

(12) 按照专利合作条约所公布的国际申请

更正本

(19) 世界知识产权组织
国际局

(43) 国际公布日
2024 年 12 月 12 日 (12.12.2024)



(10) 国际公布号
WO 2024/251229 A9

(51) 国际专利分类号:
C12N 9/22 (2006.01) *C12N 15/113* (2010.01)
C07K 19/00 (2006.01)

(21) 国际申请号: PCT/CN2024/097935

(22) 国际申请日: 2024 年 6 月 7 日 (07.06.2024)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
202310684902.0 2023年6月9日 (09.06.2023) CN

(71) 申请人: 益杰立科(上海)生物科技有限公司(EPIGENIC THERAPEUTICS, INC.) [CN/CN];
中国上海市浦东新区中国(上海)自由贸易试验区巴圣路160号4号楼一单元5楼, Shanghai 200131 (CN)。

(72) 发明人: 冯争艳(FENG, Zhengyan); 中国上海市浦东新区中国(上海)自由贸易试验区巴圣路160号4号楼一单元5楼, Shanghai 200131 (CN)。吴垒磊(WU, Leilei); 中国上海市浦东新区中国(上海)自由贸易试验区巴圣路160号4号楼一单元5楼, Shanghai 200131 (CN)。毛少帅(MAO, Shaoshuai); 中国上海市浦东新区中国(上海)自由贸易试验区巴圣路160号4号楼一单元5楼, Shanghai 200131 (CN)。杨长青(YANG, Changqing); 中国上海市浦东新区中国(上海)自由贸易试验区巴圣路160号4号楼一单元5楼, Shanghai 200131 (CN)。朱琦(ZHU, Qi); 中国上海市浦东新区中国(上海)自由贸易试验区巴圣路160号4号楼一单元5楼, Shanghai 200131 (CN)。管婧雯(GUAN, Jingwen); 中国上海市浦东新区中国(上海)自由贸易试验区巴圣路160号4号楼一单元5楼, Shanghai 200131 (CN)。臧赢(ZANG, Ying); 中国上海市浦东新区中国(上海)自由贸易试验区巴圣路160号4号楼一单元5楼, Shanghai 200131 (CN)。

(74) 代理人: 上海巅石知识产权代理事务所(普通合伙)(SHANGHAI DIANSHI PARTNERS, P.C.);
中国上海市浦东新区盛夏路608号3号楼103室张琤, Shanghai 201210 (CN)。

(81) 指定国(除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW。

(84) 指定国(除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, ME, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

- 包括国际检索报告(条约第21条(3))。
- 包括说明书序列列表部分(细则5.2(a))。
- 包括关于许可根据细则91.1对明显错误进行更正的信息(细则48.2(i))。

(48) 更正本的公布日:
2025 年 2 月 20 日 (20.02.2025)

(15) 更正内容:
见 2025 年 2 月 20 日 (20.02.2025) 公布的公告

(54) Title: CAS ENZYME AND SYSTEM AND USE THEREOF

(54) 发明名称: Cas酶及其系统和应用

(57) Abstract: The present application relates to the field of biomedicine, and specifically relates to a novel Cas enzyme, a system thereof and the use thereof.

(57) 摘要: 涉及生物医药领域, 具体地涉及一种新型Cas酶及其系统和应用。

WO 2024/251229 A9

Cas 酶及其系统和应用

技术领域

本申请涉及生物医药领域，具体地涉及一种 Cas 酶及其系统和应用。

背景技术

基因组测序技术和分析方法的最新进展显著加速了对不同领域的生物活动(从原核生物合成途径到人类病理)的遗传基础的理解。为了充分理解和评估基因测序技术所产生的巨量信息，需要在基因组和表观基因组操作技术的规模、效率和易用性等方面进行相应的提高。这些新的基因组和表观基因组工程改造技术将加速许多领域的新应用的发展，包括生物技术、农业和人类治疗学。

成簇规律间隔短回文重复序列(CRISPR)和 CRISPR 相关联(Cas)基因,统称为 CRISPR-Cas 或 CRISPR/Cas 系统,目前已被理解为对细菌和古细菌提供抗噬菌体感染的免疫。原核生物适应性免疫的 CRISPR-Cas 系统是一组极其多样的蛋白质效应子、非编码元件以及基因座结构,其中一些例子已经被工程改造并适应于产生重要的生物技术。参与宿主防御的系统的组分包括一种或多种能够修饰 DNA 或 RNA 的效应子蛋白和负责将这些蛋白活性靶向噬菌体 DNA 或 RNA 上特定序列的 RNA 指导元件,这些指导元件可被重新编程以靶向可替代的 DNA 或 RNA 靶。

CRISPR-Cas 系统大致可分为两类:1 类系统由多个效应子蛋白构成,2 类系统由单个效应子蛋白构成,所述单个效应子蛋白与 RNA 指导物复合以靶向 DNA 或 RNA 底物。2 类系统的单亚基效应子组成为工程改造和应用转换提供了更简单的组件集,并且迄今为止一直是可编程效应子的重要来源。以 CRISPR-Cas9 为例的 2 类 CRISPR-Cas 系统的表征和工程改造为基因组编辑和其他方面的多样广泛生物技术应用铺平了道路。然而,除了当前的通过其独特的性质实现了新的应用的 CRISPR-Cas 系统之外,仍然需要开发强有力的基因组工程工具,即用于修饰核酸和多核苷酸(即 DNA、RNA 或其任何杂合体、衍生物或修饰物)的替代性的可编程效应子和系统。

发明内容

一方面，本申请提供一种分离的 Cas 酶，所述 Cas 酶包含 SEQ ID NOs: 1-58 中任一项所示的氨基酸序列或与所述 SEQ ID NOs: 1-58 中任一项所示的氨基酸序列具有至少约 80%同一性的序列。例如，所述 Cas 酶包含与所述 SEQ ID NOs: 1-58 中任一项所示的氨基酸序列具有约 80%、约 81%、约 82%、约 83%、约 84%、约 85%、约 86%、约 87%、约 88%、约 89%、约 90%、约 91%、约 92%、约 93%、约 94%、约 95%、约 96%、约 97%、约 98%、约 99%、或约 100%同一性的序列。例如，本申请提供了包含如下表所示的氨基酸序列的 Cas 酶：

SEQ ID NO:	本申请编号	氨基酸序列 (*可为任意氨基酸)
1	EpiCas001	MNNNGETKCYGNNLDRFIGLREVNKTLRNELIPTDSTRENIEKNGHIISEDEL RAEMGHELKAIMDDYYRNMIEETLSRKCINWSTLFAAMDSVARSANKK DADKILEKVQKEKREEIHKLFEQGNFEKLFKAELVSDVLPDFIGKYEGYS DEERISKLGTIKLFKKFMTSFKAYFDNRKNVFSKENISTSICHRVVNDNAFIF WDNLNAFKKIDKYAHDVETIEGEISGKLGWKISQIYSVDFFGMVIPQSGI TFYNEVSGEINKYMNIYCQKEKQNSRQFKLRHLKQILCIGETSFEIPEKYE SDEEVYQSVNDFIGMIKNNDIISRLEHLAENVNKYDADKIYINSKSYENVSK CIGGTWNTIKDCLRLWYNENIRGSGKNKEKKVGKAVEEDKYKSIAMLNQL LENYSDGDGRIYNVQLYVEHITKIIIGDYKPEELVYDANIRVIEQDEKAEEIK SVLDIYIAILHWIKAFIIEDAVEKD TDFYIEIEEIFDKVQSIIPLYNMVRNYTT QKPYKQEKIKLKFNTPTLAEGWHKSKEYSNNAILQRDGKY YLGIFNAKNK PDKKIIEGCSQKKNKADYAKMVYYLLPKPHMNLPHIFIKSKEAVKNYGLS NYIVEGYDKDKHIKSSNGFDIKYCRDLIDYFKDCISKHPEWKNFNFSFTDTA LYNDIGEFYNEIKRQGYKIEWTYISEEEINRLDENGQIYLFQIYNKDFAEGSK GTPNLHTLYLKNLFSEDNLKDPVFMLNGNTEVFFRKSSI KPVVHKKGSVL VNRNYIDLNGNRITPEKEYKEIYEYCNGIGNVNLSESAMKY YDMAKHYT ATMDIVKD YRYTVDKFFIHMPMTINFNAESNNNINELALNFVAKNDNIHIG VDRGERNLIYISVIDSKGNIIKQKSYNVVNDYDYKSKLKEREYDRGDARKN WKRINNIKELKEGYLSQVVHEIAQLVIQYNAILAMEDLNYGFKRGRFKVER QVYQKFENMLINKLNYLVDKRKNAD EAGGV LKGYQLTYIPERLKNMGRQ CGIIFYVPAAYTSKIDPTTG FVNLFNFNAITSQEGRRK FINKFKHIRYNLMN

		LFEFEFDYNDFITHNTQIAKTEWSVYTNGPRIKKERVNGKWNDGYEADITE VMKQKLSDNGIDYADGGDILTRIAEADDKLVSGIFEIFRLTVQLRNSKSESC GEDYDRIISPVLNESNRFYDSDEYSGENEILPVDADANGAYCIALKGLYEVEV KQIRENWVEGDKLSGDLLKISNADWFDIFIQNKRYE*
2	EpiCas002	MNIKFNKFTGVKPVSKTLRNALIPTEFTMVNLERNGIISADELRAEKRQELK KIMDDYYRDLIERTLAENHSINWHILFELMKSKMHGDEKDTAKKLENEQK SKREEIHRLEAEQEEFGKIFSACLVSILPGFVASYAMYNDedrTEKAEVV KLFNKFMSSTEFFENRKNVFSKENISTSLCHRIVNDNALIFLNNMDSYKRI MENFKSESELVSKGLETMLGEWKISQIYSEDFFGMTMTQSGIEFYNDICGEL NLVLNLYCQKSKKNSARFKMKRLHKQILCTEKSSFEISYKYDNDKDVTYS VNGFIKNIKNNIIRLLKLTETEQYDLNKIYISSQSYETVSIAIGGTWETIR NCLTSWYDDNIKSGSKAKENKIKAVVKADKYQSIARVDYLTETYNASGTD LQKAESYLSHIVQIIGDYEPEELYFDESIHLIENGDKISEIKKILDIYIDVVHWI KTFEISESVEKEPEFYELDDIYDKILDAVPLYNKVRNYVTQKPYNQEKFK LNFDSPTLANGWSKSKERDNNAILSRDGKYLLGVFNVRNKPDKLEIGHS KPKNETDYAKMVYNLLPGASKMLPKVFLSKKGIDNFKPSDYILEGYIKKK HIKSSETFDIKYCRDLIDYFKNSISRHPWKNFGFNFTETSKYNDISEFYKEV EKQGYKIEWTYITEEDINKLDEAGQIYLFQIFNKDFAVGSKGTPNLHTLYLK NLFSKENLDNVVLKLNGEAEFFRRSSIKTPVIHKKGSILVNRCYVDADDG KRKPIPEKEYMEIYNYCNGKRDTELSERKYYNISQRYEATMDIVKDYRY SVDKLFHLPITINFKATENKNVNDMVLEYIAKSDDIHIIGIDRGERNLIYVS VIDMHGHIVMQKSYNVVNEYDYKTKLKECEQGRDAARKNWKEIGKIKEL KEGYLSMVIHEITQLIIEYNAIVMMEDLNFGRFKRGRFKVERQVYQKFENML INKLNLYVDKHKAVDEPGLLNGYQLTYIPDRMSELGRQCGIIFYVPAAYT SKIDPTTGFFNAFRFNLLTSAESRRDFLNRFKCIKYNNGSMNMFEDFDYNDF ITHNTTIARTQWKVFTNGGRIKNIISKSGWNGTKKIDDITALMKEIFDSRKID FSDGGNLLPEILAYEGADKDKFIVELLDIFKLTVQLRNSMPDGSIDYDRIIS PVINENNTFYDSDMYTEPNSPLPIDADANGAYCIALKGLYEVENQIKENWKE GETFPKSVLKLNNANWFDVQNKRYE*
3	EpiCas003	MNLAFLNRGYNYGMINRIERGPFSELGDQQLLEQTRRLAANQRCLEVHILD

		HLDEIDRRGLALRRGFSSLFDYAVRELRFSDAAAQRRIQAMRLCRRHGCV RASLQSGELGMTSAAQLETTFAGAERAERKSHTGHRNVGQGRGPDQDQD SVGSTRQLGLPAAPGEADEQGERGHRSAPEGAPPAAECSGDARERGHSA PEGAAPAGEYTGEVAARTTPAVAVAVASRSGAQPPGGVHAGALLPRPELN GNAAPVQAPSRAPVAAPGAPGAGAGAGPAAPASDEGPAAQLLPTEFSAS TPEPEPVTAPAE LRHTASRAPEPRDDPFAELVPGTVSAAPLLHPQRQRELIE QAAGMSTRQVASLLAAAPEVLPPRDTLRVAPDRYTLKVSIDQECEQGL RLKDLMAHLDPRMSWGD LVARVVREAVARHDP RGGGRGQRRKGADSA GASPRRTPKETRAAGAAGGPARRVGESTGREPNRTGAAGAAGGPAQREDE SSGRAQNRTRA AVAPVAAAARRRGGATSAPQSAPAVLVDGARAGGGPAGA PAAPPAPPGDTAATERQNPIEVL TSAAPPSQSAPAAPSDGAHAGGAPAGAS AATSAPQSAPAAQLD GARAGRVPSGAPAATSAPQFASGALLDGSNAGAGV PGAPVATAAPQPALGGSPDGGPSGASAGASAGAAGGASADGFSARAAGPG ASVATSAPKSVLGGFPDGGPPDAAAGAAAHTSREPVHSRRQSPRRRHIPAA VRRHIWHRDGGRC CYRDPLTGRRCTSAHLLQIDHLPVVAEGGGPEPSNLA LRCFAHHRMRHGYAPAASPQPP I*
4	EpiCas004	MDGLARVDRAVTACLAAPAWALPEEALVEALDEAHRLGQRLAVLERVLV REVDGRGLARAQGATSTAGWLRDRLRVS IQSARRLVADAGWLDAADDPG TTDDPGTTGDPSGTTDDPGTTGDPSGTRTYASVEPGTGIGAAGAPSPADAL PDAAGRTGLDGA VTTGPGPGAGGAGGAGGADRAGAPGLSAAAGCGTAA AGPVLREALATGRVNLEQARI IAGAVESVRKEAGPDTADKALYTLLADAR RFEPRALRALADRIL WHVAPDAAEEADRAALRRADDEA QRRRH LTL SATD DGRMRLRGLLDSETSATLLAALDPLTRPAGLGDDRTSGQRRHDALGDICR LALRTGELPDNGGEP AQIVVTTEYAALTARLTAVGLPAARQPGATSGTRSS IARPSPEPAPAERGPT EPTPAEQSGSIESTPAERRPTEHTPAERRRTERMPAER GAARRTQSGIGRLRRTTHERGGCGPRPPQRAAHNCGQPGR TGRAPAAGSG QRRPERPVPRIRNGTAGTTGTATGRWQRPRGAPTETPTETAVSGGPPTPQSS PPADPSTLGTQRGSPTTDQRPASTDRAPATPVHEATAPAQDALAVQDAAA TVQRSPRLDHGCATADQQQDPGRDQDPGRDRAPGRDRHSRGPDRDQWQ ARDRDHGRQGRGLGPGRLDNDLRLTPATVRR LACDAAILPAVLGSQGQV

		LDLGRQRRFLTGPLRRALVLRDRGCAFPGCDRPPRWCDGHHIQHWADGG ATALHNAVLLCRHHHRVIHQDDWTVRIAADGHPEFVPPAWIDPEQRPRRN RYHRR*
5	EpiCas005	MVVIECTVPDCEFKTTDASEAVAIALLANHNLVHRNEAQPRRLPANTAPK GPKLDRPKIDVGVTMEEWNIFTRRWEVFRAGSGIDENSAPSQLFQCAGTEL GDSILKANPNAVNESLTQLMEAMRSLAIIPIATCVLRTELLQMRQERDESR SFTAKVRGKAETCAYNAKCSCSRPVDYTDHVIRDVIINGLYDNEIRREVLGI DGILEKPINQVIAVVESKEMARNALPPSPSLSAVSSFQRQKTNSPHAPSPPO ADQAKEATCPGCQKTFKLFTEGARGWNKKPHQLCIDCHRVNRRARRPQR QPQQRSPQAPT VHANQAEP I AQISSLHGETKQDTRPTTHATTAKRPVRLDH HIFSKGEWKRARFRDHPRVPITISLARKGRHRKETADTPDIHAAVSAIADT GAQSDLWSLADFLACGFSREELQPIRPSLTAANNSPISIEGAFFAKLSATSYS GQSTTSFSMVYVSSSVRDMYLSYDSSLNLDLLPHCFPRHDDVTEASNERRP ASQNKTRPNYTSPTTNAARIVNEGCTKPASEHDVTCSCPQRETTPPRPSELP FPCTPANNTKMKAWLLNRYASSTFNTCPHRALPCMEGPPVEIHLDP TAKPT ACHTPANIPLHWQKRVYEDLLRDEALGVVERVPYGEPTWCHRMVVTRK HDGSPRRTVDLSPLNKYCQRETYATESPFQLARRIPNDTWKTVTDAWNGY HSVPLRSSDRHLTTFITPFGRWRYTRAPQGFLSSGDRYNRRFDAILAEFERK ERCVDDTVHYDHDLEQHWWR TIEFLTRVGQAGIVLNPDKFQFAGRSVDFA GFHVSDHTIEPLPKYLDGIREFPTPTSTTDIRSWFGLVNQVSNYAQLRDLMA PFKPFLSPRHKFLWSSELEEFQASKEKIVKAIREGVEIFDIHKRTCLRPDWS KRGIRYFLLQQHCSCSSGMPDCCPGGWGITLAGSRFLSPAEQRYAAIEGEA LAVAWGLEQTRYFTQGCDNLVIITDHKPLVKIFGDRTLDEISNSRLFRLKQR TLPWRFEIHHLPGKSNYAADATSRHPSQSNTEEKAPLCPANPPDAAEIALM TAIRDTAQELGAISWTLLAQETAADARMGNLLHSLEQGTQLNKDDPTLAN LGPIYESIYQEEGVLMYQDRVVIPPSLRQRVLQSLHAAHQGTSSMEQRARSI VYWPGMSNDIRVTREKCHDCNRNAPSQARTPPLPTAPPSTPFEAVFADFFN YKGYHYLVVGDRLSGWVEVLSSPATTNLAGSTGLIRHLRTFFATFGVPEEL SSDGGPEFTAKATEDFLRTWDVKQRLSSVSFPQSNGRAEVAVKTAKRLLLS NTGPTGSLDQDRFLRAILQLRNTPDPCSLSPAQIIFGRPLRDTLAFVNRLEK

		FSNPHIRPLWRQAWAAKEDALRTRISRTTESLREHARPLRPLAIGTRVFLQN QQGPNPNKWDRSGVVVEAAGHDQYRVKVDGSGRVTLNRRLRAYTPA LPAIPQAPSGAPAKTPSMAPDAPSSPSQTCHHTNRTATPATTPGVPTPDLAT AALALQVSPEEPDAAEEKQGPSTPEAMTLPSPPNAAARPRRTIRRPKQFEPE TGTWTE*
6	EpiCas006	MASIASLAPDGWLERKELSDGVSDDGLSDDDESLLERDTLETIRDSLATNP DEIDGLLGRPSLGSDGLSPRTPSRRREFLGERKGSKTRLSPTPPAAVPAPTC APASTKNLWRGFALFAMAMACAALSLQRSPELVEYGPSRRLQTSFGERVD EAADGATSRRLQNSNQFCAARDGWEAFVGTARAGEAFGGEVCLLASTSTE SSTLYGSHAAYSPPMRVRAEFLAADDLTRQLGLVFGVPAWPDAANLPATG YACVLEHKVTRWSSKSRLVLRANADYDKNPDWINFKRSRWFDFSEPNAW YEATAAIDDTSAIAGKVGADVSCVRRDDGGAVVASVRKEDRWVMAHP GKVGIFYSERDDLGERFVRALEVDERAPEEPSFCCYWSPTSDMCGDCPDP QTDGWCAASRARCEGCGGTFCDAPTPPKEVPGRCCYDGAGCGSCSASDD DGFCAKSESNCGSCGGFFCSLAPTPAPTRVPSRPTTPSPTAAPTVAUGHPT AAPATYAPSSSEPSSKPTTSTPSAQQGCACEEGPYEYDGKTFEGCTSYGWS VPWCGTTNCGICDGLVSTGCWDECEETYAPSSSEPSPKPTTSTPSAAPQGFP AIPLPTSATSPPTPTVFLTNADFRYGTYLITNPGVYVLREDIVFAPETPKMMP PLDSEYPPDRGFWLGFFAAISVAADDVVVDLNGHEIRSSREFLLRQRFFSII ELADKPKFANSKPPQFGALDAPLVPARRVHVKSGTLGLSSHMGHGNENE DVTVTDVVIRDFETGGVQLNGASRVRLERVTVGPSLGHEHSAGVPALAT LSQATNLLRVAEGTRHERQEEMVLLSAAVERFVQETLAGEDTSASVFASE AAGYAGLPDGSATVWIPTTGLGGPHQTSELSSSVKSTSIRLIFGRIDCSRRVL EKQPKSLCRNRRIREN*
7	EpiCas007	MAGNHFKGSDSGRPAVPPRPNGAGKAGTPGGAGQGGSGKRVSPGETAMF TALYEQSQKAKKTGRARGNVFAGGATFASQPSGAPSVANNAGAANAAN GAGNVNAGKADNNTPSAGGAGLTGNLGTIYTGASETGTFARLDDSGAEFA GSGSKEDYYDFGLNYGSDASSSSTSDGLVRHRHRKGERKKRHTGAIIAAV VAVLLVALGASGFMLLNSAKTVKSEAKEAVEIVGGLKDKVTSGDFSTLPD DAKKIDELCDSMKAETSSPLWAAASFIPVYGSDINAARTMIDALSDVSSNA

		LVPMA DNLSQATPGKLFQDGMINV SALQAVADSLSSSSKVFKSANEKVQG IGDTHISQVTELVDKVKDGFATLNGAVDAAEKVAPVLPQMLGANGQTRN YLMYAMNNVEIRACGGFGGSQGLISVTDGQMSIGEFVPRIGLSEDEAVESV DEEDEALFGNHSNLYNSGNTYSPDWPRNSQRVAALWKSQYQQDVGIVG IDPVFLQYLLGLVGNVSLPDGTVVDGTNAAKVLMHDVYWNYPVEESDGIF ASVASAAFDKILGGIGDVDVTKLVSAVERGAEEGRLIAWMRNDDEQNAIK ETGIDASLPDPDDPSADPVAGVYFNNLSFSKLDWYLNADTQIGQGKNGDG TCSYRITVTLTNIMTQEEAGKLPDYVAASAPDAARDDERLNVSLFAPTGGN ISDLTVEGTQFGLGAATWHGIPFYSGTVDLHAGETTTITYTLTSAEAGDKP LTLRQTPTCQAARDSASA*
8	EpiCas008	MEAGRDFIAFGQDVQGTDCCRGAGGAGLAGPAGRSGAGAKNVRPPCQTV GMEAGRDFIPFGQVVEGTDRDCCSGAGGAGLADAGAAGVSGVGGVLAG VRQLCAGVVEDSALWDLGAASDFAGGVVEISRAVEYLQLVAASAVDRAR TQSAAAAAPGMSWTTGWRDEDPASDENS GTTAPVPVEPPSAAATGSDDG YRNTVEFLRARLRIGAAEARGRLALGAGVLP RQGLAGRSEAAVHAELGAA VAAGAVGSRAAAVIEAALGRVRHVGDAAAAARMEHALTRTAAENDTDF LARVARSWAEAMDQDGAEPSEELLRQRQGAFIRRQRNGLHHLEIFATAEQ FEQLVTVMNTATNPRTGTGSGTGAGHGAMNAPAADTPEDAGHGSEPATD GPAGPAAVQEPAQDWAPAATGGAGWETGPAAGAGWIVGPAAVSASDTR AGVVSTSAAGATGDEFVSDPGPGYGFPGVPDPGPGDHRAEAGDTGAGGL DRRTRAQKQLDGLVGACQISLAAGTLPASGGLRPQVMVTIDYRDLLARLG NGTGTTSDSAAPAGTATDCPVHMAGTVDVAGTVDMAGTVDMAGTVGAA GSGRRTGGLPGTGSLIFTGPVTDSTIRKIACDADIIPVLLGTADRVLDIGRAS RIFPPHIRKALIARDQGCAPGCTIPASWCEAHHIDYWSRGGSTGTENGTL CSHHHHLIHKHWTIRVRAGIPWFIPPPHIDPAQTPRRNNYFHPARLDPAA*
9	EpiCas009	MPNRKVTLTAFSRMDHRVEALEEGLNEVRASIAGVQATMRESYANIVALL EKHLGKSIATEGGTSEKKDKPESSEVKNNVTLEFRQSAKKVELPSFDGDDP AGWISRAEVYFRVQGTTPAVKVSLAQLCMDGLTIHFFNSILRENEDLSWEE LKEALLERYGGHGEQDVYEQLTDLKQEAIDDYITEFEYLVAEIPRLPEKQY RGYFLHGLKPEIKGRVRSFAALSEMRRTKLLQVARVVEREVEGGSELKPN

		RGPRFGFGSGSHKNGSQSDWVMVRRREGSRNEG VKSNTDGP KGDRKRIGP RDRGYTQLTYNELMDYKQKGLCFKCKNPFGPMHRCPAKELRVLIVDDVE NEEDEERILAVEVGEEDEK GEMSILNLHHITHETHHTMKFQGTIRGVEVLI LVDSGAIHNFISQKLVHQMDWPIEPTAEMNVKLGNGFQISTQGVCRGLEM CIGDLKMNP EMHLFELGGIDVVLGIDWLKTLGDTITNWKQQTMSFWKDKS WVTLKGHEGCRRQTV ALQSILNRPKPTSQGLLWEMDNVKHKGKGVNELS IQQQAEMEVLFNKFEHVFQTP TGLPPKRN RDHAITLTDGHGPVNV RPYPYP HHHKNEIEKQVKEMLTAGIIRHSTSSFS SPVILVKKKDGTWRMCIDYRALN KATIPDKFPISVIEELLDELHGARYFSKLDL KSGYHQVRVKDS DVGKTA FRT HEGHYEFLVMPFGLTNAPATFQGLMNDIFRPLL RKNVLVFFDNILVCSKD WETHLHHLEEVLRLLAAHSLVANQKKCYFGQRSIEYLGHMITEEGVAVDS RKVDSVLSWPQPKNVKRVRGFLGLTGYRKFIRDY GKI AKPLTELTKKDG FKWGV EAQKAFDELKKRLTTAPVLTLPNFSKEFVIECDASGSGIGAILMQD KKPIAYFSKALGVRNLTKSAYEKELMVVVLAIQHWRPYLLGRKFVVSTNQ KSLKQLLQ*
10	EpiCas010	MIKDFASLPPSHQHKL SLQCQLSTIDDH QTHVIPVLIDSGCTDSIIDEAFVRQ HNISTKPLPTRVIVSNADGTKNRTGPLTEYVTLRLTVAGRTELMDFLVTGQ REARILLGHDWLRRTNPINWQTN TITYPSDDTEYHPEDGDRVIAVDLQSR IDKISVESNV SQTKSQQIAEESERAKSTLTFRDVVPPFYHHFEPVFTKASFDK LPPRRVWDHAIELLPEAKGLDCKTYPLNRDEQIHLDEFLKENLETGRIRVSK SPYAS PFFFIKKKDGSLRPVQDYRKLNALT IKNRYTLPLISELINRLKGARFF TKLDVRWGYN NVRIKEGDEHKA AFRTNRGLFEPLVMFFGLTNSPATFQTM MDRLLRDLIN EGSVIVYLD DIMIFTDSLEEHRKIVSKVLQILADNQLYLKPE KCEFEKKEDEYLGMISHDTIRMDPLKVQAVAEWPIPRTKKELQSFLGFSNF YRRFIRDY GKI AKPLTILTGKVDWEWAEDQQRAFDVLKKTITTSPTLAIPND EDPFKVECDASNYATGAVLSQQQEEKWRPVAFLSKSMNPAERNYAIWDK EMLAVMHALEEW RHLLMGAKHQFTIFTDHQNLTYFRDPQKVNRRQARW MDTLARYDFKLEHVKG SANTKADLLSRRVDHPRGERDNEGQVLLPQLAE ANIYTMEDVLEEHYLDELLASQHKLDDYVVRKLKEDPKVWKKRDGIVWI GNEKEDELTLGSFENKIYVPQDPKLRARIIEEHDTSLSGHPGERKTHEGVN

		RSFWWPSLRRDIRKYVAGCEICQRTKSHRQKPHAPLHPNPIDRPWQIITTD LIGALPESNGYDAIQVFVDRFSKMAHFVPTHVTLDLLGYAKQLRDSVIRLH GIPEKIISDRGPQFMSNFTKDFYQTLGIEPGLSTGYHPQTDGQSERINQDLEH YLRTFINHRQDDWADWLAIAEFTYNNHLHSATNQTPFFVNYGSHPRKGTG QKPSGRSQAEDFATSMTKLHEQTRQSLLLANQNLLKKYYDRKRTEAREY HKGDKVWLEGTHISTDRPMKKLDDKRFGPFTVKAKVGPSAYLLDVPRSW KAIHPVFNEILLSPYHTPEFPNQPRTSKPPPIQVDGAVEYEAKEIIASRPTKLK GAKLDYFVHWKGYPTERTWEPDHHLKNPSLITTFHRKHPAAPRVIASAA LRFRPYENFTTTTEKPRLFDWITNRFTDCGAKILGTRI*
11	EpiCas011	MSITDEYFQLTTKYRAQYGERTIVLLQVGAF FEVYGLKNPRTKVVTGSQIA EFAQICQLSVSEKNVCVEQDHVVMAGFRDYTEKYLQKLSENDYTAVVY TQTKEGKNVTRQLHAIYSAGTYISYDNDSSQQITNHIMCIWIEKYQALDKS ARLSCGVATANIFTGKSEIFEYQTAYYMNPTTFDELERCVSTYCPSEVLVLS TTLPQDVVQTIVQYVGITHPSRHIFCASNM TDAVKNCTQQKYIHHMLTTFF GEESQQLCAEFSNNEIATQAFCYLLHFVQEHNPD LVRKIEIPRFSSQSKTML LANHTLRQLNIIDDL SASGKQAGQYSSVLSFLNKCASPMGRRLFQQQLIAP TFDQEWLQKEYNMTEMLQPDTEPMIPMLRKQLATIRDIEKICRQIVLHKI YPDSIAHLYDSLQTIQQIHVCFAELPKVLEYLHHSTEKIDQVLHTMQTYLLI DRCRGIQSVQSFPHSIIQPGVSSSLDKIIQTYEENLQHFREIHAF FNKIMRAKE GGDAEYIKIHTTDKSGVSLQITKKRATVLKSLKGLADPKIQLSGGLYFCVE DIKFSSASTSADEIEIPCLKNICKALLTSEESMNQEIAKSFQDFLTTFENLHLE NLEKCALFVAKLDVLHCKAYLASNYNYCKPTIVADHQKSFVDARDLRHV LIEHIQINETYVPNDFCLNLPDQPLGVLLYGTNAV GKTSLIRALGIAVIM AQ AGLYVPCSAFTYQPYQSIFSRILGNDNLFKGLSTFAVEMSELRMILRMADES SLILGDELCSGTETESALSIFTAGLLHLHAKGASFVFATHFHEILKFEEMQQ MPKLALKHMAVHYDPQRDCLVYDRKLQDGP GNRMYGLEVCKSLYLPDD FLQSALAIRNKYYPTSRGELTHHTSTYNTQKIRGICEMCKTNMGEEIHLQP QKEANDKGFIGTIHKNHPANLMTLCEACHLKQHKSSITQKRKKT TAGYAV L*
12	EpiCas012	MEIFISDNTLNMSKKEDVEPVETNKIMKEYLDLTKQYTKIYGEKTIVLLQV

		GAFFELYGLRDDKGTTILESQIEDICIICQLNYADKKLMYNSKQVIMAGFREY MLDKNIPKIIIEAGYTAVVYVQEKDGKNIKRVLHSVCSQSTYLSCETDSSPV MTNNIMCIWLDTFKPMDRNVSTVSKTRDTIVYGVSTINIFTGKSSMFEYQT PFILNPTTFDELERYVSVFSPSEILFVSSLEKEINRTILQYIGVKCPSIHIYDSR DMENKTIACSKQVYVQHMLSTFFGEEAMNVCREFQMNTMATQAFCYLL NYIQEHDKNIVRKISIPEFNNTSDRLLLANHTLKQLNIIDDSSIDGKKTGNLS SVLSLLNKCCSPIGKRNFKMMLLPSTNETWLQQEYDMIEYMLQSENIYFI DMFRKQLSQIRDIEKICRQLLVLKVYPSSIYQLYKTIGVIHNLHICLAENKKI TDYLCKDISKEQINSYHYINDLCLQINEFLENHFVLD DCKNIQSLTSFENNII KKGISQPLDDLIKHHYEKIAIFEFVKKQLNEFMYSNGQNRDIEFIKTHETESK GLFLQITKTRANILKNYIKAKLEKDADYSFHNKELQFSFYLKDIKFGVASSA NEEIHPILOKICKELITIKDEIDVVIRETYEILKKIEDKWLPSEQLANYVA NVDILQCKTYIAKNYKYCKPEIVKSEKSFVSAQELRHCLIEHIQQNEIYVAN DICLGNEKSGMLLYGTNAVVGKTSFIRALGISIIMASGMYVPCSRFQFKPYT AIFSRILGNDNIFKGLSTFAVEMSELRIILKMADKRSILGDELCSGTETESAL SIFVAGLMDLHDKDASFIFATHFHEIINYDEIKMMNKLSLNHMSVIYDREK DALVYDRKLDGPGTKTYGLEVCKSLYLTDEFLDKAYAIRNKYFPETRGE LNFKSSTYNVNKIRSICEVCKNEIAEETHHLEEQHTADTDGFIGHFHKNHKA NLVSICKNCHDKIHHDNSGKNTLNGKQRKKTTKGYDFM*
13	EpiCas013	MLLTEEYLEEQKKLELKYGKNVVLFMELGSFYEIYEANGIGKVKEISNILNI LQTKKNKNKPESQRNPLMCGVPSVSLDKHLNRLAQEEKWTIAIMSQTATD NGFERKITEIISPGTNVRNNIKDDTNSYLA SLIDKNTKGIYSCGLSLIDVTTG NIKVYEGYGVHNDKEKILDEVNLILENYNIKELILLDNLDSNKLYYKFPSIN MKTSNKSIHYKIQLTITETFNKDFSELGLEQSPYIIDS LVGLLEFLIDHNKSLVV KLNSITFINNKEFMYLGNNPLKQLNIIGDDFSVEKLINKGISAIGRRYIREQIL NPIINADEINNRYKYSLSLSKSNKLPTIEKSLKNIYDIERLSRKLELKTIKPFEL GNLVTSLYASSDIKNIKMLD TLTFESYINLDLIKDYNSIHWNFFKDEQLM TLKTELELIMDLSNYEFDSFIKYSKDGYYYEFTNKKYENLSFTSTIEHTRK LKNSIKLFTSELSSINTRRLVIESKMMEISNVLFKDVLDIFDITIINIEINKIQRI EFYLNNTKLLAKKYCIPTILDDVENIFDVTLNRHPLVEHLEQEVFVPNSIKF

		NKENKGYLLYGQNSSGKTVLSKSGIAIILAQSGFFVPAESLAFTIKDSLFTRI IGADNINKGLSTFAVEMLELKNILNRSTPDSMIIGDEISHGTETTSGVSIVAS AILHFNAPFIIATHLHQLDSEINDIKEKPFIFHILALTFDSTNSLIYNRTIQKG KGSSYGLEFAKSLDMPDFLDKAYAIRKEIADDISGLEQLTISKPSSYNSKL FLGTCYLCDNKGEEVHHIKHQHLADDDGFIDTLRKNHKSNNLMVLCSSCHN KQHEKKI*
14	EpiCas014	MSAKKGFSMIDEYFNIYNEKIKEYGKDIAVLIECGSFMEIYQIDNNEEKIGN AYEISKILNICYANKNGDLSNNNRSYPNFVGFTTNSMSKYLPILENNYTVI LVNQLETNGNKKGKLIQRGVTAVYSKCLQPLEMDTNADNQLLYISLDVKL IHSNVLKRNKTTMYELNTSICCVNNFTNSIEIYENNYTFTKDCIENCLNDFD HLLYKFFSKDINIHLVGENIINFESSELLKFFNSRYDNMRFHNFDKQKHKEYI KTDYQNEYFKKVYNHINFGLLTPIEYLNITTMPSLSIVNLMYTIDFGRHDL KYIRNLSIPKMMNDSHFLALDTLNQLNIDKGVFDIVNCTKTAIGKRYLK SLLCKPLKNPNEIEYRYKLTDELESIINTNNEIDQLLSCHDFEKLHRKMSLEV LHPFEFEKLHTCYKNVNKLINVLSNYPLFHQIIPNNVKLVNFNEYINSYMET FDFDKLKKYNLNTSKDEISNYFLKGVIVELDEIQLKIDNITEEIERMRVKYN TICENNCKTKNPDIYIKIAYTETEGYSFSCTKIKYQTLINAIKESSNFKIKQTN NIVKFFTPELVDLSNKLTLFKDLLKNKINLHYISTLKLYYYNYHDLFTSMTT FIQILDVVNSNLKCKNKYNYCRPIDDNENDSSFFKAVGLRHPHIEVITQSNK YIPNDIVLDDNNIGILLYGLNSSGKSSLLRAIGIAVILAQCGLFVPCNSFIYKP FHTVISQVDLSDNMYIGKSSFITEMAGLKKILDLSGKNTLVLADELCRGTET YSAISIVCSTMIKLIQSNTKFFFTTHLHQIPQLKRINTLKSIDICHLSVTIKNSNI IFNRKIEKTSGDPLYGLEVCKTIINDNDFINDAFEIRNELTRNNTQILSHKKS NYNKKKITSYCEICNSCNSLETHHITEQYLADKNGIMPDGNHKNLNLVLT LCHDCHLKVTLGKIKTRGYIQTINGRILDYNTFTEDTTEKV*
15	EpiCas015	MTTEHKKYEYLFKNKSLNEEYLLLQEKYEKIYNKDFTVVLMMQVGSFHEIY ATNDRGYDLHKLSNINLNIIVSKKNKNPEVSEKNPYMIGFPILATPKFINILV DNGFHVLIKIDQVTPPPNPKRAITQIYSPGSYIEGISSYDSNYILSIYIEEIKQMN NSYILFAGLSIIDLSVGKCIHETFTSKDDEKYSLDEVLFYFNPNKEIIVYY KNLITYSENELLYLELNDKKYLISEFSDNEYTKIYQNNFLKEIYNLETCTMT

		HIEELDLERINNGRLSFILLNLYCKKMNSTILNNIYKPEFYLNNTKYLHYGNN ALKQLNVQTYNSNEKSLFDVINYTSTSMGKRYLNFNICNPISDINILNDRYD KISELKEKNIDLSNKLKKILDIERLHRKISLQNLNPHNFVDLHISYSTILSLYE DIKKTCLNKIYNEKVKMNLNENYISYYSRFLNDEMKEYNLNDITNSFYNN INEKIDILQKSIKKEYKLIENLRDELEKFIDCKKKDYFNSFNEDKSPIQINYNE KDKYHFSTTSKRGSIKNNIKNNSKYNYSYKFKDQSNMTKITCTYVDNIS DKLVILNDKIRPLIKNTYLNLDKEMYEKYNDLYLKINNIIAEIDFINSGSICSI RNKYCKPKIILKDKSFMNVKSIRHPIEQICSNKYVPHNLNLDDNQIGILLYG LNSAGKSSLMKSIGLNVILAQIGYFVSASEFVFSKYSLFTRISNIDNIYKGLS SFALELVELNAILKRSGSNTLVLSDEVCKGTEYNSALIIVSAMIKMLVESNT SFISATHLHELTKLDAIKCISKIGIYHISVKYINNIIIFNRLLKKGNGTEEYGL DFAKYMINKNKFIDICNNMCKDINNITANKSKYNNKLIVDKCQICKSIKLE THHIEFQKNTDEYGFIIKKTNLNHIHKDHISNLVILCSECHDDIHNKIMKGY EDTIKGSILKYKRSKCEEEKYNDDEIKFIINKPNISQKKVKDLFEKQFKKNI TIKNISKIINNKY*
16	EpiCas016	MYNIIYLNINIVMTLIEEYFEFQDKYTKIYGEDTIVLMEVGSFFEIYQVGDLG KAKEIASVLNMILTKKNKNINEVTFKNPLMTGFPTVSLEKNVSRINAQSKY TVVLVKQRGQSPNITREVDEILSPGIQYRINTFKNNFLANIYVEYYRKNMFG IGISVVDITTGENYLMDSVSLEGFPDNPLSEIQQMVTKYGPSEVIVSMVDFE SAESEATRIRVLLDEMIGKYSWKSCDSANYKVSFLNEMFSEIYKVKSIPTI EYLGLEHNQLAGNALIDSIGFIQKNNPCILDKMPRPEYINCAKYLTNLNDAL SQLNVVDNNKKANRKGSLFELLNTTVTAAGGRLYGRRLSNPISSKEDLNAI YAEIDLFISDSKFAEVRREIAPILDINRIWRRLVMRKGDPMDIWNLYTSIKIL ADMKNDFSWIPNLENLSNSAQSAKNYIESIFDISRIEDSYLSNLEDNIFKKEN PILDKLEAELAEIKYKSQIIAEMFANHSFFTEKTKLVYTDKEGFYIETTPSRY KQYLGLNTAPKSDILAGFDVYKNKSKVRLTNDDLRLKLSKVIKIQKMSILI KEMFVKEVDEMYSAGDDVESISAFISHIDVMAANAYNSVNLSLFKPELID NEEGYSSVECHALRHPIVERLTHSEYIQNDVLFSSNNKYGKVIFGVNSTGKST LLKATGLAIIMAAQAGMHVPSKMKLSPYQSLYTRITSGDDLFSNKSTFTVEM LELKQILEKADHNSMVLADELSHGTTETNSGVSILAAATINTLSKRKATFLFTT

		HLHQITGVRSIMDNKSIDLSHLNVNINPITGVIEYDRKLKDGAGESVYGILV AKGLGLADDFINDATSILNEVKSNRTQLDAITENNTNKYNSNKNVNKNCEIC GFQSSETHHIKEQADFKNQKYLEIQKNAKYNLVNVCNDCHKKIHS GKIVN LMYKQTTEGVELKIT*
17	EpiCas017	MPSLIKEYFKIQEESEKKFGDKTVIVMHVGGFYEVYSYEGLGKADVANE CNMMLTSKDKSKEISITNPQMCGFPLASLLRYMYFLTSGGFTCVTVKQDEK VKEKRTIDRIFSPGTCVDDSESSIVASIVADDGCIASVWIDLTSGKVMFEEIY NDNEAYRREELTRIIETCGCKEIIICMSKNGIMYDDAHEIDFCNIYEHISHQND VISRAYPNDSTVSPIEFIDMEMFHIARVSLILLIEWVEDHHQSLIQRIRIPERIN RKSLALHNTSVHQLQLVSQDKKSLFDVINKTNTSMGKMMLLKQTLLNPYSD VSKLNGMYEDIERISKVDVEVVKKVLKNIGNVDRLARRIANKSINLEDIRH LILCFENGLDIIEHMRELSVHNMISENDVKRCYIAWHSTIDMEKDV FREGV YDDVDILKKERDTIMYMM DANAKRFSKTLGVNEG VKFELHKGTYGLITTP GRAKTLKNRVGNVVFQADCKQKVISTPDIQSLIHKLHGAIDAYDECME EKMKEFVEILYSYHEIICTFGTYIAQVDVLLSHYLVAKAFNYTKPVVKEGD KSYITGKNIRHAIVERLNGNTLYMGNDVDLSNEKGLLLYGVNGSGKSCFS KSVGICVIMAQIGMFVPAETFEFCPYERIFTRISSEDNIYKGQSSFYVEMME MNTILRFANEKSLIIGDEICKGTENISAI SIVASMCNMISKIGATFVFATHLHG LVDLEPIKQIIDTIQIKHIGVTIESGVVTYDRHIKNGEGEHLYGLEIATHILGN KEFASLALETRNALLKKTTPPKSKYNARVLKKVCHICGRKDEQLDTHHIN YQCDAHENIKNTKGNLVVLCKKCHDDEHHGFLSIKGWKDTSKGVILDYEF VSEK*
18	EpiCas018	MPQLLVDEYIELTNEYRKKYGEKTIIWMQTGSFYEVYAWNEEDEQIKVSQ NILQIRITTKKSGNNRSCLMAGFPDHSYKRFEKRLLENY TIVYVDQIKKDP IERGVTKISSPGCVFDSDDDSKESILMSVLIEIEDNDFYGYTYIYDSNIGEIRM LSIKNEENQ TIDHIFDNIGDIISENRVNEVLLNIVSEQYIKLPETNSKILFHYN KIDKIDELLKPLLQNDILSKYFKKYLSIYENVIDSLELSKCLPQEIVNLIYML DFLKEHEPIFVKNLQRPIVDGPNQIKYMEFYNHALQKLQVISNEGNDLIKFI DKTLTYGGSKKLKYLLKNPSNNIEELKKRYKSVDL FVN NRQLLNETKTYL KICDLHRFYRRFAIGKINVYNDIPRIVNINENINILINIFKNVDNSIYWLPNDK

		IIQNFKDYCGDLIYTFCIDNCKSSGNVFNKDIVPELDDLKSESLNQELESIRTYLCNLINEPIYLKNTDKDGFFETTKKRGTKIAKQLKENNKFDLEISNSTSVSKISSAGVKKISVQLCDLQTKIIDLTQKKVSEKVLFFYYDTYFESSIATIIESMAWLDVYYSSYSKSAIEWNYVKPNLVECENSYIDAKQLRHPIIEQLLSIEKQAFIPNDIRIGFKESYLLYGVNSVGKSSLLKSVGLSVIMAQAGMFVPASDYKLCPYDKLIVRIGNTDNLFEAHSSFICEIREANQCVSNTNNRSLVLADEFCASTERESATQIVTSMLLWLNEKKSSFIFATHLFELIEAPIVNVKICHLKVRIESDSL VFDRKLTSGPPTERNYGILVAKKIFKNNKFLKMMDRQCQTKTSTRTDINRSRYNRGVLLSECAICAYSPIDEMSLPLDTHHINMQCMANNDGFIGNYHKDIVSNLVVLCKKCHIMAHNGKLNKICWEQRDCGNVLNIEIN*
19	EpiCas019	MLHPRLQEMLAARGIRRLYTHQAQAANLALGGAHVVVVPTASGKTLCYQLPVLDRLRLRRPQSRALYLFTKALAQDQHHGLTQDLLEAGTDAAAMIYDGDTPQAQRRRAARRIAQVLITNPDMLHVGILPYHTQWRELLAELDYVVIDE MHAYRGVFGSHVAHVLRRLQRLCAFYGSAPQFILT SATIANPAELATNLIG EPVEVVAESGAPSGERHVALVNPPVIEARLGIRRSPLVDARQLALS LAAQGEQVVCFTSRSLAVEQLVIALRERAGRVGLAPEAIRGYRAGYLAAERREIEA GLREGRVRCVVATSALELGIDIGGLTVAVLVGYPGSVASTWQRIGRVGRG QAPSA AFLIASASPLDQYLM AHPEHLFGGAEAAHINPDNLYVRS GHEQALYELPMAEGEAAQDDDRRAILAALEDAGEARLSRGRFWWSGQGFPAGEI SLRSAEAAARVSIVARRAEGRDITIGEVERSSAPRWLHEEAIYLHEGQPYLV RTLDLEGGVAQVEPATTDYFTVASEREEIEIERVVASTDGERMTTAHGEIRL TRQVTGYRRQRFETQEV LGWGVVDLPEQVILTS AFWLGVKPELVARLDAQ GLWTADRGGARGPDWEQQRD RARLRDGYRCAHCGAPERPERAHDVHHI VPFREFGWEPGRNDAYRYANRLANLVTLCPRCHQLAERQVALGSTLSDM GHVLRYLAPLWLLCDPEDIGIHADSAGRRAERPTLFIYDQLPGGAGLS DQLPRLIVPLLKSAAQVVS DCP CETGCPACIGPVLAPAIGRKRRAVGLLEALAW*
20	EpiCas020	MSWNEEITPLAALLQRWRQDPAFRRRVAAWARAPAREAQWAPIPEDLHPRWREALRQAGFSMLYTHQVEAYRAARSGAHVALFTGTASGKTAFLLPILQALCEDPQT TALFIYPTKALAH DQLAVLRAWLSALRLPAAALPYDGDTPGEARRFVRERGRLLL TNPDMLHLGILPQHPRWAAFLSHLRFVVIDEAHVYRG

		VFGSHMVNVLRRRLRLAAAYGASPQFLLASATVANGPELAEGLVEGSVRII EEDGAPYGERHFILYRPPLDDAATGLRQSALTAAGLIWTLLEHGIPTIAFA RSRMAVERLLRMVRSAWAERGGDPEEVHGYRGGYLPEERRAIEKGLREG RIRAVIATSALELGM DIGTLEACVMVGYPGTIASVRQQAGRAGRRAPLSVA IMVLTATPLDEYLAGHPEFLFRNPPEAARFDPDNLAILADHLRCAA FERPLD PGE PFARTLDTGWIVERLIADGEPLYRNPTDGRVYFVADRYPAQGVSLRTS TGDRVKIRLGAGKGGRIIGEVD RPSAPRLVHPGAVYLHEGAAYRIVSVDW EQGEAMAEPFDGDEITEPIVEARWYLDAERARRTEPWGIGAWGEVTIVHQ VVG YRRLRWPGGELLGWA EVETPEEILSTTAWWLTIELGLVERLKEEGIQI GPLDYGPEWEAIRRQVRARDGYRCRVCGAPEPPGQRHDVHHLRPLRTFRD PAEAYALENLITVCRRCHRRLEQMLGTRSALSGLAHL LHGLAPIFLMSDPA DLGLLVD PESREGGGPTVLLYEGVPGGVGFAAHLWAVQEALLEANLEV V RRCPCERGCPACVGPPGEMPD LKGETRRLLECIFAAD EAAAASKVQ*
21	EpiCas021	MADLERLLERLRDPTFGPNVTAWRRREGRPARYGETPAWLDPRLQEALR AGGIERLYTHQVAALQAVRQGEHTVVASGTASGKSLCHHLPVLQALLEDP SDCALYLFPTKALAH DQEAVLRALVERIPDL PDSAKPFPASYDGDTPGQAR RAIRERV RIVLSNP DMLHAGILPSHPHWAPFLHHLRYVILDELHTYRGLFGS HIANLLRRLRRVCAFYGSRPQFLCASATIANPGDLA GRLIGEPATVIEEDGA PRGPMHLLLYNPPLD PALGIRRSALQEARRLAVRLIGAGLQTVVFGRSRR GVELLTYLRQDLPPGLR PALRGYRGGYLPALRREIEAGLRSGEVRGVVAT NALELGV DIGTLAACIMAGYPGSIAS TWQQAGRAGR RDQPGLAVLVAGA NPLDQYLVTHPDYFFGRSPEQARINPDNPFILAAHLRCSAFELAE DGAPW GTPAPPDLLEDLVREGSLLHEGG RYYWIGVGSPAREVSLRTASGDTVTHIEE SSDGIGTIGQVDRPGAAMLVHEQAVYIHEGTFYRVERLDWEAGLAYVRPA QVDYYTQAHLRVDWRVLSEEAQAPQGRGSRHTGEVEVHAQATAYRRVH LHTQESLGWGAISLPPQEFRTQAFWLTLPWPAQEHLRPDYGPNWEEQRRR ARERDRFTCRHCGAREEVAGREHDVHHLRPFRTFGYIPGENDAYLEANRL ENLVTLCRVCHARVDHGPRQETVAALQGLGHLLAHVAPLYLLCDRGDIRE TTRLRALPEGRPVVLIYDDCPGGAGLSPTLFSCADALLQTARERLQECPCA GGCPSCIGPPEEPARDKVIHILEFLLEEEAPPPGTA*

22	EpiCas022	MNKLWQHPNPDWVPVSHFGE LRYESIDVVMARWDWDKLIWGKVRRFVF NLQKRIYKATKSGRYRKAKSLMYLLQNSYYATLLAVRTVTTENFRGKCFC ARAKTFTSEGKRTAGVDRV KANTPKRKMALVKDMLDTIQRGWDKYRPM PAKRIYIPKANGKPKVSEAKLRPLGIPTIKDRAMQAVTKIALEPFYEAKFES CSYGFRPAMGCHDAIDRITAALLHKQKWVLDADIKGCFDNIDHKFLASQID THWRPIVKRWLKAGYMEKSDYNQTNMGTPQGGTISPLLANIALDQMETD LIEHLRGIKGWKKKIGVTTVSIVEKTEKPQKEKKTQKKPKKANNKGTKEPQ KKQKNVYEGAKKSYKCRRNLLIDL VRYADDFVVLHKDKEVIEESKRFIGK WLEVRGLTLSEEKTKIVHSTEGDFDLGHHIRHYENRIKGT YKLKLLKGTPTE QRRAKATHVLRVEPTKDKIKGHYRVIGDTIWKLKAATPSELIRILQPKISGW ANYYKTVHSSEAFGKLDHLLWKRLYQWAQRRHPNKGKRWVTNKYFATV NGRRWTF AEIVGKGKDKKIVKHLKGYSKHKERTGSYARVGYDRSYYDGD TAYWAERLSKGYGNITPSKAKMLKRQDGICPVCKCRFGNEDLPEAHHIHP RHLGGEDKYTNLVLVHLHCHDKIHRENPGIKGKLVSQTSPELQGFLNNGM APAFQAEGMKPPANTSTVLRTGKGAGIITPASDEMKS PVTSTVLRTGKGA GIITPAPDDNKET*
23	EpiCas023	MDTLQLSNPDGRVRSQHSLTGLPSDKKHRDHTDSSGRLVPSCYRVINSSD MCDWDWDSIIWGKIRRLVFNLQKRIYKLT KAGKYRKARS LMYLLQHSHY AMLLGVRKVTTENKGGKTPGVDKVKVTTPRQKMNLVVDVLDTIQMGWD KYEAKPAKRIYIPKANGKERPLGIPTVLDRAMQAVTKSALEPFYEAKFESA SYGFRPGMGCHDAIDKIASVLLTKQKWILDADIKGFFDNIDHKFIVEQTDK HWRPIVKKWLKAGYMEENKYNPTKN GTPQGGIISPLLANIALDQMETCLIE HLRGIEGWKTRIGNTKISIITDKKTQKKYESRKHLQLSVIRYADDFVVIHEE KEVIEESERFIKEWLGLRGLTLSEEKTKIGNETDGFDFLGHHIRHYENRIKGT YKAKLLNGTPTEQRKAVTPHVL RVEPTKAKVNAHYQELSDIIGKMKATTP SELIGILQPKISGFANYYRTVHSSKA FSRLDHLLWKRLYQWALRRHSNKGK RWVAAKYFCTVKGRKWTFAEIKGKGKDREIVKSLKGYSKHKVRTGSHAR VG YDRSYYDGD TAYWAGRLSKGYGNISPSKAKLLKRQDGLCPVCKCALT NEDLPEAHHIKPRQYGGLWEFKNMVLLHLHCHDKIHREYGEQKRKLCKG TREYDELVKKGVFVSPEVEAKYGPDI AKMRKGPEVNKGSSKILRLGSDPGII

		SRDKDTPVQETHSNGNNEDELIQVKSWSGEEKMIPRSKMYRKTkrVRD KITNKWKLVKMWFCK*
24	EpiCas024	MNILWQHNPSEVNVFARAQKQFTSEWIPVSHLDRHRYESIESVMAKWD WDKLVWGKVRRLVFNLQKRIYKATKAGRYSKARNLMYLLQNSYYATLL AVRTVTTENK GKRTAGVDRVKVNTPKRKMALVKDMLDTIQRGWDKYRP MPAKRIYIPKANGKLRPLGIPTIKDRAMQAVTKIALEPYEAKFESCSYGFR PAMGCHDAIDKIA GTLLNKQKWVLDADIKGCFDNIDHKFLASQLLPLQHW RPIVKRWLKAGYMAKSEATKKSEYYRTDMGTPQGGTISPLLANIALDQME TDLIEHLRSLKGWTKIGTTTIGIDINKTKKSYKSRKNLKIDLVR YADDFV VLHEEKEVIEESKRFIGKWLEIRGLTLSEEKTKIVHSTEGFDLGHHIRHYEN RIKGT YKLKLLNGTPTEQRRAKATHVLRVEPTKEKIKGHYQDISDTIWKAK AATPVQLIALLQPKISGWANY YKAVHSSEAFNKLDMLLWRRL YQWACRR HPHKGKRWVANKYFATVSGRKWTF AEIVGKGKDRKIVKY LKGYSKHKER TGSHARVGYDRSY YDGD TVYWAERLSKGYGNITPSKAKMLKRQDGICPV CKCRFGNEDLPEAH HHPRHLGGEDKYTNLVLVHLHCHDKIHRENPGIKG KLVGQTSSELQGFLNNGMAPAFQAEGMKPIGTSTVLRIGKDAGIITPVSDG NKET*
25	EpiCas025	MNILWQHNP EWIPVSHLGRYRYESIESVMAKWDWDKLIWGKVRRLVFN LQKRIYKATKSGRYSKARNLMYLLQNSYYATLLAVRTVTTENK GKRTAG VDRVKANTPKRKMALVKDVLD TIQRGWDKYRPM PAKRIYIPKANGKLRP LGIPTIKDRAMQAVTKIALEPFYEAKFESCSYGFRPAMGCHDAIEKIAAALL KKQKWVLDADIKGCFDNIDHKFLAVQIDTHWRPIVKRWLKAGYMEKSDY NQTNMGTPQGGTISPLLANIALDQMETN LIEHLRGIKGWKKKIGTTTVSIVV GGKKAKRPQKENEKSQKEPKKANNGTKESQDKPKKVYEGA EKS YKARK HLHIDLVR YADDFVVLHEDKEVIEESKRFIGKWLEVRGLTLSEEKTKIVHST DGFDFLGHHIRHYENRVKGT YKLKLLNGTPTEQRRAKATHVLRVEPTKEK VKLHYREISDTIWKLKAATPVQLIAILQPKISGWANY YKTVHSSRAFGKLD MLLWRRL YQWACRRHPNKGKRWVANKYFATVNGRKWTF AQIVGKGKD RKIVKY LKGYSKHKERTGSYAKVGYDRSY YDGD TAYWAERLSKGYGNIT PSKAKMLKRQDGICPVCKCRFGNEDLPEAH HHPRHLGGEDTYLNLVLVH

		LHCHDKIHRENKGIGKGVVSQTSPELQGVLNNGMAPAFQANGMKTPTLTD TVFRVGEGAGIITPALDEKKPNADTSTVLRTGKCAGIITPAPDDNKET*
26	EpiCas026	MKQNGIDEITIKSSGEINPKSGTNTRILYEHQKNAMECLDIMNKEKAYSTLV VLPTGGGKTYTASMWLLKNAIDRKKKILWIAHRQMLLEQAAESFQKFAY AEVIPHISSFHFRIVSGASSHDQTIHIRPTDQLLIVSKDSIGRNLQRLDAWMA GENEIYMIVDEAHHSTAKTYRKVISYIKEKVPNVKLIGLTATPFRTAEEEEQG LLAKIYTDGVQNGQAVHGDIGITYQIGLKELINRRILSKPVFESYYTDEFYQ KELGLKAWESIQHLDMIPDDLAKKMAESSARNKLIVETYQANQEKEYGQTI LFAVSVEHAIQLTAQFRKAGVAADYVVSSIKDMITGVTLSTREENEKKINAY RVGDLQVLINVNILTEGVDLPKTQTVFLTRPTVSTILMTQMIGRALRGTA GGTANAYIVSFIDAWDEHIAWVNPESLFVGNNDFQENAYEQKQRDLRMIA ISKIEEFARILDDSTDTSLLEKVPFTQRIPVGMYAFTYLEENGIEHAYQVM VYNSSQKAYENMMDSLPDLFQSFHAESEYLSEHQLMEMEKQCRDTFFCGE MVPPYEPKDVHILKYYAQYESVPKFYTFDEVDRSKLDVGMIAQHIWAED MGERKKAHEYIDKIWNDAADDNMLRLLFFGKKLYFLRQLDIELMKIRYPDIYEE TANVVYETRKLEELSLHDIGKVNPKLEMQLRDAAFEKAKTSDGKYQCAV CGMKNRSKTCFHVHDIIPMNQGGKSVPENLQILCRQCNGIKGDQI*
27	EpiCas027	MARDYIEVKSFGKVFPKKGANTRMPYEHQRDAMENLDIINKSNKFSTMIV LPTGGGKTYTASMWLLRNAIDKKKKILWIAHRQILLDQAAESFQKFAYSE VIPHIPFFQYRIVSGASCHDRVINIQAALDNLIIISKDSIGRNLNGLDVWLKDE KEIFIIVDEAHHSTAKTYRKVIEYVKSXVSNVKLIGLTATPFRTASDEQGLL AKIYEDGVENGKAIHGNIGITYQIGLKELINHRILSKPIFESYYTDEMYGESL GLNSWKSQIHLNLPEDVAEQMADSAARNKLIVNTYKAKQKEYGQTIVFA INMVHAVHLSALFNKEGIKSEFIVSDVRDSVTGVRISREENERKLEEYRSGN VKVLNVNINILTEGVDPQTKTVFLARPTVSTVLMQTMIGRALRGEAAGGT TLAYIVSFVDNWNHIAWVNPDLSFGNNEFTDNPVEHVKHLMIAISKI EEFASVLDDSDTTMLEKVPFIQRIPIGMYAFTYLEENGVDVFSYQVMVYDS TRKAYEDLMASLPQLFKIFDCKDEYPSSEELLEIEAEHCRDTFFLGDMIPPYE EKDCLRILKYYAQYEIPPQFYLFDDIDRNKLDVAVIAKYIWDEDMGQRRKT EYINSIWENSDENLLRLLFFGRKLYFWRQLDIELTKLSNPNIYKKEENVQLGI

		RKLEDLSLFEIGKVNPEMEKNLRDTAFEKAKDSEGNICACCGQKRTDRIF FHVDHIIPMNKGGKTVPENLQILCRQCNGMKGDKFA*
28	EpiCas028	MVAINKRALSEEDIKMKYITPAIEKAGWDIKTQVRAEYTFDGRIVRGNL VARGKKKRADYILSYKPNMPLAIEAKDNKKPLGAGMQQALEYGEILDIPF IYSSNGDGFLEHDKFTGKERVLSLDEFPSNELWERYKKGKGKIQPHEEKI EDYFFDMSGKTPRYYQRIAINRTIEAIAKGQKRILLVMATGTGKTYTAFQII YRLWKSRTKKRILFLADRNLIDQTMANDFKHFGDKMIKVNRRKNISKAHEI YLALYQSMTGSEEWQDIYKQFSPDFFDLVIVDECHRGSAARDNSAWREVLE YFSSATQIGLTATPKETRDVSTSTYFGEPVYTYSLKQGIEDGFLAPYKVVRV SLDKDLEGYRPAKGKVDKYGNLIEDRIYNVTDFDKNLILEKRTEIVAKKITE FLKKTNRFDKTIVFCVDIEHAERMQRVLINENSDLVKENPKYIMRITGDDEE GKAQLDNFIDPASKYPTIVTTSKLLTTGVDAQTCCKLIVLDSNINSMTFEKQII GRGTRIKEEYGYFFTIMDFRGVTRLFADPEFDGDPVVVYEPKKGEDIVPP AEGDYGDFGDSVDGTVGEPDSDENETGTRIKYVVDDEVMVNERVEYY DKDGKLVTESLKDYTKKNIRNEYADLNEFINRWNSDKKSAIEELKERGV FLDAIKEEVGKDLDEFDLILHLAFDQKPLTRKERAYNVRKRNYFSKYSDVA REVLEALLDKYMNEGITELESMEILKLDPFRKIGTPKKIIFAGGKEKYIQT KELEREIYA*
29	EpiCas029	MSNPGNSVLSTRYRDARTLVRRVCRLPGEDRAPFKEKVARLREHFEQFNV DVSELCQWFMGLRKAHAAHAANAATSPAGFGPLGDFLLVPALDGVDA ESECNRWRLAVFDAVAGFRDLTAPIREAVPPALHAAIDATARNPHTNPK NVAMKRLFERLRTLDSAHRLVLLKSAAEWVVARVYRRGMENWVRQREEW EKEKTAWETARPQLTPEVRERFTQVFRSLVWDRDKRPGVTNKRPRICPYE RLKGNLDNCIYAGQKGHGPLCWKYAGFKKTREEQFSRFNDRGFEDAAE FLARRKDGVRREHVLRLQMFRGRSADEQRFTDNWSAYLRALSLNEDTVVE HGRPHCLRIGETFEKSKCEWNPHTELCKQYKRALDNPNGFDEETLRLEP LYRQWRRDYLAGPRKPSFRYPSSRDLPMPKIFGAGFHEIDFRSIVRLRLDD MSAGQWIEFGFTPWPRGYKPSRDEIKAAGRVTSVHVHFIGTRARLGFRFSV RPPKPSRFTCSQDELDELRSRCFPRQAQDQQLDAARQRLDLSFPGGAEAAK RNMRLLAVDLGETGACAAVYQGTTHQTDCLAIKINRLYAGRPDALEPD

		VKGRPSKEKRKFDKGDPRGVRKEHVARHLARIAQKTAETQHPRRQPAVP QTVTPGAHDLRSLKRHTAWMIRDWTRLNAAQIVKVAEDQRCNLIVFESLR GFLAPGYDKLCPKEEDKKRWLAMFAYGRVRRKVVEKAVERGMRVVTVP YYKSSQVCSTCGAVQENTNLHRENKRDRKFICEKNTCAVELNSDANAARV LARIFWGEITLSPDEKTVRRS*
30	EpiCas030	MTRIKYRQERTLVRGLQRLPNQDKAEFNKNVLGLRRHFEQFNLNVAELCQ WLISFKPENPGSV CETKLFWEFMLEPENFCPDYDRGPDWLRHEIFVLAAG WRTFEDIKAHNMPESLFESIKIASSKPRSKTAEALFIRLKSLEASHVMVLLKS ASEWLSTRYVRQSENWKQNEQEWQKKKA AWEDKHPELTPAIRDKYNAIF VELGISEKRPHVCSWKLS ENKDNCDWAGERKNIGDKWISHSDLCIKYHE FARKLRSKQRQHFVDNANQYLELRRYPQWTRDMAMNGLFKNVPLARN WFRNAWTNYLNALNILETTILENYSGLPHCEKLSDECVFKKHTDNCRRY KLLLGEKLSNQERELEETTYREWRRFLAPPNKPFLRYPSAQKLPTPKLFGR GYDLDFTRHVVKLRLDDMPADNFVSFGFKPWPRGYDKKPG EINITSVHV HFIGTRARVGRFAVPHSDSRFSVSQDKIDELRSGGFPGKSQDQEFLNEARQ RLLDGMNENQKSALRIMAVDLGTHRAAAFFTGCIFNKAKLLKLLKIDLL TEPKTDTTKPEKLSADEKKIQREKGLTQHHVGKHFETLEARTKEISKRQN MKMAPSDDTPDIVGDHDLRHLTSHIRRMIRDWVRLNARQITELAEENVD LIVFESMRGFRAPGYDKLDLKKKRRLAFFAYGQIRRKVAEKAVERGMRVI TVPYFKSSQICACGRSQNDKNKLRDNKWKQSFQCEFSDCNYKTHSDENA ARVLGRVFWGEITLPTD*
31	EpiCas032	MQQGCDSRFDTTIKEQEVSLFNTYTKYVDCVNKLPAQLVDADFSKLKLS MVIANALEVLNGIKKTV DNDKKYLLLSGGFKLPEVVDVLARLTSNVRQDG AIEIANFGLNSRAVRKRDILNSKKNELDEIESELGALTNKIKKSAQKPKKEDI NSKGRLTQERNSLVAEIKKLQAMNKYGLIFRRKNCDGDFEYFLALVDRDH MENLSDKSKFKSAQSTGYQYMEYNCLKLKSLLOIFESDS DNLSGVRKLER YGTVKPLFLSKEHSRELEKSKTLIAWLDENIQKIVDTIAGTTNPKITYTEYG NTDAAKNAADVWRDFERHFNRQSYSISWCHLNFEKLKTTKSVELFQIANK DHSIHPDFTSDERDRKKLNRSSGRPNLFTTYLDNVLTDEIGNHYQLLGEGV EYFKHKCGAEPSTPRRKSIDRY YVEFRVAFVPNKSESWEETNSSVKNAISQS

		DSPKHFIGIDRGENAIANIVVIDENNRAILKQDMSCLAIFSDITPHSAKKS _{YIF} LTGDEPVYIATKSATDALTDKTEIKKLS _{DVM} FTEYRD _{KLY} KLPPSED _{KKV} IPSKSSPCLDEMNI _{EDL} KKLISKLESPEYN _{PAS} RKMYLAKTNTPIIDYVYIRR YNILLQAKYMLGKKQVRV _{TDS} NFEQEKIQRKEYLLHATNSIKDNYSAYF VHHVCKLAEKYNAYIVFENLNKKNGDTREIN _{FLS} GEQISGLTPKQQNQLKG VSFYVNMEQKLINKLNYLATKDAQSHSLTLKGQFTQSIDYKGVITSMTNEQ INLITSPKKQE _{QVE} ALKSGLGKGKNCYQIGCILYVDAANTSKECCNCGFS VNTALTRCERTDEGLS _{FGS} SSNTSYKWGDTRLTERGIDESFDLNNKSDGVK LSVRPGKVVN _{AKL} STRDVYDPISCLACGFDGRGTGHDHLHENDNDAYQKI KLLQTCDELA _{AFNI} AKRGGKLIKQEEV*
32	EpiCas033	MNYSDFVGLYSISKTLRFELKPTPATKLLDDSGVKETDKIRHEHFKQAKK ILDEVHREFINDALKNVSLSTSDLQDYIKLTLDKASKKADVQKAQKKIFSS _F EQAYKITNYEWQEKLDAKKFFLQSPGIQLEIASQLAEKIAKNLEMNEDDVK NTLKQFDKFSTYFKGYAENRKN _{CYS} FDGKATEIANRTIAQN _{LIT} FYKNIEKF NKYYSGNSLLADDVSKYFLENNY _{NQ} FITQQSIAQNYNVKIGELNYIINQER QKQKKDAKANLPLFKVLYKQILGDATKEVPFEFIKDDAEYKAIHNLKDK MPKWIEEIEDHFVSQAKQLQNPEGVYL _{GKK} AIN _{TLL} QKYISIDVRPLINEEL PELAKNKSSKTDDDES _{LRN} QFISVKDFFEALDKVQE _{QSE} TILRSDESAALL PSQFIWNKLCNDIEKDVKVAKDSAPKIDTSN _{KNS} VKSALDSFNDLFRMYSY FELRYKNQPVAVMDEDAKFYQIFNAEDTPFANEKDKRFSQVYDLIRNYLT KKQYSLDKWKLNFDCSILLGGWDLNKEPEKHGVLLRHKQQYYLAIDNSA KNAFDKKENLDLYTAGADDWQKMEYKLLPGANKMLPKCLFSKSFRENFG VPSDIDVLYKNDEYKKGDKFDINKMHKLIDFYKDAIKKYEGWSMYEFVFS DTKTYQSIDQFYHEVDRFAYRLDFVNINN _{QVL} SKMNENGQIYLFEISNKDW LKEARGKANLHTQYFLNTFRGGDFKLNGEAEIFFRPKSIEAKVDKERKSKH EIIENRRFSEDKILFHPITFNRNSKLSSNFAFNQHAIERIKTNGVN _{AVI} GIDR GEKHLAYVAVVDKNGKLLDKPVSLNQIHSICKGRDFPKDYFDELKKREDE RRSNRQDWD _{AVRRI} DKLDGYVGHCVKQVVDLMVKHNAVVALENL _{NFG} FKQGRSKIERSTYNQFEQALLKKLQYVVDKNKDDDEIFGVQKGVQLAPPD VAPGKISNHMGFVFYVDPSY _{YTS} AIDPKTGYRQHVRLLDRVKTKNFSKFVD

		DAFDKIYFKDDKLIFEFNWRGLAKAQNGLGVSKNKKTDIKQDKKWGITLD VPRTVFHKNVAGKSETEIVNPQQLLICSLNEAGIKLNEDIKLQLVKNTNLSS KFVKSFVYCFNLANKLRNQIDGEDAIIISPVPNGFDSRNQKLDGKYKWNG DANGVYNIARKGAILLQKLNKASTPKEFNCKVSSEEDYDGFVGGEK*
33	EpiCas034	MIFNCILTEHQETQNTKVLVEKFVNDFKDLMGDMQKQEEYLAENYGVDK SLYKIESIKNEGDEKSKMDTATFYIYPYFDFDKNLSIEECIKRVDSFIGDKKL NRNVEELIGVSDNAGGFSFYFNGLQLLQNNRIDEIKNAILQASSLWNGKE NELQERLEFLNKKAKQVLPVRIVKNWHEYRSDIGGSIQSWVSNLRQDIEI KKLLYGYKEKKVKRGGEIEEHDIEGHIQEIEKVLGDAELGGYKQYFENKED EGGIRQTEEVIAKVQKLKHLDDDEKEKIDTEQPKINPAILNDYRDVLADFR NTFNHLYQKRYGLEEQDEGIEQKDEKGKKIQTVNKKYPKLFKELPKIPKFL GAVKIEEKGVYDKYICSYNGLKKGVEFLGGVTFTGSKRIDDYTYRNDSS KDKEEIEKARKRSIERVQRVLENILRLYRSTDAKQLKGQKIKSLLAKYCKDI SLLDGEYNYIYRNERYARDKRGQKIELKRVDDELVEDDFIQGLISDLQSFE WRLDNIKQGKWYERVAGWGALVEIEKIKIGLMAWFYEISVPHNFQEIEKN KLHFRKVEAIVKRYNQDEVLDNKEKNTIIQQGIVSEMGRGMIARMSKQEIIA RYVVQPVNAEQKYPLVMDEGKYYIHPKEILDKDEIKEQPVMIDISKEKDL TEEKNSRIKVRRDSLFAITSSKYQTQFLNNSLNRQLEKWQNLNPQLSEYSFI VEDTHKIKWVQDGDSYTPNCDEVLLSRLFVSVPFKIGVENKDKTVAIKQQ NKFIGIDIGEYGVAYYGLDGSDFSAPFVAGFFFQPDMRKIREGIAKNKERQ QAGTFTNSNTQIKRVRDNATTKIRNQVHQLVVENNALPIYEKEVSASFESGS GKISAIYNSIKRSDVYAESKADDMEHKLWGN DKIMVGAEIGAYATSYMC SVCYESIYAYIAKDDYEREYEVIDIVGEQKQIKDKWNATVKFDIEGKEMYA YIEQKDKPVGKGDTFNGKGTINFVRKYERPPLDVVRKIERLPSAIQQELNEC TKKERVKSNGEKEKISVWENRRGSQGVFRCAFCNKCSDADVQASMWIAL KGYLQNVKNAVQKKENKEQYKMYEGSIGKEVVGEIAKGFSQGGKEQVRF ILDFAKDNSIMLR*
34	EpiCas035	MATKTYRARLSSIDNRTQQA VVRTHYFFNHCLRQMIERYLDMRKGYGPF CKQLGDL LLLTRSH TDAHGIMDQLTRGAPSTQTDQEWAVLVKHHHREYGP LFLQHEGFVVDGVVVHTKPQGRHEPTPGRLAVTAKFWHQICDMACAFL

		KSNQELMAVWRKDKTEWLKTKNAWEVDNPEFMTFWNGPYADFERACEA ARVAAQEAAGQTPTARKRDSRVRGKRIDRWHLWYEWLV AHPETVGWRQ RAEAGDFKPVPDAVQAAIRKKTKRQDKYVGKYLDWLRENNPELKALDSL RRTYVRRYLRFKRPPTLTLPSANKHPYWFTLELNEFYKGADFEKGCVKMR LIDCDADGLEMKWFDVRMGCDHRLKPAWRKEKFASDGRYPPYLGGKIG RKLD RPAECAADRKAGVKGVKLVLDKGRRELFFTLVEQDCPPRLKFKKTK NRRCTADNLF DAGDQRVPIRVM SVDLGIRHIGAYAICEGTKQDDHWEMK YLKKGLLS DSSLPELGQIRLHDRHLKKLSRQGDAPTGERNFVALQDHRTG MADDRFKKAAHLLIETARRAGVDIIVFEKLDTLKPDAFDERWINRQLRDM NRRQIVTHVAQQAGEFGLLCKDDISPWLTSQVCSRCYRPGWRFSIKPKAPY KEQCSRAACQDFGYPIWDPGGHLFRCPHCGYKVNADINAAGNLA AKFFG HGFWFDNNLSRDKASGCFIWTDDQRRCFDAREAFEQWAAEASRRQALKE APF*
35	EpiCas036	MATKSYRARILTDSRLAAALDRTHVVFVESLKQMINTYLRMQNGKFGPDH KKLAQIMLSRSNTFAHGVMDQITRDQPTSTLDEEWTDLARRIHKT TGPLFL QAERFATVKNRAIHTKSRGKVIPSPETLAVPAKFWHQVCD SASAYIRSNRE LMQQWRKDRAAWLKDKNEWQQKHPEFMQFYNGPYQNFLKLCDDDRITS QLAAEQQPTASKNNRPRKTGKR FARWHLWYKW LSENPEIIEWRNKASAS DFKTVTDDVRKQIITKYPQQNKYITRLLDWLEDNNPELKTLENLRRTYVKK FDSFKRPPTLTLPSYRHPYWFTMELDQFYKKADFENG TIQLLLIDEDDDG NWFFNWMPASLKPD PRLVPSWRAETFETEGRFPPYLGGKIGKKLSRPAPTD AERKAGIAGAKLMIKNNRSELLFTVFEQDCPPRVKWAKTKNRKCPADNAF SSDGKTRKPLRILSIDLGIRHIGAFALTQGTRNDSAWQTESLKKGIINSPSIPP LRQVRRHDYDLKRKR RRHGKPKVGQRSNANLQAHR TNMAQDRFKKGAS AIVSLAREHSADLILFENLHSLKFSAFDERWMNRQLRDMNRRHIVELVSEQ APEFGITVKDDINPWMTSRICSNCNLPGRFSMKKKNPYREKLPREKCTDF GYPVWEPGGHLFRCPHCDHRVNADINAAANLAN KFFGLGYWNNGLKYD AETKTFTVHTDKKT PPLIFKPRPQFDLWADSVKTRKQLGPD PF*
36	EpiCas037	MENKNCNFDKFTKQYQVSKTVRFALTPIGRTEEHIIEKQYIEAAQRKDQAY KIVKPVIDKKLCVMIDDLTHCEKQDWEKLNELTLQYQNDKTSVEALGEQ

		QGKIRKNISKEFENSIEYKNFFGKDSEKLFKTFPQCLEEINASEADKEAVKE FQKFkTYFNFLTVRKDIFNDDSKQNTIPYRIVNENFMIFAGNRKTFSNIVK LIPDASAEIAKEGIERQEWEFYDVQNIDSCFEPDSFQQCMSQLGIQKYNFIV GLVNSYMNLYIQQHPEALEVKRSRLKLRMLHKQILSDRTNPSWLPEQFAE GEEGEKQIYEAIQALEDDLKNHNFQEKFDNLLQLIDTENPRIYIAASEMPRV SSALNYGWNGLNNVRMNVLQELNSEAKIEKILKQDVSLNELSDALNKYAD IYKKEQVPTLQQYIEYGKELLQDCAVACNDYHASIRNNVERLNQNEKLIEV LKTYLDCYQRMHFMNVFVVGDELDDKDVDFYAELDELIEDISQIPLYNKV RNYVTRKAYS�DKIRIMFEKSDFLGGWGQDFDTKKALIFEKDGLYYIGHIEK KYTNTDVQYLYEEIEEGNKAVHFVYNFQKIDNKNFPRIFIYSKGTNYAPAV RKYNLP IQNIIDIYNERKFQTDYKKVNEQEYYESLEKMIDYFKDGIMKNET YRKHFHFNWKPSNEYENIKEFYDDTDNSCFLLEKEEINFNHLKEQADQGKIY LFQISSKDFNKGSTGTPNLQTMYWRELFSQDNCKEGIIKLCGGASIYMRDV SIKQPIIHKKDSWLINKWYMEEGQNKVIPDDTYVKFSQIVQGRMSEAELNV EEQQLWNSGLIQKKKATHDITKDRRFTRKQYMLHAPLIINYKQQDIPTPRR FNEEVRSFLKNNPNINIIGIaRGEKNLIYITVIDQKGNILSGMQKSFNQIEEKG ANSRIVDYYGKLQSVEDRHTARKNWKQIGTIRELKEGYLSQVVHEIAQL MIQYNAIIVMaNLNMGFKKGRMKVEKNYQRFKMLIDKMNYLAFKKNIE GKAINPYEIGGVMNGYQLTNKFTNFADMGSQNGFIFYVPAAYTSVIDPVTG FVDVFQRAEFKTGDFLQKFEDISWDDKEQSFVFTFDYQKFKCNGTCYQNK WSLYADVNRIETVKKNNQVYRQEQCNPNQRLIELFDSKGIAYRDGHNLVN DLGNYDSKLLSDILYNFRLILQLRNSMRDYSTGEVIDYIASPVMYHRERFDS QKKNPSLPLDADANGAYHIALKGLMCLQKINEYADPDGKMDIKNMKITNE EWFKFMQTRRESNV*
37	EpiCas038	MRRFFMTKMTDFTKLYQLSHTLKFELRPIGKTSENQNSGLLEEDFKRSQD YVEVKSVLDEVHKKFLQDVLSDCDLSWQELADALKIFQVNKDRKVLEDV QKKYRKLLADRFNKNKFFKTLTEATPSKLFKSLAEPDMAERKVETFKRF ACYFKGYQETRQNIYSAEAQQTAAAYRAVNENFVKYLRVVELYENFRKN FPELAEIEGRRTAELTNGELSQILRIENYNRFIAQSGIEYLNSLIAQINSAVNL YRQAHNVKLSFVPLLFKQILSDREQAFTLERIEKDADLDEKLRAFCE RHIGN

		VEIRGETGDLFSRLKVLLTTLNDDCELYIDAASLSAIGNALTGNWSCFRDA MTAYANETLKNKSLREKYCKQPVYSFSEIKSWRVVRTDDDDGNVQPVCLSD YWHGEKTLELFAAEKILNRNILEMVRVPHDKNFRERKNDVALVKDYLDA VVEILHLLKPLNVGSENGGDLDLLGILTEFYEALSPVIPLYNQIRNYVTQKP SDIGKIKLMFNNPTLADGWDLNKEKDNTAVIFCRNGMYYL GIMNPKSKTD FNSLATDSSTDSYQKMVYKQLQNPQNMLPKVFFSAEKIKFFAPSEELLRRY EAGEYKKGEKFDLAYCHELIDFFKRSIDRHEKWSKFEFKFSETSAYSGINEF YREVAEQGYKMSFVNIPVKAVDQLVEDGKLFLFQIWNKDFSEKSTGTPNK FTLYWKALFDRQNLSDVVFKLNGEAEFLREPAITEPVVHRPGGKLVNKT VVAAVDGNNGVAIRNSIPDAVYTEIYKHVNGMEEGTLSPAAREHLNRFPRL DWEPMGLEATVKRTVVKDVKFQLVKDKRFTERKFSFHVPI TVNFKSPRL KGAKLNEMVWEFLKDNPDVNVIGIaRGERNLIYVNCIDLKGRILLQKSFNIV GGFDYHQKLDLRERERGGARKSWNEIGTIKELKAGYLSGVVYEIARLMVE YNAIVVMaDLNGGFKRSRTGIEKQIYQKFEKALIDKLNLYLFKNESDWHVP GGLLNGYQLTGPLESFAKLKGQCGWLFYVPAAYTSKIDPSTGFVSLFKLSN YTNTTEKRKAFFENFESIVYQAEQDAFVFAFDYKKFKDVENCWKTSEVYS ARRRLVYSRRSRQIEEINPTRLIKRALDKRGVVLSDGFDLKHYITSVEAVNE NAAFFKDIFDAFGTTLQMRNSSGEEDYIESPVLNENGQFFDSRKAGKEWPR DADSN GAYHIALKGLLRMKSGQTAKLKRLDWFEFVQRRGFITE*
38	EpiCas039	MNDLSEFVGLFPVSKTIK WELKPVGMTAERMREAGFLDRDRSKRDAYTV VKSMADRFYRGYIDERLSTLDLDWNHLGDAMSALKESGETKDKRELENIK KSFKDSISKHIFTKKDDDEDIFSAKVLARLRGTETDDEEVEALEKFDRFAGYF QDFFVNRRVFADTVRGSCAYRIVEDNFPRHFENCRKYQIIRRDNP ELILELK RSMEGYLETDPDVLFSIDHFGKLLRQRDIDAYNRVVGSVFRSESDNSRGFN PILNEFHQRNPDSPKMLMQKLQKQILSEPDSGSFRMRNYENEGQMFEDLQ NLWTGIETNRLLEIPASVLEEGEFDKSGVFISKAGVNTLSNWMFSDWERIG GLLQIFKINEMGDSASKSRKKIDKWLDSDSFSLDVLNRAVADET DIGIVPFL SRRLSDAAQQAREAYLGLEGCA SMRREQYQGRVKTFVDSLRLRYRDILKLF SYDPDLPLDRGFYDTFQFVYEEMASAYRAYNRCRSFITRKEYSTEKFKMSF DCPHLADGWDLNKESYNSFLFRRDGLYYLGIRNPAWRYEDFSVKADPEC

		DDFYEKAVYKQLQDPVKMLPKVFFSKKGLDKYRPPERILEGYENGLHRQG DSFDLGFCHELIDFYRDAISQNRDWDFDFKFKETSEYSSISEFYQDVESQG YQMRFIRIPAQTIDEMVDEGRLLLFQMYNKDYSRDKTPGGKPNLHTMYWE AAMSEDNLRDVRIKINGYAELFWRDASIKDGTVHRKGSVLVNRIGKDRKPI PSEAYYDLCRYCNGIRDDL GAGKGYEDQIVCKNSRYDIEKDRRYCHDRMF LHIPLTFNFKSDGKTDVNRKVLEWIRRNDDVKFLGIaRGERNLIYLSLVDSD GRILWQRSLNTIVQKDGRGNTRALDYHEKLA VREGERRNAQQSWGPMDA IKNLKEGYVSAAVHEICRIAVEEKAVIViaNLNYGFKNSRKKVEKQTYQTFE RMLINKLQNLAFKDAEPGSPGSLAPYQLCGPIDSMEDIPNHCGMLIHVTA AFTSRIDPVTGFVS VFNLSDITNNAQRADFLSRMTSIRYNQDEDEVFEFSFDY GDFRRLTPKVRKWTVCTYGD RYKYNRKDRTLVRISPTTMIKDALTKAGI RYEDGSDIRKEISGSDAVRQVFEAFEAILDMRVTTREDEDVIVSPVRDSSGRH FVSGTSGPGLPQDADANGAYCIA LKGLLLARKIRASDEDRLPLLNNNDWV EFMQTRVM*
39	EpiCas040	MIEQFVNRYPLSKTLRFG LIPQGNTENFKIMQILEKDEALARNYGKAKEII DRYHKFFINSCLADFH L TDLEKYALLYEPNKTEEEVAAMDETAKTMRQA ISGLFCKENKAIYDRLSKKDLIENFAMSVIETDEERTVLESFRKCTVYFTGF NQNRNLNMYTGEGKSTEIAHRLINQNL PKFLDNARVAKTILAALPKNDLDQI DVDFYDIFGINTQNIFEIPFYNQTLTQAGIDRYNQIIGGYTCSDGRKIKGYNE YINHFIQETKQKL PKFKPLYKQILSDRETISFIPKAFQSDSEVLSALRNFYIKK NVETDKSIRDVITELKKLFQEMDSFDEKGIYITNGVAITDLSQAVCDNWFVI KKAMQAEYDAEHSSKNRNEDYEKKRGNDLKKVKSYSLSALQRYMDAEM EDKKS LAEYIKTHTTQLCDLTASAYERVRH LLETEYSPTKKLTADADAIGRI KDFLDPIMEVFHF AKIFLGSGEEAEKDDVFY GIFSPLYDQLADIIEIYDKVRN YVTQKPYKTDKIRLIFERGD FL SGWAQPTESQAKEAHLMEKDGRYYLFITV TAINKADWERALY GKD NFQCACHIEYYKQKPDNKNTPRLFIRSKGDHYAP AVEKYNLP IQEVIDLYDKGY YKAEFKKKDQKLYYTS LRKLIDYFKLGFTK HENYKDFAF AWKESDAYENITEFYQDTMNACYRLERVTVNWN GIMDLVE KGRGFLFQIYNKDFSDHTKGTPNLHTMYFKMLFDERNLENGCYRLLGGAE MFYRKASLKLSETTIHPKGVA IKNKNPRSDRAERVLDYDI IKDKRYTETQFS

		LHFPIMMNCLAQGNPKLNLDIRKALLKSEHQYIIGIaRGERNLLYVCVINEK GEIVEQMSLNEIIGDNGYTVDYHELDRKEKERDAARKSWG TIENIKELKE GYMSQVIHKICELVLKYDAVIAMaDLNGGFMNSRIKIEKQVYQKFEKMLTE KLSYLVDKKADPMSDGGLLRAYQLTNKPDPNKSVQHGMQDGIVFYIPAW CTSKIDPTTGFVDLLKPKYTNKEEAKDFFKRFQKICFNEVENMFEFHFHYP DFQRGITAGNTDWTVYTNGSRIRTFRNPEKNNQWDNTDIVLTD AFKALFA EYAIPCDADMKEAILAQDSPKFFKELCGLLSFTLQMRNSITGRTDVDYIISP VKNAAGR FYDSRKAASDLPCDADANGAYNIARKVLLFIRRLQAQEE SNLM NKDLEKITNAQWLSFVQNGHE*
40	EpiCas041	MTKMTDFTKLYQLSHTLKFELRPIGKTSEN LQNSGLLEEDFKRSQDYVEVK SVLDEVHKKFLQDVLSDCDLSWQELADALKIFQVNKDRKVLEDVQKKYR KLLADRFNKNKFFKTLTEATPSKLFKSLLAEPDMAERKVETFKRFACYFKG YQETRKNISAE AQQTAAAYRAVNENFVKYLRVVELYENFRKNFPELAE E IGRRTAELTDGELSQILRIENYNRFIAQSGIEYLN SLIAQINS AVNLYRQAHN VKLSFVPLLFKQILSDREQVFTLERIEKDADLDEKLRAFCEQHIGNVEIRGET GDLFSRLKVLLTTLNDDCEL CIDAASLSAIGNALSGNWSCFRDAMTAYAN ETLKNKSLREKYCKQPVYSFSEIKSWRVVRTDDDGNVQPVCLSDYWHGE KALELFATEKILRNKILEMVRVPHDKNFRERKNDVALVKDYLD AVVEILH LLKPLNVGSENGGDL DLLGILTEFYEALSPVIPLYNQIRNYVTQKPSDIGKIK LMFSNPTLADGWDLNKEKDNTAVIFYRNGMYYL GIMNPKSKTDFDNLAT DSSSDCYQKMVYKLLPGPNKMLPKVFFADKNIDFFAPSEELLRRYKAGEH LKGEKFDLAYCHKLIDFFKRSITRHEEWSKFDFQFSET SAYAGIDEFYREVA EQGYKMSFVNIPVKA VDQLVEDGKLFLFQIWNKDFSEKSTGTPNKFTLYW KALFDRQNLSDVVFKLNGEAEFLREPAITEPVVHRSGEKL VNKT VVA AV DGNGMAIRNSIPDAVYTEIYKHVNGMEEGTLSPAAREHLNRFPRLDWE PG MGLEATVKRTVVKDVKFQLVKDKRFTERKFSFHPITVNFQSPGLKGAKL NEMVWEFLKDNPDVNVIGIaRGERNLIYVNCIDLKGRILLQKSFNIVGGFDY HQLDMRERERERGQARKSWNEIGTIKELKAGYLSGIVYEIARLMVEYNAIV VMaDLNGGFKRSRTGIEKQIYQKFEKALIDKLN YLVFKSESDWHVPGGLLN GYQLTGPLESFAKL GKQCGWLFYVPAAYTSKIDPSTGFVSLFKLSNYTNTE

		KRKAFENFESIVYQTEQDAFVFAFDYKKFKDVESCWKTSWEVYSARRRL VYGRRSRQIEEINPTRLIKRALEQRGVVLSDGFDLKHYITSVEAVNENAAFF KDIFDAFGTTLQMRNSSGEEDYIESPVLNENGQFFDSRKAGKEWPRDADSN GAYHIALKGLLRMKSGQTAKLKRLDWFEFVQRRGFITE*
41	EpiCas042	MRNRYSKILCLLLLACCLPQEGNAFWPFKKKKKEDKRLRKLIAQNLKAH PLVDKLVKNAFGKNGYLETLPLGKEEKEAVKVFAGFGGFFNNYNKNRE NYFSTEEKSTAIANRIVNENFSKHFSNVEIVTKIQKEVPELIQIVEAQFKGYD AIFTVNGYNMALSQAGIDTYNEMVAIWNKEANLYAQKAGKLPDGHPLKK KRNYLLSALFKQIGSEKEHLIQIDRFDGDEEVIEALTGVKKMLQEADVFEK LNMLVEDMENWDYSKIYLSAQSLSNVSVFLNNLYEDERENSWNYLDNVL REKWQIELQGKKKGTDLEEAIKRRKQSFYSIEELQEAVNAIEETDKCYNVS KWLLGAMKSERVIEEKKKDVEDFCTQWKNERNSLKETDITALKEYLEQWI FLARYCKSFYANGIEKKEKDEAFYHILEDVLYVLDEVIFYNKNRNYVTK KPYSLEKMHLKFGHNELANGWSVNKEGNYGTAILRRNGKYLAITNSLN KKMSIPTQLESTGNNYEKMVLNVFPNVFRMIPKCTTGRNDVKSCFERKEPN EYFFIDTPEFVNPFKVTREEYELNKITYDGVKKWQSDYSKNTQDEKGYKE AVTKWQFCMRFLQSYKSTAIYDYSTLQQPEKYETVDSFYHDVEKILYECH FEYVPANKIEQLEEEGRIFLFQIYNKDFSENRRPDSKKNLHTLYWEALFSEE NRKAKVIQLNGKAEIFRREKSIEHPIVHKAGEVLVNKRTKDGEPIPDDIYKD LSNYFNGRNVTSEKEEYKECLDKVYTSTKKYDITKDKRFTETKYEFHVPIT LNYQADGVKYLNQKILHVLDRDNPVNIIGLaRGERNLISYVVLNREGKIVN NQQGSFNIVGKMDYQKKLYQKEKNRDKERKTWKNIETIKDLKEGYISQVV HELTDMAIRNNAIIVMaDLNFGFKRGRTKVERQVYQKFELALLKKLHYLV TDKTEGEAMLKPGGVVLQGYQLAREVKTLEIGKQCGCVFYVPPGYTSKID PTTGFVDVFNMSGVTNREKKKAFFEKFDNMFYDEKRDMFGFSFNYEKFTT YQSSYRNDWTVYSNGSKYVWNSLKRDELIDVTKEKLLFEKYAIDYRNE ALFEQIMSQDTDKNNADFWNKLFWYFRVLLRLRNSSDELQIVSPVLNQN GEFFETPKKITEKSYLSDYPMDADTNGAYHIALKGVYLIQEKIADSVDLN NKLPKDFYKISNAEWFMRQKEK*
42	EpiCas043	MVSNKEKTPTYNLSLTNLYSVSKTLRFELRPQYSTLDHINEDQIVGKGEELK

		NHYKTFKKILDQVFSRIINDSLDKTYLDQKYISTYQDLAFKHRDQLTDKDR AELKTLKETLKKQIDKSLDHKDKKAIFSDPVKFLLDNESEFTDLIGDNRPSIE AFNRQKGYLSGYLQNRANIFDHTTNETSVAFRIVEENLAIFLNNRITLQHFL EKVAGKDGLLKSLEALFQLGFKLKLEDLLSLDYFNRTLSQSGIDQYNLLIS GKTLEDGAKLQGINEILNQYLKQHQUEEKLHKIKLKQLYKQILSESKTESFTL DFVEDNKGLAAMFLQFIDFANQLIEEKMPLLGMIQELKNSSVSAELLSRLH LNRKNIKRLSNFIYKDYGYIEQALEENFLSTVEGKITKKVLGEHRKQDTFTI HEILVALQKQQYEKDGTTLESADRLFLPSIVDFVYQNLNCEYSTLLEKVRSE KQPLLELFKERQLLEDQDTESHSSKYPDHPFNDHEIKIITTLDFYKNLQSN LALLQISDENLKL DSEFYSEFDEFYQGLKNIIPVYNKSRNFLTKKPFSTEKTK LIFNSPQLLDGWSKSKESDCLGTIFIKDGKYYIGIINSATNAKKTLEPHNFA NFDQSQYFEKMNLFFLSDLKRDFPKKYFSEKWHNQYPIPADLREKYDYRYI DEHKDERKNDLEYHHQLIAYYQDCLKKDTIEWQIYQFKYKSPEEYSDINEF LSELTPTNTYKMEFTKIPAEYIKKLVDGKLYFFQIYSKDFSEFAKGKPNLHT LYLKAVFDQKNAEEFNYNKYISGSAEIFYRPASIETRVTHPKNQPIKNKNKN NPKIESIFQYDLCKDRRYMSDKFFLHLPVELNRTPLLPNDSLVNSMVNQVV SSRNQNYFLGlaRGERHLYLVLIDQSGKIIKQQTNLNQTSSYQEKSNQTV VVTDYHDLLNDKEKIRKKNLQEWQSVENIKELKAGYLSNVVNEIGKLIAE YQPVIMLaNLNTGFKNSRIKIEKQVYQKFEKALIDKCNFYMRKDIDPSTIGG LYHALQLTKEYSKQYNGKQNGHIIYIPASYTSNIDPTTGFI SAFVQTRYENV EKAKALIEKFNDITYDADESLFCFTVDYKKFSPEAKLWQQNIWQIYTNGDR IYTFRDKEEWQSKNYILVEEFKDLFTKYQINYNRDLKTQILSQTDAGFFKQF LFLRLTLQMRNSRTADSNEADIDTKKRENDYIISPVKNQYGKIFYDSRQDY MDWPENADANGAYNIARKGLIMLKLHLKEGLPEKRICDISTEEWVQFVEEL NK*
43	EpiCas044	MTRVFIFCLLYLKTKYAPGVHKEGKMKINAFINCYSMSKTLRFKLVPEYE TEKNLLEKGFLDRDKLRADDYDLMKKVIDKYHKKHFIDKALEGFKFDLLQD YAEAFYSQSADDDGKKLEEIKKMCKELATCFQSKQDEFKLLDKKELVEKL IPAAEFIEDEEKDIAKRFGFTTYFTGFNENRQNLAAELKHGTIAFRLIEEN LPAFLYNCKKGVKIFEELDEGDVETLNNELGEILSIGNVKDILNVEYYNKTL

		TQNGIDVYNQIIGGYTQEDGTKIKGVNEYVNLYNQTHDKKLPSLAKLKKQI LSDSYSLSFLPAKFNDDESELLLSLKKFYSMVNEETGLSVEKAIQEMRNVFL HIDDCDLHNVFIDAKFINKVSNDVFGNWSVLIDGINAEYEKLNPFNGKNLG NYEEKRKAFLNKIESYSVDALQAYSGKEEKIADYVQKRAVELCDSVACAY ENMSNKNVINAREGKVKLYQDDEKTEIHKTFLDVQEFKKFAEMFCYDGTD GDTTFYGEFANYYGQIAEIPLYNKCRNYLTKKPYSEDKIKINFDNSQLLDG WDANKEKDYLTVLLLKGNNYYLGILDKNHKNVLIKDVPEKGQGEPCKFK MIYKLLPGPEKMLPKVVFSAKNIGFYNPDRILRRENGTFKTGDGFNIDDC HKLIDFYKECISKNEWDKTFGFRFKETSEYKNIGQFYNEVKEQGYKISFTDI PESYVKDLVNDGKLYLFRANKDFSPYSKGKKNLHTMYFEGIFDPENIKEK VYALNGGGELFFRCASLNYDKPTHKPNVPIKNKTYEFRTDNAKKETSTFEY DLIKDKRYTKDQYTLHCPVTLNFKERGIERINDLVQRSLRESDDNYVIGIaR GERNLIYISVIDGKGKIVEQFSMNNLSSGNDVSIDFHKMLETRERERDASRK NWNTIDNIKDLKQGYLSYVVKKICDLVVKYDAIVAMaDLNVGFKHGREKF ERQVYQKFEKALVDKMSYIVNKNASPHSAGGLFRAYQLTNKKYNEKEKQ NGFIFYVRAWNTSKIDPTTGfVNMPLKYQSKEKSKEFFDKFEDIFYDENK DMFGFTFRYDDFGINIDHKNEWTAYSNGERIITVRNSFGKWDKAKIALTPA FKKLFDDYNVDCRGDVKRQIMNVDDKDFFVRLYKLLSYTMQLRNSDDAD DYILSPVVNAEGMFFDSRNSDGSPLCADANGAYHIAKKAMWAIGKIEA DEESFKKTSLAIDNKTWLEFVQKA*
44	EpiCas045	MVSNNENTPTYNSLTNLYSVSKTLRFELKPQYSTLDHIRDDQIVDKGEDLK NHKTFKKILDQVFSRIINDSLDKTYLDQKYISTYQDLAFKHRDQLTEKDR AELKTLKETLKKQIDKSLDHKDKKAIFSDPVKFLLDNESEFADLIGDNRPSI EAFNRQKGYLSGYLQNRANIFDRTTNETSVAFRIVEENLAIFLNNRITLQHF LEKVTDKDGLLKSLEALSQLGFKLKLEDLLSLDYFNRTLSSQSGIDQYNLLI SGKTLENGTKLQGINEILNQYLQQHQEEKLHKIKLKQLYKQILSESKTESFT LDFVEDNKGLAAMLLQFIDFANQLIEEKMPLLGMIQELKNSSVSVEFLSRL HLNRKNIKRLSNFIYKDYGYIEQSLEENFLSTIEGKITKKVLEEHHKQDTFTF HEILVALQKQQYEEKNGTLESADQLLLPGIVNFIYQNLDCCECSTLLEKARSEK QPLLDLFKEKQILLEDQNTESHASKYSDRPFNDHEIKTIKTTLDFYKNLQSNL

		AIFRISDENLKL DSEFYSEFDEFYQGLKNIIPVYNKSRNFLTKKPFSTEKTKLI FN SPQLLDGWSKSKESDCLGTIFIKDGKYYIGIINSVTNVKKTLEPHNFANI DQSQYFEKMNLFFLSDLKRDFPKKYFSEKWH SQYPIPADLRKKYDYRIDE HKDERKNDLEYHHQLIAYYQDCLKK DTEWQIYQFKYKSPEEYSDVNEFLS ELTPSTYKMEFTKIPAEYIKKL VDDGKLYFFQIYSKDFSEFAKGKPNLHTLY LKAVFDQKNAEEFN YNYKISGSAEIFYRPASIEVRVTHPKNQPIKNKNKNN PKTESVFQYDLYKDRRYMSDKFFLHLPVELNRTPLLPNDSSVNSMVNQVV SSRNQNYFLGIaRGERHLIYLV LIDQNGKIIKQQTLNQITSSYQEKSNQTV E VVTDYHDLDDKEKL RKKNLQEWQSVENIKELKAGYLSNVVNEIGKIIAE YQPVIMLaNLNTGFKNSRIKIEKQVYQKFEKALIDKFNYFMRKDIDPSTIGG LYHALQLTKEYSKQYNGKQNGHIYYIPASYTSNIDPTTG FISA FVQTRYENV EKAKALIEKFNDITYDADES LFCFTADYKKFGPEAKLWQQNVWQIYTNGD RIYTFRDKEEWQSKSYILVEEFKDLFTKYQINYHNDLKTQILSQT DAGFFKQ FLFLLRLTLQMRNSRTTDSNEADIDTKKRENDYIISPVKNQY GKFYDSRQD YMEWPENADANGAYNIARKGLIMLKHLKEGLPEKRICDISTEEWVQFVEE LNK*
45	EpiCas046	MVSEKENTPTFNSLTNLYSVSKTLRFELRPQYSTLDHIKDDQIVDKGEELK NHYKTFKKILDQVFSRIINDSLDKTYLDQKYISTYQDLVFKHRDQLTDKDR AELKALKETLKKQIDKSLDHKDKKAIFSDPVKFLIDNESDFADLIGNRPSI EAFNRQKGYLSGYLQNRANIFDHTTNETSVAFRIVEENLAIFLNNRLTLQH F LEKVADKDGLLKS LQETLSQLGFKLKLEDLLSLDYFNRTL SQPGIDQYNLLI SGKALEDGKKMQGINEILNQYLQQHQEEKLHKIKLKQLYKQILSESKTESF TLDFVEDNKGLAAMLLQFIDFVNKLIEEKMPLLDMIQGLKDSSVSSEFLSRL YFERKNIKRLSNFIYKDYGYIEQSLEENFLSTIEGKITKKALEEHRKQDTFTI HEILVALQKQQYEKDG TLESADHLLLPGVVDFLYQNLDCKHSTLLEKVGS EKQPLLDLFNEKQLLEGQDAESHASKYSDRPFNDHEIKVVKTALDFYKNL QSNFAIFQITDENLKL DSEFYSEFDEFYQGLKNIIPVYNKSRNFLTKKPFSTE KTKLIFNNPQLLDGWSKSKESDCLSTIFIKDGKYYVGIINSATNAKKILFEPY NFANLDQNQYFEKMNLFFLSDLKRDFPKKYFSEKWHNQHPVPADLREKY DYRIDEHKDERKNDLKYHHQLIAYYQDCLKK DTEWQIYQFKYKAPEEYS

		DVNEFLSELTPTNTYKMEFTKIPAEYIKKLVDDGKLYFFQIYSKDFSEFAK GK PNLHTLYLKAVFDQRNAEEFNYNKYISGSAEIFYRPASIETRVTHPKNQPIK NKNKNNPKAESVFQYDLCKDRRYMSDKFFLHLPIELNRIPLLANDSSVNSM VNQVVSSRNQNYFLGIaRGERHLYLVLIDQNGRIIKQQTNLNQTSSYQEKA NNQTVEVITDYHDLNDKEKLRKKNLQEWQSVENIKELKAGYLSNVVNEI GKIIVEYQPVIMLaNLNTGFKNSRIKIEKQVYQKFEKALIDKFNYFMRKDLD SSAIGGLYHALQLTKEYSKQYNGKQNGIYYIPASYTSNIDPTTGFISAFIQT RYENVEKTKSLIEKFNDITYDAEESLFCFSADYKKFSPEAKLWQQTIWQIYT NGDRIYTFKNKEEWQSKSYILVEEFKDLFAKYHIDYCRDLKAQILSQT DAS FFKQFLFLLRLTLQMRNSRTTELNGTDADTKKRENDYIISPVKNQYGKFYD SRKDYVDWPENADANGAYNIARKGLIMLKHLKEGLPEKRICDISTEEWVQ FVEELNK*
46	EpiCas048	MKSIYENFIGLESKNLT LRFALNPEAKTQENLKLYWDKLRDEERDRA YPIV KKILDKEYQQLISEGLKLL ENQNVLDWTELA EYIRTS DLSKKKKEDKRLRK LIAQNLKAHPLVDKLVKNAFGKNGYLETLPLGKEEKEAVKV FAGFGGFF NNYNKNRENYFSTEEKSTAIANRIVNENFSKHFSNVEIVTKIQKEVPELIQIV EAQFKGYDAIFTVNGYNMALSQAGIDTYNEMVAIWNKEANLYAQKAGKL PDGHPLKKRNYLLSALFKQIGSEKEHLIQIDRFDGD EEVIEALTGVKKML QEADVFEKLNMLVEDMENWDYSKIYLSAQSLSNVSVFLNNLYEDERENS WNYLDNVLREKWQIELQGKKKGTDLEEAI RKKKQSFYSIEELQEAVNAIEE TDKCYNVSKWLLGAMKSERVIEEKKKDVEDFCTQWKNERNSLKETDIT AL KEYLEQWIFLARYCKSFYANGIEKKEKDEAFYHILEDVLYVLDEVIYFYNK VRNYVTKKPYSLEKMHLKFGHNELANGWSVNKEGNYGTAILRRNGKY YL AITNSLNKKMSIPTQLESTGNNYEKMVLNVFPNVFRMIPKCTTGRNDVKSC FERKEPNEYFFIDTPEFVNPFKVTREEYELNKITYDGVKKWQSDYSKNTQD EKGYKEAVTKWIQFCMRFLQSYKSTAIYDYSTLQQPEKYETVDSFYHDVE KILYECHFEYVPANKIEQLEEEGRIFLFQIYNKDFSENRRPDSKKNLHTLYW EALFSEENRKAKVIQLNGKAEIFRREKSIEHPIVHKAGEVLV NKRTKDGEPI PDDIYKDLSNYFN GRNVTSEKEEYKECLDKVYTSTKKYDITKDKRFTETKY EFHVPITLNYQADGVKYL NQKILHVLRDNPDVNIIGLaRGERNLISYVVLNR

		EGKIVNNQQGSFNIVGKMDYQKKLYQKEKNRDKERKTWKNJETIKDLKEG YISQVVHELTDMAIRNNAIIVMaDLNFGFKRGRTKVERQVYQKFELALLKK LHYLVTDKTEGEAMLKPGGV LQGYQLAREVKT LKEIGKQCGCVFYVPPG YTSKIDPTTG FVDVFNMSGVTNREKKKAFFEKFDNMFYDEKRDMFGFSFN YEKFTTYQSSYRNDWTVYSNGSKYVWNSLKRTDELIDVT KELKLLFEKYA IDYRNEALFEQIMSQD TDKNNADFWNKLFWYFRVLLRLRNSSDELDQIVSP VLNQNGEFFETPKKITEKSYLSDYPMDADTNGAYHIALKGVYLIQEKIAD E SVDLDNKL PKDFYKISNAEWF MFRQKEK*
47	EpiCas049	MEKFPRCSHLKV KQHCGPFVEWYEDLQSDRQKYVSLFFPWIGIGYGERKD EGDDIDFISPEQTLERRKAELGKLEELVVR FVTVMLKGGSSASSCYSLFYED YKGKDTAKKSTSKVNF IKALPLLSKKEIEDRV LGDNQLSLINLDEWMELGQ SGLSADDIWDKVSQRLVTGERSLGVDVRHTL KIRCQRPRDRGYCVV LME MVARQLRSCYECYKNHNLNELQELREKTTKQITESRDDLFSLVCDLATV LTE QDSGLSERTLFSGRKQALKPQGKPSKIQAIVDEL RQPQYSPLLEADDKELR QAYKAWNSQRRLRDRETRRPYPCRPNWRNDYQVSLGLTGRGGFWMM DGSDVVVDLQDCGRLTCSPSHYFGDLAIEGQLHQYDIRYRNKLKRRGKKL RKPPAYGEWIEATIKEIGIQRKGGDYYVYLLPVISHDSTNFQA AKKFFAAA APKQEIIDALPDEMKV LALDLNISNPVCSVKARIGRALDGPLSVLDY GQGEI VDGPKILVTDGPRCVALIELHSDVFFVKDAIREYKRTHGVAFNGAGMDDE MLERLSSIPSPSPSVRCQIQAWVAAIGKRLHLLHYGMRREGYKNLGEAIRL LAAMDAYDSMLQSYQRIHLKKGQSLPRQAKNDHSRAGLRETITRRIAYAI AQEARDCNVVFVEDLKADFDADNDNALLRLFAVKTLIRYIQEALEKIGV GLVLVAKDGTSKTNPMNGDLGWRDRNNKEHLYFELNGEAVRCHADLVA SLNVLLRGLDHSVCPYSFVTKDTNKKPGRADQKDKGYGKRLKRFFKDRFG SSNLRFVAKPDGEVTATTRRPKDALIGERVYYHGGKLV TSEVHYRMENEI KHLLTLKSPAKELSLTEDGIRSYENFTHV
48	EpiCas050	MVTPESQSDAKKAFADALFAGQEAA YGSFYAKYTPLPFPRKDSALGRLTA ELGTIGDEKYAKALKLNGLDPDEWAGYRNRAIAGEPPDALWEEIKVKVGV TNKATLASRGVGNLPTLRPDSKSYATALAYMLSSSLSSWEECAKLHRARR EEIAAKAAVPDSMTEEYALVCQLAQALEEAGSGLTAKMLNGAIREMAKT

		DSVYMVAITELADPKYKPLLALDKEDRARLMDTWQAERKLSTSLAAPKRP LFDKDYQIPFGMTGRGESFALRTDDSGRLWVKIGDHEEFSCAASDYFGNL KVEALPSDKGYKITYRKMVKNRRRQRAGYKSSVGGPLIEGEVKQILLQRR MNRRGKESFYVYLPHSIAVKDDRRHLSWFFKRADWKDGKGLEELPDQLIV SGIDLNIDQPVTVTKAIVSKNYTEGEINCPYGGGHLTGKPRSVCPDTPISQE AKLTHRKMRLLEIVRESKACHKAGTPLSSEGSERLKILYADISKPPCRNNR ESIQVVYSSLKKTYYRLKYELMQQGANDVSGCIRMLQLDDTHRALTACFD AFSGKPALAGRNFVGVARKKTSKGTARLNFREHVSHLYANAVLRACHGSD VVFIEDLSTSFdGNAHNSLKRLFAPGQLIDAITSLLESHGIAVVTVDPSGTSR RHPITGDIWYRDSMKKRELVLIDGELVSVNDDLAASINVLTGLNHSVVP HKFKVDPsGRVNVKEDGGKRLKVFLREKGLSGRLPVGSADAASKKAKA LTVYPSSSGLLSDRQVFEVLEQIKHLLAAGCKPRPVELMAGPGETYAAFSL KEVLEWSKSKSSEDLRHTGLGDETG
49	EpiCas051	MSNIKTTNILIPKQQDNYFEIENMLDRQNKIANDVFKLLLTyQGHILDQHFF GEDSKEINFIQKRQELYLVAKQADPLVTEDVFNLTCDGFVNQIASKSNMSN DLFKAFEVAYEELSEMKSyVAGKSEAKKLACKLKQWGVLPsFDNHIGEIK SGLSNAFGRFKSWYECDKMTQENFQKHVDSKSAYLEEMGDVISDRFQEFa NELGNVYNsKGLDNFDYFGGKFLLFYEQILKPSFEKGEVCYSGSIEFSNGK KQQYSIYGGIIDLlFKYKELWETKSWSKKVKRKIETVEQCTFSRFYNIFEDF KYLRKNKQRAGFREFSSTSSASICLSNSNIKYDITQVDGDRNNFYLTINGK QIQCRYRTKKIKYNKTTVEKTVENKCYLDNLEIEKQIKNDKPTGNYIFKFD VNKKRSNIATVNEPKIFKNRKGiyISLCMNIDLDDGKYDENKNIAFYlKTK YFNTYTKEADTLPLLKNKTVRIMGVDTGIKPIYTySILDYKNGKITDPNIHG KLEGKYEFTNYFNMKSLKNVIRDLPKYFQTDDHKELFKSHLFNDLTfTRDE LFDELKNFNtLREYKSRKNKGCLIRREIKSYRHKIGRMKDTNFKSDVKDIT NVFGIVKYIDSYKQVCSSWNYIGETVETKTQPNFAFADSHLAKYRENlKD NMIKTLVFNIIDEAKKHdVDILAIEDFSSEYSTSTKRDNVLKALFAPSQIQSEI ERCAKKEGIALMTVTRHYTSQVVFGKDIFGYRPKENDKTKLKWIENGEIKF VDADINAAKNIAAKLALKDVSPYTLDieGDGKYFKSSKWDKSKVFKKVND KFVNVSdQPKNTKTKMKIFLNGDNYITKEKKNELVEEIKNYKIKT

50	EpiCas052	MKSIKSIKSIKTKVVKNNELKLIELSTWCSSICEQLERYIFILGGKQIHDRDG VVVLGDGAVERKIYCKKDKSLIAACEVVYKHFTDKSSKSRTFGSWFLGGKS EGDNTNKGRKSTKEKTEKQIAKQIADKKELTDSLQLLWDKKLLPFPIDNKG YDFINTPRAKSYKWAITKTIHAKIKSYNEQCVETKKEYDALNAEINTYKTI FSGYSEKDIDDLQKFVDICEANNHRINYKFISFLKRKDLNFDEQTGKYRKE GKWIQHKNNGKEVKSKYSMKDEIVEALYKYKSLTKNDVSVLCNEHQKEDE MGKVVHYNMKRYSDLLFRKKNKKEIPSYTKISLATSKIELGLNNVKYNVE QVEDKLIWTICDQTGKDIQFVTVYTRKKEDNRTNGKKGAGFYKGKHHQL EDLKIVPVGDIGTYDISFKVNGKRPFTGTLKEPNIICRGGKVVFVQMPININID KTLNDARKKVLAYAYRETYSGSVNGKKQKMIKIENSKIAESLKS LGRDAIVL GVDLGLRGLATAVVSHNGKNETVKSSQYIKGDIVEWEKYRVFNDNIREVK KYIFLTKKSYTATTEEYTEFYKECSKPEQDYLDLSTKYDKDNVKNLNEKYT KNAWSVSKMFEDVSKMFETLKQDRLKYDIFNMPYWAASLKNYMSLMK SYNYVGVDIKVSKEYMSKYQSLYNNIKEDYAKKIGSYIVQLAVAKNCDIIV LEELKSNLGSVDRKSKRDNEMSLMWNCGRIKTHVENMAKDYGMFIDEVP EYGTSQVYHKTGNYGYRDEDNREIFWYEDNKDVAYIHADENAAINIAKRF LSQHTDNSSFSVILKGDAYYLNIA SNSKRMRAAALKTFGDLNKPFKINAND KNGNLYKKTRIFKSDSRWIGVNDKDLYIEHIKSLRNLRVRQ
51	EpiCas053	MNHFTQNLAVTKIDSSVQTLYDDFQEDFRKWYSILCSWAKVPYRVEDDN GNITHVHAKQRAAELSSLYSAQEQLLIERFMRVFLGASKEASSYS LFYDK YAPNGAKKTAALKEFLGEIPTLDDSQVEKKANTNKNCKRITVKEWRDFAA QFNKLGIDAIWGLKEKITSPDRGVGLQVRGFVSECKHPENAGYCRQLC QKLTEQVKSTVRICELYIQUETADLQSEIGNKTKQPEVKKCFDLLCQLSDEL DVNSGLSGRVLFQSIKDIQGGKDS DRKNKVCRTVMSDDKYLCLIDSDKKIV QRAFQAWKLSITLKKRQEKLYPKQRSHQNDVPMFGLTGLASFVSPDGS KFTFSEGSVETKPSLYFDGLRFQKNKHSFDMYRVAYASRRTKKGRRPKR VGR LRKATLKQVELFNKGGKFVIKLPHGIELPEHRAKLATFLMCADPKEK GYSKGGLVDEFRALAIDVNLSNPLAACVADVSKHGDGPIQALDYGRAKLC VDPFFLVANTPVCQGQLIKLCKECKAVVQAIKEFKECQAAGSDFS FETDLFLP AGHSPRHRIQSKVQELSRVLRDWKFKLQNTNQVAEMIRLLELMDAYQSMI

		SSYQRIHLKAGQTLPPQKKNSTKRANYRKDISRRLAFVVADYAQSLNVDV VFGEDLKADFDRDKNTNSLSRLFAGGMFLQHLENALAIRGIAFVKCAKNG TSITDPVSGVFGWRDKFNKERLWIERDGELGWVNSDHAACLNVLVGLM HSVVPYRVYVKAGRKTDDVEVFEEDHPIGPRKKRFFELCFGENKPKFEVV ENRVILDNAGTYQGFIYAHCLDAMQSEKQHWAEYDRIKELALAETNPKQF GLTPKKGEGYRNFCQPDIEHQVEVVGNGKRVAAATKQILESLADEKIEAT
52	EpiCas054	MSQEGEVKFKSNSDMKDRGVVTGYDYSLINAFCSKYTNLWEERDDYDST YLDKGGRKRQACIMTDRIYYISHMIKRLKYMTKKHVNPLFEQGKKAPCQ LFGKNYMSYTLRQDEETGTFFESFTDPTSESEGRAVKKFNASLAKNGYLNN LKIVESSNKDGVYKVSFNTAVENKDTLTIQTTYTYGYIKEPRIRFNLEKNVTY LDLCISGIESDNDTFNKILSEEQKENKEAIFKGFSSYFKASYAKGRDGSQKG SLDVDVFRAMGVDLGLNPLAALSIYEVKKSDKGEDIGGQVKIKHLTDHK SSDINKANKDCRVKKAKRTLKFIEGVKGLISVANHHHCNYLEDVEATGEGK EVNVGYMEMLCGDVIENHTPFKDLEDYMKYIENNLGYVDRKEWKKKTSD WVVP SLIQQAKKKISEMRREYFDAEDKHYYHTTKVNHVCNLQLHKIQAVKE MISTLNSFSTLGLNDEEKQAHADNGRPCADLWEYINGLRNYLSKGIASIIIN HALRNSVNIIFIEDLDMKASSFDSKEENKLKSLWSASQIMDYLEKFAAKHTI AVQEVNKFMTSQFSSENGTLGVRGASGNKSQFLYEDKGEVCDVDS DINAA RNIAIRGITRHSNMPMFVKVASDKEFILEMPKAAGMQKLSSMMKYLDSD NPRYVKFVKANNGTLKASVLEKEEFNKLKKQAIAWKKDNKDADSEVSILR SGKHFYIRKELEAIVKAKLCPEVEEETVKAK
53	EpiCas055	MATRTSHKSSKIRILPTGTTKKALVEYNLLFHDQVAKIMDLLTMRGVVYV LRDKDGEHEITATQIREAQWKLINSGRDSGDCIRLSQSIIDMMAQIYGGIME TGNPTKYRANLYLKKTNSGNEIGQGRSSASFMLFKDVSRGLSRPEFTRQIK DKNGKIKKPF AWATEERLMELYDAVRNDEMQUIEVEAERKSLEPDVTMEN LVKLRKV GILPLPLPGGHLYQGRGLGKLLAMPFDAAVGRLKSHYQANLAA VEARAALDKVVADKKDNVPAPLYALFLEFAEECRQNGFGVSKRFAKKLH GLNIPKTD FEPGDPFIVGLEIHKQYKYRPLWPQIGTLAGVYLAGVKAECTKG YAAITIADAYRSPALIRFGITNYARFEAEVNDQQGIDFSLQIPVDDQWTTVT VSGMPSNYFQNP ELIRQTSNYALSFTKGNRQQRRMGVILKEPLLTVRYSSD

		PEQEPRFYVHLAYNYDKTSYRMIGLDKIRAFFSSNLSKGAKLPHGTPYGLR TLSIDLGIRPIISSHLMELVSPHDIEAAQESRNGIEPIEVPDMGFARLVDSREM GGVWNSGLVGRMIDASRRCKSARHMILVSKLMRDGKPLNTHQQFLIEKY VKMYGVFLPPDDWARKDIIGEFIRQILADYRVIKDTMAKRNHPDHKGKFP DIICSDVFRMLFLIKDVISMLKSWARYKRKTGDKEIQGQNPTELNQYWRY NNLKDDLLKKIVHSLVRYADEHGVHLLLEDITGNRDSTKHRKENSMAFI WSPTLIRTRLEEEFEACGMLVQAINPSFTSQTHYMTGEFGFRDPRQKECLYF KREGRLYLIDSDGNSSMGLQVRFWTRHRDLWTLKTVKARRDRAAKLLKL DPKRAECTFVAKPTFEYAKSYLKRNLASPVAAFVDDNGNFILHKLGVNA ANQLTEGQERGETFFRHGTDWITHASHKKRMSAIEAMVEKVKKKGDKMT EADLVRLKV
54	EpiCas056	MSSDVVRPYNTKLLPDNRKHNMVVDTMNWCDNACSVFFDLFVGLYGGV KIGDVTNFLTTEKQDTTDDADLLCAINWFRAMPVSNPDVVGPEQLVVRFR DYSGHNPSPMASMYLSAPVNPRTWVDCRFLYQQMCQTLGANLDVDIST LVQTGLLPIQKDTQLNAYSFCNSVFGGKEDPSAKVTSYTAANALSGA TNPIITYEDYRQILSSAITNRPSRKGAKDILAADLDSTKTGVLPGNYVSQRIA MYQASAASKARGVVLNRLRVKAYLQNLVGLNYICTTWAAPMLRAVGDI KSKNSNNYRFVGERLERQQEIDNLLAVSNIRSAQAILQQFRANESNQFYIKK RHIGKGLPSLYDMWDLMDKEGIDNYIQSLKGDKTERLPIRELLQIYPHRK SVSAELFLSAAELTDVEDRQNRRCMVHPTVRGSLVINFGKSSTIFGAITPPDT LIKFRGEMVPAGQTGMIWFTMHLVDGGKWVKHHIPCHSSRYFEEIYAWK KDLPTLPFRRSLYGYAIGNKISDTRKIDLRKKASKQYLRTLENMTHNVE FDPTTQFMVDQELNVTITARVKKIPPVRDVPVGSRLMGLNKNQTIRDCYSI WERVEANTSGSFQHPVVQDVWLRLVADGHVTSTVNGPNGQIDQMSYSGP DVANCGNWMQERSNFITSVDDADQSIMKHFLWNNRVKLYAWNCVYAGT LMKAISKQMTTANRNLFRNEILSFVGGFRGCRGLGSLAQVSLDLLRRMKSVI SCYFSRNNAFTIEERSAFDKELMELFESLDNKMNNKRKEKVNRTLSSIFQV AKQHNVVRLVVTKNLQTPDANVRSMNQSVDWCAKIEEKIVDLGNVI GIYVDLVEALDLSHLDPFYIYAKGDVGKECQFDEVESSLLPIHTSMFRDWH GLVNBKGRTTTTPIYHSALRSFASHYGLRFDDLARIKAAELAKMLPDQKVFI

		HRGGRFFMSTYPVTSTSRPVSVDGRQRFLNSGDMVAPVNTMLRGFRARS
55	EpiCas057	MGITMSDITFRPYKSKLNPNERKMPMVVDTMNWCDNACSVFFDLFVGLY GGVKIGDVTNFLTQKQDTTDDADLLCAINWFRAMPVSNPDVAVGPEQLVV RFRDYSGHNPSPMASMYLSAPVNPRTWVDCRFLYQQMCQTLGANLDD ISTLVQTGLLPKQKDTQLNAYSFCSNVFGEVGEDPSAKVTSYTAANALS GATNPIITYEDYRQILSSAITNRPSPKGAKDILAADLDSTKTGVLPNGYVSQ RIAMYQASAASKARGVVLNRLRVKAYLQNLVGLNYICTTWAAPMLRAV GDIKSKNSNNYRFVGERLERQQEIDNLLAVSNIRSAQAILQQFRANESNQFY IKKRHIGKGLPSLYDMWDLMDKEGIDNYIQSLKGDKTERLPIRELLQIYYP HRKSVSAELFLSAAELTDVEDRQNRMCVHPTVRGSLVINFGKSSTIFGAITP PDTLIKFRGEMVPAGQTGMIWFTMHLVDGGKWVKHHIPCHSSRYFEEIYA WKKDLPTLPFRRSLYGYAIGNKISDTRKIDLRKKASKQYLRTLENMTHN VEFDPTTQFMVDQELNVTITARVKKIPPVRDVPVGSRLMGLNKNQTIRDCY SIWERVEANTSGSFQHPVVQDVWLRLVADGHVTSTVNGPNGQIDQMSYS GPDVANCGNWMQERSNFITSVDDADQSIMKHFLWNNRVKLYAWNCVYA GTLMKASKQMTTANRNLFNEILSFVGGFRFGCRLGSLAQVSLDLLRRMKS VISCYFSRNNAFTEERSAFDKELMELFESLDNKMNNKRKEKVNRTLSSIFQ VAKQHNVVRLVVTKNLQTPDANVRSMNQRSVDWCARKIEEKIVDLGNV IGIYVDLVEALDLSHLDPFIIYIAKGDVGKECQFDEVESSLLPIHTSMFRDW HGLVNKGRTTTPYHSALRSFASHYGLRFDDLARIKAAELAKMLPDQKVFI PHRGGRFFMSTYPVTSTSRPVSVDGRQRFLNSGDMVAPVNTMLRGFRARS S
56	EpiCas058	MGITMSDITFRPYKSKLNPNERKMPMARNKKLEGLVKSYKFQAKNVDI MIDILEEYSEYYNRVSNWINDNLMTLTIGDIALKCNTSQSVGYMQIALSDD YKQLPMYKLFEEKYPSTHCLNLLYAYITNNNIDNYTGNWGLSTTNRYKY GYINLVINNYRSKFANLNFVVKRSKIDINSNVEEIEKQCLYERIRLGLIKIED WKKHIDSLKRQEINEKALHRFNILSDYFNKNGESINERQSLLTIEQLQKFNG CQMKNKSMRIQVQNFKLNEKPNSLGYILQIHLNKKLFNIELLGNRQIRKGN KDNYTTLVDIINNHHGIDIIFKLENNKLYVILTSTDITFNKTITTPNKICGIDINI KHALLNTSERDSMNIQGYVNLYKRVLENDNFVSLLSKEEYETFSNMASYL

		TFCPIEFNYLYSRISNDEYTEKERLFSKILIDIQKEYENKKGCDLEKNYIHSV NKLRAKIIAYHKLKTIYTQKKQEYDIKMGFVDDSTLSKDGMDKRRFESPFN NTDCAKQIKSKLDNVLQDIEGCRKNIITYVYKIFETNGYVTIGLENLDNSNF EYHKPLCSIKSLLNYHKLLNKTIEQVKENNSVADLIEKDYYIFTDENKNIID AKFSDKGFYKINKSWFINQLIKVLHFASIKNYFITLSNNGKTQIVLVPSEFTS QINSLTHKLPMIKNDKGENVRPTKFRVRKTQEKHYNGLNADINATLNIKYI VENEWRNVLCKSPNKKEYKDARGYNFPTLIATQKHYNVSKQLEKLGIE FIQ*
57	EpiCas059	MEETRTPEKIIACQQDRSAFFKLGLPQKTTSIYHLMRCSKRYIDNVSETIS NNIEKAAMTSDNHIFRSALKEEYKKIYREKFYPDVTGQPFDPKGPDNTHIV ERTVENLCGQYISALKGVMKVRKPKWQDDIFEDES DKVIAEMLMASGNE ADIHPKLEKKIASAKKKDKMDEWQKVLDRYDYIKSKTDIEWTYTTLLYDH IARECARSMFEDGNPKFCYTNGNKFQVDELWGSERVEY EYEKYLH CRTLEAYWNEENKYFFLENQKKSCKTYEARQIIDHFIENEDAVKAETEEY MAKVAKEFGENGCKFR TENILYLVNSVSINKKQDEERCPIFKVTDIPQYA EAGIDRIGECKYVIKLLFTVGERYGT TYMPEAFSGLYDRRTNQR YDLDR ACKGSYVYLP IYISDKDSRKWSDRLKNTFTMSAFSISERKTKKDGRVLCVN LPYTHIQHKPRNYFKVYPTYGVDINVKHS LFSTNATNLNGYIDVNGELIDF YNSIGRPDKLKNINLAVTLGDKGGTAGNHTN FLLVSDLGDWDDHGKKNR DRFVKREDGTYYSRPDLLADGLNFLKKKYS GDRKSYDYINTTIAFRNNQT RLRCIDKEWNRKRAYDEEHGKPEK GERSGFYPTEDAERLTMERQALRRE IAENTRQVIA YMLWLFQENG IKSIALEHLDCGSWKNFRDQLKPKWFCNNC KYWAEKYGIPYEATKY YRSNRKYFDIEDGELIPENVKLNALGEAENDKRY FQTSIAKAVGMSSIKQMFSIVSQNYDIQVNYTEQAYTSQTDSRNGTVDKDF RVSQECYIGSDGTKMNCDFNA AVNIADRISDDTLREEFNKELPMKEFVKY GATQIVSRNRRWNAENAERV TMAAFQKRKNADGGAHTAAVKPTSKTEES SLKLVAE
58	EpiCas060	MIDRNETHLSKKSTRSVEFKLAVPNRIAPVIDEMLDLSITMSNAISKDISEHL TDWVQRAEVTKEQKRPYDYLYASEKEGIRYWYKHSGLRDKMYFDFCSKY FGYPVLKNKNAHFLTRTIENVAAAYEGRFTRFQEETKIRATDQSQLTEEEE

	LLKKNIKEILNIMCSLKPGQTVDDQILVRRSKTGKNQEKYQKLLSFYVDVTS DEKSDKYINFDILFADM CYRNSATFILKTMEISVDDSGVVNFGMV SPLVEL PTKIGSSEEATMRHILENRSVRLINEVNDMIFELTVAQYIKSSCRYRFVDELT EKANSKDSVFHKNLLMYYTSHRDKVDEGYKQYLHSVNESFGKKQIVWKN STLKLN YEEKVEKDRQVPIKMISSIKQFADLGITSISGCKYVLRVHMILDDE KKFKDQHLFCDL YDKRTGRRFDLLDHETKGYAYIPIMVPENDRR TYDISDF DCQMTEISSRDGMYFAKLTRDIDYVKPVNRFGEQVCGVDINIRHAMIDVS ERSIMNSEKYVDVEGFII DYFRNVKNLPSVADDLVTNRYLGEAE EAGKHTF FGLMSSINKSISRETGGLKNRFVKNVDGTYTILPDCFSEPLNVLMKKYEQGS LQYRYVASVKALRVQLAQMHVLCMEHDDRRQVYDNL YGKPYIDDH PYE DPDFVY AHRKLHQKIETNTQFIVSYAIQLIQEAGFSCISVEQIDKKNLGKML NLKSAKV FYNGCIYASSENGQAIADTRYFRTNCRYFDYPSDLNLF DIDIHKL NGVGEAYNINAKLMNSLGKYVGFASF KHMFAHIGTAHDVQVV FVEPAFTS QVDSKTGELRADKKEMRPYHHTYKHTDGT VSHADENGAINAKKIVTEEIL RTRLLKENS VKDIMEYGGVQLERSGKSINGVRKVYEEVYAEKNGLV*
--	--

在某些实施方式中，所述 Cas 酶具有能够结合靶 DNA 链的催化活性结构域和/或切割所述靶 DNA 链的催化活性结构域。

在某些实施方式中，所述催化活性结构域包含一个或多个氨基酸的改变，从而使得所述 Cas 酶仅具有结合靶 DNA 链的活性，或者具有结合靶 DNA 链的活性和切割所述靶 DNA 单链的活性。

另一方面，本申请提供一种融合分子，所述融合分子包含本申请所述的 Cas 酶和一个或多个异源功能结构域。

在某些实施方式中，所述一个或多个异源功能结构域能够调控一种或多种基因产物的表达。

在某些实施方式中，所述一个或多个异源功能结构域直接或间接地融合在所述 Cas 酶上。

在某些实施方式中，所述一个或多个异源功能结构域选自自解旋酶、核酸酶、解旋酶-核酸酶、DNA 甲基转移酶、DNA 羟甲基化酶、组蛋白甲基化酶、组蛋白去甲基化酶、组蛋白乙酰转移酶、组蛋白去乙酰化酶、磷酸酶、激酶、转录(共)活化物、转录阻遏物、DNA 结合蛋白、DNA 结构蛋白、标志物蛋白、报告物蛋白、荧光蛋白、配体结合蛋白、信号肽、亚细胞定位序列、抗体表位和亲和纯化标签。

在某些实施方式中，所述一个或多个异源功能结构域具有以下活性中的一种或多种：甲基酶活性、脱甲基酶活性、脱氨酶活性、转录激活活性、转录阻抑活性、转录释放因子活性、逆转录酶活性、组蛋白修饰活性、RNA 切割活性和核酸结合活性。

另一方面，本申请提供一种工程化的、可编程的、非天然存在的 CRISPR-Cas 系统，所述系统包含本申请所述的 Cas 酶或本申请所述的融合分子，和一种或多种指导 RNA，所述一种或多种指导 RNA 在细胞中靶向编码一种或多种基因产物的核酸分子的基因座，从而指导所述 Cas 酶或所述融合分子结合和/或切割所述编码一种或多种基因产物的核酸分子的基因座；并且，所述 Cas 酶或所述融合分子与所述指导 RNA 不共同天然存在。

另一方面，本申请提供一种工程化的、非天然存在的载体系统，所述载体系统包含一种或多种载体，所述一种或多种载体包括：a) 第一调节元件，所述第一调节元件可操作地连接到一种或多种指导 RNA 上，所述一种或多种指导 RNA 能够与编码一种或多种基因产物的核酸分子的基因座中的靶序列杂交，和 b) 第二调节元件，所述第二调节元件可操作地连接到本申请所述的 Cas 酶或本申请所述的融合分子上，其中，所述组分 a) 和所述组分 b) 位于所述载体系统的相同或不同载体上，且所述指导 RNA 在细胞中靶向所述编码一种或多种基因产物的核酸分子的基因座，从而指导所述 Cas 酶或所述融合分子结合和/或切割所述编码一种或多种基因产物的核酸分子的基因座；并且，所述 Cas 酶或所述融合分子与所述指导 RNA 不共同天然存在。

在某些实施方式中，所述一种或多种基因产物的表达被改变。

在某些实施方式中，所述基因产物的表达被降低或者被增多。

在某些实施方式中，所述基因产物是一种蛋白质。

在某些实施方式中，所述细胞是真核细胞。

在某些实施方式中，所述真核细胞是哺乳动物细胞。例如，所述哺乳动物细胞包括但不限于鼠类、猴、人、农畜、体育用动物和宠物的细胞。

在某些实施方式中，所述哺乳动物细胞是人类细胞。

在某些实施方式中，所述 Cas 酶是经密码子优化的，用以在真核细胞中进行表达。

在某些实施方式中，所述指导 RNA 包含融合到 tracr 序列上的指导序列。

在某些实施方式中，所述指导 RNA 包含直接重复 (Direct repeat) 序列和间隔 (Spacer) 序列，其中所述间隔序列与所述指导 RNA 靶向的核酸分子结合。

在某些实施方式中，所述直接重复序列的长度为 10 个至 70 个核苷酸。

在某些实施方式中，所述直接重复序列的长度为 31 个至 36 个核苷酸。

在某些实施方式中，所述直接重复序列包含 SEQ ID NO: 63-88 和 90-99 中任一项所示的核苷酸序列，或者包含与 SEQ ID NO: 63-88 和 90-99 中任一项所示的核苷酸序列具有至少 95% 序列同一性的核苷酸序列。

在某些实施方式中，所述间隔序列的长度为 16 个至 24 个核苷酸。

在某些实施方式中，所述指导 RNA 靶向的核酸分子包含能够与所述间隔序列互补配对的核苷酸序列。

在某些实施方式中，所述系统的所述载体或所述 Cas 酶还包含一个或多个核定位序列 (NLS)。

在某些实施方式中，所述系统通过递送系统被引入所述细胞中，所述递送系统选自病毒粒子、脂质体、脂质纳米颗粒、电穿孔、显微注射和缀合。

另一方面，本申请提供改变一种或多种基因产物的表达的方法，所述方法包括向包含和表达编码所述一种或多种基因产物的核酸分子的细胞中引入一种工程化的、非天然存在的 CRISPR-Cas 系统，所述系统包含本申请所述的 Cas 酶或本申请所述的融合分子，和一种或多种指导 RNA，所述一种或多种指导 RNA 靶向所述编码一种或多种基因产物的核酸分子的基因座，从而指导所述 Cas 酶或所述融合分子结合和/或切割所述基因座，由此改变所述一种或多种基因产物的表达；并且，所述 Cas 酶或所述融合分子与所述指导 RNA 不共同天然存在。

另一方面，本申请提供改变一种或多种基因产物的表达的方法，所述方法包括向包含和表达编码所述一种或多种基因产物的核酸分子的细胞中引入一种工程化的、非天然存在的载体系统，所述载体系统包含一种或多种载体，所述一种或多种载体包括：a) 第一调节元件，所述第一调节元件可操作地连接到一种或多种指导 RNA 上，所述一种或多种指导一种或多种指导 RNA 能够与所述编码一种或多种基因产物的核酸分子的基因座中的靶序列杂交，和 b) 第二调节元件，所述第二调节元件可操作地连接到本申请所述的 Cas 酶或本申请所述的融合分子上，其中，所述组分 a) 和所述组分 b) 位于所述载体系统的相同或不同载体上，且所述指导 RNA 在所述细胞中靶向所述编码一种或多种基因产物的核酸分子的基因座，从而指导所述 Cas 酶或所述融合分子结合和/或切割所述基因座，由此改变所述一种或多种基因产物的表达；并且，所述 Cas 酶或所述融合分子与所述指导 RNA 不共同天然存在。

在某些实施方式中，所述基因产物的表达被降低或者被增多。

在某些实施方式中，所述基因产物是一种蛋白质。

在某些实施方式中，所述细胞是真核细胞。

在某些实施方式中，所述真核细胞是哺乳动物细胞。例如，所述哺乳动物细胞包括但不

限于鼠类、猴、人、农畜、体育用动物和宠物的细胞。

在某些实施方式中，所述哺乳动物细胞是人类细胞。

在某些实施方式中，所述 Cas 酶是经密码子优化的，用以在真核细胞中进行表达。

在某些实施方式中，所述指导 RNA 包含融合到 tracr 序列上的指导序列。

在某些实施方式中，所述指导 RNA 包含直接重复（Direct repeat）序列和间隔（Spacer）序列，其中所述间隔序列与所述指导 RNA 靶向的核酸分子结合。

在某些实施方式中，所述直接重复序列的长度为 10 个至 70 个核苷酸。

在某些实施方式中，所述直接重复序列的长度为 31 个至 36 个核苷酸。

在某些实施方式中，所述直接重复序列包含 SEQ ID NO: 63-88 和 90-99 中任一项所示的核苷酸序列，或者包含与 SEQ ID NO: 63-88 和 90-99 中任一项所示的核苷酸序列具有至少 95% 序列同一性的核苷酸序列。

在某些实施方式中，所述间隔序列的长度为 16 个至 24 个核苷酸。

在某些实施方式中，所述指导 RNA 靶向的核酸分子包含能够与所述间隔序列互补配对的核苷酸序列。

在某些实施方式中，所述系统的所述载体或所述 Cas 酶还包含一个或多个核定位序列（NLS）。

在某些实施方式中，所述方法包括通过递送系统将所述 CRISPR-Cas 系统或所述载体系统引入到所述细胞中，所述递送系统选自病毒粒子、脂质体、脂质纳米颗粒、电穿孔、显微注射和缀合。

另一方面，本申请提供编码本申请所述的 Cas 酶、本申请所述的融合分子或本申请所述的 CRISPR-Cas 系统的核酸。

另一方面，本申请提供一种细胞，所述细胞包含本申请所述的 Cas 酶、本申请所述的融合分子、本申请所述的 CRISPR-Cas 系统、本申请所述的载体系统和/或本申请所述的核酸。

另一方面，本申请提供一种试剂盒，所述试剂盒包含本申请所述的 Cas 酶、本申请所述的融合分子、本申请所述的 CRISPR-Cas 系统、本申请所述的载体系统、本申请所述的核酸和/或本申请所述的细胞。

在某些实施方式中，所述试剂盒还包含用于放置所述 Cas 酶、所述融合分子、所述 CRISPR-Cas 系统、所述载体系统、所述核酸和/或所述细胞的容器，以及用法说明书。

本领域技术人员能够从下文的详细描述中容易地洞察到本申请的其它方面和优势。下文的详细描述中仅显示和描述了本申请的示例性实施方式。如本领域技术人员将认识到的，本

申请的内容使得本领域技术人员能够对所公开的具体实施方式进行修改而不脱离本申请所涉及发明的精神和范围。相应地，本申请的附图和说明书中的描述仅仅是示例性的，而非为限制性的。

附图说明

本申请所涉及的发明的具体特征如所附权利要求书所显示。通过参考下文中详细描述的实施方式和附图能够更好地理解本申请所涉及发明的特点和优势。对附图简要说明如下：

图 1A-1B 显示的是本申请所述 Cas 酶在真核细胞内的切割活性的测定流程示意图和荧光结果以及 Cas 酶的 gRNA 结合所需要的 PAM 基序。

图 2A-2B 显示的是本申请所述 Cas 酶在真核细胞内的切割活性的测定流程示意图以及不同 Cas 酶的绿色荧光细胞对比结果。

图 3A-3B 显示的是本申请所述 Cas 酶的体外切割活性及其 PAM 序列的测定流程示意图以及测序结果。

图 4 显示的是在内源位点检测本申请所述 Cas 酶活性的位点信息以及扩增检测结果。

图 5A-5C 显示的是本申请所述 Cas 酶的不同变体在真核细胞内进行切割活性测定后的 GFP 荧光结果。

图 6A-6B 显示的是对本申请所述 Cas 酶结合不同 DR (Direct repeat) 区域优化方案处理过的 sgRNA 在真核细胞内进行切割活性测定后的绿色荧光细胞比例结果。

图 7 显示的是本申请所述 Cas 酶结合不同 Spacer 长度的 sgRNA 在真核细胞内进行切割活性测定后的绿色荧光细胞比例结果。

图 8A-8B 显示的是本申请所述 Cas 酶的不同变体在真核细胞内进行切割活性测定后的绿色荧光细胞比例结果。图 8C 显示的是本申请所述 Cas 酶的不同变体在不同 PAM 报告系统中的切割活性测定结果 (绿色荧光细胞比例)。

图 9A-9C 显示的是包含本申请所述 Cas 酶的胞嘧啶碱基编辑器的结构示意图，以及其在 EMX1 和 VEGFA 两种内源位点上检测碱基编辑效率的结果。

图 10 显示的是包含本申请所述 Cas 酶的基因激活表观工具在 CXCR4 内源位点上检测靶位点基因表达的激活效果。

具体实施方式

以下由特定的具体实施例说明本申请发明的实施方式，熟悉此技术的人士可由本说明书所公开的内容容易地了解本申请发明的其他优点及效果。

术语定义

在本申请中，术语“同一性”可与“同源性”互换地使用，其通常是指两个或多个多肽分子或者两个或多个核酸分子序列之间的关系，该关系通过比较它们的序列确定。在本领域中，“同一性”还指核酸分子或者多肽序列相关性的程度，这可以通过两个或多个核苷酸或两个或多个氨基酸的序列之间的匹配来确定。在本申请中，氨基酸序列的同一性百分比(%)被定义为在比对序列并在必要时引入缺口来达到最大百分比序列同一性，而不将任意保守取代视为序列同一性的部分之后，候选序列中与参考多肽序列中的氨基酸残基相同的氨基酸残基占残基总数的百分比。可以以本领域技术之内的多种方式来达到以测定百分比氨基酸序列同一性为目的的比对，例如，使用公开可得的计算机软件，如 BLAST、BLAST-2、ALIGN 或 Megalign(DNASTAR)软件。本领域技术人员可以确定用于比对的适当参数，包括在所比较的序列的全长内达到最大比对所需的任意算法。在某些实施方式中，多肽分子或核酸分子序列的同一性百分比(%)的计算还可以基于序列突变类型确定残基总数。突变类型包括在序列任一端或两端的插入(延伸)、在序列任一端或两端的缺失(截短)、一个或多个氨基酸/核苷酸的置换/替代、在序列内部的插入、在序列内部的缺失。举多肽的氨基酸序列为例(核苷酸序列同理)，如果突变类型为以下中的一种或多种：一个或多个氨基酸/核苷酸的置换/替代、在序列内部的插入和在序列内部的缺失，则残基总数以比较的分子中较大者来计算。如果突变类型还包括在序列任一端或两端的插入(延伸)或在序列任一端或两端的缺失(截短)或在序列内部的插入和在序列内部的缺失，则在任一端或两端或内部插入或缺失的氨基酸的数量(例如，在两端插入或缺失的数量小于 20 个)并不计入残基总数中。在计算同一性百分比时，可将正在比较的序列以产生序列之间最大匹配的方式比对，并通过特定算法解决比对中的空位(如果存在的话)。

在本申请中，“催化活性结构域”是指 Cas 蛋白(酶)内展现显著二级结构含量的可鉴别或可确定的保守结构实体且该保守结构为 Cas 蛋白(酶)实现例如结合和/或切割多核苷酸功能的区域。示例性的催化活性结构域可以是 Cas9 家族的蛋白酶，其具有两个催化活性结构域，一个是类 HNH 的，它的作用是切断与指导 RNA 配对的单链多核苷酸(目标链)，另一个结构域是类 RuvC 的，它的作用是切断目标链的互补链。

在本申请中，术语“结合”(例如，关于多肽或蛋白酶的靶 DNA 结合(催化活性)结构域)通常是指大分子之间(例如，蛋白质与核酸之间)的非共价相互作用。当在非共价相互作用的状态下，大分子被称作“缔合”或“相互作用”或“结合”(例如，当分子 X 被称作与分子 Y 相互作用时，意指分子 X 以非共价方式结合分子 Y)。应了解，不是所有的结合相互作用组分都需要为序列特异性的(例如，与 DNA 骨架中的磷酸酯残基接触)，但结合相互作用的一些部分可为序列特异性的。

在申请中，术语“融合分子”通常是指至少两个部分组成的分子(bipartite molecule)，其包含本申请的酶(蛋白质或肽)，该酶与至少一个其他部分偶联，从而形成单个实体。该酶和该至少一个其他部分可以由接头分开，或者可以直接偶联。该至少一个其他部分可以在 N 端、C 端或末端氨基酸外的任意氨基酸处与本申请的酶融合。其他部分可以与已经包含在该融合分子中的部分融合。本领域的技术人员充分了解用于确定本申请的融合分子中的部分的测定最佳顺序和/或组合。通常，当融合分子包含本申请的酶和至少一种其他肽时，该术语不包括这样的融合分子，在该融合分子中，融合产生天然存在的肽。

在本申请中，术语“异源”通常是指分别不存在于天然核酸或蛋白质中的核苷酸或多肽序列。在本申请的某些实施方式中，术语“异源功能结构域”可指除了本申请所述的 Cas 酶之外的约或多于约 1 个、2 个、3 个、4 个、5 个、6 个、7 个、8 个、9 个、10 个或更多个结构域或本申请所述的融合分子的一部分。可以包含在本申请所述的融合分子中或者可以与本申请所述的 Cas 酶融合的异源功能结构域的实例包括但不限于，表位标签、报告基因序列、以及具有下列活性的一个或多个的蛋白质结构域：甲基酶活性、脱甲基酶活性、转录激活活性、转录阻抑活性、转录释放因子活性、组蛋白修饰活性、RNA 切割活性和核酸结合活性。表位标签的非限制性实例包括组氨酸(His)标签、V5 标签、FLAG 标签、流感病毒血凝素(HA)标签、Myc 标签、VSV-G 标签、和硫氧还蛋白(Trx)标签。报告基因的实例包括，但不限于，谷胱甘肽-S-转移酶(GST)、辣根过氧化物酶(HRP)、氯霉素乙酰转移酶(CAT)、 β -半乳糖苷酶、 β -葡萄糖醛酸糖苷酶、萤光素酶、绿色荧光蛋白(GFP)、HcRed、DsRed、青荧光蛋白(CFP)、黄色荧光蛋白(YFP)、以包括蓝色荧光蛋白(BFP)的自发荧光蛋白。Cas 酶可以融合到编码一种蛋白质或蛋白质片段的基因序列上，所述蛋白质或蛋白质片段结合 DNA 分子或结合其他细胞分子，其包括，但不限于，麦芽糖结合蛋白(MBP)、S-tag、Lex A DNA 结合结构域(DBD)融合物、GAL4 DNA 结合结构域融合物、以及单纯疱疹病毒(HSV)BP16 蛋白融合物。可以形成包含 Cas 酶的融合分子的一部分的另外的结构域描述于 US 20110059502 中，通过引用将其并入本文。

在本申请中，术语“表达”通常是指以此从 DNA 模板转录成多核苷酸(如转录成 mRNA 或其他 RNA 转录物)的过程和/或转录的 mRNA 随后以此翻译成肽、多肽或蛋白质的过程。转录物和编码的多肽可以总称为“基因产物”。如果多核苷酸来源于基因组 DNA，表达可以包括真核细胞中 mRNA 的剪接。

在本申请中，术语“多核苷酸”、“核苷酸”、“核苷酸序列”、“核酸”和“寡核苷酸”可互换地使用。它们通常是指具有任何长度的核苷酸的聚合形式，是脱氧核糖核苷酸或核糖核苷酸、或其类似物。多核苷酸可具有任何三维结构，并且可以执行已知或未知的任何功能。以下是多核苷酸的非限制性实例：基因或基因片段的编码区或非编码区、根据连接分析定义多个座位(一个座位)、外显子、内含子、信使 RNA(mRNA)、转运 RNA、核糖体 RNA、短发夹 RNA(shRNA)、micro-RNA(miRNA)、核酶、cDNA、重组多核苷酸、分支多核苷酸、质粒、载体、任何序列的分离的 DNA、任何序列的分离的 RNA、核酸探针、和引物。多核苷酸可以包含一个或多个经修饰的核苷酸，如甲基化的核苷酸和核苷酸类似物。如果存在，可以在聚合物组装之前或之后进行核苷酸结构的修饰。核苷酸的序列可以被非核苷酸组分中断。多核苷酸可以在聚合后，如通过与标记的组分缀合来进一步修饰。

在本申请中，术语“非天然存在的”和“工程化的”可互换地使用。当它们意指核酸分子、多肽、或其组合和其系统时，通常表示该核酸分子或多肽至少基本上从它们在自然界中或如发现于自然界中的与其结合的至少另一种组分游离出来。

在本申请中，术语“载体”通常是指一种核酸分子，它能够运送与其连接的另一种核酸分子。载体包括但不限于，单链、双链、或部分双链的核酸分子；包括一个或多个自由端、无自由端(例如环状的)的核酸分子；包括 DNA、RNA、或两者的核酸分子；以及本领域已知的其他多种多样的多核苷酸。一种类型的载体是“质粒”，其是指其中可以例如通过标准分子克隆技术插入另外的 DNA 片段的环状双链 DNA 环。另一种类型的载体是病毒载体，其中病毒衍生的 DNA 或 RNA 序列存在于用于包装病毒(例如，逆转录病毒、复制缺陷型逆转录病毒、腺病毒、复制缺陷型腺病毒、以及腺相关病毒)的载体中。病毒载体还包含由用于转染到一种宿主细胞中的病毒携带的多核苷酸。某些载体(例如，具有细菌复制起点的细菌载体和附加型哺乳动物载体)能够在它们被导入的宿主细胞中自主复制。其他载体(例如，非附加型哺乳动物载体)在引入宿主细胞后整合到该宿主细胞的基因组中，并且由此与该宿主基因组一起复制。而且，某些载体能够指导它们可操作连接的基因的表达。这样的载体在此被称为“表达载体”。在重组 DNA 技术中使用的普通表达载体通常是质粒形式。重组的表达载体可包含处于适合于在宿主细胞中的核酸表达的形式之本申请的核酸，这意味着这些重组表达载体包含基

于待用于表达的宿主细胞而选择的一种或多种调节元件，所述调节元件可操作地连接至待表达的核酸序列。在重组表达载体内，“可操作地连接”旨在表示感兴趣的核苷酸序列以一种允许该核苷酸序列的表达的方式被连接至该一种或多种调节元件(例如，处于一种体外转录/翻译系统中或当该载体被引入到宿主细胞中时，处于该宿主细胞中)。

在本申请中，术语“调节元件”通常旨在包括启动子、增强子、内部核糖体进入位点(IRES)、和其他表达控制元件(例如转录终止信号，如多聚腺苷酸化信号和多聚 U 序列)。这样的调节序列例如描述于戈德尔(Goeddel)，《基因表达技术：酶学方法》(GENE EXPRESSION TECHNOLOGY:METHODS IN ENZYMOLOGY)185，学术出版社(Academic Press)，圣地亚哥，加利福尼亚州(1990)中。在某些实施方式中，调节元件可包括指导一个核苷酸序列在许多类型的宿主细胞中的组成型表达的那些序列以及指导该核苷酸序列只在某些宿主细胞中表达的那些序列(例如，组织特异型调节序列)。组织特异型启动子可主要指导在感兴趣的期望组织中的表达，所述组织例如肌肉、神经元、骨、皮肤、血液、特定的器官(例如肝脏、胰腺)、或特殊的细胞类型(例如淋巴细胞)。调节元件还可以时序依赖性方式(如以细胞周期依赖性 or 发育阶段依赖性方式)指导表达，该方式可以是或者可以不是组织或细胞类型特异性的。在某些实施方式中，一个载体可包含一个或多个 pol III 启动子(例如 1、2、3、4、5、或更多个 pol III 启动子)、一个或多个 pol II 启动子(例如 1、2、3、4、5、或更多个 pol II 启动子)、一个或多个 pol I 启动子(例如 1、2、3、4、5、或更多个 pol I 启动子)、或其组合。pol III 启动子的实例包括但不限于 U6 和 H1 启动子。pol II 启动子的实例包括但不限于逆转录劳斯肉瘤病毒(RSV)LTR 启动子(任选地具有 RSV 增强子)、巨细胞病毒(CMV)启动子(任选地具有 CMV 增强子)、SV40 启动子、二氢叶酸还原酶启动子、 β -肌动蛋白启动子、磷酸甘油激酶(PGK)启动子、和 EF1 α 启动子。术语“调节元件”还可以涵盖的是增强子元件，如 WPRE、CMV 增强子、在 HTLV-I 的 LTR 中的 R-U5' 片段、SV40 增强子；以及在兔 β -珠蛋白的外显子 2 与 3 之间的内含子序列(《美国国家科学院院刊》(Proc.Natl.Acad.Sci.USA.)，第 78(3)卷，第 1527-31 页，1981)。

本领域技术人员将理解，表达载体的设计可取决于诸如待转化的宿主细胞的选择、所希望的表达水平等因素。一种载体可以被引入到宿主细胞中而由此产生转录物、蛋白质、或肽，包括由如本申请所述的核酸编码的融合分子或酶(例如，规律间隔成簇短回文重复序列(CRISPR)转录物、蛋白质、酶、其突变体形式、其融合分子或融合蛋白，等等)。有利的载体包括慢病毒以及腺相关病毒，并且也可选择此类型的载体以靶向具体类型的细胞。

在本申请中，术语“密码子优化”通常是指通过用在宿主细胞的基因中更频繁地或者最频繁地使用的密码子代替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 个、或更多个密码子同时维持该天然氨基酸序列而修饰一个核酸序列以便增强在感兴趣宿主细胞中的表达的方法。不同的物种对于具有特定氨基酸的某些密码子展示出特定的偏好。密码子偏好性(在生物之间的密码子使用的差异)经常与信使 RNA(mRNA)的翻译效率相关,而该翻译效率则被认为依赖于(除其他之外)被翻译的密码子的性质和特定的转运 RNA(tRNA)分子的可用性。细胞内选定的 tRNA 的优势一般反映了最频繁用于肽合成的密码子。因此,可以将基因定制为基于密码子优化在给定生物中的最佳基因表达。密码子利用率表可以容易地获得,例如在密码子使用数据库(“Codon Usage Database”)中,并且这些表可以通过不同的方式调整适用。例如,参见中村 Y.(Nakamura Y.)等人, Codon usage tabulated from the international DNA sequence databases: status for the year 2000,《核酸研究》(Nucl.Acids Res.)28:292(2000 年)。用于密码子优化特定的序列以便在特定的宿主细胞中表达的计算机算法也是可获得的,如 Gene Forge(Aptagen 公司, Jacobus, PA),也是可获得的。在某些实施方式中,在编码 Cas 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 个、或更多个、或所有密码子)相对应于对于特定氨基酸最频繁使用的密码子。

在本申请中,术语“指导 RNA”与“guide RNA”、“gRNA”可互换地使用,其通常是指促进将 RNA 指导的核酸酶或其他效应分子(通常与 gRNA 分子复合)特异性指导到靶序列上的一组核酸分子。在自然界中,crRNA 和 tracrRNA 通常作为两个独立的 RNA 分子存在,组成 gRNA。术语“tracrRNA”通常是指可与 Cas 核酸酶结合的支架型 RNA,术语“crRNA”也称为 CRISPR RNA,通常是指与所靶向的目标 DNA 互补的一段核苷酸序列。crRNA 和 tracrRNA 也可以融合成为单链,此时 gRNA 也可称为单链指导 RNA (sgRNA),sgRNA 已成为本领域技术人员在 CRISPR 技术中使用的 gRNA 的最常见的形式,因此术语“sgRNA”和“gRNA”在本文中可具有相同的含义。sgRNA 可以人工合成,也可以在体外或体内由 DNA 模板制备。sgRNA 可以结合 Cas 核酸酶,也可以靶向目标 DNA,其可引导 Cas 核酸酶切割与 gRNA 互补的 DNA 位点。

如本申请所使用的,crRNA 通常包含介导靶识别的间隔(Spacer)序列和与 CRISPR-Cas 效应蛋白形成复合物的直接重复序列(本文中也称为“Direct repeat”或“DR 序列”)。在某些情况下,间隔序列(也称导向序列)是与靶序列具有足够互补性从而与所述靶序列杂交并引导 CRISPR-Cas 系统复合物与所述靶序列的特异性结合的任何多核苷酸序列。在某些实施方案中,当最佳比对时,间隔序列与其相应靶序列之间的互补程度为至少 50%、至少 60%、至少 70%、至少 80%、至少 90%、至少 95%、至少 96%、至少 97%、至少 98%或至少 99%。确定最佳比对在本领域的普通技术人员的能力范围内。例如,存在公开和可商购的比对算法和程

序,例如但不限于 ClustalW、Matlab 中的 Smith-Waterman 算法、Bowtie、Geneious、Biopython 以及 SeqMan。

在某些实施方案中,所述间隔序列在长度上为至少 5 个、至少 10 个、至少 15 个、至少 16 个、至少 17 个、至少 18 个、至少 19 个、至少 20 个、至少 21 个、至少 22 个、至少 23 个、至少 24 个、至少 25 个、至少 26 个、至少 27 个、至少 28 个、至少 29 个、至少 30 个、至少 35 个、至少 40 个、至少 45 个或至少 50 个核苷酸。在某些实施方案中,所述间隔序列在长度上为不超过 50 个、45 个、40 个、35 个、30 个、25 个、24 个、23 个、22 个、21 个、20 个、15 个、10 个或更少个核苷酸。在某些实施方案中,所述间隔序列在长度上为 10-30 个、15-25 个、15-22 个、16-24 个、19-25 个、或 19-22 个核苷酸。在某些优选的实施方案中,所述间隔序列的长度为 20 个核苷酸。

在某些实施方案中,所述直接重复序列在长度上为至少 10 个、至少 15 个、至少 16 个、至少 17 个、至少 18 个、至少 19 个、至少 20 个、至少 21 个、至少 22 个、至少 23 个、至少 24 个、至少 25 个、至少 26 个、至少 27 个、至少 28 个、至少 29 个、至少 30 个、至少 31 个、至少 32 个、至少 33 个、至少 34 个、至少 35 个、至少 40 个、至少 45 个、至少 50 个、至少 55 个、至少 56 个、至少 57 个、至少 58 个、至少 59 个、至少 60 个、至少 61 个、至少 62 个、至少 63 个、至少 64 个、至少 65 个或至少 70 个核苷酸。在某些实施方案中,所述直接重复序列在长度上为不超过 70 个、65 个、64 个、63 个、62 个、61 个、60 个、59 个、58 个、57 个、56 个、55 个、50 个、45 个、40 个、35 个、34 个、33 个、32 个、31 个、30 个、29 个、28 个、27 个、26 个、25 个、24 个、23 个、22 个、21 个、20 个、15 个、10 个或更少个核苷酸。在某些实施方案中,所述直接重复序列在长度上为 55-70 个核苷酸,例如 55-65 个核苷酸,例如 60-65 个核苷酸,例如 62-65 个核苷酸,例如 63-64 个核苷酸。在某些实施方案中,所述直接重复序列在长度上为 15-40 个核苷酸,例如 15-25 个核苷酸,例如 20-30 个核苷酸,例如 22-36 个核苷酸,例如 31 个核苷酸。

在本申请中,术语“包含”和“包括”可互换地使用,其通常是指包括明确指定的特征,但不排除其他要素。术语“至少”通常是指包含本数的情况。

在本申请中,术语“约”或“大约”通常由本领域普通技术人员确定的针对特定值的可接受误差范围内,其部分取决于怎样测量或确定该值,即,测量系统的限制。例如,“约”或“大约”可意指按照本领域的实践在 1 或超过 1 个标准偏差内。或者,“约”或“大约”可意指至多 10%或 20%的范围(即,±10%或±20%)。

在本申请中,术语“选自”通常是指包括选择的对象以及其所有组合。例如“选自(:)A、

B 和 C”意指包括 A、B 和 C 的所有组合，例如，A、B、C、A+B、A+C、B+C 或 A+B+C。

不欲被任何理论所限，下文中的实施例仅仅是为了阐释本申请的酶及其融合分子、方法和用途等，而不适用于限制本申请发明的范围。

实施例

实施例 1

测定本申请的 Cas 酶在真核细胞内的切割活性

本实施例采用带有蓝-绿光报告系统的稳转 293T 细胞测试本申请提供的 Cas 酶的活性。如图 1 所示，该报告系统带有持续表达的 CMV 启动子、编码蓝色荧光蛋白的序列、编码绿色荧光蛋白的序列和插入在中间的 gRNA 靶向序列；gRNA 靶向序列的两边带有随机的 N 碱基序列，可以用来筛选不同的原型间隔区相邻基序（PAM）偏好的 Cas 酶。

本申请的 Cas 酶切割活性将通过报告系统显现的荧光结果测定，未发生切割的报告系统将稳定地表达蓝色荧光蛋白从而发出蓝色荧光，而编码绿色荧光蛋白的序列前由于存在终止密码子而使得绿色荧光蛋白不能表达，报告系统因而不会发出绿色荧光。只有在 gRNA 靶向序列附近发生 DNA 切割后，细胞在修复切口过程中发生移码突变，才能在蓝色荧光蛋白稳定表达的同时表达出绿色荧光蛋白。图 1A 的荧光结果表明，在转染了本申请提供的 Cas 酶（SEQ ID NO: 2）后，具有切割活性的实验组细胞会产生明显的表达绿色荧光的细胞群。

将绿色荧光的细胞群用流式分选富集，提取基因组后扩增 target 区域，通过高通量测序就能确定本申请提供的 Cas 酶的 gRNA 结合所需要的 PAM 基序（如图 1B 所示）。

在本实施例中，编码报告系统的示例性核苷酸序列如下所示（粗体为蓝色和绿色荧光蛋白序列，斜体为连接序列，下划线为 2A 剪切肽序列，粗体下划线为示例性的包含随机 N 碱基的 gRNA 靶向序列（SEQ ID NO: 59））：

ATGAGCGAGCTGATTAAGGAGAACATGCACATGAAGCTGTACATGGAGGGGCAC
CGTGGACAACCATCACTTCAAGTGCACATCCGAGGGGCGAAGGCAAGCCCTACGAG
GGCACCCAGACCATGAGAATCAAGGTGGTTCGAGGGCGGCCCTCTCCCCTTCGCCT
TCGACATCCTGGCTACTAGCTTCCTCTACGGCAGCAAGACCTTCATCAACCACACC
CAGGGCATCCCCGACTTCTTCAAGCAGTCCTTCCCTGAGGGCTTCACATGGGAGA
GAGTCACCACATACGAAGACGGGGGCGTGCTGACCGCTACCCAGGACACCAGCCT
CCAGGACGGCTGCCTCATCTACAACGTCAAGATCAGAGGGGTGAACTTCACATCC
AACGGCCCTGTGATGCAGAAGAAAACACTCGGCTGGGAGGCCTTCACCGAGACG

CTGTACCCCGCTGACGGCGGCCTGGAAGGCAGAAACGACATGGCCCTGAAGCTC
GTGGGCGGGAGCCATCTGATCGCAAACATCAAGACCACATATAGATCCAAGAAACC
CGCTAAGAACCTCAAGATGCCTGGCGTCTACTATGTGGACTACAGACTGGAAAGA
ATCAAGGAGGCCAACAACGAGACCTACGTGAGCAGCACGAGGTGGCAGTGGCC
AGATACTGCGACCTCCCTAGCAAACCTGGGGCACAAGCTTAATGGATCCGGCGCAAC
AAACTTCtctctgctgaaacaagccggagatgTCGAAGAGAATCCTGGACCGGCTGAGCAGGCTCGTT
GGGAAGCAGCCTCCAAAGAAACCATCAAGAAAACCACCAAGCCCTGTCCCCGCTGCCATGTA
CCAGTGGAATAAATGGAGGCTGCATGCACATGAAGTGTCCGCAGCCCCAGTGCAGGCTCG
AGTGGTGCTGGAACGTGTGGCTGCGAGcccCCGGCNNNNNNCAAGTCCCCTCATCCCCAC
AGANNNNNTCgggTtaGGGGACCACTGGTTCGACGTGAGCGGTTCTGGCGAAGGGGCGAG
GTTCCTTGCTGACCTGTGGCGATGTAGAAGAAAATCCCGGCCCCGGATCCGTGAGCAA
GGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGA
CGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTA
CGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCCGTGCCCTGG
CCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCG
ACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCCGAAGGCTACGTCCA
GGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTG
AAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCA
AGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACAGCCACAA
CGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCC
GCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACAC
CCCCATCGGCGACGGCCCCGTGCTGCTGCCCCGACAACCACTACCTGAGCACCCAG
TCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGT
TCGTGACCGCCGCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAA (SEQ ID

NO: 61)。

实施例 2

测定本申请的 Cas 酶在真核细胞内的切割活性强弱

将本申请提供的 Cas 酶 (SEQ ID NOs: 2、47-51 和 54)、对应的 gRNA 和带有 gRNA 靶向位点的报告质粒共转染到 HEK293T 细胞中,通过改造报告质粒上的靶向序列,可以快速构

建不同 Cas 酶所需的报告系统。如图 2A 所示，未发生切割的报告系统细胞只表达蓝色荧光蛋白，绿色荧光蛋白因为被隔断而无法表达，此时报告系统仅发出蓝色荧光。在 gRNA 靶向位点发生切割后，细胞会通过同源重组修复 DNA，从而获得完整的绿色荧光蛋白表达框，从而使报告系统发出绿色荧光。

具有更强切割活性的 Cas 酶会产生更高比例的 DNA 切割，从而得到更多拷贝的修复后的绿色荧光蛋白，细胞显示出更强的绿色荧光，因此本实施例通过比较绿色荧光细胞的比例和荧光强度，快速比较不同 Cas 酶的切割活性强弱，其比较结果如图 2B 所示。

在本实施例中，编码报告系统的示例性核苷酸序列如下所示（粗体为蓝色和绿色荧光蛋白序列，斜体为连接序列，下划线为 2A 剪切肽序列；粗体下划线为示例性的 gRNA 靶向序列（SEQ ID NO: 60），其中该序列两端的 TTTG 和 TGG 将会根据不同 Cas 酶的 PAM 偏好设置成不同序列）：

ATGAGCGAGCTGATTAAGGAGAACATGCACATGAAGCTGTACATGGAGGGGCAC
CGTGGACAACCATCACTTCAAGTGCACATCCGAGGGGCGAAGGCAAGCCCTACGAG
GGCACCCAGACCATGAGAATCAAGGTGGTCGAGGGGCGGCCCTCTCCCCTTCGCCT
TCGACATCCTGGCTACTAGCTTCCCTCTACGGCAGCAAGACCTTCATCAACCACACC
CAGGGCATCCCCGACTTCTTCAAGCAGTCCTTCCCTGAGGGCTTCACATGGGAGA
GAGTCACCACATACGAAGACGGGGGCGTGCTGACCGCTACCCAGGACACCAGCCT
CCAGGACGGCTGCCTCATCTACAACGTCAAGATCAGAGGGGTGAACTTCACATCC
AACGGCCCTGTGATGCAGAAGAAAACACTCGGCTGGGAGGCCTTCACCGAGACG
CTGTACCCCGCTGACGGCGGCCTGGAAGGCAGAAACGACATGGCCCTGAAGCTC
GTGGGCGGGAGCCATCTGATCGCAAACATCAAGACCACATATAGATCCAAGAAACC
CGCTAAGAACCTCAAGATGCCTGGCGTCTACTATGTGGACTACAGACTGGAAAGA
ATCAAGGAGGCCAACAACGAGACCTACGTCGAGCAGCACGAGGTGGCAGTGGCC
AGATACTGCGACCTCCCTAGCAAACCTGGGGCACAAGCTTAATGGATCCGGCGCAAC
AACTTCTctctgctgaaacaagccggagatgTCGAAGAGAATCCTGGACCGGCTgtgagcaagggcgaggag
cTGTTACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCC
ACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGA
CCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCCGTGCCCTGGCCCACCCTCGT
GACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAG
CAGCACGACTTCTTCAAGTCCGCCATGCCCCGAAGGCTACGTCCAGGAGCGCACCA

TCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGG
CGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGG
CAACATCCTGGGGCACAAGCTGGAGTACAACACTACAACAGCCACAACGTCTATATCA
TGGCCGACAAGCAGAAGAACGGGCATCAAGGTGAACTTCAAGATCCGCCACAACAT
CGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGG
CGACGGCCCCGTGctgctgcccgacaaccacTGA~~AtaaggatccTTTG~~AAGTCCCCCTCATCCCCACA
GTGGTtaggatatcggaagctgacctgaaGTTTCATCTGCACCACCGGCAAGCTGCCCCGTGCCCT
GGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCC
CGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCCGAAGGCTACGTC
CAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGG
TGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTT
CAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACACTACAACAGCCAC
AACGTCTATATCATGGCCGACAAGCAGAAGAACGGGCATCAAGGTGAACTTCAAGAT
CCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAAC
ACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCCGACAACCACTACCTGAGCACCC
AGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGA
GTTCGTGACCGCCGCGGGATCACTCTCG~~catgga~~cgagctgtacAAGtaa (SEQ ID NO:
62)。

实施例 3

测定本申请的 Cas 酶的体外切割活性及其 PAM 序列

如图 3A 所示的本实施例的实验设计流程图，将本申请提供的 Cas 酶用大肠杆菌表达并纯化，并与体外转录的 gRNA (gRNA 靶向序列为 SEQ ID NO: 59，且其两边带有随机的 N 碱基序列用以测定 PAM 偏好) 一起与包含 PAM 文库的质粒文库反应，有切割活性的 Cas 酶会将质粒文库切断，形成线性化 DNA。通过体外连接，将接头片段连接到 DNA 线性化后的文库片段上，然后通过特异的 PCR 引物将载体和连接上接头的文库片段扩增，并进行高通量测序。

图 3B 所示的分析测序结果显示了有切割活性的 Cas 酶对于 PAM 基序的偏好性，通过这些结果可表明，EpiCas037 (SEQ ID NO: 36)、EpiCas044 (SEQ ID NO: 43)、EpiCas040 (SEQ ID NO: 39) 和 EpiCas045 四个蛋白 (SEQ ID NO: 44) 都具有切割活性，而且都偏好富含 T 的

PAM 基序。

实施例 4

在内源位点检测本申请的 Cas 酶活性

根据实施例 3 确定的不同 Cas 酶的 PAM 基序，在 HEK293T 的内源位点找到符合条件的位点，构建 gRNA（gRNA 靶向序列为 SEQ ID NO: 89）。将本申请提供的 Cas 酶（SEQ ID NOs: 44）和对应的 gRNA 共同转染到 HEK293T 细胞，转染 3 天后通过流式分选富集转染阳性的细胞，裂解细胞后扩增目的区域。将扩增后的 PCR 片段和 T7 内切酶反应，如果 Cas 酶切割了目的位点，将会形成各种不同的随机修复的序列，这些序列会被 T7 酶切割产生目的条带以外的条带。因此，图 4 所示的结果表明该 Cas 酶在内源位点具有切割活性。

实施例 5

测定本申请的 Cas 酶突变体在真核细胞内的切割活性强弱

通过定点突变，将本申请提供的 Cas 酶（SEQ ID NO: 2）的不同位点突变成目标氨基酸（见下表），再进行如实施例 2 所述的检测。其中，gRNA 的 DR（direct repeat）区域序列和 Spacer 序列分别如 SEQ ID NO: 64 和 60 所示。图 5A 所示的结果表明，m1、m2、m4、m5、m7、m11、m15 突变体活性相较于野生型均有显著提升。再将这些有效突变叠加后进行相同方法的活性检测，图 5B 所示的结果表明，包含由两种选自下表的突变叠加形成的双突变体能进一步提升活性，部分双突变体的活性与 AsCas12a 相当；图 5C 所示的结果表明，将有效的单突变和双突变再进一步叠加所形成的三突变体和四突变体仍然保持原有的活性，且活性与 AsCas12a 相当。用于对照的 AsCas12a 的氨基酸序列如 SEQ ID NO: 127 所示。

EpiCas002-m1	S128R
EpiCas002-m2	E168R
EpiCas002-m3	K171R
EpiCas002-m4	E177R
EpiCas002-m5	D190R
EpiCas002-m6	E248R
EpiCas002-m7	L259R
EpiCas002-m8	S295R
EpiCas002-m9	A352R

EpiCas002-m10	V496R
EpiCas002-m11	N527R
EpiCas002-m12	S532R
EpiCas002-m13	K533R
EpiCas002-m14	D536R
EpiCas002-m15	E672R
EpiCas002-m16	G922R
EpiCas002-m17	E926R
EpiCas002-m18	E967R
EpiCas002-m19	N980R

实施例 6

CRISPR/Cas 编辑系统的优化

本实施例将对包含本申请提供的 Cas 酶（SEQ ID NO: 57）以及 gRNA 的编辑系统进行优化。其中，对 Cas 酶活性的检测采用如实施例 2 所述的报告系统和检测方法。

gRNA 优化:

通过分子克隆将 gRNA 的 DR（direct repeat）区域改变成 36nt（SEQ ID NO: 87）、31nt（SEQ ID NO: 91）、22nt（SEQ ID NO: 90），图 6A 所示的结果表明，31nt 长度的 DR 即能维持 100%的 Cas 酶活性。当通过分子克隆改变 DR 区域的碱基对以提高其稳定性时（DR 为 31.1 至 31.8 时，其核苷酸序列如 SEQ ID NO: 92-99 所示），图 6B 所示的结果表明，大部分改动对活性影响不大，而 31.8 的改动能够最优选地提升活性。本实施例还通过分子克隆改变了 gRNA 用于靶向 DNA 的间隔区（Spacer）长度，用以提高其结合效率，从而增强 Cas 酶活性。如图 7 所示，当 Spacer 长度从 24nt 变化至 16nt（SEQ ID NO: 100-103，60 和 104-107），本申请所提供的 Cas 酶均能维持活性。

Cas 酶突变体:

通过定点突变，将本申请提供的 Cas 酶（SEQ ID NO: 57）的不同位点突变成目标氨基酸（见下表），再进行如实施例 2 所述的检测。其中，gRNA 的 DR（direct repeat）区域序列和 Spacer 序列分别如 SEQ ID NO: 91 和 60 所示。图 8A-8B 所示的结果表明，多数单一突变对 Cas 酶活性的影响不大，如果将其中对活性优化效果较好的单一突变进行叠加，可以发现所形成的双突变体仍能保持较高的活性。

本实施例还在实施例 2 所述的报告系统基础上，将 TTTG 分别突变成 TTTH、TTVG、TVTG 和 VTTG（H 代表 A/C/T， V 代表 A/C/G，3 个质粒等比例混合后转染），构建出新的 PAM 报告系统，通过图 8C 所示的结果发现，部分突变体能够拓宽 PAM 识别的范围，尤其是在识别 VTTG、TTVG 的 PAM 的靶点上，相较于野生型，本申请提供的 Cas 酶活性获得了显著提升。

EpiCas059-m1	K57R
EpiCas059-m2	Y258R
EpiCas059-m3	K264R
EpiCas059-m4	K272R
EpiCas059-m5	D621R
EpiCas059-m6	Q649R
EpiCas059-m7	E681R
EpiCas059-m8	Y754R
EpiCas059-m9	S767R
EpiCas059-m10	Q770R
EpiCas059-m11	D823R
EpiCas059-m12	K850R
EpiCas059-m13	V861R
EpiCas059-m14	K186R
EpiCas059-m15	T197R
EpiCas059-m16	V242R
EpiCas059-m17	K304R
EpiCas059-m18	T319R
EpiCas059-m19	E397R
EpiCas059-m20	Y364R
EpiCas059-m21	S421R
E piCas059-m22	V453R
EpiCas059-m23	P627R
EpiCas059-m24	K9R
EpiCas059-m25	I10R
EpiCas059-m26	C39R
EpiCas059-m27	N53R
EpiCas059-m28	H65R

EpiCas059-m29	S69R
EpiCas059-m30	E81R
EpiCas059-m31	K95R
EpiCas059-m32	D98R
EpiCas059-m33	V103R
EpiCas059-m34	G112R
EpiCas059-m35	I115R
EpiCas059-m36	G120R
EpiCas059-m37	K123R
EpiCas059-m38	A153N
EpiCas059-m39	A153R
EpiCas059-m40	T346S
EpiCas059-m41	T346R
EpiCas059-m42	D347N
EpiCas059-m43	D347R
EpiCas059-m44	D183Y/T458K/ G528C
EpiCas059-m45	E144K/S633N
EpiCas059-m46	E144K/D183Y/ T458K/G528C/ S633N
EpiCas059-m47	E144K/T458K

实施例 7

在内源位点检测失活 Cas 酶（dCas）效果

dCas-CBE 的碱基编辑效果:

将本申请提供的 Cas 酶（SEQ ID NO: 57）进行失活处理，然后与 rAPOBEC1、UGI 融合 (图 9A), 从而获得基于本申请提供的 Cas 酶的胞嘧啶碱基编辑器, 并将其与对应的 gRNA (DR 序列如 SEQ ID NO: 91 所示, 靶向 EMX1 位点和 VEGFA 位点的 Spacer 序列分别如 SEQ ID NO: 108 和 109 所示) 共转染到 HEK293T 细胞中, 转染 48 小时后分选转染阳性的细胞, 提取基因组, 扩增目的位点, 通过 Sanger 测序确定碱基编辑效率, 结果如图 9B-9C 所示。

其中, dEpiCas059m46-CBE 编辑器的氨基酸序列如下所示 (SEQ ID NO: 116, 斜体为 NLS, 斜粗体为 rAPOBEC1, 粗体为 dEpiCas059, 斜体下划线为 UGI, 下划线为 P2A 剪切肽, 以及<>内为 mCherry 标记物; 其质粒序列为 SEQ ID NO: 117) :

MKRTADGSEFESPKKKRKV***SS*****ETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINW**
GGRHSIWRHTSQNTNKHVEVNFIEKFTT***ERY*****FCPNTRCSITWFLSWSPCGECSRAITEFLS**

RYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPTSNEAH
WPRYPHLWVRLYVLELYCIHLGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLK
SGSETPGTSESATPESGSEETRTPEDIACQQDRSAFFKLGKLPQKTTSIYHLMRCSKRY
IDNVSETISNNIEKAAMTSDNHIFRSALKEEYKKIYREKFYPDVTGQPFDPKGPDNTHI
VERTVENLCGQYISALKGVMKVRKPKWQDDIFEDES DKVIKMLMASGNEADIHPKL
EKKIASAKKKDKMDEWQKVLDRIYYIKSKTDIEWTYTTLLYDHIARECARSMFEDGN
KPKFCYTNGNKFQKQVDELWGSEVEYEEKYLHCRTLEAYWNEENKYFFLENQK
KSKKTYEARQIIDHFIENEDAVKAETEEYMAKVAKEFGENGCKFRTENILYLVNSVSI
NKKQDEERCPIFKVTDIPQYAEAGIDRIGECKYVIKLLFTVGERYGTTYMPEAFSGLY
DRRTNQRYDLIDRACKGSYVYLPIYISDKDSRKWSDRLKNTFTMSAFSISERKTKKDG
RVLCVNLPYKHIQHKPRNYFKVYPTYGVAINVKHSLFSTNATNLNGYIDVNGELIDFYN
SIGRPDKLKNINLAVTLGDKCGTAGNHTNFLLVSDLGWDHGGKKRNDRFVKREDGT
YYSRPDLLADGLNFLKKKYSGDRKSYDYINTTIAFRNNQTRLRCIDKEWNRKRAYD
EEHGKPEKGERNGFYPTEDAERLTMERQALRREIAENTRQVIAYMLWLFQENGIXIA
LAHLDCGSWKNFRDQLKPKWFCNNCKYWAKEYGIPYEATKYYSNRKYFDIEDGELI
PENVKLNALGEAENDKRYFQTSIAKAVGMSSIKQMFISVSQNYDIQVNYTEQAYTSQT
DSRNGTVDKDFRVSQECYIGSDGTMNCAFNAAVNIADRISDDTLREEFNKELPMKEF
VKYGATQIVSRNRRWNAENAERTMAAFQKRKNADGGAHTAAVKPTS KTEESSKLKLV
AEPKKKRKVSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDEN
VMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSPKKKRKVGSKRTGSGATNFSLLKQAGDVE
ENPGPVN<MVSKGEEDNMAIIEKFMRFKVHMEGSVNGHEFEIEGEGEGRPYEGTQTAKLK
VTKGGPLPFAWDILSPQFMYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVT
QDSSLQDGEFIYKVKLRGTNFPDGPVMQKKTMGWEASSERMYPEDGALKGEIKQRLKLK
DGGHYDAEVKTTYKAKKPVQLPGAYNVNIKLDITSHNEDYTIVEQYERAEGRHSTGGMDE
LYK>*

dAsCas12a-CBE 编辑器的氨基酸序列如下所示 (SEQ ID NO: 118, 斜体为 NLS, 斜粗体为 rAPOBEC1, 粗体为 dAsCas12a, 斜体下划线为 UGI, 下划线为 P2A 剪切肽, 以及<>内为 mCherry 标记物; 其质粒序列为 SEQ ID NO: 119):

MKRTADGSEFESPKKKRKVSSETGPVAVDPTLRRRIEPHEFEVFFDPRELKRETCCLLYEINW
GGRHSIWRHTSQNTNKHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLS

RYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPTSNEAH
WPRYPHLWVRLYVLELYCIHLGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLK
SGSETPGTSESATPESGSSKLEKFTNCYSLSKTLRFKAIPVGKTQENIDNKRLLEVEDEKR
AEDYKGVKKLLDRYYLSFINDVLHSIKLKNLNNYISLFRKKTRTEKENKELENLEINLR
KEIAKAFKGNEGYKSLFKKDIIETILPEFLDDKDEIALVNSFNGFTTAFTGFFDNRENMF
SEEAKSTSIAFRCINENLTRYISNMDIFEKVDAIFDKHEVQEIKEKILNSDYDVEDFFEGE
FFNFVLTQEGIDVYNAIIGGFVTESGEKIKGLNEYINLYNQKTKQKLPKFKPLYKQVLS
DRESLSFYGEGYTSDEEVLEVFRNTLNKNSEIFSSIKKLEKLFKNFDEYSSAGIFVKNGP
AISTISKDIFGEWNVIRDKWNAEYDDIHLKKKAVVTEKYEDDRRKSFKKIGSFSLEQLQ
EYADADLSVVEKLKEIIIQKVDEIYKVYGSSEKLFDAADFVLEKSLKKNDVVAIMKDLL
DSVKSFENYIKAFFGEGKETNRDESFYGDFVLAYDILLKVDHIYDAIRNYVTQKPYSKD
KFKLYFQNPQFMGGWDKDKETDYRATILRYGSKYYLAIMDKKYAKCLQKIDKDDVN
GNYEKINYKLLPGPNKMLPKVFFSKKWMAYYNPSEDIQKIYKNGTFKKGDMFNLND
CHKLIDFFKDSISRYPKWSNAYDFNFSETEKYKDIAGFYREVEEQGYKVSFESASKKEV
DKLVEEGKLYMFQIYNKDFSDKSHGTPNLHTMYFKLLFDENNHGQIRLSGGAELFMR
RASLKKEELVHPANSPIANKNPDNPCKTTTLSYDVYKDKRFSEDQYELHIPIAINKCP
KNIFKINTEVRVLLKHDDNPYVIGIARGERNLLYIVVVDGKGNIVEQYSLNEIINNFNIGI
RIKTDYHSLLDKKEKERFEARQNWTSIENIKELKAGYISQVVHKICELVEKYDAVIALA
DLNSGFKNRSRVKVEKQVYQKFELMLIDKLNMYVDKKSNPCATGGALKGYQITNKFE
SFKSMSTQNGFIFYIPAWLTSKIDPSTGFVNLLKTKYTSIADSKKFISSFDRIMYVPEEDL
FEFALDYKNFSRTDADYIKKWLYSYGNRIRIFRNPKKNNVFDWEEVCLTSAYKELFN
KYGINYQQGDIRALLCEQSDKAFYSSFMALMSLMLQMRNSITGRTDVAFLISPVKNSD
GIFYDSRNYEAQENAILPKNADANGAYNIARKVLWAIGQFKKAEDEKLDKVKIAISNK
EWLEYAQTSVKHPKKKRKVSGGSTTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDIL
VHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSPKKKRKVGSKRTGSGATN
FSLKQAGDVEENPGPVN<MVSKGEEDNMAIIEKFMRFKVHMEGSVNGHEFEIEGEGEGR
PYEGTQTAKLKVTKGGPLPFAWDILSPQFMYGSKAYVKHPADIPDYLKLSFPEGFKWERVM
NFEDGGVVTVTQDSSLQDGEFIYKVKLRTNFPSDGPVMQKKTMGWEASSERMYPEDGA
LKGEIKQRLKLDGGHYDAEVKTTYKAKKPVQLPGAYNVNIKLDITSHNEDYTIVEQYER
AEGRHSTGGMDELYK>*

dCas-SPH 的基因激活效果:

将本申请提供的 Cas 酶 (SEQ ID NO: 55) 进行失活处理, 然后与 10×GCN4 融合, 用以招募 scFV-P65-HSF1 的融合肽, 从而获得基于本申请提供的 Cas 酶的基因激活工具。该工具的原理基于 GCN4 能够自发与 scFV 相互识别并结合, 进而将具有转录激活功能的 P65 和 HSF1 效应物富集到 Cas 酶的靶位点附近, 继而激活靶位点的基因表达。

将转录激活工具和靶向 CXCR4 的 gRNA (Spacer 序列如 SEQ ID NOs: 120-123 所示, 该 4 条 gRNA 被等比例混合) 共转染到 HEK293T 细胞中, 转染 48 小时后, 收集细胞, 用 PE anti-human CXCR4 抗体染色 (BioLegendg, 306506), 通过流式细胞仪检测 PE 通道荧光强度, 用以反映 CXCR4 的表达强度。将转染阳性群的 PE 平均荧光强度, 除以转染阴性群的 PE 平均荧光强度, 得到激活效率 (MFI fold change), 用以代表不同激活工具的激活强度 (图 10), 进而代表不同工具的 DNA 结合效果。

其中, scFV-P65-HSF1 融合肽的氨基酸序列如下所示 (SEQ ID NO: 124, 斜体为 NLS, 斜粗体为 P65 和 HSF1, 粗体为 scFV, 斜体下划线为 HA 标签, 下划线为连接肽, 以及<>内为 Flag 标记物):

MGPDIVMTQSPSSLSASVGDRVITITCRSSTGAVTTSNYASWVQEKPGLFKGLIGGTN
NRAPGVPSRFSGLIGDKATLTISLQPEDFATYFCALWYSNHWVFGQGTKVELKRGG
GSGGGGSGGGGSSGGGSEVKLLESGGGLVQPGGSLKLSCAVSGFSLTDYGVNWVRQ
APGRGLEWIGVIWGDGITDYN SALKDRFIISKDNGKNTVYLQMSKVRSDDTALYYCVT
GLFDYWGGQGTTLVTVSSYPYDVDPDYAGGGGGSGGGGSGGGGSGGGGSPKKKRKVAAAPSG
QISNQALALAPSSAPVLAQTMVPSSAMVPLAQPPAPAPVLTGPPQSLAPVPKSTQAGEGT
LSEALLHLQFDADEDLGALLGNSTDPGVFTDLASVDNSEFQQLLNQGVSMHSTAEPMLM
EYPEAITRLVTGSQRPPDPAPTPLGTSGLPNGLSGDEDFSSIADMDFSALLSQISSSGQGGG
GSGFSVDTSALLDLFSPSVTPDMSLPDLDSLASIQELLSPQEPPRPPEAENSSPDSGKQLV
HYTAQPLFLLDPGSVDTGNDLPVLFELGEGSYFSEGDGFAEDPTISLLTGSEPPKAKDPTV
SASEGRGSLLT<CDYKDDDDK>*

dEpiCas057-10×GCN4 融合肽的氨基酸序列如下所示 (SEQ ID NO: 125, 斜体为 NLS, 粗体为 dEpiCas057, 以及<>内为 GCN4):

MKRTADGSEFESPKKKRKVHRRRRVREGITMSDTTFRPYKSKLNPNERKMPMVVDTMN
WCDNACSVFFDLFVGLYGGVKIGDVTNFLTEKQDTTTDADLLCAINWFRAMPVSNPD
AVVGPEQLVVRFRDYSGHNPSPMASMYLSAPVNPRYTWVDCRFLYQQMCQTLGANL

DVDISTLVQTGLLPIQKDTQLNAYSFCSNVFGEGVKEDPSAKVTSYTAAANALSGATNP
IITYEDYRQILSSAITNRPSRKGAKDILAADLDSTKTGVLPGNYVSQRIAMYQASAASK
ARGVVLPNRLRVKAYLQNLVGLNYICTTWAAPMLRAVGDIKSKNSNNYRFVGERLER
QQEIDNLLAVSNIRSAQAILQQFRANESNQFYIKKRHIGKGLPSLYDMWDLMDKEGI
DNYIQSLKGDKTERLPIRELLQIHYPHRKSVSAAELFLSAAELTDVEDRQNRRCMVHPTVR
GSLVINFGKSSTIFGAITPPDTLIKFRGEMVPAGQTGMIWFTMHLVDGGKWVKHHIPC
HSSRYFEEIYAWKKDLPTLPFPRRSLYGYAIGNKISDTRKIDLRRKKASKQYLRTLENM
THNVEFDPTTQFMVDQELNVTITARVKKIPPVRDVPVGSRLMGLNKNQTIRDCYSIWE
RVEANTSGSFQHPVVQDVWLRLVADGHVTSTVNGPNGQIDQMSYSGPDVANCNWM
QERSNFITSVDDADQSIMKHFLWNNRVKLYAWNCVYAGTLMKAISKQMTTANRNLF
NEILSFVGGFRFGCRLGSLAQVSLDLLRRMKSVISCYFSRNNAFTEERSAFDKELMELF
ESLDNKMNNKRKEKVNRTLSSIFQVAKQHNVVRLVVTKNLQTPDANVR SVMNQRSVD
WCARKIEEKIVDLGNVIGIYVDLVEALDLSHLDPFYIYAKGDVGKECQFDEVESSLLPI
HTSMFRDWHGLVNKGRTTTPYHSALRSFASHYGLRFDDLARIKAAELAKMLPDQKV
FIPHRGGRFFMSTYPVTSTSRPVSYDGRQRFLNSGAMVAPVNTMLRGFERRARSHHHHH
HPAAKRVKLDDYKDHDGDYKDHDIDYKDDDDKGSM<EELLSKNYHLENEVARLKK>GSG
SG<EELLSKNYHLENEVARLKK>GSGSG<EELLSKNYHLENEVARLKK>GSGSG<EELLSKN
YHLENEVARLKK>GSGSG<EELLSKNYHLENEVARLKK>GSGSG<EELLSKNYHLENEVAR
LKK>GSGSG<EELLSKNYHLENEVARLKK>GSGSG<EELLSKNYHLENEVARLKK>GSGSG
<EELLSKNYHLENEVARLKK>GSGSG<EELLSKNYHLENEVARLKK>GSGSGTAVNIGGGT
GPMDLQRPQGGGGPKKKRKVGSGVRLQRR*

融合肽 scFV-P65-HSF1 和 dEpiCas057-10×GCN4 可由 SEQ ID NO: 126 所示的质粒序列一同表达。

以上实验证明了通过突变失去切割活性的本申请提供的 Cas 酶适用于碱基编辑以及表观修饰编辑的应用场景，并且不限于其他基于 DNA 靶向的应用场景，如基因激活、基因沉默、染色体成像等等。

权 利 要 求 书

1. 一种分离的 Cas 酶，所述 Cas 酶包含 SEQ ID NOs: 1-58 中任一项所示的氨基酸序列或与所述 SEQ ID NOs: 1-58 中任一项所示的氨基酸序列具有至少约 80% 同一性的序列。
2. 根据权利要求 1 所述的 Cas 酶，所述 Cas 酶具有能够结合靶 DNA 链的催化活性结构域和/或切割所述靶 DNA 链的催化活性结构域。
3. 根据权利要求 1-2 中任一项所述的 Cas 酶，所述催化活性结构域包含一个或多个氨基酸的改变，从而使得所述 Cas 酶仅具有结合靶 DNA 链的活性，或者具有结合靶 DNA 链的活性和切割所述靶 DNA 单链的活性。
4. 一种融合分子，所述融合分子包含权利要求 1-3 中任一项所述的 Cas 酶和一个或多个异源功能结构域。
5. 根据权利要求 4 所述的融合分子，所述一个或多个异源功能结构域能够调控一种或多种基因产物的表达。
6. 根据权利要求 4-5 中任一项所述的融合分子，所述一个或多个异源功能结构域直接或间接地融合在所述 Cas 酶上。
7. 根据权利要求 4-6 中任一项所述的融合分子，所述一个或多个异源功能结构域选自自解旋酶、核酸酶、解旋酶-核酸酶、DNA 甲基转移酶、DNA 羟甲基化酶、组蛋白甲基化酶、组蛋白去甲基化酶、组蛋白乙酰转移酶、组蛋白去乙酰化酶、磷酸酶、激酶、转录(共)活化物、转录阻遏物、DNA 结合蛋白、DNA 结构蛋白、标志物蛋白、报告物蛋白、荧光蛋白、配体结合蛋白、信号肽、亚细胞定位序列、抗体表位和亲和纯化标签。
8. 根据权利要求 4-7 中任一项所述的融合分子，所述一个或多个异源功能结构域具有以下活性中的一种或多种：甲基酶活性、脱甲基酶活性、脱氨酶活性、转录激活活性、转录抑制活性、转录释放因子活性、逆转录酶活性、组蛋白修饰活性、RNA 切割活性和核酸结合活性。
9. 一种工程化的、可编程的、非天然存在的 CRISPR-Cas 系统，所述系统包含权利要求 1-3 中任一项所述的 Cas 酶或权利要求 4-8 中任一项所述的融合分子，和一种或多种指导 RNA，所述一种或多种指导 RNA 在细胞中靶向编码一种或多种基因产物的核酸分子的基因座，从而指导所述 Cas 酶或所述融合分子结合和/或切割所述编码一种或多种基因产物的核酸分子的基因座；并且，所述 Cas 酶或所述融合分子与所述指导 RNA 不共同天然存在。
10. 一种工程化的、非天然存在的载体系统，所述载体系统包含一种或多种载体，所述一种或多种载体包括：
 - a) 第一调节元件，所述第一调节元件可操作地连接到一种或多种指导 RNA 上，所述一种或多种指导 RNA 能够与编码一种或多种基因产物的核酸分子的基因座中的靶序列杂交，

和

b) 第二调节元件, 所述第二调节元件可操作地连接到权利要求 1-3 中任一项所述的 Cas 酶或权利要求 4-8 中任一项所述的融合分子上,

其中, 所述组分 a) 和所述组分 b) 位于所述载体系统的相同或不同载体上, 且所述指导 RNA 在细胞中靶向所述编码一种或多种基因产物的核酸分子的基因座, 从而指导所述 Cas 酶或所述融合分子结合和/或切割所述编码一种或多种基因产物的核酸分子的基因座; 并且, 所述 Cas 酶或所述融合分子与所述指导 RNA 不共同天然存在。

11. 根据权利要求 9-10 中任一项所述的系统, 所述一种或多种基因产物的表达被改变。
12. 根据权利要求 9-11 中任一项所述的系统, 所述基因产物的表达被降低或者被增多。
13. 根据权利要求 9-12 中任一项所述的系统, 所述基因产物是一种蛋白质。
14. 根据权利要求 9-13 中任一项所述的系统, 所述细胞是真核细胞。
15. 根据权利要求 9-14 中任一项所述的系统, 所述真核细胞是哺乳动物细胞。
16. 根据权利要求 9-15 中任一项所述的系统, 所述哺乳动物细胞是人类细胞。
17. 根据权利要求 9-16 中任一项所述的系统, 所述 Cas 酶是经密码子优化的, 用以在真核细胞中进行表达。
18. 根据权利要求 9-17 中任一项所述的系统, 所述指导 RNA 包含融合到 tracr 序列上的指导序列。
19. 根据权利要求 9-18 中任一项所述的系统, 所述指导 RNA 包含直接重复 (Direct repeat) 序列和间隔 (Spacer) 序列, 其中所述间隔序列与所述指导 RNA 靶向的核酸分子结合。
20. 根据权利要求 19 所述的系统, 所述直接重复序列的长度为 10 个至 70 个核苷酸。
21. 根据权利要求 19 或 20 所述的系统, 所述直接重复序列的长度为 31 个至 36 个核苷酸。
22. 根据权利要求 19-21 中任一项所述的系统, 所述直接重复序列包含 SEQ ID NO: 63-88 和 90-99 中任一项所示的核苷酸序列, 或者包含与 SEQ ID NO: 63-88 和 90-99 中任一项所示的核苷酸序列具有至少 95% 序列同一性的核苷酸序列。
23. 根据权利要求 19 所述的系统, 所述间隔序列的长度为 16 个至 24 个核苷酸。
24. 根据权利要求 19 或 23 所述的系统, 所述指导 RNA 靶向的核酸分子包含能够与所述间隔序列互补配对的核苷酸序列。
25. 根据权利要求 9-24 中任一项所述的系统, 所述系统的所述载体或所述 Cas 酶还包含一个或多个核定位序列 (NLS)。
26. 根据权利要求 9-25 中任一项所述的系统, 所述系统通过递送系统被引入所述细胞中, 所述递送系统选自病毒粒子、脂质体、脂质纳米颗粒、电穿孔、显微注射和缀合。

27. 改变一种或多种基因产物的表达的方法，所述方法包括向包含和表达编码所述一种或多种基因产物的核酸分子的细胞中引入一种工程化的、非天然存在的 CRISPR-Cas 系统，所述系统包含权利要求 1-3 中任一项所述的 Cas 酶或权利要求 4-8 中任一项所述的融合分子，和一种或多种指导 RNA，所述一种或多种指导 RNA 靶向所述编码一种或多种基因产物的核酸分子的基因座，从而指导所述 Cas 酶或所述融合分子结合和/或切割所述基因座，由此改变所述一种或多种基因产物的表达；并且，所述 Cas 酶或所述融合分子与所述指导 RNA 不共同天然存在。
28. 改变一种或多种基因产物的表达的方法，所述方法包括向包含和表达编码所述一种或多种基因产物的核酸分子的细胞中引入一种工程化的、非天然存在的载体系统，所述载体系统包含一种或多种载体，所述一种或多种载体包括：
- a) 第一调节元件，所述第一调节元件可操作地连接到一种或多种指导 RNA 上，所述一种或多种指导 RNA 能够与所述编码一种或多种基因产物的核酸分子的基因座中的靶序列杂交，和
- b) 第二调节元件，所述第二调节元件可操作地连接到权利要求 1-3 中任一项所述的 Cas 酶或权利要求 4-8 中任一项所述的融合分子上，
- 其中，所述组分 a) 和所述组分 b) 位于所述载体系统的相同或不同载体上，且所述指导 RNA 在所述细胞中靶向所述编码一种或多种基因产物的核酸分子的基因座，从而指导所述 Cas 酶或所述融合分子结合和/或切割所述基因座，由此改变所述一种或多种基因产物的表达；并且，所述 Cas 酶或所述融合分子与所述指导 RNA 不共同天然存在。
29. 根据权利要求 27 或 28 所述的方法，所述基因产物的表达被降低或者被增多。
30. 根据权利要求 27-29 中任一项所述的方法，所述基因产物是一种蛋白质。
31. 根据权利要求 27-30 中任一项所述的方法，所述细胞是真核细胞。
32. 根据权利要求 27-31 中任一项所述的方法，所述真核细胞是哺乳动物细胞。
33. 根据权利要求 27-32 中任一项所述的方法，所述哺乳动物细胞是人类细胞。
34. 根据权利要求 27-33 中任一项所述的方法，所述 Cas 酶是经密码子优化的，用以在真核细胞中进行表达。
35. 根据权利要求 27-34 中任一项所述的方法，所述指导 RNA 包含融合到 tracr 序列上的指导序列。
36. 根据权利要求 27-35 中任一项所述的方法，所述指导 RNA 包含直接重复 (Direct repeat) 序列和间隔 (Spacer) 序列，其中所述间隔序列与所述指导 RNA 靶向的核酸分子结合。
37. 根据权利要求 36 所述的方法，所述直接重复序列的长度为 10 个至 70 个核苷酸。

38. 根据权利要求 36 或 37 所述的方法，所述直接重复序列的长度为 31 个至 36 个核苷酸。
39. 根据权利要求 36-38 中任一项所述的方法，所述直接重复序列包含 SEQ ID NO: 63-88 和 90-99 中任一项所示的核苷酸序列，或者包含与 SEQ ID NO: 63-88 和 90-99 中任一项所示的核苷酸序列具有至少 95% 序列同一性的核苷酸序列。
40. 根据权利要求 36 所述的方法，所述间隔序列的长度为 16 个至 24 个核苷酸。
41. 根据权利要求 36 或 40 所述的方法，所述指导 RNA 靶向的核酸分子包含能够与所述间隔序列互补配对的核苷酸序列。
42. 根据权利要求 27-41 中任一项所述的方法，所述系统的所述载体或所述 Cas 酶还包含一个或多个核定位序列 (NLS)。
43. 根据权利要求 27-42 中任一项所述的方法，所述方法包括通过递送系统将所述 CRISPR-Cas 系统或所述载体系统引入到所述细胞中，所述递送系统选自病毒粒子、脂质体、脂质纳米颗粒、电穿孔、显微注射和缀合。
44. 编码权利要求 1-3 中任一项所述的 Cas 酶、权利要求 4-8 中任一项所述的融合分子或权利要求 9 和 11-26 中任一项所述的 CRISPR-Cas 系统的核酸。
45. 一种细胞，所述细胞包含权利要求 1-3 中任一项所述的 Cas 酶、权利要求 4-8 中任一项所述的融合分子、权利要求 9 和 11-26 中任一项所述的 CRISPR-Cas 系统、权利要求 10-26 中任一项所述的载体系统和/或权利要求 44 所述的核酸。
46. 一种试剂盒，所述试剂盒包含权利要求 1-3 中任一项所述的 Cas 酶、权利要求 4-8 中任一项所述的融合分子、权利要求 9 和 11-26 中任一项所述的 CRISPR-Cas 系统、权利要求 10-26 中任一项所述的载体系统、权利要求 44 所述的核酸和/或权利要求 33 所述的细胞。
47. 根据权利要求 46 所述的试剂盒，所述试剂盒还包含用于放置所述 Cas 酶、所述融合分子、所述 CRISPR-Cas 系统、所述载体系统、所述核酸和/或所述细胞的容器，以及用法说明书。

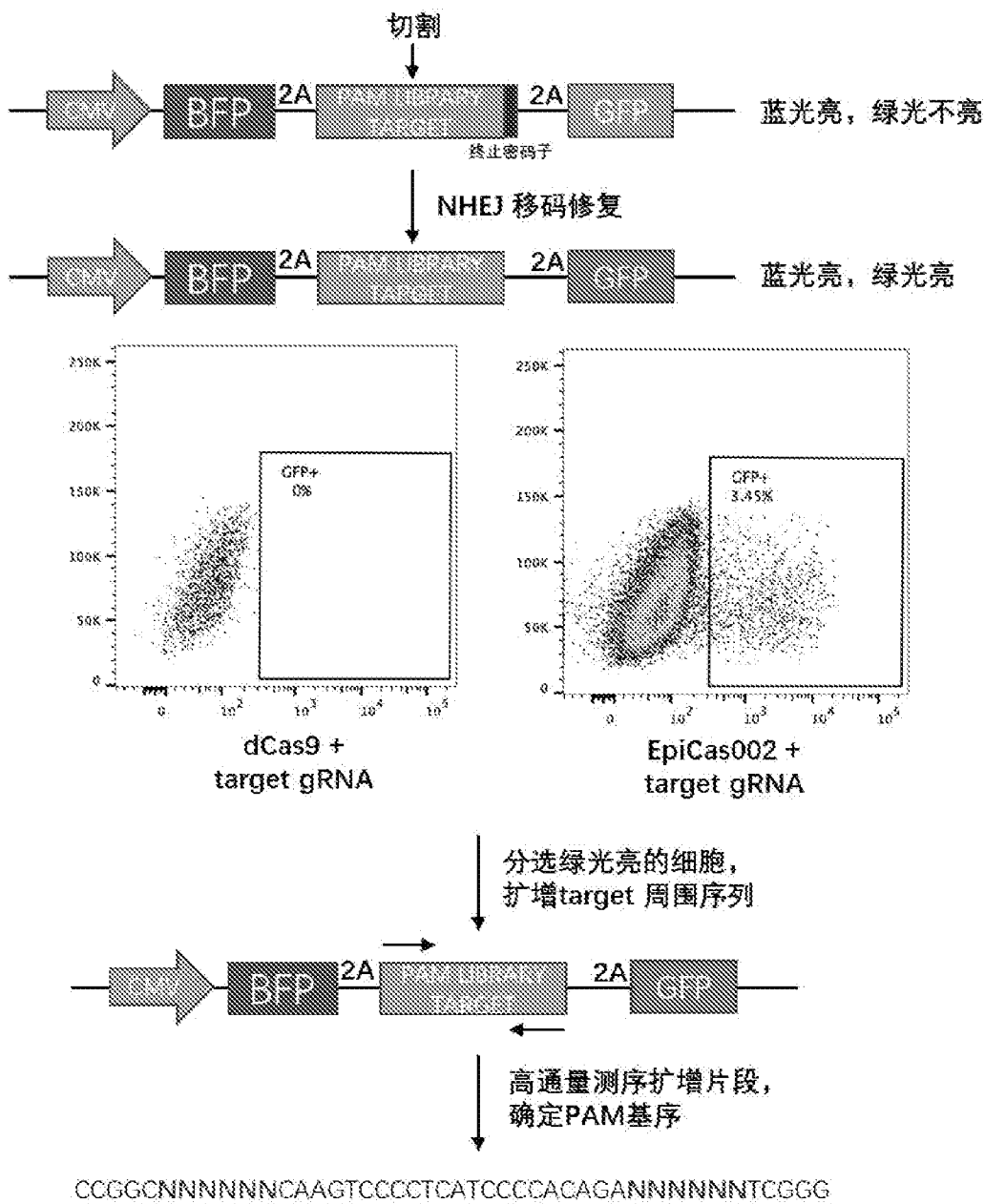


图 1A

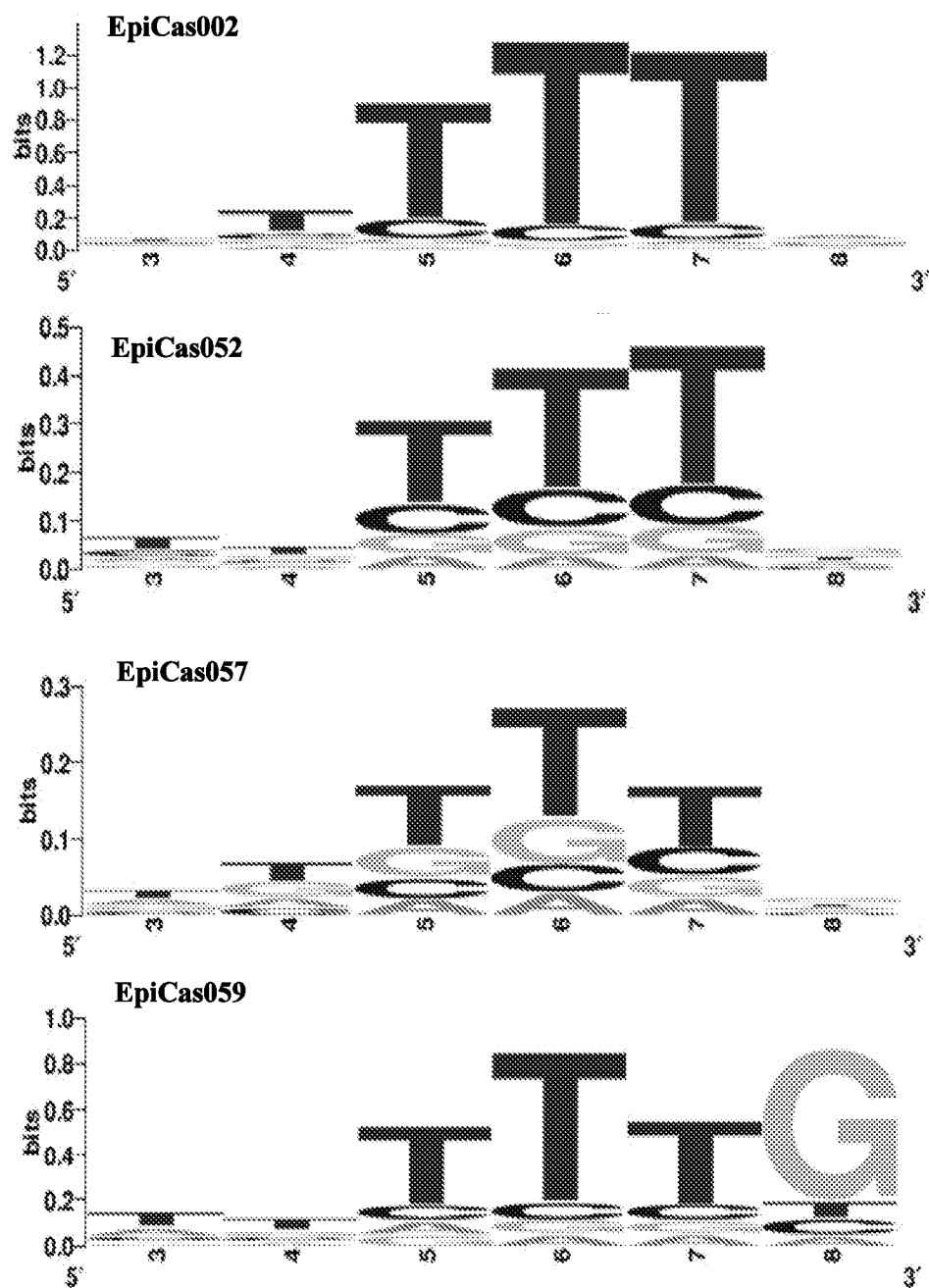


图 1B

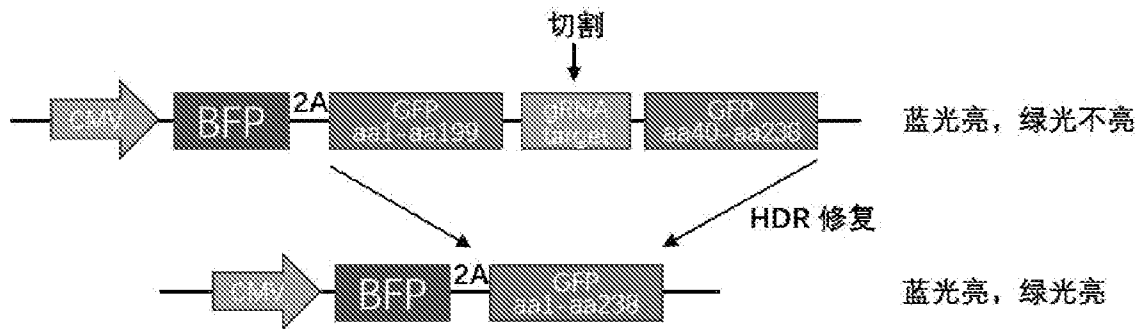


图 2A

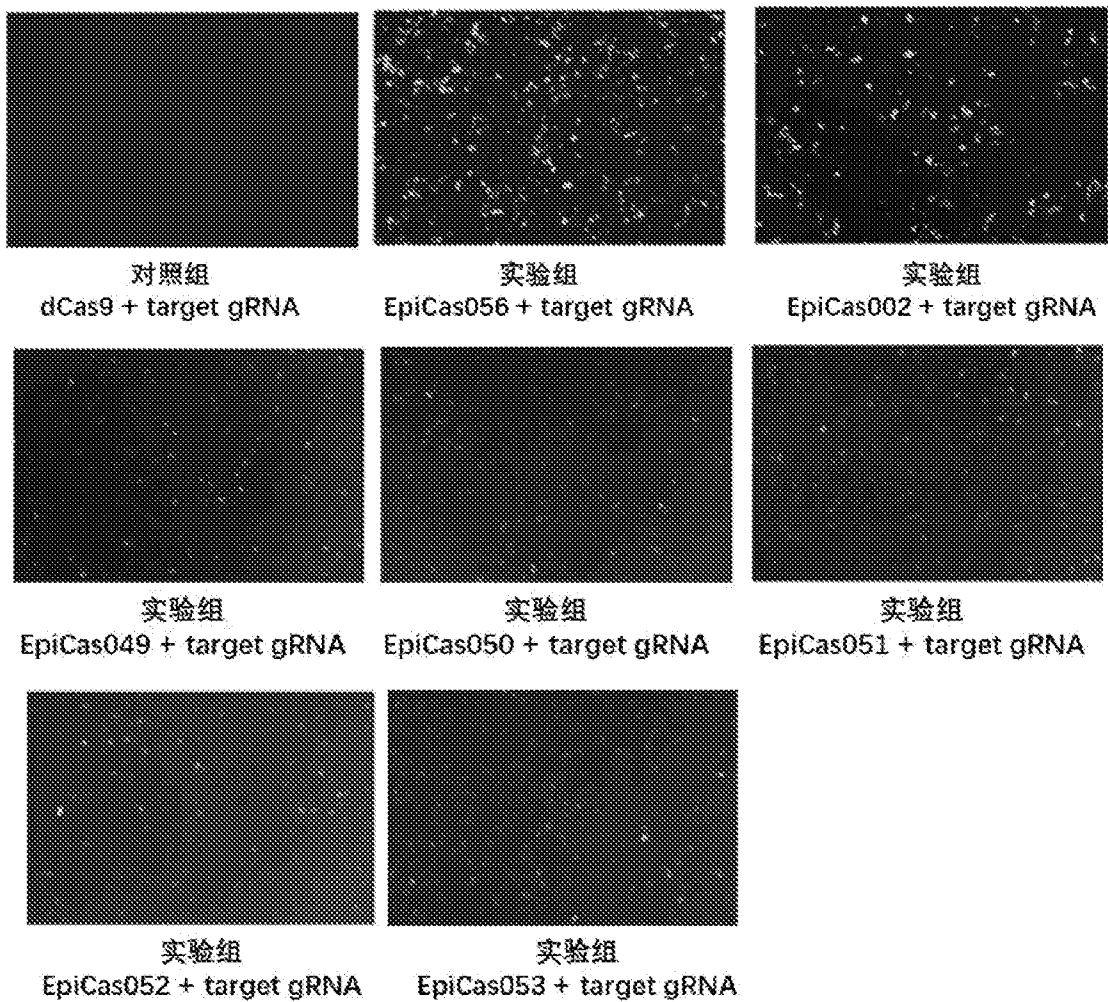


图 2B

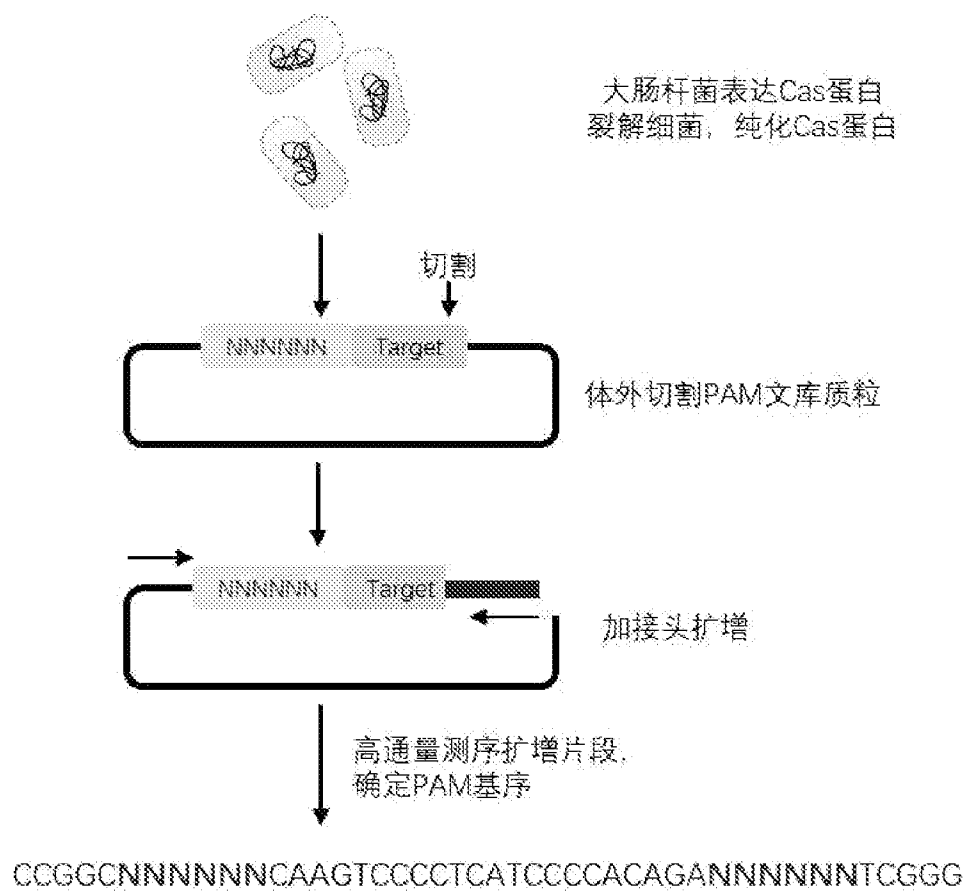


图 3A

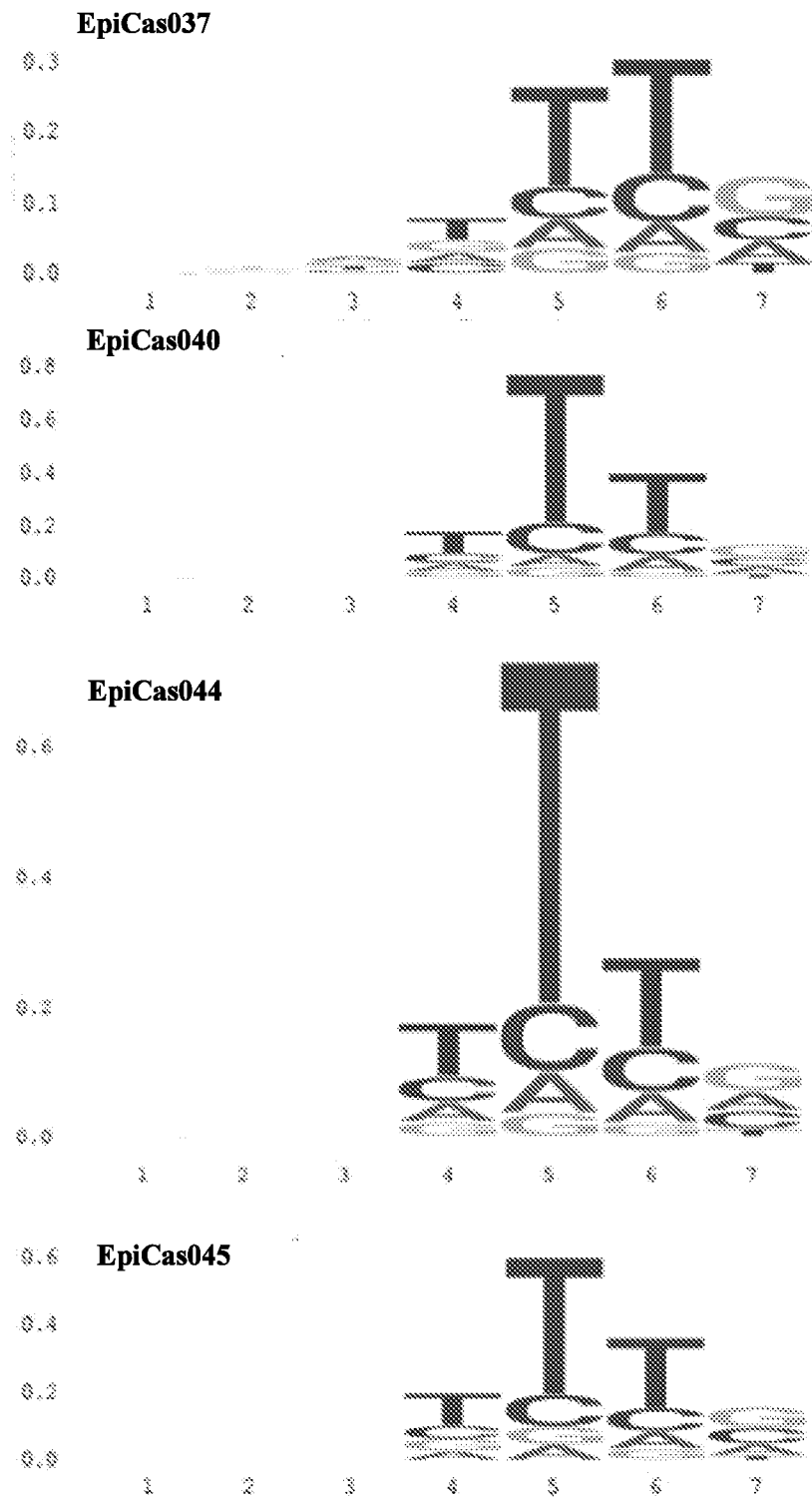
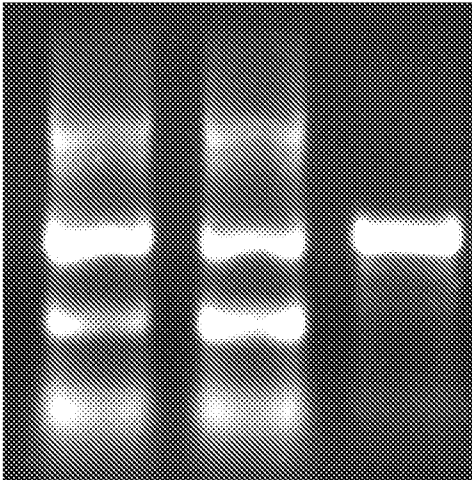


图 3B

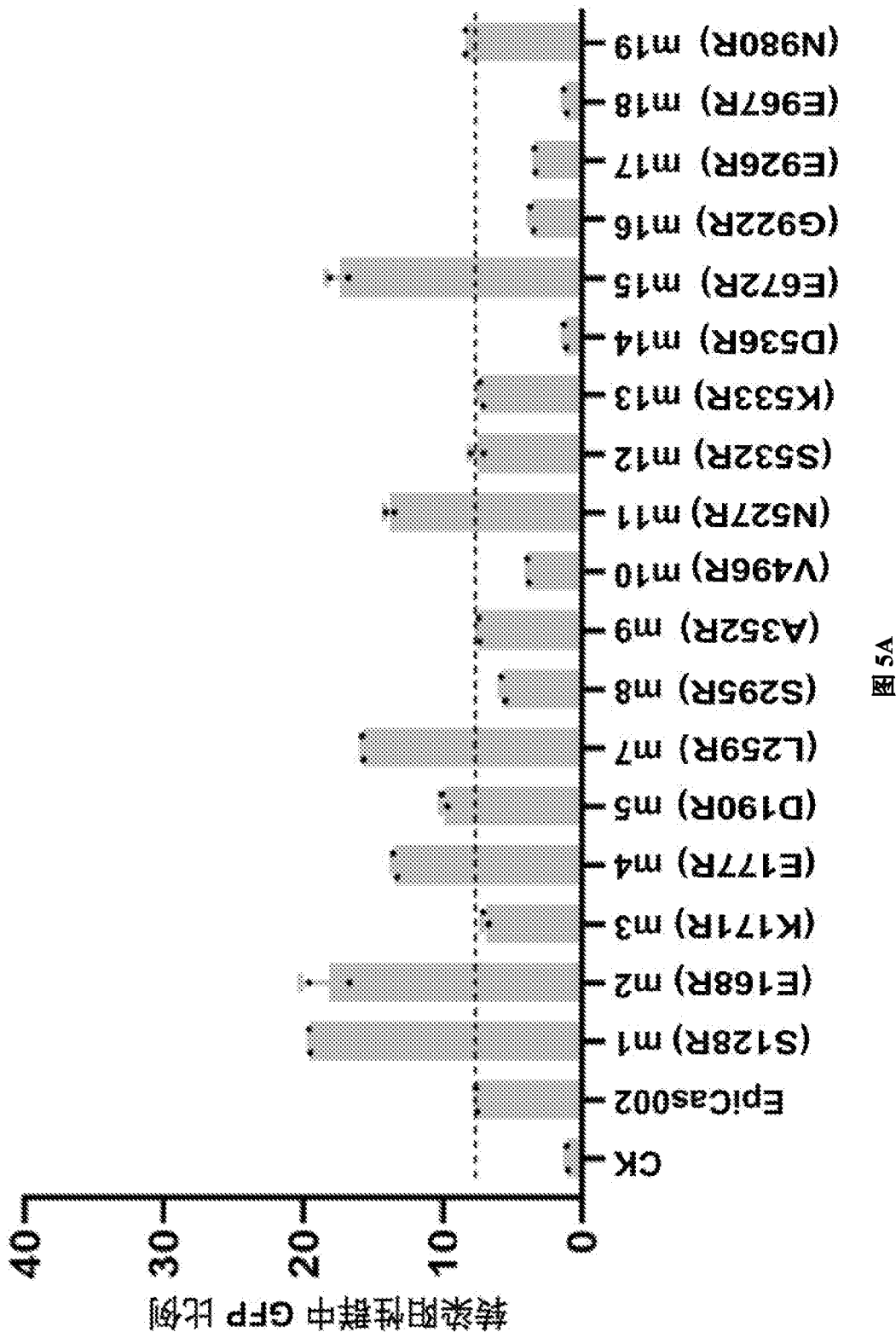
\\ACCCTCTGGGGACCGTTTGAGGAGTGTTTCAGTCTCCGTGAACGTTCCCTTAGCACTCTGCC
TTGGGAGACCCCTGGCAAACCTCCTCACAAGTCAGAGGCACCTTGCAAAGGAATCGTGAGACGG

DNMT1 target



EpiCas045 + target gRNA AsCas12a + target gRNA EpiCas045 + NT gRNA

图 4



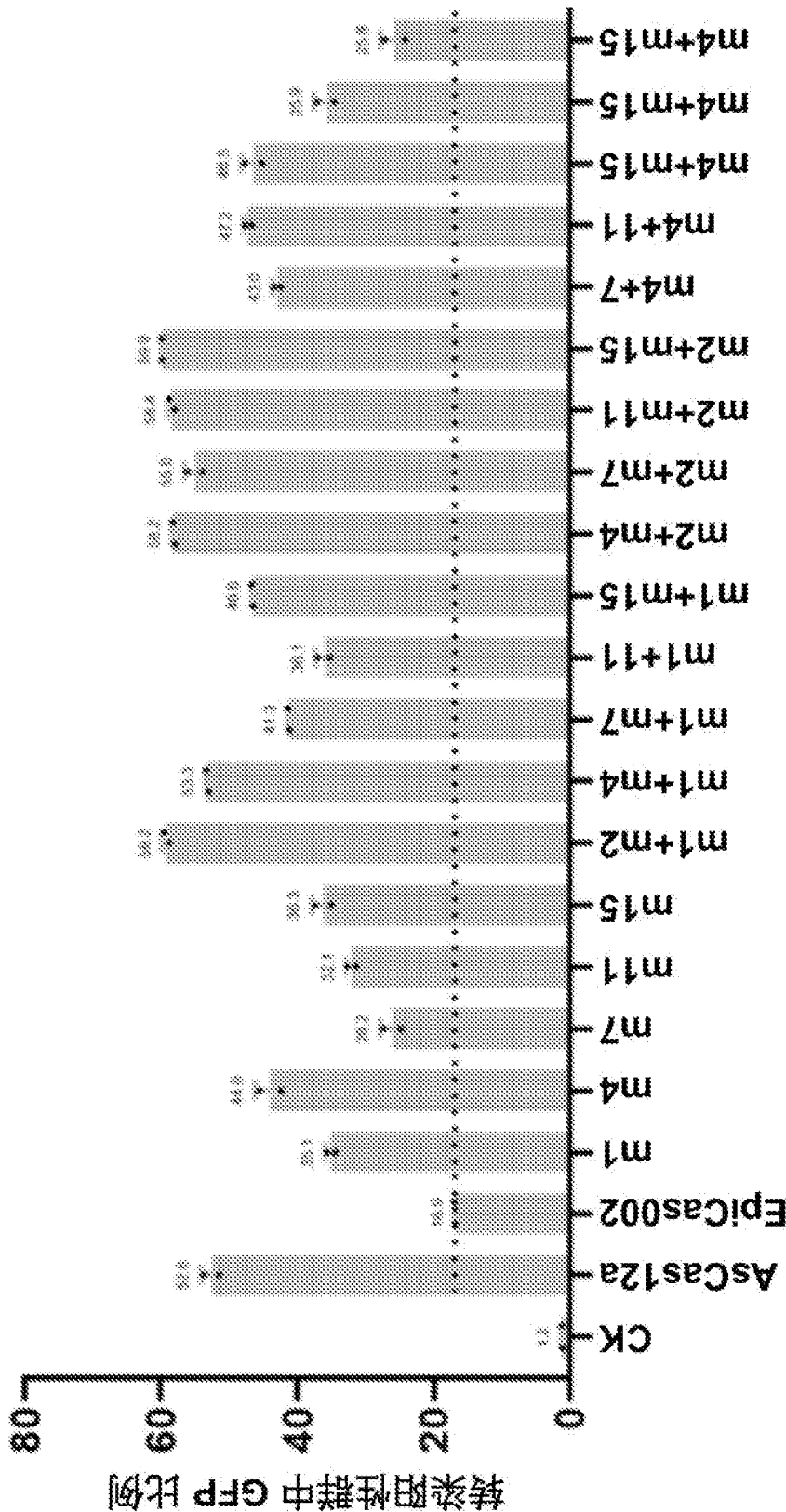


图 5B

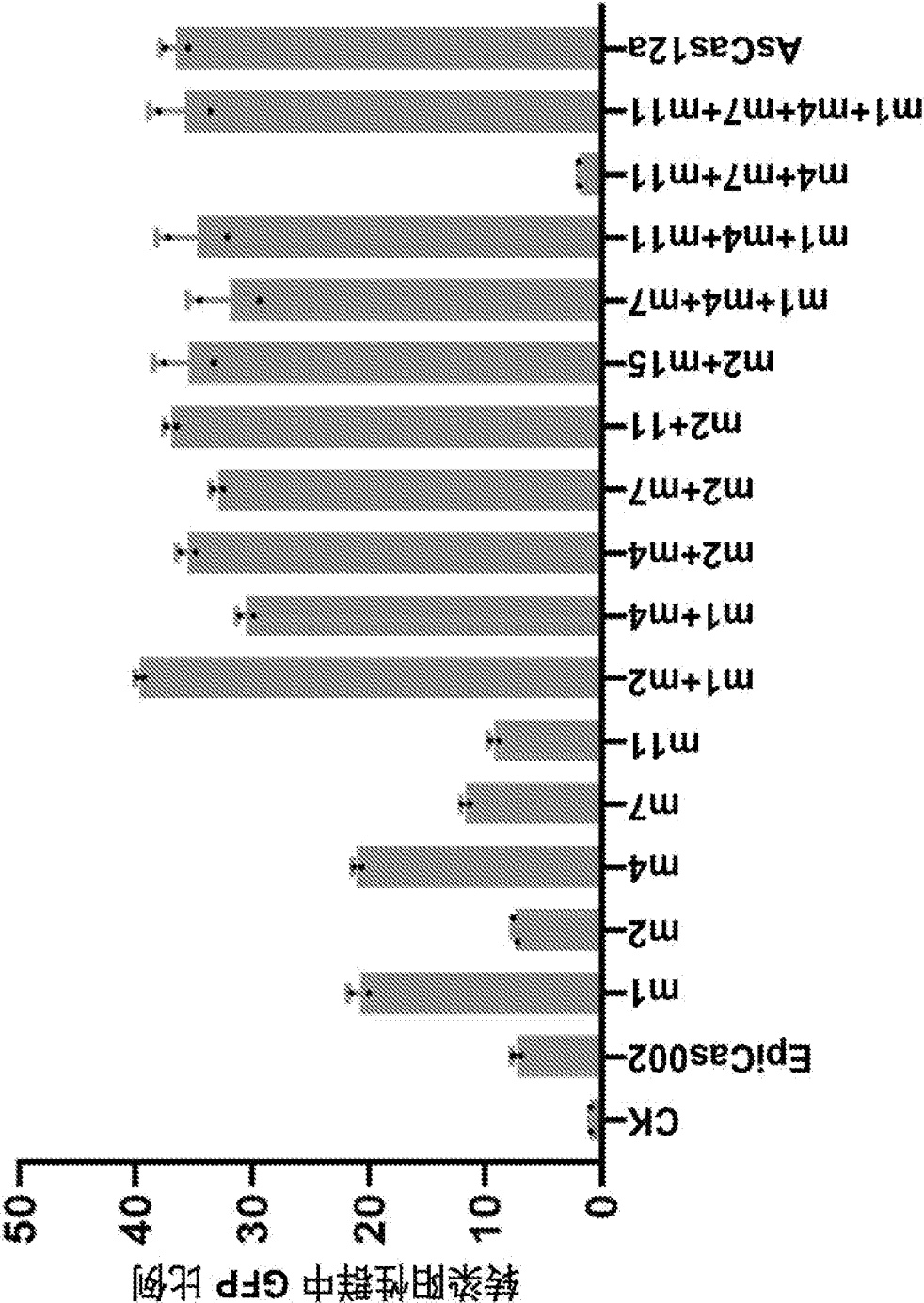


图 5C

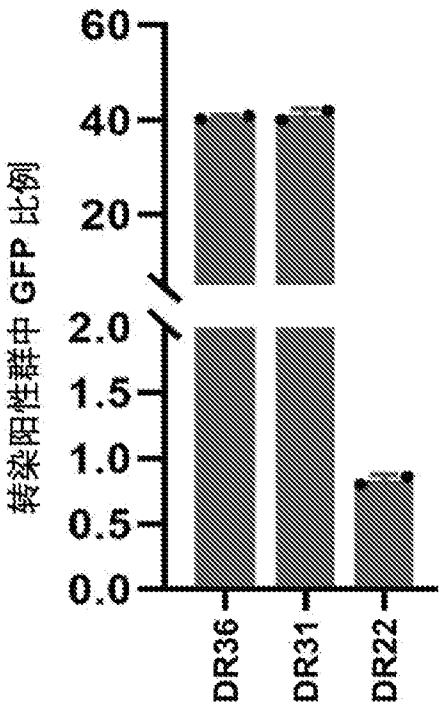


图 6A

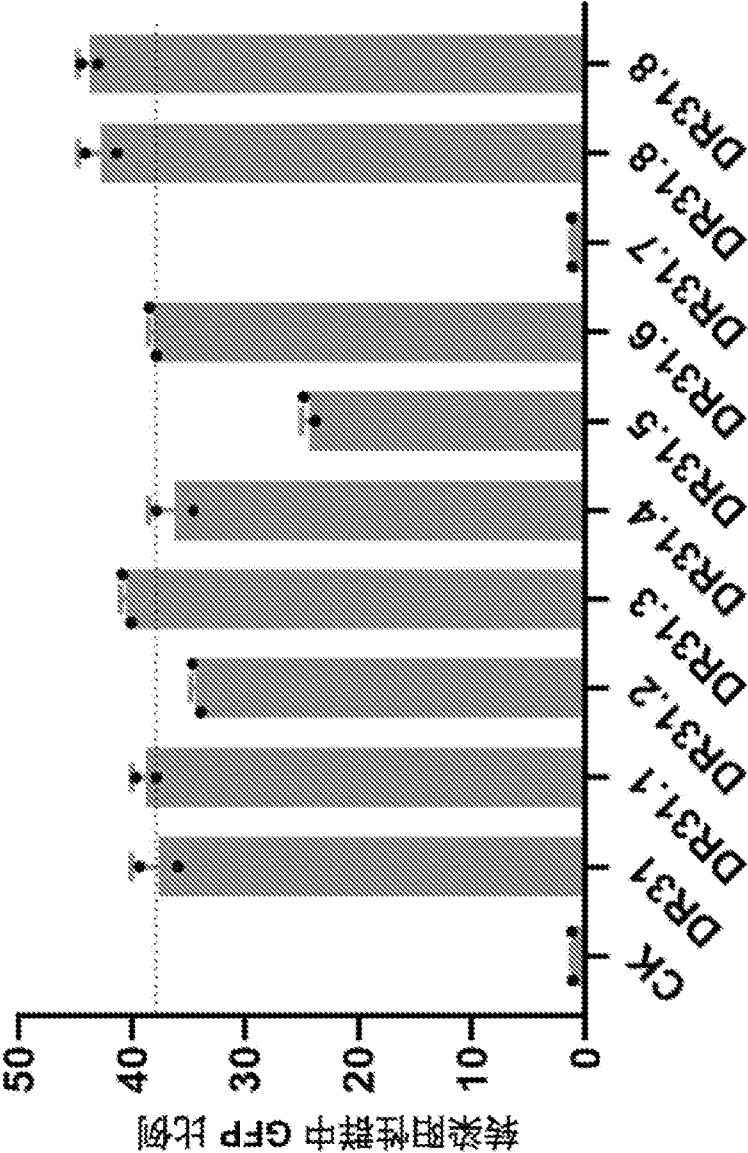
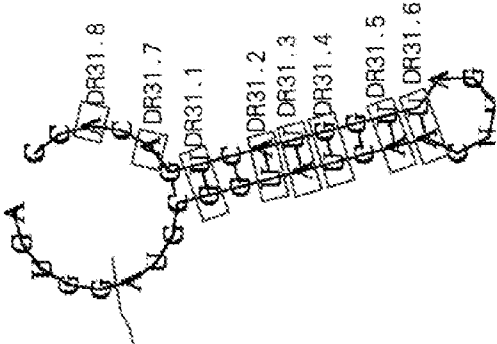


图 6B



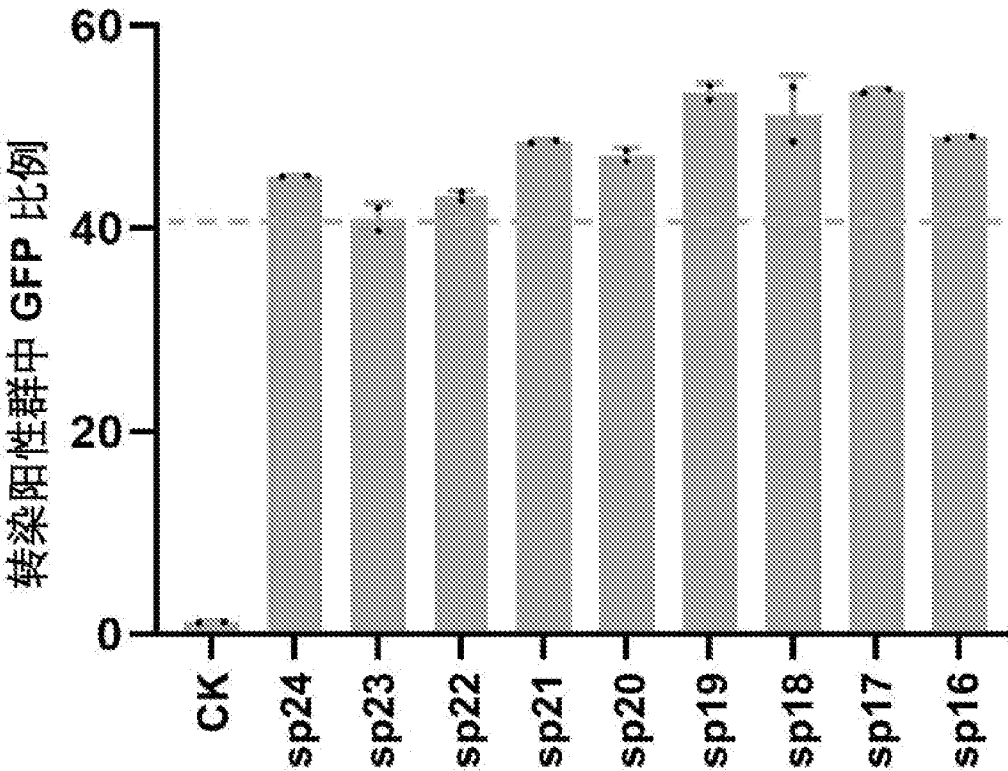


图 7

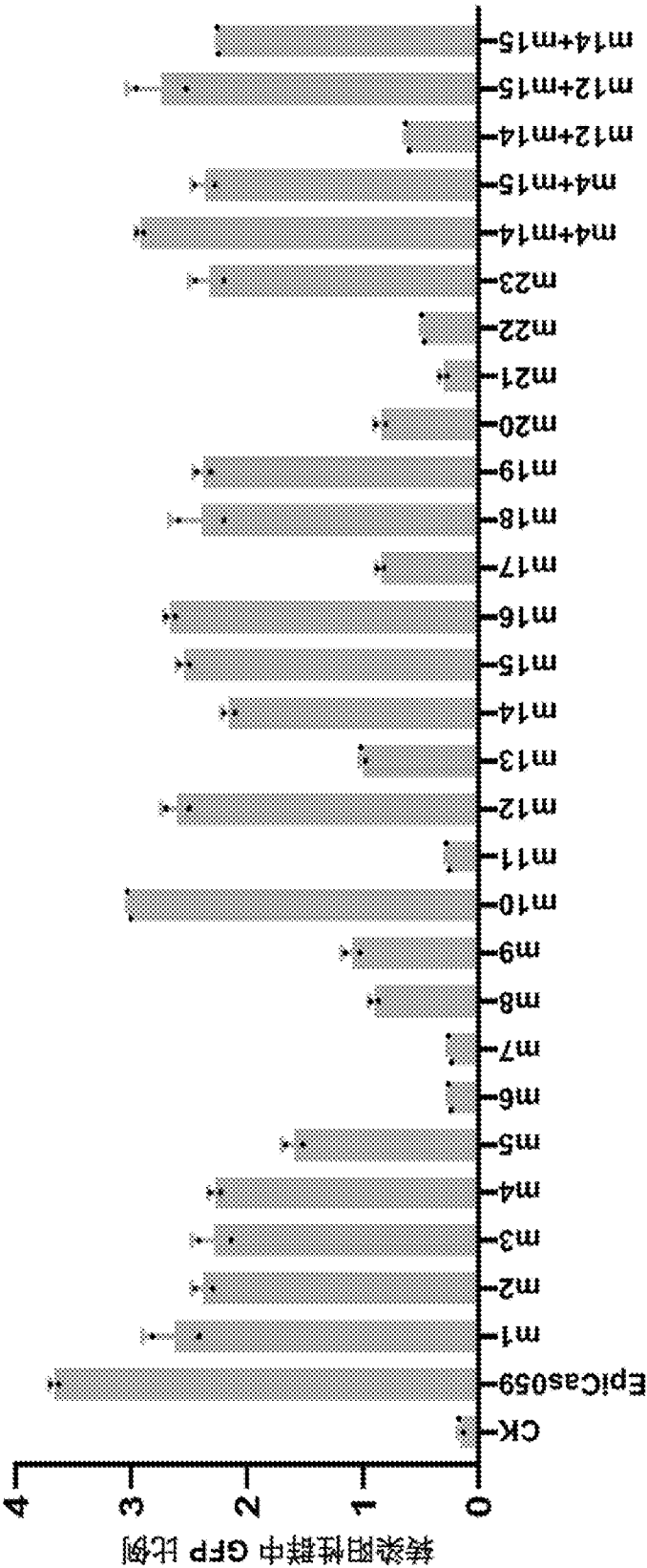


图 8A

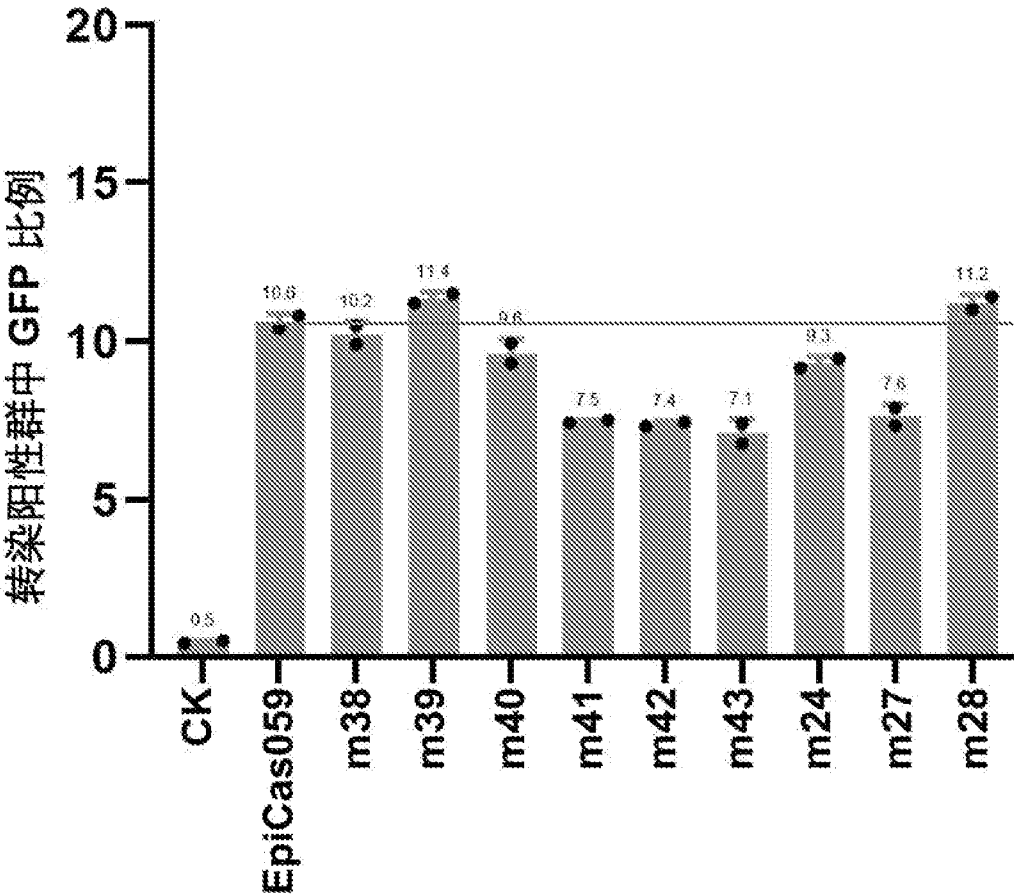


图 8B

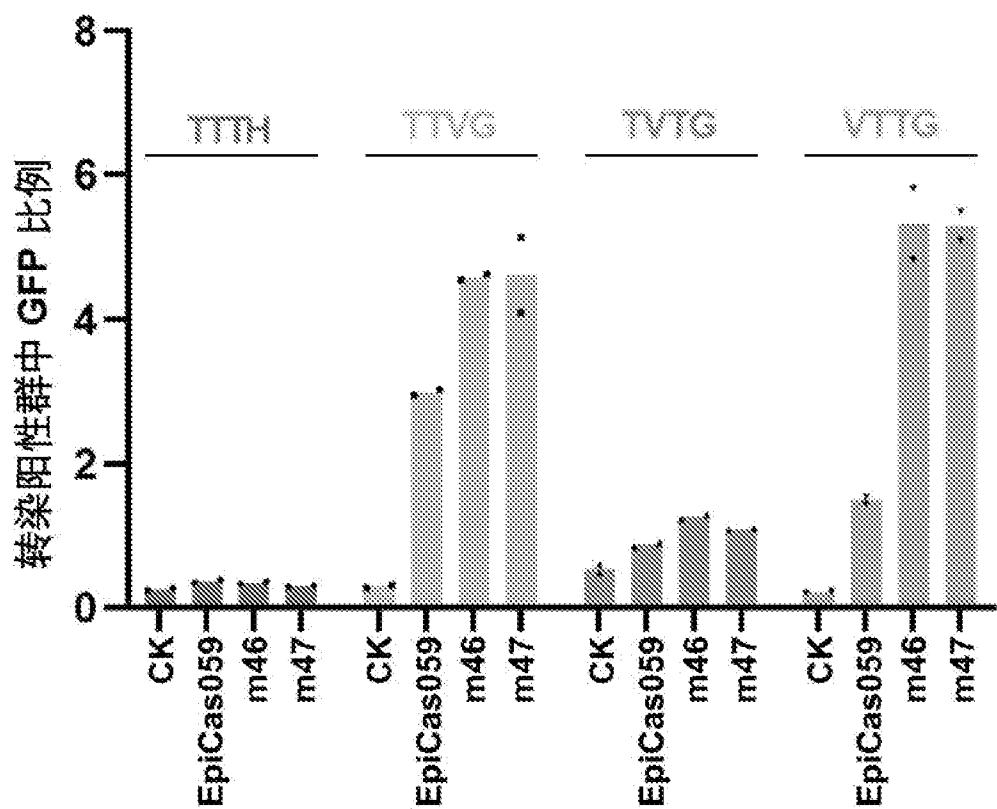


图 8C

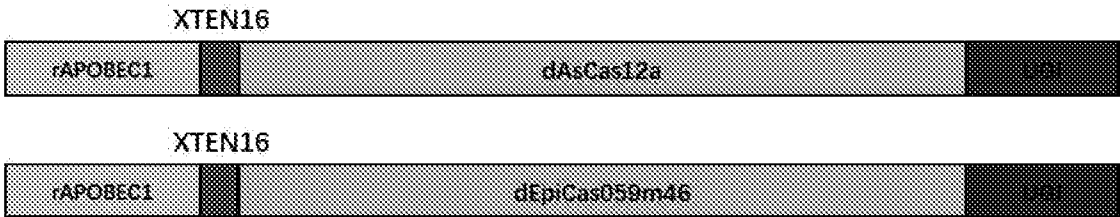


图 9A

EMX1 位点：
TGGTTGC₇C₈C₉AC₁₁C₁₂C₁₃TAGTC₁₈ATTGG

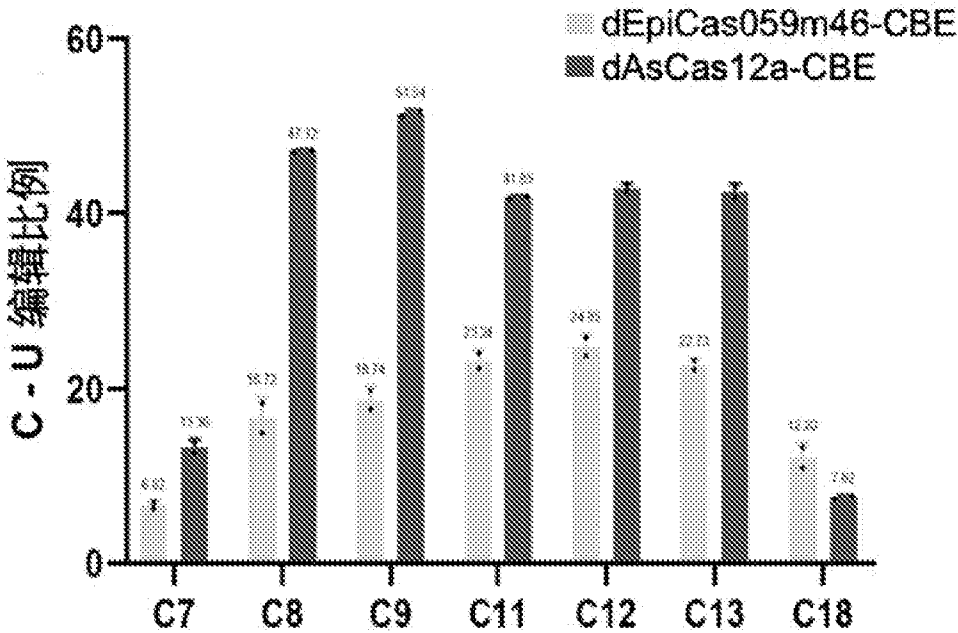


图 9B

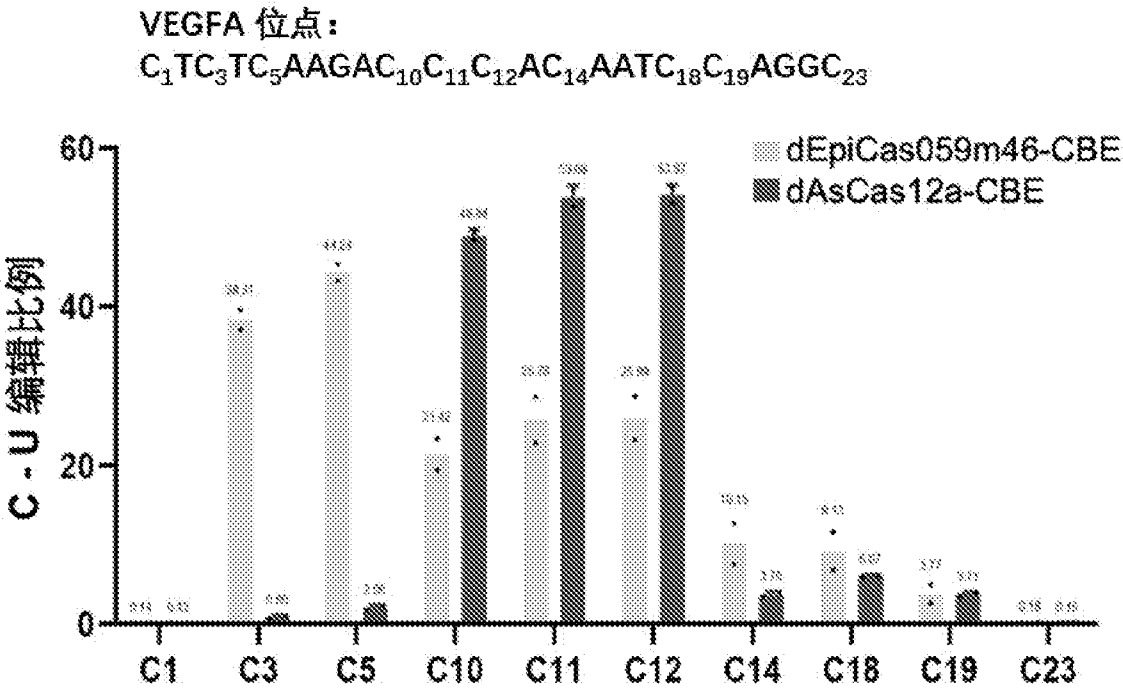


图 9C

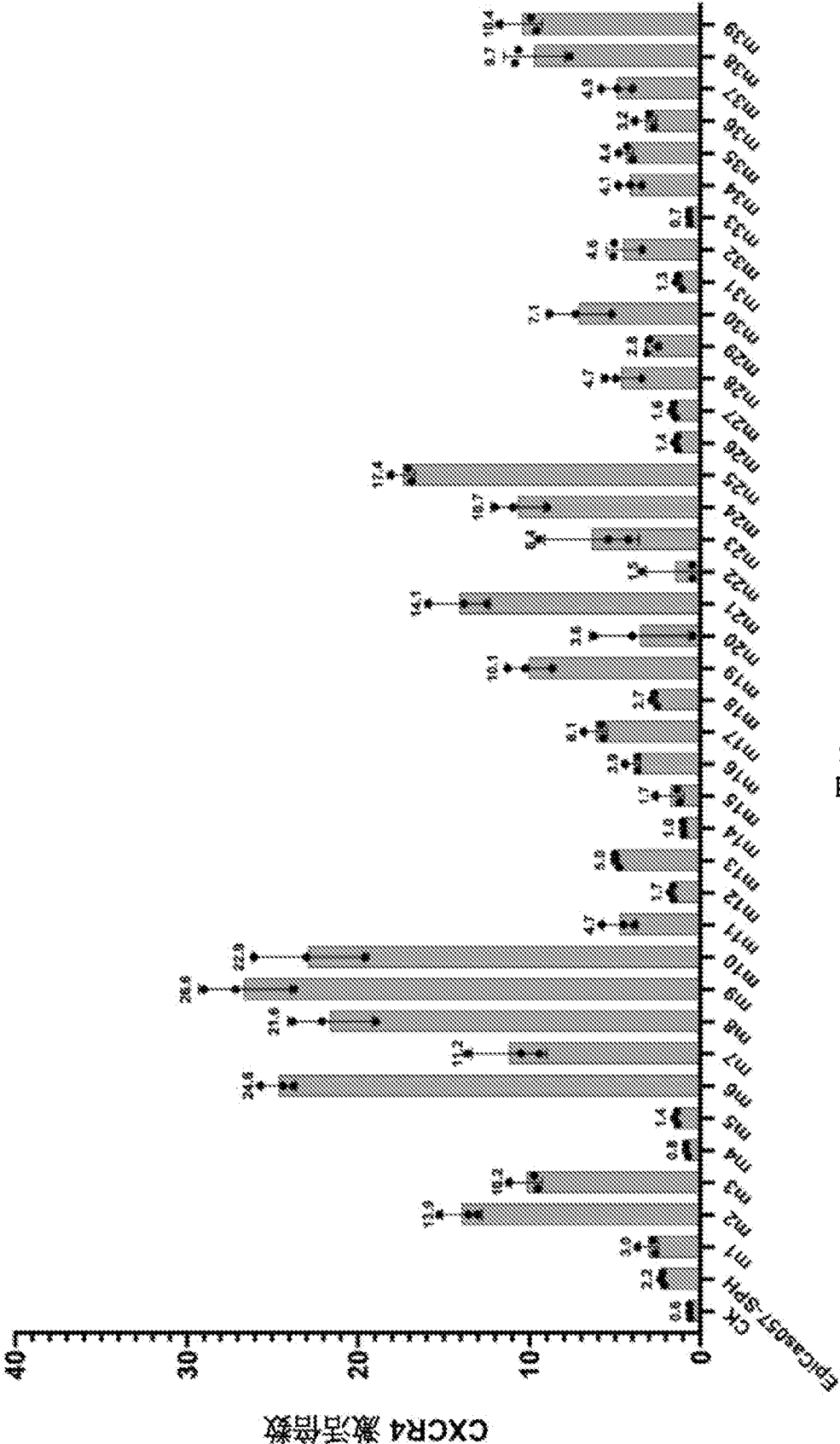


图 10

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2024/097935

A. CLASSIFICATION OF SUBJECT MATTER

C12N 9/22(2006.01)i; C07K 19/00(2006.01)i; C12N 15/113(2010.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC: C12N C07K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNTXT, ENTXT, ENTXTC, VEN, CNKI, NCBI, 万方数据库, WANFANG DATABASE, pubmed, ISI web of science, stn: ascas12a, bfp, cas12a, Cas12i2, Cpf1, gfp, nhej, type V, Cas酶, 成簇规律间隔短回文重复序列, CRISPR, SEQ ID NOS: 1-58, 融合, 结构域

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	LOU, Y. C. "MAG: Type V CRISPR-Associated Protein Cas12a/Cpf1 [Clostridium sp.]" <i>GenBank: MBS6461344.1</i> , 16 May 2021 (2021-05-16), sequence part	1-3
Y	LOU, Y. C. "MAG: Type V CRISPR-Associated Protein Cas12a/Cpf1 [Clostridium sp.]" <i>GenBank: MBS6461344.1</i> , 16 May 2021 (2021-05-16), sequence part	4-47
Y	CN 114517190 A (SHANDONG SHUNFENG BIOTECHNOLOGY CO., LTD.) 20 May 2022 (2022-05-20) claims 1-19, and description, paragraph 0034	4-47
Y	CN 112041444 A (ARBOR BIOTECHNOLOGIES, INC.) 04 December 2020 (2020-12-04) claims 1-7	4-47
A	CN 113930413 A (CHINA AGRICULTURAL UNIVERSITY) 14 January 2022 (2022-01-14) claims 1-20	1-47



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"D" document cited by the applicant in the international application

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

06 September 2024

Date of mailing of the international search report

12 September 2024

Name and mailing address of the ISA/CN

China National Intellectual Property Administration (ISA/
CN)
China No. 6, Xitucheng Road, Jimenqiao, Haidian District,
Beijing 100088

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2024/097935

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2021222140 A1 (G+FLAS LIFE SCIENCES et al.) 22 July 2021 (2021-07-22) claims 1-20	1-47
A	US 11174470 B2 (MAMMOTH BIOSCIENCES, INC.) 16 November 2021 (2021-11-16) claims 1-10	1-47
A	CN 116144631 A (HUAZHONG AGRICULTURAL UNIVERSITY) 23 May 2023 (2023-05-23) claims 1-10	1-47
A	TOTH, E. et al. "Improved LbCas12a Variants with Altered PAM Specificities further Broaden the Genome Targeting Range of Cas12a Nucleases" <i>Nucleic Acids Research</i> , Vol. 48, No. 7, 28 February 2020 (2020-02-28), abstract	1-47

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2024/097935

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:
 - a. ☒ forming part of the international application as filed.
 - b. ☐ furnished subsequent to the international filing date for the purposes of international search (Rule 13ter.1(a)),
☐ accompanied by a statement to the effect that the sequence listing does not go beyond the disclosure in the international application as filed.
2. ☐ With regard to any nucleotide and/or amino acid sequence disclosed in the international application, this report has been established to the extent that a meaningful search could be carried out without a WIPO Standard ST.26 compliant sequence listing.
3. Additional comments:

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2024/097935

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	114517190	A	20 May 2022	WO	2022166895	A1	11 August 2022
CN	112041444	A	04 December 2020	JP	2021516970	A	15 July 2021
				JP	7457653	B2	28 March 2024
				HUE	063005	T2	28 December 2023
				WO	2019178427	A1	19 September 2019
				US	2022098579	A1	31 March 2022
				ES	2952978	T3	07 November 2023
				US	2023183688	A1	15 June 2023
				US	11912992	B2	27 February 2024
				EP	3765615	A1	20 January 2021
				EP	3765615	B1	28 June 2023
				JP	2024019739	A	09 February 2024
				DE	202019005567	U1	16 February 2021
				SI	3765615	T1	30 October 2023
				CA	3093334	A1	19 September 2019
				US	2021123047	A1	29 April 2021
				US	11168324	B2	09 November 2021
				EP	4257696	A2	11 October 2023
				EP	4257696	A3	24 January 2024
				AU	2019236210	A1	10 September 2020
				US	2020063126	A1	27 February 2020
				US	10808245	B2	20 October 2020
				US	2020407715	A1	31 December 2020
				PT	3765615	T	28 August 2023
				DK	3765615	T3	21 August 2023
				PL	3765615	T3	13 November 2023
				LT	3765615	T	10 October 2023
				HRP	20230828	T1	10 November 2023
				US	2020407716	A1	31 December 2020
				FI	3765615	T3	29 August 2023
CN	113930413	A	14 January 2022	None			
US	2021222140	A1	22 July 2021	EP	3833751	A2	16 June 2021
				EP	3833751	A4	27 April 2022
				WO	2020030984	A2	13 February 2020
				WO	2020030984	A3	16 July 2020
				US	11434478	B2	06 September 2022
				CA	3109083	A1	13 February 2020
US	11174470	B2	16 November 2021	US	2022119788	A1	21 April 2022
				WO	2020142754	A2	09 July 2020
				WO	2020142754	A3	17 September 2020
				EP	3931313	A2	05 January 2022
				US	2021009974	A1	14 January 2021
CN	116144631	A	23 May 2023	None			

第I栏 核苷酸和/或氨基酸序列(续第1页第1.c项)

1. 关于国际申请中所公开的任何核苷酸和/或氨基酸序列,国际检索是基于下列序列表进行的:
- a. ☒ 作为国际申请的一部分提交的:

b. ☐ 为国际检索的目的在国际申请日之后提交 (细则13之三.1(a)) ,
☐ 附有说明序列表不超出所提交国际申请公开范围的声明。
2. ☐ 本报告是在没有收到符合WIPO ST.26标准的序列表的情况下,考虑了国际申请中披露的任何核苷酸和/或氨基酸序列,在可进行有意义检索的范围内做出的。
3. 补充意见:

A. 主题的分类

C12N 9/22(2006.01)i; C07K 19/00(2006.01)i; C12N 15/113(2010.01)i

按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

IPC: C12N C07K

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用))

CNTEXT,ENTXT,ENTXTC,VEN,CNKI,NCBI,万方数据库,pubmed,ISI web of science,stm:ascas12a,bfp,cas12a,Cas12i2,Cpf1,gfp,nhej,type V,Cas酶,成簇规律间隔短回文重复序列,CRISPR,SEQ ID NOs: 1-58,融合,结构域

C. 相关文件

类型*	引用文件, 必要时, 指明相关段落	相关的权利要求
X	LOU, Y. C. "MAG: type V CRISPR-associated protein Cas12a/Cpf1 [Clostridium sp.]" GenBank: MBS6461344.1, 2021年5月16日 (2021 - 05 - 16), 序列部分	1-3
Y	LOU, Y. C. "MAG: type V CRISPR-associated protein Cas12a/Cpf1 [Clostridium sp.]" GenBank: MBS6461344.1, 2021年5月16日 (2021 - 05 - 16), 序列部分	4-47
Y	CN 114517190 A (山东舜丰生物科技有限公司) 2022年5月20日 (2022 - 05 - 20) 权利要求1-19, 说明书第0034段	4-47
Y	CN 112041444 A (阿伯生物技术公司) 2020年12月4日 (2020 - 12 - 04) 权利要求1-7	4-47
A	CN 113930413 A (中国农业大学) 2022年1月14日 (2022 - 01 - 14) 权利要求1-20	1-47
A	US 2021222140 A1 (G+FLAS LIFE SCIENCES 等) 2021年7月22日 (2021 - 07 - 22) 权利要求1-20	1-47

☒ 其余文件在C栏的续页中列出。

☒ 见同族专利附件。

* 引用文件的具体类型:

"A" 认为不特别相关的表示了现有技术一般状态的文件

"D" 申请人在国际申请中引证的文件

"E" 在国际申请日的当天或之后公布的在先申请或专利

"L" 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的)

"O" 涉及口头公开、使用、展览或其他方式公开的文件

"p" 公布日先于国际申请日但迟于所要求的优先权日的文件

"T" 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件

"X" 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性

"Y" 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性

"&" 同族专利的文件

国际检索实际完成的日期

2024年9月6日

国际检索报告邮寄日期

2024年9月12日

ISA/CN的名称和邮寄地址

中国国家知识产权局
中国北京市海淀区蓟门桥西土城路6号 100088

授权官员

赵鹏

电话号码 (+86) 010-53961965

PCT/ISA/210 表(第2页) (2022年7月)

C. 相关文件		
类 型*	引用文件，必要时，指明相关段落	相关的权利要求
A	US 11174470 B2 (MAMMOTH BIOSCIENCES, INC.) 2021年11月16日 (2021 - 11 - 16) 权利要求1-10	1-47
A	CN 116144631 A (华中农业大学) 2023年5月23日 (2023 - 05 - 23) 权利要求1-10	1-47
A	TOTH, E. 等. "Improved LbCas12a variants with altered PAM specificities further broaden the genome targeting range of Cas12a nucleases" Nucleic Acids Research, 第48卷, 第7期, 2020年2月28日 (2020 - 02 - 28), 摘要	1-47

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2024/097935

检索报告引用的专利文件			公布日 (年/月/日)	同族专利			公布日 (年/月/日)
CN	114517190	A	2022年5月20日	WO	2022166895	A1	2022年8月11日
CN	112041444	A	2020年12月4日	JP	2021516970	A	2021年7月15日
				JP	7457653	B2	2024年3月28日
				HUE	063005	T2	2023年12月28日
				WO	2019178427	A1	2019年9月19日
				US	2022098579	A1	2022年3月31日
				ES	2952978	T3	2023年11月7日
				US	2023183688	A1	2023年6月15日
				US	11912992	B2	2024年2月27日
				EP	3765615	A1	2021年1月20日
				EP	3765615	B1	2023年6月28日
				JP	2024019739	A	2024年2月9日
				DE	202019005567	U1	2021年2月16日
				SI	3765615	T1	2023年10月30日
				CA	3093334	A1	2019年9月19日
				US	2021123047	A1	2021年4月29日
				US	11168324	B2	2021年11月9日
				EP	4257696	A2	2023年10月11日
				EP	4257696	A3	2024年1月24日
				AU	2019236210	A1	2020年9月10日
				US	2020063126	A1	2020年2月27日
				US	10808245	B2	2020年10月20日
				US	2020407715	A1	2020年12月31日
				PT	3765615	T	2023年8月28日
				DK	3765615	T3	2023年8月21日
				PL	3765615	T3	2023年11月13日
				LT	3765615	T	2023年10月10日
				HRP	20230828	T1	2023年11月10日
				US	2020407716	A1	2020年12月31日
				FI	3765615	T3	2023年8月29日
CN	113930413	A	2022年1月14日	无			
US	2021222140	A1	2021年7月22日	EP	3833751	A2	2021年6月16日
				EP	3833751	A4	2022年4月27日
				WO	2020030984	A2	2020年2月13日
				WO	2020030984	A3	2020年7月16日
				US	11434478	B2	2022年9月6日
				CA	3109083	A1	2020年2月13日
US	11174470	B2	2021年11月16日	US	2022119788	A1	2022年4月21日
				WO	2020142754	A2	2020年7月9日
				WO	2020142754	A3	2020年9月17日
				EP	3931313	A2	2022年1月5日
				US	2021009974	A1	2021年1月14日
CN	116144631	A	2023年5月23日	无			