Innovation, Sciences et Développement économique Canada
Office de la Propriété Intellectuelle du Canada

Innovation, Science and Economic Development Canada
Canadian Intellectual Property Office

CA 2907198 C 2019/12/10

(11)(21) **2 907 198**

(12) **BREVET CANADIEN**
CANADIAN PATENT

(13) **C**

(86) **Date de dépôt PCT/PCT Filing Date:** 2014/03/14
(87) **Date publication PCT/PCT Publication Date:** 2014/09/18
(45) **Date de délivrance/Issue Date:** 2019/12/10
(85) **Entrée phase nationale/National Entry:** 2015/09/15
(86) **N° demande PCT/PCT Application No.:** US 2014/028630
(87) **N° publication PCT/PCT Publication No.:** 2014/144288

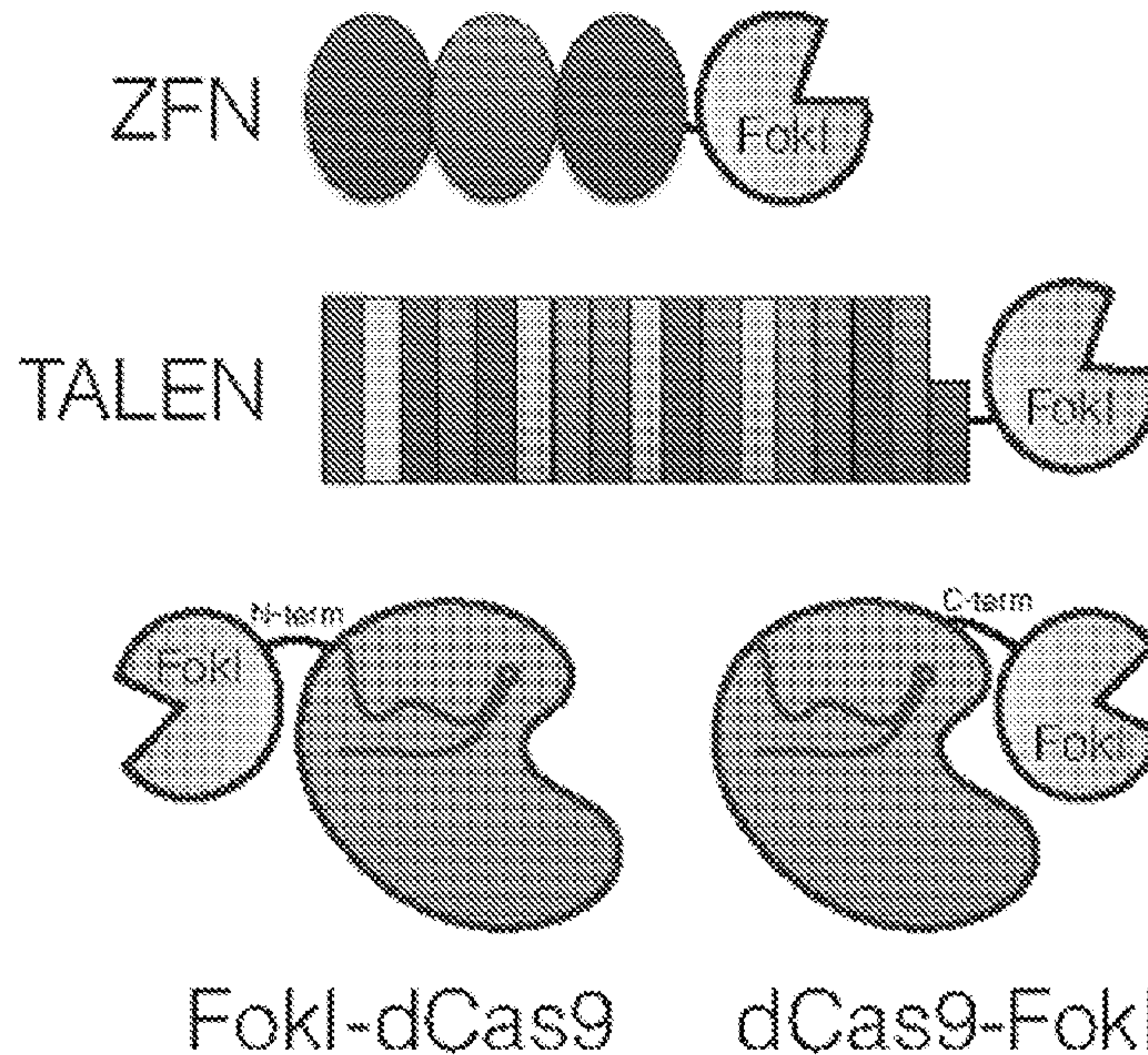
(51) **Cl.Int./Int.Cl.** *C12N 9/22* (2006.01), *A61K 38/46* (2006.01), *C07K 19/00* (2006.01), *C12N 15/09* (2006.01), *C12N 15/55* (2006.01), *C12N 15/62* (2006.01)

(72) **Inventeurs/Inventors:**
JOUNG, J. KEITH, US;
TSAI, SHENGDAR, US

(73) **Propriétaire/Owner:**
THE GENERAL HOSPITAL CORPORATION, US

(74) **Agent:** SMART & BIGGAR LLP

(54) **Titre :** UTILISATION DE NUCLEASES FOKI A GUIDAGE ARN (RFN) POUR AUGMENTER LA SPECIFICITE POUR LA MODIFICATION D'UN GENOME A GUIDAGE ARN
(54) **Title:** USING RNA-GUIDED FOKI NUCLEASES (RFNs) TO INCREASE SPECIFICITY FOR RNA-GUIDED GENOME EDITING



(57) **Abrégé/Abstract:**
Methods for increasing spe- cificity of RNA-guided genome editing, e.g., editing using CRISPR/Cas9 systems.

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

(19) World Intellectual Property Organization
International Bureau

(10) International Publication Number
WO 2014/144288 A1

(43) International Publication Date
18 September 2014 (18.09.2014)

(51) International Patent Classification:
C12P 21/06 (2006.01) *A61K 38/43* (2006.01)
C12P 19/34 (2006.01) *A61K 38/47* (2006.01)

(21) International Application Number: PCT/US2014/028630

(22) International Filing Date: 14 March 2014 (14.03.2014)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/799,647 15 March 2013 (15.03.2013) US
61/838,178 21 June 2013 (21.06.2013) US
61/838,148 21 June 2013 (21.06.2013) US
61/921,007 26 December 2013 (26.12.2013) US

(71) Applicant: THE GENERAL HOSPITAL CORPORATION [US/US]; 55 Fruit Street, Boston, Massachusetts 02114 (US).

(72) Inventors: JOUNG, J. Keith; 1 Magnolia Way, Winchester, Massachusetts 01890 (US). TSAI, Shengdar; 32 Prospect Street, Apartment 1, Charlestown, Massachusetts 02129 (US).

(74) Agents: KUGLER DEYOUNG, Janice et al.; Fish & Richardson P.C., P.O. Box 1022, Minneapolis, Minnesota 55440-1022 (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.

[Continued on next page]

(54) Title: USING RNA-GUIDED FOKI NUCLEASES (RFNS) TO INCREASE SPECIFICITY FOR RNA-GUIDED GENOME EDITING

(57) Abstract: Methods for increasing specificity of RNA-guided genome editing, e.g., editing using CRISPR/Cas9 systems.

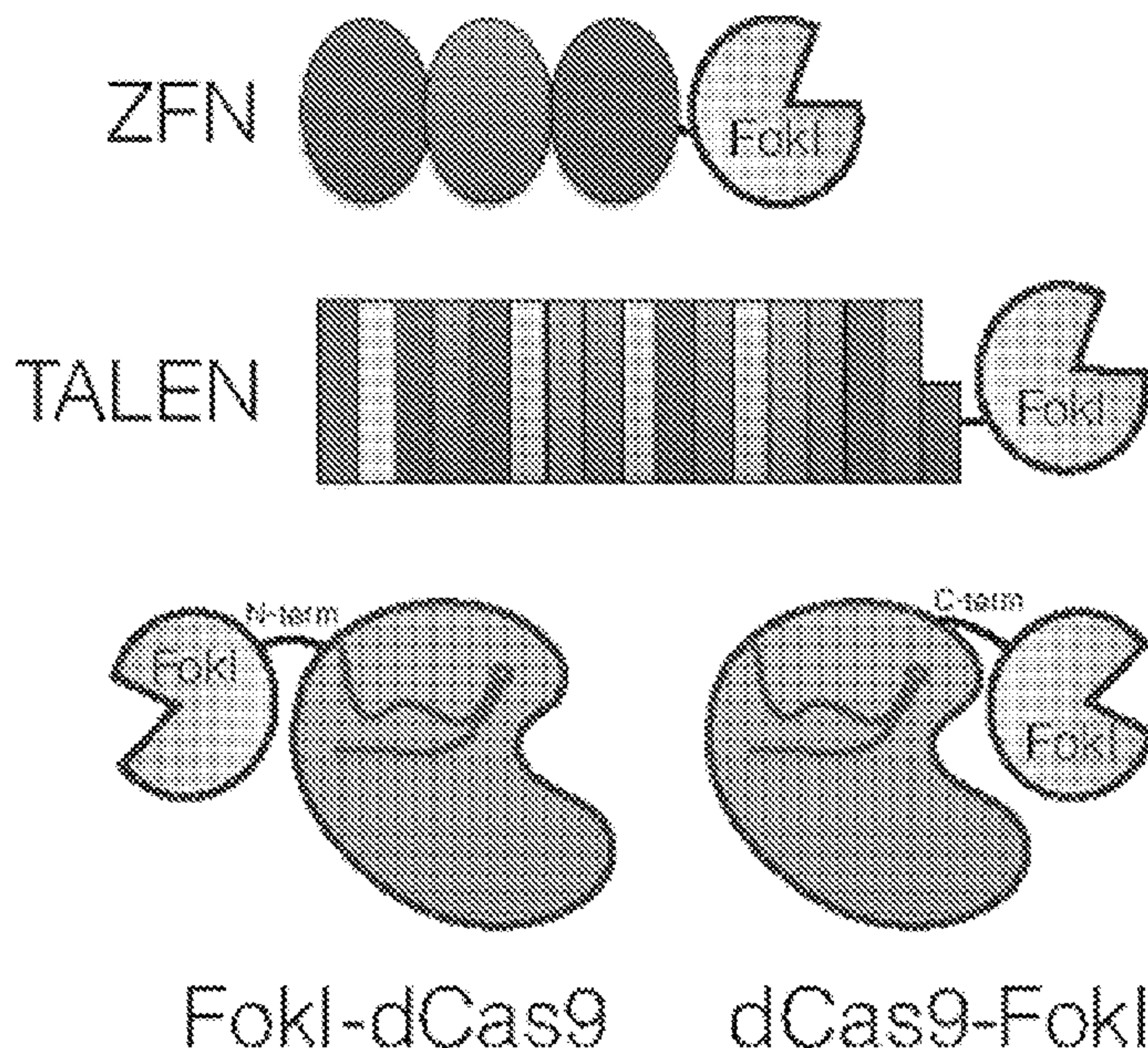


FIG. 5A

WO 2014/144288 A1

(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))*

60412-4906

Using RNA-guided FokI Nucleases (RFNs) to Increase Specificity for RNA-Guided Genome Editing

CLAIM OF PRIORITY

This application claims priority under 35 USC §119(e) to U.S. Patent Application Serial Nos. 61/799,647, filed on March 15, 2013; 61/838,178, filed on June 21, 2013; 61/838,148, filed on June 21, 2013, and 61/921,007, filed on December 26, 2013.

FEDERALLY SPONSORED RESEARCH OR DEVELOPMENT

This invention was made with Government support under Grant Nos. DP1 GM105378 awarded by the National Institutes of Health. The Government has certain rights in the invention.

TECHNICAL FIELD

Methods for increasing specificity of RNA-guided genome editing, e.g., editing using CRISPR/Cas9 systems, using RNA-guided FokI Nucleases (RFNs), e.g., FokI-dCas9 fusion proteins.

BACKGROUND

Recent work has demonstrated that clustered, regularly interspaced, short palindromic repeats (CRISPR)/CRISPR-associated (Cas) systems (Wiedenheft et al., Nature 482, 331-338 (2012); Horvath et al., Science 327, 167-170 (2010); Terns et al., Curr Opin Microbiol 14, 321-327 (2011)) can serve as the basis genome editing in bacteria, yeast and human cells, as well as *in vivo* in whole organisms such as fruit flies, zebrafish and mice (Wang et al., Cell 153, 910-918 (2013); Shen et al., Cell Res (2013); Dicarlo et al., Nucleic Acids Res (2013); Jiang et al., Nat Biotechnol 31, 233-239 (2013); Jinek et al., Elife 2, e00471 (2013); Hwang et al., Nat Biotechnol 31, 227-229 (2013); Cong et al., Science 339, 819-823 (2013); Mali et al., Science 339, 823-826 (2013c); Cho et al., Nat Biotechnol 31, 230-232 (2013); Gratz et al., Genetics 194(4):1029-35 (2013)). The Cas9 nuclease from *S. pyogenes* (hereafter simply Cas9) can be guided via base pair complementarity between the first 20 nucleotides of an engineered gRNA and the complementary strand of a target genomic DNA sequence

of interest that lies next to a protospacer adjacent motif (PAM), e.g., a PAM matching the sequence NGG or NAG (Shen et al., Cell Res (2013); Dicarlo et al., Nucleic Acids Res (2013); Jiang et al., Nat Biotechnol 31, 233-239 (2013); Jinek et al., Elife 2, e00471 (2013); Hwang et al., Nat Biotechnol 31, 227-229 (2013); Cong et al., Science
5 339, 819-823 (2013); Mali et al., Science 339, 823-826 (2013c); Cho et al., Nat Biotechnol 31, 230-232 (2013); Jinek et al., Science 337, 816-821 (2012)). Previous studies performed *in vitro* (Jinek et al., Science 337, 816-821 (2012)), in bacteria (Jiang et al., Nat Biotechnol 31, 233-239 (2013)) and in human cells (Cong et al., Science 339, 819-823 (2013)) have shown that Cas9-mediated cleavage can, in some
10 cases, be abolished by single mismatches at the gRNA/target site interface, particularly in the last 10-12 nucleotides (nts) located in the 3' end of the 20 nucleotide (nt) gRNA complementarity region.

SUMMARY

Many studies have shown that CRISPR-Cas nucleases can tolerate up to five
15 mismatches and still cleave; it is hard to predict the effects of any given single or combination of mismatches on activity. Taken together, these nucleases can show significant off-target effects but it can be challenging to predict these sites. Described herein are methods for increasing the specificity of genome editing using the CRISPR/Cas system, e.g., using RNA-guided FokI Nucleases (RFNs), e.g., FokI-
20 Cas9 or FokI-dCas9-based fusion proteins.

In a first aspect, the invention provides FokI-dCas9 fusion proteins, comprising a FokI catalytic domain sequence fused to the terminus, e.g., the N terminus, of dCas9, optionally with an intervening linker, e.g., a linker of from 2-30 amino acids, e.g., 4-12 amino acids, e.g., Gly_4Ser. In some embodiments, the FokI
25 catalytic domain comprises amino acids 388–583 or 408-583 of SEQ ID NO:4. In some embodiments, the dCas9 comprises mutations at the dCas9 comprises mutations at D10, E762, H983, or D986; and at H840 or N863; e.g., at: (i) D10A or D10N; and (ii) H840A, H840Y or H840N.

In another aspect, the invention provides nucleic acids encoding these fusion
30 proteins, vector comprising the nucleic acids, and host cells harboring or expressing the nucleic acids, vectors, or fusion proteins.

In another aspect, the invention provides methods for inducing a sequence-specific break in a double-stranded DNA molecule, e.g., in a genomic sequence in a cell, the method comprising expressing in the cell, or contacting the cell with, the FokI-dCas9 fusion protein described herein, and:

5 (a) two single guide RNAs, wherein each of the two single guide RNAs include sequences that are each complementary to one strand of the target sequence such that using both guide RNAs results in targeting both strands (i.e., one single guide RNA targets a first strand, and the other guide RNA targets the complementary strand), and FokI cuts each strand resulting in a pair of nicks on opposite DNA strands, thereby 10 creating a double-stranded break, or

(b) a tracrRNA and two crRNAs wherein each of the two crRNAs include sequences that are complementary to one strand of the target sequence such that using both crRNAs results in targeting both strands (i.e., one crRNA targets a first strand, and the other crRNA targets the complementary strand), and FokI cuts each strand 15 resulting in a pair of nicks on opposite DNA strands, thereby creating a double-stranded break.

In another aspect, the invention provides methods for increasing specificity of RNA-guided genome editing in a cell, the method comprising contacting the cell with an RNA-guided FokI Nuclease (RFN) fusion protein described herein.

20 The method may further comprise expressing in the cell, or contacting the cell with, (a) two single guide RNAs, wherein each of the two single guide RNAs include sequences that are each complementary to one strand of the target sequence such that using both guide RNAs results in targeting both strands (i.e., one single guide RNA targets a first strand, and the other guide RNA targets the complementary strand), and 25 FokI cuts each strand resulting in a pair of nicks on opposite DNA strands, thereby creating a double-stranded break, or

(b) a tracrRNA and two crRNAs wherein each of the two crRNAs include sequences that are complementary to one strand of the target sequence such that using both crRNAs results in targeting both strands (i.e., one crRNA targets a first strand, 30 and the other crRNA targets the complementary strand), and FokI cuts each strand resulting in a pair of nicks on opposite DNA strands, thereby creating a double-stranded break.

In some embodiments, the two target genomic sequences (i.e., the sequences to which the target complementarity regions of the crRNA or single guide RNAs are complementary) are spaced 10-20 base pairs apart, preferably 13-17 base pairs apart.

In some embodiments, an indel mutation is induced between the two target
5 sequences.

In some embodiments, the specificity of RNA-guided genome editing in a cell is increased.

In some embodiments, there is provided an *in vitro* method of inducing a sequence-specific break in a genomic sequence in a cell, the method comprising expressing in
10 the cell, or contacting the cell with, the RNA-guided FokI Nuclease (RFN) fusion protein as described herein, and guide RNAs that direct the RFN to two target genomic sequences.

In another aspect, the invention provides a kit comprising: an RNA-guided FokI Nuclease (RFN) fusion protein comprising a FokI catalytic domain sequence fused to the amino terminus of a catalytically inactive *Streptococcus pyogenes* CRISPR-associated 9
15 (dCas9) protein comprising an amino acid sequence that is at least 95% identical to the amino acid sequence of SEQ ID NO: 5, wherein said catalytically inactive *S. pyogenes* Cas9 has point mutations at amino acid residues corresponding to positions (i) D10, E762, H983, or D986, and (ii) H840 or N863 of *S. pyogenes* Cas9, an intervening linker from 2 to 30 amino acids, and two guide RNAs that direct said RFN fusion protein to a first target genomic
20 sequences, wherein the guide RNAs that direct said RFN fusion protein to said first target genomic sequence and said second target genomic sequence are spaced 10 to 20 nucleotides apart, and said first target genomic sequence comprises a protospacer adjacent motif (PAM) recognition sequence positioned upstream of said first target genomic sequence and said second target genomic sequence comprises a PAM recognition sequence positioned
25 downstream of said second target genomic sequence.

In another aspect, the invention provides a composition comprising: a nucleic acid encoding an RNA-guided FokI Nuclease (RFN) fusion protein comprising a FokI catalytic domain sequence fused to the amino terminus of a catalytically inactive

Streptococcus pyogenes CRISPR-associated 9 (dCas9) protein comprising an amino acid sequence that is at least 95% identical to the amino acid sequence of SEQ ID NO: 5, wherein said catalytically inactive *S. pyogenes* Cas9 has point mutations at amino acid residues corresponding to positions (i) D10, E762, H983, or D986, and (ii) H840 or N863 of
5 *S. pyogenes* Cas9, an intervening linker from 2 to 30 amino acids; and a nucleic acid encoding two guide RNAs that direct said RFN fusion protein to a first target genomic sequences, wherein the guide RNAs that direct said RFN fusion protein to said first target genomic sequence and said second target genomic sequence are spaced 10 to 20 nucleotides apart, and said first target genomic sequence comprises a protospacer adjacent motif (PAM) recognition
10 sequence positioned upstream of said first target genomic sequence and said second target genomic sequence comprises a PAM recognition sequence positioned downstream of said second target genomic sequence.

In another aspect, the invention provides a composition comprising: an RNA-guided FokI Nuclease (RFN) fusion protein comprising a FokI catalytic domain sequence
15 fused to the amino terminus of a catalytically inactive *Streptococcus pyogenes* CRISPR-associated 9 (dCas9) protein comprising an amino acid sequence that is at least 95% identical to the amino acid sequence of SEQ ID NO: 5, wherein said catalytically inactive *S. pyogenes* Cas9 has point mutations at amino acid residues corresponding to positions (i) D10, E762, H983, or D986, and (ii) H840 or N863 of *S. pyogenes* Cas9, an intervening linker from 2 to 30
20 amino acids; and two guide RNAs that direct said RFN fusion protein to a first target genomic sequences, wherein the guide RNAs that direct said RFN fusion protein to said first target genomic sequence and said second target genomic sequence are spaced 10 to 20 nucleotides apart, and said first target genomic sequence comprises a protospacer adjacent motif (PAM) recognition sequence positioned upstream of said first target genomic sequence and said
25 second target genomic sequence comprises a PAM recognition sequence positioned downstream of said second target genomic sequence.

In some embodiments, there is provided use for inducing a sequence-specific break in a genomic sequence in a cell of the RNA-guided FokI Nuclease (RFN) fusion protein as described herein, and guide RNAs that direct the RFN to two target genomic sequences.

In some embodiments, there is provided use for increasing specificity of RNA-guided genome editing in a cell of an RNA-guided FokI Nuclease (RFN) fusion protein as described herein.

Unless otherwise defined, all technical and scientific terms used herein have
5 the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Methods and materials are described herein for use in the present invention; other, suitable methods and materials known in the art can also be used. The materials, methods, and examples are illustrative only and not intended to be limiting. In case of conflict, the present specification, including definitions, will control.

10 Other features and advantages of the invention will be apparent from the following detailed description and figures, and from the claims.

DESCRIPTION OF DRAWINGS

Figure 1: Schematic illustrating a gRNA/Cas9 nuclease complex bound to its target DNA site. Scissors indicate approximate cleavage points of the Cas9 nuclease on the
15 genomic DNA target site. Note the numbering of nucleotides on the guide RNA proceeds in an inverse fashion from 5' to 3'.

Figure 2A: Schematic illustrating the rationale for truncating the 5' complementarity region of a gRNA. Thick grey lines = target DNA site, thin dark grey line structure = gRNA, grey oval = Cas9 nuclease, black lines indicate base pairing between
20 gRNA and target DNA site.

Figure 2B: Schematic overview of the EGFP disruption assay. Repair of targeted Cas9-mediated double-stranded breaks in a single integrated EGFP-PEST reporter gene by error-prone NHEJ-mediated repair leads to frame-shift mutations that disrupt the coding sequence and associated loss of fluorescence in cells.

Figures 2C-F: Activities of RGNs harboring sgRNAs bearing (C) single mismatches, (D) adjacent double mismatches, (E) variably spaced double mismatches, and (F) increasing numbers of adjacent mismatches assayed on three different target sites in the EGFP reporter gene sequence. Mean activities of replicates (see Online Methods) are shown, normalized to the activity of a perfectly matched gRNA. Error bars indicate standard errors of the mean. Positions mismatched in each gRNA are highlighted in grey in the grid below. Sequences of the three EGFP target sites were as follows:

EGFP Site 1 GGGCACGGGCAGCTTGCCGGTGG (SEQ ID NO:1)
EGFP Site 2 GATGCCGTTCTTCTGCTTGTCGG (SEQ ID NO:2)
EGFP Site 3 GGTGGTGCAGATGAACTTCAGGG (SEQ ID NO:3)

Figure 2G: Mismatches at the 5' end of the gRNA make CRISPR/Cas more sensitive more 3' mismatches. The gRNAs Watson-Crick base pair between the RNA&DNA with the exception of positions indicated with an "m" which are mismatched using the Watson-Crick transversion (i.e. EGFP Site#2 M18-19 is mismatched by changing the gRNA to its Watson-Crick partner at positions 18 & 19. Although positions near the 5' of the gRNA are generally very well tolerated, matches in these positions are important for nuclease activity when other residues are mismatched. When all four positions are mismatched, nuclease activity is no longer detectable. This further demonstrates that matches at these 5' position can help compensate for mismatches at other more 3' positions. Note these experiments were performed with a non-codon optimized version of Cas9 which can show lower absolute levels of nuclease activity as compared to the codon optimized version.

Figure 2H: Efficiency of Cas9 nuclease activities directed by gRNAs bearing variable length complementarity regions ranging from 15 to 25 nts in a human cell-based U2OS EGFP disruption assay. Expression of a gRNA from the U6 promoter requires the presence of a 5' G and therefore it was only possible to evaluate gRNAs harboring certain lengths of complementarity to the target DNA site (15, 17, 19, 20, 21, 23, and 25 nts).

Figure 3A: Efficiencies of EGFP disruption in human cells mediated by Cas9 and full-length or shortened gRNAs for four target sites in the *EGFP* reporter gene. Lengths of complementarity regions and corresponding target DNA sites are shown. Ctrl = control gRNA lacking a complementarity region.

Figure 3B: Efficiencies of targeted indel mutations introduced at seven different human endogenous gene targets by matched standard and tru-RGNs. Lengths of complementarity regions and corresponding target DNA sites are shown. Indel frequencies were measured by T7EI assay. Ctrl = control gRNA lacking a
5 complementarity region.

Figure 3C: DNA sequences of indel mutations induced by RGNs using a tru-gRNA or a matched full-length gRNA targeted to the EMX1 site. The portion of the target DNA site that interacts with the gRNA complementarity region is highlighted in grey with the first base of the PAM sequence shown in lowercase. Deletions are
10 indicated by dashes highlighted in grey and insertions by italicized letters highlighted in grey. The net number of bases deleted or inserted and the number of times each sequence was isolated are shown to the right.

Figure 3D: Efficiencies of precise HDR/ssODN-mediated alterations introduced at two endogenous human genes by matched standard and tru-RGNs.
15 %HDR was measured using a *BamH*I restriction digest assay (see the Experimental Procedures for Example 2). Control gRNA = empty U6 promoter vector.

Figure 3E: U2OS.EGFP cells were transfected with variable amounts of full-length gRNA expression plasmids (top) or tru-gRNA expression plasmids (bottom) together with a fixed amount of Cas9 expression plasmid and then assayed for
20 percentage of cells with decreased EGFP expression. Mean values from duplicate experiments are shown with standard errors of the mean. Note that the data obtained with tru-gRNA matches closely with data from experiments performed with full-length gRNA expression plasmids instead of tru-gRNA plasmids for these three EGFP target sites.

25 Figure 3F: U2OS.EGFP cells were transfected with variable amount of Cas9 expression plasmid together with variable amounts of full-length gRNA expression plasmids (top) or tru-gRNA expression plasmids (bottom) (amounts determined for each tru-gRNA from the experiments of Figure 3E). Mean values from duplicate experiments are shown with standard errors of the mean. Note that the data obtained
30 with tru-gRNA matches closely with data from experiments performed with full-length gRNA expression plasmids instead of tru-gRNA plasmids for these three EGFP target sites. The results of these titrations determined the concentrations of plasmids used in the EGFP disruption assays performed in Examples 1 and 2.

Figures 4A-C. RNA-guided FokI nucleases and a CRISPR/Cas Subtype Ypest protein 4 (Csy4)-based multiplex gRNA expression system.

(a) Schematic overview of RNA-guided FokI nucleases. Two FokI-dCas9 fusion proteins are recruited to adjacent target sites by two different gRNAs in order
5 to facilitate FokI dimerization and DNA cleavage.

(b) Schematic overview of a Csy4-based multiplex gRNA expression system. Two gRNAs (with any 5' end nucleotide) are co-expressed in a single transcript from a U6 promoter with each gRNA flanked by Csy4 recognition sites. Csy4 cleaves and releases gRNAs from the transcript. The Csy4 recognition site remains at the 3' end
10 of the gRNA with a Csy4 nuclease bound to that site.

(c) Validation of the multiplex, Csy4-based system. Two gRNAs targeted to adjacent sites in *EGFP* were expressed in a single RNA transcript using the Csy4-based system in human U2OS.EGFP cells together with Csy4 and Cas9 nucleases. Sequences of indel mutations induced in these cells are shown. The wild-type
15 sequence is shown in the top with both target sites highlighted in grey and PAM sequences shown as underlined text. Deletions are indicated by dashes against gray background and insertions by lowercase letters against a grey background. To the right of each sequence, the sizes of insertions (+) or deletions (Δ) are specified.

Figures 5A-F. Design and optimization of RNA-guided FokI nucleases.
20 (a) Schematic illustrations of a ZFN, TALEN, FokI-dCas9 fusion, and dCas9-FokI fusion.

(b) Screening the EGFP disruption activities of FokI-dCas9 fusion with gRNA pairs targeted to half-sites in one of two orientations: PAMs in (left panel) and PAMs out (right panel). Half-sites were separated by spacer sequences of variable lengths
25 ranging from 0 to 31 bps. EGFP disruption was quantified by flow cytometry, n = 1. Corresponding data for the dCas9-FokI fusion and the same gRNA pairs is shown in Fig. 5E.

(c) Additional assessment of FokI-dCas9-mediated EGFP disruption activities on target sites with half-sites oriented with their PAMs out and with spacer lengths
30 ranging from 10 to 20 bp. EGFP disruption was quantified by flow cytometry. Error bars indicate standard errors of the mean (s.e.m.), n = 2.

(d) Mean EGFP disruption values of the data from (c) grouped according to spacer length. Error bars represent s.e.m.

(e) These plots show the results of a screen for dCas9-FokI activity in EGFP disruption assay in the U2OS.EGFP cells with 60 gRNA pairs with spacings of 0-31 bp and PAM in and PAM out orientations.

(f) Sequences of FokI-dCas9 induced mutations in U2OS cells are shown.
5 The 23-nt target sequence bound by Cas9 or FokI-dCas9 is labeled in grey. The protospacer adjacent motif or PAM sequence is labeled in boldface with underlining. Deletions are marked with dashes on a light grey background. Insertions are highlighted in grey. The net number of bases inserted or deleted are indicated in a column directly to the right of the sequences.

10 Figures 6A-D. Dimerization of FokI-dCas9 RFNs is required for efficient genome editing activity.

(a) EGFP disruption activities of two RFN pairs assessed in the presence of correctly targeted gRNA pairs (to *EGFP* sites 47 and 81) and pairs in which one or the other of the gRNAs has been replaced with another gRNA targeted to a non-
15 *EGFP* sequence (in the *VEGFA* gene). EGFP disruption was quantified by flow cytometry. EGFP, Enhanced Green Fluorescent Protein; VEGFA, Vascular Endothelial Growth Factor A. Error bars represent standard errors of the mean (s.e.m.), n = 3.

(b) Quantification of mutagenesis frequencies by T7EI assay performed with
20 genomic DNA from the same cells used in the EGFP disruption assay of (a). Error bars represent s.e.m., n = 3.

(c) Activities of RFNs targeted to sites in the *APC*, *MLH1* and *VEGFA* genes. For each target, we co-expressed FokI-dCas9 with a pair of cognate gRNAs, only one gRNA for the "left" half-site, or only one gRNA for the "right" half-site. Rates of
25 mutagenesis were measured by T7E1 assay. APC, Adenomatous polyposis coli; MLH1, mutL homolog 1; VEGFA, Vascular Endothelial Growth Factor A. Error bars represent s.e.m., n = 3.

(d) Mutagenesis frequencies of RFNs targeted to VEGFA site 1 at the on-target site and at five previously known off-target (OT) sites for one of the gRNAs
30 used to target VEGFA site 1. Frequencies of mutation were determined by deep sequencing. Each value reported was determined from a single deep sequencing library prepared from genomic DNA pooled from three independent transfection experiments. The value shown for the on-target VEGFA site 1 (marked with an

asterisk) is the same as the one shown in Fig. 4a below and is only shown here for ease of comparison with the values presented in this figure.

Figures 7A-B. Mutagenic activities of a Cas9 nickase or FokI-dCas9 co-expressed with a single gRNA.

5 (a) Indel mutation frequencies induced by FokI-dCas9 (left bars) or Cas9 nickase (middle bars) in the presence of one or two gRNAs targeted to six different human gene sites. For each gene target, we assessed indel frequencies with both gRNAs, only one gRNA for the "left" half-site, or only the other gRNA for the "right" half-site. Mutation frequencies were determined by deep sequencing. Each 10 indel frequency value reported was determined from a single deep sequencing library prepared from genomic DNA pooled from three independent transfection experiments. VEGFA, Vascular Endothelial Growth Factor A; DDB2, Damage-Specific DNA Binding Protein 2; FANCF, Fanconi Anemia, Complementation Group F; FES, Feline Sarcoma Oncogene; RUNX 1, Runt-Related Transcription Factor 1.

15 (b) Data from (a) presented as a fold-reduction in the indel frequency comparing values obtained for each target site with a gRNA pair to each of the single gRNA experiments or to the control experiment (no gRNA and no Cas9 nickase or FokI-dCas9). This fold-reduction was calculated for both FokI-dCas9 (left bars in each pair, lighter grey) and Cas9 nickase (right bars in each pair, darker grey).

20 Figures 8A-C: Single Cas9 nickases can introduce point mutations with high efficiencies into their target sites.

Frequencies of different point mutations found at each position in half-sites targeted by single gRNAs for (a) *VEGFA*, (b) *FANCF*, and (c) *RUNX1* gene targets in the presence of FokI-dCas9, Cas9 nickase, or a tdTomato control. Mutation 25 frequencies were determined by deep sequencing. Each point mutation value reported was determined from a single deep sequencing library prepared from genomic DNA pooled from three independent transfection experiments. Note that the genomic DNA used for these experiments was isolated from the same cells analyzed for indel mutations in Figs. 7A-B. VEGFA, Vascular Endothelial Growth Factor A; FANCF, 30 Fanconi Anemia, Complementation Group F; RUNX 1, Runt-Related Transcription Factor 1.

DETAILED DESCRIPTION

CRISPR RNA-guided nucleases (RGNs) have rapidly emerged as a facile and efficient platform for genome editing. Although Marraffini and colleagues (Jiang et al., Nat Biotechnol 31, 233-239 (2013)) recently performed a systematic investigation of Cas9 RGN specificity in bacteria, the specificities of RGNs in human cells have not been extensively defined. Understanding the scope of RGN-mediated off-target effects in human and other eukaryotic cells will be critically essential if these nucleases are to be used widely for research and therapeutic applications. The present inventors have used a human cell-based reporter assay to characterize off-target cleavage of Cas9-based RGNs. Single and double mismatches were tolerated to varying degrees depending on their position along the guide RNA (gRNA)-DNA interface. Off-target alterations induced by four out of six RGNs targeted to endogenous loci in human cells were readily detected by examination of partially mismatched sites. The off-target sites identified harbor up to five mismatches and many are mutagenized with frequencies comparable to (or higher than) those observed at the intended on-target site. Thus RGNs are highly active even with imperfectly matched RNA-DNA interfaces in human cells, a finding that might confound their use in research and therapeutic applications.

The results described herein reveal that predicting the specificity profile of any given RGN is neither simple nor straightforward. The *EGFP* reporter assay experiments show that single and double mismatches can have variable effects on RGN activity in human cells that do not strictly depend upon their position(s) within the target site. For example, consistent with previously published reports, alterations in the 3' half of the gRNA/DNA interface generally have greater effects than those in the 5' half (Jiang et al., Nat Biotechnol 31, 233-239 (2013); Cong et al., Science 339, 819-823 (2013); Jinek et al., Science 337, 816-821 (2012)); however, single and double mutations in the 3' end sometimes also appear to be well tolerated whereas double mutations in the 5' end can greatly diminish activities. In addition, the magnitude of these effects for mismatches at any given position(s) appears to be site-dependent. Comprehensive profiling of a large series of RGNs with testing of all possible nucleotide substitutions (beyond the Watson-Crick transversions used in our *EGFP* reporter experiments) may help provide additional insights into the range of potential off-targets. In this regard, the recently described bacterial cell-based method

of Marraffini and colleagues (Jiang et al., Nat Biotechnol 31, 233-239 (2013)) or the *in vitro*, combinatorial library-based cleavage site-selection methodologies previously applied to ZFNs by Liu and colleagues (Pattanayak et al., Nat Methods 8, 765-770 (2011)) might be useful for generating larger sets of RGN specificity profiles.

5 Despite these challenges in comprehensively predicting RGN specificities, it was possible to identify *bona fide* off-targets of RGNs by examining a subset of genomic sites that differed from the on-target site by one to five mismatches. Notably, under conditions of these experiments, the frequencies of RGN-induced mutations at many of these off-target sites were similar to (or higher than) those observed at the 10 intended on-target site, enabling the detection of mutations at these sites using the T7EI assay (which, as performed in our laboratory, has a reliable detection limit of ~2 to 5% mutation frequency). Because these mutation rates were very high, it was possible to avoid using deep sequencing methods previously required to detect much lower frequency ZFN- and TALEN-induced off-target alterations (Pattanayak et al., 15 Nat Methods 8, 765-770 (2011); Perez et al., Nat Biotechnol 26, 808-816 (2008); Gabriel et al., Nat Biotechnol 29, 816-823 (2011); Hockemeyer et al., Nat Biotechnol 29, 731-734 (2011)). Analysis of RGN off-target mutagenesis in human cells also confirmed the difficulties of predicting RGN specificities – not all single and double mismatched off-target sites show evidence of mutation whereas some sites with as 20 many as five mismatches can also show alterations. Furthermore, the *bona fide* off-target sites identified do not exhibit any obvious bias toward transition or transversion differences relative to the intended target sequence.

Although off-target sites were seen for a number of RGNs, identification of these sites was neither comprehensive nor genome-wide in scale. For the six RGNs 25 studied, only a very small subset of the much larger total number of potential off-target sequences in the human genome was examined. Although examining such large numbers of loci for off-target mutations by T7EI assay is neither a practical nor a cost-effective strategy, the use of high-throughput sequencing in future studies might enable the interrogation of larger numbers of candidate off-target sites and 30 provide a more sensitive method for detecting *bona fide* off-target mutations. For example, such an approach might enable the unveiling of additional off-target sites for the two RGNs for which we failed to uncover any off-target mutations. In addition, an improved understanding both of RGN specificities and of any epigenomic factors

(e.g., DNA methylation and chromatin status) that may influence RGN activities in cells might also reduce the number of potential sites that need to be examined and thereby make genome-wide assessments of RGN off-targets more practical and affordable.

A number of strategies can be used to minimize the frequencies of genomic off-target mutations. For example, the specific choice of RGN target site can be optimized; given that off-target sites that differ at up to five positions from the intended target site can be efficiently mutated by RGNs, choosing target sites with minimal numbers of off-target sites as judged by mismatch counting seems unlikely to be effective; thousands of potential off-target sites that differ by four or five positions within the 20 bp RNA:DNA complementarity region will typically exist for any given RGN targeted to a sequence in the human genome. It is also possible that the nucleotide content of the gRNA complementarity region might influence the range of potential off-target effects. For example, high GC-content has been shown to stabilize RNA:DNA hybrids (Sugimoto et al., Biochemistry 34, 11211-11216 (1995)) and therefore might also be expected to make gRNA/genomic DNA hybridization more stable and more tolerant to mismatches. Additional experiments with larger numbers of gRNAs will be needed to assess if and how these two parameters (numbers of mismatched sites in the genome and stability of the RNA:DNA hybrid) influence the genome-wide specificities of RGNs. However, it is important to note that even if such predictive parameters can be defined, the effect of implementing such guidelines would be to further restrict the targeting range of RGNs.

One potential general strategy for reducing RGN-induced off-target effects might be to reduce the concentrations of gRNA and Cas9 nuclease expressed in the cell. This idea was tested using the RGNs for *VEGFA* target sites 2 and 3 in U2OS.EGFP cells; transfecting less gRNA- and Cas9-expressing plasmid decreased the mutation rate at the on-target site but did not appreciably change the relative rates of off-target mutations. Consistent with this, high-level off-target mutagenesis rates were also observed in two other human cell types (HEK293 and K562 cells) even though the absolute rates of on-target mutagenesis are lower than in U2OS.EGFP cells. Thus, reducing expression levels of gRNA and Cas9 in cells is not likely to provide a solution for reducing off-target effects. Furthermore, these results also

suggest that the high rates of off-target mutagenesis observed in human cells are not caused by overexpression of gRNA and/or Cas9.

The finding that significant off-target mutagenesis can be induced by RGNs in three different human cell types has important implications for broader use of this
5 genome-editing platform. For research applications, the potentially confounding effects of high frequency off-target mutations will need to be considered, particularly for experiments involving either cultured cells or organisms with slow generation times for which the outcrossing of undesired alterations would be challenging. One way to control for such effects might be to utilize multiple RGNs targeted to different
10 DNA sequences to induce the same genomic alteration because off-target effects are not random but instead related to the targeted site. However, for therapeutic applications, these findings clearly indicate that the specificities of RGNs will need to be carefully defined and/or improved if these nucleases are to be used safely in the longer term for treatment of human diseases.

15 **Methods for Improving Specificity**

As shown herein, CRISPR-Cas RNA-guided nucleases based on the *S. pyogenes* Cas9 protein can have significant off-target mutagenic effects that are comparable to or higher than the intended on-target activity (Example 1). Such off-target effects can be problematic for research and in particular for potential
20 therapeutic applications. Therefore, methods for improving the specificity of CRISPR-Cas RNA guided nucleases (RGNs) are needed.

As described in Example 1, Cas9 RGNs can induce high-frequency indel mutations at off-target sites in human cells (see also Cradick et al., 2013; Fu et al., 2013; Hsu et al., 2013; Pattanayak et al., 2013). These undesired alterations can occur
25 at genomic sequences that differ by as many as five mismatches from the intended on-target site (see Example 1). In addition, although mismatches at the 5' end of the gRNA complementarity region are generally better tolerated than those at the 3' end, these associations are not absolute and show site-to-site-dependence (see Example 1 and Fu et al., 2013; Hsu et al., 2013; Pattanayak et al., 2013). As a result,
30 computational methods that rely on the number and/or positions of mismatches currently have limited predictive value for identifying *bona fide* off-target sites. Therefore, methods for reducing the frequencies of off-target mutations remain an

important priority if RNA-guided nucleases are to be used for research and therapeutic applications.

Dimerization is an attractive potential strategy for improving the specificity of Cas9 nucleases. This is distinct from a paired Cas9 nickase approach, which is not a true dimeric system. Paired nickases work by co-localizing two Cas9 nickases on a segment of DNA, thereby inducing high efficiency genome editing via an undefined mechanism. Because dimerization is not required for enzymatic activity, single Cas9 nickases can also induce indels with high efficiencies at certain sites (via an unknown mechanism) and can therefore potentially cause unwanted off-target mutations in the genome.

Thus, one strategy to improve the specificity of RGNs is fusing a FokI endonuclease domain to a catalytically inactive form of Cas9 bearing the D10A and H840A mutations (also known as dCas9). FokI nuclease domain functions as a dimer and therefore two subunits must be recruited to the same local piece of DNA in order to induce a double-stranded break. In this configuration (**Figure 9A** and **Example 2**), two FokI-dCas9 fusions are recruited in an appropriate configuration using two different gRNAs to yield a double-stranded break. Thus, in this system, the FokI-dCas9 fusions would bind to a site that is twice as long as that of a single RGN and therefore this system would be expected to be more specific.

Therefore provided herein are FokI-dCas9 fusion proteins, wherein the FokI sequence is fused to dCas9 (preferably to the amino-terminal end of dCas9, but also optionally to the carboxy terminus), optionally with an intervening linker, e.g., a linker of from 2-30 amino acids, e.g., 4-12 amino acids, e.g., Gly_4Ser (SEQ ID NO:23) or $(Gly_4Ser)_3$. In some embodiments, the fusion proteins include a linker between the dCas9 and the FokI domains. Linkers that can be used in these fusion proteins (or between fusion proteins in a concatenated structure) can include any sequence that does not interfere with the function of the fusion proteins. In preferred embodiments, the linkers are short, e.g., 2-20 amino acids, and are typically flexible (i.e., comprising amino acids with a high degree of freedom such as glycine, alanine, and serine). In some embodiments, the linker comprises one or more units consisting of GGGS (SEQ ID NO:22) or GGGGS (SEQ ID NO:23), e.g., two, three, four, or more repeats of the GGGS (SEQ ID NO:22) or GGGGS (SEQ ID NO:23) unit. Other linker sequences can also be used.

Also described herein is a RNA-guided FokI nuclease platform in which dimerization, rather than just co-localization, is required for efficient genome editing activity. These nucleases can robustly mediate highly efficient genome editing in human cells and can reduce off-target mutations to undetectable levels as judged by sensitive deep sequencing methods. Also described is an efficient system for expressing pairs of gRNAs with any 5' end nucleotide, a method that confers a wider targeting range on the RFN platform. Finally, monomeric Cas9 nickases generally introduce more undesirable indels and point mutations than the nucleases described herein in the presence of a single gRNA. These results define a robust, user-friendly nuclease platform with the specificity advantages of a well-characterized dimeric architecture and an improved mutagenesis profile relative to paired Cas9 nickases, features that will be important for research or therapeutic applications requiring the highest possible genome editing precision.

Thus a new RNA-guided FokI Nuclease (RFN) platform is described herein for performing robust and highly specific genome editing in human cells. RFNs require two gRNAs for activity and function as dimers. Surprisingly, the engineering of an active RFN required fusion of the FokI nuclease domain to the amino-terminal end of the dCas9 protein, an architecture different from ZFNs and TALENs in which the FokI domain is fused to the carboxy-terminal end of engineered zinc finger or transcription activator-like effector repeat arrays. RFNs also require that the half-sites bound by each Fok-dCas9/gRNA complex have a particular relative orientation (PAMs out) with a relatively restricted intervening spacer length of 14 to 17 bps (although activity may be possible at additional spacings but with less consistent success).

The dimeric nature of RFNs provides important specificity advantages relative to standard monomeric Cas9 nucleases. In an ideal dimeric system, little to no activity will be observed with monomers on half-sites. The present data demonstrate that FokI-dCas9 directed by a single gRNA induces very little or no mutagenesis at RFN half-sites. 12 single gRNAs (for six RFN target sites) were tested with co-expressed FokI-dCas9 and indels were observed at very low frequencies (range of 0.0045% to 0.47%), in some cases at levels as low as background rates observed in control cells in which there was no expression of gRNA or nuclease. Given that the FokI nuclease domain functions as a dimer, it is presumed that any indels observed

with a single gRNA are likely due to recruitment of a FokI-dCas9 dimer to the DNA. Regardless of mechanism, given that only very low level mutagenesis was observed when FokI-dCas9 was tested with single gRNAs at 12 on-target half-sites, it is very unlikely that any mutagenesis will be induced at partially mismatched, off-target half-sites. Indeed, an RFN targeted to *VEGFA* did not induce detectable mutations at known off-target sites of one of the gRNAs as judged by deep sequencing.

Because RFNs are a true dimeric system, they possess a number of important advantages over paired nickase technology, which depends on co-localization but does not require dimerization. First, the direct comparisons herein show that single Cas9 nickases generally induce indel mutations with greater efficiencies than do FokI-dCas9 fusion proteins directed by the same individual gRNAs. Second, monomeric Cas9 nickases can also induce base pair substitutions in their target half-sites with high efficiencies, a previously unknown mutagenic side-effect that we uncovered in this study. Again, the direct comparisons show that monomeric Cas9 nickases induce these unwanted point mutations at substantially higher rates than FokI-dCas9 fusions guided by the same single gRNAs. Third, paired Cas9 nickases show greater promiscuity in the orientation and spacing of target half-sites than dimeric RFNs and therefore have a greater potential range of sites at which off-target mutations might be induced. Paired nickase half-sites can be oriented with their PAMs in or PAMs out and with spacer sequences ranging in length from 0 to 1000 bps (Ran et al., Cell 154, 1380-1389 (2013); Mali et al., Nat Biotechnol 31, 833-838 (2013); Cho et al., Genome Res (2013)). This promiscuity exists because the genome editing activities of Cas9 nickases do not depend on dimerization of the enzyme but rather just co-localization of the two nicks. By contrast, RFNs are much more stringent in their specificities -- half-sites must have their PAMs out and must be spaced apart by 14 to 17 bps, due to the requirement for two appropriately positioned FokI cleavage domains for efficient cleavage.

FokI

FokI is a type IIs restriction endonuclease that includes a DNA recognition domain and a catalytic (endonuclease) domain. The fusion proteins described herein can include all of FokI or just the catalytic endonuclease domain, e.g., amino acids 388–583 or 408-583 of GenBank Acc. No. AAA24927.1, e.g., as described in Li et al., Nucleic Acids Res. 39(1): 359–372 (2011); Cathomen and Joung, Mol. Ther. 16:

1200–1207 (2008), or a mutated form of FokI as described in Miller et al. Nat Biotechnol 25: 778–785 (2007); Szczepek et al., Nat Biotechnol 25: 786–793 (2007); or Bitinaite et al., Proc. Natl. Acad. Sci. USA. 95:10570–10575 (1998).

An exemplary amino acid sequence of FokI is as follows:

```
        10         20         30         40         50         60
MFLSMVSKIR TFGWVQNPGK FENLKRVVQV FDRNSKVHNE VKNIKIPTLV KESKIQKELV

        70         80         90        100        110        120
AIMNQHDLIY TYKELVGTGT SIRSEAPCDA IIQATIADQG NKKGYIDNWS SDGFLRWAHA

       130        140        150        160        170        180
LGFIEYINKS DSFVITDVGL AYSKSADGSA IEKEILIEAI SSYPPAIRIL TLLEDGQHLT

       190        200        210        220        230        240
KFDLGKNLGF SGESGFTSLP EGILLDTLAN AMPKDKGEIR NNWEGSSDKY ARMIGGWLDK

       250        260        270        280        290        300
LGLVKQGKKE FIIPTLGKPD NKEFISHAFK ITGEGLKVLR RAKGSTKFTR VPKRVYWEML

       310        320        330        340        350        360
ATNLTDKEYV RTRRALILEI LIKAGSLKIE QIQDNLKKLG FDEVIETIEN DIKGLINTGI

       370        380        390        400        410        420
FIEIKGRFYQ LKDHILQFVI PNRGVTKQLV KSELEEKKSE LRHKLKYVPH EYIELIEIAR

       430        440        450        460        470        480
NSTQDRILEM KVMEFFMKVY GYRGKHLGGS RKPDGAIYTV GSPIDYGVIV DTKAYSGGYN

       490        500        510        520        530        540
LPIGQADEMQ RYVEENQTRN KHINPNEWWK VYPSSVTEFK FLFVSGHFKG NYKAQLTRLN

       550        560        570        580
HITNCNGAVL SVEELLIGGE MIKAGTLTLE EVRRKFNNGE INF (SEQ ID NO:4)
```

An exemplary nucleic acid sequence encoding FokI is as follows:

```
ATGTTTTTGAGTATGGTTTCTAAAATAAGAACTTTCGGTTGGGTTCAAAATCCAGGTAAA
TTTGAGAATTTAAAACGAGTAGTTCAAGTATTTGATAGAAATTCTAAAGTACATAATGAA
GTGAAAAATATAAAGATACCAACCCTAGTCAAAGAAAGTAAGATCCAAAAAGAACTAGTT
GCTATTATGAATCAACATGATTTGATTTATACATATAAAGAGTTAGTAGGAACAGGAACT
TCAATACGTTCAGAAGCACCATGCGATGCAATTATTCAAGCAACAATAGCAGATCAAGGA
AATAAAAAAGGCTATATCGATAATTGGTCATCTGACGGTTTTTTGCGTTGGGCACATGCT
TTAGGATTTATTGAATATATAAATAAAAGTGATTCTTTTGTAATAACTGATGTTGGACTT
GCTTACTCTAAATCAGCTGACGGCAGCGCCATTGAAAAAGAGATTTTGATTGAAGCGATA
TCATCTTATCCTCCAGCGATTCGTATTTTAACTTTGCTAGAAGATGGACAACATTTGACA
AAGTTTGATCTTGGCAAGAATTTAGGTTTTAGTGGAGAAAGTGGATTTACTTCTCTACCG
GAAGGAATTCTTTTAGATACTCTAGCTAATGCTATGCCTAAAGATAAAGGCGAAATTCGT
AATAATTGGGAAGGATCTTCAGATAAGTACGCAAGAATGATAGGTGGTTGGCTGGATAAA
CTAGGATTAGTAAAGCAAGGAAAAAAAGAATTTATCATTCCTACTTTGGGTAAGCCGGAC
AATAAAGAGTTTATATCCCACGCTTTTAAAATTACTGGAGAAGGTTTGAAAGTACTGCGT
CGAGCAAAAGGCTCTACAAAATTTACACGTGTACCTAAAAGAGTATATTGGGAAATGCTT
GCTACAAACCTAACCGATAAAGAGTATGTAAGAACAAGAAGAGCTTTGATTTTAGAAATA
TTAATCAAAGCTGGATCATTAAAAATAGAACAAATACAAGACAACTTGAAGAAATTAGGA
TTTGATGAAGTTATAGAAACTATTGAAAATGATATCAAAGGCTTAATTAACACAGGTATA
TTTATAGAAATCAAAGGGCGATTTTATCAATTGAAAGACCATATTCTTCAATTTGTAATA
CCTAATCGTGGTGTGACTAAGCAACTAGTCAAAAGTGAACTGGAGGAGAAGAAATCTGAA
CTTCGTCATAAATTGAAATATGTGCCTCATGAATATATTGAATTAATTGAAATTGCCAGA
```

```
AATTCCACTCAGGATAGAATTCTTGAAATGAAGGTAATGGAATTTTTTATGAAAGTTTAT
GGATATAGAGGTAAACATTTGGGTGGATCAAGGAAACCGGACGGAGCAATTTATACTGTC
GGATCTCCTATTGATTACGGTGTGATCGTGGATACTAAAGCTTATAGCGGAGGTTATAAT
CTGCCAATTGGCCAAGCAGATGAAATGCAACGATATGTCGAAGAAAATCAAACACGAAAC
AAACATATCAACCCTAATGAATGGTGGAAAGTCTATCCATCTTCTGTAACGGAATTTAAG
TTTTTATTTGTGAGTGGTCACTTTAAAGGAAACTACAAAGCTCAGCTTACACGATTAAAT
CATATCACTAATTGTAATGGAGCTGTTCTTAGTGTAGAAGAGCTTTTAATTGGTGGAGAA
ATGATTAAAGCCGGCACATTAACCTTAGAGGAAGTGAGACGGAAATTTAATAACGGCGAG
ATAAACTTTTAA (SEQ ID NO:5)
```

In some embodiments, the FokI nuclease used herein is at least about 50% identical SEQ ID NO:4, e.g., to amino acids 388–583 or 408-583 of SEQ ID NO:4. These variant nucleases must retain the ability to cleave DNA. In some embodiments, the nucleotide sequences are about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 99% or 100% identical to amino acids 388–583 or 408-583 of SEQ ID NO:4. In some embodiments, any differences from amino acids 388–583 or 408-583 of SEQ ID NO:4 are in non-conserved regions.

To determine the percent identity of two sequences, the sequences are aligned for optimal comparison purposes (gaps are introduced in one or both of a first and a second amino acid or nucleic acid sequence as required for optimal alignment, and non-homologous sequences can be disregarded for comparison purposes). The length of a reference sequence aligned for comparison purposes is at least 50% (in some embodiments, about 50%, 55%, 60%, 65%, 70%, 75%, 85%, 90%, 95%, or 100% of the length of the reference sequence is aligned). The nucleotides or residues at corresponding positions are then compared. When a position in the first sequence is occupied by the same nucleotide or residue as the corresponding position in the second sequence, then the molecules are identical at that position. The percent identity between the two sequences is a function of the number of identical positions shared by the sequences, taking into account the number of gaps, and the length of each gap, which need to be introduced for optimal alignment of the two sequences.

The comparison of sequences and determination of percent identity between two sequences can be accomplished using a mathematical algorithm. For purposes of the present application, the percent identity between two amino acid sequences is determined using the Needleman and Wunsch ((1970) J. Mol. Biol. 48:444-453) algorithm which has been incorporated into the GAP program in the GCG software package, using a Blossum 62 scoring matrix with a gap penalty of 12, a gap extend penalty of 4, and a frameshift gap penalty of 5.

60412-4906

Cas9

A number of bacteria express Cas9 protein variants. The Cas9 from *Streptococcus pyogenes* is presently the most commonly used; some of the other Cas9 proteins have high levels of sequence identity with the *S. pyogenes* Cas9 and use the same guide RNAs. Others are more diverse, use different gRNAs, and recognize different PAM sequences as well (the 2-5 nucleotide sequence specified by the protein which is adjacent to the sequence specified by the RNA). Chylinski et al. classified Cas9 proteins from a large group of bacteria (RNA Biology 10:5, 1–12; 2013), and a large number of Cas9 proteins are listed in supplementary figure 1 and supplementary table 1 thereof. Additional Cas9 proteins are described in Esvelt et al., Nat Methods. 2013 Nov; 10(11):1116-21 and Fonfara et al., "Phylogeny of Cas9 determines functional exchangeability of dual-RNA and Cas9 among orthologous type II CRISPR-Cas systems." Nucleic Acids Res. 2013 Nov 22. [Epub ahead of print] doi:10.1093/nar/gkt1074.

Cas9 molecules of a variety of species can be used in the methods and compositions described herein. While the *S. pyogenes* and *S. thermophilus* Cas9 molecules are the subject of much of the disclosure herein, Cas9 molecules of, derived from, or based on the Cas9 proteins of other species listed herein can be used as well. In other words, while the much of the description herein uses *S. pyogenes* and *S. thermophilus* Cas9 molecules, Cas9 molecules from the other species can replace them. Such species include those set forth in the following table, which was created based on supplementary figure 1 of Chylinski et al., 2013.

Alternative Cas9 proteins	
GenBank Acc No.	**Bacterium**
303229466	*Veillonella atypica ACS-134-V-Col7a*
34762592	*Fusobacterium nucleatum subsp. vincentii*
374307738	*Filifactor alocis ATCC 35896*
320528778	*Solobacterium moorei F0204*
291520705	*Coprococcus catus GD-7*
42525843	*Treponema denticola ATCC 35405*
304438954	*Peptoniphilus duerdenii ATCC BAA-1640*
224543312	*Catenibacterium mitsuokai DSM 15897*
24379809	*Streptococcus mutans UA159*
15675041	*Streptococcus pyogenes SF370*
16801805	*Listeria innocua Clip11262*
116628213	*Streptococcus thermophilus LMD-9*
323463801	*Staphylococcus pseudintermedius ED99*
352684361	*Acidaminococcus intestini RyC-MR95*

Alternative Cas9 proteins	
GenBank Acc No.	**Bacterium**
302336020	*Olsenella uli DSM 7084*
366983953	*Oenococcus kitaharae DSM 17330*
310286728	*Bifidobacterium bifidum S17*
258509199	*Lactobacillus rhamnosus GG*
300361537	*Lactobacillus gasseri JV-V03*
169823755	*Finegoldia magna ATCC 29328*
47458868	*Mycoplasma mobile 163K*
284931710	*Mycoplasma gallisepticum str. F*
363542550	*Mycoplasma ovipneumoniae SC01*
384393286	*Mycoplasma canis PG 14*
71894592	*Mycoplasma synoviae 53*
238924075	*Eubacterium rectale ATCC 33656*
116627542	*Streptococcus thermophilus LMD-9*
315149830	*Enterococcus faecalis TX0012*
315659848	*Staphylococcus lugdunensis M23590*
160915782	*Eubacterium dolichum DSM 3991*
336393381	*Lactobacillus coryniformis subsp. torquens*
310780384	*Ilyobacter polytropus DSM 2926*
325677756	*Ruminococcus albus 8*
187736489	*Akkermansia muciniphila ATCC BAA-835*
117929158	*Acidothermus cellulolyticus 11B*
189440764	*Bifidobacterium longum DJO10A*
283456135	*Bifidobacterium dentium Bd1*
38232678	*Corynebacterium diphtheriae NCTC 13129*
187250660	*Elusimicrobium minutum Pei191*
319957206	*Nitratifractor salsuginis DSM 16511*
325972003	*Sphaerochaeta globus str. Buddy*
261414553	*Fibrobacter succinogenes subsp. succinogenes*
60683389	*Bacteroides fragilis NCTC 9343*
256819408	*Capnocytophaga ochracea DSM 7271*
90425961	*Rhodopseudomonas palustris BisB18*
373501184	*Prevotella micans F0438*
294674019	*Prevotella ruminicola 23*
365959402	*Flavobacterium columnare ATCC 49512*
312879015	*Aminomonas paucivorans DSM 12260*
83591793	*Rhodospirillum rubrum ATCC 11170*
294086111	*Candidatus Puniceispirillum marinum IMCC1322*
121608211	*Verminephrobacter eiseniae EF01-2*
344171927	*Ralstonia syzygii R24*
159042956	*Dinoroseobacter shibae DFL 12*
288957741	*Azospirillum sp- B510*
92109262	*Nitrobacter hamburgensis X14*
148255343	*Bradyrhizobium sp- BTAi1*
34557790	*Wolinella succinogenes DSM 1740*
218563121	*Campylobacter jejuni subsp. jejuni*
291276265	*Helicobacter mustelae 12198*

Alternative Cas9 proteins	
GenBank Acc No.	**Bacterium**
229113166	*Bacillus cereus Rock1-15*
222109285	*Acidovorax ebreus TPSY*
189485225	*uncultured Termite group 1*
182624245	*Clostridium perfringens D str.*
220930482	*Clostridium cellulolyticum H10*
154250555	*Parvibaculum lavamentivorans DS-1*
257413184	*Roseburia intestinalis L1-82*
218767588	*Neisseria meningitidis Z2491*
15602992	*Pasteurella multocida subsp. multocida*
319941583	*Sutterella wadsworthensis 3 1*
254447899	*gamma proteobacterium HTCC5015*
54296138	*Legionella pneumophila str. Paris*
331001027	*Parasutterella excrementihominis YIT 11859*
34557932	*Wolinella succinogenes DSM 1740*
118497352	*Francisella novicida U112*

The constructs and methods described herein can include the use of any of those Cas9 proteins, and their corresponding guide RNAs or other guide RNAs that are compatible. The Cas9 from *Streptococcus thermophilus* LMD-9 CRISPR1 system has also been shown to function in human cells in Cong et al (Science 339, 819 (2013)). Cas9 orthologs from *N. meningitides* are described in Hou et al., Proc Natl Acad Sci U S A. 2013 Sep 24;110(39):15644-9 and Esvelt et al., Nat Methods. 2013 Nov;10(11):1116-21. Additionally, Jinek et al. showed in vitro that Cas9 orthologs from *S. thermophilus* and *L. innocua*, (but not from *N. meningitidis* or *C. jejuni*, which likely use a different guide RNA), can be guided by a dual *S. pyogenes* gRNA to cleave target plasmid DNA, albeit with slightly decreased efficiency.

In some embodiments, the present system utilizes the Cas9 protein from *S. pyogenes*, either as encoded in bacteria or codon-optimized for expression in mammalian cells, containing mutations at D10, E762, H983, or D986 and H840 or N863, e.g., D10A/D10N and H840A/H840N/H840Y, to render the nuclease portion of the protein catalytically inactive; substitutions at these positions could be alanine (as they are in Nishimasu al., Cell 156, 935–949 (2014)) or they could be other residues, e.g., glutamine, asparagine, tyrosine, serine, or aspartate, e.g.,, E762Q, H983N, H983Y, D986N, N863D, N863S, or N863H (Figure 1C). The sequence of the catalytically inactive *S. pyogenes* Cas9 that can be used in the methods and compositions described herein is as follows; the exemplary mutations of D10A and H840A are in bold and underlined.

10 20 30 40 50 60
MDKKYSIGL**A** IGTNSVGWAV ITDEYKVPSK KFKVLGNTDR HSIKKNLIGA LLFDSGETAE

70 80 90 100 110 120
ATRLKRTARR RYTRRKNRIC YLQEIFSNEM AKVDDSFFHR LEESFLVEED KKHERHPIFG

130 140 150 160 170 180
NIVDEVAYHE KYPTIYHLRK KLVDSTDKAD LRLIYLALAH MIKFRGHFLI EGDLNPDNSD

190 200 210 220 230 240
VDKLFIQLVQ TYNQLFEENP INASGVDAKA ILSARLSKSR RLENLIAQLP GEKKNGLFGN

250 260 270 280 290 300
LIALSLGLTP NFKSNFDLAE DAKLQLSKDT YDDDLDNLLA QIGDQYADLF LAAKNLSDAI

310 320 330 340 350 360
LLSDILRVNT EITKAPLSAS MIKRYDEHHQ DLTLLKALVR QQLPEKYKEI FFDQSKNGYA

370 380 390 400 410 420
GYIDGGASQE EFYKFIKPIL EKMDGTEELL VKLNREDLLR KQRTFDNGSI PHQIHLGELH

430 440 450 460 470 480
AILRRQEDFY PFLKDNREKI EKILTFRIPY YVGPLARGNS RFAWMTRKSE ETITPWNFEE

490 500 510 520 530 540
VVDKGASAQS FIERMTNFDK NLPNEKVLPK HSLLYEYFTV YNELTKVKYV TEGMRKPAFL

550 560 570 580 590 600
SGEQKKAIVD LLFKTNRKVT VKQLKEDYFK KIECFDSVEI SGVEDRFNAS LGTYHDLLKI

610 620 630 640 650 660
IKDKDFLDNE ENEDILEDIV LTLTLFEDRE MIEERLKTYA HLFDDKVMKQ LKRRRYTGWG

670 680 690 700 710 720
RLSRKLINGI RDKQSGKTIL DFLKSDGFAN RNFMQLIHDD SLTFKEDIQK AQVSGQGDSL

730 740 750 760 770 780
HEHIANLAGS PAIKKGILQT VKVVDELVKV MGRHKPENIV IEMARENQTT QKGQKNSRER

790 800 810 820 830 840
MKRIEEGIKE LGSQILKEHP VENTQLQNEK LYLYYLQNGR DMYVDQELDI NRLSDYDVD**A**

850 860 870 880 890 900
IVPQSFLKDD SIDNKVLTRS DKNRGKSDNV PSEEVVKKMK NYWRQLLNAK LITQRKFDNL

910 920 930 940 950 960
TKAERGGLSE LDKAGFIKRQ LVETRQITKH VAQILDSRMN TKYDENDKLI REVKVITLKS

970 980 990 1000 1010 1020
KLVSDFRKDF QFYKVREINN YHHAHDAYLN AVVGTALIKK YPKLESEFVY GDYKVYDVRK

1030 1040 1050 1060 1070 1080
MIAKSEQEIG KATAKYFFYS NIMNFFKTEI TLANGEIRKR PLIETNGETG EIVWDKGRDF

1090 1100 1110 1120 1130 1140
ATVRKVLSMP QVNIVKKTEV QTGGFSKESI LPKRNSDKLI ARKKDWDPKK YGGFDSPTVA

1150 1160 1170 1180 1190 1200
YSVLVVAKVE KGKSKKLKSV KELLGITIME RSSFEKNPID FLEAKGYKEV KKDLIIKLPK

```
      1210       1220       1230       1240       1250       1260
YSLFELENGR KRMLASAGEL QKGNELALPS KYVNFLYLAS HYEKLKGSPE DNEQKQLFVE

      1270       1280       1290       1300       1310       1320
QHKHYLDEII EQISEFSKRV ILADANLDKV LSAYNKHRDK PIREQAENII HLFTLTNLGA

      1330       1340       1350       1360
PAAFKYFDTT IDRKRYTSTK EVLDATLIHQ SITGLYETRI DLSQLGGD  (SEQ ID NO:5)
```

In some embodiments, the Cas9 nuclease used herein is at least about 50% identical to the sequence of *S. pyogenes* Cas9, i.e., at least 50% identical to SEQ ID NO:5. In some embodiments, the nucleotide sequences are about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 99% or 100% identical to SEQ ID NO:5. In some embodiments, any differences from SEQ ID NO:5 are in non-conserved regions, as identified by sequence alignment of sequences set forth in Chylinski et al., RNA Biology 10:5, 1–12; 2013 (e.g., in supplementary figure 1 and supplementary table 1 thereof); Esvelt et al., Nat Methods. 2013 Nov;10(11):1116-21 and Fonfara et al., Nucl. Acids Res. (2014) 42 (4): 2577-2590. [Epub ahead of print 2013 Nov 22] doi:10.1093/nar/gkt1074. Identity is determined as set forth above.

Guide RNAs (gRNAs)

Guide RNAs generally speaking come in two different systems: System 1, which uses separate crRNA and tracrRNAs that function together to guide cleavage by Cas9, and System 2, which uses a chimeric crRNA-tracrRNA hybrid that combines the two separate guide RNAs in a single system (referred to as a single guide RNA or sgRNA, see also Jinek et al., Science 2012; 337:816–821). The tracrRNA can be variably truncated and a range of lengths has been shown to function in both the separate system (system 1) and the chimeric gRNA system (system 2). For example, in some embodiments, tracrRNA may be truncated from its 3' end by at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35 or 40 nts. In some embodiments, the tracrRNA molecule may be truncated from its 5' end by at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35 or 40 nts. Alternatively, the tracrRNA molecule may be truncated from both the 5' and 3' end, e.g., by at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15 or 20 nts on the 5' end and at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35 or 40 nts on the 3' end. See, e.g., Jinek et al., Science 2012; 337:816–821; Mali et al., Science. 2013 Feb 15;339(6121):823-6; Cong et al., Science. 2013 Feb 15;339(6121):819-23; and Hwang and Fu et al., Nat Biotechnol. 2013 Mar;31(3):227-9; Jinek et al., Elife 2, e00471 (2013)). For System 2, generally the longer length chimeric gRNAs have

shown greater on-target activity but the relative specificities of the various length gRNAs currently remain undefined and therefore it may be desirable in certain instances to use shorter gRNAs. In some embodiments, the gRNAs are complementary to a region that is within about 100-800 bp upstream of the transcription start site, e.g., is within about 500 bp upstream of the transcription start site, includes the transcription start site, or within about 100-800 bp, e.g., within about 500 bp, downstream of the transcription start site. In some embodiments, vectors (e.g., plasmids) encoding more than one gRNA are used, e.g., plasmids encoding, 2, 3, 4, 5, or more gRNAs directed to different sites in the same region of the target gene.

Cas9 nuclease can be guided to specific 17-20 nt genomic targets bearing an additional proximal protospacer adjacent motif (PAM), e.g., of sequence NGG, using a guide RNA, e.g., a single gRNA or a tracrRNA/crRNA, bearing 17-20 nts at its 5' end that are complementary to the complementary strand of the genomic DNA target site. Thus, the present methods can include the use of a single guide RNA comprising a crRNA fused to a normally trans-encoded tracrRNA, e.g., a single Cas9 guide RNA as described in Mali et al., Science 2013 Feb 15; 339(6121):823-6, with a sequence at the 5' end that is complementary to the target sequence, e.g., of 25-17, optionally 20 or fewer nucleotides (nts), e.g., 20, 19, 18, or 17 nts, preferably 17 or 18 nts, of the complementary strand to a target sequence immediately 5' of a protospacer adjacent motif (PAM), e.g., NGG, NAG, or NNGG. In some embodiments, the single Cas9 guide RNA consists of the sequence:

(X_{17-20})GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCG(X_N) (SEQ ID NO:6);

(X_{17-20})GUUUUAGAGCUAUGCUGAAAAGCAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUC(X_N) (SEQ ID NO:7);

(X_{17-20})GUUUUAGAGCUAUGCUGUUUUGGAAACAAAACAGCAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUC(X_N) (SEQ ID NO:8);

(X_{17-20})GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC(X_N) (SEQ ID NO:9),

(X_{17-20})GUUUAAGAGCUAGAAAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC(SEQ ID NO:10);

(X_{17-20})GUUUUAGAGCUAUGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC (SEQ ID

NO:11); or
(X_{17-20})GUUUAAGAGCUAUGCUGGAAACAGCAUAGCAAGUUUAAAUAAGG CUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC (SEQ ID NO:12);

5 wherein X_{17-20} is the nucleotide sequence complementary to 17-20 consecutive nucleotides of the target sequence. DNAs encoding the single guide RNAs have been described previously in the literature (Jinek et al., Science. 337(6096):816-21 (2012) and Jinek et al., Elife. 2:e00471 (2013)).

The guide RNAs can include X_N which can be any sequence, wherein N (in
10 the RNA) can be 0-200, e.g., 0-100, 0-50, or 0-20, that does not interfere with the binding of the ribonucleic acid to Cas9.

In some embodiments, the guide RNA includes one or more Adenine (A) or Uracil (U) nucleotides on the 3' end. In some embodiments the RNA includes one or more U, e.g., 1 to 8 or more Us (e.g., U, UU, UUU, UUUU, UUUUU, UUUUUU,
15 UUUUUUU, UUUUUUUU) at the 3' end of the molecule, as a result of the optional presence of one or more Ts used as a termination signal to terminate RNA PolIII transcription.

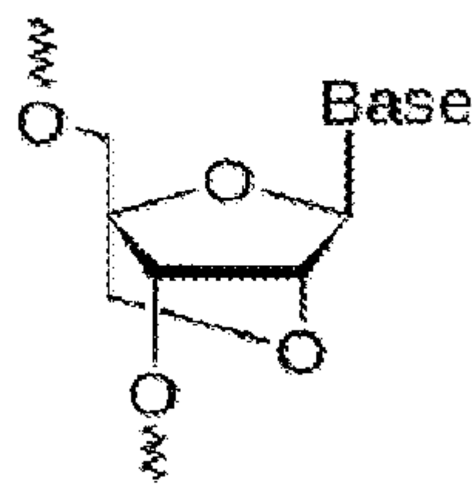
Although some of the examples described herein utilize a single gRNA, the methods can also be used with dual gRNAs (e.g., the crRNA and tracrRNA found in
20 naturally occurring systems). In this case, a single tracrRNA would be used in conjunction with multiple different crRNAs expressed using the present system, e.g., the following:
(X_{17-20})GUUUUAGAGCUA (SEQ ID NO:13);
(X_{17-20})GUUUUAGAGCUAUGCUGUUUUG (SEQ ID NO:14); or
25 (X_{17-20})GUUUUAGAGCUAUGCU (SEQ ID NO:15); and a tracrRNA sequence. In this case, the crRNA is used as the guide RNA in the methods and molecules described herein, and the tracrRNA can be expressed from the same or a different DNA molecule. In some embodiments, the methods include contacting the cell with a tracrRNA comprising or consisting of the sequence
30 GGAACCAUUCAAAACAGCAUAGCAAGUUAAAAUAAGGCUAGUCCGUUA UCAACUUGAAAAAGUGGCACCGAGUCGGUGC (SEQ ID NO:16) or an active portion thereof (an active portion is one that retains the ability to form complexes with Cas9 or dCas9). In some embodiments, the tracrRNA molecule may be truncated

from its 3' end by at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35 or 40 nts. In another embodiment, the tracrRNA molecule may be truncated from its 5' end by at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35 or 40 nts. Alternatively, the tracrRNA molecule may be truncated from both the 5' and 3' end, e.g., by at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15 or 20 nts on the 5' end and at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35 or 40 nts on the 3' end. Exemplary tracrRNA sequences in addition to SEQ ID NO:8 include the following:
UAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC (SEQ ID NO:17) or an active portion thereof; or
AGCAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC (SEQ ID NO:18) or an active portion thereof.

In some embodiments wherein (X_{17-20})GUUUUAGAGCUAUGCUGUUUUG (SEQ ID NO:14) is used as a crRNA, the following tracrRNA is used:
GGAACCAUUCAAAACAGCAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC (SEQ ID NO:16) or an active portion thereof. In some embodiments wherein (X_{17-20})GUUUUAGAGCUA (SEQ ID NO:13) is used as a crRNA, the following tracrRNA is used:
UAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC (SEQ ID NO:17) or an active portion thereof. In some embodiments wherein (X_{17-20})GUUUUAGAGCUAUGCU (SEQ ID NO:15) is used as a crRNA, the following tracrRNA is used:
AGCAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC (SEQ ID NO:18) or an active portion thereof.

In some embodiments, the gRNA is targeted to a site that is at least three or more mismatches different from any sequence in the rest of the genome in order to minimize off-target effects.

Modified RNA oligonucleotides such as locked nucleic acids (LNAs) have been demonstrated to increase the specificity of RNA-DNA hybridization by locking the modified oligonucleotides in a more favorable (stable) conformation. For example, 2'-O-methyl RNA is a modified base where there is an additional covalent linkage between the 2' oxygen and 4' carbon which when incorporated into oligonucleotides can improve overall thermal stability and selectivity (**Formula I**).



Formula I – Locked Nucleic Acid

Thus in some embodiments, the tru-gRNAs disclosed herein may comprise one or more modified RNA oligonucleotides. For example, the truncated guide RNAs molecules described herein can have one, some or all of the 17-18 or 17-19 nts 5' region of the guideRNA complementary to the target sequence are modified, e.g., locked (2'-O-4'-C methylene bridge), 5'-methylcytidine, 2'-O-methyl-pseudouridine, or in which the ribose phosphate backbone has been replaced by a polyamide chain (peptide nucleic acid), e.g., a synthetic ribonucleic acid.

In other embodiments, one, some or all of the nucleotides of the tru-gRNA sequence may be modified, e.g., locked (2'-O-4'-C methylene bridge), 5'-methylcytidine, 2'-O-methyl-pseudouridine, or in which the ribose phosphate backbone has been replaced by a polyamide chain (peptide nucleic acid), e.g., a synthetic ribonucleic acid.

In some embodiments, the single guide RNAs and/or crRNAs and/or tracrRNAs can include one or more Adenine (A) or Uracil (U) nucleotides on the 3' end.

Existing Cas9-based RGNs use gRNA-DNA heteroduplex formation to guide targeting to genomic sites of interest. However, RNA-DNA heteroduplexes can form a more promiscuous range of structures than their DNA-DNA counterparts. In effect, DNA-DNA duplexes are more sensitive to mismatches, suggesting that a DNA-guided nuclease may not bind as readily to off-target sequences, making them comparatively more specific than RNA-guided nucleases. Thus, the guide RNAs usable in the methods described herein can be hybrids, i.e., wherein one or more deoxyribonucleotides, e.g., a short DNA oligonucleotide, replaces all or part of the gRNA, e.g., all or part of the complementarity region of a gRNA. This DNA-based molecule could replace either all or part of the gRNA in a single gRNA system or alternatively might replace all of part of the crRNA and/or tracrRNA in a dual crRNA/tracrRNA system. Such a system that incorporates DNA into the complementarity region should more reliably target the intended genomic DNA

sequences due to the general intolerance of DNA-DNA duplexes to mismatching compared to RNA-DNA duplexes. Methods for making such duplexes are known in the art, See, e.g., Barker et al., BMC Genomics. 2005 Apr 22;6:57; and Sugimoto et al., Biochemistry. 2000 Sep 19;39(37):11270-81.

5 In addition, in a system that uses separate crRNA and tracrRNA, one or both can be synthetic and include one or more modified (e.g., locked) nucleotides or deoxyribonucleotides.

In a cellular context, complexes of Cas9 with these synthetic gRNAs could be used to improve the genome-wide specificity of the CRISPR/Cas9 nuclease system.

10 The methods described can include expressing in a cell, or contacting the cell with, a Cas9 gRNA plus a fusion protein as described herein.

Expression Systems

In order to use the fusion proteins described, it may be desirable to express them from a nucleic acid that encodes them. This can be performed in a variety of 15 ways. For example, the nucleic acid encoding the guide RNA can be cloned into an intermediate vector for transformation into prokaryotic or eukaryotic cells for replication and/or expression. Intermediate vectors are typically prokaryote vectors, e.g., plasmids, or shuttle vectors, or insect vectors, for storage or manipulation of the nucleic acid encoding the fusion proteins for production of the fusion proteins. The 20 nucleic acid encoding the fusion proteins can also be cloned into an expression vector, for administration to a plant cell, animal cell, preferably a mammalian cell or a human cell, fungal cell, bacterial cell, or protozoan cell.

To obtain expression, a sequence encoding a fusion protein is typically subcloned into an expression vector that contains a promoter to direct transcription. 25 Suitable bacterial and eukaryotic promoters are well known in the art and described, e.g., in Sambrook et al., Molecular Cloning, A Laboratory Manual (3d ed. 2001); Kriegler, Gene Transfer and Expression: A Laboratory Manual (1990); and Current Protocols in Molecular Biology (Ausubel et al., eds., 2010). Bacterial expression systems for expressing the engineered protein are available in, e.g., *E. coli*, *Bacillus* 30 sp., and *Salmonella* (Palva et al., 1983, Gene 22:229-235). Kits for such expression systems are commercially available. Eukaryotic expression systems for mammalian cells, yeast, and insect cells are well known in the art and are also commercially available.

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

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

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

Expression vectors containing regulatory elements from eukaryotic viruses are often used in eukaryotic expression vectors, e.g., SV40 vectors, papilloma virus vectors, and vectors derived from Epstein-Barr virus. Other exemplary eukaryotic vectors include pMSG, pAV009/A+, pMTO10/A+, pMAMneo-5, baculovirus pDSVE, and any other vector allowing expression of proteins under the direction of the SV40 early promoter, SV40 late promoter, metallothionein promoter, murine mammary

tumor virus promoter, Rous sarcoma virus promoter, polyhedrin promoter, or other promoters shown effective for expression in eukaryotic cells.

The vectors for expressing the guide RNAs can include RNA Pol III promoters to drive expression of the guide RNAs, e.g., the H1, U6 or 7SK promoters. These human promoters allow for expression of gRNAs in mammalian cells following plasmid transfection. Alternatively, a T7 promoter may be used, e.g., for in vitro transcription, and the RNA can be transcribed in vitro and purified. Vectors suitable for the expression of short RNAs, e.g., siRNAs, shRNAs, or other small RNAs, can be used. With the Cys4-based multiplex system described in Figure 4B, multiple gRNAs can be expressed in a single transcript (driven by a RNA Pol II or Pol III promoter) and then cleaved out from that larger transcript.

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

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

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

Any of the known procedures for introducing foreign nucleotide sequences into host cells may be used. These include the use of calcium phosphate transfection, polybrene, protoplast fusion, electroporation, nucleofection, liposomes, microinjection, naked DNA, plasmid vectors, viral vectors, both episomal and integrative, and any of the other well-known methods for introducing cloned genomic

DNA, cDNA, synthetic DNA or other foreign genetic material into a host cell (see, e.g., Sambrook et al., supra). It is only necessary that the particular genetic engineering procedure used be capable of successfully introducing at least one gene into the host cell capable of expressing the gRNA.

5 The present invention includes the vectors and cells comprising the vectors.

EXAMPLES

The invention is further described in the following examples, which do not limit the scope of the invention described in the claims.

Example 1. Assessing specificity of RNA-guided endonucleases

10 CRISPR RNA-guided nucleases (RGNs) have rapidly emerged as a facile and efficient platform for genome editing. This example describes the use of a human cell-based reporter assay to characterize off-target cleavage of Cas9-based RGNs.

Materials and Methods

The following materials and methods were used in Example 1.

15 **Construction of guide RNAs**

DNA oligonucleotides harboring variable 20 nt sequences for Cas9 targeting were annealed to generate short double-strand DNA fragments with 4 bp overhangs compatible with ligation into BsmBI-digested plasmid pMLM3636. Cloning of these annealed oligonucleotides generates plasmids encoding a chimeric +103 single-chain
20 guide RNA with 20 variable 5' nucleotides under expression of a U6 promoter (Hwang et al., Nat Biotechnol 31, 227-229 (2013); Mali et al., Science 339, 823-826 (2013)). pMLM3636 and the expression plasmid pJDS246 (encoding a codon optimized version of Cas9) used in this study are both available through the non-profit plasmid distribution service Addgene (addgene.org/crispr-cas).

25 **EGFP Activity Assays**

U2OS.EGFP cells harboring a single integrated copy of an *EGFP-PEST* fusion gene were cultured as previously described (Reyon et al., Nat Biotech 30, 460-465 (2012)). For transfections, 200,000 cells were Nucleofected with the indicated amounts of gRNA expression plasmid and pJDS246 together with 30 ng of a Td-
30 tomato-encoding plasmid using the SE Cell Line 4D-Nucleofector™ X Kit (Lonza) according to the manufacturer's protocol. Cells were analyzed 2 days post-transfection using a BD LSRII flow cytometer. Transfections for optimizing

gRNA/Cas9 plasmid concentration were performed in triplicate and all other transfections were performed in duplicate.

PCR amplification and sequence verification of endogenous human genomic sites

5 PCR reactions were performed using Phusion Hot Start II high-fidelity DNA polymerase (NEB). Most loci amplified successfully using touchdown PCR (98 °C, 10 s; 72–62 °C, −1 °C/cycle, 15 s; 72 °C, 30 s]10 cycles, [98 °C, 10 s; 62 °C, 15 s; 72 °C, 30 s]25 cycles). PCR for the remaining targets were performed with 35 cycles at a constant annealing temperature of 68 °C or 72 °C and 3% DMSO or 1M betaine, if
10 necessary. PCR products were analyzed on a QIAXCEL capillary electrophoresis system to verify both size and purity. Validated products were treated with ExoSap-IT (Affymetrix) and sequenced by the Sanger method (MGH DNA Sequencing Core) to verify each target site.

Determination of RGN-induced on- and off-target mutation frequencies
15 **in human cells**

For U2OS.EGFP and K562 cells, 2 x 10^5 cells were transfected with 250 ng of gRNA expression plasmid or an empty U6 promoter plasmid (for negative controls), 750 ng of Cas9 expression plasmid, and 30 ng of td-Tomato expression plasmid using the 4D Nucleofector System according to the manufacturer's instructions (Lonza). For
20 HEK293 cells, 1.65 x 10^5 cells were transfected with 125 ng of gRNA expression plasmid or an empty U6 promoter plasmid (for the negative control), 375 ng of Cas9 expression plasmid, and 30 ng of a td-Tomato expression plasmid using Lipofectamine LTX reagent according to the manufacturer's instructions (Life Technologies). Genomic DNA was harvested from transfected U2OS.EGFP,
25 HEK293, or K562 cells using the QIAamp DNA Blood Mini Kit (QIAGEN), according to the manufacturer's instructions. To generate enough genomic DNA to amplify the off-target candidate sites, DNA from three Nucleofections (for U2OS.EGFP cells), two Nucleofections (for K562 cells), or two Lipofectamine LTX transfections was pooled together before performing T7EI. This was done twice for
30 each condition tested, thereby generating duplicate pools of genomic DNA representing a total of four or six individual transfections. PCR was then performed using these genomic DNAs as templates as described above and purified using

Ampure XP beads (Agencourt) according to the manufacturer's instructions. T7EI assays were performed as previously described (Reyon et al., 2012, *supra*).

DNA sequencing of NHEJ-mediated indel mutations

Purified PCR products used for the T7EI assay were cloned into Zero Blunt TOPO vector (Life Technologies) and plasmid DNAs were isolated using an alkaline lysis miniprep method by the MGH DNA Automation Core. Plasmids were sequenced using an M13 forward primer (5' – GTAAAACGACGGCCAG – 3' (SEQ ID NO:19)) by the Sanger method (MGH DNA Sequencing Core).

Example 1a. Single Nucleotide Mismatches

To begin to define the specificity determinants of RGNs in human cells, a large-scale test was performed to assess the effects of systematically mismatching various positions within multiple gRNA/target DNA interfaces. To do this, a quantitative human cell-based enhanced green fluorescent protein (EGFP) disruption assay previously described (see Methods above and Reyon et al., 2012, supra) that enables rapid quantitation of targeted nuclease activities **(Fig. 2B)** was used. In this assay, the activities of nucleases targeted to a single integrated *EGFP* reporter gene can be quantified by assessing loss of fluorescence signal in human U2OS.EGFP cells caused by inactivating frameshift insertion/deletion (indel) mutations introduced by error prone non-homologous end-joining (NHEJ) repair of nuclease-induced double-stranded breaks (DSBs) **(Fig. 2B)**. For the studies described here, three ~100 nt single gRNAs (sgRNAs) targeted to different sequences within *EGFP* were used, as follows:

EGFP Site 1 GGGCACGGGCAGCTTGCCGGTGG (SEQ ID NO:1)
EGFP Site 2 GATGCCGTTCTTCTGCTTGTCGG (SEQ ID NO:2)
EGFP Site 3 GGTGGTGCAGATGAACTTCAGGG (SEQ ID NO:3)

Each of these sgRNAs can efficiently direct Cas9-mediated disruption of EGFP expression (see **Example 1e and 2a**, and **FIGs. 3E (top) and 3F (top)**).

In initial experiments, the effects of single nucleotide mismatches at 19 of 20 nucleotides in the complementary targeting region of three *EGFP*-targeted sgRNAs were tested. To do this, variant sgRNAs were generated for each of the three target sites harboring Watson-Crick transversion mismatches at positions 1 through 19 (numbered 1 to 20 in the 3' to 5' direction; see **Fig. 1**) and the abilities of these

various sgRNAs to direct Cas9-mediated *EGFP* disruption in human cells tested (variant sgRNAs bearing a substitution at position 20 were not generated because this nucleotide is part of the U6 promoter sequence and therefore must remain a guanine to avoid affecting expression.)

5 For EGFP target site #2, single mismatches in positions 1 – 10 of the gRNA have dramatic effects on associated Cas9 activity (**Fig. 2C**, middle panel), consistent with previous studies that suggest mismatches at the 5' end of gRNAs are better tolerated than those at the 3' end (Jiang et al., Nat Biotechnol 31, 233-239 (2013); Cong et al., Science 339, 819-823 (2013); Jinek et al., Science 337, 816-821 (2012)).
10 However, with EGFP target sites #1 and #3, single mismatches at all but a few positions in the gRNA appear to be well tolerated, even within the 3' end of the sequence. Furthermore, the specific positions that were sensitive to mismatch differed for these two targets (**Fig. 2C**, compare top and bottom panels) – for example, target site #1 was particularly sensitive to a mismatch at position 2 whereas
15 target site #3 was most sensitive to mismatches at positions 1 and 8.

Example 1b. Multiple Mismatches

To test the effects of more than one mismatch at the gRNA/DNA interface, a series of variant sgRNAs bearing double Watson-Crick transversion mismatches in adjacent and separated positions were created and the abilities of these sgRNAs to
20 direct Cas9 nuclease activity were tested in human cells using the EGFP disruption assay. All three target sites generally showed greater sensitivity to double alterations in which one or both mismatches occur within the 3' half of the gRNA targeting region. However, the magnitude of these effects exhibited site-specific variation, with target site #2 showing the greatest sensitivity to these double mismatches and target
25 site #1 generally showing the least. To test the number of adjacent mismatches that can be tolerated, variant sgRNAs were constructed bearing increasing numbers of mismatched positions ranging from positions 19 to 15 in the 5' end of the gRNA targeting region (where single and double mismatches appeared to be better tolerated).

Testing of these increasingly mismatched sgRNAs revealed that for all three
30 target sites, the introduction of three or more adjacent mismatches results in significant loss of RGN activity. A sudden drop off in activity occurred for three different EGFP-targeted gRNAs as one makes progressive mismatches starting from position 19 in the 5' end and adding more mismatches moving toward the 3' end.

Specifically, gRNAs containing mismatches at positions 19 and 19+18 show essentially full activity whereas those with mismatches at positions 19+18+17, 19+18+17+16, and 19+18+17+16+15 show essentially no difference relative to a negative control (Figure 2F). (Note that we did not mismatch position 20 in these variant gRNAs because this position needs to remain as a G because it is part of the U6 promoter that drives expression of the gRNA.)

Additional proof of that shortening gRNA complementarity might lead to RGNs with greater specificities was obtained in the following experiment: for four different EGFP-targeted gRNAs (**Figure 2H**), introduction of a double mismatch at positions 18 and 19 did not significantly impact activity. However, introduction of another double mismatch at positions 10 and 11 then into these gRNAs results in near complete loss of activity. Interestingly introduction of only the 10/11 double mismatches does not generally have as great an impact on activity.

Taken together, these results in human cells confirm that the activities of RGNs can be more sensitive to mismatches in the 3' half of the gRNA targeting sequence. However, the data also clearly reveal that the specificity of RGNs is complex and target site-dependent, with single and double mismatches often well tolerated even when one or more mismatches occur in the 3' half of the gRNA targeting region. Furthermore, these data also suggest that not all mismatches in the 5' half of the gRNA/DNA interface are necessarily well tolerated.

In addition, these results strongly suggest that gRNAs bearing shorter regions of complementarity (specifically ~17 nts) will be more specific in their activities. We note that 17 nts of specificity combined with the 2 nts of specificity conferred by the PAM sequence results in specification of a 19 bp sequence, one of sufficient length to be unique in large complex genomes such as those found in human cells.

Example 1c. Off-Target Mutations

To determine whether off-target mutations for RGNs targeted to endogenous human genes could be identified, six sgRNAs that target three different sites in the *VEGFA* gene, one in the *EMX1* gene, one in the *RNF2* gene, and one in the *FANCF* gene were used. These six sgRNAs efficiently directed Cas9-mediated indels at their respective endogenous loci in human U2OS.EGFP cells as detected by T7 Endonuclease I (**T7EI**) assay (**Methods** above). For each of these six RGNs, we then examined dozens of potential off-target sites (ranging in number from 46 to as many

as 64) for evidence of nuclease-induced NHEJ-mediated indel mutations in U2OS.EGFP cells. The loci assessed included all genomic sites that differ by one or two nucleotides as well as subsets of genomic sites that differ by three to six nucleotides and with a bias toward those that had one or more of these mismatches in the 5' half of the gRNA targeting sequence. Using the T7EI assay, four off-target sites (out of 53 candidate sites examined) for *VEGFA* site 1, twelve (out of 46 examined) for *VEGFA* site 2, seven (out of 64 examined) for *VEGFA* site 3 and one (out of 46 examined) for the *EMX1* site were readily identified. No off-target mutations were detected among the 43 and 50 potential sites examined for the *RNF2* or *FANCF* genes, respectively. The rates of mutation at verified off-target sites were very high, ranging from 5.6% to 125% (mean of 40%) of the rate observed at the intended target site. These *bona fide* off-targets included sequences with mismatches in the 3' end of the target site and with as many as a total of five mismatches, with most off-target sites occurring within protein coding genes. DNA sequencing of a subset of off-target sites provided additional molecular confirmation that indel mutations occur at the expected RGN cleavage site (**Figs. 8A-C**).

Example 1d. Off-Target Mutations in Other Cell Types

Having established that RGNs can induce off-target mutations with high frequencies in U2OS.EGFP cells, it was next sought to determine whether these nucleases would also have these effects in other types of human cells. U2OS.EGFP cells had been chosen for initial experiments because these cells were previously used to evaluate the activities of TALENs[15] but human HEK293 and K562 cells have been more widely used to test the activities of targeted nucleases. Therefore, the activities of the four RGNs targeted to *VEGFA* sites 1, 2, and 3 and the *EMX1* site were also assessed in HEK293 and K562 cells. Each of these four RGNs efficiently induced NHEJ-mediated indel mutations at their intended on-target site in these two additional human cell lines (as assessed by T7EI assay), albeit with somewhat lower mutation frequencies than those observed in U2OS.EGFP cells. Assessment of the 24 off-target sites for these four RGNs originally identified in U2OS.EGFP cells revealed that many were again mutated in HEK293 and K562 cells with frequencies similar to those at their corresponding on-target site. As expected, DNA sequencing of a subset of these off-target sites from HEK293 cells provided additional molecular evidence that alterations are occurring at the expected genomic loci. It is not known for certain

why in HEK293 cells four and in K562 cells eleven of the off-target sites identified in U2OS.EGFP cells did not show detectable mutations. However, many of these off-target sites also showed relatively lower mutation frequencies in U2OS.EGFP cells. Therefore, mutation rates of these sites in HEK293 and K562 cells may be falling below the reliable detection limit of our T7EI assay (~2-5%) because RGNs generally appear to have lower activities in HEK293 and K562 cells compared with U2OS.EGFP cells in our experiments. Taken together, the results in HEK293 and K562 cells provide evidence that the high-frequency off-target mutations we observe with RGNs will be a general phenomenon seen in multiple human cell types.

Example 1e. Titration of gRNA- and Cas9-expressing plasmid amounts used for the EGFP disruption assay

Single guide RNAs (sgRNAs) were generated for three different sequences (EGFP SITES 1-3, shown above) located upstream of *EGFP* nucleotide 502, a position at which the introduction of frameshift mutations via non-homologous end-joining can robustly disrupt expression of EGFP (Maeder, M.L. et al., Mol Cell 31, 294-301 (2008); Reyon, D. et al., Nat Biotech 30, 460-465 (2012)).

For each of the three target sites, a range of gRNA-expressing plasmid amounts (12.5 to 250 ng) was initially transfected together with 750 ng of a plasmid expressing a codon-optimized version of the Cas9 nuclease into our U2OS.EGFP reporter cells bearing a single copy, constitutively expressed *EGFP-PEST* reporter gene. All three RGNs efficiently disrupted EGFP expression at the highest concentration of gRNA plasmid (250 ng) (**Fig. 3E (top)**). However, RGNs for target sites #1 and #3 exhibited equivalent levels of disruption when lower amounts of gRNA-expressing plasmid were transfected whereas RGN activity at target site #2 dropped immediately when the amount of gRNA-expressing plasmid transfected was decreased **(Fig. 3E (top))**.

The amount of Cas9-encoding plasmid (range from 50 ng to 750 ng) transfected into our U2OS.EGFP reporter cells was titrated EGFP disruption assayed. As shown in **Fig. 3F (top)**, target site #1 tolerated a three-fold decrease in the amount of Cas9-encoding plasmid transfected without substantial loss of EGFP disruption activity. However, the activities of RGNs targeting target sites #2 and #3 decreased immediately with a three-fold reduction in the amount of Cas9 plasmid transfected **(Fig. 3F (top))**. Based on these results, 25ng/250ng, 250ng/750ng, and 200ng/750ng

of gRNA-/Cas9-expressing plasmids were used for *EGFP* target sites #1, #2, and #3, respectively, for the experiments described in Examples 1a-1d.

The reasons why some gRNA/Cas9 combinations work better than others in disrupting EGFP expression is not understood, nor is why some of these combinations are more or less sensitive to the amount of plasmids used for transfection. Although it is possible that the range of off-target sites present in the genome for these three sgRNAs might influence each of their activities, no differences were seen in the numbers of genomic sites that differ by one to six bps for each of these particular target sites (**Table 1**) that would account for the differential behavior of the three sgRNAs.

Table 1
Numbers of off-target sites in the human genome for six RGNs targeted to endogenous human genes and three RGNs targeted to the *EGFP* reporter gene

Target Site	Number of mismatches to on-target site						
	0	1	2	3	4	5	6
Target 1 (*VEGFA* Site 1)	1	1	4	32	280	2175	13873
Target 2 (*VEGFA* Site 2)	1	0	2	35	443	3889	17398
Target 3 (*VEGFA* Site 3)	1	1	17	377	6028	13398	35517
Target 4 (*EMX*)	1	0	1	18	276	2309	15731
Target 5 (*RNF2*)	1	0	0	6	116	976	7443
Target 6 (*FANCF*)	1	0	1	18	271	1467	9551
EGFP Target Site #1	0	0	3	10	156	1365	9755
EGFP Target Site #2	0	0	0	11	96	974	7353
EGFP Target Site #3	0	0	1	14	165	1439	10361

Off-target sites for each of the six RGNs targeted to the VEGFA, RNF2, FANCF, and EMX1 genes and the three RGNs targeted to EGFP Target Sites #1, #2 and #3 were identified in human genome sequence build GRCh37. Mismatches were only allowed for the 20 nt region to which the gRNA anneals and not to the PAM sequence.

Example 2: Using pairs of guideRNAs with FokI-dCas9 fusion proteins

Monomeric CRISPR-Cas9 nucleases are widely used for targeted genome editing but can induce unwanted off-target mutations with high frequencies. This example describes new dimeric RNA-guided FokI Nucleases (**RFN**s) that recognize an extended, double-length sequence and that strictly depend on two single guide RNAs (**gRNA**s) for cleavage activity. RFNs can robustly edit DNA sequences in endogenous human genes with high efficiencies. Additionally, a method for expressing gRNAs bearing any 5’ end nucleotide is described, a critical advance that

gives dimeric RFNs a useful targeting range. In direct comparisons, monomeric Cas9 nickases generally induce unwanted indels and unexpected focal point mutations with higher frequencies than RFNs directed by a matched single gRNA. RFNs combine the ease of CRISPR RNA-based targeting with the specificity enhancements of
5 dimerization and provide an important new platform for research and therapeutic applications that require highly precise genome editing.

Materials and Methods

The following materials and methods were used in Example 2.

Single and multiplex gRNA expression plasmids

10 Plasmids encoding single or multiplex gRNAs were assembled in a single-step ligation of annealed target site oligosduplexes (Integrated DNA Technologies) and a constant region oligoduplex (for multiplex gRNAs) with BsmBI-digested Csy4-flanked gRNA backbone (pSQT1313; Addgene).

Multiplex gRNA encoding plasmids were constructed by ligating: 1) annealed
15 oligos encoding the first target site, 2) phosphorylated annealed oligos encoding crRNA, tracrRNA, and Csy4-binding site, and 3) annealed oligos encoding the second target site, into a U6-Csy4site-gRNA plasmid backbone digested with BsmBI Type IIs restriction enzyme. Csy4 RNA binding sites were attached to the 3' and 5' ends of a gRNA sequence and expressed with Cas9 in cells. The Csy4 RNA binding site
20 sequence 'GUUCACUGCCGUAUAGGCAGCUAAGAAA' (SEQ ID NO:20) was fused to the 5' and 3' end of the standard gRNA sequence.

GUUCACUGCCGUAUAGGCAGNNNNNNNNNNNNNNNNNNNNGUUUUAGAGCUA
GAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCA
CCGAGUCGGUGCGUUCACUGCCGUAUAGGCAGNNNNNNNNNNNNNNNNNNNN
25 GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUU
GAAAAAGUGGCACCGAGUCGGUGCGUUCACUGCCGUAUAGGCAG (SEQ ID NO:21)

This sequence is a multiplex gRNA sequence flanked by Csy4 sites (underlined). Functionally, encoding these in multiplex on one transcript should have the same
30 result as encoding them separately. Although all pairs of Csy4-flanked sgRNAs were expressed in a multiplex context in the experiments described herein, the sgRNAs can be encoded in multiplex sgRNAs separated by Csy4 sites encoded on one transcript as well as individual sgRNAs that have an additional Csy4 sequence. In this sequence, the first N20 sequence represents the sequence complementary to one strand of the

target genomic sequence, and the second N20 sequence represents the sequence complementary to the other strand of the target genomic sequence.

A plasmid encoding the Csy4 recognition site containing gRNA was co-transfected with plasmid encoding Cas9 and Csy4 proteins separated by a '2A' peptide linkage. The results showed that gRNAs with Csy4 sites fused to the 5' and 3' ends remained capable of directing Cas9-mediated cleavage in human cells using the U2OS-EGFP disruption assay previously described. Thus, Csy4 RNA binding sites can be attached to 3' end of a gRNA sequence and complexes of these Csy4 site-containing gRNAs with Cas9 remain functional in the cell.

In some experiments, a construct encoding Csy4-T2A-FokI-dCas9 was used. The sequences of the FokI-dCas9 fusions are shown below, and include a GGGGS (SEQ ID NO:23) linker (underlined) between the FokI and dCas9 and a nuclear localization sequence.

FokI-dCas9 amino acid sequence (FokI-G4S-dCas9-nls-3XFLAG)

```
MQLVKSELEEKKSELRHKLKYVPHEYIELIEIARNSTQDRILEMKVMEFFMKVYGYR
GKHLGGSRKPDGAIYTVGSPIDYGVIVDTKAYSGGYNLPIGQADEMQRYVEENQTRN
KHINPNEWWKVYPSSVTEFKFLFVSGHFKGNYKAQLTRLNHITNCNGAVLSVEELLI
GGEMIKAGTLTLEEVRRKFNNGEINFGGGGSDKKYSIGLAIGTNSVGWAVITDEYKV
PSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQE
IFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKK
LVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE
NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSN
FDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEI
TKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQ
EEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQ
EDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVD
KGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFL
SGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDL
LKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRR
RYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKA
QVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQ
TTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVD
QELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNY
WRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRM
NTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTA
LIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLA
NGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESI
LPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGIT
IMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGN
ELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVI
LADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTS
```

TKEVLDATLIHQSITGLYETRIDLSQLGGDGSPKKKRKVSSDYKDHDGDYKDHDIDY
KDDDDK (SEQ ID NO:24)

FokI-dCas9 nucleotide sequence (FokI-G4S-dCas9-nls-3XFLAG)

ATGCAACTAGTCAAAAGTGAACTGGAGGAGAAGAAATCTGAACTTCGTCATAAATTG
AAATATGTGCCTCATGAATATATTGAATTAATTGAAATTGCCAGAAATTCCACTCAG
GATAGAATTCTTGAAATGAAGGTAATGGAATTTTTTATGAAAGTTTATGGATATAGA
GGTAAACATTTGGGTGGATCAAGGAAACCGGACGGAGCAATTTATACTGTCGGATCT
CCTATTGATTACGGTGTGATCGTGGATACTAAAGCTTATAGCGGAGGTTATAATCTG
CCAATTGGCCAAGCAGATGAAATGCAACGATATGTCGAAGAAAATCAAACACGAAAC
AAACATATCAACCCTAATGAATGGTGGAAAGTCTATCCATCTTCTGTAACGGAATTT
AAGTTTTTATTTGTGAGTGGTCACTTTAAAGGAAACTACAAAGCTCAGCTTACACGA
TTAAATCATATCACTAATTGTAATGGAGCTGTTCTTAGTGTAGAAGAGCTTTTAATT
GGTGGAGAAATGATTAAAGCCGGCACATTAACCTTAGAGGAAGTCAGACGGAAATTT
AATAACGGCGAGATAAACTTTGGTGGCGGTGGATCCGATAAAAAGTATTCTATTGGT
TTAGCCATCGGCACTAATTCCGTTGGATGGGCTGTCATAACCGATGAATACAAAGTA
CCTTCAAAGAAATTTAAGGTGTTGGGGAACACAGACCGTCATTCGATTAAAAAGAAT
CTTATCGGTGCCCTCCTATTCGATAGTGGCGAAACGGCAGAGGCGACTCGCCTGAAA
CGAACCGCTCGGAGAAGGTATACACGTCGCAAGAACCGAATATGTTACTTACAAGAA
ATTTTTAGCAATGAGATGGCCAAAGTTGACGATTCTTTCTTTCACCGTTTGGAAGAG
TCCTTCCTTGTCGAAGAGGACAAGAAACATGAACGGCACCCCATCTTTGGAAACATA
GTAGATGAGGTGGCATATCATGAAAAGTACCCAACGATTTATCACCTCAGAAAAAAG
CTAGTTGACTCAACTGATAAAGCGGACCTGAGGTTAATCTACTTGGCTCTTGCCCAT
ATGATAAAGTTCCGTGGGCACTTTCTCATTGAGGGTGATCTAAATCCGGACAACTCG
GATGTCGACAAACTGTTCATCCAGTTAGTACAAACCTATAATCAGTTGTTTGAAGAG
AACCCTATAAATGCAAGTGGCGTGGATGCGAAGGCTATTCTTAGCGCCCGCCTCTCT
AAATCCCGACGGCTAGAAAACCTGATCGCACAATTACCCGGAGAGAAGAAAAATGGG
TTGTTCGGTAACCTTATAGCGCTCTCACTAGGCCTGACACCAAATTTTAAGTCGAAC
TTCGACTTAGCTGAAGATGCCAAATTGCAGCTTAGTAAGGACACGTACGATGACGAT
CTCGACAATCTACTGGCACAAATTGGAGATCAGTATGCGGACTTATTTTTGGCTGCC
AAAAACCTTAGCGATGCAATCCTCCTATCTGACATACTGAGAGTTAATACTGAGATT
ACCAAGGCGCCGTTATCCGCTTCAATGATCAAAAGGTACGATGAACATCACCAAGAC
TTGACACTTCTCAAGGCCCTAGTCCGTCAGCAACTGCCTGAGAAATATAAGGAAATA
TTCTTTGATCAGTCGAAAAACGGGTACGCAGGTTATATTGACGGCGGAGCGAGTCAA
GAGGAATTCTACAAGTTTATCAAACCCATATTAGAGAAGATGGATGGGACGGAAGAG
TTGCTTGTAAAACTCAATCGCGAAGATCTACTGCGAAAGCAGCGGACTTTCGACAAC
GGTAGCATTCCACATCAAATCCACTTAGGCGAATTGCATGCTATACTTAGAAGGCAG
GAGGATTTTTATCCGTTCCTCAAAGACAATCGTGAAAAGATTGAGAAAATCCTAACC
TTTCGCATACCTTACTATGTGGGACCCCTGGCCCGAGGGAACTCTCGGTTCGCATGG
ATGACAAGAAAGTCCGAAGAAACGATTACTCCATGGAATTTTGAGGAAGTTGTCGAT
AAAGGTGCGTCAGCTCAATCGTTCATCGAGAGGATGACCAACTTTGACAAGAATTTA
CCGAACGAAAAAGTATTGCCTAAGCACAGTTTACTTTACGAGTATTTCACAGTGTAC
AATGAACTCACGAAAGTTAAGTATGTCACTGAGGGCATGCGTAAACCCGCCTTTCTA
AGCGGAGAACAGAAGAAAGCAATAGTAGATCTGTTATTCAAGACCAACCGCAAAGTG
ACAGTTAAGCAATTGAAAGAGGACTACTTTAAGAAAATTGAATGCTTCGATTCTGTC
GAGATCTCCGGGGTAGAAGATCGATTTAATGCGTCACTTGGTACGTATCATGACCTC
CTAAAGATAATTAAAGATAAGGACTTCCTGGATAACGAAGAGAATGAAGATATCTTA
GAAGATATAGTGTTGACTCTTACCCTCTTTGAAGATCGGGAAATGATTGAGGAAAGA
CTAAAAACATACGCTCACCTGTTCGACGATAAGGTTATGAAACAGTTAAAGAGGCGT

CGCTATACGGGCTGGGGACGATTGTCGCGGAAACTTATCAACGGGATAAGAGACAAG
CAAAGTGGTAAAACTATTCTCGATTTTCTAAAGAGCGACGGCTTCGCCAATAGGAAC
TTTATGCAGCTGATCCATGATGACTCTTTAACCTTCAAAGAGGATATACAAAAGGCA
CAGGTTTCCGGACAAGGGGACTCATTGCACGAACATATTGCGAATCTTGCTGGTTCG
CCAGCCATCAAAAAGGGCATACTCCAGACAGTCAAAGTAGTGGATGAGCTAGTTAAG
GTCATGGGACGTCACAAACCGGAAAACATTGTAATCGAGATGGCACGCGAAAATCAA
ACGACTCAGAAGGGGCAAAAAAACAGTCGAGAGCGGATGAAGAGAATAGAAGAGGGT
ATTAAAGAACTGGGCAGCCAGATCTTAAAGGAGCATCCTGTGGAAAATACCCAATTG
CAGAACGAGAAACTTTACCTCTATTACCTACAAAATGGAAGGGACATGTATGTTGAT
CAGGAACTGGACATAAACCGTTTATCTGATTACGACGTCGATGCCATTGTACCCCAA
TCCTTTTTGAAGGACGATTCAATCGACAATAAAGTGCTTACACGCTCGGATAAGAAC
CGAGGGAAAAGTGACAATGTTCCAAGCGAGGAAGTCGTAAAGAAAATGAAGAACTAT
TGGCGGCAGCTCCTAAATGCGAAACTGATAACGCAAAGAAAGTTCGATAACTTAACT
AAAGCTGAGAGGGGTGGCTTGTCTGAACTTGACAAGGCCGGATTTATTAAACGTCAG
CTCGTGGAAACCCGCCAAATCACAAAGCATGTTGCACAGATACTAGATTCCCGAATG
AATACGAAATACGACGAGAACGATAAGCTGATTCGGGAAGTCAAAGTAATCACTTTA
AAGTCAAAATTGGTGTCGGACTTCAGAAAGGATTTTCAATTCTATAAAGTTAGGGAG
ATAAATAACTACCACCATGCGCACGACGCTTATCTTAATGCCGTCGTAGGGACCGCA
CTCATTAAGAAATACCCGAAGCTAGAAAGTGAGTTTGTGTATGGTGATTACAAAGTT
TATGACGTCCGTAAGATGATCGCGAAAAGCGAACAGGAGATAGGCAAGGCTACAGCC
AAATACTTCTTTTATTCTAACATTATGAATTTCTTTAAGACGGAAATCACTCTGGCA
AACGGAGAGATACGCAAACGACCTTTAATTGAAACCAATGGGGAGACAGGTGAAATC
GTATGGGATAAGGGCCGGGACTTCGCGACGGTGAGAAAAGTTTTGTCCATGCCCCAA
GTCAACATAGTAAAGAAAACTGAGGTGCAGACCGGAGGGTTTTCAAAGGAATCGATT
CTTCCAAAAAGGAATAGTGATAAGCTCATCGCTCGTAAAAAGGACTGGGACCCGAAA
AAGTACGGTGGCTTCGATAGCCCTACAGTTGCCTATTCTGTCCTAGTAGTGGCAAAA
GTTGAGAAGGGAAAATCCAAGAAACTGAAGTCAGTCAAAGAATTATTGGGGATAACG
ATTATGGAGCGCTCGTCTTTTGAAAAGAACCCCATCGACTTCCTTGAGGCGAAAGGT
TACAAGGAAGTAAAAAAGGATCTCATAATTAAACTACCAAAGTATAGTCTGTTTGAG
TTAGAAAATGGCCGAAAACGGATGTTGGCTAGCGCCGGAGAGCTTCAAAAGGGGAAC
GAACTCGCACTACCGTCTAAATACGTGAATTTCCTGTATTTAGCGTCCCATTACGAG
AAGTTGAAAGGTTCACCTGAAGATAACGAACAGAAGCAACTTTTTGTTGAGCAGCAC
AAACATTATCTCGACGAAATCATAGAGCAAATTTCGGAATTCAGTAAGAGAGTCATC
CTAGCTGATGCCAATCTGGACAAAGTATTAAGCGCATACAACAAGCACAGGGATAAA
CCCATACGTGAGCAGGCGGAAAATATTATCCATTTGTTTACTCTTACCAACCTCGGC
GCTCCAGCCGCATTCAAGTATTTTGACACAACGATAGATCGCAAACGATACACTTCT
ACCAAGGAGGTGCTAGACGCGACACTGATTCACCAATCCATCACGGGATTATATGAA
ACTCGGATAGATTTGTCACAGCTTGGGGGTGACGGATCCCCCAAGAAGAAGAGGAAA
GTCTCGAGCGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTAC
AAGGATGACGATGACAAGTGA (SEQ ID NO:25)

Alternatively, a human codon optimized version of the construct was used, which contained both N- and C-terminal nuclear localization signals, as shown below.

Nls-FokI-dCas9-nls amino acid sequence

MPKKKRKVSSQLVKSELEEKKSELRHKLKYVPHEYIELIEIARNSTQDRILEMKVME
FFMKVYGYRGKHLGGSRKPDGAIYTVGSPIDYGVIVDTKAYSGGYNLPIGQADEMQR
YVEENQTRNKHINPNEWWKVYPSSVTEFKFLFVSGHFKGNYKAQLTRLNHITNCNGA

VLSVEELLIGGEMIKAGTLTLEEVRRKFNNGEINFGGGGSDKKYSIGLAIGTNSVGW
AVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRR
KNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKY
PTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV
QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSL
GLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS
DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA
GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLG
ELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETIT
PWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVT
EGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFN
ASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDD
KVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSL
TFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENI
VIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYL
QNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSE
EVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH
VAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDA
YLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMN
FFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQ
TGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLK
SVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLA
SAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQ
ISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDT
TIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGSPKKKRKVSSDYKDHDG
DYKDHDIDYKDDDDK (SEQ ID NO:26)

Nls-FokI—dCas9-nls nucleotide sequence

ATGCCTAAGAAGAAGCGGAAGGTGAGCAGCCAACTTGTGAAGTCTGAACTCGAGGAG
AAAAAATCAGAGTTGAGACACAAGTTGAAGTACGTGCCACACGAATACATCGAGCTT
ATCGAGATCGCCAGAAACAGTACCCAGGATAGGATCCTTGAGATGAAAGTCATGGAG
TTCTTTATGAAGGTCTACGGTTATAGAGGAAAGCACCTTGGCGGTAGCAGAAAGCCC
GATGGCGCCATCTATACTGTCGGATCTCCTATCGATTATGGGGTGATCGTGGATACC
AAAGCTTACTCAGGCGGGTACAACTTGCCCATAGGACAAGCCGACGAGATGCAGCGG
TATGTCGAAGAGAACCAGACGCGCAACAAGCACATCAACCCCAATGAATGGTGGAAA
GTGTACCCAAGTAGTGTGACTGAGTTCAAGTTCCTGTTTGTCTCCGGCCACTTTAAG
GGCAATTATAAAGCTCAGCTCACTAGACTCAATCACATCACAAACTGCAACGGAGCT
GTGTTGTCAGTGGAGGAGCTCCTGATTGGAGGCGAGATGATCAAAGCCGGCACCCTT
ACACTGGAGGAGGTGCGGCGGAAGTTCAACAATGGAGAGATCAACTTCGGTGGCGGT
GGATCCGATAAAAAGTATTCTATTGGTTTAGCCATCGGCACTAATTCCGTTGGATGG
GCTGTCATAACCGATGAATACAAAGTACCTTCAAAGAAATTTAAGGTGTTGGGGAAC
ACAGACCGTCATTCGATTAAAAAGAATCTTATCGGTGCCCTCCTATTCGATAGTGGC
GAAACGGCAGAGGCGACTCGCCTGAAACGAACCGCTCGGAGAAGGTATACACGTCGC
AAGAACCGAATATGTTACTTACAAGAAATTTTTAGCAATGAGATGGCCAAAGTTGAC
GATTCTTTCTTTCACCGTTTGGAAGAGTCCTTCCTTGTCGAAGAGGACAAGAAACAT
GAACGGCACCCCATCTTTGGAAACATAGTAGATGAGGTGGCATATCATGAAAAGTAC
CCAACGATTTATCACCTCAGAAAAAAGCTAGTTGACTCAACTGATAAAGCGGACCTG
AGGTTAATCTACTTGGCTCTTGCCCATATGATAAAGTTCCGTGGGCACTTTCTCATT
GAGGGTGATCTAAATCCGGACAACTCGGATGTCGACAAACTGTTCATCCAGTTAGTA
CAAACCTATAATCAGTTGTTTGAAGAGAACCCTATAAATGCAAGTGGCGTGGATGCG
AAGGCTATTCTTAGCGCCCGCCTCTCTAAATCCCGACGGCTAGAAAACCTGATCGCA
CAATTACCCGGAGAGAAGAAAAATGGGTTGTTCGGTAACCTTATAGCGCTCTCACTA
GGCCTGACACCAAATTTTAAGTCGAACTTCGACTTAGCTGAAGATGCCAAATTGCAG

CTTAGTAAGGACACGTACGATGACGATCTCGACAATCTACTGGCACAAATTGGAGAT
CAGTATGCGGACTTATTTTTGGCTGCCAAAAACCTTAGCGATGCAATCCTCCTATCT
GACATACTGAGAGTTAATACTGAGATTACCAAGGCGCCGTTATCCGCTTCAATGATC
AAAAGGTACGATGAACATCACCAAGACTTGACACTTCTCAAGGCCCTAGTCCGTCAG
5 CAACTGCCTGAGAAATATAAGGAAATATTCTTTGATCAGTCGAAAAACGGGTACGCA
GGTTATATTGACGGCGGAGCGAGTCAAGAGGAATTCTACAAGTTTATCAAACCCATA
TTAGAGAAGATGGATGGGACGGAAGAGTTGCTTGTAAAACTCAATCGCGAAGATCTA
CTGCGAAAGCAGCGGACTTTCGACAACGGTAGCATTCCACATCAAATCCACTTAGGC
GAATTGCATGCTATACTTAGAAGGCAGGAGGATTTTTATCCGTTCCTCAAAGACAAT
10 CGTGAAAAGATTGAGAAAATCCTAACCTTTCGCATACCTTACTATGTGGGACCCCTG
GCCCGAGGGAACTCTCGGTTCGCATGGATGACAAGAAAGTCCGAAGAAACGATTACT
CCATGGAATTTTGAGGAAGTTGTCGATAAAGGTGCGTCAGCTCAATCGTTCATCGAG
AGGATGACCAACTTTGACAAGAATTTACCGAACGAAAAAGTATTGCCTAAGCACAGT
TTACTTTACGAGTATTTCACAGTGTACAATGAACTCACGAAAGTTAAGTATGTCACT
15 GAGGGCATGCGTAAACCCGCCTTTCTAAGCGGAGAACAGAAGAAAGCAATAGTAGAT
CTGTTATTCAAGACCAACCGCAAAGTGACAGTTAAGCAATTGAAAGAGGACTACTTT
AAGAAAATTGAATGCTTCGATTCTGTCGAGATCTCCGGGGTAGAAGATCGATTTAAT
GCGTCACTTGGTACGTATCATGACCTCCTAAAGATAATTAAAGATAAGGACTTCCTG
GATAACGAAGAGAATGAAGATATCTTAGAAGATATAGTGTTGACTCTTACCCTCTTT
20 GAAGATCGGGAAATGATTGAGGAAAGACTAAAAACATACGCTCACCTGTTCGACGAT
AAGGTTATGAAACAGTTAAAGAGGCGTCGCTATACGGGCTGGGGACGATTGTCGCGG
AAACTTATCAACGGGATAAGAGACAAGCAAAGTGGTAAAACTATTCTCGATTTTCTA
AAGAGCGACGGCTTCGCCAATAGGAACTTTATGCAGCTGATCCATGATGACTCTTTA
ACCTTCAAAGAGGATATACAAAAGGCACAGGTTTCCGGACAAGGGGACTCATTGCAC
25 GAACATATTGCGAATCTTGCTGGTTCGCCAGCCATCAAAAAGGGCATACTCCAGACA
GTCAAAGTAGTGGATGAGCTAGTTAAGGTCATGGGACGTCACAAACCGGAAAACATT
GTAATCGAGATGGCACGCGAAAATCAAACGACTCAGAAGGGGCAAAAAAACAGTCGA
GAGCGGATGAAGAGAATAGAAGAGGGTATTAAAGAACTGGGCAGCCAGATCTTAAAG
GAGCATCCTGTGGAAAATACCCAATTGCAGAACGAGAAACTTTACCTCTATTACCTA
30 CAAAATGGAAGGGACATGTATGTTGATCAGGAACTGGACATAAACCGTTTATCTGAT
TACGACGTCGATGCCATTGTACCCCAATCCTTTTTGAAGGACGATTCAATCGACAAT
AAAGTGCTTACACGCTCGGATAAGAACCGAGGGAAAAGTGACAATGTTCCAAGCGAG
GAAGTCGTAAAGAAAATGAAGAACTATTGGCGGCAGCTCCTAAATGCGAAACTGATA
ACGCAAAGAAAGTTCGATAACTTAACTAAAGCTGAGAGGGGTGGCTTGTCTGAACTT
35 GACAAGGCCGGATTTATTAAACGTCAGCTCGTGGAAACCCGCCAAATCACAAAGCAT
GTTGCACAGATACTAGATTCCCGAATGAATACGAAATACGACGAGAACGATAAGCTG
ATTCGGGAAGTCAAAGTAATCACTTTAAAGTCAAAATTGGTGTCGGACTTCAGAAAG
GATTTTCAATTCTATAAAGTTAGGGAGATAAATAACTACCACCATGCGCACGACGCT
TATCTTAATGCCGTCGTAGGGACCGCACTCATTAAGAAATACCCGAAGCTAGAAAGT
40 GAGTTTGTGTATGGTGATTACAAAGTTTATGACGTCCGTAAGATGATCGCGAAAAGC
GAACAGGAGATAGGCAAGGCTACAGCCAAATACTTCTTTTATTCTAACATTATGAAT
TTCTTTAAGACGGAAATCACTCTGGCAAACGGAGAGATACGCAAACGACCTTTAATT
GAAACCAATGGGGAGACAGGTGAAATCGTATGGGATAAGGGCCGGGACTTCGCGACG
GTGAGAAAAGTTTTGTCCATGCCCCAAGTCAACATAGTAAAGAAAACTGAGGTGCAG
45 ACCGGAGGGTTTTCAAAGGAATCGATTCTTCCAAAAAGGAATAGTGATAAGCTCATC
GCTCGTAAAAAGGACTGGGACCCGAAAAAGTACGGTGGCTTCGATAGCCCTACAGTT
GCCTATTCTGTCCTAGTAGTGGCAAAAGTTGAGAAGGGAAAATCCAAGAAACTGAAG
TCAGTCAAAGAATTATTGGGGATAACGATTATGGAGCGCTCGTCTTTTGAAAAGAAC
CCCATCGACTTCCTTGAGGCGAAAGGTTACAAGGAAGTAAAAAAGGATCTCATAATT
50 AAACTACCAAAGTATAGTCTGTTTGAGTTAGAAAATGGCCGAAAACGGATGTTGGCT

AGCGCCGGAGAGCTTCAAAAGGGGAACGAACTCGCACTACCGTCTAAATACGTGAAT
TTCCTGTATTTAGCGTCCCATTACGAGAAGTTGAAAGGTTCACCTGAAGATAACGAA
CAGAAGCAACTTTTTGTTGAGCAGCACAAACATTATCTCGACGAAATCATAGAGCAA
ATTTCGGAATTCAGTAAGAGAGTCATCCTAGCTGATGCCAATCTGGACAAAGTATTA
AGCGCATACAACAAGCACAGGGATAAACCCATACGTGAGCAGGCGGAAAATATTATC
CATTTGTTTACTCTTACCAACCTCGGCGCTCCAGCCGCATTCAAGTATTTTGACACA
ACGATAGATCGCAAACGATACACTTCTACCAAGGAGGTGCTAGACGCGACACTGATT
CACCAATCCATCACGGGATTATATGAAACTCGGATAGATTTGTCACAGCTTGGGGGT
GACGGATCCCCCAAGAAGAAGAGGAAAGTCTCGAGCGACTACAAAGACCATGACGGT
GATTATAAAGATCATGACATCGATTACAAGGATGACGATGACAAGTGA (SEQ ID NO:27)

Tissue culture and transfections

All cell culture experiments were carried out in HEK 293 cells, U2OS cells, or in U2OS cells harboring a stably integrated, single-copy, destabilized EGFP gene (U2OS.EGFP cells). Cell lines were cultured in Advanced DMEM (Life Technologies) supplemented with 10% FBS, 2 mM GlutaMax (Life Technologies) and penicillin/streptomycin at 37C with 5% CO2. Additionally, U2OS.EGFP cells were cultured in the presence of 400 μg/ml of G418.

U2OS cells and U2OS.EGFP cells were transfected using the DN-100 program of a Lonza 4D-Nucleofector according to the manufacturer's instructions. In initial FokI-dCas9 activity screens and focused spacer length analysis experiments, 750 ng of pCAG-Csy4-FokI-dCas9-nls nuclease plasmid and 250 ng of gRNA encoding plasmids were transfected together with 50 ng tdTomato expression plasmid (Clontech) as a transfection control. In all other experiments in U2OS and U2OS.EGFP cells, 975 ng of human codon optimized pCAG-Csy4-T2A-nls-hFokI-dCas9-nls (SQT1601) or pCAG-Cas9-D10A nickase (NW3) were transfected along with 325 ng of gRNA vector and 10 ng of Td tomato expression plasmid and analyzed 3 days after transfection. HEK293 cells were transfected with 750 ng of nuclease plasmid, 250 ng of gRNA expression plasmid and 10 ng of Td tomato, using Lipofectamine (Life Technologies) according to the manufacturer's instructions and analyzed for NHEJ-mediated mutagenesis 3 days after transfection.

Single transfections were performed for the initial spacer activity screen, and duplicate transfections for the focused spacer length analysis. All other transfections were performed in triplicate.

EGFP disruption assay

The EGFP disruption assay was performed as previously described (see Example 1 and Reyon et al., Nat Biotech 30, 460-465 (2012)) using U2OS.EGFP reporter cells. Cells were assayed for EGFP and tdTomato expression using an BD Biosciences LSR II or Fortessa FACS analyzer.

Quantification of nuclease- or nickase-induced mutation rates by T7EI assay

T7E1 assays were performed as previously described (Reyon et al., Nat Biotech 30, 460-465 (2012)). Briefly, genomic DNA was isolated 72 hours post transfection using the Agencourt DNAdvance Genomic DNA Isolation kit (Beckman Coulter Genomics) according to the manufacturer's instructions with a Sciclone G3 liquid-handling workstation (Caliper). PCR reactions to amplify genomic loci were performed using Phusion Hot-start Flex DNA polymerase (New England Biolabs). Samples were amplified using a two-step protocol (98 °C, 30 sec; (98 °C, 7 sec; 72 °C, 30 sec) x 35; 72 °C, 5 min) or a touchdown PCR protocol ((98 °C, 10 s; 72–62 °C, −1 °C/cycle, 15 s; 72 °C, 30 s) × 10 cycles, (98 °C, 10 s; 62 °C, 15 s; 72 °C, 30 s) × 25 cycles). 200 ng of purified PCR amplicons were denatured, hybridized, and treated with T7 Endonuclease I (New England Biolabs). Mutation frequency was quantified using a Qiaxcel capillary electrophoresis instrument (Qiagen) as previously described (Reyon et al., Nat Biotech 30, 460-465 (2012)).

Sanger sequencing of mutagenized genomic DNA

The same purified PCR products used for T7EI assay were Topo-cloned (Life Technologies) and plasmid DNA of individual clones was isolated and sequenced using an M13 reverse primer (5′-GTAAAACGACGGCCAG-3′; SEQ ID NO:19).

Illumina Library Preparation and Analysis

Short 200-350 bp PCR products were amplified using Phusion Hot-start FLEX DNA polymerase. PCR products were purified using Ampure XP beads (Beckman Coulter Genomics) according to manufacturer's instructions. Dual-indexed TruSeq Illumina deep sequencing libraries were prepared using a high-throughput library preparation system (Kapa Biosystems) on a Sciclone G3 liquid-handling workstation. Final adapter-ligated libraries were quantified using a Qiaxcel capillary

electrophoresis instrument (Qiagen). 150 bp paired end sequencing was performed on an Illumina MiSeq Sequencer by the Dana-Farber Cancer Institute Molecular Biology Core.

MiSeq paired-end reads were mapped to human genome reference GChr37 using bwa. Reads with an average quality score >30 were analyzed for insertion or deletion mutations that overlapped the intended target or candidate off-target nuclease binding site. Mutation analyses were conducted using the Genome Analysis Toolkit (GATK) and Python.

Off-target search algorithm:

A target-site matching algorithm was implemented that looks for matches with less than a specified number of mismatches in a sliding window across the human genome.

Example 2a. Rationale for designing dimeric RNA-guided nucleases

It was hypothesized that a single platform combining the specificity advantages of dimerization with the ease of Cas9 targeting could be developed. To do this, the well-characterized, dimerization-dependent FokI nuclease domain was fused to a RNA-guided catalytically inactive Cas9 (dCas9) protein. It was hoped that, like FokI-containing ZFNs and TALENs, dimers of these fusions might mediate sequence-specific DNA cleavage when bound to target sites composed of two "half-sites" with a certain length "spacer" sequence between them (**Fig. 4A**). Such fusions were hypothesized to have enhanced specificity because they should require two gRNAs for activity (**Fig. 4A**) and because a single gRNA would presumably be too inefficient or unable to recruit the two FokI-containing fusion proteins required for DNA cleavage. It was hypothesized that such a dimeric system would show improved specificity relative to standard monomeric Cas9 nucleases and also would potentially possess important specificity advantages over the paired nickase system in which single nickases can still exert unwanted mutagenic effects.

Example 2b. Multiplex expression of gRNAs without 5'-end nucleotide limitations

The targeting range for a dimeric RNA-guided nuclease would be low using existing gRNA expression methods. Two sequence requirements typically restrict the targeting range of a dCas9 monomer: the requirement for a PAM sequence of 5'-

NGG that is specified by the dCas9 and a requirement for a G nucleotide at the 5' end of the gRNA imposed by the use of a U6 promoter in most expression vectors. If, however, the requirement for the 5' G in the gRNA could be relieved, then the targeting range would improve by 16-fold.

5 To develop a multiplex system that would allow for the expression of gRNAs with any 5' nucleotide, a plasmid was constructed from which two gRNAs, each flanked by cleavage sites for the Csy4 ribonuclease (Haurwitz et al., Science 329, 1355-1358 (2010)), can be expressed within a single RNA transcribed from a U6 promoter (**Fig. 4B**). Csy4 would be expected to process this transcript thereby
10 releasing the two gRNAs. Based on the known mechanism of Csy4-mediated cleavage ((Haurwitz et al., Science 329, 1355-1358 (2010); Sternberg et al., RNA 18, 661-672 (2012)), each processed gRNA should retain a Csy4 recognition site on its 3' end with a Csy4 protein bound to that site (**Fig. 4B**). In this configuration, it should be possible to express gRNAs with any 5' nucleotide. This system was tested by
15 using it to express two gRNAs targeted to sites within the *EGFP* reporter gene. Co-expression of this transcript together with Csy4 and Cas9 nucleases in human cells led to the introduction of indel mutations at both *EGFP* target sites as well as of deletion of the sequence between these sites (**Fig. 4C**). These experiments suggest that both gRNAs are being processed from the single parental RNA transcript and both are
20 capable of directing Cas9 nuclease activities in human cells.

Example 2c. Construction and optimization of dimeric RNA-guided nucleases

Two different hybrid proteins harboring the FokI nuclease domain and the dCas9 protein were constructed: one in which the FokI nuclease domain is fused to the carboxy-terminus of dCas9 (**dCas9-FokI**) and the other in which it is fused to the
25 amino-terminus (**FokI-dCas9**) (**Fig. 5A**). The dCas9-FokI protein is analogous in architecture to ZFNs and TALENs (**Fig. 5A**). To ascertain whether either or both of these fusions could mediate site-specific cleavage of DNA, a well-established human cell-based assay that can rapidly and easily quantify the introduction of NHEJ-mediated indels into an *EGFP* reporter gene was used (the EGFP disruption assay
30 described above in Example 1). Because the geometry of the half-sites required for efficient cleavage was not known, 60 pairs of gRNAs targeted to various sites in *EGFP* were designed. The two half-sites targeted by each of these gRNA pairs were oriented such that both of their PAM sequences are either directly adjacent to the

spacer sequence (the "PAM in" orientation) or positioned at the outer boundaries of the full-length target site (the "PAM out" orientation) **(Fig. 5B)**. In addition, the spacer sequence was also varied in length from 0 to 31 bps **(Fig. 5B** and **Table 2**).

TABLE 2

FokI-dCas9 EGFP Pair #	Name	Target Start Position (+)	Sequence (+) sites	SEQ ID NO:	Sequence (-) sites	SEQ ID NO:	Edge-to-edge 'spacer' distance	PAM
1	EGFP site 1	74	GAGCTGGACGGCGACGTAAACG G	28.	CGCCGGACACGCTGAACTTGTGG	29.	0	in
2	EGFP site 2	174	CCGGCAAGCTGCCCGTGCCCTG G	30.	GGTCAGGGTGGTCACGAGGGTGG	31.	1	in
3	EGFP site 3	37	CGAGGAGCTGTTCACCGGGGTG G	32.	CCGTCCAGCTCGACCAGGATGGG	33.	2	in
4	EGFP site 4	37	CGAGGAGCTGTTCACCGGGGTG G	34.	GCCGTCCAGCTCGACCAGGATGG	35.	3	in
5	EGFP site 5	174	CCGGCAAGCTGCCCGTGCCCTG G	36.	GTAGGTCAGGGTGGTCACGAGGG	37.	4	in
6	EGFP site 6	34	GGGCGAGGAGCTGTTCACCGGG G	38.	CCGTCCAGCTCGACCAGGATGGG	39.	5	in
7	EGFP site 7	33	AGGGCGAGGAGCTGTTCACCGG G	40.	CCGTCCAGCTCGACCAGGATGGG	41.	6	in
8	EGFP site 8	32	AAGGGCGAGGAGCTGTTCACCG G	42.	CCGTCCAGCTCGACCAGGATGGG	43.	7	in
9	EGFP site 9	32	AAGGGCGAGGAGCTGTTCACCG G	44.	GCCGTCCAGCTCGACCAGGATGG	45.	8	in
10	EGFP site 10	106	CAGCGTGTCCGGCGAGGGCGAG G	46.	CTTCAGGGTCAGCTTGCCGTAGG	47.	9	in
11	EGFP site 11	34	GGGCGAGGAGCTGTTCACCGGG G	48.	CGTCGCCGTCCAGCTCGACCAGG	49.	10	in
12	EGFP site 12	33	AGGGCGAGGAGCTGTTCACCGG G	50.	CGTCGCCGTCCAGCTCGACCAGG	51.	11	in
13	EGFP site 13	32	AAGGGCGAGGAGCTGTTCACCG G	52.	CGTCGCCGTCCAGCTCGACCAGG	53.	12	in
14	EGFP site 14	155	CTGAAGTTCATCTGCACCACCG G	54.	GTGGTCACGAGGGTGGGCCAGGG	55.	13	in

TABLE 2

FokI-dCas9 EGFP Pair #	Name	Target Start Position (+)	Sequence (+) sites	SEQ ID NO:	Sequence (-) sites	SEQ ID NO:	Edge-to-edge 'spacer' distance	PAM
15	EGFP site 15	101	AAGTTCAGCGTGTCCGGCGAGGG	56.	CTTCAGGGTCAGCTTGCCGTAGG	57.	14	in
16	EGFP site 16	100	CAAGTTCAGCGTGTCCGGCGAGG	58.	CTTCAGGGTCAGCTTGCCGTAGG	59.	15	in
17	EGFP site 17	58	GGTGCCCATCCTGGTCGAGCTGG	60.	CGCCGGACACGCTGAACTTGTGG	61.	16	in
18	EGFP site 18	74	GAGCTGGACGGCGACGTAAACGG	62.	GGCATCGCCCTCGCCCTCGCCGG	63.	17	in
19	EGFP site 19	307	GGAGCGCACCATCTTCTTCAAGG	64.	CTCGAACTTCACCTCGGCGCGGG	65.	18	in
20	EGFP site 20	155	CTGAAGTTCATCTGCACCACCGG	66.	GTCAGGGTGGTCACGAGGGTGGG	67.	19	in
21	EGFP site 21	95	GGCCACAAGTTCAGCGTGTCCGG	68.	CTTCAGGGTCAGCTTGCCGTAGG	69.	20	in
22	EGFP site 22	203	CTCGTGACCACCCTGACCTACGG	70.	CGTGCTGCTTCATGTGGTCGGGG	71.	21	in
23	EGFP site 23	174	CCGGCAAGCTGCCCGTGCCCTGG	72.	GCTGAAGCACTGCACGCCGTAGG	73.	22	in
24	EGFP site 24	107	AGCGTGTCCGGCGAGGGCGAGGG	74.	GGTGGTGCAGATGAACTTCAGGG	75.	23	in
25	EGFP site 25	106	CAGCGTGTCCGGCGAGGGCGAGG	76.	GGTGGTGCAGATGAACTTCAGGG	77.	24	in
26	EGFP site 26	49	CACCGGGGTGGTGCCCATCCTGG	78.	CGCCGGACACGCTGAACTTGTGG	79.	25	in
27	EGFP site 27	122	GGCGAGGGCGATGCCACCTACGG	80.	GGGCACGGGCAGCTTGCCGGTGG	81.	26	in
28	EGFP site 28	203	CTCGTGACCACCCTGACCTACGG	82.	AGAAGTCGTGCTGCTTCATGTGG	83.	27	in

TABLE 2

FokI-dCas9 EGFP Pair #	Name	Target Start Position (+)	Sequence (+) sites	SEQ ID NO:	Sequence (-) sites	SEQ ID NO:	Edge-to-edge 'spacer' distance	PAM
29	EGFP site 29	337	CAACTACAAGACCCGCGCCGAGG	84.	CGATGCCCTTCAGCTCGATGCGG	85.	28	in
30	EGFP site 30	62	CCCATCCTGGTCGAGCTGGACGG	86.	GGCATCGCCCTCGCCCTCGCCGG	87.	29	in
31	EGFP site 31	100	CAAGTTCAGCGTGTCCGGCGAGG	88.	GGTGGTGCAGATGAACTTCAGGG	89.	30	in
32	EGFP site 32	74	GAGCTGGACGGCGACGTAAACGG	90.	GACCAGGATGGGCACCACCCCGG	91.	0	out
33	EGFP site 33	314	ACCATCTTCTTCAAGGACGACGG	92.	CGCTCCTGGACGTAGCCTTCGGG	93.	1	out
34	EGFP site 34	122	GGCGAGGGCGATGCCACCTACGG	94.	CGCCGGACACGCTGAACTTGTGG	95.	2	out
35	EGFP site 35	275	TTCAAGTCCGCCATGCCCGAAGG	96.	GTCGTGCTGCTTCATGTGGTCGG	97.	3	out
36	EGFP site 36	275	TTCAAGTCCGCCATGCCCGAAGG	98.	TCGTGCTGCTTCATGTGGTCGGG	99.	4	out
37	EGFP site 37	95	GGCCACAAGTTCAGCGTGTCCGG	100.	CGTCGCCGTCCAGCTCGACCAGG	101.	5	out
38	EGFP site 38	203	CTCGTGACCACCCTGACCTACGG	102.	CCAGGGCACGGGCAGCTTGCCGG	103.	6	out
39	EGFP site 39	463	CAGCCACAACGTCTATATCATGG	104.	TGTACTCCAGCTTGTGCCCCAGG	105.	7	out
40	EGFP site 40	95	GGCCACAAGTTCAGCGTGTCCGG	106.	GCCGTCCAGCTCGACCAGGATGG	107.	9	out
41	EGFP site 41	95	GGCCACAAGTTCAGCGTGTCCGG	108.	CCGTCCAGCTCGACCAGGATGGG	109.	10	out
42	EGFP site 42	101	AAGTTCAGCGTGTCCGGCGAGGG	110.	CGTCGCCGTCCAGCTCGACCAGG	111.	11	out

TABLE 2

FokI-dCas9 EGFP Pair #	Name	Target Start Position (+)	Sequence (+) sites	SEQ ID NO:	Sequence (-) sites	SEQ ID NO:	Edge-to-edge 'spacer' distance	PAM
43	EGFP site 43	350	CGCGCCGAGGTGAAGTTCGAGG G	112.	GCCGTCGTCCTTGAAGAAGATGG	113.	12	out
44	EGFP site 44	174	CCGGCAAGCTGCCCGTGCCCTG G	114.	CTTCAGGGTCAGCTTGCCGTAGG	115.	13	out
45	EGFP site 45	100	CAAGTTCAGCGTGTCCGGCGAG G	116.	GCCGTCCAGCTCGACCAGGATGG	117.	14	out
46	EGFP site 46	100	CAAGTTCAGCGTGTCCGGCGAG G	118.	CCGTCCAGCTCGACCAGGATGGG	119.	15	out
47	EGFP site 47	101	AAGTTCAGCGTGTCCGGCGAGG G	120.	CCGTCCAGCTCGACCAGGATGGG	121.	16	out
48	EGFP site 48	107	AGCGTGTCCGGCGAGGGCGAGG G	122.	CGTCGCCGTCCAGCTCGACCAGG	123.	17	out
49	EGFP site 49	155	CTGAAGTTCATCTGCACCACCG G	124.	GGCATCGCCCTCGCCCTCGCCGG	125.	18	out
50	EGFP site 50	106	CAGCGTGTCCGGCGAGGGCGAG G	126.	GCCGTCCAGCTCGACCAGGATGG	127.	20	out
51	EGFP site 51	95	GGCCACAAGTTCAGCGTGTCCG G	128.	GACCAGGATGGGCACCACCCCGG	129.	21	out
52	EGFP site 52	107	AGCGTGTCCGGCGAGGGCGAGG G	130.	CCGTCCAGCTCGACCAGGATGGG	131.	22	out
53	EGFP site 53	337	CAACTACAAGACCCGCGCCGAG G	132.	GCGCTCCTGGACGTAGCCTTCGG	133.	23	out
54	EGFP site 54	337	CAACTACAAGACCCGCGCCGAG G	134.	CGCTCCTGGACGTAGCCTTCGGG	135.	24	out
55	EGFP site 55	397	GCTGAAGGGCATCGACTTCAAG G	136.	CCTCGAACTTCACCTCGGCGCGG	137.	25	out
56	EGFP site 56	100	CAAGTTCAGCGTGTCCGGCGAG G	138.	GACCAGGATGGGCACCACCCCGG	139.	26	out

TABLE 2

FokI-dCas9 EGFP Pair #	Name	Target Start Position (+)	Sequence (+) sites	SEQ ID NO:	Sequence (-) sites	SEQ ID NO:	Edge-to-edge 'spacer' distance	PAM
57	EGFP site 57	101	AAGTTCAGCGTGTCCGGCGAGGG	140.	GACCAGGATGGGCACCACCCCGG	141.	27	out
58	EGFP site 58	400	GAAGGGCATCGACTTCAAGGAGG	142.	CCTCGAACTTCACCTCGGCGCGG	143.	28	out
59	EGFP site 59	337	CAACTACAAGACCCGCGCCGAGG	144.	CTGGACGTAGCCTTCGGGCATGG	145.	29	out
60	EGFP site 60	307	GGAGCGCACCATCTTCTTCAAGG	146.	AGAAGTCGTGCTGCTTCATGTGG	147.	31	out
61	EGFP site 61	100	CAAGTTCAGCGTGTCCGGCGAGG	148.	CGTCGCCGTCCAGCTCGACCAGG	149.	10	out
62	EGFP site 62	286	CATGCCCGAAGGCTACGTCCAGG	150.	AGAAGTCGTGCTGCTTCATGTGG	151.	10	out
63	EGFP site 63	337	CAACTACAAGACCCGCGCCGAGG	152.	TGAAGAAGATGGTGCGCTCCTGG	153.	10	out
64	EGFP site 64	382	GGTGAACCGCATCGAGCTGAAGG	154.	CCTCGAACTTCACCTCGGCGCGG	155.	10	out
65	EGFP site 65	275	TTCAAGTCCGCCATGCCCGAAGG	156.	GCTTCATGTGGTCGGGGTAGCGG	157.	11	out
66	EGFP site 66	349	CCGCGCCGAGGTGAAGTTCGAGG	158.	GCCGTCGTCCTTGAAGAAGATGG	159.	11	out
67	EGFP site 67	382	GGTGAACCGCATCGAGCTGAAGG	160.	CTCGAACTTCACCTCGGCGCGGG	161.	11	out
68	EGFP site 68	383	GTGAACCGCATCGAGCTGAAGGG	162.	CCTCGAACTTCACCTCGGCGCGG	163.	11	out
69	EGFP site 69	520	CAAGATCCGCCACAACATCGAGG	164.	GATGCCGTTCTTCTGCTTGTCGG	165.	11	out
70	EGFP site 70	383	GTGAACCGCATCGAGCTGAAGGG	166.	CTCGAACTTCACCTCGGCGCGGG	167.	12	out

TABLE 2

FokI-dCas9 EGFP Pair #	Name	Target Start Position (+)	Sequence (+) sites	SEQ ID NO:	Sequence (-) sites	SEQ ID NO:	Edge-to-edge 'spacer' distance	PAM
71	EGFP site 71	415	CAAGGAGGACGGCAACATCCTGG	168.	TCAGCTCGATGCGGTTCACCAGG	169.	13	out
72	EGFP site 72	286	CATGCCCGAAGGCTACGTCCAGG	170.	GTCGTGCTGCTTCATGTGGTCGG	171.	14	out
73	EGFP site 73	415	CAAGGAGGACGGCAACATCCTGG	172.	CAGCTCGATGCGGTTCACCAGGG	173.	14	out
74	EGFP site 74	416	AAGGAGGACGGCAACATCCTGGG	174.	TCAGCTCGATGCGGTTCACCAGG	175.	14	out
75	EGFP site 75	101	AAGTTCAGCGTGTCCGGCGAGGG	176.	GCCGTCCAGCTCGACCAGGATGG	177.	15	out
76	EGFP site 76	286	CATGCCCGAAGGCTACGTCCAGG	178.	TCGTGCTGCTTCATGTGGTCGGG	179.	15	out
77	EGFP site 77	416	AAGGAGGACGGCAACATCCTGGG	180.	CAGCTCGATGCGGTTCACCAGGG	181.	15	out
78	EGFP site 78	417	AGGAGGACGGCAACATCCTGGGG	182.	TCAGCTCGATGCGGTTCACCAGG	183.	15	out
79	EGFP site 79	524	ATCCGCCACAACATCGAGGACGG	184.	GATGCCGTTCTTCTGCTTGTCGG	185.	15	out
80	EGFP site 80	106	CAGCGTGTCCGGCGAGGGCGAGG	186.	CGTCGCCGTCCAGCTCGACCAGG	187.	16	out
81	EGFP site 81	174	CCGGCAAGCTGCCCGTGCCCTGG	188.	CAGGGTCAGCTTGCCGTAGGTGG	189.	16	out
82	EGFP site 82	286	CATGCCCGAAGGCTACGTCCAGG	190.	CGTGCTGCTTCATGTGGTCGGGG	191.	16	out
83	EGFP site 83	417	AGGAGGACGGCAACATCCTGGGG	192.	CAGCTCGATGCGGTTCACCAGGG	193.	16	out
84	EGFP site 84	427	CAACATCCTGGGGCACAAGCTGG	194.	CGATGCCCTTCAGCTCGATGCGG	195.	16	out

TABLE 2

FokI-dCas9 EGFP Pair #	Name	Target Start Position (+)	Sequence (+) sites	SEQ ID NO:	Sequence (-) sites	SEQ ID NO:	Edge-to-edge 'spacer' distance	PAM
85	EGFP site 85	397	GCTGAAGGGCATCGACTTCAAG G	196.	GTCGCCCTCGAACTTCACCTCGG	197.	20	out

Surprisingly, the dCas9-FokI protein did not show detectable EGFP disruption activity when co-expressed with any of the 60 gRNA pairs in human U2OS.EGFP cells (**Fig. 5E**). However, screening of the FokI-dCas9 protein with the same 60 gRNA pairs did reveal EGFP disruption activity on target sites composed of half-sites in the PAM out orientation and with spacer lengths of 13 to 17 bps and of 26 bps (approximately one turn of the DNA helix more than the 13-17 bp spacer lengths) (**Fig. 5B**). Testing of FokI-dCas9 on an additional 25 target DNA sites with spacer lengths ranging from 10 to 20 bps and with half-sites in the PAM out orientation demonstrated efficient cleavage on targets with spacer lengths of 13 to 18 bps (**Figs. 5C-D**). In these experiments, one site each was tested for spacer lengths of 17 or 18 bps and not all sites with a 13 bp spacer length showed activity. Analysis of a subset of successfully targeted sites by T7EI analysis and Sanger sequencing further confirmed the presence of indels at the intended location. Thus FokI-dCas9 can be directed by two appropriately positioned gRNAs to efficiently cleave a full-length target site of interest. For simplicity, the complex of two FokI-dCas9 fusions and two gRNAs are referred to herein as RNA-guided FokI Nucleases (**RFN**s).

To extend the initial findings with the EGFP reporter gene and to ascertain whether RFNs could be used to perform routine genome editing of endogenous human genes, gRNA pairs were designed for 12 different target sites in nine different human genes (**Table 2**). Eleven of the 12 RFNs tested introduced indels with high efficiencies (range of 3 to 40%) at their intended target sites in human U2OS.EGFP cells as judged by T7EI assay (**Table 2**). Similar results were obtained with these same 12 RFN pairs in HEK293 cells (**Table 2**). Sanger sequencing of successfully targeted alleles from U2OS.EGFP cells revealed the introduction of a range of indels (primarily deletions) at the expected cleavage site (**Fig. 5F**). The high success rate and high efficiencies of modifications observed in two different human cell lines demonstrate the robustness of RFNs for modifying endogenous human genes.

Example 2d. RFNs possess extended specificities for their cleavage sites

To test whether RFNs possess enhanced recognition specificities associated with dimerization, whether these nucleases strictly depend upon the presence of both gRNAs in a pair was examined. In an ideal dimeric system, single gRNAs should not be able to efficiently direct FokI-dCas9-induced indels. To perform an initial test, two pairs of gRNAs directed to two target sites in *EGFP* were used that had been

shown to efficiently direct FokI-dCas9-induced indels to their target sites (EGFP sites 47 and 81) in human U2OS.EGFP cells (**Fig. 5C**). Replacement of one or the other gRNA in each of these two pairs with a gRNA targeted to an unrelated site in *VEGFA* resulted in reduction of EGFP disruption activity (**Fig. 6A**) and reduction of targeted mutations to undetectable levels as judged by T7EI assays (**Fig. 6B**). Similarly, the effects of using only one of each of the two gRNAs were tested using pairs that efficiently introduce FokI-dCas9-mediated indels in the human *APC*, *MLH1*, and *VEGFA* genes (**Table 2**) and again observed loss of detectable RFN-induced indels by T7EI assay (**Fig. 6C**). These results demonstrate that efficient induction of genome editing by an RFN requires two gRNAs with appropriate complementarity to the full-length target site.

Given that the activities of our RFNs depend on the expression of two gRNAs, it was hoped that their mutagenic effects on known off-target sites of one of the single gRNAs in the pair should be negligible. Performing these direct comparisons requires knowing the off-target sites for a monomeric Cas9 nuclease guided by a single gRNA that itself can also serve one of the two gRNAs needed to target a dimeric RFN. Although very few monomeric Cas9 nuclease off-target sites have been defined in the literature, five off-target sites had been previously identified for one of the gRNAs we used to target the dimeric RFN site in the human *VEGFA* gene (Example 1). Deep sequencing was used to ascertain whether these five off-target sites showed evidence of mutations in cells in which the *VEGFA*-targeted RFNs had been expressed (these are the same cells we used for the T7EI assay shown in **Fig. 6C**). The frequency of indel mutations at all five off-target sites was indistinguishable from background (**Fig. 6D and Table 3**). These results demonstrate that the use of RFNs can essentially eliminate the off-target effects originally induced by Cas9 nuclease and a single gRNA and are consistent with our observation that a single gRNA expressed with FokI-dCas9 does not efficiently induce indels. Although, at present, it is not possible to perform these direct comparisons on additional sites – such experiments will have to await the identification of off-target sites for more single gRNA sites that can also target a half-site for a dimeric RFN, it was concluded that dimeric RFNs have enhanced specificities relative to standard monomeric Cas9 nucleases.

TABLE 3

Gene name	left target sequence	SEQ ID NO:	right target sequence	SEQ ID NO:	Endogenous Sequence of RFN target sites in U2OS or 293 cells (spacer sequence in lowercase)	SEQ ID NO:
APC1	CCAGAAGTACGAGCGCCGCCCGG	198.	TGGCAGGTGAGTGAGGCTGCAGG	199.	CCGGGCGGCGCTCGTACTTCTGGccactgggcgagcgtcTGGCAGGTGAGTGAGGCTGCAGG	200.
BRCA1	GAATACCCATCTGTCAGCTTCGG	201.	GGCGGAACCTGAGAGGCGTAAGG	202.	CCGAAGCTGACAGATGGGTATTCtttgacggggggtaggGGCGGAACCTGAGAGGCGTAAGG	203.
DDB2	AATATTCAAGCAGCAGGCACAGG	204.	CTCGCGCAGGAGGCTGCAGCGGG	205.	CCTGTGCCTGCTGCTTGAATATTtccgcctttagggtgCTCGCGCAGGAGGCTGCAGCGGG	206.
EMX1	CCCAAAGCCTGGCCAGGGAGTGG	207.	GCCCCACAGGGCTTGAAGCCCGG	208.	CCACTCCCTGGCCAGGCTTTGGGgaggcctggagtcatgGCCCCACAGGGCTTGAAGCCCGG	209.
FANCF - site 1	CCCTACTTCCGCTTTCACCTTGG	210.	GGAATCCCTTCTGCAGCACCTGG	211.	CCAAGGTGAAAGCGGAAGTAGGGccttcgcgcacctcatGGAATCCCTTCTGCAGCACCTGG	212.
FANCF - site 2	CGCTCCAGAGCCGTGCGAATGGG	213.	TGGAGGCAAGAGGGCGGCTTTGG	214.	CCCATTCGCACGGCTCTGGAGCGgcggctgcacaaccagTGGAGGCAAGAGGGCGGCTTTGG	215.
FES	CGAGGAGACTGGGGACTGTAGGG	216.	CCAGCTGCTGCCTTGCCTCCAGG	217.	CCCTACAGTCCCCAGCCTCCTCGtcccatgcctccgtctCCAGCTGCTGCCTTGCCTCCAGG	218.
GLI1	CATAGCTACTGATTGGTGGTGGG	219.	CCGGCCCCTCCCCAGTCAGGGGG	220.	CCCACCACCAATCAGTAGCTATGgcgagccctgctgtctCCGGCCCCTCCCCAGTCAGGGGG	221.
MLH1	GGAAACGTCTAGATGCTCAACGG	222.	CAAAATGTCGTTCGTGGCAGTGG	223.	CCGTTGAGCATCTAGACGTTTCCttggctcttctggcgcCAAAATGTCGTTCGTGGCAGGGG	224.
RARA1	CTGTTGCTGGCCATGCCAAGCGG	225.	CCTGGGGGCGGGCACCTCAATGG	226.	CCGCTTGGCATGGCCAGCAACAGcagctcctgcccgacaCCTGGGGGCGGGCACCTCAATGG	227.
RUNX	TTCGGAGCGAAAACCAAGACAGG	228.	GAGTCCCCCGCCTTCAGAAGAGG	229.	CCTGTCTTGGTTTTCGCTCCGAAggtaaaagaaatcattGAGTCCCCCGCCTTCAGAAGAGG	230.
SS18	GGCCCGGTCGACTCCGGGCCCGG	231.	TGCTGGGAATCAGCAGTGTTTGG	232.	CCGGGCCCGGAGTCGACCGGGCCgaggcggaggcgggccTGCTGGGAATCAGCAGTGTTTGG	233.
VEGFA - site 1	GGGTGGGGGGAGTTTGCTCCTGG	234.	TCCCTCTTTAGCCAGAGCCGGGG	235.	CCAGGAGCAAACTCCCCCCACCCcctttccaaagcccatTCCCTCTTTAGCCAGAGCCGGGG	236.
VEGFA - site 2	GCCGCCGGCCGGGGAGGAGGTGG	237.	GGCGAGCCGCGGGCAGGGGCCGG	238.	CCACCTCCTCCCCGGCCGGCGGcggacagtggacgcggcGGCGAGCCGCGGGCAGGGGCCGG	239.
VEGFA - site 3	CCGTCTGCACACCCCGGCTCTGG	240.	CTCGGCCACCACAGGGAAGCTGG	241.	CCAGAGCCGGGGTGTGCAGACGGcagtcactagggggcgCTCGGCCACCACAGGGAAGCTGG	242.

Table 3

Gene name	primer 1 used for T7E1 assay	SEQ ID NO:	primer 2 used for T7E1 assay	SEQ ID NO:	DMSO added (yes/no)	Thermo-cycler protocol	amplicon size (bp)	sizes of cleavage products in T7E1 (bp)
APC1	GGCTGTGGGAAGCCAGCAAC	243.	AAGCCAGGGGCCAACTGGAG	244.	no	touchdown	634	447/187
BRCA1	GCGCGGGAATTACAGATAAATTAAAA	245.	AGGTCCCATCCTCTCATACATACCA	246.	no	touchdown	751	454/297
DDB2	ACCGCCCCTTGGCACCAC	247.	CGGAGCTCATCTGCTTCCTGT	248.	no	touchdown	627	456/171
EMX1	GGAGCAGCTGGTCAGAGGGG	249.	GGGAAGGGGGACACTGGGGA	250.	yes	two-step	729	480/249
FANCF - site 1	GCCCTACATCTGCTCTCCCTCCA	251.	GGGCCGGGAAAGAGTTGCTG	252.	no	touchdown	634	361/273
FANCF - site 2	GCCCTACATCTGCTCTCCCTCCA	253.	GGGCCGGGAAAGAGTTGCTG	254.	no	touchdown	634	466/168
FES	GGGGAGGGAGGCTCCAGGTT	255.	GGCACAATGGCTCCCAAGCA	256.	no	touchdown	633	395/238
GLI1	CCTTACCCCTCCCCTCACTCA	257.	AGAAGGGCGGGCCAGACAGT	258.	no	touchdown	869	590/279
MLH1	ATATCCTTCTAGGTAGCGGGCAGTAGCC	259.	TCTCGGGGGAGAGCGGTAAA	260.	no	touchdown	610	332/278
RARA1	CCCAGGAAAAGTGCCAGCTCA	261.	TGATGGTCACCCCAACTGGA	262.	no	touchdown	632	335/297
RUNX	AAGGCGGCGCTGGCTTTTT	263.	CCAGCACAACTTACTCGCACTTGA	264.	no	touchdown	626	389/237
SS18	GGGATGCAGGGACGGTCAAG	265.	GCCGCCCCATCCCTAGAGAAA	266.	no	touchdown	629	455/174
VEGFA - site 1	TCCAGATGGCACATTGTCAG	267.	AGGGAGCAGGAAAGTGAGGT	268.	no	touchdown	531	338/193
VEGFA - site 2	AGAGAAGTCGAGGAAGAGAGAG	269.	CAGCAGAAAGTTCATGGTTTCG	270.	yes	touchdown	756	482/274
VEGFA - site 3	TCCAGATGGCACATTGTCAG	271.	AGGGAGCAGGAAAGTGAGGT	272.	no	touchdown	531	288/243

TABLE 3

Gene name	primer 1 used for deep sequencing	SEQ ID NO:	primer 2 used for deep sequencing	SEQ ID NO:
APC1				
BRCA1				
DDB2	CGATGGCTCCCAAGAAACGC	273.	GCAGGTAGAATGCACAGCCG	274.
EMX1				
FANCF - site 1	GCCCAGAGTCAAGGAACACG	275.	AGGTAGTGCTTGAGACCGCC	276.
FANCF - site 2	CATCCATCGGCGCTTTGGTC	277.	CCGGGAAAGAGTTGCTGCAC	278.
FES	CTCCCCGTCTGCAGTCCATC	279.	CCTGCAGGGACATGTGGTGA	280.
GLI1				
MLH1				
RARA1				
RUNX	TAGGGCTAGAGGGGTGAGGC	281.	CCGAGGTGAAACAAGCTGCC	282.
SS18				
VEGFA - site 1	ATGAGGGCTCCAGATGGCAC	283.	TTCACCCAGCTTCCCTGTGG	284.
VEGFA - site 2				
VEGFA - site 3				

Example 2e. Monomeric Cas9 nickases induce higher rates of mutagenesis than single gRNA/FokI-dCas9 complexes

As noted above, an important weakness of the paired Cas9 nickase approach is that single monomeric nickases can introduce indel mutations with high frequencies at certain target sites (See Example 1 and Ran et al., Cell 154, 1380-1389 (2013); Mali et al., Nat Biotechnol 31, 833-838 (2013); Cho et al., Genome Res (2013); and Mali et al., Science 339, 823-826 (2013)). This lack of dimerization-dependence in the paired Cas9 nickase system is a potential source of off-target effects because the two monomeric nickases can each create unwanted indel mutations elsewhere in the genome. It was hypothesized that because RFNs introduce alterations using a dimerization-dependent FokI nuclease, these fusions should generally show less undesirable indel activity in the presence of only one gRNA compared to what is observed with monomeric Cas9 nickases.

To test this hypothesis, the activities of FokI-dCas9 and Cas9 nickase were compared in the presence of a single gRNA at six dimeric human gene target sites (a total of 12 half-sites; **Table 4**). These particular sites were chosen because monomeric Cas9 nickases directed by just one and/or the other gRNA in a pair could induce indel mutations at these targets. Using deep sequencing, the genome editing activities of FokI-dCas9 or Cas9 nickase were assessed in the presence of both or only one or the other gRNAs. Both FokI-dCas9 and Cas9 nickase induced indels at all six target sites with high efficiencies in the presence of two gRNAs (**Table 5**). As hypothesized, monomeric Cas9 nickases directed by the 12 single gRNAs induced indels with frequencies ranging from 0.0048% to 3.04% (**Fig. 7A** and **Table 5**). By contrast, FokI-dCas9 directed by the same 12 single gRNAs induced indels at lower frequencies ranging from 0.0045% to 0.473% (**Fig. 7A** and **Table 5**). Comparing these data directly, FokI-dCas9 induced indels with lower frequencies than Cas9 nickase for 10 of the 12 single gRNAs (**Fig. 7A** and **Table 5**). In addition, FokI-dCas9 showed greater fold-reductions in indel frequencies than Cas9 nickase at 11 of the 12 half-sites when comparing paired gRNA rates to single gRNA rates (**Fig. 7B**).

Table 4

Chromosome	Position	Site	FokI-dCas9 Indels	FokI-dCas9 Total	FokI-dCas9 Indel Frequency (%)	tdTomato control indels	tdTomato Total	tdTomato Indel Frequency (%)
6	43737290	VEGFA site 1	35000	150158	23.30878	10	258108	0.00387
15	65637531	OT1-3	1	169681	0.00058	1	139847	0.00071
12	131690182	OT1-4	4	190111	0.00210	5	139762	0.00357
12	1988060	OT1-6	3	258976	0.00115	2	178162	0.00112
1	99347645	OT1-11	4	235853	0.00169	4	186287	0.00214
17	39796322	OT1-30	1	261605	0.00038	1	286850	0.00034

Table 5. Deep sequencing of FokI-dCas9, Cas9n, and tdTomato controls at 6 sites, with single and pairs of gRNAs (same data as presented in Fig. 7).

Nuclease Type or Control	Site	guideRNA	Chromosome	Position	Indel	Totals	Percentages
FokI-dCas9	VEGFA site 1	both	6	43737290	35000	150158	23.3088
FokI-dCas9	VEGFA site 1	left	6	43737290	5	95476	0.0052
FokI-dCas9	VEGFA site 1	right	6	43737290	9	91962	0.0098
FokI-dCas9	DDB2	both	11	47236820	11303	50062	22.5780
FokI-dCas9	DDB2	left	11	47236820	311	85726	0.3628
FokI-dCas9	DDB2	right	11	47236820	153	95050	0.1610
FokI-dCas9	FANCF site 1	both	11	22647331	65846	195311	33.7134
FokI-dCas9	FANCF site 1	left	11	22647331	19	27487	0.0691
FokI-dCas9	FANCF site 1	right	11	22647331	845	225154	0.3753
FokI-dCas9	FANCF site 2	both	11	22647138	27743	120314	23.0588
FokI-dCas9	FANCF site 2	left	11	22647138	989	205832	0.4805
FokI-dCas9	FANCF site 2	right	11	22647138	142	165130	0.0860
FokI-dCas9	FES	both	15	91428181	14260	125912	11.3254
FokI-dCas9	FES	left	15	91428181	4	143877	0.0028
FokI-dCas9	FES	right	15	91428181	7	145495	0.0048
FokI-dCas9	RUNX1	both	21	36421217	61057	136164	44.8408
FokI-dCas9	RUNX1	left	21	36421217	222	162636	0.1365
FokI-dCas9	RUNX1	right	21	36421217	109	169122	0.0645
Cas9n	VEGFA site 1	both	6	43737290	14294	99036	14.4331
Cas9n	VEGFA site 1	left	6	43737290	573	82316	0.6961
Cas9n	VEGFA site 1	right	6	43737290	315	101957	0.3090
Cas9n	DDB2	both	11	47236820	6673	31168	21.4098
Cas9n	DDB2	left	11	47236820	1680	56019	2.9990
Cas9n	DDB2	right	11	47236820	172	42424	0.4054
Cas9n	FANCF site 1	both	11	22647331	66827	193111	34.6055
Cas9n	FANCF site 1	left	11	22647331	1565	109029	1.4354
Cas9n	FANCF site 1	right	11	22647331	2457	109289	2.2482
Cas9n	FANCF site 2	both	11	22647138	17007	111468	15.2573
Cas9n	FANCF site 2	left	11	22647138	120	100591	0.1193
Cas9n	FANCF site 2	right	11	22647138	1063	93162	1.1410
Cas9n	FES	both	15	91428181	16529	126597	13.0564
Cas9n	FES	left	15	91428181	6	125196	0.0048
Cas9n	FES	right	15	91428181	23	46102	0.0499
Cas9n	RUNX1	both	21	36421217	80029	216800	36.9137
Cas9n	RUNX1	left	21	36421217	1106	108670	1.0178
Cas9n	RUNX1	right	21	36421217	2169	121413	1.7865
tdTomato controls (-)	VEGF site 1	none	6	43737290	29	313517	0.0092

Table 5. Deep sequencing of FokI-dCas9, Cas9n, and tdTomato controls at 6 sites, with single and pairs of gRNAs (same data as presented in Fig. 7).

Nuclease Type or Control	Site	guideRNA	Chromosome	Position	Indel	Totals	Percentages
tdTomato controls (-)	FANCF site 1	none	11	22647331	18	578378	0.0031
tdTomato controls (-)	FANCF site 2	none	11	22647138	81	393821	0.0206
tdTomato controls (-)	FES	none	15	91428181	21	410620	0.0051
tdTomato controls (-)	DDB2	none	11	47236820	14	165314	0.0085
tdTomato controls (-)	RUNX1	none	21	36421217	13	511977	0.0025

The deep sequencing experiments also uncovered a previously undescribed and unexpected side-effect of certain monomeric Cas9 nickases: the introduction of point mutations at particular positions within their target sites. Cas9 nickase co-
5 expressed with a single gRNA for the "right" half-site of the VEGFA target induced base substitutions at position 15 of the recognition site at a frequency of 10.5% (**Fig. 8A**). Similar results were observed with Cas9 nickase and single gRNAs directed to the "right" half-site of FANCF target site 1 (mutation frequency of 16.3% at position 16) (**Fig. 8B**) or to the "right" half-site of the RUNX1 target site (mutation frequency
10 of 2% at position 17) (**Fig. 8C**). Point mutations at these positions were not observed above background levels in control samples in which no Cas9 nickase or gRNA are expressed in the cell (**Figs. 8A-8C**). Interestingly, for two of the three sites at which this hypermutation was observed, most of the substitutions observed are C to G transversions on the non-target DNA strand. The positions at which these point
15 mutations were observed fell within a strand-separated region of the target site that has been observed to be susceptible to P1 nuclease *in vitro* in a dCas9/gRNA/target DNA complex. Importantly, these point mutations occur at much lower frequencies (five to 100-fold lower) in cells that express FokI-dCas9 protein and the same gRNAs (**Fig. 8A-C**). Overall, it was concluded that FokI-dCas9 nucleases directed by a
20 single gRNA generally induce mutagenic indel and point mutations with lower frequencies than matched single Cas9 nickases.

Example 2f. Dimeric RFNs possess a high degree of specificity

Dimeric RFNs directed by two gRNAs are not expected to induce appreciable off-target mutations in human cells. RFNs, directed by a pair of gRNAs to cleave a

full-length sequence composed of two half-sites, would be expected to specify up to 44 bps of DNA in the target site. A sequence of this length will, by chance, almost always be unique (except in certain circumstances where the target might lie in duplicated genome sequence). In addition, the most closely matched sites in the genome to this full-length site should, in most cases, possess a large number of mismatches, which in turn would be expected to minimize or abolish cleavage activity by an RFN dimer. Indeed, all sites in the human genome that bear 0 to 16 mismatches (and that allow for spacers of length 14 to 17 bps) for the 15 full-length sequences successfully targeted with RFNs in this study were identified. This analysis showed that all 15 full-length sequences were unique and that the most closely matched sites in the genome ranged from 7 to 12 mismatches (**Table 6**). Sites containing this number of mismatches should not be efficiently mutagenized by RFNs and it will be interesting in future studies to confirm this hypothesis. Overall, dimeric RFNs should possess a high degree of specificity in human cells but the ultimate characterization of specificity will await the development of unbiased methods that can comprehensively define RFN specificity across the entire genome.

Table 6. Frequencies of candidate FokI-dCas9 off-target sites in the human genome that bear a defined number of mismatches

Gene	0	7	8	9	10	11	12	13	14	15	16
APC	1					1	2	16	74	414	2254
BRCA1	1	1						5	20	164	983
DDB2	1						2	7	58	267	1335
EMX1	1			1		2	8	40	175	828	3494
FANCF	1						2	4	44	298	1639
FANCF	1						2	12	79	358	1718
FES	1					3	8	32	191	939	4505
GLI1	1					2	1	7	69	343	1711
MLH1	1						2	5	22	96	643
RARA	1				1	2	8	39	187	698	2849
RUNX1	1							3	25	145	800
SS18	1					1	2	6	39	280	1207
VEGFA-1	1				1	2	3	22	103	543	2676
VEGFA-2	1				4	9	99	447	1675	5608	18599
VEGFA-3	1						3	20	120	623	2783

References

Cheng, A.W., Wang, H., Yang, H., Shi, L., Katz, Y., Theunissen, T.W., Rangarajan, S., Shivalila, C.S., Dadon, D.B., and Jaenisch, R. Multiplexed activation of endogenous genes by CRISPR-on, an RNA-guided transcriptional activator system.
5 Cell Res 23, 1163-1171. (2013).

Cho, S.W., Kim, S., Kim, J.M. & Kim, J.S. Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. Nat Biotechnol 31, 230-232 (2013).

Cong, L. et al. Multiplex genome engineering using CRISPR/Cas systems. Science 339, 819-823 (2013).

10 Cradick, T.J., Fine, E.J., Antico, C.J., and Bao, G. CRISPR/Cas9 systems targeting beta-globin and CCR5 genes have substantial off-target activity. Nucleic Acids Res. (2013).

Dicarlo, J.E. et al. Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems. Nucleic Acids Res (2013).

15 Ding, Q., Regan, S.N., Xia, Y., Oostrom, L.A., Cowan, C.A., and Musunuru, K. Enhanced efficiency of human pluripotent stem cell genome editing through replacing TALENs with CRISPRs. Cell Stem Cell 12, 393-394. (2013).

Fisher, S., Barry, A., Abreu, J., Minie, B., Nolan, J., Delorey, T.M., Young, G., Fennell, T.J., Allen, A., Ambrogio, L., et al. A scalable, fully automated process for
20 construction of sequence-ready human exome targeted capture libraries. Genome Biol 12, R1. (2011).

Friedland, A.E., Tzur, Y.B., Esvelt, K.M., Colaiacovo, M.P., Church, G.M., and Calarco, J.A. Heritable genome editing in C. elegans via a CRISPR-Cas9 system. Nat Methods 10, 741-743. (2013).

25 Fu, Y., Foden, J.A., Khayter, C., Maeder, M.L., Reyon, D., Joung, J.K., and Sander, J.D. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. Nat Biotechnol 31, 822-826. (2013).

Gabriel, R. et al. An unbiased genome-wide analysis of zinc-finger nuclease specificity. Nat Biotechnol 29, 816-823 (2011).

30 Gilbert, L.A., Larson, M.H., Morsut, L., Liu, Z., Brar, G.A., Torres, S.E., Stern-Ginossar, N., Brandman, O., Whitehead, E.H., Doudna, J.A., et al. (2013). CRISPR-Mediated Modular RNA-Guided Regulation of Transcription in Eukaryotes. Cell 154, 442-451.

Gratz, S.J. et al. Genome engineering of Drosophila with the CRISPR RNA-guided Cas9 nuclease. Genetics (2013).

Hockemeyer, D. et al. Genetic engineering of human pluripotent cells using TALE nucleases. Nat Biotechnol 29, 731-734 (2011).

5 Horvath, P. & Barrangou, R. CRISPR/Cas, the immune system of bacteria and archaea. Science 327, 167-170 (2010).

Hsu, P.D., Scott, D.A., Weinstein, J.A., Ran, F.A., Konermann, S., Agarwala, V., Li, Y., Fine, E.J., Wu, X., Shalem, O., et al. DNA targeting specificity of RNA-guided Cas9 nucleases. Nat Biotechnol 31, 827-832. (2013).

10 Hwang, W.Y. et al. Efficient genome editing in zebrafish using a CRISPR-Cas system. Nat Biotechnol 31, 227-229 (2013).

Hwang, W.Y., Fu, Y., Reyon, D., Maeder, M.L., Kaini, P., Sander, J.D., Joung, J.K., Peterson, R.T., and Yeh, J.R. Heritable and Precise Zebrafish Genome Editing Using a CRISPR-Cas System. PLoS One 8, e68708. (2013a).

15 Jiang, W., Bikard, D., Cox, D., Zhang, F. & Marraffini, L.A. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Nat Biotechnol 31, 233-239 (2013).

Jinek, M. et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 337, 816-821 (2012).

Jinek, M. et al. RNA-programmed genome editing in human cells. Elife 2, e00471
20 (2013).

Li, D., Qiu, Z., Shao, Y., Chen, Y., Guan, Y., Liu, M., Li, Y., Gao, N., Wang, L., Lu, X., et al. Heritable gene targeting in the mouse and rat using a CRISPR-Cas system. Nat Biotechnol 31, 681-683. (2013a).

Li, W., Teng, F., Li, T., and Zhou, Q. Simultaneous generation and germline
25 transmission of multiple gene mutations in rat using CRISPR-Cas systems. Nat Biotechnol 31, 684-686. (2013b).

Maeder, M.L., Linder, S.J., Cascio, V.M., Fu, Y., Ho, Q.H., and Joung, J.K. CRISPR RNA-guided activation of endogenous human genes. Nat Methods 10, 977-979. (2013).

30 Mali, P., Aach, J., Stranges, P.B., Esvelt, K.M., Moosburner, M., Kosuri, S., Yang, L., and Church, G.M. CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. Nat Biotechnol 31, 833-838. (2013a).

Mali, P., Esvelt, K.M., and Church, G.M. Cas9 as a versatile tool for engineering biology. Nat Methods 10, 957-963. (2013b).

Mali, P. et al. RNA-guided human genome engineering via Cas9. Science 339, 823-826 (2013c).

5 Pattanayak, V., Lin, S., Guilinger, J.P., Ma, E., Doudna, J.A., and Liu, D.R. High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. Nat Biotechnol 31, 839-843. (2013).

Pattanayak, V., Ramirez, C.L., Joung, J.K. & Liu, D.R. Revealing off-target cleavage specificities of zinc-finger nucleases by in vitro selection. Nat Methods 8, 765-770

10 (2011).

Perez, E.E. et al. Establishment of HIV-1 resistance in CD4+ T cells by genome editing using zinc-finger nucleases. Nat Biotechnol 26, 808-816 (2008).

Perez-Pinera, P., Kocak, D.D., Vockley, C.M., Adler, A.F., Kabadi, A.M., Polstein, L.R., Thakore, P.I., Glass, K.A., Ousterout, D.G., Leong, K.W., et al. RNA-guided

15 gene activation by CRISPR-Cas9-based transcription factors. Nat Methods 10, 973-976. (2013).

Qi, L.S., Larson, M.H., Gilbert, L.A., Doudna, J.A., Weissman, J.S., Arkin, A.P., and Lim, W.A. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell 152, 1173-1183. (2013).

20 Ran, F.A., Hsu, P.D., Lin, C.Y., Gootenberg, J.S., Konermann, S., Trevino, A.E., Scott, D.A., Inoue, A., Matoba, S., Zhang, Y., et al. Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. Cell 154, 1380-1389. (2013).

Reyon, D. et al. FLASH assembly of TALENs for high-throughput genome editing. Nat Biotech 30, 460-465 (2012).

25 Sander, J.D., Maeder, M.L., Reyon, D., Voytas, D.F., Joung, J.K., and Dobbs, D. ZiFiT (Zinc Finger Targeter): an updated zinc finger engineering tool. Nucleic Acids Res 38, W462-468. (2010).

Sander, J.D., Ramirez, C.L., Linder, S.J., Pattanayak, V., Shoresh, N., Ku, M., Foden, J.A., Reyon, D., Bernstein, B.E., Liu, D.R., et al. In silico abstraction of zinc finger

30 nuclease cleavage profiles reveals an expanded landscape of off-target sites. Nucleic Acids Res. (2013).

Sander, J.D., Zaback, P., Joung, J.K., Voytas, D.F., and Dobbs, D. Zinc Finger Targeter (ZiFiT): an engineered zinc finger/target site design tool. Nucleic Acids Res 35, W599-605. (2007).
Shen, B. et al. Generation of gene-modified mice via Cas9/RNA-mediated gene
5 targeting. Cell Res (2013).
Sugimoto, N. et al. Thermodynamic parameters to predict stability of RNA/DNA hybrid duplexes. Biochemistry 34, 11211-11216 (1995).
Terns, M.P. & Terns, R.M. CRISPR-based adaptive immune systems. Curr Opin Microbiol 14, 321-327 (2011).
10 Wang, H. et al. One-Step Generation of Mice Carrying Mutations in Multiple Genes by CRISPR/Cas-Mediated Genome Engineering. Cell 153, 910-918 (2013).
Wiedenheft, B., Sternberg, S.H. & Doudna, J.A. RNA-guided genetic silencing systems in bacteria and archaea. Nature 482, 331-338 (2012).
Yang, L., Guell, M., Byrne, S., Yang, J.L., De Los Angeles, A., Mali, P., Aach, J.,
15 Kim-Kiselak, C., Briggs, A.W., Rios, X., et al. (2013). Optimization of scarless human stem cell genome editing. Nucleic Acids Res 41, 9049-9061.

OTHER EMBODIMENTS

It is to be understood that while the invention has been described in
20 conjunction with the detailed description thereof, the foregoing description is intended to illustrate and not limit the scope of the invention, which is defined by the scope of the appended claims. Other aspects, advantages, and modifications are within the scope of the following claims.

SEQUENCE LISTING IN ELECTRONIC FORM

In accordance with Section 111(1) of the Patent Rules, this description contains a sequence listing in electronic form in ASCII text format (file: 60412-4906 Seq 26-NOV-15 v2.txt).

A copy of the sequence listing in electronic form is available from the Canadian Intellectual Property Office.

CLAIMS:

1. An RNA-guided FokI Nuclease (RFN) fusion protein, comprising a FokI catalytic domain sequence fused to the amino terminus of a catalytically inactive CRISPR-associated 9 (dCas9).

5 2. The fusion protein of claim 1, further comprising an intervening linker of from 2-30 amino acids.

3. The fusion protein of claim 2, wherein the linker comprises Gly_4Ser.

4. The fusion protein of any one of claims 1-3, wherein the FokI catalytic domain comprises amino acids 388-583 or 408-583 of SEQ ID NO:4.

10 5. The fusion protein of any one of claims 1-4, wherein the dCas9 comprises mutations at D10, E762, H983, or D986; and at H840 or N863.

6. The fusion protein of claim 5, wherein the dCas9 comprises mutations at:

(i) D10A or D10N; and

(ii) H840A, H840Y or H840N.

15 7. The fusion protein of any one of claims 1-6, wherein the dCas9 is from *S. pyogenes*.

8. A nucleic acid encoding the fusion protein of any one of claims 1-7.

9. A vector comprising the nucleic acid of claim 8.

10. A host cell expressing the fusion protein of any one of claims 1-7.

11. An *in vitro* method of inducing a sequence-specific break in a genomic sequence in
20 a cell, the method comprising expressing in the cell, or contacting the cell with, the RNA-guided FokI Nuclease (RFN) fusion protein of any one of claims 1-7, and guide RNAs that direct the RFN to two target genomic sequences.

12. The *in vitro* method of claim 11, wherein the two target genomic sequences are spaced 0-31 nucleotides apart.

13. The *in vitro* method of claim 12, wherein the two target genomic sequences are spaced 10-20 base pairs apart.

5 14. The *in vitro* method of claim 13, wherein the two target genomic sequences are spaced 13-17 base pairs apart.

15. The *in vitro* method of any one of claims 11-14, wherein the two target sequences each have a protospacer adjacent motif (PAM) sequence at the 3' end.

16. The *in vitro* method of any one of claims 11-15, wherein the guide RNAs are:

10 (a) two single guide RNAs, wherein one single guide RNA targets a first strand, and the other guide RNA targets the complementary strand, and FokI cuts each strand resulting in a pair of nicks on opposite DNA strands, thereby creating a double-stranded break, or

(b) a tracrRNA and two crRNAs wherein one crRNA targets a first strand, and the other crRNA targets the complementary strand, and FokI cuts each strand resulting in a pair of
15 nicks on opposite DNA strands, thereby creating a double-stranded break.

17. The *in vitro* method of any one of claims 11-16, wherein each of the two guide RNAs include a complementarity region that is complementary to 17-20 nucleotides of target genomic sequence.

18. The *in vitro* method of any one of claims 11-17, wherein an indel mutation is
20 induced between the two target sequences.

19. The *in vitro* method of any one of claims 11-18, wherein the specificity of RNA-guided genome editing in a cell is increased.

20. An *in vitro* method of increasing specificity of RNA-guided genome editing in a cell, the method comprising contacting the cell with an RNA-guided FokI Nuclease (RFN) fusion protein of any one of claims 1-7.

21. A kit comprising:

5 an RNA-guided FokI Nuclease (RFN) fusion protein comprising a FokI catalytic domain sequence fused to the amino terminus of a catalytically inactive *Streptococcus pyogenes* CRISPR-associated 9 (dCas9) protein comprising an amino acid sequence that is at least 95% identical to the amino acid sequence of SEQ ID NO: 5, wherein said catalytically inactive *S. pyogenes* Cas9 has point mutations at amino acid residues corresponding to
10 positions (i) D10, E762, H983, or D986, and (ii) H840 or N863 of *S. pyogenes* Cas9, an intervening linker from 2 to 30 amino acids, and

two guide RNAs that direct said RFN fusion protein to a first target genomic sequences, wherein the guide RNAs that direct said RFN fusion protein to said first target genomic sequence and said second target genomic sequence are spaced 10 to 20 nucleotides apart, and
15 said first target genomic sequence comprises a protospacer adjacent motif (PAM) recognition sequence positioned upstream of said first target genomic sequence and said second target genomic sequence comprises a PAM recognition sequence positioned downstream of said second target genomic sequence.

22. The kit of claim 21, wherein said guide RNAs are:

20 (a) two single guide RNAs, wherein one single guide RNA targets a first strand, and a second guide RNA targets a complementary strand, and FokI cuts each strand resulting in a pair of nicks on opposite DNA strands, thereby creating a double-stranded break, or

(b) a tracrRNA and two crRNAs, wherein one crRNA targets a first strand, and a second crRNA targets a complementary strand, and FokI cuts each strand resulting in a pair of
25 nicks on opposite DNA strands, thereby creating a double-stranded break.

23. The kit of claim 21 or 22, wherein each of said two guide RNAs include a complementarity region that is complementary to 17-20 nucleotides of said first target genomic sequence and said second target genomic sequence.

24. The kit of any one of claims 21-23, wherein an indel mutation is induced between
5 said first target genomic sequence and said second target genomic sequence.

25. The kit of any one of claims 21-24, wherein said first target genomic sequence and said second target genomic sequence are spaced 13-17 nucleotides apart.

26. The kit of any one of claims 21-25, wherein said intervening linker comprises Gly4Ser.

10 27. The kit of any one of claims 21-26, wherein the said FokI catalytic domain comprises amino acid residues 388–583 or amino acid residues 408-583 the amino acid sequence of SEQ ID NO:4.

28. The kit of any one of claims 21-27, wherein said point mutations are: (i) D10A or D10N; and (ii) H840A, H840Y or H840N.

15 29. The kit of any one of claims 21-28, wherein said RNA-guided FokI Nuclease fusion protein comprises the amino acid sequence of SEQ ID NO:26.

30. A composition comprising:

a nucleic acid encoding an RNA-guided FokI Nuclease (RFN) fusion protein comprising a FokI catalytic domain sequence fused to the amino terminus of a catalytically inactive
20 *Streptococcus pyogenes* CRISPR-associated 9 (dCas9) protein comprising an amino acid sequence that is at least 95% identical to the amino acid sequence of SEQ ID NO: 5, wherein said catalytically inactive *S. pyogenes* Cas9 has point mutations at amino acid residues corresponding to positions (i) D10, E762, H983, or D986, and (ii) H840 or N863 of *S. pyogenes* Cas9, an intervening linker from 2 to 30 amino acids; and

a nucleic acid encoding two guide RNAs that direct said RFN fusion protein to a first target genomic sequences, wherein the guide RNAs that direct said RFN fusion protein to said first target genomic sequence and said second target genomic sequence are spaced 10 to 20 nucleotides apart, and said first target genomic sequence comprises a protospacer adjacent
5 motif (PAM) recognition sequence positioned upstream of said first target genomic sequence and said second target genomic sequence comprises a PAM recognition sequence positioned downstream of said second target genomic sequence.

31. The composition of claim 30, wherein said guide RNAs are:

(a) two single guide RNAs, wherein one single guide RNA targets a first strand,
10 and a second guide RNA targets a complementary strand, and FokI cuts each strand resulting in a pair of nicks on opposite DNA strands, thereby creating a double-stranded break, or

(b) a tracrRNA and two crRNAs, wherein one crRNA targets a first strand, and a second crRNA targets a complementary strand, and FokI cuts each strand resulting in a pair of nicks on opposite DNA strands, thereby creating a double-stranded break.

15 32. The composition of claim 30 or 31, wherein each of said two guide RNAs include a complementarity region that is complementary to 17-20 nucleotides of said first target genomic sequence and said second target genomic sequence.

33. The composition of any one of claims 30-32, wherein an indel mutation is induced between said first target genomic sequence and said second target genomic sequence.

20 34. The composition of any one of claims 30-33, wherein said first target genomic sequence and said second target genomic sequence are spaced 13-17 nucleotides apart.

35. The composition of any one of claims 30-34, wherein said intervening linker comprises Gly4Ser.

36. The composition of any one of claims 30-35, wherein the said FokI catalytic domain comprises amino acid residues 388–583 or amino acid residues 408-583 the amino acid sequence of SEQ ID NO:4.

37. The composition of any one of claims 30-36, wherein said point mutations are: (i)
5 D10A or D10N; and (ii) H840A, H840Y or H840N.

38. The composition of any one of claims 30-37, wherein said RNA-guided FokI Nuclease fusion protein comprises the amino acid sequence of SEQ ID NO:26.

39. A composition comprising:

an RNA-guided FokI Nuclease (RFN) fusion protein comprising a FokI catalytic
10 domain sequence fused to the amino terminus of a catalytically inactive *Streptococcus pyogenes* CRISPR-associated 9 (dCas9) protein comprising an amino acid sequence that is at least 95% identical to the amino acid sequence of SEQ ID NO: 5, wherein said catalytically inactive *S. pyogenes* Cas9 has point mutations at amino acid residues corresponding to positions (i) D10, E762, H983, or D986, and (ii) H840 or N863 of
15 *S. pyogenes* Cas9, an intervening linker from 2 to 30 amino acids; and

two guide RNAs that direct said RFN fusion protein to a first target genomic sequences, wherein the guide RNAs that direct said RFN fusion protein to said first target genomic sequence and said second target genomic sequence are spaced 10 to 20 nucleotides apart, and said first target genomic sequence comprises a protospacer adjacent motif (PAM)
20 recognition sequence positioned upstream of said first target genomic sequence and said second target genomic sequence comprises a PAM recognition sequence positioned downstream of said second target genomic sequence.

40. The composition of claim 39, wherein said guide RNAs are:

(a) two single guide RNAs, wherein one single guide RNA targets a first strand,
25 and a second guide RNA targets a complementary strand, and FokI cuts each strand resulting in a pair of nicks on opposite DNA strands, thereby creating a double-stranded break, or

(b) a tracrRNA and two crRNAs, wherein one crRNA targets a first strand, and a second crRNA targets a complementary strand, and FokI cuts each strand resulting in a pair of nicks on opposite DNA strands, thereby creating a double-stranded break.

41. The composition of claim 39 or 40, wherein each of said two guide RNAs include a complementarity region that is complementary to 17-20 nucleotides of said first target genomic sequence and said second target genomic sequence.

42. The composition of any one of claims 39-41, wherein an indel mutation is induced between said first target genomic sequence and said second target genomic sequence.

43. The composition of any one of claims 39-42, wherein said first target genomic sequence and said second target genomic sequence are spaced 13-17 nucleotides apart.

44. The composition of any one of claims 39-43, wherein said intervening linker comprises Gly4Ser.

45. The composition of any one of claims 39-44, wherein the said FokI catalytic domain comprises amino acid residues 388–583 or amino acid residues 408-583 the amino acid sequence of SEQ ID NO:4.

46. The composition of any one of claims 39-45, wherein said point mutations are: (i) D10A or D10N; and (ii) H840A, H840Y or H840N.

47. The composition of any one of claims 39-46, wherein said RNA-guided FokI Nuclease fusion protein comprises the amino acid sequence of SEQ ID NO:26.

48. Use for inducing a sequence-specific break in a genomic sequence in a cell of the RNA-guided FokI Nuclease (RFN) fusion protein of any one of claims 1-7, and guide RNAs that direct the RFN to two target genomic sequences.

49. The use of claim 48, wherein the two target genomic sequences are spaced 0-31 nucleotides apart.

50. The use of claim 49, wherein the two target genomic sequences are spaced 10-20 base pairs apart.

51. The use of claim 50, wherein the two target genomic sequences are spaced 13-17 base pairs apart.

5 52. The use of any one of claims 48-51, wherein the two target sequences each have a protospacer adjacent motif (PAM) sequence at the 3' end.

53. The use of any one of claims 48-52, wherein the guide RNAs are:

(a) two single guide RNAs, wherein one single guide RNA targets a first strand, and the other guide RNA targets the complementary strand, and FokI cuts each strand resulting in
10 a pair of nicks on opposite DNA strands, thereby creating a double-stranded break, or

(b) a tracrRNA and two crRNAs wherein one crRNA targets a first strand, and the other crRNA targets the complementary strand, and FokI cuts each strand resulting in a pair of nicks on opposite DNA strands, thereby creating a double-stranded break.

54. The use of any one of claims 48-53, wherein each of the two guide RNAs include a
15 complementarity region that is complementary to 17-20 nucleotides of target genomic sequence.

55. The use of any one of claims 48-54, wherein an indel mutation is induced between the two target sequences.

56. The use of any one of claims 48-55, wherein the specificity of RNA-guided genome
20 editing in a cell is increased.

57. Use for increasing specificity of RNA-guided genome editing in a cell of an RNA-guided FokI Nuclease (RFN) fusion protein of any one of claims 1-7.

CA 2907198 2018-02-16

CRISPR/Cas9 - RGN

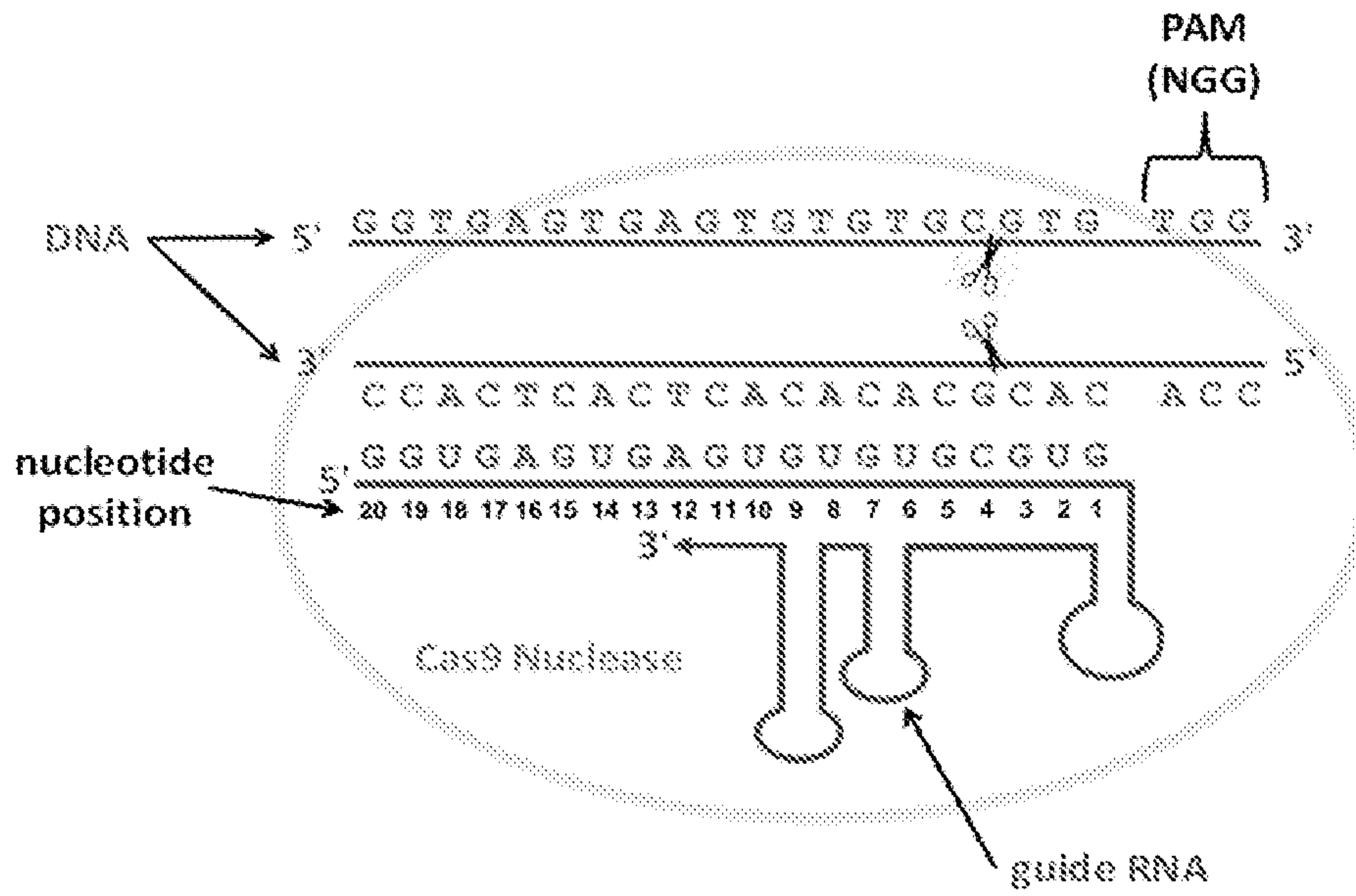


Figure 1

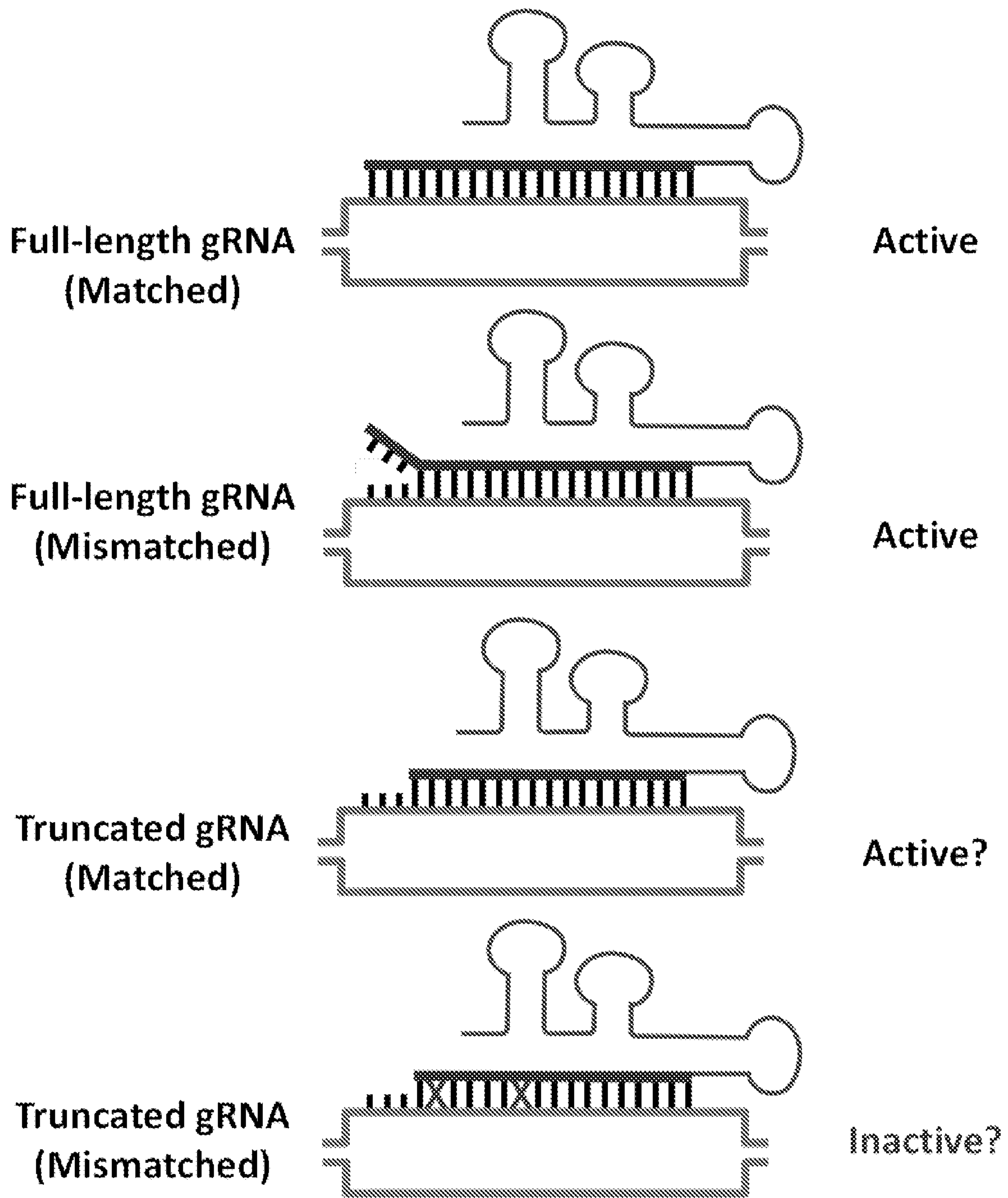


Figure 2A

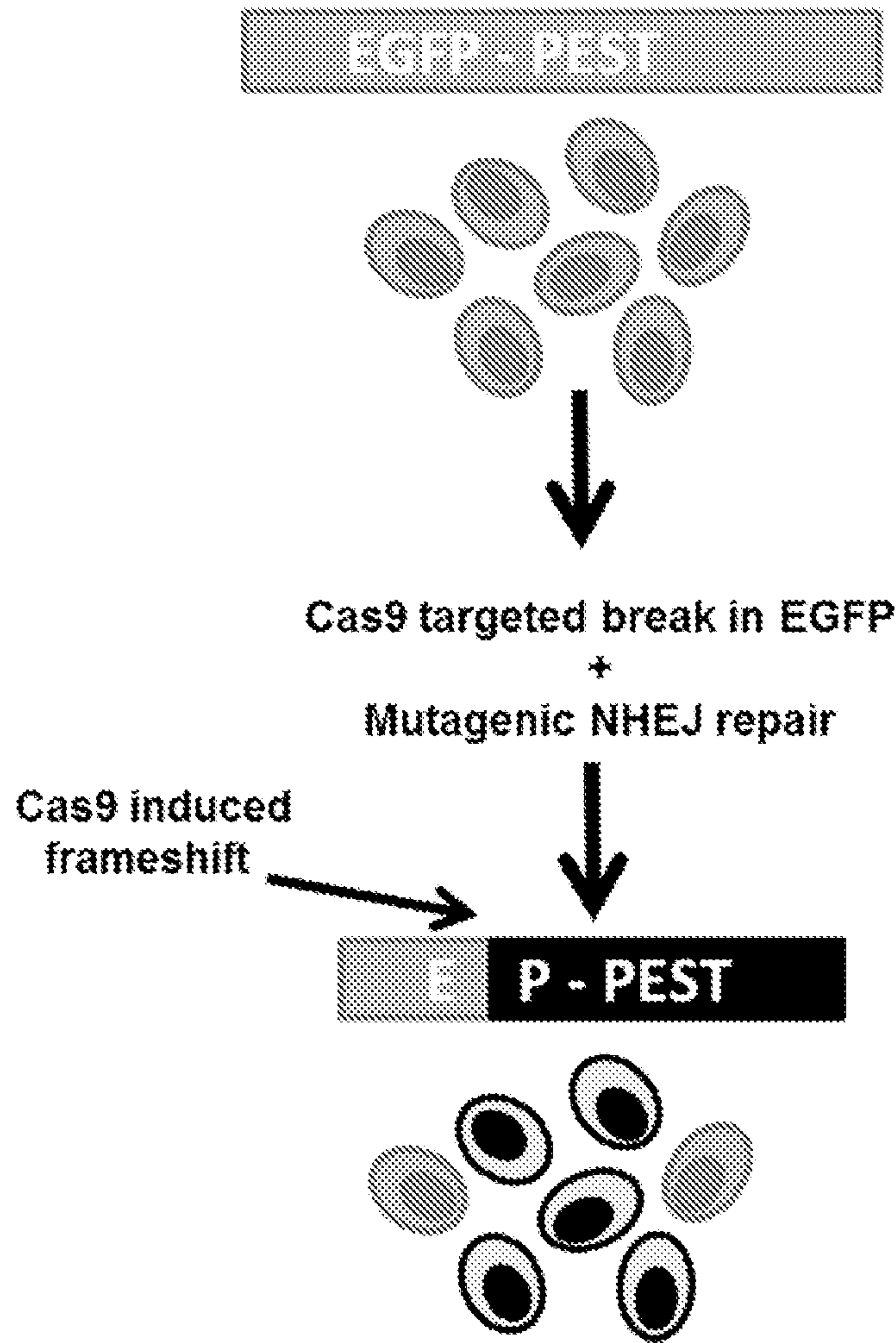


Figure 2B

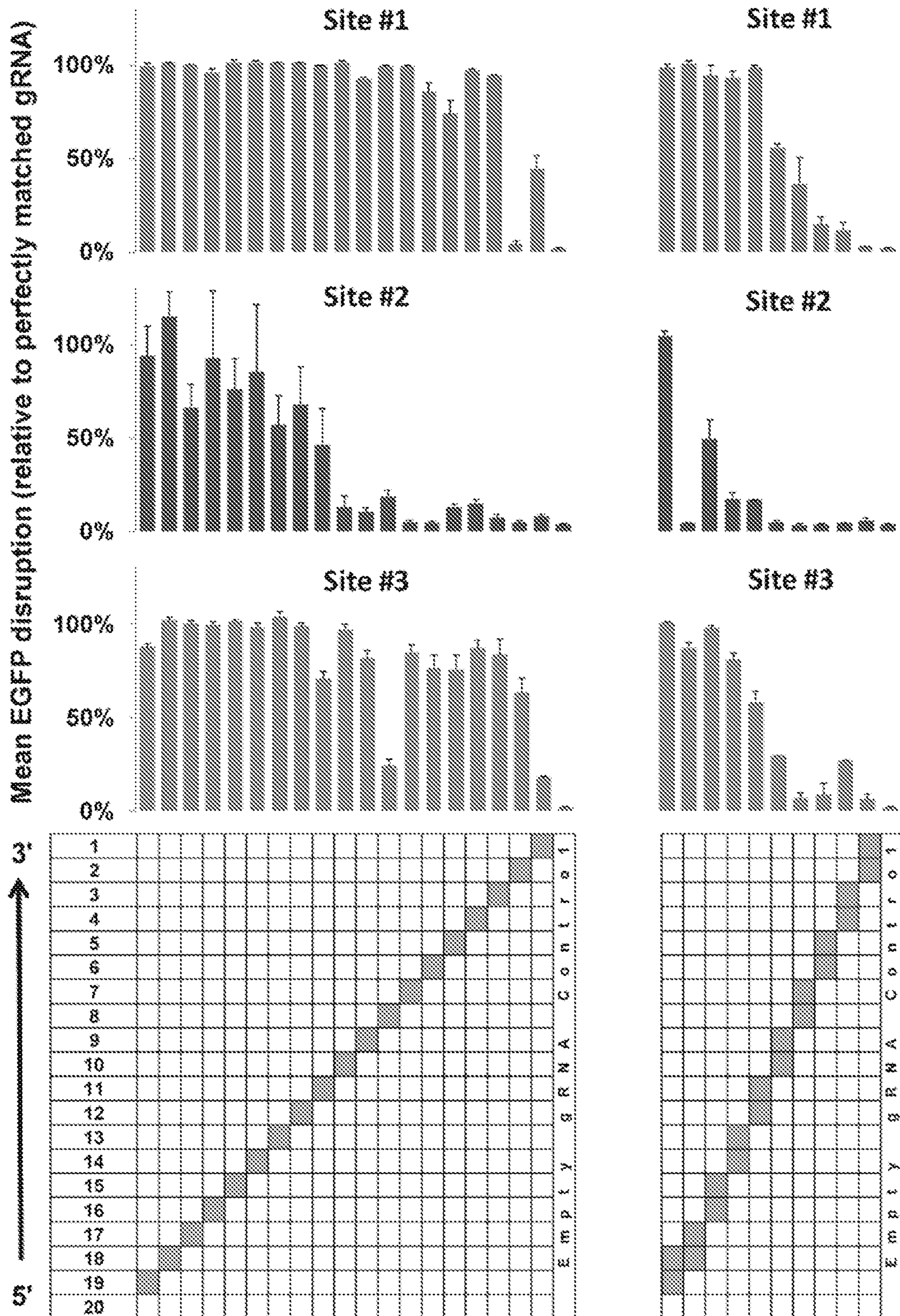


FIG. 2C FIG. 2D

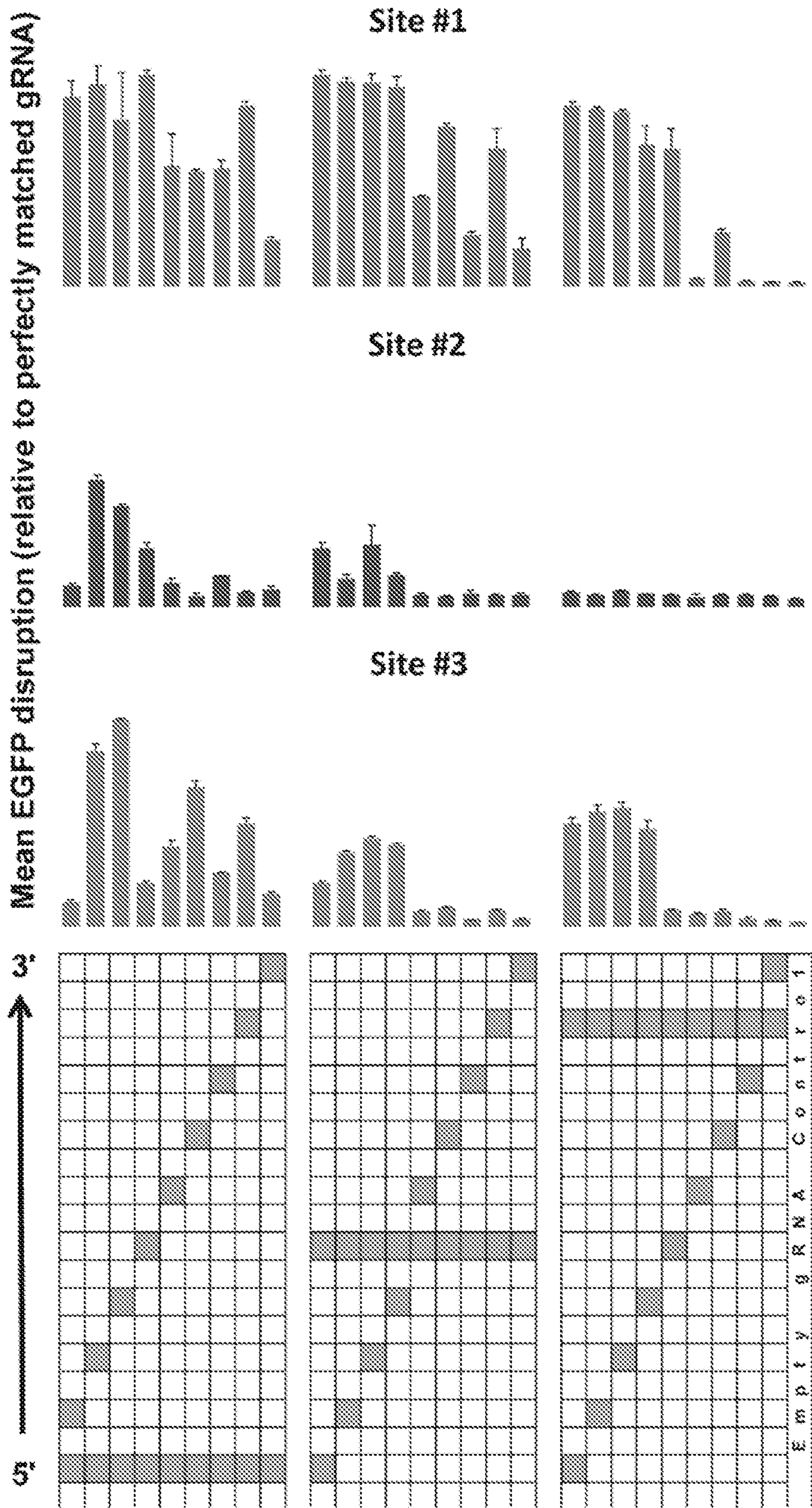


FIG. 2E

Site #1

Site #2

Site #3

Mean EGFP disruption (relative to perfectly matched gRNA)

5' 3'

Empty gRNA Control

FIG. 2F

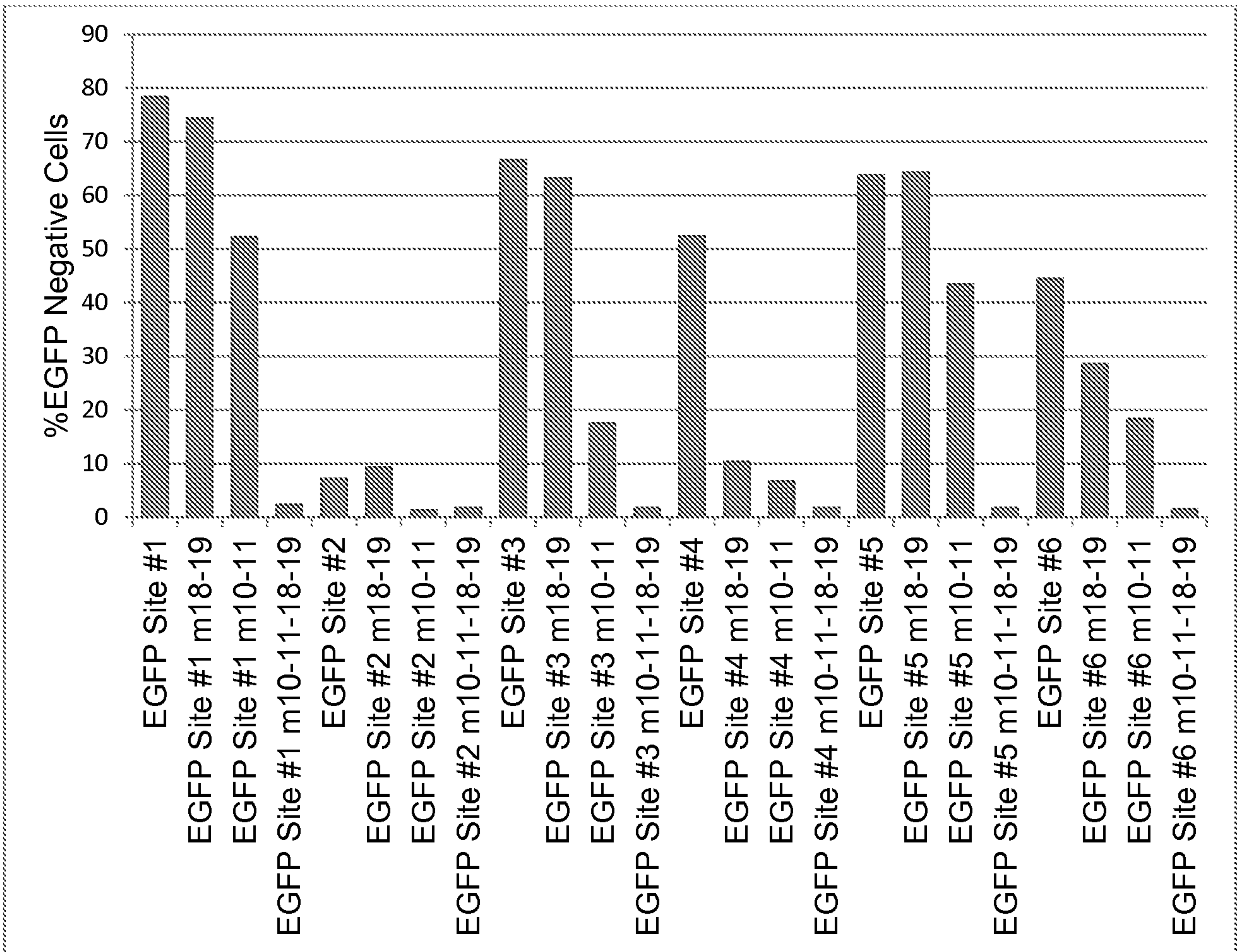


FIG. 2G

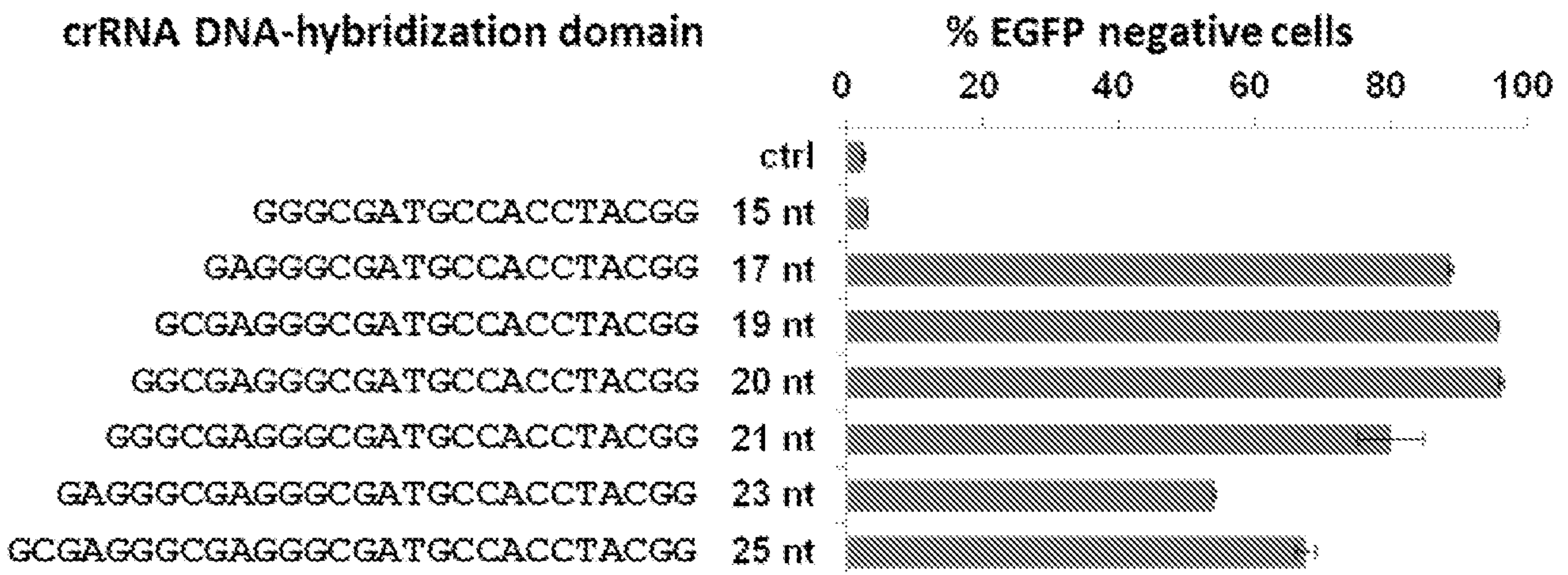


FIG. 2H

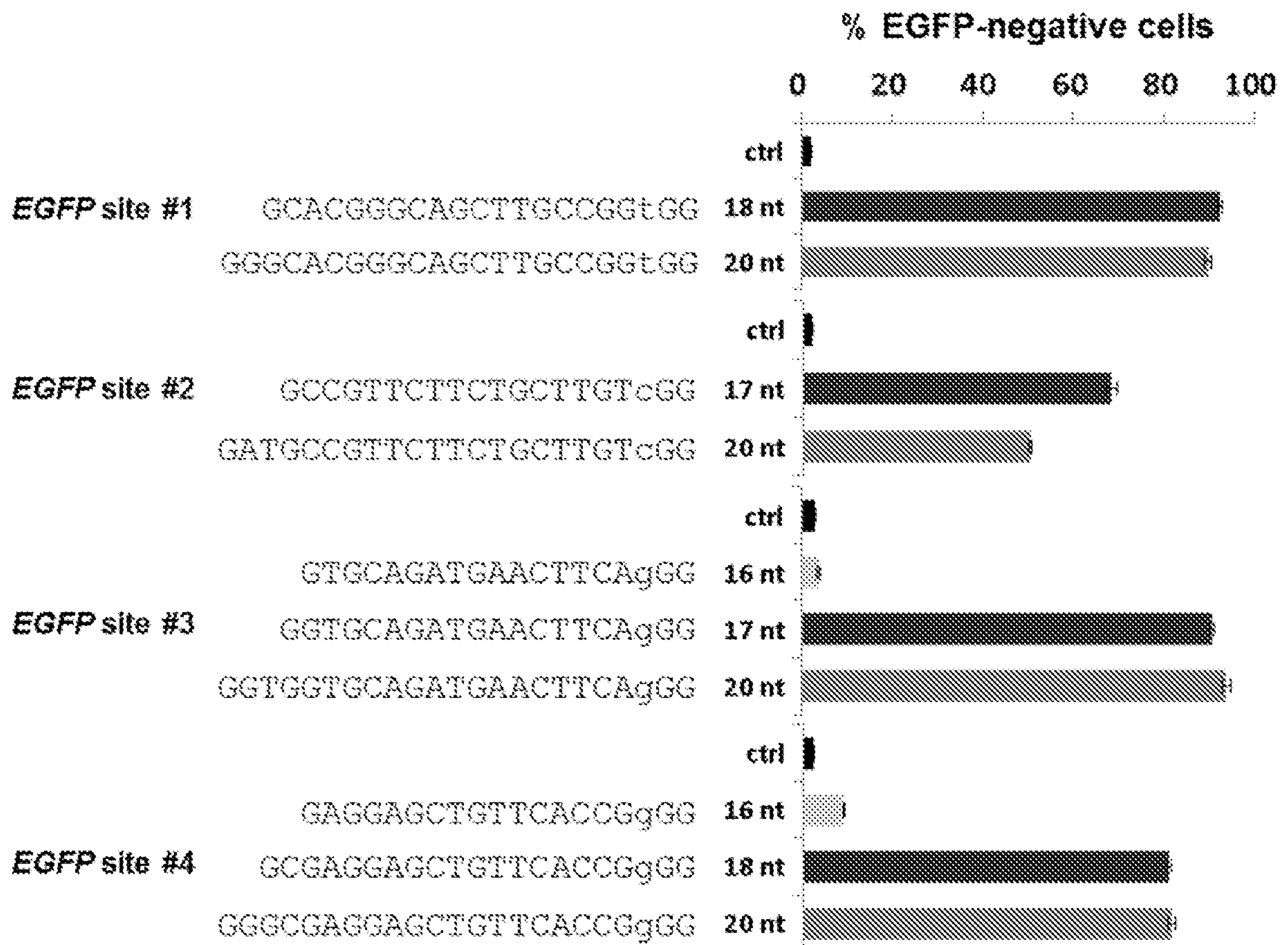


FIG. 3A

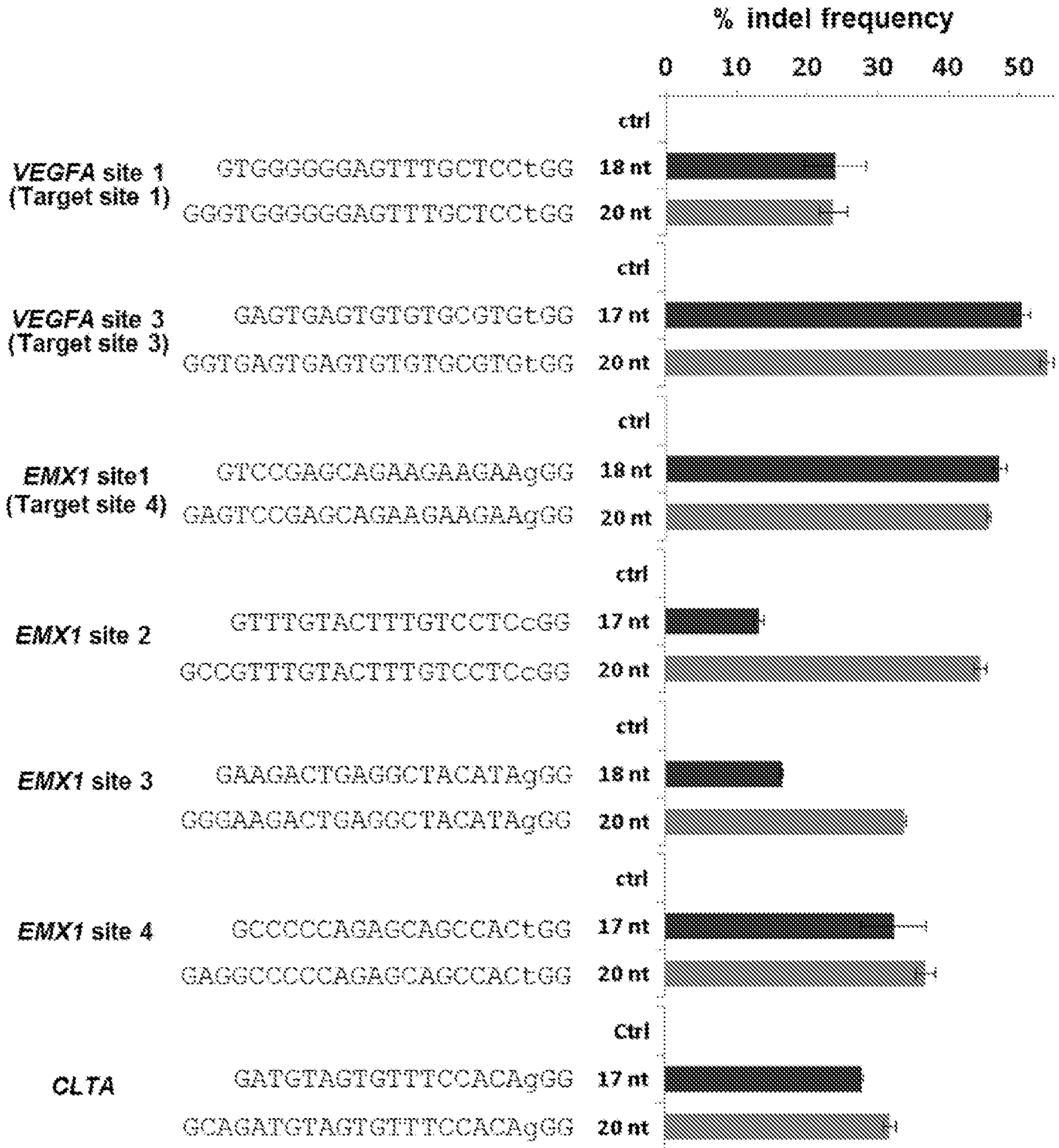


FIG. 3B

***EMX1* truncated gRNA**

```
GAAGCTGGAGGAGGAAGGGCCTGAGTCCGAGCAGAAGAAGAAgGGCTCCCATCACATCAACCGGTGG wild-type x24
GAAGCTGGAGGAGGA--------------------------------------------------->  Δ365
<-------------------------------------------------------TCAACCGGTGG Δ181
GAAGCTGGAGGAGGAAGG------------------------------------------------>  Δ138
<----------------------------------------------------------------->  Δ126
<-----------------------------------------GGGCTCCCATCACATCAACCGGTGG Δ114
GAAGCTGGAGGAGGAAGGGCCTGA------------------------------------------>  Δ101
GAAGCTGGAGGA-----------------------------------------------------GG Δ53
GAAGCTGGAGGAGGAAGGG----------------------------CCCATCACATCAACCGGTGG Δ28
GAAGCTGGAGGAGGAAGGGC---------------------------TCGCACACATCAACCGGTGG Δ27
GAAGCTGGAGGAGGAAGGGC-------------------------CTTCCATCACATCAACCGGTGG Δ25
GAAGCTGGAGGAGGAAGGGCCTGAG---------------------TCCCATCACATCAACCGGTGG Δ21 x2
GAAGCTGGAGGAGGAAGGGCCTGAGTCCGAG---------------TCCCATCACATCAACCGGTGG Δ15
GAAGCTGGAGGAGGAAGGGCCTGAGTCCGAGCAGAAG---------TCCCATCACATCAACCGGTGG Δ9
GAAGCTGGAGGAGGAAGGGCCTGAGTCCTGCCGTTTGTAG--------CCATCACATCAACCGGTGG Δ8
GAAGCTGGAGGAGGAAGGGCCTGAGTCCGAGCAGAAGA------GCTCCCATCACATCAACCGGTGG Δ6
GAAGCTGGAGGAGGAAGGGCCTGAGTCCGAGCAGAAGAA------CTCCCATCACATCAACCGGTGG Δ6
GAAGCTGGAGGAGGAAGGGCCTGAGTCCGAGC---AGAAGAAGGGCTCCCATCACATCAACCGGTGG Δ3  x3
GAAGCTGGAGGAGGAAGGGCCTGAGTCCGAGCAGAAGA--AAGGGCTCCCATCACATCAACCGGTGG Δ2
GAAGCTGGAGGAGGAAGGGCCTGAGTCCGAGCAGAAGAACAGAAGGGCTCCCATCACATCAACCGGT +2
```

***EMX1* full-length gRNA**

```
GAAGCTGGAGGAGGAAGGGCCTGAGTCCGAGCAGAAGAAGAAgGGCTCCCATCACATCAACCGGTGG wild-type x35
GAAGCTGGAGGAGGAAGGGCCTGAGTCCGAG----------------------------------->  Δ202
<----------------------------------------------------------------->  Δ115
GAA--------------------------------------------------------------->  Δ94
<----------------------------------------------------------------->  Δ79
GAAGCTGGAGG------------------------------------------------------->  Δ72
GAAGCTGGA--------------------------------------------------------GG Δ56
GAAGCTGGAGGAGGAAGGGCCTGA---------------------------------------GTGG Δ39
GAAGCTGGAGGAG--------------------------GAAGGGCTCCCATCACATCAACCGGTGG Δ26 x2
GAAGCTGGAGGAGGAAGGGCCTGAGT----------------------CCATCACATCAACCGGTGG Δ22
GAAGCTGGAGGAGGAAGGGCCTGAG---------------------TCCCATCACATCAACCGGTGG Δ21 x3
GAAGCTGGAGGAGGAAGGGCCTGAGTCCGAG------------------CATCACATCAACCGGTGG Δ18
GAAGCTGGAGGAGGAAGGGCCTGAGTCCGA--------------GCTCCCATCACATCAACCGGTGG Δ14
GAAGCTGGAGGAGGAAGGGCCTGAGTCCGAGC------AGAAGGGCTCCCATCACATCAACCGGTGG Δ6  x3
GAAGCTGGAGGAGGAAGGGCCTGAGTCCGAGC---AGAAGAAGGGCTCCCATCACATCAACCGGTGG Δ3  x3
GAAGCTGGAGGAGGAAGGGCCTGAGTCCGAGCAGA--AAGAAGGGCTCCCATCACATCAACCGGTGG Δ2  x2
GAAGCTGGAGGAGGAAGGGCCTGAGTCCGAGCAGAAGAACAGAAGGGCTCCCATCACATCAACCGGT +2
```

FIG. 3C

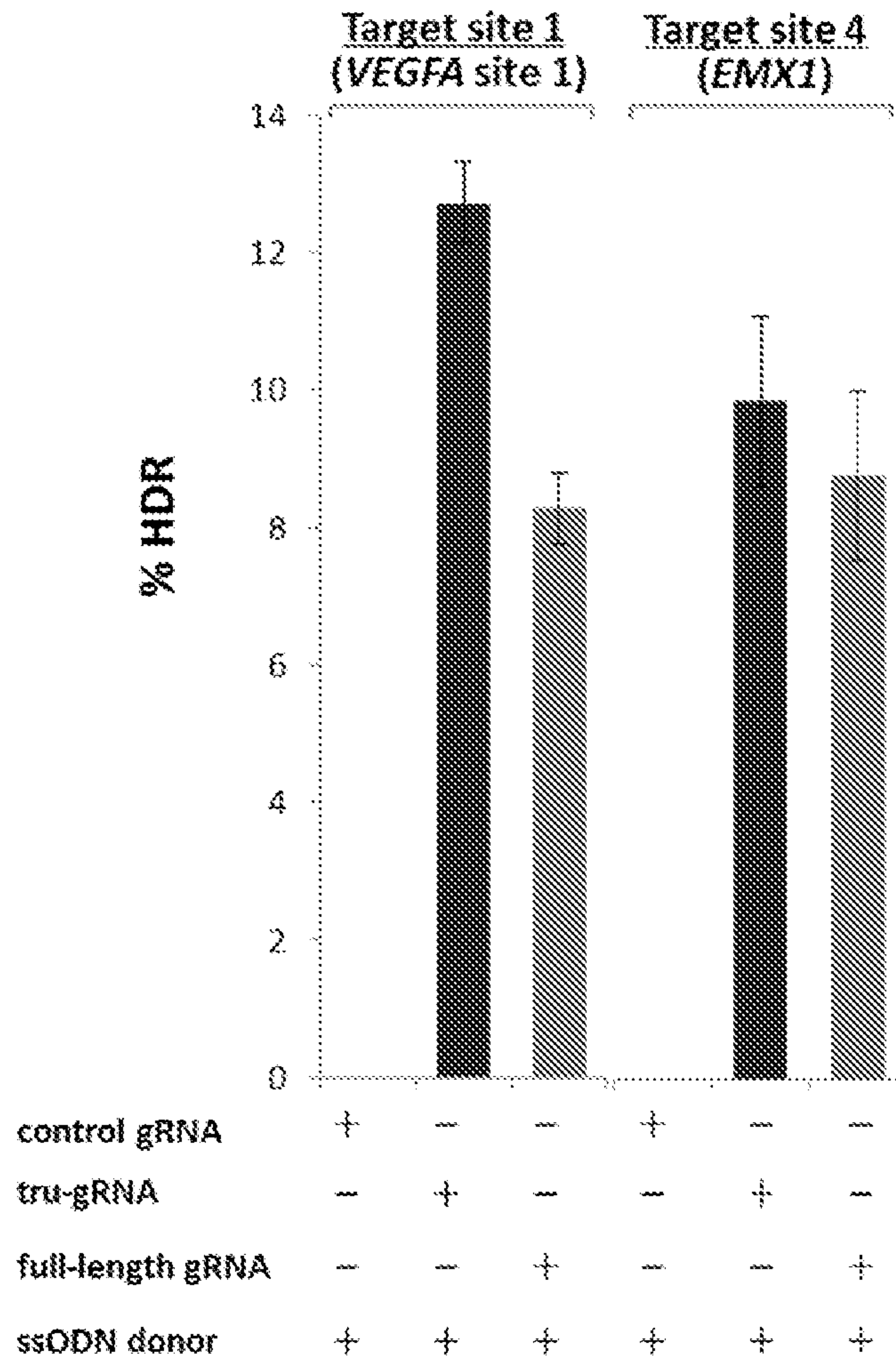
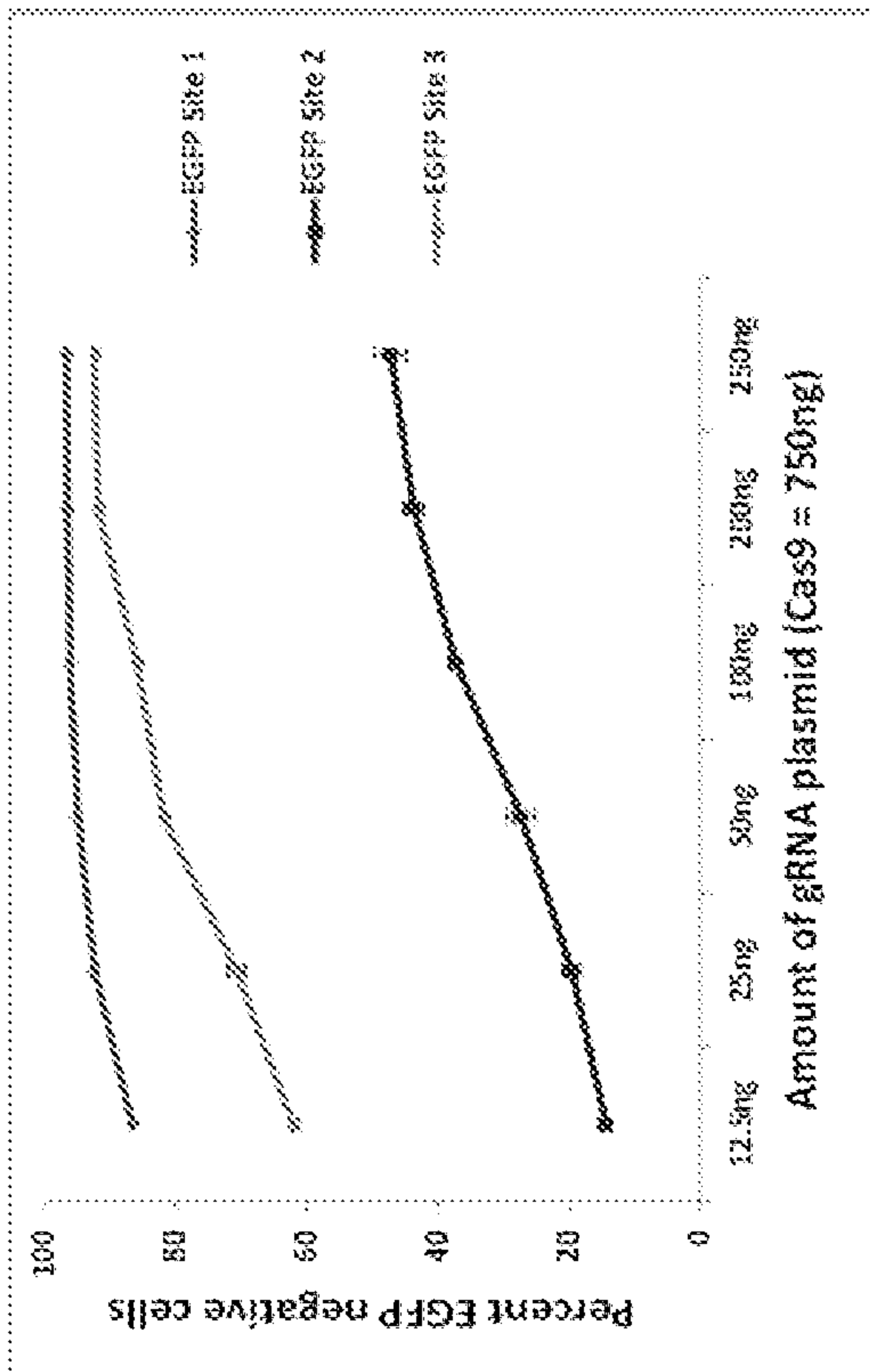


FIG. 3D



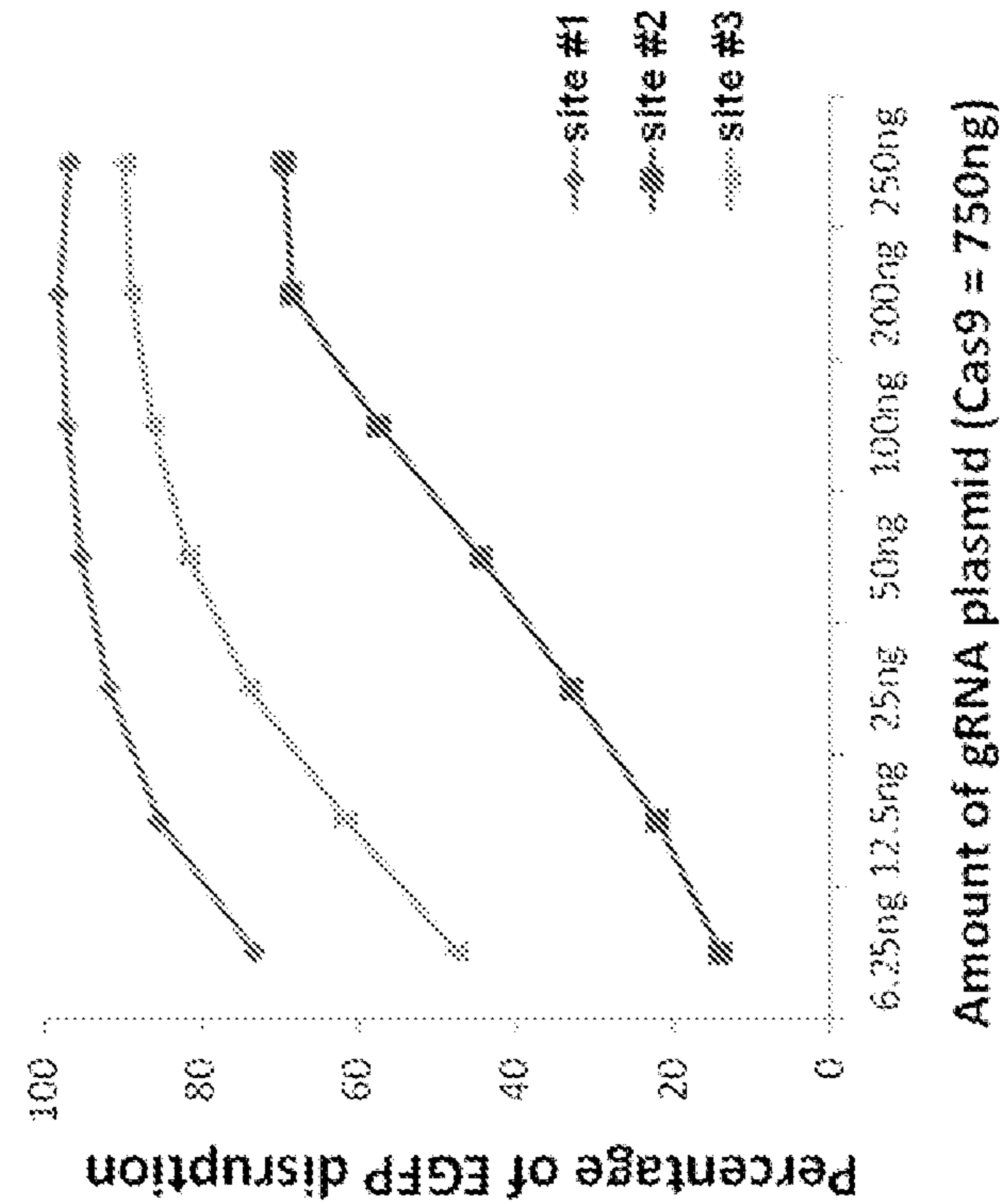
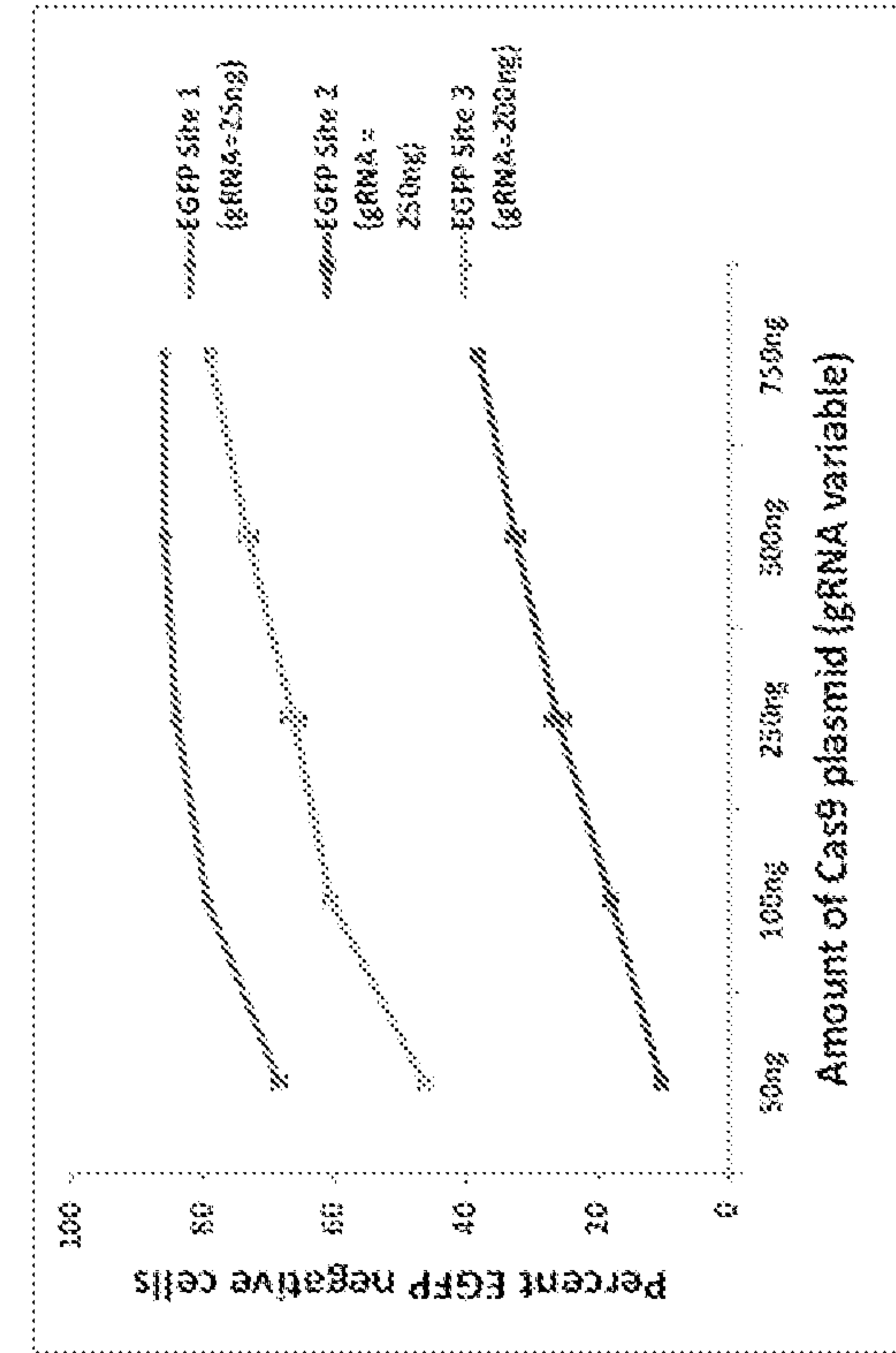


Figure 3E



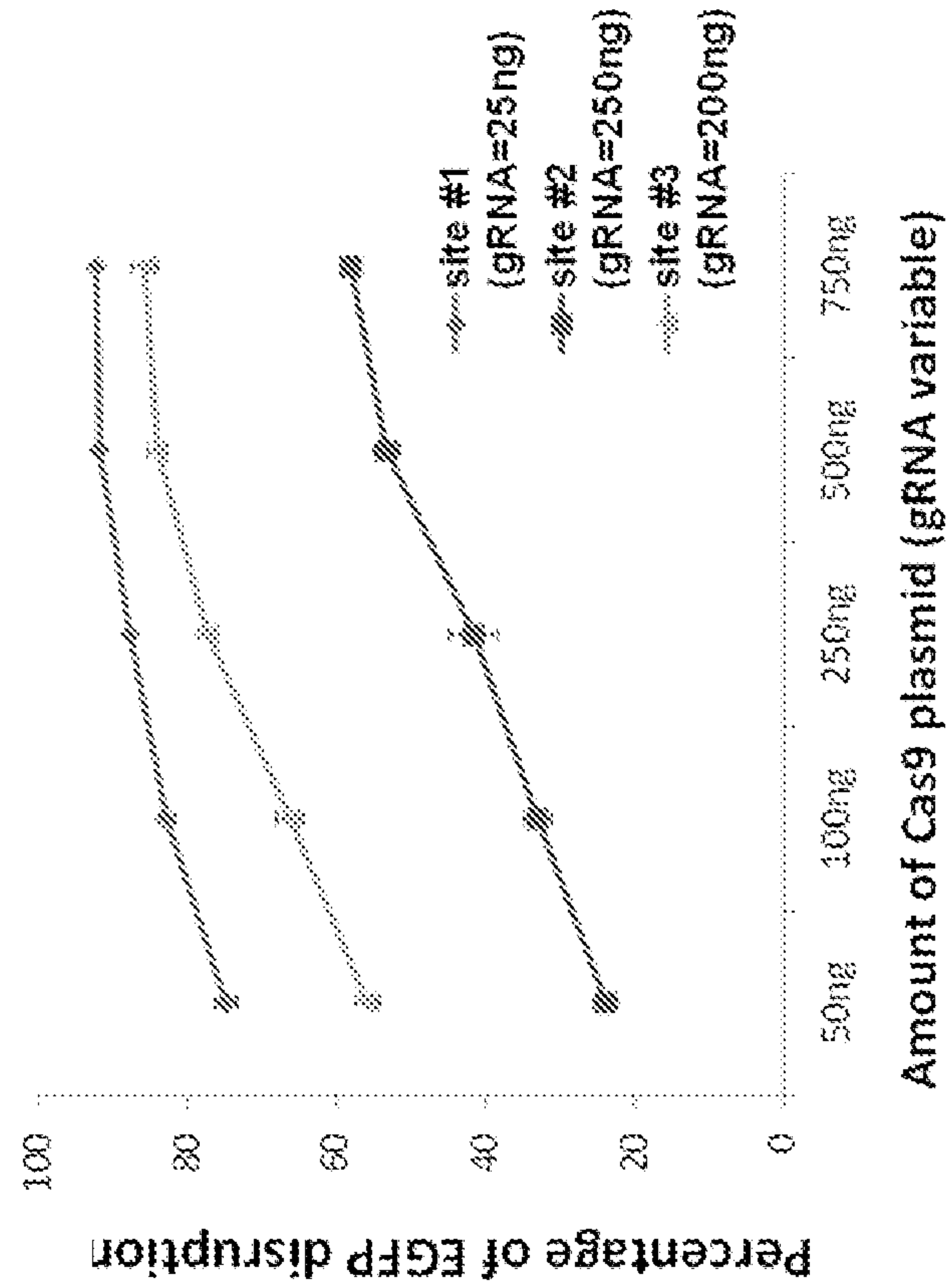


Figure 3F

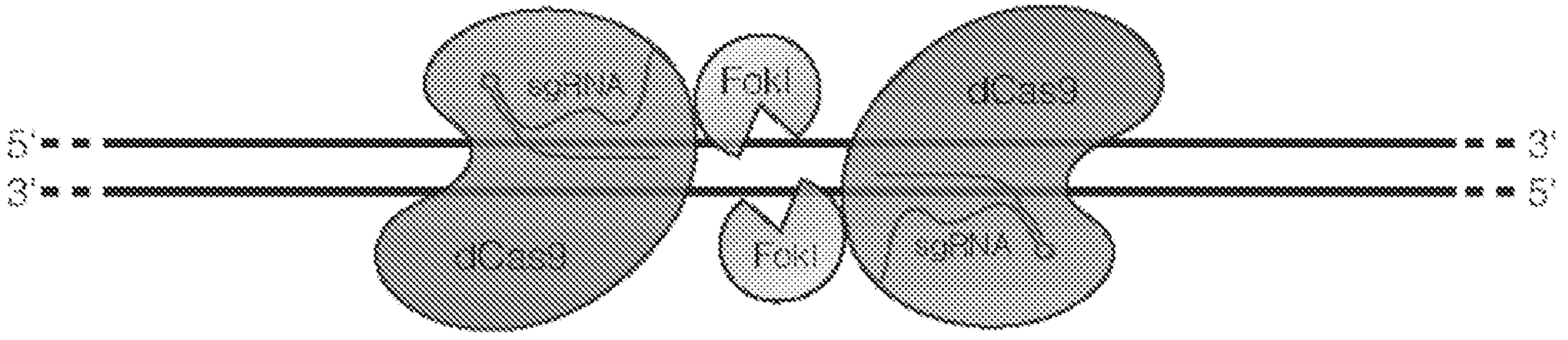


Figure 4A

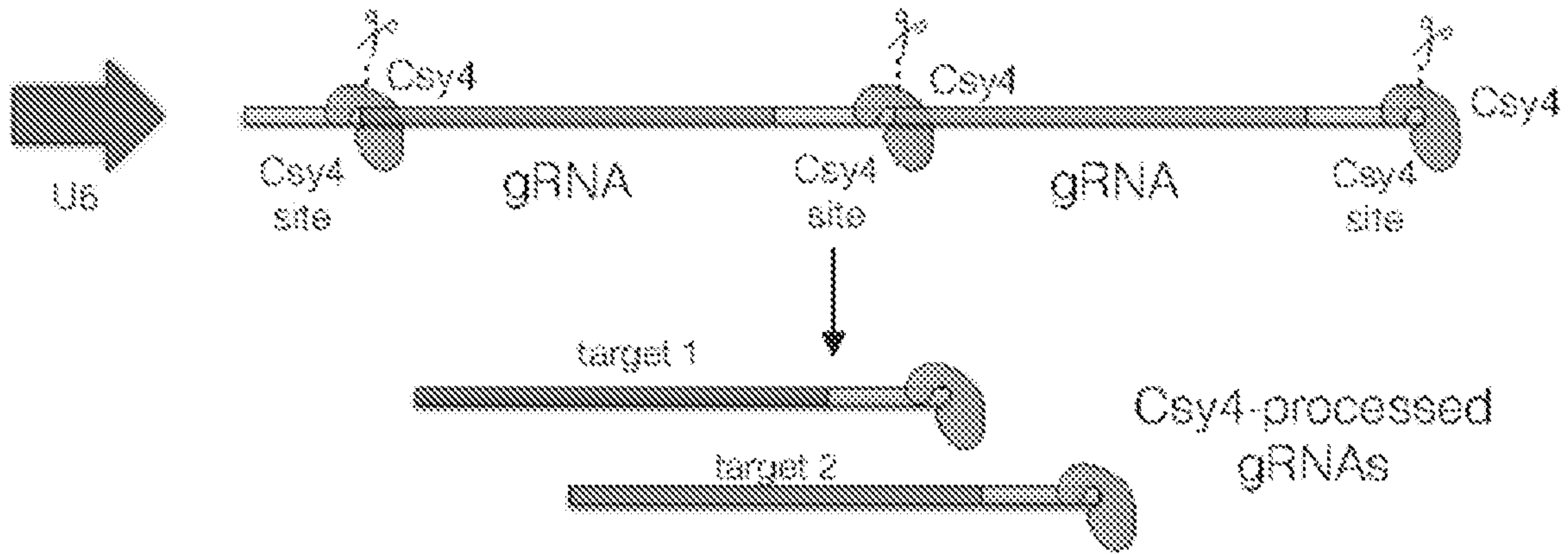


Figure 4B

EGFP

FIGURE 4C

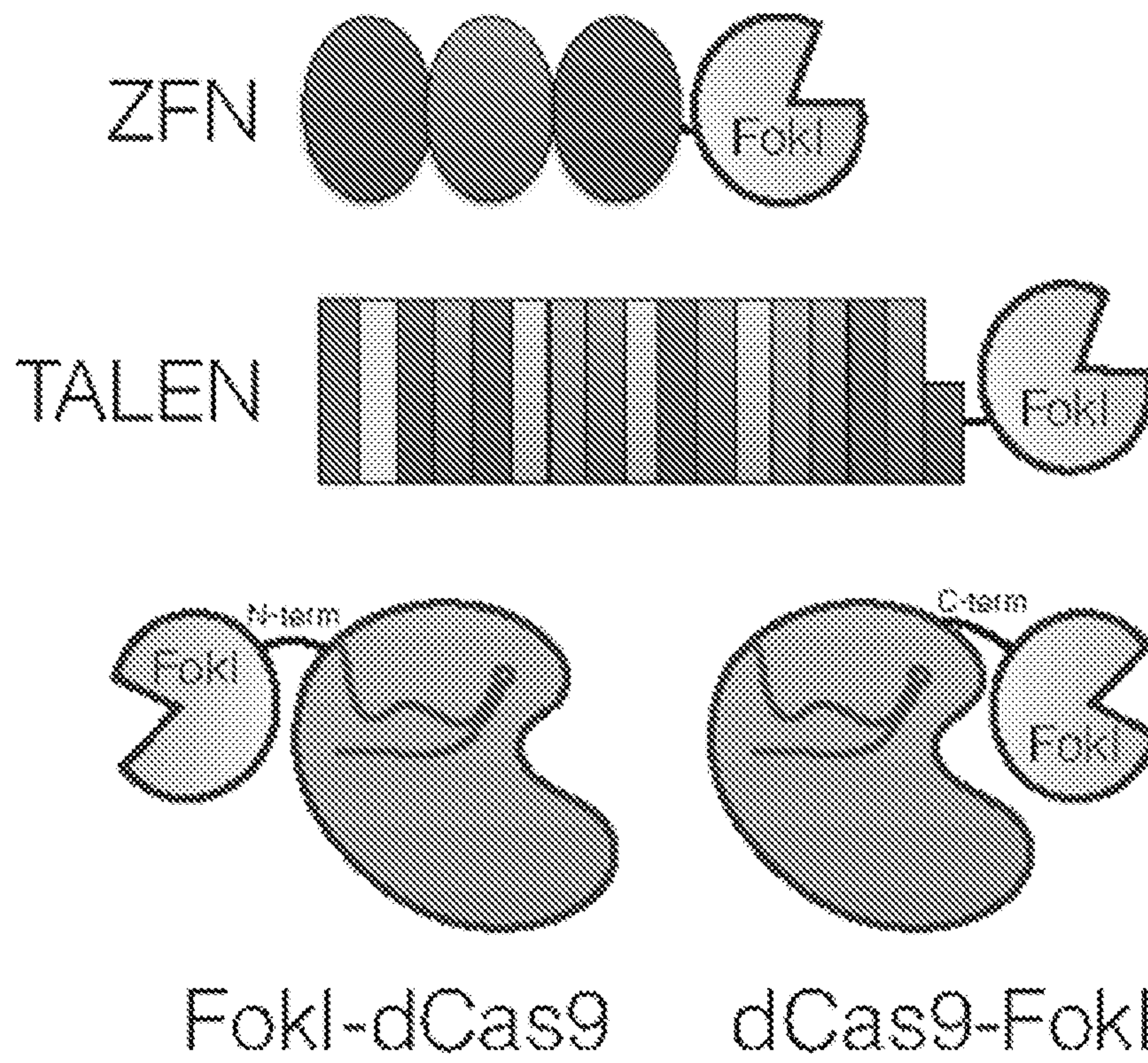
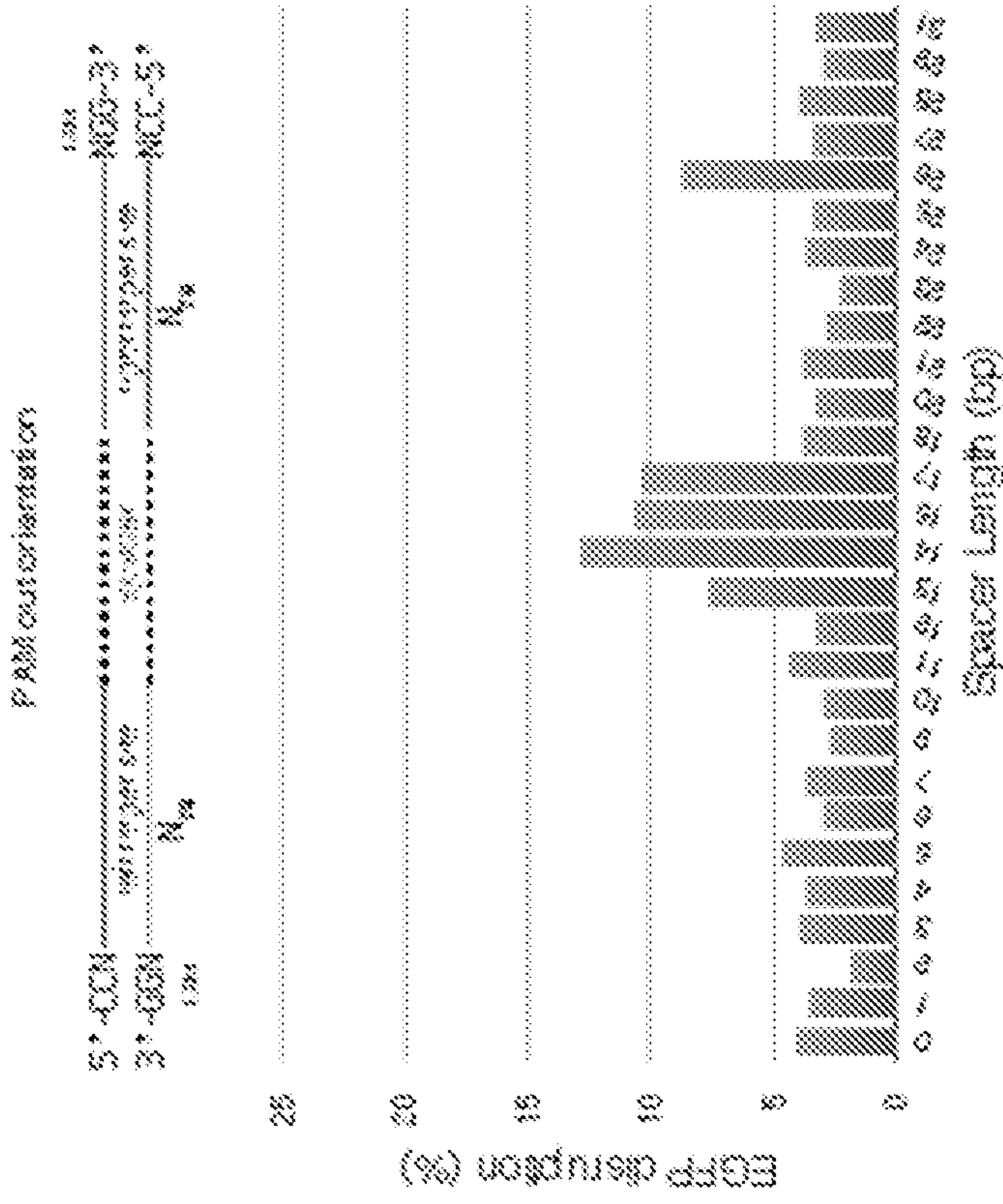


FIG. 5A



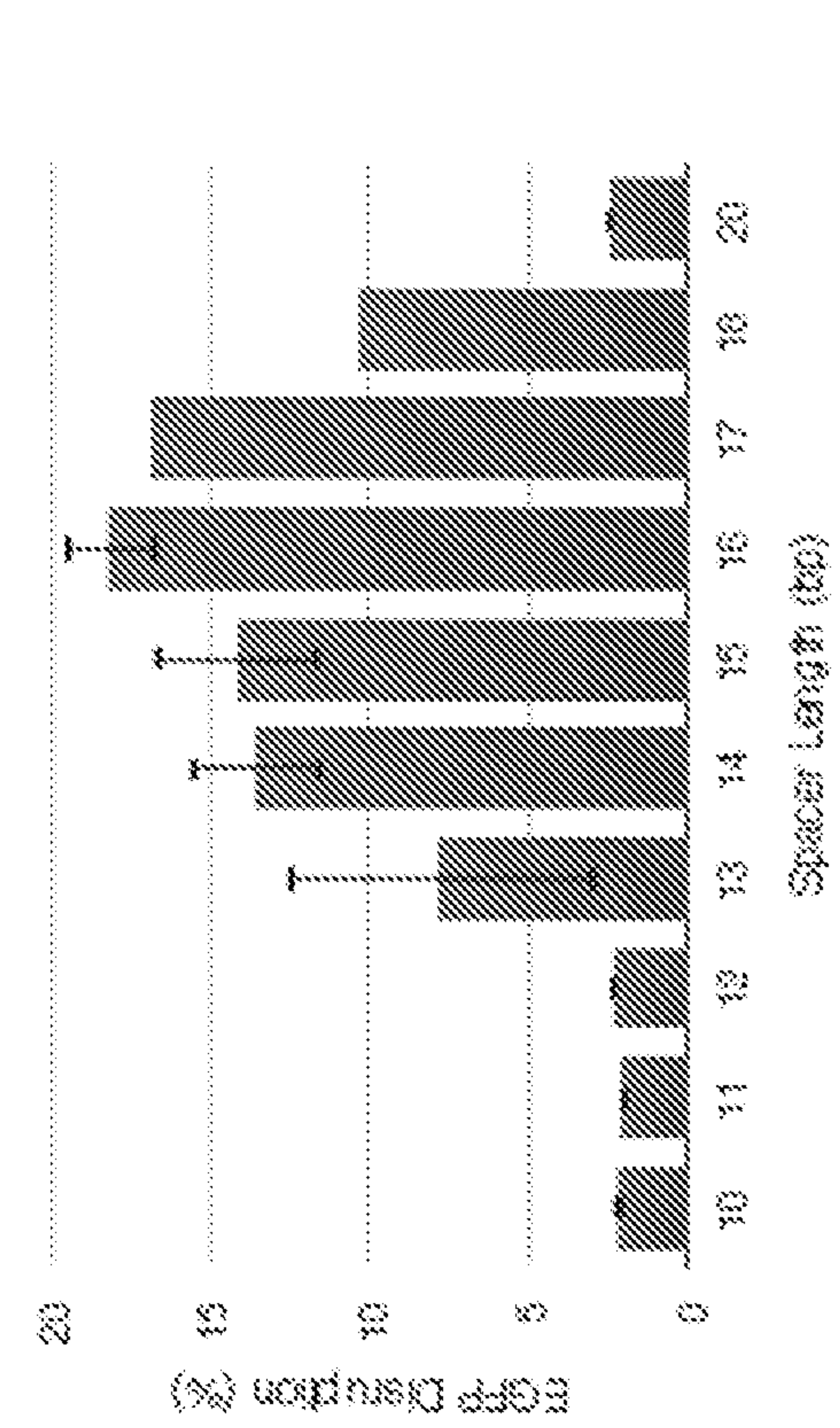


FIG. 5B

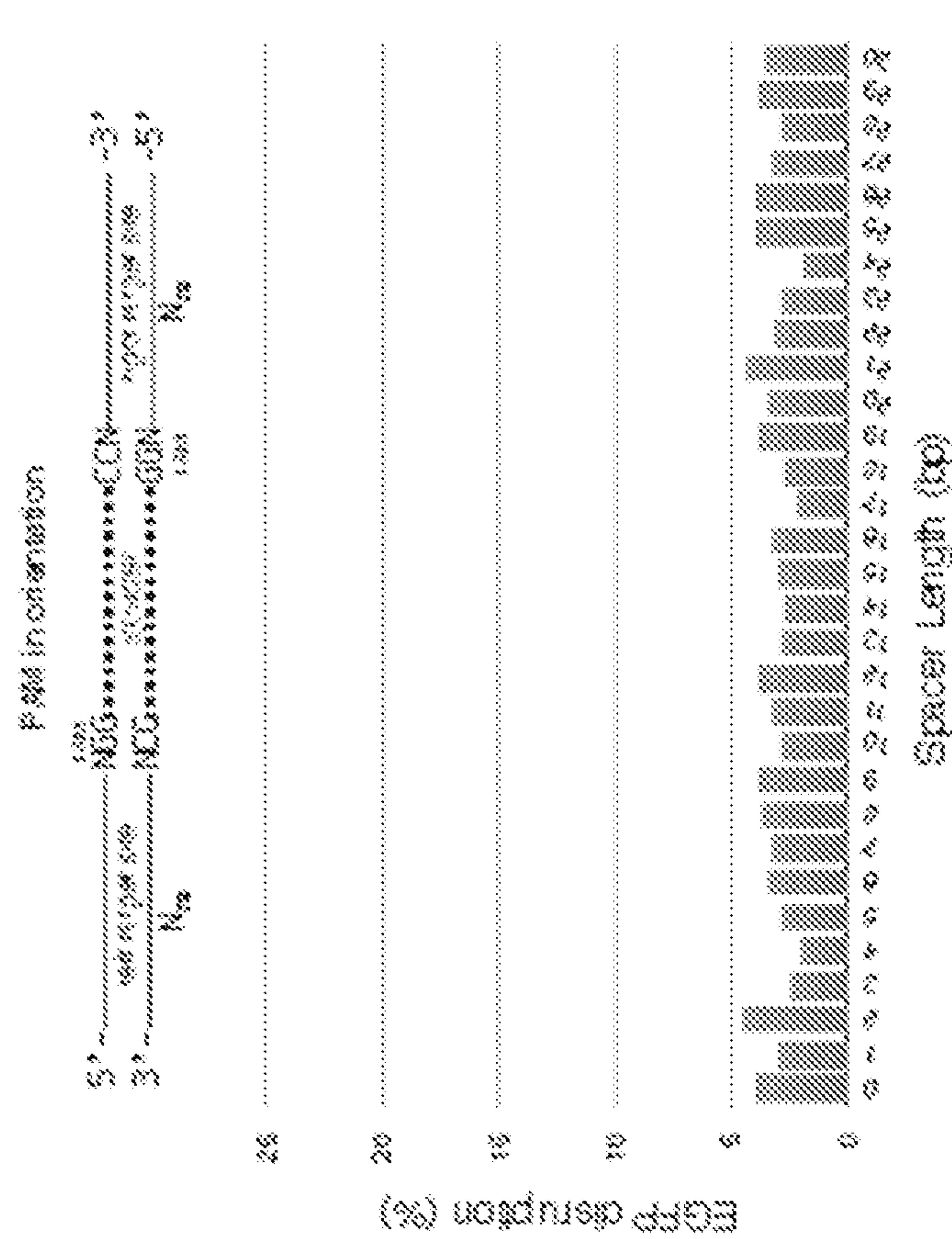


FIG. 5C

EGFP Disruption (%)

Spacer Length (bp)

FIG. 5D

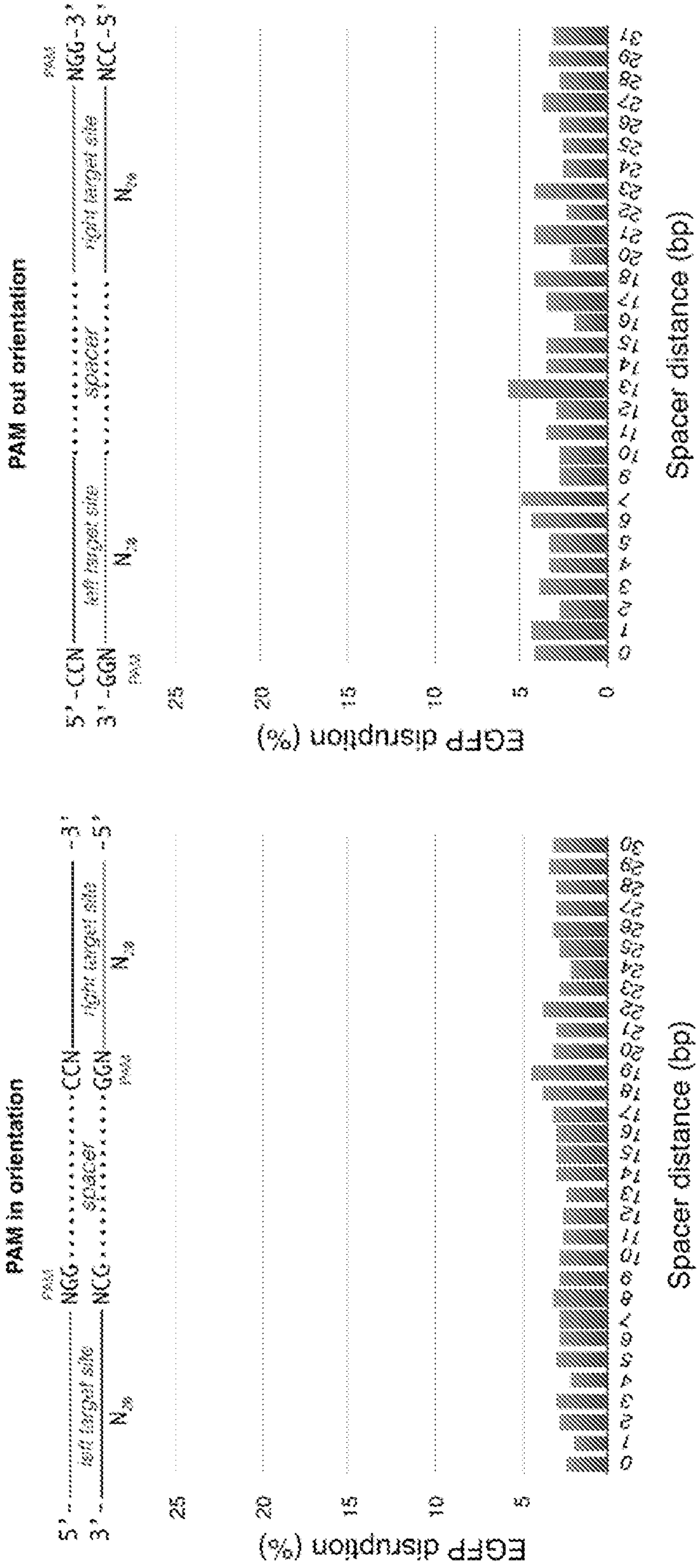


FIG. 5E

FIG. 5F-1

APC

Mutations in 9 of 23 sequences ≈ 39.1%

```
CAGGAGACGAAGAGCCCGGGCGGCGCTCGTACTTCTGGCCACTGGGCGAGCGTCTGGCAGGTGAGTGAGGCTGCAGGCATTGACGTCTCCTC WT
CAGGAGACGAAGAGCCCGGGCGGCGCTCGTACTTCTGGCCACTGGcgaAGCGTCTGGCAGGTGAGTGAGGCTGCAGGCATTGACGTCTCCTC Δ0 (Δ3 +3)
CAGGAGACGAAGAGCCCGGGCGGCGCTCGTACTTCTGGCC------------TCTGGCAGGTGAGTGAGGCTGCAGGCATTGACGTCTCCTC Δ12
CAGGAGACGAAGAGCCCGGGCGGCGCTCGTACTTCT------------AGCGTCTGGCAGGTGAGTGAGGCTGCAGGCATTGACGTCTCCTC Δ12
CAGGAGACGAAGAGCCCGGGCGGCGCTCGTACTTC----------------GTCTGGCAGGTGAGTGAGGCTGCAGGCATTGACGTCTCCTC Δ16
CAGGAGACGAAGAGCCCGGGCGGCGCTCGTACTTCTGGCCACTG-----------------TGAGTGAGGCTGCAGGCATTGACGTCTCCTC Δ17
CAGGAGACGAAGAGCCCGGGCGGCGCTCGTACTTCTGGC-------------------AGGTGAGTGAGGCTGCAGGCATTGACGTCTCCTC Δ19
CAGGAGACGAAGAGCCCGGGCGGCGCTCGT-----------------------CTGGCAGGTGAGTGAGGCTGCAGGCATTGACGTCTCCTC Δ23
CAGGAGACGAAGAGCC----------------------------------CGTCTGGCAGGTGAGTGAGGCTGCAGGCATTGACGTCTCCTC Δ34
ACCGGCGGCAGCAGG-------------------/ /-----------------------TGAGTGAGGCTGCAGGCATTGACGTCTCCTC Δ73
```

BRCA1

Mutations in 7 of 86 sequences ≈ 8.1%

```
TGGGAGAGTGGATTTCCGAAGCTGACAGATGGGTATTCTTTGAGGGGGGGTAGGGGCGGAACCTGAGAGGCGTAAGGCGTTGTGAACCCTGG WT
TGGGAGAGTGGATTTCCGAAGCTGACAGATGGGTATTCTTTGA------GTAGGGGCGGAACCTGAGAGGCGTAAGGCGTTGTGAACCCTGG Δ6
TGGGAGAGTGGATTTCCGAAGCTGACAGATGGGTAT--------GGGGGGTAGGGGCGGAACCTGAGAGGCGTAAGGCGTTGTGAACCCTGG Δ8
TGGGAGAGTGGATTTCCGAAGCTGACAGATGGGTATTCTT--------GGTAGGGGCGGAACCTGAGAGGCGTAAGGCGTTGTGAACCCTGG Δ8
TGGGAGAGTGGATTTCCGAAGCTGACAGATGGGTATTCTTTGAGGGGG----------GAACCTGAGAGGCGTAAGGCGTTGTGAACCCTGG Δ10
TGGGAGAGTGGATTTCCGAAGCTGACAGATGGGTATTCTTTG------------GGCGGAACCTGAGAGGCGTAAGGCGTTGTGAACCCTGG Δ12
TGGGAGAGTGGATTTCCGAAGCTGACAGATGGG----------------------GCGGAACCTGAGAGGCGTAAGGCGTTGTGAACCCTGG Δ22 (2x)
```

EMXI

Mutations in 2 of 91 sequences ≈ 2.2%

```
CAAGCTGGACTCTGGCCACTCCCTGGCCAGGCTTTGGGGAGGCCTGGAGTCATGGCCCCACAGGGCTTGAAGCCCGGGGGGCCGCCATTGAC WT
CAAGCTGGACTCTGGCCACTCCCTGGCCAGGCTTT----------------ATGGCCCCACAGGGCTTGAAGCCCGGGGGGCCGCCATTGAC Δ16 (2x)
```

FANCF site 1

Name/name Mutations in 9 of 29 sequences ≈ 31%

```
AGAGAGTCGCCGTCTCCAAGGTGAAAGCGGAAGTAGGGCCTTCGCGCACCTCATGGAATCCCTTCTGCAGCACCTGGATCGCTTTTCCGAGC WT
AGAGAGTCGCCGTCTCCAAGGTGAAAGCGGAAGTAG--------CGCACCTCATGGAATCCCTTCTGCAGCACCTGGATCGCTTTTCCGAGC Δ8
AGAGAGTCGCCGTCTCCAAGGTGAAAGCGGAAGTAGGGCC---------CTCATGGAATCCCTTCTGCAGCACCTGGATCGCTTTTCCGAGC Δ9
AGAGAGTCGCCGTCTCCAAGGTGAAAGCGGAAGTAGGcCCT----------CATGGAATCCCTTCTGCAGCACCTGGATCGCTTTTCCGAGC Δ10
AGAGAGTCGCCGTCTCCAAGGTGAAAGCGGAAGTAGGGCCT----------CATGGAATCCCTTCTGCAGCACCTGGATCGCTTTTCCGAGC Δ10
AGAGAGTCGCCGTCTCCAAGGTGAAAGCGGAAGTAG----------CACCTCATGGAATCCCTTCTGCAGCACCTGGATCGCTTTTCCGAGC Δ10
AGAGAGTCGCCGTCTCCAAGGTGAAAGCGGAAa-----------CGCACCTCATGGAATCCCTTCTGCAGCACCTGGATCGCTTTTCCGAGC Δ11 (Δ12 +1)
AGAGAGTCGCCGTCTCCAAGGTGAAAGCGGAAGTAGGGCCTTCGCGCA--------------TTCTGCAGCACCTGGATCGCTTTTCCGAGC Δ14
AGAGAGTCGCCGTCTCCAAGGTGAAAGCGGAAG---------------CCTCATGGAATCCCTTCTGCAGCACCTGGATCGCTTTTCCGAGC Δ15
AGAGAGTCGCCGTCTCCAAG-------------------------------CATGGAATCCCTTCTGCAGCACCTGGATCGCTTTTCCGAGC Δ31
```

FIG. 5F-2

FANCF site 2

Mutations in 13 of 71 sequences ≈ 18.3%

```
TTTGGTCGGCATGGCCCCATTCGCACGGCTCTGGAGCGGCGGCTGCACAACCAGTGGAGGCAAGAGGGCGGCTTTGGGCGGGGTCCAGTTCC WT
TTTGGTCGGCATGGCCCCATTCGCACGGCTCTGGAGCGGCGGCG---CAACCAGTGGAGGCAAGAGGGCGGCTTTGGGCGGGGTCCAGTTCC Δ3
TTTGGTCGGCATGGCCCCATTCGCACGGCTCTGGAGCGGCGGCT----AACCAGTGGAGGCAAGAGGGCGGCTTTGGGCGGGGTCCAGTTCC Δ4
TTTGGTCGGCATGGCCCCATTCGCACGGCTCTGGAGCGGCGGC--------CAGTGGAGGCAAGAGGGCGGCTTTGGGCGGGGTCCAGTTCC Δ8
TTTGGTCGGCATGGCCCCATTCGCACGGCTCTGGAGCGGCGGCTG---------TGGAGGCAAGAGGGCGGCTTTGGGCGGGGTCCAGTTCC Δ9
TTTGGTCGGCATGGCCCCATTCGCACGGCTCTGGAGCGGCGGCTG-----------GAGGCAAGAGGGCGGCTTTGGGCGGGGTCCAGTTCC Δ11
TTTGGTCGGCATGGCCCCATTCGCACGGCTCTGGAGCGGCGGga-----------GAGGCAAGAGGGCGGCTTTGGGCGGGGTCCAGTTCC Δ12 (Δ14 +2)
TTTGGTCGGCATGGCCCCATTCGCACGGCTCTGGAGCGGCGGC-------------GAGGCAAGAGGGCGGCTTTGGGCGGGGTCCAGTTCC Δ13
TTTGGTCGGCATGGCCCCATTCGCACGGCTCTGGAGCGGCGGCTG--------------GCAAGAGGGCGGCTTTGGGCGGGGTCCAGTTCC Δ14
TTTGGTCGGCATGGCCCCATTCGCACGGCTCTGGAGCGGCGGC---------------GGCAAGAGGGCGGCTTTGGGCGGGGTCCAGTTCC Δ15
TTTGGTCGGCATGGCCCCATTCGCACGGCTCTGGAG------------------TGGAGGCAAGAGGGCGGCTTTGGGCGGGGTCCAGTTCC Δ18
TTTGGTCGGCATGGCCCCATTCGCACGGCTCTGGAG-----------------------GCAAGAGGGCGGCTTTGGGCGGGGTCCAGTTCC Δ23
TTTGGTCGGCATGGCCCC----------------------------------/ /-----------------------------GTCACTGT Δ97

TTTGGTCGGCATGGCCCCATTCGCACGGCTCTGGAGCGGCGGCacactattgcaatgaaaataaatttcctttattagctagaagtgagatc +168(Δ9 +177)
Tgaaggggcttcatgatgtccccataattttggcagagggaaaaagatctcagtggtatttgtgagccagggcattggccacaccagccac
cacctttgataggcagcctgcacctgaggagtgcgAGTGGAGGCAAGAGGGCGGCTTTGGGCGGGGTCCAGTTCC
```

GLI1

Mutations in 4 of 66 sequences ≈ 6.1%

```
GTTCAACTCGATGACCCCACCACCAATCAGTAGCTATGGCGAGCCCTGCTGTCTCCGGCCCCTCCCCAGTCAGGGGGCCCCCAGTGTGGGGA WT
GTTCAACTCGATGACCCCACCACCAATCAGTAGCTAT-------CCTGCTGTCTCCGGCCCCTCCCCAGTCAGGGGGCCCCCAGTGTGGGGA Δ7
GTTCAACTCGATGACCCCACCACCAATCAGTAGCTATG-------------TCTCCGGCCCCTCCCCAGTCAGGGGGCCCCCAGTGTGGGGA Δ13
GTTCAACTCGATGACCCCACCACCAATCAGTAGCTATGGCGAGCCC-----------------CCCCAGTCAGGGGGCCCCCAGTGTGGGGA Δ17
GTTCAACTCGATGACCCCACCACCAA------------------CCTGCTGTCTCCGGCCCCTCCCCAGTCAGGGGGCCCCCAGTGTGGGGA Δ18
```

MLH1

Mutations in 4 of 20 sequences ≈ 20%

```
TGGCTGAAGGCACTTCCGTTGAGCATCTAGACGTTTCCTTGGCTCTTCTGGCGCCAAAATGTCGTTCGTGGCAGGGGTTATTCGGCGGCTGG WT
TGGCTGAAGGCACTTCCGTTGAGCATCTAGACGTTTCCT---------TGGCGCCAAAATGTCGTTCGTGGCAGGGGTTATTCGGCGGCTGG Δ9
TGGCTGAAGGCACTTCCGTTGAGCATCTA----------------TTCTGGCGCCAAAATGTCGTTCGTGGCAGGGGTTATTCGGCGGCTGG Δ16
TGGCTGAAGGCACTTCCGTTGAGCATCTAGACGTTTCCTTGG----------------ATGTCGTTCGTGGCAGGGGTTATTCGGCGGCTGG Δ16
TGGCTGAAGGCACTTCCGTTGAGCATCTAGACGTTT------------------------GTCGTTCGTGGCAGGGGTTATTCGGCGGCTGG Δ24
```

FIG. 5F-3

RARA

Mutations in 4 of 86 sequences ≈ 4.7%

```
CCCTTCTGACTGTGGCCGCTTGGCATGGCCAGCAACAGCAGCTCCTGCCCGACACCTGGGGGCGGGCACCTCAATGGGTACCCGGTGCCTCC WT
CCCTTCTGACTGTGGCCGCTTGGCATGGCCAGCAACA------CCTGCCCGACACCTGGGGGCGGGCACCTCAATGGGTACCCGGTGCCTCC Δ6
CCCTTCTGACTGTGGCCGCTTGGCATGGCCAGCAACA-------CTGCCCGACACCTGGGGGCGGGCACCTCAATGGGTACCCGGTGCCTCC Δ7
CCCTTCTGACTGTGGCCGCTTGGCATGGCCAGCA------------GCCCGACACCTGGGGGCGGGCACCTCAATGGGTACCCGGTGCCTCC Δ12
CCCTTCTGACTGTGGCCGCTTGGCATGGCCAG--------------GCCCGACACCTGGGGGCGGGCACCTCAATGGGTACCCGGTGCCTCC Δ14
```

VEGFA site 1

Mutations in 19 of 74 sequences ≈ 25.7%

```
TCAGAAATAGGGGGTCCAGGAGCAAACTCCCCCCACCCCCTTTCCAAAGCCCATTCCCTCTTTAGCCAGAGCCGGGGTGTGCAGACGGCAGT WT
TCAGAAATAGGGGGTCCAGGAGCAAACTCCCCCCACCCCCTTTCCAAAaaGCCCATTCCCTCTTTAGCCAGAGCCGGGGTGTGCAGACGGCA +2
TCAGAAATAGGGGGTCCAGGAGCAAACTCCCCCCACCCCCTTTCC---GCCCATTCCCTCTTTAGCCAGAGCCGGGGTGTGCAGACGGCAGT Δ3
TCAGAAATAGGGGGTCCAGGAGCAAACTCCCCCCACCCCCTT-----AGCCCATTCCCTCTTTAGCCAGAGCCGGGGTGTGCAGACGGCAGT Δ5
TCAGAAATAGGGGGTCCAGGAGCAAACTCCCCCCACCCCC-----AAAGCCCATTCCCTCTTTAGCCAGAGCCGGGGTGTGCAGACGGCAGT Δ5
TCAGAAATAGGGGGTCCAGGAGCAAACTCCCCCCACCCC------AAAGCCCATTCCCTCTTTAGCCAGAGCCGGGGTGTGCAGACGGCAGT Δ6  (2x)
TCAGAAATAGGGGGTCCAGGAGCAAACTCCCCCCACCCCCTTTCCA-------TTCCCTCTTTAGCCAGAGCCGGGGTGTGCAGACGGCAGT Δ7
TCAGAAATAGGGGGTCCAGGAGCAAACTCCCCCCACCCCCTTTC--------ATTCCCTCTTTAGCCAGAGCCGGGGTGTGCAGACGGCAGT Δ8
TCAGAAATAGGGGGTCCAGGAGCAAACTCCCCCCAC---------AAAGCCCATTCCCTCTTTAGCCAGAGCCGGGGTGTGCAGACGGCAGT Δ9  (2x)
TCAGAAATAGGGGGTCCAGGAGCAAACTCCCCCCACCC--------------ATTCCCTCTTTAGCCAGAGCCGGGGTGTGCAGACGGCAGT Δ14 (3x)
TCAGAAATAGGGGGTCCAGGAGCAAAC-----------------CAAAGCCCATTCCCTCTTTAGCCAGAGCCGGGGTGTGCAGACGGCAGT Δ17 (2x)
TCAGAAATAGGGGGTCCAGGAGCAAACTCCCCCCACCCCCT------------------CTTTAGCCAGAGCCGGGGTGTGCAGACGGCAGT Δ18
TCAGAAATAGGGGGTCCAGGAGCAAACTCCCCCCACCCCCTTT--------------------AGCCAGAGCCGGGGTGTGCAGACGGCAGT Δ20
TCAGAAATAGGGGGTCCAGG------------------------CAAAGCCCATTCCCTCTTTAGCCAGAGCCGGGGTGTGCAGACGGCAGT Δ24
TCAGAAATAGGGGGTCCAGGAGC-----------------------------------TCTTTAGCCAGAGCCGGGGTGTGCAGACGGCAGT Δ35
```

VEGF site 2

Mutations in 26 of 80 sequences ≈ 32.5%

```
CCCCAGCCCCAGCTACCACCTCCTCCCCGGCCGGCGGACAGTGGACGCGGCGGCGAGCCGCGGGCAGGGGCCGGAGCCCGCGCCCGGAG WT
CCCCAGCCCCAGCTACCACCTCCTCCCCGGCCGGCG--------TGGACGCGGCGGCGAGCCGCGGGCAGGGGCCGGAGCCCGCGCCCGGAG Δ8  (2x)
CCCCAGCCCCAGCTACCACCTCCTCCCCGGCCGG-------------ACGCGGCGGCGAGCCGCGGGCAGGGGCCGGAGCCCGCGCCCGGAG Δ13
CCCCAGCCCCAGCTACCACCTCCTCCCCGGC-------------TGGACGCGGCGGCGAGCCGCGGGCAGGGGCCGGAGCCCGCGCCCGGAG Δ13
CCCCAGCCCCAGCTACCACCTCCTCCCCGGCCGGCGGACAG----------------CCGCGGGCAGGGGCCGGAGCCCGCGCCCGGAG Δ16
CCCCAGCCCCAGCTACCACCTCCTCCCCGG-----------------ACGCGGCGGCGAGCCGCGGGCAGGGGCCGGAGCCCGCGCCCGGAG Δ17
CCCCAGCCCCAGCTACCACCTCCTCCCCGGCCGG-------------------------------------------AGCCCGCGCCCGaAG Δ43
CCCCAGCCCCAGCTACCACCTCCTCCCCGGCCGGCGGACAGTGGA-----------/ /--------------GGGGGTCGGGGCTCG Δ54  (3x)
CCCCAGCCCCAGCTACCACCTCCTCCCCGGCCGG-----------------------/ /--------------------------AG Δ56  (2x)
CCCCAGCCCCAGCTACCACCTCCTCCCCGGCCGGCGG------------------/ /--------------GGTGGAGGGGGTCGG Δ56  (2x)
CCCCAGCCCCAGCTACCACCTCCTCCCCGGCCGGCGGACAca--------------/ /--------------CaCGCGGCGTCGCAC Δ69  (Δ71 +2)
C------------------------------------------------------/ /------------------------GGAG Δ87  (2x)
CCCCAGCCCCAGCTACCACCTCCTCCCCGGCCGGCGGC-------------------/ /--------------CTGAAACTTTTCGTC Δ89  (7x)
CCGGAGCGCGGCGTG------------------------------------------/ /--------------GAGGGGGTCGGGGCT Δ165 (2x)
```

FIG. 5F-4

VEGFA site 3

Mutations in 31 of 77 sequences ≈ 40.3%

```
CCATTCCCTCTTTAGCCAGAGCCGGGGTGTGCAGACGGCAGTCACTAGGGGGCGCTCGGCCACCACAGGGAAGCTGGGTGAATGGAGCGAGC WT
CCATTCCCTCTTTAGCCAGAGCCGGGGTGTGCAGACGGCACT---TAGGGGGCGCTCGGCCACCACAGGGAAGCTGGGTGAATGGAGCGAGC Δ3
CCATTCCCTCTTTAGCCAGAGCCGGGGTGTGCAGACGGCA----CTAGGGGGCGCTCGGCCACCACAGGGAAGCTGGGTGAATGGAGCGAGC Δ4
CCATTCCCTCTTTAGCCAGAGCCGGGGTGTGCAGACGGCAGTCA-----GGGCGCTCGGCCACCACAGGGAAGCTGGGTGAATGGAGCGAGC Δ5
CCATTCCCTCTTTAGCCAGAGCCGGGGTGTGCAGACGGCA-----TAGGGGGCGCTCGGCCACCACAGGGAAGCTGGGTGAATGGAGCGAGC Δ5
CCATTCCCTCTTTAGCCAGAGCCGGGGTGTGCAGAC--------CTAGGGGGCGCTCGGCCACCACAGGGAAGCTGGGTGAATGGAGCGAGC Δ8  (2x)
CCATTCCCTCTTTAGCCAGAGCCGGGGTGTGCAGACGGCAG---------GGCGCTCGGCCACCACAGGGAAGCTGGGTGAATGGAGCGAGC Δ9
CCATTCCCTCTTTAGCCAGAGCCGGGGTGTGCAGAC---------TAGGGGGCGCTCGGCCACCACAGGGAAGCTGGGTGAATGGAGCGAGC Δ9  (2x)
CCATTCCCTCTTTAGCCAGAGCCGGGGTGTGCAG----------CTAGGGGGCGCTCGGCCACCACAGGGAAGCTGGGTGAATGGAGCGAGC Δ10
CCATTCCCTCTTTAGCCAGAGCCGGGGTGTGCAGA-----------AGGGGGCGCTCGGCCACCACAGGGAAGCTGGGTGAATGGAGCGAGC Δ11
CCATTCCCTCTTTAGCCAGAGCCGGGGTGTGCA-----------CTAGGGGGCGCTCGGCCACCACAGGGAAGCTGGGTGAATGGAGCGAGC Δ11
CCATTCCCTCTTTAGCCAGAGCCGGGGTGTGCAGACGGCAGT-------------TCGGCCACCACAGGGAAGCTGGGTGAATGGAGCGAGC Δ13
CCATTCCCTCTTTAGCCAGAGCCGGGGTGT--------------CTAGGGGGCGCTCGGCCACCACAGGGAAGCTGGGTGAATGGAGCGAGC Δ14
CCATTCCCTCTTTAGCCAGAGCCGGGGTGTGCAG---------------GGGCGCTCGGCCACCACAGGGAAGCTGGGTGAATGGAGCGAGC Δ15
CCATTCCCTCTTTAGCCAGAGCCGGGGTGTGCAG----------------GGCGCTCGGCCACCACAGGGAAGCTGGGTGAATGGAGCGAGC Δ16 (2x)
CCATTCCCTCTTTAGCCAGAGCCGGGGTGTGCAGACGGCAGTCACTA----------------CACAGGGAAGCTGGGTGAATGGAGCGAGC Δ16
CCATTCCCTCTTTAGCCAGAGCCGGGGTGTGCAGACGGC----------------TCGGCCACCACAGGGAAGCTGGGTGAATGGAGCGAGC Δ16
CCATTCCCTCTTTAGCCAGAGCCGGGGTGTGCAGACG-----------------CTCGGCCACCACAGGGAAGCTGGGTGAATGGAGCGAGC Δ17
CCATTCCCTCTTTAGCCAGAGCCGGGGTG-------------------GGGGCGCTCGGCCACCACAGGGAAGCTGGGTGAATGGAGCGAGC Δ19
CCATTCCCTCTTTAGCCAGAGCCGGGGTGTGC---------------------GCTCGGCCACCACAGGGAAGCTGGGTGAATGGAGCGAGC Δ21
CCATTCCCTCTTTAGCCAGAGCCGGGGTGTGCAGACGGCAGTCAC---------------------AGGGAAGCTGGGTGAATGGAGCGAGC Δ21
CCATTCCCTCTTTAGCCAGAGCCGGGGTGTGCAGACGGC---------------------CACCACAGGGAAGCTGGGTGAATGGAGCGAGC Δ21 (2x)
CCATTCCCTCTTTAGCCAGAGCCGGGGTGTGCAGACGGCAGTCA-----------------------GGGAAGCTGGGTGAATGGAGCGAGC Δ23
CCATTCCCTCTTTAGCCAGAGCCGGGGTGTGCAG------------------------GCCACCACAGGGAAGCTGGGTGAATGGAGCGAGC Δ24
CCATTCCCTCTTTAGCCAGAGCCGGGGTGTGCAGAC------------------------CACCACAGGGAAGCTGGGTGAATGGAGCGAGC Δ24
CCATTCCCTCTTTAGCCAGAGCCGGGGTGTGCAGAC---------------------------CACAGGGAAGCTGGGTGAATGGAGCGAGC Δ27
CCATTCCCTCTT----------------------------------AGTGACTGCTCGGCCACCACAGGGAAGCTGGGTGAATGGAGCGAGC Δ34 (2x)
```

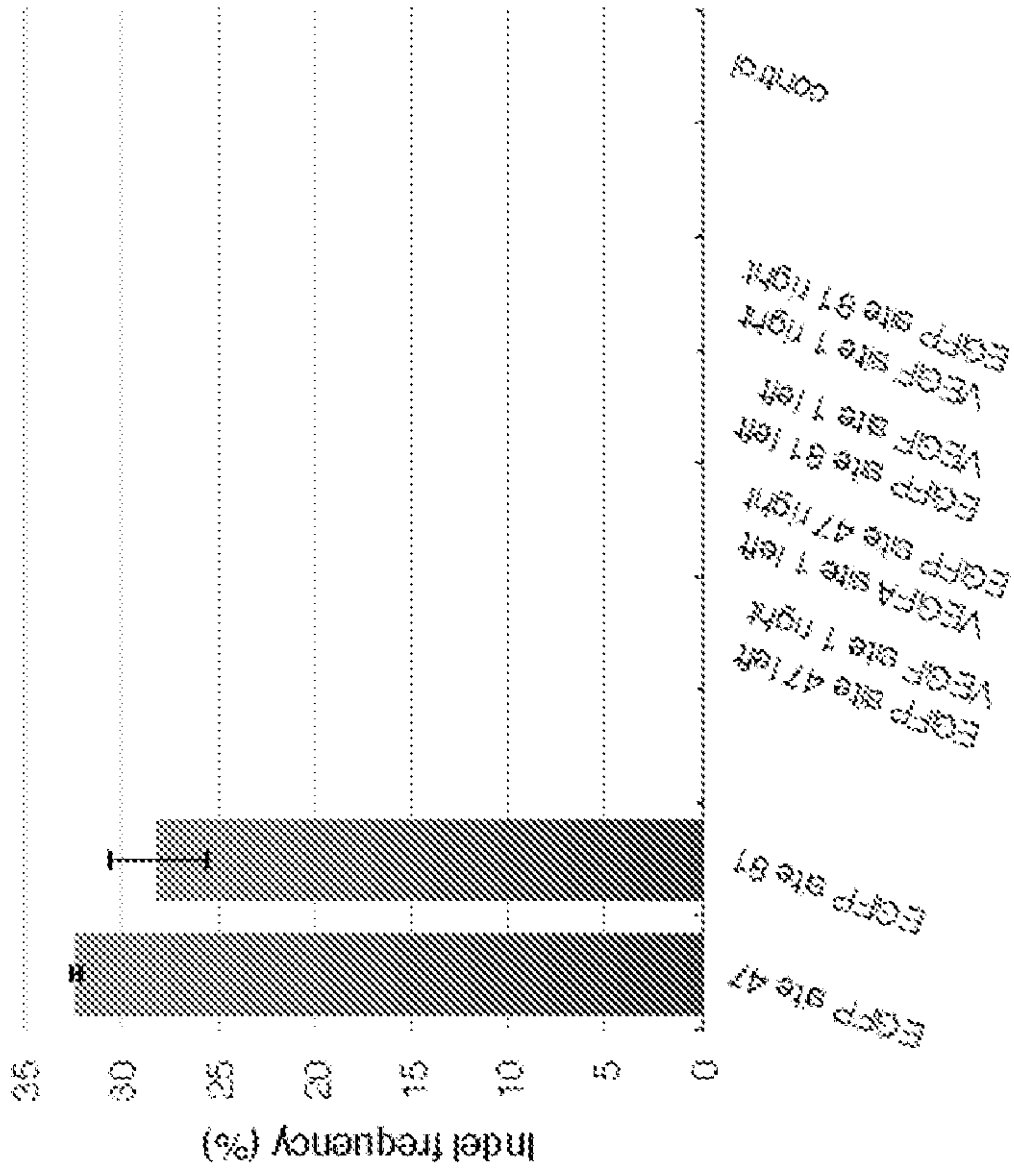


FIG. 6B

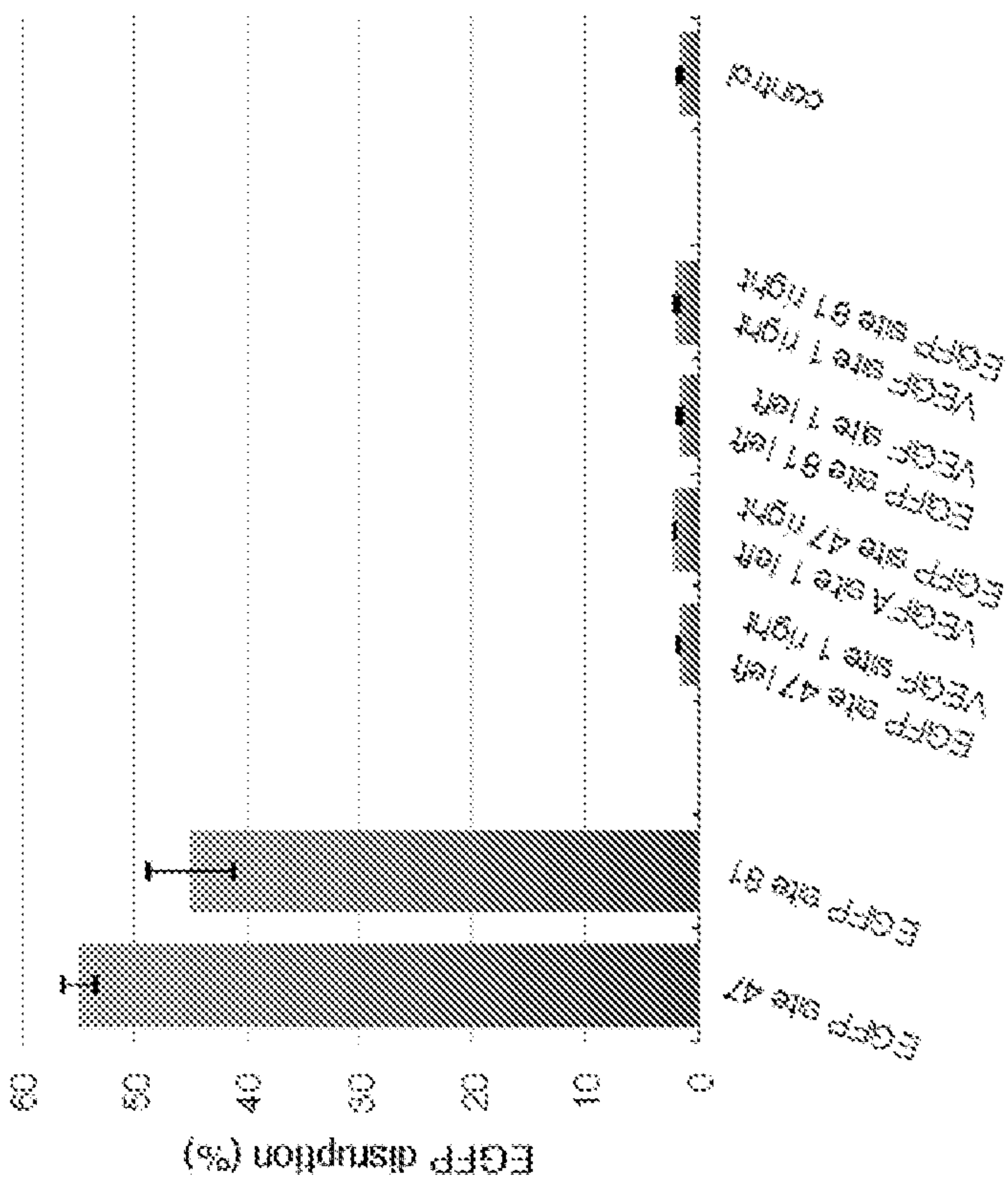


FIG. 6A

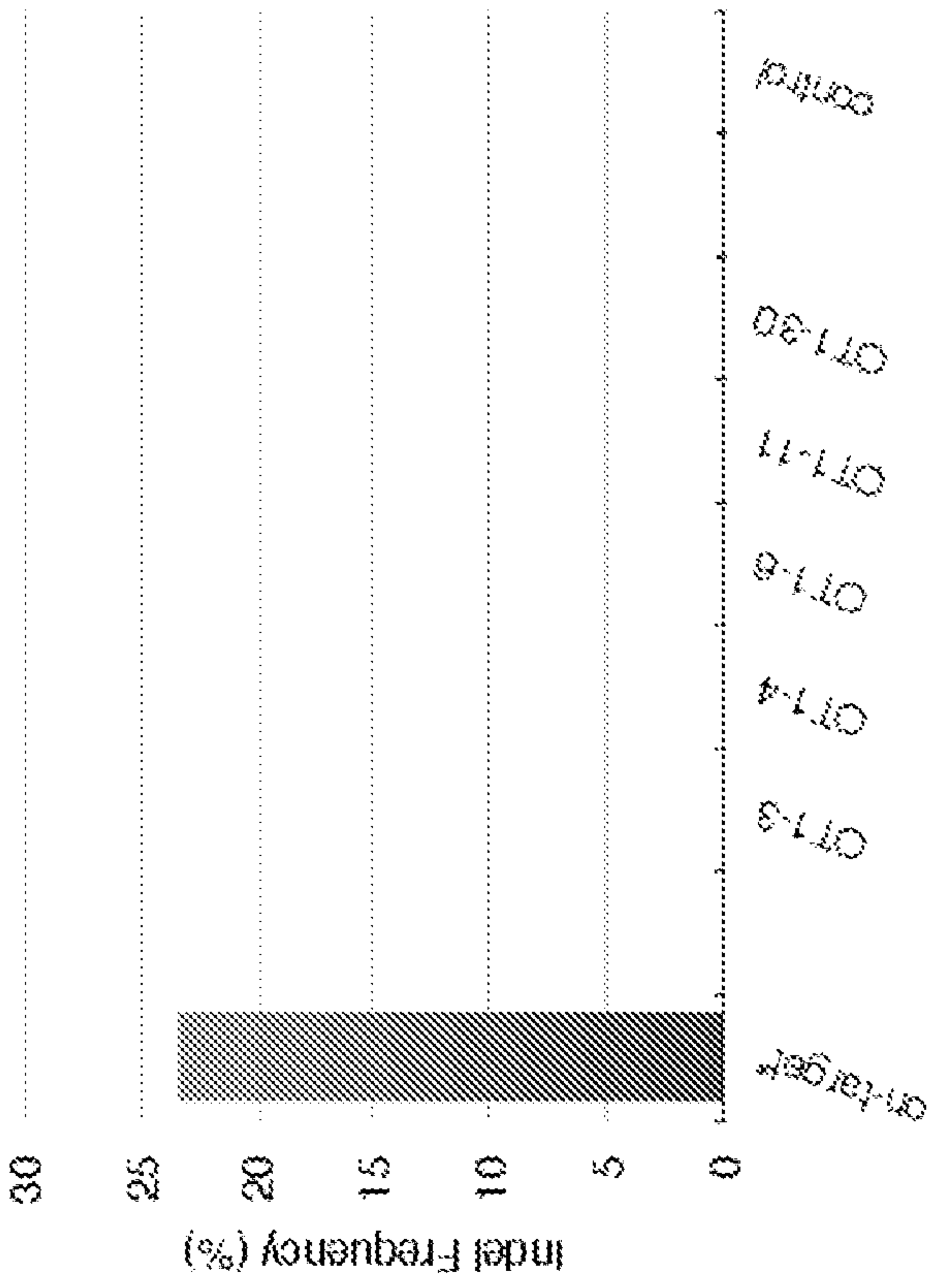


FIG. 6D

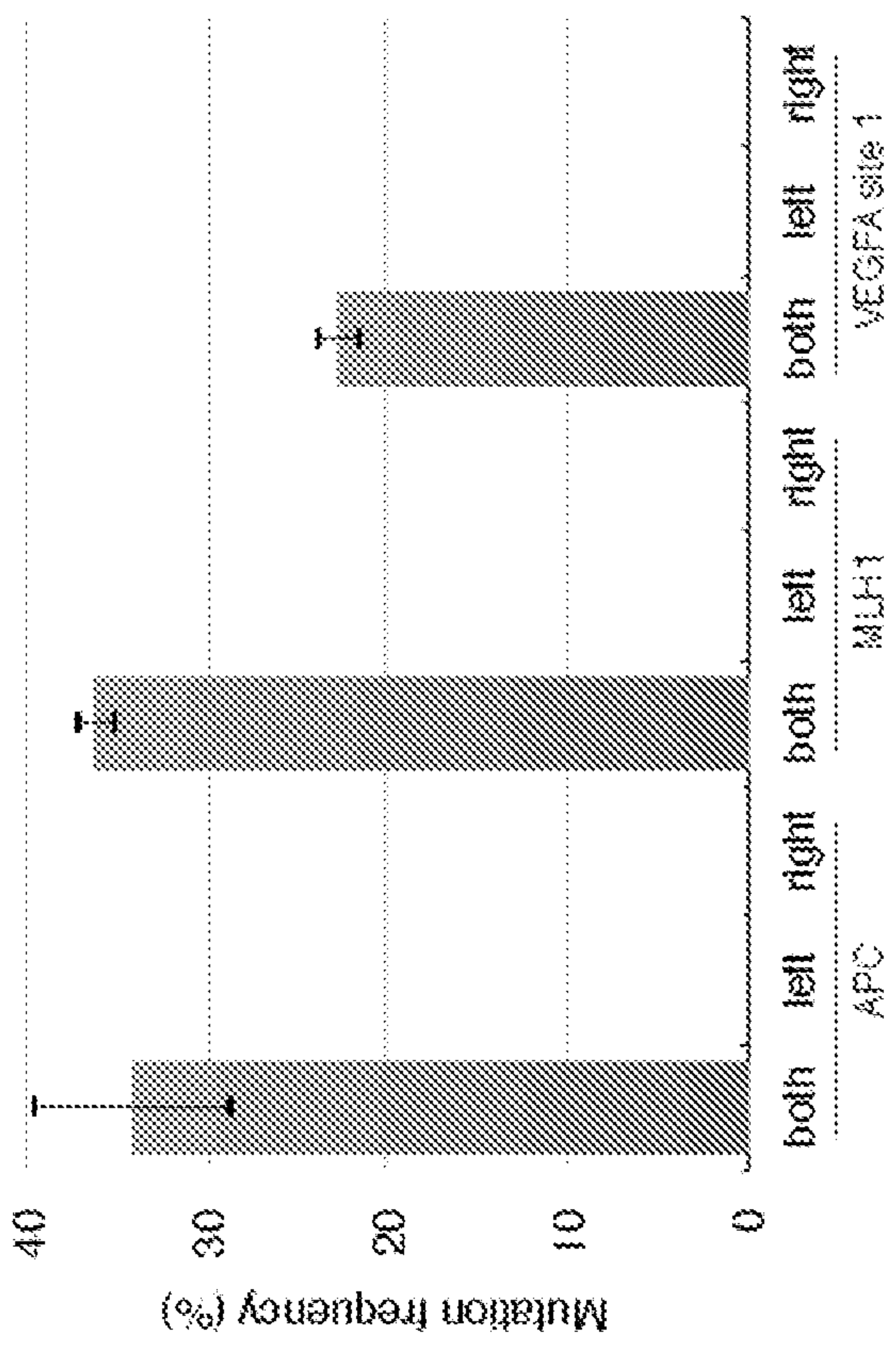


FIG. 6C

Indel frequency (%)
50
40
30
20
10
0

both left right
VEGFA site 1
DDB2
FANCF site 1
FANCF site 2
FES
RUNX1
FokI-dCas9

both left right
VEGFA site 1
DDB2
FANCF site 1
FANCF site 2
FES
RUNX1
Cas9 nickase

VEGF site 1
DDB2
FANCF site 1
FANCF site 2
FES
RUNX1
control

FIG. 7A

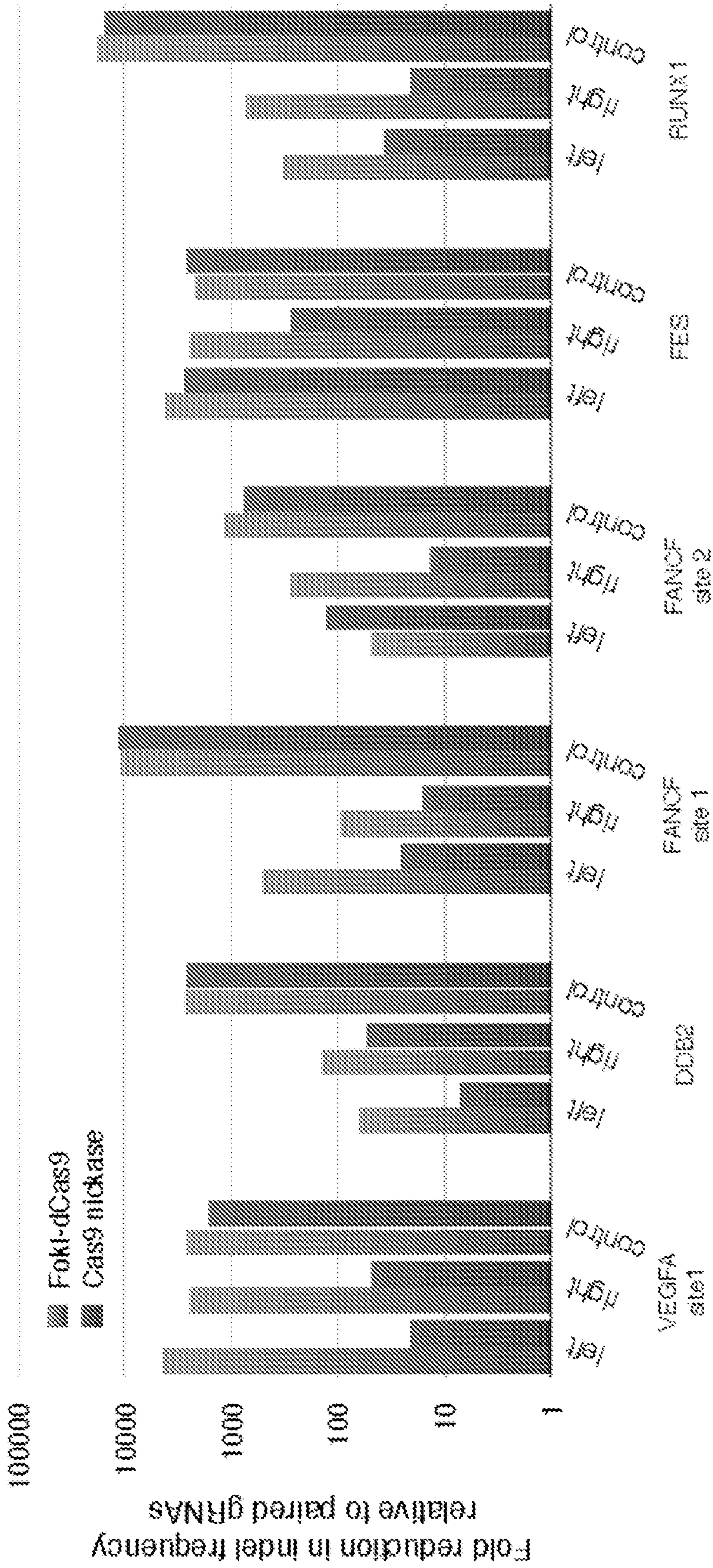
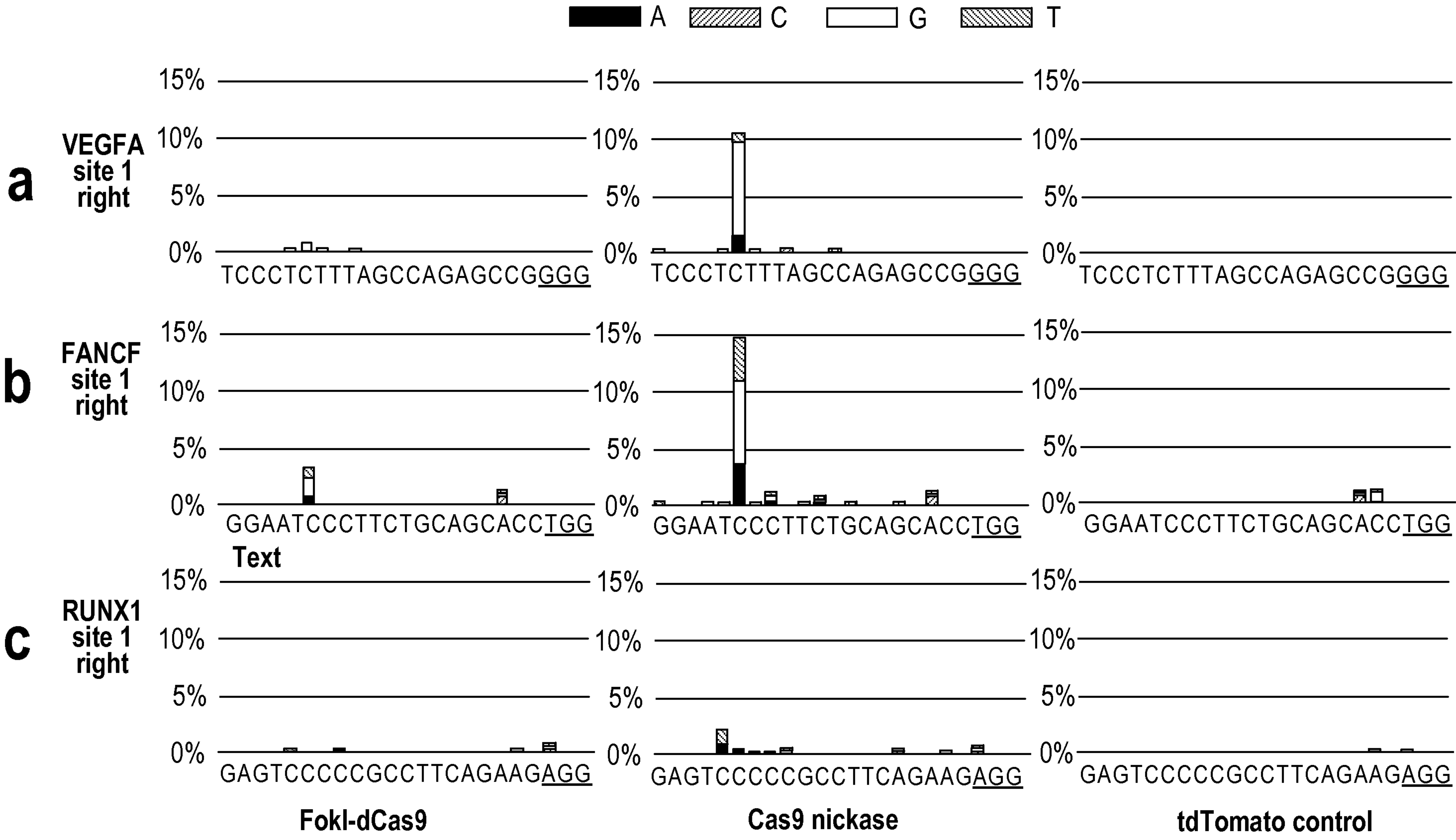
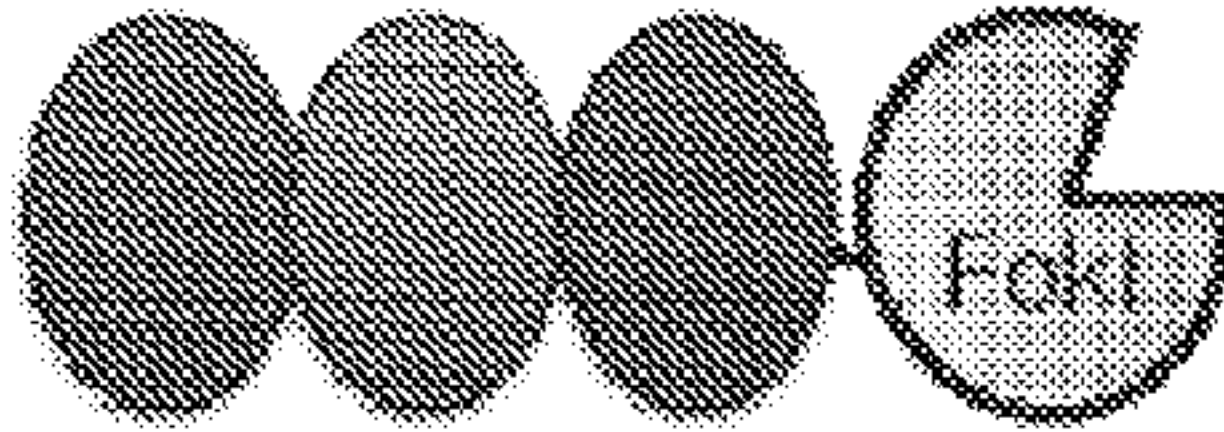


FIG. 7B

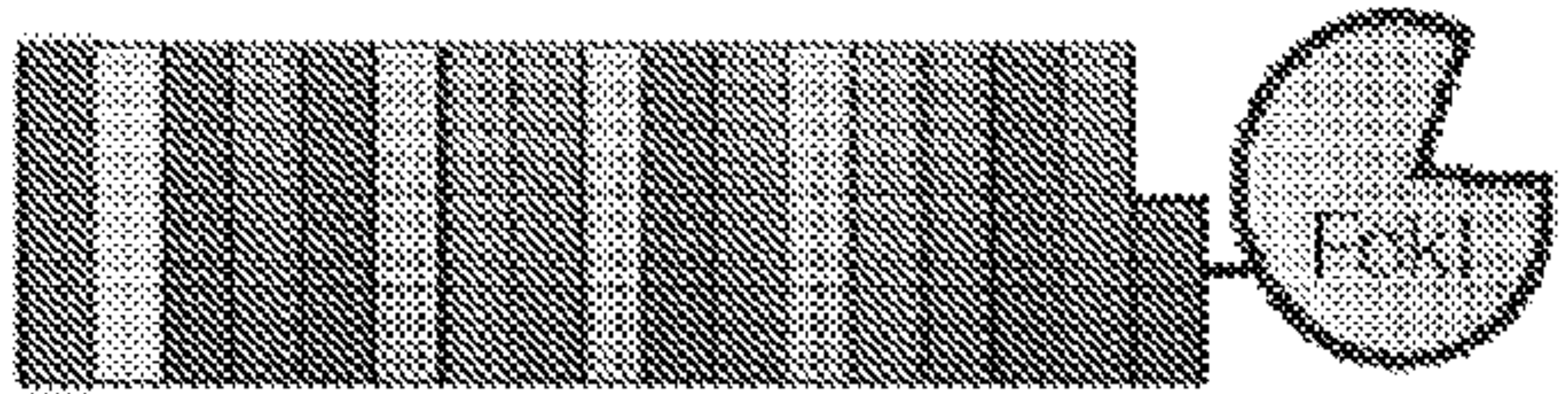


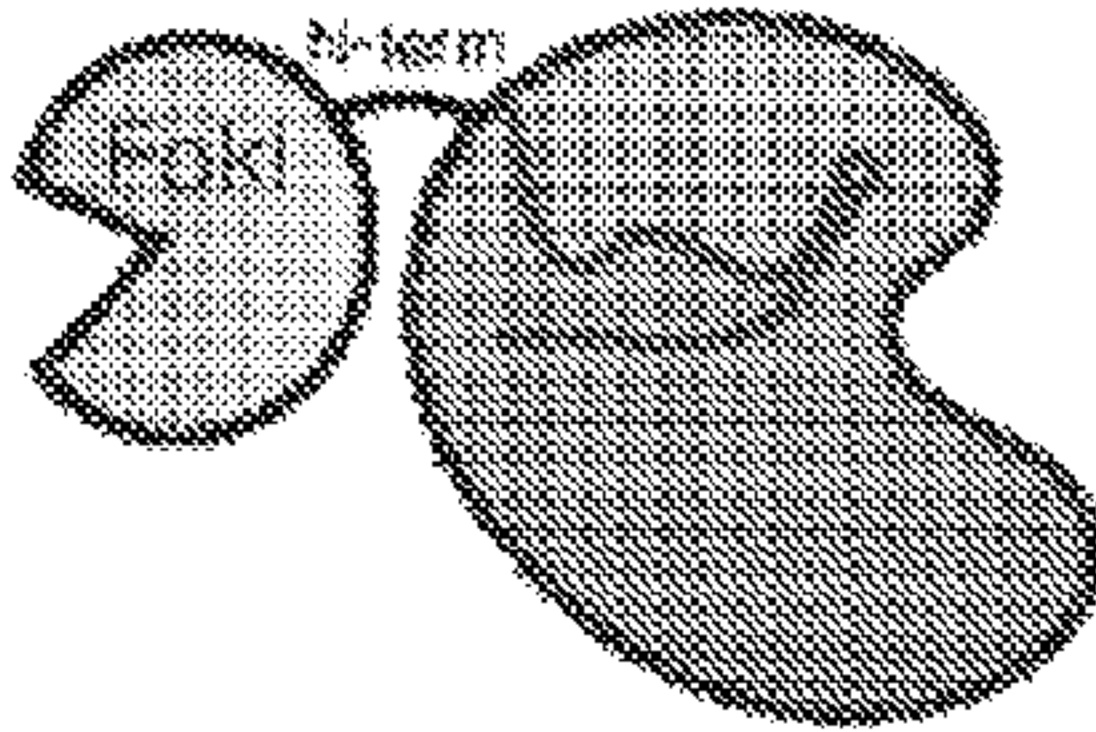
Figures 8A-C

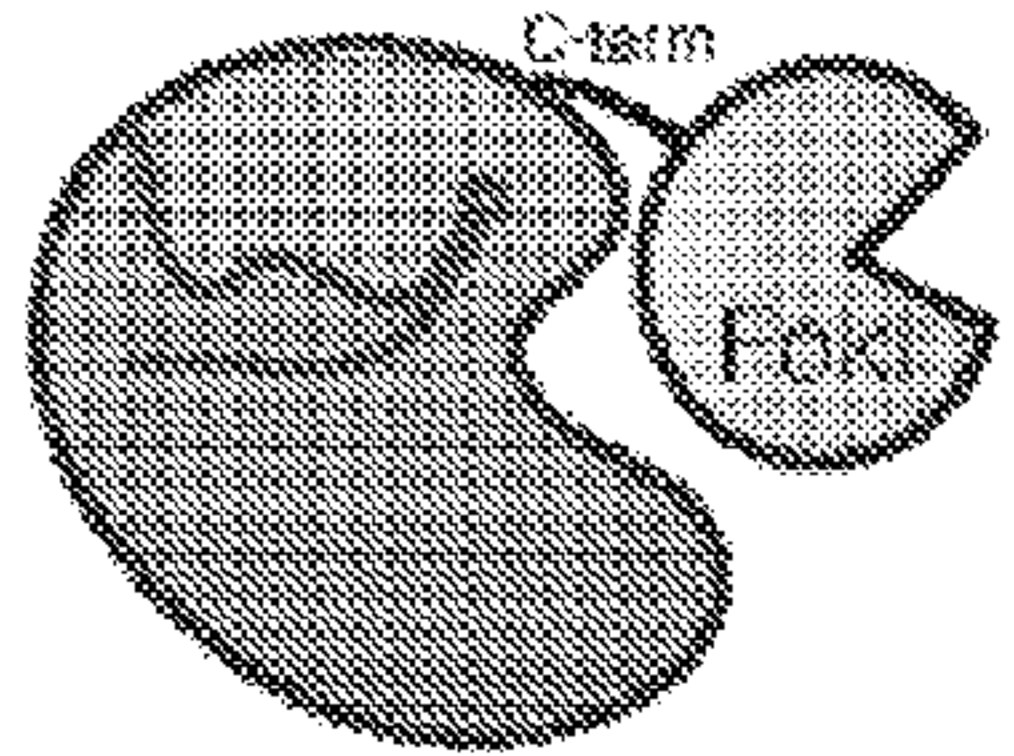
ZFN



TALEN







FokI-dCas9 dCas9-FokI