



(12) 发明专利申请

(10) 申请公布号 CN 117529555 A

(43) 申请公布日 2024. 02. 06

(21) 申请号 202280036959.3

(22) 申请日 2022.03.23

(30) 优先权数据

63/164,798 2021.03.23 US

(85) PCT国际申请进入国家阶段日

2023.11.22

(86) PCT国际申请的申请数据

PCT/US2022/021523 2022.03.23

(87) PCT国际申请的公布数据

W02022/204268 EN 2022.09.29

(71) 申请人 比姆医疗股份有限公司

地址 美国马萨诸塞州

(72) 发明人 B·蔡澈 L·巴雷拉 D·A·玻恩

M·孙

(74) 专利代理机构 北京市君合律师事务所

11517

专利代理师 吴瑜 张璐

(51) Int.Cl.

C12N 9/22 (2006.01)

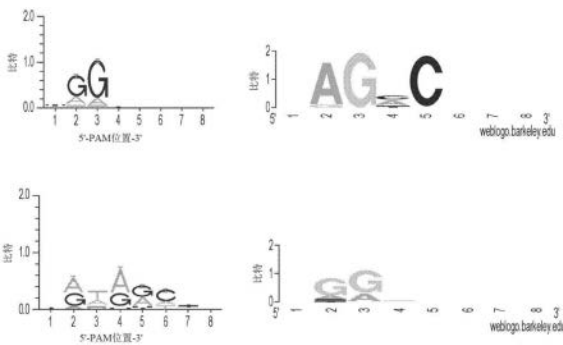
权利要求书24页 说明书121页 附图30页

(54) 发明名称

新颖CRISPR酶、方法、系统及其用途

(57) 摘要

本发明提供了用于制备和使用针对人类细胞优化的重组工程化的新颖Cas9酶的新颖系统、方法和组合物,以用于核酸靶向和操控。本发明基于来自星座链球菌、夏普氏菌属分离株RUG017、小韦荣球菌、秘鲁埃扎基拉菌、发酵乳杆菌菌株AF15-40LB菌株和嗜脲菌属马赛-P3761细菌的新颖Cas9酶的发现,所述新颖Cas9酶经密码子优化和重组生产,用于人类细胞。在一些实施例中,新颖Cas9酶可用于碱基编辑。在一些实施例中,新颖、工程化的Cas9酶用于治疗人类疾病。



1. 一种工程化的非天然存在的Cas9蛋白,其由星座链球菌Cas9、夏普氏菌Cas9、小韦荣球菌Cas9、秘鲁埃扎基拉菌Cas9、发酵乳杆菌菌株AF15-40LB Cas9或嗜脉菌属马赛-P3761 Cas9修饰。

2. 根据权利要求1所述的Cas9蛋白,其中所述星座链球菌Cas9与以下具有至少80%序列同一性:

MGKPYSIGLDIGTNSVGWAVVTDDYKVPAAKMKVLGNTDKQSIKKNLLGALLFDSGETAEATRLKRTARRRY
TRRKNRLRYLQEIFTGEMNKVDENFFQRLDDSFVLVEDDKRGEHHP IFGNIAAEVKYHDDFPTIYHLRRHLADTSKK
ADLRLVYLALAHMIKFRGHFLYEGDLKAENTDVQALFKDFVEEYDKTIEESHLSEITVDALSILTEKVS KSSRLN
LIAHYPTKKNLTLFGNLIALSLDLHPNFKTNFQLSEDAKLQFSKDTYEEDLEGFLGEVGDEYADLFASAKNLYDAI
LLSGILTVDNSTKAPLSASMVKRYEEHQDLKKLKDFIKVNAPDQYNAIFKDKNKKGYASYIESGVKQDEFYKYL
KGILLKINGSGDFLDKIDREDFLRKQRTFDNGSIPHQIHLQEMHAILRRQGEHYFPFLKENQDKIEKILTFRIPIYYV
GPLARKGSRFAWAAYKADEKITPWNFDDILDKEKSAEKFITRMTLNDLYLP EEEKVLPKHSPLYEAF TVYNELTKVK
YVNEQGEAKFFDTNMKQEIFDHVFKENRKVTKDKLLNYLNKEFEFRIVNLTGLDKENKAFNSSLGT YHDLRKILD
KSFLDDKAN EKTIEDIIQTLT LFEDREMIRQLQKYS DIFTKAQLKKLERRHYTGWGRLSYKLINGIRNKENKKT I
LDYLIDDGYANRNFMLINDDALSFKEEIARAQIIDDVDDIANVVHDLPGSPA IKGILQSVKIVDELVKVMGHN P
ANII IEMARENQTTDKGRRNSQQRLKLLQDSLKNLNDPNVNIKNVENQQQLQNDRLFLYYIQNGKDMYTGETLDINN L
SQYDIDHIIPQAFIKDNSLDNRVLTRSDKNRGKSDDVPSIEVVHEMKSFWSKLLSVKLITQRKFDNLTKAERGGLT
EEDKAGFIKRQLVETRQITKHVAQILDERFNTEFDGNKRRIRNVKII TLKSNLVSNFRKEFELYK VREINDYHHAH
DAYLNAVVGNALLLKYPQLEPEFVYGEYPKYNSYRSRSATEKFLFYSNILRFFKKEDIQTNEDGEIAWNKEKHIK
ILRKVLSYPQVNIVKKTEEQTGGSFESILPKGESDKLIPRKTKNSYWDPKKYGGF DSPVVAYSILVFADVEKGKS
KKLRKVQDMVGITIMEKKRFEKNPVDFLEQRGYRNVRL EKI IKLPKYSLFEL ENKRRRLLASAKELQKGNELVIPQ
RFTTLLYHSYRIEKDYEP EHYVEKHKDEFKELLEYISVFSRKYVLADNNLTKIEMLF SKNKDAEVSSLAKSFIS
LLTFTAFGAPAAFNFFGENIDRKRYTSVTECLNATLIHQ SITGLYETRIDL SKLGED (SEQ ID NO:1)。

3. 根据权利要求1所述的Cas9蛋白,其中所述夏普氏菌Cas9与以下具有至少80%序列同一性:

MAKNKDIRYSIGLDIGTNSVGWAVMDEHYELLKKGNHHMWGSRLFDAAEPAATRRASRSIRRRYNKRERIR
LLRDLLGDMVMEVDPTFFIRLLNVSFLDEEDKQKNLGNDYKDNYNLFIEKDFNDKTYDYKYPTIYHLRKELCENKE
KADPRLIYLALHHIVKYRGNFLKEGQSFAKVYEDIEEKL DNTLKKFMSLNDLNDLNFVDNDINSMITVLSKIYQRSK
KADDLLKIMNPTKEERAAYKEFTKALVGLKFNVSKMILAQEVKKDDKDIELDFSNDYDSTVDGLQAE LGEYIEFI
EMLHSINSWVELQDILGNNSTISAAMVEREYEEHKNDLRVLKKVIREELPDKYNEVFREDNPKLHNYLGYIKYPKNT
PVEEFY EYIKRLLAKVDTGEAREILERIDLEKFM LKQNSRTNGSIPYQM QKDEMIQI IDNQSVYYPQLKENREKLI
SILEFRIPYYFGPLNTHSEFAWIKKFEDKQKERILPWN YDQIVDIDATAEGFIERMQNTGT YFPDKPVMKN SLTV
SKFEVLNELNKIRINGKLIPVETKKELLSDLFMKNKTI TDKKLKDVLVTHQYYDTNEELKIEGYQKDLQFSTSLAP
WIDFTKIFGEINASNYQLIEKIIYDISIFEDKKILKRRLKKVYQLDDLLVDKILKLN YTGWSRLSEKLLTG IKS KN
SKETILSILENSNMNLM EIIINDESLGFKQIIEESNKKDIEGPFRYDEVKKLAGSPA IKGRIWQALLV VQEITKFMK
HEPSHIYIEFAREEQEKVRTESRIAKLQKIYKDLNLQTKEDQLVYESLKKEDAKKKIDTDALYLYLQMGKSMYSG
KPLDIDKLSTYHIDHILPRSLIKDDSLDNRVLVLPKENEWKLDSETPFEIRNKMMGFWQKLHENG LMSNKKFFSL
IRTD FNEKDKKRFINRQLVETRQI IKNVAVI INDHYTNTNVVTVRAELSHQFRERYKIYKNRDLNDLHHAH DAYIA
CILGQFIHQNFGNMDVNMIYGQYKKNYKKDVQEHNNYGFILNSMNIHFND DNSVIWDPSYIGKIKSCFCYKDVYV

TKKLEQNDAKLFDLTILPSDKNSENGVTKAKIPVNKYRKDVNKGFGSGDAPIMLAIEADKGKKHVRQVIAFPLRL
KNYNDEERIKFIEKEKNLKNVKILTEVKKNLILINHQQFFITGTNELVNATQLKLSAKNTKNLNLVDANKHNKL
ESIDDANFNEVIQELICKLQEPYISRYNSIGKEFEDSYEKINAVTKQDKLYIIIEYLIAIMSAKATQGYIKPELARE
IGTNGKNKGRIKSFTIDLNKTTFISTSVTGLFSKKYKL (SEQ ID NO:4)。

4. 根据权利要求1所述的Cas9蛋白, 其中所述小韦荣球菌Cas9与以下具有至少80%序列同一性:

MSIINFQRRGLMETQASNQLISSHLKGYPIDYFVGLDIGTSSVGWAVTNKAYELLKFRSHKMWGSRLFDEG
ESAVARRGFRSMRRRLERRKLRLKLEELFADAMAQVDPFTFMRLRESKYHYEDKTTGHSSKHILFIDKNYNDQDY
FKEYPTVYHLRSELMKSGTDDIRKLFLAVHHILKYRGNFLYEGATFDSNASTLDDVIKQALENITFNCFCDCNSAIS
SIGQILMEAGTKSDKAKAIEHLVDYIATDVTDTSSKTQKDQVKEDKKRLKAFANLVLGLNASLIDLFGSVEELE
EDLKKLQITGDTYDDKRDELAKAWSDEIYIIDDCKSVYDAIILLSIKEPGLTISESKVKAFAFKHKKDDLAILKSLK
SDRSIYNTMFKVDEKGLHNYVHYIKQGRTEETSCNREDFYKYTKKIVEGLSDSKDKEYILSQIELQILLPLQRIKD
NGVIPYQLHLEELKAILAKCGPKFPFLNEVADGFSVAEKLKMLEFRIPYVGPLNTHHNVNNGGFAWVRKASGR
VTPWNFDDKIDREKSAAAFIKNLTNKCTYLLGEDVLPKSSLLYSEFMILLNELNVRIDGKPLEKVVEHLIEAVFK
QDHKKMTKNRIEQFLKDNQYISETHKHEITGLDGEIKNDLASYRDMVRILGDGFDRSMAEEIITDITIFGESKKML
RETLRKKFASCLDDEAIKKLTKLRYRDWGRLSQKLLNGIEGCDKAGDGPETIIILMRNFSYNLMELLGDKFSFME
RIQEINAKLTEGQIVNPHDIIDDLALSPAVKRAVWQALRIVDEVAHIKKALPARIFVEVTRSNKNEKKKKDSRQKR
LSDLYAAIKKDDVLLNGLNNEIFGELKSSSLAKYDDAALRSKKLYLYTQMGRCAITGEIIELSLLNTDNYDIDHIY
PRSLTKDDSFNDNLVLCKRTANAQKSDAYPISEEIQTQKPFWTFLLKQGLISERKYERLTRITPLTADDLSGFIAR
QLVETNQSVKAATTLRRLYPGVDFVFAENVTDVRHNNFIKVRSLNHHHHAKDAYLNIIVGVNHYHERFTRNFR
AFFKKNGANRTYNLAKMFNYDVNCTNAKDGKAWDVKTSMDTVKMMDSNDVRVTKRLLQGTALADATYKATVAG
KAKDGAYIGMKTSSVFADVSKYGGMTKIKNAYSIIVQYTGKKGEVIKEIVPLPIYLTNRNTDQDLINYVASIIP
QAKDISIIYGKLCINQLVKVNGFYIYLGKTKNSKFCIDNAIQVIVSNEWIPYLVLEKFNMRKDNKDLKANVST
RALDNKHTIEVRIVEEKNIFFDYLVSKLKMPIYQKMGKNKAAELSEKGYGLFKKMSLEEQSIHLIELLNLLTNQK
TTFEVKPLGITASRSTVGSKISNQDEFKVINESITGLYSNEVTIV (SEQ ID NO:8)。

5. 根据权利要求1所述的Cas9蛋白, 其中所述秘鲁埃扎基拉菌Cas9与以下具有至少80%序列同一性:

MTKVDDYYIGLDIGTSSVGWAVTDEAYNVLFNSKKMWGVRLFDDAKTAEERRGQRGARRRLDRKKERLSLL
QDFFAAEEVAKVDPNFFLRDLNSDLYMEDKDQKLKSKYTLFNDKDFKDNFHKKYPTIHLLMDLIEDDSKKDIRLV
YLACHYLLKNRGHIFEGQKFDTKSSFENSLNELKVHLNDEYGLDLEFDNENLINILTPKLNKTAKKKELKSVIG
DTKFLKAVSAIMIGSSQKLVDLFENPEDFDDSAIKSVDFSTTSFDDKYSYELALGDKIALVNILKEIYDSSILEN
LLKEADKSKDGNKYISNAFVKYKNGHGGDLKEFKRLVRQYHKSAYFDIFRSEKVNNDNYVSYTKSSISNNKRVKANK
FTDQEAIFYKFAKKHLETIKYKINKVNGSKADLELIDGMLRDMFKNFMPKIKSSDNGVIPYQLKLMELNKILENQS
KHHEFLNVSDEYGSVCDKIASIMEFRIPYVGPLNPNNSKYAWIKKQKDSEITPWNFKDVLDSSREEFIDSLIGR
CTYLLKDEKVLPAKSLLYNEYMLNELNKLNDLPITEEMKKKIFDQLFKTRKKVTLKAVANLLKKEFNINGEILL
SGTDGDFKQGLNSYNDFAIVGDKVSDDYRDKIEEIIKLIVLYGDDKSYLQKKIKAGYGYFTDSEIKKMAGLNY
KDWGRLSKKLLTGLEGANKITGERGSIHFMREYNLNLMEKMSASFTFTEEIQKLPVDDRKLSYEMVDELYLSPS
VKRMLWQSLRIVDEIKNIMGTDSKKIFIEARGKEEVKARKESRKNQLLFYKDGKKAFFISEIGEERYSYLLSEIE
GEEENKFRWDNLYLYTQLGRCMYSLEPIDISELSSKNIYDQDHIYPKSKYDDSIENRVLVKKDLNSKKGNSYPI

PDEILNKNCYAYWKILYDKGLIGQKKYTRLTRRTGFTDDELVQFISRQIVETRQATKETANLLKTICKNSEIVYSK
AENASRFRQEFDIVKCRAVNDLHHMHDAYINIIVGNVYNTKFTKDP MNFVKKQE KARSYNLENMFKYDVKRGGYTA
WIADDEKGT VKNASIKRIRKELEGTNYRFRMNYIESGALFNATLQRKNKGSRLPKDKGPKSSIEKYGGYTNINKA
CFAVLDIKSKN KIERKLMPVERE IYAKQKNDKKLSDEIFSKYLKDRFGIEDYRVVYPVVKMRTLLKIDGSYYFITG
GSDKTLELRSALQLILPKKNEWAIKQIDKSSSENDYLTIERIQDLTEELVYNTFDIIVNKFKTSVFKKSFLNLFQDD
KIENIDFKFKSMD FKECKTLLMLVKAIRASGVRQDLKSIDLKS DYGRLLSSKTNNIGNYQEFKIIINQSITGLFENE
VDLLKL (SEQ ID NO:14)。

6. 根据权利要求1所述的Cas9蛋白,其中所述发酵乳杆菌菌株AF15-40LB Cas9与以下具有至少80%序列同一性:

MKEYHIGLDIGTSSIGWAVTDSQFKLMRIKGKTAIGVRLFEEGKTA AERRTFRTTRRRRLKRRKWRLHYLDEI
FAPHLQEVDENFLRRLKQSNIHPEPAKNQAFIGKLLFPDLLKKNERGYPTLIKMRDEL PVEQRAHYPVTNIYKLR
EAMINEDRQFDLREVYLAVHHIVKYRGHFLNNASVDKFKVGRIDFKSFNVLNEAYEELQNGEGSFTIEPSKVEKI
GQLLLDTKMRKLD RQKAVAKLLEVKVADKEETKR NKQIATAMSKLV LGYKADFATVAMANGNEWKIDLSSETSEDE
IEKFREE LSDAQNDILTEITSLFSQIMLNEIVPNGMSISESMMDRYW THERQLAEVKEYLATQPASARKEFDQVYN
KYIGQAPKEKGF DLEKGLKKILSKKENWKEIDELLKAGDFLPKQRTSANGVIPHQMHQQELDRIIEKQAKYYPWLA
TENPATGERDRHQAKYELDQLVSFRIPYYVGPLVTP EVQKATSGAKFAWAKRKEDGEITPWNLWDKIDRAESAEAF
IKRMTVKD TYLLNEDVLPANSLLYQKYNVLNELNNVRVNGRRLSVG IKGDIYTEL FKKKKT VKAGDVASLVMATR
GVNKSVEGLSDPKKFNSNLATYLDLKSIVGDKVDDNRYQMDLENIIEWRSVFEDGEIFADKLTEVEWLTDEQRSA
LVKKRYKGWGRLSKKLLTGIVDENGQRIIDLMWNTDQNF MQIVNQPVFKEQIDQLNQKAITNDGMTLRERVSVLD
DAYTSPQNKKA IWQVVRVEDIVKAVGNAPKSISIEFARNEG NKGEITRSRRTQLQKLFEDQAHEL VKDTSLTEEL
EKAPDLSDRYYFYFTQG GKMYPGDPINFDEISTKYDIDHILPQS FVKDDSLDNRLVLSRAENNKSDRVPALYA
AKMKPYWNQLLKQGLITQRKFENLTMDVDQTIKYRSLGFVKRQLVETRQVIKLTANILGSMYQEAGTDI IETRAGL
TKQLREEFDLPKVREVNDYHHAVDAYLTTFAGQYLNRRYPKLRSFFVYGEYMKFKHGS DLKLRNFNFHELMEGDK
SQGKVVDQQTGELITTRDEVADYFDWV INLKVMLISNETYEETGKYFDASHESSSLYLKNQNKSKLVVPLKNKLQ
PEYYGAYTGITQGYMVLKLLDKKGGFGVYRIPRYAADILNKCHDEVAYRNKIAEIISSDPRAPKSFEVVVPRVLK
GTFLVDGEEKFILSSYRYKV NATQLILPVSDIKLIQDNFKALKKLVNEMQTKKLI EIDNLRQVDKYKLYDINK
FRAKLHDGRSKFVELDDFGQDASKEKVI IKILRGLHFGSDLQNLKEIGFGTTP LGQFQVSEAGIRLSNTAFIIFKS
PTGLFNRKLYLKNL (SEQ ID NO:84)。

7. 根据权利要求1所述的Cas9蛋白,其中所述嗜热菌屑马赛-P3761 Cas9与以下具有至少80%序列同一性:

MEKKTNYTIGLDIGTDSVGWAVVKDDLELVKKRMKVLGNTETNYIKKNLWGSLLFESGQTAKDRRLKRVARR
RYERRRNRLTELQKIFAPAIDEVDENFFFRLNESFLVPEDKAFSKNP IFGTLGEDKTYKYTYPTIYHLRQHLADSE
EKADVRLIYLALAHMIKYRGHFLIEGKLDTEHIAINENLEQFFESYNALFSEEP IELRKEELIAIENILREKNSRT
VKEKRITSFLKDIGRANKQSPMAFITLIVGKKAKFKAAFNLEEEISLNLTD DSYDENLEILLNTIGSDFADLFDH
AQRVYNAVELAGILSGDVKNTHAKLSAQMVAMYERHKEQLKEYKSFIKANLPDQYDMTFVAPKDAQKKDLKGYAGY
IDGNMSQDSFYKFVKDQLKEVPGSEKFLDSIEKEDFLRKQRSFYNGVIPNQVHLAEMEAILDRQENYYPWLKENRE
KIIISLLTFRIPYYVGPLADGQSEFAWLERKSDEKIKPWNFSDVVDLDRSAEKFIEQLIGRDTYLPDEYVLPKSLI
YQKYMVFNELTKIAYLDERQKRMNLSSVEKKEIFETLFKKRSKVTEKQLVKFFENYLQIDNPTIFGIEDAFNADYS
TYVELAKVPGMKSMDDPDNEDLMEEIVKILTVFEDRKMRKQLEKYKERLSPEQIKELAKKH YTGWGRLSKKLLV

GIRDKETQKTILDYLVEDDNHSGGRQHLNRNLMQLINDDRLSFKKTIAELQMIDPSADLYAQVQEIAGSPAIIKKGI
LLGLKIVDEIIRVMGEKPENIVIEARENQTTARGKALSKRREAKIKEGLAALGSSLLKENLPGNADLSQRKIYLY
YTQNGKDIYLDPLDFDRLSQYDEDHIIPQSFTVDNSLDNLVLTNSSQNRGNKKDDVPSLEVVRQLAYWRSKDA
GLMTQRKFDNLTAMRGGLTDKDRERFIQRQLVETRQITKNVAKLLDMRLNDKKDEAGNKIRETNIVLLKSAMASE
FRKMFRLYKVRLENDYHHAHDAYLNAAIAINLLALYPYMAFFVYGEFRYKKKPQAEKATYEKLRQWNLIKRFGEK
QLFTPDHEDCWNKERDIKTIKKVMGYRQVNVVKKAEERTGMLFKETINGKTNKGSRIPIKKDLDPSTKYGGYIEEKM
AYYAVISYEDKKKKPGKTIVGISIMDKKEFEYDSISYLGKLGFSNPVVQIILKNYSLIAYPDGRRRYITGATKTTK
GKVELQKANQIAMEQDLVNFYIHLKNYDEISHPESYAFVQSHTDYFDRLFDSIEHYTRRFLDAETNINRLRRIYEE
EKKKDPVDIEALVASFIELLLKTSAGAPADFIEMGEAISRRRYNSMTGLFDGQVIYQSLTGLYETRMRFED (SEQ
ID NO:86)。

8. 根据权利要求2至7中任一项所述的Cas9蛋白,其包含与SEQ ID NO:1、4、8、14、84或86至少85%、至少90%、至少92%、至少95%、至少96%、至少97%、至少98%或至少99%相同的氨基酸序列。

9. 根据前述权利要求中任一项所述的Cas9蛋白,其进一步包含核定位序列(NLS)和/或FLAG、HIS或HA标签。

10. 根据权利要求9所述的Cas9蛋白,其中所述星座链球菌Cas9具有与以下至少80%相同的氨基酸序列:

MPKKKRKVGKPYISGLDIGTNSVGWAVVTDDYKVPKKMKVLGNTDKQSIKKNLLGALLFDSGETAEATRL
KRTARRRYTRRNRLRYLQEIFTGEMNKVDENFFQRLDDSFVDEDKRGHEHPIFGNIAAEVKYHDDFPTIYHLRR
HLADTSKKADLRLVYLALAHMIKFRGHFLYEGDLKAENTDVQALFKDFVEEYDKTIEESHLSEITVDALSILTEKV
SKSSRLNLI AHYPTEKKNLTFGNLIALSLDLHPNFKTNFQLSSEDAKLQFSKDTYEEDLEGFLGEVGEYADLFAS
AKNLYDAITLLSGILTVDDNSTKAPLSASMVKRYEEHQDLKKLKDFIKVNAPDQYNAIFKDKNKKGYASYIESGVK
QDEFYKYLKGILLKINGSGDFLDKIDREDFLRKQRTFDNGSIPHQIHLQEMHAILRRQGEHYPLKENQDKIEKIL
TFRIPYYVGPLARKGSRFAWAEYKADEKITPWNFDDILDKEKSAEKFITRMTLNDLYLPEEKVLPKHSPLYEAFTV
YNELTKVKYVNEQGEAKFFDTNMKQEIFDHVFKENRKVTKDKLLNYLNKEFEFRIVNLTGDKENKAFNSSLGTY
HDLRKILDKSFLDDKANAKTIEDIIQTLTLFEDREMIRQLQKYSIDIFTKAQLKKLERRHYTGWGRLSYKLINGIR
NKENKKTILDYLI DDGYANRNFMLINDDALSFKEEIAAQIIDDVDDIANVVHDLPGSPAIIKKGILQSVKIVDEL
VKVMGHN PANII IEMARENQTTDKGRRNSQQLKLLQDSLKNLDNPVNIKNVENQQLQNDRLFYYIQNGKDMYTG
ETLDINNLSQYDIDHIIPQAFIKDNSLDNRVLTRSDKNRGKSDDVPSIEVVHEMKSFWSKLLSVKLITQRKFDNLT
KAERGGTTEEDKAGFIKRLVETRQITKHVAQILDERFNTEFDGNKRRIRNVKIIITLKSNLVSFRKEFELYKVR
INDYHHAHDAYLNAVVGNAALLKYPQLEPEFVYGEYPKYNSYRSRSKATEKFLFYSNILRFFKKEDIQTNEDEIA
WNKEKHILKILRKVLSYPQVNI VKKTEEQTGGFSKESILPKGESDKLIPRKTNSYWDPKKYGGFDSPPVAYSILVF
ADVEKGKSKKLKRVQDMVGITIMEKKRFEKNPVDLFLEQGRYRNVLEKIIKLPKYSLELENKRRRLLASAKELQK
GNELVIPQRFTTLLYHSYRIEKDYEPHREYVEKHKDEFKELLEYISVFSRKYVLADNNLTKEIMLFSKNKDAEVS
SLAKSFISLLTFTAFGAPAAFNFFGENIDRKRYTSVTECLNATLIHQSI TGLYETRIDL SKLGEDGKRPAATKKAG
QAKKKKGSYPYDVPDYAYPYDVPDYAYPYDVPDYA (SEQ ID NO:2)。

10b. 根据权利要求9所述的Cas9蛋白,其中所述星座链球菌Cas9具有与以下至少80%相同的氨基酸序列:

MPKKKRKVGMGKPYISGLDIGTNSVGWAVVTDDYKVPKKMKVLGNTDKQSIKKNLLGALLFDSGETAEATR

LKRTARRRYTRRNRLRYLQEIFTGEMNKVDENFFQRLDDSLVDEDKRGEHHPIFGNIAAEVKYHDDFPTIYHLR
RHLADTSKKADLRLVYLALAHMIKFRGHFLYEGDLKAENTDVQALFKDFVEEYDKTIEESHLSEITVDALSILTEK
VSKSSRLLENLIAHYPTEKKNTLFGNLIALSLDLHPNFKTNFQLSEDAKLQFSKDTYEEDLEGFLGEVGDEYADLFA
SAKNLYDAILLSGILTVDDNSTKAPLSASVMKRYEEHQKDLKKLKDFIKVNAPDQYNAIFKDKNKKGYASYIESGV
KQDEFYKYLKGILLKINGSGDFLDKIDREDFLRKQRTFDNGSIHQIHLQEMHAILRRQGEHYPLKENQDKIEKI
LTFRIPYYVGPLARKGSRFAWAEYKADEKITPWNFDDILDKEKSAEKFITRMTLNDLYLPEEKVLPKHSPLYEAFT
VYNELTKVKYVNEQGEAKFFDTNMKQEIFDHVFKENRKVTDKLLNYLNKEFEFRIVNL TGLDKENKAFNSSLGT
YHDLRKILDKSFLDDKANEKTIEDIIQTLTLFEDREMIRQLKQYSDIFTKAQLKKLERRHYTGWGRLSYKLINGI
RNKENKKTILDYLDGIDGYANRNFMLINDDALSFKEEIARAQIIDVDVDDIANVVHDLPGSPAIIKKGILQSVKIVDE
LVKVMGHN PANII IEMARENQTTDKGRRNSQQRLKLLQDSLKNLDPVNIKNVENQQQLQNDRLFLYYIQNGKDMYT
GETLDINNLSQYDIDHIIPQAFIKDNSLDNRVLTRSDKNRGKSDDVPSIEVVHEMKSFWKLLSVKLITQRKFDNL
TKAERGGLTEEDKAGFIKRQLVETRQITKHVAQILDERFNTEFDGNKRRIRNVKII TLKSNLVS NFRKEFELYKVR
EINDYHHAHDAYLNAVVG NALLLKYPQLEPEFVYGEYPKYNSYRSRKSATEKFLFYSNILRFFKKEDIQT NEDGEI
AWNKEKHIKILRKVLSYPQVNIVKKTEEQTGGFSKESILPKGESDKLIPRKTKN SYWDPKKYGGFMQPVVAYSILV
FADVEKGKSKKL RKVQDMVGITIMEKKRFEKNPVDFLEQRGYRNVRLKIIKL PKYSLFELENKRRRLLASAKFLQ
KGNELVIPQRFTTLLYHSYRIEKDYEPHREYVEKHKDEFKELLEYISVFSRKYVLADNNLTKIEMLF SKNKDAEV
SSLAKSFISLLTFTAFGAPRAFNF FGENIARKEYRSVTECLNATLIHQ SITGLYETRIDL SKLGEDGEGADKRTAD
GSEFESP KKKRKV (SEQ ID NO:95)。

10c. 根据权利要求9所述的Cas9蛋白,其中所述星座链球菌Cas9具有与以下至少80%相同的氨基酸序列:

MPKKKRKVGMGKPYSIGLDIGTNSVGWAVVTDDYKVP AKKMKVLGNTDKQS IKNLLGALLFDSGETAEATR
LKRTARRRYTRRNRLRYLQEIFTGEMNKVDENFFQRLDDSLVDEDKRGEHHPIFGNIAAEVKYHDDFPTIYHLR
RHLADTSKKADLRLVYLALAHMIKFRGHFLYEGDLKAENTDVQALFKDFVEEYDKTIEESHLSEITVDALSILTEK
VSKSSRLLENLIAHYPTEKKNTLFGNLIALSLDLHPNFKTNFQLSEDAKLQFSKDTYEEDLEGFLGEVGDEYADLFA
SAKNLYDAILLSGILTVDDNSTKAPLSASVMKRYEEHQKDLKKLKDFIKVNAPDQYNAIFKDKNKKGYASYIESGV
KQDEFYKYLKGILLKINGSGDFLDKIDREDFLRKQRTFDNGIIPHQIHLQEMHAILRRQGEHYPLKENQDKIEKI
LTFRIPYYVGPLARKGSRFAWAEYKADEKITPWNFDDILDKEKSAEKFITRMTLNDLYLPEEKVLPKHSPLYEAFT
VYNELTKVKYVNEQGEAKFFDTNMKQEIFDHVFKENRKVTDKLLNYLNKEFEFRIVNL TGLDKENKAFNSSLGT
YHDLRKILDKSFLDDKANEKTIEDIIQTLTLFEDREMIRQLKQYSDIFTKAQLKKLERLHYTGWGRLSYKLINGI
RNKENKKTILDYLDGIDGYANRNFMLINDDALSFKEEIARAQIIDVDVDDIANVVHDLPGSPAIIKKGILQSVKIVDE
LVKVMGHN PANII IEMARENQTTDKGRRNSQQRLKLLQDSLKNLDPVNIKNVENQQQLQNDRLFLYYIQNGKDMYT
GETLDINNLSQYDIDHIIPQAFIKDNSLDNRVLTRSDKNRGKSDDVPSIEVVHEMKSFWKLLSVKLITQRKFDNL
TKAERGGLTEEDKAGFIKRQLVETRQITKHVAQILDERFNTEFDGNKRRIRNVKII TLKSNLVS NFRKEFELYKVR
EINDYHHAHDAYLNAVVG NALLLKYPQLEPEFVYGEYPKYNSYRSRKSATEKFLFYSNILRFFKKEDIQT NEDGEI
AWNKEKHIKILRKVLSYPQVNIVKKTEEQTGGFSKESILPKGESDKLIPRKTKN SYWDPKKYGGFMQPVVAYSILV
FADVEKGKSKKL RKVQDMVGITIMEKKRFEKNPVDFLEQRGYRNVRLKIIKL PKYSLFELENKRRRLLASAKFLQ
KGNELVIPQRFTTLLYHSYRIEKDYEPHREYVEKHKDEFKELLEYISVFSRKYVLADNNLTKIEMLF SKNKDAEV
SSLAKSFISLLTFTAFGAPRAFNF FGENIARKEYRSVTECLNATLIHQ SITGLYETRIDL SKLGEDGEGADKRTAD
GSEFESP KKKRKV (SEQ ID NO:96)。

11. 根据权利要求9所述的Cas9蛋白,其中所述夏普氏菌Cas9具有与以下至少80%相同的氨基酸序列:

MPKKKRKVGAKNKDIRYSIGLDIGTNSVGWAVMDEHYELLKKGNNHMMWGSRLFDAAEPAATRRASRSIRRRY
NKRERIRLLRDLLGDMVMEVDPTFFIRLLNVSFLDEEDKQKNLGN DYKDNYNLFIEKDFNDKTYDYKYPTIYHLR
KELCENKEKADPRLIYLALHHIVKYRGNFLKEGQSFAKVYEDIEEKL DNTLKKFMSLNDLNDLFVDNDINSMITVL
SKIYQRSKKADDLLKIMNPTKEERAAYKEFTKALVGLKFNVSKMILAQEVKKDDKDIELD FSNVDYDSTVDGLQAE
LGEYIEFIEMLHSINSWVELQDILGNSTISAAMVEREYEEHKNDLRVLKKVIREELPDKYNEVFREDNPKLHNYLG
YIKYPKNTPVEEFYEYIKRLLAKVDTGEAREILERIDLEKFM LKQNSRTNGSIPYQMQKDEMIQIIDNQSVYYPQL
KENREKLISILEFRIPYYFGPLNTHSEFAWIKKFEDKQKERILPWN YDQIVDIDATAEGFIERMQNTGT YFPDKPV
MAKNSLTVSKFEVLNELNKIRINGKLIPVETKKELLSDLFMKNKTITDKKLKDWLVTHQYYDTNEELKIEGYQKDL
QFSTSLAPWIDFTKIFGEINASYQLIEKIIYDISIFEDKKILKRR LKKVYQLDDLLVDKILKLN YTGWSRLSEKL
LTGIKSKNSKETILSILENSNMNLMEIINDESLGFKQII EESNKKDIEGPFRYDEVKKLAGSPA IKRGIWQALLVV
QEITKFMKHEPSHIYIEFAREEQEKVRTESRIAKLQKIYKDLNLQTKEDQLVYESLKKEDAKKKIDTDALYLYLQ
MGKSMYSGKPLDIDKLSTYHIDHILPRSLIKDDSLDNRVLVLPKENEWKLDSETPPFEIRNKMMGFQKLHENG LM
SNKKFFSLIRTD FNEKDKKRFINRQLVETRQIIKNVAVIINDHYTNTNVTVRAELSHQFRERYKIIYKNRDLNDLH
HAHDAYIACILGQFIHQNFNM DVNMIYGQYKKNYKKDVQEHNNYGFILNSMNIHFNDNSVIWDPSYIGKIKSC
FCYKDVYVTKKLEQNDAKLFDLTILPSDKNSENGVTAKIPV NKYRKDVNKYGGFSGDAPIMLAIEADKGKKHVRQ
VIAFPLRLKNYNDEERIKFIEKEKNLKNVKILTEVKKNLILINH QYFFITGTNELVNATQLKLSAKNTKNL FNLV
DANKHNKLESIDDANFNEVIQELICKLQEP IYSRYNSIGKEFEDSYEKINAVTKQDKLYIEYLI AIMS AKATQGY
IKPELAREIGTNGKNKGRIKSFTIDLNKTTFISTSVTGLFSK KYKLGKRPAATKKAGQAKKKKGSYPYDVPDYAYP
YDVPDYAYPYDVPDYA (SEQ ID NO:5)。

12. 根据权利要求9所述的Cas9蛋白,其中所述小韦荣球菌Cas9具有与以下至少80%相同的氨基酸序列:

MPKKKRKVGSIINFQRRGLMETQASNQLISSHLKGYPIKDYFVGLDIGTSSVGWAVTNKAYELLKFRSHKMW
GSRLFDEGESAVARRGFRSMRRRLERRKLRLKLEELFADAMAQVDPTFFMRLRESKYHYEDKTTGHSSKHILFID
KNYNDQDYFKEYPTVYHLRSELMKSGTDDIRKLFLAVHHILKYRGNFLYEGATFDSNASTLDDVIKQALENITFNC
FDCNSAIISSIGQILMEAGTKSDKAKAIEHLVDTYIATDVTDTSSKTQKDQVKEDKKRLKAFANLVGLNASLIDL
FGSVEELEEDLKKLQITGDTYDDKRDELAKAWSDEIYIIDDCKSVYDAIILL SIKEPGLTISESKVAFNKHKDDL
AILKSLLKSDRSIYNTMFKVDEKGLHNYVHYIKQGRTEETSCNREDFYKYTKKIVEGLSDSKDKEYILSQIELQIL
LPLQRIKDNGVIPYQLHLEELKAILAKCGPKFPFLNEVADGFSVAEKL IKMLEFRIPYYVGPLNTHHNV DNGGFAW
AVRKASGRVTPWNFDDKIDREKSAAAFIKNL TNKCTYLLGEDVLPKSSLLYSEFM LNLNNVRIDGKPLEKV VKE
HLIEAVFKQDHKKMTKNRIEQFLKDNGYISETHKHEITGLDGEIKNDLASYRDMVRILGDGFDRSMAEEIITDITI
FGESKKMLRETLRKKFASCLDDEAIKKLTKLRYRDWGRLSQKLLNGIEGCDKAGDGPETIIILMRNFSYNLMELL
GDKFSFMERIQEINAKLTEGQIVNPHDIIDDLALSPAVKRAVWQALRIVDEVAHIKKALPARIFVEVTRSNKNEKK
KKDSRQKRLSDLYAAIKKDDVLLNGLNNEIFGELKSSLAKYDDAALRSKKLYLYYTQMGRCAYTGEI IELSLNTD
NYDIDHIYPRSLTKDDSFNVLCKRTANAQSDAYPISEEIQKTQKPFWTF LKQQGLISERKYERLTRITPLTAD
DLSGFIARQLVETNQSVKAATTLLRRLYPGVDVVFVKAENVTD FRHDNNFIKVRSLNHHHHAKDAYLNIVVGNVYH
ERFTRNFRAFFKKNGANRTYNLAKMFNYDVNCTNAKD GKAWDVKTSMDTVKKMDSNDVRVTRKRLLEQT GALADAT
IYKATVAGKAKDGAYIGMTKSSVFADVSKYGGMTKIKNAYSII VQYTGKKGEVIKEIVPLPIYLTNRNTTDQDLI

NYVASIIPQAKDISIIYGKLCINQLVKVNGFYIYLGGKTNSKFCIDNAIQVIVSNEWIPYLVLEKFNNMRKDNKD
LKANVVSTRALDNKHTIEVRIVEEKNIIEFFDYLVSKLKMPIYQKMKGNKAAELSEKGYGLFKKMSLEEQSIHLIEL
LNLLTNQKTTFEVKPLGITASRSTVGSKISNQDEFKVINESITGLYSNEVTIVGKRPAATKKAGQAKKKKGSYPYD
VPDYAYPYDVPDYAYPYDVPDYA (SEQ ID NO:9)。

13. 根据权利要求9所述的Cas9蛋白,其中所述秘鲁埃扎基拉菌Cas9具有与以下至少80%相同的氨基酸序列:

MPKKKRKVGTKVKDYIIGLDIGTSSVGWAVTDEAYNVLFNSKKMWGVRLFDDAKTAEERRGQRGARRRLDR
KKERLSLLQDFFAAEEVAKVDPNFFLRLDNSDLYMEDKDQKLKSKYTLFNDKDFKDNFHKKYPTIHHLLMDLIEDD
SKKDIRLVYLACHYLLKNRGHFIFEGQKFDTKSSFENSLNELKVHLNDEYGLDLEFDNENLINILTPKLNKTAKK
KELKSVIGDTKFLKAVSAIMIGSSQKLVDLFENPEDFDDSAIKSVDFSTTSFDDKYSDYELALGDKIALVNILKEI
YDSSILENLLKEADKSKDGNKYISNAFVKKYNKHGQDLKEFKRLVRQYHKSAYFDIFRSEKVNNDVSYTKSSISN
NKRVKANKFTDQEAIFYKFAKKHLETIKYKINKVNGSKADLELIDGMLRDMEFKNFMPKIKSSDNGVIPYQLKLMEL
NKILENQSKHHEFLNVSDEYGSVCDKIASIMEFRIPYYVGPLNPNSKYAWIKKQKDSEITPWNFKD VVDLDSSREE
FIDSLIGRCTYKDEKVLPAKSLLYNEYMVLNELNNLKLNDLPITEEMKKKIFDQLFKTRKKVTLKAVANLLKKEF
NINGEILLSGTDGDFKQGLNSYNDFAIVGDKVSDDYRDKIEEIIKLIVLYGDDKSYLQKKIKAGYGKYFTDSEI
KKMAGLNYKDWGRLSKKLLTGLEGANKITGERGSI IHFMREYNLNLMELEMSASFTFTEEIQKLNVPDDRKLSYEMV
DELYLSPSVKRMLWQSLRIVDEIKNIMGTDSKKIFIEMARGKEEVKARKESRKNQLLKFYKDGKKAFISEIGEERY
SYLLSEIEGEEENKFRWDNLYLYYTQLGRCMYSLEPIDISELSSKNIYDQDHIYPKSKIYDDSIENRVLVKKDLNS
KKGNSYPIPDEILNKNCYAYWKILYDKGLIGQKKYTRLTRRTGFTDDELVQFISRQIVETRQATKETANLLKTICK
NSEIVYSKAENASRFRQEFDIVKCRVNDLHHMHDAYINIIVGNVYNTKFTKDP MNFVKKQEKARSYNLENMFKYD
VKRGGYTAWIADDEKGTVKNASIKRIRKELEGTNYRFRTRMNYIESGALFNATLQRKNKGSRPLKDKGPKSSIEKYG
GYTNINKACFAVLDIKSKNKIERKLMPVEREIIYAKQKNDKKLSDEIFSKYKDRFGIEDYRVVYPVVKMRTLKID
GSYYFITGGSDKTLELRSALQLILPKKNEWAIKQIDKSSENDYLTIERIQDLTEELVYNTFDIIVNKFKTSVFKKS
FLNLFQDDKIENIDFKFKSMDFKEKCKTLLMLVKAIRASGVRQDLKSIDLKSDYGRLLSSKTNNIGNYQEFKIIINQS
ITGLFENEVDLLKLGRPAATKKAGQAKKKKGSYPYDVPDYAYPYDVPDYA (SEQ ID NO:15)
(D12A突变体呈粗体)。

14. 根据权利要求9所述的Cas9蛋白,其中所述发酵乳杆菌菌株AF15-40LB Cas9具有与以下至少80%相同的氨基酸序列:

MPKKKRKVGKEYHIGLDIGTSSIGWAVTDSQFKLMRIKGKTAIGVRLFEEGKTAAERRTFRTTRRRRLKRRKW
RLHYLDEIFAPHLQEVDENFLRRLKQSNIPEDPAKNQAFIGKLLFPDLLKKNERGYPTLIKMRDELPEQRAHYP
VTNIYKLREAMINEDRQFDLREVYLAVHHIVKYRGHFLNNASVDKFKVGRIDFDKSFNVLNEAYEELQNGEGSFTI
EPSKVEKIGQLLLDTKMRKLD RQKAVAKLLEVKVADKEETKRNKQIATAMSKLVLYGKADFATVAMANGNEWKIDL
SSETSEDEIEKFREELESDAQNDILTEITSLFSQIMLNEIVPNGMSISESMMDRYWOTHERQLAEVKEYLATQPASAR
KEFDQVYNKYIGQAPKEKGFDEKGLKKILSKKENWKEIDELLKAGDFLPKQRTSANGVIPHQMHQQELDRIIEKQ
AKYYPWLATENPATGERDRHQAKYELDQLVSFRIPYYVGPLVTPEVQKATSGAKFAWAKRKEDGEITPWNLWDKID
RAESAEAFIKRMTVKDITYLLNEDVLPANSLLYQKYNVLNELNNVRVNGRRLSVGIKQDIYTELFKKKKTKVAGDVA
SLVMAKTRGVNKPSVEGLSDPKKFNSNLATYLDLKSIVGDKVDDNRYQMDLENIIEWRSVFEDGEIFADKLTEVEW
LTDEQRSALVKKRYKGWGRLSKKLLTGIVDENGQRIIDLMWNTDQNFMQIVNQPVFKEQIDQLNQKAITNDGMTLR
ERVESVLDDAYTSPQNKKAIWQVVRVVEDIVKAVGNAPKSI SIEFARNEGKGEITRSRRTQLQKLFEDQAHELK

DTSLTEELEKAPDLSDRYFYFTQGGKDMYTGDPIINFDEISTKYDIDHILPQSFVKDDSLDNRVLVSRAENNKSD
RVPAKLYAAKMKPYNQLLKQLITQRKFENLTMDVDQTIKYRSLGFVKRQLVETRQVIKLTANILGSMYQEAGTD
IIETRAGLTKQLREEFDLPKRVENDYHHAVDAYLTTFAGQYLNRYPKLRSFVYGEYMKFKHGSDDLKLRNFNF
HELMGDKSQGKVVDQQTGELITTRDEVADYFDWVINLKVMLISNETYEETGKYFDASHESSSLYLKNQNKSKLV
VPLKNKLQPEYYGAYTGITQGYMVILKLLDKKGGFGVYRIPRYAADILNKCHDEVAYRNKIAEIISSDPRAKSF
VVVPRVLKGTFLVDGEEKFILSSYRYKVNATQLILPVSDIKLIQDNFKALKKLNVMQTKKLEIYDNILRQVDKY
YKLYDINKFRAKLHDKRSKFVELDDFGQDASKEKVIKILRGLHFGSDLQNLKEIGFGTTPLGQFQVSEAGIRLSN
TAFIIFKSPTGLFNRKLYLKNLGRPAATKKAGQAKKKKGSYPYDVPDYAYPYDVPDYAYPYDVPDYA (SEQ ID
NO:85)。

15. 根据权利要求9所述的Cas9蛋白,其中所述嗜脉菌屑马赛-P3761 Cas9具有与以下至少80%相同的氨基酸序列:

MPKKKRKVGEEKNTYTIGLDIGTDSVGWAVVKDDLELVKKRMKVLGNTETNYIKKNLWGSLLFESGQTAKDR
RLKR二VARRRYERRRNLTELQKIFAPAIDEVDENFFRLNESFLVPEDKAFSKNPIFGTLGEDKTYTYKTYPTIY
HLRQHLADSEEKADVRLIYLALAHMIKYRGHFLIEGKLDTEHIAINENLEQFFESYNALFSEEPIELRKEELIAIE
NILREKNSRTVKEKRITSFLKDIGRANKQSPMAFITLIVGKKAKFKAAFNLEEEISLNLTDSDYDENLEILLNTI
GSDFADLFDHAQRVYNAVELAGILSGDVKNTHAKLSAQMVAMYERHKEQLKEYKSFIKANLPDQYDMTFVAPKDAQ
KKDLKGYAGYIDGNMSQDSFYKFVKDQLKEVPGSEKFLDSIEKEDFLRKQRSFYNGVIPNQVHLAEMEAILDRQEN
YYPWLKENREKIIISLLTFRIPYYVGPLADGQSEFAWLERKSDEKIKPWNFSDVVDLDRSAEKFIEQLIGRDTYLPD
EYVLPKKSLLIYQKYMVFNELTKIAYLDERQKRMNLSSEKKEIFETLFKKRSKVTEKQLVKFFENYLQIDNPTIFG
IEDAFNADYSTYVELAKVPGMKSMDDPDNEDLMEEIVKILTVFEDRKMRKQLEKYKERLSPEQIKELAKKHYTG
WGRLSKKLLVGIRDKETQKTILDYLVEDDNHSGGRQHLNRNLMQLINDDRSLFKKTI AELQ MIDPSADLYAQVQEI
AGSPAIKKGILLGLKIVDEIIRVMGEKPENIVIEMARENQTTARGKALSKRREAKIKEGLAALGSSLLKENLPGNA
DLSQRKIYLYYTQNGKDIYLDPEPLDFDRLSQYDEDDHIIPQSFTVDNSLDNLVLTNSSQNRGNKKDDVPSLEVVRQ
LAYWRSKLDAGLMTQRKFDNLTKAMRGGLTDKDRERFIQRQLVETRQITKNVAKLLDMRLNDKKDEAGNKIRETNI
VLLKSAMASEFRKMFRLYKRVRELNDYHHAHDAYLNAAIAINLLALYPYMADDFVYGEFRYKKKPAEKATYEKLRQ
WNLIKRFGEKQLFTP DHEDCWNKERDIKTIKKVMGYRQVNVVKAEEERTGMLFKETINGKTKGSRIPKIKDLPS
KYGGYIEEKMAYYAVISYEDKKKKPGKTIVGISIMDKKEFEYDSISYLGKLGFSNPVQIILKNYSLIAYPDGRRR
YITGATKTTKGKVELQKANQIAMEQDLVNFYHLKNYDEISHPESYAFVQSHTDYFDRLFDSIEHYTRRFLDAETN
INRLRRIYEEKKKDPVDIEALVASFIELLKLTSAGAPADFI FMGEAISRRRYNSMTGLFDGQVIYQSLTGLYETR
MRFEDGKRPAATKKAGQAKKKKGSYPYDVPDYAYPYDVPDYAYPYDVPDYA (SEQ ID NO:87)。

16. 根据前述权利要求中任一项所述的Cas9蛋白,其中所述Cas9蛋白的氨基酸序列包含SEQ ID NO:1、4、8、14、84或86中的至少一个、至少两个、至少三个、至少四个、至少五个、至少六个、至少七个、至少八个、至少九个或至少10个突变。

17. 根据权利要求16所述的Cas9蛋白,其中所述突变是氨基酸取代。

18. 根据前述权利要求中任一项所述的Cas9蛋白,其中所述Cas9蛋白具有切口酶活性。

18b. 根据权利要求18所述的Cas9蛋白,其中氨基酸位置处的切口酶突变对应于野生型SpCas9的一个或多个氨基酸10、12、17、762、840、854、863、982、983、984、986、987。

19. 根据权利要求16所述的Cas9蛋白,其中所述至少一个突变导致无活性的Cas9 (dCas9)。

20. 根据前述权利要求中任一项所述的Cas9蛋白,其中所述Cas9蛋白包含PAM相互作用结构域、HNH结构域和/或RuvC结构域中的至少一个氨基酸突变。

20b. 根据权利要求20所述的Cas9蛋白,其中氨基酸位置处的所述突变对应于SirCas9的RuvC结构域中的氨基酸14。

20c. 根据权利要求20所述的Cas9蛋白,其中氨基酸位置处的所述突变对应于EpeCas9的RuvC结构域中的氨基酸12。

20d. 根据权利要求20所述的Cas9蛋白,其中氨基酸位置处的所述突变对应于LfeCas9的RuvC结构域中的氨基酸9。

20e. 根据权利要求20所述的Cas9蛋白,其中氨基酸位置处的所述突变对应于PmaCas9的RuvC结构域中的氨基酸12。

20f. 根据权利要求20所述的Cas9蛋白,其中所述Cas9蛋白是超精确Cas9。

20g. 根据权利要求20所述的Cas9蛋白,其中所述Cas9蛋白包含相对于SpyCas9 (SEQ ID NO: 173) 对应于N692A、M694A、Q695A和/或H698A的突变。

20h. 根据权利要求20所述的Cas9蛋白,其中所述Cas9蛋白是高保真Cas9。

20i. 根据权利要求20所述的Cas9蛋白,其中所述Cas9蛋白包含相对于SpyCas9 (SEQ ID NO: 173) 对应于N467A、R661A、Q695A和/或Q926A的突变。

20j. 根据权利要求20所述的Cas9蛋白,其中所述Cas9蛋白是SuperFi-Cas9。

20k. 根据权利要求20所述的Cas9蛋白,其中对应于SpyCas9的Y1016、R1019、Y1010、Y1013、K1031、Q1027和/或V1018残基经突变为天冬氨酸。

21. 一种工程化的非天然存在的Cas9融合蛋白,其包含与SEQ ID NO: 1、4、8、14、84或86具有至少80%同一性的Cas9蛋白,并且其中所述Cas9蛋白与组蛋白去甲基化酶、转录活化因子或与脱氨酶融合。

21b. 根据权利要求21所述的工程化的非天然存在的Cas9融合蛋白,其进一步包含核定位序列(NLS)和/或FLAG、HIS或HA标签。

21c. 根据权利要求22所述的工程化的非天然存在的Cas9融合蛋白,其与SEQ ID NO: 2、5、9、15、85、87、95或96具有至少80%同一性。

22. 根据权利要求21所述的Cas9蛋白,其中所述Cas9蛋白与胞嘧啶脱氨酶或与腺苷脱氨酶融合。

23. 根据权利要求22所述的Cas9蛋白,其中所述Cas9蛋白与腺苷脱氨酶融合并且具有与以下项至少80%相同的氨基酸序列:

(a)

MPAAKRVKLDGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNRRVIGEGWNRAIGLHDPTAHAEIMA
LRQGGLVMQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITEGILA
DECAALLCRFFRMPRRVFNAQKKAQSSTDGSSGSETPGTSESATPESSGPKKKRKVGKPYSIGLAIGTNSVGWAV
VTDDYKVPAAKMKVLGNTDKQSIKKNLLGALLFDSGETAEATRLKRTARRRYTRRKNRLRYLQEIFTGEMNKVDEN
FFQRLDDSFVDEDKRGEHHPIFGNIAAEVKYHDDFPTIYHLRRHLADTSKKADLRLVYLALAHMIKFRGHFLYEG
DLKAENTDVQALFKDFVEEYDKTIEESHLSEITVDALSILTEKVS KSRLNLI AHYPTEKKNTLFGNLI ALSDL
HPNFKTNFQLSEDAKLQFSKDTYEEDLEGFLGEVGDEYADLFASAKNLYDAILLSGILTVDDNSTKAPLSASMVKR
YEEHQKDLKKLKDFIKVNAPDQYNAIFKDKNKKGYASYIESGVKQDEFYKYLKGILLKINGSGDFLDKIDREDFLR

KQRTFDNGSIPHQIHLQEMHAILRRQGEHYPLKENQDKIEKILTFRIPYYVGPLARKGSRFAWAEYKADEKITPW
NFDDILDKEKSAEKFITRMTLNDLYLPPEEKVLPHKSPLYEAFTVYNELTKVKYVNEQGEAKFFDTNMKQEIFDHVF
KENRKVTKDKLLNYLNKEFEFRIVNLTGDKENKAFNSSLGTYHDLRKILDKSFLDDKANেকTIEDIIQTLTLFE
DREMIRQRLQKYSDIFTKAQLKKLERRHYTGWRLSYKLINGIRNKENKKTILDYLIDDGYANRNFMLINDDALS
FKEEIARAQIIDDVDDIANVVDLPGPSAIIKKGILQSVKIVDELVKVMGHNPAIIIEMARENQTTDKGRNSQQR
LKLQDSLKNLNDPNVNIKNVENQQQLQNDRLFLYYIQNGKDMYTGETLDINNLSQYDIDHIIPQAFIKDNSLDNRVL
TRSDKNRGKSDDVPSIEVVHEMKSFWSKLLSVKLITQRKFDNLTKAERGGLTEEDKAGFIKRQLVETRQITKHVAQ
ILDERFNTEFDGNKRRIRNVKIIITLKSNLVSNFRKEFELYKVREINDYHHAHDAYLNAVVGNALLLKYPQLEPEFV
YGEYPKYNSYRSRKSATEKFLFYSNILRFFKKEDIQTNEDEGEIAWNKEKHILRKVLSYPQVNIKKTEEQTGGF
SKESILPKGESDKLIPRKTNSYWDPKKYGGFDSPVVAYSILVFADVEKGKSKLRKVQDMVGITIMEKKRFEKNP
VDFLEQRGYRNVRLKIKLPKYSLELENKRRRLASAKELQKGNELVIPQRFTTLLYHSYRIEKDYEPHEHYV
EKHKDEFKELLEYISVFSRKYVLADNNLTKEMLFSKNKDAEVSSLAKSFISLLTFTAFGAPAAFNFFGENIDRKR
YTSVTECLNATLIHQSIITGLYETRIDLKLGEDGKRPAATKKAGQAKKKKGSYPYDVPDYAYPYDVPDYAYPYDVP
DYA (SEQ ID NO:20) ;

(b)

MPAAKRVKLDGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMA
LRQGGLVMQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITEGILA
DECAALLCRFFRMPRRVFNAQKKAQSSTDGSSGSETPGTSESATPESSGPKKKRKVGAKNKDIRYSIGLAIGTNSV
GWAVMDEHYELLKKGNNHMGWSRLFDAEPAATRARSIRRRYNKRRERIRLLRDLLGDMVMEVDPTFFIRLLNV
SFLDEEDKQKNLGNDYKDNYNLIEKDFNDKTYDYKYPTIYHLRKELCENKEKADPRLIYLALHHIVKYRGNFLKE
GQSFAKVYEDIEEKLDNTLKKFMSLNDLNDLFVDNDINSMITVLSKIYQRSKKADDLLKIMNPTKEERAAYKEFTK
ALVGLKFNVS KMILAQEVKKDDKDIELDFSNDYDSTVDGLQAELEGEYIEFIEMLHSINSWVELQDILGNNSTISA
AMVERYEEHKNLRLVLKKVIREELPKYNEVFREDNPKLHNYLGYIKYPKNTPVEEFYIEYIKRLLAKVDTGEAREI
LERIDLEKFMLKQNSRTNGSIPYQMQKDEMIQIIDNQSVYYPQLKENREKLISILEFRIPYYFGPLNTHSEFAWIK
KFEDKQKERILPWNVDQIVDIDATAEGFIERMQNTGTYPDPKPVMAKNSLTVSKFEVLNENKIRINGKLIPVETK
KELLSDLFMKNKTITDKKLKDWLVTHQYYDTNEELKIEGYQKDLQFSTSLAPWIDFTKIFGEINASNYQLIEKIIY
DISIFEDKKILKRLKKVYQLDDLLVDKILKLNLTGWSRLSEKLLTGIKSKNSKETILSILENSNMNLMEIINDES
LGFKQIIIEESNKKDIEGPFYDEVKLAGSPAIIKRGIWQALLVVQEITKFMKHEPSHIYIEFAREEQEKVRESRI
AKLQKIYKDLNLQTKEDQLVYESLKKEDAKKKIDTDALYLYLQMGKSMYSGKPLDIDKLSTYHIDHILPRSLIKD
DSLNDNRVLVLPKENEWKLDSETVPFEIRNKMMGFQKLHENGLSNKKFFSLIRTDNEKDKKRFINRQLVETRQI
IKNAVAVIINDHYTNTNVTVRAELSHQFRERYKIYKNRDLNDLHHAHDAYIACILGQFIHQNFNMVDNMIYGQYK
KNYKKDVQEHNNYGFILNSMNIHFNDNSVIWDPSYIGKIKSCFCYKDVYVTKKLEQNDAKLFDLTILPSDKNSE
NGVTAKIPVKNYRKDVNKGFGSGDAPIMLAIEADKGKKHVRQVIAFPLRLKNYNDEERIKFIEKEKNLKNVKIL
TEVKKNLILINHQQFFITGTNELVNATQLKLSAKNTKNLNLVDANKHNKLESIDDANFNEVIQELICKLQEPYIY
SRYSIGKEFEDSYEKINAVTKQDKLYIIIEYLIAIMSAKATQGYIKPELAREIGTNGKNKGRIKSFTIDLNKTTFI
STSVTGLFSKKYKLGKRPAATKKAGQAKKKKGSYPYDVPDYAYPYDVPDYAYPYDVPDYA (SEQ ID NO:6) ;

(c)

MPAAKRVKLDGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMA
LRQGGLVMQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITEGILA

DECAALLCRFFRMPRRVFNAQKKAQSSTDGSSGSETPGTSESATPESSGPKKKRKVGSI INFQRRGLMETQASNQL
ISSHLKGYPIDYFVGLAIGTSSVGWAVTNKAYELLKFRSHKMWGSRLFDEGESAVARRGFRSMRRRLERRKLRLK
LLEELFADAMAQVDPTFFMRLRESKYHYEDKTTGHSSKHILFIDKNYNDQDYFKEYPTVYHLRSELMKSGTDDIRK
LFLAVHHILKYRGNFYEGATFDSNASTLDDVIKQALENITFNCFCNSAIISSIGQILMEAGKTKSDKAKAIEHLV
DTYIATDVTDTSSKTQKDQVKEDKKRLKAFANLVGLNASLIDLFGSVEELEEDLKKLQITGDTYDDKRDELAKAW
SDEIYIIDCKSVYDAI ILLSIKEPGLTISESKVKAFNKHKDDLA ILLSLLKSDRSIYNTMFKVDEKGLHNYVHYI
KQGRTEETSCNREDFYKYTKKIVEGLSDSKDKEYILSQIELQILLPLQRIKDNGVIPYQLHLEELKAILAKCGPKF
PFLNEVADGFSVAEKL IKMLEFRIPYYVGPLNTHHNVNDNGGFAVVRKASGRVTPWNFDDKIDREKSAAAFIKNLT
NKCTYLLGEDVLPKSSLLYSEFMLLNELNNVRIDGKPLEKVVKEHL IEAVFKQDHHKMTKNRIEQFLKDNGYISET
HKHEITGLDGEIKNDLASYRDMVRILGDGFRSMAEEIITDITIFGESKKMLRETLRKKFASCLDDEAIKKLTKLR
YRDWGRLSQKLLNGIEGCDKAGDGTPEII ILMRNFSYNLMELLGDKFSFMERIQEINAKLTEGQIVNPHDIIDDL
ALSPAVKRAVWQALRIVDEVAHIKKALPARIFVEVTRSNNKNEKKKKDSRQKRLSDLYAAIKKDDVLLNGLNNEIFG
ELKSSLAKEYDDAALRSKKLYLYYTQMGRCAITGEI IELSLNNTDNYDIDHIYPRSLTKDSDFDNLVLCRTANAQK
SDAYPISEEIQKTQKPFWTFKQQGLISERKYERLTRITPLTADDLSGFARQLVETNQSVAATTLRRLYPGVD
VVFVKAENVTDFRHDNNFIKVRSLNHHHHAKDAYLNIIVGVNYHERFTRNFRAFFKKNGANRTYNLAKMFNYDVNC
TNAKDGAWDVKTSMDTVKKMDSNDVRVTKRLLEQTGALADATIYKATVAGKAKDGAYIGMKTSSVFADVSKYG
GMTKIKNAYSIIVQYTGKKGEVIKEIVPLPIYLTNRNTTDQDLINVASIIPQAKDISIIYGKLCINQLVKVNGFY
YYLGGKTNSKFCIDNAIQVIVSNEWIPYLVLEKFNNMRKDNKDLKANVVSTRALDNKHTIEVRIVEEKNIEFFDY
LVSKLKMPIYQKMKGNKAAELSEKGYGLFKKMSLEEQSIHLIELLNLLTNQKTTFEVKPLGITASRSTVGSKISNQ
DEFKVINESITGLYSNEVTIVGKRPAATKAGQAKKKKGSYPYDVPDYAYPYDVPDYAYPYDVPDYA (SEQ ID
NO:10) ;

(d)

MPKKKRKVSII INFQRRGLMETQASNQLISSHLKGYPIDYFVGLAIGTSSVGWAVTNKAYELLKFRSHKMWG
SRLFDEGESAVARRGFRSMRRRLERRKLRLKLEELFADAMAQVDPTFFMRLRESKYHYEDKTTGHSSKHILFIDK
NYNDQDYFKEYPTVYHLRSELMKSGTDDIRKLFLAVHHILKYRGNFYEGATFDSNASTLDDVIKQALENITFNC
DCNSAIISSIGQILMEAGKTKSDKAKAIEHLVDTYIATDVTDTSSKTQKDQVKEDKKRLKAFANLVGLNASLIDL
GSVEELEEDLKKLQITGDTYDDKRDELAKAWSDEIYIIDCKSVYDAI ILLSIKEPGLTISESKVKAFNKHKDDLA
ILLKSLKSDRSIYNTMFKVDEKGLHNYVHYIKQGRTEETSCNREDFYKYTKKIVEGLSDSKDKEYILSQIELQILL
PLQRIKDNGVIPYQLHLEELKAILAKCGPKFPFLNEVADGFSVAEKL IKMLEFRIPYYVGPLNTHHNVNDNGGFAWA
VRKASGRVTPWNFDDKIDREKSAAAFIKNLTNKCTYLLGEDVLPKSSLLYSEFMLLNELNNVRIDGKPLEKVVKEH
LIEAVFKQDHHKMTKNRIEQFLKDNGYISETHKHEITGLDGEIKNDLASYRDMVRILGDGFRSMAEEIITDITIF
GESKKMLRETLRKKFASCLDDEAIKKLTKLRYRDWGRLSQKLLNGIEGCDKAGDGTPEII ILMRNFSYNLMELLG
DKFSFMERIQEINAKLTEGQIVNPHDIIDDLALSPAVKRAVWQALRIVDEVAHIKKALPARIFVEVTRSNNKNEKKK
KDSRQKRLSDLYAAIKKDDVLLNGLNNEIFGELKSSLAKEYDDAALRSKKLYLYYTQMGRCAITGEI IELSLNNTD
NYDIDHIYPRSLTKDSDFDNLVLCRTANAQKSDAYPISEEIQKTQKPFWTFKQQGLISERKYERLTRITPLTADD
LSGFARQLVETNQSVAATTLRRLYPGVDVVFVKAENVTDFRHDNNFIKVRSLNHHHHAKDAYLNIIVGVNYHE
RFTRNFRAFFKKNGANRTYNLAKMFNYDVNCTNAKDGAWDVKTSMDTVKKMDSNDVRVTKRLLEQTGALADATI
YKATVAGKAKDGAYIGMKTSSVFADVSKYGGMTKIKNAYSIIVQYTGKKGEVIKEIVPLPIYLTNRNTTDQDLIN
YVASIIPQAKDISIIYGKLCINQLVKVNGFYYYLGGKTNSKFCIDNAIQVIVSNEWIPYLVLEKFNNMRKDNKDL

KANVVSTRALDNKHTIEVRIVEEKNIEFFDYLVSCLKMPIYQKMKGNKAAELSEKGYGLFKKMSLEEQSIHLIELL
NLLTNQKTTFEVKPLGITASRSTVSGKISNQDEFKVINESITGLYSNEVTIVKRPAATKKAGQAKKKKSGSETPGT
SESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGG
LVMQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAA
LLCRFFRMPRRVFNAQKKAQSSTDPAAKRVKLDGSPYDVPDYAYPYDVPDYAYPYDVPDYA (SEQ ID NO:11);

(e)

MPAAKRVKLDGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMA
LRQGGLVMQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITEGILA
DECAALLCRFFRMPRRVFNAQKKAQSSTDGSSGSETPGTSESATPESSGPKKKRKVGTKVDYYIGLAIGTSSVGW
AVTDEAYNVLFNSKKMWGVRLFDDAKTAEERRGQRGARRRLDRKKERLSLLQDFFAEVAKVDPNFFLRDLNSDL
YMEDKDQKLKSKYTLFNDKDFDKDNFHKKYPTIHLLMDLIEDDSKKDIRLVYLACHYLLKNRGHFIFEGQKFDTK
SSFENSLNELKVHLNDEYGLDLEFDNENLINILTDPKLNKTAKKKELKSVIGDTKFLKAVSAIMIGSSQKLVDLFE
NPEDFDDSAIKSVDFSTTSFDDKYSDYELALGDKIALVNILKEIYDSSILENLLKEADKSKDGNKYISNAFVKKN
KHGQDLKEFKRLVRQYHKSAYFDIFRSEKVNNDYVSYTKSSISNNKRVKANKFTDQEAIFYKFAKKHLETIKYKINK
VNGSKADLELIDGMLRDMEFKNFMPKIKSSDNGVIPYQLKLMELNKILENQSKHHEFLNVSDEYGSVCDKIASIME
FRIPYYVGPLNPNSKYAWIKKQKDSEITPWNFKDVVDLSSREEFIDSLIGRCTYLKDEKVLPAKSLLYNEYMVLN
ELNNLKLNDLPITEEMKKKIFDQLFKTRKKVTLKAVANLLKKEFNINGEILLSGTDGDFKQGLNSYNDFKAIVGDK
VDSDDYRDKIEEIIKLIVLYGDDKSYLQKKIKAGYGYFTDSEIKKMAGLNYKDWGRLSKKLLTGLEGANKITGER
GSI IHFMREYNLNLMELEMSASFTFTEEIQKLNVPDDRKLSEYEMVDELYLSPSVKRMWQSLRIVDEIKNIMGTDSK
KIFIEMARGKEEVKARKESRKNQLLFYKDGKKAFISEIGEERYSYLLSEIEGEEENKFRWDNLYLYTQLGRCMY
SLEPIDISELSSKNYDQDHIYPKSKIYDSSIENRVLVKKDLNSKKGNSYIPDEILNKNCYAYWKILYDKGLIGQ
KKYTRLTRRTGFTDDELVQFISRQIVETRQATKETANLLKTICKNSEIVYSKAENASRFRQEFDIVKRAVNDLHH
MHDAYINIIVGNVYNTKFTKDPNMFVKKQEKARSYNLENMFKYDVKRGGYTAWIADDEKGTVKNASIKRIRKELEG
TNYRFRMNYIESGALFNATLQRKNKGSRLPKDKGPKSSIEKYGGYTNINKACFAVLDIKSKNKIERKLMVEREI
YAKQKNDKKLSDEIFSKYLKDRFGIEDYRVVYPVVKMRTLLKIDGSYFFITGGSCKTLELRSALQLILPKKNEWAI
KQIDKSSENDYLTIERIQDLTEELVYNTFDIIVNKFTSVFKKSFLNLFQDDKIENIDFKFKSMDFKCKTLLML
VKAIRASGVRQDLKSIDLKS DYGRLLSSKTNNIGNYQEFKIINQSITGLFENEVDLLKLGRPAATKKAGQAKKKKG
SYPYDVPDYAYPYDVPDYAYPYDVPDYA (SEQ ID NO:16);

(f)

MPKKKRKVTKVVDYYIGLAIGTSSVGWAVTDEAYNVLFNSKKMWGVRLFDDAKTAEERRGQRGARRRLDRK
KERLSLLQDFFAEVAKVDPNFFLRDLNSDLYMEDKDQKLKSKYTLFNDKDFDKDNFHKKYPTIHLLMDLIEDDS
KKDIRLVYLACHYLLKNRGHFIFEGQKFDTKSSFENSLNELKVHLNDEYGLDLEFDNENLINILTDPKLNKTAKKK
ELKSVIGDTKFLKAVSAIMIGSSQKLVDLFENPEDFDDSAIKSVDFSTTSFDDKYSDYELALGDKIALVNILKEIY
DSSILENLLKEADKSKDGNKYISNAFVKKNKHGQDLKEFKRLVRQYHKSAYFDIFRSEKVNNDYVSYTKSSISNN
KRVKANKFTDQEAIFYKFAKKHLETIKYKINKVNGSKADLELIDGMLRDMEFKNFMPKIKSSDNGVIPYQLKLMELN
KILENQSKHHEFLNVSDEYGSVCDKIASIMEFRIPYYVGPLNPNSKYAWIKKQKDSEITPWNFKDVVDLSSREEF
IDSLIGRCTYLKDEKVLPAKSLLYNEYMVLNELNNLKLNDLPITEEMKKKIFDQLFKTRKKVTLKAVANLLKKEFN
INGEILLSGTDGDFKQGLNSYNDFKAIVGDKVDSDDYRDKIEEIIKLIVLYGDDKSYLQKKIKAGYGYFTDSEIK
KMAGLNYKDWGRLSKKLLTGLEGANKITGERGSI IHFMREYNLNLMELEMSASFTFTEEIQKLNVPDDRKLSEYEMVD

ELYLSPSVKRMLWQSLRIVDEIKNIMGTDSSKIFIEMARGKEEVKARKESRKNQLLKFYKDGKKAFISEIGEERYSYLLSEIEGEEENKFRWDNLYLYTQLGRCMYSLEPIDISELSSKNIIDQDHIYPKSKIYDDSIENRVLVKKDLNSKKGNSYPIPDEILNKNCYAYWKILYDKGLIGQKKYTRLTRRTGFTDDELVQFISRQIVETRQATKETANLLKTICKNSEIVYSKAENASRFRQFEDIVKCRVNDLHHMHDAYINIIVGNVYNTKFTKDP MNFVKKQEKARSYNLENMFKYDVKRGGYTAWIADDEKGT VKNASIKRIRKELEGTNYRFRTRMNYIESGALFNATLQRKNKGSRLPKDKGPKSSIEKYGGYTNINKACFAVLDIKSKN KIERKLMPVEREIIYAKQKNDKLSDEIFSKYLKDRFGIEDYRVVYPVVKMRTLLKIDG SYYFITGGSDKTLELRSALQLILPKKNEWAIKQIDKSSENDYLTIERIQDLTEELVYNTFDIIVNKFKTSVFKKSLNLNFQDDKIENIDFKFKSMD FKECKTLLMLVKAI RASGVRQDLKSIDLKS DYGR LSSKTNNIGNYQEFKIINQSIGTGLFENEVDLLKLKRPAA TKKAGQAKKKKSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLV LNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGV RNAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTDPAAKRVKLDG SYPYDVPDYAYPYDVPDYA (SEQ ID NO:17);

(g)

MPAAKRVKLDGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLV LNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGV RNAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTDGSSGSETPGTSESATPESSGPKKKRKVGKEYHIGLAIGTSSIGWAVTDSQFKLMRIKGKTAIGVRLFE EGKTA AERRTFRTTRRRLLKRRKWRLHYLDEIFAPHLQEVDENFLRRLKQSNIHPE DPAKNQAFIGKLLFPDLLKKNERGYPTLIKMRDELPVEQRAHYPVTNIYKLREAMINEDRQFDLREVYLAVHHIVKYRGHFLNNASVDKFKVGRIDFDKSFNVLNEAYEELQNGEGSFTIEPSKVEKIGQLLLDTKMRKLD RQKAVAKLLEVKVADKEETKR NKQIATAMSKLVLYKADFATVAMANGNEWKIDLSSETSEDEIEKFREELSDAQN DILTEITS LFSQIMLNEIVPNGMSISESMMDRYW THERQLAEVKEYLATQPASARKEFDQVYNKYIGQAPKEKGF DLEKGLKKILSKKENWKEIDELLKAGDFLPKQRTSANGVIPHQMHQQELDRIIEKQAKYYPWLATENPATGERDRHQAKYELDQLVSFRIPYYVGPLVTPEVQKATSGAKFAWAKRKEDGEITPWNLWDKIDRAESAEAFIKRMTVKD TYLLNEDVLPANSLLYQKYNVLNELNNVRVNGRRLSVGIKQDIYTELFKKKTKVAGDVASLVMAKTRGVNKPSVEGLSDPKKFNSNLATYLDLKSIVGDKVDDNRYQMDLENIIEWRSVFEDGEIFADKLTEVEWLTDEQRSALVKKRYKGWGRLSKLLTGIVDENGQRIIDL MWNTDQNF MQIVNQPVFKEQIDQLNQKAITNDGMTLRERVESVLDDAYTSPQNKKAIWQVVRVEDIVKAVGNAPKSSISIEFARNEGNGGEITRSRRTQLQKLFEDQAHELVKDTSLTEELEKAPDLSDRYFYFTQGGKDMYTGDPINFDEISTKYDIDHILPQS FVKDDSLDNRLVSRAENKKS DRVP AKLYAAKMKPYWNQLLKQGLITQRKFENLTMDVDQTIKYRSLGFVKRQLVETRQVIKLTANILGSMYQEAGTDI IETRAGLTKQLREEFDLPKREVNDYHHAVDAYLTTFAGQYLNRRYPKLR SFFVYGEYMKFKHGS DLKLRNFNF FHELMEGDKSQGKVVDQQTGELITTRDEVADYFDWVINLKVMLISNETYEETGKYFDASHESSLYLKNQNKSKLVVPLKNKLQPEYYGAYTGITQGYMVLKLLDKKGGFGVYRIPRYAADILNKCHDEVAYRNKIAEIISSDPRAPKSFEVVVPRVLKGTFLVDGEEKFILSSYRYKVNATQLILPVSDIKLIQDNFKALKKLN VEMQTKKLEIYDNILRQVDKYYKLYDINKFRAKLHDGRSKFVELDDFGQDASK EKVI I KILRGLHFGSDLQNLKEIGFGTTPLGQFQVSEAGIRLSNTAFIIFKSPTGLFNRKLYLKNLGRPAATKKA GQAKKKKGSYPYDVPDYAYPYDVPDYA (SEQ ID NO:88);

(h)

MPKKKRKVKEYHIGLAIGTSSIGWAVTDSQFKLMRIKGKTAIGVRLFE EGKTA AERRTFRTTRRRLLKRRKWRLHYLDEIFAPHLQEVDENFLRRLKQSNIHPE DPAKNQAFIGKLLFPDLLKKNERGYPTLIKMRDELPVEQRAHYPVTNIYKLREAMINEDRQFDLREVYLAVHHIVKYRGHFLNNASVDKFKVGRIDFDKSFNVLNEAYEELQNGEGSFTI

EPSKVEKIGQLLLDTKMRKLDQRQKAVAKLLEVKVADKEETKRNKQIATAMSKLVLGYKADFATVAMANGNEWKIDL
SSETSEDEIEKFREELSDAQNDILTEITSLFSQIMLNEIVPNGMSISESMMDRYWOTHERQLAEVKEYLATQPASAR
KEFDQVYNKYIGQAPKEKGFDLEKGLKKILSKKENWKEIDELLKAGDFLPKQRTSANGVIPHQMHQQELDRIIEKQ
AKYYPWLATENPATGERDRHQAKYELDQLVSFRIPYYVGPLVTPVEVQKATSGAKFAWAKRKEDGEITPWNLWDKID
RAESAEAFIKRMTVKDITYLLNEDVLPANSLLYQKYNVLNELLNNRVNVRRLSVGIKQDIYTELFKKKKTKVAGDVA
SLVMAKTRGVNKPVSVEGLSDPKKFNSNLATYLDLKSIVGDKVDDNRYQMDLENIIEWRSVFEDGEIFADKLTEVEW
LTDEQRSALVKKRYKGWGRLSKKLLTGIVDENGQRIIDL MWNTDQNFMIQVNPVFKEQIDQLNQKAITNDGMTLR
ERVESVLDDAYTSPQNKKAIWQVVRVVEDIVKAVGNAPKSISIEFARNEGNGKEITRSRRTQLQKLFEDQAHELVK
DTSLTEELEKAPDLSDRYYFYFTQGGKDMYTGDPI NFDEISTKYDIDHILPQS FVKDDSLDNRVLSRAENNKSD
RVPAKLYAAKMKPYNQLLKQGLITQRKFENLTMDVDQTIKYRSLGFVKRQLVETRQVIKLTANILGSMYQEAGTD
I IETRAGLTQKLEEFDLPKVREVNDYHHAVDAYLTTFAGQYLNRRYPKLSFFVYGEYMKFKHGSCLKLRNFNF
HELMEGDKSQGKVVDQQTGELITTRDEVADYFDWVINLKVMLISNETYEETGKYFDASHSSLYLKNQNKSKLV
VPLKNKLQPEYYGAYTGITQGYMILKLLDKKGGFGVYRIPRYAADILNKCHDEVAYRNKIAEIISSDPRAKPSFE
VVVPRVLKGTFLVDGEEKFILSSYRYKV NATQLILPVSDIKLIQDNFKALKKLVNEMQTKKLEIYDNILRQVDKY
YKLYDINKFRAKLHDGRSKFVELDDFGQDASKEKVIKILRGLHFGSDLQNLKEIGFGTTPLGQFQVSEAGIRLSN
TAFIIFKSPTGLFNRKLYLKNLKRPAATKKAGQAKKKKSGSETPGTSESATPESSGSEVEFSHEYWMRHATLAKR
ARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLYDATLYVTFEPCVMCAGAMIHS
RIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTDPAAKRV
KLDGSYPYDVPDYAYPYDVPDYAYPYDVPDYA (SEQ ID NO:89);或

(i)

MSEVEFSHEYWMRHATLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQN
YRLYDATLYSTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLCRF
FRMPRRVFNAQKKAQSSTDSSGSSGGSSGSETPGTSESATPESSGSSGGSGKPYSIGLAIGTNSVGWAVVTDYK
VPAKKMKVLGNTDKQSIKKNLLGALLFDSGETAEATRLKRTARRRYTRRNRLRYLQEIFTGEMNKVDENFFQRLD
DSFLVDEDKRGEHHPIFGNIAAEVKYHDDFPTIYHLRRHLADTSKKADRLVYLALAHMIKFRGHFLYEGDLKAEN
TDVQALFKDFVEEYDKTIEESHLSEITVDALSILTEKVS KSRLNLI AHYPTEKKNTLFGNLIALSLDLHPNFKT
NFQLSEDAKLQFSKDTYEEDLEGFLGEVGDEYADLFASAKNLYDAILLSGILTVDDNSTKAPLSASMVKRYEEHQK
DLKKLKDFIKVNAPDQYNAIFKDKNKKGYASYIESGVKQDEFYKYLKGI LKINGSGDFLDKIDREDFLRKQRTFD
NGSIPHQIHLQEMHAILRRQGEHYPLKENQDKIEKILTFRIPYYVGPLARKGSRFAWAEYKADEKITPWNFDDIL
DKEKSAEKFITRMTLNDLYLP EKVLPKHSPLYEFTVYNELTKVKYVNEQGEAKFFDTNMKQEIFDHVFKENRKV
TKDKLLNYLNKEFEFRIVNL TGLDKENKAFNSSLGTYHDLRKILDKSFLDDKAN EKTIEDIIQTLTLFEDREMIR
QRLQKYSDFITKAQLKKLERRHYTGWGRLSYKLINGIRNKENKKTILDYLI DDGYANRNFMLINDDALSFKEEIA
RAQIIDDVDDIANVVHDLPGSPA I KKGILQSVKIVDELVKVMGHN PANII IEMARENQTTDKGRRNSQQRLKLQD
SLKNLDNPVNIKNVENQQQLQNDRLFLYYIQNGKDMYTGETLDINNLSQYDIDHIIPQAFIKDNSLDNRVLTRSDKN
RGKSDDVPSIEVVHEMKSFWSKLLSVKLITQRKFDNLTKAERGGLTEEDKAGFIKRQLVETRQITKHVAQILDERF
NTEFDGNKRRI RNVKIITLKS NLVSNFRKEFELYKVREINDYHHAHDAYLNAVVG NALLLKYPQLEPEFVYGEYPK
YNSYRSRKSATEKFLFYSNILRFFKKEDIQTNEDGEIAWNKEKHIKILRKVLSYPQVNIVKKTEEQTGGFSKESIL
PKGESDKLIPRKTKNSYWDPKKYGGFMQPVVAYSILVFADVEKGKSKKL RKVQDMVGITIMEKKRFEKNPVDFLEQ
RGYRNVRL EKI IKLPKYSLFELENKRRRL LASAKFLQKGNELVIPQRFTTLLYHSYRIEKDYEP EHYVEKHKDE

FKELLEYISVFSRKYVLADNNLTKIEMLFSSKNKDAEVSSLAKSFISLLTFTAFGAPRAFNFGENIARKEYRSVTE
CLNATLIHQSIITGLYETRIDLSKLGEDGEGADKRTADGSEFESPCKKKRKV (SEQ ID NO:98);

(j)

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVNLRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQN
YRLYDATLYSTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLCRF
FRMPRRVFNAQKKAQSSDSSGSSGGSSGSETPGTSESATPESSGGSSGGSGKPYSIGLAIGTNSVGWAVVTDYK
VPAKKMKVLGNTDKQSIKKNLLGALLFDSGETAEATRLKRTARRRYTRRKNRLRYLQEIFTGEMNKVDENFFQRLD
DSFLVDEDKRGEHHPIFGNIAAEVKYHDDFPTIYHLRRHLADTSKKADLRLVYLALAHMIKFRGHFLYEGDLKAEN
TDVQALFKDFVEEYDKTIEESHLSEITVDALSILTEKVS KSRLNLI AHYPTEKKNTLFGNLI ALSLDLHPNFKT
NFQLSEDAKLQFSKDTYEEDLEGFLGEVGDEYADLFASAKNLYDAILLSGILTVDDNSTKAPLSASMVKRYEEHQK
DLKKLKDFIKVNAPDQYNAIFKDKNKKGYASYIESGVKQDEFYKYLKGILLKINGSGDFLDKIDREDFLRKQRTFD
NGIIPHQIHLQEMHAILRRQGEHYPFLKENQDKIEKILTFRIPYYVGPLARKGSRFAWAEYKADEKITPWNFDDIL
DKEKSAEKFITRMTLNDLYLPEEKVLPKHSPLYEAFTVYNELTKVKYVNEQGEAKFFDTNMKQEIFDHVFKENRKV
TKDKLLNYLNKEFEFRIVNL TGLDKENKAFNSSLGTYHDLRKILDKSFLDDKAN EKTIEDIIQTTLTFEDREMIR
QRLQKYSDFITKAQLKKLERLHYTGWGRLSYKLINGIRNKENKKTILDYLDGYNRNFMLINDDALSFKEEIA
RAQIIDDVDDIANVVHDLPGSPAIIKKGILQSVKIVDELVKVMGHN PANII IEMARENQTTDKGRRNSQQRLKLLQD
SLKNLDNPVNIKNVENQQQLQNDRLFLYYIQNGKDMYTGETLDINNLSQYDIDHIIPQAFIKDNSLDNRVLTRSDKN
RGKSDDVPSIEVVHEMKSFWSKLLSVKLITQRKFDNLTKAERGGLTEEDKAGFIKRQLVETRQITKHVAQILDERF
NTEFDGNKRRIRNVKIIITLKS NLVSNFRKEFELYK VREINDYHHAHDAYLNAVVG NALLLKYPQLEPEFVYGEYPK
YNSYRSRKSATEKFLFYSNILRFFKKEDIQTNEDEGEIAWNKEKHIKILRKVLSYPQVNI VKKTEEQTGGFSKESIL
PKGESDKLIPRKTKN SYWDPK KYGGFMQPVVAYSILVFADVEKGKSKKL RKVQDMVGITIMEKKRFEKNPVDFLEQ
RGYRNVRL EKI IKLPKYSLFELENKRRRL LASAKFLQKGNELVIPQRFTTLLYHSYRIEKDYEP EHYVEKHKDE
FKELLEYISVFSRKYVLADNNLTKIEMLFSSKNKDAEVSSLAKSFISLLTFTAFGAPRAFNFGENIARKEYRSVTE
CLNATLIHQSIITGLYETRIDLSKLGEDGEGADKRTADGSEFESPCKKKRKV (SEQ ID NO:99)。

24. 根据权利要求22所述的Cas9蛋白,其中所述Cas9蛋白与胞嘧啶脱氨酶融合并且具有与以下项至少80%相同的氨基酸序列:

(a)

MPAAKRVKLDTSEKGPSTGDPTLRRRIESWEFDVFYDPRELRKETCLLYEIKWGMSRKIWRSSGKNTTNHVE
VNFIKKFTSERRFHSSISCSITWFLSWSPCWECSQAIREFLSQHPGVTLVIIYVARLFWHMDQRNRQGLRDLVNSGV
TIQIMRASEYYHCWRNFVNYPGDEAHWPQYPPLWMLYALELHCIIISLPPCLKISRWRQNLAFRRLHLQNCHY
QTIPPHILLATGLIHPSVTWRLKSGGSSGGSSGSETPGTSESATPESSGGSSGGSPCKKKRKVGKPYSIGLAIGTN
SVGWAVVTDYKVPAAKKMKVLGNTDKQSIKKNLLGALLFDSGETAEATRLKRTARRRYTRRKNRLRYLQEIFTGEM
NKVDENFFQRLDSDFLVDEDKRGEHHPIFGNIAAEVKYHDDFPTIYHLRRHLADTSKKADLRLVYLALAHMIKFRG
HFLYEGDLKAENTDVQALFKDFVEEYDKTIEESHLSEITVDALSILTEKVS KSRLNLI AHYPTEKKNTLFGNLI
ALS LD LHPNFKT NFQLSEDAKLQFSKDTYEEDLEGFLGEVGDEYADLFASAKNLYDAILLSGILTVDDNSTKAPLS
ASMVKRYEEHQKDLKKLKDFIKVNAPDQYNAIFKDKNKKGYASYIESGVKQDEFYKYLKGILLKINGSGDFLDKID
REDFLRKQRTFDNGSIPHQIHLQEMHAILRRQGEHYPFLKENQDKIEKILTFRIPYYVGPLARKGSRFAWAEYKAD
EKITPWNFDDILDKEKSAEKFITRMTLNDLYLPEEKVLPKHSPLYEAFTVYNELTKVKYVNEQGEAKFFDTNMKQE
IFDHVFKENRKVTKDKLLNYLNKEFEFRIVNL TGLDKENKAFNSSLGTYHDLRKILDKSFLDDKAN EKTIEDIIQ

TLTLFEDREMIRQRLQKYSDIFTKAQLKKLERRHYTGWGRLSYKLINGIRNKENKKTILDYLIDDDGYANRNFMLI
NDDALSFKEEIIARAQIIDDVDDIANVVHDLPGSPAIIKKGILQSVKIVDELVKVMGHN PANII IEMARENQTTDKGR
RNSQQRLKLLQDSLKNLDPVNIKNVENQQQLQNDRLFLYYIQNGKDMYTGETLDINNLSQYDIDHIIPQAFIKDNS
LDNRVLTRSDKNRGKSDDVPSIEVVHEMKSFWSKLLSVKLITQRKFDNLTKAERGGLTEEDKAGFIKRQLVETRQI
TKHVAQILDERFNTEFDGNKRRIRNVKIIITLKSNLVSNFRKEFELYKVVREINDYHHAHDAYLNAVVGNAALLKYPQ
LEPEFVYGEYPKYNSYRSRKSATEKFLFYSNILRFFKKEDIQTNEDEI IAWNKEKHIKILRKVLSYPQVNI VKKTE
EQTGGFSKESILPKGESDKLIPRKTNSYWDPKKYGGFDSPVVAYSILVFADVEKGKSKKLKRVQDMVGITIMEKK
RFEKNPVDLFLEQGRYRNVRLKIIKLPKYSLFELENKRRRLASAKELQKGNELVIPQRFTTLLYHSYRIEKDYEP
EHREYVEKHKDEFKELLEYSVFSRKYVLADNNLTKIEMLFSSKNKDAEVSSSLAKSFISLLTFTAFGAPAAFNFFGE
NIDRKRYTSVTECLNATLIHQSIITGLYETRIDLSKLGEDGKRPAATKKAGQAKKKKGSSGGSGGSGGSTNLSDIIE
KETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLS
GGSGGSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPW
ALVIQDSNGENKIKMLYPYDVPDYAYPYDVPDYAYPYDVPDYA (SEQ ID NO:21) ;

(b)

MPAAKRVKLDTSEKGPSTGDPTLRRRIESWEFDVFDPRELRKETCLLYEIKWGMSRKIWRSSGKNTTNHVE
VNFIIKFTSERRFHSSISCSITWFLSWSPCWECSQAIREFLSQHPGVTLVIIYVARLFWHMDQRNRQGLRDLVNSGV
TIQIMRASEYYHCWRNFVNYPGDEAHWPQYPPLWMLLYALELHCIILSLPPCLKISRWRQNH LAFFRLHLQNCHY
QTIPPHILLATGLIHPSVTWRLKSGGSSGGSSGSETPGTSESATPESSGGSSGGSPKKKRKVGSIINFQRRGLMET
QASNQLISSHLKGYPIDYFVGLAIGTSSVGWAVTNKAYELLKFRSHKMWGSRLFDEGESAVARRGFRSMRRRLER
RKLRLKLEELFADAMAQVDPTFFMRLRESKYHYEDKTTGHSSKHILFIDKNYNDQDYFKEYPTVYHLRSELMKSG
TDDIRKLFLAVHHILKYRGNFLYEGATFDSNASTLDDVIKQALENITFNCFCDCNSAIISSIGQILMEAGTKSDKAK
AIEHLVDTYIATDVTDTSSKTQKDQVKEDKKRLKAFANLVGLNASLIDLFGSVEELEEDLKKLQITGDTYDDKRD
ELAKAWSDEIYIIDDCKSVYDAIILLISKEPGLTISESKVKAFNKHKDDLA ILKSLKSDRSIYNTMFKVDEKGLH
NYVHYIKQGRTEETS CNREDFYKYTKKIVEGLSDSKDKEYILSQIELQILLPLQRIKDNQVIPYQLHLEELKAILA
KCGPKFPFLNEVADGFSVAEKLIKMLEFRIPYVGPLNTHHNVNDNGGFAVVRKASGRVTPWNFDDKIDREKSAAA
FIKNLTNKCTYLLGEDVLPKSSLLYSEFMLLNELNNVRIDGKPLEKVVEHLIEAVFKQDHKKMTKNRIEQFLKDN
GYISETHKHEITGLDGEIKNDLASYRDMVRILGDGFD RSMAEEIITDITIFGESKKMLRETLRKKFASCLDDEAIK
KLTKLRYRDWGRLSQKLLNGIEGCDKAGDGT PETIIILMRNFSYNLMELLGDKFSFMERIQEINAKLTEGQIVNPH
DIIDDLALSPAVKRAVWQALRIVDEVAHIKKALPARIFVEVTRS NKNEKKKKDSRQKRLSDLYAAIKDDVLLNGL
NNEIFGELKSSLA KYDDAALRSKKLYLYTQMGRCAYTGEI IELSLNNTDNYDIDHIYPRSLTKDDSF DNVLVCKR
TANAQKSDAYPISEEIQKTQKPFWTF LKQQGLISERKYERLTRITPLTADDLSGFIARQLVETNQSVKAATTLRR
LYPGVDVVVFKAENVTD FRHDNNFIKVRSLNHHHAKDAYLNIVVG NVYHERFTRNFRAFFKKNGANRTYNLAKMF
NYDVNCTNAKD GKAWDVKTSMDTVKKMDSNDVRVTKRLEQT GALADATIYKATVAGKAKDGAYIGMKT KSSVFA
DVSKYGGMTKIKNAYSII VQYTGKKGEVIKEIVPLPIYLTNRNTTDQDLIN YVASIIPQAKDISIIYGKLCINQLV
KVNGFYYYLGGKTNSKFCIDNAIQVIVSNEWIPYLKVLEKFNNMRKDNKDLKANVVSTRALDNKHTIEVRIVEEKN
IEFFDYLVSKLKMPIYQKMKGNKAAELSEKGYGLFKKMSLEEQSIHLIELLNLLTNQKTTFEVKPLGITASRSTVG
SKISNQDEFKVINESITGLYSNEVTIVGKRPAATKKAGQAKKKKGSSGGSGGSGGSTNLSDIIEKETGKQLVIQES
ILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGSGSGSGGSTNL
SDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENK

IKMLYPYDVPDYAYPYDVPDYAY (SEQ ID NO:12);

(c)

MPAAKRVKLDTSEKGPSTGDPTLRRRIESWEFDVFYDPRELRKETCLLYEIKWGMSRKIWRSSGKNTTNHVE
VNFICKFTSERRFHSSISCSITWFLSWSPCWECSQAIREFLSQHPGVTLVIIYVARLFWHMDQRNRQGLRDLVNSGV
TIQIMRASEYYHCWRNFVNYPGDEAHWPQYPPLWMLYALELHCIILSLPPCLKISRRWQNHLAFFRLHLQNCY
QTIPPHILLATGLIHPSVTWRLKSGGSSGGSSGSETPGTSESATPESSGGSSGGSPKKKRKVGTKVKDYYIGLAIG
TSSVGWAVTDEAYNVLFNSKKMWGVRLFDDAKTAEERRGQRGARRRLDRKKERLSLLQDFFAEVAKVDPNFFLR
LDNSDLYMEDKDQKLKSKYTLFNDKDFKDNFHKKYPTIHLLMDLIEDDSKKDIRLVYLACHYLLKNRGHFIFEG
QKFDTKSSFENSLNELKVHLNDEYGLDLEFDNENLINILTDPKLNKTAKKKELKSVIGDTKFLKAVSAIMIGSSQK
LVDLFENPEDFDDSAIKSVDFSTTSFDDKYSYELALGDKIALVNILKEIYDSSIENLLKEADKSKDGNKYISNA
FVKKYNKHGQDLKEFKRLVRQYHKSAYFDIFRSEKVNNDNYVSYTKSSISNNKRVKANKFTDQEAIFYKFAKKHLETI
KYKINKVNGSKADLELIDGMLRDMEFKNFMPKIKSSDNVIPYQLKLMELNKILENQSKHHEFLNVSDEYGSVCDK
IASIMEFRIPYYVGPLNPNSKYAWIKKQKDSEITPWNFKDVVDLDSSREEFIDSLIGRCTYLKDEKVLPKASLLYN
EYMLNELNNLKLNDLPITEEMKKKIFDQLFKTRKKVTLKAVANLLKKEFNINGEILLSGTDGDFKQGLNSYNDFK
AIVGDKVSDDYRDKIEEIKLIVLYGDDKSYLQKKIKAGYGYFTDSEIKKMAGLNYKDWGRLSKKLLTGLEGAN
KITGERGSI IHFMREYNLNMELMSASFTFTEEIQKLNVPDDRKLSYEMVDELYLSPSVKRMLWQSLRIVDEIKNI
MGTDSSKIFIEMARGKEEVKARKESRKNQLLFYKDGKKAFISEIGEERYSYLLSEIEGEEENKFRWDNLYLYTQ
LGRCMYSLEPIDISELSSKNIYDQDHIYPKSKIYDDSIENRVLVKKDLNSKKGNSYIPDEILNKNCYAYWKILYD
KGLIGQKKYTRLTRRTGFTDDEL VQFISRQIVETRQATKETANLLKTICKNSEIVYSKAENASRFRQEFDIVKCRA
VNDLHHMHDAYINIIVGNVYNTKFTKDP MNFVKKQEKARSYNLENMFKYDVKRGGYTAWIADDEKGTVKNASIKRI
RKELEGTNYRFRTRMNYIESGALFNATLQRKNKGSRLPKDKGPKSSIEKYGGYTNINKACFAVLDIKSKNKIERKLM
PVEREYAKQKNDKKLSDEIFSKYLKDRFGIEDYRVVYPVVKMRTLLKIDGSYYFITGGSDKTLELRSALQLILPK
KNEWAIKQIDKSSENDYLTIERIQDLTEELVYNTFDIIVNKFKTSVFKKSFLNLFQDDKIENIDFKFKSMDFKKC
KTLLMLVKAIRASGVRQDLKSIDLKSDYGRLLSSTNNIGNYQEFKIINQSITGLFENEVDLLKLGRPAATKKAGQ
AKKKKGSSGGSSGGSSGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTS
DAPEYKPWALVIQDSNGENKIKMLSGGSSGGSSGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILV
HTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLYPYDVPDYAYPYDVPDYAY (SEQ ID NO:18);

(d)

MPAAKRVKLDTSEKGPSTGDPTLRRRIESWEFDVFYDPRELRKETCLLYEIKWGMSRKIWRSSGKNTTNHVE
VNFICKFTSERRFHSSISCSITWFLSWSPCWECSQAIREFLSQHPGVTLVIIYVARLFWHMDQRNRQGLRDLVNSGV
TIQIMRASEYYHCWRNFVNYPGDEAHWPQYPPLWMLYALELHCIILSLPPCLKISRRWQNHLAFFRLHLQNCY
QTIPPHILLATGLIHPSVTWRLKSGGSSGGSSGSETPGTSESATPESSGGSSGGSPKKKRKVGKEYHIGLAIGTSS
IGWAVTDSQFKLMRIKGKTAIGVRLFEEGKTAERRTFRTRRRRLKRRKWRLHYLDEIFAPHLQEVDENFLRRLKQ
SNIHPEDPAKNQAFIGKLLFPDLLKKNRGYPTLIKMRDELPEQRAHYPTNIYKLREAMINEDRQFDLREVYLA
VHHIVKYRGHFLNNASVDKFKVGRIDFKSFNVLNEAYEELQNGEGSFTIEPSKVEKIGQLLLDTKMRKLDQKAV
AKLLEVKVADKEETKRNQIATAMSKLVLYKADFATVAMANGNEWKIDLSSETSEDEIEKFREELSDAQNDILTE
ITSLSQIMLNEIVPNGMSISESMMDRYWTHERRQLAEVKEYLATQPASARKEFDQVYNKYIGQAPKEKGFDEKGL
KKILSKKENWKEIDELLKAGDFLPKQRTSANGVIPHQMHQQLDRIIEKQAKYYPWLATENPATGERDRHQAKYEL
DQLVSFRIPYYVGPLVTPEVQKATSGAKFAWAKRKEDGEITPWNLDKIDRAESAEAFIKRMTVKDITYLLNEDVLP

ANSLLYQKYNVLNELNNVRVNGRRLSVGIKQDIYTELFKKKKTVKAGDVASLVMATRGNKPSVEGLSDPKKFNS
NLATYLDLKSIVGDKVDDNRYQMDLENI IEWRSVFEDGEIFADKLTEVEWLTDEQRSALVKKRYKGWGRLSKLLT
GIVDENGQRIIDL MWNTDQNFMQIVNQPVFKEQIDQLNQKAITNDGMTLRERVESVLDDAYTSPQNKKAIWQVVRV
VEDIVKAVGNAPKSSISIEFARNEGNGEITRSRRTQLQKLFEDQAHSLVKDTSLTEELEKAPDLSDRYFFYFTQGG
KDMYTGDPI NFDEISTKYDIDHILPQS FVKDDSLDNRLVSRAENKKS DRVPAKLYAAKMKPYWNQLLKQGLITQ
RK FENLTMDVDQTIKYRSLGFVKRQLVETRQVIKLTANILGSMYQEAGTDI IETRAGLTKQLREEFDLPKREVND
YHHA VDAYLTTFAGQYLNRRYPKLSFFVYGEYMKFKHGS DLKLRNFNFHELMEGDKSQGKVVDQQTGELITTRD
EVADYFDWVINLKVMLISNETYEETGKYFDASHESSLYLKNQNKSKLVVPLKNKLQPEYYGAYTGITQGYM VIL
KLLDKKGGFGVYRIPRYAADILNKCHDEVAYRNKIAEIISSDPRAPKSFEVVPVRLKGTFLVDGEEKFILSSYRY
KVNATQLILPVSDIKLIQDNFKALKKLVNEMQTKKLEIYDNILRQVDKYYKLYDINKFRAKLHDGRSKFVELDDF
GQDASKEKVI IKILRGLHFGSDLQNLKEIGFGTTPLGQFQVSEAGIRLSNTAFIIFKSPTGLFNRKLYLKNLGKRP
AATKKAGQAKKKKGSSGGSGGSGGSTNLSDI IEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTD
ENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGSGSGSGGSTNLSDI IEKETGKQLVIQESILMLPEEVEEVIGN
KPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLYPYDVPDYAYPYDVPDYAY (SEQ ID
NO:90) ;

(e)

MPAAKRVKLDTSEKGPSTGDPTLRRRIESWEFDVFDPRELRKETCLLYEIKWGMSRKIWRSSGKNTTNHVE
VNF IKKFTSERRFHSSISCSITWFLSWSPCWECSQAIREFLSQHPGVTLVIYVARLFWHMDQRNRQGLRDLVNSGV
TIQIMRASEYYHCWRNFVNYPGDEAHWPQYPPLWMLLYALELHCIILSLPPCLKISRWRQNH LAFFRLHLQNCY
QTIPPHILLATGLIHPSVTWRLKSGGSSGGSSGSETPGTSESATPESSGGSSGGSPKKKRKVGEKKTNYTIGLAIG
TDSVGWAVVKDDLELVKKRMKVLGNTETNYIKKNLWGSLLFESGQTAKDRRLKRVARRRYERRRNLTELQKIFAP
AIDEVDENFFFRLNESFLVPEDKAFSKNPIFGTLGEDKTYKYPTIYHLRQHLADSEEKADVRLIYLALAHMIKY
RGHFLIEGKLDTEHIAINENLEQFFESYNALFSEPIELRKEELIAIENILREKNSRTVKEKRITSFLKDIGRANK
QSPMMAFITLIVGKKAKFKAAFNLEEEISLNLTDSDYDENLEILLNTIGSDFADLFDHAQRVYNAVELAGILSGDV
KNTHAKLSAQMVAMYERHKEQLKEYKSFIKANLPDQYDMTFVAPKDAQKKDLKGYAGYIDGNMSQDSFYKFVKDQL
KEVPGSEKFLDSIEKEDFLRKQRSFYNGVIPNQVHLAEMEA ILDRQENYYPWLKENREKIIISLLTFRIPYYVGPLA
DGQSEFAWLERKSDEKIPWNFSDVVDLDRSAEKFIEQLIGRDTYLPDEYVLPKKS LIYQKYMVFNELTKIAYLDE
RQKRMNLSSVEKKEIFETLFKKRSKVTEKQLVKFFENYLQIDNPTIFGIEDAFNADYSTYVELAKVPGMKSMDDP
DNEDLMEEIVKILTVFEDRKMRRKQLEKYKERLSPEQIKELAKKHGTGWGRLSKLLVGIRDKETQKTILDYLVED
DNHSGGRQHLNRNLMQLINDDRLSFKKTI AELQ MIDPSADLYAQVQEIAGSPA I KKGILLGLKIVDEIIRVMGEKP
ENIV IEMARENQTTARGKALSKRREAKIKEGLAALGSSLLKENLPGNADLSQRKIYLYYTQNGKDIYLDEPLDFDR
LSQYDEDHIIPQSFTVDNSLDNLVLTNSSQNRGNKKDDVPSLEVVRQLAYWRS LK DAGLMTQRKFDNLTKAMRG
LTDKDRERFIQRQLVETRQITKNVAKLLDMRLNDKKDEAGNKIRETNIVLLKSAMASEFRKMFRLYKVRELNDYHH
AHDAYLNAAI AINLLALYPYMA DDFVYGEFRYKKKPQAEKATYEKL RQWNLIKRFGEKQLFTPDHEDCWNKERDIK
TIKKVMGYRQVNVVKKAEERTGMLFKETINGKTNKGSRIPIKKDLDP SKYGGYIEEKMAYYAVISYEDKKKKPGKT
IVGISIMDKKEFEYDSISYLGKLGFSNPVQIILKNYS LIAYPDGRRRYITGATKTTKGKVELQKANQIAMEQDLV
NFIYHLKNYDEISHPESYAFVQSHTDYFDR LFD SIEHYTRRFLDAETNINRLRRIYEEKKKKDPVDIEALVASFIE
LLKLT SAGAPADFI FMGEAISRRRYNSMTGLFDGQVIYQSLTGLYETRMRFEDGKRPAATKKAGQAKKKKGSSGGS
GGSGGSTNLSDI IEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALV

IQDSNGENKIKMLSGGSGGSGGSTNLSDIIEKETGKQLVIQESILMLPEEEVEEVIGNKPESDILVHTAYDESTDEN
VMLLTSDAPEYKPWALVIQDSNGENKIKMLYPYDVPDYAYPYDVPDYAY (SEQ ID NO:93);

(f)

MPAAKRVKLDTNLSDIIEKETGKQLVIQESILMLPEEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAP
EYKPWALVIQDSNGENKIKMLSGGSGGSGGSTNLSDIIEKETGKQLVIQESILMLPEEEVEEVIGNKPESDILVHTA
YDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSGGSGGSPKKKRKVEKKNYITIGLAIGTDSVGWAV
VKDDLELVKKRMKVLGNTETNYIKKNLWGSLLFESGQTAKDRRLKRVARRRYERRRNRLTELQKIFAPAIDEVDEN
FFFRLNESFLVPEDKAFSKNPIFGTLGEDKTYKYPTIYHLRQHLADSEEKADVRLIYLALAHMIKYRGHFLIEG
KLDTEHIAINENLEQFFESYNALFSEPIELRKEELIAIENILREKNSRTVKEKRITSFLKDIGRANKQSPMMAFI
TLIVGKKAKFKAAFNLEEEISLNLTDDSYDENLEILLNTIGSDFADLFDHAQRVYNAVELAGILSGDVKNTHAKLS
AQMVAMYERHKEQLKEYKSFIKANLPDQYDMTFVAPKDAQKKDLKGYAGYIDGNMSQDSFYKFVKDQLKEVPGSEK
FLDSIEKEDFLRKQRSFYNGVIPNQVHLAEMEAILDRQENYYPWLKENREKIIISLLTFRIPYVVGPLADGQSEFAW
LERKSDEKIKPWNFSDVVDLDRSAEKFIEQLIGRDTYLPDEYVLPKKSLIYQKYMVFNELTKIAYLDERQKRMNLS
SVEKKEIFETLFFKKRSKVTEKQLVKFFENYLQIDNPTIFGIEDAFNADYSTYVELAKVPGMKSMDDPDNEDLMEE
IVKILTVFEDRKMRRKQLEKYKERLSPEQIKELAKKHGTGWRLSKKLLVGIRDKETQKTILDYLVEDDNHSGGRQ
HLNRNLMQLINDRLSFKKTI AELQ MIDPSADLYAQVQEIAGSPA IKKGILLGLKIVDEIIRVMGEKPENIVIEMA
RENQTTARGKALS KRREAKIKEGLAALGSSLLKENLPGNADLSQRKIYLYYTQNGKDIYLDEPLDFDRLS QYDEDH
IIPQSFTVDNSLDNLVLTNSSQNRGNKKDDVPSLEVVRQLAYWRS LK DAGLMTQRKFDNLTKAMRGGLTDKDRER
FIQRQLVETRQITKNVAKLLDMRLNDKKDEAGNKIRETNIVLLKSAMASEFRKMFRLYK VRELNDYHHAHDAYLNA
AIAINLLALYPYMADDFVYGEFRYKKKPQAEKATYEKLRQWNLIKRFGEKQLFTPDHEDCWNKERDIKTIKKVMGY
RQVNVVKKAEERTGMLFKETINGKTNKGSRIPIKKDLDP SKYGGYIEEKMAYYAVISYEDKKKKPGKTIVGISIMD
KKEFEYDSISYLGKLGSNPVQIILKNYSLIAYPDGRRRYITGATKTTKGKVELQKANQIAMEQDLVNFYIHLKN
YDEISHPESYAFVQSHTDYFDRLFDSIEHYTRRFLDAETNINRLRRIYEEEEKKKDPVDIEALVASFIELLKLTSAG
APADFI FMGEAISRRRYNSMTGLFDGQVIYQSLTGLYETRMRFEDKRPAA TKKAGQAKKKKGSSGSGSSGSSGSET
PGTSESATPESSGSSGSGSTSEKGPSTGDPTLRRRIESWEFDFYDPREL RKETCLLYEIKWGMSRKIWRSSGKNT
TNHVEVNF IKKFTSERRFHSSISCSITWFLSWSPCWECSQAIREFLSQHPGVTLVIYVARLFWHMDQRNRQGLRDL
VNSGVTIQIMRASEYYHCWRNFVNYPGDEAHWPQYPPLWMMLYALELHCIILSLPPCLKISRRWQNHLA FFRHLHL
QNCHYQTIPPHILLATGLIHPSVTWRYPYDVPDYAYPYDVPDYAYPYDVPDYA (SEQ ID NO:94)。

25. 根据权利要求2所述的Cas9蛋白,其中所述Cas9蛋白识别包含5'-NGG-3'的PAM序列。

26. 根据权利要求3所述的Cas9蛋白,其中所述Cas9蛋白识别包含5'-NAGHC-3'的PAM序列,其中H是腺嘌呤、胞嘧啶或胸腺嘧啶。

27. 根据权利要求4所述的Cas9蛋白,其中所述Cas9蛋白识别包含5'-NRHRRH-3'的PAM序列,其中H是腺嘌呤、胞嘧啶或胸腺嘧啶,并且R是腺嘌呤或鸟嘌呤。

28. 根据权利要求5或权利要求7所述的Cas9蛋白,其中所述Cas9蛋白识别包含5'-NGG-3'的PAM序列。

29. 根据权利要求6所述的Cas9蛋白,其中所述Cas9蛋白识别包含5'-NNAAA-3'的PAM序列。

29b. 根据权利要求10b或10c所述的Cas9蛋白,其中所述Cas9蛋白识别包含5'-NGG-3'

的PAM序列。

30.一种核酸,其编码根据前述权利要求中任一项所述的Cas9蛋白。

31.根据权利要求30所述的核酸,其中所述核酸经密码子优化以用于在哺乳动物细胞中表达。

32.根据权利要求31所述的核酸,其中所述核酸经密码子优化以用于在人类细胞中表达。

33.一种真核细胞,其包含根据权利要求29中任一项所述的Cas9蛋白。

34.根据权利要求33所述的真核细胞,其中所述细胞是人类细胞。

35.一种切割真核细胞中的靶核酸的方法,其包括:

使所述细胞与根据权利要求1至29中任一项所述的Cas9和RNA向导或编码所述RNA向导的核酸接触,其中所述RNA向导包括能够与所述靶核酸杂交的同向重复序列和间隔子序列,并且

其中所述Cas9蛋白能够与所述RNA向导结合并导致与所述RNA向导互补的靶核酸序列中的断裂。

36.一种改变真核细胞中的靶核酸表达的方法,其包括:

使所述细胞与根据权利要求1至29中任一项所述的Cas9和RNA向导或编码所述RNA向导的核酸接触,其中所述RNA向导包括能够与所述靶核酸杂交的同向重复序列和间隔子序列,并且

其中所述Cas9蛋白能够与所述RNA向导结合并导致与所述RNA向导互补的靶核酸序列中的断裂。

37.一种改变真核细胞中的靶核酸表达的方法,其包括:

使所述细胞与根据权利要求1至29中任一项所述的Cas9和RNA向导或编码所述RNA向导的核酸接触,其中所述RNA向导包括能够与所述靶核酸杂交的同向重复序列和间隔子序列,并且

其中所述Cas9蛋白能够与所述RNA向导结合并编辑与所述RNA向导互补的靶核酸序列。

38.一种修饰真核细胞中的靶核酸的方法,其包括:

使所述细胞与根据权利要求1至29中任一项所述的Cas9和RNA向导或编码所述RNA向导的核酸接触,其中所述RNA向导包括能够与所述靶核酸杂交的同向重复序列和间隔子序列,并且

其中所述Cas9蛋白能够与所述RNA向导结合并编辑与所述RNA向导互补的靶核酸序列。

39.根据权利要求37或38所述的方法,其中所述Cas9蛋白是无活性的Cas9(dCas9)。

40.根据权利要求39所述的方法,其中所述dCas9与脱氨酶融合。

41.根据权利要求35至40中任一项所述的方法,其中所述RNA向导包含crRNA和tracrRNA。

42.根据权利要求35至39中任一项所述的方法,其中所述RNA向导包含sgRNA。

43.根据权利要求42所述的方法,其中与星座链球菌Cas9一起使用的所述sgRNA包含支架,所述支架包含与以下具有至少约80%同一性的序列:

5'-GUUUUAGAGCUGUGCUGUUUAAACAACACAGCAAGUUAUUUUAAGGCUUUGUCCGUACUCAAGCUUGCAAAAGCGUGCACCGAUUCGGUGCU-3'(SEQ ID NO:3)。

44. 根据权利要求42所述的方法, 其中与夏普氏菌Cas9一起使用的所述sgRNA包含支架, 所述支架包含与以下具有至少约80%同一性的序列:

5'-GUUUUAGAGUUGUGUUAUUGAAAAUAACACAACGAGUUAUUUAAAAGCUUAUGCUUAAAUGCCAGCUUUGCUGGUGUCAUUUAGAUGACUUUACUAAGGUUGCUUCGGCAACCUUUUU-3' (SEQ ID NO:7)。

45. 根据权利要求42所述的方法, 其中与小韦荣球菌Cas9一起使用的所述sgRNA包含支架, 所述支架包含与以下具有至少约80%同一性的序列:

5'-GUUUGAGAGUAGUGUGAAAAACAUAACGAGUUCAAAUACAAUUAUUUACAAUGCCUUCGGGCUGCCCGACGUAGGGCACCUCUCUCAAUUCUUCGGAUUGAGUU-3' (SEQ ID NO:13)。

46. 根据权利要求42所述的方法, 其中与秘鲁埃扎基拉菌Cas9一起使用的所述sgRNA包含支架, 所述支架包含与以下具有至少约80%同一性的序列:

5'-GUUUGAGAGUUAUGUAAUUGAAAAUUAACUGACGAGUUCAAAUAAAAUUAUUAUCAAACCGCCUAUUUAUAGGCCGCAGAUUCUGCAUUAUGCUUGCUAUUGCAAGCUU-3' (SEQ ID NO:19)。

47. 根据权利要求42所述的方法, 其中与发酵乳杆菌菌株AF15-40LB Cas9一起使用的所述sgRNA包含支架, 所述支架包含与以下具有至少约80%同一性的序列:

5'-GUCUUGGAGAGUGUGAAAAACACUCAUAGUCAAGAUAACGAGUGGUUUUCCACGAGUUAUUACUUUUGAGGUCUUAUAUGGCCCAUACAUA AAAAGGAGUCGGAUUUCCGGCUCCUUUCUU-3' (SEQ ID NO:95)。

48. 根据权利要求42所述的方法, 其中与嗜脲菌屑马赛-P3761 Cas9一起使用的所述sgRNA包含支架, 所述支架包含与以下具有至少约80%同一性的序列:

5'-GUUUUAGAGCCAUGUAGAAAAACAUAUGCAAGUUAUUUAAAUAAGGCUUUGUCCGUAAUCAACUUGAAAAAGUGGCGCUGUUUCGGCGCUU-3' (SEQ ID NO:96)。

49. 根据权利要求41所述的方法, 其中所述crRNA包含长度介于约16与26个核苷酸之间的向导序列。

50. 根据权利要求49所述的方法, 其中所述crRNA包含长度介于18与24个核苷酸之间的向导序列。

51. 根据权利要求35或36所述的方法, 其中所述靶核酸中的所述断裂是单链断裂或双链断裂。

52. 根据权利要求51所述的方法, 其中所述靶核酸中的所述断裂是单链断裂。

53. 根据权利要求34或35所述的方法, 其中所述Cas9蛋白是切割所述靶核酸序列的两条链的核酸酶, 或者是切割所述靶核酸序列的一条链的切口酶。

54. 根据权利要求34至53中任一项所述的方法, 其中所述靶核酸位于前间隔序列邻近基序(PAM)序列的5'处。

55. 根据权利要求34至54中任一项所述的方法, 其中所述Cas9可操作地连接至启动子序列以用于在真核细胞中表达, 并且其中所述向导RNA可操作地连接至启动子序列以用于在真核细胞中表达。

56. 根据权利要求55所述的方法, 其中所述真核细胞是人类细胞。

57. 根据权利要求55所述的方法, 其中所述启动子序列是真核启动子或病毒启动子。

58. 一种工程化的非天然存在的CRISPR-Cas系统, 其包括:

RNA向导或编码所述RNA向导的核酸, 其中所述RNA向导包括能够与靶核酸杂交的同向重复序列和间隔子序列; 以及

经密码子优化的CRISPR相关(Cas)蛋白,其与SEQ ID NO:1、4、8、14、84或86具有至少80%序列同一性,并且其中所述Cas蛋白能够与所述RNA向导结合并导致与所述RNA向导互补的靶核酸序列中的断裂。

58b.根据权利要求58所述的工程化的非天然存在的CRISPR-Cas系统,其中所述经密码子优化的CRISPR相关(Cas)蛋白进一步包含核定位序列(NLS)和/或FLAG、HIS或HA标签。

58c.根据权利要求59所述的工程化的非天然存在的CRISPR-Cas系统,其中所述经密码子优化的CRISPR相关(Cas)蛋白与SEQ ID NO:2、5、9、15、85、87、95或96具有至少80%序列同一性,并且其中所述Cas蛋白能够与所述RNA向导结合并导致与所述RNA向导互补的靶核酸序列中的断裂。

59.一种工程化的非天然存在的CRISPR-Cas系统,其包括:

RNA向导或编码所述RNA向导的核酸,其中所述RNA向导包括能够与靶核酸杂交的同向重复序列和间隔子序列;以及

经密码子优化的CRISPR相关(Cas)蛋白,其与SEQ ID NO:1、4、8、14、84或86具有至少80%序列同一性;

其中所述Cas蛋白与脱氨酶融合,并且其中Cas蛋白融合物能够与所述RNA向导结合并编辑与所述RNA向导互补的靶核酸序列。

59b.根据权利要求59所述的工程化的非天然存在的CRISPR-Cas系统,其中所述经密码子优化的CRISPR相关(Cas)蛋白进一步包含核定位序列(NLS)和/或FLAG、HIS或HA标签。

59c.根据权利要求59b所述的工程化的非天然存在的CRISPR-Cas系统,其中所述经密码子优化的CRISPR相关(Cas)蛋白与SEQ ID NO:2、5、9、15、85、87、95或96具有至少80%序列同一性,其中所述Cas蛋白与脱氨酶融合,并且其中所述Cas蛋白融合物能够与所述RNA向导结合并编辑与所述RNA向导互补的所述靶核酸序列。

60.根据权利要求59所述的系统,其中Cas9蛋白是无活性的Cas9(dCas9)。

61.根据权利要求58至60中任一项权利要求所述的系统,其中所述RNA向导包含crRNA和tracrRNA。

62.根据权利要求58至60中任一项所述的系统,其中所述RNA向导包含sgRNA。

63.根据权利要求62所述的系统,其中与星座链球菌Cas9一起使用的所述sgRNA包含支架,所述支架包含与以下具有至少约80%同一性的序列:

5'-GUUUUAGAGCUGUGCUGUUUAAACAACACAGCAAGUUAUUAAAGGCUUUGUCCGUACUCAAGCUUGCAAAGCGUGCACCGAUUCGGUGCU-3'(SEQ ID NO:3)。

64.根据权利要求62所述的系统,与夏普氏菌Cas9一起使用的所述sgRNA包含支架,所述支架包含与以下具有至少约80%同一性的序列:

5'-GUUUUAGAGUUGUGUUAUUGAAAAUAACACAACGAGUUAUUAAAGCUUAUGCUUAAAUGCCAGCUUUGCUGGUGUCAUUAGAUGACUUUACUAAAGGUUGCUUCGGCAACCUUUUU-3'(SEQ ID NO:7)。

65.根据权利要求62所述的系统,其中与小韦荣球菌Cas9一起使用的所述sgRNA包含支架,所述支架包含与以下具有至少约80%同一性的序列:

5'-GUUUGAGAGUAGUGUGAAAACAUACGAGUUCAAAUACAAAUAAUUUACAAUGCCUUCGGGCUGCCCCGACGUAGGGCACCUACUCUCAAUUCUUCGGAAUUGAGUU-3'(SEQ ID NO:13)。

66.根据权利要求62所述的系统,其中与秘鲁埃扎基拉菌Cas9一起使用的所述sgRNA包

含支架,所述支架包含与以下具有至少约80%同一性的序列:

5'-GUUUGAGAGUUAUGUAAUUGAAAAUUAACAUGACGAGUUCAAAUAAAAUUUAUCAAACCGCCUAUUU
AUAGGCCGAGAUUUCUGCAUUAUGCUUGCUAUUGCAAGCUU-3' (SEQ ID NO:19)。

67. 根据权利要求62所述的系统,其中与发酵乳杆菌菌株AF15-40LB Cas9一起使用的所述sgRNA包含支架,所述支架包含与以下具有至少约80%同一性的序列:

5'-GUCUUGGAGAGUGUGAAAAACACUCAUAGUCAAGAUCAAACGAGUGGUUUUCCACGAGUUAUUACUUUU
GAGGUCUUAUAUGGCCCAUACAUAAGGAGUCGGAAUUUCCGGCUCCUUUCUU-3' (SEQ ID NO:95)。

68. 根据权利要求62所述的系统,其中与嗜脉菌屑马赛-P3761 Cas9一起使用的所述sgRNA包含支架,所述支架包含与以下具有至少约80%同一性的序列:

5'-GUUUUAGAGCCAUGUAGAAAUACAUUGCAAGUUAAGGCUUUGUCCGUAUUAACUUGAAAAAGU
GGCGCUGUUUCGGCGCUU-3' (SEQ ID NO:96)。

69. 根据权利要求58至68中任一项所述的系统,其中所述Cas蛋白可操作地连接至启动子序列以用于在真核细胞中表达,并且其中所述向导RNA可操作地连接至启动子序列以用于在真核细胞中表达。

70. 根据权利要求69所述的系统,其中所述真核细胞是人类细胞。

71. 根据权利要求70所述的系统,其中所述启动子序列是真核启动子序列。

72. 一种核酸,其编码根据权利要求58至71中任一项所述的系统。

73. 一种载体,其包含根据权利要求58至72中任一项所述的系统。

74. 根据权利要求73所述的载体,其中所述载体是质粒载体或病毒载体。

75. 根据权利要求74所述的载体,其中所述病毒载体是腺相关病毒(AAV)载体或慢病毒载体。

76. 根据权利要求75所述的载体,其中所述病毒载体是AAV载体。

77. 根据权利要求76所述的载体,其中多于一种AAV载体用于封装根据权利要求59至71所述的系统。

78. 一种治疗有此需要的受试者的病症或疾病的方法,所述方法包括向所述受试者施用根据权利要求58至71中任一项所述的系统,

其中所述向导RNA和与病况或疾病相关的靶核酸的至少10个核苷酸互补;

其中所述Cas蛋白与所述向导RNA相关;

其中所述向导RNA与所述靶核酸结合;

其中所述Cas蛋白导致所述靶核酸中的断裂,任选地其中Cas9是与脱氨酶融合的无活性的Cas9(dCas9)并导致所述靶核酸中的一个或多个碱基编辑,从而治疗所述病症或疾病。

79. 根据权利要求78所述的方法,其中所述向导RNA与约18-24个核苷酸互补。

80. 根据权利要求79所述的方法,其中所述向导RNA与20个核苷酸互补。

81. 一种碱基编辑器,其包括根据权利要求16至19中任一项所述的融合蛋白。

82. 根据权利要求81所述的碱基编辑器,其包括腺苷脱氨酶结构域或胞苷脱氨酶结构域。

82b. 根据权利要求81所述的碱基编辑器,其包括腺苷脱氨酶结构域和胞苷脱氨酶结构域。

83. 一种编辑多核苷酸的核碱基的方法,所述方法包括使所述多核苷酸和与一种或多

种向导RNA复合的根据权利要求81所述的碱基编辑器接触,其中所述碱基编辑器包括腺苷脱氨酶结构域,并且其中所述一种或多种向导RNA靶向所述碱基编辑器以实现所述多核苷酸中的A • T至G • C的改变。

84. 一种编辑多核苷酸的核碱基的方法,所述方法包括使所述多核苷酸和与一种或多种向导RNA复合的根据权利要求81所述的碱基编辑器接触,其中所述碱基编辑器包括胞苷脱氨酶结构域,并且其中所述一种或多种向导RNA靶向所述碱基编辑器以实现所述多核苷酸中的C • G至T • A的改变。

85. 根据权利要求83或84所述的方法,其中所述编辑导致靶多核苷酸序列中的小于50%的插入缺失形成。

86. 根据权利要求83至85中任一项所述的方法,其中所述编辑产生点突变。

新颖CRISPR酶、方法、系统及其用途

[0001] 相关申请的交叉引用

[0002] 本申请要求于2021年3月23日提交的美国临时专利申请系列号63/164,798的优先权,该美国临时专利申请出于所有目的通过引用以其整体并入本文。

背景技术

[0003] 来自原核成簇规则间隔短回文重复序列(CRISPR)和CRISPR相关蛋白(CRISPR-Cas)系统的酶已被利用作为可重复编程和高度特异性的基因组编辑工具,以供用于真核生物。除了基因组编辑和切割外,CRISPR-Cas9还可用于将效应分子定位到基因组上的特定位置,从而通过多种机制进行遗传调控和表观遗传调控以及转录调节。

[0004] 然而,不同的基因组和基因组靶标需要多种工具来进行有效的基因工程,并且仍然需要通过发现和工程化能够识别和靶向不同序列的新颖Cas蛋白来扩展CRISPR工具箱。

[0005] 虽然CRISPR-Cas9系统可用于敲除基因或修饰基因的表达,但某些种类的基因编辑需要对靶基因进行精确修饰,诸如编辑该基因内的单个碱基。这种精确的修饰仍然是一个挑战,并需要不同的基因编辑工具包来实现广泛多种靶基因中的精确基因组修饰。

发明内容

[0006] 对独特前间隔序列邻近基序(PAM)具有特异性的新颖Cas9酶的鉴定允许扩展用于基因编辑的可用工具。本发明尤其提供了工程化的非天然存在的新颖Cas9酶,其由星座链球菌(*Streptococcus constellatus*)、夏普氏菌属(*Sharpea* spp.)分离株RUG017、小韦荣球菌(*Veillonella parvula*)、秘鲁埃扎基拉菌(*Ezakiella peruensis*)、发酵乳杆菌菌株(*Lactobacillus fermentum* strain) AF15-40LB和嗜脲菌属马赛(*Peptoniphilus* sp. Marseille) -P3761细菌分离。本发明部分地基于令人类惊讶的发现,即从不同细菌中发现的识别特定PAM序列的新颖Cas9酶可以经工程化以用于在真核细胞(例如,人类、植物等)中表达。因此,所描述的Cas9酶及其变体在真核生物中具有功能性。随之提供的实例显示了在人类细胞中使用具有不同PAM识别序列的工程化的非天然Cas9酶来靶向不同的基因组位点。例如,从星座链球菌、秘鲁埃扎基拉菌和嗜脲菌属马赛-P3761中工程化而成的Cas9识别共有PAM序列5'-NGG-3'。由从夏普氏菌属分离株RUG017中分离出的Cas9识别的共有PAM序列是5'-NAGHC-3'。由从小韦荣球菌中分离出的Cas9识别的共有PAM序列经鉴定为5'-NRHRRH-3'。由从发酵乳杆菌菌株AF15-40LB中分离出的Cas9识别的共有PAM序列经鉴定为5'-NNAAA-3'。(H=A、C或T;R=A或G)。

[0007] 在一个方面,本文提供了一种工程化的非天然存在的Cas9蛋白,其由星座链球菌Cas9、夏普氏菌Cas9、小韦荣球菌Cas9、秘鲁埃扎基拉菌Cas9、发酵乳杆菌菌株AF15-40LB Cas9或嗜脲菌属马赛-P3761Cas9修饰。

[0008] 在一些实施例中,星座链球菌Cas9蛋白与以下具有至少80%序列同一性:

[0009] MGKPYSIGLDIGTNSVGWAVVTDDYKVPKKMKVLGNTDKQSIKKNLLGALLFDSGETAEATRLKRTARRRYTRRKNRLRYLQEIFTGEMNKVDENFFQRLDDSFVLVDEDKRGEHHPIFGNIAAEVKYHDDFPTIYHLRRHLAD

TSKKADLRLVYLALAHMIKFRGHFLYEGDLKAENTDVQALFKDFVEEYDKTIEESHLSEITVDALSILTEKVS KSS
RLENLIAHYPTKKNLTLFGNLIALSLDLHPNFKTNFQLSEDAKLQFSKDTYEEDLEGFLGEVGEYADLFASAKNL
YDAILLSGILTVDDNSTKAPLSASMVKRYEEHQDLKKLKDFIKVNAPDQYNAIFKDKNKKGYASYIESGVKQDEF
YKYLKGILLKINGSGLDFDKIDREDFLRKQRTFDNGSIPHQIHLQEMHAILRRQGEHYPFLKENQDKIEKILTFRI
PPYVGPLARKGSRFAWAEYKADEKIPWNFDDILDKEKSAEKFITRMTLNDLYLPPEEKVLPKHS PLYEAF TVYNE
LTKVKYVNEQGEAKFFDTNMKQEIFDHVFKENRKVTCKDLLNYLNKEFEFRIVNL TGLDKENKAFNSSLGTYHDL
RKILDKSFLDDKANAKTIEDIIQTTLTFEDREMIRQLQKYSDFITKAQLKKLERRHYTGWGRLSYKLINGIRNKE
NKKTILDYLI DDGYANRNFMLINDDALSFKEE IARAQIIDDVDDIANVVHDLPGSPA I KKGILQSVKIVDELVKV
MGHN PANI I IEMARENQTTDKGRRNSQQRLKLLQDSLKNLDPVNIKNVENQQQLQNDRLFLYYIQNGKDMYTGETL
DINNLSQYDIDHIIPQAFIKDNSLDNRVLTRSDKNRGKSDDVPSIEVVHEMKSFW SKLLSVKLITQRKFDNLTAE
RGGLTEEDKAGFIKRQLVETRQITKHVAQILDERFNTEFDGNKRRIRNVKII TLKSNLVS NFRKEFELYK VREIND
YHHAHDAYLNAVVG NALLLKYPQLEPEFVYGEYPKYNSYRSRKSATEKFLFYSNILRFFKKEDIQTNEDGEIAWNK
EKHIKILRKVLSYPQVNI VKKTEEQTGGFSKESILPKGESDKLIPRKTKNSYWDPKKYGGFSDPVVAYSILVFADV
EKGKSKKL RKVQDMVGITIMEKKRFEKNPVDFLEQRGYRNVRL EKI IKLPKYSLFELENKRRRLLASAKELQKGNE
LVIPQRFTTLLYHSYRIEKDYEPHREYVEKHKDEFKELLEYISVFSRKYVLADNNLT KIEMLFSKNKDAEVS SL
AKSFISLLTFTAFGAPAAFNFFGENIDRKRYTSVTECLNATLIHQSI TGLYETRIDL SKLGED (SEQ ID NO:1)。

[0010] 在一些实施例中, 夏普氏菌Cas9蛋白与以下具有至少80%序列同一性:

[0011] MAKNKDIRYSIGLDIGTNSVGWAVMDEHYELLKKGNNHMGWSRLFDAAEPAATRRASRSIRRRYNKRR
ERIRLLRDLLGDMVMEVDPTFFIRLLNVSFLDEEDKQKNLGN DYKDNYNLFIEKDFNDKTYDYKPTIYHLRKELC
ENKEKADPRLIYLALHHIVKYRGNFLKEGQSF AKVYEDIEEKL DNTLKKFMSLNDLNDL FVDNDINSMITVLSKIY
QRSKKADDLLKIMNPTKEERAAYKEFTKALVGLKFNVSKMILAQEVKKDDKDIELDFSNDYDSTVDGLQAE LGEY
IEFTIEMLSINSWVELQDILGNNSTISAAMVERYEEHKNDLRVLKKVIREELPDKYNEVFREDNPKLHNYLGYIKY
PKNTPVEEFY EYIKRLLAKVDTGEAREILERIDLEKFM LKQNSRTNGSIPYQM QKDEMIQI IDNQSVYYPQLKENR
EKLISILEFRIPYYPGLNTHSEFAWIKKFEDKQKERILPWN YDQIVDIDATAEGFIERMQNTGT YFPDKPVMKN
SLTVSKFEVLNENLKIRINGKLIPVETKKELLSDFMKNKTITDKKLKDWLVTHQYYDTNEELKIEGYQKDLQFST
SLAPWIDFTKIFGEINASNYQLIEKIIYDISIFEDKKILKRRLKKVYQLDDLLVDKILKLNYTGWSRLSEKLLTGI
KSKNSKETILSILENSNMNLM EIIINDESLGFKQII EESNKKDIEGPF RYDEVKKLAGSPA I KRGIWQALLV VQEIT
KFMKHEP SHIYIEFAREEQE KVRTESRIAKLQKIYKDLNLQTKEDQLVYESLKKEDAKKKIDTDALYLYLQMGKS
MYS GKPLDIDKLSTYHIDHILPRSLIKDDSLDNRVLVLPKENEWKLDSETVPFEIRNKMMGFWQKLHENG LMSNKK
FFSLIRTD FNEKDKKRFINRQLVETRQIIKNVAVIINDHYTNTNVTVRAELSHQFRERYKIYKNRDLNDLHHAHD
AYIACILGQFIHQNF GNM DVNMIYGQYKKNYKKDVOEHNNYGFILNSMNIHFND DNSVIWDPSYIGKIKSCFCYK
DVYVTKKLEQNDAKLFDLTILPSDKNSENGVTAKIPVNKYRKDVN KYGGFSGDAPIMLAIEADKGKKHVRQVIAF
PLRLKNYNDEERIKFIEKEKNLKNVKILTEVKKNLILINH QYFFITGTNELVNATQLKLSAKNTKNL FNLVDANK
HNKLESIDDANFNEVIQELICKLQEP IYSRYNSIGKEFEDSYEKINAVTKQDKLYII EYLI AIMS AKATQGYIKPE
LAREIGTNGKNKGRIKSFTIDLNKTTFISTSVTGLFSK KYKL (SEQ ID NO:4)。

[0012] 在一些实施例中, 小韦荣球菌Cas9蛋白与以下具有至少80%序列同一性:

[0013] MSIINFQRRGLMETQASNQLISSHLKGYPIKDYFVGLDIGTSSVGWAVTNKAYELLKFRSHKMWGSRL
FDEGESAVARRGFRSMRRRLERRKLRLKLEELFADAMAQVDPTFFMRLRESKYHYEDKTTGHSSKHILFIDKNYN
DQDYFKEYPTVYHLRSELMKSGTDDIRKLFLAVHHILKYRGNFLYEGATFDSNASTLDDVIKQALENITFNCFCDCN

SAISSIGQILMEAGKTKSDKAKAIEHLVDTYIATDVTDTSSKTQKDQVKEDKKRLKAFANLVLGLNASLIDLFSGV
EELEEDLKKLQITGDTYDDKRDELAKAWSDEIYIIDDCKSVYDAIILLSIKEPGLTISESKVKA FNKHKDDLAILK
SLLKSDRSIYNTMFKVDEKGLHNYVHYIKQGRTEETSCNREDFYKYTKKIVEGLSDSKDKEYILSQIELQILLPLQ
RIKDNGVIPYQLHLEELKAILAKCGPKFPFLNEVADGFSVAEKLKMLEFRIPYYVGPLNTHHNVDNGGFAWVRK
ASGRVTPWNFDDKIDREKSAAAFIKNL TNKCTYLLGEDVLPKSSLLYSEFMLLNELNNVRIDGKPLEKVVEHLIE
AVFKQDHKKMTKNRIEQFLKDNGYISETHKHEITGLDGEIKNDLASYRDMVRILGDGFDRSMAEEIITDITIFGES
KKMLRETLRKKFASCLDDEAIKKLTKLRYRDWGRLSQKLLNGIEGCDKAGDGPETIIILMRNFSYNLMELLGDKF
SFMERIQEINAKLTEGQIVNPHDIIDDLALSPAVKRAVWQALRIVDEVAHIKKALPARIFVEVTRSNKNEKKKKDS
RQKRLSDLYAAIKDDVLLNGLNNEIFGELKSSLA KYDDAALRSKKLYLYYTQMGRCA YTGEEI IELSLLNTDNYDI
DHIYPRSLTKDDSF DNVLVCKRTANAQKSDAYPISEEIQKTQKPFWTF LKQQGLISERKYERLTRITPLTADDLSG
FIARQLVETNQSVKAATTLRRLYPGV DVVFVKAENVTD FRHDNNFIKVRSLNHHHHAKDAYLNI VGVN VYHERFT
RNFRAFFKKNGANRTYNLAKMFNYDVNCTNAKDGKAWDVKTSMDTVKKMMSNDVRVTKR LLEQT GALADATIYKA
TVAGKAKDGAYIGMKTSSVFADVSKYGGMTKIKNAYSII VQYTGKKGEVIKEIVPLPIYL TNRNTTDQDLIN YVA
SII PQAKDISIIYGKLCINQLVKVNGFYYYLGGKTNSKFCIDNAIQVIVSNEWIPYLKVLEKFNNMRKDNKDLKAN
VVSTRALDNKHTIEVRIVEEKNIEFFDYLVS KLKMPIYQKMKGNKAAELSEKGYGLFKKMSLEEQSIHLIELLNL
TNQKTTFEVKPLGITASRSTVGSKISNQDEFKVINESITGLYSNEVTIV (SEQ ID NO:8)。

[0014] 在一些实施例中, 秘鲁埃扎基拉菌Cas9蛋白与以下具有至少80%序列同一性:

[0015] MTKVKDYYIGLDIGTSSVGWAVTDEAYNV LKFNSKKMWGVRLFDDAKTAEERRGQRGARRRLDRKKER
LSLLQDFFAAEEVAKVDPNFFLR LDNSDLYMEDKDQKLKSKYTLFNDKDFDKNFHKKYPTIHHLLMDLIEDDSKKD
IRLVYLACHYLLKNRGHFI FEGQKFDTKSSFENSLNELKVHLNDEYGLDLEFDNENLINILTPKLNKTAKKKELK
SVIGDTKFLKAVSAIMIGSSQKLVDLFENPEDFDDSAIKSVDFSTTSFDDKYSYELALGDKIALVNILKEIYDSS
ILENLLKEADKSKDG NKYISNAFVKKYNKHGQDLKEFKRLVRQYHKSAYFDIFRSEK VNDNYVSYTKSSISNNKRV
KANKFTDQEA FYKFAKKHLETIKYKINKVNGSKADLELIDGMLRDMEFKNFMPKIKSSDNGVIPYQLKLMELNKIL
ENQ SKHHEFLNVSDEYGSVCDKIASIMEFRIPYYVGPLNPNSKYAWIKKOKDSEITPWNFKDVVDLDSSREEFID
SLIGRCTYLKDEKVL PKASLLYNEYMVLNELNKLNDLPITEEMKKKIFDQLFKTRKKVTLKAVANLLKKEFNIN
GEILLSGTDGDFKQGLNSYND FKAIVGDKVDSDDYRDKIEEIKLIVLYGDDKSYLQKKIKAGYGYFTDSEIKKM
AGLNYKDWGRLSKKLLTGLEGANKITGERGSI IHFMREYNLNMELMSASFTFTEEIQKLN PVDDRKLSYEMVDEL
YLSPSVKRWLWQSLRIVDEIKNIMGTD SKKIFIEMARGKEEVKARKESRKNQLLK FYKD GKAFISEIGEERYSYL
LSEIEGEEENKFRWDNLYLYYTQLGRCMYSLEPIDISELSSKN IYDQDHIYPKSKIYDDSIENRVLVKKDLNSKKG
NSYPIPDEILNKNCYAYWKILYDKGLIGQKKYTRLTRRTGFTDDELVQFISRQIVETRQATKETANLLKTICKNSE
IVYSKAENASRFRQEFDIVKCRVNDLHHMHDAYINIIVGNVYNTKFTKDP MNFVKKQEKARSYNLENMFKYDVKR
GGYTAWIADDEKGT VKNASIKRIRKELEGTNYRFRTRMNYIESGALFNATLQRKNKGS RPLKDKGPKSSIEKYGGYT
NINKACFAVLDIKSKNKIERKLMPVERE IYAKQKNDKKLSDEIFSKYLKDRFGIEDYRVVYPVVKMRTLLKIDGSY
YFITGGSDKTLELRSALQLILPKKNEWAIKQIDKSSENDYLTIERIQDLTEELVYNTFDIIVNKFKTSVFKKSFLN
LFQDDKIENIDFKFKSMD FKECKTLLMLVKAIRASGVRQDLKSIDLKSDYGR LSSKTNNIGNYQEFKIINQ
SITGLFENEVDLLKL (SEQ ID NO:14)

[0016] 在一些实施例中, 发酵乳杆菌Cas9蛋白与以下具有至少80%序列同一性:

[0017] MKEYHIGLDIGTSSIGWAVTDSQFKLMRIKGKTAIGVRLFE EGKTA AERRTFRTTRRRRLKRRKWRLHY
LDEIFAPHLQEVDENFLRRLKQSN IHPEDPAKNQAFIGKLLFPDLLKKNERGYPTLIKMRDEL PVEQRAHYPTNI

YKLREAMINEDRQFDLREVYLAVHHIVKYRGHFLNNASVDKFKVGRIDFDKSFNVLNEAYEELQNGEGSFTIEPSK
VEKIGQLLLDTKMRKLDQKAVAKLLEVKVADKEETKRNKQIATAMSKLVLGKADFATVAMANGNEWKIDLSSET
SEDEIEKFREELESDAQNDILTEITSLSFSQIMLNEIVPNGMSISESMMDRYWOTHERQLAEVKEYLATQPASARKEFD
QVYNKYIGQAPKEKGFDFLEKGLKKILSKKENWKEIDELLKAGDFLPKQRTSANGVIPHQMHQQELDRIIEKQAKYY
PWLATENPATGERDRHQAKYELDQLVSFRIPYYVGPLVTPEVQKATSGAKFAWAKRKEDGEITPWNLWDKIDRAES
AEAFIKRMTVKDITYLLNEDVLPANSLLYQKYNVLNELLNNVRVNGRRLSVGIKQDIYTELFKKKKTKVAGDVASLVM
AKTRGVNKPSVEGLSDPKKFNLSNLATYLDLKSIVGDKVDDNRYQMDLENIIEWRSVFEDGEIFADKLTEVEWLTD
QRSALVKKRYKGGWGRLSKKLLTGIVDENGQRIIDLMWNTDQNFMIQVNPVFKEQIDQLNQKAITNDGMTLRERVE
SVLDDAYTSPQNKAIWQVVRVEDIVKAVGNAPKSISIEFARNEGKGEITRSRRTQLKQLFEDQAHELKDTSL
TEELEKAPDLSDRYYFYFTQGGKDMYTGDPINFDEISTKYDIDHILPQSFKDDSLDNRLVSRAENKKS DRVPA
KLYAAKMKPYNQLLKQGLITQRKFENLTMDVDQTIKYRSLGFVKRQLVETRQVIKLTANILGSMYQEAGTDIIET
RAGLTKQLREEFDLPKVREVNDYHHAVDAYLTTFAGQYLNRRYPKLSFFVYGEYMKFKHGSCLKLRNFNFFHELM
EGDKSQGKVVDQQTGELITTRDEVADYFDWVINLKVMLISNETYEETGKYFDASHESLKLKNQNKSKLVVPLK
NKLQPEYYGAYTGITQGYMVLKLLDKKGGFGVYRIPRYAADILNKCHDEVAYRNKIAEIISSDPRAPKSFEVVVP
RVLKGTFLVDGEEKFILSSYRYKVNATQLILPVS DIKLIQDNFKALKKLNVMQTKKLEIYDNILRQVDKYYKLY
DINKFRAKLHDGRSKFVELDDFGQDASKEKVIKILRGLHFGSDLQNLKEIGFGTTPLGQFQVSEAGIRLSNTAFI
IFKSPTGLFNRKLYLKNL (SEQ ID NO:84)。

[0018] 在一些实施例中,嗜脉菌属马赛-P3761Cas9蛋白与以下具有至少80%序列同一性:

[0019] MEKKTNYTIGLDIGTDSVGWAVVKDDLELVKKRMKVLGNTETNYIKKNLWGSLLFESGQTAKDRRLKR
VARRRYERRRNLTELQKIFAPAIDEVDENFFRLNESFLVPEDKAFSKNPIFGTLGEDKTYKYTYPTIYHLRQHL
ADSEEKADVRLIYLALAHMIKYRGHFLIEGKLDTEHIAINENLEQFFESYNALFSEEPIELRKEELIAIENILREK
NSRTVKEKRITSFLKDIGRANKQSPMAFITLIVGKKAKFKAAFNLEEEISLNLTDSDYDENLEILLNTIGSDFAD
LFDHAQRVYNAVELAGILSGDVKNTHAKLSAQMVAMYERHKEQLKEYKSFIKANLPDQYDMTFVAPKDAQKKDLKG
YAGYIDGNMSQDSFYKFVKDQLKEVPGSEKFLDSIEKEDFLRKQRSFYNGVIPNQVHLAEMEAILDRQENYYPWK
ENREKIIISLLTFRIPYYVGPLADGQSEFAWLERKSDEKIKPWNFSDVVDLDRSAEKFIEQLIGRDTYLPDEYVLPK
KSLIYQKYMVFNELTKIAYLDERQKRMNLSSVEKKEIFETLFFKKRSKVTEKQLVKFFENYLQIDNPTIFGIEDAFN
ADYSTYVELAKVPGMKSMDDPDNEDLMEEIVKILTVEFDRKMRRKQLEKYKERLSPEQIKELAKKHYTGWGRLSK
KLLVGIRDKETQKTILDYLVEDDNHSGGRQHLNRNLMQLINDRSLFKKTI AELQMIDPSADLYAQVQEIAGSPAI
KKGILLGLKIVDEIIRVMGEKPENIVIEMARENQTTARGKALSKRREAKIKEGLAALGSSLLKENLPGNADLSQRK
IYLYYTQNGKDIYLDEPLDFDRLSQYDEDHIIPQSFTVDNSLDNLVLTNSSQNRGNKKDDVPSLEVVRQLAYWRS
LKDAGLMTQRKFDNLTKAMRGGLTDKDRERFIQRQLVETRQITKNVAKLLDMRLNDKKDEAGNKIRETNIVLLKSA
MASEFRKMFRLYKVRELNDYHHHAHDAYLNAAIAINLLALYPYMAADFVYGEFRYKKKPQAEKATYEKLQWNLIK
RFGKQLFTPDHEDCWNKERDIKTIKKVMGYRQVNVVKKAEERTGMLFKETINGKTNKGSRIPIKKDLPSKYGGYI
EEKMAYYAVISYEDKKKKPGKTIVGISIMDKKEFEYDSISYLGKLGFSNPVVQIILKNYSLIAYPDGRRRYITGAT
KTTKGKVELQKANQIAMEQDLVNFYHLKNYDEISHPESYAFVQSHTDYDFDLFDSIEHYTRRFLDAETNINRLRR
IYEEEEKKDPVDIEALVASFIELLKLTSAGAPADFI FMGEAISRRRYNSMTGLFDGQVIYQ SLTGLYETRMRFED
(SEQ ID NO:86)。

[0020] 在一些实施例中,Cas9蛋白包含与SEQ ID NO:1、4、8、14、84或86至少85%、至少

90%、至少92%、至少95%、至少96%、至少97%、至少98%或至少99%相同的氨基酸序列。

[0021] 在一些实施例中,Cas9蛋白进一步包含核定位序列(NLS)和/或FLAG、HIS或HA标签。

[0022] 在一些实施例中,星座链球菌Cas9具有与以下至少80%相同的氨基酸序列:

[0023] MPKKKRKVGKPYISGLDIGTNSVGWAVVTDYKVPKMKVGLGNTDKQSIKKNLLGALLFDSGETAE
ATRLKRTARRRYTRRNRLRYLQEIFTGEMNKVDENFFQRLDDSFLVDEDKRGEHHPIFGNIAAEVKYHDDFPTIY
HLRRHLADTSKKADLRLVYLALAHMIKFRGHFLYEGDLKAENTDVQALFKDFVEEYDKTIEESHLSEITVDALSIL
TEKVS KSSRLNLI AHYPTEKKNTLFGNLI ALSLDLHPNFKTNFQLSEDAKLQFSKDTYEEDLEGFLGEVGDEYAD
LFASAKNLYDAILLSGILTVDNSTKAPLSASMVKRYEEHQDLKKLKDFIKVNAPDQYNAIFKDKNKKGYASYIE
SGVKQDEFYKYLKGILLKINGS GDFLDKIDREDFLRKQRTFDNGSIPHQIHLQEMHAILRRQGEHYPFLKENQDKI
EKILTFRIPYYVGPLARKGSRFAWAEYKADEKITPWNFDDILDKEKSAEKFITRMTLNDLYLPEEKVLPKHSPLYE
AFTVYNELTKVKYVNEQGEAKFFDTNMKQEIFDHVFKENRKVTDKLLNYLNKEFEFRIVNLTGLDKENKAFNSS
LGTYHDLRKILDKSFLDDKANETIEDIIQTTLTFEDREMIRQLQKYSDFITKAQLKKLERRHYTGWRLSYKLI
NGIRNKENKKTILDYLI DDGYANRNFMLINDDALSFKEEIARAQIIDDVDDIANVVHDLPGSPAIKKGILQSVKI
VDELVKVMGHN PANI I IEMARENQTTDKGRRNSQQRLKLLQDSLKNLDNPVNIKNVENQQQLQNDRLFLYYIQNGKD
MYTGETLDINNLSQYDIDHIIPQAFIKDNSLDNRVLTRSDKNRGKSDDVPSIEVVHEMKSFSWKLSSVKLITQRKF
DNLTKAERGGLTEEDKAGFIKRQLVETRQITKHVAQILDERFNTEFDGNKRRIRNVKIITLKS NLVS NFRKEFELY
KVREINDYHHAHDAYLNAVVG NALLKYPQLEPEFVYGEYPKYNSYRSRSATEKFLFYSNILRFFKKEDIQTNE
GEIAWNKEKHILRKVLSYPQVNI VKKTEEQTGGFSKESILPKGESDKLIPRKTNSYWDPKKYGGFDSPVVAYS
ILVFADVEKGKSKKL RKVQDMVGITIMEKKRFEKNPVDFLEQRGYRNVRLKIIKLPKYSLFELENKRRRL LASAK
ELQKGNELVIPQRFTTLLYHSYRIEKDYEPHREYVEKHKDEFKELLEYISVFSRKYVLADNNLT KIEMLFSKNKD
AEVSSLAKSFISLLTFTAFGAPAAFNFFGENIDRKRYTSVTECLNATLIHQSI TGLYETRIDL SKLGEDGKRPAAT
KKAGQAKKKKGSYPYDVPDYAYPYDVPDYAYPYDVPDYA (SEQ ID NO:2)。

[0024] 在一些实施例中,夏普氏菌Cas9具有与以下至少80%相同的氨基酸序列:

[0025] MPKKKRKVGAKNKDIRYSIGLDIGTNSVGWAVMDEHYELLKKGNNHMGSRFLDAAEPAATRRASRSI
RRRYNKRERIRLLRDLLGDMVMEVDPTFFIRLLNVSFLDEEDKQKNLGN DYKDNYNLFIEKDFNDKTYDYKPTI
YHLRKELCENKEKADPRLIYLALHHIVKYRGNFLKEGQSF AKVYEDIEEKLDNTLKKFMSLNDLNDLFVDNDINSM
ITVLSKIYQRSKADLLKIMNPTKEERAAYKEFTKALVGLKFNVSKMILAQEVKKDDKDIELDFSNDYDSTVDG
LQAELGEYIEFIEMLSIN SWVELQDILGNNTISAAMVERYEEHKNDLRVLKKVIREELPDKYNEVFREDNPKLH
NYLGYIKYPKNTPVEEFY EYIKRLLAKVDTEAREILERIDLEKFMLKQNSRTNGSIPYQMQKDEMIQIIDNQSVY
YPQLKENREKLISILEFRIPYYFGPLNTHSEFAWIKKFEDKQKERILPWN YDQIVDIDATAEGFIERMQNTGTYPF
DKPVMANKSLTVSKFEVLNENLKIRINGKLIPVETKKELLSDFMKNKTI TDKKLKD WLTHQYYDTNEELKIEGY
QKDLQFSTSLAPWIDFTKIFGEINASNYQLIEKIIYD ISIFEDKKILKRRLKKVYQLDDLLVDKILKLN YTGWSR
LSEKLLTGIKSKNSKETILSILENSNMNLMEIINDESLGFKQIIIESNKKDIEGPFYRDEVKKL AGSPA I KRG I WQ
ALLVVQEITKFMKHEPSHIYIEFAREEQEKVRTESRIAKLQKIYKDLNLQTKEDQLVYESLKKEDAKKKIDTDALY
LYYLQMGKSMYSGKPLDIDKLSTYHIDHILPRSLIKDDSLDNRVLVLPKENEWKLDSETVPFEIRNKMMGFWQKLH
ENGLMSNKKFFSLIRTD FNEKDKKRFINRQLVETRQIIKNVAVIINDHYTNTNVTVRAELSHQFRERYKIYKNRD
LNDLHHAHDAYIACILGQFIHQNFNMVDNMIYGQYKKNYKKDVQEHNNYGFILNSMNIHFNDDNSVIWDPSYIG
KIKSCFCYKDVVYTKKLEQNDAKLFDLTILPSDKNSENGVTAKIPVNKYRKDVN KYGGFSGDAPIMLAIEADKKG

KHVRQVIAFPLRLKKNYNDEERIKFIEKEKNLKNVKILTEVKKNQLILINHQYFFITGTNELVNATQLKLSAKNTKN
LFNLVDANKHNKLESIDDANFNEVIQELICKLQEPIYSRYNSIGKEFEDSYEKINAVTKQDKLYIIIEYLIAIMSAK
ATQGYIKPELAREIGTNGKNKGRIKSFTIDLNKTTFISTSVTGLFSKKYKLGKRPAATKKAGQAKKKKGSYPYDVP
DYAYPYDVPDYAYPYDVPDYA (SEQ ID NO:5)。

[0026] 在一些实施例中,小韦荣球菌Cas9具有与以下至少80%相同的氨基酸序列:

[0027] MPKKKRKVGSIINFQRRGLMETQASNQLISSHLKGYPIKDYFVGLDIGTSSVGWAVTNKAYELLKFRS
HKMWGSRLFDEGESAVARRGFRSMRRRLERRKLRLKLEELFADAMAQVDPFTFFMRLRESKYHYEDKTTGHSSKHI
LFIDKNYNDQDYFKEYPTVYHLRSELMSKSGTDDIRKLFLAVHHILKYRGNFLYEGATFDSNASTLDDVIKQALENI
TFNCFDCNSAIISSIGQILMEAGTKSDKAKAIEHLVDTYIATDVTDTSSKTQKDQVKEDKKRLKAFANLVLGLNAS
LIDLFGSVEELEEDLKKLQITGDTYDDKRDELAKAWSDEIYIIDDCKSVYDAIILLSIKEPGLTISESKVKAFNKH
KDDLAILKSLKSDRSIYNTMFKVDEKGLHNYVHYIKQGRTEETSCNREDFYKYTKKIVEGLSDSKDKEYILSQIE
LQILLPLQRIKDNGVIPYQLHLEELKAILAKCGPKFPFLNEVADGFSVAEKLKMLEFRIPYYVGPLNTHHNVNNG
GFAWAVRKASGRVTPWNFDDKIDREKSAAAFIKNLTNKCTYLLGEDVLPKSSLLYSEFMLLNELNNVRIDGKPLEK
VVKEHLIEAVFKQDHKKMTKNRIEQFLKDNGYISETHKHEITGLDGEIKNDLASYRDMVRILGDGFDRSMAEEIIT
DITIFGESKKMLRETLRKKFASCLDDEAIIKLTCLRYRDWGRLSQKLLNGIEGCDKAGDGTPEIIILMRNFSYNL
MELLGDKFSFMERIQEINAKLTEGQIVNPHDIIDDLALSPAVKRAVWQALRIVDEVAHIKKALPARIFVEVTRSNK
NEKKKKDSROKRLSDLYAAIKKDDVLLNGLNNEIFGELKSSLAKYDDAALRSKKLYLYYTQMGRCAYTGEIIELSL
LNTDNYDIDHIYPRSLTKDDSFNVLCKRTANAQKSDAYPISEEIQKTQKPFWTFLKQQGLISERKYERLTRITP
LTADDLSGFIARQLVETNQSVKAATTLRRLYPGVDVVFVKAENVTDFRHDNNFIKVRSLNHHHHAKDAYLNIVVG
NVYHERFTRNFRAFFKKNGANRTYNLAKMFNYDVNCTNAKDGGKAWDVKTSMDTVKKMMSNDVRVTKRLLLEQTGAL
ADATIYKATVAGKAKDGAYIGMKTSSVFADVSKYGGMTKIKNAYSIIIVQYTGKKGEVIKEIVPLPIYLTNRNTTD
QDLINYVASIIPQAKDISIIYGKLCINQLVKVNGFYIYLGGKTNSKFCIDNAIQVIVSNEWIPYLVLEKFNNMRK
DNKDLKANVVSTRALDNKHTIEVRIVEEKNIIEFFDYLVSKLKMPIYQKMKGNKAAELSEKGYGLFKKMSLEEQSIH
LIELLNLLTNQKTTFEVKPLGITASRSTVSGKISNQDEFKVINESITGLYSNEVTIVGKRPAATKKAGQAKKKKGS
YPYDVPDYAYPYDVPDYAYPYDVPDYA (SEQ ID NO:9)。

[0028] 在一些实施例中,秘鲁埃扎基拉菌Cas9具有与以下至少80%相同的氨基酸序列:

[0029] MPKKKRKVGTKVKDYIIGLDIGTSSVGWAVTDEAYNVLFNSKKMWGVRLFDDAKTAEERRGQRGARR
RLDRKKERLSLLQDFFAEVAVKDPNFFLRLDNSDLYMEDKDQKLKSKYTLFNDKDFKDNFHKKYPTIHHLLMDL
IEDDSKKDIRLVYLACHYLLKNRGHFIFEGQKFDTKSSFENSLNELKVHLNDEYGLDLEFDNENLINILTPKLNK
TAKKKELKSVIGDTKFLKAVSAIMIGSSQKLVDLFENPEDFDDSAIKSVDFSTTSFDDKYSDYELALGDKIALVNI
LKEIYDSSILENLLKEADKSKDGKNYISNAFVKKYNKHGQDLKEFKRLVRQYHKSAYFDIFRSEKVNNDNVSYTKS
SISNNKRVKANKFTDQEAIFYKFAKKHLETIKYKINKVNGSKADLELIDGMLRDMEFKNFMPKIKSSDNGVIPYQLK
LMELNKILENQSKHHEFLNVSDEYGSVCDKIASIMEFRIPYYVGPLNPNSKYAWIKKQKDSEITPWNFKDVVDLDS
SREEFIDSLIGRCTYLKDEKVLPAKSLLYNEYMVLNELNNLKLNDLPITEEMKKKIFDQLFKTRKKVTLKAVANLL
KKEFNINGEILLSGTDGDFKQGLNSYNDFAIVGDKVSDDDYRDKIEEIIKLIVLYGDDKSYLQKKIKAGYGYFT
DSEIKKMAGLNYKDWGRLSKKLLTGLEGANKITGERGSI IHFMREYNLNLMELSASFTFTEEIQKLNVPDDRKLS
YEMVDELYLSPSVKRMLWQSLRIVDEIKNIMGTDSKKIFIEMARGKEEVKARKESRKNQLLKFYKDGKKAFISEIG
EERYSYLLSEIEGEEENKFRWDNLYLYYTQLGRCMYSLEPIDISELSSKNIYDQDHIYPKSKIYDDSIENRVLVKK
DLNSKKGNSYPIPEILNKNCYAYWKILYDKGLIGQKKYTRLTRRTGFTDDELVQFISRQIVETRQATKETANLLK

TICKNSEIVYSKAENASRFRQEFDIVKCRAVNDLHHMHDAYINIIVGNVYNTKFTKDP MNFVKKQEKARSYNLENM
FKYDVKRGGYTAWIADDEKGT VKNASIKRIRKELEGTNYRFRTRMNYIESGALFNATLQRKNKGSRLPKDKGPKSSI
EKYGGYTNINKACFAVLDIKSKN KIERKLMPVEREIIYAKQKNDKKLSDEIFSKYLKDRFGIEDYRVVYPVVKMRTL
LKIDGSYYFITGGSDKTLELRSALQLILPKKNEWAIKQIDKSSSENDYLTIERIQDLTEELVYNTFDIIVNKFKTSV
FKKSFLNLFQDDKIENIDFKFKSMD FKEKCKTLLMLVKAIRASGVRQDLKSIDLKSDYGR LSSKTNNIGNYQEFKI
INQSITGLFENEVDLLKLGRPAATKKAGQAKKKKGSYPYDVPDYAYPYDVPDYAYPYDVPDYA (SEQ ID NO:
15)。

[0030] 在一些实施例中,发酵乳杆菌菌株AF15-40LB Cas9具有与以下至少80%相同的氨基酸序列:

[0031] MPKKKRKVGKEYHIGLDIGTSSIGWAVTDSQFKLMRIKGKTAIGVRLFEEGKTA AERRTRFTRRRRLK
RRKWRHLHYLDEIFAPHLQEVDENFLRRLKQSNIHPEDPAKNQAFIGKLLFPDLLKKNERGYPTLIKMRDEL PVEQR
AHYPVTNIYKLREAMINEDRQFDLREVYLAVHHIVKYRGHFLNNASVDKFKVGRIDFKSFNVLNEAYEELQNGEG
SFTIEPSKVEKIGQLLLDTKMRKLD RQKAVAKLLEVKVADKEETKR NKQIATAMSKLV LGYKADFATVAMANGNEW
KIDLSSETSEDEIEKFREE LSDAQNDILTEITSLFSQIMLNEIVPNGMSISESMMDRYW THERQLAEVKEYLATQP
ASARKEFDQVYNYIGQAPKEKGF DLEKGLKKILSKKENWKEIDELLKAGDFLPKQRTSANGVIPHQMHQQELDRI
IEKQAKYYPWLATENPATGERDRHQA KYELDQLVSFRIPYYVGPLVTPEVQKATSGAKFAWAKRKEDGEITPWNLW
DKIDRAESAEAFIKRMTVKD TYLLNEDVLPANSLLYQKYNVLNELNNVRVNGRRLSVG IKQDIYTELFKKKKTVKA
GDVASLVMATR GVNKPSVEGLSDPKKFNSNLATYLDLKSIVGDKVDDNRYQMDLENIIEWRSVFEDGEIFADKLT
EVEWL TDEQRSALVKKRYKGWGR LSKLLTGIVDENGQRIIDLMWNTDQNF MQIVNQPVFKEQIDQLNQKAITNDG
MTLRERVESVLDDAYTSPQNKKAIWQVVRVVEDIVKAVGNAPKSI SIEFARNEG NKGEITRSRRTQLQKLFEDQAH
ELVKDTSLTEELEKAPDLSDRYYFYFTQGGKDMYT GDPINFDEISTKYDIDHILPQSFVKDDSLDN RVLVSRAENN
KKSDRVP AKLYAAKM KPYWNQLLKQGLITQRKFENLTMDVDQTIKYRSLGFV KRQLVETRQVIKLTANILGSMYQE
AGTDIIETRAGLTQQLREEFDLPK VREVNDYHHA VDAYLTTFAGQYLNRRYPKLRSFFVYGEYMKFKHGS DLKLRN
FNFFHELMEGDKSQGKVVDQQTGELITRDEVADYFDWVINLKVMLISNETYEETGKYFDASHES SLYLKNQNKK
SKLVVPLKNKLQPEYYGAYTGITQGYMVILKLLDKKGGFGVYRIPRYAADILNKCHDEVAYRNKIAEIISSDPRAP
KSFEVVVPRVLKGTFLVDGEEKFILSSYRYKV NATQLILPVSDIKLIQDNFKALKKLVEMQTKKLEIYDNILRQ
VDKYYKLYDINKFRAKLHDGRSKFVELDDFGQDASKEKVI IKILRGLHFGSDLQNLKEIGFGTTPLGQFQVSEAGI
RLSNTAFIIFKSPTGLFNRKLYLKNLGRPAATKKAGQAKKKKGSYPYDVPDYAYPYDVPDYAYPYDVPDYA (SEQ
ID NO:85)。

[0032] 在一些实施例中,嗜热菌属马赛-P3761Cas9具有与以下至少80%相同的氨基酸序列:

[0033] MPKKKRKVGEEKTNYTIGLDIGTDSVGWAVVKDDLELVKKRMKVLGNTETNYIKKNLWGSLLFESGQT
AKDRRLKRVARRRYERRRNRLTELQKIFAPAIDEVDENFFFR LNESFLVPEDKAFSKNPIFGTLGEDKTY YKYPT
IYHLRQHLADSEEKADVRLIYLALAHMIKYRGHFLIEGKLDTEHIAINENLEQFFESYNALFSEEP IELRKEELIA
IENILREKNSRTVKEKRITSFLKDIGRANKQSPMAFITLIVGKKAKFKAAFNLEEEISLNLTD DSYDENLEILLN
TIGSDFADLFDHAQRVYNAVELAGILSGDVKNTHAKLSAQMVAMYERHKEQLKEYKSF IKANLPDQYDMTFVAPKD
AQKKDLKGYAGYIDGNMSQDSFYKFVKDQLKEVPGSEKFLDSIEKEDFLRKQRSFYNGVIPNQVHLAEMEAILDRQ
ENYYPW LKENREKIIISLLTFRIPYYVGPLADGQSEFAWLERKSDEKIKPWNFSDVVDLDRSAEFIEQLIGRDTYL
PDEYVLPKKS LIYQKYMVFNELTKIAYLDERQKRMNLSSVEKKEIFETL FKKRSKVTEKQLVKFFENYLQIDNPTI

FGIEDAFNADYSTYVELAKVPGMKSMDDPDNEDLMEEIVKILTVFEDRKMRRKQLEKYKERLSPEQIKELAKKHY
TGWGRLSKKLLVGIRDKETQKTILDYLVEDDNHSGGRQHLNRNLMQLINDDRLSFKKTIAELQMIDPSADLYAQVQ
EIAGSPAIIKKGILLGLKIVDEIIRVMGEKPENIVIEMARENQTTARGKALSKRREAKIKEGLAALGSSLLKENLPG
NADLSQRKIYLYYTQNGKDIYLDEPLDFDRLSQYDEDDHIIPQSFTVDNSLDNLVLTNSSQNRGNKKDDVPSLEVVN
RQLAYWRSCLKDAGLMTQRKFDNLTKAMRGGLTDKDRERFIQRQLVETRQITKNVAKLLDMRLNDKKDEAGNKIRET
NIVLLKSAMASEFRKMFRLYKVRELNDYHHAHDAYLNAAIAINLLALYPYMADDFVYGEFRYKKKPQAEKATYEKL
RQWNLIKRFGEKQLFTPDHEDCWNKERDIKTIKKVMGYRQVNVVKKAEERTGMLFKETINGKTNKGSRIPIKKDL
PSKYGGYIEEKMAYYAVISYEDKKKKPGKTIVGISIMDKKEFEYDSISYLGKLGFSNPVVQIILKNYSLIAYPDGR
RRYITGATKTTKGKVELQKANQIAMEQDLVNFYHLKNYDEISHPESYAFVQSHTDYDFDRLFDSIEHYTRRFLDAE
TNINRLRRIYEEEEKKKDPVDIEALVASFIELLKLTSAGAPADFIFMGEAISRRRYNSMTGLFDGQVIYQSLTGLYE
TRMRFEDGKRPAATKKAGQAKKKKGSYPYDVPDYAYPYDVPDYAYPYDVPDYA (SEQ ID NO:87)。

[0034] 在一些实施例中,Cas9蛋白的氨基酸序列包含SEQ ID NO:1、4、8、14、84或86中的至少一个、至少两个、至少三个、至少四个、至少五个、至少六个、至少七个、至少八个、至少九个或至少10个突变。

[0035] 在一些实施例中,该突变是氨基酸取代。

[0036] 在一些实施例中,Cas9蛋白具有切口酶活性。

[0037] 在一些实施例中,本文提供了一种Cas9蛋白,其中该Cas9蛋白在对应于野生型SpCas9的一个或多个氨基酸10、12、17、762、840、854、863、982、983、984、986、987的氨基酸位置处包含切口酶突变。

[0038] 在一些实施例中,至少一个突变产生无活性的Cas9(dCas9)。

[0039] 在一些实施例中,Cas9蛋白包含PAM相互作用结构域、HNH结构域和/或RuvC结构域中的至少一个氨基酸突变。

[0040] 在一些实施例中,本文提供了一种Cas9蛋白,其中氨基酸位置处的突变对应于SirCas9的RuvC结构域中的氨基酸14。

[0041] 在一些实施例中,本文提供了一种Cas9蛋白,其中氨基酸位置处的突变对应于EpeCas9的RuvC结构域中的氨基酸12。

[0042] 在一些实施例中,本文提供了一种Cas9蛋白,其中氨基酸位置处的突变对应于LfeCas9的RuvC结构域中的氨基酸9。

[0043] 在一些实施例中,本文提供了一种Cas9蛋白,其中氨基酸位置处的突变对应于PmaCas9的RuvC结构域中的氨基酸12。

[0044] 在一些实施例中,Cas9蛋白进一步包含核定位序列(NLS)和/或FLAG、HIS或HA标签。

[0045] 在一个方面,本文提供了一种工程化的非天然存在的Cas9融合蛋白,其包含与SEQID NO:1、4、8、14、84或86具有至少80%同一性的Cas9蛋白,并且其中该Cas9蛋白与组蛋白去甲基化酶、转录活化因子或脱氨酶融合。

[0046] 在一些实施例中,本文提供了一种工程化的非天然存在的Cas9融合蛋白,其进一步包含核定位序列(NLS)和/或FLAG、HIS或HA标签。

[0047] 在一些实施例中,本文提供了一种工程化的非天然存在的Cas9融合蛋白,其与SEQID NO:2、5、9、15、85、87、95或96具有至少80%同一性。

[0048] 在一些实施例中,Cas9蛋白与胞嘧啶脱氨酶或与腺苷脱氨酶融合。

[0049] 在一些实施例中,Cas9蛋白与腺苷脱氨酶融合并且具有与以下项至少80%相同的氨基酸序列:

[0050] (a)

[0051] MPAARKVKLDGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHA
EIMALRQGGLVMQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITE
GILADECAALLCRFFRMPRRVFNAQKKAQSSTDGSSGSETPGTSESATPESSGPKKKRKVGKPYSIGLAIGTNSV
GWAVVTDDYKVPAAKMKVLGNTDKQSIKKNLLGALLFDSGETAEATRLKRTARRRYTRRNRLRYLQEIFTGEMNK
VDENFFQRLDDSLVDEDKRGEHHPFGNIAAEVKYHDDFPTIYHLRRHLADTSKKADLRLVYLALAHMIKFRGHF
LYEGDLKAENTDVQALFKDFVEEYDKTIEESHLSEITVDALSILTEKVS KSRLNLI AHYPTEKKNTLFGNLIAL
SLDLHPNFKTNFQLSEDAKLQFSKDTYEEDLEGFLGEVGDEYADLFASAKNLYDAILLSGILTVDDNSTKAPLSAS
MVKRYEEHQKDLKKLKDFIKVNAPDQYNAIFKDKNKKGYASYIESGVKQDEFYKYLKGILLKINGS GDFLDKIDRE
DFLRKQRTFDNGSIPHQIHLQEMHAILRRQGEHYPLKENQDKIEKILTFRIPYYVGPLARKGSRFAWAEYKADEK
ITPWNFDDILDKEKSAEKFITRMTLNDLYLPEEKVLPKHSPLYEAFTVYNELTKVKYVNEQGEAKFFDTNMKQEIF
DHVFKENRKVTDKLLNLYLNKEFEFRIVNLTGLDKENKAFNS SLGTYHDLRKILDKSFLDDKAN EKTIEDIIQT
LTLFEDREMIRQLQKYSDIFTKAQLKKLERRHYTGWGRLSYKLINGIRNKENKKTILDYLI DDGYANRNFMLIN
DDALSFKEEIARAQIIDDVDDIANVVHDLPGSPAIKKGILQSVKIVDELVKVMGHN PANII IEMARENQTTDKGRR
NSQQRLKLLQDSLKNLDNPVNIKNVENQQQLQNDRLFLYYIQNGKDMYTGETLDINNLSQYDIDHIIPQAFIKDNSL
DNRVLTRSDKNRGKSDDVPSIEVVHEMKSFSWKSLLSVKLITQRKFDNLTKAERGGLTEEDKAGFIKRQLVETRQIT
KHVAQILDERFNTEFDGNKRRIRNVKIIITLKS NLVSNFRKEFELYK VREINDYHHAHDAYLNAVVG NALLLKYPQL
EPEFVYGEYPKYNSYRSRSATEKFLFYSNILRFFKKEDIQTNEDEGEIAWNKEKH IKILRKVLSYPQVNIVKKTEE
QTGGFSKESILPKGESDKLIPRKTKNSYWDPKKYGGFDSPVVAYSILVFADVEKGKSKKL RKVQDMVGITIMEKKR
FEKNPVDFLEQRGYRNVRL EKI IKLPKYSLFELENKRRRLASAKELQKGNELVIPQRFTTLLYHSYRIEKDYEP
HREYVEKHKDEFKELLEYISVFSRKYVLADNNLT KIEMLFSKNKDAEVSSLA KSFISLLTFTAFGAPAAFNFFGEN
IDRKRYTSVTECLNATLIHQSI TGLYETRIDL SKLGEDGKRPAATKKAGQAKKKKGSYPYDVPDYAYPYDVPDYAY
PYDVPDYA (SEQ ID NO:20);

[0052] (b)

[0053] MPAARKVKLDGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHA
EIMALRQGGLVMQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITE
GILADECAALLCRFFRMPRRVFNAQKKAQSSTDGSSGSETPGTSESATPESSGPKKKRKVGAKNKDIRYSIGLAIG
TNSVGWAVMDEHYELLKKGNHMMWGSRLFDAAEPAATTRASRSIRRRYNKRREIRLLRDLLGDMVMEVDPTFFIR
LLNVSFLDEEDKQKNLGNDYKDYNLFI EKDFNDKTYDYKYPTIYHLRKELCENKEKADPRLIYLALHHIVKYRGN
FLKEGQSFAKVYEDIEEKLDNTLKKFMSLNDLNDLFDVNDINSMITVLSKIYQRSKKADDLLKIMNPTKEERAAYK
EFTKALVGLKFNVS KMILAQEVKKDDKDIELDFSNDYDSTVDGLQAE LGEYIEFI EMLHSINSWVELQDILGNS
TISAAMVERYEEHKNDLRLV LKKVIREELPDKYNEVFREDNPKLHNYLGYIKYPKNTPVEEFY EYIKRLLAKVDTGE
AREILERIDLEKFMLKQNSRTNGSIPYQMQDEMIQIIDNQSVYYPQLKENREKLISILEFRIPYYFGPLNTHSEF
AWIKKFEDKQKERILPWN YDQIVDIDATAEGFIERMQNTGTYPDPKPVMAKNSLTVSKFEVLNENKIRINGKLIP
VETKKELSLDFMKNKNTITDKKLKDWLVTHQYYDTNEELKIEGYQKDLQFSTSLAPWIDFTKIFGEINASNYQLIE
KIIYDISIFEDKKILKRRLKKVYQLDDLLVDKILKLN YTGWSRLSEKLLTG IKS KNSKETILSILENSNMNLMEII

NDESLGFKQII EESNKKDIEGPFYDEVKKLAGSPA I KRGIWQALLVVQEITKFMKHEPSHIYIEFAREEQEKVRT
ESRIAKLQKIYKDLNLQTKEDQLVYESLKKEDAKKKIDTDALYLYLQMGKSMYSGKPLDIDKLSTYHIDHILPRS
LIKDDSLDNRVLVLPKENEWKLDSETVPFEIRNKMMGFQKLHENG LMSNKKFFSLIRTD FNEKDKKRFINRQLVE
TRQIIKNVAVIINDHYTNTNVTVRAELSHQFRERYKIYKNRDLNDLHHAHDAYIACILGQFIHQNFNMVDNMIY
GQYKKNYKKDVQEHNNGYFILNSMNIHFNDDNSVIWDPSYIGKIKSCFCYKDVYVTKKLEQNDAKLFDLTILPSD
KNSGVTKAKIPVKNYRKDVNKGFGSGDAPIMLAIEADKGKKHVRQVIAFPLRLKNYNDEERIKFIEKEKNLKN
VKILTEVKKNLILINHGYFFITGTNELVNATQLKLSAKNTKNLNLVDANKHNKLESID DANFNEVIQELICKLQ
EPIYSRYNSIGKEFEDSYEKINAVTKQDKLYII EYLI AIMS AKATQGYIKPELAREIGTNGKNKGRIKSFTIDLNK
TTFISTSVTGLFSKKYKLGKRPAA TKKAGQAKKKKGSYPYDVPDYAYPYDVPDYAYPYDVPDYA (SEQ ID NO:
6) ;

[0054] (c)

[0055] MPAAKRVKLDGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLV LNNRVIGEGWNRAIGLHDPTAHA
EIMALRQGGLVMQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITE
GILADECAALLCRFFRMPRRVFNAQKKAQSSTDGSSGSETPGTSESATPESSGPKKKRKVGSI INFQRRGLMETQA
SNQLISSHLKGYPIDYFVGLAIGTSSVGWAVTNKAYELLKFRSHKMWGSRLFDEGESAVARRGFRSMRRRLERRK
LRLKLEELFADAMAQVDPTFFMRLRESKYHYEDKTTGHSSKHILFIDKNYNDQDYFKEYPTVYHLRSELMKSGTD
DIRKLFLAVHHILKYRGNFLYEGATFDSNASTLDDVIKQALENITFNCFCDCNSA ISSIGQILMEAGTKSDKAKAI
EHLVDTYIATD TVDTSSKTQKDQVKEDKKRLKAFANVLGLNASLIDLFGSVEELEEDLKKLQITGDTYDDKRDEL
AKAWSDEIYIIDCKSVYDAI ILLS IKEPGLTISESKVKAFNKHKDDLAILKSLLKSDRSIYNTMFKVDEKGLHNY
VHYIKQRTEETSCNREDFYKYTKKIVEGLSDSKDKEYILSQIELQILLPLQRIKDNQVPIPYQLHLEELKAILAKC
GPKFPFLNEVADGFSVAEKLKMLEFRIPYVGPLNTHHNVNDNGGFAVVRKASGRVTPWNFDDKIDREKSAAFI
KNLTNKCTYLLGEDVLPKSSLLYSEFMLLNELNNVRIDGKPLEKVVKEHLIEAVFKQDHKKMTKNRIEQFLKDNGY
ISETHKHEITGLDGEIKNDLASYRDMVRILGDGFDRSMAEEIITDITIFGESKMLRETLRKKFASCLDDEAIKKL
TKLRYRDWGRLSQKLLNGIEGCDKAGDGT PETIIILMRNFSYNLMELLGDKFSFMERIQEINAKLTEGQIVNPHDI
IDDLALSPAVKRAVWQALRIVDEVAHIKKALPARIFVEVTRSNEKKEKKKDSRQKRLSDLYAAIKKDDVLLNGLNN
EIFGELKSSLA KYDDAALRSKKLYLYTQMGRCAYTGEI IELSLN TDNYDIDHIYPRSLTKDDSF DNVLVCKRTA
NAQKSDAYPISEEIQKTQKPFWTF LKQQLISERKYERLTRITPLTADDLSGFIARQLVETNQSVKAATTLLRRLY
PGVDVVVFVKAENVTD FRHDNNFIKVRSLNHHHHA KDAYLNI VVG NVYHERFTRNFRAFFKKNGANRTYNLAKMFNY
DVNCTNAKD GKAWDVKTSMDTVKKMMSNDVRVTKRLLEQT GALADATIYKATVAGKAKDGAYIGMKT KSSVFADV
SKYGGMTKIKNAYSII VQYTGKKGEVIKEIVPLPIYLTNRNTTDQDLIN YVASIIPQAKDISIIYGKLCINQLVKV
NGFYYYLGGKTNSKFCIDNAIQVIVSNEWIPYLKVLEKFNNMRKDNKDLKANVVSTRALDNKHTIEVRIVEEKNIE
FFDYLVSKLKMPIYQKMGKNAAELSEKGYGLFKMSLEEQSIHLIELLNLLTNQKTTFEVKPLGITASRSTVGSK
ISNQDEFKVINESITGLYSNEVTIVGKRPAATKKAGQAKKKKGSYPYDVPDYAYPYDVPDYAYPYDVPDYA (SEQ
ID NO: 10) ;

[0056] (d)

[0057] MPKKRKVSI INFQRRGLMETQASNQLISSHLKGYPIDYFVGLAIGTSSVGWAVTNKAYELLKFRSH
KMWGSRLFDEGESAVARRGFRSMRRRLERRKLRLKLEELFADAMAQVDPTFFMRLRESKYHYEDKTTGHSSKHIL
FIDKNYNDQDYFKEYPTVYHLRSELMKSGTDDIRKLFLAVHHILKYRGNFLYEGATFDSNASTLDDVIKQALENIT
FNCFCDCNSA ISSIGQILMEAGTKSDKAKAIEHLVDTYIATD TVDTSSKTQKDQVKEDKKRLKAFANVLGLNASL

IDLFGSVEELEEDLKKLQITGDTYDDKRDELAKAWSDEIYIIDDCKSVYDAIILLISKEPGLTISESKVKAFNKHK
DDLAILKSLLKSDRSIYNTMFKVDEKGLHNYVHYIKQGRTEETSCNREDFYKYTKKIVEGLSDSKDKEYILSQIEL
QILLPLQRIKDNGVIPYQLHLEELKAILAKCGPKFPFLNEVADGFSVAEKLKMLEFRIPYYVGPLNTHHNVNDNGG
FAWAVRKASGRVTPWNFDDKIDREKSAAAFIKNLTNKCTYLLGEDVLPKSSLLYSEFMLLNELNNVRIDGKPLEKV
VKEHLIEAVFKQDHHKMTKNRIEQFLKDNGYISETHKHEITGLDGEIKNDLASYRDMVRILGDGFDRSMAEEIITD
ITIFGESKKMLRETLRKKFASCLDDEAIIKKLTCLRDRDWGRLSQKLLNGIEGCDKAGDGTPEIIILMRNFSYNLM
ELLGDKFSFMERIQEINAKLTEGQIVNPHDIIDDLALSPAVKRAVWQALRIVDEVAHIKKALPARIFVEVTRSNN
EKKKKDSRQKRLSDLYAAIKKDDVLLNGLNNEIFGELKSSLAKEYDDAALRSKKLYLYYTQMGRCAYTGEIIELSL
NTDNYDIDHIYPRSLTKDDSFNVLCKRTANAQKSDAYPISEEIQKTQKPFWTFKQQGLISERKYERLTRITPL
TADDLSGFIARQLVETNQSVAATTLLRRLYPGVDDVVFKAENVTDFRHDNNFIKVRSLNHHHHAKDAYLNIVVGN
VYHERFTRNFRAFFKKGANRTYNLAKMFNYDVNCTNAKDGAWDVKTSMDEVKMMDSNDVRVTKRLLQETGALA
DATIYKATVAGKAKDGAYIGMKTSSVFADVSKYGGMTKIKNAYSIIVQYTGKKGEVIKEIVPLPIYLTNRNTTDQ
DLINYVASIIPQAKDISIIYGKLCINQLVKVNGFYYYLGGKTSNKFCIDNAIQVIVSNEWIPYLVLEKFNMRKD
NKDLKANVVSTRALDNKHTIEVRIVEEKNIEFFDYLVSCLKMPIYQKMKGNKAAELSEKGYGLFKKMSLEEQSIHL
IELLNLLTNQKTTFEVKPLGITASRSTVGSKISNQDEFKVINESITGLYSNEVTIVKRPAAATKAGQAKKKKSGSE
TPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNRRVIGEGWNRAIGLHDPTAHAEIMAL
RQGGLVMQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITEGILAD
ECAALLCRFFRMPRRVFNAOKKAQSSTDPAAKRVKLDGSYPYDVPDYAYPYDVPDYAYPYDVPDYA (SEQ ID NO:
11) ;

[0058] (e)

[0059] MPAARVKLDGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNRRVIGEGWNRAIGLHDPTAHA
EIMALRQGGLVMQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITE
GILADECAALLCRFFRMPRRVFNAOKKAQSSTDGSSGSETPGTSESATPESSGPKKKRKVGTKVKDYIIGLAIGTS
SVGWAVTDEAYNVLFKNSKKMWGVRLFDDAKTAEERRGQRGARRRLDRKKERLSLLQDFFAEVAKVDPNFFLRDL
NSDLYMEDKDQKLKSKYTLFNDKDFDKNFHKKYPTIHLLMDLIEDDSKKDIRLVYLACHYLLKNRGHFIFEGQK
FDTKSSFENSLNELKVHLNDEYGLDLEFDNENLINILTPKLNKTAKKKELKSVIGDTKFLKAVSAIMIGSSQKL
DLFENPEDFDDSAIKSVDFSTTSFDDKYSYELALGDKIALVNILKEIYDSSILENLLKEADKSKDGKNKYSNAFV
KKYNKHGQDLKEFKRLVRQYHKSAYFDIFRSEKVNNDNYVSYTKS SISNNKRVKANKFTDQEAIFYKFAKKHLETIK
YKINKVNGSKADLELIDGMLRDMEFKNFMPKIKSSDNGVIPYQLKLMELNKILENQSKHHEFLNVSDEYGSVCDKI
ASIMEFRIPYYVGPLNPNSKYAWIKKQKDSEITPWNFKDVVLDSSREEFIDSLIGRCTYLKDEKVLPAKSLLYNE
YMVNLNLLKLNLDLPITEEMKKKIFDQLFKTRKKVTLKAVANLLKKEFNINGEILLSGTDGDFKQGLNSYNDFKA
IVGDKVDSDDYRDKIEEIIKLIVLYGDDKSYLQKKIKAGYGKYFTDSEIKKMAGLNYKDWGRLSKKLLTGLEGANK
ITGERGSI IHFMREYNLNLMELSASFTEETEEIQKLNVPDDRKLSYEMVDELYLSPSVKRMLWQSLRIVDEIKNIM
GTDSKKIFIEMARGKEEVKARKESRKNQLLKFYKDGKKAFISEIGEERYSYLLSEIEGEEENKFRWDNLYLYYTQL
GRCMYSLEPIDISELSSKNIYDQDHIYPKSKIYDDSIENRVLVKKDLNSKKGNSYIPDEILNKNCYAYWKILYDK
GLIGQKKYTRLTRRTGFTDDELVQFISRQIVETRQATKETANLLKTICKNSEIVYSKAENASRFRQEFDIVKCRAV
NDLHHMHDAYINIIVGNVYNTKFTKDP MNFVKKQEKARSYNLENMFKYDVKRGGYTAWIADDEKGTVKNASIKRIR
KELEGTNYRFRMNYIESGALFNATLQRKNKGSRLPKDKGPKSSIEKYGGYTNINKACFAVLDIKSKNKIERKLMP
VEREIIYAKQKNDKKLSDEIFSKYLKDRFGIEDYRVVYPVVKMRTLLKIDGSYYFITGGSDKTLELRSALQLILPKK

NEWAIKQIDKSSSENDYLTIERIQDLTEELVYNTFDIIVNKFKTSVFKKSFLNLFQDDKIENIDFKFKSMDFKCK
TLLMLVKAIRASGVRQDLKSIDLKSDYGRLLSSKTNNIGNYQEFKIIINQSITGLFENEVDLLKLGRPAATKKAGQA
KKKKSYPYDVPDYAYPYDVPDYAYPYDVPDYA (SEQ ID NO:16);

[0060] (f)

[0061] MPKKKRKVTKVVDYYIGLAIGTSSVGWAVTDEAYNVLFKNSKKMWGVRLFDDAKTAEERRGQRGARR
LDRKKERLSLLQDFFAEVAKVDPNFFLRDLNSDLYMEDKDQKLKSKYTLFNDKDFKDNFHKKYPTIHLLMDLI
EDDSKKDIRLVYLACHYLLKNRGHFIFEGQKFDTKSSFENSLNELKVHLNDEYGLDLEFDNENLINILTPKLNKT
AKKKELKSVIGDTKFLKAVSAIMIGSSQKLVDLFENPEDFDDSAIKSVDFSTTSFDDKYSYELALGDKIALVNIL
KEIYDSSILENLLKEADKSKDGKYNISNAFVKYKNGHQDLKEFKRLVRQYHKSAYFDIFRSEKVNNDNYVSYTKSS
ISNNKRVKANKFTDQEAIFYKFAKKHLETIKYKINKVNGSKADLELIDGMLRDMEFKNFMPKIKSSDNGVIPYQLKL
MELNKILENQSKEHFLNVSDEYGSVCDKIASIMEFRIPYYVGPLNPNSKYAWIKKQKDEITPWNFKDVVDLDSS
REEFIDSLIGRCTYLKDEKVLPAKSLLYNEYMVLNELLNKLNDLPITEEMKKKIFDQLFKTRKKVTLKAVANLLK
KEFNINGEILLSGTDGDFKQGLNSYNDFAIVGDKVSDDYRDKIEEIKLIVLYGDDKSYLQKKIKAGYGYFTD
SEIKKMAGLNYKDWGRLSKKLLTGLEGANKITGERGSIHFMREYNLNMELMSASFTFTEEIQKLNPVDDRKLSY
EMVDELYLSPSVKRWLWQSLRIVDEIKNIMGTDSKKIFIEARGKEEVKARKESRKNQLLFYKDGKKAFISEIGE
ERYSYLLSEIEGEEENKFRWDNLYLYTQLGRCMYSLEPIDISELSSKNIYDQDHIYPKSKIYDDSIENRVLVKKD
LNSKKGNSYPIPEILNKNCYAYWKILYDKGLIGQKKYTRLTRRTGFTDDELVQFISRQIVETRQATKETANLLKT
ICKNSEIVYKAENASRFRQEFDIVKRAVNDLHHMHDAYINIIVGNVYNTKFTKDPNMFVKKQEKARSYNLENMF
KYDVKRGGYTAWIADDEKGTVKNASIKRIRKELEGTNYRFRTRMNYIESGALFNATLQRKNKGSRLPKDKGPKSSIE
KYGGYTNINKACFAVLDIKSKNKKIERKLMPVEREYAKQKNDKKLSDEIFSKYLKDRFGIEDYRVVYPVVKMRTLL
KIDGSYYFITGGSCKTLELRSALQLILPKKNEWAIKQIDKSSSENDYLTIERIQDLTEELVYNTFDIIVNKFKTSVF
KKSFLNLFQDDKIENIDFKFKSMDFKCKTLLMLVKAIRASGVRQDLKSIDLKSDYGRLLSSKTNNIGNYQEFKII
NQSITGLFENEVDLLKLGRPAATKKAGQAKKKKSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDER
EVPVGAVLVNLRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRV
VFGVRNAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTDPAARVKLDGS
YPYDVPDYAYPYDVPDYAYPYDVPDYA (SEQ ID NO:17);

[0062] (g)

[0063] MPAARVKLDGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVNLRVIGEGWNRAIGLHDPTAHA
EIMALRQGGLVMQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITE
GILADECAALLCRFFRMPRRVFNAQKKAQSSTDGSSGSETPGTSESATPESSGPKKKRKGKEYHIGLAIGTSSIG
WAVTDSQFKLMRIKGKTAIGVRLFEEGKTAERTRFTTRRLKRRKWRLHYLDEIFAPHLQEVDFLRLKQSN
IHPEDPAKNQAFIGKLLFPDLLKKNRGYPTLIKMRDELPVEQRAHYPTNIYKLREAMINEDRQFDLREVYLAVH
HIVKYRGHFLNNASVDKFKVGRIDFKSFNVLNEAYEELQNGEGSFTIEPSKVEKIGQLLLDTKMRKLDQKAVAK
LLEVKVADKEETKRNQIATAMSKLVLYKADFATVAMANGNEWKIDLSSETSEDEIEKFREELSDAQNDILTEIT
SLFSQIMLNEIVPNGMSISESMMDRYWTHERRQLAEVKEYLATQPASARKEFDQVYNKYIGQAPKEKGFDEKGLKK
ILSKKENWKEIDELLKAGDFLPKQRTSANGVIPHQMHQQLDRIIEKQAKYYPWLATENPATGERDRHQAKYELDQ
LVSFRIPIYYVGPLVTPEVQKATSGAKFAWAKRKEDGEITPWNLWDKIDRAESAEAFIKRMTVKDITYLLNEDVLPAN
SLLYQKYNVLNELLNVRVNGRRLSVGIKQDIYTELFKKKTVKAGDVASLVMAKTRGVNKPVSVEGLSDPKKFNSNL
ATYLDLKSIVGDKVDDNRYQMDLENIIEWRSVFEDGEIFADKLTEVEWLTDQRSALVKKRYKGWGRLSKKLLTGI

VDENGQRI IDLMWNTDQNFMQIVNQPVFKEQIDQLNQKAITNDGMTLRERVESVLDDAYTSPQNKKAIWQVVRVVE
DIVKAVGNAPKSSISIEFARNEGKGEITRSRRTQLQKLFEDQAH ELVKDTSLTEELEKAPDLSDRYYFYFTQGGKD
MYTGDPINFDEISTKYDIDHILPQSFVKDDSLDNRVLVSRAENNKSDRVP AKLYAAKMKPYWNQLLKQGLITQRK
FENLTMDVDQTIKYRSLGFVKRQLVETRQVIKLTANILGSMYQEAGTDIIETRAGLTKQLREEFDLPKVREVNDYH
HAVDAYLTTFAGQYLNRRYPKLSFFVYGEYMKFKHGS DLKLRNFNFHELMEGDKSQGKVVDQQTGELITTRDEV
ADYFDWVINLKVMLISNETYEETGKYFDASHESSSLYLKNQNKSKLVVPLKNKLQPEYYGAYTGITQGYMVILKL
LDKKGFGVYRIPRYAADILNKCHDEVAYRNKIAEIISSDPRAPKSFEVVVPRVLKGTFLVDGEEKFILSSYRYKV
NATQLILPVSDIKLIQDNFKALKKLVNEMQTKKLEIYDNILRQVDKYYKLYDINKFRAKLHDGRSKFVELDDFGQ
DASKEKVI IKILRGLHFGSDLQNLKEIGFGTTPLGQFQVSEAGIRLSNTAFIIFKSPTGLFNRKLYLKNLGKRPAA
TKKAGQAKKKKGSYPYDVPDYAYPYDVPDYAYPYDVPDYA (SEQ IDNO:88) ;

[0064] (h)

[0065] MPKKKRKVGKEYHIGLAIGTSSIGWAVTDSQFKLMRIKGKTAIGVRLFEEGKTA AERTFRTRRRRLK
RRKWRHLHYLDEIFAPHLQEVDENFLRRLKQSNIHPEDPAKNQAFIGKLLFPDLLKKNERGYPTLIKMRDEL PVEQR
AHYPVTNIYKLREAMINEDRQFDLREVYLAVHHIVKYRGHFLNNASVDKFKVGRIDFDKSFNVLNEAYEELQNGEG
SFTIEPSKVEKIGQLLLDTKMRKLD RQKAVAKLLEVKVADKEETKRNKQIATAMSKLV LGYKADFATVAMANGNEW
KIDLSSETSEDEIEKFREE LSDAQNDILTEITSLFSQIMLNEIVPNGMSISESMMDRYW THERQLAEVKEYLATQP
ASARKEFDQVYNYIGQAPKEKGFDEKGLKKILSKKENWKEIDELLKAGDFLPKQRTSANGVIPHQMHQQELDRI
IEKQAKYYPWLATENPATGERDRHQAKYELDQLVSFRIPYYVGPLVTPEVQKATSGAKFAWAKRKEDGEITPWNLW
DKIDRAESAEAFIKRMTVKDTYLLNEDVLPANSLLYQKYNVLNELNNVRVNGRRLSVG IKQDIYTELFKKKKTVKA
GDVASLVMAKTRGVNKP SVEGLSDPKKFNSNLATYLDLKSIVGDKVDDNRYQMDLENIIEWRSVFEDGEIFADKLT
EVEWLTDEQRSALVKKRYKGWGRLSKLLTGIVDENGQRI IDLMWNTDQNFMQIVNQPVFKEQIDQLNQKAITNDG
MTLRERVESVLDDAYTSPQNKKAIWQVVRVVEDIVKAVGNAPKSSISIEFARNEGKGEITRSRRTQLQKLFEDQAH
ELVKDTSLTEELEKAPDLSDRYYFYFTQGGKDMYTGDPI NFDEISTKYDIDHILPQSFVKDDSLDNRVLVSRAEN
KKSDRVP AKLYAAKMKPYWNQLLKQGLITQRKFENLTMDVDQTIKYRSLGFVKRQLVETRQVIKLTANILGSMYQE
AGTDIIETRAGLTKQLREEFDLPKVREVNDYHHAVDAYLTTFAGQYLNRRYPKLSFFVYGEYMKFKHGS DLKLRN
FNFHELMEGDKSQGKVVDQQTGELITTRDEVADYFDWVINLKVMLISNETYEETGKYFDASHESSSLYLKNQNK
SKLVVPLKNKLQPEYYGAYTGITQGYMVILKLLDKKGFGVYRIPRYAADILNKCHDEVAYRNKIAEIISSDPRAP
KSFEVVVPRVLKGTFLVDGEEKFILSSYRYKVNATQLILPVSDIKLIQDNFKALKKLVNEMQTKKLEIYDNILRQ
VDKYYKLYDINKFRAKLHDGRSKFVELDDFGQDASKEKVI IKILRGLHFGSDLQNLKEIGFGTTPLGQFQVSEAGI
RLSNTAFIIFKSPTGLFNRKLYLKNLKRPAATKKAGQAKKKKSGSETPGTSESATPESSGSEVEFSHEYWMRHALT
LAKRARDEREVPVGAVLV LNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLYDATLYVTFEPCVMCAGA
MIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTDPA
AKRVKLDGSYPYDVPDYAYPYDVPDYAYPYDVPDYA (SEQ ID NO:89) ;

[0066] (i)

[0067] MPAAKRVKLDGSEVEFSHEYWMRHALT LAKRARDEREVPVGAVLV LNNRVIGEGWNRAIGLHDPTAHA
EIMALRQGGLVMQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITE
GILADECAALLCRFFRMPRRVFNAQKKAQSSTDGSSGSETPGTSESATPESSGPKKKRKVG EKKNTYITIGLAIGTD
SVGWAVVKDDLELVKKRMKVLGNTETNYIKKNLWGSLLFESGQTA KDRRLKRVARRRYERRRNRLTELQKIFAPAI
DEVDENFFFRLNESFLVPEDKAFSKNPIFGTLGEDKTYKYPTIYHLRQHLADSEEKADVRLIYLALAHMIKYRG

HFLIEGKLDTEHIAINENLEQFFESYNALFSEPIELRKEELIAIENILREKNSRTVKEKRITSFLKDIGRANKQS
PMMAFITLIVGKKAKFKAAPNLEEEISLNLTDSDYDENLEILLNTIGSDFADLFDHAQRVYNAVELAGILSGDVKN
THAKLSAQMVAMYERHKEQLKEYKSFIKANLPDQYDMTFVAPKDAQKKDLKGYAGYIDGNMSQDSFYKFVKDQLKE
VPGSEKFLDSIEKEDFLRKQRSFYNGVIPNQVHLAEMEAILDRQENYYPWLKENREKIIISLLTFRIPYYVGPLADG
QSEFAWLERKSDEKIKPWNFSDVVDLDRSAEKFIEQLIGRDTYLPDEYVLPKKSLLIYQKYMVFNELTKIAYLDERQ
KRMNLSSVEKKEIFETLFFKKRSKVTEKQLVKFFENYLQIDNPTIFGIEDAFNADYSTYVELAKVPGMKSMDDPDN
EDLMEEIVKILTVFEDRKMRRKQLEKYKERLSPEQIKELAKKHGTGWGRLSKKLLVGIRDKETQKTILDYLVEDDN
HSGGRQHLNRNLMQLINDDRLSFKKTIAELQMIDPSADLYAQVQEIAGSPAIIKKGILLGLKIVDEIIRVMGEKPEN
IVIEMARENQTTARGKALSRRREAKIKEGLAALGSSLLKENLPGNADLSQRKIYLYYTQNGKDIYLDEPLDFDRLS
QYDEDHIIPQSFTVDNSLDNLVLTNSSQNRGNKKDDVPSLEVVRQLAYWRSKLDAGLMTQRKFDNLTKAMRGGLT
DKDRERFIQRQLVETRQITKNVAKLLDMRLNDKKDEAGNKIRETNIVLLKSAMASEFRKMFRLYKVRNLDYHHAH
DAYLNAAIAINLLALYPYMADDFVYGEFRYKKKPQAEKATYEKLRQWNLIKRFGEKQLFTPDHEDCWNKERDIKTI
KKVMGYRQVNVVKKAEERTGMLFKETINGKTNKGSRIPIKKDLDP SKYGGYIEEKMAYYAVISYEDKKKKPGKTIV
GISIMDKKEFEYDSISYLGKLGFSNPVQIILKNYSLIAYPDGRRRYITGATKTTKGKVELQKANQIAMEQDLVNF
IYHLKNYDEISHPESYAFVQSHTDYFDRLFDSIEHYTRRFLDAETNINRLRRIYEEEEKKDPVDIEALVASFIELL
KLTSAGAPADFI FMGEAISRRRYNSMTGLFDGQVIYQSLTGLYETRMRFEDGKRPAATKKAGQAKKKKGSYPYDVP
DYAYPYDVPDYAYPYDVPDYA (SEQ ID NO:91);

[0068] (j)

[0069] MPKKKRKVEKKTNYTIGLAIGTDSVGWAVVKDDLELVKKRMKVLGNTETNYIKKNLWGSLLFESGQTA
KDRRLKRVARRRYERRRNRLTELQKIFAPAIDEVDENFFFRLNESFLVPEDKAFSKNPIFUTLUEDKTYKYPTI
YHLRQHLADSEEKADVRLIYLALAHMIKYRGHFLIEGKLDTEHIAINENLEQFFESYNALFSEPIELRKEELIAI
ENILREKNSRTVKEKRITSFLKDIGRANKQSPMAFITLIVGKKAKFKAAPNLEEEISLNLTDSDYDENLEILLNT
IGSDFADLFDHAQRVYNAVELAGILSGDVKNTHAKLSAQMVAMYERHKEQLKEYKSFIKANLPDQYDMTFVAPKDA
QKKDLKGYAGYIDGNMSQDSFYKFVKDQLKEVPGSEKFLDSIEKEDFLRKQRSFYNGVIPNQVHLAEMEAILDRQE
NYYYPWLKENREKIIISLLTFRIPYYVGPLADGQSEFAWLERKSDEKIKPWNFSDVVDLDRSAEKFIEQLIGRDTYLP
DEYVLPKKSLLIYQKYMVFNELTKIAYLDERQKRMNLSSVEKKEIFETLFFKKRSKVTEKQLVKFFENYLQIDNPTIF
GIEDAFNADYSTYVELAKVPGMKSMDDPDNEDLMEEIVKILTVFEDRKMRRKQLEKYKERLSPEQIKELAKKHGT
GWGRLSKKLLVGIRDKETQKTILDYLVEDDNHSGGRQHLNRNLMQLINDDRLSFKKTIAELQMIDPSADLYAQVQEI
IAGSPAIIKKGILLGLKIVDEIIRVMGEKPENIVIEMARENQTTARGKALSRRREAKIKEGLAALGSSLLKENLPGN
ADLSQRKIYLYYTQNGKDIYLDEPLDFDRLS QYDEDHIIPQSFTVDNSLDNLVLTNSSQNRGNKKDDVPSLEVVR
QLAYWRSKLDAGLMTQRKFDNLTKAMRGGLTDKDRERFIQRQLVETRQITKNVAKLLDMRLNDKKDEAGNKIRETN
IVLLKSAMASEFRKMFRLYKVRNLDYHHAH DAYLNAAIAINLLALYPYMADDFVYGEFRYKKKPQAEKATYEKLR
QWNLIKRFGEKQLFTPDHEDCWNKERDIKTIKKVMGYRQVNVVKKAEERTGMLFKETINGKTNKGSRIPIKKDLDP
SKYGGYIEEKMAYYAVISYEDKKKKPGKTIVGISIMDKKEFEYDSISYLGKLGFSNPVQIILKNYSLIAYPDGRR
RYITGATKTTKGKVELQKANQIAMEQDLVNF IYHLKNYDEISHPESYAFVQSHTDYFDRLFDSIEHYTRRFLDAET
NINRLRRIYEEEEKKDPVDIEALVASFIELLKLTSAGAPADFI FMGEAISRRRYNSMTGLFDGQVIYQSLTGLYET
RMRFEDKRPAATKKAGQAKKKKSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVN
NRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGA
AGSLMDVLHHPGMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTDPAAKRVKLDGSYPYDVPDYAYP

YDVDPDYAYPYDVPDYA (SEQ ID NO:92)。

[0070] 在一些实施例中,Cas9蛋白与胞嘧啶脱氨酶融合并且具有与以下项至少80%相同的氨基酸序列:

[0071] (a)

[0072] MPAARKVKLDTSEKGPSTGDPTLRRRIESWEFDVFYDPRELRKETCLLYEIKWGMSRKIWRSSGKNTT
NHVEVNFICKFTSERRFHSSISCSITWFLSWSPCWECSQAIREFLSQHPGVTLVIIYVARLFWHMDQRNRQGLRDLV
NSGVTIQIMRASEYYHCWRNRFVNYPPGDEAHWPQYPPLWMMLYALELHCIILSLPPCLKISRRWQNHLAFFRLHLQ
NCHYQTIPPHILLATGLIHPSVTWRLKSGGSSGGSSGSETPGTSESATPESSGGSSGGSPKKKRKVGKPYSTGLA
IGTNSVGWAVVTDDYKVPKMKVLGNTDKQSIIKKNLLGALLFDSGETAEATRLKRTARRRYTRRKNRLRYLQEIF
TGEMNKVDENFFQRLDDSFVDEDKRGEHPIFGNIAAEVKYHDDFPTIYHLRRHLADTSKKADLRLVYLALAHMI
KFRGHFLYEGDLKAENTDVQALFKDFVEEYDKTIEESHLSIEITVDALSILTEKVS KSRLNLI AHYPTEKKNTLF
GNLIALSLDLHPNFKTNFQLSEDAKLQFSKDTYEEDLEGFLGEVGDEYADLFASAKNLYDAILLSGILTVDNSTK
APLSASMVKRYEEHQKDLKKLKDFIKVNAPDQYNAIFKDKNKKGYASYIESGVKQDEFYKYLKGILLKINGSDFL
DKIDREDFLRKQRTFDNGSIPHQIHLQEMHAILRRQGEHYFPFLKENQDKIEKILTFRIPIYYVGPLARKGSRFAWAE
YKADEKITPWNFDDILDKEKSAEKFITRMTLNDLYLPPEEKVLPKHSPLYEAFTVYNELTKVKYVNEQGEAKFFDTN
MKQEIFDHVFKENRKVTKDKLLNYLNKEFEFRIVNL TGLDKENKAFNSSLGTYHDLRKILDKSFLDDKAN EKTIE
DIIQTTLTFEDREMIRQLQKYS DIFTKAQLKKLERRHYTGWGRLSYKLINGIRNKENKKTILDYLI DDGYANRNF
MQLINDDALSFKEE IARAQI IDDVDDIANVVHDLPGSPA I KKGILQSVKIVDELVKVMGHN PANI I IEMARENQTT
DKGRRNSQQRLKLLQDSLKNLNDPNVNIKNVENQQLQNDRLFLYYIQNGKDMYTGETLDINNLSQYDIDHIIPQAFI
KDNSLDNRVLTRSDKNRGKSDDVPSIEVVHEMKSFW SKLLSVKLITQRKFDNLTKAERGGLTEEDKAGFIKRQLVE
TRQITKHVAQILDERFNTEFDGNKRRIRNVKIITLKS NLVSNFRKEFELYKVREINDYHHAHDAYLNAVVGNA LLL
KYPQLEPEFVYGEYPKYNSYRSRKSATEKFLFYSNILRFFKKEDIQTNE DGEIAWNKEKH I KILRKVLSYPQVNIV
KKTEEQTGGFSKESILPKGESDKLIPRKTKNSYWDPKKYGGFDPV VAYSILVFADVEKGKSKKL RKVQDMVGITI
MEKKRFEKNPVDFLEQRGYRNVLEKIIKLPKYSLFEL ENKRRRL LASAKELQKGNELVIPQRFTTLLYHSYRIEK
DYEPEHREYVEKHKDEFKELLEYISVFSRKYVLADNNLT KIEMLFSKNKDAEVSSLAKSFISLLTFTA FGAPAAFN
FFGENIDRKRYTSVTECLNATLIHQSI TGLYETRIDL SKLGEDGKRPAATKKAGQAKKKKGSSGGSGGSGGSTNLS
DII EKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKP WALVIQDSNGENKI
KMLSGGSGGSGGSTNLS DII EKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPE
YKP WALVIQDSNGENKIKMLYPYDVPDYAYPYDVPDYAYPYDVPDYA (SEQ ID NO:21);

[0073] (b)

[0074] MPAARKVKLDTSEKGPSTGDPTLRRRIESWEFDVFYDPRELRKETCLLYEIKWGMSRKIWRSSGKNTT
NHVEVNFICKFTSERRFHSSISCSITWFLSWSPCWECSQAIREFLSQHPGVTLVIIYVARLFWHMDQRNRQGLRDLV
NSGVTIQIMRASEYYHCWRNRFVNYPPGDEAHWPQYPPLWMMLYALELHCIILSLPPCLKISRRWQNHLAFFRLHLQ
NCHYQTIPPHILLATGLIHPSVTWRLKSGGSSGGSSGSETPGTSESATPESSGGSSGGSPKKKRKVGSIINFQRRG
LMETQASNQLISSHLKGYPIDYFVGLAIGTSSVGWAVTNKAYELLKFRSHKMWGSRLFDEGESAVARRGFRSMRR
RLERRKLRLKLLLEELFADAMAQVDP TFFMRLRESKYHYEDKTTGHSSKHILFIDKNYNDQDYFKEYPTVYHLRSEL
MKSGTDDIRKLFLAVHHILKYRGNFLYEGATFDSNASTLDDVIKQALENITFNCFCDCNSA ISSIGQILMEAGKTKS
DKAKAIEHLVDTYIATD T VDTSSKTQKDQVKEDKKRLKAFANLVGLNASLIDLFGSVEELEEDLKKLQITGDTYD
DKRDELAKAWSDEIYIIDDCKSVYDAIILL SIKEPGLTISESKVKAFNKHKDDLAILKSLKSDRSIYNTMFKVDE

KGLHNYVHYIKQGRTEETSCNREDFYKYTKKIVEGLSDSKDKEYILSQIELQILLPLQRIKDNVPIPYQLHLEELK
AILAKCGPKFPFLNEVADGFSVAEKLKMLEFRIPYYVGPLNTHHNVNDNGGFAVVRKASGRVTPWNFDDKIDREK
SAAAFIKNLTNKCTYLLGEDVLPKSSLLYSEFMLLNELNNVRIDGKPLEKVVKEHLIEAVFKQDHKKMTKNRIEQF
LKDNGYISETHKHEITGLDGEIKNDLASYRDMVRILGDGDFDRSMAEEIITDITIFGESKKMLRETLRKKFASCLDD
EAIKKLTKLRYRDWGRLSQKLLNGIEGCDKAGDGTPETIIILMRNFSYNLMELLGDKFSFMERIQEINAKLTEGQI
VNPHDIIDDLALSPAVKRAVWQALRIVDEVAHIKKALPARIFVEVTRSNNKNEKKKKDSRQKRLSDLYAAIKKDDVL
LNLNNEIFGELKSSSLAKYDDAALRSKKLYLYTQMGRCAYTGEIIELSLLNTDNYDIDHIYPRSLTKDDSFNVLV
LCKRTANAQKSDAYPISEEIQKTQKPFWTFLLKQQGLISERKYERLTRITPLTADDLSGFIARQLVETNQSVAATT
LLRRLYPGVDVVFVKAENVTDFRHDNNFIKVRSLNHHHAKDAYLNIIVGVNHYHERFTRNFRAFFKKNGANRTYNL
AKMFNYDVNCTNAKDGAWDVKTSMDTVKKMDSNDVRVTKRLLEQTGALADATIIYKATVAGKAKDGAYIGMKTGS
SVFADVSKYGGMTKIKNAYSIIIVQYTGKKGEVIKEIVPLPIYLTNRNTTDQDLINVASIIPQAKDISIIYGKLCI
NQLVKVNGFYYYLGGKTNSKFCIDNAIQVIVSNEWIPYLKVLEKFNNMRKDNKDLKANVSTRALDNKHTIEVRIV
EEKNIEFFDYLVSKLKMPIYQKMGKNAEELSEKGYGLFKKMSLEEQSIHLIELLNLLTNQKTTFEVKPLGITASR
STVGSKISNQDEFKVINESITGLYSNEVTIVGKRPAATKKAGQAKKKKSSGGSSGGSGGSTNLSDIIEKETGKQLV
IQESILMLPEEVVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSGSGG
STNLSDIIEKETGKQLVIQESILMLPEEVVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSN
GENKIKMLYPYDVPDYAYPYDVPDYAY (SEQ ID NO:12);

[0075] (c)

[0076] MPAARVKLDTSEKGPSTGDPTLRRRIESWEFDVFDPRELRKETCLLYEIKWGMSRKIWRSSGKNTT
NHVEVNFIKKFTSERRFHSSISCSITWFLSWSPCWECSQAIREFLSQHPGVTLVIIYVARLFWHMDQRNRQGLRDLV
NSGVTIQIMRASEYYHCWRNFVNYPGDEAHWPQYPPLWMLYALELHCIILSLPPCLKISRRWQNHAFRLHLQ
NCHYQTIPPHILLATGLIHPSVTWRLKSGSSGGSSGSETPGTSESATPESSGGSSGGSPKKRKVGTKVKDYYIG
LAIGTSSVGWAVTDEAYNVLFNSKKMWGVRLFDDAKTAEERRGQRGARRRLDRKKERLSLLQDFFAEVAKVDPN
FFLRDLNSDLYMEDKDQKLKSKYTLFNDKDFKDNFHKKYPTIHHLLMDLIEDDSKKDIRLVYLACHYLLKNRGHF
IFEGQKFDTKSSFENSLNELKVHLNDEYGLDLEFDNENLINILTDPKLNKTAKKKELKSVIGDTKFLKAVSAIMIG
SSQKLVDLFENPEDFDDSAIKSVDFSTTSFDDKYSYELALGDKIALVNILKEIYDSSILENLLKEADKSKDGNKY
ISNAFVKKNKHGQDLKEFKRLVRQYHKSAYFDIFRSEKVNNDYVSYTKSSISNNKRVKANKFTDQEAIFYKFAKKH
LETIKYKINKVNGSKADLELIDGMLRDMEFKNFMPKIKSSDNVPIPYQLKLMELNKILENQSKEHFEFLNVSDEYGS
VCDKIASIMEFRIPYYVGPLNPNSKYAWIKKQKDSEITPWNFKDVVDLDSSREEFIDSLIGRCTYLKDEKVLPAKAS
LLYNEYMVLNELNNLKLNDLPITEEMKKKIFDQLFKTRKKVTLKAVANLLKKEFNINGEILLSGTDGDFKQGLNSY
NDFKAIVGDKVDSDDYRDKIEEIIKLIVLYGDDKSYLQKKIKAGYGYFTDSEIKKMAGLNYKDWGRLSKKLLTGL
EGANKITGERGSIHFMREYNLNMELMSASFTEETIEQKLPVDDRKLSEYEMVDELYLSPSVKRMLWQSLRIVDE
IKNIMGTDSKKIFIEMARGKEEVKARKESRKNQLLFYKDGKKAFISEIGEERYSYLLSEIEGEEENKFRWDNLYL
YYTQLGRCMYSLEPIDISELSSKNIYDQDHIYPKSKIYDDSIENRVLVKKDLNSKKGNSYIPDEILNKNCYAYWK
ILYDKGLIGQKKYTRLTRRTGFTDDELVQFISRQIVETRQATKETANLLKTICKNSEIVYSKAENASRFRQEFDIV
KCRAVNDLHHMHDAYINIIVGNVYNTKFTKDPNMFVKKQEKARSYNLENMFKYDVKRGGYTAWIADDEKGTVKNAS
IKRIRKELEGTNYRFRMNYIESGALFNATLQRKNKGSRLPKDKGPKSSIEKYGGYTNINKACFAVLDIKSKNKIE
RKLMPEVEREIIYAKQKNDKKLSDEIFSKYLKDRFGIEDYRVVYPVVKMRTLLKIDGSYYFITGGSDKTELELSALQL
ILPKKNEWAIKQIDKSSENDYLTIERIQDLTEELVYNTFDIIVNKFKTSVFKKSFLNLFQDDKIENIDFKFKSMDF

KEKCKTLLMLVKAIRASGVRQDLKSIDLKSDYGRLLSSKTNNIGNYQEFKIINQSITGLFENEVDLLKLGRPAATK
KAGQAKKKKGSSGGSGGSGGSTNLSIDIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVM
LLTSDAPEYKPWALVIQDSNGENKIKMLSGGSGGSGGSTNLSIDIEKETGKQLVIQESILMLPEEVEEVIGNKPES
DILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLYPYDVPDYAYPYDVPDYAY (SEQ ID NO:
18) ;

[0077] (d)

[0078] MPAARKVKLDTSEKGPSTGDPTLRRRIESWEFDVFYDPRELKRETCCLLYEIKWGMSRKIWRSSGKNTT
NHVEVNFIIKFTSERRFHSSISCSITWFLSWSPCWECSQAIREFLSQHPGVTLVIIYVARLFWHMDQRNRQGLRDLV
NSGVTIQIMRASEYYHCWRNFVNYPPGDEAHWPQYPPLWMMLYALELHCIILSLPPCLKISRRWQNLHAFRLHLQ
NCHYQTIPPHILLATGLIHPSVTWRLKSGGSGGSGGSETPGTSESATPESSGGSGGSPKKKRKVGEKEYHIGLAI
GTSSIGWAVTDSQFKLMRIKGTAGVRLFEEGKTAERTRFTRRRRLKRRKWRLHYLDEIFAPHLQEVDENFLR
RLKQSNIHPEPDAKNQAFIGKLLFPDLLKKNRGYPTLIKMRDELPVEQRAHYPVTNIYKLREAMINEDRQFDLRE
VYLAVHHIVKYRGHFLNNASVDKFKVGRIDFKSFNVLEAYEELQNGEGSFTIEPSKVEKIGQLLLDTKMRKLDL
QKAVAKLLEVKVADKEETKRKQIATAMSKLVLYKADFATVAMANGNEWKIDLSSETSEDEIEKFREELSDAQN
ILTEITSLSFSQIMLNEIVPNGMSISESMMDRYWTHERRQLAEVKEYLATQPASARKEFDQVYNKYIGQAPKEKGF
DL EKGLKKILSKKENWKEIDELLKAGDFLPKQRTSANGVIPHQMHQQELDRIIEKQAKYYPWLATENPATGERDRHQA
KYELDQLVSFRIPYYVGPLVTPEVQKATSGAKFAWAKRKEDGEITPWNLDKIDRAESAEAFIKRMTVKDITYLLNE
DVL PANSLLYQKYNVLNELNNVRVNGRRLSVGIKQDIYTELFKKKKTVKAGDVASLVMKTRGVNKPVEGLSDPK
KFNSNLATYLDLKSIVGDKVDDNRYQMDLENIIEWRSVFEDGEIFADKLTEVEWLTDEQRSALVKKRYKGWGRLSK
KLLTGIVDENGQRIIDLWNTDQNFMIQVNPVFKEQIDQLNQKAITNDGMTLRERVESVLDDAYTSPQNKKAIWQ
VVRVVEDIVKAVGNAPKSISIEFARNEGKGEITRSRRTQLQKLFEDQAHELVKDTSLTEELEKAPDLSDRYYFYF
TQGGKDMYTGDPIINFDEISTKYDIDHILPQSFVKDDSLDNRLVSRANENKKS DRVPAKLYAAKMKPYWNQLLKQG
LITQRKFENLTMDVDQTIKYRSLGFVKRQLVETRQVIKLTANILGSMYQEAGTDIETRAGLTKQLREEFDLPKVR
EVNDYHHAVDAYLTTFAGQYLNRRYPKLRSFFVYGEYMKFKHGSCLKLRNFNFHHELMEGDKSQGKVVDQQTGELI
TTRDEVADYFDWVINLKVMLISNETYEETGKYFDASHESSLYLKNQNKSKLVVPLKNKLQPEYYGAYTGITQGY
MVILKLLDKKGGFGVYRIPRYAADILNKCHDEVAYRNKIAEIISSDPAPKSFVVVPRVLKGTFLVDGEEKFILS
SYRYKVNATQLILPVSDIKLIQDNFKALKKLNEMQTKKLEIYDNILRQVDKYYKLYDINKFRAKLHDGRSKFVE
LDDFGQDASKEKVIKILRGLHFGSDLQNLKEIGFGTTPLGQFQVSEAGIRLSNTAFIIFKSPTGLFNRKLYLKNL
GKRPAATKKAGQAKKKKGSSGGSGGSGGSTNLSIDIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYD
ESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSGGSGGSTNLSIDIEKETGKQLVIQESILMLPEEVEE
VIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLYPYDVPDYAYPYDVPDYAY (SEQ
ID NO:90) ;

[0079] (e)

[0080] MPAARKVKLDTSEKGPSTGDPTLRRRIESWEFDVFYDPRELKRETCCLLYEIKWGMSRKIWRSSGKNTT
NHVEVNFIIKFTSERRFHSSISCSITWFLSWSPCWECSQAIREFLSQHPGVTLVIIYVARLFWHMDQRNRQGLRDLV
NSGVTIQIMRASEYYHCWRNFVNYPPGDEAHWPQYPPLWMMLYALELHCIILSLPPCLKISRRWQNLHAFRLHLQ
NCHYQTIPPHILLATGLIHPSVTWRLKSGGSGGSGGSETPGTSESATPESSGGSGGSPKKKRKVGEKKTNYTIG
LAIGTDSVGWAVVKDDLELVKKRMKVLGNTETNYIKKNLWGSLLFESGQTAKDRRLKRVARRRYERRRNRLTELQK
IFAPAIDEVDENFFFRLNESFLVPEDKAFSKNPIFGTLGEDKTYKYTYPTIYHLRQHLADSEEKADVRLIYLALAH

MIKYRGHFLIEGKLDTEHIAINENLEQFFESYNALFSEEPIELRKEELIAIENILREKNSRTVKEKRITSFLKDIG
RANKQSPMAFITLIVGKKAKFKAAFNLEEEISLNLTDSDYDENLEILLNTIGSDFADLFDHAQRVYNAVELAGIL
SGDVKNTHAKLSAQMVAMYERHKEQLKEYKSFIKANLPDQYDMTFVAPKDAQKKDLKGYAGYIDGNMSQDSFYKVF
KDQLKEVPGSEKFLDSIEKEDFLRKQRSFYNGVIPNQVHLAEMEAILDRQENYYPWLKENREKIIISLLTFRIPYYV
GPLADGQSEFAWLERKSDEKIKPWNFSDVVDLDRSAEKFIEQLIGRDTYLPDEYVLPKKSLEYQKYMVFNELTKIA
YLDERQKRMNLSSVEKKEIFETLFFKKRSKVTEKQLVKFFENYLQIDNPTIFGIEDAFNADYSTYVELAKVPGMKSM
MDDPDNEDLMEEIVKILTVFEDRKMRRKQLEKYKERLSPEQIKELAKKHGTGWGRLSKKLLVGIRDKETQKTILDY
LVEDDNHSGGRQHLNRNLMQLINDDRLSFKKTI AELQMIDPSADLYAQVQEIAGSPAIIKKGILLGLKIVDEIIRVM
GEKPENIV IEMARENQTTARGKALSKRREAKIKEGLAALGSSLLKENLPGNADLSQRKIYLYYTQNGKDIYLDEPL
DFDRLSQYDEDHIIPQSFTVDNSLDNLVLTNSSQNRGNKKDDVPSLEVVRQLAYWRSCLKDAGLMTQRKFDNLTKA
MRGGLTDKDRERFIQRQLVETRQITKNVAKLLDMRLNDKKDEAGNKIRETNIVLLKSAMASEFRKMFRLYKVR
ELNDYHHAHDAYLNAAIAINLLALPYMADDFVYGEFRYKKKPQAEKATYEKLQWNLIKRFGEKQLFTPDHEDCWNKE
RDIKTIKKVMGYRQVNVVKKAEERTGMLFKETINGKTNKGSRIPIKKDLDP SKYGGYIEEKMAYYAVISYEDKKKK
PGKTIVGISIMDKKEFEYDSISYLGKLGFSNPVVQIILKNYSLIAYPDGRRRYITGATKTTKGKVELQKANQIAME
QDLVNFIIYHLKNYDEISHPESYAFVQSHTDYDFRLFDSIEHYTRRFLDAETNINRLRRIYEEEEKKDPVDIEALVA
SFIELLKLTSAGAPADFI FMGEAISRRRYNSMTGLFDGQVIYQSLTGLYETRMRFEDGKRPAATKKAGQAKKKKGS
SGSGSGSGGSTNLSDIIEKETGKQLVIQESILMLPEEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKP
WALVIQDSNGENKIKMLSGSGSGSGGSTNLSDIIEKETGKQLVIQESILMLPEEEVEEVIGNKPESDILVHTAYDES
TDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLPYDPDYAYPYDPDYAY (SEQ ID NO:93);或

[0081] (f)

[0082] MPAARVKLDTNLSDIIEKETGKQLVIQESILMLPEEEVEEVIGNKPESDILVHTAYDESTDENVMLLT
SDAPEYKPWALVIQDSNGENKIKMLSGSGSGSGGSTNLSDIIEKETGKQLVIQESILMLPEEEVEEVIGNKPESDIL
VHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGSGSGSGSPKKRKRVEKKNYITGLAIGTDSV
GWAVVKDDLELVKKRMKVLGNTETNYIKKNLWGSLLFESGQTAKDRRLKRVARRRYERRRNRLTELQKIFAPAIDE
VDENFFFRLNESFLVPEDKAFSKNPIFGTLGEDKTYKYPTIYHLRQHLADSEEKADVRLIYLALAHMIKYRGHF
LIEGKLDTEHIAINENLEQFFESYNALFSEEPIELRKEELIAIENILREKNSRTVKEKRITSFLKDIGRANKQSPM
MAFITLIVGKKAKFKAAFNLEEEISLNLTDSDYDENLEILLNTIGSDFADLFDHAQRVYNAVELAGILSGDVKNTH
AKLSAQMVAMYERHKEQLKEYKSFIKANLPDQYDMTFVAPKDAQKKDLKGYAGYIDGNMSQDSFYKFKDQLKEVP
GSEKFLDSIEKEDFLRKQRSFYNGVIPNQVHLAEMEAILDRQENYYPWLKENREKIIISLLTFRIPYYVGPLADGQS
EFAWLERKSDEKIKPWNFSDVVDLDRSAEKFIEQLIGRDTYLPDEYVLPKKSLEYQKYMVFNELTKIAYLDERQKR
MNLSSVEKKEIFETLFFKKRSKVTEKQLVKFFENYLQIDNPTIFGIEDAFNADYSTYVELAKVPGMKSMDDPDNED
LMEEIVKILTVFEDRKMRRKQLEKYKERLSPEQIKELAKKHGTGWGRLSKKLLVGIRDKETQKTILDYLVEDDNHS
GGRQHLNRNLMQLINDDRLSFKKTI AELQMIDPSADLYAQVQEIAGSPAIIKKGILLGLKIVDEIIRVMGEKPENIV
IEMARENQTTARGKALSKRREAKIKEGLAALGSSLLKENLPGNADLSQRKIYLYYTQNGKDIYLDEPLDFDRLSQY
DEDHIIPQSFTVDNSLDNLVLTNSSQNRGNKKDDVPSLEVVRQLAYWRSCLKDAGLMTQRKFDNLTKAMRGGLTDK
DRERFIQRQLVETRQITKNVAKLLDMRLNDKKDEAGNKIRETNIVLLKSAMASEFRKMFRLYKVR
ELNDYHHAHDAYLNAAIAINLLALPYMADDFVYGEFRYKKKPQAEKATYEKLQWNLIKRFGEKQLFTPDHEDCWNKERDIKTIKK
VMGYRQVNVVKKAEERTGMLFKETINGKTNKGSRIPIKKDLDP SKYGGYIEEKMAYYAVISYEDKKKKPGKTIVGI
SIMDKKEFEYDSISYLGKLGFSNPVVQIILKNYSLIAYPDGRRRYITGATKTTKGKVELQKANQIAMEQDLVNFIIY

HLKNYDEISHPESYAFVQSHTDYFDRLFDSIEHYTRRFLDAETNINRLRRIYEEKKKDPVDIEALVASFIELLKL
TSAGAPADFI FMGEAISRRRYSMTGLFDGQVIYQSLTGLYETRMRFEDKRPAATKKAGQAKKKKGSSGGSSGGSS
GSETPGTSESATPESSGGSSGGSTSEKGPSTGDPTLRRRIESWEFDVFYDPRELKRETCCLLYEIKWGMSRKIWRSS
GKNTTNHVEVNFIIKFTSERRFHSSISCSITWFLSWSPCWECSQAIREFLSQHPGVTLVIYVARLFWHMDQRNRQG
LRDLVNSGVTIQIMRASEYYHCWRNFVNYPGDEAHWPQYPPLWMMLYALELHCIIISLPPCLKISRRWQNHLAFF
RLHLQNCHYQTIPPHILLATGLIHPSVTWRYPYDVPDYAYPYDVPDYAYPYDVPDYA (SEQ ID NO:94)。

[0083] 在一些实施例中,星座链球菌Cas9蛋白识别包含5'-NGG-3'的PAM序列。

[0084] 在一些实施例中,星座链球菌Cas9蛋白识别包含5'-NGC-3'的PAM序列。

[0085] 在一些实施例中,本文公开的Cas9蛋白(例如,SirCas9、VapCas9、EpeCas9、LfeCas9或PmaCas9)识别包含5'-NGC-3'的PAM序列。

[0086] 在一些实施例中,夏普氏菌Cas9蛋白识别包含5'-NAGHC-3'的PAM序列,其中H=A、C或T。

[0087] 在一些实施例中,小韦荣球菌Cas9蛋白识别包含5'-NRHRRH-3'的PAM序列,其中H是腺嘌呤、胞嘧啶或胸腺嘧啶,并且R是腺嘌呤或鸟嘌呤。

[0088] 在一些实施例中,秘鲁埃扎基拉菌Cas9蛋白识别包含5'-NGG-3'的PAM序列。

[0089] 在一些实施例中,发酵乳杆菌菌株AF15-40LB Cas9蛋白识别包含5'-NGG-3'的PAM序列。

[0090] 在一些实施例中,嗜脲菌属马赛-P3761 Cas9蛋白识别包含5'-NNAAA-3'的PAM序列。

[0091] 在一些实施例中,提供了一种核酸,其编码该Cas9蛋白。

[0092] 在一些实施例中,核酸经密码子优化以用于在哺乳动物细胞中表达。

[0093] 在一些实施例中,核酸经密码子优化以用于在人类细胞中表达。

[0094] 在一些实施例中,提供了一种真核细胞,其包含该Cas9蛋白。

[0095] 在一些实施例中,细胞是人类细胞。在一些实施例中,细胞是植物细胞。

[0096] 在一个方面,提供了一种切割真核细胞中的靶核酸的方法,该方法包括:使该细胞与如本文所述的Cas9和RNA向导或编码该RNA向导的核酸接触,其中该RNA向导包括能够与该靶核酸杂交的同向重复序列和间隔子序列,并且其中该Cas9蛋白能够与该RNA向导结合并导致与该RNA向导互补的靶核酸序列中的断裂。

[0097] 在一个方面,提供了一种改变真核细胞中的靶核酸表达的方法,该方法包括:使该细胞与如本文所述的Cas9和RNA向导或编码该RNA向导的核酸接触,其中该RNA向导包括能够与该靶核酸杂交的同向重复序列和间隔子序列,并且其中该Cas9蛋白能够与该RNA向导结合并导致与该RNA向导互补的靶核酸序列中的断裂。

[0098] 在一个方面,提供了一种改变真核细胞中的靶核酸表达的方法,该方法包括:使该细胞与如本文所述的Cas9和RNA向导或编码该RNA向导的核酸接触,其中该RNA向导包括能够与该靶核酸杂交的同向重复序列和间隔子序列,并且其中该Cas9蛋白能够与该RNA向导结合并编辑与该RNA向导互补的靶核酸序列。

[0099] 在一个方面,提供了一种修饰真核细胞中的靶核酸的方法,该方法包括:使该细胞与如本文所述的Cas9和RNA向导或编码该RNA向导的核酸接触,其中该RNA向导包括能够与该靶核酸杂交的同向重复序列和间隔子序列,并且其中该Cas9蛋白能够与该RNA向导结合

并编辑与该RNA向导互补的靶核酸序列。

[0100] 在一些实施例中,Cas9蛋白是无活性的Cas9 (dCas9)。

[0101] 在一些实施例中,dCas9与脱氨酶融合。

[0102] 在一些实施例中,RNA向导包含crRNA和tracrRNA。

[0103] 在一些实施例中,RNA向导包含sgRNA。

[0104] 在一些实施例中,与星座链球菌Cas9一起使用的sgRNA包含支架,该支架包含与以下具有至少约80%同一性的序列:

[0105] 5'-GUUUUAGAGCUGUGCUGUUUAAACAACACAGCAAGUAAAAUAAGGCUUUGUCCGUACUCAAGCUUGCAAAAGCGUGCACC GAUUCGGUGCU-3' (SEQ ID NO:3)。

[0106] 在一些实施例中,与夏普氏菌Cas9一起使用的sgRNA包含支架,该支架包含与以下具有至少约80%同一性的序列:

[0107] 5'-GUUUUAGAGUUGUGUUAUUGAAAAUAACACAACGAGUAAAAUAAAGCUUAUGCUUAAAUGCCAGCUUUGCUGGUGUCAUUUAGAUGACUUUACUAAGGUUGCUUCGGCAACCUUUUU-3' (SEQ ID NO:7)。

[0108] 在一些实施例中,与小韦荣球菌Cas9一起使用的sgRNA包含支架,该支架包含与以下具有至少约80%同一性的序列:

[0109] 5'-GUUUGAGAGUAGUGUGAAAACAUUACGAGUUCAAAUACAAAUAAUUUACAAUGCCUUCGGGCUGCCCAGCUAGGGCACCUCUCUCAAUUCUUCGGAAUUGAGUU-3' (SEQ ID NO:13)。

[0110] 在一些实施例中,与秘鲁埃扎基拉菌Cas9一起使用的sgRNA包含支架,该支架包含与以下具有至少约80%同一性的序列:

[0111] 5'-GUUUGAGAGUUAUGUAAUUGAAAAUACAUGACGAGUUCAAAUAAAAUUUAUUCAAACCGCCUAAUUUAUAGGCCGCAGAUUUCUGCAUUAUGCUUGCUAUUGCAAGCUU-3' (SEQ ID NO:19)。

[0112] 在一些实施例中,与发酵乳杆菌菌株AF15-40LB Cas9一起使用的sgRNA包含支架,该支架包含与以下具有至少约80%同一性的序列:

[0113] 5'-GUCUUGGAGAGUGUGAAAACACUCAUAGUCAAGAUCAAACGAGUGGUUUUCCACGAGUUAUUACUUUUGAGGUCUUAUUGGCCCAUACAUAAAAAGGAGUCGGAAUUUCCGGCUCCUUUUCUU-3' (SEQ ID NO:95)。

[0114] 在一些实施例中,与嗜脲菌属马赛-P3761 Cas9一起使用的sgRNA包含支架,该支架包含与以下具有至少约80%同一性的序列:

[0115] 5'-GUUUUAGAGCCAUGUAGAAUACAUGCAAGUAAAAUAAGGCUUUGUCCGUAAUCAACUUGAAAAGUGGCGCUGUUUCGGCGCUU-3' (SEQ ID NO:96)。

[0116] 在一些实施例中,crRNA包含长度介于约16与26个核苷酸之间的向导序列。

[0117] 在一些实施例中,crRNA包含长度介于18与24个核苷酸之间的向导序列。

[0118] 在一些实施例中,靶核酸中的断裂是单链断裂或双链断裂。

[0119] 在一些实施例中,靶核酸中的断裂是单链断裂。

[0120] 在一些实施例中,Cas9蛋白是切割靶核酸序列的两条链的核酸酶。在一些实施例中,Cas9是切割靶核酸序列的一条链的切口酶。

[0121] 在一些实施例中,靶核酸位于前间隔序列邻近基序(PAM)序列的5'处。

[0122] 在一些实施例中,Cas9可操作地连接至启动子序列以用于在真核细胞中表达,并且其中向导RNA可操作地连接至启动子序列以用于在真核细胞中表达。

- [0123] 在一些实施例中,真核细胞是人类细胞。
- [0124] 在一些实施例中,启动子序列是真核启动子或病毒启动子。
- [0125] 在一个方面,本文提供了一种工程化的非天然存在的CRISPR-Cas系统,其包括:RNA向导或编码该RNA向导的核酸,其中该RNA向导包含能够与靶核酸杂交的同向重复序列和间隔子序列;以及经密码子优化的CRISPR相关(Cas)蛋白,其与SEQ ID NO:1、4、8、14、84或86具有至少80%序列同一性,并且其中该Cas蛋白能够与该RNA向导结合并导致与该RNA向导互补的靶核酸序列中的断裂。
- [0126] 在一些实施例中,本文提供了一种工程化的非天然存在的CRISPR-Cas系统,其包括经密码子优化的CRISPR相关(Cas)蛋白,其与SEQ ID NO:2、5、9、15、85、87、95或96具有至少80%序列同一性,并且其中该Cas蛋白能够与该RNA向导结合并导致与该RNA向导互补的靶核酸序列中的断裂。
- [0127] 在一个方面,本文提供了一种工程化的非天然存在的CRISPR-Cas系统,其包括:RNA向导或编码该RNA向导的核酸,其中该RNA向导包含能够与靶核酸杂交的同向重复序列和间隔子序列;以及经密码子优化的CRISPR相关(Cas)蛋白,其与SEQ ID NO:1、4、8、14、84或86具有至少80%序列同一性;其中该Cas蛋白与脱氨酶融合,并且其中Cas蛋白融合物能够与该RNA向导结合并编辑与该RNA向导互补的靶核酸序列。
- [0128] 在一些实施例中,工程化的非天然存在的CRISPR-Cas系统包括经密码子优化的CRISPR相关(Cas)蛋白,该经密码子优化的CRISPR相关蛋白进一步包含核定位序列(NLS)和/或FLAG、HIS或HA标签。
- [0129] 在一些实施例中,工程化的非天然存在的CRISPR-Cas系统包括与SEQ ID NO:2、5、9、15、85、87、95或96具有至少80%序列同一性的经密码子优化的CRISPR相关(Cas)蛋白,其中该Cas蛋白与脱氨酶融合,并且其中Cas蛋白融合物能够与该RNA向导结合并编辑与该RNA向导互补的靶核酸序列。
- [0130] 在一个实施例中,Cas9蛋白是无活性的Cas9(dCas9)。
- [0131] 在一个实施例中,RNA向导包含crRNA和tracrRNA。
- [0132] 在一个实施例中,RNA向导包含sgRNA。
- [0133] 在一个实施例中,Cas蛋白可操作地连接至启动子序列以用于在真核细胞中表达,并且其中向导RNA可操作地连接至启动子序列以用于在真核细胞中表达。
- [0134] 在一个实施例中,真核细胞是人类细胞。
- [0135] 在一个实施例中,启动子序列是真核启动子序列。
- [0136] 在一个实施例中,提供了一种编码本文所述的系统的核酸。
- [0137] 在一个实施例中,提供了一种包含本文所述的系统的载体。
- [0138] 在一个实施例中,载体是质粒载体或病毒载体。
- [0139] 在一个实施例中,病毒载体是腺相关病毒(AAV)载体或慢病毒载体。
- [0140] 在一个实施例中,病毒载体是AAV载体。
- [0141] 在一个实施例中,多于一种AAV载体用于封装该系统。
- [0142] 在一个实施例中,治疗对此有需要的受试者的病症或疾病的方法包括向该受试者施用本文所述的系统,其中该向导RNA和与病况或疾病相关的靶核酸的至少10个核苷酸互补;其中该Cas蛋白与该向导RNA缔合;其中该向导RNA与该靶核酸结合;其中该Cas蛋白导致

靶核酸中的断裂,任选地其中Cas9是与脱氨酶融合的无活性的Cas9(dCas9)并产生该靶核酸中的一个或多个碱基编辑,从而治疗该病症或疾病。

[0143] 在一些实施例中,向导RNA与约18-24个核苷酸互补。

[0144] 在一些实施例中,向导RNA与20个核苷酸互补。

[0145] 在一些实施例中,碱基编辑器包括融合蛋白。

[0146] 在一些实施例中,碱基编辑器包括腺苷脱氨酶结构域或胞苷脱氨酶结构域。

[0147] 在一些实施例中,本文提供了一种编辑多核苷酸的核碱基的方法,该方法包括使该多核苷酸和与一种或多种向导RNA复合的碱基接触,其中碱基编辑器包括腺苷脱氨酶结构域,并且其中该一种或多种向导RNA靶向该碱基编辑器以实现该多核苷酸中的A • T至G • C的改变。

[0148] 在一些实施例中,本文提供了一种编辑多核苷酸的核碱基的方法,该方法包括使该多核苷酸和与一种或多种向导RNA复合的碱基编辑器接触,其中碱基编辑器包括胞苷脱氨酶结构域,并且其中该一种或多种向导RNA靶向该碱基编辑器以实现该多核苷酸中的C • G至T • A的改变。

[0149] 在一些实施例中,编辑引起靶多核苷酸序列中的小于50%的插入缺失形成。

[0150] 在一些实施例中,编辑产生点突变。

[0151] 定义

[0152] 为了更容易理解本发明,下面首先定义某些术语。用于以下术语和其它术语的另外的定义在整个说明书中阐述。

[0153] 一(A)或一(An):本文使用的冠词“一(a)”和“一(an)”是指冠词的语法对象中的一个或多个(即,至少一个)。举例来说,“要素”意指一个要素或多个要素。

[0154] 大约(Approximately)或约(about):如本文中所用,术语“大约”或“约”在应用于所关注的一个或多个值时是指与所述参考值类似的值。在某些实施方案中,术语“大约”或“约”是指处于在任一方向上(大于或小于)所述参考值的25%、20%、19%、18%、17%、16%、15%、14%、13%、12%、11%、10%、9%、8%、7%、6%、5%、4%、3%、2%、1%或更小的值的范围,除非另有说明或从上下文可以明显看出(除非该数字将超过一个可能的值的100%)。

[0155] 与.....相关联(Associated with):如果一个事件或实体的存在、水平和/或形式与另一个事件或实体的存在、水平和/或形式有关,则两个事件或实体彼此“相关联”,如所述术语在本文所使用。例如,如果特定实体(例如,多肽)的存在、水平和/或形式与疾病、病症或病状的发病率和/或易感性有关(例如,跨相关群体),则所述特定实体被认为与特定疾病、病症或病状相关。在一些实施例中,如果两个或更多个实体直接或间接相互作用,使得其在物理上彼此接近和保持彼此接近,则其在物理上彼此“缔合”。在一些实施方案中,在物理上彼此相关的两个或更多个实体彼此共价连接;在一些实施方案中,在物理上彼此相关的两个或更多个实体彼此不共价连接,而是非共价关联,例如借助于氢键、凡得瓦尔力相互作用(van der Waals interaction)、疏水性相互作用、磁性和其组合相关联。

[0156] 碱基编辑器:“碱基编辑器(BE)”或“核碱基编辑器(NBE)”意指结合多核苷酸并具有核碱基修饰活性的药剂。在各个实施例中,碱基编辑器包括核碱基修饰多肽(例如,脱氨酶)和多核苷酸可编程核苷酸结合结构域以及向导多核苷酸(例如,向导RNA)。在各个实施

例中,药剂是生物分子复合物,该生物分子复合物包括具有碱基编辑活性的蛋白质结构域,即,能够修饰核酸分子(例如,DNA)内的碱基(例如,A、T、C、G或U)的结构域。在一些实施例中,所述多核苷酸可编程DNA结合结构域融合或连接到脱氨酶结构域。在一个实施例中,所述药剂是包括一个或多个具有碱基编辑活性的结构域的融合蛋白。在另一实施例中,具有碱基编辑活性的蛋白质结构域与向导RNA连接(例如,通过向导RNA上的RNA结合基序和与脱氨酶融合的RNA结合结构域连接)。在一些实施例中,所述具有碱基编辑活性的结构域能够使核酸分子内的碱基脱氨基。在一些实施例中,所述碱基编辑器能够使DNA分子内的一个或多个碱基脱氨基。在一些实施例中,所述碱基编辑器能够使DNA内的胞嘧啶(C)或腺苷(A)脱氨基。在一些实施例中,所述碱基编辑器能够使DNA内的胞嘧啶(C)和腺苷(A)脱氨基。在一些实施例中,所述碱基编辑器是胞苷碱基编辑器(CBE)。在一些实施例中,碱基编辑器是腺苷碱基编辑器(ABE)。在一些实施例中,所述碱基编辑器是腺苷碱基编辑器(ABE)和胞苷碱基编辑器(CBE)。在一些实施方案中,碱基编辑器为与腺苷脱氨酶融合的可核酸酶活性的Cas9(dCas9)。在一些实施方案中,碱基编辑器与碱基切除修复抑制剂(例如UGI结构域或dISN结构域)融合。在一些实施方案中,融合蛋白包括与脱氨酶融合的Cas9切口酶和碱基切除修复抑制剂,如UGI或dISN结构域。在其它实施例中,所述碱基编辑器是无碱基碱基编辑器。碱基编辑器的细节在国际PCT申请第PCT/2017/045381号(WO2018/027078)和第PCT/US2016/058344号(WO2017/070632)中进行了描述,所述国际PCT申请中的每一个都通过引用整体并入本文。还参见Komor,A.C.等人,“Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage”*Nature* 533,420-424(2016); Gaudelli,N.M.等人,“Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage”*Nature* 551,464-471(2017);Komor,A.C.等人,“Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity”*Science Advances* 3:eaa04774(2017);以及Rees,H.A.等人,“Base editing:precision chemistry on the genome and transcriptome of living cells.”*Nat Rev Genet.*2018年12月;19(12):770-788.doi:10.1038/s41576-018-0059-1,这些文献的全部内容通过引用并入本文。

[0157] 碱基编辑活性:“碱基编辑活性”意指用于以化学方式改变多核苷酸内的碱基。在一个实施例中,第一碱基转化为第二碱基。在一个实施例中,碱基编辑活性是胞苷脱氨酶活性,例如将靶标C•G转化为T•A。在另一实施例中,碱基编辑活性是腺苷或腺嘌呤脱氨酶活性,例如将A•T转化为G•C。在另一实施例中,碱基编辑活性是胞嘧啶或胞苷脱氨酶活性,例如将靶标C•G转化为T•A,以及腺苷或腺嘌呤脱氨酶活性,例如将A•T转化为G•C。

[0158] 碱基编辑器系统:术语“碱基编辑器系统”是指用于编辑靶核苷酸序列的核碱基的系统。在各个实施例中,碱基编辑器(BE)系统包括(1)多核苷酸可编程核苷酸结合结构域(例如,Cas9)、脱氨酶结构域和胞苷脱氨酶结构域,其用于将靶核苷酸序列中的核碱基脱氨基;以及(2)一种或多种向导多核苷酸(例如,向导RNA),其与多核苷酸可编程核苷酸结合结构域结合。在各个实施例中,碱基编辑器(BE)系统包括选自腺苷脱氨酶或胞苷脱氨酶的核碱基编辑器结构域,以及具有核酸序列特异性结合活性的结构域。在一些实施例中,碱基编辑器系统包括(1)碱基编辑器(BE),所述碱基编辑器包括多核苷酸可编程DNA结合结构域和脱氨酶结构域,其用于将靶核苷酸序列中的一个或多个核碱基脱氨基;以及(2)一种或多种

向导RNA与多核苷酸可编程DNA结合结构域结合。在一些实施例中,所述多核苷酸可编程核苷酸结合结构域是多核苷酸可编程DNA结合结构域。在一些实施例中,所述碱基编辑器是胞苷碱基编辑器(CBE)。在一些实施例中,所述碱基编辑器是腺嘌呤或腺苷碱基编辑器(ABE)。在一些实施例中,所述碱基编辑器是腺嘌呤或腺苷碱基编辑器(ABE)或胞苷碱基编辑器(CBE)。

[0159] 在一些实施例中,多核苷酸可编程核苷酸结合结构域可以通过与脱氨酶结构域非共价相互作用或缔合来将脱氨酶结构域靶向靶核苷酸序列。例如,在一些实施例中,核碱基编辑组分,例如脱氨酶组分,可以包括另外的异源部分或结构域,其能够与作为多核苷酸可编程核苷酸结合结构域的一部分的另外的异源部分或结构域相互作用、缔合或能够与其形成复合物。在一些实施例中,另外的异源部分可以能够与多肽结合、相互作用、缔合或形成复合物。在一些实施例中,另外的异源部分可以能够与多核苷酸结合、相互作用、缔合或形成复合物。在一些实施例中,另外的异源部分可以能够与向导多核苷酸结合。在一些实施例中,另外的异源部分可以能够与多肽接头结合。另外的异源部分可以是蛋白质结构域。在一些实施例中,另外的异源部分可以是K同源(KH)结构域、MS2外壳蛋白结构域、PP7外壳蛋白结构域、SfMu Com外壳蛋白结构域、无菌(steril) α 基序、端粒酶Ku结合基序和Ku蛋白、端粒酶Sm7结合基序和Sm7蛋白或RNA识别基序。

[0160] 生物活性:如本文所用,短语“生物活性”是指在生物系统中,且尤其在生物体中具有活性的任何药剂的特征。例如,当施用于生物体时对所述生物体具有生物效应的药剂被视为是生物活性的。在特定实施例中,当肽具有生物活性时,所述肽的共享肽的至少一种生物活性的部分通常被称为“生物活性”部分。

[0161] 切割:如本文所用,切割是指由本文所述的CRISPR系统的核酸酶产生的靶核酸中的断裂。在一些实施例中,切割事件是双链DNA断裂。在一些实施例中,切割事件是单链DNA断裂。在一些实施例中,切割事件是单链RNA断裂。在一些实施例中,切割事件是双链RNA断裂。

[0162] 互补:如本文所用,互补是指形成Watson-Crick碱基配对以使得A碱基与T配对并且C碱基与G配对,或与第二核酸链上的碱基形成非传统碱基配对的核酸链。换句话说,其是指在适当的条件下彼此杂交的核酸。

[0163] 成簇间隔短回文重复序列(CRISPR)相关(Cas)系统:如本文所用,CRISPR-Cas9系统是指参与CRISPR效应子的表达或指导CRISPR效应子活性的核酸和/或蛋白质,包括编码CRISPR效应子的序列、RNA向导以及来自CRISPR基因座的其他序列和转录物。在一些实施例中,CRISPR系统是工程化、非天然存在的CRISPR系统。在一些实施例中,CRISPR系统的组分可以包含编码系统的一种或多种组分的核酸(例如,载体)、蛋白质形式的一种或多种组分或其组合。

[0164] CRISPR阵列:如本文所用,术语“CRISPR阵列”是指包括CRISPR重复序列和间隔子的核酸(例如,DNA)区段,该区段以第一个CRISPR重复序列的第一个核苷酸开始并且以最后一个(末端)CRISPR重复序列的最后一个核苷酸结束。通常,CRISPR阵列中的每个间隔子位于两个重复序列之间。如本文所用,术语“CRISPR重复序列”或“CRISPR同向重复序列”或“同向重复序列”是指多个短的同向重复序列,所述短的直接重复序列在CRISPR阵列内显示非

常少或无序列变异。

[0165] CRISPR相关蛋白 (Cas): 如本文所用, 术语“CRISPR相关蛋白”、“CRISPR效应子”、“效应子”或“CRISPR酶”是指执行酶活性或结合由RNA向导指定的核酸上的靶位点的蛋白。在不同实施例中, CRISPR效应子具有核酸核酸内切酶活性、切口酶活性、核酸外切酶活性、转座酶活性和/或切除活性。在一些实施例中, Cas是高精度Cas。在一些实施例中, Cas是高保真Cas。在一些实施例中, Cas是SuperFi-Cas。在一些实施例中, 高精度、高保真和SuperFi-Cas如以下中所描述: Bravo, J. 等人 *Structural basis for mismatch surveillance by CRISPR-Cas9* Nature, 603, 2022年3月。

[0166] crRNA: 如本文所用, 术语“CRISPR RNA”或“crRNA”是指包括由CRISPR效应子使用以靶向特定核酸序列的向导序列的RNA分子。通常, crRNA含有介导靶标识别的序列和与tracrRNA形成双链体的序列。在一些实施例中, crRNA:tracrRNA双链体与CRISPR效应子结合。

[0167] 离体: 如本文所用, 术语“离体”是指在细胞或组织中发生的事件, 这些细胞或组织在多细胞生物体外而不是在多细胞生物体内生长。

[0168] 功能等同物或功能类似物: 如本文所用, 术语“功能等同物”或“功能类似物”在氨基酸序列的功能衍生物的上下文中表示保留与原始序列的生物活性 (功能或结构) 基本上相似的生物活性 (功能或结构) 的分子。功能衍生物或等同物可以是天然衍生物或是合成制备的。示例性功能衍生物包含具有一个或多个氨基酸的取代、缺失或添加的氨基酸序列, 条件是蛋白质的生物活性是保守的。取代氨基酸理想地具有与所述取代氨基酸相似的化学物理性质。理想的相似化学物理性质包含电荷、体积、疏水性、亲水性等的相似性。

[0169] 半衰期: 如本文所用, 术语“半衰期”是量 (诸如蛋白质浓度或活性) 下降到其在时间段开始时测量的值的一半所需的时间。

[0170] 改善、增加或减少: 如本文所用, 术语“改善”、“增加”或“减少”或其语法等效物表示相对于基线测量值的值, 该基线测量值诸如在开始本文所述的治疗之前在同一体中的测量值, 或在不存在本文所述的治疗的情况下在对照受试者 (或多个对照受试者) 中的测量值。“对照受试者”是患有与正在治疗的受试者相同形式的疾病的受试者, 所述受试者与正在接受治疗的受试者的年龄大致相同。

[0171] 抑制: 如本文所用, 术语“抑制 (inhibition)”、“抑制 (inhibit)”和“抑制 (inhibiting)”是指减少或降低所关注的蛋白质或基因的活性和/或表达的过程或方法。通常, 抑制蛋白质或基因是指将蛋白质或基因的表达或相关活性降低至少10%或更多, 例如, 20%、30%、40%或50%、60%、70%、80%、90%或更多, 或表达或相关活性降低大于1倍、2倍、3倍、4倍、5倍、10倍、50倍、100倍或更多, 如通过本文所述或本领域公认的一种或多种方法测量的。

[0172] 杂交: 如本文所用, 术语“杂交”是指两个或更多个核酸通过两个核酸的碱基之间的氢键合 (通过Watson-Crick配对)、Hoogsteen结合或其他序列特异性结合而彼此结合的反应。能够与另一序列杂交的序列被称为该序列的“互补序列”, 并且称为“互补的”或显示出“互补性”。

[0173] 插入缺失: 如本文所用, 术语“插入缺失”是指核酸序列中碱基的插入或缺失。它通常会导致突变并且是遗传变异的一种常见形式。

[0174] 体外:如本文所用,术语“体外”是指发生在人工环境,例如测试管或反应容器中、细胞培养物等中而不是多细胞生物体内的事件。

[0175] 体内:如本文所用,术语“体内”是指在多细胞生物体(诸如人类或非人类动物)内发生的事件。在基于细胞的系统的情况下,该术语可用于指代在活细胞(与例如体外系统相反)内发生的事件。

[0176] 接头:术语“接头”是指用于接合两个或更多个实体的任何手段、实体或部分。在一些实施例中,接头是共价接头。在一些实施例中,接头是非共价接头。共价接头的实例包括共价键或共价附接至待连接的一种或多种蛋白质或结构域的接头部分。在一些实施例中,接头是非共价键,例如通过金属中心(诸如铂原子)的有机金属键。接合可以是永久的或可逆的。对于共价连接,可以使用各种官能团,诸如酰胺基团,包括碳酸衍生物、醚、酯(包括有机酯和无机酯)、氨基、氨基甲酸酯、脲等。为了提供连接,可以通过氧化、羟基化、取代、还原等修饰结构域以提供偶联位点。缀合方法是本领域技术人员众所周知的并且涵盖于本发明中以供使用。接头部分包括但不限于化学接头部分,或例如肽接头部分(接头序列)。应当理解,不显著降低RNA结合结构域和效应结构域的功能的修饰是优选的。

[0177] 突变:如本文所用,术语“突变”具有本领域的普通含义,并且包括例如点突变、取代、插入、缺失、倒置和缺失。

[0178] 寡核苷酸:如本文所用,术语“寡核苷酸”通常是指单链或双链DNA的介于约5与约100个核苷酸之间的多核苷酸。寡核苷酸也称为“寡聚物(oligomer)”或“寡聚物(oligo)”,并且可以从基因中分离出来,或化学合成。

[0179] PAM:术语“PAM”或“前间隔序列邻近基序”是指短核酸序列(长度通常为2-6个碱基对),该序列位于诸如CRISPR-Cas9的CRISPR系统所靶向以进行切割的核酸区域之后。Cas核酸酶切割需要PAM并且通常在切割位点下游3-4个核苷酸处发现PAM。

[0180] 多肽:如本文所用,术语“多肽”是指通过肽键连接在一起的连续氨基酸链。所述术语用于指任何长度的氨基酸链,但本领域普通技术人员将理解所述术语不限于长链并且可以指包括通过肽键连接在一起的两个氨基酸的最小链。如本领域技术人员所知,可以加工和/或修饰多肽。如本文所用,术语“多肽”和“肽”可互换使用。

[0181] 预防:如本文所用,当与疾病、病症和/或病况的发生结合使用时,术语“预防(prevent)”或“预防(prevention)”是指降低发生疾病、病症和/或病况的风险。

[0182] 蛋白质:如本文所用,术语“蛋白质”是指一种或多种作为离散单元发挥作用的多肽。如果单个多肽为离散功能单元且不需要与其它多肽进行永久或暂时的物理缔合以形成离散功能单元,那么术语“多肽”和“蛋白质”可以互换地使用。如果离散功能单位由多于一种物理上彼此缔合的多肽组成,则术语“蛋白质”是指物理偶联并作为离散单位一起发挥作用的多种多肽。

[0183] 参考:“参考”实体、系统、数量、条件集合等是与本文所述的测试实体、系统、数量、条件集合等进行比较的实体、系统、数量、条件集合等。例如,在一些实施例中,“参考”抗体是如本文所述未经工程化的对照抗体。

[0184] RNA向导:术语RNA向导是指促进本文所述的蛋白质靶向靶核酸的RNA分子。示例性“RNA向导”或“向导RNA”包含但不限于crRNA或crRNA与同源tracrRNA的组合。后者可能是独立的RNA或使用连接子(sgRNA)融合为单个RNA。在一些实施例中,RNA向导经工程化以包含

化学或生物化学修饰,在一些实施例中,RNA向导可以包含一个或多个核苷酸。

[0185] 受试者:如本文所用,术语“受试者”意指需要对其进行诊断、预后或治疗的任何受试者。例如,受试者可以是哺乳动物,例如人类或非人类灵长类动物(诸如猿、猴、红毛猩猩或黑猩猩)、狗、猫、豚鼠、兔、大鼠、小鼠、马、牛或奶牛。

[0186] sgRNA:术语“sgRNA”或“单向导RNA”是指含有(i)向导序列(crRNA序列)和(ii) Cas9核酸酶募集序列(tracrRNA)的单向导RNA。

[0187] 基本同一性:短语“基本同一性”在本文中用于指氨基酸序列或核酸序列之间的比较。如本领域普通技术人员将理解的,如果两个序列在对应位置含有相同残基,则它们通常被认为是“基本上相同的”。如本领域所熟知的,可以使用多种算法中的任何一种来比较氨基酸或核酸序列,所述算法包含商业计算机程序中可用的那些,如用于核苷酸序列的BLASTN和用于氨基酸序列的BLASTP、有空位的BLAST和PSI-BLAST。此类例示性程序描述于以下中:Altschul等人,Basic local alignment search tool,J.Mol.Biol.,215(3):403-410,1990;Altschul等人,Methods in Enzymology;Altschul等人,Nucleic Acids Res.25:3389-3402,1997;Baxevanis等人,Bioinformatics:A Practical Guide to the Analysis of Genes and Proteins,Wiley,1998;以及Misener等人(编),Bioinformatics Methods and Protocols(Methods in Molecular Biology,第132卷),Humana Press,1999。除了鉴定相同序列之外,上述程序通常提供对同一性程度的指示。在一些实施例中,如果在相关延伸段的残基上至少50%、55%、60%、65%、70%、75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或更多的对应残基相同,则认为两个序列基本上相同。在一些实施例中,所述相关延伸段是完整序列。在一些实施例中,所述相关延伸段为至少10、15、20、25、30、35、40、45、50、55、60、65、70、75、80、85、90、95、100、125、150、175、200、225、250、275、300、325、350、375、400、425、450、475、500个或更多个残基。

[0188] 靶核酸:如本文所用,术语“靶核酸”是指CRISPR-Cas9系统所结合的任何长度的核苷酸(寡核苷酸或多核苷酸),即脱氧核糖核苷酸、核糖核苷酸或其类似物。靶核酸可以具有三维结构,可以包含编码区或非编码区,可以包含外显子、内含子、mRNA、tRNA、rRNA、siRNA、shRNA、miRNA、核酶、cDNA、质粒、载体、外源序列、内源序列。靶核酸可以包括经修饰的核苷酸,所述经修饰的核苷酸包含甲基化核苷酸或核苷酸类似物。靶核酸可以散布有非核酸组分。靶核酸不限于单链、双链或多链DNA或RNA、基因组DNA、cDNA、DNA-RNA杂交体或包括嘌呤和嘧啶碱基或其它天然的、经化学或生化修饰的、非天然的或衍生的核苷酸碱基的聚合物。

[0189] 治疗有效量:如本文所用,术语“治疗有效量”是指以适用于任何医学治疗的合理益处/风险比赋予经治疗受试者治疗效果的治疗性分子(例如,本文描述的工程化抗体)的量。疗效可以是客观的(即,可通过某一测试或标记测量)或主观的(即,受试者给出效果的指示或感受到效果)。特定说来,“治疗有效量”是指有效治疗、改善或预防特定疾病或病况,或表现出可检测的治疗或预防作用的治疗性分子或组合物的量,诸如通过改善与疾病相关的症状、预防或延迟疾病发作和/或还减轻疾病症状的严重程度或频率。可以用可包括多个单位剂量的给药方案施用治疗有效量。对于任何特定的治疗性分子,治疗有效量(和/或有效给药方案内的适当单位剂量)可以变化,例如,取决于施用途径,以及与其它药剂的组合。此外,针对任何特定受试者的特定治疗有效量(和/或单位剂量)可能取决于多种因素,包括所治疗的病症和病症的严重程度;所采用的特定药剂的活性;所采用的特定组合物;受试者

的年龄、体重、总体健康状况、性别和饮食；施用时间、施用途和/或所采用的特定治疗性分子的排泄率或代谢率；治疗的持续时间；以及医学领域众所周知的类似因素。

[0190] tracrRNA：如本文所用，术语“tracrRNA”或“反式活化crRNA”是指包括形成CRISPR相关蛋白结合特定靶核酸所需结构的序列的RNA。

[0191] 治疗：如本文所用，术语“治疗(treatment)”(还有“治疗(treat)”或“治疗(treating)”)是指治疗性分子(例如，本文所述的CRISPR-Cas治疗性蛋白或系统)的任何施用，该治疗性分子部分或完全减轻、改善、缓解、抑制疾病、病症和/或病况的一或多种症状或特征，延迟疾病、病症和/或病况的一或多种症状或特征的发作，降低疾病、病症和/或病况的一或多种症状或特征的严重程度和/或降低疾病、病症和/或病况的一或多种症状或特征的发病率。此类治疗可以是针对未出现相关疾病、病症和/或病况的病征的受试者和/或针对仅出现疾病、病症和/或病况的早期病征的受试者。或者或另外，这种治疗可以是针对出现相关疾病、病症和/或病况的一种或多种确定病征的个体。

附图说明

[0192] 附图仅用于说明目的；而非用于限制。

[0193] 图1A是显示由人类密码子优化的星座链球菌Cas9识别的共有PAM基序的图。图1B是显示由人类密码子优化的夏普氏菌属分离株RUG017Cas9识别的共有PAM基序的图。图1C是显示由人类密码子优化的小韦荣球菌Cas9识别的共有PAM基序的图。图1D是显示由人类密码子优化的秘鲁埃扎基拉菌识别的共有PAM基序的图。图1E是显示由人类密码子优化的发酵乳杆菌菌株AF15-40LB识别的共有PAM基序的图。图1F是显示由人类密码子优化的嗜脲菌属马赛-P3761识别的共有PAM基序的图。

[0194] 图2A是显示使用Geneious软件针对人类密码子优化的星座链球菌ScoCas9所预测的sgRNA的RNA折叠结构的示意图。图2A描绘了包含SEQ ID NO:3的sgRNA。图2B是显示使用Geneious软件针对人类密码子优化的夏普氏菌属分离株RUG017SirCas9所预测的sgRNA的RNA折叠结构的示意图。图2B描绘了包含SEQ ID NO:7的sgRNA。图2C是显示使用Geneious软件针对人类密码子优化的小韦荣球菌VapCas9所预测的sgRNA的RNA折叠结构的示意图。图2C描绘了包含SEQ ID NO:13的sgRNA。图2D是显示使用Geneious软件针对人类密码子优化的秘鲁埃扎基拉菌EpeCas9所预测的sgRNA的RNA折叠结构的示意图。图2D描绘了包含SEQ ID NO:19的sgRNA。图2E是显示使用Geneious软件针对人类密码子优化的发酵乳杆菌菌株AF15-40LB LfeCas9所预测的sgRNA的RNA折叠结构的示意图。图2E描绘了包含SEQ ID NO:95的sgRNA。图2F是显示使用Geneious软件针对人类密码子优化的嗜脲菌属马赛-P3761 PmaCas9所预测的sgRNA的RNA折叠结构的示意图。图2F描绘了包含SEQ ID NO:96的sgRNA。

[0195] 图3是显示人类密码子优化的ScoCas9在HEK293T细胞中的离体切割活性的例示性结果的图。该图的y轴显示使用各种向导RNA获得的插入缺失频率，这些向导RNA靶向与对应于PAM共有基序的序列相邻的富含A的基因组测试位点(参见图1A)。

[0196] 图4A是显示在N末端处与腺嘌呤碱基编辑器(ABE)或胞嘧啶碱基编辑器(CBE)融合的ScoCas9 D10A突变体的构建体的示意图。图4B是显示插入缺失频率和用包含ABE的碱基编辑器实现的腺嘌呤至鸟嘌呤碱基(A至G)转化百分比的结果的图，该ABE与ScoCas9 D10A突变体的N末端融合。针对靶向富含A的基因组测试位点的各种向导RNA(x轴；表8)绘制A至G

转化百分比(y轴),这些富含A的基因组测试位点与对应于PAM共有基序的序列相邻(参见图1A)。图4C是显示插入缺失频率和用包含ABE的碱基编辑器实现的胞嘧啶至胸腺嘧啶碱基(C至T)转化百分比的结果的图,该ABE与ScoCas9 D10A突变体的N末端融合。针对靶向富含C的基因组测试位点的各种向导RNA(x轴;表8)绘制C至T转化百分比(y轴),这些富含C的基因组测试位点与对应于PAM共有基序的序列相邻(参见图1A)。

[0197] 图5A是显示WT SirCas9和在N末端处与腺嘌呤碱基编辑器(ABE)融合的SirCas9 D14A突变体的构建体的示意图。图5B是显示插入缺失频率和用包含ABE的碱基编辑器实现的A至G转化的结果的图,该ABE与SirCas9 D14A突变体的N末端融合。针对靶向富含A的基因组测试位点的各种向导RNA(x轴;表9)绘制A至G转化百分比(y轴),这些富含A的基因组测试位点与对应于PAM共有基序的序列相邻(参见图1B)。

[0198] 图6A是显示WT VapCas9和在N末端处与腺嘌呤碱基编辑器(ABE)或胞嘧啶碱基编辑器(CBE)融合的VapCas9 D38A突变体的构建体的示意图。图6B是显示插入缺失频率、用包含ABE的碱基编辑器实现的A至G转化和用包含CBE的碱基编辑器实现的C至T转化的结果的图,该ABE与VapCas9 D38A突变体的N末端融合,该CBE与VapCas9 D38A的N末端融合。针对靶向富含A的基因组测试位点的各种向导RNA(x轴;表10)绘制A至G转化百分比(y轴),这些富含A的基因组测试位点与对应于PAM共有基序的序列相邻(参见图1C)。针对靶向富含C的基因组测试位点的各种向导RNA(x轴;表10)绘制C至T转化百分比(y轴),这些富含C的基因组测试位点与对应于PAM共有基序的序列相邻(参见图1C)。

[0199] 图7A是显示与VapCas9的N末端或与VapCas9的C末端融合的ABE的构建体的示意图。图7B是显示用包含与VapCas9的N末端融合的ABE和与VapCas9的C末端融合的ABE的碱基编辑器实现的A至G转化的比较的图。针对靶向富含A的基因组测试位点的各种向导RNA(x轴;表11)绘制A至G转化百分比(y轴),这些富含A的基因组测试位点与对应于PAM共有基序的序列相邻(参见图1C)。

[0200] 图8A是显示WT EpeCas9和在N末端处与ABE和CBE融合的EpeCas9 D38A突变体的构建体的示意图。图8B是显示插入缺失频率、用包含ABE的碱基编辑器实现的A至G转化和用包含CBE的碱基编辑器实现的C至T转化的结果的图,该ABE与EpeCas9 D38A突变体的N末端融合,该CBE与EpeCas9 D38A的N末端融合。针对靶向富含A的基因组测试位点的各种向导RNA(x轴;表12)绘制A至G转化百分比(y轴),这些富含A的基因组测试位点与对应于PAM共有基序的序列相邻(参见图1D)。针对靶向富含C的基因组测试位点的各种向导RNA(x轴;表12)绘制C至T转化百分比(y轴),这些富含C的基因组测试位点与对应于PAM共有基序的序列相邻(参见图1D)。

[0201] 图9A是显示WT LfeCas9和在N末端处与ABE和CBE融合的LfeCas9 D9A突变体的示意图。图9B是显示利用LfeCas9的插入缺失频率的结果的图。图9C是显示用包含ABE的碱基编辑器实现的A至G转化的结果的图,该ABE与LfeCas9 D9A突变体的N末端融合。针对靶向富含A的基因组测试位点的各种向导RNA(x轴;表13)绘制A至G转化百分比(y轴),这些富含A的基因组测试位点与对应于PAM共有基序的序列相邻(参见图1E)。图9D是显示用包含CBE的碱基编辑器实现的C至T转化的结果的图,该CBE与LfeCas9 D9A突变体的N末端融合。针对靶向富含C的基因组测试位点的各种向导RNA(x轴;表13)绘制C至T转化百分比(y轴),这些富含C的基因组测试位点与对应于PAM共有基序的序列相邻(参见图1E)。

[0202] 图10A是显示WT PmaCas9以及在N末端和C末端处与ABE和CBE融合的PmaCas9 D12A突变体的示意图。图10B是显示用包含ABE或CBE的碱基编辑器实现的A至G或C至T转化的结果的图,该ABE或CBE与PmaCas9 D12A突变体的N末端或C末端融合。针对靶向富含A的基因组测试位点的各种向导RNA(x轴;表14)绘制A至G转化百分比(y轴),这些富含A的基因组测试位点与对应于PAM共有基序的序列相邻(参见图1F)。针对靶向富含C的基因组测试位点的各种向导RNA(x轴;表14)绘制C至T转化百分比(y轴),这些富含C的基因组测试位点与对应于PAM共有基序的序列相邻(参见图1F)。

[0203] 图11A是显示通过用两种ScoCas9-NGC变体ScoCas9-NGC-v1和ScoCas9-NGC-v2(x轴)转染细胞测量的插入缺失频率(y轴;插入缺失频率%)的例示性结果的图。还显示了未经转染的细胞对照。

[0204] 图11B是显示用A至G碱基编辑器(ABE)转染的HEK293T细胞中的例示性A至G转化(y轴;A至G转化%)的图,该A至G碱基编辑器包括经工程化以识别NGC PAM基序的ScoCas9-NGC变体、ScoCas9-NGC-v1和ScoCas9-NGC-v2(x轴)。不识别NGC的ScoCas9-NGC变体没有显示出A至G的转化。SpyCas9-NGC对照载体显示出A至G的编辑。还显示了未经转染的细胞对照。

具体实施方式

[0205] 成簇规则间隔短回文重复序列(CRISPR)最初被发现作为细菌和古细菌中的适应性免疫系统,然后可经过工程化以在活细胞和生物体中产生靶向DNA断裂。在细胞DNA修复过程期间,可能会引入各种DNA变化。不同且不断扩展的CRISPR工具箱允许可编程基因组编辑、表观基因组编辑和转录组调控。

[0206] 基于CRISPR-Cas系统的Cas基因组织以及组成蛋白的序列和结构,这些CRISPR-Cas系统包括三种主要类型(I、II和III)。三种CRISPR系统中的每一种以以下独特的Cas基因表征:Cas3,I型中的靶标降解核酸酶/解旋酶;Cas9,II型中的RNA结合和靶标降解核酸酶;Cas10,III型中的用于多种功能的大蛋白。三种CRISPR类型的不同之处还在于其相关联的效应子复合物。I型Cas系统与级联效应子复合物相关联,II型效应子复合物由单个Cas9和一个或多个RNA分子组成,并且III型干扰复合物进一步分为III-A型(靶向DNA的Csm复合物)和III-B型(靶向RNA的Cmr复合物)。Cas蛋白是所有CRISPR-Cas系统中效应子复合物的的重要组成部分。

[0207] 当前的基因组编辑技术集中于II类CRISPR-Cas系统,该系统含有用于DNA切割的单蛋白效应子核酸酶,特别是Cas9(一种双RNA向导的核酸酶,其需要CRISPR RNA(crRNA)和tracrRNA两者并且含有HNH和RuvC核酸酶结构域两者),以及Cas12a(一种单RNA向导的核酸酶,其仅需要crRNA并且含有单个RuvC结构域)。

[0208] 在以下部分中详细描述了本发明的各个方面。部分的使用并不旨在限制本发明。每个部分可以适用于本发明的任何方面。在本申请中,除非另外说明,否则“或者”的使用是指“和/或”。

[0209] 工程化的非天然存在的Cas9蛋白

[0210] 本文描述工程化的非天然存在的Cas9蛋白,其由获自星座链球菌(ScoCas9)、夏普氏菌属分离株RUG017(SirCas9)、小韦荣球菌(VapCas9或VpaCas9,在本文可互换使用)、秘鲁埃扎基拉菌(EpeCas9)、发酵乳杆菌(LfeCas9)和嗜脲菌属马赛-P3761(PmaCas9)细菌的

WT Cas9修饰。

[0211] 在一些实施例中,本文所述的工程化的非天然存在的Cas9蛋白包含与SEQ ID NO: 1、4、8、14、84或86至少60% (例如,60%、65%、70%、75%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或更多) 相同的氨基酸序列。在一些实施例中,Cas9蛋白与SEQ ID NO:1、4、8、14、84或8680%相同。在一些实施例中,Cas9蛋白的氨基酸序列与SEQ ID NO:1、4、8、14、84或86相同。下表1中提供了例示性Cas9氨基酸序列。

[0212] 表1. 例示性Cas9氨基酸序列

[0213]

野生型星座链球菌 Cas9	
	MGKPYSIGLDIGTNSVGWAVVTDDYKVPAAKKMKVLGNTDKQSIKKNLLGALLFDSDG ETAEATRLKRTARRRYTRRKNNRLRYLQEFTGEMNKVDENFFQRLDDSFVDEDKR GEHHPIFGNIAAEVKYHDDFPTIYHLRRHLADTSKKADLRLVYLALAHMIKFRGHFL YEGDLKAENTDVQALFKDFVEEYDKTIEESHLSEITVDALSILTEKVSXSRLLENLIAH YPTEKKNLTFGNLIALSLDLHPNFKTNFQLSEDAKLQFSKDTYEEDLEGFLGEVGEY ADLFASAKNLYDAILLSGILTVDDNSTKAPLSASMVKRYEEHQDKLKKLKDFIKVNA PDQYNAIFKDKNKKGYASYIESGVKQDEFYKYLKGILLKINGSGDFLDKIDREDFLRK QRTFDNGSIPHQIHLQEMHAILRRQGEHYPLKENQDKIEKILTFRIPPYYVGPLARKGS RFAWAEYKADEKITPWNFDDILDKEKSAEKFITRMTLNDLYLPEEKVLPKHSPLYEA FTVYNELTKVKYVNEQGEAKFFDTNMKQEIFDHVFKENRKVTKDKLLNYLNKEFEE FRIVNLTGLDKENKAFNSSLGTYHDLRKILDKSFLDDKANKEKTIEDIIQTLTLFEDREM IRQLQKYSDIFTKAQLKKLERRHYTGWGRLSYKLINGIRNKENKKTILDYLIIDGYA NRNFMQLINDDALSFKEEIARAQIIDDVDDIANVVHDLPGSPAIKKGILQSVKIVDELV KVMGHNPANIIEMARENQTTDKGRRNSQQRLKLLQDSLKNLDNPVNIKNVENQQL QNDRFLFLYYIQNGKDMYTGETLDINNLSQYDIDHIIPQAFIKDNSLDNRVLTRSDKNR GKSDDVPSIEVVHEMKSFWSKLLSVKLITQRKFDNLTAKAERGGLTEEDKAGFIKRQL VETRQITKHVAQILDERFNTEFDGNKRRIRNVKIITLKSNLVSNFRKEFELYKVVREIND YHHAHDAYLNAVVGNALLLKYPQLEPEFVYGEYPKYNSYRSRKSATEKFLFYSNILR

FFKKEDIQTNEDGEIAWNKEKHIKILRKVLSYPQVNIVKKTEEQTGGFSKESILPKGES
DKLIPRKTNSYWDPKKYGGFDSPVVAYSILVFADVEKGKSKKLRLKVQDMVGITIM
EKKRFEKNPVDLFLEQRGYRNVRLLEKIIKLPKYSLFELENKRRRLLASAKELQKGNELV
IPQRFTTLLYHSYRIEKDYEPHREYVEKHKDEFKELLEYISVFSRKVVLADNNLTKE
MLFSKNKDAEVSSLAKSFISLLTFTAFGAPAAFNFFGENIDRKRYTSVTECLNATLIHQ
SITGLYETRIDLKLGED (SEQ ID NO: 1)。

具有核定位信号(NLS)和接头的星座链球菌 Cas9

MPKKKRKVGGKPYSIGLDIGTNSVGWAVVTDDYKVPKMKVLGNTDK
QSIKKNLLGALLFDSGETAEATRLKRTARRRYTRRKNRLRYLQEIFTGEMN
KVDENFFQRLDDSFVDEDKRGEHHPIFGNIAAEVKYHDDFPTIYHLRRHL
ADTSKKADRLVYLALAHMIKFRGHFLYEGDLKAENTDVQALFKDFVEEY
DKTIEESHLSEITVDALSILTEKVS KSRLLENLIAHYPTEKKNTLFGNLIASL
DLHPNFKTNFQLSEDAKLQFSKDTYEEDLEGFLGEVGEYADLFASAKNL
YDAILLSGILTVDDNSTKAPLSASMVKRYEEHQKDLKKLKDFIKVNAPDQY
NAIFKDKNKKGYASYIESGVKQDEFYKYLKGILLKINGSGDFLDKIDREDFL
RKQRTFDNGSIPHQIHLQEMHAILRRQGEHYFLKENQDKIEKILTFRIPIYYV
GPLARKGSRFAWAEYKADEKITPWNFDDILDKEKSAEKFITRMTLNDLYLP
EEKVLPKHSPLYEAFTVYNELTKVKYVNEQGEAKFFDTNMKQEIFDHVFK
ENRKVTKDKLLNYLNKEFEEFRIVNLTGLDKENKAFNSSLGTYHDLRKILD
KSFLDDKANNEKTIEDIIQTLTLFEDREMIRQRLQKYS DIFTKAQLKKLERRH
YTGWGRLSYKLINGIRNKENKKTILDY LIDDGYANRNFMLINDDALS FKE
EIARAQIIDDVDDIANVVHDLPGSPAIKKGILQSVKIVDELVKVMGHNPANII
IEMARENQTTDKGRRNSQQRLKLLQDSLKNLDNPVNIKNVENQQQLQNDRL
FLYYIQNGKDMYTGETLDINNLSQYDIDHIIPQAFIKDNSLDNRVLTRSDKN
RGKSDDVPSIEVVHEMKSFWSKLLSVKLITQRKFDNLTKAERGGLTEEDKA
GFIKRQLVETRQITKHVAQILDERFNTEFDGNKRRIRNVKIITLKS NLVSNFR
KEFELYKVREINDYHHAHDAYLNAVVGNALLLKYPQLEPEFVYGEYPKYN
SYRSRKSATEKFLFYSNILRFFKKEDIQTNEDGEIAWNKEKHIKILRKVLSYP
QVNIVKKTEEQTGGFSKESILPKGESDKLIPRKTNSYWDPKKYGGFDSPV
VAYSILVFADVEKGKSKKLRLKVQDMVGITIMEKKRFEKNPVDLFLEQRGYR
NVRLLEKIIKLPKYSLFELENKRRRLLASAKELQKGNELVIPQRFTTLLYHSY

[0214]

RIEKDYEPHREYVEKHKDEFKELLEYISVFSRKYVLADNNLTKEIEMLF SKN
KDAEVSSLAKSFISLLTFTAFGAPAAFNFFGENIDRKRYTSVTECLNATLIHQ
SITGLYETRIDL SKLGED GKRPAATKKAGQAKKKKGSYPYDVPDYAYPYDV
PDYAYPYDVPDYA (SEQ ID NO: 2)。

野生型夏普氏菌 Cas9

[0215]

MAKNKDIRYSIGLDIGTNSVGWAVMDEHYELLKKGNHHMWGSRLFDAAE
PAATTRASRSIRRRYNKRERIRLLRDLLGDMVMEVDPTFFIRLLNVSFLDE
EDKQKNLGNDYKDNYNLFIEKDFNDKTYDYKYPTIYHLRKELCENKEKAD
PRLIYLALHHIVKYRGNFLKEGQSFAKVYEDIEEKLDNTLKKFMSLNDLDN
LFVDNDINSMITVLSKIYQRSKKADDLLKIMNPTKEERAAYKEFTKALVGL
KFNVSKMILAQEVKKDDKDIELDFSNVDYDSTVDGLQAEUGEYIEFIEM LH
SINSWVELQDILGNNSTISAAMVERYEEHKNDLRVLKKVIREELPKYNEV
FREDNPKLHNYLGYIKYPKNTPVVEEFYIYKRLAKVDTGEAREILERIDLE
KFMLKQNSRTNGSIPYQM QKDEMIQIIDNQSVYYPQLKENREKLISILEFRIP
YYFGPLNTHSEFAWIKKFEDKQKERILPWN YDQIVDIDATAEGFIERMQNT
GTYPDPKPVMAKN SLTVSKFEVLNELNKIRINGKLIPVETKKELLSDLFMK
NKTITDKKLKDWLVTHQYYDTNEELKIEGYQKDLQFSTSLAPWIDFTKIFG
EINASNYQLIEKIIYDISIFEDKKILKRRLKKVYQLDDLLVDKILKLNYTGWS
RLSEKLLTG IKSNSKETILSILENSNMNLMEIINDESLGFKQIIEESNKKDIE
GPFYDEVKKLAGSPA IKRGIWQALLVVQEITKFMKHEPSHIYIEFAREEQE
KVRTESRIAKLQKIYKDLNLQTKEDQLVYESLKKEDAKKKIDTDALYLYYL
QMGKSMYSGKPLDIDKLSTYHIDHILPRSLIKDDSLDNRVLVLPKENEWKL
DSETPFEIRNKMMGF WQKLHENGLMSNKKFFSLIRTDFNEKDKKRFINRQ
LVETRQIIKNVAVIINDHYTNTNVTVRAELSHQFRERYKIYKNRDLNDLH
HAHDAYIACILGQFIHQNFNM DVNMIYGQYKKNYKKDVQEHN NYGFILN
SMNHIHFND DNSVIWDPSYIGKIKSCFCYKDVYVTKKLEQNDAKLFDLTIL
PSDKNSENGVT KAKIPVNKYRKDVN KYGGFSGDAPIMLAIEADKGKKHVR
QVIAFPLRLKNYNDEERIKFIEKEKNLKNVKILTEVKKNQ LILINH QYFFITG
TNELVNATQLKLSAKNTKNL FNLVDANKHNKLESIDDANFNEVIQELICKL
QEPIYSRYNSIGKEFEDSYEKINAVTKQDKLYIIEYLIAMSAKATQGYIKPE
LAREIGTNGKNKGRIKSFTIDLNKTTFISTSVTGLFSKKYKL (SEQ ID NO:

[0216]

4)。
具有核定位信号(NLS)和接头的夏普氏菌 Cas9
MPKKKRKV <u>G</u> AKNKDIRYSIGLDIGTNSVGWAVMDEHYELLKKGNHHMWGSRLFD AAEPAATTRASRSIRRRYNKRERIRLLRDLLGDMVMEVDPTFFIRLLNVSLDEEDK QKNLGNDYKDNYNLFIEKDFNDKTTYDYKYPTIYHLRKELCENKEKADPRLIYLALH HIVKYRGNFLKEGQSFQSAKVYEDIEEKLDNTLKKFMSLNDLDNLFVDNDINSMITVLS KIYQRSKKADDLLKIMNPTKEERAAAYKEFTKALVGLKFNVSKMILAQEVKKDDKDIE LDFSNDYDSTVDGLQAEUGEYIEFIEMLSINSWVELQDILGNNSTISAAMVERYEE HKNDLRVLKKVIREELPDKYNEVFREDNPKLHNYLGYIKYPKNTPVEEFYIYIKRLL AKVDTGEAREILERIDLEKFMLKQNSRTNGSIPYQMOKDEMIQIIDNQSVYYPQLKEN REKLISILEFRIPYYFGPLNTHSEFAWIKKFEDKQKERILPWNVDQIVDIDATAEGFIER MQNTGTYPDPKPVMAKNSLTVSKFEVLNELNKIRINGKLIPVETKKELLSDFMKNK TITDKKLKDWLVTHQYYDTNEELKIEGYQKDLQFSTSLAPWIDFTKIFGEINASNYQL IEKIIYDISIFEDKKILKRRLKKVYQLDDLLVDKILKLNVTGWSRLSEKLLTGKSKNSK ETILSILENSNMNLMEIINDESLGFKQIIEESNKKDIEGPFRYDEVKKLAGSPAIRGIW QALLVVQEITKFMKHEPSHIYIEFAREEQEKVRTESRIAKLQKIYKDLNLQTKEDQLV YESLKKEDAKKKIDTDALYYLQMGKSMYSGKPLDIDKLSTYHIDHILPRSLIKDDS LDNRVLVLPKENEWKLDSETVPFEIRNKMMGFWQKLHENGMSNKKFFSLIRTDNF EKDKKRFINRQLVETRQIIKNVAVIINDHYTNTNVVTVRAELSHQFRERYKIYKNRDL NDLHHAHDAYIACILGQFIHQNFQNMVDVNMIIYGQYKKNYKKDVQEHNNYGFILNS MNHIFNDDNSVIWDPSYIGKIKSCFCYKDVYVTKKLEQNDAKLFDLTILPSDKNSEN GVTKAKIPVNKYRKDVNKYGGFSGDAPIMLAIEADKGKKHVRQVIAFPLRLKNYND EERIKFIEKEKNLKNVKILTEVKKNQLILINHQYFFITGTNELVNATQLKLSAKNTKNL FNLVDANKHNKLESIDDANFNEVIQELICKLQEPISRYNSIGKEFEDSYEKINAVTKQ DKLYIIEYLIAMSAKATQGYIKPELAREIGTNGKNKGRIKSFTIDLNKTTFISTSVTGLF SKKYKL <u>G</u> KRPAATKKAGQAKKKK <u>G</u> <i>SYPDVPDYAYPYDVPDYAYPYDVPDYA</i> (SEQ ID NO: 5)。
野生型小韦荣球菌 Cas9
MSIINFQRRGLMETQASNQLISSHLKGYPIKDYFVGLDIGTSSVGWAVTNK AYELLKFRSHKMWGSRLFDEGESAVARRGFRSMRRRLERRKLRLKLEEL FADAMAQVDPTFFMRLRESKYHYEDKTTGHSSKHILFIDKNYNDQDYFKE YPTVYHLRSELMKSGTDDIRKLFLAVHHILKYRGNFLYEGATFDSNASTLD

[0217]

DVIKQALENITFNCFCDCNSAIISSIGQILMEAGKTKSDKAKAIEHLVDITYIATD
 TVDTSSKTQKDQVKEDKKRLKAFANLVLGLNASLIDLFGSVEELEEDLKKL
 QITGDTYDDKRDELAKAWSDEIYIIDDCKSVYDAIILLSIKEPGLTISESKVK
 AFNKHKDDLAILKSLLKSDRSIYNTMFKVDEKGLHNYVHYIKQGRTEETSC
 NREDFYKYTKKIVEGLSDSKDKEYILSQIELQILLPLQRIKDNGVIPYQLHLE
 ELKAILAKCGPKFPFLNEVADGFSVAEKLKMLEFRIPYYVGPLNTHHNVD
 NGGFAWAVRKASGRVTPWNFDDKIDREKSAAAFIKNLTNKCTYLLGEDVL
 PKSSLLYSEFMLLNELNNVRIDGKPLEKVVEHLIEAVFKQDHHKMTKNRI
 EQFLKDNGYISETHKHEITGLDGEIKNDLASYRDMVRILGDGFDRSMAEEII
 TDITIFGESKKMLRETLRKKFASCLDDEAIKKLTKLRYRDWGRLSQKLLNGI
 EGCDKAGDGTPEIILMRNFSYNLMELLGDKFSFMERIQEINAKLTEGQIV
 NPHDIIDDLALSPAVKRAVWQALRIVDEVAHIKKALPARIFVEVTRSNKNE
 KKKKDSRQKRLSDLYAAIKKDDVLLNGLNNEIFGELKSSLAKYDDAALRS
 KKLYLYYTQMGRCAYTGEIILSLLNTDNYDIDHIYPRSLTKDDSFNVLVC
 KRTANAQKSDAYPISEEIQKTQKPFWTFLKQQGLISERKYERLTRITPLTAD
 DLSGFIARQLVETNQSVKAATTLLRRLYPGVDVVFVKAENVTDVFRHDNNFI
 KVRSLNHHHHAKDAYLNIVVGNVYHERFTRNFRAFFKKNGANRTYNLAK
 MFNYDVNCTNAKDGGKAWDVKTSMDTVKKMMDSNDVRVTKRLLLEQTGA
 LADATIIYKATVAGKAKDGAYIGMKTSSVFADVSKYGGMTKIKNAYSIIV
 QYTGGKGEVIKEIVPLPIYLTNRNTTDQDLINYVASIIPQAKDISIIYGKLCIN
 QLVKVNGFYYYLGGKTNSKFCIDNAIQVIVSNEWIPYLVLEKFNNMRKD
 NKDLKANVVSTRALDNKHTIEVRIVEEKNIEFFDYLVSKLKMPIYQKMKG
 NAAELSEKGYGLFKKMSLEEQSIHLIELLNLLTNQKTTFEVKPLGITASRST
 VGSKISNQDEFKVINESITGLYSNEVTIV (SEQ ID NO: 8)。

具有核定位信号(NLS)和接头的小韦荣球菌 Cas9

MPKKKRKVGSIIINFQRRGLMETQASNQLISSHLKGYPIKDIFYVGLDIGTSS
 VGWAVTNKAYELLKFRSHKMWGSRLFDEGESAVARRGFRSMRRRLERRK
 LRLKLEELFADAMAQVDPTFFMRLRESKYHYEDKTTGHSSKHILFIDKNY
 NDQDYFKEYPTVYHLRSELMKSGTDDIRKLFLAVHHILKYRGNFLYEGATF
 DSNASTLDDVIKQALENITFNCFCDCNSAIISSIGQILMEAGKTKSDKAKAIEHL
 VDTYIATDTVDTSSKTQKDQVKEDKKRLKAFANLVLGLNASLIDLFGSVEE

[0218]

LEEDLKKLQITGDTYDDKRDELAKAWSDEIYIIDDCKSVYDAIILLSIKEPGL
 TISESKVKAFNKHKDDLAILKSLLKSDRSIYNTMFKVDEKGLHNYVHYIKQ
 GRTEETSCNREDFYKYTKKIVEGLSDSKDKEYILSQIELQILLPLQRIKDNGV
 IPYQLHLEELKAILAKCGPKFPFLNEVADGFSVAEKLKMLEFRIPYYVGPL
 NTHHNVDNGGFAWAVRKASGRVTPWNFDDKIDREKSAAAFIKNLTNKCT
 YLLGEDVLPKSSLLYSEFMLLNELNNVRIDGKPLEKVVEHLIEAVFKQDH
 KKMTKNRIEQFLKDNGYISETHKHEITGLDGEIKNDLASYRDMVRILGDGF
 DRSMAREEITDITIFGESKKMLRETLRKKFASCLDDEAIKKLTKLRYRDWGR
 LSQKLLNGIEGCDKAGDGTPEITHILMRNFSYNLMELLGDKFSFMERIQEINA
 KLTEGQIVNPHDIIDDLALSPAVKRAVWQALRIVDEVAHIKKALPARIFVEV
 TRSNKNEKKKKDSRQKRLSDLYAAIKKDDVLLNGLNNEIFGELKSSLAKY
 DDAALRSKKLYLYYTQMGRCAYTGEIIELSLLNTDNYDIDHIYPRSLTKDD
 SFDNLVLCKRTANAQKSDAYPISEEIQKTQKPFWTFLLKQQGLISERKYERLT
 RITPLTADDLSGFIARQLVETNQSVKAATTLLRRLYPGVDVVFVKAENVTD
 FRHDNNFIKVRSLNHHHHAKDAYLNIVVGNVYHERFTRNFRAFFKKNGAN
 RTYNLAKMFNYDVNCTNAKDGAWDVKTSMDTVKKMMDSDNDVRVTKR
 LLEQTGALADATYKATVAGKAKDGAYIGMKTSSVFADVSKYGGMTKIK
 NAYSIIVQYTGKKGEVIKEIVPLPIYLTNRNTTDQDLINYVASIIPQAKDISIY
 GKLCINQLVKVNGFYYYLGGKTNSKFCIDNAIQVIVSNEWIPYLVLEKFN
 NMRKDNKDLKANVVSTRALDNKHTIEVRIVEEKNIEFFDYLVSKLKMPIYQ
 KMKGNKAAELSEKGYGLFKKMSLEEQSIHLIELLNLLTNQKTTFEVKPLGI
 TASRSTVGSKISNQDEFKVINESITGLYSNEVTIVGKRPAATKKAGQAKKK
KGSYPYDVPDYAYPYDVPDYAYPYDVPDYA (SEQ ID NO: 9)。

野生型 秘鲁埃扎基拉菌 Cas9

MTKVKDYYIGLDIGTSSVGWAVTDEAYNVLKFNSKKMWGVRLFDDAKTA
 EERRGQRGARRRLDRKKERLSLLQDFFAEVAKVDPNFFLRDLNDSLIME
 DKDQKLKSKYTLFNDKDFKDNFHKKYPTIHHLLMDLIEDDSKKDIRLVY
 LACHYLLKNRGHFIFEGQKFDTKSSFENSLNELKVHLNDEYGLDLEFDNEN
 LINILTPKLNKTAKKKELKSVIGDTKFLKAVSAIMIGSSQKLVDLFENPEDF
 DDSAIKSVDFSTTSFDDKYSDYELALGDKIALVNILKEIYDSSILENLLKEAD
 KSKDGNKYISNAFVKKYNKHGQDLKEFKRLVRQYHKSAYFDIFRSEKVND

[0219]

NYVSYTKSSISNNKRVKANKFTDQEAFYKFAKKHLETIKYKINKVNGSKA
 DLELIDGMLRDMEFKNFMPKIKSSDNGVIPYQLKLMELNKILENQSKHHEF
 LNVSDEYGSVCDKIASIMEFRIPYYVGPLNPNSKYAWIKKQKDSEITPWNF
 KDVVDLDSREEFIDSLIGRCTYLKDEKVLPKASLLYNEYMVLNELNNLKL
 NDLPIEEMKKKIFDQLFKTRKKVTLKAVANLLKKEFNINGEILLSGTDGDF
 KQGLNSYNDFKAIVGDKVDSDDYRDKIEEIIKLIVLYGDDKSYLQKKIKAG
 YGKYFTDSEIKKMAGLNYKDWGRLSKKLLTGLEGANKITGERGSIIHFMRE
 YNLNLMELMSASFTFTEEIQKLNPVDDRKLSYEMVDELYLSPSVKRMLWQ
 SLRIVDEIKNIMGTDSKKIFIEMARGKEEVKARKESRKNQLLKFKYKDGKKA
 FISEIGEERYSYLLSEIEGEEENKFRWDNLYLYYTQLGRCMYSLEPIDISELS
 SKNIYDQDHIYPKSKIYDDSIENRVLVKKDLNSKKGNSYPIPDEILNKNCYA
 YWKILYDKGLIGQKKYTRLTRRTGFTDDELVQFISRQIVETRQATKETANL
 LKTICKNSEIVYSKAENASRFRQEFDIVKCRAVNDLHHMHDAYINIIVGNV
 YNTKFTKDP MNFVKKQEKARSYNLENMFKYDVKRGGYTAWIADDEKGT
 VKNASIKRIRKELEGTNRYRFRMNYIESGALFNATLQRKNKGSRPLKDKGP
 KSSIEKYGGYTNINKACFAVLDIKSKNKIERKLMPVEREYAKQKNDKKLS
 DEIFSKYLKDRFGIEDYRVVYPVVKMRTLLKIDGSYYFITGGSDKTLELRSA
 LQLILPKKNEWAIKQIDKSSENDYLTIERIQDLTEELVYNTFDIIVNKFKTSV
 FKKSFLNLFQDDKIENIDFKFKSMDFKCKTLLMLVKAIRASGVRQDLKSI
 DLKSDYGRLLSSKTNNIGNYQEFKIINQSITGLFENEVDLLKL (SEQ ID NO:
 14)。

具有核定位信号(NLS)和接头的**秘鲁埃扎基拉菌** Cas9

MPKKKRKVGTKVKDYYIGLDIGTSSVGWAVTDEAYNVLKFNSKKMWGV
 RLFDDAKTAEERRGQRGARRRLDRKKERLSLLQDFFAEVAKVDPNFFLR
 LDNSDLYMEDKDQKLKSKYTLFNDKDFDKDNFHKKYPTIHHLLMDLIEDD
 SKKDIRLVYLACHYLLKNRGHFIFEGQKFDTKSSFENSLNELKVHLNDEYG
 LDLEFDNENLINILTDPKLNKTAKKKELKSVIGDTKFLKAVSAIMIGSSQKL
 VDLFENPEDFDDSAIKSVDFSTTSFDDKYSDYELALGDKIALVNILKEIYDSS
 ILENLLKEADKSKDGNKYISNAFVKKYNKHGQDLKEFKRLVRQYHKSAYF
 DIFRSEKVNDNYVSYTKSSISNNKRVKANKFTDQEAFYKFAKKHLETIKYK
 INKVNGSKADLELIDGMLRDMEFKNFMPKIKSSDNGVIPYQLKLMELNKIL

[0220]

ENQSKHHEFLNVSDEYGSVCDKIASIMEFRIPYYVGPLNPNSKYAWIKKQK
DSEITPWNFKDVVDLDSSREEFIDSLIGRCTYKDEKVLPKASLLYNEYMVL
NELNNLKLNDLPITEEMKKKIFDQLFKTRKKVTLKAVANLLKKEFNINGEIL
LSGTDGDFKQGLNSYNDFKAIVGDKVDSDDYRDKIEEIIKLIVLYGDDKSY
LQKKIKAGYGKYFTDSEIKKMAGLNYKDWGRLSKKLLTGLEGANKITGER
GSIIHFMREYNLNLMEI MSASFTFTEEIQKLN PVDDRKL SYEMVDELYLSPS
VKRMLWQSLRIVDEIKNIMGTDSKKIFIEMARGKEEVKARKESRKNQLLKF
YKDGKKAFISEIGEERYSYLLSEIEGEEENKFRWDNLYLYYTQLGRCMYSL
EPIDISELSSKNIYDQDHIYPKSKIYDDSIENRVLVKKDLNSKKGNSYPIPDEI
LNKNCYAYWKILYDKGLIGQKKYTRLTRRTGFTDDEL VQFISRQIVETRQA
TKETANLLKTICKNSEIVYSKAENASRFRQEFDIVKCRAVNDLHHMHDAYI
NIIVGNVYNTKFTKDP MNFVKKQEKARSYNLENMFKYDVKRGGYTAWIA
DDEKGT VKNASIKRIRKELEGTNYRFRMNYIESGALFNATLQRKNKGSRP
LKDKGPKSSIEKYGGYTNINKACFAVL DIKSKNKIERKLMPVEREIYAKQK
NDKKLSDEIFSKYLKDRFGIEDYRVVYPVVKMRTLLKIDGSYYFITGGSDK
TLELRSALQLILPKKNEWAIKQIDKSSENDYLTIERIQDLTEELVYNTFDIIV
NKFKTSVFKKSFNLNFQDDKIENIDFKFSMD FKEKCKTLLMLVKAIRASG
VRQDLKSIDLKS DYGR LSSKTN NIGNYQEFKIINQSITGLFENEVDLLKL **GK**
RPAATKKAGQAKKKKGSYPYDVPDYAYPYDVPDYAYPYDVPDYA (SEQ ID
NO: 15)。

野生型发酵乳杆菌菌株 AF15-40LB Cas9

MKEYHIGLDIGTSSIGWAVTDSQFKLMRIKGKTAIGVRLFEEGKTAAERRT
FRTRRRRLKRRKWRLHYLDEIFAPHLQEVDENFLRRLKQSNHPEDPAKNQ
AFIGKLLFPDLLKKNERGYPTLIKMRDEL PVEQRAHYPVTNIYKLREAMINE
DRQFDLREVYLAVHHIVKYRGHFLNNASVDKFKVGRIDFDKSFNVLNEAY
EELQNGEGSFTIEPSKVEKIGQLLLDTKMRKLD RQKAVAKLLEVKVADKEE
TKR NKQIATAMSKLV LGYKADFATVAMANGNEWKIDLSSETSEDEIEKFR
EELSDAQNDILTEITSLFSQIMLNEIVPNGMSISESMMDRYW THERQLAEVK
EYLATQPASARKEFDQVYNKYIGQAPKEKGFDLEKGLKKILSKKENWKEI
DELLKAGDFLPKQRTSANGVIPHQMHQQELDR IIEKQAKYYPWLATENPA
TGERDRHQAKYELDQLVSFRIPYYVGPLVTPEVQKATSGAKFAWAKRKED

[0221]

GEITPWNLWDKIDRAESAFAFIKMTVKDITYLLNEDVLPANSLLYQKYNV
 LNELNNVRVNGRRLSVGIKQDIYTELFKKKKTVKAGDVASLVMATRGRV
 NKPSVEGLSDPKKFNSNLATYLDLKSIVGDKVDDNRYQMDLENIIWRSVF
 EDGEIFADKLTEVEWL TDEQRSALVKKRYKGWGRLSKKLLTGIVDENGQR
 IIDLMWNTDQNFMQIVNQPVFKEQIDQLNQKAITNDGMTLRERVESVLDD
 AYTSPQNKKAIWQVVRVVEDIVKAVGNAPKSSISIEFARNEGNGEITRSRR
 TQLQKLFEDQAHEL VKDTSLTEELEKAPDLSDRYFYFTQGGKDMYT GDP
 INFDEISTKYDIDHILPQS FVKDDSLDNRLVVSRAENNKSDRVPKLYAAK
 MKPYWNQLLKQGLITQRKFENLTMDVDQTIKYRSLGFVKRQLVETRQVIK
 LTANILGSMYQEAGTDIIETRAGLTKQLREEFDLPKVREVNDYHHAVDAYL
 TTFAGQYLNRRYPKLRSFFVYGEYMKFKHGS DLKLRNFNFHHELMEGDKS
 QGKVVDQQTGELITRDEVADYFDWVINLKVMLISNETYEETGKYFDASH
 ESSSLYLKNQNKSKSLVPLKNKLQPEYYGAYTGITQGYMVILKLLDKKG
 GFGVYRIPRYAADILNKCHDEVAYRNKIAEIISSDPRAPKSFEVVVPRVLKG
 TFLVDGEEKFILSSYRYKV NATQLILPVSDIKLIQDNFKALKKLNVMQTKK
 LIEIYDNILRQVDKYYKLYDINKFRAKLHDGRSKFVELDDFGQDASKEKVII
 KILRGLHFGSDLQNLKEIGFGTTPLGQFQVSEAGIRLSNTAFIIFKSPTGLFNR
 KLYLKNL (SEQ ID NO: 84)。

具有核定位信号(NLS)和接头的发酵乳杆菌菌株 AF15-40LB Cas9

MPKKKRKVVGKEYHIGLDIGTSSIGWAVTDSQFKLMRIKGKTAIGVRLFEE
 GKTA AERRTFRTTRRLKRRKWRLHYLDEIFAPHLQEVDENFLRRLKQSN
 HPEDPAKNQAFIGKLLFPDLLKKNERGYPTLIKMRDELPVEQRAHYPVTNI
 YKLREAMINEDRQFDLREVYLAVHHIVKYRGHFLNNASVDKFKVGRIDFD
 KSFNVLNEAYEELQNGEGSFTIEPSKVEKIGQLLLDTKMRKLDRQKAVAKL
 LEVKVADKEETKRNKQIATAMSKLV LGYKADFATVAMANGNEWKIDLSS
 ETSEDEIEKFREELSDAQNDILTEITSLFSQIMLNEIVPNGMSISESMMDRYW
 THERQLAEVKEYLATQPASARKEFDQVYNKYIGQAPKEKGF DLEKGLKKI
 LSKKENWKEIDELLKAGDFLPKQRTSANGVIPHQMHQQELDRIIEKQAKYY
 PWLATENPATGERDRHQAKYELDQLVSFRIPYYVGPLVTPEVQKATSGAK
 FAWAKRKEDGEITPWNLWDKIDRAESAFAFIKMTVKDITYLLNEDVLPAN
 SLLYQKYNVLNELNNVRVNGRRLSVGIKQDIYTELFKKKKTVKAGDVASL

[0222]

VMAKTRGVNKPSVEGLSDPKKFNSNLATYLDLKSIVGDKVDDNRYQMDL
 ENIIEWRSVFEDGEIFADKLTEVEWLTDEQRSALVKKRYKGWGRLSKKLLT
 GIVDENGQRIIDLMWNTDQNFMQIVNQPVFKEQIDQLNQKAITNDGMTLR
 ERVESVLDDAYTSPQNKKAIWQVVRVVEDIVKAVGNAPKSISIEFARNEGN
 KGEITRSRRTQLQKLFEDQAHELVKDTSLTEELEKAPDLSDRYFYFTQGG
 KDMYTGDPIINFDEISTKYDIDHILPQSFVKDDSLDNRVLVSRAENNNKSDR
 VPAKLYAAKMMPYWNQLLKQGLITQRKFENLTMDVDQTIKYRSLGFVKR
 QLVETRQVIKLTANILGSMYQEAGTDIIETRAGLTKQLREEFDLPKVREVND
 YHHAVDAYLTTFAGQYLNRRYPKLRSSFVYGEYMKFKHGSCLKLRNFNFF
 HELMEGDKSQGKVVDQQTGELITTRDEVADYFDWVINLKVMLISNETYEE
 TGKYFDASHESSELYLKNQNKSKLVVPLKNKLQPEYYGAYTGITQGYMV
 ILKLLDKKGGFGVYRIPRYAADILNKCHDEVAYRNKIAEIISSDPRAPKSFEV
 VVPRVLKGTFLVDGEEKFILSSYRYKVNATQLILPVSDIKLIQDNFKALKKL
 NVEMQTKKLEIYDNILRQVDKYYKLYDINKFRAKLHDGRSKFVELDDFG
 QDASKEKVIKILRGLHFGSDLQNLKEIGFGTTPLGQFQVSEAGIRLSNTAFII
 FKSPTGLFNRKLYLKNL

GKRPAATKKAGQAKKKKGSYPYDVPDYAYPYDVPDYAYPYDVPDYA (SEQ
 ID NO: 85)。

野生型嗜脉菌属马赛-P3761 Cas9

MEKKTNYTIGLDIGTDSVGWAVVKDDLELVKKRMKVLGNTETNYIKKNL
 WGSLLFESGQTAKDRRLKRVARRRYERRRNRLTELQKIFAPAIDEVDENFF
 FRLNESFLVPEDKAFSKNPIFGTLGEDKTYKTYPTIYHLRQHLADSEEKAD
 VRLIYLALAHMIKYRGHFLIEGKLDTEHIAINENLEQFFESYNALFSEPIEL
 RKEELIAIENILREKNSRTVKEKRITSFLKDIGRANKQSPMMAFITLIVGKKA
 KFKA AFNLEEEISLNLTDSDYDENLEILLNTIGSDFADLFDHAQRVYNAVEL
 AGILSGDVKNTHAKLSAQMVAMYERHKEQLKEYKSFIKANLPDQYDMTF
 VAPKDAQKKDLKGYAGYIDGNMSQDSFYKFVKDQLKEVPGSEKFLDSIEK
 EDFLRKQRSFYNGVIPNQVHLAEMEAILDRQENYYPWLKENREKIISLLTFR
 IPYYVGPLADGQSEFAWLERKSDEKIKPWNFSDVVDLDRSAEKFIEQLIGR
 DTYLPDEYVLPKKS LIYQKYMVFNELTKIAYLDERQKRMNLSSVEKKEIFE
 TLFKKRSKVTEKQLVKFFENYLQIDNPTIFGIEDAFNADYSTYVELAKVPG

MKSMMDDPDNEDLMEEIVKILTVFEDRKMRRKQLEKYKERLSPEQIKELA
KKHYTGWGRLSKKLLVGIRDKETQKTILDYLVEDDNHSGGRQHLNRNLM
QLINDDRLSFKKTIAELQMIDPSADLYAQVQEIAGSPAIIKKGILLGLKIVDEII
RVMGEKPENIVIEMARENQTTARGKALSKRREAKIKEGLAALGSSLLKENL
PGNADLSQRKIYLYYTQNGKDIYLDDEPLDFDRLSQYDEDEHIIPQSFTVDNSL
DNLVLTNSSQNRGNKKDDVPSLEVVRQLAYWRSCLKDAGLMTQRKFDNL
TKAMRGGLTDKDRERFIQRQLVETRQITKNVAKLLDMRLNDKKDEAGNKI
RETNIVLLKSAMASEFRKMFRLYKVRELNDYHHAHDAYLNAAIAINLLAL
YPYMADDFVYGEFRYKKKPKQAEKATYEKLRQWNLIKRFGEKQLFTPDHE
DCWNERDIKTIKKVMGYRQVNVVKKAEERTGMLFKETINGKTNKGSRI
IKKDLDPISKYGGYIEEKMAYYAVISYEDKKKKPGKTIVGISIMDKKEFEYD
SISYLGKLGFSNPVVQIILKNYSLIAYPDGRRRYITGATKTTKGKVELQKAN
QIAMEQDLVNFYHLKNYDEISHPESYAFVQSHTDYFDRLFDSIEHYTRRFL
DAETNINRLRRIYEEKKKDPVDIEALVASFIELLKLTSAGAPADFIFMGEAI
SRRRYNSMTGLFDGQVIYQSLTGLYETRMRFED (SEQ ID NO: 86)。

[0223]

具有核定位信号(NLS)和接头的嗜热菌属马赛-P3761 Cas9

MPKKKRKVGEKKTNYTIGLDIGTDSVGWAVVKDDLELVKKRMKVLGNT
ETNYIKKNLWGSLLFESGQTAKDRRLKRVARRRYERRRNRLTELQKIFAPA
IDEVDENFFFRLNESFLVPEDKAFSKNPIFGTLGEDKTYKYKTYPTIYHLRQH
LADSEEKADVRLIYLALAHMIKYRGHFLIEGKLDTEHIAINENLEQFFESYN
ALFSEEPIELRKEELIAIENILREKNSRTVKEKRITSFLKDIGRANKQSPMMA
FITLIVGKKAKFKAAFNLEEEISLNLTDSDYDENLEILLNTIGSDFADLFDHA
QRVYNAVELAGILSGDVKNTHAKLSAQMVAMYERHKEQLKEYKSFIKAN
LPDQYDMTFVAPKDAQKKDLKGYAGYIDGNMSQDSFYKFVKDQLKEVPG
SEKFLDSIEKEDFLRKQRSFYNGVIPNQVHLAEMEAILDRQENYYPWLKEN
REKIISLLTFRIPYYVGPLADGQSEFAWLERKSDEKIKPWNFSDVVDLDRSA
EKFIEQLIGRDTYLPDEYVLPKKSLEYQKYMVFNELTKIAYLDERQKRMNL
SSVEKKEIFETLFKKRSKVTEKQLVKFFENYLQIDNPTIFGIEDAFNADYSTY
VELAKVPGMKSMDDPDNEDLMEEIVKILTVFEDRKMRRKQLEKYKERL
SPEQIKELAKKHYTGWGRLSKKLLVGIRDKETQKTILDYLVEDDNHSGGR
QHLNRNLMQLINDDRLSFKKTIAELQMIDPSADLYAQVQEIAGSPAIIKKGIL

[0224]

LGLKIVDEIIRVMGEKPENIVIAMARENQTTARGKALSKRREAKIKEGLAAL
 GSSLLKENLPGNADLSQRKIYLYYTQNGKDIYLDDEPLDFDRLS QYDEDHIIP
 QSFTVDNSLDNLVLTNSSQNRRGNKKDDVPSLEVVRQLAYWRS LKDAGL
 MTQRKFDNLTKAMRGGLTDKDRERFIQRQLVETRQITKNVAKLLDMRLN
 DKKDEAGNKIRETNIVLLKSAMASEFRKMFRLYKVRELNDYHHAHDAYL
 NAAIAINLLALYPYMADDFVYGEFRYKKKPKQAEKATYEKLRQWNLIKRFG
 EKQLFTP DHEDCWNKERDIKTIKKVMGYRQVNVVKKAEERTGMLFKETIN
 GKTNKGSRIPKIKDLDP SKYGGYIEEK MAYYAVISYEDKKKKPGKTIVGISI
 MDKKEFEYDSISYLGKLGFSNPVVQIILKNYSLIAYPDGRRRRYITGATKTTK
 GKVELQKANQIAMEQDLVNFIYHLKNYDEISHPESYAFVQSHTDYFDRLFD
 SIEHYTRRFLDAETNINRLRRIYEEEEKKDPVDIEALVASFIELLLKLT SAGAP
 ADFIFMGEAISR RRRYNSMTGLFDGQVIYQSLTGLYETRMRFEDKKRPAAT
KKAGQAKKKKGSYPYDVPDYAYPYDVPDYAYPYDVPDYA (SEQ ID NO: 87)。

[0225] NLS(粗体), 可以用不同的NLS取代

[0226] 接头(带下划线), 可以去除或扩增

[0227] 3xHA标签(斜体), 可以用不同的标签取代

[0228] 在一些实施例中, Cas9蛋白包含关于SEQ ID NO: 1、4、8、14、84或86的一个或多个突变。例如, Cas9蛋白的氨基酸序列包含SEQ ID NO: 1、4、8、14、84或86中的至少一个、至少两个、至少三个、至少四个、至少五个、至少六个、至少七个、至少八个、至少九个、至少十个突变。各种突变是本领域已知的并且包括例如氨基酸取代。

[0229] 在一些实施例中, Cas9的两个或更多个催化结构域(RuvC1、RuvCII、RuvCIII)经突变以产生缺乏核酸切割活性的无活性或“死”Cas9(dCas9)。在一些实施例中, 一个或多个突变位于PAM相互作用结构域、HNH结构域和/或RuvC结构域中。在一些实施例中, Cas9经突变以将DNA切割活性降低至相对于其非突变形式小于约25%、15%、10%、5%、1%、0.1%、0.01%或更低。

[0230] 在一些实施例中, 提供了Cas9的切口酶突变体型式。在一些实施例中, 切口酶突变体在RuvC结构域和/或HNH结构域中具有一个或多个氨基酸取代。已知关于SpCas9(酿脓链球菌)的多个切口酶突变并且包括例如野生型SpCas9的氨基酸位置10、12、17、762、840、854、863、982、983、984、986、987中的一个或多个处的突变。例如, 对应于SpCas9中的D10A的天冬氨酸至丙氨酸的取代导致切口酶的产生。在一些实施例中, 本文所述的Cas9具有导致切口酶产生的一个或多个突变。在一些实施例中, 本文所述的Cas9在对应于SpCas9的氨基酸10、12、17、762、840、854、863、982、983、984、986、987中的一者或多者的氨基酸位置处具有一个或多个突变。

[0231] 在一些实施例中, 突变是ScoCas9的RuvC结构域中的天冬氨酸至丙氨酸的取代(D10A)。在一些实施例中, 突变是SirCas9的RuvC结构域中的天冬氨酸至丙氨酸的取代(D14A)。在一些实施例中, 突变是VapCas9的RuvC结构域中的天冬氨酸至丙氨酸的取代(D38A)(例如, 对应于SpCas9中的D10A)。在一些实施例中, 突变是EpeCas9的RuvC结构域中

的天冬氨酸至丙氨酸的取代 (D12A)。在一些实施例中,突变是LfeCas9的RuvC结构域中的天冬氨酸至丙氨酸的取代 (D9A)。在一些实施例中,突变是PmaCas9的RuvC结构域中的天冬氨酸至丙氨酸的取代 (D12A)。

[0232] 在一些实施例中,突变是ScoCas9的RuvC结构域中的天冬氨酸至甘氨酸的取代 (D10G)。在一些实施例中,突变是SirCas9的RuvC结构域中的天冬氨酸至甘氨酸的取代 (D14G)。在一些实施例中,突变是VapCas9的RuvC结构域中的天冬氨酸至甘氨酸的取代 (D38G) (例如,对应于SpCas9中的D10G)。在一些实施例中,突变是EpeCas9的RuvC结构域中的天冬氨酸至甘氨酸的取代 (D12G)。在一些实施例中,突变是LfeCas9的RuvC结构域中的天冬氨酸至甘氨酸的取代 (D9A)。在一些实施例中,突变是PmaCas9的RuvC结构域中的天冬氨酸至甘氨酸的取代 (D12G)。

[0233] 在一些实施例中,本文所述的此类一个或多个突变将Cas9转化为Cas9的无活性或“死”型式 (dCas9)。因此,在一些实施例中,Cas9蛋白包含抑制Cas9切割DNA双链体的两条链的能力的一个或多个突变。

[0234] 在一些实施例中,当与向导RNA共表达时,死Cas9产生DNA识别复合物,其可以特异性地干扰转录延伸、RNA聚合酶结合或转录因子结合。在一些实施例中,死Cas9用于将各种功能的效应蛋白特异性靶向特异性核酸靶位点。

[0235] 在一些实施例中,高保真Cas9变体包括增强的特异性,其使脱靶切割最小化。在一些实施例中,使用工程化变体,例如‘超精确Cas9’ (对应于SpyCas9的N692A、M694A、Q695A和/或H698A突变) 和/或‘高保真Cas9’ (对应于SpyCas9的N467A、R661A、Q695A和/或Q926A突变),其包含主要在REC3结构域内的突变并实现更高的特异性和保真度。高保真变体会降低Cas9稳定错配和减少脱靶DNA切割的能力。在一些实施例中,特异性的增加伴随着在靶切割效率约100倍的损失。在一些实施例中,使用SuperFi-Cas9,其是保持与野生型Cas9相当的在靶切割率的高保真变体。在一些实施例中,SuperFi-Cas9包含RuvC环中的突变。在一些实施例中,突变抑制扭结构象的形成,该扭结构象有利于gRNA-TS双链体的后续切割。在一些实施例中,对应于SpyCas9的Y1016、R1019、Y1010、Y1013、K1031、Q1027和/或V1018残基经突变为例如天冬氨酸。(Bravo, J. 等人 Structural basis for mismatch surveillance by CRISPR-Cas9 Nature, 603, 2022年3月)。

[0236] 工程化的非天然存在的Cas9具有与SEQ ID NO. 2、5、9、15、85、87、95或96的Cas9氨基酸序列至少80% (例如,80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或更高) 相同的氨基酸序列。

[0237] 在一些实施例中,工程化的非天然存在的Cas9在针对人类细胞进行密码子优化 (例如,针对表达、稳定性等进行密码子优化) 的核酸分子中编码。

[0238] 下表2中提供了具有核定位信号 (NLS) 和接头的例示性Cas9序列。

[0239] 表2. 具有NLS和接头的例示性Cas9序列

[0240]

具有核定位信号(NLS)和接头的 *ScoCas9* 序列

MPKKRKRVGGKPYSIGLDIGTNSVGWAVVTDDYKVPAAKMKVLGNTDKQSIKKN
LLGALLFDSGETAEATRLKRTARRRYTRRKNRLRYLQEIFTGEMNKVDENFFQRLD
DSFLVDEDEKRGEHHPFGNIAAEVKYHDDFPTIYHLRRHLADTSKKADLRLVYLAL
AHMIKFRGHFLYEGDLKAENTDVQALFKDFVEEYDKTIEESHLSEITVDALSILTEK
VSKSSRLLENLIAHYPTEKKNTLFGNLIASLDLHPNFKTNFQLSEDAKLQFSKDTYE
EDLEGFLGEVGDEYADLFASAKNLYDAILLSGILTVDDNSTKAPLSASMVKRYEEH
QKDLKKLKDFIKVNAPDQYNAIFKDKNKKGYASYIESGVKQDEFYKYLKGILLKIN
GSGDFLDKIDREDFLRKQRTFDNGSIPHQIHLQEMHAILRRQGEHYPFLKENQDKIE
KILTRIPYYVGPLARKGSRFAWAEYKADEKITPWNFDDILDKEKSAEKFITRMTLN
DLYLPPEEKVLPKHSPLYEAFTVYNELTKVKYVNEQGEAKFFDTNMKQEIFDHVFKE
NRKVTKDKLLNYLNKEFEEFRIVNLTGLDKENKAFNSSLGTYHDLRKILDKSFLDD
KANEKTIEDIIQTLTLFEDREMIRQLQKYSDIFTKAQLKKLERRHYTGWGRLSYKL
INGIRNKENKKTILDYLIDDGYANRNFMLINDDALSFKEEIARAQIIDDVDDIANV
VHDLPGSPAIAKKGILQSVKIVDELVKVMGHNPANIIEMARENQTTDKGRRNSQQRL
KLLQDSLKNLDNPVNIKNVENQQLQNDRLFLYYIQNGKDMYTGETLDINNLSQYDI
DHIIPQAFIKDNSLDNRVLTRSDKNRGKSDDVPSIEVVHEMKSFWSKLLSVKLITQR
KFDNLTKAERGGLTEEDKAGFIKRQLVETRQITKHVAQILDERFNTEFDGNKRRIRN
VKIITLKSNLVSNFRKEFELYKVREINDYHHAHDAYLNAVVGNAALLLKYPQLEPEFV
YGEYPKYNSYRSRKSATEKFLFYSNILRFFKKEDIQTNEDGEIAWNKEKHILRKV
LSYPQVNIVKKTEEQTGGFSKESILPKGESDKLIPRKTKN SYWDPKKYGGFDSPVVA
YSILVFADVEKGKSKKLARKVQDMVGITIMEKKRFEKNPVDFLEQRGYRNVRLKII

[0241]

KLPKYSLFELENKRRRLASAKELQKGNELVIPQRFTTLLYHSYRIEKDYEPHEHREY
VEKHKDEFKELLEYISVFSRKYVLADNNLTKEIEMLFSSKNKDAEVSSLAKSFISLLTFT
AFGAPAAFNFFGENIDRKRYTSVTECLNATLIHQSIITGLYETRIDLKSLGEDGKRPAAT
TKKAGQAKKKKGSYPDVDPDYAYPYDVPDYAYPYDVPDYA (SEQ ID NO: 2)。

具有核定位信号(NLS)和接头的夏普氏菌 Cas9 序列

MPKKKRKVGAKNKDIRYSIGLDIGTNSVGWAVMDEHYELLKKGNHHMWGSRLF
DAAEPAATTRASRSIRRRYNKRERIRLLRDLLGDMVMEVDPTFFIRLLNVSLDEE
DKQKNLGNDYKDNYNLFIEKDFNDKTYDYKYPTIYHLRKELCENKEKADPRLIYL
ALHHIVKYRGNFLKEGQSFAKVYEDIEEKLDNTLKKFMSLNDLDNLFVDNDINSMI
TVLSKIYQSKKADDLLKIMNPTKEERAAYKEFTKALVGLKFNVSKMILAQEVKKD
DKDIELDFSNDYDSTVDGLQAELGEYIEFIEMLSINSWVELQDILGNNSTISAAM
VERYEEHKNDLRVLKKVIREELPKYNEVFREDNPKLHNYLGYIKYPKNTPVVEEFY
EYIKRLLAKVDTGEAREILERIDLEKFMLKQNSRTNGSIPYQMOKDEMIQIIDNQSV
YYPQLKENREKLISILEFRIPYYFGPLNTHSEFAWIKKFEDKQKERILPWNVDQIVDI
DATAEGFIERMQNTGTYPDKPVMAKNSLTVSKFEVLNELNKIRINGKLIPVETKKE
LLSDFLMKNKTITDKKLKDWLVTHQYYDTNEELKIEGYQKDLQFSTSLAPWIDFTK
IFGEINASNYQLIEKIIYDISIFEDKKILKRRLLKKVYQLDDLLVDKILKLNYTGWSRLS
EKLLTGIKSKNSKETILSILENSNMNLMEIINDESLGFKQIIEESNKKDIEGPFYDEV
KKLAGSPAIRGIWQALLVVQEITKFMKHEPSHIYIEFAREEQEKVRESRIAKLQKI
YKDLNLQTKEDQLVYESLKKEDAKKKIDTDALYLYYLQMGKSMYSGKPLDIDKLS
TYHIDHILPRSLIKDDSLDNRLVLPKENEWKLDSETPFEIRNKMMGFQKLHEN
GLMSNKKFFSLIRTDNFNEKDKKRFINRQLVETRQIIKNVAVIINDHYTNTNVTVRAE
LSHQFRERYKIYKNRDLNDLHHAHDAYIACILGQFIHQNFNMDVNMIIYGQYKKN
YKKDVQEHNNYGFILNSMNHIFNDNSVIWDPSYIGKIKSCFCYKDVYVTKKLE
QNDAKLFDLTILPSDKNSENGVTAKIPVNKYRKDVNKGFGSGDAPIMLAIEADK
GKKHVRQVIAFPLRLKNYNDEERIKFIEKEKNLKNVKILTEVKKNQLILINHQYFFIT
GTNELVNATQLKLSAKNTKNLNLVDANKHNKLESIDDANFNEVIQELICKLQEPY
SRYNSIGKEFEDSYEKINAVTKQDKLYIIIEYLIAIMSAKATQGYIKPELAREIGTNGK
NKGRIKSFTIDLNKTTFISTSVTGLFSKKYKLGKRPAATTKKAGQAKKKKGSYPDV
PDYAYPYDVPDYAYPYDVPDYA (SEQ ID NO: 5)。

具有核定位信号(NLS)和接头的小韦荣球菌 Cas9 序列

MPKKKRKVGSIINFQRRGLMETQASNQLISSHLKGYPIKDYFVGLDIGTSSVGWAV
TNKAYELLKFRSHKMWGSRLFDEGESAVARRGFRSMRRRLERRKLRLKLEELFAD
AMAQVDPTFFMRLRESKYHYEDKTTGHSSKHILFIDKNYNDQDYFKEYPTVYHLR

[0242]

SELMKSGTDDIRKLFLAVHHILKYRGNFLYEGATFDSNASTLDDVIKQALENITFNC
 FDCNSAISSIGQILMEAGKTKSDKAKAIEHLVDTYIATDTVDTSSKTQKDQVKEDKK
 RLKAFANLVLGLNASLIDLFGSVEELEEDLKKLQITGDTYDDKRDELAKAWSDEIYI
 IDDCSVYDAIILLSIKEPGLTISESKVKAFNKHKDDLAILKSLKSDRSIYNTMFKV
 DEKGLHNYVHYIKQGRTEETSCNREDFYKYTKKIVEGLSDSKDKEYILSQIELQILL
 PLQRIKDNGVIPYQLHLEELKAILAKCGPKFPFLNEVADGFSVAEKLKMLEFRIPYY
 VGPLNTHHNVDNGGFAWVRKASGRVTPWNFDDKIDREKSAAAFIKNLTNKCTYL
 LGEDVLPKSSLLYSEFMLLNELNNVRIDGKPLEKVVEHLIEAVFKQDHKKMTKNR
 IEQFLKDNGYISETHKHEITGLDGEIKNDLASYRDMVRILGDGFDRSMAEEIITDITIF
 GESKKMLRETLRKKFASCLDDEAIKKLTCLRYRDWGRLSQKLLNGIEGCDKAGDG
 TPETIIILMRNFSYNLMELLGDKFSFMERIQEINAKLTEGQIVNPHDIIDDLALSPAVK
 RAVWQALRIVDEVAHIKKALPARIFVEVTRSNNKNEKKKKDSRQKRLSDLYAAIKKD
 DVLLNGLNNEIFGELKSSLAKEYDDAALRSKKLYLYYTQMGRCAYTGEIHELSSLNT
 DNYDIDHIYPRSLTKDDSFNVLCKRTANAQKSDAYPISEEIQKTQKPFWTFLKQQ
 GLISERKYERLTRITPLTADDLSGFIARQLVETNQSVKAATTLLRRLYPGVDVVFVKA
 ENVTDFRHDNNFIKVRSLNHHHHAKDAYLNIVVGNVYHERFTRNFRAFFKKNGAN
 RTYNLAKMFNYDVNCTNAKD GKAWDVKTSMDTVKKMMDSNDVRVTKRLLEQT
 GALADATIYKATVAGKAKDGAYIGMKTSSVFADVSKYGGMTKIKNAYSIIVQYTG
 KKGEVIKEIVPLPIYL TNRN TTDQDLIN YVASIIPQAKDISIIYGKLCINQLVKVNGFY
 YYLGGKTNSKFCIDNAIQVIVSNEWIPYLKVLEKFNNMRKDNKDLKANVVSTRAL
 DNKHTIEVRIVEEKNIEFFDYLVSKLKMPIYQKMKGNKAAELSEKGYGLFKKMSLE
 EQSIHLIELLNLLTNQKTTFEVKPLGITASRSTVGSKISNQDEFKVINESITGLYSNEV
 TIVGKRPAAATKKAGQAKKKKGSYPDVDPDYAYPYDVPDYAYPYDVPDYA (SEQ ID
 NO: 9)。

具有核定位信号(NLS)和接头的秘鲁埃扎基拉菌 Cas9 序列

MPKKKRKVGTKVKDYIIGLDIGTSSVGWAVTDEAYNVLFNSKKMWGVRLFDD
 AKTAEERRGQRGARRRLDRKKERLSLLQDFFAEVAKVDPNFFLRDLNSDLYMED
 KDQKLKSKYTLFNDKDFDKNFHKKYPTIHLLMDLIEDDSKKDIRLVYLACHYL
 LKNRGHFIFEGQKFDTKSSFENSLNELKVHLNDEYGLDLEFDNENLINILTDPKLNK
 TAKKKELKSVIGDTKFLKAVSAIMIGSSQKLVDLFENPEDFDDSAIKSVDFSTTSFDD
 KYSDYELALGDKIALVNILKEIYDSSILENLLKEADKSKDGKNKYISNAFVKKYNKH
 GQDLKEFKRLVRQYHKSAYFDIFRSEK VNDNYVS YTKSSISNNKRVKANKFTDQEA
 FYKFAKKHLETIKYKINKVNGSKADLELIDGMLRDMEFKNFMPKIKSSDNGVIPYQ
 LKLMELNKILENQSKHHEFLNVSDEYGSVCDKIASIMEFRIPYYVGPLNPNSKYAWI

[0243]

KKQKDSEITPWNFKDVVDLDSSREEFIDSLIGRCTYKDEKVLPAKSLLYNEYMVL
NELNNLKLNDLPITEEMKKKIFDQLFKTRKKVTLKAVANLLKKEFNINGEILLSGTD
GDFKQGLNSYNDFKAIVGDKVDSDDYRDKIEEIIKLIVLYGDDKSYLQKKIKAGYG
KYFTDSEIKKMAGLNYKDWGRLSKKLLTGLEGANKITGERGSIIHFMREYNLNLME
LMSASFTFTEEIQKLNVPDDRKLSYEMVDELYLSPSVKRMLWQSLRIVDEIKNIMGT
DSKKIFIEMARGKEEVKARKESRKNQLLKFKYKDGKKAFISEIGEERYSYLLSEIEGE
EENKFRWDNLYLYYTQLGRCMYSLEPIDISELSSKNIYDQDHIYPKSKIYDDSIENRV
LVKKDLNSKKGNSYPIPDEILNKNCYAYWKILYDKGLIGQKKYTRLTRRTGFTDDEL
VQFISRQIVETRQATKETANLLKTICKNSEIVYSKAENASRFRQEFDIVKCRAVNDLH
HMHDAYINIIVGNVYNTKFTKDP MNFVKKQEKARSYNLENMFKYDVKRGGYTAW
IADDEKGTVKNASIKRIRKELEGTNYRFRMNYIESGALFNATLQRKNKGSRPLKD
KGPKSIEKYGGYTNINKACFAVLDIKSKNKIERKLMVEREIIYAKQKNDKKLSDEI
FSKYLKDRFGIEDYRVVYPVVKMRTLLKIDGSYYFITGGSDKTLELRSALQLILPKK
NEWAIKQIDKSSENDYLTIERIQDLTEELVYNTFDIIVNKFKTSVFKKSFNLNFQDDKI
ENIDFKFKSMDFKCKTLLMLVKAIRASGVRQDLKSIDLKSDYGRLLSSKTNNIGN
YQEFKIINQSITGLFENEVDLLKLGKRPAATKKAGQAKKKKGSYPYDVPDYAYPYD
VPDYAYPYDVPDYA (SEQ ID NO: 15)。

具有核定位信号(NLS)和接头的发酵乳杆菌菌株 AF15-40LB Cas9 序列

MPKKKRKVGKEYHIGLDIGTSSIGWAVTDSQFKLMRIKGKTAIGVRLFEEGKTAAE
RRTFRTTRRLKRRKWRLHYLDEIFAPHLQEVDENFLRRLKQSNHPEDPAKNQAFI
GKLLFPDLLKKNERGYPTLIKMRDELPVEQRAHYPVTNIYKLREAMINEDRQFDLR
EVYLAVHHIVKYRGHFLNNASVDKFKVGRIDFDKSFNVLNEAYEELQNGEGSFTIE
PSKVEKIGQLLLDTKMRKLD RQKAVAKLLEVKVADKEETKRNKQIATAMSKLVLG
YKADFATVAMANGNEWKIDLSSETSEDEIEKFREELSDAQNDILTEITSLFSQIMLNE
IVPNGMSISESMMDRYW THERQLAEVKEYLATQPASARKEFDQVYNKYIGQAPKE
KGFDEKGLKKILSKKENWKEIDELLKAGDFLPKQRTSANGVIPHQMHQQELDRII
EKQAKYYPWLATENPATGERDRHQAKYELDQLVSFRIPYYVGPLVTPEVQKATSGA
KFAWAKRKEDGEITPWNLWDKIDRAESAEAFIKRMTVKD TYLLNEDVLPANSLLYQ
KYNVLNELNNVRVNGRRLSVGIKQDIYTELFKKKKTVKAGDVASLVMATRGN
KPSVEGLSDPKKFNSNLATYLDLKSIVGDKVDDNRYQMDLENIIEWRSVFEDGEIFA
DKLTEVEWLTDEQRSALVKKRYKGWGRLSKKLLTGIVDENGQRIIDLMWNTDQNF
MQIVNQPVFKEQIDQLNQKAITNDGMTLRERVESVLDDAYTSPQNKKAIWQVVRV
VEDIVKAVGNAPKSSISIEFARNEGKGEITRSRRTQLQKLFEDQAHEL VKDTSLTEEL
EKAPDLSDRYYFYFTQGGKDMYTGD PINFDEISTKYDIDHILPQSFVKDDSLDNRLV

VSRAENNKKSDRVPKLYAAKMKPYWNQLLKQGLITQRKFENLTMDVDQTIKYRS
LGFVKRQLVETRQVIKLTANILGSMYQEAGTDIETRAGLTKQLREEFDLPKVREVN
DYHHAVDAYLTTFAGQYLNRRYPKLRSFFVYGEYMKFKHGSDDLKLRNFFHELM
EGDKSQGKVVDQQTGELITTRDEVADYFDWVINLKVMLISNETYEETGKYFDASH
ESSSLYLKNQNKKSKLVVPLKNKLQPEYYGAYTGITQGYMVILKLLDKKGGFGVY
RIPRYAADILNKCHDEVAYRNKIAEIISSDPRAPKSFEVVVPRVLKGTFLVDGEEKFIL
SSYRYKVNATQLILPVSDIKLIQDNFKALKKLVEMQTKKLEIYDNILRQVDKYYK
LYDINKFRAKLHDGRSKFVELDDFGQDASKEKVIKILRGLHFGSDLQNLKEIGFGT
TPLGQFQVSEAGIRLSNTAFIIFKSPTGLFNRKLYLKNL

GKRPAATKKAGQAKKKKGS*YPYDVPDYAYPYDVPDYAYPYDVPDYA* (SEQ ID NO: 85)。

具有核定位信号(NLS)和接头的嗜脉菌属马赛-P3761 Cas9 序列

MPKKKRKVGEKKTNYTIGLDIGTDSVGWAVVKDDLELVKKRMKVLGNTETNYIK
KNLWGSLLFESGQTAKDRRLKRVARRRYERRRNRLTELQKIFAPAIDEVDENFFRL
NESFLVPEDKAFSKNPIFGTLGEDKTYKYKTYPTIYHLRQHLADSEEKADVRILIYAL
AHMIKYRGHFLIEGKLDTEHIAINENLEQFFESYNALFSEEPILRKEELIAIENILRE
KNSRTVKEKRITSFLKDIGRANKQSPMMAFITLIVGKKAKFKAAFNLEEEISLNLTD
DSYDENLEILLNTIGSDFADLFDHAQRVYNAVELAGILSGDVKNTHAKLSAQMVAM
YERHKEQLKEYKSFIKANLPDQYDMTFVAPKDAQKKDLKGYAGYIDGNMSQDSFY
KFVKDQLKEVPGSEKFLDSIEKEDFLRKQRSFYNGVIPNQVHLAEMEAILDRQENY
YPWLKENREKIIISLLTFRIPYYVGPLADGQSEFAWLERKSDEKIKPWNFSDVVDLDR
SAEKFIEQLIGRDTYLPDEYVLPKKSLEYQKYMVFNELTKIAYLDERQKRMNLSSVE
KKEIFETLFKKRSKVTEKQLVKFFENYLQIDNPTIFGIEDAFNADYSTYVELAKVPG
MKSMMDDPDNEDLMEEIVKILTVFEDRKMRKQLEKYKERLSPEQIKELAKKHYT
GWGRLSKKLLVGIRDKETQKTILDYLVEDDNHSGGRQHLNRNLMQLINDDRLSFK
KTIAELQMIDPSADLYAQVQEIAGSPAIKKGILLGLKIVDEIIRVMGEKPENIVIEMAR
ENQTTARGKALSKRREAKIKEGLAALGSSLLKENLPGNADLSQRKIYLYYTQNGKD
IYLDEPLDFDRLSQYDEDHIIPQSFTVDNSLDNLVLTNSSQNRGNKKDDVPSLEVVN
RQLAYWRSLKDAGLMTQRKFDNLTKAMRGGLTDKDRERFIQRQLVETRQITKNVA
KLLDMRLNDKKDEAGNKIRETNIVLLKSAMASEFRKMFRLYKVRELNDYHHAHD
AYLNAAIAINLLALYPYMADDFVYGEFRYKKKPKAEKATYEKLRQWNLIKRFGEK
QLFTPDHEDCWNKERDIKTIKKVMGYRQVNVVKKAEERTGMLFKETINGKTNKGS
RIPIKKDLDPISKYGGYIEEKMAYYAVISYEDKKKKPGKTIVGISIMDKKEFEYDSISY
LGKLGFSNPVVQIILKNYSLIAYPDGRRRYITGATKTTKGKVELQKANQIAMEQDLV

[0244]

NFIYHLKNYDEISHPESYAFVQSHTDYFDRLFDSIEHYTRRFLDAETNINRLRRIYEE
EKKKDPVDIEALVASFIELLKLTSAAGAPADFIEMGEAISRRRYNSMTGLFDGQVIYQS
LTGLYETRMRFEDGKRPAATKKAGQAKKKKGSYPYDVDPDYAYPYDVDPDY
AYPYDVDPDY (SEQ ID NO: 87)。

具有核定位信号(NLS)和接头的 *ScoCas9* 变体序列(*ScoCas9*-NGC-v1)

MPKKKRKVGMGKPYSIGLDIGTNSVGWAVVTDDYKVPAKKMKVLGNTDKQSIK
KNLLGALLFDSGETAEATRLKRTARRRYTRRKNRLRYLQEIFTGEMNKVDENFFQR
LDDSFLVDEDEKRGHHPIFGNIAAEVKYHDDFPTIYHLRRHLADTSKKADLRLVYL
ALAHMIKFRGHFLYEGDLKAENTDVQALFKDFVEEYDKTIEESHLSEITVDALSILT
EKVSKSSRLNLI AHYPTEKKNTLFGNLIASLDLHPNFKTNFQLSEDAKLQFSKDT
YEEDLEGFLGEVGDYADLFASAKNLYDAILLSGILTVDDNSTKAPLSASMVKRYEE
HQKDLKKLKDFIKVNAPDQYNAIFKDKNKKGYASYIESGVKQDEFYKYLKGILLKI
NGSGDFLDKIDREDFLRKQRTFDNGSIPHQIHLQEMHAILRRQGEHYPFLKENQDKI
EKILTFRIPIYVVGPLARKGSRFAWA EYKADEKITPWNFDDILDKEKSAEKFITRMTL
NDLYLP EEEKVLPKHSPLYEAFTVYNELTKVKYVNEQGEAKFFDTNMKQEIFDHVFK
ENRKVTKDKLLNYLNKEFEFRIVNLTGLDKENKAFNSSLGTYHDLRKILDKSFLD
DKANEKTIEDIIQTTLTFEDREMIRQLQKYS DIFTKAQLKKLERRHYTGWGRLSYK
LINGIRNKENKKTILDY LIDDGYANRNF MQLINDDALSFKEEIARAQIIDDVDDIANV
VHDLPGSPA IKKGILQSVKIVDELVKVMGHN PANIIEMARENQTTDKGRRNSQQR L
KLLQDSLKNLDNPVNIKNVENQQLQNDRLFLYYIQNGKDMYTGETLDINNLSQYDI
DHII PQAFIKDNSLDNRVLTRSDKNRGKSDDVPSIEVVHEMKSFW SKLLSVKLITQR
KFDNLTKAERGGLTEEDKAGFIKRQLVETRQITKHVAQILDERFNTEFDGNKRRIRN
VKIITLKS NLVSNFRKEFELYKVREINDYHHAHDAYLNAVVG NALLLKYPQLEPEFV
YGEY PKYNSYRSRKSATEKFLFY SNILRFFKKEDIQTNEDGEIAWNKEKHIKILRKV
LSYPQVNIVKKTEEQTGGFSKESILPKGESDKLIPRKTKNSYWDPKKYGGFMQPVV
AYSILVFADVEKGKSKLRKVQDMVGITIMEKKRFEKNPVDFLEQRGYRNVRLEKI
IKLPKYSLFELENKRRRL LASAKFLQKGNELVIPQRFTTLLYHSYRIEKDYEP EHREY
VEKHKDEFKELLEYISVFSRKYVLADNNLT KIEMLF SKNKDAEVSS LAKSFISLLTFT
AFGAPRAFNF GENIARKEYRSVTECLNATLIHQ SITGLYETRIDL SKLGEDGEGAD
KRTADGSEFESPKKKRKV (SEQ ID NO: 95)

具有核定位信号(NLS)和接头的 *ScoCas9* 序列(*ScoCas9*-NGC-v2)

MPKKKRKVGMGKPYSIGLDIGTNSVGWAVVTDDYKVPAKKMKVLGNTDKQSIK
KNLLGALLFDSGETAEATRLKRTARRRYTRRKNRLRYLQEIFTGEMNKVDENFFQR
LDDSFLVDEDEKRGHHPIFGNIAAEVKYHDDFPTIYHLRRHLADTSKKADLRLVYL

[0245]

[0246]

ALAHMIKFRGHFLYEGDLKAENTDVQALFKDFVEEYDKTIEESHLSEITVDALSILT
 EKVSKSSRLLENLIAHYPTKEKNTLFGNLIASLDLHPNFKTNFQLSEDAKLQFSKDT
 YEEDLEGFLGEVGDEYADLFASAKNLYDAILLSGILTVDDNSTKAPLSASMVKRYEE
 HQKDLKKLKDFIKVNAPDQYNAIFKDKNKKGYASYIESGVKQDEFYKYLKGILLKI
 NGSGDFLDKIDREDFLRKQRTFDNGIIPHQIHLQEMHAILRRQGEHYPFLKENQDKI
 EKILTFRIPYYVGPLARKGSRFAWAEEKADEKITPWNFDDILDKEKSAEKFITRMTL
 NDLYLPEEKVLPKHSPLYEAFTVYNELTKVKYVNEQGEAKFFDTNMKQEIFDHVFK
 ENRKVTKDKLLNYLNKEFEFRIVNLTGLDKENKAFNSSLGTYHDLRKILDKSFLD
 DKANEKTIEDIIQTLTLFEDREMIRQLQKYS DIFTKAQLKKLERLHYTGWGRLSYK
 LINGIRNKENKKTILDY LIDDGYANRNF MQ LINDDALSFKEEIARAQIIDDVDDIANV
 VHDLPGSPAIAKKGILQSVKIVDELVKVMGHN PANIIE MARENQTTDKGRNSQQR
 LLLQDSLKNLDNPVNIKNVENQQLQNDRLFLYYIQNGKDMYTGETLDINNLSQYDI
 DHIPQAFIKDNSLDNRVLTRSDKNRGKSDDVPSIEVVHEMKSFWSKLLSVKLITQR
 KFDNLTKAERGGLTEEDKAGFIKRQLVETRQITKHVAQILDERFNTEFDGNKRRIRN
 VKIITLKS NLVSNFRKEFELYKVREINDYHHAHDAYLNAVVG NALLLKYPQLEPEFV
 YGEYPKYNSYRSRKSATEKFLFYSNILRFFKKEDIQTNEEDGEIAWNKEKHILRKV
 LSY PQVNIVKKTEEQTGGFSKESILPKGESDKLIPRKTKNSYWDPKKYGGFMQPVV
 AYSILVFADVEKGKSKLRKVQDMVGITIMEKKRFEKNPVDFLEQRGYRNVRLEKI
 IKLPKYSLFELENKRRRLASAKFLQKGNELVIPQRFTTLLYHSYRIEKDYEPHREY
 VEKHKDEFKELLEYISVFSRKYVLADNNLTKIEMLF SKNKDAEVSSLAKSFISLLTFT
 AFGAPRAFNF FGENIARKEYRSVTECLNATLIHQ SITGLYETRIDL SKLGEDGEGAD
KRTADGSEFESP KKKRKV (SEQ ID NO: 96)

[0247] NLS(粗体), 可以用不同的NLS取代

[0248] 接头(带下划线), 可以去除或扩增

[0249] 在一些实施例中, 工程化的非天然存在的Cas9包括标签。如本领域技术人员将显而易见, 取决于目的, 多个标签可与Cas9变体(例如, 3xHA标签)融合。

[0250] 不同物种呈现密码子偏倚(即生物体在密码子使用方面的差异), 这与通过利用信使RNA(mRNA)中的密码子进行mRNA翻译的效率有关, 这些mRNA中的密码子与特定生物体中该密码子的tRNA种类的丰度相对应。本领域的各种方法可用于计算机优化, 包括例如通过使用软件优化。在一些实施例中, 密码子优化是指通过用在宿主细胞的基因中更频繁使用或最频繁使用的密码子替换天然序列的至少一个密码子(例如1、2、3、4、5、10、15、20、25、50或更多个密码子)同时保持天然氨基酸序列来修饰核酸序列以增强在所关注宿主细胞中的表达。

[0251] 在一些实施例中, 本文所述的Cas9蛋白经密码子优化。这种类型的优化是本领域已知的, 并且需要外源衍生的DNA的突变以模拟预期宿主生物体或细胞的密码子偏好, 同时编码相同的蛋白质。因此, 密码子发生变化, 但编码的蛋白质保持不变。密码子优化会提高特定物种的可溶性蛋白质水平并增加活性和编辑效率。密码子优化还导致翻译和蛋白质表达增加。

[0252] 在一些实施例中,Cas9蛋白经密码子优化以用于在真核细胞中表达。在一些实施例中,Cas9蛋白经密码子优化以用于在人类细胞中表达。

[0253] 前间隔序列邻近基序 (PAM)

[0254] 每种Cas核酸内切酶仅在非靶向(即互补)DNA链上存在特定序列(称为前间隔序列邻近基序(PAM))的情况下才与其靶序列结合。从不同细菌物种中分离出的Cas核酸酶可识别不同的PAM序列。例如,SpCas9核酸酶(来自酿脓链球菌)切割PAM序列5'-NGG-3'(其中"N"可以是任何核苷酸碱基)的上游,SaCas9(来自金黄色葡萄球菌)识别靶标中的PAM序列5'-NNGRR(N)-3'。因此,不同Cas蛋白可以靶向的基因组中的位置受到独特PAM序列的位置限制。

[0255] 本文公开的由星座链球菌和秘鲁埃扎基拉菌和嗜脲菌属马赛-P3761物种工程化得到的Cas9蛋白识别共有PAM序列5'-NGG-3'。本文公开的由星座链球菌和秘鲁埃扎基拉菌和嗜脲菌属马赛-P3761物种工程化得到的Cas9蛋白识别共有PAM序列5'-NGG-3'。在一些实施例中,本文公开的Cas9蛋白经工程化以识别共有PAM序列5'-NGC-3'。例示性实施例描述于下文中并且应是非限制性的。在一些实施例中,来自星座链球菌的Cas9蛋白经工程化以识别共有PAM序列5'-NGC-3'。在一些实施例中,NGC PAM变体包括相对于ScoCas9(SEQ ID NO:1)选自或对应于D1117M、S118Q、E1201F、A1299R、D1309A、R1312E和T1314R(统称为"MQFRAER")的一个或多个氨基酸取代。在一些实施例中,NGC PAM变体包括相对于天然存在的SpyCas9(SEQ ID NO:173)选自或对应于D1135M、S1136Q、G1218K、E1219F、A1322R、D1332A、R1335E和T1337R(统称为"MQKFRAER")的一个或多个氨基酸取代。在一些实施例中,可以对SirCas9、VapCas9、EpeCas9、LfeCas9或PmaCas9进行类似或相应的氨基酸取代。

[0256] 酿脓链球菌Cas9(SpyCas9;GenBank:QSG91308.1)

[0257] MDKKYSIGLDIGTNSVGWAVITDDYKVP SKKFKVLGNTDRHS IKKNLIGALLFDSGETAEATRLKRTA
RRRYTRRKNRICYLQEIFS NEMAKVDDSFHRL EESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLV
STDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSR
RLENLIAQLPG EKKNGLFGNLIALSLGLTPNFKS NFDLAEDAKLQLSKD TYDDDLDNLLAQIGDQYADLFLAAKNL
SDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHA ILRRQEDFYFPFLKDNREKIEKILTFRI
PYVVGPLARGNSRFAMTRKSEETITPWNFE EVVDKGASAQ SFIERMTNFDKNLPNEKVLPHKSLLEYFTVYNE
LTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLK
IIKDKDFLDNEENEDI EDIVLTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQS
GKTILDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA IKKGILQTVKVVDELVKVM
GRHKPENIVIE MARENQTTQKGQKNSRERMKRIE EGIKELGSQILKEHPVENTQLQNEKLYLYLQNGRDMYVDQE
LDINRLSDYD VDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKA
ERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKS KLVSDFRKDFQFYKVR E IN
NYHHAHDAYLNAVVG TALIKKYPKLESEFVYGDYKVYDVRKMI AKSEQEIGKAT AKYFFYSNIMNFFKTEITLANG
EIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKY
GGFDSPTVAYSVLVVAKEKGKSKKLKSVKELLGITIMERS SFEKNPIDFLEAKGYKEVKKDLIIKLPKYS LFELE
NGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLK GSPEDNEQKQLFVEQHKHYLDEIIIEQISEFSKRVILA
DANLDKVL SAYNKH RDKPIREQAENI IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQ SITGLYETR

IDLSQLGGD (SEQ ID NO:173)。

[0258] 在一些实施例中,本文所述的Cas9蛋白不与任何其他PAM序列结合或表现出活性。

[0259] RNA向导

[0260] RNA向导包括与靶序列互补的多核苷酸序列。RNA向导与靶核酸序列杂交并指导CRISPR复合物与靶核酸的序列特异性结合。在一些实施例中,RNA向导与靶核酸序列具有50%、60%、70%、75%、80%、85%、90%、95%、96%、97%、98%、99%或100%的互补性。

[0261] 在一些实施例中,RNA向导的长度为约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或更多个核苷酸。在一些实施例中,RNA向导的长度为约18-24个核苷酸。在一些实施例中,RNA向导与靶核酸序列中的约18-24个核苷酸互补。例如,RNA向导与靶核酸序列中的约18、19、20、21、22、23或24个核苷酸互补。在一些实施例中,RNA向导与约18-22个核苷酸互补。在一些实施例中,RNA向导与约18-21个核苷酸互补。在一些实施例中,RNA向导与约18-20个核苷酸互补。在一些实施例中,RNA向导与靶核酸序列中的20个核苷酸互补。

[0262] 可以设计RNA向导以靶向任何靶序列。使用用于比对序列的任何算法确定最佳比对,所述算法包含尼德曼-翁施算法 (Needleman-Wunsch algorithm)、史密斯-沃特曼算法 (Smith-Waterman algorithm)、伯罗斯-惠勒算法 (Burrows-Wheeler algorithm)、ClustlW、ClustlX、BLAST、Novoalign、SOAP、Maq和ELAND。

[0263] 在一些实施例中,RNA向导被靶向细胞基因组内的独特靶序列。在一些实施例中,RNA向导被设计成缺少PAM序列。在一些实施例中,使用包括mFold或Geneious在内的折叠算法将RNA向导序列设计成具有最佳二级结构。在一些实施例中,RNA向导的表达可以在诱导型启动子下进行,例如激素诱导型、四环素或强力霉素诱导型、阿拉伯糖诱导型或光诱导型启动子。

[0264] 在一些实施例中,CRISPR系统包括一种或多种RNA向导,例如crRNA、tracrRNA和/或sgRNA。因此,在一些实施例中,RNA向导包含crRNA。在一些实施例中,RNA向导包含tracrRNA。在一些实施例中,RNA向导包含sgRNA。在一些实施例中,CRISPR系统包括多个RNA向导,包含1、2、3、4、5、6、7、8、9、10、15或更多个RNA向导。

[0265] 在一些实施例中,RNA向导包括crRNA。在一些实施例中,CRISPR系统包括多个crRNA,包含2-15个crRNA。在一些实施例中,crRNA是前体crRNA (前crRNA),其包括同向重复序列、间隔序列子和同向重复序列。在一些实施例中,crRNA是经加工或成熟的crRNA,其包括截短的同向重复序列。

[0266] 在一些实施例中,CRISPR相关蛋白切割前crRNA以形成经加工或成熟的crRNA。

[0267] 在一些实施例中,CRISPR相关蛋白与成熟crRNA形成复合物,并且间隔子序列将该复合物靶向靶核酸中的互补序列。在一些实施例中,RNA向导包含能够在适当条件下与靶核酸杂交的同向重复序列和间隔子序列。

[0268] 在一些实施例中,crRNA的间隔子长度可以在约15至50个核苷酸的范围内。在一些实施例中,RNA向导的间隔子长度为至少16个核苷酸、至少17个核苷酸、至少18个核苷酸、至少19个核苷酸、至少20个核苷酸、至少21个核苷酸或至少22个核苷酸。在一些实施例中,间隔子长度为15至17个核苷酸 (例如,15、16或17个核苷酸)、17至20个核苷酸 (例如,17、18、19或20个核苷酸)、20至24个核苷酸 (例如,20、21、22、23或24个核苷酸)、23至25个核苷酸 (例

如,23、24或25个核苷酸)、24至27个核苷酸、27至30个核苷酸、30至45个核苷酸(例如,30、31、32、33、34、35、36、37、38、39、40、41、42、43、44或45个核苷酸)、30或35至40个核苷酸、41至45个核苷酸、45至50个核苷酸(例如,45、46、47、48、49或50个核苷酸)或更长。

[0269] 在一些实施例中, RNA向导包含长度介于约16至26个核苷酸之间的同向重复 (DR) 序列。例如, 在一些实施例中, DR的长度为约16个核苷酸。在一些实施例中, DR的长度为约17个核苷酸。在一些实施例中, DR的长度为约18个核苷酸。在一些实施例中, DR的长度为约19个核苷酸。在一些实施例中, DR的长度为约20个核苷酸。在一些实施例中, DR的长度为约21个核苷酸。在一些实施例中, DR的长度为约22个核苷酸。在一些实施例中, DR的长度为约23个核苷酸。在一些实施例中, DR的长度为约24个核苷酸。在一些实施例中, DR的长度为约25个核苷酸。在一些实施例中, DR的长度为约26个核苷酸。

[0270] 在一些实施例中, crRNA包含核苷酸向导序列和DR序列。核苷酸向导序列的长度可以介于约18与24个核苷酸之间。因此, 在一些实施例中, 核苷酸向导序列的长度为约18个核苷酸。在一些实施例中, 核苷酸向导序列的长度为约19个核苷酸。在一些实施例中, 核苷酸向导序列的长度为约20个核苷酸。在一些实施例中, 核苷酸向导序列的长度为约21个核苷酸。在一些实施例中, 核苷酸向导序列的长度为约22个核苷酸。在一些实施例中, crRNA包含长度为约22个核苷酸的核苷酸向导序列和长度为约22个核苷酸的同向重复序列。

[0271] 在一些实施例中, crRNA序列可经修饰为“死crRNA”、“死向导”或“死向导序列”, 它们可以与CRISPR相关蛋白形成复合物并结合特定靶标而没有任何实质性核酸酶活性。

[0272] 在一些实施例中, crRNA可以在糖磷酸骨架或碱基中被化学修饰。在一些实施例中, crRNA可以使用2'-O-甲基、2'-F或锁核酸进行修饰, 以提高核酸酶抗性 or 碱基配对。在一些实施例中, crRNA可以含有经修饰的碱基, 诸如2-硫脲二烯或N6-甲基腺苷。

[0273] 在一些实施例中, crRNA与其他寡核苷酸、肽、蛋白质、标签、染料或聚乙二醇缀合。

[0274] 在一些实施例中, crRNA可包括适体或核糖开关序列, 其可由于三维结构而结合特定的靶分子。

[0275] 在一些实施例中, 反式活化RNA (tracrRNA) 与crRNA缔合以促进与Cas9蛋白形成复合物。在一些实施例中, tracrRNA序列的长度为约或不超过约5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、25、30、40、50、60、70、80、90、100或更多个核苷酸。在一些实施例中, tracrRNA的长度为约70个核苷酸。

[0276] 在一些实施例中, tracrRNA和crRNA包含在称为单向导RNA (sgRNA) 的单个转录物中。在一些实施例中, sgRNA包括在tracrRNA与sgRNA之间的环。

[0277] 在一些实施例中, 所述环形成序列的长度为3、4、5或更多个核苷酸。在一些实施例中, 该环具有序列GAAA、AAAG、CAAA、AAAC、UUUU、UUAUUAU、UUA、UUU和/或AAUCA。在一些实施例中, 该环具有序列GAAA。在一些实施例中, 该环具有序列AAAG。在一些实施例中, 该环具有序列CAAA。在一些实施例中, 该环具有序列AAAC。在一些实施例中, 该环具有序列AAUCA。在一些实施例中, 该环具有序列UUUU。在一些实施例中, 该环具有序列UUAUUAU。在一些实施例中, 该环具有序列UUA。在一些实施例中, 该环具有序列UUU。在一些实施例中, 该环具有序列AAUCA。

[0278] 在一些实施例中, tracrRNA和crRNA形成发夹环。在一些实施例中, sgRNA具有至少两个或更多个发夹。在一些实施例中, sgRNA具有两个、三个、四个或五个发夹。

[0279] 在一些实施例中,sgRNA包括转录终止序列,该序列包括包含六个核苷酸的polyT序列。

[0280] 在一些实施例中,sgRNA包含与以下具有至少80%同一性的序列:ScoCas9的5'-G UUUUAGAGCUGUGUGUUUAAACAACACAGCAAGUAAAAUAAGGCUUUGUCCGUACUC (SEQ ID NO:3)、

[0281] SirCas9的5'-GUUUUAGAGUUGUGUUAUUGAAAAUAACACAACGAGUAAAA

[0282] AUAAGCUUAUGCUUAAAUGC CAGCUUUGCUGGUGUCAUUUAGAUGACUUUACUAAGGUUGCUUCGG CAACCUUUUU-3' (SEQ ID NO:7)、

[0283] VapCas9的5'-GUUUGAGAGUAGUGUGAAAACAUUACGAGUUCAAAAUACAAA

[0284] UUAUUUACAAUGCCUUCGGGCGUCCCGACGUAGGGCACCUCUCUCAAUUCUUCGGAAUUGAGUU-3' (SEQ ID NO:13)、

[0285] EpeCas9的5'-GUUUGAGAGUUAUGUAAUUGAAAAUUAUGACGAGUUCAA

[0286] AUAUUUUUUAUUCAAACCGCCUAAUUUAUAGGCCGCGAGAUGUUCUGCAUUAUGCUUGCUAUUGCAAGC UU-3' (SEQ ID NO:19)、

[0287] LfeCas9的5'-GUCUUGGAGAGUGUGAAAAACACUCAUAGUCAAGAUCAAACG

[0288] AGUGGUUUUCCACGAGUUAUUACUUUUGAGGUCUUAUAUGGCCCAUACAUAUAAAAGGAGUCGGAAUUU CCGGCUCUUUUUCU-3' (SEQ ID NO:95) 以及

[0289] PmaCas9的5'-GUUUUAGAGCCAUGUAGAAUACAUUGCAAGUAAAAUAAGG

[0290] CUUUGUCCGUAUACAACUUGAAAAAGUGGCGCUGUUUCGGCGCUU-3' (SEQ ID NO:96)。

[0291] 将向导RNA添加到Cas9的5'端。在一些实施例中,sgRNA包含与SEQ ID NO:3、7、13、19、95或96具有80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或更高同一性的序列。在一些实施例中,sgRNA包含与SEQ ID NO:3、7、13、19、95或96相同的序列。

[0292] 在一些实施例中,tracrRNA是单独的转录物,不与crRNA序列包含在同一转录物中。

[0293] Cas9融合蛋白

[0294] 在一些实施例中,Cas9酶与一个或多个异源蛋白质结构域融合。在一些实施例中,Cas9酶与超过约1、2、3、4、5、6、7、8、9、10或更多个蛋白质结构域融合。在一些实施例中,异源蛋白质结构域与Cas9酶的C末端融合。在一些实施例中,异源蛋白质结构域与Cas9酶的N末端融合。在一些实施例中,异源蛋白质结构域在Cas9酶的C末端与N末端之间内部融合。在一些实施例中,内部融合在Cas9 RuvCI、RuvC II、RuvCIII、HNH、REC I或PAM相互作用结构域内进行。

[0295] Cas9蛋白可以直接或间接地连接至另一蛋白结构域。在一些实施例中,合适的CRISPR系统含有接合Cas9蛋白和异源蛋白的接头或间隔子。氨基酸接头或间隔子通常被设计为柔性的或在两个蛋白质部分之间插入结构,诸如 α -螺旋。接头或间隔子可以相对较短,或可以较长。通常,接头或间隔子在长度上含有例如1-100 (例如,1-100、5-100、10-100、20-100、30-100、40-100、50-100、60-100、70-100、80-100、90-100、5-55、10-50、10-45、10-40、10-35、10-30、10-25、10-20) 个氨基酸。在一些实施例中,接头或间隔子的长度等于或长于1、2、3、4、5、6、7、8、9、10、15、20、25、30、35、40、45、50、55、60、65、70、75、80、85、90、95或100个氨基酸。通常,较长的接头可以减少空间位阻。在一些实施例中,接头将包含甘氨酸和丝

氨酸残基的混合物。在一些实施例中,接头可以另外包含苏氨酸、脯氨酸和/或丙氨酸残基。

[0296] 在一些实施例中,Cas9蛋白与具有酶活性、表观遗传修饰活性、RNA切割活性、核酸结合活性、转录调节活性的细胞定位信号、表位标签、报告基因和蛋白质结构域融合。在一些实施例中,Cas9蛋白与核定位序列(NLS)、FLAG标签、HIS标签和/或HA标签融合。

[0297] 合适的融合配偶体包括但不限于提供甲基转移酶活性、去甲基化酶活性、乙酰转移酶活性、去乙酰化酶活性、激酶活性、磷酸酶活性、泛素连接酶活性、去泛素化活性、腺苷酸化活性、去腺苷酸化活性、SUMO化活性、去SUMO化活性、核糖基化活性、去核糖基化活性、肉豆蔻酰化活性、去肉豆蔻酰化活性、整合酶活性、转座酶活性、重组酶活性、聚合酶活性、连接酶活性、解旋酶活性或核酸酶活性的多肽,其中任何一种都可以修饰DNA或DNA相关多肽(例如,组蛋白或DNA结合蛋白)。在一些实施例中,Cas9蛋白与组蛋白去甲基化酶、转录活化因子或脱氨酶融合。

[0298] 其他合适的融合配偶体包括但不限于边界元件(例如,CTCF)、提供外周募集的蛋白质及其片段(例如核纤层蛋白A、核纤层蛋白B等)和蛋白质对接元件(例如,FKBP/FRB、Pill/Ablyl等)。

[0299] 在特定实施例中,Cas9与胞苷或腺苷脱氨酶结构域融合,例如以用于碱基编辑。在一些实施例中,Cas9与腺嘌呤和胞嘧啶碱基编辑器(ACBE或CABE)融合,其中ACBE或CABE通过将TadA的异二聚体和活化诱导的胞苷脱氨酶(AID)与Cas9切口酶(nCas9)的N末端和C末端融合来产生。在一些实施例中,ACBE或CABE同时诱导同一靶位点处的C至T和A至G碱基编辑。Xie,J等人ACBE,a new base editor for simultaneous C-to-T and A-to-G substitutions in mammalian systems.BMC Biology(18:131),2020)

[0300] 在一些实施例中,术语“胞苷脱氨酶”和“胞嘧啶脱氨酶”可以互换使用。在某些实施例中,胞苷脱氨酶结构域可以与本文所述的任何胞苷脱氨酶具有70%、75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或更高的序列同一性。在一些实施例中,胞苷脱氨酶结构域具有胞苷脱氨酶活性(例如,将C转化为U)。在某些实施例中,腺苷脱氨酶结构域可以与本文所述的任何腺苷脱氨酶具有70%、75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或更高的序列同一性。在一些实施例中,腺苷脱氨酶结构域具有腺苷脱氨酶活性(例如,将A转化为I)。在一些实施例中,术语“腺苷脱氨酶”和“腺嘌呤脱氨酶”可以互换使用。

[0301] 在一些实施例中,胞苷脱氨酶可包含载脂蛋白B mRNA编辑复合物(APOBEC)家族脱氨酶的全部或一部分。APOBEC是进化上保守的胞苷脱氨酶的家族。该家族的成员是C至U编辑酶。APOBEC样蛋白的N末端结构域是催化结构域,而C末端结构域是假催化结构域。更具体地,催化结构域是锌依赖性胞苷脱氨酶结构域并且对于胞苷脱氨作用很重要。APOBEC家族成员包括APOBEC1、APOBEC2、APOBEC3A、APOBEC3B、APOBEC3C、APOBEC3D(“APOBEC3E”现指此)、APOBEC3F、APOBEC3G、APOBEC3H、APOBEC4和活化诱导的(胞苷或胞嘧啶)脱氨酶。在一些实施例中,掺入融合蛋白中的脱氨酶包含APOBEC1脱氨酶的全部或一部分。在一些实施例中,掺入融合蛋白中的脱氨酶包含APOBEC2脱氨酶的全部或一部分。在一些实施例中,掺入融合蛋白中的脱氨酶包含APOBEC3脱氨酶的全部或一部分。在一些实施例中,掺入融合蛋白中的脱氨酶包含APOBEC3A脱氨酶的全部或一部分。在一些实施例中,掺入融合蛋白中的脱氨酶包含APOBEC3B脱氨酶的全部或一部分。在一些实施例中,掺入融合蛋白中的脱氨酶包

含APOBEC3C脱氨酶的全部或一部分。在一些实施例中,掺入融合蛋白中的脱氨酶包含APOBEC3D脱氨酶的全部或一部分。在一些实施例中,掺入融合蛋白中的脱氨酶包含APOBEC3E脱氨酶的全部或一部分。在一些实施例中,掺入融合蛋白中的脱氨酶包含APOBEC3F脱氨酶的全部或一部分。在一些实施例中,掺入融合蛋白中的脱氨酶包含APOBEC3G脱氨酶的全部或一部分。在一些实施例中,掺入融合蛋白中的脱氨酶包含APOBEC3H脱氨酶的全部或一部分。在一些实施例中,掺入融合蛋白中的脱氨酶包含APOBEC4脱氨酶的全部或一部分。在一些实施例中,掺入融合蛋白中的脱氨酶包含活化诱导的脱氨酶(AID)的全部或一部分。在一些实施例中,掺入融合蛋白中的脱氨酶包含胞苷脱氨酶1(CDA1)的全部或一部分。应当理解,融合蛋白可以包含来自任何合适的生物体(例如,人类或大鼠)的脱氨酶。在一些实施例中,融合蛋白的脱氨酶结构域来自人类、黑猩猩、大猩猩、猴、牛、狗、大鼠或小鼠。在一些实施例中,融合蛋白的脱氨酶结构域源自大鼠(例如,大鼠APOBEC1)。在一些实施例中,脱氨酶结构域是人类APOBEC1。在一些实施例中,脱氨酶结构域是pmCDA1。以下提供了例示性胞苷脱氨酶的序列。

[0302] pmCDA1(海七鳃鳗(*Petromyzon marinus*))

[0303] MTDAEYVRIHEKLDIYTFKKQFFNNKKSVSRCYVLFELKRRGERRACFWGYAVNKPQSGTERGIHAE
IFSIRKVEEYLRDNPGQFTINWYSSWSPCADCAEKILEWYNQELRGNGHTLKIWACKLYYEKNARNQIGLWNL
RDNGVGLNVMVSEHYQCCRKIFIQSSHNQLNENRWLEKTLKRAEKRRSELSIMIQVKILHTTKSPAV (SEQ ID NO:
22)

[0304] 人类AID:

[0305] MDSLLMNRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKGCHVELLFLRYISDW
DLDPGRCYRVTWFTSWSPCYDCARHVADFLRGPNLSLRIFTARLYFCEDRKAEPGLRRLHRAGVQIAIMTFKAPV
(SEQ ID NO:23)

[0306] 人类AID:

MDSLLMNRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKGCH

[0307] VELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGPNLSLRIFTARLYF
CEDRKAEPGLRRLHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAWEGLHENSURL
SRQLRRILLPLYEVDDLRLDAFRTLGL (下划线:核定位序列;双下划线:核输出
信号) (SEQ ID NO:24)

[0308] 小鼠AID:

[0309] MDSLLMKQKKFLYHFKNVRWAKGRHETLYCYVVKRRDSATSCSLDFGHLRNKSGCH
VELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVAEFLRWPNLSLRIFTARLYF
CEDRKAEPGLRRLHRAGVQIGIMTFKDYFYCWNTFVENRERTFKAWEGLHENSURL
RQLRRILLPLYEVDDLRLDAFRMLGF (下划线:核定位序列;双下划线:核输出
信号) (SEQ ID NO:25)

[0310] 犬AID:

MDSLLMKQRKFLYHFKNVRWAKGRHETLYCYVVKRRDSATSFSLDFGHLRNKSGCH

[0311] VELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGYPNLSLRIFAARLYF
CEDRKAEPGLRRLHRAGVQIAIMTFKDYFYCWNTFVENREKTFKAWEGLHENSURL
SRQLRRILLPLYEVDDLRLDAFRTLGL (下划线:核定位序列;双下划线:核输出
信号)

信号) (SEQ ID NO:26)

[0312] 牛AID:

MDSLLKKQRQFLYQFKNVRWAKGRHETLYCYVVKRRDSPTSFSLDFGHLRNKAGCHV

[0313] ELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGYPNLSLRIFTARLYFC
DKERKAEPEGLRRLHRAGVQIAIMTFKDYFCWNTFVENHERTFKAWEGLHENSURL
SRQLRRILLPLYEVDDLRLDAFRTLGL (下划线:核定位序列;双下划线:核输出

信号) (SEQ ID NO:27)

[0314] 大鼠AID:

MAVGSKPKAALVGPHWERERIWCFLCSTGLGTQQTGQTSRWLRPAATQDPVSPPRSLL
MKQRKFLYHFKNVRWAKGRHETLYCYVVKRRDSATSFSLDFGYLRNKSGCHVELLFL

[0315] RYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGYPNLSLRIFTARLTGWGALP
AGLMSPARPSDYFCWNTFVENHERTFKAWEGLHENSURLSRRLRRILLPLYEVDDLRL
DAFRTLGL (SEQ ID NO: 28)

[0316] (下划线:核定位序列;双下划线:核输出信号)

[0317] clAID(家犬(Canis lupus familiaris)):

[0318] MDSLLMKQRKFLYHFKNVRWAKGRHETLYCYVVKRRDSATSFSLDFGHLRNKSGCHVELLFLRYISDW
DLDPGRCYRVTWFTSWSPCYDCARHVADFLRGYPNLSLRIFAARLYFCEDRKAPEGLRRLHRAGVQIAIMTFKDY
FYCWNTFVENREKTFKAWEGLHENSURLSRQLRRILLPLYEVDDLRLDAFRTLGL (SEQ ID NO:29)

[0319] btAID(家牛(Bos taurus)):

[0320] MDSLLKKQRQFLYQFKNVRWAKGRHETLYCYVVKRRDSPTSFSLDFGHLRNKAGCHVELLFLRYISDW
DLDPGRCYRVTWFTSWSPCYDCARHVADFLRGYPNLSLRIFTARLYFCEDRKAPEGLRRLHRAGVQIAIMTFKD
YFCWNTFVENHERTFKAWEGLHENSURLSRQLRRILLPLYEVDDLRLDAFRTLGL (SEQ ID NO:30)

[0321] mAID(小家鼠(Mus musculus)):

[0322] MDSLLMNRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKGCHVELLFLRYISDW
DLDPGRCYRVTWFTSWSPCYDCARHVADFLRGYPNLSLRIFTARLYFCEDRKAPEGLRRLHRAGVQIAIMTFKDY
FYCWNTFVENHERTFKAWEGLHENSURLSRQLRRILLPLYEVDDLRLDAFRTLGL (SEQ ID NO:31)

[0323] rAPOBEC-1(挪威鼠(Rattus norvegicus)):

[0324] MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFI
E KFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTFLFIYIARLYHHADPRNRQGLRDLIS SGVTIQI
MTEQESGYCWRNFVNYSNEAHWPYPHLLWRLYVLELYCIILGLPPCLNLRKQPQLTFFTIALQSCHYQRLP
PHILWATGLK (SEQ ID NO:32)

[0325] maAPOBEC-1(金仓鼠(Mesocricetus auratus)):

[0326] MSSETGPVVVDPTLRRRIEPHEFDQFQDQGLRKETCLLYEIRWGGRHNIWRHTGQNTSRHVEINFI
E KFTSERYFYPSTRCSIVWFLSWSPCGECSKAITEFLSGHPNVTFLFIYAARLYHHTDQQRNRQGLRDLISRGVTIRIM
TEQEYCYCWRNFVNYPSPNEVYWPYPNLWMRLYALELYCIHLGLPPCLKIKRRHQYPLTFFRLNLQSCHYQRIIP
HILWATGFI (SEQ ID NO:33)

[0327] ppAPOBEC-1(婆罗州猩猩(Pongo pygmaeus)):

[0328] MTSEKGPSTGDPTLRRRIESWEFDVFYDPRELRKETCLLYEIKWGMSRKIWRSGKNTTNHVEVNFI
E KKFTSERRFHSSISCSITWFLSWSPCWECSQAIREFLSQHPGVTLVIIYVARLFWHMDQRNRQGLRDLVNS GVTIQ

IMRASEYYHCWRNFVNYPPGDEAHWPQYPPLWMMLYALELHCIILSLPPCLKISRRWQNHLAFFRLHLQNCHYQTI
PPHILLATGLIHPSVTWR (SEQ ID NO:34)

[0329] ocAPOBEC1 (欧洲兔 (*Oryctolagus cuniculus*)):

[0330] MASEKGPSNKDYTLRRRIEPWEFEVFFDPQELRKEACLLYEIKWGASSKTWRSSGKNTTNHVEVNFL
KLTSEGR LGPSTCCSITWFLSWSPCWECSMAIREFLSQHPGVTLIIFVARLFQHMDRRNRQGLKDLVTS
GVTVRVM
SVSEYCYCWENFVNYPPGKAAQWPRYPWRMLMYALELYCIIILGLPPCLKISRRHQKQLTFFSLTPQYCHYKMIPP
YILLATGLLQPSVPWR (SEQ ID NO:35)

[0331] mdAPOBEC-1 (短尾负鼠 (*Monodelphis domestica*)):

[0332] MNSKTGPSVGATLRRRIKPWEFVAFNPNQELRKETCLLYEIKWGNQNIWRHSNNTSQHAEINFMEK
FTAERHFNSSVRCSITWFLSWSPCWECSKAIRKFLDHYPNVTLAIFISRLYWHMDQQHRQGLKELVHSGVTIQIMS
YSEYHYCWENFVDYPQGEEDYWPKYPYLWIMLYVLELHCIILGLPPCLKISGSHSNQLALFSLDLQDCHYQKIPYN
VLVATGLVQPFVTWR (SEQ ID NO:36)

[0333] ppAPOBEC-2 (婆罗州猩猩):

[0334] MAQKEEAAAATEAASQNGEDLENLDDPEKLKELIELPPFEIVTGERLPANFFKFQFRNVEYSSGRNKT
FLCYVVEAQGGGQVQASRGYLEDEHAAAHAEAFNTILPAFDPALRYNVTWYVSSSPCAACADRIKTLSTKTN
LRLILVGR LFMWEELEIQDALKKLKEAGCKLRIMKPQDFEYVWQNFVEQEEGESKAFQPWEDIQENFLYEEKLA
DILK (SEQ ID NO:37)

[0335] btAPOBEC-2 (家牛):

[0336] MAQKEEAAAAAEPASQNGEEVENLEDPEKLKELIELPPFEIVTGERLPAHYFKFQFRNVEYSSGRNKT
FLCYVVEAQSGGGQVQASRGYLEDEHATNHAEAFNSIMPTFDPALRYMVTWYVSSSPCAACADRIVKTLNKT
LRLILVGR LFMWEEPEIQAALRKLKEAGCRLRIMKPQDFEYIWWQNFVEQEEGESKAFEPWEDIQENFLYEEKLA
DILK (SEQ ID NO:38)

[0337] mAPOBEC-3- (1) (家鼠):

[0338] MQPQRLGPRAGMPFCLGCSHRKCYSPIRNLISQETFKFHFKNLGYAKGRKDTFLCYE VTRKDCDSP
VSLHHGVFKNKDNIHAEICFLYWFHDKVLKVLSPREEFKITWYMSWSPCFECAEQIVRFLATHHNLSDIFSSRLY
NVQDPETQQNL CRLVQEGAQVAAMDLYEFKKCWKKFVDNGGRRFRPWKRLLTNFRYQDSKLQEILPCYISVPSS
SSTLSNICLTKGLPETRFWVEGRRMDPLSEEFYSQFYNQRVKHLCCYHRMKPYLCYQLEQFNGQAPLKGCLLSEK
GKQHAAILFLDKIRSMELSQVTITCYLTWSPCPNCAWQLAAFKRDRPDILHIYTSRLYFHWKRPFQKGLCSLWQS
GILVDVMDLPQFTDCWTNFVNPKRPFWPWKGLEIISRRTQRRLRRIKESWGLQDLVNDFGNLQLGPPMS (SEQ ID
NO:39)

[0339] 小鼠APOBEC-3- (2):

MGPFC LGCSHRKCYSPIRNLISQETFKFHFKNLGYAKGRKDTFLCYEVTRKDCDSPVSL
HHGVFKNKDNIHAEICFLYWFHDKVLKVLSPREEFKITWYMSWSPCFECAEQIVRFLATH
HNLSDIFSSRLYNVQDPETQQNL CRLVQEGAQVAAMDLYEFKKCWKKFVDNGGRRF

[0340] RPWKRLLTNFRYQDSKLQEILPCYIPVSSSSSTLSNICLTKGLPETRFCVEGRRMDPLS
EEEFYSQFYNQRVKHLCCYHRMKPYLCYQLEQFNGQAPLKGCLLSEK GKQHAAILFLDKIRSMELSQVTITCYLTWSPCPNCAWQLAAFKRDRPDILHIYTSRLYFHWKRPFQKGLC
SLWQSGILVDVMDLPQFTDCWTNFVNPKRPFWPWKGLEIISRRTQRRLRRIKESWGLQDLVNDFGNLQLGPPMS (斜体:核酸编辑结构域) (SEQ ID NO:40)

[0341] 大鼠APOBEC-3:

[0342] MGPFCCLGCSHRKCYSPIRNLISQETFKFHFKNRLRYAIDRKDTFLCYEVTRKDCDSPVSL
HHGVFKNKDNIHAEICFLYWFHDKVLKVLSPREEFKITWYMSWSPCFECAEQVLRFLATH
HNLSLDIFSSRLYNIRDPENQQNLCLRVQEGAQVAAMDLYEFKKCWKKFVDNGGRRFR
PWKKLLTNFRYQDSKLQEILRPCYIPVPSSSSSTLSNICLTKGLPETRFCVERRRVHLLSE
EEFYSSQFYNQRVKHLCCYYHGVKPYLCYQLEQFNGQAPLKGCLLSEKGGKQHAEILFLDK
IRSMELSQVIITCYLTWSPCPNCAWQLAAFKRDRPDILHIYTSRLYFHWKRPFPQKGLCSL
WQSGILVDVMDLPQFTDCWTNFVNPKRPFWPWKGLEIISRRTQRRLLHRIKESWGLQDL
VNDFGNLQLGPPMS (斜体:核酸编辑结构域) (SEQ ID NO:41)

[0343] hAPOBEC-3A (智人):

[0344] MEASPASGRHLMDFHFTSNFNNGIGRHKTYLCYEVERLDNGTSVKMDQHRGFLHNQAKNLLCGFYG
RHAELRFLDLVPSLQLDPAQIYRVTFISWSPCFSWGCAEVRAFLQENTHVRLRIFAARIYDYDPLYKEALQMLR
DAGAQVSIMTYDEFKHCWDTFVDHQGCPFPQWDGLDEHSQALSGRLRAILQNQGN (SEQ ID NO:42)

[0345] hAPOBEC-3F (智人):

[0346] MKPHFRNTVERMYRDTFSYNFYNRPILSRRNTVWLCYEVKTKGPSRPRLDAKIFRGQVYSQPEHHAEM
CFLSWFCGNQLPAYKCFQITWVFSWTPCPDCVAKLAEFLEHPNVTLTISAARLYYYWERDYRRALCRLSQAGARV
KIMDDEEFAYCWENFVYSEGQPFMPWYKFDNYAFLHRTLKEILRNPMEMYPHIFYFHFKNLRKAYGRNESWLFC
TMEVVKHHSPVSWKRGVFRNQVDPETHCHAERCFLSWFCDDILSPNTNYEVTWYTSWSPCECAGEVAEFLARHSN
VNLTIFTARLYYFWDTDYQEGRLSLSQEGASVEIMGYKDFKYCWENFVYNDDEPFKPKWGLKYNFLFLDSKLQEILE
(SEQ ID NO:43)

[0347] 恒河猴APOBEC-3G:

[0348] MVEPMDPRTFVSNFNNRPILSGLNTVWLCCEVKTCDPSGPPLDAKIFQGKVYSKAKYHPEMRFLRWFI
KWRQLHHDQEYKVTWYVSWSPCTRCANSVATFLAKDPKVTLTIFVARLYYFWKPDYQQALRILCQKRGGPHATMKI
MNYNEFQDCWNKFVDGRGKPFKPRNNLPKHYTELLQATLGELLRHLMDPGTFTSNFNKPWVSGQHETYLCYKVERL
HNDTWVPLNQHRGFLRNQAPNIHGFPGKGRHAELCFDLIPFWKLDGQQYRVTCFTSWSPCFSCAQEMAKFISNNEH
VSLCIFAARIYDDQGRYQEGRLALHRDGAKIAMNYSEFEYCWDTFVDRQGRPFQPWDGLDEHSQALSGRLRAI (斜
体:核酸编辑结构域;下划线:细胞质定位信号) (SEQ ID NO:44)

[0349] 黑猩猩APOBEC-3G:

MKPHFRNPVERMYQDTFSDNFYNRPILSHRNTVWLCYEVKTKGPSRPPLDAKIFRGQV
YSKLKYHPEMRFFHWFSKWRKLHRDQEYEVTWYISWSPCTKCTRVDATFLAEDPKVTLT
[0350] FVARLYYFWDPDYQEALRSLCQKRDGPRATMKIMNYDEFQHCWSKFVYSQRELFEFW
NNLPKYIILLHIMLGEILRHSMPTFTSNFNELWVRGRHETYLCYEVERLHNDTWV
LLNQRRGFLCNQAPHKHGFLEGRHAELCFDVIPIFWKLDLHQDYRVTCFTSWSPCFSCA
QEMAKFISNNKHVSLCIFAARIYDDQGRYQEGRLALHRDGAKIAMNYSEFEYCWDTFVDRQGRPFQPWDGLDEHSQALSGRLRAI (斜
体:核酸编辑结构域;下划线:细胞质定位信号) (SEQ ID NO:45)

[0351] (斜体:核酸编辑结构域;下划线:细胞质定位信号)

[0352] 绿猴APOBEC-3G:

- [0353] MNPQIRNMVEQMEPDIFVYYFNNRPILSGRNTVWLCYEVKTKDPSGPPLDANIFQGKL
YPEAKDHPKEMKFLHWFRKWRQLHRDQEYEV TWYVSWSPCTRCANSVATFLAEDPKVTLT
IFVARLYYFWKPDYQQALRILCQERGGPHATMKIMNYNEFQHCWNEFVDGQGKPKFP
RKNLPKH YTL LHATLGELLRHVMDPGTFTSNFNNKPWVSGQRETYLCYKVERSHNDT
WVLLNQHRGFLRNQAPDRHGFPKGRHAELCFLDLIPFWKLDDQYRVTCFTSWSPCFSC
AQKMAKFISNNKHVSLCIFAARIYDDQGRCQEGLRTLHRDGA KIAVMNYSEFEYCWD
TFVDRQGRPFQPWDGLDEHSQALSGRLRAI (SEQ ID NO: 46)
- [0354] (斜体:核酸编辑结构域;下划线:细胞质定位信号)
- [0355] 人类APOBEC-3G:
MKPHFRNTVERMYRDTFSYNFYNNRPILSRRNTVWLCYEVKTKGPSRPPLDAKIFRGQV
YSELKYHPEMRFFHWFSKWRKLHRDQEYEV TWYISWSPCTKCTRDMATFLAEDPKVTLT
IFVARLYYFWDPDYQEALRSLCQKRDGPRATMKIMNYDEFQHCWSKFVYSQRELFEF
- [0356] WNNLPKYIILLHIMLGEILRHSM DPPTFTFNFNNEPWVRGRHET YLCYEVERMHNDT
WVLLNQRRGFLCNQAPHKHGFLEGRHAELCFLDVIPFWKLDDQDYRVTCFTSWSPCFSC
CAQEMAKFISNNKHVSLCIFTARIYDDQGRCQEGLRTLA EAGAKISIMTYSEFKHCWDT
FVDHQGCPFQPWDGLDEHSQDL SGRLRAILQNQEN (SEQ ID NO: 47)
- [0357] (斜体:核酸编辑结构域;下划线:细胞质定位信号)
- [0358] 人类APOBEC-3F:
MKPHFRNTVERMYRDTFSYNFYNNRPILSRRNTVWLCYEVKTKGPSRPRLDAKIFRGQV
YSQPEHHAEMCFLSWFCGNQLPAYKCFQITWFWVSWTPCPDCVAKLA EFLAEHPNVTLTIS
AARLYYYWERDYRRALCRLSQAGARVKIMDDEEFAYCWENFVYSEGQPFMPWYKFD
- [0359] DNYAFLHRTLKEILRNPMEAMYPHIFYHFKNLRKAYGRNESWLCFTMEVVKHHSPVS
WKRGVFRNQVDPETHCHAERCFLSWFCDDILSPNTNYEVTWYTSWSPCECAGEVAEFL
ARHSNVNLTIFTARLYYFWDTDYQEGLRSLSQEGASVEIMGYKDFKYCWENFVYND
EPFKPWKGLKYNFLFLDSKLQEILE (SEQ ID NO: 48)
- [0360] (斜体:核酸编辑结构域)
- [0361] 人类APOBEC-3B:
MNPQIRNPMERMYRDTFYDNFENEPILYGRSYTWLCYEVKIKRGRSNLLWDTGVFRG
- [0362] QVYFKPQYHAEMCFLSWFCGNQLPAYKCFQITWFWVSWTPCPDCVAKLA EFLSEHPNVTL
TISAARLYYYWERDYRRALCRLSQAGARVTIMDYEEFAYCWENFVYNEGQQFMPWY
KFDENYAFLHRTLKEILRYLMDPDFTFTFNFNNDPLVLRRTYLCYEVERLDNGTWVL
MDQHMGFLCNEAKNLLCGFYGRHAELRFLDLVPSLQLDPAQIYRVTWFIWSPCFSWGC
- [0363] AGEVRAFLQENTHVRLRIFAARIYDYDPLYKEALQMLRDAGA QVSIMTYDEFEYCWD
TFVYRQGC PFQPWDGLEEHSQALSGRLRAILQNQGN (SEQ ID NO: 49)
- [0364] (斜体:核酸编辑结构域)
- [0365] 大鼠APOBEC-3B:
- [0366] MQPQGLGPNAGMPVCLGCSHRRPYSPIRNPLKKLYQQTFYHFKNVRYAWGRKNNFLCYEVNGMDCA
LPVPLRQGVFRKQGHIAELCFIYWFHDKVLRVLSPEEFKVTWYMSWSPCSKCAEQVARFLAAHRNLSLAIFSSR
LYYYLRNPNYQQKLCRLIQEGVHVAAMDLP EFKKCWKNKFVDNDGQPF RPWMRLRINFSFYDCKLQE IFSRMNLLRE

DVFYLFQFNNSHRVKPVQNRYRRKSYLCYQLERANGQEPLKGYLLYKKGEQHV EILFLEKMRSMELSQVRITCYLT
WSPCPNCARQLAAFKKDHDPDLILRIYTSRLYFWRKKFQKGLCTLWRSGLHVDVMDLPQFADCWTNFVNPQRPF
NELEKNSWRIQRRLRRIKESWGL (SEQ ID NO: 50)

[0367] 牛APOBEC-3B:

[0368] MDGWEVAFRSGTVLKAGVLGVSMTEGWAGSGHPGQACVWTPGTRNTMNLREVLFKQQFGNQPRVPA
PYYRRKTYLCYQLKQRNDLTDRGCFRNNKQRHAERFIDKINSLDNPSQSYKIICYITWSPCPNCANELVNFITR
NNHLKLEIFASRLYFHWIKSFKMGLQDLQAGISVAVMTHTEFEDCWEQFVDNQSRPFQPWDKLEQYSASIRRLQ
RILTAPI (SEQ ID NO: 51)

[0369] 黑猩猩APOBEC-3B:

[0370] MNPQIRNPMEMYQRTFYNFENEPILYGRSYTWLCYEVKIRRGHSNLLWDTGVFRGQMYSQPEHHA
MCFLSWFCGNQLSAYKCFQITWFSWTPCPDCVAKLAKFLAHPNVTLTISAARLYYWERDYRRALCRLSQAGAR
VKIMDDEEFAYCWENFVYNQGPMPWYKFDDNYAFLHRTLKEIIRHLMPTFTFNFNDPLVLRHQTYLCYEV
ERLDNGTWVLMQHMGLCNEAKNLLCGFYGRHAELRFLDLVPSLQLDPAQIYRVTFISWSPCFSWGCAGQVRAF
LQENTHVRLRIFAARIYDYDPLYKEALQMLRDAGAQVSIMTYDEFEYCWDTFVYRQGC PFQPWDGLEHSQALSGR
LRAILQVRASSLCMVPHRPPPPQSPGPCLPLCSEPPLGSLPTGRPAPSLPFLLTASFSPPPASLPPLPSLSLS
PGHLPVPSFHSALTSCSIQPPCSSRIRETGWASVSKEGRDLG (SEQ ID NO: 52)

[0371] 人类APOBEC-3C:

MNPQIRNPMKAMYPGTFYFQFKNLWEANDRNETWLCFTVEGIKRRSVVSWKTGVFR

[0372] NQVDSETH*CHAERCFLSWFCDDILSPNTKYQVTWYTSWSPCPDC*AGEVAEFLARHSNVN
LTIFTARLYYFQYPCYQEGRLSLSQEGVAVEIMDYEDFKYCWENFVYNDNEPFKPWKG

[0373] LKTNFRLKRLRESLQ (SEQ ID NO: 53)

[0374] (斜体:核酸编辑结构域)

[0375] 大猩猩APOBEC-3C

MNPQIRNPMKAMYPGTFYFQFKNLWEANDRNETWLCFTVEGIKRRSVVSWKTGVFR

[0376] NQVDSETH*CHAERCFLSWECDDILSPNTNYQVTWYTSWSPCPEC*AGEVAEFLARHSNVN
TIFTARLYYFQD TDYQEGRLSLSQEGVAVKIMDYKDFKYCWENFVYNDDEPFKPWKG
LKYNFRFLKRLQEILE (SEQ ID NO: 54)

[0377] (斜体:核酸编辑结构域)

[0378] 人类APOBEC-3A:

MEASPASGRHLMDPHFTSNFNNGIGRHKTYLCYEVERLDNGTSVKMDQHRGFLHN

[0379] QAKNLLCGFYGRHAELRFLDLVPSLQLDPAQIYRVTFISWSPCFSWGCAGEVRAFLQEN
THVRLRIFAARIYDYDPLYKEALQMLRDAGAQVSIMTYDEFKHCWDTFVDHQGCPFQ
PWDGLDEHSQALSGRLRAILQNQGN (SEQ ID NO: 55)

[0380] (斜体:核酸编辑结构域)

[0381] 恒河猴APOBEC-3A:

MDGSPASRPHLMDPNTFTFNFNDLSVRGRHQTYLCYEVERLDNGTWVPMDERRGF

[0382] LCNKAKNVPCGDYGHVELRFLCEVPSWQLDPAQTYRVTFISWSPCFRRGCAGQVRVF
LQENKHVRLRIFAARIYDYDPLYQEALRTL RDAGAQVSIMTYEEFKHCWDTFVDRQGR
PFQPWDGLDEHSQALSGRLRAILQNQGN (SEQ ID NO: 56)

- [0383] (斜体:核酸编辑结构域)
- [0384] 牛APOBEC-3A:
MDEYTFTENFNNQGWPSKTYLCYEMERLDGDATIPLDEYKGFVRNKGLDQPEKPCHA
ELYFLGKIHSWNLDNRNQHYRLTCFISWSPCYDCAQKLTTFLENHHISLHILASRIYTHNR
- [0385] FGCHQSGLCCELQAAGARITIMTFEDFKHCWETFVDHKGKPFQPWEGLNVKSQALCTE
LQAILKTQQN (SEQ ID NO: 57)
- [0386] (斜体:核酸编辑结构域)
- [0387] 人类APOBEC-3H:
MALLTAETFRLQFNNKRRRLRPYYPRKALLCYQLTPQNGSTPTRGYFENKKKCHAEIC
FINEIKSMGLDETQCYQVTCYLTWSPCSSCAWELVDFIKAHDHLNLGIFASRLYYHWCKP
- [0388] QQKGLRLLCGSQVPVEVMGFPKFAADCWENFVDHEKPLSFNPYKMLEELDKNSRAIKR
RLERIKIPGVRAQGRYMDILCDAEV (SEQ ID NO: 58)
- [0389] (斜体:核酸编辑结构域)
- [0390] 恒河猴APOBEC-3H:
MALLTAKTFSLQFNNKRRVKNPYYPRKALLCYQLTPQNGSTPTRGHLKNKKKDHAIEIR
FINKIKSMGLDETQCYQVTCYLTWSPCPCAGELVDFIKAHRHLNLRIFASRLYYHWRP
- [0391] NYQEGLLLLCGSQVPVEVMGLPEFTDCWENFVDHKEPPSFNPSEKLEELDKNSQAIKR
RLERIKSRSDVLENGLRSLQLGPVTPSSSIRNSR (SEQ ID NO: 59)
- [0392] 人类APOBEC-3D:
MNPQIRNPMERMYRDTFYDNFENEPILYGRSYTWLCYEVKIKRGRSNLLWDTGVFRGP
VLPKRQSNHRQEYFRFENHAEMCFLSWFCGNRLPANRRFQITWFWVSWNPCLPCVVVKVT
KFLAEHPNVTLTISAARLYYYRDRDWRWVLLRLHKAGARVKIMDYEDFAYCWENFVC
- [0393] NEGQPFMPWYKFDDNYASLHRTLKEILRNPMEMYPHIFYFHFNLLKACGRNESWL
CFTMEVTKHHSVFRKRGVFRNQVDPETHCHAERCFLSWFCDDILSPNTNYEVTWYTSW
SPCPECAGEVAEFLARHSNVNLTIFTARLCYFWDTDYQEGLCSLSQEGASVKIMGYKD
FVSCWKNFVYSDDPEFPKWKGLQTNFRLLKRRRLREILQ (SEQ ID NO: 60)
- [0394] (斜体:核酸编辑结构域)
- [0395] 人类APOBEC-1:
- [0396] MTSEKGPSTGDPTLRRRIEPWEFDVFYDPRELRKEACLLYEIKWGMSRKIWRSSGKNTTNHVEVNF
IKFTSERDFHPSMCSITWFLSWSPCWECSQAIREFLSRHPGVTLVIIYVARLFWHMDQQNRQGLRDLVNSGVTIQIM
RASEYYHCWRNFVNYPPGDEAHWPQYPPLWMMLYALELHICIILSLPPCLKISRRWQNHLTFFRLHLQNCHYQTIPP
HILLATGLIHPSVAWR (SEQ ID NO: 61)
- [0397] 小鼠APOBEC-1:
- [0398] MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSVWRHTSQNTSNHVEVNFLE
KFTTERYFRPNTRCSITWFLSWSPCGECSRAITEFLSRHPYVTLFIYIARLYHHTDQNRNQGLRDLISSGVTIQIM
TEQEYCYCWRNFVNYPPSNEAYWPRYPHLWKLYVLELYCIIILGLPPCLKILRRKQPQLTFFTITLQTCHYQRIPP
HLLWATGLK (SEQ ID NO: 62)
- [0399] 大鼠APOBEC-1:
- [0400] MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIE

KFTTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIM
TEQESGYCWRNFVNYSNEAHWPYPHLLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPP
HILWATGLK (SEQ ID NO:63)

[0401] 人类APOBEC-2:

[0402] MAQKEEAAVATEAASQNGEDLENLDDPEKLKELIELPPFEIVTGERLPANFFKFQFRNVEYSSGRNKT
FLCYVVEAQGGQVQASRGYLEDEHAAHAEEAFFNTILPAFDPALRYNVTWYVSSSPCAACADRIKTLSTKTN
LRLILVGRFLMWEEPEIQAALKKLKEAGCKLRIMKPQDFEYVWQNFVEQEEGESKAFQPWEDIQENFLYYEEKLA
DILK (SEQ ID NO:64)

[0403] 小鼠APOBEC-2:

[0404] MAQKEEAAEAAAPASQNGDDLENLEDPEKLKELIDLPPFEIVTGVRLPVNFFKFQFRNVEYSSGRNKT
FLCYVVEVQSKGGQAQATQGYLEDEHAGAHAAEEAFFNTILPAFDPALKYNVTWYVSSSPCAACADRILKTLSTKTN
LRLILVSRLFMWEEPEVQAALKKLKEAGCKLRIMKPQDFEYIWQNFVEQEEGESKAFEPWEDIQENFLYYEEKLA
DILK (SEQ ID NO:65)

[0405] 大鼠APOBEC-2:

[0406] MAQKEEAAEAAAPASQNGDDLENLEDPEKLKELIDLPPFEIVTGVRLPVNFFKFQFRNVEYSSGRNKT
FLCYVVEAQSKGGQVQATQGYLEDEHAGAHAAEEAFFNTILPAFDPALKYNVTWYVSSSPCAACADRILKTLSTKTN
LRLILVSRLFMWEEPEVQAALKKLKEAGCKLRIMKPQDFEYLWQNFVEQEEGESKAFEPWEDIQENFLYYEEKLA
DILK (SEQ ID NO:66)

[0407] 牛APOBEC-2:

[0408] MAQKEEAAAAAEPASQNGEEVENLEDPEKLKELIELPPFEIVTGERLPAHYFKFQFRNVEYSSGRNKT
FLCYVVEAQSKGGQVQASRGYLEDEHATNHAAEEAFFNSIMPTFDPALRYMVTWYVSSSPCAACADRIVKTLNKTKN
LRLILVGRFLMWEEPEIQAALRKLKEAGCRLRIMKPQDFEYIWQNFVEQEEGESKAFEPWEDIQENFLYYEEKLA
DILK (SEQ ID NO:67)

[0409] 海七鳃鳗CDA1 (pmCDA1):

[0410] MTDAEYVRIHEKLDIYTFKKQFFNNKKSVSRCYVLFELKRRGERRACFWGYAVNKPQSGTERGIHAE
IFSIRKVEEYLRDNPQGFTINWYSSWSPCADCAEKILEWYNQELRGNGHTLKIWACKLYYEKNARNQIGLWNLRDN
GVGLNMVSEHYQCCRKIFIQSSHNQLNENRWLEKTLKRAEKRRSELSFMIQVKILHTTKSPAV (SEQ ID NO:
68)

[0411] 人类APOBEC3G D316R D317R:

[0412] MKPHFRNTVERMYRDTFSYNFYNRPILSRRNTVWLCYEVKTKGPSRPPLDAKIFRGQVYSELKYHP
RFFHWFSKWRKLHRDQEYEVWYISWSPCTKCTRDMAFLAEDPKVTLTIFVARLYYFWDPDYQEALRSLCQKRDG
PRATMKFNYDEFQHCWSKFVYSQRELFEPPWNNLPKYIILLHFMGEILRHSMPPPTFTFNFNNEPWVRGRHET
YLCYEVRMHNDTWVLLNQRRGFLCNQAPHKHGFLEGRHAELCFLDVIPFWKLDLDQDYRVTCFTSWSPCF
SQAQEMAKFISKHVSCLIFTARIYRRQGRCQEGLRTLAEAGAKISFTYSEFKHCWDTFVDHQGCPFPWDGL
DEHSQDLSGRLRAILQNQEN (SEQ ID NO:69)

[0413] 人类APOBEC3G链A:

[0414] MDPPTFTFNFNNEPWVRGRHETLYCYEVRMHNDTWVLLNQRRGFLCNQAPHKHGFLEGRHAELCFLDV
IPFWKLDLDQDYRVTCFTSWSPCFSQAQEMAKFISKHVSCLIFTARIYDDQGRCQEGLRTLAEAGAKISFTYSE
FKHCWDTFVDHQGCPFPWDGLDEHSQDLSGRLRAILQ (SEQ ID NO:70)

[0415] 人类APOBEC3G链A D120R D121R:

[0416] MDPPTFTFNFNNEPWVRGRHETYLCEVERMHNDTWVLLNQRRGFLCNQAPHKHGFLEGRHAELCFLD
VIPFWKLDLDQDYRVTCFTSWSPCFSCAQEMAKFISKKNHVS LCIFTARIYRRQGRCQEGLRTLAEAGAKISFMTY
SEFKHCWDTFVDHQGCPFPWDGLDEHSQDLSGRLRAILQ (SEQ ID NO:71)

[0417] hAPOBEC-4 (智人):

[0418] MEPIYEEYLANHGTIVKPYWLSFSLDCSNCPYHIRTGEEARVSLTEFCQIFGFPYGTTFPQTKHLTF
YELKTSSGSLVQKGHASSCTGNYIHPESMLFEMNGYLD SAIYNND S I R H I I L Y S N N S P C N E A N H C C I S K M Y N F L I T
Y P G I T L S I Y F S Q L Y H T E M D F P A S A W N R E A L R S L A S L W P R V L S P I S G G I W H S V L H S F I S G V S G S H V F Q P I L T G R A L
A D R H N A Y E I N A I T G V K P Y F T D V L L Q T K R N P N T K A Q E A L E S Y P L N N A F P G Q F F Q M P S G Q L Q P N L P P D L R A P V V F V L V
P L R D L P P M H M G Q N P N K P R N I V R H L N M P Q M S F Q E T K D L G R L P T G R S V E I V E I T E Q F A S S K E A D E K K K K K G K K (SEQ
ID NO:72)

[0419] mAPOBEC-4 (家鼠):

[0420] MDSLLMKQKKFLYHFKNVRWAKGRHETYL CYVVKRRDSATSCSLDFGHLRNKSGCHVELLFLRYISDW
DLDPGRCYRVTWFTSWSPCYDCARHVAEFLRWNPNSLRIFTARLYFCEDRKAEP EGLRRLHRAGVQIGIMTFKDY
FYCWNTFVENRERTFKAW EGLHENS VRLTRQLRRILLPLYEVD DL R D A F R M L G F (SEQ ID NO:73)

[0421] rAPOBEC-4 (挪威鼠):

[0422] MEPLYEEYLTHSGTIVKPYWLSVSLNCTNCPYHIRTGEEARVPYTEFHQTFGFPWSTYPQTKHLTFY
ELRSSSGNLIQKGLASNCTGSHTHPESMLFERDGYLD S L I F H D S N I R H I I L Y S N N S P C D E A N H C C I S K M Y N F L M N Y
P E V T L S V F F S Q L Y H T E N Q F P T S A W N R E A L R G L A S L W P Q V T L S A I S G G I W Q S I L E T F V S G I S E G L T A V R P F T A G R T L
T D R Y N A Y E I N C I T E V K P Y F T D A L H S W Q K E N Q D Q K V W A A S E N Q P L H N T T P A Q W Q P D M S Q D C R T P A V F M L V P Y R D L P P
I H V N P S P Q K P R T V V R H L N T L Q L S A S K V K A L R K S P S G R P V K K E E A R K G S T R S Q E A N E T N K S K W K K Q T L F I K S N I C H L
L E R E Q K K I G I L S S W S V (SEQ ID NO:74)

[0423] mfAPOBEC-4 (食蟹猴):

[0424] MEPTYEEYLANHGTIVKPYWLSFSLDCSNCPYHIRTGEEARVSLTEFCQIFGFPYGTTPQTKHLTF
YELKTSSGSLVQKGHASSCTGNYIHPESMLFEMNGYLD S A I Y N N D S I R H I I L Y C N N S P C N E A N H C C I S K V Y N F L I T
Y P G I T L S I Y F S Q L Y H T E M D F P A S A W N R E A L R S L A S L W P R V L S P I S G G I W H S V L H S F V S G V S G S H V F Q P I L T G R A L
T D R Y N A Y E I N A I T G V K P F F T D V L L H T K R N P N T K A Q M A L E S Y P L N N A F P G Q S F Q M T S G I P P D L R A P V V F V L L P L R D L
P P M H M G Q D P N K P R N I I R H L N M P Q M S F Q E T K D L E R L P T R R S V E T V E I T E R F A S S K Q A E E K T K K K K G K K (SEQ ID
NO:75)

[0425] pmCDA-1 (海七鳃鳗):

[0426] MAGYECVRVSEKLD F D T F E F Q F E N L H Y A T E R H R T Y V I F D V K P Q S A G G R S R R L W G Y I I N N P N V C H A E L I
L M S M I D R H L E S N P G V Y A M T W Y M S W S P C A N C S S K L N P W L K N L L E E Q G H T L T M H F S R I Y D R D R E G D H R G L R G L K H V S N
S F R M G V V G R A E V K E C L A E Y V E A S R R T L T W L D T T E S M A A K M R R K L F C I L V R C A G M R E S G I P L H L F T L Q T P L L S G R V V
W W R V (SEQ ID NO:76)

[0427] pmCDA-2 (海七鳃鳗):

[0428] MELREVVD CALAS CVRHEPLSRVAFLRCFAAPSQKPRGT V I L F Y V E G A G R G V T G G H A V N Y N K Q G T S I H
A E V L L L S A V R A A L L R R R R C E D G E E A T R G C T L H C Y S T Y S P C R D C V E Y I Q E F G A S T G V R V V I H C C R L Y E L D V N R R R S E
A E G V L R S L S R L G R D F R L M G P R D A I A L L L G G R L A N T A D G E S G A S G N A V W T E T N V V E P L V D M T G F G D E D L H A Q V Q R N K
Q I R E A Y A N Y A S A V S L M L G E L H V D P D K F P F L A E F L A Q T S V E P S G T P R E T R G R P R G A S S R G P E I G R Q R P A D F E R A L G A

YGLFLHPRIVSREADREEIKRDLIVVMRKHNYQGP (SEQ ID NO:77)

[0429] pmCDA-5 (海七鳃鳗) :

[0430] MAGDENVRVSEKLDFTDFEFQFENLHYATERHRTYVIFDVKPKSAGGRSRRLWGYI INNPVCHAELI
LMSMIDRHLESNPGVYAMTWYMSWSPCANCSSKLNPLKNLLEEQGHITLMMHFSRIYDRDREGDHRGLRGLKHVSNSFR
MGVVGRAEVKE CLAEYVEASRRTLTLWLDTTESMAAKMRRKLCILVRCAGMRESGMPLHLFT (SEQ ID NO:78)

[0431] yCD (酿酒酵母) :

[0432] MVTGGMASKWDQKGMDIAYEEAALGYKEGGVPIGGCLINNKGSVLGRGHNMRFQKGSATLHGEISTL
ENCGRLEGKVYKDTTLYTTLSPCDMCTGAIIMYGIPRCVVGENVNFKSKGEKYLQTRGHEVVVDDERCKKIMKQF
IDERPDQWFEDIGE (SEQ ID NO:79) rAPOBEC-1 (8177-186) :

[0433] MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELKTCCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIE
KFTTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTFLFIYIARLYHHADPRNRQGLRDLIS SGVTIQI
MTEQESGYCWRNRFVNYSNEAHWPYPHLLWVRGLPPCLNLRKQPKLTFFITIALQSCHYQRLPPHILWATGLK
(SEQ ID NO:80)

[0434] rAPOBEC-1 (8202-213) :

[0435] MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELKTCCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIE
KFTTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTFLFIYIARLYHHADPRNRQGLRDLISSGVTIQIM
TEQESGYCWRNRFVNYSNEAHWPYPHLLWVRGLVLELYCIILGLPPCLNLRKQPKQHYQRLPPHILWATGLK
(SEQ ID NO:81)

[0436] 小鼠APOBEC-3:

MGPFCGLGCSHRKCYSPIRNLISQETFKFHFKNLGYAKGRKDTFLCYEVTRKDCDSPVSL
HHGVFKNKDNIHAEICFLYWFHDKVLKVLSPREEFKITWYMSWSPCFECAEQIVRFLATH
HNLSLDIFSSRLYNVQDPETQQNLCLRVQEGAQVAAMDLYEFKKCWKKFVDNGGRRF
RPWKRLLTNFRYQDSKLQEILRPCYIPVPSSSSSTLSNICLTGLPETRFCVEGRRMDPLS
[0437] EEEFYSQFYNQVRVKHLCYYHRMKPYLCYQLEQFNGQAPLKGCLLSEKKGQHAEILFLD
KIRSMELSQVTITCYLTWSPCPNCAWQLAAFKRDRPDILHIYTSRLYFHWKRPFQKGLC
SLWQSGILVDVMDLPQFTDCWTNMFVNPKRPFWPWKGLEIISRRTQRRRLRIKESWGLQ
DLVNDFGNLQLGPPMS (SEQ ID NO: 82)

[0438] (斜体:核酸编辑结构域)

[0439] 在一些实施例中,腺苷脱氨酶可包含腺苷脱氨酶ADAR (例如,ADAR1或ADAR2)的全部或一部分。在另一实施例中,腺苷脱氨酶可包含腺苷脱氨酶ADAT的全部或一部分。在一些实施例中,腺苷脱氨酶可包含来自大肠杆菌 (EcTadA) 的ADAT的全部或一部分,其包含以下突变中的一者或多者:D108N、A106V、D147Y、E155V、L84F、H123Y、I157F,或另一种腺苷脱氨酶中的对应突变。腺苷脱氨酶可以源自任何合适的生物体 (例如,大肠杆菌)。在一些实施例中,腺苷脱氨酶来自大肠杆菌、金黄色葡萄球菌、伤寒沙门氏菌、腐败希瓦氏菌、流感嗜血杆菌、新月柄杆菌或枯草芽孢杆菌。在一些实施例中,腺苷脱氨酶来自大肠杆菌。在一些实施例中,腺嘌呤脱氨酶是天然存在的腺苷脱氨酶,其包括对应于本文提供的任何突变 (例如,ecTadA中的突变) 的一个或多个突变。任何同源蛋白质中的相应残基可以通过例如序列比对和同源残基测定来鉴定。任何天然存在的腺苷脱氨酶 (例如,与ecTadA具有同源性) 中对应于本文描述的任何突变 (例如,ecTadA中鉴定的任何突变) 的突变可以相应地产生。在特

定实施例中,TadA是PCT/US2017/045381(WO 2018/027078)中描述的 TadA 中的任一种,该文献通过引用以其整体并入本文。通过多轮进化和选择鉴定出对单链DNA具有所需腺苷脱氨酶活性的突变(例如,TadA*7.10=来自第七轮进化的变体10),如表3中所示。

[0440] 表3.TadA变体的基因型

TadA	23	26	36	37	48	49	51	72	84	87	105	108	123	125	142	145	147	152	155	156	157	16
0.1	W	R	H	N	P		R	N	L	S	A	D	H	G	A	S	D	R	E	I	K	K
0.2	W	R	H	N	P		R	N	L	S	A	D	H	G	A	S	D	R	E	I	K	K
1.1	W	R	H	N	P		R	N	L	S	A	N	H	G	A	S	D	R	E	I	K	K
1.2	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	D	R	E	I	K	K
2.1	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
2.2	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
2.3	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
2.4	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
2.5	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
2.6	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
2.7	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
2.8	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
2.9	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
2.10	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
2.11	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
[0441] 2.12	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
3.1	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
3.2	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
3.3	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
3.4	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
3.5	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
3.6	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
3.7	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
3.8	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
4.1	W	R	H	N	P		R	N	L	S	V	N	H	G	N	S	Y	R	V	I	K	K
4.2	W	G	H	N	P		R	N	L	S	V	N	H	G	N	S	Y	R	V	I	K	K
4.3	W	R	H	N	P		R	N	F	S	V	N	Y	G	N	S	Y	R	V	F	K	K
5.1	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
5.2	W	R	H	S	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	T
5.3	W	R	L	N	P		L	N	I	S	V	N	Y	G	A	C	Y	R	V	I	N	K
5.4	W	R	H	S	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	T
5.5	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K

[0442]

TadA	23	26	36	37	48	49	51	72	84	87	105	108	123	125	142	145	147	152	155	156	157	16
5.6	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
5.7	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
5.8	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
5.9	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
5.10	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
5.11	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
5.12	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
5.13	W	R	H	N	P		L	D	F	S	V	N	Y	A	A	S	Y	R	V	F	K	K
5.14	W	R	H	N	S		L	N	F	C	V	N	Y	G	A	S	Y	R	V	F	K	K
6.1	W	R	H	N	S		L	N	F	S	V	N	Y	G	N	S	Y	R	V	F	K	K
6.2	W	R	H	N	T	V	L	N	F	S	V	N	Y	G	N	S	Y	R	V	F	N	K
6.3	W	R	L	N	S		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
6.4	W	R	L	N	S		L	N	F	S	V	N	Y	G	N	C	Y	R	V	F	N	K
6.5	W	R	L	N	I	V	L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
6.6	W	R	L	N	T	V	L	N	F	S	V	N	Y	G	N	C	Y	R	V	F	N	K
7.1	W	R	L	N	A		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
7.2	W	R	L	N	A		L	N	F	S	V	N	Y	G	N	C	Y	R	V	F	N	K
7.3	I	R	L	N	A		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
7.4	R	R	L	N	A		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
7.5	W	R	L	N	A		L	N	F	S	V	N	Y	G	A	C	Y	H	V	F	N	K
7.6	W	R	L	N	A		L	N	I	S	V	N	Y	G	A	C	Y	P	V	I	N	K
7.7	L	R	L	N	A		L	N	F	S	V	N	Y	G	A	C	Y	P	V	F	N	K
7.8	I	R	L	N	A		L	N	F	S	V	N	Y	G	N	C	Y	R	V	F	N	K
7.9	L	R	L	N	A		L	N	F	S	V	N	Y	G	N	C	Y	P	V	F	N	K
7.10	R	R	L	N	A		L	N	F	S	V	N	Y	G	A	C	Y	P	V	F	N	K

[0443] 在一些实施例中,TadA作为单体或二聚体(例如,野生型大肠杆菌TadA和工程化TadA变体的异二聚体)提供。在一些实施例中,腺苷脱氨酶是第八代TadA*8变体,如下表4中所示。

[0444] 表4:TadA8*腺苷脱氨酶变体

[0445]

腺苷脱氨酶	腺苷脱氨酶描述
TadA*8.1	单体_TadA*7.10 + Y147T
TadA*8.2	单体_TadA*7.10 + Y147R
TadA*8.3	单体_TadA*7.10 + Q154S
TadA*8.4	单体_TadA*7.10 + Y123H
TadA*8.5	单体_TadA*7.10 + V82S
TadA*8.6	单体_TadA*7.10 + T166R
TadA*8.7	单体_TadA*7.10 + Q154R
TadA*8.8	单体_TadA*7.10 + Y147R_Q154R_Y123H
TadA*8.9	单体_TadA*7.10 + Y147R_Q154R_I76Y
TadA*8.10	单体_TadA*7.10 + Y147R_Q154R_T166R
TadA*8.11	单体_TadA*7.10 + Y147T_Q154R
TadA*8.12	单体_TadA*7.10 + Y147T_Q154S
TadA*8.13	单体_TadA*7.10 + H123H_Y147R_Q154R_I76Y
TadA*8.14	异二聚体_(WT) + (TadA*7.10 + Y147T)
TadA*8.15	异二聚体_(WT) + (TadA*7.10 + Y147R)
TadA*8.16	异二聚体_(WT) + (TadA*7.10 + Q154S)
TadA*8.17	异二聚体_(WT) + (TadA*7.10 + Y123H)
TadA*8.18	异二聚体_(WT) + (TadA*7.10 + V82S)
TadA*8.19	异二聚体_(WT) + (TadA*7.10 + T166R)
TadA*8.20	异二聚体_(WT) + (TadA*7.10 + Q154R)
TadA*8.21	异二聚体_(WT) + (TadA*7.10 + Y147R_Q154R_Y123H)
TadA*8.22	异二聚体_(WT) + (TadA*7.10 + Y147R_Q154R_I76Y)
TadA*8.23	异二聚体_(WT) + (TadA*7.10 + Y147R_Q154R_T166R)
TadA*8.24	异二聚体_(WT) + (TadA*7.10 + Y147T_Q154R)
TadA*8.25	异二聚体_(WT) + (TadA*7.10 + Y147T_Q154S)

	腺苷脱氨酶	腺苷脱氨酶描述
[0446]	TadA*8.26	异二聚体_(WT) + (TadA*7.10 + H123H_Y147T_Q154R_I76Y)

[0447] 在一些实施例中,腺苷脱氨酶是第九代TadA*9变体,其在选自以下的氨基酸位置处含有改变:TadA变体的21、23、25、38、51、54、70、71、72、72、94、124、133、138、139、146和158,如以下参考序列中所示:

	10	20	30	40	50
	MSEVEFSHEY	WMRHALTLAK	<u>RARDE</u> REVPV	GAVLVLN <u>NRV</u>	IGEGWNRAIG
	60	70	80	90	100
[0448]	<u>LHDP</u> TAHAEI	MALRQGGLV <u>M</u>	<u>QNY</u> RLIDATL	YVTFEPCVMC	AGAM <u>MI</u> HSRIG
	110	120	130	140	150
	RVVFGVRNAK	TGAAGSLMDV	LHY <u>P</u> GMNHRV	EIT <u>E</u> GILAD <u>E</u>	CAALL <u>C</u> YFFR
	160				
	MPRQVFNA <u>A</u> QK KAQSSTD (SEQ ID NO: 83)				

[0449] 在一个实施例中,腺苷脱氨酶变体在选自以下的两个或更多个氨基酸位置处含有改变:上述TadA参考序列的21、23、25、38、51、54、70、71、72、94、124、133、138、139、146和158。在另一实施例中,腺苷脱氨酶变体含有选自以下的一个或多个(例如,2、3、4个)改变:SEQ ID NO.1的R21N、R23H、E25F、N38G、L51W、P54C、M70V、Q71M、N72K、Y73S、M94V、P124W、T133K、D139L、D139M、C146R和A158K。在其他实施例中,腺苷脱氨酶变体进一步含有以下改变中的一者或多者:Y147T、Y147R、Q154S、Y123H和Q154R。在另外其他实施例中,腺苷脱氨酶变体含有相对于上述TadA参考序列的改变的组合,该改变的组合选自以下:E25F+V82S+Y123H、T133K+Y147R+Q154R;E25F+V82S+Y123H+Y147R+Q154R;L51W+V82S+Y123H+C146R+Y147R+Q154R;Y73S+V82S+Y123H+Y147R+Q154R;P54C+V82S+Y123H+Y147R+Q154R;N38G+V82T+Y123H+Y147R+Q154R;N72K+V82S+Y123H+D139L+Y147R+Q154R;E25F+V82S+Y123H+D139M+Y147R+Q154R;Q71M+V82S+Y123H+Y147R+Q154R;E25F+V82S+Y123H+T133K+Y147R+Q154R;E25F+V82S+Y123H+Y147R+Q154R;V82S+Y123H+P124W+Y147R+Q154R;L51W+V82S+Y123H+C146R+Y147R+Q154R;P54C+V82S+Y123H+Y147R+Q154R;Y73S+V82S+Y123H+Y147R+Q154R;N38G+V82T+Y123H+Y147R+Q154R;R23H+V82S+Y123H+Y147R+Q154R;R21N+V82S+Y123H+Y147R+Q154R;V82S+Y123H+Y147R+Q154R+A158K;N72K+V82S+Y123H+D139L+Y147R+Q154R;E25F+V82S+Y123H+D139M+Y147R+Q154R;M70V+V82S+M94V+Y123H+Y147R+Q154R;Q71M+V82S+Y123H+Y147R+Q154R;E25F+I76Y+V82S+Y123H+Y147R+Q154R;I76Y+V82T+Y123H+Y147R+Q154R;N38G+I76Y+V82S+Y123H+Y147R+Q154R;R23H+I76Y+V82S+Y123H+Y147R+Q154R;P54C+I76Y+V82S+Y123H+Y147R+Q154R;R21N+I76Y+V82S+Y123H+Y147R+Q154R;I76Y+V82S+Y123H+D138M+Y147R+Q154R;Y72S+I76Y+V82S+Y123H+Y147R+Q154R;E25F+I76Y+V82S+Y123H+Y147R+Q154R;I76Y+V82T+Y123H+Y147R+Q154R;N38G+I76Y+V82S+Y123H+Y147R+Q154R;R23H+I76Y+V82S+Y123H+Y147R+Q154R;P54C+I76Y+V82S+Y123H+Y147R+Q154R;R21N+I76Y+V82S+Y123H+Y147R+Q154R;I76Y+V82S+Y123H+D138M+Y147R+Q154R;Y72S+I76Y+V82S+Y123H+Y147R+Q154R;和V82S+Q154R;N72K_V82S+Y123H+Y147R+Q154R;Q71M_V82S+Y123H+Y147R+Q154R;

V82S+Y123H+T133K+Y147R+Q154R;V82S+Y123H+T133K+Y147R+Q154R+A158K;M70V+Q71M+N72K+V82S+Y123H+Y147R+Q154R;N72K_V82S+Y123H+Y147R+Q154R;Q71M_V82S+Y123H+Y147R+Q154R;M70V+V82S+M94V+Y123H+Y147R+Q154R;V82S+Y123H+T133K+Y147R+Q154R;V82S+Y123H+T133K+Y147R+Q154R+A158K;和M70V+Q71M+N72K+V82S+Y123H+Y147R+Q154R。在一些实施例中,脱氨酶或其他多肽序列缺乏甲硫氨酸,例如当作为融合蛋白的组分包括时。这可以改变位置的编号。然而,技术人员将理解,此类相应的突变是指相同的突变,例如Y73S和Y72S以及D139M和D138M。

[0450] 在一些实施例中,Cas9与核定位序列融合,这些核定位序列包括SV40大T抗原的NLS、核质蛋白、c-myc、hRNPA1 M9、来自输入蛋白- α 的IBB结构域、肌瘤T蛋白的NLS、人类p53、c-abl IV、流感病毒NS1、肝炎病毒 δ 抗原、小鼠Mx1、人类聚(ADP-核糖)聚合酶、类固醇激素受体(人类)糖皮质激素。

[0451] 在一些实施例中,Cas9蛋白与表位标签融合,这些表位标签包括但不限于血凝素(HA)标签、组氨酸(His)标签、FLAG标签、Myc标签、V5标签、VSV-G标签、SNAP标签、硫氧还蛋白(Trx)标签。

[0452] 在一些实施例中,Cas9与报告基因融合,这些报告基因包括但不限于谷胱甘肽-S-转移酶(GST)、辣根过氧化物酶(HRP)、氯霉素转移酶(CAT)、HcRed、DsRed、青色荧光蛋白、黄色荧光蛋白和蓝色荧光蛋白、绿色荧光蛋白(GFP),包括增强型式或超折叠GFP,以及报告基因的其他经修饰型式。

[0453] 在一些实施例中,工程化Cas9蛋白的血清半衰期通过与异源蛋白融合而增加,这些异源蛋白诸如人类血清白蛋白、转铁蛋白、人类IgG和/或唾液酸化肽,诸如(绒毛膜促性腺激素 β 链的)羧基末端肽(CTP)。

[0454] 在一些实施例中,工程化Cas9蛋白的血清半衰期通过与不稳定结构域融合而缩短,这些不稳定结构域包括但不限于联会蛋白、泛素、FKBP12-L106P和/或二氢叶酸还原酶。

[0455] 提供增加或降低的稳定性的合适的融合配偶体包括但不限于降解决定子序列。本领域普通技术人员容易将降解决定子理解为控制其所属蛋白质的稳定性的氨基酸序列。例如,包含降解决定子序列的蛋白质的稳定性至少部分地由降解决定子序列控制。在一些情况下,合适的降解决定子是组成型的,使得该降解决定子独立于实验控制而对蛋白质稳定性发挥其影响(即,该降解决定子不是药物可诱导的、温度可诱导的等)。在一些情况下,该降解决定子为变体Cas9多肽提供可控的稳定性,使得变体Cas9多肽可以根据所需条件“开启”(即稳定)或“关闭”(即不稳定、降解)。例如,如果降解决定子是温度敏感性降解决定子,则变体Cas9多肽可以在低于阈值温度(例如,42℃、41℃、40℃、39℃、38℃、37℃、36℃、35℃、34℃、33℃、32℃、31℃、30℃等)时发挥作用(即,“开启”,稳定)但在高于阈值温度时不起作用(即“关闭”,降解)。作为另一实例,如果降解决定子是药物可诱导的降解决定子,则药物的存在或不存在可以将蛋白质从“关闭”(即,不稳定)状态切换到“开启”(即,稳定)状态,或反之亦然。例示性药物可诱导的降解决定子衍生自FKBP12蛋白。降解决定子的稳定性由与降解决定子结合的小分子的存在或不存在控制。

[0456] 合适的降解决定子的实例包括但不限于由Shield-1、DHFR、生长素和/或温度控制的那些降解决定子。合适的降解决定子的非限制性实例是本领域已知的(例如,Dohmen等人,Science,1994.263(5151):第1273-1276页;Heat-inducibledegron:a method for

constructing temperature-sensitive mutants;Schoeber等人,Am J Physiol Renal Physiol.2009年1月;296(1):F204-11:Conditional fast expression and function of multimeric TRPV5 channels using Shield-1;Chu等人,Bioorg Med Chem Lett.2008年11月15日;18(22):5941-4:Recent progress with FKBP-derived destabilizing domains;Kanemaki,Pflugers Arch.2012年12月28日:Frontiers of protein expression control with conditional degrons;Yang等人,Mol Cell.2012年11月30日;48(4):487-8:Titivated for destruction:the methyl degron;Barbour等人,Biosci Rep.2013年1月18日;33(1):Characterization of the bipartite degron that regulates ubiquitin-independent degradation of thymidylate synthase;以及Greussing等人,J Vis Exp.2012年11月10日;(69):Monitoring of ubiquitin-proteasome activity in living cells using a Degron(dgn)-destabilized green fluorescent protein (GFP)-based reporter protein;所有这些文献都通过引用以其整体并入本文)。

[0457] 例示性降解决定子序列已在细胞和动物中得到充分表征和测试。因此,将死Cas9与降解决定子序列融合产生“可调节的”和“可诱导的”死Cas9多肽。

[0458] 本文描述的任何融合配偶体可以以任何期望的组合使用。作为说明这一点的一个非限制性实例,Cas9融合蛋白可包含用于检测的YFP序列、用于稳定性的降解决定子序列和用于增加靶DNA转录的转录活化因子序列。此外,dCas9融合蛋白中可使用的融合配偶体的数量是无限的。在一些情况下,Cas9融合蛋白包含一种或多种(例如,两种或更多种、三种或更多种、四种或更多种、或者五种或更多种)异源序列。

[0459] 靶核酸

[0460] 靶核酸是DNA分子、RNA分子,其为单链、双链或多链DNA或RNA、基因组DNA、cDNA、DNA-RNA杂交体,或包括嘌呤和嘧啶碱基或其他天然的、经化学或生化修饰的、非天然的或衍生的核苷酸碱基(脱氧核糖核苷酸、核糖核苷酸或其类似物)的聚合物。靶核酸可以具有三维结构,可以包括编码区或非编码区,可以包括外显子、内含子、mRNA、tRNA、rRNA、siRNA、shRNA、miRNA、核酶、cDNA、质粒、载体、外源序列、内源序列。靶核酸可以包括经修饰的核苷酸,所述经修饰的核苷酸包含甲基化核苷酸或核苷酸类似物。在一些实施例中,靶核酸可以散布有非核酸组分。

[0461] 靶核酸由CRISPR-Cas9系统识别并结合Cas9。在一些实施例中,其由于Cas9的结合而被修饰或切割或具有改变的表达。靶核酸含有特定的可识别PAM基序,例如5'-NGG-3'、5'-NGC-3'、5'-NAGHC-3'、5'-NRHRRH-3'或5'-NNAAA-3'(H=A、C或T;R=A或G)。

[0462] 重组基因技术

[0463] 根据本公开,可以采用本领域技术范围内的常规分子生物学、微生物学和重组DNA技术。此类技术描述于文献中(参见例如,Sambrook,Fritsch&Maniatis,Molecular Cloning:A Laboratory Manual,第二版(1989)Cold Spring Harbor Laboratory Press,Cold Spring Harbor,N.Y.;DNA Cloning:A Practical Approach,第I卷和第II卷(D.N.Glover编1985);Oligonucleotide Synthesis(M.J.Gait编1984);Nucleic Acid Hybridization(B.D.Hames&S.J.Higgins编(1985));Transcription And Translation(B.D.Hames&S.J.Higgins编(1984));Animal Cell Culture(R.I.Freshney编(1986));Immobilized Cells and Enzymes(IRL Press,(1986));B.Perbal,A Practical Guide To

Molecular Cloning(1984);F.M.Ausubel等人(编),Current Protocols in Molecular Biology,John Wiley&Sons,Inc.(1994)。

[0464] 基因(诸如编码多肽(诸如本文所述的工程化Cas9酶)的核酸)的重组表达可以包括构建含有编码该多肽的核酸的表达载体。一旦获得多核苷酸,就可以使用本领域已知的技术通过重组DNA技术产生用于多肽生产的载体。可以使用已知方法构建含有多肽编码序列和适当的转录和翻译控制信号的表达载体。这些方法包括例如体外重组DNA技术、合成技术和体内基因重组。

[0465] 可以通过常规技术将表达载体转移至宿主细胞,然后可以通过常规技术培养经感染的细胞以产生多肽。

[0466] 在一些实施例中,编码靶向DNA的RNA和/或Cas9蛋白的核苷酸序列可操作地连接至控制元件,例如转录控制元件,诸如启动子。转录控制元件可以在真核细胞(例如哺乳动物细胞);或原核细胞(例如细菌或古细菌细胞)中发挥功能。在一些实施例中,真核细胞是人类细胞。在一些实施例中,编码靶向DNA的RNA和/或新颖Cas9蛋白的核苷酸序列可操作地连接至允许经编码的核苷酸序列在原核细胞和真核细胞两者中表达的多个控制元件。

[0467] 启动子可以是组成型活性启动子(即,组成型处于活性/“启动”状态的启动子),其可以是诱导型启动子(即,状态为活性/“启动”、或非活性/“关断”的启动子,受外部刺激(例如特定温度、化合物或蛋白质的存在)控制),其可能是空间受限性启动子(即转录控制元件、增强子等)(例如组织特异性启动子、细胞类型特异性启动子等),并且其可以是时间受限性启动子(即,启动子在胚胎发育的特定阶段期间或生物过程的特定阶段(例如,小鼠的毛囊周期)期间处于“启动”状态或“关断”状态)。

[0468] 合适的启动子可源自病毒并因此可称为病毒启动子,或者它们可源自任何生物体,包括原核或真核生物体。合适的启动子可用于驱动利用任何RNA聚合酶的表达(例如, pol I、pol II、pol III)。示例性启动子包括但不限于SV40早期启动子、小鼠乳腺肿瘤病毒长末端重复(LTR)启动子;腺病毒主要晚期启动子(Ad MLP);单纯疱疹病毒(HSV)启动子、巨细胞病毒(CMV)启动子(诸如CMV立即早期启动子区(CMVIE)、劳斯肉瘤病毒(RSV)启动子、人U6小核启动子(U6)(Miyagishi等人,Nature Biotechnology 20,497-500(2002))、增强型U6启动子(例如,Xia等人,Nucleic Acids Res.2003年9月1日;31(17))和/或人类HI启动子(HI)。

[0469] 诱导型启动子的实例包括但不限于T7 RNA聚合酶启动子、T3 RNA聚合酶启动子、异丙基- β -D-硫代吡喃半乳糖苷(IPTG)调节型启动子、乳糖诱导型启动子、热休克启动子、四环素调节型启动子(例如Tet-ON、Tet-OFF等)、类固醇调节型启动子、金属调节型启动子、雌激素受体调节型启动子等。因此,诱导型启动子可以由分子调节,这些分子包括但不限于多西环素、RNA聚合酶(例如,T7 RNA聚合酶)、雌激素受体和/或雌激素受体融合物。

[0470] 在一些实施例中,启动子是空间受限性启动子(即,细胞类型特异性启动子、组织特异性启动子等),使得在多细胞生物体中,启动子在特定细胞子集中具有活性(即“启动”)。空间受限性启动子也可以称为增强子、转录控制元件、控制序列等。可以使用任何方便的空间受限性启动子并且选择合适的启动子(例如,脑特异性启动子、在神经元子集中驱动表达的启动子、在种系中驱动表达的启动子、在肺中驱动表达的启动子、在肌肉中驱动表达的启动子、在胰腺胰岛细胞中驱动表达的启动子等)将取决于生物体。因此,取决于生物

体,空间受限性启动子可用于调节编码主题定点多肽的核酸在广泛多种不同组织和细胞类型中的表达。一些空间受限性启动子也在时间上受到限制,使得启动子在胚胎发育的特定阶段期间或生物过程的特定阶段(例如,毛囊周期)期间处于“启动”状态或“关断”状态。

[0471] 出于说明目的,空间受限性启动子的实例包括但不限于神经元特异性启动子、脂肪细胞特异性启动子、心肌细胞特异性启动子、平滑肌特异性启动子、光感受器特异性启动子等。神经元特异性空间受限性启动子包括但不限于神经元特异性烯醇化酶(NSE)启动子、芳香族氨基酸脱羧酶(AADC)启动子、神经丝启动子、突触蛋白启动子、thy-1启动子、血清素受体启动子、酪氨酸羟化酶启动子(TH)、GnRH启动子、L7启动子、DNMT启动子、脑啡肽启动子、髓磷脂碱性蛋白(MBP)启动子、Ca²⁺钙调蛋白依赖性蛋白激酶II- α (CamKIIa)启动子和/或CMV增强子/血小板衍生生长因子- β 启动子。

[0472] 脂肪细胞特异性空间受限性启动子包括但不限于aP2基因启动子/增强子,例如人类aP2基因的-5.4kb至+21bp的区域、葡萄糖转运体-4(GLUT4)启动子、脂肪酸转位酶(FAT/CD36)启动子、硬脂酰辅酶A去饱和酶-1(SCD1)启动子、瘦素启动子以及脂联素启动子、脂肪酶启动子和/或抵抗素启动子。

[0473] 心肌细胞特异性空间受限性启动子包括但不限于源自以下基因的控制序列:肌球蛋白轻链2、 α -肌球蛋白重链、AE3、心肌肌钙蛋白C和/或心肌肌动蛋白。

[0474] 平滑肌特异性空间受限性启动子包括但不限于SM22a启动子、平滑肌蛋白启动子和/或 α -平滑肌肌动蛋白启动子。

[0475] 光感受器特异性空间受限性启动子包括但不限于视紫红质启动子、视紫红质激酶启动子、 β 磷酸二酯酶基因启动子、色素性视网膜炎基因启动子、光感受器间类视黄醇结合蛋白(IRBP)基因增强子和/或IRBP基因启动子。

[0476] CRISPR-Cas9的基因编辑用途

[0477] 本文所述的CRISPR-Cas9系统可用于基因编辑,这可以导致基因沉默事件或所需靶基因表达中的表达改变(例如,增加或减少)。因此,在一些实施例中,本文所述的CRISPR-Cas9系统用于改变靶核酸表达的方法中。在一些实施例中,本文所述的CRISPR-Cas9系统用于修饰所需靶细胞中的靶核酸的方法中。在一些实施例中,本发明提供对真核细胞中的靶核酸进行位点特异性修饰以实现基因表达中所需修饰的方法。

[0478] 在一些实施例中,本发明提供一种工程化的非天然存在的CRISPR-Cas系统,其包括:RNA向导或编码该RNA向导的核酸,其中该RNA向导包含能够与靶核酸杂交的同向重复序列和间隔子序列;以及经密码子优化的CRISPR相关(Cas)蛋白,其与SEQ ID NO:1、4、8、14、84或86具有至少80%序列同一性,并且其中该Cas蛋白能够与该RNA向导结合并导致与该RNA向导互补的靶核酸序列中的断裂。

[0479] 在一些实施例中,本发明提供工程化的非天然存在的CRISPR-Cas系统,其包括:RNA向导或编码该RNA向导的核酸,其中该RNA向导包含能够与靶核酸杂交的同向重复序列和间隔子序列;以及经密码子优化的CRISPR相关(Cas)蛋白,其与SEQ ID NO:1、4、8、14、84或86具有至少80%序列同一性;其中该Cas蛋白与脱氨酶融合,并且其中Cas蛋白融合物能够与该RNA向导结合并编辑与该RNA向导互补的靶核酸序列。

[0480] 在一些实施例中,本文提供一种工程化的非天然存在的CRISPR-Cas系统,其包括经密码子优化的CRISPR相关(Cas)蛋白,该经密码子优化的CRISPR相关蛋白进一步包含核

定位序列(NLS)和/或FLAG、HIS或HA标签。

[0481] 在一些实施例中,本文提供了一种工程化的非天然存在的Cas9融合蛋白,其进一步包含核定位序列(NLS)和/或FLAG、HIS或HA标签。

[0482] 在一些实施例中,本文提供了一种工程化的非天然存在的Cas9融合蛋白,其与SEQ ID NO:2、5、9、15、85、87、95或96具有至少80%同一性。

[0483] 在一些实施例中,本发明提供一种改变真核细胞中的靶核酸表达的方法,该方法包括:使该细胞与本文所述的Cas9和RNA向导或编码该RNA向导的核酸接触,其中该RNA向导包括能够与该靶核酸杂交的同向重复序列和间隔子序列,并且其中该Cas9蛋白能够与该RNA向导结合并导致与该RNA向导互补的靶核酸序列中的断裂。

[0484] 在一些实施例中,本发明提供一种改变真核细胞中的靶核酸表达的方法,该方法包括:使该细胞与本文所述的Cas9和RNA向导或编码该RNA向导的核酸接触,其中该RNA向导包括能够与该靶核酸杂交的同向重复序列和间隔子序列,并且其中该Cas9蛋白能够与该RNA向导结合并编辑与该RNA向导互补的靶核酸序列。

[0485] 在一些实施例中,本发明一种修饰真核细胞中的靶核酸的方法,该方法包括:使该细胞与本文所述的Cas9和RNA向导或编码该RNA向导的核酸接触,其中该RNA向导包括能够与该靶核酸杂交的同向重复序列和间隔子序列,并且其中该Cas9蛋白能够与该RNA向导结合并编辑与该RNA向导互补的靶核酸序列。

[0486] 因此,在一些实施例中,Cas蛋白与SEQ ID NO:1、4、8、14、84或86具有约80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%同一性。在一些实施例中,Cas蛋白与SEQ ID NO:1、4、8、14、84或86相同。

[0487] 用于CRISPR-Cas9系统和方法的适合的向导RNA、Cas9突变和融合蛋白如本公开全文所述。

[0488] 在一个方面,该方法包括使CRISPR-Cas9与靶核酸结合以及实现靶核酸的切割。在一些实施例中,CRISPR-Cas9系统通过引入双链断裂来切割靶DNA或RNA双链体。在一些实施例中,CRISPR-Cas9系统通过引入单链断裂或切口来切割靶DNA或RNA。

[0489] 在一些实施例中,CRISPR-Cas9方法或系统包括具有以位点特异性方式修饰靶DNA的效应子的融合蛋白,其中修饰活性包括甲基转移酶活性、去甲基化酶活性、乙酰转移酶活性、去乙酰化酶活性、激酶活性、磷酸酶活性、泛素连接酶活性、去泛素化活性、腺苷酸化活性、去腺苷酸化活性、SUMO化活性、去SUMO化活性、核糖基化活性、去核糖基化活性、肉豆蔻酰化活性、去肉豆蔻酰化活性、整合酶活性、转座酶活性、重组酶活性、聚合酶活性、连接酶活性、解旋酶活性或核酸酶活性,其中任何一种都可以修饰DNA或DNA相关多肽(例如,组蛋白或DNA结合蛋白)。

[0490] 在一些实施例中,CRISPR-Cas9方法或系统包括融合蛋白与可以通过化学修饰核苷酸碱基来编辑DNA序列的酶,这些酶包括可以修饰腺苷或胞嘧啶碱基并充当位点特异性碱基编辑器的脱氨酶。例如,通常使用RNA作为底物的APOBEC1胞苷脱氨酶在其与Cas9融合时可以靶向单链和双链DNA,从而将胞苷直接转化为尿苷,并且ADAR酶将腺苷脱氨基为肌苷。因此,使用脱氨酶的‘碱基编辑’能够将一个靶DNA碱基可编程转化为另一个。各种碱基编辑器在本领域中是已知的并且可以用于本文所述的方法和系统中。例示性碱基编辑器描

述于例如Rees和Liu Nature Review Genetics,2018,19(12):770-788中,其内容并入本文。因此,在一些实施例中,本文所述的Cas9酶(ScoCas9、SirCas9、VapCas9、EpeCas9、LfeCas9、PmaCas9)是核碱基编辑器的组分。在一些实施例中,碱基编辑器是腺嘌呤脱氨酶TadA8或TadA9。

[0491] 在一些实施例中,碱基编辑导致引入终止密码子来沉默基因。在一些实施例中,碱基编辑通过改变氨基酸序列来改变蛋白质功能。

[0492] 在一些实施例中,CRISPR-Cas9方法或系统包括通过与组蛋白融合来对靶DNA进行表观遗传修饰。在一些实施例中,CRISPR-Cas9系统包括通过与表观遗传修饰酶(诸如读取蛋白、写入蛋白或擦除蛋白)融合来对靶DNA进行表观遗传修饰。在一些实施例中,CRISPR-Cas9系统包括与组蛋白修饰酶融合以改变靶DNA的选定区域中的组蛋白修饰模式。组蛋白修饰可以以多种不同的方式(包括甲基化、乙酰化、泛素化、磷酸化以及许多不同的组合)发生,从而导致DNA的结构变化。在一些实施例中,组蛋白修饰导致转录阻遏或活化。

[0493] 在一些实施例中,CRISPR-Cas9方法或系统通过与转录活化蛋白或转录阻遏蛋白、小分子/药物响应性转录调节子、诱导型转录调节子融合来增加或减少转录,而调节靶DNA的转录。在一些实施例中,CRISPR-Cas9系统用于控制靶编码mRNA(即蛋白质编码基因)的表达,其中结合引起基因表达增加或减少。

[0494] 在一些实施例中,CRISPR-Cas9方法或系统用于通过编辑遗传调控元件(诸如启动子或增强子)来控制基因调控。

[0495] 在一些实施例中,CRISPR-Cas9方法或系统用于控制靶标非编码RNA(包括tRNA、rRNA、snoRNA、siRNA、miRNA和长ncRNA)的表达。

[0496] 在一些实施例中,CRISPR-Cas9方法或系统用于染色质环结构的靶向工程化。调控基因组区域之间的染色质环的靶向工程化提供了一种用于操纵内源染色质结构并能够形成新的增强子-启动子连接以克服遗传缺陷或抑制异常增强子-启动子连接的手段。

[0497] 在一些实施例中,CRISPR-Cas9用于活细胞成像。荧光标记的Cas9靶向重复基因组区域(诸如着丝粒和端粒),以跟踪整个细胞周期中的天然染色质基因座,并确定3D核空间中转录活性区域和转录非活性区域的差异定位。

[0498] 在一些实施例中,CRISPR-Cas9方法或系统用于通过插入有益的临床变体或抑制突变来校正致病性突变。

[0499] 核碱基编辑器

[0500] 本文公开了用于编辑、修饰或改变包含Cas9的多核苷酸的靶核苷酸序列的新颖碱基编辑器或核碱基编辑器。本文描述的是包含多核苷酸可编程核苷酸结合结构域(例如,Cas9)和核碱基编辑结构域(例如,腺苷脱氨酶)的核碱基编辑器或碱基编辑器。多核苷酸可编程核苷酸结合结构域(例如,Cas9)当与结合的向导多核苷酸(例如,gRNA)结合时,可以特异性结合于靶多核苷酸序列(即,通过结合的向导核酸的碱基与靶多核苷酸序列的碱基之间的互补碱基配对进行),且从而将碱基编辑器定位到需要编辑的靶核酸序列。在一些实施例中,靶多核苷酸序列包含单链DNA或双链DNA。在一些实施例中,靶多核苷酸序列包含RNA。在一些实施例中,靶多核苷酸序列包含DNA-RNA杂合体。由于大多数已知的与人类疾病相关的遗传变异都是点突变,因此需要能够更有效且更干净地产生精确点突变的方法。本文提供的碱基编辑系统提供一种新的方式来提供基因组编辑,而不产生双链DNA断裂、不需要供

体DNA模板并且不会诱导过量的随机插入和缺失。

[0501] 本文提供的碱基编辑器能够修饰特定的核苷酸碱基而不产生显著比例的插入缺失。如本文所用,术语“插入缺失”是指核酸内核苷酸碱基的插入或缺失。此类插入或缺失可导致基因编码区内发生移码突变。在一些实施例中,需要产生有效修饰(例如,产生突变或脱氨基)核酸内的特定核苷酸而不在靶核苷酸序列中产生大量插入或缺失(即,插入缺失)的碱基编辑器。在某些实施例中,本文提供的任何碱基编辑器能够产生相对于插入缺失更大比例的预期修饰(例如,点突变或脱氨基)。

[0502] 在一些实施例中,本文提供的任何碱基编辑器系统导致靶多核苷酸序列中小于50%、小于40%、小于30%、小于20%、小于19%、小于18%、小于17%、小于16%、小于15%、小于14%、小于13%、小于12%、小于11%、小于10%、小于9%、小于8%、小于7%、小于6%、小于5%、小于4%、小于3%、小于2%、小于1%、小于0.9%、小于0.8%、小于0.7%、小于0.6%、小于0.5%、小于0.4%、小于0.3%、小于0.2%、小于0.1%、小于0.09%、小于0.08%、小于0.07%、小于0.06%、小于0.05%、小于0.04%、小于0.03%、小于0.02%或小于0.01%的插入缺失形成。

[0503] 本公开的一些方面基于以下认识:本文提供的任何碱基编辑器能够在核酸(例如,受试者基因组内的核酸)中有效地产生预期突变,诸如点突变,而不会产生大量非预期突变,诸如非预期点突变。在一些实施例中,本文提供的任何碱基编辑器能够产生至少0.01%的预期突变(即至少0.01%的碱基编辑效率)。在一些实施例中,本文提供的任何碱基编辑器能够产生至少0.01%、1%、2%、3%、4%、5%、10%、15%、20%、25%、30%、40%、45%、50%、60%、70%、80%、90%、95%或99%的预期突变。

[0504] 在一些实施例中,本文提供的碱基编辑器能够产生大于1:1的预期点突变与插入缺失的比率。在一些实施例中,本文提供的碱基编辑器能够产生至少1.5:1、至少2:1、至少2.5:1、至少3:1、至少3.5:1、至少4:1、至少4.5:1、至少5:1、至少5.5:1、至少6:1、至少6.5:1、至少7:1、至少7.5:1、至少8:1、至少8.5:1、至少9:1、至少10:1、至少11:1、至少12:1、至少13:1、至少14:1、至少15:1、至少20:1、至少25:1、至少30:1、至少40:1、至少50:1、至少100:1、至少200:1、至少300:1、至少400:1、至少500:1、至少600:1、至少700:1、至少800:1、至少900:1或至少1000:1或更高的预期点突变与插入缺失的比率。

[0505] 可以使用例如如以下中所述的任何适合的方法来确定预期突变和插入缺失的数量:国际PCT申请号PCT/2017/045381(WO2018/027078)和PCT/US2016/058344(WO2017/070632);Komor,A.C.等人,“Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage”*Nature* 533,420-424(2016);Gaudelli,N.M.等人,“Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage”*Nature* 551,464-471(2017);和Komor,A.C.等人,“Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity”*Science Advances* 3:eaao4774(2017);其全部内容通过引用特此并入。

[0506] 在一些实施例中,为了计算插入缺失频率,扫描测序读段以寻找与两个10-bp序列的精确匹配,这些序列侧接可能出现插入缺失的窗口的两侧。如果未定位到精确匹配,则将从分析中排除该读段。如果该插入缺失窗口的长度与参考序列完全匹配,则该读段被归类

为不含有插入缺失。如果插入缺失窗口比参考序列长或短两个或更多个碱基,则测序读段分别被归类为插入或缺失。在一些实施例中,本文提供的碱基编辑器可以限制核酸区域中插入缺失的形成。在一些实施例中,该区域位于碱基编辑器靶向的核苷酸或碱基编辑器靶向的核苷酸的2、3、4、5、6、7、8、9或10个核苷酸内的区域处。

[0507] 在靶核苷酸区域处形成的插入缺失的数量可以取决于核酸(例如,细胞基因组内的核酸)暴露于碱基编辑器的时间量。在一些实施例中,在靶核苷酸序列(例如,细胞基因组内的核酸)暴露于碱基编辑器至少1小时、至少2小时、至少6小时、至少12小时、至少24小时、至少36小时、至少48小时、至少3天、至少4天、至少5天、至少7天、至少10天或至少14天后,确定插入缺失的数量或比例。应理解,如本文所述的碱基编辑器的特征可以应用于任何融合蛋白,或使用本文提供的融合蛋白的方法。

[0508] 治疗应用

[0509] 本文所述的CRISPR-Cas9方法或系统可以具有多种治疗应用。因此,在一些实施例中,提供一种治疗对此有需要的受试者的病症或疾病的方法,该方法包括向该受试者施用如本文所述的包括Cas9的CRISPR-Cas9系统,其中该向导RNA和与病况或疾病相关的靶核酸的至少10个核苷酸互补;其中该Cas蛋白与该向导RNA缔合;其中该向导RNA与该靶核酸结合;其中该Cas蛋白导致靶核酸中的断裂,任选地其中Cas9是与脱氨酶融合的无活性的Cas9(dCas9)并产生该靶核酸中的一个或多个碱基编辑,从而治疗该病症或疾病。

[0510] 在一些实施例中,CRISPR-Cas9方法或系统可用于治疗各种疾病和病症,例如,遗传病症(例如,单基因疾病)、可以通过核酸酶活性治疗的疾病和各种癌症等。

[0511] 在一些实施例中,本文所述的CRISPR方法或系统可用于编辑靶核酸,从而修饰靶核酸(例如,通过插入、缺失或突变一个或多个核酸残基修饰)。例如,在一些实施例中,本文所述的CRISPR系统包括外源供体模板核酸(例如,DNA分子或RNA分子),该外源供体模板核酸包括所需的核酸序列。在解决用本文所述的CRISPR系统诱导的切割事件后,细胞的分子机制将利用外源供体模板核酸来修复和/或解决切割事件。可替代地,细胞的分子机制可以利用内源模板来修复和/或解决切割事件。在一些实施例中,本文所述的CRISPR系统可用于改变导致插入、缺失和/或点突变的靶核酸。在一些实施例中,所述插入是无痕插入(即,将预期的核酸序列插入靶核酸中,从而导致在解决切割事件后没有另外的非预期的核酸序列)。供体模板核酸可以是双链或单链核酸分子(例如,DNA或RNA)。在一些实施例中,本文所述的CRISPR方法或系统包括核碱基编辑器。例如,在一些实施例中,本文所述的Cas9蛋白与具有核碱基编辑活性的多肽融合。

[0512] 在一个方面,本文所述的CRISPR方法或系统可用于治疗由RNA、毒性RNA和/或突变RNA(例如,剪接缺陷或截短)的过表达引起的疾病。

[0513] 在一些实施例中,本文所述的CRISPR方法或系统还可以靶向影响RNA依赖性功能的反式作用突变,这些突变会引起各种疾病。

[0514] 在一些实施例中,本文所述的CRISPR方法或系统还可用于靶向破坏顺式作用剪接密码的突变,这些突变可导致剪接缺陷和疾病。

[0515] 本文所述的CRISPR方法或系统可进一步用于抗病毒活性,特别是针对RNA病毒。CRISPR相关蛋白可以使用经选择用于靶向病毒RNA序列的适合的RNA向导来靶向病毒RNA。

[0516] 本文所述的CRISPR方法或系统还可用于治疗受试者(例如,人类受试者)的癌症。

例如,本文所述的CRISPR相关蛋白可用靶向RNA分子的crRNA编程,该RNA分子异常(例如,包括点突变或经选择性剪接)并可见于癌细胞中,以诱导癌细胞中的细胞死亡(例如,通过细胞凋亡)。

[0517] 此外,本文所述的CRISPR方法或系统还可用于治疗受试者的传染病。例如,本文所述的CRISPR相关蛋白可用靶向RNA分子的crRNA编程,该RNA分子通过感染原(例如,细菌、病毒、寄生虫或原生动物)表达,以靶向并诱导感染原细胞中的细胞死亡。CRISPR系统还可用于治疗细胞内感染原感染宿主受试者的细胞的疾病。通过对CRISPR相关蛋白进行编程以使其靶向感染原基因所编码的RNA分子,可以靶向感染有感染原的细胞并诱导细胞死亡。

[0518] 此外,体外RNA传感测定可用于检测特定的RNA底物。CRISPR相关蛋白可用于在活细胞中的基于RNA的传感。应用的实例是通过感测例如疾病特异性RNA进行诊断。

[0519] 在期望将多核苷酸序列插入靶DNA序列的应用中,还向细胞提供包括待插入的供体序列的多核苷酸。“供体序列”或“供体多核苷酸”意指待插入在由定点修饰多肽诱导的切割位点处的核酸序列。所述供体多核苷酸将在切割位点处与基因组序列含有足够的同源性,例如与侧接切割位点(例如,在切割位点的约50个碱基或更少以内,例如在约30个碱基内、在约15个碱基内、在约10个碱基内、在约5个碱基内,或紧侧接切割位点)的核苷酸序列具有70%、80%、85%、90%、95%或100%的同源性,以支持其与和其具有同源性的基因组序列之间的同源性定向修复。供体和基因组序列之间具有序列同源性的大约为25、50、100或200个核苷酸或超过200个核苷酸(或介于10与200或更多个核苷酸之间的任何整数值)将支持同源性定向修复。供体序列可以是任意长度,例如10个核苷酸或更多、50个核苷酸或更多、100个核苷酸或更多、250个核苷酸或更多、500个核苷酸或更多、1000个核苷酸或更多、5000个核苷酸或更多等。

[0520] 所述供体序列通常与其替换的基因组序列不同。相反,所述供体序列可以含有相对于基因组序列的至少一个或多个单碱基改变、插入、缺失、倒置或重排,只要存在足够的同源性以支持同源性定向修复即可。在一些实施例中,所述供体序列包括由两个同源区域侧接的非同源序列,使得靶DNA区域与两个侧接序列之间的同源定向修复导致非同源序列在靶区域处的插入。供体序列还可以包括载体骨架,所述载体骨架含有与所关注的DNA区域不同源并且不打算插入所关注的DNA区域的序列。通常,供体序列的同源区域与需要重组的基因组序列将具有至少50%的序列同一性。在某些实施例中,存在60%、70%、80%、90%、95%、98%、99%或99.9%的序列同一性。根据供体多核苷酸的长度,可以存在介于1%与100%序列同一性之间的任何值。

[0521] 与基因组序列相比,供体序列可以包括某些序列差异,例如限制性位点、核苷酸多态性、可选择标志物(例如,耐药基因、荧光蛋白、酶等)等,所述序列差异可以用于评估供体序列在切割位点处的成功插入,或在一些情况下可以用于其它目的(例如,表示在靶向基因组基因座处的表达)。在一些情况下,如果定位于编码区中,则此类核苷酸序列差异将不会改变氨基酸序列,或将使沉默氨基酸发生改变(即,不影响蛋白质的结构或功能的改变)。可替代地,这些序列差异可以包含侧接重组序列,如FLP、loxP序列等,所述重组序列可以在稍后的时间被激活以去除标志物序列。

[0522] 所述供体序列可以作为单链DNA、单链RNA、双链DNA或双链RNA提供给细胞。它可以线性或圆形形式引入细胞中。如果以线性形式引入,可以通过本领域技术人员已知的方法

保护供体序列的末端(例如,防止核酸外切降解)。例如,将一个或多个双脱氧核苷酸残基添加到线性分子的3'末端及/或将自互补寡核苷酸连接到一个或两个末端。用于保护外源多核苷酸免于降解的另外的方法包含但不限于添加末端氨基以及使用经过修饰的核苷酸间键,例如硫代磷酸酯、氨基磷酸酯和0-甲基核糖或脱氧核糖残基。作为保护线性供体序列末端的替代方案,可以在同源区域之外包含额外长度的序列,所述同源区域可以在不影响重组的情况下被降解。可以将供体序列作为载体分子的一部分引入细胞,所述载体分子具有额外序列,例如复制起点、启动子和编码抗生素抗性的基因。此外,供体序列可以作为裸核酸、作为与如脂质体或泊洛沙姆等药剂复合的核酸引入,或者可以通过病毒(例如,腺病毒、AAV)递送,如上文对于编码靶向DNA的RNA和/或定点修饰多肽和/或供体多核苷酸的核酸所述。

[0523] 按照上述方法,可以离体切割和修饰,即“基因修饰”所关注的DNA区域。在一些实施例中,当将可选择标志物插入所关注的DNA区域时,通过将经基因修饰的细胞与剩余细胞群分离,可以使细胞群富集包括经基因修饰的细胞。在富集之前,“经基因修饰的”细胞可能仅占细胞群的约1%或更多(例如,2%或更多、3%或更多、4%或更多、5%或更多、6%或更多、7%或更多、8%或更多、9%或更多、10%或更多、15%或更多,或20%或更多)。“经基因修饰的”细胞的分离可以通过适于所使用的可选择标志物的任何方便的分离技术来实现。例如,如果已插入荧光标记物,则可以通过荧光激活细胞分选来使细胞分离,而如果已插入细胞表面标记物,则可以通过亲和分离技术将细胞与异质群体分离,例如磁分离、亲和层析、用附着到固体基质上的亲和试剂进行“淘选”或其它方便的技术。提供精确分离的技术包括荧光激活细胞分选仪,该荧光激活细胞分选仪可以具有不同程度的复杂性,诸如多色通道、低角度和钝角光散射检测通道、阻抗通道等。可以通过采用与死细胞相关联的染料(例如碘化丙啶)来针对死细胞选择这些细胞。可以采用不会过度损害经基因修饰的细胞的活力的任何技术。以此方式实现包括经修饰的DNA的细胞高度富集的细胞组合物。“高度富集”是指经过遗传修饰的细胞将为细胞组合物的70%或更多、75%或更多、80%或更多、85%或更多、90%或更多,例如细胞组合物的约95%或更多或98%或更多。换言之,所述组合物可以是基本上纯的经基因修饰的细胞的组合物。

[0524] 通过本文所述的方法产生的经基因修饰的细胞可以立即使用。可替代地,所述细胞可以在液氮温度下冷冻并长期储存,解冻并能够重复使用。在这种情况下,细胞通常会被冷冻在10%二甲亚砜(DMSO)、50%血清、40%缓冲介质或如本领域中常用的其它某种溶液中以将细胞保存在这种冷冻温度下,并且以本领域中已知的用于解冻冷冻培养细胞的方式解冻。

[0525] 经基因修饰的细胞可以在各种培养条件下体外培养。所述细胞可以在培养物中扩增,即在促进其增殖的条件下生长。培养基可以是液体或半固体,例如含有琼脂、甲基纤维素等。细胞群可以悬浮在适当的营养培养基中,诸如Iscove改良的DMEM或RPMI 1640,该营养培养基通常补充有胎牛血清(约5%-10%)、

[0526] L-谷氨酰胺、硫醇,特别是2-巯基乙醇和抗生素,例如青霉素和链霉素。培养物可以含有调节性T细胞对其有应答的生长因子。如本文所定义的,生长因子是能够通过对跨膜受体的特定作用促进细胞在培养物或完整组织中存活、生长和/或分化的分子。生长因子包含多肽和非多肽因子。

[0527] 可以将已经以这种方式进行经基因修饰的细胞移植到受试者体内,用于基因治疗等目的,例如治疗疾病或作为抗病毒、抗病原体或抗癌治疗剂,用于在农业中生产经基因修饰的生物体,或用于生物学研究。受试者可以是新生儿、青少年或成人。特别关注的是哺乳动物受试者。以用本方法处理的哺乳动物物种包含犬和猫;马;牛;绵羊等和灵长类动物,特别是人类。动物模型,特别是小型哺乳动物(例如小鼠、大鼠、豚鼠、仓鼠、兔形目动物(例如兔子)等)可以用于实验研究。

[0528] 细胞可以单独或与合适的底物或基质一起提供给受试者,例如,以支持它们在被移植的组织中的生长和/或组织。通常,将施用至少 1×10^3 个细胞,例如 5×10^3 个细胞、 1×10^4 个细胞、 5×10^4 个细胞、 1×10^5 个细胞、 1×10^6 个细胞或更多。可以通过以下任何途径将细胞引入受试者:肠胃外、皮下、静脉内、颅内、脊髓内、眼内或进入脊髓液。细胞可以通过注射、导管等引入。为了产生转基因动物(例如,转基因小鼠)的目的,也可以将细胞引入胚胎(例如,胚胎)中。

[0529] 向受试者施用治疗的次数可以变化。将经基因修饰的细胞引入受试者可能是一次性事件;但在某些情况下,这种治疗可能会在有限的时间段内引起改善并且需要进行一系列持续的重复治疗。在其它情况下,在观察到效果之前,可能需要多次施用经基因修饰的细胞。确切的方案取决于疾病或病状、疾病的阶段和正在治疗的个体受试者的参数。

[0530] 在本发明的其它方面,靶向DNA的RNA和/或定点修饰多肽和/或供体多核苷酸用于体内修饰细胞DNA,同样用于基因治疗等目的,例如用于治疗疾病或作为抗病毒剂、抗病原体剂或抗癌剂,用于在农业中生产经基因修饰的生物体,或用于生物学研究。在这些体内实施例中,将靶向DNA的RNA和/或定点修饰多肽和/或供体多核苷酸直接施用于个体。靶向DNA的RNA和/或定点修饰多肽和/或供体多核苷酸可以通过本领域中用于将肽、小分子和核酸施用给受试者的许多众所周知的方法中的任一种施用。可以将靶向DNA的RNA和/或定点修饰多肽和/或供体多核苷酸掺入多种调配物中。更具体地,本发明的靶向DNA的RNA和/或定点修饰多肽和/或供体多核苷酸可以通过与适当的药学上可接受的载体或稀释剂组合而调配成药物组合物。

[0531] 药物制剂是包含一种或多种存在于药学上可接受的载体中的靶向DNA的RNA和/或定点修饰多肽和/或供体多核苷酸的组合物。“药学上可接受的媒剂”可以是由联邦或州政府的监管机构批准的或者在美国

[0532] 药典或其它公认药典中列出的在如人类等哺乳动物中使用的媒剂。术语“媒剂”是指本发明的化合物与其一起调配用于向哺乳动物施用的稀释剂、佐剂、赋形剂或载体。此类药物媒剂可以是脂质,例如脂质体,例如脂质体树状物;液体,如水和油,包含石油、动物、植物或合成来源的油,如花生油、大豆油、矿物油、芝麻油等,盐水;阿拉伯树胶、明胶、淀粉糊、滑石、角蛋白、胶体二氧化硅、尿素等。另外,可以使用助剂、稳定剂、增稠剂、润滑剂和着色剂。药物组合物可以调配成固体、半固体、液体或气体形式的制剂,如片剂、胶囊、粉末、颗粒、软膏、溶液、栓剂、注射剂、吸入剂、凝胶、微球和气雾剂。因此,DNA靶向RNA和/或定点修饰多肽和/或供体多核苷酸的施用可以多种方式实现,包含口服、口腔、直肠、肠胃外、腹膜内、皮内、经皮、气管内、眼内等施用。活性剂在施用后可以是全身性的,也可以通过使用局部施用、壁内施用或使用植入物来定位,所述植入物起到在植入部位保留活性剂量的作用。所述活性剂可以被调配成立即活性或可以被调配成缓释。

[0533] 对于一些病状,特别是中枢神经系统病状,可能需要调配药剂以穿过血脑屏障(BBB)。一种通过血脑屏障(BBB)递送药物的策略需要通过如甘露醇或白三烯等渗透性手段或生化地通过使用血管活性物质如缓激肽来破坏BBB。使用BBB开放将特定药剂靶向脑肿瘤的潜力也是一种选择。当组合物通过血管内注射施用,BBB破坏剂可以与本发明的治疗组合物共同施用。其它通过BBB的策略可能需要使用内源性转运系统,包含Caveolin-1介导的转胞吞作用、如葡萄糖和氨基酸载体等载体介导的转运蛋白、受体介导的胰岛素或转铁蛋白转胞吞作用以及主动外排转运蛋白,如p-糖蛋白。活性转运部分也可以与用于本发明的治疗化合物缀合以促进穿过血管内皮壁的转运。

[0534] 可替代地,BBB后面的治疗剂的药物递送可以通过局部递送,例如通过鞘内递送。

[0535] 通常,提供有效量的靶向DNA的RNA和/或定点修饰多肽和/或供体多核苷酸。如上文关于离体方法所讨论的,体内靶向DNA的RNA和/或定点修饰多肽和/或供体多核苷酸的有效量或有效剂量是诱导两个同源序列之间观察到的重组量相对于阴性对照,例如与空载体或无关多肽接触的细胞增加2倍或更多的量。重组的量可以通过例如如上文所述且本领域已知的任何方便的方法来测量。待施用的靶向DNA的RNA和/或定点修饰多肽和/或供体多核苷酸的有效量或有效剂量的计算在本领域普通技术人员的技能范围内,并且对于本领域技术人员来说是常规的。待施用的最终量将取决于施用途径和待治疗的病症或病状的性质。

[0536] 给予特定患者的有效量将取决于多种因素,其中一些因素因患者而异。有能力的临床医生将能够确定有效量的治疗剂以施用于患者,以根据需要停止或逆转疾病状况的进展。利用LD50动物数据和所述药剂可用的其它信息,临床医生可以根据施用途径确定个体的最大安全剂量。例如,假设治疗组合物被施用到更大的体液中,静脉内施用的剂量可能大于鞘内施用的剂量。类似地,可以以更高剂量或以重复剂量施用快速从体内清除的组合物,以维持治疗浓度。利用普通技能,有能力的临床医生将能够在常规临床试验过程中优化特定治疗剂的剂量。

[0537] 为了包含在药物中,靶向DNA的RNA和/或定点修饰多肽和/或供体多核苷酸可以从合适的商业来源获得。作为一般建议,每剂肠胃外施用的靶向DNA的RNA和/或定点修饰多肽和/或供体多核苷酸的总药理学有效量将处于可以通过剂量应答曲线测量的范围内。

[0538] 基于靶向DNA的RNA和/或定点修饰多肽和/或供体多核苷酸的疗法,即用于治疗性施用的靶向DNA的RNA和/或定点修饰多肽和/或供体多核苷酸的制剂,必须是无菌的。通过无菌过滤膜(例如,0.2 μ m膜)过滤容易实现无菌。治疗性组合物通常放置在具有无菌入口的容器中,例如,静脉注射溶液袋或具有可通过皮下注射针刺穿的塞子的小瓶。基于靶向DNA的RNA和/或定点修饰多肽和/或供体多核苷酸的疗法可以作为水溶液或作为冻干调配物储存在单位或多剂量容器(例如,密封的安瓿或小瓶)中用于重组。作为冻干调配物的实例,10-mL小瓶填充有5mL无菌过滤的1% (w/v) 化合物水溶液,并将所得混合物冻干。通过使用抑菌注射用水重构冻干化合物来制备输注溶液。

[0539] 取决于期望的调配物,药物组合物可以包含药学上可接受的、无毒的稀释剂载体,所述载体被定义为通常用于调配用于动物或人类施用的药物组合物的载体。对稀释剂进行选择以免影响所述组合物的生物活性。此类稀释剂的实例是蒸馏水、缓冲水、生理盐水、PBS、林格氏溶液(Ringer's solution)、葡萄糖溶液和汉克氏溶液(Hank's solution)。另外,药物组合物或调配物可以包含其它载体、佐剂、或无毒的、非治疗性的、非免疫原性的稳

定剂、赋形剂等。所述组合物还可以包含接近生理条件的另外的物质,如pH调节剂和缓冲剂、毒性调节剂、润湿剂和去污剂。

[0540] 所述组合物还可以包含多种稳定剂中的任何一种,如抗氧化剂。当药物组合物包含多肽时,所述多肽可以与增强多肽的体内稳定性或以其它方式增强其药理学性质(例如,增加多肽的半衰期,降低其毒性并提高溶解度或吸收)的各种众所周知的化合物复合。此类修饰剂或络合剂的实例包含硫酸盐、葡萄糖酸盐、柠檬酸盐和磷酸盐。组合物的核酸或多肽也可以与增强其体内属性的分子复合。此类分子包含例如碳水化合物、多胺、氨基酸、其它肽、离子(例如,钠、钾、钙、镁、锰)和脂质。

[0541] 可以施用药物组合物用于预防和/或治疗性处理。可以根据细胞培养物和/或实验动物中的标准药理学程序确定活性成分的毒性和治疗功效,包含例如确定LD50(对50%群体致死的剂量)和ED50(对50%群体治疗有效的剂量)。毒性作用与治疗效果之间的剂量比为治疗指数并且所述治疗指数可以被表示为比率LD50/ED50。优选展现出大的治疗指数的疗法。

[0542] 从细胞培养和/或动物研究获得的数据可以用于调配人类使用的一系列剂量。活性成分的剂量通常在包含低毒性ED50在内的循环浓度范围内。根据所采用的剂型和所利用的施用通路,剂量可以在此范围内变化。

[0543] 用于调配药物组合物的组分优选地具有高纯度并且基本上不含潜在有害污染物(例如,至少是国家食品(NF)级,通常至少是分析级,并且更通常至少是药物级)。此外,用于体内使用的组合物通常是无菌的。就必须在使用前合成给定化合物而言,所得产物通常基本上不含任何潜在的毒性剂,特别是任何可能存在于合成或纯化过程中的内毒素。用于肠胃外施用的组合物也是无菌的、基本上等渗的并且在GMP条件下制备。

[0544] 递送系统

[0545] 本文所述的CRISPR系统或其组分、其核酸分子和/或编码或提供其组分的核酸分子、CRISPR相关蛋白或RNA向导可以通过各种递送系统,诸如载体(例如质粒和递送载体)递送。下文描述例示性实施例。CRISPR系统(例如,包括Cas9,其包含本文所述的核碱基编辑器)可以在包含在病毒载体中的核酸上被编码。病毒载体可包括慢病毒、腺病毒、逆转录病毒和腺相关病毒(AAV)。可以基于应用选择病毒载体。例如,AAV由于其温和的免疫原性而常用于体内基因递送。腺病毒因为它们所诱导的强烈的免疫原性反应而通常用作疫苗。病毒载体的封装容量可能限制可包装到载体中的碱基编辑器的大小。例如,AAV的封装容量为约4.5kb,包括两个145个碱基的反向末端重复序列(ITR)。

[0546] AAV是一种小型单链DNA依赖性病毒,属于细小病毒科。4.7kb野生型(wt) AAV基因组由两个基因构成,该两个基因分别编码四个复制蛋白和三个衣壳蛋白,并且该基因组在两侧上各侧接有145bp反向末端重复序列(ITR)。病毒粒子由三种衣壳蛋白即Vp1、Vp2和Vp3构成,该三种衣壳蛋白以1:1:10比率从同一开放阅读框但由差异剪接(Vp1)和替代翻译起始位点(分别为Vp2和Vp3)产生。Vp3是病毒粒子中最丰富的亚基并且参与细胞表面处的受体识别,从而定义病毒的向性。已在Vp1的独特N末端中鉴定出在病毒感染性中发挥作用的磷脂酶结构域。

[0547] 类似于wtAAV,重组AAV(rAAV)利用顺式作用的145-bp ITR侧接载体转基因盒,为外源DNA的封装提供高达4.5kb。感染后,rAAV可以表达本发明的融合蛋白,并且通过以游离

体形式存在于环状头-尾串联体中来持续存在而不整合到宿主基因组中。尽管rAAV在体外和体内成功使用该系统的实例有很多,但当基因编码序列的长度在大小上等于或大于wtAAV基因组时,有限的封装容量限制了AAV介导的基因递送的使用。

[0548] AAV载体的小封装容量使得大量超过此大小的基因的递送和/或大型生理调控元件的使用具有挑战性。这些挑战可以例如通过将待递送的蛋白质分成两个或更多个片段来解决,其中N末端片段与分裂的内含肽-N融合,并且C末端片段与分裂的内含肽-C融合。然后将这些片段封装到两个或更多个AAV载体中。如本文所用,“内含肽”是指连接侧接N末端和C末端外显肽(例如,待接合的片段)的自剪接蛋白质内含子(例如,肽)。某些内含肽用于接合异源蛋白质片段的用途描述例如于Wood等人, *J. Biol. Chem.* 289(21);14512-9(2014)中。例如,当与单独的蛋白质片段融合时,内含肽IntN和IntC相互识别,将自身剪接出并同时连接它们所融合的蛋白质片段的侧接N末端和C末端外显肽,从而由两个蛋白质片段重建全长蛋白质。其他适合的内含肽对于本领域技术人员来说将是显而易见的。

[0549] 在一些实施例中,本发明的CRISPR系统的长度可以变化。在一些实施例中,蛋白质片段的长度范围为2个氨基酸至约1000个氨基酸。在一些实施例中,蛋白质片段的长度范围为约5个氨基酸至约500个氨基酸。在一些实施例中,蛋白质片段的长度范围为约20个氨基酸至约200个氨基酸。在一些实施例中,蛋白质片段的长度范围为约10个氨基酸至约100个氨基酸。其他长度的适合的蛋白质片段对于本领域技术人员来说将是显而易见的。

[0550] 在一些实施例中,核酸酶(例如,Cas9)的一部分或片段与内含肽融合。核酸酶可与内含肽的N末端或C末端融合。在一些实施例中,融合蛋白的一部分或片段与内含肽融合并与AAV衣壳蛋白融合。内含肽、核酸酶和衣壳蛋白可以以任何排列(例如,核酸酶-内含肽-衣壳、内含肽-核酸酶-衣壳、衣壳-内含肽-核酸酶等)融合在一起。在一些实施例中,内含肽的N末端与融合蛋白的C末端融合,并且内含肽的C末端与AAV衣壳蛋白的N末端融合。

[0551] 在一个实施例中,通过将大的转基因表达盒分割成两个单独的半部(5'端和3'端,或头和尾)来产生双AAV载体,其中盒的每一半被封装在(<5kb的)单个AAV载体中。然后,在通过两个双AAV载体共感染同一细胞,随后:(1) 5'基因组和3'基因组之间的同源重组(HR)(双AAV重叠载体);(2) ITR介导的5'基因组和3'基因组的尾-头串联体化(双AAV反式剪接载体);或(3)这两种机制的组合(双AAV杂交载体),实现全长转基因表达盒的重新组装。双AAV载体在体内的使用引起全长蛋白质的表达。双AAV载体平台的使用代表了一种针对大小>4.7kb的转基因的有效且可行的基因转移策略。

[0552] 所公开的用于设计包括本文描述的Cas9的CRISPR系统的策略可用于产生能够封装到病毒载体中的CRISPR系统。使用基于RNA或DNA病毒的系统来递送碱基编辑器利用了高度进化的过程,用于将病毒靶向培养物或宿主内的特定细胞并将病毒有效载荷运输到细胞核或宿主细胞基因组。病毒载体可以直接施用于培养物、患者(体内)内的细胞,或者它们可以用于在体外处理细胞,并且可任选地将经修饰的细胞施用于患者(离体)。传统的基于病毒的系统可以包括用于基因转移的逆转录病毒、慢病毒、腺病毒、腺相关病毒和单纯疱疹病毒载体。通过逆转录病毒、慢病毒和腺相关病毒基因转移方法可以在宿主基因组中整合,通常引起插入的转基因长期表达。此外,已在许多不同的细胞类型和靶组织中观察到高转导效率。

[0553] 逆转录病毒的嗜性可以通过掺入外源包膜蛋白来改变,从而扩大靶细胞的潜在靶

标群体。慢病毒载体是逆转录病毒载体,其能够转导或感染非分裂细胞,并且通常会产生高病毒滴度。因此,逆转录病毒基因转移系统的选择将取决于靶组织。逆转录病毒载体包含具有至多6kb至10kb外源序列包装能力的顺式作用长末端重复序列。最小的顺式作用LTR足以复制和包装载体,然后该载体用于将治疗基因整合到靶细胞中以提供永久性的转基因表达。广泛使用的逆转录病毒载体包括基于鼠白血病病毒(MuLV)、长臂猿白血病病毒(GaLV)、猿免疫缺陷病毒(SIV)、人类免疫缺陷病毒(HIV)及其组合的那些逆转录病毒载体(参见例如,Buchscher等人,J.Virol.66:2731-2739(1992);Johann等人,J.Virol.66:1635-1640(1992);Sommerfelt等人,Virol.176:58-59(1990);Wilson等人,J.Virol.63:2374-2378(1989);Miller等人,J.Virol.65:2220-2224(1991);PCT/US94/05700)。

[0554] 逆转录病毒载体,尤其是慢病毒载体,可能需要小于给定长度的多核苷酸序列以有效整合到靶细胞中。例如,与较小尺寸的逆转录病毒载体相比,长度大于9kb的逆转录病毒载体可能导致病毒滴度较低。在一些方面,本公开的CRISPR系统(例如,包括本文公开的Cas9)具有足够的大小,以便使得能够通过逆转录病毒载体有效封装和递送到靶细胞中。在一些情况下,Cas9的大小使得即使当与向导核酸和/或可靶向核酸酶系统的其他组分一起表达时也允许有效的封装和递送。

[0555] 在优选瞬时表达的应用中,可以使用基于腺病毒的系统。基于腺病毒的载体能够在许多细胞类型中具有非常高的转导效率并且不需要细胞分裂。利用此类载体,已经获得了高滴度和高水平的表达。该载体可以在相对简单的系统中大量生产。腺相关病毒("AAV")载体还可用于用靶核酸转导细胞,例如,在核酸和肽的体外生产中,以及用于体内和离体基因疗法程序(参见例如,West等人,Virology 160:38-47(1987);美国专利号4,797,368;WO 93/24641;Kotin,Human Gene Therapy 5:793-801(1994);Muzyczka,J.Clin.Invest.94:1351(1994)。重组AAV载体的构建在许多出版物中都有描述,包括美国专利号5,173,414;Tratschin等人,Mol.Cell.Biol.5:3251-3260(1985);Tratschin等人,Mol.Cell.Biol.4:2072-2081(1984);Hermonat&Muzyczka,PNAS 81:6466-6470(1984);和Samulski等人,J.Virol.63:03822-3828(1989)。

[0556] 因此,本文所述的CRISPR系统(例如,包括本文公开的Cas9)可以用病毒载体递送。碱基编辑器系统的一种或多种组分可以在一种或多种病毒载体上被编码。例如,碱基编辑器和向导核酸可以在单个病毒载体上被编码。在其他情况下,碱基编辑器和向导核酸在不同的病毒载体上被编码。在任一情况下,碱基编辑器和向导核酸可以各自可操作地连接至启动子和终止子。

[0557] 在病毒载体上编码的组分的组合可以通过所选病毒载体的负荷物大小限制来确定。

[0558] 碱基编辑器的非病毒递送

[0559] 用于CRISPR的非病毒递送方法也是可用的。一类重要的非病毒核酸载体是纳米颗粒,其可以是有机的或无机的。纳米颗粒在本领域中是众所周知的。任何合适的纳米颗粒设计都可以用于递送基因组编辑系统组分或编码此类组分的核酸。例如,有机(例如,脂质和/或聚合物)纳米颗粒可以适合用作本公开的某些实施例中的递送媒剂。用于纳米颗粒调配物和/或基因转移的示例性脂质示于表5(下文)中。

[0560] 表5

用于基因转移的脂质			
[0561]	脂质	缩略语	特征
	1,2-二油酰基-sn-甘油-3-磷脂酰胆碱	DOPC	辅助
	1,2-二油酰基-sn-甘油-3-磷脂酰乙醇胺	DOPE	辅助
	用于基因转移的脂质		
[0562]	脂质	缩略语	特征
	胆固醇		辅助
	N-[1-(2,3-二油烯基氧基)丙基]N,N,N-三甲基氯化铵	DOTMA	阳离子
	1,2-二油酰氧基-3-三甲基铵-丙烷	DOTAP	阳离子
	二十八烷基酰胺基甘氨酸精胺	DOGS	阳离子
	N-(3-氨丙基)-N,N-二甲基-2,3-双(十二烷氧基)-1-溴化丙胺	GAP-DLRIE	阳离子
	十六烷基三甲基溴化铵	CTAB	阳离子
	6-月桂氧基己基鸟氨酸	LHON	阳离子
	1-(2,3-二油酰氧基丙基)-2,4,6-三甲基吡啶鎓	2Oc	阳离子
	2,3-二油烯基氧基-N-[2(精胺甲酰氨基-乙基)-N,N-二甲基-1-丙铵三氟乙酸酯	DOSPA	阳离子
	1,2-二油酰基-3-三甲基铵-丙烷	DOPA	阳离子
	N-(2-羟乙基)-N,N-二甲基-2,3-双(十二烷氧基)-1-溴化丙胺	MDRIE	阳离子
	二吡啶氧丙基二甲基羟乙基溴化铵	DMRI	阳离子
	3β-[N-(N',N'-二甲氨基乙烷)-氨基甲酰基]胆固醇	DC-Chol	阳离子
	双-胍基-tren-胆固醇	BGTC	阳离子
	1,3-二脱氧-2-(6-羧基-精酰基)-丙酰胺	DOSPER	阳离子
	二甲基二十八烷基溴化铵	DDAB	阳离子
	二-十八基氨基甘氨酸精胺	DSL	阳离子
	外消旋-[(2,3-双十八烷氧基丙基)(2-羟乙基)]-二甲基氯化铵	CLIP-1	阳离子
	外消旋-[2(2,3-二十六烷氧基丙基-氧甲氧基)乙基]溴化十六烷基三甲铵	CLIP-6	阳离子
	乙基二吡啶磷脂酰胆碱	EDMPC	阳离子
	1,2-二硬脂氧基-N,N-二甲基-3-氨基丙烷	DSDMA	阳离子
	1,2-二肉豆蔻酰基-三甲基铵丙烷	DMTAP	阳离子
	O,O'-二肉豆蔻基-N-赖氨酸天冬氨酸	DMKE	阳离子
	1,2-二硬脂酰基-sn-甘油-3-乙基磷胆碱	DSEPC	阳离子

用于基因转移的脂质		
脂质	缩略语	特征
N-棕榈酰 D-赤型-鞘氨醇氨基甲酰-精胺	CCS	阳离子
N-叔丁基-N0-十四烷基-3-十四烷基氨基丙脒	diC14-脒	阳离子
十八碳烯氧基[乙基-2-十七碳烯基-3 羟乙基]氯化咪唑啉	DOTIM	阳离子
N1-胆甾氧羰基-3,7-二氮杂壬烷-1,9-二胺	CDAN	阳离子
[0563] 2-(3-[双(3-氨基-丙基)-氨基]丙氨基)-N-二十四烷基氨基甲酰基甲乙酰胺	RPR209120	阳离子
1,2-二亚油酰氧基-3-二甲氨基丙烷	DLinDMA	阳离子
2,2-二亚油基-4-二甲基氨基乙基-[1,3]-二氧戊环	DLin-KC2-D MA	阳离子
二亚油基-甲基-4-二甲基氨基丁酸酯	DLin-MC3-D MA	阳离子

[0564] 表6列出了用于基因转移和/或纳米颗粒调配物的示例性聚合物。

[0565] 表6

用于基因转移的聚合物	
聚合物	缩略语
聚(乙二醇)	PEG
聚乙烯亚胺	PEI
二硫代双(琥珀酰亚胺基丙酸酯)	DSP
二甲基-3,3'-二硫代二丙亚胺酸	DTBP
聚(乙烯亚胺)二氨基甲酸酯	PEIC
[0566] 聚(L-赖氨酸)	PLL
组氨酸修饰的 PLL	
聚(N-乙烯基吡咯烷酮)	PVP
聚(丙烯亚胺)	PPI
聚(酰胺基胺)	PAMAM
聚(酰胺乙烯胺)	SS-PAEI
三乙烯四胺	TETA
聚(β -氨基酯)	

用于基因转移的聚合物

聚合物	缩略语
聚(4-羟基-L-脯氨酸酯)	PHP
聚(烯丙基胺)	
聚(α -[4-氨基丁基]-L-乙醇酸)	PAGA
聚(D,L-乳酸-共-乙醇酸)	PLGA
聚(N-乙基-4-乙烯基溴化吡啶)	
聚(磷腈)	PPZ
聚(磷酸酯)	PPE
聚(磷酸胺)	PPA
聚(N-2-羟丙基甲基丙烯酰胺)	pHPMA
聚(2-(二甲氨基)甲基丙烯酸乙酯)	pDMAEMA
聚(2-氨基乙基丙烯磷酸酯)	PPE-EA
壳聚糖	
半乳糖基化壳聚糖	
N-十二烷基化壳聚糖	
组蛋白	
胶原	
右旋糖酐-精胺	D-SPM

[0568] 表7总结了用于编码本文所述Cas9的多核苷酸的递送方法。

[0569] 表7

递送	载体/模式	递送到非分裂细胞中	表达持续时间	基因组整合	递送的分子类型
物理	(例如, 电穿孔、粒子枪、磷酸钙转染)	是	短暂的	否	核酸和蛋白质
病毒	逆转录病毒	否	稳定的	是	RNA
	慢病毒	是	稳定的	是/在具有修饰的情况下	RNA
				否	

递送	载体/模式	递送到非分		表达持续		递送的分子类型
		裂细胞中	时间	基因组整合		
[0571]	腺病毒	是	短暂的	否		DNA
	腺相关病毒 (AAV)	是	稳定的	否		DNA
	牛痘病毒	是	非常短暂	否		DNA
	单纯性疱疹病毒	是	稳定的	否		DNA
	非病毒 阳离子脂质体	是	短暂的	取决于递送物质		核酸和蛋白质
	聚合物纳米颗粒	是	短暂的	取决于递送物质		核酸和蛋白质
	生物非病毒 减毒细菌	是	短暂的	否		核酸
	递送媒介 经工程化的噬菌体	是	短暂的	否		核酸
	哺乳动物病毒样颗粒	是	短暂的	否		核酸
	生物脂质体：红细胞血影和外泌体	是	短暂的	否		核酸

[0572] 在另一方面,递送基因组编辑系统组件或编码此类组件的核酸,例如,核酸结合蛋白,诸如例如Cas9或其变体,其任选地与具有生物活性的多肽(例如,核碱基编辑器)融合,以及靶向感兴趣的基因组核酸序列的gRNA,可以通过将核糖核蛋白(RNP)递送至细胞来实现。所述RNP包括与靶向gRNA复合的核酸结合蛋白,例如Cas9。可以使用已知方法将RNP递送至细胞,这些方法诸如电穿孔、核转染或阳离子脂质介导的方法,例如,如以下所报导:Zuris,J.A.等人,2015,Nat.Biotechnology,33(1):73-80。RNP有利于在CRISPR碱基编辑系统中使用,特别是对于难以转染的细胞,例如原代细胞。另外,RNP还可以缓解细胞中蛋白质表达可能出现的困难,特别是当真核启动子(例如,可以用于CRISPR质粒的CMV或EF1A)表达不佳时。有利地,RNP的使用不需要将外源DNA递送到细胞中。此外,由于包括核酸结合蛋白和gRNA复合物的RNP会随着时间的推移而降解,因此使用RNP有可能限制脱靶效应。以与基于质粒的技术类似的方式,RNP可以用于递送结合蛋白(例如,Cas9变体)和指导同源定向修复(HDR)。

[0573] 用于驱动CRISPR系统的启动子(例如,包括本文所述的Cas9)可以包括AAV ITR。这对于消除对可能占用载体空间的额外启动子元件的需要是有利的。释放的额外空间可以用

于驱动其它元件的表达,如向导核酸或可选择标志物。ITR活性相对较弱,因此可以用于降低因所选核酸酶过度表达而导致的潜在毒性。

[0574] 任何合适的启动子都可以用于驱动Cas9和适当时向导核酸的表达。对于普遍表达,可以使用的启动子包括CMV、CAG、CBh、PGK、SV40、铁蛋白重链或轻链等。对于脑或其他CNS细胞表达,合适的启动子可以包括:用于所有神经元的突触蛋白I、用于兴奋性神经元的CaMKII α 、用于GABA能神经元的GAD67或GAD65或VGAT等。对于肝细胞表达,合适的启动子包括白蛋白启动子。对于肺细胞表达,合适的启动子可以包含SP-B。对于内皮细胞,合适的启动子可以包含ICAM。对于造血细胞,合适的启动子可以包含IFN β 或CD45。对于成骨细胞,合适的启动子可以包含OG-2。

[0575] 在一些情况下,本公开的Cas9具有足够小的尺寸以允许单独的启动子驱动同一核酸分子内的碱基编辑器和相容的向导核酸的表达。例如,载体或病毒载体可以包括与编码碱基编辑器的核酸可操作地连接的第一启动子和与向导核酸可操作地连接的第二启动子。

[0576] 用于驱动向导核酸表达的启动子可以包括:Pol III启动子,诸如U6或H1使用Pol II启动子和内含子盒来表达gRNA腺相关病毒(AAV)。

[0577] 本文所述的具有或不具有一种或多种向导核酸的Cas9可以使用腺相关病毒(AAV)、慢病毒、腺病毒或其他质粒或病毒载体类型来递送,特别是使用来自以下的调配物和剂量:例如美国专利第8,454,972号(腺病毒的调配物、剂量)、美国专利第8,404,658号(AAV的调配物、剂量)和美国专利第5,846,946号(DNA质粒的调配物、剂量);以及涉及慢病毒、AAV和腺病毒的临床试验和关于这些临床试验的出版物。例如,对于AAV,施用途径、调配物和剂量可以如美国专利第8,454,972号中和涉及AAV的临床试验中那样。对于腺病毒,施用途径、调配物和剂量可以如美国专利第8,404,658号和涉及腺病毒的临床试验中那样。对于质粒递送,施用途径、调配物和剂量可以如美国专利第5,846,946号和涉及质粒的临床研究中那样。剂量可以基于或外推到平均70kg的个体(例如,成年男性),并且可以针对不同体重和物种的患者、受试者、哺乳动物进行调整。施用频率在医疗或兽医从业者(例如,医师、兽医)的范围内,这取决于常见因素,包含年龄、性别、一般健康状况、患者或受试者的其他状况以及正在解决的特定病况或症状。病毒载体可以注射到所关注的组织中。对于细胞类型特异性碱基编辑,碱基编辑器和任选的向导核酸的表达可以由细胞类型特异性启动子驱动。

[0578] 对于体内递送,AAV可能优于其他病毒载体。在一些情况下,AAV具有低毒性,这可能是由于纯化方法不需要对可以激活免疫反应的细胞颗粒进行超离心。在一些情况下,AAV导致插入诱变的可能性很小,因为它未整合到宿主基因组中。

[0579] AAV的包装限制为4.5或4.75Kb。大于4.5或4.75Kb的构建体可以导致病毒产量显著降低。例如,SpCas9相当大,基因本身超过4.1Kb,这使其很难包装到AAV中。因此,本公开的实施例包含使用长度比常规Cas9短的公开的Cas9。

[0580] AAV可以是AAV1、AAV2、AAV5或其任何组合。可以根据待靶向的细胞选择AAV的类型;例如,可以选择AAV血清型1、2、5或混合衣壳AAV1、AAV2、AAV5或其任何组合用于靶向脑或神经元细胞;并且可以选择AAV4用于靶向心脏组织。AAV8可用于递送到肝脏。关于这些细胞的某些AAV血清型的列表可见于:Grimm,D.等人,J.Virol.82:5887-5911(2008))。

[0581] 慢病毒是复杂的逆转录病毒,其能够在有丝分裂细胞和有丝分裂后细胞中感染并

表达其基因。最常见的慢病毒是人类免疫缺陷病毒 (HIV), 该病毒使用其他病毒的包膜糖蛋白来靶向广泛范围的细胞类型。

[0582] 慢病毒可以如下制备。克隆pCasES10 (其含有慢病毒转移质粒骨架) 后, 将低传代 ($p=5$) 的HEK293FT接种在T-75烧瓶中, 以在转染前一天在含有10%胎牛血清且不含抗生素的DMEM中达到50%汇合。20小时后, 将培养基更换为OptiMEM (无血清) 培养基, 并在4小时后进行转染。用10 μ g慢病毒转移质粒 (pCasES10) 和以下封装质粒转染细胞: 5 μ g pMD2.G (VSV-g假型) 和7.5 μ g psPAX2 (gag/pol/rev/tat)。可在4mL OptiMEM中用阳离子脂质递送剂 (50 μ l Lipofectamine 2000和100 μ l Plus试剂) 进行转染。6小时后, 将培养基更换为含有10%胎牛血清的不含抗生素的DMEM。这些方法在细胞培养期间使用血清, 但优选为无血清的方法。

[0583] 慢病毒可以如下纯化。48小时后收获病毒上清液。首先清除上清液中的碎片, 并通过0.45 μ m低蛋白结合 (PVDF) 过滤器进行过滤。然后将它们在超速离心机中以24,000rpm旋转2小时。将病毒集结粒在4 $^{\circ}$ C下重悬于50 μ l DMEM中过夜。然后将它们等分并立即在-80 $^{\circ}$ C下冷冻。

[0584] 在另一实施例中, 还考虑了基于马传染性贫血病毒 (EIAV) 的最小非灵长类动物慢病毒载体。在另一实施例中, RetinoStat[®], 一种表达血管抑制蛋白内皮抑素和血管抑素的基于马传染性贫血病毒的慢病毒基因疗法载体, 其预期通过视网膜下注射递送。在另一实施例中, 考虑使用自失活慢病毒载体。

[0585] 系统的任何RNA, 例如向导RNA或编码Cas9的mRNA, 可以以RNA的形式递送。编码Cas9的mRNA可以使用体外转录产生。例如, 可以使用含有以下元件的PCR盒合成Cas9 mRNA: T7启动子、任选的kozak序列 (GCCACC)、核酸酶序列和3'UTR, 诸如来自 β 珠蛋白-polyA尾部的3'UTR。该盒可用于通过T7聚合酶进行的转录。向导多核苷酸 (例如, gRNA) 还可以使用体外转录从含有T7启动子、后跟序列“GG”和向导多核苷酸序列的盒中转录。

[0586] 为了增强表达并降低可能的毒性, 可以修饰Cas9序列和/或向导核酸以包括一种或多种修饰的核苷, 例如使用假-U或5-甲基-C进行修饰。

[0587] 一些实施例中的公开内容包括修饰细胞或生物体的方法。细胞可以是原核细胞或真核细胞。细胞可以是哺乳动物细胞。哺乳动物细胞可以是非人类灵长类动物、牛、猪、啮齿动物或小鼠细胞。通过本公开的碱基编辑器、组合物和方法向细胞引入的修饰可以使得细胞和细胞的后代被改变以改善生物产物 (诸如抗体、淀粉、醇或其他所需的细胞输出) 的生产。通过本公开的方法向细胞引入的修饰可以使得细胞和细胞的后代包括改变, 其使所产生的生物产物改变。

[0588] 该系统可以包括一种或多种不同的载体。在一方面, Cas9经密码子优化以表达期望的细胞类型, 优选真核细胞, 优选哺乳动物细胞或人类细胞。

[0589] 一般而言, 密码子优化是指通过用在宿主细胞的基因中更频繁使用或最频繁使用的密码子替换天然序列的至少一个密码子 (例如约或多于约1、2、3、4、5、10、15、20、25、50或更多个密码子) 同时保持天然氨基酸序列来修饰核酸序列以增强在所关注宿主细胞中的表达的方法。不同的物种对特定氨基酸的某些密码子表现出特定的偏倚。密码子偏倚 (生物体之间密码子使用的差异) 通常与信使RNA (mRNA) 的翻译效率相关, 而该信使RNA的翻译效率又被认为尤其取决于待翻译密码子的特性以及特定转移RNA (tRNA) 分子的可用性。细胞中

所选tRNA的优势通常反映了肽合成中最频繁使用的密码子。因此,可以基于密码子优化来定制基因以在给定生物体中实现最佳基因表达。密码子使用表很容易获得,例如,可在www.kazusa.or.jp/codon/(2002年7月9日访问)上的“密码子使用数据库”中获得,并且这些表可以以多种方式进行改编。参见Nakamura,Y.等人“Codon usage tabulated from the international DNA sequence databases:status for the year 2000”Nucl.Acids Res.28:292(2000)。用于对特定序列进行密码子优化以在特定宿主细胞中表达的计算机算法也是可用的,诸如Gene Forge (Aptagen;Jacobus,Pa.)也是可用的。在一些实施例中,编码工程化核酸酶的序列中的一个或多个密码子(例如,1、2、3、4、5、10、15、20、25、50或更多个密码子,或所有密码子)对应于特定氨基酸最频繁使用的密码子。

[0590] 封装细胞通常用于形成能够感染宿主细胞的病毒颗粒。此类细胞包括封装腺病毒和psi.2细胞的293细胞或封装逆转录病毒的PA317细胞。基因疗法中使用的病毒载体通常是通过生产将核酸载体封装到病毒颗粒中的细胞系来产生。载体通常含有封装和随后整合到宿主中所需的最小病毒序列,其他病毒序列被用于待表达的多核苷酸的表达盒替换。缺失的病毒功能通常由封装细胞系反式提供。例如,用于基因疗法的AAV载体通常仅具有来自AAV基因组的ITR序列,这些序列是封装和整合到宿主基因组中所必需的。病毒DNA可以被封装在细胞系中,该细胞系含有编码其他AAV基因的辅助质粒,即rep和cap,但缺少ITR序列。该细胞系还可能感染了腺病毒作为辅助。辅助病毒可能促进AAV载体的复制和来自辅助质粒的AAV基因的表达。在一些情况下,由于缺乏ITR序列,辅助质粒并没有被大量封装。可以通过例如热处理来减少腺病毒的污染,腺病毒对热处理比AAV更敏感。

[0591] 药物组合物

[0592] 本公开的其他方面涉及包含CRISPR系统(例如,包括本文公开的Cas9)的药物组合物。如本文所用,术语“药物组合物”是指经调配用于药物用途的组合物。在一些实施例中,药物组合物进一步包含药学上可接受的载体。在一些实施例中,药物组合物包含另外的试剂(例如,用于特异性递送、增加半衰期,或其他治疗性化合物)。

[0593] 如此处所用,术语“药学上可接受的载体”意指药学上可接受的材料、组合物或媒介,诸如液体或固体填充剂、稀释剂、赋形剂、制造助剂(例如,润滑剂、滑石粉、硬脂酸镁、硬脂酸钙或硬脂酸锌、或硬脂酸),或溶剂囊封材料,其参与将化合物从身体的一个部位(例如,递送部位)携带或运输到另一部位(例如,器官、组织或身体的一部分)。药学上可接受的载体在与调配物的其他成分相容并且对受试者的组织无害的意义上是“可接受的”(例如,生理相容的、无菌的、生理pH等)。

[0594] 可用作药学上可接受的载体的材料的一些非限制性实例包括:(1)糖,诸如乳糖、葡萄糖和蔗糖;(2)淀粉,诸如玉米淀粉和马铃薯淀粉;(3)纤维素及其衍生物,诸如羧甲基纤维素钠、甲基纤维素、乙基纤维素、微晶纤维素和醋酸纤维素;(4)粉末状黄蓍胶;(5)麦芽;(6)明胶;(7)润滑剂,诸如硬脂酸镁、十二烷基硫酸钠和滑石粉;(8)赋形剂,诸如可可脂和栓剂蜡;(9)油,诸如花生油、棉籽油、红花油、芝麻油、橄榄油、玉米油和大豆油;(10)二醇,诸如丙二醇;(11)多元醇,诸如甘油、山梨醇、甘露醇和聚乙二醇(PEG);(12)酯,诸如油酸乙酯和月桂酸乙酯;(13)琼脂;(14)缓冲剂,诸如氢氧化镁和氢氧化铝;(15)海藻酸;(16)无热原水;(17)等渗盐水;(18)林格氏溶液(Ringer's solution);(19)乙醇;(20)pH缓冲溶液;(21)聚酯、聚碳酸酯和/或聚酐;(22)膨胀剂,诸如多肽和氨基酸;(23)血清醇,诸如乙

醇;以及(23)药物调配物中采用的其他无毒相容物质。润湿剂、着色剂、脱模剂、包衣剂、甜味剂、调味剂、芳香剂、防腐剂和抗氧化剂也可以存在于调配物中。诸如“赋形剂”、“载体”、“药学上可接受的载体”、“媒剂”等术语在本文中可互换使用。

[0595] 药物组合物可包含一种或多种pH缓冲化合物以将调配物的pH维持在反映生理pH的预定水平,诸如在约5.0至约8.0的范围内。用于水性液体调配物中的pH缓冲化合物可以是氨基酸或氨基酸混合物,诸如组氨酸、或氨基酸(诸如组氨酸和甘氨酸)的混合物。替代地,pH缓冲化合物优选是将调配物的pH维持在预定水平(诸如在约5.0至约8.0的范围内)且不螯合钙离子的试剂。此类pH缓冲化合物的说明性实例包括但不限于咪唑和乙酸根离子。pH缓冲化合物可以以适合将调配物的pH维持在预定水平的任何量存在。

[0596] 药物组合物还可以含有一种或多种渗透调节剂,即,将调配物的渗透特性(例如,张力、渗透浓度和/或渗透压)调节至接受者个体的血流和血细胞可接受的水平的化合物。渗透调节剂可以是不螯合钙离子的试剂。渗透调节剂可以是本领域技术人员已知或可获得的调节调配物的渗透特性的任何化合物。本领域技术人员可以凭经验判定给定渗透调节剂用于本发明调配物的适用性。适合类型的渗透调节剂的说明性实例包括但不限于:盐,诸如氯化钠和乙酸钠;糖,诸如蔗糖、右旋糖和甘露醇;氨基酸,诸如甘氨酸;以及这些试剂和/或试剂类型中的一种或多种的混合物。渗透调节剂可以以足以调节调配物的渗透特性的任何浓度存在。

[0597] 在一些实施例中,药物组合物经调配用于递送至受试者,例如以用于基因编辑。施用本文所述的药物组合物的适合途径包括但不限于:局部、皮下、经皮、皮内、病灶内、关节内、腹膜内、膀胱内、经粘膜、牙龈、齿内、耳蜗内、经鼓膜、器官内、硬膜外、鞘内、肌内、静脉内、血管内、骨内、眼周、肿瘤内、脑内和脑室内施用。

[0598] 在一些实施例中,将本文所述的药物组合物局部施用至患病部位。在一些实施例中,本文所述的药物组合物通过注射、借助于导管、借助于栓剂或借助于植入物施用至受试者,该植入物是多孔、无孔或凝胶状材料,包括膜,诸如硅橡胶膜,或纤维。

[0599] 在其他实施例中,本文所述的药物组合物在受控释放系统中递送。在一个实施例中,可以使用泵(参见例如,Langer,1990,Science 249:1527-1533;Sefton,1989,CRC Crit.Ref.Biomed.Eng.14:201;Buchwald等人,1980,Surgery 88:507;Saudek等人,1989,N.Engl.J.Med.321:574)。在另一实施例中,可以使用聚合材料。(参见例如,Medical Applications of Controlled Release(Langer和Wise编,CRC Press,Boca Raton,Fla.,1974);Controlled Drug Bioavailability,Drug Product Design and Performance(Smolens和Ball编,Wiley,New York,1984);Ranger和Peppas,1983,Macromol.Sci.Rev.Macromol.Chem.23:61。还参见Levy等人,1985,Science 228:190;During等人,1989,Ann.Neurol.25:351;Howard等人,1989,J.Neurosurg.71:105。)例如,在Langer,同上文献中讨论了其他受控释放系统。

[0600] 在一些实施例中,根据例程程序将药物组合物调配为适于静脉内施用或皮下施用至受试者(例如,人类)的组合物。在一些实施例中,用于通过注射施用的药物组合物是以无菌等渗用途作为增溶剂和局部麻醉剂(诸如利多卡因)以缓解注射部位处的疼痛的溶液。一般地,将成分分开或混合在一起以单位剂型提供,例如,作为在气密封容器中的干燥冻干粉或无水浓缩剂,该气密封容器诸如指示活性剂的量的安瓿或小药囊。在通过输注施用药物

的情况下,可以用含有无菌药用级水或盐水的输注瓶分配。在药物组合物通过注射施用的情况下,可以提供无菌注射用水或盐水的安瓿,使得可以在施用之前混合成分。

[0601] 用于全身施用的药物组合物可以是液体,例如无菌盐水、乳酸林格氏溶液或汉克氏溶液。此外,药物组合物可以呈固体形式并在使用前立即重新溶解或悬浮。还考虑了冻干形式。药物组合物可以包含在脂质颗粒或囊泡内,诸如脂质体或微晶,其也适合于肠胃外施用。颗粒可以具有任何适合的结构,诸如单层或多层,只要其中包含组合物即可。化合物可以被包埋在含有融合脂质二油酰磷脂酰乙醇胺(DOPE)、低水平(5-10mol%)阳离子脂质的“稳定的质粒-脂质颗粒”(SPLP)中,并通过聚乙二醇(PEG)涂层进行稳定化(Zhang Y.P.等人, *Gene Ther.* 1999, 6:1438-47)。带正电荷的脂质,诸如N-[1-(2,3-二油酰氧基)丙基]-N,N,N-三甲基-甲基硫酸铵或“DOTAP”对于此类颗粒和囊泡是特别优选的。此类脂质颗粒的制备是众所周知的。参见例如,美国专利第4,880,635号、第4,906,477号、第4,911,928号、第4,917,951号、第4,920,016号和第4,921,757号;其中的每一个都通过引用并入本文。

[0602] 例如,本文描述的药物组合物可以作为单位剂量施用或封装。当关于本公开的药物组合物使用时,术语“单位剂量”是指适合作为受试者的单位剂量的物理上离散的单位,每个单位含有经计算以产生与所需稀释剂(即载体或媒剂)结合的期望治疗效果的预定量的活性物质。

[0603] 此外,药物组合物可以以药物试剂盒形式提供,该药物试剂盒包括:(a)含有冻干形式的本发明化合物的容器和(b)含有药学上可接受的稀释剂(例如,用于重构或稀释本发明冻干化合物的无菌稀释剂)的第二容器。任选地,与此类容器相关联的可以是监管药品或生物制品的制造、使用或销售的政府机构规定的形式的通知,该通知反映了制造、使用或销售机构对于人类施用的批准。

[0604] 在另一方面,包括含有可用于治疗上述疾病的材料的制品。在一些实施例中,制品包括容器和标签。合适的容器包含例如瓶子、小瓶、注射器和试管。容器可以由各种材料如玻璃或塑料形成。在一些实施例中,容器容纳可有效治疗本文所述的疾病的组合物并且可以具有无菌进入端口。例如,容器可以是静脉内溶液袋或具有可被皮下注射针刺穿的塞子的小瓶。组合物中的活性剂是本发明的化合物。在一些实施例中,容器上或与容器相关联的标签指示该组合物用于治疗所选疾病。制品可以进一步包括第二容器,该第二容器包括药学上可接受的缓冲液,诸如磷酸盐缓冲盐水、林格氏溶液或葡萄糖溶液。其可以进一步包括从商业和用户角度来看所需的材料,包括其他缓冲液、稀释剂、过滤器、针头、注射器和带有使用说明的包装插页。

[0605] 在一些实施例中,CRISPR系统(例如,包括本文所述的Cas9)作为药物组合物的一部分提供。在一些实施例中,药物组合物包含本文提供的任何融合蛋白(例如,包括本文所述的包含LubCas9的核碱基编辑器)。在一些实施例中,药物组合物包含本文提供的任何复合物。在一些实施例中,药物组合物包含核糖核蛋白复合物,该核糖核蛋白复合物包含与gRNA和阳离子脂质形成复合物的RNA向导的核酸酶(例如,Cas9)。在一些实施例中,药物组合物包含gRNA、核酸可编程DNA结合蛋白、阳离子脂质和药学上可接受的赋形剂。药物组合物可以任选地包含一种或多种另外的治疗活性物质。

[0606] 试剂盒

[0607] 在一个方面,本发明提供含有上述方法和组合物中公开的任何一种或多种元件的

试剂盒。在一些实施例中,试剂盒包括载体系统和使用该试剂盒的说明书。在一些实施例中,载体系统包括用于插入向导序列的一个或多个插入位点,其中当表达时,向导序列指导CRISPR复合物与真核细胞中的靶序列的序列特异性结合,其中CRISPR复合物包含CRISPR酶,其与(1)与靶序列杂交的向导序列和(2)与tracr序列杂交的序列复合;和/或(b)第二调控元件,其可操作地连接至编码所述CRISPR酶的酶编码序列,该酶编码序列包含核定位序列。元件可以各别地或以组合形式提供,并且可以提供在任何适合的容器中,诸如小瓶、瓶子或管。在一些实施例中,试剂盒包括呈一种或多种语言形式的说明书,例如呈多于一种语言形式的说明书。

[0608] 在一些实施例中,试剂盒包括核碱基编辑器。例如,在一些实施例中,试剂盒包括核碱基编辑器,其包括本文所述的Cas9酶(ScoCas9、SirCas9、VapCas9、EpeCas9、LfeCas9、PmaCas9)。

[0609] 在一些实施例中,试剂盒包括在利用本文所述的一种或多种元件的方法中使用的一种或多种试剂。试剂可以提供在任何适合的容器中。例如,试剂盒可以提供一种或多种反应或储存缓冲液。试剂可以以可用于特定测定的形式提供,或者以需要在使用前添加一种或多种其他组分的形式(例如以浓缩物或冻干形式)提供。缓冲液可以是任何缓冲液,包括但不限于碳酸钠缓冲液、碳酸氢钠缓冲液、硼酸盐缓冲液、Tris缓冲液、MOPS缓冲液、HEPES缓冲液及其组合。在一些实施例中,缓冲液是碱性的。在一些实施例中,缓冲液具有约7至约10的pH。在一些实施例中,试剂盒包括一种或多种对应于向导序列的寡核苷酸,其用于插入到载体中以便可操作地连接向导序列和调控元件。在一些实施例中,试剂盒包括同源重组模板多核苷酸。

[0610] 本文提及的所有出版物、专利申请、专利和其它参考文献通过引用整体并入本文。另外,材料、方法和实例仅是说明性的并且不旨在进行限制。除非另外定义,否则本文中使用的所有技术术语和科学术语具有与本发明所属领域的普通技术人员通常理解的相同的含义。尽管与本文所述的方法和材料类似或等效的方法和材料可以用于实践或测试本发明,但本文描述了适合的方法和材料。

[0611] 实例

[0612] 以下实例描述了制备和实施本发明的一些优选模式。然而,应当理解,这些实例仅用于说明目的,并不意味着限制本发明的范围。

[0613] 实例1. 针对新颖Cas9酶的筛选、新颖Cas9酶的发现和优化

[0614] 此实例描述了用于发现新颖Cas9酶的筛选。如本文所述,使用该筛选,分离并优化来自星座链球菌、夏普氏菌属分离株RUG017、小韦荣球菌、秘鲁埃扎基拉菌、发酵乳杆菌菌株AF15-40LB和嗜脲菌属马赛-P3761细菌的新颖Cas9酶。

[0615] 在发现新Cas9酶(识别新颖PAM序列)的寻找过程中,使用生物信息学筛选来寻找其他酶以扩增CRISPR的靶向范围。该筛选利用来自酿脓链球菌、金黄色葡萄球菌、嗜热链球菌和新凶手弗朗西丝菌(*F. novicida.*)的Cas9的种子序列。使用BLAST的tblastn变型进行生物信息学分析,考虑BLAST命中的e值阈值为 $1e-6$ 。简而言之,经选择用于测试的基因座是在存在来自其他物种的Cas9蛋白的情况下保持完整的基因座。选择基因座使其在CRISPR阵列内具有大于三个间隔子,并且在Cas9的5'有大于1kb的内源序列且在CRISPR阵列的3'有大于300nt。使用这种方法,从不同的细菌物种中鉴定出新颖Cas9酶,并经密码子优化以在

人类细胞中表达。然后重组生产新颖工程化Cas9酶并加以测试。

[0616] 实例2. 鉴定来自星座链球菌、夏普氏菌属分离株RUG017、小韦荣球菌、秘鲁埃扎基拉菌、发酵乳杆菌菌株AF15-40LB和嗜脲菌属-马赛-P3761细菌的新颖Cas9酶的3'PAM共有基序

[0617] 此实例说明了最初从星座链球菌、夏普氏菌属分离株RUG017、小韦荣球菌、秘鲁埃扎基拉菌、发酵乳杆菌菌株AF15-40LB和嗜脲菌属马赛-P3761物种中分离出来的人类密码子优化Cas9的前间隔序列邻近基序(PAM)序列的鉴定。

[0618] 使用体外PAM鉴定测定测试了人类、密码子优化Cas9对PAM序列的识别。将带有随机PAM序列的质粒文库与从不同细菌中分离出来的Cas9一起孵育。对未切割的质粒进行纯化和测序,以鉴定被切割的特定PAM基序。由星座链球菌Cas9识别的共有PAM序列经鉴定为5'-NGG-3'(图1A)。由夏普氏菌属分离株RUG017 Cas9识别的共有PAM序列经鉴定为5'-NAGHC-3'(图1B)。由小韦荣球菌Cas9识别的共有PAM序列经鉴定为5'-NRHRRH-3'(H=A、C或T;R=A或G)(图1C)。由秘鲁埃扎基拉菌Cas9识别的共有PAM序列经鉴定为5'-NGG-3'(图1D)。由发酵乳杆菌菌株AF15-40LB Cas9识别的共有PAM序列经鉴定为5'-NNAAA-3'(图1E)。由嗜脲菌属马赛-P3761 Cas9识别的共有PAM序列经鉴定为5'-NGG-3'(图1F)。

[0619] 实例3. 预测来自星座链球菌、夏普氏菌属分离株RUG017、小韦荣球菌、秘鲁埃扎基拉菌、发酵乳杆菌菌株AF15-40LB和嗜脲菌属马赛-P3761细菌的新颖Cas9酶的sgRNA的RNA折叠结构

[0620] 此实例展示了与新颖Cas9酶一起使用的包含crRNA和tracrRNA的例示性sgRNA的预测RNA折叠结构。

[0621] 对源自异源表达Cas9 Crispr基因座的大肠杆菌菌株的RNA进行小RNA测序。简而言之,首先将大肠杆菌重悬于Trizol中,然后在匀浆器中用氧化锆/二氧化硅珠匀浆细菌,持续三个1min的循环,从静止期细菌中分离出RNA。从匀浆样品中纯化总RNA,经DNase处理并用T4多核苷酸激酶进行3'去磷酸化,并去除rRNA。RNA文库由rRNA耗尽的RNA制备,并针对小RNA选择大小。

[0622] 对于RNA测序,将转录物用大肠杆菌Poly(A)聚合酶加poly-A尾,使用T4 RNA连接酶1与5'RNA衔接子连接并反转录,然后使用带有条形码的引物对cDNA进行PCR扩增,并在MiSeq上测序。基于相关联的条形码鉴定来自每个样品的读段,并使用BWA与参考序列进行比对。使用Picard工具将配对末端比对用于提取转录物序列,并使用Geneious软件对序列进行分析。

[0623] RNA折叠基于来自Geneious 11.1.2软件的预测。单个sgRNA转录物使crRNA与tracrRNA融合,模仿引导位点特异性Cas9活性所需的双RNA结构。与来自星座链球菌的ScoCas9一起使用的嵌合sgRNA的预测RNA折叠结构示于图2A中,与来自夏普氏菌属分离株RUG017的SirCas9一起使用的sgRNA示于图2B中,与来自小韦荣球菌的VapCas9一起使用的sgRNA示于图2C中,与来自秘鲁埃扎基拉菌的EpeCas9一起使用的sgRNA示于图2D中,与来自发酵乳杆菌菌株AF15-40LB的LfeCas9一起使用的sgRNA示于图2E中,并且与来自嗜脲菌属马赛-P3761的PmaCas9一起使用sgRNA示于图2F中。

[0624] 实例4. HEK293T细胞中利用WT ScoCas9的离体切割活性

[0625] 此实例说明了HEK293T细胞中利用来自星座链球菌的WT ScoCas9的离体核酸切割

活性。

[0626] 将HEK293T细胞铺在96孔板中。铺板后24小时,用含有Cas9和向导RNA(表10)的表达载体转染细胞。转染后72小时收获细胞并提取总DNA。

[0627] 进行深度测序来表征HEK293T细胞中的插入缺失模式。简而言之,使用两轮PCR来扩增例示性靶标(表8),以将Illumina衔接子以及独特的条形码添加到靶标扩增子。将PCR产物在2%凝胶上运行并提取凝胶。汇集样品,进行定量,并制备cDNA文库并在MiSeq上进行测序。通过深度测序确定插入缺失频率(图3)。

[0628] 表8. 例示性向导RNA序列和PAM序列

[0629]	ID (Sco/Pma)	5'→3'向导序列	3' PAM
	向导 2	GAAACAATGATAACAAGACC (SEQ ID NO: 97)	TGG
	向导 3	GTGGCCCCCTGTGCCCAGCCC (SEQ ID NO: 98)	TGG
	向导 4	GTCCCAAATATGTAGCTGTT (SEQ ID NO: 99)	TGG
	向导 6	GCTCCCATCACATCAACCGG (SEQ ID NO: 100)	TGG
	向导 7	GATGTCACCTCCAATGACTA (SEQ ID NO: 101)	GGG
	向导 9	GTTGAAGATGAAGCCCAGAG (SEQ ID NO: 102)	CGG
	向导 10	GCCAACACCAACCAGAACTT (SEQ ID NO: 103)	GGG
	向导 11	TGCTGCACACAGCAGGCCTT (SEQ ID NO: 104)	TGG

[0630] 数据显示,WT ScoCas9实现2%-32%之间的插入缺失频率。向导RNA 2和9产生超过30%的插入缺失突变,而向导RNA 11产生约2%的插入缺失突变。

[0631] 实例5.通过Cas9酶与腺嘌呤碱基编辑器(ABE)或胞苷碱基编辑器(CBE)N末端融合进行碱基编辑

[0632] 此实例说明了与腺嘌呤碱基编辑器(ABE)或胞苷碱基编辑器(CBE)融合的Cas9酶的碱基转换效率。

[0633] 简而言之,每96孔铺板25,000个HEK293T细胞。铺板后24h转染100ng Cas9表达质粒和100ng向导表达质粒。转染后5天收获细胞并提取DNA。

[0634] 进行深度测序以表征HEK293T细胞中A至G的转化或C至T的转化。使用两轮PCR区域扩增例示性靶标,以将Illumina衔接子以及独特的条形码添加到靶标扩增子。将PCR产物在2%凝胶上运行并提取凝胶。汇集样品,进行定量,并制备cDNA文库并在MiSeq上进行测序。通过对N末端以及C末端TadA8融合构建体进行深度测序来确定A至G转化的百分比。通过对N末端以及C末端ppAPOBEC1融合构建体进行深度测序来确定C至T转化的百分比。

[0635] 图4A显示在N末端处与ABE或CBE融合的ScoCas9的构建体的示意图。表9显示与ScoCas9一起使用的向导RNA序列。图4B显示插入缺失突变和通过使ScoCas9与腺嘌呤碱基编辑器(ABE)进行N末端融合实现的靶向腺嘌呤至鸟嘌呤转化百分比的图(图4B),其针对人类细胞系(HEK293T)中的基因组位点。图4C显示插入缺失突变和通过使ScoCas9与胞苷碱基编辑器进行N末端融合实现的靶向胞嘧啶至胸腺嘧啶转化百分比的图(图4C),其针对人类细胞系(HEK293T)中的基因组位点。

[0636] 表9. 与ScoCas9一起使用的向导RNA序列和PAM序列

[0637]	ID (Sco)	5'→3'向导序列	3' PAM
	向导 1	GAACACAAAGCATAGACTGC (SEQ ID NO: 105)	GGG
	向导 2	GAAACAATGATAACAAGACC (SEQ ID NO: 106)	TGG
	向导 3	GTGGCCCCTGTGCCCAGCCC (SEQ ID NO: 107)	TGG
	向导 4	GTCCCAAATATGTAGCTGTT (SEQ ID NO: 108)	TGG
	向导 5	AGAGGGACACACAGATCTAT (SEQ ID NO: 109)	TGG
	向导 6	GCTCCCATCACATCAACCGG (SEQ ID NO: 110)	TGG
	向导 7	GATGTCACCTCCAATGACTA (SEQ ID NO: 111)	GGG
	向导 8	GGGCAACCACAAACCCACGA (SEQ ID NO: 112)	GGG
	向导 9	GTTGAAGATGAAGCCCAGAG (SEQ ID NO: 113)	CGG
[0638]	向导 10	GCCAACACCAACCAGAACTT (SEQ ID NO: 114)	GGG
	向导 11	TGCTGCACACAGCAGGCCTT (SEQ ID NO: 115)	TGG
	向导 12	GTGCCAGAAACAGGGGTGAC (SEQ ID NO: 116)	GGG

[0639] 图5A显示WT SirCas9以及在N末端处与ABE融合的SirCas9 (“D14A”突变体)的构建体的示意图。表10显示与SirCas9一起使用的例示性NAGMC向导RNA序列。图5B显示插入缺失突变和通过使SirCas9与腺嘌呤碱基编辑器 (ABE) 进行N末端融合实现的靶向腺嘌呤至鸟嘌呤转化百分比的图 (图5B), 其针对人类细胞系 (HEK293T) 中的基因组位点。

[0640] 表10. 与SirCas9一起使用的向导RNA序列和PAM序列

[0641]

ID (Sir)	5'→3'序列	3' PAM
向导 1	CCTGCCTCAGCTGCTCACTT (SEQ ID NO: 117)	GAGCC
向导 2	AAACGGTCCCCAGAGGGTTC (SEQ ID NO: 118)	TAGAC
向导 3	GCCACCGGTTGATGTGATGG (SEQ ID NO: 119)	GAGCC
向导 4	AAGTGGTCCCAGGCCTCAGC (SEQ ID NO: 120)	CAGCC
向导 5	AGAGAAAATGAACTTTCAA (SEQ ID NO: 121)	AAGCC
向导 6	CCAAACCCAACTCCATCTAC (SEQ ID NO: 122)	CAGCC
向导 7	GGTCCTTGAATTGCAGTATC (SEQ ID NO: 123)	TAGCC
向导 8	GCATAGACTGCGGGGCGGGC (SEQ ID NO: 124)	CAGCC
向导 9	GGAAACTGGAACACAAAGCA (SEQ ID NO: 125)	TAGAC
向导 10	GACAGCATGTGGTAATTTTC (SEQ ID NO: 126)	CAGCC
向导 11	GCCCCCGGAACTCTGTCCA (SEQ ID NO: 127)	GAGAC
向导 12	TCGACCCCCACCAAGGTTCA (SEQ ID NO: 128)	CAGCC

[0642] 图6A显示构建体的示意图,这些构建体显示WT VapCas9以及在N末端处与ABE或CBE融合的VapCas9(“D38A”突变体)。表11显示与VapCas9一起使用的例示性NRHRRH[其中H是腺嘌呤、胞嘧啶或胸腺嘧啶,并且R是腺嘌呤或鸟嘌呤]向导RNA序列。图6B显示插入缺失突变和通过使VapCas9与腺嘌呤碱基编辑器(ABE)进行N末端融合实现的靶向腺嘌呤至鸟嘌呤转化百分比以及通过使VapCas9与胞嘧啶碱基编辑器(CBE)进行N末端融合实现的靶向胞嘧啶至胸腺嘧啶转化百分比的图(图6B),其针对人类细胞系(HEK293T)中的基因组位点。

[0643] 表11. 与VapCas9一起使用的向导RNA序列和PAM序列

[0644]

ID (Vap)	5'→3'序列	3' PAM
向导 1	TGTTAACAGCTGACCCAATA (SEQ ID NO: 129)	AGTGGC
向导 2	GTTACTCGCCTGTCAAGTGG (SEQ ID NO: 130)	CGTGAC
向导 3	GGGCTCCCATCACATCAACC (SEQ ID NO: 131)	GGTGGC
向导 4	GCTTTGGGGAGGCCTGGAGT (SEQ ID NO: 132)	CATGGC
向导 5	TAGCTGCCAATGACTATAGC (SEQ ID NO: 133)	AATAGC
向导 6	TTAAAATAGGATCTACATCA (SEQ ID NO: 134)	CGTAAC
向导 7	GAATCCTGCCATACACTTTG (SEQ ID NO: 135)	AATAGC
向导 8	CTGCGGGGCGGGCCAGCCTG (SEQ ID NO: 136)	AATAGC
向导 9	ACATTGTCAGAGGGACACAC (SEQ ID NO: 137)	TGTGGC
向导 10	AGCAACTCCAGTCCCAAATA (SEQ ID NO: 138)	TGTAGC
向导 11	GTGGTGGCCGAGCGCCCCCT (SEQ ID NO: 139)	AGTGAC
向导 12	CATTCACCCAGCTTCCCTGT (SEQ ID NO: 140)	GGTGGC

[0645] 图7A显示构建体的示意图,这些构建体显示ABE与VapCas9进行的N末端融合以及ABE与VapCas9进行的C末端融合。图7B显示通过与腺嘌呤碱基编辑器 (ABE) 进行的N末端融合和C末端融合实现的靶向腺嘌呤至鸟嘌呤转化百分比的图。

[0646] 图8A显示构建体的示意图,这些构建体显示ABE和CBE与EpeCas9进行的N末端融合。表12显示与EpeCas9一起使用的例示性向导RNA序列。图8B显示插入缺失突变的图、通过与ABE进行的N末端融合实现的靶向腺嘌呤至鸟嘌呤转化百分比以及通过与CBE的进行的N末端融合实现的靶向胞嘧啶至胸腺嘧啶转化百分比的图。

[0647] 表12. 与EpeCas9一起使用的向导RNA序列和PAM序列

[0648]

ID (Epe)	序列	PAM
向导 1	GAACACAAAGCATAGACTGC (SEQ ID NO: 141)	GGG
向导 2	GAAACAATGATAACAAGACC (SEQ ID NO: 142)	TGG
向导 3	GTGGCCCCTGTGCCCAGCCC (SEQ ID NO: 143)	TGG
向导 4	GTCCCAAATATGTAGCTGTT (SEQ ID NO: 144)	TGG
向导 5	AGAGGGACACACAGATCTAT (SEQ ID NO: 145)	TGG
向导 6	GCTCCCATCACATCAACCGG (SEQ ID NO: 146)	TGG

[0649]

向导 7	GATGTCACCTCCAATGACTA (SEQ ID NO: 147)	GGG
向导 8	GGGCAACCACAAACCCACGA (SEQ ID NO: 148)	GGG
向导 9	GTTGAAGATGAAGCCCAGAG (SEQ ID NO: 149)	CGG
向导 10	GCCAACACCAACCAGAACTT (SEQ ID NO: 150)	GGG
向导 11	TGCTGCACACAGCAGGCCTT (SEQ ID NO: 151)	TGG
向导 12	GTGCCAGAAACAGGGGTGAC (SEQ ID NO: 152)	GGG

[0650] 图9A显示构建体的示意图,这些构建体显示WT LfeCas9和在N末端处与ABE和CBE融合的LfeCas9 D9A突变体。表13显示与LfeCas9一起使用的例示性向导RNA序列。图9B显示示出用LfeCas9实现的插入缺失突变频率的结果的图。图9C显示通过使LfeCas9与腺嘌呤碱基编辑器进行的N末端融合实现的靶向腺嘌呤至鸟嘌呤转化的图。图9D显示用包含CBE的碱基编辑器实现的靶向胞嘧啶至胸腺嘧啶转化的图,该CBE与LfeCas9 D9A突变体的N末端融合。

[0651] 表13. 与LfeCas9一起使用的向导RNA序列和PAM序列

[0652]

<u>ID (Lfe)</u>	<u>序列</u>	<u>PAM</u>
向导 1	TCACGGAGACTGAACACTCC (SEQ ID NO: 153)	TCAAA
向导 2	GTAACAGACATGGACCATCA (SEQ ID NO: 154)	GGAAA
向导 3	GGGAGGGAGGGGCACAGATG (SEQ ID NO: 155)	AGAAA
向导 4	TGTGGTTCCAGAACCGGAGG (SEQ ID NO: 156)	ACAAA
向导 5	AATGAGAGAAAATGAACTT (SEQ ID NO: 157)	TCAAA
向导 6	GGCCATCAAGGATGCCACG (SEQ ID NO: 158)	AGAAA
向导 7	AAATTGTCCAGCCCCATCTG (SEQ ID NO: 159)	TCAAA
向导 8	CCTGTAAAGGAACTGGAAC (SEQ ID NO: 160)	ACAAA
向导 9	TACATGAAGCAACTCCAGTC (SEQ ID NO: 161)	CCAAA
向导 10	AAACTCCCCCACCCTTT (SEQ ID NO: 162)	CCAAA
向导 11	GAGTTGGGTTTGGTGCTCAA (SEQ ID NO: 163)	TGAAA
向导 12	GCGGGCCAGCCTGAATAGCT (SEQ ID NO: 164)	GCAAA

[0653] 图10A显示构建体的示意图,这些构建体显示WT PmaCas9以及在N末端和C末端处与ABE和CBE融合的PmaCas9 D12A突变体。图10B显示示出用包含ABE或CBE的碱基编辑器实现的A至G或C至T转化的结果的图,该ABE或CBE与PmaCas9 D12A突变体的N末端或C末端融合。

[0654] 表14. 与PmaCas9一起使用的向导RNA序列和PAM序列

[0655]

<u>ID (Pma)</u>	<u>序列</u>	<u>PAM</u>
向导 2	GAAACAATGATAACAAGACC (SEQ ID NO: 165)	TGG
向导 3	GTGGCCCCTGTGCCCAGCCC (SEQ ID NO: 166)	TGG
向导 4	GTCCCAAATATGTAGCTGTT (SEQ ID NO: 167)	TGG
向导 6	GCTCCCATCACATCAACCGG (SEQ ID NO: 168)	TGG
向导 7	GATGTCACCTCCAATGACTA (SEQ ID NO: 169)	GGG
向导 9	GTTGAAGATGAAGCCCAGAG (SEQ ID NO: 170)	CGG
向导 10	GCCAACACCAACCAGAACTT (SEQ ID NO: 171)	GGG
向导 11	TGCTGCACACAGCAGGCCTT (SEQ ID NO: 172)	TGG

[0656] 表15公开用于碱基编辑功能的例示性Cas9腺苷或腺嘌呤和胞嘧啶或胞苷碱基编辑器的序列。

[0657] 表15. 例示性Cas9腺苷或腺嘌呤以及胞嘧啶或胞苷碱基编辑器的序列

[0658]

<u>序列 ID 号(描述)</u>	<u>DNA 切割测定的组成部分</u>
腺嘌呤脱氨酶，与切口酶 <i>ScoCas9</i> 的 N 末端融合的 <i>TadA8.13m</i> 切口酶 (ABE-n<i>ScoCas9</i>, D10A 突变体) 的氨基酸序列	
<u>MPAAKRVKLDGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNR</u> <u>AIGLHDPTAHAEIMALRQGGLVMQNYRLYDATLYVTTFEPCVMCAGAMIHRSRIGRVVFGV</u> <u>RNAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQS</u> <u>STDGSSGSETPGTSESATPESSGPKKKRKVGGKPYSIGLAIGTNSVGWAVVTDDYK</u> VPAKKMKVLGNTDKQSIKKNLLGALLFDSGETAEATRLKRTARRRYTRRKNRLRY LQEIFTGEMNKVDENFFQRLDDSFVDEDKRGHHPIFGNIAAEVKYHDDFPTIYH LRRHLADTSKKADLRLVYLALAHMIKFRGHFLYEGDLKAENTDVQALFKDFVEEY DKTIEESHLSEITVDALSILTEKVS KSSRLNIAHYPTEKKNTLFGNLIALSLDLHPN FKTNFQLSEDAKLQFSKDTYEEDLEGFLGEVGDEYADLFASAKNLYDAILLSGILTV DDNSTKAPLSASMVKRYEEHQKDLKKLKDFIKVNAPDQYNAIFKDKNKKGYASYI ESGVKQDEFYKYLKGILLKINGSGDFLDKIDREDFLRKQRTFDNGSIPHQIHLQEMH AILRRQGEHYPFLKENQDKIEKILTRIPYYVGPLARKGSRFAWA EYKADEKITPWN FDDILDKEKSAEKFITRMTLNDLYLPEEKVLPKHSPLYEAFTVYNELTKVKYVNEQ GEAKFFDTNMKQEIFDHVFKENRKVTKDKLLNYLNKEFEFRIVNLTGLDKENKA	

FNSSLGTYHDLRKILDKSFLDDKANNEKTIEDIIQTLTLFEDREMIRQRLQKYSDIFTK
AQLKKLERRHYTGWGRLSYKLINGIRNKENKKTILDYLIDDGYANRNFMLINDD
ALSFKEEIARAQIIDDVDDIANVVHDLPGSPAIIKKGILQSVKIVDELVKVMGHN PANI
IEMARENQTTDKGRRNSQQLKLLQDSLKNLNDNPVNIKNVENQQLQNDRLFLYYI
QNGKDMYTGETLDINNLSQYDIDHIIPQAFIKDNSLDNRVLTRSDKNRGKSDDVPSI
EVVHEMKSFSKLLSVKLITQRKFDNLTKAERGGTEEDKAGFIKRQLVETRQITK
HVAQILDERFNTEFDGNKRRIRNVKIITLKS NLVSNFRKEFELYKVREINDYHHAHD
AYLNAVVG NALLLKYPQLEPEFVYGEY PKYNSYRSRKSATEKFLFY SNILRFFKKE
DIQTNEEDGEIAWNKEKHIKILRKVLSYPQVNIVKKTEEQTGGFSKESILPKGESDKLI
PRKTKNSYWDPKKYGGFDSPVVAYSILVFADVEKGKSKKL RKVQDMVGITIMEKK
RFEKNPVDFLEQRGYRNVRLKIIKLPKYSLFELNKRRL LASAKELQKGNELVIP
QRFTTLLYHSYRIEKDYEPHREYVEKHKDEFKELLEYISVFSRKYVLADNNLT KIE
MLFSKNKDAEVSSLAKSFISLLTFTA FGAPAAFNFFGENIDRKRYTSVTECLNATLIH
QSITGLYETRIDL SKLGEDGKRPAATKKAGQAKKKKGSYPYDVPDYAYPYDVPDYAY
PYDVPDYA (SEQ ID NO: 20)

胞苷脱氨酶，与切口酶 nScoCas9 的 N 末端融合的 ppAPOBEC1 (CBE-nScoCas9,
D10A 突变体) 的氨基酸序列

[0659]

MPAAKRVKLDTSEKGPSTGDPTLRRRIESWEFDVFYDPRELKETCLLYEIKWGMSRK
IWRSSGKNNTTNHVEVNFIIKFTSERRFHSSISCSITWFLSWSPCWECSOAIREFLSQHPGV
TLVIYVARLFWHMDQRNRQGLRDLVNSGVTIQIMRASEYYHCWRNFVNYPPGDEAHWP
QYPPLWMMLYALELHCHILSLPPCLKISRRWQNH LAFFRLHLQNCHYQTIPPHILLATGLI
HPSVTWRLKSGGSSGGSSGSETPGTSESATPESSGGSSGGSPKKKRKVGGKPYSIGL
AIGTNSVGWAVVTDDYKVP AKKMV LGNTDKQSIKKNLLGALLFDSGETAEATRL
KRTARRRYTRRKNRLRYLQEIFTGEMNKVDENFFQRLDDSFLVDEDKRGEHHPFG
NIAAEVKYHDDFPTIYHLRRHLADTSKKADLRLVYLALAHMIKFRGHFLYEGDLK
AENTDVQALFKDFVEEYDKTIEESHLS EITVDALSILTEKVS KSRL ENLIAHYPT EK
KNTLFGNLIALSLDLHPNFKT NFQLSEDAKLQFSKDTYEEDLEGFLGEVGDEYADL
FASAKNLYDAILLSGILTVDNSTKAPLSASMVKRYEEHQKDLKKLKDFIKVNAPD
QYNAIFKDKNKKGYASYIESGVKQDEFYKYLKGILLKINGSGDFLDKIDREDFLRK
QRTFDNGSIPHQIHLQEMHAILRRQGEHY PFLKENQDKIEKILTFRIPYYVGPLARK
GSRFAWA EYKADEKITPWNFDDILDKEKSAEKFITRMTLNDLYLPEEKVLPKHSPLY
EAF TVYNELTKVKYVNEQGEAKFFDTNMKQEIFDHVFKENRKVTKDKLLNYLNK
EFEEFRIVNLTGLDKENKAFNSSLGTYHDLRKILDKSFLDDKANNEKTIEDIIQTLTLF
EDREMIRQRLQKYSDIFTKAQLKKLERRHYTGWGRLSYKLINGIRNKENKKTILDY

LIDDDGYANRNFMLINDDALSFKEEIARAQIIDDVDDIANVVHDLPGSPAIIKKGILQS
 VKIVDELVKVMGHNPANIIHEMARENQTTDKGRRNSQQRLKLLQDSLKNLDNPVNI
 KNVENQQQLQNDRLFLYYIQNGKDMYTGETLDINNLSQYDIDHIIPQAFIKDNSLDN
 RVLTRSDKNRGKSDDVPSIEVVHEMKSFSWKLKSVKLITQRKFDNLTKAERGGLTE
 EDKAGFIKRQLVETRQITKHVAQILDERFNTEFDGNKRRIRNVKIITLKSNLVSNFRK
 EFELYKVREINDYHHAHDAYLNAVVGNALLLKYPQLEPEFVYGEYPKYNSYRSRK
 SATEKFLFYSNILRFFKKEDIQTNEEDGEIAWNKEKHILRKVLSYPQVNIVKKTEEQ
 TGGFSKESILPKGESDKLIPRKTNSYWDPKKYGGFDSPVVAYSILVFADVEKGKSK
 KLRKVQDMVGITIMEKKRFEKNPVDFLEQRGYRNVRLKIIKLPKYSLELENKRR
 RLLASAKELQKGNELVIPQRFTTLLYHSYRIEKDYEPHREYVEKHKDEFKELLEIYI
 SVFSRKYVLADNNLTKIEMLFSSKNKDAEVSSLAKSFISLLTFTAFGAPAAFNFFGENI
 DRKRYTSVTECLNATLIHQSIITGLYETRIDLSKLGEDGKRPAATKKAGQAKKKKG
SSGGSGSGSGSTNLSDIIEKETGKQLVIOESILMLPEEVEEVIGNKPESDILVHTAYD
ESTDENVMMLTSDAPEYKPWALVIODSNGENKIKMLSGSGSGSGSTNLSDIIEKE
TGKQLVIOESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMMLTSDAPEYKPW
ALVIODSNGENKIKMLYPYDVDPDYAYPYDVDPDYAYPYDVDPDYA (SEQ ID NO: 21)

[0660]

腺嘌呤脱氨酶,与切口酶 *SirCas9* 的 N 末端融合的 TadA8.13m(ABE-n*SirCas9*, D14A 突变体)的氨基酸序列。

MPAAKRVKLDGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVNNRVIGEGWNR
AIGLHDPTAHAEIMALRQGGLVMQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGV
RNAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQS
STDGSSGSETPGTSESATPESSGPKKKRKVGAKNKDIRYSIGLAIGTNSVGWAVMD
 EHYELLKKGNHMMWGSRLFDAAEPAATRRASRSIRRRYNKRRERIRLLRDLLGDM
 VMEVDPTFFIRLLNVSFLDEEDKQKQNLGNDYKDNYNLFIEKDFNDKTYDYDKYPTIY
 HLRKELCENKEKADPRLIYLALHHIVKYRGNFLKEGQSFQAKVYEDIEEKLDNTLKK
 FMSLNDLDNLFDNDINSMITVLSKIYQRSKKADDLLKIMNPTKEERAAYKEFTKA
 LVGLKFNVSKMILAQEVKKDDKDIELDFSNVDYDSTVDGLQAEIGEYIEFIEMLS
 INSWVELQDILGNNSTISAAMVERYEEHKNDLRVLKKVIREELPDKYNEVFREDNP
 KLHNYLGYIKYPKNTPVEEFYIYIKRLLAKVDTGEAREILERIDLEKFMLKQNSRT
 NGSIPYQMOKDEMIQIHDNQSVYYPQLKENREKLISILEFRIPYFYGPLNTHSEFAWI
 KKFEDKQKERILPWNVDQIVDIDATAEGFIERMQNTGTYPDKPVMKNSLTVSKF
 EVLNELNKIRINGKLIPVETKKELLSDFMKNKTITDKKLKDWLVTHQYYDTNEEL
 KIEGYQKDLQFSTSLAPWIDFTKIFGEINASNYQLIEKIIYDISIFEDKKILKRRLLKKV
 YQLDDLLVDKILKLNYTGWSRLSEKLLTGIKSKNSKETILSILENSNMNLMEIINDES

LGFKQIIEESNKKDIEGPFRYDEVKKLAGSPAIKRGIWQALLVVQEITKFMKHEPSHI
YIEFAREEQEKVRTESRIAKLQKIYKDLNLQTKEDQLVYESLKKEDAKKKIDTDAL
YLYYLQMGKSMYSGKPLDIDKLSTYHIDHILPRSLIKDDSLDNRVLVLPKENEWKL
DSETVPFEIRNKMMGFWQKLHENGMSNKKFFSLIRTDNFNEKDKKRFINRQLVETR
QIIKNVAVIINDHYTNTNVTVRAELSHQFRERYKIYKNRDLNDLHHAHDAYIACIL
GQFIHQNFNMDVNMIIYGQYKKNYKKDVQEHNNYGFILNSMNIHFNDNSVIW
DPSYIGKIKSCFCYKDVYVTKKLEQNDAKLFDLTILPSDKNSENGVTAKIPVNKY
RKDVNKYGGFSGDAPIMLAIEADKGKKHVRQVIAFPLRLKNYNDEERIKFIEKEKN
LKNVKILTEVKKNQLILINHQYFFITGTNELVNATQLKLSAKNTKNLNFNLVDANKH
NKLESIDDANFNEVIQELICKLQEPIYSRYNSIGKEFEDSYEKINAVTKQDKLYIIEYL
IAIMSAKATQGYIKPELAREIGTNGKNKGRIKSFTIDLNKTTFISTSVTGLFSKKYKL
GKRPAATKKAGQAKKKKGSYPYDVPDYAYPYDVPDYAYPYDVPDYA (SEQ ID NO:
6)

腺嘌呤脱氨酶，与切口酶 *VapCas9* 的 N 末端融合的 TadA8.13m (ABE-nVapCas9,
D38A 突变体) 的氨基酸序列

[0661]

MPAAKRVKLDGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVNNRVIGEGWNR
AIGLHDPTAHAEIMALRQGGLVMQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGV
RNAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQS
STDGSSGSETPGTSESATPESSGPKKKRKVGSIINFQRRGLMETQASNQLISSHLKG
YPIKDYFVGLAIGTSSVGWAVTNKAYELLKFRSHKMWGSRLFDEGESAVARRGFRS
MRRRLERRKLRLKLEELFADAMAQVDPTFFMRLRESKYHYEDKTTGHSSKHILFI
DKNYNDQDYFKEYPTVYHLRSELMKSGTDDIRKLFLAVHHILKYRGNFLYEGATF
DSNASTLDDVIKQALENITFNCFCDCNSAIISSIGQILMEAGKTKSDKAKAIEHLVDTYI
ATDVTDTSSKTQKDQVKEDKKRLKAFANLVLGLNASLIDLFSGSVEELEEDLKKLQI
TGDTYDDKRDELAKAWSDEIYIIDDCKSVYDAIILLSIKEPGLTISESKVKAFNKHK
DDLAILKSLLKSDRSIYNTMFKVDEKGLHNYVHYIKQGRTEETSCNREDFYKYTKK
IVEGLSDSKDKEYILSQIELQILLPLQRIKDNGVIPYQLHLEELKAILAKCGPKFPFLN
EVADGFSVAEKLIKMLEFRIPYYVGPLNTHHNVDNNGFAWAVRKASGRVTPWNFD
DKIDREKSAAAFIKNLTNKCTYLLGEDVLPKSSLLYSEFMLLNELNNVRIDGKPLEK
VVKEHLIEAVFKQDHKKMTKNRIEQFLKDNNGYISETHKHEITGLDGEIKNDLASR
DMVRILGDGFDRSMAEEIITDITIFGESKKMLRETLRKKFASCLDDEAIKKLTKLRY
RDWGRLSQKLLNGIEGCDKAGDGPETHILMRNFSYNLMELLGDKFSFMERIQEIN
AKLTEGQIVNPHDIIDDLALSPAVKRAVWQALRIVDEVAHIKKALPARIFVEVTRSN
KNEKKKKDSRQKRLSDLYAAIKKDDVLLNGLNNEIFGELKSSLAKYDDAALRSKK

LYLYYTQMGRCAYTGEIHELSSLNTDNYDIDHIYPRSLTKDDSFNVLVCKRTANAQ
KSDAYPISEEIQKTQKPFWTFLKQQGLISERKYERLTRITPLTADDLSGFIARQLVETN
QSVKAATTLLRRLYPGVDVVFVKAENVTDFRHDNNFIKVRSLNHHHHAKDAYLNI
VVGNVYHERFTRNFRAFFKKNGANRTYNLAKMFNYDVNCTNAKD GKAWDVKTS
MDTVKKMMDSNDVRVTKRLL EQTGALADATYKATVAGKAKDGAYIGMKT KSSV
FADVSKYGGMTKIKNAYSIIVQYTGKKGEVIKEIVPLPIYLTNRNTTDQDLINYVASI
IPQAKDISIIYGKLCINQLVKVNGFYYYLGGKTNSKFCIDNAIQVIVSNEWIPYLKVL
EKFNNMRKDNKDLKANVVSTRALDNKHTIEVRIVEEKNIEFFDYLVSKLKMPIYQK
MKGNKAAELSEKGYGLFKKMSLEEQSIHLIELLNLLTNQKTTFEVKPLGITASRSTV
GSKISNQDEFKVINESITGLYSNEVTIVGKRPAATKKAGQAKKKKGSYPYDVPDYA
YPYDVPDYA YPYDVPDYA (SEQ ID NO: 10)

腺嘌呤脱氨酶，与切口酶 *VapCas9* 的 C 末端融合的 TadA8.13m (nVapCas9-ABE8, D38A 突变体) 的氨基酸序列

[0662]

MPKKKRKVSIIINFQRRGLMETQASNQLISSHLKGYPIKD YFVGLAIGTSSVGWAVT
NKAYELLKFRSHKMWGSRLFDEGESAVARRGFRSMRRRLERRKLRLKLEELFAD
AMAQVDPTFFMRLRESKYHYEDKTTGHSSKHILFIDKNYNDQDYFKEYPTVYHLR
SELMKSGTDDIRKLFLAVHHILKYRGNFLYEGATFDSNASTLDDVIKQALENITFNC
FDCNSAIISSIGQILMEAGKTKSDKAKAIEHLVDTYIATD TVDTSSKTQKDQVKEDK
KRLKAFANLVGLNASLIDLFGSVEELEEDLKKLQITGDTYDDKRDELAKAWSDEI
YIIDDCKSVYDAIILL SIKEPGLTISESKVKA FNKHKDDLAILKSLLKSDRSIYNTMFK
VDEKGLHNYVHYIKQGRTEETSCNREDFYKYTKKIVEGLSDSKDKEYILSQIELQIL
LPLQRIKDNGVIPYQLHLEELKAILAKCGPKFPFLNEVADGFSVAEKLIKMLEFRIPY
YVGPLNTHHNVDNNGGFAWVRKASGRVTPWNFDDKIDREKSAAAFIKNLTNKCTY
LLGEDVLPKSSLLYSEFMLLNELNNVRIDGKPLEKV VKEHLIEAVFKQDHKKMTKN
RIEQFLKDNGYISETHKHEITGLDGEIKNDLAS YRDMVRILGDGFDRSMAEEIITDIT
IFGESKKMLRETLRKKFASCLDDEAIKKLTKLRYRDWGRLSQKLLNGIEGCDKAGD
GTPETIIILMRNFSYNLMELLGDKFSFMERIQEINAKLTEGQIVNPHDIIDDLALSPAV
KRAVWQALRIVDEVAHIKKALPARIFVEVTRSNKNEKKKKDSRQKRLSDLYAAIKK
DDVLLNGLNNEIFGELKSSLAKYDDAALRSKKLYLYYTQMGRCAYTGEIHELSSLN
TDNYDIDHIYPRSLTKDDSFNVLVCKRTANAQKSDAYPISEEIQKTQKPFWTFLKQ
QGLISERKYERLTRITPLTADDLSGFIARQLVETNQSVKAATTLLRRLYPGVDVVFV
KAENVTD FRHDNNFIKVRSLNHHHHAKDAYLNI VVGNVYHERFTRNFRAFFKKN
ANRTYNLAKMFNYDVNCTNAKD GKAWDVKTSMDTVKKMMDSNDVRVTKRLL
EQTGALADATYKATVAGKAKDGAYIGMKT KSSVFADVSKYGGMTKIKNAYSIIVQY

TGKKGEVIKEIVPLPIYLNRNTTDQDLINYVASIIPQAKDISIIYGKLCINQLVKVNG
 FYYYLGKGTNSKFCIDNAIQVIVSNEWIPYLKVLEKFNNMRKDNKDLKANVVSTR
 ALDNKHTIEVRIVEEKNIEFFDYLVSKLKMPIYQKMKGNKAAELSEKGYGLFKKMS
 LEEQSIHLIELLNLLTNQKTTFEVKPLGITASRSTVGSKISNQDEFKVINESITGLYSN
 EVTIVKRPAATKKAGQAKKKKSGSETPGTSESATPESSGSEVEFSHEYWMRHALTL
AKRARDEREVPVGAVLVLNRRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLY
DATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITE
GILADECAALLCRFFRMPRRVFNAQKKAQSSTDPAAKRVKLDGSYPYDVPDYAYPYDV
PDYAYPYDVPDYA (SEQ ID NO: 11)

胞苷脱氨酶，与切口酶 *VapCas9* 的 N 末端融合的 ppAPOBEC1 (CBE-n*VapCas9*,
 D38A 突变体) 的氨基酸序列

MPAAKRVKLDTSEKGPSTGDPTLRRRIESWEFDVFYDPRELKTCCLYEIKWGMSRK
IWRSSGKNNTTNHVEVNFIIKFTSERRFHSSISCSITWFLSWSPCWECSQAIREFLSQHPGV
TLVIYVARLFWHMDQNRNQGLRDLVNSGVTIQIMRASEYYHCWRNFVNYPPGDEAHWP
QYPPLWMMLYALELHCILSLPPCLKISRRWQNHLAFFRLHLQONCHYQTIPPHILLATGLI
HPSVTWRLKSGGSSGGSSGSETPGTSESATPESSGGSSGGSPKKKRKVGSIIINFQRR
 GLMETQASNQLISSHLKGYPIDYFVGLAIGTSSVGWAVTNKAYELLKFRSHKMW
 GSRLFDEGESAVARRGFRSMRRRLERRKLRLKLEELFADAMAQVDPTFFMRLRES
 KYHYEDKTTGHSSKHILFIDKNYNDQDYFKEYPTVYHLRSELMKSGTDDIRKLFLA
 VHHILKYRGNFLYEGATFDSNASTLDDVIKQALENITFNCFCDCNSAIISSIGQILMEAG
 KTKSDKAKAIEHLVDTYIATDVTVDTSSTQKQKQVKEDEKRLKAFANLVGLNASLI
 DLFGSVEELEEDLKKLQITGDTYDDKRDELAKAWSDEIYIIDDCKSVYDAIILLISKE
 PGLTISESKVKAFNKHKDDLAILKSLKSDRSIYNTMFKVDEKGLHNYVHYIKQGR
 TEETSCNREDFYKYTKKIVEGLSDSKDKEYILSQIELQILLPLQRIKDNQGVIPYQLHL
 EELKAILAKCGPKFPFLNEVADGFSVAEKLKMLEFRIPYYVGPLNTHHNVNDNGGFA
 WAVRKASGRVTPWNFDDKIDREKSAAAFIKNLTKCTYLLGEDVLPKSSLLYSEFM
 LLNELNNVRIDGKPLEKVVEHLIEAVFKQDHKKMTKNRIEQFLKDNGYISETHKH
 EITGLDGEIKNDLASYRDMVRILGDGFDRSMAEEIITDITIFGESKKMLRETLRKKFA
 SCLDDEAIKKLTKLRYRDWGRLSQKLLNGIEGCDKAGDGTPEITHILMRNFSYNLM
 ELLGDKFSFMERIQEINAKLTEGQIVNPHDIIDDLALSPAVKRAVWQALRIVDEVAHI
 KKALPARIFVEVTRSNKNEKKKKDSRQKRLSDLYAAIKKDDVLLNGLNNEIFGELK
 SSLAKYDDAALRSKKLYLYYTQMGRCAYTGEIHELNLNTDNYDIDHIYPRSLTKDD
 SFDNLVLCKRTANAQKSDAYPISEEIQKTQKPFWTFLLKQQLISERKYERLTRITPLT
 ADDLSGFIARQLVETNQSVKAATTLLRRLYPGVDVVFVKAENVTDHRHDNNFIKVR

[0663]

SLNHHHHAKDAYLNIVVGNVYHERFTRNFRAFFKKNGANRTYNLAKMFNYDVNC
TNAKD GKAWDVKTSMDTVKKMMDSDNDVRVTKRLLEQTGALADATIYKATVAGK
AKDGAYIGMKT KSSVFADVSKYGGMTKIKNAYSIIVQYTGKKGEVIKEIVPLPIYLT
NRNTTDQDLINYVASIIPQAKDISIIYGKLCINQLVKVNGFYYYLGGKTNSKFCIDNA
IQVIVSNEWIPYLVLEKFNNMRKDNKDLKANVVSTRALDNKHTIEVRIVEEKNIE
FFDYLVSKLKMPIYQKMKGNKAAELSEKGYGLFKKMSLEEQSIHLIELLNLLTNQK
TTFEVKPLGITASRSTVGSKISNQDEFKVINESITGLYSNEVTIVG**KRPAATKKAGQ**
AKKKKGSSGGSGGSGGSTNLSDIIEKETGKQLVIOESILMLPEEVEEVIGNKPESDI
LVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSGGSGGSTNL
SDIIEKETGKQLVIOESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSD
APEYKPWALVIQDSNGENKIKMLYPYDVDPDYAYPYDVDPDYAY (SEQ ID NO: 12)

腺嘌呤脱氨酶，与切口酶 EpeCas9 的 N 末端融合的 TadA8.13m (ABE-nEpeCas9, D12A 突变体) 的氨基酸序列

[0664]

MPAAKRVKLDGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVNNRVIGEGWNR
AIGLHDPTAHAEIMALRQGGLVMQNYRLYDATLYVTTFEPCVMCAGAMIHSRIGRVVFGV
RNAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQS
STDGSSGSETPGTSESATPESSG**PKKKRK**VGTKVKDYYIGL**A**IGTSSVGWAVTDEA
YNVLKFNSKKMWGVRLFDDAKTAEERRGQRGARRRLDRKKERLSLLQDFFAEV
AKVDPNFFLRDLNSDLYMEDKDQKLKSKYTLFNDKDFKDNFHKKYPTIHHLLM
DLIEDDSKKDIRLVYLACHYLLKNRGHFIFEGQKFDTKSSFENSLNELKVHLNDEY
GLDLEFDNENLINILTDPKLNKTAKKKELKSVIGDTKFLKAVSAIMIGSSQKLVDLFE
NPEDFDDSAIKSVDFSTTSFDDKYSDYELALGDKIALVNILKEIYDSSILENLLKEAD
KSKDGNKYISNAFVKKYNKHGQDLKEFKRLVRQYHKSAYFDIFRSEKVNDNYVSY
TKSSISNNKRVKANKFTDQEAIFYKFAKKHLETIKYKINKVNGSKADLELIDGMLRD
MEFKNFMPKIKSSDNGVIPYQLKLMELNKILENQSKHHEFLNVSDEYGSVCDKIAS
IMEFRIPYYVGPLNPNSKYAWIKKQKDSEITPWNFKDVVDLDSSREEFIDSLIGRCT
YLKDEKVLPAKSLLYNEYMVLNELNNLKLNDLPITEEMKKKIFDQLFKTRKKVTL
KAVANLLKKEFNINGEILLSGTDGDFKQGLNSYNDFKAIVGDKVDSDDYRDKIEEII
KLIVLYGDDKSYLQKKIKAGYGKYFTDSEIKKMAGLNYKDWGRLSKKLLTGLEGA
NKITGERGSIIHFMREYNLNLMEIAMSASFTEETEEIQKLNVPDDRKLSEYEMVDELYLS
PSVKRMLWQSLRIVDEIKNIMGTDSKKIFIEMARGKEEVKARKESRKNQLLKFYKD
GKKAFISEIGEERYSYLLSEIEGEEENKFRWDNLYLYYTQLGRCMYSLEPIDISELSS
KNIYDQDHIYPKSKIYDDSIENRVLVKKDLNSKKGNSYPIPDEILNKNCYAYWKILY
DKGLIGQKKYTRLTRRTGFTDDELVQFISRQIVETRQATKETANLLKTICKNSEIVYS

KAENASRFRQEFDIVKCRAVNDLHHMHDAYINIIVGNVYNTKFTKDPMNFVKKQE
KARSYNLENMFKYDVKRGGYTAWIADDEKGTVKNASIKRIRKELEGTNRYFTRM
NYIESGALFNATLQRKNKGSRPLKDKGPKSSIEKYGGYTNINKACFAVLDIKSKNKI
ERKLMPVEREIIYAKQKNDKKLSDEIFSKYLKDRFGIEDYRVVYPVVKMRTLTKIDG
SYYFITGGSDKTLELRSALQLILPKKNEWAIKQIDKSSENDYLTIERIQDLTEELVYN
TFDIIVNKFKTSVFKKSFLNLFQDDKIENIDFKFSMDFKCKTLLMLVKAIRASG
VRQDLKSIDLKSDYGRLLSSKTNNIGNYQEFKIINQSITGLFENEVDLLKL**GK**RPAAT
KKAGQAKKKKGSYPYDVPDYAYPYDVPDYAYPYDVPDYA (SEQ ID NO: 16)

腺嘌呤脱氨酶，与切口酶 EpeCas9 的 C 末端融合的 TadA8.13m (nEpeCas9-ABE8, D12A 突变体) 的氨基酸序列

[0665]

MPKKKRKVTKVKDYYIGL**A**IGTSSVGWAVTDEAYNVLKFNSKKMWGVRLFDDA
KTAEERRGQRGARRRLDRKKERLSLLQDFFAEVAKVDPNFFLRDLNSDLYMEDK
DQKLKSKYTLFNDKDFKDNFHKKYPTIHLLMDLIEDDSKKDIRLVYLACHYLL
KNRGHFIFEGQKFDTKSSFENSLNELKVHLNDEYGLDLEFDNENLINILTPKLNKT
AKKKELKSVIGDTKFLKAVSAIMIGSSQKLVDLFENPEDFDDSAIKSVDFSTTSFDD
KYSDYELALGDKIALVNILKEIYDSSILENLLKEADKSKDGNKYISNAFVKKYNKH
GQDLKEFKRLVRQYHKSAYFDIFRSEK VNDNYVSYTKSSISNNKRVKANKFTDQEA
FYKFAKKHLETIKYKINKVNGSKADLELIDGMLRDMEFKNFMPKIKSSDNGVIPYQ
LKLMELNKILENQSKHHEFLNVSDEYGSVCDKIASIMEFRIPYYVGPLNPNSKYAWI
KKQKDSEITPWNFKDVVDLDSSREEFIDSLIGRCTYKDEKVLPAKSLLYNEYMVL
NELNNLKLNDLPITEEMKKKIFDQLFKTRKKVTLKAVANLLKKEFNINGEILLSGTD
GDFKQGLNSYNDFKAIVGDKVDSDDYRDKIEEIIKLIVLYGDDKSYLQKKIKAGYG
KYFTDSEIKKMAGLNYKDWGRLSKKLLTGLEGANKITGERGSIHFMREYNLNL
ELMSASFTFTEEIQKLNPVDDRKLSYEMVDELYLSPSVKRMLWQSLRIVDEIKNIM
GTDSKKIFIEMARGKEEVKARKESRKNQLLKFKYKDGGKAFISEIGEERYSYLLSEIE
GEEENKFRWDNLYLYYTQLGRCMYSLEPIDISELSSKNIYDQDHIYPSKSIYDDSIEN
RVLVKKDLNSKKGNSYPIPDEILNKNCYAYWKILYDKGLIGQKKYTRLTRRTGFTD
DELVQFISRQIVETRQATKETANLLKTICKNSEIVYSKAENASRFRQEFDIVKCRAV
DLHHMHDAYINIIVGNVYNTKFTKDPMNFVKKQE KARSYNLENMFKYDVKRGGY
TAWIADDEKGTVKNASIKRIRKELEGTNRYFTRMNYIESGALFNATLQRKNKGSRP
LKDKGPKSSIEKYGGYTNINKACFAVLDIKSKNKIERKLMPVEREIIYAKQKNDKKL
SDEIFSKYLKDRFGIEDYRVVYPVVKMRTLTKIDGSYYFITGGSDKTLELRSALQLIL
PKKNEWAIKQIDKSSENDYLTIERIQDLTEELVYNTFDIIVNKFKTSVFKKSFLNLFQ
DDKIENIDFKFSMDFKCKTLLMLVKAIRASGVRQDLKSIDLKSDYGRLLSSKTN

NIGNYQEFKIINQSITGLFENEVDLLKLKRPAATKKAGQAKKKKSGSETPGTSESA
TPESSGSEVEFSHEYWMRHALTLAKRRADEREVPVGAVLVLNNRVIGEGWNRAIGLHDP
TAHAEIMALROGGLVMQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTG
AAGSLMDVLHHPGMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTDPA
KRVKLDGSYPYDVPDYAYPYDVPDYAYPYDVPDYA (SEQ ID NO: 17)

腺嘌呤脱氨酶，与切口酶 EpeCas9 的 C 末端融合的 TadA8.13m (nEpeCas9-ABE8, D12A 突变体) 的氨基酸序列

MPAAKRVKLDTSEKGPSTGDPTLRRRIESWEFDVFYDPRELKTCCLYEIKWGMSRK
IWRSSGKNTTNHVEVNFIIKKFTSERRFHSSISCSITWFLSWSPCWECSQAIREFLSQHPGV
TLVIYVARLFWHMDQRNRQGLRDLVNSGVTIQIMRASEYYHCWRNFVNYPPGDEAHWP
QYPPLWMMLYALELHCILSLPPCLKISRRWQNHLAFFRLHLQONCHYQOTIPPHILLATGLI
HPSVTWRLKSGGSSGGSSGSETPGTSESATPESSGGSSGGSPKKKRKVGTTKVKDY
YIGLAIGTSSVGWAVTDEAYNVLKFNKKMWGVRLFDDAKTAEERRGQRGARRRL
DRKKERLSLLQDFFAEVAKVDPNFFLRDLNNDLYMEDKDQKLKSKYTLFNDKDF
KDKNFHKKYPTIHLLMDLIEDDSKKDIRLVYLACHYLLKNRGHFIFEGQKFDTKS
SFENSLNELKVHLNDEYGLDLEFDNENLINILTDPKLNKTAKKKELKSVIGDTKFLK
AVSAIMIGSSQKLVDLFENPEDFDDSAIKSVDFSTTSFDDKYSDYELALGDKIALVNI
LKEIYDSSILENLLKEADKSKDGNKYISNAFVKKYNKHGQDLKEFKRLVRQYHKSA
YFDIFRSEK VNDNYVSYTKSSISNNKRVKANKFTDQEA FYKFAKKHLETIKYKINK
VNGSKADLELIDGMLRDMEFKNFMPKIKSSDNGVIPYQLKLMELNKILENQSKHH
EFLNVSDEYGSVCDKIASIMEFRIPYVGPLNPNSKYAWIKKQKDSEITPWNFKDV
VDLDSSREEFIDSLIGRCTYLKDEKVLPAKSLLYNEYMVLNELLNLKLNLDLPITEEM
KKKIFDQLFKTRKKVTLKAVANLLKKEFNINGEILLSGTDGDFKQGLNSYNDFKAI
VGDKVDSDDYRDKIEEIIKLIVLYGDDKSYLQKKIKAGYGYFTDSEIKKMAGLNY
KDWGRLSKKLLTGLEGANKITGERGSIHFMREYNLNLMELEMSASFTFTEEIQKLN
PVDDRKLSYEMVDELYLSPSVKRMLWQSLRIVDEIKNIMGTDSKKIFIEMARGKEE
VKARKESRKNQLLK FYKDGGKAFISEIGEERYSYLLSEIEGEEENKFRWDNLYLYY
TQLGRCMYSLEPIDISELSSKNYDQDHIYPKSKIYDDSIENRVLVKKDLNSKKGNS
YPIPDEILNKNCYAYWKILYDKGLIGQKKYTRLTRRTGFTDDELVQFISRQIVETRQA
TKETANLLKTICKNSEIVYSKAENASRFRQEFDIVKCRAVNDLHHMHDAYINIIVGN
VYNTKFTKDPMNFVKKQEKARSYNLENMFKYDVKRGGYTAWIADDEKGTVKNA
SIKRIRKELEGTNYRFRMNYIESGALFNATLQRKNKGSRLPKDKGPKSSIEKYGGY
TNINKACFAVLDIKSKNKIERKLMPVEREIIYAKQKNDKKLSDEIFSKYLKDRFGIED
YRVVYPVVKMRTLKIDGSYYFITGGSDKTLELRSALQLILPKKNEWAIKQIDKSSE

[0666]

NDYLTIERIQDLTEELVYNTFDIIVNKFKTSVFKKSFLNLFQDDKIENIDFKFKSMDF
KEKCKTLLMLVKAIRASGVRQDLKSIDLKSDYGRLLSSKTNNIGNYQEFKIINQSITG
LFENEVDLLKLGKRPAATKKAGQAKKKKGSSGGSGGSGGSTNLSDIIEKETGKQLV
IOESILMLPEEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQD
SNGENKIKMLSGGSGGSGGSTNLSDIIEKETGKQLVIOESILMLPEEEVEEVIGNKPE
SDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLYPYDVPDYAYP
YDVPDYAY (SEQ ID NO: 18)

腺嘌呤脱氨酶，与切口酶 *LfeCas9* 的 N 末端融合的 TadA8.13m 切口酶
(ABE-nLfeCas9, D9A 突变体) 的氨基酸序列

MPAAKRVKLDGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNR
AIGLHDPTAHAEIMALRQGGLVMQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGV
RNAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQS
STDGSSGSETPGTSESATPESSG**PKKKRKV**GKEYHIGLAIGTSSIGWAVTDSQFKLM
RIKGKTAIGVRLFEEGKTAAERTFRTRRRLLKRRKWRLHYLDEIFAPHLQEVDENF
LRRLLKQSNHPEDPAKNQAFIGKLLFPDLLKKNERGYPTLIKMRDELPEQRAHYP
VTNIYKLREAMINEDRQFDLREVYLAVHHIVKYRGHFLNNASVDKFKVGRIDFDK
SFNVLNEAYEELQNGEGSFTIEPSKVEKIGQLLLDTKMRKLDQRQKAVAKLLEVKVA
DKEETKRNKQIATAMSKLVLGKADFATVAMANGNEWKIDLSSETSEDEIEKFREE
LSDAQNDILTEITSLFSQIMLNEIVPNGMSISESMMDRYWTHERQLAEVKEYLATQP
ASARKEFDQVYNKYIGQAPKEKGFDEKGLKKILSKKENWKEIDELLKAGDFLPK
QRTSANGVIPHQMHQQELDRIIEKQAKYYPWLATENPATGERDRHQAKYELDQLV
SFRIPYYVGPLVTPEVQKATSGAKFAWAKRKEDGEITPWNLWDKIDRAESAEAFIK
RMTVKDTYLLNEDVLPANSLLYQKYNVLNELNNVRVNGRRLSVGIKQDIYTELFK
KKKTVKAGDVASLVMMAKTRGVNKPVSVEGLSDPKKFNSNLATYLDLKSIVGDKVD
DNRYQMDLENIIEWRSVFEDGEIFADKLTEVEWLTDEQRSALVKKRYKGWGRLSK
KLLTGIVDENGQRIIDLMWNTDQNFMQIVNQPVFKEQIDQLNQKAITNDGMTLRER
VESVLDDAYTSPQNKKAIWQVVRVVEDIVKAVGNAPKSISIEFARNEGNGKGEITRSR
RTQLQKLFDQAHVLKDTSLTEELEKAPDLSDRYFYFTQGGKDMYTGDPIFNDE
ISTKYDIDHILPQSFKDDSLDNRVLVSRAENNKKSDRVPKLYAAKMKPYWNQLL
KQGLITQRKFENLTMDVDQTIKYRSLGFVKRQLVETRQVIKLTANILGSMYQEAGT
DIIETRAGLTKQLREEFDLPKVREVNNDYHHAVDAYLTTFAGQYLNRRYPKLRSFFV
YGEYMKFKHGSCLKLRNFNFHHELMEGDKSQGKVVDQQTGELITTRDEVADYFD
WVINLKVMLISNETYEETGKYFDASHESSSLYLKNQNKSKLVPLKNKLQPEYY
GAYTGITQGYMVILKLLDKKGGFGVYRIPRYAADILNKCHDEVAYRNKIAEIISSDP

[0667]

RAPKSFEVVVPRVLKGTFLVDGEEKFILSSYRYKVNATQLILPVSDIKLIQDNFKALK
KLNVMQTKKLEIYDNILRQVDKYYKLYDINKFRAKLHDGRSKFVELDDFGQDA
SKEKVIHKILRGLHFGSDLQNLKEIGFGTTPLGQFQVSEAGIRLSNTAFIIFKSPTGLF
NRKLYLKNLGKRPAAATKKAGQAKKKKGSYPYDVDPDYAYPYDVDPDYAYPYDVDPDYA
(SEQ ID NO: 88)

腺嘌呤脱氨酶，与切口酶 *LfeCas9* 的 C 末端融合的 TadA8.13m 切口酶
(n*LfeCas9*-ABE, D9A 突变体) 的氨基酸序列

MPKKKRKVGKEYHIGLAIGTSSIGWAVTDSQFKLMRIKGKTAIGVRLFEEGKTAAE
RRTFRTRRRRLKRRKWRLHYLDEIFAPHLQEVDENFLRRLKQSNHPEDPAKNQAFI
GKLLFPDLLKKNERGYPTLIKMRDELVEQRAHYPVTNIYKLREAMINEDRQFDLR
EVYLAVHHIVKYRGHFLNNASVDKFKVGRIDFDKSFNVLNEAYEELQNGEGSFTIE
PSKVEKIGQLLLDTKMRKLDQRQKAVAKLLEVKVADKEETKRNKQIATAMSKLVLG
YKADFATVAMANGNEWKIDLSSETSEDEIEKFEELSDAQNDILTEITSLFSQIMLNE
IVPNGMSISESMMDRYWOTHERQLAEVKEYLATQPASARKEFDQVYNKYIGQAPKE
KGFDEKGLKKILSKKENWKEIDELLKAGDFLPKQRTSANGVIPHQMHQQELDRII
EKQAKYYPWLATENPATGERDRHQAKYELDQLVSFRIPYYVGPLVTPEVQKATSGA
KFAWAKRKEDGEITPWNLWDKIDRAESAEAFIKRMTVKDTYLLNEDVLPANSLLY
QKYNVLNELNNVRVNGRRLSVGIKQDIYTELFKKKKTKVKAGDVASLVMMAKTRGV
NKPSVEGLSDPKKFNSNLATYLDLKSIVGDKVDDNRYQMDLENIIEWRSVFEDGEI
FADKLTEVEWLTDEQRSALVKKRYKGWGRLSKLLTGIVDENGQRIIDLMWNTDQ
NFMQIVNQPVFKEQIDQLNQKAITNDGMTLRERVESVLDDAYTSPQNKKAIWQVV
RVVEDIVKAVGNAPKSSISIEFARNEGNGGEITRSRRTQLQKLFEDQAHELVKDTSLTE
ELEKAPDLSDRYFYFTQGGKDMYTGDPINFDEISTKYDIDHILPQSFKDDSLDNR
VLVSRAENNKKSDRVPKLYAAKMMPYWNQLLKQGLITQRKFENLTMDVDQTIKY
RSLGFVKRQLVETRQVIKLTANILGSMYQEAGTDIIETRAGLTQKLFEEFDLPKVRE
VNDYHHAVDAYLTTFAGQYLNRRYPKLRSFFVYGEYMKFKHGSDDLKLRNFFHHE
LMEGDKSQGKVVDQQTGELITRDEVADYFDWVINLKVMLISNETYEETGKYFDA
SHESSSLYLKNQNKSKLVVPLKNKLQPEYYGAYTGITQGYMVILKLLDKKGGFG
VYRIPRYAADILNKCHDEVAYRNKIAEIISSDPRAPKSFEVVVPRVLKGTFLVDGEEK
FILSSYRYKVNATQLILPVSDIKLIQDNFKALKKLNVMQTKKLEIYDNILRQVDK
YYKLYDINKFRAKLHDGRSKFVELDDFGQDASKEKVIHKILRGLHFGSDLQNLKEIG
FGTTPLGQFQVSEAGIRLSNTAFIIFKSPTGLFNRKLYLKNL**KRPAAATKKAGQAKK**
KKSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVNNRV
IGEGWNRAIGLHDPTAHAEIMALROGGLVMQNYRLYDATLYVTFEPCVMCAGAMIHSRI

[0668]

GRVVFQVRNAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLCRFFRMPPRRVFN
AOKKAQSSTDPAAKRVKLDGSYPDVDPDYAYPYDVPDYAYPYDVPDYA (SEQ ID NO:
89)

胞苷脱氨酶，与切口酶 LfeCas9 的 N 末端融合的 ppAPOBEC1 (CBE-nLfeCas9,
D9A 突变体) 的氨基酸序列

MPAAKRVKLDTSEKGPSTGDPTLRRRIESWEFDVFYDPRELKTCCLLYEIKWGMSRK
IWRSSGKNTTNHVEVNFIIKFTSERRFHSSISCSITWFLSWSPCWECSQAIREFLSOHPGV
TLVIYVARLFWHMDQRNRQGLRDLVNSGVTIQIMRASEYYHCWRNFVNYPPGDEAHWP
QYPPLWMMLYALELHCILSLPPCLKISRRWQNHLAFFRLHLQNCYQITIPPHILLATGLI
HPSVTWRLKSGGSSGGSSGSETPGTSESATPESSGGSSGGSPKKKRKVGKEYHIGLA
IGTSSIGWAVTDSQFKLMRIKKGKTAIGVRLFEEGKTAAERRTFRTTTRRLKRRKWRL
HYLDEIFAPHLQEVDENFLRRLKQSNHPEDPAKNQAFIGKLLFPDLLKKNERGYPT
LIKMRDELPEQRAHYPVTNIYKLREAMINEDRQFDLREVYLAVHHIVKYRGHFL
NNASVDKFKVGRIDFDKSFNVLNEAYEELQNGEGSFTIEPSKVEKIGQLLLDTKMR
KLDRQKAVAKLLEVKVADKEETKRNKQIATAMSKLVLGKADFATVAMANGNEW
KIDLSSETSEDEIEKFREELSDAQNIDILTEITSLFSQIMLNEIVPNGMSISESMMDRYW
THERQLAEVKEYLATQPASARKEFDQVYNKYIGQAPKEKGFDLEKGLKKILSKKE
NWKEIDELLKAGDFLPKQRTSANGVIPHQMHQQELDRIIEKQAKYYPWLATENPAT
GERDRHQAKYELDQLVSFRIPYYVGPLVTPEVQKATSGAKFAWAKRKEDGEITPW
NLWDKIDRAESAFAFIKRMVTKDITYLLNEDVLPANSLYQKYNVLNELNNVRVNG
RRLSVGIKQDIYTELFKKKKTVKAGDVASLVMAKTRGVNKPVSVEGLSDPKKFNSN
LATYLDLKSIVGDKVDDNRYQMDLENIEWRSVFEDGEIFADKLTEVEWLTDEQRS
ALVKKRYKGWGRLSKKLLTGIVDENGQRIIDLMWNTDQNFMQIVNQPVFKEQIDQ
LNQKAITNDGMTLRERVESVLDDAYTSPQNKKAIWQVVRVVEDIVKAVGNAPKXSIS
IEFARNEGNGKEITRSRRTQLQKLFEDQAHELVKDTSLTEELEKAPDLSDRYFYFT
QGGKDMYTGDPIINFDEISTKYDIDHILPQS FVKDDSLDNRLVLSRAENNNKSDRVP
AKLYAAKMKPYNQQLKQGLITQRKFENLTMDVDQTIKYRSLGFVQRQLVETRQV
IKLTANILGSMYQEAGTDIIETRAGLTKQLREEFDLPKVREVNDYHHAVDAYLTTF
GQYLNRRYPKLRSFFVYGEYMKFKHGS DLKLRNFNFFHELMEGDKSQGKVVDQQ
TGELITRDEVADYFDWVINLKVMLISNETYEETGKYFDASHESSSLYLKNQNKKS
KLVVPLKNKLQPEYYGAYTGITQGYMVILKLLDKKGGFGVYRIPRYAADILNKCH
DEVAYRNKIAEIISSDPRAPKSFEVVVPRVLKGTFLVDGEEKFILSSYRYKVNATQLI
LPVSDIKLIQDNFKALKKLVNEMQTKKLEIYDNILRQVDKYYKLYDINKFRAKLH
DGRSKFVELDDFGQDASKEKVIKILRGLHFGSDLQNLKEIGFGTTPLGQFQVSEAG

[0669]

IRLSNTAFIIFKSPTGLFNRLKLYLKNLGKRPAATKKAGQAKKKKGSSGGSGGSGGS
TNLSDIIEKETGKQLVIOESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLT
SDAPEYKPWALVIQDSNGENKIKMLSGGSGGSGGSTNLSDIIEKETGKQLVIOESIL
MLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGEN
KIKMLYPYDVPDYAYPYDVPDYAY (SEQ ID NO: 90)

腺嘌呤脱氨酶，与切口酶 *PmaCas9* 的 N 末端融合的 TadA8.13m 切口酶 (ABE-*PmaCas9*, D12A 突变体) 的氨基酸序列

MPAAKRVLDGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNR
AIGLHDPTAHAEIMALRQGGLVMQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGV
RNAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQS
STDGSSGSETPGTSESATPESSGPKKKRKVGEKKTNYTIGLAIGTDSVGWAVVKDD
LELVKKRMKVLGNTETNYIKKNLWGSLLFESGQTAKDRRLKRVARRRYERRRNRL
TELQKIFAPAIDEVDENFFRLNESFLVPEDKAFSKNPIFGTLGEDKTYKYTYPTIYH
LRQHLADSEEKADVRLIYLALAHMIKYRGHFLIEGKLDTEHIAINENLEQFFESYNA
LFSEPIELRKEELIAIENILREKNSRTVKEKRITSFLKDIGRANKQSPMMAFITLIVG
KKAKFKAAFNLEEEISLNLTDSDYDENLEILLNTIGSDFADLFDHAQRVYNAVELAG
ILSGDVKNTHAKLSAQMVAMYERHKEQLKEYKSFIKANLPDQYDMTFVAPKDAQ
KKDLKGYAGYIDGNMSQDSFYKFVKDQLKEVPGSEKFLDSIEKEDFLRKQRSFYN
GVIPNQVHLAEMEAILDRQENYYPWLENREKIISLLTFRIPYYVGPLADGQSEFAW
LERKSDEKIKPWNFSDVVDLDRSAEKFIEQLIGRDTYLPDEYVLPKKS LIYQKYMV
FNELTKIAYLDERQKRMNLSSVEKKEIFETLFKKRSKVTEKQLVKFFENYLQIDNPTI
FGIEDAFNADYSTYVELAKVPGMKSMMDDPDNEDLMEEIVKILT VFEDRKMRRKQ
LEKYKERLSPEQIKELAKKH YTGWGRLSKKLLVGIRDKETQKTILDYLVEDDNHSG
GRQHLNRNLMQLINDRLSFKKTIAELQ MIDPSADLYAQVQEIAGSPA IKKGILLGL
KIVDEIIRVMGEK PENIVIAMARENQTTARGKALSKRREAKIKEGLAALGSSLLKEN
LPGNADLSQRKIYLYYTQNGKDIYLDEPLDFDRLS QYDEDHIIPQSFTVDNSLDNLV
LTNSSQNRGNKKDDVPSLEV VNRQLAYWRSLKDAGLMTQRKFDNLTKAMRGGLT
DKDRERFIQRQLVETRQITKNVAKLLDMRLNDKKDEAGNKIRETNIVLLKSAMASE
FRKMFRLYKVRELNDYHHAHDAYLNAAIAINLLALYPYMADDFVYGEFRYKKKP
QAEKATYEKL RQWNLIKRFGEKQLFTP DHEDCWNKERDIKTIKKVMGYRQVNVV
KKAERTGMLFKETINGKT NKGSRIPK KDL DPSKYGGYIEEK MAYYAVISYEDKK
KKPGKTIVGISIMDKKEFEYDSISYLGKLGFSNPVVQIILKNYSLIAYPDGRRRYITG
ATKTTKGKVELQKANQIAMEQDLVNFIYHLKNYDEISHPESYAFVQSHTDYFDRLF
DSIEHYTRRFLDAETNINRLRRIYEEKKKDPVDIEALVASFIELLKLTSAGAPADFIF

[0670]

[0671]

MGEAISRRRYNSMTGLFDGQVIYQSLTGLYETRMRFEDKRPAATKKAGQAKKK
KGSYPYDVPDYAYPYDVPDYAYPYDVPDYA (SEQ ID NO: 91)

腺嘌呤脱氨酶，与切口酶 *PmaCas9* 的 C 末端融合的 TadA8.13m 切口酶
 (n*PmaCas9*-ABE, D12A 突变体) 的氨基酸序列

MPKKKRKVEKKTNYTIGLAIGTDSVGWAVVKDDLELVKKRMKVLGNTETNYIKK
 NLWGSLLFESGQTAKDRRLKRVARRRYERRRNRLTELQKIFAPAIDEVDENFFFRLN
 ESFLVPEDKAFSKNPIFGTLGEDKTYKYTYPTIYHLRQHLADSEEKADVRLIYLALA
 HMIKYRGHFLIEGKLDTEHIAINENLEQFFESYNALFSEEPIELRKEELIAIENILREK
 NSRTVKEKRITSFLKDIGRANKQSPMMAFITLIVGKKAKFKAAFNLEEEISLNLTD
 SYDENLEILLNTIGSDFADLFDHAQRVYNAVELAGILSGDVKNTHAKLSAQMVAM
 YERHKEQLKEYKSIKANLPDQYDMTFVAPKDAQKKDLKGYAGYIDGNMSQDSF
 YKFVKDQLKEVPGSEKFLDSIEKEDFLRKQRSFYNGVIPNQVHLAEMEAILDRQEN
 YYPWLKENREKIISLLTFRIPYYVGPLADGQSEFAWLERKSDEKIKPWNFSQVVDLD
 RSAEKFIEQLIGRDTYLPDEYVLPKKSIIYQKYMVFNELTKIAYLDERQKRMNLSSV
 EKKEIFETLFKKRSKVTEKQLVKFFENYLQIDNPTIFGIEDAFNADYSTYVELAKVP
 GMKSMMDDPDNEDLMEEIVKILTVEFDRKMRRKQLEKYKERLSPEQIKELAKKHY
 TGWGRLSKKLLVGIRDKETQKTILDYLVEDDNHSGGRQHLNRNLMQLINDDRLSF
 KKTIAELQMIDPSADLYAQVQEIAGSPAIIKKGILLGLKIVDEIIRVMGEKPENIVIEMA
 RENQTTARGKALSKRREAKIKEGLAALGSSLLKENLPGNADLSQRKIYLYYTQNGK
 DIYLDEPLDFDRLSQYDEDHIIPQSFTVDNSLDNLVLTNSSQNRGNKKDDVPSLEVV
 NRQLAYWRSKLDAGLMTQRKFDNLTKAMRGGLTDKDRERFIQRQLVETRQITKNV
 AKLLDMRLNDKKDEAGNKIRETNIVLLKSAMASEFRKMFRLYKVRELNDYHHAH
 DAYLNAAIAINLLALYPYMADDFVYGEFRYKKKPQAEKATYEKLRQWNLIKRFGE
 KQLFTPDHEDCWNKERDIKTIKKVMGYRQVNVVKKAEERTGMLFKETINGKTNK
 GSRIPIKKDLDPISKYGGYIEEKMAYYAVISYEDKKKKPGKTIVGISIMDKKEFEYDSI
 SYLGKLGFSNPVVQIILKNYSIIAYPDGRRRYITGATKTTKGKVELQKANQIAMEQ
 DLVNFIIYHLKNYDEISHPESYAFVQSHTDYFDRLFDSIEHYTRRFLDAETNINRLRRI
 YEEEEKKDPVDIEALVASFIELLKLTSAGAPADFIFMGEAISRRRYNSMTGLFDGQVI
 YQSLTGLYETRMRFEDKRPAATKKAGQAKKKKSGSETPGTSESATPESSGSEVEFS
HEYWMRHALTLAKRARDEREVPVGAVLVLNRRVIGEGWNRAIGLHDPTAHAEIMALRO
GGLVMQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHH
PGMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTDPAAKRVKLDG
SYPYDVPDYAYPYDVPDYAYPYDVPDYA (SEQ ID NO: 92)

胞苷脱氨酶，与切口酶 *PmaCas9* 的 N 末端融合的 ppAPOBEC1 (CBE-n*PmaCas9*,

[0672]

D12A 突变体) 的氨基酸序列

MPAAKRVKLDTSEKGPSTGDPTLRRRIESWEFDVFDYDPRELKRETCLLYEIKWGMSRK
IWRSSGKNTTNHVEVNFIIKFTSERRFHSSISCSITWFLSWSPCWECQAIREFLSQHPGV
TLVIYVARLFWHMDQRNRQGLRDLVNSGVITQIMRASEYYHCWRNFVNYPPGDEAHWP
OYPPLWMMLYALELHCILSLPPCLKISRRWQNHLAFFRLHLQNCYQOTIPPHILLATGLI
HPSVTWRLKSGGSSGGSSGSETPGTSESATPESSGGSSGGSPKKKRKVGEKKTNYTI
 GLAIGTDSVGWAVVKDDLELVKKRMKVLGNTETNYIKKNLWGSLLFESGQTAKDR
 RLKRVARRRYERRRNRLTELQKIFAPAIDEVDENFFFRLNESFLVPEDKAFSKNPIFG
 TLGEDKTYKYTYPTIYHLRQHLADSEEKADVRLIYLALAHMIKYRGHFLIEGKLD
 EHIAINENLEQFFESYNALFSEEPIELRKEELIAIENILREKNSRTVKEKRITSFLKD
 IGRANKQSPMMAFITLIVGKKAKFKAAFNLEEEISLNLTDSDSYDENLEILLNTIGSDFA
 DLFDHAQRVYNAVELAGILSGDVKNTHAKLSAQMVAMYERHKEQLKEYKSFKA
 NLPDQYDMTFVAPKDAQKKDLKGYAGYIDGNMSQDSFYKFVKDQLKEVPGSEKF
 LDSIEKEDFLRKQRSFYNGVIPNQVHLAEMEAILDRQENYYPWLKENREKIISLLTF
 RIPYYVGPLADGQSEFAWLERKSDEKIKPWNFSDVVDLDRSAEKFIEQLIGRDTYLP
 DEYVLPKKSIIYQKYMVFNELTKIAYLDERQKRMNLSSVEKKEIFETLFKKRSKVT
 EKQLVKFFENYLQIDNPTIFGIEDAFNADYSTYVELAKVPGMKSMMDDPDNEDLM
 EEIVKILTVFEDRKMRRKQLEKYKERLSPEQIKELAKKHGTGWGRLSKKLLVGIRD
 KETQKTILDYLVEDDNHSGGRQHLNRNLMQLINDRLSFKKTIAELQ MIDPSADLY
 AQVQEIAGSPAIIKKGILLGLKIVDEIIRVMGEKPENIVIAMARENQTTARGKALSKRR
 EAKIKEGLAALGSSLLKENLPGNADLSQRKIYLYYTQNGKDIYLDEPLDFDRLSQY
 DEDHIIPQSFTVDNSLDNLVLTNSSQNRGNKKDDVPSLEVVRQLAYWRSKLDAGL
 MTQRKFDNLTKAMRGGLTDKDRERFIQRQLVETRQITKNVAKLLDMRLNDKKDEA
 GNKIRETNIVLLKSAMASEFRKMFRLYKVRELNDYHHAHDAYLNAAIAINLLALYP
 YMADDFVYGEFRYKKKPKAEKATYEKLRQWNLIKRFGEKQLFTP DHEDCWNER
 DIKTIKKVMGYRQVNVVKKAEERTGMLFKETINGKTNKGSRIPIKKDLDP SKYGG
 YIEEKMAYYAVISYEDKKKKPGKTIVGISIMDKKEFEYDSISYLGKLGFSNPVVQIIL
 KNYSLIAYPDGRRRYITGATKTTKGKVELQKANQIAMEQDLVNFIYHLKNYDEISH
 PESYAFVQSHTDYFDRLFDSIEHYTRRFLDAETNINRLRRIYEEEEKKDPVDIEALVA
 SFIELLKLTSAGAPADFIFMGEAISRRRYNSMTGLFDGQVIYQSLTGLYETRMRFED
GKRPAATKKAGQAKKKKGSSGGSGSGGSTNLSDIIEKETGKQLV IQESILMLPE
EVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKM
LSGGSGSGSGGSTNLSDIIEKETGKQLV IQESILMLPEEVEEVIGNKPESDILVHTAYD
ESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLYPYDVDPDYAYPYDVDPDYAY

(SEQ ID NO: 93)

胞苷脱氨酶，与切口酶 PmaCas9 的 C 末端融合的 ppAPOBEC1 (nPmaCas9-CBE, D12A 突变体) 的氨基酸序列

MPAAKRVKLDTNLSDIIEKETGKQLVIOESILMLPEEVEEVIGNKPESDILVHTAYD
ESTDENVMLLTSDAPEYKPWALVIODSNGENKIKMLSGSGSGSGGSTNLSDIIEKE
TGKQLVIOESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPW
ALVIODSNGENKIKMLSGSGSGSGSGSPKKKRKVEKKTNYTIGLAIGTDSVGWAVV
KDDLELVKKRMKVLGNTETNYIKKNLWGSLLFESGQTAKDRRLKRVARRRYERRR
NRLTELQKIFAPAIDEVDENFFRLNESFLVPEDKAFSKNPIFGTLGEDKTYKYKTYPTI
YHLRQHLADSEEKADVRLIYLALAHMIKYRGHFLIEGKLDTEHIAINENLEQFFESY
NALFSEPIELRKEELIAIENILREKNSRTVKEKRITSFLKDIGRANKQSPMMAFITLI
VGKKAKFKAAFNLEEEISLNLTDSDSYDENLEILLNTIGSDFADLFDHAQRVYNAVEL
AGILSGDVKNTHAKLSAQMVAMYERHKEQLKEYKSFIKANLPDQYDMTFVAPKD
AQKKDLKGYAGYIDGNMSQDSFYKFVKDQLKEVPGSEKFLDSIEKEDFLRKQRSF
YNGVIPNQVHLAEMEAILDRQENYYPWLKENREKIISLLTFRIPYYVGPLADGQSEF
AWLERKSDEKIKPWNFSDVVDLDRSAEFIEQLIGRDTYLPDEYVLPKKSIIYQKY
MVFNELTKIAYLDERQKRMNLSSVEKKEIFETLFKKRSKVTEKQLVKFFENYLQID
NPTIFGIEDAFNADYSTYVELAKVPGMKSMDDPDNEDLMEEIVKILTVFEDRKM
RRKQLEKYKERLSPEQIKELAKKHGTGWGRLSKKLLVGIRDKETQKTILDYLVEDD
NHSGGRQHLNRNLMQLINDDRLSFKKTIAELQMIDPSADLYAQVQEIAGSPAIIKKGI
LLGLKIVDEIIRVMGEKPENIVIMARENQTTARGKALSKRREAKIKEGLAALGSSL
LKENLPGNADLSQRKIYLYYTQNGKDIYLDEPLDFDRLSQYDEDHIIPQSFTVDNSL
DNLVLTNSSQNRGNKKDDVPSLEVVRQLAYWRSKLDAGLMTQRKFDNLTKAMR
GGLTDKDRERFIQRQLVETRQITKNVAKLLDMRLNDKKDEAGNKIRETNIVLLKSA
MASEFRKMFRLYKVRELNDYHHAHDAYLNAAIAINLLALYPYMADDFVYGEFRY
KKKPQAEKATYEKLRQWNLIKRFGEKQLFTPDHEDCWNKERDIKTIKKVMGYRQ
VNVVKKAEERTGMLFKETINGKTNKGSRIPIKKDLDPSTYGGYIEEKMAYYAVISY
EDKKKKPGKTIVGISIMDKKEFEYDSISYLGKLGFSNPVVQIILKNYSIIAYPDGRRR
YITGATKTTKGKVELQKANQIAMEQDLVNFYHLKNYDEISHPESYAFVQSHTDYF
DRLFDSIEHYTRRFLDAETNINRLRIYEEKKKDPVDIEALVASFIELLKLTSAAGP
ADFIFMGEAISRRRYNSMTGLFDGQVIYQSLTGLYETRMRFED**KRPAATKKAGQA**
KKKKGSSSGSGSGSGSGSETPGTSESATPESSGSGSGSGSTSEKGPSTGDPTLRRRIESW
EFDVFYDPRELKRETCLLYEIKWGMSRKIWRSSGKNTTNHVEVNFIIKFTSERRFHSSIS
CSITWFLSWSPCWECSQAIREFLSQHPGVTLVIYVARLFWHMDORNROGLRDLVNSGVT

[0673]

IQIMRASEYYHCWRNFVNYPGDEAHWPQYPPLWMMLYALELHCHLSLPPCLKISRWW
QNHLAFFRLHLQONCHYQOTIPPHILLATGLIHPSVTWRYPYDVPDYAYPYDVPDYAYPYD
PDYA (SEQ ID NO: 94)

[0674]

- [0675] 接头(带下划线,无斜体或加粗)
- [0676] TadA8 (ABE) 或ppABOBEC1 (CBE) (斜体和带下划线)
- [0677] 切口酶突变:ScoCas9中的D10A突变、SirCas9中的D14A突变、VapCas9中的D38A、EpeCas9中的D12A、LfeCas9中的D9A、PmaCas9中的D12A(粗体和斜体)
- [0678] 2xUGI(粗体、斜体和带下划线)
- [0679] 3xHA标签(斜体),可以用不同的标签取代
- [0680] 实例6.工程化星座链球菌ScoCas9)NGC PAM变体
- [0681] 此实例说明了识别NGC PAM变体的ScoCas9变体的工程化。
- [0682] 简而言之,工程化得到两个变体:ScoCas9-NGC-v1,其含有用于NGC PAM识别的氨基酸取代;和ScoCas9-NGC-v2,其含有用于NGC PAM识别的氨基酸取代和增强SpyCas9活性的额外氨基酸取代。通过酿脓链球菌SpyCas9和星座链球菌ScoCas9之间的结构比较来鉴定氨基酸残基。ScoCas9-NGC-v1的氨基酸序列(SEQ ID NO:95)包含相较于野生型ScoCas9序列的以下突变:D1117M、S118Q、E1201F、A1299R、D1309A、R1312E、T1314R。ScoCas9-NGC-v2的氨基酸序列(SEQ ID NO:96)包含相较于野生型ScoCas9序列的以下突变:S409I、R655L、D1117M、S118Q、E1201F、A1299R、D1309A、R1312E、T1314R。

星座链球菌(ScoCas9)变体(ScoCas9-NGC-v1)的氨基酸序列

[0683]

MPKKRRKVGMGKPYSIGLDIGTNSVGWAVVTDDYKVPKMKMKVLGNTDKQSIK
KNLLGALLFDSGETAEATRLKRTARRRYTRRKNRLRYLQEIFTGEMNKVDENFFQ
RLDDSFVLVDEDKRGEHHPIFGNIAAEVKYHDDFPTIYHLRRHLADTSKKADLRLV
YLALAHMIKFRGHFLYEGDLKAENTDVQALFKDFVEEYDKTIEESHLSEITVDALS
ILTEKVSKSSRLENLIAHYPTEKKNTLFGNLIASLDLHPNFKTNFQLSEDAKLQFS
KDTYEEDLEGFLGEVGDEYADLFASAKNLYDAILLSGILTVDDNSTKAPLSASMVK
RYEEHQKDLKKLKDFIKVNAPDQYNAIFKDKNKKGYASYIESGVKQDEFYKYLK
GILLKINGSGDFLDKIDREDFLRKQRTFDNGSIPHQIHLQEMHAILRRQGEHYPFLK
ENQDKIEKILTRIPYYVGPLARKGSRFAWAEYKADEKITPWNFDDILDKEKSAEK
FITRMTLNDLYLPEEKVLPKHSPLYEAFTVYNELTKVKYVNEQGEAKFFDTNMKQ
EIFDHVFKENRKVTKDKLLNYLNKEFEEFRIVNLTGLDKENKAFNSSLGTYHDLR
KILDKSFLDDKAN EKTIEDIIQTLTLFEDREMIRQRLQKYSDIFTKAQLKKLERRHY

TGWGRLSYKLINGIRNKENKKTILDYLIDDGYANRNFMLINDDALSFKEEIARAQ
IIDDVDDIANVVDHDLPGSPAIIKKGILQSVKIVDELVKVMGHNPANIIEMARENQTT
DKGRRNSQQRLKLLQDSLKNLDNPVNIKNVENQQQLQNDRLFLYYIQNGKDMYTG
ETLDINNLSQYDIDHIIPQAFIKDNSLDNRVLTRSDKNRGKSDDVPSIEVVHEMKSF
WSKLLSVKLITQRKFDNLTKAERGGLTEEDKAGFIKRQLVETRQITKHVAQILDER
FNTEFDGNKRRIRNVKIITLKSNIIVSNFRKEFELYKVBREINDYHHAHDAYLNAVVG
NALLLKYPQLEPEFVYGEYPKYNSYRSRKSATEKFLFYSNILRFFKKEDIQTNEDG
EIAWNKEKHILRKVLSYPQVNIVKKTEEQTGGFSKESILPKGESDKLIPRKTNS
YWDPKKYGGFMQPVVAYSILVFADVEKGKSKLRKVQDMVGITIMEKKRFEKNP
VDFLEQRGYRNVRLKIIKLPKYSLELENKRRRLLASAKFLQKGNELVIPQRFTTL
LYHSYRIEKDYEPHREYVEKHKDEFKELLEYISVFSRKYVLADNNLTKEIMLFSK
NKDAEVSSLAKSFISLLTFTAFGAPRAFNFFGENIARKEYRSVTECLNATLIHQSITE
LYETRIDLKSLGEDG**EGADKRTADGSEFESPKKKRKV** (SEQ ID NO: 95)

星座链球菌(ScoCas9)变体(ScoCas9-NGC-v2)的氨基酸序列

[0684]

MPKKKKRVGMGKPYSIGLDIGTNSVGWAVVTDDYKVPKKMKVLGNTDKQSIK
KNLLGALLFDSGETAEATRLKRTARRRYTRRKNRLRYLQEIFTGEMNKVDENFFQ
RLDDSFVLVDEDKRGEHHPIFGNIAAEVKYHDDFPTIYHLRRHLADTSKKADLRLV
YLALAHMIKFRGHFLYEGDLKAENTDVQALFKDFVEEYDKTIEESHLSEITVDALS
ILTEKVSKSSRLNIAHYPTEKKNTLFGNLIASLDLHPNFKTNFQLSEDAKLQFS
KDTYEEDLEGFLGEVGDEYADLFASAKNLYDAILLSGILTVDDNSTKAPLSASMVK
RYEEHQKDLKKLKDFIKVNAPDQYNAIFKDKNKKGYASYIESGVKQDEFYKYLK
GILLKINGSGDFLDKIDREDFLRKQRTFDNGIIPHQIHLQEMHAILRRQGEHYPLK
ENQDKIEKILTRIPYYVGPLARKGSRFAWAEYKADEKITPWNFDDILDKEKSAEK
FITRMTLNDLYLPEEKVLPKHSPLYEAFTVYNELTKVKYVNEQGEAKFFDTNMKQ
EIFDHVFKENRKVTKDKLLNYLNKEFEFRIVNLTGLDKENKAFNSSLGTYHDLR
KILDKSFLDDKANIEKTIEDIIQTLTLFEDREMIRQLKQYSDIFTKAQLKKLERLHY
TGWGRLSYKLINGIRNKENKKTILDYLIDDGYANRNFMLINDDALSFKEEIARAQ
IIDDVDDIANVVDHDLPGSPAIIKKGILQSVKIVDELVKVMGHNPANIIEMARENQTT
DKGRRNSQQRLKLLQDSLKNLDNPVNIKNVENQQQLQNDRLFLYYIQNGKDMYTG
ETLDINNLSQYDIDHIIPQAFIKDNSLDNRVLTRSDKNRGKSDDVPSIEVVHEMKSF
WSKLLSVKLITQRKFDNLTKAERGGLTEEDKAGFIKRQLVETRQITKHVAQILDER
FNTEFDGNKRRIRNVKIITLKSNIIVSNFRKEFELYKVBREINDYHHAHDAYLNAVVG
NALLLKYPQLEPEFVYGEYPKYNSYRSRKSATEKFLFYSNILRFFKKEDIQTNEDG
EIAWNKEKHILRKVLSYPQVNIVKKTEEQTGGFSKESILPKGESDKLIPRKTNS

[0685]

YWDPKKYGGFMQPVVAYSILVFADVEKGKSKKLRKVQDMVGITIMEKKRFEKNP
 VDFLEQRGYRNVRLEKIIKLPKYSLFELENKRRRLLASAKFLQKGNELVIPQRFTTL
 LYHSYRIEKDYEPHEHREYVEKHKDEFKELLEYISVFSRKYVLADNNLTKEI MLFSK
 NKDAEVSSLAKSFISLLTFTAFGAPRAFNF FGENIARKEYRSVTECLNATLIHQ SITG
 LYETRIDL SKLGEDG**GADKRTADGSEFESPKKKRKV** (SEQ ID NO: 96)

[0686] NLS(粗体,斜体)

[0687] 接头(粗体,带下划线)

[0688] 使用ScoCas9-NGC变体靶向通过慢病毒介导的插入随机整合到HEK293T细胞的基因组中的基因组基因座并测试核酸酶和碱基编辑活性。

[0689] 简而言之,将HEK293T细胞铺在96孔板中。铺板后24小时,用含有ScoCas9-NGC变体和向导RNA序列ATCGACAAGAAAGGGACTGA (SEQ ID NO:97) 的表达载体转染细胞。ScoCas9变体识别例示性NGC 3'PAM序列AGC。转染后72小时收获细胞并提取总DNA。

[0690] 进行深度测序来表征HEK293T细胞中的插入缺失模式。使用两轮PCR扩增例示性靶标,以将Illumina衔接子以及独特的条形码添加到靶标扩增子。将PCR产物在2%凝胶上运行并提取凝胶。汇集样品,进行定量,并制备cDNA文库并在MiSeq上进行测序。转染后4天通过深度测序确定插入缺失频率。

[0691] 结果显示了两种ScoCas9-NGC变体的核酸酶活性。用ScoCas9-NGC-v1和ScoCas9-NGC-v2实现了约20%-35%之间的插入缺失频率(图11A)。

[0692] 由ScoCas9-NGC变体与ABE碱基编辑器构建融合物。

[0693]

与腺嘌呤碱基编辑器融合的 ScoCas9 变体(ABE-nScoCas9-NGC-v1)的氨基酸序列

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNRRVIGEGWNRAIGLHDPTAHAE
 IMALRQGGGLVMQNYRLYDATLYSTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLM
 DVLHHPGMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTD**SGGSSGGSS**
GSETPGTSESATPESSGGSSGSGKPYSIGLAIGTNSVGWAVVTDDYKVPKKMK
 VLGNTDKQSIKKNLLGALLFDSGETAEATRLKRTARRRYTRRKNRLRYLQEIFTGE
 MNKVDENFFQRLDDSFLVDEDKRGEHHPFGNIAAEVKYHDDFPTIYHLRRHLADT
 SKKADLRLVYLALAHMIKFRGHFLYEGDLKAENTDVQALFKDFVEEYDKTIEESHL
 SEITVDALSILTEKVSKSSRLNIAHYPTKNTLFGNLIALLSLDLHPNFKTNFQLSE
 DAKLQFSKDTYEEDLEGFLGEVGDEYADLFASAKNLYDAILLSGILTVDDNSTKAPL
 SASMVKRYEEHQKDLKKLKDFIKVNAPDQYNAIFKDKNKKGYASYIESGVKQDEF
 YKYLKGILLKINGSGDFLDKIDREDFLRKQRTFDNGSIPHQIHLQEMHAILRRQGEH

[0694]

YPFLKENQDKIEKILTRIPYYVGPLARKGSRFAWAHEYKADEKITPWNFDDILDKEK
 SAEKFITRMTLNDLYLPEEKVLPKHSPLYEAFTVYNELTKVKYVNEQGEAKFFDTN
 MKQEIFDHVFKENRKVTKDKLLNYLNKEFEFRIVNLTGLDKENKAFNSSLGTYHD
 LRKILDKSFLDDKANIEKTIEDIIQTTLTFEDREMIRQLQKYSIDFTKAQLKKLERRH
 YTGWGRLSYKLINGIRNKENKKTILDYLIDDGYANRNFMLINDDALSFKKEIARA
 QIIDDVDDIANVVHDLPGSPAIIKKGILQSVKIVDELVKVMGHNPNANIIEMARENQTT
 DKGRRNSQQRLKLLQDSLKNLDNPVNIKNVENQQQLQNDRLFLYYIQNGKDMYTGE
 TLDINNLSQYDIDHIIPQAFIKDNLSDNRVLTRSDKNRGKSDDVPSIEVVHEMKSFW
 KLLSVKLITQRKFDNLTKAERGGLTEEDKAGFIKRQLVETRQITKHVAQILDERFNTE
 FDGNKRRIRNVKIITLKSNNVSNFRKEFELYKVREINDYHHAHDAYLNAVVGNAALL
 KYPQLEPEFVYGEYPKYNSYRSRKSATEKFLFYSNILRFFKKEDIQTNEDEIAWNK
 EKHILRKVLSTYPQVNIIVKKTEEQTGGFSKESILPKGESDKLIPRKTNSYWDPKK
 YGGFMQPVVAYSILVFADVEKGKSKKLKRVQDMVGITIMEKKRFEKNPVDFLEQR
 GYRNVRLKIIKLPKYSLFELENKRRRLASAKFLQKGNELVIPQRFTTLLYHSYRIE
 KDYEPEHREYVEKHKDEFKELLEYISVFSRKYVLADNNLTKEIEMLFSSKNKDAEVSS
 LAKSFISLLTFTAFGAPRAFNFGENIARKEYRSVTECLNATLIHQSIITGLYETRIDLSK
 LGEDGE**GADKRTADGSEFESPKKKRKV** (SEQ ID NO: 98)

与腺嘌呤碱基编辑器融合的 ScoCas9 变体(ABE-nScoCas9-NGC-v2)的氨基酸序列

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAFLVLNNRVIGEGWNRAIGLHDPTAHAE
IMALRQGGLVMQNYRLYDATLYSTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLM
*DVLHHPGMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTD***SGGSSGGSS**
GSETPGTSESATPESSGGSSGGSSGKPYSIGLAIGTNSVGWAVVTDDYKVPKKMK
 VLGNTDKQSIKKNLLGALLFDSGETAEATRLKRTARRRYTRRKNRLRYLQEIFTGE
 MNKVDENFFQRLDDSFVDEDKRGEHHPIFGNIAAEVKYHDDFPTIYHLRRHLADT
 SKKADLRVLYLALAHMIKFRGHFLYEGDLKAENTDVQALFKDFVEEYDKTIEESHL
 SEITVDALSILTEKVSRSRLENLIAHYPTEKKNTLFGNLIALSLDLHPNFKTNFQLSE
 DAKLQFSKDTYEEDLEGFLGEVGDEYADLFASAKNLYDAILLSGILTVDDNSTKAPL
 SASMVKRYEEHQKDLKKLKDFIKVNAPDQYNAIFKDKNKKGYASYIESGVKQDEF
 YKYLKGILLKINGSGDFLDKIDREDFLRKQRTFDNGIIPHQIHLQEMHAILRRQGEH
 YPFLKENQDKIEKILTRIPYYVGPLARKGSRFAWAHEYKADEKITPWNFDDILDKEK
 SAEKFITRMTLNDLYLPEEKVLPKHSPLYEAFTVYNELTKVKYVNEQGEAKFFDTN
 MKQEIFDHVFKENRKVTKDKLLNYLNKEFEFRIVNLTGLDKENKAFNSSLGTYHD
 LRKILDKSFLDDKANIEKTIEDIIQTTLTFEDREMIRQLQKYSIDFTKAQLKKLERLH
 YTGWGRLSYKLINGIRNKENKKTILDYLIDDGYANRNFMLINDDALSFKKEIARA

[0695]

QIIDDVDDIANVVHDLPGSPAIIKKGILQSVKIVDELVKVMGHNPNANIIEMARENQTT
DKGRRNSQQRLKLLQDSLKNLDNPVNIKNVENQQQLQNDRLFLYYIQNGKDMYTGE
TLDINNLSQYDIDHIIPQAFIKDNSLDNRVLTRSDKNRGKSDDVPSIEVVHEMKSFW
KLLSVKLITQRKFDNLTKAERGGTEEDKAGFIKRQLVETRQITKHVAQILDERFNTE
FDGNKRRIRNVKIITLKSNLVSNFRKEFELYKVREINDYHHAHDAYLNAVVGNALLL
KYPQLEPEFVYGEYPKYNSYRSRKSATEKFLFYSNILRFFKKEDIQTNEEDGEIAWNK
EKHIKILRKVLSYPQVNIVKKTEEQTGGFSKESILPKGESDKLIPRKTKN SYWDPKK
YGGFMQPVVAYSILVFADVEKKGSKKLRKVQDMVGITIMEKKRFEKNPVDFLEQR
GYRNVRLKIIKLPKYSLFELENKRRRLLASAKFLQKGNELVIPQRFTTLLYHSYRIE
KDYPEHREYVEKHKDEFKELLEYISVFSRKYVLADNNLTKIEMLF SKNKDAEVSS
LAKSFISLLTFTAFGAPRAFNFFGENIARKEYRSVTECLNATLIHQ SITGLYETRIDLK
LGEDGE***GADKRTADGSEFESPKKKRKV*** (SEQ ID NO: 99)

[0696] 接头(粗体,带下划线)

[0697] TadA8 (ABE) (斜体)

[0698] NLS(粗体,斜体)

[0699] 还进行深度测序来表征HEK293T细胞中的A至G的转化(图11B)。转染后4天通过NGS测量腺嘌呤至鸟嘌呤(A至G)的转化。结果显示了利用两种ABE-nScoCas9-NGC变体的碱基编辑活性。两种变体都显示了约20%-30%之间的A至G的转化。使用识别NGG的ScoCas9作为阴性对照,并且没有显示出碱基编辑。SpyCas9用作阳性对照并且显示出约40%的A至G的转化。

[0700] 总体而言,结果显示,经工程化以识别NGC PAM序列的ScoCas9变体可以执行核酸酶以及碱基编辑活动。

[0701] 等同物和范围

[0702] 本领域中的技术人员仅使用常规实验将认识到或能够确定本文所描述的本发明具体实施例的许多等同物。本发明的范围不旨在限于上述描述,而是如随附权利要求书所述。

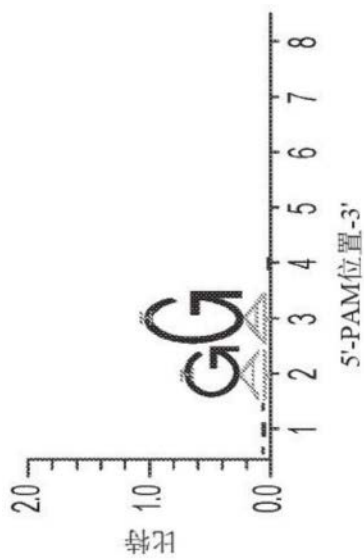


图1A

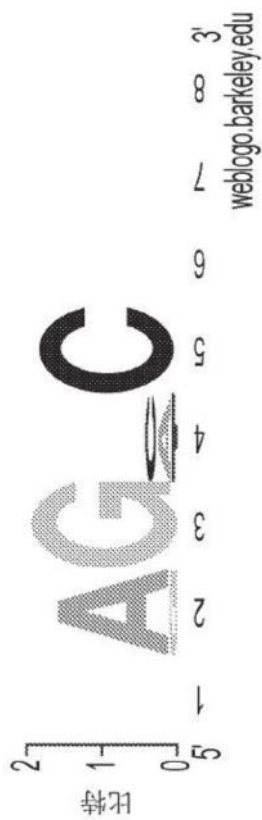


图1B

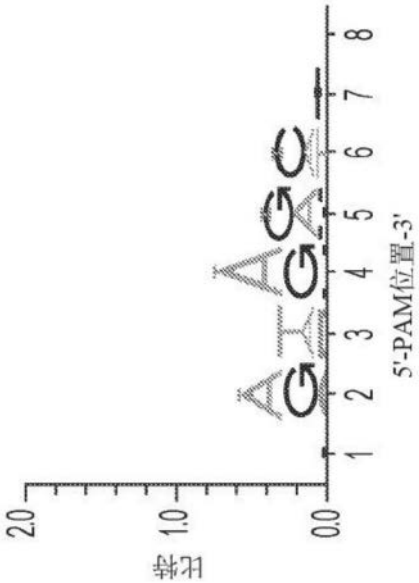


图1C

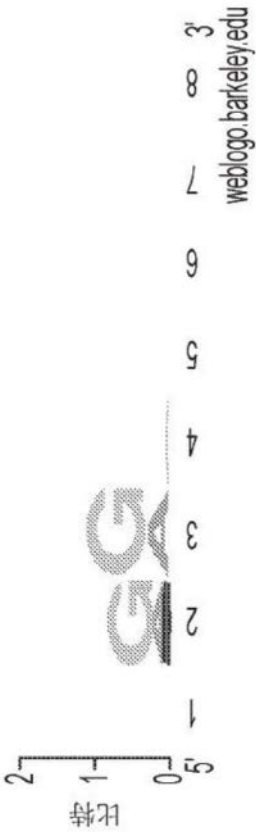


图1D

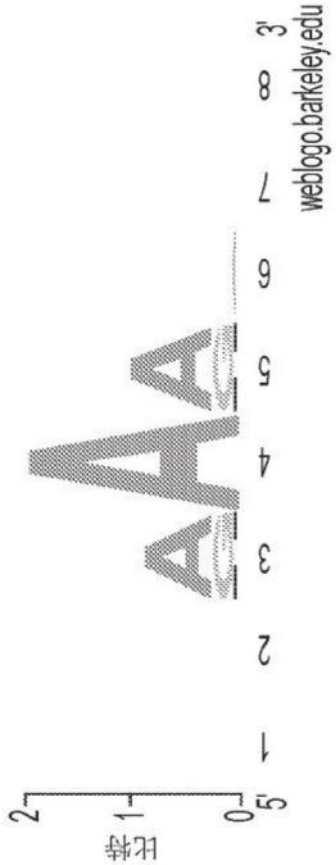


图1E

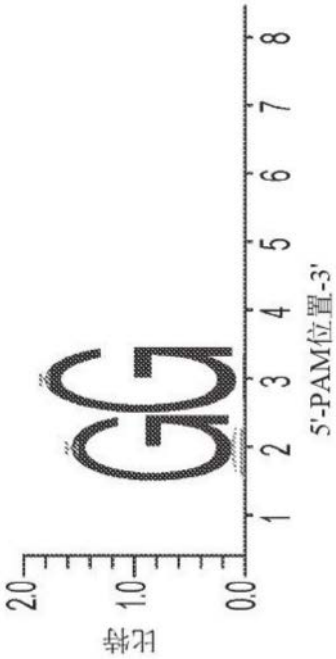


图1F

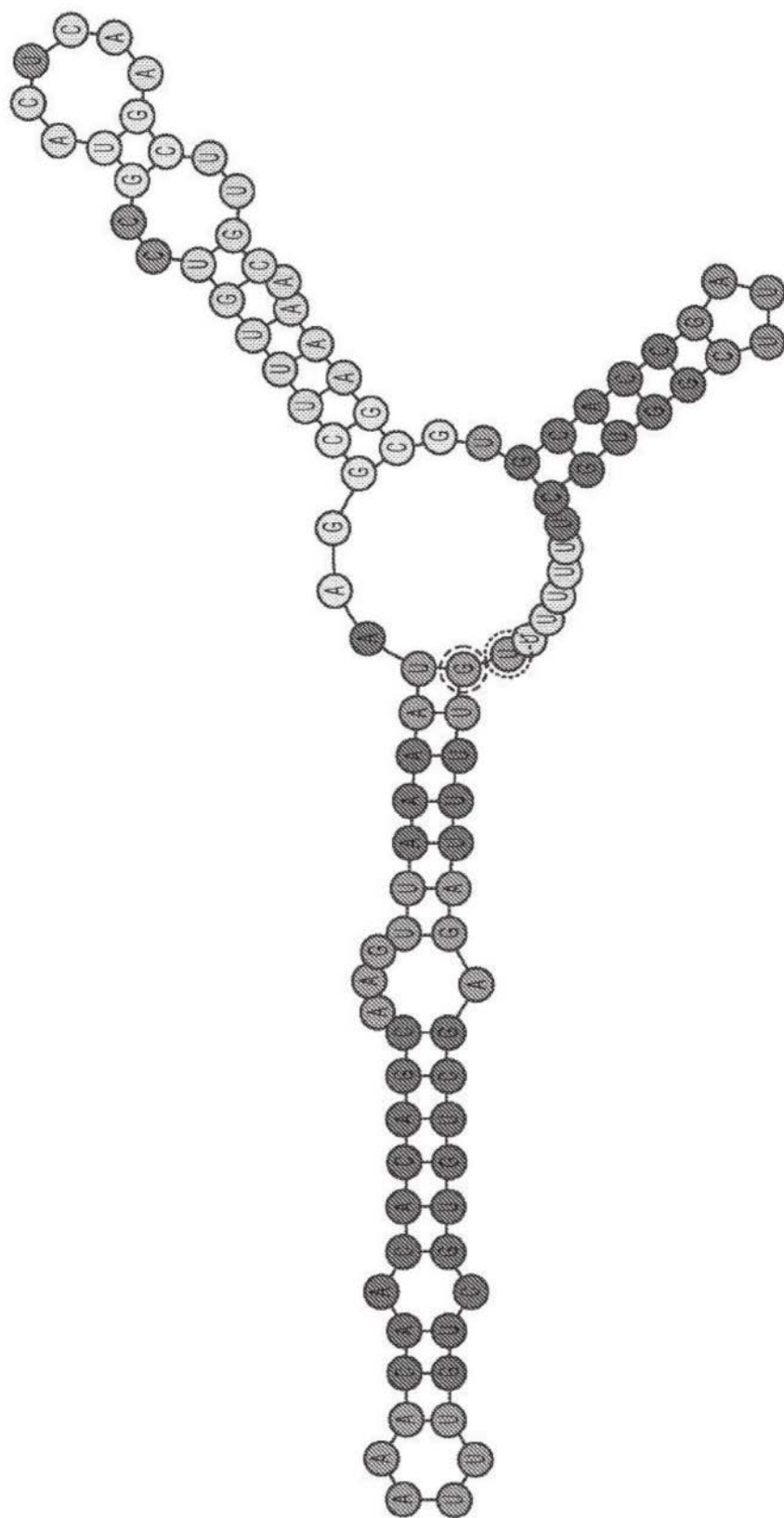


图2A

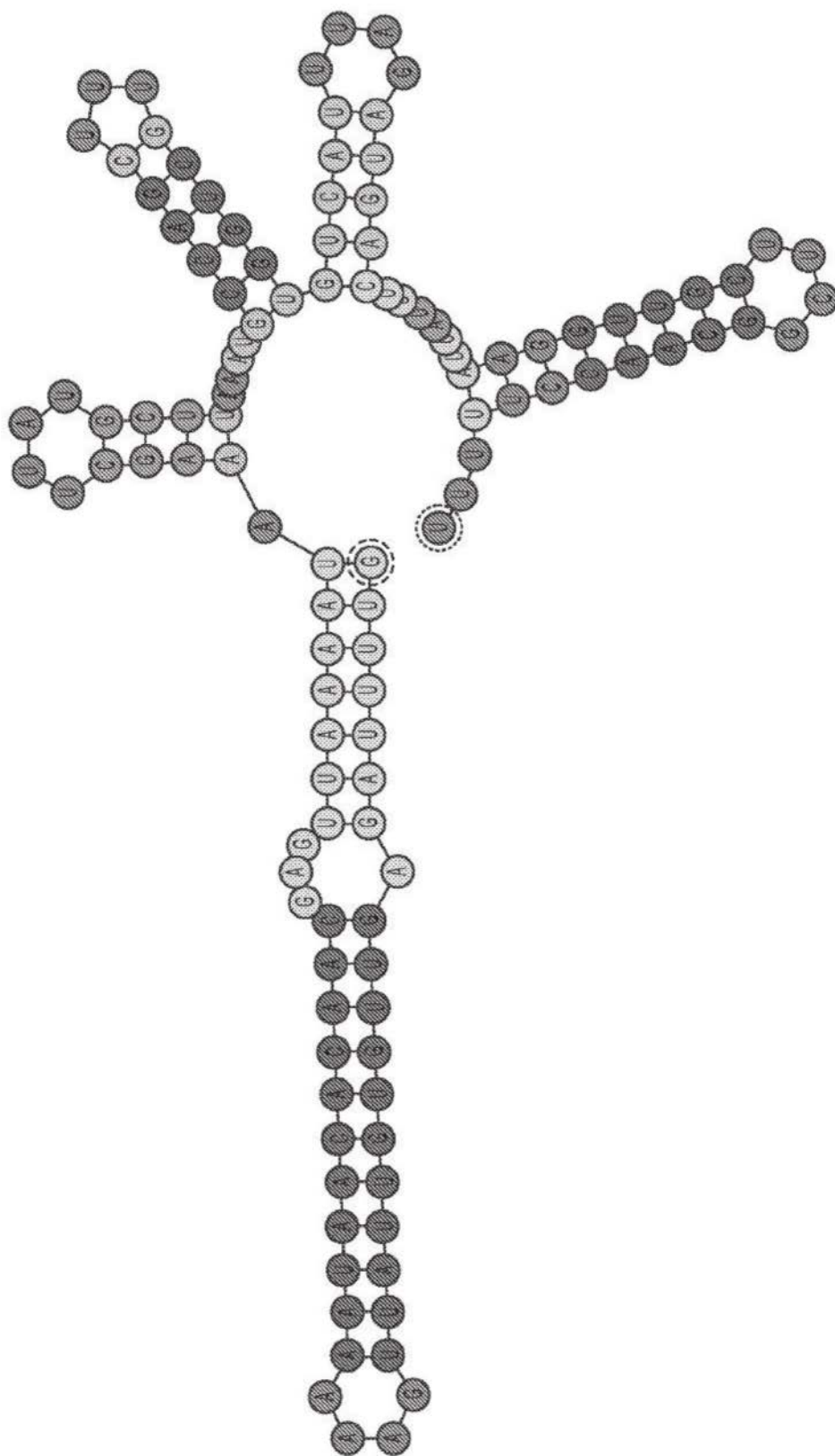


图2B

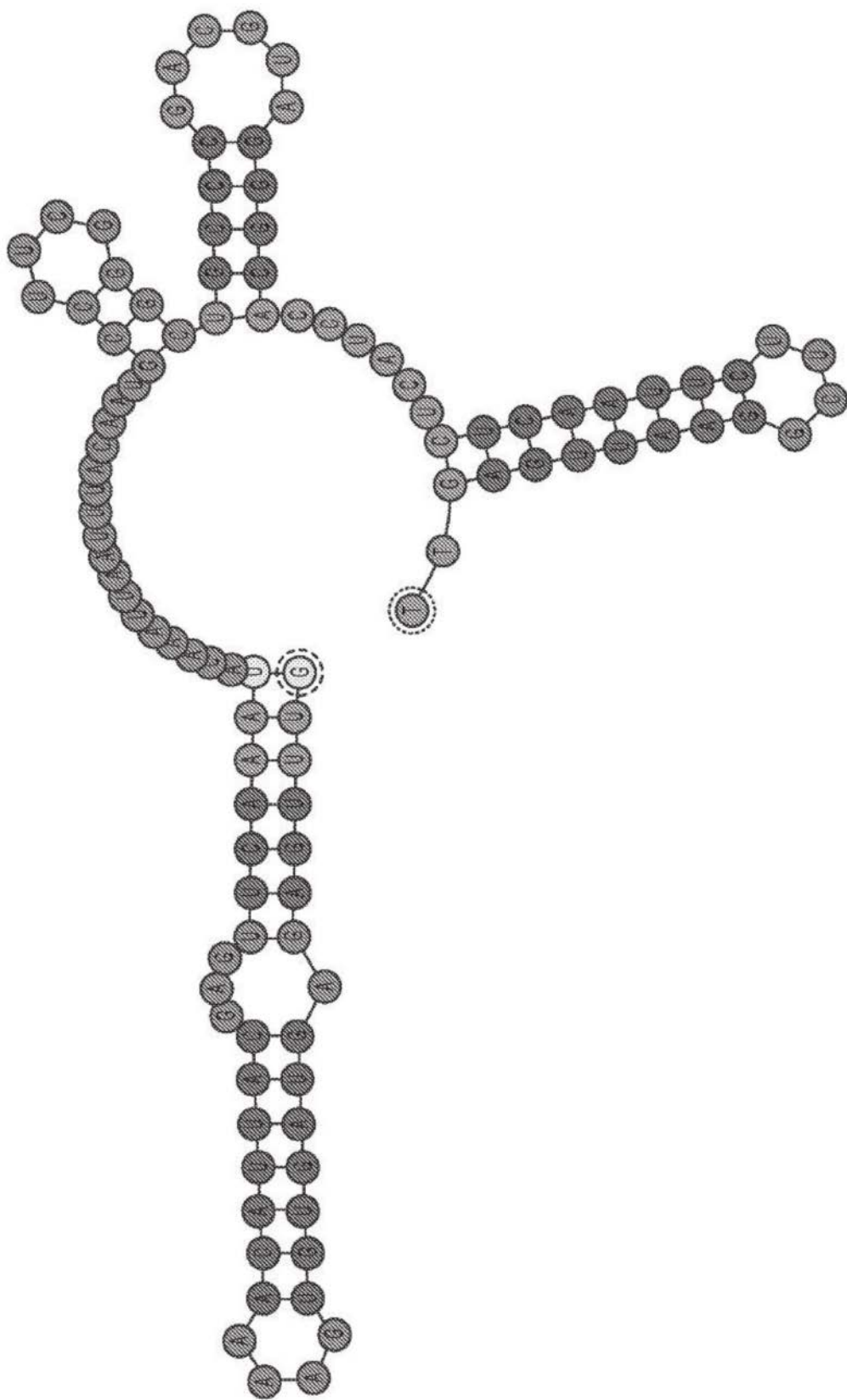


图2C



图2D

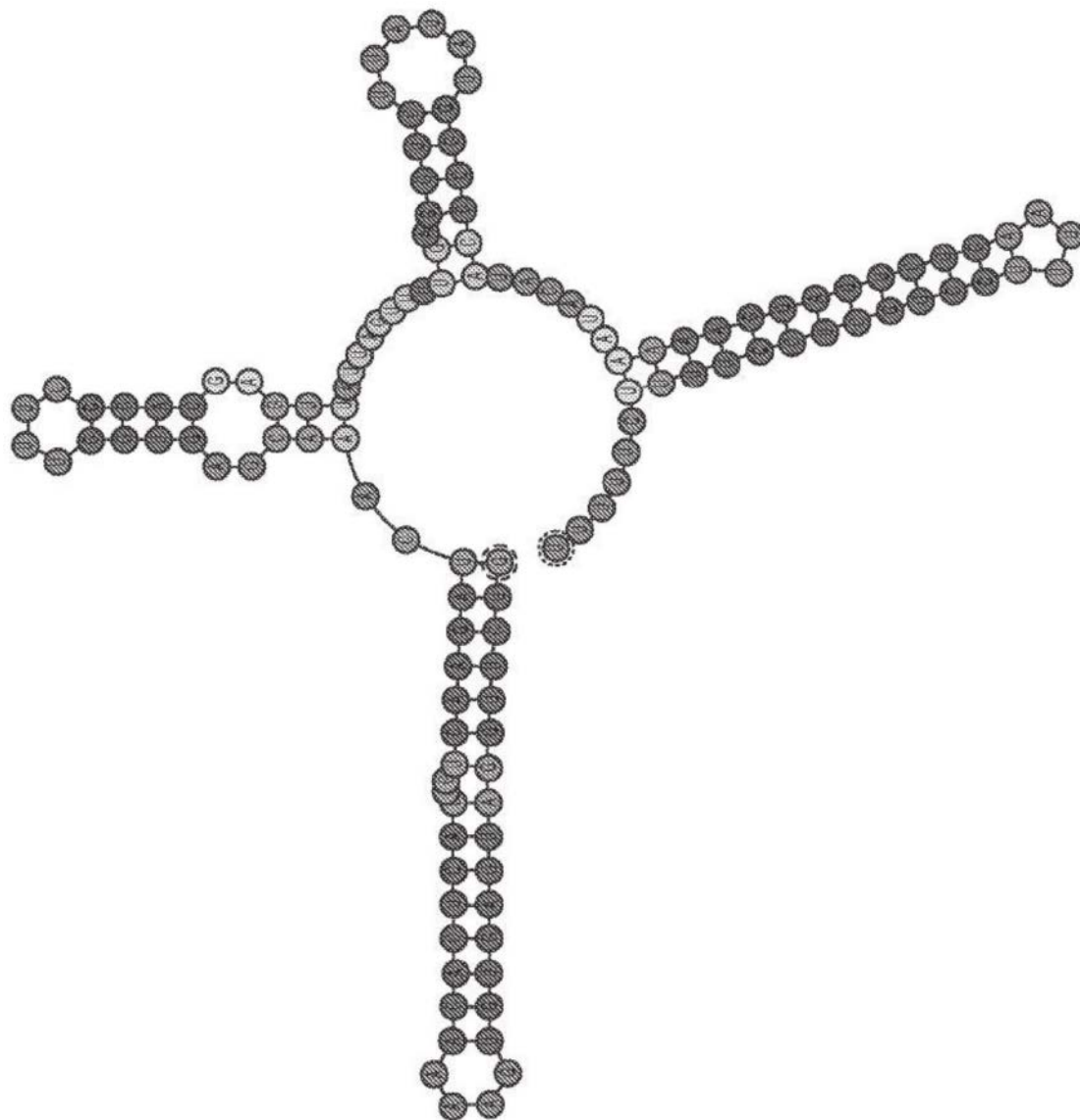


图2E

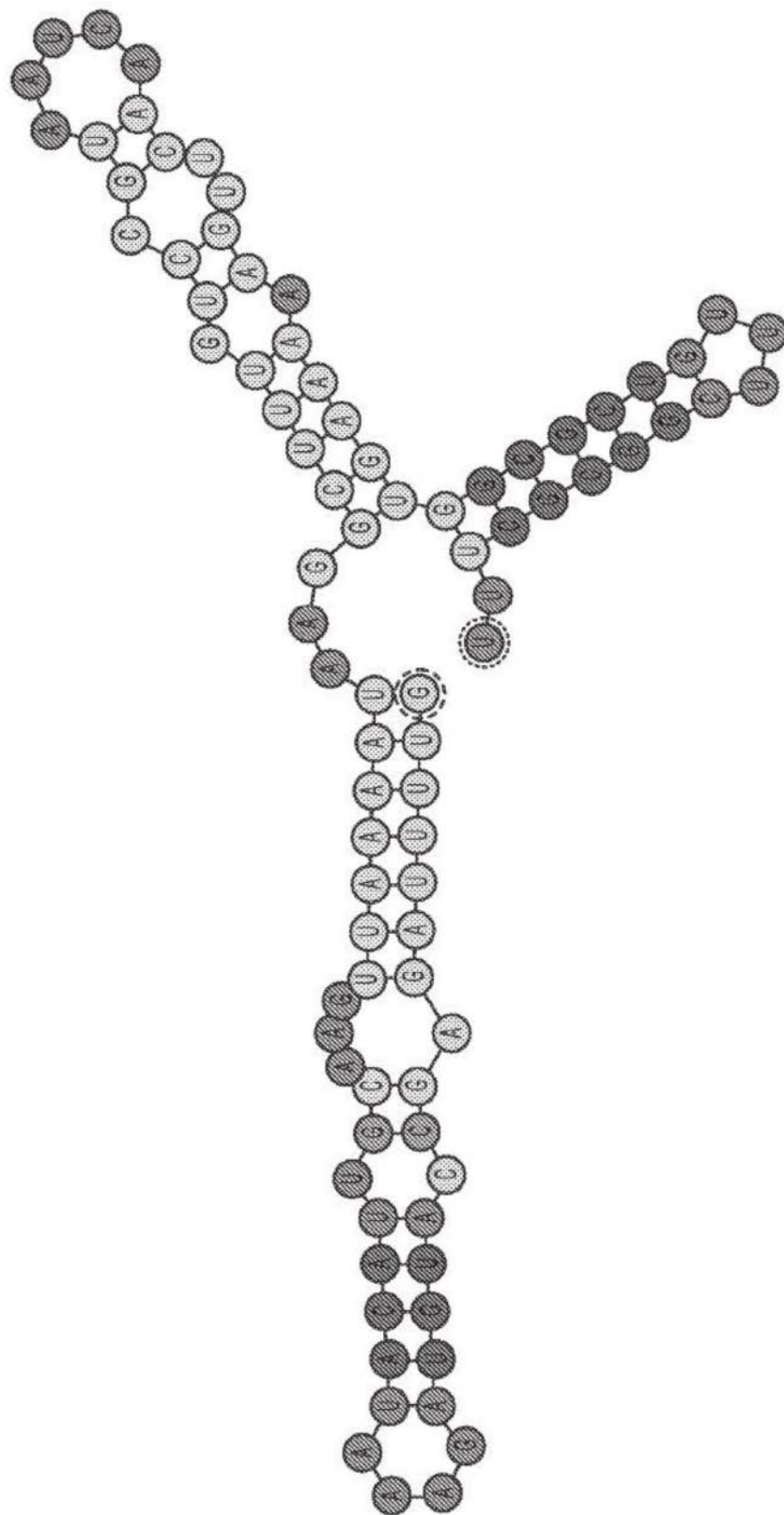


图2F

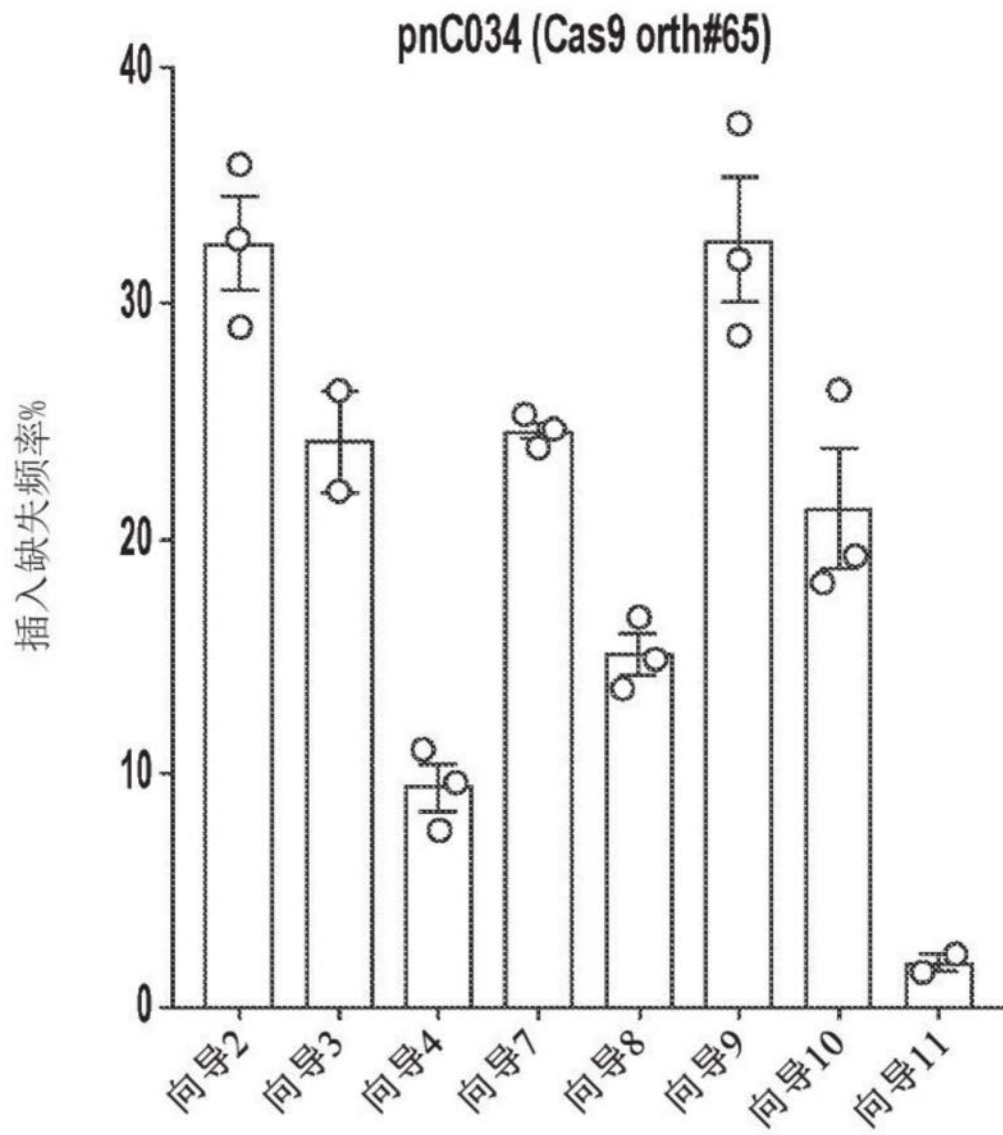


图3

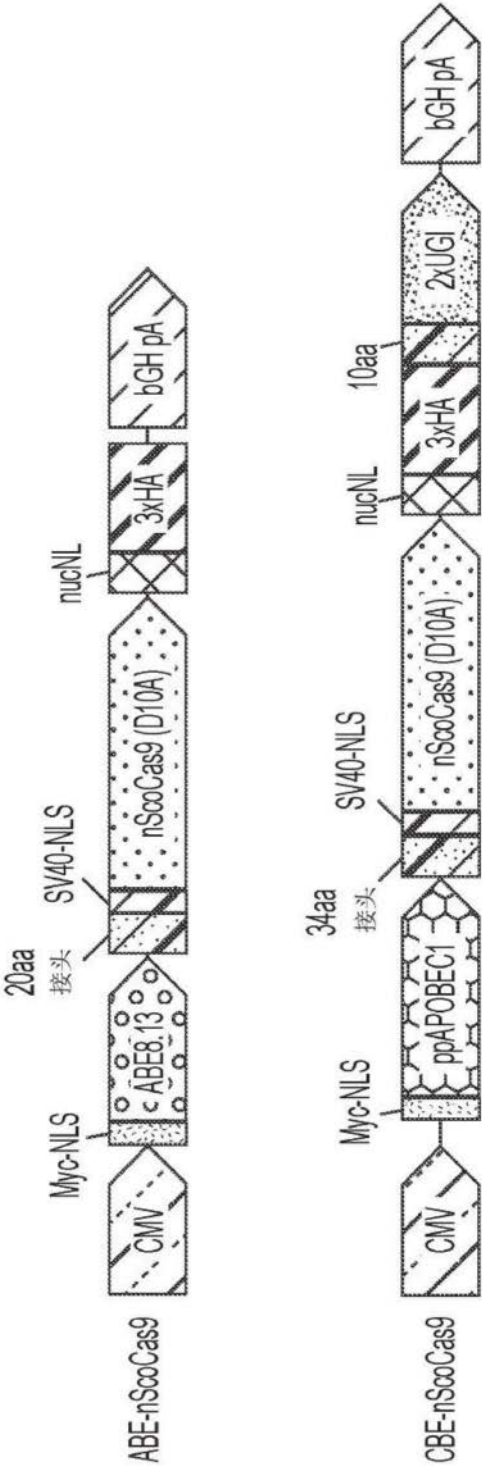


图4A

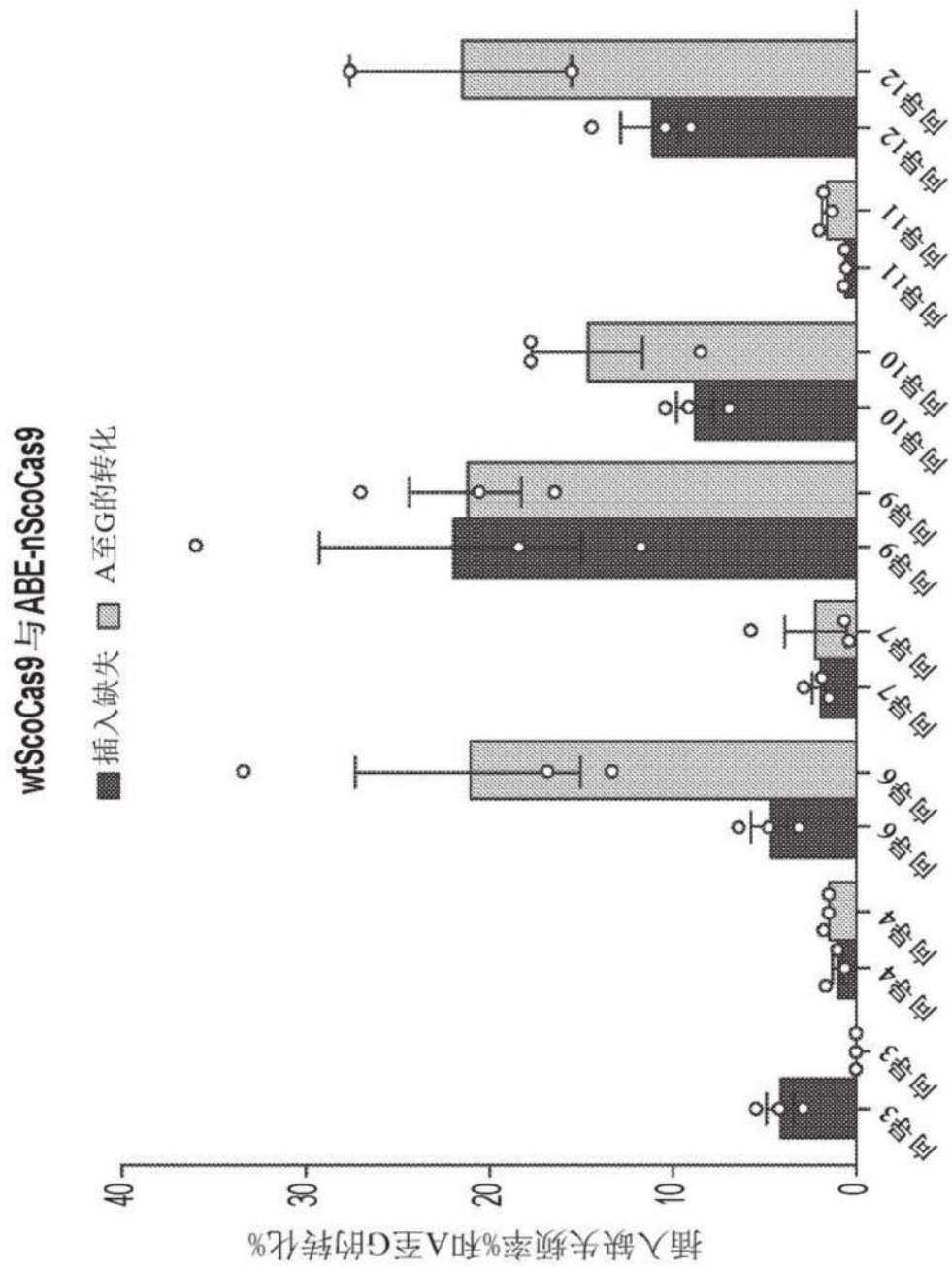


图4B

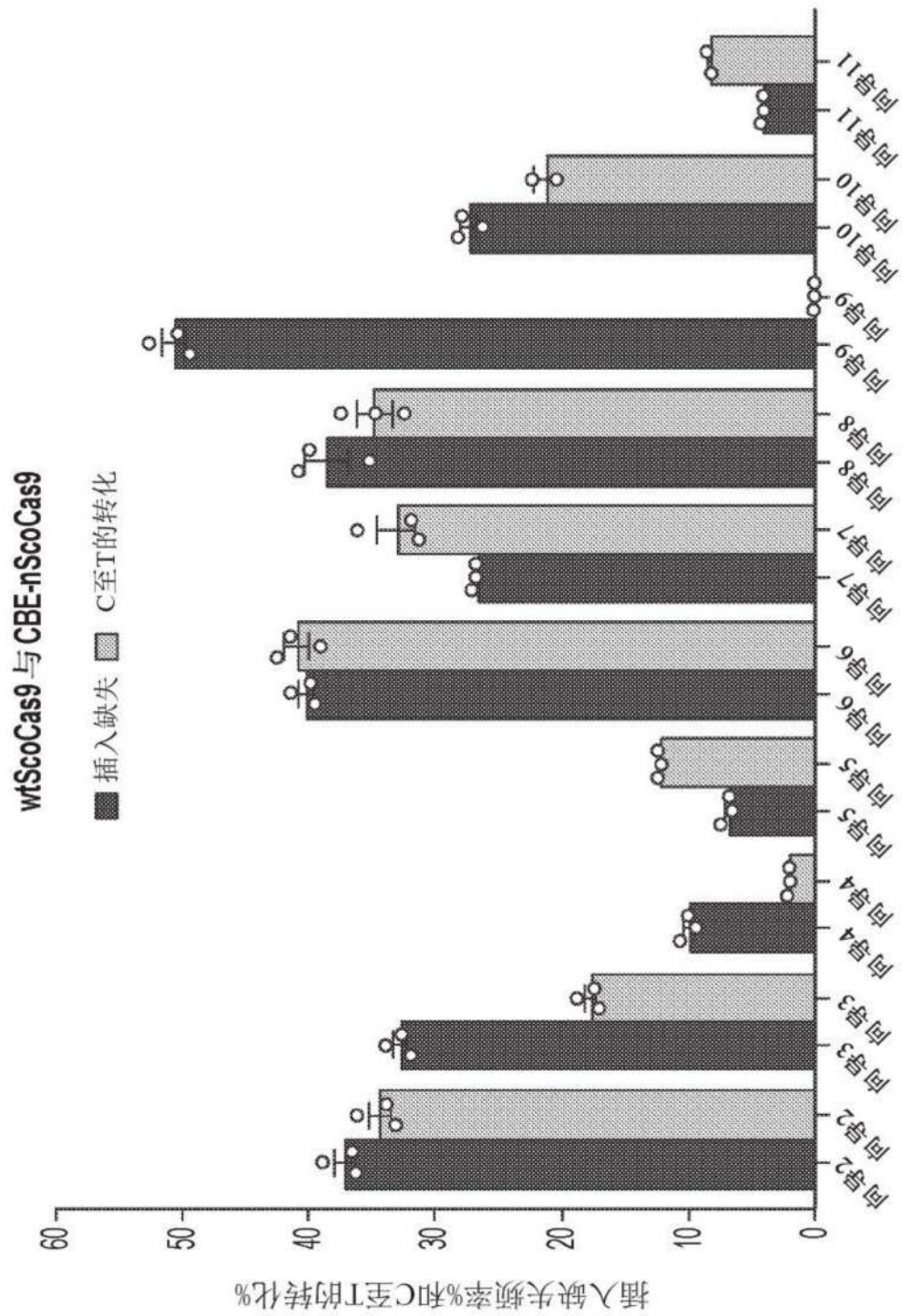


图4C

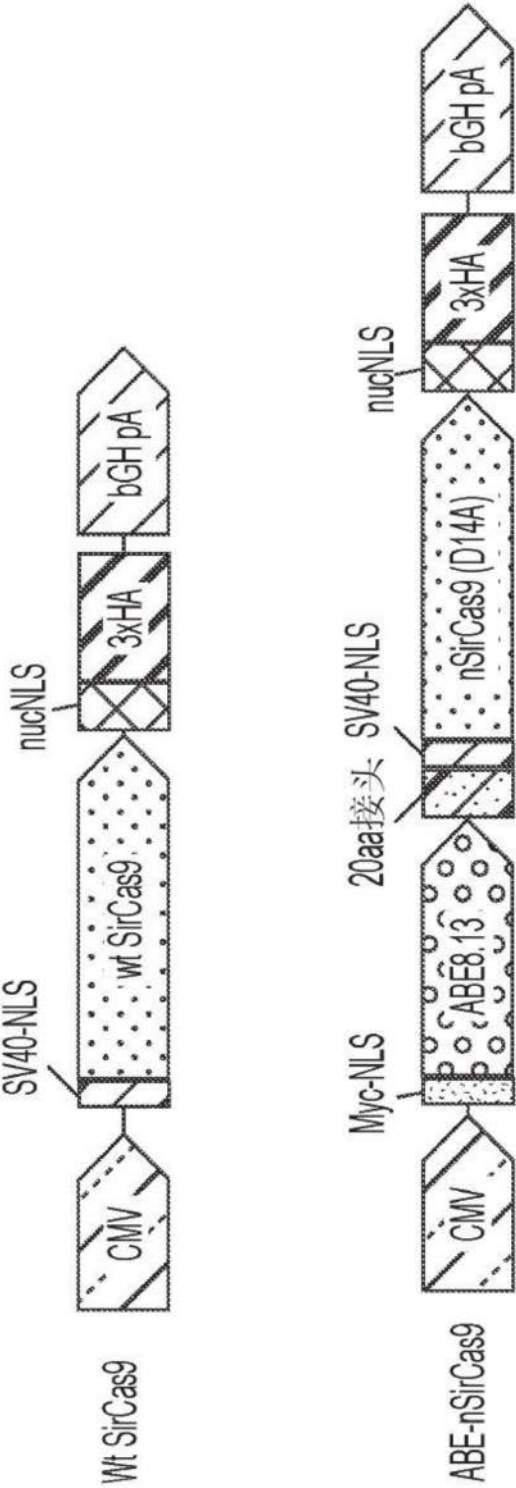


图5A

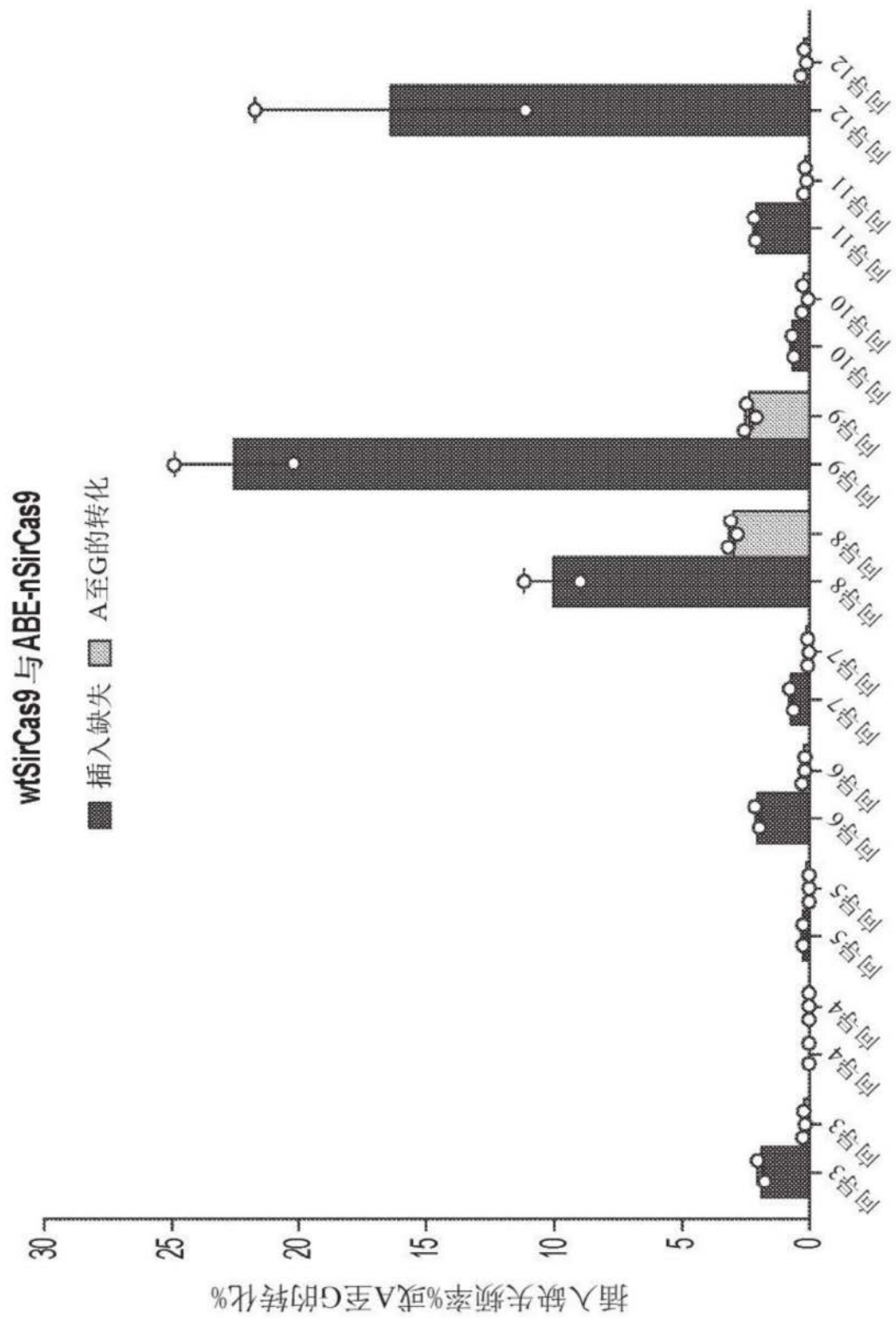


图5B

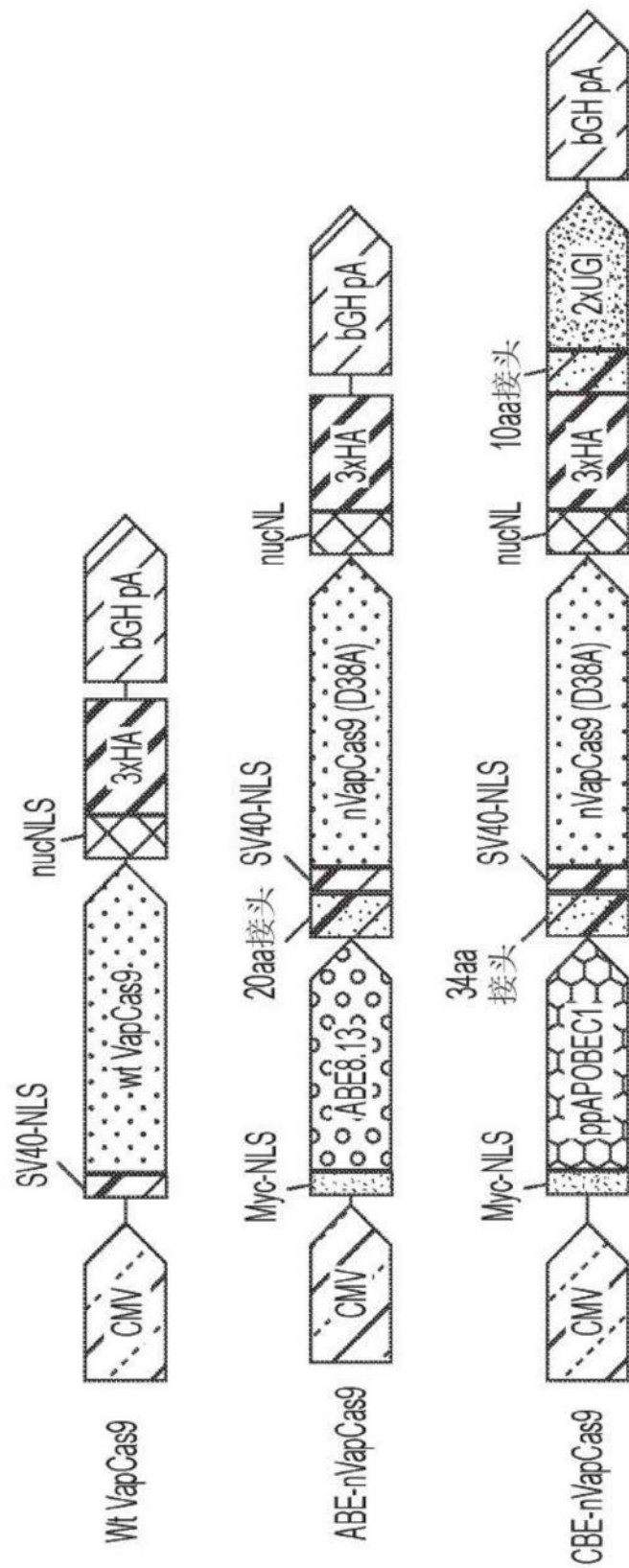


图6A

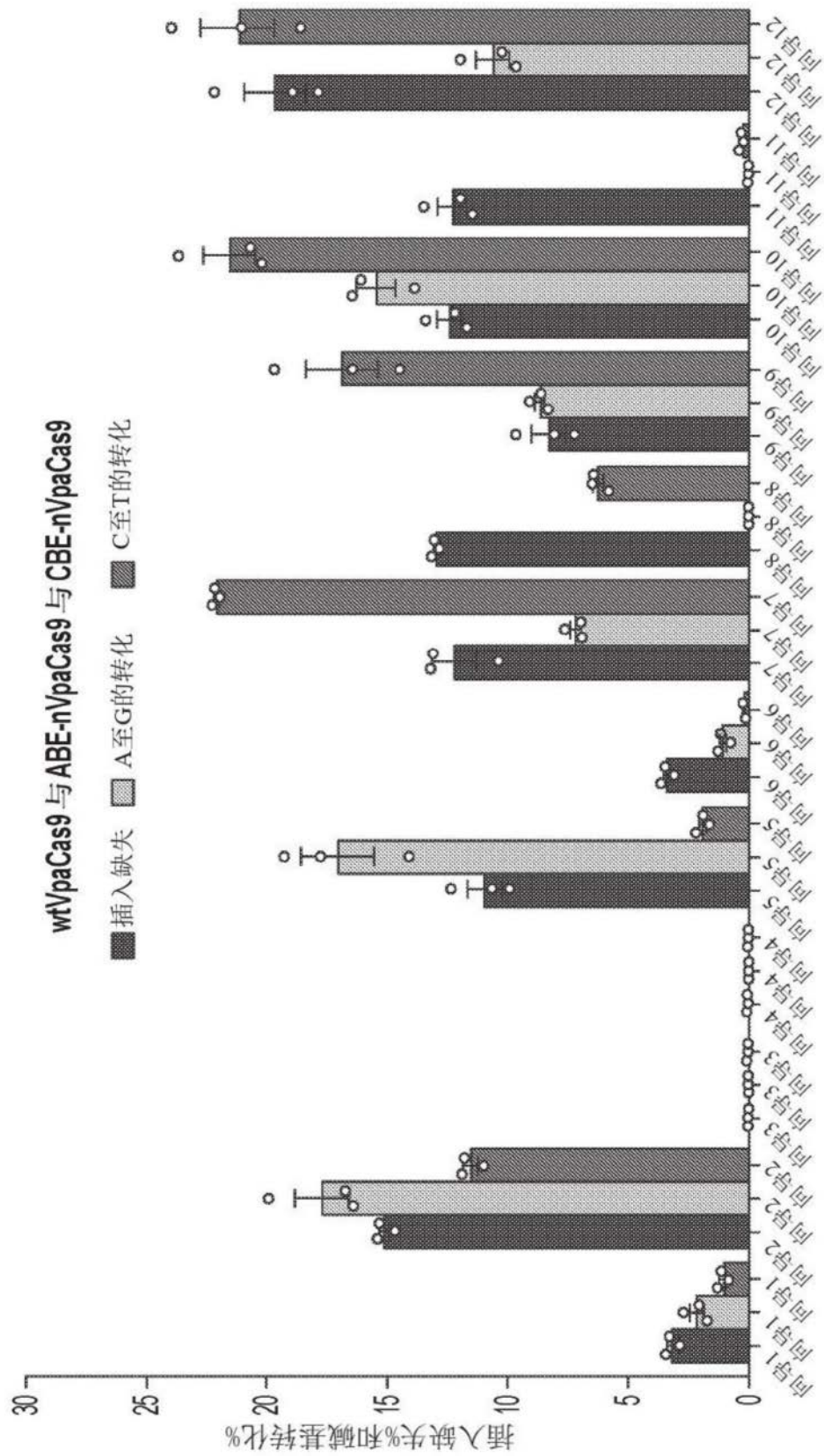


图6B

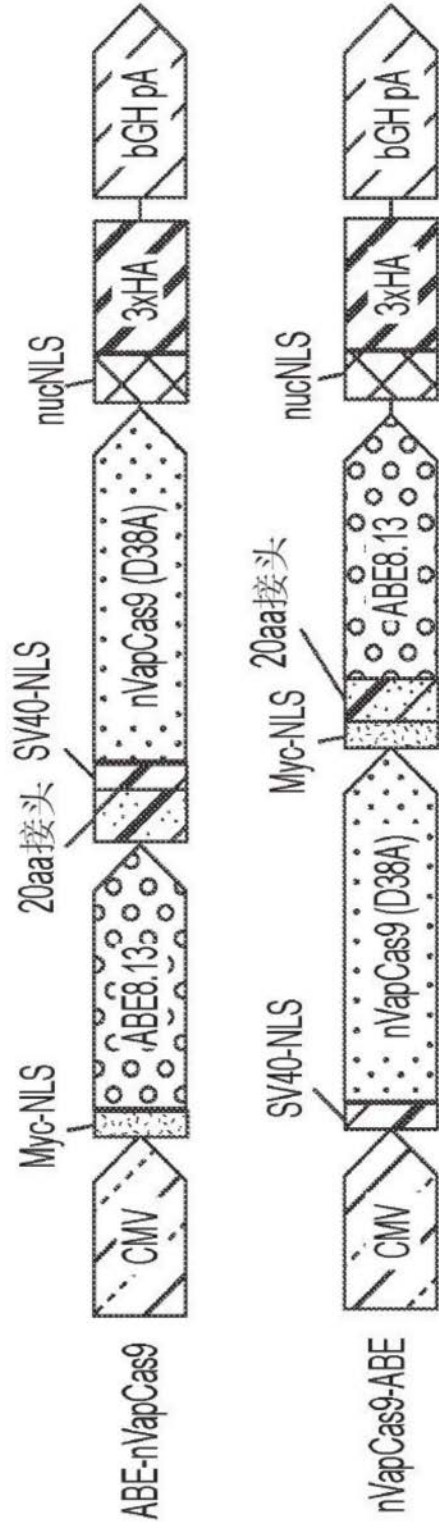


图7A

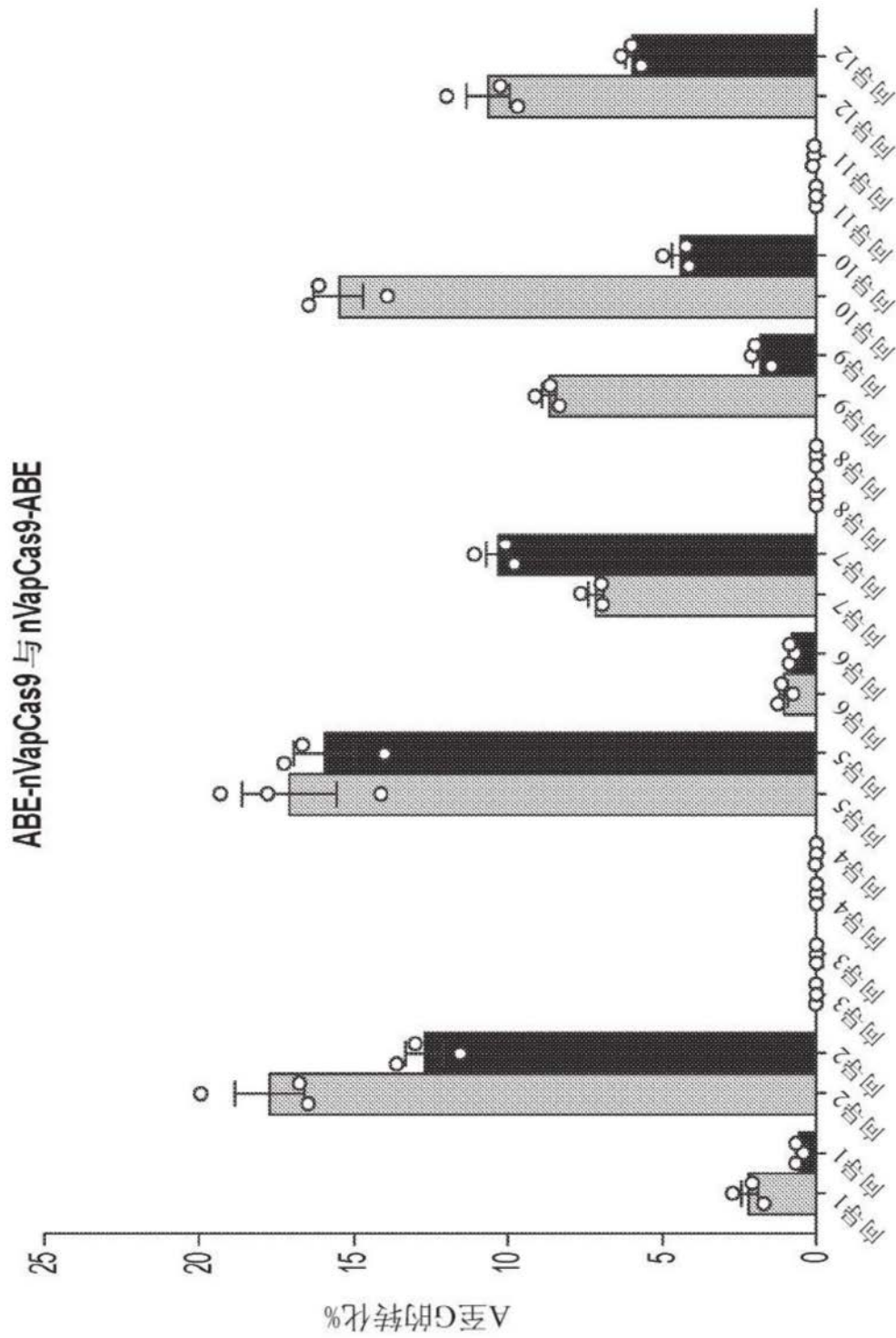


图7B

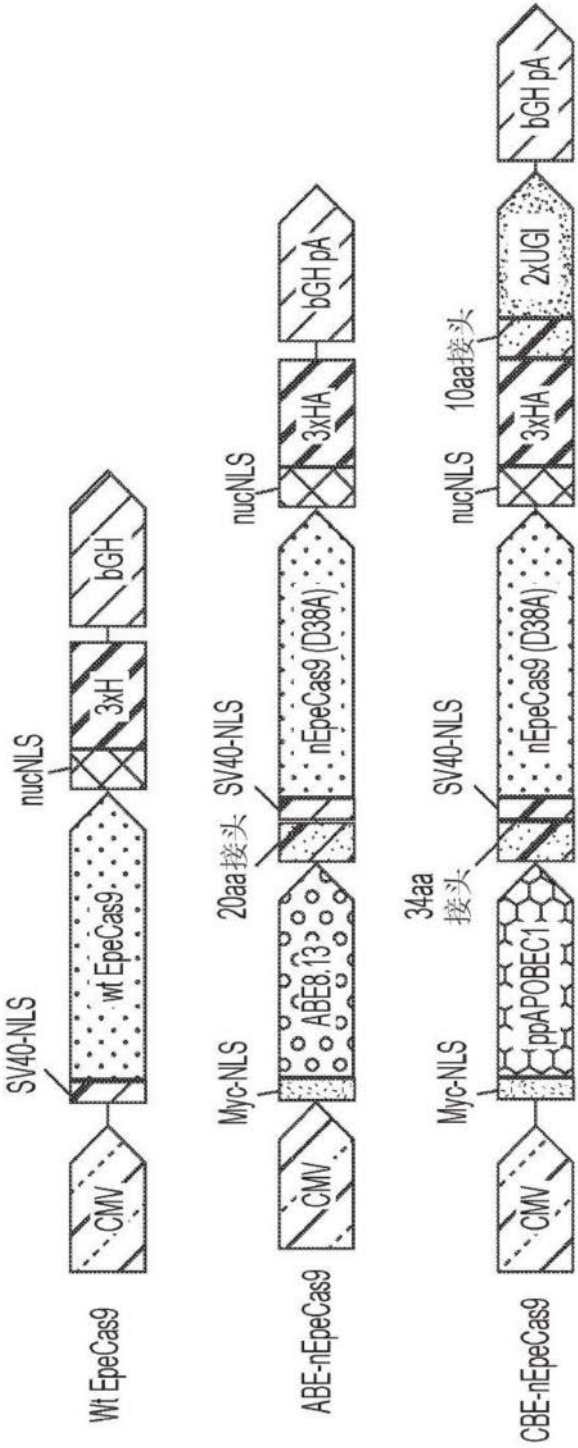


图8A

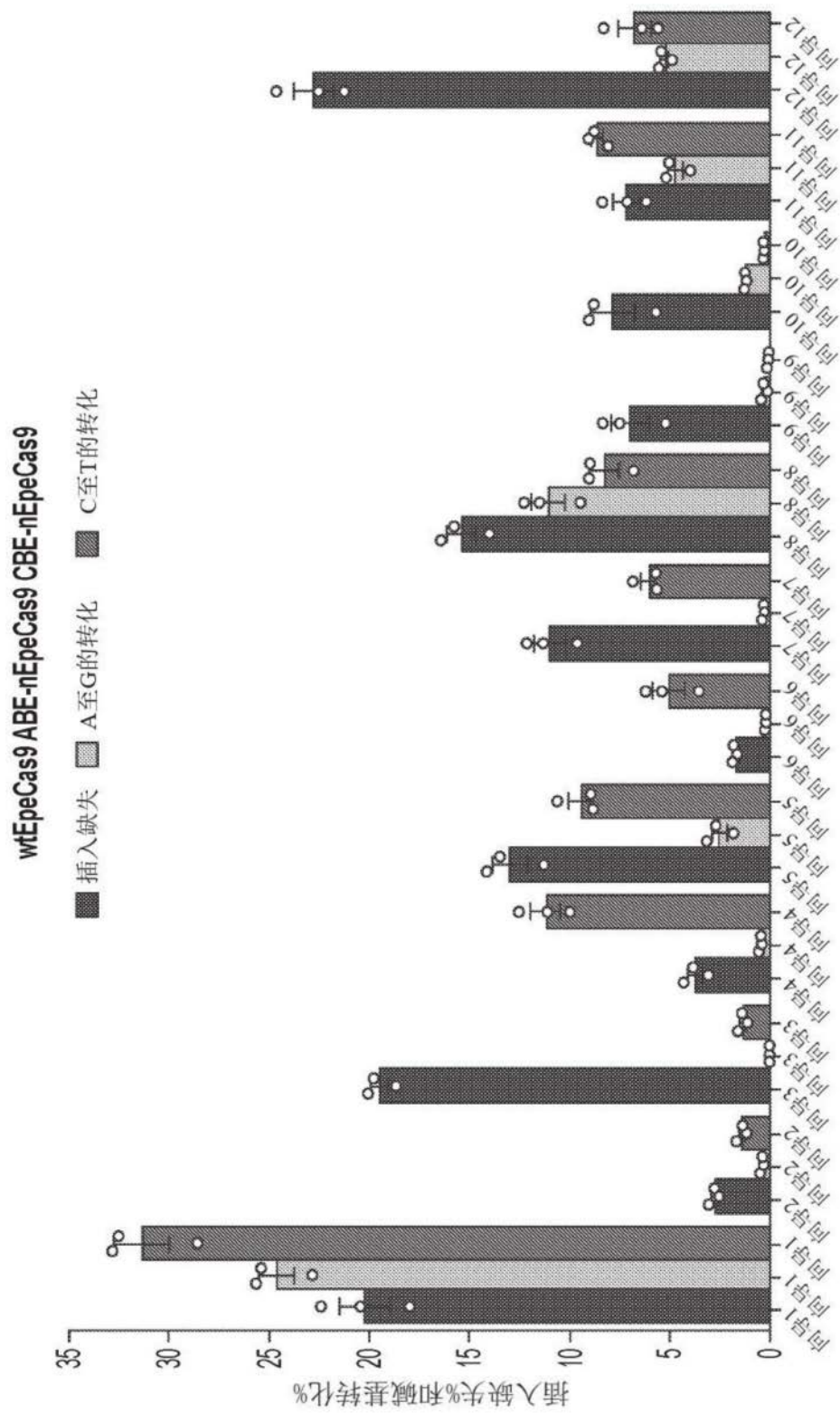


图8B

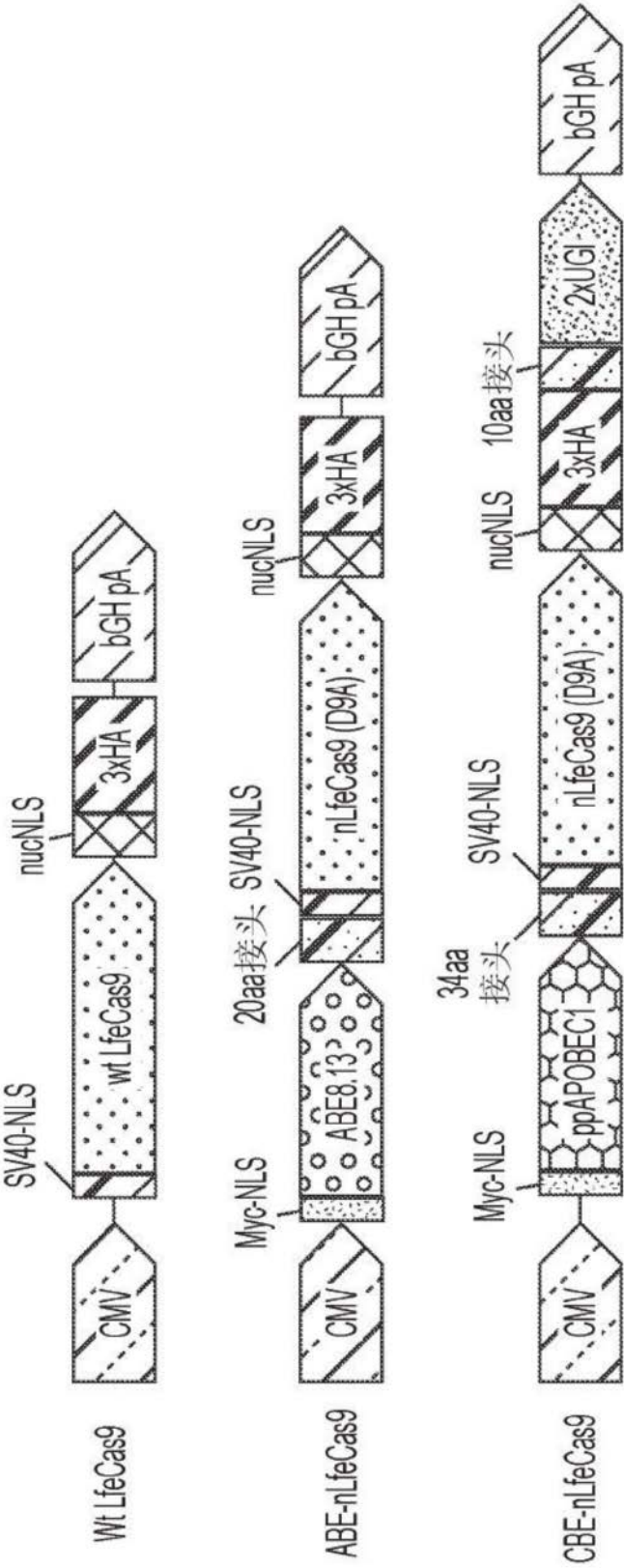


图9A

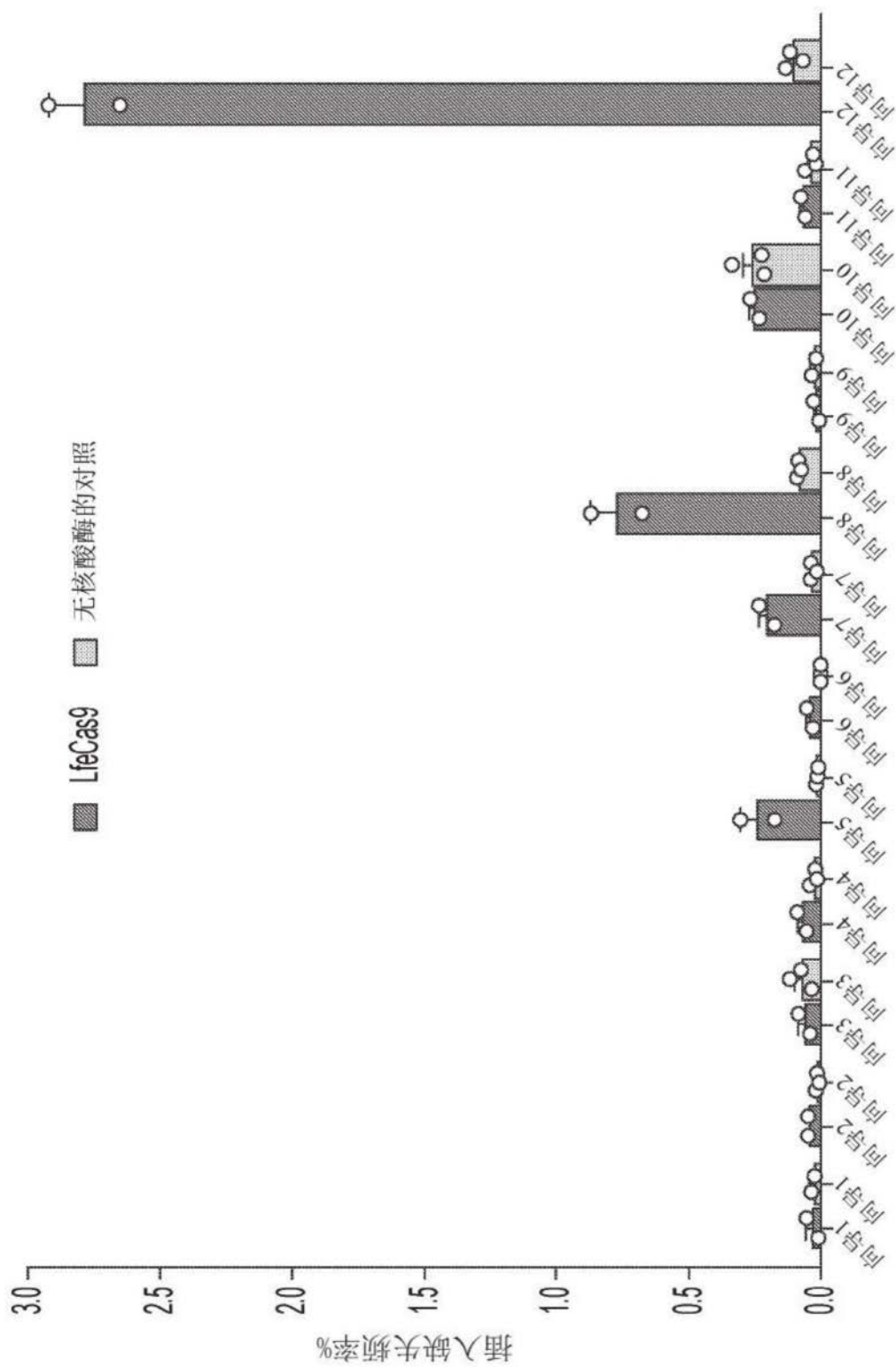


图9B

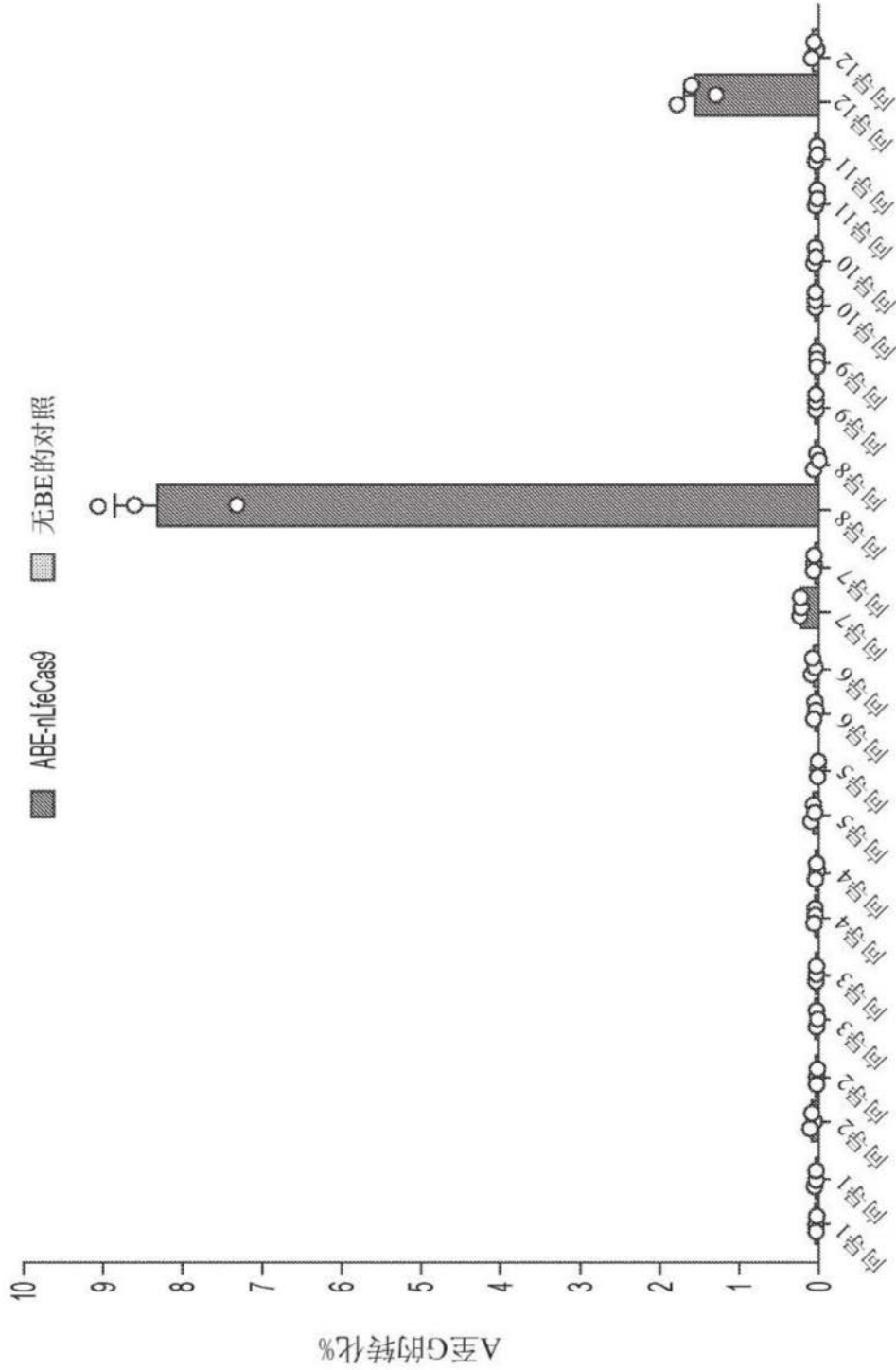


图9C

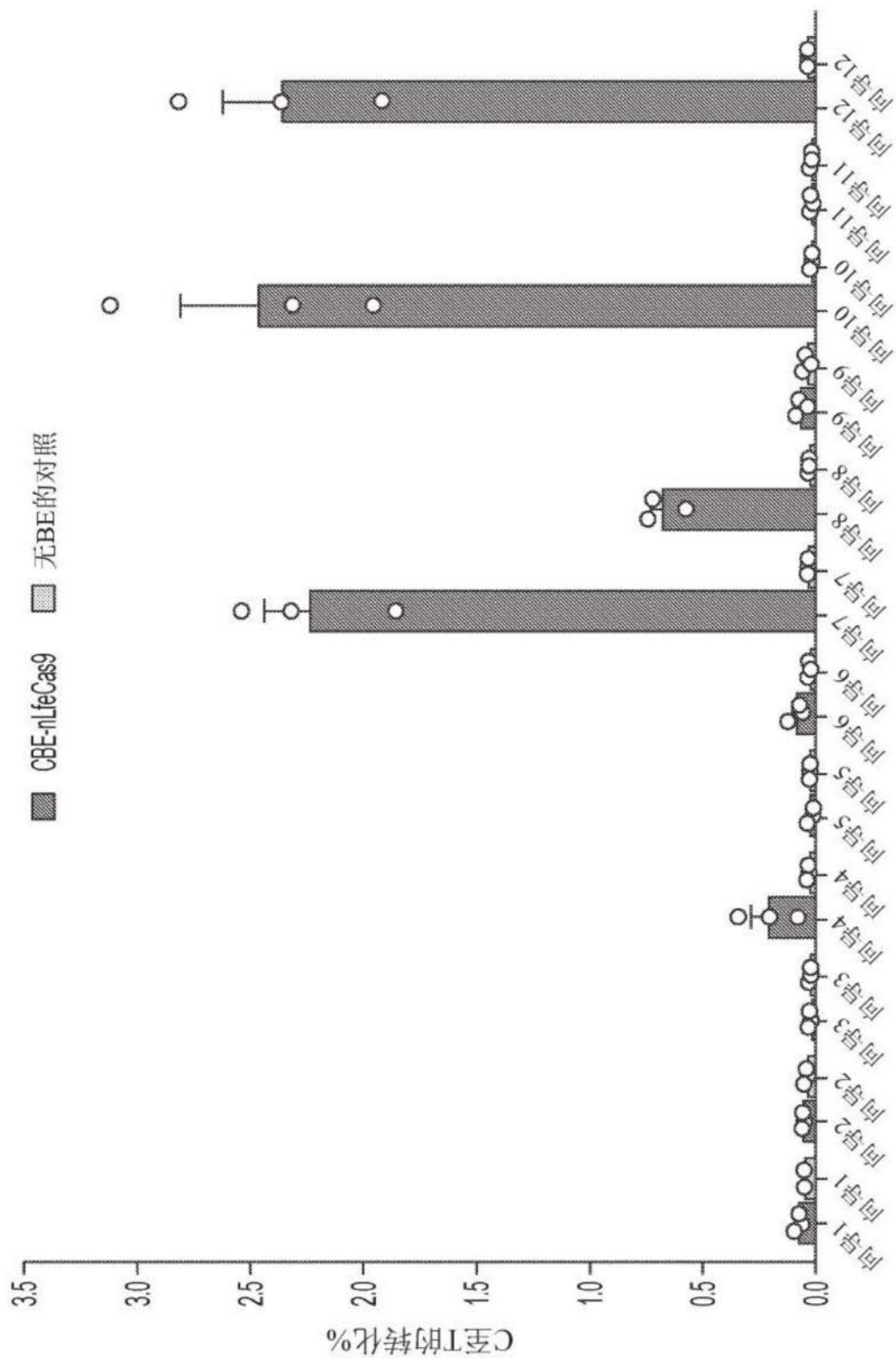


图9D

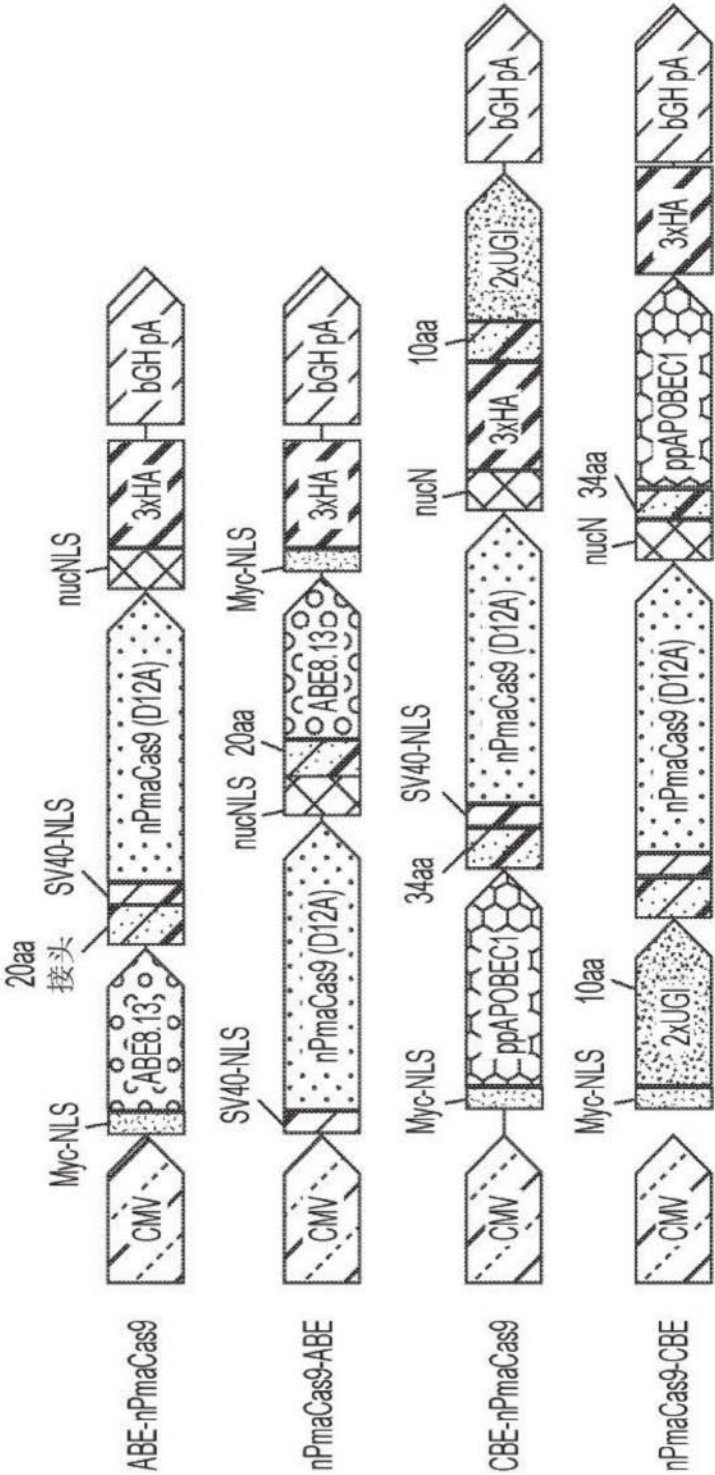


图10A

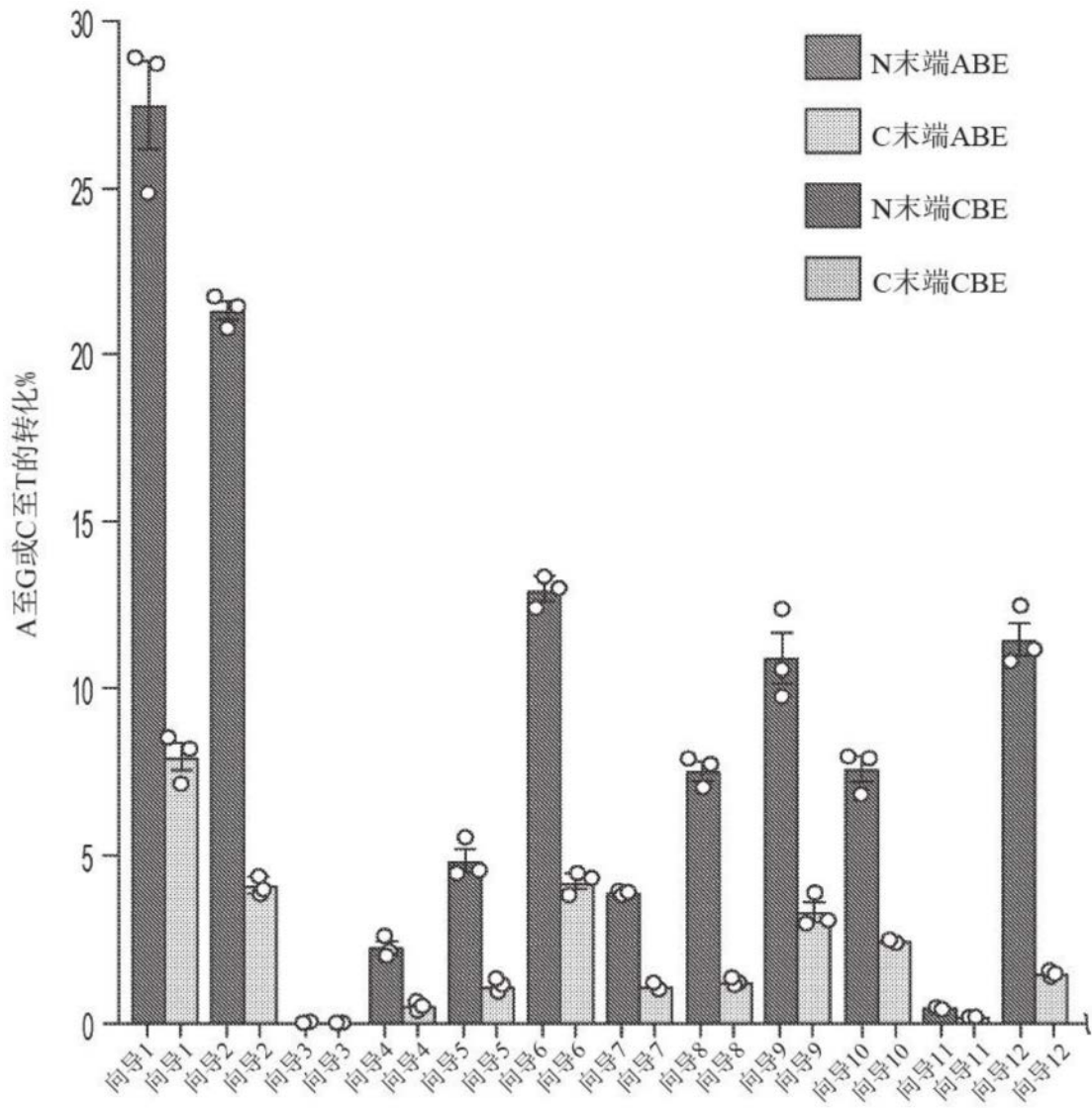


图10B

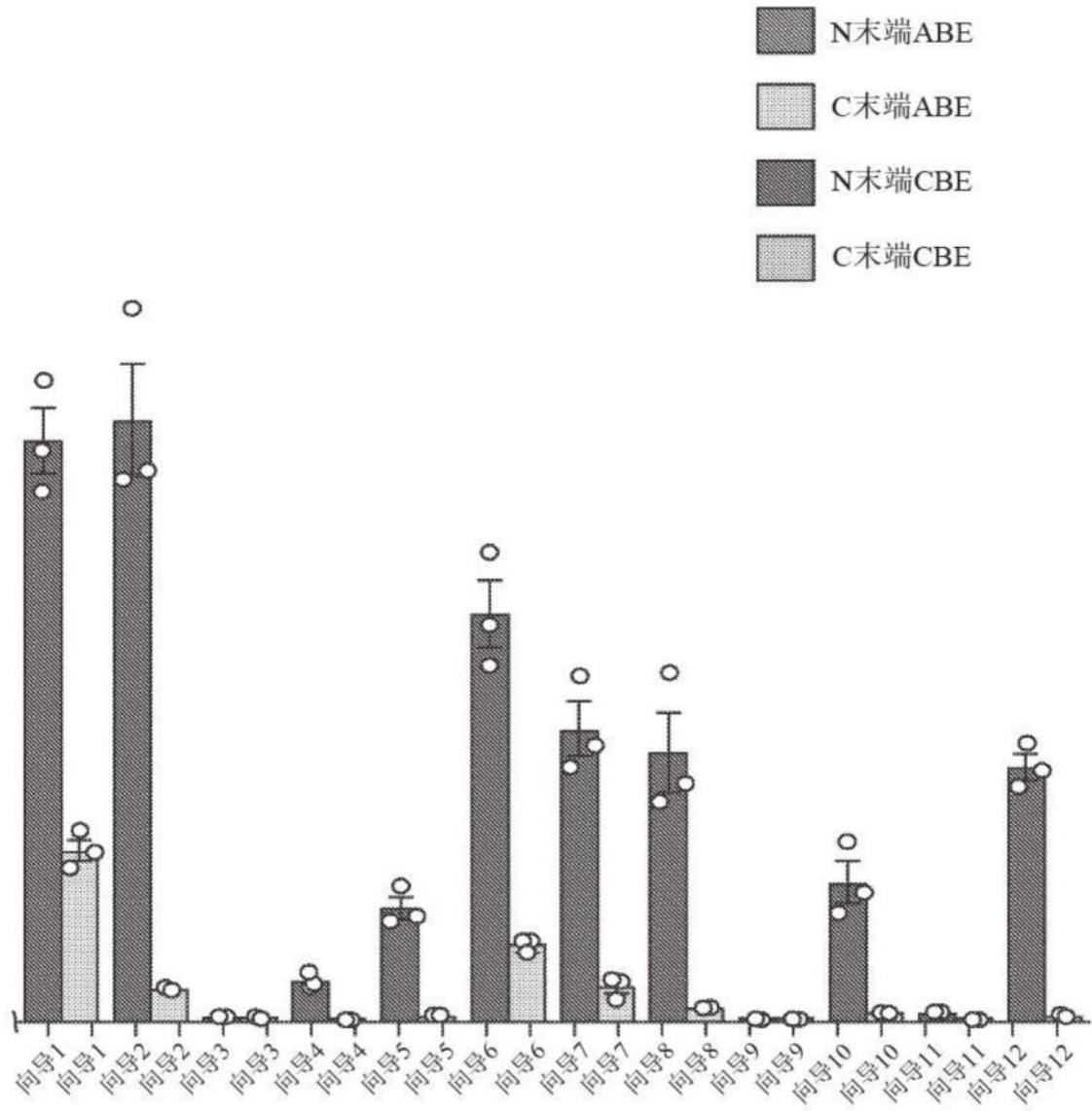


图10B接续

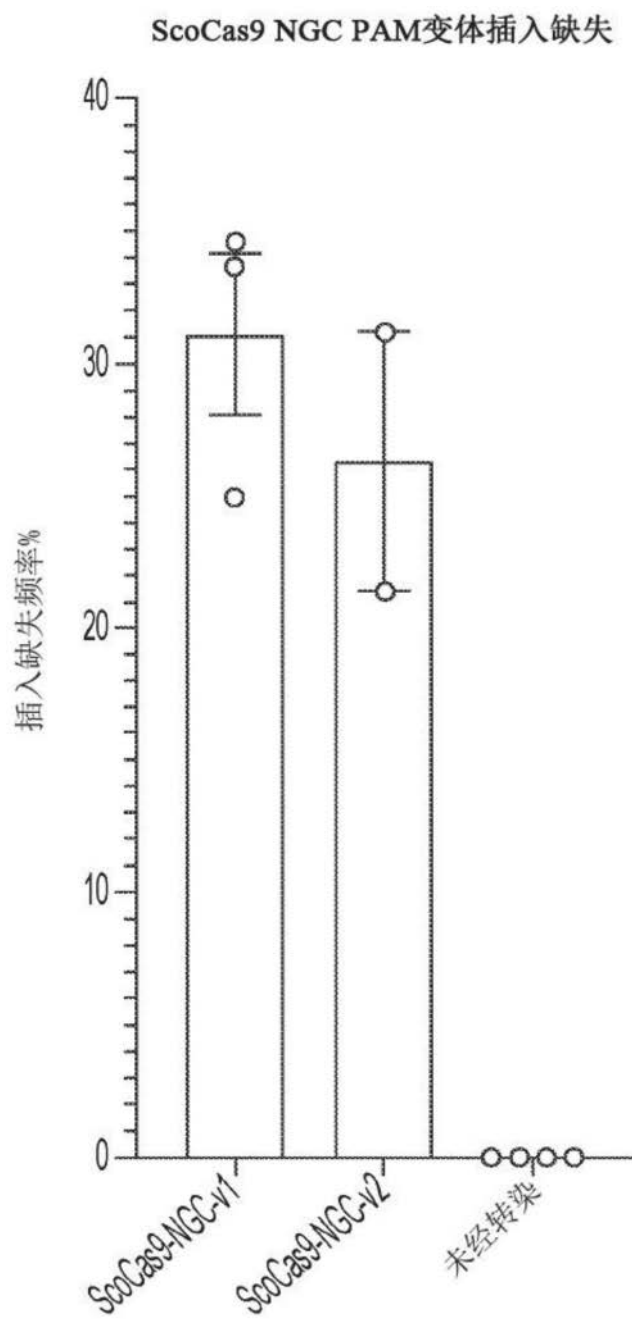


图11A

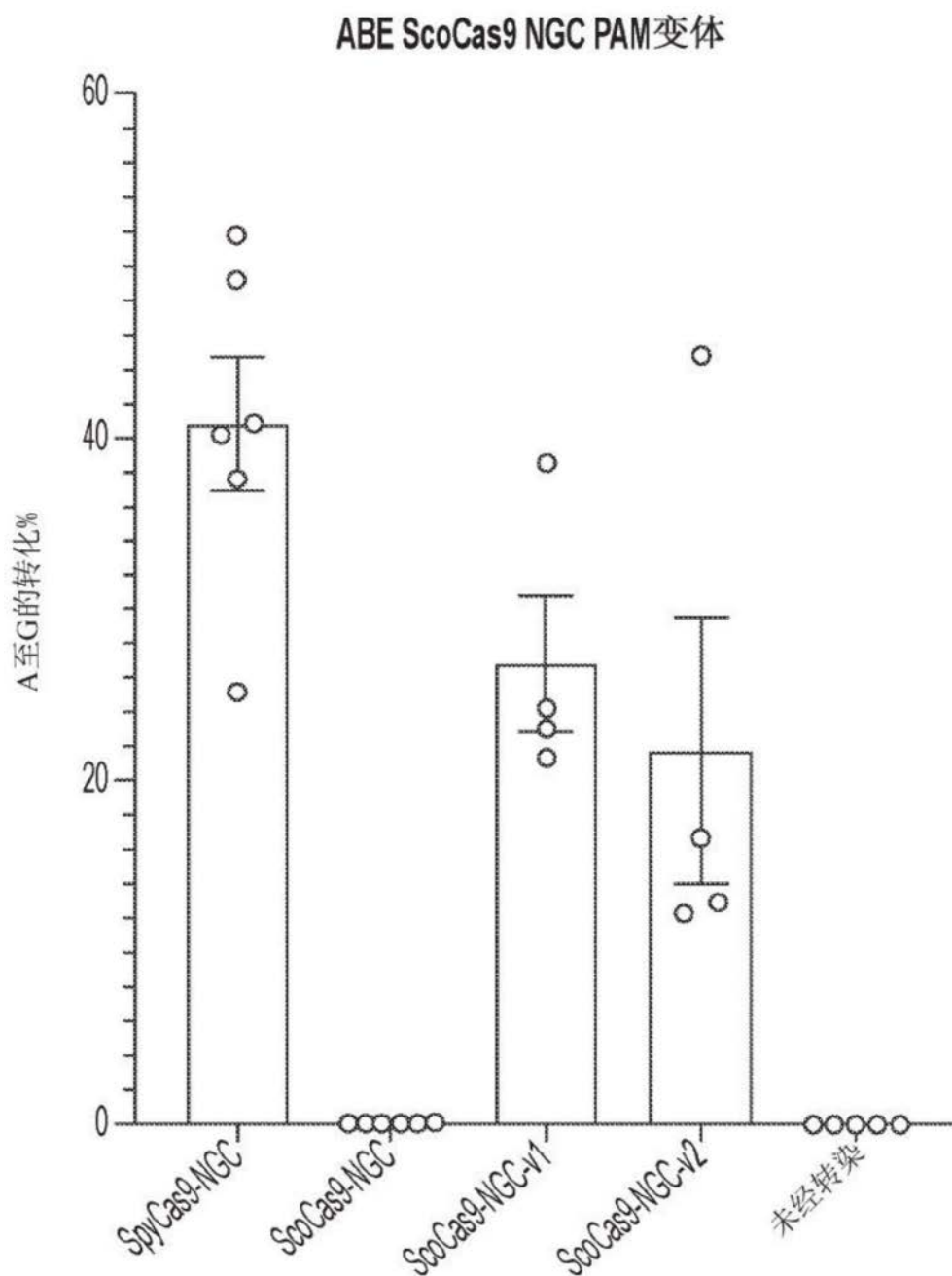


图11B