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

Innovation, Science and Economic Development Canada
Canadian Intellectual Property Office

CA 2915842 C 2022/11/29

(11)(21) 2 915 842

(12) BREVET CANADIEN
CANADIAN PATENT

(13) C

(86) Date de dépôt PCT/PCT Filing Date: 2014/06/10
(87) Date publication PCT/PCT Publication Date: 2014/12/24
(45) Date de délivrance/Issue Date: 2022/11/29
(85) Entrée phase nationale/National Entry: 2015/12/16
(86) N° demande PCT/PCT Application No.: US 2014/041804
(87) N° publication PCT/PCT Publication No.: 2014/204726
(30) Priorités/Priorities: 2013/06/17 (US61/836,123); 2013/07/17 (US61/847,537); 2013/08/05 (US61/862,355); 2013/08/28 (US61/871,301); 2013/12/12 (US PCT/US2013/0074667); 2013/12/12 (US61/915,325); 2014/04/15 (US61/979,733)

(51) Cl.Int./Int.Cl. *C12N 15/63* (2006.01), *C12N 15/10* (2006.01), *C12N 9/22* (2006.01), *C40B 40/08* (2006.01)

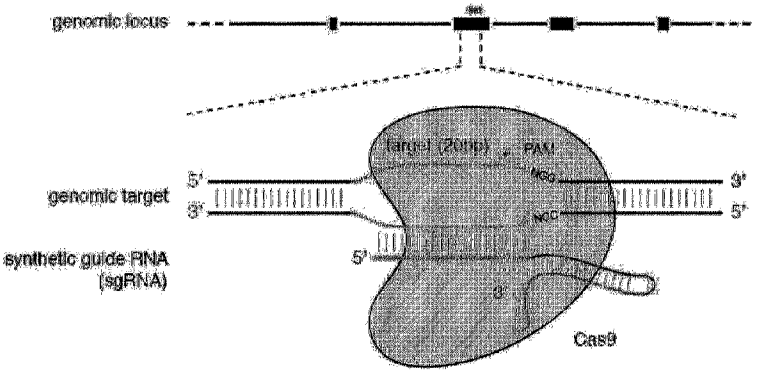
(72) Inventeurs/Inventors:
ZHANG, FENG, US;
CONG, LE, US;
RAN, FEI, US

(73) Propriétaires/Owners:
THE BROAD INSTITUTE, INC., US;
MASSACHUSETTS INSTITUTE OF TECHNOLOGY, US;
PRESIDENT AND FELLOWS OF HARVARD COLLEGE, US

(74) Agent: MBM INTELLECTUAL PROPERTY LAW LLP

(54) Titre : ADMINISTRATION ET UTILISATION DE SYSTEMES CRISPR-CAS, VECTEURS ET COMPOSITIONS POUR LE CIBLAGE ET LE TRAITEMENT DU FOIE

(54) Title: DELIVERY AND USE OF THE CRISPR-CAS SYSTEMS, VECTORS AND COMPOSITIONS FOR HEPATIC TARGETING AND THERAPY



(57) Abrégé/Abstract:
The invention provides for delivery, engineering and optimization of systems, methods, and compositions for manipulation of sequences and/or activities of target sequences. Provided are deliver)' systems and tissues or organ which are targeted as sites for delivery. Also provided are vectors and vector systems some of which encode one or more components of a CRISPR complex, as well as methods for the design and use of such vectors. Also provided are methods of directing CRISPR complex formation in eukaryotic cells to ensure enhanced specificity for target recognition and avoidance of toxicity and to edit or modify a target site in a genomic locus of interest to alter or improve the status of a disease or a condition.

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

CORRECTED VERSION

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
24 December 2014 (24.12.2014)

WIPO | PCT

(10) International Publication Number
WO 2014/204726 A9

(51) International Patent Classification:
C12N 15/63 (2006.01) *C40B 40/08* (2006.01)
C12N 15/10 (2006.01) *C12N 9/22* (2006.01)

(21) International Application Number:
PCT/US2014/041804

(22) International Filing Date:
10 June 2014 (10.06.2014)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

61/836,123	17 June 2013 (17.06.2013)	US
61/847,537	17 July 2013 (17.07.2013)	US
61/862,355	5 August 2013 (05.08.2013)	US
61/871,301	28 August 2013 (28.08.2013)	US
PCT/US2013/0074667	12 December 2013 (12.12.2013)	US
61/915,325	12 December 2013 (12.12.2013)	US
61/979,733	15 April 2014 (15.04.2014)	US

(71) Applicants: **THE BROAD INSTITUTE INC.** [US/US]; 7 Cambridge Center, Cambridge, MA 02142 (US). **MASSACHUSETTS INSTITUTE OF TECHNOLOGY** [US/US]; 77 Massachusetts Ave., Cambridge, MA 02142 (US).

(72) Inventors; and
(71) Applicants : **ZHANG, Feng** [US/US]; 100 Pacific Street, Apt. 11, Cambridge, MA 02139 (US). **CONG, Le** [CN/US]; 375 Harvard Street, Apt. 14, Cambridge, MA 02138 (US). **RAN, Fei** [US/US]; 129 St. Botolph Street, Apt. 2, Boston, MA 02115 (US).

(74) Agents: **LU, Deborah, L.** et al.; Vedder Price P.C., 1633 Broadway, New York, NY 10019 (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).

[Continued on next page]

(54) Title: DELIVERY AND USE OF THE CRISPR-CAS SYSTEMS, VECTORS AND COMPOSITIONS FOR HEPATIC TARGETING AND THERAPY

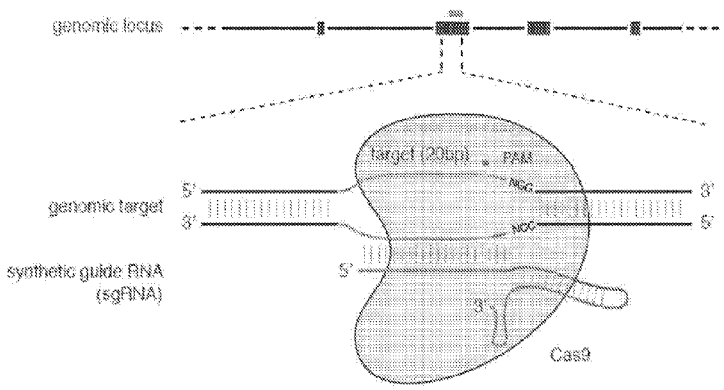


FIG. 1

(57) Abstract: The invention provides for delivery, engineering and optimization of systems, methods, and compositions for manipulation of sequences and/or activities of target sequences. Provided are deliver)' systems and tissues or organ which are targeted as sites for delivery. Also provided are vectors and vector systems some of which encode one or more components of a CRISPR complex, as well as methods for the design and use of such vectors. Also provided are methods of directing CRISPR complex formation in eukaryotic cells to ensure enhanced specificity for target recognition and avoidance of toxicity and to edit or modify a target site in a genomic locus of interest to alter or improve the status of a disease or a condition.

WO 2014/204726 A9

WO 2014/204726 A9

Published:

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

(48) Date of publication of this corrected version:
5 March 2015

(15) Information about Correction:
see Notice of 5 March 2015

DEMANDE OU BREVET VOLUMINEUX

LA PRÉSENTE PARTIE DE CETTE DEMANDE OU CE BREVET COMPREND PLUS D'UN TOME.

CECI EST LE TOME 1 DE 6
CONTENANT LES PAGES 1 À 72

NOTE : Pour les tomes additionels, veuillez contacter le Bureau canadien des brevets

JUMBO APPLICATIONS/PATENTS

THIS SECTION OF THE APPLICATION/PATENT CONTAINS MORE THAN ONE VOLUME

THIS IS VOLUME 1 OF 6
CONTAINING PAGES 1 TO 72

NOTE: For additional volumes, please contact the Canadian Patent Office

NOM DU FICHIER / FILE NAME :

NOTE POUR LE TOME / VOLUME NOTE:

DELIVERY AND USE OF THE CRISPR-CAS SYSTEMS, VECTORS AND COMPOSITIONS FOR HEPATIC TARGETING AND THERAPY

[0001]

[0002]

FIELD OF THE INVENTION

[0003] The present invention generally relates to the delivery, engineering, optimization and therapeutic applications of systems, methods, and compositions used for the control of gene expression involving sequence targeting, such as genome perturbation or gene-editing, that relate to Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) and components thereof. In particular, the present invention relates to aspects related to delivery to the liver, for gene therapy of liver conditions, understanding liver or liver tissue gene function and the creation of liver models. Liver or liver tissue includes parenchymal cells commonly referred to as hepatocytes. Liver or Liver tissue can also be liver cells that are non-parenchymal cells, especially as such cells constitute 40% of the total number of liver cells even though only 6.5%

Date Reçue/Date Received 2020-10-05

of its volume; and, examples of such non-parenchymal cells liver cells or tissue include sinusoidal hepatic endothelial cells, Kupffer cells and hepatic stellate cells. Cells of the liver express one or more liver gene product(s). Advantageously the invention is practiced with respect to hepatocytes or liver or liver tissue comprising hepatocytes.

STATEMENT AS TO FEDERALLY SPONSORED RESEARCH

[0004] This invention was made with government support under the NIH Pioneer Award (1DP1MH100706) awarded by the National Institutes of Health. The government has certain rights in the invention.

BACKGROUND OF THE INVENTION

[0005] Recent advances in genome sequencing techniques and analysis methods have significantly accelerated the ability to catalog and map genetic factors associated with a diverse range of biological functions and diseases. Precise genome targeting technologies are needed to enable systematic reverse engineering of causal genetic variations by allowing selective perturbation of individual genetic elements, as well as to advance synthetic biology, biotechnological, and medical applications. Although genome-editing techniques such as designer zinc fingers, transcription activator-like effectors (TALEs), or homing meganucleases are available for producing targeted genome perturbations, there remains a need for new genome engineering technologies that are affordable, easy to set up, scalable, and amenable to targeting multiple positions within the eukaryotic genome.

SUMMARY OF THE INVENTION

[0006] The CRISPR-Cas system does not require the generation of customized proteins to target specific sequences but rather a single Cas enzyme can be programmed by a short RNA molecule to recognize a specific DNA target. Adding the CRISPR-Cas system to the repertoire of genome sequencing techniques and analysis methods may significantly simplify the methodology and accelerate the ability to catalog and map genetic factors associated with a diverse range of biological functions and diseases. To utilize the CRISPR-Cas system effectively for genome editing without deleterious effects, it is critical to understand aspects of engineering, optimization and cell-type/tissue/organ specific delivery of these genome engineering tools, which are aspects of the claimed invention.

Date Reçue/Date Received 2020-10-05

[0007] There exists a pressing need for alternative and robust systems and techniques for nucleic acid sequence targeting with a wide array of applications. Aspects of this invention address this need and provide related advantages. An exemplary CRISPR complex comprises a CRISPR enzyme complexed with a guide sequence hybridized or hybridizable to a target sequence within the target polynucleotide. The guide sequence is linked to a tracr mate sequence, which in turn hybridizes to a tracr sequence.

[0008] In one aspect, the invention provides methods for using one or more elements of a CRISPR-Cas system. The CRISPR complex of the invention provides an effective means for modifying a target polynucleotide. The CRISPR complex of the invention has a wide variety of utilities including modifying (e.g., deleting, inserting, translocating, inactivating, activating) a target polynucleotide in a multiplicity of cell types in various tissues and organs. As such the CRISPR complex of the invention has a broad spectrum of applications in, e.g., gene or genome editing, gene therapy, drug discovery, drug screening, disease diagnosis, and prognosis. *In vivo*, *in vitro* and *ex vivo* uses are envisaged.

[0009] Aspects of the invention relate to Cas9 enzymes having improved liver-targeting specificity in a CRISPR-Cas9 system having guide RNAs having optimal activity, smaller in length than wild-type Cas9 enzymes and nucleic acid molecules coding therefor, and chimeric Cas9 enzymes, as well as methods of improving the targeting specificity of a Cas9 enzyme or of designing a CRISPR-Cas9 system comprising designing or preparing guide RNAs having optimal activity and/or selecting or preparing a Cas9 enzyme having a smaller size or length than wild-type Cas9 whereby packaging a nucleic acid coding therefor into a delivery vector is more advanced as there is less coding therefor in the delivery vector than for wild-type Cas9, and/or generating chimeric Cas9 enzymes.

[0010] Also provided are uses of the present sequences, vectors, enzymes or systems, in medicine. Also provided are uses of the same in gene or genome editing. This is in relation to liver tissues or cells, whether *in* or *ex vivo*,

[0011] In an additional aspect of the invention, a Cas9 enzyme may comprise one or more mutations and may be used as a generic DNA binding protein with or without fusion to a functional domain. The mutations may be artificially introduced mutations or gain- or loss-of-function mutations. The mutations may include but are not limited to mutations in one of the catalytic domains (D10 and H840) in the RuvC and HNH catalytic domains, respectively.

Date Reçue/Date Received 2020-10-05

Further mutations have been characterized and may be used in one or more compositions of the invention. In one aspect of the invention, the mutated Cas9 enzyme may be fused to a protein domain, e.g., such as a transcriptional activation domain. In one aspect, of the invention, the transcriptional activation domain may be VP64. In other aspects of the invention, the transcriptional repressor domain may be KRAB or SID4X. Other aspects of the invention relate to the mutated Cas 9 enzyme being fused to domains which include but are not limited to a transcriptional activator, repressor, a recombinase, a transposase, a histone remodeler, a demethylase, a DNA methyltransferase, a cryptochrome, a light inducible/controllable domain or a chemically inducible/controllable domain.

[0012] In a further embodiment, the invention provides for methods to generate mutant tracrRNA and direct repeat sequences or mutant chimeric guide sequences that allow for enhancing performance of these RNAs in cells. Aspects of the invention also provide for selection of said sequences.

[0013] Aspects of the invention also provide for methods of simplifying the cloning and delivery of components of the CRISPR complex. In the preferred embodiment of the invention, a suitable promoter, such as the U6 promoter, is amplified with a DNA oligo and added onto the guide RNA. The resulting PCR product can then be transfected into cells to drive expression of the guide RNA. Aspects of the invention also relate to the guide RNA being transcribed *in vitro* or ordered from a synthesis company and directly transfected.

[0014] In one aspect, the invention provides for methods to improve activity by using a more active polymerase. In a preferred embodiment, the expression of guide RNAs under the control of the T7 promoter is driven by the expression of the T7 polymerase in the cell. In an advantageous embodiment, the cell is a eukaryotic cell. In a preferred embodiment the eukaryotic cell is a human cell. In a more preferred embodiment the human cell is a patient specific cell.

[0015] In one aspect, the invention provides for methods of reducing the toxicity of Cas enzymes. In certain aspects, the Cas enzyme is any Cas9 as described herein, for instance any naturally-occurring bacterial Cas9 as well as any chimaeras, mutants, homologs or orthologs. In a preferred embodiment, the Cas9 is delivered into the cell in the form of mRNA. This allows for the transient expression of the enzyme thereby reducing toxicity. In another preferred

Date Reçue/Date Received 2020-10-05

embodiment, the invention also provides for methods of expressing Cas9 under the control of an inducible promoter, and the constructs used therein.

[0016] In another aspect, the invention provides for methods of improving the *in vivo* applications of the CRISPR-Cas system. In the preferred embodiment, the Cas enzyme is wildtype Cas9 or any of the modified versions described herein, including any naturally-occurring bacterial Cas9 as well as any chimaeras, mutants, homologs or orthologs. An advantageous aspect of the invention provides for the selection of Cas9 homologs that are easily packaged into viral vectors for delivery. Cas9 orthologs typically share the general organization of 3-4 RuvC domains and a HNH domain. The 5' most RuvC domain cleaves the non-complementary strand, and the HNH domain cleaves the complementary strand. All notations are in reference to the guide sequence.

[0017] The catalytic residue in the 5' RuvC domain is identified through homology comparison of the Cas9 of interest with other Cas9 orthologs (from S. pyogenes type II CRISPR locus, S. thermophilus CRISPR locus 1, S. thermophilus CRISPR locus 3, and Franciscilla novicida type II CRISPR locus), and the conserved Asp residue (D10) is mutated to alanine to convert Cas9 into a complementary-strand nicking enzyme. Similarly, the conserved His and Asn residues in the HNH domains are mutated to Alanine to convert Cas9 into a non-complementary-strand nicking enzyme. In some embodiments, both sets of mutations may be made, to convert Cas9 into a non-cutting enzyme.

[0018] In some embodiments, the CRISPR enzyme is a type I or III CRISPR enzyme, preferably a type II CRISPR enzyme. This type II CRISPR enzyme may be any Cas enzyme. A preferred Cas enzyme may be identified as Cas9 as this can refer to the general class of enzymes that share homology to the biggest nuclease with multiple nuclease domains from the type II CRISPR system. Most preferably, the Cas9 enzyme is from, or is derived from, spCas9 or saCas9. By derived, Applicants mean that the derived enzyme is largely based, in the sense of having a high degree of sequence homology with, a wildtype enzyme, but that it has been mutated (modified) in some way as described herein.

[0019] It will be appreciated that the terms Cas and CRISPR enzyme are generally used herein interchangeably, unless otherwise apparent. As mentioned above, many of the residue numberings used herein refer to the Cas9 enzyme from the type II CRISPR locus in *Streptococcus pyogenes*. However, it will be appreciated that this invention includes many more

Date Reçue/Date Received 2020-10-05

Cas9s from other species of microbes, such as SpCas9, SaCas9, St1Cas9 and so forth. Further examples are provided herein. The skilled person will be able to determine appropriate corresponding residues in Cas9 enzymes other than SpCas9 by comparison of the relevant amino acid sequences. Thus, where a specific amino acid replacement is referred to using the SpCas9 numbering, then, unless the context makes it apparent this is not intended to refer to other Cas9 enzymes, the disclosure is intended to encompass corresponding modifications in other Cas9 enzymes. SaCas9 is particularly preferred.

[0020] An example of a codon optimized sequence, in this instance optimized for humans (i.e. being optimized for expression in humans) is provided herein, e.g., see the SaCas9 human codon optimized sequence. Whilst this is preferred, it will be appreciated that other examples are possible, and codon optimization for a host species other than human, or for codon optimization for specific organs such as the brain, is known.

[0021] In further embodiments, the invention provides for methods of enhancing the function of Cas9 by generating chimeric Cas9 proteins. Chimeric Cas9 proteins chimeric Cas9s may be new Cas9 containing fragments from more than one naturally occurring Cas9. These methods may comprise fusing N-terminal fragments of one Cas9 homolog with C-terminal fragments of another Cas9 homolog. These methods also allow for the selection of new properties displayed by the chimeric Cas9 proteins.

[0022] It will be appreciated that in the present methods, where the organism is an animal or a plant, the modification may occur *ex vivo* or *in vitro*, for instance in a cell culture and in some instances not *in vivo*. In other embodiments, it may occur *in vivo*.

[0023] In one aspect, the invention provides a method of modifying an organism or a non-human organism by manipulation of a target sequence in a genomic locus of interest comprising: delivering a non-naturally occurring or engineered composition comprising :

A) - I. a CRISPR-Cas system chimeric RNA (chiRNA) polynucleotide sequence, wherein the polynucleotide sequence comprises:

(a) a guide sequence capable of hybridizing to a target sequence in a eukaryotic cell,

(b) a tracr mate sequence, and

(c) a tracr sequence, and

II. a polynucleotide sequence encoding a CRISPR enzyme comprising at least one or more nuclear localization sequences,

Date Reçue/Date Received 2020-10-05

wherein (a), (b) and (c) are arranged in a 5' to 3' orientation,

wherein when transcribed, the tracr mate sequence hybridizes to the tracr sequence and the guide sequence directs sequence-specific binding of a CRISPR complex to the target sequence, and

wherein the CRISPR complex comprises the CRISPR enzyme complexed with (1) the guide sequence that is hybridized or hybridizable to the target sequence, and (2) the tracr mate sequence that is hybridized or hybridizable to the tracr sequence and the polynucleotide sequence encoding a CRISPR enzyme is DNA or RNA,

or

(B) I. polynucleotides comprising:

(a) a guide sequence capable of hybridizing to a target sequence in a eukaryotic cell, and

(b) at least one or more tracr mate sequences,

II. a polynucleotide sequence encoding a CRISPR enzyme, and

III. a polynucleotide sequence comprising a tracr sequence,

wherein when transcribed, the tracr mate sequence hybridizes to the tracr sequence and the guide sequence directs sequence-specific binding of a CRISPR complex to the target sequence, and

wherein the CRISPR complex comprises the CRISPR enzyme complexed with (1) the guide sequence that is hybridized or hybridizable to the target sequence, and (2) the tracr mate sequence that is hybridized or hybridizable to the tracr sequence, and the polynucleotide sequence encoding a CRISPR enzyme is DNA or RNA.

[0024] In some embodiments, the second alternative above is preferred. The first alternative is particularly preferred, however, in most but not all aspects of the disclosure.

[0025] It will be appreciated that the present application is directed to the liver, whether that is the organ per se or a tissue within it or simply one or more liver cells, e.g., hepatocytes. Primary hepatocytes are preferred. The liver cells may be comprised within a vertebrate animal, either a patient (in the sense of an animal in need of CRISPR-directed gene therapy) or a model organism, or may be in cell culture , an organoid or other ex vivo tissue, such a "liver on a chip" for instance where hepatocytes are seeded and grown on a scaffold. Harvested hepatocytes from un-transplanted organs are also a useful target. With the development of 3-D printing techniques being applied to biology, printed tissues are within grasp and it is entirely feasible that liver cells or tissues printed in this way to create an organoid or onto a chip could also be targeted.

Date Reçue/Date Received 2020-10-05

[0026] Thus, provided is a model organism comprising liver cells, such as hepatocytes, to which the present CRISPR-Cas system has been delivered. Similarly, also provided is an *ex vivo* collection of two or more liver cells, such as hepatocytes, to which the present CRISPR-Cas system has been delivered. Such collections may include liver organs, liver organoids, liver cells populating a scaffold (as.g., such as 'liver on a chip'). Methods of creating such models or colelctions are also provided.

[0027] In particular, such liver cells may express, or may comprise polynucleotides capable of expressing, a Cas enzyme. As discussed herein, this has the advantage of providing a ready model for interrogating gene function through gene perturbation, including knock down. This is particularly useful in studying conditions of the liver, such as amyloidosis and others those listed herein, as well as broader conditions such as obesity, where liver is only one of the affect components in the body.

[0028] Methods of interrogating liver gene function are also provided herein. These typically comprise delivering to liver cells, either *in* or *ex vivo*, the CRISPR-Cas system. However, if the cells already comprise Cas, whether expressed as a protein or encoded by polynucleotides already comprised within the cells, then only the CRISPR polynucleotide needs to be delivered. The method may include extraction from and, optionally, re-insertion back into the liver. By delivering, it is meant actually physical delivery of the polynucleotides to the nucleus of the cell, but also transfection. Therefore, delivery should also be read as including transfection unless otherwise apparent.

[0029] Also provided is a method of inducing gene perturbation in one or more liver cells, comprising transducing a first population of cells with a CRISPR-Cas system according to the present invention to thereby alter the genome of the first population of cells to obtain a second population of cells. The method may be *ex vivo* or *in vitro*, for instance in a cell culture or in an *ex vivo* or *in vitro* model (such as an organoid or 'liver on a chip'). Alternatively, the method may be *in vivo*, in which case it may also include isolating the first population of cells from the subject, and transplanting the second population of cells (back) into the subject. Gene perturbation may be for one or more, or two or more, or three or more, or four or more genes. The gene perturbation may be a reduction in gene function (i.e. activity in the encoded gene product). This may be, for instance, induced through alteration of the genome of the first population of cells to obtain the second population of cells, wherein the second population of

Date Reçue/Date Received 2020-10-05

cells has a defective genotype, such as a monogenic condition, which is absent in the first population of cells. This may require a corresponding repair template , as discussed herein, to provide the defective sequence or it may be through induction of a DSB. In particular, the gene perturbation is a gene knockdown.

[0030] Alternatively, the gene perturbation may be an increase in gene function (i.e. activity in the encoded gene product). This may be, for instance, induced through alteration of the genome of the first population of cells to obtain the second population of cells, wherein the first population of cells has a defective genotype, such as a monogenic condition, which is absent in (i.e. corrected for in) the second population of cells. This may require a corresponding repair template, as discussed herein, to provide the corrected sequence.

[0031] If multiplexing is used, then a mixture of reduction of one or more genes and increase of one or more genes is envisaged. This may be achieved through provision of one or more of the guides (in the multiplex) and corresponding repair templates may be used to reduce function, whilst one or more of the guides and their corresponding templates may be used to increase function.

[0032] Also provided is a method of interrogating function of one or more genes in one or more liver cells, comprising determining changes in expression of the one or more genes in the first populations of liver cells, inducing said gene perturbation in said first population to provide said second population with an altered genome (or genotype), and determining changes in expression of the one or more genes in the second population of liver cells, thereby interrogating the function of the one or more genes.

[0033] Also provided is a model and a method of creating said model. The model may be an animal comprising a liver (an *in vivo* model) or it may be an *ex vivo* or *in vitro* model, such as a liver organoid or 'liver on a chip' or the collection of liver cells, such as a on a scaffold, as described herein. The liver cells of either model will preferably be transfected with Cas9. Accordingly, there is specifically provided a model comprising one or more liver cells comprising the CRISPR enzyme, preferably a Cas9 such as Sa or SpCas9. The model cells may have been transfected or transduced with the second regulatory element provided herein, which is second regulatory element operably linked to an enzyme-coding sequence encoding a CRISPR enzyme comprising at list one or more nuclear localization sequences (NLSs). The model may be, as described above, an *in vivo* model or it may be an *ex vivo* or *in vitro* model. Such a model

Date Reçue/Date Received 2020-10-05

allows rapid interrogation of function of one or more genes, as only the CRISPR-Cas system polynucleotide sequence (comprising one or more guide sequences targeting said one or more genes) needs to be delivered to perturb the function of said gene. In other words, methods of interrogating gene function in such models may comprise only delivery of the CRISPR-Cas system polynucleotide sequence (comprising the one or more guide sequences), the Cas (CRISPR enzyme) having already been provided in the cell(s) of the model. Methods of creating such models are also provided, comprising transducing or transfecting one or more liver cells in a first population of liver cells with a second regulatory element operably linked to an enzyme-coding sequence encoding a CRISPR enzyme comprising at list one or more nuclear localization sequences (NLSs) as described herein to thereby provide one or more second population of liver cells comprising or expressing the CRISPR enzyme.

[0034] Methods of creating gene perturbed models, in particular gene knock down models, are also provided. These methods may typically comprise inducing gene perturbation in one or more genes, as described herein, in a first population of cells to thereby provide a second population of cells with an altered genome (or genotype). The second population of cells may then be seeded into a scaffold or onto a chip, for instance, to thereby provide an *ex vivo* or *in vitro* model. Alternatively, the second population of may be comprised within an *in vivo* animal.

[0035] Methods of gene therapy are also envisaged. For instance, correction of one or more deficient genotypes (for example single point mutations) is achievable through the use of the present CRISPR-Cas system in the liver cells discussed herein (including the models). Monogenic conditions associated with the liver are particularly preferred and are exemplified herein, see Example 38 where the CRISPR-Cas9 system target was ApoB, a lipid metabolism gene, was effective at inducing a phenotypic change in vivo. Compositons for use in gene therapy are also provided.

[0036] Although various Cas enzymes are envisaged, Cas9 is particularly preferred and we have shown particular efficacy in the liver for SaCa9. Tracr sequence from Sa is also preferred if the Cas enzyme is an Sa Cas enzyme. A suitable PAM in such circumstance is NNGRR. For S. pyogenes Cas9 or derived enzymes, a suitable PAM is 5'-NRG.

[0037] Although one guide may be used, so-called multiplexing with two, three, four or more guides, is particularly useful in interrogation of gene function and model creation (to provide

Date Reçue/Date Received 2020-10-05

multiple gene knock downs), but also in gene therapy where multiple defective genotypes are to be corrected (either multiple errors in a single gene or, more likely, multiple errors spread across several genes). Alternatively, multiplexing with two guides is useful in a dual nickase approach to reduce off-target effects or simply selection of multiple targets within one gene to ensure Cas recruitment. Triple and quadruple guides are preferred. Reference to gene herein is made interchangeably with genomic locus.

[0038] The intron approach described here is also useful in this regard, where the guide is positioned within the Cas intron.

[0039] Preferred means of delivery include the methods described by Kanasty below, such as LNP, especially where only the guide is to be delivered or it is to be delivered alone. However, viral vectors including lentiviral and AAV are generally preferred for the liver as they have been successful to date. Of these, AAV is preferred and especially serotype 8, with AAV2/8 shown to be effective.

[0040] Some preferred targets, to the extent that they are present in or conditions of the liver are metabolic disorders, such as any one of: Amyloid neuropathy (TTR, PALB); Amyloidosis (APOA1, APP, AAA, CVAP, AD1, GSN, FGA, LYZ, TTR, PALB); Cirrhosis (KRT18, KRT8, CIRH1A, NAIC, TEX292, KIAA1988); Cystic fibrosis (CFTR, ABCC7, CF, MRP7); Glycogen storage diseases (SLC2A2, GLUT2, G6PC, G6PT, G6PT1, GAA, LAMP2, LAMPB, AGL, GDE, GBE1, GYS2, PYGL, PFKM); Hepatic adenoma, 142330 (TCF1, HNF1A, MODY3), Hepatic failure, early onset, and neurologic disorder (SCOD1, SCO1), Hepatic lipase deficiency (LIPC), Hepatoblastoma, cancer and carcinomas (CTNNB1, PDGFRL, PDGRL, PRLTS, AXIN1, AXIN, CTNNB1, TP53, P53, LFS1, IGF2R, MPRI, MET, CASP8, MCH5; Medullary cystic kidney disease (UMOD, HNFJ, FJHN, MCKD2, ADMCKD2); Phenylketonuria (PAH, PKU1, QDPR, DHPR, PTS); Polycystic kidney and hepatic disease (FCYT, PKHD1, ARPKD, PKD1, PKD2, PKD4, PKDTS, PRKCSH, G19P1, PCLD, SEC63). Other preffered targets include any one or more of include one or more of: PCSK9; Hmgcr; SERPINA1; ApoB; and.or LDL.

[0041] It will be appreciated that methods of altering expression in the liver do not involve alteration of the germline, which may be excluded on moral grounds. In fact, although transfection of stem cells is envisage and certainly preferred in some embodiments, primary

Date Reçue/Date Received 2020-10-05

hepatocytes are particularly preferred, particularly where they may show or be stimualred to show some regeneration.

[0042] Type II CRISPRS are particulalrly preferred, especially for use in eukaryotes, as in the present case, where livers are only found in eukaryotes, particularly vertbertate anmimals, in any case.

[0043] Use of the CRISPR-Cas systems to invoke a phenotypic change is a particular advantage, especially *in vivo*. We have shown this in the present application.

[0044] Where therapeutic applications are envisaged, or for other genome engineering in the liver, then where a correction is required it will be appreciated that following nicking or cleavage of the genomic DNA target, then correction via the HDR pathway is preferred. For gene knockdown, NHEJ is advantageous, however, correction via the HDR pathway is preferred for therapy. In such circumstances, it is preferable to deliver a repair template. This is most preferably ssDNA although RNA via a retroviral vector to provide a corresponding DNA template is also possible. The skilled person can readily put the invention into practice from the herein teachings contributing to the knowledge in the art; and in this regard mention is made that the skilled person from the herein teachings contributing to the knowledge in the art can readily appreciate and implement considerations as to homologous arm length. Mention is made of patent applications and publications including herein inventor Zhang, including those cited herein. The repair template is preferably co-delivered with one or more elements of the CRISPR-Cas system.

[0045] Also provided is a method of altering expression of at least one liver gene product comprising introducing into a eukaryotic liver cell, for example a hepatocyte, containing and expressing a DNA molecule having a target sequence and encoding the gene product, an engineered, non-naturally occurring Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-CRISPR associated (Cas) (CRISPR-Cas) system comprising one or more vectors comprising:

a) a first regulatory element operable in a eukaryotic cell operably linked to at least one nucleotide sequence encoding a CRISPR-Cas system guide RNA that hybridizes with the target sequence, and

b) a second regulatory element operable in a eukaryotic cell operably linked to a nucleotide sequence encoding a Type-II Cas9 protein,

Date Reçue/Date Received 2020-10-05

wherein components (a) and (b) are located on same or different vectors of the system, whereby the guide RNA targets the target sequence and the Cas9 protein cleaves the DNA molecule, whereby expression of the at least one liver gene product is altered; and, wherein the Cas9 protein and the guide RNA do not naturally occur together.

[0046] Reference below to targets will be ubnderstood to be hepatic targets or genes otherwise exprerssed in the liver unless otherwise apparent.

[0047] Any or all of the polynucleotide sequence encoding a CRISPR enzyme, guide sequence, tracr mate sequence or tracr sequence, may be RNA. The polynucleotides comprising the sequence encoding a CRISPR enzyme, the guide sequence, tracr mate sequence or tracr sequence may be RNA and may be delivered via liposomes, nanoparticles, exosomes, microvesicles, or a gene-gun.

[0048] It will be appreciated that where reference is made to a polynucleotide, which is RNA and is said to 'comprise' a feature such a tracr mate sequence, the RNA sequence includes the feature. Where the polynucleotide is DNA and is said to comprise a feature such a tracr mate sequence, the DNA sequence is or can be transcribed into the RNA including the feature at issue. Where the feature is a protein, such as the CRISPR enzyme, the DNA or RNA sequence referred to is, or can be, translated (and in the case of DNA transcribed first).

[0049] Accordingly, in certain embodiments the invention provides a method of modifying the liver of an organism, e.g., mammal including human or a non-human mammal or organism by manipulation of a target sequence in a genomic locus of interest comprising delivering a non-naturally occurring or engineered composition comprising a viral or plasmid vector system comprising one or more viral or plasmid vectors operably encoding a composition for expression thereof, wherein the composition comprises: (A) a non-naturally occurring or engineered composition comprising a vector system comprising one or more vectors comprising I. a first regulatory element operably linked to a CRISPR-Cas system chimeric RNA (chiRNA) polynucleotide sequence, wherein the polynucleotide sequence comprises (a) a guide sequence capable of hybridizing to a target sequence in a eukaryotic cell, (b) a tracr mate sequence, and (c) a tracr sequence, and II. a second regulatory element operably linked to an enzyme-coding sequence encoding a CRISPR enzyme comprising at least one or more nuclear localization sequences (or optionally at least one or more nuclear localization sequences as some embodiments can involve no NLS), wherein (a), (b) and (c) are arranged in a 5' to 3' orientation,

Date Reçue/Date Received 2020-10-05

wherein components I and II are located on the same or different vectors of the system, wherein when transcribed, the tracr mate sequence hybridizes to the tracr sequence and the guide sequence directs sequence-specific binding of a CRISPR complex to the target sequence, and wherein the CRISPR complex comprises the CRISPR enzyme complexed with (1) the guide sequence that is hybridized or hybridizable to the target sequence, and (2) the tracr mate sequence that is hybridized or hybridizable to the tracr sequence, or (B) a non-naturally occurring or engineered composition comprising a vector system comprising one or more vectors comprising I. a first regulatory element operably linked to (a) a guide sequence capable of hybridizing to a target sequence in a eukaryotic cell, and (b) at least one or more tracr mate sequences, II. a second regulatory element operably linked to an enzyme-coding sequence encoding a CRISPR enzyme, and III. a third regulatory element operably linked to a tracr sequence, wherein components I, II and III are located on the same or different vectors of the system, wherein when transcribed, the tracr mate sequence hybridizes to the tracr sequence and the guide sequence directs sequence-specific binding of a CRISPR complex to the target sequence, and wherein the CRISPR complex comprises the CRISPR enzyme complexed with (1) the guide sequence that is hybridized or hybridizable to the target sequence, and (2) the tracr mate sequence that is hybridized or hybridizable to the tracr sequence. In some embodiments, components I, II and III are located on the same vector. In other embodiments, components I and II are located on the same vector, while component III is located on another vector. In other embodiments, components I and III are located on the same vector, while component II is located on another vector. In other embodiments, components II and III are located on the same vector, while component I is located on another vector. In other embodiments, each of components I, II and III is located on different vectors. The invention also provides a viral or plasmid vector system as described herein.

[0050] Preferably, the vector is a viral vector, such as a lenti- or baculo- or preferably adeno-viral/adeno-associated viral vectors, but other means of delivery are known (such as yeast systems, microvesicles, gene guns/means of attaching vectors to gold nanoparticles) and are provided. In some embodiments, one or more of the viral or plasmid vectors may be delivered via liposomes, nanoparticles, exosomes, microvesicles, or a gene-gun.

[0051] By manipulation of a target sequence, Applicants also mean the epigenetic manipulation of a target sequence. This may be of the chromatin state of a target sequence, such

Date Reçue/Date Received 2020-10-05

as by modification of the methylation state of the target sequence (i.e. addition or removal of methylation or methylation patterns or CpG islands), histone modification, increasing or reducing accessibility to the target sequence, or by promoting 3D folding.

[0052] It will be appreciated that where reference is made to a method of modifying an organism or mammal including human or a non-human mammal or organism by manipulation of a target sequence in a genomic locus of interest, this may apply to the organism (or mammal) as a whole or just a single cell or population of cells from that organism (if the organism is multicellular). In the case of humans, for instance, Applicants envisage, *inter alia*, a single cell or a population of cells and these may preferably be modified *ex vivo* and then re-introduced. In this case, a biopsy or other tissue or biological fluid sample may be necessary. Stem cells are also particularly preferred in this regard. But, of course, *in vivo* embodiments are also envisaged.

[0053] In certain embodiments the invention provides a method of treating or inhibiting a condition caused by a defect in a target sequence in a genomic locus of interest in a subject (e.g., mammal or human) or a non-human subject (e.g., mammal) in need thereof comprising modifying the subject or a non-human subject by manipulation of the target sequence and wherein the condition is susceptible to treatment or inhibition by manipulation of the target sequence comprising providing treatment comprising: delivering a non-naturally occurring or engineered composition comprising an AAV or lentivirus vector system comprising one or more AAV or lentivirus vectors operably encoding a composition for expression thereof, wherein the target sequence is manipulated by the composition when expressed, wherein the composition comprises: (A) a non-naturally occurring or engineered composition comprising a vector system comprising one or more vectors comprising I. a first regulatory element operably linked to a CRISPR-Cas system chimeric RNA (chiRNA) polynucleotide sequence, wherein the polynucleotide sequence comprises (a) a guide sequence capable of hybridizing to a target sequence in a eukaryotic cell, (b) a tracr mate sequence, and (c) a tracr sequence, and II. a second regulatory element operably linked to an enzyme-coding sequence encoding a CRISPR enzyme comprising at least one or more nuclear localization sequences (or optionally at least one or more nuclear localization sequences as some embodiments can involve no NLS) wherein (a), (b) and (c) are arranged in a 5' to 3' orientation, wherein components I and II are located on the same or different vectors of the system, wherein when transcribed, the tracr mate sequence hybridizes to the tracr sequence and the guide sequence directs sequence-specific binding of a

Date Reçue/Date Received 2020-10-05

CRISPR complex to the target sequence, and wherein the CRISPR complex comprises the CRISPR enzyme complexed with (1) the guide sequence that is hybridized or hybridizable to the target sequence, and (2) the tracr mate sequence that is hybridized or hybridizable to the tracr sequence, or (B) a non-naturally occurring or engineered composition comprising a vector system comprising one or more vectors comprising I. a first regulatory element operably linked to (a) a guide sequence capable of hybridizing to a target sequence in a eukaryotic cell, and (b) at least one or more tracr mate sequences, II. a second regulatory element operably linked to an enzyme-coding sequence encoding a CRISPR enzyme, and III. a third regulatory element operably linked to a tracr sequence, wherein components I, II and III are located on the same or different vectors of the system, wherein when transcribed, the tracr mate sequence hybridizes to the tracr sequence and the guide sequence directs sequence-specific binding of a CRISPR complex to the target sequence, and wherein the CRISPR complex comprises the CRISPR enzyme complexed with (1) the guide sequence that is hybridized or hybridizable to the target sequence, and (2) the tracr mate sequence that is hybridized or hybridizable to the tracr sequence. In some embodiments, components I, II and III are located on the same vector. In other embodiments, components I and II are located on the same vector, while component III is located on another vector. In other embodiments, components I and III are located on the same vector, while component II is located on another vector. In other embodiments, components II and III are located on the same vector, while component I is located on another vector. In other embodiments, each of components I, II and III is located on different vectors. The invention also provides a viral (e.g. AAV or lentivirus) vector system as described herein. , and can be part of a vector system as described herein.

[0054] Some methods of the invention can include inducing expression. The organism or subject is a eukaryote (including mammal including human) or a non-human eukaryote or a non-human animal or a non-human mammal, provided it has a liver or hepatic function. In some embodiments, the organism or subject is a non-human animal, and may be an arthropod, for example, an insect, or may be a nematode. In some methods of the invention the organism or subject is a mammal or a non-human mammal. A non-human mammal may be for example a rodent (preferably a mouse or a rat), an ungulate, or a primate. In some methods of the invention the viral vector is an AAV or a lentivirus, and can be part of a vector system as described herein. In some methods of the invention the CRISPR enzyme is a Cas9. In some methods of the

Date Reçue/Date Received 2020-10-05

invention the expression of the guide sequence is under the control of the T7 promoter and is driven by the expression of T7 polymerase.

[0055] The invention in some embodiments comprehends a method of delivering a CRISPR enzyme comprising delivering to a cell mRNA encoding the CRISPR enzyme. In some of these methods the CRISPR enzyme is a Cas9.

[0056] The invention also provides methods of preparing the vector systems of the invention, in particular the viral vector systems as described herein. The invention in some embodiments comprehends a method of preparing the AAV of the invention comprising transfecting plasmid(s) containing or consisting essentially of nucleic acid molecule(s) coding for the AAV into AAV-infected cells, and supplying AAV rep and/or cap obligatory for replication and packaging of the AAV. In some embodiments the AAV rep and/or cap obligatory for replication and packaging of the AAV are supplied by transfecting the cells with helper plasmid(s) or helper virus(es). In some embodiments the helper virus is a poxvirus, adenovirus, herpesvirus or baculovirus. In some embodiments the poxvirus is a vaccinia virus. In some embodiments the cells are mammalian cells. And in some embodiments the cells are insect cells and the helper virus is baculovirus. In other embodiments, the virus is a lentivirus.

[0057] The invention further comprehends a composition of the invention or a CRISPR enzyme thereof (including or alternatively mRNA encoding the CRISPR enzyme) for use in medicine or in therapy. In some embodiments the invention comprehends a composition according to the invention or a CRISPR enzyme thereof (including or alternatively mRNA encoding the CRISPR enzyme) for use in a method according to the invention. In some embodiments the invention provides for the use of a composition of the invention or a CRISPR enzyme thereof (including or alternatively mRNA encoding the CRISPR enzyme) in *ex vivo* gene or genome editing. In certain embodiments the invention comprehends use of a composition of the invention or a CRISPR enzyme thereof (including or alternatively mRNA encoding the CRISPR enzyme) in the manufacture of a medicament for *ex vivo* gene or genome editing or for use in a method according of the invention. The invention comprehends in some embodiments a composition of the invention or a CRISPR enzyme thereof (including or alternatively mRNA encoding the CRISPR enzyme), wherein the target sequence is flanked at its 3' end by a PAM (protospacer adjacent motif) sequence comprising 5'-motif, especially where the Cas9 is (or is derived from) *S. pyogenes* or *S. aureus* Cas9. For example, a suitable PAM is

Date Reçue/Date Received 2020-10-05

5'-NRG or 5'-NNGRR (where N is any Nucleotide) for SpCas9 or SaCas9 enzymes (or derived enzymes), respectively, as mentioned below.

[0058] It will be appreciated that SpCas9 or SaCas9 are those from or derived from *S. pyogenes* or *S. aureus* Cas9. It may of course, be mutated or otherwise changed from the wild type to suit the intended use, as described herein. The dual nciakse D10A mutant or variant is preferred, especially in combination with two overlapping guides directed as oppositing sites on differneing strands of the same chromosome.

[0059] Apects of the invention comprehend improving the specificity of a CRISPR enzyme, e.g. Cas9, mediated gene targeting and reducing the likelihood of off-target modification by the CRISPR enzyme, e.g. Cas9. The invention in some embodiments comprehends a method of modifying an organism or a non-human organism by minimizing off-target modifications by manipulation of a first and a second target sequence on opposite strands of a DNA duplex in a genomic locus of interest in a cell comprising delivering a non-naturally occurring or engineered composition comprising :

[0060] I. a first CRISPR-Cas system chimeric RNA (chiRNA) polynucleotide sequence, wherein the first polynucleotide sequence comprises:

(a) a first guide sequence capable of hybridizing to the first target sequence,

(b) a first tracr mate sequence, and

(c) a first tracr sequence,

[0061] II. a second CRISPR-Cas system chiRNA polynucleotide sequence, wherein the second polynucleotide sequence comprises:

(a) a second guide sequence capable of hybridizing to the second target sequence,

(b) a second tracr mate sequence, and

(c) a second tracr sequence, and

[0062] III. a polynucleotide sequence encoding a CRISPR enzyme comprising at least one or more nuclear localization sequences and comprising one or more mutations, wherein (a), (b) and (c) are arranged in a 5' to 3' orientation, wherein when transcribed, the first and the second tracr mate sequence hybridize to the first and second tracr sequence respectively and the first and the second guide sequence directs sequence-specific binding of a first and a second CRISPR complex to the first and second target sequences respectively, wherein the first CRISPR complex comprises the CRISPR enzyme complexed with (1) the first guide sequence that is hybridized or

Date Reçue/Date Received 2020-10-05

hybridizable to the first target sequence, and (2) the first tracr mate sequence that is hybridized or hybridizable to the first tracr sequence, wherein the second CRISPR complex comprises the CRISPR enzyme complexed with (1) the second guide sequence that is hybridized or hybridizable to the second target sequence, and (2) the second tracr mate sequence that is hybridized or hybridizable to the second tracr sequence, wherein the polynucleotide sequence encoding a CRISPR enzyme is DNA or RNA, and wherein the first guide sequence directs cleavage of one strand of the DNA duplex near the first target sequence and the second guide sequence directs cleavage of the other strand near the second target sequence inducing a double strand break, thereby modifying the organism or the non-human organism by minimizing off-target modifications.

[0063] In some methods of the invention any or all of the polynucleotide sequence encoding the CRISPR enzyme, the first and the second guide sequence, the first and the second tracr mate sequence or the first and the second tracr sequence, is/are RNA. In further embodiments of the invention the polynucleotides comprising the sequence encoding the CRISPR enzyme, the first and the second guide sequence, the first and the second tracr mate sequence or the first and the second tracr sequence, is/are RNA and are delivered via liposomes, nanoparticles, exosomes, microvesicles, or a gene-gun. In certain embodiments of the invention, the first and second tracr mate sequence share 100% identity and/or the first and second tracr sequence share 100% identity. In some embodiments, the polynucleotides may be comprised within a vector system comprising one or more vectors. In preferred embodiments of the invention the CRISPR enzyme is a Cas9 enzyme, e.g. SpCas9. In an aspect of the invention the CRISPR enzyme comprises one or more mutations in a catalytic domain, wherein the one or more mutations are selected from the group consisting of D10A, E762A, H840A, N854A, N863A and D986A. In a highly preferred embodiment the CRISPR enzyme has the D10A mutation. In preferred embodiments, the first CRISPR enzyme has one or more mutations such that the enzyme is a complementary strand nicking enzyme, and the second CRISPR enzyme has one or more mutations such that the enzyme is a non-complementary strand nicking enzyme. Alternatively the first enzyme may be a non-complementary strand nicking enzyme, and the second enzyme may be a complementary strand nicking enzyme.

[0064] In preferred methods of the invention the first guide sequence directing cleavage of one strand of the DNA duplex near the first target sequence and the second guide sequence

Date Reçue/Date Received 2020-10-05

directing cleavage of the other strand near the second target sequence results in a 5' overhang. In embodiments of the invention the 5' overhang is at most 200 base pairs, preferably at most 100 base pairs, or more preferably at most 50 base pairs. In embodiments of the invention the 5' overhang is at least 26 base pairs, preferably at least 30 base pairs or more preferably 34-50 base pairs. Most preferably, the overlap is between 5 and -1 base pairs.

[0065] The invention in some embodiments comprehends a method of modifying an organism or a non-human organism by minimizing off-target modifications by manipulation of a first and a second target sequence on opposite strands of a DNA duplex in a genomic locus of interest in a cell comprising delivering a non-naturally occurring or engineered composition comprising a vector system comprising one or more vectors comprising

[0066] I. a first regulatory element operably linked to

(a) a first guide sequence capable of hybridizing to the first target sequence, and

(b) at least one or more tracr mate sequences,

[0067] II. a second regulatory element operably linked to

(a) a second guide sequence capable of hybridizing to the second target sequence, and

(b) at least one or more tracr mate sequences,

[0068] III. a third regulatory element operably linked to an enzyme-coding sequence encoding a CRISPR enzyme, and

[0069] IV. a fourth regulatory element operably linked to a tracr sequence,

[0070] wherein components I, II, III and IV are located on the same or different vectors of the system, when transcribed, the tracr mate sequence hybridizes to the tracr sequence and the first and the second guide sequence direct sequence-specific binding of a first and a second CRISPR complex to the first and second target sequences respectively, wherein the first CRISPR complex comprises the CRISPR enzyme complexed with (1) the first guide sequence that is hybridized or hybridizable to the first target sequence, and (2) the tracr mate sequence that is hybridized or hybridizable to the tracr sequence, wherein the second CRISPR complex comprises the CRISPR enzyme complexed with (1) the second guide sequence that is hybridized or hybridizable to the second target sequence, and (2) the tracr mate sequence that is hybridized or hybridizable to the tracr sequence, wherein the polynucleotide sequence encoding a CRISPR enzyme is DNA or RNA, and wherein the first guide sequence directs cleavage of one strand of the DNA duplex near the first target sequence and the second guide sequence directs cleavage of

Date Reçue/Date Received 2020-10-05

the other strand near the second target sequence inducing a double strand break, thereby modifying the organism or the non-human organism by minimizing off-target modifications.

[0071] The invention also provides a vector system as described herein. The system may comprise one, two, three or four different vectors. Components I, II, III and IV may thus be located on one, two, three or four different vectors, and all combinations for possible locations of the components are herein envisaged, for example: components I, II, III and IV can be located on the same vector; components I, II, III and IV can each be located on different vectors; components I, II, II I and IV may be located on a total of two or three different vectors, with all combinations of locations envisaged, etc.

[0072] In some methods of the invention any or all of the polynucleotide sequence encoding the CRISPR enzyme, the first and the second guide sequence, the first and the second tracr mate sequence or the first and the second tracr sequence, is/are RNA. In further embodiments of the invention the first and second tracr mate sequence share 100% identity and/or the first and second tracr sequence share 100% identity. In preferred embodiments of the invention the CRISPR enzyme is a Cas9 enzyme, e.g. SpCas9. In an aspect of the invention the CRISPR enzyme comprises one or more mutations in a catalytic domain, wherein the one or more mutations are selected from the group consisting of D10A, E762A, H840A, N854A, N863A and D986A. In a highly preferred embodiment the CRISPR enzyme has the D10A mutation. In preferred embodiments, the first CRISPR enzyme has one or more mutations such that the enzyme is a complementary strand nicking enzyme, and the second CRISPR enzyme has one or more mutations such that the enzyme is a non-complementary strand nicking enzyme. Alternatively the first enzyme may be a non-complementary strand nicking enzyme, and the second enzyme may be a complementary strand nicking enzyme. In a further embodiment of the invention, one or more of the viral vectors are delivered via liposomes, nanoparticles, exosomes, microvesicles, or a gene-gun.

[0073] In preferred methods of the invention the first guide sequence directing cleavage of one strand of the DNA duplex near the first target sequence and the second guide sequence directing cleavage of other strand near the second target sequence results in a 5' overhang. In embodiments of the invention the 5' overhang is at most 200 base pairs, preferably at most 100 base pairs, or more preferably at most 50 base pairs. In embodiments of the invention the 5'

Date Reçue/Date Received 2020-10-05

overhang is at least 26 base pairs, preferably at least 30 base pairs or more preferably 34-50 base pairs.

[0074] The invention in some embodiments comprehends a method of modifying a genomic locus of interest by minimizing off-target modifications by introducing into a cell containing and expressing a double stranded DNA molecule encoding a gene product of interest an engineered, non-naturally occurring CRISPR-Cas system comprising a Cas protein having one or more mutations and two guide RNAs that target a first strand and a second strand of the DNA molecule respectively, whereby the guide RNAs target the DNA molecule encoding the gene product and the Cas protein nicks each of the first strand and the second strand of the DNA molecule encoding the gene product, whereby expression of the gene product is altered; and, wherein the Cas protein and the two guide RNAs do not naturally occur together.

[0075] In preferred methods of the invention the Cas protein nicking each of the first strand and the second strand of the DNA molecule encoding the gene product results in a 5' overhang. In embodiments of the invention the 5' overhang is at most 200 base pairs, preferably at most 100 base pairs, or more preferably at most 50 base pairs. In embodiments of the invention the 5' overhang is at least 26 base pairs, preferably at least 30 base pairs or more preferably 34-50 base pairs.

[0076] Embodiments of the invention also comprehend the guide RNAs comprising a guide sequence fused to a tracr mate sequence and a tracr sequence. In an aspect of the invention the Cas protein is codon optimized for expression in a eukaryotic cell, preferably a mammalian cell or a human cell. In further embodiments of the invention the Cas protein is a type II CRISPR-Cas protein, e.g. a Cas 9 protein. In a highly preferred embodiment the Cas protein is a Cas9 protein, e.g. SpCas9. In aspects of the invention the Cas protein has one or more mutations selected from the group consisting of D10A, E762A, H840A, N854A, N863A and D986A. In a highly preferred embodiment the Cas protein has the D10A mutation.

[0077] Aspects of the invention relate to the expression of the gene product being decreased or a template polynucleotide being further introduced into the DNA molecule encoding the gene product or an intervening sequence being excised precisely by allowing the two 5' overhangs to reanneal and ligate or the activity or function of the gene product being altered or the expression of the gene product being increased. In an embodiment of the invention, the gene product is a protein.

Date Reçue/Date Received 2020-10-05

[0078] The invention also comprehends an engineered, non-naturally occurring CRISPR-Cas system comprising a Cas protein having one or more mutations and two guide RNAs that target a first strand and a second strand respectively of a double stranded DNA molecule encoding a gene product in a cell, whereby the guide RNAs target the DNA molecule encoding the gene product and the Cas protein nicks each of the first strand and the second strand of the DNA molecule encoding the gene product, whereby expression of the gene product is altered; and, wherein the Cas protein and the two guide RNAs do not naturally occur together.

[0079] In aspects of the invention the guide RNAs may comprise a guide sequence fused to a tracr mate sequence and a tracr sequence. In an embodiment of the invention the Cas protein is a type II CRISPR-Cas protein. In an aspect of the invention the Cas protein is codon optimized for expression in a eukaryotic cell, preferably a mammalian cell or a human cell. In further embodiments of the invention the Cas protein is a type II CRISPR-Cas protein, e.g. a Cas 9 protein. In a highly preferred embodiment the Cas protein is a Cas9 protein, e.g. SpCas9. In aspects of the invention the Cas protein has one or more mutations selected from the group consisting of D10A, E762A, H840A, N854A, N863A and D986A. In a highly preferred embodiment the Cas protein has the D10A mutation.

[0080] Aspects of the invention relate to the expression of the gene product being decreased or a template polynucleotide being further introduced into the DNA molecule encoding the gene product or an intervening sequence being excised precisely by allowing the two 5' overhangs to reanneal and ligate or the activity or function of the gene product being altered or the expression of the gene product being increased. In an embodiment of the invention, the gene product is a protein.

[0081] The invention also comprehends an engineered, non-naturally occurring vector system comprising one or more vectors comprising:

a) a first regulatory element operably linked to each of two CRISPR-Cas system guide RNAs that target a first strand and a second strand respectively of a double stranded DNA molecule encoding a gene product,

b) a second regulatory element operably linked to a Cas protein,

wherein components (a) and (b) are located on same or different vectors of the system, whereby the guide RNAs target the DNA molecule encoding the gene product and the Cas protein nicks each of the first strand and the second strand of the DNA molecule encoding the gene product,

Date Reçue/Date Received 2020-10-05

whereby expression of the gene product is altered; and, wherein the Cas protein and the two guide RNAs do not naturally occur together.

[0082] In aspects of the invention the guide RNAs may comprise a guide sequence fused to a tracr mate sequence and a tracr sequence. In an embodiment of the invention the Cas protein is a type II CRISPR-Cas protein. In an aspect of the invention the Cas protein is codon optimized for expression in a eukaryotic cell, preferably a mammalian cell or a human cell. In further embodiments of the invention the Cas protein is a type II CRISPR-Cas protein, e.g. a Cas 9 protein. In a highly preferred embodiment the Cas protein is a Cas9 protein, e.g. SpCas9. In aspects of the invention the Cas protein has one or more mutations selected from the group consisting of D10A, E762A, H840A, N854A, N863A and D986A. In a highly preferred embodiment the Cas protein has the D10A mutation.

[0083] Aspects of the invention relate to the expression of the gene product being decreased or a template polynucleotide being further introduced into the DNA molecule encoding the gene product or an intervening sequence being excised precisely by allowing the two 5' overhangs to reanneal and ligate or the activity or function of the gene product being altered or the expression of the gene product being increased. In an embodiment of the invention, the gene product is a protein. In preferred embodiments of the invention the vectors of the system are viral vectors. In a further embodiment, the vectors of the system are delivered via liposomes, nanoparticles, exosomes, microvesicles, or a gene-gun.

[0084] In one aspect, the invention provides a method of modifying a target polynucleotide in a liver cell. In some embodiments, the method comprises allowing a CRISPR complex to bind to the target polynucleotide to effect cleavage of said target polynucleotide thereby modifying the target polynucleotide, wherein the CRISPR complex comprises a CRISPR enzyme complexed with a guide sequence hybridized or hybridizable to a target sequence within said target polynucleotide, wherein said guide sequence is linked to a tracr mate sequence which in turn hybridizes to a tracr sequence. In some embodiments, said cleavage comprises cleaving one or two strands at the location of the target sequence by said CRISPR enzyme. In some embodiments, said cleavage results in decreased transcription of a target gene. In some embodiments, the method further comprises repairing said cleaved target polynucleotide by homologous recombination with an exogenous template polynucleotide, wherein said repair results in a mutation comprising an insertion, deletion, or substitution of one or more nucleotides

Date Reçue/Date Received 2020-10-05

of said target polynucleotide. In some embodiments, said mutation results in one or more amino acid changes in a protein expressed from a gene comprising the target sequence. In some embodiments, the method further comprises delivering one or more vectors to said eukaryotic cell, wherein the one or more vectors drive expression of one or more of: the CRISPR enzyme, the guide sequence linked to the tracr mate sequence, and the tracr sequence. In some embodiments, said vectors are delivered to the eukaryotic cell in a subject. In some embodiments, said modifying takes place in said eukaryotic cell in a cell culture. In some embodiments, the method further comprises isolating said eukaryotic cell from a subject prior to said modifying. In some embodiments, the method further comprises returning said eukaryotic cell and/or cells derived therefrom to said subject.

[0085] In one aspect, the invention provides a method of modifying expression of a polynucleotide in a liver cell. In some embodiments, the method comprises allowing a CRISPR complex to bind to the polynucleotide such that said binding results in increased or decreased expression of said polynucleotide; wherein the CRISPR complex comprises a CRISPR enzyme complexed with a guide sequence hybridized or hybridizable to a target sequence within said polynucleotide, wherein said guide sequence is linked to a tracr mate sequence which in turn hybridizes to a tracr sequence. In some embodiments, the method further comprises delivering one or more vectors to said eukaryotic cells, wherein the one or more vectors drive expression of one or more of: the CRISPR enzyme, the guide sequence linked to the tracr mate sequence, and the tracr sequence.

[0086] In one aspect, the invention provides a method of generating a model liver cell comprising a mutated disease gene. In some embodiments, a disease gene is any gene associated with an increase in the risk of having or developing a disease. In some embodiments, the method comprises (a) introducing one or more vectors into a eukaryotic cell, wherein the one or more vectors drive expression of one or more of: a CRISPR enzyme, a guide sequence linked to a tracr mate sequence, and a tracr sequence; and (b) allowing a CRISPR complex to bind to a target polynucleotide to effect cleavage of the target polynucleotide within said disease gene, wherein the CRISPR complex comprises the CRISPR enzyme complexed with (1) the guide sequence that is hybridized or hybridizable to the target sequence within the target polynucleotide, and (2) the tracr mate sequence that is hybridized or hybridizable to the tracr sequence, thereby generating a model eukaryotic cell comprising a mutated disease gene. In some embodiments,

Date Reçue/Date Received 2020-10-05

said cleavage comprises cleaving one or two strands at the location of the target sequence by said CRISPR enzyme. In some embodiments, said cleavage results in decreased transcription of a target gene. In some embodiments, the method further comprises repairing said cleaved target polynucleotide by homologous recombination with an exogenous template polynucleotide, wherein said repair results in a mutation comprising an insertion, deletion, or substitution of one or more nucleotides of said target polynucleotide. In some embodiments, said mutation results in one or more amino acid changes in a protein expression from a gene comprising the target sequence.

[0087] In one aspect the invention provides for a method of selecting one or more liver cell(s) by introducing one or more mutations in a gene in the one or more cell (s), the method comprising: introducing one or more vectors into the cell (s), wherein the one or more vectors drive expression of one or more of: a CRISPR enzyme, a guide sequence linked to a tracr mate sequence, a tracr sequence, and an editing template; wherein the editing template comprises the one or more mutations that abolish CRISPR enzyme cleavage; allowing homologous recombination of the editing template with the target polynucleotide in the cell(s) to be selected; allowing a CRISPR complex to bind to a target polynucleotide to effect cleavage of the target polynucleotide within said gene, wherein the CRISPR complex comprises the CRISPR enzyme complexed with (1) the guide sequence that is hybridized or hybridizable to the target sequence within the target polynucleotide, and (2) the tracr mate sequence that is hybridized or hybridizable to the tracr sequence, wherein binding of the CRISPR complex to the target polynucleotide induces cell death, thereby allowing one or more prokaryotic cell(s) in which one or more mutations have been introduced to be selected. In a preferred embodiment, the CRISPR enzyme is Cas9. Aspects of the invention allow for selection of specific cells without requiring a selection marker or a two-step process that may include a counter-selection system.

[0088] In one aspect, the invention provides for methods of modifying a target polynucleotide in a liver cell. In some embodiments, the method comprises allowing a CRISPR complex to bind to the target polynucleotide to effect cleavage of said target polynucleotide thereby modifying the target polynucleotide, wherein the CRISPR complex comprises a CRISPR enzyme complexed with a guide sequence hybridized or hybridizable to a target sequence within said target polynucleotide, wherein said guide sequence is linked to a tracr mate sequence which in turn hybridizes to a tracr sequence.

Date Reçue/Date Received 2020-10-05

[0089] In other embodiments, this invention provides a method of modifying expression of a polynucleotide in a liver cell. The method comprises increasing or decreasing expression of a target polynucleotide by using a CRISPR complex that binds to the polynucleotide.

[0090] Where desired, to effect the modification of the expression in a cell, one or more vectors comprising a tracr sequence, a guide sequence linked to the tracr mate sequence, a sequence encoding a CRISPR enzyme is delivered to a cell. In some methods, the one or more vectors comprises a regulatory element operably linked to an enzyme-coding sequence encoding said CRISPR enzyme comprising a nuclear localization sequence; and a regulatory element operably linked to a tracr mate sequence and one or more insertion sites for inserting a guide sequence upstream of the tracr mate sequence. When expressed, the guide sequence directs sequence-specific binding of a CRISPR complex to a target sequence in a cell. Typically, the CRISPR complex comprises a CRISPR enzyme complexed with (1) the guide sequence that is hybridized or hybridizable to the target sequence, and (2) the tracr mate sequence that is hybridized or hybridizable to the tracr sequence.

[0091] In some methods, a target polynucleotide can be inactivated to effect the modification of the expression in a cell. For example, upon the binding of a CRISPR complex to a target sequence in a cell, the target polynucleotide is inactivated such that the sequence is not transcribed, the coded protein is not produced, or the sequence does not function as the wild-type sequence does. For example, a protein or microRNA coding sequence may be inactivated such that the protein is not produced.

[0092] In certain embodiments, the CRISPR enzyme comprises one or more mutations selected from the group consisting of D10A, E762A, H840A, N854A, N863A or D986A and/or the one or more mutations is in a RuvC1 or HNH domain of the CRISPR enzyme or is a mutation as otherwise as discussed herein. In some embodiments, the CRISPR enzyme has one or more mutations in a catalytic domain, wherein when transcribed, the tracr mate sequence hybridizes to the tracr sequence and the guide sequence directs sequence-specific binding of a CRISPR complex to the target sequence, and wherein the enzyme further comprises a functional domain. In some embodiments, the functional domain is a transcriptional activation domain, preferably VP64. In some embodiments, the functional domain is a transcription repression domain, preferably KRAB. In some embodiments, the transcription repression domain is SID, or concatemers of SID (eg SID4X). In some embodiments, the functional domain is an epigenetic

Date Reçue/Date Received 2020-10-05

modifying domain, such that an epigenetic modifying enzyme is provided. In some embodiments, the functional domain is an activation domain, which may be the P65 activation domain.

[0093] In some embodiments, the CRISPR enzyme is a type I or III CRISPR enzyme, but is preferably a type II CRISPR enzyme. This type II CRISPR enzyme may be any Cas enzyme. A Cas enzyme may be identified as Cas9 as this can refer to the general class of enzymes that share homology to the biggest nuclease with multiple nuclease domains from the type II CRISPR system. Most preferably, the Cas9 enzyme is from, or is derived from, spCas9 or saCas9. By derived, Applicants mean that the derived enzyme is largely based, in the sense of having a high degree of sequence homology with, a wildtype enzyme, but that it has been mutated (modified) in some way as described herein.

[0094] It will be appreciated that the terms Cas and CRISPR enzyme are generally used herein interchangeably, unless otherwise apparent. As mentioned above, many of the residue numberings used herein refer to the Cas9 enzyme from the type II CRISPR locus in *Streptococcus pyogenes*. However, it will be appreciated that this invention includes many more Cas9s from other species of microbes, such as SpCas9, SaCa9, St1Cas9 and so forth.

[0095] An example of a codon optimized sequence, in this instance optimized for humans (i.e. being optimized for expression in humans) is provided herein, see the SaCas9 human codon optimized sequence. Whilst this is preferred, it will be appreciated that other examples are possible and codon optimization for a host species is known.

[0096] Preferably, delivery is in the form of a vector which may be a viral vector, such as a lenti- or baculo- or preferably adeno-viral/adeno-associated viral vectors, but other means of delivery are known (such as yeast systems, microvesicles, gene guns/means of attaching vectors to gold nanoparticles) and are provided. A vector may mean not only a viral or yeast system (for instance, where the nucleic acids of interest may be operably linked to and under the control of (in terms of expression, such as to ultimately provide a processed RNA) a promoter), but also direct delivery of nucleic acids into a host cell. While in herein methods the vector may be a viral vector and this is advantageously an AAV, other viral vectors as herein discussed can be employed, such as lentivirus. For example, baculoviruses may be used for expression in insect cells. These insect cells may, in turn be useful for producing large quantities of further vectors, such as AAV or lentivirus vectors adapted for delivery of the present invention. Also envisaged

Date Reçue/Date Received 2020-10-05

is a method of delivering the present CRISPR enzyme comprising delivering to a cell mRNA encoding the CRISPR enzyme. It will be appreciated that in certain embodiments the CRISPR enzyme is truncated, and/or comprised of less than one thousand amino acids or less than four thousand amino acids, and/or is a nuclease or nickase, and/or is codon-optimized, and/or comprises one or more mutations, and/or comprises a chimeric CRISPR enzyme, and/or the other options as herein discussed. AAV and lentiviral vectors are preferred.

[0097] In certain embodiments, the target sequence is flanked or followed, at its 3' end, by a PAM suitable for the CRISPR enzyme, typically a Cas and in particular a Cas9.

[0098] For example, a suitable PAM is 5'-NRG or 5'-NNGRR for SpCas9 or SaCas9 enzymes (or derived enzymes), respectively.

[0099] It will be appreciated that SpCas9 or SaCas9 are those from or derived from *S. pyogenes* or *S. aureus* Cas9.

[00100] Some points in the present application are summarised below:

[00101] **AAV2/8**

[00102] Preferred delivery for the CRISPR-Cas system is through a viral vector. This vector may be a lentiviral vector or an AAV vector, as discussed at some length herein. Whet we have particularly showed is that AAV is a preferred example of a viral vector. Within that, we gone on to show that AAV8 and in particular AAV2/8 (AAV8 packaged with AAV2 packaging signal ITR) is useful in delivery to the liver, especially in vivo.

[00103] **Phenotypic Changes seen In Vivo**

[00104] As discussed elsewhere, we have been able to show, in vivo, that phenotypic change can be detected. This is a significant step forward as a deficiency often levelled at RNAi is that no lasting effect is seen. With the present invention, phenotypic change can be seen in the liver for the first time. A preferred arrangement to achieve this is to use that in Example 36. Important elements of this are preferred alone or in combination, namely:

- Sa Cas9;
- Use of a chimeric guide RNA comprising the guide, tracr sequence and tracr mate;
- For the tracr sequence, Sa tracr is preferable to recruit the Sa Cas9;
- AAV8 or more preferably AAV2/8;
- For experimental purposes, Rosa26 is a useful negative control;

Date Reçue/Date Received 2020-10-05

- Although use of the CMV promoter in an AAV vector is helpful, use of a liver-specific promoter such as TBG is particularly effective;
- The target or targets may be wide-ranging as CRISPR has been shown to have broad applicability across targets, once they guides are successfully delivered and the Css9 enzymes are suitably expressed. However, preferred targets in the liver (against which the guides may be designed) nevertheless include one or more of: PCSK9; Hmgcr; SERPINA1; ApoB; and.or LDL.

[00105] Accordingly, in some embodiments it is particularly preferred that the Cas enzyme is an Sa Cas9. Preferably, the CRISPRS-Cas polynucleotide sequence is chimeric and preferably includes an Sa tracr where the Cas9 is an Sa Cas9. A viral vector may be used which is preferably AAV2/8. Furthermore, a liver-specific promoter is ideal and a preferred example is TBG. All of these may be used in combination to provide a chimeric CRISPRS-Cas polynucleotide sequence including an Sa tracr, wherein the Cas9 is an SaCas9, and the vector is AAV2/8, with at least the Cas9 under the control of a liver-specific such as TBG. Any of the above targets may be sued with this system, in particular ApoB due to its importance in obesity.

[00106] Yin and Anderson's later Nature Biotech Paper (NBT 2884, referenced herein) provides further support for the in vivo phenotypic changes that we have already shown.

[00107] Additional data that we provide in Example 37, then adds further support by demonstrating efficient in vivo editing of somatic liver tissue via Cas9. Moreover, delivery via AAV2/8 and the use of an SaCas9 again show the usefulness of this particular approach in vivo. The preferred ApoB was again targeted.

[00108] Later examples 38 and 39 show excellent in vivo data for effecicacy in inducing a phenotypic change in vivo: specifically ApoB, a lipid metabolism gene, whilst Example 40 shows the applicability of the technique to post-mitotic cells, of which liver is an important example. Example 41 shows that multiple epitope tags are preferable for detection purposes.

[00109] Although viral vectors are preferred, in some embodiments, the use of cell penetrating peptides is a viable alternative and so is also preferred.

[00110] Accordingly, it is an object of the invention to not encompass within the invention any previously known product, process of making the product, or method of using the product such that Applicants reserve the right and hereby disclose a disclaimer of any previously known product, process, or method. It is further noted that the invention does not intend to encompass within the scope of the invention any product, process, or making of the product or method of

Date Reçue/Date Received 2020-10-05

using the product, which does not meet the written description and enablement requirements of the USPTO (35 U.S.C. §112, first paragraph) or the EPO (Article 83 of the EPC), such that Applicants reserve the right and hereby disclose a disclaimer of any previously described product, process of making the product, or method of using the product.

[00111] It is noted that in this disclosure and particularly in the claims and/or paragraphs, terms such as "comprises", "comprised", "comprising" and the like can have the meaning attributed to it in U.S. Patent law; e.g., they can mean "includes", "included", "including", and the like; and that terms such as "consisting essentially of" and "consists essentially of" have the meaning ascribed to them in U.S. Patent law, e.g., they allow for elements not explicitly recited, but exclude elements that are found in the prior art or that affect a basic or novel characteristic of the invention.

[00112] These and other embodiments are disclosed or are obvious from and encompassed by, the following Detailed Description.

BRIEF DESCRIPTION OF THE DRAWINGS

[00113] The novel features of the invention are set forth with particularity in the appended claims. A better understanding of the features and advantages of the present invention will be obtained by reference to the following detailed description that sets forth illustrative embodiments, in which the principles of the invention are utilized, and the accompanying drawings of which:

[00114] **Figure 1** shows a schematic model of the CRISPR system. The Cas9 nuclease from Streptococcus pyogenes (yellow) is targeted to genomic DNA by a synthetic guide RNA (sgRNA) consisting of a 20-nt guide sequence (blue) and a scaffold (red). The guide sequence base-pairs with the DNA target (blue), directly upstream of a requisite 5'-NGG protospacer adjacent motif (PAM; magenta), and Cas9 mediates a double-stranded break (DSB) ~3 bp upstream of the PAM (red triangle).

[00115] **Figure 2A-F** shows an exemplary CRISPR system, a possible mechanism of action, an example adaptation for expression in eukaryotic cells, and results of tests assessing nuclear localization and CRISPR activity.

[00116] **Figure 3A-D** shows results of an evaluation of SpCas9 specificity for an example target.

Date Reçue/Date Received 2020-10-05

[00117] **Figure 4A-G** show an exemplary vector system and results for its use in directing homologous recombination in eukaryotic cells.

[00118] **Figure 5** provides a table of protospacer sequences and summarizes modification efficiency results for protospacer targets designed based on exemplary S. pyogenes and S. thermophilus CRISPR systems with corresponding PAMs against loci in human and mouse genomes. Cells were transfected with Cas9 and either pre-crRNA/tracrRNA or chimeric RNA, and analyzed 72 hours after transfection. Percent indels are calculated based on Surveyor assay results from indicated cell lines (N=3 for all protospacer targets, errors are S.E.M., N.D. indicates not detectable using the Surveyor assay, and N.T. indicates not tested in this study).

[00119] **Figure 6A-C** shows a comparison of different tracrRNA transcripts for Cas9-mediated gene targeting.

[00120] **Figure 7** shows a schematic of a surveyor nuclease assay for detection of double strand break-induced micro-insertions and –deletions.

[00121] **Figure 8A-B** shows exemplary bicistronic expression vectors for expression of CRISPR system elements in eukaryotic cells.

[00122] **Figure 9A-C** shows histograms of distances between adjacent S. pyogenes SF370 locus 1 PAM (NGG) (Figure 9A) and S. thermophilus LMD9 locus 2 PAM (NNAGAAW) (Figure 9B) in the human genome; and distances for each PAM by chromosome (Chr) (Figure 9C).

[00123] **Figure 10A-D** shows an exemplary CRISPR system, an example adaptation for expression in eukaryotic cells, and results of tests assessing CRISPR activity.

[00124] **Figure 11A-C** shows exemplary manipulations of a CRISPR system for targeting of genomic loci in mammalian cells.

[00125] **Figure 12A-B** shows the results of a Northern blot analysis of crRNA processing in mammalian cells.

[00126] **Figure 13A-B** shows an exemplary selection of protospacers in the human PVALB and mouse Th loci.

[00127] **Figure 14** shows example protospacer and corresponding PAM sequence targets of the S. thermophilus CRISPR system in the human EMX1 locus.

[00128] **Figure 15** provides a table of sequences for primers and probes used for Surveyor, RFLP, genomic sequencing, and Northern blot assays.

Date Reçue/Date Received 2020-10-05

[00129] **Figure 16A-C** shows exemplary manipulation of a CRISPR system with chimeric RNAs and results of SURVEYOR assays for system activity in eukaryotic cells.

[00130] **Figure 17A-B** shows a graphical representation of the results of SURVEYOR assays for CRISPR system activity in eukaryotic cells.

[00131] **Figure 18** shows an exemplary visualization of some S. pyogenes Cas9 target sites in the human genome using the UCSC genome browser.

[00132] **Figure 19A-D** shows a circular depiction of the phylogenetic analysis revealing five families of Cas9s, including three groups of large Cas9s (~1400 amino acids) and two of small Cas9s (~1100 amino acids).

[00133] **Figure 20A-F** shows the linear depiction of the phylogenetic analysis revealing five families of Cas9s, including three groups of large Cas9s (~1400 amino acids) and two of small Cas9s (~1100 amino acids).

[00134] **Figure 21A-D** shows genome editing via homologous recombination. (a) Schematic of SpCas9 nickase, with D10A mutation in the RuvC I catalytic domain. (b) Schematic representing homologous recombination (HR) at the human EMX1 locus using either sense or antisense single stranded oligonucleotides as repair templates. Red arrow above indicates sgRNA cleavage site; PCR primers for genotyping (Tables J and K) are indicated as arrows in right panel. (c) Sequence of region modified by HR. d, SURVEYOR assay for wildtype (wt) and nickase (D10A) SpCas9-mediated indels at the EMX1 target 1 locus (n=3). Arrows indicate positions of expected fragment sizes.

[00135] **Figure 22A-B** shows single vector designs for SpCas9.

[00136] **Figure 23** shows a graph representing the length distribution of Cas9 orthologs.

[00137] **Figure 24A-M** shows sequences where the mutation points are located within the SpCas9 gene.

[00138] **Figure 25A** shows the Conditional Cas9, Rosa26 targeting vector map.

[00139] **Figure 25B** shows the Constitutive Cas9, Rosa26 targeting vector map.

[00140] **Figure 26** shows a schematic of the important elements in the Constitutive and Conditional Cas9 constructs.

[00141] **Figure 27** shows delivery and *in vivo* mouse brain Cas9 expression data.

[00142] **Figure 28** shows RNA delivery of Cas9 and chimeric RNA into cells (A) Delivery of a GFP reporter as either DNA or mRNA into Neuro-2A cells. (B) Delivery of Cas9 and chimeric

Date Reçue/Date Received 2020-10-05

RNA against the Icam2 gene as RNA results in cutting for one of two spacers tested. (C) Delivery of Cas9 and chimeric RNA against the F7 gene as RNA results in cutting for one of two spacers tested.

[00143] **Figure 29** shows how DNA double-strand break (DSB) repair promotes gene editing. In the error-prone non-homologous end joining (NHEJ) pathway, the ends of a DSB are processed by endogenous DNA repair machineries and rejoined together, which can result in random insertion/deletion (indel) mutations at the site of junction. Indel mutations occurring within the coding region of a gene can result in frame-shift and a premature stop codon, leading to gene knockout. Alternatively, a repair template in the form of a plasmid or single-stranded oligodeoxynucleotides (ssODN) can be supplied to leverage the homology-directed repair (HDR) pathway, which allows high fidelity and precise editing.

[00144] **Figure 30A-C** shows anticipated results for HDR in HEK and HUES9 cells. (a) Either a targeting plasmid or an ssODN (sense or antisense) with homology arms can be used to edit the sequence at a target genomic locus cleaved by Cas9 (red triangle). To assay the efficiency of HDR, we introduced a HindIII site (red bar) into the target locus, which was PCR-amplified with primers that anneal outside of the region of homology. Digestion of the PCR product with HindIII reveals the occurrence of HDR events. (b) ssODNs, oriented in either the sense or the antisense (s or a) direction relative to the locus of interest, can be used in combination with Cas9 to achieve efficient HDR-mediated editing at the target locus. A minimal homology region of 40 bp, and preferably 90 bp, is recommended on either side of the modification (red bar). (c) Example of the effect of ssODNs on HDR in the EMX1 locus is shown using both wild-type Cas9 and Cas9 nickase (D10A). Each ssODN contains homology arms of 90 bp flanking a 12-bp insertion of two restriction sites.

[00145] **Figure 31A-C** shows the repair strategy for Cystic Fibrosis delta F508 mutation.

[00146] **Figure 32A-B** (a) shows a schematic of the GAA repeat expansion in FXN intron 1 and (b) shows a schematic of the strategy adopted to excise the GAA expansion region using the CRISPR/Cas system.

[00147] **Figure 33** shows a screen for efficient SpCas9 mediated targeting of Tet1-3 and Dnmt1, 3a and 3b gene loci. Surveyor assay on DNA from transfected N2A cells demonstrates efficient DNA cleavage by using different gRNAs.

Date Reçue/Date Received 2020-10-05

[00148] **Figure 34** shows a strategy of multiplex genome targeting using a 2-vector system in an AAV1/2 delivery system. Tet1-3 and Dnmt1, 3a and 3b gRNA under the control of the U6 promoter. GFP-KASH under the control of the human synapsin promoter. Restriction sides shows simple gRNA replacement strategy by subcloning. HA-tagged SpCas9 flanked by two nuclear localization signals (NLS) is shown. Both vectors are delivered into the brain by AAV1/2 virus in a 1:1 ratio.

[00149] **Figure 35** shows verification of multiplex DNMT targeting vector #1 functionality using Surveyor assay. N2A cells were co-transfected with the DNMT targeting vector #1 (+) and the SpCas9 encoding vector for testing SpCas9 mediated cleavage of DNMTs genes family loci. gRNA only (-) is negative control. Cells were harvested for DNA purification and downstream processing 48 h after transfection.

[00150] **Figure 36** shows verification of multiplex DNMT targeting vector #2 functionality using Surveyor assay. N2A cells were co-transfected with the DNMT targeting vector #1 (+) and the SpCas9 encoding vector for testing SpCas9 mediated cleavage of DNMTs genes family loci. gRNA only (-) is negative control. Cells were harvested for DNA purification and downstream processing 48 h after transfection.

[00151] **Figure 37** shows schematic overview of short promoters and short polyA versions used for HA-SpCas9 expression in vivo. Sizes of the encoding region from L-ITR to R-ITR are shown on the right.

[00152] **Figure 38** shows schematic overview of short promoters and short polyA versions used for HA-SaCas9 expression in vivo. Sizes of the encoding region from L-ITR to R-ITR are shown on the right.

[00153] **Figure 39** shows expression of SpCas9 and SaCas9 in N2A cells. Representative Western blot of HA-tagged SpCas9 and SaCas9 versions under the control of different short promoters and with or short polyA (spA) sequences. Tubulin is loading control. mCherry (mCh) is a transfection control. Cells were harvested and further processed for Western blotting 48 h after transfection.

[00154] **Figure 40** shows screen for efficient SaCas9 mediated targeting of Tet3 gene locus. Surveyor assay on DNA from transfected N2A cells demonstrates efficient DNA cleavage by using different gRNAs with NNGGGT PUM sequence. GFP transfected cells and cells expressing only SaCas9 are controls.

Date Reçue/Date Received 2020-10-05

[00155] **Figure 41** shows expression of HA-SaCas9 in the mouse brain. Animals were injected into dentate gyri with virus driving expression of HA-SaCas9 under the control of human Synapsin promoter. Animals were sacrificed 2 weeks after surgery. HA tag was detected using rabbit monoclonal antibody C29F4 (Cell Signaling). Cell nuclei stained in blue with DAPI stain.

[00156] **Figure 42** shows expression of SpCas9 and SaCas9 in cortical primary neurons in culture 7 days after transduction. Representative Western blot of HA-tagged SpCas9 and SaCas9 versions under the control of different promoters and with bgh or short polyA (spA) sequences. Tubulin is loading control.

[00157] **Figure 43** shows LIVE/DEAD stain of primary cortical neurons 7 days after transduction with AAV1 particles carrying SpCas9 with different promoters and multiplex gRNAs constructs (example shown on the last panel for DNMTs). Neurons after AAV transduction were compared with control untransduced neurons. Red nuclei indicate permeabilized, dead cells (second line of panels). Live cells are marked in green color (third line of panels).

[00158] **Figure 44** shows LIVE/DEAD stain of primary cortical neurons 7 days after transduction with AAV1 particles carrying SaCas9 with different promoters. Red nuclei indicate permeabilized, dead cells (second line of panels). Live cells are marked in green color (third line of panels).

[00159] **Figure 45** shows comparison of morphology of neurons after transduction with AAV1 virus carrying SpCas9 and gRNA multiplexes for TETs and DNMTs genes loci. Neurons without transduction are shown as a control.

[00160] **Figure 46** shows verification of multiplex DNMT targeting vector #1 functionality using Surveyor assay in primary cortical neurons. Cells were co-transduced with the DNMT targeting vector #1 and the SpCas9 viruses with different promoters for testing SpCas9 mediated cleavage of DNMTs genes family loci.

[00161] **Figure 47** shows in vivo efficiency of SpCas9 cleavage in the brain. Mice were injected with AAV1/2 virus carrying gRNA multiplex targeting DNMT family genes loci together with SpCas9 viruses under control of 2 different promoters: mouse Mecp2 and rat Map1b. Two weeks after injection brain tissue was extracted and nuclei were prepped and sorted using FACS, based on the GFP expression driven by Synapsin promoter from gRNA multiplex

Date Reçue/Date Received 2020-10-05

construct. After gDNA extraction Surveyor assay was run. + indicates GFP positive nuclei and – control, GFP-negative nuclei from the same animal. Numbers on the gel indicate assessed SpCas9 efficiency.

[00162] **Figure 48** shows purification of GFP-KASH labeled cell nuclei from hippocampal neurons. The outer nuclear membrane (ONM) of the cell nuclear membrane is tagged with a fusion of GFP and the KASH protein transmembrane domain. Strong GFP expression in the brain after one week of stereotactic surgery and AAV1/2 injection. Density gradient centrifugation step to purify cell nuclei from intact brain. Purified nuclei are shown. Chromatin stain by Vybrant® DyeCycle™ Ruby Stain is shown in red, GFP labeled nuclei are green. Representative FACS profile of GFP+ and GFP- cell nuclei (Magenta: Vybrant® DyeCycle™ Ruby Stain, Green: GFP).

[00163] **Figure 49** shows efficiency of SpCas9 cleavage in the mouse brain. Mice were injected with AAV1/2 virus carrying gRNA multiplex targeting TET family genes loci together with SpCas9 viruses under control of 2 different promoters: mouse Mecp2 and rat Map1b. Three weeks after injection brain tissue was extracted, nuclei were prepped and sorted using FACS, based on the GFP expression driven by Synapsin promoter from gRNA multiplex construct. After gDNA extraction Surveyor assay was run. + indicates GFP positive nuclei and – control, GFP-negative nuclei from the same animal. Numbers on the gel indicate assessed SpCas9 efficiency.

[00164] **Figure 50** shows GFP-KASH expression in cortical neurons in culture. Neurons were transduced with AAV1 virus carrying gRNA multiplex constructs targeting TET genes loci. The strongest signal localize around cells nuclei due to KASH domain localization.

[00165] **Figure 51** shows (top) a list of spacing (as indicated by the pattern of arrangement for two PAM sequences) between pairs of guide RNAs. Only guide RNA pairs satisfying patterns 1, 2, 3, 4 exhibited indels when used with SpCas9(D10A) nickase. (bottom) Gel images showing that combination of SpCas9(D10A) with pairs of guide RNA satisfying patterns 1, 2, 3, 4 led to the formation of indels in the target site.

[00166] **Figure 52** shows a list of U6 reverse primer sequences used to generate U6-guide RNA expression casssettes. Each primer needs to be paired with the U6 forward primer "gcactgagggcctatttcccatgattc" to generate amplicons containing U6 and the desired guide RNA.

Date Reçue/Date Received 2020-10-05

[00167] **Figure 53** shows a Genomic sequence map from the human Emx1 locus showing the locations of the 24 patterns listed in Figure 33.

[00168] **Figure 54** shows on (right) a gel image indicating the formation of indels at the target site when variable 5' overhangs are present after cleavage by the Cas9 nickase targeted by different pairs of guide RNAs. on (left) a table indicating the lane numbers of the gel on the right and various parameters including identifying the guide RNA pairs used and the length of the 5' overhang present following cleavage by the Cas9 nickase.

[00169] **Figure 55** shows a Genomic sequence map from the human Emx1 locus showing the locations of the different pairs of guide RNAs that result in the gel patterns of Fig. 54 (right) and which are further described in Example 35.

[00170] **Figure 56** shows a Representative Surveyor Gel showing genomic cleavage by SaCas9.

[00171] **Figure 57** shows Genome Cleavage Efficiency of PAM Sequences (All targets).

[00172] **Figure 58** shows Genome Cleavage Efficiency of PAM Sequences (Cleaved targets)

[00173] **Figure 59** shows Genome Cleavage Efficiency of PAM Sequences (All targets, discard low-efficiency and orphan targets).

[00174] **Figure 60** shows Genome Cleavage Efficiency of PAM Sequences (Cleaved targets, discard low-efficiency and orphan targets).

[00175] **Figure 61** shows a Sequence Logo for Working Cleaved Spacers & PAMs (New endogenous genome test showing that T is not required).

[00176] **Figure 62** shows Liver Tissue Slice Immunohistochemistry Staining Image from AAV-CMV-EGFP and AAV-CMV-SaCas9-U6-sgRNA(Pcsk9) injected animal (Verification of SaCas9 protein expression, 2 weeks post injection).

[00177] **Figure 63** shows Cleavage of Liver Tissue by SaCas9 delivered via tail-vein injection of AAV2/8 virus(1 week time points)

[00178] **Figure 64** shows a Time Course Assay for Cleavage of Liver Tissue by SaCas9 delivered via tail-vein injection of AAV2/8 (AAV2/8-SaCas9-U6-sgRNA(Pcsk9)) virus

[00179] **Figure 65** shows screening for functional CRISPR/Cas targets in human 293FT cells after delivery of SaCas9 and U6-sgRNA cassette targeting human SERPINA1 gene loci, followed by surveyor assay and gel analysis of 12 of the total 24 different spacer designs of sgRNA-expressing dsDNA targeting human SERPINA1 gene, the DNA Ladder is to the left.

Date Reçue/Date Received 2020-10-05

[00180] **Figure 66** shows gel analysis of 12 samples, for each of the 6 spacer designs of sgRNA-expressing dsDNA were co-transfected with SaCas9 plasmid into Mouse Hepatocyte cell line, two replica were placed next to each other. The DNA Ladder is to the left.

[00181] **Figure 67** shows Acute dissected liver tissue from mouse injected with TBG version vs. CMV version of EGFP (6 days post injection, GFP channel image, 10X).

[00182] **Figure 68A-B** shows (A) Design of AAV vector for packaging of SaCas9 and guide RNA expression systems with the ubiquitous mammalian CMV promoter for delivery into a wide range of tissues. (B) Design of AAV vector for packaging of SaCas9 and guide RNA expression systems with the liver-specific TBG promoter for targeting hepatocytes in vivo. ITR, AAV inverted terminal repeats. hSaCas9, human codon optimized SaCas9. NLS, nuclear localization signal. HA, Human influenza hemagglutinin derived tag. bGHpA, bovine growth hormone polyadenylation signal. U6, human U6 promoter. sgRNA, single-guide RNA.

[00183] **Figure 69A-B** shows (A) Surveyor assay results showing genomic modification rate for liver tissues from mouse injected with AAV2/8 expressing SaCas9 targeting mouse Pcsk9 gene or control AAV2/8 virus expressing EGFP reporter gene. All samples were taken 1 wk after tail vein injection. (B) Statistics summarizing cleavage efficiency from all three time points collected from mouse injected with AAV2/8 expressing either SaCas9 targeting mouse Pcsk9 gene.

[00184] **Figure 70A-D** shows a biochemical screen for small Cas9 orthologs. (a) Phylogenetic tree of Cas9 orthologs, with subfamily and sizes (amino acids) indicated. Conserved nuclease domains are highlighted in colored boxes, black residues represent conserved sequences. (b) Schematic illustrating in vitro cleavage-based method used to identify protospacer adjacent motifs (PAMs). (c) Consensus PAMs for eight Cas9 orthologs from sequencing of cleaved fragments. (d) Biochemical cleavage reaction using orthologs and sgRNAs targeting different loci bearing the putative PAMs (shown in red). Red triangles indicate cleavage fragments.

[00185] **Figure 71A-F** shows *in vitro* characterization of *Staphylococcus aureus* Cas9. (a) Schematic showing the structure of S. aureus sgRNA. Indels vary depending on (b) length of guide sequence or (c) repeat:anti-repeat duplex. (d) Consensus PAM for SaCas9 in HEK 293FT cells. Pooled indel values for all putative PAM 4-base combinations (top, $n \geq 3$) and overall sequence logo ($n = 116$, bottom) are shown. SpCas9 and SaCas9 cleavage efficiency comparison for e, genomic target sites and f, genome-wide off-target loci (error bars indicate Wilson

Date Reçue/Date Received 2020-10-05

intervals). Off-target (OT) sequences with significant indels are highlighted above graph. n = 3, error bars S.E.M unless otherwise noted; N.D. not detectable.

[00186] **Figure 72A-E** shows AAV delivery of *S. aureus* Cas into live animals. (a) Schematics illustrating AAV single-vector system (top) and experimental timeline (bottom). (b) Mouse Pcsk9 locus showing SaCas9 target locations. Guide sequences are highlighted in blue with PAM in magenta. (c) Time course of liver tissue indel formation at targets 1 and 6 post injection of AAV2/8 particles (up to 2 animals each; error bars represent liver tissue pieces). (d) Indel formation at target 6 at 1 and 3 weeks post-injection. Each lane represents a piece of liver tissue. Red triangles indicate cleavage fragments. (e) Representative chromatogram and indels generated by SaCas9 *in vivo*.

[00187] **Figure 73** shows a schematic of CRISPR-Cas loci of six orthologs from two subfamilies of Type II CRISPR-Cas systems. Spacer or "guide" sequences are shown in blue, followed by direct repeat (gray). Predicted tracrRNAs are shown in red, and folded based on the Constraint Generation RNA folding model.

[00188] **Figure 74** shows a stacked bar graph indicating the fraction of targets cleaved at 2, 3, 4, or 5-bp upstream of PAM for each Cas9 ortholog; all Cas9s cleave most frequently at 3-bp upstream of PAM (red triangle).

[00189] **Figure 75A-B** shows: (a) SURVEYOR assays showing indel formation at human endogenous loci from co-transfection of Cas9 orthologs and sgRNA in HEK 293FT cells. (b) SaCas9 cleaves multiple targets with high efficiency. PAM sequences for individual targets are shown above each lane, with consensus sequences for each Cas9 highlighted in red. Red triangles indicate cleaved fragments.

[00190] **Figure 76A-B** shows: (a) histograms of distances between adjacent *Staphylococcus aureus* subsp. aureus Type II CRISPR PAM (NNGRR) in the human genome (GRCh38). (b) Distances for each PAM by chromosome.

[00191] **Figure 77A-B** shows the location of SaCas9 targets and PAMs within the mouse Pcsk9 gene locus. b, Indels produced at target sites from transfection of mouse liver hepatoma (Hepa1-6) cell line. Red arrows indicate cleavage sites.

[00192] Figure 78A shows that guide (target) 1 induced the highest percentage of indels in ApoB.

Date Reçue/Date Received 2020-10-05

[00193] Figure 78B shows the results of a Surveyor nuclease gel assay for indel formation efficiency, 4 weeks post-injection.

[00194] Figure 79B shows oil red staining to detect hepatic lipid accumulation phenotype in vivo following AAV-Cas9-sgRNA delivery. The scale bar in each square represents 20 micrometres.

[00195] Figure 80 shows that 21 nucleotudes nts/base pairs (bp), represented by the grey bars is the optimal spacer length, at least compared to 20 or 22 base pairs (represented by the black and the white bars, respectively) across a range of targets and within two different genes (AAVS1 and EMX1).

[00196] Figure 81 shows whether a guide sequence could be inserted into the Cas9 intronic sequence

[00197] Figure 82 shows that the full-length H1 promoter (grey bar) is still weaker than U6 promoter (black bar), as the U6 shows increased indel percentage formation for each target tested.

[00198] Figure 83 shows that short H1 promoter is weaker than the full-length H1

[00199] Figure 84 shows distance between the 5' ends of two guide sequences in a construct measured in relation to the cleavage efficiency of the D10A SaCAs9 double nickase.

[00200] Figure 85 (Example 40) shows CRISPR-Cas9 system delivery and targeting of *Mecp2* locus in the mouse brain. (a) AAV-SpCas9 and AAV-SpGuide(*Mecp2*) expression vectors. The sgRNA vector contains encoding sequence of the GFP-KASH fusion protein for identification of transduced neurons. (**b**) Expression of HA-Cas9 and GFP-KASH in the dorsal dentate gyrus (DG) of mouse hippocampus. Scale bar, 100 μm. (**c**) Quantification of cells efficiently targeted by the dual-vector Cas9-CRISPR system. (**d**) Graphical representation of the mouse *Mecp2* locus showing Cas9 target location; sgRNA indicated in blue. PAM sequence marked in purple. Representative mutation patterns detected by sequencing of *Mecp2* locus were shown below: green - wild-type sequence; red dashes - deleted bases; red bases: insertion or mutations; red arrowhead indicates CRISPR-Cas9 cutting site. (**e**) SURVEYOR™ assay gel showing modification of the *Mecp2* locus, 2 weeks after AAV delivery in the DG region. (**f**) Western blot analysis of MeCP2 protein expression in the targeted brain region and quantification of MeCP2 protein levels in dorsal DG (*t*-test, $^{**}p<0.001$, n=4 from 3 animals, error bars: s.e.m.). (**g**) Images of the dorsal DG region, 2 weeks after CRISPR-Cas9 targeting of *Mecp2* locus. Scale bar, 150

Date Reçue/Date Received 2020-10-05

μm. (**h**) Quantification of MeCP2 positive cells population within all detected cells (DAPI staining) in the targeted brain region in compare to control collateral site (*t*-test, ****$p<0.0001$, n=290 and 249 cells from 2 animals, respectively; error bars: s.e.m). (ITR – inverted terminal repeat; HA – hemagglutinin tag; NLS – nuclear localization signal; spA – synthetic polyadenylation signal; U6 – PolIII promoter; sgRNA – single guide RNA; hSyn – human synapsin 1 promoter; GFP- green fluorescent protein; KASH - Klarsicht, ANC1, Syne Homology nuclear transmembrane domain; bGH pA – bovine growth hormone polyadenylatio signal; WPRE – Woodchuck Hepatitis virus posttranscriptional regulatory element).

[00201] Figure 86 (Example 40) shows analysis of gene expression in Cas9-mediated MeCP2 knockdown neurons. (a) Strategy for cell nuclei purification of CRISPR-Cas9 targeted cells from the mouse brain. (**b**) Hierarchical clustering of differentially expressed genes (*t*-test, $p<0.01$, n=19 populations of sorted nuclei from 8 animals) detected by RNAseq. Relative log2(TPM+1) expression levels of genes are normalized for each row and displayed in red-blue color scale. Each column represents a population of targeted 100 neuronal nuclei FACS sorted from the isolated, dentate gyrus population of cells, either from control or *Mecp2* sgRNA transduced animals, as indicated.

[00202] Figure 87 (Example 40) shows cell-autonomous defects in cellular response properties of neurons after CRISPR-mediated MeCP2 knockdown. (a) Cartoon showing *in vivo* experiment configuration from mouse visual cortex and visual stimulation parameter. GFP^+ neuron is shown. Scale bar, 20 μm. (b) Cartoon showing recording configuration in layer 2/3 excitatory neurons that receive both contra- and ipsilateral eye specific input. Genome modified GFP^+ cells are in green whereas unmodified cells are in gray. Normalized spike shape shows regular spiking excitatory neurons. (c,d) Average OSI (c) and evoked FR (d) were measured from GFP^+ cells expressing *Mecp2* and control sgRNA, respectively (*t*-test, *$p<0.05$; numbers in graph indicate numbers of recorded cells; n=2-3 animals; error bars: s.e.m).

[00203] Figure 88 (Example 40) shows simultaneous, multiplex gene editing in the mouse brain. (a) Schematic illustration of CRISPR-Cas9 system designed for multiplex genome targeting. (b) Graphical representation of targeted DNMT mouse loci. Guide RNAs are indicated in blue. PAM sequences are marked in purple. (c) SURVEYOR™ assay gel showing modification of DNMTs loci in FACS sorted GFP-KASH positive cells, 4 weeks after AAV delivery in the DG region. (d) Deep sequencing-based analysis of DNMTs loci modification in

Date Reçue/Date Received 2020-10-05

single cells, showing co-occurrence of modification in multiple loci. (e) Western blot analysis for Dnmt3a and Dnmt1 proteins after in vivo delivery of CRISPR-Cas9 system targeting DNMT family genes (top). Western blot quantification of Dnmt3a and Dnmt1 protein levels in DG after in vivo CRISPR-Cas9 targeting (bottom; t-test, **p<0.001, *p<0.05, Dnmt3a: n=7; Dnmt1: n=5 from 5 animals; error bars: s.e.m). (f) Contextual learning deficits, 8 weeks after targeting of DNMT genes using SpCas9 in the DG region of hippocampus, tested in training and altered context (t-test, ***p<0.0001, n=18 animals, 2 independent experiments; error bars: s.e.m).

[00204] Figure 89 (Example 40) shows cloning and expression of HA-tagged SpCas9 (HA-SpCas9) for AAV packaging. (a) Schematic overview of different cloning strategies to minimize SpCas9 expression cassette size using short rat *Map1b* promotor (pMap1b), a truncated version of the mouse *Mecp2* promoter (pMecp2) and a short polyA motif (spA). (**b**) Western blot analysis of primary cortical neuron culture expressing HA-SpCas9 using different SpCas9 expression cassettes. (**c**) *Mecp2* promoter drives HA-SpCas9 (red) expression in neurons (Map1b, NeuN; arrows) but not in astroglia (GFAP, arrowheads). Co-expressioin of HA-SpCas9 with GFP-KASH is shown (bottom). Nuclei were labeled with DAPI (blue). Scale bars, 20 μm. (**d**) Schematic overview of GFP-labeling. Enhanced green fluorescent protein (GFP) fused to the nuclear transmembrane KASH domain and integration of GFP-KASH to the outer nuclear membrane is illustrated. (**e**) Co-infection efficiency calculation, showing populations of cell expressing both HA-SpCas9 and GFP-KASH (n=973 neurons from 3 cultures; error bars: s.e.m). (**f**) Cells were stained with LIFE/DEAD® kit 7 days after virus delivery. Quantification of $DAPI^+$ and dead ($DEAD^+$) cells (control n=518 $DAPI^+$ nuclei; SpCas9/GFP-KASH n=1003 $DAPI^+$ nuclei from 2 cultures; error bars: s.e.m). (ITR – inverted terminal repeat; HA – hemagglutinin tag; NLS – nuclear localization signal; spA – synthetic polyadenylation signal; U6 – PolIII promoter; sgRNA – single guide RNA; hSyn – human synapsin 1 promoter; GFP- green fluorescent protein; KASH - Klarsicht, ANC1, Syne Homology nuclear transmembrane domain; bGH pA – bovine growth hormone polyadenylation signal; WPRE – Woodchuck Hepatitis virus posttranscriptional regulatory element).

[00205] Figure 90 (Example 40) shows targeting of *Mecp2* in Neuro-2a cells. (a) *Mecp2* targeting sequences and corresponding protospacer adjacent motifs (PAM). (b) Evaluation of 6 *Mecp2* sgRNAs co-transfected with SpCas9 into Neuro-2a cells. Locus modification efficiencies were analyzed 48 h after transfection using SURVEYOR™ assay.

Date Reçue/Date Received 2020-10-05

[00206] Figure 91 (Example 40) shows CRISPR-SpCas9 targeting of *Mecp2* in primary cortical neurons. (a) Immunofluorescent staining of MeCP2 (red) in cultured neurons 7 days after AAV-CRISPR transduction (green, GFP-KASH). Nuclei were labeled with DAPI (blue). Scale bar, 20 μm. (**b**) Evaluation of *Mecp2* locus targeting using SpCas9 or dSpCas9, together with *Mecp2* sgRNA or control (targeting bacterial *lacZ* gene) sgRNA, using SURVEYOR™ assay gel. (**c**) Quantification of MeCP2 positive nuclei in targeted population of neurons (GFP^+). (**d**) Western blot of MeCP2 protein levels after CRISPR-SpCas9 targeting of *Mecp2* locus and quantification of MeCP2 protein levels (*t*-test, $^{**}p<0.001$, n=5 from 3 cultures, error bars: s.e.m).

[00207] Figure 92 (Example 40) shows morphological changes in dendritic tree of neurons after SpCas9-mediated MeCP2 knockdown *in vitro*. (a) Reduced complexity of dendritic tree in neurons after CRISPR-SpCas9 targeting of *Mecp2* locus. Scale bar, 20 μm. (b) Changes in dendritic spines morphology in neurons targeted with SpCas9 and *Mecp2* sgRNA. Scale bar, 10 μm. Morphology of cells was visualized with co-transfection with mCherry construct. Cells for morphology analysis were chosen based on the result of Mecp2 staining. (c) Dendritic tree morphology assessed with number of dendritic ends and (d) Sholl analysis (*t*-test, $^{***}p<0.0001$, n=40 from 2 cultures). (e) Spine density quantification (*t*-test, $^{***}p<0.0001$, n=40 from 2 cultures, error bars: s.e.m).

[00208] Figure 93 (Example 40) shows RNAseq of neuronal nuclei from control animals and SpCas9-mediated Mecp2 knockdown. Box plot presenting the number of detected genes across the RNA-seq libraries (19 libraries each of 100 nuclei taken from control sgRNA or *Mecp2* sgRNA transduced nuclei; n=4 animals/group) per quantile of expression level. All genes are divided to 10 quantiles by their mean log2(TPM+1) expression level, then for each quantile the number of genes that are detected (log2(TPM+1)>2) was counted in each sample. The three target sequences shown are SEQ ID NO: 2 , SEQ ID NO: 3 and SEQ ID NO: 4 , for Dnmt3a, Dnmt1 and Dnmt3b, respectively.

[00209] Figure 94 (Example 40) shows multiplex genome targeting of DNMT family members *in vitro*. (a) *Dnmt3a*, *Dnmt1* and *Dnmt3b* targeting sequences and corresponding protospacer adjacent motifs (PAM). (b) SURVEYOR™ nuclease assay analysis of Neuro-2a cells 48 hours after transfection with SpCas9 and DNMT 3xsgRNA vector targeting *Dnmt3a*, *Dnmt1* and *Dnmt3b* loci. Efficient genome editing of all three targeted genes is shown.

Date Reçue/Date Received 2021-09-28

[00210] Figure 95 (Example 40) shows next generation sequencing of targeted *Dnmt3a*, *Dnmt1* and *Dnmt3b* loci. Examples of sequencing results of mutated *Dnmt3a* (a), *Dnmt1* (b) and *Dnmt3b* (c) loci after *in vivo* delivery of SpCas9 and DNMT 3xsgRNA into the mouse dentate gyrus. Green: wild-type sequence, red dashes: deleted bases, red bases: insertion or mutations. Red arrowheads indicate CRISPR-SpCas9 cutting site. The full sequences used in this figure are provided as SEQ ID NO: 2, SEQ ID NO: 3, and SEQ ID NO: 4 for the Dnmt3a, the Dnmt1 and the Dnmt3b loci, respectively. They are:

[00211] SEQ ID NO: 2 (Dnmt3a)

CCT CCG TGT CAG CGA CCC ATG CCA A

[00212] SEQ ID NO: 3 (Dnmt1)

CCA GCG TCG AAC AGC TCC AGC CCG

[00213] SEQ ID NO: 4 (Dnmt3b)

AGA GGG TGC CAG CGG GTA TAT GAG G

[00214] Figure 96 shows SaCas9 protein sequences are codon optimized ("reopt") and have their ubiquitination signals removed ("reopt(Ub)") for enhanced expression. Protein blots against FLAG- and HA-tagged SaCas9 show approximately 2-fold increased expression of optimized SaCas9 (reopt, #2-4) relative to the original constructs (#0, 5, and 6), and similar level as SpCas9 (SpCas9 330, top bar left panel; SpCas9 414, top bar right panel). The addition of 3xHA tagging (right panel #6) improves detection signal over the 1xHA tag (right panel #5) by ~2 fold.

[00215] Figure 97 shows indel efficiency using sgRNAs transcribed by U6 promoter as-is (grey, left hand bar for each number of nts) or appending a "G" (blue, right hand bar for each number of nts and with a thicker border) to 5'-most position of sgRNA for SaCas9. Total sgRNA spacer lengths (including G) are indicated on the x-axis. Graph represents with aggregated data from 5 sgRNAs.

[00216] Figure 98 shows optimization of sgRNA spacer length (x axis). Graphs show indel formation with different lengths of sgRNA spacer in HEK (left) and Hepa (right) cells.

[00217] The figures herein are for illustrative purposes only and are not necessarily drawn to scale.

DETAILED DESCRIPTION OF THE INVENTION

[00218] With respect to general information on CRISPR-Cas Systems: Reference is made to US provisional patent applications 61/758,468; 61/802,174; 61/806,375; 61/814,263; 61/819,803

Date Reçue/Date Received 2021-09-28

and 61/828,130, filed on January 30, 2013; March 15, 2013; March 28, 2013; April 20, 2013; May 6, 2013 and May 28, 2013 respectively. Reference is also made to US provisional patent application 61/836,123, filed on June 17, 2013. Reference is also made to US provisional patent applications 61/736,527 and 61/748,427, filed on December 12, 2012 and January 2, 2013, respectively. Reference is also made to US provisional patent application 61/791,409, filed on March 15, 2013. Reference is also made to US provisional patent application 61/799,800, filed March 15, 2013. Reference is also made to US provisional patent applications 61/835,931, 61/835,936, 61/836,127, 61/836, 101, 61/836,080 and 61/835,973, each filed June 17, 2013. Further reference is made to US provisional patent applications 61/862,468 and 61/862,355 filed on August 5, 2013; 61/871,301 filed on August 28, 2013; 61/960,777 filed on September 25, 2013 and 61/961, 980 filed on October 28, 2013. Further reference is made to US provisional patent application 61/915,325, filed on December 12, 2013. Each of these applications, and all documents cited therein or during their prosecution ("appln cited documents") and all documents cited or referenced in the appln cited documents, together with any instructions, descriptions, product specifications, and product sheets for any products mentioned therein or in any document therein

may be employed in the practice of the invention.

|00219| Also with respect to general information on CRISPR-Cas Systems, mention is made of:

- *Multiplex genome engineering using CRISPR/Cas systems*. Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., & Zhang, F. *Science* Feb 15;339(6121):819-23 (2013);
- *RNA-guided editing of bacterial genomes using CRISPR-Cas systems*. Jiang W., Bikard D., Cox D., Zhang F, Marraffini LA. Nat Biotechnol Mar;31(3):233-9 (2013);
- *One-Step Generation of Mice Carrying Mutations in Multiple Genes by CRISPR/Cas-Mediated Genome Engineering*. Wang H., Yang H., Shivalila CS., Dawlaty MM., Cheng AW., Zhang F., Jaenisch R. Cell May 9;153(4):910-8 (2013);

Date Reçue/Date Received 2020-10-05

- *Optical control of mammalian endogenous transcription and epigenetic states.* Konermann S, Brigham MD, Trevino AE, Hsu PD, Heidenreich M, Cong L, Platt RJ, Scott DA, Church GM, Zhang F. Nature. 2013 Aug 22;500(7463):472-6. doi: 10.1038/Nature12466. Epub 2013 Aug 23;
- *Double Nicking by RNA-Guided CRISPR Cas9 for Enhanced Genome Editing Specificity.* Ran, FA., Hsu, PD., Lin, CY., Gootenberg, JS., Konermann, S., Trevino, AE., Scott, DA., Inoue, A., Matoba, S., Zhang, Y., & Zhang, F. *Cell* Aug 28. pii: S0092-8674(13)01015-5. (2013);
- *DNA targeting specificity of RNA-guided Cas9 nucleases.* Hsu, P., Scott, D., Weinstein, J., Ran, FA., Konermann, S., Agarwala, V., Li, Y., Fine, E., Wu, X., Shalem, O., Cradick, TJ., Marraffini, LA., Bao, G., & Zhang, F. *Nat Biotechnol* doi:10.1038/nbt.2647 (2013);
- *Genome engineering using the CRISPR-Cas9 system.* Ran, FA., Hsu, PD., Wright, J., Agarwala, V., Scott, DA., Zhang, F. Nature Protocols Nov;8(11):2281-308. (2013);
- *Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells.* Shalem, O., Sanjana, NE., Hartenian, E., Shi, X., Scott, DA., Mikkelson, T., Heckl, D., Ebert, BL., Root, DE., Doench, JG., Zhang, F. *Science* Dec 12. (2013). [Epub ahead of print];
- *Crystal structure of cas9 in complex with guide RNA and target DNA.* Nishimasu, H., Ran, FA., Hsu, PD., Konermann, S., Shehata, SI., Dohmae, N., Ishitani, R., Zhang, F., Nureki, O. *Cell* Feb 27. (2014). 156(5):935-49;
- *Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells.* Wu X., Scott DA., Kriz AJ., Chiu AC., Hsu PD., Dadon DB., Cheng AW., Trevino AE., Konermann S., Chen S., Jaenisch R., Zhang F., Sharp PA. *Nat Biotechnol.* (2014) Apr 20. doi: 10.1038/nbt.2889, and
- *Development and Applications of CRISPR-Cas9 for Genome Engineering*, Hsu et al, Cell 157, 1262-1278 (June 5, 2014) (Hsu 2014),

discussed briefly below:

- Cong *et al.* engineered type II CRISPR/Cas systems for use in eukaryotic cells based on both *Streptococcus thermophilus* Cas9 and also *Streptoccocus pyogenes* Cas9 and demonstrated that Cas9 nucleases can be directed by short RNAs to induce precise cleavage of DNA in human and mouse cells. Their study further showed that Cas9 as converted into a nicking enzyme can be used to facilitate homology-directed repair in eukaryotic cells with minimal mutagenic activity. Additionally, their study demonstrated that multiple guide sequences can be encoded into a

Date Reçue/Date Received 2020-10-05

single CRISPR array to enable simultaneous editing of several at endogenous genomic loci sites within the mammalian genome, demonstrating easy programmability and wide applicability of the RNA-guided nuclease technology. This ability to use RNA to program sequence specific DNA cleavage in cells defined a new class of genome engineering tools. These studies further showed that other CRISPR loci are likely to be transplantable into mammalian cells and can also mediate mammalian genome cleavage. Importantly, it can be envisaged that several aspects of the CRISPR/Cas system can be further improved to increase its efficiency and versatility.

- Jiang *et al.* used the clustered, regularly interspaced, short palindromic repeats (CRISPR)–associated Cas9 endonuclease complexed with dual-RNAs to introduce precise mutations in the genomes of *Streptococcus pneumoniae* and *Escherichia coli.* The approach relied on dual-RNA:Cas9-directed cleavage at the targeted genomic site to kill unmutated cells and circumvents the need for selectable markers or counter-selection systems. The study reported reprogramming dual-RNA:Cas9 specificity by changing the sequence of short CRISPR RNA (crRNA) to make single- and multinucleotide changes carried on editing templates. The study showed that simultaneous use of two crRNAs enabled multiplex mutagenesis. Furthermore, when the approach was used in combination with recombineering, in *S. pneumoniae*, nearly 100% of cells that were recovered using the described approach contained the desired mutation, and in *E. coli,* 65% that were recovered contained the mutation.
- Konermann *et al.* addressed the need in the art for versatile and robust technologies that enable optical and chemical modulation of DNA-binding domains based CRISPR Cas9 enzyme and also Transcriptional Activator Like Effectors
- As discussed in the present specification, the Cas9 nuclease from the microbial CRISPR-Cas system is targeted to specific genomic loci by a 20 nt guide sequence, which can tolerate certain mismatches to the DNA target and thereby promote undesired off-target mutagenesis. To address this, Ran *et al.* described an approach that combined a Cas9 nickase mutant with paired guide RNAs to introduce targeted double-strand breaks. Because individual nicks in the genome are repaired with high fidelity, simultaneous nicking *via* appropriately offset guide RNAs is required for double-stranded breaks and extends the number of specifically recognized bases for target cleavage. The authors demonstrated that using paired nicking can reduce off-target activity by 50- to 1,500-fold in cell lines and to facilitate gene knockout in mouse zygotes without

Date Reçue/Date Received 2020-10-05

sacrificing on-target cleavage efficiency. This versatile strategy enables a wide variety of genome editing applications that require high specificity.

- Hsu *et al.* characterized SpCas9 targeting specificity in human cells to inform the selection of target sites and avoid off-target effects. The study evaluated >700 guide RNA variants and SpCas9-induced indel mutation levels at >100 predicted genomic off-target loci in 293T and 293FT cells. The authors that SpCas9 tolerates mismatches between guide RNA and target DNA at different positions in a sequence-dependent manner, sensitive to the number, position and distribution of mismatches. The authors further showed that SpCas9-mediated cleavage is unaffected by DNA methylation and that the dosage of SpCas9 and sgRNA can be titrated to minimize off-target modification. Additionally, to facilitate mammalian genome engineering applications, the authors reported providing a web-based software tool to guide the selection and validation of target sequences as well as off-target analyses.
- Ran *et al.* described a set of tools for Cas9-mediated genome editing *via* non-homologous end joining (NHEJ) or homology-directed repair (HDR) in mammalian cells, as well as generation of modified cell lines for downstream functional studies. To minimize off-target cleavage, the authors further described a double-nicking strategy using the Cas9 nickase mutant with paired guide RNAs. The protocol provided by the authors experimentally derived guidelines for the selection of target sites, evaluation of cleavage efficiency and analysis of off-target activity. The studies showed that beginning with target design, gene modifications can be achieved within as little as 1–2 weeks, and modified clonal cell lines can be derived within 2–3 weeks.
- Shalem *et al.* described a new way to interrogate gene function on a genome-wide scale. Their studies showed that delivery of a genome-scale CRISPR-Cas9 knockout (GeCKO) library targeted 18,080 genes with 64,751 unique guide sequences enabled both negative and positive selection screening in human cells. First, the authors showed use of the GeCKO library to identify genes essential for cell viability in cancer and pluripotent stem cells. Next, in a melanoma model, the authors screened for genes whose loss is involved in resistance to vemurafenib, a therapeutic that inhibits mutant protein kinase BRAF. Their studies showed that the highest-ranking candidates included previously validated genes NF1 and MED12 as well as novel hits NF2, CUL3, TADA2B, and TADA1. The authors observed a high level of consistency

Date Reçue/Date Received 2020-10-05

between independent guide RNAs targeting the same gene and a high rate of hit confirmation, and thus demonstrated the promise of genome-scale screening with Cas9.

- Nishimasu *et al.* reported the crystal structure of *Streptococcus pyogenes* Cas9 in complex with sgRNA and its target DNA at 2.5 A° resolution. The structure revealed a bilobed architecture composed of target recognition and nuclease lobes, accommodating the sgRNA:DNA heteroduplex in a positively charged groove at their interface. Whereas the recognition lobe is essential for binding sgRNA and DNA, the nuclease lobe contains the HNH and RuvC nuclease domains, which are properly positioned for cleavage of the complementary and non-complementary strands of the target DNA, respectively. The nuclease lobe also contains a carboxyl-terminal domain responsible for the interaction with the protospacer adjacent motif (PAM). This high-resolution structure and accompanying functional analyses have revealed the molecular mechanism of RNA-guided DNA targeting by Cas9, thus paving the way for the rational design of new, versatile genome-editing technologies.
- Wu *et al.* mapped genome-wide binding sites of a catalytically inactive Cas9 (dCas9) from *Streptococcus pyogenes* loaded with single guide RNAs (sgRNAs) in mouse embryonic stem cells (mESCs). The authors showed that each of the four sgRNAs tested targets dCas9 to between tens and thousands of genomic sites, frequently characterized by a 5-nucleotide seed region in the sgRNA and an NGG protospacer adjacent motif (PAM). Chromatin inaccessibility decreases dCas9 binding to other sites with matching seed sequences; thus 70% of off-target sites are associated with genes. The authors showed that targeted sequencing of 295 dCas9 binding sites in mESCs transfected with catalytically active Cas9 identified only one site mutated above background levels. The authors proposed a two-state model for Cas9 binding and cleavage, in which a seed match triggers binding but extensive pairing with target DNA is required for cleavage.
- Hsu 2014 is a review article that discusses generally CRISPR-Cas9 history from yogurt to genome editing, including genetic screening of cells, that is in the information, data and findings of the applications in the lineage of this specification filed prior to June 5, 2014. The general teachings of Hsu 2014 do not involve the specific models, animals of the instant specification.

[00220] The invention relates to the engineering and optimization of systems, methods and compositions used for the control of gene expression involving sequence targeting, such as

Date Reçue/Date Received 2020-10-05

genome perturbation or gene-editing, that relate to the CRISPR-Cas system and components thereof. In advantageous embodiments, the Cas enzyme is Cas9.

[00221] The CRISPRS-Cas polynucleotide sequence is generally referred to herein as the guide, or even as guide RNA (sgRNA), although it will be appreciated that this terminology was not as commonplace previously. Furthermore, reference is made herein to a CRISPR-Cas9 system, although it will be appreciated that the invention can be broadly practiced as to any CRISPR-Cas system. Advantageously the Cas has a nuclease function either to induce a DSB, a nick or a double nick. Cas9 is preferred and SaCas9 is particularly preferred.

[00222] Example 38 showed that both genotypic and, crucially, phenotypic changes are seen with CRISPR-Cas systems. Not only that, but the CRISPR-Cas9 system was effective at inducing a phenotypic change *in vivo*.

[00223] Specifically, the target was ApoB, a lipid metabolism gene. What is so encouraging is that ApoB can be said to be the "gold-standard" in liver delivery, and is widely used in mouse models of obesity.

[00224] Delivery was via intravenous injection. An AAV vector was used, as well as a Liver-specific promoter (TBG) for Cas9.

[00225] Delivery through expression from a viral vector as seen here is an improvement over Anderson/Yin's (NBT 2884) use of hydrodynamic delivery as the delivery method, because hydrodynamic delivery requires several mls of fluid to be injected which is stressful on the murine body and can be fatal. Hydrodynamic delivery is best suited for delivery of plasmid (naked) DNA, whereas we have shown that packaging the guide and Cas9 sequences within a viral delivery vector is preferable in terms of greatly increased efficiency. Indeed, only relatively small volumes need to be introduced, and this can be done intravenously (i.v.), which is likely to be much more acceptable therapeutically.

[00226] What was particularly encouraging was that not only was a genotypic change seen in a "gold-standard" gene for liver such as ApoB, but phenotypic changes were also recorded. Previous work with PCSK9 had shown genotypic, but not phenotypic changes, so the phenotypic changes seen with ApoB validate the plausibility of CRISPR delivery to, and its ability to effect phenotypic change in, the Liver. This is in combination with the more therapeutically acceptable means of delivery (i.v. compared to hydrodynamic delivery). As such, viral delivery of CRISPR-Cas9 system (guide and Cas9) is preferred, especially intravenously).

Date Reçue/Date Received 2020-10-05

[00227] Potential targets include: PCSK9, HMGCR, APOB, LDLR, ANGPTL3, F8, F9/FIX, AAT, FAH, HPD, TAT, ATP7B, UGT1A1, OTC, ARH.

[00228] Accordingly, provided are methods of inducing a phenotypic change *in vivo* comprising administering the CRISPR-Cas9 system to the target cells, for instance the liver. Suitable delivery routes are described herein but i.v. injection is preferred in some embodiments. Viral vectors are preferred, particularly AAV, in particular AAV serotype 2/8.

[00229] Also provided is a CRISPR-Cas9 system comprising one or more guides targeting lipid metabolism genes, for instance ApoB. Methods of treating obesity, comprising administering said CRISPR-Cas9 system, are also envisaged. A mouse model comprising one or more liver gene knock down(s), especially of lipid metabolism gene(s), for instance including ApoB, are preferred.

[00230] Liver specific promoters for the Cas9 will be apparent but may include those listed above. A preferred example is TBG.

[00231] As shown in Example 39, the guide may be 18-23 nucleotides in length. It may be 18-22, or 19-22, or 18-21, 20-22, but is preferably 22 and most preferably 21 nucleotides in length.

[00232] Also provided is proof of principle of successful packaging of a guide sequence into a SaCas9 intron. Accordingly, the CRISPR-Cas9 systems, wherein one or more guide sequences are packaged (positioned or inserted) into a Cas9 intron, are preferred.

[00233] The H1 promoter can be used and may be preferable in some circumstances.

[00234] Expanding on the work by Ran (Cell, 154, 21 Aug 2013), the degree of overlap in the dual guide approach using a D10A Double-Nickase was investigated. Optimal results were shown between -5 and +1 bp (5' to 5'). Accordingly, it is prefer to use a dual guide approach to minimise off target effects. These preferably overlap, or come close to overlapping, at their 5' ends, on different stands of DNA at the genomic target . Preferably, the overlap is in the range of -5 to +1 bp. In these instances, it will be appreciated that the Cas9 is a double nickase, such as the preferred D10A variant.

[00235] Example 40 provides, *inter alia*: a first demonstration of successful AAV-mediated Cas9 delivery in vivo as well as efficient genome modification in post-mitotic neurons; for the development of a nuclear tagging technique which enables easy isolation of neuronal nuclei from Cas9 and sgRNA-expressing cells; a demonstration of applications toward RNAseq analysis of

Date Reçue/Date Received 2020-10-05

neuronal transcriptome; how electrophysiological studies and other techniques can be integrated with Cas9-mediated genome perturbation to determine phenotypic changes; how electrophysiological studies and other techniques can be integrated with Cas9-mediated genome perturbation to determine phenotypic changes; how electrophysiological studies and other techniques can be integrated with Cas9-mediated genome perturbation to determine phenotypic changes; and a demonstration of multiplex targeting and the ability to study gene function on rodent behavior using Cas9-mediated genome editing.

[00236] The present invention provides: understanding and testing of gene function, including the creation and testing of models; including as to gene therapy and hence gene therapy, gene therapy methods and uses for gene therapy are within the ambit of the skilled person from this diclsoure.

[00237] An additional aspect, discussed further below, is in relation to a method for Nuclear Tagging.

[00238] It will be appreciated that reference to CRISPR-Cas9 systems herein is a short-hand for referring to the Cas9 enzymes provided herein in combination with the guides or guides used to target one or more genomic sequences. (And that the invention can also be broadly considered as to CRISPR-Cas systems.) Reference to guide(s) includes sgRNA, as well as the chimeric polynucleotide sequences described herein which comprises the guide sequences capabale of hybridising to target sequences in the genome of the subject, a tracr mate sequence and a tracr sequence.

[00239] The data essentially shows phenotypic changes resulting from gene knock down using two separate CRISPR-Cas9 systems according to the invention (guide RNA in combination with a Cas9 enzyme), in this case to successfully perturb gene function. The chosen tissue was brain tissue, but the results provide proof of principle for a wide range of post-mitotic tissues. This is an important distinction, because previous work has focussed on dividing cells (i.e. pre-mitotic).

[00240] In other words, whereas SpCas9 has been broadly used to engineer dividing cells, we demonstrate that SpCas9 can also be used to engineer the genome of postmitotic neurons. This is done with high efficiency via NHEJ-mediated indel generation to create knock downs, but therapeutic uses involving correction via the HDR mechanism (upon provision of a repair

Date Reçue/Date Received 2020-10-05

template) are also envisaged. Both are dependent on successful delivery and functional expression of the Cas9 and RNA guide or guides, which is shown here.

[00241] The fact that genotypic changes induced by the CRISPR-Cas9 systems then results in a phenotypic change is also important for both of the above areas (gene function and gene therapy).

[00242] The first CRISPR-Cas9 system employed guide sequences directed at (targeting) *Mecp2*. A dual vector CRISPR-Cas9 system, with one vector comprising the guide and one comprising the Cas9, was successfully employed showing further proof of principle for such dual vector systems. The dual vector CRISPR-Cas9 system was successfully delivered, via stereotactical injection, to two separate locations in the brain, specifically the Hippocampal dentate gyrus and the visual cortex. In both cases, gene perturbation of the same gene, *Mecp2,* was seen indicating that the dual vector system was successfully delivered and acted as expected, with transcription and functional activity in the Cas9 enzyme (in this case an SpCas9), and successful recruitment of the Cas9 to the genomic target sequence by the guide sequences.

[00243] AAV-mediated in vivo delivery of SpCas9 and sgRNA provides a rapid and powerful technology for achieving precise genomic perturbations within intact neural circuits. As such, the vector used was an AAV vector, adding further evidence for their use in general and in dual vector CRISPR-Cas9 systems in particular, especially in post-mitotic cells and tissues, and in particular in the brain.

[00244] It will of course be appreciated that the choice of promoter is important in achieving expression from the CRISPR-Cas9 system, in particular the Cas9 or both guide(s) and Cas9. Suitable examples for cell and cell lifecycle stage specificity can be determined from the literature. Nevertheless, we provide some non-limiting examples: TBG, a liver-specific promoter and is used here to drive expression of SaCas9; the H1 promoter; a truncated H1 promoter; the U6 promoter. Also, as guides do not necessarily need a specific promoter, one or more guides could similarly packaged into a/the Cas9 intron.

[00245] The second CRISPR-Cas9 system used included a multiplex approach. One key advantage of the SpCas9 system is its ability to facilitate multiplex genome editing. This second system successfully targeted three or more genes from the same family (in this case, *Dmnt1, 3a* and *3b*) by including suitable guides and resulted in stable knockouts of multiple genes. This has broad implications for probing the function of not only individual genes, but also whole gene

Date Reçue/Date Received 2020-10-05

families, in the tissues of living animals. This is particularly important for tissues such as the brain where this has not been possible before, or could only be achieved through long years of classical genetics. Applicants have shown that single or multiple gene perturbation (even complete knock down) can occur in post-mitotic cells in a normal animal. However, this could equally apply to a model organism (for instance one *already* carrying a gene mutation or perturbation or comprising altered expression of some kind) or a transgenic organism, lending a quick alternative to existing methods of producing model organisms and using model organisms to understand gene function. Further guides (and/or whole CRISPR-Cas9 systems) could be employed to make later rounds of gene perturbations and/or reinstatements (restoring gene function for instance by correction of the perturbed gene through provision, for instance, of a repair template, such as ssDNA suitable for HDR) within the same organism.

[00246] In fact, in general, SpCas9-mediated targeting of single or multiple genes can recapitulate morphological, electrophysiological, and behavioral phenotypes observed using classical, more time-consuming genetic mouse models.

[00247] Alternatively to knocking down whole gene families or related genes, the data here also provides proof of principle that simultaneous knock down or three or more unrelated genes is equally feasible. This is applicable across all tissues, but is particularly strongly presented in respect of post-mitotic tissues, especially the brain.

[00248] Another useful aspect of the work is that it showed that a combined, or integrated, approach could be taken to studying gene function, employing CRISPR to effect a genotypic change and then using classical tools such as electrophysiology (particularly relevant to brain and CNS tissue), biochemical, sequencing, electrophysiological, and/or behavioral readouts to establish what, if any, phenotypic changes result from the genotypic change induced by the CRISPR-Cas9 system. For example in the brain, this allows us to study the function of individual as wells as groups of genes in neural processes and their roles in brain disorders *in vivo*.

[00249] The successful perturbation of genes in this work is equally applicable to correction or reinstatement of gene function, i.e. the use of CRISPR-Cas9 systems in gene therapy. This is particularly in relation to targeting post-mitotic cells, especially the brain.

[00250] In general, the use of CRISPR-Cas9 systems show improvements over existing techniques such as Zn fingers, which take a long time to design and produce and cannot

Date Reçue/Date Received 2020-10-05

multiplex and shRNA, which has too many off-target effects whereas CRISPR off-target effects can be minimised by using double-nicakse approaches.

[00251] **Targeting of Tissues**

[00252] The new work supports the use of CRISPR-Cas9 systems to target genes in post-mitotic cells through delivery of the CRISPR-Cas9 system to the appropriate location (i.e. to cells within the organs or tissues of interest). Preferred tissues are within the following organs:

[00253] Kidney;

[00254] Digestive System including the stomach, pancreas, duodenum, ileum and/or colon;

[00255] Heart;

[00256] Lung;

[00257] Brain, in particular neurones, and/or CNS in general;

[00258] Eye, including retinal tissue;

[00259] Ear, including the inner ear;

[00260] Skin;

[00261] Muscle;

[00262] Bone; and/or

[00263] Liver in general, although this is excluded in some embodiments as it is also the subject of a separate application.

[00264] It will be appreciated that many of the above may comprise pre-mitotic cells, but that this aspect of the invention is directed to post-mitotic cells or tissues within those organs.

[00265] In particular, we prefer that the organ is the kidney or the brain. Within the brain, the data specifically shows delivery to the Hippocampal dentate gyrus and the visual cortex, which are preferred tissues, although other tissues including any one or more of the following: primary motor cortex, primary auditoty cortex, primary somatosensory cortex, cerebellum, main olfactory bulb, prefrontal cortex, endopiriform nucleus, amygdala, substantia nigra, striatum, pallidum, thalamus, hypothalamus, Parabranchial nucleus, superior olivary complex, cochlear nuclei, mammillary nuclei, are also preferred in some embodiments. Liver tissue are also preferred in some embodiments.

[00266] Cells from the brain, and neurones in particular, are especially preferred.

[00267] The choice of promoter to drive expression of the CRISPR-Cas9 system, especially the Cas9 is important, as mentioned above. To be considered when selecting a promoter are the

Date Reçue/Date Received 2020-10-05

cell cycle stage (early/late) and the cell type as promoters will be specific for one of more cell types and cell-cycle stages. Suitable promoters may include any one or more of the following, in some embodiments: Suitable promoters may include any one or more of the following, in some embodiments:

Cell Type	**Promoter**
Excitatory neurons	CamkII
Fast spiking interneurons	Parvalbumin
All interneurons	vGAT
Dopaminoceptive neurons	DR1
Dopaminoceptive neurons	DR2
Astroglia	GFAP
Activated neurons	Arc

[00268] The dual vector CRISPR-Cas9 system used in targeting the brain, in particular the Hippocampal dentate gyrus, packaged SpCas9 and sgRNA expression cassettes on two separate viral vectors. Cas9s, in particular SpCAs9s, are therefore preferably delivered by adenoviral vectors, especially AAV (i.e. as AAV-SpCas9). Guides are preferably delivered as sgRNA expression cassettes by adenoviral vectors, especially AAV (i.e. as AAV-SpGuide). A preferred route for this tissue (the Hippocampal dentate gyrus) and for the brain in general is stereotactical injection.

[00269] **Understanding And Testing Of Gene Function, And The Creation And Use Of Models To So Do**

[00270] Conditions include Huntington's, but essentially include any condition found in post-mitotic cells and especially those that may benefit from being studied *in vivo* or lack a useful model.

[00271] As mentioned above, CRISPR-Cas9 systems can be used to interrogate the function of one or more genes in post-mitotic cells. This may be achieved through delivery and expression of the CRISPR-Cas9 system to the post-mitotic cell, wherein the guide(s) of the CRISPR-Cas9 system are designed to recruit the Cas9 to the genomic target or targets of interest. Equally, where the Cas9 is already comprised within the post-mitotic cell, protein (transcribed)

Date Reçue/Date Received 2020-10-05

form, then delivery of the guides to the post-mitotic cell will suffice. Where the Cas9 is already comprised within the post-mitotic cell, in polynucleotide (untranscribed), then delivery of the guides to the post-mitotic cell as well as induction of transcription of the Cas9 polynucleotide will be necessary. Having the Cas9 under the control of an inducible or repressible promoter, such as the tet (tetracycline) on-off system may be advantageous here.

[00272] One aspect that is particularly promising is the integration of CRISPR techniques with phenotypic assays to determine the phenotypic changes, if any, resulting from gene perturbations, especially knock downs. For instance, Example 40 shows what can be achieved with targeted genomic perturbations coupled with quantitative readouts to provide insights into the biological function of specific genomic elements. In particular, Cas9-mediated in vivo genome editing in the brain can also be coupled with electrophysiological recording to study the effect of genomic perturbation on specific cell types or circuit components. In a broader sense, use of the CRISPR-Cas9 systems (to provide Cas9-mediated genomic perturbations) can be combined with biochemical, sequencing, electrophysiological, and behavioral analysis to study the function of the targeted genomic element.

[00273] Thus in one aspect, there is provided: a method of interrogating the function of one or more genes in a post-mitotic cell, comprising:

inducing a deficient genotype or gene knock down as described below; and

determining changes in expression of the one or more genes in the condition thereby interrogating the function of the one or more genes.

[00274] Optionally, the method may also include:

transplanting the second population of cells into the subject thereby inducing the condition associated with the deficient genotype or gene knock down. This may be prior to the determining step.

[00275] The following applies broadly to appropriate aspects of the invention. The cell may be in a subject, such as a human, animal or model organism, so that gene function is interrogated *in vivo*. However, it is also envisaged that the cell may be ex vivo, for instance in a cell culture or in a model organ or organoid. In some embodiments, the method may include isolation a first population of cells from the subject, optionally culturing them and transducing them with one or more CRISPR-Cas9 systems. Further optional culturing may follow. Transplantation of the transduced cells back into the subject may then occur.

Date Reçue/Date Received 2020-10-05

[00276] The cell may be from any of the tissues or organs described herein. The brain is one preferred example, providing for said method of interrogating the function of one or more genes, wherein the post-mitotic cell is a brain cell, for instance a neurone. Particularly in vivo, this allows for the interrogation of gene function on animal behaviour. The animal is preferably a mammal, for instance a rodent. Given the complexity of the nervous system, which consists of intricate networks of heterogeneous cell types, being able to efficiently edit the genome of neurons in vivo enables direct testing of gene function in relevant cell types embedded in native contexts. This is supported by our data where knockout mice showed impaired memory consolidation when tested under trained context conditions Our results demonstrate that CRIPSR-Cas9-mediated knockout of DNMT family members in dentate gyrus neurons is sufficient to probe the function of genes in behavioral tasks

[00277] This shows the versatility of Cas9s in facilitating targeted gene knockout in the mammalian brain in vivo, for studying genes functions and, in particular, for dissection of neuronal circuits. Introducing stable knockouts of multiple genes in the brain of living animals will have potentially far-reaching applications, such as causal interrogation of multigenic mechanisms in physiological and neuropathological conditions.

[00278] The specifics of this work are that we chose the mouse Mecp2 promoter (235 bp, pMecp2)7 and a minimal polyadenylation signal (48 bp, spA) based on their ability to achieve sufficient levels of SpCas9 expression in cultured primary mouse cortical neurons. Mecp2 gene, plays a principal role in Rett syndrome, a type of autism spectrum disorder. To target Mecp2, we first designed several sgRNAs targeting exon 3 of the mouse Mecp2 gene and evaluated their efficacy using Neuro-2a cells. The most efficient sgRNA was identified using the SURVEYOR nuclease assay. The delivery was via stereotactical injection of a mixture (1:1 ratio) of high titer AAV-SpCas9 and AAV-SpGuide. We also successfully tested the possibility of multiplex genome editing in the brain we designed a multiplex sgRNA expression vector consisting of three sgRNAs in tandem, along with GFP-KASH for nuclei labelling.

[00279] Thus, also provided are methods of inducing conditions involving one or more gene knockdowns in a post-mitotic cell. Examples of such conditions are numerous, but may include Rett syndrome, as exemplified. Suitable promoters will be apparent, and the Mecp2 promoter is ideal for Rett syndrome. One way to select a promoter to drive expression of the CRISPR-Cas9 system, in particular the Cas9, is to select the promoter for the gene of interest.

Date Reçue/Date Received 2020-10-05

[00280] Thus in one aspect, there is provided: A method of inducing a conditions involving one or more deficient genes (or genotyoes) or gene knockdowns in a post-mitotic cell, comprising:

transducing a first population of cells with a non-naturally occurring or engineered composition comprising a vector system comprising one or more vectors comprising

a first regulatory element operably linked to a CRISPR-Cas system chimeric RNA (chiRNA) polynucleotide sequence, wherein the polynucleotide sequence comprises

one or more, preferably three or more, guide sequences capable of hybridizing to three or more target sequences in genome of the subject,

a tracr mate sequence, and

a tracr sequence, and

a second regulatory element operably linked to an enzyme-coding sequence encoding a CRISPR enzyme comprising at list one or more nuclear localization sequences (NLSs), wherein (a), (b) and (c) are arranged in a 5' to 3' orientation,

wherein components I and II are located on the same or different vectors of the system, wherein when transcribed, the tracr mate sequence hybridizes to the tracr sequence and the guide sequence direct sequence-specific binding of CRISPR complexes to the target sequence,

wherein the CRISPR complex comprises the CRISPR enzyme complexed with (1) the guide sequence that is hybridized or hybridizable to the target sequence, and (2) the tracr mate sequence that is hybridized or hybridizable to the tracr sequence,

wherein the CRISPR enzyme alters the genome of the first population of cells to obtain a second population of cells bearing the one or more deficient genes or knocked down genes.

[00281] Optionally, the method may also include:

isolating a first population of cells from the subject.

[00282] Optionally, the method may also include:

transplanting the second population of cells into the subject thereby inducing the proliferative condition.

[00283] This essentially involves inducing a non-functional (which include partially non-functional) genotype into the target cell, to thereby provide a model for study (including future restoration of the functional genotype).

Date Reçue/Date Received 2020-10-05

[00284] CRISPR-Cas9 systems can also be used to facilitate the study of gene functions in cellular assays by enabling targeted knockout in post-mitotic neurons.

[00285] Methods for delivering nucleotides to neuronal cells are well known and reviewed in The Journal of Neuroscience, by Karra and Dahm (5 May 2010, 30(18): 6171-6177; doi: 10.1523/JNEUROSCI.0183-10.2010). Examples include electrical transfection methods (such as electroporation, nucleofection, and single-cell electroporation); chemical transfection methods (such as Ca2+ phosphate co/precipitation and lipofection); viral delivery (such as Adenoviral, Adeno-Associated Virus (AAV), Lentiviral and Herpes Simplex Virus); and physical transfection methods (such as microinjection and biolistics (DNA-coated gold particles). All of these can be used for delivery of the CRISPR-Cas9 system, but lipofection or viral methods are preferred, especially AAV or Lentiviral.

[00286] **Models**

[00287] Models are provided with single or multiple genes knocked down. An example would be a rodent model for Rett syndrome, a Mecp2 knock down. Others include Dmnt family knock downs, specifically Dmnt1, 3a and 3b knock downs. As such, models studying neurological conditions are provided. All that needs to be done is to identify the target genes of interest, design suitable guide(s) and include these in a suitable CRISPR-Cas9 system and deliver it to the post-mitotic cell(s) whether in vivo or ex vivo, as required. For instance, the models may have altered dendritic tree morphology and/or spine density are provided.

[00288] As mentioned above, models tissues are also provided, such as oraganoids or "Liver on a chip" or non-liver equivalents thereof such as ear, kidney and brain tissues, for instance on a chip or supported in a scaffold. Animal models and model tissues are preferred. These may be already transformed with Cas9 so that they comprise Cas9 in nucleotide or protein form, as mentioned above. These have the advantage that Cas9 does not need to be delivered alongside the guide(s) and this in turn may allow for a much greater degree of (guide) multiplexing to be accommodated within the delivery vectors. Again, use of inducible or repressible systems such as tet-on or tet-off, may be advantageous here.

[00289] Models obtainable using the CRISPR-Cas9 system are herein described and within the ambit of the skilled person from this disclosure and the knowledge in the art. Due to the versatility of the CRISPR-Cas9 system, the range of possible models, whether human, rodent,

Date Reçue/Date Received 2020-10-05

mammalian or otherwise is hugely diverse and this can be established by simple selection of appropriates guide(s). Methods of creating such models are also provided.

[00290] **Gene Therapy**

[00291] The data in Example 40 focuses on gene perturbation, primarily knock down. Gene knock down is likely to be only a small, if important, part of the total quorum of possible applications of CRISPR-Cas9 systems to gene therapy. As already shown in the Yin and Anderson paper (Nature Biotech 2884 published online 30 March 2014), a functional phenotype can be restored following correction of a deficient mutation in hereditary tyrosinemia type I (HTI), an otherwise fatal condition caused by mutation of fumarylacetoacetate hydrolase (FAH) (G to A in the last nucleotide in exon 8) which causes skipping of exon 8 during splicing and results in the formation of a truncated, unstable FAH protein, leading to accumulation of toxic metabolites. Correction of the A mutation back to the wild-type G geneotype resulted in a restored phenotype.

[00292] As such, the approaches taken in the present work can plausibly be applied to gene therapy. In particular, the dual vector approach, the nuclear tagging approach, the specifics of the brain delivery (the form of injection, the promoters and/or viral vectors used), as well as the multiplexing (use of multiple guides for multiple targets either within the same or within different genes) could equally be applied to correctional gene therapy (i.e. where a deficient genotype is corrected) as to the exemplified gene knock down. The main difference between correctional gene therapy and gene knock down is that in order to correct a deficient genotype, such as a point mutation (for instance in Cystic Fibrosis, see ref Schwank *et al*, Cell Stem Cell 13, 653-658 5 Dec 2013), it is advantageous to provide a repair template to stimulate the HDR mechanism and ideally provide a suitable Cas9 nickase as well.

[00293] Accordingly, the present vectors preferably target post-mitotic cells. Where the guide or guides target a deficient genotype, are preferably also provided with a repair template, for instance ssDNA corresponding to the corrected sequence (a genotype providing functional phenotype). Repair templates are described herein. The Cas9 may be provided in the same or a different vector from the guide or guides. The vectors are preferably viral vectors, more preferably adenoviral vectors and most preferably AAV vectors. Delivery to the cells is preferably by intravenous injection or by stereotactic injection, as appropriate. The selection of the promoter can also be important and preferred examples are provided herein.

Date Reçue/Date Received 2020-10-05

[00294] Methods of treating genetic diseases or conditions caused by, or associated with, a deficient genotype in post-mitotic cells are provided, comprising delivery of the CRISPR-Cas9 system to the appropriate cell. A deficient genotype may be the non-wild type genotype. In particular, single point mutations and/or monogenic disorders are especially suited to treatment using CRISPR-Cas9 systems. Where multiple genes require editing or correcting, then a multiplex approach may be used to target them all simultaneously. Alternatively, two or more rounds of different CRISPR-Cas9 systems could be envisaged. Preferably, the wild-type genotype is corrected for. It does not necessarily have to be the most common genotype, provided that function is restored or improved in the phenotype.

[00295] An example of a restored phenotype is the restoration of hearing to restore VGLUT3 function and hence hearing in the inner ear of mice (Omar Akil, Rebecca P. Seal, Kevin Burke, Chuansong Wang, Aurash Alemi, Matthew During, Robert H. Edwards, Lawrence R. Lustig. Restoration of Hearing in the VGLUT3 Knockout Mouse Using Virally Mediated Gene Therapy.*Neuron*, 2012; 75 (2): 283 DOI: 10.1016/j.neuron.2012.05.019). This was using AAV-mediated delivery of VGLUT3 itself, but it is entirely plausible that CRISPR-Cas9 system could also be used, preferably also using AAV vectors, to target the cells of the inner ear and correct the non-functional VGLUT3 genotype, with similar phenotypic consequences, namely restoration of hearing. As such, delivery of the CRISPR-Cas9 system to the inner ear, preferably using AAV vectors, is preferred, thus treating hearing loss. Indeed, restoration of gene function in sensory organs such as the eye, including the retina, nose and ear (particularly the inner ear) is preferred.

[00296] A relatively recent overview, which includes a discussion of disorders in post-mitotic tissues (eye, ear and beyond) is Kaufmann *et al* (EMBO Mol Med (2013(, 5, p1642-1661). This confirms the usefulness of AAV in the correction of monogenic disorders in post-mitotic tissues. It states that "in combination with other characteristics such as low inflammatory activity, they have shown to have an excellent safety profile and are therefore highly attractive tools for *in vivo* gene therapy. Indeed, Glybera® is a recombinant AAV for direct intramuscular injection…" The paper, with citations, reviews gene therapy in the retina, central nervous system, liver, skeletal and cardiac muscle as target tissues. And, with citations, indicates that "initial studies exploited the prototype AAV serotype 2 vector, the portfolio of AAV vectors has recently been

Date Reçue/Date Received 2020-10-05

expanded to include additional serotypes and even engineered capsids."

[00297] **RNAseq Analysis of the Transcriptome**

[00298] The combination of SpCas9-mediated genome perturbation and population level RNAseq analysis provides a way to characterize transcriptional regulation and suggest genes that may be important to specific functions or disease processes in the cells under consideration. In particular, the cells are from the brain, in particular neurones. Fast-acting techniques such as a CRISPR-Cas9 system are advantageous in studying the transcriptome, which is, by its nature, transient. As such, the use of CRISPR-Cas9 systems according to the present invention in analysis of the transcriptome (RNAseq) are provided.

[00299] **Nuclear Tagging Method**

[00300] To facilitate immunofluorescence identification of SpCas9-expressing neurons, we tagged SpCas9 with a HA-epitope tag (derived from human influenza hemagluttinin, a general epitope tag widely used in expression vectors).

[00301] For the AAV-SpGuide vector, we packaged an U6-sgRNA expression cassette as well as the green fluorescent protein (GFP)-fused with the KASH nuclear trans-membrane domain driven by the human Synapsin I promoter. The GFP-KASH fusion protein directs GFP to the outer nuclear membrane and enables fluorescence-based identification and purification of intact neuronal nuclei transduced by AAV-SpGuide.

[00302] Accordingly, the vectors of the present invention are preferably adapted in a similar fashion. Thus, the vectors are provided wherein the Cas9 is tagged with an epitope tag, such as the HA-epitope tag. The Cas9 may be any of the Cas9s described herein, for instance Sp or SaCas9 and may be any variant (such as D10A double nickase etc), provide that it is or can be tagged appropriately.

[00303] The vectors of the present invention may also be adapted so that the guide RNA is packaged within an expression cassette, which comprises:

a reporter protein; and

optionally, a suitable promoter for the guide RNA, such as U6;

wherein the reporter protein is fused with a nuclear trans-membrane domain operably linked to a suitable promoter therefor.

Date Reçue/Date Received 2020-10-05

[00304] The reporter protein is preferably a fluorescent protein, for instance one of green, red or yellow fluorescent proteins (GFP, RFP, YFP) and so forth.

[00305] Examples of nuclear trans-membrane domains include KASH-like domains, Sun2 domains, LEM domains. In some preferred embodiments, the nuclear trans-membrane domain is the KASH nuclear trans-membrane. Preferably, the promoter for the trans-membrane domain is the human Synapsin I promoter; see also documents cited herein.

[00306] This tagging approach may be used within single or dual vector systems, but preferably within dual vector systems as space is limited in single vector systems and the need for separate tags lessened as well.

[00307] Furthermore, each aspect of this tagging technique can be used independently of the other, so that epitope tagging of the Cas9 can be used alone, or the reporter/fluorescent protein cassette approach can be used alone, or more preferably both can be used together.

[00308] Multiple or repeat epitope Tags are preferred for the Cas9. In particular, a triple epitope tag was shown in Example 41 to improve detection. The tag is preferably a repeat, more preferably a triple repeat. HA is a preferred Cas9 epitope tag. A triple HA epitope tag is, therefore, preferred in some embodiments.

[00309] Kanasty and Anderson (Nature Materials, Vol 12 Nov 2013) is a useful review, initially submitted on 11 March 2013 and published online on 23 October 2013 of delivery of RNAi. Due to the similarities between RNAi and CRISPR guide sequences, the teaching of this and other art in respect of RNAi is informative for the mechanisms of delivering the guides in our CRISPR-Cas9 system. Some of the techniques described are also be suitable for delivery of the Cas9 as well. In some instance is may be useful to deliver the guides of our CRISPR-Cas9 system separately from the Cas9. This may be as part of a dual-vector delivery system, where the vectors are considered in the broadest light as simply any means of delivery, rather than specifically viral vectors. It is envisaged that the Cas9 may be delivered via a viral vector and that guides specific to genomic targets are delivered separately. As discussed herein, the guides could be delivered via the same vector types as the Cas9, for example a dual-vector system where the Cas9 is delivered in an AAV vector and the guide(s) are delivered in a separate AAV vector. This can be done substantially contemporaneously (i.e. co-delivery), but it could also be done at separate points in time, separated even by weeks or months. For example, if a first round of CRISPR-Cas9 systems have been delivered, but then it is subsequently required to provide

Date Reçue/Date Received 2020-10-05

further guides, then the original Cas9 which is hopefully still functional in the target cells may be re-used. If the Cas9 is under the control of an inducible promoter, then induction of transcription of new CAs9 in the target cells is preferred. Equally, if a CAs9-expressing model provided for herein is used, then only delivery of guide(s) is necessary. Accordingly, where delivery of guide(s) is required separately from Cas9, then it may be delivered in much the same way as RNAi. As such, the review by Kanasty is helpful in pointing out a number of known approaches that are suitable, with particular focus on the liver, although the means of delivery are generally appropriate for a broad range of cells. Examples include:

- "Liposomal delivery system, as well as siRNA conjugated to lipophilic molecules, interact with serum lipoproteins and subsequently gain entry into hepatocytes that take up those lipoproteins;"
- PEGylation;
- Conjugates such as:
    a. Dynamic Polyconjugates (DPCs, 10nm nanoparticles), which have been shown to deliver RNAi to successfully supress ApoB (thereby crossing over with our work on targeting ApoB via a CRISPR-Cas9 system); and
    b. Triantennary GalNAc conjugates
    c. are "both highly effective" especially GalNAc;
    - Other nanoparticles include:
        d. Cyclodextrin Polymer nanoparticles (CDP), including additional formulation components such as adamantine-PEG (AD-PEG) and adamantine-PEG-transferrin (AD-PEG-Tf) ;
        e. Lipid Nanoparticles (LNP), including cationic or ionisable lipids, shielding lipids, cholesterol and endogenous or exogenous targeting ligands. An example of an endogenous targeting ligand is Retinol Binding protein (RBP) useful for targeting hepatic and pancreatic stellate cells, which express the RBP receptor. An example of an exogenous targeting ligand is GalNac, which also targets the liver via the asialoglycoprotein receptor on hepatocytes. A combined approach is seen in Anlylams ALN-VSP;
- "Fenestrations in the liver endothelium allow molecules 100-200 nm in diameter to diffuse out of the bloodstream and gain access to the hepatocytes and other liver cells";

Date Reçue/Date Received 2020-10-05

- Ligands such as GalNAc are suitable for delivery to non-parenchymal liver cells expressing the mannose receptor, and to hepatocytes where conjugation of suitable siRNA to a GalNAc ligand has been shown to successfully supress PCSK9; and

Oligonucleotide nanoparticles (ONPs) composed of composed of complimentary DNA fragments designed to hybridise into a pre-defined 3D structure. Using suitable 3' overhand sequences, 6 siRNA strands could be attached to each particle, even at a specified position. The hydrodynamic diameter was about 29nm.

[00310] These approaches are preferred in some embodiments for delivery of at least the guides for a CRISPR-Cas9 system. Especially preferred are Dynamic Polyconjugates or the use of an endogenous targeting ligands such as Retinol Binding protein or exogenous targeting ligands such as GalNac.

[00311] An advantage of the present methods is that the CRISPR system avoids off-target binding and its resulting side effects. This is achieved using systems arranged to have a high degree of sequence specificity for the target DNA.

[00312] Cas9

[00313] Cas9 optimization may be used to enhance function or to develop new functions, one can generate chimeric Cas9 proteins. Examples that the Applicants have generated are provided in Example 6. Chimeric Cas9 proteins can be made by combining fragments from different Cas9 homologs. For example, two example chimeric Cas9 proteins from the Cas9s described herein. For example, Applicants fused the N-term of St1Cas9 (fragment from this protein is in bold) with C-term of SpCas9. The benefit of making chimeric Cas9s include any or all of: reduced toxicity; improved expression in eukaryotic cells; enhanced specificity; reduced molecular weight of protein, for example, making the protein smaller by combining the smallest domains from different Cas9 homologs; and/or altering the PAM sequence requirement.

[00314] The Cas9 may be used as a generic DNA binding protein. For example, and as shown in Example 7, Applicants used Cas9 as a generic DNA binding protein by mutating the two catalytic domains (D10 and H840) responsible for cleaving both strands of the DNA target. In order to upregulate gene transcription at a target locus Applicants fused a transcriptional activation domain (VP64) to Cas9. Other transcriptional activation domains are known. As shown in Example 17, transcriptional activation is possible. As also shown in Example 17, gene

Date Reçue/Date Received 2020-10-05

repression (in this case of the beta-catenin gene) is possible using a Cas9 repressor (DNA-binding domain) that binds to the target gene sequence, thus repressing its activity.

[00315] Cas9 and one or more guide RNA can be delivered using adeno associated virus (AAV), lentivirus, adenovirus or other plasmid or viral vector types, in particular, using formulations and doses from, for example, US Patents Nos. 8,454,972 (formulations, doses for adenovirus), 8,404,658 (formulations, doses for AAV) and 5,846,946 (formulations, doses for DNA plasmids) and from clinical trials and publications regarding the clinical trials involving lentivirus, AAV and adenovirus. For examples, for AAV, the route of administration, formulation and dose can be as in US Patent No. 8,454,972 and as in clinical trials involving AAV. For Adenovirus, the route of administration, formulation and dose can be as in US Patent No. 8,404,658 and as in clinical trials involving adenovirus. For plasmid delivery, the route of administration, formulation and dose can be as in US Patent No 5,846,946 and as in clinical studies involving plasmids. Doses may be based on or extrapolated to an average 70 kg individual, and can be adjusted for patients, subjects, mammals of different weight and species. Frequency of administration is within the ambit of the medical or veterinary practitioner (e.g., physician, veterinarian), depending on usual factors including the age, sex, general health, other conditions of the patient or subject and the particular condition or symptoms being addressed.

[00316] The viral vectors can be injected into the tissue of interest. For cell-type specific genome modification, the expression of Cas9 can be driven by a cell-type specific promoter. For example, liver-specific expression might use the Albumin promoter and neuron-specific expression might use the Synapsin I promoter.

[00317] Transgenic animals and plants

[00318] Transgenic animals (models) are also provided and the following applies equally to *ex vivo* model tissues and collections of tissues, such as organoids, liver on a chip and so forth. Preferred examples include animals comprising Cas9, in terms of polynucleotides encoding Cas9 or the protein itself. Mice, rats and rabbits are preferred. To generate transgenic mice with the constructs, as exemplified herein one may inject pure, linear DNA into the pronucleus of a zygote from a pseudo pregnant female, e.g. a CB56 female. Founders may then be identified, genotyped, and backcrossed to CB57 mice. The constructs may then be cloned and optionally verified, for instance by Sanger sequencing. Knock outs are envisaged where for instance one or more genes are knocked out in a model. However, are knockins are also envisaged (alone or

Date Reçue/Date Received 2020-10-05

in combination). An example knockin Cas9 mouse was generated and this is exemplified, but Cas9 knockins are preferred. To generate a Cas9 knock in mice one may target the same constitutive and conditional constructs to the Rosa26 locus, as described herein (Figs. 25A-B and 26). Methods of US Patent Publication Nos. 20120017290 and 20110265198 assigned to Sangamo BioSciences, Inc. directed to targeting the Rosa locus may be modified to utilize the CRISPR Cas system of the present invention. In another embodiment, the methods of US Patent Publication No. 20130236946 assigned to Cellectis directed to targeting the Rosa locus may also be modified to utilize the CRISPR Cas system of the present invention.

[00319] Utility of the conditional Cas9 mouse: Applicants have shown in 293 cells that the Cas9 conditional expression construct can be activated by co-expression with Cre. Applicants also show that the correctly targeted R1 mESCs can have active Cas9 when Cre is expressed. Because Cas9 is followed by the P2A peptide cleavage sequence and then EGFP Applicants identify successful expression by observing EGFP. Applicants have shown Cas9 activation in mESCs. This same concept is what makes the conditional Cas9 mouse so useful. Applicants may cross their conditional Cas9 mouse with a mouse that ubiquitously expresses Cre (ACTB-Cre line) and may arrive at a mouse that expresses Cas9 in every cell. It should only take the delivery of chimeric RNA to induce genome editing in embryonic or adult mice. Interestingly, if the conditional Cas9 mouse is crossed with a mouse expressing Cre under a tissue specific promoter, there should only be Cas9 in the tissues that also express Cre. This approach may be used to edit the genome in only precise tissues by delivering chimeric RNA to the same tissue.

[00320] As mentioned above, transgenic animals are also provided.In this regard, transgenic animals, especially mammals such as livestock (cows, sheep, goats and pigs), but also poultry and edible insects, are preferred.

[00321] Adeno associated virus (AAV)

[00322] In terms of *in vivo* delivery, AAV is advantageous over other viral vectors for a couple of reasons:

[00323] Low toxicity (this may be due to the purification method not requiring ultra centrifugation of cell particles that can activate the immune response)

[00324] Low probability of causing insertional mutagenesis because it doesn't integrate into the host genome.

Date Reçue/Date Received 2020-10-05

[00325] AAV has a packaging limit of 4.5 or 4.75 Kb. This means that Cas9 as well as a promoter and transcription terminator have to be all fit into the same viral vector. Constructs larger than 4.5 or 4.75 Kb will lead to significantly reduced virus production. SpCas9 is quite large, the gene itself is over 4.1 Kb, which makes it difficult for packing into AAV. Therefore embodiments of the invention include utilizing homologs of Cas9 that are shorter. For example:

[00326]

Species	Cas9 Size
Corynebacter diphtheriae	3252
Eubacterium ventriosum	3321
Streptococcus pasteurianus	3390
Lactobacillus farciminis	3378
Sphaerochaeta globus	3537
Azospirillum B510	3504
Gluconacetobacter diazotrophicus	3150
Neisseria cinerea	3246
Roseburia intestinalis	3420
Parvibaculum lavamentivorans	3111
Staphylococcus aureus	3159
Nitratifractor salsuginis DSM 16511	3396
Campylobacter lari CF89-12	3009
Streptococcus thermophilus LMD-9	3396

[00327] These species are therefore, in general, preferred Cas9 species. Applicants have shown delivery and *in vivo* mouse brain Cas9 expression data.

[00328] Two ways to package Cas9 coding nucleic acid molecules, e.g., DNA, into viral vectors to mediate genome modification *in vivo* are preferred:

[00329] To achieve NHEJ-mediated gene knockout:

[00330] Single virus vector:

Vector containing two or more expression cassettes:

Promoter-Cas9 coding nucleic acid molecule -terminator

Promoter-gRNA1-terminator

Date Reçue/Date Received 2020-10-05

Promoter-gRNA2-terminator

Promoter-gRNA(N)-terminator (up to size limit of vector)

[00331] Double virus vector:

Vector 1 containing one expression cassette for driving the expression of Cas9

Promoter-Cas9 coding nucleic acid molecule-terminator

Vector 2 containing one more expression cassettes for driving the expression of one or more guideRNAs

Promoter-gRNA1-terminator

Promoter-gRNA(N)-terminator (up to size limit of vector)

[00332] To mediate homology-directed repair. In addition to the single and double virus vector approaches described above, an additional vector is used to deliver a homology-direct repair template.

[00333] Promoter used to drive Cas9 coding nucleic acid molecule expression can include:

[00334] AAV ITR can serve as a promoter: this is advantageous for eliminating the need for an additional promoter element (which can take up space in the vector). The additional space freed up can be used to drive the expression of additional elements (gRNA, etc.). Also, ITR activity is relatively weaker, so can be used to reduce toxicity due to over expression of Cas9.

[00335] For ubiquitous expression, can use promoters: CMV, CAG, CBh, PGK, SV40, Ferritin heavy or light chains, etc.

[00336] For brain expression, can use promoters: SynapsinI for all neurons, CaMKIIalpha for excitatory neurons, GAD67 or GAD65 or VGAT for GABAergic neurons, etc.

[00337] For liver expression, can use Albumin promoter.

[00338] For lung expression, can use SP-B.

[00339] For endothelial cells, can use ICAM.

[00340] For hematopoietic cells can use IFNbeta or CD45.

[00341] For Osteoblasts can use OG-2.

[00342] Promoter used to drive guide RNA can include:

Pol III promoters such as U6 or H1

Use of Pol II promoter and intronic cassettes to express gRNA

[00343] As to AAV, the AAV can be AAV1, AAV2, AAV5 or any combination thereof. One can select the AAV of the AAV with regard to the cells to be targeted; e.g., one can select AAV

Date Reçue/Date Received 2020-10-05

serotypes 1, 2, 5 or a hybrid capsid AAV1, AAV2, AAV5 or any combination thereof for targeting brain or neuronal cells; and one can select AAV4 for targeting cardiac tissue. AAV8 is useful for delivery to the liver. The above promoters and vectors are preferred individually.

[00344] RNA delivery is also a useful method of *in vivo* delivery. Fig. 27 shows delivery and *in vivo* mouse brain Cas9 expression data. It is possible to deliver Cas9 and gRNA (and, for instance, HR repair template) into cells using liposomes or nanoparticles. Thus delivery of the CRISPR enzyme, such as a Cas9 and/or delivery of the RNAs of the invention may be in RNA form and via microvesicles, liposomes or nanoparticles. For example, Cas9 mRNA and gRNA can be packaged into liposomal particles for delivery *in vivo*. Liposomal transfection reagents such as lipofectamine from Life Technologies and other reagents on the market can effectively deliver RNA molecules into the liver.

[00345] Enhancing NHEJ or HR efficiency is also helpful for delivery. It is preferred that NHEJ efficiency is enhanced by co-expressing end-processing enzymes such as Trex2 (Dumitrache et al. Genetics. 2011 August; 188(4): 787–797). It is preferred that HR efficiency is increased by transiently inhibiting NHEJ machineries such as Ku70 and Ku86. HR efficiency can also be increased by co-expressing prokaryotic or eukaryotic homologous recombination enzymes such as RecBCD, RecA.

[00346] Various means of delivery are described herein, and further discussed in this section.

[00347] Viral delivery: The CRISPR enzyme, for instance a Cas9, and/or any of the present RNAs, for instance a guide RNA, can be delivered using adeno associated virus (AAV), lentivirus, adenovirus or other viral vector types, or combinations thereof. Cas9 and one or more guide RNAs can be packaged into one or more viral vectors. In some embodiments, the viral vector is delivered to the tissue of interest by, for example, an intramuscular injection, while other times the viral delivery is via intravenous, transdermal, intranasal, oral, mucosal, or other delivery methods. Such delivery may be either via a single dose, or multiple doses. One skilled in the art understands that the actual dosage to be delivered herein may vary greatly depending upon a variety of factors, such as the vector chose, the target cell, organism, or tissue, the general condition of the subject to be treated, the degree of transformation/modification sought, the administration route, the administration mode, the type of transformation/modification sought, etc.

Date Reçue/Date Received 2020-10-05

DEMANDE OU BREVET VOLUMINEUX

LA PRÉSENTE PARTIE DE CETTE DEMANDE OU CE BREVET COMPREND PLUS D'UN TOME.

CECI EST LE TOME 1 DE 6
CONTENANT LES PAGES 1 À 72

NOTE : Pour les tomes additionels, veuillez contacter le Bureau canadien des brevets

JUMBO APPLICATIONS/PATENTS

THIS SECTION OF THE APPLICATION/PATENT CONTAINS MORE THAN ONE VOLUME

THIS IS VOLUME 1 OF 6
CONTAINING PAGES 1 TO 72

NOTE: For additional volumes, please contact the Canadian Patent Office

NOM DU FICHIER / FILE NAME :

NOTE POUR LE TOME / VOLUME NOTE:

THE EMBODIMENTS OF THE INVENTION FOR WHICH AN EXCLUSIVE PROPERTY OR PRIVILEGE IS CLAIMED ARE DEFINED AS FOLLOWS:

1. A composition comprising a pharmaceutically acceptable carrier and a Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-CRISPR associated (Cas) (CRISPR-Cas) system comprising: a Cas9 comprising at least one nuclear localization sequence, a guide sequence designed to hybridize to a hepatic target sequence expressed in liver and associated with a gene involved in a liver disease or disorder, a tracr mate sequence and a tracr sequence, or one or more polynucleotides encoding said CRISPR-Cas system, for use in treating the liver disease or disorder *in vivo* in an organism.

2. The composition for use according to claim 1, wherein expression of a gene product is altered by the CRISPR-Cas system in liver, thereby producing a phenotypic change.

3. The composition for use according to claim 1 or 2, for treating or inhibiting a condition caused by a defect in the hepatic target sequence expressed in liver.

4. The composition for use according to any one of claims 1 to 3, said composition comprising one or more viral vectors comprising
(a) a first regulatory element operably linked to a polynucleotide sequence encoding the guide sequence, the tracr mate sequence, and the tracr sequence, and
(b) a second regulatory element operably linked to a polynucleotide sequence encoding the Cas9 comprising at least one nuclear localization sequence,
wherein components a) and b) are located on same or different vectors.

5. The composition for use according to any one of claims 1 to 3, said composition comprising:
(a) a polynucleotide sequence comprising, from 5' to 3', the guide sequence, the tracr mate sequence, and the tracr sequence, and
(b) a polynucleotide sequence encoding the Cas9 and at least one nuclear localization sequence,

Date Reçue/Date Received 2021-09-28

wherein the tracr mate sequence hybridizes to the tracr sequence, wherein the guide sequence directs sequence-specific binding of a CRISPR complex comprising the Cas9 to the target sequence, and wherein the polynucleotide sequence encoding the Cas9 is DNA or RNA.

6. The composition for use according to any one of claims 1 to 3, said composition comprising:

(a) a polynucleotide sequence comprising the guide sequence and the tracr mate sequence,

(b) a polynucleotide sequence encoding the Cas9 and at least one nuclear localization sequence, and

(c) a polynucleotide sequence comprising the tracr sequence,

wherein the tracr mate sequence hybridizes to the tracr sequence, wherein the guide sequence directs sequence-specific binding of a CRISPR complex comprising the Cas9 to the target sequence, and wherein the polynucleotide sequence encoding a Cas9 is DNA or RNA.

7. The composition for use according to any one of claims 1 to 6, wherein expression of two or more hepatic gene products is altered.

8. The composition for use according to any one of claims 1 to 5, wherein the guide sequence is fused to the tracr mate sequence and the tracr sequence.

9. The composition for use according to any one of claims 1 to 8, wherein the polynucleotide sequence encoding the Cas9 is codon optimized for expression in the organism.

10. The composition for use according to any one of claims 1 to 9, wherein the organism is a mammal.

11. The composition for use according to any one of claims 1 to 10, wherein the Cas9 is *S. pyogenes* or *S. aureus* Cas9.

Date Reçue/Date Received 2021-09-28

12. The composition for use according to claim 4, wherein the one or more viral vectors is an adeno-associated virus (AAV) or lentiviral vector.

13. The composition for use according to claim 12, wherein the AAV is an AAV2/8 vector.

14. The composition for use according to any one of claim 4 to 6, wherein a) and b) are located on the same viral vector.

15. The composition for use according to claim 14, wherein said viral vector is an adeno-associated virus (AAV) vector and said Cas9 is *S. aureus* Cas9.

16. The composition for use according to any one of claim 1 to 15, wherein the Cas9 comprises one or more mutations selected from D10A, E762A, H840A, N854A, N863A and D986A with reference to the position numbering of *S. pyogenes* Cas9 and lacks the ability to cleave one or both strands of a target polynucleotide containing the target sequence.

17. The composition for use according to claim 16, wherein the Cas9 is linked to one or more heterologous protein domains.

18. The composition for use according to any one of claims 1 to 17, wherein the hepatic target sequence comprises a polynucleotide sequence encoding PCSK9, A1AT, SERPINA1, HMGCR, APOB, LDLR, ANGPTL3, F8, F9/FIX, AAT, FAH, HPD, TAT, ATP7B, UGT1A1, OTC, or ARH.

19. The composition for use according to any one of claims 1 to 18, wherein the polynucleotide sequence encoding the Cas9 is operably linked to a liver-specific promoter.

20. The composition for use according to any one of claims 1 to 18, wherein the polynucleotide sequence encoding the Cas9 is operably linked to a TBG promoter.

Date Reçue/Date Received 2021-09-28

21. The composition for use according to any one of claims 1 to 20, wherein the liver disease or disorder is AAT deficiency, hypercholesterolemia, or hyperlipidemia.

22. The composition for use according to any one of claims 1 to 21, wherein said use is the correction of one or more deficient genotypes.

23. The composition for use according to any one of claims 1 to 22, wherein said CRISPR-Cas system is comprised in a lipid particle as RNA polynucleotides.

24. The composition for use according to claim 23, wherein the lipid particle is a stable nucleic acid-lipid particle (SNALP).

25. The composition for use according to any one of claims 1 to 22, wherein said CRISPR-Cas system is comprised in the adeno-associated virus (AAV) 2/8 vector.

26. The composition for use according to any one of claims 1 to 25, wherein the composition is formulated for systemic or intravenous administration.

27. Use of a Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-CRISPR associated (Cas) (CRISPR-Cas) system comprising: a Cas9 comprising at least one nuclear localization sequence, a guide sequence designed to hybridize to a hepatic target sequence expressed in liver and associated with a gene involved in a liver disease or disorder, a tracr mate sequence and a tracr sequence, or one or more polynucleotides encoding said CRISPR-Cas system to treat the liver disease or disorder *in vivo* in an organism.

28. The use according to claim 27, wherein upon said use expression of a gene product is altered by the CRISPR-Cas system in liver, thereby producing a phenotypic change.

29. The use according to claim 27 or 28, wherein the CRISPR-Cas system is for delivery to liver cells in an organism.

Date Reçue/Date Received 2021-09-28

30. The use according to any one of claims 27 to 29, wherein the liver disease or disorder is caused by a defect in the hepatic target sequence expressed in liver.

31. The use according to any one of claims 27 to 29, wherein the use comprises use of one or more viral vectors comprising

(a) a first regulatory element operably linked to a polynucleotide sequence encoding the guide sequence, the tracr mate sequence, and the tracr sequence, and

(b) a second regulatory element operably linked to a polynucleotide sequence encoding the Cas9 comprising at least one nuclear localization sequence,

wherein components a) and b) are located on same or different vectors.

32. The use according to any one of claims 27 to 31, said use comprises use of a composition comprising:

(a) a polynucleotide comprising, from 5' to 3' the guide sequence, the tracr mate sequence, and the tracr sequence, and

(b) a polynucleotide encoding the Cas9 and at least one nuclear localization sequence,

wherein the tracr mate sequence hybridizes to the tracr sequence, wherein the guide sequence directs sequence-specific binding of a CRISPR complex comprising the Cas9 to the target sequence, and wherein the polynucleotide encoding the Cas9 is DNA or RNA.

33. The use according to any one of claims 27 to 31, said use comprising use of:

(a) a polynucleotide comprising the guide sequence and the tracr mate sequence,

(b) a polynucleotide encoding the Cas9 and at least one nuclear localization sequence, and

(c) a polynucleotide comprising the tracr sequence,

wherein the tracr mate sequence hybridizes to the tracr sequence, wherein the guide sequence directs sequence-specific binding of a CRISPR complex comprising the Cas9 to the target sequence, and wherein the polynucleotide sequence encoding a Cas9 is DNA or RNA.

Date Reçue/Date Received 2021-09-28

34. The use according to any one of claims 27 to 33, wherein upon said use expression of two or more hepatic gene products is altered.

35. The use according to any one of claims 27 to 34, wherein the guide sequence is fused to the tracr mate sequence and the tracr sequence.

36. The use according to any one of claims 27 to 34, wherein the polynucleotide encoding the Cas9 is codon optimized for expression in the organism.

37. The use according to any one of claims 27 to 36, wherein the organism is a mammal.

38. The use according to any one of claims 27 to 37, wherein the Cas9 is *S. pyogenes* or *S. aureus* Cas9.

39. The use according to claim 31, wherein the one or more viral vectors is an adeno-associated virus (AAV) or lentiviral vector.

40. The use according to claim 39, wherein the adeno-associated virus (AAV) or lentiviral vector is an AAV2/8 vector.

41. The use according to any one of claim 31 to 40, wherein a) and b) are located on the same viral vector.

42. The use according to claim 41, wherein said viral vector is an AAV vector and said Cas9 is *S. aureus* Cas9.

43. The use according to any one of claim 27 to 42, wherein the Cas9 comprises one or more mutations selected from D10A, E762A, H840A, N854A, N863A and D986A with reference to the position numbering of *S. pyogenes* Cas9 and lacks the ability to cleave one or both strands of a target polynucleotide containing the target sequence.

Date Reçue/Date Received 2021-09-28

44. The use according to claim 43, wherein the Cas9 is linked to one or more heterologous protein domains.

45. The use according to any one of claims 27 to 44, wherein the hepatic target sequence comprises a polynucleotide sequence encoding PCSK9, A1AT, SERPINA1, HMGCR, APOB, LDLR, ANGPTL3, F8, F9/FIX, AAT, FAH, HPD, TAT, ATP7B, UGT1A1, OTC, or ARH.

46. The use according to any one of claims 27 to 45, wherein the polynucleotide sequence encoding the Cas9 is operably linked to a liver-specific promoter.

47. The use according to any one of claims 27 to 44, wherein the polynucleotide sequence encoding the Cas9 is operably linked to a TBG promoter.

48. The use according to any one of claims 27 to 47, wherein the liver disease or disorder is AAT deficiency, hypercholesterolemia, or hyperlipidemia.

49. The use according to any one of claims 27 to 48, wherein upon said use one or more deficient genotypes are corrected.

50. The use according to any one of claims 27 to 49, wherein said CRISPR-Cas system is formulated as RNA polynucleotides in a lipid particle.

51. The use according to claim 50, wherein the lipid particle is a stable nucleic acid-lipid particle (SNALP).

52. The use according to any one of claims 27 to 49, wherein said CRISPR-Cas system is formulated as adeno-associated virus (AAV)2/8 vector.

Date Reçue/Date Received 2021-09-28

53. The use according to any one of claims 27 to 49, wherein the CRISPR-Cas system is formulated for systemic or intravenous administration.

54. Use of a Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-CRISPR associated (Cas) (CRISPR-Cas) system comprising: a Cas9 comprising at least one nuclear localization sequence, a guide sequence designed to hybridize to a hepatic target sequence expressed in liver and associated with a gene involved in a liver disease or disorder, a tracr mate sequence and a tracr sequence, or one or more polynucleotides encoding said CRISPR-Cas system in the manufacture of a medicament to treat the liver disease or disorder *in vivo* in an organism.

Date Reçue/Date Received 2021-09-28

DEMANDE OU BREVET VOLUMINEUX

LA PRÉSENTE PARTIE DE CETTE DEMANDE OU CE BREVET COMPREND PLUS D'UN TOME.

CECI EST LE TOME 1 DE 2
CONTENANT LES PAGES 1 À 77

NOTE : Pour les tomes additionels, veuillez contacter le Bureau canadien des brevets

JUMBO APPLICATIONS/PATENTS

THIS SECTION OF THE APPLICATION/PATENT CONTAINS MORE THAN ONE VOLUME

THIS IS VOLUME 1 OF 2
CONTAINING PAGES 1 TO 77

NOTE: For additional volumes, please contact the Canadian Patent Office

NOM DU FICHIER / FILE NAME :

NOTE POUR LE TOME / VOLUME NOTE:

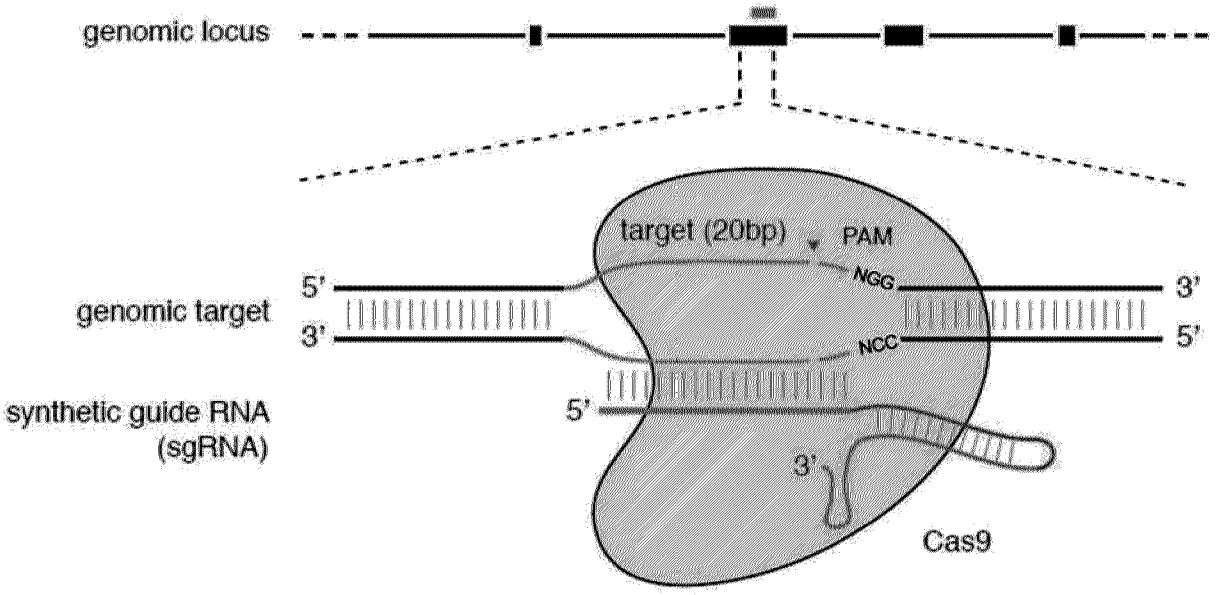


FIG. 1

Date Reçue/Date Received 2021-09-28

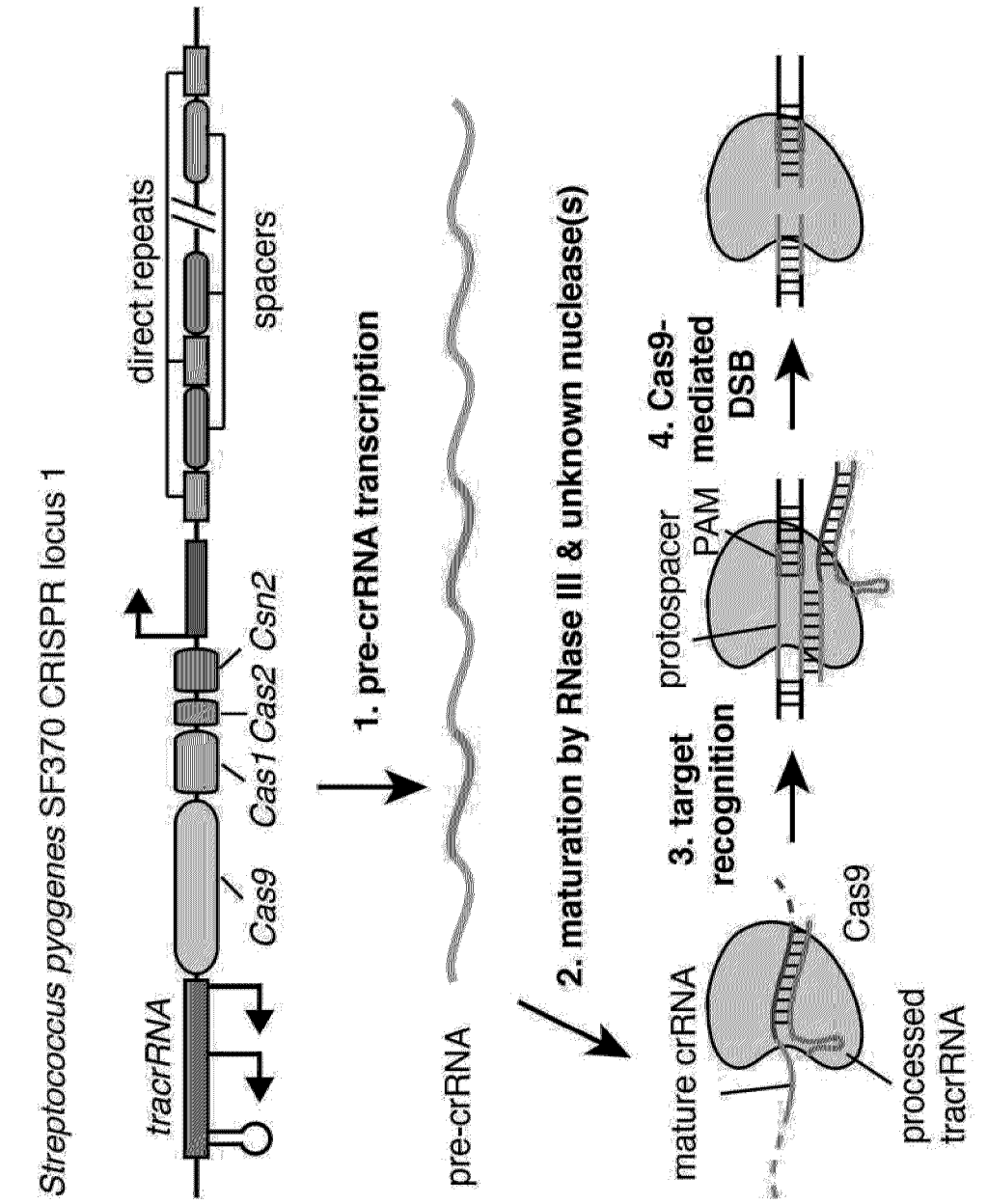

A
Streptococcus pyogenes SF370 CRISPR locus 1
tracrRNA
Cas9
Cas1
Cas2
Csn2
direct repeats
spacers
1. pre-crRNA transcription
pre-crRNA
2. maturation by RNase III & unknown nuclease(s)
mature crRNA
Cas9
processed tracrRNA
3. target recognition
protospacer
PAM
4. Cas9-mediated DSB

FIG. 2A

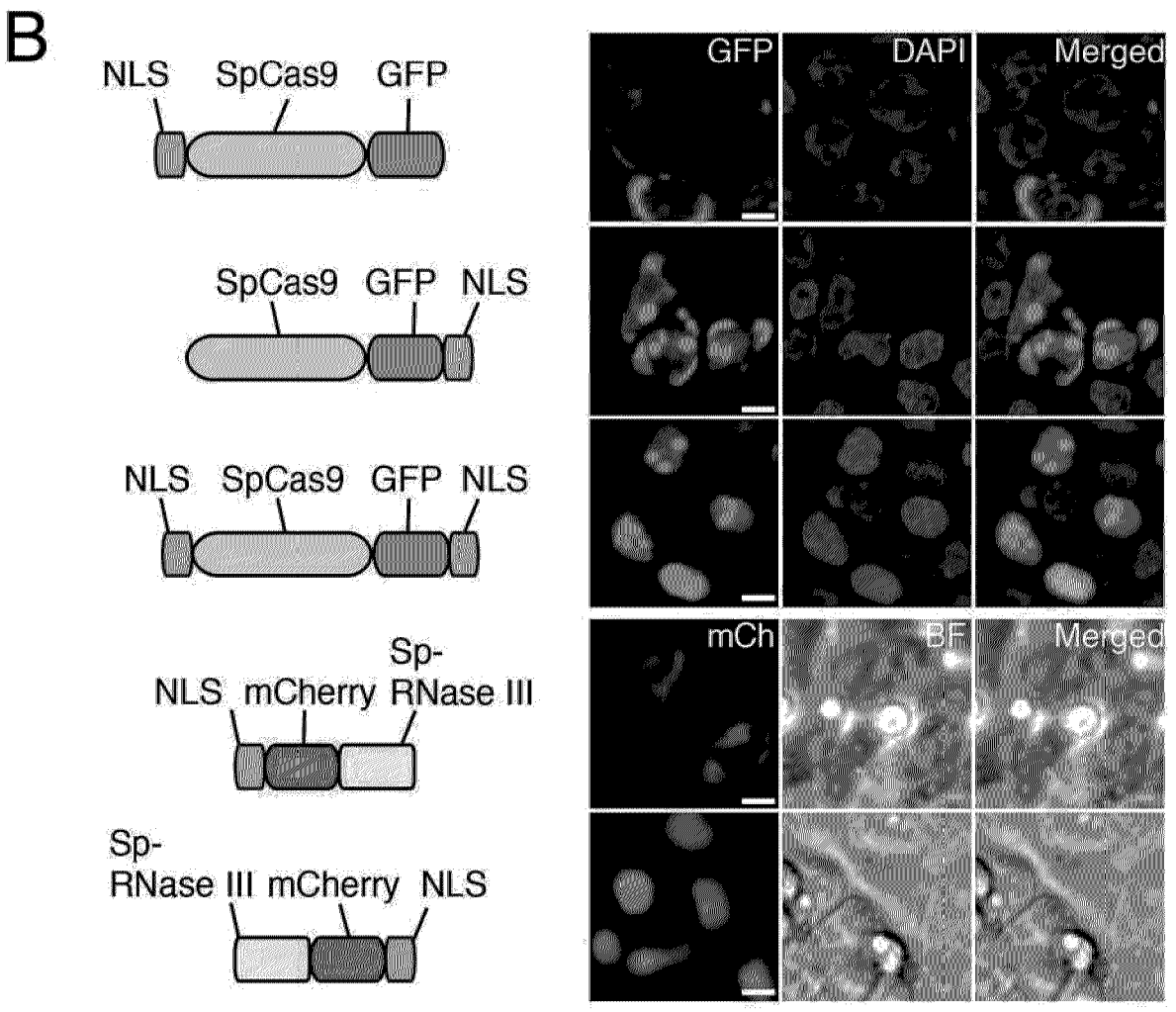

B
NLS SpCas9 GFP
SpCas9 GFP NLS
NLS SpCas9 GFP NLS
Sp-
NLS mCherry RNase III
Sp-
RNase III mCherry NLS
GFP
DAPI
Merged
mCh
BF
Merged

FIG. 2B

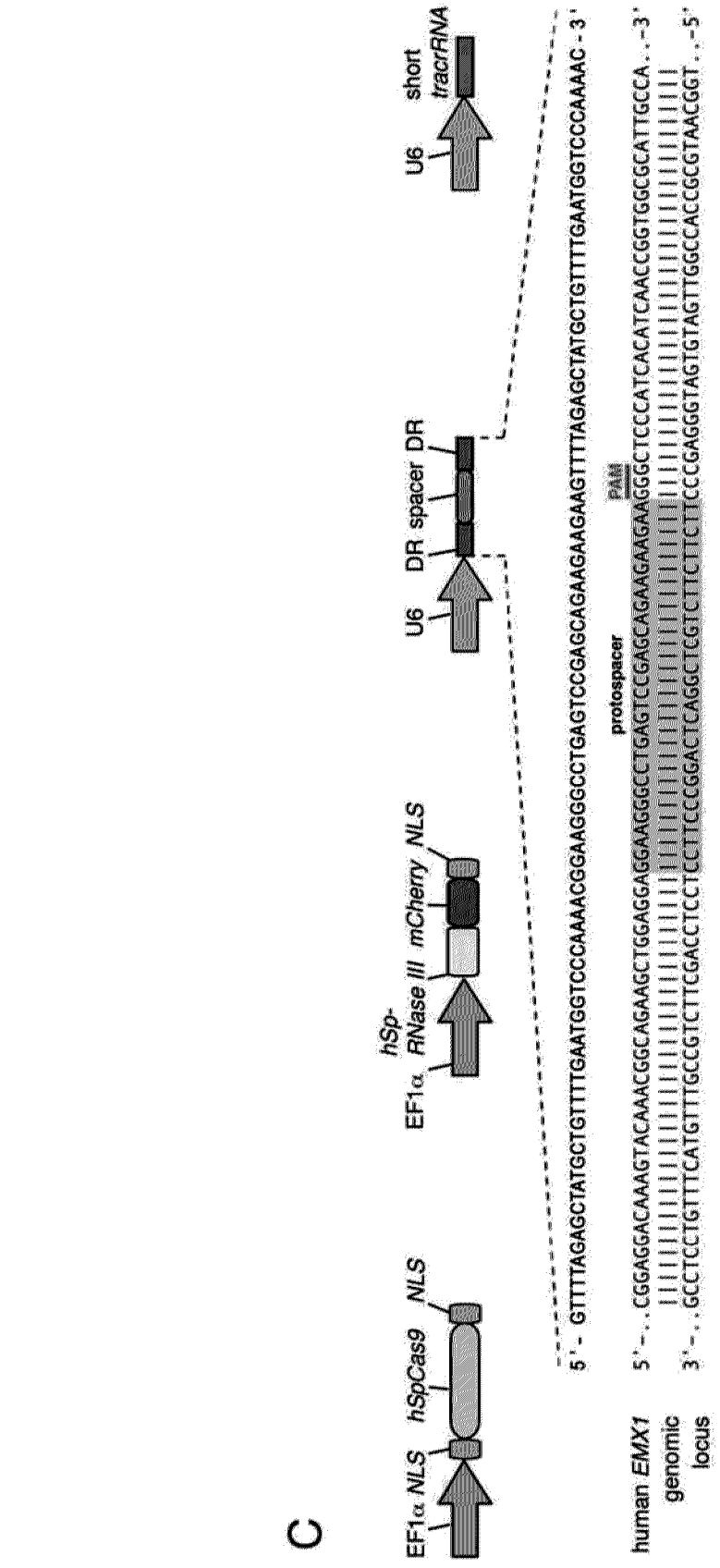

C
EF1α NLS hSpCas9 NLS
EF1α hSp-RNase III mCherry NLS
U6 DR spacer DR
U6 short tracrRNA
5'- GTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAACGGAAGGGCCTGAGTCCGAGCAGAAGAAGAAGTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAAC - 3'
protospacer
PAM
human EMX1 genomic locus
5'-..CGGAGGACAAAGTACAAACGGCAGAAGCTGGAGGAGGAAGGGCCTGAGTCCGAGCAGAAGAAGAAGGGCTCCCATCACATCAACCGGTGGCGCATTGCCA..-3'
3'-..GCCTCCTGTTTCATGTTTGCCGTCTTCGACCTCCTTCCCGGACTCAGGCTCGTCTTCTTCTTCCCGAGGGTAGTGTAGTTGGCCACCGCGTAACGGT..-5'

FIG. 2C

D

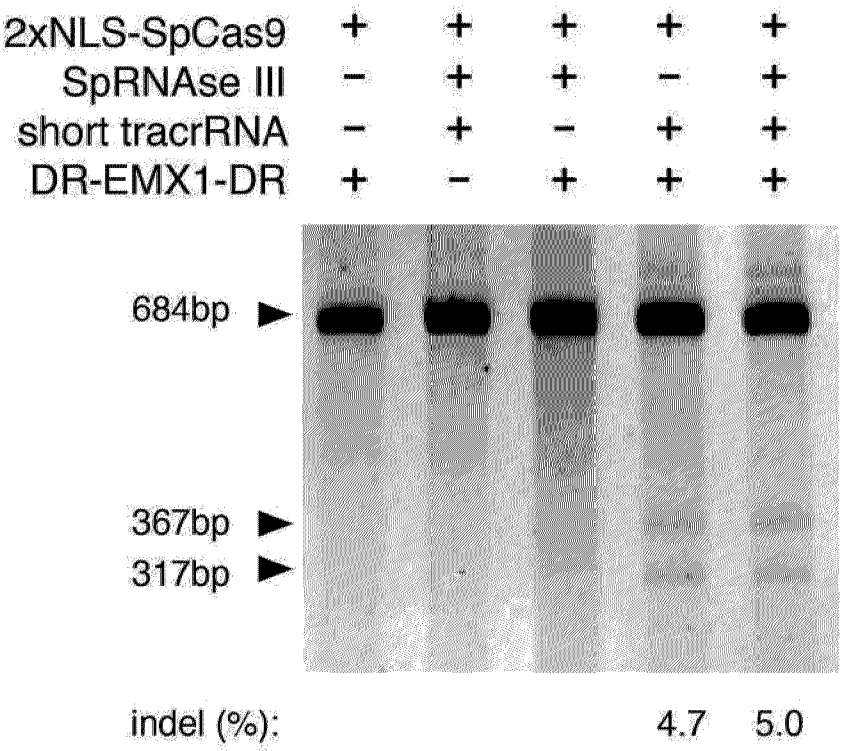
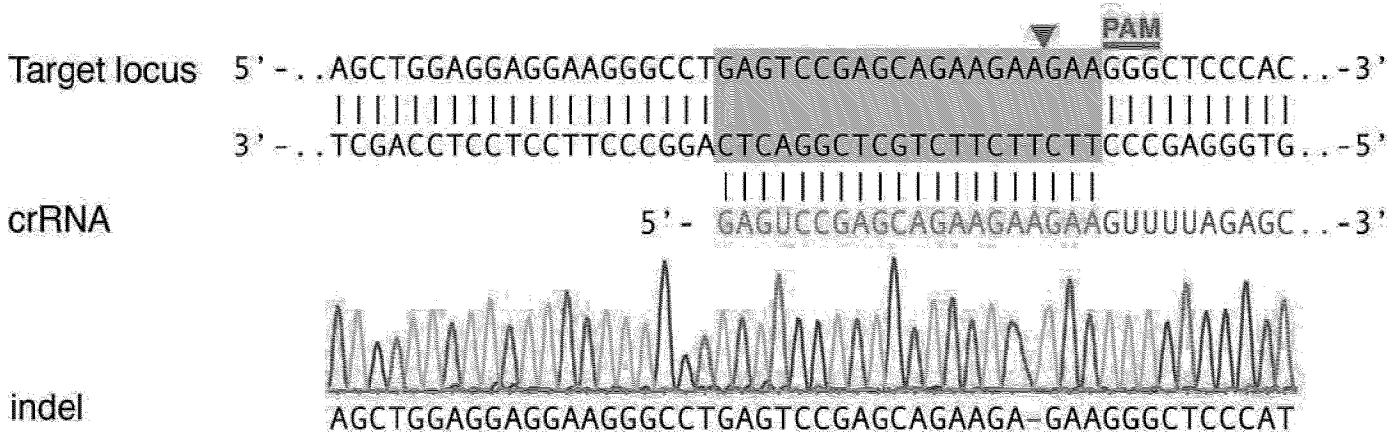

2xNLS-SpCas9 + + + + +
SpRNAse III − + + − +
short tracrRNA − + − + +
DR-EMX1-DR + − + + +
684bp
367bp
317bp
indel (%): 4.7 5.0

FIG. 2D

Date Reçue/Date Received 2021-09-28

E



F

human *EMX1* protospacer target (mutation in 5 of 43 sequenced clones = 11.6%)

WT	5'-..CTGGAGGAGGAAGGGCCTGAGTCCGAGCAGAAG AAGAAGGGCTCCCATCACAT..-3'
Δ1	CTGGAGGAGGAAGGGCCTGAGTCCGAGCAGAAG--AGAAGGGCTCCCATCACAT
+1	CTGGAGGAGGAAGGGCCTGAGTCCGAGCAGAAGAAAGAAGGGCTCCCATCACAT
Δ3	CTGGAGGAGGAAGGGCCTGAGTCCGAGCAGAAG----AAGGGCTCCCATCACAT
m1, Δ6	CTGGAGGAGGAAGGGCCTGAGCCCGAGCAGAAG-------GGCTCCCATCACAT

FIG. 2E-F

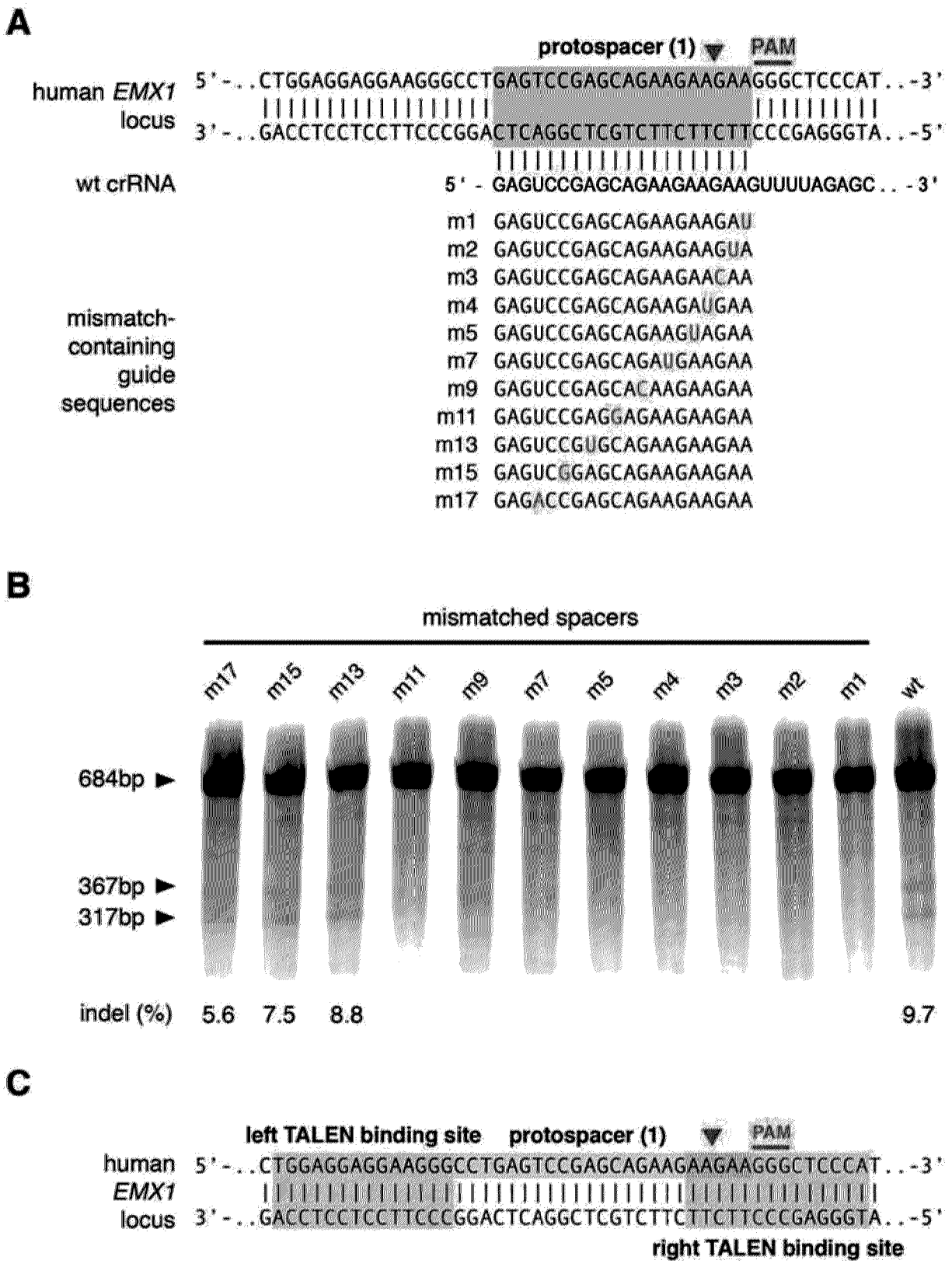

A
protospacer (1)
PAM
human EMX1 locus
5'-..CTGGAGGAGGAAGGGCCTGAGTCCGAGCAGAAGAAGAAGGGCTCCCAT..-3'
3'-..GACCTCCTCCTTCCCGGACTCAGGCTCGTCTTCTTCTTCCCGAGGGTA..-5'
wt crRNA
5' - GAGUCCGAGCAGAAGAAGAAGUUUUAGAGC..-3'
mismatch-containing guide sequences
m1 GAGUCCGAGCAGAAGAAGAU
m2 GAGUCCGAGCAGAAGAAGUA
m3 GAGUCCGAGCAGAAGAACAA
m4 GAGUCCGAGCAGAAGAUGAA
m5 GAGUCCGAGCAGAAGUAGAA
m7 GAGUCCGAGCAGAUGAAGAA
m9 GAGUCCGAGCACAAGAAGAA
m11 GAGUCCGAGGAGAAGAAGAA
m13 GAGUCCGUGCAGAAGAAGAA
m15 GAGUCGGAGCAGAAGAAGAA
m17 GAGACCGAGCAGAAGAAGAA
B
mismatched spacers
m17
m15
m13
m11
m9
m7
m5
m4
m3
m2
m1
wt
684bp
367bp
317bp
indel (%)
5.6
7.5
8.8
9.7
C
left TALEN binding site
protospacer (1)
PAM
human EMX1 locus
5'-..CTGGAGGAGGAAGGGCCTGAGTCCGAGCAGAAGAAGAAGGGCTCCCAT..-3'
3'-..GACCTCCTCCTTCCCGGACTCAGGCTCGTCTTCTTCTTCCCGAGGGTA..-5'
right TALEN binding site

FIG. 3A-C

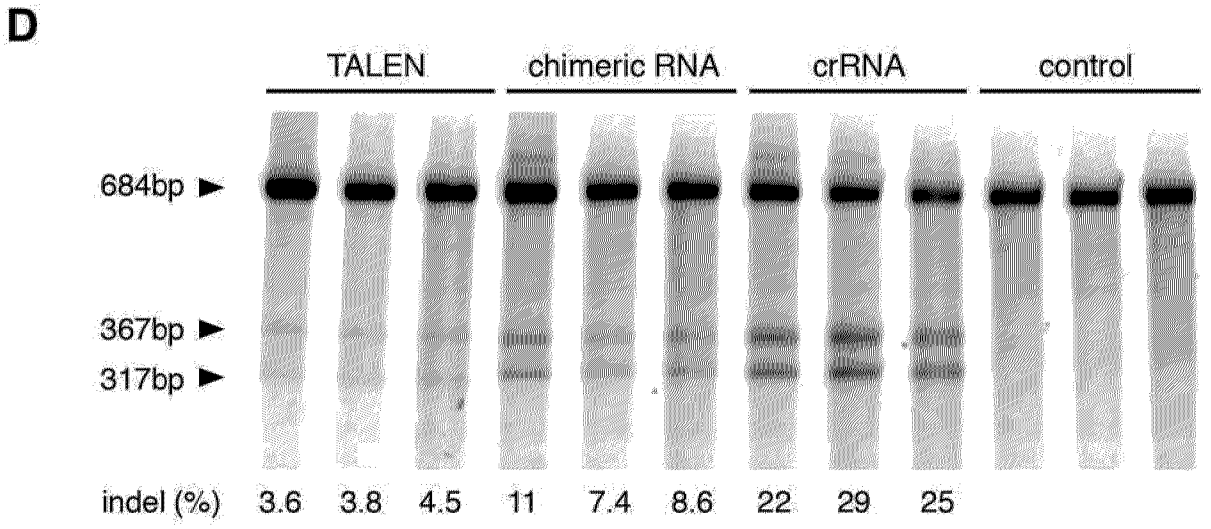

D
TALEN
chimeric RNA
crRNA
control
684bp
367bp
317bp
indel (%)
3.6
3.8
4.5
11
7.4
8.6
22
29
25

FIG. 3D

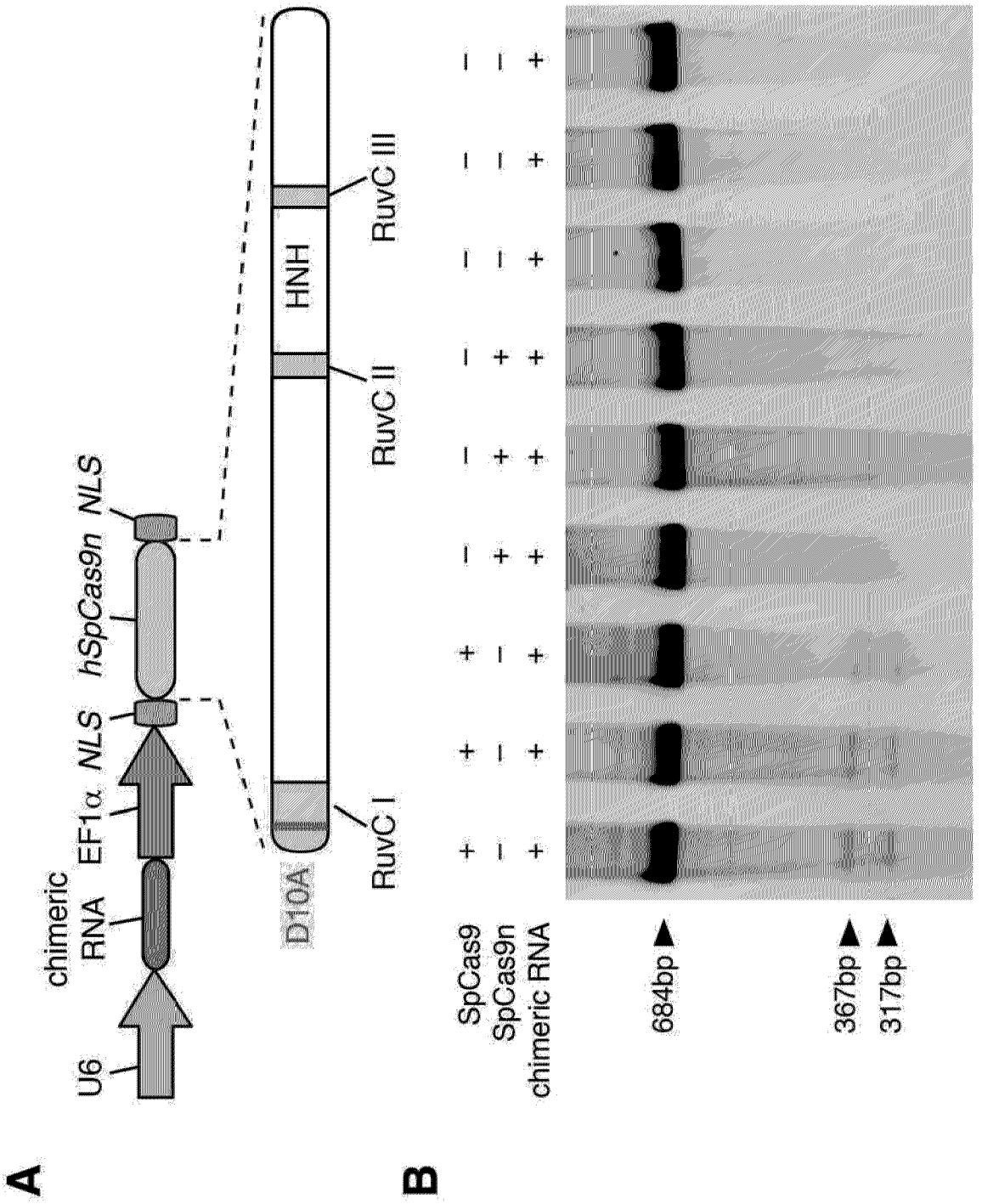

A
U6
chimeric RNA
EF1α
NLS
hSpCas9n
NLS
D10A
RuvC I
RuvC II
HNH
RuvC III
B
SpCas9
SpCas9n
chimeric RNA
684bp
367bp
317bp
indel (%)
5.3
4.6
3.5

FIG. 4A-B

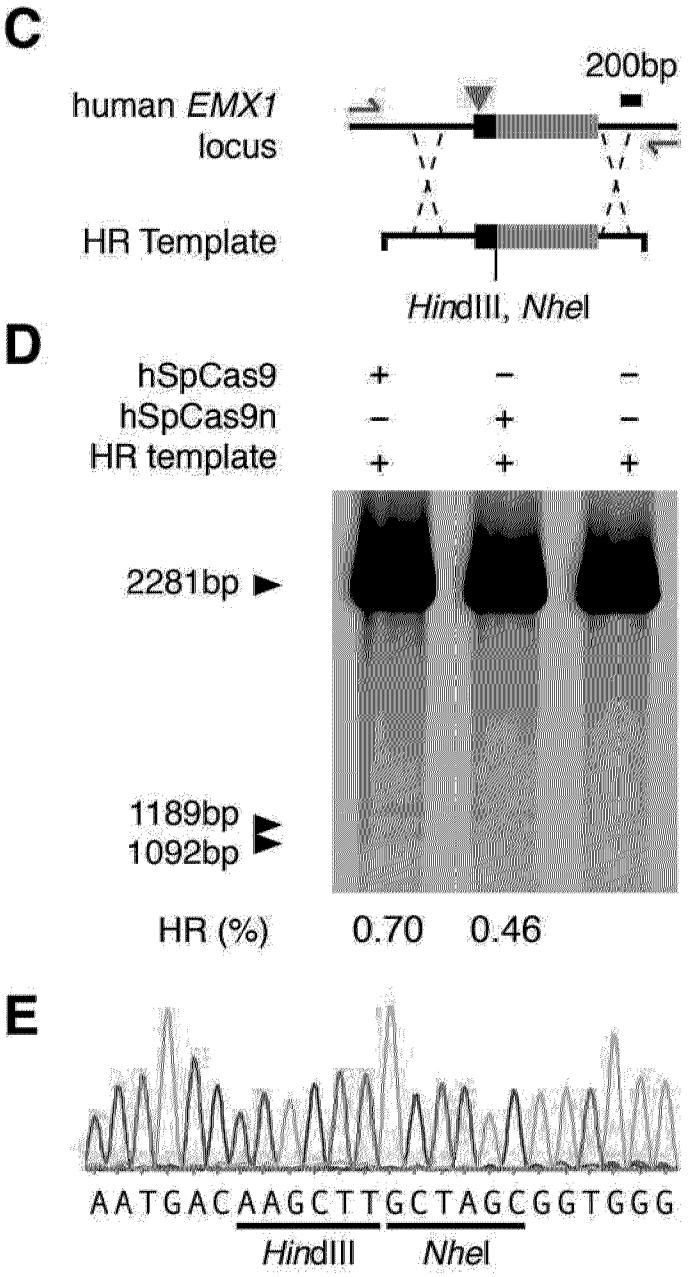

C
200bp
human EMX1
locus
HR Template
HindIII, NheI
D
hSpCas9
hSpCas9n
HR template
2281bp
1189bp
1092bp
HR (%)
0.70
0.46
E
AATGACAAGCTTGCTAGCGGTGGG
HindIII
NheI

FIG. 4C-E

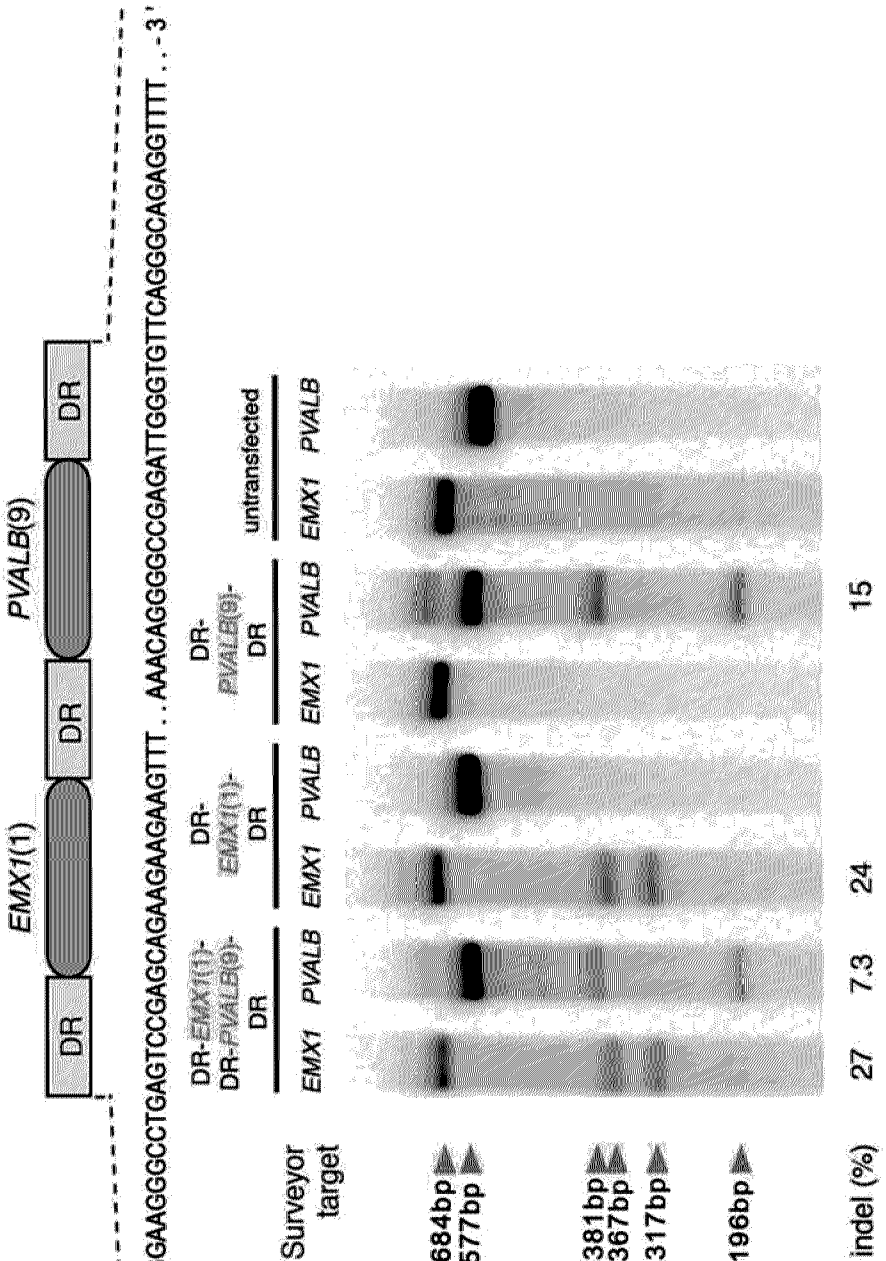

F
crRNA array design
EMX1(1)
PVALB(9)
DR
DR
DR
DR
5' - . . AAAACGGAAGGGCCTGAGTCCGAGCAGAAGAAGAAGTTT . . AAACAGGGGCCGAGATTGGGTGTTCAGGGCAGAGGTTTT . . - 3'
DR-EMX1(1)-DR-PVALB(9)-DR
DR-EMX1(1)-DR
DR-PVALB(9)-DR
untransfected
EMX1 PVALB EMX1 PVALB EMX1 PVALB EMX1 PVALB
Surveyor target
684bp
577bp
381bp
367bp
317bp
196bp
indel (%) 27 7.3 24 15

FIG. 4F

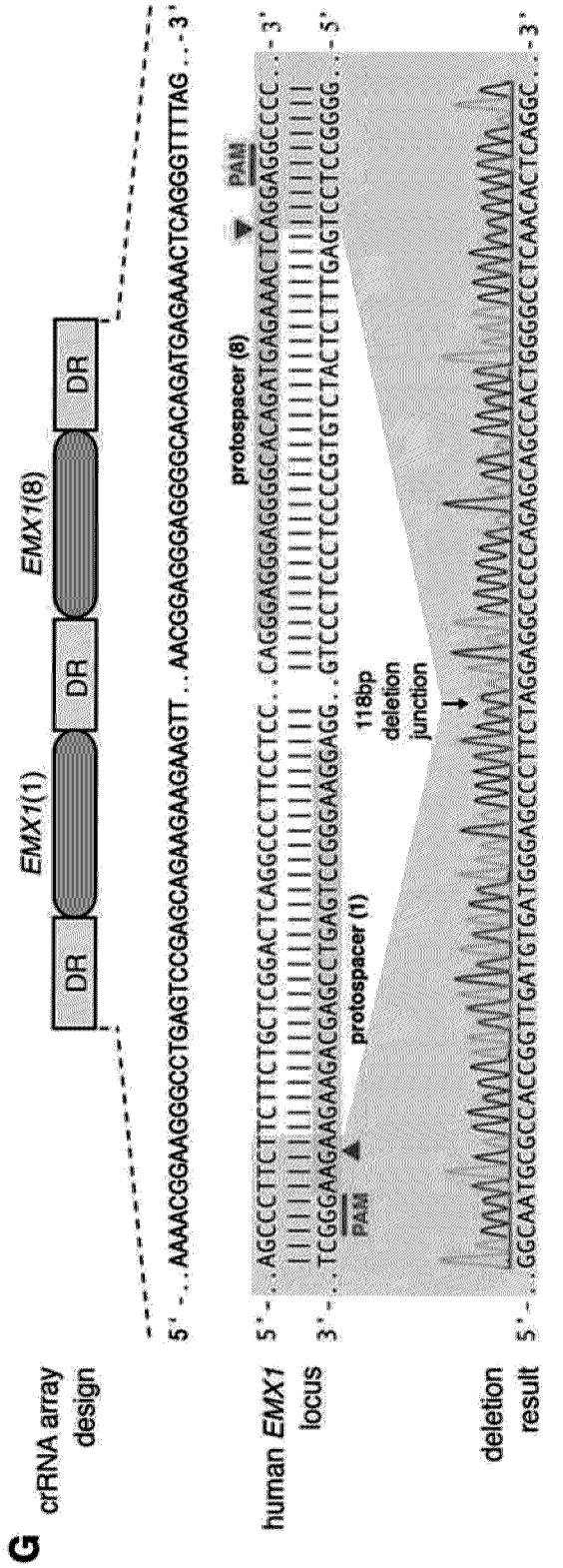

G
crRNA array design
EMX1(1)
EMX1(8)
DR
DR
DR
5' - . . AAAACGGAAGGGCCTGAGTCCGAGCAGAAGAAGAAGTT . . AACGGAGGGAGGGGCACAGATGAGAAACTCAGGGTTTTAG . . - 3'
human EMX1 locus
protospacer (8)
PAM
5' - . . AGCCCTTCTTCTTCTGCTCGGACTCAGGCCCTTCCTCC . . CAGGGAGGGAGGGGCACAGATGAGAAACTCAGGAGGCCCC . . - 3'
3' - . . TCGGGAAGAAGAAGACGAGCCTGAGTCCGGGAAGGAGG . . GTCCCTCCCTCCCCGTGTCTACTCTTTGAGTCCTCCGGGG . . - 5'
protospacer (1)
PAM
118bp deletion junction
deletion result
5' - . . GGCAATGCGCCACCGGTTGATGTGATGGGAGCCCTTCTAGGAGGCCCCAGAGCAGCCACTGGGGCCTCAACACTCAGGC . . - 3'

FIG. 4G

Date Reçue/Date Received 2021-09-28

Cas9	target species	gene	protospacer ID	protospacer sequence (5' to 3')	PAM	strand	cell line tested	% indel (pre-crRNA + tracrRNA)	% indel (chimeric RNA)
S. pyogenes SF370 type II CRISPR	*Homo sapiens*	*EMX1*	1	GGAAGGGCCTGAGTCCGAGCAGAAGAAGAA	GGG	+	293FT	20 ± 1.8	6.7 ± 0.62
		EMX1	2	CATTGGAGGTGACATCGATGTCCTCCCCAT	TGG	–	293FT	2.1 ± 0.31	N.D.
		EMX1	3	GGACATCGATGTCACCTCCAATGACTAGGG	TGG	+	293FT	14 ± 1.1	N.D.
		EMX1	4	CATCGATGTCCTCCCCATTGGCCTGCTTCG	TGG	–	293FT	11 ± 1.7	N.D.
		EMX1	5	TTCGTGGCAATGCGCCACCGGTTGATGTGA	TGG	–	293FT	4.3 ± 0.46	2.1 ± 0.51
		EMX1	6	TCGTGGCAATGCGCCACCGGTTGATGTGAT	GGG	–	293FT	4.0 ± 0.66	0.41 ± 0.25
		EMX1	7	TCCAGCTTCTGCCGTTTGTACTTTGTCCTC	CGG	–	293FT	1.5 ± 0.12	N.D.
		EMX1	8	GGAGGGAGGGGCACAGATGAGAAACTCAGG	AGG	–	293FT	7.8 ± 0.83	2.3 ± 1.2
	Homo sapiens	*PVALB*	9	AGGGGCCGAGATTGGGTGTTCAGGGCAGAG	AGG	+	293FT	21 ± 2.6	6.5 ± 0.32
		PVALB	10	ATGCAGGAGGGTGGCGAGAGGGGCCGAGAT	TGG	+	293FT	N.D.	N.D.
		PVALB	11	GGTGGCGAGAGGGGCCGAGATTGGGTGTTC	AGG	+	293FT	N.D.	N.D.
	Mus musculus	*Th*	12	CAAGCACTGAGTGCCATTAGCTAAATGCAT	AGG	–	Neuro2A	27 ± 4.3	4.1 ± 2.2
		Th	13	AATGCATAGGGTACCACCCACAGGTGCCAG	GGG	–	Neuro2A	4.8 ± 1.2	N.D.
		Th	14	ACACACATGGGAAAGCCTCTGGGCCAGGAA	AGG	+	Neuro2A	11.3 ± 1.3	N.D.
S. thermophilus LMD-9 CRISPR1	*Homo sapiens*	*EMX1*	15	GGAGGAGGTAGTATACAGAAACACAGAGAA	GTAGAAT	–	293FT	14 ± 0.88	N.T.
		EMX1	16	AGAATGTAGAGGAGTCACAGAAACTCAGCA	CTAGAAA	–	293FT	7.8 ± 0.77	N.T.

FIG. 5

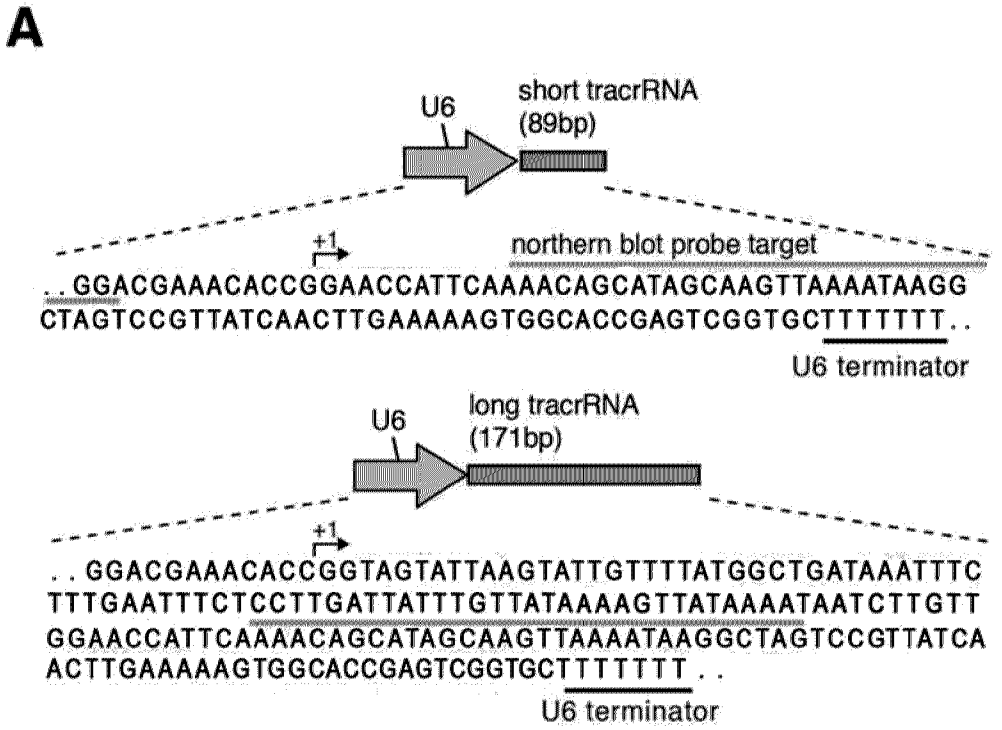

A
U6
short tracrRNA
(89bp)
+1
northern blot probe target
..GGACGAAACACCGGAACCATTCAAAACAGCATAGCAAGTTAAAATAAGG
CTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTTT..
U6 terminator
U6
long tracrRNA
(171bp)
+1
..GGACGAAACACCGGTAGTATTAAGTATTGTTTTATGGCTGATAAATTTC
TTTGAATTTCTCCTTGATTATTTGTTATAAAAGTTATAAAATAATCTTGTT
GGAACCATTCAAAACAGCATAGCAAGTTAAAATAAGGCTAGTCCGTTATCA
ACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTTT..
U6 terminator
B
SpCas9 + + + +
long tracrRNA + + – –
short tracrRNA – – + +
SpRNase III + + + +
DR-EMX1(1)-DR + + + +
684bp
367bp
317bp
indel (%) 0.60 14 11

FIG. 6A-B

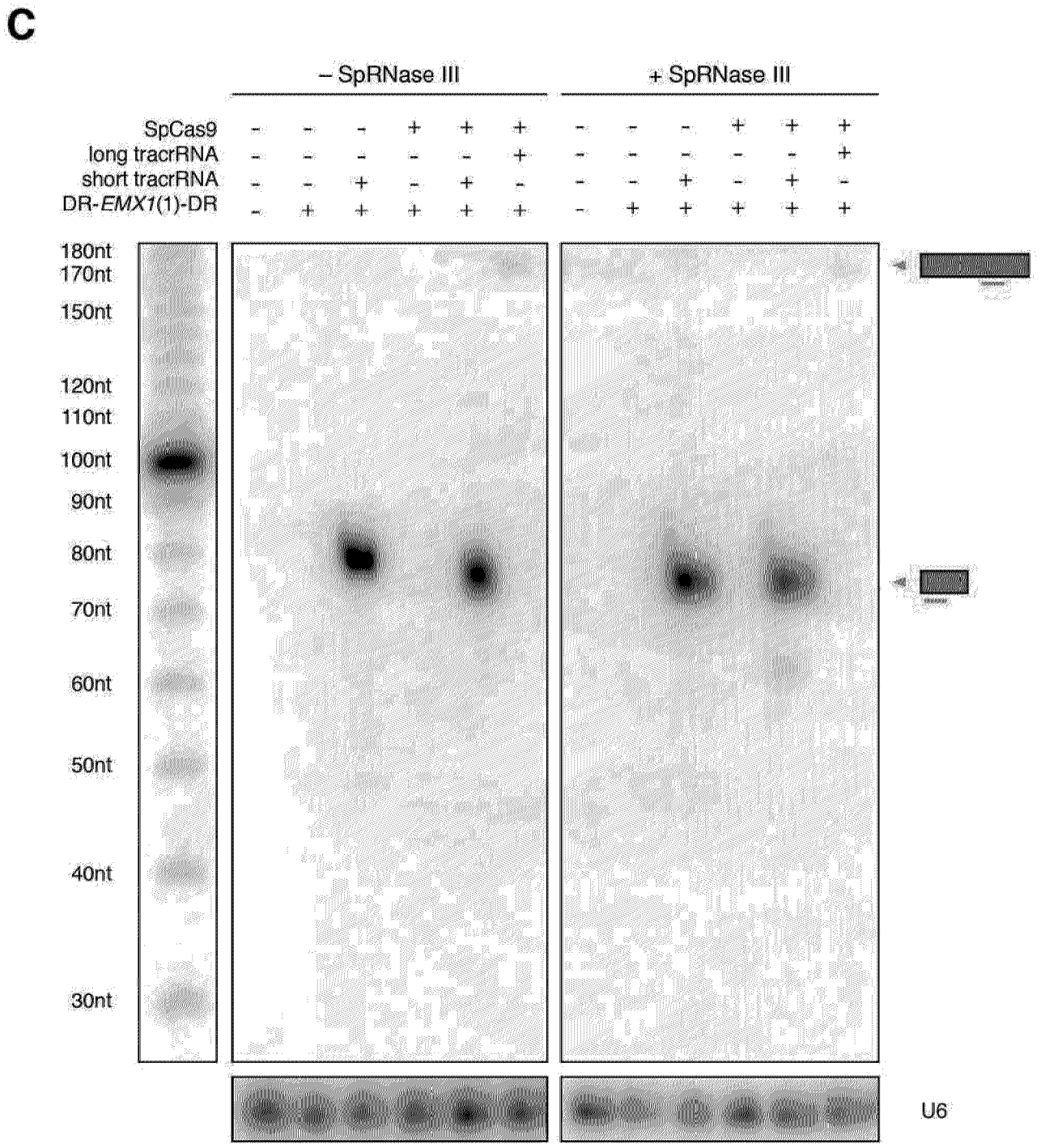
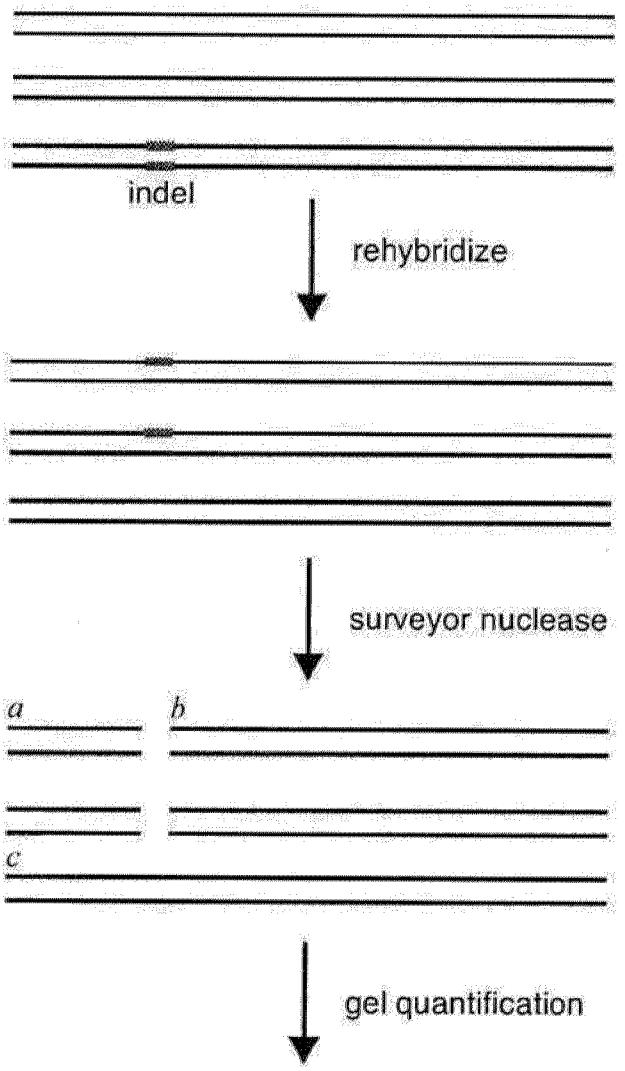

C
– SpRNase III
+ SpRNase III
SpCas9
long tracrRNA
short tracrRNA
DR-EMX1(1)-DR
180nt
170nt
150nt
120nt
110nt
100nt
90nt
80nt
70nt
60nt
50nt
40nt
30nt
U6

FIG. 6C



$$\% \text{ indel} = \left(1 - \sqrt{1 - (a + b)/(a + b + c)}\right) * 100$$

FIG. 7

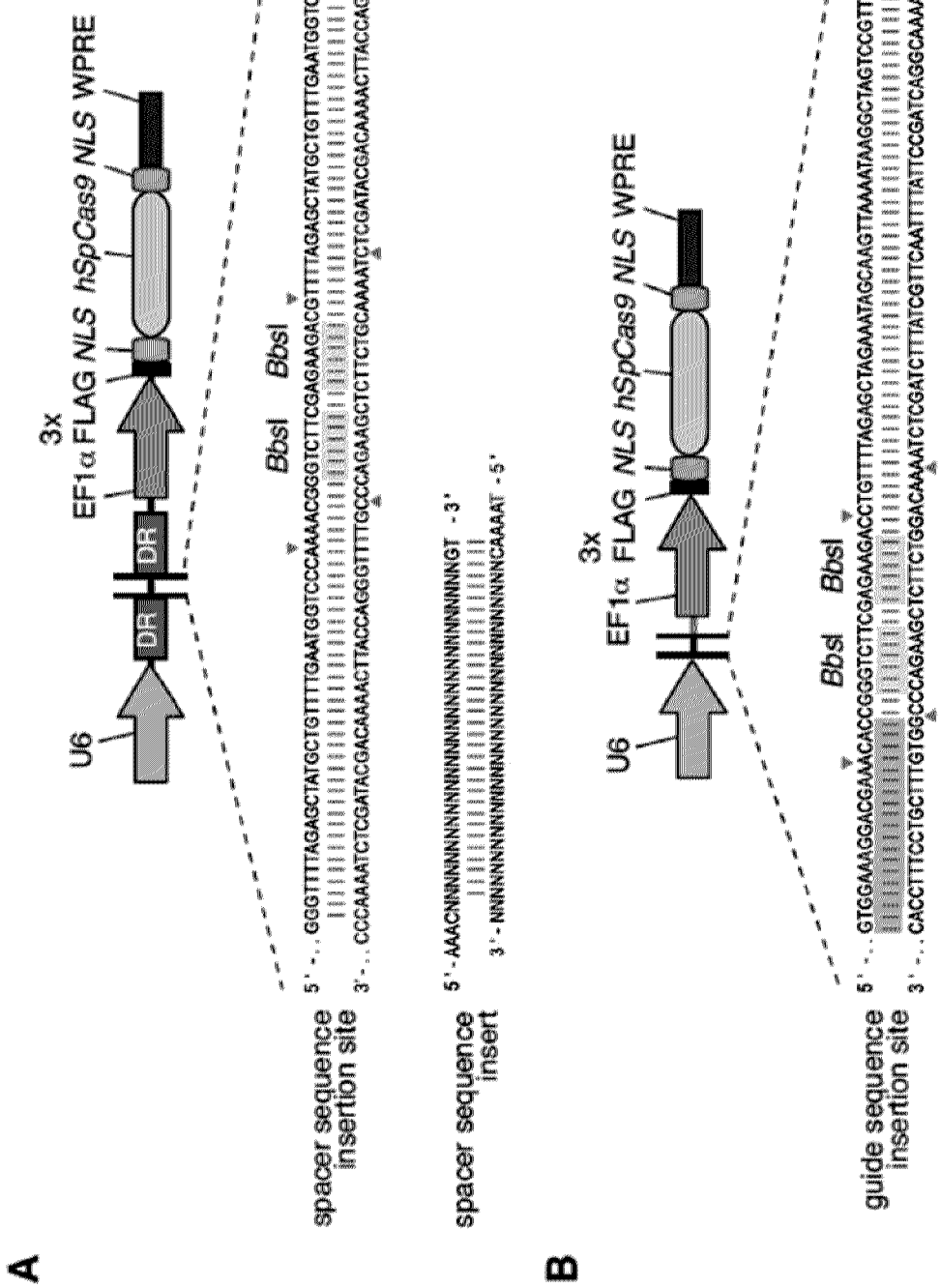
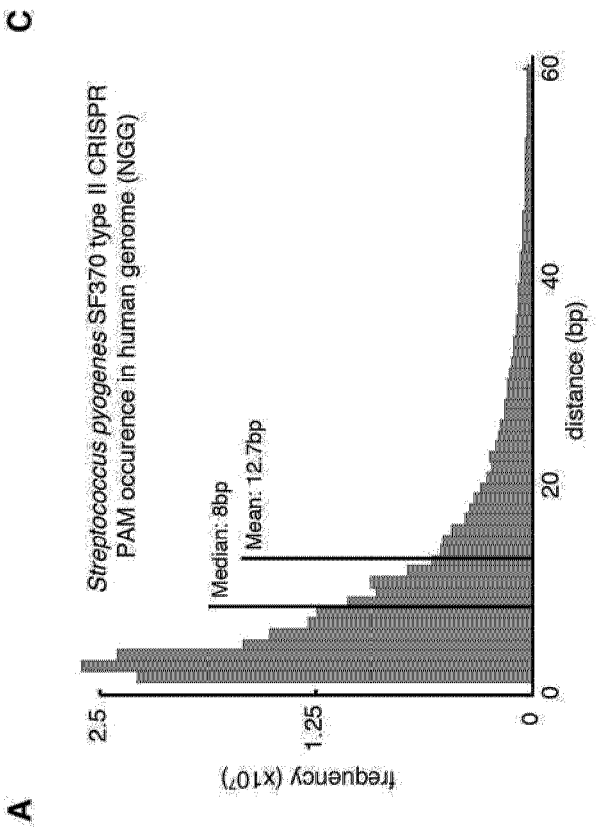

A
U6
DR
DR
3x
EF1α FLAG NLS hSpCas9 NLS WPRE
BbsI
BbsI
spacer sequence insertion site
5' - GGGTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAACGGGTCTTCGAGAAGACGTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAACTTTT - 3'
3' - CCCAAAATCTCGATACGACAAAACTTACCAGGGTTTTGCCCAGAAGCTCTTCTGCAAAATCTCGATACGACAAAACTTACCAGGGTTTTGAAAA - 5'
spacer sequence insert
5' - AAACNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNGT - 3'
3' - NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNCAAAAT - 5'
B
U6
3x
EF1α FLAG NLS hSpCas9 NLS WPRE
BbsI
BbsI
guide sequence insertion site
5' - GTGGAAAGGACGAAACACCGGGTCTTCGAGAAGACCTGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTTTT - 3'
3' - CACCTTTCCTGCTTTGTGGCCCAGAAGCTCTTCTGGACAAAATCTCGATCTTTATCGTTCAATTTTATTCCGATCAGGCAAAAA - 5'
guide sequence insert
5' - CACCGNNNNNNNNNNNNNNNNNNN - 3'
3' - CNNNNNNNNNNNNNNNNNNNCAAA - 5'

FIG. 8A-B



C

	NGG		NNAGAAW	
Chr	median	mean	median	mean
1	7	12.8	67	115.8
2	8	12.7	64	100.8
3	8	13.0	63	98.5
4	9	14.0	61	94.5
5	8	13.1	63	97.9
6	8	13.1	63	98.5
7	8	12.4	64	102.9
8	8	12.8	64	100.9
9	7	13.9	65	120.5
10	7	12.1	66	107.0
11	7	12.0	65	105.8
12	8	12.4	65	103.5
13	8	13.6	62	94.6
14	8	12.0	65	101.5
15	7	11.5	68	107.7
16	7	11.7	74	136.8
17	6	10.3	76	127.9
18	8	13.4	63	101.8
19	6	9.4	82	145.4
20	7	11.1	72	121.8
21	7	13.4	64	111.4
22	6	9.2	85	140.3
X	8	13.2	63	99.0
Y	8	29.2	62	223.7

FIG. 9A-C

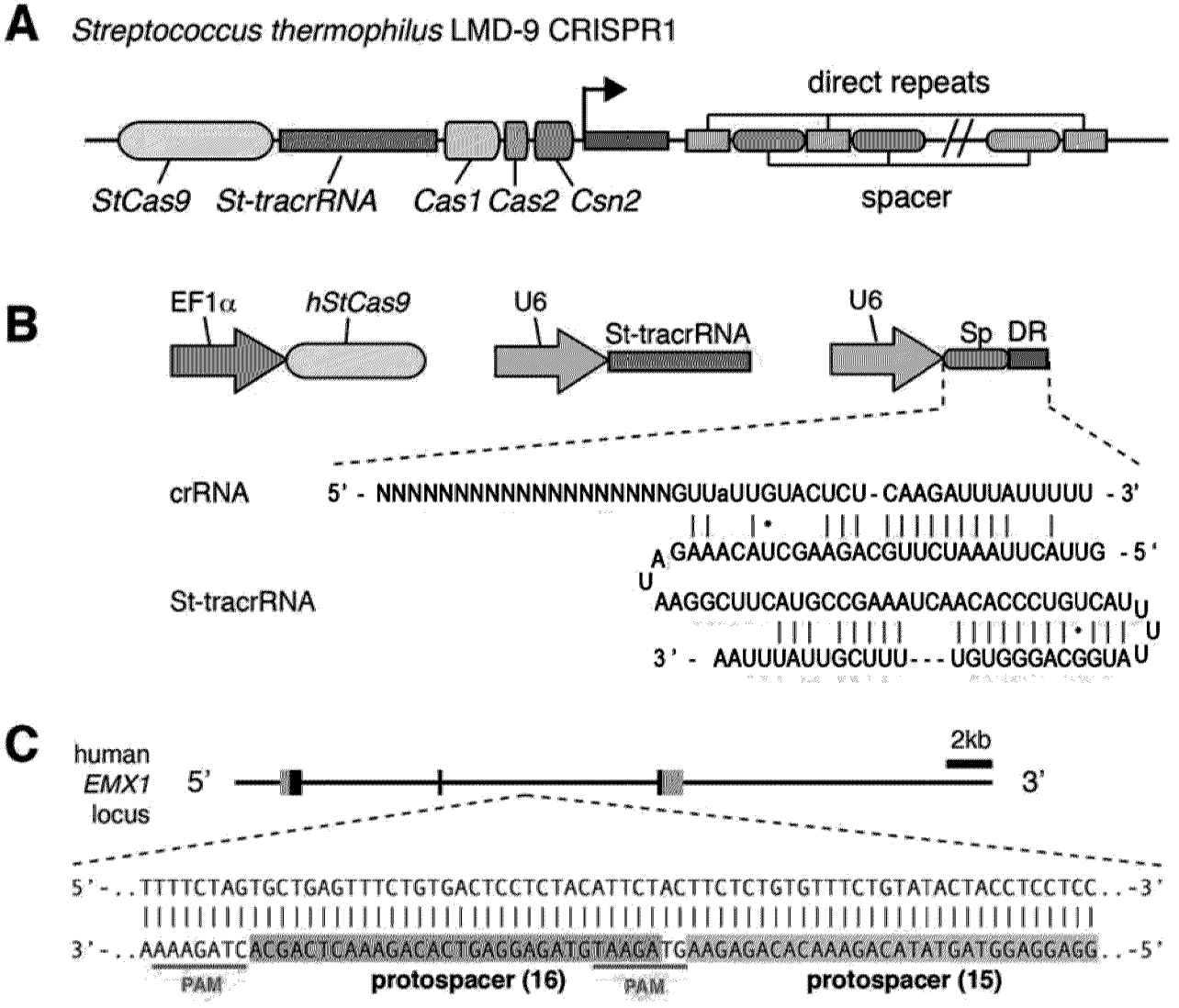

A
Streptococcus thermophilus LMD-9 CRISPR1
direct repeats
StCas9
St-tracrRNA
Cas1
Cas2
Csn2
spacer
B
EF1α
hStCas9
U6
St-tracrRNA
U6
Sp
DR
crRNA
5' - NNNNNNNNNNNNNNNNNNNNGUUaUUGUACUCU - CAAGAUUUAUUUUU - 3'
A GAAACAUCGAAGACGUUCUAAAUUCAUUG - 5'
U
St-tracrRNA
AAGGCUUCAUGCCGAAAUCAACACCCUGUCAU U
U
3' - AAUUUAUUGCUUU - - - UGUGGGACGGUA U
C
human
EMX1
locus
5'
2kb
3'
5'-..TTTTCTAGTGCTGAGTTTCTGTGACTCCTCTACATTCTACTTCTCTGTGTTTCTGTATACTACCTCCTCC..-3'
3'-..AAAAGATCACGACTCAAAGACACTGAGGAGATGTAAGATGAAGAGACACAAAGACATATGATGGAGGAGG..-5'
PAM
protospacer (16)
PAM
protospacer (15)

FIG. 10A-C

D

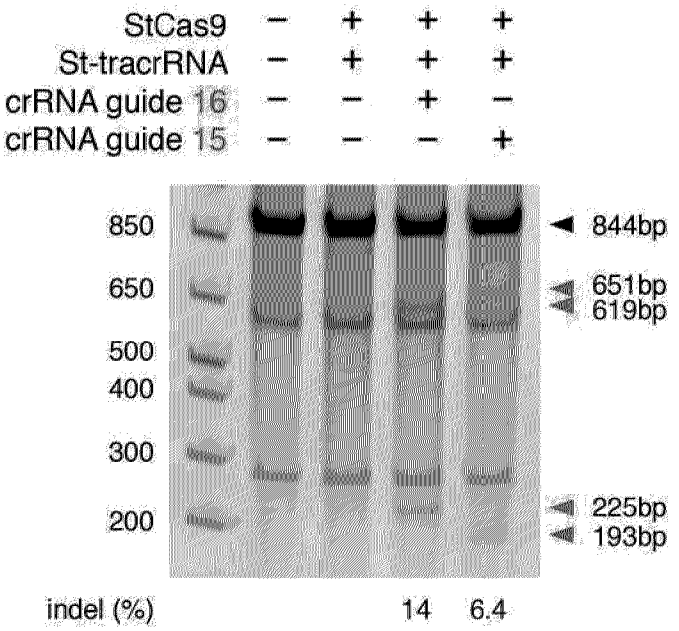

StCas9 − + + +
St-tracrRNA − + + +
crRNA guide 16 − − + −
crRNA guide 15 − − − +
850
650
500
400
300
200
844bp
651bp
619bp
225bp
193bp
indel (%)
14
6.4

FIG. 10D

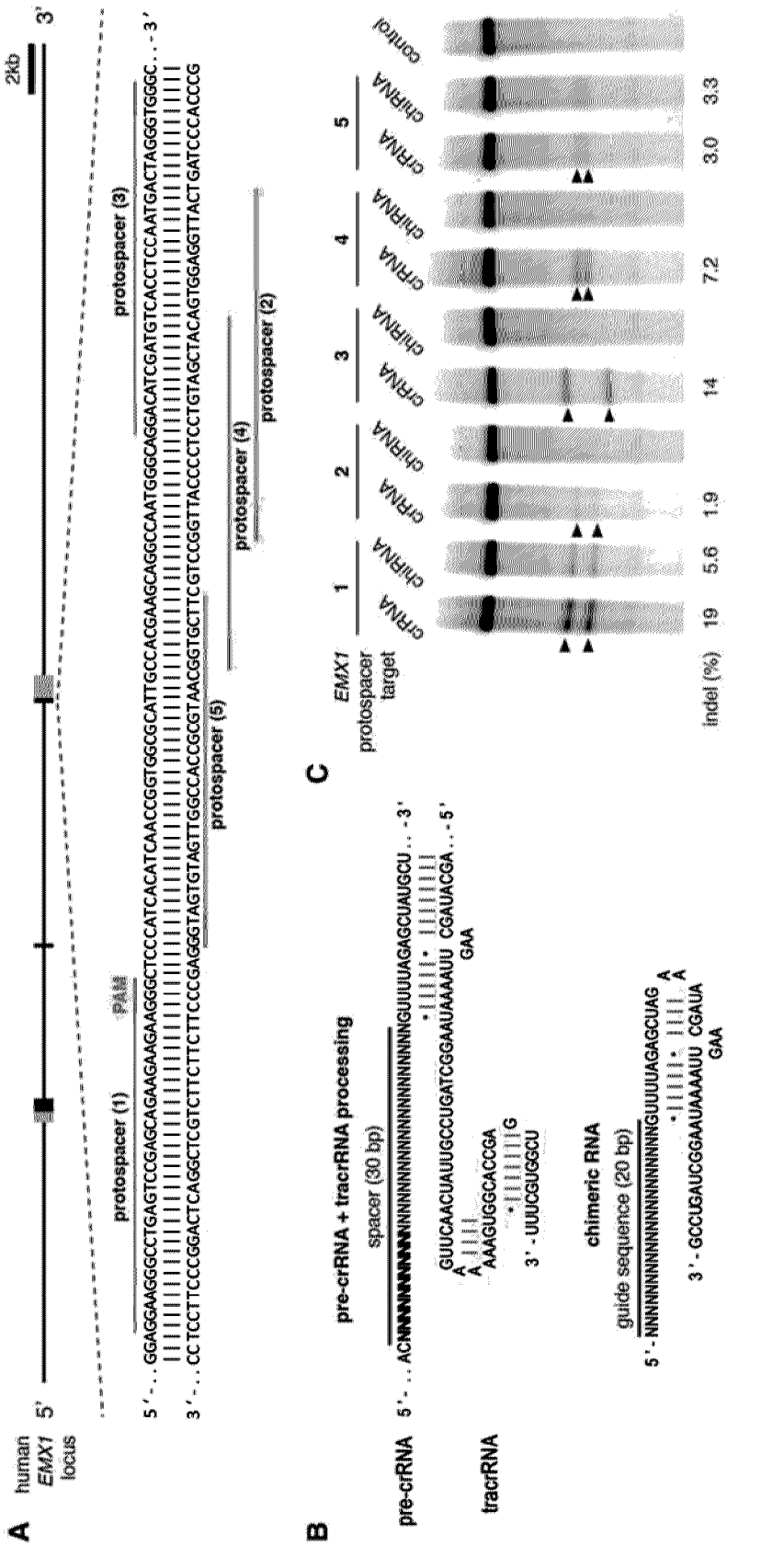

A
human EMX1 locus
5'
3'
2kb
protospacer (1)
PAM
protospacer (3)
protospacer (5)
protospacer (4)
protospacer (2)
B
pre-crRNA + tracrRNA processing
spacer (30 bp)
pre-crRNA
tracrRNA
chimeric RNA
guide sequence (20 bp)
C
EMX1 protospacer target
crRNA
tracrRNA
control
indel (%)
19
5.6
1.9
14
7.2
3.0
3.3

FIG. 11A-C

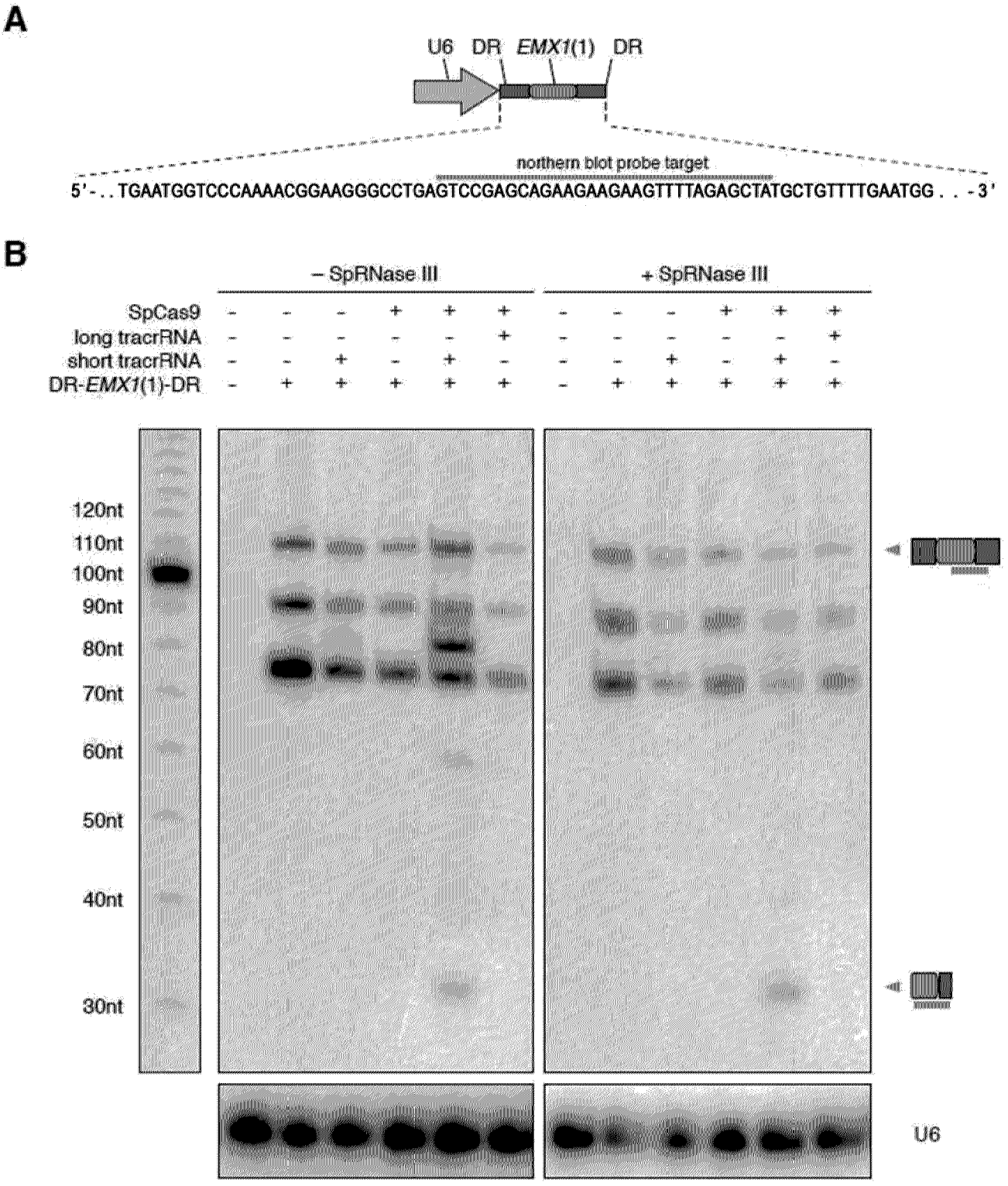

A
U6 DR EMX1(1) DR
northern blot probe target
5'- . . TGAATGGTCCCAAAACGGAAGGGCCTGAGTCCGAGCAGAAGAAGAAGTTTTAGAGCTATGCTGTTTTGAATGG . . - 3'
B
– SpRNase III
+ SpRNase III
SpCas9
long tracrRNA
short tracrRNA
DR-EMX1(1)-DR
120nt
110nt
100nt
90nt
80nt
70nt
60nt
50nt
40nt
30nt
U6

FIG. 12A-B

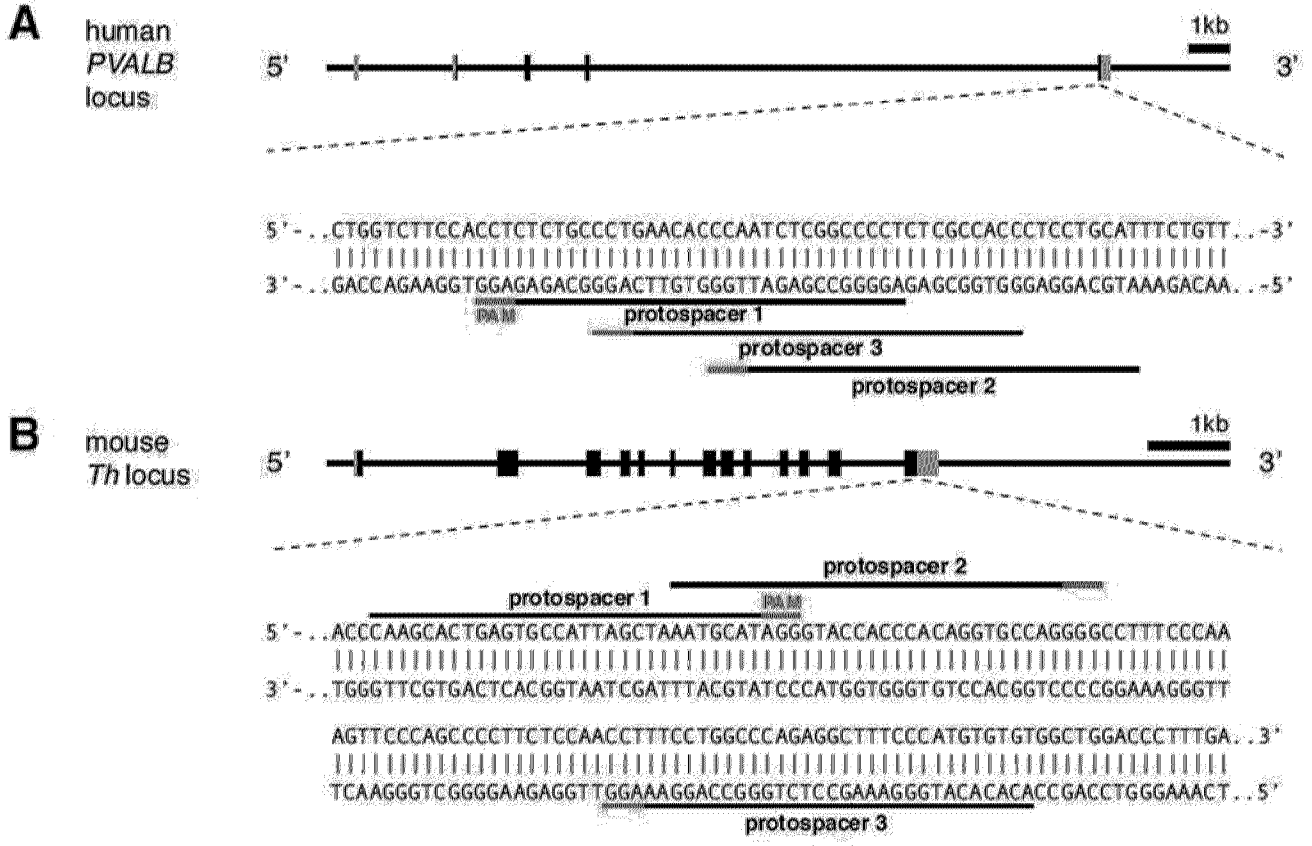

A
human
PVALB
locus
5'
1kb
3'
5'-..CTGGTCTTCCACCTCTCTGCCCTGAACACCCAATCTCGGCCCCTCTCGCCACCCTCCTGCATTTCTGTT..-3'
3'-..GACCAGAAGGTGGAGAGACGGGACTTGTGGGTTAGAGCCGGGGAGAGCGGTGGGAGGACGTAAAGACAA..-5'
PAM
protospacer 1
protospacer 3
protospacer 2
B
mouse
Th locus
5'
1kb
3'
protospacer 2
protospacer 1
PAM
5'-..ACCCAAGCACTGAGTGCCATTAGCTAAATGCATAGGGTACCACCCACAGGTGCCAGGGGCCTTTCCCAA
3'-..TGGGTTCGTGACTCACGGTAATCGATTTACGTATCCCATGGTGGGTGTCCACGGTCCCCGGAAAGGGTT
AGTTCCCAGCCCCTTCTCCAACCTTTCCTGGCCCAGAGGCTTTCCCATGTGTGTGGCTGGACCCTTTGA..3'
TCAAGGGTCGGGGAAGAGGTTGGAAAGGACCGGGTCTCCGAAAGGGTACACACACCGACCTGGGAAACT..5'
protospacer 3

FIG. 13A-B

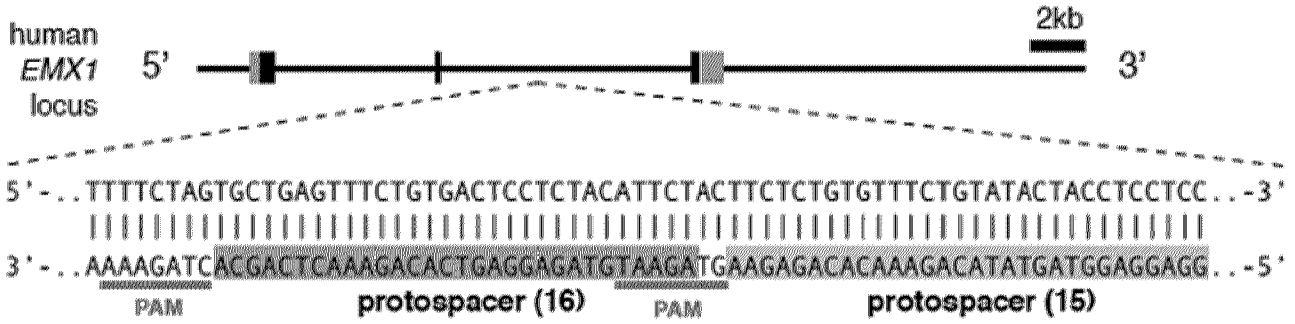

human
EMX1
locus
5'
2kb
3'
5'-..TTTTCTAGTGCTGAGTTTCTGTGACTCCTCTACATTCTACTTCTCTGTGTTTCTGTATACTACCTCCTCC..-3'
3'-..AAAAGATCACGACTCAAAGACACTGAGGAGATGTAAGATGAAGAGACACAAAGACATATGATGGAGGAGG..-5'
PAM
protospacer (16)
PAM
protospacer (15)

FIG. 14

Primer name	Assay	Genomic Target	Primer sequence
Sp-EMX1-F	SURVEYOR assay, sequencing	*EMX1*	AAAACCACCCTTCTCTCTGGC
Sp-EMX1-R	SURVEYOR assay, sequencing	*EMX1*	GGAGATTGGAGACACGGAGAG
Sp-PVALB-F	SURVEYOR assay, sequencing	*PVALB*	CTGGAAAGCCAATGCCTGAC
Sp-PVALB-R	SURVEYOR assay, sequencing	*PVALB*	GGCAGCAAACTCCTTGTCCT
Sp-Th-F	SURVEYOR assay, sequencing	*Th*	GTGCTTTGCAGAGGCCTACC
Sp-Th-R	SURVEYOR assay, sequencing	*Th*	CCTGGAGCGCATGCAGTAGT
St-EMX1-F	SURVEYOR assay, sequencing	*EMX1*	ACCTTCTGTGTTTCCACCATTC
St-EMX1-R	SURVEYOR assay, sequencing	*EMX1*	TTGGGGAGTGCACAGACTTC
Sp-EMX1-RFLP-F	RFLP, sequencing	*EMX1*	GGCTCCCTGGGTTCAAAGTA
Sp-EMX1-RFLP-R	RFLP, sequencing	*EMX1*	AGAGGGGTCTGGATGTCGTAA
Pb_EMX1_sp1	Northern Blot Probe	Not applicable	TAGCTCTAAAACTTCTTCTTCTGCTCGGAC
Pb_tracrRNA	Northern Blot Probe	Not applicable	CTAGCCTTATTTTAACTTGCTATGCTGTTT

FIG. 15

a

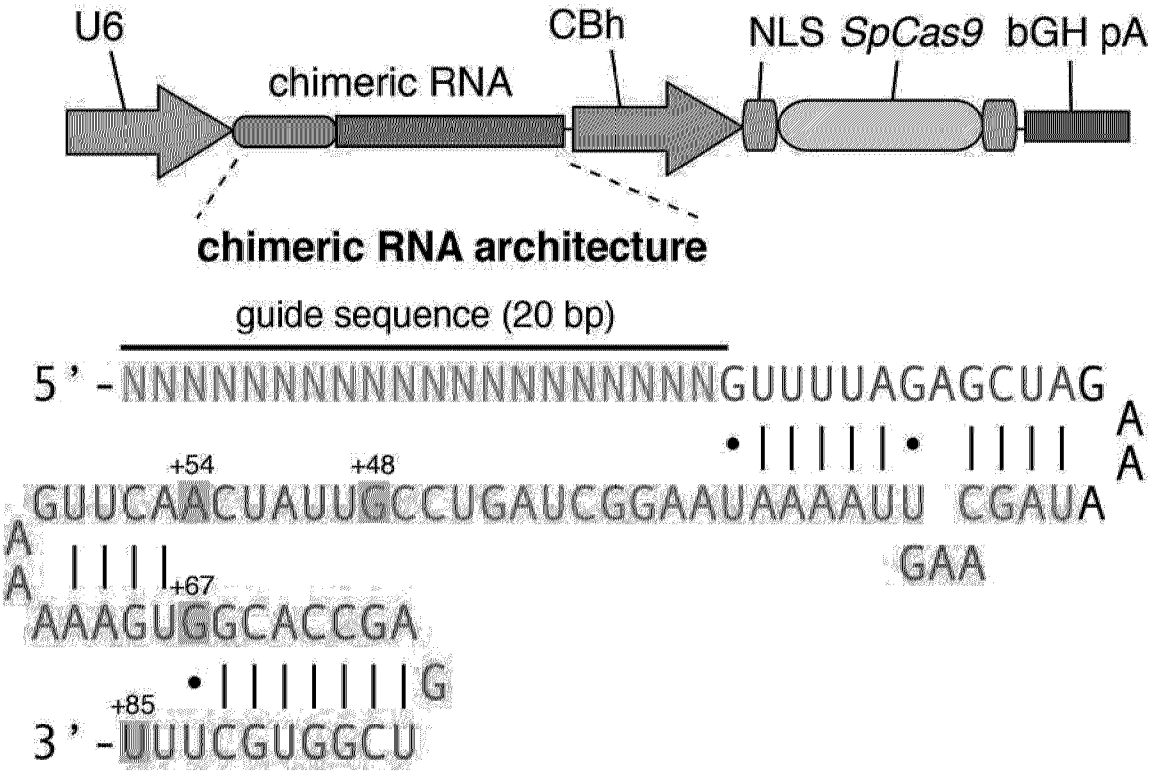

U6
chimeric RNA
CBh
NLS SpCas9 bGH pA
chimeric RNA architecture
guide sequence (20 bp)
5'-NNNNNNNNNNNNNNNNNNNNGUUUUAGAGCUAG
A
A
+54
+48
GUUCAACUAUUGCCUGAUCGGAAUAAAAUU CGAUA
A
A
GAA
+67
AAAGUGGCACCGA
+85
3'-UUUCGUGGCU

FIG. 16A

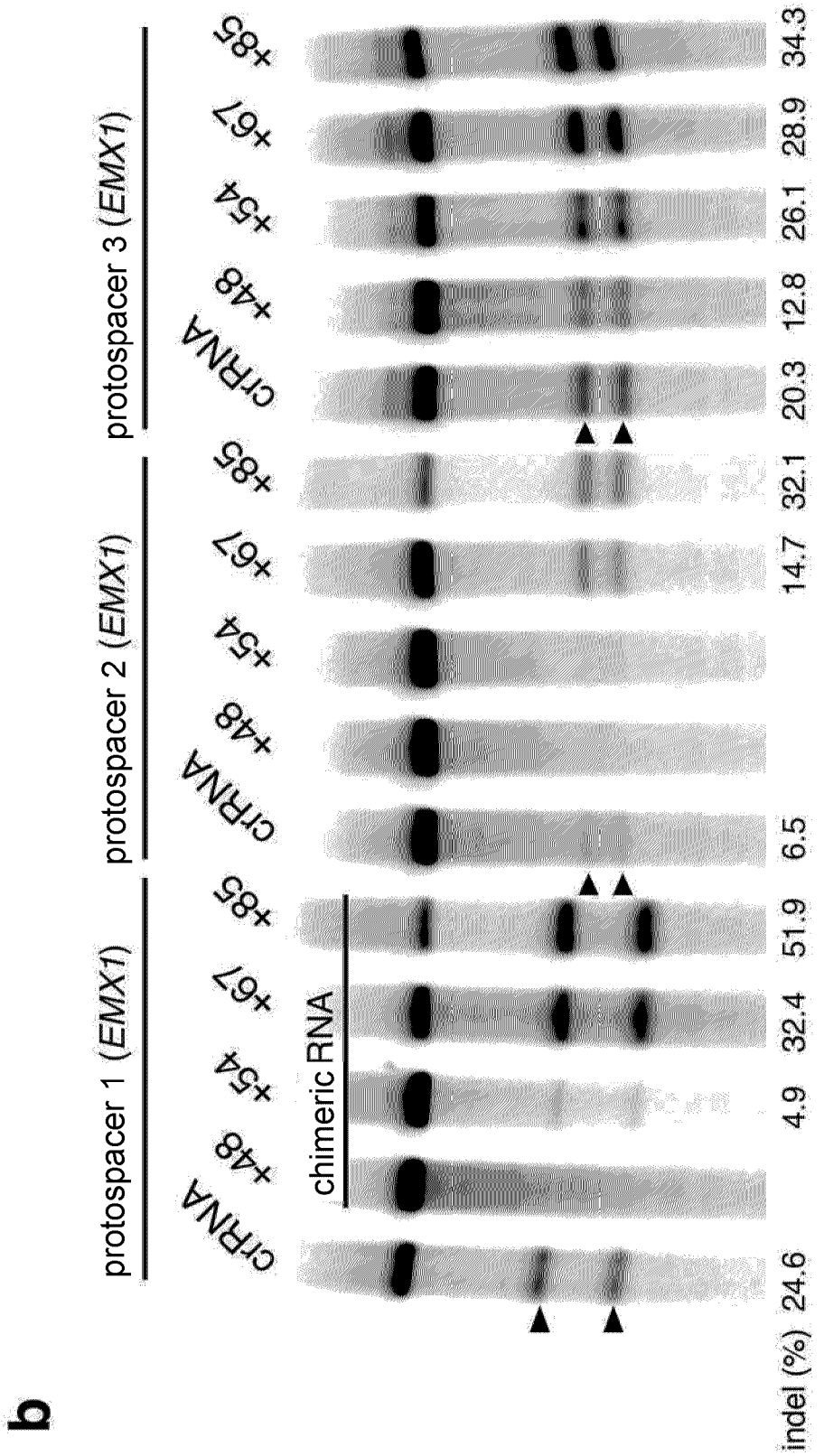

b
protospacer 1 (EMX1)
protospacer 2 (EMX1)
protospacer 3 (EMX1)
crRNA
+48
+54
+67
+85
chimeric RNA
indel (%)
24.6
4.9
32.4
51.9
6.5
14.7
32.1
20.3
12.8
26.1
28.9
34.3

FIG. 16B

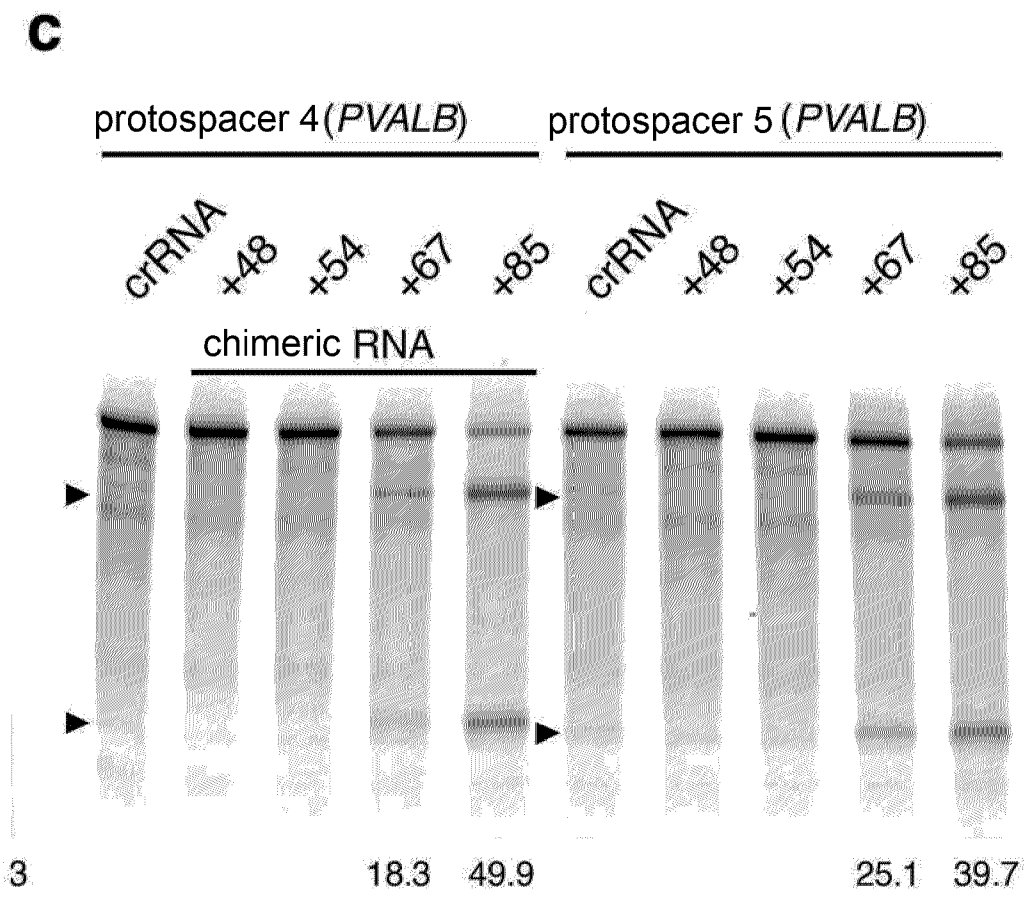

c
protospacer 4 (PVALB)
protospacer 5 (PVALB)
crRNA
+48
+54
+67
+85
crRNA
+48
+54
+67
+85
chimeric RNA
3
18.3
49.9
25.1
39.7

FIG. 16C

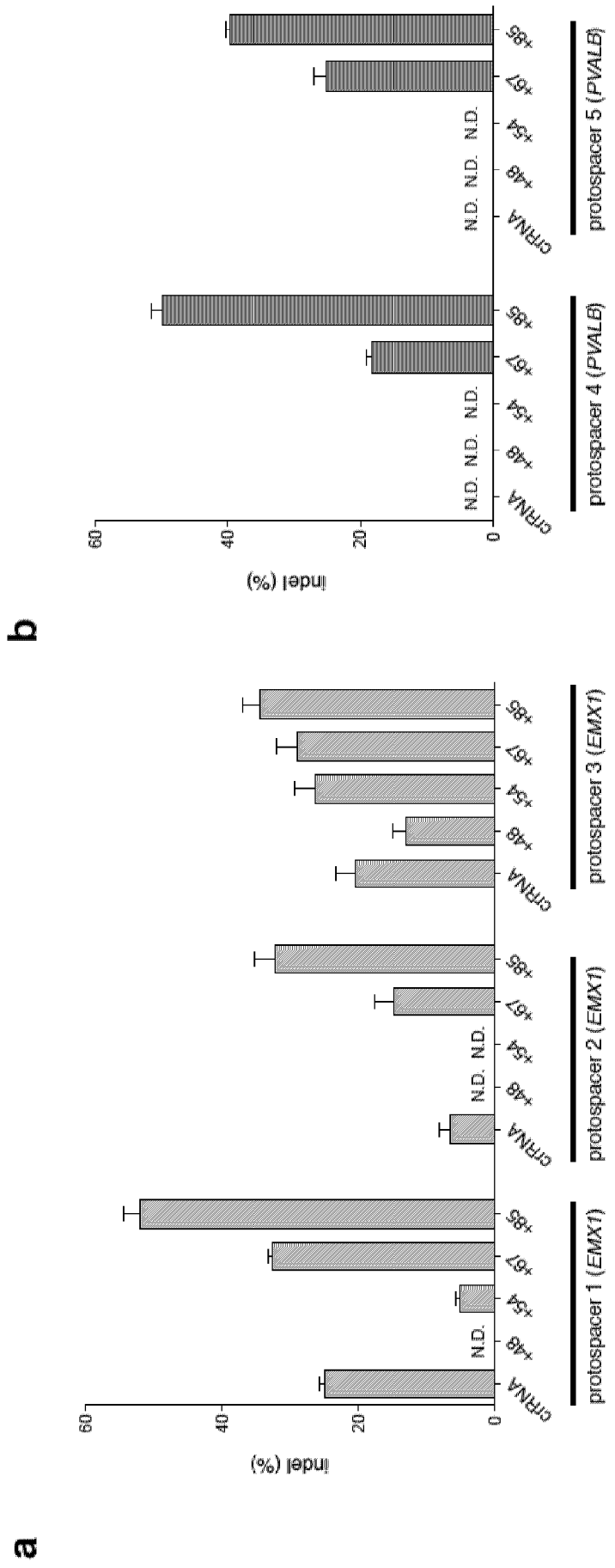

a
b
Indel (%)
0
20
40
60
crRNA
+48
+54
+67
+85
N.D.
protospacer 1 (EMX1)
protospacer 2 (EMX1)
protospacer 3 (EMX1)
protospacer 4 (PVALB)
protospacer 5 (PVALB)

FIG. 17A-B

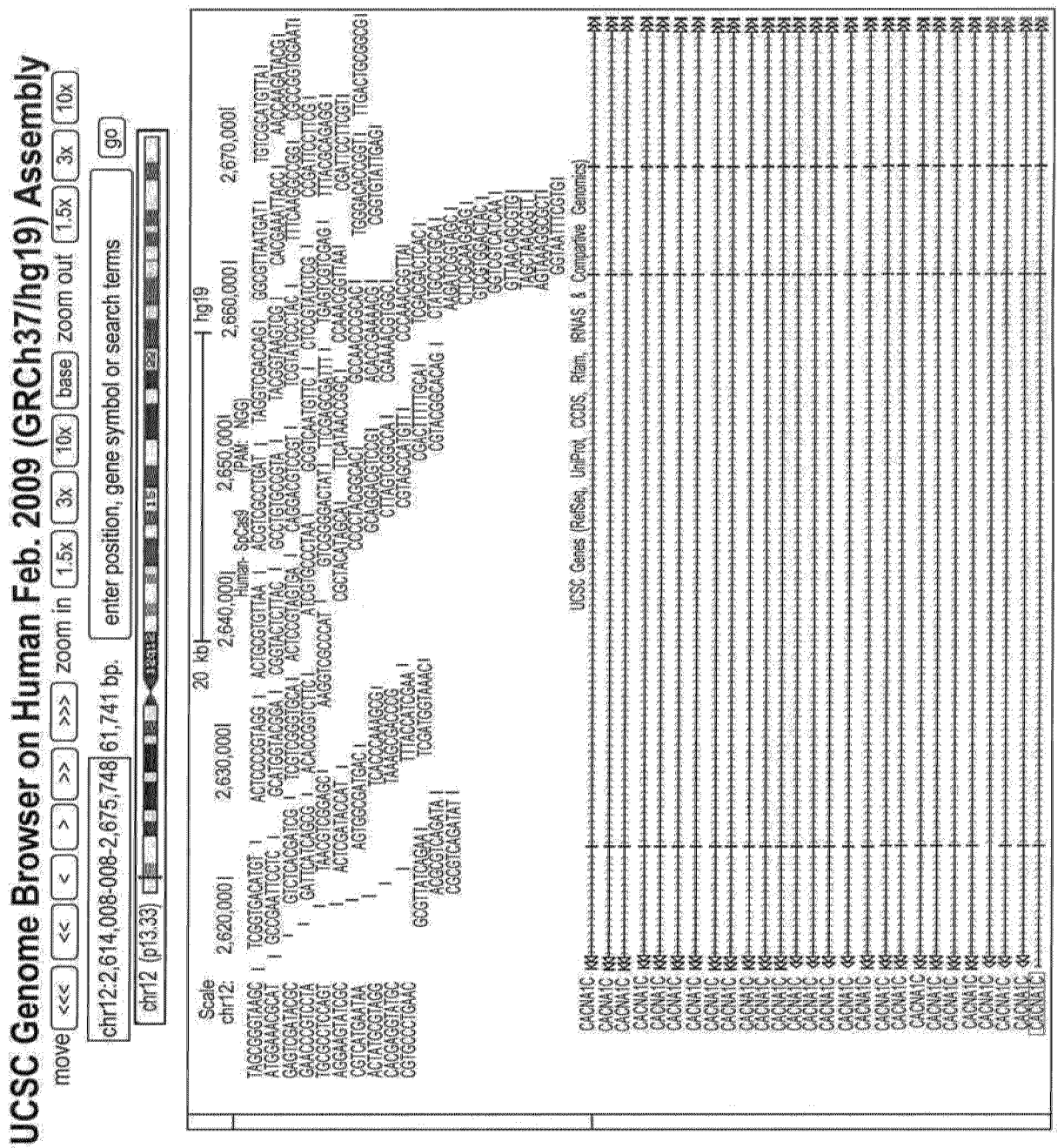

UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly
move <<< << < > >> >>> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x
chr12:2,614,008-008-2,675,748 61,741 bp.
enter position, gene symbol or search terms
go
chr12 (p13.33)
20 kb
hg19
Scale
chr12:
2,620,000
2,630,000
2,640,000
2,650,000
2,660,000
2,670,000
Human-SpCas9 (PAM: NGG)
UCSC Genes (RefSeq, UniProt, CCDS, Rfam, tRNAS & Comparative Genomics)
CACNA1C

FIG. 18

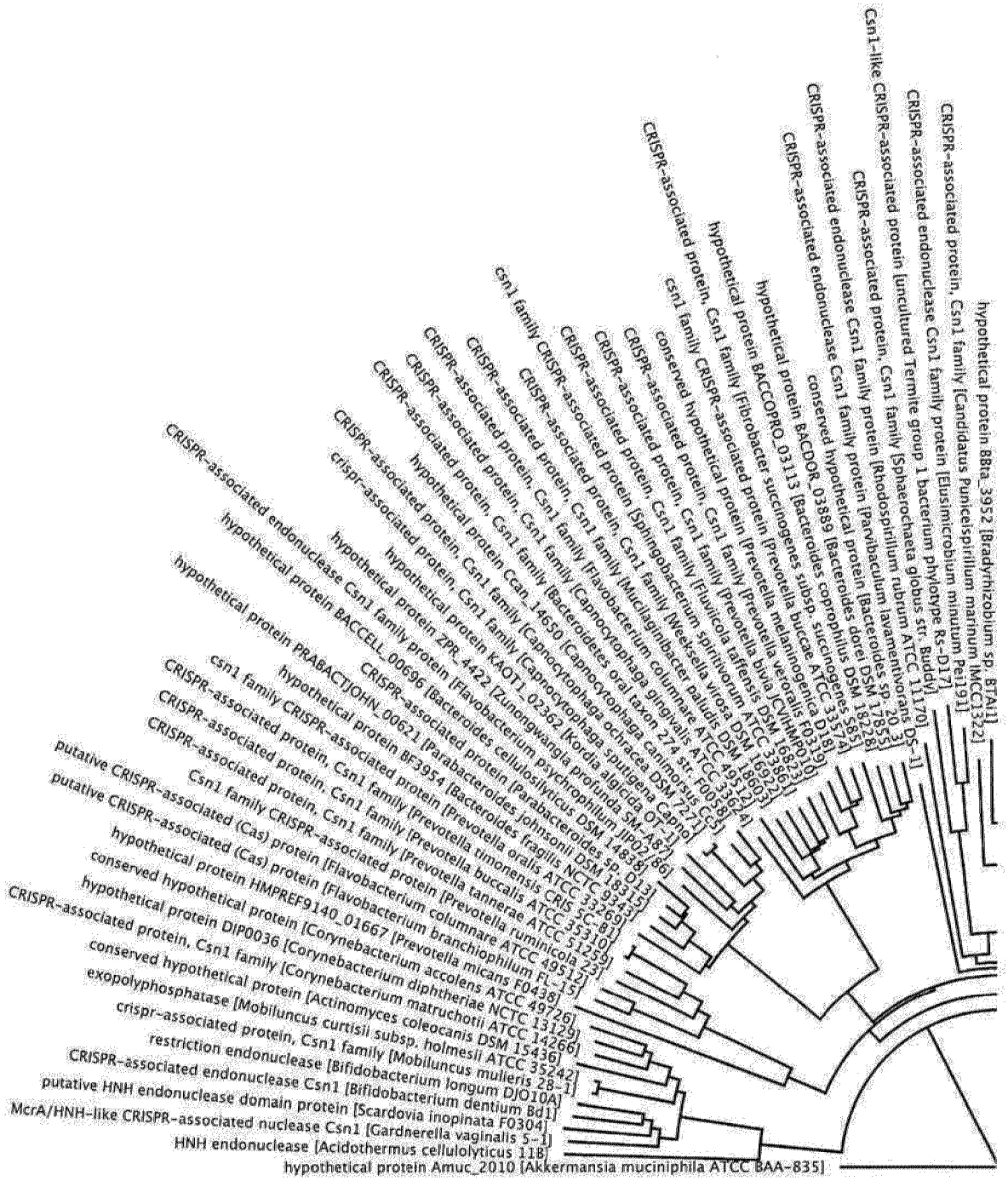

hypothetical protein BBta_3952 [Bradyrhizobium sp. BTAi1]
CRISPR-associated protein, Csn1 family [Candidatus Puniceispirillum marinum IMCC1322]
CRISPR-associated endonuclease Csn1 family protein [Elusimicrobium minutum Pei191]
Csn1-like CRISPR-associated protein [uncultured Termite group 1 bacterium phylotype Rs-D17]
CRISPR-associated protein, Csn1 family protein [Sphaerochaeta globus str. Buddy]
CRISPR-associated endonuclease Csn1 family protein [Parvibaculum lavamentivorans DS-1]
CRISPR-associated protein, Csn1 family protein [Rhodospirillum rubrum ATCC 11170]
CRISPR-associated endonuclease Csn1 family protein [Bacteroides dorei DSM 18628]
conserved hypothetical protein BACDOR_03889 [Bacteroides coprophilus DSM 18228]
hypothetical protein BACCOPRO_03113 [Bacteroides subsp. succinogenes S85]
CRISPR-associated endonuclease Csn1 family protein [Prevotella melaninogenica ATCC 33574]
hypothetical protein [Fibrobacter succinogenes subsp. succinogenes D18]
CRISPR-associated protein, Csn1 family [Prevotella veroralis F0319]
csn1 family CRISPR-associated protein, Csn1 family [Prevotella bivia JCVIHMP022]
conserved hypothetical protein, Csn1 family [Prevotella taffensis DSM 16822]
CRISPR-associated protein, Csn1 family [Fluviicola taffensis DSM 16823]
CRISPR-associated protein, Csn1 family [Flavobacterium spiritivorum ATCC 33861]
CRISPR-associated protein [Sphingobacterium columnare ATCC 43617]
CRISPR-associated protein, Csn1 family [Mucilaginibacter paludis DSM 18603]
CRISPR-associated protein, Csn1 family [Capnocytophaga ginivalis ATCC 33624]
csn1 family CRISPR-associated protein, Csn1 family [Weeksella virosa DSM 16922]
CRISPR-associated protein, Csn1 family [Capnocytophaga ochracea DSM 7271]
hypothetical protein, Csn1 family [Capnocytophaga sputigena Capno]
hypothetical protein Ccan_14650 [Capnocytophaga canimorsus Cc5]
hypothetical protein ZPR_4422 [Zunongwangia profunda SM-A87]
hypothetical protein KAOT1_02362 [Kordia algicida OT-1]
crispr-associated protein, Csn1 family [Flavobacterium psychrophilum JIP02/86]
CRISPR-associated protein [Parabacteroides johnsonii DSM 18315]
hypothetical protein BACCELL_00696 [Bacteroides cellulosilyticus DSM 14838]
CRISPR-associated endonuclease Csn1 family protein [Bacteroides fragilis NCTC 9343]
hypothetical protein PRABACTJOHN_00621 [Parabacteroides sp. D13]
CRISPR-associated protein, Csn1 family [Prevotella oralis ATCC 33269]
csn1 family CRISPR-associated protein, Csn1 family [Prevotella timonensis CRIS 5C-B1]
hypothetical protein BF3954 [Bacteroides thetaiotaomicron VPI-5482]
CRISPR-associated protein, Csn1 family [Prevotella buccalis ATCC 35310]
CRISPR-associated protein, Csn1 family [Prevotella tannerae ATCC 51259]
putative CRISPR-associated (Cas) protein [Flavobacterium columnare ATCC 49512]
Csn1 family CRISPR-associated protein [Prevotella ruminicola 23]
putative CRISPR-associated (Cas) protein [Flavobacterium branchiophilum FL-15]
hypothetical protein HMPREF9140_01667 [Prevotella micans F0438]
conserved hypothetical protein [Corynebacterium accolens ATCC 49726]
hypothetical protein DIP0036 [Corynebacterium diphtheriae NCTC 13129]
CRISPR-associated protein, Csn1 family [Corynebacterium matruchotii ATCC 14266]
conserved hypothetical protein [Actinomyces coleocanis DSM 15436]
exopolyphosphatase [Mobiluncus curtisii subsp. holmesii ATCC 35242]
crispr-associated protein, Csn1 family [Mobiluncus mulieris 28-1]
restriction endonuclease [Bifidobacterium longum DJO10A]
CRISPR-associated endonuclease Csn1 [Bifidobacterium dentium Bd1]
putative HNH endonuclease domain protein [Scardovia inopinata F0304]
McrA/HNH-like CRISPR-associated nuclease Csn1 [Gardnerella vaginalis 5-1]
HNH endonuclease [Acidothermus cellulolyticus 11B]
hypothetical protein Amuc_2010 [Akkermansia muciniphila ATCC BAA-835]

FIG. 19A


hypothetical protein Nham_4054 [Nitrobacter hamburgensis X14]
CRISPR-associated protein, Csn1 family [Azospirillum sp. B510]
hypothetical protein [Dinoroseobacter shibae DFL 12]
Conserved hypothetical protein SUS17_3938 [Sphingomonas sp. S17]
hypothetical protein HDN1F_34120 [gamma proteobacterium HdN1]
CRISPR-associated protein, Csn1 family [Ralstonia syzygii R24]
crispr-associated endonuclease Csn1 family [Nitrosomonas sp. AL212]
CRISPR-associated protein, Csn1 family [Acidovorax avenae subsp. avenae ATCC 19860]
hypothetical protein WS1445 [Wolinella succinogenes DSM 1740]
CRISPR-associated protein, Csn1 family [Alicyclophilus denitrificans BC]
crispr-associated protein, Csn1 family [Alicyclophilus denitrificans K601]
CRISPR-associated protein [Helicobacter cinaedi CCUG 18818]
hypothetical protein CBL_1425 [Campylobacter canadensis MIT 98-5491]
CRISPR-associated protein [Campylobacter jejuni subsp. doyle 269.97]
hypothetical protein CRISPR-associated protein [Campylobacter jejuni subsp. jejuni 81116]
CRISPR-associated protein [Helicobacter mustelae 12198]
CRISPR-associated protein, Csn1 family [Helicobacter cinaedi CCUG 18818]
hypothetical protein CBL_1425 [Campylobacter canadensis MIT 98-5491]
CRISPR-associated protein [Campylobacter jejuni subsp. doyle 269.97]
hypothetical protein RPC_4489 [Rhodopseudomonas palustris BisB18]
CRISPR-associated protein, Csn1 family [Gluconacetobacter diazotrophicus PAl 5]
CRISPR-associated protein GDI_2123 [Gluconacetobacter diazotrophicus PAl 5]
CRISPR-associated protein, Csn1 family [Rhodopseudomonas palustris BisB18]
CRISPR-associated protein, Csn1 family [Rhodopseudomonas palustris BisB5]
CRISPR-associated protein, Csn1 family [Methylocystis sp. ATCC 49242]
CRISPR-associated protein, Csn1 family [Methylosinus trichosporium OB3b]
CRISPR-associated protein, Csn1 family [uncultured Termite group 1 bacterium phylotype Rs-D17]
CRISPR-associated protein, Csn1 family [Clostridium cellulolyticum H10]
CRISPR-associated protein, Csn1 family [Clostridium perfringens D str. JGS1721]
CRISPR-associated protein, Csn1 family [Ilyobacter polytropus DSM 2926]
CRISPR-associated protein, Csn1 family [Bacillus cereus Rock1-15]
CRISPR-associated protein, Csn1 family [Lactobacillus coryniformis subsp. torquens KCTC 3535]
CRISPR-associated protein, Csn1 family [Phascolarctobacterium succinatutens YIT 12067]
CRISPR-associated protein, Csn1 family [Roseburia intestinalis L1-82]
CRISPR-associated protein, Csn1 family [Roseburia inulinivorans DSM 16841]
CRISPR-associated protein, Csn1 family [Aminomonas paucivorans DSM 12260]
CRISPR-associated protein, Csn1 family [Acidovorax ebreus TPSY]
CRISPR-associated protein ROSEINA2194_03455 [Roseburia inulinivorans DSM 16841]
CRISPR-associated protein, Csn1 family [Simonsiella muelleri ATCC 29453]
CRISPR-associated protein, Csn1 family [Neisseria flavescens SK114]
CRISPR-associated protein, Csn1 family [Neisseria cinerea ATCC 14685]
CRISPR-associated protein, Csn1 family [Neisseria lactamica 020-06]
CRISPR-associated protein cas9/csn1, subtype II/nmemi [Simonsiella muelleri ATCC 29453]
CRISPR-associated protein, Csn1 family [Neisseria meningitidis 053442]
CRISPR-associated protein, Csn1 family [Neisseria meningitidis Z2491]
hypothetical protein NLA_17660 [Neisseria meningitidis alpha14]
hypothetical protein NMCC_0397 [Neisseria meningitidis ATCC BAA-1200]
hypothetical protein NMA0631 [Neisseria meningitidis ATCC 23330]
putative CRISPR-associated protein [Neisseria bacilliformis ATCC 43330]
hypothetical protein HMPREF9123_0020 [Neisseria bacilliformis ATCC BAA-1200]
CRISPR-associated endonuclease Csn1 family protein [Kingella kingae ATCC 23330]
hypothetical protein PM1127 [Pasteurella multocida subsp. multocida str. Pm70]
CRISPR-associated endonuclease Csn1 family protein [Actinobacillus minor NM305]
hypothetical protein PARA_18570 [Haemophilus parainfluenzae T3T1]
CRISPR-associated protein, Csn1 family [Actinobacillus pleuropneumoniae serovar 10 str. D13039]
CRISPR-associated endonuclease Csn1 family protein [Actinobacillus succinogenes 130Z]
CRISPR-associated protein, Csn1 family [Ruminococcus albus 8]
CRISPR-associated protein, csn1 family [Nitratifractor salsuginis DSM 16511]
hypothetical protein MMOB0330 [Mycoplasma mobile 163K]

FIG. 19B

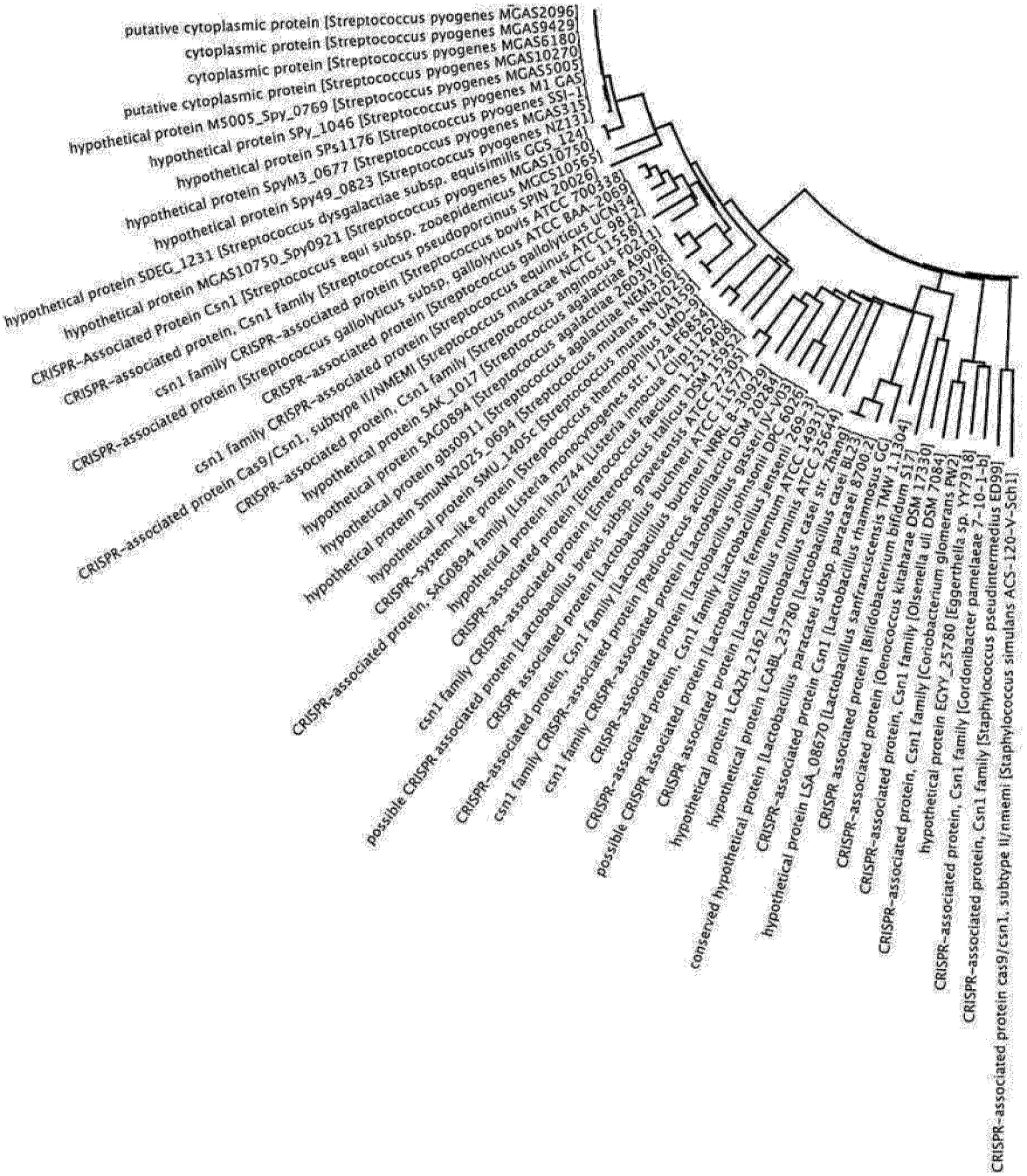

putative cytoplasmic protein [Streptococcus pyogenes MGAS2096]
cytoplasmic protein [Streptococcus pyogenes MGAS9429]
cytoplasmic protein [Streptococcus pyogenes MGAS6180]
putative cytoplasmic protein [Streptococcus pyogenes MGAS10270]
hypothetical protein MS005_Spy_0769 [Streptococcus pyogenes MGAS5005]
hypothetical protein SPy_1046 [Streptococcus pyogenes M1 GAS]
hypothetical protein SPs1176 [Streptococcus pyogenes SSI-1]
hypothetical protein SpyM3_0677 [Streptococcus pyogenes MGAS315]
hypothetical protein Spy49_0823 [Streptococcus pyogenes NZ131]
hypothetical protein SDEG_1231 [Streptococcus dysgalactiae subsp. equisimilis GGS_124]
hypothetical protein MGAS10750_Spy0921 [Streptococcus pyogenes MGAS10750]
CRISPR-Associated Protein Csn1 [Streptococcus equi subsp. zooepidemicus MGCS10565]
CRISPR-associated protein, Csn1 family [Streptococcus pseudoporcinus SPIN 20026]
csn1 family CRISPR-associated protein [Streptococcus gallolyticus subsp. gallolyticus ATCC BAA-2069]
CRISPR-associated protein [Streptococcus gallolyticus subsp. gallolyticus ATCC 700338]
CRISPR-associated protein [Streptococcus equinus ATCC 9812]
CRISPR-associated protein, Csn1 family [Streptococcus macacae NCTC 11558]
csn1 family CRISPR-associated protein [Streptococcus agalactiae F021_1]
CRISPR-associated Cas9/Csn1, subtype II/NMEMI [Streptococcus agalactiae A909]
CRISPR-associated protein, Csn1 family [Streptococcus agalactiae 2603V/R]
hypothetical protein SAK_1017 [Streptococcus agalactiae NEM316]
hypothetical protein gbs0911 [Streptococcus agalactiae NN2025]
hypothetical protein SmuNN2025_0694 [Streptococcus mutans UA159]
hypothetical protein SMU_1405c [Streptococcus mutans NN2025]
CRISPR-system-like protein [Streptococcus thermophilus LMD-9]
CRISPR-associated protein, SAG0894 family [Listeria monocytogenes str. 1/2a F6854]
hypothetical protein lin2744 [Listeria innocua Clip11262]
CRISPR-associated protein [Enterococcus faecium 1231408]
csn1 family CRISPR-associated protein [Lactobacillus brevis subsp. gravesensis ATCC 15305]
CRISPR-associated protein, Csn1 family [Lactobacillus buchneri NRRL B-30929]
csn1 family CRISPR-associated protein [Pediococcus acidilactici DSM 20284]
CRISPR-associated protein, Csn1 family [Lactobacillus johnsonii DPC 6026]
CRISPR-associated protein [Lactobacillus gasseri JV-V03]
possible CRISPR associated protein [Lactobacillus fermentum ATCC 14931]
CRISPR associated protein LCAZH_2162 [Lactobacillus ruminis ATCC 25644]
hypothetical protein LCABL_23780 [Lactobacillus casei str. Zhang]
hypothetical protein [Lactobacillus paracasei subsp. paracasei 8700:2]
conserved hypothetical protein LSA_08670 [Lactobacillus sanfranciscensis TMW 1.1304]
CRISPR-associated protein Csn1 [Lactobacillus casei BL23]
CRISPR-associated protein [Bifidobacterium bifidum S17]
CRISPR-associated protein [Oenococcus kitaharae DSM 17330]
CRISPR-associated protein, Csn1 family [Olsenella uli DSM 7084]
hypothetical protein EGYY_25780 [Eggerthella sp. YY7918]
CRISPR-associated protein, Csn1 family [Gordonibacter pamelaeae 7-10-1-b]
CRISPR-associated protein, Csn1 family [Staphylococcus pseudintermedius ED99]
CRISPR-associated protein cas9/csn1, subtype II/nmemi [Staphylococcus simulans ACS-120-V-Sch1]

FIG. 19C

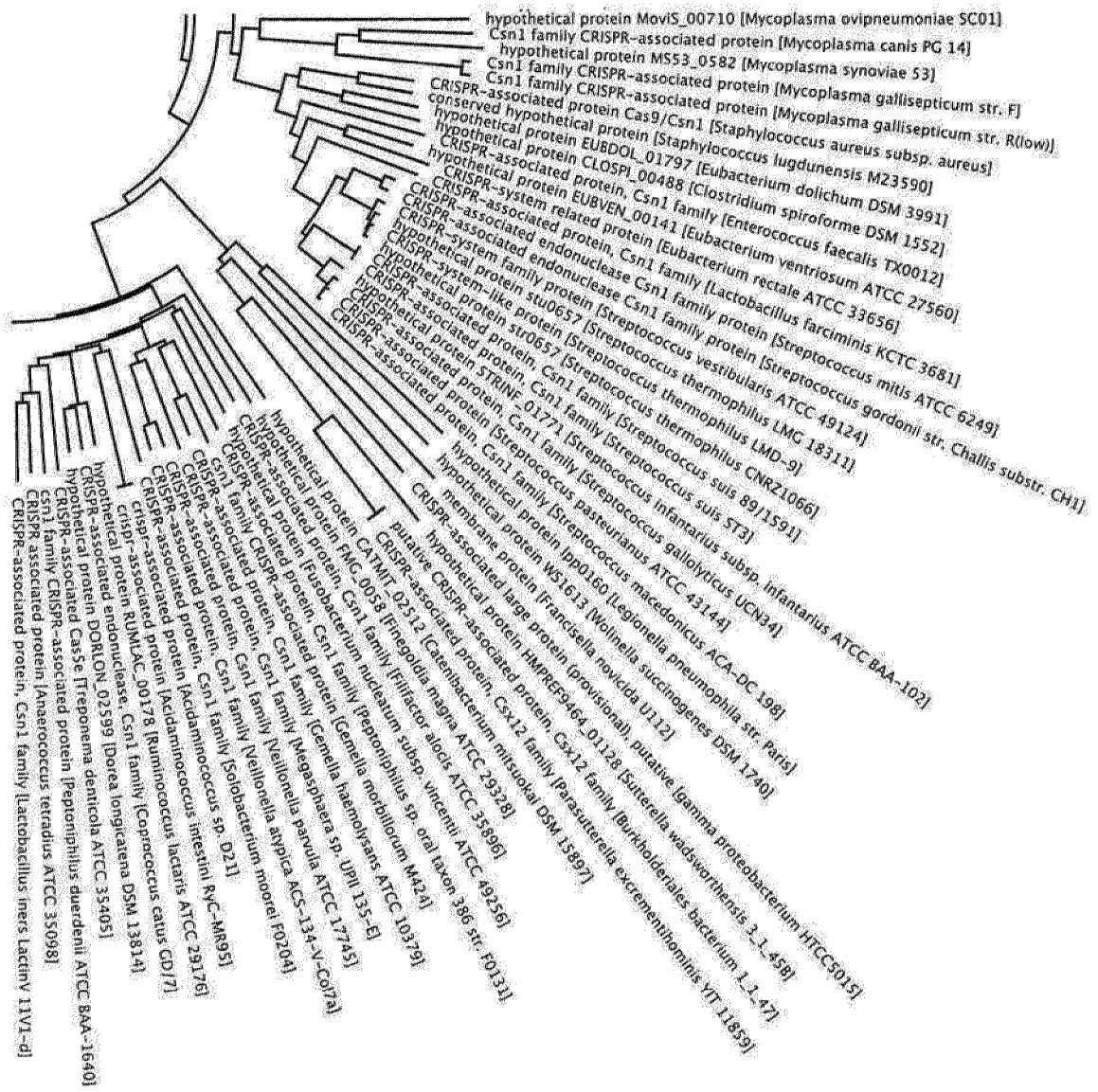

hypothetical protein MoviS_00710 [Mycoplasma ovipneumoniae SC01]
Csn1 family CRISPR-associated protein [Mycoplasma canis PG 14]
hypothetical protein MS53_0582 [Mycoplasma synoviae 53]
Csn1 family CRISPR-associated protein [Mycoplasma gallisepticum str. F]
CRISPR-associated protein Cas9/Csn1 [Staphylococcus aureus subsp. aureus]
conserved hypothetical protein [Staphylococcus lugdunensis M23590]
hypothetical protein EUBDOL_01797 [Eubacterium dolichum DSM 3991]
hypothetical protein CLOSPI_00488 [Clostridium spiroforme DSM 1552]
CRISPR-associated protein, Csn1 family [Enterococcus faecalis TX0012]
CRISPR-system related protein [Eubacterium ventriosum ATCC 27560]
CRISPR-associated protein, Csn1 family protein [Lactobacillus farciminis KCTC 3681]
CRISPR-associated endonuclease Csn1 family protein [Streptococcus mitis ATCC 6249]
hypothetical protein EUBVEN_00141 [Eubacterium rectale ATCC 33656]
CRISPR-system family protein [Streptococcus gordonii str. Challis substr. CH1]
hypothetical protein stu0657 [Streptococcus vestibularis ATCC 49124]
CRISPR-system-like protein str0657 [Streptococcus thermophilus LMG 18311]
hypothetical protein STRINF_01771 [Streptococcus thermophilus LMD-9]
CRISPR-associated protein, Csn1 family [Streptococcus thermophilus CNRZ1066]
CRISPR-associated protein, Csn1 family [Streptococcus suis 89/1591]
CRISPR-associated protein, Csn1 family [Streptococcus infantarius subsp. infantarius ATCC BAA-102]
CRISPR-associated protein, Csn1 family [Streptococcus pasteurianus ATCC 43144]
membrane protein lpp0160 [Legionella pneumophila str. Paris]
hypothetical protein WS1613 [Molinella succinogenes DSM 1740]
hypothetical protein HMPREF9464_01128 [Sutterella wadsworthensis 3_1_45B]
CRISPR-associated large protein (provisional), putative (gamma proteobacterium HTCC5015]
hypothetical protein Csx12 family [Parasutterella excrementihominis YIT 11859]
CRISPR-associated protein, Csn1 family [Burkholderiales bacterium 1_1_47]
putative CRISPR-associated protein, Csx12 family [Catenibacterium mitsuokai DSM 15897]
hypothetical protein FMG_0058 [Finegoldia magna ATCC 29328]
CRISPR-associated protein, Csn1 family [Peptoniphilus sp. oral taxon 386 str. F0131]
hypothetical protein [Fusobacterium nucleatum subsp. vincentii ATCC 49256]
CRISPR-associated protein, Csn1 family [Gemella haemolysans ATCC 10379]
hypothetical protein [Megasphaera sp. UPII 135-E]
CRISPR-associated protein, Csn1 family [Veillonella atypica ACS-134-V-Col7a]
CRISPR-associated protein, Csn1 family [Veillonella parvula ATCC 17745]
Csn1 family [Solobacterium moorei F0204]
CRISPR-associated protein, Csn1 family [Acidaminococcus intestini RyC-MR95]
CRISPR-associated protein [Acidaminococcus sp. D21]
CRISPR-associated protein [Ruminococcus lactaris ATCC 29176]
crispr-associated protein RUMLAC_00178 [Ruminococcus lactaris ATCC 29176]
crispr-associated endonuclease, Csn1 family [Coprococcus catus GD/7]
hypothetical protein DORLON_02599 [Dorea longicatena DSM 13814]
CRISPR-associated Cas5e [Treponema denticola ATCC 35405]
hypothetical protein [Peptoniphilus duerdenii ATCC BAA-1640]
CRISPR-associated protein [Anaerococcus tetradius ATCC 35098]
Csn1 family CRISPR-associated protein [Lactobacillus iners LactinV 11V1-d]
CRISPR-associated protein, Csn1 family [Lactobacillus iners LactinV 11V1-d]

FIG. 19D

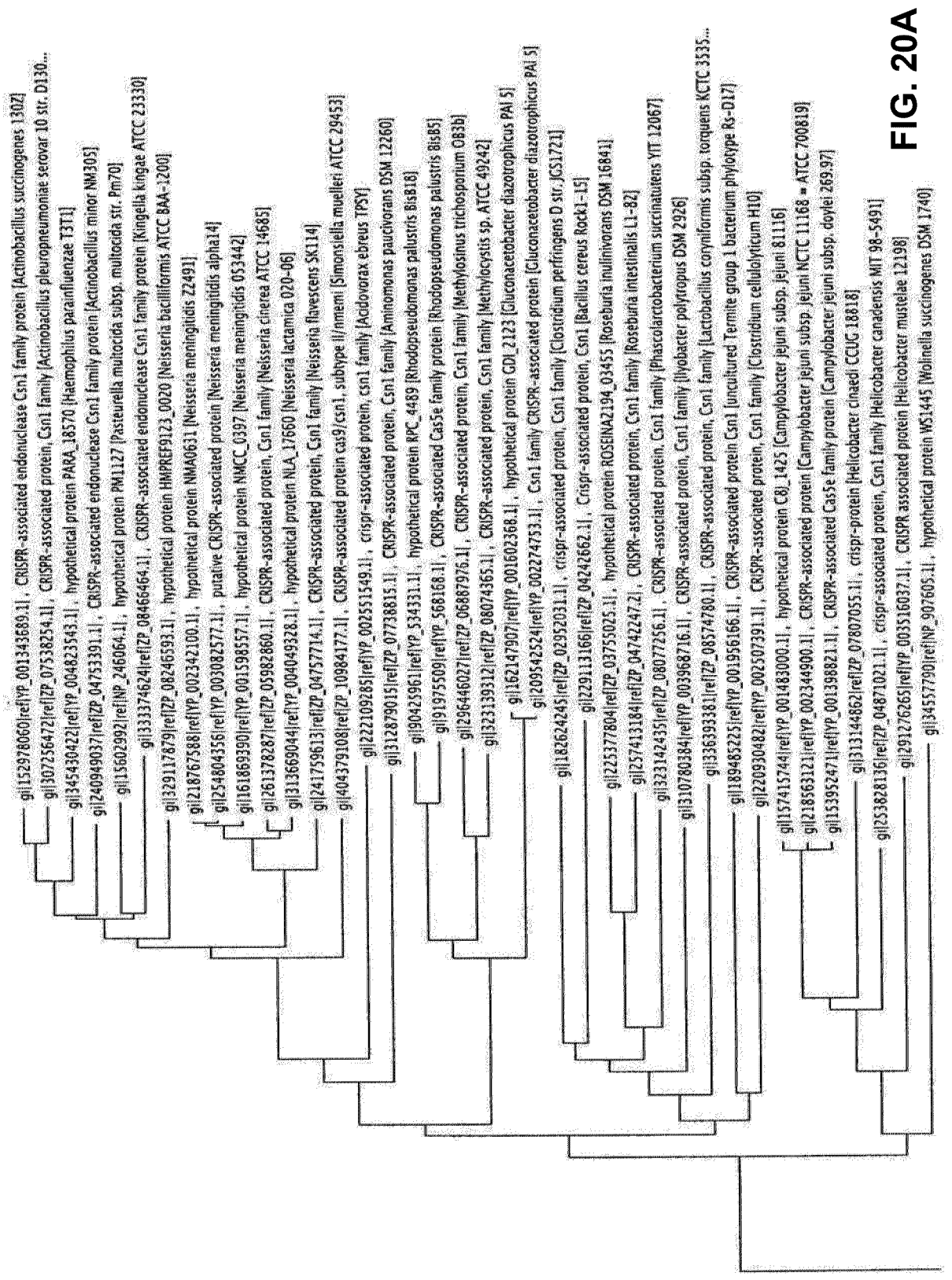

gi|152978060|ref|YP_001343689.1|, CRISPR-associated endonuclease Csn1 family protein [Actinobacillus succinogenes 130Z]
gi|307256472|ref|ZP_07538254.1|, CRISPR-associated protein, Csn1 family [Actinobacillus pleuropneumoniae serovar 10 str. D130...
gi|345430422|ref|YP_004823543.1|, hypothetical protein PARA_18570 [Haemophilus parainfluenzae T3T1]
gi|240949037|ref|ZP_04753391.1|, CRISPR-associated endonuclease Csn1 family protein [Actinobacillus minor NM305]
gi|15602992|ref|NP_246064.1|, hypothetical protein PM1127 [Pasteurella multocida subsp. multocida str. Pm70]
gi|333374624|ref|ZP_08466464.1|, CRISPR-associated endonuclease Csn1 family protein [Kingella kingae ATCC 23330]
gi|329117879|ref|ZP_08246593.1|, hypothetical protein HMPREF9123_0020 [Neisseria bacilliformis ATCC BAA-1200]
gi|218767588|ref|YP_002342100.1|, hypothetical protein NMA0631 [Neisseria meningitidis Z2491]
gi|254804356|ref|YP_003082577.1|, putative CRISPR-associated protein [Neisseria meningitidis alpha14]
gi|161869390|ref|YP_001598557.1|, hypothetical protein NMCC_0397 [Neisseria meningitidis 053442]
gi|261378287|ref|ZP_05982860.1|, CRISPR-associated protein, Csn1 family [Neisseria cinerea ATCC 14685]
gi|313669044|ref|YP_004049328.1|, hypothetical protein NLA_17660 [Neisseria lactamica 020-06]
gi|241759613|ref|ZP_04757714.1|, CRISPR-associated protein, Csn1 family [Neisseria flavescens SK114]
gi|404379108|ref|ZP_10984177.1|, CRISPR-associated protein cas9/csn1, subtype II/nmeni [Simonsiella muelleri ATCC 29453]
gi|222109285|ref|YP_002551549.1|, crispr-associated protein, csn1 family [Acidovorax ebreus TPSY]
gi|312879015|ref|ZP_07738815.1|, CRISPR-associated protein, Csn1 family [Aminomonas paucivorans DSM 12260]
gi|90425961|ref|YP_534331.1|, hypothetical protein RPC_4489 [Rhodopseudomonas palustris BisB18]
gi|91975509|ref|YP_568168.1|, CRISPR-associated Cas5e family protein [Rhodopseudomonas palustris BisB5]
gi|296446027|ref|ZP_06887976.1|, CRISPR-associated protein, Csn1 family [Methylosinus trichosporium OB3b]
gi|323139312|ref|ZP_08074365.1|, CRISPR-associated protein, Csn1 family [Methylocystis sp. ATCC 49242]
gi|162147907|ref|YP_001602368.1|, hypothetical protein GDI_2123 [Gluconacetobacter diazotrophicus PAl 5]
gi|209542524|ref|YP_002274753.1|, Csn1 family CRISPR-associated protein [Gluconacetobacter diazotrophicus PAl 5]
gi|18624245|ref|ZP_02952031.1|, crispr-associated protein, Csn1 family [Clostridium perfringens D str. JGS1721]
gi|229113166|ref|ZP_04242662.1|, Crispr-associated protein, Csn1 [Bacillus cereus Rock1-15]
gi|225377804|ref|ZP_03755025.1|, hypothetical protein ROSEINA2194_03455 [Roseburia inulinivorans DSM 16841]
gi|257413184|ref|ZP_04742247.2|, CRISPR-associated protein, Csn1 family [Roseburia intestinalis L1-82]
gi|323142435|ref|ZP_08077256.1|, CRISPR-associated protein, Csn1 family [Phascolarctobacterium succinatutens YIT 12067]
gi|310780384|ref|YP_003968716.1|, CRISPR-associated protein, Csn1 family [Ilyobacter polytropus DSM 2926]
gi|336393381|ref|ZP_08574780.1|, CRISPR-associated protein, Csn1 family [Lactobacillus coryniformis subsp. torquens KCTC 3535...
gi|189485225|ref|YP_001956166.1|, CRISPR-associated protein Csn1 [uncultured Termite group 1 bacterium phylotype Rs-D17]
gi|220930482|ref|YP_002507391.1|, CRISPR-associated protein, Csn1 family [Clostridium cellulolyticum H10]
gi|157415744|ref|YP_001483000.1|, hypothetical protein C8J_1425 [Campylobacter jejuni subsp. jejuni 81116]
gi|218563121|ref|YP_002344900.1|, CRISPR-associated protein [Campylobacter jejuni subsp. jejuni NCTC 11168 = ATCC 700819]
gi|153952471|ref|YP_001398821.1|, CRISPR-associated Cas5e family protein [Campylobacter jejuni subsp. doylei 269.97]
gi|313144662|ref|ZP_07807055.1|, crispr-protein [Helicobacter cinaedi CCUG 18818]
gi|253828136|ref|ZP_04871021.1|, crispr-associated protein, Csn1 family [Helicobacter canadensis MIT 98-5491]
gi|291276265|ref|YP_003516037.1|, CRISPR associated protein [Helicobacter mustelae 12198]
gi|345577790|ref|NP_907605.1|, hypothetical protein WS1445 [Wolinella succinogenes DSM 1740]

FIG. 20A

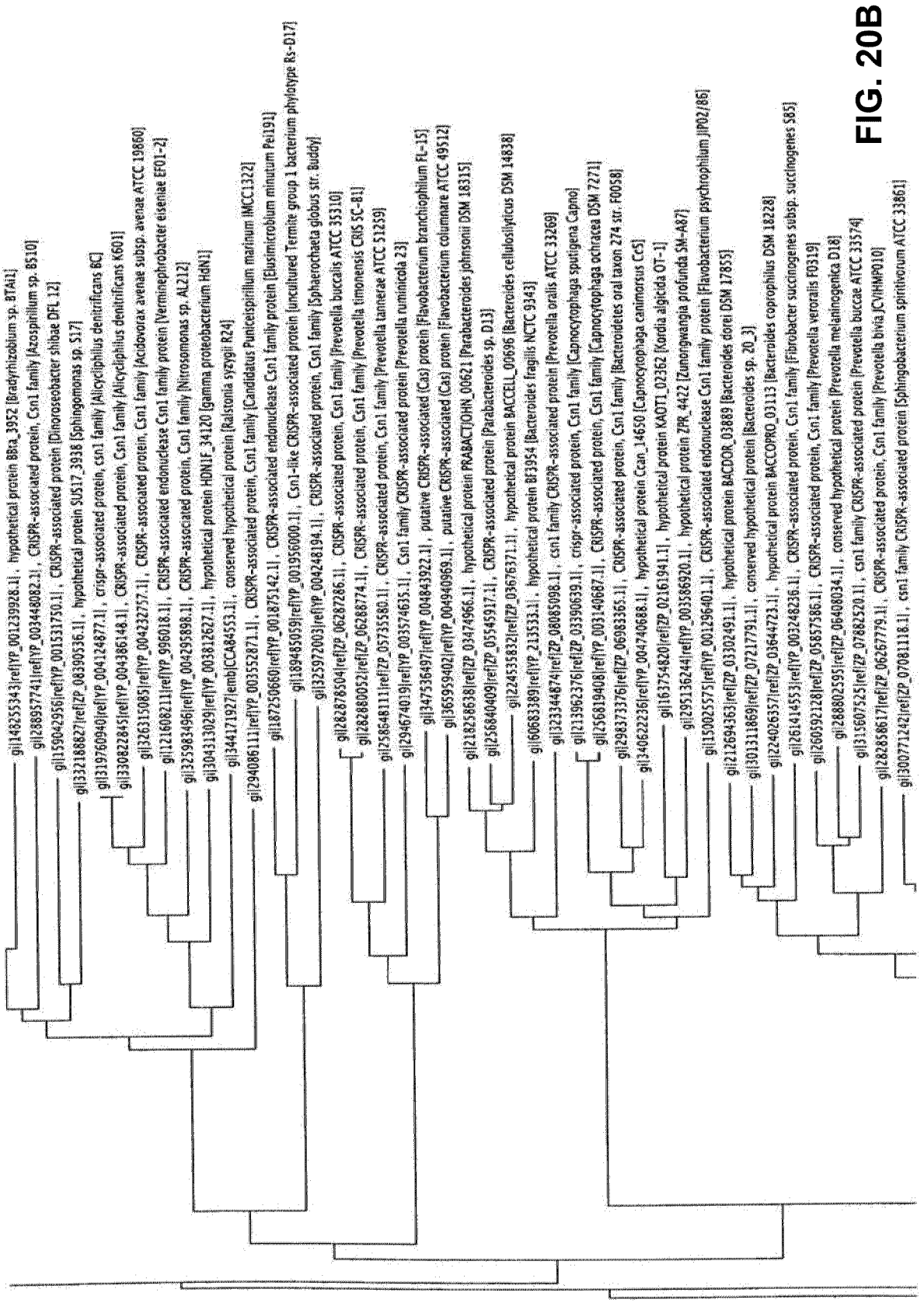

gi|148255343|ref|YP_001239928.1| , hypothetical protein BBta_3952 [Bradyrhizobium sp. BTAi1]
gi|238957741|ref|YP_003448082.1| , CRISPR-associated protein, Csn1 family [Azospirillum sp. B510]
gi|159042956|ref|YP_001531750.1| , CRISPR-associated protein [Dinoroseobacter shibae DFL 12]
gi|332188827|ref|ZP_08390536.1| , hypothetical protein SU17_3938 [Sphingomonas sp. S17]
gi|319760940|ref|YP_004124877.1| , crispr-associated protein, csn1 family [Alicycliphilus denitrificans BC]
gi|330822845|ref|YP_004386148.1| , CRISPR-associated protein, Csn1 family [Alicycliphilus denitrificans K601]
gi|326315085|ref|YP_004232757.1| , CRISPR-associated protein, Csn1 family [Acidovorax avenae subsp. avenae ATCC 19860]
gi|121608211|ref|YP_996018.1| , CRISPR-associated endonuclease Csn1 family protein [Verminephrobacter eiseniae EF01-2]
gi|325983496|ref|YP_004295898.1| , CRISPR-associated protein, Csn1 family [Nitrosomonas sp. AL212]
gi|304313029|ref|YP_003812627.1| , hypothetical protein HDN1F_34120 [gamma proteobacterium HdN1]
gi|344171927|emb|CCA84553.1| , conserved hypothetical protein [Ralstonia syzygii R24]
gi|294086111|ref|YP_003552871.1| , CRISPR-associated protein, Csn1 family [Candidatus Puniceispirillum marinum IMCC1322]
gi|187250660|ref|YP_001875142.1| , CRISPR-associated endonuclease Csn1 family protein [Elusimicrobium minutum Pei191]
gi|189485059|ref|YP_001956000.1| , Csn1-like CRISPR-associated protein [uncultured Termite group 1 bacterium phylotype Rs-D17]
gi|325972003|ref|YP_004248194.1| , CRISPR-associated protein, Csn1 family [Sphaerochaeta globus str. Buddy]
gi|282878504|ref|ZP_06287286.1| , CRISPR-associated protein, Csn1 family [Prevotella buccalis ATCC 35310]
gi|282880052|ref|ZP_06288774.1| , CRISPR-associated protein, Csn1 family [Prevotella timonensis CRIS 5C-B1]
gi|258548111|ref|ZP_05735580.1| , CRISPR-associated protein, Csn1 family [Prevotella tannerae ATCC 51259]
gi|294674019|ref|YP_003574635.1| , Csn1 family CRISPR-associated protein [Prevotella ruminicola 23]
gi|347536497|ref|YP_004843922.1| , putative CRISPR-associated (Cas) protein [Flavobacterium branchiophilum FL-15]
gi|365959402|ref|YP_004940969.1| , putative CRISPR-associated (Cas) protein [Flavobacterium columnare ATCC 49512]
gi|218258638|ref|ZP_03474966.1| , hypothetical protein PRABACTJOHN_00621 [Parabacteroides johnsonii DSM 18315]
gi|256840409|ref|ZP_05545917.1| , CRISPR-associated protein [Parabacteroides sp. D13]
gi|224535832|ref|ZP_03676371.1| , hypothetical protein BACCELL_00696 [Bacteroides cellulosilyticus DSM 14838]
gi|60683893|ref|YP_213533.1| , hypothetical protein BF3954 [Bacteroides fragilis NCTC 9343]
gi|323344874|ref|ZP_08085098.1| , csn1 family CRISPR-associated protein [Prevotella oralis ATCC 33269]
gi|213962376|ref|ZP_03390639.1| , crispr-associated protein, Csn1 family [Capnocytophaga sputigena Capno]
gi|256819408|ref|YP_003140687.1| , CRISPR-associated protein, Csn1 family [Capnocytophaga ochracea DSM 7271]
gi|298373376|ref|ZP_06983365.1| , CRISPR-associated protein, Csn1 family [Bacteroidetes oral taxon 274 str. F0058]
gi|340622236|ref|YP_004740688.1| , hypothetical protein Ccan_14650 [Capnocytophaga canimorsus Cc5]
gi|163754820|ref|ZP_02161941.1| , hypothetical protein KAOT1_02362 [Kordia algicida OT-1]
gi|295136244|ref|YP_003586920.1| , hypothetical protein ZPR_4422 [Zunongwangia profunda SM-A87]
gi|150025575|ref|YP_001296401.1| , CRISPR-associated endonuclease Csn1 family protein [Flavobacterium psychrophilum JIP02/86]
gi|212694363|ref|ZP_03302491.1| , hypothetical protein BACDOR_03889 [Bacteroides dorei DSM 17855]
gi|301311869|ref|ZP_07217791.1| , conserved hypothetical protein [Bacteroides sp. 20_3]
gi|224026357|ref|ZP_03644723.1| , hypothetical protein BACCOPRO_03113 [Bacteroides coprophilus DSM 18228]
gi|261414553|ref|YP_003248236.1| , CRISPR-associated protein, Csn1 family [Fibrobacter succinogenes subsp. succinogenes S85]
gi|260592128|ref|ZP_05857586.1| , CRISPR-associated protein, Csn1 family [Prevotella veroralis F0319]
gi|288802595|ref|ZP_06408034.1| , conserved hypothetical protein [Prevotella melaninogenica D18]
gi|315607525|ref|ZP_07882520.1| , csn1 family CRISPR-associated protein [Prevotella buccae ATCC 33574]
gi|282858617|ref|ZP_06267779.1| , CRISPR-associated protein, Csn1 family [Prevotella bivia JCVIHMP010]
gi|300771242|ref|ZP_07081118.1| , csn1 family CRISPR-associated protein [Sphingobacterium spiritivorum ATCC 33861]

FIG. 20B

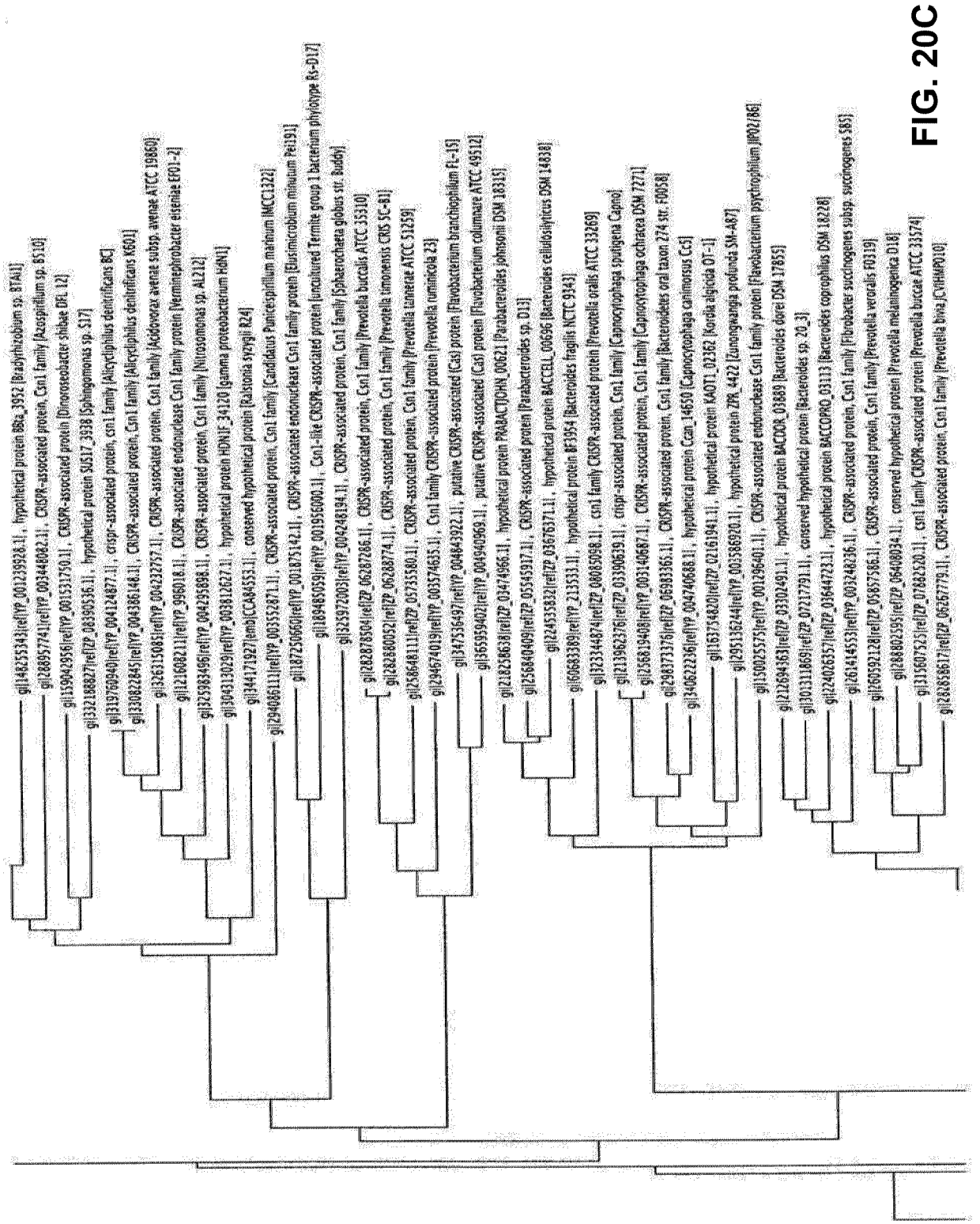

gi|148255343|ref|YP_001239928.1|, hypothetical protein BBta_3952 [Bradyrhizobium sp. BTAi1]
gi|288957741|ref|YP_003440082.1|, CRISPR-associated protein, Csn1 family [Azospirillum sp. B510]
gi|115904956|ref|YP_001531750.1|, CRISPR-associated protein [Dinoroseobacter shibae DFL 12]
gi|332188827|ref|ZP_08390536.1|, hypothetical protein SUS17_3938 [Sphingomonas sp. S17]
gi|319769940|ref|YP_004124877.1|, crispr-associated protein, csn1 family [Alicycliphilus denitrificans BC]
gi|330822845|ref|YP_004386148.1|, CRISPR-associated protein, Csn1 family [Alicycliphilus denitrificans K601]
gi|326315085|ref|YP_004232757.1|, CRISPR-associated protein, Csn1 family [Acidovorax avenae subsp. avenae ATCC 19860]
gi|121608211|ref|YP_998018.1|, CRISPR-associated endonuclease Csn1 family protein [Nermineophrobacter eiseniae EF01-2]
gi|325983496|ref|YP_004295898.1|, CRISPR-associated protein, Csn1 family [Nitrosomonas sp. AL212]
gi|304313029|ref|YP_003812627.1|, hypothetical protein HDNIF_34120 [gamma proteobacterium HdN1]
gi|344171927|emb|CCA84553.1|, conserved hypothetical protein [Ralstonia syzygii R24]
gi|294086111|ref|YP_003552871.1|, CRISPR-associated protein, Csn1 family [Candidatus Puniceispirillum marinum IMCC1322]
gi|187250660|ref|YP_001875142.1|, CRISPR-associated endonuclease Csn1 family protein [Businicrobium minutum Pei191]
gi|118948059|ref|YP_001956000.1|, Csn1-like CRISPR-associated protein [uncultured Termite group 1 bacterium phylotype Rs-D17]
gi|325972003|ref|YP_004248194.1|, CRISPR-associated protein, Csn1 family [Sphaerochaeta globus str. Buddy]
gi|262837504|ref|ZP_06287266.1|, CRISPR-associated protein, Csn1 family [Prevotella buccalis ATCC 35310]
gi|282860052|ref|ZP_06288774.1|, CRISPR-associated protein, Csn1 family [Prevotella timonensis CRIS 5C-B1]
gi|258648111|ref|ZP_05735580.1|, CRISPR-associated protein, Csn1 family [Prevotella tannerae ATCC 51259]
gi|294674019|ref|YP_003574635.1|, Csn1 family CRISPR-associated protein [Prevotella ruminicola 23]
gi|347536497|ref|YP_004843922.1|, putative CRISPR-associated (Cas) protein [Flavobacterium branchiophilum FL-15]
gi|356595402|ref|YP_004940969.1|, putative CRISPR-associated (Cas) protein [Flavobacterium columnare ATCC 49512]
gi|218258638|ref|ZP_03474966.1|, hypothetical protein PRABACTJOHN_00621 [Parabacteroides johnsonii DSM 18315]
gi|256849409|ref|ZP_05545917.1|, CRISPR-associated protein [Parabacteroides sp. D13]
gi|224535532|ref|ZP_03676371.1|, hypothetical protein BACCELL_00696 [Bacteroides cellulosilyticus DSM 14838]
gi|60683389|ref|YP_213533.1|, hypothetical protein BF3954 [Bacteroides fragilis NCTC 9343]
gi|323344874|ref|ZP_08085098.1|, csn1 family CRISPR-associated protein [Prevotella oralis ATCC 33269]
gi|211962376|ref|ZP_03390619.1|, crispr-associated protein, Csn1 family [Capnocytophaga sputigena Capno]
gi|256819408|ref|YP_003140687.1|, CRISPR-associated protein, Csn1 family [Capnocytophaga ochracea DSM 7271]
gi|298373376|ref|ZP_06983365.1|, CRISPR-associated protein, Csn1 family [Bacteroidetes oral taxon 274 str. F0058]
gi|340622396|ref|YP_004740688.1|, hypothetical protein Ccan_14650 [Capnocytophaga canimorsus Cc5]
gi|161754820|ref|ZP_02161941.1|, hypothetical protein KAOT1_02162 [Kordia algicida OT-1]
gi|295113624|ref|YP_003586920.1|, hypothetical protein ZPR_4422 [Zunongwangia profunda SM-A87]
gi|150025575|ref|YP_001296401.1|, CRISPR-associated endonuclease Csn1 family protein [Flavobacterium psychrophilum JIP02/86]
gi|212694363|ref|ZP_03302491.1|, hypothetical protein BACCOP_03883 [Bacteroides dorei DSM 17855]
gi|301311869|ref|ZP_07217791.1|, conserved hypothetical protein [Bacteroides sp. 20_3]
gi|224026157|ref|ZP_03644723.1|, hypothetical protein BACCOPRO_03113 [Bacteroides coprophilus DSM 18228]
gi|261414553|ref|YP_003248236.1|, CRISPR-associated protein, Csn1 family [Fibrobacter succinogenes subsp. succinogenes S85]
gi|260592128|ref|ZP_05857586.1|, CRISPR-associated protein, Csn1 family [Prevotella veroralis F0319]
gi|288802595|ref|ZP_06408034.1|, conserved hypothetical protein [Prevotella melaninogenica D18]
gi|315507525|ref|ZP_07862520.1|, csn1 family CRISPR-associated protein [Prevotella buccae ATCC 33574]
gi|282858617|ref|ZP_06267779.1|, CRISPR-associated protein, Csn1 family [Prevotella bivia JCVIHMP010]

FIG. 20C

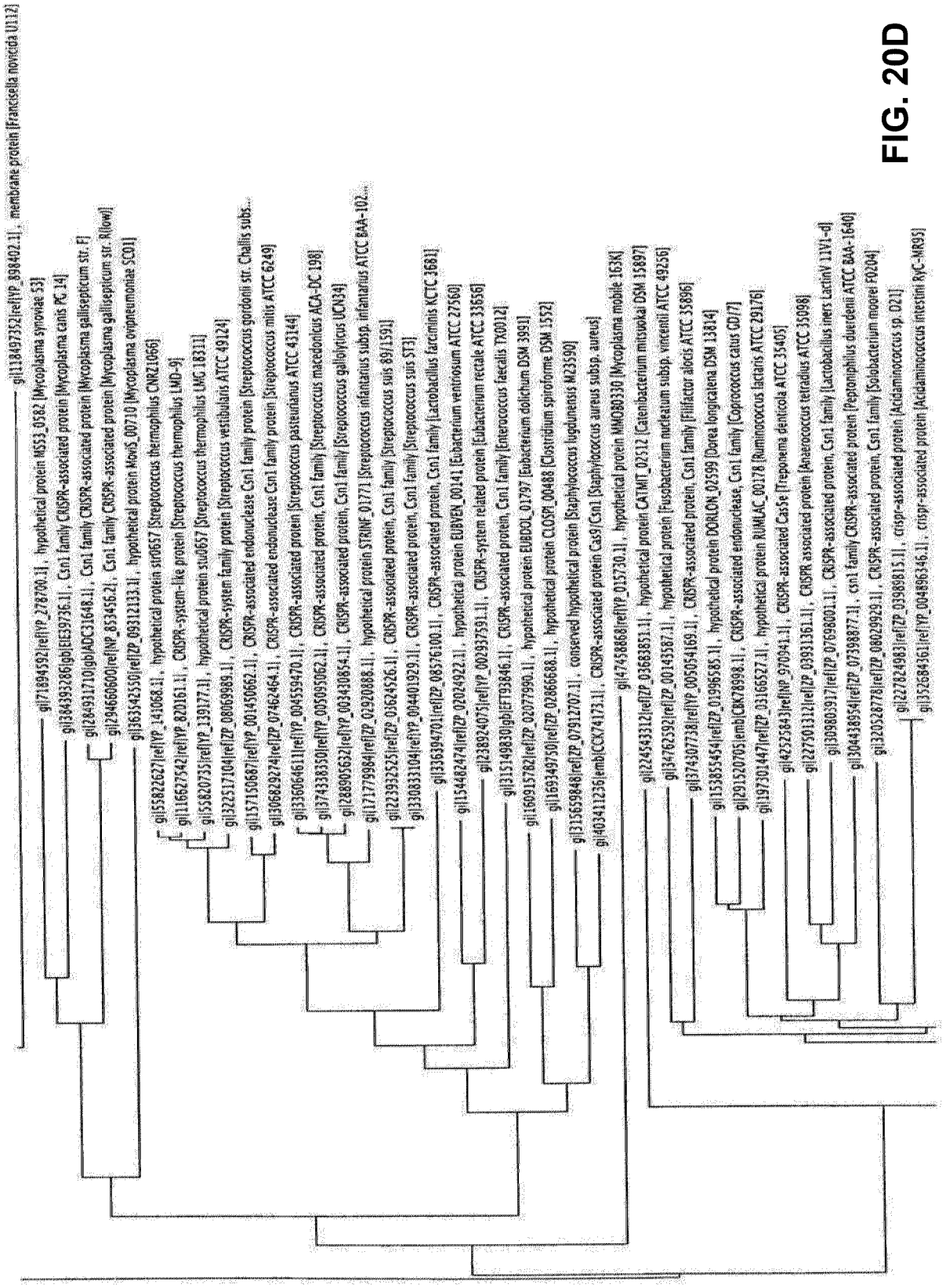

gi|118497352|ref|YP_898402.1|, membrane protein [Francisella novicida U112]
gi|71894592|ref|YP_278700.1|, hypothetical protein MSS3_0582 [Mycoplasma synoviae 53]
gi|384393286|gb|EIE39736.1|, Csn1 family CRISPR-associated protein [Mycoplasma canis PG 14]
gi|284931719|gb|ADC31648.1|, Csn1 family CRISPR-associated protein [Mycoplasma gallisepticum str. F]
gi|294668600|ref|NP_853456.2|, Csn1 family CRISPR-associated protein [Mycoplasma gallisepticum str. R(low)]
gi|363542550|ref|ZP_09312133.1|, hypothetical protein HMPREF9336_00710 [Mycoplasma ovipneumoniae SC01]
gi|15826627|ref|YP_141068.1|, hypothetical protein str0657 [Streptococcus thermophilus CNRZ1066]
gi|116627542|ref|YP_820161.1|, CRISPR-system-like protein [Streptococcus thermophilus LMD-9]
gi|15820735|ref|YP_139177.1|, hypothetical protein stu0657 [Streptococcus thermophilus LMG 18311]
gi|322517104|ref|ZP_08069988.1|, CRISPR-system family protein [Streptococcus vestibularis ATCC 49124]
gi|115750687|ref|YP_001450662.1|, CRISPR-associated endonuclease Csn1 family protein [Streptococcus gordonii str. Challis subs...
gi|306829274|ref|ZP_07462464.1|, CRISPR-associated endonuclease Csn1 family protein [Streptococcus mitis ATCC 6249]
gi|336064611|ref|YP_004559470.1|, CRISPR-associated protein [Streptococcus pasteurianus ATCC 43144]
gi|374338350|ref|YP_005095062.1|, CRISPR-associated protein, Csn1 family [Streptococcus macedonicus ACA-DC 198]
gi|288905632|ref|YP_003430854.1|, CRISPR-associated protein, Csn1 family [Streptococcus gallolyticus UCN34]
gi|171779984|ref|ZP_02920886.1|, hypothetical protein STRINF_01771 [Streptococcus infantarius subsp. infantarius ATCC BAA-102...
gi|223932525|ref|ZP_03624526.1|, CRISPR-associated protein, Csn1 family [Streptococcus suis 89/1591]
gi|330833104|ref|YP_004401929.1|, CRISPR-associated protein, Csn1 family [Streptococcus suis ST3]
gi|336394701|ref|ZP_08576100.1|, CRISPR-associated protein, Csn1 family [Lactobacillus farciminis KCTC 3681]
gi|154482474|ref|ZP_02024922.1|, hypothetical protein EUBVEN_00141 [Eubacterium ventriosum ATCC 27560]
gi|223892405|ref|YP_002937591.1|, CRISPR-system related protein [Eubacterium rectale ATCC 33656]
gi|315149830|gb|EFT93846.1|, CRISPR-associated protein, Csn1 family [Enterococcus faecalis TX0012]
gi|160915782|ref|ZP_02077990.1|, hypothetical protein EUBDOL_01797 [Eubacterium dolichum DSM 3991]
gi|169349750|ref|ZP_02866688.1|, hypothetical protein CLOSPI_00488 [Clostridium spiroforme DSM 1552]
gi|315659848|ref|ZP_07912707.1|, conserved hypothetical protein [Staphylococcus hyguenensis M235/90]
gi|403411236|emb|CCK74173.1|, CRISPR-associated protein Cas9/Csn1 [Staphylococcus aureus subsp. aureus]
gi|474588868|ref|YP_015730.1|, hypothetical protein MMOB0330 [Mycoplasma mobile 163K]
gi|224543312|ref|ZP_03683851.1|, hypothetical protein CATMIT_02512 [Catenibacterium mitsuokai DSM 15897]
gi|346762592|ref|ZP_00143587.1|, hypothetical protein [Fusobacterium nucleatum subsp. vincentii ATCC 49256]
gi|374307738|ref|YP_005054169.1|, CRISPR-associated protein, Csn1 family [Filifactor alocis ATCC 35896]
gi|153855454|ref|ZP_01996585.1|, hypothetical protein DORLON_02599 [Dorea longicatena DSM 13814]
gi|291520705|emb|CBK78998.1|, CRISPR-associated endonuclease, Csn1 family [Coprococcus catus GD/7]
gi|197301447|ref|ZP_03166527.1|, hypothetical protein RUMLAC_00178 [Ruminococcus lactaris ATCC 29176]
gi|142522643|ref|NP_970941.1|, CRISPR-associated Cas5e [Treponema denticola ATCC 35405]
gi|227501312|ref|ZP_03931361.1|, CRISPR associated protein [Anaerococcus tetradius ATCC 35098]
gi|309803917|ref|ZP_07698001.1|, CRISPR-associated protein, Csn1 family [Lactobacillus iners LactinV 11V1-d]
gi|304438954|ref|ZP_07398877.1|, csn1 family CRISPR-associated protein [Peptoniphilus duerdenii ATCC BAA-1640]
gi|320528778|ref|ZP_08029929.1|, CRISPR-associated protein, Csn1 family [Solobacterium moorei F0204]
gi|227824983|ref|ZP_03988815.1|, crispr-associated protein [Acidaminococcus sp. D21]
gi|352684361|ref|YP_004896346.1|, crispr-associated protein [Acidaminococcus intestini RyC-MR95]

FIG. 20D

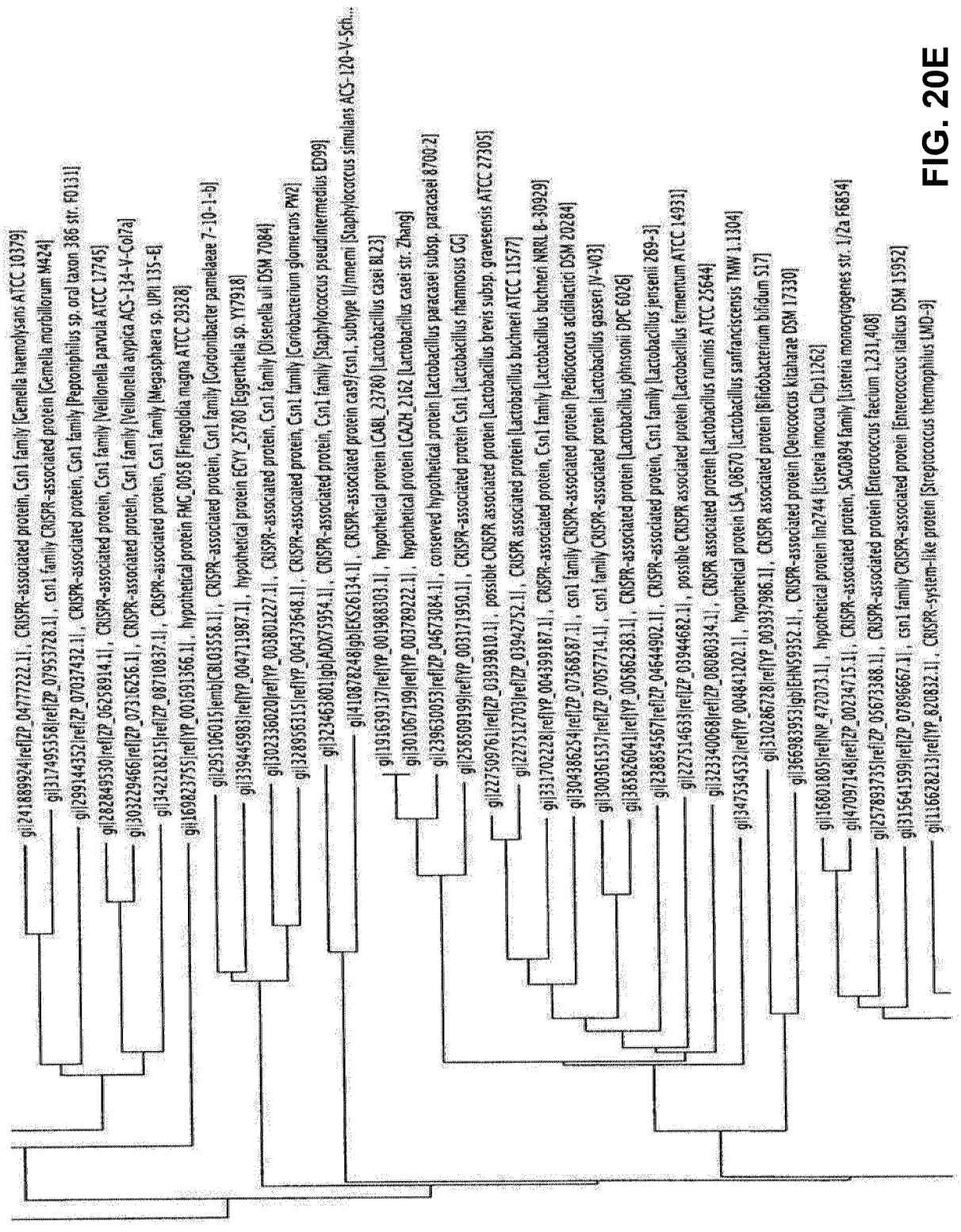

gi|241889924|ref|ZP_04777222.1| , CRISPR-associated protein, Csn1 family [Gemella haemolysans ATCC 10379]
gi|317495358|ref|ZP_07953728.1| , csn1 family CRISPR-associated protein [Gemella morbillorum M424]
gi|299144352|ref|ZP_07037432.1| , CRISPR-associated protein, Csn1 family [Peptoniphilus sp. oral taxon 386 str. F0131]
gi|282849530|ref|ZP_06258914.1| , CRISPR-associated protein, Csn1 family [Veillonella parvula ATCC 17745]
gi|303229466|ref|ZP_07316256.1| , CRISPR-associated protein, Csn1 family [Veillonella atypica ACS-134-V-Col7a]
gi|342218215|ref|ZP_08710837.1| , CRISPR-associated protein, Csn1 family [Megasphaera sp. UPII 135-E]
gi|160823755|ref|YP_001691366.1| , hypothetical protein PMC_0058 [Finegoldia magna ATCC 29328]
gi|295106015|emb|CBL03558.1| , CRISPR-associated protein, Csn1 family [Cordonibacter pamelaeae 7-10-1-b]
gi|339445983|ref|YP_004711987.1| , hypothetical protein EGYY_25780 [Eggerthella sp. YY7918]
gi|302336020|ref|YP_003801227.1| , CRISPR-associated protein, Csn1 family [Olsenella uli DSM 7084]
gi|328956315|ref|YP_004373648.1| , CRISPR-associated protein, Csn1 family [Coriobacterium glomerans PW2]
gi|323463801|gb|ADX75954.1| , CRISPR-associated protein, Csn1 family [Staphylococcus pseudintermedius ED99]
gi|410878248|gb|EKS26134.1| , CRISPR-associated protein cas9/csn1, subtype II/nmemi [Staphylococcus simulans ACS-120-V-Sch...
gi|191639137|ref|YP_001968303.1| , hypothetical protein LCABL_23780 [Lactobacillus casei BL23]
gi|301067199|ref|YP_003789222.1| , hypothetical protein LCAZH_2162 [Lactobacillus casei str. Zhang]
gi|239630053|ref|ZP_04673084.1| , conserved hypothetical protein [Lactobacillus paracasei subsp. paracasei 8700:2]
gi|258509199|ref|YP_003171950.1| , CRISPR-associated protein Csn1 [Lactobacillus rhamnosus GG]
gi|227509761|ref|ZP_03939810.1| , possible CRISPR associated protein [Lactobacillus brevis subsp. gravesensis ATCC 27305]
gi|227512703|ref|ZP_03942752.1| , CRISPR associated protein [Lactobacillus buchneri ATCC 11577]
gi|331702228|ref|YP_004399187.1| , CRISPR-associated protein, Csn1 family [Lactobacillus buchneri NRRL B-30929]
gi|304386254|ref|ZP_07368587.1| , csn1 family CRISPR-associated protein [Pediococcus acidilactici DSM 20284]
gi|300361537|ref|ZP_07057714.1| , csn1 family CRISPR-associated protein [Lactobacillus gasseri JV-V03]
gi|385826041|ref|YP_005862383.1| , CRISPR-associated protein [Lactobacillus johnsonii DPC 6026]
gi|238854567|ref|ZP_04644902.1| , CRISPR-associated protein, Csn1 family [Lactobacillus jensenii 269-3]
gi|227514633|ref|ZP_03944682.1| , possible CRISPR associated protein [Lactobacillus fermentum ATCC 14931]
gi|323340068|ref|ZP_08080334.1| , CRISPR associated protein [Lactobacillus ruminis ATCC 25644]
gi|347534532|ref|YP_004841202.1| , hypothetical protein LSA_08670 [Lactobacillus sanfranciscensis TMW 1.1304]
gi|310286728|ref|YP_003937986.1| , CRISPR associated protein [Bifidobacterium bifidum S17]
gi|366983953|gb|EHN59352.1| , CRISPR-associated protein [Pediococcus kitaharae DSM 17330]
gi|168801805|ref|NP_472073.1| , hypothetical protein lin2744 [Listeria innocua Clip11262]
gi|47097148|ref|ZP_00234715.1| , CRISPR-associated protein, SAG0894 family [Listeria monocytogenes str. 1/2a F6854]
gi|257893735|ref|ZP_05673388.1| , CRISPR-associated protein [Enterococcus faecium 1,231,408]
gi|315641599|ref|ZP_07896667.1| , csn1 family CRISPR-associated protein [Enterococcus italicus DSM 15952]
gi|116628213|ref|YP_820832.1| , CRISPR-system-like protein [Streptococcus thermophilus LMD-9]

FIG. 20E

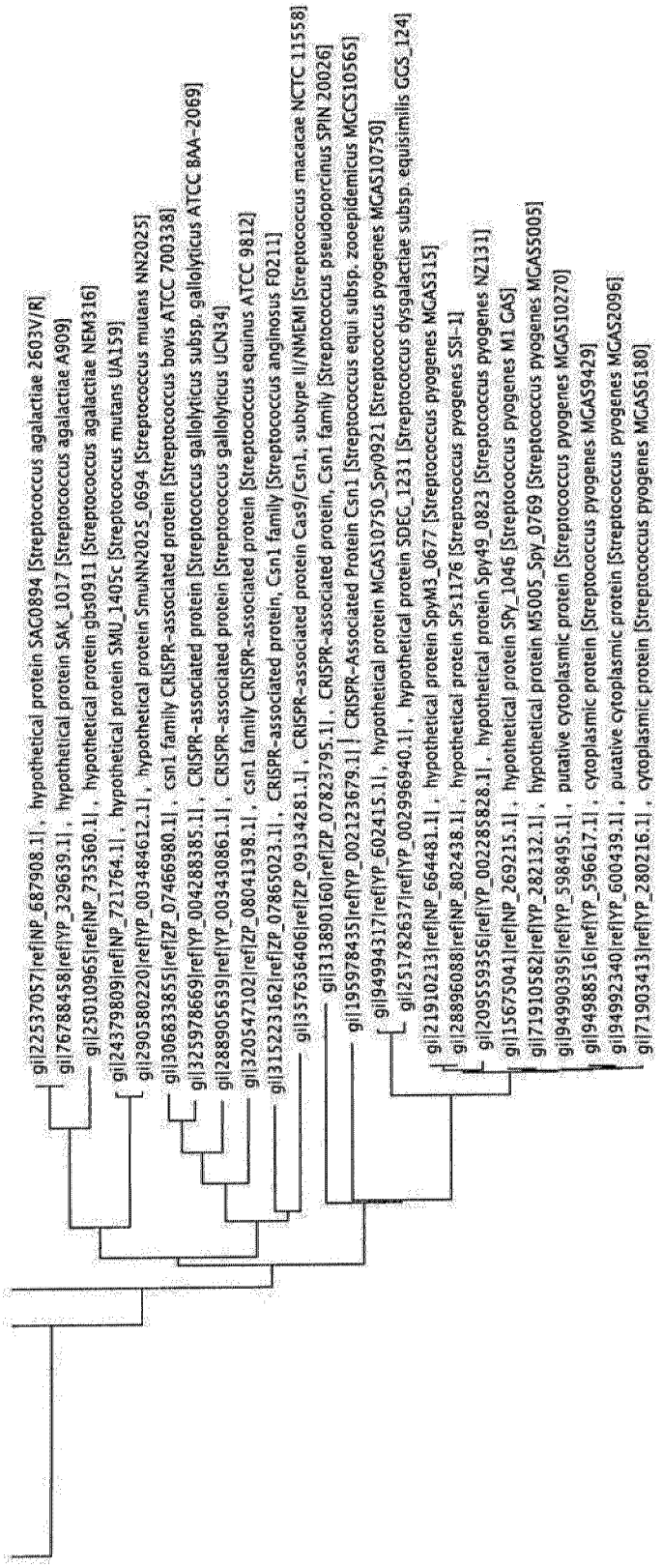

gi|22537057|ref|NP_687908.1| , hypothetical protein SAG0894 [Streptococcus agalactiae 2603V/R]
gi|76788458|ref|YP_329639.1| , hypothetical protein SAK_1017 [Streptococcus agalactiae A909]
gi|25010965|ref|NP_735360.1| , hypothetical protein gbs0911 [Streptococcus agalactiae NEM316]
gi|24379809|ref|NP_721764.1| , hypothetical protein SMU_1405c [Streptococcus mutans UA159]
gi|29053802 20|ref|YP_003484612.1| , hypothetical protein SmuNN2025_0694 [Streptococcus mutans NN2025]
gi|306833855|ref|ZP_07466980.1| , csn1 family CRISPR-associated protein [Streptococcus bovis ATCC 700338]
gi|325978669|ref|YP_004288385.1| , CRISPR-associated protein [Streptococcus gallolyticus subsp. gallolyticus ATCC BAA-2069]
gi|288905639|ref|YP_003430861.1| , CRISPR-associated protein [Streptococcus gallolyticus UCN34]
gi|320547102|ref|ZP_08041398.1| , csn1 family CRISPR-associated protein [Streptococcus equinus ATCC 9812]
gi|315223162|ref|ZP_07865023.1| , CRISPR-associated protein, Csn1 family [Streptococcus anginosus F0211]
gi|357636406|ref|ZP_09134281.1| , CRISPR-associated protein Cas9/Csn1, subtype II/NMEMI [Streptococcus macacae NCTC 11558]
gi|313890160|ref|ZP_07823795.1| , CRISPR-associated protein, Csn1 family [Streptococcus pseudoporcinus SPIN 20026]
gi|19597843 5|ref|YP_002123679.1| CRISPR-Associated Protein Csn1 [Streptococcus equi subsp. zooepidemicus MGCS10565]
gi|94994317|ref|YP_602415.1| , hypothetical protein MGAS10750_Spy0921 [Streptococcus pyogenes MGAS10750]
gi|251782637|ref|YP_002996940.1| , hypothetical protein SDEG_1231 [Streptococcus dysgalactiae subsp. equisimilis GGS_124]
gi|21910213|ref|NP_664481.1| , hypothetical protein SpyM3_0677 [Streptococcus pyogenes MGAS315]
gi|28896088|ref|NP_802438.1| , hypothetical protein SPs1176 [Streptococcus pyogenes SSI-1]
gi|209559356|ref|YP_002285828.1| , hypothetical protein Spy49_0823 [Streptococcus pyogenes NZ131]
gi|15675041|ref|NP_269215.1| , hypothetical protein SPy_1046 [Streptococcus pyogenes M1 GAS]
gi|71910582|ref|YP_282132.1| , hypothetical protein M5005_Spy_0769 [Streptococcus pyogenes MGAS5005]
gi|94990395|ref|YP_598495.1| , putative cytoplasmic protein [Streptococcus pyogenes MGAS10270]
gi|94988516|ref|YP_596617.1| , cytoplasmic protein [Streptococcus pyogenes MGAS9429]
gi|94992340|ref|YP_600439.1| , putative cytoplasmic protein [Streptococcus pyogenes MGAS2096]
gi|71903413|ref|YP_280216.1| , cytoplasmic protein [Streptococcus pyogenes MGAS6180]

FIG. 20F

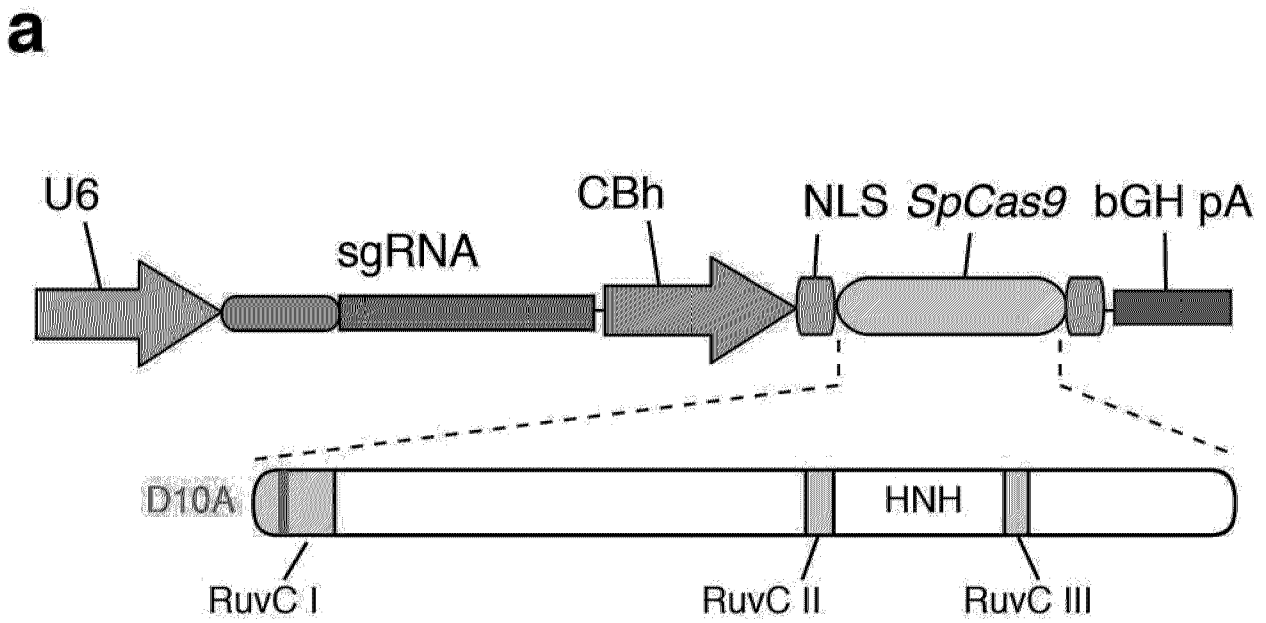

a
U6
sgRNA
CBh
NLS
SpCas9
bGH pA
D10A
HNH
RuvC I
RuvC II
RuvC III

FIG. 21A

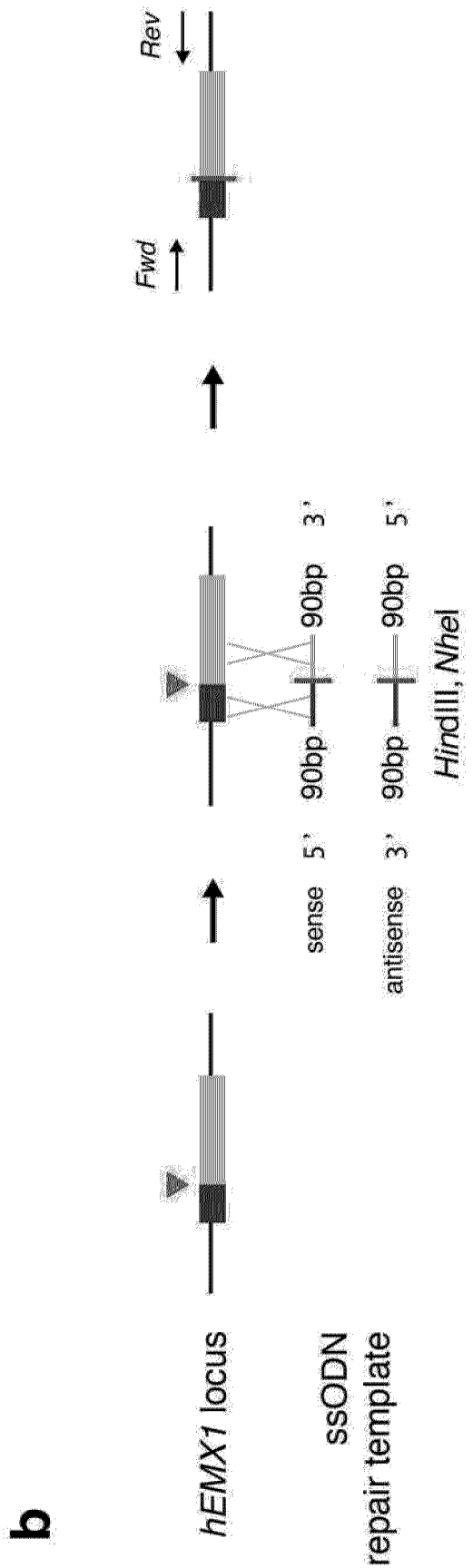

b
hEMX1 locus
ssODN
repair template
sense
5'
90bp
3'
antisense
3'
90bp
5'
HindIII, NheI
Fwd
Rev

FIG. 21B

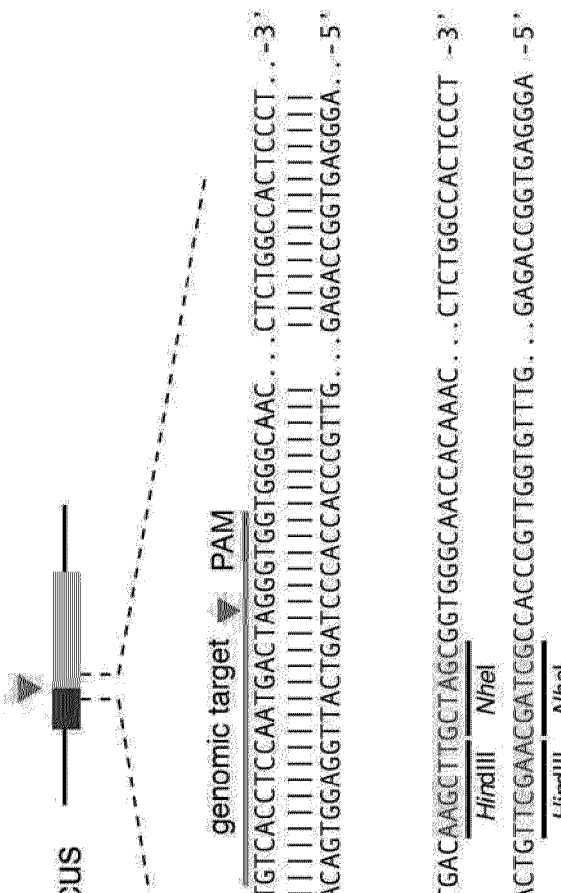

C
hEMX1 locus
genomic target
PAM
5'-...CAGAAGAAGAAGGGC...CCAATGGGGAGGACATCGATGTCACCTCCAATGACTAGGGTGGTGGGCAAC...CTCTGGCCACTCCCT...-3'
3'-...GTCTTCTTCTTCCCG...GGTTACCCCTCCTGTAGCTACAGTGGAGGTTACTGATCCCACCACCCGTTG...GAGACCGGTGAGGGA...-5'
sense 5'- CAGAAGAAGAAGGGC...ACATCGATGTCACCTCCAATGACAAGCTTGCTAGCGGTGGGCAACCACAAAC...CTCTGGCCACTCCCT -3'
HindIII NheI
antisense 3'- GTCTTCTTCTTCCCG...TGTAGCTACAGTGGAGGTTACTGTTCGAACGATCGCCACCCGTTGGTGTTTG...GAGACCGGTGAGGGA -5'
HindIII NheI

FIG. 21C

d

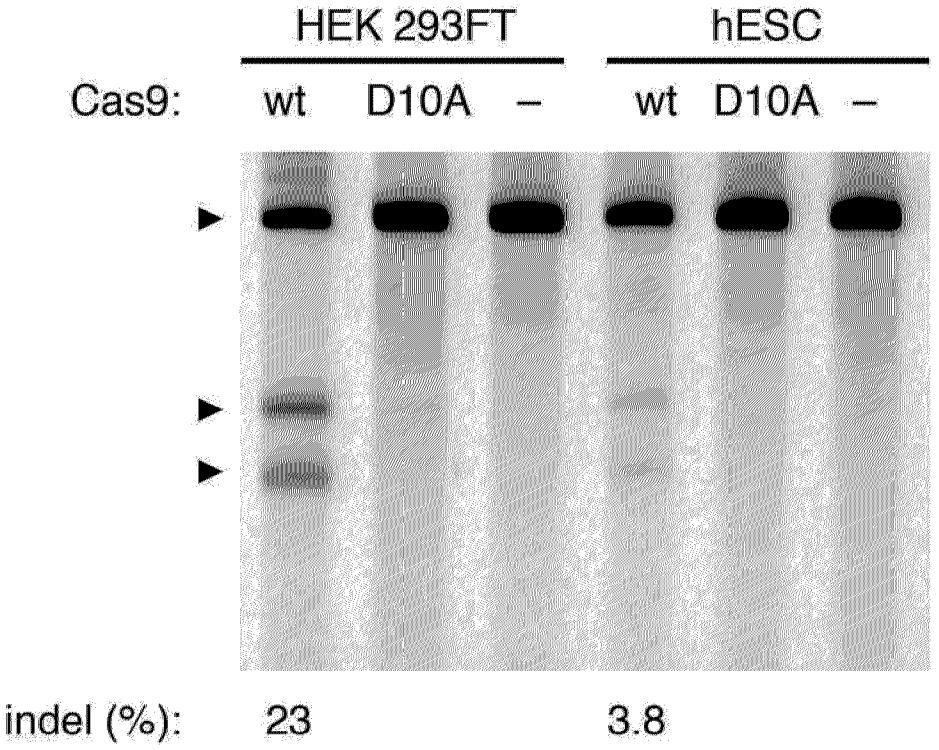


FIG. 21D

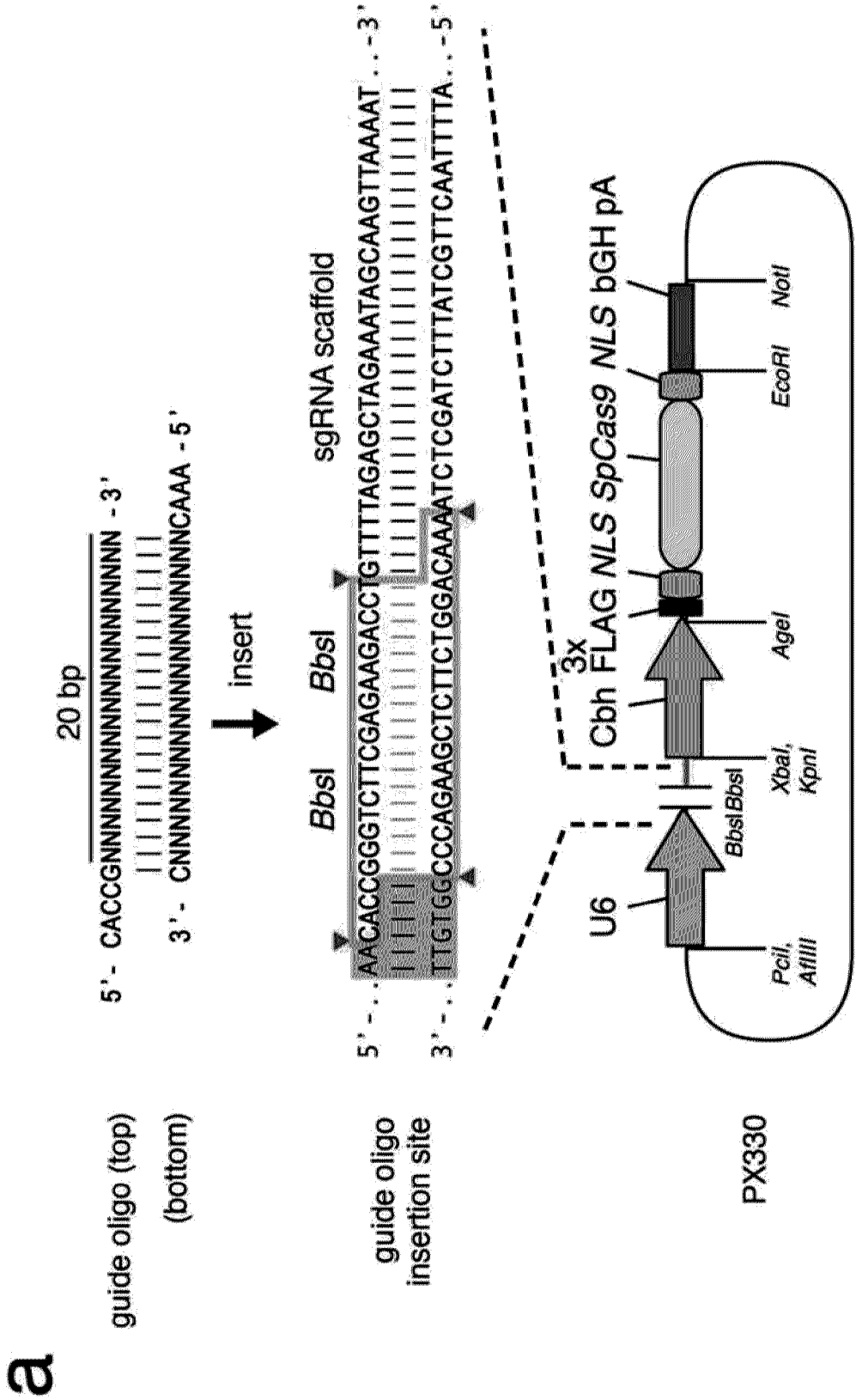

a
guide oligo (top)
(bottom)
20 bp
5'- CACCGNNNNNNNNNNNNNNNNNNN - 3'
3'- CNNNNNNNNNNNNNNNNNNNNCAAA - 5'
insert
guide oligo insertion site
BbsI
BbsI
sgRNA scaffold
5'- . . AACACCGGGTCTTCGAGAAGACCTGTTTTAGAGCTAGAAATAGCAAGTTAAAAT . . -3'
3'- . . TTGTGGCCCAGAAGCTCTTCTGGACAAAATCTCGATCTTTATCGTTCAATTTTA . . -5'
PX330
U6
3x
Cbh FLAG NLS SpCas9 NLS bGH pA
PciI, AflIII
BbsI BbsI
XbaI, KpnI
AgeI
EcoRI
NotI

FIG. 22A

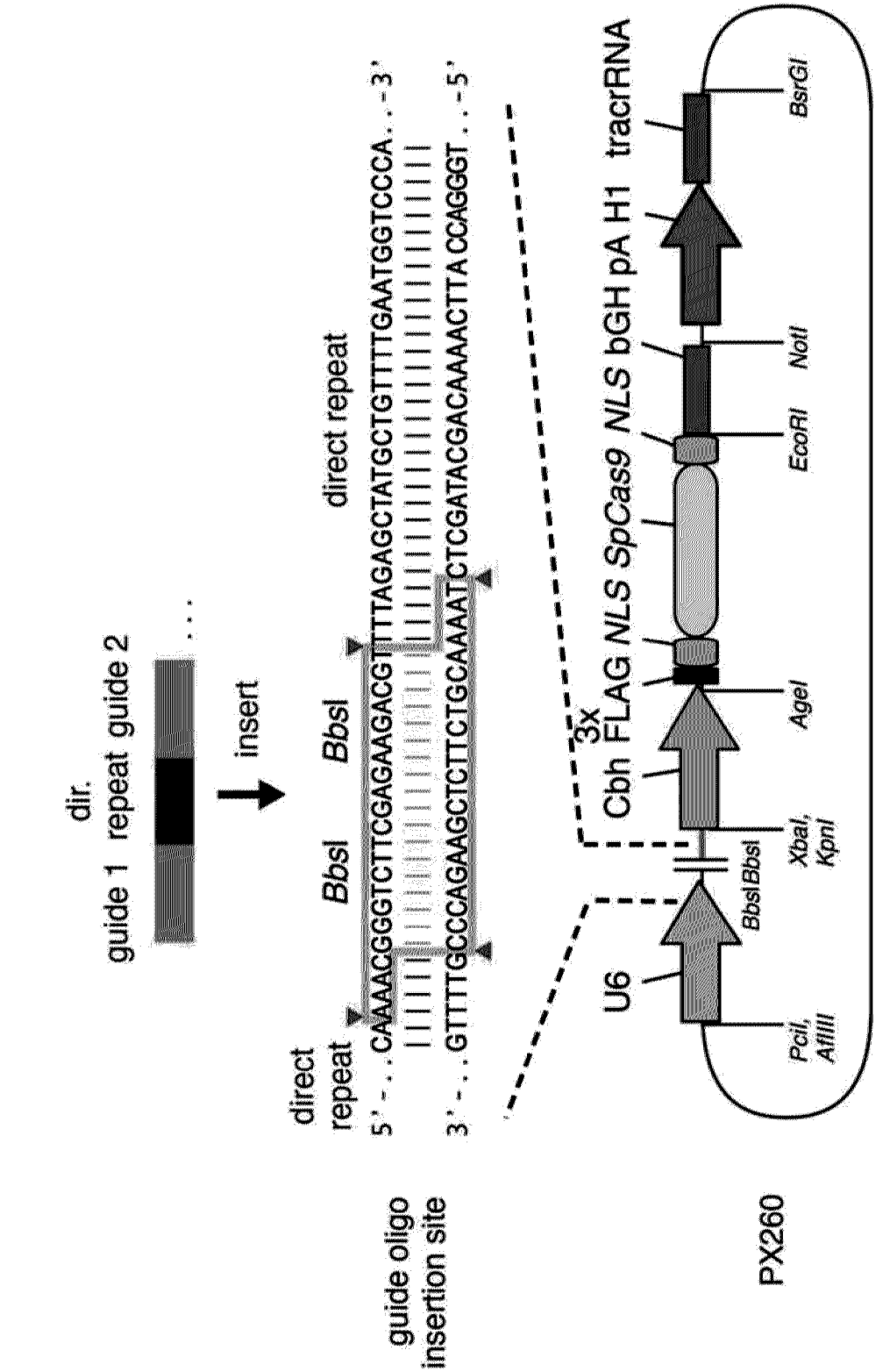

b
dir.
guide 1 repeat guide 2
insert
BbsI BbsI
direct repeat
direct repeat
5' - . . CAAACGGGTCTTCGAGAAGACGTTTTAGAGCTATGCTGTTTTGAATGGTCCCA. . - 3'
3' - . . GTTTTGCCCAGAAGCTCTTCTGCAAAATCTCGATACGACAAAACTTACCAGGGT . . - 5'
guide oligo insertion site
U6
3x
Cbh FLAG NLS SpCas9 NLS bGH pA H1 tracrRNA
PciI, AflIII
BbsI BbsI
XbaI, KpnI
AgeI
EcoRI
NotI
BsrGI
PX260

FIG. 22B

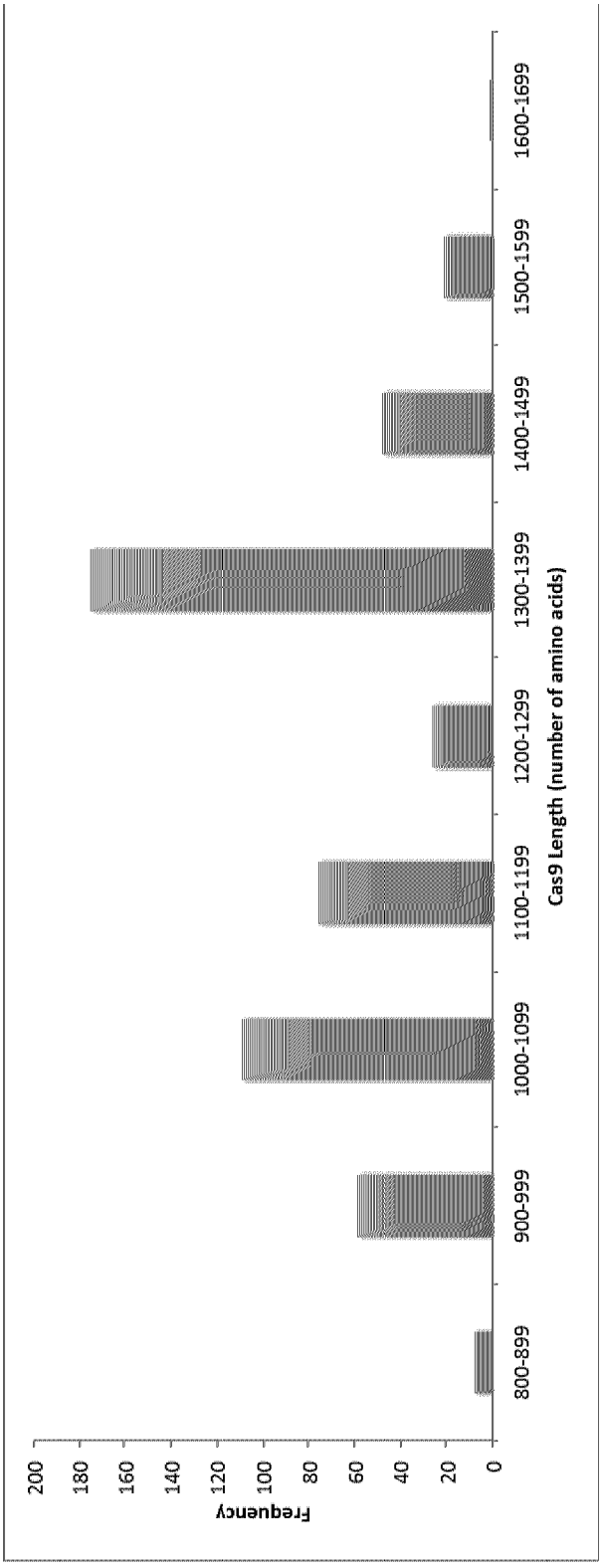

Frequency
0
20
40
60
80
100
120
140
160
180
200
800-899
900-999
1000-1099
1100-1199
1200-1299
1300-1399
1400-1499
1500-1599
1600-1699
Cas9 Length (number of amino acids)

FIG. 23

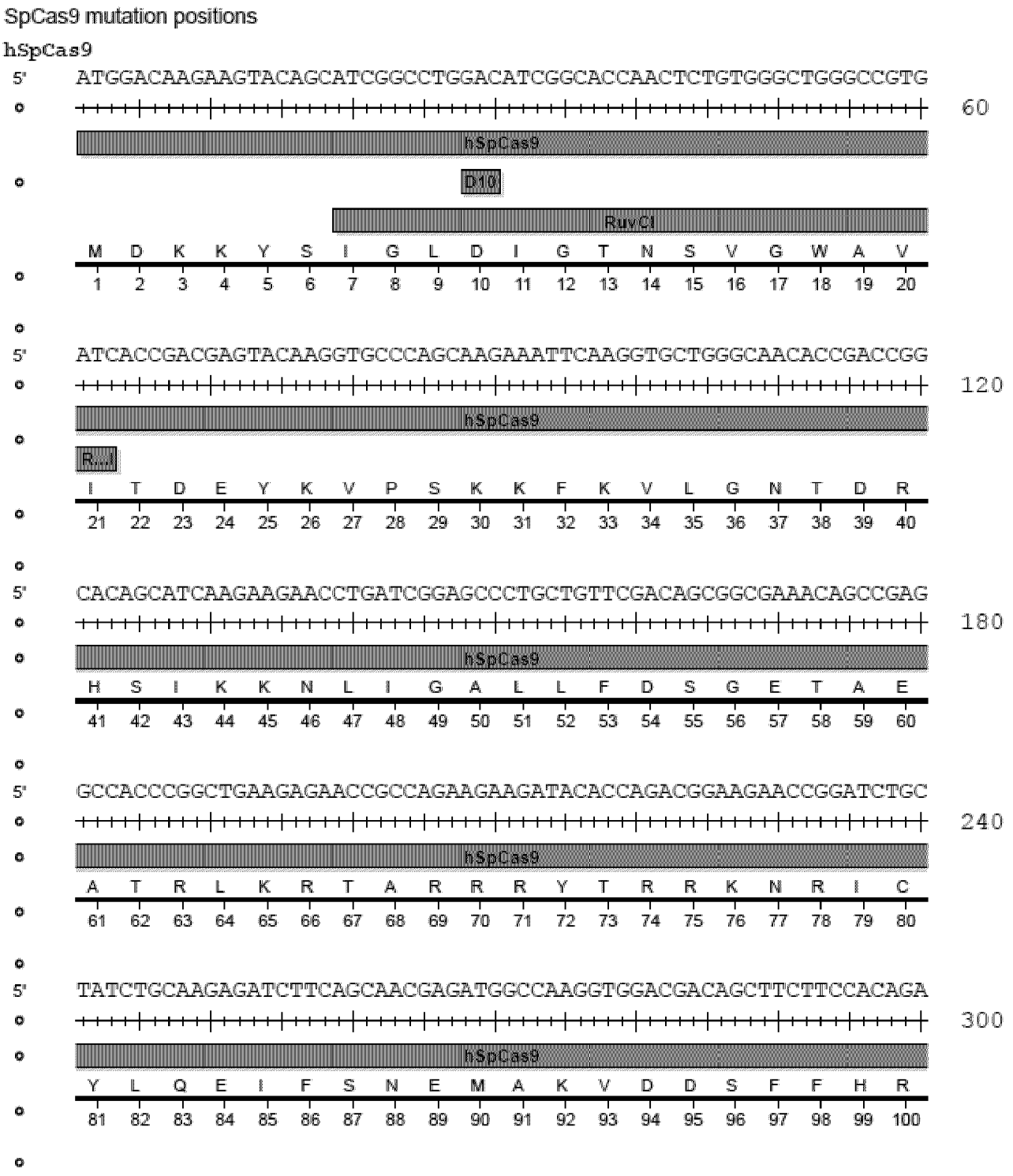

SpCas9 mutation positions
hSpCas9
5' ATGGACAAGAAGTACAGCATCGGCCTGGACATCGGCACCAACTCTGTGGGCTGGGCCGTG
60
hSpCas9
D10
RuvCI
M D K K Y S I G L D I G T N S V G W A V
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
5' ATCACCGACGAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACCGG
120
hSpCas9
R...I
I T D E Y K V P S K K F K V L G N T D R
21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40
5' CACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTCGACAGCGGCGAAACAGCCGAG
180
hSpCas9
H S I K K N L I G A L L F D S G E T A E
41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60
5' GCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACCAGACGGAAGAACCGGATCTGC
240
hSpCas9
A T R L K R T A R R R Y T R R K N R I C
61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80
5' TATCTGCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGA
300
hSpCas9
Y L Q E I F S N E M A K V D D S F F H R
81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100

FIG. 24A

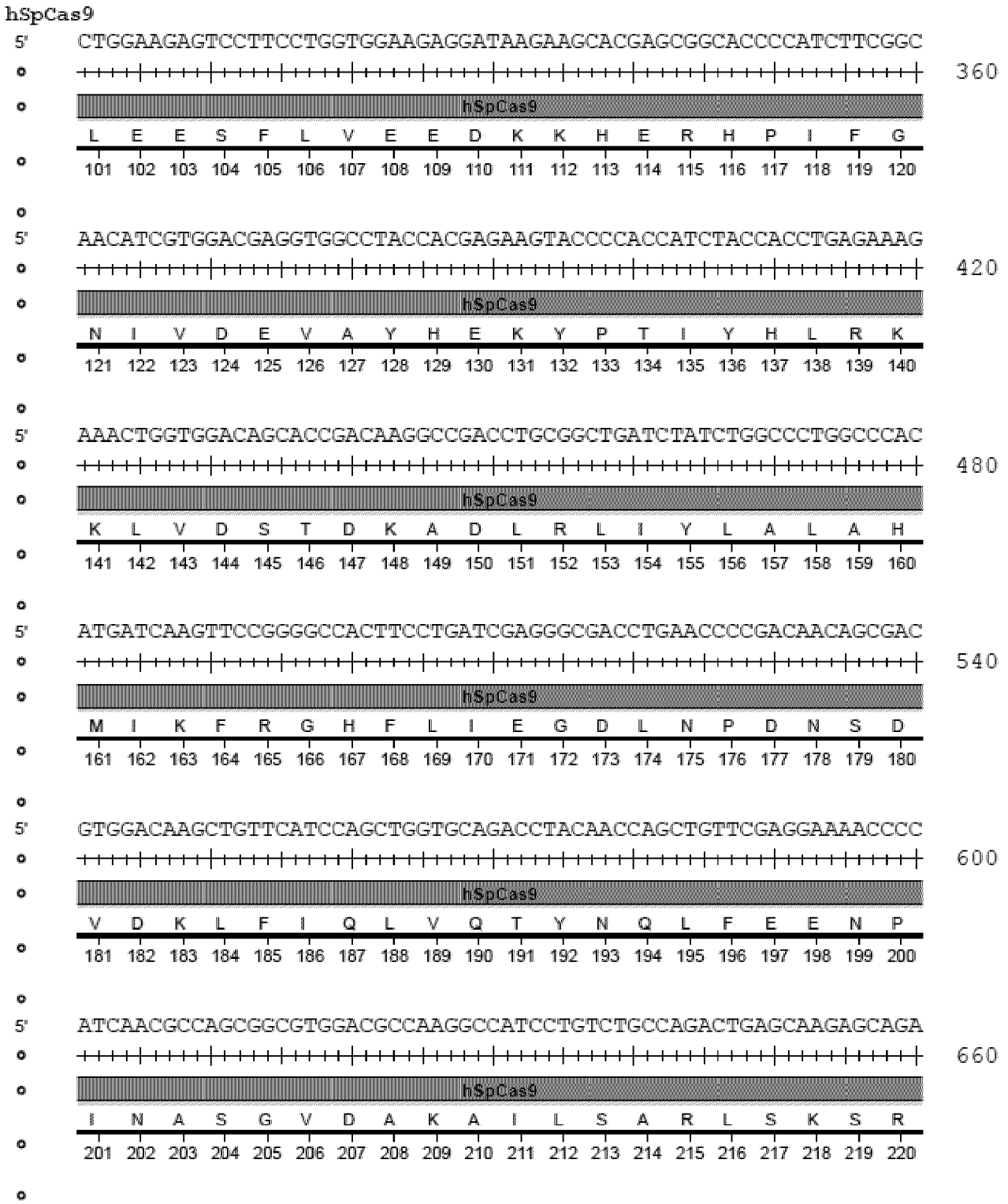

hSpCas9
5' CTGGAAGAGTCCTTCCTGGTGGAAGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGC
360
hSpCas9
L E E S F L V E E D K K H E R H P I F G
101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120
5' AACATCGTGGACGAGGTGGCCTACCACGAGAAGTACCCCACCATCTACCACCTGAGAAAG
420
hSpCas9
N I V D E V A Y H E K Y P T I Y H L R K
121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140
5' AAACTGGTGGACAGCACCGACAAGGCCGACCTGCGGCTGATCTATCTGGCCCTGGCCCAC
480
hSpCas9
K L V D S T D K A D L R L I Y L A L A H
141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160
5' ATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGACAACAGCGAC
540
hSpCas9
M I K F R G H F L I E G D L N P D N S D
161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180
5' GTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAAAACCCC
600
hSpCas9
V D K L F I Q L V Q T Y N Q L F E E N P
181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200
5' ATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGA
660
hSpCas9
I N A S G V D A K A I L S A R L S K S R
201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220

FIG. 24B

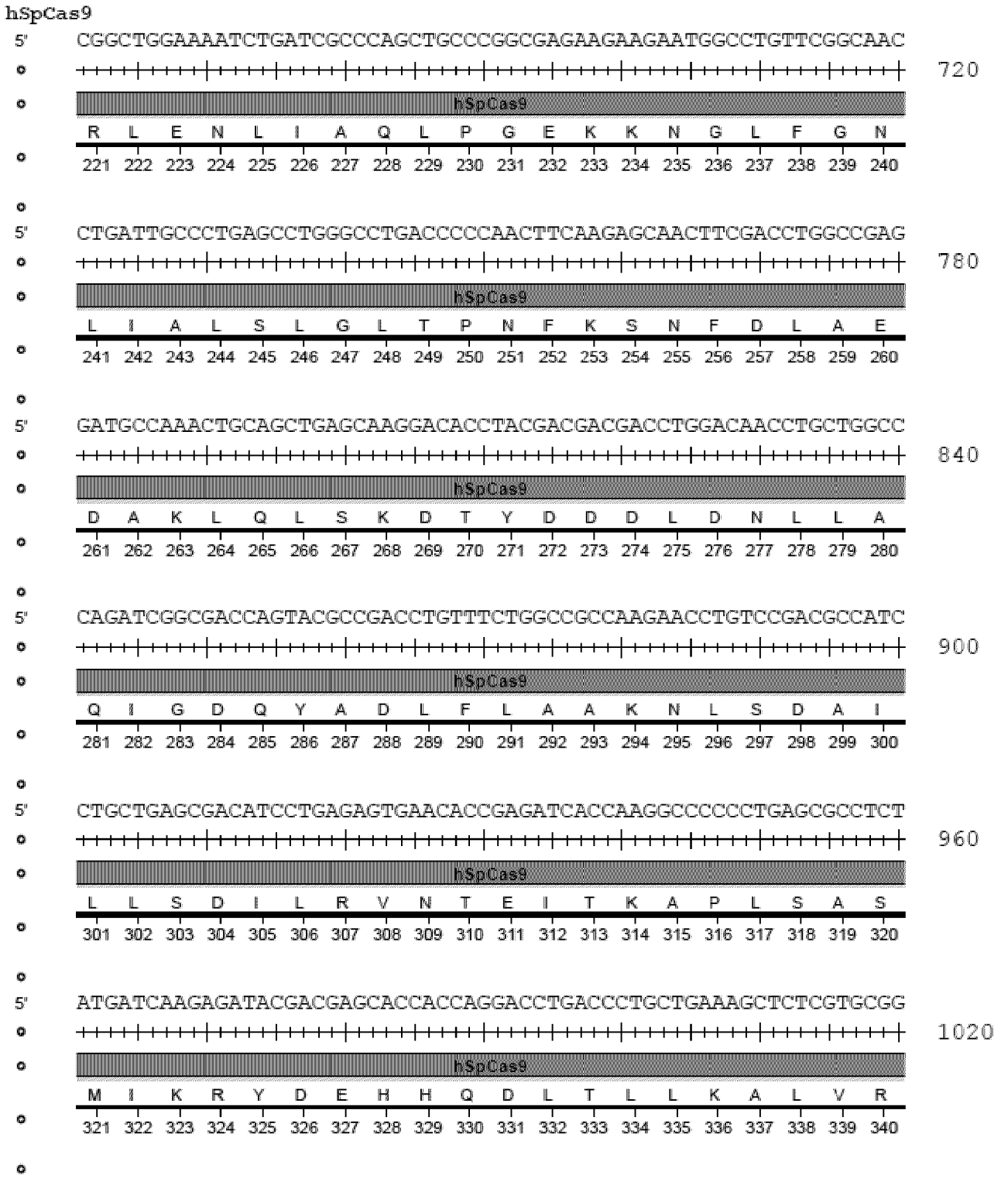

hSpCas9
5' CGGCTGGAAAATCTGATCGCCCAGCTGCCCGGCGAGAAGAAGAATGGCCTGTTCGGCAAC
720
hSpCas9
R L E N L I A Q L P G E K K N G L F G N
221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240
5' CTGATTGCCCTGAGCCTGGGCCTGACCCCCAACTTCAAGAGCAACTTCGACCTGGCCGAG
780
hSpCas9
L I A L S L G L T P N F K S N F D L A E
241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260
5' GATGCCAAACTGCAGCTGAGCAAGGACACCTACGACGACGACCTGGACAACCTGCTGGCC
840
hSpCas9
D A K L Q L S K D T Y D D D L D N L L A
261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280
5' CAGATCGGCGACCAGTACGCCGACCTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCATC
900
hSpCas9
Q I G D Q Y A D L F L A A K N L S D A I
281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300
5' CTGCTGAGCGACATCCTGAGAGTGAACACCGAGATCACCAAGGCCCCCCTGAGCGCCTCT
960
hSpCas9
L L S D I L R V N T E I T K A P L S A S
301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320
5' ATGATCAAGAGATACGACGAGCACCACCAGGACCTGACCCTGCTGAAAGCTCTCGTGCGG
1020
hSpCas9
M I K R Y D E H H Q D L T L L K A L V R
321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340

FIG. 24C

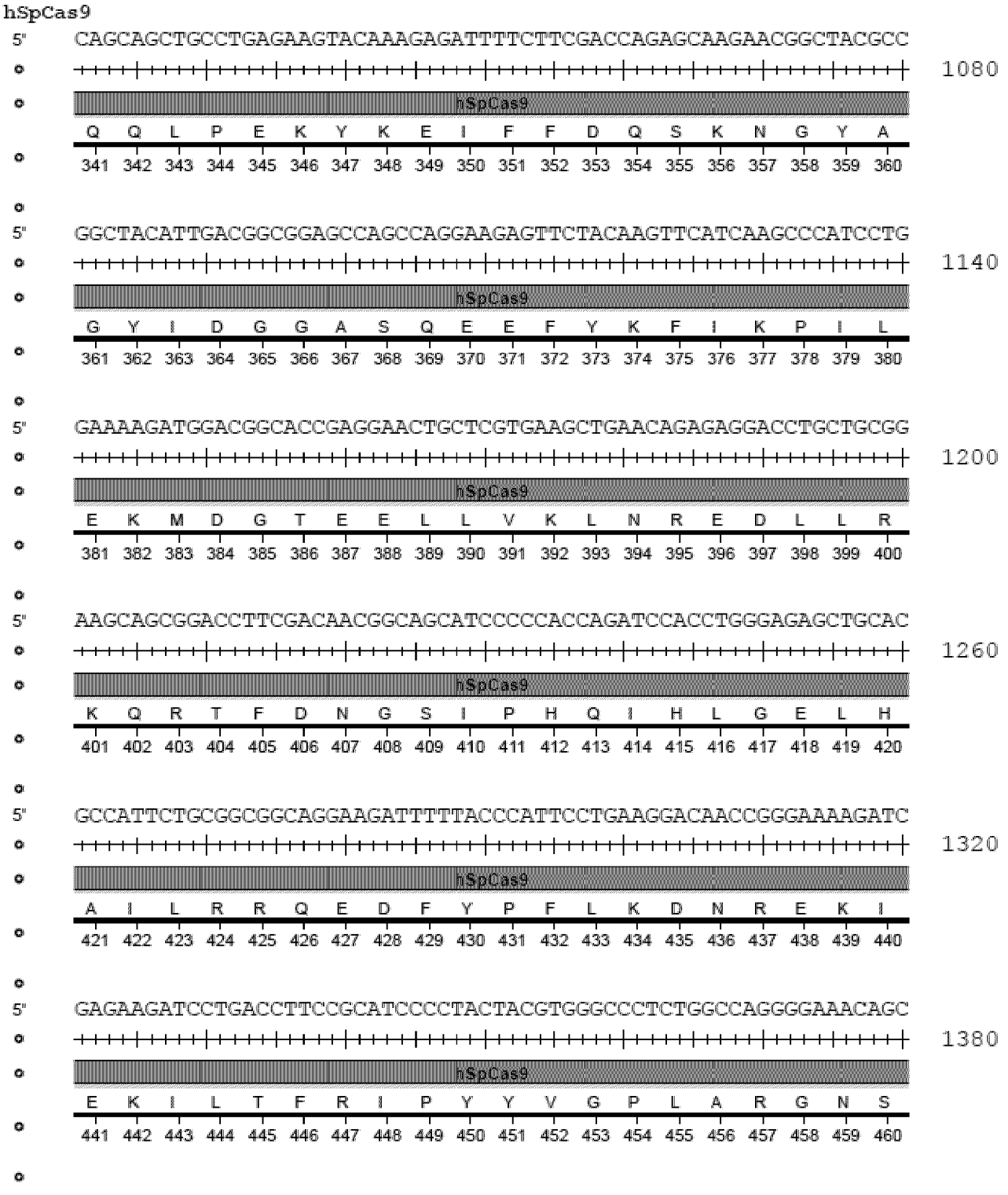

hSpCas9
5' CAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCC
1080
hSpCas9
Q Q L P E K Y K E I F F D Q S K N G Y A
341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360
5' GGCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATCCTG
1140
hSpCas9
G Y I D G G A S Q E E F Y K F I K P I L
361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380
5' GAAAAGATGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGG
1200
hSpCas9
E K M D G T E E L L V K L N R E D L L R
381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400
5' AAGCAGCGGACCTTCGACAACGGCAGCATCCCCCACCAGATCCACCTGGGAGAGCTGCAC
1260
hSpCas9
K Q R T F D N G S I P H Q I H L G E L H
401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420
5' GCCATTCTGCGGCGGCAGGAAGATTTTTACCCATTCCTGAAGGACAACCGGGAAAAGATC
1320
hSpCas9
A I L R R Q E D F Y P F L K D N R E K I
421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440
5' GAGAAGATCCTGACCTTCCGCATCCCCTACTACGTGGGCCCTCTGGCCAGGGGAAACAGC
1380
hSpCas9
E K I L T F R I P Y Y V G P L A R G N S
441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460

FIG. 24D

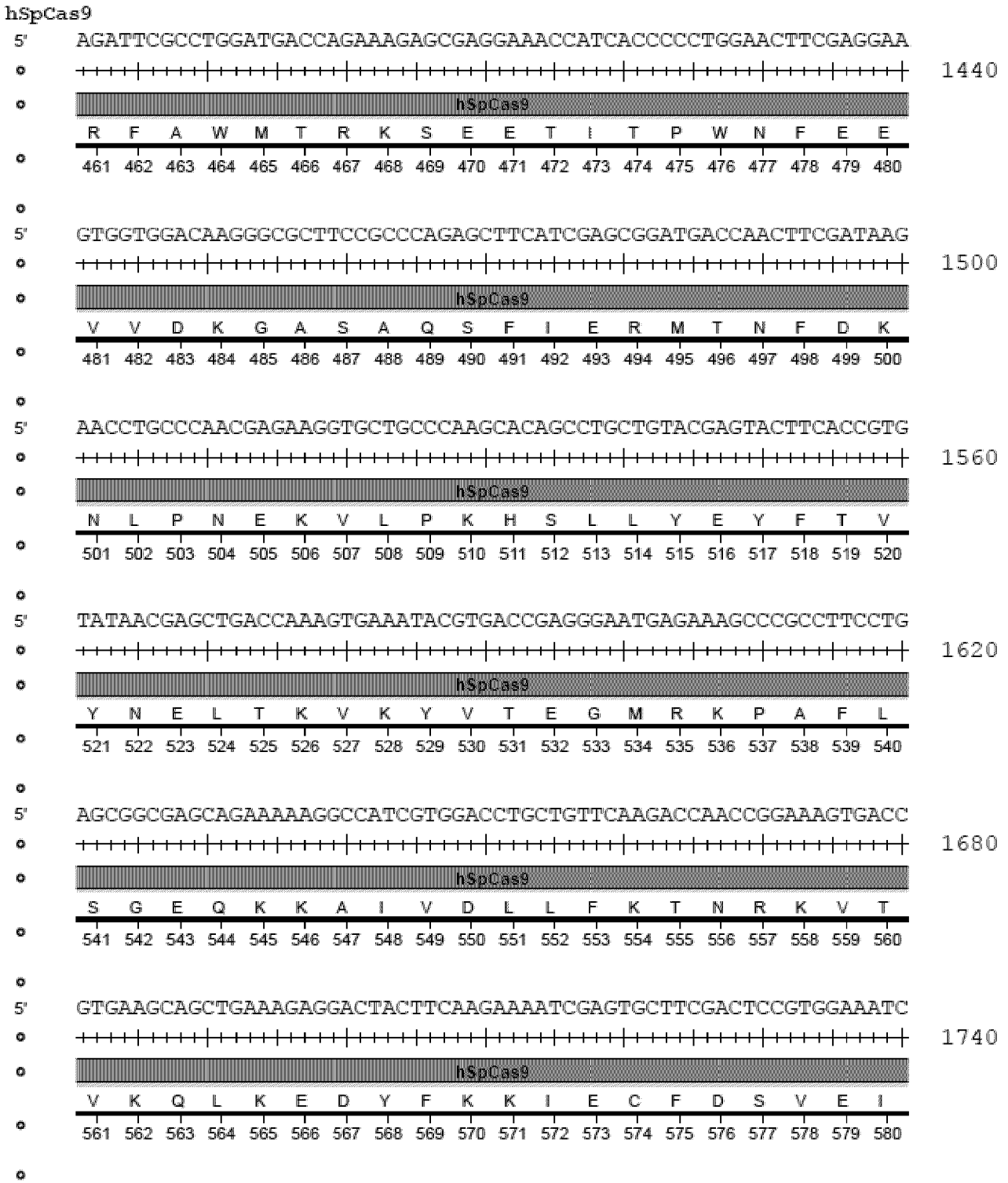

hSpCas9
5' AGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATCACCCCCTGGAACTTCGAGGAA
1440
hSpCas9
R F A W M T R K S E E T I T P W N F E E
461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480
5' GTGGTGGACAAGGGCGCTTCCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAG
1500
hSpCas9
V V D K G A S A Q S F I E R M T N F D K
481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500
5' AACCTGCCCAACGAGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTG
1560
hSpCas9
N L P N E K V L P K H S L L Y E Y F T V
501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520
5' TATAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTCCTG
1620
hSpCas9
Y N E L T K V K Y V T E G M R K P A F L
521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540
5' AGCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACC
1680
hSpCas9
S G E Q K K A I V D L L F K T N R K V T
541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560
5' GTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACTCCGTGGAAATC
1740
hSpCas9
V K Q L K E D Y F K K I E C F D S V E I
561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580

FIG. 24E

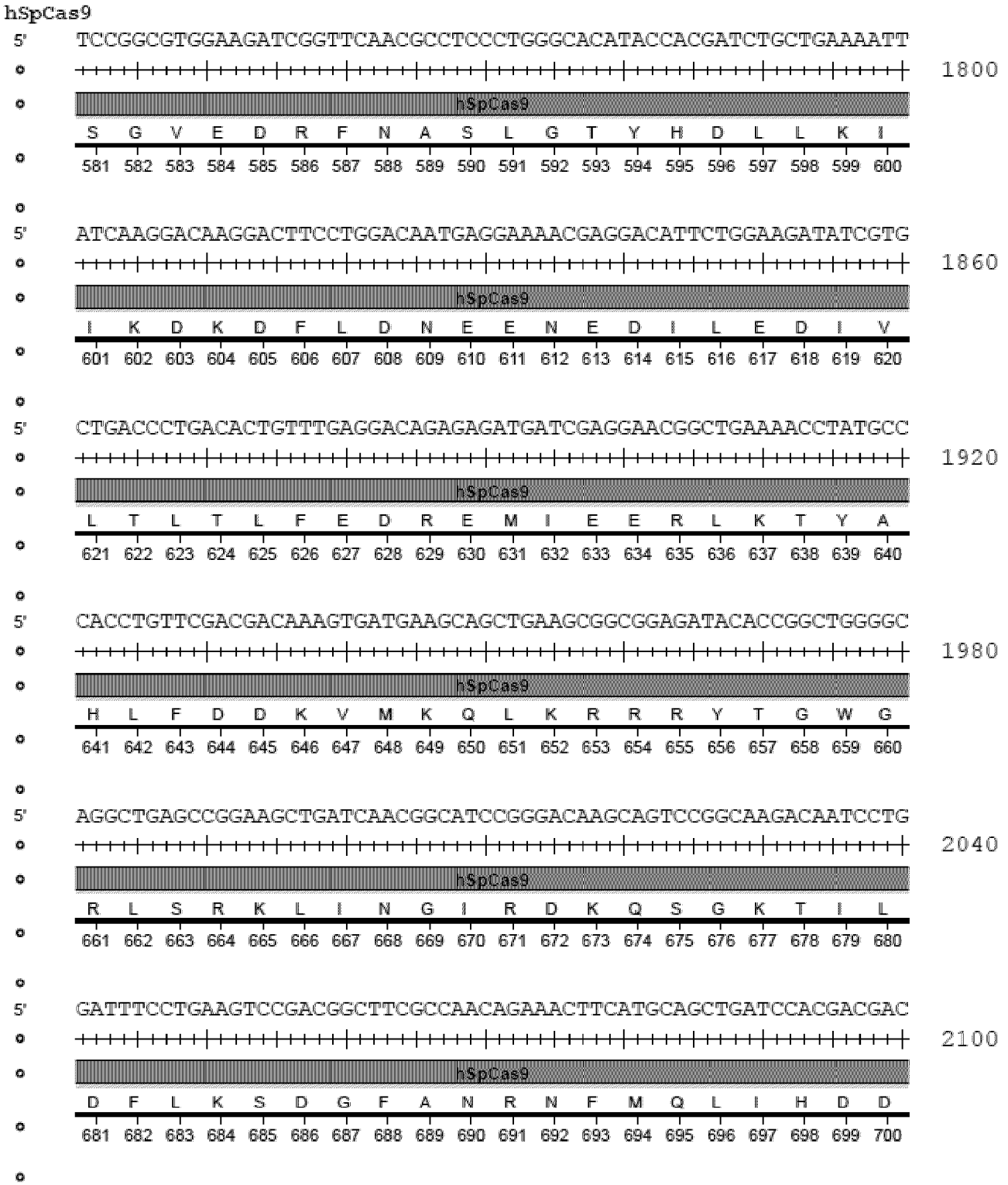

hSpCas9
5' TCCGGCGTGGAAGATCGGTTCAACGCCTCCCTGGGCACATACCACGATCTGCTGAAAATT
1800
hSpCas9
S G V E D R F N A S L G T Y H D L L K I
581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600
5' ATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACGAGGACATTCTGGAAGATATCGTG
1860
hSpCas9
I K D K D F L D N E E N E D I L E D I V
601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620
5' CTGACCCTGACACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCTATGCC
1920
hSpCas9
L T L T L F E D R E M I E E R L K T Y A
621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640
5' CACCTGTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGC
1980
hSpCas9
H L F D D K V M K Q L K R R R Y T G W G
641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660
5' AGGCTGAGCCGGAAGCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTG
2040
hSpCas9
R L S R K L I N G I R D K Q S G K T I L
661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680
5' GATTTCCTGAAGTCCGACGGCTTCGCCAACAGAAACTTCATGCAGCTGATCCACGACGAC
2100
hSpCas9
D F L K S D G F A N R N F M Q L I H D D
681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700

FIG. 24F

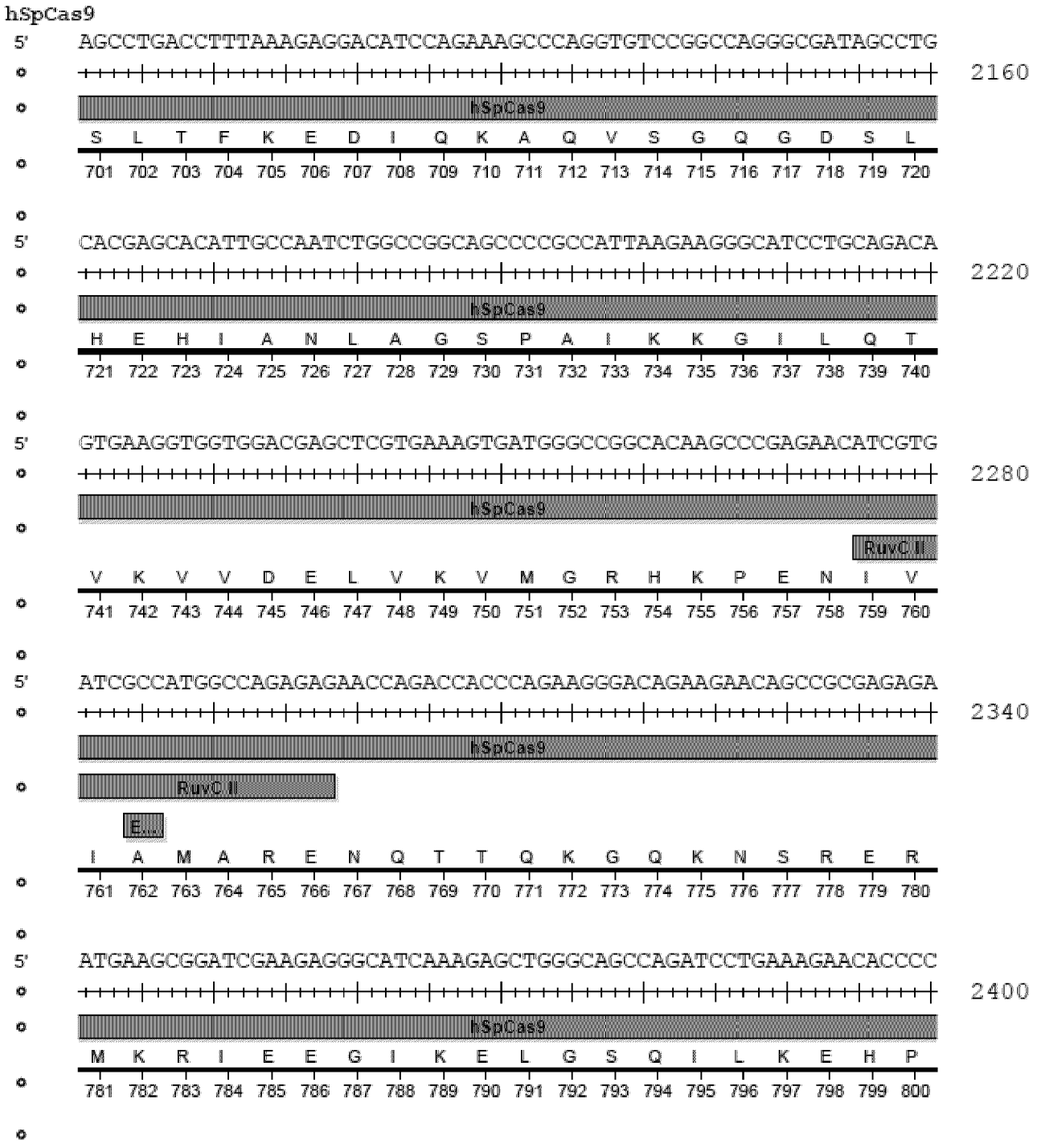

hSpCas9
5' AGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGGCGATAGCCTG
2160
hSpCas9
S L T F K E D I Q K A Q V S G Q G D S L
701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720
5' CACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACA
2220
hSpCas9
H E H I A N L A G S P A I K K G I L Q T
721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740
5' GTGAAGGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGAGAACATCGTG
2280
hSpCas9
RuvC II
V K V V D E L V K V M G R H K P E N I V
741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760
5' ATCGCCATGGCCAGAGAGAACCAGACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGA
2340
hSpCas9
RuvC II
E...
I A M A R E N Q T T Q K G Q K N S R E R
761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780
5' ATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAACACCCC
2400
hSpCas9
M K R I E E G I K E L G S Q I L K E H P
781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800

FIG. 24G

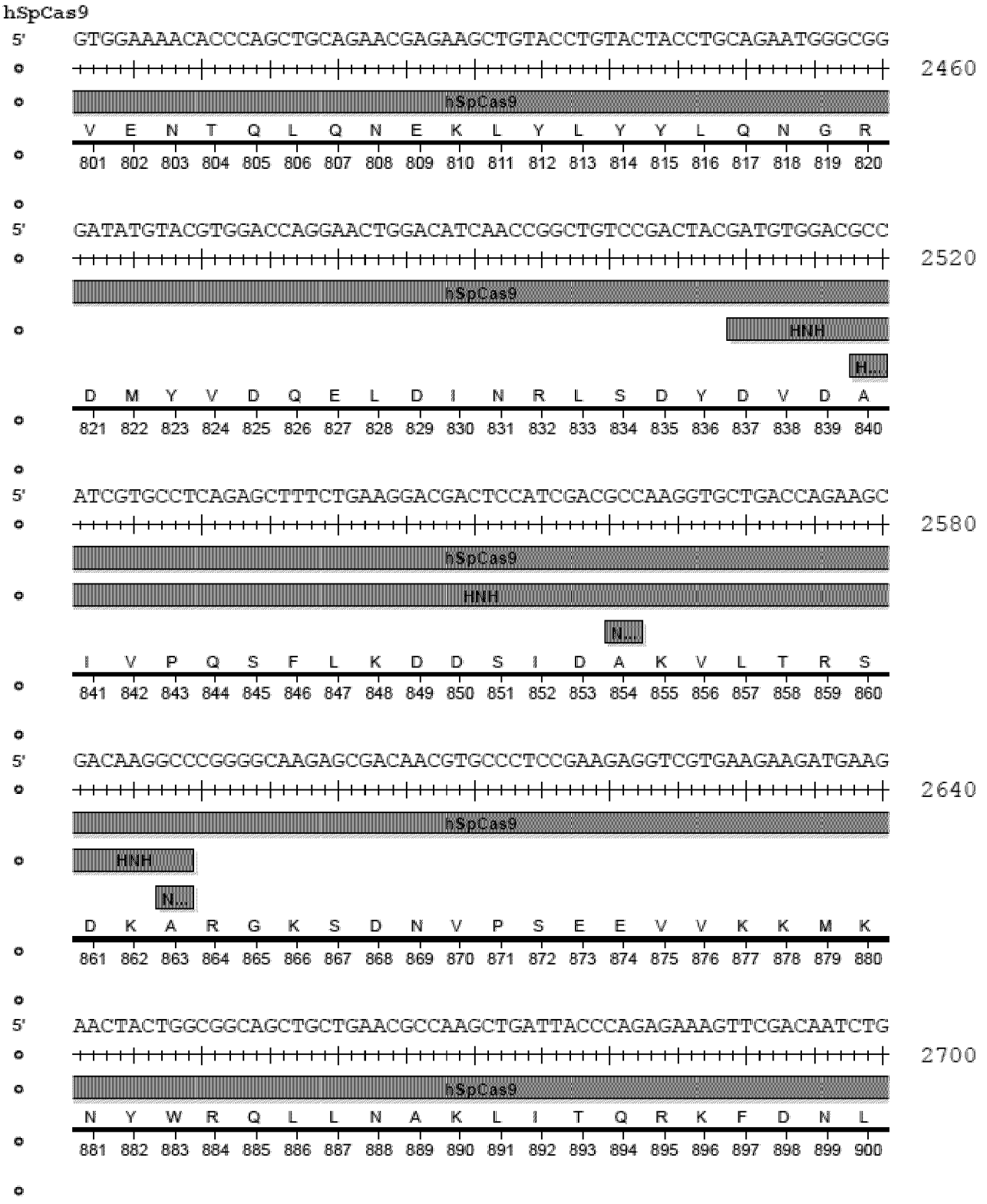

hSpCas9
5' GTGGAAAACACCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAATGGGCGG
2460
hSpCas9
V E N T Q L Q N E K L Y L Y Y L Q N G R
801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820
5' GATATGTACGTGGACCAGGAACTGGACATCAACCGGCTGTCCGACTACGATGTGGACGCC
2520
hSpCas9
HNH
H...
D M Y V D Q E L D I N R L S D Y D V D A
821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840
5' ATCGTGCCTCAGAGCTTTCTGAAGGACGACTCCATCGACGCCAAGGTGCTGACCAGAAGC
2580
hSpCas9
HNH
N...
I V P Q S F L K D D S I D A K V L T R S
841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860
5' GACAAGGCCCGGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTCGTGAAGAAGATGAAG
2640
hSpCas9
HNH
N...
D K A R G K S D N V P S E E V V K K M K
861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880
5' AACTACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTTCGACAATCTG
2700
hSpCas9
N Y W R Q L L N A K L I T Q R K F D N L
881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900

FIG. 24H

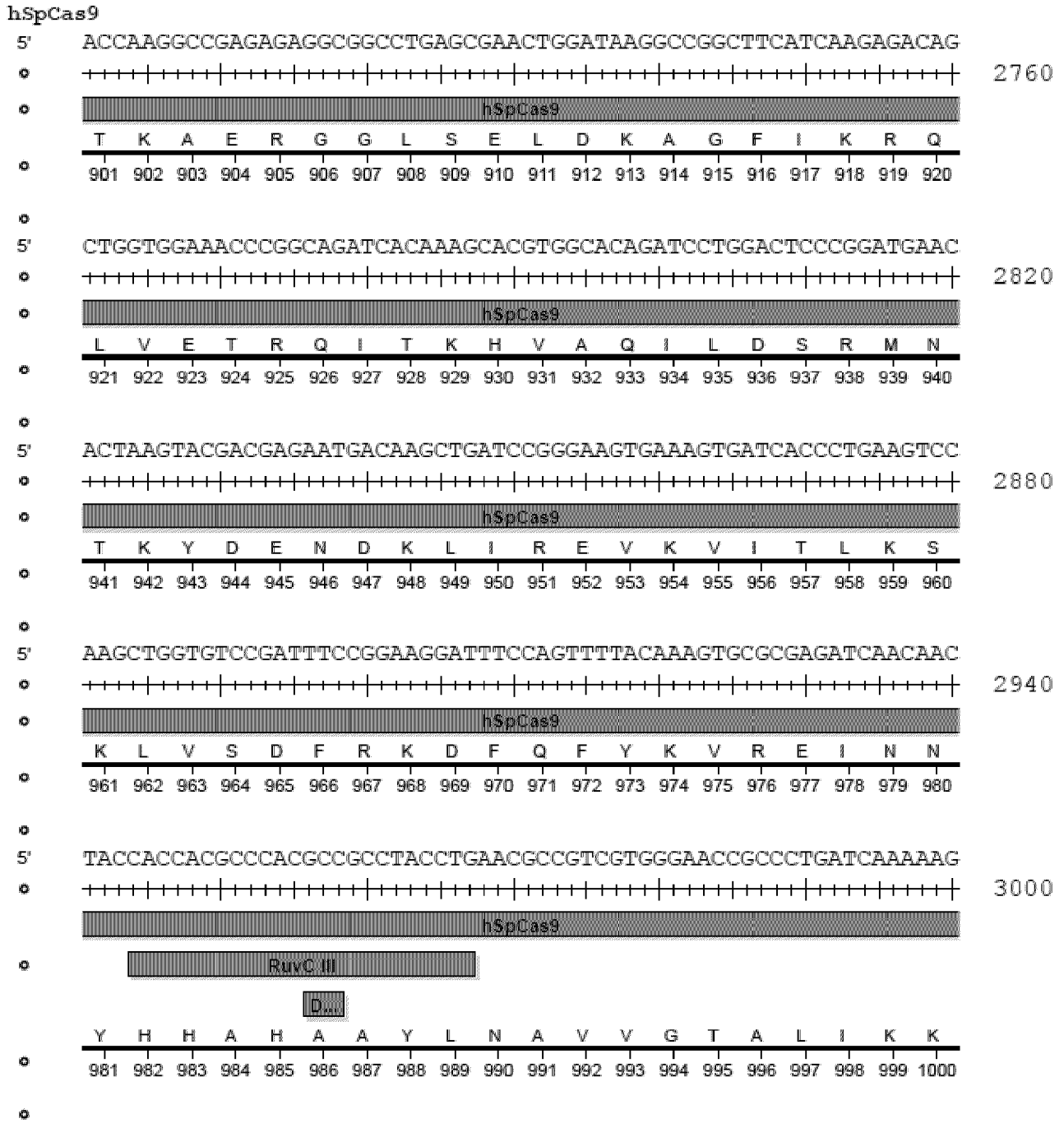

hSpCas9
5' ACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAG
2760
hSpCas9
T K A E R G G L S E L D K A G F I K R Q
901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920
5' CTGGTGGAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAAC
2820
hSpCas9
L V E T R Q I T K H V A Q I L D S R M N
921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940
5' ACTAAGTACGACGAGAATGACAAGCTGATCCGGGAAGTGAAAGTGATCACCCTGAAGTCC
2880
hSpCas9
T K Y D E N D K L I R E V K V I T L K S
941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960
5' AAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTTTACAAAGTGCGCGAGATCAACAAC
2940
hSpCas9
K L V S D F R K D F Q F Y K V R E I N N
961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980
5' TACCACCACGCCCACGCCGCCTACCTGAACGCCGTCGTGGGAACCGCCCTGATCAAAAAG
3000
hSpCas9
RuvC III
D...
Y H H A H A A Y L N A V V G T A L I K K
981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000

FIG. 24I

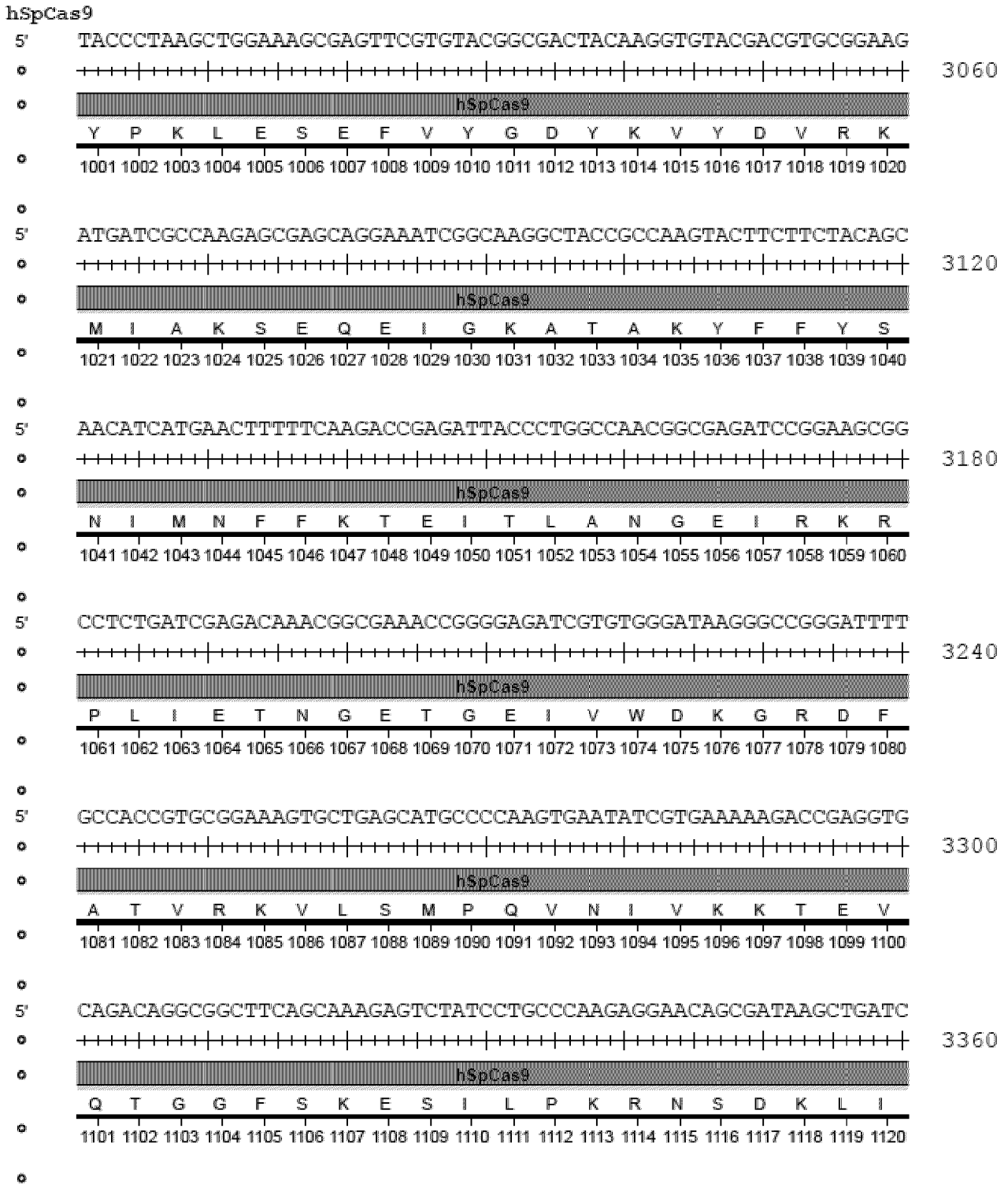

hSpCas9
5' TACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAG
3060
hSpCas9
Y P K L E S E F V Y G D Y K V Y D V R K
1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020
5' ATGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGC
3120
hSpCas9
M I A K S E Q E I G K A T A K Y F F Y S
1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040
5' AACATCATGAACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGG
3180
hSpCas9
N I M N F F K T E I T L A N G E I R K R
1041 1042 1043 1044 1045 1046 1047 1048 1049 1050 1051 1052 1053 1054 1055 1056 1057 1058 1059 1060
5' CCTCTGATCGAGACAAACGGCGAAACCGGGGAGATCGTGTGGGATAAGGGCCGGGATTTT
3240
hSpCas9
P L I E T N G E T G E I V W D K G R D F
1061 1062 1063 1064 1065 1066 1067 1068 1069 1070 1071 1072 1073 1074 1075 1076 1077 1078 1079 1080
5' GCCACCGTGCGGAAAGTGCTGAGCATGCCCCAAGTGAATATCGTGAAAAAGACCGAGGTG
3300
hSpCas9
A T V R K V L S M P Q V N I V K K T E V
1081 1082 1083 1084 1085 1086 1087 1088 1089 1090 1091 1092 1093 1094 1095 1096 1097 1098 1099 1100
5' CAGACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCCAAGAGGAACAGCGATAAGCTGATC
3360
hSpCas9
Q T G G F S K E S I L P K R N S D K L I
1101 1102 1103 1104 1105 1106 1107 1108 1109 1110 1111 1112 1113 1114 1115 1116 1117 1118 1119 1120

FIG. 24J

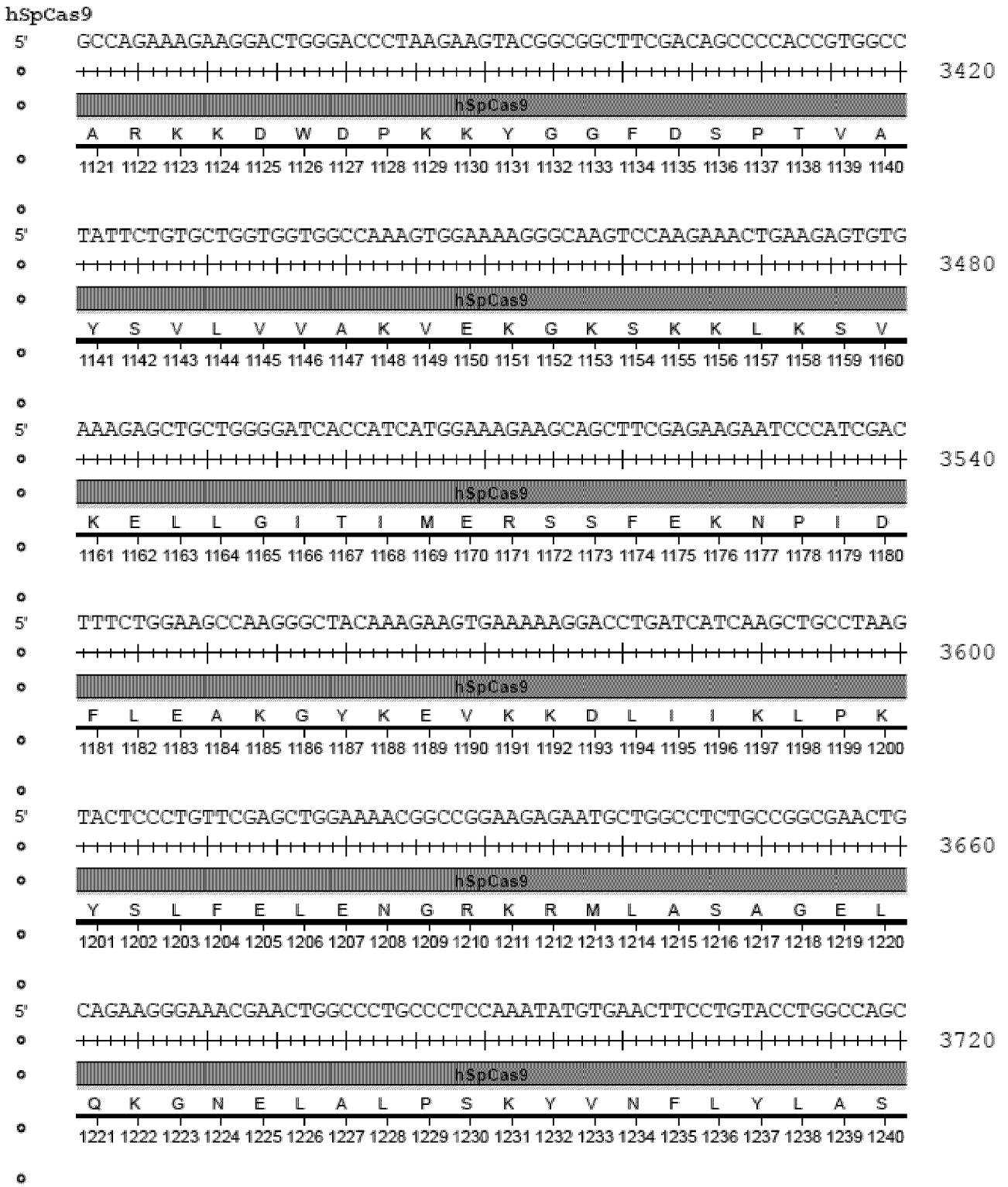

hSpCas9
5' GCCAGAAAGAAGGACTGGGACCCTAAGAAGTACGGCGGCTTCGACAGCCCCACCGTGGCC
3420
hSpCas9
A R K K D W D P K K Y G G F D S P T V A
1121 1122 1123 1124 1125 1126 1127 1128 1129 1130 1131 1132 1133 1134 1135 1136 1137 1138 1139 1140
5' TATTCTGTGCTGGTGGTGGCCAAAGTGGAAAAGGGCAAGTCCAAGAAACTGAAGAGTGTG
3480
hSpCas9
Y S V L V V A K V E K G K S K K L K S V
1141 1142 1143 1144 1145 1146 1147 1148 1149 1150 1151 1152 1153 1154 1155 1156 1157 1158 1159 1160
5' AAAGAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCCCATCGAC
3540
hSpCas9
K E L L G I T I M E R S S F E K N P I D
1161 1162 1163 1164 1165 1166 1167 1168 1169 1170 1171 1172 1173 1174 1175 1176 1177 1178 1179 1180
5' TTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAG
3600
hSpCas9
F L E A K G Y K E V K K D L I I K L P K
1181 1182 1183 1184 1185 1186 1187 1188 1189 1190 1191 1192 1193 1194 1195 1196 1197 1198 1199 1200
5' TACTCCCTGTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTG
3660
hSpCas9
Y S L F E L E N G R K R M L A S A G E L
1201 1202 1203 1204 1205 1206 1207 1208 1209 1210 1211 1212 1213 1214 1215 1216 1217 1218 1219 1220
5' CAGAAGGGAAACGAACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGC
3720
hSpCas9
Q K G N E L A L P S K Y V N F L Y L A S
1221 1222 1223 1224 1225 1226 1227 1228 1229 1230 1231 1232 1233 1234 1235 1236 1237 1238 1239 1240

FIG. 24K

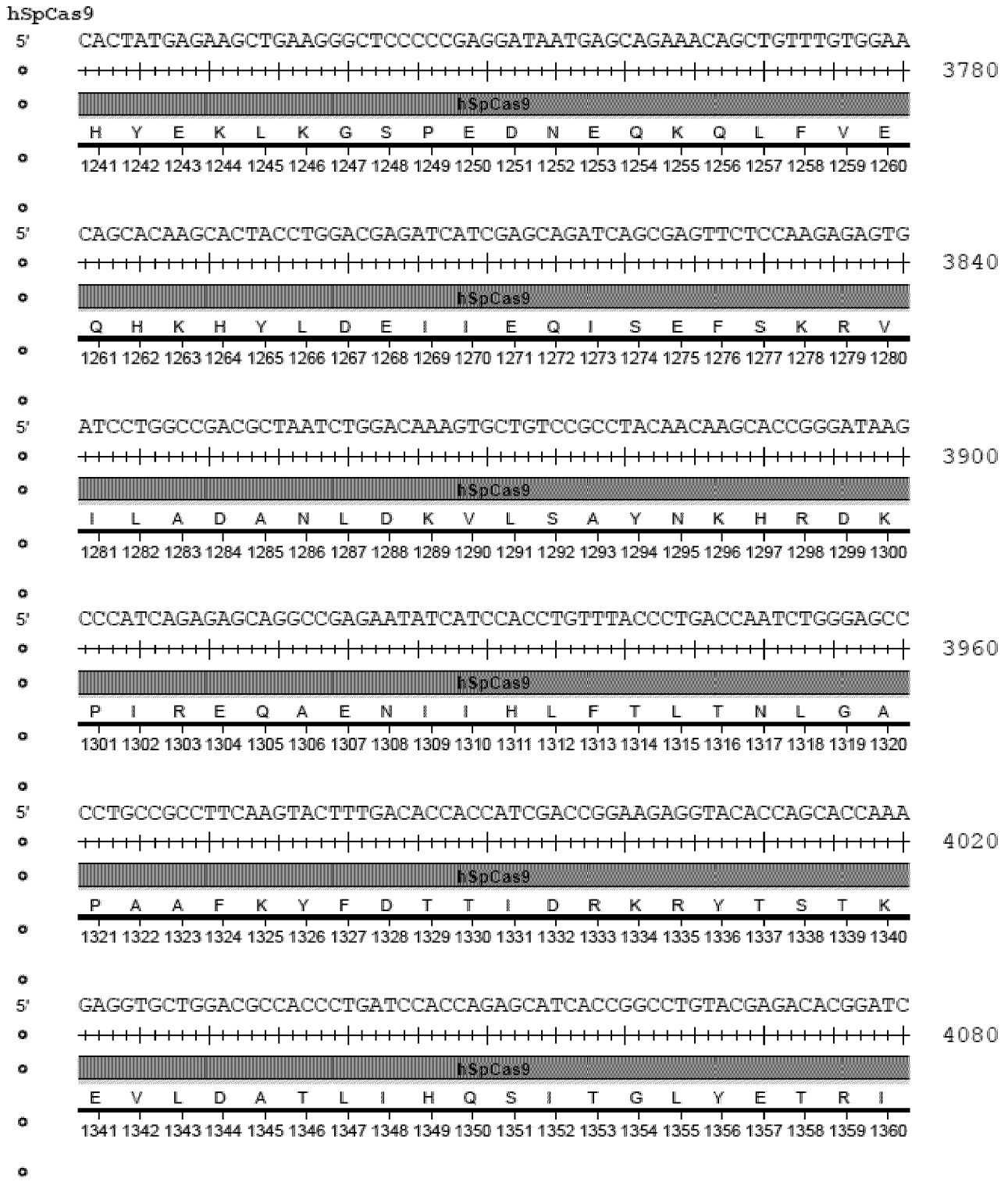

hSpCas9
5' CACTATGAGAAGCTGAAGGGCTCCCCCGAGGATAATGAGCAGAAACAGCTGTTTGTGGAA
3780
hSpCas9
H Y E K L K G S P E D N E Q K Q L F V E
1241 1242 1243 1244 1245 1246 1247 1248 1249 1250 1251 1252 1253 1254 1255 1256 1257 1258 1259 1260
5' CAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGAGTG
3840
hSpCas9
Q H K H Y L D E I I E Q I S E F S K R V
1261 1262 1263 1264 1265 1266 1267 1268 1269 1270 1271 1272 1273 1274 1275 1276 1277 1278 1279 1280
5' ATCCTGGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCTACAACAAGCACCGGGATAAG
3900
hSpCas9
I L A D A N L D K V L S A Y N K H R D K
1281 1282 1283 1284 1285 1286 1287 1288 1289 1290 1291 1292 1293 1294 1295 1296 1297 1298 1299 1300
5' CCCATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACCCTGACCAATCTGGGAGCC
3960
hSpCas9
P I R E Q A E N I I H L F T L T N L G A
1301 1302 1303 1304 1305 1306 1307 1308 1309 1310 1311 1312 1313 1314 1315 1316 1317 1318 1319 1320
5' CCTGCCGCCTTCAAGTACTTTGACACCACCATCGACCGGAAGAGGTACACCAGCACCAAA
4020
hSpCas9
P A A F K Y F D T T I D R K R Y T S T K
1321 1322 1323 1324 1325 1326 1327 1328 1329 1330 1331 1332 1333 1334 1335 1336 1337 1338 1339 1340
5' GAGGTGCTGGACGCCACCCTGATCCACCAGAGCATCACCGGCCTGTACGAGACACGGATC
4080
hSpCas9
E V L D A T L I H Q S I T G L Y E T R I
1341 1342 1343 1344 1345 1346 1347 1348 1349 1350 1351 1352 1353 1354 1355 1356 1357 1358 1359 1360

FIG. 24L

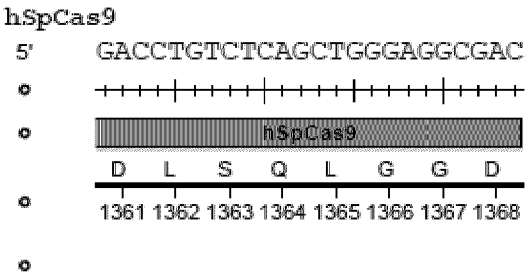

hSpCas9
5'
GACCTGTCTCAGCTGGGAGGCGAC
4104
hSpCas9
D L S Q L G G D
1361 1362 1363 1364 1365 1366 1367 1368

FIG. 24M

Conditional Cas9, Rosa26 targeting vector map

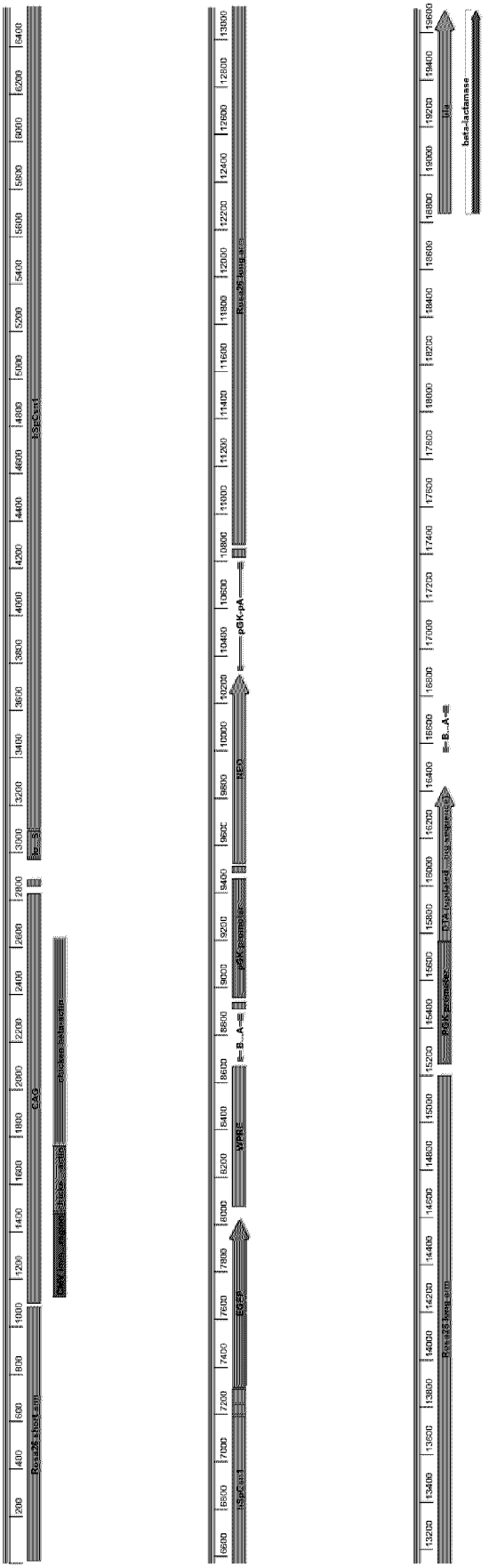


FIG. 25A

Constitutive Cas9, Rosa26 targeting vector map

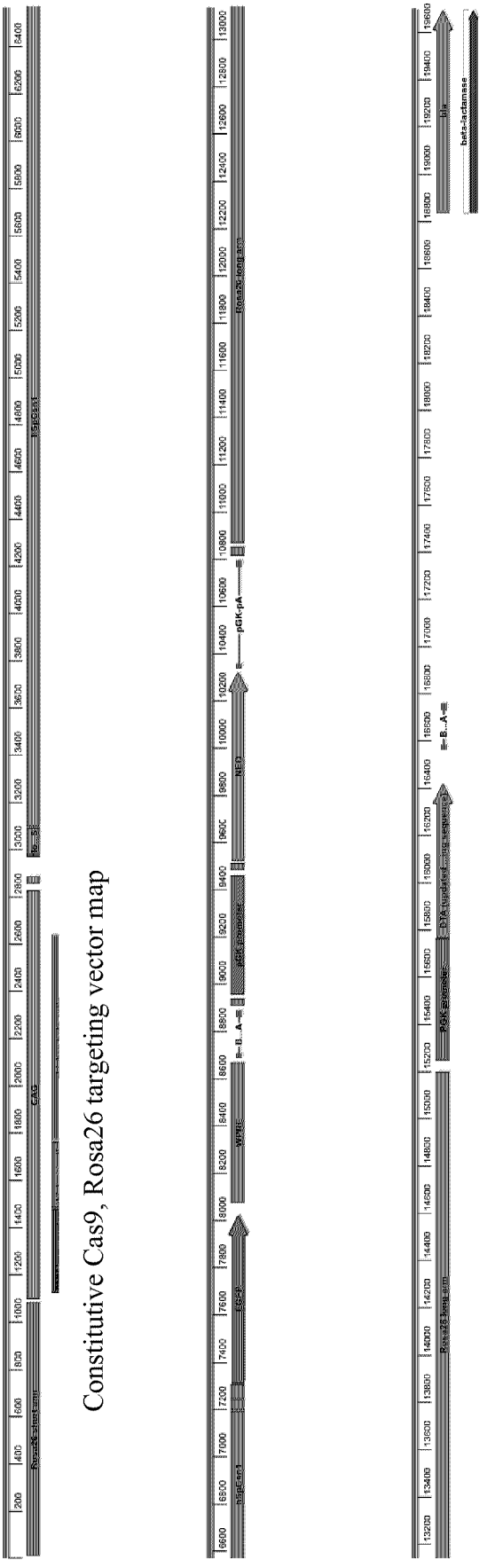


FIG. 25B

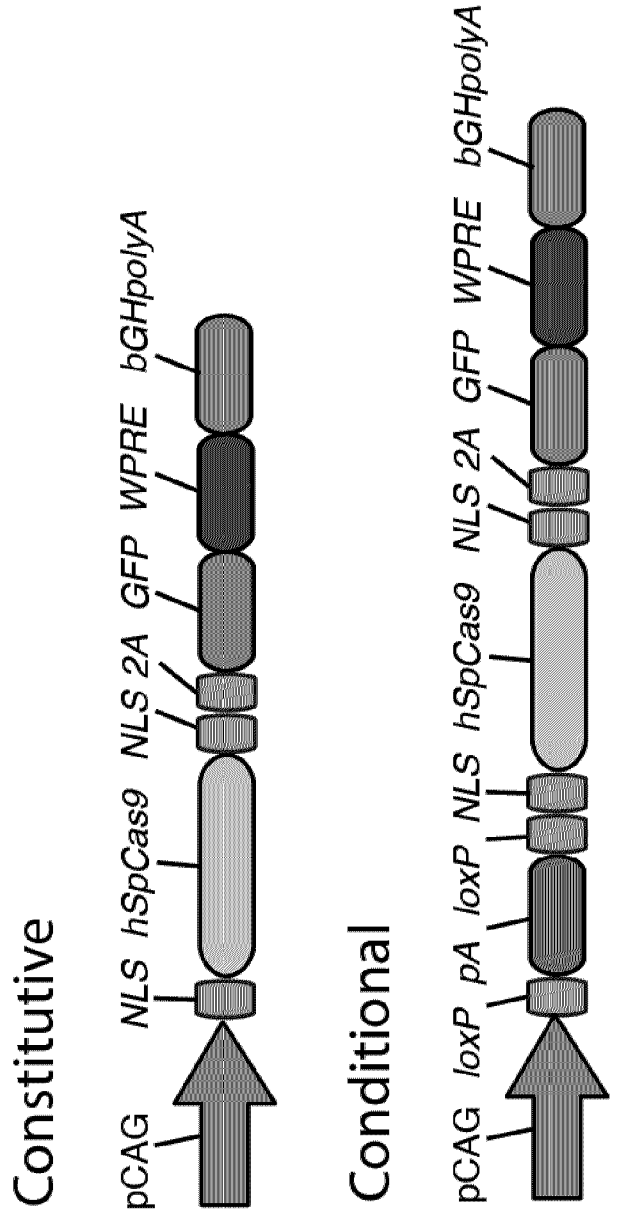
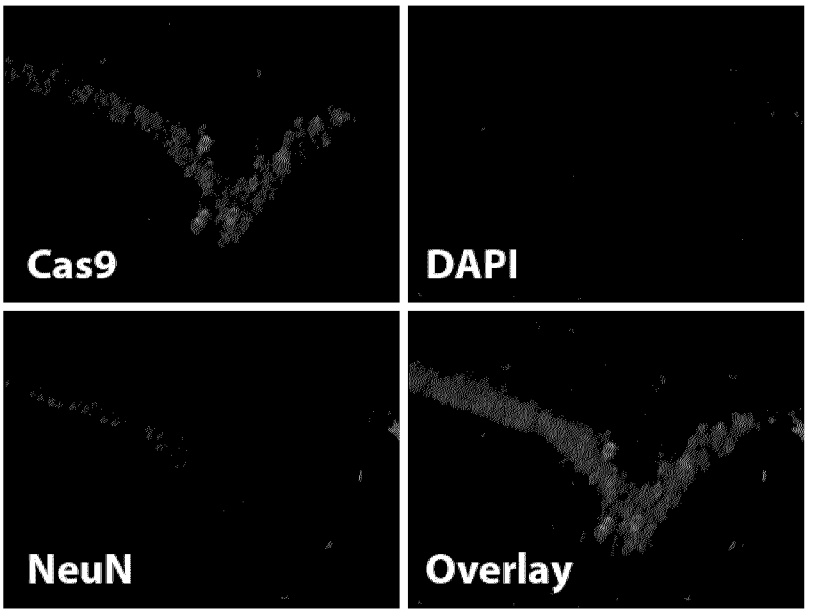

Constitutive
pCAG
NLS
hSpCas9
NLS
2A
GFP
WPRE
bGHpolyA
Conditional
pCAG
loxP
pA
loxP
NLS
hSpCas9
NLS
2A
GFP
WPRE
bGHpolyA

FIG. 26

Cas9 Expression in Mouse Hippocampus (AAV)



Cas9 Expression in Mouse Cortex (AAV)

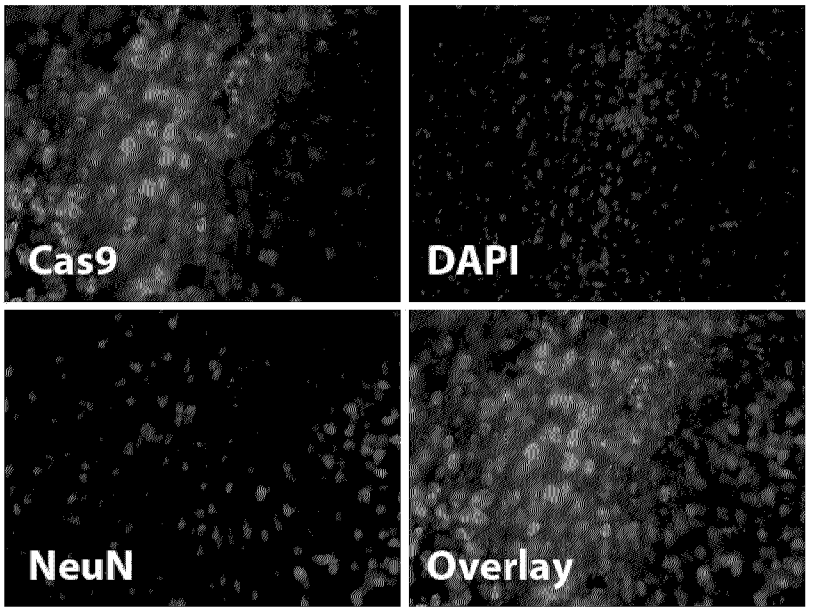


FIG. 27

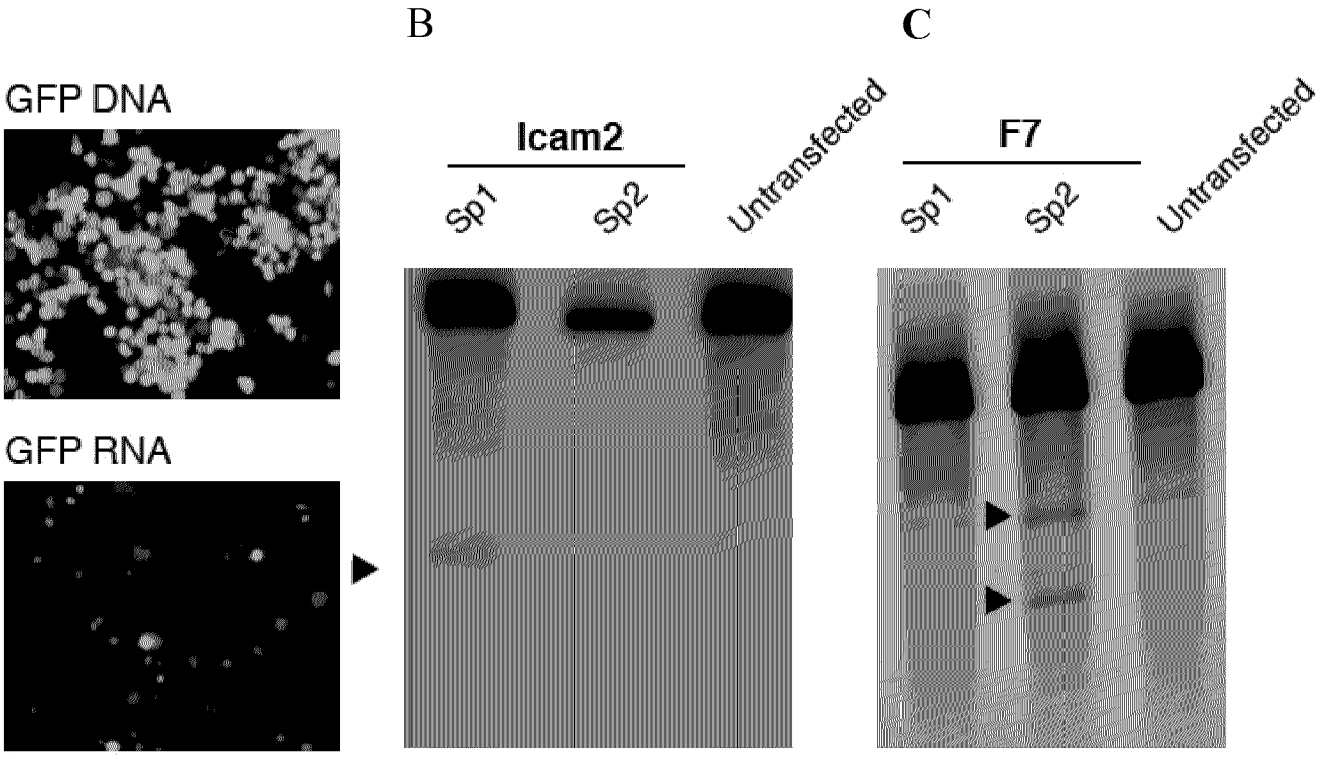

A
B
C
GFP DNA
GFP RNA
Icam2
Sp1
Sp2
Untransfected
F7
Sp1
Sp2
Untransfected

FIG. 28

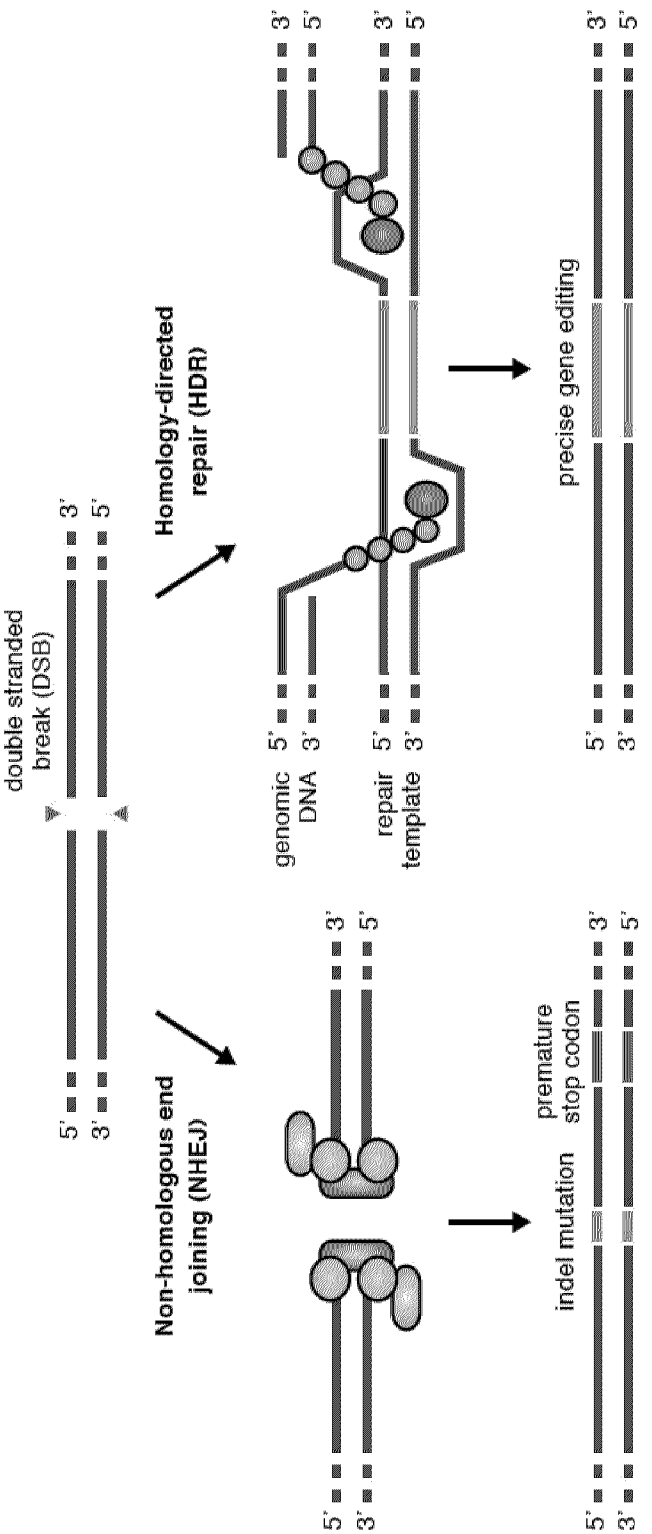

double stranded break (DSB)
5'
3'
3'
5'
Homology-directed repair (HDR)
genomic DNA
repair template
precise gene editing
Non-homologous end joining (NHEJ)
indel mutation
premature stop codon

FIG. 29

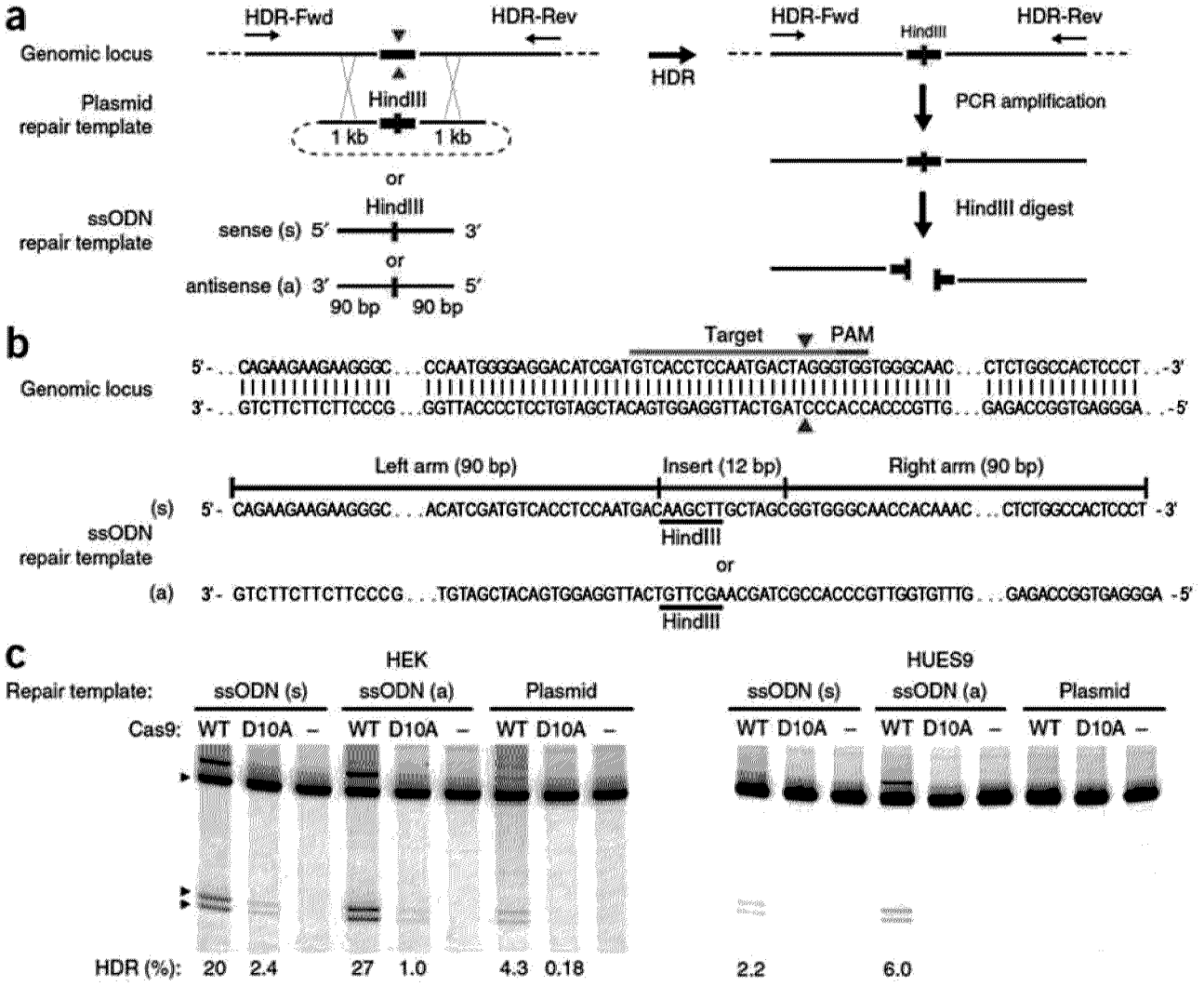

a
HDR-Fwd
HDR-Rev
Genomic locus
HindIII
Plasmid repair template
1 kb
1 kb
or
HindIII
ssODN repair template
sense (s) 5′
3′
or
antisense (a) 3′
5′
90 bp
90 bp
HDR
HDR-Fwd
HindIII
HDR-Rev
PCR amplification
HindIII digest
b
Target
PAM
Genomic locus
5′- . . CAGAAGAAGAAGGGC . . . CCAATGGGGAGGACATCGATGTCACCTCCAATGACTAGGGTGGTGGGCAAC . . . CTCTGGCCACTCCCT . . -3′
3′- . . GTCTTCTTCTTCCCG . . . GGTTACCCCTCCTGTAGCTACAGTGGAGGTTACTGATCCCACCACCCGTTG . . . GAGACCGGTGAGGGA . . -5′
Left arm (90 bp)
Insert (12 bp)
Right arm (90 bp)
(s)
5′- CAGAAGAAGAAGGGC . . . ACATCGATGTCACCTCCAATGACAAGCTTGCTAGCGGTGGGCAACCACAAAC . . . CTCTGGCCACTCCCT -3′
HindIII
ssODN repair template
or
(a)
3′- GTCTTCTTCTTCCCG . . . TGTAGCTACAGTGGAGGTTACTGTTCGAACGATCGCCACCCGTTGGTGTTTG . . . GAGACCGGTGAGGGA -5′
HindIII
c
HEK
HUES9
Repair template:
ssODN (s)
ssODN (a)
Plasmid
ssODN (s)
ssODN (a)
Plasmid
Cas9:
WT D10A –
WT D10A –
WT D10A –
WT D10A –
WT D10A –
WT D10A –
HDR (%):
20
2.4
27
1.0
4.3
0.18
2.2
6.0

FIG. 30A-C

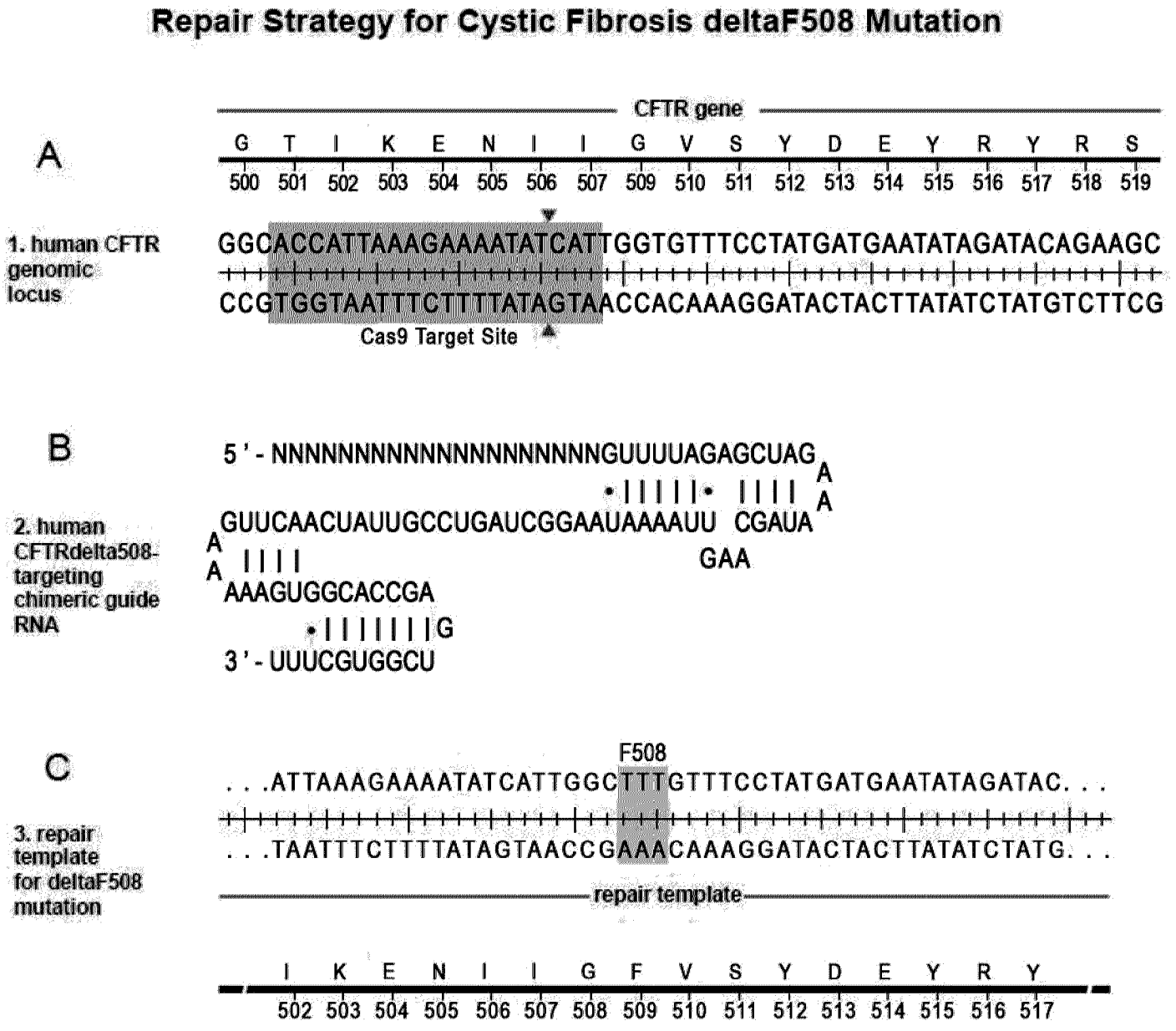
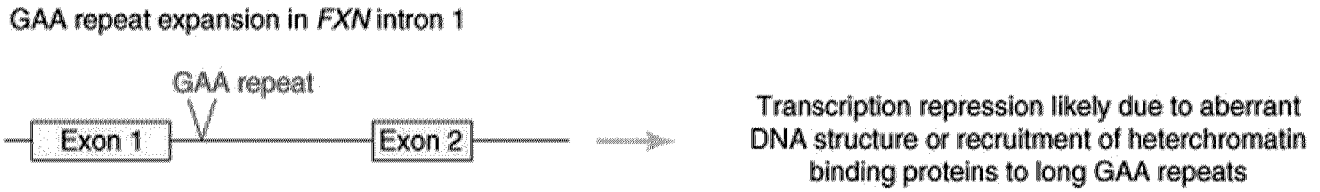

Repair Strategy for Cystic Fibrosis deltaF508 Mutation
A
CFTR gene
G T I K E N I I G V S Y D E Y R Y R S
500 501 502 503 504 505 506 507 509 510 511 512 513 514 515 516 517 518 519
1. human CFTR genomic locus
GGCACCATTAAAGAAAATATCATTGGTGTTTCCTATGATGAATATAGATACAGAAGC
CCGTGGTAATTTCTTTTATAGTAACCACAAAGGATACTACTTATATCTATGTCTTCG
Cas9 Target Site
B
5 ' - NNNNNNNNNNNNNNNNNNNGUUUUAGAGCUAG
GUUCAACUAUUGCCUGAUCGGAAUAAAAUU CGAUA
GAA
AAAGUGGCACCGA
3 ' - UUUCGUGGCU
2. human CFTRdelta508-targeting chimeric guide RNA
C
F508
. . .ATTAAAGAAAATATCATTGGCTTTGTTTCCTATGATGAATATAGATAC. . .
. . .TAATTTCTTTTATAGTAACCGAAACAAAGGATACTACTTATATCTATG. . .
repair template
I K E N I I G F V S Y D E Y R Y
502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517
3. repair template for deltaF508 mutation

FIG. 31

A


GAA repeat expansion in *FXN* intron 1
GAA repeat
Exon 1
Exon 2
Transcription repression likely due to aberrant DNA structure or recruitment of heterchromatin binding proteins to long GAA repeats

B

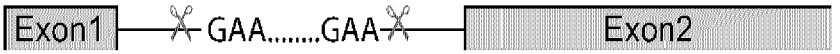

Exon1
GAA.....GAA
Exon2

FIG. 32

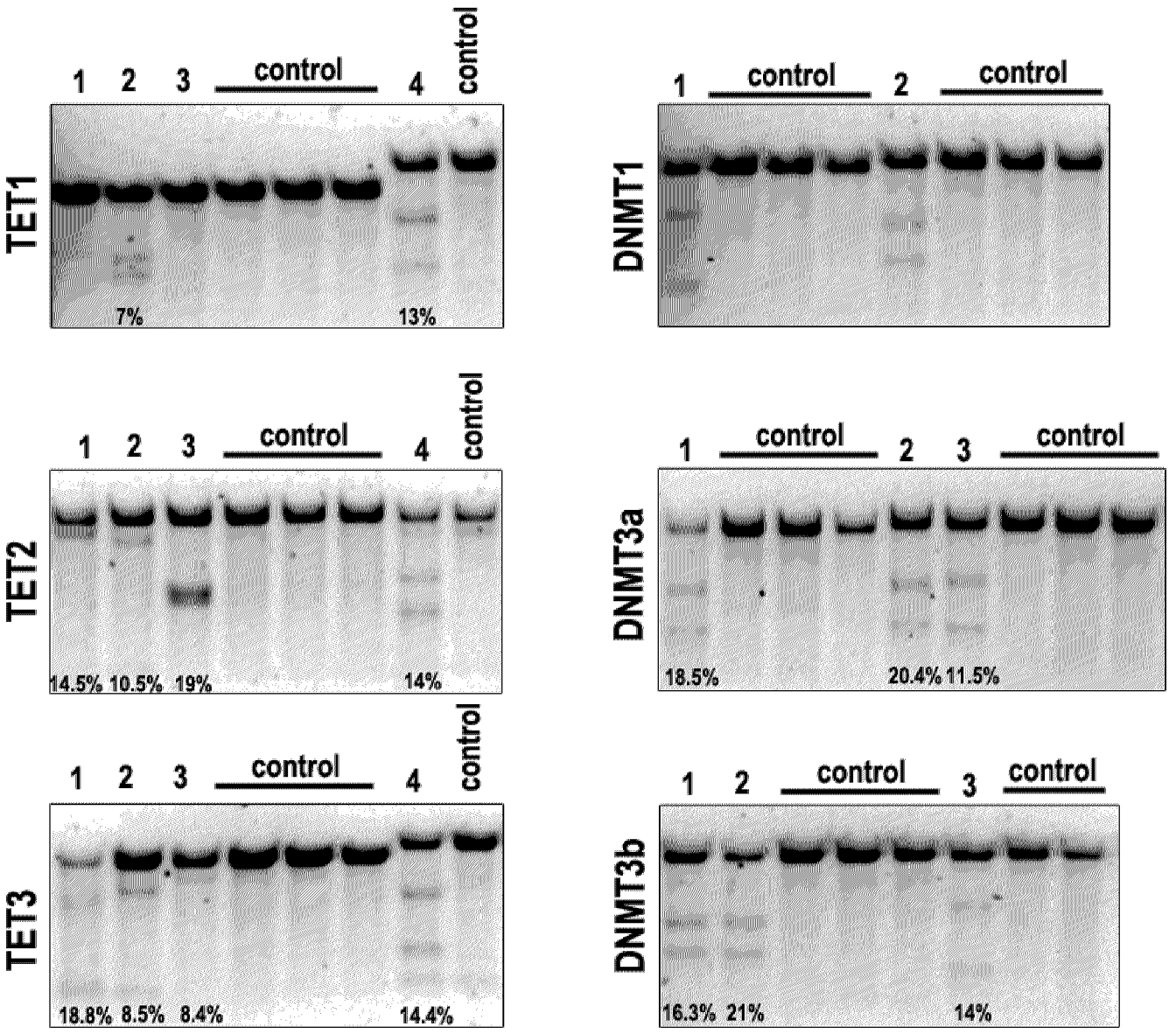

TET1
1 2 3 control 4 control
7%
13%
DNMT1
1 control 2 control
TET2
1 2 3 control 4 control
14.5% 10.5% 19%
14%
DNMT3a
1 control 2 3 control
18.5%
20.4% 11.5%
TET3
1 2 3 control 4 control
18.8% 8.5% 8.4%
14.4%
DNMT3b
1 2 control 3 control
16.3% 21%
14%

FIG. 33

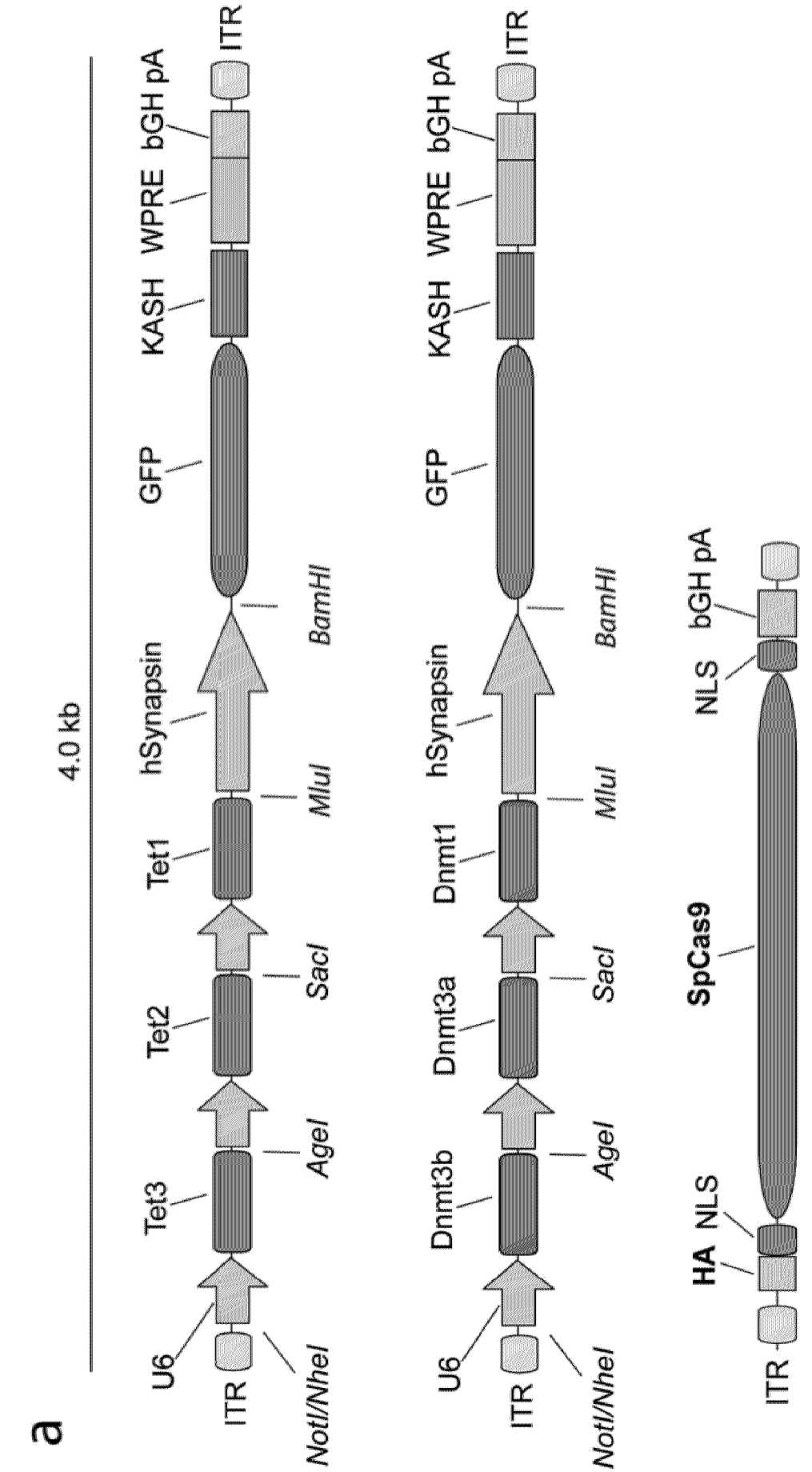

a
4.0 kb
ITR
U6
NotI/NheI
Tet3
AgeI
Tet2
SacI
Tet1
MluI
hSynapsin
GFP
BamHI
KASH
WPRE
bGH pA
ITR
ITR
U6
NotI/NheI
Dnmt3b
AgeI
Dnmt3a
SacI
Dnmt1
MluI
hSynapsin
GFP
BamHI
KASH
WPRE
bGH pA
ITR
ITR
HA
NLS
SpCas9
NLS
bGH pA

FIG. 34

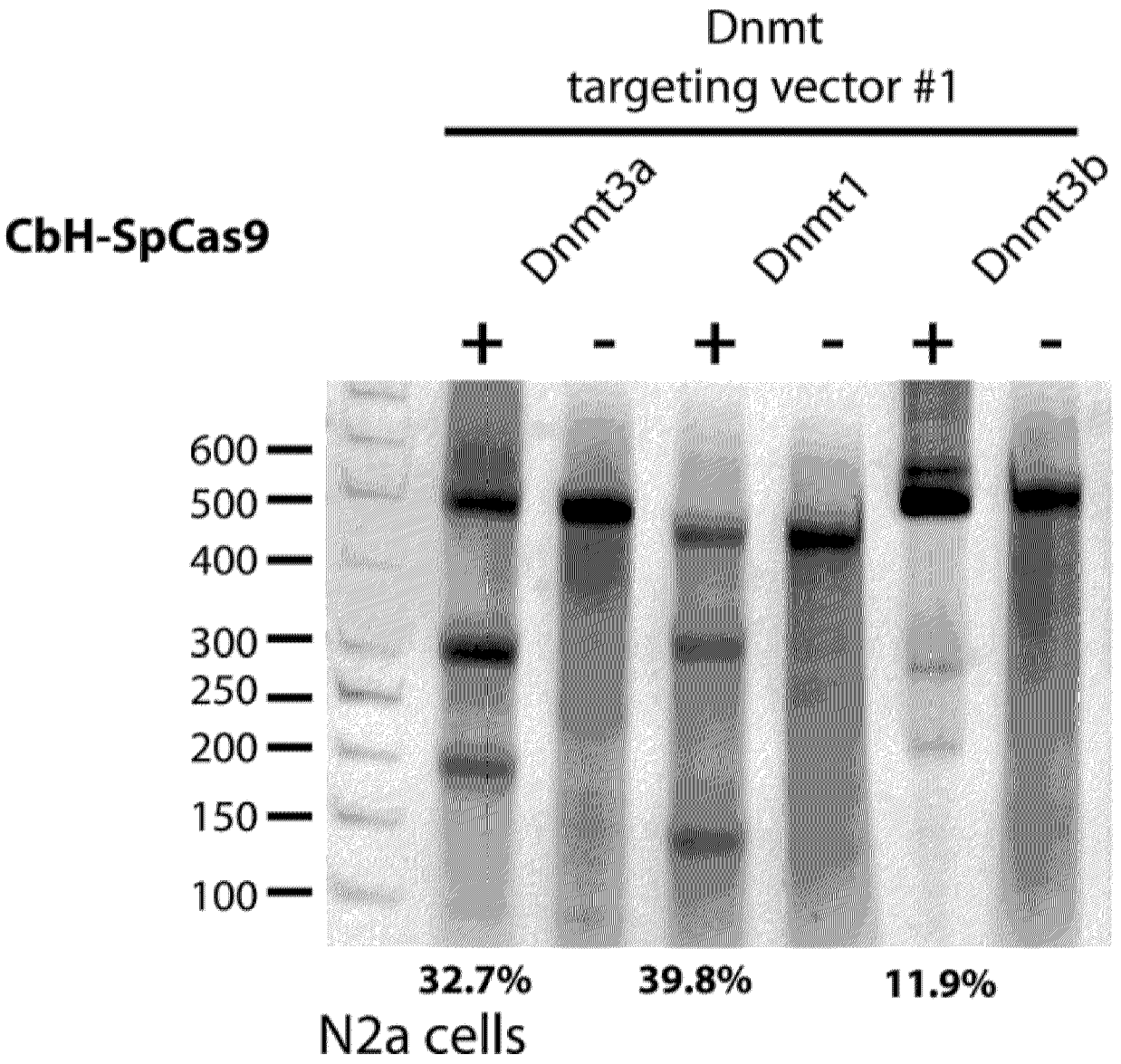

Dnmt
targeting vector #1
CbH-SpCas9
Dnmt3a
Dnmt1
Dnmt3b
+ - + - + -
600
500
400
300
250
200
150
100
32.7%
39.8%
11.9%
N2a cells

FIG. 35

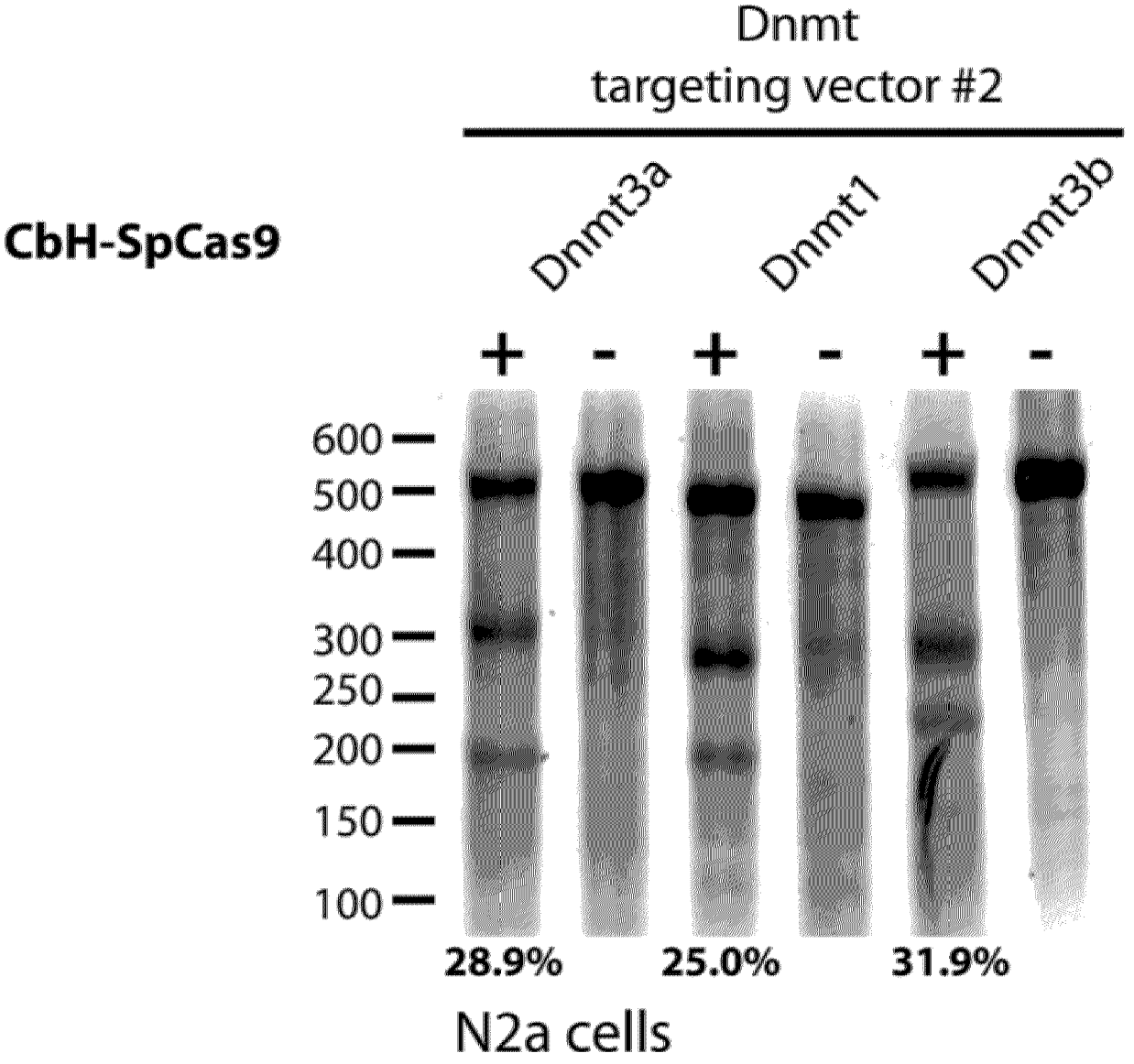

Dnmt
targeting vector #2
CbH-SpCas9
Dnmt3a
Dnmt1
Dnmt3b
+ - + - + -
600
500
400
300
250
200
150
100
28.9%
25.0%
31.9%
N2a cells

FIG. 36

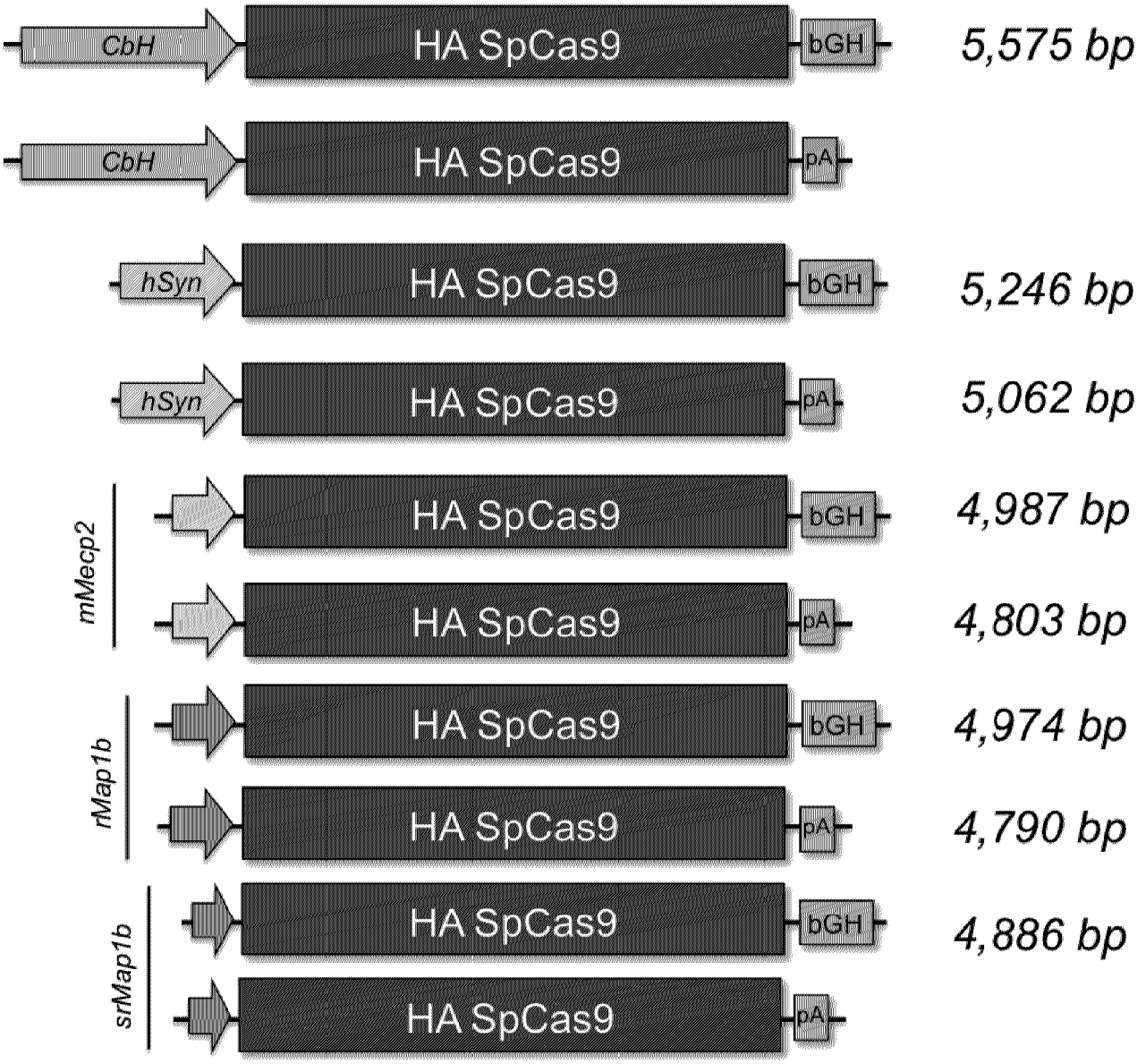

CbH
HA SpCas9
bGH
5,575 bp
CbH
HA SpCas9
pA
hSyn
HA SpCas9
bGH
5,246 bp
hSyn
HA SpCas9
pA
5,062 bp
mMecp2
HA SpCas9
bGH
4,987 bp
HA SpCas9
pA
4,803 bp
rMap1b
HA SpCas9
bGH
4,974 bp
HA SpCas9
pA
4,790 bp
srMap1b
HA SpCas9
bGH
4,886 bp
HA SpCas9
pA

FIG. 37

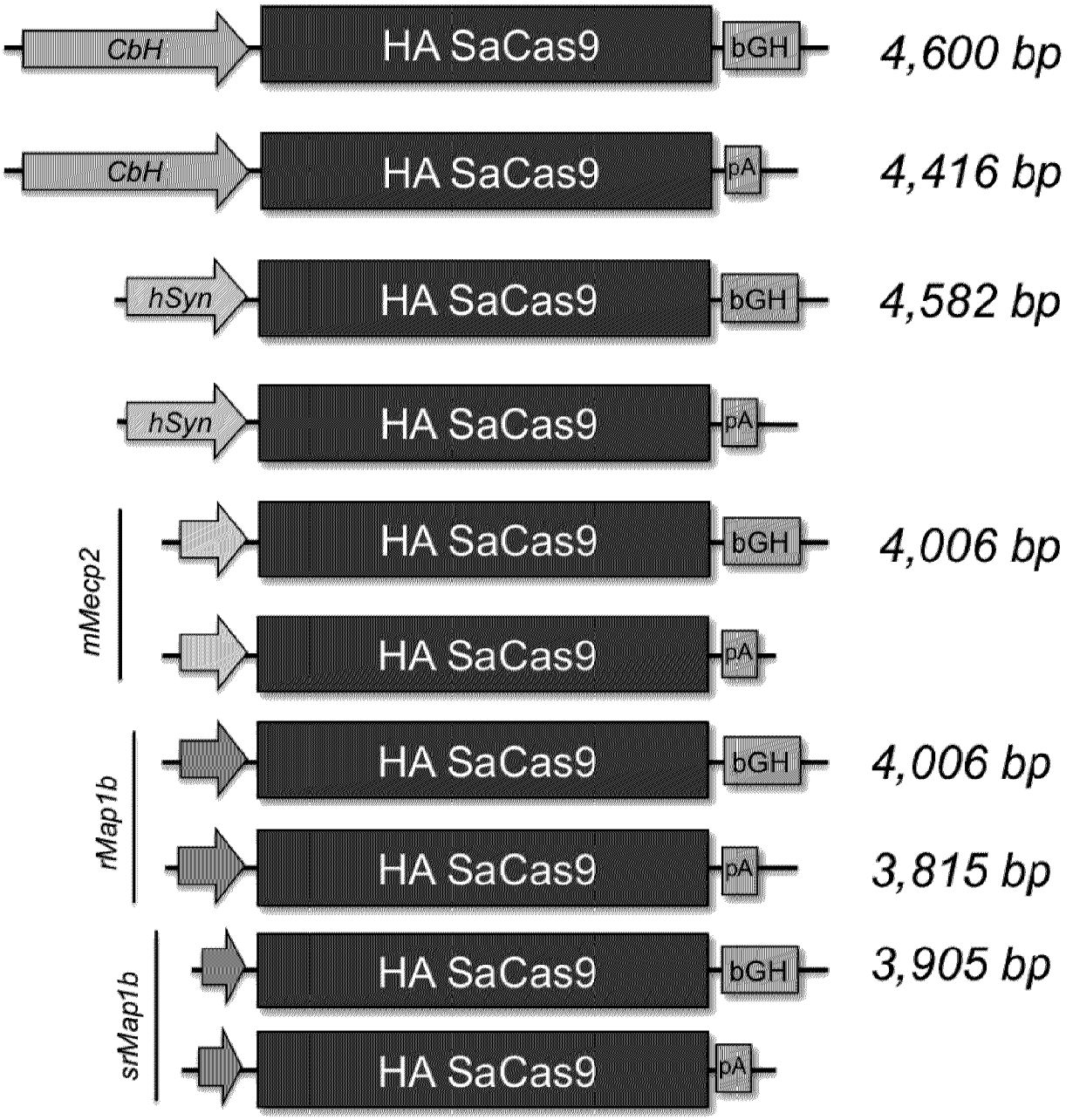

CbH
HA SaCas9
bGH
4,600 bp
CbH
HA SaCas9
pA
4,416 bp
hSyn
HA SaCas9
bGH
4,582 bp
hSyn
HA SaCas9
pA
mMecp2
HA SaCas9
bGH
4,006 bp
HA SaCas9
pA
rMap1b
HA SaCas9
bGH
4,006 bp
HA SaCas9
pA
3,815 bp
srMap1b
HA SaCas9
bGH
3,905 bp
HA SaCas9
pA

FIG. 38

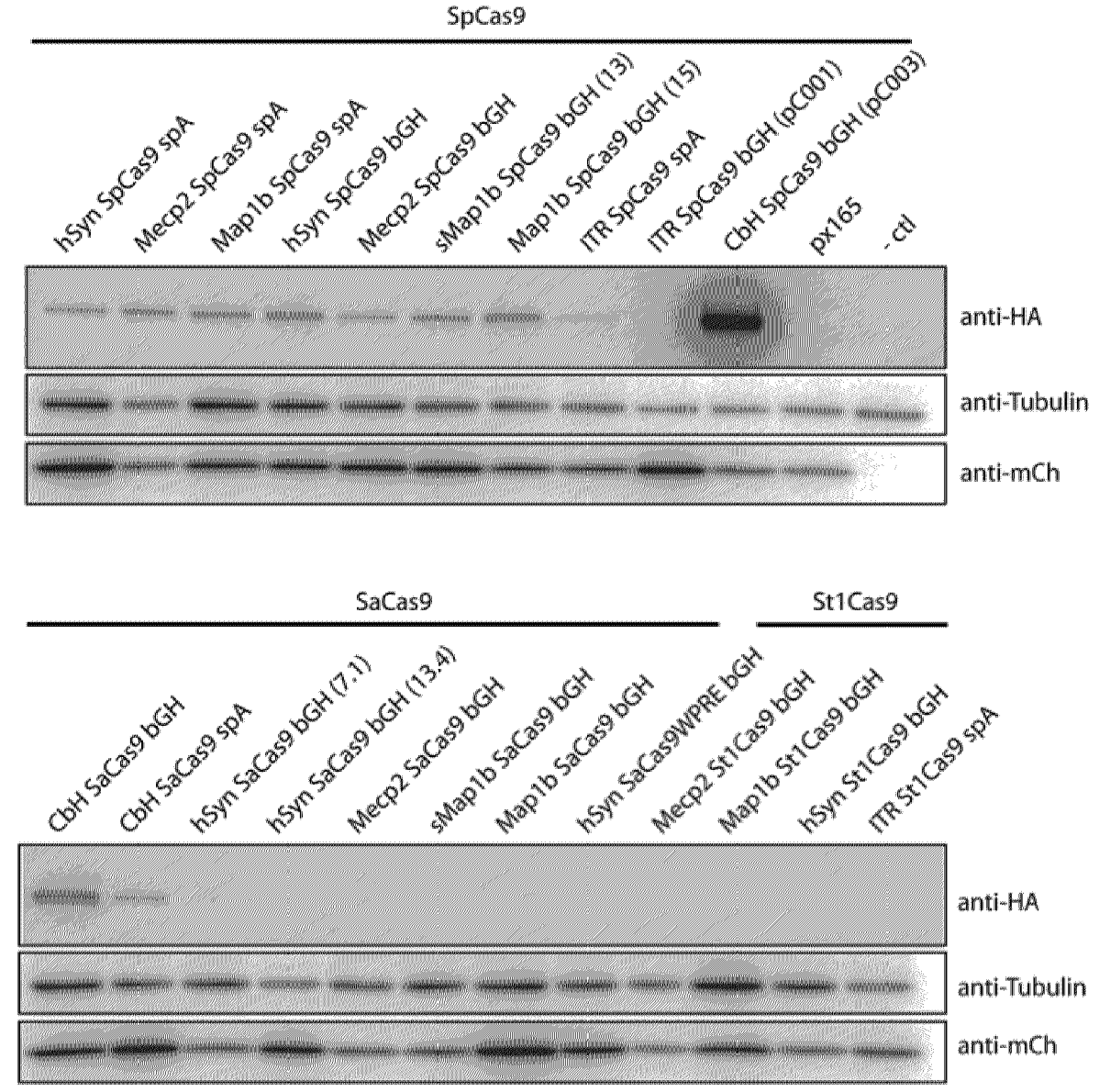

Expression of SpCas9 & SaCas9 in N2a cells
SpCas9
hSyn SpCas9 spA
Mecp2 SpCas9 spA
Map1b SpCas9 spA
hSyn SpCas9 bGH
Mecp2 SpCas9 bGH
sMap1b SpCas9 bGH (13)
Map1b SpCas9 bGH (15)
ITR SpCas9 spA
ITR SpCas9 bGH (pC001)
CbH SpCas9 bGH (pC003)
px165
- ctl
anti-HA
anti-Tubulin
anti-mCh
SaCas9
St1Cas9
CbH SaCas9 bGH
CbH SaCas9 spA
hSyn SaCas9 bGH (7.1)
hSyn SaCas9 bGH (13.4)
Mecp2 SaCas9 bGH
sMap1b SaCas9 bGH
Map1b SaCas9 bGH
hSyn SaCas9WPRE bGH
Mecp2 St1Cas9 bGH
Map1b St1Cas9 bGH
hSyn St1Cas9 bGH
ITR St1Cas9 spA
anti-HA
anti-Tubulin
anti-mCh

FIG. 39

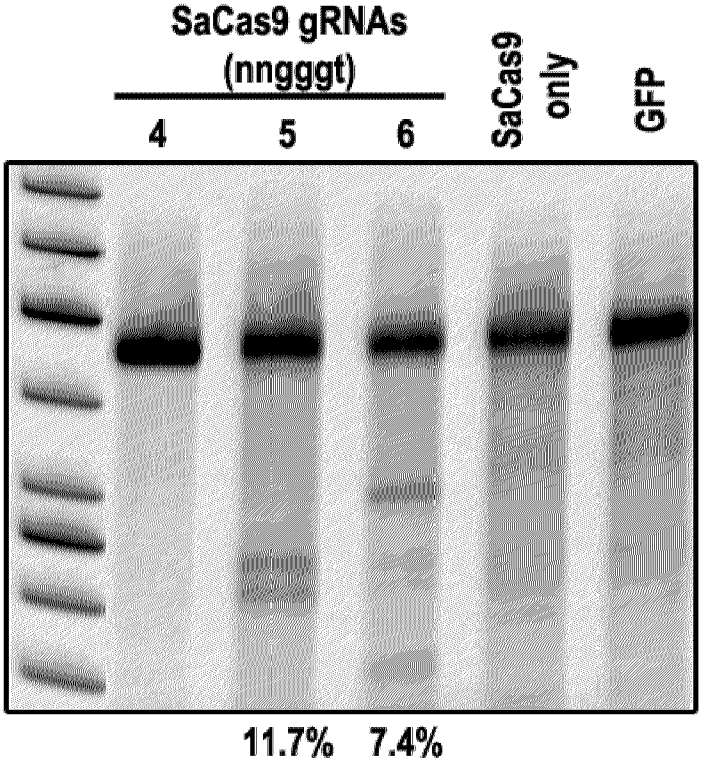


FIG. 40

DEMANDE OU BREVET VOLUMINEUX

LA PRÉSENTE PARTIE DE CETTE DEMANDE OU CE BREVET COMPREND PLUS D'UN TOME.

CECI EST LE TOME 1 DE 2
CONTENANT LES PAGES 1 À 77

NOTE : Pour les tomes additionels, veuillez contacter le Bureau canadien des brevets

JUMBO APPLICATIONS/PATENTS

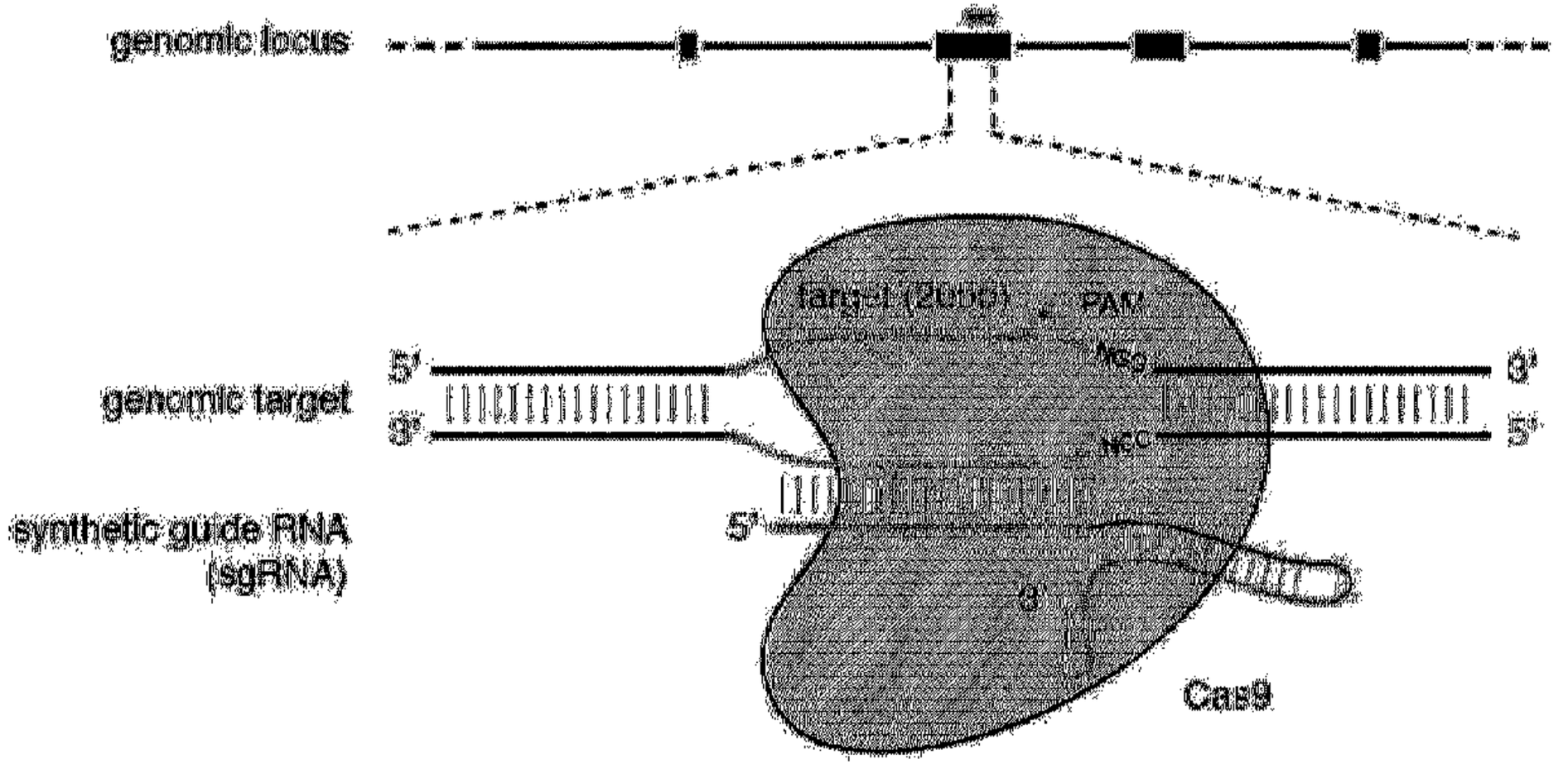
THIS SECTION OF THE APPLICATION/PATENT CONTAINS MORE THAN ONE VOLUME

THIS IS VOLUME 1 OF 2
CONTAINING PAGES 1 TO 77

NOTE: For additional volumes, please contact the Canadian Patent Office

NOM DU FICHIER / FILE NAME :

NOTE POUR LE TOME / VOLUME NOTE:


genomic locus
genomic target
synthetic guide RNA (sgRNA)
5'
3'
target 1 (20nt)
PAM
NGG
NCC
5'
3'
Cas9