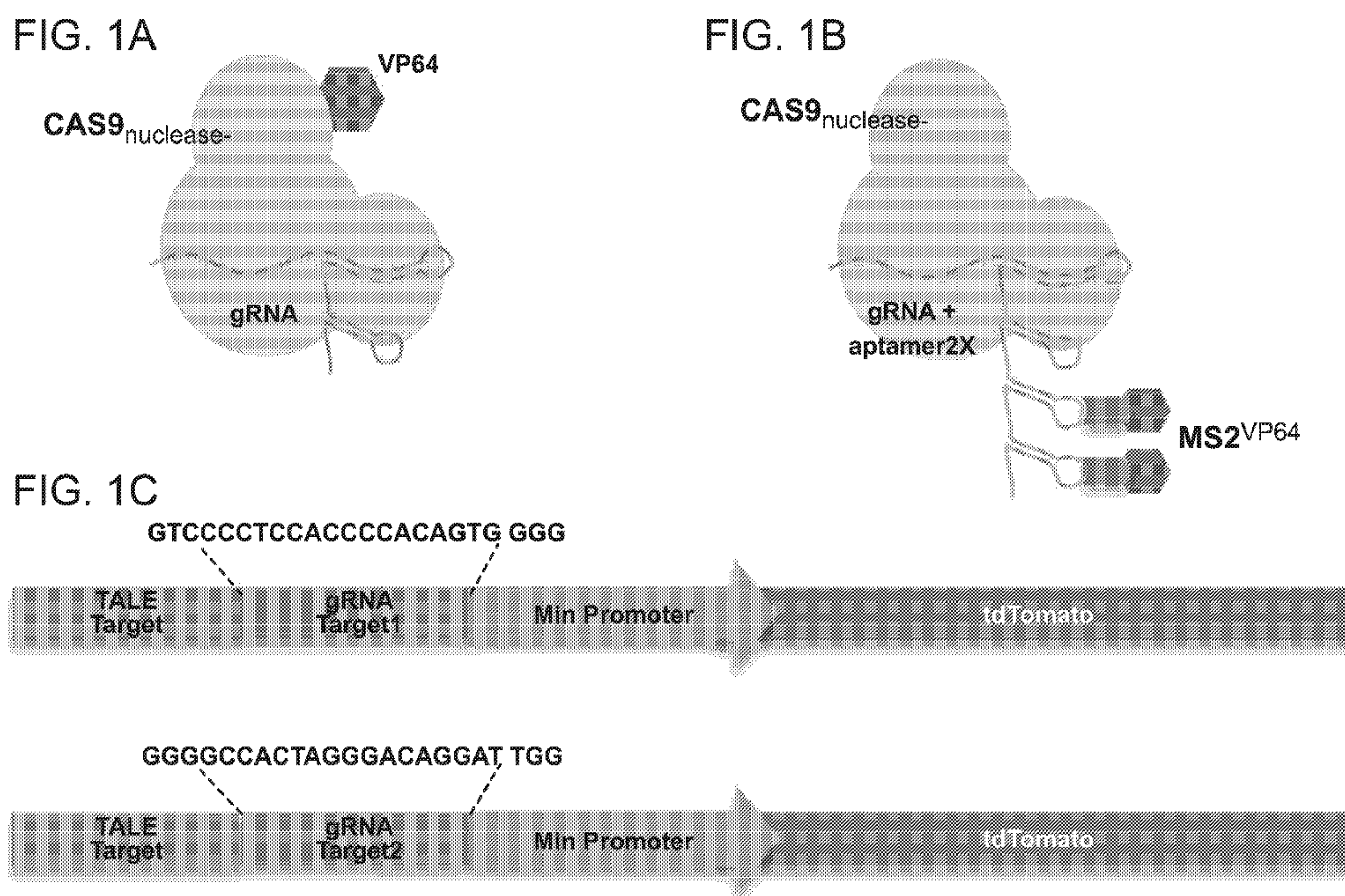




(86) **Date de dépôt PCT/PCT Filing Date:** 2014/06/04
(87) **Date publication PCT/PCT Publication Date:** 2014/12/11
(85) **Entrée phase nationale/National Entry:** 2015/12/04
(86) **N° demande PCT/PCT Application No.:** US 2014/040868
(87) **N° publication PCT/PCT Publication No.:** 2014/197568
(30) **Priorité/Priority:** 2013/06/04 (US61/830,787)

(51) **Cl.Int./Int.Cl.** **A61K 48/00** (2006.01),
A61K 31/7105 (2006.01), **A61K 38/17** (2006.01),
C07H 21/02 (2006.01), **C12N 15/00** (2006.01),
C12N 5/07 (2010.01)
(71) **Demandeur/Applicant:**
PRESIDENT AND FELLOWS OF HARVARD COLLEGE,
US
(72) **Inventeurs/Inventors:**
CHURCH, GEORGE M., US;
MALI, PRASHANT G., US;
ESVELT, KEVIN M., US
(74) **Agent:** KIRBY EADES GALE BAKER

(54) **Titre : REGULATION DE LA TRANSCRIPTION A GUIDAGE ARN**
(54) **Title: RNA-GUIDED TRANSCRIPTIONAL REGULATION**



(57) **Abrégé/Abstract:**

Methods of modulating expression of a target nucleic acid in a cell are provided including introducing into the cell a first foreign nucleic acid encoding one or more RNAs complementary to DNA, wherein the DNA includes the target nucleic acid, introducing into the cell a second foreign nucleic acid encoding a nuclease-null Cas9 protein that binds to the DNA and is guided by the one or more RNAs, introducing into the cell a third foreign nucleic acid encoding a transcriptional regulator protein or domain, wherein the one or more RNAs, the nuclease-null Cas9 protein, and the transcriptional regulator protein or domain are expressed, wherein the one or more RNAs, the nuclease-null Cas9 protein and the transcriptional regulator protein or domain co-localize to the DNA and wherein the transcriptional regulator protein or domain regulates expression of the target nucleic acid.



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

(19) World Intellectual Property

Organization

International Bureau



(10) International Publication Number

WO 2014/197568 A3

(43) International Publication Date

11 December 2014 (11.12.2014)

WIPO | PCT

(51) International Patent Classification:

A61K 48/00 (2006.01) C07H 21/02 (2006.01)

C12N 15/00 (2006.01)

(21) International Application Number:

PCT/US2014/040868

(22) International Filing Date:

4 June 2014 (04.06.2014)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

61/830,787

4 June 2013 (04.06.2013)

US

(71) Applicant: **PRESIDENT AND FELLOWS OF HARVARD COLLEGE** [US/US]; 17 Quincy Street, Cambridge, Massachusetts 02138 (US).

(72) Inventors: **CHURCH, George M.**; 218 Kent Street, Brookline, Massachusetts 02446 (US). **MALI, Prashant G.**; 88 Beacon Street, Somerville, Massachusetts 02143 (US). **ESVELT, Kevin M.**; 20 Towbridge Street, Apt. 2, Cambridge, Massachusetts 02138 (US).

(74) Agent: **IWANICKI, John P.**; 28 State Street, Suite 1800, Boston, Massachusetts 02109 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY,

BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, 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, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

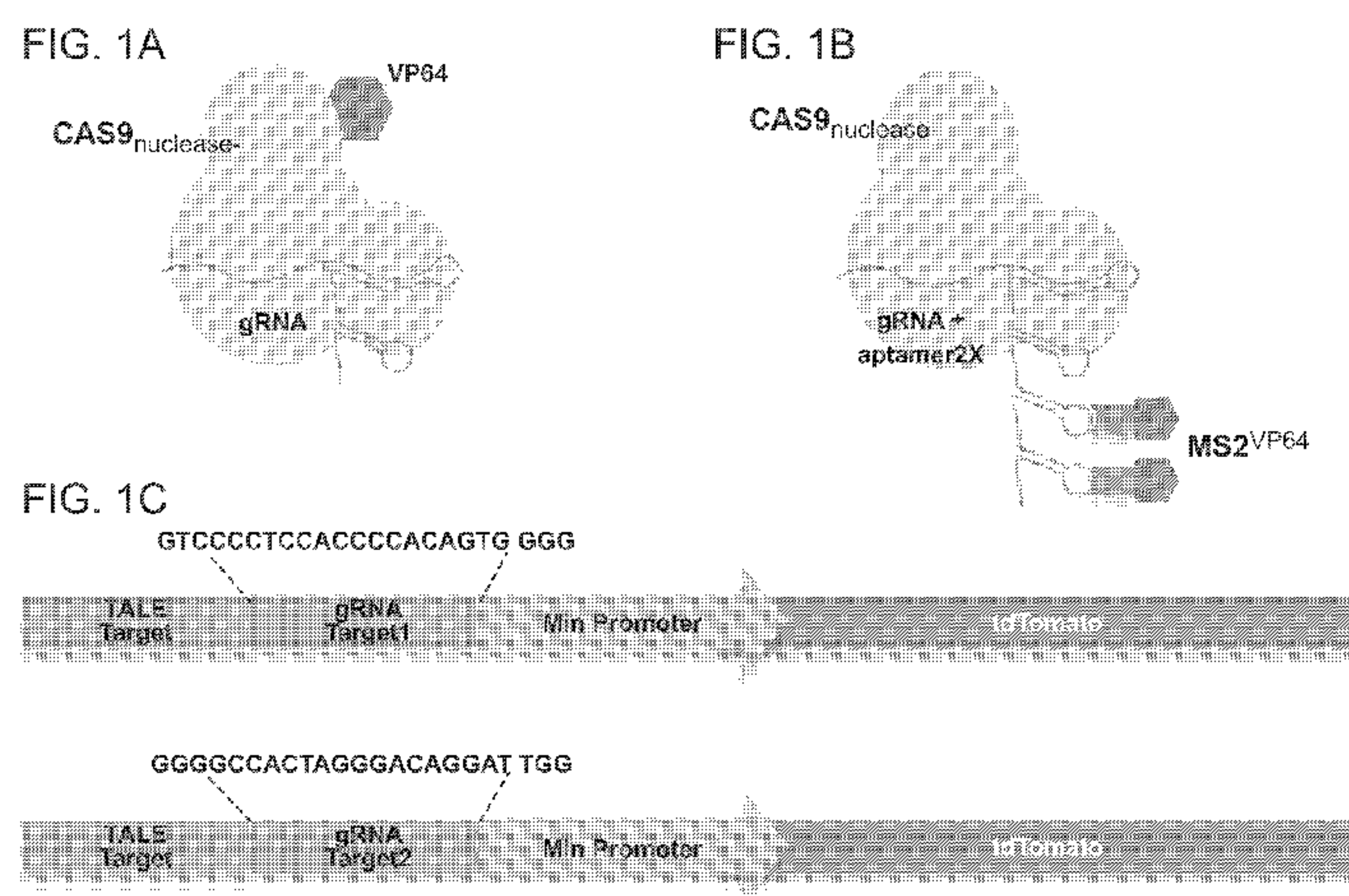
Published:

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

(88) Date of publication of the international search report:

12 March 2015

(54) Title: RNA-GUIDEDED TRANSCRIPTIONAL REGULATION



(57) Abstract: Methods of modulating expression of a target nucleic acid in a cell are provided including introducing into the cell a first foreign nucleic acid encoding one or more RNAs complementary to DNA, wherein the DNA includes the target nucleic acid, introducing into the cell a second foreign nucleic acid encoding a nuclease-null Cas9 protein that binds to the DNA and is guided by the one or more RNAs, introducing into the cell a third foreign nucleic acid encoding a transcriptional regulator protein or domain, wherein the one or more RNAs, the nuclease-null Cas9 protein, and the transcriptional regulator protein or domain are expressed, wherein the one or more RNAs, the nuclease-null Cas9 protein and the transcriptional regulator protein or domain co-localize to the DNA and wherein the transcriptional regulator protein or domain regulates expression of the target nucleic acid.

WO 2014/197568 A3

RNA-GUIDED TRANSCRIPTIONAL REGULATION

RELATED APPLICATION DATA

This application claims priority to U.S. Provisional Patent Application No. 61/830,787 filed on June 4, 2013 and is hereby incorporated herein by reference in its entirety for all purposes.

5

STATEMENT OF GOVERNMENT INTERESTS

This invention was made with government support under Grant No. P50 HG005550 from the National Institutes of health and DE-FG02-02ER63445 from the Department of Energy. The government has certain rights in the invention.

BACKGROUND

10 Bacterial and archaeal CRISPR-Cas systems rely on short guide RNAs in complex with Cas proteins to direct degradation of complementary sequences present within invading foreign nucleic acid. See Deltcheva, E. et al. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* 471, 602-607 (2011); Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proceedings of the National Academy of Sciences of the United States of America* 109, E2579-2586 (2012); Jinek, M. et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337, 816-821 (2012); Sapranas, R. et al. The *Streptococcus thermophilus* CRISPR/Cas system provides immunity in *Escherichia coli*. *Nucleic acids research* 39, 9275-9282 (2011); and Bhaya, D., Davison, M. & Barrangou, R. 15 CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. *Annual review of genetics* 45, 273-297 (2011). A recent *in vitro* reconstitution of the *S. pyogenes* type II CRISPR system demonstrated that crRNA ("CRISPR RNA") fused to a normally trans-encoded tracrRNA ("trans-activating CRISPR RNA") is sufficient to direct Cas9 protein to sequence-specifically cleave target DNA sequences matching the crRNA. Expressing a gRNA 20 homologous to a target site results in Cas9 recruitment and degradation of the target DNA. See H. Deveau *et al.*, Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *Journal of Bacteriology* 190, 1390 (Feb, 2008).

SUMMARY

Aspects of the present disclosure are directed to a complex of a guide RNA, a DNA 30 binding protein and a double stranded DNA target sequence. According to certain aspects, DNA binding proteins within the scope of the present disclosure include a protein that forms a complex

with the guide RNA and with the guide RNA guiding the complex to a double stranded DNA sequence wherein the complex binds to the DNA sequence. This aspect of the present disclosure may be referred to as co-localization of the RNA and DNA binding protein to or with the double stranded DNA. In this manner, a DNA binding protein-guide RNA complex may be used to
5 localize a transcriptional regulator protein or domain at target DNA so as to regulate expression of target DNA.

According to certain aspects, a method of modulating expression of a target nucleic acid in a cell is provided including introducing into the cell a first foreign nucleic acid encoding one or more RNAs (ribonucleic acids) complementary to DNA (deoxyribonucleic acid), wherein the DNA
10 includes the target nucleic acid, introducing into the cell a second foreign nucleic acid encoding an RNA guided nuclease-null DNA binding protein that binds to the DNA and is guided by the one or more RNAs, introducing into the cell a third foreign nucleic acid encoding a transcriptional regulator protein or domain, wherein the one or more RNAs, the RNA guided nuclease-null DNA binding protein, and the transcriptional regulator protein or domain are expressed, wherein the one
15 or more RNAs, the RNA guided nuclease-null DNA binding protein and the transcriptional regulator protein or domain co-localize to the DNA and wherein the transcriptional regulator protein or domain regulates expression of the target nucleic acid.

According to one aspect, the foreign nucleic acid encoding an RNA guided nuclease-null DNA binding protein further encodes the transcriptional regulator protein or domain fused to the
20 RNA guided nuclease-null DNA binding protein. According to one aspect, the foreign nucleic acid encoding one or more RNAs further encodes a target of an RNA-binding domain and the foreign nucleic acid encoding the transcriptional regulator protein or domain further encodes an RNA-binding domain fused to the transcriptional regulator protein or domain.

According to one aspect, the cell is a eukaryotic cell. According to one aspect, the cell is a
25 yeast cell, a plant cell or an animal cell. According to one aspect, the cell is a mammalian cell.

According to one aspect, the RNA is between about 10 to about 500 nucleotides. According to one aspect, the RNA is between about 20 to about 100 nucleotides.

According to one aspect, the transcriptional regulator protein or domain is a transcriptional activator. According to one aspect, the transcriptional regulator protein or domain upregulates
30 expression of the target nucleic acid. According to one aspect, the transcriptional regulator protein or domain upregulates expression of the target nucleic acid to treat a disease or detrimental condition. According to one aspect, the target nucleic acid is associated with a disease or detrimental condition.

According to one aspect, the one or more RNAs is a guide RNA. According to one aspect,
35 the one or more RNAs is a tracrRNA-crRNA fusion. According to one aspect, the guide RNA includes a spacer sequence and a tracer mate sequence. The guide RNA may also include a tracr

sequence, a portion of which hybridizes to the tracer mate sequence. The guide RNA may also include a linker nucleic acid sequence which links the tracer mate sequence and the tracer sequence to produce the tracerRNA-crRNA fusion. The spacer sequence binds to target DNA, such as by hybridization.

5 According to one aspect, the guide RNA includes a truncated spacer sequence. According to one aspect, the guide RNA includes a truncated spacer sequence having a 1 base truncation at the 5' end of the spacer sequence. According to one aspect, the guide RNA includes a truncated spacer sequence having a 2 base truncation at the 5' end of the spacer sequence. According to one aspect, the guide RNA includes a truncated spacer sequence having a 3 base truncation at the 5' end of the spacer sequence. According to one aspect, the guide RNA includes a truncated spacer sequence having a 4 base truncation at the 5' end of the spacer sequence. Accordingly, the spacer sequence may have a 1 to 4 base truncation at the 5' end of the spacer sequence.

15 According to certain embodiments, the spacer sequence may include between about 16 to about 20 nucleotides which hybridize to the target nucleic acid sequence. According to certain embodiments, the spacer sequence may include about 20 nucleotides which hybridize to the target nucleic acid sequence.

According to certain aspects, the linker nucleic acid sequence may include between about 4 and about 6 nucleic acids.

20 According to certain aspects, the tracer sequence may include between about 60 to about 500 nucleic acids. According to certain aspects, the tracer sequence may include between about 64 to about 500 nucleic acids. According to certain aspects, the tracer sequence may include between about 65 to about 500 nucleic acids. According to certain aspects, the tracer sequence may include between about 66 to about 500 nucleic acids. According to certain aspects, the tracer sequence may include between about 67 to about 500 nucleic acids. According to certain aspects, the tracer sequence may include between about 68 to about 500 nucleic acids. According to certain aspects, the tracer sequence may include between about 69 to about 500 nucleic acids. According to certain aspects, the tracer sequence may include between about 70 to about 500 nucleic acids. According to certain aspects, the tracer sequence may include between about 80 to about 500 nucleic acids. According to certain aspects, the tracer sequence may include between about 90 to about 500 nucleic acids. According to certain aspects, the tracer sequence may include between about 100 to about 500 nucleic acids.

35 According to certain aspects, the tracer sequence may include between about 60 to about 200 nucleic acids. According to certain aspects, the tracer sequence may include between about 64 to about 200 nucleic acids. According to certain aspects, the tracer sequence may include between about 65 to about 200 nucleic acids. According to certain aspects, the tracer sequence may include between about 66 to about 200 nucleic acids. According to certain aspects, the tracer sequence may

include between about 67 to about 200 nucleic acids. According to certain aspects, the tracer sequence may include between about 68 to about 200 nucleic acids. According to certain aspects, the tracer sequence may include between about 69 to about 200 nucleic acids. According to certain aspects, the tracer sequence may include between about 70 to about 200 nucleic acids. According to certain aspects, the tracer sequence may include between about 80 to about 200 nucleic acids. According to certain aspects, the tracer sequence may include between about 90 to about 200 nucleic acids. According to certain aspects, the tracer sequence may include between about 100 to about 200 nucleic acids.

An exemplary guide RNA is depicted in Figure 5B.

According to one aspect, the DNA is genomic DNA, mitochondrial DNA, viral DNA, or exogenous DNA.

According to certain aspects, a method of modulating expression of a target nucleic acid in a cell is provided including introducing into the cell a first foreign nucleic acid encoding one or more RNAs (ribonucleic acids) complementary to DNA (deoxyribonucleic acid), wherein the DNA includes the target nucleic acid, introducing into the cell a second foreign nucleic acid encoding an RNA guided nuclease-null DNA binding protein of a Type II CRISPR System that binds to the DNA and is guided by the one or more RNAs, introducing into the cell a third foreign nucleic acid encoding a transcriptional regulator protein or domain, wherein the one or more RNAs, the RNA guided nuclease-null DNA binding protein of a Type II CRISPR System, and the transcriptional regulator protein or domain are expressed, wherein the one or more RNAs, the RNA guided nuclease-null DNA binding protein of a Type II CRISPR System and the transcriptional regulator protein or domain co-localize to the DNA and wherein the transcriptional regulator protein or domain regulates expression of the target nucleic acid.

According to one aspect, the foreign nucleic acid encoding an RNA guided nuclease-null DNA binding protein of a Type II CRISPR System further encodes the transcriptional regulator protein or domain fused to the RNA guided nuclease-null DNA binding protein of a Type II CRISPR System. According to one aspect, the foreign nucleic acid encoding one or more RNAs further encodes a target of an RNA-binding domain and the foreign nucleic acid encoding the transcriptional regulator protein or domain further encodes an RNA-binding domain fused to the transcriptional regulator protein or domain.

According to one aspect, the cell is a eukaryotic cell. According to one aspect, the cell is a yeast cell, a plant cell or an animal cell. According to one aspect, the cell is a mammalian cell.

According to one aspect, the RNA is between about 10 to about 500 nucleotides. According to one aspect, the RNA is between about 20 to about 100 nucleotides.

According to one aspect, the transcriptional regulator protein or domain is a transcriptional activator. According to one aspect, the transcriptional regulator protein or domain upregulates

expression of the target nucleic acid. According to one aspect, the transcriptional regulator protein or domain upregulates expression of the target nucleic acid to treat a disease or detrimental condition. According to one aspect, the target nucleic acid is associated with a disease or detrimental condition.

5 According to one aspect, the one or more RNAs is a guide RNA. According to one aspect, the one or more RNAs is a tracrRNA-crRNA fusion.

According to one aspect, the DNA is genomic DNA, mitochondrial DNA, viral DNA, or exogenous DNA.

10 According to certain aspects, a method of modulating expression of a target nucleic acid in a cell is provided including introducing into the cell a first foreign nucleic acid encoding one or more RNAs (ribonucleic acids) complementary to DNA (deoxyribonucleic acid), wherein the DNA includes the target nucleic acid, introducing into the cell a second foreign nucleic acid encoding a nuclease-null Cas9 protein that binds to the DNA and is guided by the one or more RNAs, introducing into the cell a third foreign nucleic acid encoding a transcriptional regulator protein or domain, wherein the one or more RNAs, the nuclease-null Cas9 protein, and the transcriptional regulator protein or domain are expressed, wherein the one or more RNAs, the nuclease-null Cas9 protein and the transcriptional regulator protein or domain co-localize to the DNA and wherein the transcriptional regulator protein or domain regulates expression of the target nucleic acid.

20 According to one aspect, the foreign nucleic acid encoding a nuclease-null Cas9 protein further encodes the transcriptional regulator protein or domain fused to the nuclease-null Cas9 protein. According to one aspect, the foreign nucleic acid encoding one or more RNAs further encodes a target of an RNA-binding domain and the foreign nucleic acid encoding the transcriptional regulator protein or domain further encodes an RNA-binding domain fused to the transcriptional regulator protein or domain.

25 According to one aspect, the cell is a eukaryotic cell. According to one aspect, the cell is a yeast cell, a plant cell or an animal cell. According to one aspect, the cell is a mammalian cell.

According to one aspect, the RNA is between about 10 to about 500 nucleotides. According to one aspect, the RNA is between about 20 to about 100 nucleotides.

30 According to one aspect, the transcriptional regulator protein or domain is a transcriptional activator. According to one aspect, the transcriptional regulator protein or domain upregulates expression of the target nucleic acid. According to one aspect, the transcriptional regulator protein or domain upregulates expression of the target nucleic acid to treat a disease or detrimental condition. According to one aspect, the target nucleic acid is associated with a disease or detrimental condition.

35 According to one aspect, the one or more RNAs is a guide RNA. According to one aspect, the one or more RNAs is a tracrRNA-crRNA fusion.

According to one aspect, the DNA is genomic DNA, mitochondrial DNA, viral DNA, or exogenous DNA.

According to one aspect a cell is provided that includes a first foreign nucleic acid encoding one or more RNAs complementary to DNA, wherein the DNA includes a target nucleic acid, a second foreign nucleic acid encoding an RNA guided nuclease-null DNA binding protein, and a third foreign nucleic acid encoding a transcriptional regulator protein or domain wherein the one or more RNAs, the RNA guided nuclease-null DNA binding protein and the transcriptional regulator protein or domain are members of a co-localization complex for the target nucleic acid.

According to one aspect, the foreign nucleic acid encoding an RNA guided nuclease-null DNA binding protein further encodes the transcriptional regulator protein or domain fused to an RNA guided nuclease-null DNA binding protein. According to one aspect, the foreign nucleic acid encoding one or more RNAs further encodes a target of an RNA-binding domain and the foreign nucleic acid encoding the transcriptional regulator protein or domain further encodes an RNA-binding domain fused to the transcriptional regulator protein or domain.

According to one aspect, the cell is a eukaryotic cell. According to one aspect, the cell is a yeast cell, a plant cell or an animal cell. According to one aspect, the cell is a mammalian cell.

According to one aspect, the RNA is between about 10 to about 500 nucleotides. According to one aspect, the RNA is between about 20 to about 100 nucleotides.

According to one aspect, the transcriptional regulator protein or domain is a transcriptional activator. According to one aspect, the transcriptional regulator protein or domain upregulates expression of the target nucleic acid. According to one aspect, the transcriptional regulator protein or domain upregulates expression of the target nucleic acid to treat a disease or detrimental condition. According to one aspect, the target nucleic acid is associated with a disease or detrimental condition.

According to one aspect, the one or more RNAs is a guide RNA. According to one aspect, the one or more RNAs is a tracrRNA-crRNA fusion.

According to one aspect, the DNA is genomic DNA, mitochondrial DNA, viral DNA, or exogenous DNA.

According to certain aspects, the RNA guided nuclease-null DNA binding protein is an RNA guided nuclease-null DNA binding protein of a Type II CRISPR System. According to certain aspects, the RNA guided nuclease-null DNA binding protein is a nuclease-null Cas9 protein.

According to one aspect, a method of altering a DNA target nucleic acid in a cell is provided that includes introducing into the cell a first foreign nucleic acid encoding two or more RNAs with each RNA being complementary to an adjacent site in the DNA target nucleic acid, introducing into the cell a second foreign nucleic acid encoding at least one RNA guided DNA

binding protein nickase and being guided by the two or more RNAs, wherein the two or more RNAs and the at least one RNA guided DNA binding protein nickase are expressed and wherein the at least one RNA guided DNA binding protein nickase co-localizes with the two or more RNAs to the DNA target nucleic acid and nicks the DNA target nucleic acid resulting in two or more adjacent nicks.

According to one aspect, a method of altering a DNA target nucleic acid in a cell is provided that includes introducing into the cell a first foreign nucleic acid encoding two or more RNAs with each RNA being complementary to an adjacent site in the DNA target nucleic acid, introducing into the cell a second foreign nucleic acid encoding at least one RNA guided DNA binding protein nickase of a Type II CRISPR System and being guided by the two or more RNAs, wherein the two or more RNAs and the at least one RNA guided DNA binding protein nickase of a Type II CRISPR System are expressed and wherein the at least one RNA guided DNA binding protein nickase of a Type II CRISPR System co-localizes with the two or more RNAs to the DNA target nucleic acid and nicks the DNA target nucleic acid resulting in two or more adjacent nicks.

According to one aspect, a method of altering a DNA target nucleic acid in a cell is provided that includes introducing into the cell a first foreign nucleic acid encoding two or more RNAs with each RNA being complementary to an adjacent site in the DNA target nucleic acid, introducing into the cell a second foreign nucleic acid encoding at least one Cas9 protein nickase having one inactive nuclease domain and being guided by the two or more RNAs, wherein the two or more RNAs and the at least one Cas9 protein nickase are expressed and wherein the at least one Cas9 protein nickase co-localizes with the two or more RNAs to the DNA target nucleic acid and nicks the DNA target nucleic acid resulting in two or more adjacent nicks.

According to the methods of altering a DNA target nucleic acid, the two or more adjacent nicks are on the same strand of the double stranded DNA. According to one aspect, the two or more adjacent nicks are on the same strand of the double stranded DNA and result in homologous recombination. According to one aspect, the two or more adjacent nicks are on different strands of the double stranded DNA. According to one aspect, the two or more adjacent nicks are on different strands of the double stranded DNA and create double stranded breaks. According to one aspect, the two or more adjacent nicks are on different strands of the double stranded DNA and create double stranded breaks resulting in nonhomologous end joining. According to one aspect, the two or more adjacent nicks are on different strands of the double stranded DNA and are offset with respect to one another. According to one aspect, the two or more adjacent nicks are on different strands of the double stranded DNA and are offset with respect to one another and create double stranded breaks. According to one aspect, the two or more adjacent nicks are on different strands of the double stranded DNA and are offset with respect to one another and create double stranded breaks resulting in nonhomologous end joining. According to one aspect, the method

further includes introducing into the cell a third foreign nucleic acid encoding a donor nucleic acid sequence wherein the two or more nicks results in homologous recombination of the target nucleic acid with the donor nucleic acid sequence.

According to one aspect, a method of altering a DNA target nucleic acid in a cell is provided including introducing into the cell a first foreign nucleic acid encoding two or more RNAs with each RNA being complementary to an adjacent site in the DNA target nucleic acid, introducing into the cell a second foreign nucleic acid encoding at least one RNA guided DNA binding protein nickase and being guided by the two or more RNAs, and wherein the two or more RNAs and the at least one RNA guided DNA binding protein nickase are expressed and wherein the at least one RNA guided DNA binding protein nickase co-localizes with the two or more RNAs to the DNA target nucleic acid and nicks the DNA target nucleic acid resulting in two or more adjacent nicks, and wherein the two or more adjacent nicks are on different strands of the double stranded DNA and create double stranded breaks resulting in fragmentation of the target nucleic acid thereby preventing expression of the target nucleic acid.

According to one aspect, a method of altering a DNA target nucleic acid in a cell is provided including introducing into the cell a first foreign nucleic acid encoding two or more RNAs with each RNA being complementary to an adjacent site in the DNA target nucleic acid, introducing into the cell a second foreign nucleic acid encoding at least one RNA guided DNA binding protein nickase of a Type II CRISPR system and being guided by the two or more RNAs, and wherein the two or more RNAs and the at least one RNA guided DNA binding protein nickase of a Type II CRISPR System are expressed and wherein the at least one RNA guided DNA binding protein nickase of a Type II CRISPR System co-localizes with the two or more RNAs to the DNA target nucleic acid and nicks the DNA target nucleic acid resulting in two or more adjacent nicks, and wherein the two or more adjacent nicks are on different strands of the double stranded DNA and create double stranded breaks resulting in fragmentation of the target nucleic acid thereby preventing expression of the target nucleic acid.

According to one aspect, a method of altering a DNA target nucleic acid in a cell is provided including introducing into the cell a first foreign nucleic acid encoding two or more RNAs with each RNA being complementary to an adjacent site in the DNA target nucleic acid, introducing into the cell a second foreign nucleic acid encoding at least one Cas9 protein nickase having one inactive nuclease domain and being guided by the two or more RNAs, and wherein the two or more RNAs and the at least one Cas9 protein nickase are expressed and wherein the at least one Cas9 protein nickase co-localizes with the two or more RNAs to the DNA target nucleic acid and nicks the DNA target nucleic acid resulting in two or more adjacent nicks, and wherein the two or more adjacent nicks are on different strands of the double stranded DNA and create double

stranded breaks resulting in fragmentation of the target nucleic acid thereby preventing expression of the target nucleic acid.

According to one aspect, a cell is provided including a first foreign nucleic acid encoding two or more RNAs with each RNA being complementary to an adjacent site in a DNA target nucleic acid, and a second foreign nucleic acid encoding at least one RNA guided DNA binding protein nickase, and wherein the two or more RNAs and the at least one RNA guided DNA binding protein nickase are members of a co-localization complex for the DNA target nucleic acid.

According to one aspect, the RNA guided DNA binding protein nickase is an RNA guided DNA binding protein nickase of a Type II CRISPR System. According to one aspect, the RNA guided DNA binding protein nickase is a Cas9 protein nickase having one inactive nuclease domain.

According to one aspect, the cell is a eukaryotic cell. According to one aspect, the cell is a yeast cell, a plant cell or an animal cell. According to one aspect, the cell is a mammalian cell.

According to one aspect, the RNA includes between about 10 to about 500 nucleotides.

According to one aspect, the RNA includes between about 20 to about 100 nucleotides.

According to one aspect, the target nucleic acid is associated with a disease or detrimental condition.

According to one aspect, the two or more RNAs are guide RNAs. According to one aspect, the two or more RNAs are tracrRNA-crRNA fusions.

According to one aspect, the DNA target nucleic acid is genomic DNA, mitochondrial DNA, viral DNA, or exogenous DNA.

Further features and advantages of certain embodiments of the present invention will become more fully apparent in the following description of embodiments and drawings thereof, and from the claims.

BRIEF DESCRIPTION OF THE DRAWINGS

The patent or application file contains drawings executed in color. Copies of this patent or patent application publication with the color drawings will be provided by the Office upon request and payment of the necessary fee. The foregoing and other features and advantages of the present embodiments will be more fully understood from the following detailed description of illustrative embodiments taken in conjunction with the accompanying drawings in which:

Figure 1A and Figure 1B are schematics of RNA-guided transcriptional activation. Figure 1C is a design of a reporter construct. Figure 1D shows data demonstrating that Cas9N-VP64 fusions display RNA-guided transcriptional activation as assayed by both fluorescence-activated cell sorting (FACS) and immunofluorescence assays (IF). Figure 1E shows assay data by FACS and IF demonstrating gRNA sequence-specific transcriptional activation from reporter constructs

in the presence of Cas9N, MS2-VP64 and gRNA bearing the appropriate MS2 aptamer binding sites. Figure 1F depicts data demonstrating transcriptional induction by individual gRNAs and multiple gRNAs.

Figure 2A depicts a methodology for evaluating the landscape of targeting by Cas9-gRNA complexes and TALEs. Figure 2B depicts data demonstrating that a Cas9-gRNA complex is on average tolerant to 1-3 mutations in its target sequences. Figure 2C depicts data demonstrating that the Cas9-gRNA complex is largely insensitive to point mutations, except those localized to the PAM sequence. Figure 2D depicts heat plot data demonstrating that introduction of 2 base mismatches significantly impairs the Cas9-gRNA complex activity. Figure 2E depicts data demonstrating that an 18-mer TALE reveals is on average tolerant to 1-2 mutations in its target sequence. Figure 2F depicts data demonstrating the 18-mer TALE is, similar to the Cas9-gRNA complexes, largely insensitive to single base mismatched in its target. Figure 2G depicts heat plot data demonstrating that introduction of 2 base mismatches significantly impairs the 18-mer TALE activity.

Figure 3A depicts a schematic of a guide RNA design. Figure 3B depicts data showing percentage rate of non-homologous end joining for off-set nicks leading to 5' overhangs and off-set nicks leading to 3' overhangs. Figure 3C depicts data showing percentage rate of targeting for off-set nicks leading to 5' overhangs and off-set nicks leading to 3' overhangs.

Figure 4A is a schematic of a metal coordinating residue in RuvC PDB ID: 4EP4 (blue) position D7 (left), a schematic of HNH endonuclease domains from PDB IDs: 3M7K (orange) and 4H9D (cyan) including a coordinated Mg-ion (gray sphere) and DNA from 3M7K (purple) (middle) and a list of mutants analyzed (right). Figure 4B depicts data showing undetectable nuclease activity for Cas9 mutants m3 and m4, and also their respective fusions with VP64. Figure 4C is a higher-resolution examination of the data in Figure 4B.

Figure 5A is a schematic of a homologous recombination assay to determine Cas9-gRNA activity. Figure 5B depicts guide RNAs with random sequence insertions and percentage rate of homologous recombination

Figure 6A is a schematic of guide RNAs for the OCT4 gene. Figure 6B depicts transcriptional activation for a promoter-luciferase reporter construct. Figure 6C depicts transcriptional activation via qPCR of endogenous genes.

Figure 7A is a schematic of guide RNAs for the REX1 gene. Figure 7B depicts transcriptional activation for a promoter-luciferase reporter construct. Figure 7C depicts transcriptional activation via qPCR of endogenous genes.

Figure 8A depicts in schematic a high level specificity analysis processing flow for calculation of normalized expression levels. Figure 8B depicts data of distributions of percentages of binding sites by numbers of mismatches generated within a biased construct library. *Left:*

Theoretical distribution. *Right:* Distribution observed from an actual TALE construct library. Figure 8C depicts data of distributions of percentages of tag counts aggregated to binding sites by numbers of mismatches. *Left:* Distribution observed from the positive control sample. *Right:* Distribution observed from a sample in which a non-control TALE was induced.

5 Figure 9A depicts data for analysis of the targeting landscape of a Cas9-gRNA complex showing tolerance to 1-3 mutations in its target sequence. Figure 9B depicts data for analysis of the targeting landscape of a Cas9-gRNA complex showing insensitivity to point mutations, except those localized to the PAM sequence. Figure 9C depicts heat plot data for analysis of the targeting landscape of a Cas9-gRNA complex showing that introduction of 2 base mismatches significantly
10 impairs activity. Figure 9D depicts data from a nuclease mediated HR assay confirming that the predicted PAM for the *S. pyogenes* Cas9 is NGG and also NAG.

 Figure 10A depicts data from a nuclease mediated HR assay confirming that 18-mer TALEs tolerate multiple mutations in their target sequences. Figure 10B depicts data from analysis of the targeting landscape of TALEs of 3 different sizes (18-mer, 14-mer and 10-mer). Figure 10C
15 depicts data for 10-mer TALEs show near single-base mismatch resolution. Figure 10D depicts heat plot data for 10-mer TALEs show near single-base mismatch resolution.

 Figure 11A depicts designed guide RNAs. Figure 11B depicts percentage rate of non-homologous end joining for various guide RNAs.

DETAILED DESCRIPTION

20 Embodiments of the present disclosure are based on the use of DNA binding proteins to co-localize transcriptional regulator proteins or domains to DNA in a manner to regulate a target nucleic acid. Such DNA binding proteins are readily known to those of skill in the art to bind to DNA for various purposes. Such DNA binding proteins may be naturally occurring. DNA binding proteins included within the scope of the present disclosure include those which may be guided by
25 RNA, referred to herein as guide RNA. According to this aspect, the guide RNA and the RNA guided DNA binding protein form a co-localization complex at the DNA. According to certain aspects, the DNA binding protein may be a nuclease-null DNA binding protein. According to this aspect, the nuclease-null DNA binding protein may result from the alteration or modification of a DNA binding protein having nuclease activity. Such DNA binding proteins having nuclease
30 activity are known to those of skill in the art, and include naturally occurring DNA binding proteins having nuclease activity, such as Cas9 proteins present, for example, in Type II CRISPR systems. Such Cas9 proteins and Type II CRISPR systems are well documented in the art. See

Makarova et al., *Nature Reviews, Microbiology*, Vol. 9, June 2011, pp. 467-477 including all supplementary information hereby incorporated by reference in its entirety.

Exemplary DNA binding proteins having nuclease activity function to nick or cut double stranded DNA. Such nuclease activity may result from the DNA binding protein having one or
 5 more polypeptide sequences exhibiting nuclease activity. Such exemplary DNA binding proteins may have two separate nuclease domains with each domain responsible for cutting or nicking a particular strand of the double stranded DNA. Exemplary polypeptide sequences having nuclease activity known to those of skill in the art include the McrA-HNH nuclease related domain and the RuvC-like nuclease domain. Accordingly, exemplary DNA binding proteins are those that in
 10 nature contain one or more of the McrA-HNH nuclease related domain and the RuvC-like nuclease domain. According to certain aspects, the DNA binding protein is altered or otherwise modified to inactivate the nuclease activity. Such alteration or modification includes altering one or more amino acids to inactivate the nuclease activity or the nuclease domain. Such modification includes removing the polypeptide sequence or polypeptide sequences exhibiting nuclease activity, i.e. the
 15 nuclease domain, such that the polypeptide sequence or polypeptide sequences exhibiting nuclease activity, i.e. nuclease domain, are absent from the DNA binding protein. Other modifications to inactivate nuclease activity will be readily apparent to one of skill in the art based on the present disclosure. Accordingly, a nuclease-null DNA binding protein includes polypeptide sequences modified to inactivate nuclease activity or removal of a polypeptide sequence or sequences to
 20 inactivate nuclease activity. The nuclease-null DNA binding protein retains the ability to bind to DNA even though the nuclease activity has been inactivated. Accordingly, the DNA binding protein includes the polypeptide sequence or sequences required for DNA binding but may lack the one or more or all of the nuclease sequences exhibiting nuclease activity. Accordingly, the DNA binding protein includes the polypeptide sequence or sequences required for DNA binding but may
 25 have one or more or all of the nuclease sequences exhibiting nuclease activity inactivated.

According to one aspect, a DNA binding protein having two or more nuclease domains may be modified or altered to inactivate all but one of the nuclease domains. Such a modified or altered DNA binding protein is referred to as a DNA binding protein nickase, to the extent that the DNA binding protein cuts or nicks only one strand of double stranded DNA. When guided by
 30 RNA to DNA, the DNA binding protein nickase is referred to as an RNA guided DNA binding protein nickase.

An exemplary DNA binding protein is an RNA guided DNA binding protein of a Type II CRISPR System which lacks nuclease activity. An exemplary DNA binding protein is a nuclease-null Cas9 protein. An exemplary DNA binding protein is a Cas9 protein nickase.

35 In *S. pyogenes*, Cas9 generates a blunt-ended double-stranded break 3bp upstream of the protospacer-adjacent motif (PAM) via a process mediated by two catalytic domains in the protein:

an HNH domain that cleaves the complementary strand of the DNA and a RuvC-like domain that cleaves the non-complementary strand. See Jinke et al., *Science* 337, 816-821 (2012) hereby incorporated by reference in its entirety. Cas9 proteins are known to exist in many Type II CRISPR systems including the following as identified in the supplementary information to

5 Makarova et al., *Nature Reviews, Microbiology*, Vol. 9, June 2011, pp. 467-477: *Methanococcus maripaludis* C7; *Corynebacterium diphtheriae*; *Corynebacterium efficiens* YS-314; *Corynebacterium glutamicum* ATCC 13032 Kitasato; *Corynebacterium glutamicum* ATCC 13032 Bielefeld; *Corynebacterium glutamicum* R; *Corynebacterium kroppenstedtii* DSM 44385; *Mycobacterium abscessus* ATCC 19977; *Nocardia farcinica* IFM10152; *Rhodococcus erythropolis*

10 PR4; *Rhodococcus jostii* RHA1; *Rhodococcus opacus* B4 uid36573; *Acidothermus cellulolyticus* 11B; *Arthrobacter chlorophenolicus* A6; *Kribbella flavida* DSM 17836 uid43465; *Thermomonospora curvata* DSM 43183; *Bifidobacterium dentium* Bd1; *Bifidobacterium longum* DJO10A; *Slackia heliotrinireducens* DSM 20476; *Persephonella marina* EX H1; *Bacteroides fragilis* NCTC 9434; *Capnocytophaga ochracea* DSM 7271; *Flavobacterium psychrophilum* JIP02

15 86; *Akkermansia muciniphila* ATCC BAA 835; *Roseiflexus castenholzii* DSM 13941; *Roseiflexus* RS1; *Synechocystis* PCC6803; *Elusimicrobium minutum* Pei191; uncultured Termite group 1 bacterium phylotype Rs D17; *Fibrobacter succinogenes* S85; *Bacillus cereus* ATCC 10987; *Listeria innocua*; *Lactobacillus casei*; *Lactobacillus rhamnosus* GG; *Lactobacillus salivarius* UCC118; *Streptococcus agalactiae* A909; *Streptococcus agalactiae* NEM316; *Streptococcus*

20 *agalactiae* 2603; *Streptococcus dysgalactiae equisimilis* GGS 124; *Streptococcus equi zooepidemicus* MGCS10565; *Streptococcus gallolyticus* UCN34 uid46061; *Streptococcus gordonii* Challis subst CH1; *Streptococcus mutans* NN2025 uid46353; *Streptococcus mutans*; *Streptococcus pyogenes* M1 GAS; *Streptococcus pyogenes* MGAS5005; *Streptococcus pyogenes* MGAS2096; *Streptococcus pyogenes* MGAS9429; *Streptococcus pyogenes* MGAS10270; *Streptococcus*

25 *pyogenes* MGAS6180; *Streptococcus pyogenes* MGAS315; *Streptococcus pyogenes* SSI-1; *Streptococcus pyogenes* MGAS10750; *Streptococcus pyogenes* NZ131; *Streptococcus thermophiles* CNRZ1066; *Streptococcus thermophiles* LMD-9; *Streptococcus thermophiles* LMG 18311; *Clostridium botulinum* A3 Loch Maree; *Clostridium botulinum* B Eklund 17B; *Clostridium botulinum* Ba4 657; *Clostridium botulinum* F Langeland; *Clostridium cellulolyticum* H10;

30 *Finegoldia magna* ATCC 29328; *Eubacterium rectale* ATCC 33656; *Mycoplasma gallisepticum*; *Mycoplasma mobile* 163K; *Mycoplasma penetrans*; *Mycoplasma synoviae* 53; *Streptobacillus moniliformis* DSM 12112; *Bradyrhizobium* BTAi1; *Nitrobacter hamburgensis* X14; *Rhodopseudomonas palustris* BisB18; *Rhodopseudomonas palustris* BisB5; *Parvibaculum lavamentivorans* DS-1; *Dinoroseobacter shibae* DFL 12; *Gluconacetobacter diazotrophicus* Pal 5

35 FAPERJ; *Gluconacetobacter diazotrophicus* Pal 5 JGI; *Azospirillum* B510 uid46085; *Rhodospirillum rubrum* ATCC 11170; *Diaphorobacter* TPSY uid29975; *Verminephrobacter*

eiseniae EF01-2; *Neisseria meningitides* 053442; *Neisseria meningitides* alpha14; *Neisseria meningitides* Z2491; *Desulfovibrio salexigens* DSM 2638; *Campylobacter jejuni* doylei 269 97; *Campylobacter jejuni* 81116; *Campylobacter jejuni*; *Campylobacter lari* RM2100; *Helicobacter hepaticus*; *Wolinella succinogenes*; *Tolomonas auensis* DSM 9187; *Pseudoalteromonas atlantica* T6c; *Shewanella pealeana* ATCC 700345; *Legionella pneumophila* Paris; *Actinobacillus succinogenes* 130Z; *Pasteurella multocida*; *Francisella tularensis* novicida U112; *Francisella tularensis* holarctica; *Francisella tularensis* FSC 198; *Francisella tularensis* tularensis; *Francisella tularensis* WY96-3418; and *Treponema denticola* ATCC 35405. Accordingly, aspects of the present disclosure are directed to a Cas9 protein present in a Type II CRISPR system, which has been rendered nuclease null or which has been rendered a nickase as described herein.

The Cas9 protein may be referred by one of skill in the art in the literature as Csn1. The *S. pyogenes* Cas9 protein sequence that is the subject of experiments described herein is shown below. See Deltcheva et al., *Nature* 471, 602-607 (2011) hereby incorporated by reference in its entirety.

MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAE
 ATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHERHPIFG
 NIVDEVAYHEKYPTIYHLRKKLV DSTDKADLRLLYLA LAHMIKFRGHFLIEGDLNPDNSD
 VDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLLENLIAQLPGEKKNGLFGN
 LIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
 LLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA
 GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELH
 AILRRQEDFY PFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEE
 VVDKGASASQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFL
 SGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
 IKDKDFLDNEENEDILEDIVLTTLTFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG
 RLSRKLINGIRDKQSGKTILDFLKSDGFANRNF MQLIHDDSLTFKEDIQKAQVSGQGDSL
 HEHLANLAGSPAIKKGILQTVKVVDDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRER
 MKRIE EGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDH
 IVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
 TKAERGGLSELDKAGFIKRQLVETROITKHVAQILDSRMNTKYDENDKLIREVKVITLKS
 KLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVG TALIKKYPKLESEFVYGDYKVYDVR
 K
 MIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDF
 ATVRKVLSMPQVNIKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVA
 YSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIKLPK
 YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLGSPEDNEQKQLFVE

QHKHYLDEHIEQISEFSKRVLADANLDKVL SAYNKH RDKPIREQAENIIHLFTLTNLGA
 PAAFKYFDITIDRKRYTSTKEVL DATLIHQ SITGLYETRIDLSQLGGD-

According to certain aspects of methods of RNA-guided genome regulation described
 5 herein, Cas9 is altered to reduce, substantially reduce or eliminate nuclease activity. According to
 one aspect, Cas9 nuclease activity is reduced, substantially reduced or eliminated by altering the
 RuvC nuclease domain or the HNH nuclease domain. According to one aspect, the RuvC nuclease
 domain is inactivated. According to one aspect, the HNH nuclease domain is inactivated.
 According to one aspect, the RuvC nuclease domain and the HNH nuclease domain are inactivated.
 10 According to an additional aspect, Cas9 proteins are provided where the RuvC nuclease domain
 and the HNH nuclease domain are inactivated. According to an additional aspect, nuclease-null
 Cas9 proteins are provided insofar as the RuvC nuclease domain and the HNH nuclease domain are
 inactivated. According to an additional aspect, a Cas9 nickase is provided where either the RuvC
 nuclease domain or the HNH nuclease domain is inactivated, thereby leaving the remaining
 15 nuclease domain active for nuclease activity. In this manner, only one strand of the double
 stranded DNA is cut or nicked.

According to an additional aspect, nuclease-null Cas9 proteins are provided where one or
 more amino acids in Cas9 are altered or otherwise removed to provide nuclease-null Cas9 proteins.
 According to one aspect, the amino acids include D10 and H840. See Jinke et al., *Science* 337,
 20 816-821 (2012). According to an additional aspect, the amino acids include D839 and N863.
 According to one aspect, one or more or all of D10, H840, D839 and H863 are substituted with an
 amino acid which reduces, substantially eliminates or eliminates nuclease activity. According to
 one aspect, one or more or all of D10, H840, D839 and H863 are substituted with alanine.
 According to one aspect, a Cas9 protein having one or more or all of D10, H840, D839 and H863
 25 substituted with an amino acid which reduces, substantially eliminates or eliminates nuclease
 activity, such as alanine, is referred to as a nuclease-null Cas9 or Cas9N and exhibits reduced or
 eliminated nuclease activity, or nuclease activity is absent or substantially absent within levels of
 detection. According to this aspect, nuclease activity for a Cas9N may be undetectable using
 known assays, i.e. below the level of detection of known assays.

30 According to one aspect, the nuclease null Cas9 protein includes homologs and orthologs
 thereof which retain the ability of the protein to bind to the DNA and be guided by the RNA.
 According to one aspect, the nuclease null Cas9 protein includes the sequence as set forth for
 naturally occurring Cas9 from *S. pyogenes* and having one or more or all of D10, H840, D839 and
 H863 substituted with alanine and protein sequences having at least 30%, 40%, 50%, 60%, 70%,
 35 80%, 90%, 95%, 98% or 99% homology thereto and being a DNA binding protein, such as an
 RNA guided DNA binding protein.

According to one aspect, the nuclease null Cas9 protein includes the sequence as set forth for naturally occurring Cas9 from *S. pyogenes* excepting the protein sequence of the RuvC nuclease domain and the HNH nuclease domain and also protein sequences having at least 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 98% or 99% homology thereto and being a DNA binding protein, such as an RNA guided DNA binding protein. In this manner, aspects of the present disclosure include the protein sequence responsible for DNA binding, for example, for co-localizing with guide RNA and binding to DNA and protein sequences homologous thereto, and need not include the protein sequences for the RuvC nuclease domain and the HNH nuclease domain (to the extent not needed for DNA binding), as these domains may be either inactivated or removed from the protein sequence of the naturally occurring Cas9 protein to produce a nuclease null Cas9 protein.

For purposes of the present disclosure, Figure 4A depicts metal coordinating residues in known protein structures with homology to Cas9. Residues are labeled based on position in Cas9 sequence. *Left*: RuvC structure, PDB ID: 4EP4 (blue) position D7, which corresponds to D10 in the Cas9 sequence, is highlighted in a Mg-ion coordinating position. *Middle*: Structures of HNH endonuclease domains from PDB IDs: 3M7K (orange) and 4H9D (cyan) including a coordinated Mg-ion (gray sphere) and DNA from 3M7K (purple). Residues D92 and N113 in 3M7K and 4H9D positions D53 and N77, which have sequence homology to Cas9 amino acids D839 and N863, are shown as sticks. *Right*: List of mutants made and analyzed for nuclease activity: Cas9 wildtype; Cas9_{m1} which substitutes alanine for D10; Cas9_{m2} which substitutes alanine for D10 and alanine for H840; Cas9_{m3} which substitutes alanine for D10, alanine for H840, and alanine for D839; and Cas9_{m4} which substitutes alanine for D10, alanine for H840, alanine for D839, and alanine for N863.

As shown in Figure 4B, the Cas9 mutants: m3 and m4, and also their respective fusions with VP64 showed undetectable nuclease activity upon deep sequencing at targeted loci. The plots show the mutation frequency versus genomic position, with the red lines demarcating the gRNA target. Figure 4C is a higher-resolution examination of the data in Figure 4B and confirms that the mutation landscape shows comparable profile as unmodified loci.

According to one aspect, an engineered Cas9-gRNA system is provided which enables RNA-guided genome regulation in human cells by tethering transcriptional activation domains to either a nuclease-null Cas9 or to guide RNAs. According to one aspect of the present disclosure, one or more transcriptional regulatory proteins or domains (such terms are used interchangeably) are joined or otherwise connected to a nuclease-deficient Cas9 or one or more guide RNA (gRNA). The transcriptional regulatory domains correspond to targeted loci. Accordingly, aspects of the present disclosure include methods and materials for localizing transcriptional regulatory domains to targeted loci by fusing, connecting or joining such domains to either Cas9N or to the gRNA.

According to one aspect, a Cas9N-fusion protein capable of transcriptional activation is provided. According to one aspect, a VP64 activation domain (see Zhang et al., *Nature Biotechnology* 29, 149-153 (2011) hereby incorporated by reference in its entirety) is joined, fused, connected or otherwise tethered to the C terminus of Cas9N. According to one method, the transcriptional regulatory domain is provided to the site of target genomic DNA by the Cas9N protein. According to one method, a Cas9N fused to a transcriptional regulatory domain is provided within a cell along with one or more guide RNAs. The Cas9N with the transcriptional regulatory domain fused thereto bind at or near target genomic DNA. The one or more guide RNAs bind at or near target genomic DNA. The transcriptional regulatory domain regulates expression of the target gene. According to a specific aspect, a Cas9N-VP64 fusion activated transcription of reporter constructs when combined with gRNAs targeting sequences near the promoter, thereby displaying RNA-guided transcriptional activation.

According to one aspect, a gRNA-fusion protein capable of transcriptional activation is provided. According to one aspect, a VP64 activation domain is joined, fused, connected or otherwise tethered to the gRNA. According to one method, the transcriptional regulatory domain is provided to the site of target genomic DNA by the gRNA. According to one method, a gRNA fused to a transcriptional regulatory domain is provided within a cell along with a Cas9N protein. The Cas9N binds at or near target genomic DNA. The one or more guide RNAs with the transcriptional regulatory protein or domain fused thereto bind at or near target genomic DNA. The transcriptional regulatory domain regulates expression of the target gene. According to a specific aspect, a Cas9N protein and a gRNA fused with a transcriptional regulatory domain activated transcription of reporter constructs, thereby displaying RNA-guided transcriptional activation.

The gRNA tethers capable of transcriptional regulation were constructed by identifying which regions of the gRNA will tolerate modifications by inserting random sequences into the gRNA and assaying for Cas9 function. gRNAs bearing random sequence insertions at either the 5' end of the crRNA portion or the 3' end of the tracrRNA portion of a chimeric gRNA retain functionality, while insertions into the tracrRNA scaffold portion of the chimeric gRNA result in loss of function. See Figure 5A-B summarizing gRNA flexibility to random base insertions. Figure 5A is a schematic of a homologous recombination (HR) assay to determine Cas9-gRNA activity. As shown in Figure 5B, gRNAs bearing random sequence insertions at either the 5' end of the crRNA portion or the 3' end of the tracrRNA portion of a chimeric gRNA retain functionality, while insertions into the tracrRNA scaffold portion of the chimeric gRNA result in loss of function. The points of insertion in the gRNA sequence are indicated by red nucleotides. Without wishing to be bound by scientific theory, the increased activity upon random base insertions at the 5' end may be due to increased half-life of the longer gRNA.

To attach VP64 to the gRNA, two copies of the MS2 bacteriophage coat-protein binding RNA stem-loop were appended to the 3' end of the gRNA. See Fusco et al., *Current Biology*: CB13, 161-167 (2003) hereby incorporated by reference in its entirety. These chimeric gRNAs were expressed together with Cas9N and MS2-VP64 fusion protein. Sequence-specific transcriptional activation from reporter constructs was observed in the presence of all 3 components.

Figure 1A is a schematic of RNA-guided transcriptional activation. As shown in Figure 1A, to generate a Cas9N-fusion protein capable of transcriptional activation, the VP64 activation domain was directly tethered to the C terminus of Cas9N. As shown in Figure 1B, to generate gRNA tethers capable of transcriptional activation, two copies of the MS2 bacteriophage coat-protein binding RNA stem-loop were appended to the 3' end of the gRNA. These chimeric gRNAs were expressed together with Cas9N and MS2-VP64 fusion protein. Figure 1C shows design of reporter constructs used to assay transcriptional activation. The two reporters bear distinct gRNA target sites, and share a control TALE-TF target site. As shown in Figure 1D, Cas9N-VP64 fusions display RNA-guided transcriptional activation as assayed by both fluorescence-activated cell sorting (FACS) and immunofluorescence assays (IF). Specifically, while the control TALE-TF activated both reporters, the Cas9N-VP64 fusion activates reporters in a gRNA sequence specific manner. As shown in Figure 1E, gRNA sequence-specific transcriptional activation from reporter constructs only in the presence of all 3 components: Cas9N, MS2-VP64 and gRNA bearing the appropriate MS2 aptamer binding sites was observed by both FACS and IF.

According to certain aspects, methods are provided for regulating endogenous genes using Cas9N, one or more gRNAs and a transcriptional regulatory protein or domain. According to one aspect, an endogenous gene can be any desired gene, referred to herein as a target gene. According to one exemplary aspect, genes target for regulation included ZFP42 (REX1) and POU5F1 (OCT4), which are both tightly regulated genes involved in maintenance of pluripotency. As shown in Figure 1F, 10 gRNAs targeting a ~5kb stretch of DNA upstream of the transcription start site (DNase hypersensitive sites are highlighted in green) were designed for the REX1 gene. Transcriptional activation was assayed using either a promoter-luciferase reporter construct (see Takahashi et al., *Cell* 131 861-872 (2007) hereby incorporated by reference in its entirety) or directly via qPCR of the endogenous genes.

Figure 6A-C is directed to RNA-guided OCT4 regulation using Cas9N-VP64. As shown in Figure 6A, 21 gRNAs targeting a ~5kb stretch of DNA upstream of the transcription start site were designed for the OCT4 gene. The DNase hypersensitive sites are highlighted in green. Figure 6B shows transcriptional activation using a promoter-luciferase reporter construct. Figure 6C shows transcriptional activation directly via qPCR of the endogenous genes. While introduction of

individual gRNAs modestly stimulated transcription, multiple gRNAs acted synergistically to stimulate robust multi-fold transcriptional activation.

Figure 7A-C is directed to RNA-guided REX1 regulation using Cas9N, MS2-VP64 and gRNA+2X-MS2 aptamers. As shown in Figure 7A, 10 gRNAs targeting a ~5kb stretch of DNA upstream of the transcription start site were designed for the REX1 gene. The DNase hypersensitive sites are highlighted in green. Figure 7B shows transcriptional activation using a promoter-luciferase reporter construct. Figure 7C shows transcriptional activation directly via qPCR of the endogenous genes. While introduction of individual gRNAs modestly stimulated transcription, multiple gRNAs acted synergistically to stimulate robust multi-fold transcriptional activation. In one aspect, the absence of the 2X-MS2 aptamers on the gRNA does not result in transcriptional activation. See Maeder et al., *Nature Methods* 10, 243-245 (2013) and Perez-Pinera et al., *Nature Methods* 10, 239-242 (2013) each of which are hereby incorporated by reference in its entirety.

Accordingly, methods are directed to the use of multiple guide RNAs with a Cas9N protein and a transcriptional regulatory protein or domain to regulate expression of a target gene.

Both the Cas9 and gRNA tethering approaches were effective, with the former displaying ~1.5–2 fold higher potency. This difference is likely due to the requirement for 2-component as opposed to 3-component complex assembly. However, the gRNA tethering approach in principle enables different effector domains to be recruited by distinct gRNAs so long as each gRNA uses a different RNA-protein interaction pair. See Karyer-Bibens et al., *Biology of the Cell / Under the Auspices of the European Cell Biology Organization* 100, 125-138 (2008) hereby incorporated by reference in its entirety. According to one aspect of the present disclosure, different target genes may be regulated using specific guide RNA and a generic Cas9N protein, i.e. the same or a similar Cas9N protein for different target genes. According to one aspect, methods of multiplex gene regulation are provided using the same or similar Cas9N.

Methods of the present disclosure are also directed to editing target genes using the Cas9N proteins and guide RNAs described herein to provide multiplex genetic and epigenetic engineering of human cells. With Cas9-gRNA targeting being an issue (see Jiang et al., *Nature Biotechnology* 31, 233-239 (2013) hereby incorporated by reference in its entirety), methods are provided for in-depth interrogation of Cas9 affinity for a very large space of target sequence variations. Accordingly, aspects of the present disclosure provide direct high-throughput readout of Cas9 targeting in human cells, while avoiding complications introduced by dsDNA cut toxicity and mutagenic repair incurred by specificity testing with native nuclease-active Cas9.

Further aspects of the present disclosure are directed to the use of DNA binding proteins or systems in general for the transcriptional regulation of a target gene. One of skill in the art will

readily identify exemplary DNA binding systems based on the present disclosure. Such DNA binding systems need not have any nuclease activity, as with the naturally occurring Cas9 protein. Accordingly, such DNA binding systems need not have nuclease activity inactivated. One exemplary DNA binding system is TALE. As a genome editing tool, usually TALE-FokI dimers are used, and for genome regulation TALE-VP64 fusions have been shown to be highly effective. According to one aspect, TALE specificity was evaluated using the methodology shown in Figure 2A. A construct library in which each element of the library comprises a minimal promoter driving a dTomato fluorescent protein is designed. Downstream of the transcription start site m, a 24bp (A/C/G) random transcript tag is inserted, while two TF binding sites are placed upstream of the promoter: one is a constant DNA sequence shared by all library elements, and the second is a variable feature that bears a 'biased' library of binding sites which are engineered to span a large collection of sequences that present many combinations of mutations away from the target sequence the programmable DNA targeting complex was designed to bind. This is achieved using degenerate oligonucleotides engineered to bear nucleotide frequencies at each position such that the target sequence nucleotide appears at a 79% frequency and each other nucleotide occurs at 7% frequency. See Patwardhan et al., *Nature Biotechnology* 30, 265-270 (2012) hereby incorporated by reference in its entirety. The reporter library is then sequenced to reveal the associations between the 24bp dTomato transcript tags and their corresponding 'biased' target site in the library element. The large diversity of the transcript tags assures that sharing of tags between different targets will be extremely rare, while the biased construction of the target sequences means that sites with few mutations will be associated with more tags than sites with more mutations. Next, transcription of the dTomato reporter genes is stimulated with either a control-TF engineered to bind the shared DNA site, or the target-TF that was engineered to bind the target site. The abundance of each expressed transcript tag is measured in each sample by conducting RNAseq on the stimulated cells, which is then mapped back to their corresponding binding sites using the association table established earlier. The control-TF is expected to excite all library members equally since its binding site is shared across all library elements, while the target-TF is expected to skew the distribution of the expressed members to those that are preferentially targeted by it. This assumption is used in step 5 to compute a normalized expression level for each binding site by dividing the tag counts obtained for the target-TF by those obtained for the control-TF.

As shown in Figure 2B, the targeting landscape of a Cas9-gRNA complex reveals that it is on average tolerant to 1-3 mutations in its target sequences. As shown in Figure 2C, the Cas9-gRNA complex is also largely insensitive to point mutations, except those localized to the PAM sequence. Notably this data reveals that the predicted PAM for the *S. pyogenes* Cas9 is not just NGG but also NAG. As shown in Figure 2D, introduction of 2 base mismatches significantly impairs the Cas9-gRNA complex activity, however only when these are localized to the 8-10 bases

nearer the 3' end of the gRNA target sequence (in the heat plot the target sequence positions are labeled from 1-23 starting from the 5' end).

The mutational tolerance of another widely used genome editing tool, TALE domains, was determined using the transcriptional specificity assay described herein. As shown in Figure 2E, the TALE off-targeting data for an 18-mer TALE reveals that it can tolerate on average 1-2 mutations in its target sequence, and fails to activate a large majority of 3 base mismatch variants in its targets. As shown in Figure 2F, the 18-mer TALE is, similar to the Cas9-gRNA complexes, largely insensitive to single base mismatched in its target. As shown in Figure 2G, introduction of 2 base mismatches significantly impairs the 18-mer TALE activity. TALE activity is more sensitive to mismatches nearer the 5' end of its target sequence (in the heat plot the target sequence positions are labeled from 1-18 starting from the 5' end).

Results were confirmed using targeted experiments in a nuclease assay which is the subject of Figure 10A-C directed to evaluating the landscape of targeting by TALEs of different sizes. As shown in Figure 10A, using a nuclease mediated HR assay, it was confirmed that 18-mer TALEs tolerate multiple mutations in their target sequences. As shown in Figure 10B, using the approach described in Fig. 2, the targeting landscape of TALEs of 3 different sizes (18-mer, 14-mer and 10-mer) was analyzed. Shorter TALEs (14-mer and 10-mer) are progressively more specific in their targeting but also reduced in activity by nearly an order of magnitude. As shown in Figure 10C and 10D, 10-mer TALEs show near single-base mismatch resolution, losing almost all activity against targets bearing 2 mismatches (in the heat plot the target sequence positions are labeled from 1-10 starting from the 5' end). Taken together, these data imply that engineering shorter TALEs can yield higher specificity in genome engineering applications, while the requirement for FokI dimerization in TALE nuclease applications is essential to avoid off-target effect. See Kim et al., *Proceedings of the National Academy of Sciences of the United States of America* 93, 1156-1160 (1996) and Pattanayak et al., *Nature Methods* 8, 765-770 (2011) each of which are hereby incorporated by reference in its entirety.

Figure 8A-C is directed to high level specificity analysis processing flow for calculation of normalized expression levels illustrated with examples from experimental data. As shown in Figure 8A, construct libraries are generated with a biased distribution of binding site sequences and random sequence 24bp tags that will be incorporated into reporter gene transcripts (top). The transcribed tags are highly degenerate so that they should map many-to-one to Cas9 or TALE binding sequences. The construct libraries are sequenced (3rd level, left) to establish which tags co-occur with binding sites, resulting in an association table of binding sites vs. transcribed tags (4th level, left). Multiple construct libraries built for different binding sites may be sequenced at once using library barcodes (indicated here by the light blue and light yellow colors; levels 1-4, left). A construct library is then transfected into a cell population and a set of different Cas9/gRNA or

TALE transcription factors are induced in samples of the populations (2nd level, right). One sample is always induced with a fixed TALE activator targeted to a fixed binding site sequence within the construct (top level, green box); this sample serves as a positive control (green sample, also indicated by a + sign). cDNAs generated from the reporter mRNA molecules in the induced samples are then sequenced and analyzed to obtain tag counts for each tag in a sample (3rd and 4th level, right). As with the construct library sequencing, multiple samples, including the positive control, are sequenced and analyzed together by appending sample barcodes. Here the light red color indicates one non-control sample that has been sequenced and analyzed with the positive control (green). Because only the transcribed tags and not the construct binding sites appear in each read, the binding site vs. tag association table obtained from construct library sequencing is then used to tally up total counts of tags expressed from each binding site in each sample (5th level). The tallies for each non-positive control sample are then converted to normalized expression levels for each binding site by dividing them by the tallies obtained in the positive control sample. Examples of plots of normalized expression levels by numbers of mismatches are provided in Figures 2B and 2E, and in Figure 9A and Figure 10B. Not covered in this overall process flow are several levels of filtering for erroneous tags, for tags not associable with a construct library, and for tags apparently shared with multiple binding sites. Figure 8B depicts example distributions of percentages of binding sites by numbers of mismatches generated within a biased construct library. *Left*: Theoretical distribution. *Right*: Distribution observed from an actual TALE construct library. Figure 8C depicts example distributions of percentages of tag counts aggregated to binding sites by numbers of mismatches. *Left*: Distribution observed from the positive control sample. *Right*: Distribution observed from a sample in which a non-control TALE was induced. As the positive control TALE binds to a fixed site in the construct, the distribution of aggregated tag counts closely reflects the distribution of binding sites in Figure 8B, while the distribution is skewed to the left for the non-control TALE sample because sites with fewer mismatches induce higher expression levels. *Below*: Computing the relative enrichment between these by dividing the tag counts obtained for the target-TF by those obtained for the control-TF reveals the average expression level versus the number of mutations in the target site.

These results are further reaffirmed by specificity data generated using a different Cas9-gRNA complex. As shown in Figure 9A, a different Cas9-gRNA complex is tolerant to 1-3 mutations in its target sequence. As shown in Figure 9B, the Cas9-gRNA complex is also largely insensitive to point mutations, except those localized to the PAM sequence. As shown in Figure 9C, introduction of 2 base mismatches however significantly impairs activity (in the heat plot the target sequence positions are labeled from 1-23 starting from the 5' end). As shown in Figure 9D, it

was confirmed using a nuclease mediated HR assay that the predicted PAM for the *S. pyogenes* Cas9 is NGG and also NAG.

According to certain aspects, binding specificity is increased according to methods described herein. Because synergy between multiple complexes is a factor in target gene activation by Cas9N-VP64, transcriptional regulation applications of Cas9N is naturally quite specific as individual off-target binding events should have minimal effect. According to one aspect, off-set nicks are used in methods of genome-editing. A large majority of nicks seldom result in NHEJ events, (see Certo et al., *Nature Methods* 8, 671-676 (2011) hereby incorporated by reference in its entirety) thus minimizing the effects of off-target nicking. In contrast, inducing off-set nicks to generate double stranded breaks (DSBs) is highly effective at inducing gene disruption. According to certain aspects, 5' overhangs generate more significant NHEJ events as opposed to 3' overhangs. Similarly, 3' overhangs favor HR over NHEJ events, although the total number of HR events is significantly lower than when a 5' overhang is generated. Accordingly, methods are provided for using nicks for homologous recombination and off-set nicks for generating double stranded breaks to minimize the effects of off-target Cas9-gRNA activity.

Figure 3A-C is directed to multiplex off-set nicking and methods for reducing the off-target binding with the guide RNAs. As shown in Figure 3A, the traffic light reporter was used to simultaneously assay for HR and NHEJ events upon introduction of targeted nicks or breaks. DNA cleavage events resolved through the HDR pathway restore the GFP sequence, whereas mutagenic NHEJ causes frameshifts rendering the GFP out of frame and the downstream mCherry sequence in frame. For the assay, 14 gRNAs covering a 200bp stretch of DNA: 7 targeting the sense strand (U1-7) and 7 the antisense strand (D1-7) were designed. Using the Cas9D10A mutant, which nicks the complementary strand, different two-way combinations of the gRNAs were used to induce a range of programmed 5' or 3' overhangs (the nicking sites for the 14 gRNAs are indicated). As shown in Figure 3B, inducing off-set nicks to generate double stranded breaks (DSBs) is highly effective at inducing gene disruption. Notably off-set nicks leading to 5' overhangs result in more NHEJ events as opposed to 3' overhangs. As shown in Figure 3C, generating 3' overhangs also favors the ratio of HR over NHEJ events, but the total number of HR events is significantly lower than when a 5' overhang is generated.

Figure 11A-B is directed to Cas9D10A nickase mediated NHEJ. As shown in Figure 11A, the traffic light reporter was used to assay NHEJ events upon introduction of targeted nicks or double-stranded breaks. Briefly, upon introduction of DNA cleavage events, if the break goes through mutagenic NHEJ, the GFP is translated out of frame and the downstream mCherry sequences are rendered in frame resulting in red fluorescence. 14 gRNAs covering a 200bp stretch of DNA: 7 targeting the sense strand (U1-7) and 7 the antisense strand (D1-7) were designed. As shown in Figure 11B, it was observed that unlike the wild-type Cas9 which results in DSBs and

robust NHEJ across all targets, most nicks (using the Cas9D10A mutant) seldom result in NHEJ events. All 14 sites are located within a contiguous 200bp stretch of DNA and over 10-fold differences in targeting efficiencies were observed.

According to certain aspects, methods are described herein of modulating expression of a target nucleic acid in a cell that include introducing one or more, two or more or a plurality of foreign nucleic acids into the cell. The foreign nucleic acids introduced into the cell encode for a guide RNA or guide RNAs, a nuclease-null Cas9 protein or proteins and a transcriptional regulator protein or domain. Together, a guide RNA, a nuclease-null Cas9 protein and a transcriptional regulator protein or domain are referred to as a co-localization complex as that term is understood by one of skill in the art to the extent that the guide RNA, the nuclease-null Cas9 protein and the transcriptional regulator protein or domain bind to DNA and regulate expression of a target nucleic acid. According to certain additional aspects, the foreign nucleic acids introduced into the cell encode for a guide RNA or guide RNAs and a Cas9 protein nickase. Together, a guide RNA and a Cas9 protein nickase are referred to as a co-localization complex as that term is understood by one of skill in the art to the extent that the guide RNA and the Cas9 protein nickase bind to DNA and nick a target nucleic acid.

Cells according to the present disclosure include any cell into which foreign nucleic acids can be introduced and expressed as described herein. It is to be understood that the basic concepts of the present disclosure described herein are not limited by cell type. Cells according to the present disclosure include eukaryotic cells, prokaryotic cells, animal cells, plant cells, fungal cells, archael cells, eubacterial cells and the like. Cells include eukaryotic cells such as yeast cells, plant cells, and animal cells. Particular cells include mammalian cells. Further, cells include any in which it would be beneficial or desirable to regulate a target nucleic acid. Such cells may include those which are deficient in expression of a particular protein leading to a disease or detrimental condition. Such diseases or detrimental conditions are readily known to those of skill in the art. According to the present disclosure, the nucleic acid responsible for expressing the particular protein may be targeted by the methods described herein and a transcriptional activator resulting in upregulation of the target nucleic acid and corresponding expression of the particular protein. In this manner, the methods described herein provide therapeutic treatment.

Target nucleic acids include any nucleic acid sequence to which a co-localization complex as described herein can be useful to either regulate or nick. Target nucleic acids include genes. For purposes of the present disclosure, DNA, such as double stranded DNA, can include the target nucleic acid and a co-localization complex can bind to or otherwise co-localize with the DNA at or adjacent or near the target nucleic acid and in a manner in which the co-localization complex may

have a desired effect on the target nucleic acid. Such target nucleic acids can include endogenous (or naturally occurring) nucleic acids and exogenous (or foreign) nucleic acids. One of skill based on the present disclosure will readily be able to identify or design guide RNAs and Cas9 proteins which co-localize to a DNA including a target nucleic acid. One of skill will further be able to
5 identify transcriptional regulator proteins or domains which likewise co-localize to a DNA including a target nucleic acid. DNA includes genomic DNA, mitochondrial DNA, viral DNA or exogenous DNA.

Foreign nucleic acids (i.e. those which are not part of a cell's natural nucleic acid composition) may be introduced into a cell using any method known to those skilled in the art for
10 such introduction. Such methods include transfection, transduction, viral transduction, microinjection, lipofection, nucleofection, nanoparticle bombardment, transformation, conjugation and the like. One of skill in the art will readily understand and adapt such methods using readily identifiable literature sources.

Transcriptional regulator proteins or domains which are transcriptional activators include
15 VP16 and VP64 and others readily identifiable by those skilled in the art based on the present disclosure.

Diseases and detrimental conditions are those characterized by abnormal loss of expression of a particular protein. Such diseases or detrimental conditions can be treated by upregulation of the particular protein. Accordingly, methods of treating a disease or detrimental condition are
20 provided where the co-localization complex as described herein associates or otherwise binds to DNA including a target nucleic acid, and the transcriptional activator of the co-localization complex upregulates expression of the target nucleic acid. For example upregulating PRDM16 and other genes promoting brown fat differentiation and increased metabolic uptake can be used to treat metabolic syndrome or obesity. Activating anti-inflammatory genes are useful in
25 autoimmunity and cardiovascular disease. Activating tumor suppressor genes is useful in treating cancer. One of skill in the art will readily identify such diseases and detrimental conditions based on the present disclosure.

The following examples are set forth as being representative of the present disclosure. These examples are not to be construed as limiting the scope of the present disclosure as these and
30 other equivalent embodiments will be apparent in view of the present disclosure, figures and accompanying claims.

EXAMPLE I

Cas9 Mutants

Sequences homologous to Cas9 with known structure were searched to identify candidate mutations in Cas9 that could ablate the natural activity of its RuvC and HNH domains. Using HHpred (world wide website toolkit.tuebingen.mpg.de/hhpred), the full sequence of Cas9 was queried against the full Protein Data Bank (January 2013). This search returned two different HNH endonucleases that had significant sequence homology to the HNH domain of Cas9; PacI and a putative endonuclease (PDB IDs: 3M7K and 4H9D respectively). These proteins were examined to find residues involved in magnesium ion coordination. The corresponding residues were then identified in the sequence alignment to Cas9. Two Mg-coordinating side-chains in each structure were identified that aligned to the same amino acid type in Cas9. They are 3M7K D92 and N113, and 4H9D D53 and N77. These residues corresponded to Cas9 D839 and N863. It was also reported that mutations of PacI residues D92 and N113 to alanine rendered the nuclease catalytically deficient. The Cas9 mutations D839A and N863A were made based on this analysis. Additionally, HHpred also predicts homology between Cas9 and the N-terminus of a *Thermus thermophilus* RuvC (PDB ID: 4EP4). This sequence alignment covers the previously reported mutation D10A which eliminates function of the RuvC domain in Cas9. To confirm this as an appropriate mutation, the metal binding residues were determined as before. In 4EP4, D7 helps to coordinate a magnesium ion. This position has sequence homology corresponding to Cas9 D10, confirming that this mutation helps remove metal binding, and thus catalytic activity from the Cas9 RuvC domain.

EXAMPLE II

Plasmid Construction

The Cas9 mutants were generated using the Quikchange kit (Agilent technologies). The target gRNA expression constructs were either (1) directly ordered as individual gBlocks from IDT and cloned into the pCR-BluntII-TOPO vector (Invitrogen); or (2) custom synthesized by Genewiz; or (3) assembled using Gibson assembly of oligonucleotides into the gRNA cloning vector (plasmid #41824). The vectors for the HR reporter assay involving a broken GFP were constructed by fusion PCR assembly of the GFP sequence bearing the stop codon and appropriate fragment assembled into the EGIP lentivector from Addgene (plasmid #26777). These lentivectors were then used to establish the GFP reporter stable lines. TALENs used in this study were constructed using standard protocols. See Sanjana et al., *Nature Protocols* 7, 171-192 (2012) hereby incorporated by reference in its entirety. Cas9N and MS2 VP64 fusions were performed

using standard PCR fusion protocol procedures. The promoter luciferase constructs for OCT4 and REX1 were obtained from Addgene (plasmid #17221 and plasmid #17222).

EXAMPLE III

Cell culture and Transfections

5 HEK 293T cells were cultured in Dulbecco's modified Eagle's medium (DMEM, Invitrogen) high glucose supplemented with 10% fetal bovine serum (FBS, Invitrogen), penicillin/streptomycin (pen/strep, Invitrogen), and non-essential amino acids (NEAA, Invitrogen). Cells were maintained at 37°C and 5% CO₂ in a humidified incubator.

Transfections involving nuclease assays were as follows: 0.4×10⁶ cells were transfected
10 with 2μg Cas9 plasmid, 2μg gRNA and/or 2μg DNA donor plasmid using Lipofectamine 2000 as per the manufacturer's protocols. Cells were harvested 3 days after transfection and either analyzed by FACS, or for direct assay of genomic cuts the genomic DNA of ~1 X 10⁶ cells was extracted using DNAeasy kit (Qiagen). For these PCR was conducted to amplify the targeting region with genomic DNA derived from the cells and amplicons were deep sequenced by MiSeq Personal
15 Sequencer (Illumina) with coverage >200,000 reads. The sequencing data was analyzed to estimate NHEJ efficiencies.

For transfections involving transcriptional activation assays: 0.4×10⁶ cells were transfected with (1) 2μg Cas9N-VP64 plasmid, 2μg gRNA and/or 0.25μg of reporter construct; or (2) 2μg Cas9N plasmid, 2μg MS2-VP64, 2μg gRNA-2XMS2aptamer and/or 0.25μg of reporter construct.
20 Cells were harvested 24-48hrs post transfection and assayed using FACS or immunofluorescence methods, or their total RNA was extracted and these were subsequently analyzed by RT-PCR. Here standard taqman probes from Invitrogen for OCT4 and REX1 were used, with normalization for each sample performed against GAPDH.

For transfections involving transcriptional activation assays for specificity profile of Cas9-gRNA complexes and TALEs: 0.4×10⁶ cells were transfected with (1) 2μg Cas9N-VP64 plasmid, 2μg gRNA and 0.25μg of reporter library; or (2) 2μg TALE-TF plasmid and 0.25μg of reporter library; or (3) 2μg control-TF plasmid and 0.25μg of reporter library. Cells were harvested 24hrs post transfection (to avoid the stimulation of reporters being in saturation mode). Total RNA extraction was performed using RNAeasy-plus kit (Qiagen), and standard RT-pcr performed using
30 Superscript-III (Invitrogen). Libraries for next-generation sequencing were generated by targeted pcr amplification of the transcript-tags.

EXAMPLE IV

Computational and Sequence Analysis for Calculation of Cas9-TF and TALE-TF Reporter
Expression Levels

The high-level logic flow for this process is depicted in Figure 8A, and additional details
5 are given here. For details on construct library composition, see Figures 8A (level 1) and 8B.

Sequencing: For Cas9 experiments, construct library (Figure 8A, level 3, left) and reporter gene
cDNA sequences (Figure 8A, level 3, right) were obtained as 150bp overlapping paired end reads
on an Illumina MiSeq, while for TALE experiments, corresponding sequences were obtained as
51bp non-overlapping paired end reads on an Illumina HiSeq.

10 Construct library sequence processing: *Alignment:* For Cas9 experiments, novoalign V2.07.17
(world wide website novocraft.com/main/index/php) was used to align paired reads to a set of
250bp reference sequences that corresponded to 234bp of the constructs flanked by the pairs of 8bp
library barcodes (see Figure 8A, 3rd level, left). In the reference sequences supplied to novoalign,
the 23bp degenerate Cas9 binding site regions and the 24bp degenerate transcript tag regions (see
15 Figure 8A, first level) were specified as Ns, while the construct library barcodes were explicitly
provided. For TALE experiments, the same procedures were used except that the reference
sequences were 203bp in length and the degenerate binding site regions were 18bp vs. 23bp in
length. *Validity checking:* Novoalign output for comprised files in which left and right reads for
each read pair were individually aligned to the reference sequences. Only read pairs that were both
20 uniquely aligned to the reference sequence were subjected to additional validity conditions, and
only read pairs that passed all of these conditions were retained. The validity conditions included:
(i) Each of the two construct library barcodes must align in at least 4 positions to a reference
sequence barcode, and the two barcodes must to the barcode pair for the same construct library.
(ii) All bases aligning to the N regions of the reference sequence must be called by novoalign as
25 As, Cs, Gs or Ts. Note that for neither Cas9 nor TALE experiments did left and right reads overlap
in a reference N region, so that the possibility of ambiguous novoalign calls of these N bases did
not arise. (iii) Likewise, no novoalign-called inserts or deletions must appear in these regions. (iv)
No Ts must appear in the transcript tag region (as these random sequences were generated from As,
Cs, and Gs only). Read pairs for which any one of these conditions were violated were collected in
30 a rejected read pair file. These validity checks were implemented using custom perl scripts.

Induced sample reporter gene cDNA sequence processing: *Alignment:* SeqPrep (downloaded from
world wide website github.com/jstjohn/SeqPrep) was first used to merge the overlapping read pairs
to the 79bp common segment, after which novoalign (version above) was used to align these 79bp
common segments as unpaired single reads to a set of reference sequences (see Figure 8A, 3rd
35 level, right) in which (as for the construct library sequencing) the 24bp degenerate transcript tag

was specified as Ns while the sample barcodes were explicitly provided. Both TALE and Cas9 cDNA sequence regions corresponded to the same 63bp regions of cDNA flanked by pairs of 8bp sample barcode sequences. *Validity checking:* The same conditions were applied as for construct library sequencing (see above) except that: (a) Here, due prior SeqPrep merging of read pairs, validity processing did not have to filter for unique alignments of both reads in a read pair but only for unique alignments of the merged reads. (b) Only transcript tags appeared in the cDNA sequence reads, so that validity processing only applied these tag regions of the reference sequences and not also to a separate binding site region.

Assembly of table of binding sites vs. transcript tag associations: Custom perl was used to generate these tables from the validated construct library sequences (Figure 8A, 4th level, left). Although the 24bp tag sequences composed of A, C, and G bases should be essentially unique across a construct library (probability of sharing = $\sim 2.8e-11$), early analysis of binding site vs. tag associations revealed that a non-negligible fraction of tag sequences were in fact shared by multiple binding sequences, likely mainly caused by a combination of sequence errors in the binding sequences, or oligo synthesis errors in the oligos used to generate the construct libraries. In addition to tag sharing, tags found associated with binding sites in validated read pairs might also be found in the construct library read pair reject file if it was not clear, due to barcode mismatches, which construct library they might be from. Finally, the tag sequences themselves might contain sequence errors. To deal with these sources of error, tags were categorized with three attributes: (i) *safe* vs. *unsafe*, where *unsafe* meant the tag could be found in the construct library rejected read pair file; *shared* vs. *nonshared*, where *shared* meant the tag was found associated with multiple binding site sequences, and *2+* vs. *1-only*, where *2+* meant that the tag appeared at least twice among the validated construct library sequences and so presumed to be less likely to contain sequence errors. Combining these three criteria yielded 8 classes of tags associated with each binding site, the most secure (but least abundant) class comprising only *safe*, *nonshared*, *2+* tags; and the least secure (but most abundant) class comprising all tags regardless of safety, sharing, or number of occurrences.

Computation of normalized expression levels: Custom perl code was used to implement the steps indicated in Figure 8A, levels 5-6. First, tag counts obtained for each induced sample were aggregated for each binding site, using the binding site vs. transcript tag table previously computed for the construct library (see Figure 8C). For each sample, the aggregated tag counts for each binding site were then divided by the aggregated tag counts for the positive control sample to generate normalized expression levels. Additional considerations relevant to these calculations included:

1. For each sample, a subset of “novel” tags were found among the validity-checked cDNA gene sequences that could not be found in the binding site *vs.* transcript tag association table. These tags were ignored in the subsequent calculations.

2. The aggregations of tag counts described above were performed for each of the eight classes of tags described above in binding site *vs.* transcript tag association table. Because the binding sites in the construct libraries were biased to generate sequences similar to a central sequence frequently, but sequences with increasing numbers of mismatches increasingly rarely, binding sites with few mismatches generally aggregated to large numbers of tags, while binding sites with more mismatches aggregated to smaller numbers. Thus, although use of the most secure tag class was generally desirable, evaluation of binding sites with two or more mismatches might be based on small numbers of tags per binding site, making the secure counts and ratios less statistically reliable even if the tags themselves were more reliable. In such cases, all tags were used. Some compensation for this consideration obtains from the fact that the number of separate aggregated tag counts for n mismatching positions grew with the number of combinations of mismatching positions (equal to $\binom{4}{n} 3^n$), and so dramatically increases with n ; thus the averages of aggregated tag counts for different numbers n of mismatches (shown in Figs. 2b, 2e, and in Figures 9A and 10B) are based on a statistically very large set of aggregated tag counts for $n \geq 2$.

3. Finally, the binding site built into the TALE construct libraries was 18bp and tag associations were assigned based on these 18bp sequences, but some experiments were conducted with TALEs programmed to bind central 14bp or 10bp regions within the 18bp construct binding site regions. In computing expression levels for these TALEs, tags were aggregated to binding sites based on the corresponding regions of the 18bp binding sites in the association table, so that binding site mismatches outside of this region were ignored.

EXAMPLE V

RNA-guided SOX2 and NANOG Regulation Using Cas9_N-VP64

The sgRNA (aptamer-modified single guide RNA) tethering approach described herein allows different effector domains to be recruited by distinct sgRNAs so long as each sgRNA uses a different RNA-protein interaction pair, enabling multiplex gene regulation using the same Cas9_N-protein. For the Figure 12A SOX2 and Figure 12B NANOG genes, 10 gRNAs were designed targeting a ~1kb stretch of DNA upstream of the transcription start site. The DNase hypersensitive sites are highlighted in green. Transcriptional activation via qPCR of the endogenous genes was assayed. In both instances, while introduction of individual gRNAs modestly stimulated transcription, multiple gRNAs acted synergistically to stimulate robust multi-fold transcriptional

activation. Data are means $\pm$ SEM (N=3). As shown in Figure 12A-B, two additional genes, SOX2 and NANOG, were regulated via sgRNAs targeting within an upstream $\sim$ 1kb stretch of promoter DNA. The sgRNAs proximal to the transcriptional start site resulted in robust gene activation.

5

EXAMPLE VI

Evaluating the Landscape of targeting by Cas9-gRNA Complexes

Using the approach described in Fig. 2, the targeting landscape of two additional Cas9-gRNA complexes (Figure 13A-C) and (Figure 13D-F) was analyzed. The two gRNAs have vastly different specificity profiles with gRNA2 tolerating up to 2-3 mismatches and gRNA3 only up to 1. These aspects are reflected in both the one base mismatch (Figure 13B, 13E) and two base mismatch plots (Figure 13C, 13F). In Figure 13C and 13F, base mismatch pairs for which insufficient data were available to calculate a normalized expression level are indicated as gray boxes containing an 'x', while, to improve data display, mismatch pairs whose normalized expression levels are outliers that exceed the top of the color scale are indicated as yellow boxes containing an asterisk '*'. Statistical significance symbols are: *** for $P < .0005/n$, ** for $P < .005/n$, * for $P < .05/n$, and N.S. (Non-Significant) for $P \geq .05/n$, where n is the number of comparisons (refer Table 2).

10

15

EXAMPLE VII

Validations, Specificity of Reporter Assay

As shown in Figure 14A-C, specificity data was generated using two different sgRNA:Cas9 complexes. It was confirmed that the assay was specific for the sgRNA being evaluated, as a corresponding mutant sgRNA was unable to stimulate the reporter library. Figure 14A: The specificity profile of two gRNAs (wild-type and mutant; sequence differences are highlighted in red) were evaluated using a reporter library designed against the wild-type gRNA target sequence. Figure 14B: It was confirmed that this assay was specific for the gRNA being evaluated (data re-plotted from Fig. 13D), as the corresponding mutant gRNA is unable to stimulate the reporter library. Statistical significance symbols are: *** for $P < .0005/n$, ** for $P < .005/n$, * for $P < .05/n$, and N.S. (Non-Significant) for $P \geq .05/n$, where n is the number of comparisons (refer Table 2). Different sgRNAs can have different specificity profiles (Figures 13A, 13D), specifically, sgRNA2 tolerates up to 3 mismatches and sgRNA3 only up to 1. The greatest sensitivity to mismatches was localized to the 3' end of the spacer, albeit mismatches at other positions were also observed to affect activity.

20

25

30

EXAMPLE VIII

Validations, Single and Double-base gRNA Mismatches

As shown in Figure 15A-D, it was confirmed by targeted experiments that single-base mismatches within 12 bp of the 3' end of the spacer in the assayed sgRNAs resulted in detectable targeting. However, 2 bp mismatches in this region resulted in significant loss of activity. Using a nuclease assay, 2 independent gRNAs were tested: gRNA2 (Figure 15A-B) and gRNA3 (Figure 15C-D) bearing single or double-base mismatches (highlighted in red) in the spacer sequence versus the target. It was confirmed that single-base mismatches within 12bp of the 3' end of the spacer in the assayed gRNAs result in detectable targeting, however 2bp mismatches in this region result in rapid loss of activity. These results further highlight the differences in specificity profiles between different gRNAs consistent with the results in Figure 13. Data are means $\pm$ SEM (N=3).

EXAMPLE IX

Validations, 5' gRNA truncations

As shown in Figure 16A-D, truncations in the 5' portion of the spacer resulted in retention of sgRNA activity. Using a nuclease assay, 2 independent gRNA were tested: gRNA1 (Figure 16A-B) and gRNA3 (Figure 16C-D) bearing truncations at the 5' end of their spacer. It was observed that 1-3bp 5' truncations are well tolerated, but larger deletions lead to loss of activity. Data are means $\pm$ SEM (N=3).

EXAMPLE X

Validations, *S. pyogenes* PAM

As shown in Figure 17A-B, it was confirmed using a nuclease mediated HR assay that the PAM for the *S. pyogenes* Cas9 is NGG and also NAG. Data are means $\pm$ SEM (N=3). According to an additional investigation, a generated set of about 190K Cas9 targets in human exons that had no alternate NGG targets sharing the last 13 nt of the targeting sequence was scanned for the presence of alternate NAG sites or for NGG sites with a mismatch in the prior 13 nt. Only 0.4% were found to have no such alternate targets.

EXAMPLE XI

Validations, TALE Mutations

Using a nuclease mediated HR assay (Figure 18A-B) it was confirmed that 18-mer TALEs tolerate multiple mutations in their target sequences. As shown in Figure 18A-B certain mutations in the

middle of the target lead to higher TALE activity, as determined via targeted experiments in a nuclease assay.

EXAMPLE XII

TALE Monomer Specificity Versus TALE Protein Specificity

5 To decouple the role of individual repeat-variable diresidues (RVDs), it was confirmed that choice of RVDs did contribute to base specificity but TALE specificity is also a function of the binding energy of the protein as a whole. Figure 19A-C shows a comparison of TALE monomer specificity versus TALE protein specificity. Figure 19A: Using a modification of approach described in Fig. 2, the targeting landscape of 2 14-mer TALE-TFs bearing a contiguous set of 6
10 NI or 6 NH repeats was analyzed. In this approach, a reduced library of reporters bearing a degenerate 6-mer sequence in the middle was created and used to assay the TALE-TF specificity. Figure 19B-C: In both instances, it was noted that the expected target sequence is enriched (i.e. one bearing 6 As for NI repeats, and 6 Gs for NH repeats). Each of these TALEs still tolerate 1-2 mismatches in the central 6-mer target sequence. While choice of monomers does contribute to
15 base specificity, TALE specificity is also a function of the binding energy of the protein as a whole. According to one aspect, shorter engineered TALEs or TALEs bearing a composition of high and low affinity monomers result in higher specificity in genome engineering applications and FokI dimerization in nuclease applications allows for further reduction in off-target effects when using shorter TALEs.

20

EXAMPLE XIII

Off-set Nicking, Native Locus

Figure 20A-B shows data related to off-set nicking. In the context of genome-editing, off-set nicks were created to generate DSBs. A large majority of nicks do not result in non-homologous end joining (NHEJ) mediated indels and thus when inducing off-set nicks, off-target single nick events
25 will likely result in very low indel rates. Inducing off-set nicks to generate DSBs is effective at inducing gene disruption at both integrated reporter loci and at the native AAVS1 genomic locus. Figure 20A: The native AAVS1 locus with 8 gRNAs covering a 200bp stretch of DNA was targeted: 4 targeting the sense strand (s1-4) and 4 the antisense strand (as1-4). Using the Cas9D10A mutant, which nicks the complementary strand, different two-way combinations of the
30 gRNAs was used to induce a range of programmed 5' or 3' overhangs. Figure 20B: Using a Sanger sequencing based assay, it was observed that while single gRNAs did not induce detectable NHEJ events, inducing off-set nicks to generate DSBs is highly effective at inducing gene disruption. Notably off-set nicks leading to 5' overhangs result in more NHEJ events as opposed to 3'

overhangs. The number of Sanger sequencing clones is highlighted above the bars, and the predicted overhang lengths are indicated below the corresponding x-axis legends.

EXAMPLE XIV

Off-set Nicking, NHEJ Profiles

- 5 Figure 21A-C is directed to off-set nicking and NHEJ profiles. Representative Sanger sequencing results of three different off-set nicking combinations is shown with positions of the targeting gRNAs highlighted by boxes. Furthermore, consistent with the standard model for homologous recombination (HR) mediated repair, engineering of 5' overhangs via off-set nicks generated more robust NHEJ events than 3' overhangs (Figure 3B). In addition to a stimulation of NHEJ, robust
- 10 induction of HR was observed when the 5' overhangs were created. Generation of 3' overhangs did not result in improvement of HR rates (Figure 3C).

EXAMPLE XV

Table 1

gRNA Targets for Endogenous Gene Regulation

- 15 Targets in the REX1, OCT4, SOX2 and NANOG promoters used in Cas9-gRNA mediated activation experiments are listed.

gRNA Name	gRNA Target
REX1 1	ctggcgggatcaactcgcggtt agg
REX1 2	cctcgycctccaaaagtgcct agg
REX1 3	acgctgattcctgcagatca ggg
REX1 4	ccagggaatacgtatccacca ggg
REX1 5	gccacacccaagcgatcaaa tgg
REX1 6	aaataatacatttotaaggta agg
REX1 7	gctactggggaggctgaggc agg
REX1 8	tagcaatacagtcacattaa tgg
REX1 9	ctcatgtgatcccccgctct cgg
REX1 10	ccgggcagagagtgaacggc cgg
OCT4 1	ttccttccctctctccgctgc tgg
OCT4 2	tctctgcgaagcccctggag agg
OCT4 3	aatgcagttgcgcagtgccag tgg
OCT4 4	ccctcagcctcctaaagtgcct ggg
OCT4 5	gagtcacaaatcctctttact agg
OCT4 6	gagtgctctggatttgaggata agg
OCT4 7	cagcacctcatctcccagtg agg
OCT4 8	tctaaaacccagggaatcat ggg
OCT4 9	cacaagggcagccagggatcc agg
OCT4 10	gatggcaagctgagaaacac tgg
OCT4 11	tgaaatgcacgcatacaatt agg
OCT4 12	ccagtcacagacctggccttc tgg
OCT4 13	cccagaaaaacagaccctga agg
OCT4 14	aagggttgagcacttyttta ggg
OCT4 15	atgtctgagttttgggtcgag agg
OCT4 16	ggccacctgaaggggaagta ggg
OCT4 17	tggcagtcctactcctgaaga tgg
OCT4 18	ggcacagtgccagaggtctg tgg
OCT4 19	taaaaataaaaaaactaaca ggg
OCT4 20	tctytgggggacctgcactg agg
OCT4 21	ggccagaggtcaaggtagt ggg
SOX2 1	cacgacccgaaaacctttctta cgg
SOX2 2	gttgaatgaagacagtcctag tgg
SOX2 3	taagaacagagcaagttacg tgg
SOX2 4	tytaaygtaagayaggyagay cgg
SOX2 5	tyacacaccaaactcctgcac tgg
SOX2 6	tttaccccacttccttcgaaa agg
SOX2 7	gtggctggcaggctggctct ggg
SOX2 8	ctcccccggcctccccgcg cgg
SOX2 9	cbaaacccggcagcgyaggtc ggg
SOX2 10	aggagccgcgcgcgcgtgat tgg
NANOG 1	cacacacacccacacgagat ggg
NANOG 2	gaagaaagctaaagagccaga ggg
NANOG 3	atgagaatttcaataacctc agg
NANOG 4	tcccgctctgttgcccaggc tgg
NANOG 5	cagacaccccaccaccatggy tgg
NANOG 6	tcccaatttactyggattac agg
NANOG 7	tgatttaaaagtttgaaaacg tgg
NANOG 8	tctagtccccacctagtct ggg
NANOG 9	gattaactgagaattcacaa ggg
NANOG 10	cgccaggagggtggtgtcta agg

EXAMPLE XVI

Table 2

5 Summary of Statistical Analysis of Cas9-gRNA and TALE Specificity Data

Table 2(a) P-values for comparisons of normalized expression levels of TALE or Cas9-VP64 activators binding to target sequences with particular numbers of target site mutations. Normalized expression levels have been indicated by boxplots in the figures indicated in the Figure

column, where the boxes represent the distributions of these levels by numbers of mismatches from the target site. P-values were computed using t-tests for each consecutive pair of numbers of mismatches in each boxplot, where the t-tests were either one sample or two sample t-tests (see Methods). Statistical significance was assessed using Bonferroni-corrected P-value thresholds, where the correction was based on the number of comparisons within each boxplot. Statistical significance symbols are: *** for $P < .0005/n$, ** for $P < .005/n$, * for $P < .05/n$, and N.S. (Non-Significant) for $P \geq .05/n$, where n is the number of comparisons. Table 2(b) Statistical characterization of seed region in Figure 2D: $\log_{10}(P\text{-values})$ indicating the degree of separation between expression values for Cas9N VP64+gRNA binding to target sequences with two mutations for those position pairs mutated within candidate seed regions at the 3' end of the 20bp target site vs. all other position pairs. The greatest separation, indicated by the largest $-\log_{10}(P\text{-values})$ (highlighted above), is found in the last 8-9bp of the target site. These positions may be interpreted as indicating the start of the "seed" region of this target site. See the section "Statistical characterization of seed region" in Methods for information on how the P-values were computed.

15

a

Figure	Expression level comparison:		t-test	P-value	Symbol
	mutations vs.	mutations			
2b	0	1	1-samp	7.8E-05	**
	1	2	2-samp	1.4E-06	***
	2	3	2-samp	4.0E-61	***
	3	4	2-samp	0	***
	4	5	2-samp	0	***
	5	6	2-samp	1.0E-217	***
	6	7	2-samp	1.7E-43	***
	7	8	2-samp	3.7E-52	N.S.
2c	0	1	1-samp	8.9E-51	N.S.
	1	2	2-samp	1.9E-56	***
	2	3	2-samp	5.0E-147	***
	3	4	2-samp	0	***
	4	5	2-samp	0	***
	5	6	2-samp	4.2E-63	***
	6	7	2-samp	1.6E-03	*
	7	8	2-samp	4.7E-01	N.S.
S7a	0	1	1-samp	5.2E-02	N.S.
	1	2	2-samp	2.8E-05	***
	2	3	2-samp	5.5E-21	***
	3	4	2-samp	1.4E-58	***
	4	5	2-samp	8.5E-101	***
	5	6	2-samp	6.8E-94	***
	6	7	2-samp	1.8E-61	***
	7	8	2-samp	5.1E-24	***
S7d and S8d	0	1	1-samp	2.3E-18	***
	1	2	2-samp	2.4E-08	***
	2	3	2-samp	6.2E-54	***
	3	4	2-samp	4.0E-141	***
	4	5	2-samp	1.9E-20	***
	5	6	2-samp	1.2E-53	*
	6	7	2-samp	3.8E-55	***
	7	8	2-samp	9.4E-51	N.S.
S8c	0	1	1-samp	7.2E-53	N.S.
	1	2	2-samp	5.0E-51	N.S.
	2	3	2-samp	3.9E-34	***
	3	4	2-samp	8.5E-153	***
	4	5	2-samp	8.6E-76	***
	5	6	2-samp	1.6E-03	*
	6	7	2-samp	7.1E-01	N.S.
	7	8	2-samp	7.8E-02	N.S.
S13a (left)	0	1	1-samp	7.3E-01	N.S.
	1	2	2-samp	2.4E-06	***
	2	3	2-samp	7.2E-140	***
	3	4	2-samp	0	***
	4	5	2-samp	0	***
	5	6	2-samp	1.0E-72	***
	6	7	2-samp	4.0E-03	*
S13a (middle)	0	1	1-samp	9.4E-02	N.S.
	1	2	2-samp	5.2E-09	***
	2	3	2-samp	7.9E-86	***
	3	4	2-samp	2.9E-53	***
	4	5	2-samp	3.5E-10	***
S13a (right)	0	1	1-samp	1.3E-13	***
	1	2	2-samp	1.1E-04	***
	2	3	2-samp	3.7E-08	***

b

seed start position	Number position pairs		-log10 P-value
	both in seed	not both in seed	
2	171	19	3.11
3	153	37	1.46
4	136	54	2.01
5	120	70	3.34
6	105	85	5.65
7	91	99	7.34
8	78	112	6.61
9	66	124	7.10
10	55	135	9.72
11	45	145	9.83
12	36	154	10.44
13	28	162	10.72
14	21	169	8.97
15	15	175	5.61
16	10	180	3.34
17	6	184	2.26
18	3	187	1.16

EXAMPLE XVII

Sequences of Proteins and RNAs in the Examples

A. Sequences of the Cas9_N-VP64 activator constructs based on the m4 mutant are displayed below.

5 Three versions were constructed with the Cas9_{m4}^{VP64} and Cas9_{m4}^{VP64}N fusion protein formats

showing highest activity. Corresponding vectors for the m3 and m2 mutants (Figure 4A) were also constructed (NLS and VP64 domains are highlighted).

>Cas9_{m4}^{VP64}

```

gccaccATGGACAAGAAGTACTCCATTGGGCTCGCTATCGGCACAAACAGCGTCGGCTGG
5  GCCGTCATTACGGACGAGTACAAGGTGCCGAGCAAAAAATTCAAAGTTCTGGGCAAT
  ACCGATCGCCACAGCATAAAGAAGAACCTCATTGGCGCCCTCCTGTTCTGACTCCGGGG
  AGACGGCCGAAGCCACGCGGCTCAAAAGAACAGCACGGCGCAGATATACCCGCAGAA
  AGAATCGGATCTGCTACCTGCAGGAGATCTTTAGTAATGAGATGGCTAAGGTGGATGA
  CTCTTTCTTCCATAGGCTGGAGGAGTCCTTTTTTGGTGGAGGAGGATAAAAAGCACGAG
10  CGCCACCCAATCTTTGGCAATATCGTGGACGAGGTGGCGTACCATGAAAAGTACCCAA
  CCATATATCATCTGAGGAAGAAGCTTGTAGACAGTACTGATAAGGCTGACTTGCGGTT
  GATCTATCTCGCGCTGGCGCATATGATCAAATTTCGGGGACACTTCCTCATCGAGGGG
  GACCTGAACCCAGACAACAGCGATGTCGACAAACTCTTTATCCAACCTGGTTCAGACTT
  ACAATCAGCTTTTTCGAAGAGAACCCGATCAACGCATCCGGAGTTGACGCCAAAGCAA
15  TCCTGAGCGCTAGGCTGTCCAAATCCCGGCGGCTCGAAAACCTCATCGCACAGCTCCC
  TGGGGAGAAGAAGAACGGCCTGTTTGGTAATCTTATCGCCCTGTCACTCGGGCTGACC
  CCAACTTTAAATCTAACTTCGACCTGGCCGAAGATGCCAAGCTTCAACTGAGCAAAG
  ACACCTACGATGATGATCTCGACAATCTGCTGGCCCAGATCGGCGACCAGTACGCAGA
  CCTTTTTTTTGGCGGCAAAGAACCTGTCAGACGCCATTCTGCTGAGTGATATTCTGCGAG
20  TGAACACGGAGATCACCAAAGCTCCGCTGAGCGCTAGTATGATCAAGCGCTATGATG
  AGCACCACCAAGACTTGACTTTGCTGAAGGCCCTTGTCAGACAGCAACTGCCTGAGAA
  GTACAAGGAAATTTTCTTCGATCAGTCTAAAAATGGCTACGCCGGATACATTGACGGC
  GGAGCAAGCCAGGAGGAATTTTACAAATTTATTAAGCCCATCTTGGAAAAAATGGAC
  GGCACCGAGGAGCTGCTGGTAAAGCTTAACAGAGAAGATCTGTTGCGCAAACAGCGC
25  ACTTTCGACAATGGAAGCATCCCCCACCAGATTCACCTGGGCGAACTGCACGCTATCC
  TCAGGCGGCAAGAGGATTTCTACCCCTTTTTTGAAAGATAACAGGGAAAAGATTGAGA
  AAATCCTCACATTTTCGGATACCCTACTATGTAGGCCCCCTCGCCCGGGGAAATTCCAG
  ATTCGCGTGGATGACTCGCAAATCAGAAGAGACCATCACTCCCTGGAACCTTCGAGGAA
  GTCGTGGATAAGGGGGCCTCTGCCCAGTCCTTCATCGAAAGGATGACTAACTTTGATA
30  AAAATCTGCCTAACGAAAAGGTGCTTCCTAAACACTCTCTGCTGTACGAGTACTTCAC
  AGTTTATAACGAGCTCACCAAGGTCAAATACGTCACAGAAGGGATGAGAAAGCCAGC
  ATTCCTGTCTGGAGAGCAGAAGAAAGCTATCGTGGACCTCCTCTTCAAGACGAACCGG
  AAAGTTACCGTGAAACAGCTCAAAGAAGACTATTTCAAAAAGATTGAATGTTTCGACT
  CTGTTGAAATCAGCGGAGTGGAGGATCGCTTCAACGCATCCCTGGGAACGTATCACGA
35  TCTCCTGAAAATCATTAAGACAAGGACTTCCTGGACAATGAGGAGAACGAGGACAT

```

TCTTGAGGACATTGTCCTCACCTTACGTTGTTTGAAGATAGGGAGATGATTGAAGAA
 CGCTTGAAAACCTTACGCTCATCTCTTCGACGACAAAGTCATGAAACAGCTCAAGAGGC
 GCCGATATACAGGATGGGGGCGGCTGTCAAGAAAACCTGATCAATGGGATCCGAGACA
 AGCAGAGTGGAAAGACAATCCTGGATTTTCTTAAGTCCGATGGATTTGCCAACCGGAA
 5 CTTTCATGCAGTTGATCCATGATGACTCTCTCACCTTTAAGGAGGACATCCAGAAAGCA
 CAAGTTTCTGGCCAGGGGGACAGTCTTCACGAGCACATCGCTAATCTTGCAGGTAGCC
 CAGCTATCAAAAAGGGAATACTGCAGACCGTTAAGGTCGTGGATGAACTCGTCAAAG
 TAATGGGAAGGCATAAGCCCGAGAATATCGTTATCGAGATGGCCCGAGAGAACCAAA
 CTACCCAGAAGGGACAGAAGAACAGTAGGGAAAGGATGAAGAGGATTGAAGAGGGT
 10 ATAAAAGAAGCTGGGGTCCCAAATCCTTAAGGAACACCCAGTTGAAAACACCCAGCTT
 CAGAATGAGAAGCTCTACCTGTACTACCTGCAGAACGGCAGGGACATGTACGTGGAT
 CAGGAACTGGACATCAATCGGCTCTCCGACTACGACGTGGCTGCTATCGTGCCCCAGT
 CTTTTCTCAAAGATGATTCTATTGATAATAAAGTGTTGACAAGATCCGATAAAgcTAGA
 GGGAAGAGTGATAACGTCCCCTCAGAAGAAGTTGTCAAGAAAATGAAAAATTATTGG
 15 CGGCAGCTGCTGAACGCCAAACTGATCACACAACGGAAGTTCGATAATCTGACTAAG
 GCTGAACGAGGTGGCCTGTCTGAGTTGGATAAAGCCGGCTTCATCAAAGGCAGCTTG
 TTGAGACACGCCAGATCACCAAGCACGTGGCCCAAATTCTCGATTACGCATGAACAC
 CAAGTACGATGAAAATGACAAACTGATTCGAGAGGTGAAAGTTATTACTCTGAAGTCT
 AAGCTGGTCTCAGATTTTCAGAAAGGACTTTTCAGTTTTATAAGGTGAGAGAGATCAACA
 20 ATTACCACCATGCGCATGATGCCTACCTGAATGCAGTGGTAGGCACTGCACTTATCAA
 AAAATATCCCAAGCTTGAATCTGAATTTGTTTACGGAGACTATAAAGTGTACGATGTT
 AGGAAAATGATCGCAAAGTCTGAGCAGGAAATAGGCAAGGCCACCGCTAAGTACTTC
 TTTTACAGCAATATTATGAATTTTTTTCAAGACCGAGATTACACTGGCCAATGGAGAGA
 TTCGGAAGCGACCACTTATCGAAACAAACGGAGAAACAGGAGAAATCGTGTGGGACA
 25 AGGGTAGGGATTTTCGCGACAGTCCGGAAGGTCCTGTCCATGCCGCAGGTGAACATCGT
 TAAAAAGACCGAAGTACAGACCGGAGGCTTCTCCAAGGAAAGTATCCTCCCGAAAAG
 GAACAGCGACAAGCTGATCGCACGCAAAAAAGATTGGGACCCCAAGAAATACGGCGG
 ATTCGATTCTCCTACAGTCGCTTACAGTGTACTGGTTGTGGCCAAAGTGGAGAAAGGG
 AAGTCTAAAAAACTCAAAGCGTCAAGGAACTGCTGGGCATCACAATCATGGAGCGA
 30 TCAAGCTTCGAAAAAAACCCCATCGACTTTCTCGAGGCGAAAGGATATAAAGAGGTC
 AAAAAAGACCTCATCATTAAAGCTTCCCAAGTACTCTCTCTTTGAGCTTGAAAACGGCC
 GGAAACGAATGCTCGCTAGTGCGGGCGAGCTGCAGAAAGGTAACGAGCTGGCACTGC
 CCTCTAAATACGTTAATTTCTTGTATCTGGCCAGCCACTATGAAAAGCTCAAAGGGTCT
 CCCGAAGATAATGAGCAGAAGCAGCTGTTTCGTGGAACAACACAAACACTACCTTGAT
 35 GAGATCATCGAGCAAATAAGCGAATTCTCCAAAAGAGTGATCCTCGCCGACGCTAAC
 CTCGATAAGGTGCTTTCTGCTTACAATAAGCACAGGGATAAGCCCATCAGGGAGCAGG

CAGAAAACATTATCCACTTGTTTACTCTGACCAACTTGGGCGCGCCTGCAGCCTTCAA
 GTACTTCGACACCACCATAGACAGAAAGCGGTACACCTCTACAAAGGAGGTCCTGGA
 CGCCACACTGATTCATCAGTCAATTACGGGGCTCTATGAAACAAGAATCGACCTCTCT
 CAGCTCGGTGGAGACAGCAGGGCTGACCCCAAGAAGAAGAGGAAGGTGGAGGCCAG
 5 CGGTTCCGGACGGGCTGACGCATTGGACGATTTTGATCTGGATATGCTGGGAAGTGAC
 GCCCTCGATGATTTTGACCTTGACATGCTTGGTTCGGATGCCCTTGATGACTTTGACCT
 CGACATGCTCGGCAGTGACGCCCTTGATGATTTTCGACCTGGACATGCTGATTA ACTCT
 AGATGA

10 >Cas9_{m4}^{VP64}N Sequences

gccaccATGCCCAAGAAGAAGAGGAAGGTGGGAAGGGGGATGGACAAGAAGTACTCCA
 TTGGGCTCGCTATCGGCACAAACAGCGTCGGCTGGGCGGTCATTACGGACGAGTACAA
 GGTGCCGAGCAAAAAATTCAAAGTTCTGGGCAATACCGATCGCCACAGCATAAAGAA
 GAACCTCATTGGCGCCCTCCTGTTCGACTCCGGGGAGACGGCCGAAGCCACGCGGCTC
 15 AAAAGAACAGCACGGCGCAGATATACCCGCAGAAAGAATCGGATCTGCTACCTGCAG
 GAGATCTTTAGTAATGAGATGGCTAAGGTGGATGACTCTTTCTTCCATAGGCTGGAGG
 AGTCCTTTTTTGGTGGAGGAGGATAAAAAGCACGAGCGCCACCCAATCTTTGGCAATAT
 CGTGGACGAGGTGGCGTACCATGAAAAGTACCCAACCATATATCATCTGAGGAAGAA
 GCTTGTAGACAGTACTGATAAGGCTGACTTGCGGTTGATCTATCTCGCGCTGGCGCAT
 20 ATGATCAAATTTTCGGGGACACTTCCTCATCGAGGGGGACCTGAACCCAGACAACAGC
 GATGTCGACAAACTCTTTATCCA ACTGGTTCAGACTTACAATCAGCTTTTTCGAAGAGA
 ACCCGATCAACGCATCCGGAGTTGACGCCAAAGCAATCCTGAGCGCTAGGCTGTCCAA
 ATCCCGGCGGCTCGAAAACCTCATCGCACAGCTCCCTGGGGAGAAGAAGAACGGCCT
 GTTTGGTAATCTTATCGCCCTGTCACTCGGGCTGACCCCCAACTTTAAATCTAACTTCG
 25 ACCTGGCCGAAGATGCCAAGCTTCAACTGAGCAAAGACACCTACGATGATGATCTCG
 ACAATCTGCTGGCCCAGATCGGCGACCAGTACGCAGACCTTTTTTTTGGCGGCAAAGAA
 CCTGTCAGACGCCATTCTGCTGAGTGATATTCTGCGAGTGAACACGGAGATCACCAAA
 GCTCCGCTGAGCGCTAGTATGATCAAGCGCTATGATGAGCACCACCAAGACTTGACTT
 TGCTGAAGGCCCTTGTCAGACAGCAACTGCCTGAGAAGTACAAGGAAATTTTCTTCGA
 30 TCAGTCTAAAAATGGCTACGCCGGATACATTGACGGCGGAGCAAGCCAGGAGGAATT
 TTACAAATTTATTAAGCCCATCTTGGA AAAAATGGACGGCACCGAGGAGCTGCTGGTA
 AAGCTTAACAGAGAAGATCTGTTGCGCAAACAGCGCACTTTCGACAATGGAAGCATC
 CCCCACCAGATTACCTGGGCGAACTGCACGCTATCCTCAGGCGGCAAGAGGATTTCT
 ACCCCTTTTTTGAAAGATAACAGGGAAAAGATTGAGAAAATCCTCACATTTTCGGATACC

CTACTATGTAGGCCCCCTCGCCCGGGGAAATTCCAGATTCGCGTGGATGACTCGCAAA
 TCAGAAGAGACCATCACTCCCTGGAACCTCGAGGAAGTCGTGGATAAGGGGGCCTCT
 GCCCAGTCCTTCATCGAAAGGATGACTAACTTTGATAAAAATCTGCCTAACGAAAAGG
 TGCTTCCTAAACACTCTCTGCTGTACGAGTACTTCACAGTTTATAACGAGCTCACCAAG
 5 GTCAAATACGTCACAGAAGGGATGAGAAAGCCAGCATTCCTGTCTGGAGAGCAGAAG
 AAAGCTATCGTGGACCTCCTCTTCAAGACGAACCGGAAAGTTACCGTGAAACAGCTCA
 AAGAAGACTATTTCAAAAAGATTGAATGTTTCGACTCTGTTGAAATCAGCGGAGTGGA
 GGATCGCTTCAACGCATCCCTGGGAACGTATCACGATCTCCTGAAAATCATTAAAGAC
 AAGGACTTCCTGGACAATGAGGAGAACGAGGACATTCTTGAGGACATTGTCCTCACCC
 10 TTACGTTGTTTGAAGATAGGGAGATGATTGAAGAACGCTTGAAAACCTTACGCTCATCT
 CTTGACGACAAAAGTCATGAAACAGCTCAAGAGGCGCCGATATACAGGATGGGGGCG
 GCTGTCAAGAAAACCTGATCAATGGGATCCGAGACAAGCAGAGTGGAAGACAATCCT
 GGATTTTCTTAAGTCCGATGGATTTGCCAACCGGAACTTCATGCAGTTGATCCATGATG
 ACTCTCTCACCTTTAAGGAGGACATCCAGAAAGCACAAGTTTCTGGCCAGGGGGGACAG
 15 TCTTCACGAGCACATCGCTAATCTTGCAGGTAGCCCAGCTATCAAAAAGGGAATACTG
 CAGACCGTTAAGGTCGTGGATGAACTCGTCAAAGTAATGGGAAGGCATAAGCCCGAG
 AATATCGTTATCGAGATGGCCCGAGAGAACCAAACTACCCAGAAGGGACAGAAGAAC
 AGTAGGGAAAGGATGAAGAGGATTGAAGAGGGTATAAAAGAACTGGGGTCCCAAAT
 CCTTAAGGAACACCCAGTTGAAAACACCCAGCTTCAGAATGAGAAGCTCTACCTGTAC
 20 TACCTGCAGAACGGCAGGGACATGTACGTGGATCAGGAACTGGACATCAATCGGCTC
 TCCGACTACGACGTGGCTGCTATCGTGCCCCAGTCTTTTCTCAAAGATGATTCTATTGA
 TAATAAAGTGTTGACAAGATCCGATAAAgcTAGAGGGAAGAGTGATAACGTCCCCTCA
 GAAGAAGTTGTCAAGAAAATGAAAAATTATTGGCGGCAGCTGCTGAACGCCAACTG
 ATCACACAACGGAAGTTCGATAATCTGACTAAGGCTGAACGAGGTGGCCTGTCTGAGT
 25 TGGATAAAGCCGGCTTCATCAAAAGGCAGCTTGTTGAGACACGCCAGATCACCAAGC
 ACGTGGCCCAAATTCTCGATTACGCATGAACACCAAGTACGATGAAAATGACAAACT
 GATTCGAGAGGTGAAAGTTATTACTCTGAAGTCTAAGCTGGTCTCAGATTTTCAGAAAG
 GACTTTCAGTTTTATAAGGTGAGAGAGATCAACAATTACCACCATGCGCATGATGCCT
 ACCTGAATGCAGTGGTAGGCACTGCACTTATCAAAAAATATCCCAAGCTTGAATCTGA
 30 ATTTGTTTACGGAGACTATAAAGTGTACGATGTTAGGAAAATGATCGCAAAGTCTGAG
 CAGGAAATAGGCAAGGCCACCGCTAAGTACTTCTTTTACAGCAATATTATGAATTTTT
 TCAAGACCGAGATTACACTGGCCAATGGAGAGATTTCGGAAGCGACCACTTATCGAAA
 CAAACGGAGAAACAGGAGAAATCGTGTGGGACAAGGGTAGGGATTTTCGCGACAGTCC
 GGAAGGTCCTGTCCATGCCGCAGGTGAACATCGTTAAAAAGACCGAAGTACAGACCG
 35 GAGGCTTCTCCAAGGAAAGTATCCTCCCGAAAAGGAACAGCGACAAGCTGATCGCAC
 GCAAAAAAAGATTGGGACCCCAAGAAATACGGCGGATTTCGATTCTCCTACAGTCGCTTA

CAGTGTACTGGTTGTGGCCAAAGTGGAGAAAGGGAAGTCTAAAAAACTCAAAAGCGT
 CAAGGAACTGCTGGGCATCACAATCATGGAGCGATCAAGCTTCGAAAAAAACCCCAT
 CGACTTTCTCGAGGCGAAAGGATATAAAGAGGTCAAAAAAGACCTCATCATTAAGCTT
 CCCAAGTACTCTCTCTTTGAGCTTGAAAACGGCCGGAAACGAATGCTCGCTAGTGCGG
 5 GCGAGCTGCAGAAAGGTAACGAGCTGGCACTGCCCTCTAAATACGTTAATTTCTTGTA
 TCTGGCCAGCCACTATGAAAAGCTCAAAGGGTCTCCCGAAGATAATGAGCAGAAGCA
 GCTGTTCGTGGAACAACACAAACACTACCTTGATGAGATCATCGAGCAAATAAGCGA
 ATTCTCCAAAAGAGTGATCCTCGCCGACGCTAACCTCGATAAGGTGCTTTCTGCTTAC
 AATAAGCACAGGGATAAGCCCATCAGGGAGCAGGCAGAAAACATTATCCACTTGTTT
 10 ACTCTGACCAACTTGGGCGCGCCTGCAGCCTTCAAGTACTTCGACACCACCATAGACA
 GAAAGCGGTACACCTCTACAAAGGAGGTCCTGGACGCCACACTGATTCATCAGTCAAT
 TACGGGGCTCTATGAAACAAGAATCGACCTCTCTCAGCTCGGTGGAGACAGCAGGGCT
 GACCCCAAGAAGAAGAGGAAGGTGGAGGCCAGCGGTTCCGGACGGGGCTGACGCATTG
 GACGATTTTGATCTGGATATGCTGGGAAGTGACGCCCTCGATGATTTTGACCTTGACA
 15 TGCTTGGTTTCGGATGCCCTTGATGACTTTGACCTCGACATGCTCGGCAGTGACGCCCTT
 GATGATTTGACCTGGACATGCTGATTAACCTCTAGATGA

>Cas9_{m4}^{VP64}C

gccaccATGGACAAGAAGTACTCCATTGGGCTCGCTATCGGCACAAACAGCGTCGGCTGG
 20 GCCGTCATTACGGACGAGTACAAGGTGCCGAGCAAAAAATTCAAAGTTCTGGGCAAT
 ACCGATCGCCACAGCATAAAGAAGAACCTCATTGGCGCCCTCCTGTTCGACTCCGGGG
 AGACGGCCGAAGCCACGCGGCTCAAAAGAACAGCACGGCGCAGATATACCCGCAGAA
 AGAATCGGATCTGCTACCTGCAGGAGATCTTTAGTAATGAGATGGCTAAGGTGGATGA
 CTCTTTCTTCCATAGGCTGGAGGAGTCCTTTTTTGGTGGAGGAGGATAAAAAGCACGAG
 25 CGCCACCCAATCTTTGGCAATATCGTGGACGAGGTGGCGTACCATGAAAAGTACCCAA
 CCATATATCATCTGAGGAAGAAGCTTGTAGACAGTACTGATAAGGCTGACTTGCGGTT
 GATCTATCTCGCGCTGGCGCATATGATCAAATTTGGGGGACACTTCCTCATCGAGGGG
 GACCTGAACCCAGACAACAGCGATGTCGACAAACTCTTTATCCAACCTGGTTTCAGACTT
 ACAATCAGCTTTTTCGAAGAGAACCCGATCAACGCATCCGGAGTTGACGCCAAAGCAA
 30 TCCTGAGCGCTAGGCTGTCCAAATCCCGGCGGCTCGAAAACCTCATCGCACAGCTCCC
 TGGGGAGAAGAAGAACGGCCTGTTTGGTAATCTTATCGCCCTGTCACTCGGGCTGACC
 CCCAACTTTAAATCTAACTTCGACCTGGCCGAAGATGCCAAGCTTCAACTGAGCAAAG
 ACACCTACGATGATGATCTCGACAATCTGCTGGCCCAGATCGGCGACCAGTACGCAGA
 CCTTTTTTTTGGCGGCAAAGAACCTGTCAGACGCCATTCTGCTGAGTGATATTCTGCGAG

TGAACACGGAGATCACCAAAGCTCCGCTGAGCGCTAGTATGATCAAGCGCTATGATG
 AGCACCACCAAGACTTGACTTTGCTGAAGGCCCTTGTCAGACAGCAACTGCCTGAGAA
 GTACAAGGAAATTTTCTTCGATCAGTCTAAAAATGGCTACGCCGGATACATTGACGGC
 GGAGCAAGCCAGGAGGAATTTTACAAATTTATTAAGCCCATCTTGGA AAAAATGGAC
 5 GGCACCGAGGAGCTGCTGGTAAAGCTTAACAGAGAAGATCTGTTGCGCAAACAGCGC
 ACTTTTCGACAATGGAAGCATCCCCCACCAGATTCACCTGGGCGAACTGCACGCTATCC
 TCAGGCGGCAAGAGGATTTCTACCCCTTTTTGAAAGATAACAGGGAAAAGATTGAGA
 AAATCCTCACATTTTCGGATACCCTACTATGTAGGCCCCCTCGCCCGGGGAAATTCCAG
 ATTCGCGTGGATGACTCGCAAATCAGAAGAGACCATCACTCCCTGGA ACTTCGAGGAA
 10 GTCGTGGATAAAGGGGGCCTCTGCCCAGTCCTTCATCGAAAGGATGACTAACTTTGATA
 AAAATCTGCCTAACGAAAAGGTGCTTCCTAAACACTCTCTGCTGTACGAGTACTTCAC
 AGTTTATAACGAGCTCACCAAGGTCAAATACGTCACAGAAGGGATGAGAAAGCCAGC
 ATTCCTGTCTGGAGAGCAGAAGAAAGCTATCGTGGACCTCCTCTTCAAGACGAACCGG
 AAAGTTACCGTGAAACAGCTCAAAGAAGACTATTTCAAAAAGATTGAATGTTTCGACT
 15 CTGTTGAAATCAGCGGAGTGGAGGATCGCTTCAACGCATCCCTGGGAACGTATCACGA
 TCTCCTGAAAATCATTAAAGACAAGGACTTCCTGGACAATGAGGAGAACGAGGACAT
 TCTTGAGGACATTGTCCTCACCTTACGTTGTTTGAAGATAGGGAGATGATTGAAGAA
 CGCTTGAAA ACTTACGCTCATCTCTTCGACGACAAAGTCATGAAACAGCTCAAGAGGC
 GCCGATATACAGGATGGGGGCGGCTGTCAAGAAA ACTGATCAATGGGATCCGAGACA
 20 AGCAGAGTGGAAAGACAATCCTGGATTTTCTTAAGTCCGATGGATTTGCCAACCGGAA
 CTTTCATGCAGTTGATCCATGATGACTCTCTCACCTTTAAGGAGGACATCCAGAAAGCA
 CAAGTTTCTGGCCAGGGGGACAGTCTTCACGAGCACATCGCTAATCTTGCAGGTAGCC
 CAGCTATCAAAAAGGGAATACTGCAGACCGTTAAGGTCTGTTGATGAACTCGTCAAAG
 TAATGGGAAGGCATAAGCCCGAGAATATCGTTATCGAGATGGCCCGAGAGAACCAAA
 25 CTACCCAGAAGGGACAGAAGAACAGTAGGGAAAGGATGAAGAGGATTGAAGAGGGT
 ATAAAAGAACTGGGGTCCCAAATCCTTAAGGAACACCCAGTTGAAAACACCCAGCTT
 CAGAATGAGAAGCTCTACCTGTACTACCTGCAGAACGGCAGGGACATGTACGTGGAT
 CAGGAACTGGACATCAATCGGCTCTCCGACTACGACGTGGCTGCTATCGTGCCCCAGT
 CTTTTCTCAAAGATGATTCTATTGATAATAAAGTGTTGACAAGATCCGATAAAgcTAGA
 30 GGGAAGAGTGATAACGTCCCCTCAGAAGAAGTTGTCAAGAAAATGAAAAATTATTGG
 CGGCAGCTGCTGAACGCCAAACTGATCACACAACGGAAGTTCGATAATCTGACTAAG
 GCTGAACGAGGTGGCCTGTCTGAGTTGGATAAAGCCGGCTTCATCAAAGGCAGCTTG
 TTGAGACACGCCAGATCACCAAGCACGTGGCCCAAATTCTCGATTACGCATGAACAC
 CAAGTACGATGAAAATGACAAACTGATTTCGAGAGGTGAAAGTTATTACTCTGAAGTCT
 35 AAGCTGGTCTCAGATTTTCAGAAAGGACTTTCAGTTTTATAAGGTGAGAGAGATCAACA
 ATTACCACCATGCGCATGATGCCTACCTGAATGCAGTGGTAGGCACTGCACTTATCAA

AAAATATCCCAAGCTTGAATCTGAATTTGTTTACGGAGACTATAAAGTGTACGATGTT
 AGGAAAATGATCGCAAAGTCTGAGCAGGAAATAGGCAAGGCCACCGCTAAGTACTTC
 TTTTACAGCAATATTATGAATTTTTTCAAGACCGAGATTACACTGGCCAATGGAGAGA
 TTCGGAAGCGACCACTTATCGAAACAAACGGAGAAACAGGAGAAATCGTGTGGGACA
 5 AGGGTAGGGATTTTCGCGACAGTCCGGAAGGTCCTGTCCATGCCGCAGGTGAACATCGT
 TAAAAAGACCGAAGTACAGACCGGAGGCTTCTCCAAGGAAAGTATCCTCCCGAAAAG
 GAACAGCGACAAGCTGATCGCACGCAAAAAAGATTGGGACCCCAAGAAATACGGCGG
 ATTCGATTCTCCTACAGTCGCTTACAGTGTACTGGTTGTGGCCAAAGTGGAGAAAGGG
 AAGTCTAAAAAACTCAAAAGCGTCAAGGAACTGCTGGGCATCACAATCATGGAGCGA
 10 TCAAGCTTCGAAAAAAACCCCATCGACTTTCTCGAGGCGAAAGGATATAAAGAGGTC
 AAAAAAGACCTCATCATTAAGCTTCCCAAGTACTCTCTCTTTGAGCTTGAAAACGGCC
 GGAAACGAATGCTCGCTAGTGCGGGCGAGCTGCAGAAAGGTAACGAGCTGGCACTGC
 CCTCTAAATACGTTAATTTCTTGTATCTGGCCAGCCACTATGAAAAGCTCAAAGGGTCT
 CCCGAAGATAATGAGCAGAAGCAGCTGTTCGTGGAACAACACAAACACTACCTTGAT
 15 GAGATCATCGAGCAAATAAGCGAATTCTCCAAAAGAGTGATCCTCGCCGACGCTAAC
 CTCGATAAGGTGCTTTTCTGCTTACAATAAGCACAGGGATAAGCCCATCAGGGAGCAGG
 CAGAAAACATTATCCACTTGTTTACTCTGACCAACTTGGGCGCGCCTGCAGCCTTCAA
 GTACTTCGACACCACCATAGACAGAAAGCGGTACACCTCTACAAAGGAGGTCCTGGA
 CGCCACACTGATTCATCAGTCAATTACGGGGCTCTATGAAACAAGAATCGACCTCTCT
 20 CAGCTCGGTGGAGACAGCAGGGCTGACCCCAAGAAGAAGAGGAAGGTGGAGGCCAG
 CGGTTCCGGACGGGCTGACGCATTGGACGATTTTGATCTGGATATGCTGGGAAGTGAC
 GCCCTCGATGATTTTGACCTTGACATGCTTGGTTCGGATGCCCTTGATGACTTTGACCT
 CGACATGCTCGGCAGTGACGCCCTTGATGATTTTCGACCTGGACATGCTGATTA ACTCT
 AGAGCGGCGCGCAGATCCAAAAAAGAAGAGAAAGGTAGATCCAAAAAAGAAGAGAAA
 25 GGTAGATCCAAAAAAGAAGAGAAAGGTAGATACGGCCCGCATAG

B. Sequences of the MS2-activator constructs and corresponding gRNA backbone vector with 2X
 MS2 aptamer domains is provided below (NLS, VP64, gRNA spacer, and MS2-binding RNA stem
 loop domains are highlighted). Two versions of the former were constructed with the MS2_{VP64N}
 30 fusion protein format showing highest activity.

>MS2_{VP64N}

gccaccATGGGACCTAAGAAAAAGAGGAAGGTGGCGGCCGCTTCTAGAATGGCTTCTAA
 CTTTACTCAGTTCGTTCTCGTCGACAATGGCGGAACTGGCGACGTGACTGTGCCCCA
 AGCAACTTCGCTAACGGGATCGCTGAATGGATCAGCTCTAACTCGCGTTCACAGGCTT

ACAAAGTAACCTGTAGCGTTCGTCAGAGCTCTGCGCAGAATCGCAAATACACCATCAA
 AGTCGAGGTGCCTAAAGGCGCCTGGCGTTCGTACTTAAATATGGAACTAACCATTCCA
 ATTTTCGCCACGAATTCCGACTGCGAGCTTATTGTAAAGGCAATGCAAGGTCTCCTAA
 AAGATGGAAACCCGATTCCCTCAGCAATCGCAGCAAACCTCCGGCATCTACGAGGCCA
 5 GCGGTTCCGGACGGGCTGACGCATTGGACGATTTTGATCTGGATATGCTGGGAAGTGA
 CGCCCTCGATGATTTTGACCTTGACATGCTTGGTTCCGGATGCCCTTGATGACTTTGACC
 TCGACATGCTCGGCAGTGACGCCCTTGATGATTTTCGACCTGGACATGCTGATTAACCTCT
 AGATGA

>MS2_{VP6}C

10 gccaccATGGGACCTAAGAAAAAGAGGAAGGTGGCGGCCGCTTCTAGAATGGCTTCTAA
 CTTTACTCAGTTCGTTCTCGTCGACAATGGCGGAACTGGCGACGTGACTGTCGCCCCA
 AGCAACTTCGCTAACGGGATCGCTGAATGGATCAGCTCTAACTCGCGTTCACAGGCTT
 ACAAAGTAACCTGTAGCGTTCGTCAGAGCTCTGCGCAGAATCGCAAATACACCATCAA
 AGTCGAGGTGCCTAAAGGCGCCTGGCGTTCGTACTTAAATATGGAACTAACCATTCCA
 15 ATTTTCGCCACGAATTCCGACTGCGAGCTTATTGTAAAGGCAATGCAAGGTCTCCTAA
 AAGATGGAAACCCGATTCCCTCAGCAATCGCAGCAAACCTCCGGCATCTACGAGGCCA
 GCGGTTCCGGACGGGCTGACGCATTGGACGATTTTGATCTGGATATGCTGGGAAGTGA
 CGCCCTCGATGATTTTGACCTTGACATGCTTGGTTCCGGATGCCCTTGATGACTTTGACC
 TCGACATGCTCGGCAGTGACGCCCTTGATGATTTTCGACCTGGACATGCTGATTAACCTCT
 20 AGAGCGGCCCGCAGATCCAAAAAAGAAGAGAAAGGTAGATCCAAAAAAGAAGAGAAA
 GGTAGATCCAAAAAAGAAGAGAAAGGTAGATACGGCCCGCATAG

>gRNA_{2XMS2}

TGTACAAAAAAGCAGGCTTTAAAGGAACCAATTCAGTCGACTGGATCCGGTACCAAG
 GTCGGGCAGGAAGAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACA
 25 AGGCTGTTAGAGAGATAATTAGAATTAATTTGACTGTAAACACAAAGATATTAGTACA
 AAATACGTGACGTAGAAAGTAATAATTTCTTGGGTAGTTTGCAGTTTTTAAAATTATGTT
 TTAAAATGGACTATCATATGCTTACCGTAACTTGAAAGTATTTTCGATTTCTTGGCTTTA
 TATATCTTGTGGAAAGGACGAAACACCGNNNNNNNNNNNNNNNNNNNNNNNNNNNGTTTTAGAG
 CTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACC
 30 GAGTCGGTGCTCTGCAGGTCGACTCTAGAAAACATGAGGATCACCCATGTCTGCAGTA
 TTCCCGGGTTCATTAGATCCTAAGGTACCTAATTGCCTAGAAAACATGAGGATCACCC
 ATGTCTGCAGGTCGACTCTAGAAATTTTTTCTAGAC

C. dTomato fluorescence based transcriptional activation reporter sequences are listed below (lScel control-TF target, gRNA targets, minCMV promoter and FLAG tag + dTomato sequences are highlighted).

>TF Reporter 1

```

5  TAGGGATAACAGGGTAATAGTGTCCCCTCCACCCCACAGTGGGGCGAGGTAGGCGTG
   TACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGCCTGGAG
   AATTCgccaccaigGACTACAAGGATGACGACGATAAACTTCCGGTGGCGGACTGGGTTC
   CACCGTGAGCAAGGGGCGAGGAGGTCATCAAAGAGTTTCATGCGCTTCAAGGTGCGCAT
   GGAGGGCTCCATGAACGGCCACGAGTTTCGAGATCGAGGGGCGAGGGGCGAGGGCCGCCC
10  CTACGAGGGCAGCCAGACCGCCAAGCTGAAGGTGACCAAGGGGCGGCCCCCTGCCCTT
   CGCCTGGGACATCCTGTCCCCCCAGTTTCATGTACGGCTCCAAGGCGTACGTGAAGCAC
   CCCGCCGACATCCCCGATTACAAGAAGCTGTCCTTCCCCGAGGGGCTTCAAGTGGGAGC
   GCGTGATGAACTTCGAGGACGGCGGTCTGGTGACCGTGACCCAGGACTCCTCCCTGCA
   GGACGGCACGCTGATCTACAAGGTGAAGATGCGCGGCACCAACTTCCCCCCCCGACGG
15  CCCCCTAATGCAGAAGAAGACCATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCC
   CCGCGACGGCGTGCTGAAGGGGCGAGATCCACCAGGCCCCTGAAGCTGAAGGACGGCGG
   CCACTACCTGGTGGAGTTCAAGACCATCTACATGGCCAAGAAGCCCGTGCAACTGCCC
   GGCTACTACTACGTGGACACCAAGCTGGACATCACCTCCCACAACGAGGACTACACCA
   TCGTGGAACAGTACGAGCGCTCCGAGGGGCCGCCACCACCTGTTCTGTACGGCATGGA
20  CGAGCTGTACAAGTAA

```

>TF Reporter 2

```

   TAGGGATAACAGGGTAATAGTGGGGCCACTAGGGACAGGATTGGCGAGGTAGGCGTG
   TACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGCCTGGAG
   AATTCgccaccatgGACTACAAGGATGACGACGATAAACTTCCGGTGGCGGACTGGGTTC
25  CACCGTGAGCAAGGGGCGAGGAGGTCATCAAAGAGTTTCATGCGCTTCAAGGTGCGCAT
   GGAGGGCTCCATGAACGGCCACGAGTTTCGAGATCGAGGGGCGAGGGGCGAGGGCCGCCC
   CTACGAGGGCAGCCAGACCGCCAAGCTGAAGGTGACCAAGGGGCGGCCCCCTGCCCTT
   CGCCTGGGACATCCTGTCCCCCCAGTTTCATGTACGGCTCCAAGGCGTACGTGAAGCAC
   CCCGCCGACATCCCCGATTACAAGAAGCTGTCCTTCCCCGAGGGGCTTCAAGTGGGAGC
30  GCGTGATGAACTTCGAGGACGGCGGTCTGGTGACCGTGACCCAGGACTCCTCCCTGCA
   GGACGGCACGCTGATCTACAAGGTGAAGATGCGCGGCACCAACTTCCCCCCCCGACGG
   CCCCCTAATGCAGAAGAAGACCATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCC
   CCGCGACGGCGTGCTGAAGGGGCGAGATCCACCAGGCCCCTGAAGCTGAAGGACGGCGG
   CCACTACCTGGTGGAGTTCAAGACCATCTACATGGCCAAGAAGCCCGTGCAACTGCCC

```

GGCTACTACTACGTGGACACCAAGCTGGACATCACCTCCCACAACGAGGACTACACCA
TCGTGGAACAGTACGAGCGCTCCGAGGGGCCGCCACCACTGTTCTGTACGGCATGGA
CGAGCTGTACAAGTAA

- 5 D. General format of the reporter libraries used for TALE and Cas9-gRNA specificity assays is provided below (ISeel control-TF target, gRNA/TALE target site (23bp for gRNAs and 18bp for TALEs), minCMV promoter, RNA barcode, and dTomato sequences are highlighted).

> Specificity Reporter Libraries

TAGGGATAACAGGOTAATAGTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNCGAGGTAGGCGT
10 GTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGCCTGGA
GAATTCgcccaccatgGACTACAAGGATGACGACGATAAANNNNNNNNNNNNNNNNNNNNNNNN
NNNNACTTCCGGTGGCGGACTGGGTTCACCGTGAGCAAGGGGCGAGGAGGTCATCAA
AGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGCTCCATGAACGGGCCACGAGTTCGA
GATCGAGGGGCGAGGGGCGAGGGGCCGCCCCCTACGAGGGGCACCCAGACCCGCCAAGCTGAA
15 GGTGACCAAGGGGCGGCCCCCTGCCCTTCGGCTGGGACATCCTGTCCCCCAGTTCATG
TACGGCTCCAAGGCGTACGTGAAGCACCCCGCCGACATCCCCGATTACAAGAAGCTGT
CCTTCCCCGAGGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGCGGGTCTGGT
GACCGTGACCCAGGACTCCTCCCTGCAGGACGGCACGCTGATCTACAAGGTGAAGAT
GCGGGGCACCAACTTCCCCCCCCGACGGGCCCCGTAATGCAGAAGAAGACCATGGGCTG
20 GGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGGCGAGATCCA
CCAGGCCCTGAAGCTGAAGGACGGCGGGCCACTACCTGGTGGAGTTCAAGACCATCTA
CATGGCCAAGAAGCCCGTGCAACTGCCCGGCTACTACTACGTGGACACCAAGCTGGA
CATCACCTCCCACAACGAGGACTACACCATCGTGGAACAGTACGAGCGCTCCGAGGG
CCGCCACCACTGTTCTGTACGGCATGGACGAGCTGTACAAGTAAGAATTC

25

Claims:

1. A method of modulating expression of a target nucleic acid in a cell comprising
 introducing into the cell a first foreign nucleic acid encoding one or more RNAs
 5 complementary to DNA, wherein the DNA includes the target nucleic acid,
 introducing into the cell a second foreign nucleic acid encoding a nuclease-null Cas9
 protein that binds to the DNA and is guided by the one or more RNAs,
 introducing into the cell a third foreign nucleic acid encoding a transcriptional regulator
 protein or domain,
 10 wherein the one or more RNAs, the nuclease-null Cas9 protein, and the transcriptional
 regulator protein or domain are expressed,
 wherein the one or more RNAs, the nuclease-null Cas9 protein and the transcriptional
 regulator protein or domain co-localize to the DNA and wherein the transcriptional regulator
 protein or domain regulates expression of the target nucleic acid.
- 15 2. The method of claim 1 wherein the foreign nucleic acid encoding a nuclease-null
 Cas9 protein further encodes the transcriptional regulator protein or domain fused to the nuclease-
 null Cas9 protein.
3. The method of claim 1 wherein the foreign nucleic acid encoding one or more
 RNAs further encodes a target of an RNA-binding domain and the foreign nucleic acid encoding
 20 the transcriptional regulator protein or domain further encodes an RNA-binding domain fused to
 the transcriptional regulator protein or domain.
4. The method of claim 1 wherein the cell is a eukaryotic cell.
5. The method of claim 1 wherein the cell is a yeast cell, a plant cell or an animal
 cell.
- 25 6. The method of claim 1 wherein the RNA is between about 10 to about 500
 nucleotides.
7. The method of claim 1 wherein the RNA is between about 20 to about 100
 nucleotides.
8. The method of claim 1 wherein the transcriptional regulator protein or domain is a
 30 transcriptional activator.
9. The method of claim 1 wherein the transcriptional regulator protein or domain
 upregulates expression of the target nucleic acid.
10. The method of claim 1 wherein the transcriptional regulator protein or domain
 upregulates expression of the target nucleic acid to treat a disease or detrimental condition.
- 35 11. The method of claim 1 wherein the target nucleic acid is associated with a disease
 or detrimental condition.

12. The method of claim 1 wherein the one or more RNAs is a guide RNA.

13. The method of claim 1 wherein the one or more RNAs is a tracrRNA-crRNA fusion.

14. The method of claim 1 wherein the DNA is genomic DNA, mitochondrial DNA,
5 viral DNA, or exogenous DNA.

15. A method of altering a DNA target nucleic acid in a cell comprising
introducing into the cell a first foreign nucleic acid encoding two or more RNAs with each
RNA being complementary to an adjacent site in the DNA target nucleic acid,
introducing into the cell a second foreign nucleic acid encoding at least one Cas9 protein
10 nickase having one inactive nuclease domain and being guided by the two or more RNAs, and
wherein the two or more RNAs and the at least one Cas9 protein nickase are expressed and
wherein the at least one Cas9 protein nickase co-localizes with the two or more RNAs to the DNA
target nucleic acid and nicks the DNA target nucleic acid resulting in two or more adjacent nicks.

16. The method of claim 15 wherein the two or more adjacent nicks are on the same
15 strand of the double stranded DNA.

17. The method of claim 15 wherein the two or more adjacent nicks are on the same
strand of the double stranded DNA and result in homologous recombination.

18. The method of claim 15 wherein the two or more adjacent nicks are on different
strands of the double stranded DNA.

20 19. The method of claim 15 wherein the two or more adjacent nicks are on different
strands of the double stranded DNA and create double stranded breaks.

20. The method of claim 15 wherein the two or more adjacent nicks are on different
strands of the double stranded DNA and create double stranded breaks resulting in nonhomologous
end joining.

25 21. The method of claim 15 wherein the two or more adjacent nicks are on different
strands of the double stranded DNA and are offset with respect to one another.

22. The method of claim 15 wherein the two or more adjacent nicks are on different
strands of the double stranded DNA and are offset with respect to one another and create double
stranded breaks.

30 23. The method of claim 15 wherein the two or more adjacent nicks are on different
strands of the double stranded DNA and are offset with respect to one another and create double
stranded breaks resulting in nonhomologous end joining.

24. The method of claim 15 further including introducing into the cell a third foreign
nucleic acid encoding a donor nucleic acid sequence wherein the two or more nicks results in
35 homologous recombination of the target nucleic acid with the donor nucleic acid sequence.

25. A cell comprising

a first foreign nucleic acid encoding one or more RNAs complementary to DNA, wherein the DNA includes a target nucleic acid,

a second foreign nucleic acid encoding a nuclease-null Cas9 protein, and

a third foreign nucleic acid encoding a transcriptional regulator protein or domain wherein the one or more RNAs, the nuclease-null Cas9 protein and the transcriptional regulator protein or domain are members of a co-localization complex for the target nucleic acid.

26. The cell of claim 25 wherein the foreign nucleic acid encoding a nuclease-null Cas9 protein further encodes the transcriptional regulator protein or domain fused to the nuclease-null Cas9 protein.

27. The cell of claim 25 wherein the foreign nucleic acid encoding one or more RNAs further encodes a target of an RNA-binding domain and the foreign nucleic acid encoding the transcriptional regulator protein or domain further encodes an RNA-binding domain fused to the transcriptional regulator protein or domain.

28. The cell of claim 25 wherein the cell is a eukaryotic cell.

29. The cell of claim 25 wherein the cell is a yeast cell, a plant cell or an animal cell.

30. The cell of claim 25 wherein the RNA includes between about 10 to about 500 nucleotides.

31. The cell of claim 25 wherein the RNA includes between about 20 to about 100 nucleotides.

32. The cell of claim 25 wherein the transcriptional regulator protein or domain is a transcriptional activator.

33. The cell of claim 25 wherein the transcriptional regulator protein or domain upregulates expression of the target nucleic acid.

34. The cell of claim 25 wherein the transcriptional regulator protein or domain upregulates expression of the target nucleic acid to treat a disease or detrimental condition.

35. The cell of claim 25 wherein the target nucleic acid is associated with a disease or detrimental condition.

36. The cell of claim 25 wherein the one or more RNAs is a guide RNA.

37. The cell of claim 25 wherein the one or more RNAs is a tracrRNA-crRNA fusion.

38. The cell of claim 25 wherein the DNA is genomic DNA, mitochondrial DNA, viral DNA, or exogenous DNA.

39. A cell comprising

a first foreign nucleic acid encoding two or more RNAs with each RNA being complementary to an adjacent site in a DNA target nucleic acid, and

a second foreign nucleic acid encoding at least one Cas9 protein nickase having one inactive nuclease domain, and wherein the two or more RNAs and the at least one Cas9 protein nickase are members of a co-localization complex for the DNA target nucleic acid.

40. The cell of claim 39 wherein the cell is a eukaryotic cell.

5 41. The cell of claim 39 wherein the cell is a yeast cell, a plant cell or an animal cell.

42. The cell of claim 39 wherein the RNA includes between about 10 to about 500 nucleotides.

43. The cell of claim 39 wherein the RNA includes between about 20 to about 100 nucleotides.

10 44. The cell of claim 39 wherein the target nucleic acid is associated with a disease or detrimental condition.

45. The cell of claim 39 wherein the two or more RNAs are guide RNAs.

46. The cell of claim 39 wherein the two or more RNAs are tracrRNA-crRNA fusions.

15 47. The cell of claim 39 wherein the DNA target nucleic acid is genomic DNA, mitochondrial DNA, viral DNA, or exogenous DNA.

48. A method of altering a DNA target nucleic acid in a cell comprising introducing into the cell a first foreign nucleic acid encoding two or more RNAs with each RNA being complementary to an adjacent site in the DNA target nucleic acid,

20 introducing into the cell a second foreign nucleic acid encoding at least one Cas9 protein nickase having one inactive nuclease domain and being guided by the two or more RNAs, and

wherein the two or more RNAs and the at least one Cas9 protein nickase are expressed and wherein the at least one Cas9 protein nickase co-localizes with the two or more RNAs to the DNA target nucleic acid and nicks the DNA target nucleic acid resulting in two or more adjacent nicks, and

25 wherein the two or more adjacent nicks are on different strands of the double stranded DNA and create double stranded breaks resulting in fragmentation of the target nucleic acid thereby preventing expression of the target nucleic acid.

49. A method of modulating expression of a target nucleic acid in a cell comprising

30 introducing into the cell a first foreign nucleic acid encoding one or more RNAs complementary to DNA, wherein the DNA includes the target nucleic acid,

introducing into the cell a second foreign nucleic acid encoding an RNA guided nuclease null DNA binding protein,

introducing into the cell a third foreign nucleic acid encoding a transcriptional regulator protein or domain,

35 wherein the one or more RNAs, the RNA guided nuclease null DNA binding protein, and the transcriptional regulator protein or domain are expressed,

wherein the one or more RNAs, the RNA guided nuclease null DNA binding protein and the transcriptional regulator protein or domain co-localize to the DNA and wherein the transcriptional regulator protein or domain regulates expression of the target nucleic acid.

50. The method of claim 49 wherein the foreign nucleic acid encoding an RNA guided nuclease null DNA binding protein further encodes the transcriptional regulator protein or domain fused to the RNA guided nuclease null DNA binding protein.

51. The method of claim 49 wherein the foreign nucleic acid encoding one or more RNAs further encodes a target of an RNA-binding domain and the foreign nucleic acid encoding the transcriptional regulator protein or domain further encodes an RNA-binding domain fused to the transcriptional regulator protein or domain.

52. The method of claim 49 wherein the cell is a eukaryotic cell.

53. The method of claim 49 wherein the cell is a yeast cell, a plant cell or an animal cell.

54. The method of claim 49 wherein the RNA is between about 10 to about 500 nucleotides.

55. The method of claim 49 wherein the RNA is between about 20 to about 100 nucleotides.

56. The method of claim 49 wherein the transcriptional regulator protein or domain is a transcriptional activator.

57. The method of claim 49 wherein the transcriptional regulator protein or domain upregulates expression of the target nucleic acid.

58. The method of claim 49 wherein the transcriptional regulator protein or domain upregulates expression of the target nucleic acid to treat a disease or detrimental condition.

59. The method of claim 49 wherein the target nucleic acid is associated with a disease or detrimental condition.

60. The method of claim 49 wherein the one or more RNAs is a guide RNA.

61. The method of claim 49 wherein the one or more RNAs is a tracrRNA-crRNA fusion.

62. The method of claim 49 wherein the DNA is genomic DNA, mitochondrial DNA, viral DNA, or exogenous DNA.

63. A method of altering a DNA target nucleic acid in a cell comprising introducing into the cell a first foreign nucleic acid encoding two or more RNAs with each RNA being complementary to an adjacent site in the DNA target nucleic acid,

introducing into the cell a second foreign nucleic acid encoding at least one RNA guided DNA binding protein nickase, and

wherein the two or more RNAs and the at least one RNA guided DNA binding protein nickase are expressed and wherein the at least one RNA guided DNA binding protein nickase co-localizes with the two or more RNAs to the DNA target nucleic acid and nicks the DNA target nucleic acid resulting in two or more adjacent nicks.

5 64. The method of claim 63 wherein the two or more adjacent nicks are on the same strand of the double stranded DNA.

65. The method of claim 63 wherein the two or more adjacent nicks are on the same strand of the double stranded DNA and result in homologous recombination.

10 66. The method of claim 63 wherein the two or more adjacent nicks are on different strands of the double stranded DNA.

67. The method of claim 63 wherein the two or more adjacent nicks are on different strands of the double stranded DNA and create double stranded breaks.

15 68. The method of claim 63 wherein the two or more adjacent nicks are on different strands of the double stranded DNA and create double stranded breaks resulting in nonhomologous end joining.

69. The method of claim 63 wherein the two or more adjacent nicks are on different strands of the double stranded DNA and are offset with respect to one another.

20 70. The method of claim 63 wherein the two or more adjacent nicks are on different strands of the double stranded DNA and are offset with respect to one another and create double stranded breaks.

71. The method of claim 63 wherein the two or more adjacent nicks are on different strands of the double stranded DNA and are offset with respect to one another and create double stranded breaks resulting in nonhomologous end joining.

25 72. The method of claim 63 further including introducing into the cell a third foreign nucleic acid encoding a donor nucleic acid sequence wherein the two or more nicks results in homologous recombination of the target nucleic acid with the donor nucleic acid sequence.

30 73. A cell comprising
a first foreign nucleic acid encoding one or more RNAs complementary to DNA, wherein the DNA includes a target nucleic acid,
a second foreign nucleic acid encoding an RNA guided nuclease-null DNA binding protein, and

a third foreign nucleic acid encoding a transcriptional regulator protein or domain wherein the one or more RNAs, the RNA guided nuclease-null DNA binding protein and the transcriptional regulator protein or domain are members of a co-localization complex for the target nucleic acid.

74. The cell of claim 73 wherein the foreign nucleic acid encoding an RNA guided nuclease-null DNA binding protein further encodes the transcriptional regulator protein or domain fused to the RNA guided nuclease-null DNA binding protein.

75. The cell of claim 73 wherein the foreign nucleic acid encoding one or more RNAs further encodes a target of an RNA-binding domain and the foreign nucleic acid encoding the transcriptional regulator protein or domain further encodes an RNA-binding domain fused to the transcriptional regulator protein or domain.

76. The cell of claim 73 wherein the cell is a eukaryotic cell.

77. The cell of claim 73 wherein the cell is a yeast cell, a plant cell or an animal cell.

78. The cell of claim 73 wherein the RNA includes between about 10 to about 500 nucleotides.

79. The cell of claim 73 wherein the RNA includes between about 20 to about 100 nucleotides.

80. The cell of claim 73 wherein the transcriptional regulator protein or domain is a transcriptional activator.

81. The cell of claim 73 wherein the transcriptional regulator protein or domain upregulates expression of the target nucleic acid.

82. The cell of claim 73 wherein the transcriptional regulator protein or domain upregulates expression of the target nucleic acid to treat a disease or detrimental condition.

83. The cell of claim 73 wherein the target nucleic acid is associated with a disease or detrimental condition.

84. The cell of claim 73 wherein the one or more RNAs is a guide RNA.

85. The cell of claim 73 wherein the one or more RNAs is a tracrRNA-crRNA fusion.

86. The cell of claim 73 wherein the DNA is genomic DNA, mitochondrial DNA, viral DNA, or exogenous DNA.

87. A cell comprising

a first foreign nucleic acid encoding two or more RNAs with each RNA being complementary to an adjacent site in a DNA target nucleic acid, and

a second foreign nucleic acid encoding at least one RNA guided DNA binding protein nickase, and wherein the two or more RNAs and the at least one RNA guided DNA binding protein nickase are members of a co-localization complex for the DNA target nucleic acid.

88. The cell of claim 87 wherein the cell is a eukaryotic cell.

89. The cell of claim 87 wherein the cell is a yeast cell, a plant cell or an animal cell.

90. The cell of claim 87 wherein the RNA includes between about 10 to about 500 nucleotides.

91. The cell of claim 87 wherein the RNA includes between about 20 to about 100 nucleotides.

92. The cell of claim 87 wherein the target nucleic acid is associated with a disease or detrimental condition.

5 93. The cell of claim 87 wherein the two or more RNAs are guide RNAs.

94. The cell of claim 87 wherein the two or more RNAs are tracrRNA-crRNA fusions.

95. The cell of claim 87 wherein the DNA target nucleic acid is genomic DNA, mitochondrial DNA, viral DNA, or exogenous DNA.

96. A method of altering a DNA target nucleic acid in a cell comprising
10 introducing into the cell a first foreign nucleic acid encoding two or more RNAs with each RNA being complementary to an adjacent site in the DNA target nucleic acid,

introducing into the cell a second foreign nucleic acid encoding at least one RNA guided DNA binding protein nickase, and

wherein the two or more RNAs and the at least one RNA guided DNA binding protein
15 nickase are expressed and wherein the at least one RNA guided DNA binding protein nickase co-localizes with the two or more RNAs to the DNA target nucleic acid and nicks the DNA target nucleic acid resulting in two or more adjacent nicks, and

wherein the two or more adjacent nicks are on different strands of the double stranded DNA and create double stranded breaks resulting in fragmentation of the target nucleic acid
20 thereby preventing expression of the target nucleic acid.

97. The method of claim 49 wherein the RNA guided nuclease null DNA binding protein is an RNA guided nuclease null DNA binding protein of a Type II CRISPR system.

98. The method of claim 63 wherein the at least one RNA guided DNA binding protein nickase is an RNA guided DNA binding protein nickase of a Type II CRISPR system.

25 99. The cell of claim 73 wherein the RNA guided nuclease null DNA binding protein is an RNA guided nuclease null DNA binding protein of a Type II CRISPR system.

100. The cell of claim 87 wherein the at least one RNA guided DNA binding protein nickase is an RNA guided DNA binding protein nickase of a Type II CRISPR system.

101. The method of claim 96 wherein the at least one RNA guided DNA binding
30 protein nickase is an RNA guided DNA binding protein nickase of a Type II CRISPR system.

102. A guide RNA including a spacer sequence having a 1 to 4 base 5' truncation.

103. A guide RNA being a fusion transcript of a crRNA and a tracr RNA wherein the tracr RNA includes between about 64 to about 500 nucleic acids.

104. A guide RNA being a fusion transcript of a crRNA and a tracr RNA wherein the
35 tracr RNA includes between about 65 to about 500 nucleic acids.

105. A guide RNA being a fusion transcript of a crRNA and a tracr RNA wherein the tracr RNA includes between about 66 to about 500 nucleic acids.

106. A guide RNA being a fusion transcript of a crRNA and a tracr RNA wherein the tracr RNA includes between about 67 to about 500 nucleic acids.

5 107. A guide RNA being a fusion transcript of a crRNA and a tracr RNA wherein the tracr RNA includes between about 68 to about 500 nucleic acids.

108. A guide RNA being a fusion transcript of a crRNA and a tracr RNA wherein the tracr RNA includes between about 69 to about 500 nucleic acids.

10 109. A guide RNA being a fusion transcript of a crRNA and a tracr RNA wherein the tracr RNA includes between about 70 to about 500 nucleic acids.

110. A guide RNA being a fusion transcript of a crRNA and a tracr RNA wherein the tracr RNA includes between about 80 to about 500 nucleic acids.

111. A guide RNA being a fusion transcript of a crRNA and a tracr RNA wherein the tracr RNA includes between about 90 to about 500 nucleic acids.

15 112. A guide RNA being a fusion transcript of a crRNA and a tracr RNA wherein the tracr RNA includes between about 100 to about 500 nucleic acids.

FIG. 1A

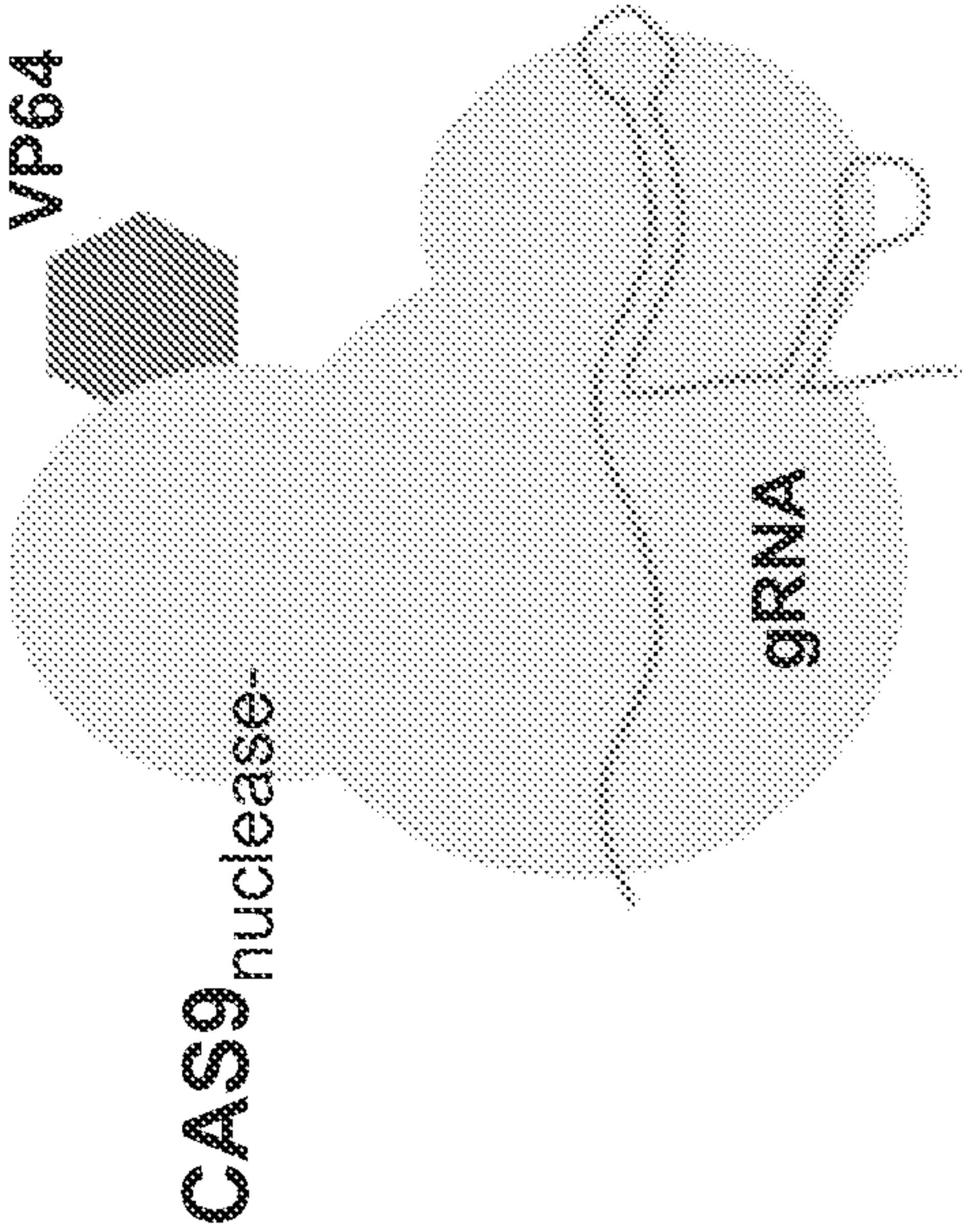


FIG. 1B

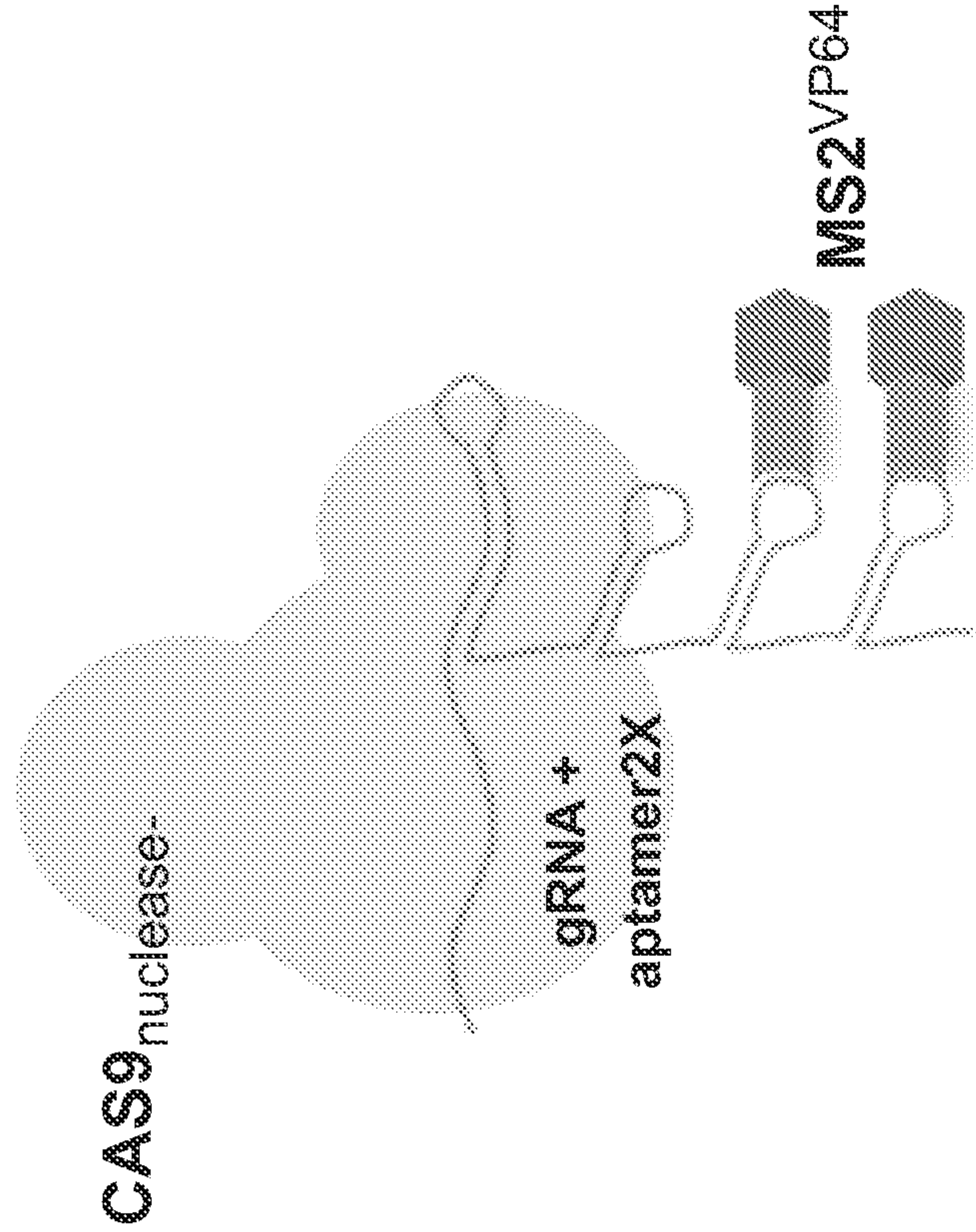


FIG. 1C

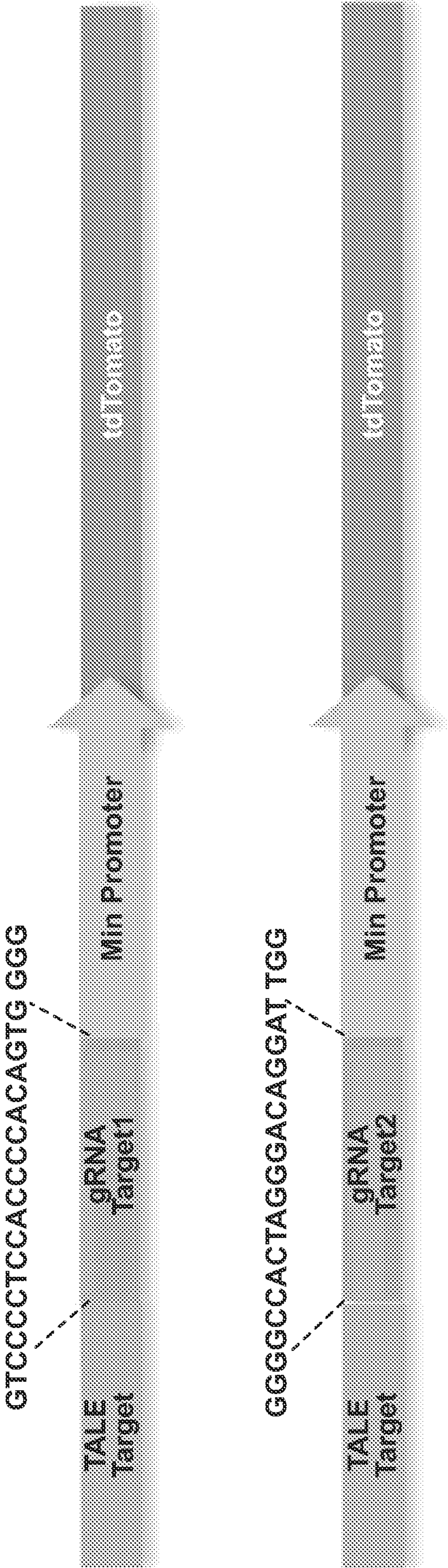


FIG. 1D-1

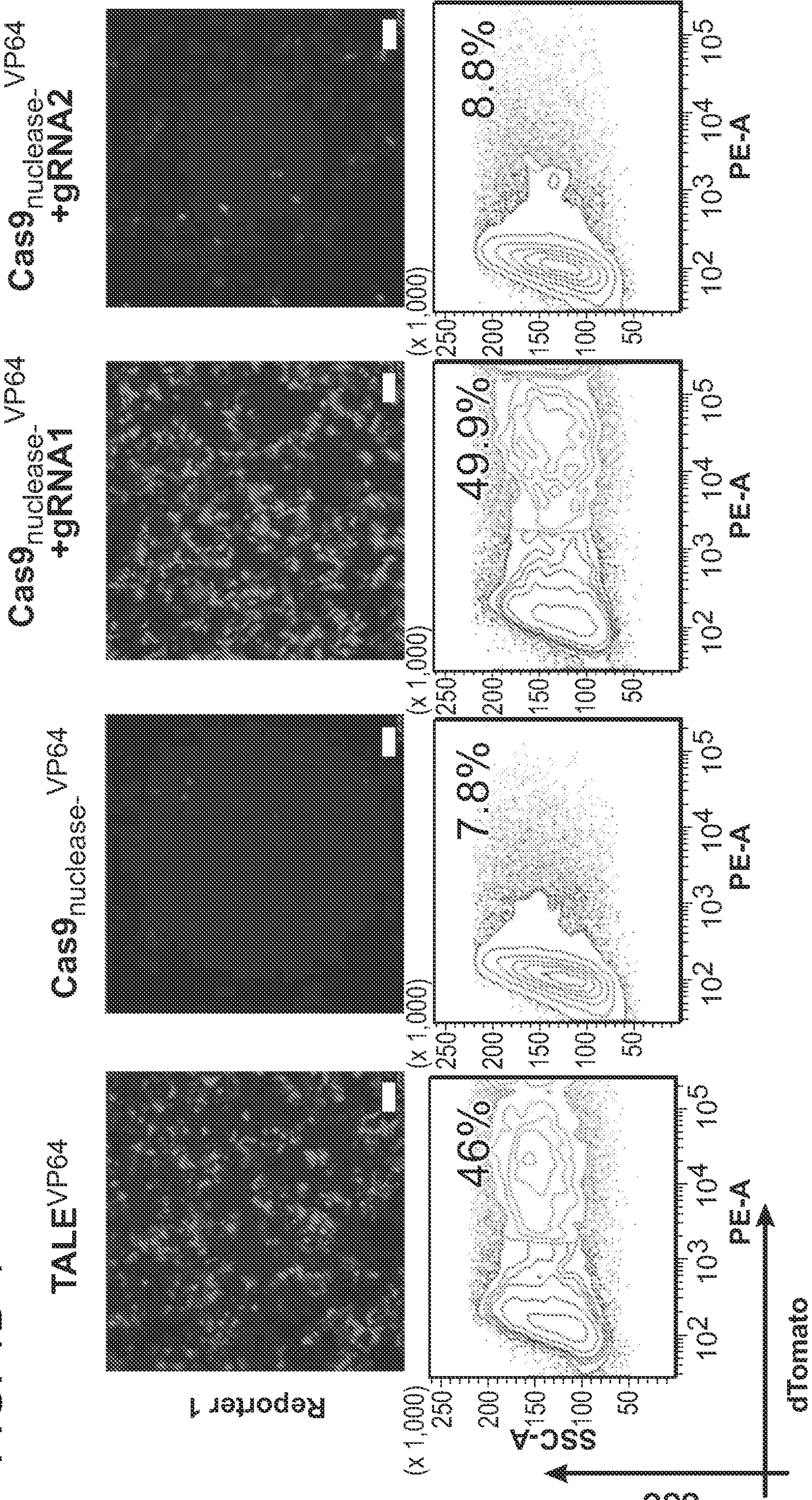


FIG. 1D-2

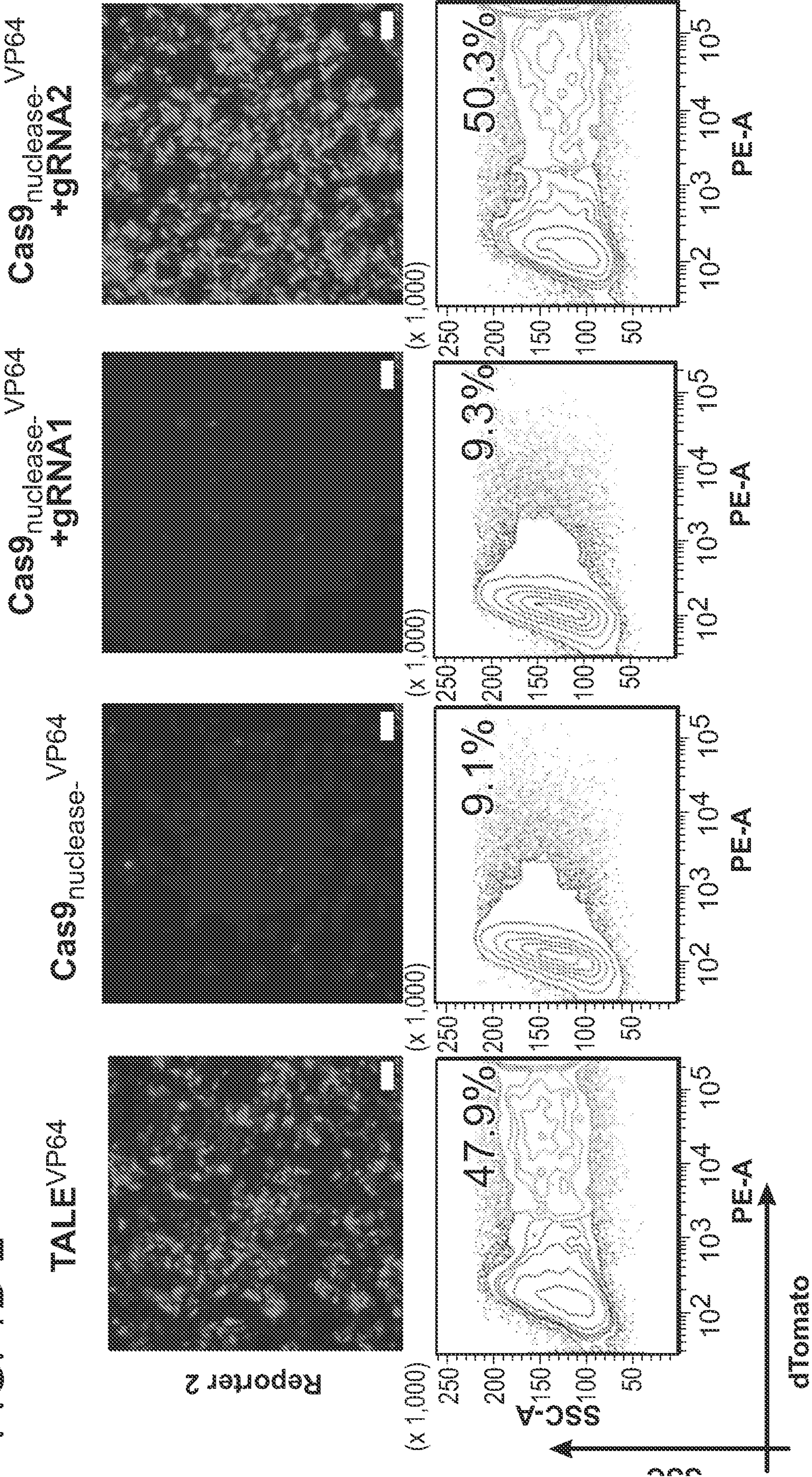


FIG. 1E-1

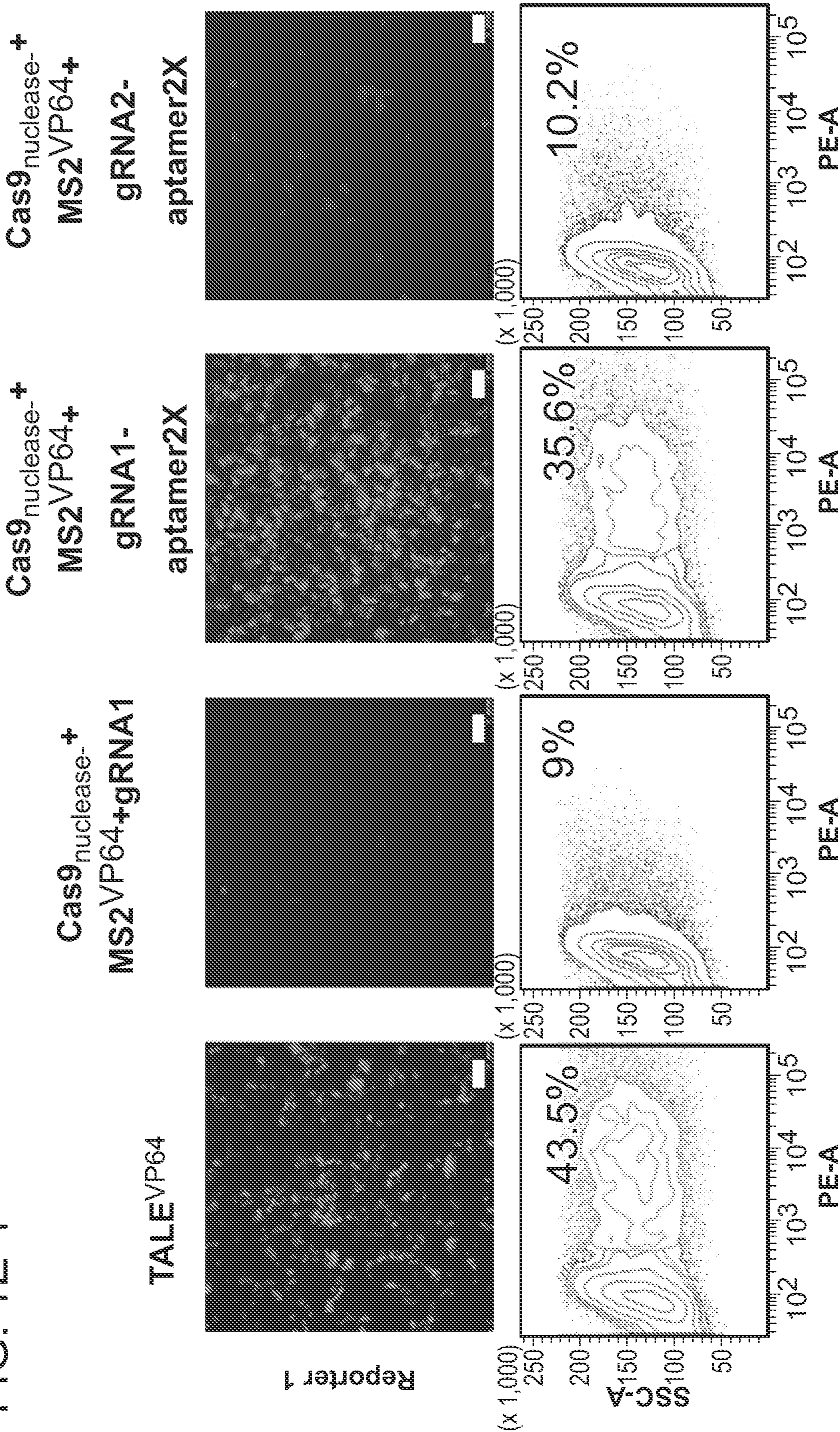


FIG. 1E-2

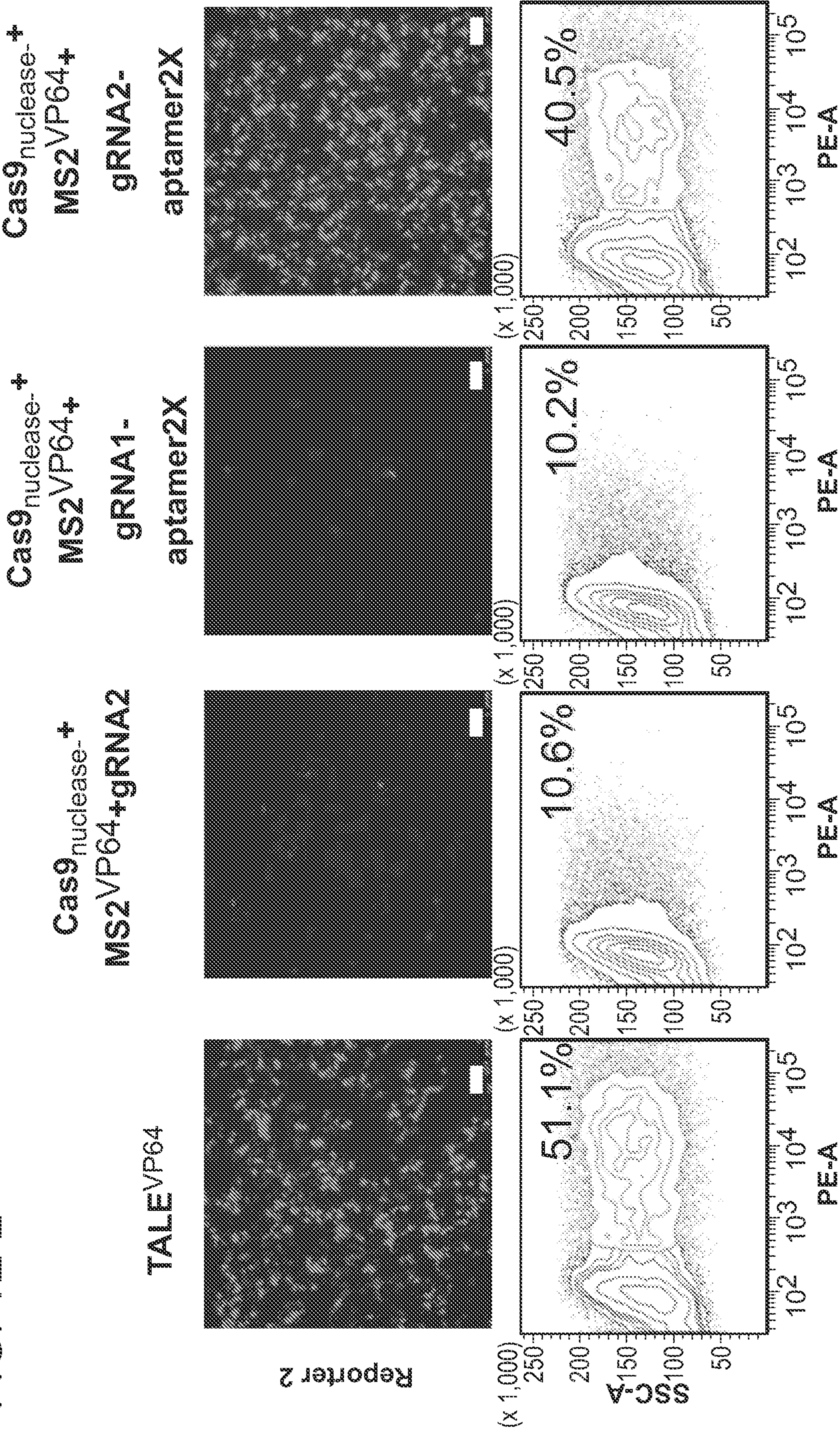


FIG. 1F

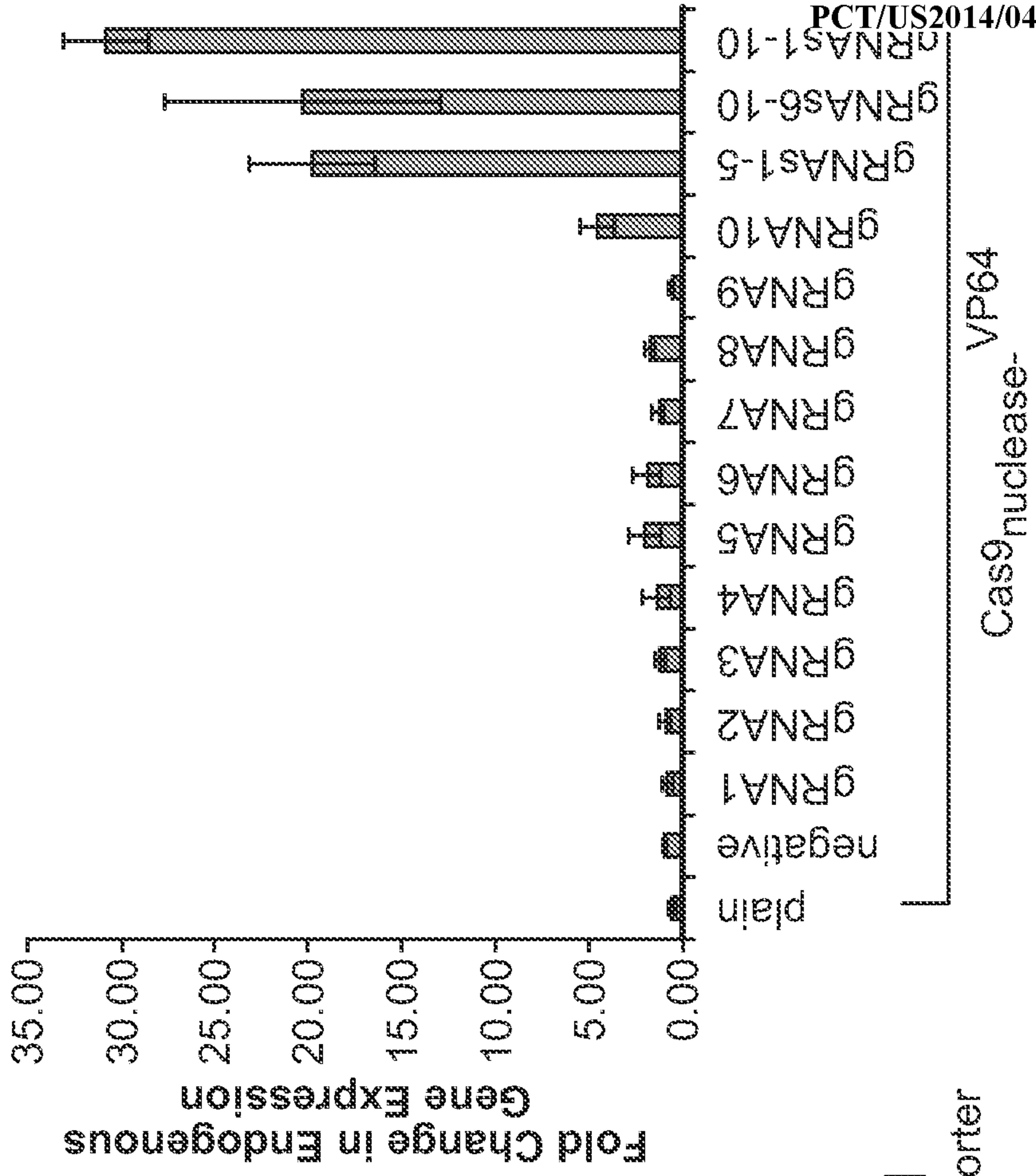
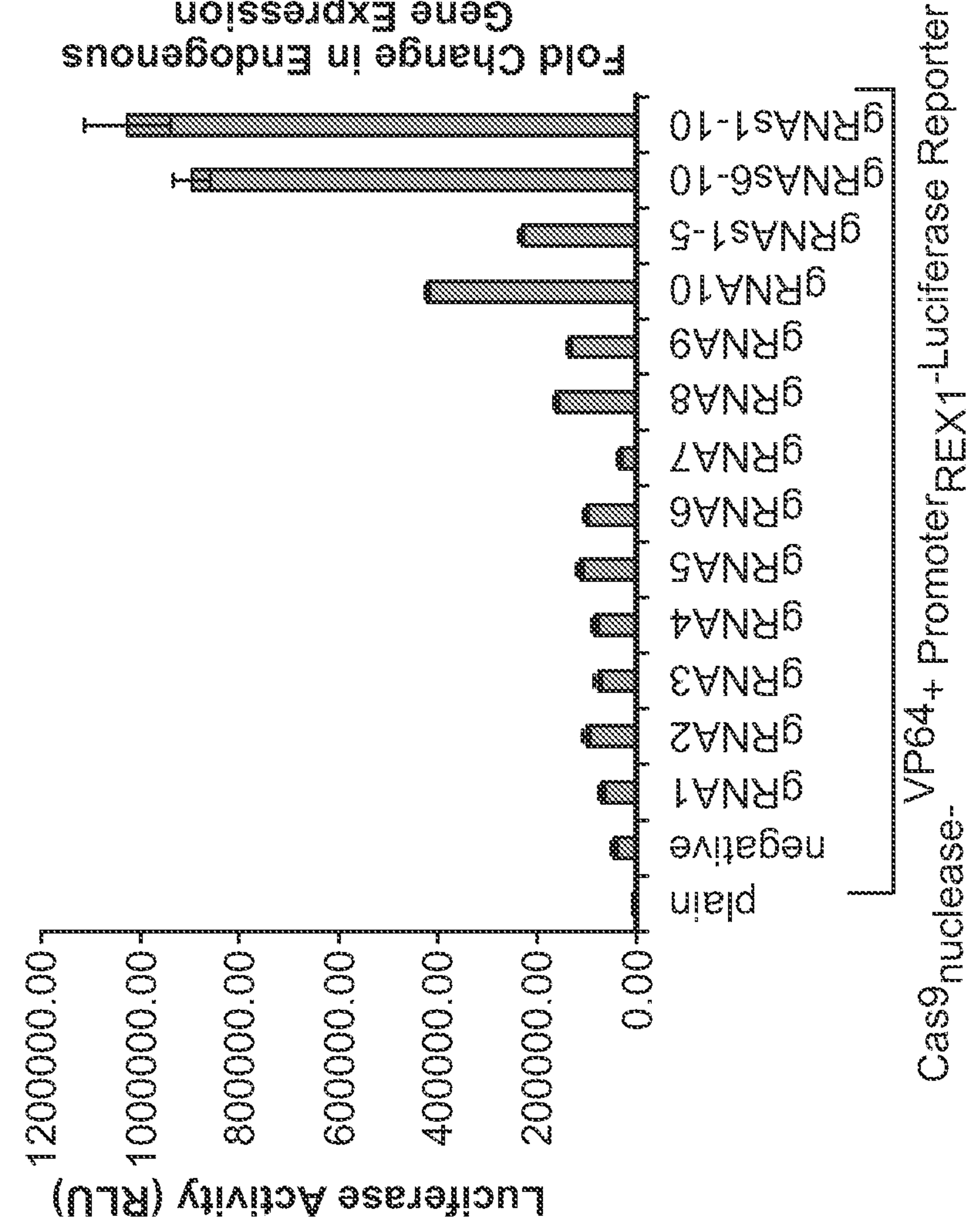
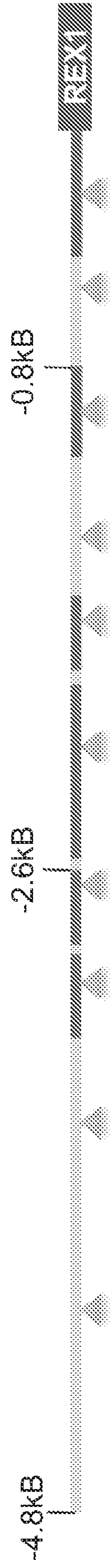
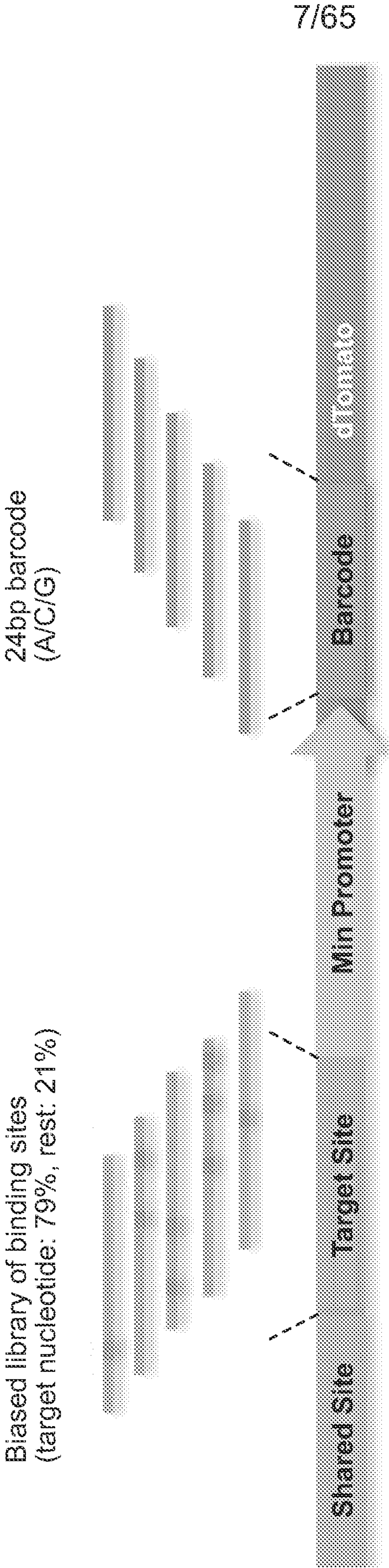


FIG. 2A



- Step 1: Map barcode to corresponding target site in the library
- Step 2: Stimulate library by either a:
- 1) control-TF that binds the shared site; or
 - 2) TALE-TF/gRNA-Cas9-TF (target-TF) that binds the target site.
- Step 3: Perform RNAseq and determine expressed barcodes for each.
- Step 4: Map back expressed barcodes to corresponding binding sites.
- Step 5: Compute relative enrichment of target-TF vs. control-TF barcodes.

8/65

FIG. 2B

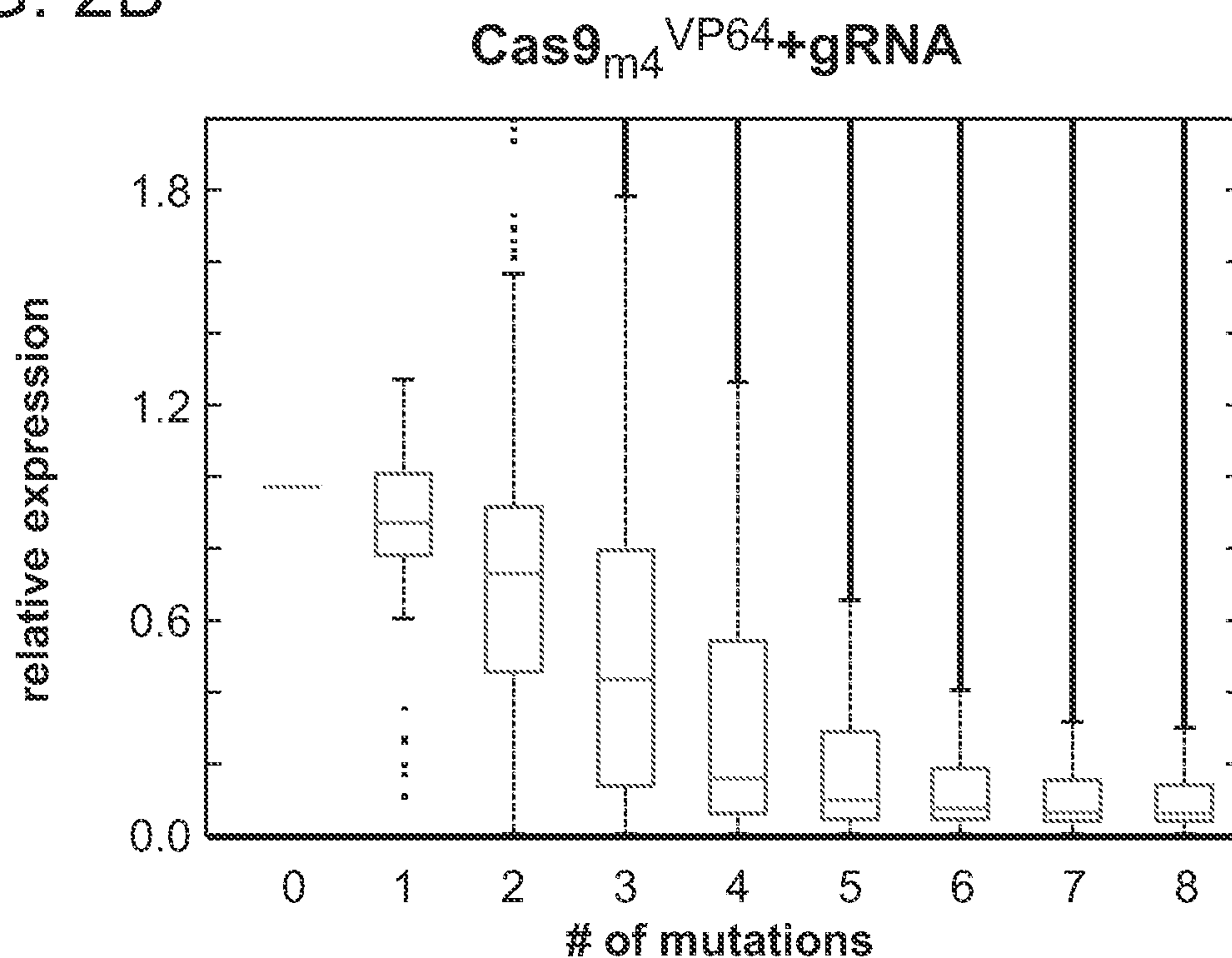
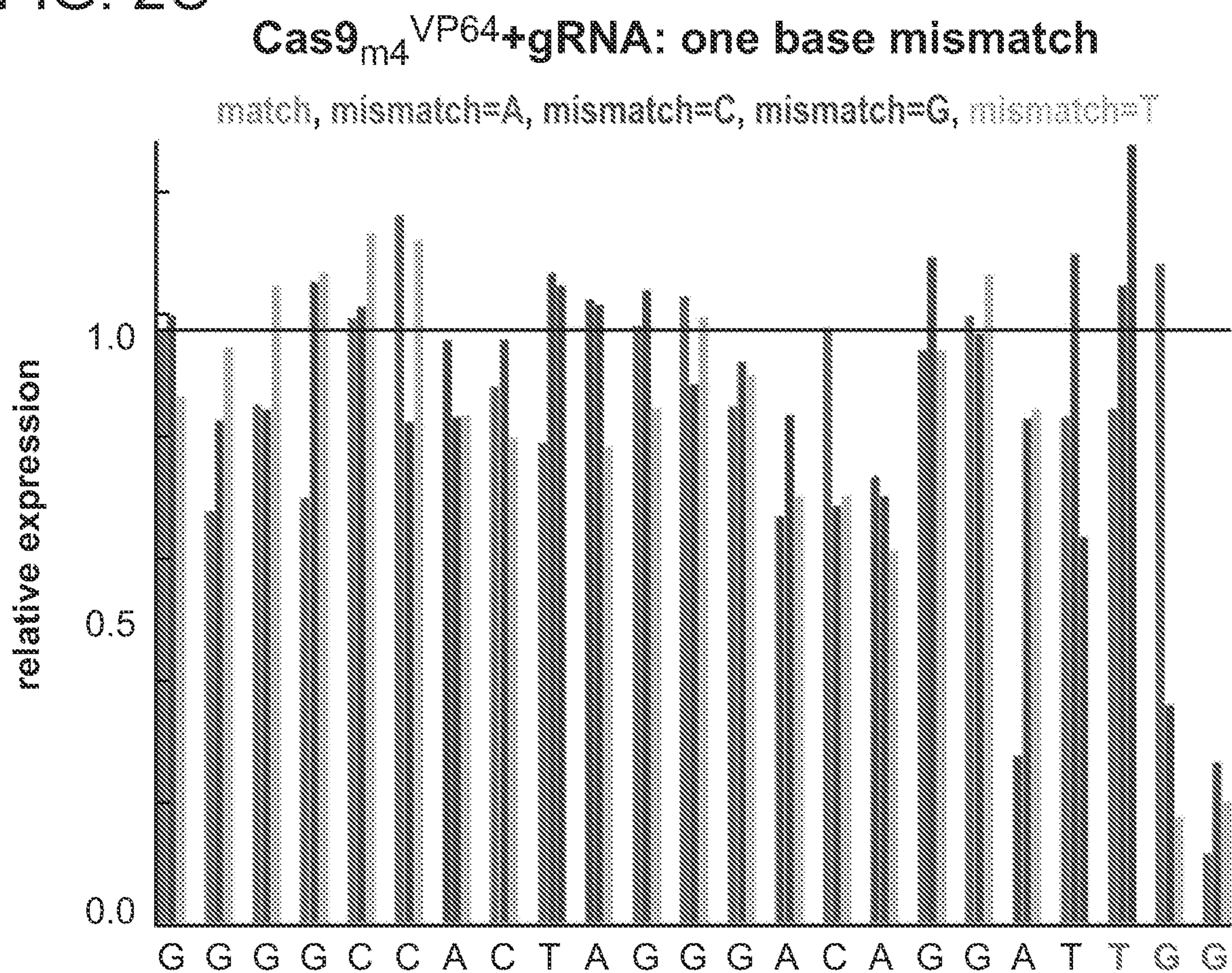


FIG. 2C



9/65

FIG. 2D

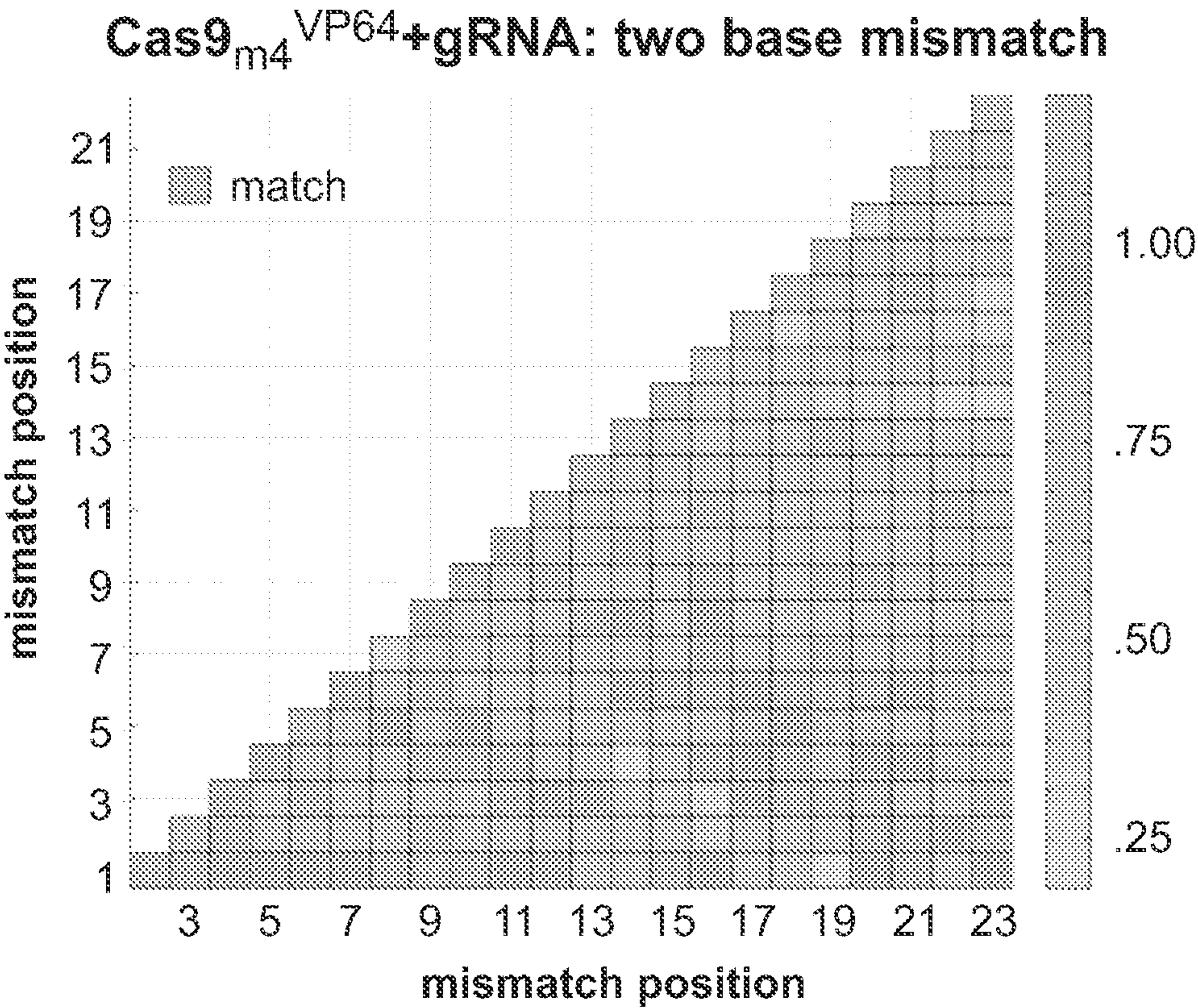
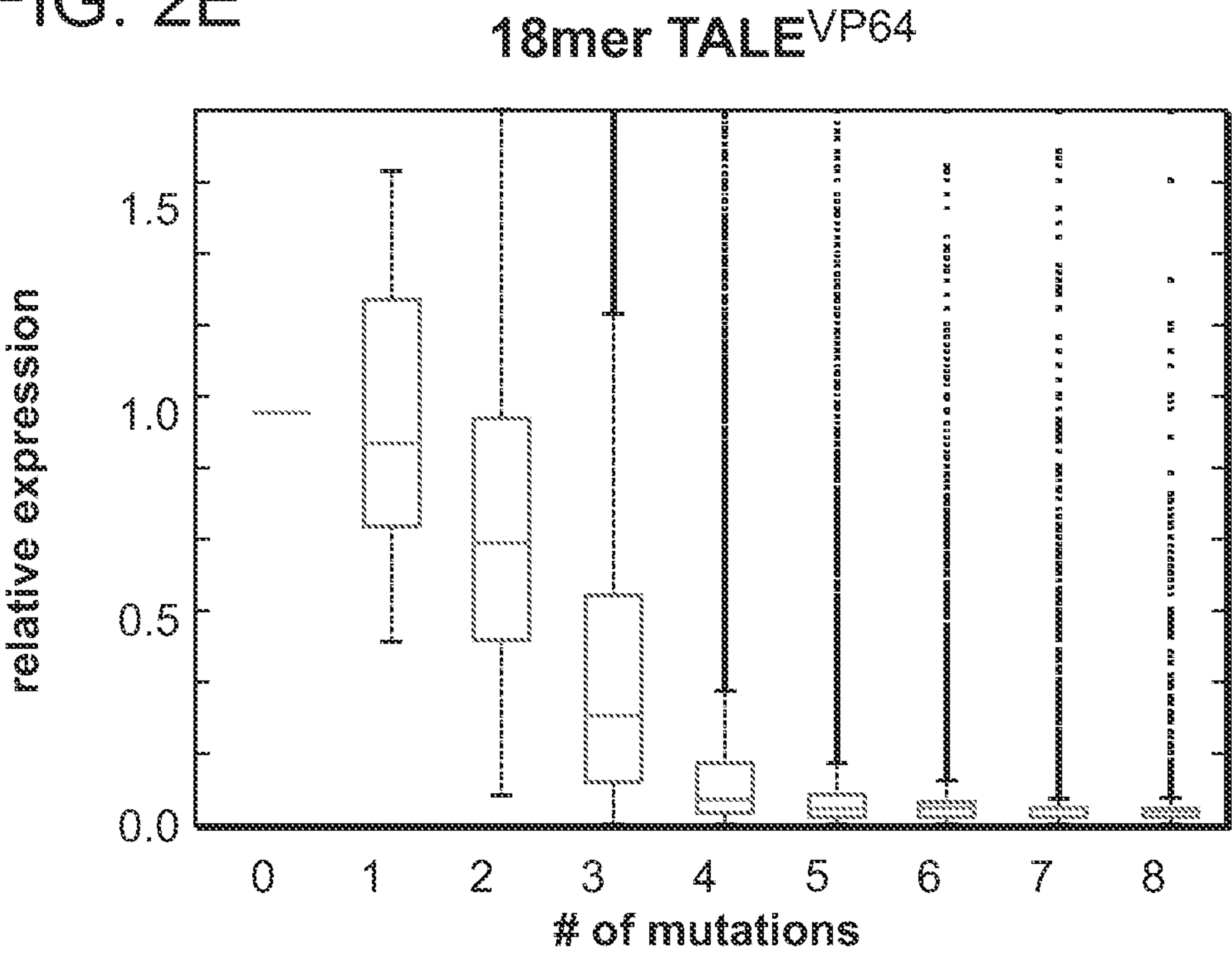


FIG. 2E



10/65

FIG. 2F

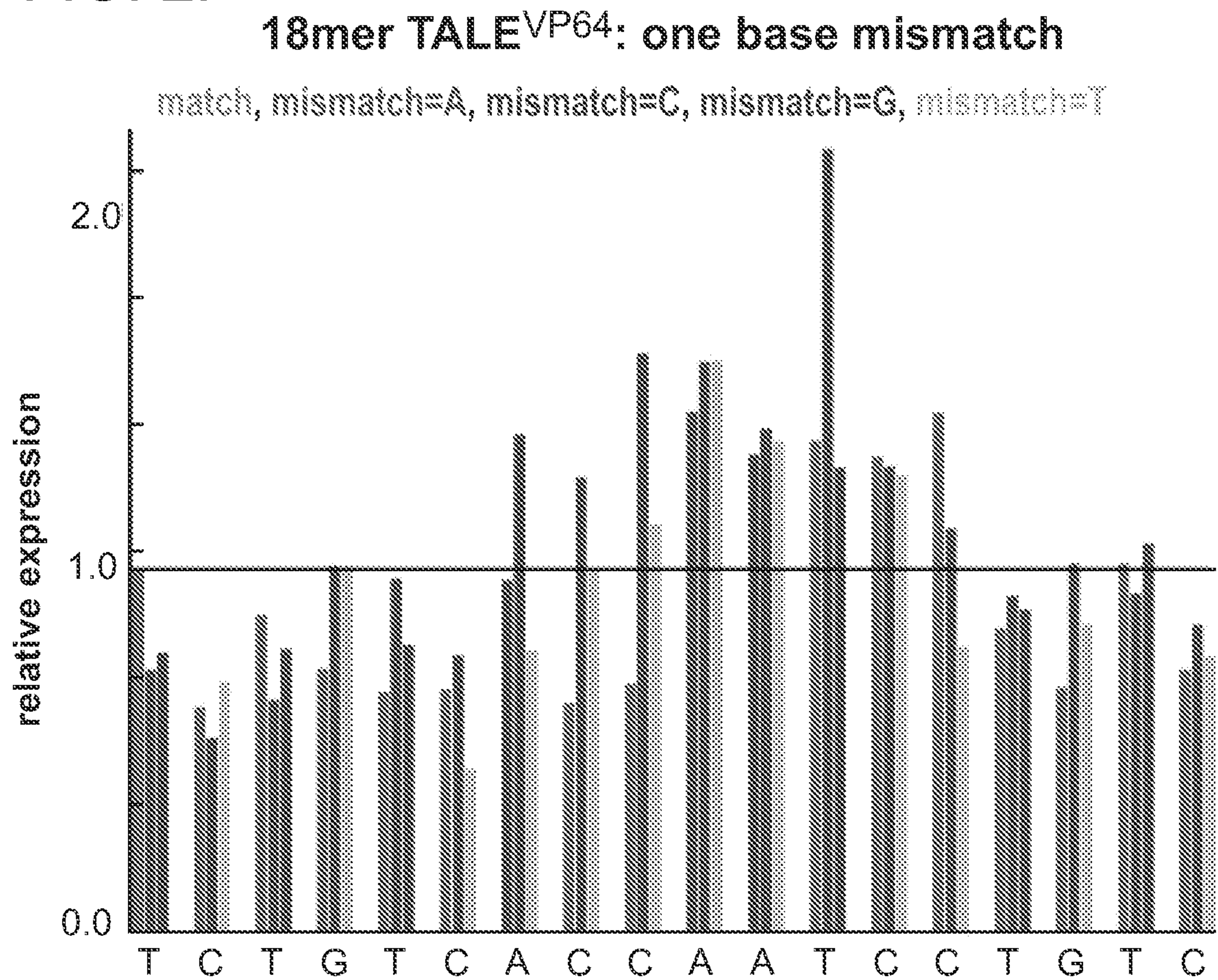


FIG. 2G

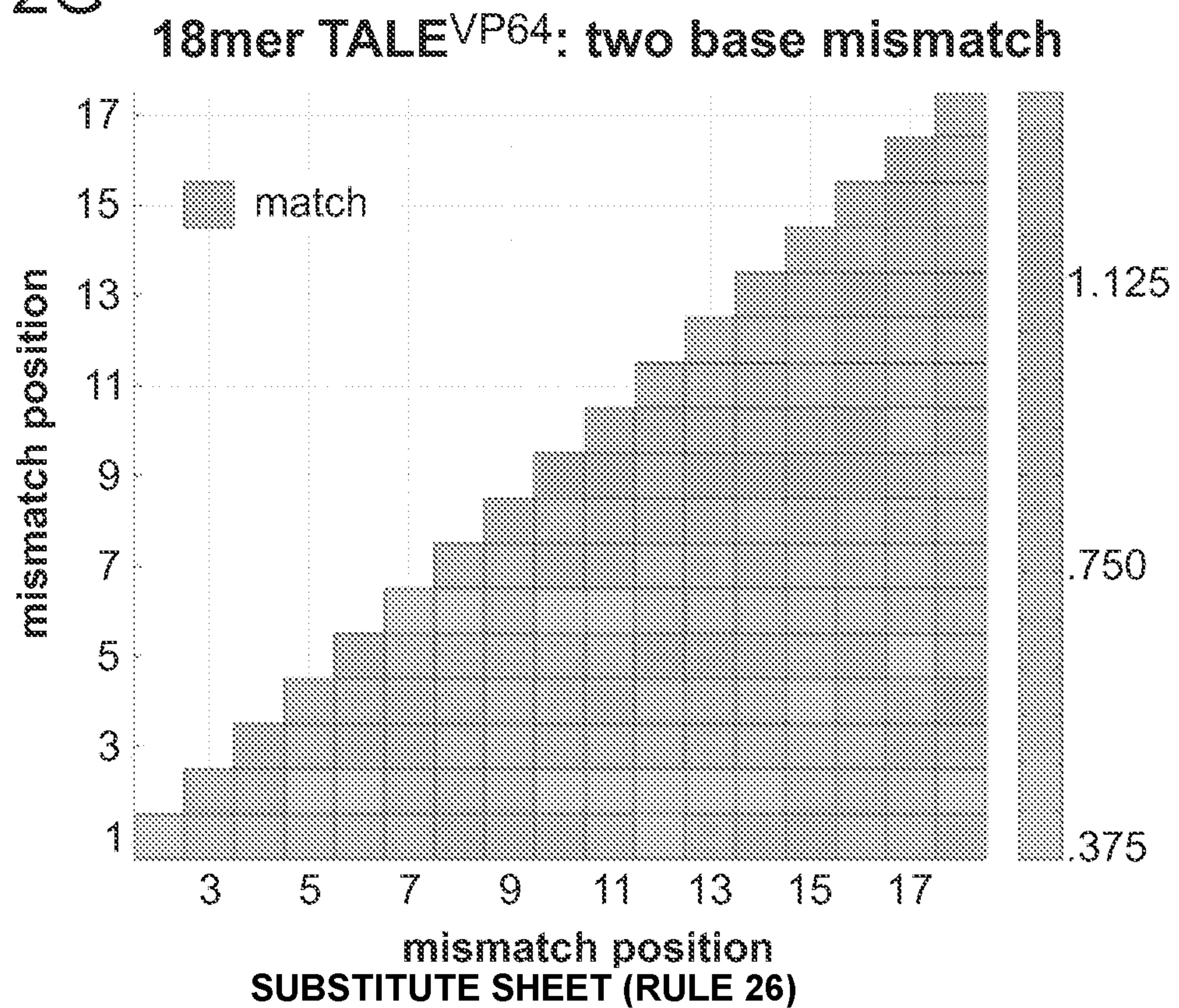


FIG. 3A

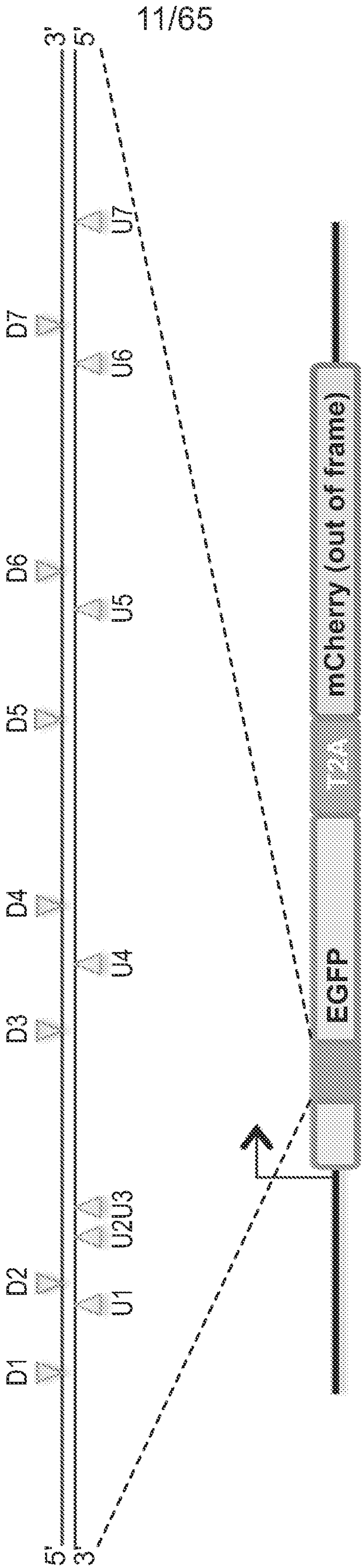
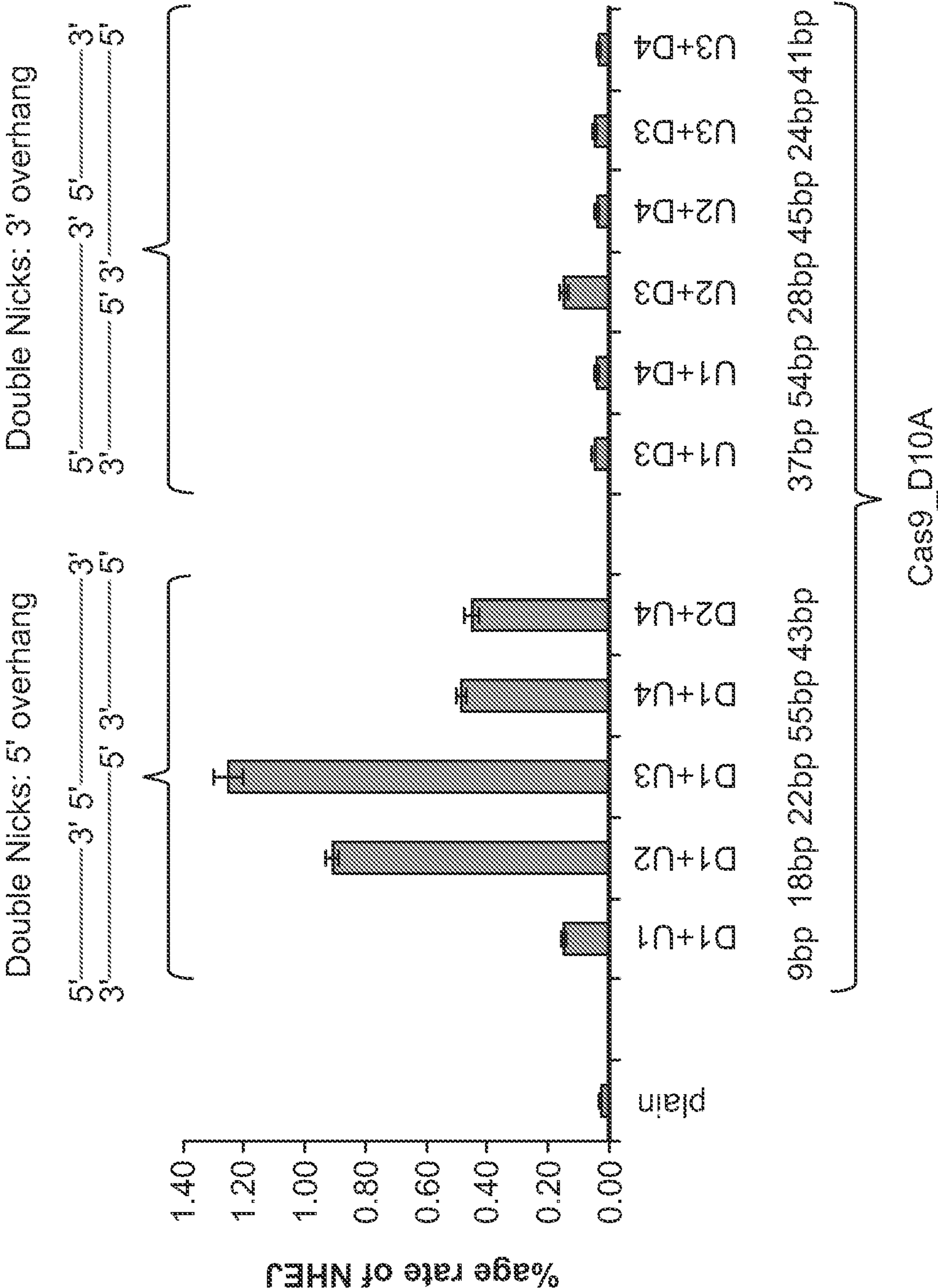
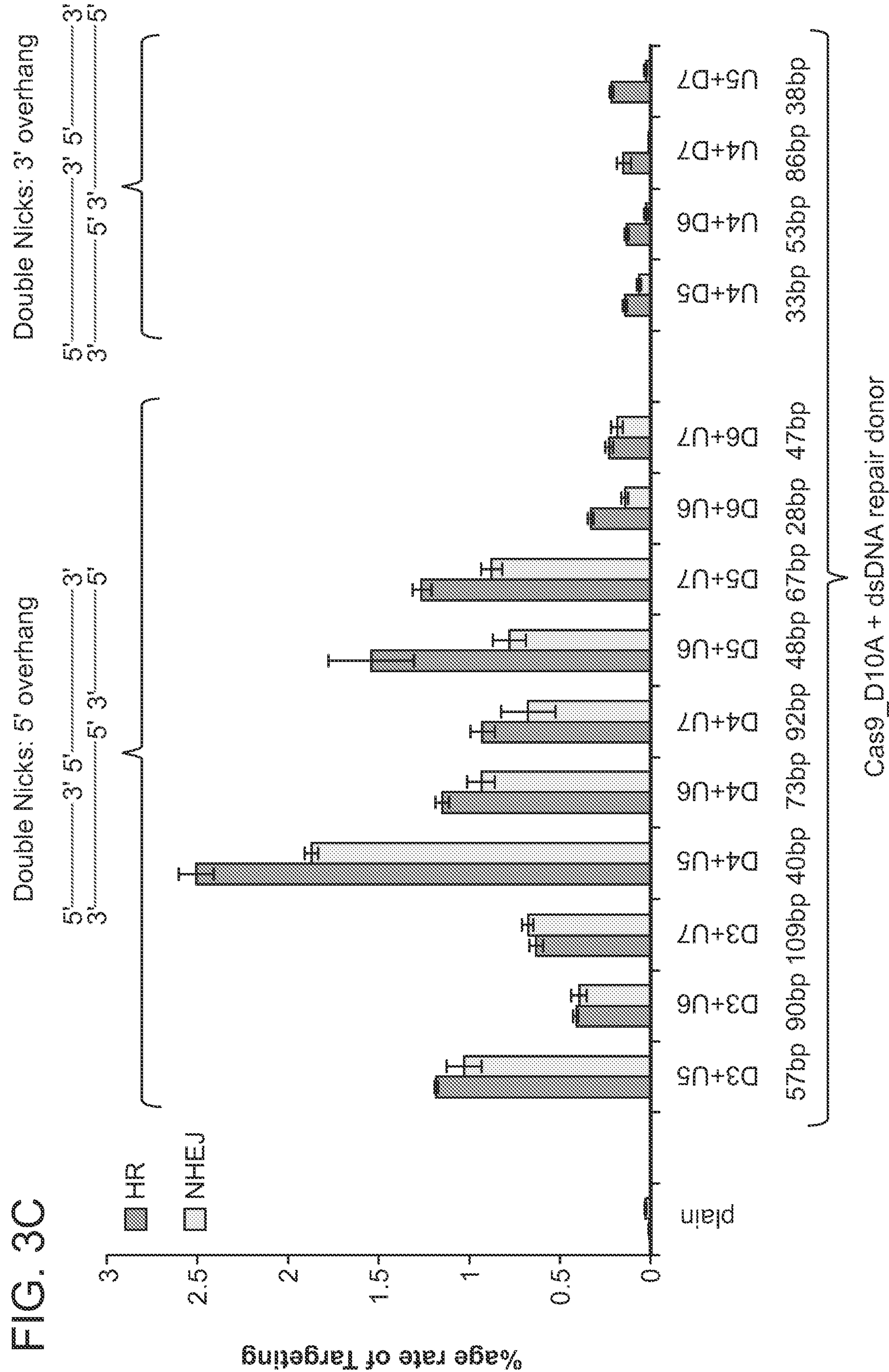


FIG. 3B





Name	Mutations
Cas9	wild-type
Cas9 _{m1}	D10A
Cas9 _{m2}	D10A+H840A
Cas9 _{m3}	D10A+D839A+H840A
Cas9 _{m4}	D10A+D839A+H840A+N863A

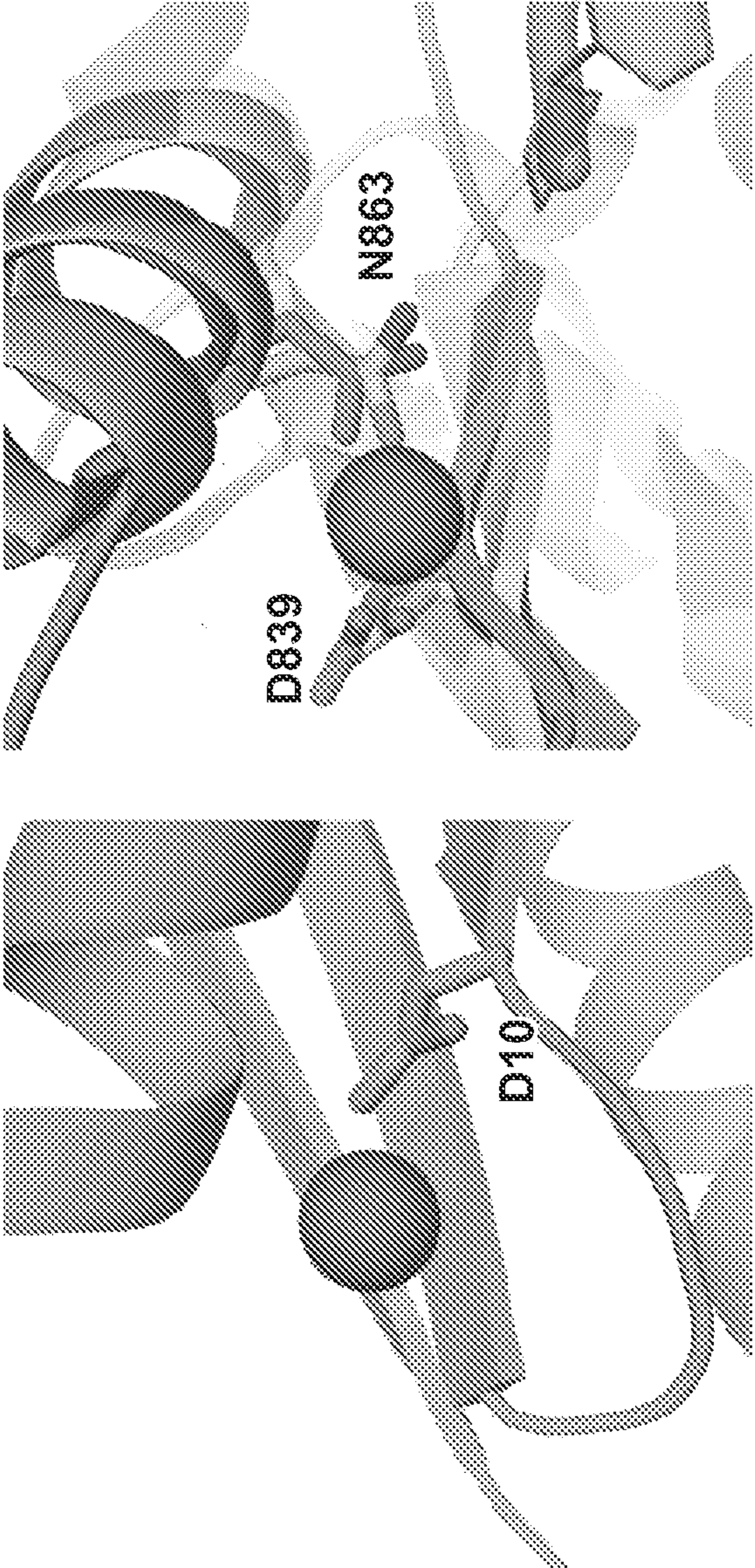


FIG. 4A

FIG. 4B

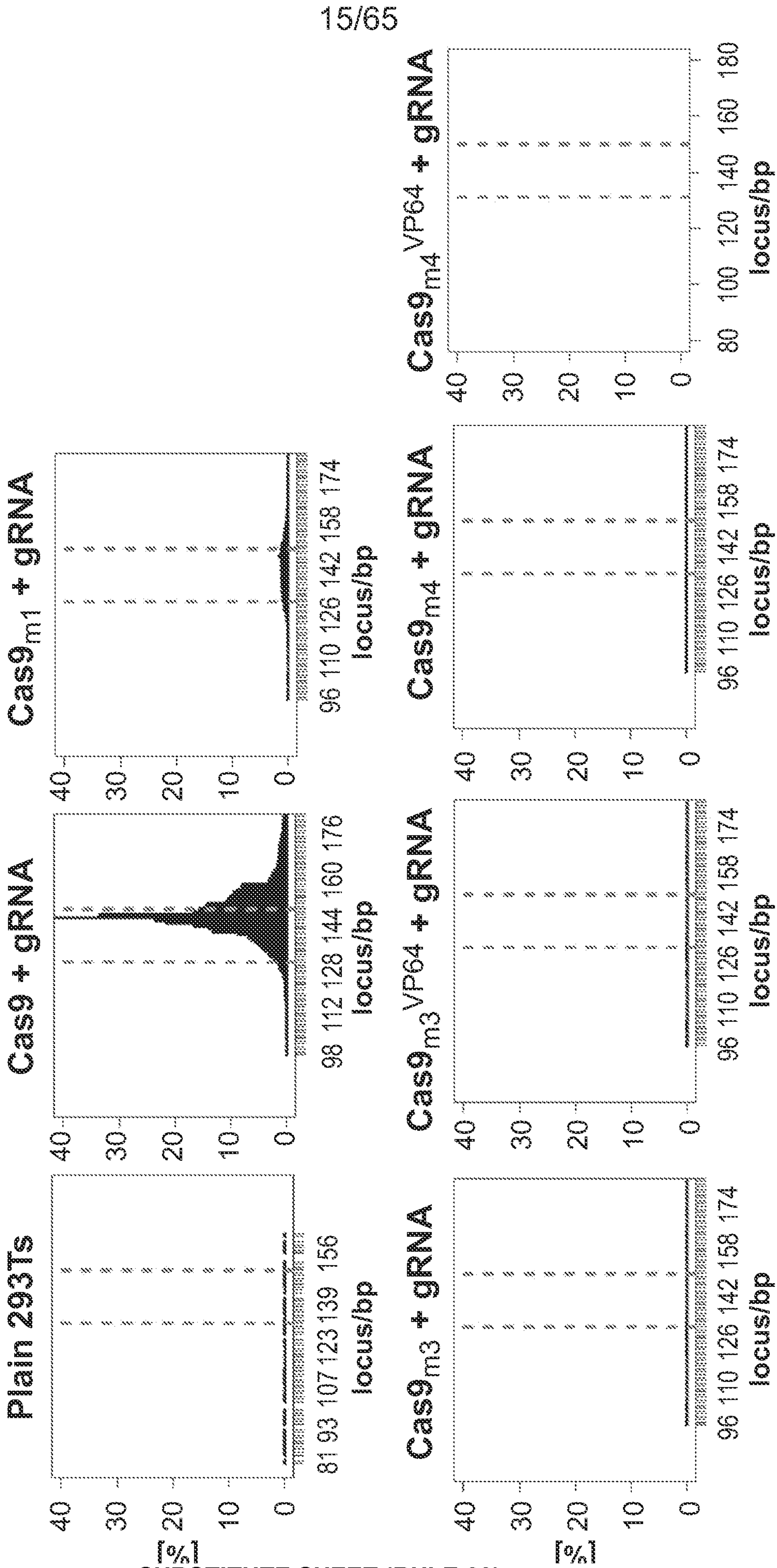
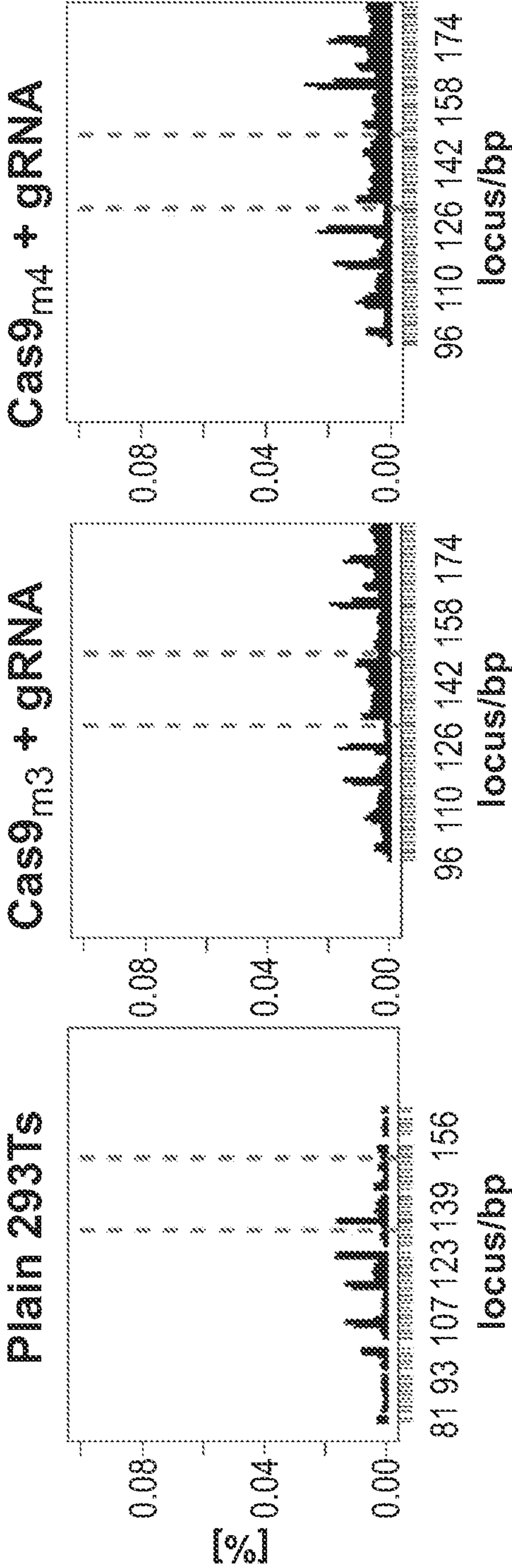


FIG. 4C



LA
G
L

gRNA Target

TAA TACTTTTATCTGTCCTCCACCCACAGTGGGCGCTAGGGGCGGATGGTTGGTGACAGAAAGCCCC

Score Target

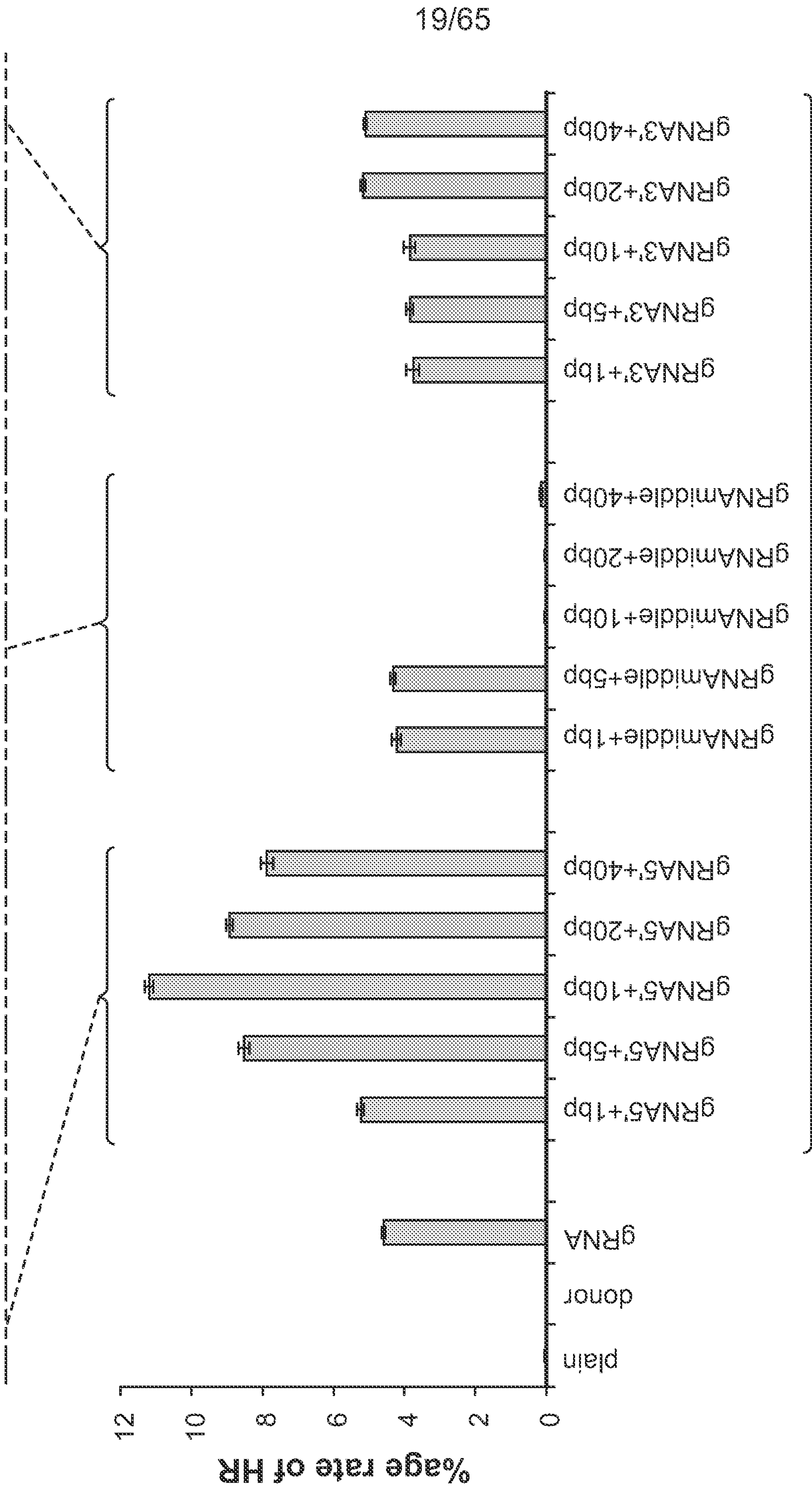
repair donor

DNA break stimulates HR

17/65

SUBSTITUTE SHEET (RULE 26)





donor + hCas9

FIG. 5B-2

FIG. 6A

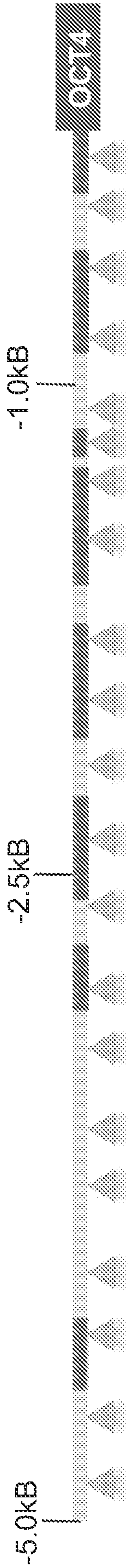


FIG. 6B

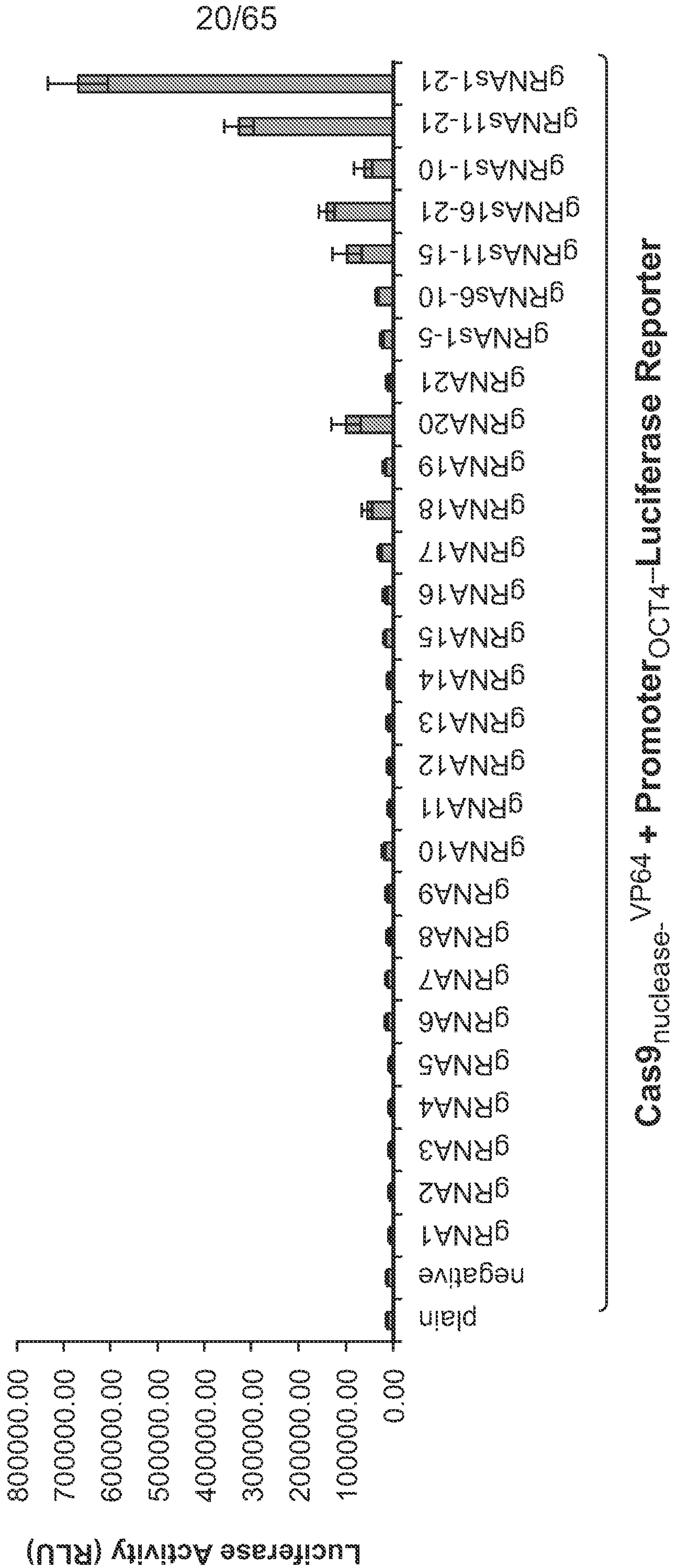


FIG. 6C

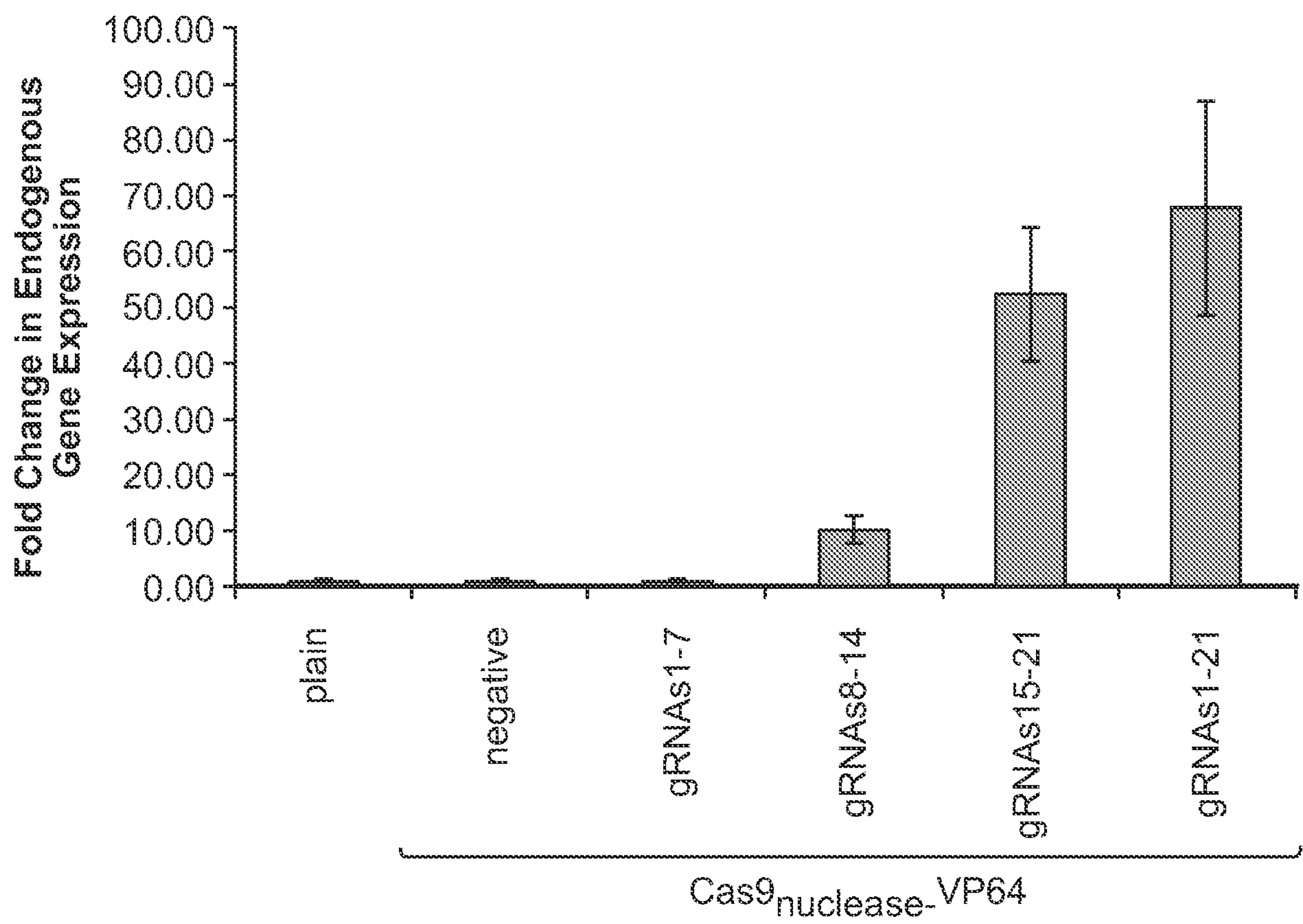


FIG. 7A

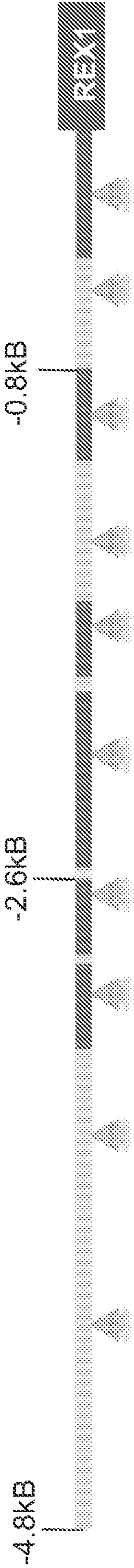


FIG. 7B

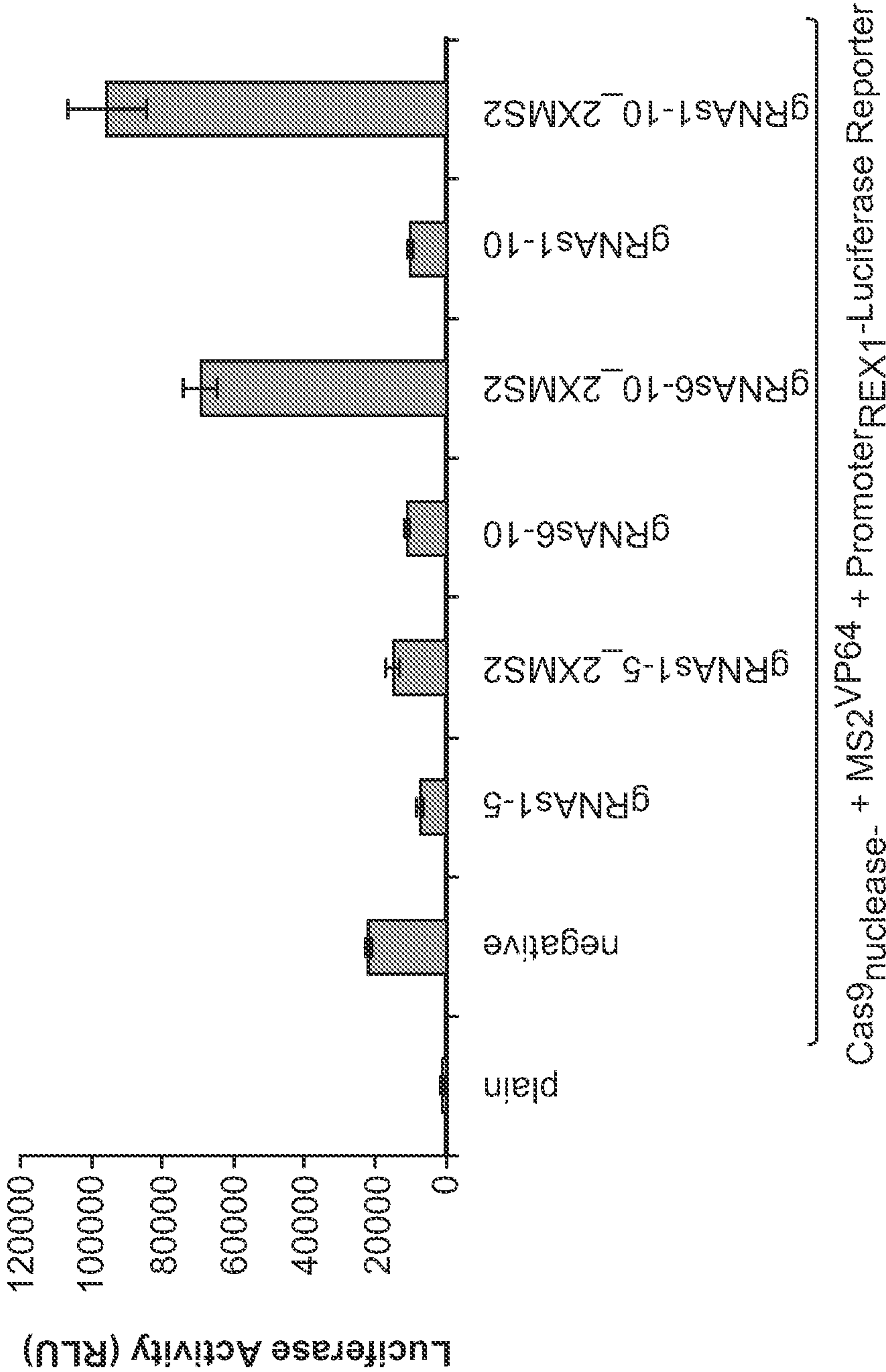
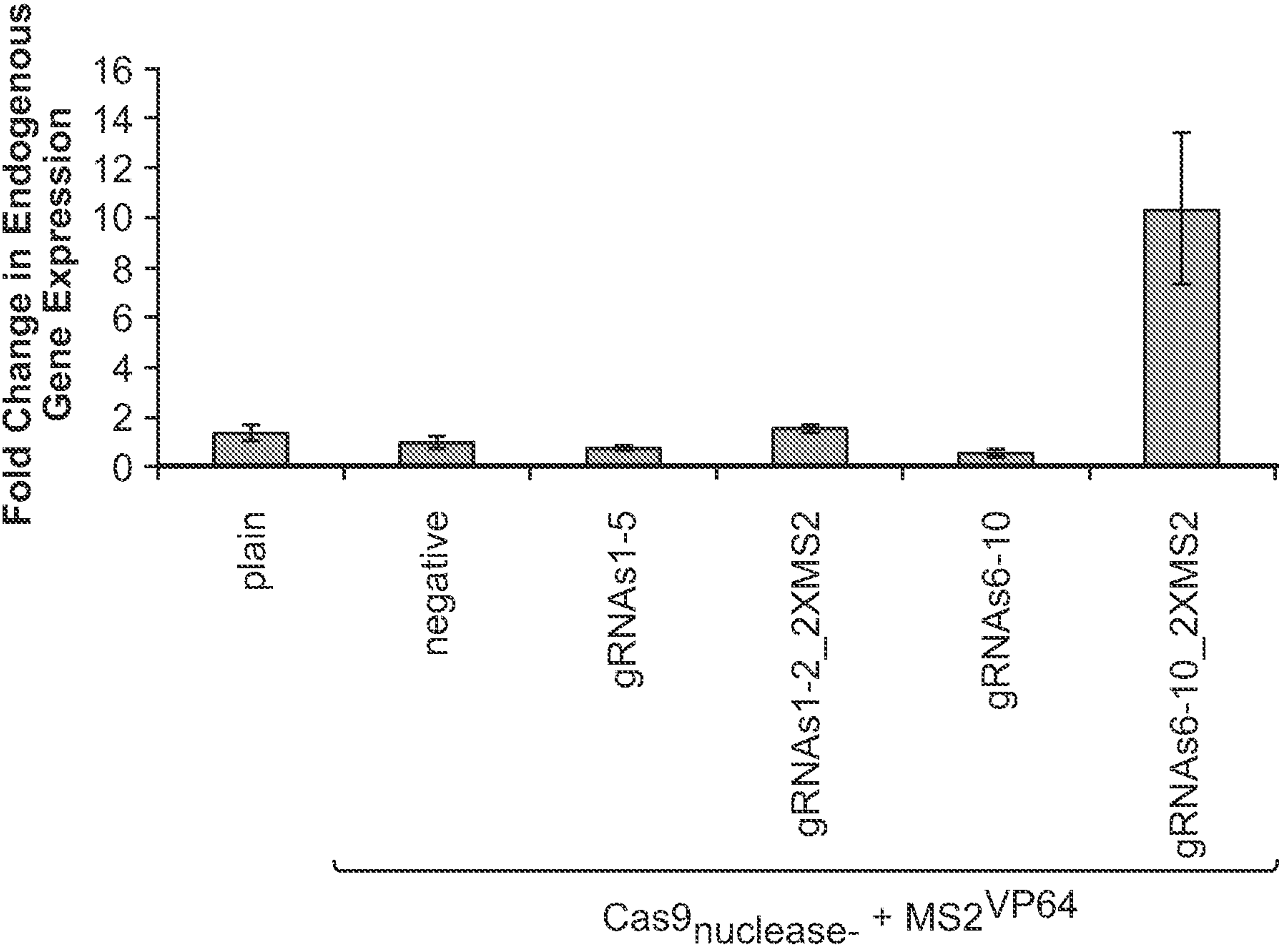
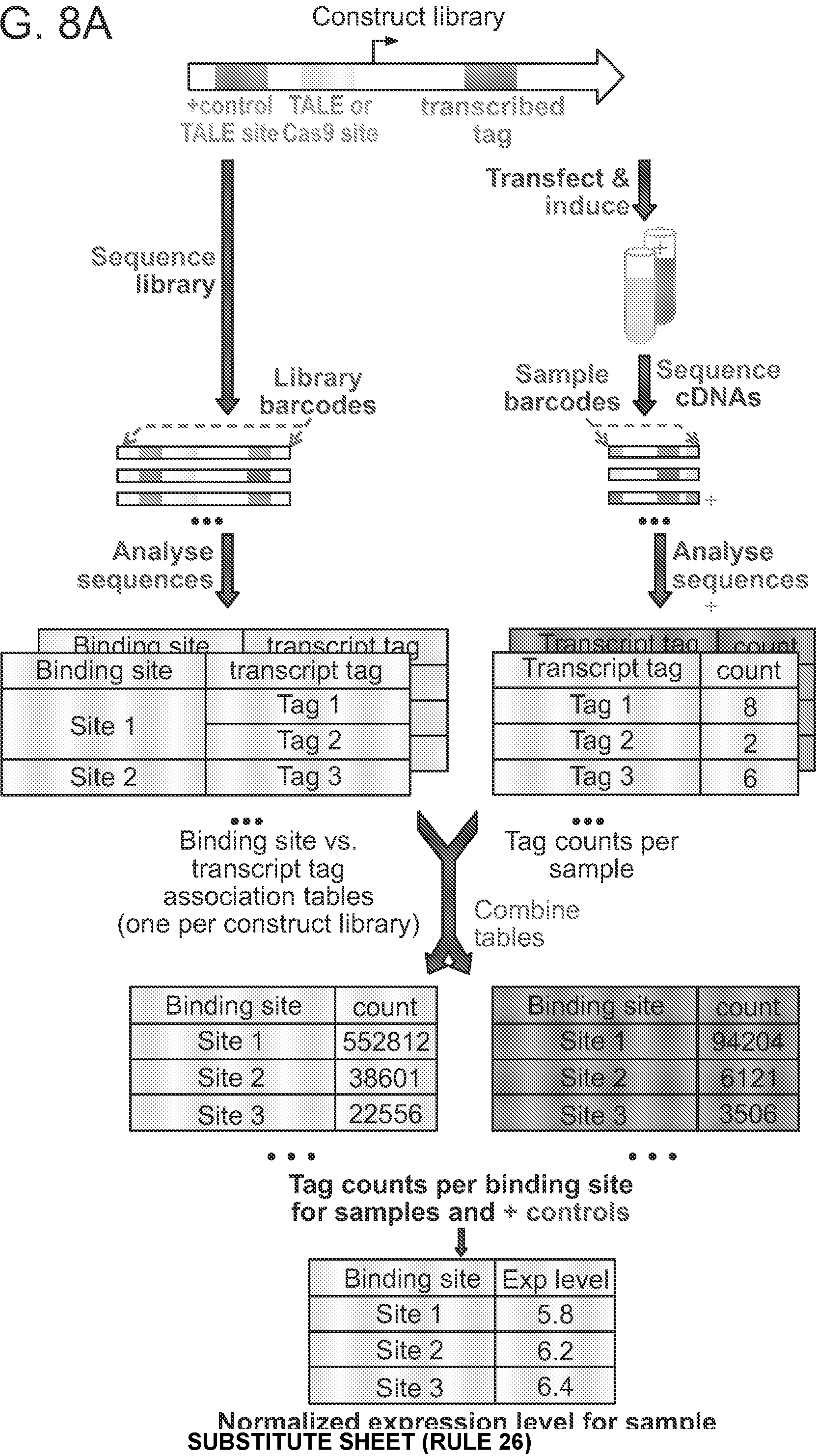


FIG. 7C



24/65

FIG. 8A



25/65

FIG. 8B

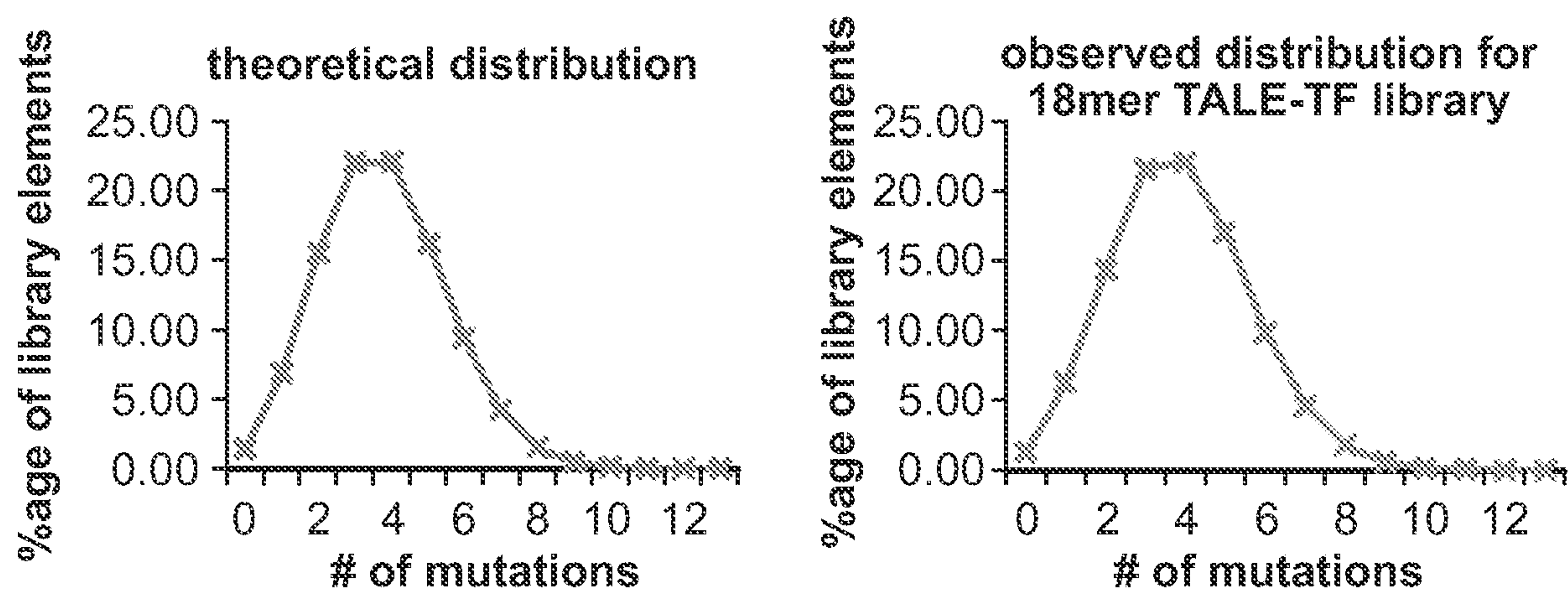


FIG. 8C

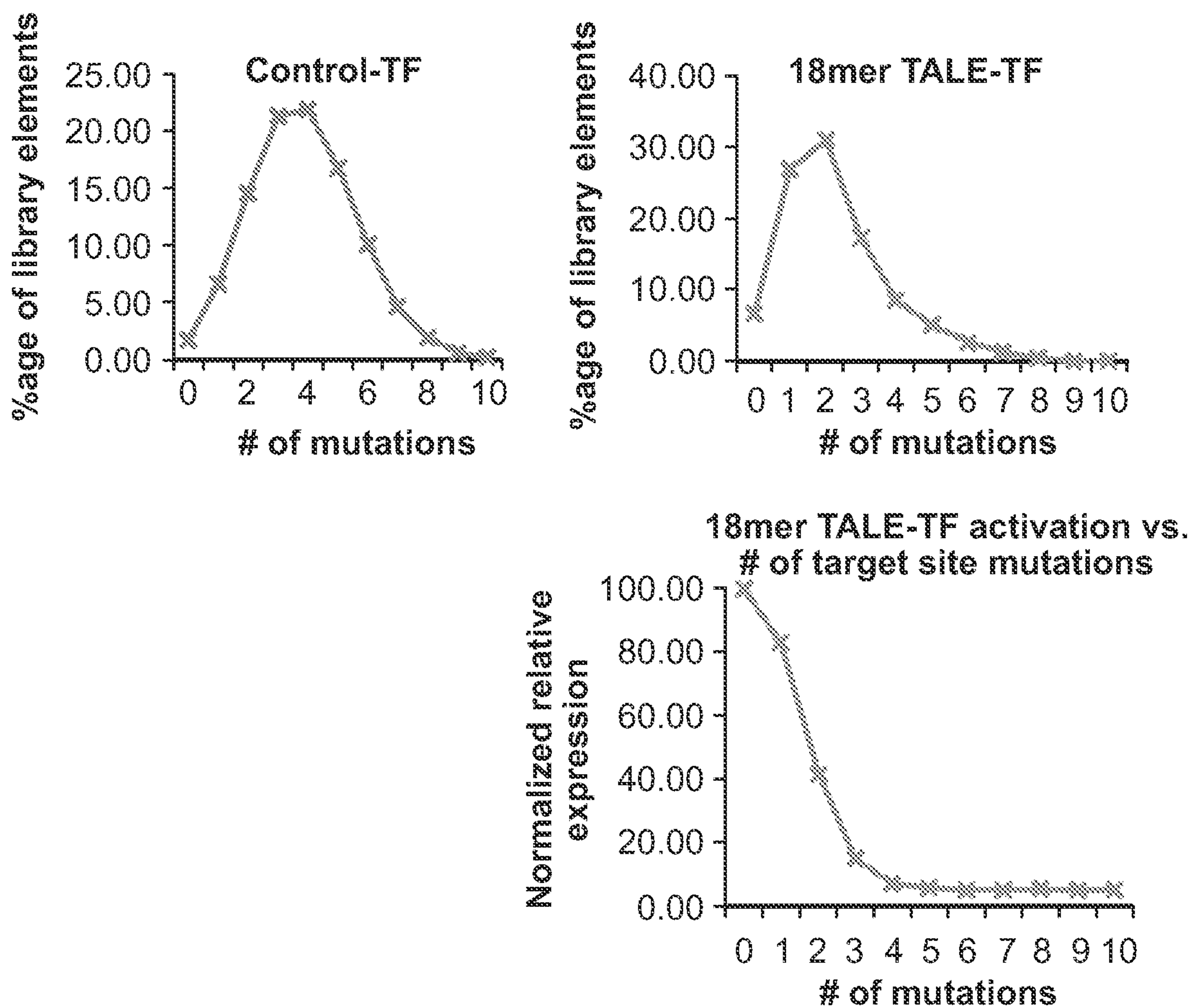


FIG. 9A

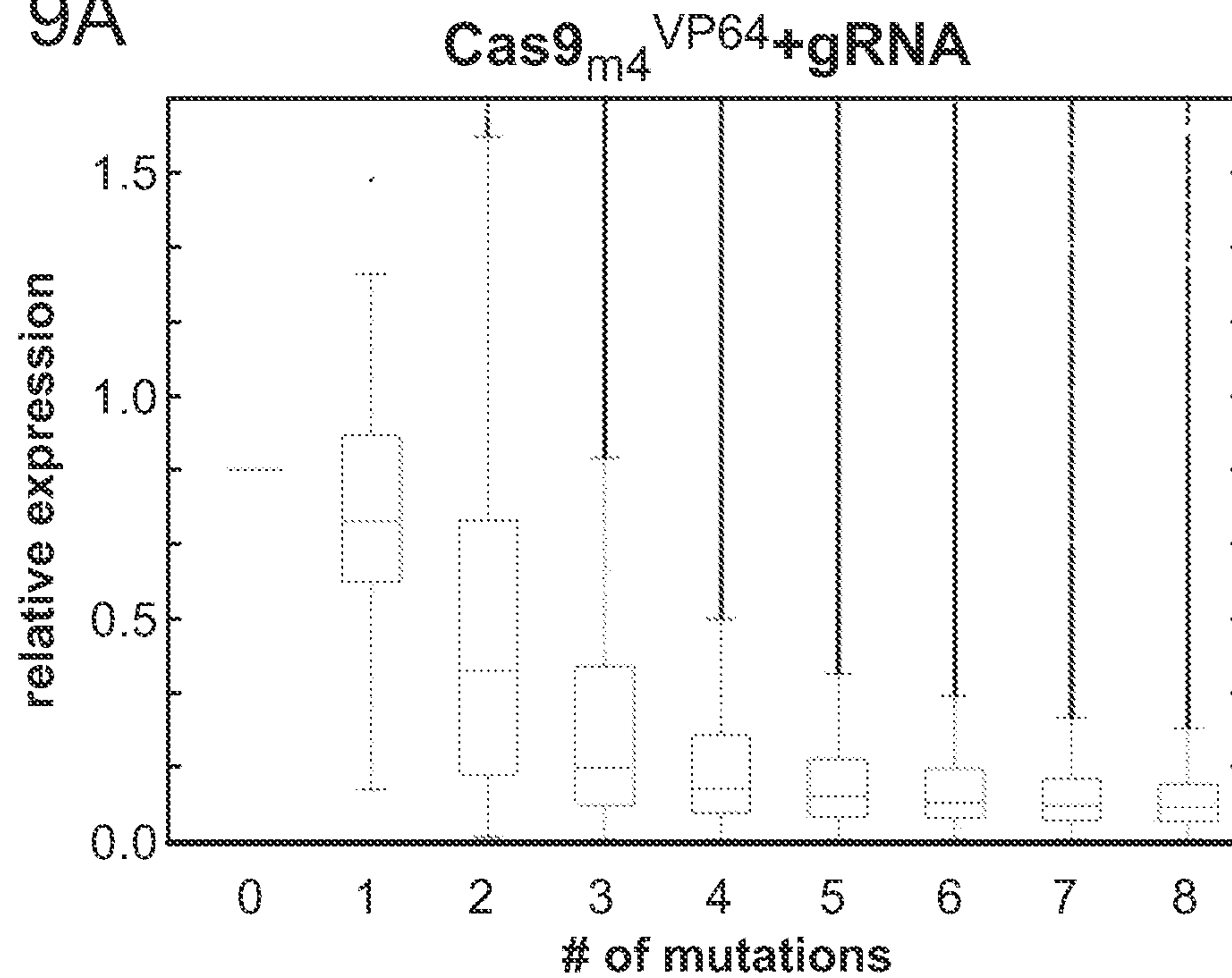
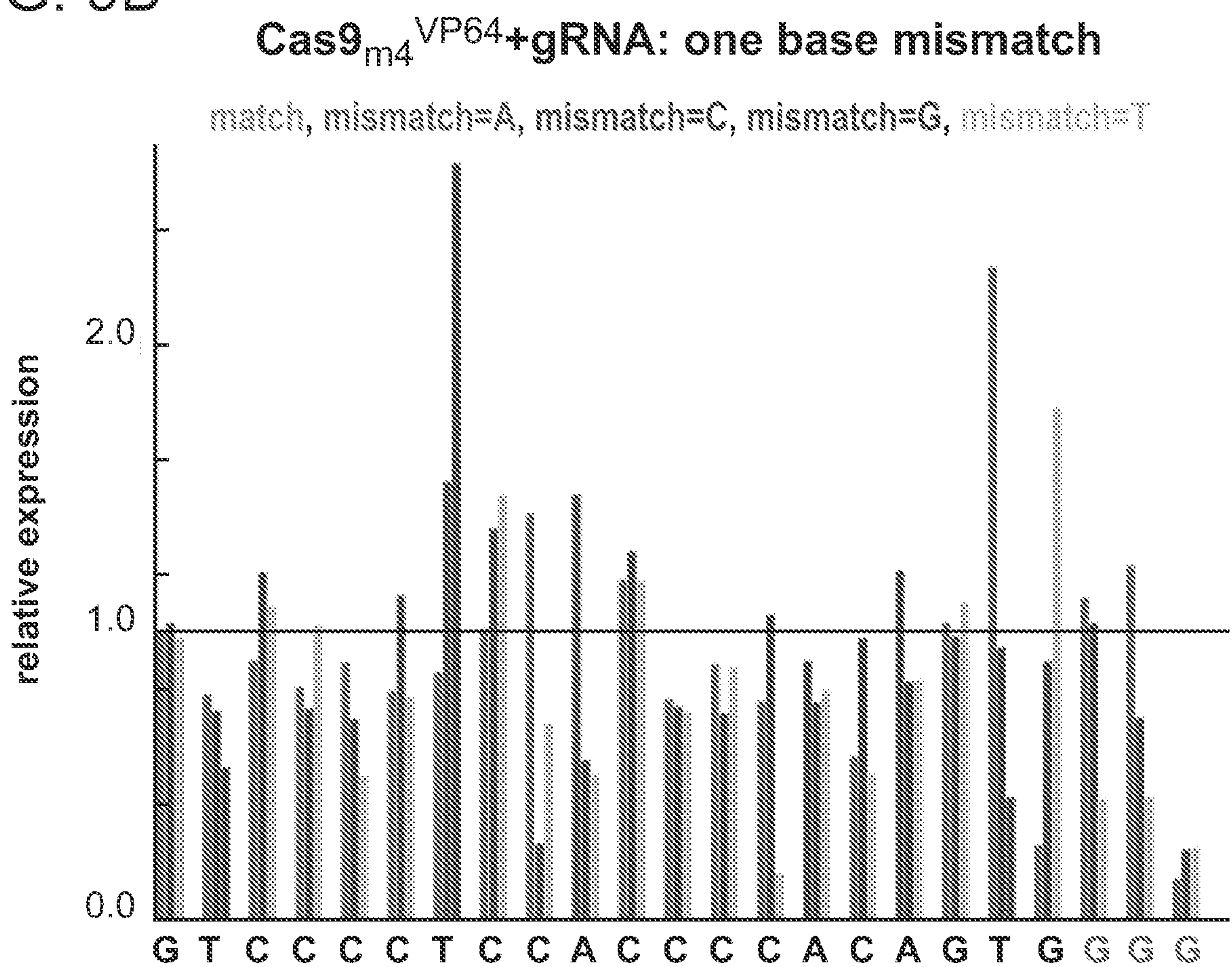


FIG. 9B



27/65

FIG. 9C

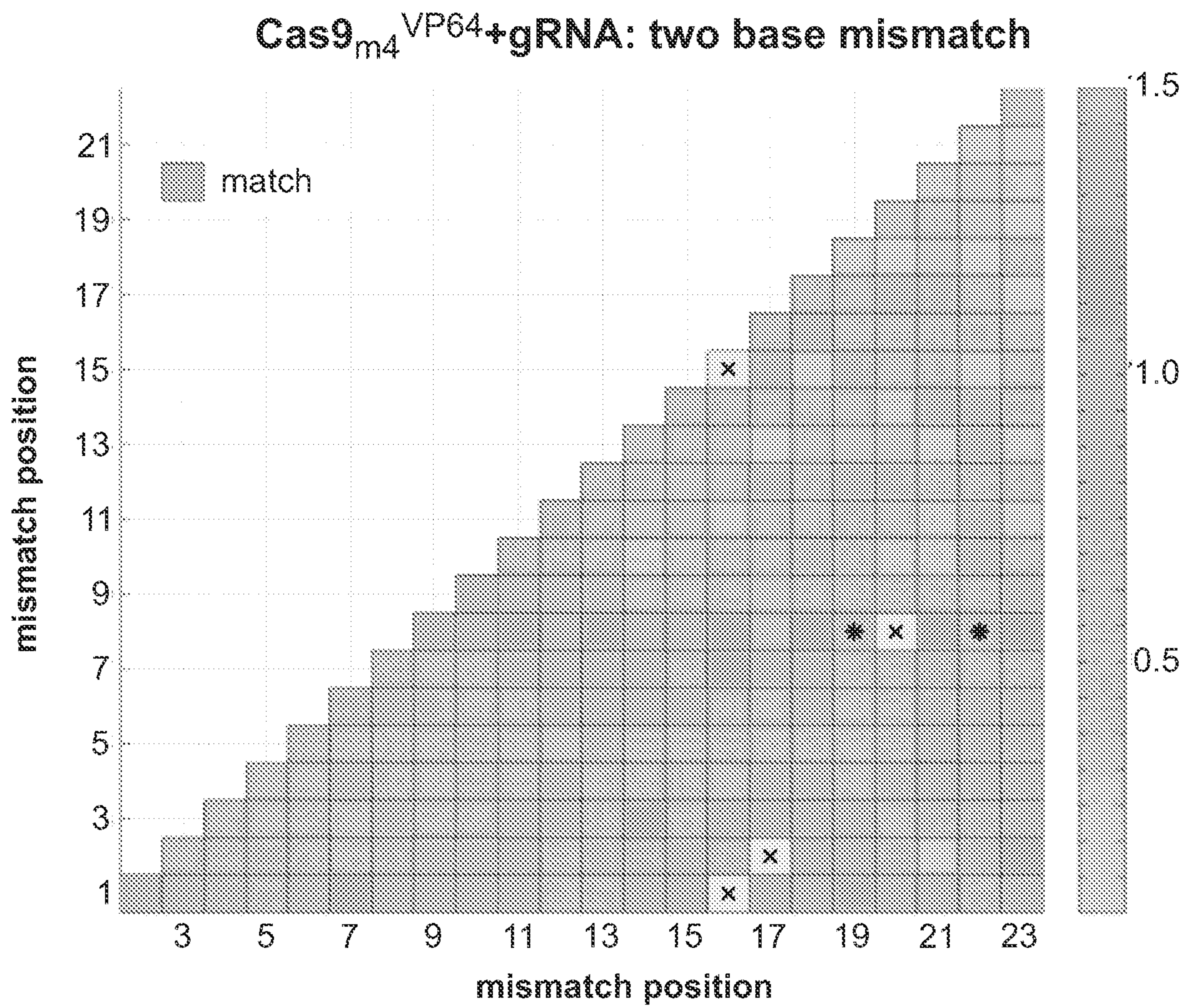


FIG. 9D

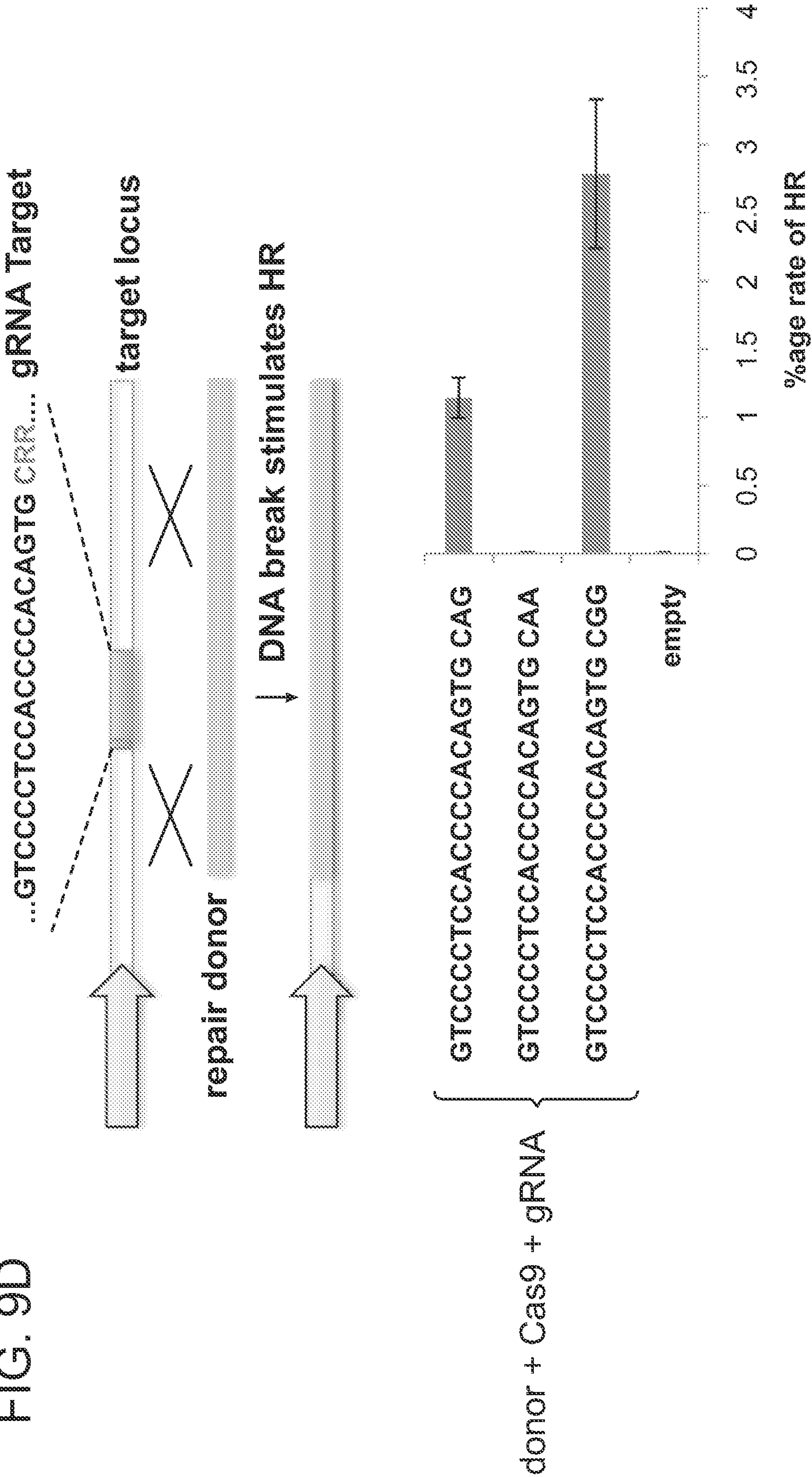


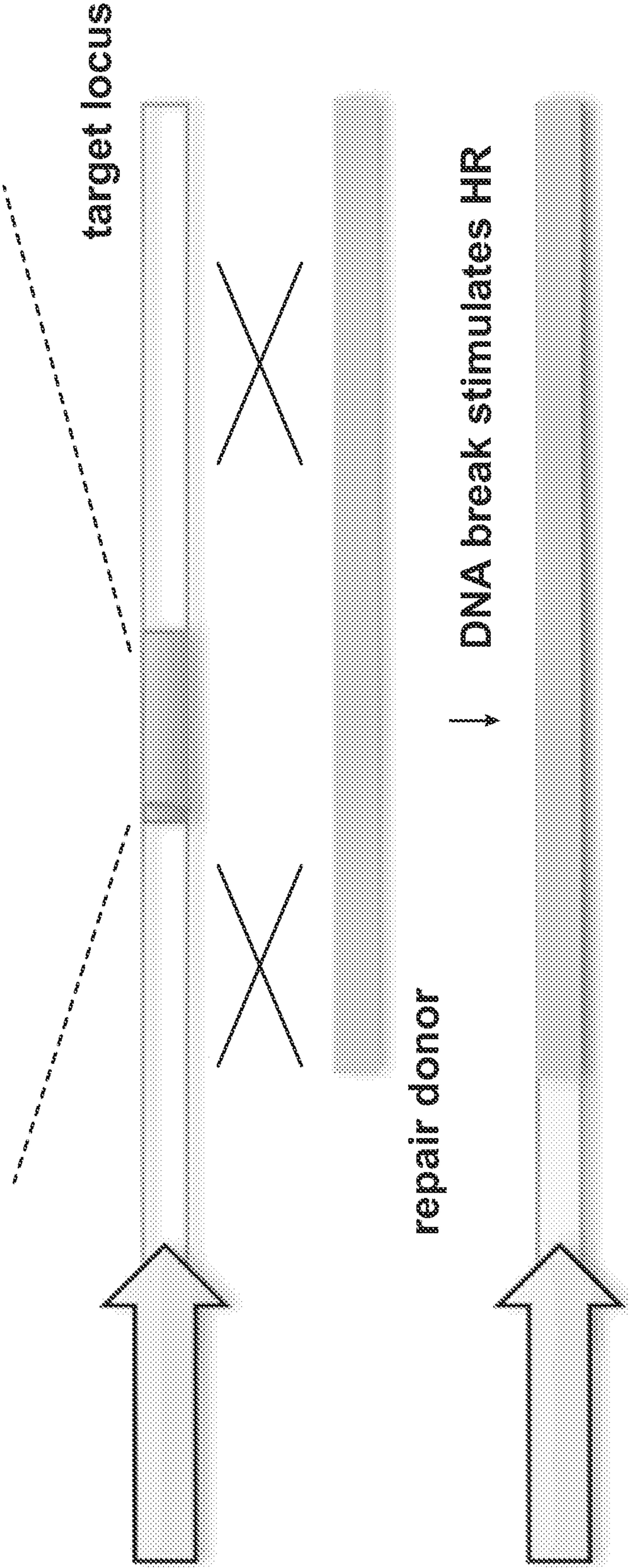
FIG. 10A-1

...TGTCCCTCCACCCACACAGTGGGGCCACTAGGGA CAGCAATTGGTCACAGAAA...

...TGTCCCTCCACCCACACAGTGGGGCCACTAGGGA CAGCAATTGGTCACAGAAA...

...TGTCCCTCCACCCACACAGTGGGGCCACTAGGGA CAGCAATTGGTCACAGAAA...

...TGTCCCTCCCTTTTTCAGTGGGGCCACTAGGGA CAGCAATTGGTCACAGAAA...



30/65

FIG. 10A-2

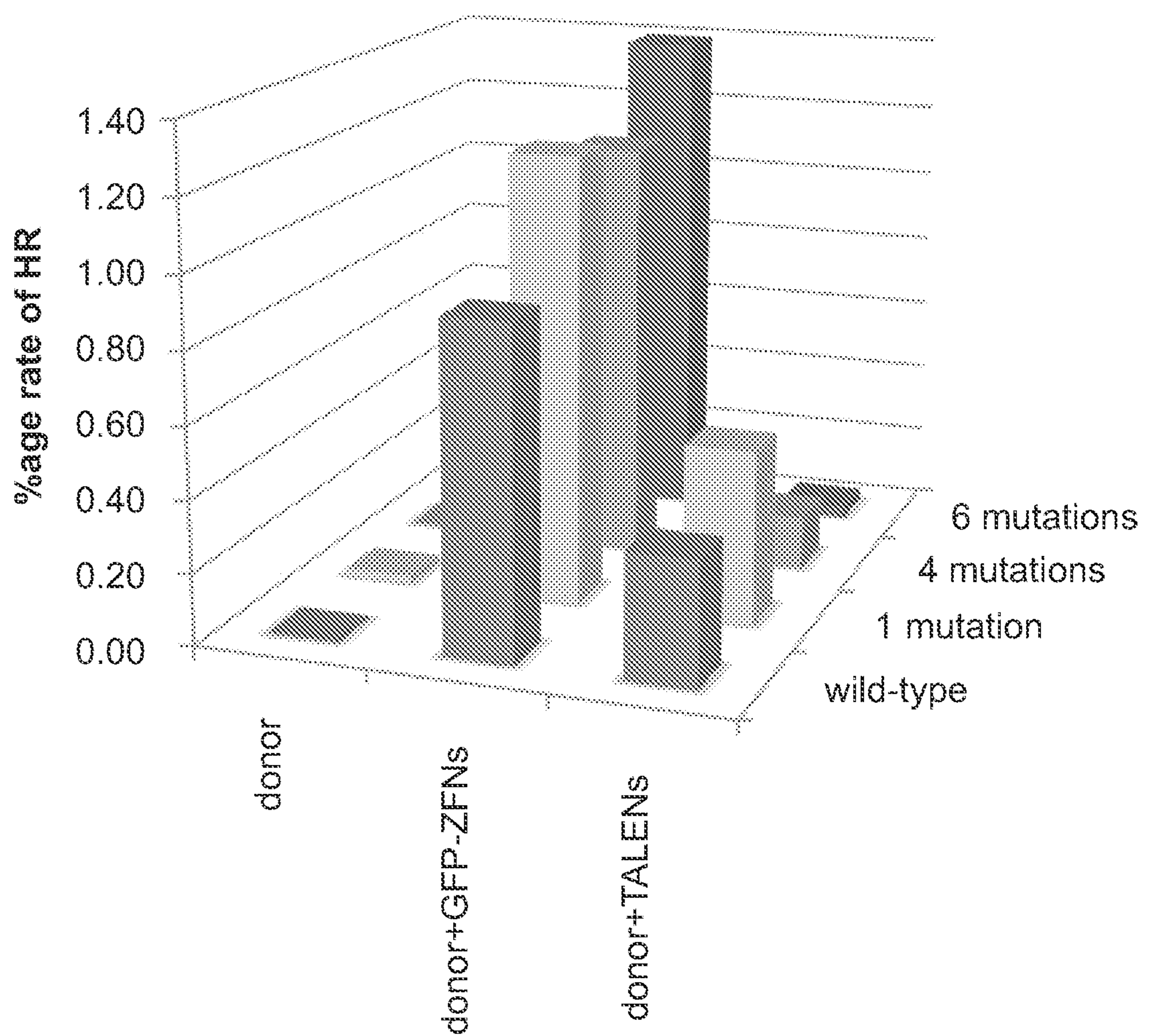
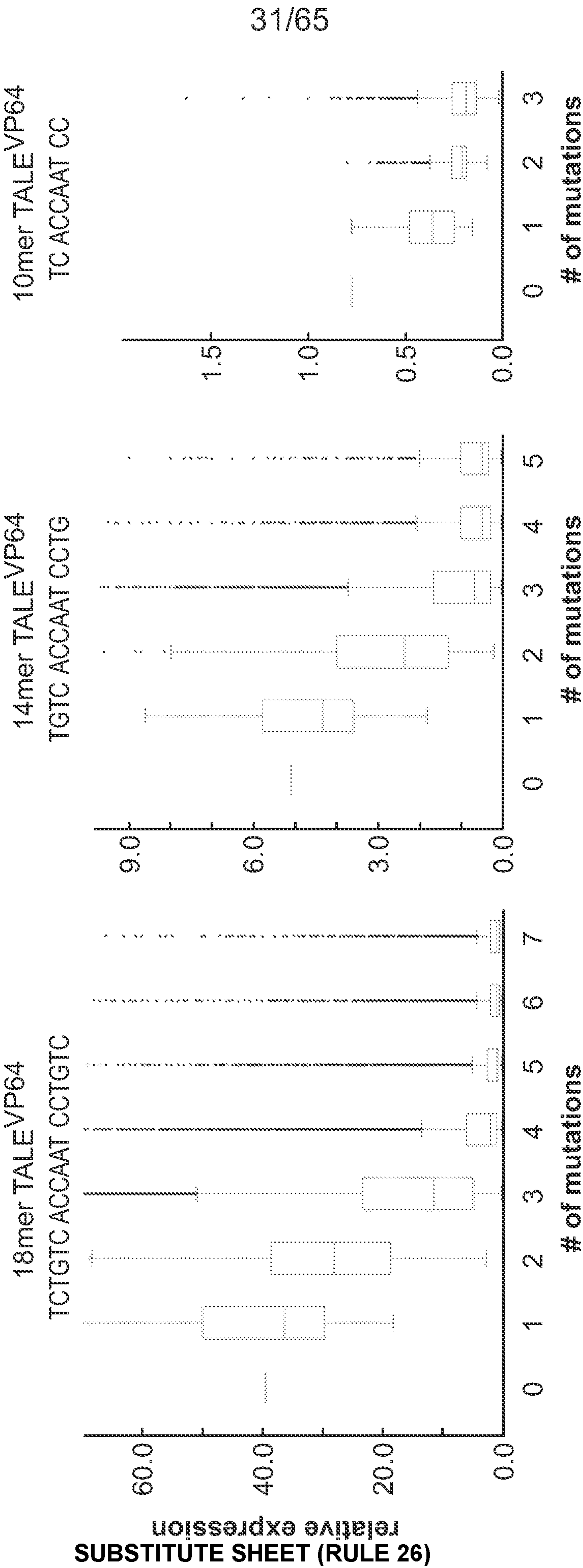


FIG. 10B



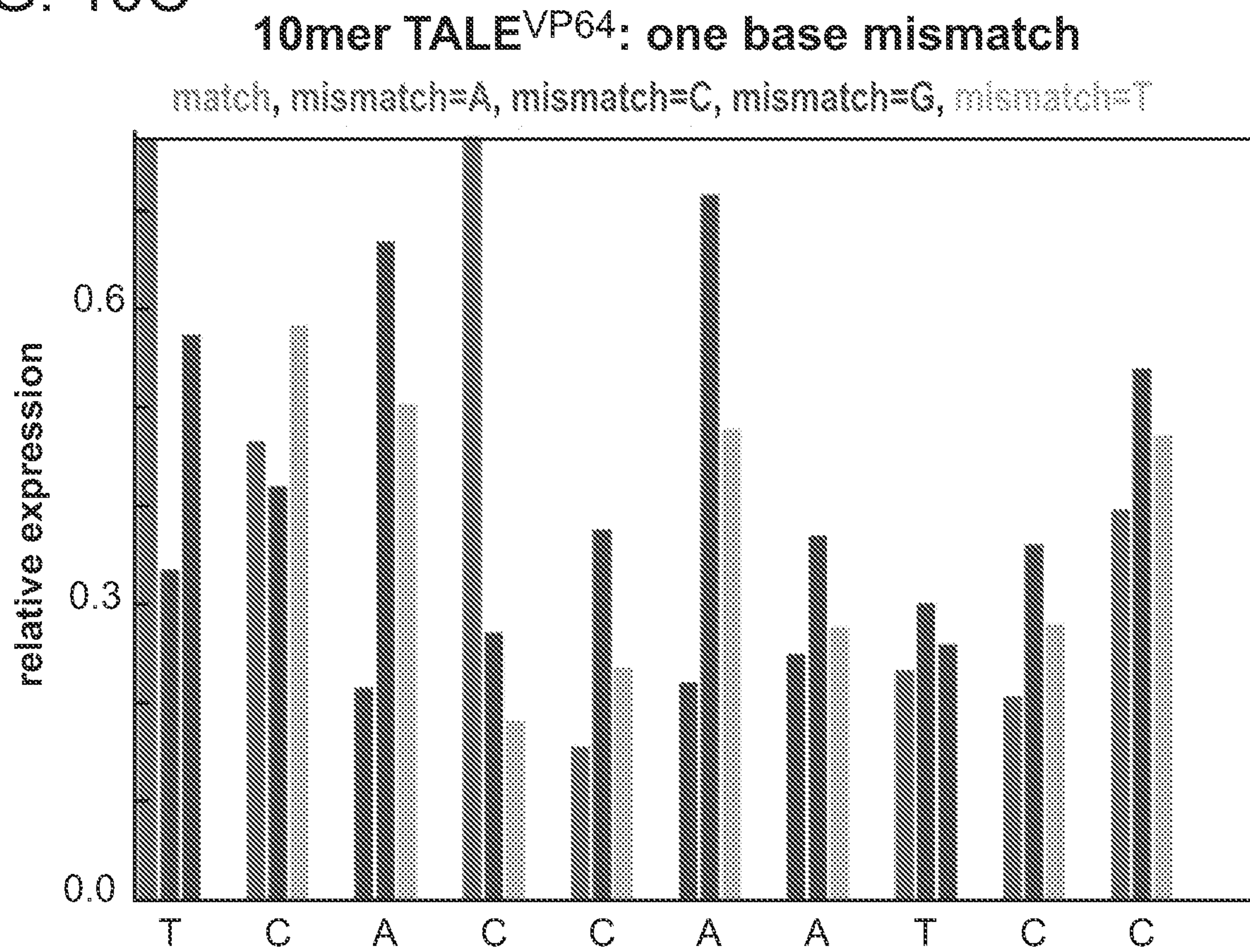


FIG. 10D

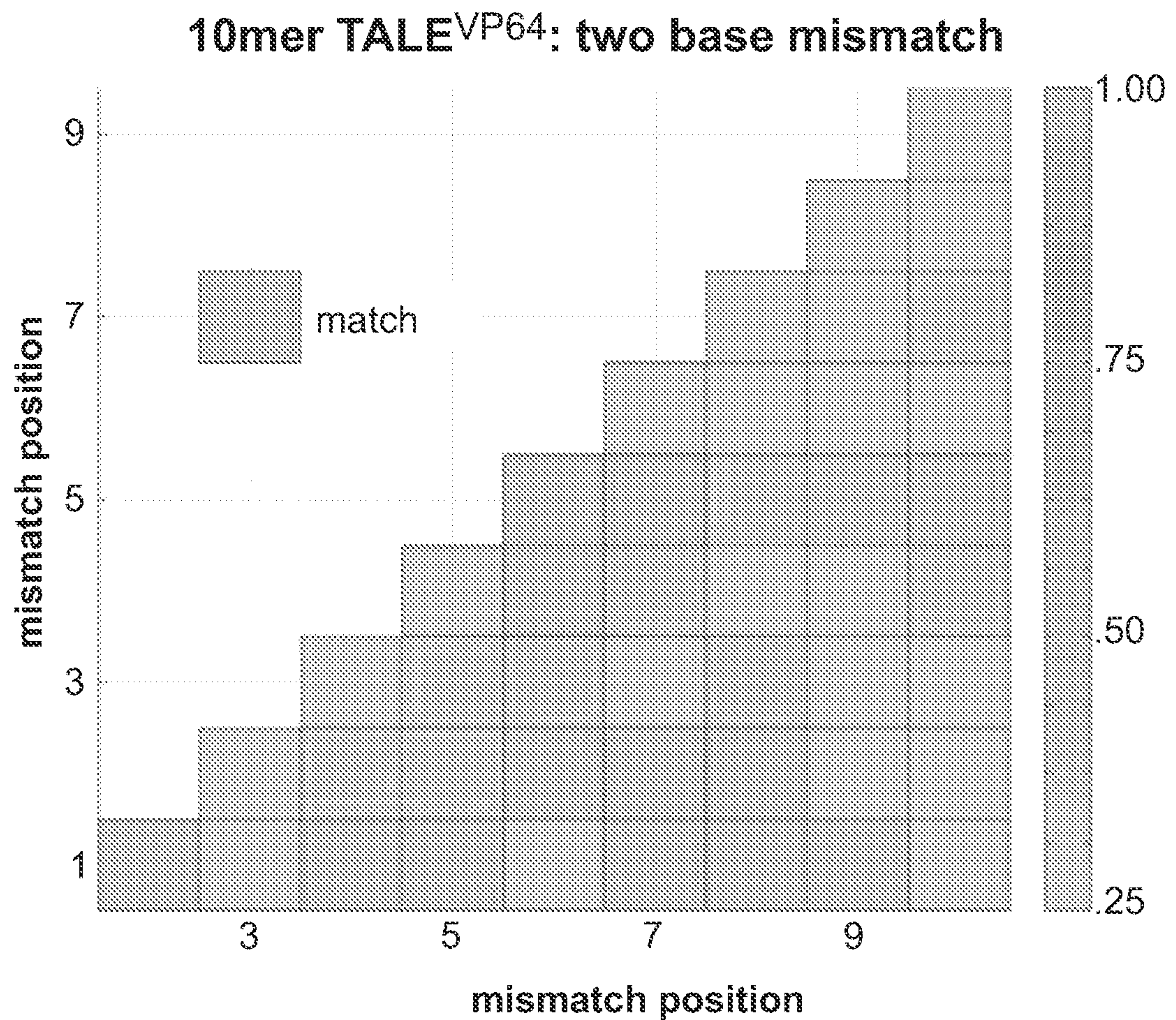
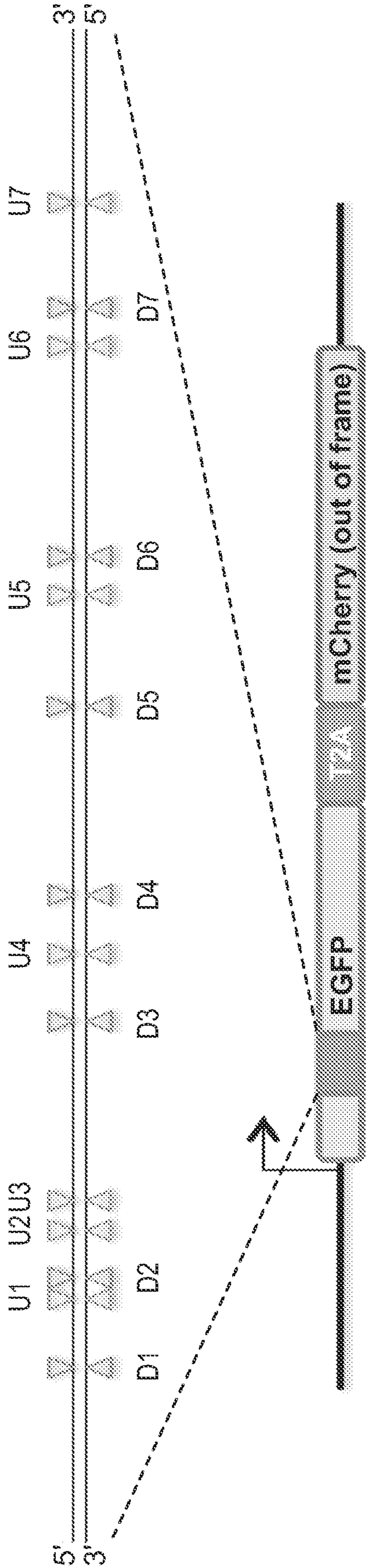
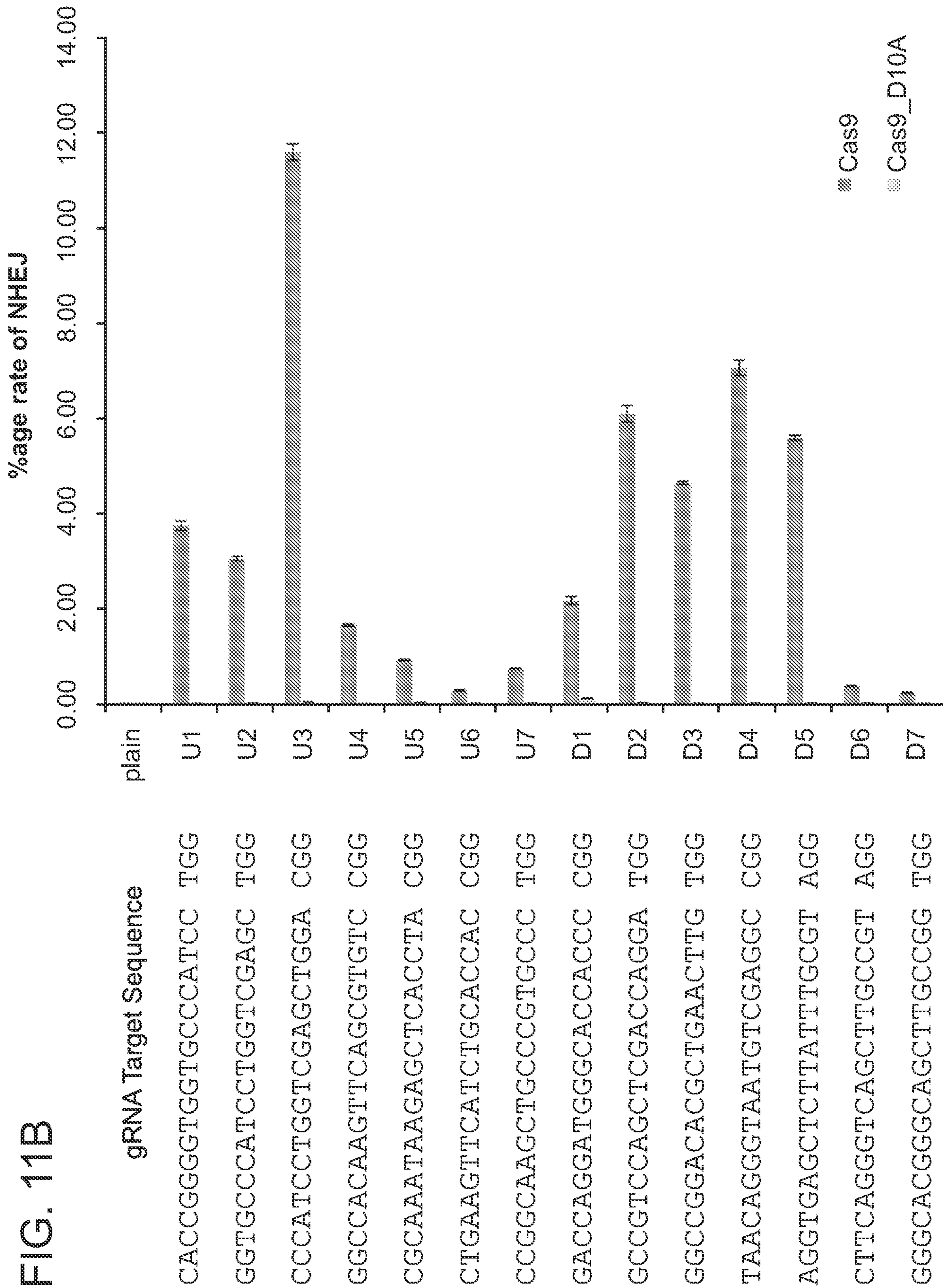


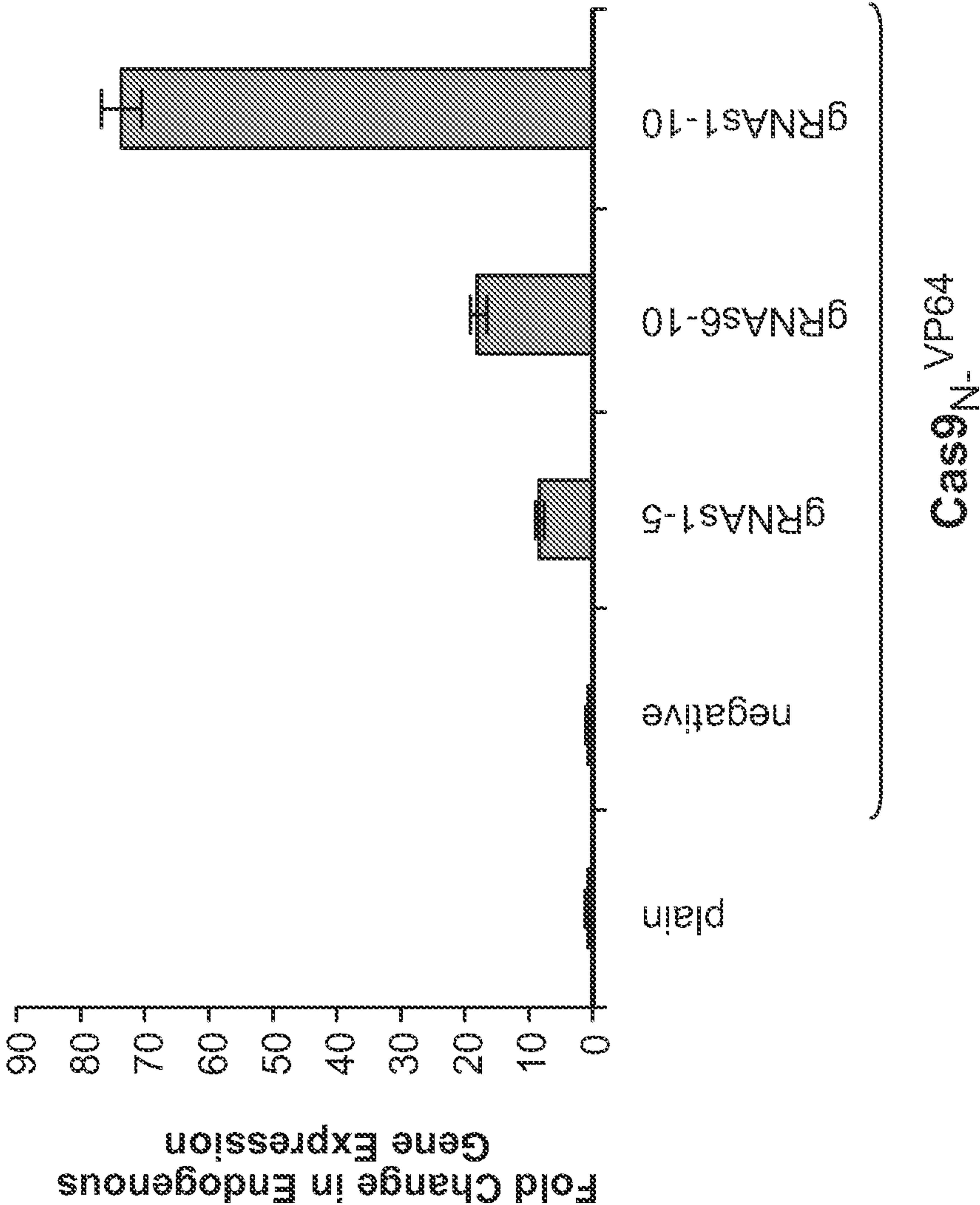
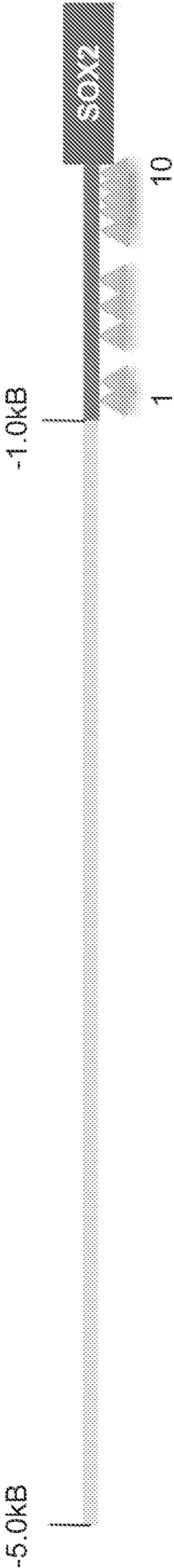
FIG. 11A





35/65

FIG. 12A



36/65

FIG. 12B

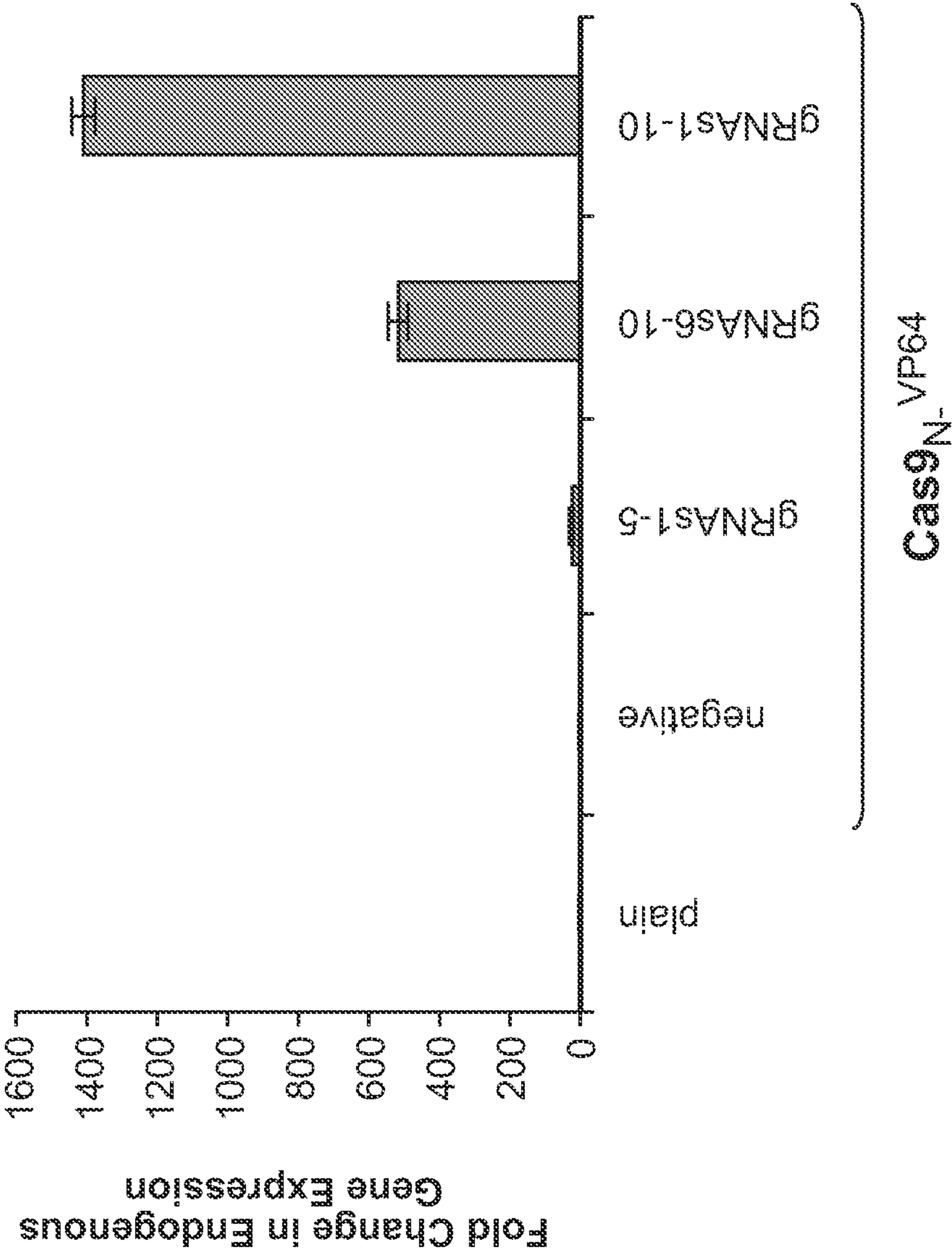
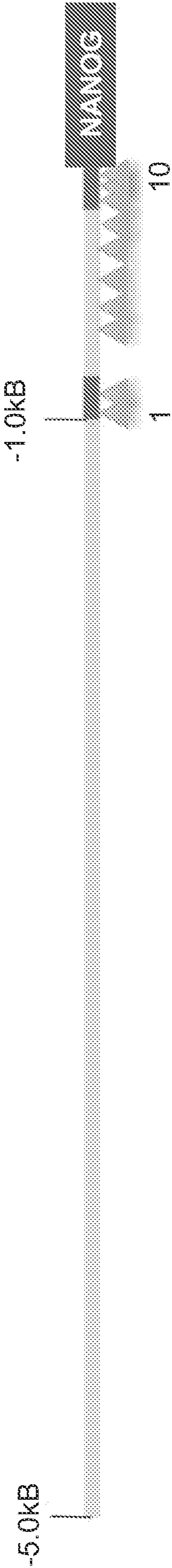


FIG. 13A

37/65

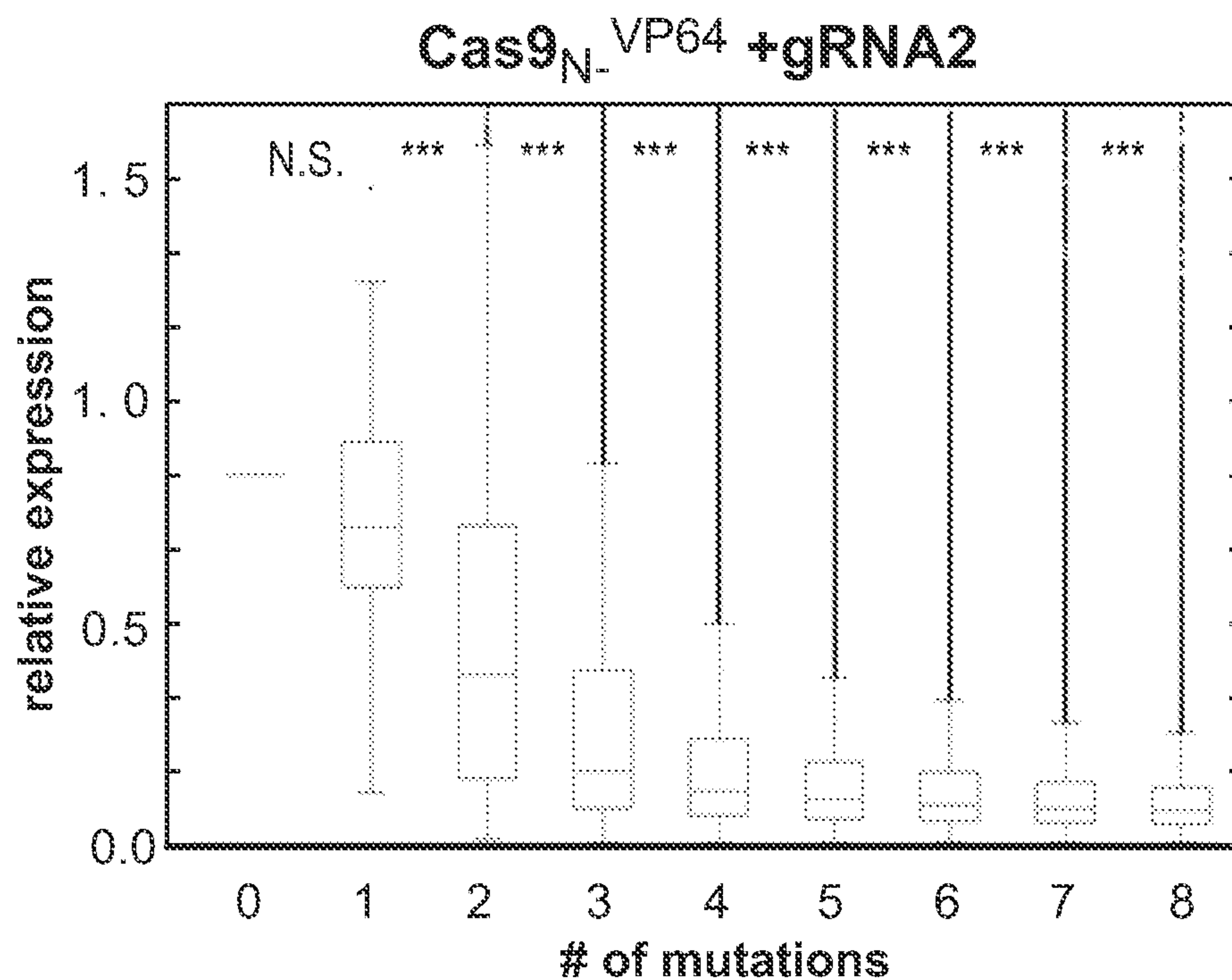
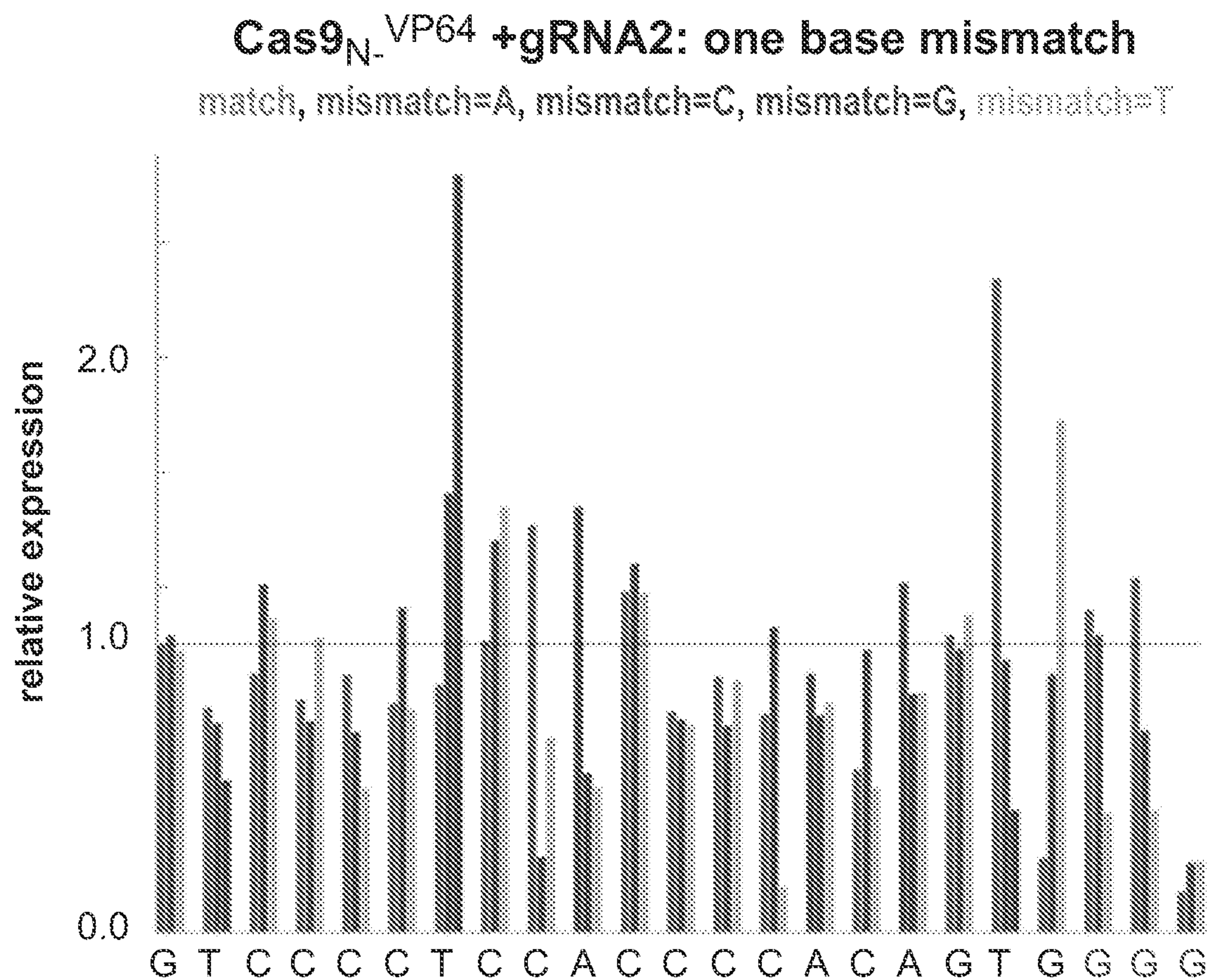


FIG. 13B



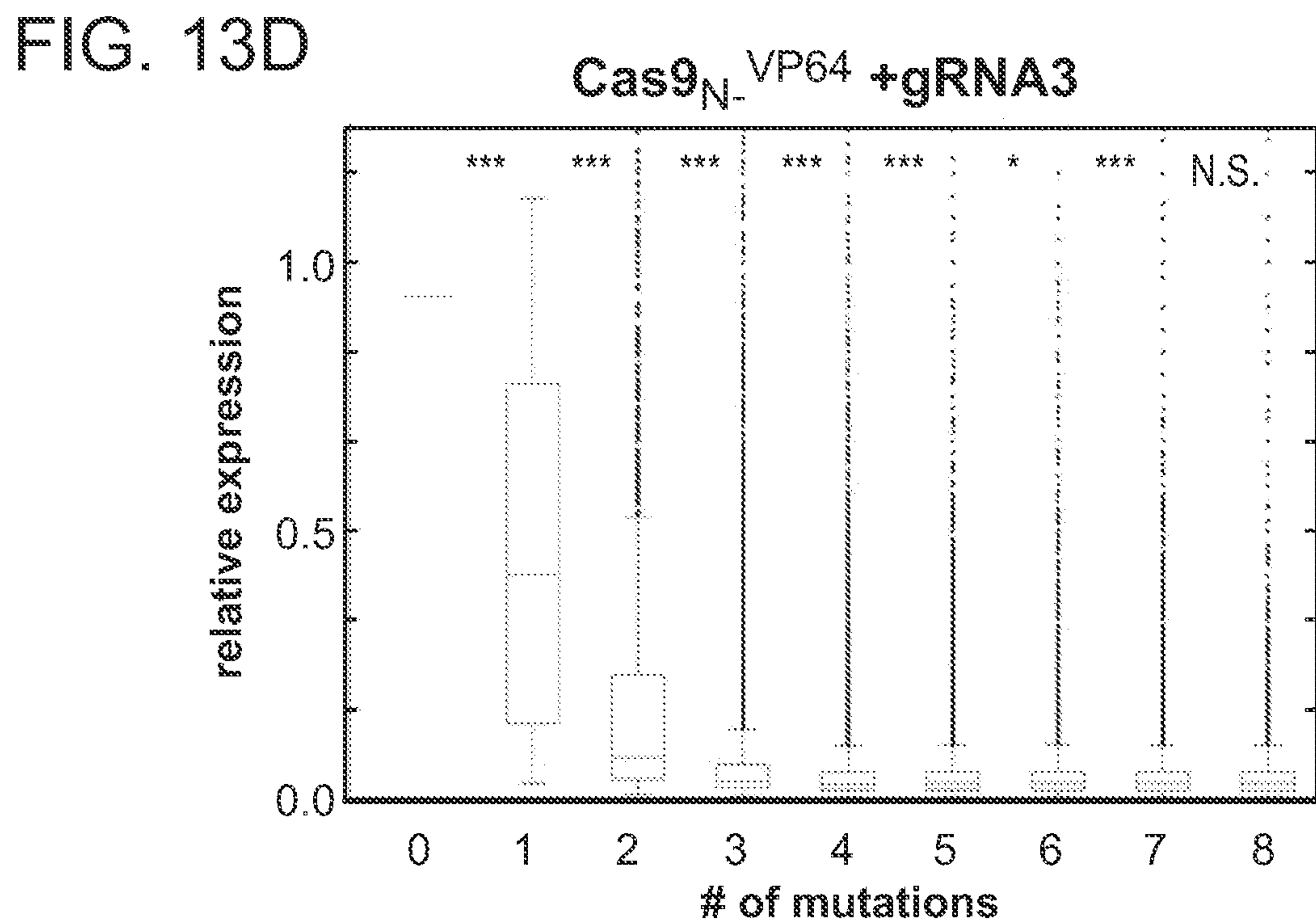
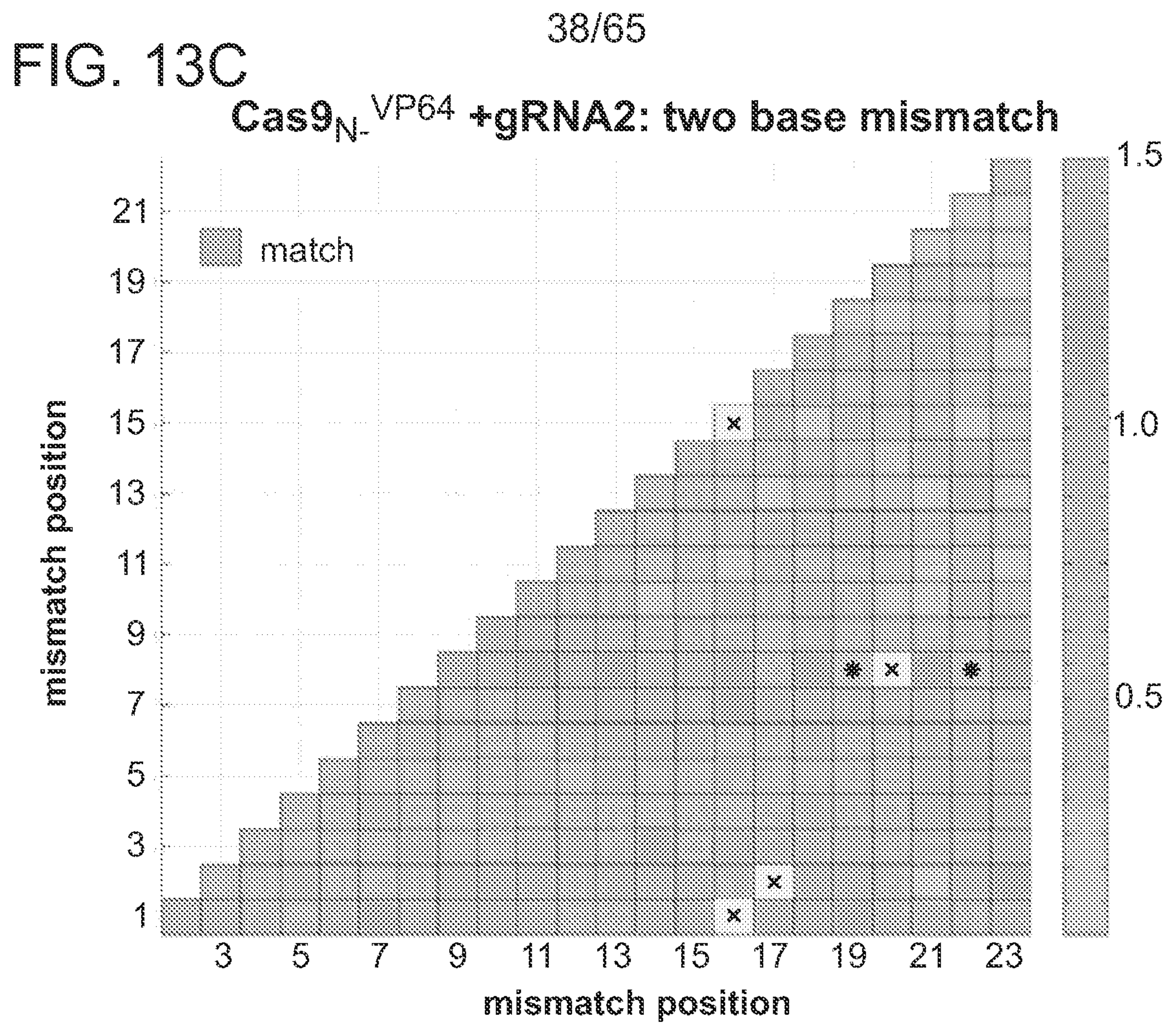


FIG. 13E

39/65

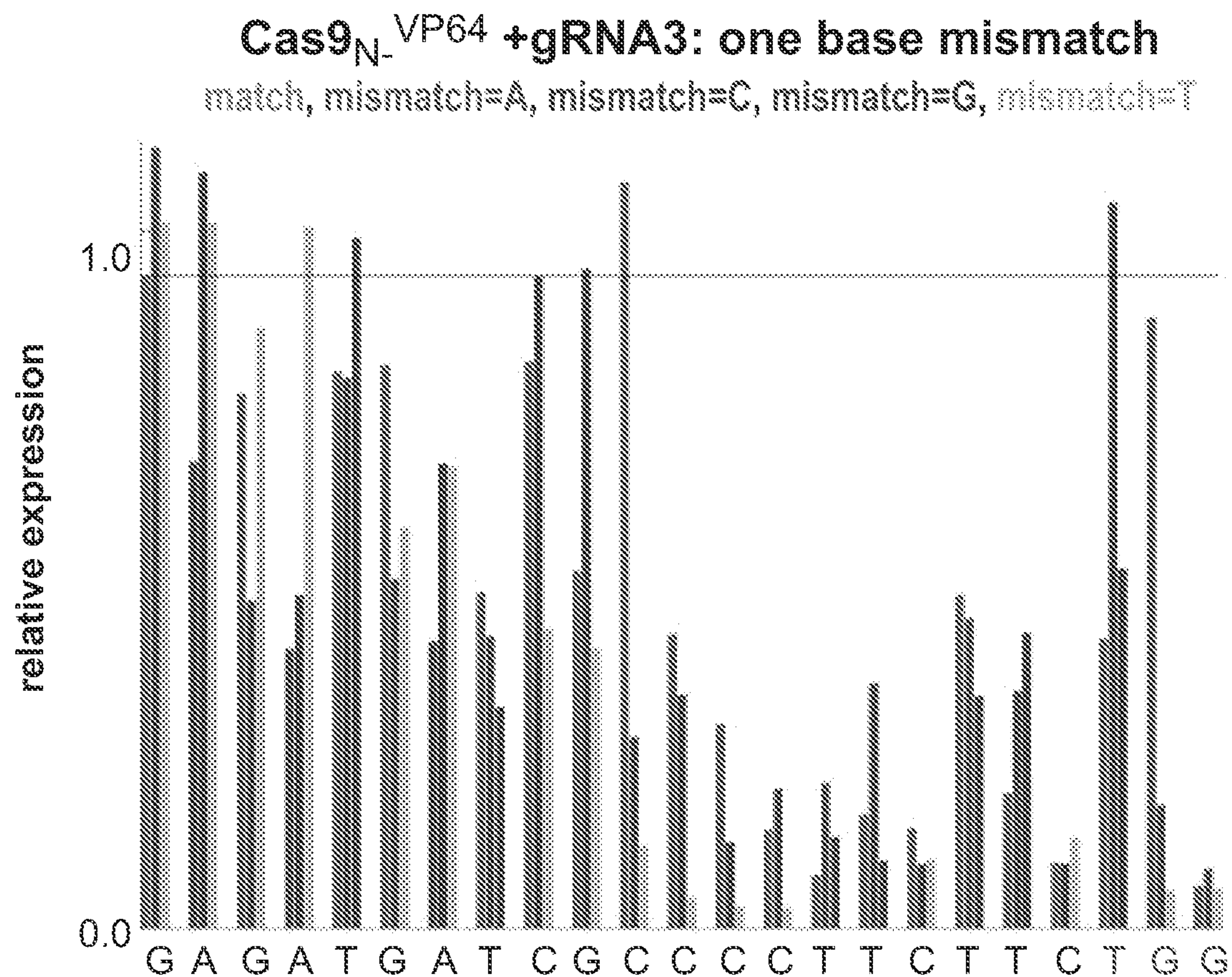


FIG. 13F

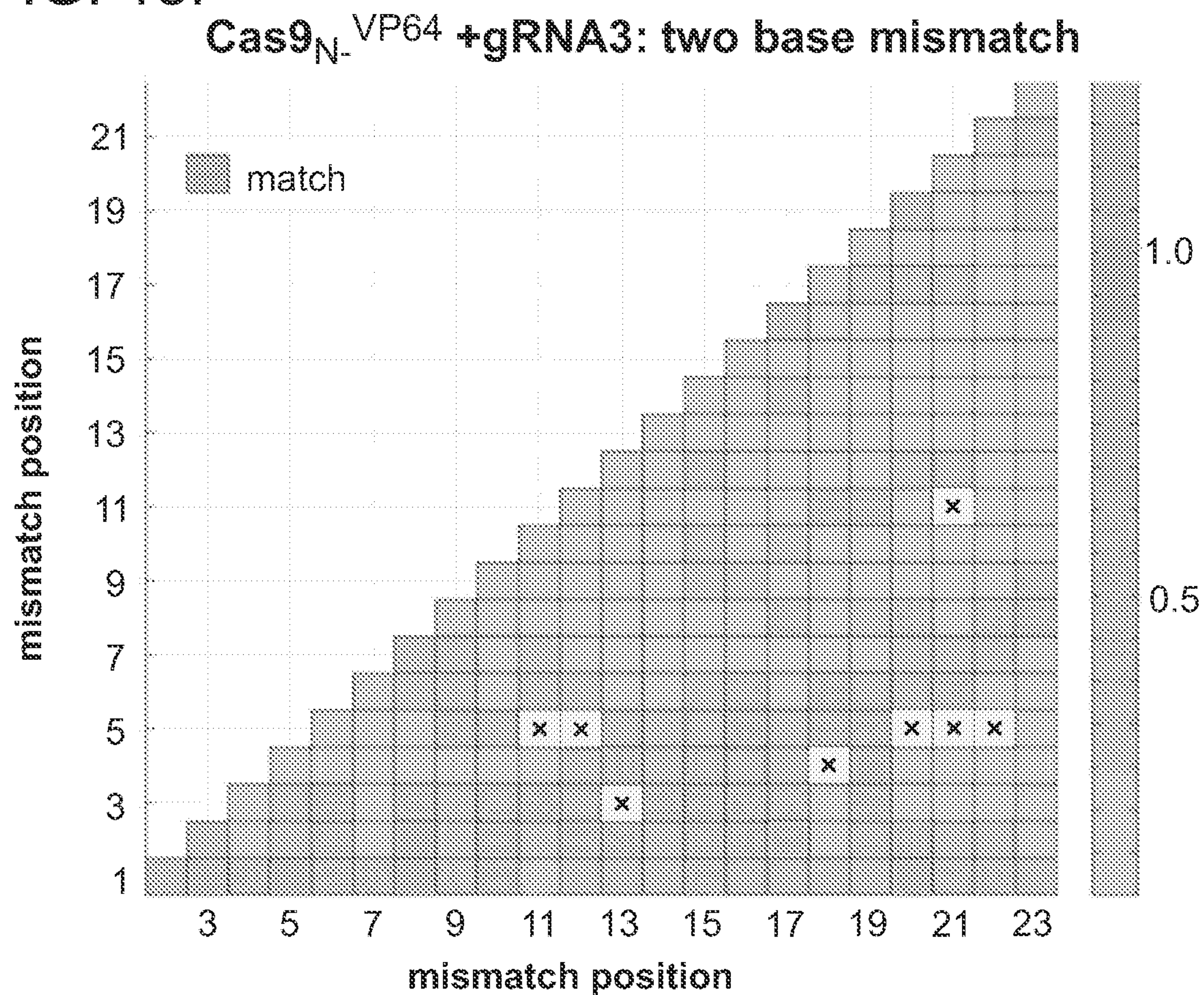


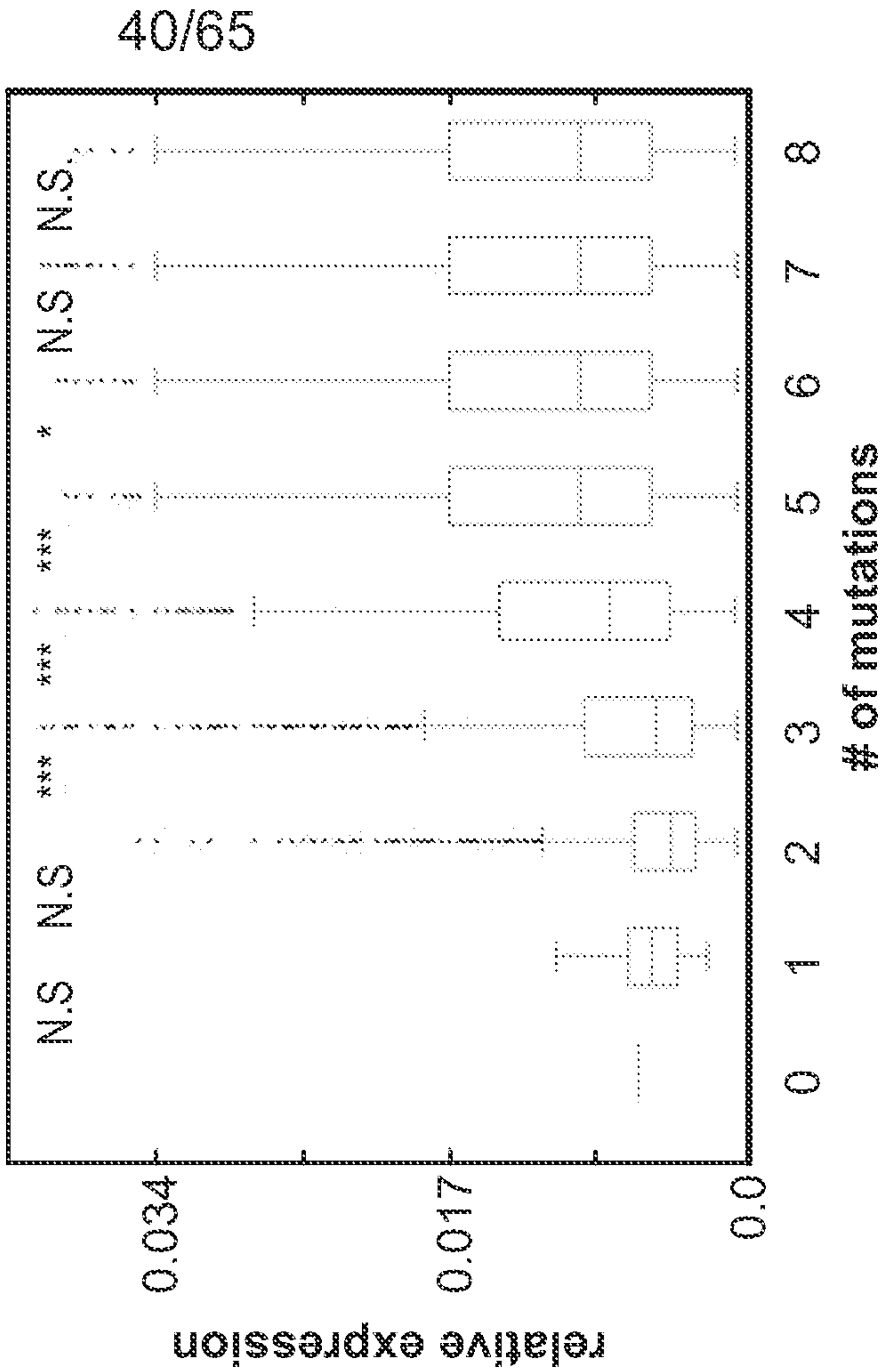
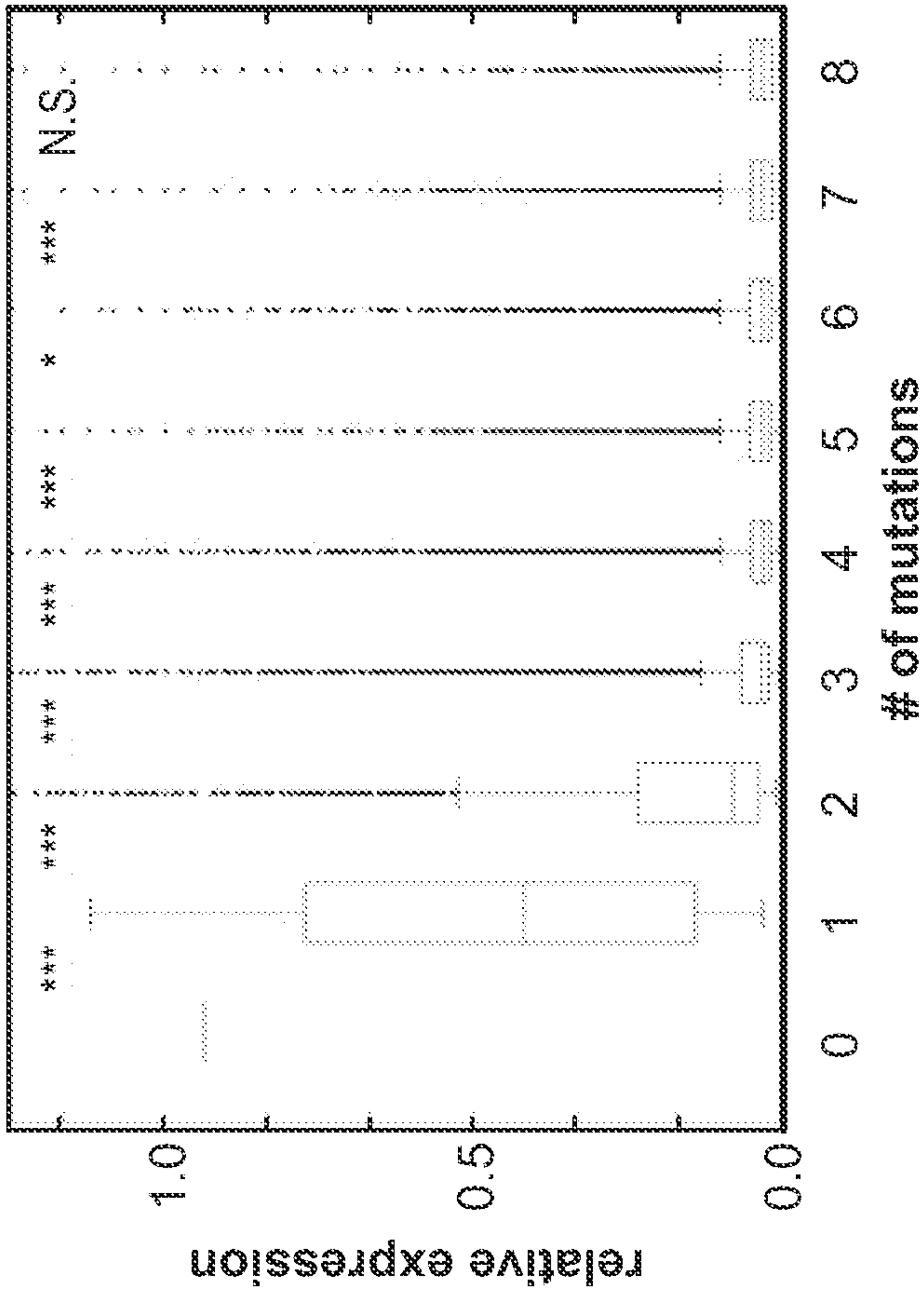
FIG. 14A

Target : GAGATGATCGCCCCCTTC TGG
gRNA3 : GAGATGATCGCCCCCTTC
gRNA3mut: GTGATGACCGCCCTTC

FIG. 14B

FIG. 14C

Cas9_{N-VP64} +gRNA3
Cas9_{N-VP64} +gRNA3mut



41/65

FIG. 15A

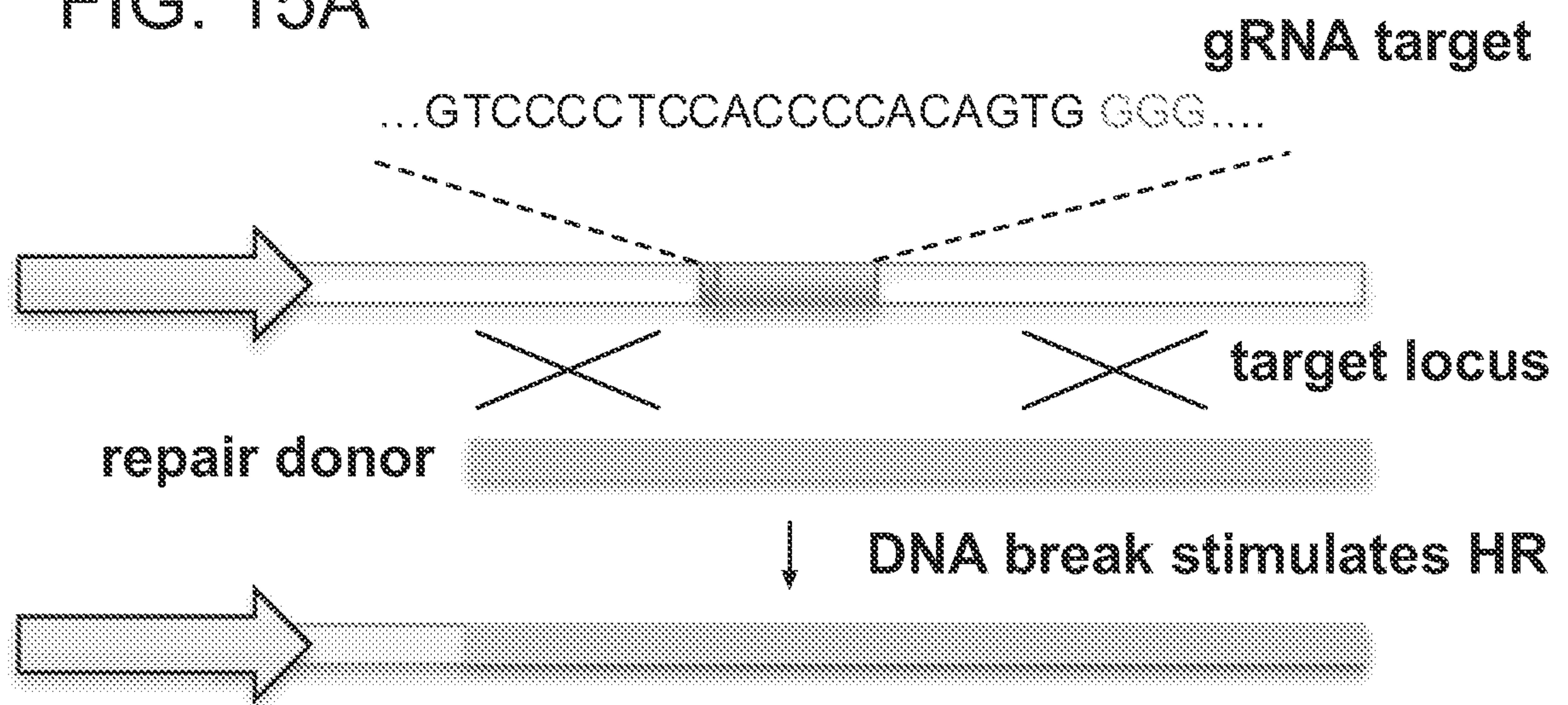
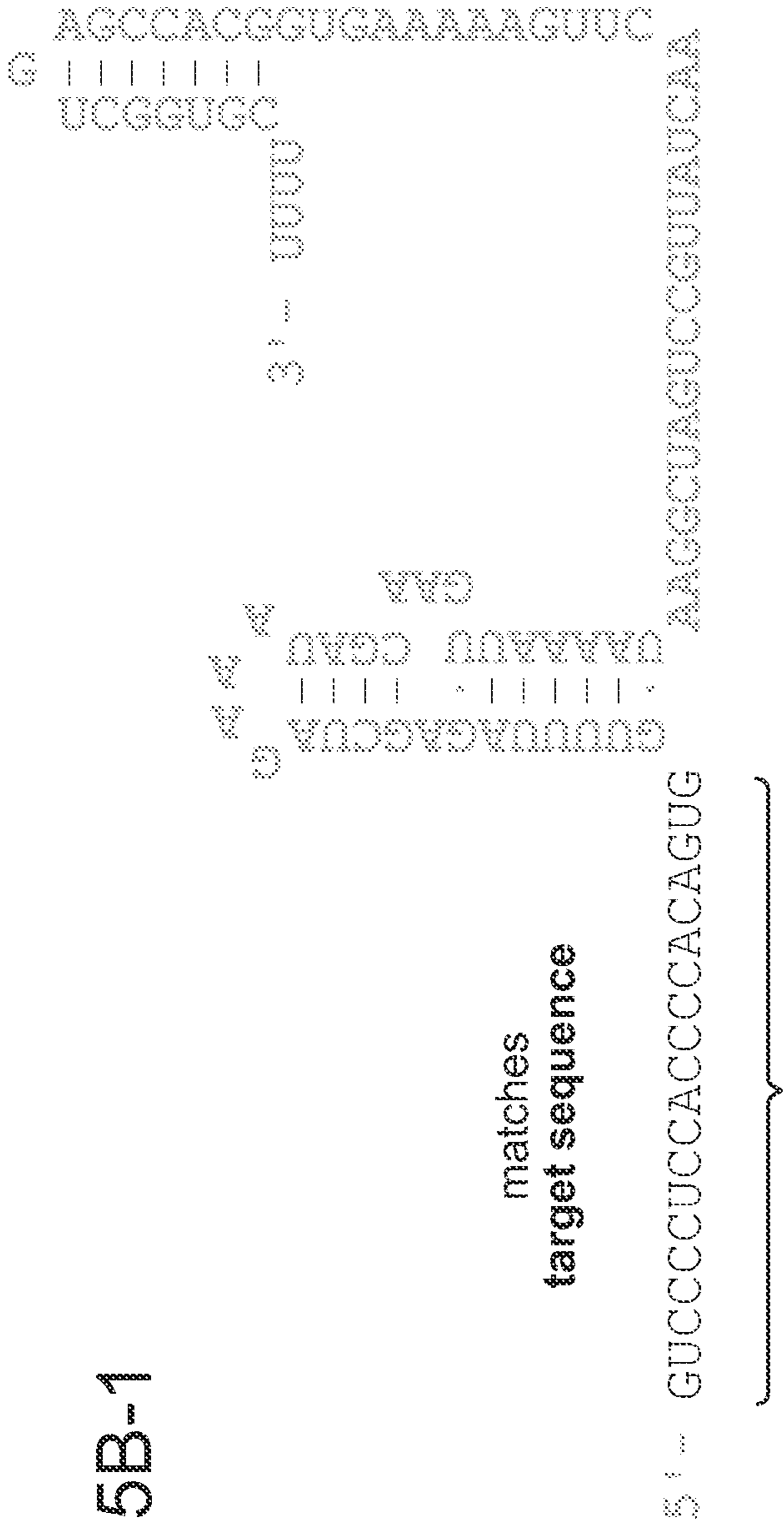
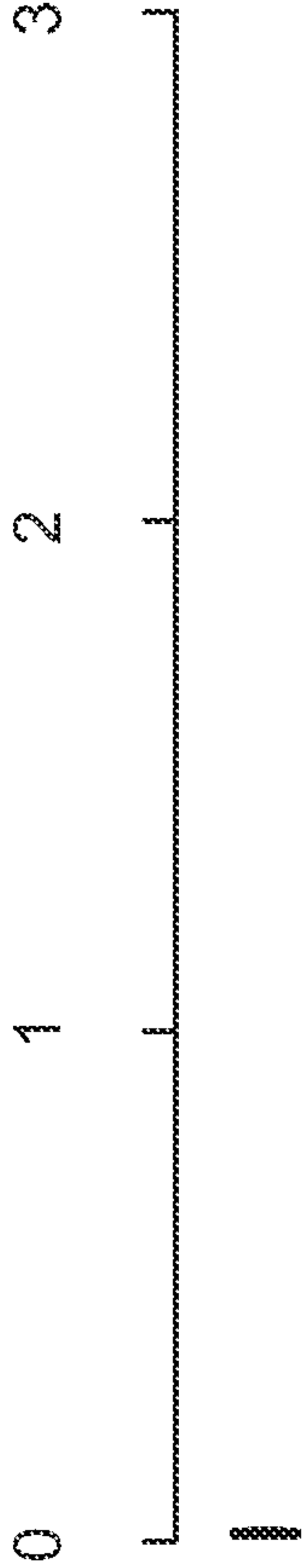


FIG. 15B-1

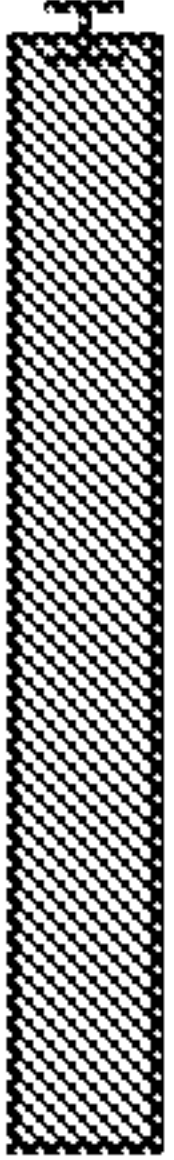


42/65

relative rate of targeting

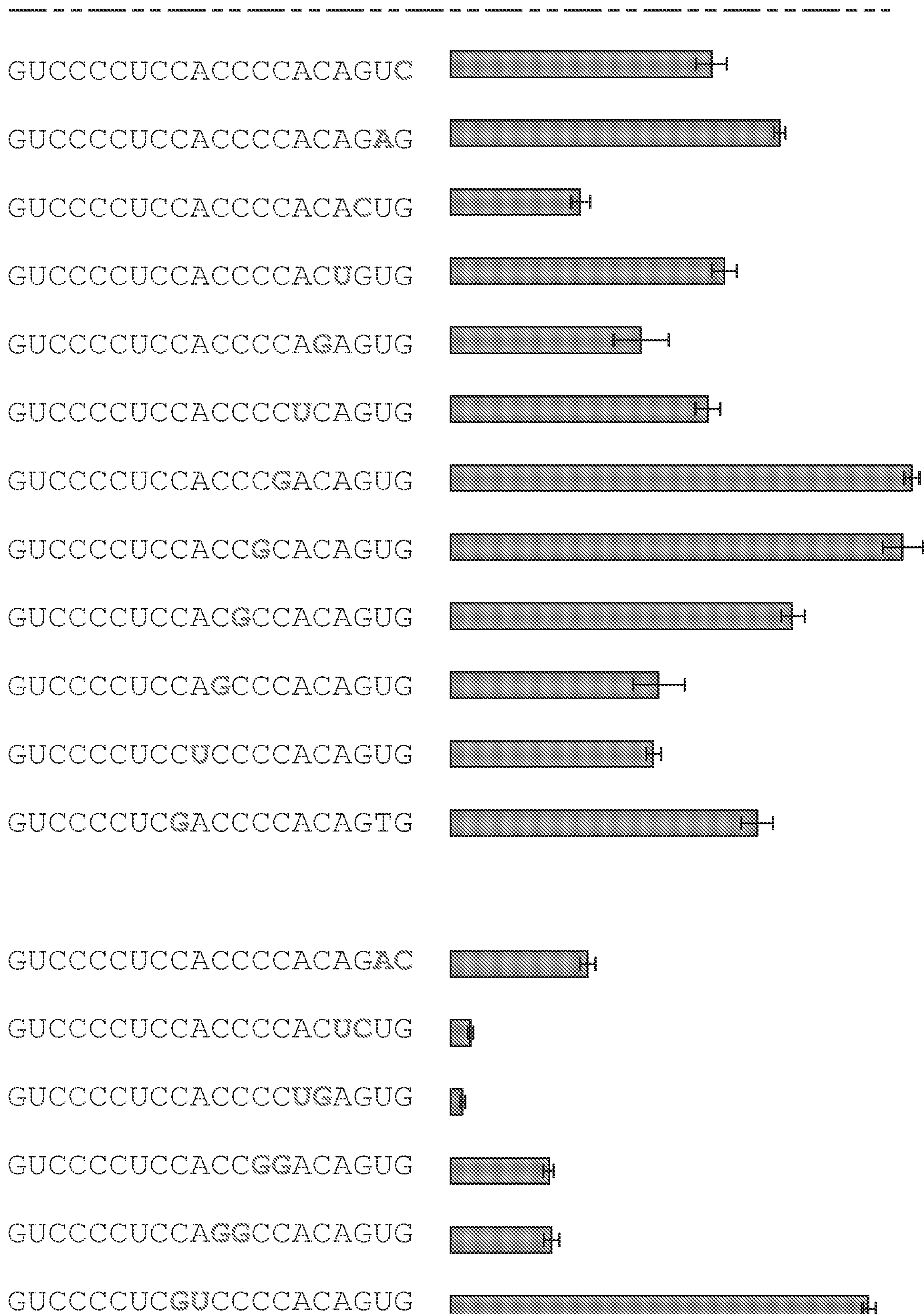


GUCCCCUCCACCCACAGUG



43/65

FIG. 15B-2



SUBSTITUTE SHEET (RULE 26)

44/65

FIG. 15C

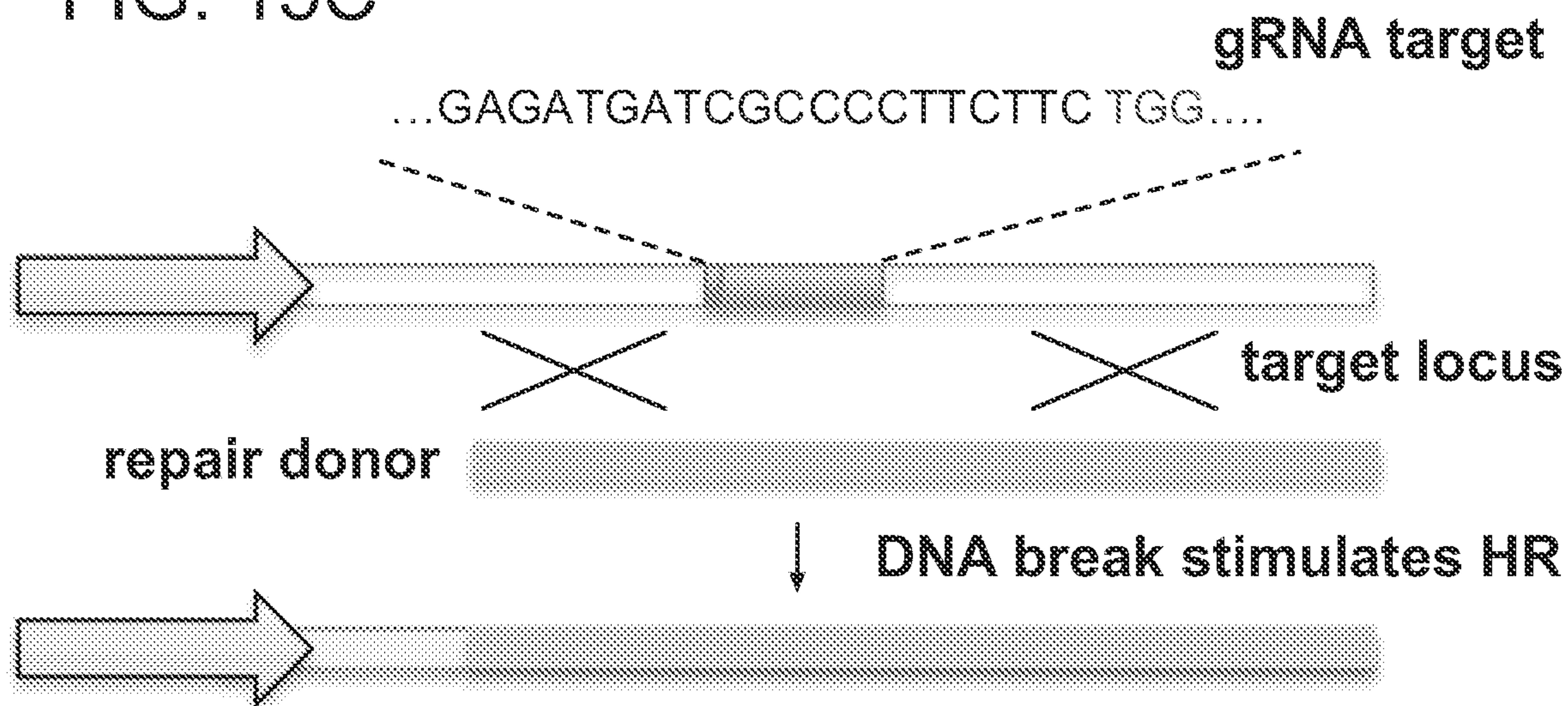
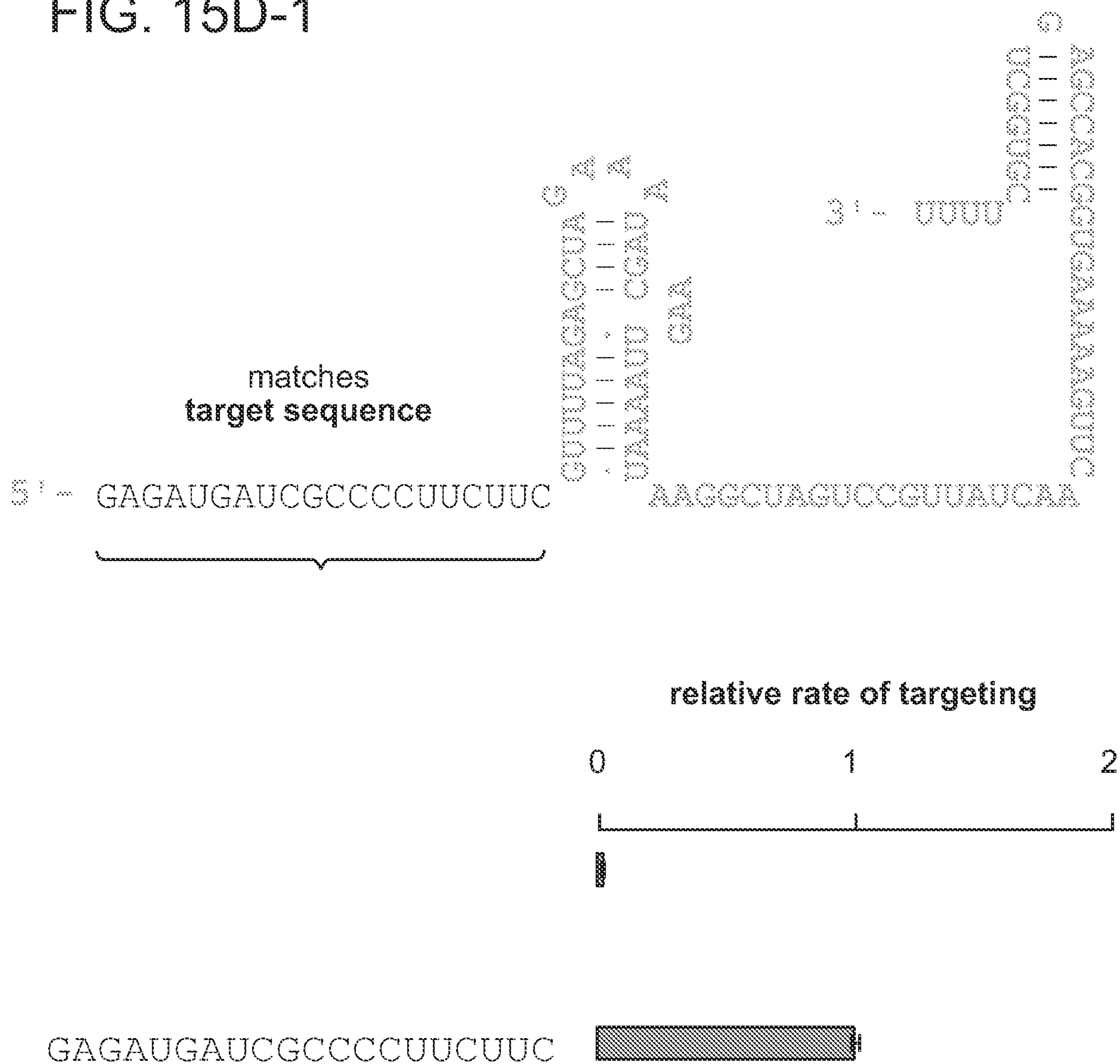
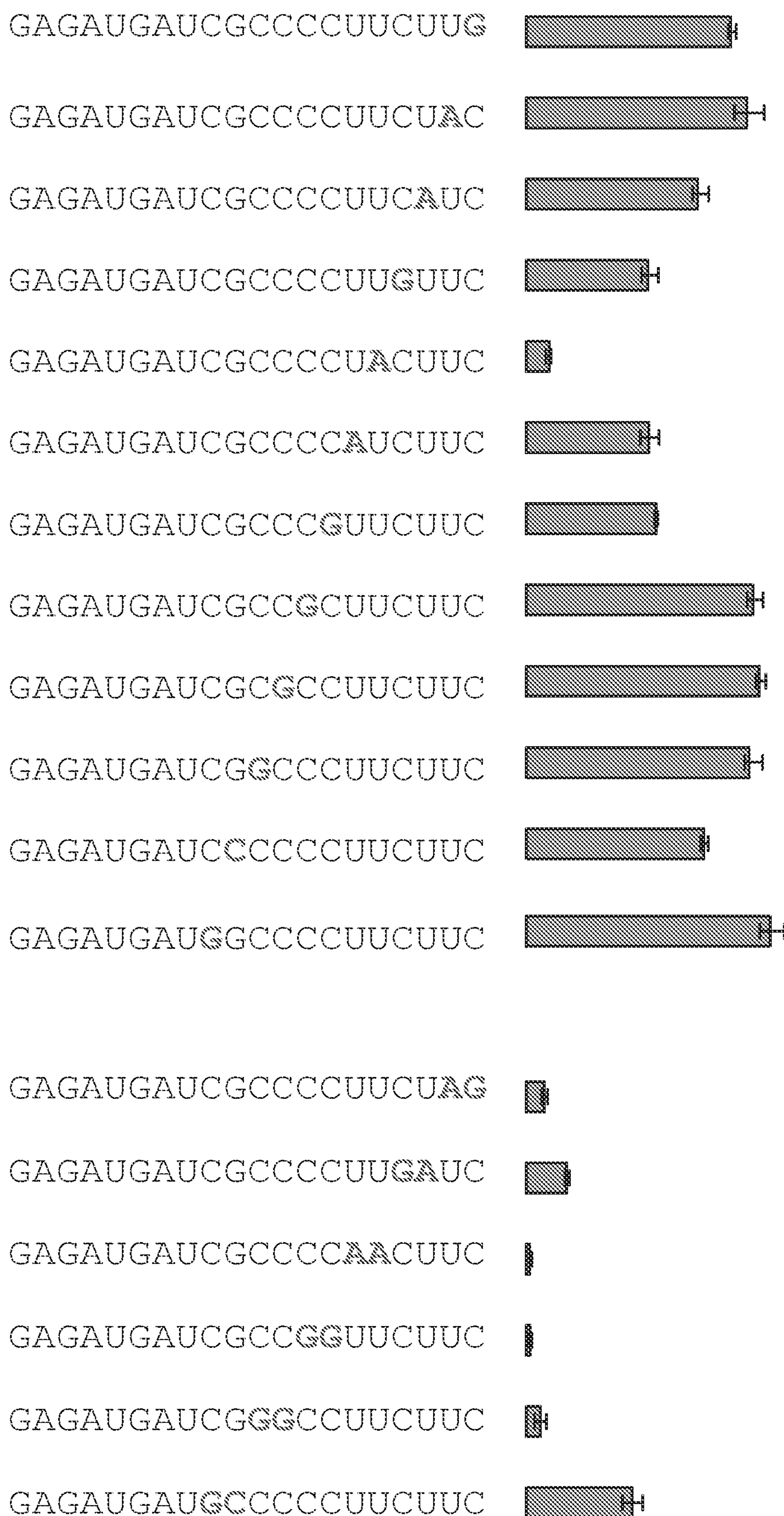


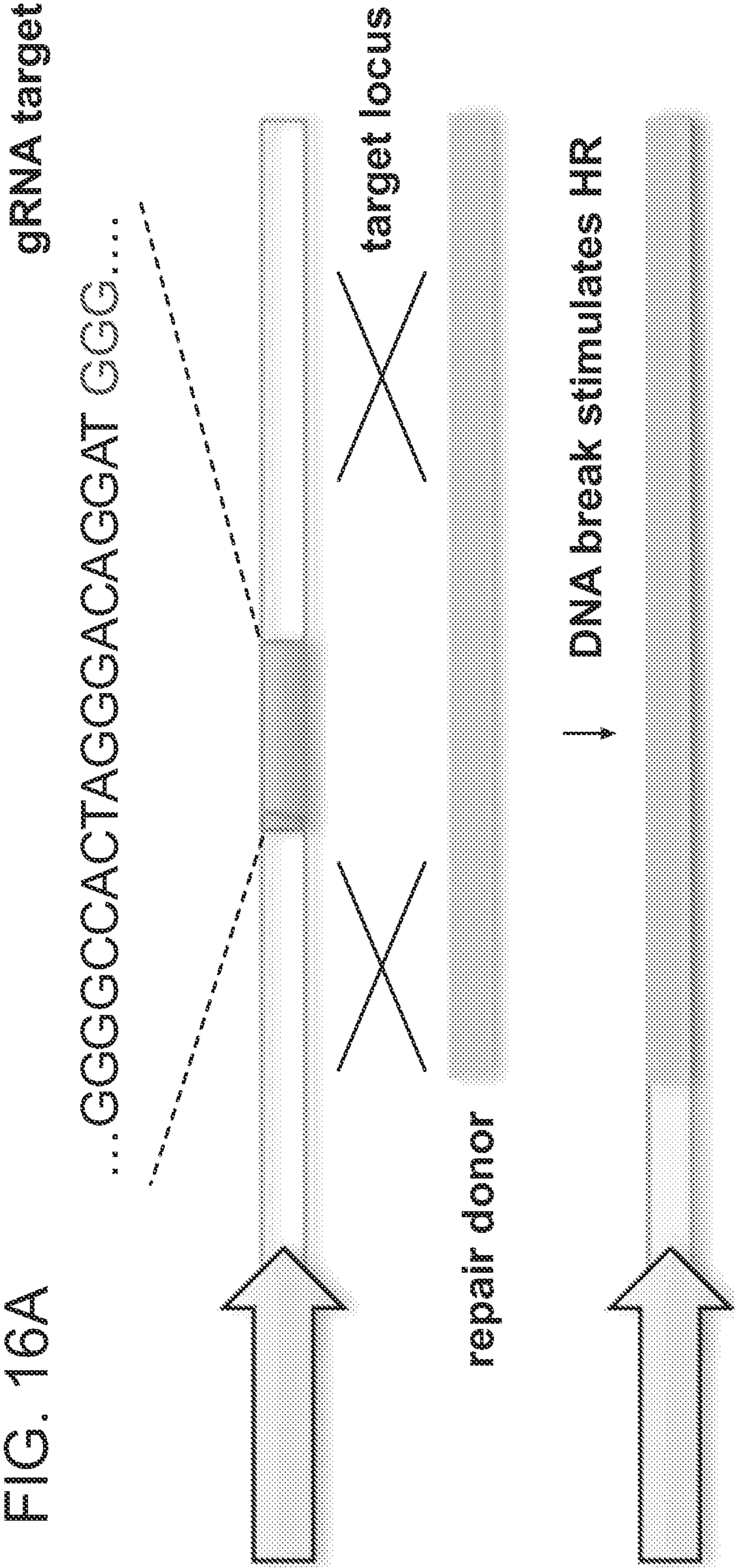
FIG. 15D-1



45/65

FIG. 15D-2





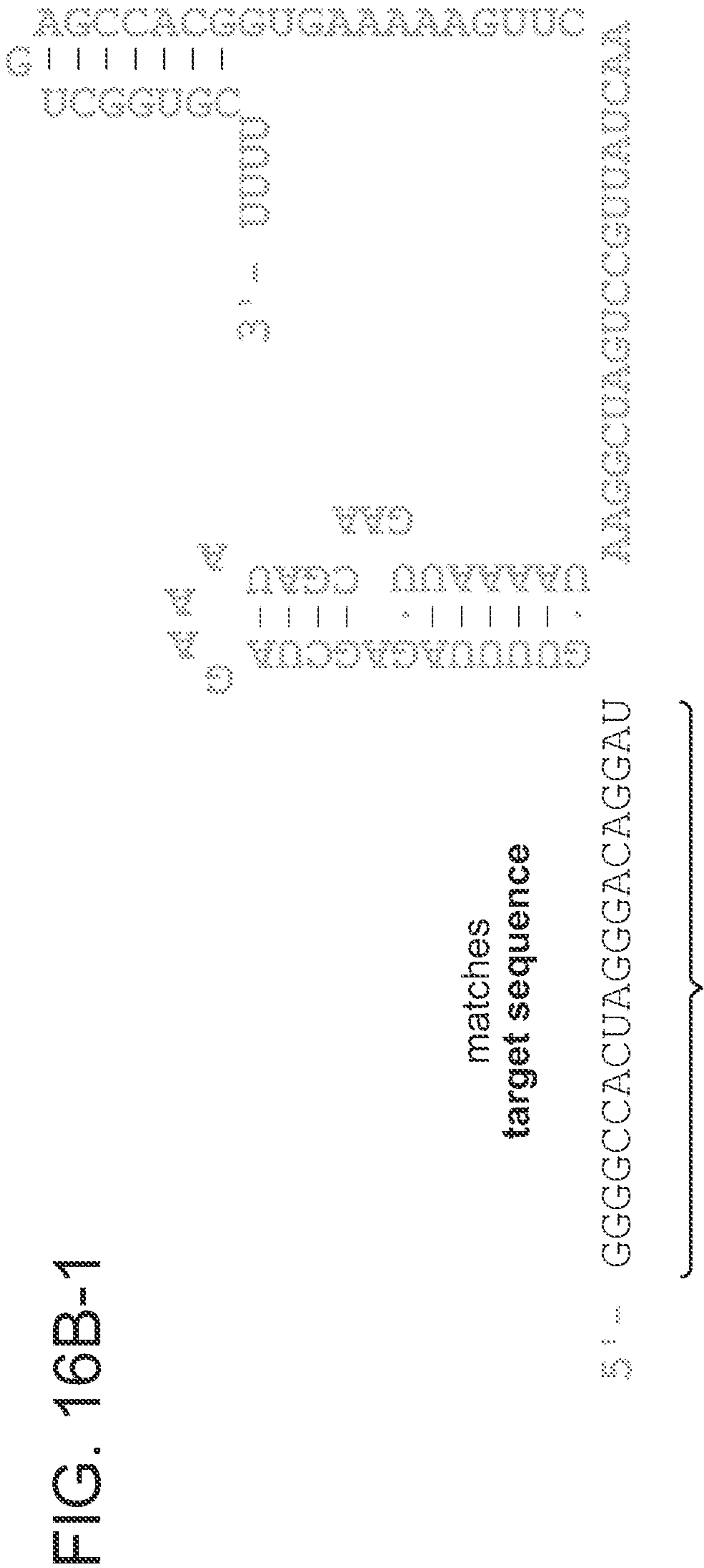
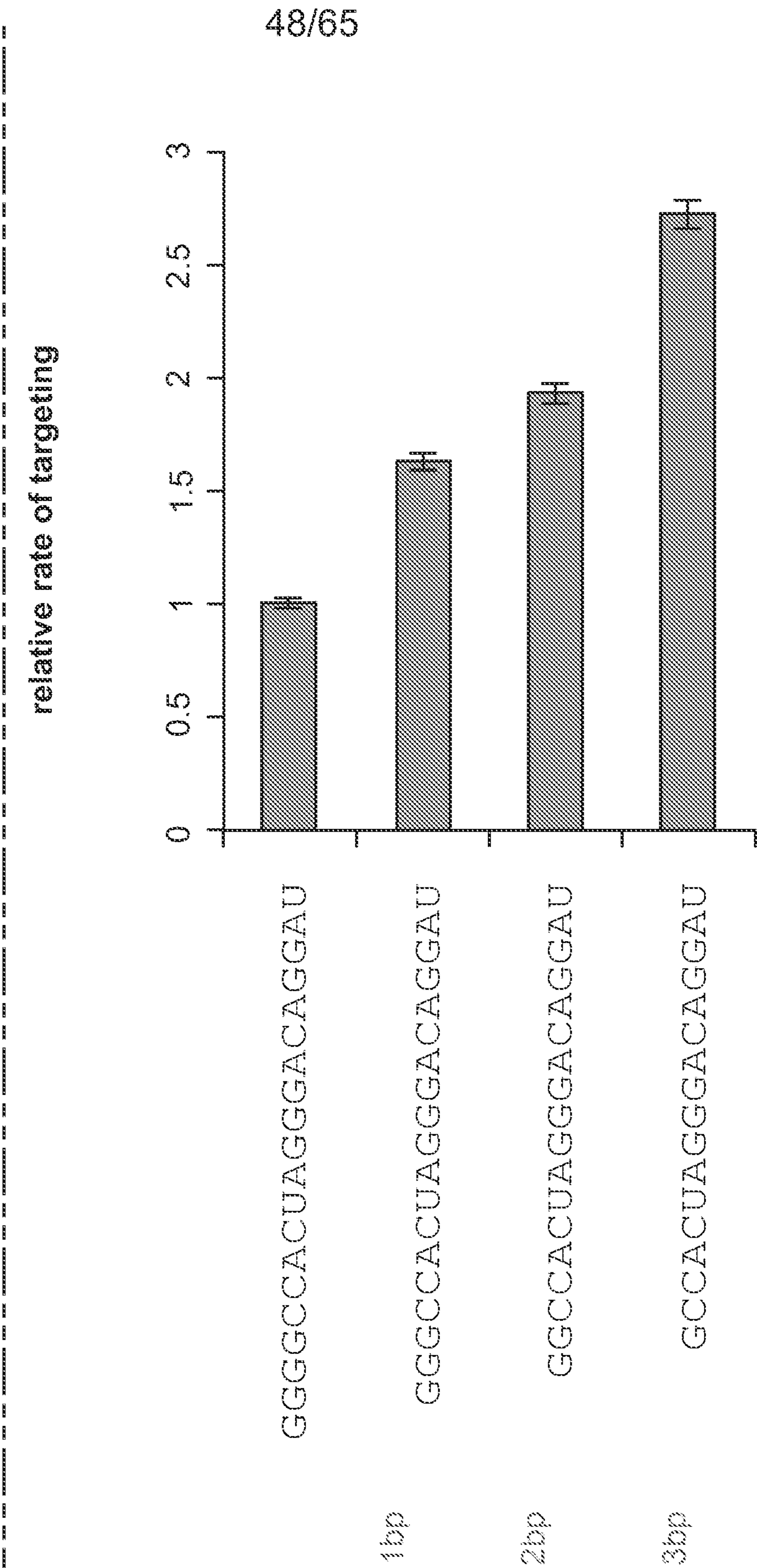
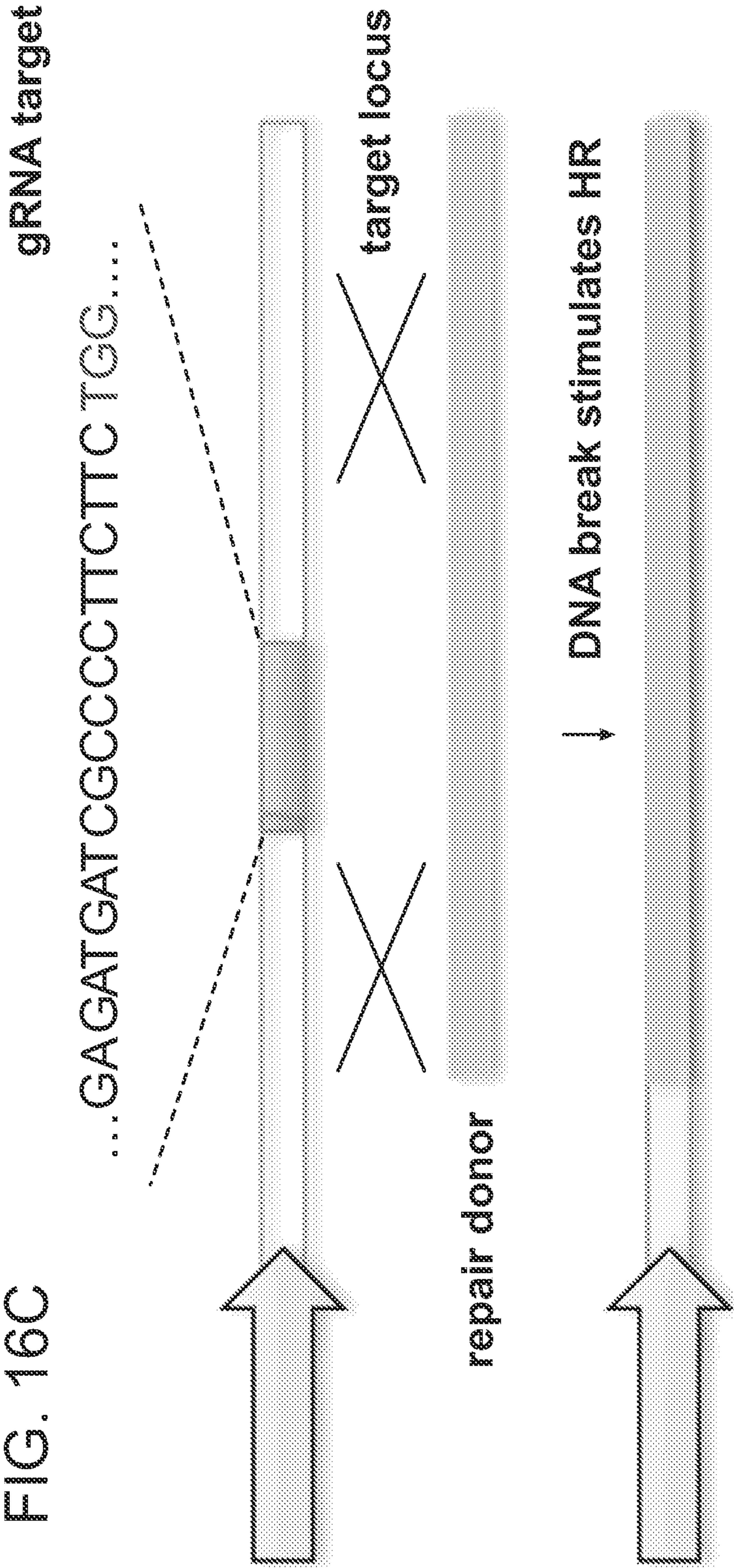


FIG. 16B-2





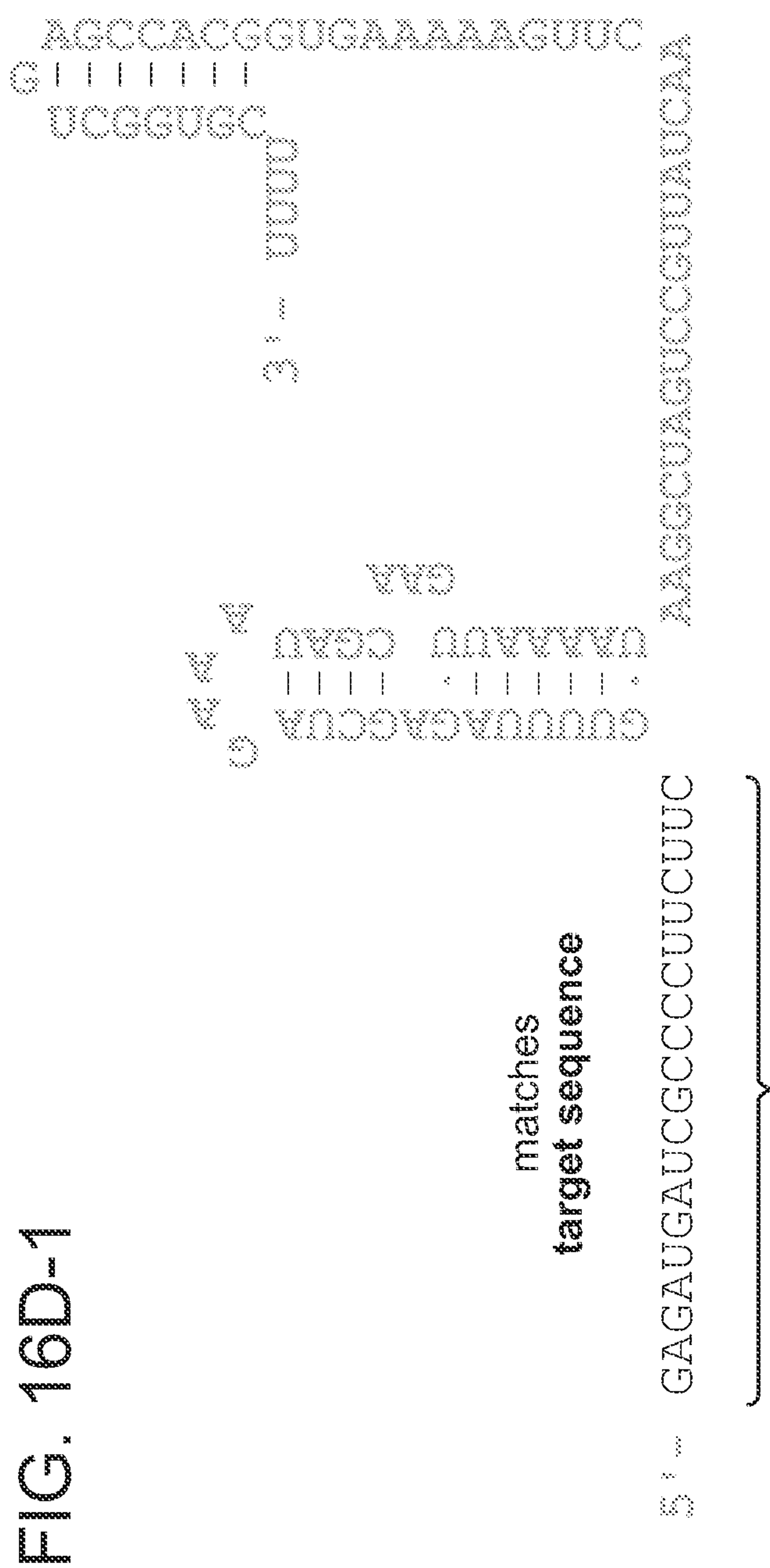
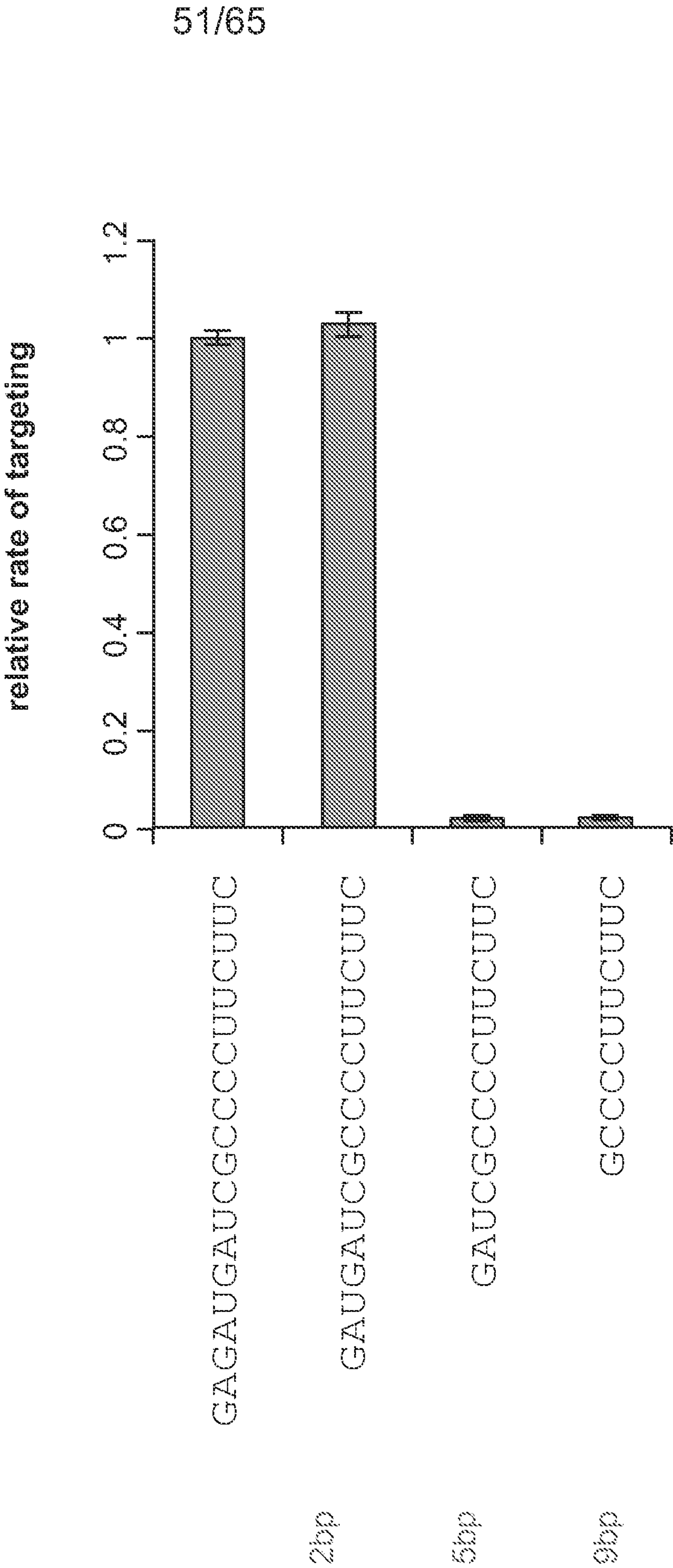


FIG. 16D-1

FIG. 16D-2



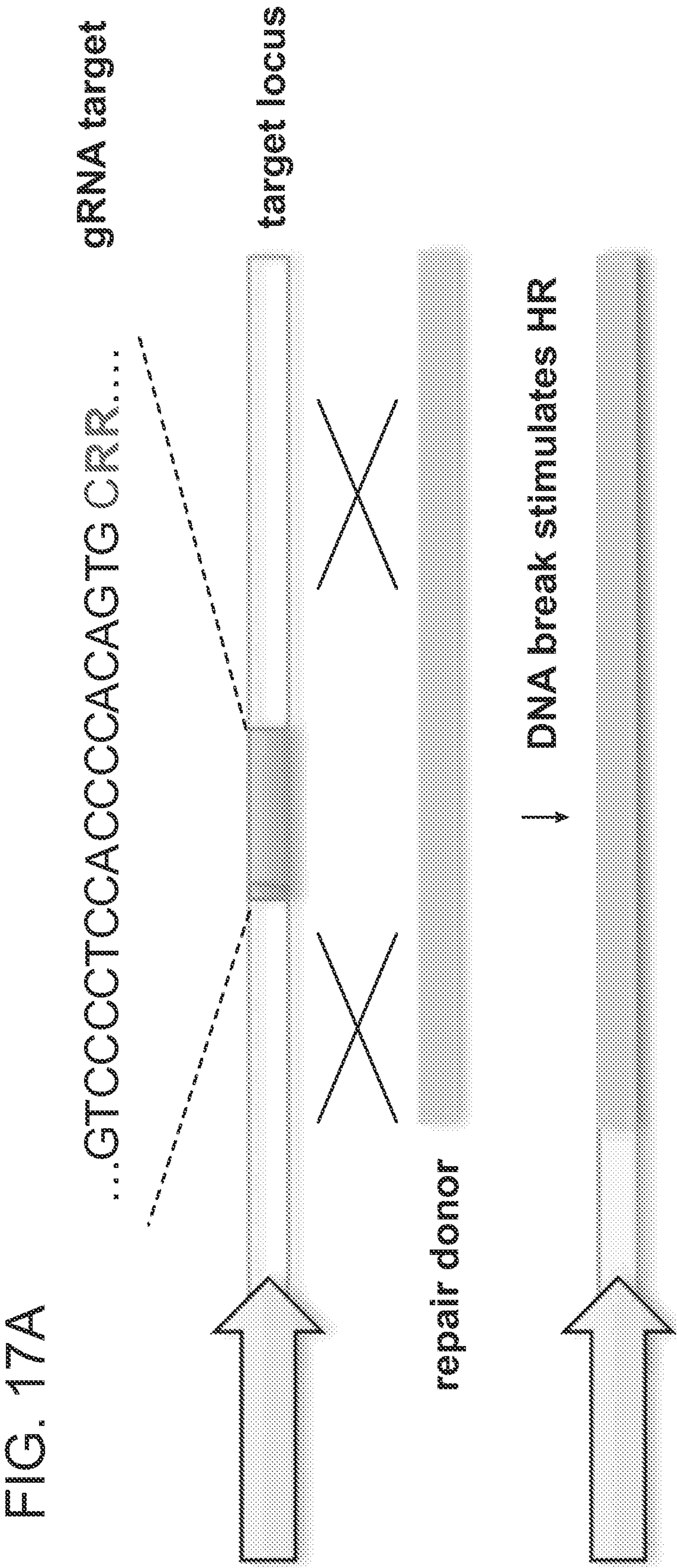


FIG. 17B

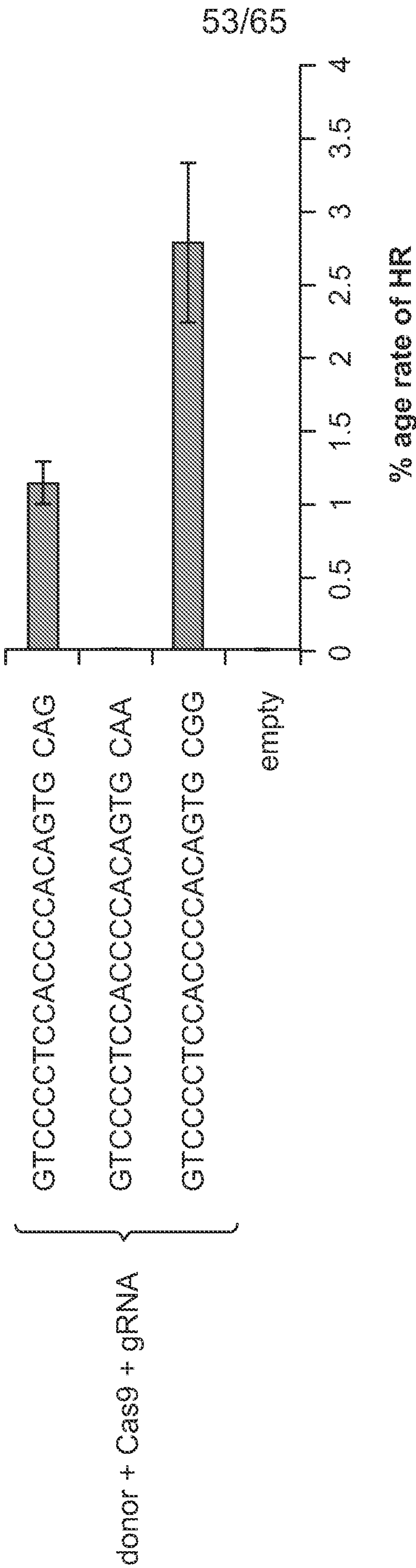
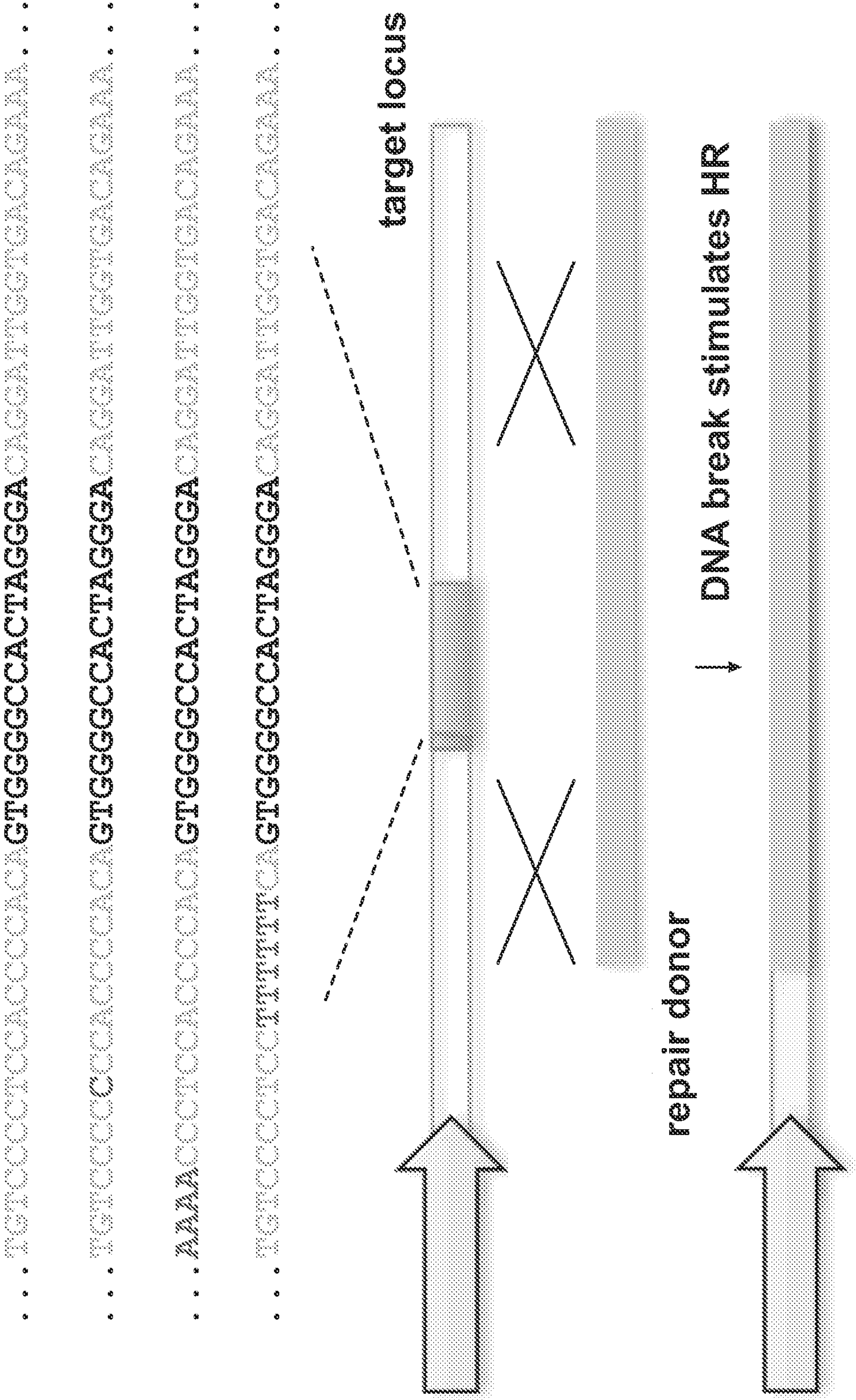


FIG. 18A



55/65

FIG. 18B

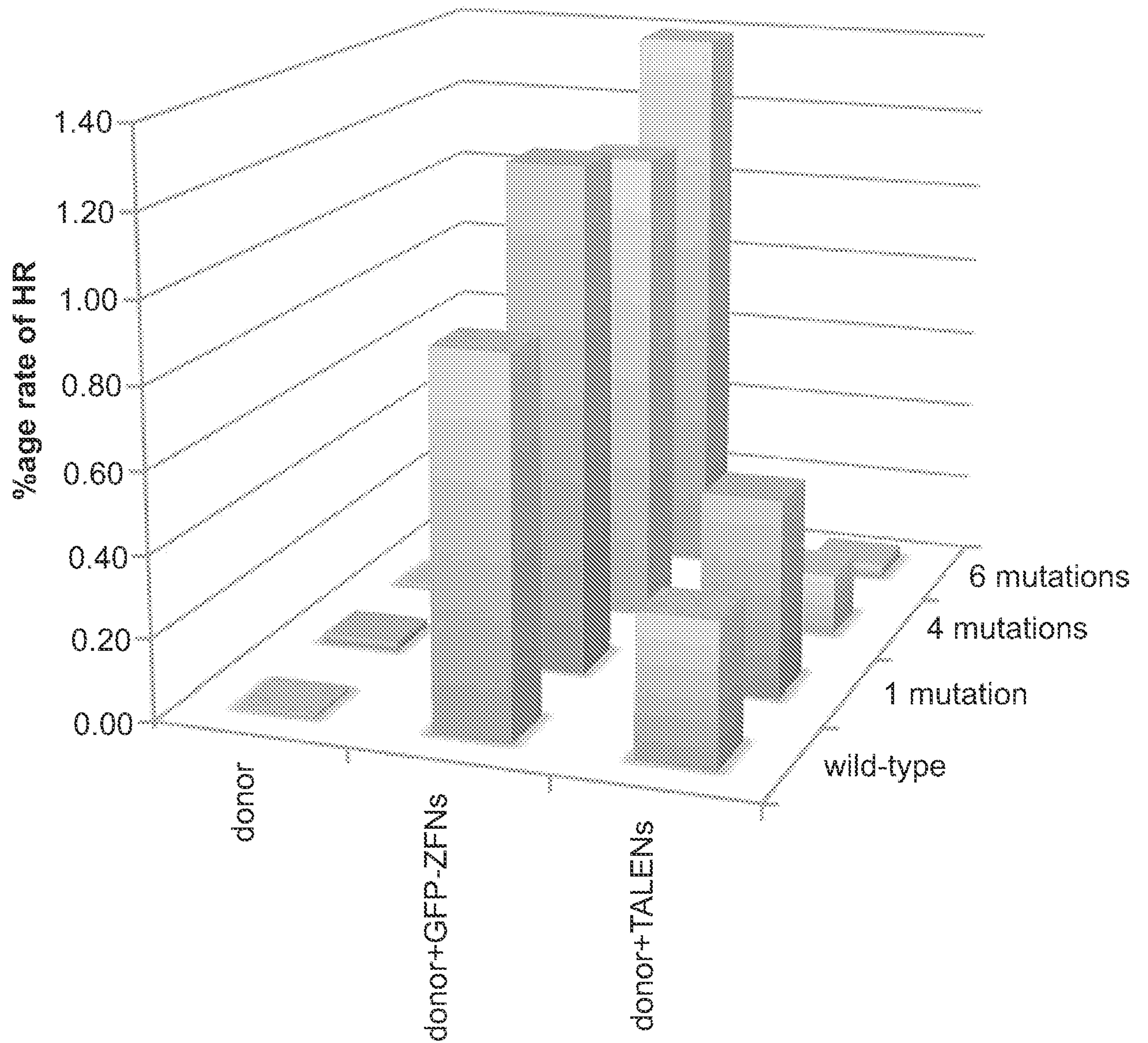


FIG. 19A

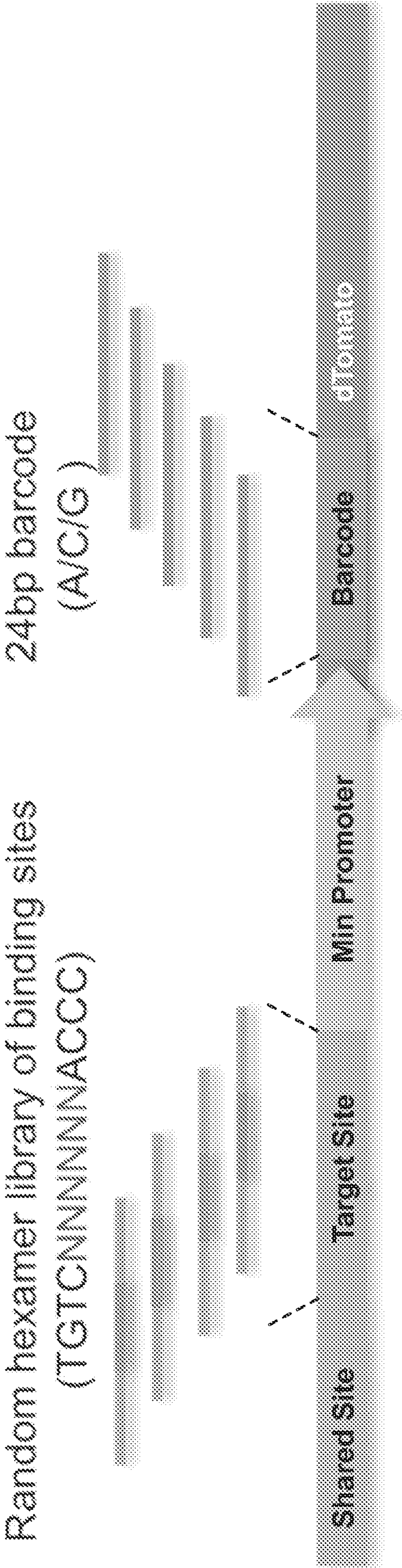
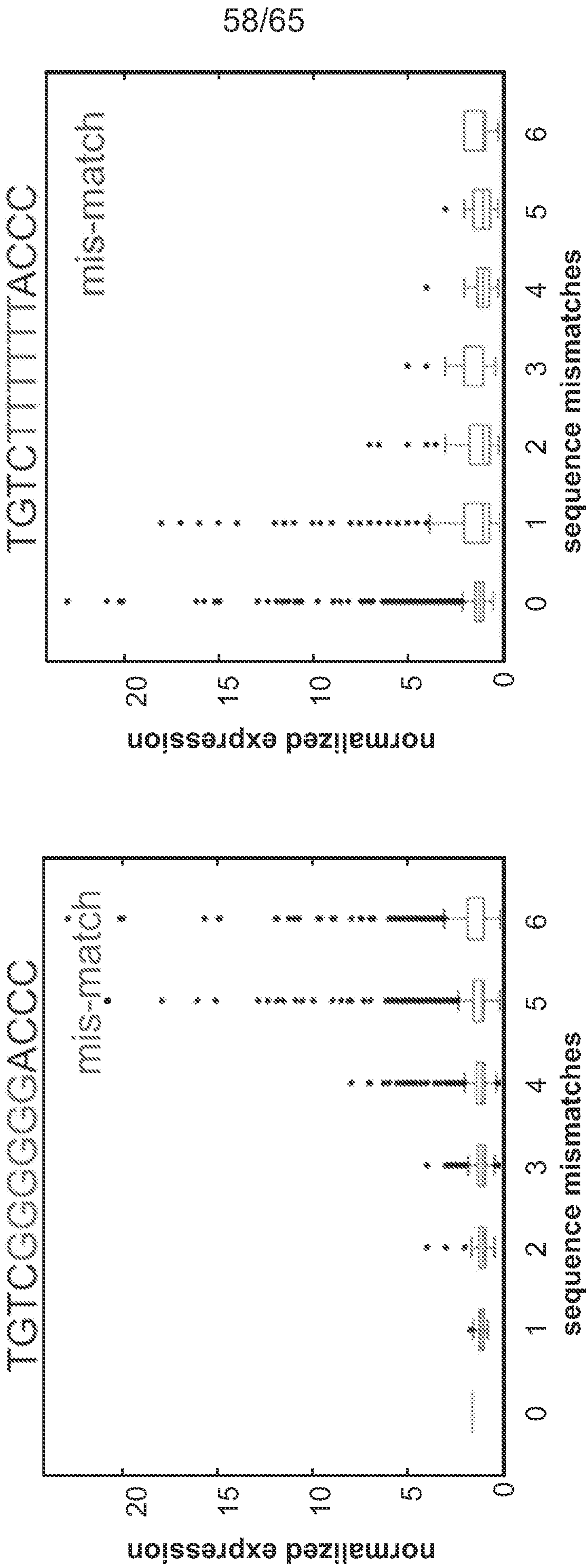


FIG. 19B-2



1997

Target TALE-TF:

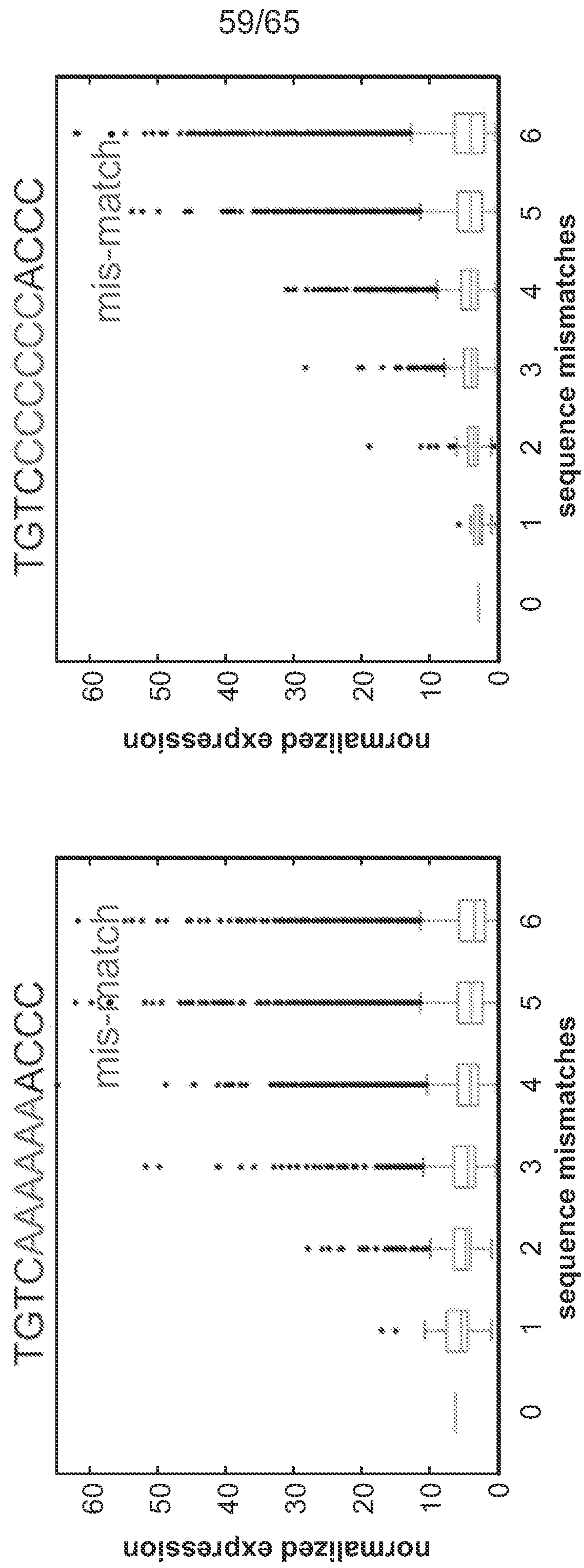


FIG. 19C-2

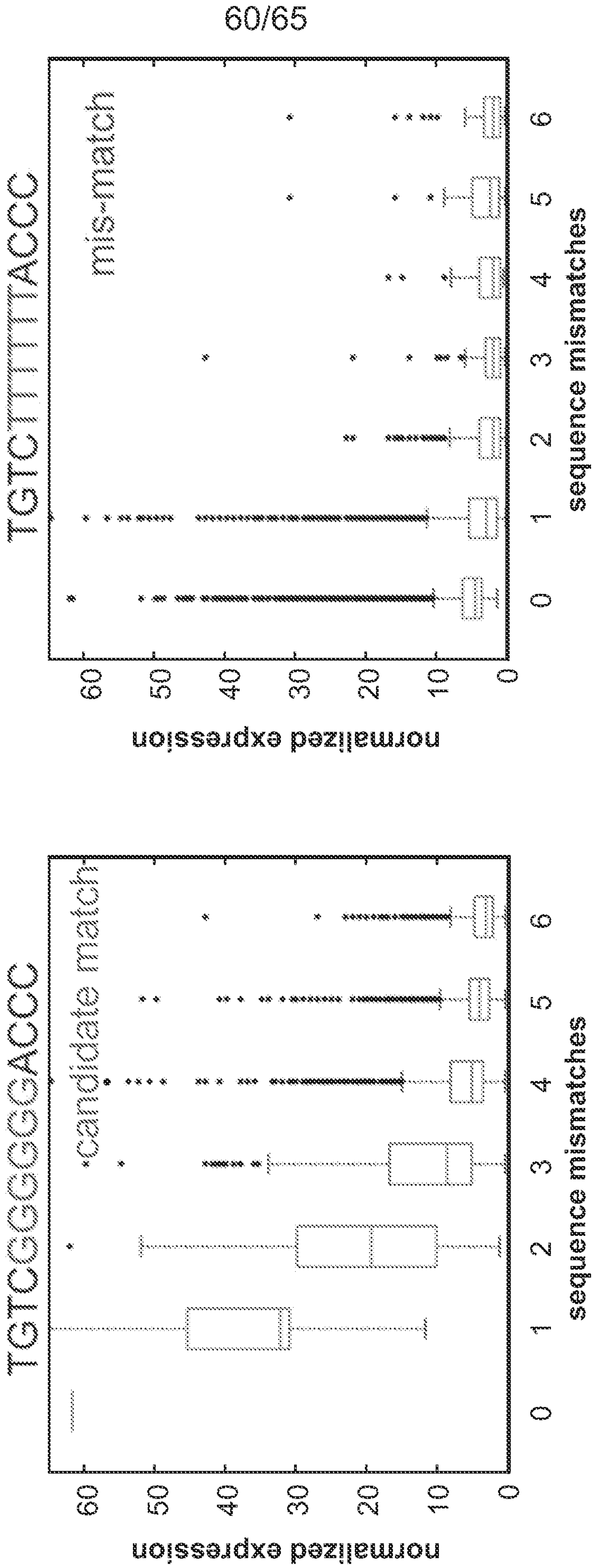
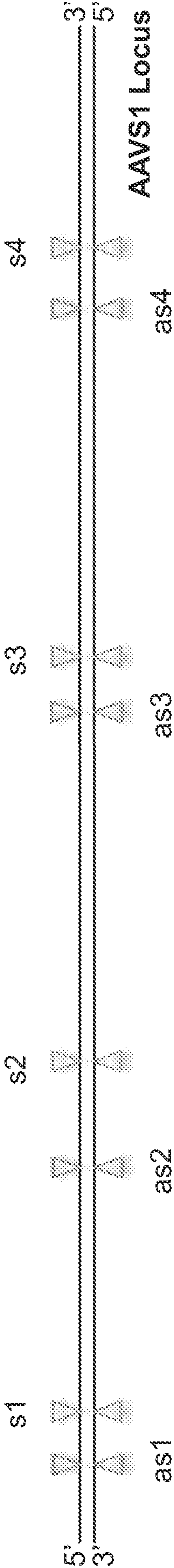


FIG. 20A



gRNA	Sequence
AAVS1__s1	GGATCCTGTGTCCCCGAGCT GGG
AAVS1__s2	GTTAATGTGGCTCTGGTTCT GGG
AAVS1__s3	GGGCCACTAGGGACAGGAT TGG
AAVS1__s4	CTTCCTAGTCTCCTGATATT GGG
AAVS1__as1	TGGTCCCAGCTCGGGACAC AGG
AAVS1__as2	AGAACCAGAGCCACATTAAAC CCG
AAVS1__as3	GTCACCAATCCTGTCCCTAG TGG
AAVS1__as4	AGACCCAAATATCAGGAGACT AGG

FIG. 20B

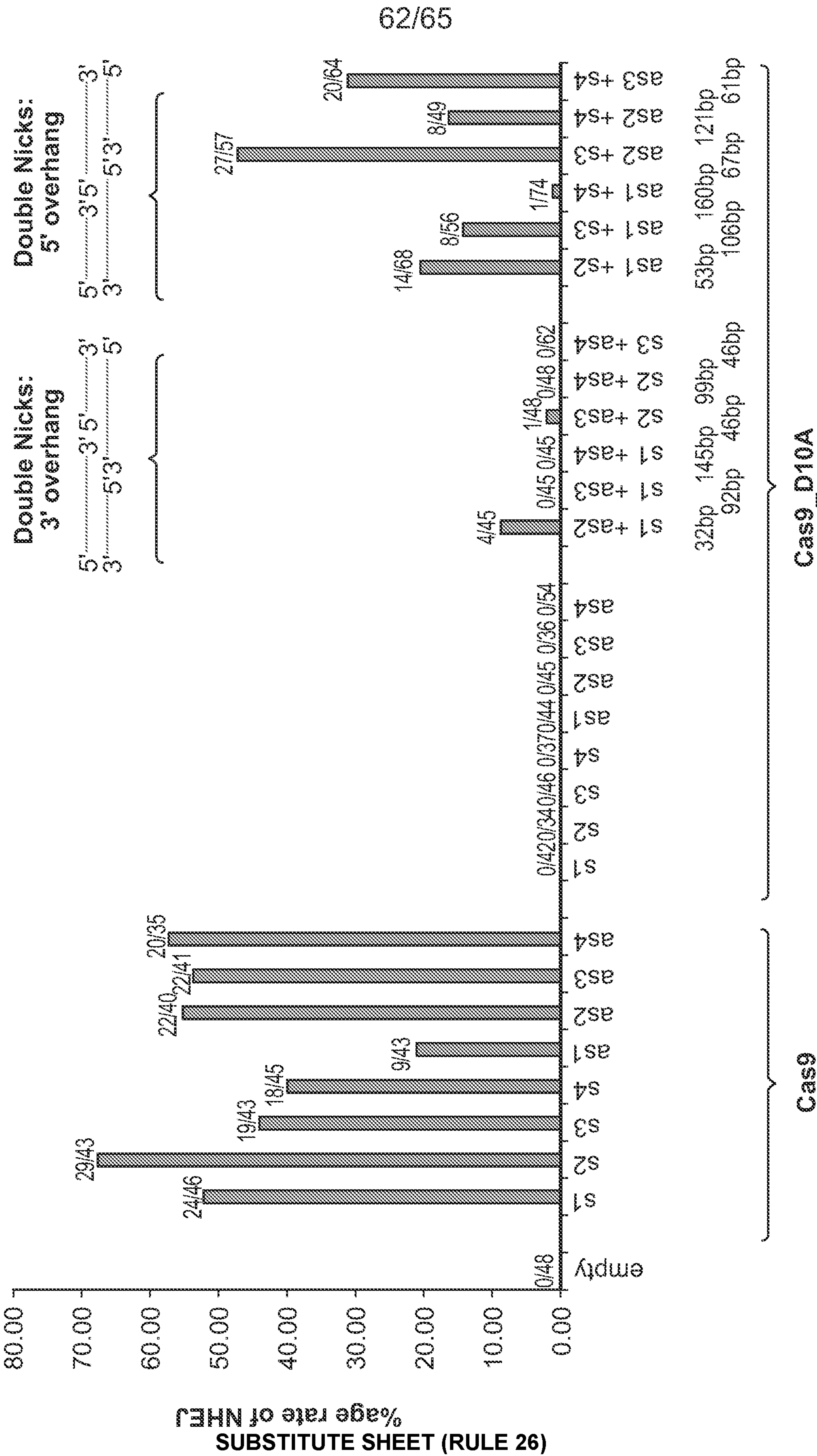


FIG. 1A

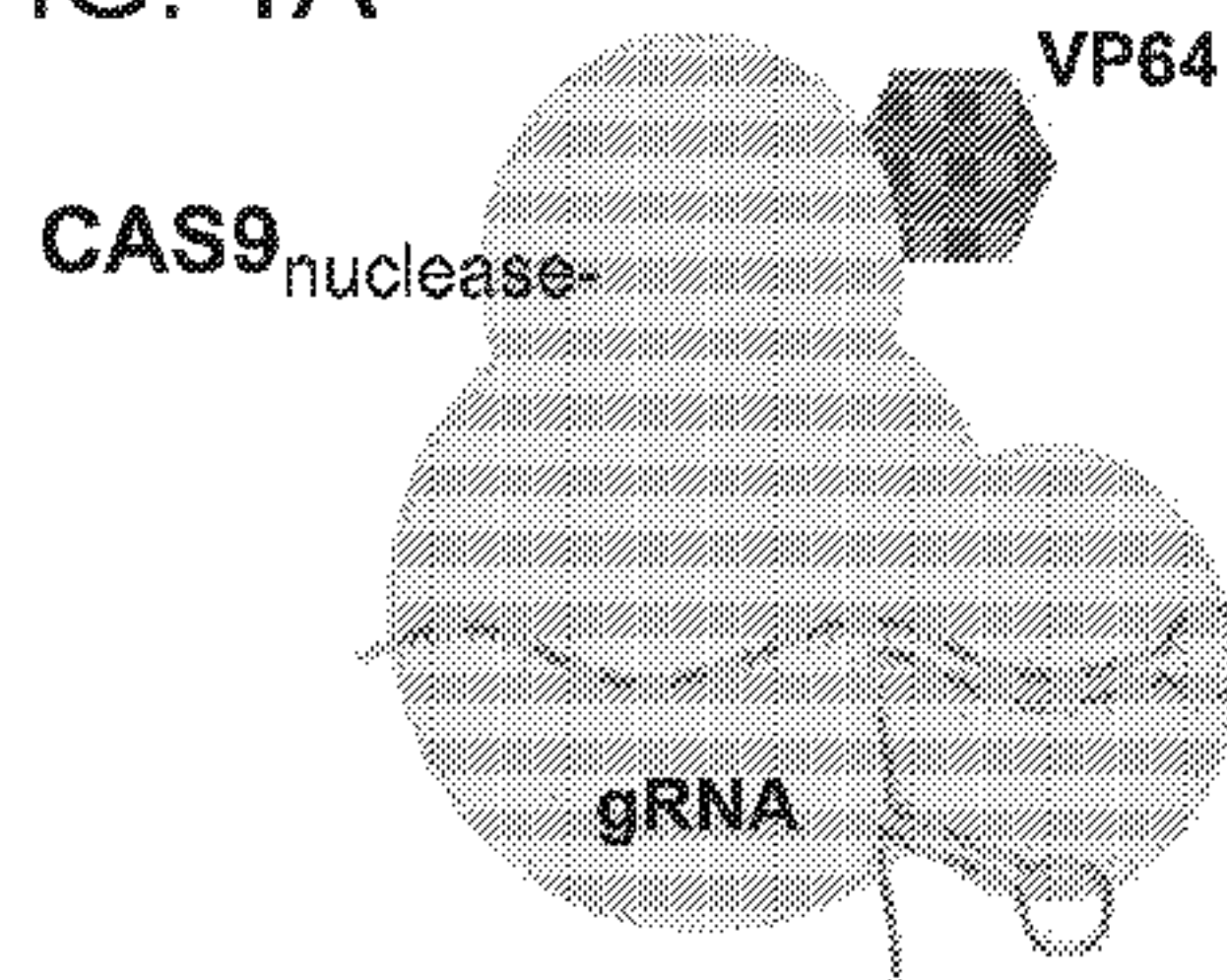


FIG. 1B

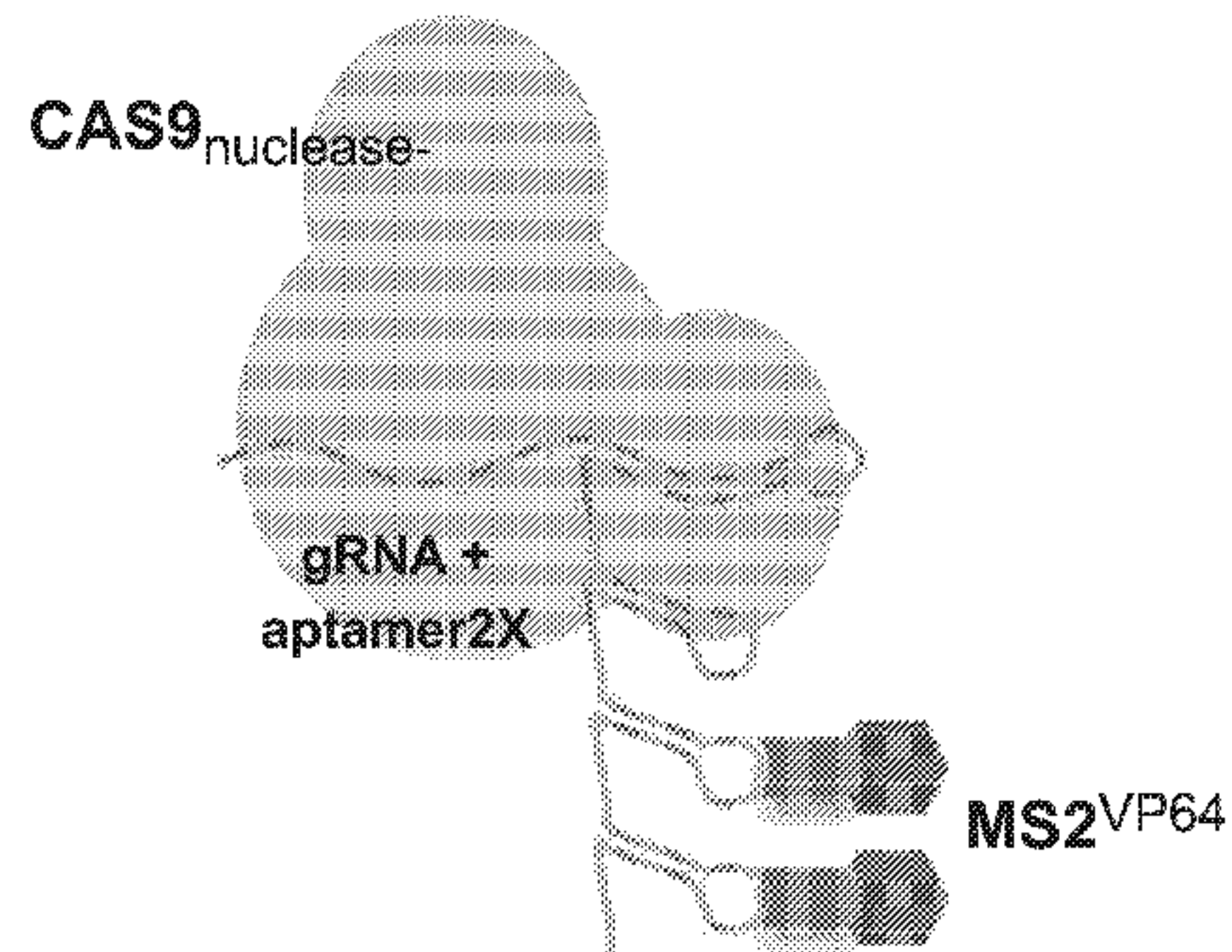


FIG. 1C

