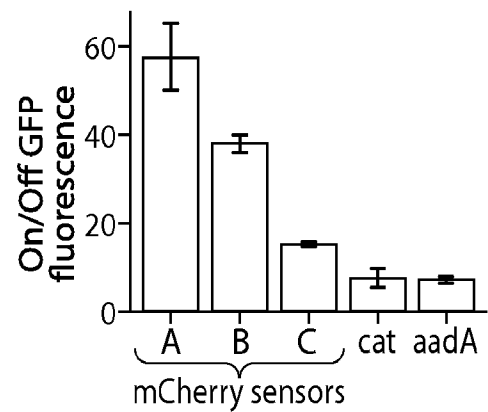


FIG. 21D

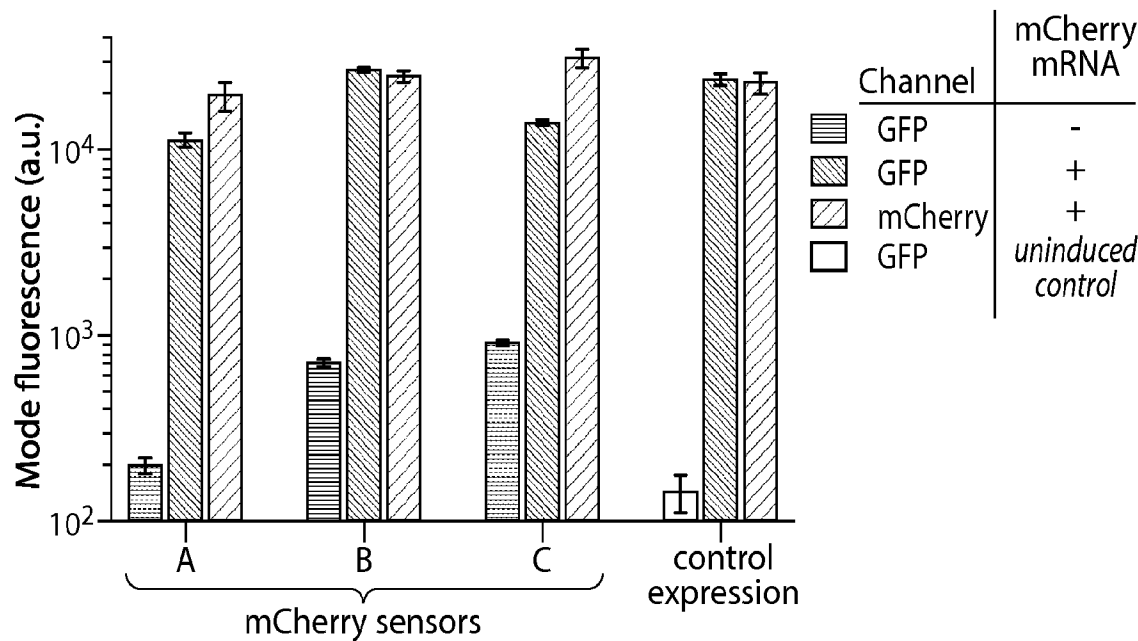


FIG. 21E

35/57

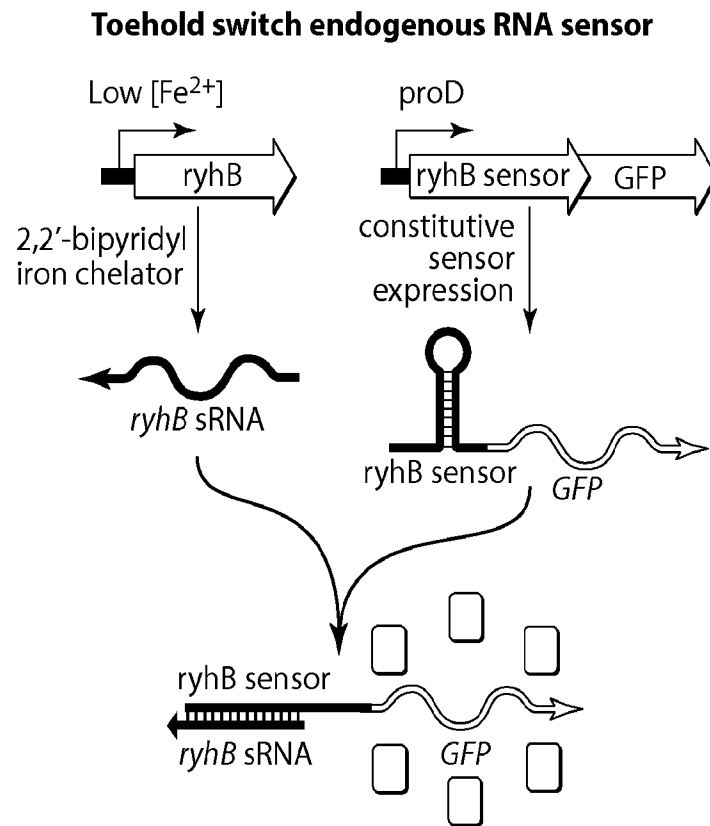


FIG. 22A

36/57

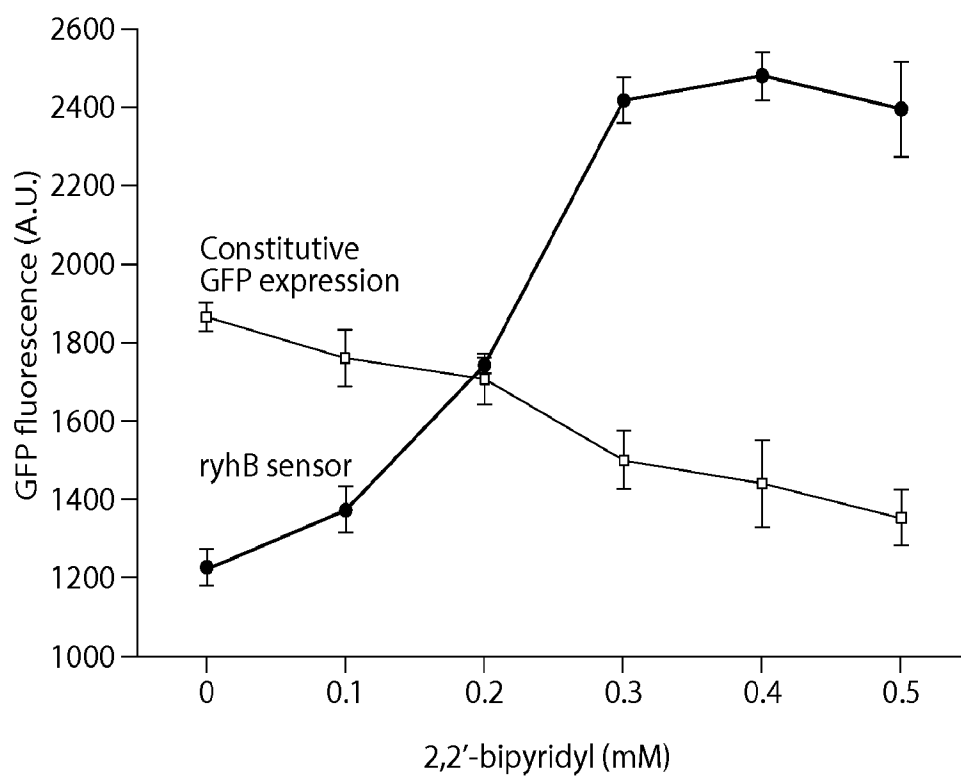


FIG. 22B

37/57

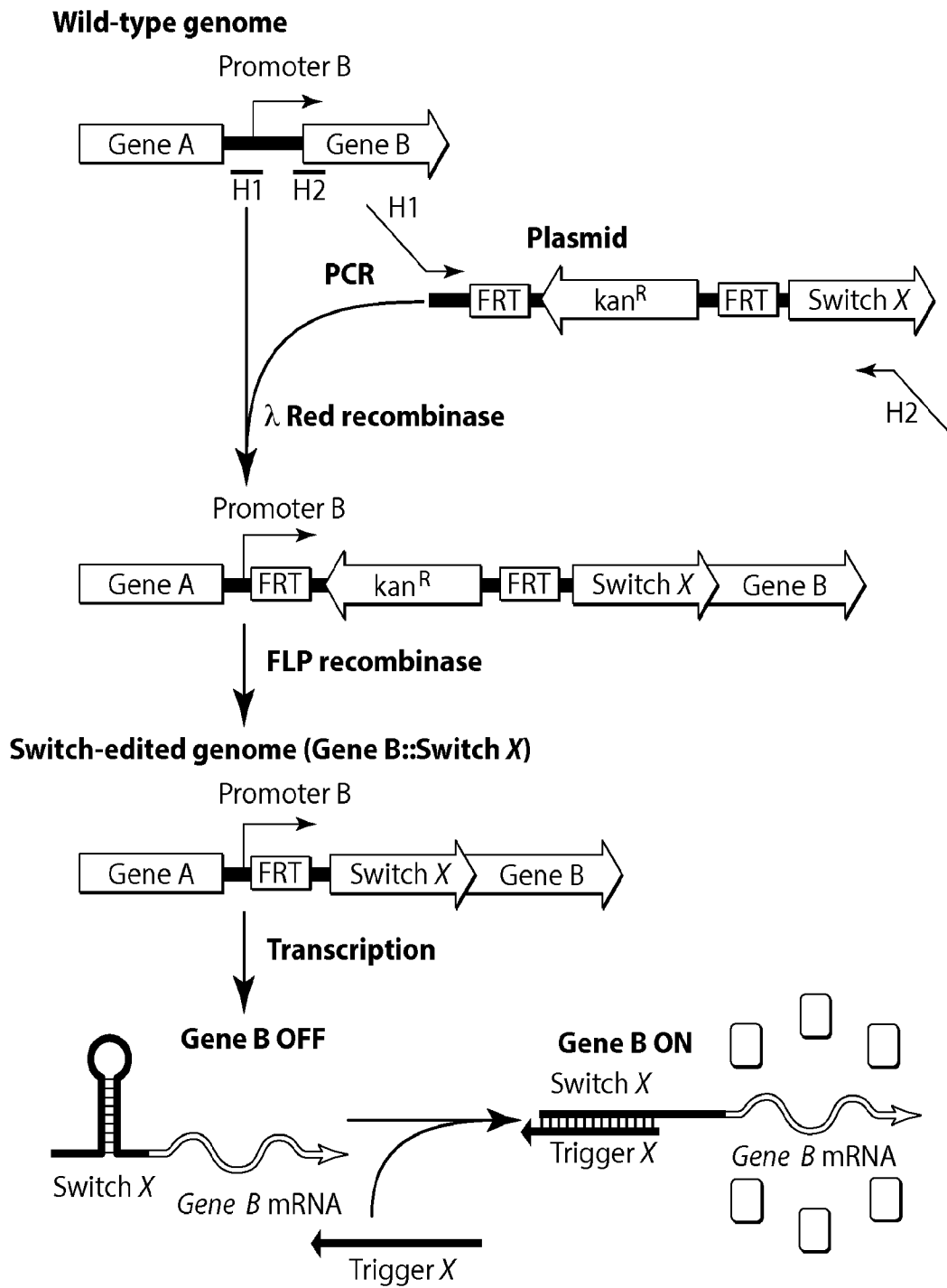


FIG. 23A

38/57

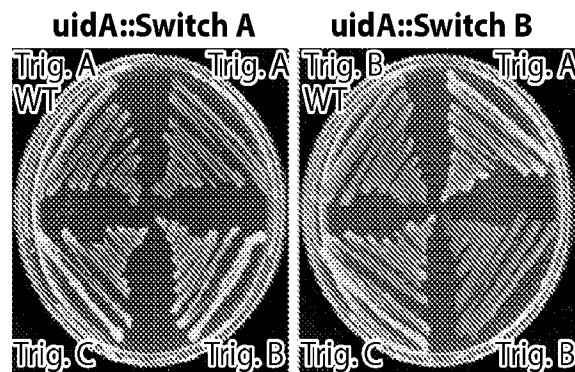


FIG. 23B

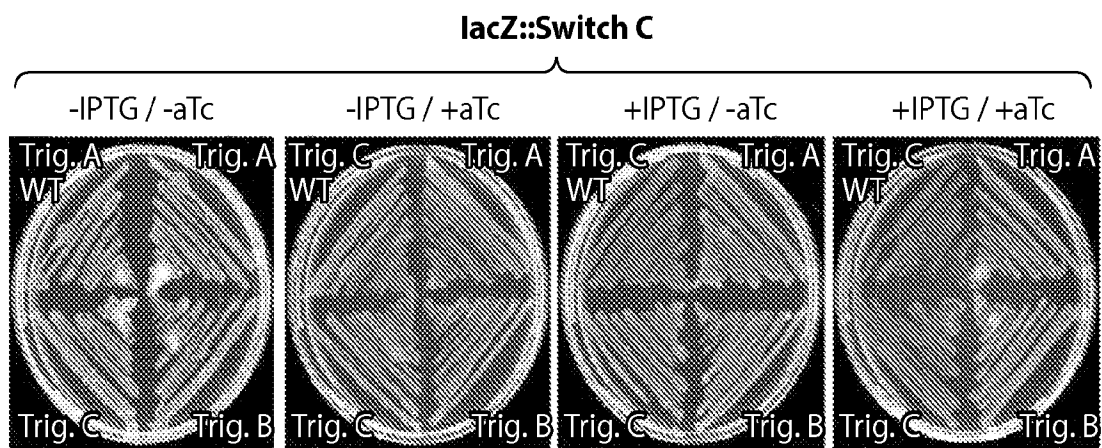


FIG. 23C

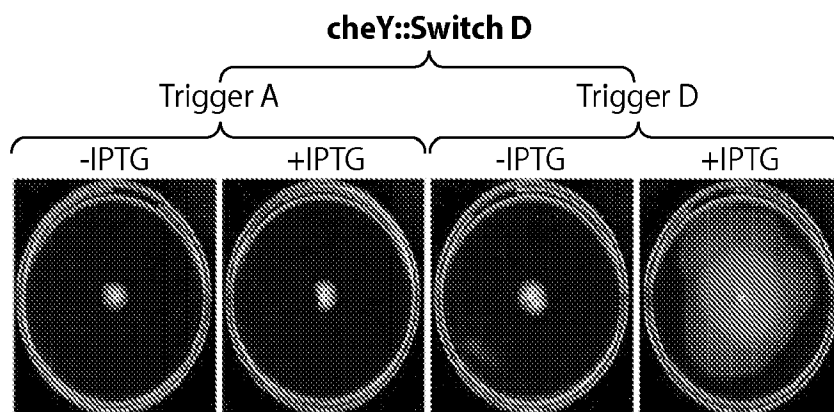


FIG. 23D

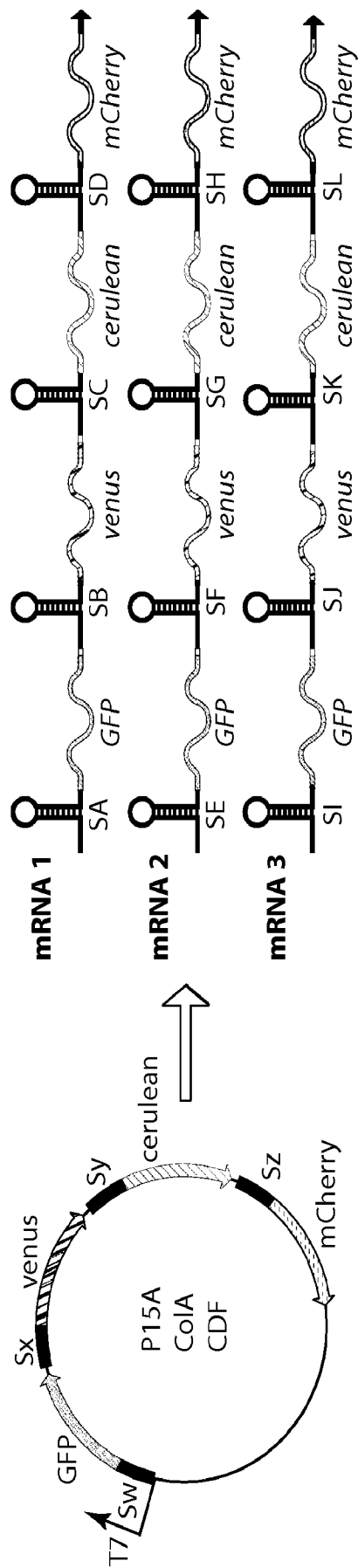


FIG. 24A

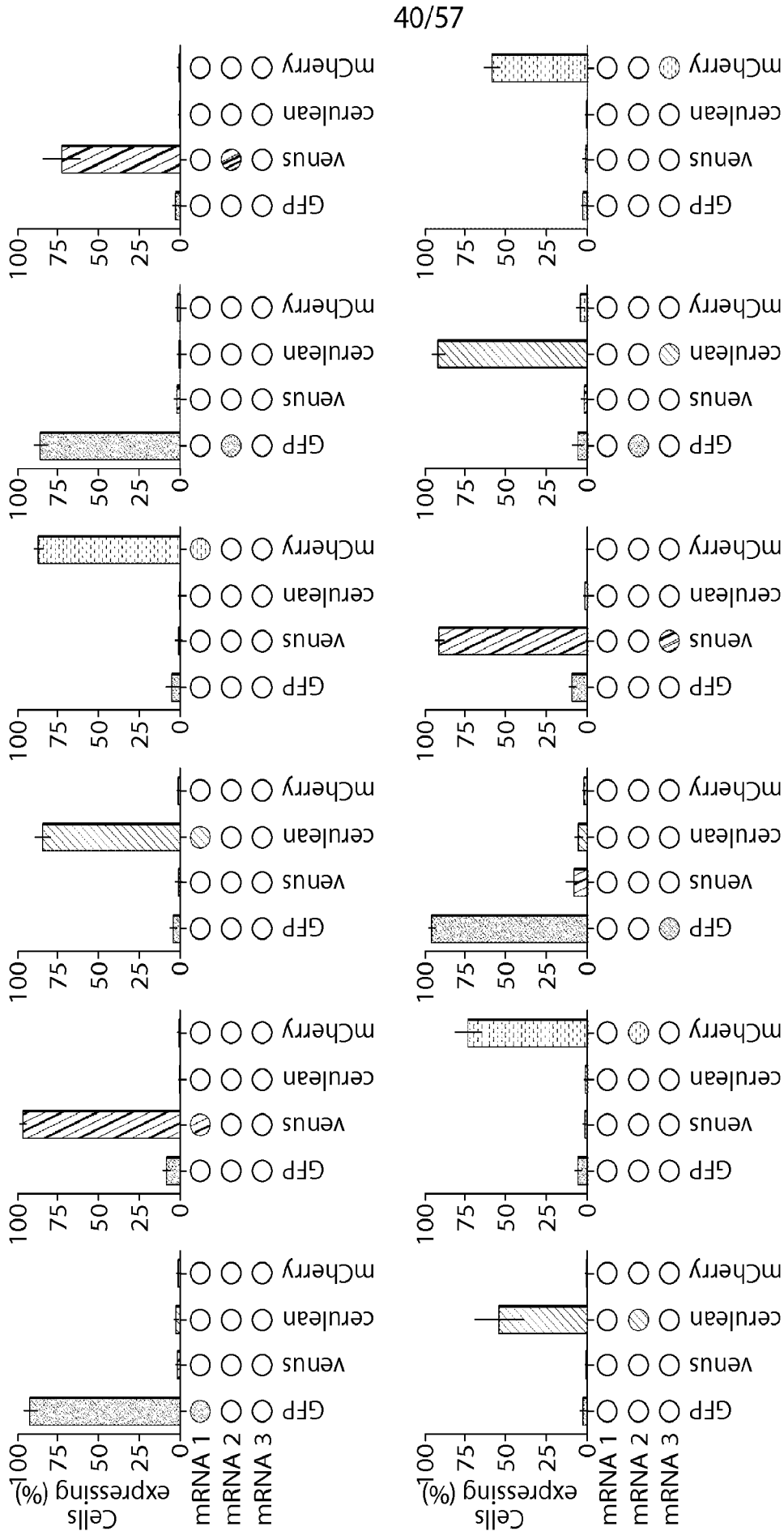


FIG. 24B

41/57

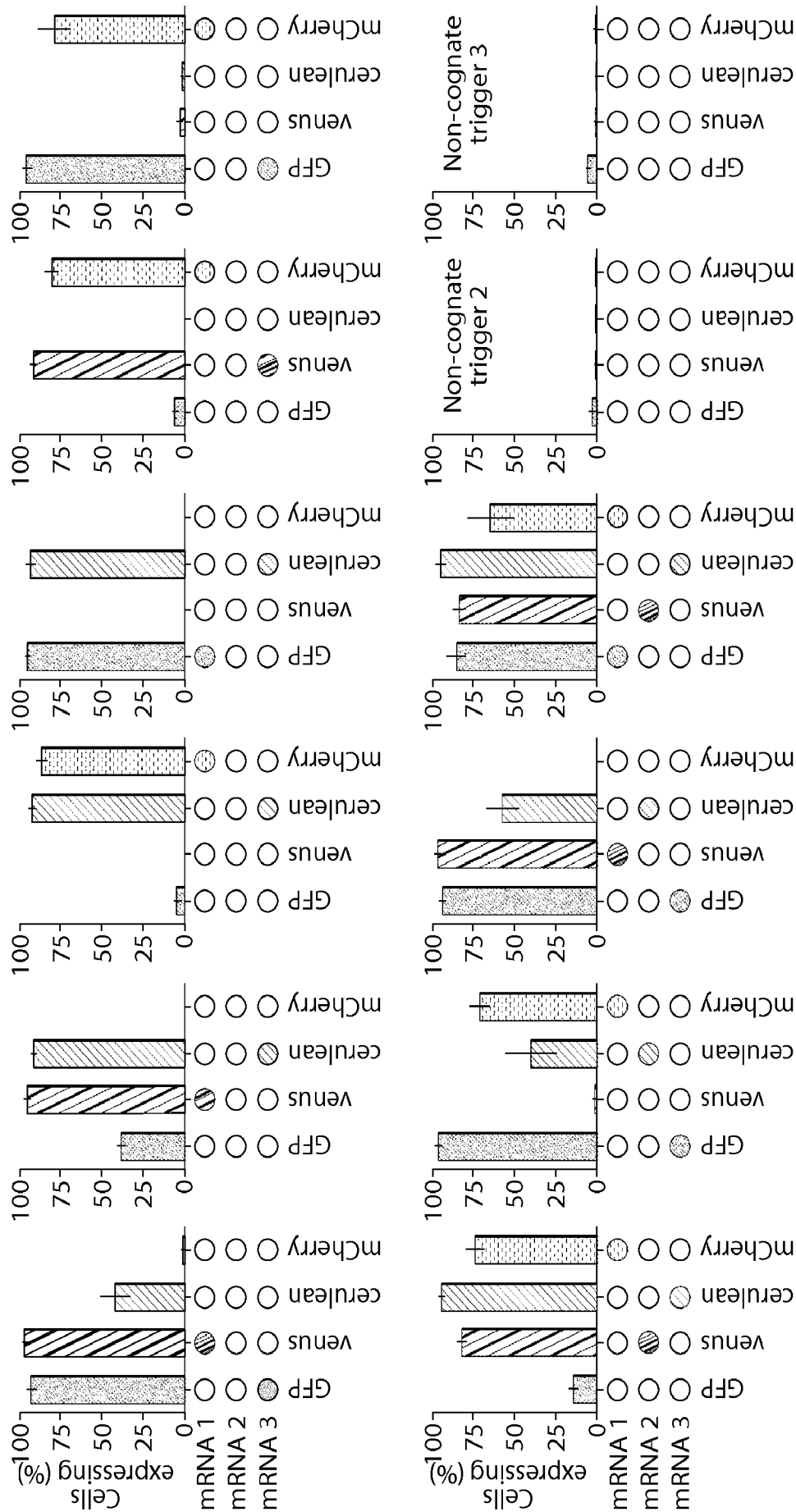


FIG. 24C

42/57

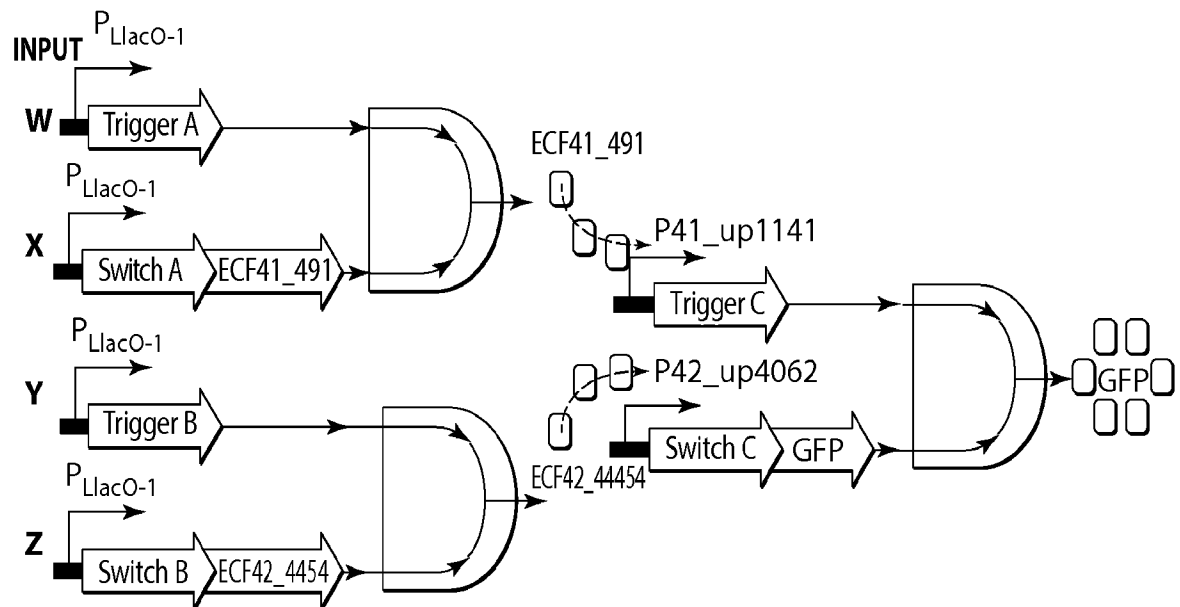


FIG. 25A

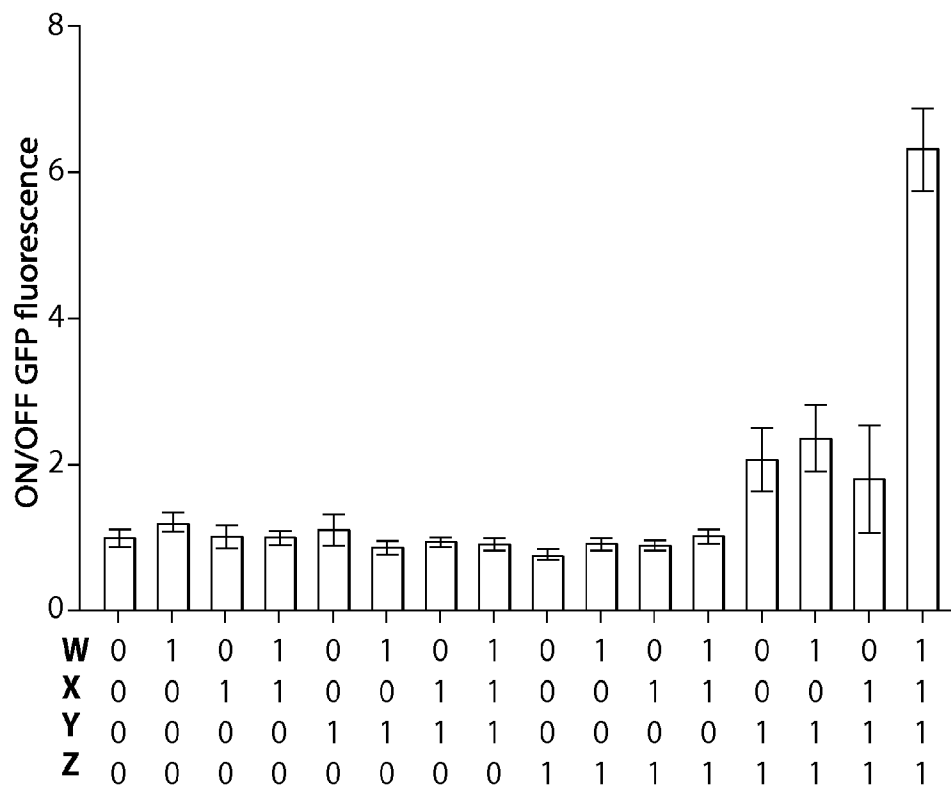


FIG. 25B

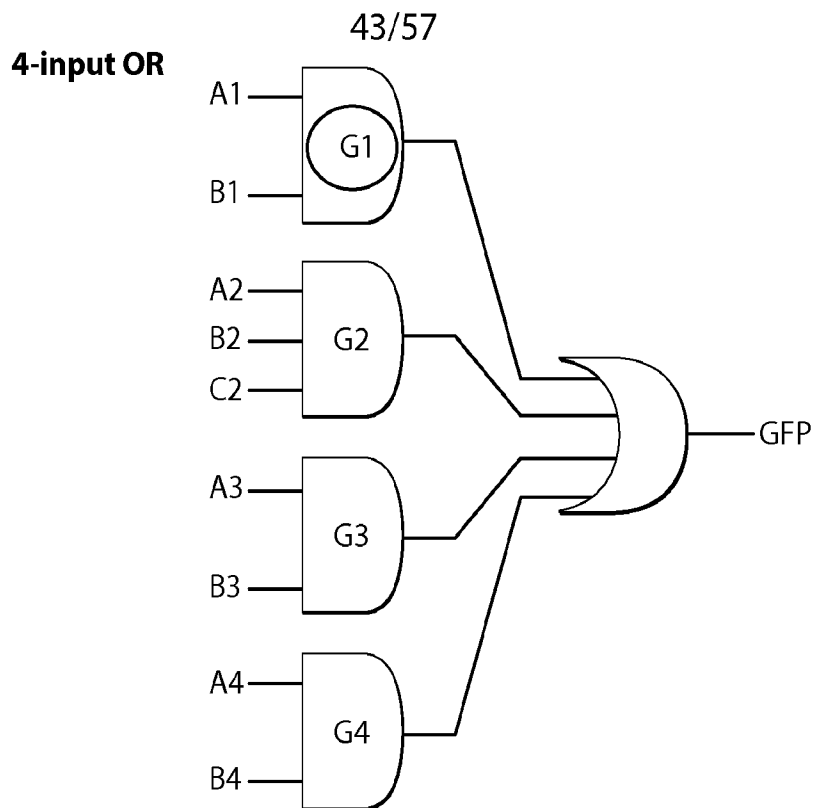


FIG. 26A

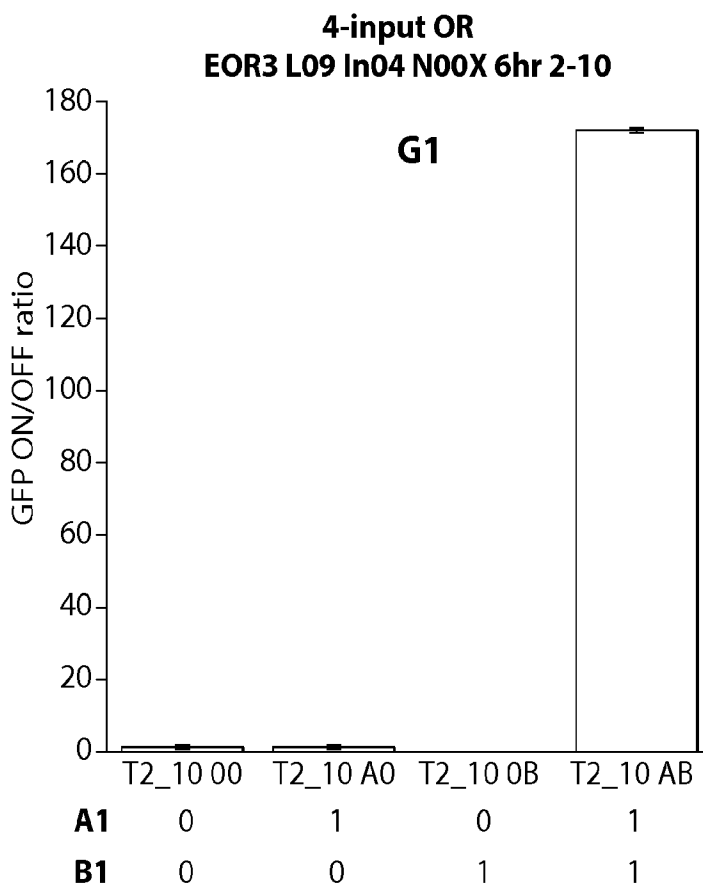
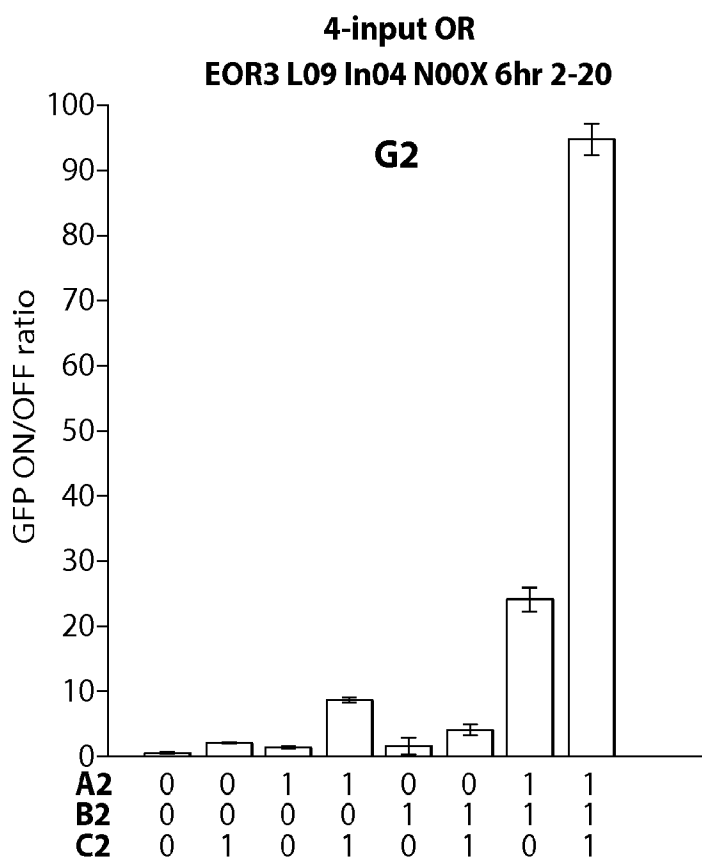
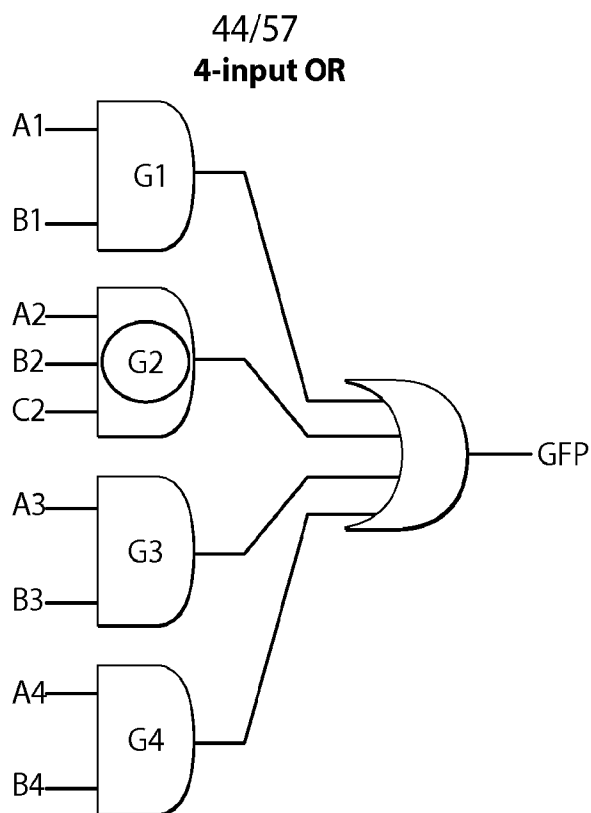


FIG. 26B

**FIG. 27B**

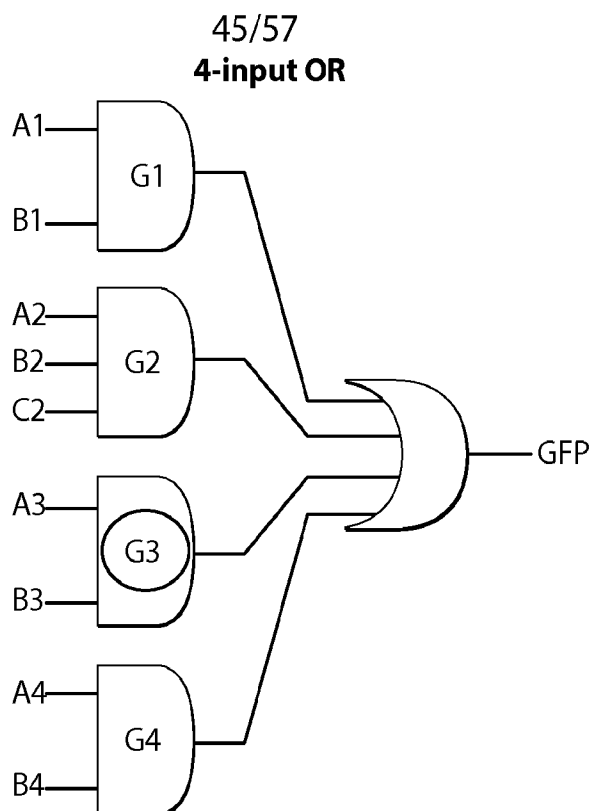


FIG. 28A

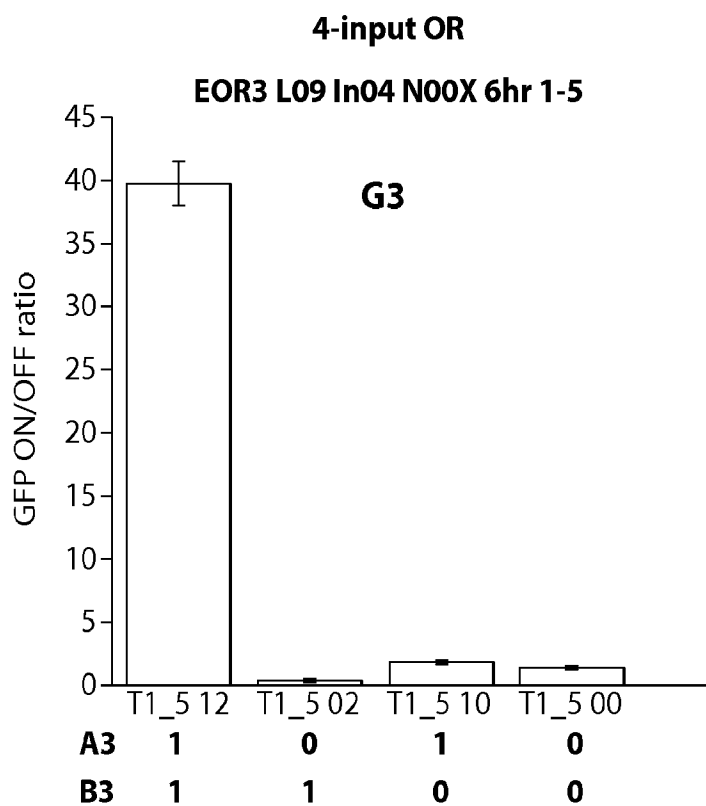


FIG. 28B

46/57

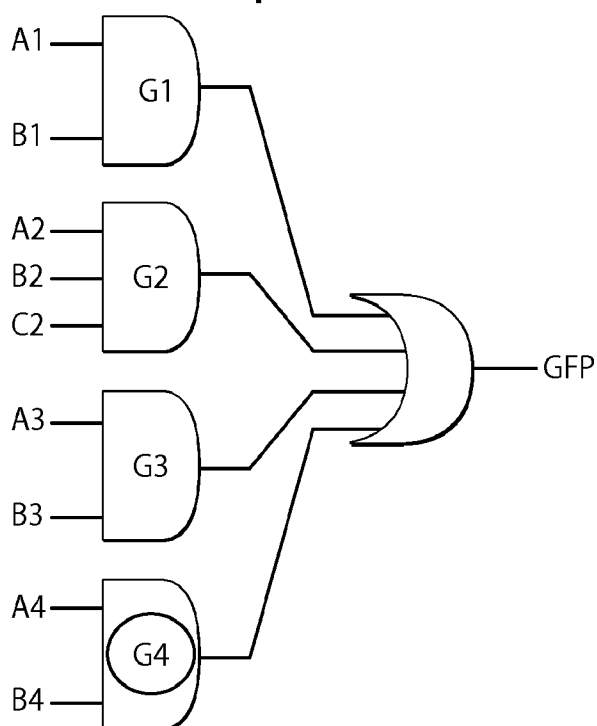
4-input OR

FIG. 29A

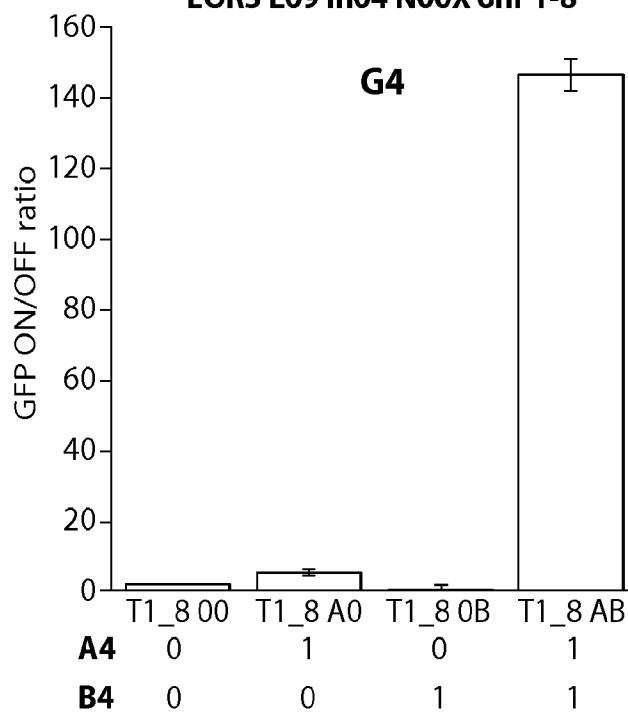
4-input OR**EOR3 L09 In04 N00X 6hr 1-8**

FIG. 29B

47/57

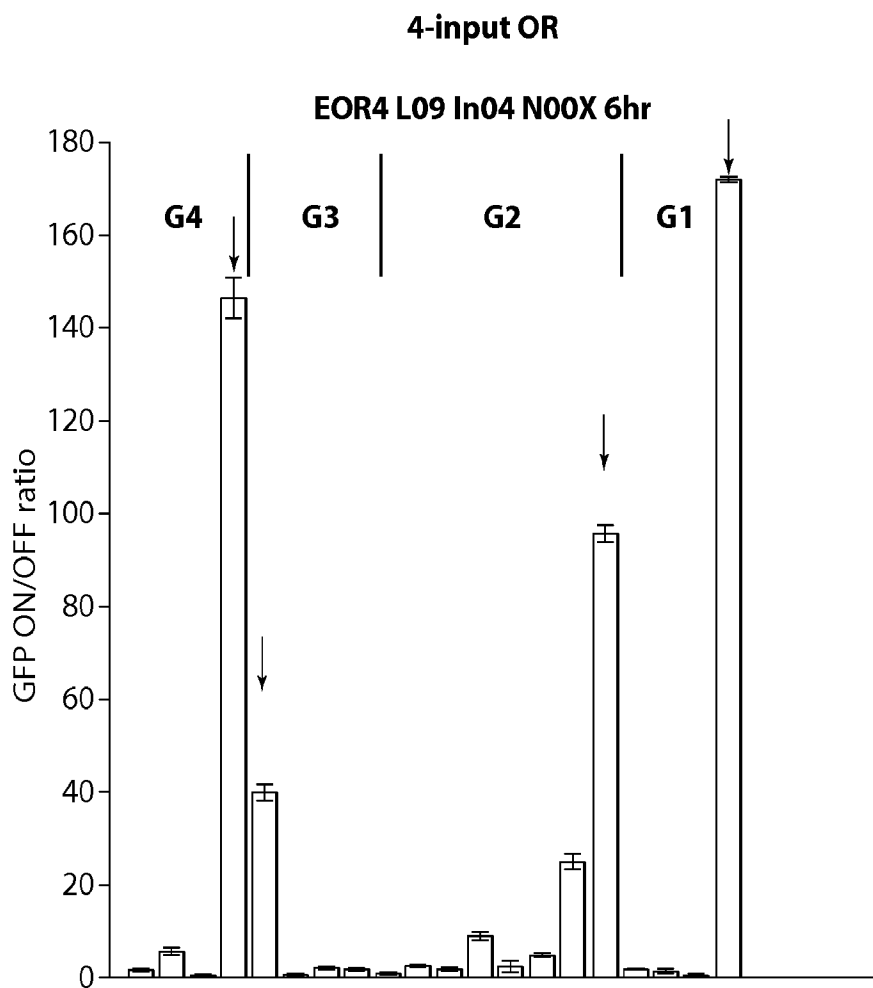


FIG. 30

48/57

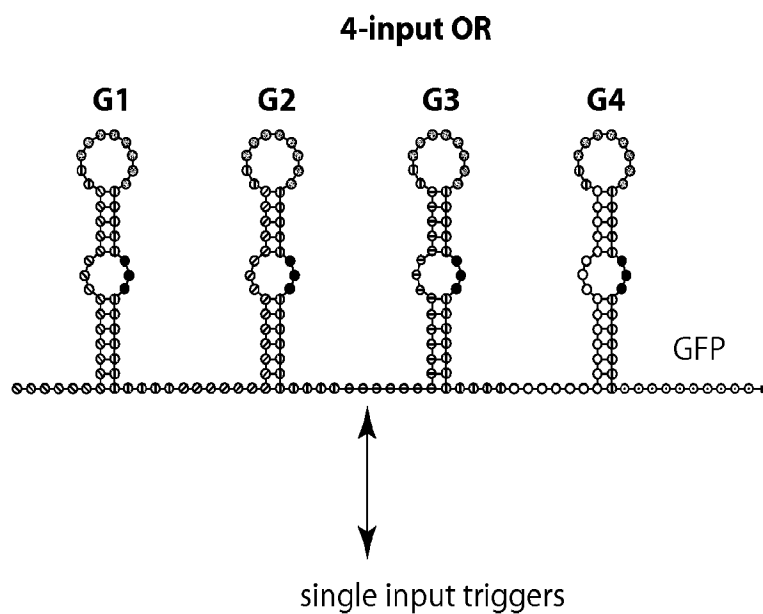


FIG. 31A

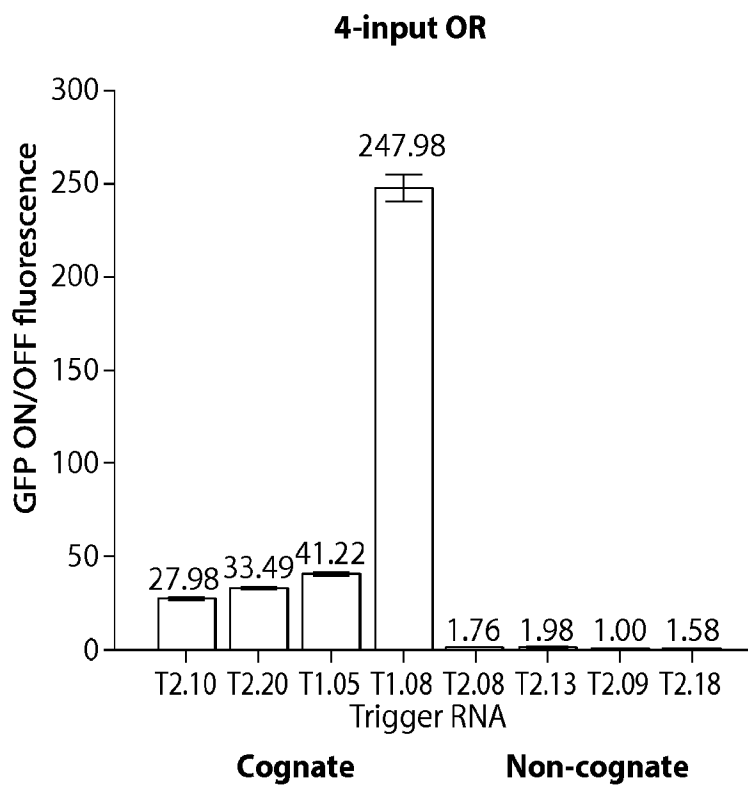


FIG. 31B

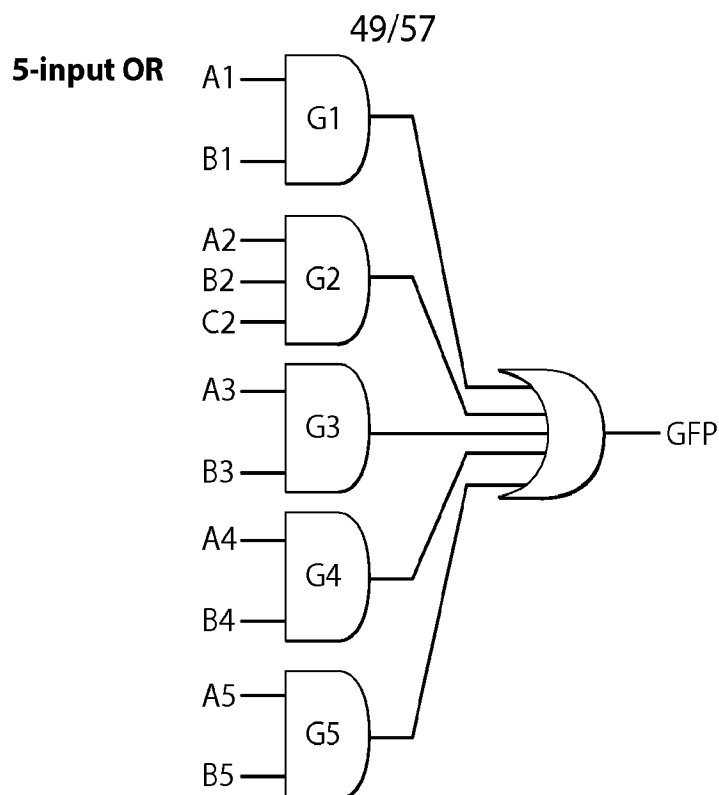


FIG. 32A

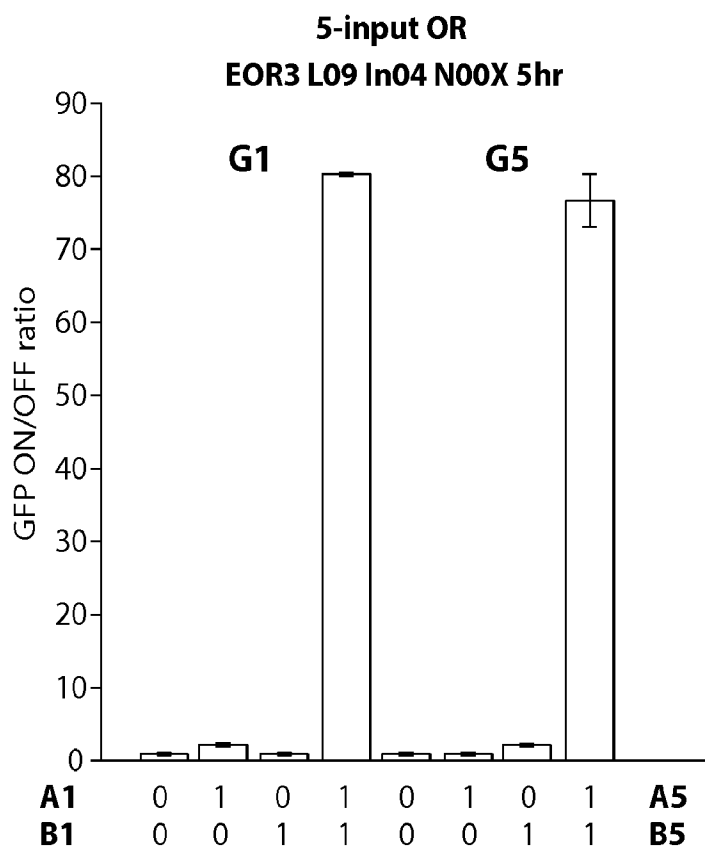


FIG. 32B

50/57

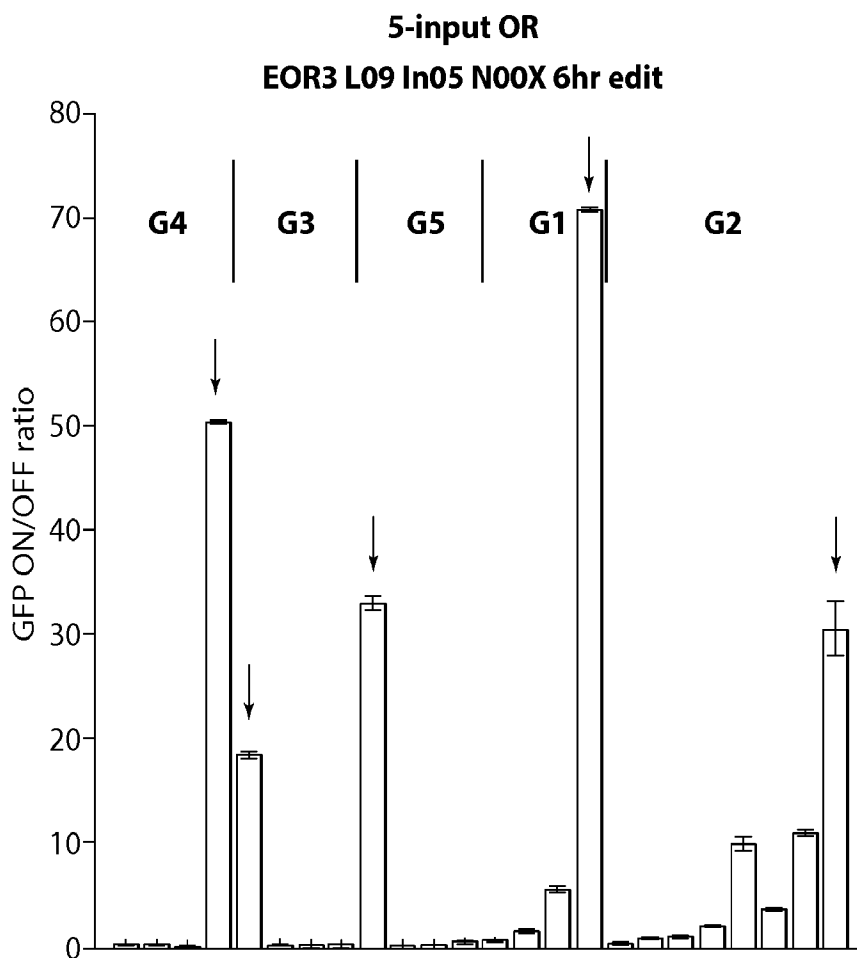


FIG. 33

51/57

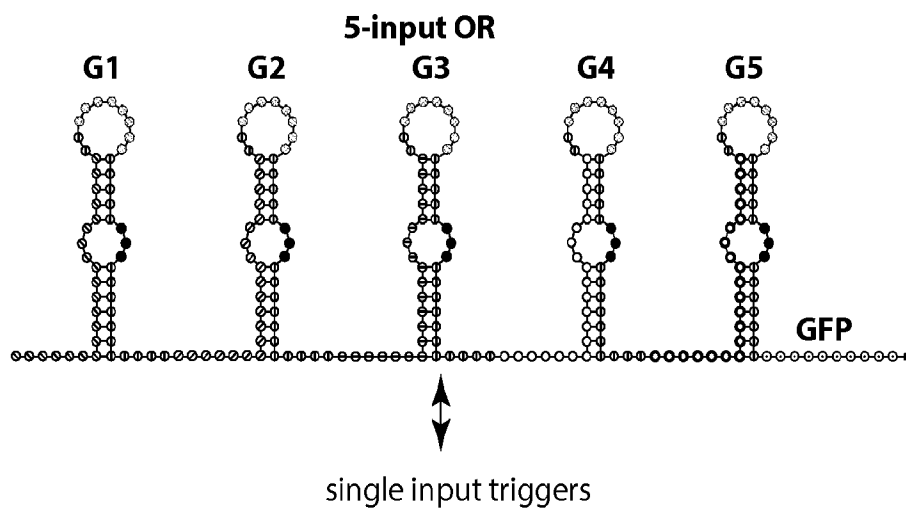


FIG. 34A

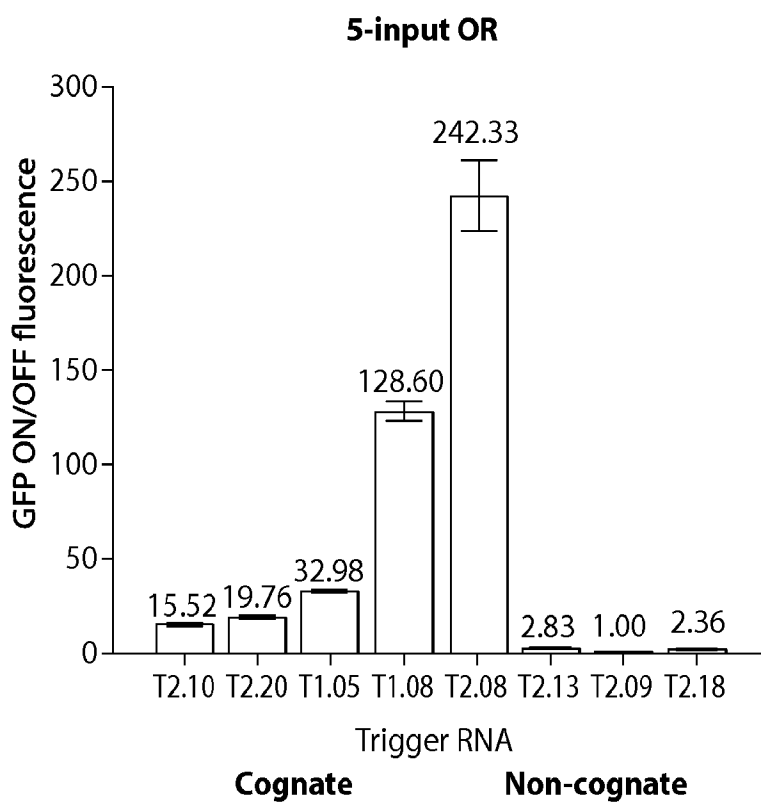


FIG. 34B

52/57

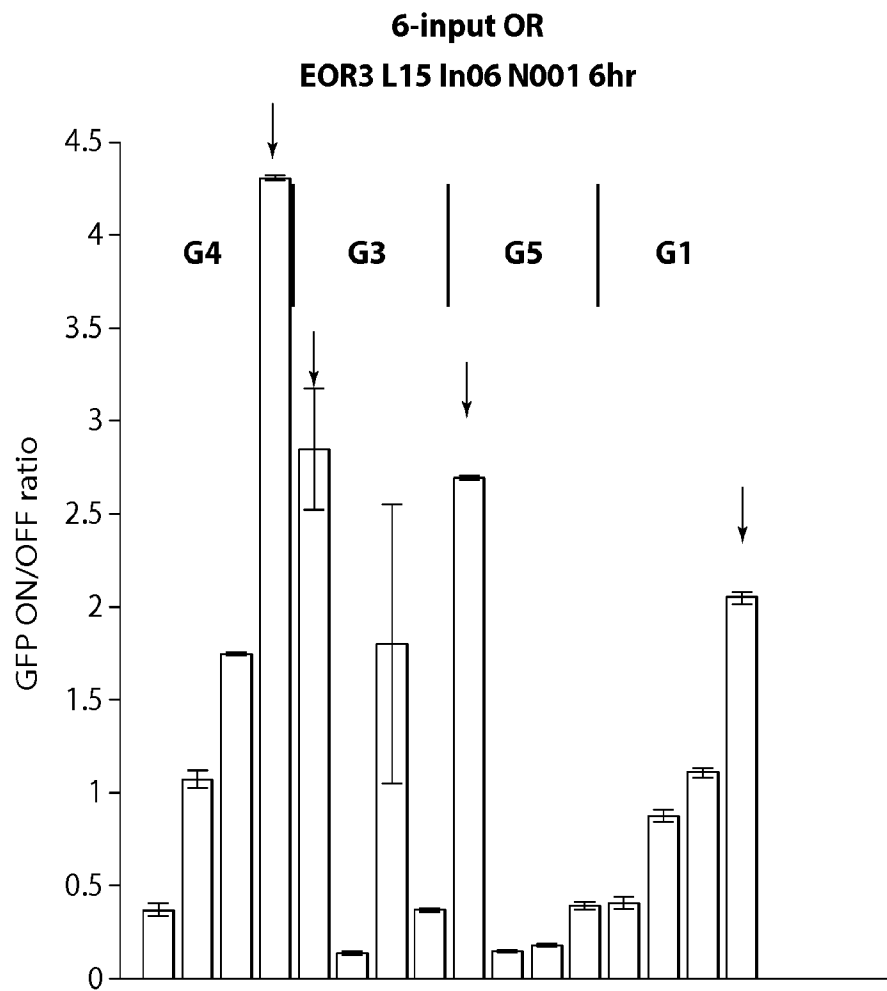


FIG. 35

53/57

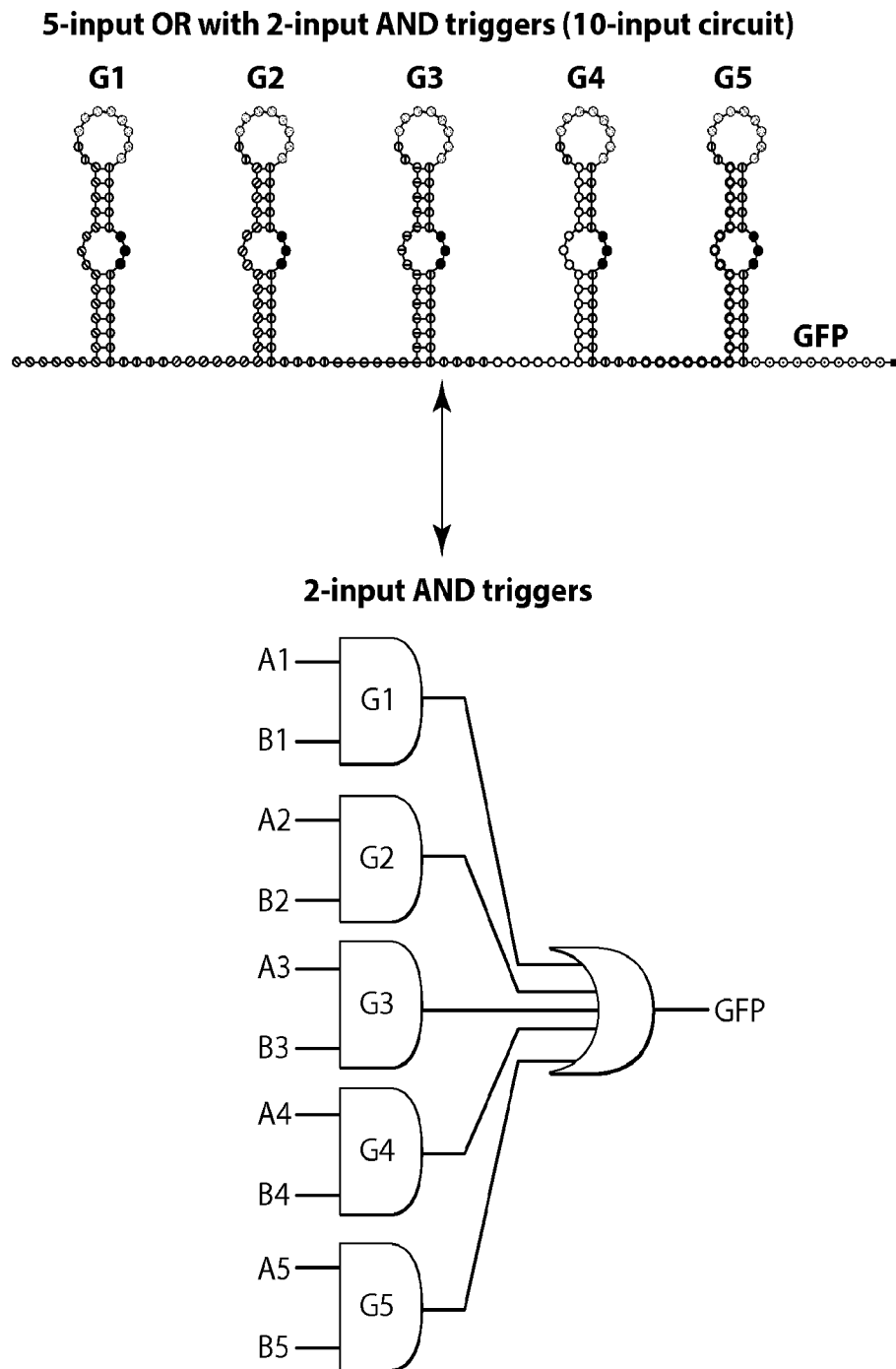
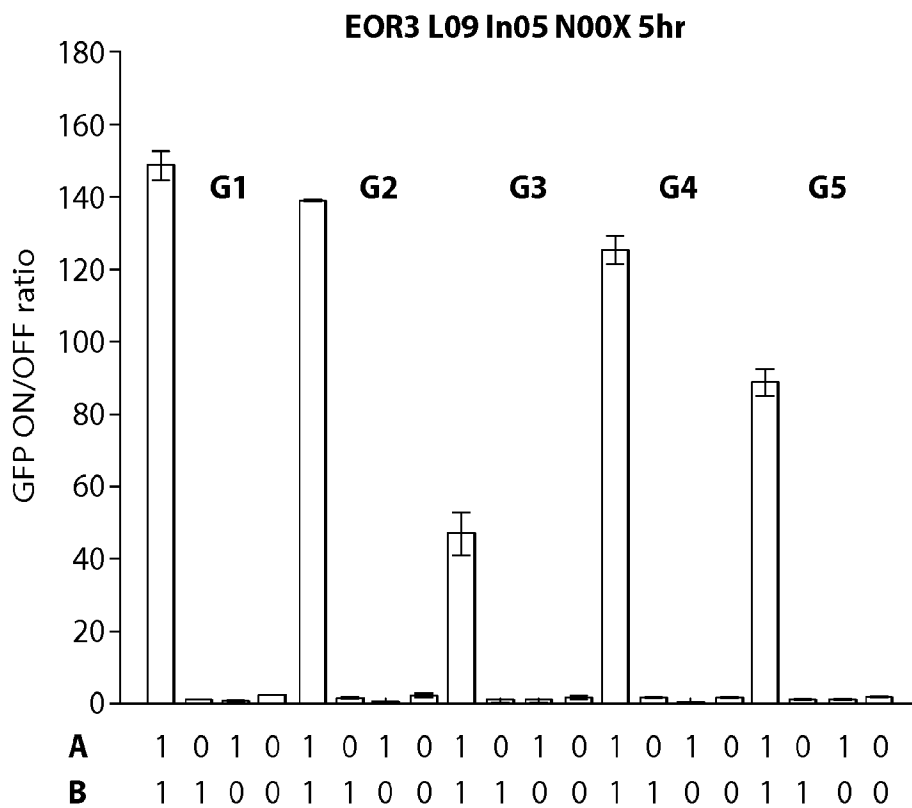
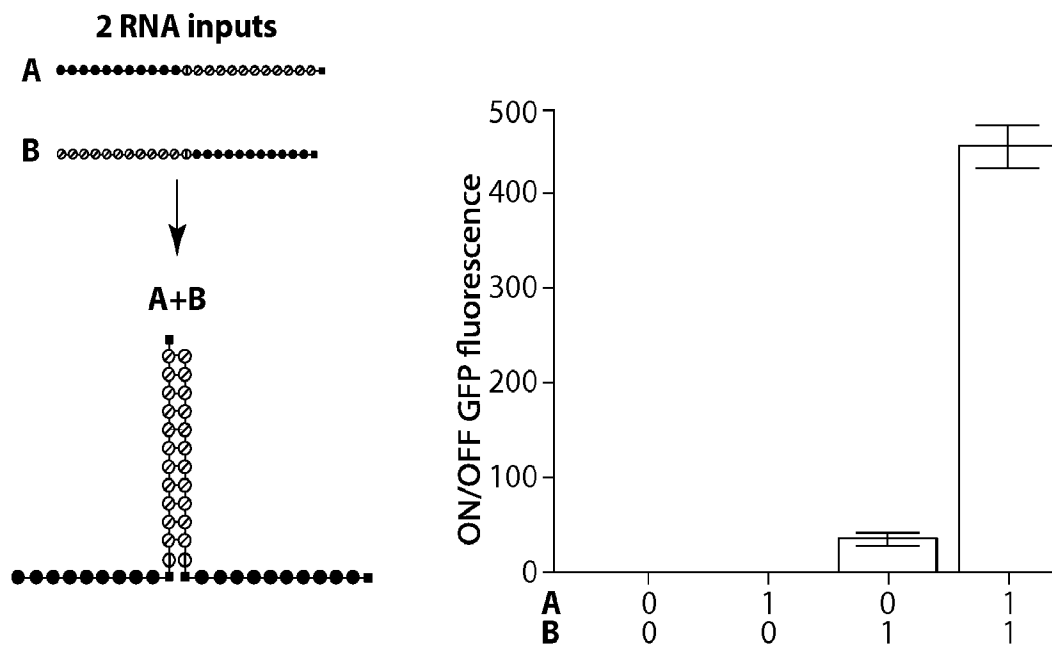


FIG. 36A

54/57

5-input OR with 2-input AND triggers (10-input circuit)**a single 2-input AND gate****FIG. 36B**

55/57

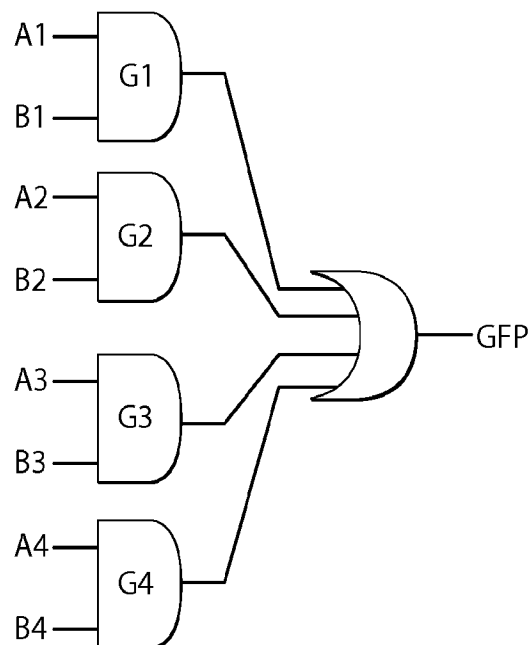
4-input OR with 2-input AND triggers (8-input circuit)

FIG. 37A

4-input OR with 2-input AND triggers (8-input circuit)

EOR3 L09 In04 N00X 5hr

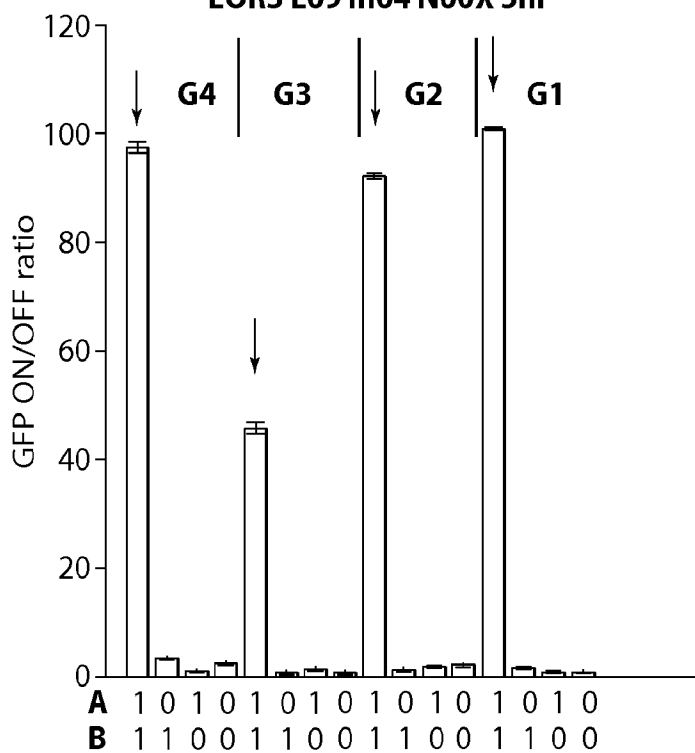


FIG. 37B

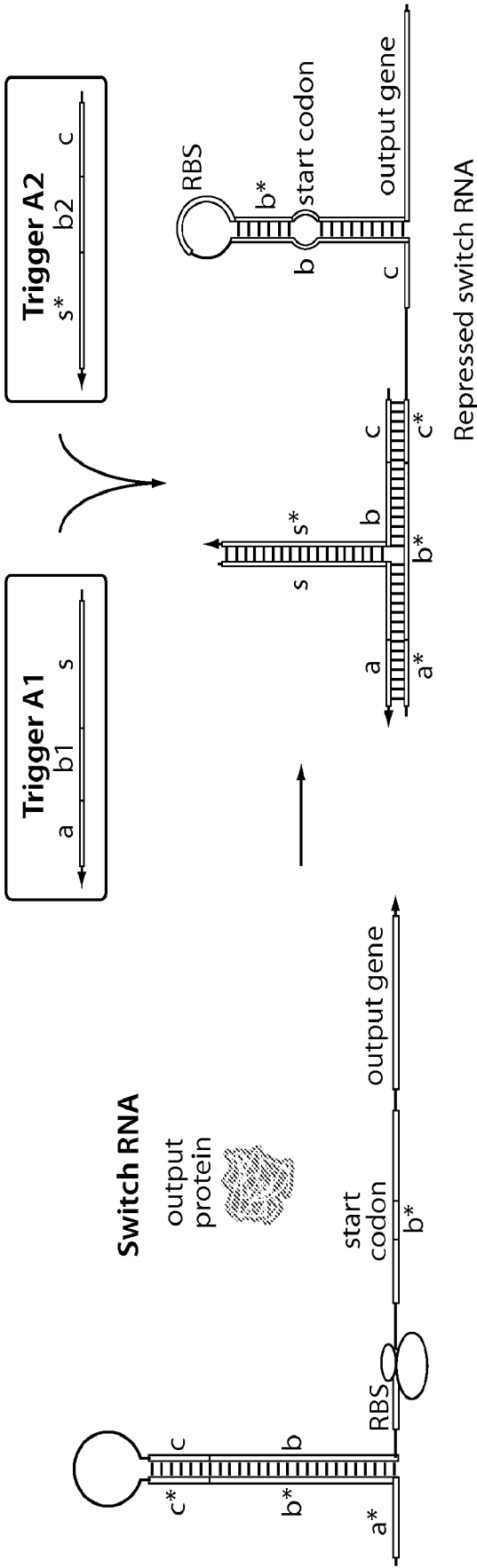


FIG. 38A

57/57

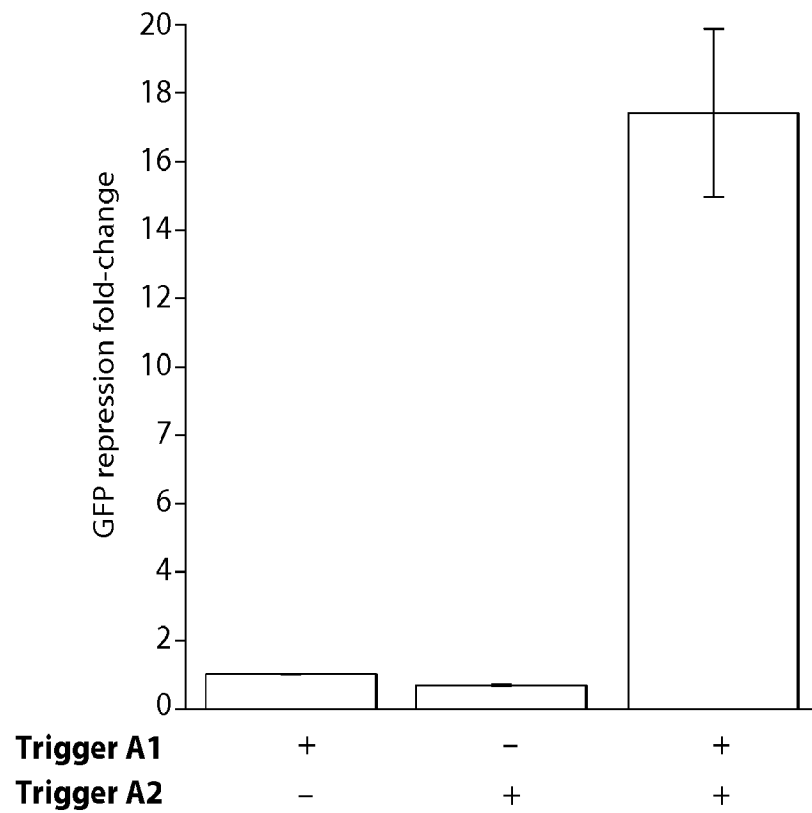


FIG. 38B

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US15/40460

A. CLASSIFICATION OF SUBJECT MATTER

IPC(8) - C12Q1/68; C12N15/11 (2015.01)

CPC - C12Q 1/68, 1/6897; C12N15/11

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): C12Q 1/68; C12N 15/11 (2015.01)

CPC: C12Q 1/6897, 1/6816; C12N 15/113, 15/11, 15/67, 2310/17, 15/115

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

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

PatSeer (US, EP, WO, JP, DE, GB, CN, FR, KR, ES, AU, IN, CA, INPADOC Data); Google; Google Scholar; PubMed; ScienceDirect; host*, cell*, riboregulator*, 'RNA', 'single-strand*', 'toehold domain', 'double-strand*', 'stem domain', 'initiation codon', 'loop domain', 'ribosom*', bind*, sit*, 'coding domain', 'taRNA*', 'trans-activating', 'Alendander Green', 'Peng Yin', 'James Collins', 'pluralit'

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X --- Y	WO 2014/074648 A2 (PRESIDENT AND FELLOWS OF HARVARD COLLEGE, et al.) May 15, 2014; page 3, lines 12-16, 31; page 4, lines 4-5, 8-12, 14-15; page 5, lines 12-14, 20, 27; page 6, lines 8, 28-32; page 7, lines 1, 5-12; page 10, line 24; page 14, lines 3-5; page 31, line 1; page 33, line 20; page 35, lines 10, 30-31; Claims 1, 27, 28, 31, 60, 62; figures 9, 12, 21	1-5, 23-26 ----- 37-41
Y	CULLER, S. et al. Reprogramming Cellular Behavior With RNA Controllers Responsive To Endogenous Proteins. Science. 26 November 2010, Vol. 330, pages 1251-1255; figure 1(C); DOI: 10.1126/science.1192128.	37-41

☐ 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

19 November 2015 (19.11.2015)

Date of mailing of the international search report

22 DEC 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

Shane Thomas

PCT Helpdesk: 571-272-4300

PCT OSP: 571-272-7774

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US15/40460

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

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

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. ☒ Claims Nos.: 6-22, 31-33, 42
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

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

This International Searching Authority found multiple inventions in this international application, as follows:
See extra sheet.

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
1-5, 23-26, 37-41

Remark on Protest

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

.-***-Continued from Box No. III: Observations Where Unity of Invention Is Lacking:

This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees must be paid.

Group I: Claims 1-5, 23-26 and 37-41 are directed toward a system comprising a host cell having, integrated or encoded into its genome, a plurality of riboregulators, each riboregulator comprising an RNA comprising (i) a single-stranded toehold domain, (ii) a fully or partially double-stranded stem domain comprising an initiation codon, (iii) a loop domain comprising a ribosome binding site, and (iv) a coding domain; and a method of detecting the presence of a plurality of RNA in a cell using said riboregulators.

Group II: Claims 27-30 and 34-36 are directed toward a method of controlling gene and/or protein expression in a cell comprising integrating or encoding a plurality of riboregulators into the genome of the cell, each riboregulator integrated upstream of a target coding sequence, and modulating expression of one or more of a plurality of trans-activating RNA (taRNA) in the cell.

The inventions listed as Groups I and II do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons: the special technical features of Group I include wherein each riboregulator detects a different endogenous RNA from all other riboregulators in the plurality, not present in Group II; the special technical features of Group II include a trans-activating RNA, not present in Group I.

Groups I and II share the technical features including a plurality of riboregulators which have been incorporated into the genome of the cell, or introduced into the cell; each riboregulator comprises an RNA comprising (i) a single-stranded toehold domain, (ii) a fully or partially double-stranded stem domain comprising an initiation codon, (iii) a loop domain comprising a ribosome binding site, and (iv) a coding domain for a protein; a reporter protein; and expression of the target coding sequence.

However, these shared technical features are previously disclosed by WO 2014/074648 A2 to President and Fellows of Harvard College et al. (hereinafter 'Harvard').

Harvard discloses a plurality of riboregulators (a plurality of crRNA riboregulators (a plurality of riboregulators; page 2, lines 25-31; page 4, lines 3-8) which have been incorporated into the genome of a cell (which are integrated (incorporated) into the host cell DNA (genome); page 29, lines 2-7), or introduced into the cell (introduced into a cell; page 5, lines 4-10); each riboregulator comprising an RNA (each riboregulator comprising an RNA; page 3, line 30 - page 4, line 2) comprising (i) a single-stranded toehold domain (comprising (i) a single-stranded toehold domain; page 3, line 30 - page 4, line 2), (ii) a fully or partially double-stranded stem domain (a fully or partially double-stranded stem domain; page 3, line 30 - page 4, line 2) comprising an initiation codon (comprising an initiation codon; page 3, line 30 - page 4, line 2), (iii) a loop domain (a loop domain; page 3, line 30 - page 4, line 2) comprising a ribosome binding site (comprising a ribosome binding site; page 3, line 30 - page 4, line 2), and (iv) a coding domain for a protein (a coding domain for a protein; page 4, lines 9-12); a reporter protein (a reporter protein; page 2, lines 25-33); and expression of the target coding sequence (expression of a downstream (target) coding sequence; page 4, lines 3-8).

Since none of the special technical features of the Groups I and II inventions is found in more than one of the inventions, and since all of the shared technical features are previously disclosed by the Harvard reference, unity of invention is lacking.