(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
10 August 2017 (10.08.2017)

WIPO | PCT

(10) International Publication Number
WO 2017/136591 A1

(51) International Patent Classification:
A61K 48/00 (2006.01) *C07H 21/04* (2006.01)
C12N 15/86 (2006.01)

(21) International Application Number:
PCT/US2017/016279

(22) International Filing Date:
2 February 2017 (02.02.2017)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/290,200 2 February 2016 (02.02.2016) US

(71) Applicants: **MEIRAGTX UK II LIMITED** [GB/GB]; 92 Britannia Walk, London, N1 7NQ (GB). **MEIRAGTX, LLC** [US/US]; 450 East 29th Street 15th Floor, New York, NY 10016 (US).

(72) Inventors: **VOLLES, Michael, J.**; 69 Bolton Street Apt. 102, Cambridge, MA 02140 (US). **DANOS, Olivier, F.**; 49 Hancock Street, Boston, MA 02114 (US). **GUO, Xuecui**; 1 Bungtown Road, Cold Spring Harbor, NY 11724 (US).

(74) Agents: **SOMERVILLE, Deborah, A.** et al.; Fox Rothschild, LLP, Princeton Pike Corporate Center, 997 Lenox Drive, Building 3, Lawrenceville, NJ 08648-2311 (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, DJ, 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, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

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

Declarations under Rule 4.17:

— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*

— *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))*

Published:

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

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

WO 2017/136591 A1

(54) Title: REGULATION OF GENE EXPRESSION THROUGH APTAMER-MODULATED POLYADENYLATION

(57) Abstract: The invention provides polynucleotide constructs for the regulation of gene expression by aptamer-based modulation of U1 small nuclear ribonucleoprotein (snRNP)-mediated suppression of polyadenylation and methods of using the constructs to regulate gene expression in response to the presence or absence of a ligand that binds the aptamer. The polynucleotide construct contains a U1 binding site in the context of a riboswitch comprising an effector region and an aptamer such that when the aptamer binds a ligand, target gene expression occurs.

REGULATION OF GENE EXPRESSION THROUGH APTAMER-MODULATED POLYADENYLATION

FIELD OF THE INVENTION

[0001] The invention provides polynucleotide constructs for the regulation of gene expression by aptamer-based modulation of U1 snRNP-mediated suppression of polyadenylation and methods of using the constructs to regulate gene expression in response to the presence or absence of a ligand that binds the aptamer. The polynucleotide construct contains a U1 binding site in the context of a riboswitch comprising an effector region and an aptamer such that when the aptamer binds a ligand, target gene expression occurs.

BACKGROUND OF THE INVENTION

[0002] Messenger RNAs (mRNAs) in eukaryotic cells are produced from pre-mRNA transcripts by extensive post-transcriptional processing, including 5′ end capping, removal of introns by splicing, and 3′ end cleavage and polyadenylation. Splicing is performed by a spliceosome, a large RNA-protein complex comprised predominantly of small nuclear ribonucleoproteins (snRNPs). The predominant (U2-type) spliceosome contains the U1, U2, U4, U6 and U5 snRNPs, whereas a much less abundant U12-type spliceosome is comprised of U11, U12, U4atac, U6atac and U5 snRNPs.

[0003] The 3′ end of almost all eukaryotic mRNAs comprises a poly(A) tail—a homopolymer of 20 to 250 adenosine residues. The poly(A) tail is added to pre-mRNA in the nucleus by cleavage and polyadenylation, a process catalyzed by a large complex of proteins. In vertebrates, addition of a poly(A) tail depends on two cis-acting RNA elements, the highly conserved AAUAAA polyadenylation sequence found upstream of the polyadenylation site and a poorly conserved GU-rich element found downstream. Addition of a poly(A) tail to mRNA protects it from degradation, among other functions.

[0004] The U1 snRNP has been shown to have splicing-independent function in pre-mRNA processing by suppressing 3′ end processing of pre-mRNA, inhibiting aberrant intronic polyadenylation signal, and ensure promoter directionality. For example, Fortes et al. (PNAS, 100:8264-69, incorporated herein by reference), placed one to three U1 binding sites in the 3′ UTR leading to reduced reporter gene expression.

SUMMARY OF THE INVENTION

[0005] In one aspect, the present invention provides a polynucleotide cassette for the regulation of the expression of a target gene comprising a riboswitch wherein the riboswitch comprises an effector region and an aptamer, wherein the effector region comprises a U1 snRNP binding site and sequence complementary to the U1 snRNP binding site. In one embodiment, the aptamer binds a small molecule ligand.

[0006] In one embodiment, the effector sequence comprises, in addition to the U1 snRNP binding site and sequence complementary to the U1 snRNP binding, additional sequence that is capable of forming a stem when the aptamer binds ligand. In one embodiment, the effector region comprises a stem-forming sequence that is 9 to 11 base pairs. In one embodiment, the effector region comprises a stem-forming sequence with one or more mismatched bases in the stem.

[0007] In one embodiment, the U1 snRNP binding site is 8 to 10 nucleotides. In one embodiment, the U1 snRNP binding site comprises the sequence CAGGTAAG. In one embodiment, the U1 snRNP binding site is selected from the group consisting of CAGGTAAGTA, CAGGTAAGT, and CAGGTAAG.

[0008] In one embodiment, the polynucleotide cassette comprises two or more riboswitches, wherein each riboswitch comprises an effector region and an aptamer, wherein the effector region comprises a U1 snRNP binding site and sequence complementary to the U1 snRNP binding site. In one embodiment, the two or more riboswitches each comprise an aptamer that binds the same ligand. In one embodiment, the two or more riboswitches comprise different aptamers that bind different ligands.

[0009] In another aspect, the present invention provides a method of modulating the expression of a target gene comprising:

[0010] (a) inserting one or more of the polynucleotide cassettes described above into the 3′ UTR of a target gene,

[0011] (b) introducing the target gene comprising the polynucleotide cassette into a cell, and

[0012] (c) exposing the cell to a ligand that specifically binds the aptamer in an amount effective to increase expression of the target gene.

[0013] In one embodiment, the ligand is a small molecule.

[0014] In one embodiment, the polynucleotide cassette is inserted about 87 and/or about 140 nucleotides 5' of the polyadenylation signal. In one embodiment, the polynucleotide cassette is inserted at one or more of about 74, about 110, or about 149 nucleotides 5' of the polyadenylation signal.

[0015] In one embodiment, two or more of the polynucleotide cassettes are inserted into the 3′ UTR of the target gene. In one embodiment, the two or more polynucleotide cassettes comprise different aptamers that specifically bind to different small molecule ligands. In one embodiment, the two or more polynucleotide cassettes comprise the same aptamer. In one embodiment, the two or more polynucleotide cassettes are inserted at different locations of the 3′ UTR of the target gene.

[0016] In another aspect, the present invention provides a target gene comprising a polynucleotide cassette described above incorporated in a vector for the expression of the target gene. In one embodiment, the vector is a viral vector. In one embodiment, the viral vector is selected from the group consisting of adenoviral vector, adeno-associated virus vector, and lentiviral vector.

[0017] In another aspect, the present invention provides a vector comprising a target gene that contains a polynucleotide cassette described herein. In one embodiment, the vector is a viral vector. In one embodiment, the viral vector is selected from the group consisting of adenoviral vector, adeno-associated virus vector, and lentiviral vector.

BRIEF DESCRIPTION OF THE FIGURES

[0018] Figures 1a-1b. U1 binding site in 3′ UTR suppresses target gene (luciferase) expression.

[0019] Figure 1a. Schematics for the insertion of U1 binding sequence in 3′ UTR at different position in pRL-SV40 vector. The wild type and mutant U1 consensus binding sequences are listed.

[0020] Figure 1b. Results of Renilla luciferase assay. HEK 293 cells were transfected with the indicated constructs containing at various positions in the 3′ UTR either the consensus U1 binding sequence or the mutant sequence. The results were expressed as mean±SD (n=3), and the reduction fold was calculated as ratio of luciferase activity from mutant construct to luciferase activity from wild type construct.

[0021] Figures 2a-2b. Effect of the length of the U1 binding sequence in the 3′ UTR on suppression of gene expression.

[0022] Figure 2a. The full length and the truncated sequences of U1 binding sites inserted at -87 position in 3′ UTR.

[0023] Figure 2b. HEK 293 cells were transfected with the indicated constructs containing full length or truncated U1 binding sequences inserted at -87 and luciferase activity was determined.

[0024] Figures 3a-3b. The effect of stem-loop structure at the U1 site in 3′ UTR on gene expression.

[0025] Figure 3a. The sequence of the stem-loop structure that embedded the U1 binding site in the stem. Mutations were made to generate a broken stem (shown on right).

[0026] Figure 3b. Results of luciferase activity assay indicate that the stem-loop structure sequestering the U1 binding site completely abolished the suppressive effect of U1 interference on gene expression.

[0027] Figures 4a-4c. Use of a theophylline aptamer to regulate U1 interference in 3′ UTR of a target gene.

[0028] Figure 4a. The schematics showing the aptamer-modulated U1 interference in 3′ UTR and gene expression. In the absence of aptamer ligand (top panel), a U1 site inserted in the 3′ UTR is available for U1 snRNP binding and U1-mediated polyadenylation interference, thus suppressing target gene expression. In the presence of aptamer ligand (lower panel), aptamer/ligand binding leads to the formation of stem which sequesters the U1 binding site from U1 snRNP binding. U1-mediated polyadenylation interference is abolished, resulting in target gene expression.

[0029] Figure 4b. The sequence of the effector region stem that sequesters the U1 binding site and connects the theophylline aptamer, generated by serial truncation of the hairpin stem and lower stem of theophylline aptamer. Shown are the stem sequences for constructs U1 theo_1, 8 and 9. Theophylline is shown as ⩓.

[0030] Figure 4c. HEK 293 cells were transfected with the indicated constructs, and treated with or without 3 mM theophylline. The induction fold was calculated as the ratio of the value of luciferase activity obtained from theophylline treated cells to the luciferase activity obtained from untreated cells.

[0031] Figures 5a-5c. Optimization of effector region stem sequence to enhance the regulatability of U1 site accessibility.

[0032] Figure 5a. The stem sequences of U1_theo_8 and U1_theo_9, with the sequential mutations listed beside corresponding nucleotides. The sequential mutations create a mismatch in the effector region stem.

[0033] Figure 5b. A total of 9 constructs were generated through sequential mutation of U1_theo_8 and were each transfected into HEK 293 cells. Transfected cells were treated with or without 3 mM theophylline and luciferase activity measured. The results were expressed as mean±SD, and the induction fold was indicated for construct U1_theo_8 to U1_8_5.

[0034] Figure 5c. A total of 8 constructs were generated through sequential mutation of U1_theo_9 and were each transfected into HEK 293 cells. Transfected cells were treated with or without 3 mM theophylline and luciferase activity measured. The results were expressed as mean±SD.

[0035] Figures 6a-6b. Effect of multiple U1_theo riboswitches in the 3′ UTR on regulation of target gene expression.

[0036] Figure 6a. A schematic showing the 4 copies of U1_theo_8_1 in the 3′ UTR of the target gene (luciferase).

[0037] Figure 6b. HEK 293 were transfected with the indicated constructs, and treated with theophylline at different concentrations. Four copies of U1_theo_8_1 in the 3′ UTR yielded a 4.3-fold induction of luciferase activity at 6 mM concentration.

[0038] Figures 7a-7c. Use of guanine aptamer to modulate U1 interference.

[0039] Figure 7a. Schematic indicating the U1_Gua sequence at different position in 3′ UTR containing SV40 early poly(A) signal.

[0040] Figure 7b. Schematic showing the effector region stem sequence in U1_Gua switch that sequesters the U1 site (in the presence of guanine ligand) and connects guanine aptamer. Guanine is shown as ⋙.

[0041] Figure 7c. Guanine aptamer regulates U1 interference with polyadenylation in response to guanine treatment. HEK 293 cells were transfected with the indicated constructs. Constructs containing the mutant U1 site and its complementary sequence were used as

control. Results were expressed as mean±SD, the induction folds are shown for each construct.

[0042] Figure 8. U1_Gua riboswitch functions in the context of human beta globin polyA sequence.

DETAILED DESCRIPTION OF THE INVENTION

[0043] The invention provides polynucleotide constructs for the regulation of gene expression by aptamer-based modulation of U1 snRNP-mediated suppression of polyadenylation and methods of using the constructs to regulate gene expression in response to the presence or absence of a ligand that binds the aptamer. The polynucleotide construct contains at least one riboswitch that contains an effector region and a sensor region. The effector region contains a U1 snRNP binding site and sequence complementary to the U1 snRNP binding site such that the two sequences are capable of forming a stem that sequesters the U1 snRNP binding site thereby preventing binding of the U1 snRNP. The effector region may contain additional sequence and its complement so that the effector region stem is longer than the U1 snRNP binding site and its complementary sequence. The sensor region comprises an RNA sequence that binds a ligand, and in response to this binding alters the conformation of the effector region. In one embodiment, the sensor region comprises an aptamer. When the aptamer ligand is not present, the effector region does not form a stem. The U1 snRNP binding site is available and is bound by a U1 snRNP, which inhibits polyadenylation of the mRNA. When the aptamer ligand is present, it binds the aptamer causing the effector region to form a stem, which prevents U1 snRNP from binding the U1 binding site. The U1 snRNP is not recruited to the 3′ UTR of the target gene mRNA and polyadenylation of the message occurs.

[0044] The gene regulation polynucleotide cassette refers to a recombinant DNA construct that, when incorporated into the DNA of a target gene in the 3′ UTR, provides the ability to regulate expression of the target gene by aptamer/ligand mediated suppression of polyadenylation by U1 snRNP. As used herein, a polynucleotide cassette or construct is a nucleic acid (e.g., DNA or RNA) comprising elements derived from different sources (e.g., different organisms, different genes from the same organism, and the like). The polynucleotide cassette comprises a riboswitch. The riboswitch in the context of the present invention contains a sensor region (e.g., an aptamer) and an effector region that together are responsible for sensing the presence of a ligand that binds the sensor region and altering the

conformation of the effector region that contains a U1 snRNP binding site and sequence that is complementary to the U1 snRNP binding site. In one embodiment, the target gene's expression is increased when the aptamer ligand is present and decreased when the ligand is absent.

[0045] Riboswitch

[0046] The term "riboswitch" as used herein refers to a regulatory segment of a RNA polynucleotide. A riboswitch in the context of the present invention contains a sensor region (e.g., an aptamer) and an effector region that together are responsible for sensing the presence of a ligand (e.g., a small molecule) and modulating the accessibility of a U1 snRNP binding site located in the effector region. In one embodiment, the riboswitch is recombinant, utilizing polynucleotides from two or more sources. The term "synthetic" as used herein in the context of a riboswitch refers to a riboswitch that is not naturally occurring. In one embodiment, the sensor and effector regions are joined by a polynucleotide linker. In one embodiment, the polynucleotide linker forms a RNA stem (i.e., a region of the RNA polynucleotide that is double-stranded).

[0047] Effector region

[0048] The effector region of the riboswitch comprises RNA sequence that, in response to ligand binding the sensor region (e.g., an aptamer), forms a stem structure (a double-stranded region) that lowers the accessibility of a U1 snRNP binding site to the U1 snRNP. The effector region comprises a U1 snRNP binding site and sequence complimentary the U1 snRNP binding site. When the aptamer binds its ligand, the effector region forms a stem and thus sequesters the U1 snRNP binding site from binding a U1 snRNP. Under certain conditions (for example, when the aptamer is not bound to its ligand), the effector region is in a context that provides access to the U1 snRNP binding site, allowing U1 snRNP to bind the mRNA and inhibit polyadenylation leading to degradation of the message.

[0049] The U1 snRNP binding site can be any polynucleotide sequence that is capable of binding the U1 snRNP, thereby recruiting the U1 snRNP to the 3′ UTR of a target gene and suppressing polyadenylation of the target gene message. In one embodiment, the U1 snRNP binding site (also referred to herein as the "U1 binding site" or "U1 site") is the consensus site CAGGTAAGTA (CAGGUAAGUA when in the mRNA). In some embodiments, the U1 snRNP binding site is a variation of this consensus sequence, including

for example sequences that are shorter or have one or more nucleotides changed from the consensus sequence. In one embodiment, the U1 snRNP binding site contains the sequence CAGGTAAG. In some embodiments, the binding site is encoded by the sequence selected from CAGGTAAGTA, CAGGTAAGT, and CAGGTAAG. The U1 snRNP binding site can be any 5′ splice site from a gene, e.g., the 5′ splice site from human DHFR exon 2 (*see* Examples 7 and 8).

[0050] The stem portion of the effector region should be of a sufficient length (and GC content) to inhibit U1 snRNP binding the U1 snRNP binding site upon ligand binding the aptamer, while also allowing access to the U1 snRNP binding site when the ligand is not present in sufficient quantities. In embodiments of the invention, the stem portion of the effector region comprises stem sequence in addition to U1 snRNP binding site and its complementary sequence. In embodiments of the invention, this additional stem sequence comprises sequence from the aptamer stem. The length and sequence of the stem portion can be modified using known techniques in order to identify stems that allow acceptable background expression of the target gene when no ligand is present and acceptable expression levels of the target gene when the ligand is present. If the stem is, for example, too long it may hide access to the U1 snRNP binding site in the presence or absence of ligand. If the stem is too short, it may not form a stable stem capable of sequestering the U1 snRNP binding site, in which case U1 snRNP will bind and inhibit polyadenylation of the message (leading to degradation of the target gene mRNA) in the presence or absence of ligand. In one embodiment, the total length of the effector region stem is between about 7 base pairs to about 20 base pairs. In some embodiments, the length of the stem is between about 8 base pairs to about 11 base pairs. In some embodiments, the length of the stem is 8 base pairs to 11 base pairs. In addition to the length of the stem, the GC base pair content of the stem can be altered to modify the stability of the stem.

[0051] In some embodiments, the effector region stem contains one or more mismatched nucleotides that do not base pair with the complementary portion of the effector region stem.

[0052] **Aptamer/Ligand**

[0053] In one embodiment, the sensor region comprises an aptamer. The term "aptamer" as used herein refers to an RNA polynucleotide that specifically binds to a ligand. The term "ligand" refers to a molecule that is specifically bound by an aptamer. In one

embodiment, the ligand is a low molecular weight (less than about 1,000 Daltons) molecule including, for example, lipids, monosaccharides, second messengers, co-factors, metal ions, other natural products and metabolites, nucleic acids, as well as most therapeutic drugs. In one embodiment, the ligand is a polynucleotide with two or more nucleotide bases.

[0054] In one embodiment, the ligand is selected from the group consisting of 8-azaguanine, adenosine 5'-monophosphate monohydrate, amphotericin B, avermectin B1, azathioprine, chlormadinone acetate, mercaptopurine, moricizine hydrochloride, N6-methyladenosine, nadide, progesterone, promazine hydrochloride, pyrvinium pamoate, sulfaguanidine, testosterone propionate, thioguanosine, tyloxapol and vorinostat.

[0055] Aptamer ligands can also be cell endogenous components that increase significantly under specific physiological/pathological conditions, such as oncogenic transformation – these may include second messenger molecules such as GTP or GDP, calcium; fatty acids, or fatty acids that are incorrectly metabolized such as 13-HODE in breast cancer (Flaherty, JT et al., Plos One, Vol. 8, e63076, 2013, incorporated herein by reference); amino acids or amino acid metabolites; metabolites in the glycolysis pathway that usually have higher levels in cancer cells or in normal cells in metabolic diseases; and cancer-associated molecules such as Ras or mutant Ras protein, mutant EGFR in lung cancer, indoleamine-2,3-dioxygenase (IDO) in many types of cancers. Endogenous ligands include progesterone metabolites in breast cancer as disclosed by JP Wiebe (Endocrine-Related Cancer (2006) 13:717–738, incorporated herein by reference). Endogenous ligands also include metabolites with increased levels resulting from mutations in key metabolic enzymes in kidney cancer such as lactate, glutathione, kynurenine as disclosed by Minton, DR and Nanus, DM (Nature Reviews, Urology, Vol. 12, 2005, incorporated herein by reference).

[0056] Aptamers have binding regions that are capable of forming complexes with an intended target molecule (i.e., the ligand). The specificity of the binding can be defined in terms of the comparative dissociation constants (Kd) of the aptamer for its ligand as compared to the dissociation constant of the aptamer for unrelated molecules. Thus, the ligand is a molecule that binds to the aptamer with greater affinity than to unrelated material. Typically, the Kd for the aptamer with respect to its ligand will be at least about 10-fold less than the Kd for the aptamer with unrelated molecules. In other embodiments, the Kd will be at least about 20-fold less, at least about 50-fold less, at least about 100-fold less, and at least about 200-fold less. An aptamer will typically be between about 15 and about 200

nucleotides in length. More commonly, an aptamer will be between about 30 and about 100 nucleotides in length.

[0057] The aptamers that can be incorporated as part of the riboswitch can be a naturally occurring aptamer, or modifications thereof, or aptamers that are designed de novo and/or screened through systemic evolution of ligands by exponential enrichment (SELEX) or other screening methods. Examples of aptamers that bind small molecule ligands include, but are not limited to theophylline, dopamine, sulforhodamine B, cellobiose, kanamycin A, lividomycin, tobramycin, neomycin B, viomycin, chloramphenicol, streptomycin, cytokines, cell surface molecules, and metabolites. For a review of aptamers that recognize small molecules, *see, e.g.*, Famulok, Science 9:324-9 (1999) and McKeague, M. & DeRosa, M.C. J. Nuc. Aci. 2012 (both of which are incorporated herein by reference). In another embodiment, the aptamer is a complementary polynucleotide.

[0058] Methods for identifying aptamer/ligand

[0059] In one embodiment, the aptamer is designed to bind a particular small molecule ligand. Methods for designing and selecting aptamers that bind particular ligands are disclosed in 62/370,599, incorporated herein by reference. Other methods for screening aptamers include, for example, SELEX. Methods for designing aptamers that selectively bind a small molecule using SELEX are disclosed in, e.g., U.S. Patent Nos. 5,475,096, 5,270,163, and Abdullah Ozer, et al. Nuc. Aci. 2014, which are incorporated herein by reference. Modifications of the SELEX process are described in U.S. Patent Nos. 5,580,737 and 5,567,588, which are incorporated herein by reference.

[0060] Selection techniques for identifying aptamers generally involve preparing a large pool of DNA or RNA molecules of the desired length that contain a region that is randomized or mutagenized. For example, an oligonucleotide pool for aptamer selection might contain a region of 20-100 randomized nucleotides flanked by regions of defined sequence that are about 15-25 nucleotides long and useful for the binding of PCR primers. The oligonucleotide pool is amplified using standard PCR techniques, or other means that allow amplification of selected nucleic acid sequences. The DNA pool may be transcribed *in vitro* to produce a pool of RNA transcripts when an RNA aptamer is desired. The pool of RNA or DNA oligonucleotides is then subjected to a selection based on their ability to bind specifically to the desired ligand. Selection techniques include, for example, affinity chromatography, although any protocol which will allow selection of nucleic acids based on

their ability to bind specifically to another molecule may be used. Selection techniques for identifying aptamers that bind small molecules and function within a cell may involve cell based screening methods. In the case of affinity chromatography, the oligonucleotides are contacted with the target ligand that has been immobilized on a substrate in a column or on magnetic beads. The oligonucleotide is preferably selected for ligand binding in the presence of salt concentrations, temperatures, and other conditions which mimic normal physiological conditions. Oligonucleotides in the pool that bind to the ligand are retained on the column or bead, and nonbinding sequences are washed away. The oligonucleotides that bind the ligand are then amplified (after reverse transcription if RNA transcripts were utilized) by PCR (usually after elution). The selection process is repeated on the selected sequences for a total of about three to ten iterative rounds of the selection procedure. The resulting oligonucleotides are then amplified, cloned, and sequenced using standard procedures to identify the sequences of the oligonucleotides that are capable of binding the target ligand. Once an aptamer sequence has been identified, the aptamer may be further optimized by performing additional rounds of selection starting from a pool of oligonucleotides comprising a mutagenized aptamer sequence.

[0061] In vivo aptamer screening may be used following one or more rounds of in vitro selection (e.g., SELEX). For example, Konig, J. et al. (RNA. 2007, 13(4):614–622, incorporated herein by reference) describe combining SELEX and a yeast three-hybrid system for in vivo selection of aptamer.

[0062] **Target genes**

[0063] The gene regulation cassette of the present invention is a platform that can be used to regulate the expression of any target gene that can be expressed in a target cell, tissue or organism. The term “target gene” refers to a polynucleotide that is introduced into a cell and is capable of being transcribed into RNA and translated and/or expressed under appropriate conditions. Alternatively, the target gene is endogenous to the target cell and the gene regulation cassette of the present invention is positioned into the target gene (for example into the 5′ or 3′ UTR of an endogenous target gene). An example of a target gene is a polynucleotide encoding a therapeutic polypeptide. In another embodiment, the target gene encodes an RNA such as a miRNA, rRNA, small or long noncoding RNAs, short hairpin RNA (shRNA) and any other regulatory RNAs. In one embodiment, the target gene is exogenous to the cell in which the recombinant DNA construct is to be transcribed. In

another embodiment, the target gene is endogenous to the cell in which the recombinant DNA construct is to be transcribed.

[0064] The target gene according to the present invention may be a gene encoding a protein, or a sequence encoding a non-protein coding RNA. The target gene may be, for example, a gene encoding a structural protein, an enzyme, a cell signaling protein, a mitochondrial protein, a zinc finger protein, a hormone, a transport protein, a growth factor, a cytokine, an intracellular protein, an extracellular protein, a transmembrane protein, a cytoplasmic protein, a nuclear protein, a receptor molecule, an RNA binding protein, a DNA binding protein, a transcription factor, translational machinery, a channel protein, a motor protein, a cell adhesion molecule, a mitochondrial protein, a metabolic enzyme, a kinase, a phosphatase, exchange factors, a chaperone protein, and modulators of any of these. In embodiments, the target gene encodes erythropoietin (Epo), human growth hormone (hGH), transcription activator-like effector nucleases (TALEN), human insulin, CRISPR associated protein 9 (cas9), or an immunoglobulin (or portion thereof), including, e.g., a therapeutic antibody.

[0065] **Expression Constructs**

[0066] The present invention contemplates the use of a recombinant vector for introduction into target cells of a polynucleotide encoding a target gene and containing the gene regulation cassette described herein. In many embodiments, the recombinant DNA construct of this invention includes additional DNA elements including DNA segments that provide for the replication of the DNA in a host cell and expression of the target gene in that cell at appropriate levels. The ordinarily skilled artisan appreciates that expression control sequences (promoters, enhancers, and the like) are selected based on their ability to promote expression of the target gene in the target cell. "Vector" means a recombinant plasmid, yeast artificial chromosome (YAC), mini chromosome, DNA mini-circle or virus (including virus derived sequences) that comprises a polynucleotide to be delivered into a host cell, either in vitro or in vivo. In one embodiment, the recombinant vector is a viral vector or a combination of multiple viral vectors.

[0067] Viral vectors for the aptamer-mediated expression of a target gene in a target cell, tissue, or organism are known in the art and include adenoviral (AV) vectors, adeno-associated virus (AAV) vectors, retroviral and lentiviral vectors, and Herpes simplex type 1 (HSV1) vectors.

[0068] Adenoviral vectors include, for example, those based on human adenovirus type 2 and human adenovirus type 5 that have been made replication defective through deletions in the E1 and E3 regions. The transcriptional cassette can be inserted into the E1 region, yielding a recombinant E1/E3-deleted AV vector. Adenoviral vectors also include helper-dependent high-capacity adenoviral vectors (also known as high-capacity, "gutless" or "gutted" vectors), which do not contain viral coding sequences. These vectors, contain the cis-acting elements needed for viral DNA replication and packaging, mainly the inverted terminal repeat sequences (ITR) and the packaging signal (Ψ). These helper-dependent AV vector genomes have the potential to carry from a few hundred base pairs up to approximately 36 kb of foreign DNA.

[0069] Recombinant adeno-associated virus "rAAV" vectors include any vector derived from any adeno-associated virus serotype, including, without limitation, AAV-1, AAV-2, AAV-3, AAV-4, AAV-5, AAV-7 and AAV-8, AAV-9, AAV-10, and the like. rAAV vectors can have one or more of the AAV wild-type genes deleted in whole or in part, preferably the Rep and/or Cap genes, but retain functional flanking ITR sequences. Functional ITR sequences are retained for the rescue, replication, packaging and potential chromosomal integration of the AAV genome. The ITRs need not be the wild-type nucleotide sequences, and may be altered (e.g., by the insertion, deletion or substitution of nucleotides) so long as the sequences provide for functional rescue, replication and packaging.

[0070] Alternatively, other systems such as lentiviral vectors can be used in embodiments of the invention. Lentiviral-based systems can transduce non-dividing as well as dividing cells making them useful for applications targeting, for examples, the non-dividing cells of the CNS. Lentiviral vectors are derived from the human immunodeficiency virus and, like that virus, integrate into the host genome providing the potential for long-term gene expression.

[0071] Polynucleotides, including plasmids, YACs, minichromosomes and minicircles, carrying the target gene containing the gene regulation cassette can also be introduced into a cell or organism by nonviral vector systems using, for example, cationic lipids, polymers, or both as carriers. Conjugated poly-L-lysine (PLL) polymer and polyethylenimine (PEI) polymer systems can also be used to deliver the vector to cells. Other methods for delivering the vector to cells includes hydrodynamic injection and electroporation and use of ultrasound, both for cell culture and for organisms. For a review

of viral and non-viral delivery systems for gene delivery see Nayerossadat, N. et al. (Adv Biomed Res. 2012; 1:27; incorporated herein by reference).

[0072] Methods of Modulating Expression of a Target Gene

[0073] In one aspect, this invention provides a method of modulating expression of a target gene (e.g., a therapeutic gene), by (a) inserting the gene regulation cassette of the present invention into the 3′ UTR of a target gene; (b) introducing the target gene comprising the gene regulation cassette into a cell; and (c) exposing the cell to a ligand that binds the aptamer. In one embodiment, the ligand is a small molecule. In aspects, expression of the target gene in target cells confers a desired property to a cell into which it was introduced, or otherwise leads to a desired therapeutic outcome.

[0074] In one embodiment, one or more gene regulation cassettes are inserted into the 3′ untranslated region of the target gene. In one embodiment, a single gene regulation cassette is inserted into the 3′ UTR of a target gene. In other embodiments 2, 3, 4, or more gene regulation cassettes are inserted in the target gene. In one embodiment, two gene regulation cassettes are inserted into the target gene. When multiple gene regulation cassettes are inserted into a target gene, they each can contain the same aptamer such that a single ligand can be used to modulate ribonuclease cleavage of the multiple cassettes and thereby modulate target gene expression. In other embodiments, multiple gene regulation cassettes are inserted into a target gene, each can contain a different aptamer so that exposure to multiple different small molecule ligands modulates target gene expression. In other embodiments, multiple gene regulation cassettes are inserted into a target gene, each containing different ribonuclease substrate sequences. This may be useful in reducing recombination and improving ease of incorporation into viral vectors.

[0075] The polynucleotide cassettes of the present invention are effective at modulating target gene expression when the polynucleotide cassette is located in the 3′ UTR of the target gene at any location upstream (5′) of the polyadenylation signal (e.g., AATAAA). The polynucleotide cassette is also effective at modulating target gene expression by blocking polyadenylation when different poly(A) signal sequences are present, for example the SV40 early versus late poly(A) signals (*see, e.g.*, Examples 4-6, 8). In one embodiment, the at least one polynucleotide cassette of the present invention is inserted about 87 or about 140 nucleotides 5' of the polyadenylation sequence. In one embodiment, a polynucleotide cassette of the present invention is inserted at, or near, both locations. In

other embodiments, a polynucleotide cassette is inserted at one or more of about 74, about 110, or about 149 nucleotides 5′ of the polyadenylation signal.

[0076] The polynucleotide cassette of the present invention can be used in combination with other mechanisms for the regulation of expression of the target gene. In one embodiment, a polynucleotide cassette of the present invention is used in combination with a gene regulation cassette that modulates target gene expression by aptamer-mediated regulation of alternative splicing as described in WO 2016/126747, incorporated herein by reference.

[0077] **Methods of Treatment and Pharmaceutical Compositions**

[0078] One aspect of the invention provides a method of regulating the level of a therapeutic protein delivered by gene therapy. In this embodiment, the "target gene" may encode the therapeutic protein. The "target gene" may encode a protein that is endogenous or exogenous to the cell.

[0079] The therapeutic gene sequence containing the regulatory cassette with aptamer-driven riboswitch is delivered to target cells in vitro or ex vivo, e.g., by a vector. The cell specificity of the "target gene" may be controlled by promoter or other elements within the vector. Delivery of the vector construct containing the target gene and the polynucleotide cassette, and the transfection of the target tissues resulting in stable transfection of the regulated target gene, is often the first step in producing the therapeutic protein.

[0080] However, due to the presence of the regulatory cassette within the target gene sequence, the target gene is not expressed at significant levels, i.e., it is in the "off state" in the absence of the specific ligand that binds to the aptamer contained within in the regulatory cassette riboswitch. Only when the aptamer specific ligand is administered (or otherwise present in sufficient quantities) is the target gene expression activated.

[0081] The delivery of the vector construct containing the target gene and the delivery of the activating ligand generally are separated in time. The delivery of the activating ligand will control when the target gene is expressed, as well as the level of protein expression. The ligand may be delivered by a number of routes including, but not limited to, oral, intramuscular (IM), intravenous (IV), intraocular, or topically.

[0082] The timing of delivery of the ligand will depend on the requirement for activation of the target gene. For example, if the therapeutic protein encoded by the target

gene is required constantly, an oral small molecule ligand may be delivered daily, or multiple times a day, to ensure continual activation of the target gene, and thus continual expression of the therapeutic protein. If the target gene has a long acting effect, the inducing ligand may be dosed less frequently.

[0083] This invention allows the expression of the therapeutic transgene to be controlled temporally, in a manner determined by the temporal dosing of the ligand specific to the aptamer within the riboswitch of the regulatory polynucleotide cassette. The increased expression of the therapeutic transgene only on ligand administration, increases the safety of a gene therapy treatment by allowing the target gene to be off in the absence of the ligand.

[0084] Different aptamers can be used to allow different ligands to activate target genes. In certain embodiments of the invention, each therapeutic gene containing a regulatory cassette will have a specific aptamer within the cassette that will be activated by a specific small molecule. This means that each therapeutic gene can be activated only by the ligand specific to the aptamer housed within it. In these embodiments, each ligand will only activate one therapeutic gene. This allows for the possibility that several different "target genes" may be delivered to one individual and each will be activated on delivery of the specific ligand for the aptamer contained within the regulatory cassette housed in each target gene.

[0085] This invention allows any therapeutic protein whose gene can be delivered to the body (such as erythropoietin (EPO) or a therapeutic antibody) to be produced by the body when the activating ligand is delivered. This method of therapeutic protein delivery may replace the manufacture of such therapeutic proteins outside of the body which are then injected or infused, e.g., antibodies used in cancer or to block inflammatory or autoimmune disease. The body containing the regulated target gene becomes the biologics manufacturing factory, which is switched on when the gene-specific ligand is administered.

[0086] Dosing levels and timing of dosing of a therapeutic protein may be important to therapeutic effect. For example, in the delivery of AVASTIN (anti-VEGF antibody) for cancer. The present invention increases the ease of dosing in response to monitoring for therapeutic protein levels and effects.

[0087] In one embodiment, the target gene may encode a nuclease that can target and edit a particular DNA sequence. Such nucleases include Cas9, zinc finger containing nucleases, or TALENs. In the case of these nucleases, the nuclease protein may be required

for only a short period of time that is sufficient to edit the target endogenous genes. However, if an unregulated nuclease gene is delivered to the body, this protein may be present for the rest of the life of the cell. In the case of nucleases, there is an increasing risk of off-target editing the longer the nuclease is present. Regulation of expression of such proteins has a significant safety advantage. In this case, vector containing the nuclease target gene containing a regulatory cassette could be delivered to the appropriate cells in the body. The target gene is in the "off" state in the absence of the cassette-specific ligand, so no nuclease is produced. Only when the activating ligand is administered, is the nuclease produced. When sufficient time has elapsed allowing sufficient editing to occur, the ligand will be withdrawn and not administered again. Thus, the nuclease gene is thereafter in the "off" state and no further nuclease is produced and editing stops. This approach may be used to correct genetic conditions, including a number of inherited retinopathies such as LCA10 caused by mutations in CEP290 and Stargardt's Disease caused by mutations in ABCA4.

[0088] Administration of a regulated target gene encoding a therapeutic protein which is activated only on specific ligand administration may be used to regulate therapeutic genes to treat many different types of diseases, e.g., cancer with therapeutic antibodies, immune disorders with immune modulatory proteins or antibodies, metabolic diseases, rare diseases such as PNH with anti-C5 antibodies or antibody fragments as the regulated gene, or ocular angiogenesis with therapeutic antibodies, and dry AMD with immune modulatory proteins.

[0089] A wide variety of specific target genes, allowing for the treatment of a wide variety of specific diseases and conditions, are suitable for use in the present invention. For example, insulin or an insulin analog (preferably human insulin or an analog of human insulin) may be used as the target gene to treat type I diabetes, type II diabetes, or metabolic syndrome; human growth hormone may be used as the target gene to treat children with growth disorders or growth hormone-deficient adults; erythropoietin (preferably human erythropoietin) may be used as the target gene to treat anemia due to chronic kidney disease, anemia due to myelodysplasia, or anemia due to cancer chemotherapy.

[0090] The present invention may be especially suitable for treating diseases caused by single gene defects such as cystic fibrosis, hemophilia, muscular dystrophy, thalassemia, or sickle cell anemia. Thus, human β-, γ-, δ-, or ζ-globin may be used as the target gene to treat β-thalassemia or sickle cell anemia; human Factor VIII or Factor IX may be used as the target gene to treat hemophilia A or hemophilia B.

[0091] The ligands used in the present invention are generally combined with one or more pharmaceutically acceptable carriers to form pharmaceutical compositions suitable for administration to a patient. Pharmaceutically acceptable carriers include solvents, binders, diluents, disintegrants, lubricants, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, generally used in the pharmaceutical arts. Pharmaceutical compositions may be in the form of tablets, pills, capsules, troches, and the like, and are formulated to be compatible with their intended route of administration. Examples of routes of administration include parenteral, e.g., intravenous, intradermal, intranasal, subcutaneous, oral, inhalation, transdermal (topical), transmucosal, and rectal.

[0092] The pharmaceutical compositions comprising ligands are administered to a patient in a dosing schedule such that an amount of ligand sufficient to desirably regulate the target gene is delivered to the patient. When the ligand is a small molecule and the dosage form is a tablet, capsule, or the like, preferably the pharmaceutical composition comprises from 0.1 mg to 10 g of ligand; from 0.5 mg to 5 g of ligand; from 1 mg to 1 g of ligand; from 2 mg to 750 mg of ligand; from 5 mg to 500 mg of ligand; or from 10 mg to 250 mg of ligand.

[0093] The pharmaceutical compositions may be dosed once per day or multiple times per day (e.g., 2, 3, 4, 5, or more times per day). Alternatively, pharmaceutical compositions may be dosed less often than once per day, e.g., once every 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, or 14 days, or once a month or once every few months. In some embodiments of the invention, the pharmaceutical compositions may be administered to a patient only a small number of times, e.g., once, twice, three times, etc.

[0094] The present invention provides a method of treating a patient in need of increased expression of a therapeutic protein encoded by a target gene, the method comprising administering to the patient a pharmaceutical composition comprising a ligand for an aptamer, where the patient previously had been administered a recombinant DNA comprising the target gene, where the target gene contains a gene regulation cassette of the present invention that provides the ability to regulate expression of the target gene by the ligand of the aptamer by alternative splicing of pre-mRNA of the target gene, thereby increasing expression of the therapeutic protein.

[0095] **Articles of Manufacture and Kits**

[0096] Also provided are kits or articles of manufacture for use in the methods described herein. In aspects, the kits comprise the compositions described herein (e.g., for compositions for delivery of a vector comprising the target gene containing the gene regulation cassette) in suitable packaging. Suitable packaging for compositions (such as ocular compositions for injection) described herein are known in the art, and include, for example, vials (such as sealed vials), vessels, ampules, bottles, jars, flexible packaging (e.g., sealed Mylar or plastic bags), and the like. These articles of manufacture may further be sterilized and/or sealed.

[0097] The present invention also provides kits comprising compositions described herein and may further comprise instruction(s) on methods of using the composition, such as uses described herein. The kits described herein may further include other materials desirable from a commercial and user standpoint, including other buffers, diluents, filters, needles, syringes, and package inserts with instructions for performing the administration, including e.g., any methods described herein. For example, in some embodiments, the kit comprises rAAV for expression of the target gene comprising the gene regulation cassette of the present invention, a pharmaceutically acceptable carrier suitable for injection, and one or more of: a buffer, a diluent, a filter, a needle, a syringe, and a package insert with instructions for performing the injections. In some embodiments, the kit is suitable for intraocular injection, intramuscular injection, intravenous injection and the like.

[0098] "Homology" and "homologous" as used herein refer to the percent of identity between two polynucleotide sequences or between two polypeptide sequences. The correspondence between one sequence to another can be determined by techniques known in the art. For example, homology can be determined by a direct comparison of two polypeptide molecules by aligning the sequence information and using readily available computer programs. Two polynucleotide or two polypeptide sequences are "substantially homologous" to each other when, after optimally aligned with appropriate insertions or deletions, at least about 80%, at least about 85%, at least about 90%, and at least about 95% of the nucleotides or amino acids, respectively, match over a defined length of the molecules, as determined using the methods above.

[0099] "Percent sequence identity" with respect to a reference polypeptide or nucleic acid sequence is defined as the percentage of amino acid residues or nucleotides in a candidate sequence that are identical with the amino acid residues or nucleotides in the reference polypeptide or nucleic acid sequence, after aligning the sequences and introducing

gaps, if necessary, to achieve the maximum percent sequence identity. Alignment for purposes of determining percent amino acid or nucleic acid sequence identity can be achieved in ways known to the ordinarily-skilled artisan, for example, using publicly available computer software programs including BLAST, BLAST-2, ALIGN or Megalign (DNASTAR) software.

[0100] The term "polynucleotide" or "nucleic acid" as used herein refers to a polymeric form of nucleotides of any length, either ribonucleotides or deoxyribonucleotides. Thus, this term includes, but is not limited to, single-, double- or multi- stranded DNA or RNA, genomic DNA, cDNA, DNA-RNA hybrids, or a polymer comprising purine and pyrimidine bases, or other natural, chemically or biochemically modified, non-natural, or derivatized nucleotide bases.

[0101] "Heterologous" or "exogenous" means derived from a genotypically distinct entity from that of the rest of the entity to which it is compared or into which it is introduced or incorporated. For example, a polynucleotide introduced by genetic engineering techniques into a different cell type is a heterologous polynucleotide (and, when expressed, can encode a heterologous polypeptide). Similarly, a cellular sequence (e.g., a gene or portion thereof) that is incorporated into a viral vector is a heterologous nucleotide sequence with respect to the vector.

[0102] The below table contains a listing of the DNA sequences of the constructs described herein as well as other sequences as described. The U1 binding site and its mutant sequence are in grey highlighted lower case letters; the stem-loop structure (SL) is in underlined lower case letters; aptamer sequences are in wave underlined lower case letters; poly(A) signal sequence is in thick underlined lower case letters; and coding sequences for luciferase genes are in upper case letters.

SEQ ID NO.:	Description	Sequence
1	Consensus U1 binding site	CAGGTAAGTA
2	Mutant U1 binding site	CATGGAACTA
3	U1-87 wt in pRL-SV40	TCAAATCGTTCGTTGAGCGAGTTCTCAAAAATGAA CAATAAttctagagcggccgcttcgagcagacatgataagatacattgatgagt ttggacaaaccaggtaagtacaactagaatgcagtgaaaaaaatgctttatttgtgaa atttgtgatgctattgctttatttgtaaccattataagctgcaataaacaagttaacaacaa c
4	U1-87 mut in pRL-	TCAAATCGTTCGTTGAGCGAGTTCTCAAAAATGAA

SEQ ID NO.:	Description	Sequence
	SV40	CAATAAttctagagcggccgcttcgagcagacatgataagatacattgatgagt ttggacaaaccatggnactacaactagaatgcagtgaaaaaaatgctttatttgtgaaa tttgtgatgctattgctttatttgtaaccattataagctgcaataaacaagttaacaacaac
5	U1-140 wt in pRL-SV40	TCAAATCGTTCGTTGAGCGAGTTCTCAAAAATGAA CAATAAttctagacaggtaagtagcggccgcttcgagcagacatgataagata cattgatgagtttggacaaaccacaactagaatgcagtgaaaaaaatgctttatttgtg aaatttgtgatgctattgctttatttgtaaccattataagctgcaataaacaagttaacaac aac
6	U1-140 mut in pRL-SV40	TCAAATCGTTCGTTGAGCGAGTTCTCAAAAATGAA CAATAAttctagacatggnactagcggccgcttcgagcagacatgataagata cattgatgagtttggacaaaccacaactagaatgcagtgaaaaaaatgctttatttgtg aaatttgtgatgctattgctttatttgtaaccattataagctgcaataaacaagttaacaac aac
7	2xU1-140 wt in pRL-SV40	TCAAATCGTTCGTTGAGCGAGTTCTCAAAAATGAA CAATAAttctagacaggtaagtaaccaaacaggtaagtagcggccgcttcgag cagacatgataagatacattgatgagtttggacaaaccacaactagaatgcagtgaa aaaaatgctttatttgtgaaatttgtgatgctattgctttatttgtaaccattataagctgca ataaacaagttaacaacaac
8	2xU1-140 mut in pRL-SV40	TCAAATCGTTCGTTGAGCGAGTTCTCAAAAATGAA CAATAAttctagacatggnactaaccaaacatggaactagcggccgcttcgag cagacatgataagatacattgatgagtttggacaaaccacaactagaatgcagtgaa aaaaatgctttatttgtgaaatttgtgatgctattgctttatttgtaaccattataagctgca ataaacaagttaacaacaac
9	3xU1-140 wt in pRL-SV40	TCAAATCGTTCGTTGAGCGAGTTCTCAAAAATGAA CAATAAttctagacaggtaagtaaccaaacaggtaagtaaccaaacaggtaag tagcggccgcttcgagcagacatgataagatacattgatgagtttggacaaaccaca actagaatgcagtgaaaaaaatgctttatttgtgaaatttgtgatgctattgctttatttgta accattataagctgcAATAAAcaagttaacaacaac
10	3xU1_140 mut in pRL-SV40	TCAAATCGTTCGTTGAGCGAGTTCTCAAAAATGAA CAATAAttctagacatggnactaaccaaacatggaactaaccaaacatggaac tagcggccgcttcgagcagacatgataagatacattgatgagtttggacaaaccaca actagaatgcagtgaaaaaaatgctttatttgtgaaatttgtgatgctattgctttatttgta accattataagctgcAATAAAcaagttaacaacaac
11	87-9 in pRL-SV40	TCAAATCGTTCGTTGAGCGAGTTCTCAAAAATGAA CAATAAttctagagcggccgcttcgagcagacatgataagatacattgatgagt ttggacaaaccaggtaagtcaactagaatgcagtgaaaaaaatgctttatttgtgaaat ttgtgatgctattgctttatttgtaaccattataagctgcaataaacaagttaacaacaac
12	87-8 in pRL-SV40	TCAAATCGTTCGTTGAGCGAGTTCTCAAAAATGAA CAATAAttctagagcggccgcttcgagcagacatgataagatacattgatgagt ttggacaaaccaggtaagcaactagaatgcagtgaaaaaaatgctttatttgtgaaatt tgtgatgctattgctttatttgtaaccattataagctgcaataaacaagttaacaacaac
13	87-7 in pRL-SV40	TCAAATCGTTCGTTGAGCGAGTTCTCAAAAATGAA CAATAAttctagagcggccgcttcgagcagacatgataagatacattgatgagt ttggacaaaccaggtaacaactagaatgcagtgaaaaaaatgctttatttgtgaaattt gtgatgctattgctttatttgtaaccattataagctgcaataaacaagttaacaacaac
14	87-6 in pRL-SV40	TCAAATCGTTCGTTGAGCGAGTTCTCAAAAATGAA CAATAAttctagagcggccgcttcgagcagacatgataagatacattgatgagt ttggacaaaccaggttcaactagaatgcagtgaaaaaaatgctttatttgtgaaatttgt

SEQ ID NO.:	Description	Sequence
		gatgctattgctttatttgtaaccattataagctgcaataaacaagttaacaacaac
15	87-5 in pRL-SV40	TCAAATCGTTCGTTGAGCGAGTTCTCAAAAATGAA CAATAAttctagagcggccgcttcgagcagacatgataagatacattgatgagt ttggacaaaccaaggtcaactagaatgcagtgaaaaaaatgctttatttgtgaaatttgt gatgctattgctttatttgtaaccattataagctgcaataaacaagttaacaacaac
16	U1 87-9 SL in pRL-SV40	TCAAATCGTTCGTTGAGCGAGTTCTCAAAAATGAA CAATAAttctagagcggccgcttcgagcagacatgataagatacattgatgagt ttggacaaaccaaacccaggtaagttaccgaaaggtaacttacctgggcaactagaa tgcagtgaaaaaaatgctttatttgtgaaatttgtgatgctattgctttatttgtaaccatta taagctgcaataaacaagttaacaacaac
17	U1 87-9 SL broken in pRL-SV40	TCAAATCGTTCGTTGAGCGAGTTCTCAAAAATGAA CAATAAttctagagcggccgcttcgagcagacatgataagatacattgatgagt ttggacaaaccaaacccaggtaagttaccgaaaaaaacaaacaaaaaacaactaga atgcagtgaaaaaaatgctttatttgtgaaatttgtgatgctattgctttatttgtaaccatt ataagctgcaataaacaagttaacaacaac
18	U1_theo_1 in pRL-SV40	TCAAATCGTTCGTTGAGCGAGTTCTCAAAAATGAA CAATAAttctagagcggccgcttcgagcagacatgataagatacattgatgagt ttggacaaaccaaacccaggtaagttaccggcgataccagccgaaaggcccttggc agcgtcggtaacttacctgggcaactagaatgcagtgaaaaaaatgctttatttgtgaa atttgtgatgctattgctttatttgtaaccattataagctgcaataaacaagttaacaacaa c
19	U1_theo_2 in pRL-SV40	TCAAATCGTTCGTTGAGCGAGTTCTCAAAAATGAA CAATAAttctagagcggccgcttcgagcagacatgataagatacattgatgagt ttggacaaaccaaacccaggtaagttacggcgataccagccgaaaggcccttggca gcgtcgtaacttacctgggcaactagaatgcagtgaaaaaaatgctttatttgtgaaat ttgtgatgctattgctttatttgtaaccattataagctgcaataaacaagttaacaacaac
20	U1_theo_3 in pRL-SV40	TCAAATCGTTCGTTGAGCGAGTTCTCAAAAATGAA CAATAAttctagagcggccgcttcgagcagacatgataagatacattgatgagt ttggacaaaccaaacccaggtaagttacgcgataccagccgaaaggcccttggcag cgtgtaacttacctgggcaactagaatgcagtgaaaaaaatgctttatttgtgaaatttg tgatgctattgctttatttgtaaccattataagctgcaataaacaagttaacaacaac
21	U1_theo_4 in pRL-SV40	TCAAATCGTTCGTTGAGCGAGTTCTCAAAAATGAA CAATAAttctagagcggccgcttcgagcagacatgataagatacattgatgagt ttggacaaaccaaacccaggtaagttagcgataccagccgaaaggcccttggcagc gttaacttacctgggcaactagaatgcagtgaaaaaaatgctttatttgtgaaatttgtg atgctattgctttatttgtaaccattataagctgcaataaacaagttaacaacaac
22	U1_theo_5 in pRL-SV40	TCAAATCGTTCGTTGAGCGAGTTCTCAAAAATGAA CAATAAttctagagcggccgcttcgagcagacatgataagatacattgatgagt ttggacaaaccaaacccaggtaagttacgataccagccgaaaggcccttggcagcg taacttacctgggcaactagaatgcagtgaaaaaaatgctttatttgtgaaatttgtgat gctattgctttatttgtaaccattataagctgcaataaacaagttaacaacaac
23	U1_theo_6 in pRL-SV40	TCAAATCGTTCGTTGAGCGAGTTCTCAAAAATGAA CAATAAttctagagcggccgcttcgagcagacatgataagatacattgatgagt ttggacaaaccaaacccaggtaagttcgataccagccgaaaggcccttggcagcga acttacctgggcaactagaatgcagtgaaaaaaatgctttatttgtgaaatttgtgatgc

SEQ ID NO.:	Description	Sequence
		tattgctttatttgtaaccattataagctgcaataaacaagttaacaacaac
24	U1_theo_7 in pRL-SV40	TCAAATCGTTCGTTGAGCGAGTTCTCAAAAATGAA CAATAAttctagagcggccgcttcgagcagacatgataagatacattgatgagt ttggacaaaccaaacccaggtaagttgataccagccgaaaggcccttggcagcaac ttacctgggcaactagaatgcagtgaaaaaaatgctttatttgtgaaatttgtgatgctat tgctttatttgtaaccattataagctgcaataaacaagttaacaacaac
25	U1_theo_8 in pRL-SV40	TCAAATCGTTCGTTGAGCGAGTTCTCAAAAATGAA CAATAAttctagagcggccgcttcgagcagacatgataagatacattgatgagt ttggacaaaccaaacccaggtaagtgataccagccgaaaggcccttggcagcactt acctgggcaactagaatgcagtgaaaaaaatgctttatttgtgaaatttgtgatgctatt gctttatttgtaaccattataagctgcaataaacaagttaacaacaac
26	U1_theo_9 in pRL-SV40	TCAAATCGTTCGTTGAGCGAGTTCTCAAAAATGAA CAATAAttctagagcggccgcttcgagcagacatgataagatacattgatgagt ttggacaaaccaaacccaggtaagtataccagccgaaaggcccttggcagacttac ctgggcaactagaatgcagtgaaaaaaatgctttatttgtgaaatttgtgatgctattgc tttatttgtaaccattataagctgcaataaacaagttaacaacaac
27	U1_theo_8_1 in pRL-SV40	TCAAATCGTTCGTTGAGCGAGTTCTCAAAAATGAA CAATAAttctagagcggccgcttcgagcagacatgataagatacattgatgagt ttggacaaaccaaacccaggtaagtgataccagccgaaaggcccttggcagcactt acctggacaactagaatgcagtgaaaaaaatgctttatttgtgaaatttgtgatgctatt gctttatttgtaaccattataagctgcaataaacaagttaacaacaac
28	U1_theo_8_2 in pRL-SV40	TCAAATCGTTCGTTGAGCGAGTTCTCAAAAATGAA CAATAAttctagagcggccgcttcgagcagacatgataagatacattgatgagt ttggacaaaccaaacccaggtaagtgataccagccgaaaggcccttggcagcactt acctgaacaactagaatgcagtgaaaaaaatgctttatttgtgaaatttgtgatgctatt gctttatttgtaaccattataagctgcaataaacaagttaacaacaac
29	U1_theo_8_3 in pRL-SV40	TCAAATCGTTCGTTGAGCGAGTTCTCAAAAATGAA CAATAAttctagagcggccgcttcgagcagacatgataagatacattgatgagt ttggacaaaccaaacccaggtaagtgataccagccgaaaggcccttggcagcactt acctaaacaactagaatgcagtgaaaaaaatgctttatttgtgaaatttgtgatgctatt gctttatttgtaaccattataagctgcaataaacaagttaacaacaac
30	U1_theo_8_4 in pRL-SV40	TCAAATCGTTCGTTGAGCGAGTTCTCAAAAATGAA CAATAAttctagagcggccgcttcgagcagacatgataagatacattgatgagt ttggacaaaccaaacccaggtaagtgataccagccgaaaggcccttggcagcactt accaaaaacaactagaatgcagtgaaaaaaatgctttatttgtgaaatttgtgatgctatt gctttatttgtaaccattataagctgcaataaacaagttaacaacaac
31	U1_theo_8_5 in pRL-SV40	TCAAATCGTTCGTTGAGCGAGTTCTCAAAAATGAA CAATAAttctagagcggccgcttcgagcagacatgataagatacattgatgagt ttggacaaaccaaacccaggtaagtgataccagccgaaaggcccttggcagcactt acaaaaacaactagaatgcagtgaaaaaaatgctttatttgtgaaatttgtgatgctatt gctttatttgtaaccattataagctgcaataaacaagttaacaacaac
32	U1_theo_8_6 in pRL-SV40	TCAAATCGTTCGTTGAGCGAGTTCTCAAAAATGAA CAATAAttctagagcggccgcttcgagcagacatgataagatacattgatgagt ttggacaaaccaaacccaggtaagtgataccagccgaaaggcccttggcagcactt aaaaaaacaactagaatgcagtgaaaaaaatgctttatttgtgaaatttgtgatgctatt gctttatttgtaaccattataagctgcaataaacaagttaacaacaac
33	U1_theo_8_7 in pRL-SV40	TCAAATCGTTCGTTGAGCGAGTTCTCAAAAATGAA CAATAAttctagagcggccgcttcgagcagacatgataagatacattgatgagt

SEQ ID NO.:	Description	Sequence
		ttggacaaaccaaacccaggtaagtgataccagccgaaaggcccttggcagcactt caaaaaacaactagaatgcagtgaaaaaaatgctttatttgtgaaatttgtgatgctatt gctttatttgtaaccattataagctgcaataaacaagttaacaacaac
34	U1_theo_8_8 in pRL-SV40	TCAAATCGTTCGTTGAGCGAGTTCTCAAAAATGAA CAATAAttctagagcggccgcttcgagcagacatgataagatacattgatgagt ttggacaaaccaaacccaggtaagtgataccagccgaaaggcccttggcagcacta caaaaaacaactagaatgcagtgaaaaaaatgctttatttgtgaaatttgtgatgctatt gctttatttgtaaccattataagctgcaataaacaagttaacaacaac
35	U1_theo_8_9 in pRL-SV40	TCAAATCGTTCGTTGAGCGAGTTCTCAAAAATGAA CAATAAttctagagcggccgcttcgagcagacatgataagatacattgatgagt ttggacaaaccaaacccaggtaagtgataccagccgaaaggcccttggcagcaca acaaaaaacaactagaatgcagtgaaaaaaatgctttatttgtgaaatttgtgatgctat tgctttatttgtaaccattataagctgcaataaacaagttaacaacaac
36	U1_theo_9_1 in pRL-SV40	TCAAATCGTTCGTTGAGCGAGTTCTCAAAAATGAA CAATAAttctagagcggccgcttcgagcagacatgataagatacattgatgagt ttggacaaaccaaacccaggtaagtataccagccgaaaggcccttggcagacttac ctggacaactagaatgcagtgaaaaaaatgctttatttgtgaaatttgtgatgctattgct ttatttgtaaccattataagctgcaataaacaagttaacaacaac
37	U1_theo_9_2 in pRL-SV40	TCAAATCGTTCGTTGAGCGAGTTCTCAAAAATGAA CAATAAttctagagcggccgcttcgagcagacatgataagatacattgatgagt ttggacaaaccaaacccaggtaagtataccagccgaaaggcccttggcagacttac ctgaacaactagaatgcagtgaaaaaaatgctttatttgtgaaatttgtgatgctattgct ttatttgtaaccattataagctgcaataaacaagttaacaacaac
38	U1_theo_9_3 in pRL-SV40	TCAAATCGTTCGTTGAGCGAGTTCTCAAAAATGAA CAATAAttctagagcggccgcttcgagcagacatgataagatacattgatgagt ttggacaaaccaaacccaggtaagtataccagccgaaaggcccttggcagacttac ctaaacaactagaatgcagtgaaaaaaatgctttatttgtgaaatttgtgatgctattgct ttatttgtaaccattataagctgcaataaacaagttaacaacaac
39	U1_theo_9_4 in pRL-SV40	TCAAATCGTTCGTTGAGCGAGTTCTCAAAAATGAA CAATAAttctagagcggccgcttcgagcagacatgataagatacattgatgagt ttggacaaaccaaacccaggtaagtataccagccgaaaggcccttggcagacttac caaaacaactagaatgcagtgaaaaaaatgctttatttgtgaaatttgtgatgctattgc tttatttgtaaccattataagctgcaataaacaagttaacaacaac
40	U1_theo_9_5 in pRL-SV40	TCAAATCGTTCGTTGAGCGAGTTCTCAAAAATGAA CAATAAttctagagcggccgcttcgagcagacatgataagatacattgatgagt ttggacaaaccaaacccaggtaagtataccagccgaaaggcccttggcagacttac aaaaacaactagaatgcagtgaaaaaaatgctttatttgtgaaatttgtgatgctattgc tttatttgtaaccattataagctgcaataaacaagttaacaacaac
41	U1_theo_9_6 in pRL-SV40	TCAAATCGTTCGTTGAGCGAGTTCTCAAAAATGAA CAATAAttctagagcggccgcttcgagcagacatgataagatacattgatgagt ttggacaaaccaaacccaggtaagtataccagccgaaaggcccttggcagacttaa aaaaacaactagaatgcagtgaaaaaaatgctttatttgtgaaatttgtgatgctattgc tttatttgtaaccattataagctgcaataaacaagttaacaacaac
42	U1_theo_9_7 in	TCAAATCGTTCGTTGAGCGAGTTCTCAAAAATGAA

SEQ ID NO.:	Description	Sequence
	pRL-SV40	CAATAAttctagagcggccgcttcgagcagacatgataagatacattgatgagt ttggacaaaccaaacccaggtaagtataccagccgaaaggcccttggcagacttca aaaaacaactagaatgcagtgaaaaaaatgctttatttgtgaaatttgtgatgctattgc tttatttgtaaccattataagctgcaataaacaagttaacaacaac
43	U1_theo_9_8 in pRL-SV40	TCAAATCGTTCGTTGAGCGAGTTCTCAAAAATGAA CAATAAttctagagcggccgcttcgagcagacatgataagatacattgatgagt ttggacaaaccaaacccaggtaagtataccagccgaaaggcccttggcagactaca aaaaacaactagaatgcagtgaaaaaaatgctttatttgtgaaatttgtgatgctattgc tttatttgtaaccattataagctgcaataaacaagttaacaacaac
44	4xU1_theo_8_1 in pRL-SV40	TCAAATCGTTCGTTGAGCGAGTTCTCAAAAATGAA CAATAAttctagagcggccgcttcgagcagacatgataagatacattgatgagt ttggacaaaccaaaccaggtaagtgataccagccgaaaggcccttggcagcactt acctggacaactagaatgcagtgaaaaaagagtttggacaaaccaaaccaggtaa gtgataccagccgaaaggcccttggcagcacttacctggacaactagaatgcagtg aaaaaaatgcgagtttggacaaaccaaaccaggtaagtgataccagccgaaagg cccttggcagcacttacctggacaactagaatgcagtgaaaaaaatgctttatttgtga gagtttggacaaaccaaaccaggtaagtgataccagccgaaaggcccttggcag cacttacctggacaactagaatgcagtgaaaaaaatgctttatttgtgaaatttgtgatg ctattgctttatttgtaaccattataagctgcaataaacaagttaacaacaac
45	U1_Gua_1 wt in pFLuc	TCATAAAGGCCAAGAAGGGCGGAAAGATCGCCGT GTAAgccataccacattaaggtaatgtataatcgcgtggatatggcacgcaagttt ctaccgggcaccgtaaatgtccgactacattactgtagaggttttacttgctttaaaaaa cctcccacacctccccctgaacctgaaacataaaatgaatgcaattgttgttgttaactt gtttattgcagcttataatggttacaaataaagcaatagcatcacaaatttcacaaataa agcatttttttcactgcattctagttgtggtttgtccaaactcatcaatgtatcttaa
46	U1_Gua_1 mut in pFLuc	TCATAAAGGCCAAGAAGGGCGGAAAGATCGCCGT GTAAgccataccacattaatggaatctataatcgcgtggatatggcacgcaagttt ctaccgggcaccgtaaatgtccgactagattcctgtagaggttttacttgctttaaaaaa cctcccacacctccccctgaacctgaaacataaaatgaatgcaattgttgttgttaactt gtttattgcagcttataatggttacaaataaagcaatagcatcacaaatttcacaaataa agcatttttttcactgcattctagttgtggtttgtccaaactcatcaatgtatcttaa
47	U1_Gua_2 wt in pFLuc	TCATAAAGGCCAAGAAGGGCGGAAAGATCGCCGT GTAAgccataccacatttgtagaggttttacttgctttaaaaaacctcccacaccta aggtaatgtataatcgcgtggatatggcacgcaagtttctaccgggcaccgtaaatgt ccgactacattacccccctgaacctgaaacataaaatgaatgcaattgttgttgttaact tgtttattgcagcttataatggttacaaataaagcaatagcatcacaaatttcacaaataa agcatttttttcactgcattctagttgtggtttgtccaaactcatcaatgtatcttaa
48	U1_Gua_2 mut in pFLuc	TCATAAAGGCCAAGAAGGGCGGAAAGATCGCCGT GTAAgccataccacatttgtagaggttttacttgctttaaaaaacctcccacaccta atggaatctataatcgcgtggatatggcacgcaagtttctaccgggcaccgtaaatgt ccgactagattcccccccctgaacctgaaacataaaatgaatgcaattgttgttgttaact tgtttattgcagcttataatggttacaaataaagcaatagcatcacaaatttcacaaataa agcatttttttcactgcattctagttgtggtttgtccaaactcatcaatgtatcttaa
49	U1_Gua_3 wt in pFLuc	TCATAAAGGCCAAGAAGGGCGGAAAGATCGCCGT GTAAgccataccacatttgtagaggttttacttgctttaaaaaacctcccacacctc cccctgaacctgaaacataaaatgaatgcaattgtaaggtaatgtataatcgcgtgga

SEQ ID NO.:	Description	Sequence
		tatggcacgcaagtttctaccgggcaccgtaaatgtccgactacattactgttgttaact tgtttattgcagcttataatggttacaaataaagcaatagcatcacaaatttcacaaataa agcatttttttcactgcattctagttgtggtttgtccaaactcatcaatgtatcttaa
50	U1_Gua_3 mut in pFLuc	TCATAAAGGCCAAGAAGGGCGGAAAGATCGCCGT GTAAgccataccacatttgtagaggttttacttgctttaaaaaacctcccacacctc cccctgaacctgaaacataaaatgaatgcaattgtaatggaatctataatcgcgtggat atggcacgcaagtttctaccgggcaccgtaaatgtccgactagattcctgttgttaact tgtttattgcagcttataatggttacaaataaagcaatagcatcacaaatttcacaaataa agcatttttttcactgcattctagttgtggtttgtccaaactcatcaatgtatcttaa
51	wtU1_Gua_HBGP A	TCATAAAGGCCAAGAAGGGCGGAAAGATCGCCGT GTAAggatccaagcttatcgataccgtcgacctcgaggtaatgtataatcgcgtg gatatggcacgcaagtttctaccgggcaccgtaaatgtccgactacattacggccca gatctaattcaccccaccagtgcaggctgcctatcagaaagtggtggctggtgtggc taatgccctggcccacaagtatcactaagctcgctttcttgctgtccaatttctattaaag gttcctttgttccctaagtccaactactaaactgggggatattatgaagggccttgagc atctggattctgcctaataaaaaacatttatttcattgcaatgatgtattt
52	mutU1_Gua_HBG PA	TCATAAAGGCCAAGAAGGGCGGAAAGATCGCCGT GTAAggatccaagcttatcgataccgtcgacctcgatggaatctataatcgcgtg gatatggcacgcaagtttctaccgggcaccgtaaatgtccgactacattacggccca gatctaattcaccccaccagtgcaggctgcctatcagaaagtggtggctggtgtggc taatgccctggcccacaagtatcactaagctcgctttcttgctgtccaatttctattaaag gttcctttgttccctaagtccaactactaaactgggggatattatgaagggccttgagc atctggattctgcctaataaaaaacatttatttcattgcaatgatgtattt

[0103] It is to be understood and expected that variations in the principles of invention herein disclosed can be made by one skilled in the art and it is intended that such modifications are to be included within the scope of the present invention. The following Examples further illustrate the invention, but should not be construed to limit the scope of the invention in any way. All references cited herein are hereby incorporated by reference in their entirety.

EXAMPLE 1

[0104] Effects on target gene expression of location and copy number of U1 binding site in 3′ UTR

[0105] Experimental procedure:

[0106] Plasmid constructs: oligonucleotides (oligos) containing either wild type consensus U1 binding sequence or its mutant sequence were synthesized (IDT). The synthesized oligos were cloned into the 3′ UTR of the pRL-SV40 luciferase expression vector

(Promega) using Gibson Assembly Cloning Kit (NEB). Constructs were purified (Qiagen) and verified by DNA sequencing (Genewiz).

[0107] Transfection: 3.5 x10^4 HEK 293 cells were plated in a 96-well flat bottom plate the day before transfection. Plasmid DNA (500 ng) was added to a tube or a 96-well U-bottom plate. Separately, TransIT-293 reagent (Mirus; 1.4 µl) was added to 50 µl Opti-mem I media (Life Technologies), and allowed to sit for 5 minutes at room temperature ("RT"). Then, 50 µl of this diluted transfection reagent was added to the DNA, mixed, and incubated at RT for 20 min. Finally, 7 µl of this solution was added to a well of cells in the 96-well plate.

[0108] Renilla luciferase assay of cultured cells: 24 hours after the media change, plates were removed from the incubator, and equilibrated to RT for several minutes on a lab bench, and the media aspirated. Glo-lysis buffer (Promega, 100 µl, RT) was added, and the plates allowed to remain at room temp for at least 5 minutes. Then, the well contents were mixed by 50 µl trituration, and 20 µl of each well was mixed with 20 µl of Renilla-glo reagent in a solid-white 384-well plate. Ten minutes later, luminescence was measured using a Tecan machine with 500 msec read time. The luciferase activity was expressed as mean relative light unit (RLU) ±SD.

[0109] Results:

[0110] In order to build a U1-polyadenylation-based gene regulation platform, a U1 binding sequence was inserted into the 3′ UTR of a target reporter gene. The 10 nucleotide ("nt") consensus U1 binding sequence (referred to herein as wildtype or wt) was placed either 87 nt (U1-87) or 140 nt (U1-140) upstream of the AATAAA poly(A) sequence in the 3 ′UTR in of the luciferase gene in the pRL-SV40 vector, as shown in Figure 1a. As shown in Figure 1b, insertion of 10-nt mutant U1 binding sequence at -87 position (U1-87 mut) did not have any suppressive effect on luciferase gene expression compared to expression from the unmodified pRL-SV40 construct. However, insertion of a 10-nt wild type U1 binding sequence at the -87 position (U1-87 wt) inhibited 98% of Renilla luciferase expression when compared to luciferase activity from the pRL-SV40 control vector. When compared with the U1-87 mut construct, the U1-87 wt generated 50-fold decrease in luciferase activity. Similarly, the mutant U1 site at -140 (U1-140 mut) did not cause inhibition of luciferase gene expression. However, the wild type U1 binding sequence at the -140 location (U1-140 wt)

resulted in 91% inhibition of luciferase activity, generating 12-fold inhibition of luciferase activity when compared to the U1-140 mut.

[0111] The inhibitory effect of either 2 copies or 3 copies of U1 binding sequence inserted at -140 in 3′ UTR was also tested. As shown in Figure 1b, 2 copies (2xU1-140 wt) or 3 copies (3xU1-140 wt) of the wild type U1 binding sequence resulted in further reduction in luciferase activity when compared to single copy of U1 site at the same location (U1-140 wt), generating 272-fold and 462-fold inhibition, respectively, when compared to their corresponding mutant constructs. Thus, these results indicate that multiple copies of U1 binding sequence have synergistic function in suppressing gene expression.

EXAMPLE 2

[0112] Effect on target gene expression of the length of U1 binding sequence in the 3′ UTR.

[0113] Experimental procedures: as described in Example 1.

[0114] Results:

[0115] Based on the results described in Example 1, the suppressive effect of a U1 binding site inserted 87 nt (U1-87) upstream of the SV40 late polyA signal sequence was further characterized. In order to determine the minimal length of the U1 binding sequence for repressing target gene expression when placed in the 3′, a series of constructs containing 3′ sequentially-truncated sequence of U1 site, as shown in Figure 2a. As shown in Figure 2b, 9-nt U1 site (87-9) functioned as well as the 10-nt U1 site (U1-87). However, when the U1 binding sequence was truncated to 8 nt, the suppressive effect of the U1 site was somewhat reduced, resulting in 3-fold less efficiency. When the U1 site sequence was 7 nt or shorter, the suppressive effect was significantly less efficient. These results indicate that U1 binding site inserted 87 nt upstream of SV40 late poly(A) signal in 3' UTR is more effective at repressing target gene expression when at least 8 nt long in order to recruit U1 snRNP efficiently to suppress polyadenylation. A 9 nt U1 binding sequence inserted 87 nt upstream of SV40 late polyA signal sequence was selected for further characterization.

EXAMPLE 3

[0116] The effect of stem-loop secondary structure formation at the U1 site in 3′ UTR on gene expression

[0117] Experimental Procedure:

[0118] The stem-loop sequences containing a wild type 9 nt U1 binding sequence were synthesized (IDT) and cloned into pRL-SV40 vector. The construct sequences were verified by DNA sequencing (Genewiz). HEK 293 cells were transfected and Renilla luciferase assay were performed as described in Example 1.

[0119] Results:

[0120] To determine the effect of a hairpin or stem-loop structure on the U1 site-mediated suppression of gene expression, the 9 nt U1 site was embedded in a stem-loop (SL) structure. In this stem-loop structure, as shown in Figure 3a, the U1 site and its complementary sequence forms the stem. A mutant sequence was made as control in which no stem is formed (SL broken). As shown in Figure 3b, embedding the 9 nt U1 site in the stem-loop structure (U1 87-9 SL) completely abolished the suppressive effect of the U1 site in the 3' UTR, whereas the control sequence that can't form stem-loop structure (U1 87-9 SL broken) showed little effect on the suppressive function of U1 site. These results support that a U1 site embedded in stem-loop structure is not accessible to U1 snRNP binding, therefore gene expression is not suppressed by U1 snRNP-mediated suppression of polyadenylation.

EXAMPLE 4

[0121] Use of theophylline aptamer to regulate target gene expression via modulating U1 interference with polyadenylation

[0122] Experimental procedures:

[0123] Oligonucleotides containing either the full length of sequences of the stem from the stem-loop structure and theophylline aptamer or the truncated sequences of the stem, were synthesized (IDT) and cloned into pRL-SV40 vector using the Gibson Assembly Cloning Kit (NEB). Construct sequences were verified by DNA sequencing (Genewiz).

[0124] Transfection and aptamer ligand treatment: HEK 293 cells were transfected as described in Example 1. Four hours after transfection, the media was aspirated, and new media with or without 3 mM theophylline was added. A Renilla luciferase assay was performed 20 to 24 hours after theophylline treatment as described in Example 1. The induction fold was expressed as the quotient of luciferase activity obtained in the presence of aptamer ligand divided by the value obtained in the absence of the aptamer ligand.

[0125] Results:

[0126] In order to regulate the accessibility of a U1 site in the 3' UTR of a target gene, and thereby regulate target gene expression, the stem sequence of the aptamer structure was linked directly to the stem of the stem-loop structure tested in Example 3. In this configuration, as illustrated in Figure 4a, insertion of aptamer sequence disrupts the formation of the stem structure, therefore, the U1 site in the 3' UTR is accessible to U1 snRNP binding. However, in the presence of aptamer ligand, aptamer/ligand binding causes a RNA structure conformational change, facilitating stem formation and sequestering U1 site from U1 snRNP binding in 3' UTR.

[0127] A synthetic theophylline aptamer was tested in this configuration by linking the lower stem sequence of the theophylline aptamer structure to the stem of the stem-loop structure, which generated a 19 base pair stem as shown in Figure 4b. We rationalized that if the stem is too long, the stem formation could be independent of the existence of aptamer sequence and occur in the absence of aptamer ligand. Conversely, if the stem is too short, even in the presence of aptamer ligand, a stable aptamer structure cannot be achieved. Therefore, to determine an optimal length of the stem (including any stem from the aptamer and the stem containing U1 binding site), serial truncations to the stem were made, generating 9 constructs with the length of the stem ranging from 19 to 11 bp. As shown in Figure 4c, with constructs U1_theo_1 through U1_theo_6, the luciferase activity was not suppressed significantly comparing to construct U1 87-9 in the absence of theophylline, suggesting the sequestering of the U1 site due to stem formation. Further, there was no increase in the luciferase activity after theophylline treatment. However, in construct 8 and 9, the luciferase activity was increased after theophylline treatment, suggesting aptamer/ligand binding further blocks U1 site accessibility, leading to increased target gene expression.

EXAMPLE 5

[0128] Optimization of the stem sequence to enhance the regulatability of U1 accessibility due to aptamer/ligand binding

[0129] Experimental procedures: As described in Example 4.

[0130] Results:

[0131] Constructs U1_theo_8 and 9 showed small increase in luciferase activity after theophylline treatment. To further enhance the regulatability of the U1 site accessibility when the aptamer is not bound by ligand, the stem structure was weakened by sequentially mutating the stem bases, as shown in Figure 5a. Though, this strategy did not improve the

induction of luciferase activity from construct U1_theo_9, as shown in Figure 5c, it improved regulatability of construct U1_theo_8. As shown in Figure 5b, in a series of 9 constructs made from U1_theo_8, constructs U1_theo_8_1 through 5 generated increased luciferase activity after theophylline treatment when compared to luciferase activity in the absence of aptamer ligand. The induction fold was increased when comparing with the induction fold of U1_theo_8 construct (1.3 fold), ranging from 1.8 to 2.6 fold. Therefore, we have built a synthetic riboswitch that regulates U1 interference with polyadenylation through aptamer/ligand binding.

EXAMPLE 6

[0132] Effects of multiple U1_theo riboswitch on regulating target gene expression

[0133] Experimental procedures:

[0134] Oligos containing 4 copies of the U1_theo_8_1 sequence in tandem were synthesized (IDT) and cloned using Gibson strategy and cloning kit (NEB). HEK 293 cells were transfected and Renilla luciferase assay was performed as described in Example 4.

[0135] Results:

[0136] Construct U1_theo_8-1 generated 1.8-fold induction after theophylline treatment, due to relatively high basal level expression of luciferase gene. In order to lower the basal level expression of the target gene, 4 copies of the U1_theo_8_1 sequence was placed in tandem in the 3' UTR, as shown in Figure 6a. Indeed, as shown in Figure 6b, 4 copies of U1-theo_8_1 did reduce the basal level expression by 83% comparing to the single copy of U1_theo_8_1. Upon theophylline treatment, the 4-copy construct increased the luciferase activity in a dose-dependent manner, generating 40% of the luciferase activity of 87-9 SL control vector and maximal 4.3-fold induction at 6 mM theophylline. These results indicate that multiple copies of the U1 site riboswitch in tandem can reduce basal expression level of the target gene, improving target gene regulatability. In the current construct configuration, the same aptamer sequence was used in all 4 copies of regulation cassettes. Different aptamer sequences can be used in each copy of the riboswitch.

EXAMPLE 7

[0137] Use of xpt-guanine aptamer to regulate target gene expression via modulating U1 interference

[0138] Experimental procedures:

[0139] Plasmid constructs: The EGFP gene in pEGFP-C1 vector was replaced with firefly luciferase gene coding sequence to generate a pFLuc vector that contains an SV40 early polyadenylation signal sequence. Oligos containing either wild type or mutant U1 binding site from the 5' splice site of human DHFR exon 2 and xpt-guanine aptamer sequence followed by the complementary sequence of the last 7 nt of the U1 site sequence were synthesized (IDT) and cloned into the pFLuc vector using the Gibson cloning strategy and kit (NEB).

[0140] Transfection and aptamer ligand treatment: HEK 293 cells were transfected as described in Example 1. Four hours after transfection, the media was aspirated, and new media with or without 500 μM guanine was added.

[0141] Firefly luciferase assay of cultured cells: Twenty-four hours after the media change, the plates were removed from the incubator, and equilibrated to RT for several minutes on a lab bench, then aspirated. Glo-lysis buffer (Promega, 100 uL, RT) was added, and the plates allowed to remain at RT for at least 5 minutes. Then, the well contents were mixed by 50 uL trituration, and 20 uL of each sample was mixed with 20 μL of bright-glo reagent (Promega) that had been diluted to 10% in glo-lysis buffer. The 96 wells were spaced on an opaque white 384-well plate. Following a 5 min incubation at RT, luminescence was measured using a Tecan machine with 500 mSec read time. The luciferase activity was expressed as mean relative light unit (RLU)±S.D.

[0142] Results:

[0143] The use of additional aptamer/ligand pair, either synthetic or natural aptamers, in this U1-interference based, gene regulation platform was tested by inserting a 10 nt U1 site, the 5' splice site from DHFR exon 2, linked with xpt-guanine aptamer at different positions in pRLuc vector, as shown in Figure 7a. In this 10 nt U1 site, only the last 7 nt were designed to base pair with its complementary sequence located at the 3' end of xpt-guanine aptamer sequence, as shown in Figure 7b. This sequence configuration was generated through a serial truncation of the stem sequence connecting the aptamer P1 stem and stem formed by the U1 site and its complementary sequence, and demonstrated high dynamic range in alternative splicing based aptamer riboswitch. In this study, we placed this sequence configuration at either 149 nt, 110 nt or 74 nt upstream of the SV40 early poly(A) signal sequence in 3' UTR region, generating U1_Guan_1, 2 and 3 constructs. In addition, a

mutant U1 site was generated as a control sequence. As shown in Figure 7c, upon guanine treatment, the luciferase activity increased 2.4, 2.0 and 1.7 fold from U1_Gua_1, 2, and 3 constructs, respectively.

EXAMPLE 8

Aptamer-modulated U1 interference is not polyA sequence specific

[0144] Experimental procedures:

[0145] Plasmid constructs: SV40 early polyadenylation sequence was replaced with human beta globin polyA sequence in pFLuc vector to generate pFLuc_HBGPA. Oligos containing either wild type or mutant U1 binding site from 5' ss of human DHFR exon 2 and xpt-guanine aptamer sequence followed by the complementary sequence of last 7 nt of U1 site sequence were synthesized (IDT) and cloned into pFLuc_HBGPA vector using Gibson cloning strategy and kit (NEB).

[0146] Transfection and luciferase assay were as described in Example 7.

[0147] Results:

[0148] We have demonstrated aptamer-modulated U1 interference with polyadenylation in the context of either SV40 late polyA sequence or SV40 early polyA sequence, as shown in Example 4-6. Further, to demonstrate the aptamer-modulated U1 interference, thus target gene expression was not limited to SV40 polyA sequences, we tested the guanine aptamer-modulated U1 interference in the context of the polyA sequence from human beta globin gene. As shown in Figure 8, upon guanine treatment, the luciferase activity increased 2.1 fold from wtU1_Gua_HBGPA construct. In contrast, the mutU1_Gua_HBGPA didn't induce luciferase activity when comparing to the untreated samples. These results demonstrate that aptamer-modulated U1 interference is not polyA sequence specific.

We Claim:

1. A polynucleotide cassette for the regulation of the expression of a target gene comprising a riboswitch wherein the riboswitch comprises an effector region and an aptamer, wherein the effector region comprises a U1 snRNP binding site and sequence complementary to the U1 snRNP binding site.
2. The polynucleotide cassette of claim 1, wherein the aptamer binds a small molecule ligand.
3. The polynucleotide cassette of claim 1 or claim 2, wherein the effector sequence comprises additional sequence that is capable of forming a stem when the aptamer binds ligand.
4. The polynucleotide cassette of claim 3, wherein the effector region comprises a stem-forming sequence that is 9 to 11 base pairs.
5. The polynucleotide cassette of claim 4, wherein the effector region comprises a stem-forming sequence with one or more mismatched bases in the stem.
6. The polynucleotide cassette according to any of claims 1 to 5, wherein the U1 snRNP binding site is 8 to 10 nucleotides.
7. The polynucleotide cassette of claims according to any 1 to 5, wherein the U1 snRNP binding site comprises the sequence CAGGTAAG.
8. The polynucleotide cassette according to any of claims 1 to 5, wherein the U1 snRNP binding site is selected from the group consisting of CAGGTAAGTA, CAGGTAAGT, and CAGGTAAG.
9. A polynucleotide cassette of claim 1, wherein the polynucleotide cassette comprises two or more riboswitches, wherein each riboswitch comprises an effector region and an aptamer, wherein the effector region comprises a U1 snRNP binding site and sequence complementary to the U1 snRNP binding site.
10. The polynucleotide cassette of claim 9, wherein the two or more riboswitches each comprise an aptamer that binds the same ligand.
11. The polynucleotide cassette of claim 9, wherein the two or more riboswitches comprise different aptamers that bind different ligands.
12. A method of modulating the expression of a target gene comprising
    a. inserting one or more of the polynucleotide cassettes of claims 1-11 into the 3′ UTR of a target gene,
    b. introducing the target gene comprising the polynucleotide cassette into a cell, and
    c. exposing the cell to a ligand that specifically binds the aptamer in an amount

effective to increase expression of the target gene.

13. The method of claim 12, wherein the ligand is a small molecule.
14. The method of claim 12 or claim 13, wherein the polynucleotide cassette is inserted about 87 or about 140 nucleotides 5' of the polyadenylation signal.
15. The method of claim 12 or claim 13, wherein the polynucleotide cassette is inserted about 74, about 110, or about 149 nucleotides 5' of the polyadenylation signal.
16. The method of claim 12 or claim 13, wherein two or more of polynucleotide cassettes are inserted into the 3′ UTR of the target gene.
17. The method of claim 14, wherein the two or more polynucleotide cassettes comprise different aptamers that specifically bind to different small molecule ligands.
18. The method of claim 14, wherein the two or more polynucleotide cassettes comprise the same aptamer.
19. The method according to any of claims 12 to 16, wherein the two or more polynucleotide cassettes are inserted at different locations of the 3′ UTR of the target gene.
20. The method according to any of claims 12 to 19, wherein the target gene comprising the polynucleotide cassette is incorporated in a vector for the expression of the target gene.
21. The method of claim 20, wherein the vector is a viral vector.
22. The method of claim 21, wherein the viral vector is selected from the group consisting of adenoviral vector, adeno-associated virus vector, and lentiviral vector.
23. A vector comprising a target gene that contains a polynucleotide cassette of claims 1-11.
24. The vector of claim 23, wherein the vector is a viral vector.
25. The vector of claim 24, wherein the viral vector is selected from the group consisting of adenoviral vector, adeno-associated virus vector, and lentiviral vector.

Figure 1a

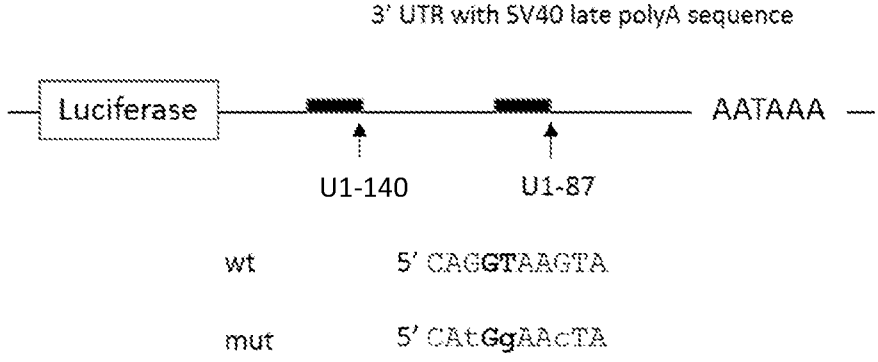
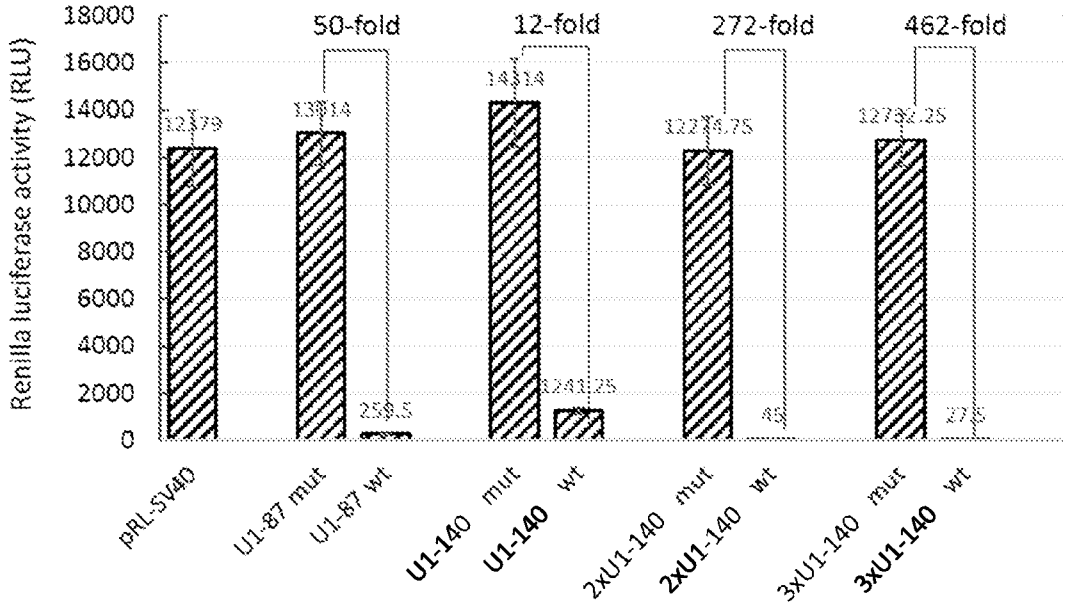


Figure 1b



Construct	Sequence	Length (nt)
U1-87	5' CAGGTAAGTA 3'	10
87-9	5' CAGGTAAGT 3'	9
87-8	5' CAGGTAAG 3'	8
87-7	5' CAGGTAA 3'	7
87-6	5' CAGGTA 3'	6
87-5	5' CAGGT 3'	5

Progressive deletion

FIG. 2a

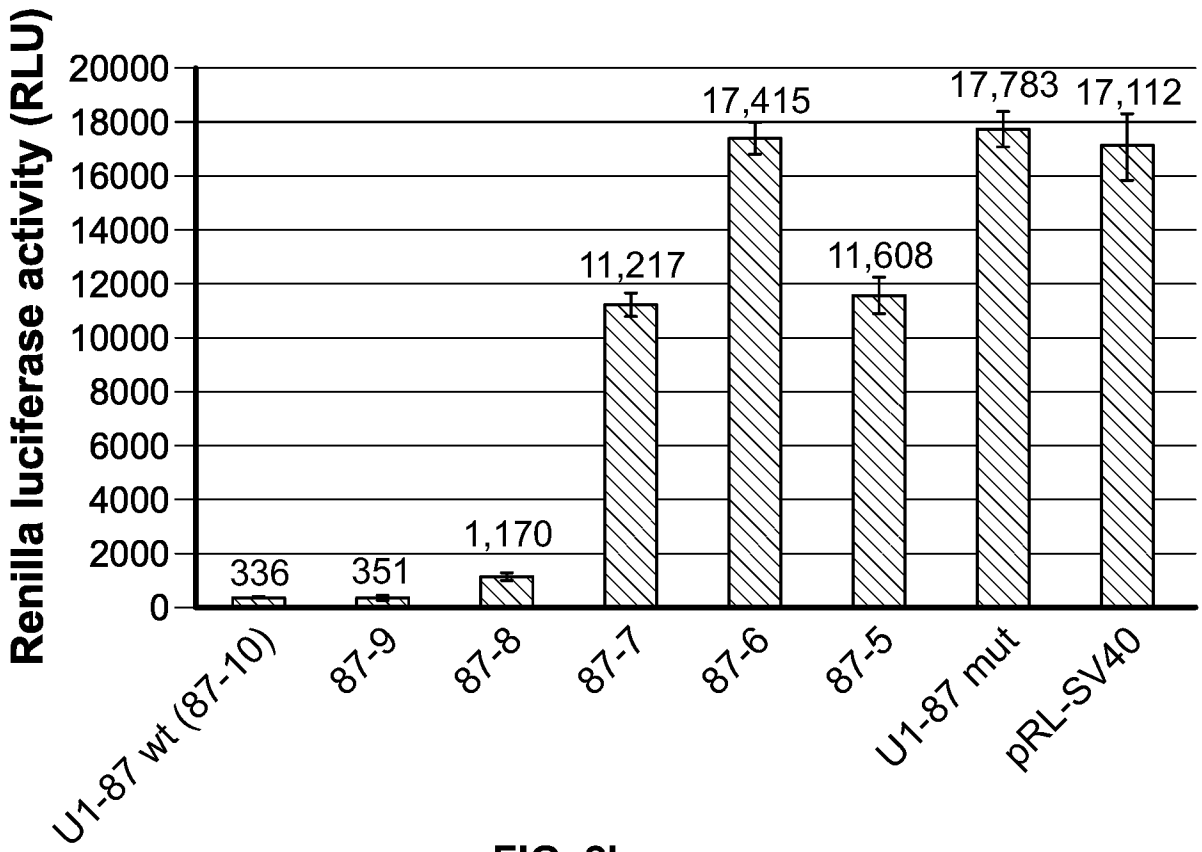
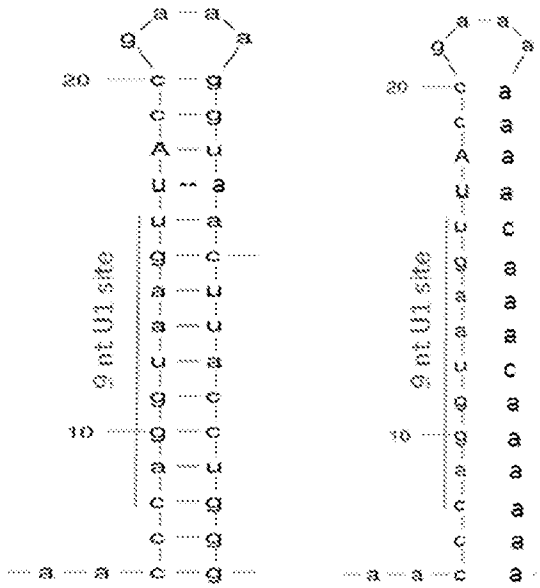


FIG. 2b

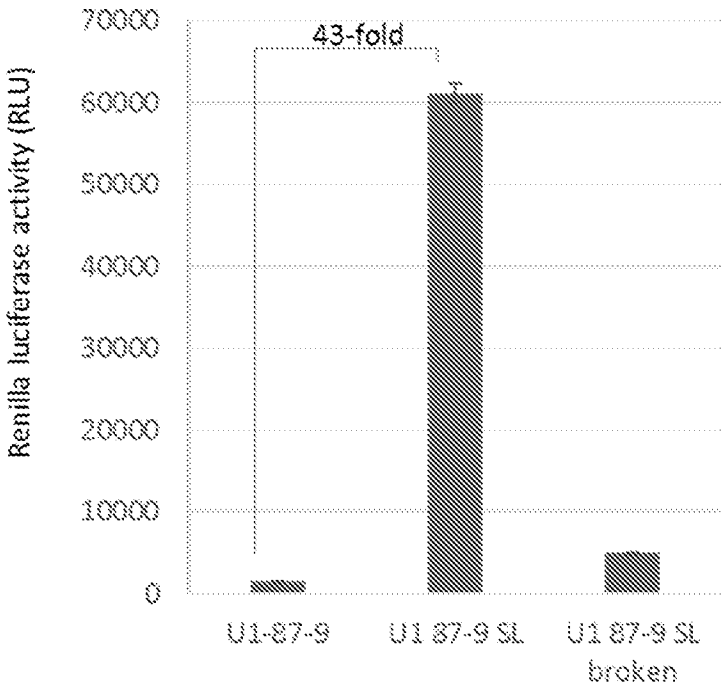
Figure 3a

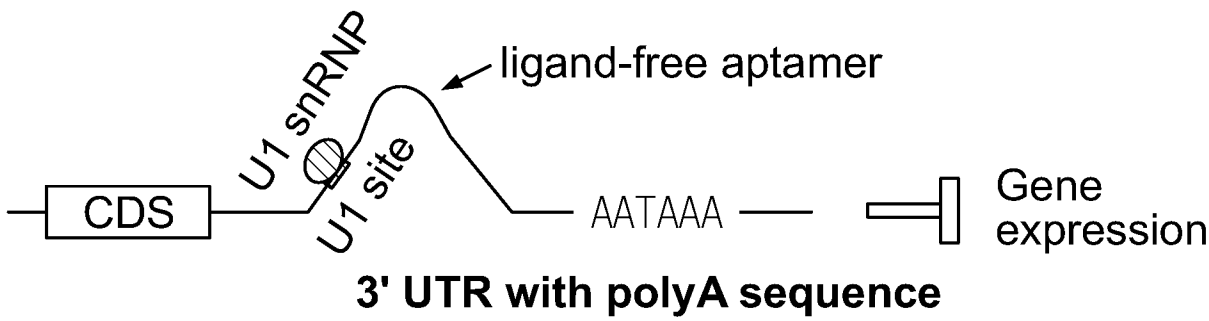


single U1 site 87-9 in stem loop

single U1 site 87-9 with broken stem

Figure 3b




U1 snRNP
ligand-free aptamer
U1 site
CDS
AATAAA
Gene expression
3' UTR with polyA sequence

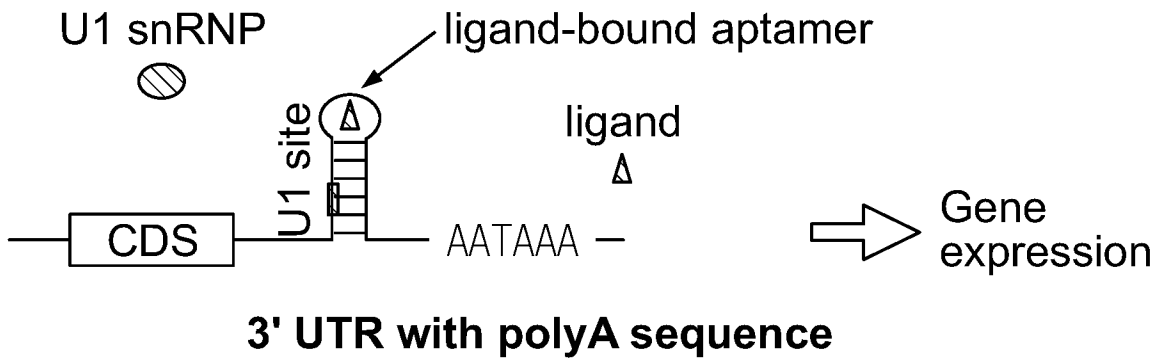

U1 snRNP
ligand-bound aptamer
U1 site
ligand
CDS
AATAAA
Gene expression
3' UTR with polyA sequence

FIG. 4a

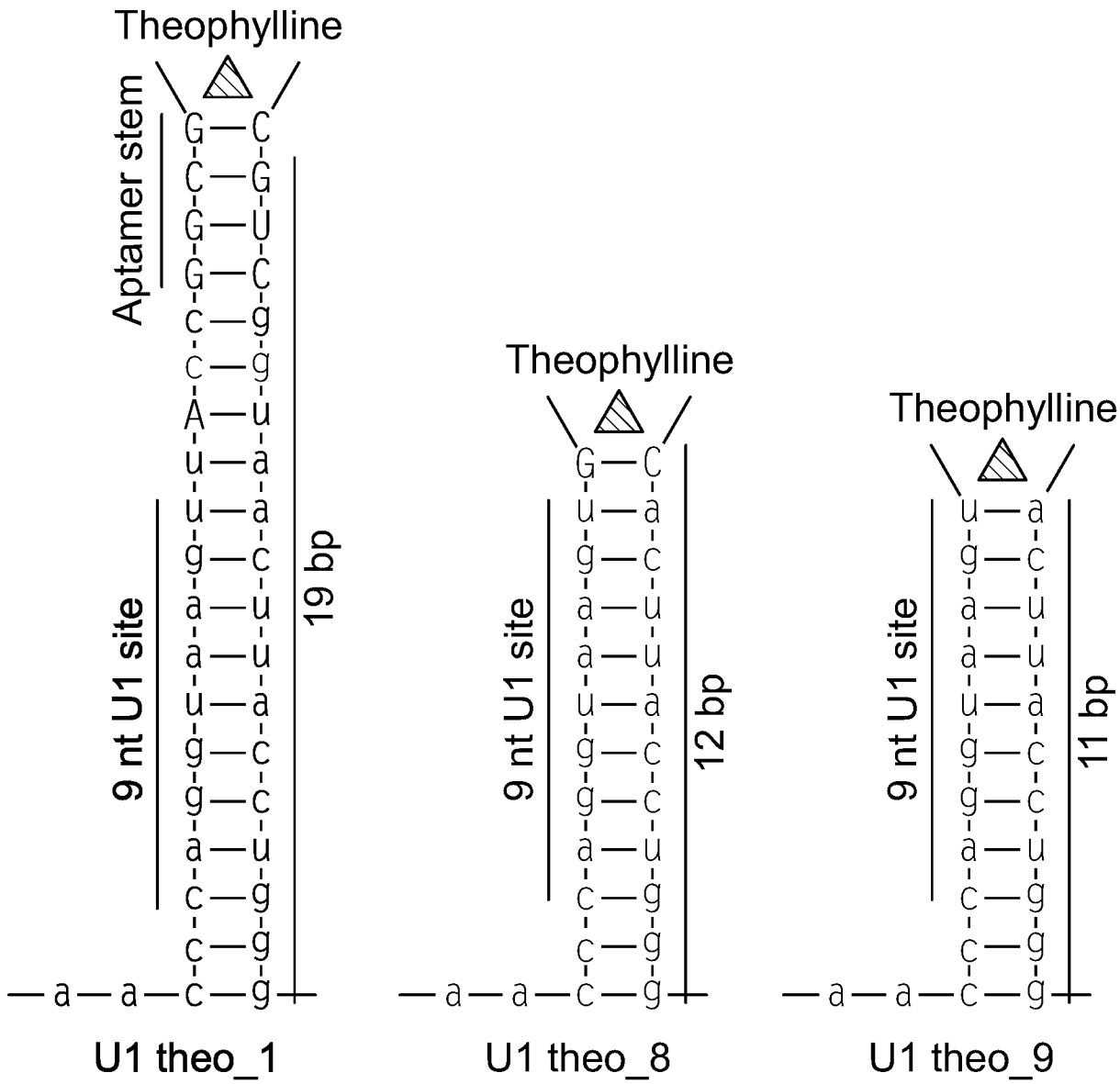

Theophylline
Aptamer stem
9 nt U1 site
19 bp
U1 theo_1
Theophylline
9 nt U1 site
12 bp
U1 theo_8
Theophylline
9 nt U1 site
11 bp
U1 theo_9

FIG. 4b

Figure 4c

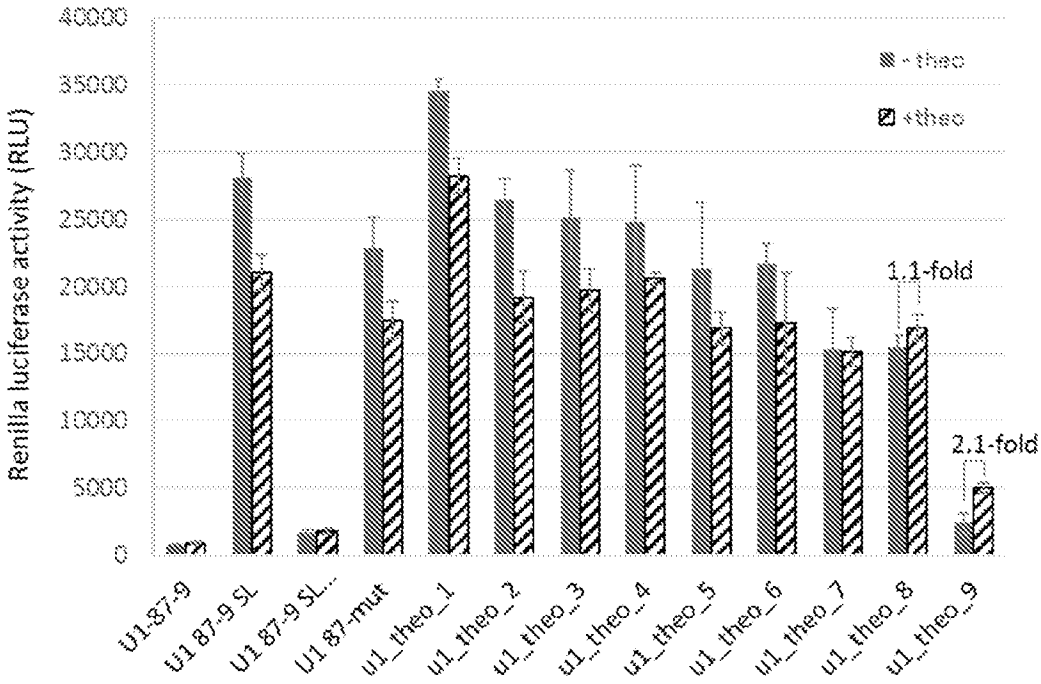

Renilla luciferase activity (RLU)
40000
35000
30000
25000
20000
15000
10000
5000
0
- theo
+theo
1.1-fold
2.1-fold
U1-87-9
U1 87-9 SL
U1 87-9 SL...
U1 87-mut
u1_theo_1
u1_theo_2
u1_theo_3
u1_theo_4
u1_theo_5
u1_theo_6
u1_theo_7
u1_theo_8
u1_theo_9

Figure 5a

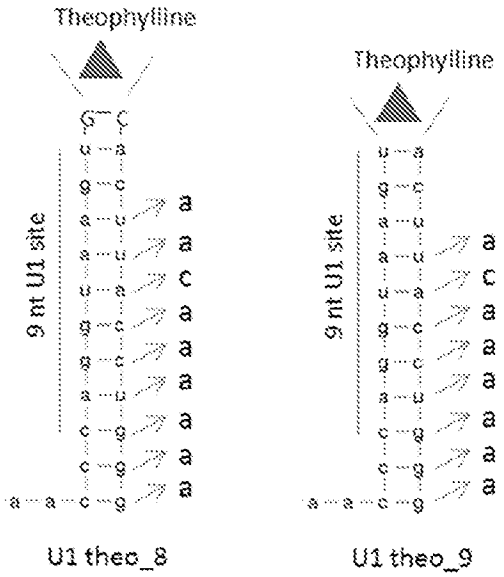


Figure 5b

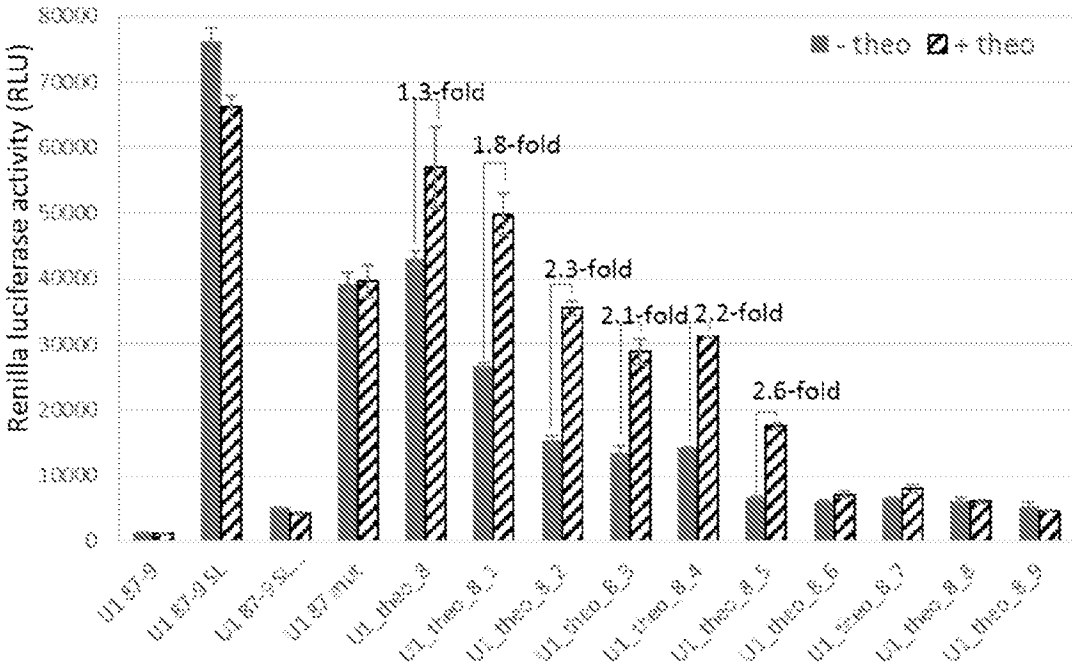
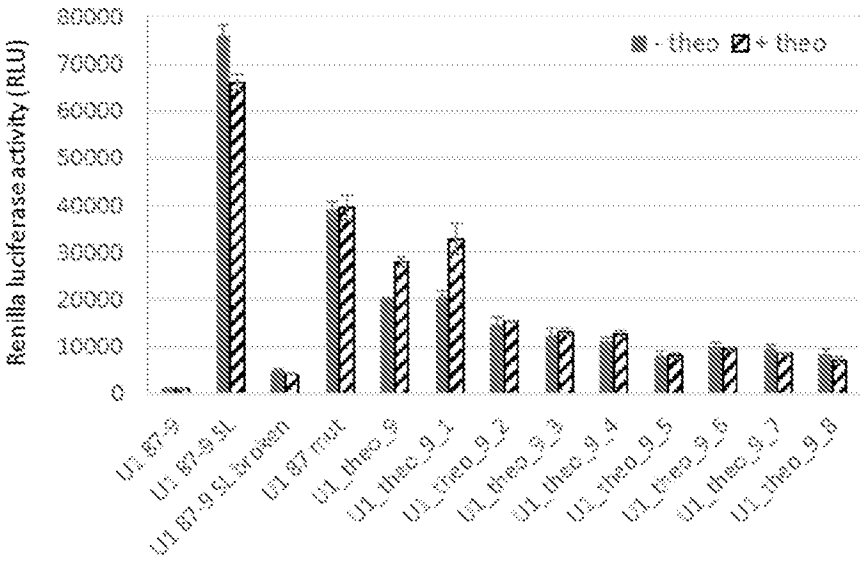


Figure 5c


Renilla luciferase activity (RLU)
80000
70000
60000
50000
40000
30000
20000
10000
0
- theo
+ theo
U1 87-9
U1 87-9 SL
U1 87-9 SL broken
U1 87 mut
U1_theo_9
U1_thec_9_1
U1_theo_9_2
U1_theo_9_3
U1_thec_9_4
U1_theo_9_5
U1_theo_9_6
U1_thec_9_7
U1_theo_9_8

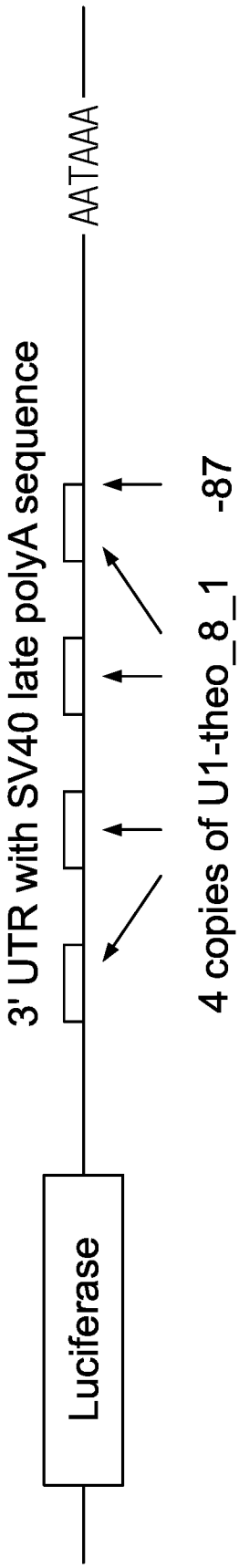

Luciferase
3' UTR with SV40 late polyA sequence
AATAAA
4 copies of U1-theo_8_1
-87

FIG. 6a

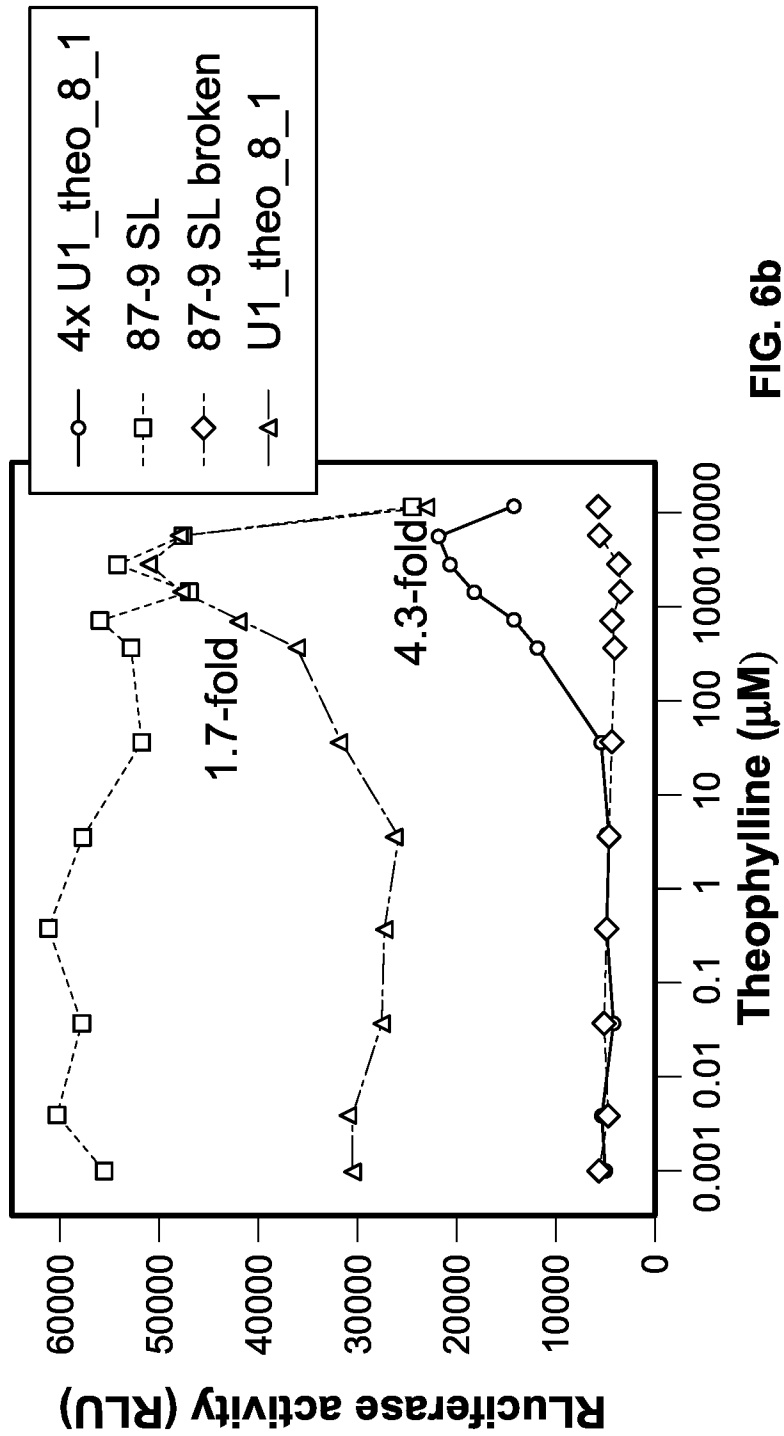

4x U1_theo_8_1
87-9 SL
87-9 SL broken
U1_theo_8_1
1.7-fold
4.3-fold
Rluciferase activity (RLU)
60000
50000
40000
30000
20000
10000
0
0.001 0.01 0.1 1 10 100 1000 10000
Theophylline (µM)

FIG. 6b

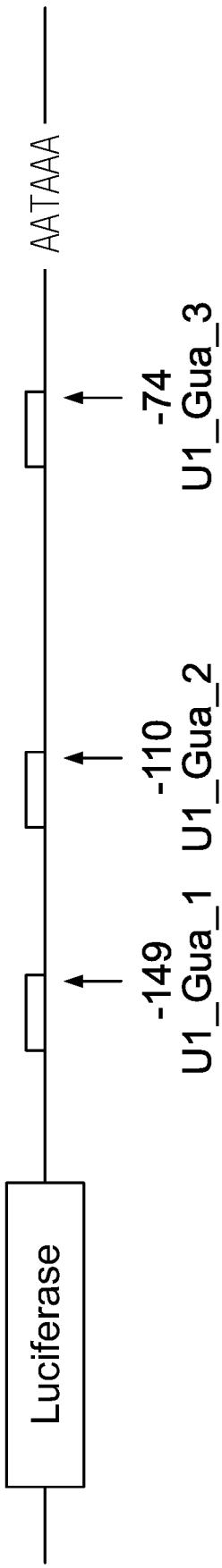

Luciferase
3' UTR with SV40 early polyA sequence
AATAAA
-149
U1_Gua_1
-110
U1_Gua_2
-74
U1_Gua_3

FIG. 7a

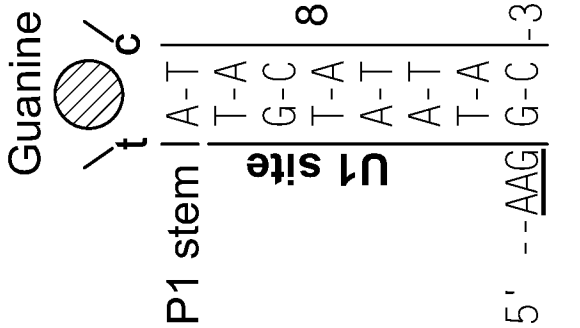

Guanine
t
c
P1 stem
A-T
T-A
G-C
T-A
A-T
A-T
T-A
G-C
U1 site
8 bp
5' --AAG
-3'

FIG. 7b

Figure 7c

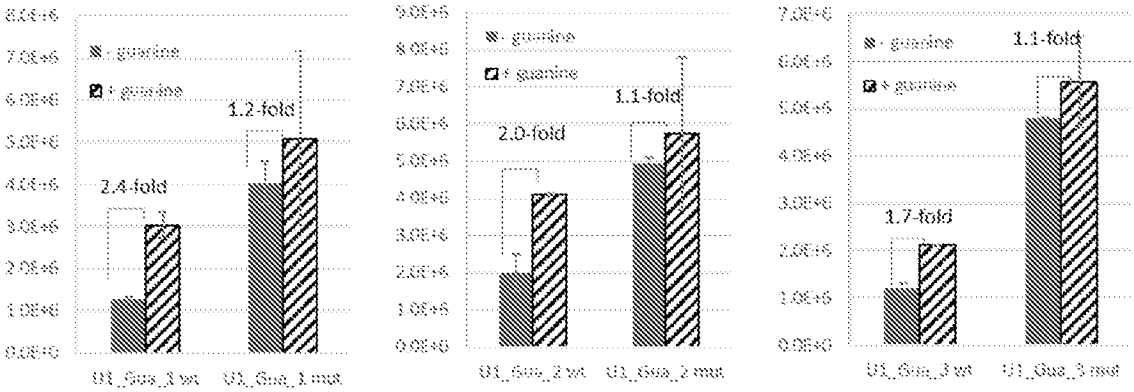
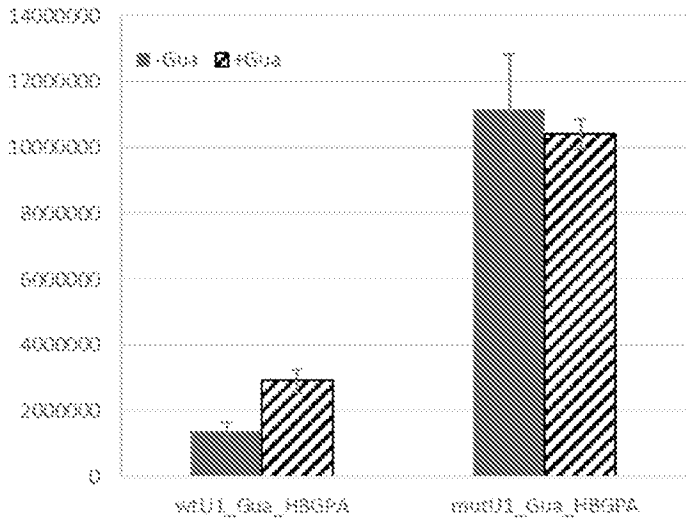

- guanine
+ guanine
2.4-fold
1.2-fold
2.0-fold
1.1-fold
1.7-fold
1.1-fold

Figure 8


14000000
12000000
10000000
8000000
6000000
4000000
2000000
0
-Gua
+Gua
wtU1_Gua_HBGPA
mutU1_Gua_HBGPA

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 17/16279

A. CLASSIFICATION OF SUBJECT MATTER
IPC(8) - A61K 48/00; C12N 15/00; C07H 21/04 (2017.01)
CPC - C12N 15/113; C12N 15/86; C12N 15/861; C12N 15/867; C12N 15/8645; C12N 15/11; C12N 15/115; C12N 2310/16; C12N 15/111

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)
See Search History Document

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

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

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 2012/0107331 A1 (STROBEL et al.) 03 May 2012 (03.05.2012), para [0009], [0023], [0032], [0036], [0039], [0040], [0041], [0068], [0072], [0075], [0076], [0077], [0079],[0081], [0087], Fig 1A, and pg 33, Table 2	1-5, 9-11
Y	US 2009/0170793 A1 (Gaur) 02 July 2009 (02.07.2009), Abstract, para [0003], [0004], [0011], [0047], [0058], [0080], [0116], and [0132]	1-5, 9-11
Y	US 2010/0173407 A1 (WYPIJEWSKI et al.) 08 July 2010 (08.07.2010), Abstract, para [0009], [0013], [0014], [0015], [0021], [0049], [0053], and SEQ ID NO: 2	1-5, 9-11
A	US 2003/0082149 A1 (ROWE et al.) 01 May 2003 (01.05.2003), para [0036], [0039], [0046], [0049], and SEQ ID NO: 2	1-5, 9-11
A	MALCA et al., The U1 snRNP Base Pairs with the 5' Splice Site within a Penta-snRNP Complex. Mol Cell Biol. 2003, Vol. 23(10), p. 3442-55. Entire documentation, especially Abstract; pg 3442, col 1; pg 3444, Fig 1; and pg 3446, Fig 3	1-5, 9-11
A	KIM et al., An artificial riboswitch for controlling pre-mRNA splicing. RNA. 2005, Vol. 11(11), p. 1667-77. Entire documentation, especially Abstract; and pg 1668, col 2, lower para, and Fig 1	1-5, 9-11

☐ 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	Date of mailing of the international search report
14 April 2017	26 MAY 2017

Name and mailing address of the ISA/US	Authorized officer:
Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-8300	Lee W. Young PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 17/16279

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.: 6-8, 12-25
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

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

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

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.