



(51) International Patent Classification:

C12N 15/66 (2006.01) C12N 15/55 (2006.01)
C12N 9/22 (2006.01) C12N 5/10 (2006.01)

(21) International Application Number:

PCT/US2014/039195

(22) International Filing Date:

22 May 2014 (22.05.2014)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

61/826,338 22 May 2013 (22.05.2013) US

(71) Applicants: NORTHWESTERN UNIVERSITY
[US/US]; 1800 Sherman Ave., Evanston, IL 60201 (US).
WISCONSIN ALUMNI RESEARCH FOUNDATION
[US/US]; 614 Walnut Street, 13th Floor, Madison, WI
53726 (US).

(72) Inventors: SONTHEIMER, Erik, J.; 628 Abbotsford
Road, Kenilworth, IL 60043 (US). ZHANG, Yan; 722
Clark Street, Apt. 310, Evanston, IL 60201 (US). MON-
DRAGON, Alfonso; 5540 North Magnolia Ave., Chicago,
IL 60640 (US). RAJAN, Rakhi; 8538 Lotus Ave., Apt.
419, Skokie, IL 60077 (US). THOMSON, James; 1807
Regent St., Madison, WI 53726 (US). HOU, Zhonggang;
4817 Sheboygan Ave., Apt. 202, Madison, WI 53705
(US).

(74) Agents: MCBRIDE, M. Scott et al.; Andrus Intellectual
Property Law, LLP, 100 East Wisconsin Avenue, Suite
1100, Milwaukee, WI 53202 (US).

(81) Designated States (unless otherwise indicated, for every
kind of national protection available): AE, AG, AL, AM,
AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY,
BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM,
DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT,
HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR,
KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME,
MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ,
OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA,
SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM,
TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM,
ZW.

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

Published:

- with international search report (Art. 21(3))
- with sequence listing part of description (Rule 5.2(a))

(54) Title: RNA-DIRECTED DNA CLEAVAGE AND GENE EDITING BY CAS9 ENZYME FROM NEISSERIA MENINGITID-
IS

(57) Abstract: Disclosed are components and methods for RNA-directed DNA cleavage and gene editing. The components include
and the methods utilize a Cas9 protein from Neisseria and one or more RNA molecules in order to direct the Cas9 protein to bind to
and optionally cleave or nick a target sequence.



RNA-DIRECTED DNA CLEAVAGE AND GENE EDITING BY CAS9 ENZYME FROM NEISSERIA MENINGITIDIS

CROSS-REFERENCE TO RELATED PATENT APPLICATIONS

[0001] The present application claims the benefit of priority under 35 U.S.C. 119(e) to U.S. Provisional Patent Application No. 61/826,338, filed on May 22, 2013, the content of which is incorporate herein by reference in its entirety.

BACKGROUND

[0002] The field of the invention relates to methods, kits, and compositions for modifying DNA. In particular, the field of the invention relates to components and methods for RNA-directed DNA cleavage and gene editing.

[0003] Genome editing has proven to be quite difficult in cells, particularly in mammalian cells. One way to improve genome-editing efficiency is to introduce a double-strand break (DSB) in the desired DNA region. DSBs stimulate the DNA repair machinery and, in the presence of a homologous repair template, greatly enhance genome editing efficiency. Currently, there are two widely used systems to introduce targeted DSBs in genomes of mammalian cells - Zinc Finger Nucleases (ZFNs) and Transcription activator-like effector nucleases (TALENs), both of which are engineered by fusing site-specific DNA recognition domains to FokI endonucleases. One major obstacle to the wide application of these two systems is the difficulty of engineering peptides that recognize specific target DNA sites. Also, for each unique target sequence, a different pair of ZFNs or TALENs has to be engineered. For ZFNs, the optimal designing algorithm is proprietary and only available through commercial sources. For TALENs, the design rules are quite straightforward, but it still takes weeks to make one pair of targeting constructs, and each must be adequately expressed and validated.

[0004] Clustered, regularly interspaced short palindromic repeats (CRISPRs) are known in the art (see Marraffini and Sontheimer, Nature Reviews Vol. 11, Mach 2010, 181-190, U.S. Published Patent Application No. 2010/0076057, and U.S. Patent No. 8,697,359, which are incorporated herein by reference in their entirety), and have been utilized for genome editing (see Cain, SciBX, Vol. 6, No. 4, January 2010, 1-7). Recently,

CRISPR RNAs (crRNAs) have been developed that direct DNA cleavage by a bacterial protein called Cas9. (See Cong *et al.*, Science, Vol. 339, February 15, 2013, 819-822; and Mali *et al.*, Science, Vol. 339, February 15, 2013, 823-826). This system requires only three components: a Cas9 endonuclease, a trans-activating CRISPR RNA (tracrRNA), and the target-specifying crRNA which hybridizes to a target DNA sequence and targets the DNA sequence for cleavage by the Cas9 endonuclease. Accordingly, nearly any genomic locus can be targeted by the same Cas9 protein, as long as a crRNA complementary to the targeted sequence is provided. Two Cas9 proteins (SpCas9 from *Streptococcus pyogenes* and StCas9 from *Streptococcus thermophilus*) have been reported as effective in genome editing, and each has its own targeting sequence requirements. However, there is a need for the identification of new systems in order to maximize the potential of CRISPR as a gene editing tool.

[0005] Here, we report a new form of Cas9 (NmCas9 from *Neisseria meningitidis*) that has distinct targeting requirements which are less likely to result in off-target effects. Furthermore, unlike SpCas9 and StCas9, NmCas9 can function with crRNAs that are embedded within longer unprocessed precursors, indicating that NmCas9 can accommodate a greater range of targeting crRNA structures and functionalities. In addition, mutant forms of NmCas9 can be used that bind DNA in an RNA-directed fashion, but that do not cleave the DNA.

SUMMARY

[0006] Disclosed are methods and components for RNA-directed DNA cleavage and gene editing. The methods utilize components including a Cas9 protein from *Neisseria* and one or more RNA molecules in order to direct the Cas9 protein to bind to and optionally cleave or nick a target sequence.

[0007] In some embodiments, the methods modify a target DNA sequence in a cell and may include: (a) expressing a Cas9 protein from a *Neisseria* species or a variant protein thereof in the cell (e.g., by transfecting the cell with a DNA molecule or an RNA molecule that expresses the Cas9 protein (*i.e.*, Cas9 mRNA)); and (b) transfecting the cell with an RNA molecule or expressing an RNA molecule in the cell from a DNA molecule, wherein the RNA molecule binds to the Cas9 protein or variant, and the RNA molecule

hybridizes to the target DNA sequence. Optionally, the Cas9 protein or variant protein has nuclease activity (*e.g.*, DNase activity and/or RNase activity) and cleaves one (*i.e.*, nicks) or both strands of the target DNA sequence. Optionally, the methods further comprise contacting the target DNA sequence with a homologous DNA fragment, wherein homologous recombination is induced between the homologous DNA fragment and the target DNA sequence (*e.g.*, homologous recombination to effect gene repair or to effect gene disruption). In further embodiments, the Cas9 protein or variant protein has no nuclease activity (*e.g.*, no DNase activity) and binds to the target DNA sequence but does not cleave the DNA sequence.

[0008] In some embodiments, the Cas9 protein or variant protein is encoded and expressed by a nucleic acid having a codon sequence that is optimized for expression in the cell. For example, the nucleic acid may have a codon sequence that is optimized for expression in an animal cell (*e.g.*, a human or non-human mammalian cell). The Cas9 protein may be expressed from an expression vector comprising a prokaryotic or eukaryotic promoter for expressing the Cas9 protein which is transfected into the cell.

[0009] Suitable Cas9 proteins may include, but are not limited to, Cas9 proteins from *Neisseria* species (*e.g.*, *Neisseria meningitidis*). Variants of Cas9 proteins may include proteins having an amino acid sequence that has at least about 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity to an amino acid sequence of a Cas9 protein. Optionally, the variant has one or more biological activities associated with the Cas9 protein (*e.g.*, nuclease activity and RNA binding activity).

[0010] The methods may be utilized to target a DNA sequence in a cell. Suitable cells may include prokaryotic cells and eukaryotic cells. In some embodiments, the methods are performed to target a DNA sequence in a stem cell (*e.g.*, an embryonic stem cell or an induced pluripotent stem cell).

[0011] The methods typically utilize an RNA molecule that comprises a sequence that hybridizes with a target DNA sequence in a cell. The RNA molecule also binds with the Cas9 protein or a variant thereof. In some embodiments, the RNA molecule comprises two molecules of duplexed RNA (*e.g.*, crRNA duplexed with tracrRNA). In other embodiments, the RNA molecule is a single RNA molecule forming a hairpin structure

(e.g., where crRNA is linked to tracrRNA via a linker such as GAAA and the crRNA and tracrRNA form the stem of the hairpin). In further embodiments, the RNA may include an RNA mimic of green fluorescent protein (GFP). As such, the RNA may be utilized to map a target DNA sequence via adding 4-hydroxybenzylidene (HPD), 3,5-dimethoxy-4-hydroxybenzylidene (DMHPD), or a 3,5-difluoro-4-hydroxybenzylidene to the cell (DFHPD), wherein the RNA binds to HPD, DMHPD, or DFHPD to form a fluorescent complex. In further embodiments, the RNA may comprise Xist RNA or fragments thereof, which may be utilized to modulate the expression of the target DNA sequence. The RNA may be transfected directly into a cell and/or may be expressed from an expression vector comprising a prokaryotic or eukaryotic promoter for expressing the RNA when the expression vector is transfected into the cell.

[0012] Also disclosed are proteins, polynucleotides, vectors, and kits for performing the disclosed methods. For example, a contemplated protein may include the *Neisseria meningitidis* Cas9 protein or a variant thereof. A contemplated polynucleotide may comprise a eukaryotic promoter operably linked to a polynucleotide sequence encoding a Cas9 protein from a *Neisseria* species or a variant thereof (e.g., a Cas9 protein fused to one or more of a nuclear localization signal (NLS), a ligand for purifying the variant protein, and a tag for identifying the variant protein). The polynucleotide may be present in a vector for propagating the polynucleotide or expressing the polynucleotide (e.g. a prokaryotic and/or eukaryotic vector).

[0013] The contemplated kits may comprise any of the presently disclosed proteins, polynucleotides, and vectors. A kit may comprise: (a) a polynucleotide for expressing a Cas9 protein from a *Neisseria* species or a variant protein thereof in a cell (e.g., as part of an expression vector comprising a eukaryotic promoter for expressing the Cas9 protein or alternatively as Cas9 mRNA); and (b) RNA that binds to the Cas9 protein or variant and RNA that hybridizes to the target DNA sequence in the cell (e.g., as a single RNA or as multiple RNAs, or as a DNA vector or vectors that expresses the single or multiple RNAs).

[0014] Also contemplated herein are cells that are transformed or transfected with the polynucleotides or vectors contemplated herein. Suitable cells may include prokaryotic and eukaryotic cells.

BRIEF DESCRIPTION OF THE DRAWINGS

[0015] Figure 1. A newly identified mode of crRNA biogenesis in *Neisseria lactamica*. (A) Differential RNA-seq (dRNA-seq)-based analysis of the minimal CRISPR/Cas system in *N. lactamica* 020-06 reveals expression of tracrRNA and crRNAs. Approximately three million cDNAs from untreated and TEX-treated RNA were sequenced and mapped to the genome. Read counts are plotted here for the CRISPR/*cas* locus. Both strands of all libraries were adjusted to the same scale (maximum of 50,000 for leading strand; minimum of -50,000 for lagging strand) that reflects a relative expression score. The number of reads obtained for tracrRNA and crRNAs are in the range of 40,000 to 50,000 each, which is comparable to the range we observe for other high-abundance classes of RNAs. (B) Top: dRNA-seq data were mapped onto the genomic region corresponding to the *tracrRNA* gene. Expression scores at each position were adjusted to the same relative scale. Bottom: sequence of the full-length 107-nt form and processed 91-nt form of tracrRNA. The arrow indicates the primary tracrRNA transcription start site (TSS) based on enrichment in the TEX+ libraries. Boxes denote the extended -10 promoter element and the anti-CRISPR-repeat region of the tracrRNA. (C) Northern analysis of *N. lactamica* 020-06 tracrRNA. OD₆₀₀ (left panel) and TEX treatment (right panel) is denoted at the top of each lane. Processed and unprocessed tracrRNAs are schematized on the right, with the RNase III processing site indicated with an arrowhead and the anti-repeat region indicated as the left-hand box. (D) Top: dRNA-seq data were mapped onto the genomic region corresponding to the CRISPR array. Expression scores at each position were adjusted to the same relative scale. Middle: the primary TSS of each CRISPR spacer-repeat unit (based on enrichment in the TEX+ libraries) is indicated by an arrow. Primary CRISPR transcripts of different lengths with likely 5' triphosphates also are indicated by arrows. Bottom: The sequence of spacer 4 and its flanking CRISPR repeats, with the putative extended -10 box (consensus sequence 5'-tgnTACAAT-3') in each single repeat. (E) Northern analysis of *N. lactamica* 020-06 crRNA, using a probe complementary to the CRISPR repeat. OD₆₀₀ (left panel) and TEX treatment (right panel)

is denoted at the top of each lane. Candidate monomeric, dimeric and trimeric crRNAs are schematized on the right, each of a predicted size consistent with bands observed on the blots.

[0016] Figure 2. CRISPR repeats contain active promoters that form the 5' ends of mature crRNAs, whereas RNase III processing forms crRNA 3' ends. (A) Top left panel: Promoter-element-containing sequences used for *gfp* fusions. Plasmids included a wild-type (wt) and mutated (mut) *N. lactamica* 020-06 CRISPR repeat (pNH13 and pNH14, respectively), a *C. jejunii* NCTC11168 CRISPR repeat (pNH18), and a positive control promoter from T7 phage A1. The promoterless control *gfp* construct (pAS00046) and a construct with three wildtype *N. lactamica* 020-06 CRISPR repeats (pNH17) are not shown. Promoter elements are indicated by boxes. P_{mut} has a TA→CC mutation in the promoter element as compared to P_{wt} . Top right panel: Flow-cytometric fluorescence intensity measurements of cells containing the transcriptional *gfp* fusions described in the left panel. Fluorescence values are expressed in arbitrary units (AU). Error bars indicate standard deviation for three independent biological replicates. Bottom panel: Fluorescence images of transcriptional *gfp* fusion strains grown on agar plates. The right image was captured in the visible light mode; the left image shows the same plate in fluorescence mode (460 nm excitation, 510 nm emission). (B) Classification of Type II CRISPR/*cas* loci. The genomic organization of representative Type II-A or Type II-B loci, as defined previously, is given on top. Below are two CRISPR/*cas* loci (including from *neisseriae*) from the newly defined Subtype II-C. (C) Gel electrophoresis of radiolabeled RNAs from *in vitro* transcription reactions using linear DNA templates from a subset of the CRISPR repeat-containing sequences given in (A). Full-length run-off transcripts (168 nt) are denoted by the arrow. The area to the right of the dotted line was imaged at lower contrast to avoid overexposure. See also Figure 8. (D) Top: Base pairing between a primary crRNA and tracrRNA. RNase III cleavage sites inferred from dRNA-seq are indicated by arrows. Bottom: Northern analysis of total RNA from *N. meningitidis* WUE2594 and its Δrnc derivative during mid-log and early stationary phase, probed for tracrRNA (left) and crRNAs (right).

[0017] Figure 3. CRISPR organization in *Neisseria*. (A) Schematic representation of CRISPR loci from seven different *Neisseria* strains. Strain names are indicated (*Nm*, *N.*

meningitidis; *Nl*, *N. lactamica*) with the total number of spacers in each strain shown in parentheses. The arrow indicates the direction of crRNA transcription. Repeats and spacers are shown as rectangles and diamonds, respectively. Unique spacers, repeats that match the consensus, and variant repeats are illustrated (see Tables 2 and 3). (B) Potential natural targets for *N. meningitidis* 8013 spacers. 9 out of 25 spacers match varying numbers of *Neisseria* genomes. For each spacer, its number in the 8013 array, the quantity of known protospacer matches, and a representative target genome (*Nm*, *N. meningitidis*; *Ng*, *N. gonorrhoea*) are listed. Protospacers and 10 flanking nts (on both sides) from the representative target genomes are aligned. Sequence similarities are indicated at the top, revealing the 5'-GATT-3' PAM consensus 5-8 nts 3' of the protospacer. The WebLogo is derived from the alignment of all *Neisseria*-matched protospacers, not just those shown here that match spacers from the *N. meningitidis* 8013 CRISPR. The PAM regions and non-consensus nucleotides are illustrated. Potential self-targeting spacers and spacers with possible prophage-like targets are illustrated. (See also Figure 9).

[0018] Figure 4. Natural transformation is limited by the native *Neisseria* CRISPR/Cas system. (A) Schematic representation of integrational vector pGCC2, and recombination between pGCC2 and the meningococcal chromosome (Tobiason and Seifert, 2010). Individual elements are not drawn to scale. (B) pGCC2 derivatives with potential targets for different *N. meningitidis* 8013 spacers (1, 8, 9, 16, 17, 18, 23, and 25) were tested by natural transformation assays using wild-type 8013 as the recipient strain. The data show log-scale plots of colony-forming units (cfu) per ml [mean $\pm$ s.e.m. (standard error of the mean) for three independent biological replicates] for total cells (left bars) and *Erm*^R transformants (right bars) from three independent experiments. See also Figure 10. (C) Top panel, sequences of a series of mutations in the pGCC2 derivative carrying the 350 nt target for spacer 9. The arrow indicates reversed orientation of the target sequence in the plasmid. Bottom panel, pGCC2 constructs containing the spacer 9 target mutations are tested by natural transformation into wild-type *N. meningitidis* 8013. Data are presented as in (B). Error bars represent s.e.m. for three independent biological replicates.

[0019] Figure 5. Type II-C CRISPR interference requires Cas9 and tracrRNA but is independent of RNase III-mediated processing. (A) Schematic representations of *N.*

meningitidis 8013 mutant strains: including *cas1*, *cas2*, *cas9* and *rnc* genes; *tracrRNA* gene; kanamycin resistance gene; and CRISPR repeats and spacers (squares and diamonds, respectively). Arrows indicate transposon insertions in the *rnc::Tn*, *cas9::Tn*, *cas1::Tn*, and *cas2::Tn* mutants. (B, C) pYZEJS040 (-) and its protospacer 25-containing derivative (+) were tested by natural transformation assays using *N. meningitidis* 8013 and its mutant derivatives as recipients. Relevant genotypes as well as the presence or absence of pGCC2-mediated *cas9* complementation are given at the bottom. The data show cfu/ml (log scale, mean $\pm$ s.e.m. for three independent biological replicates) for total cells (left bars) and chloramphenicol-resistant transformants (right bars). (See also Figure 11). (D) Total RNA from the indicated strains were subjected to northern analysis using a probe complementary to spacer 25 (top). In the lower panel, the same blot was probed for 5S RNA as a loading control. (See also Figure 12). (E) As in (D), except that a probe specific for *tracrRNA* was used.

[0020] Figure 6. *Neisseria* type II-C CRISPR/Cas limits natural transformation of meningococcal chromosomal DNA. (A) Schematic representation of the *siaA/ctrA* and *lctP/aspC* chromosomal loci in the *cas9::Tn* strain (top; with the transposon insertion indicated). Below are derivatives of the same *cas9::Tn* strain following transformation with the *ermC*-marked vector pGCC2 with or without a protospacer that matches CRISPR spacer 25, or with the *cat*-marked vector pYZEJS040 with or without a protospacer that matches CRISPR spacer 25. (B) For the left panel, gDNA from the *ermC*-marked strains shown in (A), as well as the unmarked control strain, was used in transformation assays with wild-type *N. meningitidis* 8013. Natural transformations were performed and presented as in Figure 4B. The right panel shows an analogous experiment using gDNA from the *cat*-marked strains shown in (A), as well as the unmarked control strain. Error bars represent s.e.m. for three independent biological replicates.

[0021] Figure 7. CrRNA biogenesis and CRISPR interference in *Neisseria*. Type II-C CRISPR/*cas* loci in *Neisseria* spp. initiate transcription within each spacer, driven by promoter elements embedded within each repeat. The resulting crRNAs and pre-crRNAs carry 5'-terminal triphosphates. Following *tracrRNA* annealing, RNase III can cleave both strands of the *tracrRNA*/pre-crRNA duplex (right pathway). Unexpectedly, pre-crRNA processing is not required: when RNase III is unavailable or fails to act, Cas9 can still

form functional complexes with tracrRNA and crRNA (left pathway). The naturally-encoded crRNAs target sequences present in other *Neisseria* spp. chromosomes, consistent with the high frequency of genetic exchange by natural transformation. Because Type II-C have only three protein-coding genes, lack leader sequences upstream of the CRISPR array, and do not require the host factor RNase III, they are the most streamlined natural CRISPR/Cas systems known.

[0022] Figure 8. *In vitro* transcription assay with *E. coli* σ^{70} polymerase (RNAP) holoenzyme with linear DNA templates containing either a wild-type or a mutant repeat. The complete gel of the experiment from Figure 2C is shown. The area to the right of the dotted line was imaged at lower contrast to avoid overexposure.

[0023] Figure 9. Alignment of potential natural targets for all *Neisseria* spacers revealed a putative 3' PAM. A total of 35 unique *Neisseria* spacer sequences have potential matches to varying numbers of *Neisseria* genomes or plasmids. For each of them, spacer number, the number of known protospacer matches, and one representative target genome are listed. Protospacers and 10 nts of 5' and 3' flanking sequences are aligned and the resulting sequence logos are shown on the top. A putative 3' PAM (5'-NNNGATT-3') is deduced. PAM equivalent regions in the targets are illustrated as well as nucleotides differing from consensus PAM. The five spacers with "self-targeting" potential, their spacer name and representative target genome also are indicated. Spacers with at least one phage-related potential natural target, their spacer number and representative target genome also are indicated.

[0024] Figure 10. Protospacer-containing potential targets cloned into pGCC2. For each protospacer, the number of the matching spacer in the *N. meningitidis* 8013 CRISPR array is listed. Representative protospacers and 10 flanking nts on both sides are aligned. Potential targets for spacers 16, 23, and 25, cloned from synthetic oligonucleotides, include 10 nts on both sides of the protospacers. Potential targets for spacers 8, 9, 17, and 18, cloned from PCR products, include varying numbers (81-217 nt) of 5' and 3' flanking nts, as indicated in parentheses. As a control, Protospacer 1 is cloned without any flanking sequences. The PAM regions in potential targets are illustrated as well as non-consensus nucleotides. Potential self-targeting spacers also are indicated.

[0025] Figure 11. *Neisseria* Type II-C CRISPR/Cas limits natural transformation of pYZEJS040 plasmids. (A) Schematic representation of integrational vector pYZEJS040, and recombination between *siaA-MCS-CAT-ctrA* region of pYZEJS040 and the capsule locus of meningococcal chromosomes, including: genes of the capsule biosynthesis operon (*siaA-D*) and the capsule transport operon (*ctrA-D*); *CAT* gene; cloning site (MCS); and regions required for maintenance in *E. coli*. Individual elements are not drawn to scale. (B) pYZEJS040 derivatives with potential natural targets for spacers 9 and 25 are tested by natural transformation using wild-type *N. meningitidis* 8013 as recipient strain. Natural transformations are performed and presented as in Figure 4B, except that chloramphenicol-resistant (Cm^R) transformants are scored. Error bars represent s.e.m. for three independent biological replicates. Potential targets for spacers 9 and 25 are identical to those used in Figure 4B.

[0026] Figure 12. Northern analysis of the complete crRNA pool in *N. meningitidis*. Total RNAs from the indicated strains were subjected to northern blot analysis using a probe complementary to 1-22nt of repeat (top). Bottom, the same blot probed for 5S RNA as a loading control. Size markers were indicated.

[0027] Figure 13. Functional NmCas9 can be expressed in mammalian cells. (A) Western blot analysis demonstrates that FLAG-tagged NmCas9 is expressed in 293 cells. Lane 1: Untransfected 293FT cells. Lane 2: 293FT cells transfected with FLAG-tagged NmCas9 expressing plasmid. Upper panel: anti-FLAG tag western. Lower panel: anti-GAPDH western as loading control. (B) Design of the crRNA that targets the tdTomato coding sequence; PAM sequence; crRNA spacer; and crRNA repeat. (C) The plasmid containing the tdTomato coding sequence (see section B) was linearized with *NdeI* and mixed with different combinations of tracrRNA, crRNA and cell lysate prepared from NmCas9-expressing 293FT cells. After incubation at 37°C, DNA was purified and analyzed by agarose gel electrophoresis. The diagram on the right shows the expected cleavage products and their predicted sizes. 'N' indicates inclusion of a nonspecific crRNA that does not target tdTomato. (D) Cleavage product (see Section C) was extracted from the gel and analyzed by Sanger sequencing using the primers indicated in the right panel. The cleavage site, indicated by the arrow, was inferred from the sequencing.

[0028] Figure 14. NmCas9 functions in gene disruption in human ES cells. (A-C) The localization of NmCas9 with an NLS at the N-terminus (A), C-terminus (B), or both termini (C) was analyzed by either EGFP fluorescence (A) or anti-HA immunofluorescence (B and C) in 293 cells. Scale bar: 20 μ m. (D) The localization of NmCas9 with the double NLS (see Section C) was analyzed by anti-HA immunofluorescence in human ES cells. Scale bar: 20 μ m. (E) Design of a single plasmid used for gene editing in hES cells. (F) FACS analysis of tdTomato reporter human ES cell lines after electroporation of the indicated crRNA/tracrRNA/NmCas9 constructs. The number in the plot indicates the percentage of tdTomato-fluorescence-negative cells five days after electroporation. (G) Indels introduced by the targeting CRISPR in the tdTomato negative population (see Section F) were analyzed by targeted PCR amplification and sequencing. The protospacer sequence is underlined. The numbers in parentheses indicate the number of sequenced clones containing that specific indel.

[0029] Figure 15. Specificity screen of NmCas9 system. (A) The top panel shows the crRNA sequence targeting tdTomato, with locations of single point mutations in the spacer region of the tdTomato targeting crRNA. The bottom panel shows the efficiency of each mutant at disrupting tdTomato expression. All mutants were tested as described in Figure 2, Section F. The efficiency is defined as percent tdTomato-negative cells (mutant spacer) divided by percent tdTomato-negative cells (wildtype spacer). Error bars: S.E.M. (B) Top panel depicts the locations of different mutant PAMs in the tdTomato sequence including: wildtype PAM; mutant PAM; and spacer sequence (underlined). For the bottom-most protospacer, the opposite strand was targeted, and the reverse complement sequence is therefore shown. The bottom panel shows the efficiency of targeting at each site associated with the indicated PAM, as revealed by the loss of tdTomato expression. All targeting experiments were performed as described in Figure 2, Section F. The efficiency is defined as percent tdTomato-negative cells (mutant PAM) divided by percent tdTomato-negative cells (wildtype PAM). Error bars: S.E.M.

[0030] Figure 16. Gene targeting in hESCs using NmCas9. (A) Donor DNA and crRNA design. The mismatch in the first nt of crRNA is to satisfy the requirement of the U6 promoter for a G residue at the transcription start site. (B) Phase-contrast (top) and fluorescent (bottom) images of targeted clones from H1, H9 and iPS005 line. Scale bar: 50

μ M. (C) Southern blot analysis of targeted clones of H1, H9 and iPS005 line. Genomic DNA was digested with BamHI. The Southern probe is located outside of donor DNA (see Section A). The wildtype clone should give one band of 4.2 kb and targeted heterozygous clone should give one additional band of 5.6 kb. (D) Targeted clones (see Section B) were treated with 10 μ M SB431542 and 10 ng/ml BMP4 to initiate differentiation. The EGFP signal was analyzed by FACS 3 days after differentiation to identify undifferentiated parental cells before targeting; targeted cells before differentiation; and targeted cells after differentiation.

[0031] Figure 17. Single-guide RNA (sgRNA) is able to direct NmCas9-catalyzed genome editing. *A.* Design of an sgRNA that targets tdTomato including a spacer, crRNA repeat, linker and tracrRNA. *B.* FACS analysis of tdTomato reporter human ES cell lines after electroporation of the sgRNA construct and the nmCas9 expressing plasmid. The number in the plot indicates the percentage of tdTomato-fluorescence-negative cells five days after electroporation.

[0032] Figure 18. NmCas9 D16A functions as a nickase in human pluripotent stem cells. Sequences targeted by the sgRNAs are as indicated and cellular TdTomato fluorescence was analyzed by flow cytometer 5 days after transfection.

[0033] Figure 19. NmCas9 mediates homology-directed repair using either sense or antisense ssODN. The Human ES cell line used here, which expresses both EGFP and a single mutated (deficient, 9nt deleted) copy of Tomato fluorescent protein gene, was transfected with plasmids expressing NmCas9 and gRNA targeting the mutated region of tomato gene with or without 100pmole of 160nt ssODN as repair template. Cellular fluorescence was analyzed by flow cytometry 5 days after transfection..

[0034] Figure 20. NmCas9 can be delivered in the form of mRNA instead of plasmid DNA. Human ES cells expressing both TdTomato and EGFP were transfected with NmCas9 mRNA or an NmCas9 expressing plasmid, together with a plasmid expressing sgRNA that targets tomato gene. Cellular fluorescence was analyzed by flow cytometry 5 days after transfection.

DESCRIPTION

[0035] The present invention is described herein using several definitions, as set forth below and throughout the application.

[0036] Unless otherwise specified or indicated by context, the terms “a”, “an”, and “the” mean “one or more.” For example, “a protein” or “an RNA” should be interpreted to mean “one or more proteins” or “one or more RNAs,” respectively.

[0037] As used herein, “about,” “approximately,” “substantially,” and “significantly” will be understood by persons of ordinary skill in the art and will vary to some extent on the context in which they are used. If there are uses of these terms which are not clear to persons of ordinary skill in the art given the context in which they are used, “about” and “approximately” will mean plus or minus $\leq 10\%$ of the particular term and “substantially” and “significantly” will mean plus or minus $>10\%$ of the particular term.

[0038] As used herein, the terms “include” and “including” have the same meaning as the terms “comprise” and “comprising” in that these latter terms are “open” transitional terms that do not limit claims only to the recited elements succeeding these transitional terms. The term “consisting of,” while encompassed by the term “comprising,” should be interpreted as a “closed” transitional term that limits claims only to the recited elements succeeding this transitional term. The term “consisting essentially of,” while encompassed by the term “comprising,” should be interpreted as a “partially closed” transitional term which permits additional elements succeeding this transitional term, but only if those additional elements do not materially affect the basic and novel characteristics of the claim.

[0039] Disclosed are methods that utilize and kits and compositions that comprise components for RNA-directed DNA cleavage and gene editing. The methods typically utilize and the kits and composition typically comprise a Cas9 protein, or a variant protein thereof, and RNA that hybridizes to a target DNA sequence. The Cas9 protein and RNA typically bind and form a complex with the target DNA sequence. The Cas9 protein may have nuclease activity (*e.g.*, DNase activity and/or RNase activity) and may cleave one

(i.e., nick) or both strands of the target DNA sequence. The term “nick” will be understood as an interruption in the covalent continuity of one strand of a double-stranded nucleic acid molecule. The term “nick” can also describe an enzymatic activity that results in the production of a nick in a nucleic acid molecule. The disclosed methods may be utilized for RNA-directed DNA cleavage *in vitro*, RNA-directed genome editing *in vivo*, and RNA-directed genome binding by Cas9 proteins.

[0040] CRISPR/Cas is a recently discovered, adaptive, sequence-based immune system identified in bacteria and archaea. A “Type II” CRISPR/Cas system from *Streptococcus pyogenes* SF370 has been developed into a simple eukaryotic genome editing tool. This system requires only three components: Cas9 endonuclease, a trans-activating CRISPR RNA (tracrRNA), and the target-specifying crRNA. By fusing the crRNA and tracrRNA into a single transcript referred to as an sgRNA, the machinery can be further streamlined into a two-component system. The target DNA sequence that base-pairs with the crRNA is referred to as the “protospacer.” The two nuclease domains (RuvC and HNH) of Cas9 each cleave one DNA target strand and thus induce a DSB. Cleavage by Cas9 also depends on the presence of a short motif called a protospacer adjacent motif (PAM) that flanks the target region recognized by crRNA base pairing.

[0041] The present inventors have demonstrated that the *Neisseria meningitidis* (Nm) Cas9/crRNA/tracrRNA system can work efficiently for genome editing in human embryonic stem cells (hESCs), leaving behind small insertions and deletions. They have also shown that the NmCas9-induced DSB can serve as a site of transgene insertion. They have mapped the NmCas9 cleavage site to the third and fourth base pairs of the protospacer, at the end closest to the PAM. Importantly, the Cas9/crRNA/tracrRNA system uses the same Cas9 protein and tracrRNA for every targeting event. Only one component – the crRNA – needs to be customized for each individual target, which makes the system very user-friendly.

[0042] In addition to these *in vivo* advances, the inventors have shown that recombinant NmCas9 can be expressed in *E. coli* cells and that it can catalyze crRNA-directed DNA cleavage *in vitro*. This could enable enhanced recombinant DNA capabilities.

[0043] The inventors have also demonstrated that the NmCas9 system, in its native bacterial context, has a novel feature: It can function with long, unprocessed crRNA precursors. In bacterial cells, Type II CRISPR/Cas systems generate pre-crRNAs that are cleaved by a protein called RNase III. In *N. meningitidis*, deletion of the *rnc* gene that encodes RNase III has no deleterious effect on the CRISPR pathway, unlike all other Type II systems examined to date. *In vitro* experiments have confirmed that unprocessed crRNAs can direct DNA cleavage by NmCas9. Together these indicate that NmCas9 can tolerate extensions on its cognate crRNAs without loss of function, which SpCas9 and StCas9 cannot, perhaps enabling expanded functionality by fusing the crRNAs to other useful RNAs such as RNA mimics (see Paige *et al.*, Science, 29 July 2011) and Xist RNA or fragments thereof (see Plath *et al.*, Annu Rev Genet. 2002; 36:233-78 Epub 2002 Jun 11).

[0044] The inventors have also demonstrated that the NmCas9 system has distinct PAM requirements versus Type II CRISPR/Cas systems from different bacteria. For example, for SpCas9 the PAM is 5'-NGG3', while for NmCas9, the PAM is 5'-NNNNGATT3' (in both cases the dash represents the terminal nucleotide of the crRNA-paired sequence). Thus, the presently disclosed methods will open up potential target sites that are not cleavable with existing systems. Also, the specificity for genome editing may increase with a longer PAM.

[0045] The present inventors have identified a novel Cas9 protein. As used herein, the terms "protein" or "polypeptide" or "peptide" may be used interchangeably to refer to a polymer of amino acids. Typically, a "polypeptide" or "protein" is defined as a longer polymer of amino acids, of a length typically of greater than 50, 60, 70, 80, 90, or 100 amino acids. A "peptide" is defined as a short polymer of amino acids, of a length typically of 50, 40, 30, 20 or less amino acids.

[0046] A "protein" as contemplated herein typically comprises a polymer of naturally occurring amino acids (*e.g.*, alanine, arginine, asparagine, aspartic acid, cysteine, glutamine, glutamic acid, glycine, histidine, isoleucine, leucine, lysine, methionine, phenylalanine, proline, serine, threonine, tryptophan, tyrosine, and valine). The proteins contemplated herein may be further modified *in vitro* or *in vivo* to include non-amino acid

moieties. These modifications may include but are not limited to acylation (*e.g.*, O-acylation (esters), N-acylation (amides), S-acylation (thioesters)), acetylation (*e.g.*, the addition of an acetyl group, either at the N-terminus of the protein or at lysine residues), formylation lipoylation (*e.g.*, attachment of a lipoate, a C8 functional group), myristoylation (*e.g.*, attachment of myristate, a C14 saturated acid), palmitoylation (*e.g.*, attachment of palmitate, a C16 saturated acid), alkylation (*e.g.*, the addition of an alkyl group, such as an methyl at a lysine or arginine residue), isoprenylation or prenylation (*e.g.*, the addition of an isoprenoid group such as farnesol or geranylgeraniol), amidation at C-terminus, glycosylation (*e.g.*, the addition of a glycosyl group to either asparagine, hydroxylysine, serine, or threonine, resulting in a glycoprotein). Distinct from glycation, which is regarded as a nonenzymatic attachment of sugars, polysialylation (*e.g.*, the addition of polysialic acid), glypiation (*e.g.*, glycosylphosphatidylinositol (GPI) anchor formation, hydroxylation, iodination (*e.g.*, of thyroid hormones), and phosphorylation (*e.g.*, the addition of a phosphate group, usually to serine, tyrosine, threonine or histidine).

[0047] The Cas9 proteins disclosed herein may include “wild type” Cas9 protein and variants, mutants, and derivatives thereof. As used herein the term “wild type” is a term of the art understood by skilled persons and means the typical form of an organism, strain, gene or characteristic as it occurs in nature as distinguished from mutant or variant forms. As used herein, a “variant,” “mutant,” or “derivative” refers to a protein molecule having an amino acid sequence that differs from a reference protein or polypeptide molecule. A variant or mutant may have one or more insertions, deletions, or substitutions of an amino acid residue relative to a reference molecule. A variant or mutant may include a fragment of a reference molecule. For example, a Cas9 mutant or variant molecule may have one or more insertions, deletions, or substitution of at least one amino acid residue relative to the Cas9 full-length polypeptide. The sequence of the full-length Cas9 protein from *Neisseria meningitidis* is presented as SEQ ID NO:1 and may be used as a reference in this regard.

[0048] Regarding proteins, a “deletion” refers to a change in the amino acid sequence that results in the absence of one or more amino acid residues. A deletion may remove at least 1, 2, 3, 4, 5, 10, 20, 50, 100, 200, or more amino acids residues. A

deletion may include an internal deletion and/or a terminal deletion (*e.g.*, an N-terminal truncation, a C-terminal truncation or both of a reference polypeptide).

[0049] Regarding proteins, “fragment” is a portion of an amino acid sequence which is identical in sequence to but shorter in length than a reference sequence. A fragment may comprise up to the entire length of the reference sequence, minus at least one amino acid residue. For example, a fragment may comprise from 5 to 1000 contiguous amino acid residues of a reference polypeptide, respectively. In some embodiments, a fragment may comprise at least 5, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, 150, 250, or 500 contiguous amino acid residues of a reference polypeptide. Fragments may be preferentially selected from certain regions of a molecule. The term “at least a fragment” encompasses the full length polypeptide. A fragment of a Cas9 protein may comprise or consist essentially of a contiguous portion of an amino acid sequence of the full-length Cas9 protein (SEQ ID NO:1). A fragment may include an N-terminal truncation, a C-terminal truncation, or both truncations relative to the full-length Cas9 protein.

[0050] Regarding proteins, the words “insertion” and “addition” refer to changes in an amino acid sequence resulting in the addition of one or more amino acid residues. An insertion or addition may refer to 1, 2, 3, 4, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 150, 200, or more amino acid residues. A variant of a Cas9 protein may have N-terminal insertions, C-terminal insertions, internal insertions, or any combination of N-terminal insertions, C-terminal insertions, and internal insertions.

[0051] Regarding proteins, the phrases “percent identity” and “% identity,” refer to the percentage of residue matches between at least two amino acid sequences aligned using a standardized algorithm. Methods of amino acid sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail below, generally preserve the charge and hydrophobicity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide. Percent identity for amino acid sequences may be determined as understood in the art. (*See, e.g.*, U.S. Patent No. 7,396,664, which is incorporated herein by reference in its entirety). A suite of commonly used and freely available sequence comparison algorithms is provided by the National Center for

Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST), which is available from several sources, including the NCBI, Bethesda, Md., at its website. The BLAST software suite includes various sequence analysis programs including “blastp,” that is used to align a known amino acid sequence with other amino acids sequences from a variety of databases. As described herein, variants, mutants, or fragments (*e.g.*, a Cas9 protein variant, mutant, or fragment thereof) may have 99%, 98%, 97%, 96%, 95%, 94%, 93%, 92%, 91%, 90%, 80%, 70%, 60%, or 50% amino acid sequence identity relative to a reference molecule (*e.g.*, relative to the Cas9 full-length polypeptide (SEQ ID:1)).

[0052] Regarding proteins, percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

[0053] Regarding proteins, the amino acid sequences of variants, mutants, or derivatives as contemplated herein may include conservative amino acid substitutions relative to a reference amino acid sequence. For example, a variant, mutant, or derivative protein may include conservative amino acid substitutions relative to a reference molecule. “Conservative amino acid substitutions” are those substitutions that are a substitution of an amino acid for a different amino acid where the substitution is predicted to interfere least with the properties of the reference polypeptide. In other words, conservative amino acid substitutions substantially conserve the structure and the function of the reference polypeptide. The following table provides a list of exemplary conservative amino acid substitutions which are contemplated herein:

Original Residue	Conservative Substitution
Ala	Gly, Ser
Arg	His, Lys
Asn	Asp, Glu, His
Asp	Asn, Glu
Cys	Ala, Ser
Gln	Asn, Glu, His
Glu	Asp, Gln, His
Gly	Ala
His	Asn, Arg, Glu, Lys
Ile	Leu, Val
Leu	Ile, Val
Lys	Arg, Gln, His
Met	Leu, Ile
Phe	His, Met, Leu, Trp, Tyr
Ser	Cys, Thr
Thr	Ser, Val
Trp	Phe, Tyr
Tyr	His, Phe, Trp
Val	Ile, Leu, Thr

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

[0054] The disclosed proteins, mutants, variants, or described herein may have one or more functional or biological activities exhibited by a reference polypeptide (*e.g.*, one or more functional or biological activities exhibited by wild-type Cas9 protein). For example, the disclosed Cas9 proteins, mutants, variants, or derivatives thereof may have one or more biological activities that include: binding to a single-stranded RNA, binding to a double-stranded RNA, binding to a target polynucleotide sequence, nicking a single strand of the target DNA sequence, and/or cleaving both strands of the target DNA sequence.

[0055] The disclosed Cas9 proteins may be substantially isolated or purified. The term “substantially isolated or purified” refers to amino acid sequences that are removed

from their natural environment, and are at least 60% free, preferably at least 75% free, and more preferably at least 90% free, even more preferably at least 95% free from other components with which they are naturally associated.

[0056] Also disclosed herein are polynucleotides, for example polynucleotide sequences that encode Cas9 proteins (*e.g.*, DNA that encodes a polypeptide having the amino acid sequence of SEQ ID NO:1 or a polypeptide variant having an amino acid sequence with at least about 50%, 60%, 70%, 80%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity to SEQ ID NO:1; DNA comprising the polynucleotide sequence of SEQ ID NO:2; DNA comprising the polynucleotide sequence of SEQ ID NO:3; or Cas9 mRNA). Other polynucleotides contemplated herein are RNAs that direct Cas9-mediated binding, nicking, and/or cleaving of a target DNA sequence (*e.g.*, tracrRNA, crRNA, sgRNA) and DNA that encodes such RNAs. The terms “polynucleotide,” “polynucleotide sequence,” “nucleic acid” and “nucleic acid sequence” refer to a nucleotide, oligonucleotide, polynucleotide (which terms may be used interchangeably), or any fragment thereof. These phrases also refer to DNA or RNA of genomic, natural, or synthetic origin (which may be single-stranded or double-stranded and may represent the sense or the antisense strand).

[0057] Regarding polynucleotide sequences, the terms “percent identity” and “% identity” refer to the percentage of residue matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences. Percent identity for a nucleic acid sequence may be determined as understood in the art. (*See, e.g.*, U.S. Patent No. 7,396,664, which is incorporated herein by reference in its entirety). A suite of commonly used and freely available sequence comparison algorithms is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST), which is available from several sources, including the NCBI, Bethesda, Md., at its website. The BLAST software suite includes various sequence analysis programs including “blastn,” that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called “BLAST 2

Sequences” that is used for direct pairwise comparison of two nucleotide sequences. “BLAST 2 Sequences” can be accessed and used interactively at the NCBI website. The “BLAST 2 Sequences” tool can be used for both blastn and blastp (discussed above).

[0058] Regarding polynucleotide sequences, percent identity may be measured over the length of an entire defined polynucleotide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

[0059] Regarding polynucleotide sequences, “variant,” “mutant,” or “derivative” may be defined as a nucleic acid sequence having at least 50% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the “BLAST 2 Sequences” tool available at the National Center for Biotechnology Information’s website. (See Tatiana A. Tatusova, Thomas L. Madden (1999), “Blast 2 sequences - a new tool for comparing protein and nucleotide sequences”, FEMS Microbiol Lett. 174:247-250). Such a pair of nucleic acids may show, for example, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length.

[0060] Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code where multiple codons may encode for a single amino acid. It is understood that changes in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein. For example, polynucleotide sequences as contemplated herein may encode a Cas9 protein and may be codon-optimized for expression in a particular host. In the art, codon usage frequency

tables have been prepared for a number of host organisms including humans, mouse, rat, pig, *E. coli*, plants, and other host cells.

[0061] A “recombinant nucleic acid” is a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques known in the art. The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

[0062] “Substantially isolated or purified” nucleic acid or amino acid sequences are contemplated herein. The term “substantially isolated or purified” refers to nucleic acid or amino acid sequences that are removed from their natural environment, and are at least 60% free, preferably at least 75% free, and more preferably at least 90% free, even more preferably at least 95% free from other components with which they are naturally associated.

[0063] “Transformation” or “transfected” describes a process by which exogenous nucleic acid (e.g., DNA or RNA) is introduced into a recipient cell. Transformation or transfection may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation or transfection is selected based on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection, electroporation, heat shock, lipofection, and particle bombardment.

[0064] Methods of non-viral delivery of nucleic acids include lipofection, nucleofection, microinjection, electroporation, biolistics, virosomes, liposomes, immunoliposomes, polycation or lipid:nucleic acid conjugates, naked DNA, artificial virions, and agent-enhanced uptake of DNA. Lipofection is described in e.g., U.S. Pat.

Nos. 5,049,386, 4,946,787; and 4,897,355) and lipofection reagents are sold commercially (e.g., Transfectam.TM. and Lipofectin.TM.). Cationic and neutral lipids that are suitable for efficient receptor-recognition lipofection of polynucleotides include those of Felgner, WO 91/17424; WO 91/16024. Delivery can be to cells (e.g. in vitro or ex vivo administration) or target tissues (e.g. in vivo administration). The term "transformed cells" or "transfected cells" includes stably transformed or transfected cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed or transfected cells which express the inserted DNA or RNA for limited periods of time.

[0065] The polynucleotide sequences contemplated herein may be present in expression vectors. For example, the vectors may comprise: (a) a polynucleotide encoding an ORF of a Cas9 protein; (b) a polynucleotide that expresses an RNA that directs Cas9-mediated binding, nicking, and/or cleaving of a target DNA sequence; and both (a) and (b). The polynucleotide present in the vector may be operably linked to a prokaryotic or eukaryotic promoter. "Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame. Vectors contemplated herein may comprise a heterologous promoter (e.g., a eukaryotic or prokaryotic promoter) operably linked to a polynucleotide that encodes a Cas9 protein. A "heterologous promoter" refers to a promoter that is not the native or endogenous promoter for the protein or RNA that is being expressed. For example, a heterologous promoter for a Cas9 protein of *Neisseria meningitidis* may include a eukaryotic promoter or a prokaryotic promoter that is not the native, endogenous promoter for the Cas9 protein of *Neisseria meningitidis*.

[0066] As used herein, "expression" refers to the process by which a polynucleotide is transcribed from a DNA template (such as into and mRNA or other RNA transcript) and/or the process by which a transcribed mRNA is subsequently translated into peptides, polypeptides, or proteins. Transcripts and encoded polypeptides

may be collectively referred to as "gene product." If the polynucleotide is derived from genomic DNA, expression may include splicing of the mRNA in a eukaryotic cell.

[0067] The term "vector" refers to some means by which nucleic acid (e.g., DNA) can be introduced into a host organism or host tissue. There are various types of vectors including plasmid vector, bacteriophage vectors, cosmid vectors, bacterial vectors, and viral vectors. As used herein, a "vector" may refer to a recombinant nucleic acid that has been engineered to express a heterologous polypeptide (e.g., a Cas9 protein). The recombinant nucleic acid typically includes *cis*-acting elements for expression of the heterologous polypeptide.

[0068] Any of the conventional vectors used for expression in eukaryotic cells may be used for directly introducing DNA into a subject. Expression vectors containing regulatory elements from eukaryotic viruses may be used in eukaryotic expression vectors (e.g., vectors containing SV40, CMV, or retroviral promoters or enhancers). Exemplary vectors include those that express proteins under the direction of such promoters as the SV40 early promoter, SV40 later promoter, metallothionein promoter, human cytomegalovirus promoter, murine mammary tumor virus promoter, and Rous sarcoma virus promoter. Expression vectors as contemplated herein may include prokaryotic control sequences that modulate expression of a heterologous protein (e.g. Cas9 protein. Prokaryotic expression control sequences may include constitutive or inducible promoters (e.g., T3, T7, Lac, trp, or phoA), ribosome binding sites, or transcription terminators.

[0069] The vectors contemplated herein may be introduced and propagated in a prokaryote, which may be used to amplify copies of a vector to be introduced into a eukaryotic cell or as an intermediate vector in the production of a vector to be introduced into a eukaryotic cell (e.g. amplifying a plasmid as part of a viral vector packaging system). A prokaryote may be used to amplify copies of a vector and express one or more nucleic acids, such as to provide a source of one or more proteins for delivery to a host cell or host organism. Expression of proteins in prokaryotes may be performed using *Escherichia coli* with vectors containing constitutive or inducible promoters directing the expression of either a Cas9 protein or a fusion protein comprising a Cas9 protein or a fragment thereof. Fusion vectors add a number of amino acids to a protein encoded

therein, such as to the amino terminus of the recombinant protein. Such fusion vectors may serve one or more purposes, such as: (i) to increase expression of recombinant protein; (ii) to increase the solubility of the recombinant protein; (iii) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification (*e.g.*, a His tag); (iv) to tag the recombinant protein for identification (*e.g.*, such as Green fluorescence protein (GFP) or an antigen (*e.g.*, an HA tag such as SEQ ID NOs:6 and 7) that can be recognized by a labelled antibody); (v) to promote localization of the recombinant protein to a specific area of the cell (*e.g.*, where the Cas9 protein is fused (*e.g.*, at its N-terminus or C-terminus) to a nuclear localization signal (NLS) which may include the NLS of SV40 (*e.g.*, SEQ ID NOs:4 and 5, which is a monopartite NLS), nucleoplasmin (which comprises a bipartite signal of two clusters of basic amino acids separated by a spacer of about 10 amino acids), C-myc, M9 domain of hnRNP A1, or a synthetic NLS (*e.g.*, SEQ ID NOs:8 and 9)). The importance of neutral and acidic amino acids in NLS have been studied. (*See Makkerh et al. (1996) Curr Biol 6(8):1025-1027*). Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase.

[0070] The presently disclosed methods may include delivering one or more polynucleotides, such as or one or more vectors as described herein, one or more transcripts thereof, and/or one or more proteins transcribed therefrom, to a host cell. Further contemplated are host cells produced by such methods, and organisms (such as animals, plants, or fungi) comprising or produced from such cells. In some embodiments, a CRISPR enzyme (*e.g.*, Cas9 protein) in combination with (and optionally complexed with) a guide sequence is delivered to a cell. Conventional viral and non-viral based gene transfer methods can be used to introduce nucleic acids in mammalian cells or target tissues. Such methods can be used to administer nucleic acids encoding components of a CRISPR system to cells in culture, or in a host organism. Non-viral vector delivery systems include DNA plasmids, RNA (*e.g.* a transcript of a vector described herein), naked nucleic acid, and nucleic acid complexed with a delivery vehicle, such as a

liposome. Viral vector delivery systems include DNA and RNA viruses, which have either episomal or integrated genomes after delivery to the cell.

[0071] In the methods contemplated herein, a host cell may be transiently or non-transiently transfected (*i.e.*, stably transfected) with one or more vectors described herein. In some embodiments, a cell is transfected as it naturally occurs in a subject (*i.e.*, in situ). In some embodiments, a cell that is transfected is taken from a subject (*i.e.*, explanted). In some embodiments, the cell is derived from cells taken from a subject, such as a cell line. Suitable cells may include stem cells (*e.g.*, embryonic stem cells and pluripotent stem cells). A cell transfected with one or more vectors described herein may be used to establish a new cell line comprising one or more vector-derived sequences. In the methods contemplated herein, a cell may be transiently transfected with the components of a CRISPR system as described herein (such as by transient transfection of one or more vectors, or transfection with RNA), and modified through the activity of a CRISPR complex, in order to establish a new cell line comprising cells containing the modification but lacking any other exogenous sequence.

[0072] The methods, vectors, and systems described herein may be used to produce a non-human transgenic animal or a transgenic plant or algae. Transgenic animals may include a mammal, such as a mouse, rat, or rabbit. Methods for producing transgenic plants and animals are known in the art, and generally begin with a method of cell transfection with one or more vectors as contemplated herein.

EXAMPLES

[0073] The following Examples are illustrative and are not intended to limit the scope of the claimed subject matter.

[0074] Example 1

[0075] Reference is made to Zhang *et al.*, "Processing-Independent CRISPR RNAs Limit Natural Transformation in *Neisseria meningitidis*" *Molecular Cell* 50, 488-503, May 23, 2013, the contents of which are incorporated herein by reference.

[0076] Abstract

[0077] CRISPR interference confers adaptive, sequence-based immunity against viruses and plasmids and is specified by CRISPR RNAs (crRNAs) that are transcribed and processed from spacer-repeat units. Pre-crRNA processing is essential for CRISPR interference in all systems studied thus far. Here, our studies of crRNA biogenesis and CRISPR interference in naturally competent *Neisseria* spp. reveal a unique crRNA maturation pathway in which crRNAs are transcribed from promoters that are embedded within each repeat, yielding crRNA 5' ends formed by transcription and not by processing. Although crRNA 3' end formation involves RNase III and *trans*-encoded tracrRNA, as in other Type II CRISPR systems, this processing is dispensable for interference. The meningococcal pathway is the most streamlined CRISPR/*cas* system characterized to date. Endogenous CRISPR spacers limit natural transformation, which is the primary source of genetic variation that contributes to immune evasion, antibiotic resistance, and virulence in the human pathogen *N. meningitidis*. Highlights of these new CRISPRs include the following: unlike previously described CRISPRs, each *Neisseria* repeat carries its own promoter; pre-crRNA processing is dispensable for CRISPR interference in *Neisseria* spp; CRISPR interference blocks natural transformation in the pathogen *N. meningitidis*; and *Neisseria* CRISPR/Cas systems are the most streamlined observed to date

[0078] Introduction

[0079] Clustered, regularly interspaced, short palindromic repeat (CRISPR) loci confer sequence-based, adaptive immunity against virus infection and plasmid conjugation in bacteria and archaea (Haft et al., 2005; Makarova et al., 2006; Barrangou et al., 2007; Terns and Terns, 2011; Wiedenheft et al., 2012). CRISPRs consist of short repeats separated by similarly sized, non-repetitive sequences called spacers, which are derived from previously encountered invasive sequences such as viral genomes or plasmids (Bolotin et al., 2005; Mojica et al., 2005; Pourcel et al., 2005). CRISPR loci are transcribed as long CRISPR RNA (crRNA) precursors that are processed into small crRNAs (Brouns et al., 2008; Hale et al., 2008). Pre-crRNA transcription is generally driven by promoters within "leader" sequences outside of the CRISPR array. The crRNAs assemble with CRISPR-associated (Cas) proteins into complexes that cleave complementary "protospacer" sequences within invasive nucleic acids, a phenomenon

known as CRISPR interference (Karginov and Hannon, 2010; Marraffini and Sontheimer, 2010; Terns and Terns, 2011; Wiedenheft et al., 2012). The sequence information in crRNAs is used to guide Cas complexes to their targets on viruses and plasmids, leading to their destruction (Barrangou et al., 2007; Brouns et al., 2008; Marraffini and Sontheimer, 2008; Hale et al., 2009; Garneau et al., 2010; Westra et al., 2012). Most CRISPR/Cas systems cleave incoming DNAs (Marraffini and Sontheimer, 2008; Garneau et al., 2010; Westra et al., 2012), though RNA-cleaving systems have also been identified (Hale et al., 2009, 2012; Zhang et al., 2012).

[0080] CRISPR/Cas systems have been classified into types I, II and III based primarily on their *cas* gene complement (Makarova et al., 2011a). Common to all of these three types is that the CRISPR array is transcribed as a multimeric pre-crRNA that is processed into crRNAs that each contain an individual spacer flanked on one or both sides by partial repeat sequences (Bhaya et al., 2011). However, the molecular events underlying processing dramatically differ. Whereas in Type I and III systems the processing enzymes are encoded within the CRISPR/*cas* locus, Type II systems use the host enzyme RNase III (encoded by the *rnc* gene) and a noncoding RNA called tracrRNA (Deltcheva et al., 2011). In *Streptococcus pyogenes* SF370, an *rnc* mutant abolishes the function of a Type II CRISPR/*cas* locus, indicating that pre-crRNA processing is essential (Deltcheva et al., 2011).

[0081] The importance of the Type II CRISPR/Cas pathway has been dramatically enhanced by its development into a system for RNA-guided DNA cleavage *in vitro* (Jinek et al., 2012) and genome editing *in vivo* (Jinek et al., 2013; Cho et al., 2013; Cong et al., 2013; DiCarlo et al., 2013; Hwang et al., 2013; Jiang et al., 2013; Mali et al., 2013). Our ability to exploit this new technology further will depend on a deeper understanding of the underlying molecular mechanisms, and will be increased by the characterization of systems that are as simplified and streamlined as possible. Type II CRISPR/Cas loci, which are found in bacteria but not archaea, usually contain four *cas* genes: *cas1*, *cas2*, *cas9*, and either *csn2* (subtype II-A) or *cas4* (subtype II-B) (Makarova et al., 2011b). Cas9 is the effector protein for the interference function of existing spacer sequences in the CRISPR array (Sapranaukas et al., 2011; Gasimmas et al., 2012; Jinek et al., 2012), whereas the other proteins are thought to be involved in the acquisition of new CRISPR

spacers. The tracrRNA is essential for crRNA-directed DNA recognition and Cas9-catalyzed DNA cleavage *in vitro*, even with crRNAs that bypass processing (Jinek et al., 2012). DNA targeting in both Type I and Type II systems requires not only crRNA/target complementarity but also a protospacer adjacent motif (PAM), which is a short (2-5 nt), conserved sequence just outside of the protospacer (Deveau et al., 2008; Horvath et al., 2008; Mojica et al., 2009; Semenova et al., 2011; Sashital et al., 2012).

[0082] Although CRISPR interference was originally defined as a phage defense pathway, CRISPR/Cas systems are now understood to play a broader role in limiting horizontal gene transfer (HGT) (Marraffini and Sontheimer, 2008). The three primary routes of HGT are transformation, conjugation, and phage transduction, and the latter two are well established as being subject to interference by naturally occurring spacers. Artificial means of transformation (e.g. electroporation) can also be blocked by CRISPR interference (Marraffini and Sontheimer, 2008; Deltcheva et al., 2011; Sapranaukas et al., 2011; Semenova et al., 2011), though natural transformation uses a very different DNA uptake process (Chen et al., 2005). An engineered spacer can prevent natural transformation specified by an *S. pyogenes* CRISPR/*cas* locus transplanted into *Streptococcus pneumonia* (Bikard et al., 2012). However, although this artificial system suggests that natural CRISPR/Cas contexts may do likewise, the fundamental question of whether native CRISPR/Cas systems play a role in natural transformation remains to be addressed.

[0083] Strains from the genus *Neisseria* serve as paradigms for natural transformation, as they are competent during all phases of their life cycle and use this process for frequent genetic exchange (Hamilton and Dillard, 2006). Although functional CRISPR/*cas* systems have not been identified in *Neisseria gonorrhoeae*, some strains of *Neisseria lactamica* and *Neisseria meningitidis* carry apparent Type II CRISPR/*cas* loci (Grissa et al., 2007). Meningococci are obligate human commensals that can invade the bloodstream and cerebrospinal fluid (Bratcher et al., 2012), and meningococcal disease is responsible for tens of thousands of deaths per year (Stephens et al., 2007).

[0084] Here we characterize the CRISPR pathway in neisseriae and find that it exhibits several unique features, most notably a streamlined functional architecture that

includes a previously unknown, processing-independent mode of crRNA biogenesis. Furthermore, naturally occurring spacers match sequences from other *Neisseria* genomes, including a prophage-like meningococcal disease-associated (MDA) island that correlates with invasiveness and pathogenicity (Bille et al., 2005, 2008). We find that a native meningococcal CRISPR/*cas* locus prevents natural transformation of spacer-matched sequences, suggesting that it can limit the horizontal spread of virulence genes.

[0085] Results

[0086] dRNA-seq reveals that each repeat in the *Neisseria* CRISPR carries its own promoter. We analysed all 19 sequenced *Neisseria* genomes available in the NCBI database (fifteen from *N. meningitidis*, three from *N. gonorrhoeae*, and one from *N. lactamica*) using CRISPRFinder (Grissa et al., 2007) or CRISPRi (Rousseau et al., 2009). We identified seven putative Type II CRISPR/*cas* loci: six in *N. meningitidis* strains, and one in *N. lactamica* 020-06. All were highly similar, and unlike other Type II loci characterized previously (Barrangou et al., 2007; Deltcheva et al., 2011; Magadán et al., 2012), included a set of only three predicted protein-coding genes (*cas9*, *cas1* and *cas2*) but neither *csn2* nor *cas4*. To examine the expression status of a representative locus, we performed our recently developed dRNA-seq approach (Sharma et al., 2010) on *N. lactamica* 020-06. We prepared two cDNA libraries from total RNA using a strategy that allows us to distinguish between transcripts with either primary or processed 5' ends: one library is generated from untreated RNA, whereas the other is treated with terminator exonuclease (TEX), which specifically degrades RNAs with 5'-monophosphate ends (including the abundant rRNAs and tRNAs) that are formed by processing. Primary transcripts with 5'-triphosphate ends survive TEX treatment, resulting in their relative enrichment in the TEX+ library.

[0087] Our global mapping of cDNA reads identified a tracrRNA and small crRNAs as highly abundant classes of transcripts (Figure 1A), which suggests a highly active CRISPR/Cas system. Reads obtained from the tracrRNA locus reveal the existence of two tracrRNA forms [107 nucleotides (nt) and 91 nt] (Figure 1B). The treatment with TEX eliminated the shorter tracrRNA reads, which indicates that they are products of processing as observed in a Type II-A CRISPR system (Deltcheva et al., 2011). These

sequencing results were confirmed by Northern analyses (Figure 1C). In striking contrast, crRNAs were enriched rather than depleted in the TEX-treated libraries (Figure 1D), suggesting that formation of many if not most crRNA 5' ends is processing-independent. TEX treatment did affect the relative amounts of individual crRNAs; for example, crRNAs from spacers 4, 6 and 9 exhibited very similar read counts in the untreated sample but varied over a 5-fold range in the TEX-treated sample. These results suggest that the crRNA pool contains some 5'-monophosphorylated crRNAs, the fraction of which varies from spacer to spacer. Northern blots confirmed the resistance of a population of mature, ~48-nt crRNAs, as well as a subset of longer crRNA precursors, to TEX digestion (Figure 1E). This crRNA profile indicates a different mode of crRNA biogenesis from that reported for other Type II systems (Deltcheva et al., 2011). Intriguingly, we noted that the terminal nine nucleotides of each CRISPR repeat exhibit sequence hallmarks of an extended -10 box promoter element (Hook-Barnard and Hinton, 2007) and that the TEX-resistant crRNA 5' ends map 9-10 nt downstream from the center of each such element (Figure 1D).

[0088] crRNA biogenesis in *Neisseria lactamica* depends on single promoter elements in each CRISPR repeat. The dRNA-seq results and -10 box similarity suggest that in *N. lactamica* 020-06, each CRISPR repeat carries its own minimal promoter, and that pre-crRNA transcription initiates independently within each spacer. As an initial test of this hypothesis, we designed a series of transcriptional green fluorescent protein (*gfp*) fusion constructs containing either single or multiple CRISPR repeats, introduced these constructs into *E. coli*, and analysed cellular GFP fluorescence. As shown in Figure 2A, the *gfp* fusion construct with a wild-type CRISPR repeat led to robust cellular fluorescence, whereas a two-nt substitution in the extended -10 promoter TG motif (Hook-Barnard and Hinton, 2007) reduced *gfp* expression to background levels. Constructs with *gfp* fused to three CRISPR repeat-spacer units increased the fluorescence signal almost two-fold, consistent with the possibility of increased transcription from the multiple repeats. We obtained similar results using repeat sequences from a Type II CRISPR/*cas* locus in *Campylobacter jejuni* NCTC11168 (Figure 2A); this locus, like that of *N. lactamica* 020-06, also contains only *cas9*, *cas1* and *cas2* and has CRISPR repeats that include an extended -10 box consensus. Thus, our *gfp* reporter assays prove that the putative promoter elements in each *N. lactamica* repeat are indeed active in bacterial cells,

and are likely also present in some other Type II CRISPR/*cas* loci with similar, minimized *cas* gene content (Figure 2B and Table 1). The *Neisseria* and *Campylobacter* systems are also unusual in that the crRNAs and *cas* genes are transcribed in the opposite direction, and the CRISPR arrays lack recognizable “leader” sequences with external promoters. Based on these considerations as well as independent phylogenetic analyses (Koonin and Makarova, 2013; Chylinski et al., 2013), we now consider these variant Type II CRISPR/Cas loci as members of a distinct and newly defined subtype, Type II-C (Figure 2B).

[0089] To obtain additional proof that each *N. lactamica* CRISPR repeat carries its own promoter element, we used purified *E. coli* σ^{70} RNA polymerase (RNAP) holoenzyme in *in vitro* transcription assays with linear DNA templates containing either a wild-type or a mutant repeat. A transcript of the expected length (168 nt) was obtained with the wild-type CRISPR repeat template (and with a control -10/-35 promoter construct), but not with the mutated repeat (Figures 2C and 8). This result demonstrated that the extended -10 motif in the *N. lactamica* CRISPR repeat was recognized even by the heterologous *E. coli* σ^{70} RNAP holoenzyme. These data show that crRNAs in *N. lactamica* exhibit a unique mode of biogenesis involving transcription from extended -10 class promoters located within CRISPR repeats.

[0090] RNase III is involved in 3' end formation of *Neisseria* crRNAs. The observation that crRNA 5' ends correspond to sites of transcription initiation in *N. lactamica* suggests a reduced dependence on processing relative to other CRISPR systems. To determine whether this reduced dependence extends to crRNA 3' end formation as well or if 3' processing still occurs, and to extend our studies to other *Neisseria* strains, we deleted the *rnc* gene (which encodes RNase III) in *N. meningitidis* WUE2594, and then compared the tracrRNA and crRNA populations from this Δrnc mutant with wild-type bacteria by northern analysis. As shown in Figure 2D, the Δrnc strain exhibited a complete loss of the shorter (91 nt), TEX-sensitive tracrRNA, indicating that the 5' end of this RNA is generated by RNase III-dependent processing as observed previously in a Type II-A system (Deltcheva et al., 2011). We also observed dramatic differences in the crRNA population in the Δrnc mutant: the 48-nt population is virtually abolished, and longer pre-crRNAs accumulate. This result strongly suggests that *Neisseria*

spp. crRNA 3' end formation depends upon RNase III rather than direct transcription termination.

[0091] Repeat/spacer organization and potential targets of *Neisseria* Type II-C CRISPR loci. Having defined unique features of CRISPR/Cas systems in *neisseriae*, we turned our attention towards functional analyses, beginning with an examination of CRISPR organization and spacer content. Of the 103 spacers found in the seven CRISPR-positive genomes (Figure 3A and Table 2), one is 29 nt long while all others are 30 nt. All seven CRISPRs have the same 36-nt repeat consensus, with only a few repeats that deviate from this consensus (Figure 3A and Table 3). Intriguingly, the polarity of spacer conservation is opposite to that generally observed in other CRISPR loci. Conserved spacers that are shared among multiple strains in *Neisseria* spp. (color-coded in Figure 3A) are enriched at the upstream end of the array (relative to the direction of transcription) but are far less common at the downstream end. In contrast, other CRISPRs described thus far have the most recently derived and therefore least conserved spacers at the upstream end, i.e., proximal to the promoter (Makarova et al., 2011b). This observation suggests that new spacer acquisition in *neisseriae*, and perhaps in other Type II-C loci, occurs at the downstream end. In addition, deviations from the repeat consensus are most common at the upstream end in *neisseriae* (Table 3), whereas other CRISPRs most frequently exhibit the opposite tendency. Although technical limitations have thus far precluded direct tests of spacer acquisition in *neisseriae*, several considerations make it likely that these CRISPRs are competent for the adaptive component of the pathway. First, Type II-C loci are relatively common, with many more apparent examples among sequenced bacterial genomes than Type II-B loci (Chylinski et al., 2013). It is unlikely that type II-C systems would be so widespread if they were unable to provide adaptive protection to their hosts. Second, despite the conservation of the Type II-C loci in closely related *Neisseria* strains, there are many differences in spacer content (Figure 3A). This implies that these CRISPR loci are fluid, i.e. can adapt by adding and deleting spacers. Third, reports from other systems indicate that *cas1* and *cas2* can suffice for spacer acquisition (Yosef et al., 2012). It is therefore plausible that *cas1* and *cas2* likewise suffice for adaptation in Type II-C systems.

[0092] BLASTN searches with the 83 unique spacer sequences for similar sequences in the NCBI database allowed us to identify at least one potential target for 35 (~42%) of them. For simplicity we required either a perfect match, or at most a single mismatch in the 10 protospacer nucleotides furthest from the PAM (i.e., well outside of the presumptive “seed” sequence that requires perfect complementarity for interference) (Sapranaukas et al., 2011; Semenova et al., 2011; Wiedenheft et al., 2011; Jinek et al., 2012). Figure 3B shows representative protospacer sequences that match CRISPR spacers from *N. meningitidis* strain 8013; representative protospacers that match spacers from all strains are shown in Figure 9. Protospacer alignments reveal an apparent PAM of 5'-NNNNGATT-3' (Figures 3B and 9). Of the 325 distinct candidate protospacers that match these 35 CRISPR spacers, all are in *Neisseria* sequences: 248 (76%) in *N. meningitidis* genomes, 69 (21%) in *N. gonorrhoeae* genomes, 6 (2%) in *N. lactamica* plasmids, and 1 each in *N. meningitidis* and *N. flavescens* plasmids. In some cases (shown in Figures 3B and 9), potential protospacers are present in the same genome as the targeting CRISPR spacer, suggesting the potential for autoimmunity (Stern et al., 2010). However, these protospacers all include significant deviations from the PAM consensus, perhaps explaining the apparent lack of self-targeting that is implied by the persistence of the matching protospacer. Intriguingly, 22 out of 35 CRISPR spacers (shown in Figures 3B and 9) with identifiable targets match at least one potential prophage sequence (Table 4), including the meningococcal disease-associated (MDA) island that is associated with invasiveness and pathogenicity in young adult patients (Bille et al., 2005, 2008).

[0093] CRISPR interference blocks natural transformation in *N. meningitidis*. The preponderance of protospacers in *Neisseria* spp. genomes suggests that the CRISPR/*cas* loci could interfere with natural transformation. For our functional analyses addressing this possibility, we focused on *N. meningitidis* 8013, primarily because it exhibits the most robust transformation competence in our hands. For transformation assays we used the vector pGCC2, which contains an erythromycin resistance gene (*ermC*) and polylinker inserted into sequences from the gonococcal *lctP/aspC* locus (Figure 4A). Upon transformation into *N. meningitidis*, homologous recombination into the meningococcal *lctP/aspC* locus leads to *ermC* insertion and erythromycin-resistant (*Erm*^R) transformants. We selected potential natural target sequences, including ten or

more nucleotides on both sides of the protospacer, for seven of the spacers of *N. meningitidis* 8013, namely spacers 8, 9, 16, 17, 18, 23, and 25 (Figures 3B and 10). Upon cloning them into pGCC2, the resulting vectors were used in liquid-medium natural transformation assays into wild-type 8013 cells, and transformation frequencies (antibiotic-resistant cfu ml⁻¹/total cfu ml⁻¹) were determined. For comparison we also cloned a spacer 1 target without any flanking sequences. The results are shown in Figure 4B. Empty pGCC2 exhibited a transformation frequency of 3.9×10^{-6} (Table 5), consistent with previous reports (Rusniok et al., 2009). Plasmids carrying targets for spacers 1, 16, 18 and 23 all exhibited transformation frequencies of $2-4 \times 10^{-6}$ (Table 5), i.e. comparable to that of the empty vector. The cloned protospacers in these plasmids either lack flanking *Neisseria* sequences or have flanking sequences that deviate significantly from the PAM consensus (Figures 3B and 10). In contrast, protospacers 8, 9, 17 and 25, all four of which have flanking sequences that conform to the PAM consensus, consistently failed to yield transformants, indicating that they likely elicited CRISPR interference. Protospacers matching CRISPR spacers 9 and 25 (Figure 10) cloned into a different transformation vector, pYZEJS040 (which confers chloramphenicol resistance), for targeted integration into the distinct capsule locus also yielded no transformants (Figure 11), demonstrating that the observed effect was independent of the vector, the integration locus, and the selectable marker.

[0094] To examine targeting requirements further, we generated a series of mutations in the protospacer or flanking sequences of the pGCC2-derived plasmid targeted by spacer 9 (Figure 4C). Substitutions of two consecutive nucleotides within the PAM or the seed sequence of the protospacer yielded plasmids with transformation frequencies comparable to that of the empty vector (mutants 3 and 5; Figure 4C and Table 5), indicating that interference was abolished. In contrast, two-nucleotide substitutions in a non-PAM flanking region, substitutions in non-seed protospacer positions, small deletions at the PAM-distal end of the spacer, or a wild-type protospacer cloned in the opposite orientation all had no effect on interference (mutants 1, 2, 4, 6, 7 and 8). All of these observations are consistent with previously defined characteristics of functional Type II CRISPR/Cas systems (Deveau et al., 2008; Saprunauskas et al., 2011; Gasunas et al., 2012; Jinek et al., 2012), indicating that CRISPR interference is indeed responsible for the

observed effects on transformation frequencies. The effects of single-nucleotide mutations in this apparent 4-nt PAM remain to be defined.

[0095] Genetic analysis of the *N. meningitidis* CRISPR/*cas* locus. In other Type II CRISPR/Cas systems, Cas9 is the only Cas protein that is necessary for interference specified by existing spacers (Barrangou et al., 2007; Deltcheva et al., 2011; Sapranaukas et al., 2011; Jinek et al., 2012). To investigate if Type II-C CRISPR/Cas systems exhibit the same Cas protein requirements, we introduced transposon insertion mutations in the three *cas* genes – *cas1*, *cas2*, and *cas9* – in *N. meningitidis* 8013 (Figure 5A). We also generated an unmarked, in-frame *cas9* deletion strain to avoid potential polar effects and to generate a guaranteed null allele (Figure 5A). We transformed wild-type and mutant strains with a pYZEJS040 construct carrying protospacer 25 (as in Figure 11B), and compared their transformation frequencies with those of empty pYZEJS040. As expected, the empty vector readily transformed all strains, with transformation frequencies in the range of $\sim 0.5\text{--}7 \times 10^{-5}$ (Figure 5B and Table 5). No transformants were observed when the protospacer 25 construct was used with wild-type cells, indicating effective CRISPR interference. Similarly, transposon insertion mutants in *cas1*, *cas2*, or a control irrelevant gene (*cas1::Tn*, *cas2::Tn* and *1851::Tn*, respectively) exhibited complete interference, consistent with previous results in Type II-A systems (Sapranaukas et al., 2011). In contrast, CRISPR function is abolished in both the transposon-induced (*cas9::Tn*) and deletion ($\Delta cas9$) mutations in *cas9*. The CRISPR interference defect of both alleles could be complemented with wild-type *cas9* under the control of its native promoter (integrated chromosomally via pGCC2), whereas empty pGCC2 had no effect (Figure 5B).

[0096] Previous studies of *cas9* orthologs from *S. thermophilus* and *S. pyogenes* identified two distinct active sites in RuvC-like and HNH nuclease motifs that are important for Cas9 function *in vivo* and *in vitro* (Sapranaukas et al., 2011; Gasunas et al., 2012; Jinek et al., 2012). We engineered alanine mutants in corresponding catalytic residues (D16 in the RuvC domain and H588 in the HNH domain) and tested the abilities of both single mutants to complement the interference defect of the *cas9::Tn* mutant. Both alanine mutants failed to restore interference (Figure 5B). Altogether these analyses demonstrate that the *Neisseria* Type II-C CRISPR/Cas system requires *cas9* but not *cas1*

or *cas2* for interference of natural transformation, and that the presence of intact RuvC-like and HNH motifs are essential for *cas9* function.

[0097] RNase III-catalysed pre-crRNA processing is dispensable for Type II-C CRISPR interference. Two additional loci – *tracrRNA* and *rnc* (the gene encoding RNase III) – have been shown to be essential for crRNA processing and interference in the Type II-A system of *S. pyogenes* SF370 (Deltcheva et al., 2011). The unique *Neisseria* biogenesis pathway described above, in which repeat-driven promoters yield crRNAs with unprocessed 5' ends, raises questions about the roles of *tracrRNA* and RNase III in this Type II-C system. To examine this issue, we generated *N. meningitidis* 8013 derivatives carrying a transposon-induced allele of *rnc*, or a complete deletion of either *rnc* or *tracrRNA* (Figure 5A). These strains were tested in liquid transformation assays as described above, using pYZEJS040 and its protospacer 25-bearing derivative. The results are shown in Figure 5C. The Δ *tracrRNA* strain was completely defective in CRISPR interference, but the defect was restored upon integration of a *tracrRNA* gene with its native promoter in a distinct chromosomal locus. These results are consistent with the strict requirement for *S. pyogenes* *tracrRNA* for pre-crRNA processing and interference *in vivo* (Deltcheva et al., 2011), as well as crRNA-directed, Cas9-catalyzed DNA cleavage *in vitro* (Jinek et al., 2012) and in eukaryotes (Cong et al., 2013).

[0098] Intriguingly, despite the previously demonstrated importance of the *tracrRNA* as a guide for RNase III-mediated processing, we detected no interference defect in either the *rnc::Tn* or Δ *rnc* mutants (Figure 5C). This result is in stark contrast to that observed in the Type II-A system in *S. pyogenes* SF370 (Deltcheva et al., 2011). The lack of an interference defect was observed with a vector that is targeted by an internal spacer (spacer 9; Table 5) as well as a terminal spacer (spacer 25; Figure 5C and Table 5). Northern analyses revealed clear processing defects in both crRNA (Figure 5D) and *tracrRNA* (Figure 5E) for the *N. meningitidis* 8013 *rnc::Tn* mutant, consistent with the results described above with a Δ *rnc* mutant in the WUE2594 strain background (Figure 2D). Mature, 48 nt crRNAs are virtually absent in the *rnc::Tn* mutant, and longer precursors accumulate (Figures 5D and 12). CrRNAs are also strongly depleted in *cas9* mutants, but unlike in the *rnc::Tn* mutant, pre-crRNAs do not accumulate. The latter observation suggests that Cas9 is important for crRNA stability but not processing, or that

Cas9 functions in processing and also stabilizes unprocessed precursors. We conclude that the *N. meningitidis* Type II-C CRISPR/Cas system, unlike all other CRISPR/Cas systems characterized to date, does not require pre-crRNA processing for interference activity. RNase III-catalyzed pre-crRNA processing occurs within the bacterial cell but is dispensable for interference. The tracrRNA requirement likely reflects its involvement in target DNA binding and cleavage by crRNA-programmed Cas9, as observed *in vitro* (Jinek et al., 2012).

[0099] CRISPR interference limits transformation by *Neisseria* genomic DNA. Plasmids are rare in *N. meningitidis* (van Passel et al., 2006), and *Neisseria* genomic DNA (gDNA) is thought to be the most frequent substrate for natural transformation (Hamilton and Dillard, 2006). To test whether our results with *E. coli*-isolated plasmids extend to *Neisseria* gDNA, we generated strains carrying a selectable marker tightly linked to a validated target (protospacer 25). We used the *cas9::Tn* strain to enable transformation and integration of both empty pGCC2 (Figure 4A) and protospacer 25-containing pGCC2 into the meningococcal chromosome (Figure 6A). We then isolated gDNA from these strains and used them as donors in liquid-medium transformation assays with wildtype *N. meningitidis* 8013 cells. Transformants (transformation frequency of 1.6×10^{-5} ; Table 5) were readily obtained using DNA that lacked protospacer 25 adjacent to the *ermC* marker, whereas no transformants were observed when the protospacer was present (Figure 6B, left panel). Similar results were obtained with gDNA strains carrying the *CAT* marker at the capsule locus with and without tightly-linked protospacer 25 (Figures 6 and 11A, and Table 5), indicating that the interference effect was not marker- or locus-specific. We conclude that CRISPR interference is effective against the most common natural substrate for transformation in *N. meningitidis*.

[00100] The potential target spectrum of *Neisseria* CRISPR loci. *In silico* target analysis for *N. meningitidis* 8013 CRISPR spacers is summarized in Figure 3B and in much greater detail as part of Figure 9 and 12. Nine out of the 25 spacers have potential targets that match genomic sequences of *N. meningitidis* and *N. gonorrhoeae* strains. Some of these spacers have potential targets either located within known Nf (Neisserial filamentous) prophages or in genes annotated to encode putative phage-associated proteins. For example, spacer 8 matches to genes (NMB1628 and others) encoding

putative surface antigen tspB proteins that are phage adsorption protein homologues; spacer 9 has twelve protospacer matches that are all in the intergenic regions between genes encoding phage assembly proteins (NGO1138 and others) and the transposases PivNM/irg (NGO1137 and others) in *N. gonorrhoeae* genomes; spacer 17 targets multiple genes (NMB1749 and others) encoding zonula occludens toxin family proteins known to be Nf prophage assembly/structural proteins (Figure 3B and Table 4) (Kawai et al., 2005; Skaar et al., 2005). In addition, spacers 21 and 23 both have a single match to genes encoding phage-associated proteins (Figure 3B and Table 4). Importantly, there are spacers that match to *Neisseria* genomic sequences(s) not related to known phage or mobile elements at all (to the best of our knowledge), such as spacer 14 (two matched loci, one a hypothetical protein and the other an intergenic region), spacer 25 (a predicted DNA binding protein), spacer 16 (a predicted deacetylase), and spacer 18 (hemagglutinin/hemolysin family proteins).

[00101] We also performed *in silico* target analyses in a more prophage-directed way: we examined the available literature for reported *Neisseria* prophage and mobile element sequences (Masignani et al., 2001; Braid et al., 2004; Bille et al., 2005; Kawai et al., 2005; Skaar et al., 2005; Joseph et al., 2011) and then searched them for matches to any of the 325 *Neisseria* protospacers. Overall, among all the 35 unique *Neisseria* spacers with potential targets, 63% (22/35) match a phage-related protospacer by this criterion (Table 4). We noted that apparent prophage targeting by the *N. meningitidis* WUE2594 CRISPR is particularly extensive, accounting for 69% (36/52) of all the phage-related potential matches shown in Table 4. We speculate that the presence of a functional Type II-C CRISPR/Cas system with a dozen prophage-matched spacers has contributed to the lack of Nf prophages in the *N. meningitidis* WUE2594 genome (Joseph et al., 2011 and our observations). The genes most frequently matched (67 out of the 325 protospacers) by *Neisseria* spacers are those encoding homologues of PivNG/PivNM/irg putative transposases and recombinases (Kawai et al., 2005; Skaar et al., 2005). The fact that these genes are usually adjacent to and probably functionally associated with putative Nf prophage and insertion sequences (Skaar et al., 2005; Kawai et al., 2006) suggests that *Neisseria* Type II-C CRISPR/Cas system interferes with the acquisition of Nf prophages by targeting their PivNG/PivNM/irg transposase-encoding loci.

[00102] We also observed that the candidate phage-related CRISPR targets almost exclusively belong to filamentous prophages (Table 4), including the 8kb MDA (Meningococcal Disease Associated) island associated with invasiveness and pathogenicity (Bille et al., 2005). In contrast, several Mu-like prophages (Masignani et al., 2001; Braid et al., 2004; Joseph et al., 2011) from *Neisseria* genomes had no CRISPR spacer matches (Table 4). The reasons for the difference in apparent CRISPR targeting of filamentous and Mu-like prophages are not known.

[00103] Discussion

[00104] CRISPR interference and the third major pillar of horizontal gene transfer.
CRISPR/Cas pathways have been revealed as RNA-directed immune systems that protect bacteria and archaea from phage infection and HGT (Karginov and Hannon, 2010; Marraffini and Sontheimer, 2010; Terns and Terns, 2011; Wiedenheft et al., 2012). Several dozen bacterial species are known to be competent for HGT via natural transformation. Of this subset of bacteria, *Neisseria* spp. are unusual in that their transformation competence is constitutive (Hamilton and Dillard, 2006). Only a few phages are known to infect *N. meningitidis*, and although conjugative plasmids are present in some meningococcal isolates (van Passel et al., 2006), transformation is the major mechanism for mobilization of meningococcal chromosomal loci (Moxon and Jansen, 2005). *Neisseria* genomic sequences are preferred substrates for natural transformation, given that DNA uptake is strongly promoted by a short DNA uptake sequence (DUS) that is highly overrepresented in *Neisseria* spp. chromosomes (Budroni et al., 2011). DNA exchange is so frequent that the population structures of most neisseriae are effectively panmictic, with little propensity for the emergence of clonal subpopulations (Smith et al., 1993). Frequent HGT in *N. meningitidis* is thought to promote capsule switching and other forms of antigenic variation, homology-based DNA repair, and other functions (Hamilton and Dillard, 2006). Native CRISPR/Cas systems have previously been shown to prevent phage infection (and, by inference, phage transduction) and conjugation, which constitute two of the primary routes of HGT. Our results reveal a role for a native CRISPR/Cas system in preventing the third main route of HGT, natural transformation. This is consistent with recent reports that CRISPR/Cas systems can target loci that are already established in bacterial or archaeal chromosomes (Edgar and Qimron, 2010;

Gudbergsdottir et al., 2011; Jiang et al., 2013), indicating that interference does not depend on the invasive DNA's route of entry. Similarly, an engineered, heterologous CRISPR/Cas system introduced into *Streptococcus pneumoniae* can block natural transformation during active infection in mice (Bikard et al., 2012). We find that a native CRISPR/Cas system in *N. meningitidis* can block the transformation events that can be so important for immune evasion and other critical aspects of invasiveness and pathogenicity. Intriguingly, the ability of native CRISPR systems to block natural transformation would be expected to enable the selection of spacers that discriminate against specific chromosomal loci that negatively affect the fitness of certain strains or under certain conditions.

[00105] Although relatively few phages are known to infect *N. meningitidis*, they are not unknown (Kawai et al., 2005). Several genomic islands have been identified that resemble phages and could therefore represent prophage sequences (Bille et al., 2005, 2008; Joseph et al., 2011). One such sequence, the MDA island, correlates with invasiveness and pathogenicity in young adults (Bille et al., 2005, 2008). The existence of numerous CRISPR spacers with the potential to target these sequences suggests that CRISPR interference plays a role in shaping prophage content and serves phage defense functions in *N. meningitidis*, as elsewhere. CRISPR interference could limit the spread of prophages via either transformation or infection. Accordingly, CRISPR interference could negatively correlate with meningococcal pathogenicity, as suggested in enterococci (Palmer and Gilmore, 2010) and streptococci (Bikard et al., 2012). Alternatively, meningococcal Cas9 could participate in other regulatory events that contribute to pathogenicity, as suggested very recently (Sampson et al., 2013).

[00106] It is noteworthy that many *N. meningitidis* and *N. lactamica* strains encode CRISPR systems, while strains of the closely related *N. gonorrhoeae* with clearly functional CRISPR systems have not been identified. It is believed that these organisms split in relatively recent times (<100,000 years ago), evolutionarily speaking, but exact estimates have been stymied by the large recombination frequencies in these species (Bennett et al., 2010). It is equally possible that the nasopharyngeal-localized species gained the system after the split, or that *N. gonorrhoeae* lost the CRISPR system after the split. Both pathogens have been suggested not to establish long-lasting clones and tend

towards linkage equilibrium (Buckee et al., 2008). It may not be coincidental that *N. meningitidis* carries a CRISPR system and can develop semi-clonal lineages (Bart et al., 2001), given that the CRISPR system could provide a short-term barrier to HGT. It is also possible that the co-existence of commensal *Neisseria* species such as *N. lactamica* and *N. meningitidis* in the nasal pharynx (Feil and Spratt, 2001) selects for a system that limits genetic exchange, and intriguingly, some *N. lactamica* and *N. meningitidis* isolates show large amounts of exchange while others show lower signatures of exchange (Hanage et al., 2005; Corander et al., 2012). In contrast, *N. gonorrhoeae* inhabits a niche that is devoid of other closely related bacteria that encode the DUS to allow efficient transformation of their sequences (Vazques et al., 1993; Cehovin et al., 2013).

[00107] Towards a minimal CRISPR/Cas system. In CRISPR/Cas systems investigated to date, crRNAs are transcribed from an external promoter, generating a multimeric precursor. The pre-crRNA is processed by endonucleolytic cleavage to generate mature crRNAs (Carte et al., 2008; Haurwitz et al., 2010; Gesner et al., 2011), and processing is essential for interference *in vivo* (Brouns et al., 2008; Delicheva et al., 2011; Hale et al., 2012; Westra et al., 2012). The potential presence of minimal and apparently fortuitous promoter elements has been noted within certain CRISPRs of *Sulfolobus solfataricus* P2, though they are not thought to account for the functional expression of crRNA and in fact appear to be suppressed by the repeat-binding protein Cbp1 (Deng et al., 2012). The results presented here reveal that streamlined CRISPR/Cas systems exist in which pre-crRNA processing is not essential (Figure 7). In CRISPR-containing strains of *N. meningitidis* and *N. lactamica*, as well as other species such as *C. jejuni*, the CRISPR repeats each contain an extended -10 box that drives transcription initiation within the downstream spacer. Thus, many crRNAs contain 5'-triphosphate ends that are not subject to further 5'-processing.

[00108] Like other Type II CRISPR/Cas systems, neisseriae produce a tracrRNA that apparently anneals to pre-crRNA and enables binding and cleavage by a RNase III. This reaction generates crRNA 3' ends, and *rnc* mutants accumulate multimeric crRNA precursors. However, these *rnc* mutants exhibit no interference defect, indicating that processing is not essential. In addition, while the tracrRNA is essential for interference, its role in directing processing is not, since processing is itself dispensable. This provides the

first clear indication that the tracrRNA is required for post-processing events such as target DNA binding and cleavage in bacterial cells, as it is *in vitro* (Jinek et al., 2012).

[00109] Among the three main types of CRISPR/Cas pathways, the Type II systems are the simplest ones characterized thus far, as judged by the number of components and essential steps. Both Type II-A and Type II-B systems include the CRISPR array itself, a tracrRNA, four protein-coding genes encoded within the *cas* locus, and the host factor RNase III (Deltcheva et al., 2011; Makarova et al., 2011b; Magadán et al., 2012; Chylinski et al., 2013). The *Neisseria* systems that we have characterized are even more streamlined: they do not require a separate leader sequence to drive crRNA transcription, they lack one of the four *cas/csn* genes present in Type II-A or II-B systems, and they do not require RNase III or crRNA processing. The *Neisseria* systems are among the founding members of a new CRISPR/Cas subtype (Type II-C) that is characterized by a smaller number of *cas/csn* proteins (Koonin and Makarova, 2013; Chylinski et al., 2013), and in at least some cases by repeat-embedded promoters and processing independence.

[00110] Importantly, recent reports have shown that Type II CRISPR/Cas systems can be ported into eukaryotic cells and employed for RNA-directed genome editing and genome binding, including multiplexed applications specified by multiple spacers (Jinek et al., 2012, 2013; Cho et al., 2013; Cong et al., 2013; DiCarlo et al., 2013; Hwang et al., 2013; Mali et al., 2013; Qi et al., 2013). The Cas9 effector proteins from *neisseriae* share the conserved features observed in the *S. pyogenes* and *S. thermophilus* Cas9 enzymes used in these studies (Chylinski et al., 2013). The fewer the functional requirements for the operation of such systems, the greater their versatility and applicability will be. Separately encoded crRNAs and tracrRNAs are more efficient *in vivo* than single-guide RNAs that combine essential crRNA and tracrRNA domains in the same transcript (Cong et al., 2013). Although endogenous eukaryotic activities can substitute for bacterial RNase III to process tracrRNA/pre-crRNA hybrids in human and mouse cells (Cong et al., 2013), it is not known whether RNase III will be dispensable in other eukaryotic contexts, or indeed in all mammalian cell types. Accordingly, the identification of processing-independent CRISPR/Cas systems could increase the range of applications in eukaryotic genome editing, especially in light of the potential toxicity of bacterial RNase III expression (Pines et al., 1988). Such applications will benefit from further analysis of

meningococcal Cas9 activity, including the definition of the presumptive cleavage sites relative to the PAM.

[00111] Experimental Procedures

[00112] Bacterial Strains, Plasmids, and Oligonucleotides. *N. lactamica* 020-06, *N. meningitidis* WUE2594 and 8013, and mutant derivatives thereof that were used in this study are listed in Supplemental Experimental Procedures, as are complete lists of all plasmids and DNA oligonucleotides.

[00113] Mutant Strain Construction. All mutants were confirmed by PCR and DNA sequencing. Most mutant strains were generated by transformation with appropriately constructed plasmids. For generation of the *cas9*, *rnc*, and control NMV_1851 transposon-induced alleles in the 8013 strain background, we used gDNA from the corresponding mutant in the NeMeSys collection (Rusniok et al., 2009) to transform 8013. For generation of the Δrnc derivative of 8013, we used gDNA from the WUE2594 Δrnc derivative that was initially made by a plasmid-based approach. For complementation of *cas9::Tn*, $\Delta cas9$, and $\Delta tracrRNA$ mutants, we cloned wildtype copies of the relevant gene into plasmid pGCC2 and transformed the resulting plasmids into the parental mutant strain.

[00114] RNA Extraction and Depletion of Processed RNAs. For 020-06, WUE2594 and its mutant derivatives, total RNA was extracted from frozen cell pellet lysates using the hot-phenol method described previously (Blomberg *et al.*, 1990). For depletion of processed transcripts, equal amounts of total RNA were incubated with TerminatorTM exonuclease (TEX) (Epicentre) or in buffer alone as described (Sharma et al., 2010). For 8013 and its mutant derivatives, total RNAs were extracted from frozen cell pellets using miRNeasy Mini Kit (Qiagen) with two additional steps: a ten minute initial cell lysis with lysozyme and Proteinase K, and a later on-column DNase digestion step (the RNase-Free DNase Set, Qiagen).

[00115] dRNA-seq. Libraries for Solexa sequencing (HiSeq) of cDNA were constructed by vertis Biotechnology AG, Germany (<http://www.vertis-biotech.com/>), as described previously for eukaryotic microRNA (Berezikov *et al.*, 2006) but omitting the

RNA size-fractionation step prior to cDNA synthesis. cDNA libraries were sequenced using a HiSeq 2000 machine (Illumina) in single read mode at the Max Planck Genome Centre Cologne (Cologne, Germany). Data analysis was done as described (Chao et al., 2012), with the only exception being that the minimal read length after trimming and clipping was 12 nt instead of 20 nt.

[00116] Transcriptional *gfp* Fusions. The inserts used for the construction of the transcriptional *gfp* fusion plasmids were generated with synthetic DNA oligonucleotides. *E. coli* cells were transformed with these plasmids and grown on agar plates for fluorescence imaging. To measure GFP fluorescence, single colonies were grown in LB broth, fixed, and analyzed by flow cytometry.

[00117] In vitro Transcription. Templates for *in vitro* transcription assays were PCR-generated, gel-purified 210 bp DNA fragments amplified from pNH13, pNH14, or pNH15. Transcription reactions with sigma-saturated *E. coli* RNA Polymerase holoenzyme (Epicentre) included α -[^{32} P]-ATP.

[00118] Natural Transformation. Natural transformation assays were performed as described (Duffin and Seifert, 2012). Transformation frequencies were reported as antibiotic-resistant cfu/ml divided by total cfu/ml from at least three independent experiments (mean $\pm$ s.e.m.).

[00119] Accession Numbers. The Gene Expression Omnibus (GEO) accession number for the dRNA-Seq data reported in this paper is GSE44582.

[00120] Bacterial Strains and Growth Conditions. *N. lactamica* 020-06, *N. meningitidis* WUE2594 and 8013, and mutant derivatives thereof that were used in this study are listed below.

Strain names	Relevant genotypes	Source
<i>N. lactamica</i> 020-06	Wild type	Dr. Julia Bennett
<i>N. meningitidis</i> WUE2594	Wild type	Dr. Christoph Schoen lab collection
	Δmc	This study
<i>N. meningitidis</i> 8013	Wild type	Dr. Hank Seifert lab collection
	1851::Tn	Dr. Vladimir Pelicic, genomic DNA
	cas9::Tn	Dr. Vladimir Pelicic, genomic DNA
	mc::Tn	Dr. Vladimir Pelicic, genomic DNA
	Δmc	This study

	<i>cas1::Tn</i>	This study
	<i>cas2::Tn</i>	This study
	<i>cas9::Tn/pGCC2</i>	This study
	<i>cas9::Tn /pGCC2-cas9 wt</i>	This study
	<i>cas9::Tn/pGCC2-cas9 D16A</i>	This study
	<i>cas9::Tn/pGCC2-cas9 H588A</i>	This study
	$\Delta cas9$	This study
	$\Delta cas9/pGCC2$	This study
	$\Delta cas9/pGCC2-cas9 wt$	This study
	$\Delta tracr$	This study
	$\Delta tracr/pGCC2$	This study
	$\Delta tracr/pGCC2-tracr$	This study
	<i>cas9::Tn/pGCC2-protospacer25</i>	This study
	<i>cas9::Tn/pYZEJS040</i>	This study
	<i>cas9::Tn/pYZEJS040-protospacer25</i>	This study

[00121] Strain 8013 and its derivatives were grown on GC Medium Base (GCB) (Difco) plates with appropriate antibiotics and Kellogg's supplements I and II (22.2 mM glucose, 0.68 mM glutamine, 0.45 mM co-carboxylase, 1.23 mM $\text{Fe}(\text{NO}_3)_3$; all from Sigma). Antibiotic concentrations used for 8013 were 2.5 $\mu\text{g}/\text{ml}$ for erythromycin; 50 $\mu\text{g}/\text{ml}$ for kanamycin; 50 $\mu\text{g}/\text{ml}$ for streptomycin; and 2.5 $\mu\text{g}/\text{ml}$ for chloramphenicol. 020-06, WUE2594, and derivatives thereof were grown on GC agar (Difco) with PolyViteX (bioMerieux), and with 7 $\mu\text{g}/\text{ml}$ chloramphenicol when appropriate. All solid cultures were incubated at 37°C in a 5% CO_2 humidified atmosphere.

[00122] Liquid cultures of 020-06, WUE2594 and its derivatives were grown in a 37°C shaker-incubator at 220 rpm without added CO_2 . Bacteria grown on Columbia agar plates with 5% sheep blood (bioMerieux) were harvested and a starter culture was inoculated to a final OD_{600} of 0.4 in a flask containing 10 ml of Proteose Peptone Media (PPM^+) medium supplemented with PolyViteX (bioMerieux). After one hour the starter culture was used to inoculate a flask containing 25 ml PPM^+ /PolyViteX to a final OD_{600} of 0.05. When the cultures reached mid-log (OD_{600} 0.5) or early stationary (OD_{600} 1.0) phase, 10 ml of culture were harvested. The cell samples were immediately centrifuged for 10 min at 4,000 rpm. The cell pellet was frozen in liquid N_2 and stored at -80°C until RNA extraction.

[00123] Mutant Strain Construction. All mutants were confirmed by PCR and DNA sequencing. PCRs for verifying strains or transformants were performed with Taq or OneTaq DNA Polymerases (NEB) using either 10 ng of genomic DNA (25 cycles) or 0.5

μl CLS extracts (35 cycles) as templates. Chromosomal DNAs were isolated using QIAamp DNA Mini Kit (Qiagen).

[00124] The *cas1* and *cas2* transposon-induced alleles were made by transforming 8013 with the plasmids pCR2.1/*cas1*-Kan and pCR2.1/*cas2*-Kan, respectively, followed by Kan^R selection. For generation of the *cas9*, *rnc*, and control NMV_1851 transposon-induced alleles in the 8013 strain background, we used chromosomal DNA from the corresponding mutant in the NeMeSys collection (Rusniok et al., 2009) to transform 8013, and then selected Kan^R transformants. The *cas9::Tn* mutant strain with a transposon insertion after the 604th nucleotide of the ORF was constructed with NeMeSys mutant 23/6. The *rnc::Tn* mutant strain with a transposon insertion after 574th nucleotide of the ORF was constructed with NeMeSys mutant 6/47. A control strain with a transposon insertion after the 22nd ORF nucleotide of gene NMV_1851 (which encodes a hypothetical protein) was constructed using NeMeSys mutant 73/5. The *kan*-marked Δ *tracrRNA* strain was made by transforming 8013 with plasmid pSMARTHCamp/ Δ *tracr*+Kan, followed by Kan^R selection.

[00125] The WUE2594 Δ *rnc* derivative was constructed by replacing the *rnc* gene with a kanamycin resistance cassette. WUE2594 was then transformed with the plasmid pBJ1 and Kan^R colonies were selected. The Δ *rnc* derivative of 8013 was made by transforming 8013 with genomic DNA from the WUE2594 Δ *rnc* derivative, followed by Kan^R selection.

[00126] To create the unmarked, in-frame Δ *cas9* allele (removing all ORF nts except for the five N-terminal and five C-terminal codons), we first selected a spontaneous streptomycin-resistant (Sm^R) mutant of 8013 by plating 3×10^9 wt cells on GCB plates with 50 μg/ml streptomycin, and selecting an Sm^R colony. We confirmed that it carried an A128G substitution in *rpsL*, resulting in a K34R missense mutation. We then transformed this Sm^R derivative with plasmid pSTblue-1/ Δ *cas9*/*CAT-rpsL*, in which a dual-marker cassette [*CAT* (chloramphenicol acetyltransferase) and wild-type *rpsL*] replaced most of the *cas9* ORF. The resulting Cm^R transformants are also streptomycin-sensitive (Sm^S), since the Sm^S phenotype conferred by the wildtype *rpsL* is dominant over the Sm^R phenotype conferred by the *rpsL*^{A128G} allele at the native locus. Sm^S Cm^R transformants

were then transformed with plasmid pSTblue-1/ $\Delta cas9$. Sm^R Cm^S colonies from this transformation were screened by PCR to confirm replacement of the dual marker cassette with the unmarked *cas9* deletion.

[00127] For complementation of *cas9::Tn*, $\Delta cas9$, and $\Delta tracrRNA$ mutants, we cloned wildtype copies of the relevant gene into plasmid pGCC2, transformed the resulting plasmids into the parental mutant strain, and selected erythromycin-resistant (Erm^R) transformants.

[00128] To generate strains carrying a selectable marker tightly linked to a target protospacer (as a source of chromosomal DNA for genomic transformation experiments), plasmids pGCC2 or pGCC2-MC8013spacer25 were transformed into the *cas9::Tn* strain, and Erm^R transformants were selected. Similarly, pYZEJS040 or pYZEJS040-MC8013spacer25 were transformed into the *cas9::Tn* strain, and Cm^R transformants were selected.

[00129] Plasmids. A complete list of all plasmids, as well as information on their construction, is provided at the end of this section. *E. coli* Top10 cells (Invitrogen) were used for all cloning procedures. All plasmid constructions were sequence-verified. PCR reactions for cloning were performed with Platinum Pfx DNA Polymerase (Invitrogen).

[00130] The inserts used for the construction of transcriptional *gfp* fusion plasmids pNH13, pNH14, pNH15 and pNH18 were generated by duplex formation of complementary DNA oligonucleotides. Oligonucleotide pairs were JVO9535/JVO9536 and JVO9537/JVO9538 for *Neisseria* spp. wildtype (pNH13) and mutant (pNH14) CRISPR repeat constructs, respectively; JVO9599/JVO9601 for the wildtype CRISPR repeat from *Campylobacter jejuni* NCTC11168 (pNH18); and JVO9539/JVO9540 for the -10/-35 positive control promoter from T7A1 phage (pNH15). For each DNA duplex insert, 100 nM sense oligonucleotides were annealed with equimolar amounts of antisense oligonucleotides at 95°C for 3 min, followed by slow cooling to room temperature. DNA duplexes were digested with *AatII/NheI* and cloned into *AatII/NheI*-digested pAS093. For construction of 3x CRISPR-repeat-spacer unit-*gfp* transcriptional fusion plasmid pNH17, the plasmid pAS093 was digested with *AatII/NheI* and ligated to *AatII/NheI*-digested PCR

products amplified from *N. lactamica* 020-06 chromosomal DNA with primer pairs JVO9585/JVO9548.

[00131] To generate the pBJ1 plasmid used for creating the Δrnc mutation in WUE2594, ~600 bp upstream and downstream of the *rnc* gene were amplified with the primer pairs *rnc1/rnc2* and *rnc3/rnc4*, respectively, using WUE2594 genomic DNA as template. The oligonucleotides were modified so as to introduce *Bam*HI/*Eco*RI site at the 5' and 3' ends (respectively) of the upstream fragment, and *Eco*RI/*Hind*III sites at the 5' and 3' ends (respectively) of the downstream fragment. These fragments were cloned into the pBluescript II SK(+) vector (Invitrogen) along with an *Eco*RI-digested fragment of pUC4K (GE Healthcare) containing the kanamycin cassette, yielding the knock-out plasmid (pBJ1) that contains the kanamycin cassette flanked on either side by the upstream and downstream regions of *rnc*.

[00132] Short putative targets for strain 8013 CRISPR spacers 1 (30 nts), 16 (50 nts), 23 (50 nts), and 25 (50 nts) were created by annealing synthetic oligonucleotide pairs OYZ001/OYZ002, OYZ007/OYZ008, OYZ011/OYZ012, and OYZ015/OYZ016, respectively. Longer (208, 350, 305, and 203 nt) putative targets for spacers 8, 9, 17, 18 of 8013 were PCR-amplified from the chromosomal DNAs of *N. meningitidis* strain MC58, *N. gonorrhoea* strain FA1090, and *N. meningitidis* strains MC58 and Z2491 respectively, and digested with *Aat*II and *Pac*I. Primer pairs for these PCRs were OYZ003/OYZ004, OYZ005/OYZ006, OYZ009/OYZ010, and OYZ013/OYZ014, respectively. All eight of these putative targets were ligated into pGCC2 via *Aat*II and *Pac*I sites, to create pGCC2 derivatives for interference tests.

[00133] pYZEJS040 (pSTblue-1/*siaA*+*CAT*+*ctrA*) was constructed by PCR-amplifying three individual fragments: a 562nt *siaA* fragment from 8013 chromosomal DNA using primers OYZ036/OYZ037; a 561nt *ctrA* fragment from 8013 chromosomal DNA using primers OYZ040/OYZ041; and a 1239nt *CAT* cassette from the pGCC5 vector using primers OYZ038/OYZ039. 100ng of each of the three fragments were added to a 50 μ l PCR reaction without any primers. After 15 cycles of PCR, outside primers OYZ036/OYZ041 were added and 20 more cycles were performed. The ends of the 2.3 kb fusion product *siaA-CAT-ctrA* were blunted, and the fragment was ligated into the *Eco*RV

site of pSTblue-1 to yield pYZEJS040. The pYZEJS040 derivatives used in interference tests were generated by ligating potential targets for 8013 CRISPR spacers 9 and 25 into pYZEJS040 via the *Aat*II and *Pac*I sites.

[00134] To construct plasmid pCR2.1/*cas1*-Kan, a 2.4 kb insert was PCR-amplified from the chromosomal DNA of NeMeSys strain 10/4 (Rusniok et al., 2009) using primers OYZ060/OYZ061. This insert, which contains a 1.6 kb Kan^R transposon inserted into the *cas1* gene, was cloned using Original TA Cloning Kit pCR2.1 (Invitrogen) according to the manufacturer's instructions. Similarly, plasmid pCR2.1/*cas2*-Kan was created by amplifying a 2.45 kb insert from the chromosomal DNA of NeMeSys strain 71/27 using primers OYZ052/OYZ055, and cloning that fragment into pCR2.1.

[00135] To create plasmids to be used in generating the unmarked $\Delta cas9$ mutant, genomic sequences upstream and downstream of *cas9* gene were PCR amplified, fused together via overlapping PCR and cloned into pSTblue-1. A 662 nt region containing the first 15 nt of the *cas9* ORF and 632 nt upstream of *cas9* was PCR-amplified from 8013 genomic DNA using primers OYZ066/OYZ068. Similarly, a 517nt region containing the last 15 nt of the *cas9* ORF and 487nt downstream of *cas9* was amplified using primers OYZ069/OYZ071. 100 ng of both PCR fragments were added to a 50 μ l PCR reaction without any primers. After 15 cycles of PCR, outside primers OYZ066/OYZ071 were added and 20 more PCR cycles were performed. The resulting 1.2 kb fusion product included internal *Sa*I and *Spe*I sites (originally incorporated in the primers). The ends of the fragment were blunted, and the product was ligated into the *Eco*RV site of pSTblue-1 to create plasmid pSTblue-1/ $\Delta cas9$ +*Sa*I-*Spe*I. The *Sa*I and *Spe*I sites of this plasmid were used to introduce a 1.6kb *CAT-rpsL* dual marker cassette, and resulted in pSTblue-1/ $\Delta cas9$ /*CAT-rpsL*. The plasmid pSTblue-1/ $\Delta cas9$ was generated similarly: 647 nt and 502 nt genomic fragments upstream and downstream of the *cas9* gene, including the 15 nts at each terminus of the ORF, were amplified using primers OYZ066/OYZ067 and OYZ070/OYZ071, respectively, and then fused together by overlapping PCR. The 1.2 kb fusion product was blunted and ligated into the *Eco*RV site of pSTblue-1.

[00136] To create pSMARTHCAmp/ $\Delta tracr$ +*Pme*I, genomic sequences upstream and downstream of the *tracr*RNA region were PCR amplified and fused together via

overlapping PCR. The 638nt upstream region and the 598nt downstream region were amplified from 8013 chromosomal DNA using primer pairs OYZ081/OYZ082 and OYZ083/OYZ084, respectively. 100 ng of both fragments were added to a 50 µl PCR reaction without any primers, and after 15 cycles, outside primers OYZ081/OYZ084 were added and 20 more cycles were performed. The 1.2 kb fusion product included an internal *PmeI* site (designed in the primers). The fragment was blunted and ligated into vector pSMARTHCAMP according to the instructions for the CloneSmart Cloning Kit (Lucigen). The *PmeI* site was used to insert a 1.2 kb Kan^R cassette that had been amplified from NeMeSys mutant 23/6 chromosomal DNA using primer pair OYZ085/OYZ086. This yielded plasmid pSMARTHCamp/ $\Delta tracr$ +Kan.

[00137] Complementation plasmid pGCC2/promoter+*cas9*wt was created by amplifying the *cas9* ORF and its native promoter from 8013 genomic DNA using primer pair OYZ072/OYZ073, digesting the PCR product with *AatII* and *PacI*, and then ligating it into pGCC2 via the *AatII/PacI* sites. pGCC2/promoter+tracr was created by amplifying the tracrRNA locus with its native promoter from 8013 genomic DNA using primer pair OYZ091/OYZ092, digesting the PCR product with *AatII* and *PacI*, then ligating it into pGCC2 via the *AatII/PacI* sites.

Plasmids	Relevant characteristics	Source
Plasmids for <i>in vivo</i> interference assays		
pYZEJS001	pGCC2 empty vector	Dr. Hank Seifert lab collection
pYZEJS010	pGCC2-MC8013 spacer1 target	This study
pYZEJS011	pGCC2-MC8013 spacer8 target	This study
pYZEJS012	pGCC2-MC8013 spacer9 target	This study
pYZEJS014	pGCC2-MC8013 spacer16 target	This study
pYZEJS015	pGCC2-MC8013 spacer17 target	This study
pYZEJS016	pGCC2-MC8013 spacer18 target	This study
pYZEJS017	pGCC2-MC8013 spacer23 target	This study
pYZEJS018	pGCC2-MC8013 spacer25 target	This study
pYZEJS019	pGCC2-MC8013 spacer9 target mut1	This study
pYZEJS020	pGCC2-MC8013 spacer9 target mut2	This study
pYZEJS021	pGCC2-MC8013 spacer9 target mut3	This study
pYZEJS022	pGCC2-MC8013 spacer9 target mut4	This study
pYZEJS023	pGCC2-MC8013 spacer9 target mut5	This study
pYZEJS024	pGCC2-MC8013 spacer9 target mut6	This study
pYZEJS025	pGCC2-MC8013 spacer9 target mut7	This study
pYZEJS026	pGCC2-MC8013 spacer9 target mut8	This study
pYZEJS028	pGCC5 empty vector	Dr. Hank Seifert lab collection
pYZEJS032	pSTblue-1 empty vector	Novagen
pYZEJS040	pSTblue-1/ <i>staA</i> + <i>CAT</i> + <i>ctrA</i>	This study
pYZEJS042	pYZEJS040-protospacer9	This study
pYZEJS043	pYZEJS040protospacer25	This study

Plasmids for creating the $\Delta cas9$ strain		
pYZEJS033	pSTblue-1/ $\Delta cas9$ /Sall+SpeI	This study
pYZEJS034	pSTblue-1/ $\Delta cas9$	This study
pYZEJS035	pSTblue-1/ $\Delta cas9$ /CAT+rpsL	This study
Plasmids for creating $cas1::Tn$ and $cas2::Tn$ strains		
pYZEJS037	pCR2.1/ $cas1$ -Kan	This study
pYZEJS038	pCR2.1/ $cas2$ -Kan	This study
Plasmid for creating the $cas9$ complementation strain		
pYZEJS044	pGCC2-promoter+ $cas9$ wt	This study
Plasmids for creating $\Delta tracr$ strain		
pYZEJS061	pSMARTHCAmp/ $\Delta tracr$ +PmeI	This study
pYZEJS062	pSMARTHCAmp/ $\Delta tracr$ +Kan	This study
Plasmid for creating $tracr$ complementation strain		
pYZEJS064	pGCC2-promoter+ $tracr$	This study
Plasmids for creating the Δrnc strain in <i>N. meningitidis</i> WUE2594		
pBJ1	Δrnc	This study
Plasmids for promoter cloning		
pAS093 (P_{invR} - gfp)	$invR$ transcriptional gfp fusion plasmid, expressing constitutive gfp	(Sittka et al., 2008)
pAS0046 (gfp (P_{mut}))	background control plasmid for transcriptional gfp fusion plasmid	(Sittka et al., 2007)
pNH13 (P_{wt})	<i>Neisseria</i> : CRISPR-repeat transcriptional gfp fusion plasmid	This study
pNH14 (P_{mut})	<i>Neisseria</i> : mutant CRISPR-repeat transcriptional gfp fusion plasmid	This study
pNH15 (P_{ctrl})	T7 phage promoter A1 transcriptional gfp fusion plasmid	This study
pNH17 (P_{3x})	<i>Neisseria</i> : 3x CRISPR-repeat-spacer unit transcriptional gfp fusion plasmid	This study
pNH18 ($P_{C. jejuni}$)	<i>Campylobacter jejuni</i> : CRISPR-repeat transcriptional gfp fusion plasmid	This study

[00138] RNA Extraction, Depletion of Processed RNAs, and Northern Blots. For 020-06, WUE2594 and its mutant derivatives, frozen cell pellets from liquid cultures were resuspended in lysis solution containing 800 μ l of 0.5 mg/ml lysozyme in TE buffer (pH 8.0) and 80 μ l 10% SDS. Bacterial cells were lysed by placing the samples for 1–2 minutes at 65°C in a water bath. Afterwards, total RNA was extracted from the lysates using the hot-phenol method described previously (Blomberg *et al.*, 1990). For depletion of processed transcripts, total RNA was freed of residual genomic DNA by DNase I treatment, and equal amounts of *Neisseria* RNA were incubated with Terminator 5'-phosphate-dependent exonuclease (TEX) (Epicentre) or in buffer alone as previously described (Sharma *et al.*, 2010). For northern blot analysis, 5 μ g total RNA freed of residual genomic DNA or 3 μ g of TEX treated RNA was loaded per sample. After separation by electrophoresis in 8% polyacrylamide/8.3 M urea/1x TBE gels, RNA was

transferred onto Hybond-XL membranes, and membranes were hybridized with γ - ^{32} P-ATP end-labeled oligodeoxyribonucleotide probes.

[00139] For 8013 and its mutant derivatives, cells grown overnight on GCB plates were collected, immediately treated with RNAlprotect Bacteria Reagent (Qiagen), and frozen at -80°C for storage. Total RNAs were extracted using miRNeasy Mini Kit (Qiagen) with two additional steps: a 10 min initial cell lysis in 30 mM Tris-HCl (pH 8.0)/1 mM EDTA containing 1.5 mg/ml lysozyme (Invitrogen) and 2 mg/ml Proteinase K (Fermentas), and a later on-column DNase digestion step (The RNase-Free DNase Set, Qiagen). For northern analysis, 8-10 μg of total RNA for each sample was separated by electrophoresis in a 10% polyacrylamide/8 M urea/1x TBE gel. RNAs were electroblotted overnight at 14V to a Genescreen Plus membrane (PerkinElmer) in 1x TBE, cross-linked to the membrane by UV irradiation and then soaking in 0.16M N-(3-dimethylaminopropyl)-N'-ethylcarbodiimide hydrochloride/0.13 M 1-methylimidazole (both from Sigma) (pH 8.0) at 60°C for 2h. Blots were prehybridized with 8 ml ULTRAhyb buffer (Ambion) at 60°C for 30 min, then probed at 37°C overnight with 5×10^6 cpm/ml of the appropriate DNA oligonucleotide probe. Northern probes were 5' end-labeled with γ - ^{32} P-ATP (PerkinElmer) and T4 polynucleotide kinase (NEB). The membranes were washed at room temperature twice with 2x SSC/0.1% SDS for 5 min and then twice with 1x SSC/0.1% SDS for 15 min. RNAs were then visualized by PhosphorImager detection. Similarly end-labeled *MspI*-digested pBR322 DNAs (NEB) were used as size markers.

[00140] Construction of cDNA Libraries for dRNA-seq. Total RNA was freed of residual genomic DNA by DNase I treatment. For depletion of processed transcripts, equal amounts of *Neisseria* RNA were incubated with TerminatorTM 5'-phosphate dependent exonuclease (TEX) (Epicentre) as previously described (Sharma et al., 2010). Libraries for Solexa sequencing (HiSeq) of cDNA were constructed by vertis Biotechnology AG, Germany (<http://www.vertis-biotech.com/>), as described previously for eukaryotic microRNA (Berezikov et al., 2006) but omitting the RNA size-fractionation step prior to cDNA synthesis. In brief, equal amounts of RNA samples were poly(A)-tailed using poly(A) polymerase. Then, the 5'-triphosphate structures were removed using tobacco acid pyrophosphatase (TAP). Afterwards, an RNA adapter was ligated to the 5'-phosphate of

the RNA. First-strand cDNA was synthesized by an oligo(dT)-adapter primer and MMLV reverse transcriptase. A PCR-based amplification step with a high-fidelity DNA polymerase was then used to increase the cDNA concentration to 20-30 ng/ μ l. A library-specific barcode for sequence multiplexing was included in the 3'-sequencing adapter. cDNA libraries were sequenced using a HiSeq 2000 machine (Illumina) in single read mode at the Max Planck Genome Centre Cologne (Cologne, Germany).

[00141] Read Mapping and Coverage Plot Construction. Sample preparation, sequencing (Illumina GAIIx) and data analysis was done as described (Chao et al., 2012), with the only exception being that the minimal read length after trimming and clipping was 12 nt instead of 20 nt.

[00142] Transcriptional *gfp* Fusions. *E. coli* cells were transformed with transcriptional *gfp* fusion plasmids and grown on agar plates for fluorescence imaging. To measure GFP fluorescence, single colonies were inoculated in LB broth and grown for 12 h. Cells were then fixed in 4% paraformaldehyde/1x PBS and analysed by flow cytometry.

[00143] *In vitro* Transcription. Templates for *in vitro* transcription assays were PCR-generated, gel-purified 210 bp DNA fragments amplified from pNH13, pNH14, or pNH15. Primer pairs were the forward primers used for construction of each DNA duplex insert (see above), together with reverse primer JVO155. Templates (100 ng) were incubated at 37°C in transcription buffer (40 mM Tris-HCl (pH 7.5)/100 mM KCl/10 mM MgCl₂/0.01% Triton/1 mM DTT) together with 1.5 Units sigma-saturated *E. coli* RNA Polymerase Holoenzyme (Epicentre), α -[³²P]-ATP (30 μ Ci; Hartmann-Analytic Braunschweig), and NTP mix (10 μ M ATP and 200 μ M each CTP, GTP, UTP). A negative control reaction used water in place of DNA template. 25 μ l reactions were incubated for 1, 5, 10 and 30 min. Aliquots were phenol-extracted, precipitated, denatured by heating in formamide loading dye, separated by electrophoresis in 12% sequencing gels, and analyzed with a PhosphorImager.

[00144] CRISPR prediction and *in silico* Analysis of Natural Targets. CRISPRs in sequenced *Neisseria* genomes were predicted using CRISPRfinder (<http://crispr.u-psud.fr/Server/>) (Grissa et al., 2007) and CRISPRi (<http://crispi.genouest.org/>) (Rousseau et al., 2009). Our initial predictions of *Neisseria* CRISPRs were consistent with those of

CRISPRdb (<http://crispr.u-psud.fr/crispr/>). Spacers were subjected to blastn (Basic Local Alignment Search Tool) search against the nr/nt database (<http://www.ncbi.nlm.nih.gov/>). Multiple Sequence Alignments were performed using WebLogo (<http://weblogo.berkeley.edu/logo.cgi>).

[00145] Natural Transformation. Natural transformation assays were performed in *N. meningitidis* 8013 and its mutant derivatives as described for *N. gonorrhoeae* (Duffin and Seifert, 2012). 150 ng plasmids or 100 ng chromosomal DNA was used per transformation reaction. 10 µl of serial 10-fold dilutions were spotted on GCB plate in triplicates in the presence and absence of appropriate antibiotics. 200 µl from the undiluted final transformation mixture were also plated on GCB plates with appropriate antibiotics to enhance detection. Eight representative transformants per reaction were verified by re-streaking on selective plates twice and then by PCR from CLS extract (i.e., from cells lysed in 1% Triton/20 mM Tris-HCl (pH 8.3)/2mM EDTA at 94°C for 15 min and then 20°C for 5 min). Transformation frequencies were reported as antibiotic-resistant cfu/ml divided by total cfu/ml from at least three independent experiments (mean ± s.e.m.).

[00146] References for Example 1

[00147] Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., and Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 315, 1709–1712.

[00148] Bart, A., Barnabé, C., Achtman, M., Dankert, J., van der Ende, A., and Tibayrenc, M. (2001). The population structure of *Neisseria meningitidis* serogroup A fits the predictions for clonality. *Infect. Genet. Evol.* 1, 117-122.

[00149] Bennett, J.S., Bentley, S.D., Vernikos, G.S., Quail, M.A., Cherevach, I., White, B., Parkhill, J., and Maiden, M.C.J. (2010). Independent evolution of the core and accessory gene sets in the genus *Neisseria*: insights gained from the genome of *Neisseria lactamica* isolate 020-06. *BMC Genomics* 11, 652.

- [00150] Bhaya, D., Davison, M., and Barrangou, R. (2011). CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. *Annu. Rev. Genet.* 45, 273–297.
- [00151] Bikard, D., Hatoun-Aslan, A., Mucida, D., and Marraffini, L.A. (2012). CRISPR interference can prevent natural transformation and virulence acquisition during in vivo bacterial infection. *Cell Host & Microbe* 12, 177–186.
- [00152] Bille, E., Ure, R., Gray, S.J., Kaczmarek, E.B., McCarthy, N.D., Nassif, X., Maiden, M.C.J., and Tinsley, C.R. (2008). Association of a bacteriophage with meningococcal disease in young adults. *PloS One* 3, e3885.
- [00153] Bille, E., Zahar, J.-R., Perrin, A., Morelle, S., Kriz, P., Jolley, K.A., Maiden, M.C.J., Dervin, C., Nassif, X., and Tinsley, C.R. (2005). A chromosomally integrated bacteriophage in invasive meningococci. *J. Exp. Med.* 201, 1905–1913.
- [00154] Bolotin, A., Quinquis, B., Sorokin, A., and Ehrlich, S.D. (2005). Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* 151, 2551–2561.
- [00155] Braid, M.D., Silhavy, J.L., Kitts, C.L., Cano, R.J., and Howe, M.M. (2004). Complete genomic sequence of bacteriophage B3, a Mu-like phage of *Pseudomonas aeruginosa*. *J. Bact.* 186, 6560–6574.
- [00156] Bratcher, H.B., Bennett, J.S., and Maiden, M.C.J. (2012). Evolutionary and genomic insights into meningococcal biology. *Future Microbiol.* 7, 873–885.
- [00157] Brouns, S.J.J., Jore, M.M., Lundgren, M., Westra, E.R., Slijkhuis, R.J.H., Snijders, A.P.L., Dickman, M.J., Makarova, K.S., Koonin, E. V., and Van der Oost, J. (2008). Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* 321, 960–964.

- [00158]** Buckee, C.O., Jolley, K.A., Recker, M., Penman, B., Kriz, P., Gupta, S., and Maiden, M.C.J. (2008). Role of selection in the emergence of lineages and the evolution of virulence in *Neisseria meningitidis*. *Proc. Natl. Acad. Sci. USA* *105*, 15082–15087.
- [00159]** Budroni, S., Siena, E., Dunning Hotopp, J.C., Seib, K.L., Serruto, D., Nofroni, C., Comanducci, M., Riley, D.R., Daugherty, S.C., Angiuoli, S.V., et al. (2011). *Neisseria meningitidis* is structured in clades associated with restriction modification systems that modulate homologous recombination. *Proc. Natl. Acad. Sci. USA* *108*, 4494–4499.
- [00160]** Carte, J., Wang, R., Li, H., Terns, R.M., and Terns, M.P. (2008). Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. *Genes Dev.* *22*, 3489–3496.
- [00161]** Cehovin, A., Simpson, P.J., McDowell, M.A., Brown, D.R., Noschese, R., Pallett, M., Brady, J., Baldwin, G.S., Lea, S.M., Matthews, S.J., and Pelicic, V. (2013). Specific DNA recognition mediated by a type IV pilin. *Proc. Natl. Acad. Sci. USA* *110*, 3065–3070.
- [00162]** Chao, Y., Papenfort, K., Reinhardt, R., Sharma, C.M., and Vogel, J. (2012). An atlas of Hfq-bound transcripts reveals 3' UTRs as a genomic reservoir of regulatory small RNAs. *EMBO J.* *31*, 4005–4019.
- [00163]** Chen, I., Christie, P.J., and Dubnau, D. (2005). The ins and outs of DNA transfer in bacteria. *Science* *310*, 1456–1460.
- [00164]** Cho, S.W., Kim, S., Kim, J.M., and Kim, J.-S. (2013). Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nature Biotech.* *31*, 230–232.
- [00165]** Chylinski, K., Le Rhun, A., and Charpentier, E. (2013). The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems. *RNA Biol.*, in press [ePub ahead of print (doi:10.4161/ma24321)].

- [00166] Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., et al. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science* 339, 819–823.
- [00167] Corander, J., Connor, T.R., O'Dwyer, C.A., Kroll, J.S., and Hanage, W.P. (2012). Population structure in the *Neisseria*, and the biological significance of fuzzy species. *J. Royal Soc. Interface* 9, 1208-1215.
- [00168] Deltcheva, E., Chylinski, K., Sharma, C.M., Gonzales, K., Chao, Y., Pirzada, Z.A., Eckert, M.R., Vogel, J., and Charpentier, E. (2011). CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* 471, 602–607.
- [00169] Deng, L., Kenchappa, C.S., Peng, X., She, Q., and Garrett, R.A. (2012). Modulation of CRISPR locus transcription by the repeat-binding protein CbpI in *Sulfolobus*. *Nucleic Acids Res.* 40, 2470-2480.
- [00170] Deveau, H., Barrangou, R., Garneau, J.E., Labonté, J., Fremaux, C., Boyaval, P., Romero, D.A., Horvath, P., and Moineau, S. (2008). Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *J. Bact.* 190, 1390–1400.
- [00171] DiCarlo, J.E., Norville, J.E., Mali, P., Rios, X., Aach, J., and Church, G.M. (2013). Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. *Nucleic Acids Res.* 41, 4336-4343.
- [00172] Duffin, P.M., and Seifert, H.S. (2012). Genetic transformation of *Neisseria gonorrhoeae* shows a strand preference. *FEMS Microbiol. Letters* 334, 44–48.
- [00173] Edgar, R., and Qimron, U. (2010). The *Escherichia coli* CRISPR system protects from λ lysogenization, lysogens, and prophage induction. *J. Bact.* 192, 6291–6294.
- [00174] Feil, E.J., and Spratt, B.G. (2001). Recombination and the population structures of bacterial pathogens. *Annu. Rev. Microbiol.* 55, 561-590.

- [00175] Garneau, J.E., Dupuis, M.-É., Villion, M., Romero, D.A., Barrangou, R., Boyaval, P., Fremaux, C., Horvath, P., Magadán, A.H., and Moineau, S. (2010). The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* 468, 67–71.
- [00176] Gasiunas, G., Barrangou, R., Horvath, P., and Siksnys, V. (2012). Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc. Natl. Acad. Sci. USA* 109, 2579–2586.
- [00177] Gesner, E.M., Schellenberg, M.J., Garside, E.L., George, M.M., and Macmillan, A.M. (2011). Recognition and maturation of effector RNAs in a CRISPR interference pathway. *Nature Struct. Mol. Biol.* 18, 688–692.
- [00178] Grissa, I., Vergnaud, G., and Pourcel, C. (2007). The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinformatics* 8, 172.
- [00179] Gudbergssdottir, S., Deng, L., Chen, Z., Jensen, J.V.K., Jensen, L.R., She, Q., and Garrett, R.A. (2011). Dynamic properties of the *Sulfolobus* CRISPR/Cas and CRISPR/Cmr systems when challenged with vector-borne viral and plasmid genes and protospacers. *Mol. Microbiol.* 79, 35–49.
- [00180] Haft, D.H., Selengut, J., Mongodin, E.F., and Nelson, K.E. (2005). A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLoS Comp. Biol.* 1, e60.
- [00181] Hale, C., Kleppe, K., Terns, R.M., and Terns, M.P. (2008). Prokaryotic silencing (psi)RNAs in *Pyrococcus furiosus*. *RNA* 14, 2572–2579.
- [00182] Hale, C.R., Majumdar, S., Elmore, J., Pfister, N., Compton, M., Olson, S., Resch, A.M., Glover, C.V.C., Graveley, B.R., Terns, R.M., et al. (2012). Essential features and rational design of CRISPR RNAs that function with the Cas RAMP module complex to cleave RNAs. *Mol. Cell* 45, 292–302.

- [00183] Hale, C.R., Zhao, P., Olson, S., Duff, M.O., Graveley, B.R., Wells, L., Terns, R.M., and Terns, M.P. (2009). RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell* 139, 945–956.
- [00184] Hamilton, H.L., and Dillard, J.P. (2006). Natural transformation of *Neisseria gonorrhoeae*: from DNA donation to homologous recombination. *Mol. Microbiol.* 59, 376–385.
- [00185] Hanage, W.P., Fraser, C., and Spratt, B.G. (2005). Fuzzy species among recombinogenic bacteria. *BMC Biol.* 3, 6.
- [00186] Haurwitz, R.E., Jinek, M., Wiedenheft, B., Zhou, K., and Doudna, J.A. (2010). Sequence- and structure-specific RNA processing by a CRISPR endonuclease. *Science* 329, 1355–1358.
- [00187] Hook-Barnard, I.G., and Hinton, D.M. (2007). Transcription initiation by mix and match elements: flexibility for polymerase binding to bacterial promoters. *Gene Reg. Systems Biol.* 1, 275–293.
- [00188] Horvath, P., Romero, D.A., Coûté-Monvoisin, A.-C., Richards, M., Deveau, H., Moineau, S., Boyaval, P., Fremaux, C., and Barrangou, R. (2008). Diversity, activity, and evolution of CRISPR loci in *Streptococcus thermophilus*. *J. Bact.* 190, 1401–1412.
- [00189] Hwang, W.Y., Fu, Y., Reyon, D., Maeder, M.L., Tsai, S.Q., Sander, J.D., Peterson, R.T., Yeh, J.-R.J., and Joung, J.K. (2013). Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nature Biotech.* 31, 227–229.
- [00190] Jiang, W., Bikard, D., Cox, D., Zhang, F., and Marraffini, L.A. (2013). RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nature Biotech.* 31, 233–239.
- [00191] Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337, 816–821.

- [00192] Jinek, M., East, A., Cheng, A., Lin, S., Ma, E., and Doudna, J. (2013). RNA-programmed genome editing in human cells. *eLife* 2, e00471.
- [00193] Jolley, K.A., and Maiden, M.C.J. (2010). BIGSdb: Scalable analysis of bacterial genome variation at the population level. *BMC Bioinformatics* 11, 595.
- [00194] Joseph, B., Schwarz, R.F., Linke, B., Blom, J., Becker, A., Claus, H., Goesmann, A., Frosch, M., Müller, T., Vogel, U., et al. (2011). Virulence evolution of the human pathogen *Neisseria meningitidis* by recombination in the core and accessory genome. *PloS One* 6, e18441.
- [00195] Karginov, F. V., and Hannon, G.J. (2010). The CRISPR system: small RNA-guided defense in bacteria and archaea. *Mol. Cell* 37, 7–19.
- [00196] Kawai, M., Nakao, K., Uchiyama, I., and Kobayashi, I. (2006). How genomes rearrange: genome comparison within bacteria *Neisseria* suggests roles for mobile elements in formation of complex genome polymorphisms. *Gene* 383, 52–63.
- [00197] Kawai, M., Uchiyama, I., and Kobayashi, I. (2005). Genome comparison in silico in *Neisseria* suggests integration of filamentous bacteriophages by their own transposase. *DNA Research* 12, 389–401.
- [00198] Koonin, E. V., and Makarova, K.S. (2013). CRISPR-Cas: evolution of an RNA-based adaptive immunity system in prokaryotes. *RNA Biol.*, in press [ePub ahead of print (doi:10.4161/rna24022)].
- [00199] Magadán, A.H., Dupuis, M.-È., Villion, M., and Moineau, S. (2012). Cleavage of phage DNA by the *Streptococcus thermophilus* CRISPR3-Cas system. *PLoS ONE* 7, e40913.
- [00200] Makarova, K.S., Aravind, L., Wolf, Y.I., and Koonin, E. V. (2011a). Unification of Cas protein families and a simple scenario for the origin and evolution of CRISPR-Cas systems. *Biol. Direct* 6, 38.

- [00201]** Makarova, K.S., Grishin, N. V, Shabalina, S.A., Wolf, Y.I., and Koonin, E. V (2006). A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biol. Direct* 1, 7.
- [00202]** Makarova, K.S., Haft, D.H., Barrangou, R., Brouns, S.J.J., Charpentier, E., Horvath, P., Moineau, S., Mojica, F.J.M., Wolf, Y.I., Yakunin, A.F., et al. (2011b). Evolution and classification of the CRISPR-Cas systems. *Nature Rev. Microbiol.* 9, 467–477.
- [00203]** Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. (2013). RNA-guided human genome engineering via Cas9. *Science* 339, 823–826.
- [00204]** Marraffini, L.A. and Sontheimer, E.J. (2008). CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science* 322, 1843–1845.
- [00205]** Marraffini, L.A. and Sontheimer, E.J. (2010). CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. *Nature Rev. Genet.* 11, 181–190.
- [00206]** Massignani, V., Giuliani, M.M., Tettelin, H., Comanducci, M., Rappuoli, R., and Scarlato, V. (2001). Mu-like Prophage in serogroup B *Neisseria meningitidis* coding for surface-exposed antigens. *Infect. Immun.* 69, 2580–2588.
- [00207]** Mojica, F.J.M., Díez-Villaseñor, C., García-Martínez, J., and Almendros, C. (2009). Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* 155, 733–740.
- [00208]** Mojica, F.J.M., Díez-Villaseñor, C., García-Martínez, J., and Soria, E. (2005). Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J. Mol. Evol.* 60, 174–182.
- [00209]** Moxon, E.R., and Jansen, V.A.A. (2005). Phage variation: understanding the behaviour of an accidental pathogen. *Trends Microbiol.* 13, 563–565.

- [00210] Palmer, K., and Gilmore, M. (2010). Multidrug-resistant enterococci lack CRISPR-cas. *mBio* *1*, e00227–10.
- [00211] Peng, J., Yang, L., Yang, F., Yang, J., Yan, Y., Nie, H., Zhang, X., Xiong, Z., Jiang, Y., Cheng, F., et al. (2008). Characterization of ST-4821 complex, a unique *Neisseria meningitidis* clone. *Genomics* *91*, 78–87.
- [00212] Pines, O., Yoon, H., and Inouye, M. (1988). Expression of double-stranded-RNA-specific RNase III of *Escherichia coli* is lethal to *Saccharomyces cerevisiae*. *J. Bact.* *170*, 2989–2993.
- [00213] Pourcel, C., Salvignol, G., and Vergnaud, G. (2005). CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology* *151*, 653–663.
- [00214] Qi, L.S., Larson, M.H., Gilbert, L.A., Doudna, J.A., Weissman, J.S., Arkin, A.P., and Lim, W.A. (2013). Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* *152*, 1173–1183.
- [00215] Rousseau, C., Gomet, M., Le Romancer, M., and Nicolas, J. (2009). CRISPRi: a CRISPR interactive database. *Bioinformatics* *25*, 3317–3318.
- [00216] Rusniok, C., Vallenet, D., Floquet, S., Ewles, H., Mouzé-Soulama, C., Brown, D., Lajus, A., Buchrieser, C., Médigue, C., Glaser, P., et al. (2009). NeMeSys: a biological resource for narrowing the gap between sequence and function in the human pathogen *Neisseria meningitidis*. *Genome Biol.* *10*, R110.
- [00217] Sampson, T.R., Saroj, S.D., Llewellyn, A.C., Tzeng, Y.-L., and Weiss, D.S. (2013). A CRISPR/Cas system mediates bacterial innate immune evasion and virulence. *Nature*, in press [ePub ahead of print (doi:10.1038/nature12048)].
- [00218] Sapranaukas, R., Gasiunas, G., Fremaux, C., Barrangou, R., Horvath, P., and Siksnys, V. (2011). The *Streptococcus thermophilus* CRISPR/Cas system provides immunity in *Escherichia coli*. *Nucleic Acids Res.* *39*, 9275–9282.

- [00219] Sashital, D.G., Wiedenheft, B., and Doudna, J.A. (2012). Mechanism of foreign DNA selection in a bacterial adaptive immune system. *Mol. Cell* 46, 606–615.
- [00220] Semenova, E., Jore, M.M., Datsenko, K.A., Semenova, A., Westra, E.R., Wanner, B., van der Oost, J., Brouns, S.J.J., and Severinov, K. (2011). Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proc. Natl. Acad. Sci. U.S.A.* 108, 10098–10103.
- [00221] Sharma, C.M., Hoffmann, S., Darfeuille, F., Reignier, J., Findeiss, S., Sittka, A., Chabas, S., Reiche, K., Hackermüller, J., Reinhardt, R., et al. (2010). The primary transcriptome of the major human pathogen *Helicobacter pylori*. *Nature* 464, 250–255.
- [00222] Sittka, A., Lucchini, S., Papenfort, K., Sharma, C.M., Rolle, K., Binnewies, T.T., Hinton, J.C.D., and Vogel, J. (2008). Deep sequencing analysis of small noncoding RNA and mRNA targets of the global post-transcriptional regulator, Hfq. *PLoS Genet.* 4, e1000163.
- [00223] Sittka, A., Pfeiffer, V., Tedin, K., and Vogel, J. (2007). The RNA chaperone Hfq is essential for the virulence of *Salmonella typhimurium*. *Molec. Microbiol.* 63, 193–217.
- [00224] Skaar, E.P., Lecuyer, B., Lenich, A.G., Lazio, M.P., Perkins-Balding, D., Seifert, H.S., and Karls, A.C. (2005). Analysis of the Piv recombinase-related gene family of *Neisseria gonorrhoeae*. *J. Bact.* 187, 1276–1286.
- [00225] Smith, J.M., Smith, N.H., O'Rourke, M., and Spratt, B.G. (1993). How clonal are bacteria? *Proc. Natl. Acad. Sci. USA* 90, 4384–4388.
- [00226] Stephens, D.S., Greenwood, B., and Brandtzaeg, P. (2007). Epidemic meningitis, meningococcaemia, and *Neisseria meningitidis*. *Lancet* 369, 2196–2210.
- [00227] Stern, A., Keren, L., Wurtzel, O., Amitai, G., and Sorek, R. (2010). Self-targeting by CRISPR: gene regulation or autoimmunity? *Trends Genet.* 26, 335–340.

- [00228] Terns, M.P., and Terns, R.M. (2011). CRISPR-based adaptive immune systems. *Curr. Opin. Microbiol.* *14*, 321–327.
- [00229] Tobiasson, D.M., and Seifert, H.S. (2010). Genomic content of *Neisseria* species. *J. Bact.* *192*, 2160–2168.
- [00230] van Passel, M.W.J., Van der Ende, A., and Bart, A. (2006). Plasmid diversity in *neisseriae*. *Infect. Immun.* *74*, 4892–4899.
- [00231] Vázquez, J.A., de la Fuente, L., Berron, S., O'Rourke, M., Smith, N.H., Zhou, J., and Spratt, B.G. (1993). Ecological separation and genetic isolation of *Neisseria gonorrhoeae* and *Neisseria meningitidis*. *Curr. Biol.* *3*, 567–572.
- [00232] Westra, E.R., Van Erp, P.B.G., Künne, T., Wong, S.P., Staals, R.H.J., Seegers, C.L.C., Bollen, S., Jore, M.M., Semenova, E., Severinov, K., et al. (2012). CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3. *Mol. Cell* *46*, 595–605.
- [00233] Wiedenheft, B., Van Duijn, E., Bultema, J.B., Bultema, J., Waghmare, S.P., Waghmare, S., Zhou, K., Barendregt, A., Westphal, W., Heck, A.J.R., et al. (2011). RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions. *Proc. Natl. Acad. Sci. USA* *108*, 10092–10097.
- [00234] Wiedenheft, B., Sternberg, S.H., and Doudna, J.A. (2012). RNA-guided genetic silencing systems in bacteria and archaea. *Nature* *482*, 331–338.
- [00235] Yosef, I., Goren, M.G., and Qimron, U. (2012). Proteins and DANN elements essential for the CRISPR adaptation process in *Escherichia coli*. *Nucleic Acids Res.* *40*, 5569–5576.
- [00236] Zhang, J., Rouillon, C., Kerou, M., Reeks, J., Brugger, K., Graham, S., Reimann, J., Cannone, G., Liu, H., Albers, S.-V., et al. (2012). Structure and mechanism of the CMR complex for CRISPR-mediated antiviral immunity. *Mol. Cell* *45*, 303–313.

[00237] Example 2

[00238] Reference is made to Hou *et al.*, "Efficient genome engineering in human pluripotent stem cells using Cas9 from *Neisseria meningitidis*", PNAS, vol. 110, no. 39, pp 15644-15649, September 24, 2013, the contents of which are incorporated herein by reference in its entirety.

[00239] Abstract

[00240] Genome engineering in human pluripotent stem cells holds great promise for biomedical research and regenerative medicine. Recently, an RNA-guided, DNA-cleaving interference pathway from bacteria [the Type II clustered, regularly interspaced, short palindromic repeats (CRISPR)-CRISPR-associated (Cas) pathway] has been adapted for use in eukaryotic cells, greatly facilitating genome editing. Only two CRISPR-Cas systems (from *Streptococcus pyogenes* and *Streptococcus thermophilus*), each with their own distinct targeting requirements and limitations, have been developed for genome editing thus far. Furthermore, limited information exists about homology-directed repair (HDR)-mediated gene targeting using long donor DNA templates in human pluripotent stem cells (hPSCs) with these systems. Here, using a distinct CRISPR-Cas system from *Neisseria meningitidis*, we demonstrate efficient targeting of an endogenous gene in three hPSC lines using HDR. The Cas9 RNA-guided endonuclease from *N. meningitidis* (NmCas9) recognizes a 5'-NNNNGATT-3' protospacer adjacent motif (PAM) different from those recognized by Cas9 proteins from *S. pyogenes* and *S. thermophilus* (SpCas9 and StCas9, respectively). Similar to SpCas9, NmCas9 is able to use a single-guide RNA (sgRNA) to direct its activity. Due to its distinct PAM, the *N. meningitidis* CRISPR-Cas machinery increases the sequence contexts amenable to RNA-directed genome editing.

[00241] Introduction

[00242] Human pluripotent stem cells (hPSCs) can proliferate indefinitely while maintaining the potential to give rise to virtually all human cell types (1). They are therefore invaluable for regenerative medicine, drug screening, and biomedical research. However, to realize the full potential of hPSCs, it will be necessary to manipulate their genomes in a precise, efficient manner. Historically, gene targeting in hPSCs has been

extremely difficult (2). The development of zinc-finger nucleases (ZFNs) and transcription activator-like endonucleases (TALENs) (reviewed in refs. (3) and (4)) has facilitated gene targeting in hPSCs (5-7). Nonetheless, they require the design, expression, and validation of a new pair of proteins for every targeted locus, rendering both of these platforms time-consuming and labor-intensive (8-10).

[00243] Clustered, regularly interspaced, short palindromic repeat (CRISPR) loci, along with CRISPR-associated (*cas*) genes, underlie an adaptive immune system of bacteria and archaea that defends against bacteriophages (11) and limits horizontal gene transfer (12-14). "Protospacer" sequences from invading nucleic acids are incorporated as "spacers" within CRISPRs, conferring immunity and providing a genomic memory of past invasions. CRISPR-Cas systems have been classified into three types (Types I, II and III) and numerous subtypes (15). All use short CRISPR RNAs (crRNAs) (16, 17) to specify genetic interference via the destruction of invading nucleic acids (18). The target nucleic acids are recognized by crRNA Watson-Crick pairing. Importantly, most CRISPR-Cas subtypes target DNA directly (13, 19, 20), suggesting the possibility of engineered, RNA-directed gene targeting/editing systems. The use of RNA guides for gene targeting would confer many advantages over ZFNs and TALENs, especially by obviating the need for repeated protein design/optimization. Recently, this vision has become a reality (21-31).

[00244] Type II CRISPR-Cas systems are noteworthy in that the essential targeting activities – crRNA binding, target DNA binding, R-loop formation, and double-stranded DNA cleavage – are all executed by a single polypeptide, Cas9 (32-35). In addition to crRNA and Cas9, an additional RNA, trans-acting CRISPR RNA (tracrRNA), is essential for interference in bacteria (14, 32, 36) and *in vitro* (34, 36). The tracrRNA is partially complementary to pre-crRNA repeats, leading to the formation of duplexes that are cleaved by the host factor ribonuclease III (RNase III) (32). The Type II crRNA maturation pathway was originally characterized in strains of *Streptococcus pyogenes* (32) and *Streptococcus thermophilus* (35, 36), and RNase III-catalyzed pre-crRNA processing is essential for interference in both native systems. Recent studies of a Type II CRISPR-Cas locus from *Neisseria meningitidis* revealed an intrinsically RNase III- and processing-independent system, which nonetheless requires tracrRNA (14). Importantly, crRNA-directed DNA cleavage was reconstituted *in vitro* with recombinant *S. pyogenes* Cas9

(SpCas9) (34) or *S. thermophilus* Cas9 (StCas9) (33, 36). The SpCas9 *in vitro* system enabled the development of fused crRNA-tracrRNA chimeras called single-guide RNAs (sgRNAs) that bypass processing (34). Subsequent development of eukaryotic genome editing applications has focused on sgRNAs (21-30), though separately encoded pre-crRNAs and tracrRNAs are also effective (21).

[00245] Target cleavage by many CRISPR/Cas systems, including those from Type II, require proximity to a 2-5 nucleotide (nt) sequence called a protospacer adjacent motif (PAM) (37) (38-40). Genome editing applications reported thus far have focused almost exclusively on SpCas9, which has a 5'-NGG-3' PAM. StCas9 (from the CRISPR1 locus of strain LMD-9) has also been used in eukaryotes (21), and that system has a 5'-NNAGAAW-3' PAM (W = A or T). Eukaryotic editing capabilities will benefit from the increased frequency of target sites stemming from the development of additional Cas9s with distinct PAMs.

[00246] Targeting by sgRNAs usually relies on either of two approaches. First, double-strand break (DSB) repair by nonhomologous end joining (NHEJ) can be used to generate insertions or deletions (indels) that induce frame shifts. Second, the addition of a homologous repair template can allow Cas9-induced DSBs or nicks to be repaired by homology-directed repair (HDR). The latter strategy is useful for making precise changes such as repairing mutations or inserting transgenes. Most studies thus far have relied on either NHEJ, or on HDR using short DNA fragments or oligos (24-26, 29, 31). Currently there is very limited information available on gene targeting using long DNA donor templates in hPSCs (23).

[00247] Here, we report the development of *N. meningitidis* Cas9 (NmCas9) (14) as a genome editing platform, and its application to high-efficiency targeting of an endogenous gene in hPSCs. This system uses a 24 nt proto-spacer for targeting and requires a PAM that is different from those of SpCas9 or StCas9. We have achieved ~60% targeting efficiency with two human embryonic stem cell (hESC) lines and one human induced pluripotent stem (iPS) cell line. Our work demonstrates the feasibility of using the *N. meningitidis* CRISPR/Cas system in genome editing in hPSCs using long DNA donor templates. This work also provides an alternative to the *S. pyogenes* and *S. thermophilus*

CRISPR-Cas system and expands the genomic contexts that are amenable to RNA-directed genome editing in eukaryotes.

[00248] Results

[00249] Functional Expression of NmCas9 in Mammalian Cells. Our recent work has shown that *N. meningitidis* strain 8013 has a functional type II-C CRISPR/Cas system (14), and that Cas9 is the only Cas protein required for interference activity. We set out to test whether this system could be used for efficient gene targeting in hPSCs. We cloned the open reading frame (ORF) from the 3.25 kb *cas9* gene, along with a C-terminal FLAG tag, into a mammalian expression plasmid under the control of an EF1 α promoter (Fig. 13A). This NmCas9-containing vector was transfected into 293FT cells and the expression of NmCas9 protein was analyzed by anti-FLAG western blot. As shown in Fig. 13A, full-length NmCas9 was efficiently expressed in 293FT cells. We then assayed the nuclease activity of NmCas9 expressed in mammalian cells by *in vitro* plasmid cleavage. Cell extract was prepared from 293FT cells two days after transfection with the NmCas9-containing vector (the same one as in Fig. 13A). We assembled cleavage reactions using cell extract, various *in vitro*-synthesized small RNAs, and the plasmid ptdTomato pre-linearized by NdeI (Fig. 13B). tdTomato is a fusion of two copies of the dTomato gene, each of which has one consensus PAM sequence (5'-NNNNGATT-3') (Fig. 13B). As shown in Fig. 13C, we achieved efficient plasmid cleavage only in the presence of both tracrRNA and a cognate crRNA (Fig. 13C, lane 3). The pattern of the cleavage products was consistent with two predicted cleavage sites in the PAM-proximal regions (Fig. 13C, right panel). Importantly, a non-cognate crRNA (N), which contains sequences from EGFP, did not direct NmCas9-mediated cleavage (Fig. 13C, lane 4), indicating that the specificity of the NmCas9 nuclease is indeed guided by the spacer-derived sequence in crRNAs. Additionally, plasmid cleavage is deficient when tracrRNA is absent, even in the presence of a cognate crRNA (Fig. 13C, lane 5), suggesting that tracrRNA is necessary for NmCas9 function *in vitro*. This is consistent with the tracrRNA requirement for NmCas9-mediated interference in bacterial cells (14).

[00250] Two Cas9 orthologs, SpoCas9 and StCas9, were previously demonstrated to induce blunt double-strand breaks (DSB) in their DNA targets, between the third and

fourth nucleotide counting from the PAM-proximal end of protospacers (34) (19, 33). We hypothesized that NmCas9 cleaves the DNA target in a similar way, and we tested this by mapping the NmCas9 cleavage site on ptdTomato by Sanger sequencing. Two cleavage products in Fig. 13C (the 1.5 kb and the 2.5 kb fragments) were gel-extracted and sequenced to identify the NmCas9 cleavage sites on the sense strand and the antisense strand, respectively. As expected, NmCas9 induced a blunt-end DSB between the third and fourth nucleotides counting from the PAM-proximal end of the proto-spacer (Fig. 13D).

[00251] NmCas9 Functions in RNA-Directed Gene Disruption in hPSCs. Knowing that NmCas9, without any codon optimization, can be efficiently expressed in mammalian cells and is functional *in vitro*, we next tested its utility in genome editing in hPSCs. We first monitored its localization. We transfected 293FT cells with several NmCas9 constructs with various nuclear localization signal (NLS) arrangements, and analyzed NmCas9 protein localization by either GFP fluorescence or anti-HA immunostaining. NmCas9 with NLSs on both N- and C-termini localized efficiently to the nucleus (Fig. 14C), while NmCas9 constructs with just one NLS did not (Fig. 14A and 14B). In addition, the same NmCas9 construct with two NLSs also localized to the nucleus of hESCs (Fig. 14D). We noticed that in hESCs, NmCas9, without any crRNA/tracrRNA, displayed a punctate pattern similar to the organization of the nucleolus in hESCs. It is not yet clear if this phenomenon is related to the organization of the double NLS on the protein.

[00252] To test the genome editing activity of NmCas9, we used an hESC cell reporter line that has a single copy of the tdTomato fluorescent protein gene knocked into the highly expressed DNMT3b locus (H9 DNMT3b-tdTomato), leading to tdTomato fluorescence. If NmCas9 is able to introduce a DSB in the tdTomato sequence in the genome, repair by NHEJ would likely lead to indels that disrupt tdTomato expression. Accordingly, the appearance of tdTomato-negative cells would be predicted to reflect genome-editing activity.

[00253] Human ESCs are known to have low transfection efficiencies. To achieve maximum genome-editing efficiency in hPSCs, we assembled expression cassettes of all

the necessary components (NmCas9, tracrRNA and crRNA) onto one single plasmid that contains an OriP sequence (Fig. 14E). OriP was reported to increase the transfection efficiency and plasmid stability in hPSCs if co-transfected with an RNA expressing the EBNA protein (41, 42). The encoded tracrRNA and crRNA both corresponded to the mature, processed forms as they exist in *N. meningitidis* cells (14). The resulting all-in-one plasmids were electroporated into H9 DNMT3b-tdTomato cells, and tdTomato fluorescence was monitored by FACS 4-6 days after electroporation. As shown in Fig. 14F, a sub-population (5.4%) of tdTomato negative cells became detectable only when a tdTomato-targeting crRNA was encoded on the plasmid. Importantly, for the control plasmid expressing non-targeting crRNA, only background levels (~0.1%) of tdTomato negative cells appeared (Fig. 14F), likely due to the low level of spontaneous differentiation in the culture, leading to repression of the DNMT3b promoter. The increased frequency of non-fluorescent cells in the presence of the cognate crRNA suggests successful genome editing by NmCas9.

[00254] To confirm that NmCas9 introduced a DSB at the intended genomic site, we FACS-sorted the tdTomato negative population, PCR-amplified the genomic region flanking the predicted cutting site in the 5' copy of dTomato, cloned the resulting PCR fragments and sequenced 22 of the resulting plasmids (selected at random). The sequencing results showed both insertions and deletions in the tdTomato sequence (Fig. 14G, only unique indels are shown) in 95% of the sequenced clones. Most importantly, all of these indels were centered around the NmCas9 cleavage site, indicating that the DSB occurred at the intended position (Fig. 14G).

[00255] A Chimeric sgRNA is Effective for Gene Editing in hPSCs. To simplify the NmCas9 genome editing system, we explored the possibility of substituting both crRNA and tracrRNA with a chimeric sgRNA. We fused the 5' end of the 91 nt processed tracrRNA sequence with the 3' end of the 48 nt mature crRNA using a 6 nt linker (Supp. Fig. 13A). This sgRNA was cloned under the control of the U6 promoter and electroporated into the H9 DNMT3b-tdTomato reporter cell line together with a plasmid expressing NmCas9. FACS analysis showed that this sgRNA indeed resulted in tdTomato-negative cells (Fig. 17B) at a level comparable to that achieved by the all-in-one plasmid expressing separate crRNA and tracrRNA (see Fig. 14E). These results indicated that an

sgRNA could substitute for separate crRNA and tracrRNA in directing NmCas9-mediated gene editing in hESCs.

[00256] Specificity of NmCas9 in hPSCs. We next tested the specificity of NmCas9 in mammalian cells by mutational analysis. We introduced single-nucleotide mutations at every odd-numbered position from the 1st to the 17th nt in the PAM-proximal end [spanning the cleavage site (see Figs. 13D and 14G) and the functionally critical “seed” sequence] of the spacer in the tdTomato-targeting crRNA construct (Fig. 15A, upper panel). We then measured the ability of those constructs to give rise to tdTomato-negative cells in the H9 DNMT3b-tdTomato cell line. As shown in Fig. 15A, mutations at position 1 through 9 led to background levels of tdTomato-negative cells, indicating that mismatches at these positions in the crRNA/target duplex are not tolerated by NmCas9. As for mutations at positions 11, 13, 15, some tdTomato negative cells appeared, but with an efficiency of only 10-25% of that observed with wildtype crRNA (Fig. 15A lower panel). The mismatch at position 17 was ~40% as efficient as wildtype. These results imply a crRNA/target specificity comparable to that of the SpCas9 system in mammalian cells (21).

[00257] We also investigated PAM sequence requirements for NmCas9 in human ES cells. We designed five crRNAs that use different sequences as the PAM in the tdTomato coding region (Fig. 15B) and then tested their ability to disrupt tdTomato expression in H9 DNMT3b-tdTomato cells. Four of the sites were associated with a PAM that varied from the 5'-NNNNGATT-3' consensus by only a single nt. Only a GCTT variant site was efficiently targeted, while the other four variants were severely deficient (Fig. 15B, lower panel). Our results indicate that an A-to-C mutation at the 2nd nt of the PAM could be tolerated, whereas a G-to-C mutation at the 1st position, T-to-C at the 3rd, and T-to-G at the 4th likely render the PAM variants non-functional. Interestingly, C is the second most frequent residue at the 2nd nt of the PAM in candidate bacterial protospacers (14), suggesting that GCTT might also be a natural PAM.

[00258] NmCas9 Increases Gene Targeting Efficiency in hPSCs. We next explored whether NmCas9 can increase gene-targeting efficiency in hPSCs compared to the traditional method in which no DSB was intentionally introduced at the target site. We

used a donor DNA template previously used to target the endogenous *POU5F1* (*OCT4*) gene (6) (Fig. 16A), creating a fusion of *OCT4* with *EGFP*. We designed the crRNA using the consensus PAM sequence located ~84bp downstream of the *OCT4* stop codon (Fig. 16A). Two human ES cell lines, H1 and H9, and one human iPS cell line, iPS005 (43), were used in the experiment. After puromycin selection, we were able to obtain clones for all three cell lines when plasmid expressing the *OCT4*-targeting crRNA was used. Of these clones, ~60% were correctly targeted with single insertion events (Table 7), comparable to the efficiency obtained using TALENs in a previous report with the same donor DNA (6). Fluorescent images of the targeted clones revealed the expected nuclear localization of EGFP signal due to the fusion with Oct4 protein (Fig. 16B). Southern blots using a probe outside the targeting vector's homology arm confirmed the correct integration of the donor sequence in the *OCT4* locus (Fig. 16C). Most importantly, the EGFP signals respond to differentiation cues as the endogenous Oct4 would (Fig. 16D). In a control experiment with an all-in-one plasmid expressing a non-targeting crRNA, no puromycin-resistant clones were obtained with the H1 ESC line. Only one puromycin-resistant clone each was obtained from H9 ESCs and iPS005 iPSCs, and neither clone was correctly targeted (Table 7). All of the above results indicated that the CRISPR-Cas system from *N. meningitidis* was able to generate accurately targeted clones in hPSCs with much increased efficiency compared to the traditional method.

[00259] Discussion

[00260] Genome Editing by *N. meningitidis* Cas9. In this report, we have successfully used the Type II-C CRISPR-Cas system from *N. meningitidis* to achieve both NHEJ-mediated gene editing and long DNA donor-directed gene targeting of an endogenous locus in hPSCs. The targeting efficiency we obtain with NmCas9 is comparable to that achieved with TALENs. Using the same donor construct, we were able to get ~60% targeting efficiency in all three different hPSC lines tested (Table 7), whereas the targeting efficiency of a TALEN was 48% in the one hESC line tested (6). A previous report using SpCas9 in human iPSCs achieved a targeting efficiency of 43%, close to what we observed with NmCas9 (6). However, that report only identified seven clones and did not perform further analysis to confirm the correct integration of the donor DNA sequence

only at the intended site. Therefore additional work will be needed to compare the efficiency of mammalian gene targeting using these two CRISPR-Cas systems.

[00261] CrRNA/Target Mismatch Tolerance by NmCas9 in Mammalian Cells. One potential advantage of NmCas9, relative to SpCas9, is that it might offer better targeting specificity by virtue of its longer crRNA spacer (24 vs. 20 nts) and its longer PAM (14). We chose 24 nt as the crRNA spacer length for NmCas9 because that is the length of the crRNA spacer in *N. meningitidis*. CrRNA-target mismatches distant from the PAM were tolerated to various extents for both NmCas9 (Fig. 15B) and SpCas9 (21) in mammalian cells. However, NmCas9 was more sensitive than SpCas9 to mismatches at the 13th, 15th, and 17th nts (counting from the PAM-proximal end of the proto-spacer). NmCas9 gene editing efficiencies with mismatches at those positions were no higher than 10-40% of those observed with the perfectly matched crRNA (Fig. 15B), whereas with SpCas9, mismatches at equivalent positions retained 60-90% of the non-mismatched efficiency (21).

[00262] PAM requirements in mammalian cells. One hallmark of Type II CRISPR-Cas systems is the requirement of a nearby PAM on the target sequence. This sequence varies between different Cas9 orthologs. Among Cas9 proteins validated for mammalian genome editing, PAM functional requirements have been defined for three: those from *S. pyogenes* SF370 (21-23, 32, 34), *S. thermophilus* LMD-9 (the CRISPR1 locus) (19, 21, 38), and *N. meningitidis* 8013 (Fig. 15B) (14). On one hand, the PAM requirement adds a second layer of specificity for gene targeting, beyond that afforded by spacer/protospacer complementarity. For longer PAMs (such as the NmCas9 PAM, 5'-NNNNGATT-3'), the frequency of off-target cutting events should potentially drop significantly compared to SpCas9, which requires a 5'-NGG-3' PAM. On the other hand, longer PAM requirements also constrain the frequency of targetable sites. By developing genome-editing systems using a range of Cas9 proteins with distinct PAM requirements, the genomic regions that can be targeted by CRISPR-Cas editing would expand significantly.

[00263] The results in Fig. 15B show that NmCas9 does allow limited deviation from the 5'-NNNNGATT-3' PAM. Having a variable PAM can potentially increase the flexibility during the design of targeting construct. However, it also increases the potential

of off-target cleavage. Due to the limited options afforded by the sequence of tdTomato, we only tested one nucleotide substitution in each position of the PAM domain. It is possible that additional nucleotide substitutions will also be tolerated. A detailed mutational analysis will be needed to fully understand the PAM requirements of NmCas9 in mammalian cells.

[00264] Editing the Genomes of hPSCs. Compared to two other widely used systems for enhancing gene targeting efficiency (ZFNs and TALENs), the CRISPR-Cas system offers a much simpler and more user-friendly design. For each different genomic locus to be targeted, one only needs to design a small RNA by applying simple Watson-Crick base-pairing rules. This system's ease of use will make gene targeting in hPSCs, once considered a difficult project, a routine lab technique. This simple and high efficiency gene targeting system for hPSC will also have a tremendous impact on personalized regenerative medicine. One concern with using CRISPR/Cas in human genome editing is off-target cleavage. Our work (Fig. 15A) and that of others (21, 44) has shown that the CRISPR/Cas system can tolerate mismatches within the crRNA, especially in the PAM-distal region. This raises concerns that other regions in the genome might be cleaved unintentionally. Indeed, recent work has shown various off-target cleavage rates in the human genome using SpCas9 with different sgRNAs (44). To fully understand this issue, whole-genome sequencing of cells targeted by different Cas9 proteins with different crRNA/sgRNA constructs will be needed. A potential way to get around this problem is to use a nickase, a Cas9 variant in which one nuclease domain is inactivated by a mutation (21, 34), so that off-target cleavage will have a much lower chance of generating unwanted mutations in the genome while HDR will still be stimulated.

[00265] Materials and Methods

[00266] Cell Culture. Human ESCs and iPS cells were cultured in E8™ medium (43) on Matrigel-coated tissue culture plates with daily media change at 37°C with 5% CO₂. Cells were split every 4-5 days with 0.5 mM EDTA in 1x PBS. 293FT cells were cultured similarly in DMEM/F12 media supplemented with 10% FBS.

[00267] NmCas9 DNA Transfection and *In Vitro* Plasmid Digestion. All transfections with 293FT cells were done using Fugene HD (Promega) following the

manufacturer's instructions. Cell lysate was prepared two days after transfection. Plasmid digestion using cell lysate was carried out at 37°C for 1-4 hours in digestion buffer (1x PBS with 10 mM MgCl₂). See supplemental method for a detailed procedure. To map the cleavage site of NmCas9, the digested plasmid DNA was excised from the agarose gel and purified using Gel Extraction Kit (Qiagen). The purified fragments were then sequenced to map the cleavage site.

[00268] Gene Editing in hPSCs. All plasmids used in this experiment were purified using the MaxiPrep Kit from Qiagen. Human PSCs were passaged two or three days before the experiments. Immediately before the experiment, hPSCs were individualized by Accutase® treatment, washed once with E8™ medium, and resuspended at densities of 2.5 - 6.2 x 10⁶ cells/ml in E8™ medium with 10 mM HEPES buffer (pH 7.2 - 7.5) (Life Technologies). For electroporation, 400 µl of cell suspension, 15 µg of pSimple-Cas9-Tracr-CrRNA plasmid, 5 µg of EBNA RNA, and (for those experiments involving gene targeting by HDR) 5 µg of linearized DNA template plasmid (Addgene 31939) were mixed in a 4mm cuvette (BioRad) and immediately electroporated with a BioRad Gene Pulser. Electroporation parameters were 250V, 500 µF, and infinite resistance. Cells were then plated into appropriate Matrigel coated culture dishes in E8™ medium supplemented with 10 µM ROCK inhibitor Y-27632. Media was changed the next day to E8™ medium. For those experiments involving gene editing by HDR, puromycin selection was started 4 days after electroporation. Surviving colonies were picked 4 to 6 days after selection and expanded in E8™ medium.

[00269] Plasmid Construction. The *cas9* gene from *Neisseria meningitidis* strain 8013 was PCR-amplified and cloned into the pSimpleII plasmid (an OriP containing plasmid) under the control of the EF1α promoter. Nuclear localization signals and HA tag sequences were incorporated via the PCR primers. An *N. meningitidis* BsmBI-crRNA cassette and the *N. meningitidis* tracrRNA, both under the control of U6 RNA polymerase III promoters, were synthesized as gene blocks (Integrated DNA Technologies) and cloned into pSimpleII-Cas9 via blunt end cloning, generating the pSimple-Cas9-Tracr-BsmBI plasmid that includes all elements needed for targeting. To insert specific spacer sequences into the crRNA cassette, synthetic oligonucleotides containing the desired spacer sequences were annealed to generate a duplex with overhangs compatible with

those generated by BsmBI digestion of the pSimple-Cas9-Tracr-BsmBI plasmid. The insert was then ligated into the BsmBI-digested plasmid.

[00270] NmCas9 DNA Transfection and *In Vitro* Plasmid Digestion. All transfections with 293FT cells were done using Eugene® HD (Promega) following the manufacture's instructions. Roughly 2µg plasmids and 6µl of Eugene HD were used for one well of a 6-well plate. Two days after transfection, 293FT cells expressing NmCas9 were harvested by TrypLE (Life Technologies), washed once in PBS, and then lysed in PBS by sonication. Cellular debris was cleared by centrifugation and the supernatant was used in plasmid digestion assays. For the digestions, 1µg tdTomato plasmid (Clontech) linearized by *NdeI* (New England Biolabs) was mixed with *in vitro*-transcribed tracrRNA, crRNA and 293FT cell lysate and incubated at 37°C for 1-4 hours in digestion buffer (1x PBS with 10 mM MgCl₂). DNA from the reaction mix was then purified with a PCR clean-up kit (Qiagen) and resolved by agarose gel electrophoresis. To map the cleavage site of NmCas9, the digested plasmid DNA was excised from the agarose gel and purified using Gel Extraction Kit (Qiagen). The purified fragments were then sequenced to map the cleavage site.

[00271] *In Vitro* Transcription. Synthetic oligonucleotides (Integrated DNA Technologies) containing the T7 promoter sequence and *N. meningitidis* tracrRNA or crRNA sequences were annealed to generate dsDNA templates for run-off transcription. *In vitro* transcription was done using the MegaScript T7 In Vitro Transcription kit (Ambion) following the manufacture's specifications.

[00272] Southern Blots. Genomic DNA of targeted clones is purified using PureGene core kit (Qiagen). 5 µg of genomic DNA was digested with BamHI and then resolved on a 0.8% agarose gel. DIG-labeled DNA probe synthesis, DNA gel transfer, and blot hybridization and visualization were done according to Roche's DIG application manual.

[00273] Genome editing using single-guide RNA (sgRNA). A single-guide RNA that targets tdTomato was put under the control of a U6 promoter and cloned into the EcoRV site of pstBlue-1 (Novagen). For electroporation, 7.5 µg of pstBlue-U6-sgRNA, 7.5 µg of pSimpleII-NLS-NmCas9-HA-NLS(s) and 5µg of EBNA RNA was mixed with

~1 x 10⁶ cells in a 4 mm cuvette (BioRad) and immediately electroporated with a BioRad Gene Pulser. Cells were then plated into appropriate Matrigel coated culture dishes in E8™ medium supplemented with 10 μM ROCK inhibitor Y-27632.

[00274] References for Example 2

- [00275]** 1. Thomson JA, *et al.* (1998) Embryonic stem cell lines derived from human blastocysts. *Science* 282(5391):1145-1147.
- [00276]** 2. Zwaka TP & Thomson JA (2003) Homologous recombination in human embryonic stem cells. *Nat Biotechnol* 21(3):319-321.
- [00277]** 3. Urnov FD, Rebar EJ, Holmes MC, Zhang HS, & Gregory PD (2010) Genome editing with engineered zinc finger nucleases. *Nat Rev Genet* 11(9):636-646.
- [00278]** 4. Joung JK & Sander JD (2013) TALENs: a widely applicable technology for targeted genome editing. *Nat Rev Mol Cell Biol* 14(1):49-55.
- [00279]** 5. Hockemeyer D, *et al.* (2009) Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases. *Nat Biotechnol* 27(9):851-857.
- [00280]** 6. Hockemeyer D, *et al.* (2011) Genetic engineering of human pluripotent cells using TALE nucleases. *Nat Biotechnol* 29(8):731-734.
- [00281]** 7. Zou J, *et al.* (2009) Gene targeting of a disease-related gene in human induced pluripotent stem and embryonic stem cells. *Cell Stem Cell* 5(1):97-110.
- [00282]** 8. Zhang L, *et al.* (2000) Synthetic zinc finger transcription factor action at an endogenous chromosomal site. Activation of the human erythropoietin gene. *J Biol Chem* 275(43):33850-33860.
- [00283]** 9. Cermak T, *et al.* (2011) Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res* 39(12):e82.

- [00284] 10. Porteus MH (2006) Mammalian gene targeting with designed zinc finger nucleases. *Mol Ther* 13(2):438-446.
- [00285] 11. Barrangou R, *et al.* (2007) CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 315(5819):1709-1712.
- [00286] 12. Bikard D, Hatoun-Aslan A, Mucida D, & Marraffini LA (2012) CRISPR interference can prevent natural transformation and virulence acquisition during in vivo bacterial infection. *Cell Host Microbe* 12(2):177-186.
- [00287] 13. Marraffini LA & Sontheimer EJ (2008) CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science* 322(5909):1843-1845.
- [00288] 14. Zhang Y, *et al.* (2013) Processing-Independent CRISPR RNAs Limit Natural Transformation in *Neisseria meningitidis*. *Mol Cell* 50(4):488-503.
- [00289] 15. Makarova KS, *et al.* (2011) Evolution and classification of the CRISPR-Cas systems. *Nat Rev Microbiol* 9(6):467-477.
- [00290] 16. Brouns SJ, *et al.* (2008) Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* 321(5891):960-964.
- [00291] 17. Hale C, Kleppe K, Terns RM, & Terns MP (2008) Prokaryotic silencing (psi)RNAs in *Pyrococcus furiosus*. *RNA* 14(12):2572-2579.
- [00292] 18. Wiedenheft B, Sternberg SH, & Doudna JA (2012) RNA-guided genetic silencing systems in bacteria and archaea. *Nature* 482(7385):331-338.
- [00293] 19. Garneau JE, *et al.* (2010) The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* 468(7320):67-71.
- [00294] 20. Westra ER, *et al.* (2012) CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3. *Mol Cell* 46(5):595-605.

- [00295] 21. Cong L, *et al.* (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science* 339(6121):819-823.
- [00296] 22. Jinek M, *et al.* (2013) RNA-programmed genome editing in human cells. *Elife* 2:e00471.
- [00297] 23. Mali P, *et al.* (2013) RNA-guided human genome engineering via Cas9. *Science* 339(6121):823-826.
- [00298] 24. Wang H, *et al.* (2013) One-Step Generation of Mice Carrying Mutations in Multiple Genes by CRISPR/Cas-Mediated Genome Engineering. *Cell* 153(4):910-918.
- [00299] 25. Cho SW, Kim S, Kim JM, & Kim JS (2013) Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nat Biotechnol* 31(3):230-232.
- [00300] 26. Chang N, *et al.* (2013) Genome editing with RNA-guided Cas9 nuclease in zebrafish embryos. *Cell Res* 23(4):465-472.
- [00301] 27. DiCarlo JE, *et al.* (2013) Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. *Nucleic Acids Res* 41(7):4336-4343.
- [00302] 28. Gratz SJ, *et al.* (2013) Genome engineering of *Drosophila* with the CRISPR RNA-guided Cas9 nuclease. *Genetics*.
- [00303] 29. Hwang WY, *et al.* (2013) Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat Biotechnol* 31(3):227-229.
- [00304] 30. Xiao A, *et al.* (2013) Chromosomal deletions and inversions mediated by TALENs and CRISPR/Cas in zebrafish. *Nucleic Acids Res*.
- [00305] 31. Ding Q, *et al.* (2013) Enhanced efficiency of human pluripotent stem cell genome editing through replacing TALENs with CRISPRs. *Cell Stem Cell* 12(4):393-394.

- [00306] 32. Deltcheva E, *et al.* (2011) CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* 471(7340):602-607.
- [00307] 33. Gasiunas G, Barrangou R, Horvath P, & Siksnys V (2012) Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc Natl Acad Sci U S A* 109(39):E2579-2586.
- [00308] 34. Jinek M, *et al.* (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337(6096):816-821.
- [00309] 35. Sapranas R, *et al.* (2011) The *Streptococcus thermophilus* CRISPR/Cas system provides immunity in *Escherichia coli*. *Nucleic Acids Res* 39(21):9275-9282.
- [00310] 36. Karvelis T, *et al.* (2013) crRNA and tracrRNA guide Cas9-mediated DNA interference in *Streptococcus thermophilus*. *RNA Biol* 10(5).
- [00311] 37. Shah SA, Erdmann S, Mojica FJ, & Garrett RA (2013) Protospacer recognition motifs: Mixed identities and functional diversity. *RNA Biol* 10(5).
- [00312] 38. Deveau H, *et al.* (2008) Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *J Bacteriol* 190(4):1390-1400.
- [00313] 39. Horvath P, *et al.* (2008) Diversity, activity, and evolution of CRISPR loci in *Streptococcus thermophilus*. *J Bacteriol* 190(4):1401-1412.
- [00314] 40. Mojica FJ, Diez-Villasenor C, Garcia-Martinez J, & Almendros C (2009) Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* 155(Pt 3):733-740.
- [00315] 41. Kameda T, Smuga-Otto K, & Thomson JA (2006) A severe de novo methylation of episomal vectors by human ES cells. *Biochem Biophys Res Commun* 349(4):1269-1277.

- [00316] 42. Ren C, *et al.* (2006) Establishment and applications of epstein-barr virus-based episomal vectors in human embryonic stem cells. *Stem Cells* 24(5):1338-1347.
- [00317] 43. Chen G, *et al.* (2011) Chemically defined conditions for human iPSC derivation and culture. *Nat Methods* 8(5):424-429.
- [00318] 44. Fu Y, *et al.* (2013) High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat Biotechnol.*

[00319] Example 3

[00320] Reference is made to Figures 18-22 and the Brief Descriptions provided above. In Figure 18, NmCas9 D16A is shown to function as a nickase in human pluripotent stem cells. Human ES cells expressing both TdTomato and EGFP were transfected with plasmids encoding SpCas9 D10A, NmCas9 D16A or both, along with their respective single guide RNAs (sgRNAs) that target TdTomato coding sequence. Sequences targeted by the sgRNAs are as indicated in Figure 18. Cellular TdTomato fluorescence was analyzed by flow cytometer 5 days after transfection. Significant amount (0.94%) of TdTomato negative cells (indicated by arrow in Figure 18) becomes detectable only in the presence of both SpCas9 D10A and NmCas9 D16A. This suggests that NmCas9 D16A is a functional nickase in human ES cells, and when paired with another properly positioned SpCas9 D10A nickase, could induce DSBs (result in 31nt 5' overhangs) and NHEJ.

[00321] Figure 19 demonstrates NmCas9 mediated homology-directed repair using either sense or antisense ssODN. A Human ES cell line that expresses both EGFP and a single mutated (deficient, 9nt deleted) copy of Tomato fluorescent protein gene was transfected with plasmids expressing NmCas9 and gRNA targeting the mutated region of tomato gene, with or without 100pmole of 160nt ssODN as repair template. The ssODNs corresponded to wild type tomato sequence, and were centered around the mutated region. Cellular fluorescence was analyzed by flow cytometry 5 days after transfection. Tomato

fluorescence positive cells become detectable (0.3%-0.5%) only when sense or antisense ssODN was co-transfected with NmCas9 and gRNA.

[00322] Figure 20 illustrates that NmCas9 can be delivered in the form of mRNA instead of plasmid DNA. Human ES cells expressing both TdTomato and EGFP were transfected with NmCas9 mRNA or an NmCas9 expressing plasmid, together with a plasmid expressing sgRNA that targets tomato gene. Cellular fluorescence was analyzed by flow cytometry 5 days after transfection. The results in Figure 20 shown that the Tomato fluorescence negative cell population, which indicates cleavage of the TdTomato gene, arises when NmCas9 is transfected in the forms of either plasmid (10.8%) or mRNA (23.4%). The Nm Cas9 mRNA used in this study is 5' capped and has a 3' polyA tail. Roughly 30ug of NmCas9 mRNA was transfected.

[00323] In the foregoing description, it will be readily apparent to one skilled in the art that varying substitutions and modifications may be made to the invention disclosed herein without departing from the scope and spirit of the invention. The invention illustratively described herein suitably may be practiced in the absence of any element or elements, limitation or limitations which is not specifically disclosed herein. The terms and expressions which have been employed are used as terms of description and not of limitation, and there is no intention that in the use of such terms and expressions of excluding any equivalents of the features shown and described or portions thereof, but it is recognized that various modifications are possible within the scope of the invention. Thus, it should be understood that although the present invention has been illustrated by specific embodiments and optional features, modification and/or variation of the concepts herein disclosed may be resorted to by those skilled in the art, and that such modifications and variations are considered to be within the scope of this invention. Citations to a number of patent and non-patent references are made herein. The cited references are incorporated by reference herein in their entireties. In the event that there is an inconsistency between a definition of a term in the specification as compared to a definition of the term in a cited reference, the term should be interpreted based on the definition in the specification.

Table 1. Bacterial species with predicted Type II-C CRISPR/Cas systems harboring repeats with extended -10 boxes

STRAIN CRISPR_id	START	END	# SPACER	DR consensus	-10box [tgnTAAAAAT]	DR length
<i>Helicobacter mustelae</i> 12198 ^a						
NC_013949_1	24706	25400	10	GTTTATAGCCACTTCATAAAATATGTTTATGCTAAAAAT		36
<i>Campylobacter jejuni</i> subsp- <i>jejuni</i> NCTC 11168 ^{a,c}						
NC_002163_2	1455125	1455424	04	GTTTATAGTCCCTTTTAAATTTCTTTTATGGTAAAAAT		36
<i>Neisseria meningitidis</i> Z2491 ^{a,c}						
NC_003116_10	608413	609504	16	GTTGTAGCTCCCTTTTCTCATTTTCGCAGTGTCTACAAT		36
<i>lyobacter polytropus</i> DSM 2925 ^a						
NC_014633_2	744582	745934	20	GTTGTACTTCCTTAATTTATTTAGCTATGTTACAAT		36
<i>Pasteurella multocida</i> subsp- <i>multocida</i> str- Pm70 ^a						
NC_002663_3	1322127	1322492	05	GTTGTAGTTCCTCTCTCATTTTCGCAGTGTCTACAAT		36
<i>Clostridium cellulolyticum</i> H10 ^a						
NC_011898_3	3652357	3652923	08	GTTATAGCTCCAATTCAGGCTCCCGATATGCTATAAT		36
<i>Rhodopseudomonas palustris</i> BisB18 ^a						
NC_007925_4	4995446	4996735	19	GCCGTGGCTCCCTACCGATTTCCCCCGTGGTAGGGCT		36
<i>Azospirillum</i> sp- B510 ^a						
NC_013854_7	3034463	3034649	02	GCTTCAATGAGGCCCAAGCAATTTCTGCCTGGGAAGAC		36
NC_013854_8	3035919	3038413	33	GCTTCAATGAGGCCCAAGCAATTTCTGCCTGGGAAGAC		36
<i>Candidatus Puncicepirillum marinum</i> IMCC1322 ^a						
NC_014010_1	40	1792	26	GTTGCTCTAGGGCTCTCAATCACCAGAGTGTCTATACT		36
<i>Parvibaculum lavamentivorans</i> DS-1 ^b						
tmp_1_Crispr_1	101247	104452	48	GCTGGGATTGGGGCGTCTCTCGATTGTGCTACTCT		36
^a Type II – C CRISPR/cas prediction						
^b Species with CRISPRdb entry (http://crispr.u-psud.fr/)						
^c Species without CRISPRdb entry but available genome sequences to predict CRISPR repeats						
^d experimentally confirmed extended -10 box						

Table 2. List of identical spacers shared among different *Neisseria* strains

Strain	Spacer	=	Strain	Spacer	=	Strain	Spacer
8013	Sp1		M01-240355	Sp1		Alpha 14	Sp1
	Sp3		M01-240355	Sp2			
	Sp4		M01-240355	Sp3			
	Sp5		M01-240355	Sp11			
	Sp10		8013	Sp11			
	Sp11		8013	Sp10			
	Sp12		8013	Sp13			
	Sp13		8013	Sp12			
Z2491	Sp1		WUE2594	Sp1			
	Sp2		WUE2594	Sp2		053442	Sp5
	Sp3		WUE2594	Sp3			
	Sp4		WUE2594	Sp4			
	Sp5		WUE2594	Sp5			
	Sp8		Z2491	Sp9			
	Sp9		Z2491	Sp8			
	Sp11		WUE2594	Sp6			
	Sp13		WUE2594	Sp8			
	Sp15		WUE2594	Sp10			
WUE2594	Sp16		WUE2594	Sp11			
	Sp1		Z2491	Sp1			
	Sp2		Z2491	Sp2		053442	Sp5
	Sp3		Z2491	Sp3			
	Sp4		Z2491	Sp4			
	Sp5		Z2491	Sp5			
	Sp6		Z2491	Sp11			
	Sp8		Z2491	Sp13			
	Sp10		Z2491	Sp15			
	Sp11		Z2491	Sp16			
M01-240355	Sp1		8013	Sp1		Alpha14	Sp1
	Sp2		8013	Sp3			
	Sp3		8013	Sp4			
	Sp5		053442	Sp3			
	Sp8		M01-240355	Sp13			
	Sp11		8013	Sp5			
	Sp13		M01-240355	Sp8			
053442	Sp3		M01-240355	Sp5			
	Sp5		WUE2594	Sp2		Z2491	Sp2
Alpha14	Sp1		M01-240355	Sp1		8013	Sp1
	Sp4		Alpha14	Sp5			
	Sp5		Alpha14	Sp4			

Table 3. List of *Neisseria* repeat variants

Strain name	Consensus Repeat Sequence	Repeat number	Variant Repeat Sequences*
<i>N.m.</i> 8013		R1	GTTGTAGCTCCCTTTCTCATTTTCGCAGTGCTACAAT
<i>N.m.</i> Z2491		R1	GTTGTAGCGCCCC A TTCTCATTTTCGCAGTGCTACAAT
<i>N.m.</i> WUE2594		R1	GTTGTAGCGCCCC A TTCTCATTTTCGCAGTGCTACAAT
<i>N.m.</i> M01-240335		R1 R15	GTTGTAGCGCCCC A TTCTCATTTTCGCAGTGCTACAAT GTTGTAGCTCCCTTTCTCATTTTCGCAGT A CTACAAT
<i>N.m.</i> Alpha14		R1 R4	GTTGTAGCGCCCC A TTCTCATTTTCGCAGTGCTACAAT GTTGTAGCTCCCTTTCTCAT AA T
<i>N.m.</i> 053442		R1 R2 R8, 9, 10	GTTGTAGCGCCCC A TTCTCATTTTCGCAGTGCTACAAT GTTGTAGCTCCCG A TTCTCATTTTCGCAGTGCTACAAT GTTGTAGCTCCCTTTCTC C TTTCGCAGTGCTACAAT
<i>N.I.</i> 020-06		None	
* Nucleotides differing from consensus repeat sequences are underlined and in bold.			
<i>N.m.</i> , <i>Neisseria meningitidis</i> ; <i>N.I.</i> , <i>Neisseria lactamica</i> .			

Table 4. Prophage-related potential natural targets for *Neisseria* CRISPRs

Phage-like sequences	References	Target locus_tag	Annotations (NCBI & Neisseria)	Targeted by <i>Neisseria</i> Strain name _spacer number
Meningococcal disease-associated (MCA) island in <i>N. meningitidis</i> Z2481 (NMA1792-1800), an <i>stkA</i> filamentous prophage associated with hyperinvasive isolates	(Eise et al., 2005)	NMA1792	putative phage replication initiation factor	WUE2564_sp21
		NMA1799	putative zonular occludens toxin-like protein	WUE2564_sp12
		NMA1800	putative pilin gene-inverting protein/transposase	WUE2564_sp18
NF (Nisseria filamentous) prophages in <i>N. meningitidis</i> MC58 (NMB1543-1552) (NMB1625-1635)	(Kawai et al., 2005; Joseph et al., 2011)	NMB1543	putative phage replication initiation factor	WUE2564_sp21
		NMB1544	conserved hypothetical protein	WUE2564_sp19*
		NMB1552	putative pilin gene-inverting protein (PhML)	WUE2564_sp13 WUE2564_sp18
		NMB1626	putative pilin gene-inverting protein (PhML)	WUE2564_sp13 WUE2564_sp18
		NMB1628	Putative TspB protein	8013_sp8*
		NMB1633	conserved hypothetical protein	WUE2564_sp19
NF2 prophages in <i>N. meningitidis</i> MC58 (NMB1742-1750)	(Kawai et al., 2005; Joseph et al., 2011)	NMB1749	putative zonular occludens toxin-like protein	WUE2564_sp7 053442_sp8 053442_sp8 8013_sp17
		NMB1750	putative pilin gene-inverting protein (PhML)	053442_sp7 WUE2564_sp17
		Intergenic	Between NMB1749 and NMB1750	Z2491_sp3 = WUE2564_sp3
ISNgds and nearby degenerate NF prophages in <i>N. gonorrhoeae</i> FA1090 (N84-G3: NGO1137-NGO1147 N84-G3: NGO1164-NGO1170 N84-G6: NGO1262-NGO1270 N84-G6: NGO1641-NGO1643)	(Kawai et al., 2005; Sklar et al., 2005)	NGO1137	invertase related gene 2 (irg2), putative phage associated protein	Z2491_sp2** = WUE2564_sp2* =053442_sp5* MO1-240355_sp16
		NGO1138	putative zonular occludens toxin-like, phage associated protein	053442_sp8*
		Intergenic	Between NGO1137 and NGO1138	8013_sp8
		NGO1164	invertase related gene 3 (irg3), putative phage associated protein	Z2491_sp2* = WUE2564_sp2* =053442_sp8* MO1-240355_sp16
		NGO1165	putative zonular occludens toxin-like, phage associated protein	053442_sp8*
		Intergenic	Between NGO1164 and NGO1165	8013_sp8
		NGO1262	invertase related gene 5 (irg5), putative phage associated protein	Z2491_sp2* = WUE2564_sp2* =053442_sp5* MO1-240355_sp16
		NGO1263	putative zonular occludens toxin-like, phage associated protein	053442_sp8*
		Intergenic	Between NGO1262 and NGO1263	8013_sp8
		NGO1641	invertase related gene 6 (irg6), putative phage associated protein	Z2491_sp2* = WUE2564_sp2* =053442_sp5* MO1-240355_sp16
		NGO1643	putative zonular occludens toxin-like, phage associated protein	053442_sp8*
		Intergenic	Between NGO1641 and NGO1643	8013_sp8
ISNgds/ISAgds with partially deleted NF prophages in <i>N. gonorrhoeae</i> FA1090 (N84-G1: NGO0773) (N84-G3: NGO1260) (N84-G6: NGO1703)	(Kawai et al., 2005; Sklar et al., 2005)	NGO0773	putative invertase related gene 1 (irg1)	MO1-240355_sp16
		NGO1260	putative invertase related gene 4 (irg4)	MO1-240355_sp18
		NGO1703	putative invertase related gene 8 (irg8)	MO1-240355_sp15

M1 prophages in <i>N. meningitidis</i> FAM18 (NP1-C1: <i>NMC0022-0031</i>) (NP1-C2: <i>NMC0277-0288</i>) (NP1-C3: <i>NMC1709-1718</i>) (NP1-C4: <i>NMC1861-1869</i>)	(Kawai et al., 2005)	<i>NMC0022</i>	putative transposase, pilin gene-inverting protein (PivML)	WUE2594_sp18
		<i>NMC0023</i>	putative zonular occludens toxin-like protein	WUE2594_sp12
		<i>NMC0030</i>	conserved hypothetical protein	WUE2594_sp19
		<i>NMC0278</i>	conserved hypothetical protein	WUE2594_sp19
		<i>NMC0285</i>	putative zonular occludens toxin-like protein	WUE2594_sp12
		<i>NMC0286</i>	putative invertase / transposase, putative pilin gene-inverting protein (PivML)	WUE2594_sp18
		<i>NMC1710</i>	conserved hypothetical protein	WUE2594_sp19
		<i>NMC1714</i>	conserved hypothetical protein	WUE2594_sp20
		<i>NMC1717</i>	putative zonular occludens toxin-like protein	WUE2594_sp12
		<i>NMC1718</i>	putative transposase, putative pilin gene-inverting protein (PivML)	WUE2594_sp18
		<i>NMC1862</i>	conserved hypothetical protein	WUE2594_sp19
		<i>NMC1864</i>	hypothetical integral membrane protein	Z2491_sp18 = WUE2594_sp10
		<i>NMC1868</i>	putative zonular occludens toxin-like protein	WUE2594_sp12
Meningococcal disease associated (MDA) island-like in <i>N. meningitidis</i> 053442 (<i>NMCC_0148</i> to <i>NMCC_0156</i>)	(Peng et al., 2008)	<i>NMCC_0148</i>	putative invertase/transposase	WUE2594_sp13 WUE2594_sp18
		<i>NMCC_0149</i>	conserved hypothetical protein	WUE2594_sp12
		<i>NMCC_0153</i>	conserved hypothetical protein	Z2491_sp15 = WUE2594_sp10
		<i>NMCC_0156</i>	conserved hypothetical protein	WUE2594_sp21
Mu-like prophages in <i>N. meningitidis</i> MC58 MuMenB/NeisMu1: (<i>NMB1078-1121</i> , 35kb) NeisMu2: (<i>NMB0995-0991</i> , 4kb) NMB1: (<i>NMB1002-1007</i> , 2kb)	(Braid et al., 2004)(Masignani et al., 2001)	None	N/A	None
Mu-like prophages in <i>N. meningitidis</i> Z2491 Pnm1 (<i>NMA1821-NMA1884</i> , 35kb) Pnm2 (<i>NMA1281-NMA1330</i> , 29kb) Pnm3 (<i>NMA1185-NMA1192</i> , 7kb) Pnm4/5 (<i>NMA1208-NMA1231</i> , 12kb)	(Masignani et al., 2001, Braid et al., 2004)	None	N/A	None
Only perfect matches throughout the 30-nt spacers, or 1 mismatch within the 3'-terminal 10 nts, were considered "potential targets" and listed in this table. Protospacers with a single mismatch are denoted with an asterisk (*).				

Table 5. Transformation frequencies reported in this study

Donor DNA	Recipient Strain (<i>N. meningitidis</i> 8013)	Antibiotic Selection ^(a)	Transformation Frequencies	
			Mean ^(b)	SEM ^(b)
Related to Figure 4B				
pGCC2 empty	wt	Erythromycin	3.9×10^{-6}	1.1×10^{-7}
pGCC2-protospacer 1	wt	Erythromycin	2.5×10^{-6}	8.9×10^{-8}
pGCC2-protospacer 8	wt	Erythromycin	0	0
pGCC2-protospacer 9	wt	Erythromycin	0	0
pGCC2-protospacer 16	wt	Erythromycin	2.2×10^{-6}	2.5×10^{-7}
pGCC2-protospacer 17	wt	Erythromycin	0	0
pGCC2-protospacer 18	wt	Erythromycin	2.5×10^{-6}	5.1×10^{-7}
pGCC2-protospacer 23	wt	Erythromycin	3.7×10^{-6}	1.6×10^{-6}
pGCC2-protospacer 25	wt	Erythromycin	0	0
No DNA	wt	Erythromycin	0	0
Related to Figure 4C				
pGCC2 empty	wt	Erythromycin	3.8×10^{-6}	1.5×10^{-7}
pGCC2-Ps9-wildtype	wt	Erythromycin	0	0
pGCC2-Ps9Mutant1	wt	Erythromycin	0	0
pGCC2-Ps9Mutant2	wt	Erythromycin	0	0
pGCC2-Ps9Mutant3	wt	Erythromycin	1.6×10^{-6}	5.3×10^{-7}
pGCC2-Ps9Mutant4	wt	Erythromycin	0	0
pGCC2-Ps9Mutant5	wt	Erythromycin	1.2×10^{-6}	2.4×10^{-7}
pGCC2-Ps9Mutant6	wt	Erythromycin	0	0
pGCC2-Ps9Mutant7	wt	Erythromycin	0	0
pGCC2-Ps9Mutant8	wt	Erythromycin	0	0
No DNA	wt	Erythromycin	0	0
Related to Figure 5B				
PYZEJS040	wt	Chloramphenicol	6.0×10^{-6}	1.6×10^{-6}
PYZEJS040-protospacer25	wt	Chloramphenicol	0	0
PYZEJS040	<i>cas9::Tn</i>	Chloramphenicol	3.8×10^{-6}	1.9×10^{-6}
PYZEJS040-protospacer25	<i>cas9::Tn</i>	Chloramphenicol	2.9×10^{-6}	4.4×10^{-6}
PYZEJS040	<i>cas9::Tn + cas9 wt</i>	Chloramphenicol	1.1×10^{-6}	4.9×10^{-6}
PYZEJS040-protospacer25	<i>cas9::Tn + cas9 wt</i>	Chloramphenicol	0	0
PYZEJS040	<i>cas9::Tn + cas9 D16A</i>	Chloramphenicol	5.2×10^{-6}	1.4×10^{-6}
PYZEJS040-protospacer25	<i>cas9::Tn + cas9 D16A</i>	Chloramphenicol	2.5×10^{-6}	7.7×10^{-6}
PYZEJS040	<i>cas9::Tn + cas9 H588A</i>	Chloramphenicol	2.0×10^{-6}	2.9×10^{-6}
PYZEJS040-protospacer25	<i>cas9::Tn + cas9 H588A</i>	Chloramphenicol	4.0×10^{-6}	1.7×10^{-6}
PYZEJS040	<i>cas9::Tn + empty</i>	Chloramphenicol	6.9×10^{-6}	2.5×10^{-6}
PYZEJS040-protospacer25	<i>cas9::Tn + empty</i>	Chloramphenicol	6.2×10^{-6}	1.2×10^{-6}
PYZEJS040	$\Delta cas9$	Chloramphenicol	9.5×10^{-6}	1.6×10^{-6}
PYZEJS040-protospacer25	$\Delta cas9$	Chloramphenicol	7.2×10^{-6}	2.6×10^{-6}
PYZEJS040	$\Delta cas9 + cas9 wt$	Chloramphenicol	1.9×10^{-6}	6.0×10^{-6}
PYZEJS040-protospacer25	$\Delta cas9 + cas9 wt$	Chloramphenicol	0	0
PYZEJS040	$\Delta cas9 + empty$	Chloramphenicol	9.0×10^{-6}	5.7×10^{-6}
PYZEJS040-protospacer25	$\Delta cas9 + empty$	Chloramphenicol	8.9×10^{-6}	3.1×10^{-6}
PYZEJS040	<i>cas1::Tn</i>	Chloramphenicol	3.8×10^{-6}	1.3×10^{-6}
PYZEJS040-protospacer25	<i>cas1::Tn</i>	Chloramphenicol	0	0
PYZEJS040	<i>cas2::Tn</i>	Chloramphenicol	5.0×10^{-6}	1.7×10^{-6}
PYZEJS040-protospacer25	<i>cas2::Tn</i>	Chloramphenicol	0	0
PYZEJS040	<i>1851::Tn</i>	Chloramphenicol	1.3×10^{-6}	7.9×10^{-6}
PYZEJS040-protospacer25	<i>1851::Tn</i>	Chloramphenicol	0	0
Related to Figure 5C				
PYZEJS040	wt	Chloramphenicol	2.7×10^{-6}	1.0×10^{-6}
PYZEJS040-protospacer25	wt	Chloramphenicol	0	0
PYZEJS040	<i>mcr::Tn^(c)</i>	Chloramphenicol	1.9×10^{-6}	8.2×10^{-6}
PYZEJS040-protospacer25	<i>mcr::Tn^(c)</i>	Chloramphenicol	0	0
PYZEJS040	Δmcr	Chloramphenicol	6.0×10^{-6}	2.3×10^{-6}
PYZEJS040-protospacer25	Δmcr	Chloramphenicol	0	0
PYZEJS040	$\Delta tracr$	Chloramphenicol	4.7×10^{-6}	2.4×10^{-6}
PYZEJS040-protospacer25	$\Delta tracr$	Chloramphenicol	1.4×10^{-6}	4.5×10^{-6}
PYZEJS040	$\Delta tracr + tracr$	Chloramphenicol	4.3×10^{-6}	1.8×10^{-6}
PYZEJS040-protospacer25	$\Delta tracr + tracr$	Chloramphenicol	0	0
PYZEJS040	$\Delta tracr + empty$	Chloramphenicol	3.7×10^{-6}	2.4×10^{-6}

PyZEJS040-protospacer25	Δ tracr + empty	Chloramphenicol	6.7×10^{-6}	1.3×10^{-5}
Related to Figure 6B				
gDNA of <i>cas2::Tn</i>	wt	Erythromycin	0	0
gDNA of <i>cas9::Tn/pGCC2</i>	wt	Erythromycin	1.6×10^{-5}	6.1×10^{-5}
gDNA of <i>cas9::Tn/pGCC2-Ps25</i>	wt	Erythromycin	0	0
Related to Figure 6B				
gDNA of <i>cas2::Tn</i>	wt	Chloramphenicol	0	0
gDNA of <i>cas9::Tn/pYZEJS040</i>	wt	Chloramphenicol	5.1×10^{-6}	2.6×10^{-5}
gDNA of <i>cas9::Tn/pYZEJS040-Ps25</i>	wt	Chloramphenicol	0	0
Related to Figure S4				
pYZEJS040	wt	Chloramphenicol	6.0×10^{-6}	1.6×10^{-5}
pYZEJS040-protospacer 9	wt	Chloramphenicol	0	0
pYZEJS040-protospacer 25	wt	Chloramphenicol	0	0
Related to interference assays testing internal protospacer 9 in <i>mc::Tn</i> and Δmc strains.				
pYZEJS040	wt	Chloramphenicol	7.3×10^{-6}	1.8×10^{-5}
pYZEJS040-protospacer 9	wt	Chloramphenicol	0	0
pYZEJS040	<i>mc::Tn</i>	Chloramphenicol	2.2×10^{-5}	2.5×10^{-5}
pYZEJS040-protospacer 9	<i>mc::Tn</i>	Chloramphenicol	0	0
pYZEJS040	Δ mc	Chloramphenicol	6.9×10^{-6}	3.7×10^{-5}
pYZEJS040-protospacer 9	Δ mc	Chloramphenicol	0	0
^(a) Indicates antibiotic used to select transformants.				
^(b) The average and standard error of the mean (s.e.m.) of transformation frequencies (ratio comparing transformants cfu/ml vs. total cfu/ml) from at least three independent experiments.				
^(c) The <i>mc::Tn</i> mutant of <i>N. meningitidis</i> 8013 exhibited obvious slow-growth defects.				

Table 6. Characteristics of the seven CRISPR/Cas-containing *Neisseria* strains

Strain name	Serogroup ^(a)	Country/Year ^(a)	ST ^(a)	Clonal Complex ^(a)	CRISPR_id ^(b)
<i>N. meningitidis</i>					
Z2491	A	Gambia /1983	4	ST-4 complex	NC_003116_10
8013	C	France /1989	177	ST-18 complex	NC_017501_11
WUE2594	A	Germany /1991	5	ST-5 complex	NC_017512_9
M01-240355	B	UK /2001	213	ST-213 complex	NC_017517_9
053442	C	China/2004	4821	ST-4821 complex	NC_010120_5
alpha14	NG	Germany/1999	53	ST-53 complex	NC_013016_1 to NC_013016_2 ^(c)
<i>N. lactamica</i>					
020-06	N/A	UK / 1997	640	ST-640 complex	NC_014752_11
^(a) according to <i>Neisseria</i> PubMLST database (Jolley and Maiden, 2010)					
^(b) according to CRISPRdb					
^(c) NC_013016_1 and NC_013016_2 each constitute part of our predicted CRISPR in <i>N. meningitidis</i> alpha14.					

Table 7. Summary of gene targeting efficiency using NmCas9 in hPSCs

Cell line	crRNA	Clone analyzed	Nontargeted	Targeted with additional insertions	Targeted	Targeting efficiency (%)
H1 (ES)	Nontargeting	0	0	0	0	0
	Targeting	20	5	3	12	60
H9 (ES)	Nontargeting	1	1	0	0	0
	Targeting	39	9	7	23	59
iPS005 (iPS)	Nontargeting	1	1	0	0	0
	Targeting	10	1	3	6	60

Table 8. Plasmids used in this study

Plasmid no.	Plasmid name	Description
1	pSimpleII-NmCas9-FLAG	Flag tagged NmCas9 without NLS
2	pSimpleII-NLS-NmCas9-EGFP	NmCas9 EGFP fusion with N-terminal NLS
3	pSimpleII-NmCas9-HA-NLS	NmCas9 with C-terminal HA tag and NLS
4	pSimpleII-NLS-NmCas9-HA-NLS(s)	NmCas9 with dual NLS and HA tags
5	pSimpleII-U6-tracrRNA-U6-BsmBI-NLS-NmCas9-HA-NLS(s)	All-in-one plasmid containing NmCas9, tracrRNA expression cassette and U6-BsmBI cassette
6	pSimpleII-U6-tracrRNA-U6-crRNA(tdTomato)-NLS-NmCas9-HA-NLS(s)	All-in-one plasmid containing NmCas9, tracrRNA expression cassette and tdTomato-targeting crRNA expression cassette
7	pSimpleII-U6-tracrRNA-U6-crRNA(EGFP)-NLS-NmCas9-HA-NLS(s)	All-in-one plasmid containing NmCas9, tracrRNA expression cassette and EGFP-targeting crRNA expression cassette
8	pSimpleII-U6-tracrRNA-U6-crRNA(OCT4)-NLS-NmCas9-HA-NLS(s)	All-in-one plasmid containing NmCas9, tracrRNA expression cassette and OCT4-targeting crRNA expression cassette
9	pSTBlue-1-U6-sgRNA (tdTomato)	U6-driven sgRNA targeting tdTomato

Table 9. crRNA-encoding DNA sequences used in this study

crRNA	Encoding DNA sequence
EGFP targeting crRNA	<u>gttcagcgtg</u> <u>tccggcgagggcg</u> <u>gattgt</u> <u>gtagctcccttctcatttcg</u>
OCT4 targeting crRNA	<u>Gacctgag</u> <u>tttgtccagggttt</u> <u>gttgt</u> <u>agctcccttctcatttcg</u>
tdTomato targeting crRNA (GATT PAM)	<u>gtacgtga</u> <u>agcaccgccgcgacat</u> <u>gttgt</u> <u>agctcccttctcatttcg</u>
tdTomato targeting crRNA (GATG PAM)	<u>Gccccgag</u> <u>ggttcaagtggagc</u> <u>gttgt</u> <u>agctcccttctcatttcg</u>
tdTomato targeting crRNA (GACT PAM)	<u>ggacgg</u> <u>cggtctgggac</u> <u>cgtagc</u> <u>gttgt</u> <u>agctcccttctcatttcg</u>
tdTomato targeting crRNA (GCTT PAM)	<u>gattaca</u> <u>gaagcgtctcttccc</u> <u>gttgt</u> <u>agctcccttctcatttcg</u>
tdTomato targeting crRNA (CATT PAM)	<u>Gggctcc</u> <u>agcccatggtcttct</u> <u>gttgt</u> <u>agctcccttctcatttcg</u>
tdTomato targeting crRNA (CCAA PAM)	<u>ggccgccc</u> <u>tacgagggcaccag</u> <u>gttgt</u> <u>agctcccttctcatttcg</u>

All sequences are 5' to 3', left to right. Spacer regions are in lowercase and underlined, and CRISPR repeat regions are in uppercase. In some cases, the first nucleotide of the spacer is changed to a G to satisfy the requirement of the U6 promoter.

Table 10. Primers used in this study

Name	Forward	Reverse
For cleavage site mapping in tdTomato	ATGGTGAGCAAGGGCGAGGAG	CCGGTGTCTGCGGTGCGCATGCCCCCAG
For tdTomato indel mapping	AACACTGTCCCTCTCACTGCTCCCTGCTTC	CCGGTGTCTGCGGTGCGCATGCCCCCAG
For making a Southern blot probe for OCT4	GTGATGCCACCAAGAACCTT	ACAGCAGCGGAGCAAAATAGGT

All sequences are 5' to 3', left to right.

tracrRNA and crRNA

Sequences that are cleaved off during processing are underlined.

CLAIMS

We claim:

1. A method for modifying a target DNA sequence in a cell, the method comprising:
 - (a) expressing a Cas9 protein from a *Neisseria* species or a variant protein thereof in the cell; and
 - (b) expressing or transfecting an RNA in the cell, wherein the RNA binds to the Cas9 protein or variant, and the RNA hybridizes to the target DNA sequence.
2. The method of claim 1, wherein the Cas9 protein or variant protein has nuclease activity and cleaves both strands of the target DNA sequence.
3. The method of claim 2, further comprising contacting the target DNA sequence with a homologous DNA fragment.
4. The method of claim 1, wherein the Cas9 protein or variant protein has nuclease activity and nicks a single strand of the target DNA sequence.
5. The method of claim 4, further comprising contacting the target DNA sequence with a homologous DNA fragment.
6. The method of claim 1, wherein the Cas9 protein or variant protein has no nuclease activity and binds to the target sequence.
7. The method of claim 1, wherein the Cas9 protein or variant protein is expressed from a nucleic acid having a codon sequence that is optimized for expression in the cell.
8. The method of claim 1, wherein the variant has an amino acid sequence that is at least 80% identical to a Cas9 protein from a *Neisseria* species.
9. The method of claim 1, wherein the Cas9 protein is from *Neisseria meningitidis*.

10. The method of claim 1, wherein expressing a Cas9 protein from a *Neisseria* species or a variant protein thereof in the cell comprises transfecting the cell with an expression vector that expresses the Cas9 protein from a eukaryotic promoter.

11. The method of claim 1, wherein expressing a Cas9 protein from a *Neisseria* species or a variant protein thereof in the cell comprises transfecting the cell with an mRNA that encodes the Cas9 protein.

12. The method of claim 1, wherein expressing an RNA in the cell that binds to the Cas9 protein or variant and hybridizes to the target DNA sequence comprises transfecting the cell with an expression vector that expresses the RNA from a eukaryotic promoter.

13. The method of claim 1, wherein the cell is a prokaryotic cell.

14. The method of claim 1, wherein the cell is a eukaryotic cell.

15. The method of claim 1, wherein the cell is a stem cell.

16. The method of claim 1, wherein the cell is an embryonic stem cell.

17. The method of claim 1, wherein the cell is an induced pluripotent stem cell.

18. The method of claim 1, wherein the RNA comprises two molecules of duplexed RNA.

19. The method of claim 1, wherein the RNA comprises a single RNA molecule forming a hairpin structure.

20. The method of claim 1, wherein the RNA comprises an RNA mimic of green fluorescent protein (GFP).

21. The method of claim 1, further comprising contacting the target DNA sequence with 4-hydroxybenzylidene, 3,5-dimethoxy-4-hydroxybenzylidene, or a 3,5-difluoro-4-hydroxybenzylidene.

22. The method of claim 1, wherein the RNA comprises Xist RNA.
23. A recombinant Cas9 protein from a *Neisseria* species or a variant thereof comprising a nuclear localization signal.
24. A recombinant Cas9 protein from a *Neisseria* species or a variant thereof comprising a ligand or a tag for purifying or identifying the Cas9 protein.
25. A polynucleotide encoding the protein of claim 23.
26. A cell transfected with the polynucleotide of claim 25.
27. A kit for performing the method of claim 1 comprising: (a) a vector for expressing a Cas9 protein from a *Neisseria* species or a variant protein thereof in the cell; and (b) a vector for expressing an RNA in the cell, wherein the RNA binds to the Cas9 protein or variant, and the RNA hybridizes to the target DNA sequence.
28. A kit comprising the protein of claim 23.
29. A kit comprising the polynucleotide of claim 25.
30. A kit comprising the cell of claim 26.

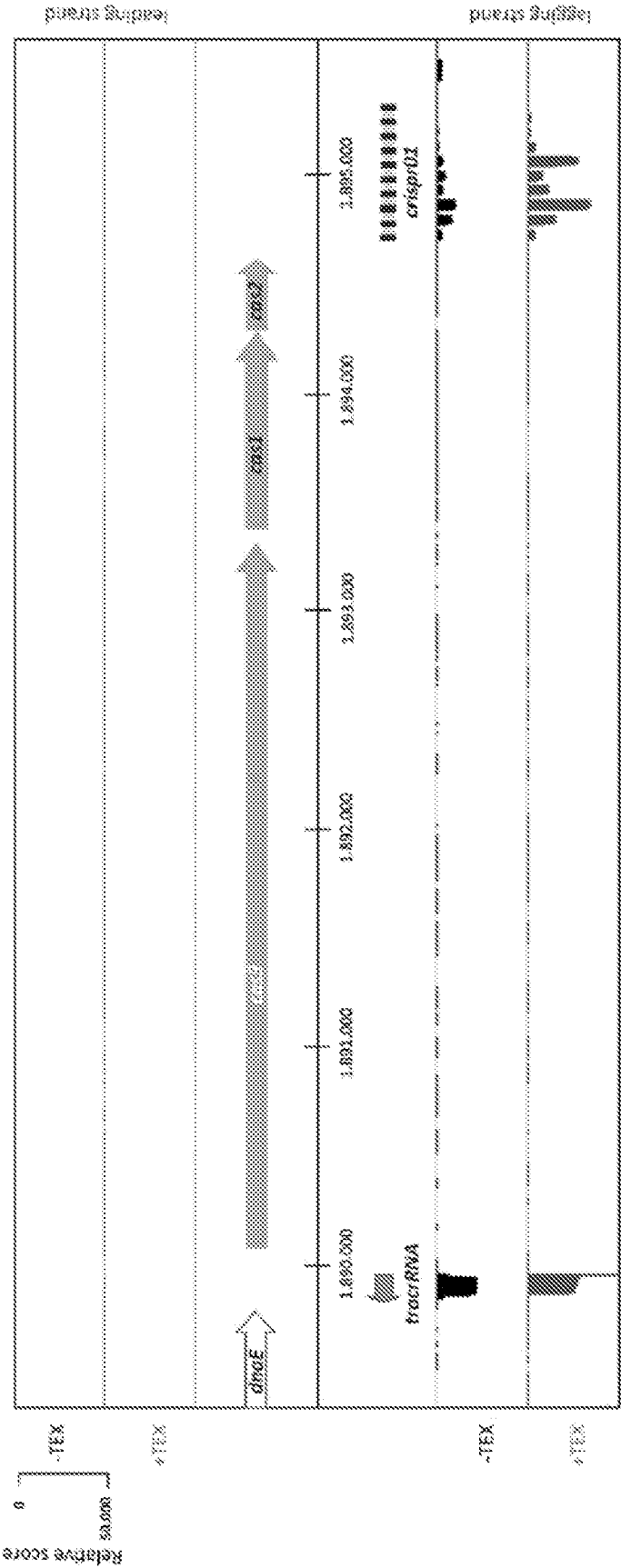
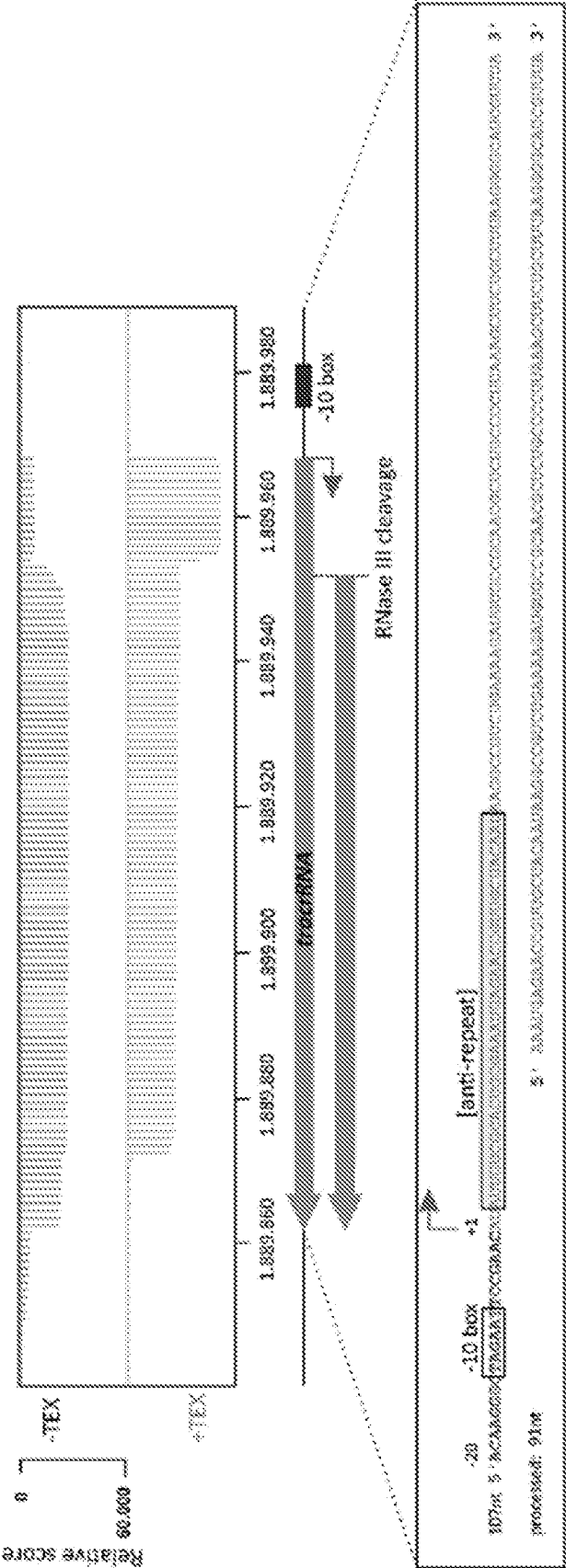


Figure 1A

Figure 1B



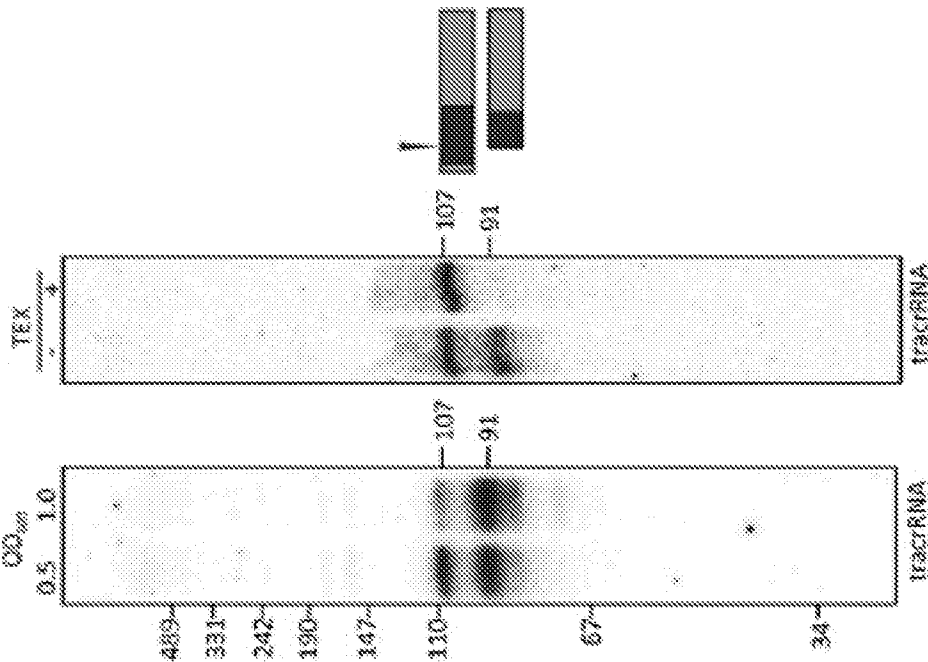
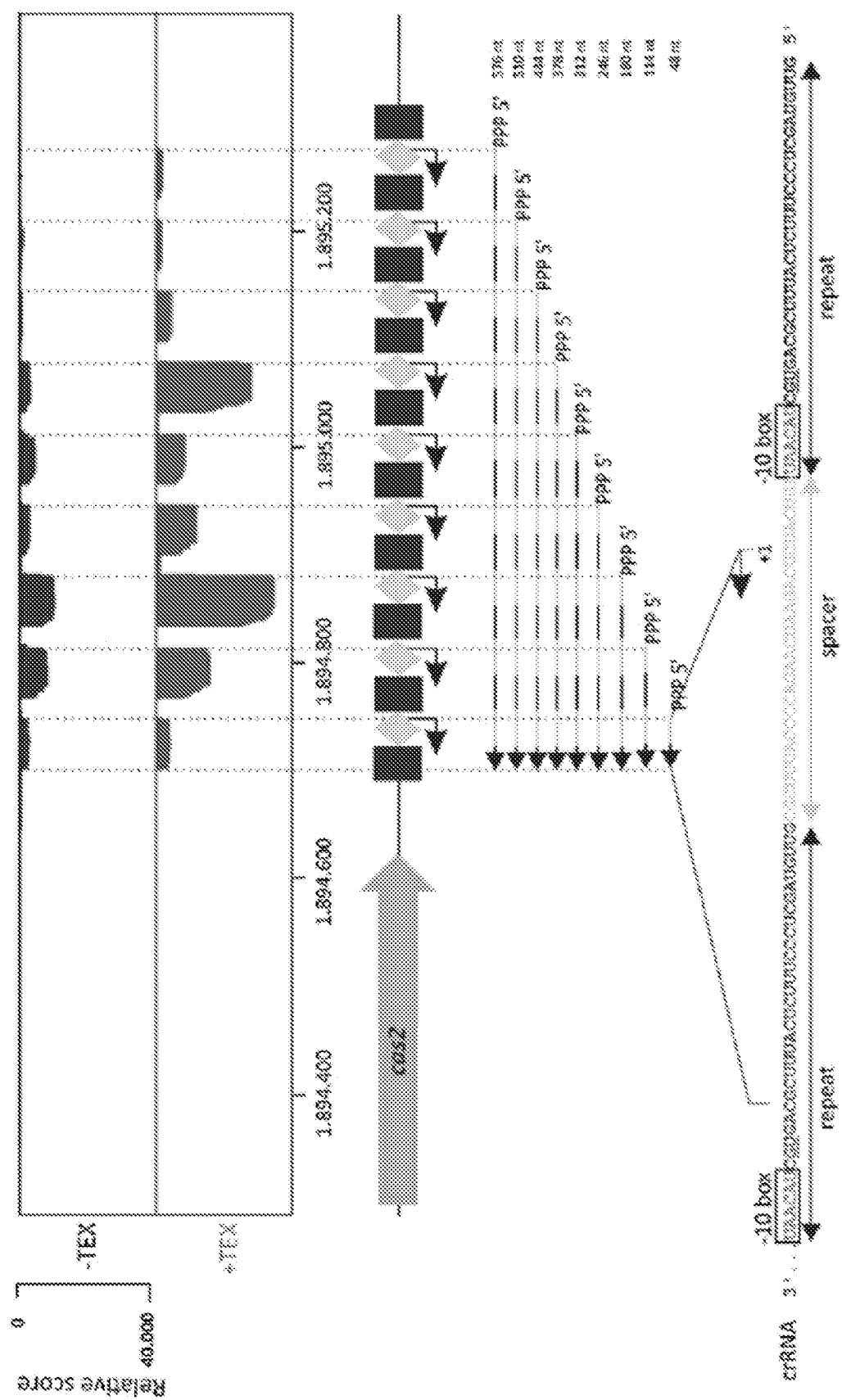


Figure 1C

Figure 1D



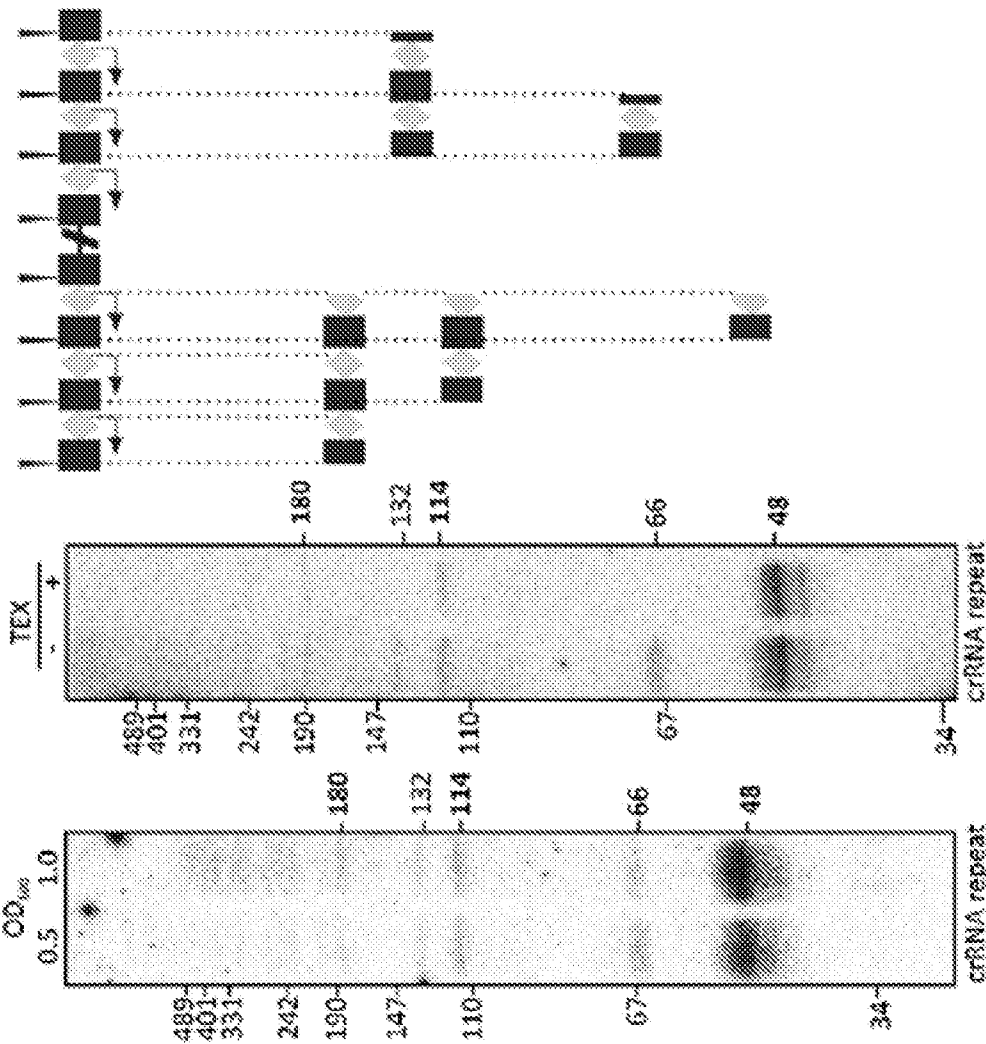
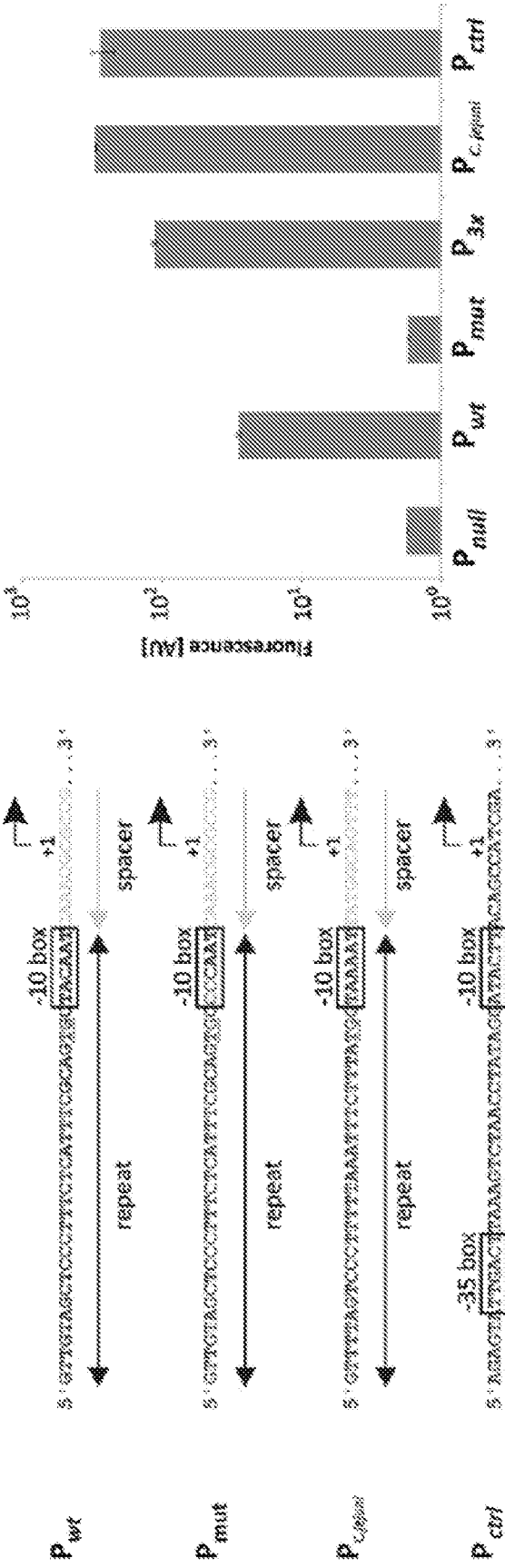


Figure 1E

Figure 2A



Fluorescence mode: Visible light mode

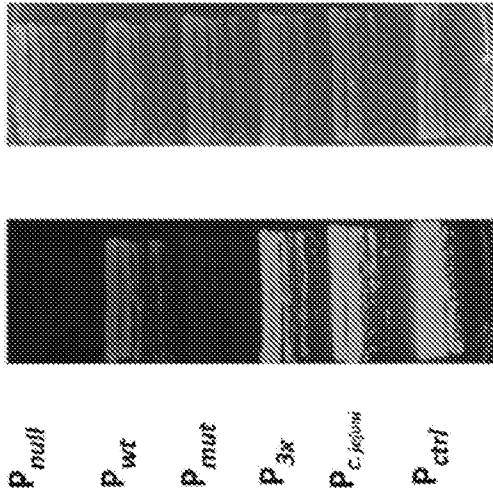


Figure 2B

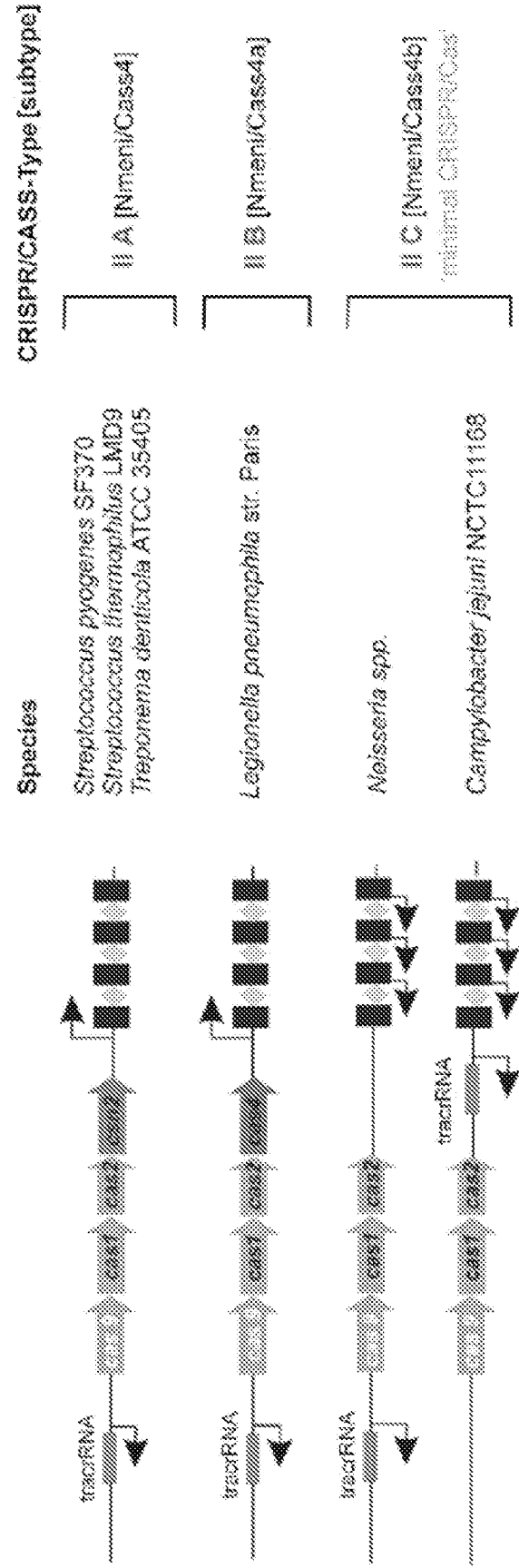


Figure 2C

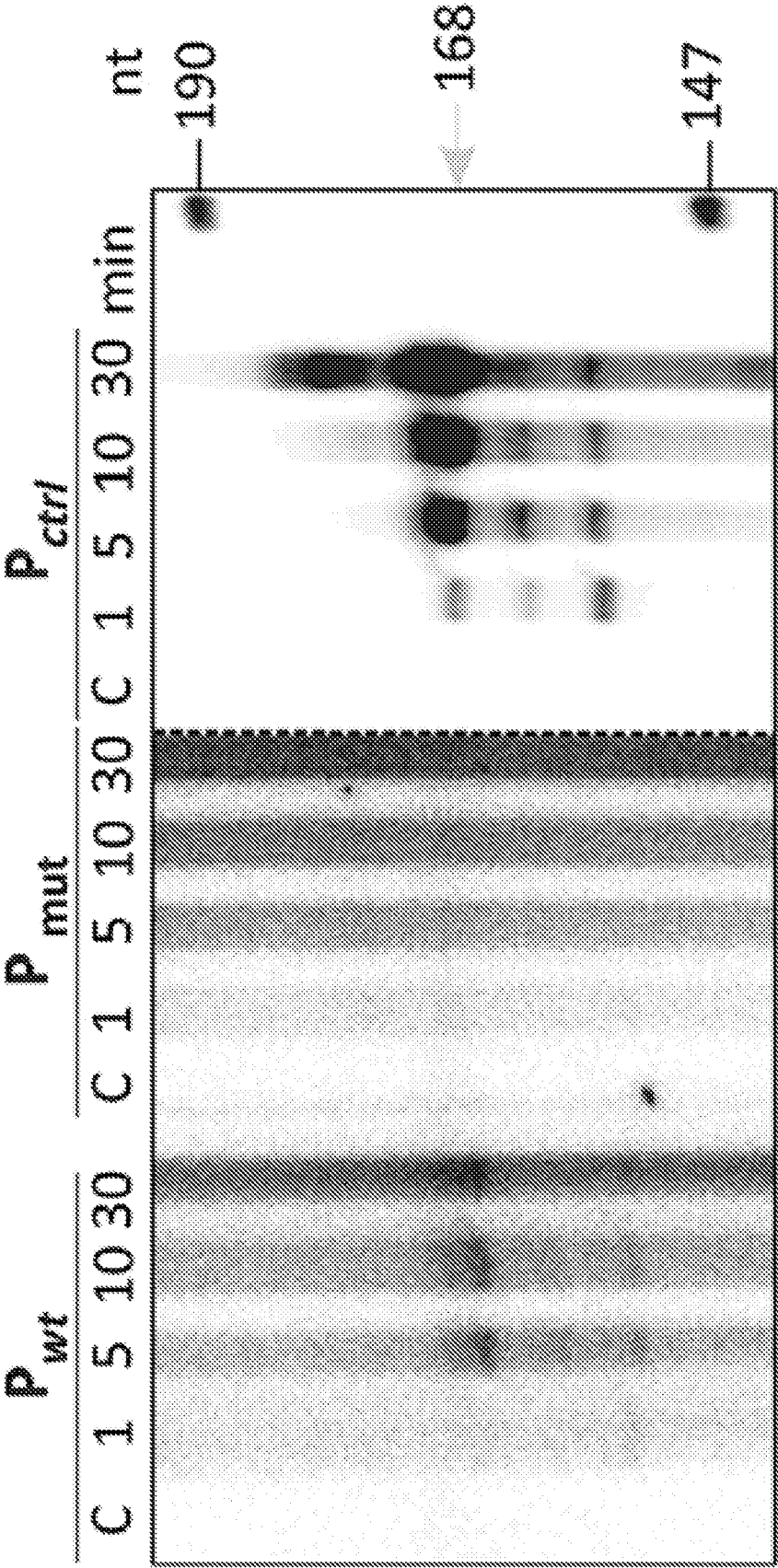


Figure 2D

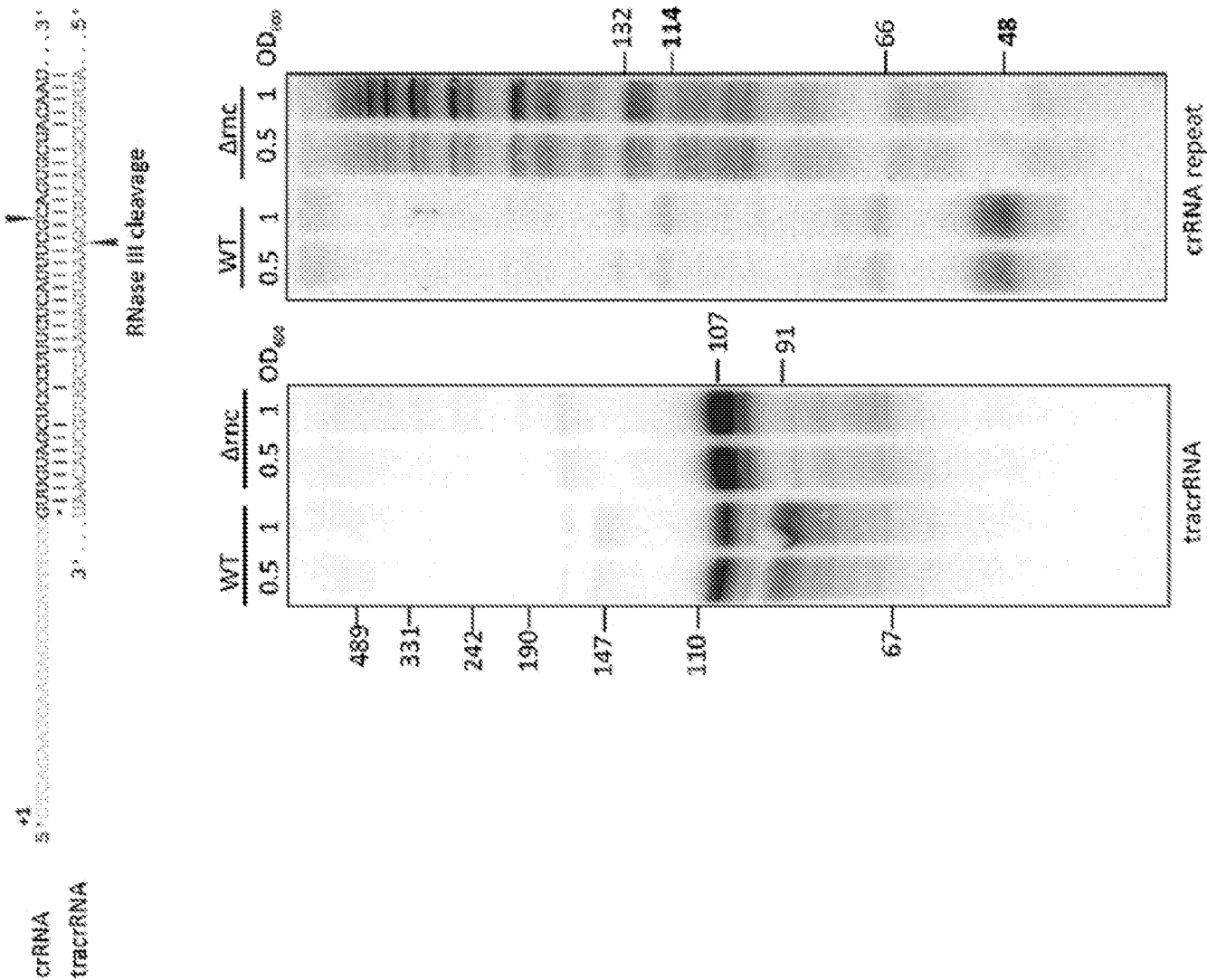


Figure 3A

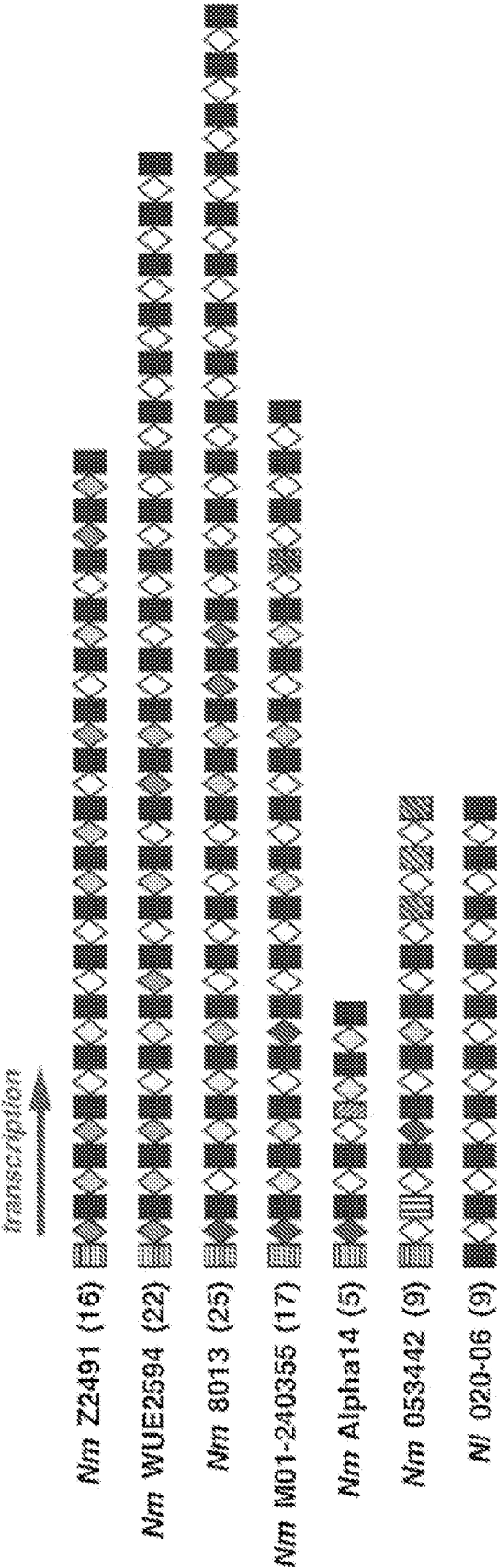
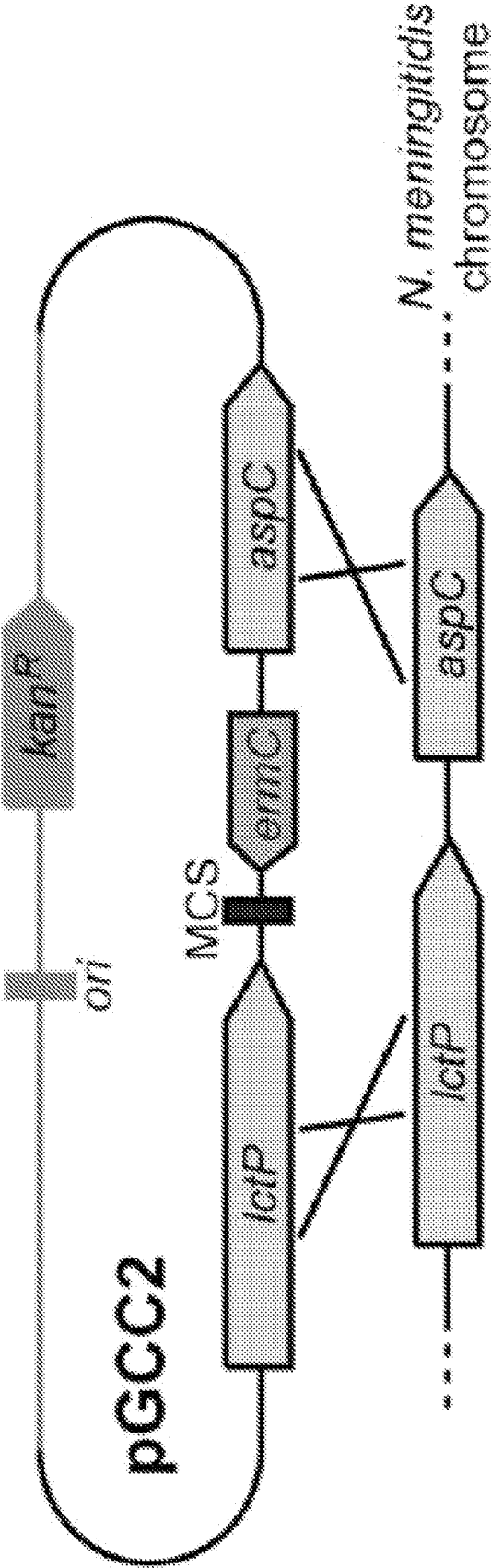


Figure 3B

					
			5' flank	Representative protospacer sequence	3' flank
8	Nm H44/76	4	GACCTGAAA	- ACGCTTACCATCAACACGAAATCGGTAAC	- TACAGATTTT
9	Ng FA1090	12	TGGTTTGTGA	- ATCATGACATCCTCAGATTAGTATTCAGA	- ATATGATTTT
14	Nm alpha710	5	GCAGAAATGCC	- CTCGCCCGGCCCTCCCCACAGGGAGGG	- AGCGGATTCG
16	Nm 8013	15	GAAATCAGCG	- ATACGGCCCTCGCACTTGTCCTCCACTCGAGCA	- AATATCTGAA
17	Nm NZ-5/33	6	GGAATGTTCC	- ACCTGTTCCGTCCTTTGCTGTTCTGTCCTTA	- TAAGGATTTGA
18	Nm 8013	18	CCCATACCGA	- CCTGAGCAATCAAAACAGCGCGGAAAATTGCT	- GTTGTCCTGCA
21	Nm WUE2594	1	GTTAAGAACG	- TAGCCTGCTATGCCGCATATAGCGATGCTT	- TCCCCTTTAA
23	Nm WUE2594	1	TTTTCAGTGT	- TCCGGCCCGTTTCATCGCGGTTCTTCCTG	- CGTTAATTTT
25	Nm alpha710	1	TACAGATTTT	- TATTCCTCATGCGCGGCCCATTACCTTTAC	- GCCGGATTCG
			5' flank	Representative protospacer sequence	3' flank
			Number of known protospacer matches		
			Representative target genome		
			Nm 8013 CRISPR spacer number		

Figure 4A



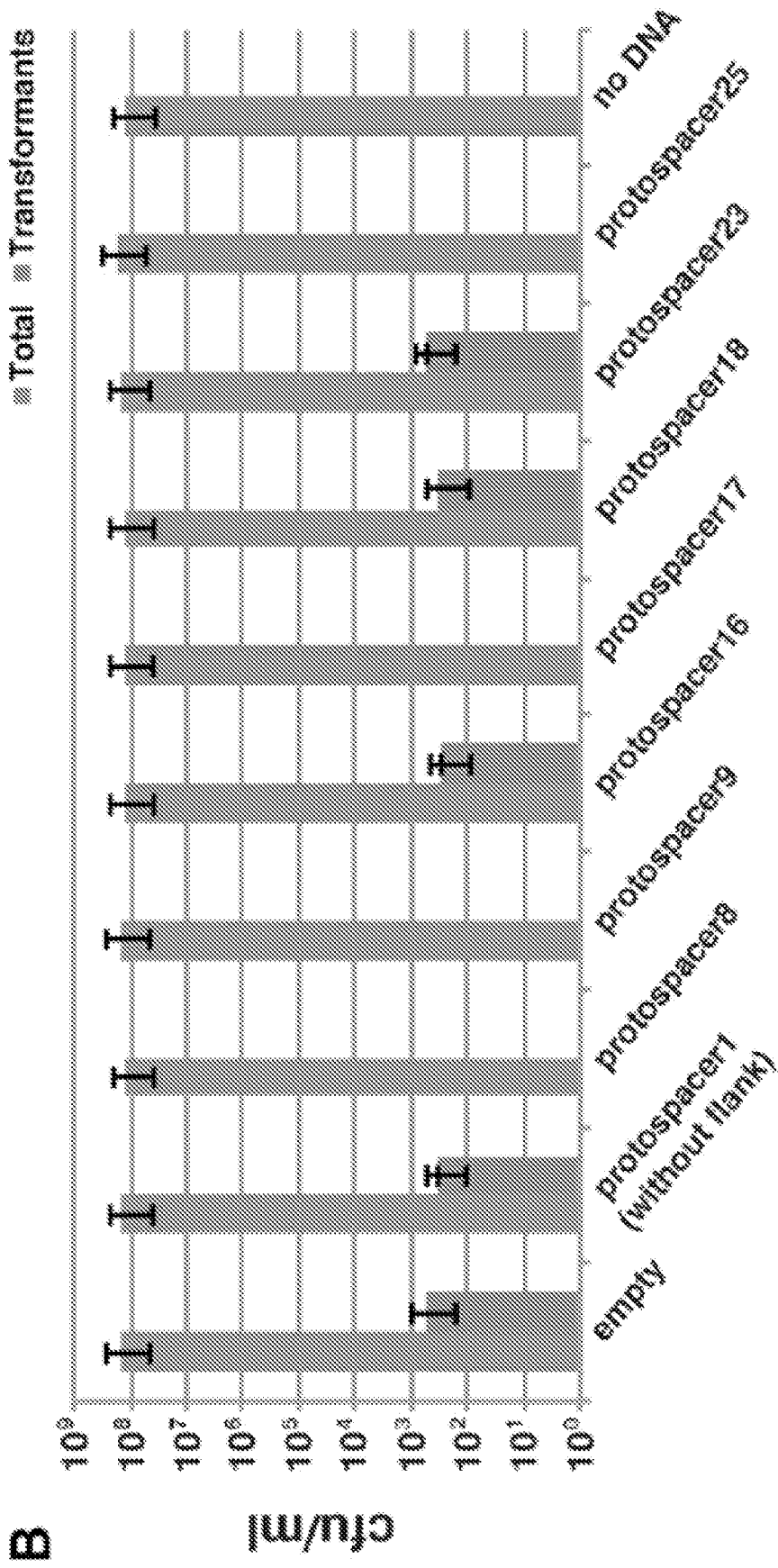


Figure 4B

Figure 4C

Protospacer 9

	Protospacer	3' flank
Wildtype	ATCATGACATCCTCAGATTAGTATTCAGA	- ATATGATTTT
Mutant 1	TACATGACATCCTCAGATTAGTATTCAGA	- ATATGATTTT
Mutant 2	ATCATGACATCCTCTCATTTAGTATTCAGA	- ATATGATTTT
Mutant 3	ATCATGACATCCTCAGATTAGTATTGTGA	- ATATGATTTT
Mutant 4	ATCATGACATCCTCAGATTAGTATTCAGA	- AATTGATTTT
Mutant 5	ATCATGACATCCTCAGATTAGTATTCAGA	- ATATGTATTTT
Mutant 6	--CATGACATCCTCAGATTAGTATTCAGA	- ATATGATTTT
Mutant 7	-----ACATCCTCAGATTAGTATTCAGA	- ATATGATTTT
Mutant 8	ATCATGACATCCTCAGATTAGTATTCAGA	- ATATGATTTT

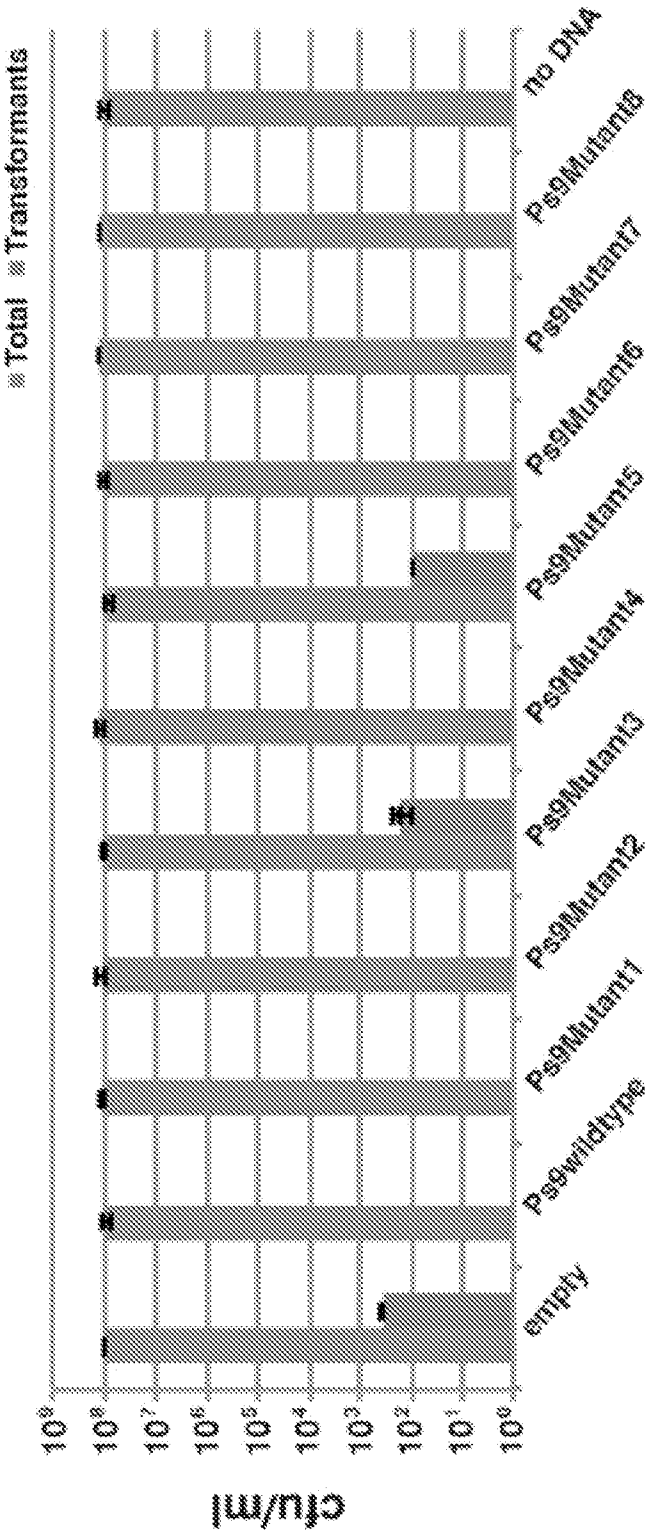


Figure 5A

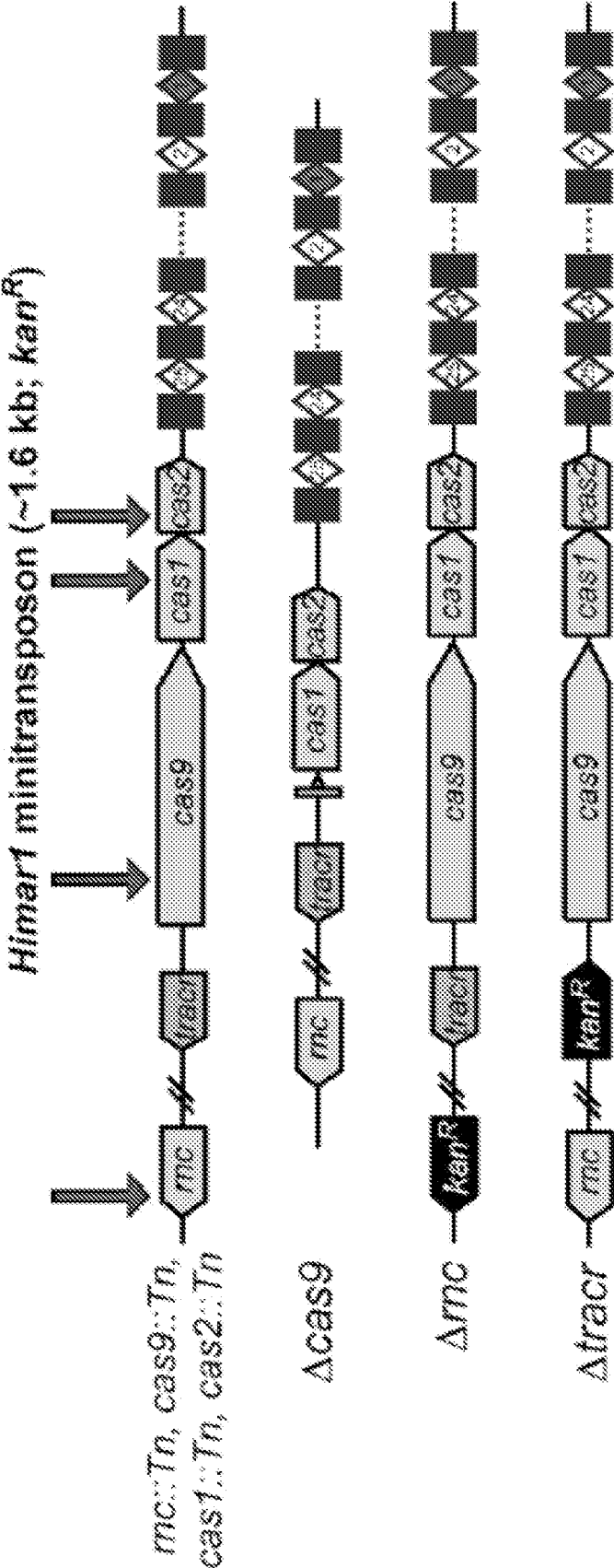


Figure 5C

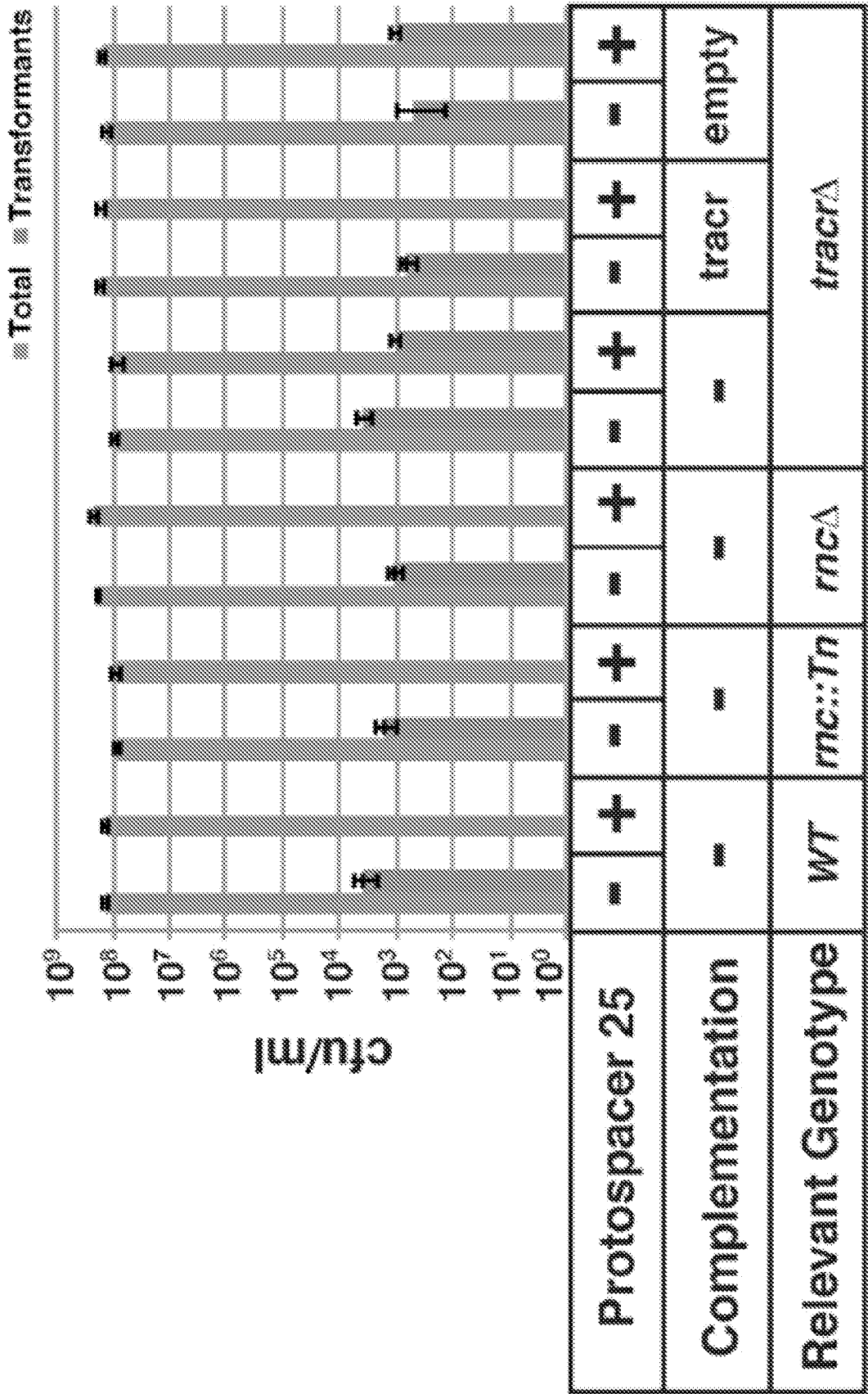


Figure 5D

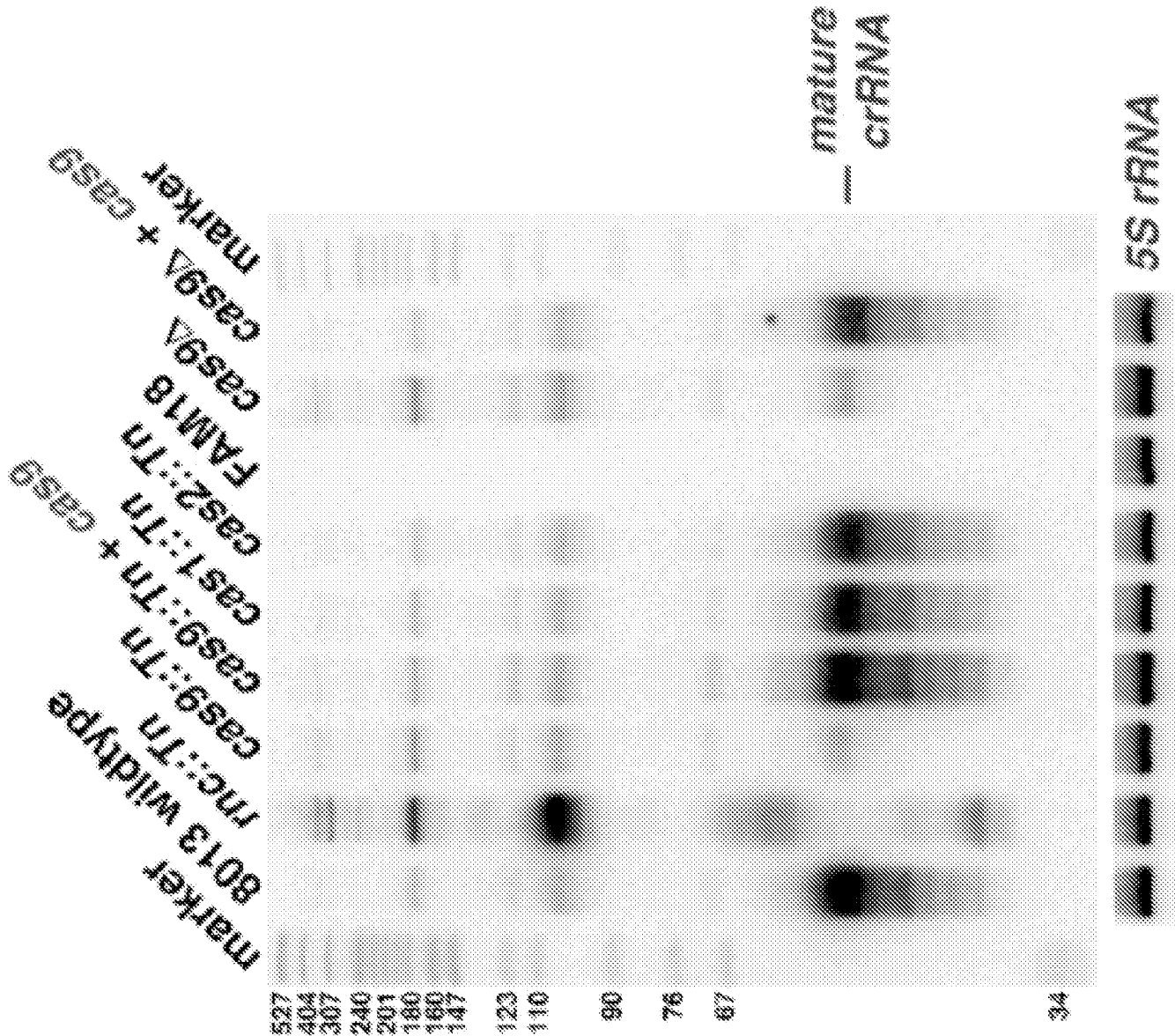


Figure 5E

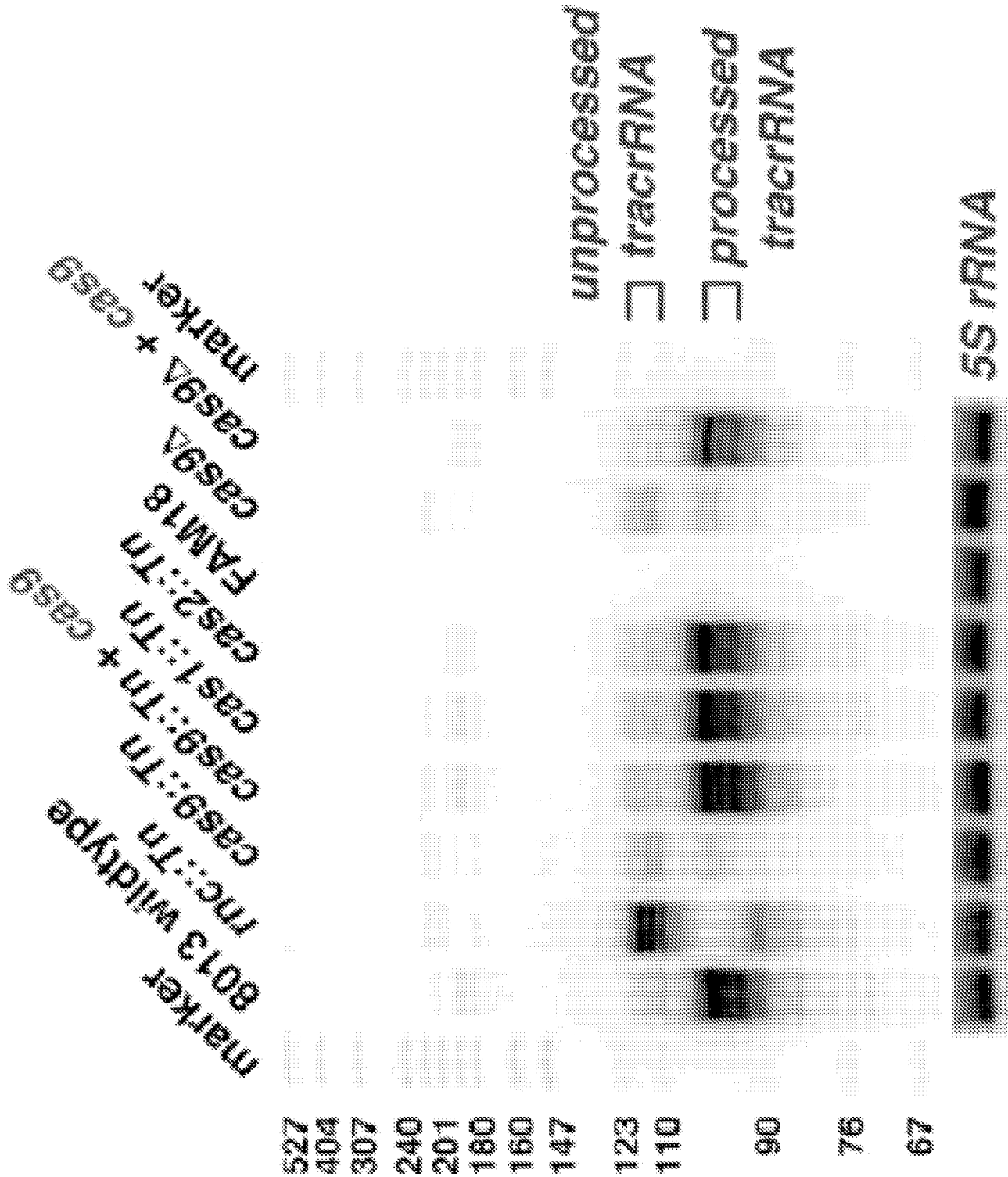


Figure 6A

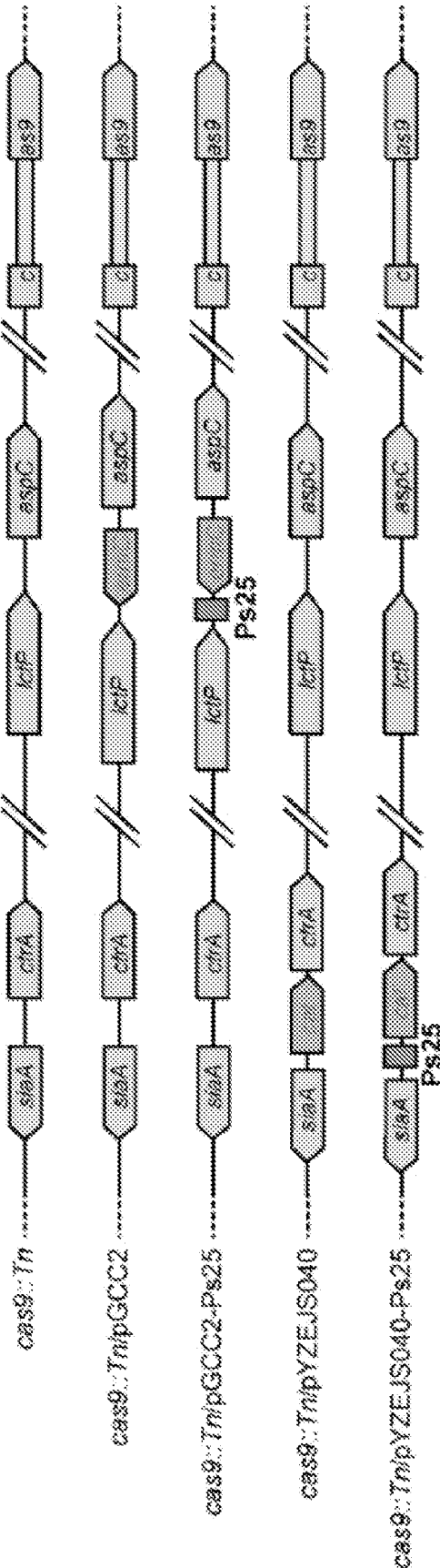


Figure 6B

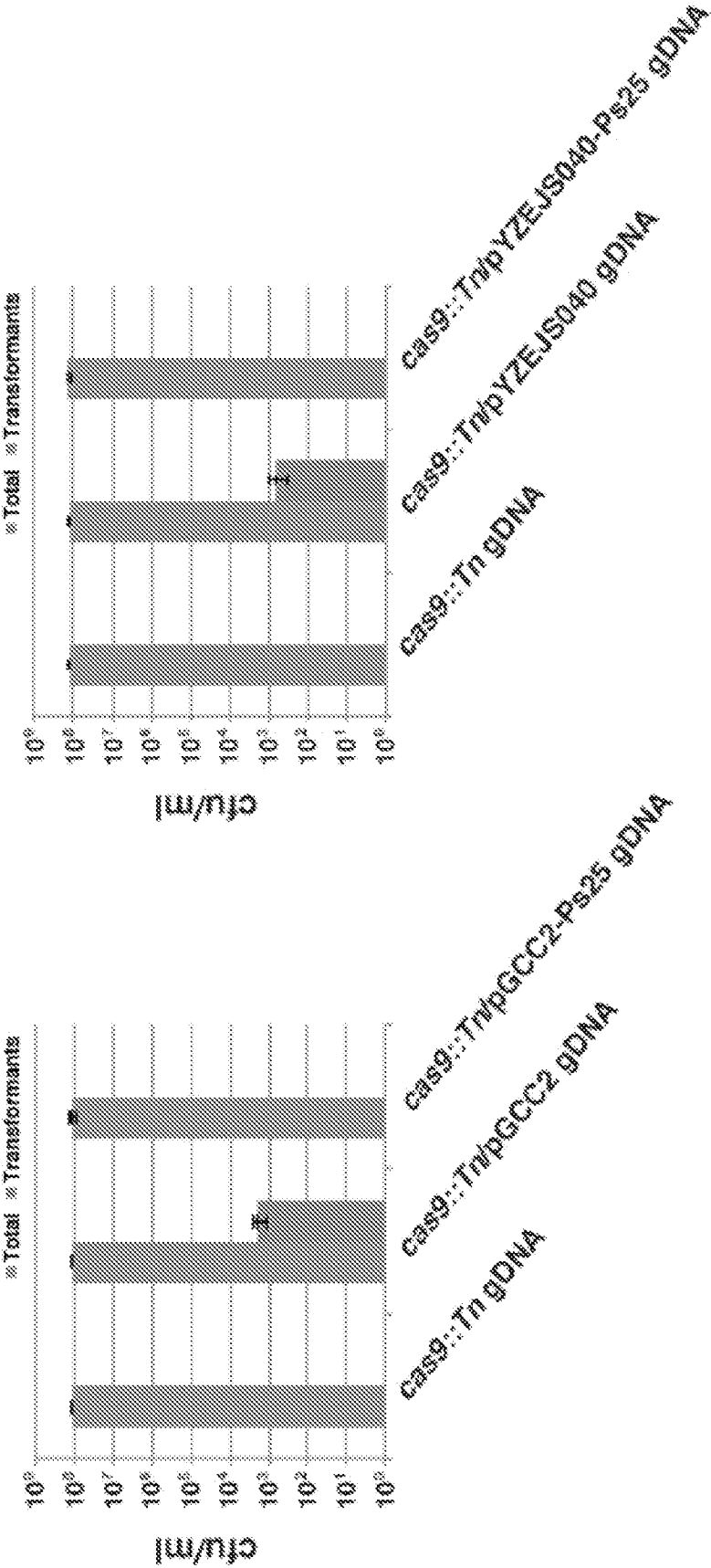
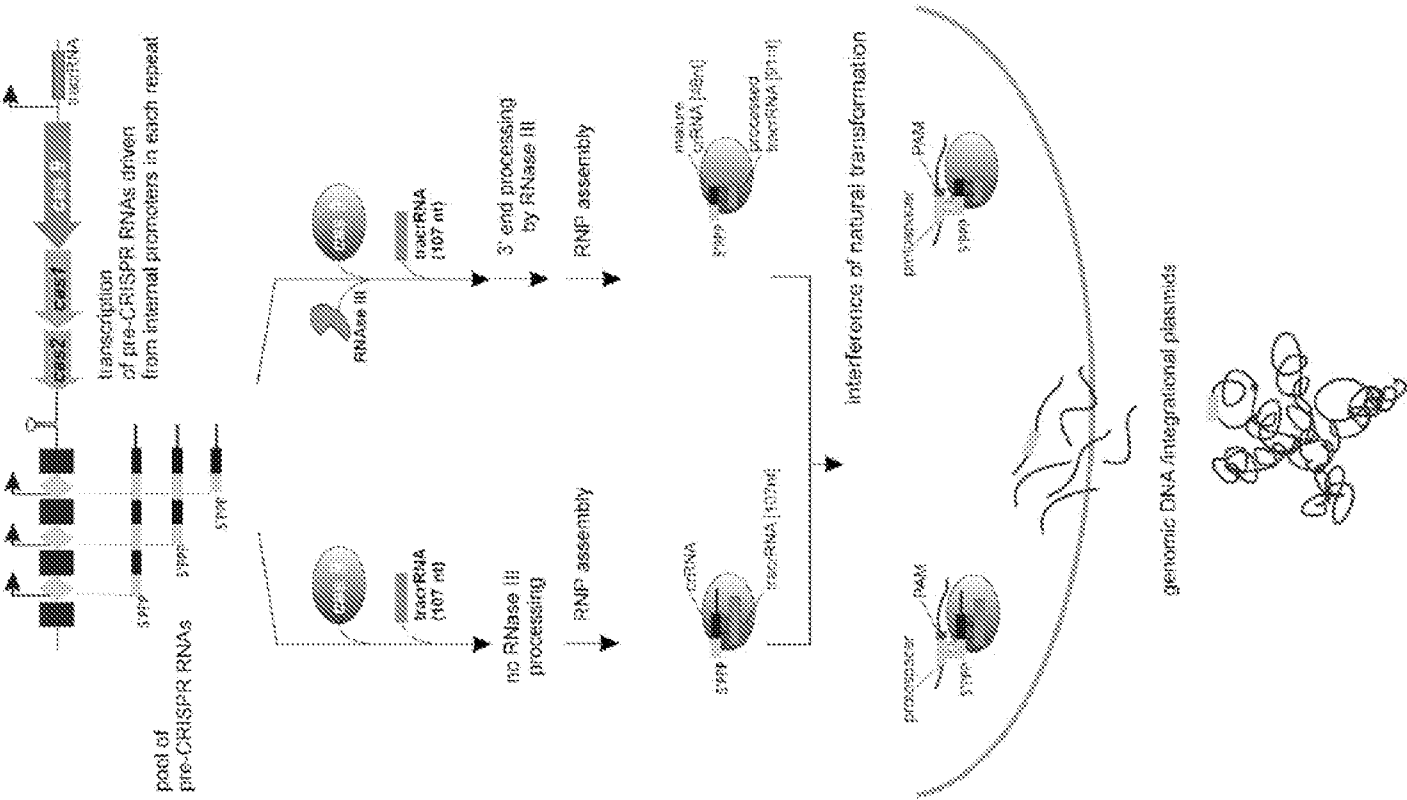


Figure 7



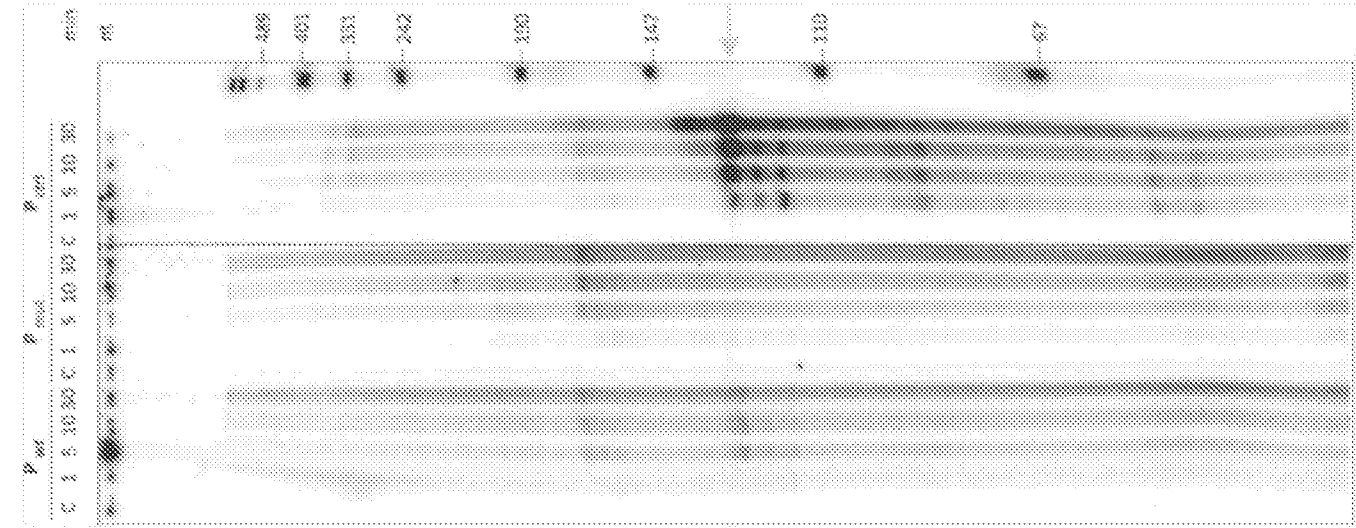


Figure 8

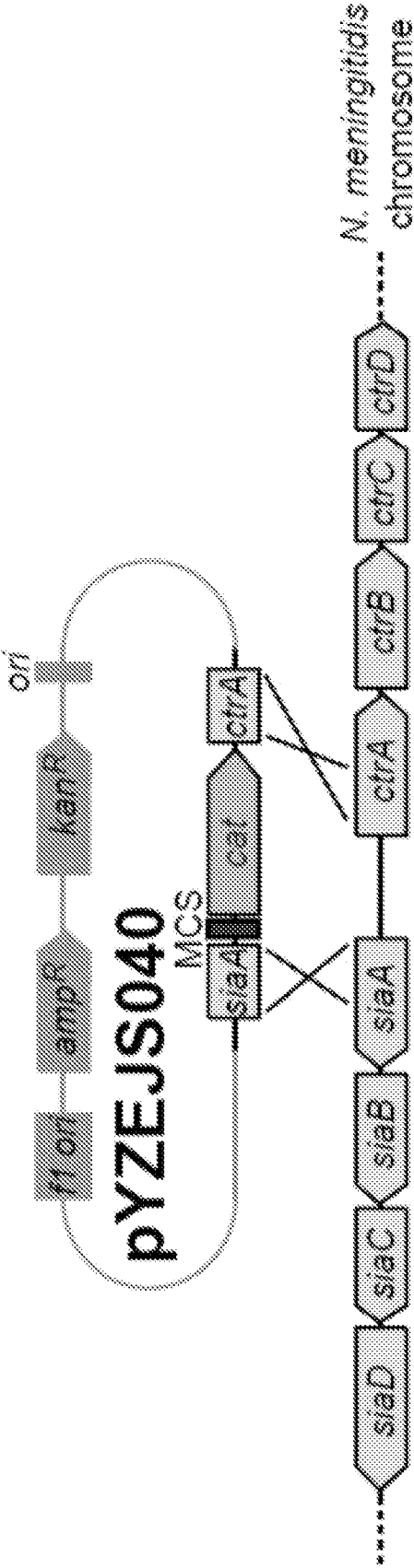


Strain	CAZAP3 strain number	Representative prepubertal matches	Number of known prepubertal matches	5' flank	3' flank
8013	8	8013-240358	15	AGCCTTACCATCAACAGAAATCGTAAAC	TATAGATTTT
8013	9	8013-240358	16	ATCATATCATCTCCAGATTTACTATTCAGA	ATATGATTTT
8013	14	8013-240358	5	CTCTCTCTGGGCTCTCCACAGAGAGAG	AGCGGATTC
8013	16	8013-240358	15	ATACGCTCTGCTACTTGTCTACTCTGAGCA	TATATCTTAA
8013	17	8013-240358	6	ACCTGTCTGCTCTTCTGCTCTGCTCTTA	TATGATTTGA
8013	18	8013-240358	16	CCGTACATCAAAACAGCGAAATTTGCT	TTCTGCTTGA
8013	21	8013-240358	1	TAGCTCTCTATGCTCATATGAGATGCTT	TCCCGCTTAA
8013	23	8013-240358	1	TCTGCTCTGCTCTCTGCTCTGCTCTGCTG	CTCTATTTTC
8013	25	8013-240358	1	TATCTCTCATGCTGCTGCTCATCTCTCTAC	CCCGGATTC
8013	2	8013-240358	5	CTCTCTCTCTTATCAAGCTGCTCTCTT	TCTCGAGAGG
8013	3	8013-240358	9	CTCTCTCTCTTATCAAGCTGCTCTCTT	CGCGAAGG
8013	12	8013-240358	1	GGTCTCTCTCTTATCAAGCTGCTCTCTT	CTCTATTTT
8013	13	8013-240358	1	CGTCTCTCTCTTATCAAGCTGCTCTCTT	TATGATTTT
8013	14	8013-240358	2	CTCTCTCTCTTATCAAGCTGCTCTCTT	ATCTGATTTT
8013	15	8013-240358	6	TACCGCTCTCTTATCAAGCTGCTCTCTT	CGATGATTTT
8013	7	8013-240358	6	ATCTCTCTCTTATCAAGCTGCTCTCTT	CGATGATTTT
8013	12	8013-240358	11	CTCTCTCTCTTATCAAGCTGCTCTCTT	CATCTGATTTT
8013	13	8013-240358	9	CTCTCTCTCTTATCAAGCTGCTCTCTT	TGAGATTTT
8013	14	8013-240358	51	CGCTCTCTCTTATCAAGCTGCTCTCTT	AGATTTTAA
8013	17	8013-240358	6	CTCTCTCTCTTATCAAGCTGCTCTCTT	CTCTGATTTT
8013	18	8013-240358	20	CTCTCTCTCTTATCAAGCTGCTCTCTT	CGCTGATTTT
8013	19	8013-240358	9	CTCTCTCTCTTATCAAGCTGCTCTCTT	CTCTGATTTT
8013	20	8013-240358	5	CTCTCTCTCTTATCAAGCTGCTCTCTT	CATCTGATTTT
8013	21	8013-240358	12	CTCTCTCTCTTATCAAGCTGCTCTCTT	AGATTTTAA
8013	23	8013-240358	19	CTCTCTCTCTTATCAAGCTGCTCTCTT	AGATTTTAA
8013	5	8013-240358	9	CTCTCTCTCTTATCAAGCTGCTCTCTT	CTCTGATTTT
8013	13	8013-240358	15	CTCTCTCTCTTATCAAGCTGCTCTCTT	CTCTGATTTT
8013	16	8013-240358	23	CTCTCTCTCTTATCAAGCTGCTCTCTT	CTCTGATTTT
8013	7	8013-240358	1	CTCTCTCTCTTATCAAGCTGCTCTCTT	TATGATTTT
8013	8	8013-240358	18	CTCTCTCTCTTATCAAGCTGCTCTCTT	AGATTTTAA
8013	9	8013-240358	6	CTCTCTCTCTTATCAAGCTGCTCTCTT	CTCTGATTTT
8013	4	8013-240358	1	CTCTCTCTCTTATCAAGCTGCTCTCTT	TATGATTTT
8013	6	8013-240358	3	CTCTCTCTCTTATCAAGCTGCTCTCTT	AGATTTTAA

Figure 10

1		TTCTATTATAAAGCTGCCCTGACTTACTCG	-	
8	(71 nt)...	GACCTCAAAA	-	ACGCTTACCAATCAACACGAAATCGTAAC
9	(136 nt)...	TGGTTTGTGA	-	ATCATGACATCCTCAGATTTAGTATTCAGA
16		GAAATCAGCG	-	ATACGGCCCTCGCACTTGTCCTCACTCGAGCA
17	(48 nt)...	GGACTGTTC	-	ACCTGTTCGGTCCCTTTGCTGTCTGTCCTTA
18	(77 nt)...	CGCATACCGA	-	CCTGAGCATCAAAAACAGCGGAAATTCGT
23		TTTGACTGT	-	TCCGGCGGTTTCATGCGCGTTCCTTGCTG
25		TACAGATTTT	-	TATTCCTCATGCGCGGCGCATTACCTTTAC
N71 8013 CHSP4 spacer number		5' flank		
		Proospacer sequence		
		3' flank		
		TACAGATTTT... (87 nt)		
		ATATGATTTT... (164 nt)		
		AATATCTGAA		
		TAAGGATTGA... (207 nt)		
		GTTGTCCTGCA... (76 nt)		
		CGTTAATTTC		
		GCCGGATTGC		

Figure 11A



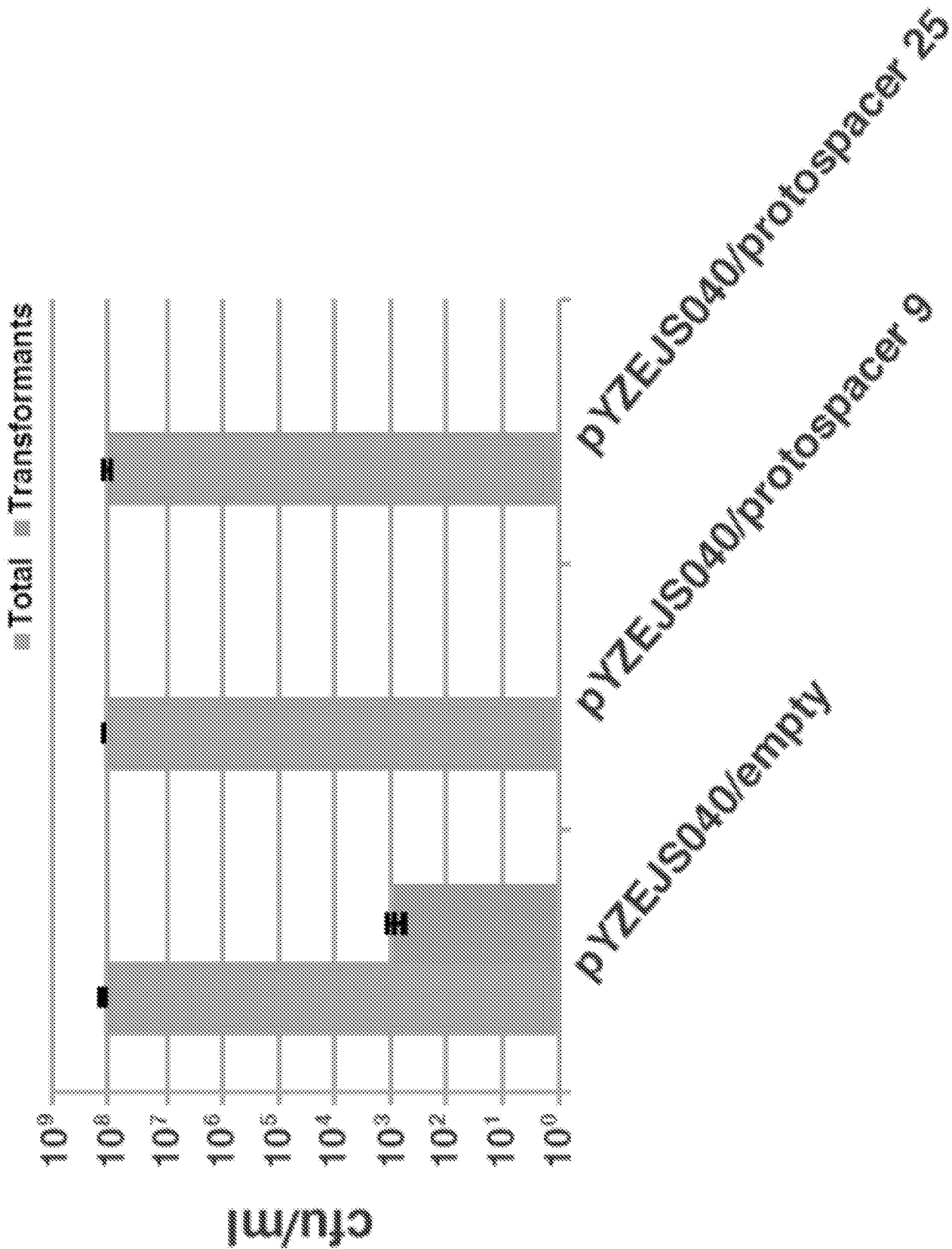


Figure 11B

Figure 12

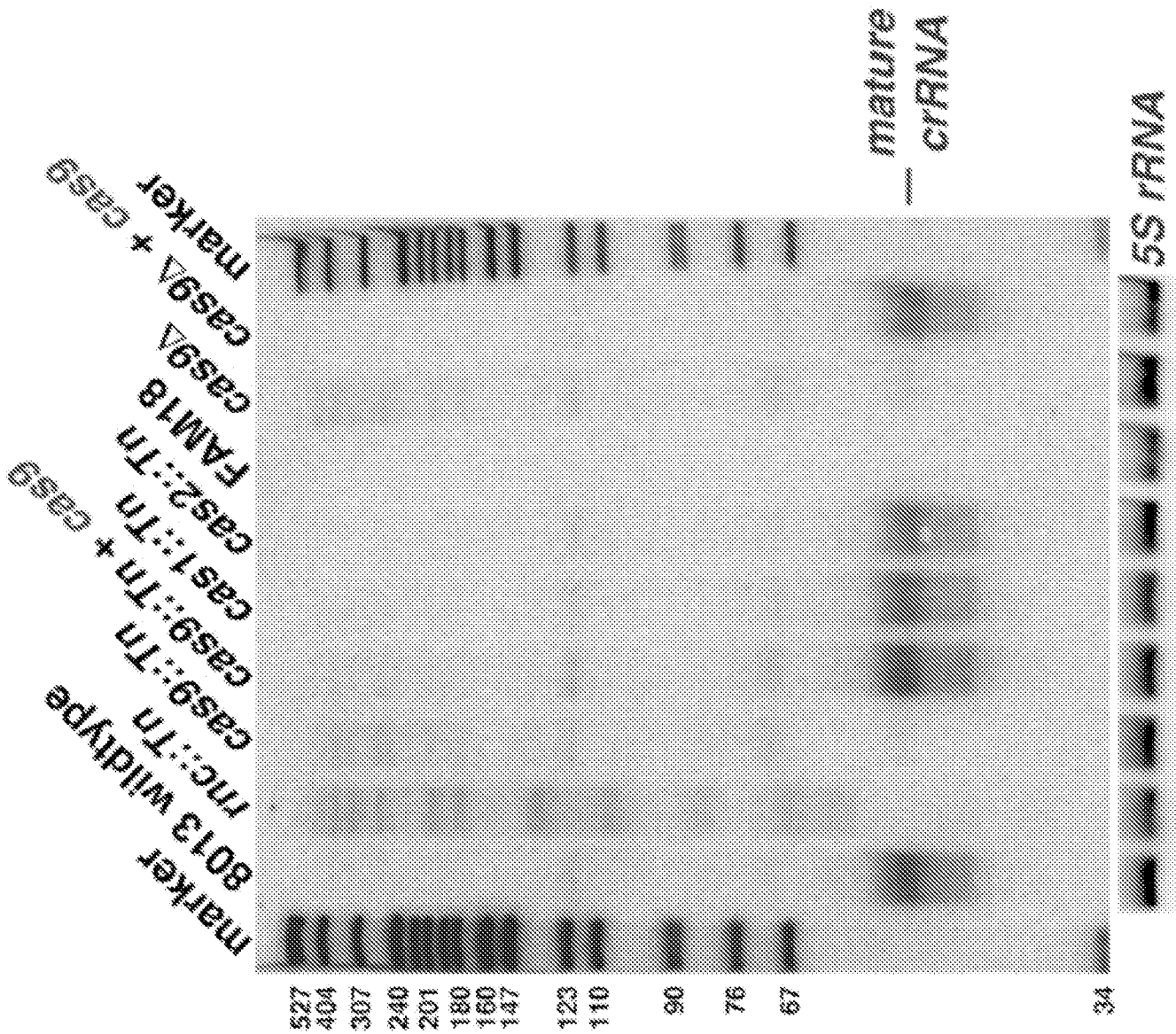


Figure 13A

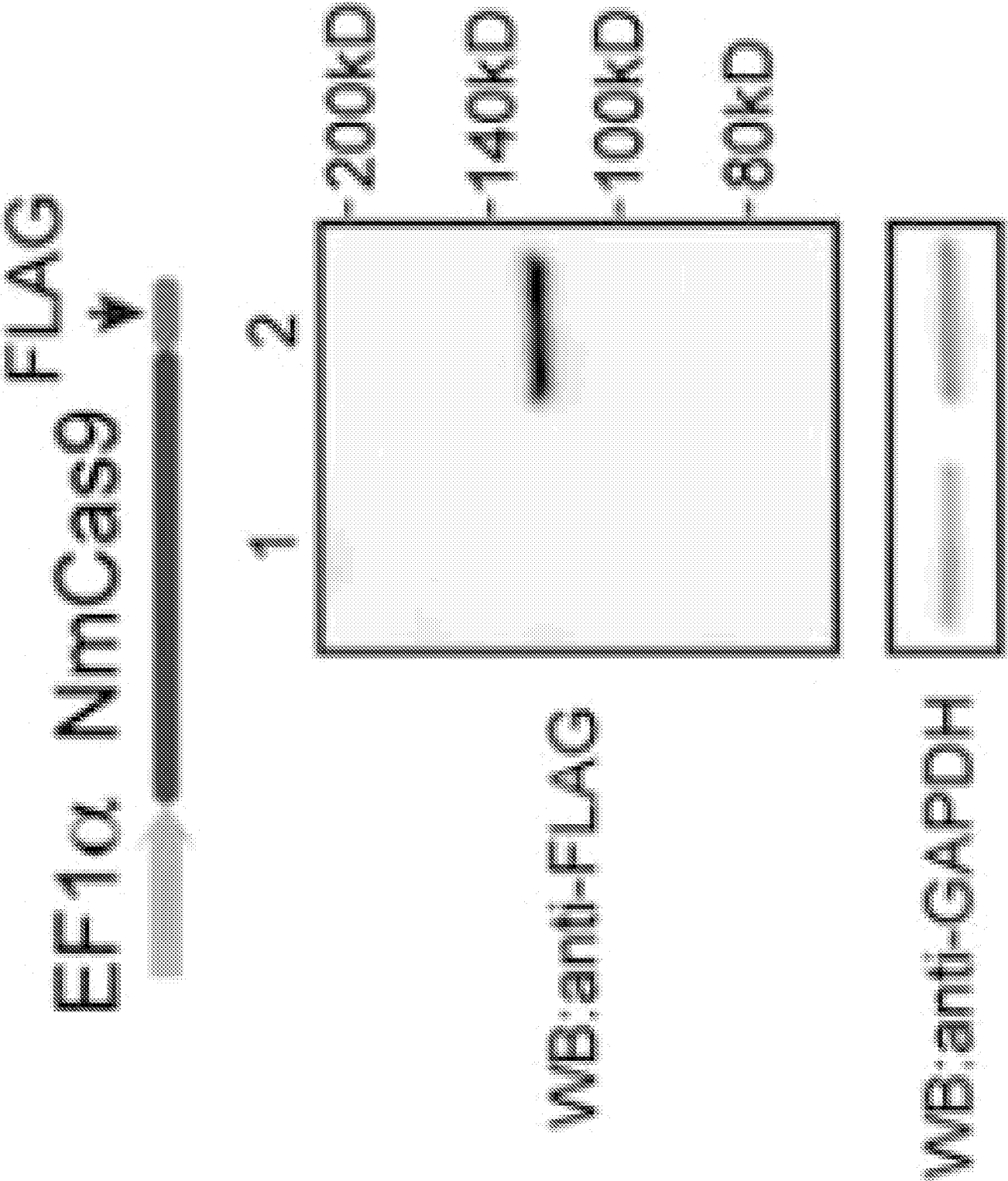
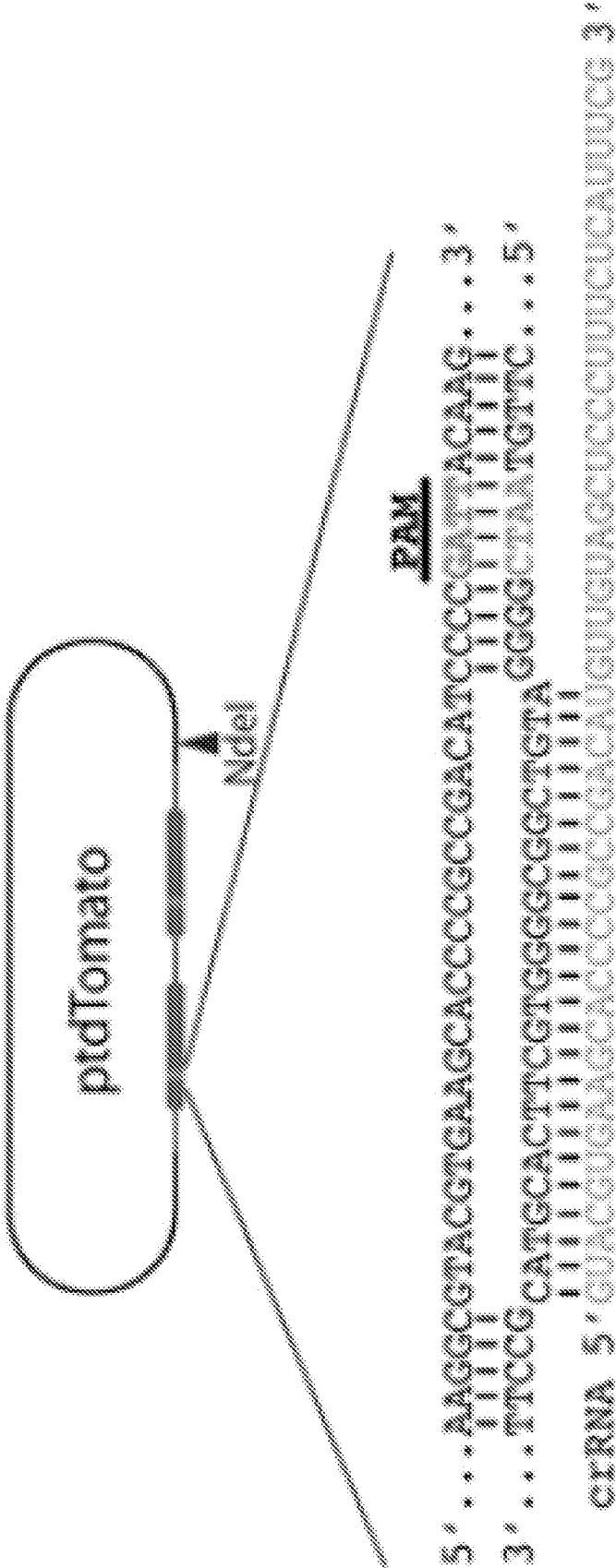


Figure 13B



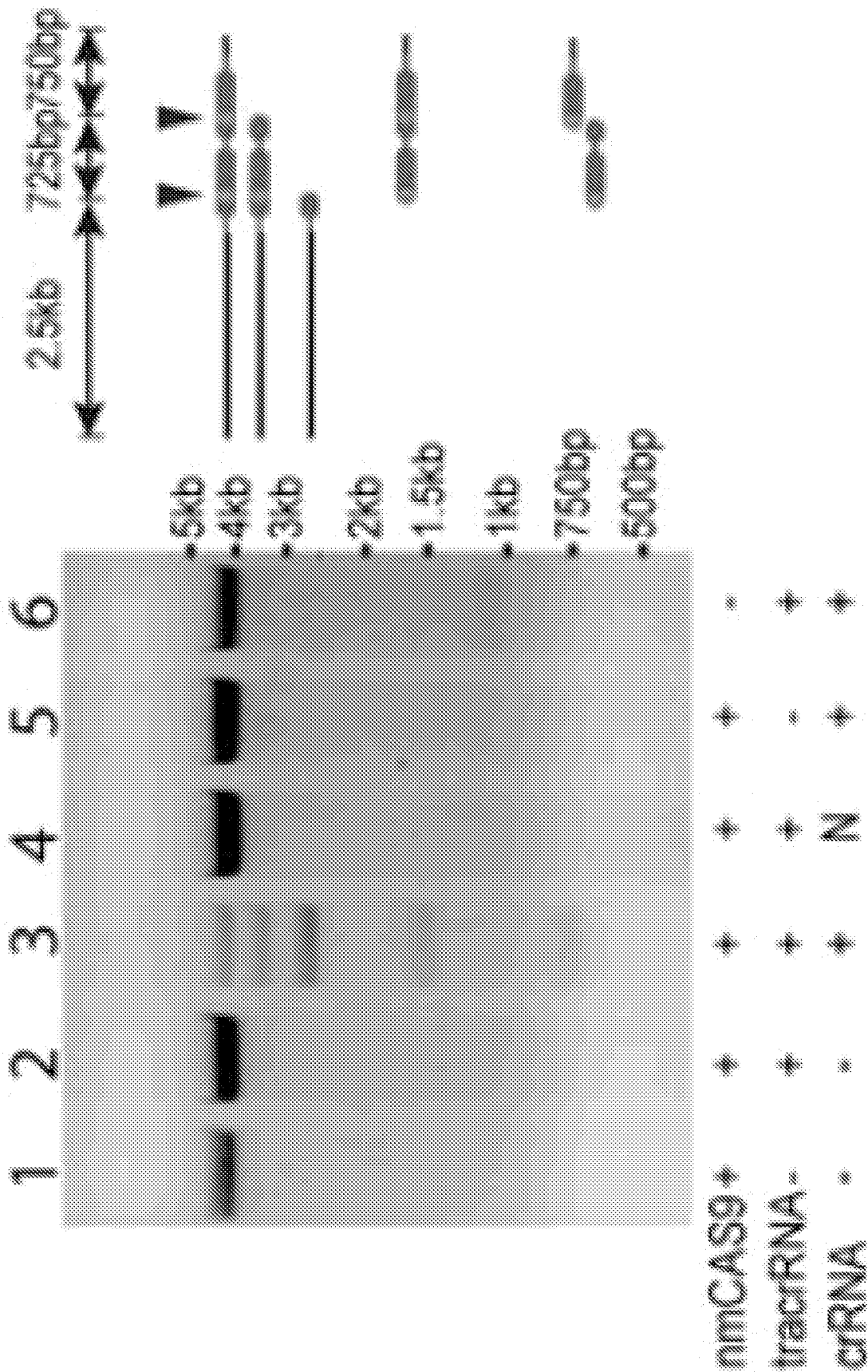


Figure 13C

Figure 13D

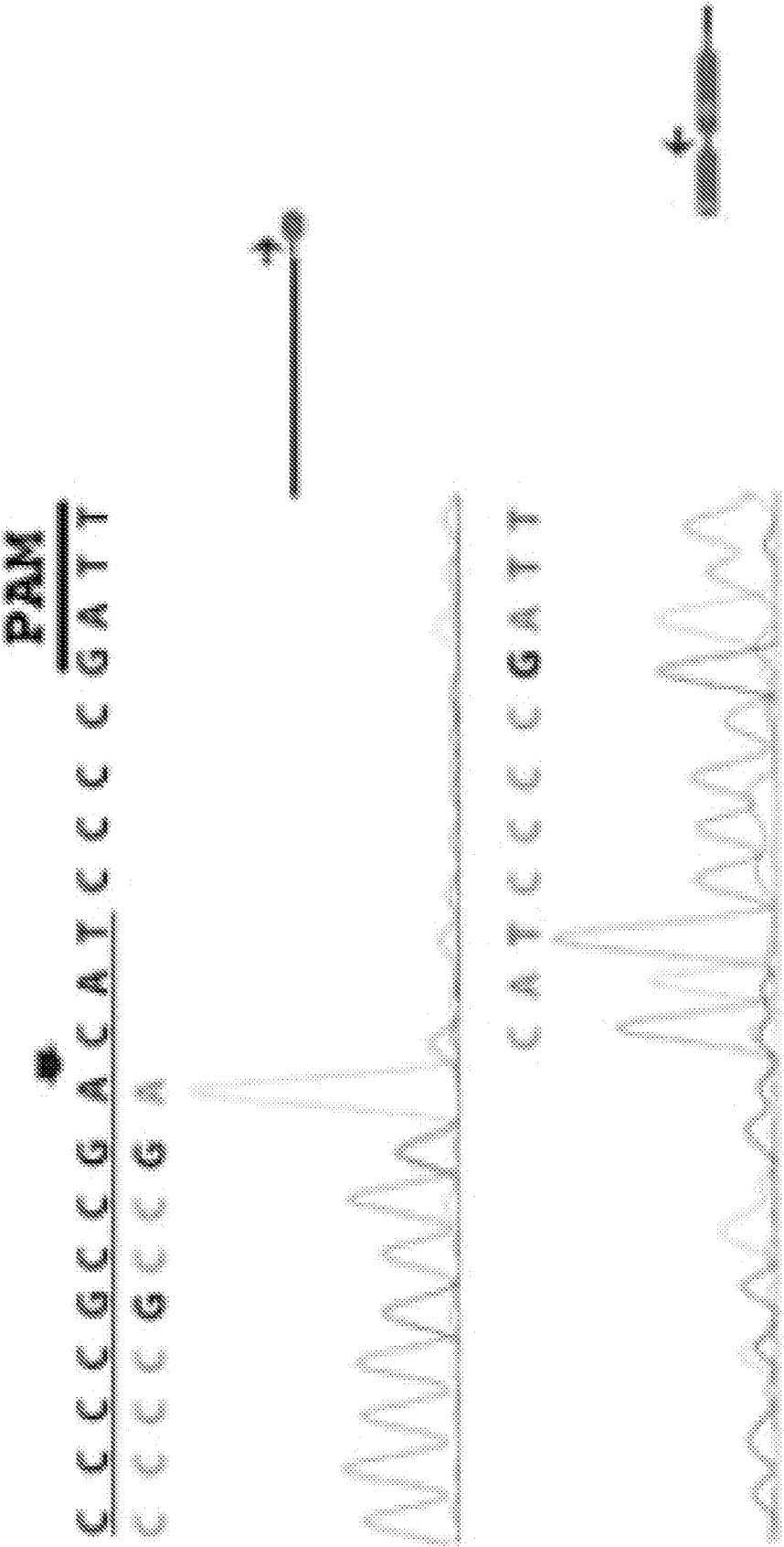
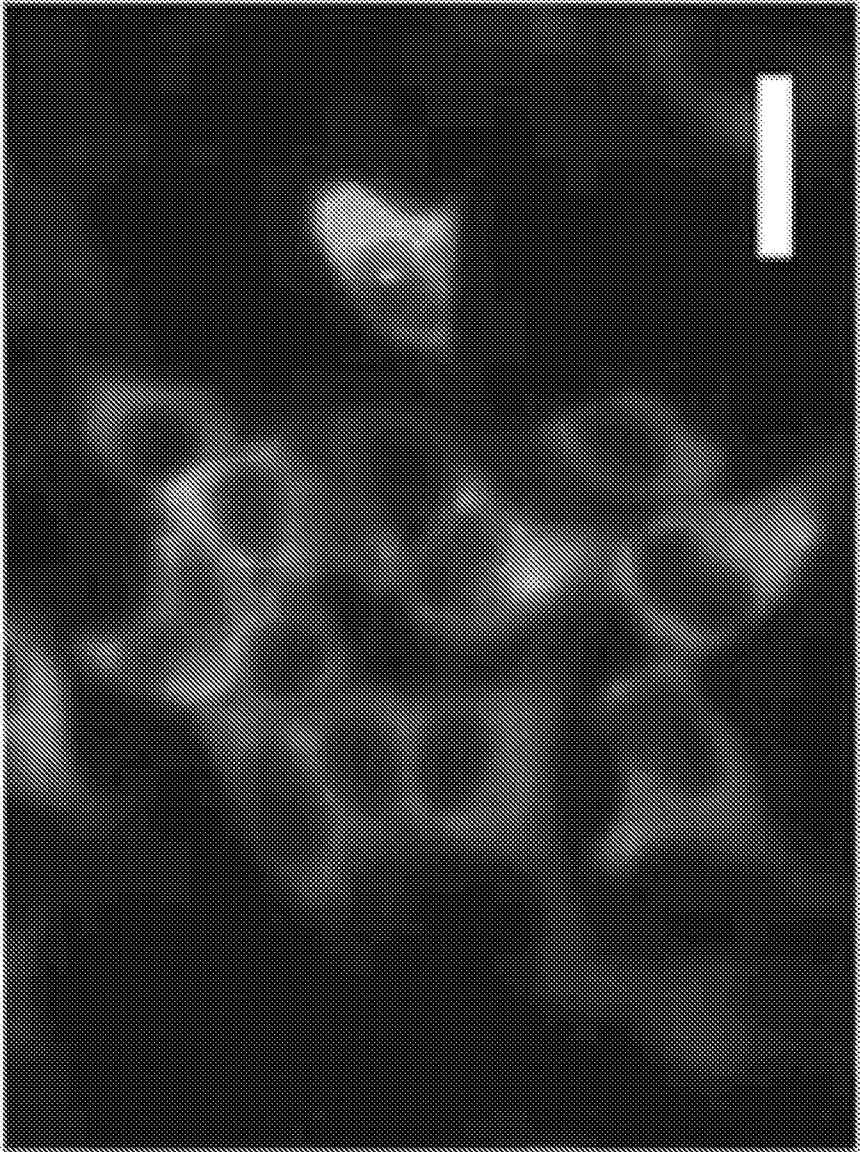
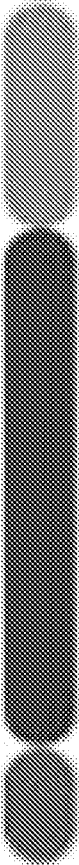


Figure 14A

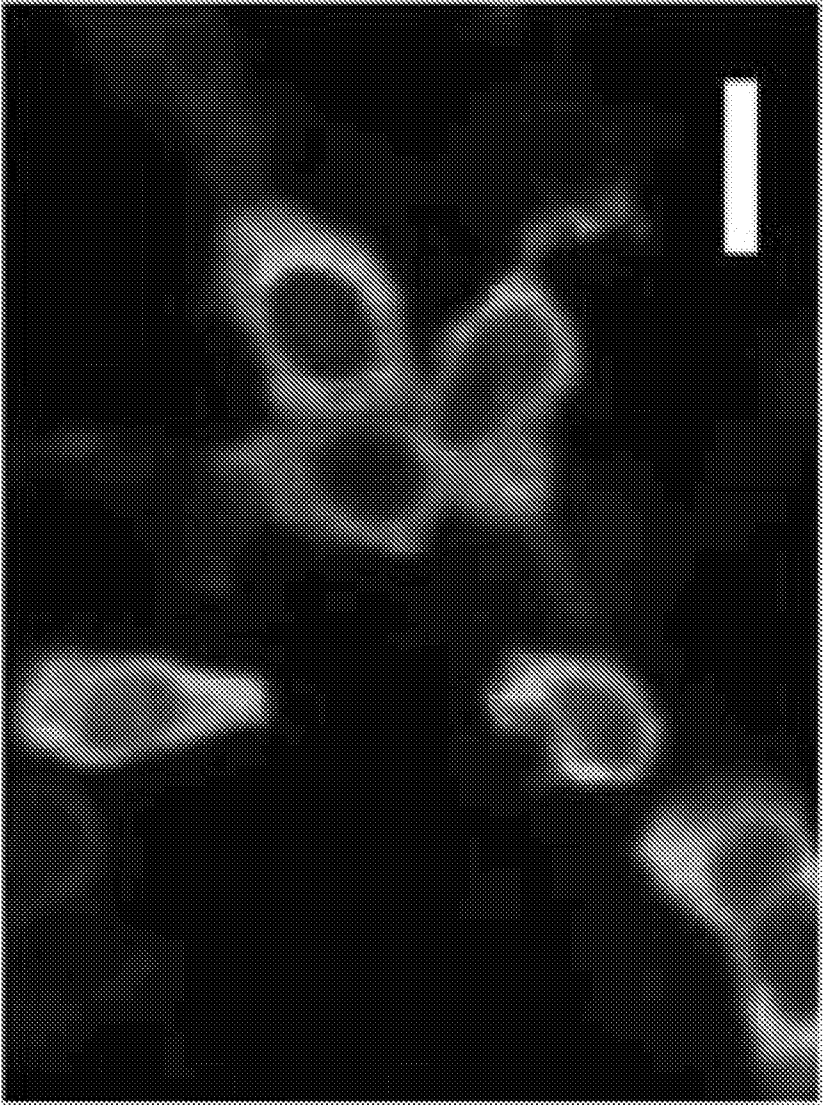
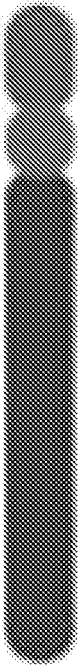
NLS NmCAS9 GFP



293FT

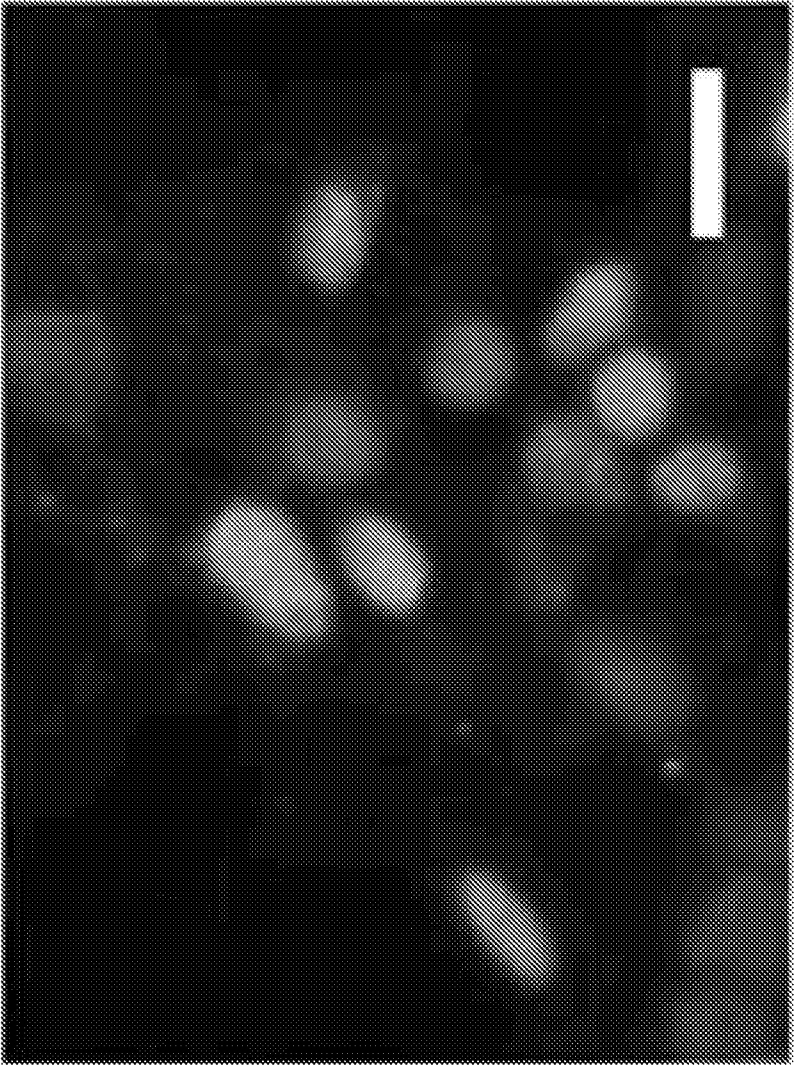
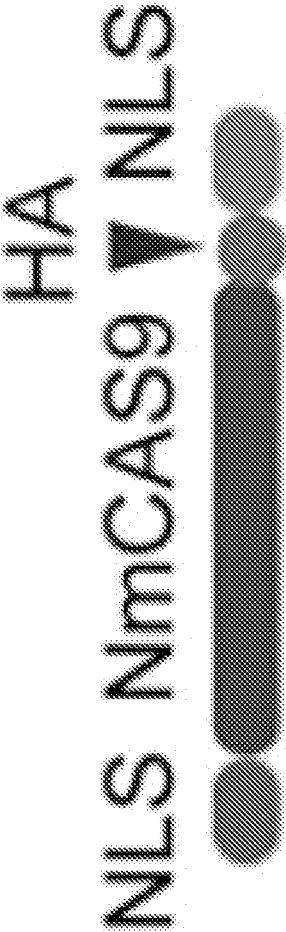
Figure 14B

HA
NmCAS9 ▼ NLS



293FT

Figure 14C



293FT

Figure 14D

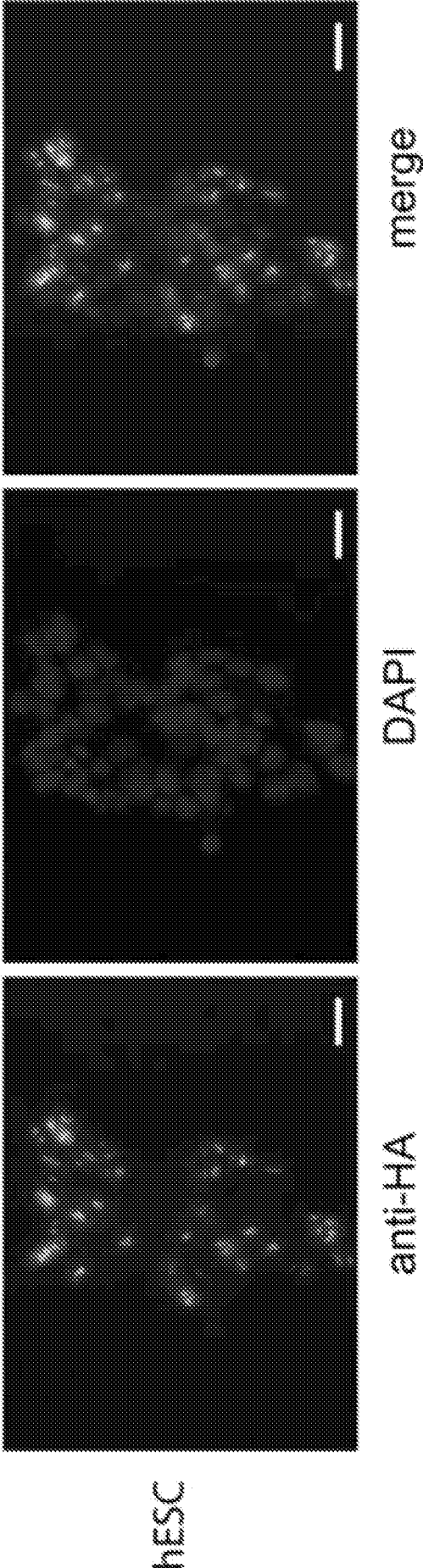


Figure 14E

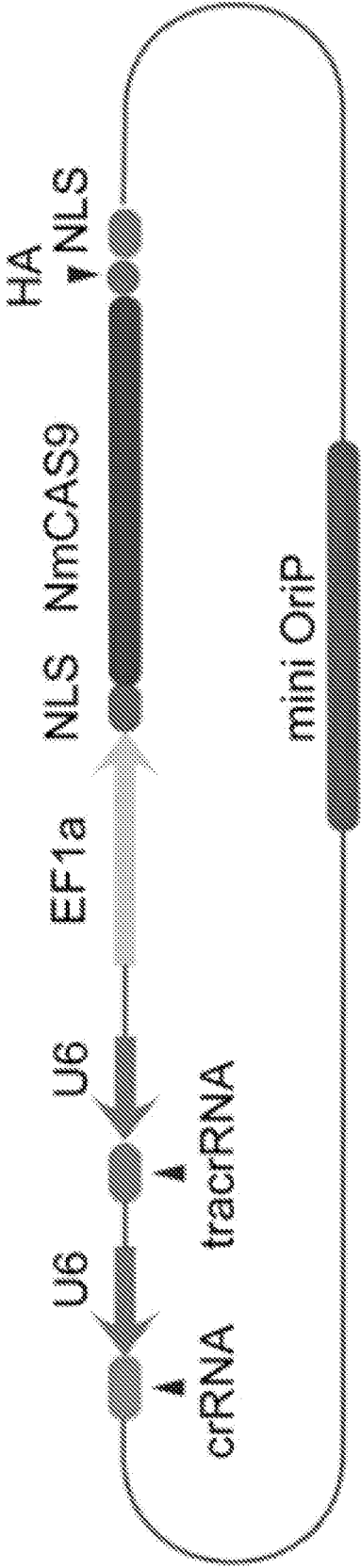


Figure 14F

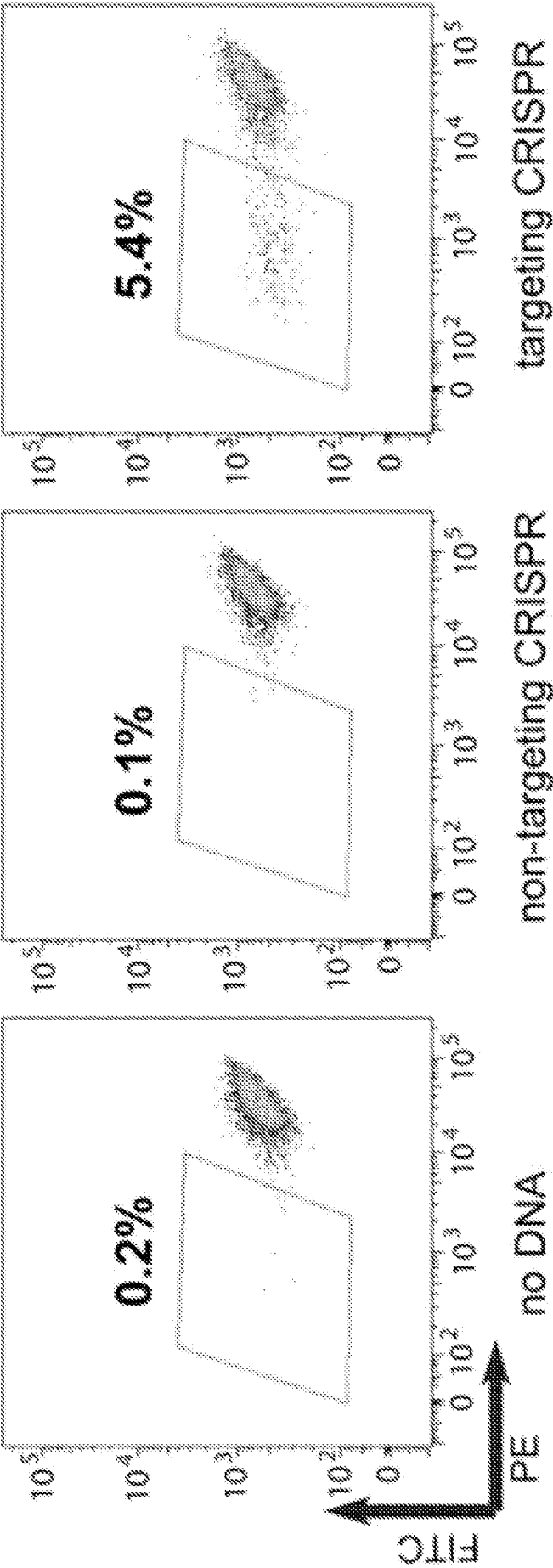
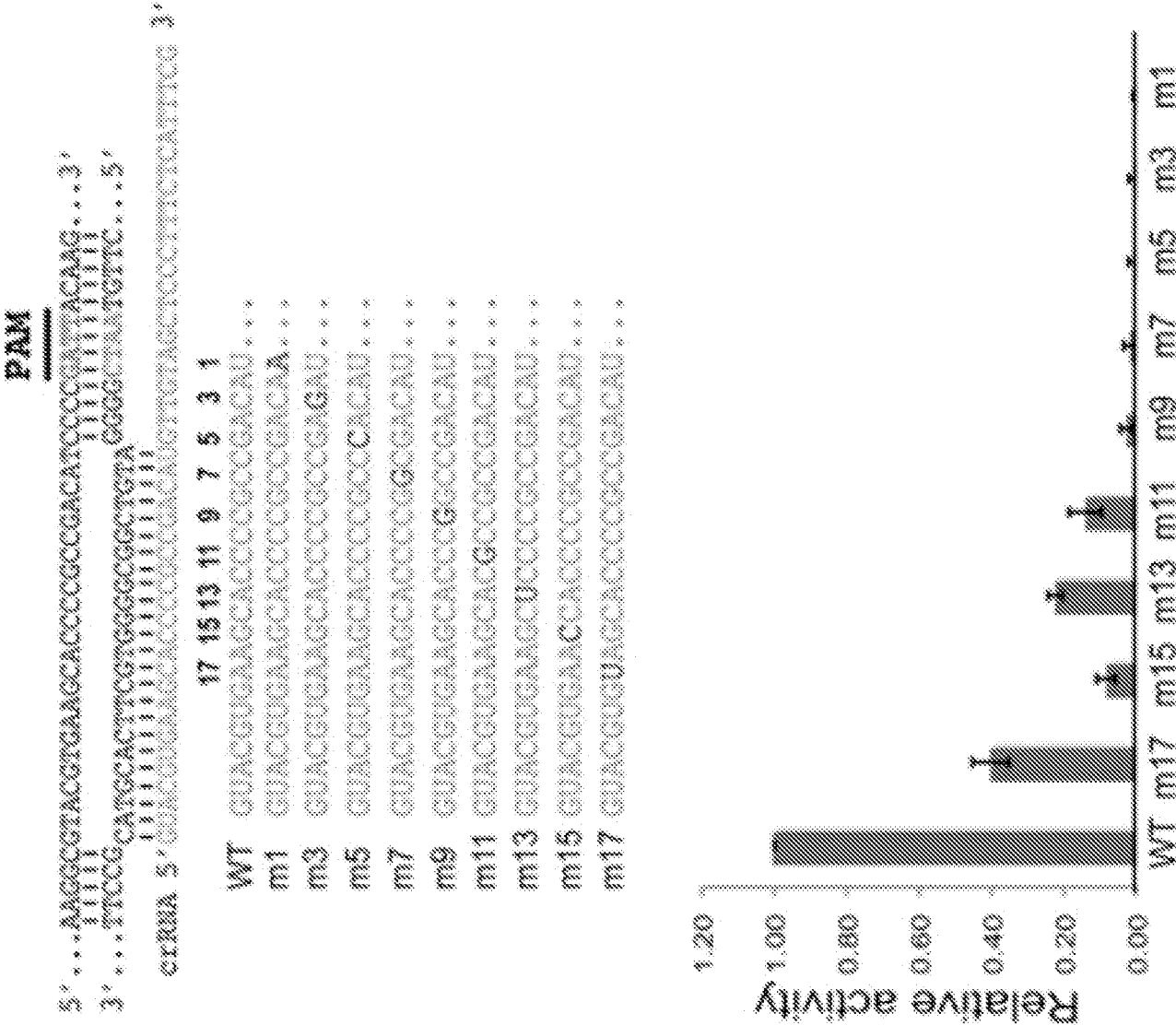


Figure 14G

Deletions		PAM
WT	CCAAGGCGTACGTGAAGCACCCCGCCG <u>ACAT</u> CCCCGATTACAA	<u>CCCCGATTACAA</u>
D1 (2)	CCAAGGCGTACGTGAAGCACCCCGCCGA--ATCCCCGATTACAA	
D1 (2)	CCAAGGCGTACGTGAAGCACCCCGCCG--CATCCCCGATTACAA	
D2 (2)	CCAAGGCGTACGTGAAGCACCCCGCCG--ATCCCCGATTACAA	
D9 (1)	CCAAGGCGTACGTGAAGCACCCCG--TCCCCGATTACAA	
D11 (1)	CCAAGGCGTACGTGAAGCA-----TCCCCGATTACAA	
D11 (1)	CCAAGGCGTACGTGAAGCAC-----CCCCGATTACAA	
Insertions		PAM
WT	CCAAGGCGTACGTGAAGCACCCCGCCG <u>ACAT</u> CCCCGATTACAA	<u>CCCCGATTACAA</u>
I1 (2)	CCAAGGCGTACGTGAAGCACCCCGCCGAG-----CATCCCCGATTACAA	
I1 (3)	CCAAGGCGTACGTGAAGCACCCCGCCGAA-----CATCCCCGATTACAA	
I1 (3)	CCAAGGCGTACGTGAAGCACCCCGCCGAC-----CATCCCCGATTACAA	
I2 (1)	CCAAGGCGTACGTGAAGCACCCCGCCGAGA-----CATCCCCGATTACAA	
I7 (1)	CCAAGGCGTACGTGAAGCACCCCGCCGATTACAAG--CATCCCCGATTACAA	
I10 (1)	CCAAGGCGTACGTGAAGCACCCCGCCGATTACAAGACATCCCCGATTACAA	
I+D (1)	CCAAGGCGTACGTGAAGTACGTGAAG-----CATCCCCGATTACAA	

Figure 15A



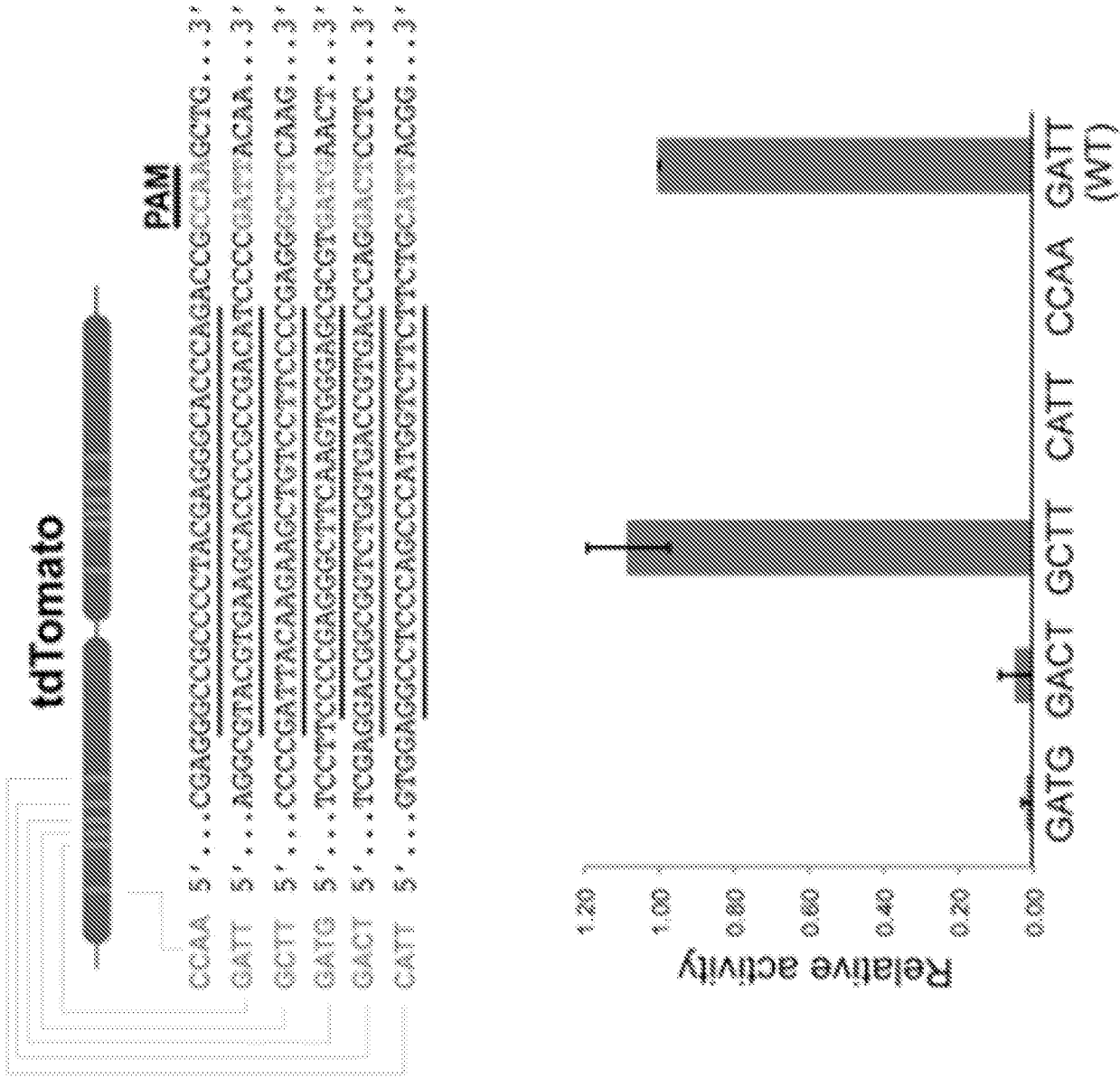


Figure 15B

Figure 16A

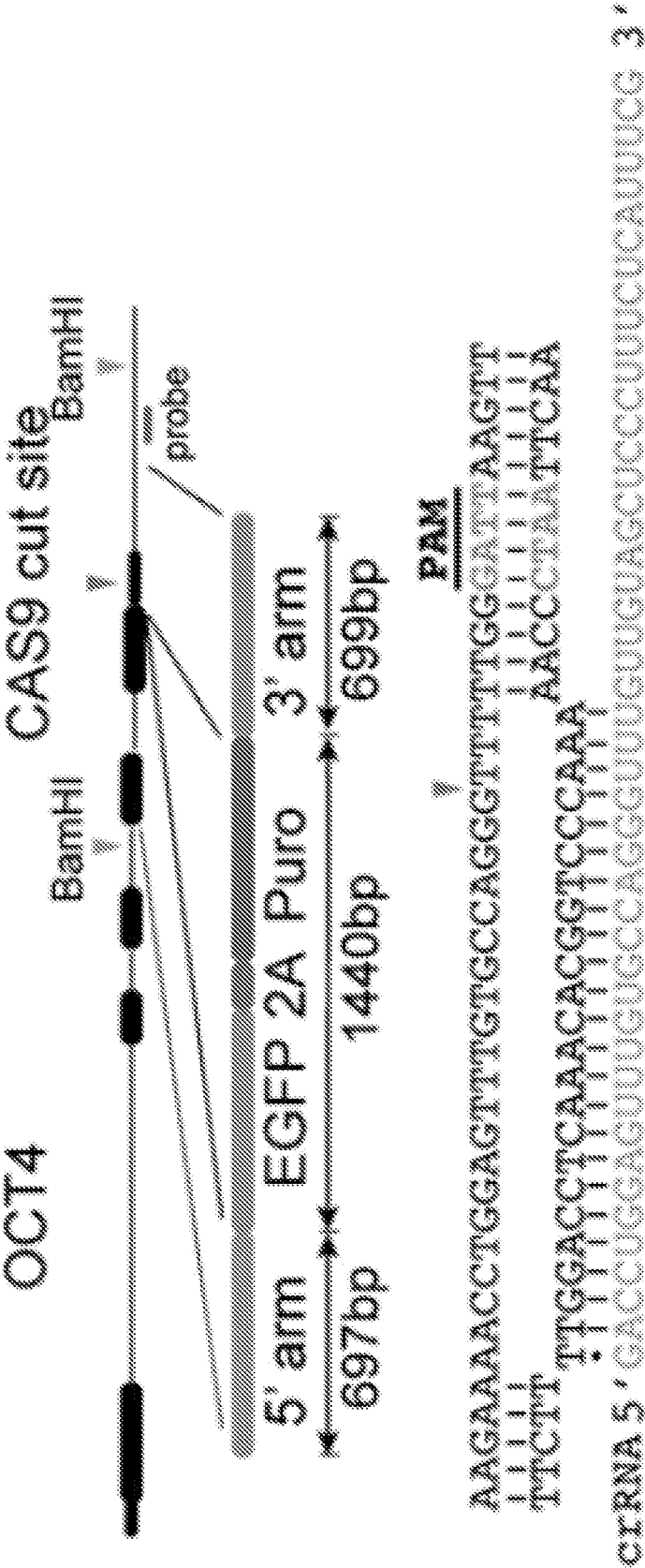


Figure 16B

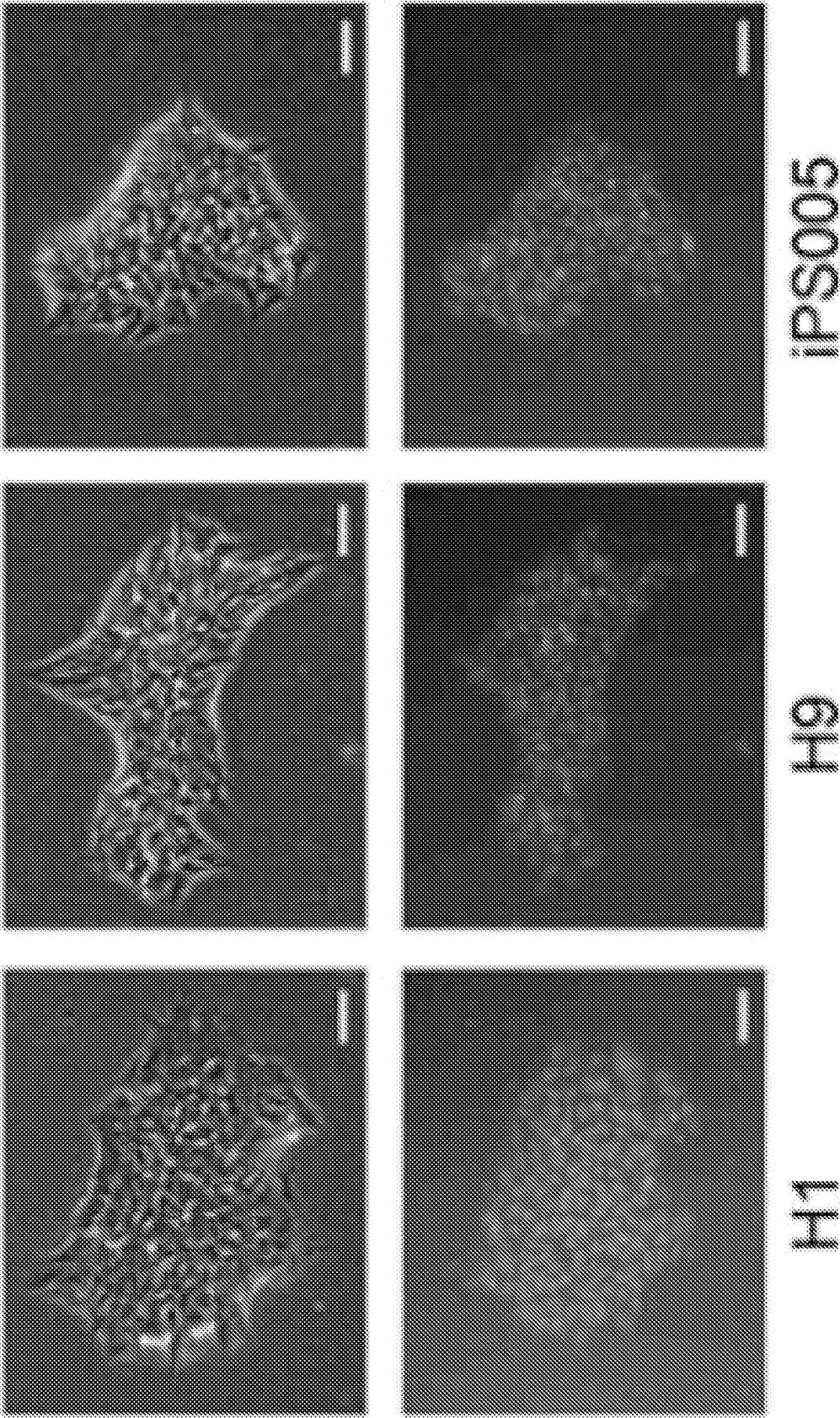


Figure 16C

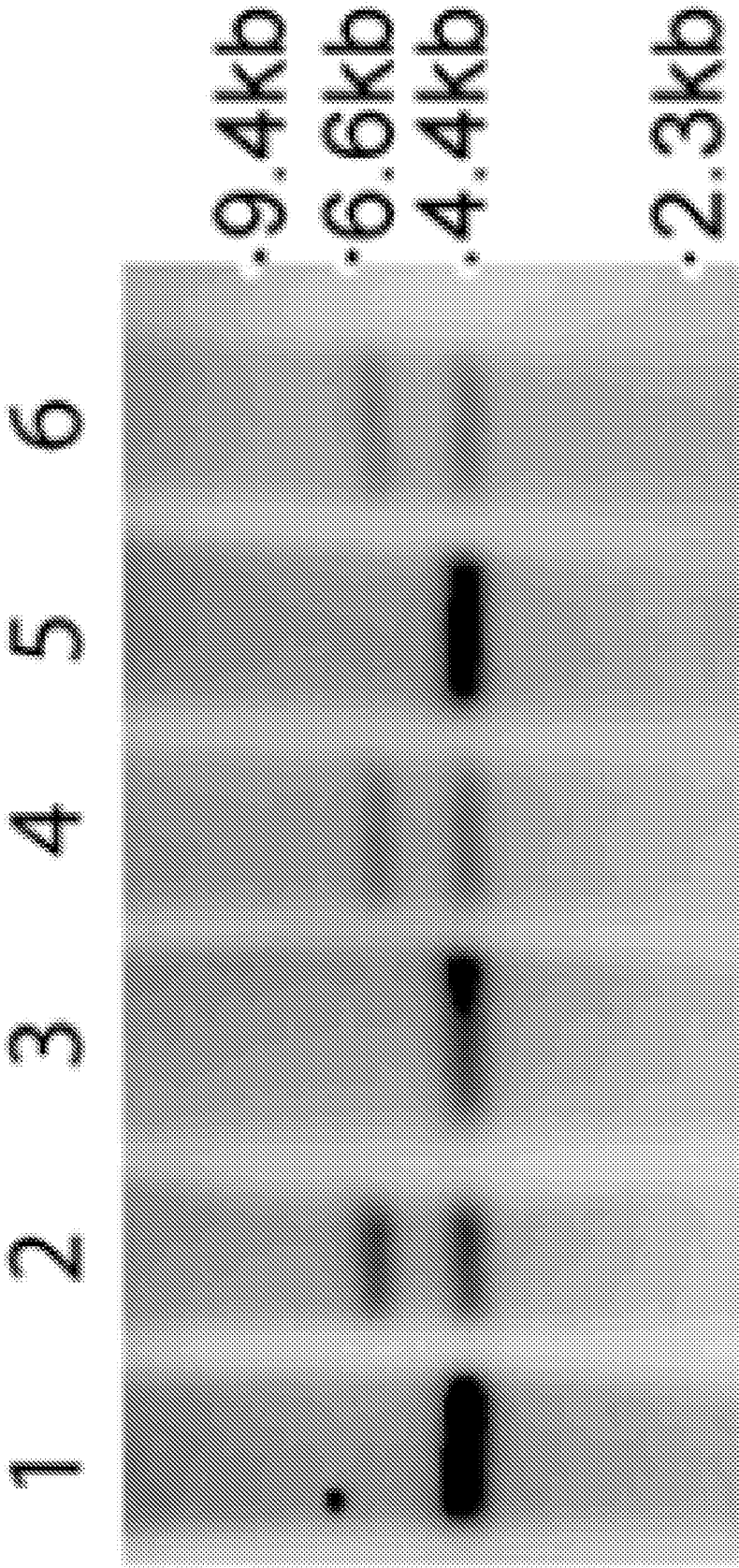


Figure 16D

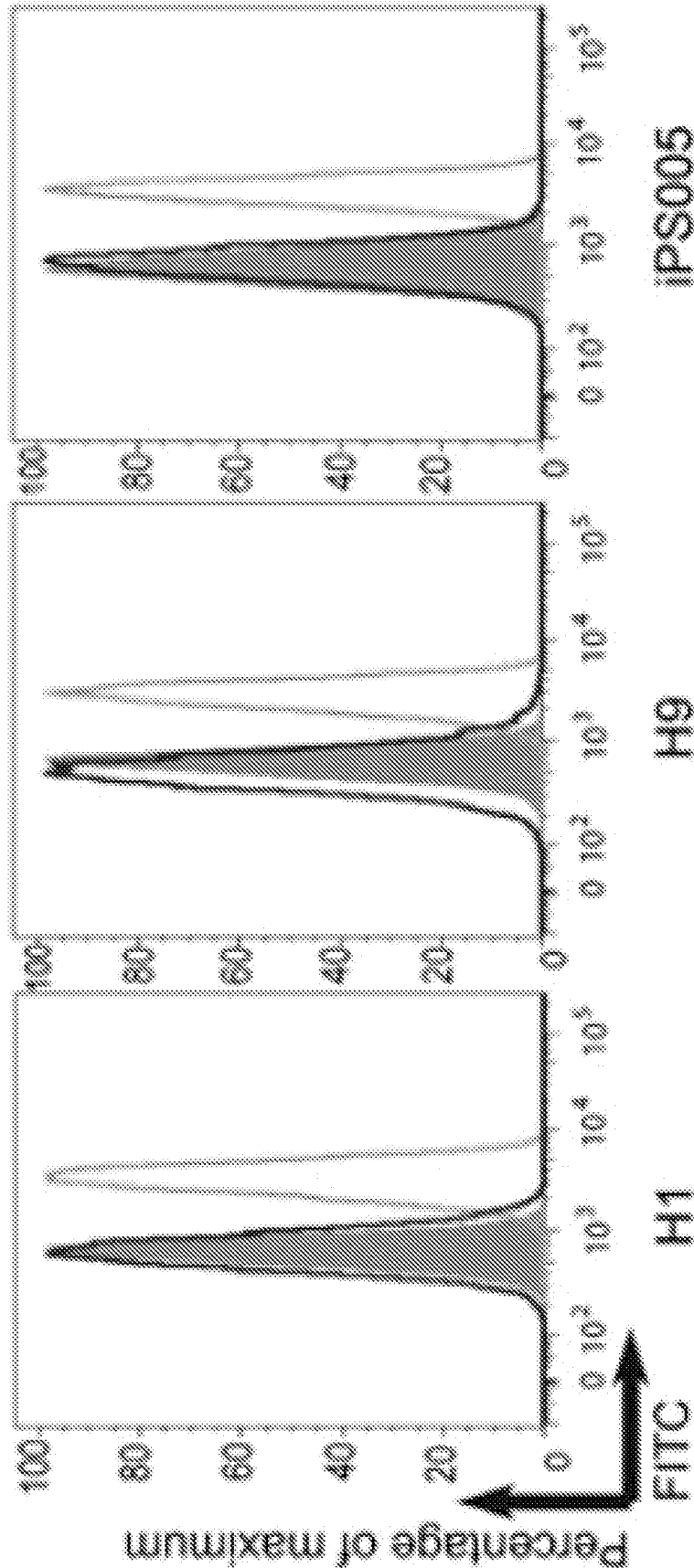


Figure 17A

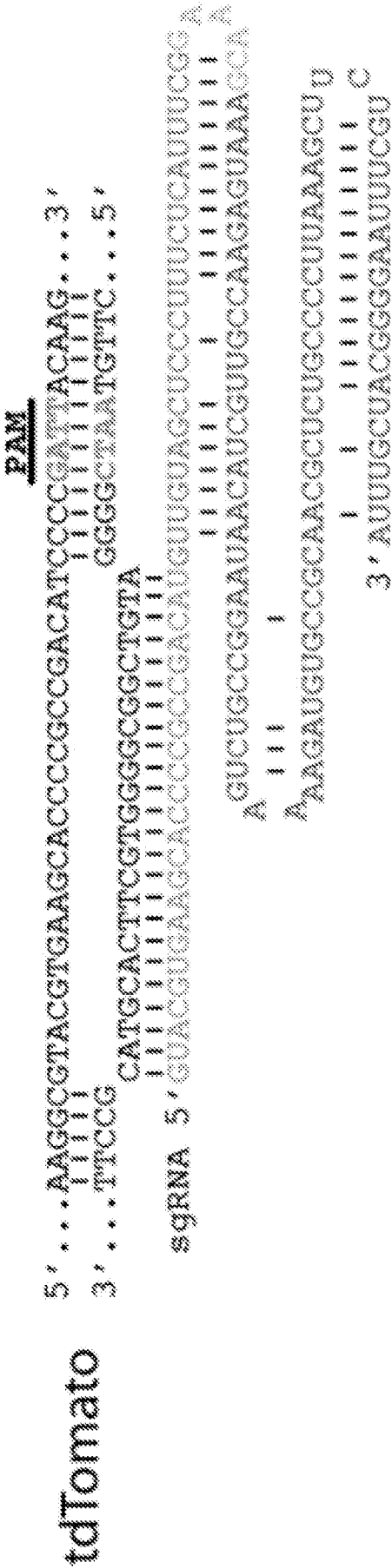


Figure 17B

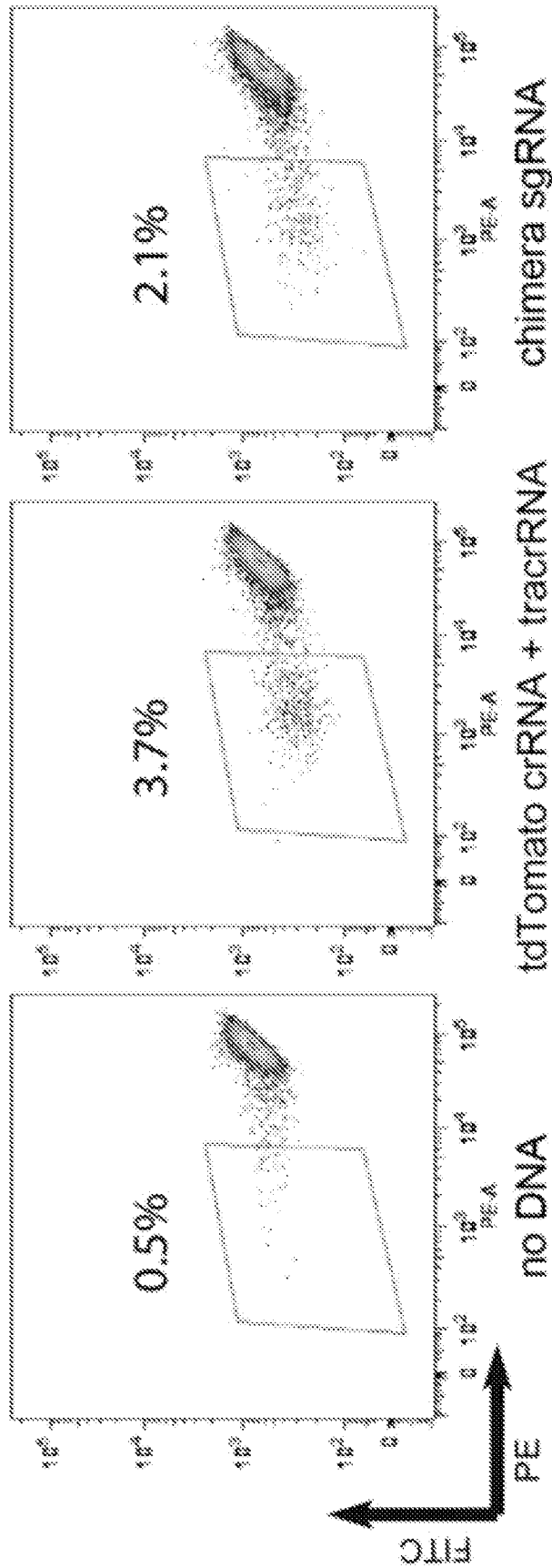


Figure 18

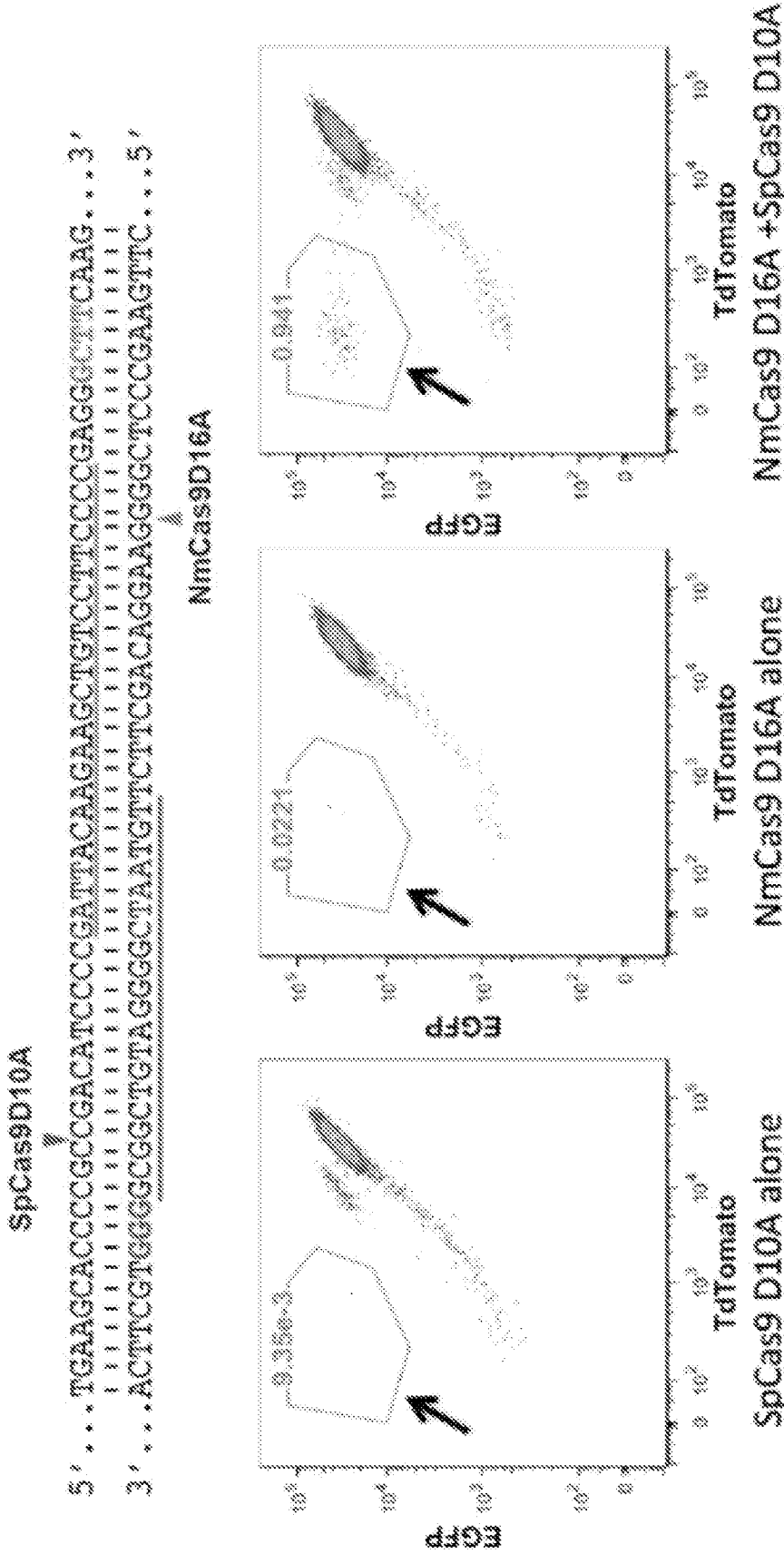
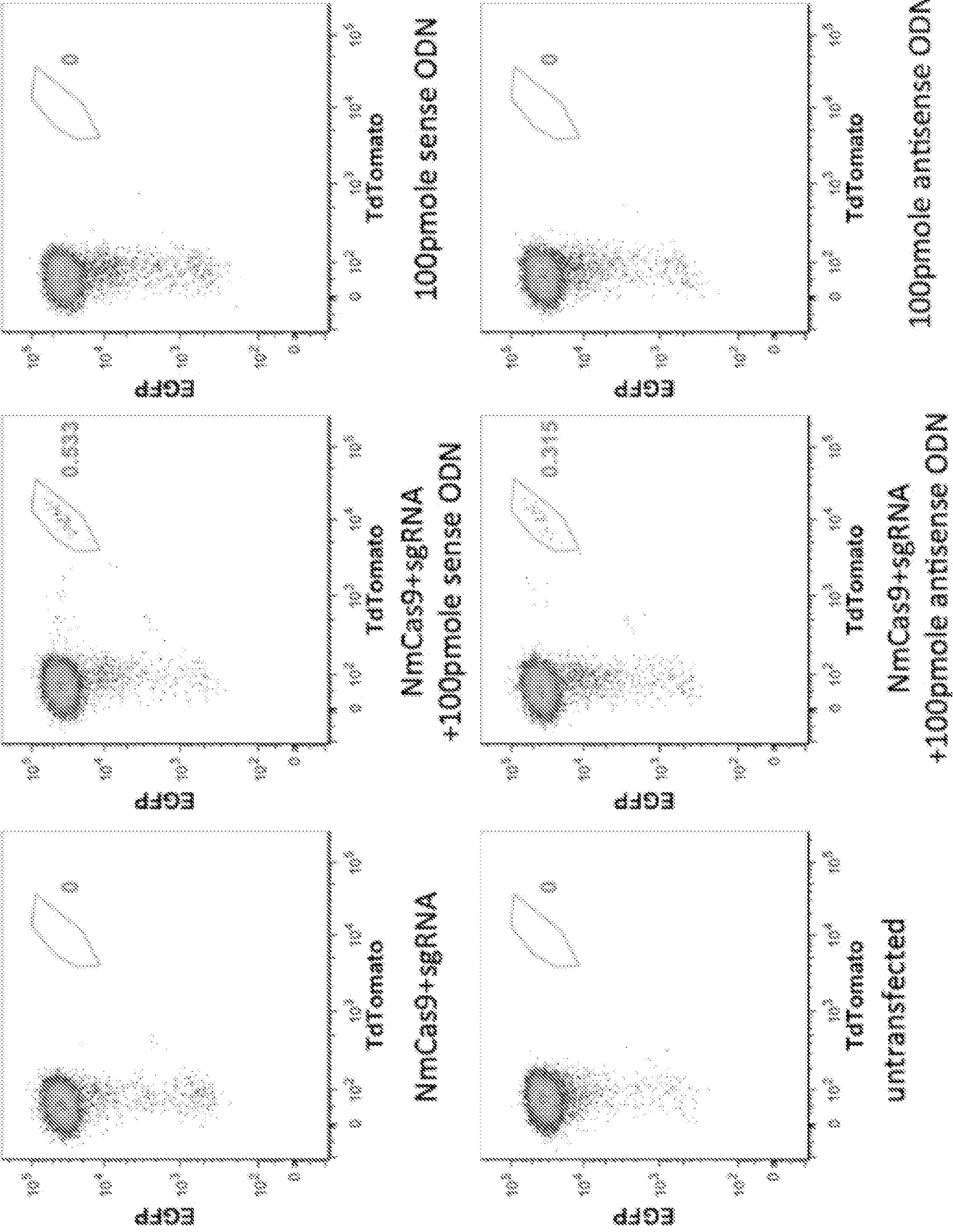


Figure 19



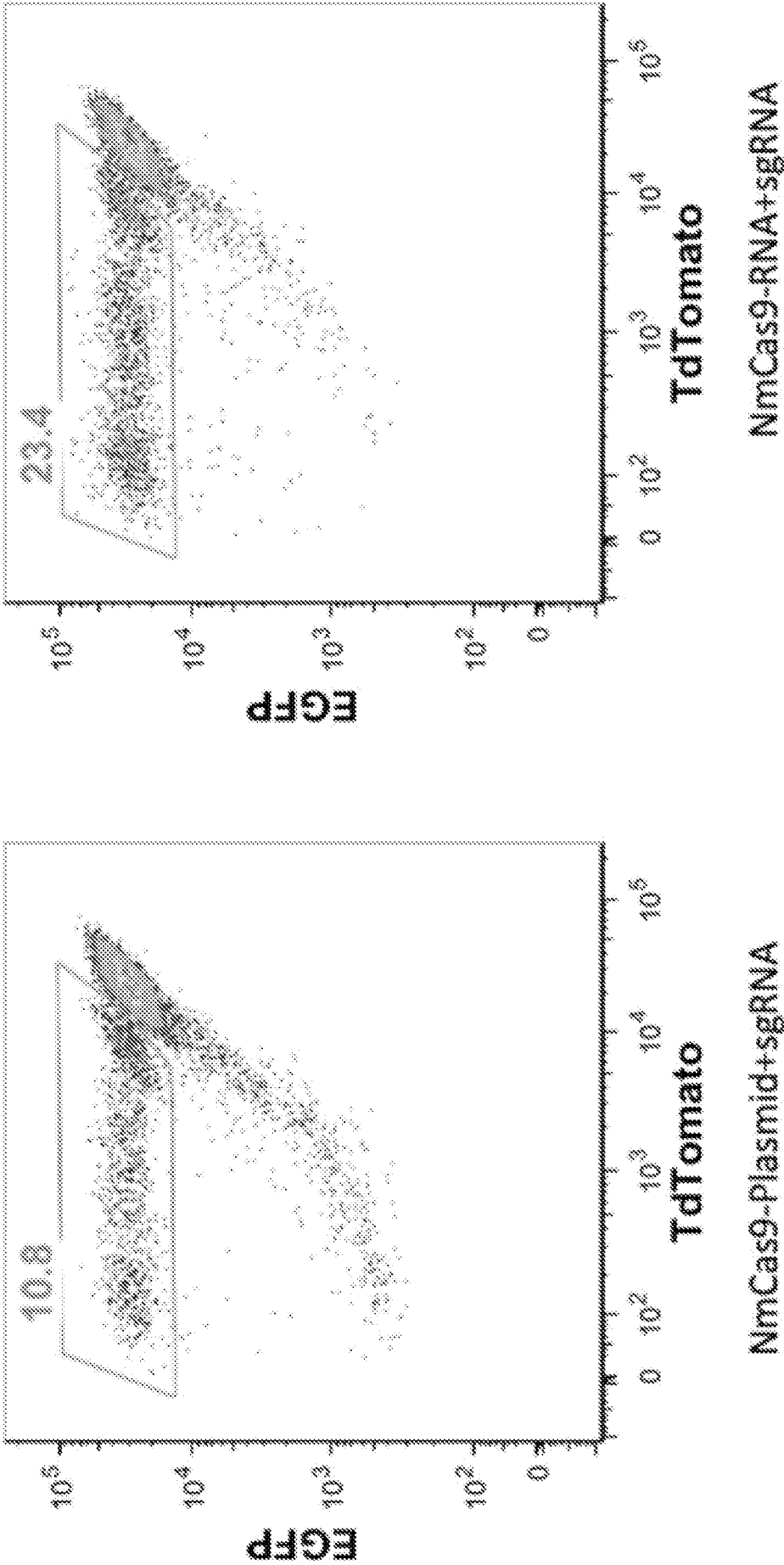


Figure 20

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 2014/039195

A. CLASSIFICATION OF SUBJECT MATTER		<i>C12N 15/66 (2006.01)</i> <i>C12N 9/22 (2006.01)</i> <i>C12N 15/55 (2006.01)</i> <i>C12N 5/10 (2006.01)</i>
According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED		
Minimum documentation searched (classification system followed by classification symbols)		
C12N 15/09, 15/63, 15/66 , 9/22, 15/55, 5/10		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)		
WIPO (Patentscope), DB Espacenet, PatFT & AppFT USPTO, RUPAT, EAPATIS, CIPO, PAJ, KIPRIS, K-PION, SIPO, DWPI, PubMed, Scopus		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	MALI P. et al. "RNA-guided human genome engineering via Cas9." Science, 15 February 2013, Vol. 339(6121): 823-826 and Supplementary materials, doi:10.1126/science.1232033, p.1-36	1-30
Y	SAMPSON T.R. et al. "A CRISPR/Cas system mediates bacterial innate immune evasion and virulence." Nature , 9 May 2013, Vol.497(7448),p. 254-258 and Supplementary material; doi:10.1038/nature12048	1-30
Y	GASIUNAS G. et al. "Cas9 ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria." Proceedings of the National Academy of Sciences, 2012; Vol.109, no.39: E2579-E2586; doi:10.1073/pnas.1208507109	13,18
Y	PAIGE J.S. et al. "RNA mimics of green fluorescent protein." Science, 2011, Vol. 333,p. 642-646; doi:10.1126/science.1207339	21
☒ Further documents are listed in the continuation of Box C. ☐ See patent family annex.		
* Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier document but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family		
Date of the actual completion of the international search		Date of mailing of the international search report
06 August 2014 (06.08.2014)		11 September 2014 (11.09.2014)
Name and mailing address of the ISA/RU: FIPS, Russia, 123995, Moscow, G-59, GSP-5, Berezhkovskaya nab., 30-1 Facsimile No. +7 (499) 243-33-37		Authorized officer T.Babakova Telephone No. 495 531 65 15

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 2014/039195

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	BELETSKII A. et al. "PNA interference mapping demonstrates functional domains in the noncoding RNA Xist." Proceedings of the National Academy of Sciences, 2001, Vol. 98, no.16, p. 9215-9220; doi:10.1073/pnas.161173098	22
A	WO 2010/075424 A2 (UNIV CALIFORNIA) 01.07.2010	1-30