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

Innovation, Science and Economic Development Canada
Canadian Intellectual Property Office

CA 2989830 A1 2016/12/22

(21) 2 989 830

(12) DEMANDE DE BREVET CANADIEN
CANADIAN PATENT APPLICATION

(13) A1

(86) Date de dépôt PCT/PCT Filing Date: 2016/06/17
(87) Date publication PCT/PCT Publication Date: 2016/12/22
(85) Entrée phase nationale/National Entry: 2017/12/15
(86) N° demande PCT/PCT Application No.: US 2016/038034
(87) N° publication PCT/PCT Publication No.: 2016/205613
(30) Priorités/Priorities: 2015/06/18 (US62/181,453); 2015/08/19 (US62/207,312); 2015/10/05 (US62/237,360); 2015/11/13 (US62/255,256); 2015/12/18 (US62/269,876)

(51) Cl.Int./Int.Cl. *C07K 14/315* (2006.01), *C12N 9/22* (2006.01)

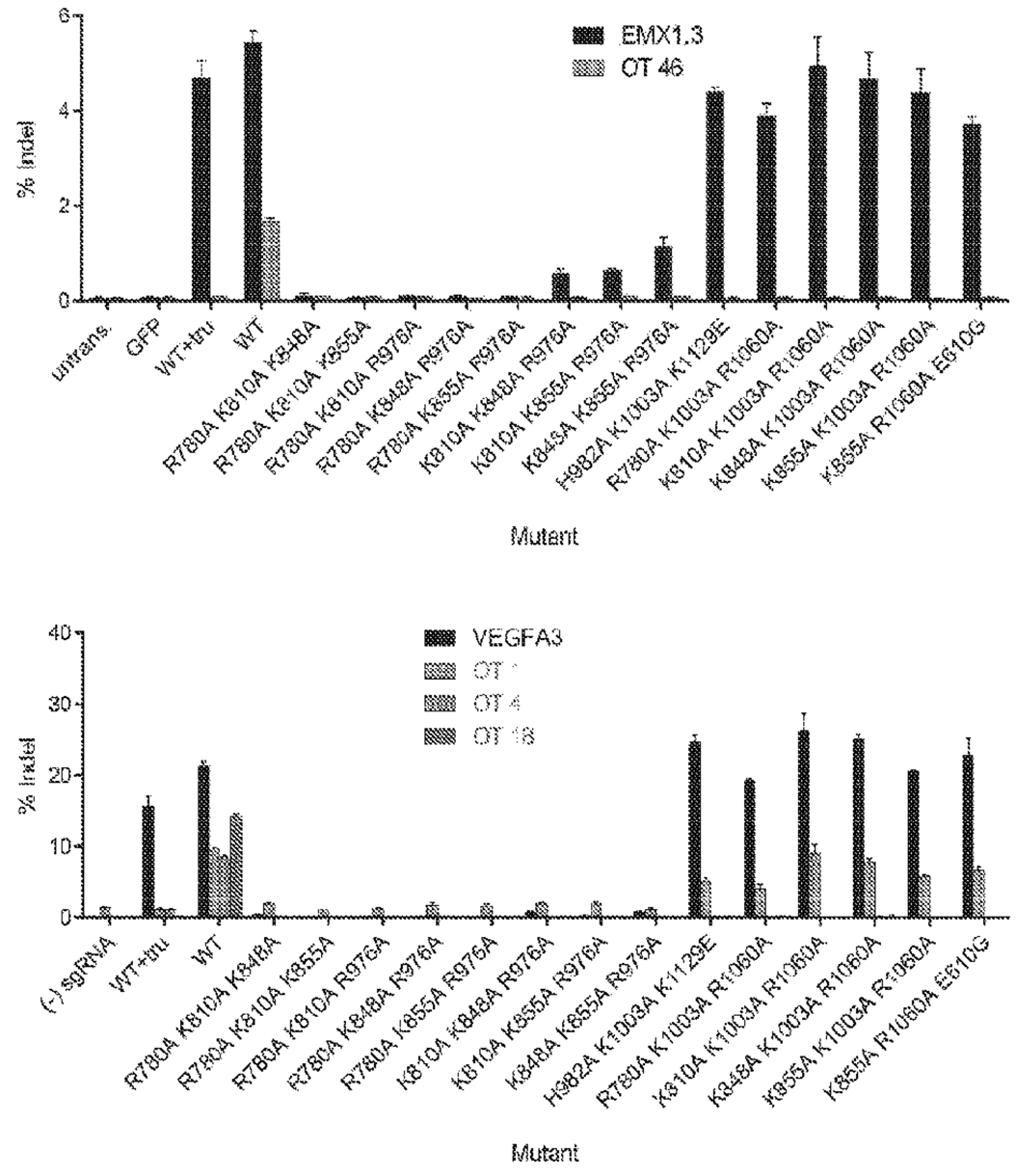
(71) Demandeurs/Applicants:
THE BROAD INSTITUTE INC., US;
MASSACHUSETTS INSTITUTE OF TECHNOLOGY, US

(72) Inventeurs/Inventors:
ZHANG, FENG, US;
GAO, LINYI, US;
ZETSCHE, BERND, US;
SLAYMAKER, IAN, US

(74) Agent: MBM INTELLECTUAL PROPERTY LAW LLP

(54) Titre : MUTATIONS D'ENZYME CRISPR QUI REDUISENT LES EFFETS NON CIBLES
(54) Title: CRISPR ENZYME MUTATIONS REDUCING OFF-TARGET EFFECTS

Fig. 5



(57) Abrégé/Abstract:
Disclosed and claimed are mutation(s) or modification(s) of the CRISPR enzyme, for example a Cas enzyme such as a Cas9, which obtain an improvement, for instance a reduction, as to off-target effects of a CRISPR-Cas or CRISPR-enzyme or CRISPR-Cas9 system or complex containing or including such a mutated or modified Cas or CRISPR enzyme or Cas9. Methods for making and using and uses of such mutated or modified Cas or CRISPR enzyme or Cas9 and systems or complexes containing the same and products from such methods and uses are also disclosed and claimed.

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
22 December 2016 (22.12.2016)

WIPO | PCT

(10) International Publication Number
WO 2016/205613 A1

(51) International Patent Classification:
C07K 14/315 (2006.01) *C12N 9/22* (2006.01)

(21) International Application Number:
PCT/US2016/038034

(22) International Filing Date:
17 June 2016 (17.06.2016)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

62/181,453	18 June 2015 (18.06.2015)	US
62/207,312	19 August 2015 (19.08.2015)	US
62/237,360	5 October 2015 (05.10.2015)	US
62/255,256	13 November 2015 (13.11.2015)	US
62/269,876	18 December 2015 (18.12.2015)	US

(71) Applicants: **THE BROAD INSTITUTE INC.** [US/US]; 415 Main Street, 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). **GAO, Linyi** [US/US]; 70 Pacific Street, #482c, Cambridge, MA 02139 (US). **ZETSCHE, Bernd** [DE/US]; 383 Washington Street, Gloucester, MA 01930 (US). **SLAYMAKER, Ian** [US/US]; 1783 Massachusetts Ave., Apt 1, Cambridge, MA 02140 (US).

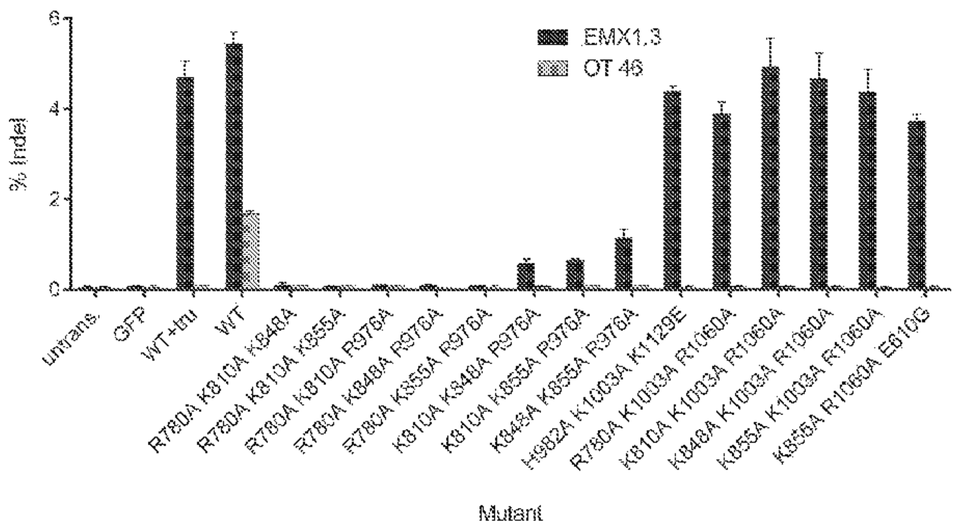
(74) Agents: **KOWALSKI, Thomas, J.** et al.; Vedder Price P.c., 1633 Broadway, New York, NY 10019 (US).

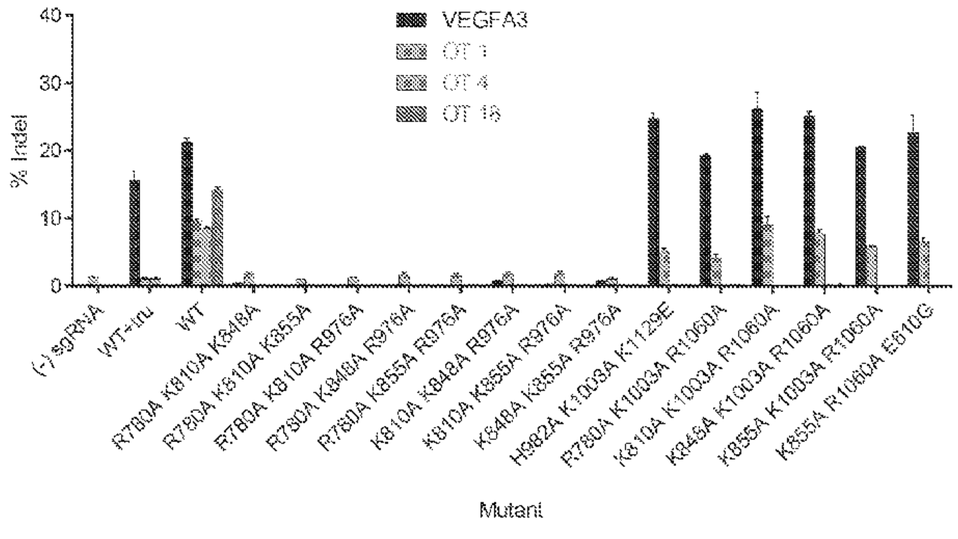
(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP,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

[Continued on next page]

(54) Title: CRISPR ENZYME MUTATIONS REDUCING OFF-TARGET EFFECTS

Fig. 5





(57) Abstract: Disclosed and claimed are mutation(s) or modification(s) of the CRISPR enzyme, for example a Cas enzyme such as a Cas9, which obtain an improvement, for instance a reduction, as to off-target effects of a CRISPR-Cas or CRISPR-enzyme or CRISPR-Cas9 system or complex containing or including such a mutated or modified Cas or CRISPR enzyme or Cas9. Methods for making and using and uses of such mutated or modified Cas or CRISPR enzyme or Cas9 and systems or complexes containing the same and products from such methods and uses are also disclosed and claimed.

WO 2016/205613 A1

WO 2016/205613 A1

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

Published:

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

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

— *with sequence listing part of description (Rule 5.2(a))*

DEMANDE OU BREVET VOLUMINEUX

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

CECI EST LE TOME 1 DE 5
CONTENANT LES PAGES 1 À 157

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 5
CONTAINING PAGES 1 TO 157

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

NOM DU FICHIER / FILE NAME :

NOTE POUR LE TOME / VOLUME NOTE:

CRISPR ENZYME MUTATIONS REDUCING OFF-TARGET EFFECTS

CROSS REFERENCE / INCORPORATION BY REFERENCE

[0001] Reference is made to U.S. provisional application Ser. No. 62/181,453, filed on June 18, 2015, U.S. provisional application Ser. No. 62/207,312, filed August 19, 2015, U.S. provisional application Ser. No. 62/237,360, filed October 5, 2015, U.S. provisional application Ser. No. 62/255,256, filed November 13, 2015 and U.S. Provisional application Ser. No. 62/269,876, filed December 18, 2015.

[0002] The foregoing application(s) and all documents cited or referenced herein ("herein cited documents"), and all documents cited or referenced in herein cited documents, together with any manufacturer's instructions, descriptions, product specifications, and product sheets for any products mentioned herein or in any document incorporated by reference herein, are hereby incorporated herein by reference, and may be employed in the practice of the invention. More specifically, all referenced documents are incorporated by reference to the same extent as if each individual document was specifically and individually indicated to be incorporated by reference.

STATEMENT AS TO FEDERALLY SPONSORED RESEARCH

[0003] This invention was made with government support under grant number MH100706 and MH110049 awarded by the National Institutes of Health. The government has certain rights in the invention.

FIELD OF THE INVENTION

[0004] The present invention generally relates to Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR), CRISPR enzyme (e.g., Cas or Cas9), CRISPR-Cas or CRISPR system or CRISPR-Cas complex, components thereof, nucleic acid molecules, e.g., vectors, involving the same and uses of all of the foregoing, amongst other aspects.

BACKGROUND OF THE INVENTION

[0005] The first publication of an enabling disclosure of how to make and use a CRISPR-Cas system in eukaryotic cells is Cong et al., Science 2013; 339:819-823 (published online 3 January 2013). The first patent filing of an enabling disclosure of how to make and use a CRISPR-Cas system in eukaryotic cells is Zhang et al., US Provisional application Serial No. 61/736,527, filed 12 December 2012, from which many patent applications claim priority, including those that have matured into seminal US Patents Nos. 8,999,641, 8,993,233, 8,945,839, 8,932,814,

8,906,616, 8,895,308, 8,889,418, 8,889,356, 8,871,445, 8,865,406, 8,795,965, 8,771,945 and 8,697,359.

SUMMARY OF THE INVENTION

[0006] Consistent with providing the breakthrough advances that enabled use of the CRISPR-Cas system in eukaryotic cells, the Zhang et al. laboratory of the Broad Institute recognized there remains a need for improved CRISPR enzymes for use in effecting modifications to target loci but which reduce or eliminate activity towards off-targets. There exists a pressing need for alternative and robust systems and techniques for reducing off-target activity of CRISPR enzymes when in complexed with guide RNAs. There also exists a pressing need for alternative and robust systems and techniques for increasing the activity of CRISPR enzymes when complexed with guide RNAs.

[0007] Several strategies to enhance Cas9 specificity have been developed, including reducing the amount of Cas9 in the cell, using Cas9 nickase mutants to create a pair of juxtaposed single-stranded DNA nicks, truncating the guide sequence at the 5' end, and using a pair of catalytically-inactive Cas9 nucleases, each fused to a FokI nuclease domain.

[0008] The inventors have surprisingly determined that modifications may be made to CRISPR enzymes which confer reduced off-target activity compared to unmodified CRISPR enzymes and/or increased target activity compared to unmodified CRISPR enzymes. Thus, provided herein are improved CRISPR enzymes which may have utility in a wide range of gene modifying applications. Also provided herein are CRISPR complexes, compositions and systems, as well as methods and uses, all comprising the herein disclosed modified CRISPR enzymes. CRISPR-Cas9 is preferred, including without limitation, SaCas9, SpCas9, and orthologs.

[0009] In an aspect, there is provided an engineered CRISPR protein, wherein the protein complexes with a nucleic acid molecule comprising RNA to form a CRISPR complex, wherein when in the CRISPR complex, the nucleic acid molecule targets one or more target polynucleotide loci, the protein comprises at least one modification compared to unmodified CRISPR, and wherein the CRISPR complex comprising the modified protein has altered activity as compared to the complex comprising the unmodified CRISPR protein. CRISPR-Cas9 is preferred, including without limitation, SaCas9, SpCas9, and orthologs. CRISPR proteins include those with enzymatic activity, for example nuclease activity.

[0010] In an aspect, the altered activity of the engineered CRISPR protein comprises an altered binding property as to the nucleic acid molecule comprising RNA or the target polynucleotide loci, altered binding kinetics as to the nucleic acid molecule comprising RNA or the target polynucleotide loci, or altered binding specificity as to the nucleic acid molecule comprising RNA or the target polynucleotide loci compared to off-target polynucleotide loci.

[0011] In certain embodiments, the altered activity of the engineered CRISPR protein comprises increased targeting efficiency or decreased off-target binding. In certain embodiments, the altered activity of the engineered CRISPR protein comprises modified cleavage activity.

[0012] In certain embodiments, the altered activity comprises increased cleavage activity as to the target polynucleotide loci. In certain embodiments, the altered activity comprises decreased cleavage activity as to the target polynucleotide loci. In certain embodiments, the altered activity comprises decreased cleavage activity as to off-target polynucleotide loci. In certain embodiments, the altered activity comprises increased cleavage activity as to off-target polynucleotide loci. Accordingly, in certain embodiments, there is increased specificity for target polynuycleotide loci as compared to off-target polynucleotide loci. In other embodiments, there is reduced specificity for target polynuycleotide loci as compared to off-target polynucleotide loci.

[0013] In an aspect of the invention, the altered activity of the engineered CRISPR protein comprises altered helicase kinetics.

[0014] In an aspect of the invention, the engineered CRISPR protein comprises a modification that alters association of the protein with the nucleic acid molecule comprising RNA, or a strand of the target polynucleotide loci, or a strand of off-target polynucleotide loci. In an aspect of the invention, the engineered CRISPR protein comprises a modification that alters formation of the CRISPR complex.

[0015] The present invention provides:

a non-naturally-occurring CRISPR enzyme, wherein:

the enzyme complexes with guide RNA to form a CRISPR complex,

when in the CRISPR complex, the guide RNA targets one or more target polynucleotide loci and the enzyme alters the polynucleotide loci, and

the enzyme comprises at least one modification,

whereby the enzyme in the CRISPR complex has reduced capability of modifying one or more off-target loci as compared to an unmodified enzyme, and/or whereby the enzyme in the CRISPR complex has increased capability of modifying the one or more target loci as compared to an unmodified enzyme.

[0016] In any such non-naturally-occurring CRISPR enzyme, the modification may comprise modification of one or more amino acid residues of the enzyme.

[0017] In any such non-naturally-occurring CRISPR enzyme, the modification may comprise modification of one or more amino acid residues located in a region which comprises residues which are positively charged in the unmodified enzyme.

[0018] In any such non-naturally-occurring CRISPR enzyme, the modification may comprise modification of one or more amino acid residues which are positively charged in the unmodified enzyme.

[0019] In any such non-naturally-occurring CRISPR enzyme, the modification may comprise modification of one or more amino acid residues which are not positively charged in the unmodified enzyme.

[0020] The modification may comprise modification of one or more amino acid residues which are uncharged in the unmodified enzyme.

[0021] The modification may comprise modification of one or more amino acid residues which are negatively charged in the unmodified enzyme.

[0022] The modification may comprise modification of one or more amino acid residues which are are hydrophobic in the unmodified enzyme.

[0023] The modification may comprise modification of one or more amino acid residues which are polar in the unmodified enzyme.

[0024] In any of the above-described non-naturally-occurring CRISPR enzymes, the enzyme may comprise a TypeII CRISPR enzyme. The enzyme may comprise a Cas9 enzyme.

[0025] In certain of the above-described non-naturally-occurring CRISPR enzymes, the modification may comprise modification of one or more residues located in a region between a RuvC domain and the HNH domain. The RuvC domain may comprise the RuvCII domain or the RuvCIII domain. The modification may comprise modification of one or more residues located in a groove.

[0026] In certain of the above-described non-naturally-occurring CRISPR enzymes, the modification may comprise modification of one or more residues located outside of a region between a RuvC domain and the HNH domain, or outside of a groove.

[0027] In certain of the above-described non-naturally-occurring CRISPR enzymes, the modification may comprise modification of one or more residues in a region which comprises:

the residues R63 to K1325 or K775 to K1325 of *Streptococcus pyogenes* Cas9 (SpCas9) or a corresponding region in another Cas9 ortholog; or

the residues K37 to K736 of *Staphylococcus aureus* Cas9 (SaCas9) or a corresponding region in another Cas9 ortholog.

[0028] In certain of the above-described non-naturally-occurring CRISPR enzymes, the modification comprises a modification of one or more residues wherein the one or more residues comprises arginine, histidine or lysine.

[0029] In any of the above-described non-naturally-occurring CRISPR enzymes, the enzyme may be modified by mutation of said one or more residues.

[0030] In certain of the above-described non-naturally-occurring CRISPR enzymes, the enzyme is modified by mutation of said one or more residues, and wherein the mutation comprises substitution of a residue in the unmodified enzyme with an alanine residue.

[0031] In certain of the above-described non-naturally-occurring CRISPR enzymes, the enzyme is modified by mutation of said one or more residues, and wherein the mutation comprises substitution of a residue in the unmodified enzyme with aspartic acid or glutamic acid.

[0032] In certain of the above-described non-naturally-occurring CRISPR enzymes, the enzyme is modified by mutation of said one or more residues, and wherein the mutation comprises substitution of a residue in the unmodified enzyme with serine, threonine, asparagine or glutamine.

[0033] In certain of the above-described non-naturally-occurring CRISPR enzymes, the enzyme is modified by mutation of said one or more residues, and wherein the mutation comprises substitution of a residue in the unmodified enzyme with alanine, glycine, isoleucine, leucine, methionine, phenylalanine, tryptophan, tyrosine or valine.

[0034] In certain of the above-described non-naturally-occurring CRISPR enzymes, the enzyme is modified by mutation of said one or more residues, and wherein the mutation comprises substitution of a residue in the unmodified enzyme with a polar amino acid residue.

[0035] In certain of the above-described non-naturally-occurring CRISPR enzymes, the enzyme is modified by mutation of said one or more residues, and wherein the mutation comprises substitution of a residue in the unmodified enzyme with an amino acid residue which is not a polar amino acid residue.

[0036] In certain of the above-described non-naturally-occurring CRISPR enzymes, the enzyme is modified by mutation of said one or more residues, and wherein the mutation comprises substitution of a residue in the unmodified enzyme with a negatively charged amino acid residue.

[0037] In certain of the above-described non-naturally-occurring CRISPR enzymes, the enzyme is modified by mutation of said one or more residues, and wherein the mutation comprises substitution of a residue in the unmodified enzyme with an amino acid residue which is not a negatively charged amino acid residue.

[0038] In certain of the above-described non-naturally-occurring CRISPR enzymes, the enzyme is modified by mutation of said one or more residues, and wherein the mutation comprises substitution of a residue in the unmodified enzyme with an uncharged amino acid residue

[0039] In certain of the above-described non-naturally-occurring CRISPR enzymes, the enzyme is modified by mutation of said one or more residues, and wherein the mutation comprises substitution of a residue in the unmodified enzyme with with an amino acid residue which is not an uncharged amino acid residue.

[0040] In certain of the above-described non-naturally-occurring CRISPR enzymes, the enzyme is modified by mutation of said one or more residues, and wherein the mutation comprises substitution of a residue in the unmodified enzyme with a hydrophobic amino acid residue

[0041] In certain of the above-described non-naturally-occurring CRISPR enzymes, the enzyme is modified by mutation of said one or more residues, and wherein the mutation comprises substitution of a residue in the unmodified enzyme with an amino acid residue which is not a hydrophobic amino acid residue.

[0042] The non-naturally-occurring CRISPR enzyme may be SpCas9 or an ortholog of SpCas9, and wherein:

the enzyme is modified by or comprises modification, e.g., comprises, consists essentially of or consists of modification by mutation of any one of the SpCas9 or SaCas9 residues listed in any one of Tables 1-7 or a corresponding residue in the Cas9 ortholog; or

the enzyme comprises, consists essentially of or consists of modification in any one (single), two (double), three (triple), four (quadruple) or more position(s) in accordance with the disclosure throughout this application, including without limitation in this Summary and/or in the Brief Description of Drawings and/or in the Detailed Description and/or in any of the Examples and/or in any of the Figures, or a corresponding residue or position in the Cas9 ortholog, e.g., an enzyme comprising, consisting essentially of or consisting of modification in any one of the Cas9 residues recited in any of this Summary and/or in the Brief Description of Drawings and/or in the Detailed Description and/or in any of the Examples and/or in any of the Figures or elsewhere herein, or a corresponding residue or position in the Cas9 ortholog. In such an enzyme, each residue may be modified by substitution with an alanine residue.

[0043] In certain of the above-described non-naturally-occurring CRISPR enzymes, the enzyme is modified by mutation of one or more residues including but not limited positions 12, 13, 63, 415, 610, 775, 779, 780, 810, 832, 848, 855, 861, 862, 866, 961, 968, 974, 976, 982, 983, 1000, 1003, 1014, 1047, 1060, 1107, 1108, 1109, 1114, 1129, 1240, 1289, 1296, 1297, 1300, 1311, and 1325 with reference to amino acid position numbering of SpCas9.

[0044] In certain of the above-described non-naturally-occurring CRISPR enzymes, the enzyme is modified by mutation and comprises one or more alanine substitutions at residues including but not limited positions 63, 415, 775, 779, 780, 810, 832, 848, 855, 861, 862, 866, 961, 968, 974, 976, 982, 983, 1000, 1003, 1014, 1047, 1060, 1107, 1108, 1109, 1114, 1129, 1240, 1289, 1296, 1297, 1300, 1311, or 1325 with reference to amino acid position numbering of SpCas9.

[0045] In certain of the above-described non-naturally-occurring CRISPR enzymes, the enzyme is modified by mutation and comprises one or more substitions of K775A, E779L, Q807A, R780A, K810A, R832A, K848A, K855A, K862A, K866A, K961A, K968A, K974A, R976A, H982A, H983A, K1000A, K1014A, K1047A, K1060A, K1003A, K1107A, S1109A, H1240A, K1289A, K1296A, H1297A, K1300A, H1311A, or K1325A.

[0046] In certain of the above-described non-naturally-occurring CRISPR enzymes, the enzyme is modified by mutation and comprises two or more substitutions, wherein the two or

more substitions include without limitation R783A and A1322T, or R780A and K810A, or ER780A and K855A, or R780A and R976A, or K848A and R976A, or K855A and R976A, and R780A and K848A, or K810A and K848A, or K848A and K855A, or K810A and K855A, or H982A and R1060A, or H982A and R1003A, or K1003A and R1060A, or R780A and H982A, or K810A and H982A, or K848A and H982A, or K855A and H982A, or R780A and K1003A, or K810A and R1003A, or K848A and K1003A, or K848A and K1007A, or R780A and R1060A, or K810A and R1060A, or K848A and R1060A, or R780A and R1114A, or K848A and R1114A, or R63A and K855A, or R63A and H982A, or H415A and R780A, or H415A and K848A, or K848A and E1108A, or K810A and K1003A, or R780A and R1060A, K810A and R1060A, or K848A and R1060A.

[0047] In certain of the above-described non-naturally-occurring CRISPR enzymes, the enzyme is modified by mutation and comprises three or more substitutions, wherein the three or more substitions include without limitation H982A, K1003A, and K1129E, or R780A, K1003A, and R1060A, or K810A, K1003A, and R1060A, or K848A, K1003A, and R1060A, or K855A, K1003A, and R1060A, or H982A, K1003A, and R1060A, or R63A, K848A, and R1060A, or T13I, R63A, and K810A, or G12D, R63A, and R1060A.

[0048] In certain of the above-described non-naturally-occurring CRISPR enzymes, the enzyme is modified by mutation and comprises four or more substitutions, wherein the four or more substitions include without limitation R63A, E610G, K855A, and R1060A, or R63A, K855A, R1060A, and E610G.

[0049]

[0050] In one preferred embodiment, the mutation in the non-naturally-occurring CRISPR enzyme is not a mutation listed in Table B. In a further preferred embodiment, the mutation in the non-naturally-occurring CRISPR enzyme is not R63A, K866A, H982A, H983A, K1107A, K1107A, KES1107-1109AG or KES1107-1109GG with reference to amino acid position numbering of SpCas9. In a further preferred embodiment, the non-naturally-occurring CRISPR enzyme is not an enzyme modified by a single mutation selected from R63A, K866A, H982A, H983A, K1107A and K1107A or an enzyme modified by a mutation selected from KES1107-1109AG and KES1107-1109GG with reference to amino acid position numbering of SpCas9.

[0051] In a preferred embodiment the above-described non-naturally-occurring CRISPR enzyme is modified by mutation of one or more residues including but not limited positions 12,

13, 415, 610, 775, 779, 780, 810, 832, 848, 855, 861, 862, 961, 968, 974, 976, 1000, 1003, 1014, 1047, 1060, 1114, 1129, 1240, 1289, 1296, 1297, 1300, 1311, and 1325 with reference to amino acid position numbering of SpCas9.

[0052] In a further preferred embodiment the above-described non-naturally-occurring CRISPR enzyme is modified by mutation and comprises one or more alanine substitutions at residues including but not limited positions 415, 775, 779, 780, 810, 832, 848, 855, 861, 862, 961, 968, 974, 976, 1000, 1003, 1014, 1047, 1060, 1114, 1129, 1240, 1289, 1296, 1297, 1300, 1311, or 1325 with reference to amino acid position numbering of SpCas9.

[0053] In a further preferred embodiment the above-described non-naturally-occurring CRISPR enzyme is modified by mutation and comprises one or more substitions of K775A, E779L, Q807A, R780A, K810A, R832A, K848A, K855A, K862A, K961A, K968A, K974A, R976A, K1000A, K1014A, K1047A, K1060A, K1003A, S1109A, H1240A, K1289A, K1296A, H1297A, K1300A, H1311A, or K1325A.

[0054] In any of the non-naturally-occurring CRISPR enzymes:

a single mismatch may exist between the target and a corresponding sequence of the one or more off-target loci; and/or

two, three or four or more mismatches may exist between the target and a corresponding sequence of the one or more off-target loci, and/or

wherein in (ii) said two, three or four or more mismatches are contiguous.

[0055] In any of the non-naturally-occurring CRISPR enzymes the enzyme in the CRISPR complex may have reduced capability of modifying one or more off-target loci as compared to an unmodified enzyme and wherein the enzyme in the CRISPR complex has increased capability of modifying the said target loci as compared to an unmodified enzyme.

[0056] In any of the non-naturally-occurring CRISPR enzymes, when in the CRISPR complex the relative difference of the modifying capability of the enzyme as between target and at least one off-target locus may be increased compared to the relative difference of an unmodified enzyme.

[0057] In any of the non-naturally-occurring CRISPR enzymes, the CRISPR enzyme may comprise one or more additional mutations, wherein the one or more additional mutations are in one or more catalytically active domains.

[0058] In such non-naturally-occurring CRISPR enzymes, the CRISPR enzyme may have reduced or abolished nuclease activity compared with an enzyme lacking said one or more additional mutations.

[0059] In some such non-naturally-occurring CRISPR enzymes, the CRISPR enzyme does not direct cleavage of one or other DNA strand at the location of the target sequence.

[0060] In some such non-naturally-occurring CRISPR enzymes, the one or more additional mutations comprise mutation of D10 of SpCas9, E762 of SpCas9, H840 of SpCas9, N854 of SpCas9, N863 of SpCas9 and/or D986 of SpCas9 or corresponding residues of other Cas9 orthologs.

[0061] In some such non-naturally-occurring CRISPR enzymes, the one or more additional mutations comprise D10A, E762A, H840A, N854A, N863A and/or D986A of SpCas9 or corresponding residues of other Cas9 orthologs.

[0062] In some such non-naturally-occurring CRISPR enzymes, the one or more additional mutations comprise two additional mutations. The two additional mutations may comprise D10A SpCas9 and H840A SpCas9, or corresponding residues of another Cas9 ortholog. In some such non-naturally-occurring CRISPR enzymes, the CRISPR enzyme may not direct cleavage of either DNA strand at the location of the target sequence.

[0063] Where the CRISPR enzyme comprises one or more additional mutations in one or more catalytically active domains, the one or more additional mutations may be in a catalytically active domain of the CRISPR enzyme comprising RuvCI, RuvCII or RuvCIII.

[0064] Without being bound by theory, in an aspect of the invention, the methods and mutations described provide for enhancing conformational rearrangement of Cas9 domains to positions that results in cleavage at on-target sits and avoidance of those conformational states at off-target sites. Cas9 cleaves target DNA in a series of coordinated steps. First, the PAM-interacting domain recognizes the PAM sequence 5' of the target DNA. After PAM binding, the first 10-12 nucleotides of the target sequence (seed sequence) are sampled for sgRNA:DNA complementarity, a process dependent on DNA duplex separation. If the seed sequence nucleotides complement the sgRNA, the remainder of DNA is unwound and the full length of sgRNA hybridizes with the target DNA strand. The nt-groove between the RuvC and HNH domains stabilizes the non-targeted DNA strand and facilitates unwinding through non-specific interactions with positive charges of the DNA phosphate backbone. RNA:cDNA and

Cas9:ncDNA interactions drive DNA unwinding in competition against cDNA:ncDNA rehybridization. Other cas9 domains affect the conformation of nuclease domains as well, for example linkers connecting HNH with RuvCII and RuvCIII. Accordingly, the methods and mutations provided encompass, without limitation, RuvCI, RuvCIII, RuvCIII and HNH domains and linkers. Conformational changes in Cas9 brought about by target DNA binding, including seed sequence interaction, and interactions with the target and non-target DNA strand determine whether the domains are positioned to trigger nuclease activity. Thus, the mutations and methods provided herein demonstrate and enable modifications that go beyond PAM recognition and RNA-DNA base pairing.

[0065] In an aspect, the invention provides Cas9 nucleases that comprise an improved equilibrium towards conformations associated with cleavage activity when involved in on-target interactions and/or improved equilibrium away from conformations associated with cleavage activity when involved in off-target interactions. In one aspect, the invention provides Cas9 nucleases with improved proof-reading function, i.e. a Cas9 nuclease which adopts a conformation comprising nuclease activity at an on-target site, and which conformation has increased unfavorability at an off-target site. Sternberg et al., Nature 527(7576):110-3, doi: 10.1038/nature15544, published online 28 October 2015. Epub 2015 Oct 28, used Förster resonance energy transfer FRET) experiments to detect relative orientations of the Cas9 catalytic domains when associated with on- and off-target DNA.

[0066] The invention further provides methods and mutations for modulating nuclease activity and/or specificity using modified guide RNAs. As discussed, on-target nuclease activity can be increased or decreased. Also, off-target nuclease activity can be increased or decreased. Further, there can be increased or decreased specificity as to on-target activity vs. off-target activity. Modified guide RNAs include, without limitation, truncated guide RNAs, dead guide RNAs, chemically modified guide RNAs, guide RNAs associated with functional domains, modified guide RNAs comprising functional domains, modified guide RNAs comprising aptamers, modified guide RNAs comprising adapter proteins, and guide RNAs comprising added or modified loops.

[0067] In an aspect, the invention also provides methods and mutations for modulating Cas9 binding activity and/or binding specificity. In certain embodiments Cas9 proteins lacking nuclease activity are used. In certain embodiments, modified guide RNAs are employed that

promote binding but not nuclease activity of a Cas9 nuclease. In such embodiments, on-target binding can be increased or decreased. Also, in such embodiments off-target binding can be increased or decreased. Moreover, there can be increased or decreased specificity as to on-target binding vs. off-target binding.

[0068] The methods and mutations which can be employed in various combinations to increase or decrease activity and/or specificity of on-target vs. off-target activity, or increase or decrease binding and/or specificity of on-target vs. off-target binding, can be used to compensate or enhance mutations or modifications made to promote other effects. Such mutations or modifications made to promote other effects in include mutations or modification to the Cas9 and or mutation or modification made to a guide RNA. In certain embodiments, the methods and mutations are used with chemically modified guide RNAs. Examples of guide RNA chemical modifications include, without limitation, incorporation of 2′-O-methyl (M), 2′-O-methyl 3′phosphorothioate (MS), or 2′-O-methyl 3′thioPACE (MSP) at one or more terminal nucleotides. Such chemically modified guide RNAs can comprise increased stability and increased activity as compared to unmodified guide RNAs, though on-target vs. off-target specificity is not predictable. (See, Hendel, 2015, Nat Biotechnol. 33(9):985-9, doi: 10.1038/nbt.3290, published online 29 June 2015). Chemically modified guide RNAs futher include, without limitation, RNAs with phosphorothioate linkages and locked nucleic acid (LNA) nucleotides comprising a methylene bridge between the 2′ and 4′ carbons of the ribose ring. The methods and mutations of the invention are used to modulate Cas9 nuclease activity and/or binding with chemically modified guide RNAs.

[0069] In an aspect, the invention provides methods and mutations for modulating binding and/or binding specificity of Cas9 proteins comprising functional domains such as nucleases, transcriptional activators, transcriptional repressors, and the like. For example, a Cas9 protein can be made nuclease-null by introducing mutations such as D10A, D839A, H840A and N863A in nuclease domains RuvC and HNH. Nuclease deficient Cas9 proteins are useful for RNA-guided target sequence dependent delivery of functional domains. The invention provides methods and mutations for modulating binding of Cas9 proteins. In one embodiment, the functional domain comprises VP64, providing an RNA-guided transcription factor. In another embodiment, the functional domain comprises Fok I, providing an RNA-guided nuclease activity. Mention is made of U.S. Pat. Pub. 2014/0356959, U.S. Pat. Pub. 2014/0342456, U.S.

Pat. Pub. 2015/0031132, and Mali, P. et al., 2013, Science 339(6121):823-6, doi: 10.1126/science.1232033, published online 3 January 2013 and through the teachings herein the invention comprehends methods and materials of these documents applied in conjunction with the teachings herein. In certain embodiments, on-target binding is increased. In certain embodiments, off-target binding is decreased. In certain embodiments, on-target binding is decreased. In certain embodiments, off-target binding is increased. Accordingly, the invention also provides for increasing or decreasing specificity of on-target binding vs. off-target binding of functionalized Cas9 binding proteins.

[0070] The use of Cas9 as an RNA-guided binding protein is not limited to nuclease-null Cas9. Cas9 enzymes comprising nuclease activity can also function as RNA-guided binding proteins when used with certain guide RNAs. For example short guide RNAs and guide RNAs comprising nucleotides mismatched to the target can promote RNA directed Cas9 binding to a target sequence with little or no target cleavage. (See, e.g., Dahlman, 2015, Nat Biotechnol. 33(11):1159-1161, doi: 10.1038/nbt.3390, published online 05 October 2015). In an aspect, the invention provides methods and mutations for modulating binding of Cas9 proteins that comprise nuclease activity. In certain embodiments, on-target binding is increased. In certain embodiments, off-target binding is decreased. In certain embodiments, on-target binding is decreased. In certain embodiments, off-target binding is increased. In certain embodiments, there is increased or decreased specificity of on-target binding vs. off-target binding. In certain embodiments, nuclease activity of guide RNA-Cas9 enzyme is also modulated.

[0071] RNA–DNA heteroduplex formation is important for cleavage activity and specificity throughout the target region, not only the seed region sequence closest to the PAM. Thus, truncated guide RNAs show reduced cleavage activity and specificity. In an aspect, the invention provides method and mutations for increasing activity and specificity of cleavage using altered guide RNAs.

[0072] The invention also demonstrates that modifications of Cas9 nuclease specificity can be made in concert with modifications to targeting range. Cas9 mutants can be designed that have increased target specificity as well as accommodating modifications in PAM recognition, for example by choosing mutations that alter PAM specificity and combining those mutations with nt-groove mutations that increase (or if desired, decrease) specificity for on-target sequences vs. off-target sequences. In one such embodiment, a PI domain residue is mutated to

accommodate recognition of a desired PAM sequence while one or more nt-groove amino acids is mutated to alter target specificity. Kleinstiver involves SpCas9 and SaCas9 nucleases in which certain PI domain residues are mutated and recognize alternative PAM sequences (see Kleinstiver et al., Nature 523(7561):481-5 doi: 10.1038/nature14592, published online 22 June 2015; Kleinstiver et al., Nature Biotechnology, doi: 10.1038/nbt.3404, published online 2 November 2015). The Cas9 methods and modifications described herein can be used to counter loss of specificity resulting from alteration of PAM recognition, enhance gain of specificity resulting from alteration of PAM recognition, counter gain of specificity resulting from alteration of PAM recognition, or enhance loss of specificity resulting from alteration of PAM recognition.

[0073] The methods and mutations can be used with any Cas9 enzyme with altered PAM recognition. Non-limiting examples of PAMs included NGG, NNGRRT, NN[A/C/T]RRT, NGAN, NGCG, NGAG, NGNG, NGC, and NGA.

[0074] In further embodiments, the methods and mutations are used modified proteins.

[0075] In any of the non-naturally-occurring CRISPR enzymes, the CRISPR enzyme may comprise one or more heterologous functional domains.

[0076] The one or more heterologous functional domains may comprise one or more nuclear localization signal (NLS) domains. The one or more heterologous functional domains may comprise at least two or more NLSs.

[0077] In certain embodiments of the invention, at least one nuclear localization signal (NLS) is attached to the nucleic acid sequences encoding the Cas9 effector proteins. In preferred embodiments at least one or more C-terminal or N-terminal NLSs are attached (and hence nucleic acid molecule(s) coding for the the Cas9 effector protein can include coding for NLS(s) so that the expressed product has the NLS(s) attached or connected). In a preferred embodiment a C-terminal NLS is attached for optimal expression and nuclear targeting in eukaryotic cells, preferably human cells. In a preferred embodiment, the codon optimized effector protein is SpCas9 or SaCas9 and the spacer length of the guide RNA is from 15 to 35 nt. In certain embodiments, the spacer length of the guide RNA is at least 16 nucleotides, such as at least 17 nucleotides. In certain embodiments, the spacer length is from 15 to 17 nt, from 17 to 20 nt, from 20 to 24 nt, eg. 20, 21, 22, 23, or 24 nt, from 23 to 25 nt, e.g., 23, 24, or 25 nt, from 24 to 27 nt, from 27-30 nt, from 30-35 nt, or 35 nt or longer. In certain embodiments of the invention, the codon optimized effector protein is SpCas9 or SaCas9 and the direct repeat length of the guide

RNA is at least 16 nucleotides. In certain embodiments, the codon optimized effector protein is FnCpf1p and the direct repeat length of the guide RNA is from 16 to 20 nt, e.g., 16, 17, 18, 19, or 20 nucleotides. In certain preferred embodiments, the direct repeat length of the guide RNA is 19 nucleotides.

[0078] The one or more heterologous functional domains comprises one or more transcriptional activation domains. A transcriptional activation domain may comprise VP64.

[0079] The one or more heterologous functional domains comprises one or more transcriptional repression domains. A transcriptional repression domain may comprise a KRAB domain or a SID domain.

[0080] The one or more heterologous functional domain may comprise one or more nuclease domains. The one or more nuclease domains may comprise Fok1.

[0081] The one or more heterologous functional domains may have one or more of the following activities: methylase activity, demethylase activity, transcription activation activity, transcription repression activity, transcription release factor activity, histone modification activity, nuclease activity, single-strand RNA cleavage activity, double-strand RNA cleavage activity, single-strand DNA cleavage activity, double-strand DNA cleavage activity and nucleic acid binding activity.

[0082] The at least one or more heterologous functional domains may be at or near the amino-terminus of the enzyme and/or at or near the carboxy-terminus of the enzyme.

[0083] The one or more heterologous functional domains may be fused to the CRISPR enzyme, or tethered to the CRISPR enzyme, or linked to the CRISPR enzyme by a linker moiety.

[0084] In any of the non-naturally-occurring CRISPR enzymes, the CRISPR enzyme may comprise a CRISPR enzyme from an organism from a genus comprising *Streptococcus*, *Campylobacter*, *Nitratifractor*, *Staphylococcus*, *Parvibaculum*, *Roseburia*, *Neisseria*, *Gluconacetobacter*, *Azospirillum*, *Sphaerochaeta*, *Lactobacillus*, *Eubacterium* or *Corynebacter*.

[0085] In any of the non-naturally-occurring CRISPR enzymes, the CRISPR enzyme may comprise a chimeric Cas9 enzyme comprising a first fragment from a first Cas9 ortholog and a second fragment from a second Cas9 ortholog, and the first and second Cas9 orthologs are different. At least one of the first and second Cas9 orthologs may comprise a Cas9 from an organism comprising *Streptococcus*, *Campylobacter*, *Nitratifractor*, *Staphylococcus*,

Parvibaculum, *Roseburia*, *Neisseria*, *Gluconacetobacter*, *Azospirillum*, *Sphaerochaeta*, *Lactobacillus*, *Eubacterium* or *Corynebacter*.

[0086] In any of the non-naturally-occurring CRISPR enzymes, a nucleotide sequence encoding the CRISPR enzyme may be codon optimized for expression in a eukaryote.

[0087] In any of the non-naturally-occurring CRISPR enzymes, the cell may be a eukaryotic cell or a prokaryotic cell; wherein the CRISPR complex is operable in the cell, and whereby the enzyme of the CRISPR complex has reduced capability of modifying one or more off-target loci of the cell as compared to an unmodified enzyme and/or whereby the enzyme in the CRISPR complex has increased capability of modifying the one or more target loci as compared to an unmodified enzyme.

[0088] The invention also provides a non-naturally-occurring, engineered composition comprising a CRISPR-Cas complex comprising any the non-naturally-occurring CRISPR enzyme described above.

[0089] The invention also provides a non-naturally-occurring, engineered composition comprising:

a delivery system operably configured to deliver CRISPR-Cas complex components or one or more polynucleotide sequences comprising or encoding said components into a cell, and wherein said CRISPR-Cas complex is operable in the cell,

CRISPR-Cas complex components or one or more polynucleotide sequences encoding for transcription and/or translation in the cell the CRISPR-Cas complex components, comprising:

(I) the non-naturally-occurring CRISPR enzyme according to any one of the preceding claims;

(II) CRISPR-Cas complex RNA comprising:

the guide sequence,

a tracr mate sequence, and

a tracr sequence,

wherein:

in the cell:

the tracr mate sequence hybridizes to the tracr sequence;

the CRISPR complex is formed;

the guide RNA targets the target polynucleotide loci and the enzyme alters the polynucleotide loci, and

the enzyme in the CRISPR complex has reduced capability of modifying one or more off-target loci as compared to an unmodified enzyme and/or whereby the enzyme in the CRISPR complex has increased capability of modifying the one or more target loci as compared to an unmodified enzyme.

[0090] In any such compositions, the delivery system may comprise a yeast system, a lipofection system, a microinjection system, a biolistic system, virosomes, liposomes, immunoliposomes, polycations, lipid:nucleic acid conjugates or artificial virions.

[0091] In any such compositions, the delivery system may comprise a vector system comprising one or more vectors, and wherein component (II) comprises a first regulatory element operably linked to a polynucleotide sequence which comprises the guide sequence, the tracr mate sequence and the tracr sequence, and wherein component (I) comprises a second regulatory element operably linked to a polynucleotide sequence encoding the CRISPR enzyme. In such compositions the guide RNA or CRISPR-Cas complex RNA may comprise a chimeric RNA.

[0092] In any such compositions, the delivery system may comprise a vector system comprising one or more vectors, and wherein component (II) comprises a first regulatory element operably linked to the guide sequence and the tracr mate sequence, and a third regulatory element operably linked to the tracr sequence, and wherein component (I) comprises a second regulatory element operably linked to a polynucleotide sequence encoding the CRISPR enzyme.

[0093] In any such compositions, the composition may comprise more than one guide RNA, and each guide RNA has a different target whereby there is multiplexing.

[0094] In any such compositions, the polynucleotide sequence(s) may be on one vector.

[0095] The invention also provides an engineered, non-naturally occurring Clustered Regularly Interspersed Short Palindromic Repeats (CRISPR)-CRISPR associated (Cas) (CRISPR-Cas) vector system comprising one or more vectors comprising:

a) a first regulatory element operably linked to a nucleotide sequence encoding a non-naturally-occurring CRISPR enzyme of any one of the inventive constructs herein; and

b) a second regulatory element operably linked to one or more nucleotide sequences encoding one or more of the guide RNAs, the guide RNA comprising a guide sequence, a tracr sequence, and a tracr mate sequence,

wherein:

components (a) and (b) are located on same or different vectors,

the tracr mate sequence hybridizes to the tracr sequence;

the CRISPR complex is formed;

the guide RNA targets the target polynucleotide loci and the enzyme alters the polynucleotide loci, and

the enzyme in the CRISPR complex has reduced capability of modifying one or more off-target loci as compared to an unmodified enzyme and/or whereby the enzyme in the CRISPR complex has increased capability of modifying the one or more target loci as compared to an unmodified enzyme.

[0096] In such a system, component (II) may comprise a first regulatory element operably linked to a polynucleotide sequence which comprises the guide sequence, the tracr mate sequence and the tracr sequence, and wherein component (II) may comprise a second regulatory element operably linked to a polynucleotide sequence encoding the CRISPR enzyme. In such a system, the guide RNA may comprise a chimeric RNA.

[0097] In such a system, component (I) may comprise a first regulatory element operably linked to the guide sequence and the tracr mate sequence, and a third regulatory element operably linked to the tracr sequence, and wherein component (II) may comprise a second regulatory element operably linked to a polynucleotide sequence encoding the CRISPR enzyme. Such a system may comprise more than one guide RNA, and each guide RNA has a different target whereby there is multiplexing. Components (a) and (b) may be on the same vector.

[0098] In any such systems comprising vectors, the one or more vectors may comprise one or more viral vectors, such as one or more retrovirus, lentivirus, adenovirus, adeno-associated virus or herpes simplex virus.

[0099] In any such systems comprising regulatory elements, at least one of said regulatory elements may comprise a tissue-specific promoter. The tissue-specific promoter may direct expression in a mammalian blood cell, in a mammalian liver cell or in a mammalian eye.

[00100] In any of the above-described compositions or systems the tracr sequence may comprise one or more protein-interacting RNA aptamers. The one or more aptamers may be located in the tetraloop and/or stemloop 2 of the tracr sequence. The one or more aptamers may be capapble of binding MS2 bacteriophage coat protein.

[00101] In any of the above-described compositions or systems the tracr sequence may be 30 or more nucleotides in length.

[00102] In any of the above-described compositions or systems the cell may a eukaryotic cell or a prokaryotic cell; wherein the CRISPR complex is operable in the cell, and whereby the enzyme of the CRISPR complex has reduced capability of modifying one or more off-target loci of the cell as compared to an unmodified enzyme and/or whereby the enzyme in the CRISPR complex has increased capability of modifying the one or more target loci as compared to an unmodified enzyme.

[00103] The invention also provides a CRISPR complex of any of the above-described compositions or from any of the above-described systems.

[00104] The invention also provides an engineered CRISPR protein, complex, composition, system, vector, cell or cell line according to the invention for use in therapy.

[00105] The invention also provides a method of modifying a locus of interest in a cell comprising contacting the cell with any of the above-described compositions or any of the above-described systems, or wherein the cell comprises any of the above-described CRISPR complexs present within the cell. In such methods the cell may be a eukaryotic cell. In such methods, an organism may comprise the cell. In such methods the organism may not be a human or other animal. The invention also provides an engineered CRISPR protein, complex, composition, system, vector, cell or cell line according to the invention for use in modifying a locus of interest in a cell. Said modifying preferably comprises contacting the cell with any of the above-described compositions or any of the above-described systems. The invention also provides a use of an engineered CRISPR protein, complex, composition, system, vector, cell or cell line according to the invention in the preparation of a medicament for modifying a locus of interest in a cell.

[00106] Any such method may be *ex vivo* or *in vitro*.

[00107] Any such method, said modifying may comprise modulating gene expression. Said modulating gene expression may comprise activating gene expression and/or repressing gene expression.

[00108] The invention also provides an engineered CRISPR protein, complex, composition, system, vector, cell or cell line according to the invention for use in modifying a locus of interest in a cell. The invention also provides a use of an engineered CRISPR protein, complex, composition, system, vector, cell or cell line according to the invention in the preparation of a medicament for modifying a locus of interest in a cell. Said modifying preferably comprises contacting the cell with any of the above-described compositions or any of the above-described systems.The invention also provides a method of treating a disease, disorder or infection in an individual in need thereof comprising administering an effective amount of any of the compositions, systems or CRISPR complexes described above. The disease, disorder or infection may comprise a viral infection. The viral infection may be HBV.

[00109] The invention also provides an engineered CRISPR protein, complex, composition, system, vector, cell or cell line according to the invention for use in the treatment of a disease, disorder or infection in an individual in need thereof. The disease, disorder or infection may comprise a viral infection. The viral infection may be HBV. The invention also provides a use of an engineered CRISPR protein, complex, composition, system, vector, cell or cell line according to the invention in the preparation of a medicament for the treatment of a disease, disorder or infection in an individual in need thereof. The disease, disorder or infection may comprise a viral infection.

[00110] The invention also provides the use of any of the compositions, systems or CRISPR complexes described above for gene or genome editing.

[00111] The invention also provides any of the compositions, systems or CRISPR complexes described above for use as a therapeutic. The therapeutic may be for gene or genome editing, or gene therapy.

[00112] 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 of a HSC, e.g., wherein the genomic locus of interest is associated with a mutation associated with an aberrant protein expression or with a disease condition or state, comprising:

delivering to an HSC, e.g., via contacting an HSC with a particle containing, a non-naturally occurring or engineered composition comprising:

I. a CRISPR-Cas system chimeric RNA (chiRNA) polynucleotide sequence, comprising:

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

(b) a tracr mate sequence, and

(c) a tracr sequence, and

II. a CRISPR enzyme, optionally comprising at least one or more nuclear localization sequences,

wherein 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 to the target sequence, and (2) the tracr mate sequence that is hybridized to the tracr sequence; and

the method may optionally include also delivering a HDR template, e.g., via the particle contacting the HSC containing or contacting the HSC with another particle containing, the HDR template wherein the HDR template provides expression of a normal or less aberrant form of the protein; wherein "normal" is as to wild type, and "aberrant" can be a protein expression that gives rise to a condition or disease state; and

optionally the method may include isolating or obtaining HSC from the organism or non-human organism, optionally expanding the HSC population, performing contacting of the particle(s) with the HSC to obtain a modified HSC population, optionally expanding the population of modified HSCs, and optionally administering modified HSCs to the organism or non-human organism.

[00113] The invention also provides an engineered CRISPR protein, complex, composition, system, vector, cell or cell line according to the invention for use in modifying an organism or a non-human organism by manipulation of a target sequence in a genomic locus of interest of a HSC. Said modifying preferably comprises

delivering to an HSC, e.g., via contacting an HSC with a particle containing, a non-naturally occurring or engineered composition comprising:

I. a CRISPR-Cas system chimeric RNA (chiRNA) polynucleotide sequence, comprising:

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

(b) a tracr mate sequence, and

(c) a tracr sequence, and

II. a CRISPR enzyme, optionally comprising at least one or more nuclear localization sequences,

wherein 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 to the target sequence, and (2) the tracr mate sequence that is hybridized to the tracr sequence. Said modifying further optionally includes delivering a HDR template, e.g., via the particle contacting the HSC containing or contacting the HSC with another particle containing, the HDR template wherein the HDR template provides expression of a normal or less aberrant form of the protein; wherein "normal" is as to wild type, and "aberrant" can be a protein expression that gives rise to a condition or disease state. Said modifying further optionally includes isolating or obtaining HSC from the organism or non-human organism, optionally expanding the HSC population, performing contacting of the particle(s) with the HSC to obtain a modified HSC population, optionally expanding the population of modified HSCs, and optionally administering modified HSCs to the organism or non-human organism.

[00114] 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 of a HSC, e.g., wherein the genomic locus of interest is associated with a mutation associated with an aberrant protein expression or with a disease condition or state, comprising: delivering to an HSC, e.g., via contacting an HSC with a particle containing, a non-naturally occurring or engineered composition comprising: I. (a) a guide sequence capable of hybridizing to a target sequence in a HSC, and (b) at least one or more tracr mate sequences, II. a CRISPR enzyme optionally having one or more NLSs, and III. a polynucleotide sequence comprising a tracr sequence, wherein 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 to the target sequence, and (2) the tracr mate sequence that is hybridized to the tracr sequence; and

the method may optionally include also delivering a HDR template, e.g., via the particle contacting the HSC containing or contacting the HSC with another particle containing, the HDR template wherein the HDR template provides expression of a normal or less aberrant form of the protein; wherein "normal" is as to wild type, and "aberrant" can be a protein expression that gives rise to a condition or disease state; and

optionally the method may include isolating or obtaining HSC from the organism or non-human organism, optionally expanding the HSC population, performing contacting of the particle(s) with the HSC to obtain a modified HSC population, optionally expanding the population of modified HSCs, and optionally administering modified HSCs to the organism or non-human organism. The invention also provides an engineered CRISPR protein, complex, composition, system, vector, cell or cell line according to the invention for use in such modifying an organism or a non-human organism by manipulation of a target sequence in a genomic locus of interest of a HSC, e.g., wherein the genomic locus of interest is associated with a mutation associated with an aberrant protein expression or with a disease condition or state.

[00115] The delivery can be of one or more polynucleotides encoding any one or more or all of the CRISPR-complex, advantageously linked to one or more regulatory elements for in vivo expression, e.g. via particle(s), containing a vector containing the polynucleotide(s) operably linked to the regulatory element(s). Any or all of the polynucleotide sequence encoding a CRISPR enzyme, guide sequence, tracr mate sequence or tracr sequence, may be RNA. 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).

[00116] In certain embodiments the invention provides a method of modifying 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 of an HSC e.g., wherein the genomic locus of interest is associated with a mutation associated with an aberrant protein expression or with a disease condition or state, comprising delivering, e.g., via contacting of a non-naturally occurring or engineered composition with the HSC, wherein the composition comprises one or more particles comprising viral, plasmid or nucleic acid molecule vector(s) (e.g. RNA) operably encoding a composition for expression thereof, wherein the composition comprises: (A) 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 CRISPR complex to the target sequence, and wherein the CRISPR complex comprises the CRISPR enzyme complexed with (1) the guide sequence that is hybridized to the target sequence, and (2) the tracr mate sequence that is hybridized 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 to the target sequence, and (2) the tracr mate sequence that is hybridized to the tracr sequence; the method may optionally include also delivering a HDR template, e.g., via the particle contacting the HSC containing or contacting the HSC with another particle containing, the HDR template wherein the HDR template provides expression of a

normal or less aberrant form of the protein; wherein "normal" is as to wild type, and "aberrant" can be a protein expression that gives rise to a condition or disease state; and optionally the method may include isolating or obtaining HSC from the organism or non-human organism, optionally expanding the HSC population, performing contacting of the particle(s) with the HSC to obtain a modified HSC population, optionally expanding the population of modified HSCs, and optionally administering modified HSCs to the organism or non-human organism. 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. The invention also provides an engineered CRISPR protein, complex, composition, system, vector, cell or cell line according to the invention for use in such modifying 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 of an HSC e.g., wherein the genomic locus of interest is associated with a mutation associated with an aberrant protein expression or with a disease condition or state, comprising delivering, e.g., via contacting of a non-naturally occurring or engineered composition with the HSC.

[00117] 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 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. 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. And the invention is especially advantageous as to HSCs.

[00118] The invention in some embodiments comprehends a method of modifying an organism or a non-human organism 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 HSC e.g., wherein the genomic locus of interest is associated with a mutation associated with an aberrant protein expression or with a disease condition or state, comprising delivering, e.g., by contacting HSCs with particle(s) comprising a non-naturally occurring or engineered composition comprising :

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,

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

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; or

IV. expression product(s) of one or more of I. to III., e.g., the the first and the second tracr mate sequence, the CRISPR enzyme;

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 to the first target sequence, and (2) the first tracr mate sequence that is hybridized to the first tracr sequence, wherein the second

CRISPR complex comprises the CRISPR enzyme complexed with (1) the second guide sequence that is hybridized to the second target sequence, and (2) the second tracr mate sequence that is hybridized 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; and the mthod may optionally include also delivering a HDR template, e.g., via the particle contacting the HSC containing or contacting the HSC with another particle containing, the HDR template wherein the HDR template provides expression of a normal or less aberrant form of the protein; wherein "normal" is as to wild type, and "aberrant" can be a protein expression that gives rise to a condition or disease state; and optionally the method may include isolating or obtaining HSC from the organism or non-human organism, optionally expanding the HSC population, performing contacting of the particle(s) with the HSC to obtain a modified HSC population, optionally expanding the population of modified HSCs, and optionally administering modified HSCs to the organism or non-human organism. 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 encoding 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; but, it is advantageous that the delivery is via a particle. 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, with reference to SpCas9 are selected from the group consisting of D10A, E762A, H840A, N854A, N863A and D986A, e.g., a 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 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 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. The invention also provides an engineered CRISPR protein, complex, composition, system, vector, cell or cell line according to the invention for use in such modifying an organism or a non-human organism 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 HSC e.g., wherein the genomic locus of interest is associated with a mutation associated with an aberrant protein expression or with a disease condition or state, comprising delivering, e.g., by contacting HSCs with particle(s) comprising a non-naturally occurring or engineered composition.

[00119] The invention in some embodiments comprehends a method of modifying an organism or a non-human organism 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 HSC e.g., wherein the genomic locus of interest is associated with a mutation associated with an aberrant protein expression or with a disease condition or state, comprising delivering, e.g., by contacting HSCs with particle(s) comprising a non-naturally occurring or engineered composition comprising :

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,

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,

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

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

V. expression product(s) of one or more of I. to IV., e.g., the the first and the second tracr mate sequence, the CRISPR enzyme;

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 to the first target sequence, and (2) the tracr mate sequence that is hybridized to the tracr sequence, wherein the second CRISPR complex comprises the CRISPR enzyme complexed with (1) the second guide sequence that is hybridized to the second target sequence, and (2) the tracr mate sequence that is hybridized 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 the other strand near the second target sequence inducing a double strand break, thereby modifying the organism or the non-human organism; and the mthod may optionally include also delivering a HDR template, e.g., via the particle contacting the HSC containing or contacting the HSC with another particle containing, the HDR template wherein the HDR template provides expression of a normal or less aberrant form of the protein; wherein "normal" is as to wild type, and "aberrant" can be a protein expression that gives rise to a condition or disease state; and optionally the method may include isolating or obtaining HSC from the organism or non-human organism, optionally expanding the HSC population, performing contacting of the particle(s) with the HSC to obtain a modified HSC population, optionally expanding the population of modified HSCs, and optionally administering modified HSCs to the organism or non-human organism. The invention also provides an engineered CRISPR protein, complex, composition, system, vector, cell or cell line according to the invention for use in such modifying an organism or a non-human organism 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 HSC e.g., wherein the genomic locus of interest is associated with a mutation associated with an aberrant

protein expression or with a disease condition or state, comprising delivering, e.g., by contacting HSCs with particle(s) comprising a non-naturally occurring or engineered composition.

[00120] 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. 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 with reference to SpCas9 are selected from the group consisting of D10A, E762A, H840A, N854A, N863A and D986A; e.g., 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; but, particle delivery is advantageous.

[00121] 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' overhang is at least 26 base pairs, preferably at least 30 base pairs or more preferably 34-50 base pairs.

[00122] The invention in some embodiments comprehends a method of modifying a genomic locus of interest in HSC e.g., wherein the genomic locus of interest is associated with a mutation associated with an aberrant protein expression or with a disease condition or state, by introducing into the HSC, e.g., by contacting HSCs with particle(s) 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 in the HSC, whereby the guide RNAs target the DNA molecule and the Cas protein nicks each of the first strand and the second strand of the DNA molecule, whereby a target in the HSC is altered; and, wherein the Cas protein and the two guide RNAs do not naturally occur together and the mthod may optionally include also delivering a HDR template, e.g., via the particle contacting the HSC containing or contacting the HSC with another particle containing, the HDR template wherein the HDR template provides expression of a normal or less aberrant form of the protein; wherein "normal" is as to wild type, and "aberrant" can be a protein expression that gives rise to a condition or disease state; and optionally the method may include isolating or obtaining HSC from the organism or non-human organism, optionally expanding the HSC population, performing contacting of the particle(s) with the HSC to obtain a modified HSC population, optionally expanding the population of modified HSCs, and optionally administering modified HSCs to the organism or non-human organism. In preferred methods of the invention the Cas protein nicking each of the first strand and the second strand of the DNA molecule 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. 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 or SaCas9. In aspects of the invention the Cas protein has one or more mutations in respect of SpCas9 selected from the group consisting of D10A, E762A, H840A, N854A, N863A and D986A; e.g., a D10A mutation. Aspects of the invention relate to the expression of a 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.

[00123] The invention in some embodiments comprehends a method of modifying a genomic locus of interest in HSC e.g., wherein the genomic locus of interest is associated with a mutation associated with an aberrant protein expression or with a disease condition or state, by introducing into the HSC, e.g., by contacting HSCs with particle(s) 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 of the HSC, and

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

c) expression product(s) of a) or b),

wherein components (a) and (b) are located on same or different vectors of the system, whereby the guide RNAs target the DNA molecule of the HSC and the Cas protein nicks each of the first strand and the second strand of the DNA molecule of the HSC; and, wherein the Cas protein and the two guide RNAs do not naturally occur together; and the method may optionally include also delivering a HDR template, e.g., via the particle contacting the HSC containing or contacting the HSC with another particle containing, the HDR template wherein the HDR template provides expression of a normal or less aberrant form of the protein; wherein "normal" is as to wild type, and "aberrant" can be a protein expression that gives rise to a condition or disease state; and optionally the method may include isolating or obtaining HSC from the organism or non-human organism, optionally expanding the HSC population, performing contacting of the particle(s) with the HSC to obtain a modified HSC population, optionally expanding the population of modified HSCs, and optionally administering modified HSCs to the organism or non-human organism. 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 or SaCas9. In aspects of the invention the Cas protein has one or more mutations with reference to

SpCas9 selected from the group consisting of D10A, E762A, H840A, N854A, N863A and D986A; e.g., the D10A mutation. Aspects of the invention relate to the expression of a 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; and particles are preferred. In one aspect, the invention provides a method of modifying a target polynucleotide in a HSC. 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 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 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 or expression product(s) thereof, e.g., via particle(s), to said HSC, 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 HSC in a subject. In some embodiments, said modifying takes place in said HSC in a cell culture. In some embodiments, the method further comprises isolating said HSC from a subject prior to said modifying. In some embodiments, the method further comprises returning said HSC and/or cells derived therefrom to said subject.

[00124] In one aspect, the invention provides a method of generating a HSC 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 or expression product(s) thereof, e.g., via particle(s), into a HSC, 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 to the target sequence within the target polynucleotide, and (2) the tracr mate sequence that is hybridized to the tracr sequence, thereby generating a HSC comprising a mutated disease gene. 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 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. In some embodiments the modified HSC is administered to an animal to thereby generate an animal model.

[00125] In one aspect, the invention provides for methods of modifying a target polynucleotide in a HSC. Also provided is an engineered CRISPR protein, complex, composition, system, vector, cell or cell line according to the invention for use in modifying a target polynucleotide in a HSC. 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 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 other embodiments, this invention provides a method of modifying expression of a polynucleotide in a eukaryotic cell that arises from an HSC. The method comprises increasing or decreasing expression of a target polynucleotide by using a

CRISPR complex that binds to the polynucleotide in the HSC; advantageously the CRISPR complex is delivered via particle(s).

[00126] In some methods, a target polynucleotide can be inactivated to effect the modification of the expression in a HSC. 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.

[00127] In some embodiments the RNA of the CRISPR-Cas system, e.g., the guide or sgRNA, can be modified; for instance to include an aptamer or a functional domain. An aptamer is a synthetic oligonucleotide that binds to a specific target molecule; for instance a nucleic acid molecule that has been engineered through repeated rounds of in vitro selection or SELEX (systematic evolution of ligands by exponential enrichment) to bind to various molecular targets such as small molecules, proteins, nucleic acids, and even cells, tissues and organisms. Aptamers are useful in that they offer molecular recognition properties that rival that of antibodies. In addition to their discriminate recognition, aptamers offer advantages over antibodies including that they elicit little or no immunogenicity in therapeutic applications. Accordingly, in the practice of the invention, either or both of the enzyme or the RNA can include a functional domain.

[00128] 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 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. In some embodiments, the functional domain comprises nuclease activity. In one such embodiment, the functional domain comprises Fok1.

[00129] The invention also provides an *in vitro* or *ex vivo* cell comprising any of the modified CRISPR enzymes, compositions, systems or complexes described above, or from any of the methods described above. The cell may be a eukaryotic cell or a prokaryotic cell. The invention also provides progeny of such cells. The invention also provides a product of any such cell or of any such progeny, wherein the product is a product of the said one or more target loci as

modified by the modified CRISPR enzyme of the CRISPR complex. The product may be a peptide, polypeptide or protein. Some such products may be modified by the modified CRISPR enzyme of the CRISPR complex. In some such modified products, the product of the target locus is physically distinct from the product of the said target locus which has not been modified by the said modified CRISPR enzyme.

[00130] The invention also provides a polynucleotide molecule comprising a polynucleotide sequence encoding any of the non-naturally-occurring CRISPR enzymes described above.

[00131] Any such polynucleotide may further comprise one or more regulatory elements which are operably linked to the polynucleotide sequence encoding the non-naturally-occurring CRISPR enzyme.

[00132] In any such polynucleotide which comprises one or more regulatory elements, the one or more regulatory elements may be operably configured for expression of the non-naturally-occurring CRISPR enzyme in a eukaryotic cell. The eukaryotic cell may be a human cell. The eukaryotic cell may be a rodent cell, optionally a mouse cell. The eukaryotic cell may be a yeast cell. The eukaryotic cell may be a chinese hamster ovary (CHO) cell. The eukaryotic cell may be an insect cell.

[00133] In any such polynucleotide which comprises one or more regulatory elements, the one or more regulatory elements may be operably configured for expression of the non-naturally-occurring CRISPR enzyme in a prokaryotic cell.

[00134] In any such polynucleotide which comprises one or more regulatory elements, the one or more regulatory elements may operably configured for expression of the non-naturally-occurring CRISPR enzyme in an in vitro system.

[00135] The invention also provides an expression vector comprising any of the above-described polynucleotide molecules. The invention also provides such polynucleotide molecule(s), for instance such polynucleotide molecules operably configured to express the protein and/or the nucleic acid component(s), as well as such vector(s).

[00136] The invention further provides for a method of making muations to a Cas9 or a mutated or modified Cas9 that is an ortholog of SaCas9 and/or SpCas9 comprising ascertaining amino acid(s) in that ortholog may be in close proximity or may touch a nucleic acid molecule, e.g., DNA, RNA, sgRNA, etc., and/or amino acid(s) analogous or corresponding to herein-identified amino acid(s) in SaCas9 and/or SpCas9 for modification and/or mutation, and

synthesizing or preparing or expressing the ortholog comprising, consisting of or consisting essentially of modification(s) and/or mutation(s) or mutating as herein-discussed, e.g., modifying, e.g., changing or mutating, a neutral amino acid to a charged, e.g., positively charged, amino acid, e.g., from alanine to, e.g., lysine. The so modified ortholog can be used in CRISPR-Cas systems; and nucleic acid molecule(s) expressing it may be used in vector or other delivery systems that deliver molecules or or encoding CRISPR-Cas system components as herein-discussed.

[00137] The invention also provides an engineered CRISPR protein, complex, composition, system, vector, cell or cell line according to the invention for use in making muations to a Cas9 or a mutated or modified Cas9 that is an ortholog of SaCas9 and/or SpCas9 comprising ascertaining amino acid(s) in that ortholog may be in close proximity or may touch a nucleic acid molecule, e.g., DNA, RNA, sgRNA, and/or amino acid(s) analogous or corresponding to herein-identified amino acid(s) in SaCas9 and/or SpCas9 for modification and/or mutation, and synthesizing or preparing or expressing the ortholog comprising, consisting of or consisting essentially of modification(s) and/or mutation(s) or mutating as herein-discussed, e.g., modifying, e.g., changing or mutating, a neutral amino acid to a charged, e.g., positively charged, amino acid, e.g., from alanine to e.g., lysine.

[00138] In an aspect, the invention provides efficient on-target activity and minimizes off target activity. In an aspect, the invention provides efficient on-target cleavage by a CRISPR protein and minimizes off-target cleavage by the CRISPR protein. In an aspect, the invention provides guide specific binding of a CRISPR protein at a gene locus without DNA cleavage. In an aspect, the invention provides efficient guide directed on-target binding of a CRISPR protein at a gene locus and minimizes off-target binding of the CRISPR protein. Accordingly, in an aspect, the invention provides target-specific gene regulation. In an aspect, the invention provides guide specific binding of a CRISPR enzyme at a gene locus without DNA cleavage. Accordingly, in an aspect, the invention provides for cleavage at one gene locus and gene regulation at a different gene locus using a single CRISPR enzyme. In an aspect, the invention provides orthogonal activation and/or inhibition and/or cleavage of multiple targets using one or more CRISPR protein and/or enzyme.

[00139] In another aspect, the present invention provides for a method of functional screening of genes in a genome in a pool of cells ex vivo or in vivo comprising the administration or

expression of a library comprising a plurality of CRISPR-Cas system guide RNAs (sgRNAs) and wherein the screening further comprises use of a CRISPR enzyme, wherein the CRISPR complex is modified to comprise a heterologous functional domain. In an aspect the invention provides a method for screening a genome comprising the administration to a host or expression in a host in vivo of a library. In an aspect the invention provides a method as herein discussed further comprising an activator administered to the host or expressed in the host. In an aspect the invention provides a method as herein discussed wherein the activator is attached to a CRISPR protein. In an aspect the invention provides a method as herein discussed wherein the activator is attached to the N terminus or the C terminus of the CRISPR protein. In an aspect the invention provides a method as herein discussed wherein the activator is attached to a sgRNA loop. In an aspect the invention provides a method as herein discussed further comprising a repressor administered to the host or expressed in the host. In an aspect the invention provides a method as herein discussed, wherein the screening comprises affecting and detecting gene activation, gene inhibition, or cleavage in the locus.

[00140] In an aspect the invention provides a method as herein discussed, wherein the host is a eukaryotic cell. In an aspect the invention provides a method or use as herein discussed, wherein the host is a mammalian cell. In an aspect the invention provides a method as herein discussed, wherein the host is a non-human eukaryote cell. In an aspect the invention provides a method as herein discussed, wherein the non-human eukaryote cell is a non-human mammal cell. In an aspect the invention provides a method as herein discussed, wherein the non-human mammal cell may be including, but not limited to, primate bovine, ovine, procine, canine, rodent, Leporidae such as monkey, cow, sheep, pig, dog, rabbit, rat or mouse cell. In an aspect the invention provides a method or use as herein discuscussed, the cell may be a a non-mammalian eukaryotic cell such as poultry bird (e.g., chicken), vertebrate fish (e.g., salmon) or shellfish (e.g., oyster, claim, lobster, shrimp) cell. In an aspect the invention provides a method or use as herein discussed, the non-human eukaryote cell is a plant cell. The plant cell may be of a monocot or dicot or of a crop or grain plant such as cassava, corn, sorghum, soybean, wheat, oat or rice. The plant cell may also be of an algae, tree or production plant, fruit or vegetable (e.g., trees such as citrus trees, e.g., orange, grapefruit or lemon trees; peach or nectarine trees; apple or pear trees; nut trees such as almond or walnut or pistachio trees; nightshade plants; plants of the genus *Brassica*; plants of the genus *Lactuca*; plants of the genus *Spinacia*; plants of the genus

Capsicum; cotton, tobacco, asparagus, carrot, cabbage, broccoli, cauliflower, tomato, eggplant, pepper, lettuce, spinach, strawberry, blueberry, raspberry, blackberry, grape, coffee, cocoa, etc).

[00141] In an aspect the invention provides a method as herein discussed comprising the delivery of the CRISPR-Cas complexes or component(s) thereof or nucleic acid molecule(s) coding therefor, wherein said nucleic acid molecule(s) are operatively linked to regulatory sequence(s) and expressed in vivo. In an aspect the invention provides a method as herein discussed wherein the expressing in vivo is via a lentivirus, an adenovirus, or an AAV. In an aspect the invention provides a method as herein discussed wherein the delivery is via a particle, a nanoparticle, a lipid or a cell penetrating peptide (CPP).

[00142] In an aspect the invention provides a pair of CRISPR-Cas complexes, each comprising a guide RNA (sgRNA) comprising a guide sequence capable of hybridizing to a target sequence in a genomic locus of interest in a cell, wherein at least one loop of each sgRNA is modified by the insertion of distinct RNA sequence(s) that bind to one or more adaptor proteins, and wherein the adaptor protein is associated with one or more functional domains, wherein each sgRNA of each CRISPR-Cas comprises a functional domain having a DNA cleavage activity. In an aspect the invention provides a paired CRISPR-Cas complexes as herein-discussed, wherein the DNA cleavage activity is due to a Fok1 nuclease.

[00143] In an aspect the invention provides a method for cutting a target sequence in a genomic locus of interest comprising delivery to a cell of the CRISPR-Cas complexes or component(s) thereof or nucleic acid molecule(s) coding therefor, wherein said nucleic acid molecule(s) are operatively linked to regulatory sequence(s) and expressed in vivo. In an aspect the invention provides a method as herein-discussed wherein the delivery is via a lentivirus, an adenovirus, or an AAV. In an aspect the invention provides a method as herein-discussed or paired CRISPR-Cas complexes as herein-discussed wherein the target sequence for a first complex of the pair is on a first strand of double stranded DNA and the target sequence for a second complex of the pair is on a second strand of double stranded DNA. In an aspect the invention provides a method as herein-discussed or paired CRISPR-Cas complexes as herein-discussed wherein the target sequences of the first and second complexes are in proximity to each other such that the DNA is cut in a manner that facilitates homology directed repair. In an aspect a herein method can further include introducing into the cell template DNA. In an aspect a herein method or herein paired CRISPR-Cas complexes can involve wherein each CRISPR-Cas

complex has a CRISPR enzyme that is mutated such that it has no more than about 5% of the nuclease activity of the CRISPR enzyme that is not mutated.

[00144] In an aspect the invention provides a library, method or complex as herein-discussed wherein the sgRNA is modified to have at least one non-coding functional loop, e.g., wherein the at least one non-coding functional loop is repressive; for instance, wherein the at least one non-coding functional loop comprises Alu.

[00145] In one aspect, the invention provides a method for altering or modifying expression of a gene product. The said method may comprise introducing into a cell containing and expressing a DNA molecule encoding the gene product an engineered, non-naturally occurring CRISPR-Cas system comprising a Cas protein and guide RNA that targets the DNA molecule, whereby the guide RNA targets the DNA molecule encoding the gene product and the Cas protein cleaves the DNA molecule encoding the gene product, whereby expression of the gene product is altered; and, wherein the Cas protein and the guide RNA do not naturally occur together. The invention comprehends the guide RNA comprising a guide sequence fused to a tracr sequence. The invention further comprehends the Cas protein being codon optimized for expression in a Eukaryotic cell. In a preferred embodiment the Eukaryotic cell is a mammalian cell and in a more preferred embodiment the mammalian cell is a human cell. In a further embodiment of the invention, the expression of the gene product is decreased.

[00146] The invention also provides an engineered CRISPR protein, complex, composition, system, vector, cell or cell line as defined herein for use in altering the expression of a genomic locus of interest in a mammalian cell. The invention also provides a use of an engineered CRISPR protein, complex, composition, system, vector, cell or cell line for the preparation of a medicament for altering the expression of a genomic locus of interest in a mammalian cell. Said altering preferably comprises contacting the cell with an engineered CRISPR protein, complex, composition, system, vector, cell or cell line of the invention and thereby delivering a vector and allowing the CRISPR-Cas complex to form and bind to target. Said altering further preferably comprises determining if the expression of the genomic locus has been altered.

[00147] In an aspect, the invention provides altered cells and progeny of those cells, as well as products made by the cells. CRISPR-Cas9 proteins and systems of the invention are used to produce cells comprising a modified target locus. In some embodiments, the method may comprise allowing a nucleic acid-targeting complex to bind to the target DNA or RNA to effect

cleavage of said target DNA or RNA thereby modifying the target DNA or RNA, wherein the nucleic acid-targeting complex comprises a nucleic acid-targeting effector protein complexed with a guide RNA hybridized to a target sequence within said target DNA or RNA. In one aspect, the invention provides a method of repairing a genetic locus in a cell. In another aspect, the invention provides a method of modifying expression of DNA or RNA in a eukaryotic cell. In some embodiments, the method comprises allowing a nucleic acid-targeting complex to bind to the DNA or RNA such that said binding results in increased or decreased expression of said DNA or RNA; wherein the nucleic acid-targeting complex comprises a nucleic acid-targeting effector protein complexed with a guide RNA. Similar considerations and conditions apply as above for methods of modifying a target DNA or RNA. In fact, these sampling, culturing and re-introduction options apply across the aspects of the present invention. In an aspect, the invention provides for methods of modifying a target DNA or RNA in a eukaryotic cell, which may be *in vivo, ex vivo* or *in vitro*. In some embodiments, the method comprises sampling a cell or population of cells from a human or non-human animal, and modifying the cell or cells. Culturing may occur at any stage *ex vivo*. Such cells can be, without limitation, plant cells, animal cells, particular cell types of any organism, including stem cells, immune cells, T cell, B cells, dendritic cells, cardiovascular cells, epithelial cells, stem cells and the like. The cells can be modified according to the invention to produce gene products, for example in controlled amounts, which may be increased or decreased, depending on use, and/or mutated. In certain embodiments, a genetic locus of the cell is repaired. The cell or cells may even be re-introduced into the non-human animal or plant. For re-introduced cells it may be preferred that the cells are stem cells.

[00148] In an aspect, the invention provides cells which transiently comprise CRISPR systems, or components. For example, CRISPR proteins or enzymes and nucleic acids are transiently provided to a cell and a genetic locus is altered, followed by a decline in the amount of one or more components of the CRISPR system. Subsequently, the cells, progeny of the cells, and organisms which comprise the cells, having acquired a CRISPR mediated genetic alteration, comprise a diminished amount of one or more CRISPR system components, or no longer contain the one or more CRISPR system components. One non-limiting example is a self-inactivating CRISPR-Cas system such as further described herein. Thus, the invention provides cells, and organisms, and progeny of the cells and organisms which comprise one or more CRISPR-Cas

system-altered genetic loci, but essentially lack one or more CRISPR system component. In certain embodiments, the CRISPR system components are substantially absent. Such cells, tissues and organisms advantageously comprise a desired or selected genetic alteration but have lost CRISPR-Cas components or remnants thereof that potentially might act non-specifically, lead to questions of safety, or hinder regulatory approval. As well, the invention provides products made by the cells, organisms, and progeny of the cells and organisms.

[00149] The invention further provides a method of improving the specificity of a CRISPR system by providing an engineered CRISPR protein according to the invention. Preferably, an engineered CRISPR protein wherein

the protein complexes with a nucleic acid molecule comprising RNA to form a CRISPR complex,

wherein when in the CRISPR complex, the nucleic acid molecule targets one or more target polynucleotide loci,

the protein comprises at least one modification compared to the unmodified protein,

wherein the CRISPR complex comprising the modified protein has altered activity as compared to the complex comprising the unmodified protein. Said at least one modification is preferably in the RuvC and/or HNH domains as described herein or in the binding groove between the HNH and RuvC domains. Preferred modifications are mutations as described herein.

[00150] The invention further provides a use of an engineered CRISPR protein according to the invention to improve the specificity of a CRISPR system, Preferably an engineered CRISPR protein

wherein the protein complexes with a nucleic acid molecule comprising RNA to form a CRISPR complex,

wherein when in the CRISPR complex, the nucleic acid molecule targets one or more target polynucleotide loci,

wherein the protein is modified to comprise at least one modification compared to the unmodified protein,

[00151] wherein the CRISPR complex comprising the modified protein has altered activity as compared to the complex comprising the unmodified protein. Said at least one modification is

preferably in the RuvC and/or HNH domains as described herein or in the binding groove between the HNH and RuvC domains. Preferred modifications are mutations as described herein.

BRIEF DESCRIPTION OF THE DRAWINGS

[00152] 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:

[00153] Figure 1A-1B provides a schematic summary, with it understood that Applicant(s) / inventor(s) are not necessarily bound by any particular theory set forth herein or in any particular Figure, including Figure 1. The Figure discusses mutation of positively charged residues binding to the non-targeted gDNA strand whereby specificity is improved. Data in the Table of the schematic summary is as follows and is as to mutations of SpCas9:

Indel %

Cas9 mutant	ON Target (EMX1)	OFF Target 1(OT25)	OFF Target 2(OT46)
WT	**24.8**	**10.5**	**8.8**
R780	22.9	0.0	0.1
K810	23.3	0.1	0.1
K848	24.3	0.1	0.1
K855	25.1	0.2	0.3
R976	15.6	0.1	0.1
H982	20.9	0.5	0.4
K1003	24.8	4.1	2.8
R1060	20.4	1.3	1.8
GFP	0.1	0.0	0.1
untrans.	0.1	0.0	0.1

With reference to the numbering of SpCas9, Figure 1A illustrates Alanine mutations that improve specificity, distributed along the non-targeting strand groove, e.g., Arg780, Lys810, Lys855, Lys848, Lys1003, Arg1060, Arg976, His982. Without wishing to be bound by any one particular theory, the mechanism proposal is that nuclease activity is inactive until the non-targeted DNA strand sterically triggers HNH conformation change; non-targeted strand binding to the groove between HNH and RuvC depends on RNA:DNA pairing; mutating DNA binding residues in the groove places more energetic demand on proper RNA:DNA pairing (Figure 1B). Using the information herein, including in Figure 1, the skilled person can readily prepare

mutants of other Cas9s (e.g., other than SpCas9) that exhibit improved or reduced off-target effects. For instance, the documents cited herein provide information on numerous orthologs to SpCas9 and SaCas9 exemplified herein. From that information, including the sequence information of those other Cas9s, one skilled in the art can, from the information in this disclosure, readily prepare analogous mutants having reduced off-target effects in Cas9 orthologs in addition to SpCas9 and SaCas9 exemplified herein. Further, documents herein provide crystal structure information as to Cas9, e.., SpCas9; and one can readily make structural comparisons between crystal structures, e.g., between the crystal structure of SpCas9 and the crystal structure of an ortholog thereto, to also readily, without undue experimentation, obtain analogous mutants having reduced off-target effects in Cas9 orthologs in addition to SpCas9. Accordingly, the invention is broadly applicable to modification(s) or mutation(s) in various Cas9 orthologs to reduce off-target effects, including but not limited to SpCas9 and SaCas9. As discussed further herein, additional or further modification of the above-described Cas9 enzymes can readily be achieved whereby the enzyme in the CRISPR complex has increased capability of modifying the one or more target loci as compared to an unmodified enzyme.

[00154] Figure 2A shows activity of modified SpCas9 enzymes as measured by % INDEL formation. 49 point mutants of SpCas9 are depicted. The target sequence of EMX1.3 is a sequence of the EMX1 gene and activity is compared against a related off-target sequence (OT 46).

[00155] Figure 2B shows activity of modified SpCas9 enzymes as measured by % INDEL formation. 49 point mutants of SpCas9 are depicted. The target sequence is a sequence of the VEGFA gene and activity is compared against two related off-target sequences (OT 1 and OT 2).

[00156] Figure 2C shows activity of modified SpCas9 enzymes as measured by % INDEL formation. The target sequence is a sequence of the VEGFA gene and activity is compared against three related off-target sequences (OT 1, OT 4, and OT 18).

[00157] Figure 3 shows activity of modified SpCas9 enzymes as measured by % INDEL formation. Point mutants demonstrating specificity with respect to off-target sequences are depicted. The target sequences are sequences of the EMX1 and VEGFA genes and activity is compared against nine related off-target sequences.

[00158] Figure 4A shows activity of modified SpCas9 enzymes as measured by % INDEL formation. Double mutants of SpCas9 are depicted. The target sequence is a sequence of the EMX1 gene and activity is compared against two related off-target sequences.

[00159] Figure 4B shows activity of modified SpCas9 enzymes as measured by % INDEL formation. Double mutants of SpCas9 are depicted. The target sequence is a sequence of the VEGFA genes and activity is compared against three related off-target sequences.

[00160] Figure 5 shows activity of modified SpCas9 enzymes as measured by % INDEL formation. 14 triple mutants of SpCas9 are depicted. The target sequences are sequences of the EMX1 and VEGFA genes and activity is compared against four related off-target sequences (OT 46, OT 1, OT 4 and OT 18).

[00161] Figure 6 shows activity of modified SaCas9 enzymes as measured by % INDEL formation. The target sequence EMX101 is a sequence of the EMX1 gene and activity is compared against three related off-target sequences (OT1, OT2 and OT3).

[00162] Figure 7 shows activity of modified SaCas9 enzymes as measured by % INDEL formation. The target sequence of EMX101 is a sequence of EMX1and the activity is compared against a related off-target sequence (OT3).

[00163] Figure 8A-8D shows a phylogenetic tree of Cas genes; from the teachings herein and the knowledge in the art, mutation(s) or modification(s) of the exemplified SpCas9 and SaCas9 can be applied to other Cas9s.

[00164] Figure 9A-9F shows 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); from the teachings herein and the knowledge in the art, mutation(s) or modification(s) of the exemplified SpCas9 and SaCas9 can be applied to other Cas9s (and thus, the invention comprehends modification(s) or mutation(s) as herein exemplified as to SpCas9 and SaCas9 across Cas9s and the families and groups of Cas9s of Figure 9).

[00165] Figure 10A-10D shows activity of modified SpCas9 enzymes as measured by % INDEL formation. Figures 10A-C show activity for target sequences EMX101, EMX1.1, EMX1.2, EMX1.3, EMX1.8, EMX1.10, DNMT1.1, DNMT1.2, DNMT1.4, DNMT1.7, VEGFA4, VEGFA5, and VEGFA3. Figure 10D shows VEGFA3 activity compared against off-target sequence OT4.

[00166] Figure 11A-11D shows activity of modified SpCas9 enzymes as measured by % INDEL formation. Figures 11A-C show activity for target sequences EMX101, EMX1.1, EMX1.2, EMX1.3, EMX1.8, EMX1.10, DNMT1.1, DNMT1.2, DNMT1.4, DNMT1.7, VEGFA4, VEGFA5, and VEGFA3. Figuer 11D shows VEGFA3 activity compared against off-target sequence OT4.

[00167] Figure 12 shows activity of modified SpCas9 enxymes as measured by % INDEL formation. The target sequence is VEGFA3 and activity is compared against four related off-target sequences (OT1, OT2, OT4 and OT18).

[00168] Figure 13 shows activity of modified SpCas9 enxymes as measured by % INDEL formation. The target sequence is VEGFA3 and activity is compared against four related off-target sequences (OT1, OT2, OT4 and OT18).

[00169] Figure 14 shows activity of modified SpCas9 enzymes as measured by % INDEL formation. 14 point mutants of SpCas9 are depicted. The target sequence is EMX1.3 and activity is compared against five related off-target sequences (OT14, OT23, OE35, OT46, and OT53).

[00170] Figure 15A-15F shows structural aspects of SpCas9 and improved specificity. Panel A is a model of target unwinding. The nt-groove between the RuvC (teal) and HNH (magenta) domains stabilize DNA unwinding through non-specific DNA interactions with the non-complementary strand. RNA:cDNA and Cas9:ncDNA interactions drive DNA unwinding (top arrow) in competition against cDNA:ncDNA rehybridization (bottom arrow). Panel B: The structure of SpCas9 (PDB ID 4UN3) showing the nt-groove situated between the HNH (magenta) and RuvC (teal) domains. The non-target DNA strand (red) was manually modeled into the nt-groove (inset). Panel C: Screen of alanine point mutants for improvement in specificity. Panel D: Assessment of top point mutants at additional off-target loci. The top five specificity conferring mutants are highlighted in red. Panel E: Combination mutants improve specificity compared to single point mutants. eSpCas9(1.0) and eSpCas9(1.1) are highlighted in red. Panel F: Screen of top point mutants and combination mutants at 10 target loci for on-target cleavage efficiency. SpCas9(K855A), eSpCas9(1.0), and eSpCas9(1.1) are highlighted in red.

[00171] Figure 16A-16C shows maintenance of on-target efficiency by spCas9 mutants. Panel A shows an assessment of efficiency of on-target cutting of SpCas9 mutants as compared to SpCas9 for 24 sgRNAs targeted to 9 genomic loci. Panel B is a Tukey plot of normalized on-

target indel formation for mutants SpCas9(K855A), eSpCas9(1.0) and eSpCas9(1.1). Panel C is a Western blot of SpCas9 and mutants using anti-SpCas9 antibody.

[00172] Figure 17A-17C shows sensitivity of spCas9 and mutants K855A, eSpCas9(1.0), and eSpCas9(1.1) to single and double base mismatches between the guide RNA and target DNA. Panel A depicts mismatched guide sequences against a VEGFA target. Panel B provides heat maps for spCas9 and three mutants showing indel % with guide sequences having a single base mismatch. Panel C shows indel formation with guide sequences containing consecutive transversion mismatches. Compared to wild type: eSpCas9(1.0) comprises K810A, K1003A, R1060A; eSpCas9(1.1) comprises K848A, K1003A, R1060A.

[00173] Figure 18A-18F shows unbiased genome-wide off-target profiles of mutants SpCas9(K855A) and eSpCas9(1.1). Panel A is a Schematic outline of the BLESS (direct *in situ* breaks labelling, enrichment on streptavidin and next-generation sequencing) workflow. Panel B shows representative BLESS sequencing for forward (red) and reverse (blue) reads mapped to the genome. Reads mapping at Cas9 cut sites have distinct shape compared to DSB hotspots. Panels C and D show Manhattan plots of genome-wide DSB clusters generated by each SpCas9 mutant using the *EMX1*(1) (panel C) and *VEGFA*(1) (panel D) targeting guides. Panels E an F depict targeted deep sequencing validation of off-target sites identified in BLESS. Off-target sites are ordered by DSB score (blue heatmap). Green heatmaps indicates sequence similarity between target and off-target sequences.

[00174] Figure 19 shows a schematic of sgRNA guided targeting and DNA unwinding. Cas9 cleaves target DNA in a series of coordinated steps. First, the PAM-interacting domain recognizes an NGG sequence 5' of the target DNA. After PAM binding, the first 10-12 nucleotides of the target sequence (seed sequence) are sampled for sgRNA:DNA complementarity, a process dependent on DNA duplex separation. If the seed sequence nucleotides complement the sgRNA, the remainder of DNA is unwound and the full length of sgRNA hybridizes with the target DNA strand. In this model, the nt-groove between the RuvC (teal) and HNH (magenta) domains stabilizes the non-targeted DNA strand and facilitates unwinding through non-specific interactions with positive charges of the DNA phosphate backbone. RNA:cDNA and Cas9:ncDNA interactions drive DNA unwinding (top arrow) in competition against cDNA:ncDNA rehybridization (bottom arrow).

[00175] Figure 20 depicts electrostatics of SpCas9 reveal non-target strand groove. (A) Crystal structure (4UN3) of SpCas9 paired with sgRNA and target DNA colored by electrostatic potential to highlight positively charged regions. Scale is from -10 to 1 keV. (B) Identical to panel (A) with HNH domain removed to reveal the sgRNA:DNA heteroduplex. (C) Crystal structure (in the same orientation as (A)) colored by domain: HNH (magenta), RuvC (teal), and PAM-interacting (PI) (beige).

[00176] Figures 21A-21D show an off-target analysis of generated mutants. Twenty-nine SpCas9 point mutants were generated and tested for specificity at (A) an *EMX1* target site and (B) two *VEGFA* target sites. Mutants combining the top residues that improved specificity were further tested at (C) *EMX1* and (D) *VEGFA*.

[00177] Figure 22A-22C provides an annotated SpCas9 amino acid sequence. Mutations of SpCas9 that altered non-targeted strand groove charges were primarily in the RuvC and HNH domains (highlighted in yellow). RuvC (cyan), bridge helix (BH, green), REC (grey), HNH (magenta), and PI (beige) domains are annotated as in Nishmasu et al.

[00178] Figure 23 shows a comparison of the specificity of K855A, eSpCas9(1.0), and eSpCas9(1.1) with truncated sgRNAs and indicates SpCas9(1.0) and eSpCas9(1.1) outperform truncated sgRNAs as a strategy for improving specificity. Indel frequency at three loci (*EMX1*(1), *VEGFA*(1) and *VEGFA*(5)) were tested at major annotated and predicted off-target sites. For both VEGFA target sites, tru-sgRNA increased indel frequency at some off-target sites and generated indels at off-targets not observed in wild type. The number of off-target sites detectable by NGS each SpCas9 mutant are listed below the heat map.

[00179] Figure 24 shows increasing positive charge in the nt-groove can result in increased cleavage at off-target sites. Point mutants SpCas9(S845K) and SpCas9(L847R) exhibited less specificity than wild-type SpCas9 at the *EMX1*(1) target site.

[00180] Figures 25A-25D depict generation of eSaCas9 through mutagenesis of the nt-groove. An improved specificity version of SaCas9 was generated similarly to eSpCas9. (A,B) Single and double amino acid mutants of residues in the groove between the RuvC and HNH domains were screened for decreased off-target cutting. (C) Mutants with improved specificity were combined to make a variant of SaCas9 that maintained on-target cutting at EMX site 7 and had significantly reduced off-target cutting. (D) Crystal structure of SaCas9 showing the groove between the HNH and RuvC domains.

[00181] Figure 26 shows a characterization of on-target efficiency for certain specificity-enhancing mutants. Three SpCas9 mutants at the phosphate lock loop (Lys1107, Glu1108, Ser1109) in the PI domain confer specificity to bases 1 and 2 of the sgRNA proximal to the PAM. These consisted of a point mutant (K1107A) and two mutants in which the Lys-Glu-Ser sequence was replaced with the dipeptides Lys-Gly (KG) and Gly-Gly (GG), respectively. Our data indicated that these mutants can substantially reduce on-target cleavage efficiency.

[00182] Figure 27 shows eSpCas9(1.1) is not cytotoxic to human cells. HEK293T cells were transfected with WT or eSpCas9(1.1) and incubated for 72 hours before measuring cell survival using the CellTiter-Glo assay which fluoresces in response ATP production by live cells.

[00183] Figure 28 shows an analysis of Nt-groove mutants with truncated guide RNAs. Truncated guide RNAs (Tru) were combined with single amino acid SpCas9 mutants and targeted to (A) *EMX1*(1) or (B) *VEGFA*(1). While most mutants targeted to *EMX1* with an 18 nt guide retained on-target efficiency, those targeted to *VEGFA*(1) with a 17 nt guide were severely compromised.

[00184] Figures 29A-29B shows selected single and double amino acid mutants. As in SpCas9, reduction of positive charges in the non-targeting strand groove enhances specificity. Reduction of positive charges can be archived by substituting positive charged amino acids with neutral or negative charged amino acids (A) or by moving the position of the positive charged amino acid inside the groove (B). Mutants of interest are K572,

[00185] Figure 30 shows improved specificity of selected mutants. CM2 exhibits strong reduction in off target activity while retaining full on target activity. CM1: R499A; Q500K; K572A. CM2: R499A; Q500K; R654A; G655R. CM3: K572A; R654A; G655R.

[00186] Figure 31 shows activation of gamma globin HBG1 locus by complexes of spCas9 or spCas9 mutants guides of length 15bp, 17bp, and 20bp. Cas9 (px165) is unmutated spCas9. dCas9 indicates inactive spCas9. Depicted single mutants ("SM") are R780A, K810A, and K848A. Depicted double mutants ("DM") are R780A/K810A, and R780A/K855A.

[00187] Figure 32 shows comparison of different programmable nuclease platforms.

[00188] Figure 33 shows Types of Therapeutic Genome Modifications. The specific type of genome editing therapy depends on the nature of the mutation causing disease. **a**, In gene disruption, the pathogenic function of a protein is silenced by targeting the locus with NHEJ. Formation of indels on the gene of interest often result in frameshift mutations that create

premature stop codons and a non-functional protein product, or non-sense mediated decay of transcripts, suppressing gene function. **b**, HDR gene correction can be used to correct a deleterious mutation. DSB is targeted near the mutation site in the presence of an exogenously provided, corrective HDR template. HDR repair of the break site with the exogenous template corrects the mutation, restoring gene function. **c**, An alternative to gene correction is gene addition. This mode of treatment introduces a therapeutic transgene into a safe-harbor locus in the genome. DSB is targeted to the safe-harbor locus and an HDR template containing homology to the break site, a promoter and a transgene is introduced to the nucleus. HDR repair copies the promoter-transgene cassette into the safe-harbor locus, recovering gene function, albeit without true physiological control over gene expression.

[00189] Figure 34 shows *Ex vivo* vs. *in vivo* editing therapy. In *ex vivo* editing therapy cells are removed from a patient, edited and then re-engrafted (top panel). For this mode of therapy to be successful, target cells must be capable of survival outside the body and homing back to target tissues post-transplantation. In vivo therapy involves genome editing of cells *in situ* (bottom panels). For *in vivo* systemic therapy, delivery agents that are relatively agnostic to cell identity or state would be used to effect editing in a wide range of tissue types. Although this mode of editing therapy may be possible in the future, no delivery systems currently exist that are efficient enough to make this feasible. *In vivo* targeted therapy, where delivery agents with tropism for specific organ systems are administered to patients are feasible with clinically relevant viral vectors.

[00190] Figures 35A-35B show a schematic representation of gene therapy *via* Cas9 Homologous Recombination (HR) vectors.

[00191] Figure 36 presents a schematic of sugar attachments for directed delivery of protein or guide, especially with GalNac.

[00192] Figures 37A, 37B, and 37C together illustrate a sequence alignment of SaCas9 and SpCas9. RUVC and HNH domain annotations for the two proteins are also shown in the three figures.

[00193] The figures herein are for illustrative purposes only and are not necessarily drawn to scale.

DETAILED DESCRIPTION OF THE INVENTION

[00194] Before the present methods of the invention are described, it is to be understood that this invention is not limited to particular methods, components, products or combinations described, as such methods, components, products and combinations may, of course, vary. It is also to be understood that the terminology used herein is not intended to be limiting, since the scope of the present invention will be limited only by the appended claims.

[00195] As used herein, the singular forms "a", "an", and "the" include both singular and plural referents unless the context clearly dictates otherwise.

[00196] The terms "comprising", "comprises" and "comprised of" as used herein are synonymous with "including", "includes" or "containing", "contains", and are inclusive or open-ended and do not exclude additional, non-recited members, elements or method steps. It will be appreciated that the terms "comprising", "comprises" and "comprised of" as used herein comprise the terms "consisting of", "consists" and "consists of", as well as the terms "consisting essentially of", "consists essentially" and "consists essentially of". 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. It may be advantageous in the practice of the invention to be in compliance with Art. 53(c) EPC and Rule 28(b) and (c) EPC. Nothing herein is intended as a promise.

[00197] The recitation of numerical ranges by endpoints includes all numbers and fractions subsumed within the respective ranges, as well as the recited endpoints.

[00198] The term "about" or "approximately" as used herein when referring to a measurable value such as a parameter, an amount, a temporal duration, and the like, is meant to encompass variations of +/-20% or less, preferably +/-10% or less, more preferably +/-5% or less, and still more preferably +/-1% or less of and from the specified value, insofar such variations are appropriate to perform in the disclosed invention. It is to be understood that the value to which the modifier "about" or "approximately" refers is itself also specifically, and preferably, disclosed.

[00199] Whereas the terms "one or more" or "at least one", such as one or more or at least one member(s) of a group of members, is clear per se, by means of further exemplification, the term encompasses inter alia a reference to any one of said members, or to any two or more of said members, such as, e.g., any ≥3, ≥4, ≥5, ≥6 or ≥7 etc. of said members, and up to all said members.

[00200] All references cited in the present specification are hereby incorporated by reference in their entirety. In particular, the teachings of all references herein specifically referred to are incorporated by reference.

[00201] Unless otherwise defined, all terms used in disclosing the invention, including technical and scientific terms, have the meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. By means of further guidance, term definitions are included to better appreciate the teaching of the present invention.

[00202] In the following passages, different aspects of the invention are defined in more detail. Each aspect so defined may be combined with any other aspect or aspects unless clearly indicated to the contrary. In particular, any feature indicated as being preferred or advantageous may be combined with any other feature or features indicated as being preferred or advantageous.

[00203] Standard reference works setting forth the general principles of recombinant DNA technology include Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, ed. Sambrook et al., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989; Current Protocols in Molecular Biology, ed. Ausubel et al., Greene Publishing and Wiley-Interscience, New York, 1992 (with periodic updates) ("Ausubel et al. 1992"); the series Methods in Enzymology (Academic Press, Inc.); Innis et al., PCR Protocols: A Guide to Methods and Applications, Academic Press: San Diego, 1990; PCR 2: A Practical Approach (M.J. MacPherson, B.D. Hames and G.R. Taylor eds. (1995); Harlow and Lane, eds. (1988) Antibodies, a Laboratory Manual; and Animal Cell Culture (R.I. Freshney, ed. (1987). General principles of microbiology are set forth, for example, in Davis, B. D. et al., Microbiology, 3rd edition, Harper & Row, publishers, Philadelphia, Pa. (1980).

[00204] Reference throughout this specification to "one embodiment" or "an embodiment" means that a particular feature, structure or characteristic described in connection with the embodiment is included in at least one embodiment of the present invention. Thus, appearances

of the phrases "in one embodiment" or "in an embodiment" in various places throughout this specification are not necessarily all referring to the same embodiment, but may. Furthermore, the particular features, structures or characteristics may be combined in any suitable manner, as would be apparent to a person skilled in the art from this disclosure, in one or more embodiments. Furthermore, while some embodiments described herein include some but not other features included in other embodiments, combinations of features of different embodiments are meant to be within the scope of the invention, and form different embodiments, as would be understood by those in the art. For example, in the appended claims, any of the claimed embodiments can be used in any combination.

[00205] In this description of the invention, reference is made to the accompanying drawings that form a part hereof, and in which are shown by way of illustration only of specific embodiments in which the invention may be practiced. It is to be understood that other embodiments may be utilized and structural or logical changes may be made without departing from the scope of the present invention. The description, therefore, is not to be taken in a limiting sense, and the scope of the present invention is defined by the appended claims.

[00206] 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 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.

[00207] Preferred statements (features) and embodiments of this invention are set herein below. Each statements and embodiments of the invention so defined may be combined with any other statement and/or embodiments unless clearly indicated to the contrary. In particular, any feature indicated as being preferred or advantageous may be combined with any other feature or features or statements indicated as being preferred or advantageous.

[00208] As used herein, the term "non-human organism" or "non-human cell" refers to an organism or cell different than or not originating from *Homo sapiens*. As used herein, the term

"non-human eukaryote" or "non-human eukaryotic cell" refers to a eukaryotic organism or cell different than or not derived from *Homo sapiens*. In preferred embodiments, such eukaryote (cell) is a non-human animal (cell), such as (a cell or cell population of a) non-human mammal, non-human primate, an ungulate, rodent (preferably a mouse or rat), rabbit, canine, dog, cow, bovine, sheep, ovine, goat, pig, fowl, poultry, chicken, fish, insect, or arthropod, preferably a mammal, such as a rodent, in particular a mouse. In some embodiments of the invention the organism or subject or cell may be (a cell or cell population derived from) an arthropod, for example, an insect, or a nematode. In some methods of the invention the organism or subject or cell is a plant (cell). In some methods of the invention the organism or subject or cell is (derived from) algae, including microalgae, or fungus. The skilled person will appreciate that the eukaryotic cells which may be transplanted or introduced in a non-human eukaryote according to the methods as referred to herein are preferably derived from or originate from the same species as the eukaryote to which they are transplanted. For example, a mouse cell is transplanted in a mouse in certain embodiment according to the methods of the invention as described herein. In certain embodiments, the eukaryote is an immunocompromized eukaryote, i.e. a eukaryote in which the immune system is partially or completely shut down. For instance, immunocompromized mice may be used in the methods according to the invention as described herein. Examples of immunocompromized mice include, but are not limited to Nude mice, RAG -/- mice, SCID (severe compromised immunodeficiency) mice, SCID-Beige mice, NOD (non-obese diabetic)-SCID mice, NOG or NSG mice, etc.

[00209] It will be understood that the CRISPR-Cas system as described herein is non-naturally occurring in said cell, i.e. engineered or exogenous to said cell. The CRISPR-Cas system as referred to herein has been introduced in said cell. Methods for introducing the CRISPR-Cas system in a cell are known in the art, and are further described herein elsewhere. The cell comprising the CRISPR-Cas system, or having the CRISPR-Cas system introduced, according to the invention comprises or is capable of expressing the individual components of the CRISPR-Cas system to establish a functional CRISPR complex, capable of modifying (such as cleaving) a target DNA sequence. Accordingly, as referred to herein, the cell comprising the CRISPR-Cas system can be a cell comprising the individual components of the CRISPR-Cas system to establish a functional CRISPR complex, capable of modifying (such as cleaving) a target DNA sequence. Alternatively, as referred to herein, and preferably, the cell comprising the

CRISPR-Cas system can be a cell comprising one or more nucleic acid molecule encoding the individual components of the CRISPR-Cas system, which can be expressed in the cell to establish a functional CRISPR complex, capable of modifying (such as cleaving) a target DNA sequence.

[00210] As used herein, the term "crRNA" or "guide RNA" or "single guide RNA" or "sgRNA" or "one or more nucleic acid components" of a Type V or Type VI CRISPR-Cas locus effector protein comprises any polynucleotide sequence having sufficient complementarity with a target nucleic acid sequence to hybridize with the target nucleic acid sequence and direct sequence-specific binding of a nucleic acid-targeting complex to the target nucleic acid sequence. In some embodiments, the degree of complementarity, when optimally aligned using a suitable alignment algorithm, is about or more than about 50%, 60%, 75%, 80%, 85%, 90%, 95%, 97.5%, 99%, or more. Optimal alignment may be determined with the use of any suitable algorithm for aligning sequences, non-limiting example of which include the Smith-Waterman algorithm, the Needleman-Wunsch algorithm, algorithms based on the Burrows-Wheeler Transform (e.g., the Burrows Wheeler Aligner), ClustalW, Clustal X, BLAT, Novoalign (Novocraft Technologies; available at www.novocraft.com), ELAND (Illumina, San Diego, CA), SOAP (available at soap.genomics.org.cn), and Maq (available at maq.sourceforge.net). The ability of a guide sequence (within a nucleic acid-targeting guide RNA) to direct sequence-specific binding of a nucleic acid -targeting complex to a target nucleic acid sequence may be assessed by any suitable assay. For example, the components of a nucleic acid-targeting CRISPR system sufficient to form a nucleic acid -targeting complex, including the guide sequence to be tested, may be provided to a host cell having the corresponding target nucleic acid sequence, such as by transfection with vectors encoding the components of the nucleic acid -targeting complex, followed by an assessment of preferential targeting (e.g., cleavage) within the target nucleic acid sequence, such as by Surveyor assay as described herein. Similarly, cleavage of a target nucleic acid sequence may be evaluated in a test tube by providing the target nucleic acid sequence, components of a nucleic acid -targeting complex, including the guide sequence to be tested and a control guide sequence different from the test guide sequence, and comparing binding or rate of cleavage at the target sequence between the test and control guide sequence reactions. Other assays are possible, and will occur to those skilled in the art. A guide sequence, and hence a nucleic acid-targeting guide RNA may be selected to target any target

nucleic acid sequence. The target sequence may be DNA. The target sequence may be any RNA sequence. In some embodiments, the target sequence may be a sequence within a RNA molecule selected from the group consisting of messenger RNA (mRNA), pre-mRNA, ribosomal RNA (rRNA), transfer RNA (tRNA), micro-RNA (miRNA), small interfering RNA (siRNA), small nuclear RNA (snRNA), small nucleolar RNA (snoRNA), double stranded RNA (dsRNA), non-coding RNA (ncRNA), long non-coding RNA (lncRNA), and small cytoplasmatic RNA (scRNA). In some preferred embodiments, the target sequence may be a sequence within a RNA molecule selected from the group consisting of mRNA, pre-mRNA, and rRNA. In some preferred embodiments, the target sequence may be a sequence within a RNA molecule selected from the group consisting of ncRNA, and lncRNA. In some more preferred embodiments, the target sequence may be a sequence within an mRNA molecule or a pre-mRNA molecule.

[00211] In some embodiments, a nucleic acid-targeting guide RNA is selected to reduce the degree secondary structure within the RNA-targeting guide RNA. In some embodiments, about or less than about 75%, 50%, 40%, 30%, 25%, 20%, 15%, 10%, 5%, 1%, or fewer of the nucleotides of the nucleic acid-targeting guide RNA participate in self-complementary base pairing when optimally folded. Optimal folding may be determined by any suitable polynucleotide folding algorithm. Some programs are based on calculating the minimal Gibbs free energy. An example of one such algorithm is mFold, as described by Zuker and Stiegler (Nucleic Acids Res. 9 (1981), 133-148). Another example folding algorithm is the online webserver RNAfold, developed at Institute for Theoretical Chemistry at the University of Vienna, using the centroid structure prediction algorithm (see e.g., A.R. Gruber et al., 2008, Cell 106(1): 23-24; and PA Carr and GM Church, 2009, Nature Biotechnology 27(12): 1151-62).

[00212] In certain embodiments, a guide RNA or crRNA may comprise, consist essentially of, or consist of a direct repeat (DR) sequence and a guide sequence or spacer sequence. In certain embodiments, the guide RNA or crRNA may comprise, consist essentially of, or consist of a direct repeat sequence fused or linked to a guide sequence or spacer sequence. In certain embodiments, the direct repeat sequence may be located upstream (i.e., 5') from the guide sequence or spacer sequence. In other embodiments, the direct repeat sequence may be located downstream (i.e., 3') from the guide sequence or spacer sequence.

[00213] In certain embodiments, the crRNA comprises a stem loop, preferably a single stem loop. In certain embodiments, the direct repeat sequence forms a stem loop, preferably a single stem loop.

[00214] In certain embodiments, the spacer length of the guide RNA is from 15 to 35 nt. In certain embodiments, the spacer length of the guide RNA is at least 15 nucleotides. In certain embodiments, the spacer length is from 15 to 17 nt, e.g., 15, 16, or 17 nt, from 17 to 20 nt, e.g., 17, 18, 19, or 20 nt, from 20 to 24 nt, e.g., 20, 21, 22, 23, or 24 nt, from 23 to 25 nt, e.g., 23, 24, or 25 nt, from 24 to 27 nt, e.g., 24, 25, 26, or 27 nt, from 27-30 nt, e.g., 27, 28, 29, or 30 nt, from 30-35 nt, e.g., 30, 31, 32, 33, 34, or 35 nt, or 35 nt or longer.

[00215] The "tracrRNA" sequence or analogous terms includes any polynucleotide sequence that has sufficient complementarity with a crRNA sequence to hybridize. In some embodiments, the degree of complementarity between the tracrRNA sequence and crRNA sequence along the length of the shorter of the two when optimally aligned is about or more than about 25%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 97.5%, 99%, or higher. In some embodiments, the tracr sequence is about or more than about 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 40, 50, or more nucleotides in length. In some embodiments, the tracr sequence and crRNA sequence are contained within a single transcript, such that hybridization between the two produces a transcript having a secondary structure, such as a hairpin. In an embodiment of the invention, the transcript or transcribed polynucleotide sequence has at least two or more hairpins. In preferred embodiments, the transcript has two, three, four or five hairpins. In a further embodiment of the invention, the transcript has at most five hairpins. In a hairpin structure the portion of the sequence 5' of the final "N" and upstream of the loop corresponds to the tracr mate sequence, and the portion of the sequence 3' of the loop corresponds to the tracr sequence.

[00216] In general, the CRISPR-Cas, CRISPR-Cas9 or CRISPR system may be as used in the foregoing documents, such as WO 2014/093622 (PCT/US2013/074667) and refers collectively to transcripts and other elements involved in the expression of or directing the activity of CRISPR-associated ("Cas") genes, including sequences encoding a Cas gene, in particular a Cas9 gene in the case of CRISPR-Cas9, a tracr (trans-activating CRISPR) sequence (e.g. tracrRNA or an active partial tracrRNA), a tracr-mate sequence (encompassing a "direct repeat" and a tracrRNA-processed partial direct repeat in the context of an endogenous CRISPR system), a guide sequence (also referred to as a "spacer" in the context of an endogenous CRISPR

system), or "RNA(s)" as that term is herein used (e.g., RNA(s) to guide Cas9, e.g. CRISPR RNA and transactivating (tracr) RNA or a single guide RNA (sgRNA) (chimeric RNA)) or other sequences and transcripts from a CRISPR locus. In general, a CRISPR system is characterized by elements that promote the formation of a CRISPR complex at the site of a target sequence (also referred to as a protospacer in the context of an endogenous CRISPR system). In the context of formation of a CRISPR complex, "target sequence" refers to a sequence to which a guide sequence is designed to have complementarity, where hybridization between a target sequence and a guide sequence promotes the formation of a CRISPR complex. The section of the guide sequence through which complementarity to the target sequence is important for cleavage activity is referred to herein as the seed sequence. A target sequence may comprise any polynucleotide, such as DNA or RNA polynucleotides. In some embodiments, a target sequence is located in the nucleus or cytoplasm of a cell, and may include nucleic acids in or from mitochondrial, organelles, vesicles, liposomes or particles present within the cell. In some embodiments, especially for non-nuclear uses, NLSs are not preferred. In some embodiments, direct repeats may be identified *in silico* by searching for repetitive motifs that fulfill any or all of the following criteria: 1. found in a 2Kb window of genomic sequence flanking the type II CRISPR locus; 2. span from 20 to 50 bp; and 3. interspaced by 20 to 50 bp. In some embodiments, 2 of these criteria may be used, for instance 1 and 2, 2 and 3, or 1 and 3. In some embodiments, all 3 criteria may be used.

[00217] In embodiments of the invention the terms guide sequence and guide RNA, i.e. RNA capable of guiding Cas to a target genomic locus, are used interchangeably as in foregoing cited documents such as WO 2014/093622 (PCT/US2013/074667). In general, a guide sequence is any polynucleotide sequence having sufficient complementarity with a target polynucleotide sequence to hybridize with the target sequence and direct sequence-specific binding of a CRISPR complex to the target sequence. In some embodiments, the degree of complementarity between a guide sequence and its corresponding target sequence, when optimally aligned using a suitable alignment algorithm, is about or more than about 50%, 60%, 75%, 80%, 85%, 90%, 95%, 97.5%, 99%, or more. Optimal alignment may be determined with the use of any suitable algorithm for aligning sequences, non-limiting example of which include the Smith-Waterman algorithm, the Needleman-Wunsch algorithm, algorithms based on the Burrows-Wheeler Transform (e.g. the Burrows Wheeler Aligner), ClustalW, Clustal X, BLAT, Novoalign

(Novocraft Technologies; available at www.novocraft.com), ELAND (Illumina, San Diego, CA), SOAP (available at soap.genomics.org.cn), and Maq (available at maq.sourceforge.net). In some embodiments, a guide sequence is about or more than about 5, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, 75, or more nucleotides in length. In some embodiments, a guide sequence is less than about 75, 50, 45, 40, 35, 30, 25, 20, 15, 12, or fewer nucleotides in length. Preferably the guide sequence is 10 30 nucleotides long. The ability of a guide sequence to direct sequence-specific binding of a CRISPR complex to a target sequence may be assessed by any suitable assay. For example, the components of a CRISPR system sufficient to form a CRISPR complex, including the guide sequence to be tested, may be provided to a host cell having the corresponding target sequence, such as by transfection with vectors encoding the components of the CRISPR sequence, followed by an assessment of preferential cleavage within the target sequence, such as by Surveyor assay as described herein. Similarly, cleavage of a target polynucleotide sequence may be evaluated in a test tube by providing the target sequence, components of a CRISPR complex, including the guide sequence to be tested and a control guide sequence different from the test guide sequence, and comparing binding or rate of cleavage at the target sequence between the test and control guide sequence reactions. Other assays are possible, and will occur to those skilled in the art.

[00218] In some embodiments of CRISPR-Cas systems, the degree of complementarity between a guide sequence and its corresponding target sequence can be about or more than about 50%, 60%, 75%, 80%, 85%, 90%, 95%, 97.5%, 99%, or 100%; a guide or RNA or sgRNA can be about or more than about 5, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, 75, or more nucleotides in length; or guide or RNA or sgRNA can be less than about 75, 50, 45, 40, 35, 30, 25, 20, 15, 12, or fewer nucleotides in length; and advantageously tracr RNA is 30 or 50 nucleotides in length. However, an aspect of the invention is to reduce off-target interactions, e.g., reduce the guide interacting with a target sequence having low complementarity. Indeed, in the examples, it is shown that the invention involves mutations that result in the CRISPR-Cas system being able to distinguish between target and off-target sequences that have greater than 80% to about 95% complementarity, e.g., 83%-84% or 88-89% or 94-95% complementarity (for instance, distinguishing between a target having 18 nucleotides from an off-target of 18 nucleotides having 1, 2 or 3 mismatches). Accordingly, in the context of the present invention the degree of complementarity between a guide sequence and

its corresponding target sequence is greater than 94.5% or 95% or 95.5% or 96% or 96.5% or 97% or 97.5% or 98% or 98.5% or 99% or 99.5% or 99.9%, or 100%. Off target is less than 100% or 99.9% or 99.5% or 99% or 99% or 98.5% or 98% or 97.5% or 97% or 96.5% or 96% or 95.5% or 95% or 94.5% or 94% or 93% or 92% or 91% or 90% or 89% or 88% or 87% or 86% or 85% or 84% or 83% or 82% or 81% or 80% complementarity between the sequence and the guide, with it advantageous that off target is 100% or 99.9% or 99.5% or 99% or 99% or 98.5% or 98% or 97.5% or 97% or 96.5% or 96% or 95.5% or 95% or 94.5% complementarity between the sequence and the guide.

[00219] In particularly preferred embodiments according to the invention, the guide RNA (capable of guiding Cas to a target locus) may comprise (1) a guide sequence capable of hybridizing to a genomic target locus in the eukaryotic cell; (2) a tracr sequence; and (3) a tracr mate sequence. All (1) to (3) may reside in a single RNA, i.e. an sgRNA (arranged in a 5' to 3' orientation), or the tracr RNA may be a different RNA than the RNA containing the guide and tracr sequence. The tracr hybridizes to the tracr mate sequence and directs the CRISPR/Cas complex to the target sequence.

[00220] The methods according to the invention as described herein comprehend inducing one or more mutations in a eukaryotic cell (in vitro, i.e. in an isolated eukaryotic cell) as herein discussed comprising delivering to cell a vector as herein discussed. The mutation(s) can include the introduction, deletion, or substitution of one or more nucleotides at each target sequence of cell(s) via the guide(s) RNA(s) or sgRNA(s). The mutations can include the introduction, deletion, or substitution of 1-75 nucleotides at each target sequence of said cell(s) via the guide(s) RNA(s) or sgRNA(s). The mutations can include the introduction, deletion, or substitution of 1, 5, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, or 75 nucleotides at each target sequence of said cell(s) via the guide(s) RNA(s) or sgRNA(s). The mutations can include the introduction, deletion, or substitution of 5, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, or 75 nucleotides at each target sequence of said cell(s) via the guide(s) RNA(s) or sgRNA(s). The mutations include the introduction, deletion, or substitution of 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, or 75 nucleotides at each target sequence of said cell(s) via the guide(s) RNA(s) or sgRNA(s). The mutations can include the introduction, deletion, or substitution of 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50,

or 75 nucleotides at each target sequence of said cell(s) via the guide(s) RNA(s) or sgRNA(s). The mutations can include the introduction, deletion, or substitution of 40, 45, 50, 75, 100, 200, 300, 400 or 500 nucleotides at each target sequence of said cell(s) via the guide(s) RNA(s) or sgRNA(s).

[00221] For minimization of toxicity and off-target effect, control of the concentration of Cas mRNA and guide RNA delivered is considered. Optimal concentrations of Cas mRNA and guide RNA can be determined by testing different concentrations in a cellular or non-human eukaryote animal model and using deep sequencing the analyze the extent of modification at potential off-target genomic loci. Alternatively, to minimize the level of toxicity and off-target effect, Cas nickase mRNA (for example S. pyogenes Cas9 with the D10A mutation) can be delivered with a pair of guide RNAs targeting a site of interest. Guide sequences and strategies to minimize toxicity and off-target effects can be as in WO 2014/093622 (PCT/US2013/074667); or, via mutation as herein.

[00222] Typically, in the context of an endogenous CRISPR system, formation of a CRISPR complex (comprising a guide sequence hybridized to a target sequence and complexed with one or more Cas proteins) results in cleavage of one or both strands in or near (e.g. within 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 50, or more base pairs from) the target sequence. Without wishing to be bound by theory, the tracr sequence, which may comprise or consist of all or a portion of a wild-type tracr sequence (e.g. about or more than about 20, 26, 32, 45, 48, 54, 63, 67, 85, or more nucleotides of a wild-type tracr sequence), may also form part of a CRISPR complex, such as by hybridization along at least a portion of the tracr sequence to all or a portion of a tracr mate sequence that is operably linked to the guide sequence.

[00223] The location of the RuvCI, RuvCII, RuvCIII and HNH domains is indicated in figure 22A-C. As used herein, the term "RuvCI domain" preferably refers to the domain comprising amino acids 1-60 of *Streptococcus pyogenes* Cas9 (SpCas9) or a corresponding region in another Cas9 ortholog or a CRISPR nuclease other than Cas9. As used herein, the term "RuvCII domain" preferably refers to the domain comprising amino acids 718-775 of *Streptococcus pyogenes* Cas9 (SpCas9) or a corresponding region in another Cas9 ortholog or a CRISPR nuclease other than Cas9. As used herein, the term "RuvCIII domain" preferably refers to the domain comprising amino acids 909-1099 of *Streptococcus pyogenes* Cas9 (SpCas9) or a corresponding region in another Cas9 ortholog or a CRISPR nuclease other than Cas9. As used herein, the term "HNH

domain" preferably refers to the domain comprising amino acids 776-908 of *Streptococcus pyogenes* Cas9 (SpCas9) or a corresponding region in another Cas9 ortholog or a CRISPR nuclease other than Cas9. The groove between the RuvC and HNH domains refers to the groove between these domain in the three-dimensional structure of a non-naturally-occurring CRISPR enzyme as described herein. Figure 25D shows the Crystal structure of SaCas9 wherein the groove between the HNH and RuvC domains in the three-dimensional structure of SaCas9 is shown.

Aptamers

[00224] One guide with a first aptamer/RNA-binding protein pair can be linked or fused to an activator, whilst a second guide with a second aptamer/RNA-binding protein pair can be linked or fused to a repressor. The guides are for different targets (loci), so this allows one gene to be activated and one repressed. For example, the following schematic shows such an approach:

Guide 1– MS2 aptamer--------MS2 RNA-binding protein--------VP64 activator; and

Guide 2 – PP7 aptamer--------PP7 RNA-binding protein--------SID4x repressor.

[00225] The present invention also relates to orthogonal PP7/MS2 gene targeting. In this example, sgRNA targeting different loci are modified with distinct RNA loops in order to recruit MS2-VP64 or PP7-SID4X, which activate and repress their target loci, respectively. PP7 is the RNA-binding coat protein of the bacteriophage Pseudomonas. Like MS2, it binds a specific RNA sequence and secondary structure. The PP7 RNA-recognition motif is distinct from that of MS2. Consequently, PP7 and MS2 can be multiplexed to mediate distinct effects at different genomic loci simultaneously. For example, an sgRNA targeting locus A can be modified with MS2 loops, recruiting MS2-VP64 activators, while another sgRNA targeting locus B can be modified with PP7 loops, recruiting PP7-SID4X repressor domains. In the same cell, dCas9 can thus mediate orthogonal, locus-specific modifications. This principle can be extended to incorporate other orthogonal RNA-binding proteins such as Q-beta.

[00226] An alternative option for orthogonal repression includes incorporating non-coding RNA loops with transactive repressive function into the guide (either at similar positions to the MS2/PP7 loops integrated into the guide or at the 3' terminus of the guide). For instance, guides were designed with non-coding (but known to be repressive) RNA loops (e.g. using the Alu repressor (in RNA) that interferes with RNA polymerase II in mammalian cells). The Alu RNA

sequence was located: in place of the MS2 RNA sequences as used herein (e.g. at tetraloop and/or stem loop 2); and/or at 3' terminus of the guide. This gives possible combinations of MS2, PP7 or Alu at the tetraloop and/or stemloop 2 positions, as well as, optionally, addition of Alu at the 3' end of the guide (with or without a linker).

[00227] The use of two different aptamers (distinct RNA) allows an activator-adaptor protein fusion and a repressor-adaptor protein fusion to be used, with different guides, to activate expression of one gene, whilst repressing another. They, along with their different guides can be administered together, or substantially together, in a multiplexed approach. A large number of such modified guides can be used all at the same time, for example 10 or 20 or 30 and so forth, whilst only one (or at least a minimal number) of Cas9s to be delivered, as a comparatively small number of Cas9s can be used with a large number modified guides. The adaptor protein may be associated (preferably linked or fused to) one or more activators or one or more repressors. For example, the adaptor protein may be associated with a first activator and a second activator. The first and second activators may be the same, but they are preferably different activators. For example, one might be VP64, whilst the other might be p65, although these are just examples and other transcriptional activators are envisaged. Three or more or even four or more activators (or repressors) may be used, but package size may limit the number being higher than 5 different functional domains. Linkers are preferably used, over a direct fusion to the adaptor protein, where two or more functional domains are associated with the adaptor protein. Suitable linkers might include the GlySer linker.

[00228] It is also envisaged that the enzyme-guide complex as a whole may be associated with two or more functional domains. For example, there may be two or more functional domains associated with the enzyme, or there may be two or more functional domains associated with the guide (via one or more adaptor proteins), or there may be one or more functional domains associated with the enzyme and one or more functional domains associated with the guide (via one or more adaptor proteins).

[00229] The fusion between the adaptor protein and the activator or repressor may include a linker. For example, GlySer linkers GGGS can be used. They can be used in repeats of 3 ($(GGGGS)_3$) or 6, 9 or even 12 or more, to provide suitable lengths, as required. Linkers can be used between the RNA-binding protein and the functional domain (activator or repressor), or

between the CRISPR Enzyme (Cas9) and the functional domain (activator or repressor). The linkers the user to engineer appropriate amounts of "mechanical flexibility".

Dead guides: Guide RNAs comprising a dead guide sequence may be used in the present invention

[00230] In one aspect, the invention provides guide sequences which are modified in a manner which allows for formation of the CRISPR complex and successful binding to the target, while at the same time, not allowing for successful nuclease activity (i.e. without nuclease activity / without indel activity). For matters of explanation such modified guide sequences are referred to as "dead guides" or "dead guide sequences". These dead guides or dead guide sequences can be thought of as catalytically inactive or conformationally inactive with regard to nuclease activity. Nuclease activity may be measured using surveyor analysis or deep sequencing as commonly used in the art, preferably surveyor analysis. Similarly, dead guide sequences may not sufficiently engage in productive base pairing with respect to the ability to promote catalytic activity or to distinguish on-target and off-target binding activity. Briefly, the surveyor assay involves purifying and amplifying a CRISPR target site for a gene and forming heteroduplexes with primers amplifying the CRISPR target site. After re-anneal, the products are treated with SURVEYOR nuclease and SURVEYOR enhancer S (Transgenomics) following the manufacturer's recommended protocols, analyzed on gels, and quantified based upon relative band intensities.

[00231] Hence, in a related aspect, the invention provides a non-naturally occurring or engineered composition Cas9 CRISPR-Cas system comprising a functional Cas9 as described herein, and guide RNA (gRNA) wherein the gRNA comprises a dead guide sequence whereby the gRNA is capable of hybridizing to a target sequence such that the Cas9 CRISPR-Cas system is directed to a genomic locus of interest in a cell without detectable indel activity resultant from nuclease activity of a non-mutant Cas9 enzyme of the system as detected by a SURVEYOR assay. For shorthand purposes, a gRNA comprising a dead guide sequence whereby the gRNA is capable of hybridizing to a target sequence such that the Cas9 CRISPR-Cas system is directed to a genomic locus of interest in a cell without detectable indel activity resultant from nuclease activity of a non-mutant Cas9 enzyme of the system as detected by a SURVEYOR assay is herein termed a "dead gRNA". It is to be understood that any of the gRNAs according to the

invention as described herein elsewhere may be used as dead gRNAs / gRNAs comprising a dead guide sequence as described herein below. Any of the methods, products, compositions and uses as described herein elsewhere is equally applicable with the dead gRNAs / gRNAs comprising a dead guide sequence as further detailed below. By means of further guidance, the following particular aspects and embodiments are provided.

[00232] The ability of a dead guide sequence to direct sequence-specific binding of a CRISPR complex to a target sequence may be assessed by any suitable assay. For example, the components of a CRISPR system sufficient to form a CRISPR complex, including the dead guide sequence to be tested, may be provided to a host cell having the corresponding target sequence, such as by transfection with vectors encoding the components of the CRISPR sequence, followed by an assessment of preferential cleavage within the target sequence, such as by Surveyor assay as described herein. Similarly, cleavage of a target polynucleotide sequence may be evaluated in a test tube by providing the target sequence, components of a CRISPR complex, including the dead guide sequence to be tested and a control guide sequence different from the test dead guide sequence, and comparing binding or rate of cleavage at the target sequence between the test and control guide sequence reactions. Other assays are possible, and will occur to those skilled in the art. A dead guide sequence may be selected to target any target sequence. In some embodiments, the target sequence is a sequence within a genome of a cell.

[00233] As explained further herein, several structural parameters allow for a proper framework to arrive at such dead guides. Dead guide sequences are shorter than respective guide sequences which result in active Cas9-specific indel formation. Dead guides are 5%, 10%, 20%, 30%, 40%, 50%, shorter than respective guides directed to the same Cas9 leading to active Cas9-specific indel formation.

[00234] As explained below and known in the art, one aspect of gRNA – Cas9 specificity is the direct repeat sequence, which is to be appropriately linked to such guides. In particular, this implies that the direct repeat sequences are designed dependent on the origin of the Cas9. Thus, structural data available for validated dead guide sequences may be used for designing Cas9 specific equivalents. Structural similarity between, e.g., the orthologous nuclease domains RuvC of two or more Cas9 effector proteins may be used to transfer design equivalent dead guides. Thus, the dead guide herein may be appropriately modified in length and sequence to reflect such

Cas9 specific equivalents, allowing for formation of the CRISPR complex and successful binding to the target, while at the same time, not allowing for successful nuclease activity.

[00235] The use of dead guides in the context herein as well as the state of the art provides a surprising and unexpected platform for network biology and/or systems biology in both in vitro, ex vivo, and in vivo applications, allowing for multiplex gene targeting, and in particular bidirectional multiplex gene targeting. Prior to the use of dead guides, addressing multiple targets, for example for activation, repression and/or silencing of gene activity, has been challenging and in some cases not possible. With the use of dead guides, multiple targets, and thus multiple activities, may be addressed, for example, in the same cell, in the same animal, or in the same patient. Such multiplexing may occur at the same time or staggered for a desired timeframe.

[00236] For example, the dead guides now allow for the first time to use gRNA as a means for gene targeting, without the consequence of nuclease activity, while at the same time providing directed means for activation or repression. Guide RNA comprising a dead guide may be modified to further include elements in a manner which allow for activation or repression of gene activity, in particular protein adaptors (e.g. aptamers) as described herein elsewhere allowing for functional placement of gene effectors (e.g. activators or repressors of gene activity). One example is the incorporation of aptamers, as explained herein and in the state of the art. By engineering the gRNA comprising a dead guide to incorporate protein-interacting aptamers (Konermann et al., "Genome-scale transcription activation by an engineered CRISPR-Cas9 complex," doi:10.1038/nature14136, incorporated herein by reference), one may assemble a synthetic transcription activation complex consisting of multiple distinct effector domains. Such may be modeled after natural transcription activation processes. For example, an aptamer, which selectively binds an effector (e.g. an activator or repressor; dimerized MS2 bacteriophage coat proteins as fusion proteins with an activator or repressor), or a protein which itself binds an effector (e.g. activator or repressor) may be appended to a dead gRNA tetraloop and/or a stem-loop 2. In the case of MS2, the fusion protein MS2-VP64 binds to the tetraloop and/or stem-loop 2 and in turn mediates transcriptional up-regulation, for example for *Neurog2*. Other transcriptional activators are, for example, VP64. P65, HSF1, and MyoD1. By mere example of this concept, replacement of the MS2 stem-loops with PP7-interacting stem-loops may be used to recruit repressive elements.

[00237] Thus, one aspect is a gRNA of the invention which comprises a dead guide, wherein the gRNA further comprises modifications which provide for gene activation or repression, as described herein. The dead gRNA may comprise one or more aptamers. The aptamers may be specific to gene effectors, gene activators or gene repressors. Alternatively, the aptamers may be specific to a protein which in turn is specific to and recruits / binds a specific gene effector, gene activator or gene repressor. If there are multiple sites for activator or repressor recruitment, it is preferred that the sites are specific to either activators or repressors. If there are multiple sites for activator or repressor binding, the sites may be specific to the same activators or same repressors. The sites may also be specific to different activators or different repressors. The gene effectors, gene activators, gene repressors may be present in the form of fusion proteins.

[00238] In an embodiment, the dead gRNA as described herein or the Cas9 CRISPR-Cas complex as described herein includes a non-naturally occurring or engineered composition comprising two or more adaptor proteins, wherein each protein is associated with one or more functional domains and wherein the adaptor protein binds to the distinct RNA sequence(s) inserted into the at least one loop of the dead gRNA.

[00239] Hence, an aspect provides a non-naturally occurring or engineered composition comprising a guide RNA (gRNA) comprising a dead guide sequence capable of hybridizing to a target sequence in a genomic locus of interest in a cell, wherein the dead guide sequence is as defined herein, a Cas9 comprising at least one or more nuclear localization sequences, wherein the Cas9 optionally comprises at least one mutation wherein at least one loop of the dead gRNA is modified by the insertion of distinct RNA sequence(s) that bind to one or more adaptor proteins, and wherein the adaptor protein is associated with one or more functional domains; or, wherein the dead gRNA is modified to have at least one non-coding functional loop, and wherein the composition comprises two or more adaptor proteins, wherein the each protein is associated with one or more functional domains.

[00240] In certain embodiments, the adaptor protein is a fusion protein comprising the functional domain, the fusion protein optionally comprising a linker between the adaptor protein and the functional domain, the linker optionally including a GlySer linker.

[00241] In certain embodiments, the at least one loop of the dead gRNA is not modified by the insertion of distinct RNA sequence(s) that bind to the two or more adaptor proteins.

[00242] In certain embodiments, the one or more functional domains associated with the adaptor protein is a transcriptional activation domain.

[00243] In certain embodiments, the one or more functional domains associated with the adaptor protein is a transcriptional activation domain comprising VP64, p65, MyoD1, HSF1, RTA or SET7/9.

[00244] In certain embodiments, the one or more functional domains associated with the adaptor protein is a transcriptional repressor domain.

[00245] In certain embodiments, the transcriptional repressor domain is a KRAB domain.

[00246] In certain embodiments, the transcriptional repressor domain is a NuE domain, NcoR domain, SID domain or a SID4X domain.

[00247] In certain embodiments, at least one of the one or more functional domains associated with the adaptor protein have one or more activities comprising methylase activity, demethylase activity, transcription activation activity, transcription repression activity, transcription release factor activity, histone modification activity, DNA integration activity RNA cleavage activity, DNA cleavage activity or nucleic acid binding activity.

[00248] In certain embodiments, the DNA cleavage activity is due to a Fok1 nuclease.

[00249] In certain embodiments, the dead gRNA is modified so that, after dead gRNA binds the adaptor protein and further binds to the Cas9 and target, the functional domain is in a spatial orientation allowing for the functional domain to function in its attributed function.

[00250] In certain embodiments, the at least one loop of the dead gRNA is tetra loop and/or loop2. In certain embodiments, the tetra loop and loop 2 of the dead gRNA are modified by the insertion of the distinct RNA sequence(s).

[00251] In certain embodiments, the insertion of distinct RNA sequence(s) that bind to one or more adaptor proteins is an aptamer sequence. In certain embodiments, the aptamer sequence is two or more aptamer sequences specific to the same adaptor protein. In certain embodiments, the aptamer sequence is two or more aptamer sequences specific to different adaptor protein.

[00252] In certain embodiments, the adaptor protein comprises MS2, PP7, Qβ, F2, GA, fr, JP501, M12, R17, BZ13, JP34, JP500, KU1, M11, MX1, TW18, VK, SP, FI, ID2, NL95, TW19, AP205, ϕCb5, ϕCb8r, ϕCb12r, ϕCb23r, 7s, PRR1.

[00253] In certain embodiments, the cell is a eukaryotic cell. In certain embodiments, the eukaryotic cell is a mammalian cell, optionally a mouse cell. In certain embodiments, the mammalian cell is a human cell.

[00254] In certain embodiments, a first adaptor protein is associated with a p65 domain and a second adaptor protein is associated with a HSF1 domain.

[00255] In certain embodiments, the composition comprises a Cas9 CRISPR-Cas complex having at least three functional domains, at least one of which is associated with the Cas9 and at least two of which are associated with dead gRNA.

[00256] In certain embodiments, the composition further comprises a second gRNA, wherein the second gRNA is a live gRNA capable of hybridizing to a second target sequence such that a second Cas9 CRISPR-Cas system is directed to a second genomic locus of interest in a cell with detectable indel activity at the second genomic locus resultant from nuclease activity of the Cas9 enzyme of the system.

[00257] In certain embodiments, the composition further comprises a plurality of dead gRNAs and/or a plurality of live gRNAs.

[00258] One aspect of the invention is to take advantage of the modularity and customizability of the gRNA scaffold to establish a series of gRNA scaffolds with different binding sites (in particular aptamers) for recruiting distinct types of effectors in an orthogonal manner. Again, for matters of example and illustration of the broader concept, replacement of the MS2 stem-loops with PP7-interacting stem-loops may be used to bind / recruit repressive elements, enabling multiplexed bidirectional transcriptional control. Thus, in general, gRNA comprising a dead guide may be employed to provide for multiplex transcriptional control and preferred bidirectional transcriptional control. This transcriptional control is most preferred of genes. For example, one or more gRNA comprising dead guide(s) may be employed in targeting the activation of one or more target genes. At the same time, one or more gRNA comprising dead guide(s) may be employed in targeting the repression of one or more target genes. Such a sequence may be applied in a variety of different combinations, for example the target genes are first repressed and then at an appropriate period other targets are activated, or select genes are repressed at the same time as select genes are activated, followed by further activation and/or repression. As a result, multiple components of one or more biological systems may advantageously be addressed together.

[00259] In an aspect, the invention provides nucleic acid molecule(s) encoding dead gRNA or the Cas9 CRISPR-Cas complex or the composition as described herein.

[00260] In an aspect, the invention provides a vector system comprising: a nucleic acid molecule encoding dead guide RNA as defined herein. In certain embodiments, the vector system further comprises a nucleic acid molecule(s) encoding Cas9. In certain embodiments, the vector system further comprises a nucleic acid molecule(s) encoding (live) gRNA. In certain embodiments, the nucleic acid molecule or the vector further comprises regulatory element(s) operable in a eukaryotic cell operably linked to the nucleic acid molecule encoding the guide sequence (gRNA) and/or the nucleic acid molecule encoding Cas9 and/or the optional nuclear localization sequence(s).

[00261] In another aspect, structural analysis may also be used to study interactions between the dead guide and the active Cas9 nuclease that enable DNA binding, but no DNA cutting. In this way amino acids important for nuclease activity of Cas9 are determined. Modification of such amino acids allows for improved Cas9 enzymes used for gene editing.

[00262] A further aspect is combining the use of dead guides as explained herein with other applications of CRISPR, as explained herein as well as known in the art. For example, gRNA comprising dead guide(s) for targeted multiplex gene activation or repression or targeted multiplex bidirectional gene activation / repression may be combined with gRNA comprising guides which maintain nuclease activity, as explained herein. Such gRNA comprising guides which maintain nuclease activity may or may not further include modifications which allow for repression of gene activity (e.g. aptamers). Such gRNA comprising guides which maintain nuclease activity may or may not further include modifications which allow for activation of gene activity (e.g. aptamers). In such a manner, a further means for multiplex gene control is introduced (e.g. multiplex gene targeted activation without nuclease activity / without indel activity may be provided at the same time or in combination with gene targeted repression with nuclease activity).

[00263] For example, 1) using one or more gRNA (e.g. 1-50, 1-40, 1-30, 1-20, preferably 1-10, more preferably 1-5) comprising dead guide(s) targeted to one or more genes and further modified with appropriate aptamers for the recruitment of gene activators; 2) may be combined with one or more gRNA (e.g. 1-50, 1-40, 1-30, 1-20, preferably 1-10, more preferably 1-5) comprising dead guide(s) targeted to one or more genes and further modified with appropriate

aptamers for the recruitment of gene repressors. 1) and/or 2) may then be combined with 3) one or more gRNA (e.g. 1-50, 1-40, 1-30, 1-20, preferably 1-10, more preferably 1-5) targeted to one or more genes. This combination can then be carried out in turn with 1) + 2) + 3) with 4) one or more gRNA (e.g. 1-50, 1-40, 1-30, 1-20, preferably 1-10, more preferably 1-5) targeted to one or more genes and further modified with appropriate aptamers for the recruitment of gene activators. This combination can then be carried in turn with 1) + 2) + 3) + 4) with 5) one or more gRNA (e.g. 1-50, 1-40, 1-30, 1-20, preferably 1-10, more preferably 1-5) targeted to one or more genes and further modified with appropriate aptamers for the recruitment of gene repressors. As a result various uses and combinations are included in the invention. For example, combination 1) + 2); combination 1) + 3); combination 2) + 3); combination 1) + 2) + 3); combination 1) + 2) +3) +4); combination 1) + 3) + 4); combination 2) + 3) +4); combination 1) + 2) + 4); combination 1) + 2) +3) +4) + 5); combination 1) + 3) + 4) +5); combination 2) + 3) +4) +5); combination 1) + 2) + 4) +5); combination 1) + 2) +3) + 5); combination 1) + 3) +5); combination 2) + 3) +5); combination 1) + 2) +5).

[00264] In an aspect, the invention provides an algorithm for designing, evaluating, or selecting a dead guide RNA targeting sequence (dead guide sequence) for guiding a Cas9 CRISPR-Cas system to a target gene locus. In particular, it has been determined that dead guide RNA specificity relates to and can be optimized by varying i) GC content and ii) targeting sequence length. In an aspect, the invention provides an algorithm for designing or evaluating a dead guide RNA targeting sequence that minimizes off-target binding or interaction of the dead guide RNA. In an embodiment of the invention, the algorithm for selecting a dead guide RNA targeting sequence for directing a CRISPR system to a gene locus in an organism comprises a) locating one or more CRISPR motifs in the gene locus, analyzing the 20 nt sequence downstream of each CRISPR motif by i) determining the GC content of the sequence; and ii) determining whether there are off-target matches of the 15 downstream nucleotides nearest to the CRISPR motif in the genome of the organism, and c) selecting the 15 nucleotide sequence for use in a dead guide RNA if the GC content of the sequence is 70% or less and no off-target matches are identified. In an embodiment, the sequence is selected for a targeting sequence if the GC content is 60% or less. In certain embodiments, the sequence is selected for a targeting sequence if the GC content is 55% or less, 50% or less, 45% or less, 40% or less, 35% or less or 30% or less. In an embodiment, two or more sequences of the gene locus are analyzed and the sequence having

the lowest GC content, or the next lowest GC content, or the next lowest GC content is selected. In an embodiment, the sequence is selected for a targeting sequence if no off-target matches are identified in the genome of the organism. In an embodiment, the targeting sequence is selected if no off-target matches are identified in regulatory sequences of the genome.

[00265] In an aspect, the invention provides a method of selecting a dead guide RNA targeting sequence for directing a functionalized CRISPR system to a gene locus in an organism, which comprises: a) locating one or more CRISPR motifs in the gene locus; b) analyzing the 20 nt sequence downstream of each CRISPR motif by: i) determining the GC content of the sequence; and ii) determining whether there are off-target matches of the first 15 nt of the sequence in the genome of the organism; c) selecting the sequence for use in a guide RNA if the GC content of the sequence is 70% or less and no off-target matches are identified. In an embodiment, the sequence is selected if the GC content is 50% or less. In an embodiment, the sequence is selected if the GC content is 40% or less. In an embodiment, the sequence is selected if the GC content is 30% or less. In an embodiment, two or more sequences are analyzed and the sequence having the lowest GC content is selected. In an embodiment, off-target matches are determined in regulatory sequences of the organism. In an embodiment, the gene locus is a regulatory region. An aspect provides a dead guide RNA comprising the targeting sequence selected according to the aforementioned methods.

[00266] In an aspect, the invention provides a dead guide RNA for targeting a functionalized CRISPR system to a gene locus in an organism. In an embodiment of the invention, the dead guide RNA comprises a targeting sequence wherein the CG content of the target sequence is 70% or less, and the first 15 nt of the targeting sequence does not match an off-target sequence downstream from a CRISPR motif in the regulatory sequence of another gene locus in the organism. In certain embodiments, the GC content of the targeting sequence 60% or less, 55% or less, 50% or less, 45% or less, 40% or less, 35% or less or 30% or less. In certain embodiments, the GC content of the targeting sequence is from 70% to 60% or from 60% to 50% or from 50% to 40% or from 40% to 30%. In an embodiment, the targeting sequence has the lowest CG content among potential targeting sequences of the locus.

[00267] In an embodiment of the invention, the first 15 nt of the dead guide match the target sequence. In another embodiment, first 14 nt of the dead guide match the target sequence. In another embodiment, the first 13 nt of the dead guide match the target sequence. In another

embodiment first 12 nt of the dead guide match the target sequence. In another embodiment, first 11 nt of the dead guide match the target sequence. In another embodiment, the first 10 nt of the dead guide match the target sequence. In an embodiment of the invention the first 15 nt of the dead guide does not match an off-target sequence downstream from a CRISPR motif in the regulatory region of another gene locus. In other embodiments, the first 14 nt, or the first 13 nt of the dead guide, or the first 12 nt of the guide, or the first 11 nt of the dead guide, or the first 10 nt of the dead guide, does not match an off-target sequence downstream from a CRISPR motif in the regulatory region of another gene locus. In other embodiments, the first 15 nt, or 14 nt, or 13 nt, or 12 nt, or 11 nt of the dead guide do not match an off-target sequence downstream from a CRISPR motif in the genome.

[00268] In certain embodiments, the dead guide RNA includes additional nucleotides at the 3'-end that do not match the target sequence. Thus, a dead guide RNA that includes the first 15 nt, or 14 nt, or 13 nt, or 12 nt, or 11 nt downstream of a CRISPR motif can be extended in length at the 3' end to 12 nt, 13 nt, 14 nt, 15 nt, 16 nt, 17 nt, 18 nt, 19 nt, 20 nt, or longer.

[00269] The invention provides a method for directing a Cas9 CRISPR-Cas system, including but not limited to a dead Cas9 (dCas9) or functionalized Cas9 system (which may comprise a functionalized Cas9 or functionalized guide) to a gene locus. In an aspect, the invention provides a method for selecting a dead guide RNA targeting sequence and directing a functionalized CRISPR system to a gene locus in an organism. In an aspect, the invention provides a method for selecting a dead guide RNA targeting sequence and effecting gene regulation of a target gene locus by a functionalized Cas9 CRISPR-Cas system. In certain embodiments, the method is used to effect target gene regulation while minimizing off-target effects. In an aspect, the invention provides a method for selecting two or more dead guide RNA targeting sequences and effecting gene regulation of two or more target gene loci by a functionalized Cas9 CRISPR-Cas system. In certain embodiments, the method is used to effect regulation of two or more target gene loci while minimizing off-target effects.

[00270] In an aspect, the invention provides a method of selecting a dead guide RNA targeting sequence for directing a functionalized Cas9 to a gene locus in an organism, which comprises: a) locating one or more CRISPR motifs in the gene locus; b) analyzing the sequence downstream of each CRISPR motif by: i) selecting 10 to 15 nt adjacent to the CRISPR motif, ii) determining the GC content of the sequence; and c) selecting the 10 to 15 nt sequence as a

targeting sequence for use in a guide RNA if the GC content of the sequence is 40% or more. In an embodiment, the sequence is selected if the GC content is 50% or more. In an embodiment, the sequence is selected if the GC content is 60% or more. In an embodiment, the sequence is selected if the GC content is 70% or more. In an embodiment, two or more sequences are analyzed and the sequence having the highest GC content is selected. In an embodiment, the method further comprises adding nucleotides to the 3' end of the selected sequence which do not match the sequence downstream of the CRISPR motif. An aspect provides a dead guide RNA comprising the targeting sequence selected according to the aforementioned methods.

[00271] In an aspect, the invention provides a dead guide RNA for directing a functionalized CRISPR system to a gene locus in an organism wherein the targeting sequence of the dead guide RNA consists of 10 to 15 nucleotides adjacent to the CRISPR motif of the gene locus, wherein the CG content of the target sequence is 50% or more. In certain embodiments, the dead guide RNA further comprises nucleotides added to the 3' end of the targeting sequence which do not match the sequence downstream of the CRISPR motif of the gene locus.

[00272] In an aspect, the invention provides for a single effector to be directed to one or more, or two or more gene loci. In certain embodiments, the effector is associated with a Cas9, and one or more, or two or more selected dead guide RNAs are used to direct the Cas9-associated effector to one or more, or two or more selected target gene loci. In certain embodiments, the effector is associated with one or more, or two or more selected dead guide RNAs, each selected dead guide RNA, when complexed with a Cas9 enzyme, causing its associated effector to localize to the dead guide RNA target. One non-limiting example of such CRISPR systems modulates activity of one or more, or two or more gene loci subject to regulation by the same transcription factor.

[00273] In an aspect, the invention provides for two or more effectors to be directed to one or more gene loci. In certain embodiments, two or more dead guide RNAs are employed, each of the two or more effectors being associated with a selected dead guide RNA, with each of the two or more effectors being localized to the selected target of its dead guide RNA. One non-limiting example of such CRISPR systems modulates activity of one or more, or two or more gene loci subject to regulation by different transcription factors. Thus, in one non-limiting embodiment, two or more transcription factors are localized to different regulatory sequences of a single gene. In another non-limiting embodiment, two or more transcription factors are localized to different

regulatory sequences of different genes. In certain embodiments, one transcription factor is an activator. In certain embodiments, one transcription factor is an inhibitor. In certain embodiments, one transcription factor is an activator and another transcription factor is an inhibitor. In certain embodiments, gene loci expressing different components of the same regulatory pathway are regulated. In certain embodiments, gene loci expressing components of different regulatory pathways are regulated.

[00274] In an aspect, the invention also provides a method and algorithm for designing and selecting dead guide RNAs that are specific for target DNA cleavage or target binding and gene regulation mediated by an active Cas9 CRISPR-Cas system. In certain embodiments, the Cas9 CRISPR-Cas system provides orthogonal gene control using an active Cas9 which cleaves target DNA at one gene locus while at the same time binds to and promotes regulation of another gene locus.

[00275] In an aspect, the invention provides an method of selecting a dead guide RNA targeting sequence for directing a functionalized Cas9 to a gene locus in an organism, without cleavage, which comprises a) locating one or more CRISPR motifs in the gene locus; b) analyzing the sequence downstream of each CRISPR motif by i) selecting 10 to 15 nt adjacent to the CRISPR motif, ii) determining the GC content of the sequence, and c) selecting the 10 to 15 nt sequence as a targeting sequence for use in a dead guide RNA if the GC content of the sequence is 30% more, 40% or more. In certain embodiments, the GC content of the targeting sequence is 35% or more, 40% or more, 45% or more, 50% or more, 55% or more, 60% or more, 65% or more, or 70% or more. In certain embodiments, the GC content of the targeting sequence is from 30% to 40% or from 40% to 50% or from 50% to 60% or from 60% to 70%. In an embodiment of the invention, two or more sequences in a gene locus are analyzed and the sequence having the highest GC content is selected.

[00276] In an embodiment of the invention, the portion of the targeting sequence in which GC content is evaluated is 10 to 15 contiguous nucleotides of the 15 target nucleotides nearest to the PAM. In an embodiment of the invention, the portion of the guide in which GC content is considered is the 10 to 11 nucleotides or 11 to 12 nucleotides or 12 to 13 nucleotides or 13, or 14, or 15 contiguous nucleotides of the 15 nucleotides nearest to the PAM.

[00277] In an aspect, the invention further provides an algorithm for identifying dead guide RNAs which promote CRISPR system gene locus cleavage while avoiding functional activation

or inhibition. It is observed that increased GC content in dead guide RNAs of 16 to 20 nucleotides coincides with increased DNA cleavage and reduced functional activation.

[00278] It is also demonstrated herein that efficiency of functionalized Cas9 can be increased by addition of nucleotides to the 3' end of a guide RNA which do not match a target sequence downstream of the CRISPR motif. For example, of dead guide RNA 11 to 15 nt in length, shorter guides may be less likely to promote target cleavage, but are also less efficient at promoting CRISPR system binding and functional control. In certain embodiments, addition of nucleotides that don't match the target sequence to the 3' end of the dead guide RNA increase activation efficiency while not increasing undesired target cleavage. In an aspect, the invention also provides a method and algorithm for identifying improved dead guide RNAs that effectively promote CRISPRP system function in DNA binding and gene regulation while not promoting DNA cleavage. Thus, in certain embodiments, the invention provides a dead guide RNA that includes the first 15 nt, or 14 nt, or 13 nt, or 12 nt, or 11 nt downstream of a CRISPR motif and is extended in length at the 3' end by nucleotides that mismatch the target to 12 nt, 13 nt, 14 nt, 15 nt, 16 nt, 17 nt, 18 nt, 19 nt, 20 nt, or longer.

[00279] In an aspect, the invention provides a method for effecting selective orthogonal gene control. As will be appreciated from the disclosure herein, dead guide selection according to the invention, taking into account guide length and GC content, provides effective and selective transcription control by a functional Cas9 CRISPR-Cas system, for example to regulate transcription of a gene locus by activation or inhibition and minimize off-target effects. Accordingly, by providing effective regulation of individual target loci, the invention also provides effective orthogonal regulation of two or more target loci.

[00280] In certain embodiments, orthogonal gene control is by activation or inhibition of two or more target loci. In certain embodiments, orthogonal gene control is by activation or inhibition of one or more target locus and cleavage of one or more target locus.

[00281] In one aspect, the invention provides a cell comprising a non-naturally occurring Cas9 CRISPR-Cas system comprising one or more dead guide RNAs disclosed or made according to a method or algorithm described herein wherein the expression of one or more gene products has been altered. In an embodiment of the invention, the expression in the cell of two or more gene products has been altered. The invention also provides a cell line from such a cell.

[00282] In one aspect, the invention provides a multicellular organism comprising one or more cells comprising a non-naturally occurring Cas9 CRISPR-Cas system comprising one or more dead guide RNAs disclosed or made according to a method or algorithm described herein. In one aspect, the invention provides a product from a cell, cell line, or multicellular organism comprising a non-naturally occurring Cas9 CRISPR-Cas system comprising one or more dead guide RNAs disclosed or made according to a method or algorithm described herein.

[00283] A further aspect of this invention is the use of gRNA comprising dead guide(s) as described herein, optionally in combination with gRNA comprising guide(s) as described herein or in the state of the art, in combination with systems e.g. cells, transgenic animals, transgenic mice, inducible transgenic animals, inducible transgenic mice) which are engineered for either overexpression of Cas9 or preferably knock in Cas9. As a result a single system (e.g. transgenic animal, cell) can serve as a basis for multiplex gene modifications in systems / network biology. On account of the dead guides, this is now possible in both in vitro, ex vivo, and in vivo.

[00284] For example, once the Cas9 is provided for, one or more dead gRNAs may be provided to direct multiplex gene regulation, and preferably multiplex bidirectional gene regulation. The one or more dead gRNAs may be provided in a spatially and temporally appropriate manner if necessary or desired (for example tissue specific induction of Cas9 expression). On account that the transgenic / inducible Cas9 is provided for (e.g. expressed) in the cell, tissue, animal of interest, both gRNAs comprising dead guides or gRNAs comprising guides are equally effective. In the same manner, a further aspect of this invention is the use of gRNA comprising dead guide(s) as described herein, optionally in combination with gRNA comprising guide(s) as described herein or in the state of the art, in combination with systems (e.g. cells, transgenic animals, transgenic mice, inducible transgenic animals, inducible transgenic mice) which are engineered for knockout Cas9 CRISPR-Cas.

[00285] As a result, the combination of dead guides as described herein with CRISPR applications described herein and CRISPR applications known in the art results in a highly efficient and accurate means for multiplex screening of systems (e.g. network biology). Such screening allows, for example, identification of specific combinations of gene activities for identifying genes responsible for diseases (e.g. on/off combinations), in particular gene related diseases. A preferred application of such screening is cancer. In the same manner, screening for treatment for such diseases is included in the invention. Cells or animals may be exposed to

aberrant conditions resulting in disease or disease like effects. Candidate compositions may be provided and screened for an effect in the desired multiplex environment. For example a patient's cancer cells may be screened for which gene combinations will cause them to die, and then use this information to establish appropriate therapies.

[00286] In one aspect, the invention provides a kit comprising one or more of the components described herein. The kit may include dead guides as described herein with or without guides as described herein.

[00287] The structural information provided herein allows for interrogation of dead gRNA interaction with the target DNA and the Cas9 permitting engineering or alteration of dead gRNA structure to optimize functionality of the entire Cas9 CRISPR-Cas system. For example, loops of the dead gRNA may be extended, without colliding with the Cas9 protein by the insertion of adaptor proteins that can bind to RNA. These adaptor proteins can further recruit effector proteins or fusions which comprise one or more functional domains.

[00288] In some preferred 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 (e.g. SID4X). In some embodiments, the functional domain is an epigenetic 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.

[00289] An aspect of the invention is that the above elements are comprised in a single composition or comprised in individual compositions. These compositions may advantageously be applied to a host to elicit a functional effect on the genomic level.

[00290] In general, the dead gRNA are modified in a manner that provides specific binding sites (e.g. aptamers) for adapter proteins comprising one or more functional domains (e.g. via fusion protein) to bind to. The modified dead gRNA are modified such that once the dead gRNA forms a CRISPR complex (i.e. Cas9 binding to dead gRNA and target) the adapter proteins bind and, the functional domain on the adapter protein is positioned in a spatial orientation which is advantageous for the attributed function to be effective. For example, if the functional domain is a transcription activator (e.g. VP64 or p65), the transcription activator is placed in a spatial orientation which allows it to affect the transcription of the target. Likewise, a transcription

repressor will be advantageously positioned to affect the transcription of the target and a nuclease (e.g. Fok1) will be advantageously positioned to cleave or partially cleave the target.

[00291] The skilled person will understand that modifications to the dead gRNA which allow for binding of the adapter + functional domain but not proper positioning of the adapter + functional domain (e.g. due to steric hindrance within the three dimensional structure of the CRISPR complex) are modifications which are not intended. The one or more modified dead gRNA may be modified at the tetra loop, the stem loop 1, stem loop 2, or stem loop 3, as described herein, preferably at either the tetra loop or stem loop 2, and most preferably at both the tetra loop and stem loop 2.

[00292] As explained herein the functional domains may be, for example, one or more domains from the group consisting of methylase activity, demethylase activity, transcription activation activity, transcription repression activity, transcription release factor activity, histone modification activity, RNA cleavage activity, DNA cleavage activity, nucleic acid binding activity, and molecular switches (e.g. light inducible). In some cases it is advantageous that additionally at least one NLS is provided. In some instances, it is advantageous to position the NLS at the N terminus. When more than one functional domain is included, the functional domains may be the same or different.

[00293] The dead gRNA may be designed to include multiple binding recognition sites (e.g. aptamers) specific to the same or different adapter protein. The dead gRNA may be designed to bind to the promoter region -1000 - +1 nucleic acids upstream of the transcription start site (i.e. TSS), preferably -200 nucleic acids. This positioning improves functional domains which affect gene activation (e.g. transcription activators) or gene inhibition (e.g. transcription repressors). The modified dead gRNA may be one or more modified dead gRNAs targeted to one or more target loci (e.g. at least 1 gRNA, at least 2 gRNA, at least 5 gRNA, at least 10 gRNA, at least 20 gRNA, at least 30 gRNA, at least 50 gRNA) comprised in a composition.

[00294] The adaptor protein may be any number of proteins that binds to an aptamer or recognition site introduced into the modified dead gRNA and which allows proper positioning of one or more functional domains, once the dead gRNA has been incorporated into the CRISPR complex, to affect the target with the attributed function. As explained in detail in this application such may be coat proteins, preferably bacteriophage coat proteins. The functional domains associated with such adaptor proteins (e.g. in the form of fusion protein) may include,

for example, one or more domains from the group consisting of methylase activity, demethylase activity, transcription activation activity, transcription repression activity, transcription release factor activity, histone modification activity, RNA cleavage activity, DNA cleavage activity, nucleic acid binding activity, and molecular switches (e.g. light inducible). Preferred domains are Fok1, VP64, P65, HSF1, MyoD1. In the event that the functional domain is a transcription activator or transcription repressor it is advantageous that additionally at least an NLS is provided and preferably at the N terminus. When more than one functional domain is included, the functional domains may be the same or different. The adaptor protein may utilize known linkers to attach such functional domains.

[00295] Thus, the modified dead gRNA, the (inactivated) Cas9 (with or without functional domains), and the binding protein with one or more functional domains, may each individually be comprised in a composition and administered to a host individually or collectively. Alternatively, these components may be provided in a single composition for administration to a host. Administration to a host may be performed via viral vectors known to the skilled person or described herein for delivery to a host (e.g. lentiviral vector, adenoviral vector, AAV vector). As explained herein, use of different selection markers (e.g. for lentiviral gRNA selection) and concentration of gRNA (e.g. dependent on whether multiple gRNAs are used) may be advantageous for eliciting an improved effect.

[00296] On the basis of this concept, several variations are appropriate to elicit a genomic locus event, including DNA cleavage, gene activation, or gene deactivation. Using the provided compositions, the person skilled in the art can advantageously and specifically target single or multiple loci with the same or different functional domains to elicit one or more genomic locus events. The compositions may be applied in a wide variety of methods for screening in libraries in cells and functional modeling in vivo (e.g. gene activation of lincRNA and identification of function; gain-of-function modeling; loss-of-function modeling; the use the compositions of the invention to establish cell lines and transgenic animals for optimization and screening purposes).

[00297] The current invention comprehends the use of the compositions of the current invention to establish and utilize conditional or inducible CRISPR transgenic cell /animals, which are not believed prior to the present invention or application. For example, the target cell comprises Cas9 conditionally or inducibly (e.g. in the form of Cre dependent constructs) and/or the adapter protein conditionally or inducibly and, on expression of a vector introduced into the

target cell, the vector expresses that which induces or gives rise to the condition of Cas9 expression and/or adaptor expression in the target cell. By applying the teaching and compositions of the current invention with the known method of creating a CRISPR complex, inducible genomic events affected by functional domains are also an aspect of the current invention. One example of this is the creation of a CRISPR knock-in / conditional transgenic animal (e.g. mouse comprising e.g. a Lox-Stop-polyA-Lox(LSL) cassette) and subsequent delivery of one or more compositions providing one or more modified dead gRNA (e.g. -200 nucleotides to TSS of a target gene of interest for gene activation purposes) as described herein (e.g. modified dead gRNA with one or more aptamers recognized by coat proteins, e.g. MS2), one or more adapter proteins as described herein (MS2 binding protein linked to one or more VP64) and means for inducing the conditional animal (e.g. Cre recombinase for rendering Cas9 expression inducible). Alternatively, the adaptor protein may be provided as a conditional or inducible element with a conditional or inducible Cas9 to provide an effective model for screening purposes, which advantageously only requires minimal design and administration of specific dead gRNAs for a broad number of applications.

[00298] In another aspect the dead guides are further modified to improve specificity. Protected dead guides may be synthesized, whereby secondary structure is introduced into the 3' end of the dead guide to improve its specificity. A protected guide RNA (pgRNA) comprises a guide sequence capable of hybridizing to a target sequence in a genomic locus of interest in a cell and a protector strand, wherein the protector strand is optionally complementary to the guide sequence and wherein the guide sequence may in part be hybridizable to the protector strand. The pgRNA optionally includes an extension sequence. The thermodynamics of the pgRNA-target DNA hybridization is determined by the number of bases complementary between the guide RNA and target DNA. By employing 'thermodynamic protection', specificity of dead gRNA can be improved by adding a protector sequence. For example, one method adds a complementary protector strand of varying lengths to the 3' end of the guide sequence within the dead gRNA. As a result, the protector strand is bound to at least a portion of the dead gRNA and provides for a protected gRNA (pgRNA). In turn, the dead gRNA references herein may be easily protected using the described embodiments, resulting in pgRNA. The protector strand can be either a separate RNA transcript or strand or a chimeric version joined to the 3' end of the dead gRNA guide sequence.

Tandem guides and uses in a multiplex (tandem) targeting approach

[00299] The inventors have shown that CRISPR enzymes as defined herein can employ more than one RNA guide without losing activity. This enables the use of the CRISPR enzymes, systems or complexes as defined herein for targeting multiple DNA targets, genes or gene loci, with a single enzyme, system or complex as defined herein. The guide RNAs may be tandemly arranged, optionally separated by a nucleotide sequence such as a direct repeat as defined herein. The position of the different guide RNAs is the tandem does not influence the activity. It is noted that the terms "CRISPR-Cas system", "CRISP-Cas complex" "CRISPR complex" and "CRISPR system" are used interchangeably. Also the terms "CRISPR enzyme", "Cas enzyme", or "CRISPR-Cas enzyme", can be used interchangeably. In preferred embodiments, said CRISPR enzyme, CRISP-Cas enzyme or Cas enzyme is Cas9, or any one of the modified or mutated variants thereof described herein elsewhere.

[00300] In one aspect, the invention provides a non-naturally occurring or engineered CRISPR enzyme, preferably a class 2 CRISPR enzyme, preferably a Type V or VI CRISPR enzyme as described herein, such as without limitation Cas9 as described herein elsewhere, used for tandem or multiplex targeting. It is to be understood that any of the CRISPR (or CRISPR-Cas or Cas) enzymes, complexes, or systems according to the invention as described herein elsewhere may be used in such an approach. Any of the methods, products, compositions and uses as described herein elsewhere are equally applicable with the multiplex or tandem targeting approach further detailed below. By means of further guidance, the following particular aspects and embodiments are provided.

[00301] In one aspect, the invention provides for the use of a Cas9 enzyme, complex or system as defined herein for targeting multiple gene loci. In one embodiment, this can be established by using multiple (tandem or multiplex) guide RNA (gRNA) sequences.

[00302] In one aspect, the invention provides methods for using one or more elements of a Cas9 enzyme, complex or system as defined herein for tandem or multiplex targeting, wherein said CRISP system comprises multiple guide RNA sequences. Preferably, said gRNA sequences are separated by a nucleotide sequence, such as a direct repeat as defined herein elsewhere.

[00303] The Cas9 enzyme, system or complex as defined herein provides an effective means for modifying multiple target polynucleotides. The Cas9 enzyme, system or complex as defined herein has a wide variety of utility including modifying (e.g., deleting, inserting, translocating,

inactivating, activating) one or more target polynucleotides in a multiplicity of cell types. As such the Cas9 enzyme, system or complex as defined herein of the invention has a broad spectrum of applications in, e.g., gene therapy, drug screening, disease diagnosis, and prognosis, including targeting multiple gene loci within a single CRISPR system.

[00304] In one aspect, the invention provides a Cas9 enzyme, system or complex as defined herein, i.e. a Cas9 CRISPR-Cas complex having a Cas9 protein having at least one destabilization domain associated therewith, and multiple guide RNAs that target multiple nucleic acid molecules such as DNA molecules, whereby each of said multiple guide RNAs specifically targets its corresponding nucleic acid molecule, e.g., DNA molecule. Each nucleic acid molecule target, e.g., DNA molecule can encode a gene product or encompass a gene locus. Using multiple guide RNAs hence enables the targeting of multiple gene loci or multiple genes. In some embodiments the Cas9 enzyme may cleave the DNA molecule encoding the gene product. In some embodiments expression of the gene product is altered. The Cas9 protein and the guide RNAs do not naturally occur together. The invention comprehends the guide RNAs comprising tandemly arranged guide sequences. The invention further comprehends coding sequences for the Cas9 protein being codon optimized for expression in a eukaryotic cell. In a preferred embodiment the eukaryotic cell is a mammalian cell, a plant cell or a yeast cell and in a more preferred embodiment the mammalian cell is a human cell. Expression of the gene product may be decreased. The Cas9 enzyme may form part of a CRISPR system or complex, which further comprises tandemly arranged guide RNAs (gRNAs) comprising a series of 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 25, 25, 30, or more than 30 guide sequences, each capable of specifically hybridizing to a target sequence in a genomic locus of interest in a cell. In some embodiments, the functional Cas9 CRISPR system or complex binds to the multiple target sequences. In some embodiments, the functional CRISPR system or complex may edit the multiple target sequences, e.g., the target sequences may comprise a genomic locus, and in some embodiments there may be an alteration of gene expression. In some embodiments, the functional CRISPR system or complex may comprise further functional domains. In some embodiments, the invention provides a method for altering or modifying expression of multiple gene products. The method may comprise introducing into a cell containing said target nucleic acids, e.g., DNA molecules, or containing and expressing target nucleic acid, e.g., DNA molecules; for instance, the target

nucleic acids may encode gene products or provide for expression of gene products (e.g., regulatory sequences).

[00305] In preferred embodiments the CRISPR enzyme used for multiplex targeting is Cas9, or the CRISPR system or complex comprises Cas9. In some embodiments, the CRISPR enzyme used for multiplex targeting is AsCas9, or the CRISPR system or complex used for multiplex targeting comprises an AsCas9. In some embodiments, the CRISPR enzyme is an LbCas9, or the CRISPR system or complex comprises LbCas9. In some embodiments, the Cas9 enzyme used for multiplex targeting cleaves both strands of DNA to produce a double strand break (DSB). In some embodiments, the CRISPR enzyme used for multiplex targeting is a nickase. In some embodiments, the Cas9 enzyme used for multiplex targeting is a dual nickase. In some embodiments, the Cas9 enzyme used for multiplex targeting is a Cas9 enzyme such as a DD Cas9 enzyme as defined herein elsewhere.

[00306] In embodiments, the Cas9 may be paired, for example as a pair of nickases, for example SaCas9 nickases (eSaCas9 nickases). Further, the Cas9 may be packaged with one or two or more guides on an AAV vector. This may be performed as described in Friedland AE et al, *Characterization of Staphylococcus aureus Cas9: a smaller Cas9 for all-in-one adeno-associated virus delivery and paired nickase applications*, Genome Biol. 2015 Nov 24;16:257. doi: 10.1186/s13059-015-0817-8., the disclosure of which is hereby incorporated by reference.

[00307] In some general embodiments, the Cas9 enzyme used for multiplex targeting is associated with one or more functional domains. In some more specific embodiments, the CRISPR enzyme used for multiplex targeting is a deadCas9 as defined herein elsewhere.

[00308] In an aspect, the present invention provides a means for delivering the Cas9 enzyme, system or complex for use in multiple targeting as defined herein or the polynucleotides defined herein. Non-limiting examples of such delivery means are e.g. particle(s) delivering component(s) of the complex, vector(s) comprising the polynucleotide(s) discussed herein (e.g., encoding the CRISPR enzyme, providing the nucleotides encoding the CRISPR complex). In some embodiments, the vector may be a plasmid or a viral vector such as AAV, or lentivirus. Transient transfection with plasmids, e.g., into HEK cells may be advantageous, especially given the size limitations of AAV and that while Cas9 fits into AAV, one may reach an upper limit with additional guide RNAs.

[00309] Also provided is a model that constitutively expresses the Cas9 enzyme, complex or system as used herein for use in multiplex targeting. The organism may be transgenic and may have been transfected with the present vectors or may be the offspring of an organism so transfected. In a further aspect, the present invention provides compositions comprising the CRISPR enzyme, system and complex as defined herein or the polynucleotides or vectors described herein. Also provides are Cas9 CRISPR systems or complexes comprising multiple guide RNAs, preferably in a tandemly arranged format. Said different guide RNAs may be separated by nucleotide sequences such as direct repeats.

[00310] Also provided is a method of treating a subject, e.g., a subject in need thereof, comprising inducing gene editing by transforming the subject with the polynucleotide encoding the Cas9 CRISPR system or complex or any of polynucleotides or vectors described herein and administering them to the subject. A suitable repair template may also be provided, for example delivered by a vector comprising said repair template. Also provided is a method of treating a subject, e.g., a subject in need thereof, comprising inducing transcriptional activation or repression of multiple target gene loci by transforming the subject with the polynucleotides or vectors described herein, wherein said polynucleotide or vector encodes or comprises the Cas9 enzyme, complex or system comprising multiple guide RNAs, preferably tandemly arranged. Where any treatment is occurring ex vivo, for example in a cell culture, then it will be appreciated that the term 'subject' may be replaced by the phrase "cell or cell culture."

[00311] Compositions comprising Cas9 enzyme, complex or system comprising multiple guide RNAs, preferably tandemly arranged, or the polynucleotide or vector encoding or comprising said Cas9 enzyme, complex or system comprising multiple guide RNAs, preferably tandemly arranged, for use in the methods of treatment as defined herein elsewhere are also provided. A kit of parts may be provided including such compositions. Use of said composition in the manufacture of a medicament for such methods of treatment are also provided. Use of a Cas9 CRISPR system in screening is also provided by the present invention, e.g., gain of function screens. Cells which are artificially forced to overexpress a gene are be able to down regulate the gene over time (re-establishing equilibrium) e.g. by negative feedback loops. By the time the screen starts the unregulated gene might be reduced again. Using an inducible Cas9 activator allows one to induce transcription right before the screen and therefore minimizes the

chance of false negative hits. Accordingly, by use of the instant invention in screening, e.g., gain of function screens, the chance of false negative results may be minimized.

[00312] In one aspect, the invention provides an engineered, non-naturally occurring CRISPR system comprising a Cas9 protein and multiple guide RNAs that each specifically target a DNA molecule encoding a gene product in a cell, whereby the multiple guide RNAs each target their specific DNA molecule encoding the gene product and the Cas9 protein cleaves the target DNA molecule encoding the gene product, whereby expression of the gene product is altered; and, wherein the CRISPR protein and the guide RNAs do not naturally occur together. The invention comprehends the multiple guide RNAs comprising multiple guide sequences, preferably separated by a nucleotide sequence such as a direct repeat and optionally fused to a tracr sequence. In an embodiment of the invention the CRISPR protein is a type V or VI CRISPR-Cas protein and in a more preferred embodiment the CRISPR protein is a Cas9 protein. The invention further comprehends a Cas9 protein being codon optimized for expression in a eukaryotic cell. In a preferred embodiment the eukaryotic cell is a mammalian cell and in a more preferred embodiment the mammalian cell is a human cell. In a further embodiment of the invention, the expression of the gene product is decreased.

[00313] In another aspect, the invention provides an engineered, non-naturally occurring vector system comprising one or more vectors comprising a first regulatory element operably linked to the multiple Cas9 CRISPR system guide RNAs that each specifically target a DNA molecule encoding a gene product and a second regulatory element operably linked coding for a CRISPR protein. Both regulatory elements may be located on the same vector or on different vectors of the system. The multiple guide RNAs target the multiple DNA molecules encoding the multiple gene products in a cell and the CRISPR protein may cleave the multiple DNA molecules encoding the gene products (it may cleave one or both strands or have substantially no nuclease activity), whereby expression of the multiple gene products is altered; and, wherein the CRISPR protein and the multiple guide RNAs do not naturally occur together. In a preferred embodiment the CRISPR protein is Cas9 protein, optionally codon optimized for expression in a eukaryotic cell. In a preferred embodiment the eukaryotic cell is a mammalian cell, a plant cell or a yeast cell and in a more preferred embodiment the mammalian cell is a human cell. In a further embodiment of the invention, the expression of each of the multiple gene products is altered, preferably decreased.

[00314] In one aspect, the invention provides a vector system comprising one or more vectors. In some embodiments, the system comprises: (a) a first regulatory element operably linked to a direct repeat sequence and one or more insertion sites for inserting one or more guide sequences up- or downstream (whichever applicable) of the direct repeat sequence, wherein when expressed, the one or more guide sequence(s) direct(s) sequence-specific binding of the CRISPR complex to the one or more target sequence(s) in a eukaryotic cell, wherein the CRISPR complex comprises a Cas9 enzyme complexed with the one or more guide sequence(s) that is hybridized to the one or more target sequence(s); and (b) a second regulatory element operably linked to an enzyme-coding sequence encoding said Cas9 enzyme, preferably comprising at least one nuclear localization sequence and/or at least one NES; wherein components (a) and (b) are located on the same or different vectors of the system. Where applicable, a tracr sequence may also be provided. In some embodiments, component (a) further comprises two or more guide sequences operably linked to the first regulatory element, wherein when expressed, each of the two or more guide sequences direct sequence specific binding of a Cas9 CRISPR complex to a different target sequence in a eukaryotic cell. In some embodiments, the CRISPR complex comprises one or more nuclear localization sequences and/or one or more NES of sufficient strength to drive accumulation of said Cas9 CRISPR complex in a detectable amount in or out of the nucleus of a eukaryotic cell. In some embodiments, the first regulatory element is a polymerase III promoter. In some embodiments, the second regulatory element is a polymerase II promoter. In some embodiments, each of the guide sequences is at least 16, 17, 18, 19, 20, 25 nucleotides, or between 16-30, or between 16-25, or between 16-20 nucleotides in length.

[00315] Recombinant expression vectors can comprise the polynucleotides encoding the Cas9 enzyme, system or complex for use in multiple targeting as defined herein in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory elements, which may be selected on the basis of the host cells to be used for expression, that is operatively-linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory element(s) in a manner that allows for expression of the nucleotide sequence (e.g., in an in vitro transcription/translation system or in a host cell when the vector is introduced into the host cell).

[00316] In some embodiments, a host cell is transiently or non-transiently transfected with one or more vectors comprising the polynucleotides encoding the Cas9 enzyme, system or complex for use in multiple targeting as defined herein. In some embodiments, a cell is transfected as it naturally occurs in a subject. In some embodiments, a cell that is transfected is taken from a subject. In some embodiments, the cell is derived from cells taken from a subject, such as a cell line. A wide variety of cell lines for tissue culture are known in the art and exemplified herein elsewhere. Cell lines are available from a variety of sources known to those with skill in the art (see, e.g., the American Type Culture Collection (ATCC) (Manassus, Va.)). In some embodiments, a cell transfected with one or more vectors comprising the polynucleotides encoding the Cas9 enzyme, system or complex for use in multiple targeting as defined herein is used to establish a new cell line comprising one or more vector-derived sequences. In some embodiments, a cell transiently transfected with the components of a Cas9 CRISPR system or complex for use in multiple targeting as described herein (such as by transient transfection of one or more vectors, or transfection with RNA), and modified through the activity of a Cas9 CRISPR system or complex, is used to establish a new cell line comprising cells containing the modification but lacking any other exogenous sequence. In some embodiments, cells transiently or non-transiently transfected with one or more vectors comprising the polynucleotides encoding the Cas9 enzyme, system or complex for use in multiple targeting as defined herein, or cell lines derived from such cells are used in assessing one or more test compounds.

[00317] The term "regulatory element" is as defined herein elsewhere.

[00318] Advantageous vectors include lentiviruses and adeno-associated viruses, and types of such vectors can also be selected for targeting particular types of cells.

[00319] In one aspect, the invention provides a eukaryotic host cell comprising (a) a first regulatory element operably linked to a direct repeat sequence and one or more insertion sites for inserting one or more guide RNA sequences up- or downstream (whichever applicable) of the direct repeat sequence, wherein when expressed, the guide sequence(s) direct(s) sequence-specific binding of the Cas9 CRISPR complex to the respective target sequence(s) in a eukaryotic cell, wherein the Cas9 CRISPR complex comprises a Cas9 enzyme complexed with the one or more guide sequence(s) that is hybridized to the respective target sequence(s); and/or (b) a second regulatory element operably linked to an enzyme-coding sequence encoding said

Cas9 enzyme comprising preferably at least one nuclear localization sequence and/or NES. In some embodiments, the host cell comprises components (a) and (b). Where applicable, a tracr sequence may also be provided. In some embodiments, component (a), component (b), or components (a) and (b) are stably integrated into a genome of the host eukaryotic cell. In some embodiments, component (a) further comprises two or more guide sequences operably linked to the first regulatory element, and optionally separated by a direct repeat, wherein when expressed, each of the two or more guide sequences direct sequence specific binding of a Cas9 CRISPR complex to a different target sequence in a eukaryotic cell. In some embodiments, the Cas9 enzyme comprises one or more nuclear localization sequences and/or nuclear export sequences or NES of sufficient strength to drive accumulation of said CRISPR enzyme in a detectable amount in and/or out of the nucleus of a eukaryotic cell.

[00320] In some embodiments, the Cas9 enzyme is a type V or VI CRISPR system enzyme. In some embodiments, the Cas9 enzyme is a Cas9 enzyme. In some embodiments, the Cas9 enzyme is derived from Francisella tularensis 1, Francisella tularensis subsp. novicida, Prevotella albensis, Lachnospiraceae bacterium MC2017 1, Butyrivibrio proteoclasticus, Peregrinibacteria bacterium GW2011_GWA2_33_10, Parcubacteria bacterium GW2011_GWC2_44_17, Smithella sp. SCADC, Acidaminococcus sp. BV3L6, Lachnospiraceae bacterium MA2020, Candidatus Methanoplasma termitum, Eubacterium eligens, Moraxella bovoculi 237, Leptospira inadai, Lachnospiraceae bacterium ND2006, Porphyromonas crevioricanis 3, Prevotella disiens, or Porphyromonas macacae Cas9, and may include further alterations or mutations of the Cas9 as defined herein elsewhere, and can be a chimeric Cas9. In some embodiments, the Cas9 enzyme is codon-optimized for expression in a eukaryotic cell. In some embodiments, the CRISPR enzyme directs cleavage of one or two strands at the location of the target sequence. In some embodiments, the first regulatory element is a polymerase III promoter. In some embodiments, the second regulatory element is a polymerase II promoter. In some embodiments, the one or more guide sequence(s) is (are each) at least 16, 17, 18, 19, 20, 25 nucleotides, or between 16-30, or between 16-25, or between 16-20 nucleotides in length. When multiple guide RNAs are used, they are preferably separated by a direct repeat sequence. In an aspect, the invention provides a non-human eukaryotic organism; preferably a multicellular eukaryotic organism, comprising a eukaryotic host cell according to any of the described embodiments. In other aspects, the invention provides a eukaryotic organism; preferably a

multicellular eukaryotic organism, comprising a eukaryotic host cell according to any of the described embodiments. The organism in some embodiments of these aspects may be an animal; for example a mammal. Also, the organism may be an arthropod such as an insect. The organism also may be a plant. Further, the organism may be a fungus.

[00321] In one aspect, the invention provides a kit comprising one or more of the components described herein. In some embodiments, the kit comprises a vector system and instructions for using the kit. In some embodiments, the vector system comprises (a) a first regulatory element operably linked to a direct repeat sequence and one or more insertion sites for inserting one or more guide sequences up- or downstream (whichever applicable) of the direct repeat sequence, wherein when expressed, the guide sequence directs sequence-specific binding of a Cas9 CRISPR complex to a target sequence in a eukaryotic cell, wherein the Cas9 CRISPR complex comprises a Cas9 enzyme complexed with the guide sequence that is hybridized to the target sequence; and/or (b) a second regulatory element operably linked to an enzyme-coding sequence encoding said Cas9 enzyme comprising a nuclear localization sequence. Where applicable, a tracr sequence may also be provided. In some embodiments, the kit comprises components (a) and (b) located on the same or different vectors of the system. In some embodiments, component (a) further comprises two or more guide sequences operably linked to the first regulatory element, wherein when expressed, each of the two or more guide sequences direct sequence specific binding of a CRISPR complex to a different target sequence in a eukaryotic cell. In some embodiments, the Cas9 enzyme comprises one or more nuclear localization sequences of sufficient strength to drive accumulation of said CRISPR enzyme in a detectable amount in the nucleus of a eukaryotic cell. In some embodiments, the CRISPR enzyme is a type V or VI CRISPR system enzyme. In some embodiments, the CRISPR enzyme is a Cas9 enzyme. In some embodiments, the Cas9 enzyme is derived from Francisella tularensis 1, Francisella tularensis subsp. novicida, Prevotella albensis, Lachnospiraceae bacterium MC2017 1, Butyrivibrio proteoclasticus, Peregrinibacteria bacterium GW2011_GWA2_33_10, Parcubacteria bacterium GW2011_GWC2_44_17, Smithella sp. SCADC, Acidaminococcus sp. BV3L6, Lachnospiraceae bacterium MA2020, Candidatus Methanoplasma termitum, Eubacterium eligens, Moraxella bovoculi 237, Leptospira inadai, Lachnospiraceae bacterium ND2006, Porphyromonas crevioricanis 3, Prevotella disiens, or Porphyromonas macacae Cas9 (e.g., modified to have or be associated with at least one DD), and may include further alteration

or mutation of the Cas9, and can be a chimeric Cas9. In some embodiments, the DD-CRISPR enzyme is codon-optimized for expression in a eukaryotic cell. In some embodiments, the DD-CRISPR enzyme directs cleavage of one or two strands at the location of the target sequence. In some embodiments, the DD-CRISPR enzyme lacks or substantially DNA strand cleavage activity (e.g., no more than 5% nuclease activity as compared with a wild type enzyme or enzyme not having the mutation or alteration that decreases nuclease activity). In some embodiments, the first regulatory element is a polymerase III promoter. In some embodiments, the second regulatory element is a polymerase II promoter. In some embodiments, the guide sequence is at least 16, 17, 18, 19, 20, 25 nucleotides, or between 16-30, or between 16-25, or between 16-20 nucleotides in length.

[00322] In one aspect, the invention provides a method of modifying multiple target polynucleotides in a host cell such as a eukaryotic cell. In some embodiments, the method comprises allowing a Cas9CRISPR complex to bind to multiple target polynucleotides, e.g., to effect cleavage of said multiple target polynucleotides, thereby modifying multiple target polynucleotides, wherein the Cas9CRISPR complex comprises a Cas9 enzyme complexed with multiple guide sequences each of the being hybridized to a specific target sequence within said target polynucleotide, wherein said multiple guide sequences are linked to a direct repeat sequence. Where applicable, a tracr sequence may also be provided (e.g. to provide a single guide RNA, sgRNA). In some embodiments, said cleavage comprises cleaving one or two strands at the location of each of the target sequence by said Cas9 enzyme. In some embodiments, said cleavage results in decreased transcription of the multiple target genes. In some embodiments, the method further comprises repairing one or more of 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 one or more of said target polynucleotides. In some embodiments, said mutation results in one or more amino acid changes in a protein expressed from a gene comprising one or more of the target sequence(s). 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 Cas9 enzyme and the multiple guide RNA sequence linked to a direct repeat sequence. Where applicable, a tracr sequence may also be provided. 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.

[00323] In one aspect, the invention provides a method of modifying expression of multiple polynucleotides in a eukaryotic cell. In some embodiments, the method comprises allowing a Cas9 CRISPR complex to bind to multiple polynucleotides such that said binding results in increased or decreased expression of said polynucleotides; wherein the Cas9 CRISPR complex comprises a Cas9 enzyme complexed with multiple guide sequences each specifically hybridized to its own target sequence within said polynucleotide, wherein said guide sequences are linked to a direct repeat sequence. Where applicable, a tracr sequence may also be provided. 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 Cas9 enzyme and the multiple guide sequences linked to the direct repeat sequences. Where applicable, a tracr sequence may also be provided.

[00324] In one aspect, the invention provides a recombinant polynucleotide comprising multiple guide RNA sequences up- or downstream (whichever applicable) of a direct repeat sequence, wherein each of the guide sequences when expressed directs sequence-specific binding of a Cas9CRISPR complex to its corresponding target sequence present in a eukaryotic cell. In some embodiments, the target sequence is a viral sequence present in a eukaryotic cell. Where applicable, a tracr sequence may also be provided. In some embodiments, the target sequence is a proto-oncogene or an oncogene.

[00325] Aspects of the invention encompass a non-naturally occurring or engineered composition that may comprise a guide RNA (gRNA) comprising a guide sequence capable of hybridizing to a target sequence in a genomic locus of interest in a cell and a Cas9 enzyme as defined herein that may comprise at least one or more nuclear localization sequences.

[00326] An aspect of the invention encompasses methods of modifying a genomic locus of interest to change gene expression in a cell by introducing into the cell any of the compositions described herein.

[00327] An aspect of the invention is that the above elements are comprised in a single composition or comprised in individual compositions. These compositions may advantageously be applied to a host to elicit a functional effect on the genomic level.

[00328] As used herein, the term "guide RNA" or "gRNA" has the leaning as used herein elsewhere and comprises any polynucleotide sequence having sufficient complementarity with a target nucleic acid sequence to hybridize with the target nucleic acid sequence and direct sequence-specific binding of a nucleic acid-targeting complex to the target nucleic acid sequence. Each gRNA may be designed to include multiple binding recognition sites (e.g., aptamers) specific to the same or different adapter protein. Each gRNA may be designed to bind to the promoter region -1000 - +1 nucleic acids upstream of the transcription start site (i.e. TSS), preferably -200 nucleic acids. This positioning improves functional domains which affect gene activation (e.g., transcription activators) or gene inhibition (e.g., transcription repressors). The modified gRNA may be one or more modified gRNAs targeted to one or more target loci (e.g., at least 1 gRNA, at least 2 gRNA, at least 5 gRNA, at least 10 gRNA, at least 20 gRNA, at least 30 g RNA, at least 50 gRNA) comprised in a composition. Said multiple gRNA sequences can be tandemly arranged and are preferably separated by a direct repeat.

[00329] Thus, gRNA, the CRISPR enzyme as defined herein may each individually be comprised in a composition and administered to a host individually or collectively. Alternatively, these components may be provided in a single composition for administration to a host. Administration to a host may be performed via viral vectors known to the skilled person or described herein for delivery to a host (e.g., lentiviral vector, adenoviral vector, AAV vector). As explained herein, use of different selection markers (e.g., for lentiviral sgRNA selection) and concentration of gRNA (e.g., dependent on whether multiple gRNAs are used) may be advantageous for eliciting an improved effect. On the basis of this concept, several variations are appropriate to elicit a genomic locus event, including DNA cleavage, gene activation, or gene deactivation. Using the provided compositions, the person skilled in the art can advantageously and specifically target single or multiple loci with the same or different functional domains to elicit one or more genomic locus events. The compositions may be applied in a wide variety of methods for screening in libraries in cells and functional modeling in vivo (e.g., gene activation of lincRNA and identification of function; gain-of-function modeling; loss-of-function modeling;

the use the compositions of the invention to establish cell lines and transgenic animals for optimization and screening purposes).

[00330] The current invention comprehends the use of the compositions of the current invention to establish and utilize conditional or inducible CRISPR transgenic cell /animals; *see, e.g.*, Platt et al., Cell (2014), 159(2): 440-455, or PCT patent publications cited herein, such as WO 2014/093622 (PCT/US2013/074667). For example, cells or animals such as non-human animals, e.g., vertebrates or mammals, such as rodents, e.g., mice, rats, or other laboratory or field animals, e.g., cats, dogs, sheep, etc., may be 'knock-in' whereby the animal conditionally or inducibly expresses Cas9 akin to Platt et al. The target cell or animal thus comprises the CRISPR enzyme (e.g., Cas9) conditionally or inducibly (e.g., in the form of Cre dependent constructs), on expression of a vector introduced into the target cell, the vector expresses that which induces or gives rise to the condition of the CRISPR enzyme (e.g., Cas9) expression in the target cell. By applying the teaching and compositions as defined herein with the known method of creating a CRISPR complex, inducible genomic events are also an aspect of the current invention. Examples of such inducible events have been described herein elsewhere.

[00331] In some embodiments, phenotypic alteration is preferably the result of genome modification when a genetic disease is targeted, especially in methods of therapy and preferably where a repair template is provided to correct or alter the phenotype.

[00332] In some embodiments diseases that may be targeted include those concerned with disease-causing splice defects.

[00333] In some embodiments, cellular targets include Hemopoietic Stem/Progenitor Cells (CD34+); Human T cells; and Eye (retinal cells) – for example photoreceptor precursor cells.

[00334] In some embodiments Gene targets include: Human Beta Globin – HBB (for treating Sickle Cell Anemia, including by stimulating gene-conversion (using closely related HBD gene as an endogenous template)); CD3 (T-Cells); and CEP920 - retina (eye).

[00335] In some embodiments disease targets also include: cancer; Sickle Cell Anemia (based on a point mutation); HBV, HIV; Beta-Thalassemia; and ophthalmic or ocular disease – for example Leber Congenital Amaurosis (LCA)-causing Splice Defect.

In some embodiments delivery methods include: Cationic Lipid Mediated "direct" delivery of Enzyme-Guide complex (RiboNucleoProtein) and electroporation of plasmid DNA.

[00336] Methods, products and uses described herein may be used for non-therapeutic purposes. Furthermore, any of the methods described herein may be applied in vitro and ex vivo.

[00337] In an aspect, provided is a non-naturally occurring or engineered composition comprising:

I. two or more CRISPR-Cas system polynucleotide sequences comprising

(a) a first guide sequence capable of hybridizing to a first target sequence in a polynucleotide locus,

(b) a second guide sequence capable of hybridizing to a second target sequence in a polynucleotide locus,

(c) a direct repeat sequence,

and

II. a Cas9 enzyme or a second polynucleotide sequence encoding it,

wherein when transcribed, the first and the second guide sequences direct sequence-specific binding of a first and a second Cas9 CRISPR complex to the first and second target sequences respectively,

wherein the first CRISPR complex comprises the Cas9 enzyme complexed with the first guide sequence that is hybridizable to the first target sequence,

wherein the second CRISPR complex comprises the Cas9 enzyme complexed with the second guide sequence that is hybridizable to the second target sequence, 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 or non-animal organism. Similarly, compositions comprising more than two guide RNAs can be envisaged e.g. each specific for one target, and arranged tandemly in the composition or CRISPR system or complex as described herein.

[00338] In another embodiment, the Cas9 is delivered into the cell as a protein. In another and particularly preferred embodiment, the Cas9 is delivered into the cell as a protein or as a nucleotide sequence encoding it. Delivery to the cell as a protein may include delivery of a Ribonucleoprotein (RNP) complex, where the protein is complexed with the multiple guides.

[00339] In an aspect, host cells and cell lines modified by or comprising the compositions, systems or modified enzymes of present invention are provided, including stem cells, and progeny thereof.

[00340] In an aspect, methods of cellular therapy are provided, where, for example, a single cell or a population of cells is sampled or cultured, wherein that cell or cells is or has been modified *ex vivo* as described herein, and is then re-introduced (sampled cells) or introduced (cultured cells) into the organism. Stem cells, whether embryonic or induce pluripotent or totipotent stem cells, are also particularly preferred in this regard. But, of course, *in vivo* embodiments are also envisaged.

[00341] Inventive methods can further comprise delivery of templates, such as repair templates, which may be dsODN or ssODN, see below. Delivery of templates may be via the cotemporaneous or separate from delivery of any or all the CRISPR enzyme or guide RNAs and via the same delivery mechanism or different. In some embodiments, it is preferred that the template is delivered together with the guide RNAs and, preferably, also the CRISPR enzyme. An example may be an AAV vector where the CRISPR enzyme is AsCas9 or LbCas9.

[00342] Inventive methods can further comprise: (a) delivering to the cell a double-stranded oligodeoxynucleotide (dsODN) comprising overhangs complimentary to the overhangs created by said double strand break, wherein said dsODN is integrated into the locus of interest; or –(b) delivering to the cell a single-stranded oligodeoxynucleotide (ssODN), wherein said ssODN acts as a template for homology directed repair of said double strand break. Inventive methods can be for the prevention or treatment of disease in an individual, optionally wherein said disease is caused by a defect in said locus of interest. Inventive methods can be conducted *in vivo* in the individual or *ex vivo* on a cell taken from the individual, optionally wherein said cell is returned to the individual.

[00343] The invention also comprehends products obtained from using CRISPR enzyme or Cas enzyme or Cas9 enzyme or CRISPR-CRISPR enzyme or CRISPR-Cas system or CRISPR-Cas9 system for use in tandem or multiple targeting as defined herein.

Escorted guides for the Cas9 CRISPR-Cas system according to the invention

[00344] In one aspect the invention provides escorted Cas9 CRISPR-Cas systems or complexes, especially such a system involving an escorted Cas9 CRISPR-Cas system guide. By

"escorted" is meant that the Cas9 CRISPR-Cas system or complex or guide is delivered to a selected time or place within a cell, so that activity of the Cas9 CRISPR-Cas system or complex or guide is spatially or temporally controlled. For example, the activity and destination of the Cas9 CRISPR-Cas system or complex or guide may be controlled by an escort RNA aptamer sequence that has binding affinity for an aptamer ligand, such as a cell surface protein or other localized cellular component. Alternatively, the escort aptamer may for example be responsive to an aptamer effector on or in the cell, such as a transient effector, such as an external energy source that is applied to the cell at a particular time.

[00345] The escorted Cas9 CRISPR-Cas systems or complexes have a gRNA with a functional structure designed to improve gRNA structure, architecture, stability, genetic expression, or any combination thereof. Such a structure can include an aptamer.

[00346] Aptamers are biomolecules that can be designed or selected to bind tightly to other ligands, for example using a technique called systematic evolution of ligands by exponential enrichment (SELEX; Tuerk C, Gold L: "Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase." Science 1990, 249:505-510). Nucleic acid aptamers can for example be selected from pools of random-sequence oligonucleotides, with high binding affinities and specificities for a wide range of biomedically relevant targets, suggesting a wide range of therapeutic utilities for aptamers (Keefe, Anthony D., Supriya Pai, and Andrew Ellington. "Aptamers as therapeutics." Nature Reviews Drug Discovery 9.7 (2010): 537-550). These characteristics also suggest a wide range of uses for aptamers as drug delivery vehicles (Levy-Nissenbaum, Etgar, et al. "Nanotechnology and aptamers: applications in drug delivery." Trends in biotechnology 26.8 (2008): 442-449; and, Hicke BJ, Stephens AW. "Escort aptamers: a delivery service for diagnosis and therapy." J Clin Invest 2000, 106:923-928.). Aptamers may also be constructed that function as molecular switches, responding to a que by changing properties, such as RNA aptamers that bind fluorophores to mimic the activity of green fluorescent protein (Paige, Jeremy S., Karen Y. Wu, and Samie R. Jaffrey. "RNA mimics of green fluorescent protein." Science 333.6042 (2011): 642-646). It has also been suggested that aptamers may be used as components of targeted siRNA therapeutic delivery systems, for example targeting cell surface proteins (Zhou, Jiehua, and John J. Rossi. "Aptamer-targeted cell-specific RNA interference." Silence 1.1 (2010): 4).

[00347] Accordingly, provided herein is a gRNA modified, e.g., by one or more aptamer(s) designed to improve gRNA delivery, including delivery across the cellular membrane, to intracellular compartments, or into the nucleus. Such a structure can include, either in addition to the one or more aptamer(s) or without such one or more aptamer(s), moiety(ies) so as to render the guide deliverable, inducible or responsive to a selected effector. The invention accordingly comprehends an gRNA that responds to normal or pathological physiological conditions, including without limitation pH, hypoxia, O_2 concentration, temperature, protein concentration, enzymatic concentration, lipid structure, light exposure, mechanical disruption (e.g. ultrasound waves), magnetic fields, electric fields, or electromagnetic radiation.

[00348] An aspect of the invention provides non-naturally occurring or engineered composition comprising an escorted guide RNA (egRNA) comprising:

an RNA guide sequence capable of hybridizing to a target sequence in a genomic locus of interest in a cell; and,

an escort RNA aptamer sequence, wherein the escort aptamer has binding affinity for an aptamer ligand on or in the cell, or the escort aptamer is responsive to a localized aptamer effector on or in the cell, wherein the presence of the aptamer ligand or effector on or in the cell is spatially or temporally restricted.

[00349] The escort aptamer may for example change conformation in response to an interaction with the aptamer ligand or effector in the cell.

[00350] The escort aptamer may have specific binding affinity for the aptamer ligand.

[00351] The aptamer ligand may be localized in a location or compartment of the cell, for example on or in a membrane of the cell. Binding of the escort aptamer to the aptamer ligand may accordingly direct the egRNA to a location of interest in the cell, such as the interior of the cell by way of binding to an aptamer ligand that is a cell surface ligand. In this way, a variety of spatially restricted locations within the cell may be targeted, such as the cell nucleus or mitochondria.

[00352] Once intended alterations have been introduced, such as by editing intended copies of a gene in the genome of a cell, continued CRISPR/Cas9 expression in that cell is no longer necessary. Indeed, sustained expression would be undesirable in certain casein case of off-target effects at unintended genomic sites, etc. Thus time-limited expression would be useful. Inducible expression offers one approach, but in addition Applicants have engineered a Self-

Inactivating Cas9 CRISPR-Cas system that relies on the use of a non-coding guide target sequence within the CRISPR vector itself. Thus, after expression begins, the CRISPR system will lead to its own destruction, but before destruction is complete it will have time to edit the genomic copies of the target gene (which, with a normal point mutation in a diploid cell, requires at most two edits). Simply, the self inactivating Cas9 CRISPR-Cas system includes additional RNA (i.e., guide RNA) that targets the coding sequence for the CRISPR enzyme itself or that targets one or more non-coding guide target sequences complementary to unique sequences present in one or more of the following: (a) within the promoter driving expression of the non-coding RNA elements, (b) within the promoter driving expression of the Cas9 gene, (c) within 100bp of the ATG translational start codon in the Cas9 coding sequence, (d) within the inverted terminal repeat (iTR) of a viral delivery vector, e.g., in an AAV genome.

[00353] The egRNA may include an RNA aptamer linking sequence, operably linking the escort RNA sequence to the RNA guide sequence.

[00354] In embodiments, the egRNA may include one or more photolabile bonds or non-naturally occurring residues.

[00355] In one aspect, the escort RNA aptamer sequence may be complementary to a target miRNA, which may or may not be present within a cell, so that only when the target miRNA is present is there binding of the escort RNA aptamer sequence to the target miRNA which results in cleavage of the egRNA by an RNA-induced silencing complex (RISC) within the cell.

[00356] In embodiments, the escort RNA aptamer sequence may for example be from 10 to 200 nucleotides in length, and the egRNA may include more than one escort RNA aptamer sequence.

[00357] It is to be understood that any of the RNA guide sequences as described herein elsewhere can be used in the egRNA described herein. In certain embodiments of the invention, the guide RNA or mature crRNA comprises, consists essentially of, or consists of a direct repeat sequence and a guide sequence or spacer sequence. In certain embodiments, the guide RNA or mature crRNA comprises, consists essentially of, or consists of a direct repeat sequence linked to a guide sequence or spacer sequence. In certain embodiments the guide RNA or mature crRNA comprises 19 nts of partial direct repeat followed by 23-25 nt of guide sequence or spacer sequence. In certain embodiments, the effector protein is a FnCas9 effector protein and requires at least 16 nt of guide sequence to achieve detectable DNA cleavage and a minimum of 17 nt of

guide sequence to achieve efficient DNA cleavage *in vitro*. In certain embodiments, the direct repeat sequence is located upstream (i.e., 5') from the guide sequence or spacer sequence. In a preferred embodiment the seed sequence (i.e. the sequence essential critical for recognition and/or hybridization to the sequence at the target locus) of the FnCas9 guide RNA is approximately within the first 5 nt on the 5' end of the guide sequence or spacer sequence.

[00358] The egRNA may be included in a non-naturally occurring or engineered Cas9 CRISPR-Cas complex composition, together with a Cas9 which may include at least one mutation, for example a mutation so that the Cas9 has no more than 5% of the nuclease activity of a Cas9 not having the at least one mutation, for example having a diminished nuclease activity of at least 97%, or 100% as compared with the Cas9 not having the at least one mutation. The Cas9 may also include one or more nuclear localization sequences. Mutated Cas9 enzymes having modulated activity such as diminished nuclease activity are described herein elsewhere.

[00359] The engineered Cas9 CRISPR-Cas composition may be provided in a cell, such as a eukaryotic cell, a mammalian cell, or a human cell.

[00360] In embodiments, the compositions described herein comprise a Cas9 CRISPR-Cas complex having at least three functional domains, at least one of which is associated with Cas9 and at least two of which are associated with egRNA.

[00361] The compositions described herein may be used to introduce a genomic locus event in a host cell, such as an eukaryotic cell, in particular a mammalian cell, or a non-human eukaryote, in particular a non-human mammal such as a mouse, in vivo. The genomic locus event may comprise affecting gene activation, gene inhibition, or cleavage in a locus. The compositions described herein may also be used to modify a genomic locus of interest to change gene expression in a cell. Methods of introducing a genomic locus event in a host cell using the Cas9 enzyme provided herein are described herein in detail elsewhere. Delivery of the composition may for example be by way of delivery of a nucleic acid molecule(s) coding for the composition, which nucleic acid molecule(s) is operatively linked to regulatory sequence(s), and expression of the nucleic acid molecule(s) in vivo, for example by way of a lentivirus, an adenovirus, or an AAV.

[00362] The present invention provides compositions and methods by which gRNA-mediated gene editing activity can be adapted. The invention provides gRNA secondary structures that

improve cutting efficiency by increasing gRNA and/or increasing the amount of RNA delivered into the cell. The gRNA may include light labile or inducible nucleotides.

[00363] To increase the effectiveness of gRNA, for example gRNA delivered with viral or non-viral technologies, Applicants added secondary structures into the gRNA that enhance its stability and improve gene editing. Separately, to overcome the lack of effective delivery, Applicants modified gRNAs with cell penetrating RNA aptamers; the aptamers bind to cell surface receptors and promote the entry of gRNAs into cells. Notably, the cell-penetrating aptamers can be designed to target specific cell receptors, in order to mediate cell-specific delivery. Applicants also have created guides that are inducible.

[00364] Light responsiveness of an inducible system may be achieved via the activation and binding of cryptochrome-2 and CIB1. Blue light stimulation induces an activating conformational change in cryptochrome-2, resulting in recruitment of its binding partner CIB1. This binding is fast and reversible, achieving saturation in <15 sec following pulsed stimulation and returning to baseline <15 min after the end of stimulation. These rapid binding kinetics result in a system temporally bound only by the speed of transcription/translation and transcript/protein degradation, rather than uptake and clearance of inducing agents. Crytochrome-2 activation is also highly sensitive, allowing for the use of low light intensity stimulation and mitigating the risks of phototoxicity. Further, in a context such as the intact mammalian brain, variable light intensity may be used to control the size of a stimulated region, allowing for greater precision than vector delivery alone may offer.

[00365] The invention contemplates energy sources such as electromagnetic radiation, sound energy or thermal energy to induce the guide. Advantageously, the electromagnetic radiation is a component of visible light. In a preferred embodiment, the light is a blue light with a wavelength of about 450 to about 495 nm. In an especially preferred embodiment, the wavelength is about 488 nm. In another preferred embodiment, the light stimulation is via pulses. The light power may range from about 0-9 mW/cm^2. In a preferred embodiment, a stimulation paradigm of as low as 0.25 sec every 15 sec should result in maximal activation.

[00366] Cells involved in the practice of the present invention may be a prokaryotic cell or a eukaryotic cell, advantageously an animal cell a plant cell or a yeast cell, more advantageously a mammalian cell.

[00367] The chemical or energy sensitive guide may undergo a conformational change upon induction by the binding of a chemical source or by the energy allowing it act as a guide and have the Cas9 CRISPR-Cas system or complex function. The invention can involve applying the chemical source or energy so as to have the guide function and the Cas9 CRISPR-Cas system or complex function; and optionally further determining that the expression of the genomic locus is altered.

[00368] There are several different designs of this chemical inducible system: 1. ABI-PYL based system inducible by Abscisic Acid (ABA) (see, e.g., http://stke.sciencemag.org/cgi/content/abstract/sigtrans;4/164/rs2), 2. FKBP-FRB based system inducible by rapamycin (or related chemicals based on rapamycin) (see, e.g., http://www.nature.com/nmeth/journal/v2/n6/full/nmeth763.html), 3. GID1-GAI based system inducible by Gibberellin (GA) (see, e.g., http://www.nature.com/nchembio/journal/v8/n5/full/nchembio.922.html).

[00369] Another system contemplated by the present invention is a chemical inducible system based on change in sub-cellular localization. Applicants also developed a system in which the polypeptide include a DNA binding domain comprising at least five or more Transcription activator-like effector (TALE) monomers and at least one or more half-monomers specifically ordered to target the genomic locus of interest linked to at least one or more effector domains are further linker to a chemical or energy sensitive protein. This protein will lead to a change in the sub-cellular localization of the entire polypeptide (i.e. transportation of the entire polypeptide from cytoplasm into the nucleus of the cells) upon the binding of a chemical or energy transfer to the chemical or energy sensitive protein. This transportation of the entire polypeptide from one sub-cellular compartments or organelles, in which its activity is sequestered due to lack of substrate for the effector domain, into another one in which the substrate is present would allow the entire polypeptide to come in contact with its desired substrate (i.e. genomic DNA in the mammalian nucleus) and result in activation or repression of target gene expression.

[00370] This type of system could also be used to induce the cleavage of a genomic locus of interest in a cell when the effector domain is a nuclease.

[00371] A chemical inducible system can be an estrogen receptor (ER) based system inducible by 4-hydroxytamoxifen (4OHT) (see, e.g., http://www.pnas.org/content/104/3/1027.abstract). A mutated ligand-binding domain of the estrogen receptor called ERT2 translocates into the

nucleus of cells upon binding of 4-hydroxytamoxifen. In further embodiments of the invention any naturally occurring or engineered derivative of any nuclear receptor, thyroid hormone receptor, retinoic acid receptor, estrogen receptor, estrogen-related receptor, glucocorticoid receptor, progesterone receptor, androgen receptor may be used in inducible systems analogous to the ER based inducible system.

[00372] Another inducible system is based on the design using Transient receptor potential (TRP) ion channel based system inducible by energy, heat or radio-wave (see, e.g., http://www.sciencemag.org/content/336/6081/604). These TRP family proteins respond to different stimuli, including light and heat. When this protein is activated by light or heat, the ion channel will open and allow the entering of ions such as calcium into the plasma membrane. This influx of ions will bind to intracellular ion interacting partners linked to a polypeptide including the guide and the other components of the Cas9 CRISPR-Cas complex or system, and the binding will induce the change of sub-cellular localization of the polypeptide, leading to the entire polypeptide entering the nucleus of cells. Once inside the nucleus, the guide protein and the other components of the Cas9 CRISPR-Cas complex will be active and modulating target gene expression in cells.

[00373] This type of system could also be used to induce the cleavage of a genomic locus of interest in a cell; and, in this regard, it is noted that the Cas9 enzyme is a nuclease. The light could be generated with a laser or other forms of energy sources. The heat could be generated by raise of temperature results from an energy source, or from nano-particles that release heat after absorbing energy from an energy source delivered in the form of radio-wave.

[00374] While light activation may be an advantageous embodiment, sometimes it may be disadvantageous especially for *in vivo* applications in which the light may not penetrate the skin or other organs. In this instance, other methods of energy activation are contemplated, in particular, electric field energy and/or ultrasound which have a similar effect.

[00375] Electric field energy is preferably administered substantially as described in the art, using one or more electric pulses of from about 1 Volt/cm to about 10 kVolts/cm under *in vivo* conditions. Instead of or in addition to the pulses, the electric field may be delivered in a continuous manner. The electric pulse may be applied for between 1 μs and 500 milliseconds, preferably between 1 μs and 100 milliseconds. The electric field may be applied continuously or in a pulsed manner for 5 about minutes.

[00376] As used herein, 'electric field energy' is the electrical energy to which a cell is exposed. Preferably the electric field has a strength of from about 1 Volt/cm to about 10 kVolts/cm or more under *in vivo* conditions (see WO97/49450).

[00377] As used herein, the term "electric field" includes one or more pulses at variable capacitance and voltage and including exponential and/or square wave and/or modulated wave and/or modulated square wave forms. References to electric fields and electricity should be taken to include reference the presence of an electric potential difference in the environment of a cell. Such an environment may be set up by way of static electricity, alternating current (AC), direct current (DC), etc, as known in the art. The electric field may be uniform, non-uniform or otherwise, and may vary in strength and/or direction in a time dependent manner.

[00378] Single or multiple applications of electric field, as well as single or multiple applications of ultrasound are also possible, in any order and in any combination. The ultrasound and/or the electric field may be delivered as single or multiple continuous applications, or as pulses (pulsatile delivery).

[00379] Electroporation has been used in both *in vitro* and *in vivo* procedures to introduce foreign material into living cells. With *in vitro* applications, a sample of live cells is first mixed with the agent of interest and placed between electrodes such as parallel plates. Then, the electrodes apply an electrical field to the cell/implant mixture. Examples of systems that perform *in vitro* electroporation include the Electro Cell Manipulator ECM600 product, and the Electro Square Porator T820, both made by the BTX Division of Genetronics, Inc (see U.S. Pat. No 5,869,326).

[00380] The known electroporation techniques (both *in vitro* and *in vivo*) function by applying a brief high voltage pulse to electrodes positioned around the treatment region. The electric field generated between the electrodes causes the cell membranes to temporarily become porous, whereupon molecules of the agent of interest enter the cells. In known electroporation applications, this electric field comprises a single square wave pulse on the order of 1000 V/cm, of about 100 .mu.s duration. Such a pulse may be generated, for example, in known applications of the Electro Square Porator T820.

[00381] Preferably, the electric field has a strength of from about 1 V/cm to about 10 kV/cm under *in vitro* conditions. Thus, the electric field may have a strength of 1 V/cm, 2 V/cm, 3 V/cm, 4 V/cm, 5 V/cm, 6 V/cm, 7 V/cm, 8 V/cm, 9 V/cm, 10 V/cm, 20 V/cm, 50 V/cm, 100

V/cm, 200 V/cm, 300 V/cm, 400 V/cm, 500 V/cm, 600 V/cm, 700 V/cm, 800 V/cm, 900 V/cm, 1 kV/cm, 2 kV/cm, 5 kV/cm, 10 kV/cm, 20 kV/cm, 50 kV/cm or more. More preferably from about 0.5 kV/cm to about 4.0 kV/cm under *in vitro* conditions. Preferably the electric field has a strength of from about 1 V/cm to about 10 kV/cm under *in vivo* conditions. However, the electric field strengths may be lowered where the number of pulses delivered to the target site are increased. Thus, pulsatile delivery of electric fields at lower field strengths is envisaged.

[00382] Preferably the application of the electric field is in the form of multiple pulses such as double pulses of the same strength and capacitance or sequential pulses of varying strength and/or capacitance. As used herein, the term "pulse" includes one or more electric pulses at variable capacitance and voltage and including exponential and/or square wave and/or modulated wave/square wave forms.

[00383] Preferably the electric pulse is delivered as a waveform selected from an exponential wave form, a square wave form, a modulated wave form and a modulated square wave form.

[00384] A preferred embodiment employs direct current at low voltage. Thus, Applicants disclose the use of an electric field which is applied to the cell, tissue or tissue mass at a field strength of between 1V/cm and 20V/cm, for a period of 100 milliseconds or more, preferably 15 minutes or more.

[00385] Ultrasound is advantageously administered at a power level of from about 0.05 W/cm^2 to about 100 W/cm^2. Diagnostic or therapeutic ultrasound may be used, or combinations thereof.

[00386] As used herein, the term "ultrasound" refers to a form of energy which consists of mechanical vibrations the frequencies of which are so high they are above the range of human hearing. Lower frequency limit of the ultrasonic spectrum may generally be taken as about 20 kHz. Most diagnostic applications of ultrasound employ frequencies in the range 1 and 15 MHz' (From Ultrasonics in Clinical Diagnosis, P. N. T. Wells, ed., 2nd. Edition, Publ. Churchill Livingstone [Edinburgh, London & NY, 1977]).

[00387] Ultrasound has been used in both diagnostic and therapeutic applications. When used as a diagnostic tool ("diagnostic ultrasound"), ultrasound is typically used in an energy density range of up to about 100 mW/cm^2 (FDA recommendation), although energy densities of up to 750 mW/cm^2 have been used. In physiotherapy, ultrasound is typically used as an energy source in a range up to about 3 to 4 W/cm^2 (WHO recommendation). In other therapeutic applications,

higher intensities of ultrasound may be employed, for example, HIFU at 100 W/cm up to 1 kW/cm^2 (or even higher) for short periods of time. The term "ultrasound" as used in this specification is intended to encompass diagnostic, therapeutic and focused ultrasound.

[00388] Focused ultrasound (FUS) allows thermal energy to be delivered without an invasive probe (see Morocz et al 1998 Journal of Magnetic Resonance Imaging Vol.8, No. 1, pp.136-142. Another form of focused ultrasound is high intensity focused ultrasound (HIFU) which is reviewed by Moussatov et al in Ultrasonics (1998) Vol.36, No.8, pp.893-900 and TranHuuHue et al in Acustica (1997) Vol.83, No.6, pp.1103-1106.

[00389] Preferably, a combination of diagnostic ultrasound and a therapeutic ultrasound is employed. This combination is not intended to be limiting, however, and the skilled reader will appreciate that any variety of combinations of ultrasound may be used. Additionally, the energy density, frequency of ultrasound, and period of exposure may be varied.

[00390] Preferably the exposure to an ultrasound energy source is at a power density of from about 0.05 to about 100 Wcm^{-2}. Even more preferably, the exposure to an ultrasound energy source is at a power density of from about 1 to about 15 Wcm^{-2}.

[00391] Preferably the exposure to an ultrasound energy source is at a frequency of from about 0.015 to about 10.0 MHz. More preferably the exposure to an ultrasound energy source is at a frequency of from about 0.02 to about 5.0 MHz or about 6.0 MHz. Most preferably, the ultrasound is applied at a frequency of 3 MHz.

[00392] Preferably the exposure is for periods of from about 10 milliseconds to about 60 minutes. Preferably the exposure is for periods of from about 1 second to about 5 minutes. More preferably, the ultrasound is applied for about 2 minutes. Depending on the particular target cell to be disrupted, however, the exposure may be for a longer duration, for example, for 15 minutes.

[00393] Advantageously, the target tissue is exposed to an ultrasound energy source at an acoustic power density of from about 0.05 Wcm^{-2} to about 10 Wcm^{-2} with a frequency ranging from about 0.015 to about 10 MHz (see WO 98/52609). However, alternatives are also possible, for example, exposure to an ultrasound energy source at an acoustic power density of above 100 Wcm^{-2}, but for reduced periods of time, for example, 1000 Wcm^{-2} for periods in the millisecond range or less.

[00394] Preferably the application of the ultrasound is in the form of multiple pulses; thus, both continuous wave and pulsed wave (pulsatile delivery of ultrasound) may be employed in any combination. For example, continuous wave ultrasound may be applied, followed by pulsed wave ultrasound, or vice versa. This may be repeated any number of times, in any order and combination. The pulsed wave ultrasound may be applied against a background of continuous wave ultrasound, and any number of pulses may be used in any number of groups.

[00395] Preferably, the ultrasound may comprise pulsed wave ultrasound. In a highly preferred embodiment, the ultrasound is applied at a power density of 0.7 Wcm^{-2} or 1.25 Wcm^{-2} as a continuous wave. Higher power densities may be employed if pulsed wave ultrasound is used.

[00396] Use of ultrasound is advantageous as, like light, it may be focused accurately on a target. Moreover, ultrasound is advantageous as it may be focused more deeply into tissues unlike light. It is therefore better suited to whole-tissue penetration (such as but not limited to a lobe of the liver) or whole organ (such as but not limited to the entire liver or an entire muscle, such as the heart) therapy. Another important advantage is that ultrasound is a non-invasive stimulus which is used in a wide variety of diagnostic and therapeutic applications. By way of example, ultrasound is well known in medical imaging techniques and, additionally, in orthopedic therapy. Furthermore, instruments suitable for the application of ultrasound to a subject vertebrate are widely available and their use is well known in the art.

[00397] The rapid transcriptional response and endogenous targeting of the instant invention make for an ideal system for the study of transcriptional dynamics. For example, the instant invention may be used to study the dynamics of variant production upon induced expression of a target gene. On the other end of the transcription cycle, mRNA degradation studies are often performed in response to a strong extracellular stimulus, causing expression level changes in a plethora of genes. The instant invention may be utilized to reversibly induce transcription of an endogenous target, after which point stimulation may be stopped and the degradation kinetics of the unique target may be tracked.

[00398] The temporal precision of the instant invention may provide the power to time genetic regulation in concert with experimental interventions. For example, targets with suspected involvement in long-term potentiation (LTP) may be modulated in organotypic or dissociated neuronal cultures, but only during stimulus to induce LTP, so as to avoid interfering with the

normal development of the cells. Similarly, in cellular models exhibiting disease phenotypes, targets suspected to be involved in the effectiveness of a particular therapy may be modulated only during treatment. Conversely, genetic targets may be modulated only during a pathological stimulus. Any number of experiments in which timing of genetic cues to external experimental stimuli is of relevance may potentially benefit from the utility of the instant invention.

[00399] The *in vivo* context offers equally rich opportunities for the instant invention to control gene expression. Photoinducibility provides the potential for spatial precision. Taking advantage of the development of optrode technology, a stimulating fiber optic lead may be placed in a precise brain region. Stimulation region size may then be tuned by light intensity. This may be done in conjunction with the delivery of the Cas9 CRISPR-Cas system or complex of the invention, or, in the case of transgenic Cas9 animals, guide RNA of the invention may be delivered and the optrode technology can allow for the modulation of gene expression in precise brain regions. A transparent Cas9 expressing organism, can have guide RNA of the invention administered to it and then there can be extremely precise laser induced local gene expression changes.

[00400] A culture medium for culturing host cells includes a medium commonly used for tissue culture, such as M199-earle base, Eagle MEM (E-MEM), Dulbecco MEM (DMEM), SC-UCM102, UP-SFM (GIBCO BRL), EX-CELL302 (Nichirei), EX-CELL293-S (Nichirei), TFBM-01 (Nichirei), ASF104, among others. Suitable culture media for specific cell types may be found at the American Type Culture Collection (ATCC) or the European Collection of Cell Cultures (ECACC). Culture media may be supplemented with amino acids such as L-glutamine, salts, anti-fungal or anti-bacterial agents such as Fungizone®, penicillin-streptomycin, animal serum, and the like. The cell culture medium may optionally be serum-free.

[00401] The invention may also offer valuable temporal precision *in vivo*. The invention may be used to alter gene expression during a particular stage of development. The invention may be used to time a genetic cue to a particular experimental window. For example, genes implicated in learning may be overexpressed or repressed only during the learning stimulus in a precise region of the intact rodent or primate brain. Further, the invention may be used to induce gene expression changes only during particular stages of disease development. For example, an oncogene may be overexpressed only once a tumor reaches a particular size or metastatic stage. Conversely, proteins suspected in the development of Alzheimer's may be knocked down only at

defined time points in the animal's life and within a particular brain region. Although these examples do not exhaustively list the potential applications of the invention, they highlight some of the areas in which the invention may be a powerful technology.

Protected guides: Enzymes according to the invention can be used in combination with protected guide RNAs

[00402] In one aspect, an object of the current invention is to further enhance the specificity of Cas9 given individual guide RNAs through thermodynamic tuning of the binding specificity of the guide RNA to target DNA. This is a general approach of introducing mismatches, elongation or truncation of the guide sequence to increase / decrease the number of complimentary bases vs. mismatched bases shared between a genomic target and its potential off-target loci, in order to give thermodynamic advantage to targeted genomic loci over genomic off-targets.

[00403] In one aspect, the invention provides for the guide sequence being modified by secondary structure to increase the specificity of the Cas9 CRISPR-Cas system and whereby the secondary structure can protect against exonuclease activity and allow for 3' additions to the guide sequence.

[00404] In one aspect, the invention provides for hybridizing a "protector RNA" to a guide sequence, wherein the "protector RNA" is an RNA strand complementary to the 5' end of the guide RNA (gRNA), to thereby generate a partially double-stranded gRNA. In an embodiment of the invention, protecting the mismatched bases with a perfectly complementary protector sequence decreases the likelihood of target DNA binding to the mismatched base pairs at the 3' end. In embodiments of the invention, additional sequences comprising an extended length may also be present.

[00405] Guide RNA (gRNA) extensions matching the genomic target provide gRNA protection and enhance specificity. Extension of the gRNA with matching sequence distal to the end of the spacer seed for individual genomic targets is envisaged to provide enhanced specificity. Matching gRNA extensions that enhance specificity have been observed in cells without truncation. Prediction of gRNA structure accompanying these stable length extensions has shown that stable forms arise from protective states, where the extension forms a closed loop with the gRNA seed due to complimentary sequences in the spacer extension and the spacer seed. These results demonstrate that the protected guide concept also includes sequences

matching the genomic target sequence distal of the 20mer spacer-binding region. Thermodynamic prediction can be used to predict completely matching or partially matching guide extensions that result in protected gRNA states. This extends the concept of protected gRNAs to interaction between X and Z, where X will generally be of length 17-20nt and Z is of length 1-30nt. Thermodynamic prediction can be used to determine the optimal extension state for Z, potentially introducing small numbers of mismatches in Z to promote the formation of protected conformations between X and Z. Throughout the present application, the terms "X" and seed length (SL) are used interchangeably with the term exposed length (EpL) which denotes the number of nucleotides available for target DNA to bind; the terms "Y" and protector length (PL) are used interchangeably to represent the length of the protector; and the terms "Z", "E", "E"" and EL are used interchangeably to correspond to the term extended length (ExL) which represents the number of nucleotides by which the target sequence is extended.

[00406] An extension sequence which corresponds to the extended length (ExL) may optionally be attached directly to the guide sequence at the 3' end of the protected guide sequence. The extension sequence may be 2 to 12 nucleotides in length. Preferably ExL may be denoted as 0, 2, 4, 6, 8, 10 or 12 nucleotides in length.. In a preferred embodiment the ExL is denoted as 0 or 4 nucleotides in length. In a more preferred embodiment the ExL is 4 nucleotides in length. The extension sequence may or may not be complementary to the target sequence.

[00407] An extension sequence may further optionally be attached directly to the guide sequence at the 5' end of the protected guide sequence as well as to the 3' end of a protecting sequence. As a result, the extension sequence serves as a linking sequence between the protected sequence and the protecting sequence. Without wishing to be bound by theory, such a link may position the protecting sequence near the protected sequence for improved binding of the protecting sequence to the protected sequence.

[00408] Addition of gRNA mismatches to the distal end of the gRNA can demonstrate enhanced specificity. The introduction of unprotected distal mismatches in Y or extension of the gRNA with distal mismatches (Z) can demonstrate enhanced specificity. This concept as mentioned is tied to X, Y, and Z components used in protected gRNAs. The unprotected mismatch concept may be further generalized to the concepts of X, Y, and Z described for protected guide RNAs.

[00409] Cas9Cas9In one aspect, the invention provides for enhanced Cas9Cas9 specificity wherein the double stranded 3' end of the protected guide RNA (pgRNA) allows for two possible outcomes: (1) the guide RNA-protector RNA to guide RNA-target DNA strand exchange will occur and the guide will fully bind the target, or (2) the guide RNA will fail to fully bind the target and because Cas9 target cleavage is a multiple step kinetic reaction that requires guide RNA:target DNA binding to activate Cas9-catalyzed DSBs, wherein Cas9 cleavage does not occur if the guide RNA does not properly bind. According to particular embodiments, the protected guide RNA improves specificity of target binding as compared to a naturally occurring CRISPR-Cas system. According to particular embodiments the protected modified guide RNA improves stability as compared to a naturally occurring CRISPR-Cas. According to particular embodiments the protector sequence has a length between 3 and 120 nucleotides and comprises 3 or more contiguous nucleotides complementary to another sequence of guide or protector. According to particular embodiments, the protector sequence forms a hairpin. According to particular embodiments the guide RNA further comprises a protected sequence and an exposed sequence. According to particular embodiments the exposed sequence is 1 to 19 nucleotides. More particularly, the exposed sequence is at least 75%, at least 90% or about 100% complementary to the target sequence. According to particular embodiments the guide sequence is at least 90% or about 100% complementary to the protector strand. According to particular embodiments the guide sequence is at least 75%, at least 90% or about 100% complementary to the target sequence. According to particular embodiments, the guide RNA further comprises an extension sequence. More particularly, the extension sequence is operably linked to the 3' end of the protected guide sequence, and optionally directly linked to the 3' end of the protected guide sequence. According to particular embodiments the extension sequence is 1-12 nucleotides. According to particular embodiments the extension sequence is operably linked to the guide sequence at the 3' end of the protected guide sequence and the 5' end of the protector strand and optionally directly linked to the 3' end of the protected guide sequence and the 3' end of the protector strand, wherein the extension sequence is a linking sequence between the protected sequence and the protector strand. According to particular embodiments the extension sequence is 100% not complementary to the protector strand, optionally at least 95%, at least 90%, at least 80%, at least 70%, at least 60%, or at least 50% not complementary to the protector strand. According to particular embodiments the guide sequence further comprises

mismatches appended to the end of the guide sequence, wherein the mismatches thermodynamically optimize specificity.

[00410] In one aspect, the invention provides an engineered, non-naturally occurring CRISPR-Cas system comprising a Cas9 protein and a protected guide RNA that targets a DNA molecule encoding a gene product in a cell, whereby the protected guide RNA targets the DNA molecule encoding the gene product and the Cas9 protein cleaves the DNA molecule encoding the gene product, whereby expression of the gene product is altered; and, wherein the Cas9 protein and the protected guide RNA do not naturally occur together. The invention comprehends the protected guide RNA comprising a guide sequence fused 3' to a direct repeat sequence. The invention further comprehends the Cas9 protein being codon optimized for expression in a Eukaryotic cell. In a preferred embodiment the Eukaryotic cell is a mammalian cell, a plant cell or a yeast cell and in a more preferred embodiment the mammalian cell is a human cell. In a further embodiment of the invention, the expression of the gene product is decreased. In some embodiments, the Cas9 enzyme is *Acidaminococcus sp. BV3L6, Lachnospiraceae bacterium* or *Francisella Novicida* Cas9, and may include mutated Cas9 derived from these organisms. The enzyme may be a further Cas9 homolog or ortholog. In some embodiments, the nucleotide sequence encoding the Cfp1 enzyme is codon-optimized for expression in a eukaryotic cell. In some embodiments, the Cas9 enzyme directs cleavage of one or two strands at the location of the target sequence. In some embodiments, the first regulatory element is a polymerase III promoter. In some embodiments, the second regulatory element is a polymerase II promoter. In general, and throughout this specification, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. Vectors include, but are not limited to, nucleic acid molecules that are single-stranded, double-stranded, or partially double-stranded; nucleic acid molecules that comprise one or more free ends, no free ends (e.g., circular); nucleic acid molecules that comprise DNA, RNA, or both; and other varieties of polynucleotides known in the art. One type of vector is a "plasmid," which refers to a circular double stranded DNA loop into which additional DNA segments can be inserted, such as by standard molecular cloning techniques. Another type of vector is a viral vector, wherein virally-derived DNA or RNA sequences are present in the vector for packaging into a virus (e.g., retroviruses, replication defective retroviruses, adenoviruses, replication defective adenoviruses, and adeno-associated viruses). Viral vectors also include

polynucleotides carried by a virus for transfection into a host cell. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (e.g., bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (e.g., non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively-linked. Such vectors are referred to herein as "expression vectors." Common expression vectors of utility in recombinant DNA techniques are often in the form of plasmids.

[00411] Recombinant expression vectors can comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory elements, which may be selected on the basis of the host cells to be used for expression, that is operatively-linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory element(s) in a manner that allows for expression of the nucleotide sequence (e.g., in an in vitro transcription/translation system or in a host cell when the vector is introduced into the host cell).

[00412] Advantageous vectors include lentiviruses and adeno-associated viruses, and types of such vectors can also be selected for targeting particular types of cells.

[00413] In one aspect, the invention provides a eukaryotic host cell comprising (a) a first regulatory element operably linked to a direct repeat sequence and one or more insertion sites for inserting one or more guide sequences downstream of the direct repeat sequence, wherein when expressed, the guide sequence directs sequence-specific binding of a CRISPR complex to a target sequence in a eukaryotic cell, wherein the CRISPR complex comprises a CRISPR enzyme complexed with the guide RNA comprising the guide sequence that is hybridized to the target sequence and/or (b) a second regulatory element operably linked to an enzyme-coding sequence encoding said Cas9 enzyme comprising a nuclear localization sequence. In some embodiments, the host cell comprises components (a) and (b). In some embodiments, component (a), component (b), or components (a) and (b) are stably integrated into a genome of the host eukaryotic cell. In some embodiments, component (a) further comprises two or more guide sequences operably linked to the first regulatory element, wherein when expressed, each of the two or more guide sequences direct sequence specific binding of a CRISPR complex to a

different target sequence in a eukaryotic cell. In some embodiments, the Cas9 enzyme directs cleavage of one or two strands at the location of the target sequence. In some embodiments, the Cas9 enzyme lacks DNA strand cleavage activity. In some embodiments, the first regulatory element is a polymerase III promoter. In some embodiments, the second regulatory element is a polymerase II promoter.

[00414] In an aspect, the invention provides a non-human eukaryotic organism; preferably a multicellular eukaryotic organism, comprising a eukaryotic host cell according to any of the described embodiments. In other aspects, the invention provides a eukaryotic organism; preferably a multicellular eukaryotic organism, comprising a eukaryotic host cell according to any of the described embodiments. The organism in some embodiments of these aspects may be an animal; for example a mammal. Also, the organism may be an arthropod such as an insect. The organism also may be a plant or a yeast. Further, the organism may be a fungus.

[00415] In one aspect, the invention provides a kit comprising one or more of the components described herein above. In some embodiments, the kit comprises a vector system and instructions for using the kit. In some embodiments, the vector system comprises (a) a first regulatory element operably linked to a direct repeat sequence and one or more insertion sites for inserting one or more guide sequences downstream of the direct repeat sequence, wherein when expressed, the guide sequence directs sequence-specific binding of a Cas9 CRISPR complex to a target sequence in a eukaryotic cell, wherein the CRISPR complex comprises a Cas9 enzyme complexed with the protected guide RNA comprising the guide sequence that is hybridized to the target sequence and/or (b) a second regulatory element operably linked to an enzyme-coding sequence encoding said Cas9 enzyme comprising a nuclear localization sequence. In some embodiments, the kit comprises components (a) and (b) located on the same or different vectors of the system. In some embodiments, component (a) further comprises two or more guide sequences operably linked to the first regulatory element, wherein when expressed, each of the two or more guide sequences direct sequence specific binding of a CRISPR complex to a different target sequence in a eukaryotic cell. In some embodiments, the Cas9 enzyme comprises one or more nuclear localization sequences of sufficient strength to drive accumulation of said Cas9 enzyme in a detectable amount in the nucleus of a eukaryotic cell. In some embodiments, the Cas9 enzyme is *Acidaminococcus sp. BV3L6, Lachnospiraceae bacterium MA2020* or *Francisella tularensis 1 Novicida* Cas9, and may include mutated Cas9

derived from these organisms. The enzyme may be a Cas9 homolog or ortholog. In some embodiments, the CRISPR enzyme is codon-optimized for expression in a eukaryotic cell. In some embodiments, the CRISPR enzyme directs cleavage of one or two strands at the location of the target sequence. In some embodiments, the CRISPR enzyme lacks DNA strand cleavage activity. In some embodiments, the first regulatory element is a polymerase III promoter. In some embodiments, the second regulatory element is a polymerase II promoter.

[00416] In one aspect, the invention provides a method of modifying a target polynucleotide in a eukaryotic 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 Cas9 enzyme complexed with protected guide RNA comprising a guide sequence hybridized to a target sequence within said target polynucleotide. In some embodiments, said cleavage comprises cleaving one or two strands at the location of the target sequence by said Cas9 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 non-homologous end joining (NHEJ)-based gene insertion mechanisms, more particularly 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 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 Cas9 enzyme, the protected guide RNA comprising the guide sequence linked to direct repeat 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.

[00417] In one aspect, the invention provides a method of modifying expression of a polynucleotide in a eukaryotic cell. In some embodiments, the method comprises allowing a Cas9 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 Cas9 enzyme complexed with a protected guide RNA comprising a guide sequence hybridized to a target sequence within said polynucleotide. 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 Cas9 enzyme and the protected guide RNA.

[00418] In one aspect, the invention provides a method of generating a model eukaryotic cell comprising a mutated disease gene. In some embodiments, a disease gene is any gene associated 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 Cas9 enzyme and a protected guide RNA comprising a guide sequence linked to a direct repeat 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 Cas9 enzyme complexed with the guide RNA comprising the sequence that is hybridized to the target sequence within the target polynucleotide, thereby generating a model eukaryotic cell comprising a mutated disease gene. In some embodiments, said cleavage comprises cleaving one or two strands at the location of the target sequence by said Cas9 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 non-homologous end joining (NHEJ)-based gene insertion mechanisms 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.

[00419] In one aspect, the invention provides a method for developing a biologically active agent that modulates a cell signaling event associated with a disease gene. In some embodiments, a disease gene is any gene associated an increase in the risk of having or developing a disease. In some embodiments, the method comprises (a) contacting a test compound with a model cell of any one of the described embodiments; and (b) detecting a change in a readout that is indicative of a reduction or an augmentation of a cell signaling event associated with said mutation in said disease gene, thereby developing said biologically active agent that modulates said cell signaling event associated with said disease gene.

[00420] In one aspect, the invention provides a recombinant polynucleotide comprising a protected guide sequence downstream of a direct repeat sequence, wherein the protected guide sequence when expressed directs sequence-specific binding of a CRISPR complex to a corresponding target sequence present in a eukaryotic cell. In some embodiments, the target sequence is a viral sequence present in a eukaryotic cell. In some embodiments, the target sequence is a proto-oncogene or an oncogene.

[00421] In one aspect the invention provides for a method of selecting one or more 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 Cas9 enzyme, a protected guide RNA comprising a guide sequence, and an editing template; wherein the editing template comprises the one or more mutations that abolish Cas9 enzyme cleavage; allowing non-homologous end joining (NHEJ)-based gene insertion mechanisms 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 Cas9 enzyme complexed with the protected guide RNA comprising a guide sequence that is hybridized to the target sequence within the target polynucleotide, wherein binding of the CRISPR complex to the target polynucleotide induces cell death, thereby allowing one or more cell(s) in which one or more mutations have been introduced to be selected. In a preferred embodiment of the invention the cell to be selected may be a eukaryotic cell. 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.

[00422] With respect to mutations of the Cas9 enzyme, when the enzyme is not FnCas9, mutations may be as described herein elsewhere; conservative substitution for any of the replacement amino acids is also envisaged. In an aspect the invention provides as to any or each or all embodiments herein-discussed wherein the CRISPR enzyme comprises at least one or more, or at least two or more mutations, wherein the at least one or more mutation or the at least two or more mutations are selected from those described herein elsewhere.

[00423] In a further aspect, the invention involves a computer-assisted method for identifying or designing potential compounds to fit within or bind to CRISPR-Cas9 system or a functional portion thereof or vice versa (a computer-assisted method for identifying or

designing potential CRISPR-Cas9 systems or a functional portion thereof for binding to desired compounds) or a computer-assisted method for identifying or designing potential CRISPR-Cas9 systems (e.g., with regard to predicting areas of the CRISPR-Cas9 system to be able to be manipulated—for instance, based on crystal structure data or based on data of Cas9 orthologs, or with respect to where a functional group such as an activator or repressor can be attached to the CRISPR-Cas9 system, or as to Cas9 truncations or as to designing nickases), said method comprising:

[00424] using a computer system, e.g., a programmed computer comprising a processor, a data storage system, an input device, and an output device, the steps of:

[00425] (a) inputting into the programmed computer through said input device data comprising the three-dimensional co-ordinates of a subset of the atoms from or pertaining to the CRISPR-Cas9 crystal structure, e.g., in the CRISPR-Cas9 system binding domain or alternatively or additionally in domains that vary based on variance among Cas9 orthologs or as to Cas9s or as to nickases or as to functional groups, optionally with structural information from CRISPR-Cas9 system complex(es), thereby generating a data set;

[00426] (b) comparing, using said processor, said data set to a computer database of structures stored in said computer data storage system, e.g., structures of compounds that bind or putatively bind or that are desired to bind to a CRISPR-Cas9 system or as to Cas9 orthologs (e.g., as Cas9s or as to domains or regions that vary amongst Cas9 orthologs) or as to the CRISPR-Cas9 crystal structure or as to nickases or as to functional groups;

[00427] (c) selecting from said database, using computer methods, structure(s)—e.g., CRISPR-Cas9 structures that may bind to desired structures, desired structures that may bind to certain CRISPR-Cas9 structures, portions of the CRISPR-Cas9 system that may be manipulated, e.g., based on data from other portions of the CRISPR-Cas9 crystal structure and/or from Cas9 orthologs, truncated Cas9s, novel nickases or particular functional groups, or positions for attaching functional groups or functional-group-CRISPR-Cas9 systems;

[00428] (d) constructing, using computer methods, a model of the selected structure(s); and

[00429] (e) outputting to said output device the selected structure(s);

[00430] and optionally synthesizing one or more of the selected structure(s);

[00431] and further optionally testing said synthesized selected structure(s) as or in a CRISPR-Cas9 system;

[00432] or, said method comprising: providing the co-ordinates of at least two atoms of the CRISPR-Cas9 crystal structure, e.g., at least two atoms of the herein Crystal Structure Table of the CRISPR-Cas9 crystal structure or co-ordinates of at least a sub-domain of the CRISPR-Cas9 crystal structure ("selected co-ordinates"), providing the structure of a candidate comprising a binding molecule or of portions of the CRISPR-Cas9 system that may be manipulated, e.g., based on data from other portions of the CRISPR-Cas9 crystal structure and/or from Cas9 orthologs, or the structure of functional groups, and fitting the structure of the candidate to the selected co-ordinates, to thereby obtain product data comprising CRISPR-Cas9 structures that may bind to desired structures, desired structures that may bind to certain CRISPR-Cas9 structures, portions of the CRISPR-Cas9 system that may be manipulated, truncated Cas9s, novel nickases, or particular functional groups, or positions for attaching functional groups or functional-group-CRISPR-Cas9 systems, with output thereof; and optionally synthesizing compound(s) from said product data and further optionally comprising testing said synthesized compound(s) as or in a CRISPR-Cas9 system.

[00433] The testing can comprise analyzing the CRISPR-Cas9 system resulting from said synthesized selected structure(s), e.g., with respect to binding, or performing a desired function.

[00434] The output in the foregoing methods can comprise data transmission, e.g., transmission of information via telecommunication, telephone, video conference, mass communication, e.g., presentation such as a computer presentation (e.g. POWERPOINT), internet, email, documentary communication such as a computer program (e.g. WORD) document and the like. Accordingly, the invention also comprehends computer readable media containing: atomic co-ordinate data according to the herein-referenced Crystal Structure, said data defining the three dimensional structure of CRISPR-Cas9 or at least one sub-domain thereof, or structure factor data for CRISPR-Cas9, said structure factor data being derivable from the atomic co-ordinate data of herein-referenced Crystal Structure. The computer readable media can also contain any data of the foregoing methods. The invention further comprehends methods a computer system for generating or performing rational design as in the foregoing methods containing either: atomic co-ordinate data according to herein-referenced Crystal

Structure, said data defining the three dimensional structure of CRISPR-Cas9 or at least one sub-domain thereof, or structure factor data for CRISPR-Cas9, said structure factor data being derivable from the atomic co-ordinate data of herein-referenced Crystal Structure. The invention further comprehends a method of doing business comprising providing to a user the computer system or the media or the three dimensional structure of CRISPR-Cas9 or at least one sub-domain thereof, or structure factor data for CRISPR-Cas9, said structure set forth in and said structure factor data being derivable from the atomic co-ordinate data of herein-referenced Crystal Structure, or the herein computer media or a herein data transmission.

[00435] A "binding site" or an "active site" comprises or consists essentially of or consists of a site (such as an atom, a functional group of an amino acid residue or a plurality of such atoms and/or groups) in a binding cavity or region, which may bind to a compound such as a nucleic acid molecule, which is/are involved in binding.

[00436] By "fitting", is meant determining by automatic, or semi-automatic means, interactions between one or more atoms of a candidate molecule and at least one atom of a structure of the invention, and calculating the extent to which such interactions are stable. Interactions include attraction and repulsion, brought about by charge, steric considerations and the like. Various computer-based methods for fitting are described further

[00437] By "root mean square (or rms) deviation", we mean the square root of the arithmetic mean of the squares of the deviations from the mean.

[00438] By a "computer system", is meant the hardware means, software means and data storage means used to analyze atomic coordinate data. The minimum hardware means of the computer-based systems of the present invention typically comprises a central processing unit (CPU), input means, output means and data storage means. Desirably a display or monitor is provided to visualize structure data. The data storage means may be RAM or means for accessing computer readable media of the invention. Examples of such systems are computer and tablet devices running Unix, Windows or Apple operating systems.

[00439] By "computer readable media", is meant any medium or media, which can be read and accessed directly or indirectly by a computer e.g., so that the media is suitable for use in the above-mentioned computer system. Such media include, but are not limited to: magnetic storage media such as floppy discs, hard disc storage medium and magnetic tape; optical storage media such as optical discs or CD-ROM; electrical storage media such as

RAM and ROM; thumb drive devices; cloud storage devices and hybrids of these categories such as magnetic/optical storage media.

[00440] The invention comprehends the use of the protected guides described herein above in the optimized functional CRISPR-Cas enzyme systems described herein.

Formation of a RISC through Guide Engineering

[00441] In some embodiments, the guide may be a protected guide (e.g. a pgRNA) or an escorted guide (e.g. an esgRNA) as described herein. Both of these, in some embodiments, make use of RISC. A RISC is a key component of RNAi. RISC (RNA-induced silencing complex) is a multiprotein, specifically a ribonucleoprotein, complex which incorporates one strand of a double-stranded RNA (dsRNA) fragment, such as small interfering RNA (siRNA) or microRNA (miRNA), which acts as a template for RISC to recognize a complementary messenger RNA (mRNA) transcript. The mRNA is thus cleaved by one of the components of the RISC.

[00442] As such, the formation of a RISC is advantageous in some embodiments. Guide RNAs according to various aspects of the present invention, including but not limited to protected and/or escorted guide RNAs, may be adapted to include RNA nucleotides that promote formation of a RISC, for example in combination with an siRNA or miRNA that may be provided or may, for instance, already be expressed in a cell. This may be useful, for instance, as a self-inactivating system to clear or degrade the guide.

[00443] Thus, the guide RNA may comprise a sequence complementary to a target miRNA or an siRNA, which may or may not be present within a cell. In this way, only when the miRNA or siRNA is present, for example through expression (by the cell or through human intervention), is there binding of the RNA sequence to the miRNA or siRNA which then results in cleavage of the guide RNA an RNA-induced silencing complex (RISC) within the cell. Therefore, in some embodiments, the guide RNA comprises an RNA sequence complementary to a target miRNA or siRNA, and binding of the guide RNA sequence to the target miRNA or siRNA results in cleavage of the guide RNA by an RNA-induced silencing complex (RISC) within the cell.

[00444] This is explained further below with specific reference to both protected and escorted guides.

[00445] RISC formation through use of Protected Guides

[00446] For example, a protected guide may be described in the following aspect: an engineered, non-naturally occurring composition comprising a Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-CRISPR associated (Cas) (CRISPR-Cas) system having a protected guide RNA (pgRNA) polynucleotide sequence comprising (a) a protector sequence, (b) a guide sequence capable of hybridizing to a target sequence in a eukaryotic cell, (c) a tracr mate sequence, and (d) a tracr sequence wherein (a), (b), (c) and (d) are arranged in a 5' to 3' orientation, wherein the protector sequence comprises two or more nucleotides that are non-complementary to the target 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, wherein the CRISPR complex comprises a Type II Cas9 protein complexed with (1) the guide sequence that is hybridized to the target sequence, and (2) the tracr mate sequence that is hybridized to the tracr sequence and wherein in the polynucleotide sequence, one or more of the guide, tracr and tracr mate sequences are modified.

[00447] In one aspect, this protected guide system is used for secondary structure protection for 5' extensions to the sgRNA. For example, Applicants extend the sgRNA such that a miRNA binding site is introduced to make the sgRNA only active when the miRNA binding site is processed and cleaved by the RISC complex machinery. This would not be possible without secondary structure protection since exonuclease processing would start from the 5' end and cut back towards the sgRNA. By adding a small secondary structure loop 5' to the added miRNA site, then miRNA may be protected from exonuclease chew back.

[00448] RISC formation through use of Escorted Guides

[00449] In another example, an escorted guide may be described. In particular, an miRNA Inducible esgRNA is envisaged. Here the escort RNA aptamer sequence is complementary to a target miRNA, so that when the target miRNA is present in a cell incorporated into the RNA-induced silencing complex (RISC), there is binding of the escort RNA aptamer sequence to the target miRNA, which results in cleavage of the esgRNA by an RNA-induced silencing complex (RISC) within the cell.

[00450] In alternative embodiments, a wide variety of primary and secondary structures may be provided at the 5' end of the esgRNA, designed so that the RISC complex is able to access the miRNA binding site. An esgRNA may have first and second linker sequences,

5' to a protector sequence. In alternative embodiments, linkers 1 and 2 may for example each independently be 0, 1, 2, 3, or 4 nucleotides long, with a protector sequence of 0, 1 or 2 nucleotides in length.

[00451] In an exemplary embodiment, induction of esgRNA targeting may be illustrated using miR-122 in a HEK.293 cell system, in which miR-122 is not expressed natively. In the absence of exogenous miR-122, the protected esgRNAs did not mediate targeted EMX1.3 nuclease activity. When exogenous miR-122 is added (100 ng / well) targeted EMX1.3 cutting was observed (as distinct cleavage artifacts visible as electrophoretic variants on gels). This demonstrates that highly expressed endogenous miRNAs can be utilized in systems that provide genetically inducible sgRNAs. Any miRNA may be used in place of miRNA122, with a corresponding sequence readily determined.

[00452] For example, an sgRNA may be linked to an "escort" RNA aptamer sequence complementary to an endogenous target miRNA. The target miRNA may form an RNA-induced silencing complex (RISC) within the cell. When the target miRNA is present in a cell there is binding of the escort RNA aptamer sequence to the target miRNA, which results in cleavage of the esgRNA by the RNA-induced silencing complex (RISC) within the cell. Cleavage of the escort releases the active sgRNA.

[00453] For example, a protected guide may be described in the following aspect: a non-naturally occurring or engineered composition comprising an escorted single CRISPR-Cas9 guide RNA (esgRNA) comprising:

an RNA guide sequence capable of hybridizing to a target sequence in a genomic locus of interest in a cell; and,

an escort RNA aptamer sequence,

wherein the escort RNA aptamer sequence comprises binding affinity for an aptamer ligand on or in the cell, or the escort RNA aptamer sequence is responsive to a localized aptamer effector on or in the cell,

wherein the presence of the aptamer ligand or effector on or in the cell is spatially or temporally restricted.

[00454] The escort RNA aptamer sequence may be complementary to a target miRNA, which may or may not be present within a cell, so that only when the target miRNA is present is there binding of the escort RNA aptamer sequence to the target miRNA which results

in cleavage of the esgRNA by an RNA-induced silencing complex (RISC) within the cell. Therefore, in some embodiments, the escort RNA aptamer sequence is complementary to a target miRNA, and binding of the escort RNA aptamer sequence to the target miRNA results in cleavage of the esgRNA by an RNA-induced silencing complex (RISC) within the cell.

[00455]

[00456] According to the invention, a nucleotide sequence encoding at least one of said guide RNA or Cas protein is operably connected in the cell with a regulatory element comprising a promoter of a gene of interest, whereby expression of at least one CRISPR-Cas system component is driven by the promoter of the gene of interest. "operably connected" is intended to mean that the nucleotide sequence encoding the guide RNA and/or the Cas is linked to the regulatory element(s) in a manner that allows for expression of the nucleotide sequence, as also referred to herein elsewhere. The term "regulatory element" is also described herein elsewhere. According to the invention, the regulatory element comprises a promoter of a gene of interest, such as preferably a promoter of an endogenous gene of interest. In certain embodiments, the promoter is at its endogenous genomic location. In such embodiments, the nucleic acid encoding the CRISPR and/or Cas is under transcriptional control of the promoter of the gene of interest at its native genomic location. In certain other embodiments, the promoter is provided on a (separate) nucleic acid molecule, such as a vector or plasmid, or other extrachromosomal nucleic acid, i.e. the promoter is not provided at its native genomic location. In certain embodiments, the promoter is genomically integrated at a non-native genomic location.

[00457] In certain embodiments, a nucleic acid encoding the guide RNA is operably connected in the cell with a regulatory element comprising a promoter of a gene of interest. In certain embodiments, a nucleic acid encoding the Cas is operably connected in the cell with a regulatory element comprising a promoter of a gene of interest. In certain embodiments a nucleic acid encoding the guide RNA is operably connected in the cell with a regulatory element comprising a promoter of a gene of interest and a nucleic acid encoding the Cas is operably connected in the cell with a regulatory element comprising a promoter of a gene of interest. In this latter case, the promoter driving the expression of the guide RNA and the Cas may be the same or may be different. In certain embodiments, a nucleic acid encoding the guide RNA and/or Cas is genomically integrated. In certain embodiments, a nucleic acid encoding the

guide RNA and/or Cas is extrachromosomal or episomal. The nucleic acid encoding the guide RNA and the nucleic acid encoding the Cas may reside on the same or different nucleic acid molecules.

[00458] The selected DNA sequences which are targeted by the guide RNA(s) according to the invention may be endogenous DNA sequences or exogenous DNA sequences. The selected DNA sequences which are targeted by the guide RNA(s), such as exogenous DNA sequences, according to the invention may be genomically integrated or may be extrachromosomal (e.g. provided on a plasmid or vector). In certain embodiments, the methods as described herein comprise introducing in the cell a vector or plasmid, by means known in the art as described herein elsewhere, said vector or plasmid comprising said selected DNA sequence and said method comprises detection of the modification of said selected DNA sequence on said vector. It will be understood that said vector or plasmid, or at least the selected DNA sequence comprised therein, may be genomically integrated, such as random integration or via homologous recombination. When the selected target DNA sequence is an endogenous sequence, it is preferred that the sequence is selected such that modification thereof has no or minimal impact on the (normal) functioning of the cell. The skilled person will readily identify such sequences by routine analysis or experimentation. In any case, it is preferred that such selected endogenous target DNA sequence does not reside in a coding sequence or ORF of a gene and/or does not reside in regulatory sequences of a gene (such as promoters, enhancers, silencers, etc.).

[00459] As described herein elsewhere, the selected target DNA sequence is modified by the action of a functional CRISPR complex (i.e. the guide RNA complexed with the Cas protein, wherein the guide RNA comprises the guide sequence, tracr mate sequence and tracr sequence in 5' to 3' orientation, wherein the tracr sequence may or may not be on the same nucleic acid molecule as the guide sequence and tracr mate sequence). As used herein, "modified" essentially corresponds to mutated, i.e. the nucleic acid sequence of the target DNA sequence is altered, as described herein elsewhere, such as comprising point mutations, deletions, substitutions, or insertions of one or more nucleotides.

[00460] However as described herein elsewhere, it will also be apparent that in certain embodiments "modified" corresponds to alterations of target loci such as the activation or repression of the transcription of a gene, methylation or demethylation of CpG sites and the like, which may not require point mutations, deletions, substitutions, or insertions of one or more

nucleotides. Furthermore as described herein elsewhere, it will also be apparent that reference to a CRISPR-Cas enzyme as "altering" or "modifying" one or more target polynucleotide loci encompasses direct alteration or modification, e.g. via the catalytic activity of the enzyme itself but also indirect alteration or modification, e.g. via a catalytic activity associated with the CRISPR-Cas enzyme such as a heterologous functional domain, e.g. a transcriptional activation domain. In addition, as will be appreciated it is intended that the one or more target polynucleotide loci which are "altered" or "modified" by the action of the CRISPR-Cas enzyme may be comprised in or adjacent the polynucleotide sequence complementary to the guide sequence portion of a guide RNA, e.g. in embodiments wherein the alteration or modification is effected by the catalytic activity of the CRISPR-Cas enzyme itself, e.g. cleavage of DNA by the nuclease activity of the CRISPR-Cas enzyme. However, also encompassed are embodiments wherein one or more target loci to be "altered" or "modified" are at a location distinct from the sequence complementary to the guide sequence portion of the guide RNA, e.g. in embodiments wherein the alteration or modification is effected via a heterologous functional domain associated with the CRISPR-Cas enzyme, e.g. activation or repression of the transcription of a gene. As such, "alteration" or "modification" (or analogous terms) of a target locus means via direct or indirect action of the CRISPR-Cas enzyme, and furthermore the "target locus" to be altered or modified and the "target sequence" which is complementary to the guide sequence portion of the guide RNA may or may not be the same.

[00461] In certain embodiments, in the methods according to the invention as described herein the CRISPR-Cas system is multiplexed, i.e. multiple different guide RNAs can be provided. Each guide RNA may target (i.e. hybridize with) a different selected DNA target. Expression of the different guide RNAs may be driven by the promoters of different genes of interest. Accordingly, in certain embodiments, the methods of the invention as described herein are methods for determining expression of more than one, such as at least two genes of interest in a cell comprising providing a cell comprising a CRISPR-Cas system, said CRISPR-Cas system comprising more than one, such as at least two guide RNAs that target a different selected DNA sequence and a Cas protein capable of modifying the selected DNA sequence; whereby each guide RNA is operably connected in the cell with a regulatory element comprising a promoter of a different gene of interest; and determining expression of said genes of interest based on detection of the modification of said respective selected DNA sequences. In certain

embodiments, more than one different guide RNA may be operably connected in the cell with a regulatory element comprising a promoter of the same gene of interest. The different guide RNAs may be provided on different nucleic acid molecules or on the same nucleic acid molecule. The respective guide RNAs may be designed such that only modification of a first selected target DNA creates or destroys a second selected target DNA.

[00462] In certain embodiments, one or more of the components of the CRISPR-Cas system may be conditionally (e.g. tissue or cell type specific) and/or inducibly (e.g. chemically inducible) expressed in the cell. Inducible and conditional expression systems are described herein elsewhere. In particular embodiments, one or more of the guide RNA(s) may be conditionally and/or inducibly expressed in the cell. In particular preferred embodiments, the Cas may be conditionally and/or inducibly expressed in the cell.

[00463] As used herein, the term “targeting” of a selected DNA sequence means that a guide RNA is capable of hybridizing with a selected DNA sequence. As uses herein, "hybridization" or “hybridizing” refers to a reaction in which one or more polynucleotides react to form a complex that is stabilized via hydrogen bonding between the bases of the nucleotide residues. The hydrogen bonding may occur by Watson Crick base pairing, Hoogstein binding, or in any other sequence specific manner. The complex may comprise two strands forming a duplex structure, three or more strands forming a multi stranded complex, a single self hybridizing strand, or any combination of these. A hybridization reaction may constitute a step in a more extensive process, such as the initiation of PGR, or the cleavage of a polynucleotide by an enzyme. A sequence capable of hybridizing with a given sequence is referred to as the "complement" of the given sequence.

[00464] As used herein, "expression" or “expressing” refers to the process by which a polynucleotide is transcribed from a DNA template (such as into and mRNA or other RNA transcript) and/or the process by which a transcribed mRNA is subsequently translated into peptides, polypeptides, or proteins. Transcripts and encoded polypeptides may be collectively referred to as "gene product." If the polynucleotide is derived from genomic DNA, expression may include splicing of the mRNA in a eukaryotic ceil. As used herein "expression" of a gene or nucleic acid encompasses not only cellular gene expression, but also the transcription and translation of nucleic acid(s) in cloning systems and in any other context.

[00465] The terms "polypeptide", "peptide" and "protein" are used interchangeably herein to refer to polymers of amino acids of any length. The polymer may be linear or branched, it may comprise modified amino acids, and it may be interrupted by non amino acids. The terms also encompass an amino acid polymer that has been modified; for example, disulfide bond formation, glycosylation, lipidation, acetylation, phosphorylation, or any other manipulation, such as conjugation with a labeling component. As used herein the term "amino acid" includes natural and/or unnatural or synthetic amino acids, including glycine and both the D or I, optical isomers, and amino acid analogs and peptidomimetics.

[00466] The terms "subject," "individual," and "patient" are used interchangeably herein to refer to a vertebrate, preferably a mammal, more preferably a human. Mammals include, but are not limited to, murines, simians, humans, farm animals, sport animals, and pets. Tissues, cells and their progeny of a biological entity obtained in vivo or cultured in vitro are also encompassed.

[00467] In certain embodiments, the methods and cells according to the invention as described herein may be used in screening methods for therapeutic agents, and/or in diagnostic methods. Candidate therapeutic agents may have a different effect of temporal expression profiles, which may be read out according to the methods as described herein.

[00468] The terms "therapeutic agent", "therapeutic capable agent" or "treatment agent" are used interchangeably and refer to a molecule or compound that confers some beneficial effect upon administration to a subject. The beneficial effect includes enablement of diagnostic determinations; amelioration of a disease, symptom, disorder, or pathological condition; reducing or preventing the onset of a disease, symptom, disorder or condition; and generally counteracting a disease, symptom, disorder or pathological condition.

[00469] As used herein, "treatment" or "treating," or "palliating" or "ameliorating" are used interchangeably. These terms refer to an approach for obtaining beneficial or desired results including but not limited to a therapeutic benefit and/or a prophylactic benefit. By therapeutic benefit is meant any therapeutically relevant improvement in or effect on one or more diseases, conditions, or symptoms under treatment. For prophylactic benefit, the compositions may be administered to a subject at risk of developing a particular disease, condition, or symptom, or to a subject reporting one or more of the physiological symptoms of a disease, even though the disease, condition, or symptom may not have yet been manifested.

[00470] As used herein, the terms "chimeric RNA", "chimeric guide RNA", "guide RNA", "single guide RNA" and "synthetic guide RNA" refer to the polynucleotide sequence comprising the guide sequence, the tracr sequence and the tracr mate sequence. The term "guide sequence" refers to the about 20bp sequence within the guide RNA that specifies the target site and may be used interchangeably with the terms "guide" or "spacer". The term "tracr mate sequence" may also be used interchangeably with the term "direct repeat(s)". The guide sequence, tracr, and tracr mate sequence may be provided on a single nucleic acid molecule. Alternatively, the guide and tracr mate sequence may be provided on a single nucleic acid molecule, whereas the tracr is provided on a separate nucleic acid molecule.

[00471] In general, the CRISPR-Cas or CRISPR system is as used in the foregoing documents, such as WO 2014/093622 (PCT/US2013/074667) and refers collectively to transcripts and other elements involved in the expression of or directing the activity of CRISPR-associated ("Cas") genes, including sequences encoding a Cas gene, a tracr (trans-activating CRISPR) sequence (e.g. tracrRNA or an active partial tracrRNA), a tracr-mate sequence (encompassing a "direct repeat" and a tracrRNA-processed partial direct repeat in the context of an endogenous CRISPR system), a guide sequence (also referred to as a "spacer" in the context of an endogenous CRISPR system), or "RNA(s)" as that term is herein used (e.g., RNA(s) to guide Cas, such as Cas9, e.g. CRISPR RNA and transactivating (tracr) RNA or a single guide RNA (sgRNA) (chimeric RNA)) or other sequences and transcripts from a CRISPR locus. In general, a CRISPR system is characterized by elements that promote the formation of a CRISPR complex at the site of a target sequence (also referred to as a protospacer in the context of an endogenous CRISPR system). In the context of formation of a CRISPR complex, "target sequence" refers to a sequence to which a guide sequence is designed to have complementarity, where hybridization between a target sequence and a guide sequence promotes the formation of a CRISPR complex. A target sequence may comprise any polynucleotide, such as DNA or RNA polynucleotides. In some embodiments, a target sequence is located in the nucleus or cytoplasm of a cell. In some embodiments, direct repeats may be identified in silico by searching for repetitive motifs that fulfill any or all of the following criteria: 1. found in a 2Kb window of genomic sequence flanking the type II CRISPR locus; 2. span from 20 to 50 bp; and 3. interspaced by 20 to 50 bp. In some embodiments, 2 of these criteria may be used, for instance 1 and 2, 2 and 3, or 1 and 3. In some embodiments, all 3 criteria may be used.

[00472] In embodiments of the invention the terms guide sequence and guide RNA, i.e. RNA capable of guiding Cas to a target genomic locus, are used interchangeably as in foregoing cited documents such as WO 2014/093622 (PCT/US2013/074667). In general, a guide sequence is any polynucleotide sequence having sufficient complementarity with a target polynucleotide sequence to hybridize with the target sequence and direct sequence-specific binding of a CRISPR complex to the target sequence. In some embodiments, the degree of complementarity between a guide sequence and its corresponding target sequence, when optimally aligned using a suitable alignment algorithm, is about or more than about 50%, 60%, 75%, 80%, 85%, 90%, 95%, 97.5%, 99%, or more. Optimal alignment may be determined with the use of any suitable algorithm for aligning sequences, non-limiting example of which include the Smith-Waterman algorithm, the Needleman-Wunsch algorithm, algorithms based on the Burrows-Wheeler Transform (e.g. the Burrows Wheeler Aligner), ClustalW, Clustal X, BLAT, Novoalign (Novocraft Technologies; available at www.novocraft.com), ELAND (Illumina, San Diego, CA), SOAP (available at soap.genomics.org.cn), and Maq (available at maq.sourceforge.net). In some embodiments, a guide sequence is about or more than about 5, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, 75, or more nucleotides in length. In some embodiments, a guide sequence is less than about 75, 50, 45, 40, 35, 30, 25, 20, 15, 12, or fewer nucleotides in length. Preferably the guide sequence is 10 30 nucleotides long. The ability of a guide sequence to direct sequence-specific binding of a CRISPR complex to a target sequence may be assessed by any suitable assay. For example, the components of a CRISPR system sufficient to form a CRISPR complex, including the guide sequence to be tested, may be provided to a host cell having the corresponding target sequence, such as by transfection with vectors encoding the components of the CRISPR sequence, followed by an assessment of preferential cleavage within the target sequence, such as by Surveyor assay as described herein. Similarly, cleavage of a target polynucleotide sequence may be evaluated in a test tube by providing the target sequence, components of a CRISPR complex, including the guide sequence to be tested and a control guide sequence different from the test guide sequence, and comparing binding or rate of cleavage at the target sequence between the test and control guide sequence reactions. Other assays are possible, and will occur to those skilled in the art.

[00473] A guide sequence, i.e. an RNA capable of guiding Cas to a genomic target locus, may be selected to target any target sequence. In some embodiments, the target sequence is a

sequence within a genome of a cell. Exemplary target sequences include those that are unique in the target genome. For example, for the S. pyogenes Cas9, a unique target sequence in a genome may include a Cas9 target site of the form MMMMMMMMMNNNNNNNNNNNNXGG where NNNNNNNNNNNNXGG (N is A, G, T, or C; and X can be anything) has a single occurrence in the genome. A unique target sequence in a genome may include an S. pyogenes Cas9 target site of the form MMMMMMMMMNNNNNNNNNNNNXGG where NNNNNNNNNNNNXGG (N is A, G, T, or C; and X can be anything) has a single occurrence in the genome. For the S. thermophilus CRISPR1 Cas9, a unique target sequence in a genome may include a Cas9 target site of the form MMMMMMMMMNNNNNNNNNNNNXXAGAAW where NNNNNNNNNNNNXXAGAAW (N is A, G, T, or C; X can be anything; and W is A or T) has a single occurrence in the genome. A unique target sequence in a genome may include an S. thermophilus CRISPR1 Cas9 target site of the form MMMMMMMMMNNNNNNNNNNNNXXAGAAW where NNNNNNNNNNNNXXAGAAW (N is A, G, T, or C; X can be anything; and W is A or T) has a single occurrence in the genome. For the S. pyogenes Cas9, a unique target sequence in a genome may include a Cas9 target site of the form MMMMMMMMMNNNNNNNNNNNNXGGXG where NNNNNNNNNNNNXGGXG (N is A, G, T, or C; and X can be anything) has a single occurrence in the genome. A unique target sequence in a genome may include an S. pyogenes Cas9 target site of the form MMMMMMMMMNNNNNNNNNNNNXGGXG where NNNNNNNNNNNNXGGXG (N is A, G, T, or C; and X can be anything) has a single occurrence in the genome. In each of these sequences "M" may be A, G, T, or C, and need not be considered in identifying a sequence as unique. In some embodiments, a guide sequence is selected to reduce the degree secondary structure within the guide sequence. In some embodiments, about or less than about 75%, 50%, 40%, 30%, 25%, 20%, 15%, 10%, 5%, 1%, or fewer of the nucleotides of the guide sequence participate in self-complementary base pairing when optimally folded. Optimal folding may be determined by any suitable polynucleotide folding algorithm. Some programs are based on calculating the minimal Gibbs free energy. An example of one such algorithm is mFold, as described by Zuker and Stiegler (Nucleic Acids Res. 9 (1981), 133-148). Another example folding algorithm is the online webserver RNAfold, developed at Institute for Theoretical Chemistry at the University of Vienna, using the centroid structure prediction algorithm (see e.g.

A.R. Gruber et al., 2008, Cell 106(1): 23-24; and PA Carr and GM Church, 2009, Nature Biotechnology 27(12): 1151-62).

[00474] In general, a tracr mate sequence includes any sequence that has sufficient complementarity with a tracr sequence to promote one or more of: (1) excision of a guide sequence flanked by tracr mate sequences in a cell containing the corresponding tracr sequence; and (2) formation of a CRISPR complex at a target sequence, wherein the CRISPR complex comprises the tracr mate sequence hybridized to the tracr sequence. In general, degree of complementarity is with reference to the optimal alignment of the tracr mate sequence and tracr sequence, along the length of the shorter of the two sequences. Optimal alignment may be determined by any suitable alignment algorithm, and may further account for secondary structures, such as self-complementarity within either the tracr sequence or tracr mate sequence. In some embodiments, the degree of complementarity between the tracr sequence and tracr mate sequence along the length of the shorter of the two when optimally aligned is about or more than about 25%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 97.5%, 99%, or higher. In some embodiments, the tracr sequence is about or more than about 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 40, 50, or more nucleotides in length. In some embodiments, the tracr sequence and tracr mate sequence are contained within a single transcript, such that hybridization between the two produces a transcript having a secondary structure, such as a hairpin. In an embodiment of the invention, the transcript or transcribed polynucleotide sequence has at least two or more hairpins. In preferred embodiments, the transcript has two, three, four or five hairpins. In a further embodiment of the invention, the transcript has at most five hairpins. In a hairpin structure the portion of the sequence 5' of the final "N" and upstream of the loop corresponds to the tracr mate sequence, and the portion of the sequence 3' of the loop corresponds to the tracr sequence Further non-limiting examples of single polynucleotides comprising a guide sequence, a tracr mate sequence, and a tracr sequence are as follows (listed 5' to 3'), where "N" represents a base of a guide sequence, the first block of lower case letters represent the tracr mate sequence, and the second block of lower case letters represent the tracr sequence, and the final poly-T sequence represents the transcription terminator: (1) NNNNNNNNNNNNNNNNNNNNgtttttgtactctcaagatttaGAAAtaaatcttgcagaagctacaaagataaggctt catgccgaaatcaacaccctgtcatttatggcagggtgttttcgttatttaaTTTTTT (SEQ ID NO: X); (2) NNNNNNNNNNNNNNNNNNNNgtttttgtactctcaGAAAtgcagaagctacaaagataaggcttcatgccgaaatca

acaccctgtcattttatggcagggtgttttcgttatttaaTTTTTT; (SEQ ID NO: X) (3) NNNNNNNNNNNNNNNNNNNNgttttgtactctcaGAAAtgcagaagctacaaagataaggcttcatgccgaaatca acaccctgtcattttatggcagggtgtTTTTTT; (SEQ ID NO: X) (4) NNNNNNNNNNNNNNNNNNNNgttttagagctaGAAAtagcaagttaaaataaggctagtccgttatcaacttgaaaa agtggcaccgagtcggtgcTTTTTT; (SEQ ID NO: X) (5) NNNNNNNNNNNNNNNNNNNNgttttagagctaGAAATAGcaagttaaaataaggctagtccgttatcaacttgaa aaagtgTTTTTTT; and (SEQ ID NO: X) (6) NNNNNNNNNNNNNNNNNNNNgttttagagctagAAATAGcaagttaaaataaggctagtccgttatcaTTTTT TTT (SEQ ID NO: X). In some embodiments, sequences (1) to (3) are used in combination with Cas9 from S. thermophilus CRISPR1. In some embodiments, sequences (4) to (6) are used in combination with Cas9 from S. pyogenes. In some embodiments, the tracr sequence is a separate transcript from a transcript comprising the tracr mate sequence.

[00475] In some embodiments, candidate tracrRNA may be subsequently predicted by sequences that fulfill any or all of the following criteria: 1. sequence homology to direct repeats (motif search in Geneious with up to 18-bp mismatches); 2. presence of a predicted Rho-independent transcriptional terminator in direction of transcription; and 3. stable hairpin secondary structure between tracrRNA and direct repeat. In some embodiments, 2 of these criteria may be used, for instance 1 and 2, 2 and 3, or 1 and 3. In some embodiments, all 3 criteria may be used.

[00476] In some embodiments, chimeric synthetic guide RNAs (sgRNAs) designs may incorporate at least 12 bp of duplex structure between the direct repeat and tracrRNA.

[00477] The RNAs to guide Cas, such as Cas9, can comprise CRISPR RNA and transactivating (tracr) RNA. The tracr mate and the tracr sequence can be connected to form a transactivating (tracer) sequence. The tracr mate and the tracr sequence can optionally be designed to form a single guide RNA (sgRNA). Indeed, it is advantageous that the RNAs to guide Cas can comprise chimeric single guide RNA (sgRNA). The tracr sequence and tracr mate sequence along the length of the shorter of the two when optimally aligned can be about or more than about 25%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 97.5%, 99%, or higher. The tracr sequence can be about or more than about 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 40, 50, or more nucleotides in length.

[00478] In a classic CRISPR-Cas systems, the degree of complementarity between a guide sequence and its corresponding target sequence can be about or more than about 50%, 60%, 75%, 80%, 85%, 90%, 95%, 97.5%, 99%, or 100%; a guide or RNA or sgRNA can be about or more than about 5, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, 75, or more nucleotides in length; or guide or RNA or sgRNA can be less than about 75, 50, 45, 40, 35, 30, 25, 20, 15, 12, or fewer nucleotides in length; and advantageously tracr RNA is 30 or 50 nucleotides in length. However, an aspect of the invention is to reduce off-target interactions, e.g., reduce the guide interacting with a target sequence having low complementarity. Indeed, in the examples, it is shown that the invention involves mutations that result in the CRISPR-Cas system being able to distinguish between target and off-target sequences that have greater than 80% to about 95% complementarity, e.g., 83%-84% or 88-89% or 94-95% complementarity (for instance, distinguishing between a target having 18 nucleotides from an off-target of 18 nucleotides having 1, 2 or 3 mismatches). Accordingly, in the context of the present invention the degree of complementarity between a guide sequence and its corresponding target sequence is greater than 94.5% or 95% or 95.5% or 96% or 96.5% or 97% or 97.5% or 98% or 98.5% or 99% or 99.5% or 99.9%, or 100%. Off target is less than 100% or 99.9% or 99.5% or 99% or 99% or 98.5% or 98% or 97.5% or 97% or 96.5% or 96% or 95.5% or 95% or 94.5% or 94% or 93% or 92% or 91% or 90% or 89% or 88% or 87% or 86% or 85% or 84% or 83% or 82% or 81% or 80% complementarity between the sequence and the guide, with it advantageous that off target is 100% or 99.9% or 99.5% or 99% or 99% or 98.5% or 98% or 97.5% or 97% or 96.5% or 96% or 95.5% or 95% or 94.5% complementarity between the sequence and the guide.

[00479] In particularly preferred embodiments according to the invention, the guide RNA (capable of guiding Cas to a target locus) may comprise (1) a guide sequence capable of hybridizing to a genomic target locus in the eukaryotic cell; (2) a tracr sequence; and (3) a tracr mate sequence. All (1) to (3) may reside in a single RNA, i.e. an sgRNA (arranged in a 5' to 3' orientation), or the tracr RNA may be a different RNA than the RNA containing the guide and tracr sequence. The tracr hybridizes to the tracr mate sequence and directs the CRISPR/Cas complex to the target sequence.

[00480] The methods according to the invention as described herein comprehend inducing one or more mutations in a eukaryotic cell (in vitro, i.e. in an isolated eukaryotic cell) as herein

discussed comprising delivering to cell a vector as herein discussed. The mutation(s) can include the introduction, deletion, or substitution of one or more nucleotides at each target sequence of cell(s) via the guide(s) RNA(s) or sgRNA(s). The mutations can include the introduction, deletion, or substitution of 1-75 nucleotides at each target sequence of said cell(s) via the guide(s) RNA(s) or sgRNA(s). The mutations can include the introduction, deletion, or substitution of 1, 5, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, or 75 nucleotides at each target sequence of said cell(s) via the guide(s) RNA(s) or sgRNA(s). The mutations can include the introduction, deletion, or substitution of 5, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, or 75 nucleotides at each target sequence of said cell(s) via the guide(s) RNA(s) or sgRNA(s). The mutations include the introduction, deletion, or substitution of 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, or 75 nucleotides at each target sequence of said cell(s) via the guide(s) RNA(s) or sgRNA(s). The mutations can include the introduction, deletion, or substitution of 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, or 75 nucleotides at each target sequence of said cell(s) via the guide(s) RNA(s) or sgRNA(s). The mutations can include the introduction, deletion, or substitution of 40, 45, 50, 75, 100, 200, 300, 400 or 500 nucleotides at each target sequence of said cell(s) via the guide(s) RNA(s) or sgRNA(s).

[00481] For minimization of toxicity and off-target effect, it will be important to control the concentration of Cas mRNA and guide RNA delivered. Optimal concentrations of Cas mRNA and guide RNA can be determined by testing different concentrations in a cellular or non-human eukaryote animal model and using deep sequencing the analyze the extent of modification at potential off-target genomic loci. Alternatively, to minimize the level of toxicity and off-target effect, Cas nickase mRNA (for example S. pyogenes Cas9 with the D10A mutation) can be delivered with a pair of guide RNAs targeting a site of interest. Guide sequences and strategies to minimize toxicity and off-target effects can be as in WO 2014/093622 (PCT/US2013/074667); or, via mutation as herein.

[00482] In some embodiments, the CRISPR system is derived advantageously from a type II CRISPR system. In some embodiments, one or more elements of a CRISPR system is derived from a particular organism comprising an endogenous CRISPR system, such as Streptococcus pyogenes. In preferred embodiments of the invention, the CRISPR system is a type II CRISPR

system and the Cas enzyme is Cas9, which catalyzes DNA cleavage. Non-limiting examples of Cas proteins include Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas9 (also known as Csn1 and Csx12), Cas10, Csy1, Csy2, Csy3, Cse1, Cse2, Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, Csf4, homologues thereof, or modified versions thereof. 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. It will be appreciated that SpCas9 or SaCas9 are those from or derived from S. pyogenes or S. aureus Cas9. 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 It will be appreciated that the terms Cas and CRISPR enzyme are generally used herein interchangeably, unless otherwise apparent. The Cas enzyme can be for instance any naturally-occurring bacterial Cas9 as well as any chimaeras, mutants, homologs or orthologs. Many of the residue numberings used herein refer to the Cas9 enzyme from the type II CRISPR locus in Streptococcus pyogenes (annotated alternatively as SpCas9 or spCas9). However, it will be appreciated that this invention includes many more Cas9s from other species of microbes, e.g., orthologs of SpCas9, or Cas9s derived from microbes in addition to S. pyogenes, e.g., SaCas9 derived from S. aureus, St1Cas9 derived from S. thermophilus and so forth. 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.

[00483] In some embodiments, the unmodified Cas has DNA cleavage activity, such as Cas9. In some embodiments, the Cas directs cleavage of one or both strands at the location of a target sequence, such as within the target sequence and/or within the complement of the target sequence. In some embodiments, the Cas directs cleavage of one or both strands within about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 50, 100, 200, 500, or more base pairs from the first or last nucleotide of a target sequence. In some embodiments, a vector encodes a Cas that is mutated to

with respect to a corresponding wild-type enzyme such that the mutated Cas lacks the ability to cleave one or both strands of a target polynucleotide containing a target sequence. For example, an aspartate-to-alanine substitution (D10A) in the RuvC I catalytic domain of Cas9 from S. pyogenes converts Cas9 from a nuclease that cleaves both strands to a nickase (cleaves a single strand). Other examples of mutations that render Cas9 a nickase include, without limitation, H840A, N854A, and N863A. As a further example, two or more catalytic domains of Cas9 (RuvC I, RuvC II, and RuvC III or the HNH domain) may be mutated to produce a mutated Cas9 substantially lacking all DNA cleavage activity. In some embodiments, a D10A mutation is combined with one or more of H840A, N854A, or N863A mutations to produce a Cas9 enzyme substantially lacking all DNA cleavage activity. In some embodiments, a Cas is considered to substantially lack all DNA cleavage activity when the DNA cleavage activity of the mutated enzyme is about no more than 25%, 10%, 5%, 1%, 0.1%, 0.01%, or less of the DNA cleavage activity of the non-mutated form of the enzyme; an example can be when the DNA cleavage activity of the mutated form is nil or negligible as compared with the non-mutated form. Thus, the Cas 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 (e.g., D10 and H840) in the RuvC and HNH catalytic domains respectively; or the CRISPR enzyme can comprise one or more mutations selected from the group consisting of D10A, E762A, H840A, N854A, N863A or D986A and/or one or more mutations in a RuvC1 or HNH domain of the Cas or has a mutation as otherwise as discussed herein. In one aspect of the invention, the Cas enzyme may be fused to a protein, e.g., a TAG, and/or an inducible/controllable domain such as a chemically inducible/controllable domain. The Cas in the invention may be a chimeric Cas proteins; e.g., a Cas having enhanced function by being a chimera. Chimeric Cas proteins may be new Cas containing fragments from more than one naturally occurring Cas. These may comprise fusions of N-terminal fragment(s) of one Cas9 homolog with C-terminal fragment(s) of another Cas homolog. The Cas can be delivered into the cell in the form of mRNA. The expression of Cas can be under the control of an inducible promoter. Where the enzyme is not SpCas9, mutations may be made at any or all residues corresponding to positions 10, 762, 840, 854, 863 and/or 986 of SpCas9 (which may be ascertained for instance by standard sequence comparison tools). In particular, any or all of the

following mutations are preferred in SpCas9: D10A, E762A, H840A, N854A, N863A and/or D986A; as well as conservative substitution for any of the replacement amino acids is also envisaged. The same (or conservative substitutions of these mutations) at corresponding positions in other Cas9s are also preferred. Particularly preferred are D10 and H840 in SpCas9. However, in other Cas9s, residues corresponding to SpCas9 D10 and H840 are also preferred. It is explicitly an object of the invention to avoid reading on known mutations. That is, the mutations known in the art to cause a Cas9 to become a nickase or a Cas9 to become "dead", e.g., have little or no, e.g., 5% or less than 5%, e.g., less than 4%, 3%, 2% or 1%, nuclease activity as compared to a non-mutated Cas9, are not intended to be within the scope of Cas9 mutations that reduce or eliminate interaction between the guide and off-target nucleic acid molecules, with Applicant reserving the right to employ provisos to exclude such known-"nickase"-or-"dead"-Cas9-resulting mutation. Indeed, the phrase "whereby the enzyme in the CRISPR complex has reduced capability of modifying one or more off-target loci as compared to an unmodified enzyme and/or whereby the enzyme in the CRISPR complex has increased capability of modifying the one or more target loci as compared to an unmodified enzyme" (or like expressions) is not intended to read upon mutations that only result in a nickase of dead Cas9 or known Cas9 mutations. HOWEVER, this is not to say that the instant invention modification(s) or mutation(s) "whereby the enzyme in the CRISPR complex has reduced capability of modifying one or more off-target loci as compared to an unmodified enzyme and/or whereby the enzyme in the CRISPR complex has increased capability of modifying the one or more target loci as compared to an unmodified enzyme" (or like expressions) cannot be combined with mutations that result in the enzyme being a nickase or dead. Such a dead enzyme can be an enhanced nucleic acid molecule binder. And such a nickase can be an enhanced nickase. For instance, changing neutral amino acid(s) in and/or near the groove and/or other charged residues in other locations in Cas9 that are in close proximity to a nucleic acid (e.g., DNA, cDNA, RNA, sgRNA to positive charged amino acid(s) may result in "whereby the enzyme in the CRISPR complex has reduced capability of modifying one or more off-target loci as compared to an unmodified enzyme and/or whereby the enzyme in the CRISPR complex has increased capability of modifying the one or more target loci as compared to an unmodified enzyme", e.g., more cutting. As this can be both enhanced on- and off-target cutting (a super cutting Cas9), using such with what is known in the art as a tru-guide or tru-sgRNAs (see, e.g.,

Fu et al., "Improving CRISPR-Cas nuclease specificity using truncated guide RNAs," Nature Biotechnology 32, 279–284 (2014) doi:10.1038/nbt.2808 Received 17 November 2013 Accepted 06 January 2014 Published online 26 January 2014 Corrected online 29 January 2014) to have enhanced on target activity without higher off target cutting or for making super cutting nickases, or for combination with a mutation that renders the Cas9 dead for a super binder.

[00484] Orthologs of SpCas9 can be used in the practice of the invention. A Cas enzyme may be identified 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 (S. pyogenes Cas9) or saCas9 (S. aureus Cas9). StCas9" refers to wild type Cas9 from S. thermophilus, the protein sequence of which is given in the SwissProt database under accession number G3ECR1. Similarly, S pyogenes Cas9 or spCas9 is included in SwissProt under accession number Q99ZW2. 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. 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. Enzymatic action by Cas9 derived from Streptococcus pyogenes or any closely related Cas9 generates double stranded breaks at target site sequences which hybridize to 20 nucleotides of the guide sequence and that have a protospacer-adjacent motif (PAM) sequence (examples include NGG/NRG or a PAM that can be determined as described herein) following the 20 nucleotides of the target sequence. CRISPR activity through Cas9 for site-specific DNA recognition and cleavage is defined by the guide sequence, the tracr sequence that hybridizes in part to the guide sequence and the PAM sequence. Without wishing to be bound by theory, it is believed that the target sequence should be associated with a PAM (protospacer adjacent motif); that is, a short sequence recognized by the CRISPR complex. The precise sequence and length requirements for the PAM differ depending on the Cas used, but PAMs are typically 2-5 base pair sequences adjacent the protospacer (that is, the target sequence. 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 Cas complexed with a guide sequence hybridized 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. More aspects of the CRISPR system are in Karginov and Hannon, The CRISPR system: small RNA-guided defense in bacteria and archaea, Mole Cell 2010, January 15; 37(1): 7. The type II CRISPR locus from Streptococcus pyogenes SF370, which contains a cluster of four genes Cas9, Cas1, Cas2, and Csn1, as well as two non-coding RNA elements, tracrRNA and a characteristic array of repetitive sequences (direct repeats) interspaced by short stretches of non-repetitive sequences (spacers, about 30bp each). In this system, targeted DNA double-strand break (DSB) is generated in four sequential steps. First, two non-coding RNAs, the pre-crRNA array and tracrRNA, are transcribed from the CRISPR locus. Second, tracrRNA hybridizes to the direct repeats of pre-crRNA, which is then processed into mature crRNAs containing individual spacer sequences. Third, the mature crRNA:tracrRNA complex directs Cas to the DNA target consisting of the protospacer and the corresponding PAM via heteroduplex formation between the spacer region of the crRNA and the protospacer DNA. Finally, Cas mediates cleavage of target DNA upstream of PAM to create a DSB within the protospacer. A pre-crRNA array consisting of a single spacer flanked by two direct repeats (DRs) is also encompassed by the term "tracr-mate sequences"). In certain embodiments, Cas may be constitutively present or inducibly present or conditionally present or administered or delivered. Cas optimization may be used to enhance function or to develop new functions, one can generate chimeric Cas proteins. And Cas may be used as a generic DNA binding protein.

[00485] Typically, in the context of an endogenous CRISPR system, formation of a CRISPR complex (comprising a guide sequence hybridized to a target sequence and complexed with one or more Cas proteins) results in cleavage of one or both strands in or near (e.g. within 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 50, or more base pairs from) the target sequence. Without wishing to be bound by theory, the tracr sequence, which may comprise or consist of all or a portion of a wild-type tracr sequence (e.g. about or more than about 20, 26, 32, 45, 48, 54, 63, 67, 85, or more nucleotides of a wild-type tracr sequence), may also form part of a CRISPR complex, such as by

hybridization along at least a portion of the tracr sequence to all or a portion of a tracr mate sequence that is operably linked to the guide sequence.

[00486] In this disclosure, the term "Cas" can mean "Cas9" or a CRISPR enzyme. In the context of the invention, a Cas9 or Cas or CRISPR enzyme is mutated or modified, "whereby the enzyme in the CRISPR complex has reduced capability of modifying one or more off-target loci as compared to an unmodified enzyme" (or like expressions); and, when reading this specification, the terms "Cas9" or "Cas" or "CRISPR enzyme and the like are meant to include mutated or modified Cas9 or Cas or CRISPR enzyme in accordance with the invention, i.e., "whereby the enzyme in the CRISPR complex has reduced capability of modifying one or more off-target loci as compared to an unmodified enzyme" (or like expressions).

Codon optimization and codon usage for expressing a Cas protein

[00487] The nucleic acid molecule encoding a Cas is advantageously codon optimized Cas. An example of a codon optimized sequence, is in this instance a sequence optimized for expression in a eukaryote, e.g., humans (i.e. being optimized for expression in humans), or for another eukaryote, animal or mammal as herein discussed; see, e.g., SaCas9 human codon optimized sequence in WO 2014/093622 (PCT/US2013/074667). 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 is known. In some embodiments, an enzyme coding sequence encoding a Cas is codon optimized for expression in particular cells, such as eukaryotic cells. The eukaryotic cells may be those of or derived from a particular organism, such as a mammal, including but not limited to human, or non-human eukaryote or animal or mammal as herein discussed, e.g., mouse, rat, rabbit, dog, livestock, or non-human mammal or primate. In some embodiments, processes for modifying the germ line genetic identity of human beings and/or processes for modifying the genetic identity of animals which are likely to cause them suffering without any substantial medical benefit to man or animal, and also animals resulting from such processes, may be excluded. In general, codon optimization refers to a process of modifying a nucleic acid sequence for enhanced expression in the host cells of interest by replacing at least one codon (e.g. about or more than about 1, 2, 3, 4, 5, 10, 15, 20, 25, 50, or more codons) of the native sequence with codons that are more frequently or most frequently used in the genes of that host cell while maintaining the native amino acid sequence.

Various species exhibit particular bias for certain codons of a particular amino acid. Codon bias (differences in codon usage between organisms) often correlates with the efficiency of translation of messenger RNA (mRNA), which is in turn believed to be dependent on, among other things, the properties of the codons being translated and the availability of particular transfer RNA (tRNA) molecules. The predominance of selected tRNAs in a cell is generally a reflection of the codons used most frequently in peptide synthesis. Accordingly, genes can be tailored for optimal gene expression in a given organism based on codon optimization. Codon usage tables are readily available, for example, at the "Codon Usage Database" available at www.kazusa.orjp/codon/ and these tables can be adapted in a number of ways. See Nakamura, Y., et al. "Codon usage tabulated from the international DNA sequence databases: status for the year 2000" Nucl. Acids Res. 28:292 (2000). Computer algorithms for codon optimizing a particular sequence for expression in a particular host cell are also available, such as Gene Forge (Aptagen; Jacobus, PA), are also available. In some embodiments, one or more codons (e.g. 1, 2, 3, 4, 5, 10, 15, 20, 25, 50, or more, or all codons) in a sequence encoding a Cas correspond to the most frequently used codon for a particular amino acid.

[00488] In certain embodiments, the methods as described herein may comprise providing a Cas transgenic cell in which one or more nucleic acids encoding one or more guide RNAs are provided or introduced operably connected in the cell with a regulatory element comprising a promoter of one or more gene of interest. As used herein, the term "Cas transgenic cell" refers to a cell, such as a eukaryotic cell, in which a Cas gene has been genomically integrated. The nature, type, or origin of the cell are not particularly limiting according to the present invention. Also the way how the Cas transgene is introduced in the cell is may vary and can be any method as is known in the art. In certain embodiments, the Cas transgenic cell is obtained by introducing the Cas transgene in an isolated cell. In certain other embodiments, the Cas transgenic cell is obtained by isolating cells from a Cas transgenic organism. By means of example, and without limitation, the Cas transgenic cell as referred to herein may be derived from a Cas transgenic eukaryote, such as a Cas knock-in eukaryote. Reference is made to WO 2014/093622 (PCT/US13/74667), incorporated herein by reference. 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. 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. By means of further example reference is made to Platt et. al. (Cell; 159(2):440-455 (2014)), describing a Cas9 knock-in mouse, which is incorporated herein by reference. The Cas transgene can further comprise a Lox-Stop-polyA-Lox(LSL) cassette thereby rendering Cas expression inducible by Cre recombinase. Alternatively, the Cas transgenic cell may be obtained by introducing the Cas transgene in an isolated cell. Delivery systems for transgenes are well known in the art. By means of example, the Cas transgene may be delivered in for instance eukaryotic cell by means of vector (e.g., AAV, adenovirus, lentivirus) and/or particle and/or nanoparticle delivery, as also described herein elsewhere.

[00489] It will be understood by the skilled person that the cell, such as the Cas transgenic cell, as referred to herein may comprise further genomic alterations besides having an integrated Cas gene or the mutations arising from the sequence specific action of Cas when complexed with RNA capable of guiding Cas to a target locus, such as for instance one or more oncogenic mutations, as for instance and without limitation described in Platt et al. (2014), Chen et al., (2014) or Kumar et al.. (2009).

Cas protein with one or more NLS(s)

[00490] In some embodiments, the Cas sequence is fused to one or more nuclear localization sequences (NLSs), such as about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more NLSs. In some embodiments, the Cas comprises about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more NLSs at or near the amino-terminus, about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more NLSs at or near the carboxy-terminus, or a combination of these (e.g. zero or at least one or more NLS at the amino-terminus and zero or at one or more NLS at the carboxy terminus). When more than one NLS is present, each may be selected independently of the others, such that a single NLS may be present in more than one copy and/or in combination with one or more other NLSs present in one or more copies. In a preferred embodiment of the invention, the Cas comprises at most 6 NLSs. In some embodiments, an NLS is considered near the N- or C-terminus when the nearest amino acid of the NLS is within about 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 40, 50, or more amino acids along the polypeptide chain from the N- or C-terminus. Non-limiting examples of NLSs include an NLS sequence derived from: the NLS of the SV40 virus large T-antigen, having the amino acid sequence PKKKRKV(SEQ ID NO: X); the NLS from

nucleoplasmin (e.g. the nucleoplasmin bipartite NLS with the sequence KRPAATKKAGQAKKKK) (SEQ ID NO: X); the c-myc NLS having the amino acid sequence PAAKRVKLD (SEQ ID NO: X) or RQRRNELKRSP(SEQ ID NO: X); the hRNPA1 M9 NLS having the sequence NQSSNFGPMKGGNFGGRSSGPYGGGGQYFAKPRNQGGY(SEQ ID NO: X); the sequence RMRIZFKNKGKDTAELRRRRVEVSVELRKAKKDEQILKRRNV (SEQ ID NO: X) of the IBB domain from importin-alpha; the sequences VSRKRPRP (SEQ ID NO: X) and PPKKARED (SEQ ID NO: X) of the myoma T protein; the sequence POPKKKPL (SEQ ID NO: X) of human p53; the sequence SALIKKKKKMAP (SEQ ID NO: X) of mouse c-abl IV; the sequences DRLRR (SEQ ID NO: X) and PKQKKRK (SEQ ID NO: X) of the influenza virus NS1; the sequence RKLKKKIKKL (SEQ ID NO: X) of the Hepatitis virus delta antigen; the sequence REKKKFLKRR (SEQ ID NO: X) of the mouse Mx1 protein; the sequence KRKGDEVDGVDEVAKKKSKK (SEQ ID NO: X) of the human poly(ADP-ribose) polymerase; and the sequence RKCLQAGMNLEARKTKK (SEQ ID NO: X) of the steroid hormone receptors (human) glucocorticoid. In general, the one or more NLSs are of sufficient strength to drive accumulation of the Cas in a detectable amount in the nucleus of a eukaryotic cell. In general, strength of nuclear localization activity may derive from the number of NLSs in the Cas, the particular NLS(s) used, or a combination of these factors. Detection of accumulation in the nucleus may be performed by any suitable technique. For example, a detectable marker may be fused to the Cas, such that location within a cell may be visualized, such as in combination with a means for detecting the location of the nucleus (e.g. a stain specific for the nucleus such as DAPI). Cell nuclei may also be isolated from cells, the contents of which may then be analyzed by any suitable process for detecting protein, such as immunohistochemistry, Western blot, or enzyme activity assay. Accumulation in the nucleus may also be determined indirectly, such as by an assay for the effect of CRISPR complex formation (e.g. assay for DNA cleavage or mutation at the target sequence, or assay for altered gene expression activity affected by CRISPR complex formation and/or Cas enzyme activity), as compared to a control no exposed to the Cas or complex, or exposed to a Cas lacking the one or more NLSs. In some embodiments, there is no NLS added or fused to the Cas protein.

Delivery of CRISPR system

[00491] Through this disclosure and the knowledge in the art, CRISPR-Cas system, specifically the novel CRISPR systems described herein, or components thereof or nucleic acid molecules thereof (including, for instance HDR template) or nucleic acid molecules encoding or providing components thereof may be delivered by a delivery system herein described both generally and in detail.

General information on vectors

[00492] In general, and throughout this specification, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. Vectors include, but are not limited to, nucleic acid molecules that are single-stranded, double-stranded, or partially double-stranded; nucleic acid molecules that comprise one or more free ends, no free ends (e.g., circular); nucleic acid molecules that comprise DNA, RNA, or both; and other varieties of polynucleotides known in the art. One type of vector is a "plasmid," which refers to a circular double stranded DNA loop into which additional DNA segments can be inserted, such as by standard molecular cloning techniques. Another type of vector is a viral vector, wherein virally-derived DNA or RNA sequences are present in the vector for packaging into a virus (e.g., retroviruses, replication defective retroviruses, adenoviruses, replication defective adenoviruses, and adeno-associated viruses). Viral vectors also include polynucleotides carried by a virus for transfection into a host cell. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (e.g., bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (e.g., non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively-linked. Such vectors are referred to herein as "expression vectors." Vectors for and that result in expression in a eukaryotic cell can be referred to herein as "eukaryotic expression vectors." Common expression vectors of utility in recombinant DNA techniques are often in the form of plasmids.

[00493] Recombinant expression vectors can comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory elements, which may be selected on the basis

of the host cells to be used for expression, that is operatively-linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory element(s) in a manner that allows for expression of the nucleotide sequence (e.g., in an in vitro transcription/translation system or in a host cell when the vector is introduced into the host cell).

[00494] The term "regulatory element" is intended to include promoters, enhancers, internal ribosomal entry sites (IRES), and other expression control elements (e.g., transcription termination signals, such as polyadenylation signals and poly-U sequences). Such regulatory elements are described, for example, in Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Regulatory elements include those that direct constitutive expression of a nucleotide sequence in many types of host cell and those that direct expression of the nucleotide sequence only in certain host cells (e.g., tissue-specific regulatory sequences). A tissue-specific promoter may direct expression primarily in a desired tissue of interest, such as muscle, neuron, bone, skin, blood, specific organs (e.g., liver, pancreas), or particular cell types (e.g., lymphocytes). Regulatory elements may also direct expression in a temporal-dependent manner, such as in a cell-cycle dependent or developmental stage-dependent manner, which may or may not also be tissue or cell-type specific. In some embodiments, a vector comprises one or more pol III promoter (e.g., 1, 2, 3, 4, 5, or more pol III promoters), one or more pol II promoters (e.g., 1, 2, 3, 4, 5, or more pol II promoters), one or more pol I promoters (e.g., 1, 2, 3, 4, 5, or more pol I promoters), or combinations thereof. Examples of pol III promoters include, but are not limited to, U6 and H1 promoters. Examples of pol II promoters include, but are not limited to, the retroviral Rous sarcoma virus (RSV) LTR promoter (optionally with the RSV enhancer), the cytomegalovirus (CMV) promoter (optionally with the CMV enhancer) [see, e.g., Boshart et al, Cell, 41:521-530 (1985)], the SV40 promoter, the dihydrofolate reductase promoter, the β-actin promoter, the phosphoglycerol kinase (PGK) promoter, and the EF1α promoter. Also encompassed by the term "regulatory element" are enhancer elements, such as WPRE; CMV enhancers; the R-U5' segment in LTR of HTLV-I (Mol. Cell. Biol., Vol. 8(1), p. 466-472, 1988); SV40 enhancer; and the intron sequence between exons 2 and 3 of rabbit β-globin (Proc. Natl. Acad. Sci. USA., Vol. 78(3), p. 1527-31, 1981). It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the

level of expression desired, etc. A vector can be introduced into host cells to thereby produce transcripts, proteins, or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein (e.g., clustered regularly interspersed short palindromic repeats (CRISPR) transcripts, proteins, enzymes, mutant forms thereof, fusion proteins thereof, etc.).

[00495] Advantageous vectors include lentiviruses and adeno-associated viruses, and types of such vectors can also be selected for targeting particular types of cells.

[00496] In some embodiments, one or more vectors driving expression of one or more elements of a nucleic acid-targeting system are introduced into a host cell such that expression of the elements of the nucleic acid-targeting system direct formation of a nucleic acid-targeting complex at one or more target sites. For example, a nucleic acid-targeting effector enzyme and a nucleic acid-targeting guide RNA could each be operably linked to separate regulatory elements on separate vectors. RNA(s) of the nucleic acid-targeting system can be delivered to a transgenic nucleic acid-targeting effector protein animal or mammal, e.g., an animal or mammal that constitutively or inducibly or conditionally expresses nucleic acid-targeting effector protein; or an animal or mammal that is otherwise expressing nucleic acid-targeting effector protein or has cells containing nucleic acid-targeting effector protein, such as by way of prior administration thereto of a vector or vectors that code for and express in vivo nucleic acid-targeting effector protein. Alternatively, two or more of the elements expressed from the same or different regulatory elements, may be combined in a single vector, with one or more additional vectors providing any components of the nucleic acid-targeting system not included in the first vector. Nucleic acid-targeting system elements that are combined in a single vector may be arranged in any suitable orientation, such as one element located 5' with respect to ("upstream" of) or 3' with respect to ("downstream" of) a second element. The coding sequence of one element may be located on the same or opposite strand of the coding sequence of a second element, and oriented in the same or opposite direction. In some embodiments, a single promoter drives expression of a transcript encoding a nucleic acid-targeting effector protein and the nucleic acid-targeting guide RNA, embedded within one or more intron sequences (e.g., each in a different intron, two or more in at least one intron, or all in a single intron). In some embodiments, the nucleic acid-targeting effector protein and the nucleic acid-targeting guide RNA may be operably linked to and expressed from the same promoter. Delivery vehicles, vectors, particles, nanoparticles, formulations and components thereof for expression of one or more elements of a

nucleic acid-targeting system are as used in the foregoing documents, such as WO 2014/093622 (PCT/US2013/074667). In some embodiments, a vector comprises one or more insertion sites, such as a restriction endonuclease recognition sequence (also referred to as a "cloning site"). In some embodiments, one or more insertion sites (e.g., about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more insertion sites) are located upstream and/or downstream of one or more sequence elements of one or more vectors. In some embodiments, a vector comprises an insertion site upstream of a tracr mate sequence, and optionally downstream of a regulatory element operably linked to the tracr mate sequence, such that following insertion of a guide sequence into the insertion site and upon expression the guide sequence directs sequence-specific binding of a nucleic acid-targeting complex to a target sequence in a eukaryotic cell. In some embodiments, a vector comprises two or more insertion sites, so as to allow insertion of a guide sequence at each site. In such an arrangement, the two or more guide sequences may comprise two or more copies of a single guide sequence, two or more different guide sequences, or combinations of these. When multiple different guide sequences are used, a single expression construct may be used to target nucleic acid-targeting activity to multiple different, corresponding target sequences within a cell. For example, a single vector may comprise about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, or more guide sequences. In some embodiments, about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more such guide-sequence-containing vectors may be provided, and optionally delivered to a cell. In some embodiments, a vector comprises a regulatory element operably linked to an enzyme-coding sequence encoding a nucleic acid-targeting effector protein. Nucleic acid-targeting effector protein or nucleic acid-targeting guide RNA or RNA(s) can be delivered separately; and advantageously at least one of these is delivered via a particle or nanoparticle complex. Nucleic acid-targeting effector protein mRNA can be delivered prior to the nucleic acid-targeting guide RNA to give time for nucleic acid-targeting effector protein to be expressed. Nucleic acid-targeting effector protein mRNA might be administered 1-12 hours (preferably around 2-6 hours) prior to the administration of nucleic acid-targeting guide RNA. Alternatively, nucleic acid-targeting effector protein mRNA and nucleic acid-targeting guide RNA can be administered together. Advantageously, a second booster dose of guide RNA can be administered 1-12 hours (preferably around 2-6 hours) after the initial administration of nucleic acid-targeting effector protein mRNA + guide RNA.

Additional administrations of nucleic acid-targeting effector protein mRNA and/or guide RNA might be useful to achieve the most efficient levels of genome modification.

General information on vector delivery

[00497] In certain aspects the invention involves vectors, e.g. for delivering or introducing in a cell Cas and/or RNA capable of guiding Cas to a target locus (i.e. guide RNA), but also for propagating these components (e.g. in prokaryotic cells). A used herein, a "vector" is a tool that allows or facilitates the transfer of an entity from one environment to another. It is a replicon, such as a plasmid, phage, or cosmid, into which another DNA segment may be inserted so as to bring about the replication of the inserted segment. Generally, a vector is capable of replication when associated with the proper control elements. In general, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. Vectors include, but are not limited to, nucleic acid molecules that are single-stranded, double-stranded, or partially double-stranded; nucleic acid molecules that comprise one or more free ends, no free ends (e.g. circular); nucleic acid molecules that comprise DNA, RNA, or both; and other varieties of polynucleotides known in the art. One type of vector is a "plasmid," which refers to a circular double stranded DNA loop into which additional DNA segments can be inserted, such as by standard molecular cloning techniques. Another type of vector is a viral vector, wherein virally-derived DNA or RNA sequences are present in the vector for packaging into a virus (e.g. retroviruses, replication defective retroviruses, adenoviruses, replication defective adenoviruses, and adeno-associated viruses (AAVs)). Viral vectors also include polynucleotides carried by a virus for transfection into a host cell. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (e.g. bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (e.g., non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively-linked. Such vectors are referred to herein as "expression vectors." Common expression vectors of utility in recombinant DNA techniques are often in the form of plasmids.

[00498] Recombinant expression vectors can comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant

expression vectors include one or more regulatory elements, which may be selected on the basis of the host cells to be used for expression, that is operatively-linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory element(s) in a manner that allows for expression of the nucleotide sequence (e.g. in an in vitro transcription/translation system or in a host cell when the vector is introduced into the host cell). With regards to recombination and cloning methods, mention is made of U.S. patent application 10/815,730, published September 2, 2004 as US 2004-0171156 A1, the contents of which are herein incorporated by reference in their entirety.

[00499] The vector(s) can include the regulatory element(s), e.g., promoter(s). The vector(s) can comprise Cas encoding sequences, and/or a single, but possibly also can comprise at least 3 or 8 or 16 or 32 or 48 or 50 guide RNA(s) (e.g., sgRNAs) encoding sequences, such as 1-2, 1-3, 1-4 1-5, 3-6, 3-7, 3-8, 3-9, 3-10, 3-8, 3-16, 3-30, 3-32, 3-48, 3-50 RNA(s) (e.g., sgRNAs). In a single vector there can be a promoter for each RNA (e.g., sgRNA), advantageously when there are up to about 16 RNA(s) (e.g., sgRNAs); and, when a single vector provides for more than 16 RNA(s) (e.g., sgRNAs), one or more promoter(s) can drive expression of more than one of the RNA(s) (e.g., sgRNAs), e.g., when there are 32 RNA(s) (e.g., sgRNAs), each promoter can drive expression of two RNA(s) (e.g., sgRNAs), and when there are 48 RNA(s) (e.g., sgRNAs), each promoter can drive expression of three RNA(s) (e.g., sgRNAs). By simple arithmetic and well established cloning protocols and the teachings in this disclosure one skilled in the art can readily practice the invention as to the RNA(s) (e.g., sgRNA(s) for a suitable exemplary vector such as AAV, and a suitable promoter such as the U6 promoter, e.g., U6-sgRNAs. For example, the packaging limit of AAV is ~4.7 kb. The length of a single U6-sgRNA (plus restriction sites for cloning) is 361 bp. Therefore, the skilled person can readily fit about 12-16, e.g., 13 U6-sgRNA cassettes in a single vector. This can be assembled by any suitable means, such as a golden gate strategy used for TALE assembly (http://www.genome-engineering.org/taleffectors/). The skilled person can also use a tandem guide strategy to increase the number of U6-sgRNAs by approximately 1.5 times, e.g., to increase from 12-16, e.g., 13 to approximately 18-24, e.g., about 19 U6-sgRNAs. Therefore, one skilled in the art can readily reach approximately 18-24, e.g., about 19 promoter-RNAs, e.g., U6-sgRNAs in a single vector, e.g., an AAV vector. A further means for increasing the number of promoters and RNAs, e.g., sgRNA(s) in a vector is to use a

single promoter (e.g., U6) to express an array of RNAs, e.g., sgRNAs separated by cleavable sequences. And an even further means for increasing the number of promoter-RNAs, e.g., sgRNAs in a vector, is to express an array of promoter-RNAs, e.g., sgRNAs separated by cleavable sequences in the intron of a coding sequence or gene; and, in this instance it is advantageous to use a polymerase II promoter, which can have increased expression and enable the transcription of long RNA in a tissue specific manner. (see, e.g., http://nar.oxfordjournals.org/content/34/7/e53.short, http://www.nature.com/mt/journal/v16/n9/abs/mt2008144a.html). In an advantageous embodiment, AAV may package U6 tandem sgRNA targeting up to about 50 genes. Accordingly, from the knowledge in the art and the teachings in this disclosure the skilled person can readily make and use vector(s), e.g., a single vector, expressing multiple RNAs or guides or sgRNAs under the control or operatively or functionally linked to one or more promoters—especially as to the numbers of RNAs or guides or sgRNAs discussed herein, without any undue experimentation.

[00500] The guide RNA(s), e.g., sgRNA(s) encoding sequences and/or Cas encoding sequences, can be functionally or operatively linked to regulatory element(s) and hence the regulatory element(s) drive expression. The promoter(s) can be constitutive promoter(s) and/or conditional promoter(s) and/or inducible promoter(s) and/or tissue specific promoter(s). The promoter can be selected from the group consisting of RNA polymerases, pol I, pol II, pol III, T7, U6, H1, retroviral Rous sarcoma virus (RSV) LTR promoter, the cytomegalovirus (CMV) promoter, the SV40 promoter, the dihydrofolate reductase promoter, the β-actin promoter, the phosphoglycerol kinase (PGK) promoter, and the EF1α promoter. An advantageous promoter is the promoter is U6.

Vector delivery

[00501] Vector delivery, e.g., plasmid, 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 any suitable vector, e.g., plasmid or viral vectors, such as 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 vectors, e.g., plasmid or viral vectors. In some embodiments, the vector, e.g., plasmid or viral vector is delivered to the tissue of interest by, for

example, an intramuscular injection, while other times the 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 choice, 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.

[00502] Such a dosage may further contain, for example, a carrier (water, saline, ethanol, glycerol, lactose, sucrose, calcium phosphate, gelatin, dextran, agar, pectin, peanut oil, sesame oil, etc.), a diluent, a pharmaceutically-acceptable carrier (e.g., phosphate-buffered saline), a pharmaceutically-acceptable excipient, and/or other compounds known in the art. The dosage may further contain one or more pharmaceutically acceptable salts such as, for example, a mineral acid salt such as a hydrochloride, a hydrobromide, a phosphate, a sulfate, etc.; and the salts of organic acids such as acetates, propionates, malonates, benzoates, etc. Additionally, auxiliary substances, such as wetting or emulsifying agents, pH buffering substances, gels or gelling materials, flavorings, colorants, microspheres, polymers, suspension agents, etc. may also be present herein. In addition, one or more other conventional pharmaceutical ingredients, such as preservatives, humectants, suspending agents, surfactants, antioxidants, anticaking agents, fillers, chelating agents, coating agents, chemical stabilizers, etc. may also be present, especially if the dosage form is a reconstitutable form. Suitable exemplary ingredients include microcrystalline cellulose, carboxymethylcellulose sodium, polysorbate 80, phenylethyl alcohol, chlorobutanol, potassium sorbate, sorbic acid, sulfur dioxide, propyl gallate, the parabens, ethyl vanillin, glycerin, phenol, parachlorophenol, gelatin, albumin and a combination thereof. A thorough discussion of pharmaceutically acceptable excipients is available in REMINGTON'S PHARMACEUTICAL SCIENCES (Mack Pub. Co., N.J. 1991) which is incorporated by reference herein.

[00503] In an embodiment herein the delivery is via an adenovirus, which may be at a single booster dose containing at least 1 x 10^5 particles (also referred to as particle units, pu) of adenoviral vector. In an embodiment herein, the dose preferably is at least about 1 x 10^6 particles (for example, about 1 x 10^6-1 x 10^{12} particles), more preferably at least about 1 x 10^7 particles, more preferably at least about 1 x 10^8 particles (e.g., about 1 x 10^8-1 x 10^{11} particles or

about 1 x 10^8-1 x 10^{12} particles), and most preferably at least about 1 x 10^0 particles (e.g., about 1 x 10^9-1 x 10^{10} particles or about 1 x 10^9-1 x 10^{12} particles), or even at least about 1 x 10^{10} particles (e.g., about 1 x 10^{10}-1 x 10^{12} particles) of the adenoviral vector. Alternatively, the dose comprises no more than about 1 x 10^{14} particles, preferably no more than about 1 x 10^{13} particles, even more preferably no more than about 1 x 10^{12} particles, even more preferably no more than about 1 x 10^{11} particles, and most preferably no more than about 1 x 10^{10} particles (e.g., no more than about 1 x 10^9 articles). Thus, the dose may contain a single dose of adenoviral vector with, for example, about 1 x 10^6 particle units (pu), about 2 x 10^6 pu, about 4 x 10^6 pu, about 1 x 10^7 pu, about 2 x 10^7 pu, about 4 x 10^7 pu, about 1 x 10^8 pu, about 2 x 10^8 pu, about 4 x 10^8 pu, about 1 x 10^9 pu, about 2 x 10^9 pu, about 4 x 10^9 pu, about 1 x 10^{10} pu, about 2 x 10^{10} pu, about 4 x 10^{10} pu, about 1 x 10^{11} pu, about 2 x 10^{11} pu, about 4 x 10^{11} pu, about 1 x 10^{12} pu, about 2 x 10^{12} pu, or about 4 x 10^{12} pu of adenoviral vector. See, for example, the adenoviral vectors in U.S. Patent No. 8,454,972 B2 to Nabel, et. al., granted on June 4, 2013; incorporated by reference herein, and the dosages at col 29, lines 36-58 thereof. In an embodiment herein, the adenovirus is delivered via multiple doses.

[00504] In an embodiment herein, the delivery is via an AAV. A therapeutically effective dosage for in vivo delivery of the AAV to a human is believed to be in the range of from about 20 to about 50 ml of saline solution containing from about 1 x 10^{10} to about 1 x 10^{10} functional AAV/ml solution. The dosage may be adjusted to balance the therapeutic benefit against any side effects. In an embodiment herein, the AAV dose is generally in the range of concentrations of from about 1 x 10^5 to 1 x 10^{50} genomes AAV, from about 1 x 10^8 to 1 x 10^{20} genomes AAV, from about 1 x 10^{10} to about 1 x 10^{16} genomes, or about 1 x 10^{11} to about 1 x 10^{16} genomes AAV. A human dosage may be about 1 x 10^{13} genomes AAV. Such concentrations may be delivered in from about 0.001 ml to about 100 ml, about 0.05 to about 50 ml, or about 10 to about 25 ml of a carrier solution. Other effective dosages can be readily established by one of ordinary skill in the art through routine trials establishing dose response curves. See, for example, U.S. Patent No. 8,404,658 B2 to Hajjar, et al., granted on March 26, 2013, at col. 27, lines 45-60.

Packaging and Promoters generally

[00505] Ways to package Cas9 coding nucleic acid molecules, e.g., DNA, into vectors, e.g., viral vectors, to mediate genome modification *in vivo* include:

[00506] To achieve NHEJ-mediated gene knockout:

Single virus vector:

- Vector containing two or more expression cassettes:
- Promoter-Cas9 coding nucleic acid molecule -terminator
- Promoter-gRNA1-terminator
- Promoter-gRNA2-terminator
- Promoter-gRNA(N)-terminator (up to size limit of vector)

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)

[00507] To mediate homology-directed repair.

- In addition to the single and double virus vector approaches described above, an additional vector may used to deliver a homology-direct repair template.

[00508] The promoter used to drive Cas9 coding nucleic acid molecule expression can include:

- 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 potential toxicity due to over expression of Cas9.
- For ubiquitous expression, can use promoters: CMV, CAG, CBh, PGK, SV40, Ferritin heavy or light chains, etc.

- For brain or other CNS expression, can use promoters: SynapsinI for all neurons, CaMKIIalpha for excitatory neurons, GAD67 or GAD65 or VGAT for GABAergic neurons, etc.
- For liver expression, can use Albumin promoter.
- For lung expression, can use SP-B.
- For endothelial cells, can use ICAM.
- For hematopoietic cells can use IFNbeta or CD45.
- For Osteoblasts can use OG-2.

[00509] The 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

Crystallization and Structure of CRISPR-Cas9

[00510] Crystallization of CRISPR-Cas9 and Characterization of Crystal Structure: The crystals can be obtained by techniques of protein crystallography, including batch, liquid bridge, dialysis, vapor diffusion and hanging drop methods. Generally, the crystals are grown by dissolving substantially pure CRISPR-Cas9 and a nucleic acid molecule to which it binds in an aqueous buffer containing a precipitant at a concentration just below that necessary to precipitate. Water is removed by controlled evaporation to produce precipitating conditions, which are maintained until crystal growth ceases. *See* Nishimasu *et al.*

[00511] Uses of the Crystals, Crystal Structure and Atomic Structure Co-Ordinates: The crystals, and particularly the atomic structure co-ordinates obtained therefrom, have a wide variety of uses. The crystals and structure co-ordinates are particularly useful for identifying compounds (nucleic acid molecules) that bind to CRISPR-Cas9, and CRISPR-Cas9s that can bind to particular compounds (nucleic acid molecules). Thus, the structure co-ordinates described herein can be used as phasing models in determining the crystal structures of additional synthetic or mutated CRISPR-Cas9s, Cas9s, nickases, binding domains. The provision of the crystal structure of CRISPR-Cas9 complexed with a nucleic acid molecule may provide the skilled artisan with a insight into CRISPR-Cas9. This insight provides a means to design modified CRISPR-Cas9s, such as by attaching thereto a functional group, such as a repressor or activator. While one can attach a functional group such as a repressor or activator to the N or C

terminal of CRISPR-Cas9, the crystal structure demonstrates that the N terminal seems obscured or hidden, whereas the C terminal is more available for a functional group such as repressor or activator. Moreover, the crystal structure demonstrates that there is a flexible loop between approximately CRISPR-Cas9 (*S. pyogenes*) residues 534-676 which is suitable for attachment of a functional group such as an activator or repressor. Attachment can be via a linker, e.g., a flexible glycine-serine (GlyGlyGlySer) or (GGGS)3 or a rigid alpha-helical linker such as (Ala(GluAlaAlaAlaLys)Ala). In addition to the flexible loop there is also a nuclease or H3 region, an H2 region and a helical region. By "helix" or "helical", is meant a helix as known in the art, including, but not limited to an alpha-helix. Additionally, the term helix or helical may also be used to indicate a c-terminal helical element with an N-terminal turn.

[00512] The provision of the crystal structure of CRISPR-Cas9 complexed with a nucleic acid molecule allows a novel approach for drug or compound discovery, identification, and design for compounds that can bind to CRISPR-Cas9 and thus the disclosure provides tools useful in diagnosis, treatment, or prevention of conditions or diseases of multicellular organisms, e.g., algae, plants, invertebrates, fish, amphibians, reptiles, avians, mammals; for example domesticated plants, animals (e.g., production animals such as swine, bovine, chicken; companion animal such as felines, canines, rodents (rabbit, gerbil, hamster); laboratory animals such as mouse, rat), and humans. Accordingly, the disclosure provides a computer-based method of rational design of CRISPR-Cas9 complexes. This rational design can comprise: providing the structure of the CRISPR-Cas9 complex as defined by some or all (e.g., at least 2 or more, e.g., at least 5, advantageously at least 10, more advantageously at least 50 and even more advantageously at least 100 atoms of the structure) co-ordinates of the Crystal Structure Table and/or in Figure(s) concerning the crystal structure; *see* Nishimasu *et al.*; providing a structure of a desired nucleic acid molecule as to which a CRISPR-Cas9 complex is desired; and fitting the structure of the CRISPR-Cas9 complex as defined by some or all co-ordinates to the desired nucleic acid molecule, including in said fitting obtaining putative modification(s) of the CRISPR-Cas9 complex as defined by some or all co-ordinates for said desired nucleic acid molecule to bind for CRISPR-Cas9 complex(es) involving the desired nucleic acid molecule. The method or fitting of the method may use the co-ordinates of atoms of interest of the CRISPR-Cas9 complex as defined by some or all co-ordinates which are in the vicinity of the active site or binding region (e.g., at least 2 or more, e.g., at least 5, advantageously at least 10,

more advantageously at least 50 and even more advantageously at least 100 atoms of the structure) in order to model the vicinity of the active site or binding region. These co-ordinates may be used to define a space which is then screened "in silico" against a desired or candidate nucleic acid molecule. Thus, the disclosure provides a computer-based method of rational design of CRISPR-Cas9 complexes. This method may include: providing the co-ordinates of at least two atoms of the Crystal Structure Table ("selected co-ordinates"); *see* Nishimasu *et al.*; providing the structure of a candidate or desired nucleic acid molecule; and fitting the structure of the candidate to the selected co-ordinates. In this fashion, the skilled person may also fit a functional group and a candidate or desired nucleic acid molecule. For example, providing the structure of the CRISPR-Cas9 complex as defined by some or all (e.g., at least 2 or more, e.g., at least 5, advantageously at least 10, more advantageously at least 50 and even more advantageously at least 100 atoms of the structure) co-ordinates; providing a structure of a desired nucleic acid molecule as to which a CRISPR-Cas9 complex is desired; fitting the structure of the CRISPR-Cas9 complex as defined by some or all co-ordinates in the Crystal Structure Table and/or in Figures concerning the crystal structure; *see* Nishimasu *et al.*; to the desired nucleic acid molecule, including in said fitting obtaining putative modification(s) of the CRISPR-Cas9 complex as defined for said desired nucleic acid molecule to bind for CRISPR-Cas9 complex(es) involving the desired nucleic acid molecule; selecting putative fit CRISPR-Cas9—desired nucleic acid molecule complex(es), fitting such putative fit CRISPR-Cas9—desired nucleic acid molecule complex(es) to the functional group (e.g., activator, repressor), e.g., as to locations for situating the functional group (e.g., positions within the flexible loop) and/or putative modifications of the putative fit CRISPR-Cas9—desired nucleic acid molecule complex(es) for creating locations for situating the functional group.

[00513] HOWEVER, knowledge of the SpCas9 crystal structure (*see* Nishimasu *et al.*) COULD NOT have predicted the reduction of off-target effects achieved by the mutations of the instant invention; or that particular mutations could achieve reduction of off-target effects, as herein disclosed. But, now that, through this disclosure, there is knowledge of mutations that provide or achieve reduction in off-target effects, the skilled person can readily apply the teachings herein, in conjunction with the knowledge of the SpCas9 crystal structure, and the knowledge of Cas9 sequences to make sequence and structural analyses across Cas9s to determine analogous amino acids that can be mutated or modified in a manner analogous hereto,

DEMANDE OU BREVET VOLUMINEUX

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

CECI EST LE TOME 1 DE 5
CONTENANT LES PAGES 1 À 157

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 5
CONTAINING PAGES 1 TO 157

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

NOM DU FICHIER / FILE NAME :

NOTE POUR LE TOME / VOLUME NOTE:

What is claimed is:

1. An engineered CRISPR protein, wherein:

the protein complexes with a nucleic acid molecule comprising RNA to form a CRISPR complex,

wherein when in the CRISPR complex, the nucleic acid molecule targets one or more target polynucleotide loci

the protein comprises at least one modification compared to the unmodified protein,

wherein the CRISPR complex comprising the modified protein has altered activity as compared to the complex comprising the unmodified protein.

2. The engineered CRISPR protein of claim 1, wherein the altered activity comprises altered binding property as to the nucleic acid molecule comprising RNA or the target polynucleotide loci, altered binding kinetics as to the nucleic acid molecule comprising RNA or the target polynucleotide loci, or altered binding specificity as to the nucleic acid molecule comprising RNA or the target polynucleotide loci compared to off-target polynucleotide loci.

3. The engineered CRISPR protein of claim 1 or 2, wherein the altered activity comprises increased targeting efficiency or decreased off-target binding.

4. The engineered CRISPR protein of any one of claims 1 to 3, wherein the altered activity comprises modified cleavage activity.

5. The engineered CRISPR protein of claim 4, wherein the modified cleavage activity comprises increased cleavage activity as to the target polynucleotide loci.

6. The engineered CRISPR protein of claim 4, wherein the modified cleavage activity comprises decreased cleavage activity as to the target polynucleotide loci.

7. The engineered CRISPR protein of any one of claims 4 to 6, wherein the modified cleavage activity comprises decreased cleavage activity as to off-target polynucleotide loci.

8. The engineered CRISPR protein of any one of claims 4 to 6, wherein the modified cleavage activity comprises increased cleavage activity as to off-target polynucleotide loci.

9. The engineered CRISPR protein of any one of the preceding claims wherein the altered activity comprises altered helicase kinetics.

10. The engineered CRISPR protein of any one of the preceding claims wherein the modified CRISPR protein comprises a modification that alters association of the protein with the nucleic acid molecule comprising RNA, or a strand of the target polynucleotide loci, or a strand of off-target polynucleotide loci.

11. The engineered CRISPR protein of any one of the preceding claims wherein the modified CRISPR protein comprises a modification that alters formation of the CRISPR complex.

12. The engineered CRISPR protein of any one of the preceding claims wherein the modified CRISPR protein comprises a modification that alters targeting of the nucleic acid molecule to the polynucleotide loci.

13. The engineered CRISPR protein of any one of claims 10 to 12, wherein the modification comprises a mutation in a region of the protein that associates with the nucleic acid molecule.

14. The engineered CRISPR protein of any one of claims 10 to 12, wherein the modification comprises a mutation in a region of the protein that associates with a strand of the target polynucleotide loci.

15. The engineered CRISPR protein of any one of claims 10 to 12, wherein the modification comprises a mutation in a region of the protein that associates with a strand of the off-target polynucleotide loci.

16. The engineered CRISPR protein of any one of claims 10 to 15, wherein the modification or mutation comprises decreased positive charge in a region of the protein that associates with the nucleic acid molecule comprising RNA, or a strand of the target polynucleotide loci, or a strand of off-target polynucleotide loci.

17. The engineered CRISPR protein of any one of claims 10 to 15, wherein the modification or mutation comprises decreased negative charge in a region of the protein that associates with the nucleic acid molecule comprising RNA, or a strand of the target polynucleotide loci, or a strand of off-target polynucleotide loci.

18. The engineered CRISPR protein of any one of claims 10 to 15, wherein the modification or mutation comprises increased positive charge in a region of the protein that associates with the nucleic acid molecule comprising RNA, or a strand of the target polynucleotide loci, or a strand of off-target polynucleotide loci.

19. The engineered CRISPR protein of any one of claims 10 to 15, wherein the modification or mutation comprises increased negative charge in a region of the protein that associates with the nucleic acid molecule comprising RNA, or a strand of the target polynucleotide loci, or a strand of off-target polynucleotide loci.

20. The engineered CRISPR protein of any one of claims 10 to 19, wherein the modification or mutation increases steric hindrance between the protein and the nucleic acid molecule comprising RNA, or a strand of the target polynucleotide loci, or a strand of off-target polynucleotide loci.

21. The engineered CRISPR protein of any one of claims 10 to 19, wherein the modification or mutation comprises a substitution of Lys, His, Arg, Glu, Asp, Ser, Gly, or Thr.

22. The engineered CRISPR protein of any one of claims 10 to 19, wherein the modification or mutation comprises a substitution with Gly, Ala, Ile, Glu, or Asp.

23. The engineered CRISPR protein of any one of claims 10 to 19, wherein the modification or mutation comprises an amino acid substitution in a binding groove.

24. The engineered CRISPR protein of any one of claims 10 to 19, wherein the binding groove is between the RuvC and HNH domains.

25. The engineered CRISPR protein of any one of claims 10 to 19, wherein the modification or mutation comprises a mutation in a RuvCI, RuvCIII, RuvCIII or HNH domain.

26. The engineered CRISPR protein of any one of claims 10 to 19, wherein the modification or mutation comprises an amino acid substitution at one or more of positions 12, 13, 63, 415, 610, 775, 779, 780, 810, 832, 848, 855, 861, 862, 866, 961, 968, 974, 976, 982, 983, 1000, 1003, 1014, 1047, 1060, 1107, 1108, 1109, 1114, 1129, 1240, 1289, 1296, 1297, 1300, 1311, and 1325 with reference to amino acid position numbering of SpCas9.

27. The engineered CRISPR protein of claim 26, wherein the modification or mutation at position 63, 415, 775, 779, 780, 810, 832, 848, 855, 861, 862, 866, 961, 968, 974, 976, 982, 983, 1000, 1003, 1014, 1047, 1060, 1107, 1108, 1109, 1114, 1129, 1240, 1289, 1296, 1297, 1300, 1311, or 1325 comprises an alanine substitution.

28. The engineered CRISPR protein of claim 26, wherein the modification or mutation at position 12 comprises an aspartic acid substitution.

29. The engineered CRISPR protein of claim 26, wherein the modification or mutation at position 13 comprises an isoleucine substitution.

30. The engineered CRISPR protein of claim 26, wherein the modification or mutation at position 610 comprises a glycine substitution.

31. The engineered CRISPR protein of claim 26, wherein the modification or mutation at position 799 comprises a leucine substitution.

32. The engineered CRISPR protein of claim 26, wherein the modification or mutation at position 1129 comprises a glutamic acid substitution.

33. The engineered CRISPR protein of claim 26, wherein the modification or mutation comprises K775A, E779L, Q807A, R780A, K810A, R832A, K848A, K855A, K862A, K961A, K968A, K974A, R976A, H983A, K1000A, K1014A, K1047A, K1060A, K1003A, S1109A, H1240A, K1289A, K1296A, H1297A, K1300A, H1311A, or K1325A.

34. The engineered CRISPR protein of claim 26, wherein the modification or mutation comprises R783A and A1322T, or R780A and K810A, or R780A and K855A, or R780A and R976A, or K848A and R976A, or K855A and R976A, or R780A and K848A, or

K810A and K848A, or K848A and K855A, or K810A and K855A, or H982A and R1060A, or H982A and R1003A, or K1003A and R1060A, or R780A and H982A, or K810A and H982A, or K848A and H982A, or K855A and H982A, or R780A and K1003A, or K810A and R1003A, or K848A and K1003A, or K848A and K1007A, or R780A and R1060A, or K810A and R1060A, or K848A and R1060A, or R780A and R1114A, or K848A and R1114A, or R63A and K855A, or R63A and H982A, or H415A and R780A, or H415A and K848A, or K848A and E1108A, or K810A and K1003A, or R780A and R1060A, or K810A and R1060A, or K848A and R1060A.

35. The engineered CRISPR protein of claim 26, wherein the modification or mutation comprises H982A, K1003A, and K1129E, or R780A, K1003A, and R1060A, or K810A, K1003A, and R1060A, or K848A, K1003A, and R1060A, or K855A, K1003A, and R1060A, or H982A, K1003A, and R1060A, or R63A, K848A, and R1060A, or T13I, R63A, and K810A, or G12D, R63A, and R1060A.

36. The engineered CRISPR protein of claim 26, wherein the modification or mutation comprises R63A, E610G, K855A, and R1060A, or R63A, K855A, R1060A, and E610G.

37. The engineered CRISPR protein of any one of the preceding claims, wherein the protein comprises a Type II CRISPR protein.

38. The engineered CRISPR protein of any one of the preceding claims, wherein the CRISPR protein comprises a CRISPR protein from an organism from a genus comprising *Streptococcus*, *Campylobacter*, *Nitratifractor*, *Staphylococcus*, *Parvibaculum*, *Roseburia*, *Neisseria*, *Gluconacetobacter*, *Azospirillum*, *Sphaerochaeta*, *Lactobacillus*, *Eubacterium* or *Corynebacter*.

39. The engineered CRISPR protein of any one of the preceding claims, wherein the protein comprises a Cas9 protein.

40. The engineered CRISPR protein of any one of the preceding claims, wherein the CRISPR protein comprises a chimeric Cas9 protein comprising a first fragment from a first Cas9

ortholog and a second fragment from a second Cas9 ortholog, and the first and second Cas9 orthologs are different.

41. The engineered CRISPR protein of claim 40, wherein at least one of the first and second Cas9 orthologs comprises a Cas9 from an organism comprising *Streptococcus*, *Campylobacter*, *Nitratifractor*, *Staphylococcus*, *Parvibaculum*, *Roseburia*, *Neisseria*, *Gluconacetobacter*, *Azospirillum*, *Sphaerochaeta*, *Lactobacillus*, *Eubacterium* or *Corynebacter*.

42. The engineered CRISPR protein of any one of the preceding claims, wherein the the CRISPR protein comprises one or more nuclear localization signal (NLS) domains.

43. The engineered CRISPR protein of any one of the preceding claims, wherein the CRISPR protein comprises at least two or more NLSs.

44. The engineered CRISPR protein of any one of the preceding claims, wherein the CRISPR protein comprises one or more heterologous functional domains.

45. The engineered CRISPR protein of claim 44, wherein the one or more heterologous functional domains comprises one or more transcriptional activation domains.

46. The engineered CRISPR protein of claim 45, wherein the transcriptional activation domain comprises VP64.

47. The engineered CRISPR protein of claim 44, wherein the one or more heterologous functional domains comprises one or more transcriptional repression domains.

48. The engineered CRISPR protein of claim 47, wherein the transcriptional repression domain comprises a KRAB domain or a SID domain.

49. The engineered CRISPR protein of claim 44, wherein the one or more heterologous functional domains comprises one or more nuclease domains.

50. The engineered CRISPR protein of claim 49, wherein a nuclease domain comprises Fok1.

51. The engineered CRISPR protein of claim 44, wherein the one or more heterologous functional domains have one or more of the following activities: methylase activity, demethylase activity, transcription activation activity, transcription repression activity, transcription release factor activity, histone modification activity, nuclease activity, single-strand RNA cleavage activity, double-strand RNA cleavage activity, single-strand DNA cleavage activity, double-strand DNA cleavage activity and nucleic acid binding activity.

52. An engineered CRISPR protein according to any one of the preceding claims, wherein the CRISPR protein is encoded by a nucleotide sequence which is codon optimized for expression in a eukaryote.

53. A composition comprising the engineered CRISPR protein of any one of the preceding claims.

54. A system comprising the engineered CRISPR protein of any one of the preceding claims and a nucleic acid molecule comprising RNA.

55. A vector system comprising one or more vectors, wherein the one or more vectors comprises:

a) a first regulatory element operably linked to a nucleotide sequence encoding the engineered CRISPR protein of any one of the preceding claims; and

b) a second regulatory element operably linked to one or more nucleotide sequences encoding one or more nucleic acid molecules comprising a guide RNA comprising a guide sequence, a tracr sequence, and a tracr mate sequence, wherein components (a) and (b) are located on same or different vectors.

56. An isolated eukaryotic cell comprising the system of claim 54 or 55.

57. A method of modulating gene expression, wherein the method comprises introducing the engineered CRISPR protein or system of any one of the preceding claims into a cell.

58. The method of claim 59, wherein the cell is a eukaryotic or a prokaryotic cell.

59. The method of claim 59, wherein the method is *ex vivo* or *in vitro*.

60. A method of treating a disease, disorder or infection in an individual in need thereof comprising administering an effective amount of the engineered CRISPR protein or composition of any one of the preceding claims.

61. A method of altering the expression of a genomic locus of interest in a mammalian cell comprising

contacting the cell with the composition of any of claims 1-55 and thereby delivering the vector and allowing the CRISPR-Cas complex to form and bind to target, and

determining if the expression of the genomic locus has been altered.

62. The method of claim 61, which comprises genomic DNA cleavage resulting in decreased transcription of a gene.

63. The method of claim 61 wherein altering expression comprises genome editing.

64. The method of claim 61, wherein altering expression comprises increasing expression of a gene product.

65. The method of claim 61, wherein altering expression comprises modification of a gene product.

66. An isolated cell having altered expression of a genomic locus from the method of any one of claim 61 - 65, wherein the altered expression is in comparison with a cell that has not been subjected to the method of altering the expression of the genomic locus.

67. An isolated cell line from the cell of claim 66.

68. A method of modifying a target locus of interest, the method comprising delivering to said locus a non-naturally occurring or engineered composition comprising a Cas9 protein of any one of claims 1 to 52, and one or more nucleic acid components, wherein the Cas9 protein forms a complex with the one or more nucleic acid components and upon binding of the

complex to a target locus of interest, the Cas9 protein induces a modification of the target locus of interest.

69. The method of claim 68, wherein the target locus of interest is within a cell.

70. The method of claim 69, wherein the cell is a eukaryotic cell.

71. The method of claim 69, wherein the cell is an animal or human cell.

72. The method of claim 69, wherein the cell is a plant cell.

73. The method of claim 68, wherein the target locus of interest is comprised in a DNA molecule *in vitro*.

74. The method of any one of claims 68-73, wherein said non-naturally occurring or engineered composition comprising a Cas9 protein and one or more nucleic acid components is delivered to the cell as one or more polynucleotide molecules.

75. The method of any one of claims 68-73, wherein Cas9 protein comprises one or more nuclear localization signal(s) (NLS(s)).

76. The method of any one of claims 74-75, wherein the one or more polynucleotide molecules are comprised within one or more vectors.

77. The method of any one of claims 74-76, wherein the one or more polynucleotide molecules comprise one or more regulatory elements operably configured to express the Cas9 protein and/or the nucleic acid component(s), optionally wherein the one or more regulatory elements comprise inducible promoters.

78. The method of any one of claims 74-77 wherein the one or more polynucleotide molecules or the one or more vectors are comprised in a delivery system.

79. The method of any one of claims 74-78, wherein the system or the one or more polynucleotide molecules are delivered via particles, vesicles, or one or more viral vectors.

80. The method of claim 79 wherein the particles comprise a lipid, a sugar, a metal or a protein.

81. The method of claim 79 wherein the vesicles comprise exosomes or liposomes.

82. The method of claim 79 wherein the one or more viral vectors comprise one or more of adenovirus, one or more lentivirus or one or more adeno-associated virus.

83. The method of any one of claims 68-79, which is a method of modifying a cell, a cell line or an organism by manipulation of one or more target sequences at genomic loci of interest.

84. A cell from the method of claim 83, or progeny thereof, wherein the cell comprises a modification not present in a cell not subjected to the method.

85. The cell of claim 84, of progeny thereof, wherein the cell not subjected to the method comprises an abnormality and the cell from the method has the abnormality addressed or corrected.

86. A cell product from the cell or progeny thereof of claim 84, wherein the product is modified in nature or quantity with respect to a cell product from a cell not subjected to the method.

87. The cell product of claim 86, wherein the cell not subjected to the method comprises an abnormality and the cell product reflects the abnormality having been addressed or corrected by the method.

88. An *in vitro*, *ex* vivo or *in vivo* host cell or cell line or progeny thereof comprising a system comprising a Cas9 protein of any one of claims 1 to 52, and one or more nucleic acid components, wherein the Cas9 protein forms a complex with the one or more nucleic acid components and upon binding of the complex to a target locus of interest, the Cas9 protein induces a modification of the target locus of interest.

89. The host cell or cell line or progeny thereof according to claim 88, wherein the cell is a eukaryotic cell.

90. The host cell or cell line or progeny thereof according to claim 89, wherein the cell is an animal cell.

91. The host cell or cell line or progeny thereof according to claim 89, wherein the cell is a human cell.

92. The host cell, cell line or progeny thereof according to claim 89, comprising a stem cell or stem cell line.

93. The host cell or cell line or progeny thereof according to claim 89, wherein the cell is a plant cell.

94. A method of producing a plant, having a modified trait of interest encoded by a gene of interest, said method comprising comprising contacting a plant cell with a system according to any one of claims 54-55 or subjecting the plant cell to a method according to claim 68-83, thereby either modifying or introducing said gene of interest, and regenerating a plant from said plant cell.

95. A method of identifying a trait of interest in a plant, said trait of interest encoded by a gene of interest, said method comprising comprising contacting a plant cell with a system according to any one of claims 54-55 or subjecting the plant cell to a method according to claim 68-83, thereby identifying said gene of interest.

96. The method of claims 95, further comprising introducing the identified gene of interest into a plant cell or plant cell line or plant germplasm and generating a plant therefrom, whereby the plant contains the gene of interest.

97. The method of claim 96 wherein the plant exhibits the trait of interest.

98. A particle comprising a system according to any one of claims 54-55.

99. The particle of claim 98, wherein the particle comprises the Cas9 protein of any one of claims 1-52 protein complexed with the guide RNA.

100. The particle of claim 98, wherein the particle comprises eSpCas9(1.1) complexed with the guide RNA.

101. The complex, nucleic acid molecule comprising RNA, or protein of claim 1, wherein the complex, nucleic acid molecule comprising RNA or protein is conjugated to at least one sugar moiety, optionally N-acetyl galactosamine (GalNAc), in particular triantennary GalNAc.

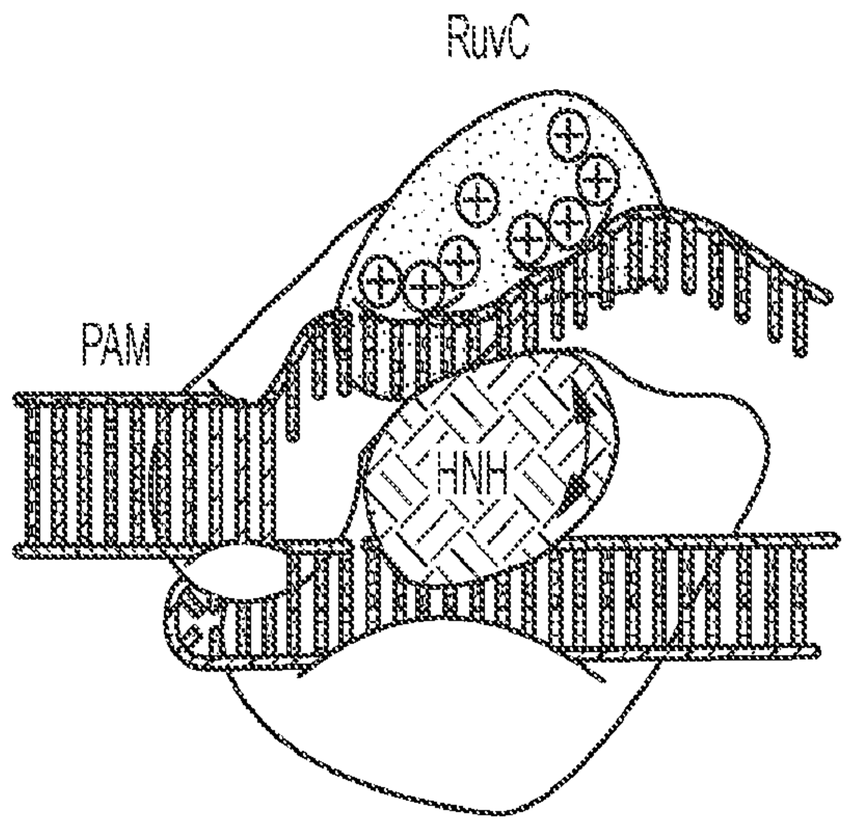
102. The complex, nucleic acid component or protein of claim 68, wherein the complex, nucleic acid component or protein is conjugated to at least one sugar moiety, optionally N-acetyl galactosamine (GalNAc), in particular triantennary GalNAc.

103. A method of improving the specificity of a CRISPR system by providing an engineered CRISPR protein having modification according to any one of claims 1-36.

104. Use of an engineered CRISPR protein to improve the specificity of a CRISPR system, wherein the CRISPR protein is modified according to any one of claims 1-36.

Mutation of positivly charged residues binding to non-targeted gDNA strand improves specificity.

Cas9 mutant	Indel % ON Target (EMX1)	OFF Target 1(OT25)	OFF Target 2(OT46)
WT	**24.8**	**10.5**	**8.8**
R780	22.9	0.0	0.1
K810	23.3	0.1	0.1
K848	24.3	0.1	0.1
K855	25.1	0.2	0.3
R976	15.6	0.1	0.1
H982	20.9	0.5	0.4
K1003	24.6	4.1	2.8
R1060	20.4	1.3	1.8
GFP	0.1	0.0	0.1
untrans.	0.1	0.0	0.1



ALA mutations that improve specificity are distributed along the non-targeting strand groove.

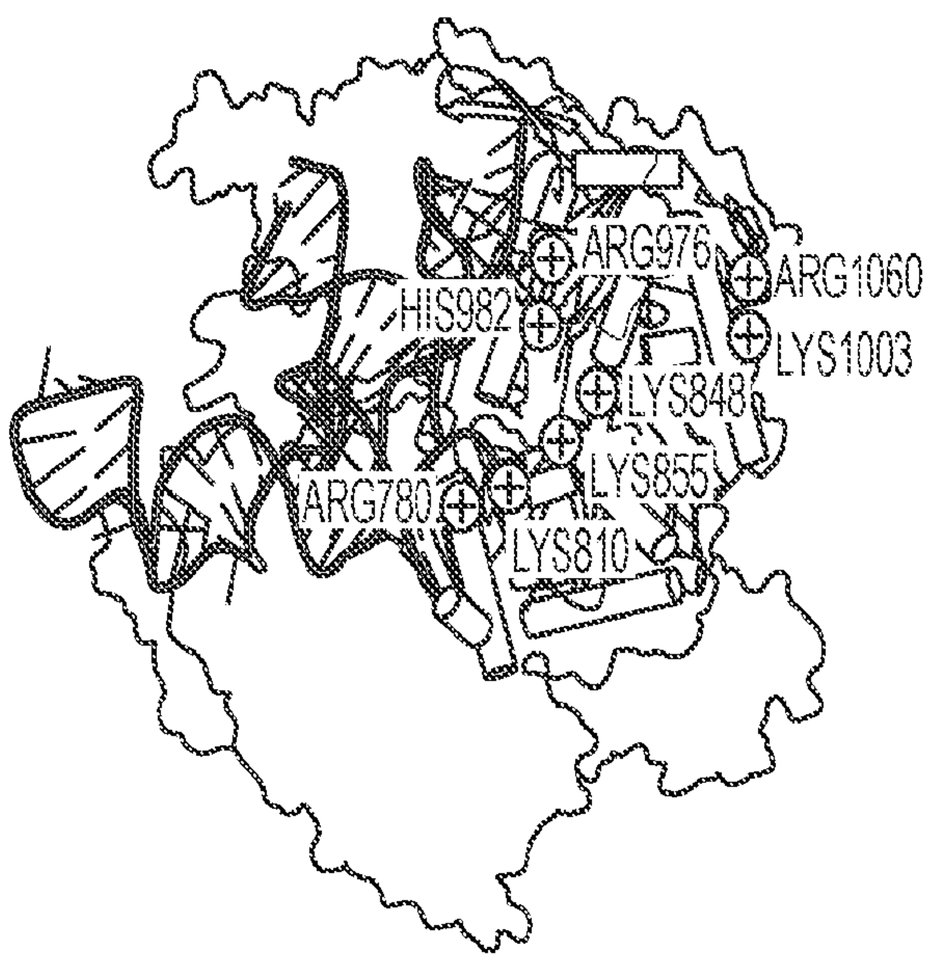


Fig. 1A

02989830 2017-12-15
CA

Mechanism proposal:
A) Nuclease activity is inactive until non-targeted DNA strand sterically triggers **HNH conformational change**
B) Non-targeted strand binding to the **groove** between HNH and RuvC depends on RNA:DNA pairing
C) Mutating DNA binding residues in the **groove** place more energetic demand on proper RNA:DNA pairing

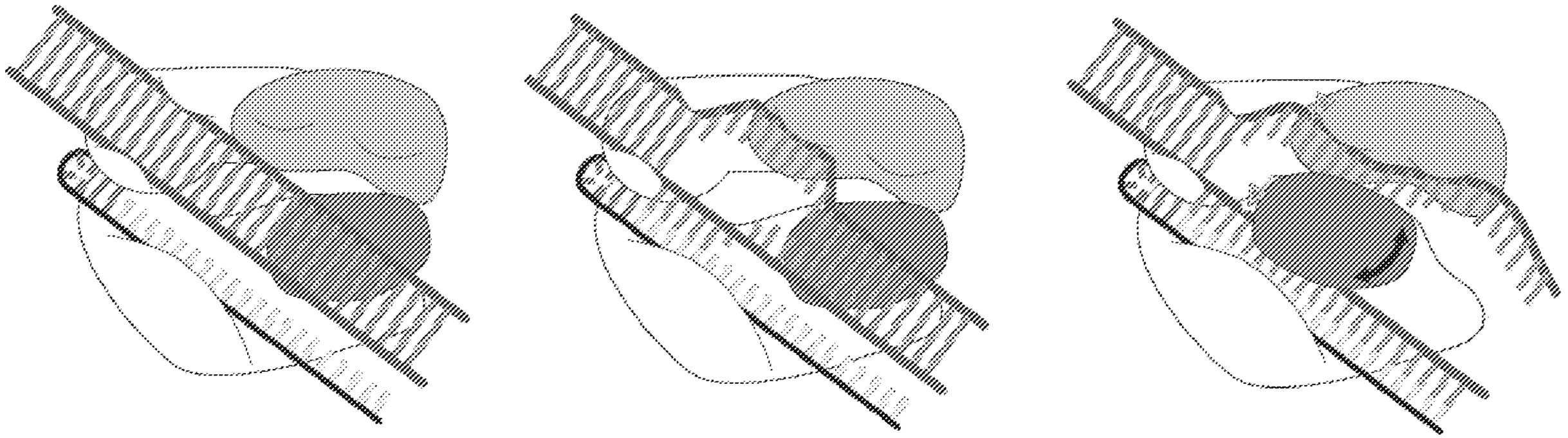


Fig. 1B

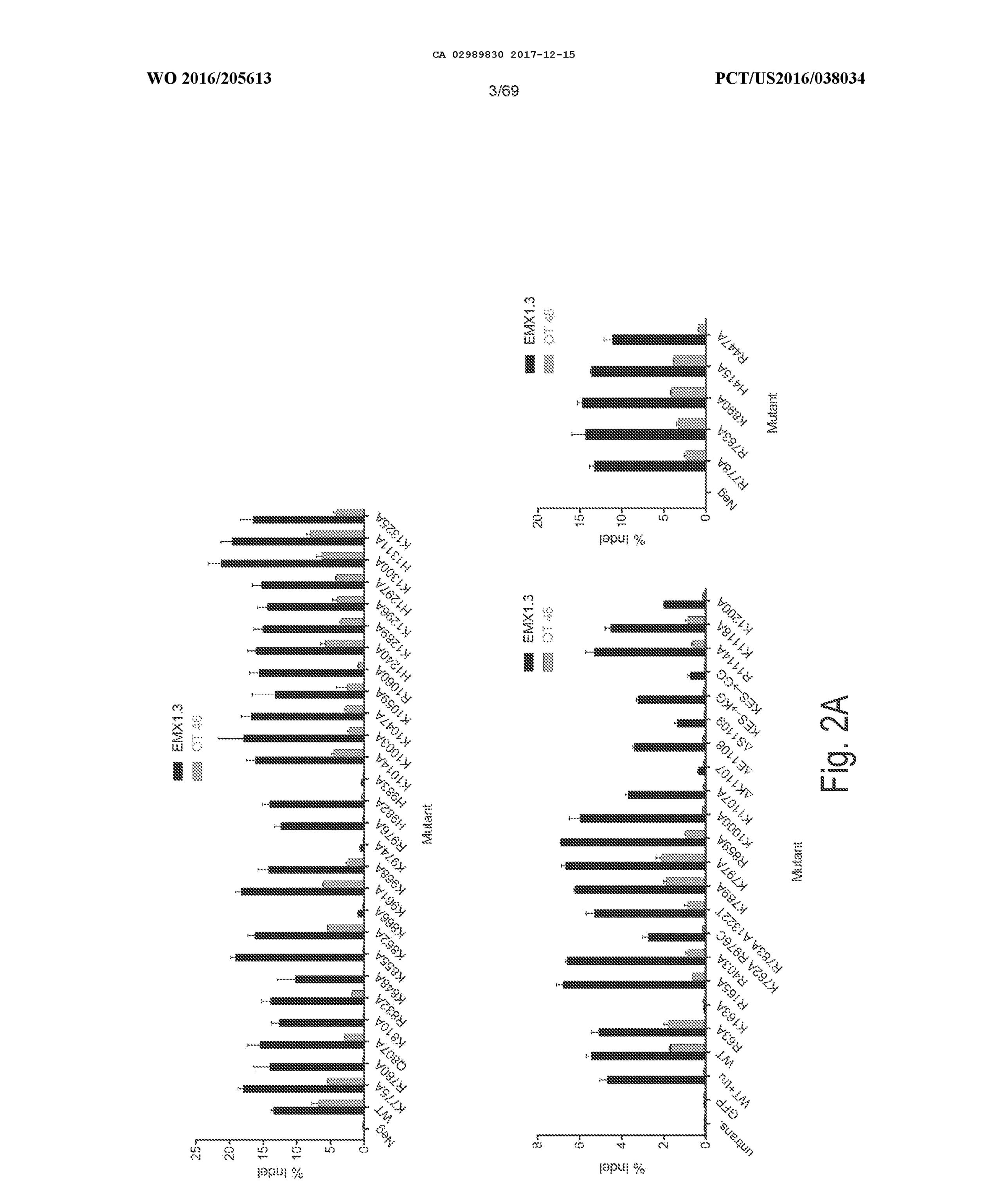
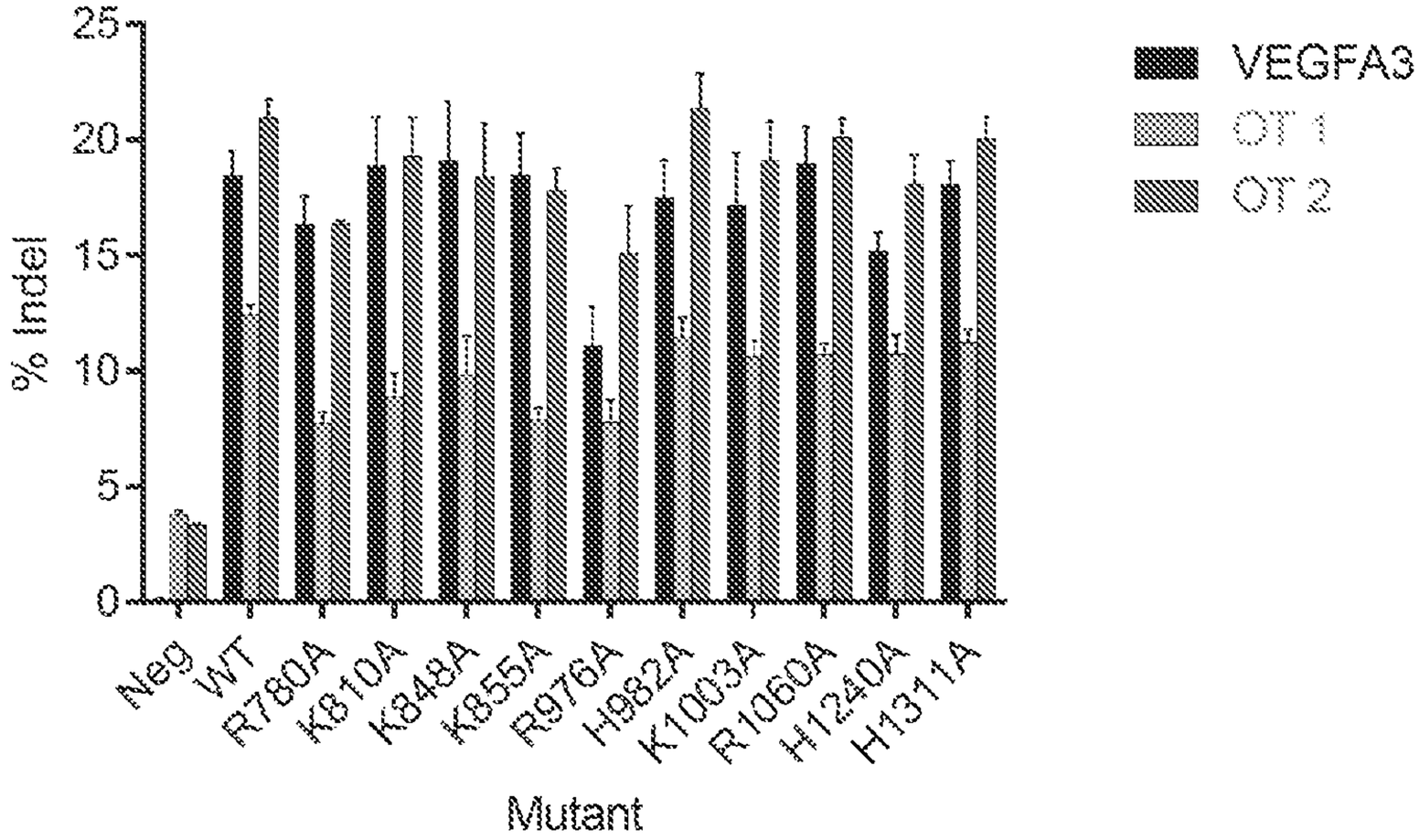

EMX1.3
OT #9
% Indel
Neg
R177A
R776A
K855A
H847A
R447A
Mutant
K1132A
H1300A
K1297A
H1296A
K1286A
H1260A
K1060A
F1059A
K1053A
K1037A
K1034A
K1014A
H982A
R976A
K974A
K958A
K961A
K866A
K862A
K855A
K836A
R832A
R810A
Q807A
K776A
K775A
WT
% Indel
EMX1.3
OT #9
Neg
R177A
R776A
K855A
H847A
R447A
Mutant
K1132A
H1300A
K1297A
H1296A
K1286A
H1260A
K1060A
F1059A
K1053A
K1037A
K1034A
K1014A
H982A
R976A
K974A
K958A
K961A
K866A
K862A
K855A
K836A
R832A
R810A
Q807A
K776A
K775A
WT
% Indel
EMX1.3
OT #9
GFP
WT+Lig
WT
R653A
K1633A
R1655A
R2032A
R976A
K776A
K1132A
H1300A
K1297A
H1296A
K1286A
H1260A
K1060A
F1059A
K1053A
K1037A
K1034A
K1014A
H982A
R976A
K974A
K958A
K961A
K866A
K862A
K855A
K836A
R832A
R810A
Q807A
K776A
K775A
WT
% Indel

Fig. 2A


% Indel
0
5
10
15
20
25
VEGFA3
OT 1
OT 2
Neg
WT
R780A
K810A
K848A
K855A
R976A
H982A
K1003A
R1060A
H1240A
H1311A
Mutant

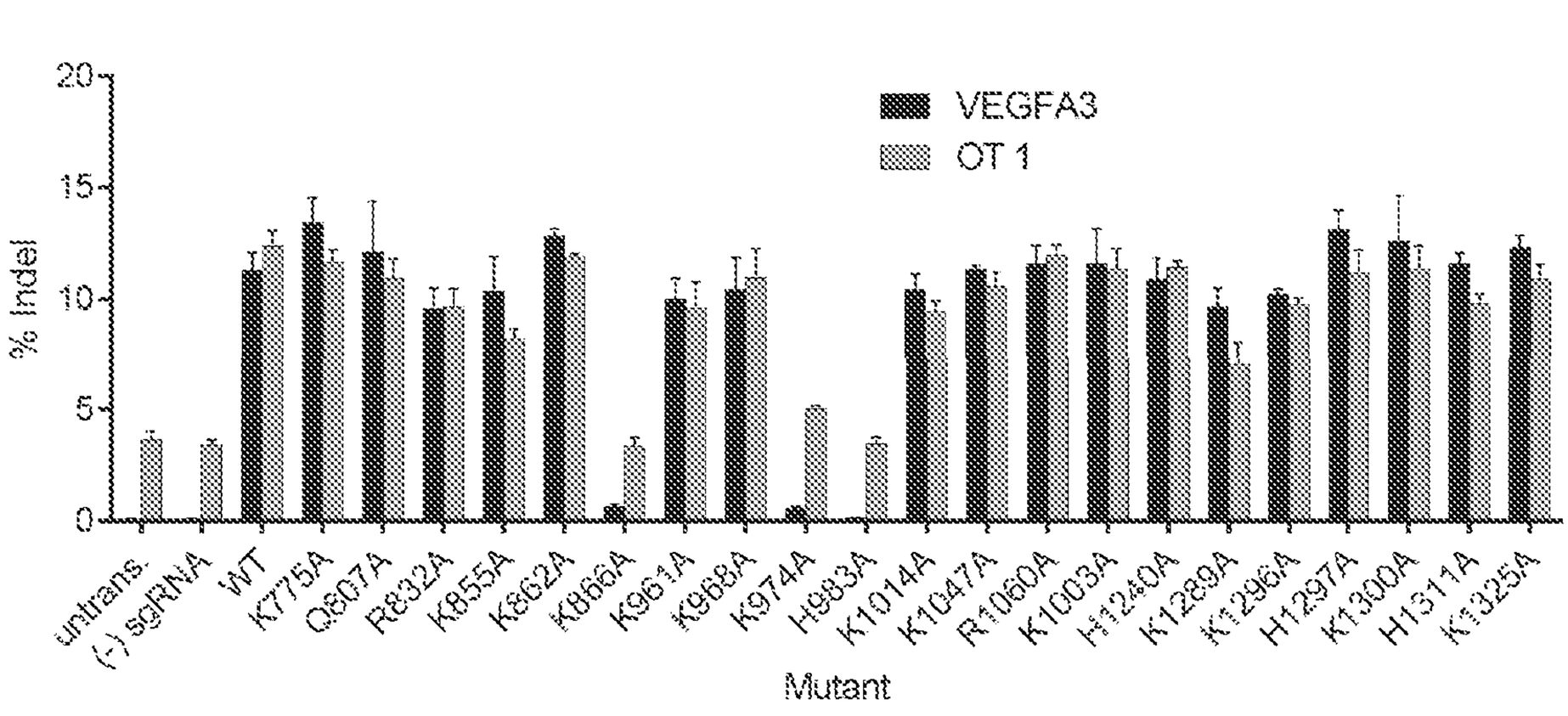

% Indel
0
5
10
15
20
VEGFA3
OT 1
untrans.
(-) sgRNA
WT
K775A
Q807A
R832A
K855A
K862A
K866A
K961A
K968A
K974A
H983A
K1014A
K1047A
R1060A
K1003A
H1240A
K1289A
K1296A
H1297A
K1300A
H1311A
K1325A
Mutant

Fig. 2B

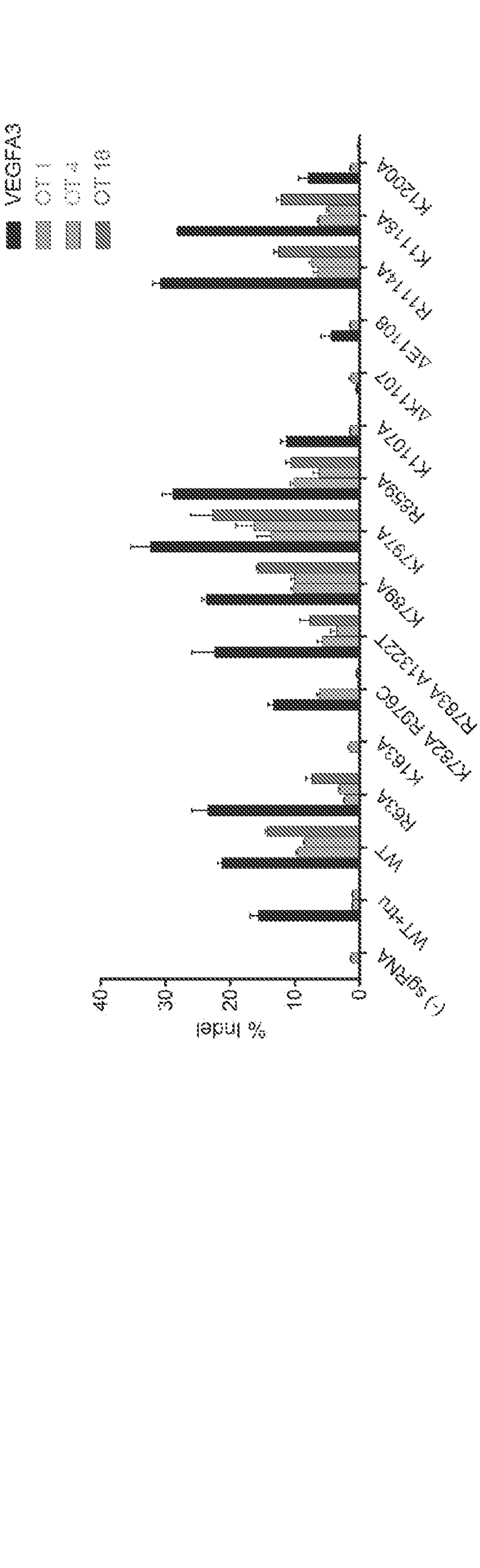

VEGFA3
OT-1
OT-4
OT-13
% indel
0
10
20
30
40
(-) sgRNA
WT+tru
WT
R165A
R403A
K100QA
Δ51-109
KES→KG
KES→QG
Mutant
VEGFA3
OT-1
OT-4
OT-13
% indel
0
5
10
15
20
25
(-) sgRNA
WT
R778A
R783A
K890A
N415A
R447A
Mutant
VEGFA3
OT-1
OT-4
OT-13
% indel
0
10
20
30
40
(-) sgRNA
WT+tru
WT
R63A
K163A
R780A A132T
K782A R978C
R783A A132T
K789A
K797A
R859A
K1107A
ΔK1107
ΔE1108
R1114A
K1118A
K1200A
Mutant

Fig. 2C

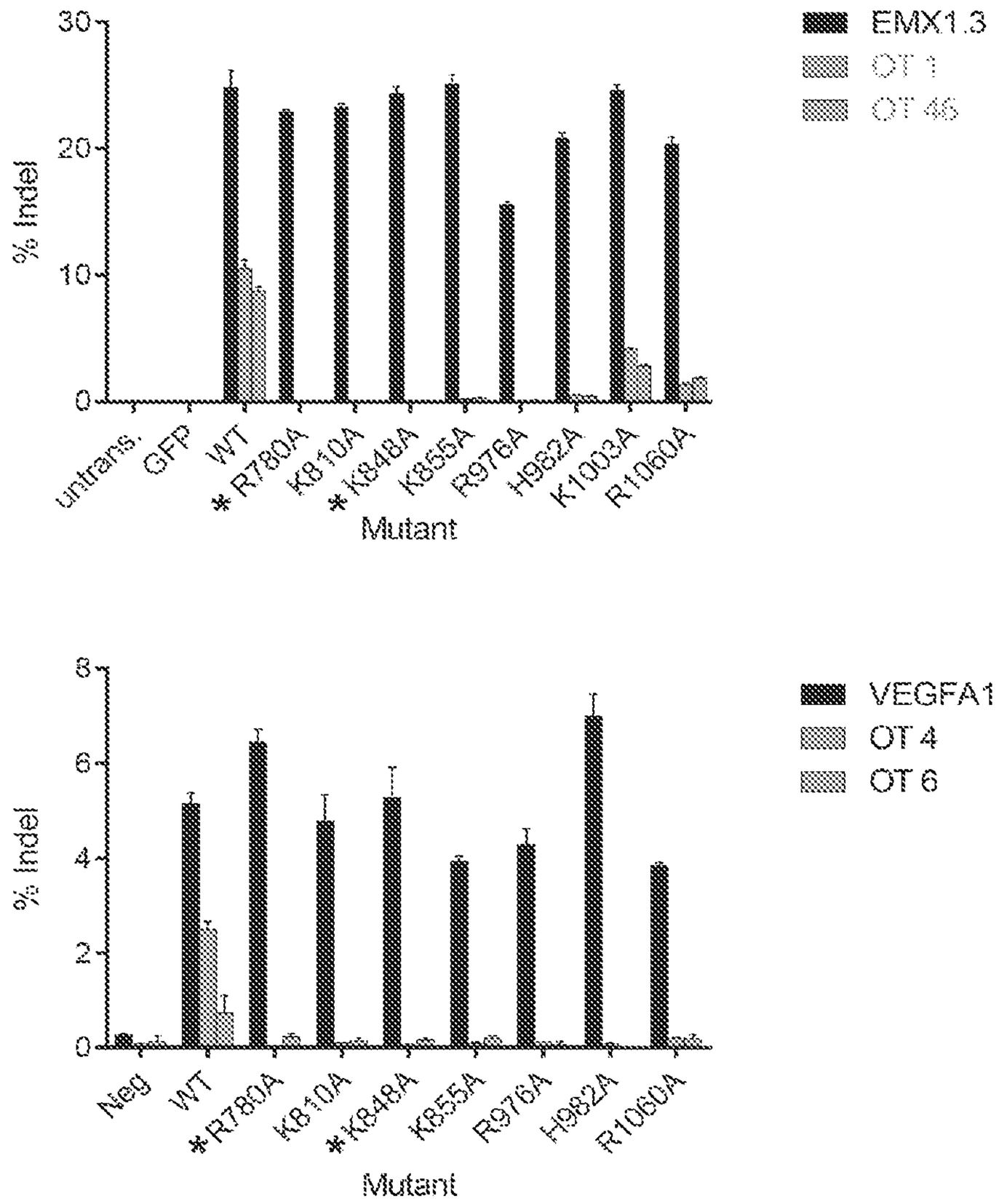
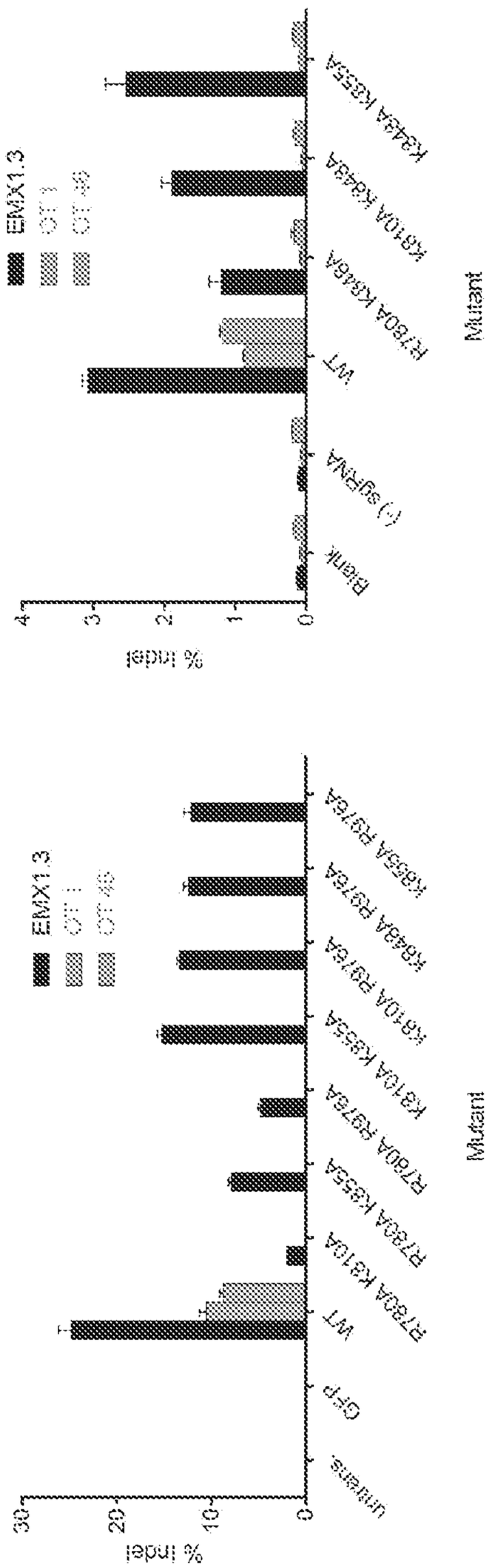
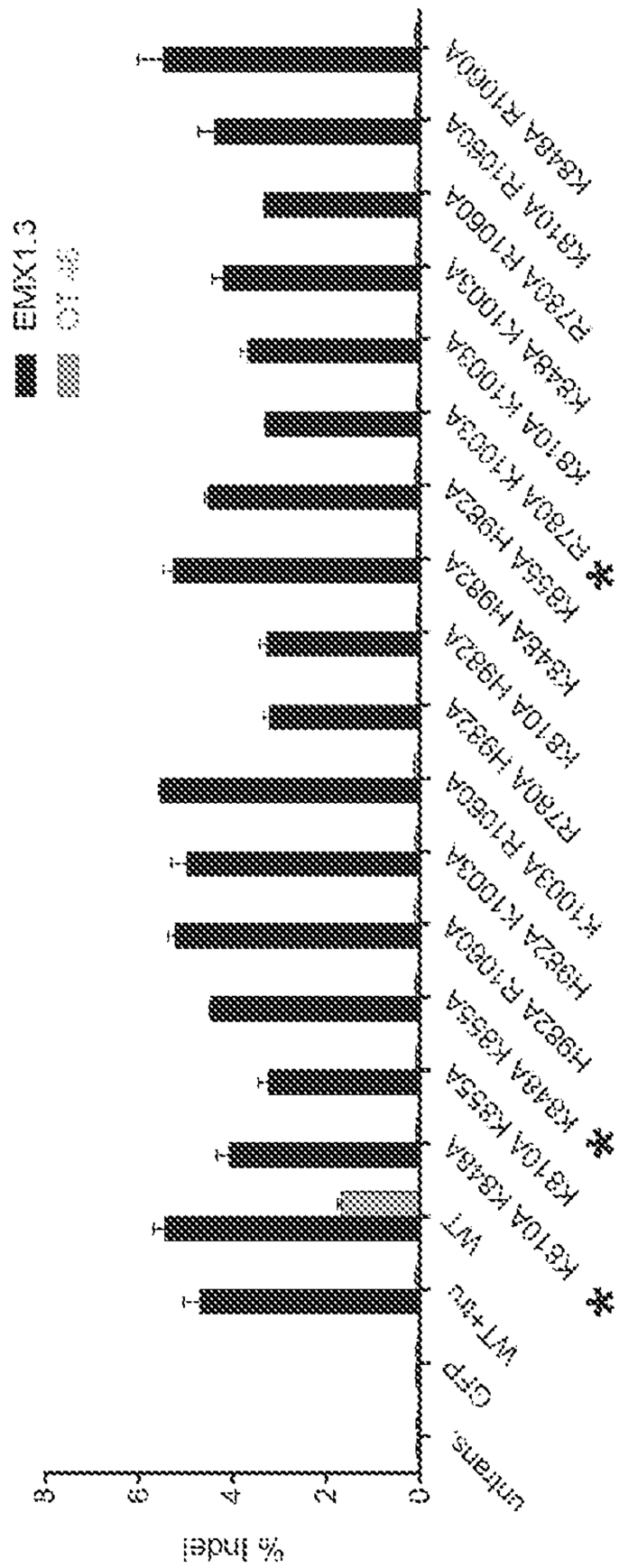

% Indel
30
20
10
0
untrans.
GFP
WT
*R780A
K810A
*K848A
K855A
R976A
H982A
K1003A
R1060A
Mutant
EMX1.3
OT 1
OT 46
% Indel
8
6
4
2
0
Neg
WT
*R780A
K810A
*K848A
K855A
R976A
H982A
R1060A
Mutant
VEGFA1
OT 4
OT 6
% Indel
40
30
20
10
0
Neg
WT
*R780A
K810A
*K848A
K855A
R976A
H982A
Mutant
VEGFA3
OT 1
OT 2
OT 4
OT 17
OT 18

Fig. 3


EMX1.3
OT 1
OT #5
% indel
0
1
2
3
4
Blank
(-) sgRNA
WT
R780A K848A
K810A K848A
K848A K855A
Mutant


EMX1.3
OT #5
% indel
0
2
4
6
8
untrans.
GFP
WT+tru
WT
K810A K848A
K810A K855A
K848A R1060A
K810A R1060A
R780A R1060A
K848A K1003A
R780A K1003A
K810A K1003A
K848A H982A
K855A H982A
R780A H982A
K848A N992A
K810A N992A
R780A R1060A
K1003A R1060A
H982A R1060A
K1003A K1060A
H982A K855A
R780A K855A
K810A K855A
R780A K810A
Mutant

EMX1.3
OT 1
OT #5
% indel
0
10
20
30
untrans.
GFP
WT
R780A K810A
R780A K855A
K810A K855A
R780A R976A
K848A R976A
K810A R976A
K855A R976A
Mutant

Fig. 4A

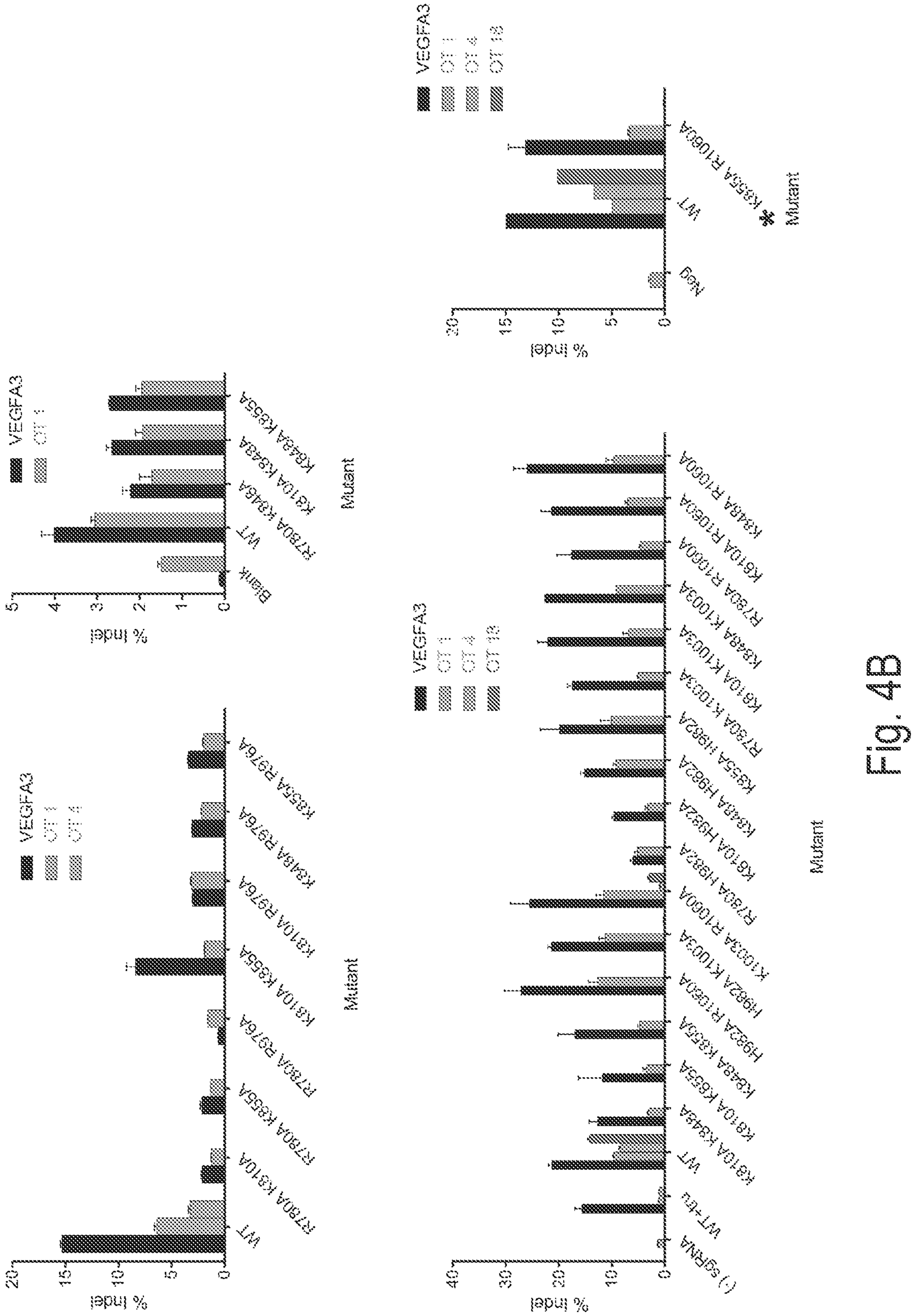

VEGFA3
OT 1
OT 2
OT 12
% indel
Neg
WT
K855A R1060A
Mutant
Blank
WT
R780A K846A
K810A K842A
K842A K855A
R780A K810A
R780A K855A
K842A P976A
K810A P976A
K855A P976A
R780A P976A
K842A H982A
K810A H982A
R780A H982A
K1003A R1060A
H982A K1003A
K842A R1060A
K810A K1003A
R780A K1003A
K842A K1003A
K810A K842A
R780A K842A
K810A K855A
R780A K855A
K842A K855A
K810A R1060A
R780A R1060A
(-) sgRNA
WT+dTu
WT
K810A K842A
K810A K855A
K842A R1060A
H982A R1060A
K1003A R1060A
R780A K1003A
K842A K1003A
K810A K1003A
R780A H982A
K855A H982A
K842A H982A
K810A H982A
R780A K810A
R780A K855A
K842A K855A
K810A R1060A
R780A R1060A

Fig. 4B

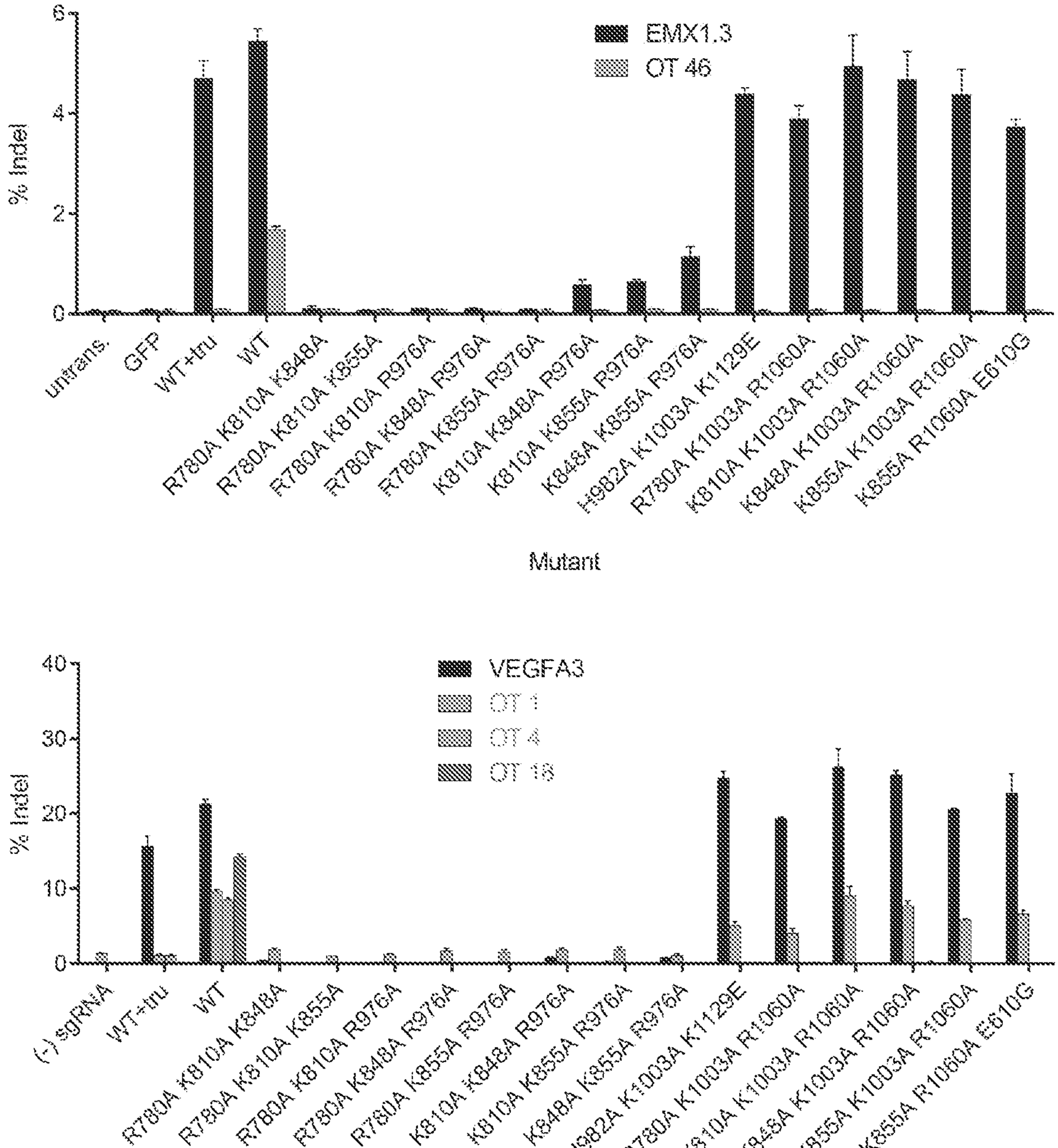
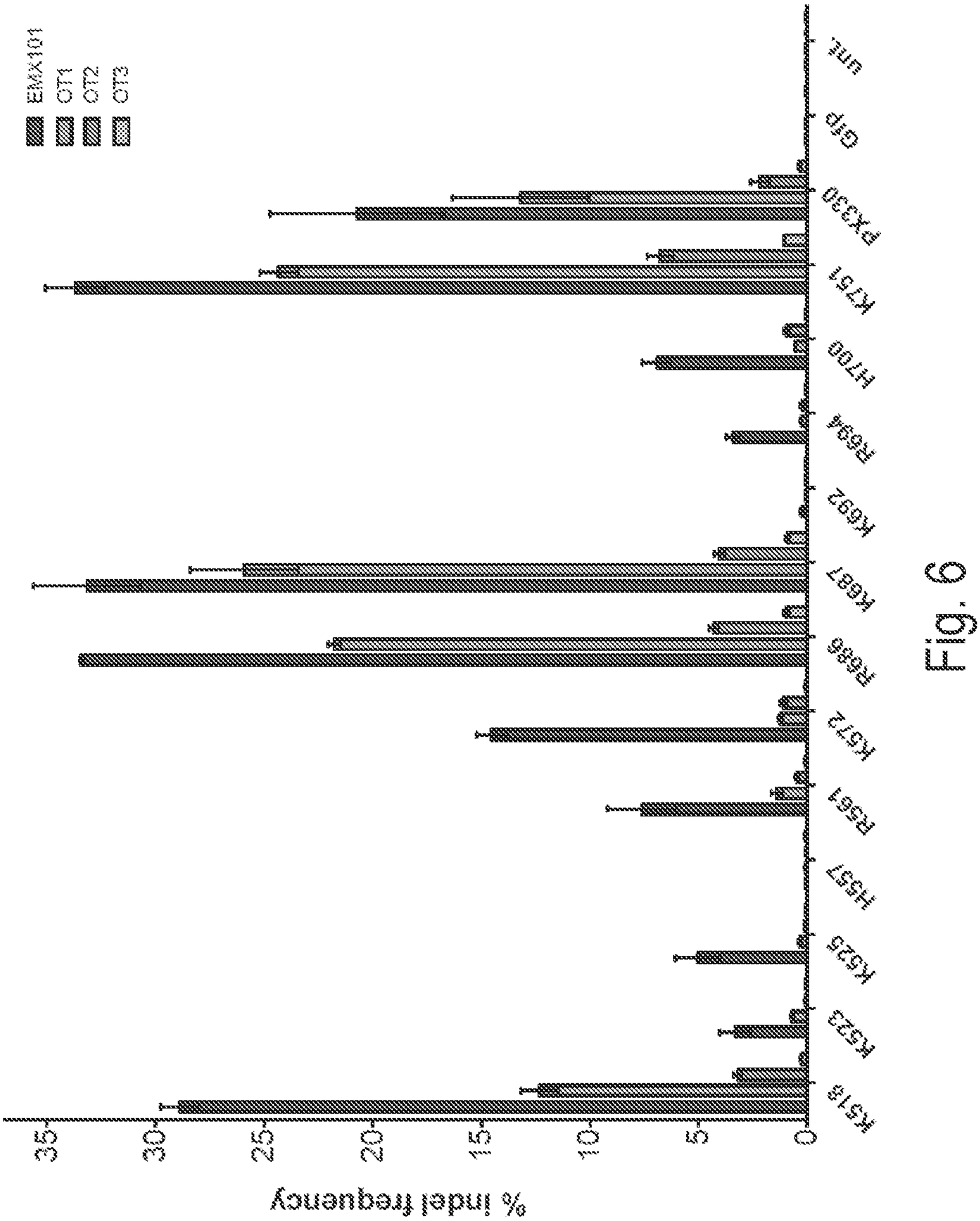

EMX1.3
OT 46
% Indel
0
2
4
6
untrans.
GFP
WT+tru
WT
R780A K810A K848A
R780A K810A K855A
R780A K810A R976A
R780A K848A R976A
R780A K855A R976A
K810A K848A R976A
K810A K855A R976A
K848A K855A R976A
H982A K1003A K1129E
R780A K1003A R1060A
K810A K1003A R1060A
K848A K1003A R1060A
K855A K1003A R1060A
K855A R1060A E610G
Mutant
VEGFA3
OT 1
OT 4
OT 16
% Indel
0
10
20
30
40
(-) sgRNA
WT+tru
WT
R780A K810A K848A
R780A K810A K855A
R780A K810A R976A
R780A K848A R976A
R780A K855A R976A
K810A K848A R976A
K810A K855A R976A
K848A K855A R976A
H982A K1003A K1129E
R780A K1003A R1060A
K810A K1003A R1060A
K848A K1003A R1060A
K855A K1003A R1060A
K855A R1060A E610G
Mutant

Fig. 5


EMX101
OT1
OT2
OT3
% indel frequency
5
0
5
0
5
0
HEK site 1
HEK site 2
HEK site 3
HEK site 4
VEGFA site 1
VEGFA site 2
VEGFA site 3
EMX1 site 1
EMX1 site 2
EMX1 site 3
EMX1 site 4
VEGFA site 4
VEGFA site 5
VEGFA site 6
VEGFA site 7
VEGFA site 8
VEGFA site 9
VEGFA site 10
VEGFA site 11
VEGFA site 12
VEGFA site 13
VEGFA site 14
VEGFA site 15
VEGFA site 16
VEGFA site

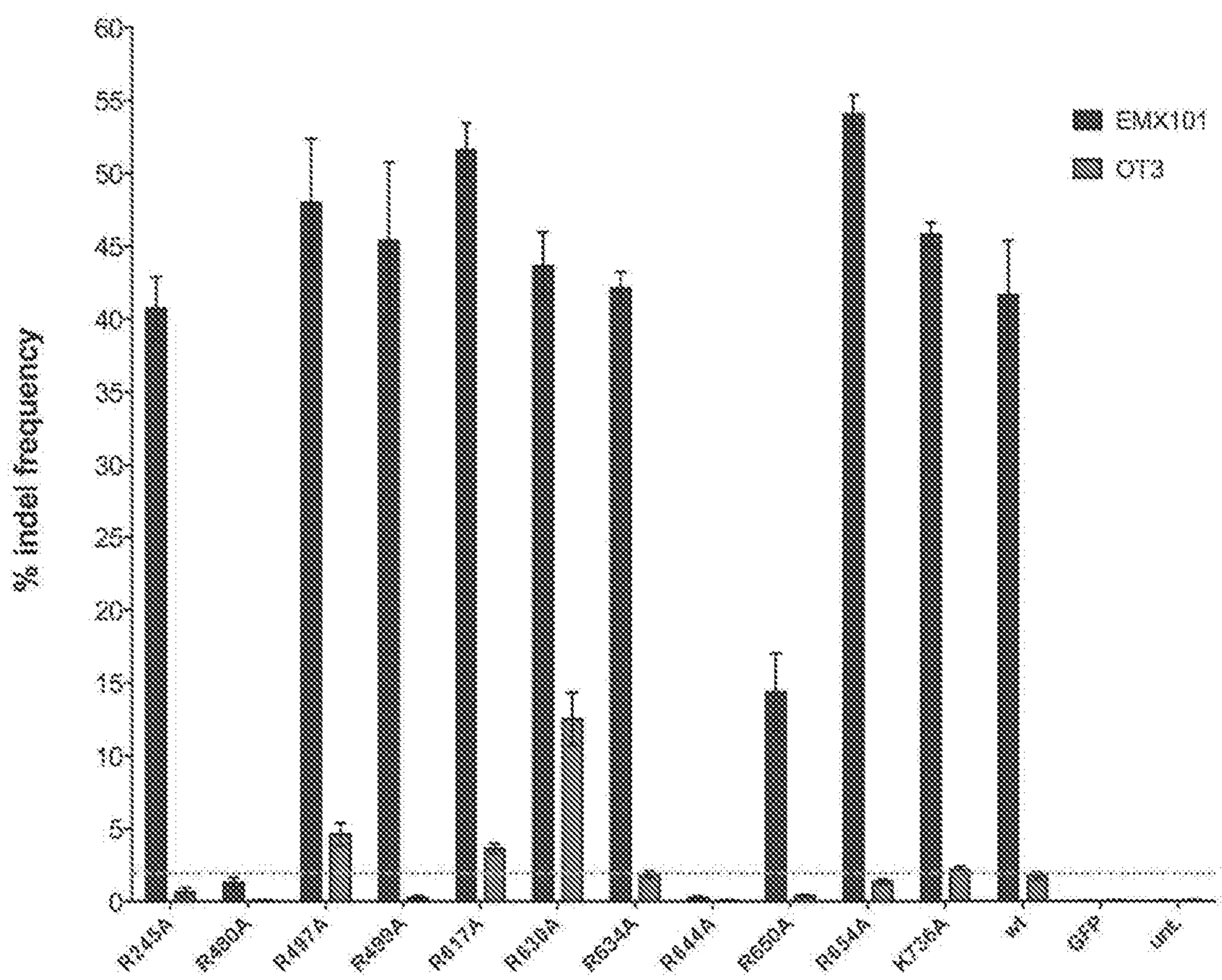
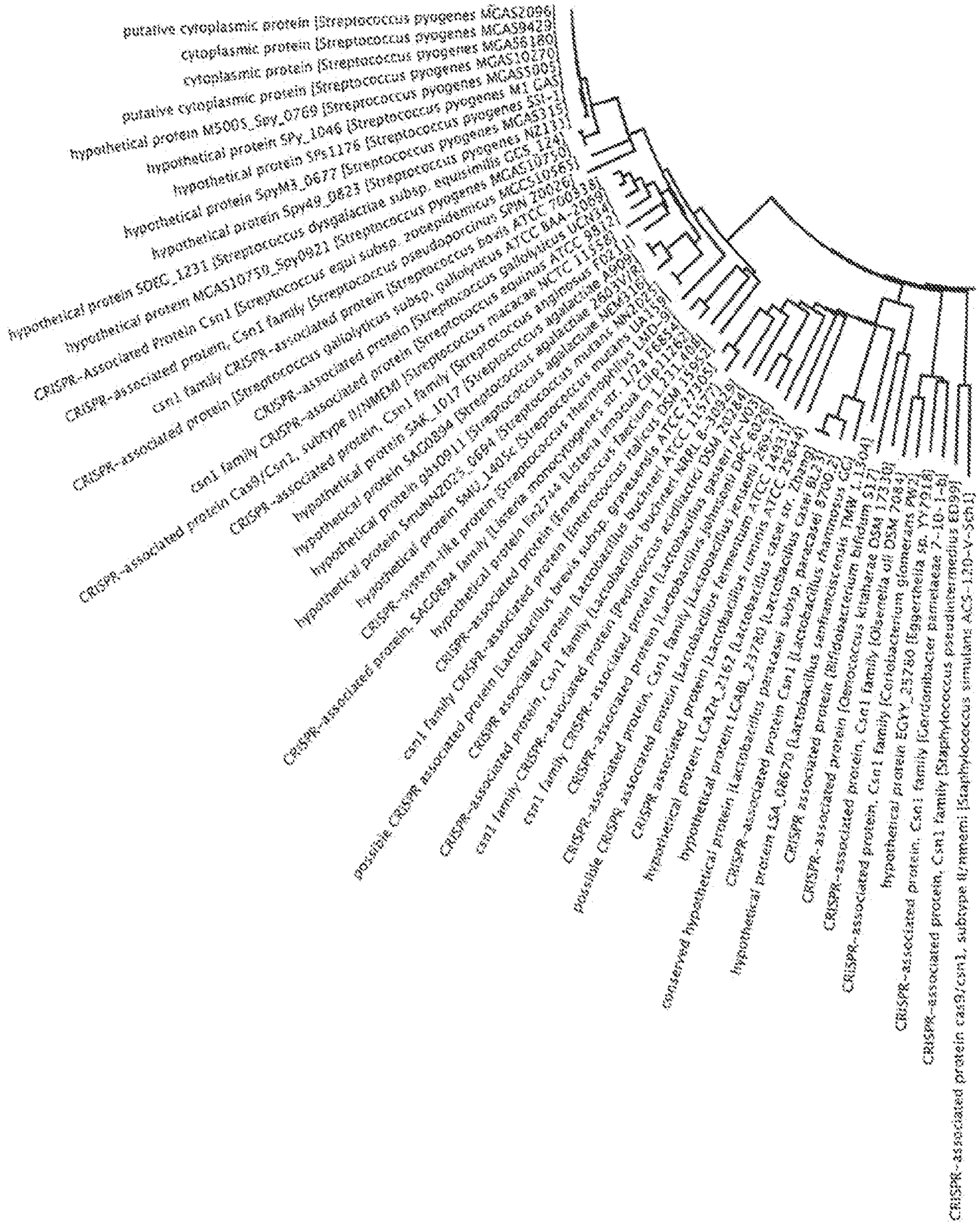
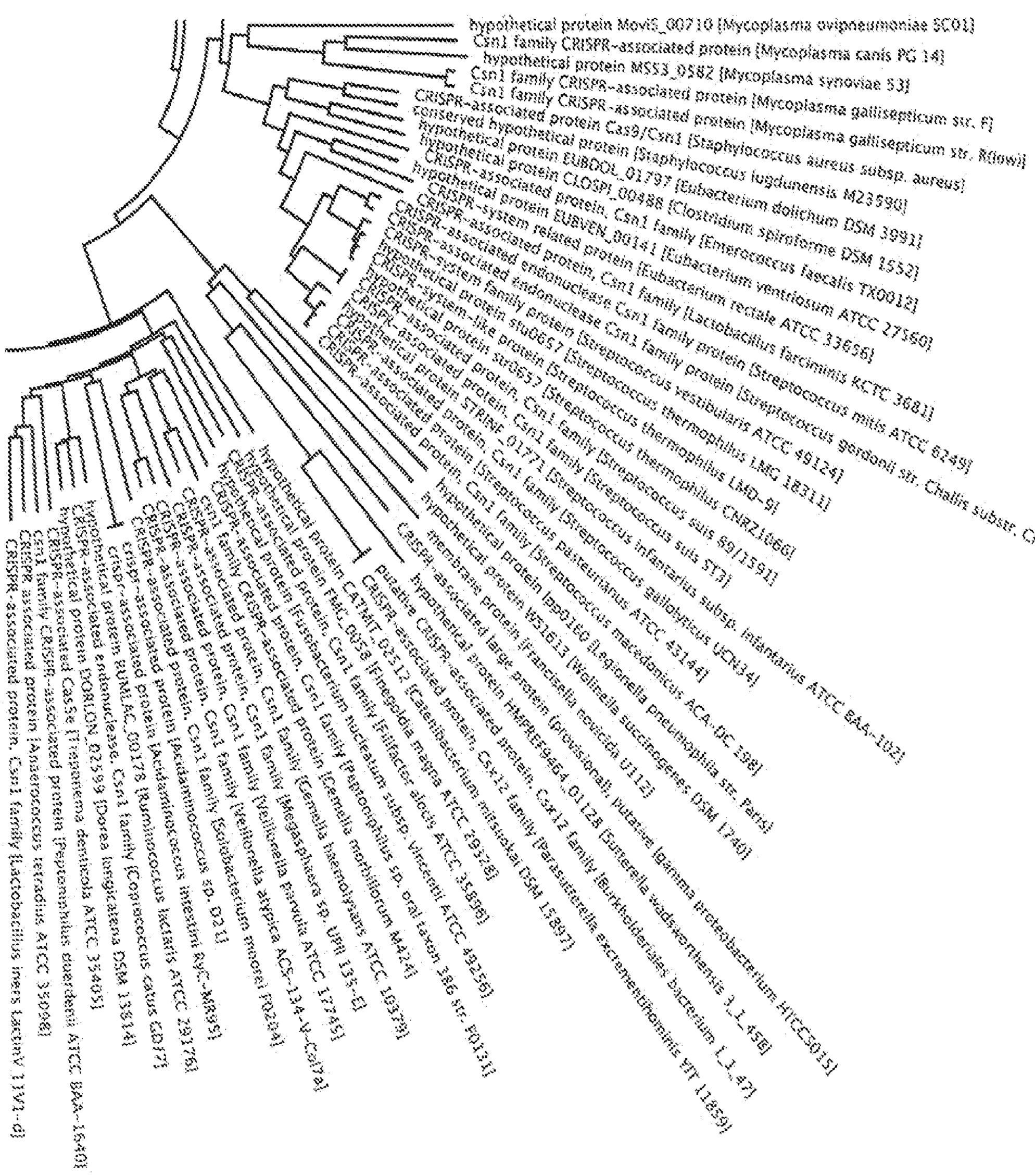

% indel frequency
0
5
10
15
20
25
30
35
40
45
50
55
60
EMX101
OT3
R445A
R480A
R447A
R469A
R457A
R626A
R634A
R444A
R650A
R454A
K775A
wt
GFP
unt.

Fig. 7


CRISPR-associated protein, Csn1 family [Candidatus Puniceispirillum marinum IMCC1322]
CRISPR-associated endonuclease Csn1 family protein [Lutimicrobium minutum Pei191]
CRISPR-associated protein [uncultured Termite group 1 bacterium phylotype Rs-D17]
Csn1-like CRISPR-associated protein [Sphaerochaeta globus str. Buddy]
CRISPR-associated protein. Csn1 family [Spirosoma linguale DSM 74]
CRISPR-associated protein [Parvibaculum lavamentivorans DS-1]
CRISPR-associated endonuclease Csn1 family protein [Rhodospirillum rubrum ATCC 11170]
conserved hypothetical protein [Bacteroides dorei DSM 17855]
CRISPR-associated endonuclease Csn1 family [Bacteroides coprocola DSM 17136]
hypothetical protein BACCOPRO_03113 [Bacteroides coprophilus DSM 18228]
hypothetical protein BACCOPRO_03869 [Bacteroides coprophilus DSM 18228]
conserved hypothetical protein [Bacteroides sp. 2_2]
CRISPR-associated protein, Csn1 family [Bacteroides sp. D1]
CRISPR-associated protein [Sphingobacterium spiritivorum ATCC 33861]
CRISPR-associated protein, Csn1 family [Bacteroides sp. 3_1_40A]
CRISPR-associated protein, Csn1 family [Flavobacterium columnare ATCC 49512]
CRISPR-associated protein, Csn1 family [Capnocytophaga ochracea DSM 7271]
Csn1 family CRISPR-associated protein [Capnocytophaga sputigena Capno]
CRISPR-associated protein, Csn1 family [Flavobacterium psychrophilum JIP02/86]
CRISPR-associated protein, Csn1 family [Prevotella buccalis ATCC 35310]
CRISPR-associated protein, Csn1 family [Prevotella ruminicola 23]
CRISPR-associated protein, Csn1 family [Prevotella tannerae ATCC 51259]
CRISPR-associated protein, Csn1 family [Prevotella micans F0438]
hypothetical protein PRABACTJOHN_00521 [Parabacteroides johnsonii DSM 18315]
hypothetical protein BF3354 [Bacteroides fragilis NCTC 9343]
CRISPR-associated protein, Csn1 family [Prevotella oralis ATCC 33269]
CRISPR-associated protein, Csn1 family [Prevotella timonensis CRIS 5C-B1]
CRISPR-associated protein, Csn1 family [Prevotella oralis ATCC 33317]
Csn1 family CRISPR-associated protein [Prevotella multisaccharivorax DSM 17128]
hypothetical protein HMPREF9140_01667 [Prevotella oralis F0302]
conserved hypothetical protein [Prevotella multiformis DSM 16608]
hypothetical protein DIP0036 [Corynebacterium diphtheriae NCTC 13129]
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 8d1]
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]
putative CRISPR-associated (Cas) protein [Flavobacterium columnare ATCC 49512]
hypothetical protein BACCELL_00696 [Bacteroides cellulosilyticus DSM 14838]
CRISPR-associated protein ZPR_4422 [Zunongwangia profunda SM-A87]
CRISPR-associated protein KAOT1_07362 [Kordia algicida OT-1]
CRISPR-associated endonuclease Csn1 family protein [Flavobacterium psychrophilum JIP02/86]
CRISPR-associated endonuclease Csn1 family protein [Capnocytophaga gingivalis ATCC 33624]
CRISPR-associated endonuclease Csn1 family protein [Capnocytophaga canimorsus Cc5]
CRISPR-associated endonuclease Csn1 family protein [Weeksella virosa DSM 16922]
CRISPR-associated endonuclease Csn1 family protein [Riemerella anatipestifer RA-YM]
CRISPR-associated endonuclease Csn1 family protein [Elizabethkingia anophelis NUHP1]
CRISPR-associated endonuclease Csn1 family protein [Elizabethkingia meningoseptica ATCC 13253]
CRISPR-associated endonuclease Csn1 family protein [Elizabethkingia meningoseptica PO19]
CRISPR-associated endonuclease Csn1 family protein [Elizabethkingia meningoseptica ATCC 33958]
CRISPR-associated endonuclease Csn1 family protein [Elizabethkingia meningoseptica G0181]
CRISPR-associated endonuclease Csn1 family protein [Elizabethkingia meningoseptica G0187]
CRISPR-associated endonuclease Csn1 family protein [Elizabethkingia meningoseptica G0185]
CRISPR-associated endonuclease Csn1 family protein [Elizabethkingia meningoseptica G0183]
CRISPR-associated endonuclease Csn1 family protein [Elizabethkingia meningoseptica G0182]
CRISPR-associated endonuclease Csn1 family protein [Elizabethkingia meningoseptica G0184]
CRISPR-associated endonuclease Csn1 family protein [Elizabethkingia meningoseptica G0186]
CRISPR-associated endonuclease Csn1 family protein [Elizabethkingia meningoseptica G0188]
CRISPR-associated endonuclease Csn1 family protein [Elizabethkingia meningoseptica G0189]
CRISPR-associated endonuclease Csn1 family protein [Elizabethkingia meningoseptica G0190]
CRISPR-associated endonuclease Csn1 family protein [Elizabethkingia meningoseptica G0191]
CRISPR-associated endonuclease Csn1 family protein [Elizabethkingia meningoseptica G0192]
CRISPR-associated endonuclease Csn1 family protein [Elizabethkingia meningoseptica


hypothetical protein Nham_4054 [Nitrobacter hamburgensis X14]
CRISPR-associated protein, Csn1 family [Azospirillum sp. B510]
hypothetical protein [Oligotropha carboxidovorans OM5]
conserved hypothetical protein B5S17_3838 [Sphingomonas sp. S17]
hypothetical protein [Rahnella aquatilis HX2]
CRISPR-associated protein HMPREF9436_34126 [gamma proteobacterium HdN1]
crispr-associated protein, Csn1 family [Nitrosomonas sp. AL212]
CRISPR-associated protein, Csn1 family [Acidovorax avenae subsp. avenae ATCC 19860]
hypothetical protein CBL_1423 [Candidatus Puniceispirillum marinum IMCC1322]
CRISPR-associated protein, Csn1 family [Verminephrobacter eiseniae EF01-2]
hypothetical protein RPC_4469 [Rhodospirillum centenum SW]
CRISPR-associated protein, Csn1 family [Methylobacterium sp. 4-46]
CRISPR-associated protein, Csn1 family [Methylobacterium extorquens DM4]
hypothetical protein CDL_2123 [Campylobacter curvus 525.92]
CRISPR-associated protein, Csn1 family [Helicobacter cinaedi CCUG 18818]
CRISPR-associated protein, Csn1 family [Campylobacter jejuni subsp. jejuni 81116]
hypothetical protein [Gluconacetobacter diazotrophicus PAL 5]
hypothetical protein [Rhodopseudomonas palustris BisB18]
hypothetical protein [Rhodopseudomonas palustris BisA53]
hypothetical protein [Rhodopseudomonas palustris BisB5]
hypothetical protein [Rhodopseudomonas palustris HaA2]
hypothetical protein [Rhodopseudomonas palustris BisA53]
hypothetical protein [Rhodopseudomonas palustris BisB18


putative cytoplasmic protein [Streptococcus pyogenes MGAS2096]
cytoplasmic protein [Streptococcus pyogenes MGAS9429]
cytoplasmic protein [Streptococcus pyogenes MGAS6180]
putative cytoplasmic protein [Streptococcus pyogenes MGAS10270]
hypothetical protein M5005_Spy_0769 [Streptococcus pyogenes MGAS5005]
hypothetical protein SPy_1046 [Streptococcus pyogenes M1 GAS]
hypothetical protein Spy1176 [Streptococcus pyogenes 5321]
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 700538]
CRISPR-associated protein [Streptococcus gallolyticus ATCC BAA-2069]
csn1 family CRISPR-associated protein [Streptococcus agalactiae GD201008-001]
CRISPR-associated protein [Streptococcus agalactiae NEM316]
hypothetical protein SAK_1017 [Streptococcus agalactiae 2603V/R]
hypothetical protein SAGO894 [Streptococcus agalactiae ATCC 27591]
hypothetical protein gbs0911 [Streptococcus agalactiae NEM316]
hypothetical protein SmuNN2025_0694 [Streptococcus mutans NN2025]
hypothetical protein SMU_1405c [Streptococcus mutans UA159]
CRISPR-system-like protein [Streptococcus mutans LJ23]
hypothetical protein SACOL2684 [Staphylococcus aureus subsp. aureus COL]
CRISPR-associated protein [Enterococcus faecium 1,231,408]
hypothetical protein [Enterococcus faecium DO]
CRISPR-associated protein [Enterococcus faecium 1,231,410]
hypothetical protein lin2744 [Listeria innocua Clip11262]
hypothetical protein [Streptococcus thermophilus LMD-9]
hypothetical protein [Streptococcus thermophilus str. 1,2a F6854]
CRISPR-associated protein [Enterococcus faecalis DS5]
hypothetical protein [Enterococcus faecalis ATCC 19433]
CRISPR-associated protein [Lactobacillus brevis subsp. gravesensis ATCC 27305]
hypothetical protein [Lactobacillus buchneri NRRL B-30929]
csn1 family [Lactobacillus buchneri ATCC 11577]
hypothetical protein [Lactobacillus actinobacteri DSM 20284]
hypothetical protein [Lactobacillus jensenii JV-V03]
csn1 family [Lactobacillus jensenii 269-3]
hypothetical protein [Lactobacillus gasseri DEC 6820]
hypothetical protein [Lactobacillus fermentum ATCC 14931]
hypothetical protein [Lactobacillus reuteri ATCC 23272]
CRISPR-associated protein [Lactobacillus casei str. Zhang]
hypothetical protein [Lactobacillus casei 8700:2]
CRISPR-associated protein [Lactobacillus paracasei subsp. paracasei 8700:2]
hypothetical protein [Lactobacillus rhamnosus GG]
CRISPR-associated protein [Lactobacillus rhamnosus LMS 2-1]
hypothetical protein [Bifidobacterium bifidum S17]
csn1 family [Bifidobacterium kitaharae DSM 17330]
hypothetical protein [Oscillibacter ruminantium GCA 000375]
csn1 family [Coriobacteriaceae bacterium 68-1-2]
hypothetical protein [Eggerthella sp. YY7918]
CRISPR-associated protein, Csn1 family [Eggerthella lenta DSM 2243]
CRISPR-associated protein, Csn1 family [Catenibacterium mitsuokai DSM 15897]
CRISPR-associated protein, Csn1 family [Streptococcus pseudoporcinus SPIN 20026


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]
Csn1 family CRISPR-associated protein [Mycoplasma gallisepticum str. R(low)]
CRISPR-associated protein Cas9/Csn1 [Staphylococcus lugdunensis M23590]
conserved hypothetical protein [Staphylococcus aureus subsp. aureus]
hypothetical protein EUBDOL_01797 [Eubacterium dolichum DSM 3991]
hypothetical protein CLOSPI_00488 [Clostridium spiroforme DSM 1552]
CRISPR-associated protein, Csn1 family [Enterococcus faecalis TX0012]
hypothetical protein EUBVEN_00141 [Eubacterium ventriosum ATCC 27560]
CRISPR-system related protein [Eubacterium rectale ATCC 33656]
CRISPR-associated protein, Csn1 family [Lactobacillus farciminis KCTC 3681]
CRISPR-associated endonuclease Csn1 family protein [Streptococcus mitis ATCC 6249]
hypothetical protein stu0657 [Streptococcus gordonii str. Challis substr. CH1]
CRISPR-system family protein [Streptococcus vestibularis ATCC 49124]
hypothetical protein stu0657 [Streptococcus thermophilus LMD-9]
CRISPR-system-like protein [Streptococcus thermophilus LMG 18311]
CRISPR-associated protein, Csn1 family [Streptococcus thermophilus CNRZ1066]
hypothetical protein STRINF_01771 [Streptococcus infantarius subsp. infantarius ATCC BAA-102]
CRISPR-associated protein, Csn1 family [Streptococcus suis 89/1591]
hypothetical protein, Csn1 family [Streptococcus suis ST3]
membrane protein [Streptococcus pasteurianus ATCC 43144]
CRISPR-associated protein, Csn1 family [Streptococcus gallolyticus UCN34]
hypothetical protein lpp0160 [Legionella pneumophila str. Paris]
hypothetical protein WS1613 [Wolinella succinogenes DSM 1740]
hypothetical protein [Francisella novicida U112]
hypothetical protein HMPREF9464_01128 [Sutterella wadsworthensis 3_1_45B]
CRISPR-associated large protein [Burkholderiales bacterium 1_1_47]
hypothetical protein HMPREF9464_01128 [Sutterella wadsworthensis 3_1_45B]
CRISPR-associated protein, Csx12 family [Parasutterella excrementihominis YIT 11859]
CRISPR-associated protein, Csx12 family [Parasutterella secunda]
hypothetical protein CATMIT_02512 [Catenibacterium mitsuokai DSM 15897]
CRISPR-associated protein [Finegoldia magna ATCC 29328]
hypothetical protein FAEFAC_02358 [Faecalibacterium prausnitzii ATCC 27768]
CRISPR-associated protein, Csx12 family [Peptoniphilus sp. oral taxon 386 str. F0131]
hypothetical protein [Peptoniphilus sp. oral taxon 836 str. F0429]
CRISPR-associated protein, Csn1 family [Peptoniphilus harei ATCC 29327]
hypothetical protein, Csn1 family [Peptoniphilus indolicus ATCC 29427]
CRISPR-associated protein, Csn1 family [Peptoniphilus lacrimalis 315-B]
hypothetical protein, Csn1 family [Gemella haemolysans ATCC 10379]
CRISPR-associated protein, Csn1 family [Gemella morbillorum M424]
hypothetical protein [Eubacterium nucleatum ATCC 49256]
CRISPR-associated protein, Csn1 family [Megasphaera sp. UPII 135-E]
hypothetical protein [Eubacterium saphenum ATCC 49989]
CRISPR-associated protein, Csn1 family [Veillonella parvula ATCC 17745]
hypothetical protein [Megasphaera micronuciformis F0428]
CRISPR-associated protein, Csn1 family [Veillonella atypica ACS-134-V-Col7a]
Csn1 family CRISPR-associated protein [Veillonella ratti]
CRISPR-associated protein, Csn1 family [Veillonella parvula DSM 2008]
CRISPR-associated protein, Csn1 family [Selenomonas sp. FOBRC9]
CRISPR-associated protein, Csn1 family [Selenomonas noxia ATCC 43541]
CRISPR-associated protein, Csn1 family [Selenomonas sputigena ATCC 35185]
hypothetical protein [Selenomonas sputigena ATCC 35185]
CRISPR-associated protein, Csn1 family [Selenomonas flueggei ATCC 43532]
CRISPR-associated protein [Selenomonas sp. FOBRC5]
CRISPR-associated protein, Csn1 family [Selenomonas sp. FOBRC6]
CRISPR-associated protein, Csn1 family [Selenomonas sp. FOBRC7]
CRISPR-associated protein, Csn1 family [Selenomonas sp. FOBRC8]
CRISPR-associated protein, Csn1 family [Selenomonas sp. FOBRC9]
CRISPR-associated protein [Selenomonas sp. oral taxon 149 str. 67H29BP]
CRISPR-associated protein, Csn1 family [Selenomonas sp. oral taxon 920 str. F0640]
CRISPR-associated protein, Csn1 family [Selenomonas sp. oral taxon 137 str. F0430]
CRISPR-associated protein, Csn1 family [Selenomonas sp. oral taxon 892]
CRISPR-associated protein, Csn1 family [Selenomonas sp. oral taxon 894

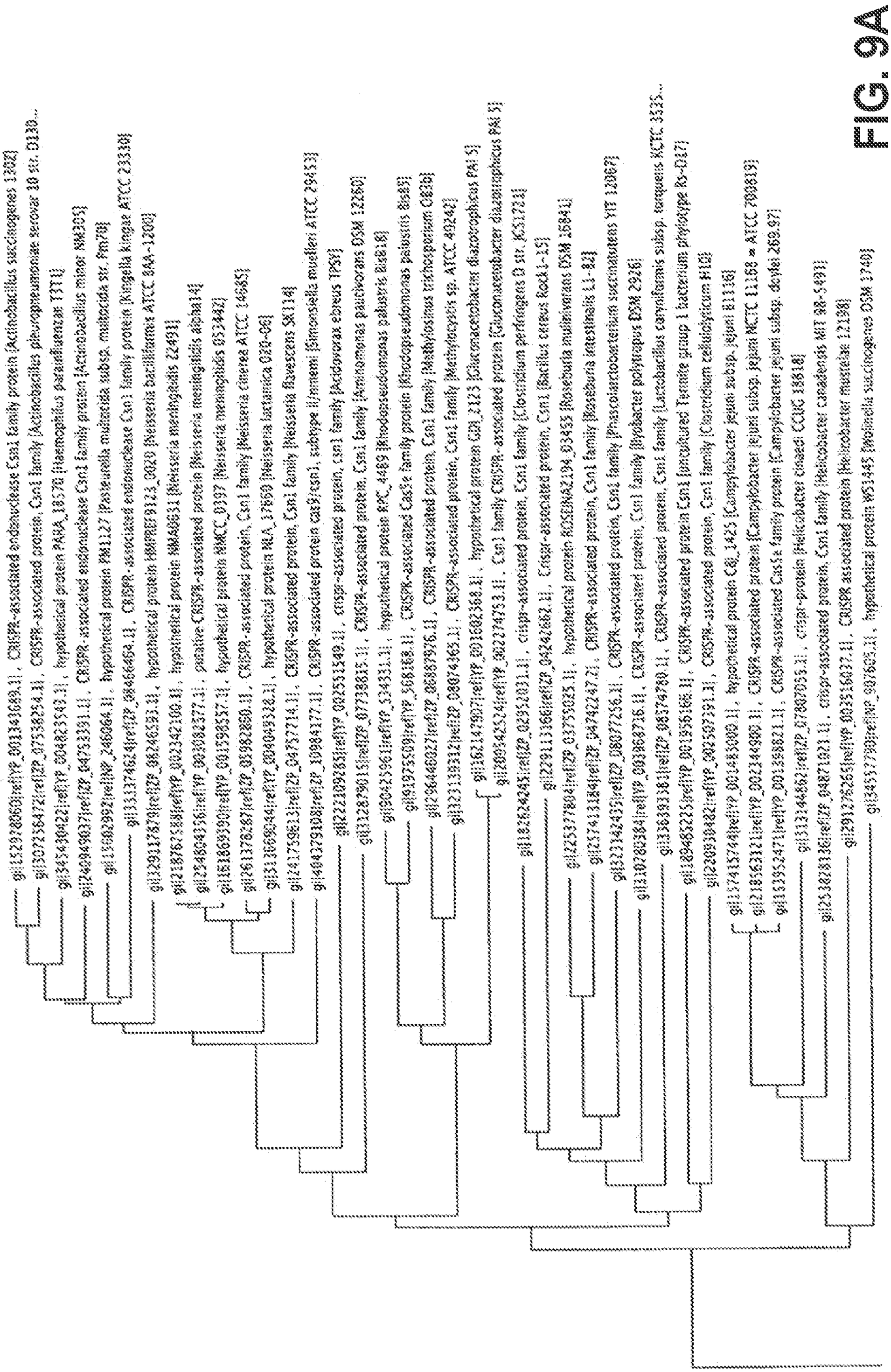

gi|152978865|ref|YP_001341669.1|, CRISPR-associated endonuclease Cas1 family protein [Actinobacillus succinogenes 13...
gi|302758472|ref|ZP_07538254.1|, CRISPR-associated protein, Cas1 family [Actinobacillus pleuropneumoniae serovar 10 str. 0330...
gi|345430422|ref|YP_004823543.1|, hypothetical protein PAM_1857G [Haemophilus parainfluenzae T3T1]
gi|240949037|ref|ZP_04753391.1|, CRISPR-associated endonuclease Cas1 family protein [Actinobacillus minor NM305]
gi|166462992|ref|NP_246064.1|, hypothetical protein PM1127 [Pasteurella multocida subsp. multocida str. Pm70]
gi|113374624|ref|ZP_08466462.1|, CRISPR-associated endonuclease Cas1 family protein [Kingella kingae ATCC 2...
gi|291178780|ref|ZP_06046231.1|, hypothetical protein HMPREF9123_0020 [Neisseria bacilliformis ATCC BAA-1200]
gi|218767588|ref|YP_002342100.1|, hypothetical protein NMA0631 [Neisseria meningitidis Z2491]
gi|218762488|ref|YP_002342100.1|, putative CRISPR-associated protein [Neisseria meningitidis alpha14]
gi|225460436|ref|YP_002882377.1|, hypothetical protein NMCC_0197 [Neisseria meningitidis 053442]
gi|261376267|ref|ZP_05982860.1|, CRISPR-associated protein, Cas1 family [Neisseria cinerea ATCC 14685]
gi|313360944|ref|ZP_07774664.1|, hypothetical protein HMPREF0179_00466 [Neisseria flavescens SK114]
gi|241759813|ref|ZP_04757714.1|, CRISPR-associated protein, Cas1 family [Neisseria flavescens SK114]
gi|448279108|ref|ZP_10984177.1|, CRISPR-associated protein cas1/csn1, subtype II/nmeni [Simonsiella muelleri ATCC 29...
gi|222109265|ref|YP_002531549.1|, crispr-associated protein, csn1 family [Acidovorax ebreus TPSY]
gi|222109265|ref|YP_002531615.1|, CRISPR-associated protein, Csn1 family [Aromonas pauloverans DSM 12260]
gi|222109265|ref|YP_002531331.1|, hypothetical protein AFC_4489 [Rhodopseudomonas palustris BisB18]
gi|222109265|ref|YP_002538168.1|, CRISPR-associated Cas5e family protein [Rhodopseudomonas palustris BisB5]
gi|222109265|ref|YP_002538168.1|, CRISPR-associated protein, Cas1 family [Methylosinus trichosporium OB3b]
gi|222109265|ref|YP_002537976.1|, CRISPR-associated protein, Csn1 family [Methylocystis sp. ATCC 49242]
gi|222109265|ref|YP_002537976.1|, CRISPR-associated protein, Csn1 family [Methylocystis sp. ATCC 49242]
gi|222109265|ref|YP_002534365.1|, Crispr-associated protein, Csn1 [Bacillus cereus Rock1-15]
gi|222109265|ref|YP_002534365.1|, hypothetical protein ROSEINA2194_03435 [Roseburia inulinivorans DSM 16841]
gi|222109265|ref|YP_002534365.1|, CRISPR-associated protein, Csn1 family [Roseburia intestinalis L1-82]
gi|222109265|ref|YP_002534365.1|, CRISPR-associated protein, Csn1 family [Phascolarctobacterium succinatutens YIT 12067]
gi|222109265|ref|YP_002534365.1|, CRISPR-associated protein, Csn1 family [Bryobacter polymorphus DSM 2926]
gi|222109265|ref|YP_002534360.1|, CRISPR-associated protein, Csn1 family [Lactobacillus carvophilus subsp. torquens KCTC 3325...
gi|222109265|ref|YP_002534361.1|, CRISPR-associated protein Csn1 [uncultured Termite group 1 bacterium phylotype Rs-D17]
gi|222109265|ref|YP_002534361.1|, CRISPR-associated protein, Csn1 family [Clostridium cellulolyticum H10]
gi|222109265|ref|YP_002534361.1|, crispr-associated protein Cmj_1425 [Campylobacter jejuni subsp. jejuni 81116]
gi|222109265|ref|YP_002534361.1|, CRISPR-associated protein [Campylobacter jejuni subsp. jejuni NCTC 11168 = ATCC 700819]
gi|222109265|ref|YP_002534361.1|, CRISPR-associated Cas5e family protein [Campylobacter jejuni subsp. doylei 269.97]
gi|222109265|ref|YP_002534361.1|, crispr-protein [Helicobacter cinaedi CCUG 18818]
gi|222109265|ref|YP_002534361.1|, CRISPR-associated protein, Csn1 family [Helicobacter canadensis MIT 98-5491]
gi|222109265|ref|YP_002534361.1|, CRISPR associated protein [Helicobacter mustelae 12198]
gi|222109265|ref|YP_002534361.1|, hypothetical protein WS1445 [Wolinella succinogenes DSM 1740]

FIG. 9A

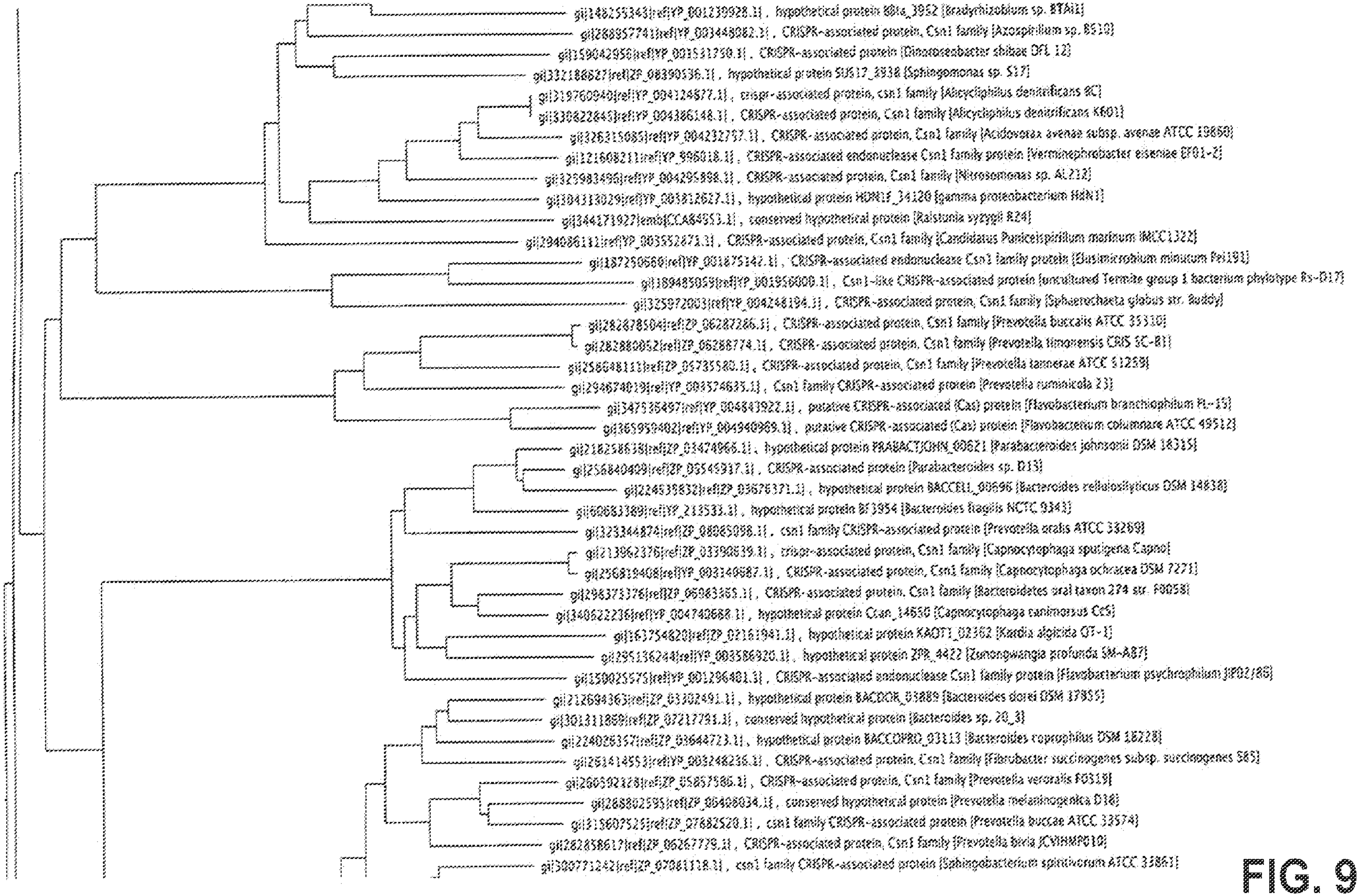
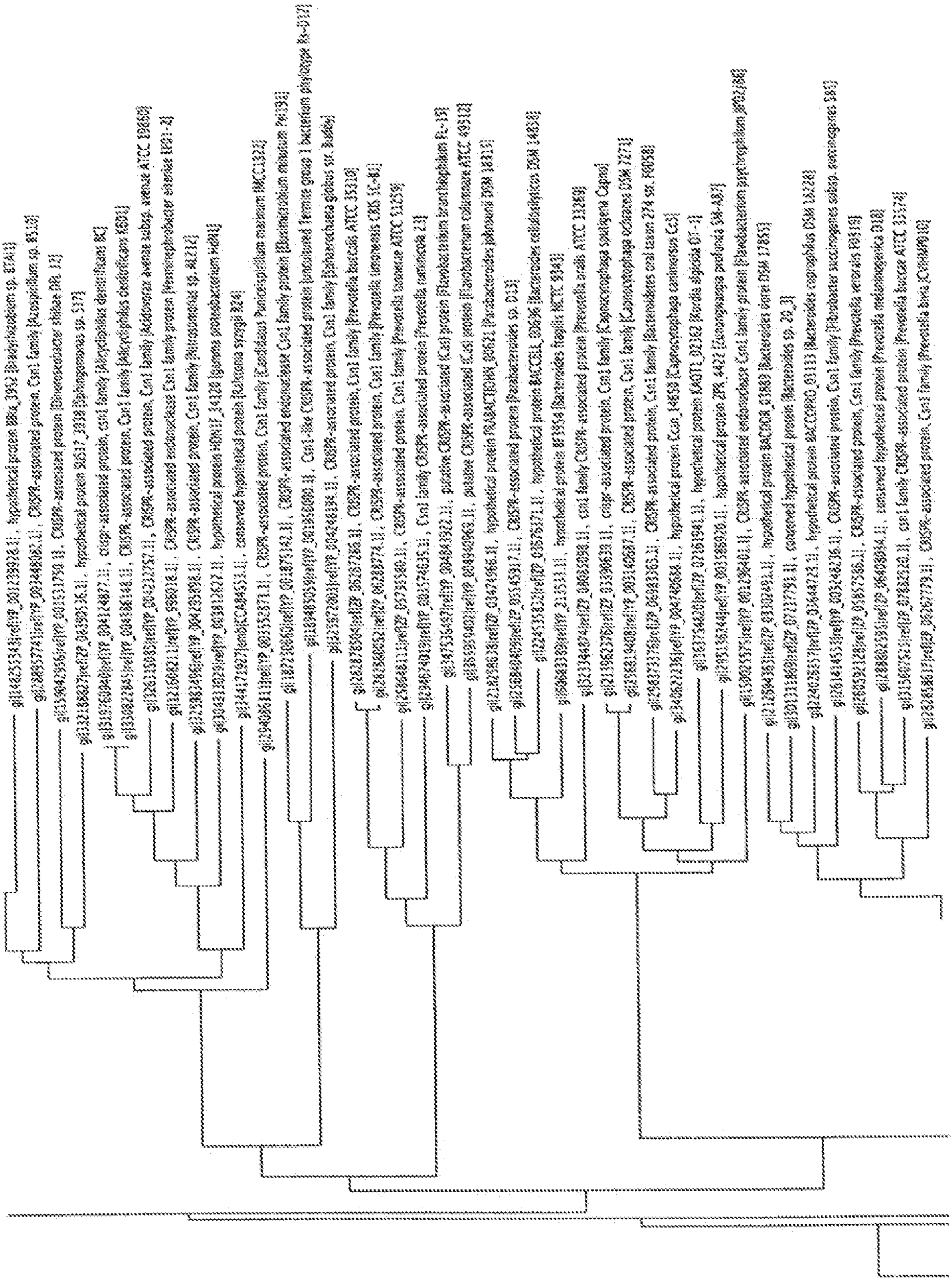


FIG. 9B

SUBSTITUTE SHEET (RULE 26)

FIG. 6C



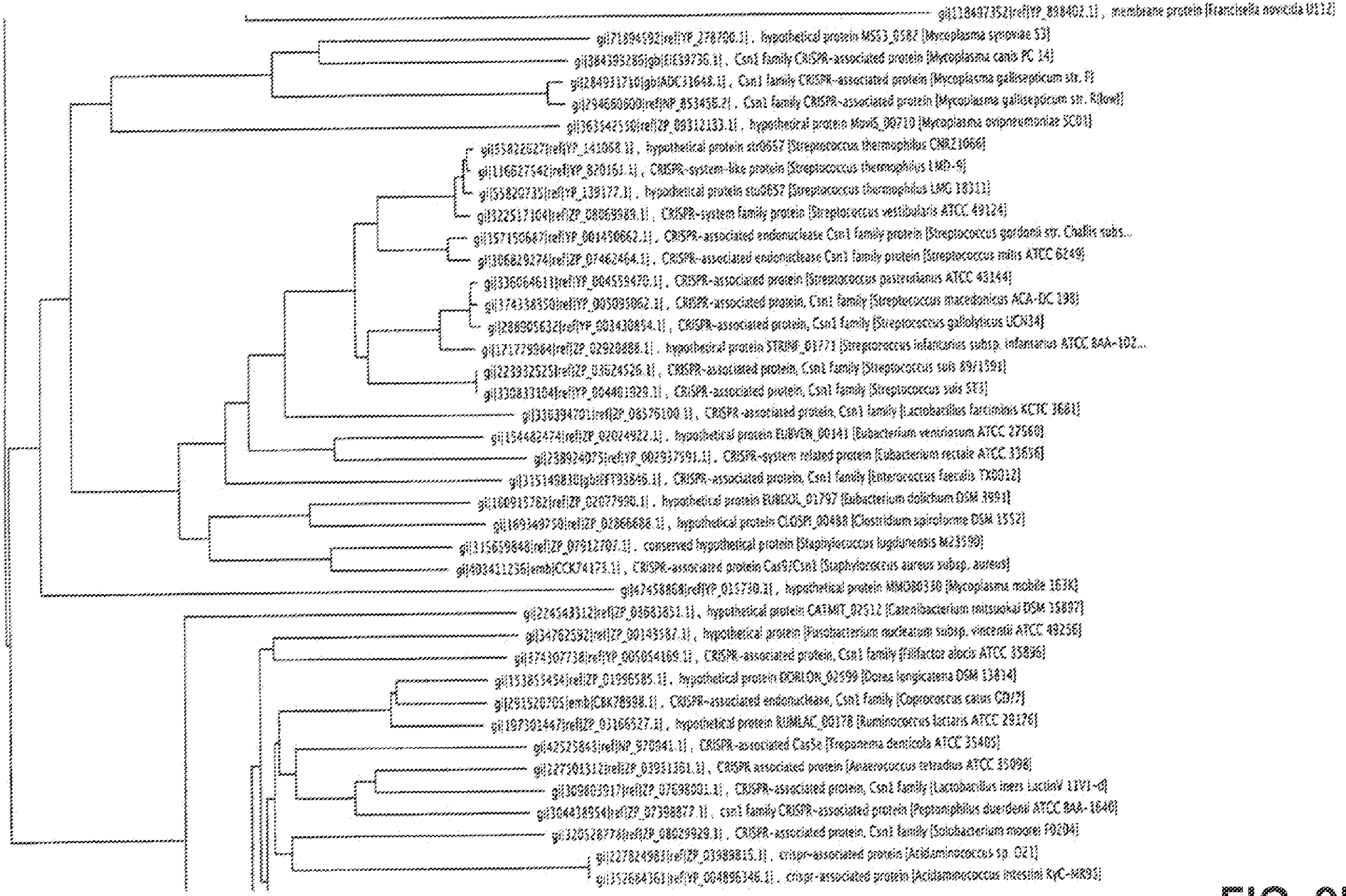


FIG. 9D

SUBSTITUTE SHEET (RULE 26)

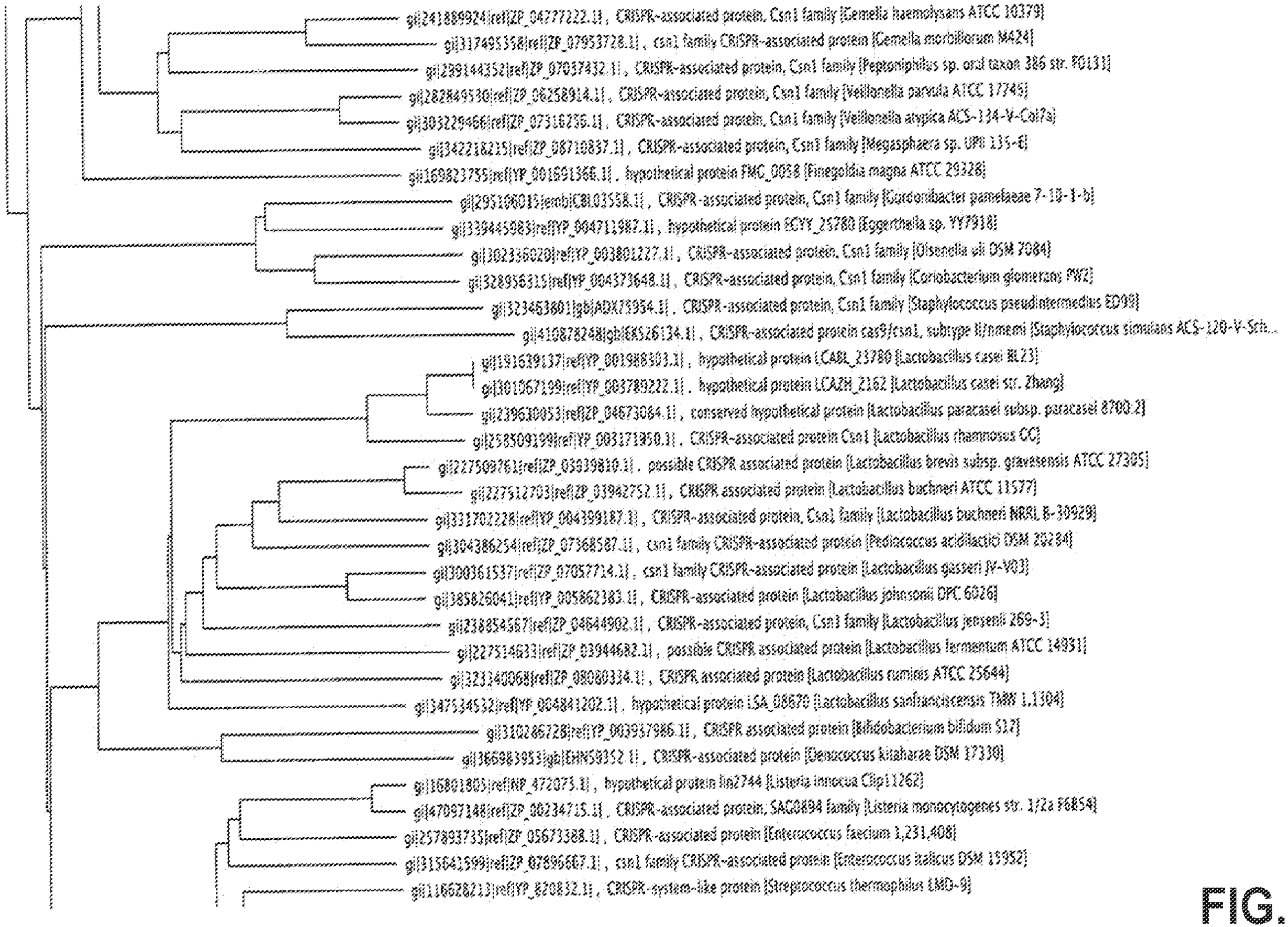


FIG. 9E

CA 02989830 2017-12-15

SUBSTITUTE SHEET (RULE 26)

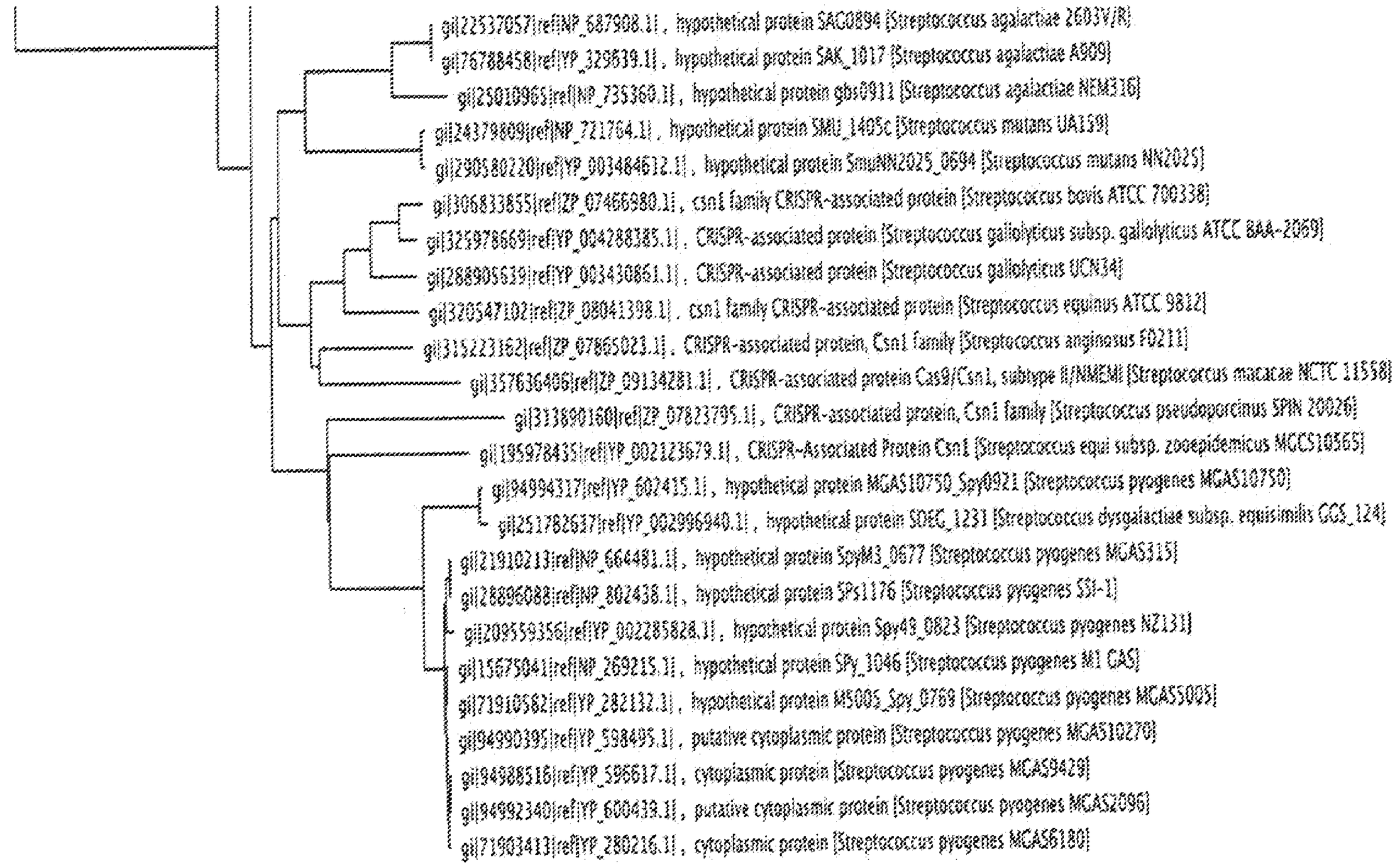


FIG. 9F

SUBSTITUTE SHEET (RULE 26)

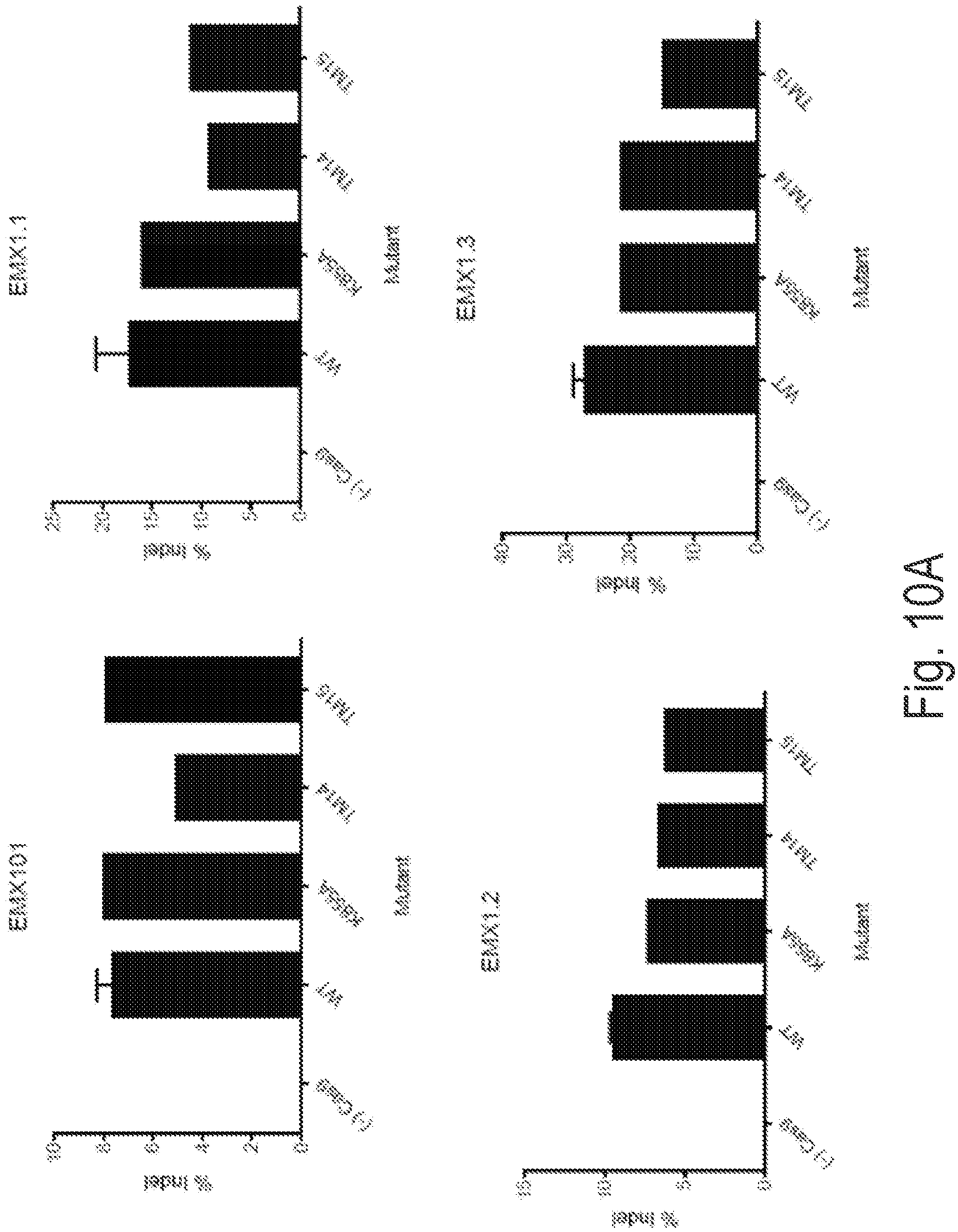
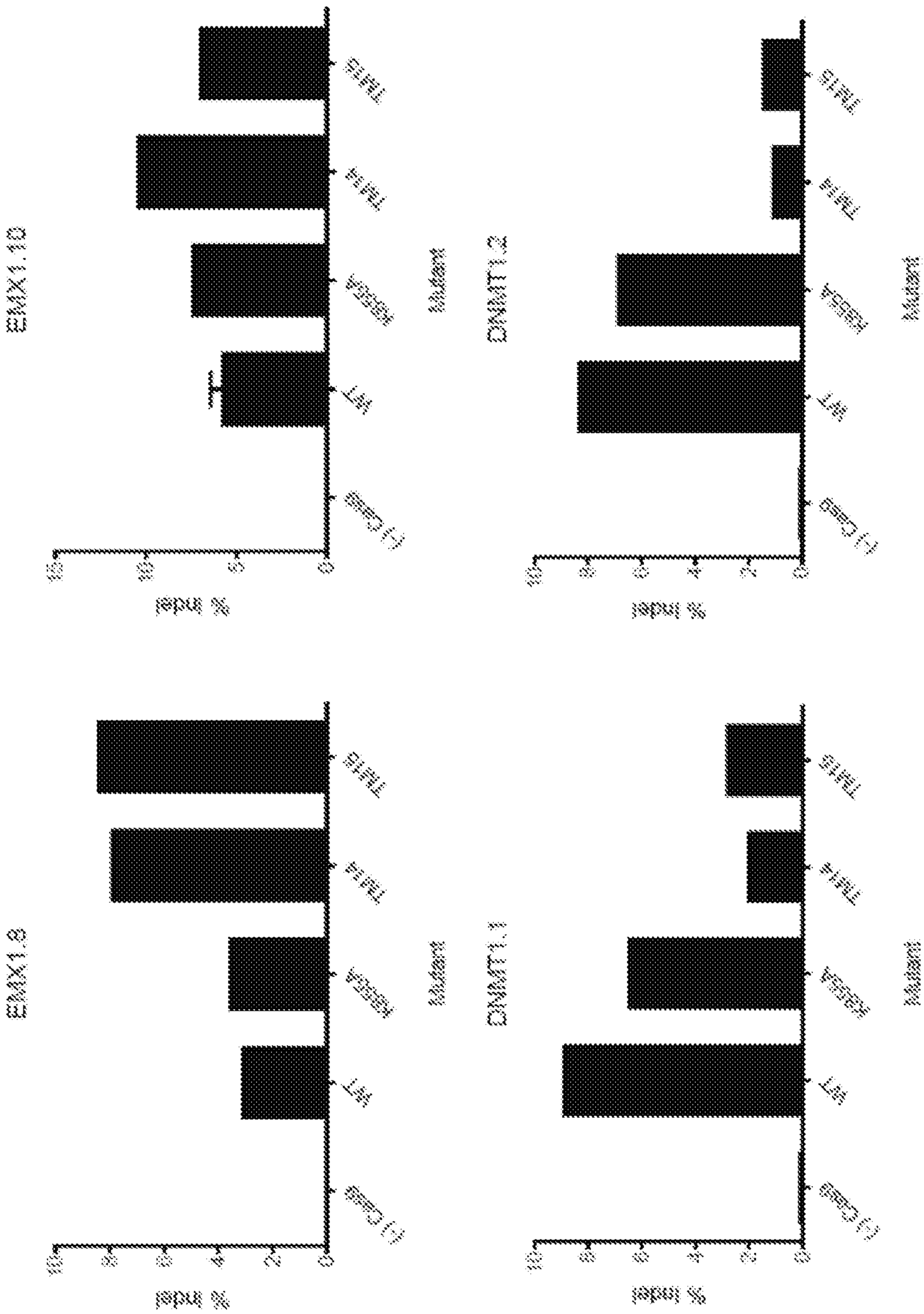

EMX1.1
% indel
25
20
15
10
5
0
SLR#1
SLR#2
SpCas9
UT
GFP (-)
Mutant
EMX1.3
% indel
60
40
20
0
SLR#1
SLR#2
SpCas9
UT
GFP (-)
Mutant
EMX1.01
% indel
80
60
40
20
0
SLR#1
SLR#2
SpCas9
UT
GFP (-)
Mutant
EMX1.2
% indel
60
40
20
0
SLR#1
SLR#2
SpCas9
UT
GFP (-)
Mutant

Fig. 10A


EMX1.10
% Indel
0
20
40
60
HEK293
K562
U87
HUES9
EMX1.9
% Indel
0
20
40
60
HEK293
K562
U87
HUES9

CRMTT1.2
% Indel
0
20
40
60
HEK293
K562
U87
HUES9
CRMTT1.1
% Indel
0
20
40
60
HEK293
K562
U87
HUES9

Fig. 10B

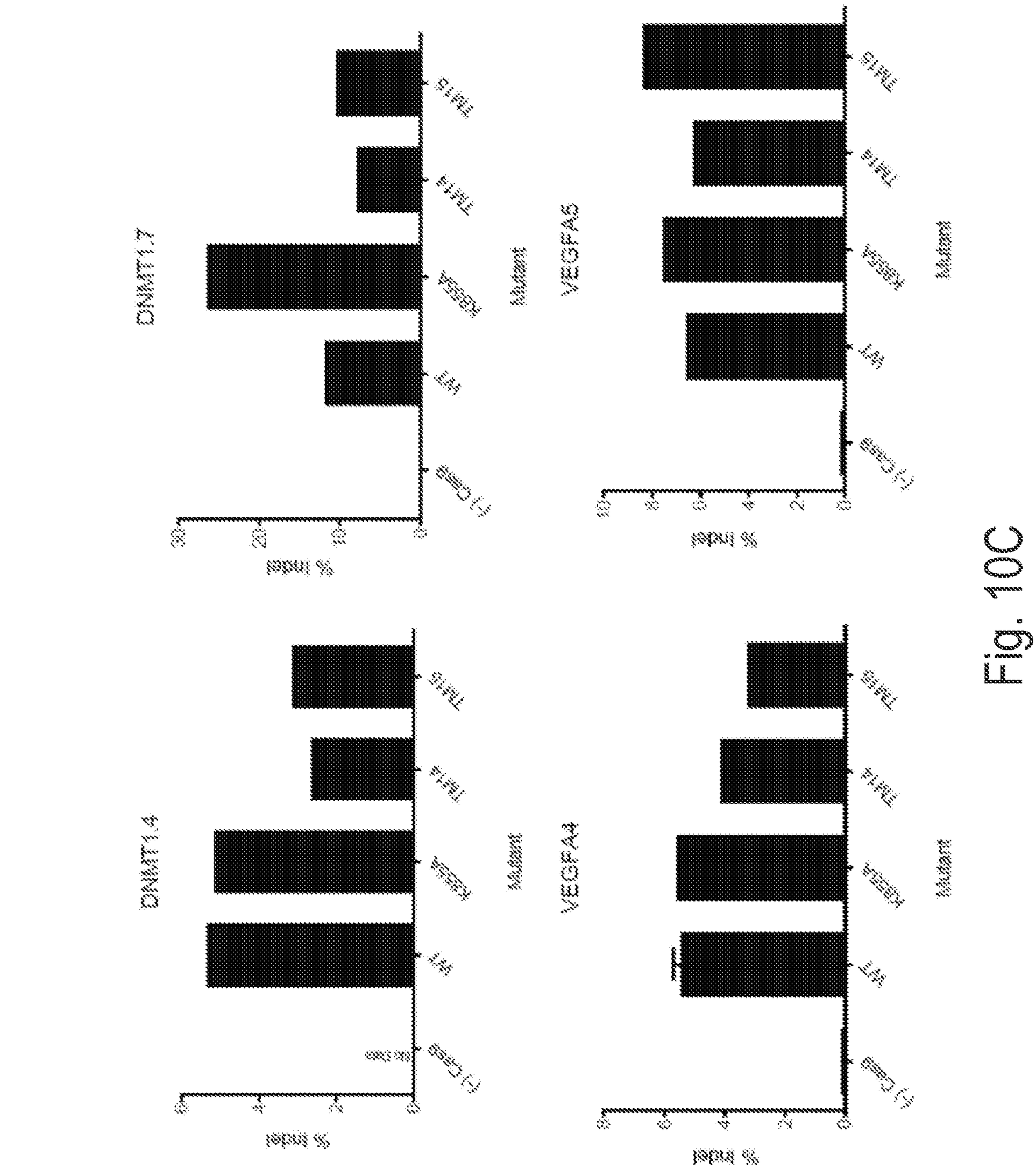

DNMT1.7
% Indel
0
20
40
60
80
EBV-LCL
Jurkat
HEK293
THP1
SH-SY5Y
DNMT1.4
% Indel
0
20
40
60
80
EBV-LCL
Jurkat
HEK293
THP1
SH-SY5Y
VEGFA5
% Indel
0
20
40
60
80
EBV-LCL
Jurkat
HEK293
THP1
SH-SY5Y
VEGFA4
% Indel
0
20
40
60
80
EBV-LCL
Jurkat
HEK293
THP1
SH-SY5Y

Fig. 1C

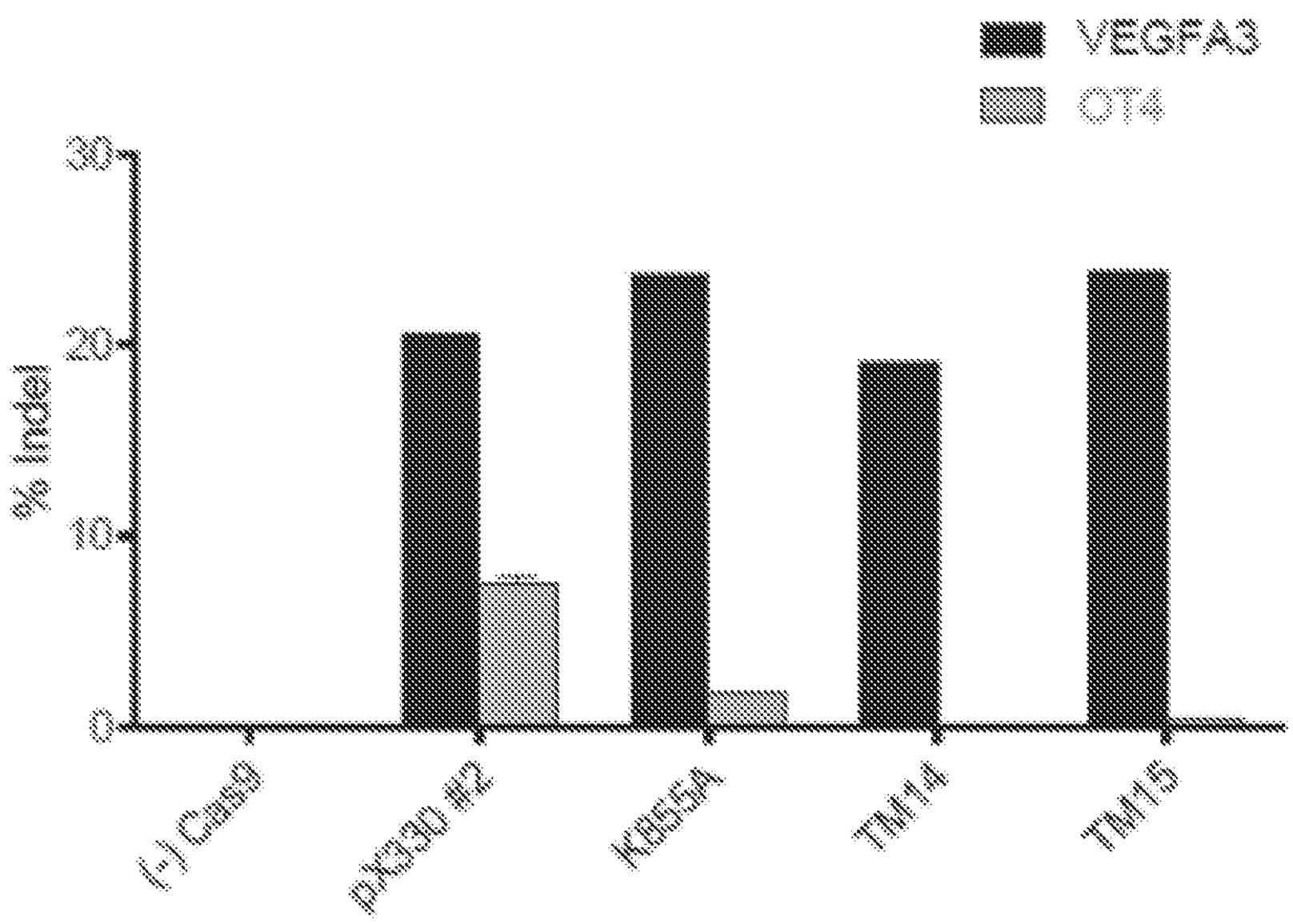


Fig. 10D

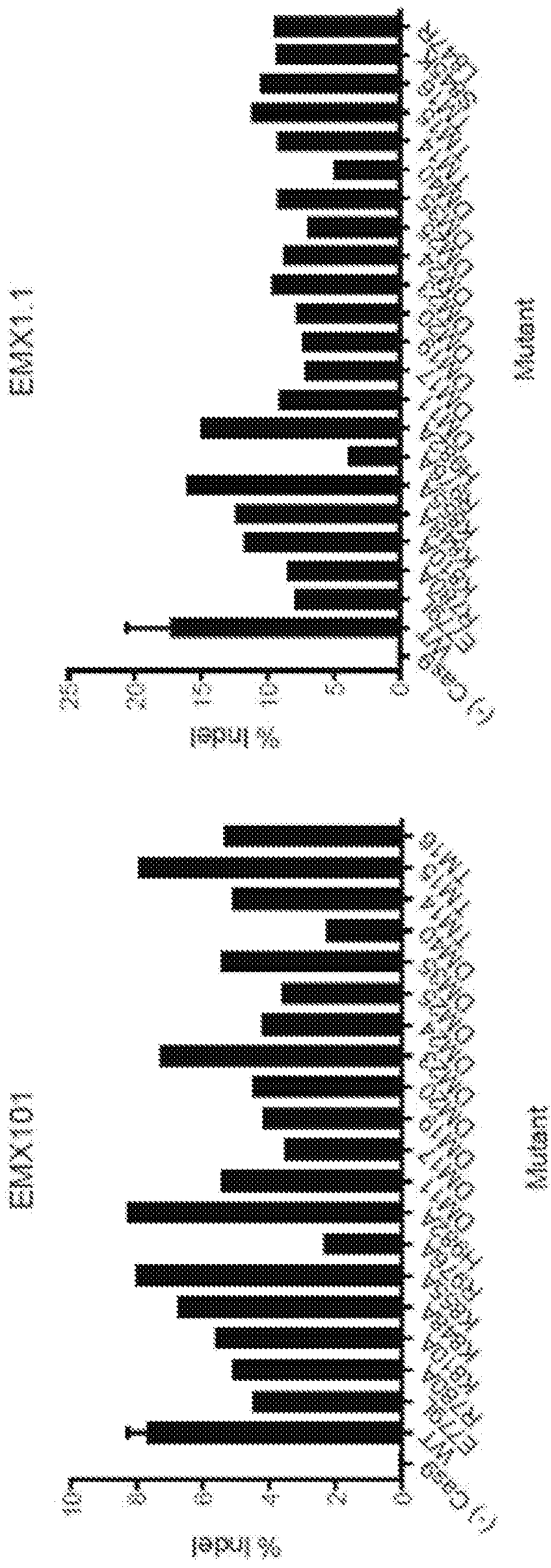
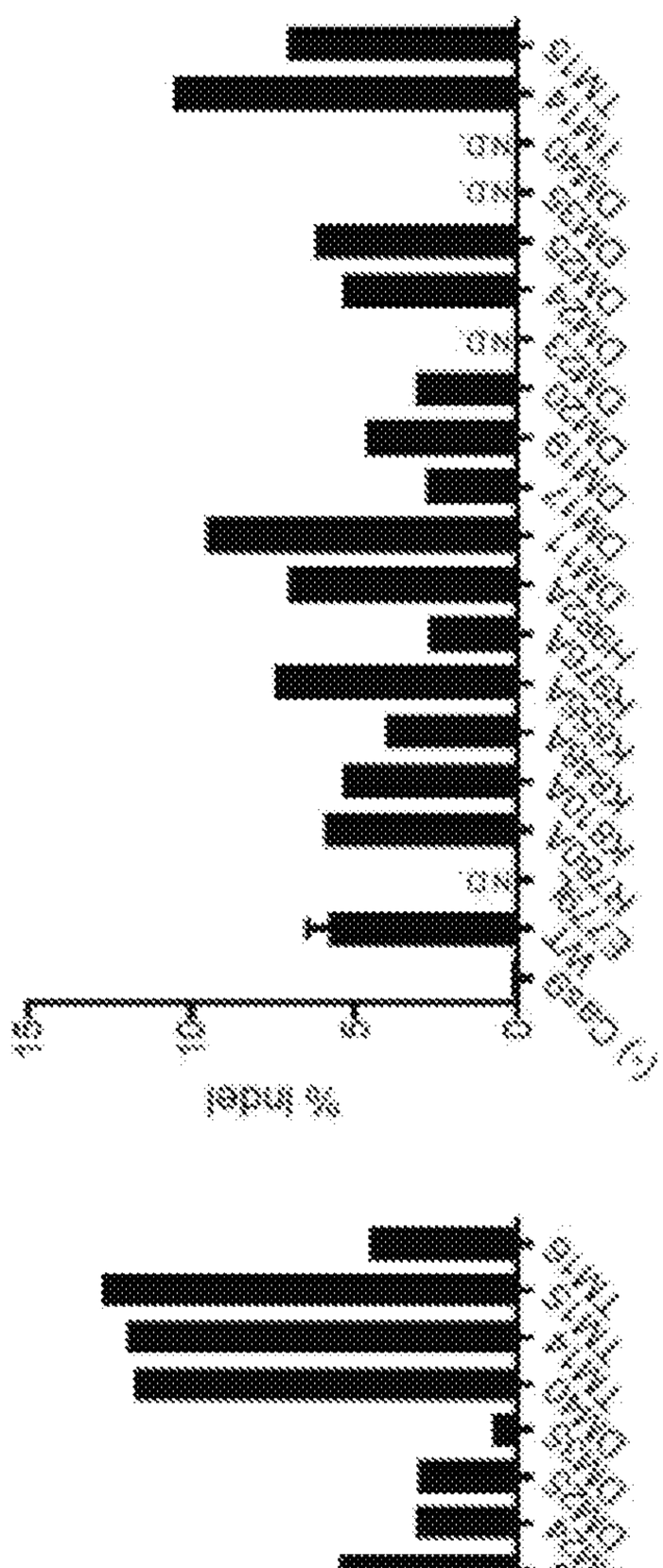

EMX1.1
% indel
Mutant
EMX1.3
% indel
Mutant
EMX1.1
% indel
Mutant
EMX1.2
% indel
Mutant

Fig. 1A


EMX1.10
Mutant
% Indel
EMX1.6
Mutant
% Indel
Input %
0
2
4
6
8
10

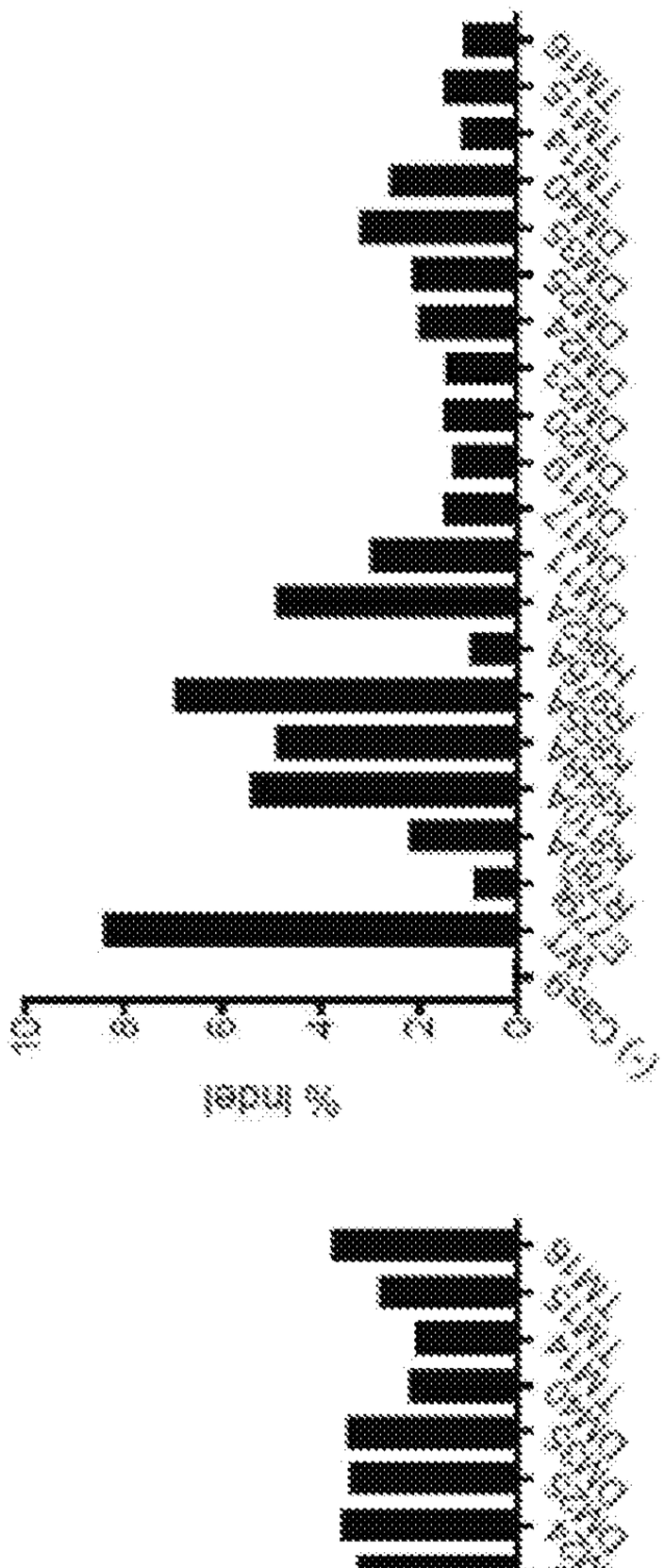

DNMT1.2
Mutant
% Indel
DNMT1.1
Mutant
% Indel
Input %
0
2
4
6
8
10

Fig. 11B

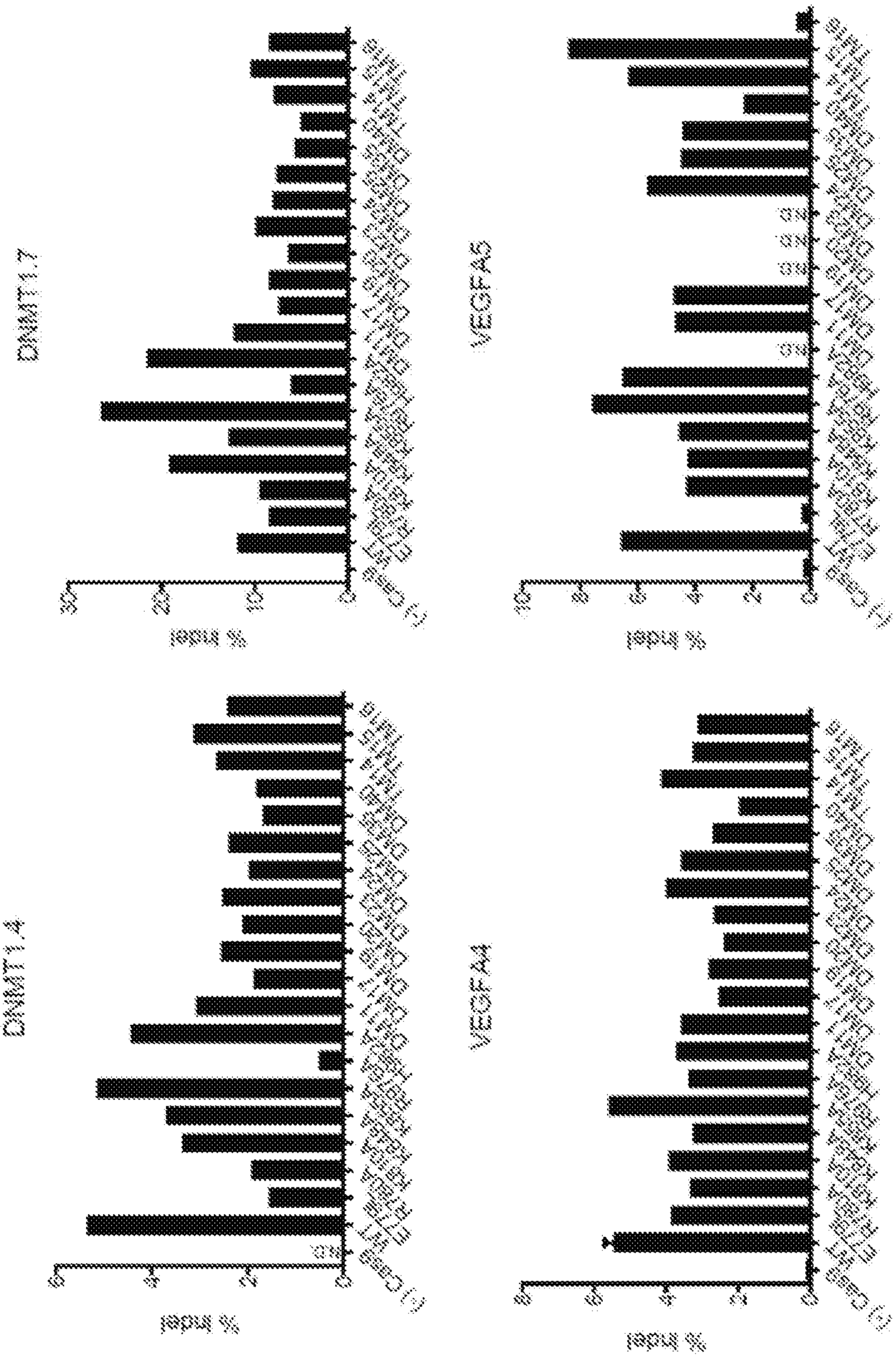


Fig. 1C

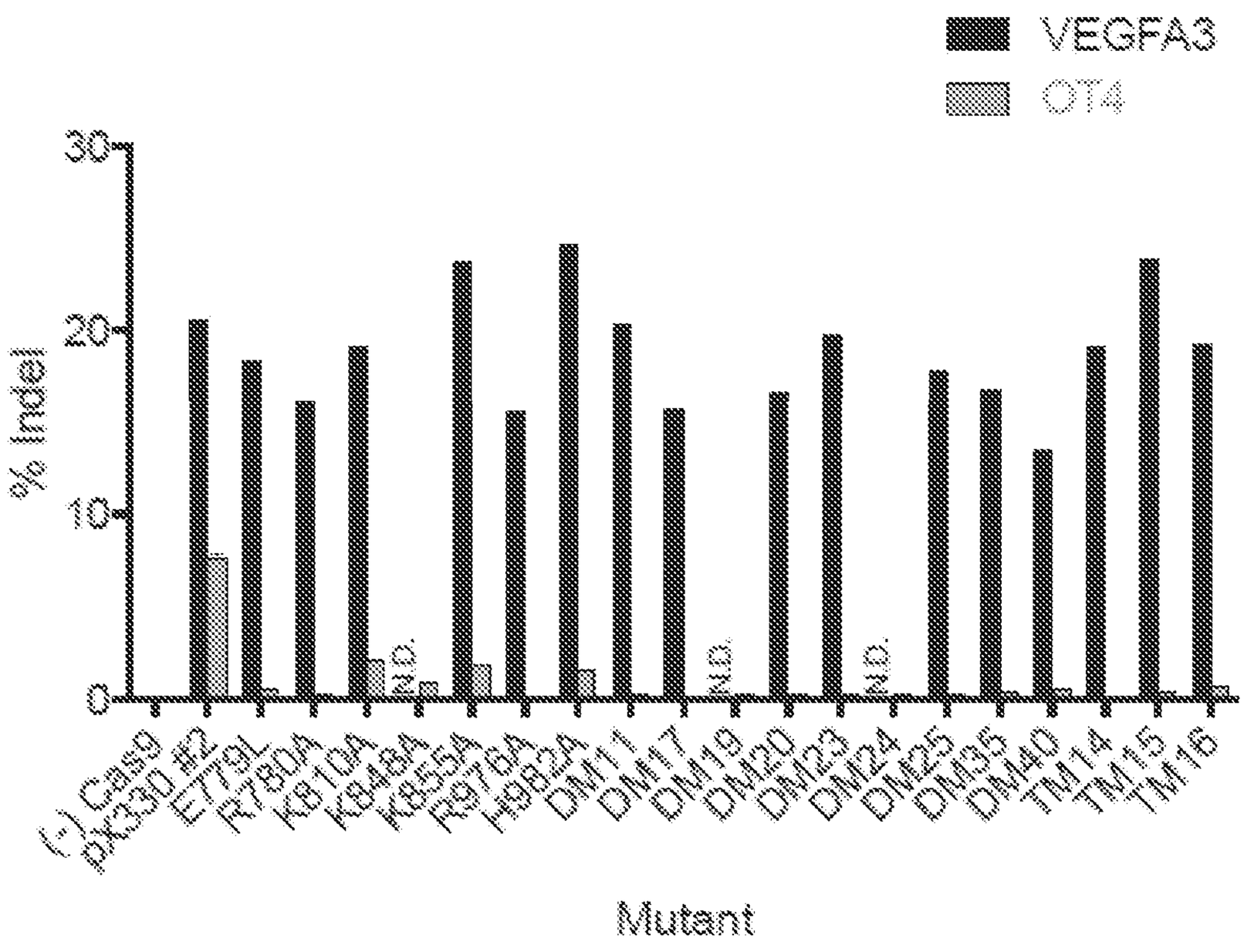

VEGFA3
OT4
% Indel
30
20
10
0
N.D.
N.D.
N.D.
(-) Cas9
pX330 #2
E779L
R780A
K810A
K848A
K855A
R976A
H982A
DM11
DM17
DM19
DM20
DM23
DM24
DM25
DM35
DM40
TM14
TM15
TM16
Mutant

Fig. 11D

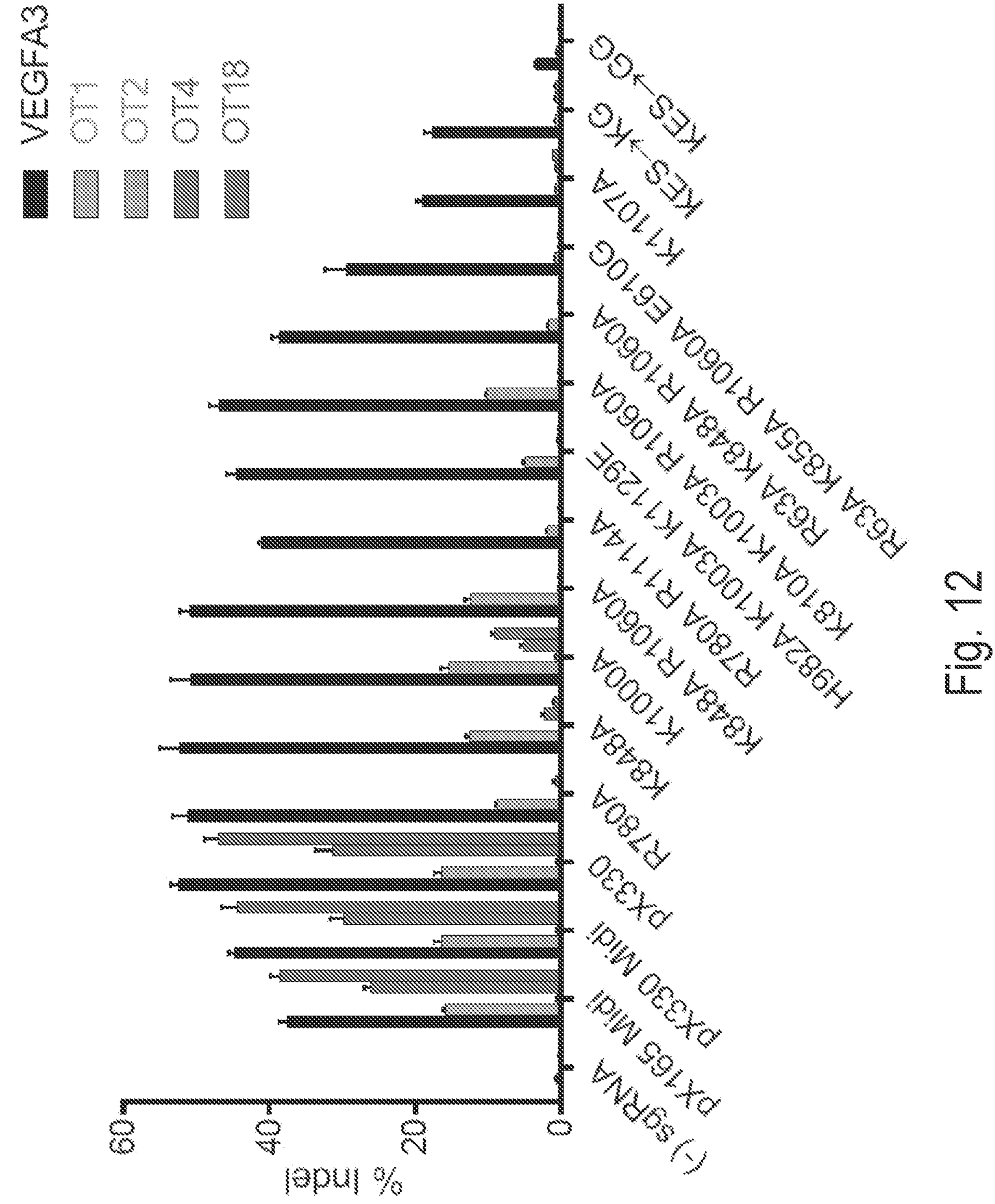

VEGFA3
OT1
OT2
OT14
OT18
% indel
0
20
40
60
SpCas9 (-)
eSpCas9 (1.1)
SpCas9-HF1
HypaCas9
xCas9 3.7
xCas9 3.6
xCas9 3.5
xCas9 3.4
xCas9 3.3
xCas9 3.2
xCas9 3.1
xCas9 3.0
xCas9 2.7
xCas9 2.6
xCas9 2.5
xCas9 2.4
xCas9 2.3
xCas9 2.2
xCas9 2.1
xCas9 2.0
xCas9 1.9
xCas9 1.8
xCas9 1.7
xCas9 1.6
xCas9 1.5
xCas9 1.4
xCas9 1.3
xCas9 1.2
xCas9 1.1
xCas9 1.0
xCas9 0.9
xCas9 0.8
xCas9 0.7
xCas9 0.6
xCas9 0.5
xCas9 0.4
xCas9 0.3
xCas9 0.2
xCas9 0.1

Fig. 12

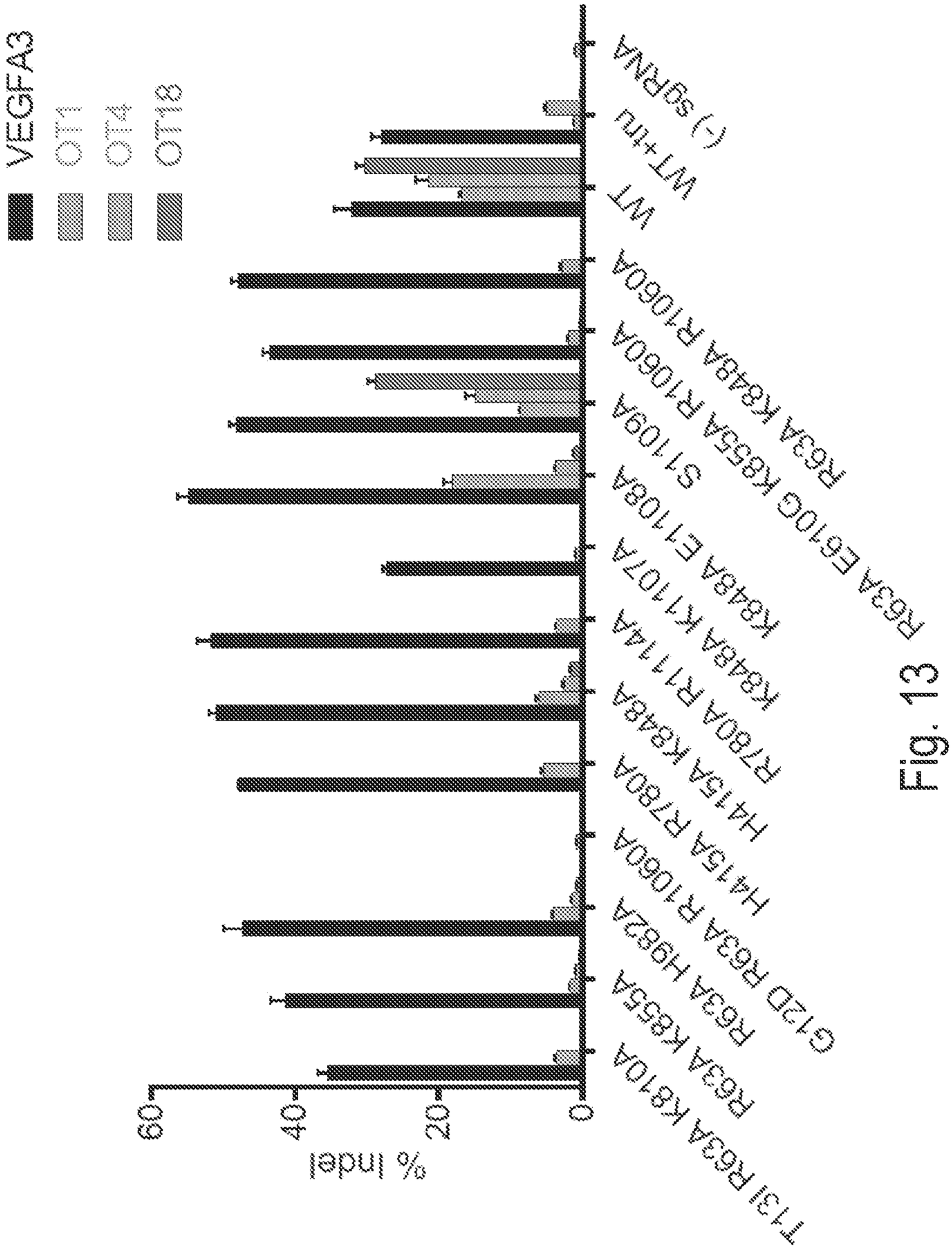
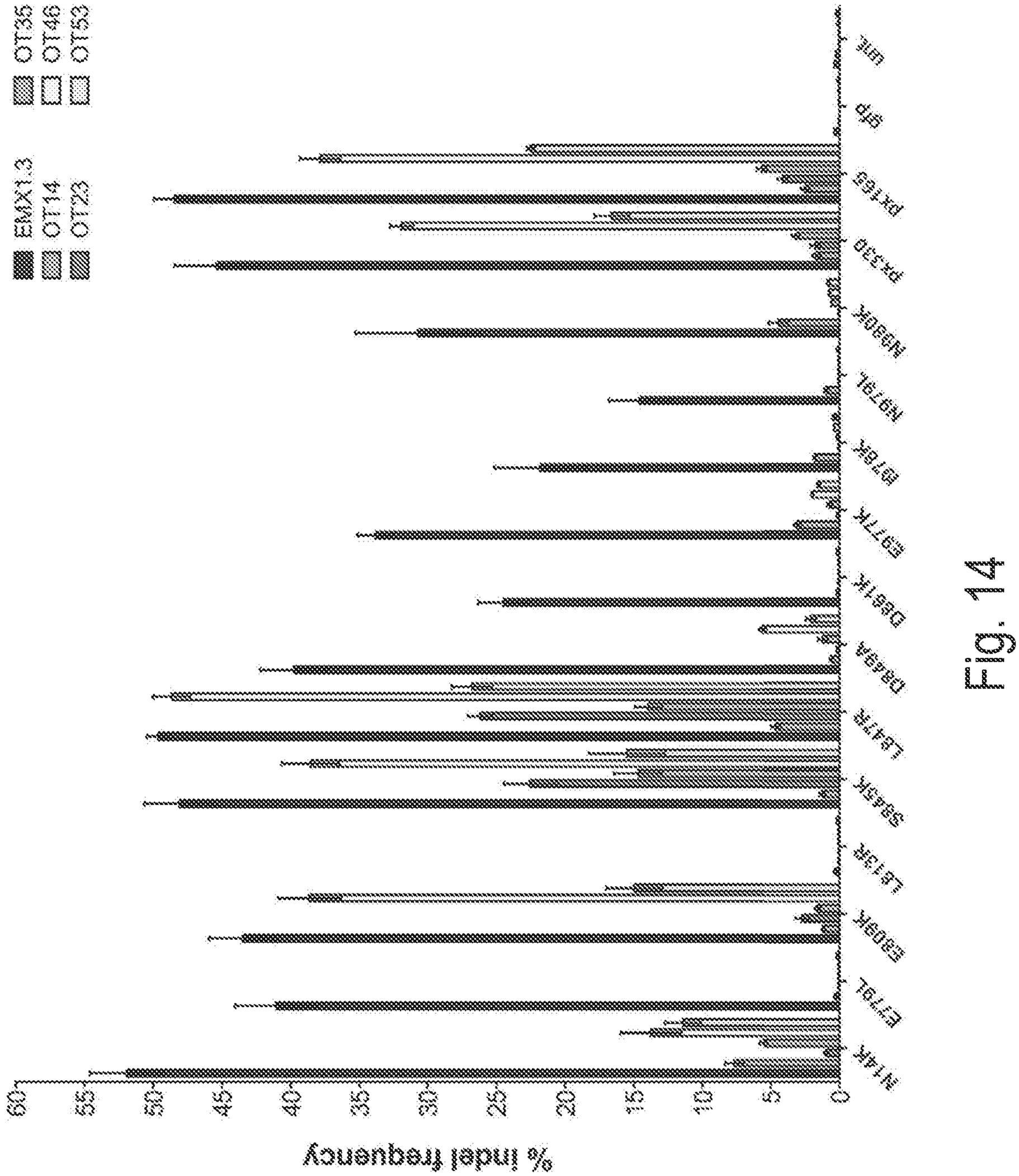

VEGFA3
OT1
OT4
OT18
% indel
0
20
40
60
VNEGS (-)
mut+LNA
LNA
W90A W84A R106A
W90A W55A R106A
S110A R106A
R63A E61OA
R63A K84A
R63A H96A
R63A R106A
H4+5A R78OA
H4+5A K84A
R78OA R113A
R78OA K84A
K84A K110A
K84A E110A
G12D R63A R106A
R63A K84A
R63A E61OA

Fig. 13


EMX1.3
OT14
OT23
OT35
OT46
OT53
UDE
dP
PX-tE3
PX-3J9
PX-3J9K
HEK293FT
HEK293
HEK293K
HEK293TK
HEK293TL
HEK293TLK
HEK293TLK
HEK293TLK

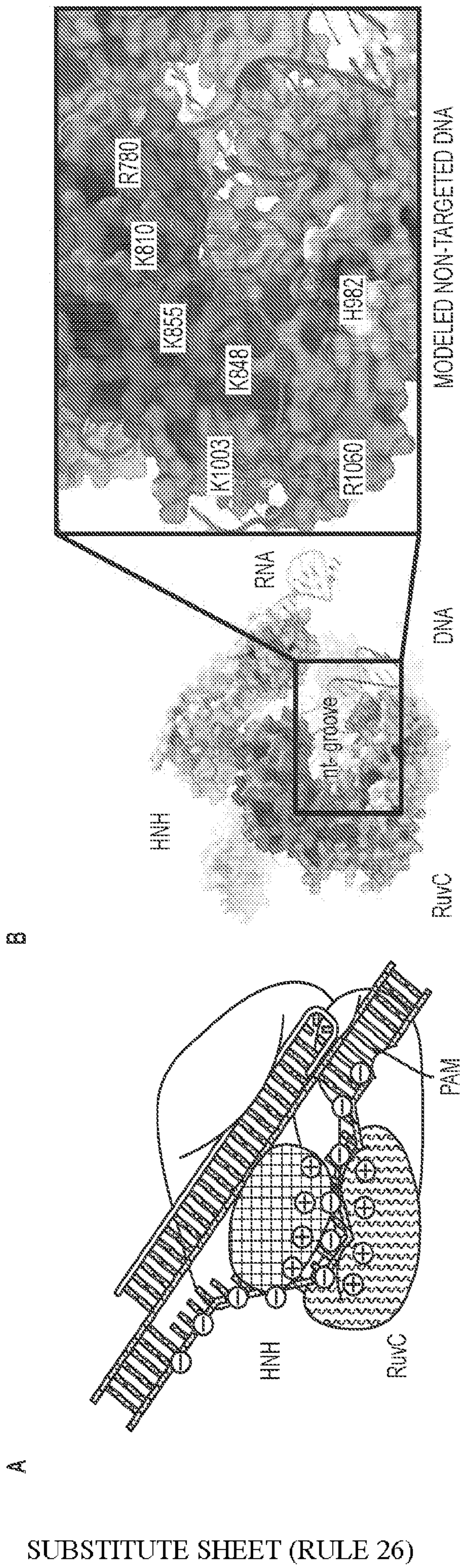

R780
K810
K855
K848
K1003
R1060
H982
MODELED NON-TARGETED DNA
DNA
RNA
nt- groove
HNH
RuvC
PAM
HNH
RuvC

Fig. 15A - 15B

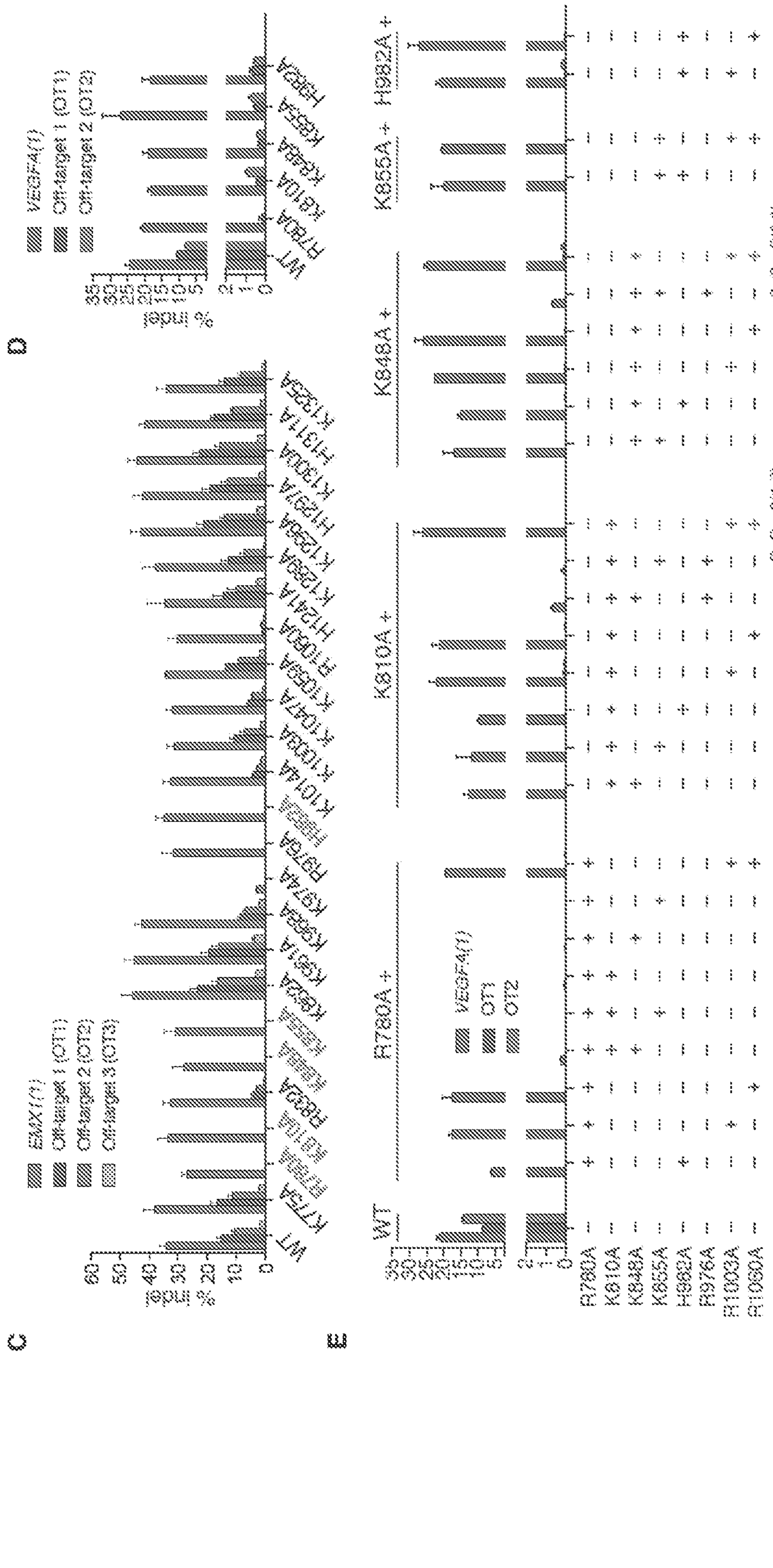

B
C
D
E
VEGFA(1)
Off-target 1 (OT1)
Off-target 2 (OT2)
Off-target 3 (OT3)
% indel
WT
R780A
K810A
K848A
K855A
H982A
R976A
R1003A
R1060A
esp Cas9 (1.1)
esp Cas9 (1.0)
HiFi Cas9
VQR
EQR
VRER
SpCas9-HF1
eSpCas9 (1.1)
K848A +
K855A +
H982A +
R780A +
K810A +
R976A +
R1003A +
R1060A +

Fig. 15C-15E

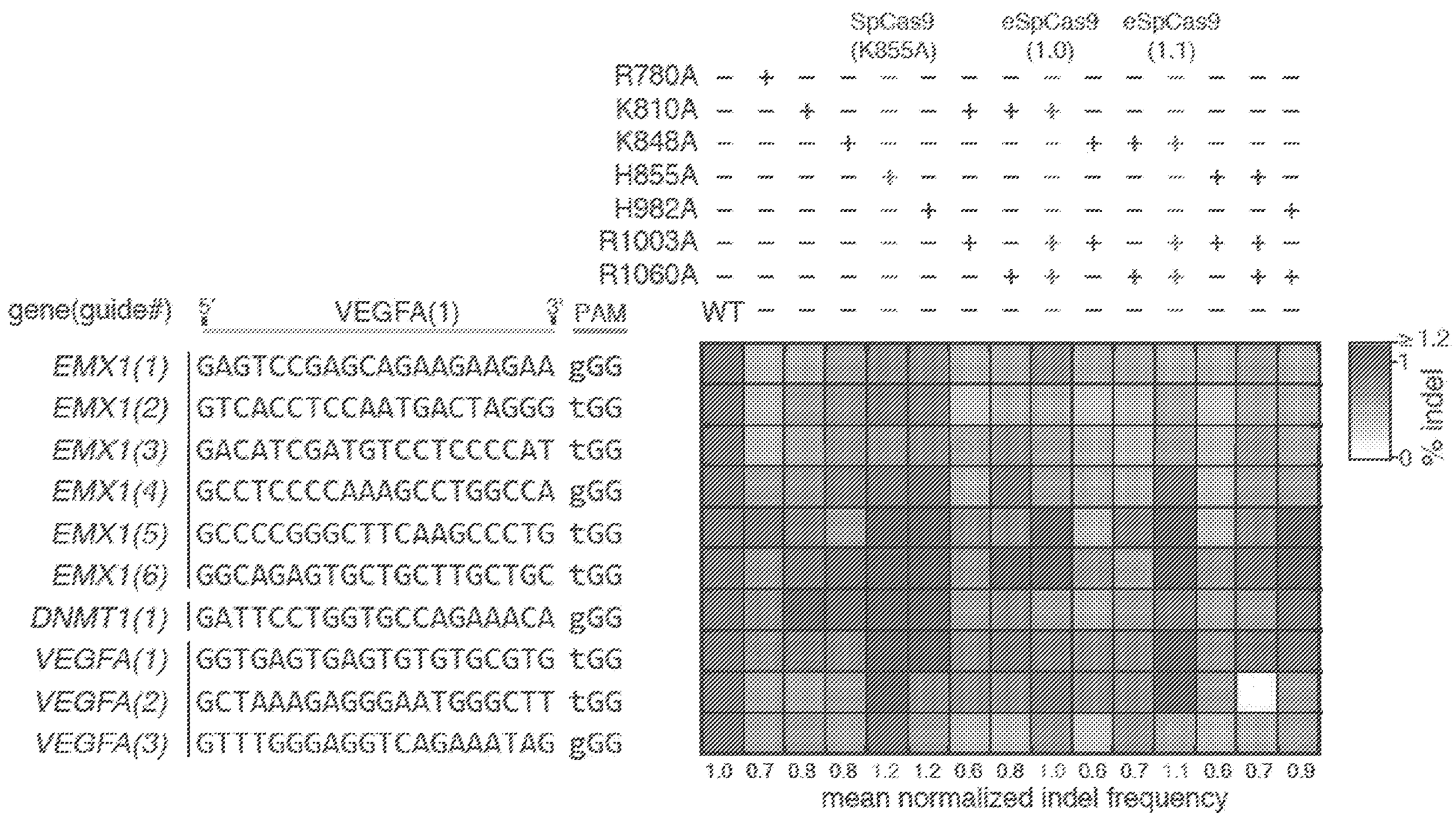

SpCas9 (K855A)
eSpCas9 (1.0)
eSpCas9 (1.1)
R780A
K810A
K848A
H855A
H982A
R1003A
R1060A
WT
gene(guide#)
5'
VEGFA(1)
3'
PAM
EMX1(1) GAGTCCGAGCAGAAGAAGAA gGG
EMX1(2) GTCACCTCCAATGACTAGGG tGG
EMX1(3) GACATCGATGTCCTCCCCAT tGG
EMX1(4) GCCTCCCCAAAGCCTGGCCA gGG
EMX1(5) GCCCCGGGCTTCAAGCCCTG tGG
EMX1(6) GGCAGAGTGCTGCTTGCTGC tGG
DNMT1(1) GATTCCTGGTGCCAGAAACA gGG
VEGFA(1) GGTGAGTGAGTGTGTGCGTG tGG
VEGFA(2) GCTAAAGAGGGAATGGGCTT tGG
VEGFA(3) GTTTGGGAGGTCAGAAATAG gGG
≥1.2
1
0
indel %
1.0 0.7 0.8 0.8 1.2 1.2 0.6 0.8 1.0 0.6 0.7 1.1 0.6 0.7 0.9
mean normalized indel frequency

Fig. 15F

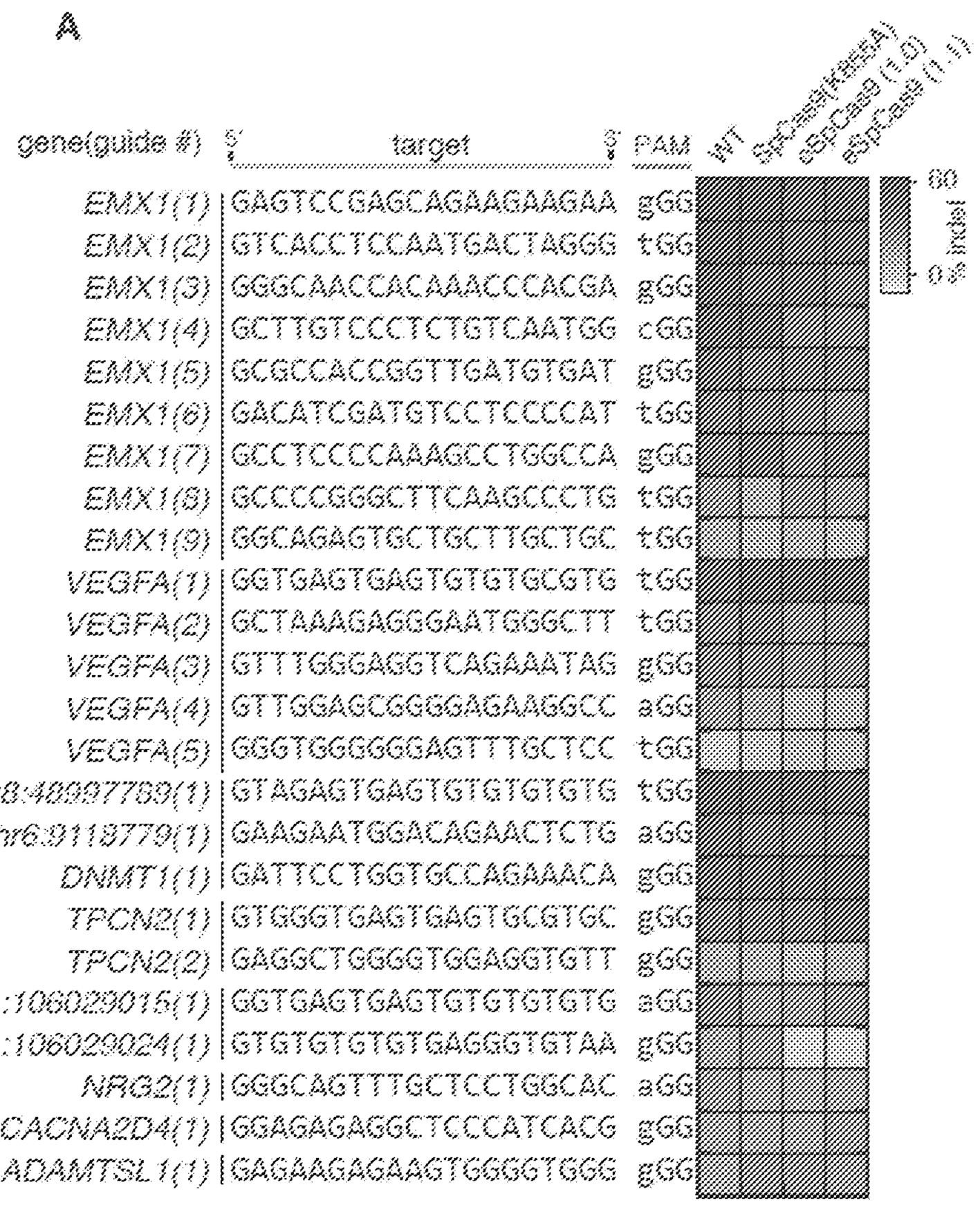

A
gene(guide #)
5'
target
3'
PAM
WT
SpCas9(K855A)
eSpCas9 (1.0)
eSpCas9 (1.1)
80
Indel
0
EMX1(1) GAGTCCGAGCAGAAGAAGAA gGG
EMX1(2) GTCACCTCCAATGACTAGGG tGG
EMX1(3) GGGCAACCACAAACCCACGA gGG
EMX1(4) GCTTGTCCCTCTGTCAATGG cGG
EMX1(5) GCGCCACCGGTTGATGTGAT gGG
EMX1(6) GACATCGATGTCCTCCCCAT tGG
EMX1(7) GCCTCCCCAAAGCCTGGCCA gGG
EMX1(8) GCCCCGGGCTTCAAGCCCTG tGG
EMX1(9) GGCAGAGTGCTGCTTGCTGC tGG
VEGFA(1) GGTGAGTGAGTGTGTGCGTG tGG
VEGFA(2) GCTAAAGAGGGAATGGGCTT tGG
VEGFA(3) GTTTGGGAGGTCAGAAATAG gGG
VEGFA(4) GTTGGAGCGGGGAGAAGGCC aGG
VEGFA(5) GGGTGGGGGGAGTTTGCTCC tGG
Chr8:48997789(1) GTAGAGTGAGTGTGTGTGTG tGG
Chr6:91118779(1) GAAGAATGGACAGAACTCTG aGG
DNMT1(1) GATTCCTGGTGCCAGAAACA gGG
TPCN2(1) GTGGGTGAGTGAGTGCGTGC gGG
TPCN2(2) GAGGCTGGGGTGGAGGTGTT gGG
hr14:106029015(1) GGTGAGTGAGTGTGTGTGTG aGG
hr14:106029024(1) GTGTGTGTGTGAGGGTGTAA gGG
NRG2(1) GGGCAGTTTGCTCCTGGCAC aGG
CACNA2D4(1) GGAGAGAGGCTCCCATCACG gGG
ADAMTSL1(1) GAGAAGAGAAGTGGGGTGGG gGG
B
normalized indel %
2.0
1.5
1.0
0.5
0.0
SpCas9(K855A)
eSpCas9 (1.0)
eSpCas9 (1.1)
C
SpCas9
(-)
WT
WT
K855A
e1.0
e1.1
SpCas9
GAPDH

Fig. 16A-16C

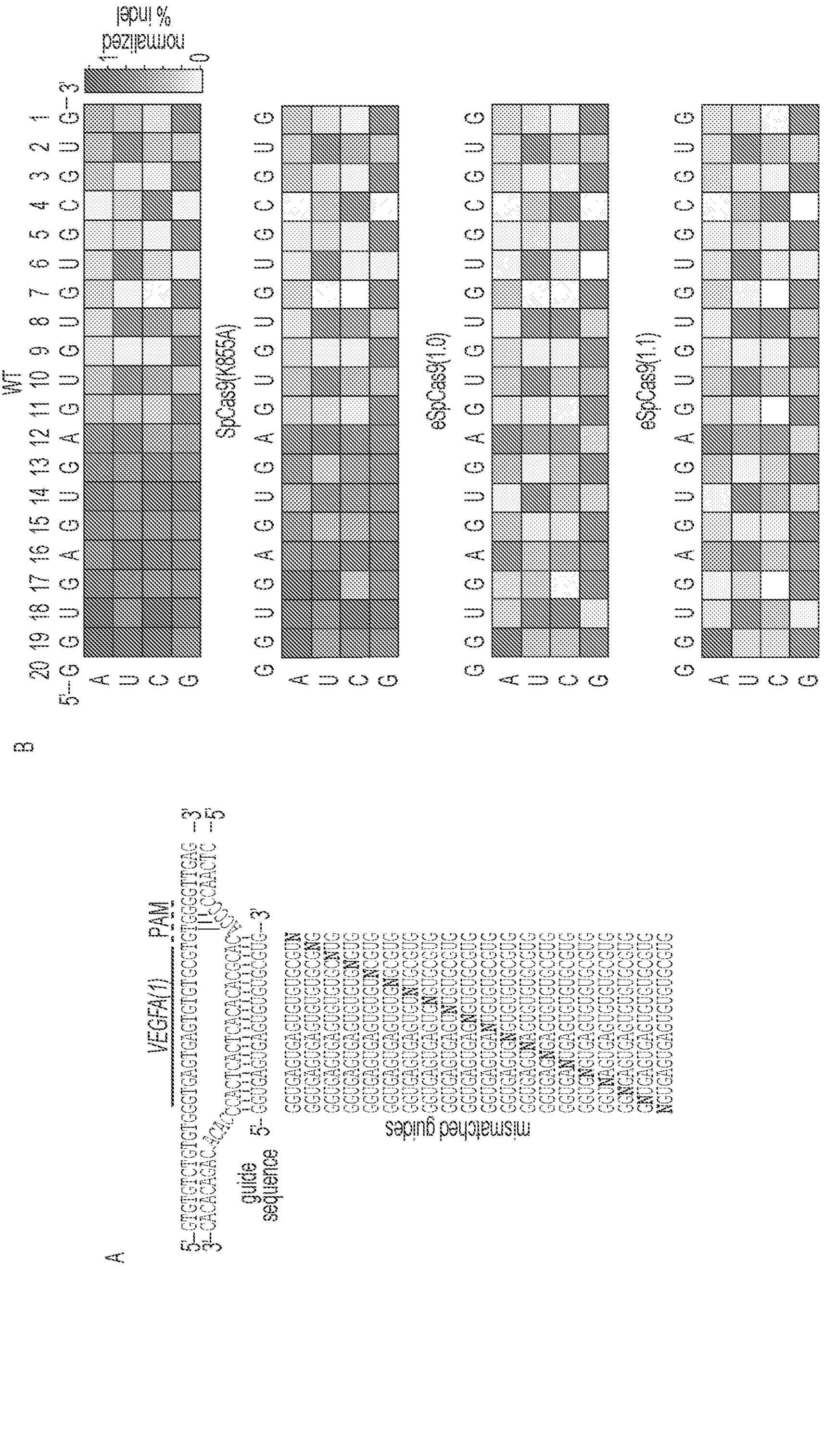

A
B
WT
SpCas9(K855A)
eSpCas9(1.0)
eSpCas9(1.1)
normalized % indel
5'--G
20 19 18 17 16 15 14 13 12 11 10 9 8 7 6 5 4 3 2 1
G G U C A G C U G C A G U G U C G U G C C G U C G--3'
A U C G
VEGFA(1)
PAM
5'-GTCTGTCTGTGTGGTGAGTGAGTGTGTGCGTGTGGGGTTGAG -3'
3'-CACACAGACACACCACTCACTCACACACGCACACCCCAACTC -5'
guide sequence
5'-GGUGAGUGAGUGUGUGCGUGUGGGGUUGAG-3'
mismatched guides

Fig. 17A-17B

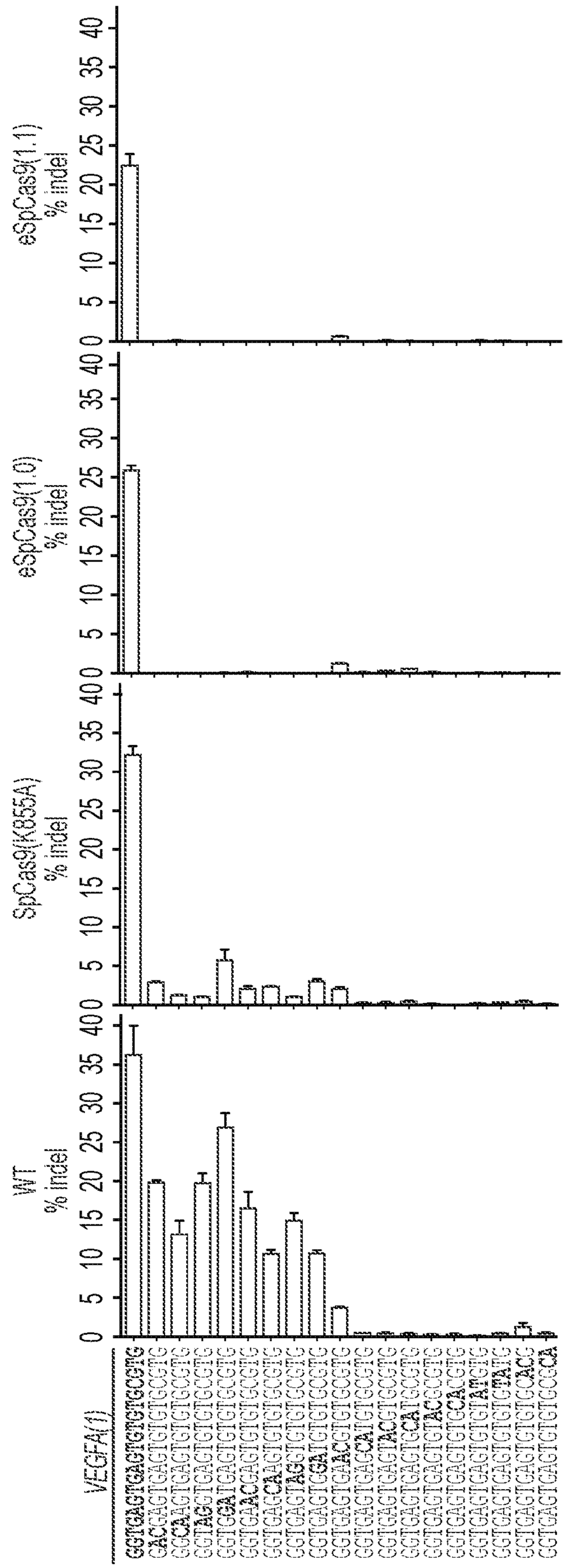
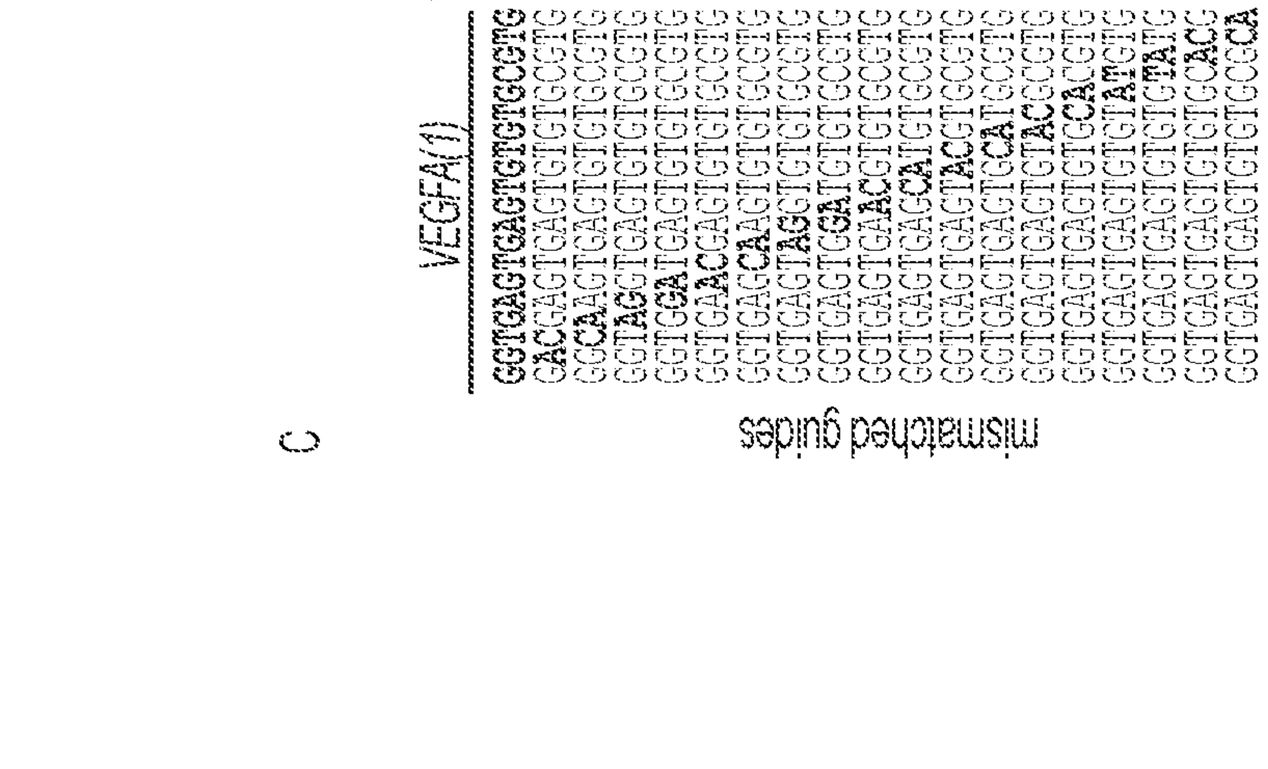
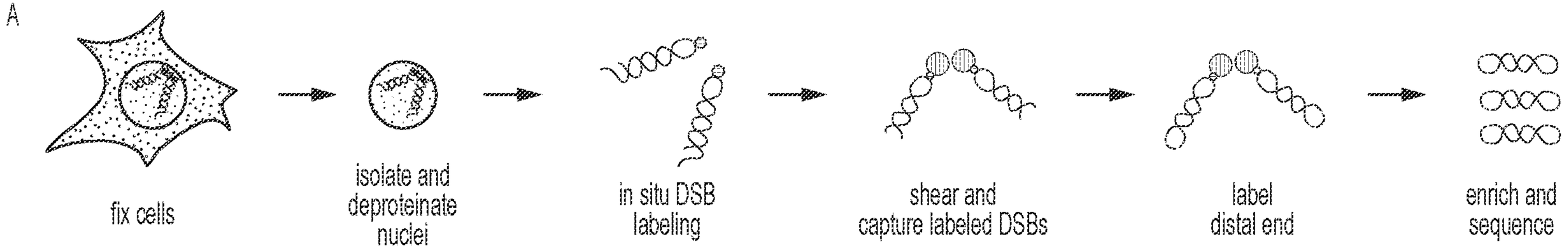

eSpCas9(1.1)
% indel
eSpCas9(1.0)
% indel
SpCas9(K855A)
% indel
WT
% indel
0
5
10
15
20
25
30
35
40

Fig. 17C


C
mismatched guides
VEGFA(1)
GGTGAGTGAGTGTGTGCGTG
GAGGAGTGAGTGTGTGCGTG
GGCAAGTGAGTGTGTGCGTG
GGTAGGTGAGTGTGTGCGTG
GGTGGATGAGTGTGTGCGTG
GGTGAACGAGTGTGTGCGTG
GGTGAGCAAGTGTGTGCGTG
GGTGAGTAGGTGTGTGCGTG
GGTGAGTGGATGTGTGCGTG
GGTGAGTGAACGTGTGCGTG
GGTGAGTGAGCATGTGCGTG
GGTGAGTGAGTACGTGCGTG
GGTGAGTGAGTGTGCATGCG
GGTGAGTGAGTGTGTACGCG
GGTGAGTGAGTGTGTGCACG
GGTGAGTGAGTGTGTATGTG
GGTGAGTGAGTGTGTCTATG
GGTGAGTGAGTGTGTGCACG
GGTGAGTGAGTGTGTGCGCA


A
fix cells
isolate and deproteinate nuclei
in situ DSB labeling
shear and capture labeled DSBs
label distal end
enrich and sequence

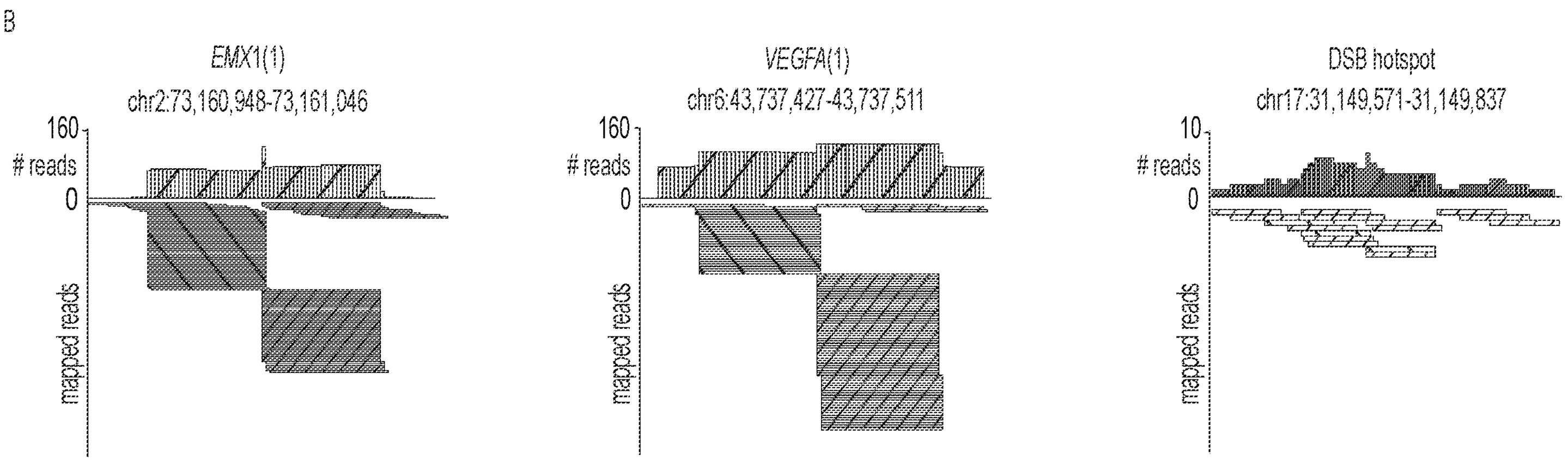

B
EMX1(1)
chr2:73,160,948-73,161,046
160
reads
0
mapped reads
VEGFA(1)
chr6:43,737,427-43,737,511
160
reads
0
mapped reads
DSB hotspot
chr17:31,149,571-31,149,837
10
reads
0
mapped reads

Fig. 18A-18B

Fig. 18C

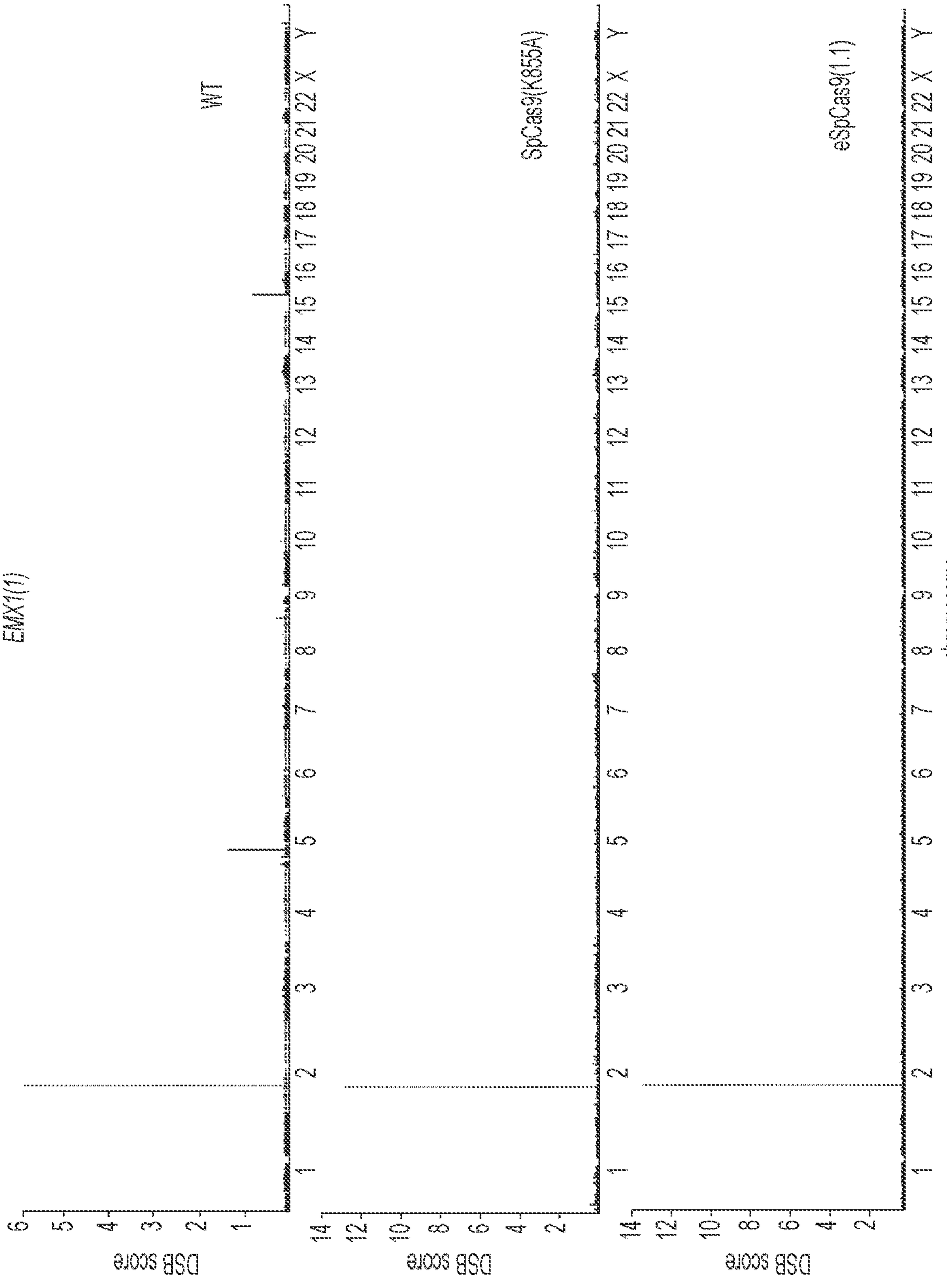
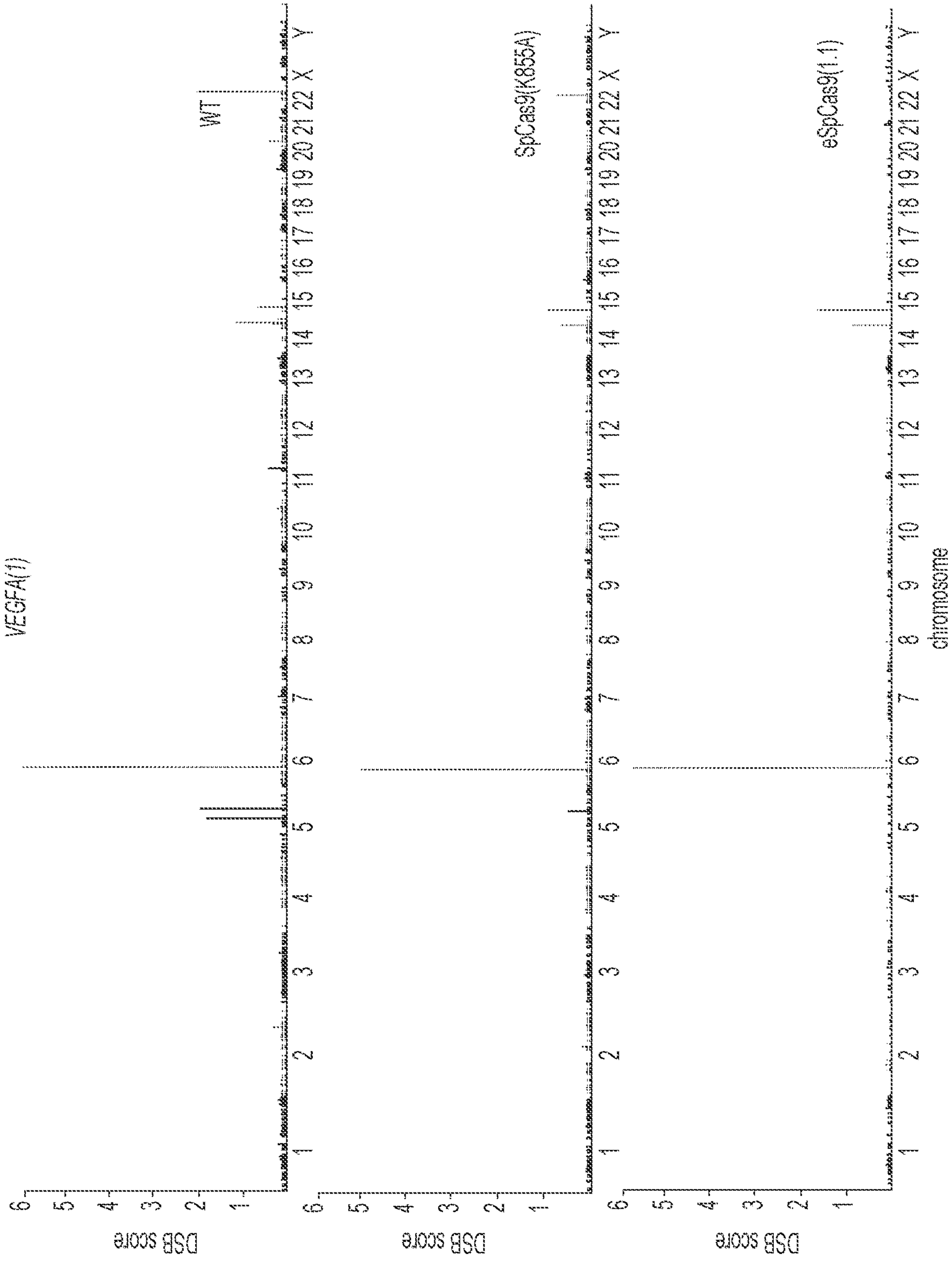


Fig. 18D


WT
SpCas9(K855A)
eSpCas9(1.1)
VEGFA(1)
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 X Y
chromosome
DSB score
1 2 3 4 5 6

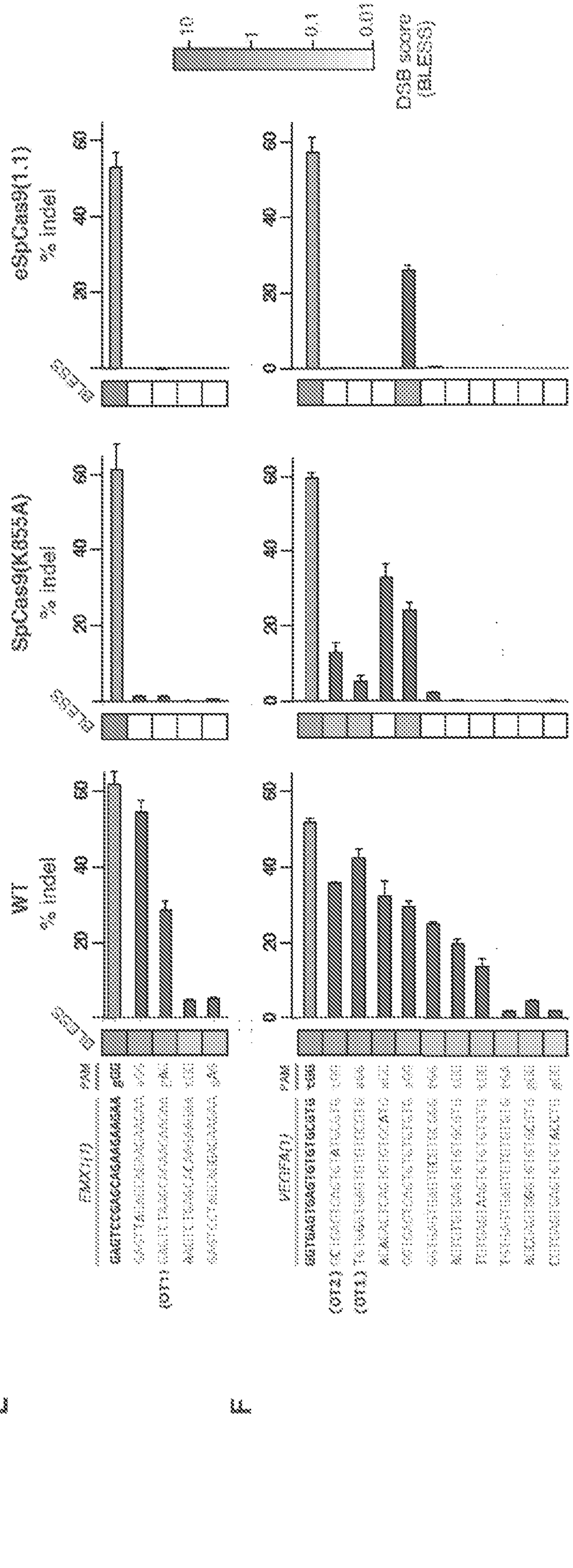

DSB score (BLESS)
10
1
0.1
0.01
eSpCas9(1.1)
% indel
SpCas9(K855A)
% indel
WT
% indel
20
40
60
0
20
40
60
SSC
GUIDE-seq
PAM
GAATCCGAGCAGAAGAAGAA
GAATCCGAGCAGAAGAAGAG
GAATCCGAGCAGAAGAAGTA
GAATCCGAGCAGAAGAAGTT
GAATCCGAGCAGAAGAAGGA
GAATCCGAGCAGAAGAAGCA
GAATCCGAGCAGAAGAAGAA
GAATCCGAGCAGAAGAAGAA
GAATCCGAGCAGAAGAAGAA
GAATCCGAGCAGAAGAAGAA
GAATCCGAGCAGAAGAAGAA
GAATCCGAGCAGAAGAAGAA
VEGFA(1)
(11)
(12)
(13)
EMX1(1)

Fig. 18E-18F

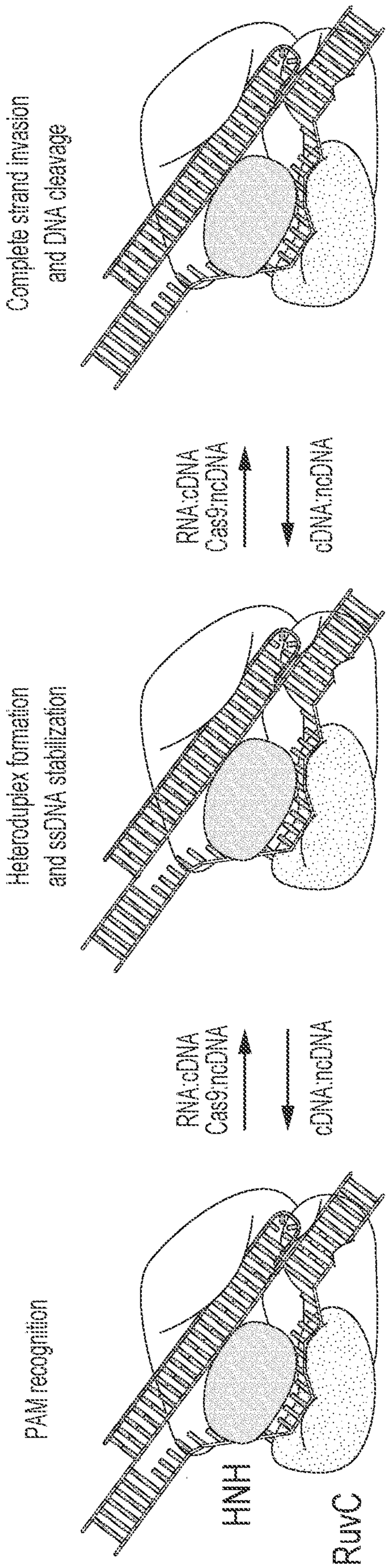

PAM recognition
HNH
RuvC
RNA:cDNA
Cas9:ncDNA
cDNA:ncDNA
Heteroduplex formation
and ssDNA stabilization
RNA:cDNA
Cas9:ncDNA
cDNA:ncDNA
Complete strand invasion
and DNA cleavage

Fig. 19

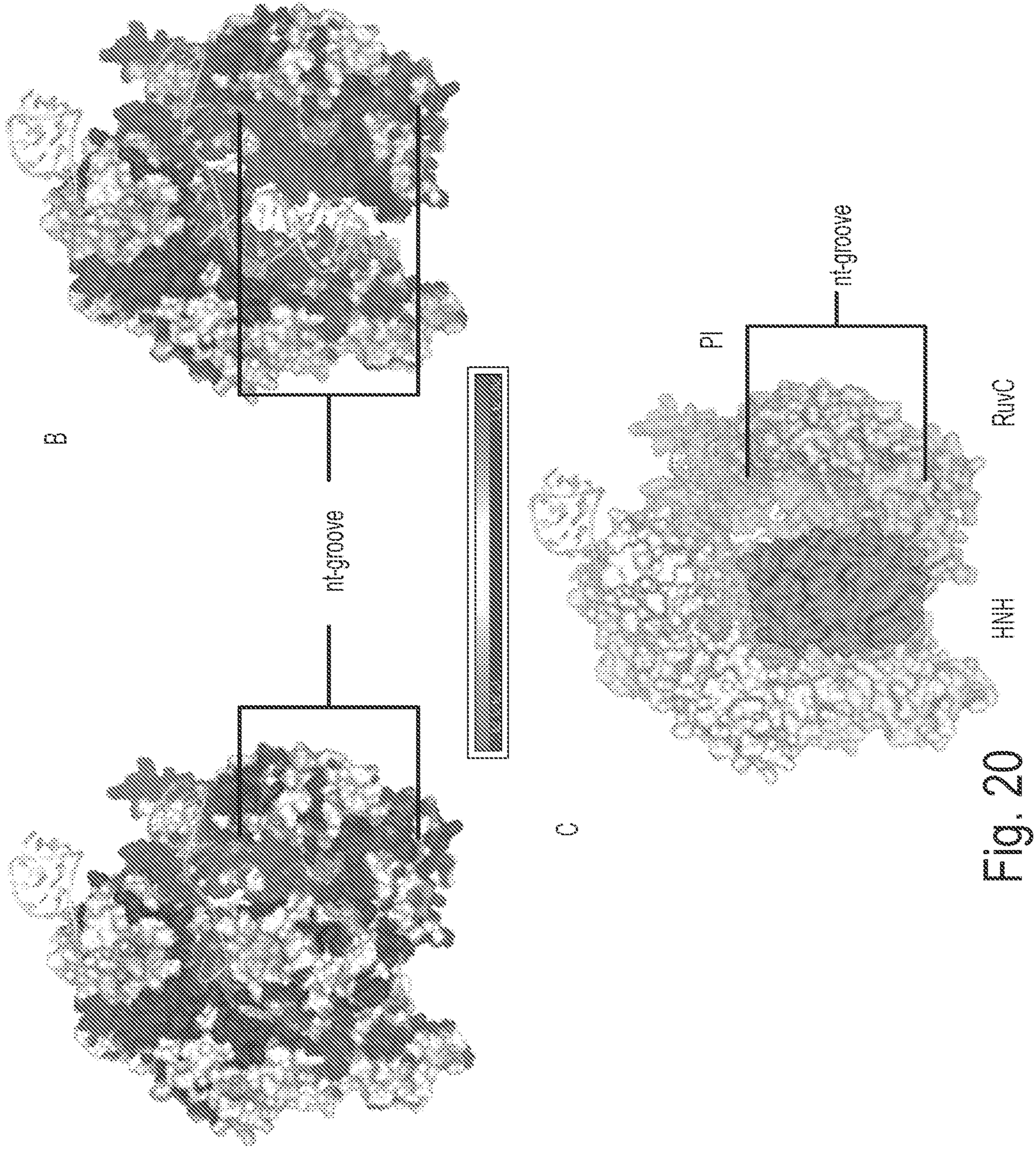


Fig. 20

A

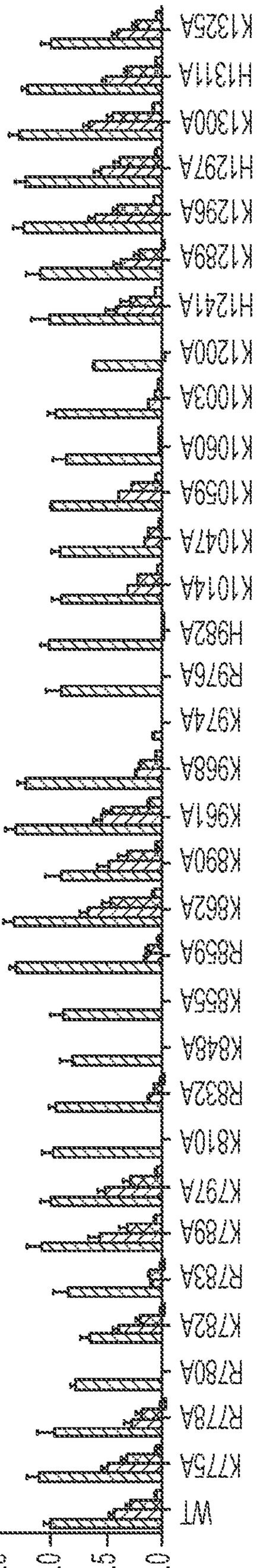

A
Normalized Indel
2.0
1.5
1.0
0.5
0.0
WT
K775A
R778A
R780A
K782A
K783A
K789A
K797A
K810A
R832A
K848A
K855A
R859A
K862A
K890A
K961A
K968A
K974A
R976A
H982A
K1014A
K1047A
K1059A
K1060A
K1003A
K1200A
H1241A
K1289A
K1296A
H1297A
K1300A
H1311A
K1325A
EMX1(1)
OT1
OT2
OT3
B
Normalized Indel
2.0
1.5
1.0
0.5
0.0
WT
K810A
K848A
K855A
R976A
H982A
K1059A
VEGFA(1)
OT1
OT2
C
Normalized Indel
2.0
1.5
1.0
0.5
0.0
WT
R780A
K810A
K848A
K855A
R976A
H982A
K1059A
R1060A
VEGFA(5)
OT1

Fig. 21A-21B

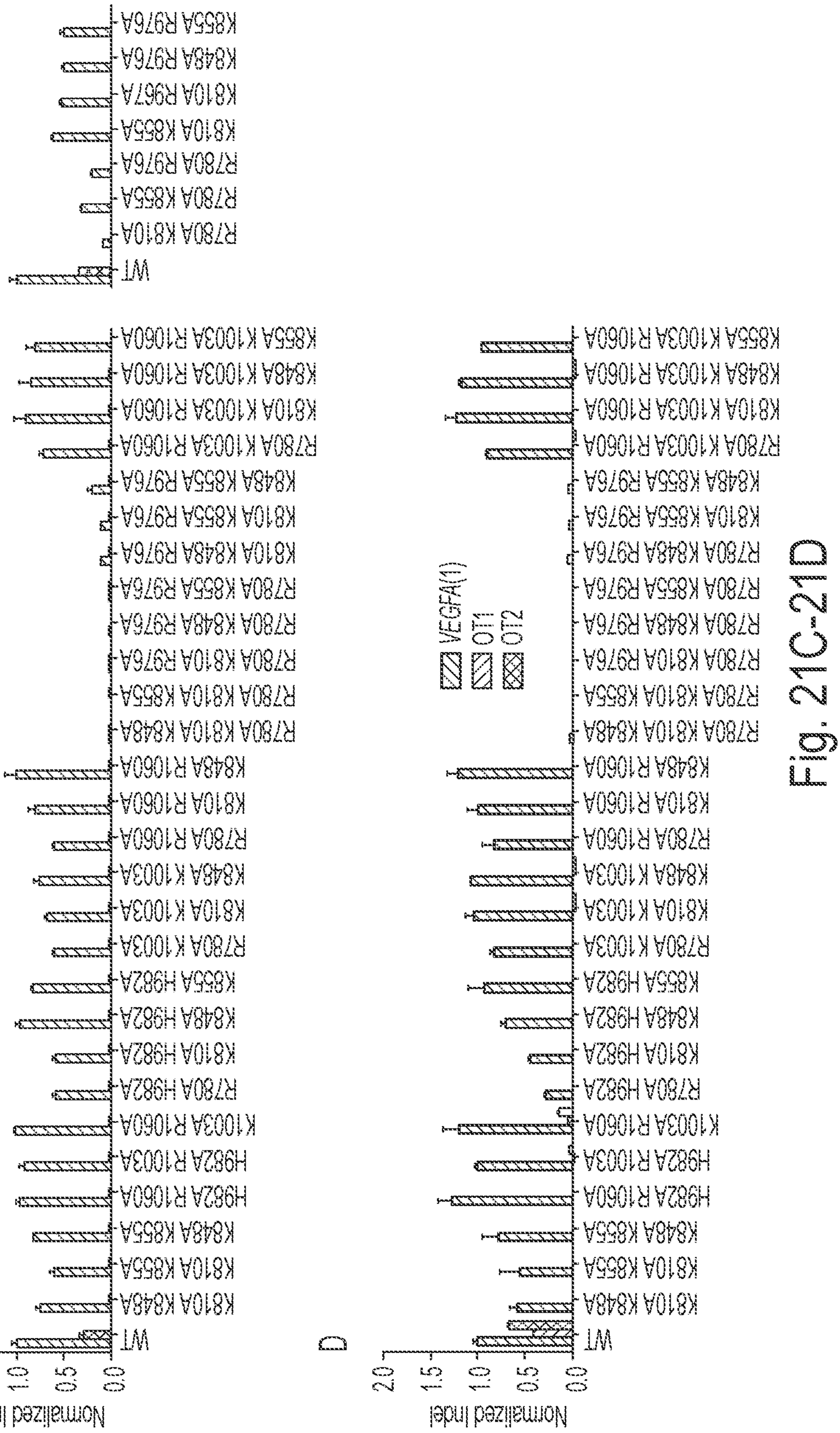

VEGFA(1)
OT1
OT2
K855A R976A
K848A R976A
K810A R967A
K810A K855A
R780A R976A
R780A K855A
R780A K810A
WT
K855A K1003A R1060A
K848A K1003A R1060A
K810A K1003A R1060A
R780A K1003A R1060A
K848A K855A R976A
K810A K855A R976A
K810A K848A R976A
R780A K855A R976A
R780A K848A R976A
R780A K810A R976A
R780A K810A K855A
R780A K810A K848A
K848A R1060A
K810A R1060A
R780A R1060A
K848A K1003A
K810A K1003A
R780A K1003A
K855A H982A
K848A H982A
K810A H982A
R780A H982A
K1003A R1060A
H982A R1003A
H982A R1060A
K848A K855A
K810A K855A
K810A K848A
WT
Normalized Indel
0.0
0.5
1.0
1.5
2.0
C
EMX1(1)
OT2
K855A K1003A R1060A
K848A K1003A R1060A
K810A K1003A R1060A
R780A K1003A R1060A
K848A K855A R976A
K810A K855A R976A
R780A K848A R976A
R780A K855A R976A
R780A K848A R976A
R780A K810A R976A
R780A K810A K855A
R780A K810A K848A
K848A R1060A
K810A R1060A
R780A R1060A
K848A K1003A
K810A K1003A
R780A K1003A
K855A H982A
K848A H982A
K810A H982A
R780A H982A
K1003A R1060A
H982A R1003A
H982A R1060A
K848A K855A
K810A K855A
K810A K848A
WT
Normalized Indel
0.0
0.5
1.0
1.5
2.0
D

Fig. 21C-21D

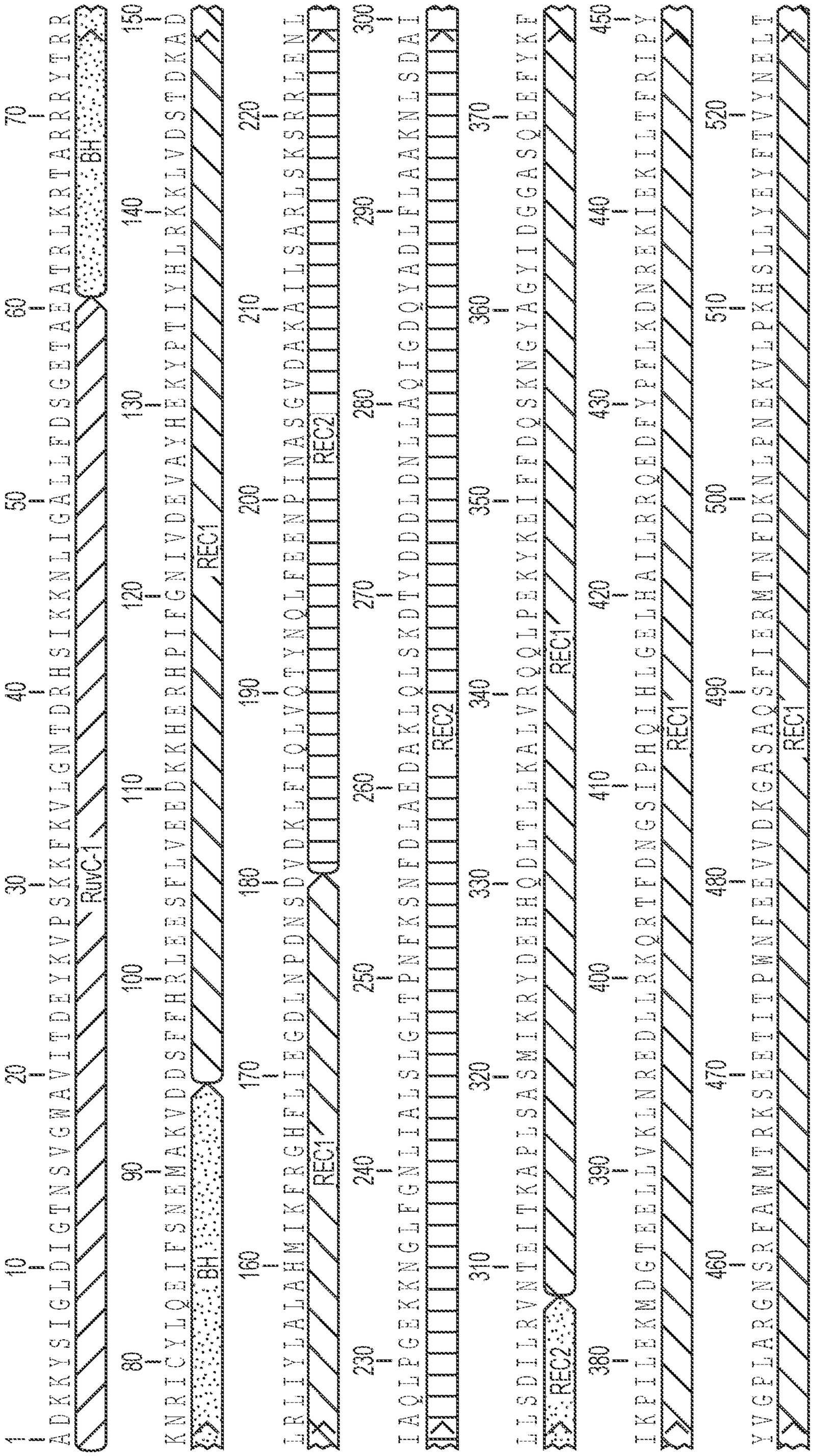

1 10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170 180 190 200 210 220 230 240 250 260 270 280 290 300 310 320 330 340 350 360 370 380 390 400 410 420 430 440 450 460 470 480 490 500 510 520
ADKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

Fig. 22A

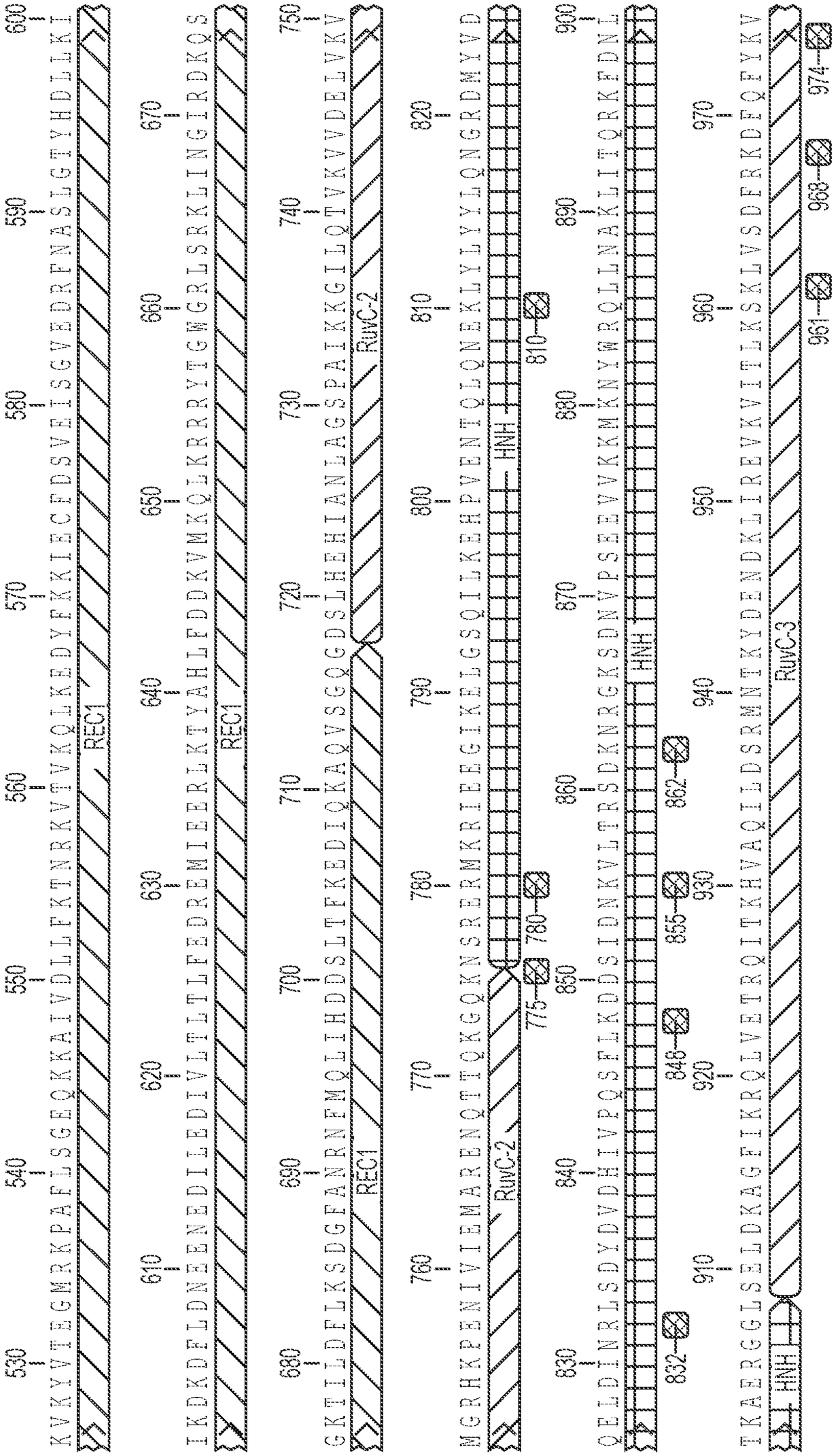


Fig. 22B

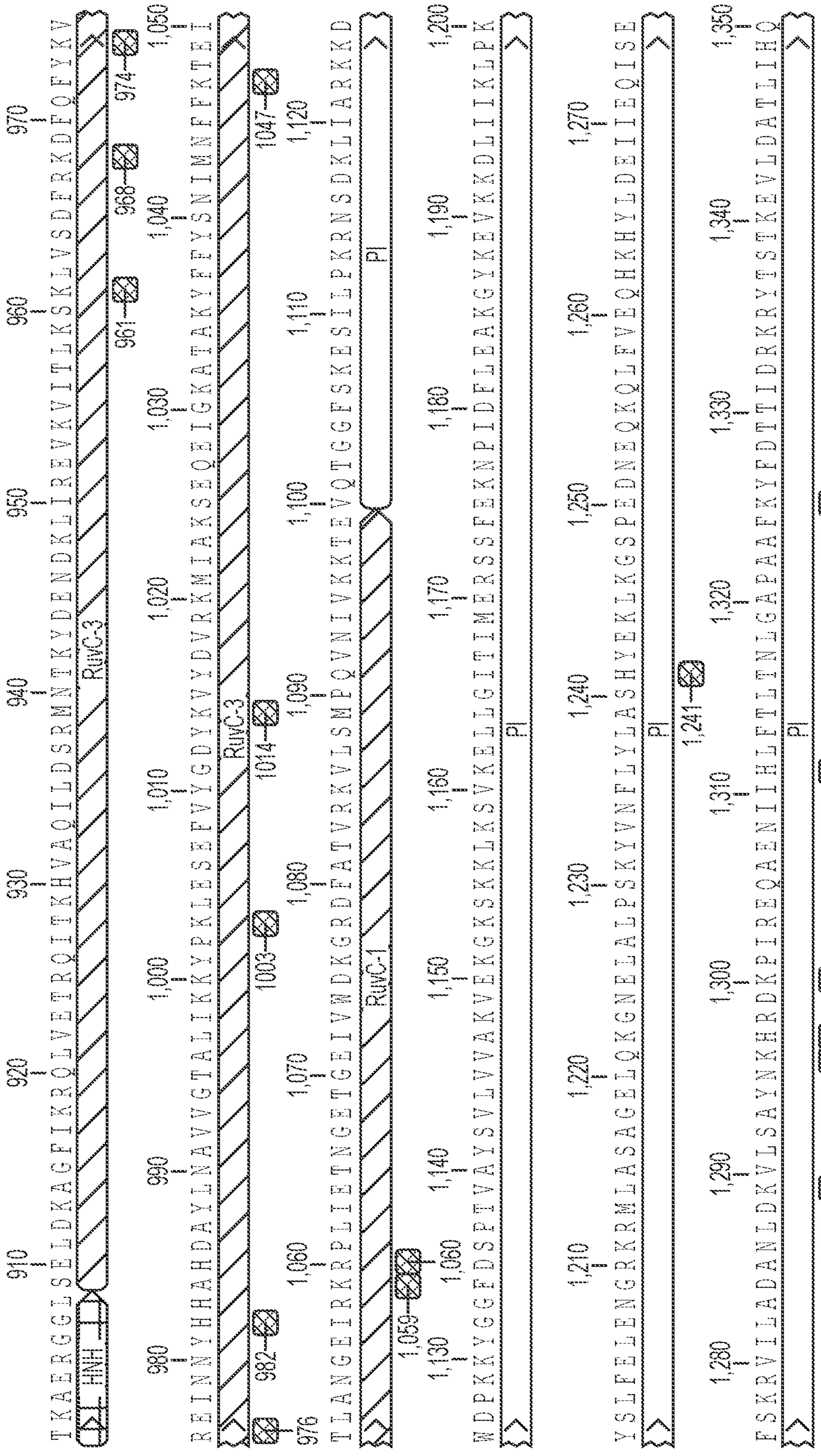
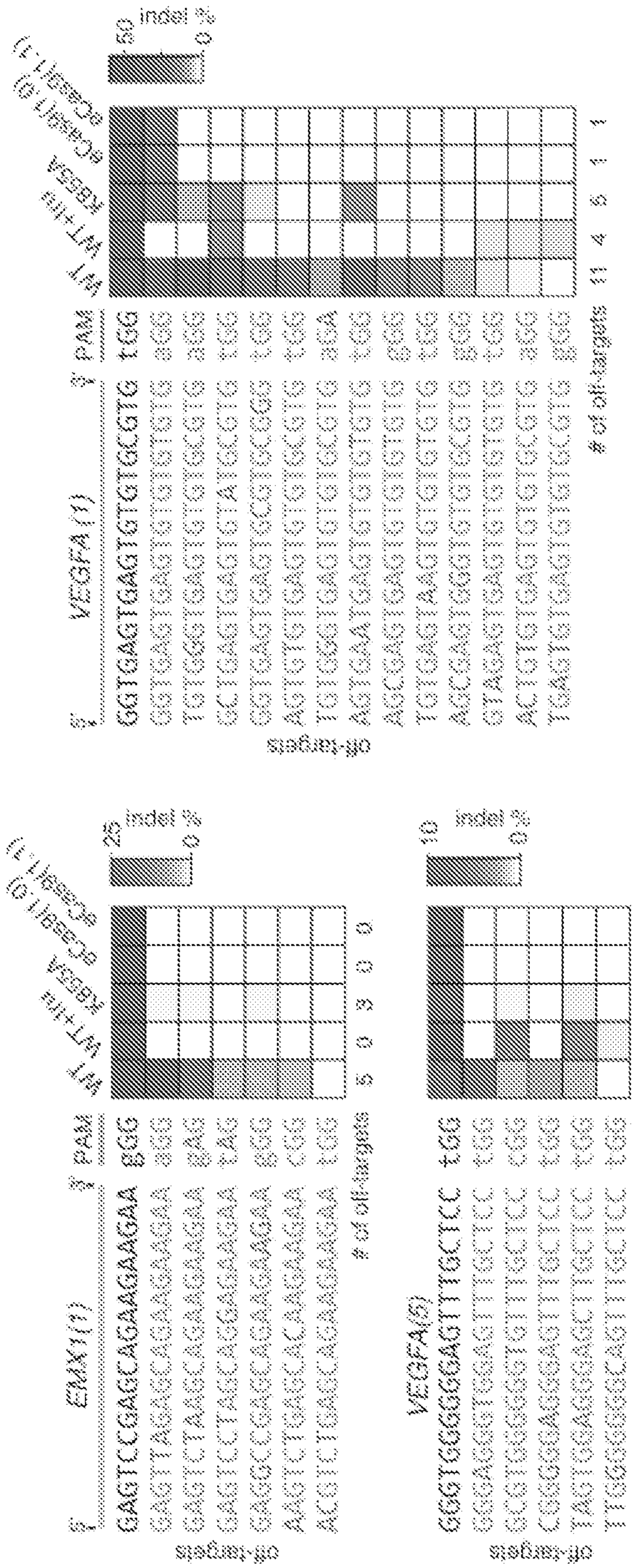

970
960
950
940
930
920
1,040
1,030
1,020
1,010
1,000
990
980
1,120
1,110
1,100
1,090
1,080
1,070
1,060
1,200
1,190
1,180
1,170
1,160
1,150
1,140
1,130
1,270
1,260
1,250
1,240
1,230
1,220
1,210
1,350
1,340
1,330
1,320
1,310
1,300
1,290
1,280
968
974
1,050
961
1047
1003
1014
1,241
1,296
1,297
1,311
1,325
1,368
1,360
1,269
RuvC-1
RuvC-3
HNH
PI

Fig. 22C


indel %
25
0
eGFP (U6-VEGFA (1))
eGFP (U6-VEGFA (2))
eGFP (U6-VEGFA (3))
eGFP (U6-VEGFA (4))
eGFP (U6-VEGFA (5))
eGFP (U6-VEGFA (6))
eGFP (U6-VEGFA (7))
eGFP (U6-VEGFA (8))
eGFP (U6-VEGFA (9))
eGFP (U6-VEGFA (10))
eGFP (U6-VEGFA (11))
eGFP (U6-VEGFA (12))
eGFP (U6-VEGFA (13))
eGFP (U6-VEGFA

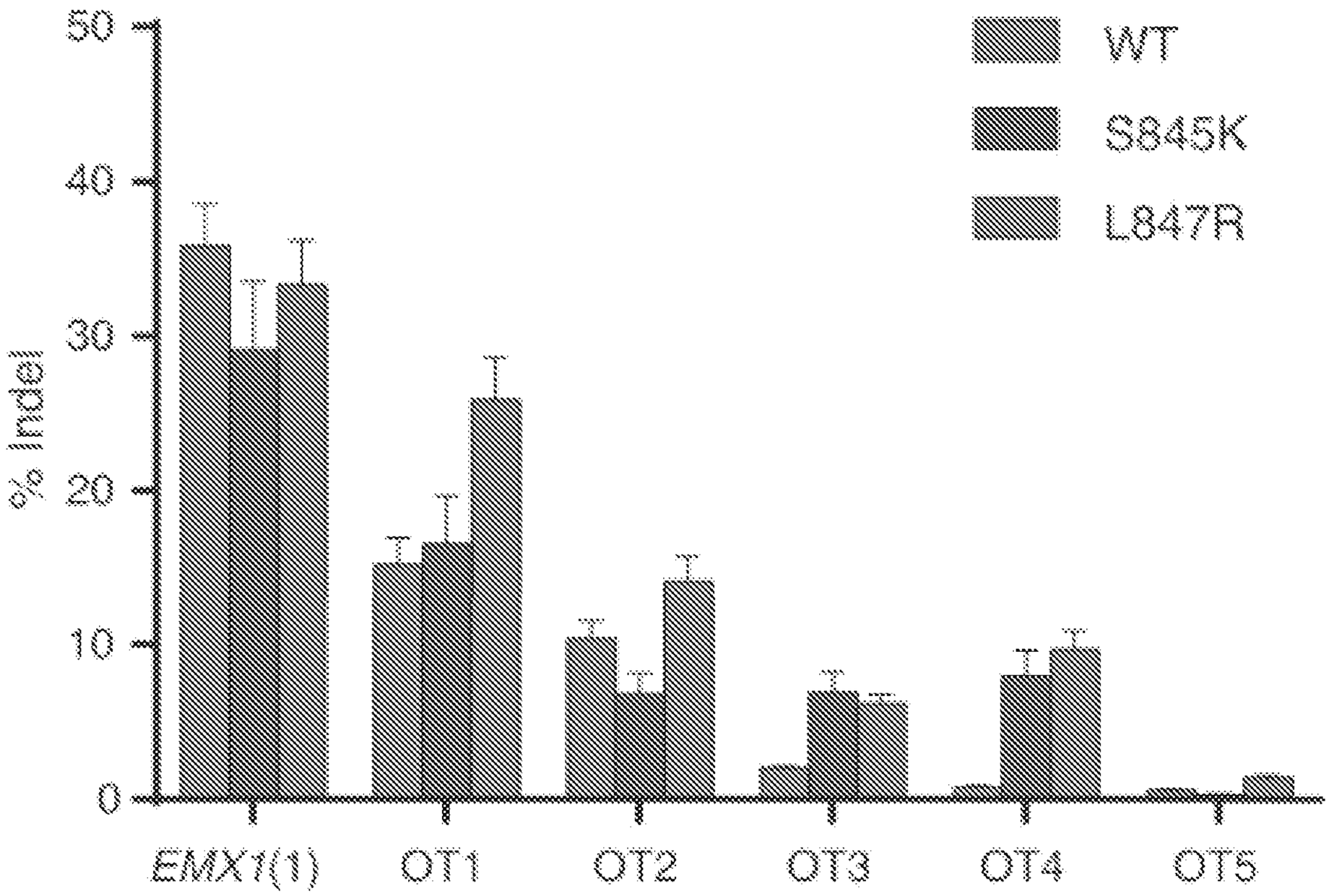


Fig. 24

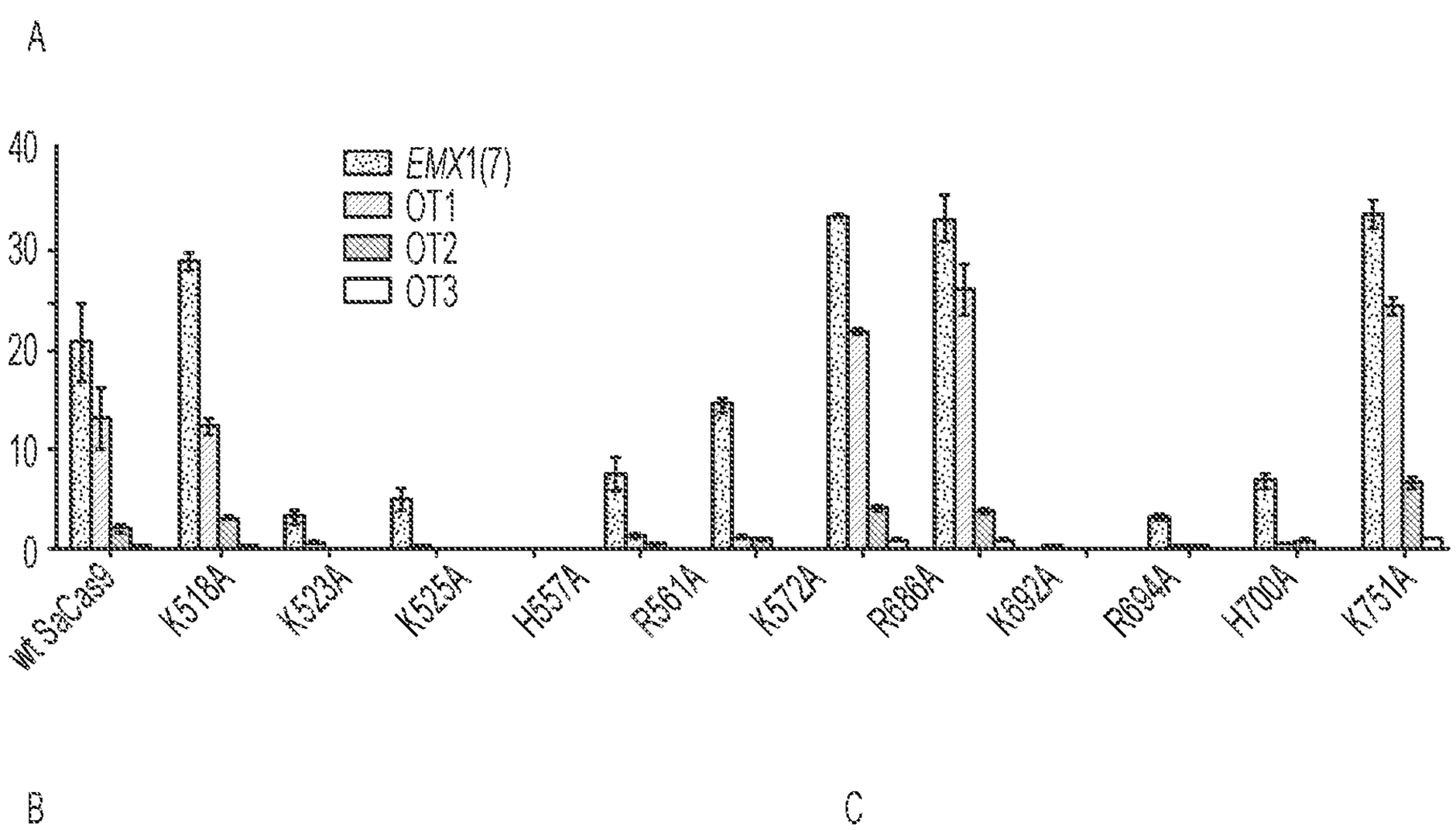

A
40
30
20
10
0
EMX1(7)
OT1
OT2
OT3
WT SaCas9
K518A
K523A
K525A
H557A
R561A
K572A
R686A
K692A
R694A
H700A
K751A
B
% indel
0
10
20
30
40
50
60
wt SaCas9
R497A
R499A
Q500K
R634A
R654A
G655R
C
% indel
0
5
10
15
20
25
30
35
40
wt SaCas9
R499A
Q500K
R499A
Q500A
R654A
G655A
K572A
K572A
R654A
G655R

Fig. 25A-25C

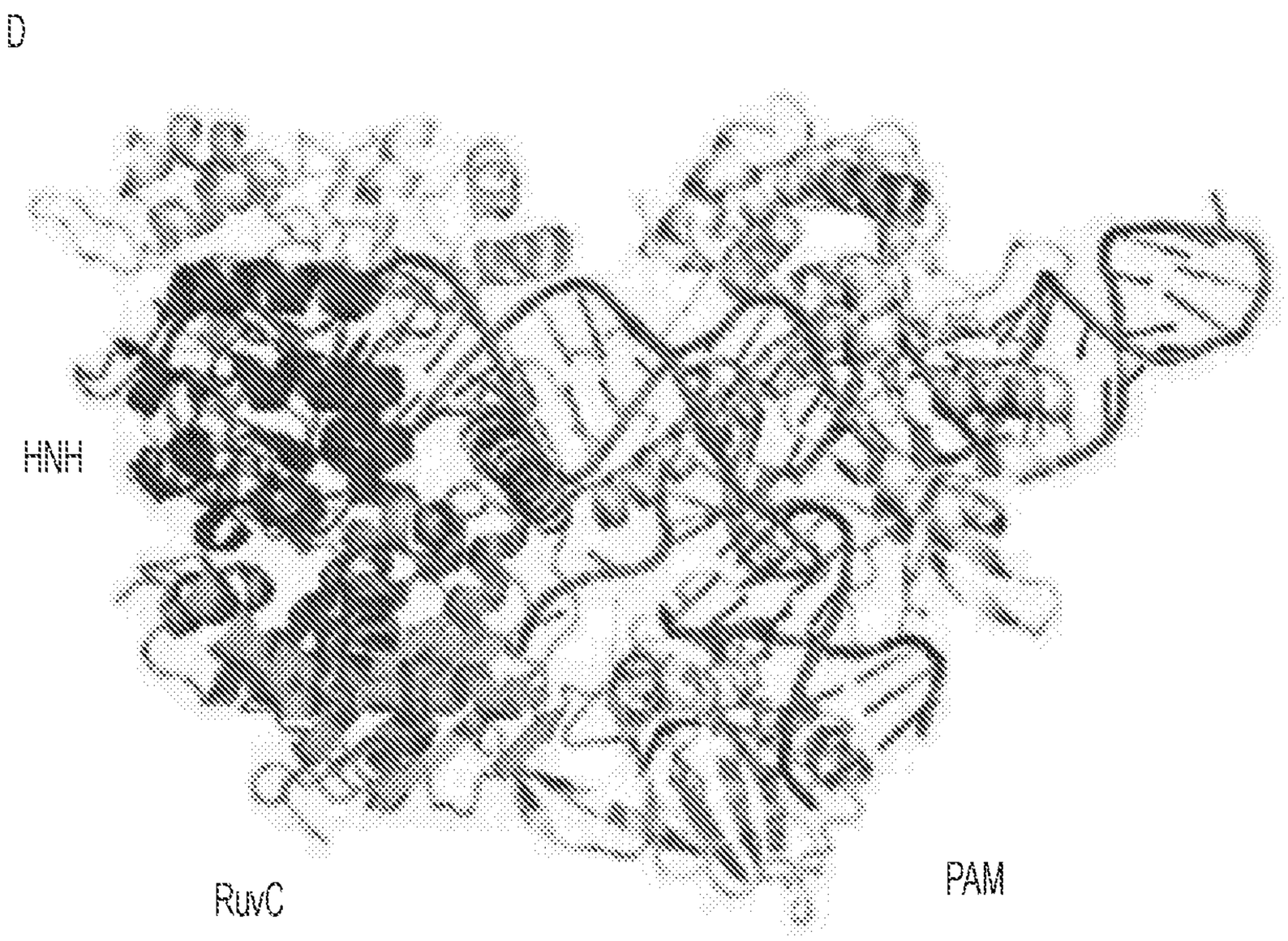


Fig. 25D

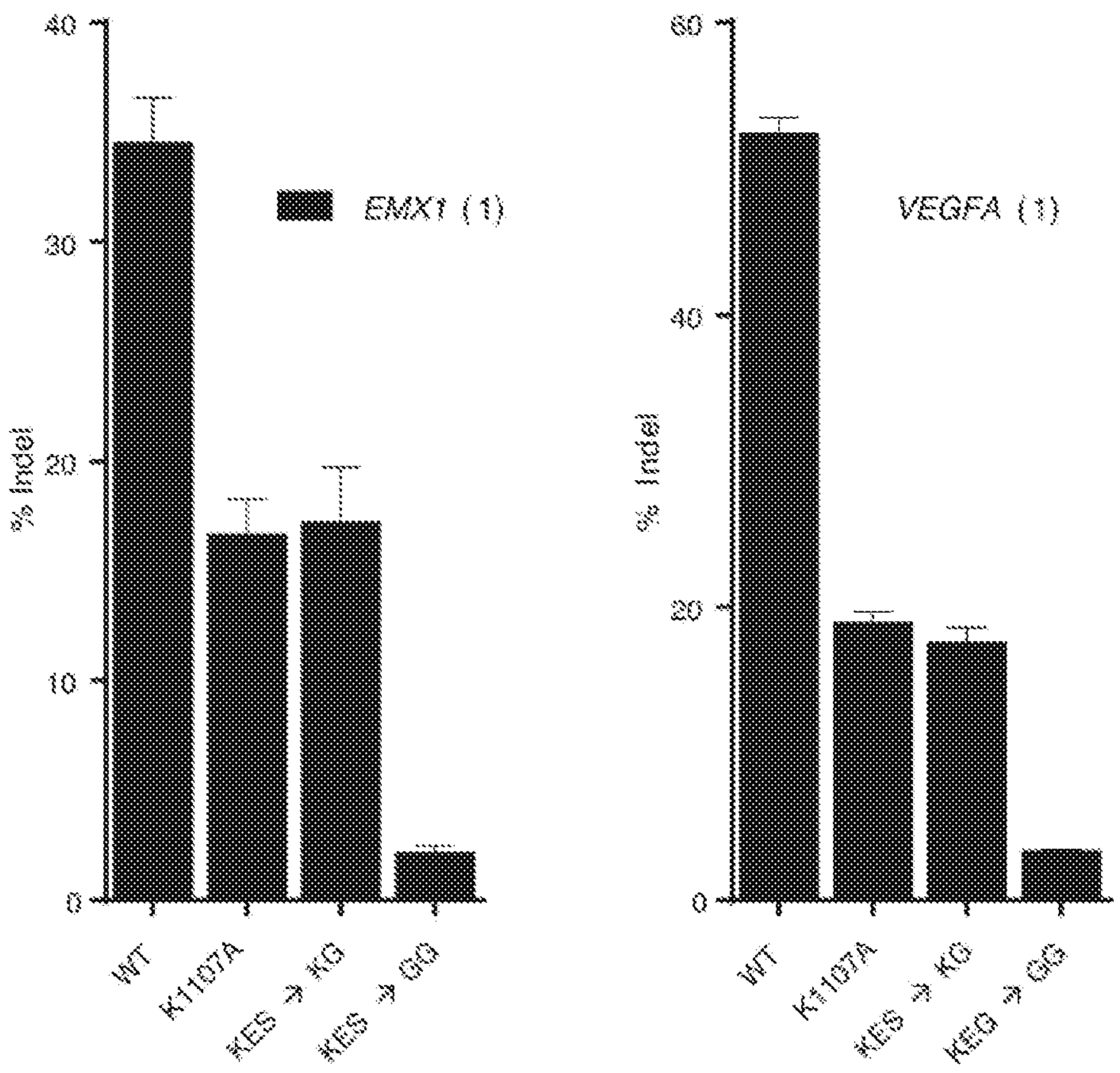

40
30
20
10
0
% Indel
EMX1 (1)
WT
K1107A
KES → KG
KES → GG
60
40
20
0
% Indel
VEGFA (1)
WT
K1107A
KES → KG
KES → GG

Fig. 26

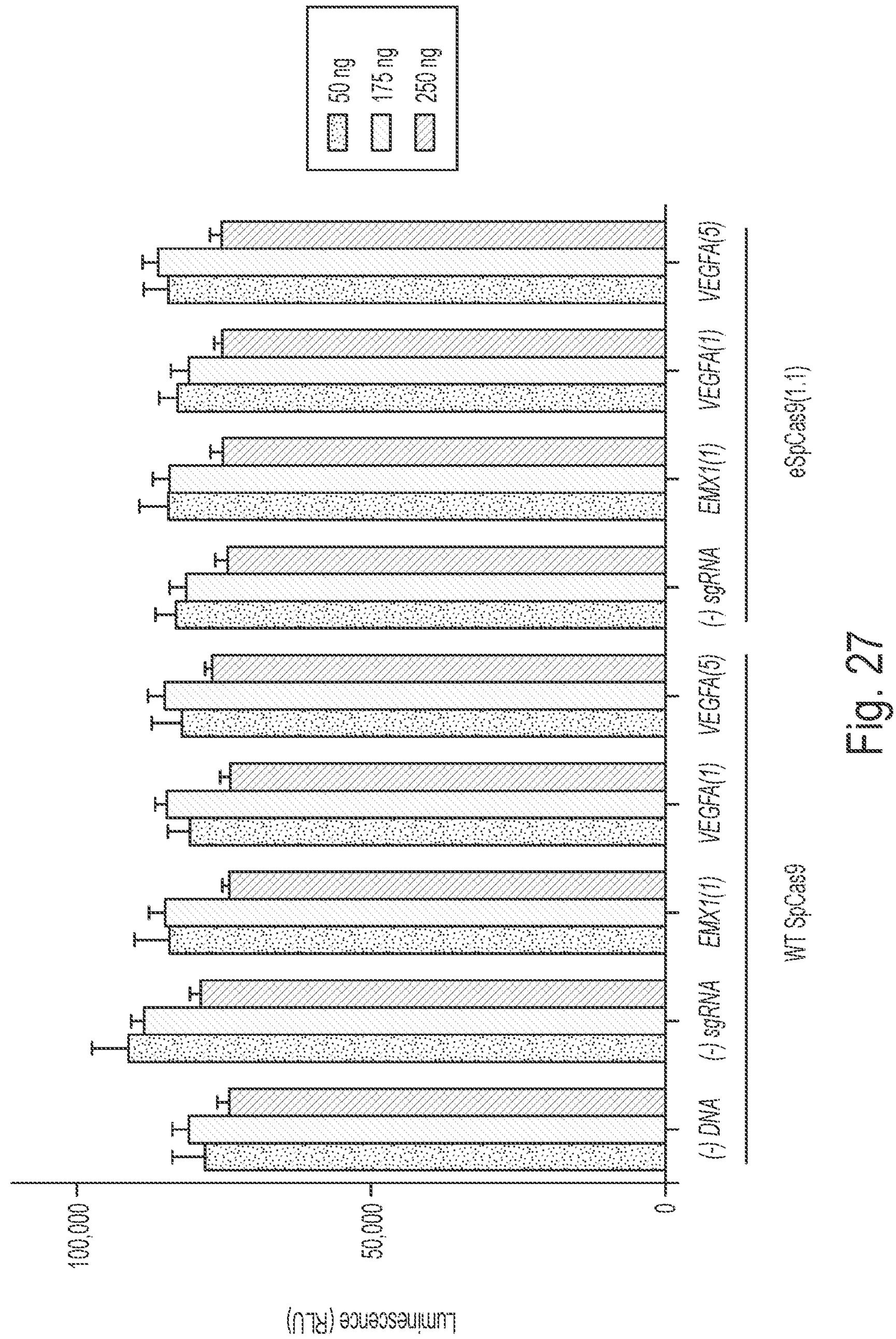

50 ng
175 ng
250 ng
(-) DNA
(-) sgRNA
EMX1(1)
VEGFA(1)
VEGFA(5)
WT SpCas9
eSpCas9(1.1)
(-) DNA
(-) sgRNA
EMX1(1)
VEGFA(1)
VEGFA(5)
100,000
50,000
0
Luminescence (RLU)

Fig. 27

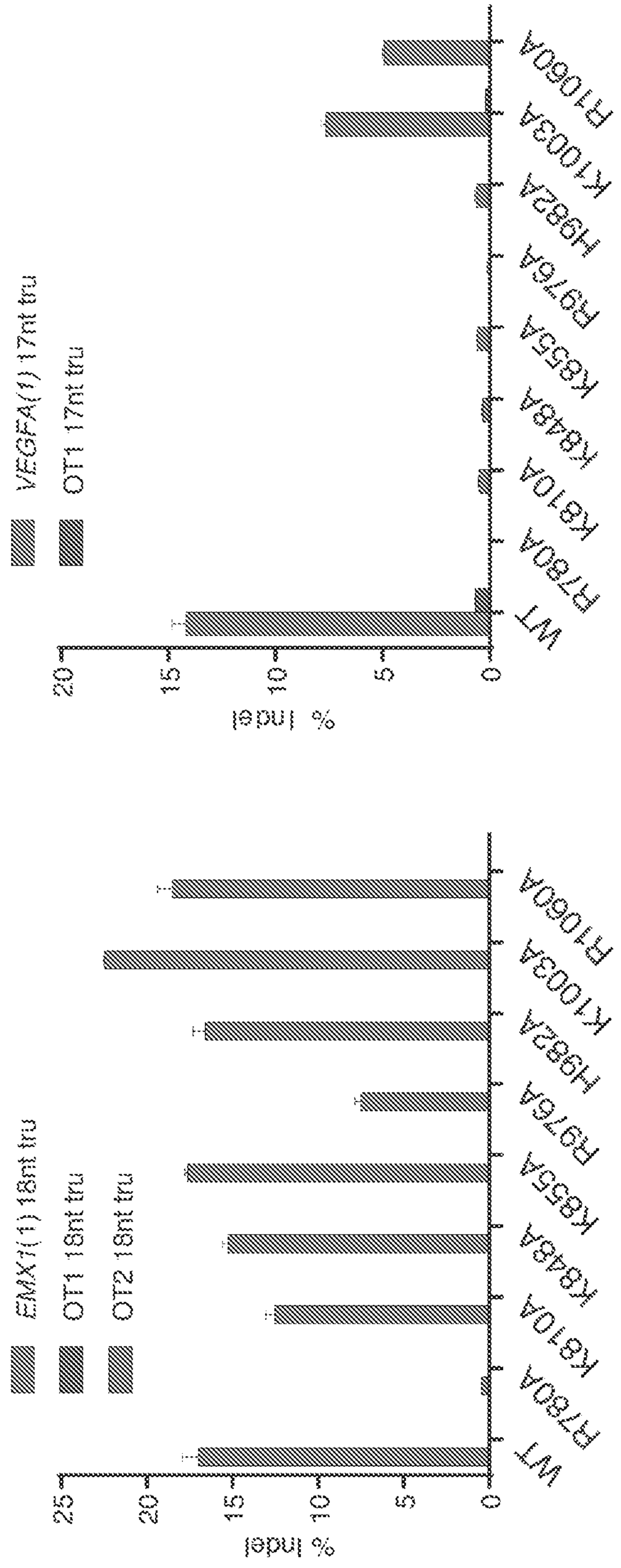

VEGFA(1) 17nt trru
OT1 17nt trru
% indel
0
5
10
15
20
WT
R780A
K810A
K848A
K855A
R976A
H982A
K1003A
R1060A
EMX1(1) 18nt trru
OT1 18nt trru
OT2 18nt trru
% indel
0
5
10
15
20
25
WT
R780A
K810A
K848A
K855A
R976A
H982A
K1003A
R1060A

Fig. 28

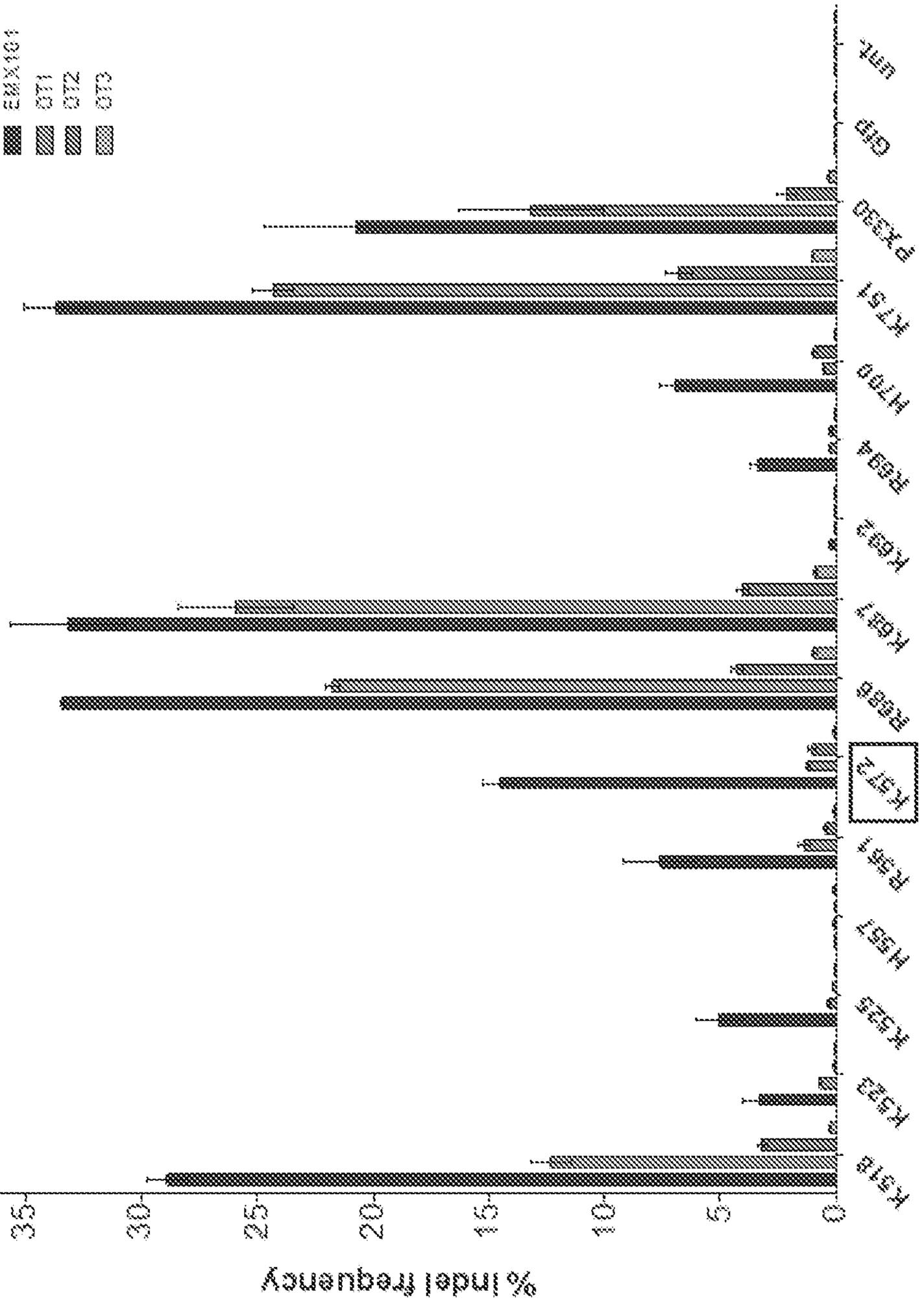

EMX1041
EMX1
EMX2
EMX3
WT
EMX3
P433Q
K434
M749
R694
K695
K523
K525
M551
P552
K512
R553
K522
R518
% indel frequency
35
30
25
20
15
10
5
0

Fig. 29A

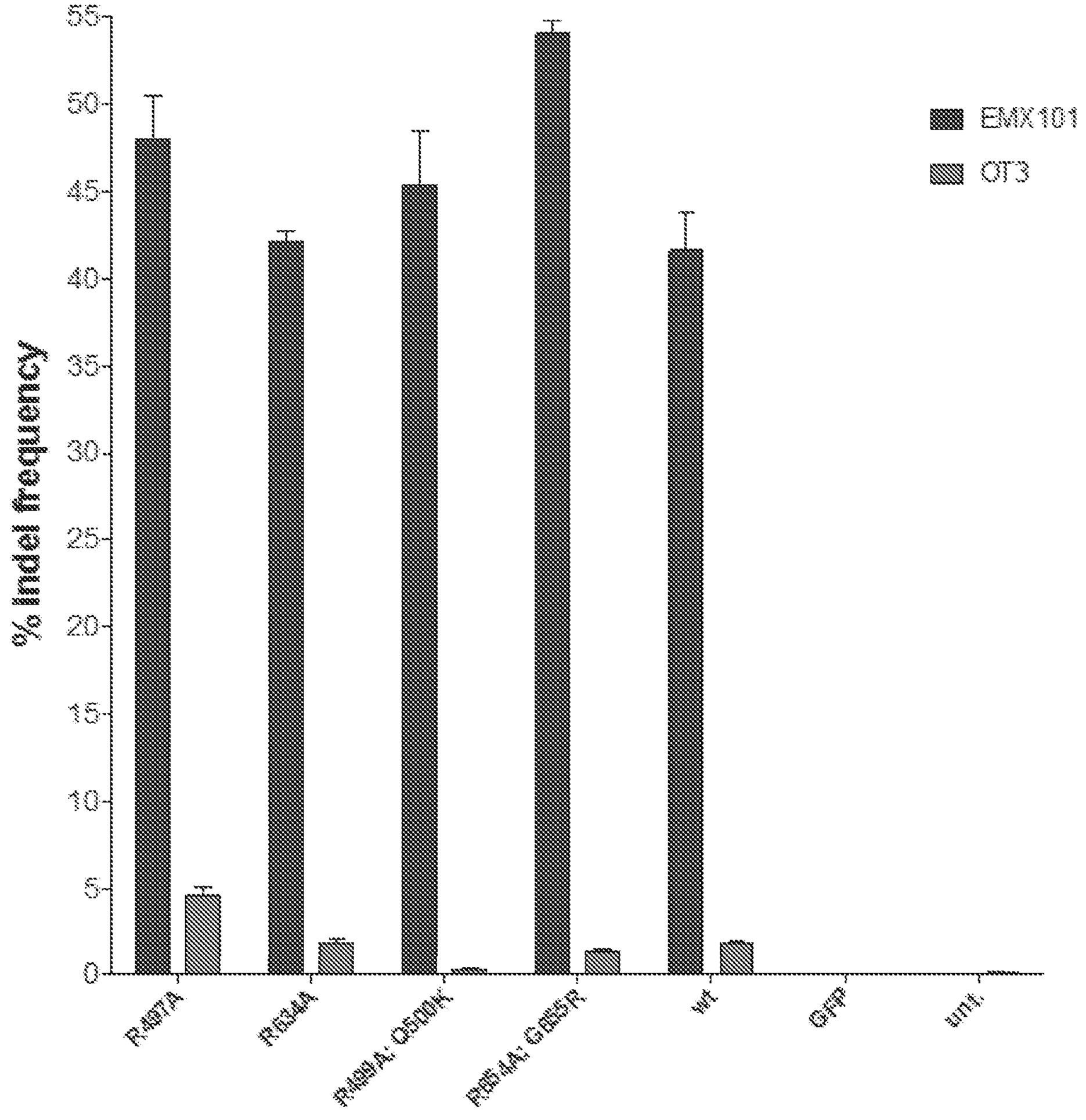

% indel frequency
0
5
10
15
20
25
30
35
40
45
50
55
EMX101
OT3
R:497A
R:634A
R:498A; Q500K
R:654A; G656R
wt
GFP
UTC

Fig. 29B

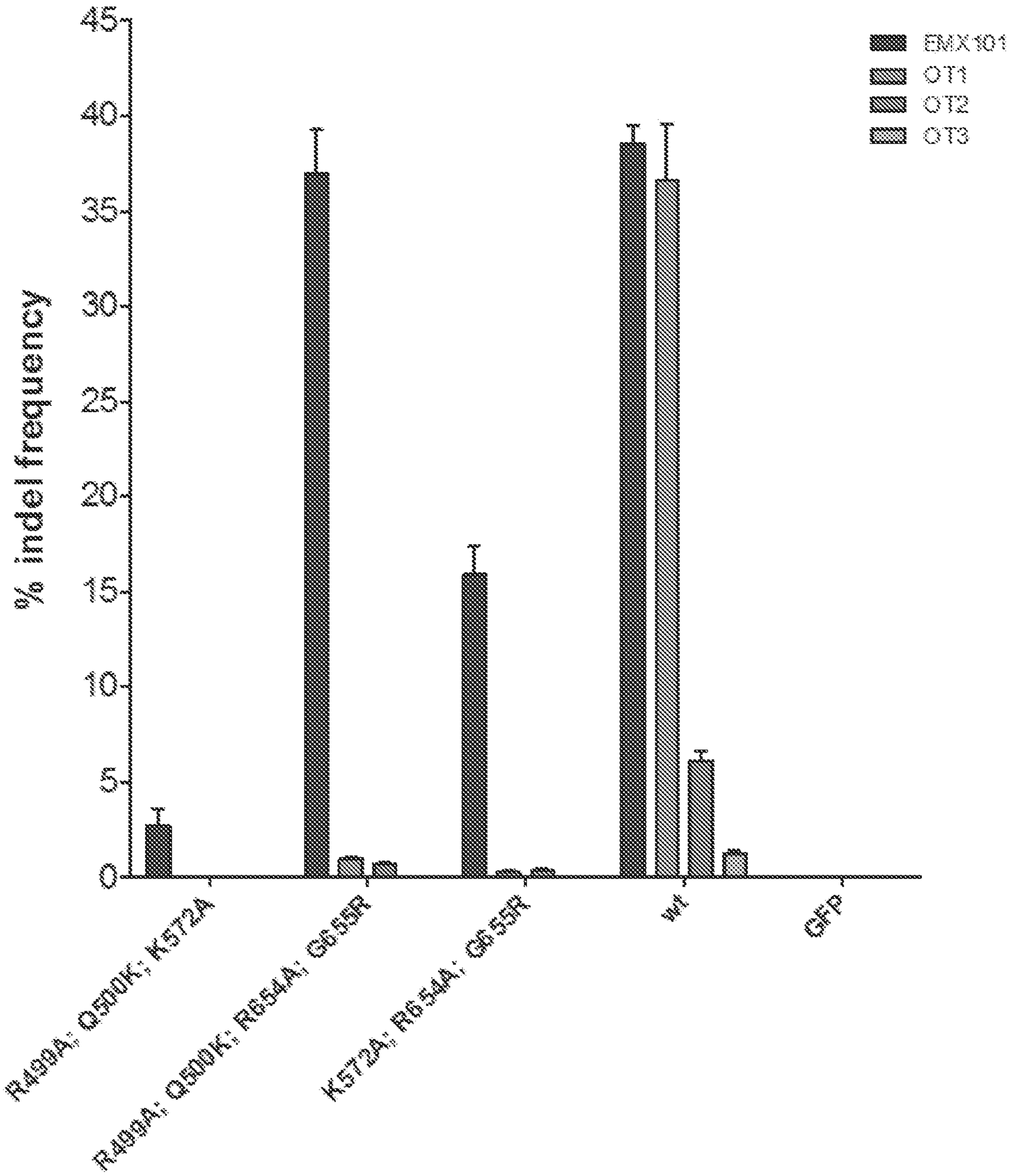

% indel frequency
0
5
10
15
20
25
30
35
40
45
EMX101
OT1
OT2
OT3
R490A; Q500K; K572A
Q500K; R654A; G655R
K572A; R654A; G655R
wt
GFP

Fig. 30

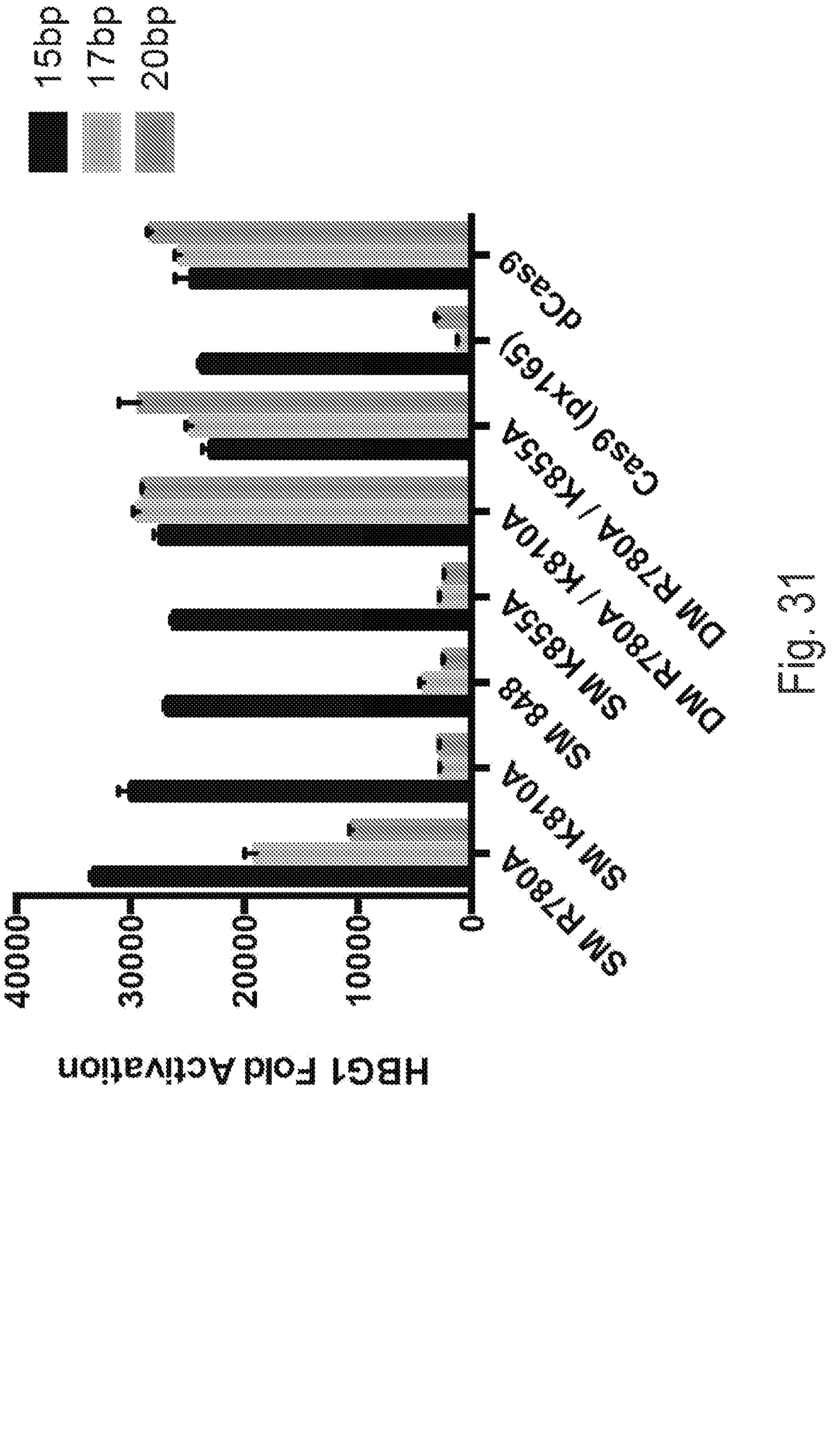

15
17
20
dCas9
Cas9 (pX165)
SM K855A / K810A
SM R780A / K810A
DM R780A / K810A
DM R780A
SM K855A
SM 848
SM K810A
SM R780A
0
10000
20000
30000
40000
1
2
3
4
HBG1 Fold Activation

Fig. 31

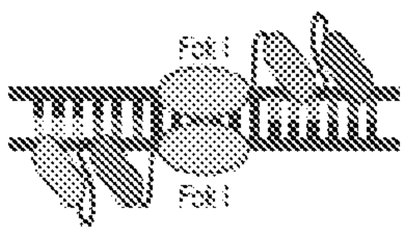
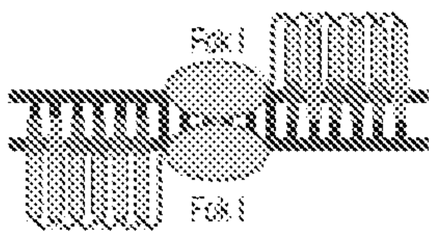
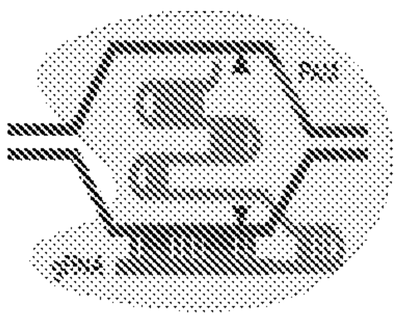
	Zinc Finger Nucleases	TALE Nucleases	CRISPR-Cas9
			
Recognition site	Typically 3 to 18 bp per ZFP; 18 to 36 bp per ZFN pair	Typically 14 to 20 bp per TALE; 28 to 40bp per TALEN	22bp (20bp guide sequence + 2bp PAM sequence); up to 44 bp for double nicking
Specificity	Small number of positional mismatches tolerated	Small number of positional mismatches tolerated	Positional and multiple consecutive mismatches tolerated
Targeting constraints	Difficult to target non-G-rich sequences	5' targeted base must be a T	Targeted sequence must precede a PAM
Ease of engineering	Difficult, may require substantial protein engineering	Moderate, requires complex molecular cloning methods	Easily re-targeted using standard cloning procedures and oligo synthesis
Immunogenicity	Likely low, as ZFs are based on human transcription factor protein scaffold	Unknown, protein derived from the bacteria Xanthomonas sp.	Unknown, protein derived from various bacterial species
Ease of *ex vivo* delivery	Relatively easy through methods such as electroporation and lentiviral transduction	Relatively easy through methods such as electroporation and lentiviral transduction	Relatively easy through methods such as electroporation and lentiviral transduction
Ease of *in vivo* delivery	Relatively easy due to small size of ZFN expression cassettes, allows use in a variety of viral vectors	Difficult due to the large size of each TALEN and repetitive nature of DNA encoding TALENs, leading to unwanted recombination events	Moderate. The commonly used Cas9 from *S. pyogenes* may be large enough to impose delivery problems, but other, smaller orthologoues exist
Ease of multiplexing	Low	Low	High

Fig. 32

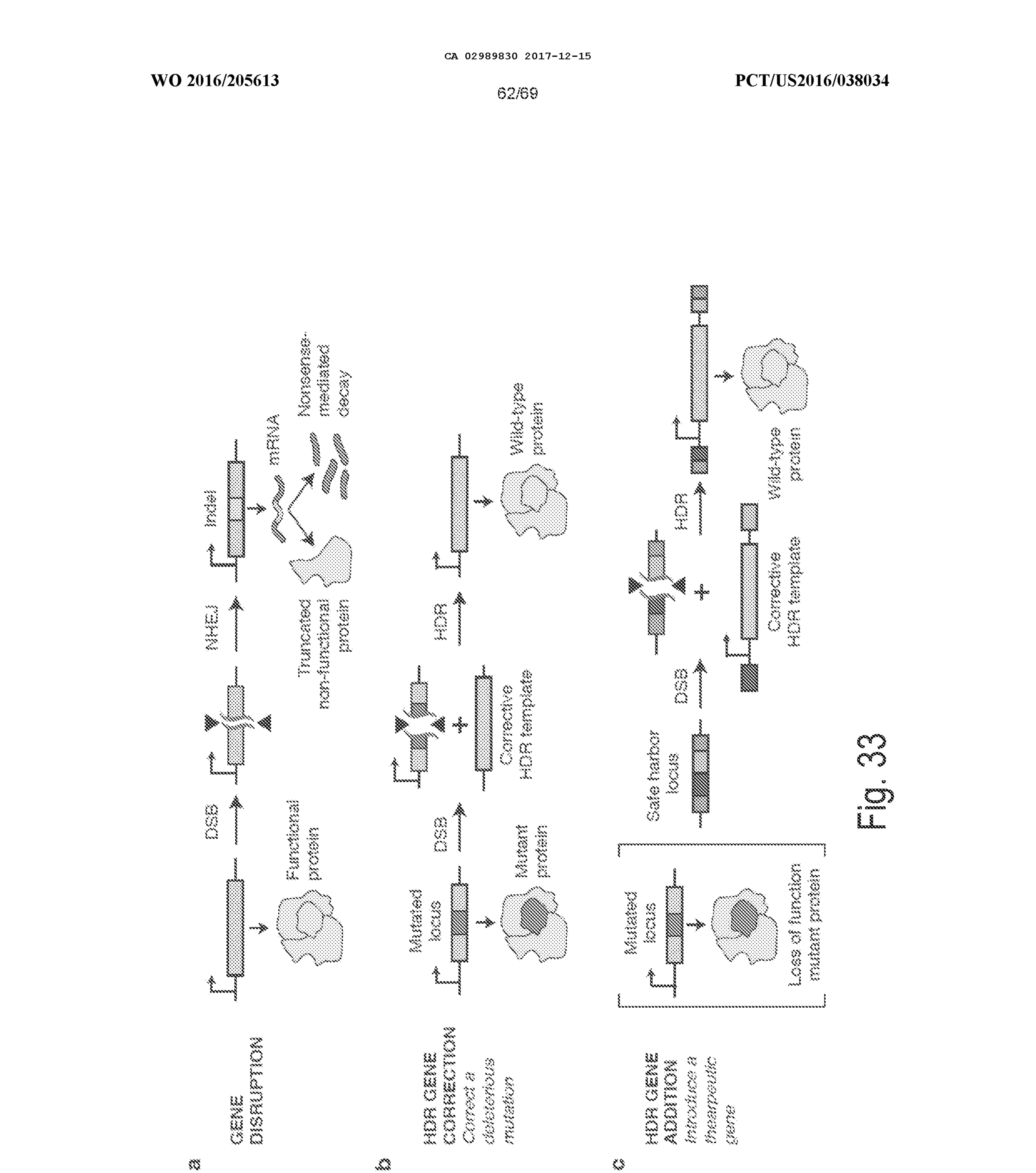

Gene disruption
DSB
Functional protein
Gene correction
Correct a deleterious mutation
Mutated locus
DSB
Mutant protein
HDR
Corrective HDR template
Wild-type protein
NHEJ
Indel
mRNA
Truncated non-functional protein
Nonsense-mediated decay
HDR gene addition
Introduce a therapeutic gene
Safe harbor locus
DSB
HDR
Corrective HDR template
Wild-type protein
Mutated locus
Loss of function mutant protein

Fig. 33

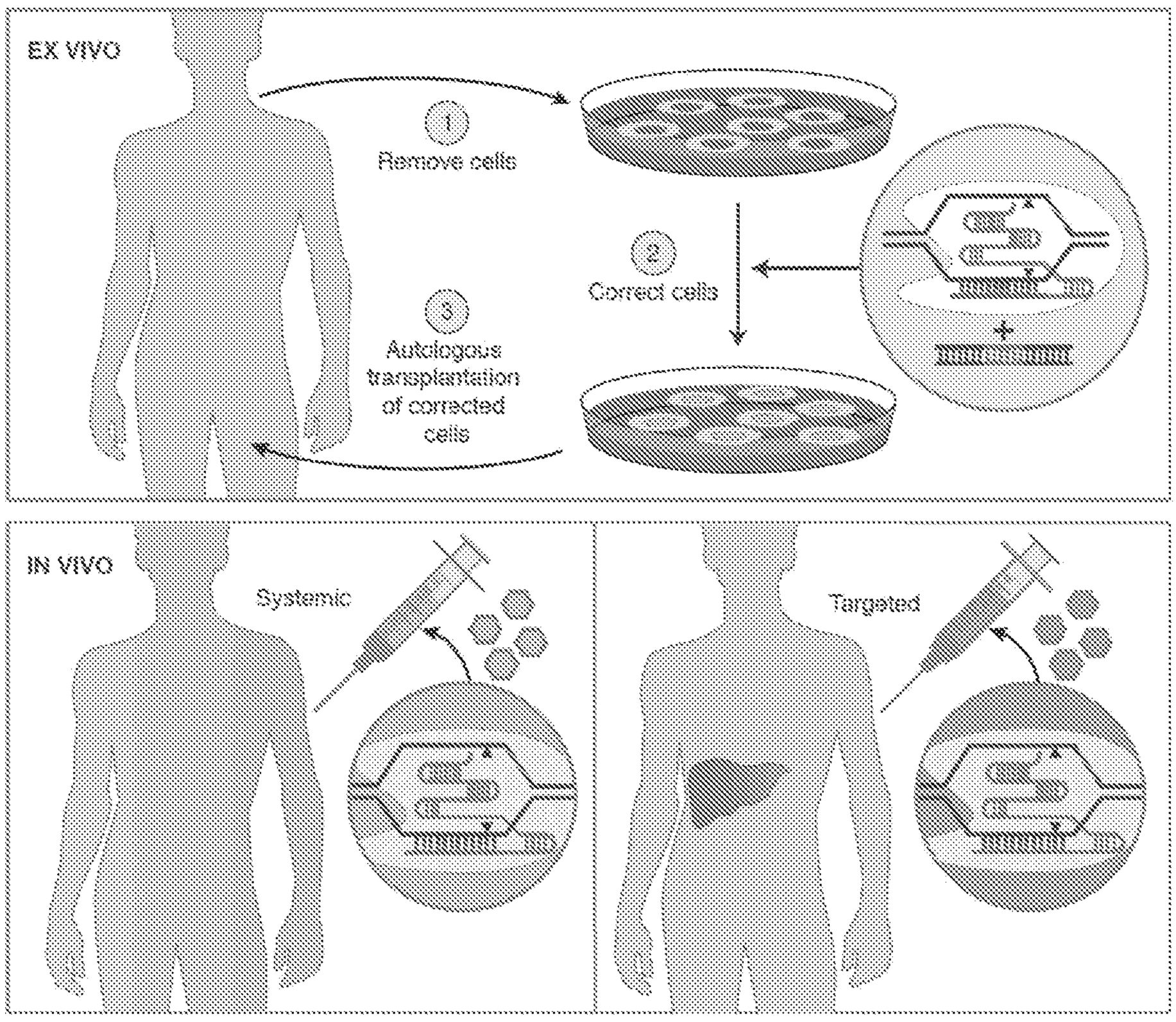

EX VIVO
1
Remove cells
2
Correct cells
3
Autologous transplantation of corrected cells
IN VIVO
Systemic
Targeted

Fig. 34

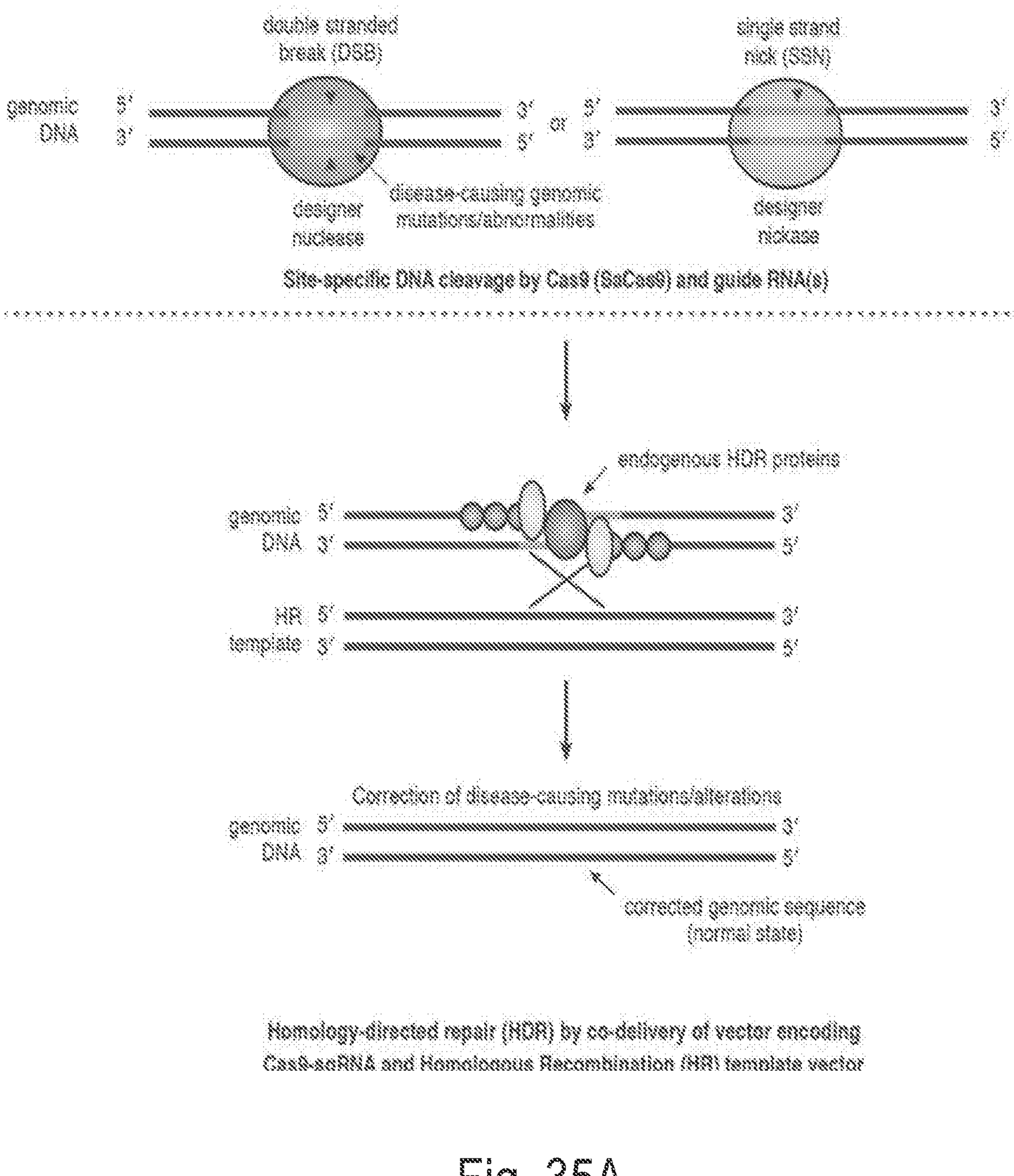

double stranded
break (DSB)
single strand
nick (SSN)
genomic
DNA
5'
3'
or
designer
nuclease
disease-causing genomic
mutations/abnormalities
designer
nickase
Site-specific DNA cleavage by Cas9 (SaCas9) and guide RNA(s)
endogenous HDR proteins
HR
template
Correction of disease-causing mutations/alterations
corrected genomic sequence
(normal state)
Homology-directed repair (HDR) by co-delivery of vector encoding
Cas9-sgRNA and Homologous Recombination (HR) template vector

Fig. 35A

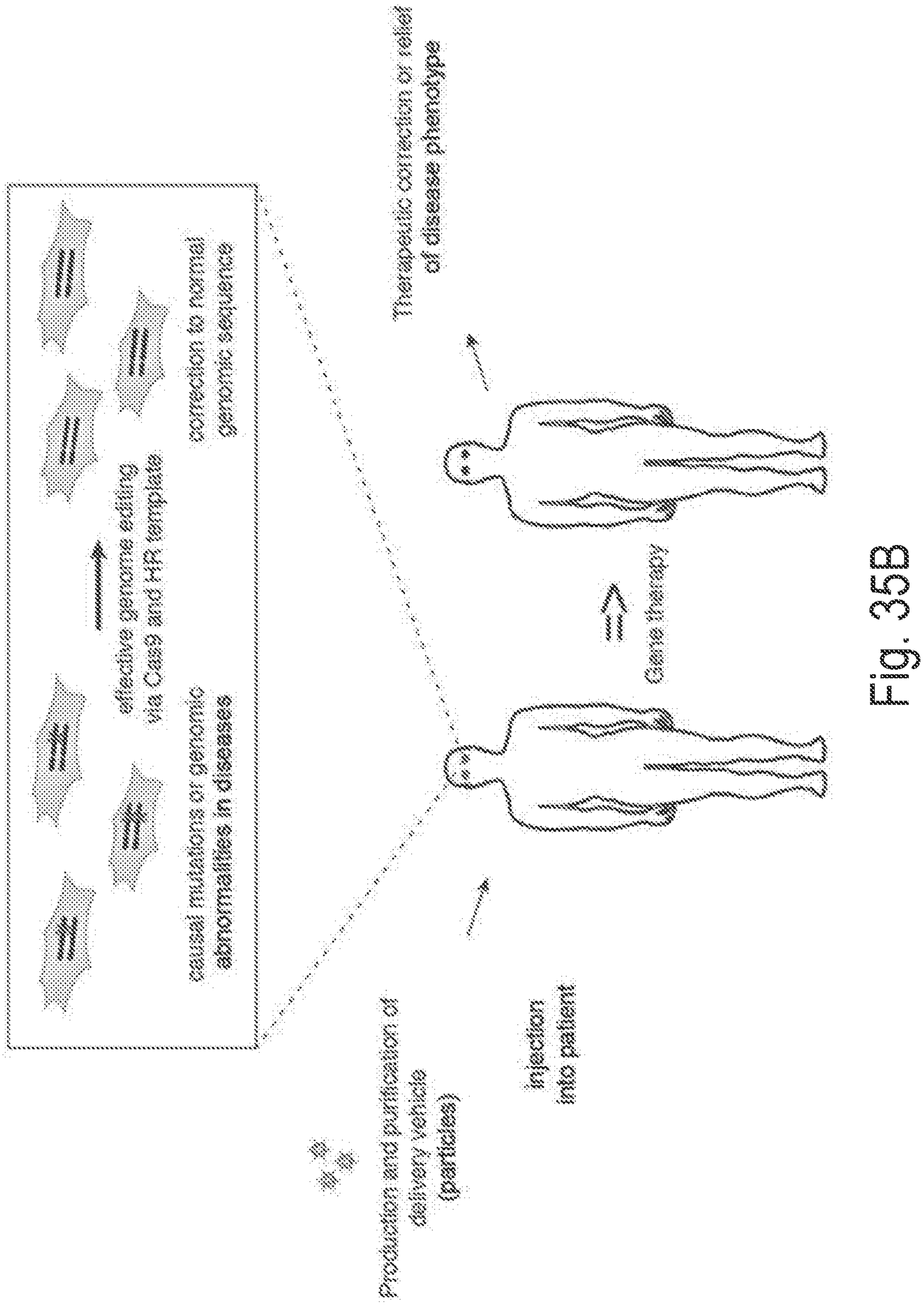

Therapeutic correction of rebalancing of disease phenotype
Genomic sequence correction to normal
Causal mutations in genomic abnormalities in diseases
Corrective gene and HA template
Cure via Cas
Therapeutic gene therapy
Injection into patient
Correction and purification of delivery vehicle (particles)

Fig. 35B

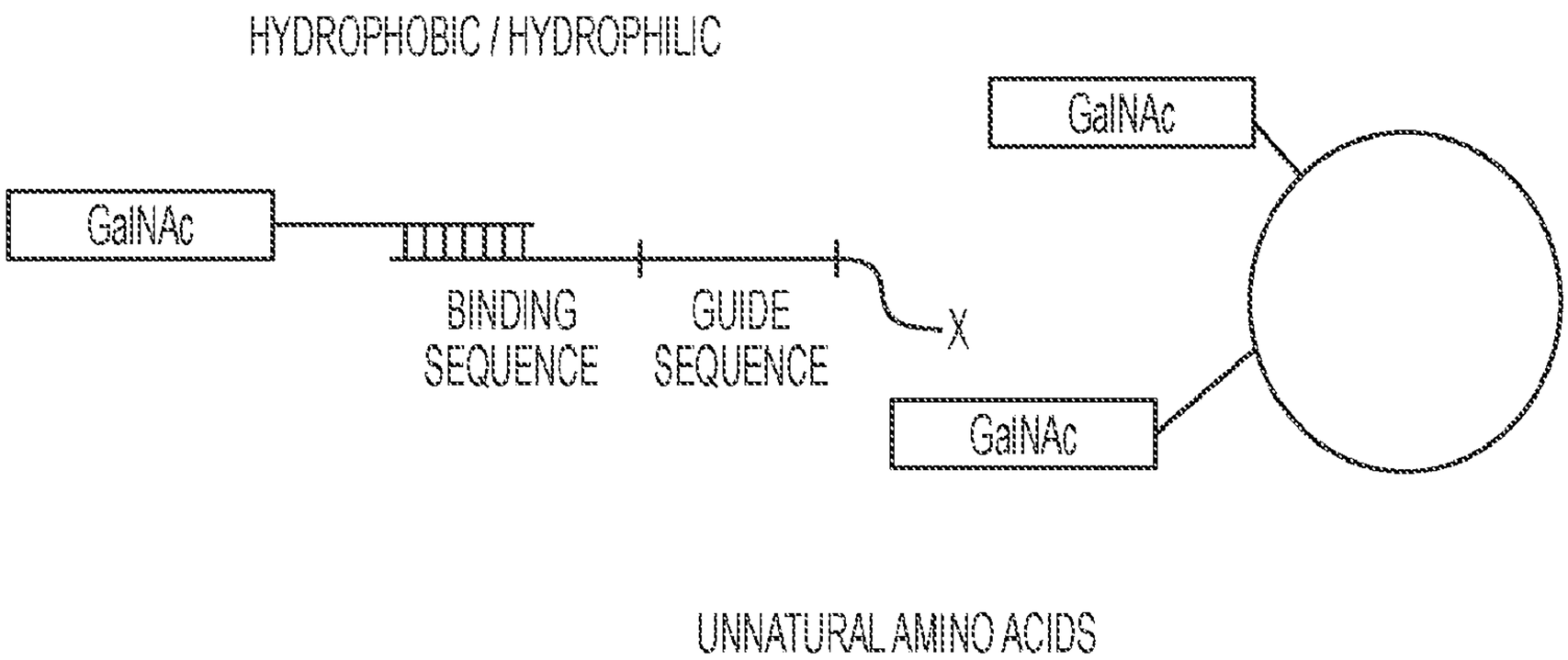

HYDROPHOBIC / HYDROPHILIC
GalNAc
GalNAc
BINDING SEQUENCE
GUIDE SEQUENCE
X
GalNAc
UNNATURAL AMINO ACIDS
GalNAc
ALKYNE
eCas9
ALKYNE
GalNAc
"CLICKABLE"
PEG
GalNAc

Fig. 36

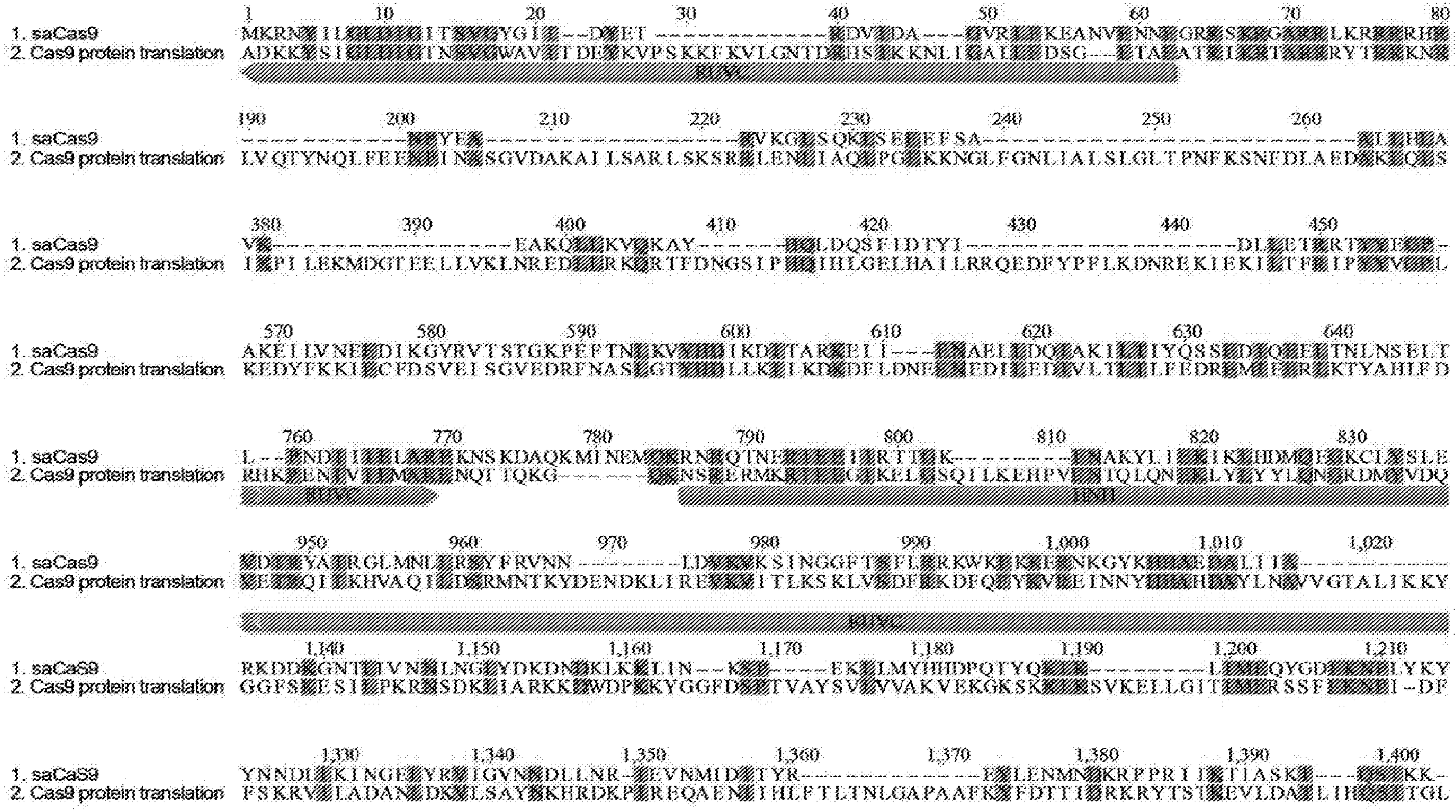

1. saCas9
2. Cas9 protein translation
80 XC
1. saCas9
2. Cas9 protein translation
1. saCas9
2. Cas9 protein translation
1. saCas9
2. Cas9 protein translation
RuvC
HNH
1. saCas9
2. Cas9 protein translation
1. saCas9
2. Cas9 protein translation
1. saCas9
2. Cas9 protein translation

FIG. 37A

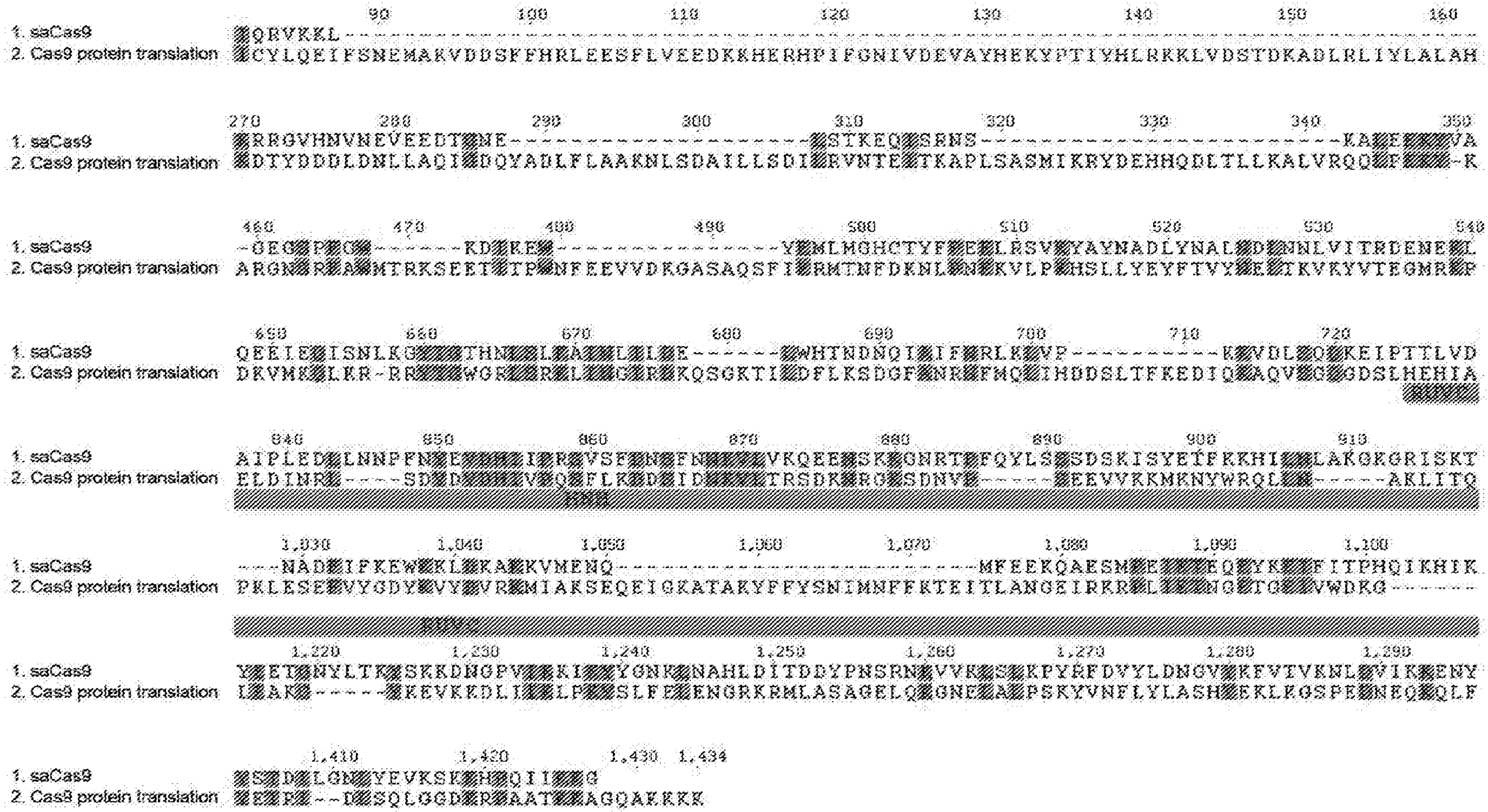

1. saCas9
2. Cas9 protein translation
90 100 110 120 130 140 150 160
1. saCas9 QRVKKL
2. Cas9 protein translation CYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAH
270 280 290 300 310 320 330 340 350
1. saCas9 RRGVHNVNEVEEDTGNE STKEQISENS KAEEVA
2. Cas9 protein translation DTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYK-K
460 470 480 490 500 510 520 530 540
1. saCas9 GEGDPDGD KDIKEE VDMLMGHCTYFEELESVEYAYNADLYNALNDLNNLVITRDENEKL
2. Cas9 protein translation ARGNLRIAIMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKP
650 660 670 680 690 700 710 720
1. saCas9 QEEIEKISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKLVP KEVDLSQQKEIPTTLVD
2. Cas9 protein translation DKVMKQLKR-RRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIA
840 850 860 870 880 890 900 910
1. saCas9 AIPLEDSLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEENSKKGNRTPFQYLSSSDSKISYETFKKHILNLAKGKGRISKT
2. Cas9 protein translation ELDINRL-----SDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVP------SEEVVKKMKNYWRQLLN-----AKLITQ
RuvC
1,030 1,040 1,050 1,060 1,070 1,080 1,090 1,100
1. saCas9 ---NADLIFKEWEKLDKAKKVMENQ MFEEKQAESMPEIETEQDYKDIFITPHQIKHIK
2. Cas9 protein translation PKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKG
HNH
1,220 1,230 1,240 1,250 1,260 1,270 1,280 1,290
1. saCas9 YATIRNYLTKDSKKDVKVLDSMQIQKYDVNKELNAHLDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKNLDVIKKENY
2. Cas9 protein translation IDAK-----EKEVKKDLIIELPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLF
1,410 1,420 1,430 1,434
1. saCas9 SEQDLQNIYEVKSKLHLQIIKKG
2. Cas9 protein translation EEQFE--DISQLGGDKRPAATKKAGQAKKKK

FIG. 37B

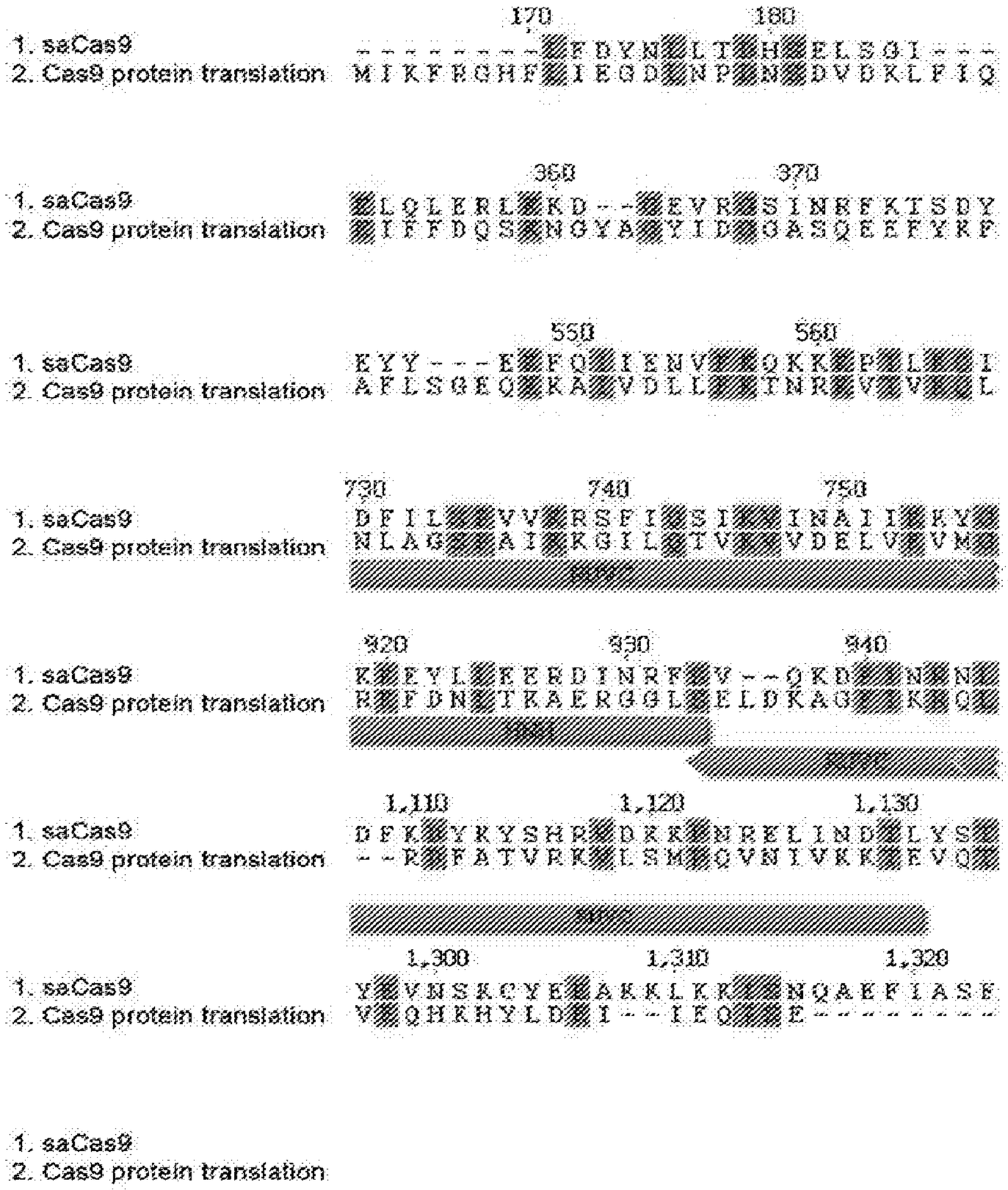
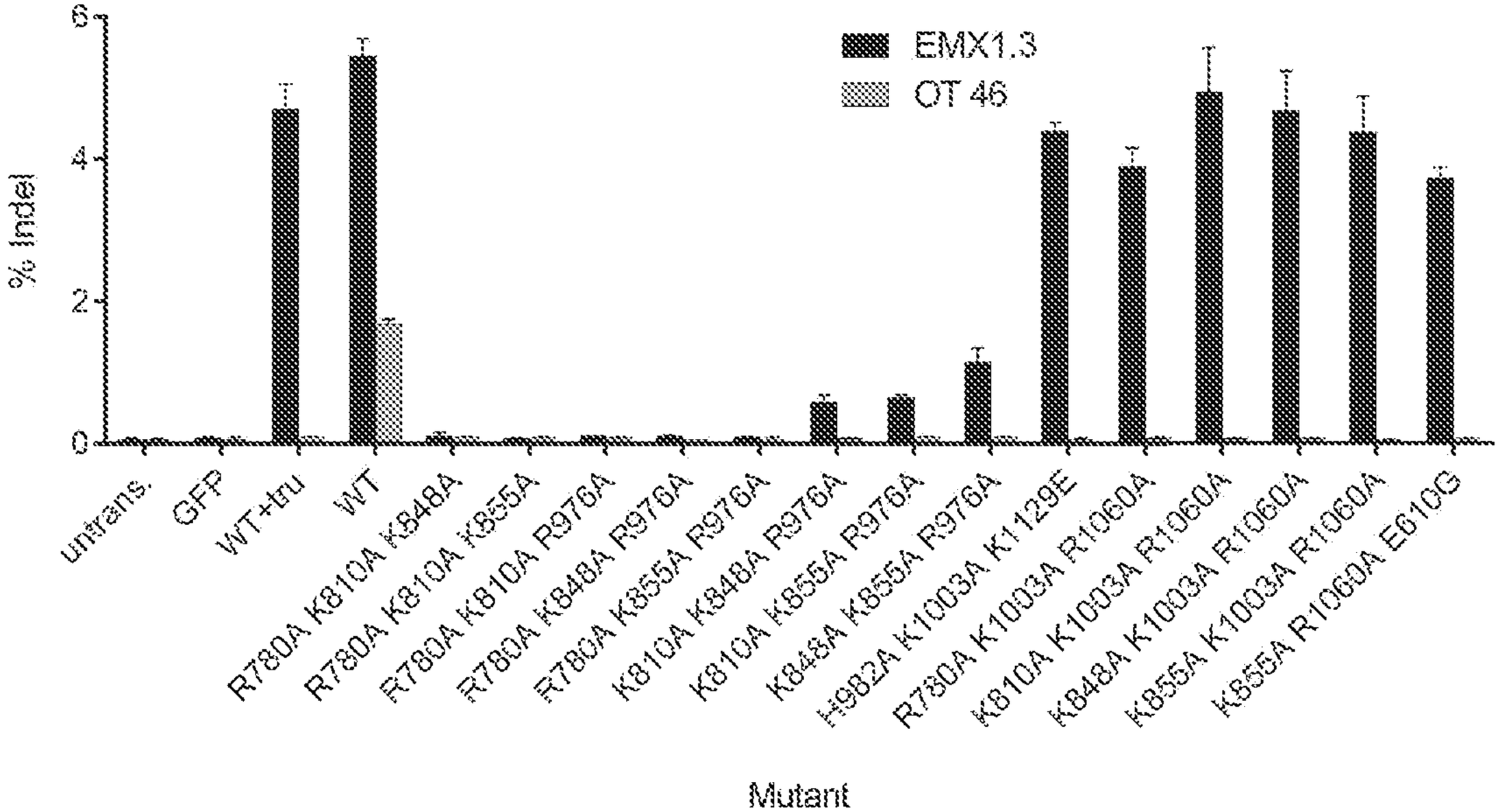
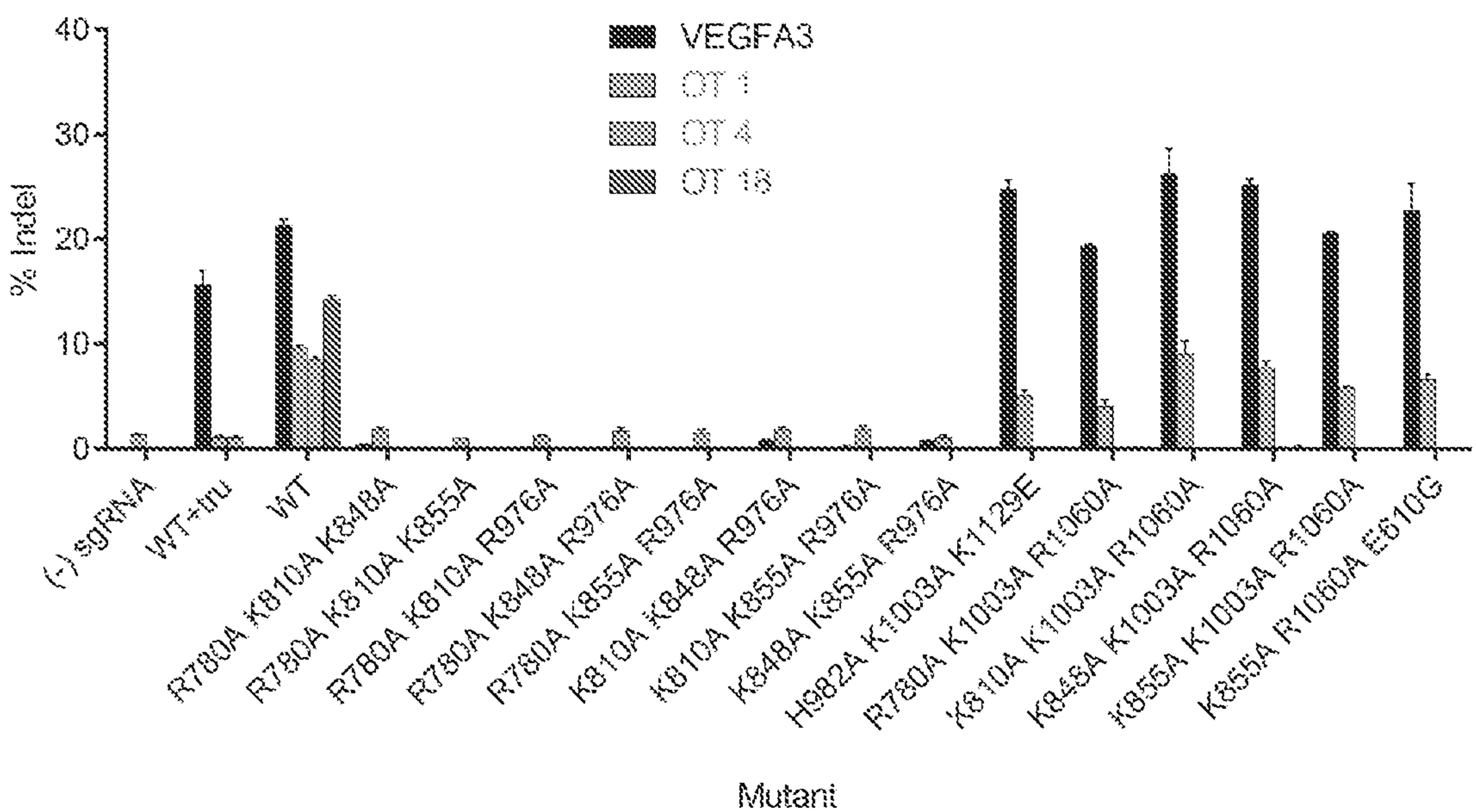


FIG. 37C

SUBSTITUTE SHEET (RULE 26)

Fig. 5


EMX1.3
OT 46
% Indel
0
2
4
6
untrans.
GFP
WT+tru
WT
R780A K810A K848A
R780A K810A K855A
R780A K810A R976A
R780A K848A R976A
R780A K855A R976A
K810A K848A R976A
K810A K855A R976A
K848A K855A R976A
H982A K1003A K1129E
R780A K1003A R1060A
K810A K1003A R1060A
K848A K1003A R1060A
K855A K1003A R1060A
K855A R1060A E610G
Mutant


VEGFA3
OT 1
OT 4
OT 18
% Indel
0
10
20
30
40
(-) sgRNA
WT+tru
WT
R780A K810A K848A
R780A K810A K855A
R780A K810A R976A
R780A K848A R976A
R780A K855A R976A
K810A K848A R976A
K810A K855A R976A
K848A K855A R976A
H982A K1003A K1129E
R780A K1003A R1060A
K810A K1003A R1060A
K848A K1003A R1060A
K855A K1003A R1060A
K855A R1060A E610G
Mutant