



(10) 授权公告号 CN 111328343 B

(45) 授权公告日 2023.11.28

(21) 申请号 201880068076.4

(22) 申请日 2018.08.22

(65) 同一申请的已公布的文献号
申请公布号 CN 111328343 A

(43) 申请公布日 2020.06.23

(30) 优先权数据

62/548,846 2017.08.22 US

62/572,963 2017.10.16 US

62/639,178 2018.03.06 US

15/937,699 2018.03.27 US

(85) PCT国际申请进入国家阶段日
2020.04.17

(86) PCT国际申请的申请数据
PCT/US2018/047597 2018.08.22

(87) PCT国际申请的公布数据
W02019/040664 EN 2019.02.28

(73) 专利权人 萨克生物研究学院
地址 美国加利福尼亚州

(72) 发明人 P·D·徐 S·康纳曼

(74) 专利代理机构 北京华睿卓成知识产权代理
事务所(普通合伙) 11436
专利代理师 程淼

(51) Int.Cl.
C12N 9/22 (2006.01)
C12N 15/55 (2006.01)

(56) 对比文件

WO 2016205711 A1,2016.12.22

CN 110914433 A,2020.03.24

Sudarsanam,P.等.hypothetical protein
EUBSIR_02687 [[Eubacterium] siraeum DSM
15702].《Genbank》.2012,

Sudarsanam,P.等.hypothetical protein
EUBSIR_02687 [[Eubacterium] siraeum DSM
15702].《Genbank》.2012,

Liang Liu等.Two Distant Catalytic
Sites Are Responsible for C2c2 RNase
Activities.《Cell》.2017,第168卷

审查员 吴至芳

权利要求书2页 说明书83页
序列表(电子公布) 附图112页

(54) 发明名称

RNA靶向方法和组合物

(57) 摘要

本公开提供了用于靶向RNA分子的CRISPR/
Cas方法和组合物,其可以用于检测、编辑或修饰
靶RNA。

1. 分离的CRISPR-Cas系统向导RNA (gRNA)或编码所述gRNA的核酸,包括一个或更多个同向重复 (DR) 序列和一个或更多个间隔子序列,

其中所述一个或更多个DR序列包括:

一个或更多个DR序列,所述DR序列由SEQ ID NO: 131的核酸序列组成;或

一个或更多个DR序列,所述DR序列由在SEQ ID NO: 131的核酸序列的5'端截短1、2、3、4、5或6个核苷酸的截短形式组成;

其中所述一个或更多个间隔子序列与一个或更多个靶RNA分子杂交,其中所述gRNA能够与Cas13d蛋白形成复合物,并且其中所述gRNA将所述复合物引导至所述一种或多种靶RNA分子。

2. 根据权利要求1所述的分离的CRISPR-Cas系统gRNA或编码所述gRNA的核酸,包括:

一个或更多个DR序列,所述DR序列由在SEQ ID NO: 131的核酸序列的5'端截短1-6个核苷酸的截短形式组成。

3. 根据权利要求1所述的分离的CRISPR-Cas系统gRNA或编码所述gRNA的核酸,包括:

一个或更多个DR序列,所述DR序列由在SEQ ID NO: 131的核酸序列的5'端截短5-6个核苷酸的截短形式组成。

4. 根据权利要求1所述的分离的CRISPR-Cas系统gRNA或编码所述gRNA的核酸,包括:

一个或更多个DR序列,所述DR序列由SEQ ID NO: 131的核酸序列组成。

5. 根据权利要求1-4任一项所述的分离的CRISPR-Cas系统gRNA或编码所述gRNA的核酸,进一步包括适配体序列。

6. 根据权利要求5所述的分离的CRISPR-Cas系统gRNA或编码所述gRNA的核酸,其中所述适配体序列包括荧光团适配体。

7. 根据权利要求5所述的分离的CRISPR-Cas系统gRNA或编码所述gRNA的核酸,其中所述适配体序列包括MS2、PP7或Q β 。

8. 根据权利要求1-4任一项所述的分离的CRISPR-Cas系统gRNA或编码所述gRNA的核酸,其中所述一个或更多个间隔子序列每个包括28-32个核苷酸。

9. 根据权利要求1-4任一项所述的分离的CRISPR-Cas系统gRNA或编码所述gRNA的核酸,其中所述一个或更多个间隔子序列每个包括10-30个核苷酸。

10. 根据权利要求1-4任一项所述的分离的CRISPR-Cas系统gRNA或编码所述gRNA的核酸,其中所述gRNA包括DR-间隔子-DR-间隔子-DR-间隔子的序列。

11. 非细菌细胞,包括根据权利要求1-10任一项所述的分离的CRISPR-Cas系统gRNA或编码所述gRNA的核酸。

12. 根据权利要求11所述的非细菌细胞,其中所述非细菌细胞是哺乳动物细胞。

13. 根据权利要求12所述的非细菌细胞,其中所述哺乳动物细胞是人细胞。

14. 根据权利要求11所述的非细菌细胞,其中所述非细菌细胞是植物细胞、昆虫细胞、真菌细胞或酵母细胞。

15. 细菌细胞,包括根据权利要求1-10任一项所述的分离的CRISPR-Cas系统gRNA或编码所述gRNA的核酸,且所述gRNA对于所述细菌细胞不是天然的。

16. 根据权利要求15所述的细菌细胞,其中所述细菌细胞是大肠杆菌。

17. 组合物,包括:

(a) 根据权利要求1-10任一项所述的分离的CRISPR-Cas系统gRNA或编码所述gRNA的核酸,

根据权利要求11-14任一项所述的非细菌细胞,或
根据权利要求15或16所述的细菌细胞,和

(b) 药学上可接受的载体。

18. 试剂盒,包括:

根据权利要求1-10任一项所述的分离的CRISPR-Cas系统gRNA或编码所述gRNA的核酸,
根据权利要求11-14任一项所述的非细菌细胞,
根据权利要求15或16所述的细菌细胞,或
根据权利要求17所述的组合物。

19. 根据权利要求18所述的试剂盒,其进一步包括递送系统、标记物或其组合。

20. 根据权利要求19所述的试剂盒,其中所述递送系统包括脂质体、颗粒、外来体、微泡、病毒载体或质粒。

21. 根据权利要求19所述的试剂盒,其中所述标记物包括肽或抗体。

22. 根据权利要求18-21任一项所述的试剂盒,进一步包括细胞培养基或生长培养基。

RNA靶向方法和组合物

[0001] 相关申请的交叉引用

[0002] 本申请要求2017年8月22日提交的美国临时申请第62/548,846号、2017年10月16日提交的美国临时申请第62/572,963号、2018年3月6日提交的美国临时申请第62/639,178号和2018年3月27日提交的美国临时申请第15/937,699号的优先权,这些申请均通过引用全文结合入本文。

技术领域

[0003] 本公开涉及用于修饰(包括检测)RNA的CRISPR/Cas系统,其利用新的Cas13d蛋白质(也称作CasR和nCas1)和向导RNA。

[0004] 对政府资助的致谢

[0005] 本发明在美国国家卫生研究院(The National Institutes of Health)授予的5DP5 0D021369-02和5R21 AG056811-02的政府资助下做出。美国政府对本发明拥有某些权利。

背景技术

[0006] 从微阵列(Schena et al.,1995)到新一代测序和单细胞研究(Shendure et al.,2017),过去20年来的技术进步已经转变了在细胞功能和疾病中转录组变化的作图(mapping)。但是,质询单个转录物动力学的功能以及在观察到的转录变化和细胞表型之间建立因果联系均要求积极控制或调节所需转录物的能力。

[0007] DNA改造技术例如CRISPR-Cas9(Doudna and Charpentier,2014;Hsu et al.,2014)使得研究者能够剖析特定遗传元件的功能或纠正致病突变。但是,研究和操纵RNA的简单且可扩展的工具显著落后于其DNA对应物。能够切割或抑制所需转录物的现有RNA干扰技术,具有明显的脱靶效应,并且由于其在内源性过程中的关键作用仍然是具有挑战性的改造靶标(Birmingham et al.,2006;Jackson et al.,2003)。因此,直接研究RNA功能作用的方法仍然有限。

[0008] RNA改造中的关键限制之一是缺乏可以易于被重新靶向并引入到靶细胞中的RNA-结合结构域。例如,MS2 RNA-结合结构域识别不变的21-核苷酸(nt)RNA序列(Peabody,1993),因此要求进行基因组修饰,以标记所需的转录物。Pumilio同源性结构域具有模块化重复序列,其中每个蛋白质模块识别单独的RNA碱基,但是它们仅可以靶向短的8nt RNA序列(Cheong and Hall,2006)。尽管之前表征的II型(Batra et al.,2017;O'Connell et al.,2014)和VI型(Abudayyeh et al.,2016;East-Seletsky et al.,2016)CRISPR-Cas系统可以被再编程,以识别20-30nt RNA,但是它们的大尺寸(~1200氨基酸,aa)使其难以包装到用于原代细胞和体内递送的AAV中。

发明内容

[0009] 本申请提供原核生物基因组的生物信息学分析,以鉴定CRISPR-Cas重复序列阵列

的序列特征,并挖掘之前未表征的可以用于RNA靶向工具的紧凑型Cas核糖核酸酶。改造的VI-D型CRISPR效应子可以用于有效地敲低(knockdown)人细胞中的内源性RNA,并操纵可变剪接,为RNA靶向应用以及进一步效应子结构域融合(作为转录组改造工具箱的部分)创造条件。

[0010] 本文提供靶向一种或多种靶RNA分子的方法,例如规律间隔成簇短回文重复序列(CRISPR)-相关的(Cas)系统-介导的RNA编辑方法。这些方法可以包括:使一种或多种靶RNA分子接触非天然(例如,在其引入的细胞或系统中并不天然地存在)或改造的CRISPR-Cas系统。这种CRISPR-Cas系统可以包括:(1)本文提供的至少一种Cas13d蛋白质或至少一种Cas13d核酸编码序列(例如编码所述至少一种Cas13d蛋白质的mRNA或载体);和(2)至少一种与所述一种或多种靶RNA分子杂交的CRISPR-Cas系统向导核酸分子(例如向导RNA,gRNA),或至少一种编码所述gRNA的核酸分子。Cas13d蛋白质与gRNA形成复合物,并且gRNA将该复合物引导至一种或多种靶RNA分子,并修饰(例如,切割、检测)所述一种或多种靶RNA分子。在一些实例中,一种或多种靶RNA分子(或含有一种或多种靶RNA分子的细胞)与包括至少一种Cas13d蛋白质和至少一种gRNA的复合物接触。在一些实例中,所述系统包括Mg²⁺。但是,在一些实例中,例如,如果不需要切割靶RNA,所述系统不包括Mg²⁺。

[0011] 在一些实例中,使一种或多种靶RNA分子与非天然或改造的CRISPR-Cas系统接触包括:例如,使用内吞作用、脂质体、颗粒、外来体、微泡、基因枪、电穿孔、病毒或其组合,向含有一种或多种靶RNA分子的细胞(例如真核细胞或原核细胞)中引入非天然或改造的CRISPR-Cas系统。在一些实例中,使一种或多种靶RNA分子与非天然或改造的CRISPR-Cas系统接触包括:使含有一种或多种靶RNA分子的无细胞系统(例如生物或环境样品,或细胞裂解物)与非天然或改造的CRISPR-Cas系统接触(例如,在检测靶RNA的诊断方法中)。

[0012] 在一些实例中,至少一种Cas13d蛋白质:(1)包括一个或多个HEPN结构域,(2)不大于150kD(例如,不大于140kD,不大于130kD,不大于120kD,例如大约90-120kD,大约100-120kD,或者大约110kD),(3)包括一个或多个突变的HEPN结构域,并且可以加工向导RNA,但不能裂解或切割一个或多个靶RNA分子,(4)包括来自原核生物基因组或宏基因组(metagenome)、肠道宏基因组、活性污泥宏基因组(an activated sludge metagenome)、厌氧消化池宏基因组(an anaerobic digester metagenome)、鸡肠道宏基因组、人肠道宏基因组、猪肠道宏基因组、牛肠道宏基因组、绵羊肠道宏基因组、山羊肠道宏基因组、水豚肠道宏基因组、灵长类肠道宏基因组、白蚁肠道宏基因组、粪便宏基因组、来自梭菌目(Order Clostridiales)或瘤胃菌科(Family Ruminococcaceae)的基因组的Cas13d直向同源物,(5)包括来自白色瘤胃球菌(*Ruminococcus albus*)、惰性真杆菌(*Eubacterium siraeum*)、黄化瘤胃球菌(*Ruminococcus flavefaciens*)XPD3002株、黄化瘤胃球菌FD-1、未培养的真杆菌属(*Eubacterium* sp)TS28-c4095、未培养的瘤胃球菌属(*Ruminococcus* sp.)、双环瘤胃球菌(*Ruminococcus bicirculans*)或瘤胃球菌属CAG57的Cas13d直向同源物,(6)包括与SEQ ID NO:1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100、101、102、103、104、105、106、107、108、109、110、111、112、113、138、147、149、153、155、158、

160、162、164、166、168、170、175、177、179、181、183、185、187、189、194、198、200、202、204、206、208、210、212、214、216、218、220、222、224、226、229、231、233、235、237、239、241、243、245、247、249、251、253、278、279、280、281、282、283、284、285、292、293、294、295或296至少80%、至少85%、至少90%、至少92%、至少95%、至少96%、至少97%、至少98%、至少99%或100%的序列同一性,或者(7)其组合(例如,1-6中任意者的组合)。在一些实例中,至少一种Cas13d蛋白质与SEQ ID NO:1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100、101、102、103、104、105、106、107、108、109、110、111、112、113、138、147、149、153、155、158、160、162、164、166、168、170、175、177、179、181、183、185、187、189、194、198、200、202、204、206、208、210、212、214、216、218、220、222、224、226、229、231、233、235、237、239、241、243、245、247、249、251、253、278、279、280、281、282、283、284、285、292、293、294、295或296具有至少80%、至少85%、至少90%、至少92%、至少95%、至少96%、至少97%、至少98%、至少99%或100%的序列同一性,并且包括至少一个SEQ ID NO:195、196或197所示的基序。在一些实例中,至少一种Cas13d蛋白质与SEQ ID NO:1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100、101、102、103、104、105、106、107、108、109、110、111、112、113、138、147、149、153、155、158、160、162、164、166、168、170、175、177、179、181、183、185、187、189、194、198、200、202、204、206、208、210、212、214、216、218、220、222、224、226、229、231、233、235、237、239、241、243、245、247、249、251、253、278、279、280、281、282、283、284、285、292、293、294、295或296具有至少80%、至少85%、至少90%、至少92%、至少95%、至少96%、至少97%、至少98%、至少99%或100%的序列同一性,并且包括至少一个SEQ ID NO:288、289、290或291所示的共有序列。在一些实例中,至少一种Cas13d蛋白质进一步包括一种或多种其他药剂(例如,是融合蛋白),例如,一个或多个亚细胞定位信号、一个或多个效应子结构域或其组合。在一些实例中,所述包括一个或多个HEPN结构域的至少一种Cas13d蛋白质为不大于1500aa,不大于1200aa,不大于1100aa,不大于1000aa,例如大约800-1500aa,大约800-1250aa或者大约850-950aa。

[0013] 另外提供了编码这些Cas13d蛋白质的分离的核酸分子,例如cDNA、基因组DNA、RNA或mRNA。这些分离的核酸分子可以是载体(例如质粒或病毒载体)的部分,并且可以与启动子或增强子元件(或二者)可操作地连接。在一些实例中,编码Cas13d蛋白质的分离的核酸分子与SEQ ID NO:124、125、126、127、128、139、140或141具有至少80%、至少85%、至少90%、至少92%、至少95%、至少96%、至少97%、至少98%、至少99%或100%的序列同一性。在一些实例中,编码至少一种Cas13d蛋白质的分离的核酸分子(其可以是载体的部分)包括至少一个经密码子优化以在真核细胞例如人细胞中表达的Cas13d蛋白质编码序列,例如,与SEQ ID NO:114、115、116、117、118、119、120、121、122、123、142、143、144或145具有至少80%、至少85%、至少90%、至少92%、至少95%、至少96%、至少97%、至少98%、至少

99%或100%的序列同一性的Cas13d编码序列。

[0014] 在一些实例中,以Cas13d-介导的方式与一种或多种靶RNA分子杂交的gRNA包括:一个或多个同向重复(Direct Repeat,DR)序列,一个或多个间隔子序列,例如一个或多个包括DR-间隔子-DR-间隔子阵列的序列。在一些实例中,一个或多个DR序列与SEQ ID NO:129、130、131、132、133、134、135、136、137、148、150、151、152、154、156、157、159、161、163、165、167、169、176、178、180、182、184、186、188、190、191、192、193、199、201、203、205、207、209、211、213、215、217、219、221、223、225、227、228、230、232、234、236、238、240、242、244、246、248、250、252或254具有至少80%、至少85%、至少90%、至少92%、至少95%、至少96%、至少97%、至少98%、至少99%或100%的序列同一性。在一实例中,gRNA包括额外的序列,例如适配体序列(例如MS2茎环)。

[0015] 在一些实例中,多个gRNA从单一阵列产生,其中每个gRNA可以是不同的,例如,靶向不同的RNA或者靶向单一RNA的多个区,或者其组合。

[0016] 提供了靶向一种或多种靶RNA分子的方法。在一些实例中,靶向整个RNA。在一些实例中,靶向RNA的部分。靶向RNA分子可以包括以下的一种或多种:切割一种或多种靶RNA分子或者使一种或多种靶RNA分子产生切口,激活一种或多种靶RNA分子,使一种或多种靶RNA分子失活,可视化或检测一种或多种靶RNA分子,标记一种或多种靶RNA分子,结合一种或多种靶RNA分子,编辑一种或多种靶RNA分子,运输一种或多种靶RNA分子,和掩蔽所述一种或多种靶RNA分子。在一些实例中,修饰一种或多种靶RNA分子包括以下的一种或多种:RNA碱基置换,RNA碱基缺失,RNA碱基插入,靶RNA中的断裂,将RNA甲基化,和将RNA去甲基化。

[0017] 在一些实例中,这些方法用于治疗疾病,例如,人的疾病。在这些实例中,一种或多种靶RNA分子与疾病有关。

[0018] 另外提供了分离的蛋白质,包括非天然的蛋白质。在一些实例中,蛋白质与SEQ ID NO:1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100、101、102、103、104、105、106、107、108、109、110、111、112、113、138、147、149、153、155、158、160、162、164、166、168、170、175、177、179、181、183、185、187、189、194、198、200、202、204、206、208、210、212、214、216、218、220、222、224、226、229、231、233、235、237、239、241、243、245、247、249、251、253、278、279、280、281、282、283、284、285、292、293、294、295或296具有至少80%、至少85%、至少90%、至少92%、至少95%、至少96%、至少97%、至少98%、至少99%或100%的序列同一性。在一些实例中,本公开涵盖的分离的蛋白质包括与SEQ ID NO:195、196、197、288、289、290或291具有至少80%、至少85%、至少90%、至少92%、至少95%、至少96%、至少97%、至少98%、至少99%或100%的序列同一性的序列。在一些实例中,分离的蛋白质是来自原核生物基因组或宏基因组、肠道宏基因组、活性污泥宏基因组、厌氧消化池宏基因组、鸡肠道宏基因组、人肠道宏基因组、猪肠道宏基因组、牛肠道宏基因组、绵羊肠道宏基因组、山羊肠道宏基因组、水豚肠道宏基因组、灵长类肠道宏基因组、白蚁肠道宏基因组、粪便宏基因组、来自梭菌目或瘤胃菌科的基因组的Cas13d直向同源物。在一些实例中,Cas13d直向同源物包括来自白色瘤胃球菌、惰性真杆菌、黄化瘤胃球菌XPD3002株、黄化瘤

胃球菌FD-1、未培养的真杆菌属TS28-c4095、未培养的瘤胃球菌属、双环瘤胃球菌或瘤胃球菌属CAG57的Cas13d直向同源物。Cas13d蛋白质可以进一步包括其他药剂或结构域(例如,是融合蛋白),例如,一个或多个亚细胞定位信号、一个或多个效应子结构域或者其组合。

[0019] 另外提供了分离的向导RNA(gRNA)分子。在一些实例中,分离的gRNA包括一个或多个同向重复(DR)序列,例如,未加工(例如,大约36nt)或加工的DR(例如,大约30nt)。在一些实例中,DR与SEQ ID NO:129、130、131、132、133、134、135、136、137、148、150、151、152、154、156、157、159、161、163、165、167、169、176、178、180、182、184、186、188、190、191、192、193、199、201、203、205、207、209、211、213、215、217、219、221、223、225、227、228、230、232、234、236、238、240、242、244、246、248、250、252或254具有至少80%、至少85%、至少90%、至少92%、至少95%、至少96%、至少97%、至少98%、至少99%或100%的序列同一性。这样的gRNA可以进一步包括一个或多个对靶RNA特异(例如,与其互补)的间隔子序列。

[0020] 另外提供了核糖核蛋白(RNP)复合物,其包括本文提供的Cas13d蛋白质和本文提供的gRNA。

[0021] 另外提供了重组细胞,其包括本文提供的任意Cas13d蛋白质(或编码Cas13d的核酸分子)、任意gRNA、任意RNP复合物或任意载体。在一实例中,细胞不是细菌细胞。在一实例中,细胞是细菌细胞。

[0022] 另外提供了组合物,其包括本文提供的任意Cas13d蛋白质(或编码Cas13d的核酸分子)、任意gRNA、任意RNP复合物、任意分离的核酸分子、任意载体或任意细胞中的一种或多种。这些组合物可以包括药学可接受的载体。

[0023] 另外提供了试剂盒。这些试剂盒可以包括本文提供的任意Cas13d蛋白质(或编码Cas13d的核酸分子)、任意gRNA、任意RNP复合物、任意分离的核酸分子、任意载体、任意细胞或任意组合物中的一种或多种。这些试剂可以是合并的,或者处于单独的容器中。

[0024] 从以下参考附图进行的详细描述中,本公开的前述和其它目的和特征将变得更加明显。

附图说明

[0025] 图1A-1B:用于鉴定RNA-靶向的2类CRISPR系统Cas13d的生物信息学流程。(A)说明CRISPR系统鉴定的计算流程的示意图。使用推定2类CRISPR基因座的最小定义,其仅要求CRISPR重复阵列和长度>750的邻近蛋白质(nearby protein)。初步搜索在来源于NCBI Genome的原核生物基因组组装(genome assemblies)上进行,随后针对无预测开放阅读框的公共宏基因组序列经由预测Cas13d蛋白质的TBLASTN进行扩展。DR,同向重复。(B)全长Cas13d效应子和宏基因组片段的系统进化(Phylogenetic)分类和比对。Cas13d效应子和宏基因组Cas13d蛋白质片段聚类成若干不同的分支,为方便理解将其加以颜色。阴影表示使用B1osum62矩阵的残基保守性。该研究中使用的全长Cas13d效应子取样自Cas13d家族的不同分支。Cas13d蛋白质和蛋白质片段的比对使用ClustalOmega 1.2.4进行,并且最大似然树的构建(maximum-likelihood tree building)用PhyML 3.2进行。

[0026] 图2A-2C:VI型CRISPR-Cas13d是单一效应子CRISPR核糖核酸酶家族。(A)本文使用的Cas13d效应子的最大似然系统进化树(maximum-likelihood phylogenetic tree),其中完整的Cas13d CRISPR基因座连同保守的HEPN RNA酶结构域一起示出。灰色矩形表示

CRISPR同向重复(DR),蓝色菱形表示间隔子序列。(B)来自未培养的瘤胃球菌属样品的异源表达的Cas13d基因座的RNA测序。对CRISPR阵列的成熟gRNA作图表明了经加工的30nt DR和14-26nt的可变间隔子长度。同向重复截短的共折叠分析表明强的发夹结构。(C)纯化的惰性真杆菌(*E. siraeum*) Cas 13d和催化死亡的Cas13d (dCas13d)蛋白质各自足以将向导阵列加工成其两个组件gRNA。加入EDTA不损害gRNA加工。‘d’, dCas13d (R295A, H300A, R849A, H854A)。

[0027] 图3A-3D:RNA-靶向的2类CRISPR效应子的系统进化分类和Cas13d家族内的序列保守性。(A)根据Blosom62,本文使用的Cas13d效应子具有HEPN基序保守性,保守残基以阴影示出。RxxxxH HEPN基序突出显示。(B)VI型CRISPR-Cas家族的最大似然树。VI型Cas13超家族效应子的平均氨基酸长度以红色示出。之前所述的2类CRISPR RNA-靶向蛋白质(Abudayyeh et al.,2017;Cox et al.,2017;East-Seletsky et al.,2017;East-Seletsky et al.,2016;Smargon et al.,2017)和Cas13d效应子的比对使用MAFFT 7.38进行,并且最大似然树的构建用PhyML 3.2进行。分支标记物和比例尺表明每个位点的置换。(C)预测的Cas13d同向重复RNA二级结构。(D)全长36nt Cas13d同向重复的序列图标(sequence logo)。

[0028] 图4A-4C:重组Cas13d蛋白质的纯化。将EsCas13d表达成N-端His-MBP融合物,并通过连续的亲和力、阳离子交换和尺寸排阻色谱纯化。通过TEV蛋白酶切割除去His-标签。(A)EsCas13d的Superdex 200柱的色谱图。(B)惰性真杆菌Cas13d的尺寸排阻色谱级分的SDS-PAGE凝胶。(C)纯化的惰性真杆菌Cas13d和dCas13d(两个HEPN基序中预测的催化残基的R295A、H300A、R849A、H854A突变)的SDS-PAGE凝胶。

[0029] 图5A-5D:通过Cas13d在体外进行可编程RNA靶向。(A)惰性真杆菌Cas13d要求匹配的向导阵列或成熟gRNA,以有效切割互补的ssRNA靶。变性凝胶示出在37°C下孵育1小时的切割反应。NT,非靶向。(B)用dCas13d置换或者加入EDTA破坏了用向导和阵列进行的Cas13d-介导的RNA靶向。‘d’,死亡的Cas13d。(C)变性凝胶,其示出Cas13d切割活性的向导-靶匹配依赖性激活。加扰(Scrambled)靶RNA(‘A’)是荧光标记的,而向导-互补的激活靶RNA(‘B’)是未标记的。通过单独除去向导RNA或互补的靶RNA,以及加入EDTA或催化失活Cas13d(示作‘d’),完全破坏了RNA切割活性。(D)Cas13d RNA酶活性的向导和靶依赖性激活模型。Cas13d:gRNA:靶RNA三元复合物能够切割互补的靶RNA或旁观者RNA(bystander RNA)。

[0030] 图6A-6H:Cas13d特性的体外表征。(A)示出gRNA间隔子截短物的长度和序列以及相对于互补ssRNA靶的间隔子位置的示意图。(B)变性凝胶,其示出具有不同间隔子长度的靶RNA的EsCas13d切割活性。(C)变性凝胶,其示出与来自图3A的12种向导配对的EsCas13d切割反应,所述向导拼接(tiling)ssRNA靶的互补ssDNA形式。(D)变性凝胶,其示出使用与相同的12种向导配对的EsCas13d的切割反应,所述向导拼接互补靶的dsDNA形式。(E)对图3A的切割效率进行定量。各个PFS碱基是拼接互补靶RNA的3种不同间隔子序列的平均。切割百分比通过切割的条带强度除以总条带强度的比例确定。平均值示以±SD,每个数据点表示一次独立的重复。(F)携带不同PFS碱基的靶RNA的Cas13d-介导的切割(给定不变的间隔子序列)。显示了对Cas13d切割效率进行定量,和表明EsCas13d切割活性的代表性变性凝胶。差异不显著(单向ANOVA,P=0.768)。切割百分比如上确定,平均值示以±SD,n=3。(G)和(H)Cas13d活性的最优温度范围。变性凝胶示出两种不同的靶RNA在16-62°C范围内的温

度下的EsCas13d切割活性。

[0031] 图7A-7B:Cas13d靶底物偏好性的表征。(A) Cas13d可以用多种向导进行一般性再编程,并且不表现出前间隔子侧翼序列(protospacer flanking sequence,PFS)要求。示出了通过EsCas13d和拼接靶RNA的12种向导的RNA切割。对照条带来自平行运行的单独凝胶。(B) Cas13d优先切割发夹回环或线性均聚重复序列中的尿嘧啶碱基,其每5nt被转换突变(transition mutation) (X)打断,以便能够进行合成。

[0032] 图8A-8D:改造的Cas13d直向同源物在人细胞内的RNA敲低活性筛选。(A) 编码改造的Cas13d效应子和向导的哺乳动物表达构建物示意图。NLS,核定位信号。pre-gRNA,含有侧翼为2个全长36nt DR的单一30nt间隔子序列的未加工的人工向导RNA。gRNA,具有单一30nt加工的DR和22nt间隔子序列的预测成熟向导RNA。(B) 使用4种pre-gRNA或gRNA的库(pool),在人HEK 293FT细胞中Cas13d直向同源物活性筛选中的mCherry蛋白质敲低的热图(heatmap)。归一化MFI,相对于非靶向条件的中位荧光强度。灰色的位置未测试,n=3。(C) 显示改造构建物的定位和表达的Cas13d的免疫荧光化学。比例尺,10 μ m。蓝色伪彩,核的DAPI染色。(D) 比较AdmCas13d(SEQ ID NO:143)和RfxCas13d(SEQ ID NO:280)直向同源物构建物的内源性B4GALNT1 mRNA的敲低,揭示了RfxCas13d-NLS(CasRx;SEQ ID NO:280)对于两种向导RNA构建物都是最有效的。使用4种向导的库进行靶向。NT,非靶向。数值为平均值 \pm SEM,n=3。

[0033] 图9A-9H:CasRx介导许多种人编码和非编码转录物的有效和特异性敲低。(A) 拼接靶转录物的多种向导RNA可以表达成单一阵列,并通过RfxCas13d-NLS(CasRx;SEQ ID NO:280)加工成相同细胞内的单独gRNA。4种向导的阵列各自经由瞬时转染在293FT细胞中通过CasRx介导靶敲低。相对于GFP载体对照的敲低通过qPCR测定,n=3。(C) CasRx靶序列和间隔子位置-匹配的shRNA的示意图。(D) 通过单独的位置匹配的shRNA和CasRx gRNA的相对靶RNA敲低。NT,非靶向。CRISPRi,用于转录阻遏的dCas9-介导的CRISPR干扰(n=3)。(E) 如通过RNA测序(n=3)所测定的,B4GALNT1靶向和非靶向(NT)之间的差异转录物水平的火山图。鉴定出542种非特异性转录物变化。(F) B4GALNT1-靶向CasRx和非靶向(NT)向导之间的差异转录物水平的火山图。靶向向导位置与(E)中所示的shRNA匹配。B4GALNT1是仅有的表现出显著变化的转录物,n=3。(G) 匹配的shRNA和CasRx向导的显著脱靶转录物扰动的小结。(H) 11种内源性转录物的CasRx靶向,其各自在293FT细胞中使用3个向导和1个非靶向(NT)向导。转录物水平相对于GFP载体对照,平均值 \pm SEM,n=3。

[0034] 图10A-10D:改造的Cas13超家族效应子的靶向敲低和剪接的比较。(A) 单独的位置匹配的gRNA在HEK 293FT细胞中对于CasRx、NLS-LwaCas13a-msfGFP(Abudayyeh et al., 2017)和PspCas13b-NES(Cox et al., 2017)的相对靶RNA敲低。NT,非靶向。数值为平均值 \pm SEM,n=3。(B) Cas13中位敲低效率的比较。每个向导RNA,n=3。****表示根据弗里德曼检验(Friedman's test), $P < 0.0001$ 。(C) 双色剪接报告子上通过无催化活性NLS-dCas13a-msfGFP的外显子排斥(exon exclusion)。向导与CasRx在图6B中报道的那些是位置匹配的。平均值 \pm SEM,n=3。(D) 通过NLS-dCas13a-msfGFP和CasRx进行剪接调节的比较。示出相对于非靶向向导的靶向外显子排斥的倍数变化。****表示根据双向ANOVA, $P < 0.0001$ 。

[0035] 图11A-11B:来自人细胞内ANXA4的CasRx和shRNA靶向的RNA测序。(A) 如通过RNA测序(n=3)所确定的,ANXA4靶向和非靶向(NT)shRNA之间的差异转录物水平的火山图。鉴定

出915种非特异性转录物变化。(B)图9B中使用的ANXA4靶向CasRx阵列和非靶向(NT)阵列之间的差异转录物水平的火山图,所述ANXA4靶向CasRx阵列含有与(A)中所示的shRNA位置匹配的向导。ANXA4是仅有的表现出显著下调的转录物, $n=3$ 。HIST2HBE是仅有的经鉴定表现出显著上调的转录物。H2B是H2AX(Du et al., 2006)的二聚配偶体,已经显示其与ANXA4相互作用(Yang et al., 2010)。

[0036] 图12A-12F: AAV递送无催化活性的dCasRx剪接效应子,以操纵可变剪接。(A) 双色外显子跳读报告子的示意图。+1和+3, 阅读框。BP, 内含子分支点(intronic branch point)-靶向向导。SA, 剪接受体位点-重叠向导。EX, 外显子向导。SD, 剪接供体位点-重叠向导。AUG, 起始密码子。UGA, 终止密码子。包括第二外显子导致dsRed的框外非荧光翻译(+3), 然后是框内mTagBFP2。靶向外显子的排除导致dsRed的框内翻译(+1), 然后是终止密码子。(B) 通过dCasRx和靶向pre-mRNA的N-端hnRNP a1-dCasRx融合蛋白的诱导外显子排斥。使用hnRNP a1的富含Gly的C-端结构域作为效应子结构域。外显子跳读效率表示成主要携带dsRed或BFP同种型的细胞的相对百分比, 其通过流式细胞术测定。NLS, 核定位信号。‘A’, 携带全部4种向导的CRISPR阵列。数值为平均值 \pm SEM, $n=3$ 。(C) 携带dCasRx和总转基因大小 < 4.3 kb的三向导阵列的AAV设计, 其包括AAV末端反向重复序列(ITR)。(D) 额颞叶痴呆(FTD)疾病模型的示意图。神经元经由来源于患者的iPSC和对照iPSC的Neurogenin-2(Ngn2)定向分化产生, 然后用dCasRx或载体对照AAV(EFS-mTagBFP2)转导。(E) FTD与推定内含子剪接增强子中的SNP有关, 所述增强子在编码tau的MAPT转录物的外显子10之后。MAPT外显子10的可变剪接导致4R tau(通过包涵)和3R tau(通过排斥)。包括所示IVS 10+16突变的内含子剪接增强子中的SNP导致增加的外显子包涵(exon inclusion), 和更高水平的4R tau。为了促进4R tau水平的降低, 将dCasRx阵列中含有的gRNA靶向外显子10剪接受体(g1)以及两个以紫色示出的推定外显子剪接增强子(g2, g3)。(F) 用AAV转导后的第14天, 通过qPCR检验分化的神经元中的相对4R/3Rtau转录物比例。FTD, 携带IVS 10+16的额颞叶痴呆细胞。数值为平均值 \pm S.D., $n=3$ 。****表示 $P<0.0001$ 。

[0037] 图13是表示使用公开的方法在人细胞中进行RNA靶向的条形图。

[0038] 图14是显示公开的Cas13d和DR如何可以实现可变剪接的示意图。

[0039] 图15A-15D是显示以下的一系列小图:(A) 结合匹配向导RNA的间隔子序列的靶后, Cas13d转化成活性RNA酶复合物。它能够切割gRNA-互补靶RNA或非互补的旁观者RNA。(B) Cas13d的靶依赖性RNA酶活性可以转化成可检测的信号, 例如, 通过切割标记的检测RNA, 该检测RNA仅在匹配Cas13d向导RNA的间隔子的靶的存在下被切割。在该实例中, 检测RNA含有荧光团‘F’和消除荧光的淬灭剂‘Q’。只有在旁观者RNA切割后, 荧光团从淬灭剂中释放, 并产生荧光。(C) 来自嗜性真杆菌的Cas13d仅在完美匹配的靶的存在下产生可见的信号, 在不同错配靶的存在下不产生可见的信号。(D) 来自黄化瘤胃球菌XPD3002株的Cas13d仅在完美匹配的靶的存在下产生可见的信号, 在不同错配靶的存在下不产生可见的信号。

[0040] 图16A-16B:(A) 7种直向同源物的比对显示出高保守性(绿色条)和低保守性(红色条)区域。选择进行缺失的区域以1-10标记。(B) 全长CasRx和CasRx缺失变体的敲低(顶部)和剪接(底部)活性。缺失变体5显示出保留了完整的活性, 表明了保留完整活性的同时缺失低保守性区域的可行性。

[0041] 图17A和17B显示细菌细胞中的ccdB靶向。(A) 引入到表达ccdB的细菌细胞中的构

建物,和(B) ccdB在不同条件下的相对表达。

[0042] 图18A-18MMM显示53种不同Cas13d蛋白质的比对(SEQ ID NO:293、183、189、220、218、222、229、216、177、200、139、179、208、166、185、202、239、294、249、210、243、212、175、164、160、295、296、187、241、140、162、155、206、181、231、128、198、237、233、253、214、224、127、235、153、125、245、247、204、226、3、126和149,从顶部到底部)。

[0043] 图19A显示了经由mCherry报告检验(mCherry reporter assay)筛选具有NLS(SEQ ID NO:25和256)和NES构建物(SEQ ID NO:287)的50种Cas13d直向同源物(参见表1)在人细胞中的敲低结果,其通过敲低效率和效应子基因大小进行作图。仅直向同源物的子集(Ga0531-NLS,Ga7274-NLS和k87-NES,分别地,SEQ ID NO:235,198和189)表现出与CasRx变体(SEQ ID NO:280)相当的活性。

[0044] 图19B是条形图,其显示了内源性CD81在HEK 293FT细胞中被多种具有NLS和NES亚细胞定位信号的Cas13d直向同源物敲低。从mCherry报告检验中选择最有活性的直向同源物(参见表1)用于第二轮活性筛选。CD81敲低经由细胞表面流式细胞术加以评价。CasRx变体RfxCas13d-NLS(SEQ ID NO:280)表现出最高的敲低,一些其他直向同源物在该内源性靶上表现出>75%的敲低。

[0045] 图20是条形图,其显示CasRx(SEQ ID NO:280;RfxCas13d-NLS设计)与PspCas13b-NES(Cox et al., Science. 2017Nov 24;358(6366):1019-1027)用于内源性CD81蛋白质敲低的剂量-反应比较。CasRx在等同的蛋白质水平(如通过Cas13-GFP表达所测量的)下表现出显著更高的活性,表明每个酶单位的活性更高,因此达到生物学/治疗上相关敲低水平所需的剂量更低。此外,CasRx显著小于PspCas13b。CasRx的小尺寸和在较低剂量下的高活性使得其可用于治疗性递送。读数在人HEK 293FT细胞中瞬时转染后3天进行。

[0046] 图21是条形图,其显示了Cas13d直向同源物CasRx和Ga0531的多种改造缺失保持了人HEK 293FT细胞中的敲低活性。

[0047] 图22是条状图,其显示了dCas13d变体(具有突变的HEPN结构域,以消除靶RNA切割)被靶向我们的双色蓝至红剪接报告子(bichromatic blue-to-red reporter)的剪接受体位点,以介导外显子排斥(如通过mTagBFP2表达的降低所测量的)。即使是对于具有高水平内源性蛋白质敲低(>70%CD81蛋白质敲低)的直向同源物,剪接调节可能是无效的。这突出了某些Cas13d直向同源物(例如CasRx和截短的CasRx变体,如CasRx-del13)的高敲低和高剪接调节活性的独特优势组合。将dCas13d变体瞬时转染到HEK293FT细胞后第3天,通过流式细胞术进行剪接读数。

[0048] 图23A-23C:用Cas13d进行RNA编辑。(A) dRfxCas13d-ADAR2效应子融合物和失活的GFP报告子的示意图。(B) GFP报告子中的提前终止密码子(premature stop codon)的靶向A到I的RNA编辑重新激活了GFP表达。非靶向的向导则不导致RNA编辑。(C)对Cas13d-介导的RNA编辑效率进行定量。

[0049] 序列表

[0050] 使用37C.F.R.1.822中定义的标准核苷酸碱基字母缩写和氨基酸三字母代码,显示了所附序列表中列出的核酸和氨基酸序列。每个核酸序列仅示出一条链,但是互补链理解成通过任意参考示出的链而包括在内。序列表以ASCII文本文件提交,于2018年8月22日生成,0.99MB,在此将其并入作为参考。在所附序列表中:

- [0051] SEQ ID NO:1是含有HEPN位点的来自惰性真杆菌的示例性Cas13d序列。
- [0052] SEQ ID NO:2是含有突变HEPN位点的来自惰性真杆菌的示例性Cas13d序列。
- [0053] SEQ ID NO:3是含有HEPN位点的来自未培养的瘤胃球菌属的示例性Cas13d序列。
- [0054] SEQ ID NO:4是含有突变HEPN位点的来自未培养的瘤胃球菌属的示例性Cas13d序列。
- [0055] SEQ ID NO:5是来自肠道_宏基因组_重叠群2791000549 (Gut_metagenome_contig2791000549)的示例性Cas13d序列。
- [0056] SEQ ID NO:6是来自肠道_宏基因组_重叠群855000317的示例性Cas13d序列。
- [0057] SEQ ID NO:7是来自肠道_宏基因组_重叠群3389000027的示例性Cas13d序列。
- [0058] SEQ ID NO:8是来自肠道_宏基因组_重叠群8061000170的示例性Cas13d序列。
- [0059] SEQ ID NO:9是来自肠道_宏基因组_重叠群1509000299的示例性Cas13d序列。
- [0060] SEQ ID NO:10是来自肠道_宏基因组_重叠群9549000591的示例性Cas13d序列。
- [0061] SEQ ID NO:11是来自肠道_宏基因组_重叠群71000500的示例性Cas13d序列。
- [0062] SEQ ID NO:12是来自人肠道宏基因组的示例性Cas13d序列。
- [0063] SEQ ID NO:13是来自肠道_宏基因组_重叠群3915000357的示例性Cas13d序列。
- [0064] SEQ ID NO:14是来自肠道_宏基因组_重叠群4719000173的示例性Cas13d序列。
- [0065] SEQ ID NO:15是来自肠道_宏基因组_重叠群6929000468的示例性Cas13d序列。
- [0066] SEQ ID NO:16是来自肠道_宏基因组_重叠群7367000486的示例性Cas13d序列。
- [0067] SEQ ID NO:17是来自肠道_宏基因组_重叠群7930000403的示例性Cas13d序列。
- [0068] SEQ ID NO:18是来自肠道_宏基因组_重叠群993000527的示例性Cas13d序列。
- [0069] SEQ ID NO:19是来自肠道_宏基因组_重叠群6552000639的示例性Cas13d序列。
- [0070] SEQ ID NO:20是来自肠道_宏基因组_重叠群11932000246的示例性Cas13d序列。
- [0071] SEQ ID NO:21是来自肠道_宏基因组_重叠群12963000286的示例性Cas13d序列。
- [0072] SEQ ID NO:22是来自肠道_宏基因组_重叠群2952000470的示例性Cas13d序列。
- [0073] SEQ ID NO:23是来自肠道_宏基因组_重叠群451000394的示例性Cas13d序列。
- [0074] SEQ ID NO:24是来自惰性真杆菌DSM 15702 (Eubacterium_siraeum_DSM_15702)的示例性Cas13d序列。
- [0075] SEQ ID NO:25是来自肠道_宏基因组_P19E0k2120140920_c369000003的示例性Cas13d序列。
- [0076] SEQ ID NO:26是来自肠道_宏基因组_重叠群7593000362的示例性Cas13d序列。
- [0077] SEQ ID NO:27是来自肠道_宏基因组_重叠群12619000055的示例性Cas13d序列。
- [0078] SEQ ID NO:28是来自肠道_宏基因组_重叠群1405000151的示例性Cas13d序列。
- [0079] SEQ ID NO:29是来自鸡_肠道_宏基因组_c298474的示例性Cas13d序列。
- [0080] SEQ ID NO:30是来自肠道_宏基因组_重叠群1516000227的示例性Cas13d序列。
- [0081] SEQ ID NO:31是来自肠道_宏基因组_重叠群1838000319的示例性Cas13d序列。
- [0082] SEQ ID NO:32是来自肠道_宏基因组_重叠群13123000268的示例性Cas13d序列。
- [0083] SEQ ID NO:33是来自肠道_宏基因组_重叠群5294000434的示例性Cas13d序列。
- [0084] SEQ ID NO:34是来自肠道_宏基因组_重叠群6415000192的示例性Cas13d序列。
- [0085] SEQ ID NO:35是来自肠道_宏基因组_重叠群6144000300的示例性Cas13d序列。

- [0086] SEQ ID NO:36是来自肠道_宏基因组_重叠群9118000041的示例性Cas13d序列。
- [0087] SEQ ID NO:37是来自活性_污泥_宏基因组_转录物_124486 (Activated_sludge_metagenome_transcript_124486)的示例性Cas13d序列。
- [0088] SEQ ID NO:38是来自肠道_宏基因组_重叠群1322000437的示例性Cas13d序列。
- [0089] SEQ ID NO:39是来自肠道_宏基因组_重叠群4582000531的示例性Cas13d序列。
- [0090] SEQ ID NO:40是来自肠道_宏基因组_重叠群9190000283的示例性Cas13d序列。
- [0091] SEQ ID NO:41是来自肠道_宏基因组_重叠群1709000510的示例性Cas13d序列。
- [0092] SEQ ID NO:42是具有HEPN结构域的来自M24_(LSQX01212483_厌氧_消化池_宏基因组) (M24_(LSQX01212483_Anaerobic_digester_metagenome))的示例性Cas13d序列。
- [0093] SEQ ID NO:43是来自肠道_宏基因组_重叠群3833000494的示例性Cas13d序列。
- [0094] SEQ ID NO:44是来自活性_污泥_宏基因组_转录物_117355的示例性Cas13d序列。
- [0095] SEQ ID NO:45是来自肠道_宏基因组_重叠群11061000330的示例性Cas13d序列。
- [0096] SEQ ID NO:46是来自绵羊肠道宏基因组的肠道_宏基因组_重叠群338000322的示例性Cas13d序列。
- [0097] SEQ ID NO:47是来自人肠道宏基因组的示例性Cas13d序列。
- [0098] SEQ ID NO:48是来自肠道_宏基因组_重叠群9530000097的示例性Cas13d序列。
- [0099] SEQ ID NO:49是来自肠道_宏基因组_重叠群1750000258的示例性Cas13d序列。
- [0100] SEQ ID NO:50是来自肠道_宏基因组_重叠群5377000274的示例性Cas13d序列。
- [0101] SEQ ID NO:51是来自肠道_宏基因组_P19E0k2120140920_c248000089的示例性Cas13d序列。
- [0102] SEQ ID NO:52是来自肠道_宏基因组_重叠群11400000031的示例性Cas13d序列。
- [0103] SEQ ID NO:53是来自肠道_宏基因组_重叠群7940000191的示例性Cas13d序列。
- [0104] SEQ ID NO:54是来自肠道_宏基因组_重叠群6049000251的示例性Cas13d序列。
- [0105] SEQ ID NO:55是来自肠道_宏基因组_重叠群1137000500的示例性Cas13d序列。
- [0106] SEQ ID NO:56是来自肠道_宏基因组_重叠群9368000105的示例性Cas13d序列。
- [0107] SEQ ID NO:57是来自肠道_宏基因组_重叠群546000275的示例性Cas13d序列。
- [0108] SEQ ID NO:58是来自肠道_宏基因组_重叠群7216000573的示例性Cas13d序列。
- [0109] SEQ ID NO:59是来自肠道_宏基因组_重叠群4806000409的示例性Cas13d序列。
- [0110] SEQ ID NO:60是来自肠道_宏基因组_重叠群10762000480的示例性Cas13d序列。
- [0111] SEQ ID NO:61是来自肠道_宏基因组_重叠群4114000374的示例性Cas13d序列。
- [0112] SEQ ID NO:62是来自黄化瘤胃球菌FD1 (Ruminococcus_flavofaciens_FD1)的示例性Cas13d序列。
- [0113] SEQ ID NO:63是来自肠道_宏基因组_重叠群7093000170的示例性Cas13d序列。
- [0114] SEQ ID NO:64是来自肠道_宏基因组_重叠群11113000384的示例性Cas13d序列。
- [0115] SEQ ID NO:65是来自肠道_宏基因组_重叠群6403000259的示例性Cas13d序列。
- [0116] SEQ ID NO:66是来自肠道_宏基因组_重叠群6193000124的示例性Cas13d序列。
- [0117] SEQ ID NO:67是来自肠道_宏基因组_重叠群721000619的示例性Cas13d序列。
- [0118] SEQ ID NO:68是来自肠道_宏基因组_重叠群1666000270的示例性Cas13d序列。
- [0119] SEQ ID NO:69是来自肠道_宏基因组_重叠群2002000411的示例性Cas13d序列。

- [0120] SEQ ID NO:70是来自白色瘤胃球菌(*Ruminococcus_albus*)的示例性Cas13d序列。
- [0121] SEQ ID NO:71是来自肠道_宏基因组_重叠群13552000311的示例性Cas13d序列。
- [0122] SEQ ID NO:72是来自肠道_宏基因组_重叠群10037000527的示例性Cas13d序列。
- [0123] SEQ ID NO:73是来自肠道_宏基因组_重叠群238000329的示例性Cas13d序列。
- [0124] SEQ ID NO:74是来自肠道_宏基因组_重叠群2643000492的示例性Cas13d序列。
- [0125] SEQ ID NO:75是来自肠道_宏基因组_重叠群874000057的示例性Cas13d序列。
- [0126] SEQ ID NO:76是来自肠道_宏基因组_重叠群4781000489的示例性Cas13d序列。
- [0127] SEQ ID NO:77是来自肠道_宏基因组_重叠群12144000352的示例性Cas13d序列。
- [0128] SEQ ID NO:78是来自肠道_宏基因组_重叠群5590000448的示例性Cas13d序列。
- [0129] SEQ ID NO:79是来自肠道_宏基因组_重叠群9269000031的示例性Cas13d序列。
- [0130] SEQ ID NO:80是来自肠道_宏基因组_重叠群8537000520的示例性Cas13d序列。
- [0131] SEQ ID NO:81是来自肠道_宏基因组_重叠群1845000130的示例性Cas13d序列。
- [0132] SEQ ID NO:82是来自肠道_宏基因组_P13E0k2120140920_c3000072的示例性Cas13d序列。
- [0133] SEQ ID NO:83是来自肠道_宏基因组_P1E0k2120140920_c1000078的示例性Cas13d序列。
- [0134] SEQ ID NO:84是来自肠道_宏基因组_重叠群12990000099的示例性Cas13d序列。
- [0135] SEQ ID NO:85是来自肠道_宏基因组_重叠群525000349的示例性Cas13d序列。
- [0136] SEQ ID NO:86是来自肠道_宏基因组_重叠群7229000302的示例性Cas13d序列。
- [0137] SEQ ID NO:87是来自肠道_宏基因组_重叠群3227000343的示例性Cas13d序列。
- [0138] SEQ ID NO:88是来自肠道_宏基因组_重叠群7030000469的示例性Cas13d序列。
- [0139] SEQ ID NO:89是来自肠道_宏基因组_重叠群5149000068的示例性Cas13d序列。
- [0140] SEQ ID NO:90是来自肠道_宏基因组_重叠群400200045的示例性Cas13d序列。
- [0141] SEQ ID NO:91是来自肠道_宏基因组_重叠群10420000446的示例性Cas13d序列。
- [0142] SEQ ID NO:92是来自新黄化瘤胃球菌XPD3002株(*new_flavefaciens*,_strain_XPD3002)的示例性Cas13d序列。
- [0143] SEQ ID NO:93是来自M26_肠道_宏基因组_重叠群698000307的示例性Cas13d序列。
- [0144] SEQ ID NO:94是来自M36_未培养的瘤胃球菌属_TS28_c40956(*M36_Uncultured_Eubacterium_sp_TS28_c40956*)的示例性Cas13d序列。
- [0145] SEQ ID NO:95是来自M12_肠道_宏基因组_P25C0k2120140920_c134000066的示例性Cas13d序列。
- [0146] SEQ ID NO:96是来自人肠道宏基因组的示例性Cas13d序列。
- [0147] SEQ ID NO:97是来自M10_肠道_宏基因组_P25C90k2120140920_c28000041的示例性Cas13d序列。
- [0148] SEQ ID NO:98是来自M11_肠道_宏基因组_P25C7k2120140920_c4078000105的示例性Cas13d序列。
- [0149] SEQ ID NO:99是来自肠道_宏基因组_P25C0k2120140920_c32000045的示例性Cas13d序列。

- [0150] SEQ ID NO:100是来自M13_肠道_宏基因组_P23C7k2120140920_c3000067的示例性Cas13d序列。
- [0151] SEQ ID NO:101是来自M5_肠道_宏基因组_P18E90k2120140920的示例性Cas13d序列。
- [0152] SEQ ID NO:102是来自M21_肠道_宏基因组_P18E0k2120140920的示例性Cas13d序列。
- [0153] SEQ ID NO:103是来自M7_肠道_宏基因组_P38C7k2120140920_c4841000003的示例性Cas13d序列。
- [0154] SEQ ID NO:104是来自双环瘤胃球菌(*Ruminococcus_bicirculans*)的示例性Cas13d序列。
- [0155] SEQ ID NO:105是示例性Cas13d序列。
- [0156] SEQ ID NO:106是示例性Cas13d共有序列。
- [0157] SEQ ID NO:107是来自M18_肠道_宏基因组_P22E0k2120140920_c3395000078的示例性Cas13d序列。
- [0158] SEQ ID NO:108是来自M17_肠道_宏基因组_P22E90k2120140920_c114的示例性Cas13d序列。
- [0159] SEQ ID NO:109是来自瘤胃球菌属_CAG57(*Ruminococcus_sp_CAG57*)的示例性Cas13d序列。
- [0160] SEQ ID NO:110是来自肠道_宏基因组_P11E90k2120140920_c43000123的示例性Cas13d序列。
- [0161] SEQ ID NO:111是来自M6_肠道_宏基因组_P13E90k2120140920_c7000009的示例性Cas13d序列。
- [0162] SEQ ID NO:112是来自M19_肠道_宏基因组_P17E90k2120140920的示例性Cas13d序列。
- [0163] SEQ ID NO:113是来自肠道_宏基因组_P17E0k2120140920_c87000043的示例性Cas13d序列。
- [0164] SEQ ID NO:114是示例性的人密码子优化的惰性真杆菌Cas13d核酸序列。
- [0165] SEQ ID NO:115是示例性的具有突变HEPN结构域的人密码子优化的惰性真杆菌Cas13d核酸序列。
- [0166] SEQ ID NO:116是示例性的具有N-端NLS的人密码子优化的惰性真杆菌Cas13d核酸序列。
- [0167] SEQ ID NO:117是示例性的具有N-和C-端NLS标签的人密码子优化的惰性真杆菌Cas13d核酸序列。
- [0168] SEQ ID NO:118是示例性的人密码子优化的未培养瘤胃球菌属Cas13d核酸序列。
- [0169] SEQ ID NO:119是示例性的具有突变HEPN结构域的人密码子优化的未培养瘤胃球菌属Cas13d核酸序列。
- [0170] SEQ ID NO:120是示例性的具有N-端NLS的人密码子优化的未培养瘤胃球菌属Cas13d核酸序列。
- [0171] SEQ ID NO:121是示例性的具有N-和C-端NLS标签的人密码子优化的未培养瘤胃

球菌属Cas13d核酸序列。

[0172] SEQ ID NO:122是示例性的人密码子优化的未培养黄化瘤胃球菌FD1 Cas13d核酸序列。

[0173] SEQ ID NO:123是示例性的具有突变HEPN结构域的人密码子优化的未培养黄化瘤胃球菌FD1 Cas13d核酸序列。

[0174] SEQ ID NO:124是来自双环瘤胃球菌的示例性Cas13d核酸序列。

[0175] SEQ ID NO:125是来自惰性真杆菌的示例性Cas13d核酸序列。

[0176] SEQ ID NO:126是来自黄化瘤胃球菌FD1的示例性Cas13d核酸序列。

[0177] SEQ ID NO:127是来自白色瘤胃球菌的示例性Cas13d核酸序列。

[0178] SEQ ID NO:128是来自黄化瘤胃球菌XPD的示例性Cas13d核酸序列。

[0179] SEQ ID NO:129是惰性真杆菌Cas13d的示例性共有DR核酸序列。

[0180] SEQ ID NO:130是瘤胃球菌属Cas13d的示例性共有DR核酸序列。

[0181] SEQ ID NO:131是黄化瘤胃球菌XPD3002株Cas13d和CasRx的示例性共有DR核酸序列。

[0182] SEQ ID NO:132-137是示例性共有DR核酸序列。

[0183] SEQ ID NO:138是7种全长Cas13d直向同源物的示例性50%共有序列。

[0184] SEQ ID NO:139是来自肠道宏基因组P1E0的示例性Cas13d核酸序列。

[0185] SEQ ID NO:140是来自厌氧消化池的示例性Cas13d核酸序列。

[0186] SEQ ID NO:141是来自瘤胃球菌属CAG:57的示例性Cas13d核酸序列。

[0187] SEQ ID NO:142是示例性的人密码子优化的未培养肠道宏基因组P1E0 Cas13d核酸序列。

[0188] SEQ ID NO:143是示例性的人密码子优化的厌氧消化池Cas13d核酸序列。

[0189] SEQ ID NO:144是示例性的人密码子优化的黄化瘤胃球菌XPD Cas13d核酸序列。

[0190] SEQ ID NO:145是示例性的人密码子优化的白色瘤胃球菌Cas13d核酸序列。

[0191] SEQ ID NO:146是瘤胃球菌属CAG:57CRISPR阵列的示例性加工。

[0192] SEQ ID NO:147是来自人肠道宏基因组序列重叠群emb|OBVH01003037.1的示例性Cas13d蛋白质序列(也发现于WGS重叠群emb|OBXZ01000094.1|和emb|OBF01000033.1)。

[0193] SEQ ID NO:148是示例性共有DR核酸序列(与SEQ ID NO:147一起)。

[0194] SEQ ID NO:149是来自重叠群tpg|DBYI01000091.1|(组装自牛肠道宏基因组的未培养黄化瘤胃球菌UBA1190)的示例性Cas13d蛋白质序列。

[0195] SEQ ID NO:150-152是示例性共有DR核酸序列(与SEQ ID NO:149一起)。

[0196] SEQ ID NO:153是来自重叠群tpg|DJXD01000002.1|(未培养瘤胃球菌组装, UBA7013, 来自绵羊肠道宏基因组)的示例性Cas13d蛋白质序列。

[0197] SEQ ID NO:154是示例性共有DR核酸序列(与SEQ ID NO:153一起)。

[0198] SEQ ID NO:155是来自重叠群OGZC01000639.1(人肠道宏基因组组装)的示例性Cas13d蛋白质序列。

[0199] SEQ ID NO:156-177是示例性共有DR核酸序列(与SEQ ID NO:155一起)。

[0200] SEQ ID NO:158是来自重叠群emb|OHBM01000764.1(人肠道宏基因组组装)的示例性Cas13d蛋白质序列。

- [0201] SEQ ID NO:159是示例性共有DR核酸序列(与SEQ ID NO:158一起)。
- [0202] SEQ ID NO:160是来自重叠群emb|OHCP01000044.1(人肠道宏基因组组装)的示例性Cas13d蛋白质序列。
- [0203] SEQ ID NO:161是示例性共有DR核酸序列(与SEQ ID NO:160一起)。
- [0204] SEQ ID NO:162是来自重叠群emb|OGDF01008514.1|(人肠道宏基因组组装)的示例性Cas13d蛋白质序列。
- [0205] SEQ ID NO:163是示例性共有DR核酸序列(与SEQ ID NO:162一起)。
- [0206] SEQ ID NO:164是来自重叠群emb|OGPN01002610.1(人肠道宏基因组组装)的示例性Cas13d蛋白质序列。
- [0207] SEQ ID NO:165是示例性共有DR核酸序列(与SEQ ID NO:164一起)。
- [0208] SEQ ID NO:166是来自重叠群NFIR01000008.1(真杆菌属An3,来自鸡肠道宏基因组)的示例性Cas13d蛋白质序列。
- [0209] SEQ ID NO:167是示例性共有DR核酸序列(与SEQ ID NO:166一起)。
- [0210] SEQ ID NO:168是来自重叠群NFLV01000009.1(真杆菌属An11,来自鸡肠道宏基因组)的示例性Cas13d蛋白质序列。
- [0211] SEQ ID NO:169是示例性共有DR核酸序列(与SEQ ID NO:168一起)。
- [0212] SEQ ID NO:171-174是示例性Cas13d基序序列。
- [0213] SEQ ID NO:175是来自重叠群OJMM01002900人肠道宏基因组序列的示例性Cas13d蛋白质序列。
- [0214] SEQ ID NO:176是示例性共有DR核酸序列(与SEQ ID NO:175一起)。
- [0215] SEQ ID NO:177是来自重叠群ODAI011611274.1肠道宏基因组序列的示例性Cas13d蛋白质序列。
- [0216] SEQ ID NO:178是示例性共有DR核酸序列(与SEQ ID NO:177一起)。
- [0217] SEQ ID NO:179是来自重叠群OIZX01000427.1的示例性Cas13d蛋白质序列。
- [0218] SEQ ID NO:180是示例性共有DR核酸序列(与SEQ ID NO:179一起)。
- [0219] SEQ ID NO:181是来自重叠群emb|OCVV012889144.1的示例性Cas13d蛋白质序列。
- [0220] SEQ ID NO:182是示例性共有DR核酸序列(与SEQ ID NO:181一起)。
- [0221] SEQ ID NO:183是来自重叠群OCTW011587266.1的示例性Cas13d蛋白质序列。
- [0222] SEQ ID NO:184是示例性共有DR核酸序列(与SEQ ID NO:183一起)。
- [0223] SEQ ID NO:185是来自重叠群emb|OGNF01009141.1的示例性Cas13d蛋白质序列。
- [0224] SEQ ID NO:186是示例性共有DR核酸序列(与SEQ ID NO:185一起)。
- [0225] SEQ ID NO:187是来自重叠群emb|OIEN01002196.1的示例性Cas13d蛋白质序列。
- [0226] SEQ ID NO:188是示例性共有DR核酸序列(与SEQ ID NO:187一起)。
- [0227] SEQ ID NO:189是来自重叠群e-k87_11092736的示例性Cas13d蛋白质序列。
- [0228] SEQ ID NO:190-193是示例性共有DR核酸序列(与SEQ ID NO:189一起)。
- [0229] SEQ ID NO:194是来自肠道_宏基因组_重叠群6893000291的示例性Cas13d序列。
- [0230] SEQ ID NO:195-197是示例性Cas13d基序序列。
- [0231] SEQ ID NO:198是来自Ga0224415_10007274的示例性Cas13d蛋白质序列。
- [0232] SEQ ID NO:199是示例性共有DR核酸序列(与SEQ ID NO:198一起)。

- [0233] SEQ ID NO:200是来自EMG_10003641的示例性Cas13d蛋白质序列。
- [0234] SEQ ID NO:201是示例性共有DR核酸序列(与SEQ ID NO:200一起)。
- [0235] SEQ ID NO:202是来自Ga0129306_1000735的示例性Cas13d蛋白质序列。
- [0236] SEQ ID NO:203是示例性共有DR核酸序列(与SEQ ID NO:202一起)。
- [0237] SEQ ID NO:204是来自Ga0129317_1008067的示例性Cas13d蛋白质序列。
- [0238] SEQ ID NO:205是示例性共有DR核酸序列(与SEQ ID NO:204一起)。
- [0239] SEQ ID NO:206是来自Ga0224415_10048792的示例性Cas13d蛋白质序列。
- [0240] SEQ ID NO:207是示例性共有DR核酸序列(与SEQ ID NO:206一起)。
- [0241] SEQ ID NO:208是来自160582958_gene49834的示例性Cas13d蛋白质序列。
- [0242] SEQ ID NO:209是示例性共有DR核酸序列(与SEQ ID NO:208一起)。
- [0243] SEQ ID NO:210是来自250twins_35838_GL0110300的示例性Cas13d蛋白质序列。
- [0244] SEQ ID NO:211是示例性共有DR核酸序列(与SEQ ID NO:210一起)。
- [0245] SEQ ID NO:212是来自250twins_36050_GL0158985的示例性Cas13d蛋白质序列。
- [0246] SEQ ID NO:213是示例性共有DR核酸序列(与SEQ ID NO:212一起)。
- [0247] SEQ ID NO:214是来自31009_GL0034153的示例性Cas13d蛋白质序列。
- [0248] SEQ ID NO:215是示例性共有DR核酸序列(与SEQ ID NO:214一起)。
- [0249] SEQ ID NO:216是来自530373_GL0023589的示例性Cas13d蛋白质序列。
- [0250] SEQ ID NO:217是示例性共有DR核酸序列(与SEQ ID NO:216一起)。
- [0251] SEQ ID NO:218是来自BMZ-11B_GL0037771的示例性Cas13d蛋白质序列。
- [0252] SEQ ID NO:219是示例性共有DR核酸序列(与SEQ ID NO:218一起)。
- [0253] SEQ ID NO:220是来自BMZ-11B_GL0037915的示例性Cas13d蛋白质序列。
- [0254] SEQ ID NO:221是示例性共有DR核酸序列(与SEQ ID NO:220一起)。
- [0255] SEQ ID NO:222是来自BMZ-11B_GL0069617的示例性Cas13d蛋白质序列。
- [0256] SEQ ID NO:223是示例性共有DR核酸序列(与SEQ ID NO:222一起)。
- [0257] SEQ ID NO:224是来自-DLF014_GL0011914的示例性Cas13d蛋白质序列。
- [0258] SEQ ID NO:225是示例性共有DR核酸序列(与SEQ ID NO:224一起)。
- [0259] SEQ ID NO:226是来自EYZ-362B_GL0088915的示例性Cas13d蛋白质序列。
- [0260] SEQ ID NO:227-228是示例性共有DR核酸序列(与SEQ ID NO:226一起)。
- [0261] SEQ ID NO:229是来自Ga0099364_10024192的示例性Cas13d蛋白质序列。
- [0262] SEQ ID NO:230是示例性共有DR核酸序列(与SEQ ID NO:229一起)。
- [0263] SEQ ID NO:231是来自Ga0187910_10006931的示例性Cas13d蛋白质序列。
- [0264] SEQ ID NO:232是示例性共有DR核酸序列(与SEQ ID NO:231一起)。
- [0265] SEQ ID NO:233是来自Ga0187910_10015336的示例性Cas13d蛋白质序列。
- [0266] SEQ ID NO:234是示例性共有DR核酸序列(与SEQ ID NO:233一起)。
- [0267] SEQ ID NO:235是来自Ga0187910_10040531的示例性Cas13d蛋白质序列。
- [0268] SEQ ID NO:236是示例性共有DR核酸序列(与SEQ ID NO:235一起)。
- [0269] SEQ ID NO:237是来自Ga0187911_10069260的示例性Cas13d蛋白质序列。
- [0270] SEQ ID NO:238是示例性共有DR核酸序列(与SEQ ID NO:237一起)。
- [0271] SEQ ID NO:239是来自MH0288_GL0082219的示例性Cas13d蛋白质序列。

- [0272] SEQ ID NO:240是示例性共有DR核酸序列(与SEQ ID NO:239一起)。
- [0273] SEQ ID NO:241是来自02.UC29-0_GL0096317的示例性Cas13d蛋白质序列。
- [0274] SEQ ID NO:242是示例性共有DR核酸序列(与SEQ ID NO:241一起)。
- [0275] SEQ ID NO:243是来自PIG-014_GL0226364的示例性Cas13d蛋白质序列。
- [0276] SEQ ID NO:244是示例性共有DR核酸序列(与SEQ ID NO:243一起)。
- [0277] SEQ ID NO:245是来自PIG-018_GL0023397的示例性Cas13d蛋白质序列。
- [0278] SEQ ID NO:246是示例性共有DR核酸序列(与SEQ ID NO:245一起)。
- [0279] SEQ ID NO:247是来自PIG-025_GL0099734的示例性Cas13d蛋白质序列。
- [0280] SEQ ID NO:248是示例性共有DR核酸序列(与SEQ ID NO:247一起)。
- [0281] SEQ ID NO:249是来自PIG-028_GL0185479的示例性Cas13d蛋白质序列。
- [0282] SEQ ID NO:250是示例性共有DR核酸序列(与SEQ ID NO:249一起)。
- [0283] SEQ ID NO:251是来自-Ga0224422_10645759的示例性Cas13d蛋白质序列。
- [0284] SEQ ID NO:252是示例性共有DR核酸序列(与SEQ ID NO:251一起)。
- [0285] SEQ ID NO:253是来自ODAI嵌合体的示例性Cas13d蛋白质序列。
- [0286] SEQ ID NO:254是示例性共有DR核酸序列(与SEQ ID NO:253一起)。
- [0287] SEQ ID NO:255是HEPN基序。
- [0288] SEQ ID NO:256和257分别是示例性Cas13d核定位信号氨基酸和核酸序列。
- [0289] SEQ ID NO:258和260分别是示例性SV40大T抗原核定位信号氨基酸和核酸序列。
SEQ ID NO:259是dCas9靶序列。
- [0290] SEQ ID NO:261是靶向ccdB的人工惰性真杆菌nCas1阵列。
- [0291] SEQ ID NO:262是完整的36nt同向重复序列。
- [0292] SEQ ID NO:263-266是间隔子序列。
- [0293] SEQ ID NO:267是靶向ccdB的人工未培养瘤胃球菌属nCas1阵列。
- [0294] SEQ ID NO:268是完整的36nt同向重复序列。
- [0295] SEQ ID NO:269-272是间隔子序列。
- [0296] SEQ ID NO:273是ccdB靶RNA序列。
- [0297] SEQ ID NO:274-277是间隔子序列。
- [0298] SEQ ID NO:278是突变的Cas13d序列,NLS-Ga_0531 (trunc) -NLS-HA。该突变体具有非保守N-端的缺失。
- [0299] SEQ ID NO:279是突变的Cas13d序列,NES-Ga_0531 (trunc) -NES-HA。该突变体具有非保守N-端的缺失。
- [0300] SEQ ID NO:280是全长Cas13d序列,NLS-RfxCas13d-NLS-HA。
- [0301] SEQ ID NO:281是突变的Cas13d序列,NLS-RfxCas13d (de15) -NLS-HA。该突变体具有氨基酸558-587的缺失。
- [0302] SEQ ID NO:282是突变的Cas13d序列,NLS-RfxCas13d (de15.12) -NLS-HA。该突变体具有氨基酸558-587和953-966的缺失。
- [0303] SEQ ID NO:283是突变的Cas13d序列,NLS-RfxCas13d (de15.13) -NLS-HA。该突变体具有氨基酸376-392和558-587的缺失。
- [0304] SEQ ID NO:284是突变的Cas13d序列,NLS-RfxCas13d (de15.12+5.13) -NLS-HA。该

突变体具有氨基酸376-392、558-587和953-966的缺失

[0305] SEQ ID NO:285是突变的Cas13d序列,NLS-RfxCas13d(del13)-NLS-HA。该突变体具有氨基酸376-392的缺失。

[0306] SEQ ID NO:286是用于编辑ADAR2表达的效应子序列。氨基酸1-969是dRfxCas13, aa 970-991是NLS序列,并且氨基酸992-1378是ADAR2_{DD}。

[0307] SEQ ID NO:287是示例性HIV NES蛋白质序列。

[0308] SEQ ID NO:288-291是示例性Cas13d基序序列。

[0309] SEQ ID NO:292是Cas13d直向同源物序列MH_4866。

[0310] SEQ ID NO:293是来自037_-_emb|OIZA01000315.1|的示例性Cas13d蛋白质序列。

[0311] SEQ ID NO:294是来自PIG-022_GL0026351的示例性Cas13d蛋白质序列。

[0312] SEQ ID NO:295是来自PIG-046_GL0077813的示例性Cas13d蛋白质序列。

[0313] SEQ ID NO:296是来自猪嵌合体(pig_chimera)的示例性Cas13d蛋白质序列。

具体实施方式

[0314] 除非另外指出,技术术语根据常规用法使用。分子生物学常用术语的定义可见于以下文献:Benjamin Lewin,Genes VII,published by Oxford University Press,1999; Kendrew et al. (eds.),The Encyclopedia of Molecular Biology,published by Blackwell Science Ltd.,1994;和Robert A.Meyers(ed.),Molecular Biology and Biotechnology:a Comprehensive Desk Reference,published by VCH Publishers, Inc.,1995;以及其他类似的参考文献。

[0315] 如本文所用,单数形式“一种”、“一个”和“该”是指单数和复数,除非上下文另有明确指示。如本文所用,术语“包含”是指“包括”。因此,“包含核酸分子”是指“包括核酸分子”,不排除其他元件。还应理解,除非另有说明,否则对于核酸给出的任何和所有碱基大小是近似的,并且出于描述性目的而提供。尽管可以使用与本文所述的那些类似或等同的许多方法和材料,但下文描述了特别合适的方法和材料。在冲突的情况下,以本说明书(包括术语的解释)为准。此外,材料、方法和实施例仅是说明性的,而无意进行限定。所有参考文献,包括专利申请和专利,均通过引用整体并入本文。

[0316] 为了便于审阅本公开的各种实施方式,提供对特定术语的以下解释:

[0317] I. 术语

[0318] 施用:通过任何有效途径向受试者提供或给予药剂,例如本文公开的Cas13d蛋白质(或Cas13d编码序列)或向导分子(或编码序列)。示例性的施用途径包括,但不限于,注射(例如皮下、肌内、皮内、腹膜内、肿瘤内和静脉内)、透皮、鼻内和吸入途径。

[0319] Cas13d(也称作CasR,用于CRISPR-相关的RNA酶):可以切割或结合RNA的RNA-向导的RNA内切核酸酶。Cas13d蛋白质包括一个或两个HEPN结构域(例如,参见SEQ ID NO:1-3、42、62、70、82、83、92、93、94、95、96、97、98、99、100、101、102、103、104、105、106、107、108、109、110、111、112、113、138、147、149、153、155、158、160、162、164、166、168、170、175、177、179、181、183、185、187、189、194、198、200、202、204、206、208、210、212、214、216、218、220、222、224、226、229、231、233、235、237、239、241、243、245、247、249、251、253、278、279、280、281、282、283、284、285、292、293、294、295和296)。天然的HEPN结构域包括序列RXXXXH(SEQ

ID NO:255)。包括突变的HEPN结构域、从而不能切割RNA但是可以加工向导RNA的Cas13d蛋白质也包括在本公开内(例如,参见SEQ ID NO:2和4)。天然Cas13d蛋白质的比对示于图18A-18MMM。此外,Cas13d蛋白质特异性识别具有特定二级结构的gRNA的同向重复序列(例如,参见FIGS.S2C)。在一实例中,Cas13d蛋白质识别和/或结合DR,所述DR具有(1)大约4-8nt的回环,(2)4-12nt的茎,茎由互补核苷酸形成,由于茎内的nt错配其可以包括小的(例如1或2bp)凸起(bulge),和(3)由未成对nt形成的凸起或突出端,其可以为大约10-14nt(例如,任一侧上5-7)。

[0320] 在一实例中,全长(非截短)Cas13d蛋白质的长度为870-1080个氨基酸。在一实例中,Cas13d蛋白质来源于来自梭菌目的细菌的基因组序列或宏基因组序列。在一实例中,Cas13d蛋白质的对应的DR序列位于包括Cas13d gRNA的分子中的间隔子序列的5'端处。在一实例中,Cas13d gRNA中的DR序列在相对于未加工的Cas13d向导阵列转录物的DR序列的5'端处被截短(例如,被截短至少1nt,至少2nt,至少3nt,至少4nt,至少5nt,例如1-3nt,3-6nt,5-7nt或5-10nt)。在一实例中,Cas13d gRNA中的DR序列在5'端处被Cas13d蛋白质截短5-7nt。在一实例中,Cas13d蛋白质可以切割靶RNA,该靶RNA在间隔子-靶双链体3'端的侧翼为A、U、G或C核糖核苷酸中任意者,在5'端的侧翼为A、U、G或C中任意者。

[0321] 在一实例中,Cas13d蛋白质与SEQ ID NO:1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100、101、102、103、104、105、106、107、108、109、110、111、112、113、138、147、149、153、155、158、160、162、164、166、168、170、175、177、179、181、183、185、187、189、194、198、200、202、204、206、208、210、212、214、216、218、220、222、224、226、229、231、233、235、237、239、241、243、245、247、249、251、253、278、279、280、281、282、283、284、285、292、293、294、295或296具有至少80%、至少85%、至少90%、至少91%、至少92%、至少95%、至少96%、至少97%、至少98%或至少99%的序列同一性。

[0322] 在一实例中,Cas13d蛋白质含有两个HEPN RNA酶结构域,该结构域含有RXXXXH氨基酸基序(SEQ ID NO:255;其中X表示任意氨基酸)。此外,Cas13d蛋白质可以包括一个或多个以下的以常用的Prosites形式书写的氨基酸基序:

[0323] 基序1:L-x(5)-[FWY]-x(3)-K-[NQS]-[ILM]-[ILMV]-x(2)-N-x(2)-[FWY]-x(2)-[AG]-x(4)-[DE]-x-D(SEQ ID NO:195)

[0324] 基序2:[FWY]-[ILV]-x(2)-[NQS]-[ILV]-x(2)-[DNST]-x(2)-F-x-Y-x(2)-[HKR]-[FHY](SEQ ID NO:196)

[0325] 基序3:Y-[CDNSV]-x(2)-R-[FWY]-x-[ADNT]-[LM]-[ST]-x(4)-[FWY](SEQ ID NO:197)。

[0326] 在一实例中,Cas13d蛋白质包括一个或多个以下的以常用的Prosites形式书写的氨基酸共有序列:

[0327] L-[DE]-G-K-E-I-N-x-[LM]-x-[CST]-[ANST]-L-I-x-K-F-[DE]-[GN]-I-x-[ADS]-[FIL]-x-[DEHK]-[AILV](SEQ ID NO:288)

[0328] R-N-[FY]-I-[AIL]-x-N-V-[ILV]-x-[NS]-x(2)-F-x-Y-[ILV]-[AIV]-[KR]-[HY]-

x(2) - [PT] - x(3) - [HKR] - x - [IL] - x(2) - N - x(2) - [ILV] - x(2) - [FY] - V - L - x(2) - [ILM] - [AGP] - [DE] - x(4) - [NQRS] (SEQ ID NO:289)

[0329] L - x - [LM] - [ST] - [IV] - x - Y - x(3) - K - [NS] - [ILM] - V - x - [IV] - N - [AST] - R - [FY] - x(2) - [AG] - [FY] - x - [ACV] - [LV] - E - R - D - x(2) - [FL] (SEQ ID NO:290)

[0330] P - F - [AGS] - Y - [CNV] - x - [AP] - R - [FY] - K - N - L - [ST] - x - [EN] - x - L - F - x - [DKR] - [DEN] (SEQ ID NO:291)

[0331] 因此,在一些实例中,与SEQ ID NO:1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100、101、102、103、104、105、106、107、108、109、110、111、112、113、138、147、149、153、155、158、160、162、164、166、168、170、175、177、179、181、183、185、187、189、194、198、200、202、204、206、208、210、212、214、216、218、220、222、224、226、229、231、233、235、237、239、241、243、245、247、249、251、253、278、279、280、281、282、283、284、285、292、293、294、295或296具有至少80%、至少85%、至少90%、至少91%、至少92%、至少95%、至少96%、至少97%、至少98%或至少99%的序列同一性的Cas13d蛋白质包括SEQ ID NO:195、196或197的基序。

[0332] 此外,在一些实例中,与SEQ ID NO:1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100、101、102、103、104、105、106、107、108、109、110、111、112、113、138、147、149、153、155、158、160、162、164、166、168、170、175、177、179、181、183、185、187、189、194、198、200、202、204、206、208、210、212、214、216、218、220、222、224、226、229、231、233、235、237、239、241、243、245、247、249、251、253、278、279、280、281、282、283、284、285、292、293、294、295或296具有至少80%、至少85%、至少90%、至少91%、至少92%、至少95%、至少96%、至少97%、至少98%或至少99%的序列同一性的Cas13d蛋白质包括SEQ ID NO:288、289、290或291的共有序列。

[0333] 互补性:核酸与另一核酸序列通过传统的沃森-克里克碱基配对或其它非传统类型形成氢键的能力。互补性百分比表示核酸分子中可以与第二核酸序列形成氢键(例如沃森-克里克碱基配对)的残基的百分比(例如10个中的5、6、7、8、9、10个为50%、60%、70%、80%、90%和100%互补)。“完美互补”是指核酸序列的所有连续残基将与第二核酸序列中相同数量的连续残基形成氢键。如本文所用,“基本上互补”是指在8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、30、35、40、45、50或更多个核苷酸的区域上至少60%、65%、70%、75%、80%、85%、90%、95%、97%、98%、99%或100%的互补性程度,或者是指在严格条件下杂交的两种核酸。

[0334] CRISPR(规律间隔成簇短回文重复序列):CRISPR RNA阵列是CRISPR系统的限定特征。术语“CRISPR”是指包括间隔以可变间隔子的恒定同向重复(DR)的阵列结构。在一些实例中,CRISPR阵列至少包括DR-间隔子-DR-间隔子(参见图1A)。该特征用于公开的计算流程

中以鉴定新的Cas13d蛋白质家族(图1A)。在细菌中,该阵列被转录为一个单一转录物(含有多个crRNA单位),然后将其通过Cas13d蛋白质和其它RNA酶加工成单独的crRNA。在约40%经测序的细菌基因组和90%经测序的古细菌中发现CRISPR。CRISPR通常与编码CRISPR相关蛋白质(例如本文提供的Cas13d蛋白)的cas基因相关。公开的CRISPR/Cas系统可以用于RNA靶向,例如检测靶RNA,在任意所需的位置处修饰靶RNA,或者在任意所需的位置处切割靶RNA。

[0335] 下调或敲低:当用于提及分子例如靶RNA的表达时,是指导致靶RNA的产生降低但在一些实例中并不完全消除靶RNA产物或靶RNA功能的任何过程。在一实例中,下调或敲低不导致可检测靶RNA表达或靶RNA活性的完全消除。在一些实例中,靶RNA是编码RNA。在一些实例中,靶RNA是非编码RNA。可以被靶向用于下调的RNA分子的具体实例包括mRNA、miRNA、rRNA、tRNA、核RNA、lincRNA、环状RNA和结构RNA。在一些实例中,靶RNA的下调或敲低包括减少靶RNA的翻译并由此可以减少相应蛋白质存在的过程。公开的CRISPR/Cas系统可以用于下调任何感兴趣的靶RNA。

[0336] 下调或敲低包括任何可检测的靶RNA的降低。在某些实例中,与对照(例如在相应的正常细胞或样品中检测到的靶RNA的量)相比,细胞或无细胞系统中可检测的靶RNA减少至少10%、至少20%、至少30%、至少40%、至少50%、至少60%、至少70%、至少75%、至少80%、至少90%、至少95%、至少96%、至少97%、至少98%或至少99%(例如减少40%至90%、40%至80%或50%至95%)。在一实例中,对照是正常细胞(例如,不包括Cas13d或向导RNA的非重组细胞)中的相对表达量。

[0337] 有效量:足以产生有益或期望结果的药剂(例如本文提供的CRISPR/Cas药剂)的量。

[0338] 治疗有效量可以根据以下一种或多种而变化:所治疗的受试者和疾病状况、受试者的体重和年龄、疾病状况的严重程度、施用方式等,其可以容易地由本领域普通技术人员确定。有益的治疗效果可以包括实现诊断确定;改善疾病、症状、病症或病理状况;减少或预防疾病、症状、病症或病状的发作;并且通常对抗疾病、症状、病症或病理状况。在一实施方式中,“有效量”是足以减轻疾病症状的量,例如减轻至少10%、至少20%、至少50%、至少70%或至少90%(与未施用治疗剂相比)。

[0339] 该术语还适用于允许本文的Cas13d和/或gRNA的表达并允许靶向(例如检测或修饰)靶RNA的剂量。

[0340] 增加或降低:分别为在量上与对照值相比统计学显著的正向或负向变化。增加是正向变化,例如与对照值相比增加至少50%、至少100%、至少200%、至少300%、至少400%或至少500%。降低是负向变化,例如与对照值相比降低至少20%、至少25%、至少50%、至少75%、至少80%、至少90%、至少95%、至少98%、至少99%或至少100%。在一些实例中,该降低小于100%,例如降低不超过90%、不超过95%或不超过99%。

[0341] 分离的:“分离的”生物组分(例如Cas13d蛋白质或核酸、gRNA或含有它们的细胞)已经与存在该组分的生物体的细胞或组织中的其它生物组分(例如其它细胞、染色体和染色体外DNA和RNA以及蛋白质)基本上分离、分开产生或从其中纯化出来。已经“分离”的核酸和蛋白质包括通过标准纯化方法纯化的核酸和蛋白质。该术语还包括通过在宿主细胞中重组表达而制备的核酸和蛋白质以及化学合成的核酸和蛋白质。分离的Cas13d蛋白质或核酸

或含有它们的细胞在一些实例中是至少50%纯的,例如至少75%、至少80%、至少90%、至少95%、至少98%或至少100%纯的。

[0342] 标记物(label):直接或间接地与另一分子(例如核酸分子)缀合以促进该分子检测的化合物或组合物。标记物的非限制性实例包括荧光(fluorescent)和发荧光(fluorogenic)部分、发色部分、半抗原、亲和标签和放射性同位素。标记物可以是直接可检测的(例如,光学可检测)或间接可检测的(例如,经由与一种或多种另外的分子相互作用,该另外的分子又是可检测的)。

[0343] 调节:RNA含量的变化。调节可以包括,但不限于,RNA活化(例如上调)、RNA抑制(例如下调)、核糖核苷酸缺失、核糖核苷酸插入、核糖核苷酸化学修饰、核糖核苷酸共价或非共价连接和/或核糖核苷酸置换。

[0344] 非天然或改造的:在此使用的术语可互换,并且表示涉及人工。当提及核酸分子或多肽时,该术语表示核酸分子或多肽至少基本上不含至少一种与它们在自然界中天然相关的和如在自然界中所发现的其它组分。此外,该术语可以表示核酸分子或多肽具有在自然界中不存在的序列。

[0345] 可操作连接(Operably linked):当第一核酸序列与第二核酸序列处于功能关系时,第一核酸序列与第二核酸序列可操作地连接。例如,如果启动子影响编码序列(例如Cas13d蛋白质的编码序列)的转录或表达,则启动子与编码序列可操作连接。通常,可操作连接的DNA序列是连续的,并且在需要连接两个蛋白质编码区的情况下,在相同的阅读框中。

[0346] 药学上可接受的载体:可用于本发明的药学上可接受的载体是常规的。Remington's Pharmaceutical Sciences(E.W.Martin,Mack Publishing Co.,Easton,PA,15th Edition(1975))描述了适合于Cas13d蛋白质或核酸分子(或使用具有公开的Cas13d蛋白质的公开的CRISPR/Cas系统修饰RNA所需的其它分子)的药物递送的组合物和制剂。

[0347] 通常,载体的性质将取决于所采用的具体施用方式。例如,肠胃外制剂通常包括可注射流体,其包括药学上和生理学上可接受的流体,例如水、生理盐水、平衡盐溶液、葡萄糖水溶液、甘油等作为媒介物。除生物中性载体之外,待施用的药物组合物还可以含有少量无毒的辅助物质,例如润湿剂或乳化剂、防腐剂和pH缓冲剂等,例如乙酸钠或脱水山梨醇单月桂酸酯。

[0348] 多肽、肽和蛋白质:是指任意长度的氨基酸聚合物。聚合物可以是线性的或分支的,其可以包括修饰的氨基酸,并可以被非氨基酸中断。该术语还包括经修饰的氨基酸聚合物;例如二硫键形成、糖基化、脂化、乙酰化、磷酸化或任何其它操纵,例如与标记组分缀合。如本文所用,术语“氨基酸”包括天然和/或非天然或合成的氨基酸,包括甘氨酸和D或L光学异构体、以及氨基酸类似物和肽模拟物。

[0349] 启动子:指导核酸转录的核酸控制序列的阵列。启动子包括转录起始位点附近的必需核酸序列。启动子还任选地包括远端增强子或阻遏物元件。“组成型启动子”是持续活性的启动子,并且不受外部信号或分子的调节。相反,“诱导型启动子”的活性受外部信号或分子(例如转录因子)调节。

[0350] 重组或宿主细胞:已经被遗传改变或者能够通过引入外源性多核苷酸例如重组质粒或载体而被遗传改变的细胞。通常,宿主细胞是其中载体可以增殖并且其核酸得以表达

的细胞。这样的细胞可以是真核的或原核的。该术语还包括主题宿主细胞的任何后代。应当理解到,所有后代可以与亲本细胞不相同,因为复制过程中可以发生突变。然而,当使用术语“宿主细胞”时,这样的后代包括在内。

[0351] 调节元件:该术语包括启动子、增强子、内部核糖体进入位点(IRES)和其它表达控制元件(例如转录终止信号,如多聚腺苷酸化信号和poly-U序列)。这些调节元件描述于,例如,Goeddel, GENE EXPRESSION TECHNOLOGY:METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990), 其通过引用整体并入本文。调节元件包括在许多类型的宿主细胞中指导核苷酸序列组成型表达的调节元件和仅在某些宿主细胞中指导核苷酸序列表达的调节元件(例如,组织特异性调节序列)。组织特异性启动子可以指导主要在感兴趣的所需组织中的表达,上述组织例如肌肉、神经元、骨、皮肤、血液、特定器官(例如肝、胰腺)或特定细胞类型(例如淋巴细胞)。调节元件还可以以依赖于时间的方式指导表达,例如,以依赖于细胞周期或者依赖于发育阶段的方式,其可以是组织或细胞类型特异性的,也可以不是组织或细胞类型特异性的。

[0352] 在一些实施方式中,本文提供的载体包括pol III启动子(例如U6和H1启动子)、pol II启动子(例如逆转录病毒劳斯氏肉瘤病毒(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)或二者。

[0353] 术语“调节元件”还包括增强子元件,例如WPRE;CMV增强子;HTLV-I的LTR中的R-U5'链段(segment);SV40增强子;和兔 β -球蛋白的外显子2和3之间的内含子序列。

[0354] 在一些实例中,Cas13d编码序列与启动子、增强子元件或二者可操作连接。

[0355] RNA编辑:一种基因工程类型,其中使用改造的核酸酶(如本文提供的Cas13d蛋白质)在生物体的基因组中插入、缺失或置换RNA分子(或RNA的核糖核苷酸),上述改造的核酸酶在RNA中的所需位置处产生位点特异性的链断裂。所诱导的断裂被修复,导致靶向突变或修复。本文公开的CRISPR/Cas方法,例如使用Cas13d的方法,可以用于编辑一种或多种靶RNA的序列,例如与癌症(例如乳腺癌、结肠癌、肺癌、前列腺癌、黑素瘤)、传染性疾病(例如HIV、肝炎、HPV和西尼罗病毒)或神经变性病症(例如亨廷顿病或ALS)相关的靶RNA。例如, RNA编辑可以用于治疗疾病或病毒感染。

[0356] RNA插入位点:被靶向或已经历了外源性多核苷酸插入的RNA位点。公开的方法包括使用公开的Cas13d蛋白质,其可以用于靶向RNA,用于在RNA插入位点处操纵。

[0357] 序列同一性/相似性:氨基酸(或核苷酸)序列之间的相似性以序列之间的相似性(或称为序列同一性)表示。序列同一性经常以同一性(或相似性或同源性)百分比来测量;百分比越高,两条序列越相似。

[0358] 用于进行比较的序列比对方法是本领域公知的。各种程序和比对算法描述于:Smith和Waterman, Adv. Appl. Math. 2:482, 1981; Needleman和Wunsch, J. Mol. Biol. 48:443, 1970; Pearson和Lipman, Proc. Natl. Acad. Sci. U.S.A. 85:2444, 1988; Higgins和Sharp, Gene 73:237, 1988; Higgins和Sharp, CABIOS 5:151, 1989; Corpet et al., Nucleic Acids Research 16:10881, 1988; 以及Pearson和Lipman, Proc. Natl. Acad. Sci. U.S.A. 85:2444, 1988. Altschul et al., Nature Genet. 6:119, 1994呈现了序列比对方法和同源性计算的详细考虑。

[0359] NCBI基本局部比对搜索工具(The NCBI Basic Local Alignment Search Tool, BLAST) (Altschul et al., J.Mol.Biol.215:403,1990)可以从几个来源获得,这些来源包括国家生物技术信息中心(the National Center for Biotechnology Information, NCBI, Bethesda, MD)和互联网,用于与序列分析程序blastp、blastn、blastx、tblastn和tblastx结合使用。如何使用该程序确定序列同一性的描述可在互联网上的NCBI网站上获得。

[0360] 本领域已知和本文公开的蛋白质和核酸序列的变体的特征通常在于,在使用NCBI Blast2.0, gapped Blastp设置为默认参数的氨基酸序列的全长比对上,具有至少约80%、至少90%、至少95%、至少96%、至少97%、至少98%或至少99%的序列同一性。为了比较大于约30个氨基酸的氨基酸序列,使用设置为默认参数的默认BLOSUM62矩阵(缺口存在值(gap existence cost)为11,每残基缺口值(gap cost)为1)来使用Blast2序列函数。当比对短肽(少于约30个氨基酸)时,比对应使用Blast2序列函数进行,使用设定为默认参数的PAM30矩阵(开放缺口9,延伸缺口1罚分)。当通过该方法评估时,与参考序列具有甚至更高相似性的蛋白质将会显示增加的同一性百分比,例如至少95%、至少98%或至少99%序列同一性。当比较少于完整序列的序列同一性时,同源物和变体将通常在10-20个氨基酸的短窗口内具有至少80%的序列同一性,并且根据它们与参考序列的相似性,可以具有至少85%或至少90%或至少95%的序列同一性。在互联网上的NCBI网站上可以获得在这些短窗口上确定序列同一性的方法。本领域技术人员将会认识到,这些序列同一性范围仅供指导;完全可能获得落在所提供范围之外的强显著性同源物。

[0361] 因此,在一实例中,Cas13d蛋白质具有与SEQ ID NO:1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100、101、102、103、104、105、106、107、108、109、110、111、112、113、138、147、19、153、155、158、160、162、164、166、168、170、175、177、179、181、183、185、187、189、194、198、200、202、204、206、208、210、212、214、216、218、220、222、224、226、229、231、233、235、237、239、241、243、245、247、249、251、253、278、279、280、281、282、283、284、285、292、293、294、295或296至少80%、至少85%、至少90%、至少91%、至少92%、至少95%、至少96%、至少97%、至少98%或至少99%的序列同一性。

[0362] 受试者:脊椎动物,例如哺乳动物,例如人。哺乳动物包括,但不限于,鼠类、猿猴、人、农场动物、运动动物和宠物。在一实施方式中,受试者是非人哺乳动物受试者,例如猴或其他非人灵长类动物、小鼠、大鼠、兔、猪、山羊、绵羊、狗、猫、马或牛。在一些实例中,受试者具有可以使用本文提供的方法治疗的病症(例如,病毒感染)或遗传疾病。在一些实例中,受试者具有可以使用本文提供的方法诊断的病症(例如,病毒感染)或遗传疾病。在一些实例中,受试者是实验室动物/生物体,例如斑马鱼、非洲爪蟾(Xenopus)、秀丽隐杆线虫(C.elegans)、果蝇(Drosophila)、小鼠、兔或大鼠。还包括体内获得或体外培养的生物实体的组织、细胞及其后代。

[0363] 治疗剂:是指一种或多种在施用给受试者后带来一些有益效果的分子或化合物。有益的治疗效果可以包括实现诊断确定;改善疾病、症状、病症或病理状况;减少或预防疾

病、症状、病症或病状的发作；并且通常对抗疾病、症状、病症或病理状况。

[0364] 转导、转化和转染：当病毒或载体将核酸分子转移到细胞中时，其“转导”细胞。当通过将核酸并入细胞基因组或通过游离型复制(episomal replication)，核酸被细胞稳定地复制时，细胞被转导入细胞中的核酸“转化”或“转染”。

[0365] 这些术语包括所有将核酸分子引入这些细胞的技术，包括用病毒载体转染、用质粒载体转化、和通过电穿孔、脂转染、粒子枪加速(particle gun acceleration)和本领域的其它方法导入裸DNA。在一些实例中，该方法是化学方法(例如磷酸钙转染)、物理方法(例如电穿孔、显微注射、粒子轰击)、融合(例如脂质体)、受体介导的胞吞作用(例如DNA-蛋白复合物、病毒包膜/衣壳-DNA复合物)和病毒如重组病毒的生物感染(Wolff, J.A., ed, Gene Therapeutics, Birkhauser, Boston, USA, 1994)。将核酸分子引入细胞中的方法是已知的(例如，参见美国专利第6,110,743号)。这些方法可以用于将细胞用公开的药剂转导，以操纵其基因组。

[0366] 转基因：外源性基因。

[0367] 治疗(Treating)、治疗(Treatment)和疗法(Therapy)：在损伤、病理或状况的减轻或改善中的任何成功或成功的标记，包括任何客观或主观参数，例如症状的减少、缓解、减轻或使状况对患者更耐受，减缓退化或衰退的速率，使退化的终点不那么衰弱，改善受试者的身体或精神健康，或延长存活时间。治疗可以通过客观或主观参数来评估；包括身体检查、血液和其它临床测试等的结果。为了预防性益处，公开的组合物可以施用于处于发展特定疾病、病症或症状的风险中的受试者，或施用于报告疾病的一种或多种生理症状的受试者，即使该疾病、病症或症状可能尚未显现。

[0368] 上调：当用于提及分子例如靶RNA的表达时，是指导致靶RNA的产生增加的任何过程。在一实例中，包括直接上调，例如，如果靶RNA参与具有其自身转录的反馈环。在一实例中，包括间接上调，例如，通过抑制性miRNA的敲低，其导致该miRNA的靶的激活。

[0369] 在一些实例中，靶RNA是编码RNA。在一些实例中，靶RNA是非编码RNA。可以被靶向上调的RNA分子的具体实例包括mRNA、miRNA、rRNA、tRNA、核RNA和结构RNA。在一些实例中，靶RNA的上调或激活包括增加靶RNA的翻译并由此可以增加相应蛋白质存在的过程。公开的CRISPR/Cas系统可以用于上调任何感兴趣的靶RNA。

[0370] 上调包括任何可检测的靶RNA的增加。在某些实例中，与对照(例如在相应的正常细胞或样品中检测到的靶RNA的量)相比，细胞或无细胞系统中可检测的靶RNA的表达增加至少20%、至少30%、至少40%、至少50%、至少60%、至少70%、至少75%、至少80%、至少90%、至少95%、至少100%、至少200%、至少400%或至少500%。在一实例中，对照是正常细胞(例如，不包括Cas13d或向导RNA的非重组细胞)中的相对表达量。

[0371] 在足以……的条件下：用于描述允许所需活性的任何环境的短语。在一实例中，所需活性是本文公开的Cas13d蛋白质的表达，其与其他必需元件组合，例如以调节靶RNA。

[0372] 载体：一种核酸分子，其中可以引入外来核酸分子而不破坏载体在宿主细胞中复制和/或整合的能力。载体包括，但不限于，单链、双链或部分双链的核酸分子；包括一个或多个游离端、无游离端(例如，环状)的核酸分子；包括DNA、RNA或二者的核酸分子；以及本领域已知的其它多核苷酸。

[0373] 载体可以包括允许其在宿主细胞中复制的核酸序列，例如复制起点。载体还可以

包括一个或多个选择标记基因或本领域已知的其他遗传元件。整合载体能够将其自身整合到宿主核酸中。表达载体是含有允许插入的一个或多个基因转录和翻译的必需调节序列的载体。

[0374] 一种类型的载体是“质粒”，其是指环状双链DNA环，其中可以插入另外的DNA片段，例如通过标准分子克隆技术插入。另一种类型的载体是病毒载体，其中病毒来源的DNA或RNA序列存在于载体中以包装到病毒（例如，逆转录病毒、复制缺陷逆转录病毒、腺病毒、复制缺陷腺病毒和腺相关病毒）中。病毒载体还包括由病毒携带的用于转染到宿主细胞中的多核苷酸。在一些实施方式中，载体是慢病毒（例如整合缺陷型慢病毒载体）或腺相关病毒（AAV）载体。

[0375] 某些载体能够在它们所导入的宿主细胞中自主复制（例如，具有细菌复制起点的细菌载体和游离型哺乳动物载体）。其它载体（例如，非游离型哺乳动物载体）在引入到宿主细胞中时整合到宿主细胞的基因组中，从而与宿主基因组一起复制。

[0376] 某些载体能够指导与它们可操作地连接的基因的表达。这些载体在此称作“表达载体”。常见的表达载体通常是质粒的形式。重组表达载体可以以适合在宿主细胞中表达核酸的形式包括本文提供的核酸（例如向导RNA[其可以从RNA序列表达或是RNA序列]、编码Cas13d蛋白质的核酸），这意味着重组表达载体包括一个或多个调节元件，该一个或多个调节元件可以基于用于表达的宿主细胞而选择，其与待表达的核酸序列可操作地连接。在重组表达载体内，“可操作地连接”是指感兴趣的核苷酸序列以允许该核苷酸序列表达的方式（例如，在体外转录/翻译系统中或当载体被引入宿主细胞时在宿主细胞中）与调节元件连接。本领域技术人员将会认识到，表达载体的设计可以取决于这些因素，例如待转化的宿主细胞的选择、所需的表达水平等。可以将载体引入到宿主细胞中，从而产生由本文所述核酸编码的转录物、蛋白质或肽，包括融合蛋白或肽（例如规律间隔成簇短回文重复序列（CRISPR）转录物、蛋白质、酶、其突变形式、其融合蛋白等）。

[0377] II. 若干实施方式的概述

[0378] 2类CRISPR-Cas系统赋予微生物多样的适应性免疫机制。本文提供了原核生物基因组和宏基因组序列的分析，以鉴定未表征的RNA引导的、RNA-靶向的CRISPR系统家族，该家族被分类成VI-D型。七种不同的直系同源物的生物化学表征和蛋白质改造产生了源自黄化瘤胃球菌XPD3002 (CasRx) 的核糖核酸酶效应子，其在人细胞中具有强活性。CasRx介导的敲低相对于跨多种内源性转录物的RNA干扰表现出高效率 and 特异性。作为最紧凑的单一效应子Cas酶之一，CasRx还可以灵活地包装到腺相关病毒中。病毒编码的无催化活性的CasRx可靶向前mRNA的顺式元件(cis-elements)，以操纵可变剪接，减轻额颞叶痴呆神经元模型中失调的tau同种型比例。本文的结果将CasRx作为可编程的RNA结合模块，用于有效地靶向细胞RNA，使得能够实现转录组改造和治疗方法的通用平台。

[0379] 在各种细菌和古细菌生命中均发现2类CRISPR系统。将CRISPR基因座的最小定义用于原核生物基因组和宏基因组序列的生物信息学挖掘（其仅要求CRISPR重复阵列和邻近蛋白质），本文提供了命名为VI型CRISPR-Cas13d的RNA靶向2类CRISPR系统的未表征的显著紧凑家族的鉴定。

[0380] 由于CRISPR系统通常作为20千碱基的基因组序列内的功能操纵子存在，因此即使是片段化的宏基因组读取也足以回收可用于生物改造目的的Cas酶。本文和其他人

(Shmakov et al., 2015) 描述的CRISPR基因组挖掘策略, 结合通过下一代测序对微生物群体进行分析的持续努力, 应当在机制上对基因组改造工具箱提供多样的补充。

[0381] 对Cas13d效应子的两种不同的核糖核酸酶特性进行了生物化学表征, 上述效应子经由不依赖于HEPN结构域的机制将CRISPR重复阵列加工成成熟向导, 随后对互补激活物RNA进行向导序列依赖性识别。这触发了HEPN介导的RNA酶活性, 使得Cas13d能够切割激活物和旁观者RNA, 这是其它RNA靶向CRISPR系统所共有的特性。Cas13d还没有表现出明显的侧翼序列要求, 并且发现其在拼接靶RNA的crRNA中均有活性, 表明了靶向任意单链RNA序列的能力。

[0382] 在人细胞中对从Cas13d家族的不同分支取样的Cas13d直向同源物的全面活性报告筛选揭示, 可以改造来自黄化瘤胃球XPD3002株 (CasRx) 的Cas13d的NLS融合物, 以在真核环境中用于可编程RNA靶向 (图8D)。CasRx融合物敲低了14种内源性mRNA和lncRNA的多样性组, 相对于RNA干扰、dCas9介导的CRISPR干扰和Cas13超家族的其他成员, 始终以有利的效率实现>90%的敲低 (图10A-10C)。此外, CasRx干扰的特异性明显比间隔子匹配的shRNA高, 与数百个RNA干扰相比, 没有可检测的脱靶变化。

[0383] CasRx是最小的双组分平台, 其包括改造的CRISPR-Cas13d效应子和相关的向导RNA, 并且可以是完全遗传编码的。由于CasRx是正交 (orthogonally) 递送的蛋白质, 因此HEPN-非活性dCasRx可以被改造成靶向特定RNA元件的柔性RNA结合模块。重要的是, 由于CasRx使用不同的核糖核酸酶活性来加工向导RNA, 因此dCasRx仍可以与重复序列阵列配对以用于多重应用。本文通过建立用于调节可变剪接和所得蛋白质同种型比例的dCasRx剪接效应子融合物, 将其应用于额颞叶痴呆的神经元模型中, 显示了这一概念的实用性。

[0384] 平均大小为930aa的Cas13d是在哺乳动物细胞中表征的最小的2类CRISPR效应子。这允许CasRx效应子结构域融合物与编码多种向导RNA的CRISPR阵列配对, 同时保持在用于原代细胞和体内递送的通用腺相关病毒 (AAV) 递送载体 (Naldini, 2015) 的包装尺寸限制内。此外, 靶向AAV递送CasRx至特定的有丝分裂期后细胞类型例如神经元, 可以介导矫正载荷 (corrective payload) 的长期表达, 这避免了永久性遗传修饰或频繁的再施用 (Chiriboga et al., 2016), 补充其它核酸靶向技术例如DNA核酸酶编辑或反义寡核苷酸。据估计RNA错误剪接疾病占遗传疾病的高达15% (Hammond和Wood, 2011), 突出了能够多重靶向的经改造的剪接效应子的潜力。本文提供的材料可以用于RNA靶向敲低和剪接, 例如活细胞标记和基因筛选, 以进行转录物成像、运输或调节。CRISPR-Cas13d和改造的变体例如CasRx总体上能够实现灵活的核酸改造、转录组相关研究和治疗, 将基因组编辑工具箱从DNA扩展到RNA。

[0385] 本文提供了靶向 (例如, 修饰、检测) 一种或多种RNA分子的方法, 例如, 规律间隔成簇短回文重复序列 (CRISPR) -CRISPR相关的 (Cas) 系统介导的RNA编辑方法。这样的方法可以包括使一种或多种靶RNA分子接触非天然或改造的 (例如引入其的细胞或系统中天然地不存在的) CRISPR-Cas系统。因此, 在一些实例中, 公开的CRISPR-Cas系统包括天然Cas13d蛋白质 (或编码序列) 和天然gRNA, 但是用于其中Cas1蛋白质 (或编码序列) 和gRNA天然地不存在的系统或细胞中。而且, gRNA分子内的间隔子序列是非天然的, 并且已经被修饰成与靶RNA分子互补。

[0386] 在一些实例中, 靶RNA是编码RNA。在一些实例中, RNA是非编码RNA。

[0387] 公开的CRISPR-Cas系统可以包括(1)至少一种Cas13d蛋白质或至少一种Cas13d核酸编码序列(例如编码至少一种Cas13d蛋白质的mRNA或载体);和(2)至少一种与靶RNA分子具有充分互补性的CRISPR-Cas系统向导核酸分子(例如gRNA)(或至少一种编码gRNA的核酸分子)。Cas13d蛋白质与gRNA形成复合物,并且gRNA将复合物引导至一种或多种靶RNA分子。这种靶向可以允许Cas13d-gRNA复合物修饰或检测一种或多种靶RNA分子。在一些实例中,使一种或多种靶RNA分子(或含有一种或多种靶RNA分子的细胞)接触包括至少一种Cas13d蛋白质和至少一种gRNA的复合物。在一些实例中,所述系统包括Mg²⁺。但是,在一些实例中,所述系统不要求Mg²⁺,例如,如果不需要靶RNA的切割。

[0388] 在一些实例中,使一种或多种靶RNA分子接触非天然或改造的CRISPR-Cas系统包括向含有一种或多种靶RNA分子的细胞(例如真核细胞或原核细胞)中引入非天然或改造的CRISPR-Cas系统,例如,使用内吞作用(例如,受体介导的内吞作用、微胞饮作用(micropinocytosis)、脂质体、颗粒、外来体、微泡、基因枪、电穿孔、病毒、RNP-抗体融合(例如,通过将Cas13d RNP拴(tethering)至抗体、抗体片段或其他靶向部分[例如ScFv、适配体、DARPin、纳米抗体(nanobodies)、亲和体(affibodies)等]上,RNP可以被内吞到细胞中,可以想象RNP可以被拴至许多其他的实体上)或其组合。因此,细胞可以被转化、转导、转染,或者以其他方式与公开的CRISPR-Cas系统的适当核酸分子接触。产生的细胞是重组细胞。在一些实例中,使一种或多种靶RNA分子接触非天然或改造的CRISPR-Cas系统包括使含有一种或多种靶RNA分子的无细胞系统(例如生物或环境样品或细胞裂解物)接触非天然或改造的CRISPR-Cas系统(例如,在检测靶RNA的诊断方法中)。

[0389] 在一些实例中,使用至少2种、至少3种、至少4种、至少5种、至少10种或至少20种不同的gRNA。例如,这些方法可以包括靶向至少2种、至少3种、至少4种、至少5种、至少10种或至少20种不同的靶RNA分子,靶向一种或多种RNA分子的至少2个、至少3个、至少4个、至少5个、至少10个或至少20个不同区域,或者其组合。

[0390] 另外提供了编码这些Cas13d蛋白质的分离的核酸分子,例如,cDNA、基因组DNA、RNA或mDNA。这些分离的核酸分子可以是载体(例如质粒或病毒载体)的部分,并且可以与启动子可操作地连接。在一些实例中,编码Cas13d蛋白质的分离的核酸分子与SEQ ID NO: 124、125、126、127、128、139、140或141具有至少80%、至少85%、至少90%、至少92%、至少95%、至少96%、至少97%、至少98%、至少99%或100%的序列同一性;或者与SEQ ID NO: 114、115、116、117、118、119、120、121、122、123、142、143、144或145至少80%、至少85%、至少90%、至少92%、至少95%、至少96%、至少97%、至少98%、至少99%或100%的序列同一性。在其他的实例中,分离的核酸分子编码与SEQ ID NO: 1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100、101、102、103、104、105、106、107、108、109、110、111、112、113、138、147、149、153、155、158、160、162、164、166、168、170、175、177、179、181、183、185、187、189、194、198、200、202、204、206、208、210、212、214、216、218、220、222、224、226、229、231、233、235、237、239、241、243、245、247、249、251、253、278、279、280、281、282、283、284、285、292、293、294、295或296具有至少85%、至少90%、至少92%、至少95%、

至少96%、至少97%、至少98%、至少99%或100%的序列同一性的Cas13d蛋白质。

[0391] 在一些实例中,编码至少一种Cas13d蛋白质的分离的核酸分子(其可以是载体的部分)包括至少一个Cas13d蛋白质编码序列,其经密码子优化以在真核细胞中表达;至少一个Cas13d蛋白质编码序列,其经密码子优化以在人细胞中表达。在一实例中,这样的Cas13d编码序列与SEQ ID NO:114、115、116、117、118、119、120、121、122、123、142、143、144或145具有至少80%、至少85%、至少90%、至少92%、至少95%、至少96%、至少97%、至少98%、至少99%或100%的序列同一性,或者与SEQ ID NO:124、125、126、127、128、139、140或141具有至少80%、至少85%、至少90%、至少92%、至少95%、至少96%、至少97%、至少98%、至少99%或100%的序列同一性。在其他的实例中,真核细胞密码子优化的核酸序列编码与SEQ ID NO:1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100、101、102、103、104、105、106、107、108、109、110、111、112、113、138、147、149、153、155、158、160、162、164、166、168、170、175、177、179、181、183、185、187、189、194、198、200、202、204、206、208、210、212、214、216、218、220、222、224、226、229、231、233、235、237、239、241、243、245、247、249、251、253、278、279、280、281、282、283、284、285、292、293、294、295或296具有至少85%、至少90%、至少92%、至少95%、至少96%、至少97%、至少98%、至少99%或100%的序列同一性的Cas13d蛋白质。

[0392] 在一些实例中,与一种或多种靶RNA分子杂交的gRNA包括一个或多个同向重复(DR)序列、一个或多个间隔子序列或者一个或多个包括DR-间隔子-DR-间隔子的序列。在一些实例中,一个或多个DR序列与SEQ ID NO:129、130、131、132、133、134、135、136、137、148、150、151、152、154、156、157、159、161、163、165、167、169、176、178、180、182、184、186、188、190、191、192、193、199、201、203、205、207、209、211、213、215、217、219、221、223、225、227、228、230、232、234、236、238、240、242、244、246、248、250、252或254具有至少80%、至少85%、至少90%、至少92%、至少95%、至少96%、至少97%、至少98%、至少99%或100%的序列同一性。在一实例中,gRNA包括额外的序列,例如适配体序列。

[0393] 在一些实例中,从单一阵列转录物加工多个gRNA,其中每个gRNA可以是不同的,例如,以靶向不同的RNA或者靶向单一RNA的多个区域。

[0394] 在一些实例中,DR在5'端截短1-10个核苷酸(例如1、2、3、4、5、6、7、8、9或10个核苷酸),例如,以表达成经预加工的成熟向导RNA。

[0395] 提供了靶向一种或多种靶RNA分子的方法。靶向RNA分子可以包括以下的一种或多种:切割一种或多种靶RNA分子或使一种或多种靶RNA分子产生切口,激活或上调一种或多种靶RNA分子,激活或抑制一种或多种靶RNA分子的翻译,使一种或多种靶RNA分子失活,可视化、标记或检测一种或多种靶RNA分子,结合一种或多种靶RNA分子,编辑一种或多种靶RNA分子,运输一种或多种靶RNA分子,和掩蔽一种或多种靶RNA分子。在一些实例中,修饰一种或多种靶RNA分子包括以下的一种或多种:RNA碱基置换、RNA碱基缺失、RNA碱基插入、靶RNA中的断裂、将RNA甲基化和将RNA去甲基化。

[0396] 在一些实例中,这些方法用于治疗疾病,例如人的疾病。在这些实例中,一种或多

种靶RNA分子与疾病相关。

[0397] 另外提供了分离的蛋白质,其包括非天然蛋白质。在一些实例中,蛋白质与SEQ ID NO:1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100、101、102、103、104、105、106、107、108、109、110、111、112、113、138、147、149、153、155、158、160、162、164、166、168、170、175、177、179、181、183、185、187、189、194、198、200、202、204、206、208、210、212、214、216、218、220、222、224、226、229、231、233、235、237、239、241、243、245、247、249、251、253、278、279、280、281、282、283、284、285、292、293、294、295或296具有至少80%、至少85%、至少90%、至少92%、至少95%、至少96%、至少97%、至少98%、至少99%或100%的序列同一性。在一些实例中,分离的蛋白质是来自原核生物基因组或宏基因组、肠道宏基因组、活性污泥宏基因组、厌氧消化池宏基因组、鸡肠道宏基因组、人肠道宏基因组、猪肠道宏基因组、牛肠道宏基因组、绵羊肠道宏基因组、山羊肠道宏基因组、水豚肠道宏基因组、灵长类肠道宏基因组、白蚁肠道宏基因组、粪便宏基因组、来自梭菌目或瘤胃菌科的基因组的Cas13d直向同源物。在一些实例中,Cas13d直向同源物包括来自白色瘤胃球菌、惰性真杆菌、黄化瘤胃球菌XPD3002株、黄化瘤胃球菌FD-1、未培养的真杆菌属TS28-c4095、未培养的瘤胃球菌属、双环瘤胃球菌或瘤胃球菌属CAG57的Cas13d直向同源物。这些蛋白质可以包括亚细胞定位信号。在一些实例中,这些蛋白质在至少一个天然HEPN结构域中包括突变。

[0398] 另外提供了分离的向导RNA(gRNA)分子。在一些实例中,分离的gRNA包括一个或多个同向重复(DR)序列,例如与SEQ ID NO:129、130、131、132、133、134、135、136、137、148、150、151、152、154、156、157、159、161、163、165、167、169、176、178、180、182、184、186、188、190、191、192、193、199、201、203、205、207、209、211、213、215、217、219、221、223、225、227、228、230、232、234、236、238、240、242、244、246、248、250、252或254具有至少80%、至少85%、至少90%、至少92%、至少95%、至少96%、至少97%、至少98%、至少99%或100%的序列同一性的序列。这种gRNA可以进一步包括一个或多个对靶RNA特异(例如与其互补)的间隔子序列。这些向导gRNA还可以任选地在DR的5'端截短1-10个核苷酸(例如1、2、3、4、5、6、7、8、9或10个核苷酸),例如,以产生经预加工的向导RNA。

[0399] 另外提供了核糖核蛋白(RNP)复合物,其包括一种或多种本文提供的Cas13d蛋白质和一种或多种本文提供的gRNA。

[0400] 另外提供了重组细胞,其包括本文提供的任意Cas13d蛋白质(或编码Cas13d的核酸分子)、任意gRNA、任意RNP复合物或任意载体。在一实例中,所述细胞不是细菌细胞。在一实例中,所述细胞是细菌细胞。

[0401] 另外提供了组合物,其包括本文提供的任意Cas13d蛋白质(或编码Cas13d的核酸分子)、任意gRNA或阵列、任意RNP复合物、任意分离的核酸分子、任意载体或任意细胞中的一种或多种。这些组合物可以包括药学可接受的载体。

[0402] 另外提供了试剂盒。这些试剂盒可以包括本文提供的任意Cas13d蛋白质(或编码Cas13d的核酸分子)、任意gRNA或阵列、任意RNP复合物、任意分离的核酸分子、任意载体、任

意细胞或任意组合物中的一种或多种。这些试剂可以是合并的,或者在单独的容器中。

[0403] 在一些实例中,通过将Cas13d蛋白质(或编码该蛋白质的核酸)与改造的RNA向导(或编码RNA向导的核酸)组合,可以针对其RNA靶对Cas13d蛋白质进行编程,所述改造的RNA向导由完整或部分的同向重复序列,后接与RNA靶互补的“间隔子”序列(或其变体,即,阵列(DR-间隔子-DR-间隔子-DR-间隔子等)或前向导RNA(DR-间隔子-DR)组成。通过突变保守的RNA酶HEPN基序(RXXXXH),Cas13d蛋白质可以被催化性失活,并转化到RNA结合模块中。本文提供了示例性Cas13d蛋白质和对应的向导(例如,SEQ ID NO:147-170、175-193和SEQ ID NO:198-254)。

[0404] A. Cas13d蛋白质

[0405] 本文提供了新的Cas13d蛋白质,例如序列表中所示的那些。SEQ ID NO:1、3、42、62、70、82、83、92、147、149、153、155、158、160、162、164、166、168、170、175、177、179、181、183、185、187、189、198、200、202、204、206、208、210、212、214、216、218、220、222、224、226、229、231、233、235、237、239、241、243、245、247、249、251、253、292、293、294、295和296提供了不同的全长蛋白质,SEQ ID NO:2、4-41、43-61、63-69、71-81、84-91、93-113、194、278、279、280、281、282、283、284和285提供了Cas13d变体和片段。这些蛋白质可以用于公开的方法、组合物和试剂盒中。

[0406] 在一些实例中,Cas13d蛋白质包括一个或多个(例如1或2个)天然HEPN结构域。在一些实例中,Cas13d蛋白质包括一个或多个突变的HEPN结构域(例如,突变Cas13d蛋白质可以加工gRNA,但不能修饰靶RNA)。在一些实例中,Cas13d蛋白质不大于150kD,不大于140kD,不大于130kD,不大于120kD,例如约90-120kD,约100-120kD,或约110kD。

[0407] 除本文提供的Cas13d蛋白质以外,本公开还包括来自原核生物基因组或宏基因组、肠道宏基因组、活性污泥宏基因组、厌氧消化池宏基因组、鸡肠道宏基因组、人肠道宏基因组、猪肠道宏基因组、牛肠道宏基因组、绵羊肠道宏基因组、山羊肠道宏基因组、水豚肠道宏基因组、灵长类肠道宏基因组、白蚁肠道宏基因组、粪便宏基因组、来自梭菌目或瘤胃菌科的基因组的Cas13d直向同源物,例如,来自白色瘤胃球菌、惰性真杆菌、黄化瘤胃球菌XPD3002株、黄化瘤胃球菌FD-1、未培养的真杆菌属TS28-c4095、未培养的瘤胃球菌属、双环瘤胃球菌或瘤胃球菌属CAG57的Cas13d直向同源物。

[0408] 在一些实例中,Cas13d蛋白质为至少800aa,至少900aa,或至少1000aa,例如,800至1200aa、850至1050aa或860-1040aa。

[0409] 1. 变体Cas13d序列

[0410] 本公开包括Cas13d蛋白质,其包括本文所提供的序列的变体(例如,SEQ ID NO:1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100、101、102、103、104、105、106、107、108、109、110、111、112、113、138、147、149、153、155、158、160、162、164、166、168、170、175、177、179、181、183、185、187、189、194、198、200、202、204、206、208、210、212、214、216、218、220、222、224、226、229、231、233、235、237、239、241、243、245、247、249、251、253、278、279、280、281、282、283、284、285、292、293、294、295或296的变体)。在一些实

例中,本文提供的Cas13d蛋白质可以含有一个或多个突变,例如单个插入、单个缺失、单个置换或其组合。在一些实例中,Cas13d蛋白质包括至少1个、至少5个、至少10个、至少15个、至少20个、至少25个、至少30个、至少40个、至少50个、至少75个、至少100个、至少200或至少300个aa插入,例如1-20个插入(例如,在N-或C-端或在蛋白质内,例如插入完整的小结构域),至少1个、至少5个、至少10个、至少15个、至少20个、至少25个、至少30个、至少40个、至少50个、至少75个、至少100个、至少200或至少300个aa缺失(例如缺失完整的小结构域),例如1-20个缺失(例如,在N-或C-端或在蛋白质内),至少1个、至少5个、至少10个、至少15个、至少20个、至少25个、至少30个aa置换,例如1-20个置换,或者其任意组合(例如,单个插入连同1-19个置换),但是保留结合与gRNA分子内的间隔子序列互补的靶RNA分子的能力和/或将向导阵列RNA转录物加工成gRNA分子的能力,和/或保留切割靶RNA的能力。在非保守区中具有缺失的示例性Cas13蛋白质显示于SEQ ID NOS:278、279、280、281、282、283、284和285中。在一些实例中,本公开提供了任意公开的Casd蛋白质的变体(例如SEQ ID NO:1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100、101、102、103、104、105、106、107、108、109、110、111、112、113、138、147、149、153、155、158、160、162、164、166、168、170、175、177、179、181、183、185、187、189、194、198、200、202、204、206、208、210、212、214、216、218、220、222、224、226、229、231、233、235、237、239、241、243、245、247、249、251、253、278、279、280、281、282、283、284、285、292、293、294、295或296),其具有1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29或30个氨基酸变化,但保留结合与gRNA分子内的间隔子序列互补的靶RNA分子的能力和/或将向导阵列RNA转录物加工成gRNA分子的能力。在一些实例中,任何公开的Cas13d蛋白质(例如SEQ ID NO:1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100、101、102、103、104、105、106、107、108、109、110、111、112、113、138、147、149、153、155、158、160、162、164、166、168、170、175、177、179、181、183、185、187、189、194、198、200、202、204、206、208、210、212、214、216、218、220、222、224、226、229、231、233、235、237、239、241、243、245、247、249、251或252,或其变体,例如SEQ ID NO:278、279、280、281、282、283、284、285、292、293、294、295或296,进一步包括1-8个氨基酸插入、1-15个氨基酸缺失、1-10个氨基酸置换或其任意组合(例如,1-15、1-4或1-5氨基酸缺失连同1-10、1-5或1-7个氨基酸置换),其保持结合与gRNA分子内的间隔子序列互补的靶RNA分子的能力和/或将向导阵列RNA转录物加工成gRNA分子的能力。在一实例中,这些变体肽可以通过使用标准方法例如定点诱变或PCR来操纵编码肽的核苷酸序列产生。这些变体也可以是化学合成的。

[0411] 在一些实例中,Cas13d蛋白质包括SEQ ID NO:195、196或197所示的基序。因此,在一些实例中,与SEQ ID NO:1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、

47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100、101、102、103、104、105、106、107、108、109、110、111、112、113、138、147、149、153、155、158、160、162、164、166、168、170、175、177、179、181、183、185、187、189、194、198、200、202、204、206、208、210、212、214、216、218、220、222、224、226、229、231、233、235、237、239、241、243、245、247、249、251、253、278、279、280、281、282、283、284、285、292、293、294、295或296具有至少80%、至少85%、至少90%、至少92%、至少95%、至少96%、至少97%、至少98%、至少99%或100%的序列同一性的Cas13d蛋白质包括至少一个SEQ ID NO:195、196或197所示的基序。

[0412] 在一些实例中,Cas13d蛋白质包括SEQ ID NO:288、289、290或291所示的保守序列。因此,在一些实例中,与SEQ ID NO:1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100、101、102、103、104、105、106、107、108、109、110、111、112、113、138、147、149、153、155、158、160、162、164、166、168、170、175、177、179、181、183、185、187、189、194、198、200、202、204、206、208、210、212、214、216、218、220、222、224、226、229、231、233、235、237、239、241、243、245、247、249、251、253、278、279、280、281、282、283、284、285、292、293、294、295或296具有至少80%、至少85%、至少90%、至少92%、至少95%、至少96%、至少97%、至少98%、至少99%或100%的序列同一性的Cas13d蛋白质包括至少一个SEQ ID NO:288、289、290或291所示的基序。

[0413] 修饰或突变的一种类型包括将氨基酸替换成具有相似生物化学特性的氨基酸残基,即,保守置换(例如,1-4、1-8、1-10或1-20个保守置换)。典型地,保守置换几乎不影响直至不影响产生的肽的活性。例如,保守置换是Cas13d蛋白质(例如SEQ ID NO:1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100、101、102、103、104、105、106、107、108、109、110、111、112、113、138、147、149、153、155、158、160、162、164、166、168、170、175、177、179、181、183、185、187、189、194、198、200、202、204、206、208、210、212、214、216、218、220、222、224、226、229、231、233、235、237、239、241、243、245、247、249、251、253、278、279、280、281、282、283、284、285、292、293、294、295或296)中氨基酸置换,其几乎不影响Cas13d蛋白质结合与gRNA分子内的间隔子序列互补的靶RNA分子的能力和/或将向导阵列RNA转录物加工成gRNA分子的能力。丙氨酸扫描可以用于鉴定Cas13d蛋白质(例如SEQ ID NO:1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100、101、102、103、104、105、106、107、108、109、110、111、112、113、138、147、149、153、155、158、160、

162、164、166、168、170、175、177、179、181、183、185、187、189、194、198、200、202、204、206、208、210、212、214、216、218、220、222、224、226、229、231、233、235、237、239、241、243、245、247、249、251、253、278、279、280、281、282、283、284、285、292、293、294、295或296)中那些氨基酸可以耐受氨基酸置换。在一实例中,当丙氨酸或其他保守氨基酸置换成1-4、1-8、1-10或1-20个天然氨基酸时,变体Cas13d蛋白质(例如1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100、101、102、103、104、105、106、107、108、109、110、111、112、113、138、147、149、153、155、158、160、162、164、166、168、170、175、177、179、181、183、185、187、189、194、198、200、202、204、206、208、210、212、214、216、218、220、222、224、226、229、231、233、235、237、239、241、243、245、247、249、251、253、278、279、280、281、282、283、284、285、292、293、294、295或296)在CRISPR/Cas系统中修饰基因表达的能力,改变不超过25%,例如不超过20%,例如不超过10%。可以置换蛋白质中的原始氨基酸并且视作为保守置换的氨基酸实例包括:Ser置换Ala;Lys置换Arg;Gln或His置换Asn;Glu-置换Asp;Ser置换Cys;Asn置换Gln;Asp置换Glu;Pro置换Gly;Asn或Gln置换His;Leu或Val置换Ile;Ile或Val置换Leu;Arg或Gln置换Lys;Leu或Ile置换Met;Met、Leu或Tyr置换Phe;Thr置换Ser;Ser置换Thr;Tyr置换Trp;Trp或Phe置换Tyr;以及Ile或Leu置换Val。

[0414] 鉴定特别适合插入、置换或缺失的区域的一种方法是靶向直向同源物之间表现出低水平保守性的氨基酸链段。这些区域在图1B中提供的Cas13d蛋白质比对的保守性图中示出。Cas13d的保守残基在图18A-18MMM中提供的蛋白质比对中加以进一步标记(由比对的保守残基下方的符号“.”“:”或“*”示出)。图16A-16B中进一步提供了缺失及其功能测试的实例。这些方法用于产生SEQ ID NO:278、279、280、281、282、283、284和285所示的Cas13d功能性缺失变体。这些变体可以进一步包括变化,例如1-30、1-20或5-30个氨基酸置换,例如保守氨基酸置换。

[0415] 另一类型的置换可以通过将一种直向同源物的部分用另一直向同源物的同源性区域交换,获得组合的“嵌合”蛋白质。这种嵌合蛋白质可以组合多种Cas13d直向同源物的有益特性。

[0416] 更实质性的变化可以通过使用不太保守的置换做出,例如,选择在保持以下特性的影响上差异更大的残基:(a)置换区域中多肽骨架的结构,例如,作为折叠或螺旋构象;(b)目标位点处多肽的电荷或疏水性;或(c)侧链的大小。通常预期产生多肽功能变化最大的置换是以下置换,其中:(a)亲水性残基例如丝氨酸或苏氨酸置换(或置换成)疏水性残基例如亮氨酸、异亮氨酸、苯丙氨酸、缬氨酸或丙氨酸;(b)半胱氨酸或脯氨酸置换(或置换成)其他残基;(c)具有带正电的侧链的残基,例如赖氨酸、精氨酸或组氨酸,置换(或置换成)带负电的残基例如谷氨酸或天冬氨酸;或(d)具有庞大侧链的残基例如苯丙氨酸置换(或置换成)没有侧链的氨基酸例如甘氨酸。

[0417] 因此,本公开提供了与SEQ ID NO:1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、

67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100、101、102、103、104、105、106、107、108、109、110、111、112、113、138、147、149、153、155、158、160、162、164、166、168、170、175、177、179、181、183、185、187、189、194、198、200、202、204、206、208、210、212、214、216、218、220、222、224、226、229、231、233、235、237、239、241、243、245、247、249、251、253、278、279、280、281、282、283、284、285、292、293、294、295或296具有至少80%、至少85%、至少90%、至少92%、至少95%、至少96%、至少97%、至少98%、至少99%或100%的序列同一性的Cas13d蛋白质或其组合(例如嵌合体)。

[0418] 在一实施例中,Cas13d蛋白质包括非天然氨基酸。

[0419] 2. 具有其他元件(Elements)的Cas13d蛋白质

[0420] Cas13d蛋白质(例如SEQ ID NO:1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100、101、102、103、104、105、106、107、108、109、110、111、112、113、138、147、149、153、155、158、160、162、164、166、168、170、175、177、179、181、183、185、187、189、194、198、200、202、204、206、208、210、212、214、216、218、220、222、224、226、229、231、233、235、237、239、241、243、245、247、249、251、253、278、279、280、281、282、283、284、285、292、293、294、295或296中任意者)可以包括其他元件或结构域,例如,在N-或C-端(或两端)。产生的蛋白质可以称作Cas13d融合蛋白。

[0421] 在一实例中,本文提供的Cas13d蛋白质(例如天然Cas13d、截短的Cas13d或具有突变HEPN结构域的Cas13d)包括亚细胞定位信号。示例性亚细胞定位信号包括细胞器定位信号,例如核定位信号(NLS)、核输出信号(NES)或线粒体定位信号。在一实例中,Cas13d蛋白质包括NLS,例如SPKKKRKVEAS(SEQ ID NO:256;例如,由AGCCCAAGAAgAAGAGaAAGGTGGAGGC CAGC,SEQ ID NO:257编码)或GPKKKRKVAAA(SV40大T抗原NLS,SEQ ID NO:258;例如,由gga cctaagaaaaagaggaaggtggcggccgct,SEQ ID NO:260编码)。可以作为Cas13d蛋白质的部分的示例性NES包括腺病毒5型E1B核输出序列、HIV核输出序列(例如,参见SEQ ID NO:287)、MAPK核输出序列或PTK2核输出序列。

[0422] 在一些实例中,至少一种Cas13d蛋白质(例如天然Cas13d、截短的Cas13d或具有突变HEPN结构域的Cas13d)进一步包括一个或多个效应子结构域。示例性效应子结构域包括蛋白质和/或酶,例如,可以切割RNA的那些(例如,PIN内切核酸酶结构域、NYN结构域、来自SOT1的SMR结构域或来自葡萄球菌核酸酶的RNA酶结构域),可以影响RNA稳定性的那些(例如,来自UPF1、EXOSC5和STAU1的三四增蛋白(tristetraprolin, TTP)或结构域),可以编辑核苷酸或核糖核苷酸的那些(例如,胞苷脱氨酶、PPR蛋白质、腺苷脱氨酶、ADAR家族蛋白质或APOBEC家族蛋白质),可以激活翻译的那些(例如,eIF4E和其它翻译起始因子、酵母多聚(A)结合蛋白或GLD2的结构域),可以抑制翻译的那些(例如,Pumilio或FBF PUF蛋白质、脱腺苷酶(deadenylase)、CAF1、Argonaute蛋白质),可以将RNA甲基化的那些(例如,来自m6A甲基转移酶因子的结构域,例如METTL14、METTL3或WTAP),可以将RNA去甲基化的那些(例如,人烷基化修复同源物5或Alkbh5),可以影响剪接的那些(例如SRSF1的RS-富含结构域、

hnRNP A1的Gly-富含结构域、RBM4的丙氨酸富含基序或DAZAP1的脯氨酸富含基序),可以进行亲和纯化或免疫沉淀的那些(例如FLAG、HA、生物素或HALO标签),以及可以进行基于邻近性的蛋白质标记和鉴定的那些(例如生物素连接酶(例如BirA)或过氧化物酶(例如APEX2),以便将与靶RNA相互作用的蛋白质生物素化)。

[0423] 在一些实例中,Cas13d蛋白质和效应子模块组合可以构成转录传感器(transcriptional sensor)。例如,转录传感器可以包括至少一种具有突变HEPN结构域(例如,SEQ ID NO:2或4)的Cas13d蛋白、至少一种含有至少一个对靶RNA特异的间隔子序列的gRNA、和效应子模块例如任选地分裂(split)荧光蛋白或探针(例如,分裂Venus荧光蛋白、分裂GFP、分裂增强GFP、分裂mCherry、分裂超折叠mCherry、和其它荧光蛋白变体例如ECFP、YFP、RFP及其衍生物或片段);任选地分裂发光蛋白或探针(例如Gaussia、萤火虫、NanoLuc或Renilla变体);任选地分裂酶(例如泛素或TEV蛋白酶);FRET相容性蛋白质对;一种或多种经由可切割接头与Cas13d融合的转录因子(例如人工GAL4、锌指(zinc finger)、转录激活因子样效应子(TALE)、CRISPR-Cas9、CRISPR-Cpf1或基于TetR的转录因子或内源性转录因子);反式剪接蛋白质以恢复其功能的分裂内含肽,例如转录因子(例如,来自海洋嗜热盐菌(*Rhodothermus marinus*)或DnaE的内含肽);经磷酸化激活的激酶-底物对(例如TYK2-STAT3);一种、两种或更多种在二聚化或多聚化时激活的单体(例如胱天蛋白酶9);或一种或多种在相互作用时诱导构象和功能变化的蛋白质。在一实例中,由于结合特定转录物,两个或多个Cas13d蛋白质和gRNA的空间接近性将会激活效应子模块,导致细胞中可检测信号或可检测活性。

[0424] 在一实例中,效应子结构域与特异性识别并结合RNA适配体的蛋白质融合,上述RNA适配体例如能附加至或插入gRNA分子的RNA适配体(例如MS2、PP7、QB和其他适配体)。由于Cas13d和gRNA复合物将引导适配体蛋白质-效应子结构域接近靶RNA,该适配体-效应子结构域融合可以用于靶向靶RNA。

[0425] 在另一实例中,适配体可以直接插入到gRNA分子中,以允许靶RNA的检测,例如荧光团适配体(例如,菠菜(Spinach)、芒果(Mango)等)。

[0426] 在一些实例中,Cas13d蛋白质(例如天然Cas13d、截短的Cas13d或具有突变HEPN结构域的Cas13d)包括纯化标签,例如HA-标签、His-标签(例如6-His)、Myc-标签、E-标签、S-tag、钙调蛋白标签、FLAG-标签、GST-标签、MBP-标签等。在一些实例中,这些标签位于Cas13d蛋白质的N-或C-端。

[0427] 在一些实例中,Cas13d蛋白质(例如天然Cas13d、截短的Cas13d或具有突变HEPN结构域的Cas13d)包括一个或多个亚细胞定位信号、效应子结构域和纯化标签。

[0428] 在一些实例中,Cas13d蛋白可以分裂成多个片段,这些片段然后被单独表达。Cas13d的这些片段可以任选地与其它蛋白质结构域融合。在一实例中,Cas13d可以分成两个半部(halves),然后与诱导型异二聚体对的两部分融合。诱导型异二聚体结合后,Cas13d的两个半部彼此募集,以形成活性蛋白。这样的系统将允许Cas13d活性的诱导型控制。可用的异二聚体对包括光照后或通过施用小分子化合物等而二聚化的两种蛋白质。异二聚体对的具体实例包括但不限于:光诱导型Magnets蛋白、光诱导型iLID-SspB对、光诱导型Cryptochrome2-CIB1二聚体和小分子诱导型FKBP蛋白质。在分裂Cas13d设计的另一实例中,可以将Cas13d蛋白质的两个半部与蛋白质反式剪接结构域融合。这种设计将实现两个

半部的分开表达,一旦在细胞内表达,两个半部便重构为全长蛋白质。这种转剪接结构域的实例包括内含肽系统。

[0429] 鉴定特别适合蛋白质分裂的区域的一种方法是鉴定直向同源物之间表现出低水平保守性的氨基酸链段。这些区域在图1B中提供的Cas13d蛋白质比对的保守性图中示出。Cas13d的保守残基在图18A-18MMM中提供的Cas13d蛋白质比对中加以进一步标记(由比对的保守残基下方的符号“.”“:”或“*”示出)。

[0430] 3. Cas13d蛋白质的生成

[0431] 在一实例中,Cas13d蛋白质在体外表达,例如在原核细胞(例如细菌,如乳杆菌(Lactobacillus)、乳球菌(Lactococcus)、芽孢杆菌(Bacillus)(如枯草芽孢杆菌(B. subtilis))、埃希氏菌(Escherichia)(如大肠杆菌(E. coli))、鼠伤寒沙门氏菌(Salmonella typhimurium)和梭菌(Clostridium))、古细菌细胞、植物或植物细胞、真菌细胞(例如链孢霉属(Neurospora))、酵母细胞(例如酵母属(Saccharomyces)或毕赤酵母属(Pichia)(如酿酒酵母(S. cerevisiae)或巴斯德毕赤酵母(P. pastoris))、乳酸克鲁维酵母(Kluyveromyces lactis))、昆虫细胞(例如SF9细胞)或哺乳动物细胞(例如293细胞或永生哺乳动物骨髓和淋巴样细胞系)中表达。一经表达,Cas13d蛋白质可被分离和/或纯化(例如使用层析或免疫学分离)。在一些实例中,Cas13d蛋白质上的标签允许蛋白质从培养基中分离。示例性的方法包括硫酸铵沉淀、亲和柱、柱层析等(通常,参见R. Scopes, Protein Purification, Springer-Verlag, N.Y., 1982)。至少约90%至95%均一性例如98%至99%均一性的基本上纯的组合物可以用于本文提供的方法中。例如,Cas13d蛋白质的纯化制剂可以用作从CRISPR/Cas系统中的核酸分子表达Cas13d蛋白质的替代方案。

[0432] 除重组方法以外,本文公开的Cas13d蛋白质还可以使用天然化学连接和/或表达的蛋白质连接来整体地或部分地构建。

[0433] B. 编码Cas13d蛋白质的核酸分子

[0434] 本公开包括编码Cas13d蛋白质的核酸分子。核酸分子包括编码Cas13d肽的DNA、基因组DNA、cDNA、mRNA和RNA序列。这些核酸分子可以包括天然的或非天然的核苷酸或核糖核苷酸。编码SEQ ID NO:1、3、42、62、70、82、83、92和104的新Cas13d蛋白质的示例性核酸分子显示于SEQ ID NO:124-128、139、140和141中。另外还提供了密码子优化的编码新Cas13d蛋白质的核酸分子,例如,经优化以在哺乳动物细胞例如人细胞中表达的核酸分子(SEQ ID NO:114-123和142-145)。例如,SEQ ID NO:114、118和122提供经优化以在人细胞中表达的核酸分子。SEQ ID NO:115、119和123提供经优化以在人细胞中表达并且编码突变HEPN位点的核酸分子。SEQ ID NO:116和120提供经优化以在人细胞中表达并且包括N-端核定位(NLS)编码序列(即SPKKKRKVEAS)的核酸分子。SEQ ID NO:117和121提供经优化以在人细胞中表达并且包括N-端和C-端NLS编码序列(即,分别为SPKKKRKVEAS, SEQ ID NO:256和GPKKKRKVAAS SEQ ID NO:258)的核酸分子。

[0435] 在一实例中,核酸序列编码与SEQ ID NO:1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100、101、102、103、104、105、106、107、108、109、110、

111、112、113、138、147、149、153、155、158、160、162、164、166、168、170、175、177、179、181、183、185、187、189、194、198、200、202、204、206、208、210、212、214、216、218、220、222、224、226、229、231、233、235、237、239、241、243、245、247、249、251、253、278、279、280、281、282、283、284、285、292、293、294、295或296具有至少60%、至少70%、至少75%、至少80%、至少90%、至少92%、至少95%、至少96%、至少97%、至少99%或至少99%的序列同一性的Cas13d蛋白质。这些核酸分子可以基于本文提供的氨基酸序列和遗传密码产生。在一实例中，Cas13d核酸序列与SEQ ID NO:124、125、126、127、128、139、140或141具有至少70%、至少80%、至少85%、至少90%、至少92%、至少95%、至少98%或至少99%的序列同一性。在一实例中，Cas13d核酸序列经优化以在哺乳动物细胞例如人细胞中表达，例如与SEQ ID NO:114、115、116、117、118、119、120、121、122、123、142、143、144或145具有至少70%、至少80%、至少85%、至少90%、至少92%、至少95%、至少98%或至少99%的序列同一性的核酸序列。

[0436] 技术人员可以容易地构建多种含有功能上等同的核酸的克隆，例如序列不同但是编码相同Cas13d蛋白质序列的核酸。编码序列中的沉默突变由遗传密码的简并性(即冗余性)导致，其中一个以上密码子可以编码相同的氨基酸残基。因此，例如，亮氨酸可以由CTT、CTC、CTA、CTG、TTA或TTG编码；丝氨酸可以由TCT、TCC、TCA、TCG、AGT或AGC编码；天冬酰胺可以由AAT或AAC编码；天冬氨酸可以由GAT或GAC编码；半胱氨酸可以由TGT或TGC编码；丙氨酸可以由GCT、GCC、GCA或GCG编码；谷氨酰胺可以由CAA或CAG编码；酪氨酸可以由TAT或TAC编码；和异亮氨酸可以由ATT、ATC或ATA编码。显示标准遗传密码的表可见于多种来源(参见，例如，Stryer,1988,Biochemistry,3rd Edition,W.H.5Freeman and Co.,NY)。

[0437] 基于遗传密码，可以产生编码任何Cas13d序列的核酸序列。在一些实例中，这样的序列经优化以在宿主或靶细胞中表达，所述细胞例如用于表达Cas13d蛋白质的宿主细胞或者其中实施公开的方法的细胞(例如，在哺乳动物细胞例如人细胞中)。特定物种的密码子偏好性和密码子使用表可以用于改造编码Cas13d的分离的核酸分子(例如，编码与SEQ ID NO:1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100、101、102、103、104、105、106、107、108、109、110、111、112、113、138、147、149、153、155、158、160、162、164、166、168、170、175、177、179、181、183、185、187、189、194、198、200、202、204、206、208、210、212、214、216、218、220、222、224、226、229、231、233、235、237、239、241、243、245、247、249、251或253具有至少80%、至少85%、至少90%、至少92%、至少95%、至少96%、至少97%、至少98%、至少99%或100%的序列同一性的核酸分子，其利用该特定物种的密码子使用偏好性)。例如，本文公开的Cas13d蛋白质可以设计成具有所关注的特定生物体优先使用的密码子。在一实例中，Cas13d核酸序列经优化以在人细胞中表达，例如，与SEQ ID NO:114、115、116、117、118、119、120、121、122、123、142、143、144或145具有至少70%、至少80%、至少85%、至少90%、至少92%、至少95%、至少98%或至少99%的序列同一性的核酸序列。

[0438] 编码Cas13d蛋白质的核酸(例如编码与SEQ ID NO:1、2、3、4、5、6、7、8、9、10、11、

12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100、101、102、103、104、105、106、107、108、109、110、111、112、113、138、147、149、153、155、158、160、162、164、166、168、170、175、177、179、181、183、185、187、189、194、198、200、202、204、206、208、210、212、214、216、218、220、222、224、226、229、231、233、235、237、239、241、243、245、247、249、251、253、278、279、280、281、282、283、284、285、292、293、294、295或296具有至少80%、至少85%、至少90%、至少92%、至少95%、至少96%、至少97%、至少98%、至少99%或100%的序列同一性的蛋白质的核酸)可以通过体外方法克隆或扩增,例如聚合酶链式反应(PCR)、连接酶链式反应(LCR)、基于转录的扩增系统(TAS)、自主序列复制系统(3SR)和QB复制酶扩增系统(QB)。此外,编码Cas13d蛋白质的核酸(例如编码与SEQ ID NO:1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100、101、102、103、104、105、106、107、108、109、110、111、112、113、138、147、149、153、155、158、160、162、164、166、168、170、175、177、179、181、183、185、187、189、194、198、200、202、204、206、208、210、212、214、216、218、220、222、224、226、229、231、233、235、237、239、241、243、245、247、249、251、253、278、279、280、281、282、283、284、285、292、293、294、295或296具有至少80%、至少85%、至少90%、至少92%、至少95%、至少96%、至少97%、至少98%、至少99%或100%的序列同一性的蛋白质的核酸)可以通过克隆技术制备。合适的克隆和测序技术的实例以及足以指导技术人员进行克隆的说明可见于Sambrook et al. (ed.), *Molecular Cloning: A Laboratory Manual* 2nd ed., vol.1-3, Cold Spring Harbor Laboratory Press, Cold Spring, Harbor, N.Y., 1989; 和 Ausubel et al., (1987) in "Current Protocols in Molecular Biology," John Wiley and Sons, New York, N.Y.

[0439] 编码Cas13d蛋白质的核酸序列(例如编码与1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100、101、102、103、104、105、106、107、108、109、110、111、112、113、138、147、149、153、155、158、160、162、164、166、168、170、175、177、179、181、183、185、187、189、194、198、200、202、204、206、208、210、212、214、216、218、220、222、224、226、229、231、233、235、237、239、241、243、245、247、249、251、253、278、279、280、281、282、283、284、285、292、293、294、295或296具有至少80%、至少85%、至少90%、至少92%、至少95%、至少96%、至少97%、至少98%、至少99%或100%的序列同一性的蛋白质的核酸序列)可以通过任何合适的方法制备,包括,例如,克隆合适的序列,或者通过例如以下的方法直接化学合成:Narang et al., *Meth. Enzymol.* 68:90-99, 1979的磷酸三酯方法;Brown et al., *Meth. Enzymol.* 68:109-151, 1979的磷酸二酯方法;Beaucage et al., *Tetra. Lett.* 22:

1859-1862,1981的二乙基胺基磷方法;Beaucage&Caruthers,Tetra.Letts.22(20):1859-1862,1981所述的固相亚磷酸三酯方法,例如,使用例如Needham-VanDevanter et al.,Nucl.Acids Res.12:6159-6168,1984中所述的自动化合成仪;和美国专利第4,458,066号的固相载体方法。化学合成产生单链寡核苷酸。通过与互补序列杂交,或者通过使用单链作为模板用DNA聚合酶聚合,可以将该单链寡核苷酸转化成双链DNA。技术人员将会认识到,尽管DNA的化学合成通常限于大约100碱基的序列,但可以通过较短序列的连接得到更长的序列。

[0440] 在一实例中,Cas13d蛋白质(例如与SEQ ID NO:1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100、101、102、103、104、105、106、107、108、109、110、111、112、113、138、147、149、153、155、158、160、162、164、166、168、170、175、177、179、181、183、185、187、189、194、198、200、202、204、206、208、210、212、214、216、218、220、222、224、226、229、231、233、235、237、239、241、243、245、247、249、251、253、278、279、280、281、282、283、284、285、292、293、294、295或296具有至少80%、至少85%、至少90%、至少92%、至少95%、至少96%、至少97%、至少98%、至少99%或100%的序列同一性的蛋白质)通过将编码Cas13d蛋白质的cDNA插入到质粒或载体中制备。可以进行插入,使得Cas13d蛋白质在框内读出,从而产生Cas13d蛋白质。

[0441] Cas13d核酸编码序列(例如与SEQ ID NO:1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100、101、102、103、104、105、106、107、108、109、110、111、112、113、138、147、149、153、155、158、160、162、164、166、168、170、175、177、179、181、183、185、187、189、194、198、200、202、204、206、208、210、212、214、216、218、220、222、224、226、229、231、233、235、237、239、241、243、245、247、249、251、253、278、279、280、281、282、283、284、285、292、293、294、295或296具有至少80%、至少85%、至少90%、至少92%、至少95%、至少96%、至少97%、至少98%、至少99%或100%的序列同一性的序列)可以插入到表达载体中,该表达载体包括,但不限于质粒、病毒或其他运载体,其可以被操纵以允许序列的插入或并入,并可以在原核细胞或真核细胞中表达。宿主可以包括微生物、酵母、昆虫、植物和哺乳动物生物体。载体可以编码选择标记,例如胸苷激酶基因或抗生素抗性基因。

[0442] 编码Cas13d蛋白质的核酸序列(例如编码与SEQ ID NO:1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100、101、102、103、104、105、106、107、108、109、110、111、112、113、138、147、149、153、155、158、160、162、164、166、168、170、175、177、179、181、183、185、187、189、194、198、200、202、204、206、208、210、212、214、216、218、

220、222、224、226、229、231、233、235、237、239、241、243、245、247、249、251、253、278、279、280、281、282、283、284、285、292、293、294、295或296具有至少80%、至少85%、至少90%、至少92%、至少95%、至少96%、至少97%、至少98%、至少99%或100%的序列同一性的蛋白质的核酸序列)可以与表达控制序列可操作地连接。连接与Cas13d编码序列可操作性连接的表达控制序列,使得Cas13d蛋白质编码序列的表达在与表达控制序列相容的条件下实现。表达控制序列包括,但不限于适当的启动子、增强子、转录终止子、Cas13d蛋白质编码基因前的起始密码子(即ATG)、内含子的剪接信号、保持该基因的正确阅读框以允许mRNA正常翻译、以及终止密码子。

[0443] 在一实施方式中,载体用于在酵母例如酿酒酵母(*S.cerevisiae*)、毕赤酵母(*P.pastoris*)或乳酸克鲁维酵母(*Kluyveromyces lactis*)中表达。用于酵母表达系统的示例性启动子包括但不限于:组成型启动子质膜 H^+ -ATP酶(PMA1)、甘油醛-3-磷酸脱氢酶(GPD)、磷酸甘油酸激酶-1(PGK1)、醇脱氢酶-1(ADH1)和多效性耐药泵(PDR5)。此外,可以使用许多诱导型启动子,例如GAL1-10(由半乳糖诱导)、PH05(由低细胞外无机磷酸诱导)和串联热休克HSE元件(由温度升高至37°C诱导)。引导响应于可滴定的诱导物的可变表达的启动子包括甲硫氨酸响应性MET3和MET25启动子和铜依赖性CUP1启动子。这些启动子中的任意者可以克隆到多拷贝(2 μ)或单拷贝(CEN)质粒中,以提供表达水平的额外控制水平。质粒可以包括用于在酵母中进行选择的营养标记(例如URA3、ADE3、HIS1和其它)和用于在细菌中繁殖的抗生素抗性(AMP)。用于在乳酸克鲁维酵母上表达的质粒是已知的,例如pKLAC1。

[0444] 也可以制备编码Cas13d的病毒载体(例如编码与SEQ ID NO:1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100、101、102、103、104、105、106、107、108、109、110、111、112、113、138、147、149、153、155、158、160、162、164、166、168、170、175、177、179、181、183、185、187、189、194、198、200、202、204、206、208、210、212、214、216、218、220、222、224、226、229、231、233、235、237、239、241、243、245、247、249、251、253、278、279、280、281、282、283、284、285、292、293、294、295或296具有至少80%、至少85%、至少90%、至少92%、至少95%、至少96%、至少97%、至少98%、至少99%或100%的序列同一性的蛋白质的病毒载体)。示例性的病毒载体包括多瘤病毒、SV40、腺病毒、牛痘(vaccinia)病毒、腺相关病毒、疱疹病毒(包括HSV和EBV)、慢病毒、辛德毕斯(Sindbis)病毒、禽类、鼠类和人类来源的 α 病毒和逆转录病毒。可以使用杆状病毒(苜蓿银纹夜蛾多核多角体病毒(*Autographa californica multinuclear polyhedrosis virus*);AcMNPV)载体,其可获自商业来源。其它合适的载体包括逆转录病毒载体、正痘病毒载体、禽痘(avipox)载体、鸡痘(fowlpox)载体、羊痘(capripox)载体、猪痘(suipox)载体、腺病毒载体、疱疹病毒载体、 α 病毒载体、杆状病毒载体、辛德毕斯病毒载体、牛痘病毒载体和脊髓灰质炎病毒(poliovirus)载体。具体的示例性载体是痘病毒载体,例如牛痘病毒、鸡痘病毒和高度减毒的牛痘病毒(MVA)、腺病毒、杆状病毒等。使用的痘病毒包括正痘病毒、猪痘病毒、禽痘病毒和羊痘病毒。正痘(orthopox)包括牛痘、鼠痘(ectromelia)和浣熊痘。使用的正痘的一实例是牛痘。禽痘包括鸡痘、金丝雀痘和鸽痘。羊痘包括山羊痘和绵羊痘。在一实例中,猪痘(suipox)是猪痘

(swinepox)。可以使用的其它病毒载体包括其它DNA病毒，例如单纯疱疹病毒和腺病毒，以及RNA病毒，例如逆转录病毒和脊髓灰质炎病毒。

[0445] 编码Cas13d蛋白质的病毒载体(例如编码与SEQ ID NO:1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100、101、102、103、104、105、106、107、108、109、110、111、112、113、138、147、149、153、155、158、160、162、164、166、168、170、175、177、179、181、183、185、187、189、194、198、200、202、204、206、208、210、212、214、216、218、220、222、224、226、229、231、233、235、237、239、241、243、245、247、249、251、253、278、279、280、281、282、283、284、285、292、293、294、295或296具有至少80%、至少85%、至少90%、至少92%、至少95%、至少96%、至少97%、至少98%、至少99%或100%的序列同一性的蛋白质的病毒载体)可以包括至少一个与和编码Cas13d蛋白质的核酸序列可操作连接的表达控制元件。表达控制元件控制和调节Cas13d核酸序列的表达。可以使用的示例性表达控制元件包括,但不限于,lac系统,噬菌体 λ 的操作子和启动子区,酵母启动子和源自多瘤病毒、腺病毒、逆转录病毒或SV40的启动子。在一实例中,启动子是CMV、U6、CBh、CMW、Cbh、EF1a。在一实例中,启动子是细胞类型特异性启动子,例如突触蛋白(synapsin)或GFAP,或诱导型启动子,例如四环素诱导型启动子。其他的可操作性元件包括,但不限于,前导序列、终止密码子、聚腺苷酸化信号和编码Cas13d蛋白质的核酸序列在宿主系统中适当转录和后续翻译所必需的任何其它序列。表达载体可以含有对于含有核酸序列的表达载体在宿主系统中转移和后续复制所必需的额外元件。这些元件的实例包括,但不限于,复制起点和选择标记。

[0446] 在一实例中,载体包括Cas13d蛋白质编码序列之后的polyA信号、用于在病毒载体中表达的WPRE信号或其组合。

[0447] 在一实例中,该方法使用编码Cas13d蛋白质的mRNA的直接递送。

[0448] C. 向导核酸分子

[0449] 本公开提供了可以用于本文提供的方法、组合物和试剂盒中的向导核酸分子,例如向导RNA(gRNA或crRNA、CRISPR(向导)RNA)。这样的分子可以包括天然的或非天然的核苷酸或核糖核苷酸(例如LNA或其它化学修饰的核苷酸或核糖核苷酸,例如以保护向导RNA免受降解)。在一些实例中,向导序列是RNA。向导核酸可以包括修饰的碱基或化学修饰(例如,参见Latorre et al., *Angewandte Chemie* 55:3548-50, 2016)。向导序列将Cas13d蛋白质引导至靶RNA,从而靶向RNA(例如,修饰或检测RNA)。

[0450] 向导分子包括一个或多个称作间隔子的区域。间隔子与靶RNA序列具有充分的互补性,以便与靶RNA杂交,并引导Cas13d蛋白质与靶RNA的序列特异性结合。因此,间隔子是向导序列的可变部分。在一些实例中,间隔子与靶RNA(或者作为靶的RNA的区域)具有100%互补性,但是间隔子与靶RNA可以具有小于100%的互补性,例如与靶RNA具有至少80%、至少85%、至少90%、至少95%、至少98%或至少99%的互补性。

[0451] 向导序列还可以包括一个或多个同向重复(DR)。DR是向导的恒定部分,其含有强二级结构(图3C),这有利于Cas13d蛋白质和向导分子之间的相互作用。每种直向同源物具有略微不同的DR序列(例如,SEQ ID NO:129、130、131、132、133、134、135、136、137、148、

150、151、152、154、156、157、159、161、163、165、167和169)。在一实例中，gRNA包括至少一个与SEQ ID NO:129、130、131、132、133、134、135、136、137、148、150、151、152、154、156、157、159、161、163、165、167、169、176、178、180、182、184、186、188、190、191、192、193、199、201、203、205、207、209、211、213、215、217、219、221、223、225、227、228、230、232、234、236、238、240、242、244、246、248、250、252或254具有至少80%、至少85%、至少90%、至少92%、至少95%、至少96%、至少97%、至少98%、至少99%或100%的序列同一性的DR序列(例如1、2、3或4个这样的DR序列)。

[0452] 在一实例中，向导序列在其5' -端包括恒定的DR，在其3' 端包括可变的间隔子。在一实例中，包括序列DR-间隔子-DR-间隔子。在一些实例中，序列DR-间隔子重复2次或更多次，例如，至少3次或至少4次。该类型的序列称作向导阵列。

[0453] 向导分子通常以多种加工状态存在。在一实例中，未加工的向导RNA是36nt的DR、后接30-32nt的间隔子。向导RNA被Cas13d自身或其他RNA酶加工(截短/修饰)成更短的“成熟”形式。在一些实施方式中，未加工的向导序列的长度为大约或者至少大约30、35、40、45、50、55、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75或更多个核苷酸(nt)。在一些实施方式中，加工的向导序列为大约44-60nt(例如40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69或70nt)。在一些实施方式中，未加工的间隔子长度为大约28-32nt(例如25、26、27、28、29、30、31、32、33、34或35nt)，而成熟(加工的)间隔子可以为10-30nt、10-25nt、14-25nt、20-22nt或14-30nt(例如10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34或35nt)。在一些实施方式中，未加工的DR为大约36nt(例如30、31、32、33、34、35、36、37、38、39、40或41nt)，而加工的DR为大约30nt(例如25、26、27、28、29、30、31、32、33、34或35nt)。

[0454] 向导序列引导CRISPR复合物与靶RNA序列特异性结合的能力可以通过任何合适的测定来评估。例如，可以将包括待测试的向导序列的足以形成CRISPR复合物的CRISPR系统的组分，提供至具有相应的靶RNA分子的宿主细胞中，例如，通过用编码CRISPR序列组分的载体转染，然后评估靶序列中优先的切割。类似地，通过提供靶RNA、CRISPR复合物的组分，包括待测试的向导序列和与测试向导序列不同的对照向导序列，并比较测试和对照向导序列反应之间在靶RNA处的结合或切割速率，可以在试管中评价靶RNA序列的切割。也可以使用其他测定，并且是本领域技术人员将会想到的。

[0455] 另外提供了载体，例如病毒载体或质粒(例如，逆转录病毒、慢病毒、腺病毒、腺相关病毒或单纯性疱疹病毒)，其包括向导核酸分子。本文公开了示例性载体。在一些实例中，向导核酸分子与启动子或表达控制元件(其实例在本申请别处提供)可操作地连接。如本文别处所述，这些载体可以包括其他元件，例如编码选择标记(例如抗生素，例如嘌呤霉素、潮霉素)或者可检测标记(例如GFP或其他荧光团)的基因。

[0456] 在一实例中，多种gRNA是阵列的部分(其可以是载体例如病毒载体或质粒的部分)。例如，包括序列DR-间隔子-DR-间隔子-DR-间隔子的向导阵列可以包括3个独特的未加工gRNA(每个DR-间隔子序列一个)。一经引入到细胞或无细胞系统中，阵列则被Cas13d蛋白质加工成3种单独的成熟gRNA。这允许将多种多路复用(multiplexing)，例如递送多种gRNA至细胞或系统，以靶向多种靶RNA或者单一靶RNA中的多个位置(或其组合)。

[0457] D. 编码Cas13d和向导核酸分子的载体

[0458] 本公开提供了载体,例如本文别处所述的质粒和病毒载体,其包括一个或多个向导分子编码序列(例如,以允许靶向一个或多个RNA分子)和一个或多个Cas13d蛋白质编码序列。这些载体可以用于本文提供的方法、组合物和试剂盒中。这些载体可以包括天然的或非天然的核苷酸或核糖核苷酸。这些载体可以包括与向导分子(其可以是包括至少两种不同向导分子的阵列的部分)和Cas13d蛋白质编码序列可操作连接的单个启动子。另外可选地,向导分子(其可以是包括至少两种不同向导分子的阵列的部分)和Cas13d蛋白质编码序列可以与不同的启动子可操作连接。在一些实例中,向导分子(其可以是包括至少两种不同向导分子的阵列的部分)和Cas13d蛋白质编码序列可以与启动子、增强子或二者可操作连接。

[0459] E. 重组细胞和无细胞系统

[0460] 提供了包括非天然Cas13d蛋白质、非天然Cas13d蛋白质编码序列、向导分子(或编码序列)或其组合的细胞。这些重组细胞可以用于本文提供的方法、组合物和试剂盒。可以将编码本文公开的Cas13d蛋白质的核酸分子和/或编码向导分子的核酸分子引入到细胞中,以产生转化的(例如重组的)细胞。在一些实例中,通过将一种或多种非天然Cas13d蛋白质和一种或多种向导分子(例如gRNA)引入到细胞中,例如作为核糖核蛋白(RNP)复合物,来产生上述这些细胞。

[0461] 类似地,提供了无细胞系统,例如由裂解的细胞产生那些(或在试管或其他容器中包括Cas13d RNP的那些,其中添加有体外转录或化学合成的靶RNA),其包括Cas13d蛋白质、Cas13d蛋白质编码序列、向导分子(或编码序列)或其组合。这些无细胞系统可以用于本文提供的方法、组合物和试剂盒。在一些实例中,将一种或多种非天然Cas13d蛋白质和一种或多种向导分子(例如gRNA)加入到无细胞系统中,例如,作为RNP复合物。

[0462] 因此,公开了含有Cas13d蛋白质(例如与SEQ ID NO:1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100、101、102、103、104、105、106、107、108、109、110、111、112、113、138、147、149、153、155、158、160、162、164、166、168、170、175、177、179、181、183、185、187、189、194、198、200、202、204、206、208、210、212、214、216、218、220、222、224、226、229、231、233、235、237、239、241、243、245、247、249、251、253、278、279、280、281、282、283、284、285、292、293、294、295或296具有至少80%、至少85%、至少90%、至少92%、至少95%、至少96%、至少97%、至少98%、至少99%或100%的序列同一性的蛋白质)的细胞和无细胞系统。类似地,提供了含有向导分子(例如具有至少一个与SEQ ID NO:129、130、131、132、133、134、135、136或137具有至少80%、至少85%、至少90%、至少92%、至少95%、至少96%、至少97%、至少98%、至少99%或100%的序列同一性的DR序列,并且在一些实例中,还具有至少一个与靶RNA互补的间隔子序列的向导分子)的细胞和无细胞系统。

[0463] 这样的重组细胞(例如,其可以用于产生无细胞系统)可以是真核的或原核的。这些细胞的实例包括,但不限于,细菌、古细菌、植物、真菌、酵母、昆虫和哺乳动物细胞,例如乳杆菌(Lactobacillus)、乳球菌(Lactococcus)、芽孢杆菌(Bacillus)(例如枯草芽孢杆菌(B. subtilis))、埃希氏菌属(Escherichia)(例如大肠杆菌)、梭状芽孢杆菌

(Clostridium)、酵母属(Saccharomyces)或毕赤酵母属(Pichia)(例如酿酒酵母或巴斯德毕赤酵母)、乳酸克鲁维酵母、鼠伤寒沙门氏菌(*Salmonella typhimurium*)、果蝇(*Drosophila*)细胞、秀丽隐杆线虫(*C.elegans*)细胞、非洲爪蟾(*Xenopus*)细胞、SF9细胞、C129细胞、293细胞、链孢霉属(*Neurospora*)和永生化哺乳动物细胞系(例如Hela细胞、髓样细胞系和淋巴样细胞系)。

[0464] 在一实例中,细胞是原核细胞,例如细菌细胞,例如大肠杆菌。

[0465] 在一实例中,细胞是真核细胞,例如哺乳动物细胞,例如人细胞。在一实例中,细胞是原代真核细胞、干细胞、肿瘤/癌细胞、循环肿瘤细胞(CTC)、血细胞(例如,T细胞、B细胞、NK细胞、Tregs等)、造血干细胞、特化免疫细胞(例如,肿瘤浸润淋巴细胞或肿瘤抑制淋巴细胞)、肿瘤微环境中的基质细胞(例如,癌症相关的成纤维细胞等)。在一实例中,细胞是中枢或外周神经系统的脑细胞(例如,神经元、星形胶质细胞、小胶质细胞、视网膜神经节细胞、视杆细胞/视锥细胞等)。

[0466] 在一实例中,细胞是生物样品的部分(或从生物样品获得),例如从受试者获得的含有基因组DNA、RNA(例如mRNA)、蛋白质或其组合的生物样本。实例包括,但不限于,外周血、血清、血浆、尿、唾液、痰、组织活检、细针抽吸物、手术标本和尸体解剖材料。这些细胞也可以用于产生无细胞系统。

[0467] 在一实例中,细胞(或无细胞系统)来自肿瘤,例如血液肿瘤(例如白血病,包括急性白血病(例如急性淋巴细胞性白血病、急性髓细胞性白血病、急性骨髓性白血病和成髓细胞性、早幼粒细胞性、骨髓单核细胞性、单核细胞性和红白血病)、慢性白血病(例如慢性髓细胞性(粒细胞性)白血病、慢性髓性白血病和慢性淋巴细胞性白血病)、真性红细胞增多症、淋巴瘤、霍奇金氏病、非霍奇金氏淋巴瘤(包括低度、中度和高度)、多发性骨髓瘤,华氏巨球蛋白血症(Waldenström's macroglobulinemia)、重链病、骨髓增生异常综合征、套细胞淋巴瘤和骨髓发育不良)或实体肿瘤(例如肉瘤和癌:纤维肉瘤、粘液肉瘤、脂肪肉瘤、软骨肉瘤、骨肉瘤和其它肉瘤,滑膜瘤、间皮瘤、尤文氏瘤(Ewing's tumor)、平滑肌肉瘤、横纹肌肉瘤、结肠癌、淋巴恶性肿瘤、胰腺癌、乳腺癌、肺癌、卵巢癌、前列腺癌、肝细胞癌、鳞状细胞癌、基底细胞癌、腺癌、汗腺瘤、皮脂腺癌、乳突癌、乳头状腺癌、髓样癌(medullary carcinoma)、支气管癌、肾细胞癌、肝癌、胆管癌、绒毛膜癌、维尔姆斯瘤(Wilms'tumor)、宫颈癌、睾丸癌、膀胱癌和CNS肿瘤(例如神经胶质瘤、星形细胞瘤、髓母细胞瘤(medulloblastoma)、颅咽管瘤、室管膜瘤、松果体瘤、成血管细胞瘤、听觉神经瘤、少突神经胶质瘤、膜瘤、黑素瘤、成神经细胞瘤和成视网膜细胞瘤)。

[0468] 在一实例中,细胞(或无细胞系统)获自环境样品,例如水、土壤或空气样品。

[0469] F. 组合物和试剂盒

[0470] 提供了包括Cas13d蛋白质、Cas13d蛋白质编码序列、向导分子(或编码序列)或其组合的组合物和试剂盒。在一实例中,组合物或试剂盒包括由一种或多种Cas13d蛋白质和一种或多种向导分子(例如gRNA)组成的RNP复合物。在一实例中,所述组合物或试剂盒包括编码Cas13d蛋白质的载体、向导分子或两者。在一实例中,组合物或试剂盒包括细胞,例如细菌细胞或真核细胞,其包括非天然Cas13d蛋白质、非天然Cas13d蛋白质编码序列、向导分子(或编码序列)或其组合。在一实例中,组合物或试剂盒包括无细胞系统,其包括非天然Cas13d蛋白质、非天然Cas13d蛋白质编码序列、向导分子(或编码序列)或其组合。

[0471] 这些组合物可以包括药学上可接受的载体(例如盐水、水、甘油、DMSO或PBS)。在一些实例中,组合物是液体、冻干粉末或冻存的。

[0472] 在一些实例中,试剂盒包括递送系统(例如脂质体、颗粒、外来体、微泡、病毒载体或质粒)和/或标记物(例如肽或抗体,其可以与Cas13d RNP直接缀合或者与含有Cas13d RNP的颗粒缀合,以引导细胞类型特异性摄取/提高内含体逃逸/实现血脑屏障跨越等)。在一些实例中,试剂盒进一步包括细胞培养基或生长培养基,例如适合于生长细菌、植物、昆虫或哺乳动物细胞的培养基。

[0473] 在一些实例中,试剂盒的这些部分位于单独的容器中。

[0474] G. 靶向RNA

[0475] 公开的Cas13d蛋白质(和编码序列)和向导分子(例如gRNA和编码序列)可以用于CRISPR/Cas系统中,以靶向一种或多种RNA分子,例如存在于样品(例如生物样品、环境样品(例如土壤、空气或水样品)等)中的那些。在一实例中,靶RNA是编码RNA。在一实例中,靶RNA是核RNA。在其他实例中,靶RNA是非编码RNA(例如功能性RNA、siRNA、microRNA、snRNA、snoRNA、piRNA、scaRNA、tRNA、rRNA、lncRNA或lincRNA)。这些RNA靶向方法可以在体外进行(例如,在细胞培养物中或在无细胞系统中),或者在体内进行(例如在生物体、胚胎或哺乳动物中)。

[0476] 本文提供的CRISPR/Cas系统包括两个通用组分:(1) Cas13d蛋白质或其编码序列(其表达可以由启动子驱动)和(2) 向导核酸分子,例如RNA(gRNA),其对靶RNA(其表达可以由启动子驱动)具有特异性。当引入到细胞(或无细胞系统中)时,例如,(1) 作为Cas13d mRNA和Cas13d gRNA,(2) 作为单一载体或质粒的部分或者分成多个载体或质粒,(3) 作为单独的Cas13d蛋白质和向导分子,或者(4) 作为Cas13d蛋白质和向导分子的RNP复合物,向导分子将Cas13d引导至靶RNA。如果Cas13d蛋白质具有天然HEPN结构域或者与携带RNA酶活性的适当效应子结构域融合,可以切割RNA。如果Cas13d蛋白质具有突变HEPN结构域,向导阵列可以加工成成熟gRNA,但不切割靶RNA。使用该系统,RNA序列容易被靶向,例如,被编辑或检测,任选地使用效应子结构域。

[0477] 1. 将Cas13d蛋白质直接引入到细胞中

[0478] 在一实例中,将Cas13d蛋白质在重组细胞例如大肠杆菌中表达,并纯化。然后,将所得纯化的Cas13d蛋白质连同对靶RNA具有特异性的适当向导分子一起,引入到细胞或生物体中,其中可以靶向一种或多种RNA。在一些实例中,Cas13d蛋白质和向导核酸分子作为单独的组分引入到目标细胞/生物体中。在其他实例中,纯化的Cas13d蛋白质与向导核酸(例如gRNA)复合,并且将该核糖核蛋白(RNP)复合物引入到目标细胞中(例如,使用转染或注射)。在一些实例中,将Cas13d蛋白质和向导分子注射到胚胎中(例如人、小鼠、斑马鱼或非爪蟾胚胎)。

[0479] 一旦Cas13d蛋白质和向导核酸分子位于细胞中,则可以靶向一种或多种RNA。

[0480] 2. 从核酸表达Cas13d

[0481] 在一实例中,Cas13d蛋白质由含有靶RNA(例如待检测或修饰的RNA)的细胞中的核酸分子表达。在一些这样的实例中,Cas13d蛋白质由载体表达,所述载体例如引入到细胞或无细胞系统中的病毒载体或质粒。这导致在细胞、生物体或系统中产生Cas13d蛋白质。此外,这些核酸分子可以在细胞/生物体/系统中与靶RNA特异性的向导核酸分子(例如gRNA)

共表达。

[0482] 在一实例中,多个质粒或载体用于RNA靶向。编码Cas13d的核酸分子可以例如在一种载体或质粒上提供,并且向导核酸分子(例如gRNA)在另一种质粒或载体上提供。可以将多种质粒或病毒载体混合,并同时或分别导入到细胞(或无细胞系统)中。

[0483] 在一些实例中,多种核酸分子由单一载体或质粒表达。例如,单一载体可以包括编码Cas13d的核酸分子,并且单独的载体可以包括向导分子。

[0484] 在一些实例中,多种不同的向导分子(例如gRNA)存在于单个阵列和/或载体上,每个靶对应一种向导分子(例如1、2、3、4、5或10种不同靶)。在一实例中,所述方法包括递送多种gRNA(例如至少2种、至少3种、至少4种、至少5种、至少10种、至少20种或至少50种不同的gRNA),上述gRNA是阵列(其可以是载体例如病毒载体或质粒的部分)的部分。一经引入到细胞或无细胞系统中,阵列被Cas13d蛋白质加工成单独的成熟gRNA。

[0485] 由载体表达的核酸分子可以处于启动子的控制下,并任选地含有选择标记(例如抗生素抗性)。

[0486] 在一些实例中,蛋白质和向导分子通过胚胎表达(例如斑马鱼或非洲爪蟾胚胎)。Cas13d蛋白质可以由注射的质粒DNA、注射的mRNA或者稳定地整合到动物基因组中的拷贝表达。gRNA可以直接注射,或者由质粒或稳定地整合到动物基因组中的拷贝表达。

[0487] 3. 靶

[0488] 通过公开的方法可以靶向一种或多种RNA,例如细胞、无细胞系统或生物体中的至少1种、至少2种、至少3种、至少4种或至少5种不同的RNA,例如1、2、3、4、5、6、7、8、9或10种不同的RNA。在一实例中,RNA与疾病相关,例如囊性纤维化、亨廷顿病、泰萨二氏病(Tay-Sachs)、脆性X综合征(Fragile X syndrome)、脆性X相关震颤/共济失调综合征、杜氏肌营养不良(Duchenne muscular dystrophy)、强直性肌营养不良、骨髓性肌萎缩、脊髓小脑性共济失调或家族性ALS。在一实例中,RNA与癌症相关(例如肺癌、乳腺癌、结肠癌、肝癌、胰腺癌、前列腺癌、骨癌、脑癌、皮肤癌(例如黑素瘤)或肾癌)。靶RNA的实例包括,但不限于,与癌症相关的那些(例如,BCR-ABL、Ras、Raf、p53、BRCA1、BRCA2、CXCR4、 β -连环蛋白、HER2和CDK4)。

[0489] 在一实例中,RNA与病毒感染有关,例如正链RNA病毒感染,例如小核糖核酸病毒(例如,蹄疫病毒科(Aphthoviridae)[例如,口蹄疫病毒(FMDV)])、心病毒科(Cardioviridae);肠道病毒科(如柯萨奇病毒、埃可病毒、肠道病毒和脊髓灰质炎病毒);鼻病毒科(Rhinoviridae)(鼻病毒);肝病毒科(Hepataviridae)(甲型肝炎病毒);囊膜病毒(Togaviruses)(其实例包括风疹;甲病毒(alphaviruses)(例如西方马脑炎病毒、东方马脑炎病毒和委内瑞拉马脑炎病毒));黄病毒(Flaviviruses)(其实例包括登革热病毒、西尼罗河病毒和日本脑炎病毒);杯状病毒科(Calciviridae)(包括诺如病毒和札幌病毒);或冠状病毒(Coronaviruses)(其实例包括SARS冠状病毒,如Urbani株);或负链RNA病毒感染,如正粘病毒(Orthomyxoviruses)(如流感病毒)、弹状病毒(Rhabdoviruses)(如狂犬病病毒)和副粘病毒(Paramyxoviruses)(其实例包括麻疹病毒、呼吸道合胞病毒或副流感病毒);或DNA病毒感染(如疱疹病毒(如水痘-带状疱疹病毒(Varicella-zoster virus),如Oka株;巨细胞病毒(cytomegalovirus);和单纯疱疹病毒(HSV)1型和2型)、腺病毒(如腺病毒1型和腺病毒41型)、痘病毒(Poxviruses)(如牛痘病毒)或细小病毒(Parvoviruses)(如细小病毒

B19)感染)。

[0490] 在一实例中, RNA与细菌感染或细菌感染的性质相关, 例如细菌抗性、持久性或抗生素抗性。这些RNA的检测可以用于诊断方法, 而在基于细胞或无细胞系统中编辑这些RNA可以用于治疗方法。

[0491] 4. 检测DNA的方法

[0492] 在一实例中, 靶向RNA的方法导致检测、可视化或标记靶RNA。例如, 通过使用至少一种具有突变HEPN结构域(例如SEQ ID NO:2或4)的Cas13d蛋白质、至少一种含有至少一个对靶RNA特异的间隔子序列和效应子模块, 靶RNA将会被Cas13d识别, 但不会被切割或产生缺口, 同时效应子模块被激活。在一些实例中, 这样的方法用于检测靶RNA。这样的方法可以在细胞或无细胞系统中使用, 以确定是否存在靶RNA, 例如在肿瘤细胞中是否存在靶RNA。在一些实例中, 细胞或无细胞系统获自组织样品、血液样品或唾液样品,

[0493] 在一实例中, 检测RNA的方法包括与荧光蛋白质或其他可检测标记物融合的Cas13d蛋白质连同含有对靶RNA特异的间隔子序列的gRNA。Cas13d与靶RNA的结合可以通过显微镜或其他成像方法可视化。在另一实例中, RNA适配体序列可以附加至gRNA分子或插入gRNA分子内, 例如MS2、PP7、QB和其他适配体。引入与荧光蛋白质或其他可检测标记融合的特异性结合这些适配体的蛋白质(例如MS2噬菌体外壳蛋白质)可以用于检测靶RNA, 因为Cas13d-gRNA-靶RNA复合物将经由适配体相互作用而被标记。

[0494] 在另一实例中, 检测RNA的方法是用于诊断或治疗的转录传感器(例如, 作为合成回路的部分)。例如, 转录传感器可以包括至少一种具有突变HEPN结构域(例如SEQ ID NO:2或4)的Cas13d蛋白质、至少一种含有至少一个对靶RNA特异的间隔子序列的gRNA, 和效应子模块, 例如任选地分裂荧光蛋白质或探针; 任选地分裂发光蛋白质或探针; 和任选地催化可检测反应的分裂酶, 例如泛素或TEV蛋白酶; FRET-相容性蛋白质对; 一种或多种经由可切割接头与Cas13d融合的转录因子; 反式剪接蛋白质以恢复其功能的分裂内含肽, 例如转录因子; 经磷酸化激活的激酶-底物对; 一种、两种或更多种经二聚或多聚激活的单体; 或者一种或多种经相互作用诱导构象和功能变化的蛋白质。在一实例中, 因结合特定转录物造成的两个或更多个Cas13d蛋白质与gRNA的空间上靠近将会激活效应子模块, 导致细胞中可检测的信号或可检测的活性。

[0495] 例如, 转录传感器可以允许检测癌症特异性转录物、炎症特异性转录物、疾病特异性转录或细胞状态特异性转录物。能够检测特定转录物的含有基于Cas13d的系统的合成回路可以编码条件逻辑(conditional logic), 例如, 要求靶检测上调或下调用于治疗应用的基因。

[0496] 在一实例中, 该方法导致与靶RNA结合的可检测药剂, 其可以被检测。例如, 可以使用两种单独的各自包括荧光团(例如GFP)的部分的Cas13d融合蛋白, 和两种具有不同间隔子序列的不同gRNA, 所述间隔子序列靶向紧邻的RNA区域。当两个半部与邻近的靶RNA结合时, 荧光团的两个部分形成完整的荧光团, 从而产生可检测的信号。

[0497] 在一实例中, 该方法导致RNA检测, 例如, 通过触发例如以下的响应: 第二基因的表达、蛋白质的修饰、蛋白质或RNA易位至不同位置、经由自杀基因诱导细胞死亡、诱导细胞增殖、实现第二功能的转基因的诱导、诱导DNA序列的永久性变化以实现存储过去转录事件的记忆、或者改变RNA以实现下拉(pulldown)。

[0498] 在一实施方式中,转录因子的两个半部可以经由分裂内含肽系统与两个单独的Cas13d连接。Cas13d蛋白质提供有两个具有不同间隔子序列的不同gRNA,所述间隔子序列靶向紧邻的RNA区域。与邻近的靶RNA结合后,分裂内含肽反式剪接重构的转录因子(TF),使得其可以易位至核,以发动靶基因或靶基因簇。在一实例中,靶基因可以是细胞中的内源性基因。在另一实例中,靶基因可以是在载体上表达的或通过基因改造引入的转基因,例如荧光蛋白或毒素。

[0499] 5. 在无细胞系统中检测RNA的方法

[0500] 在一实例中,在无细胞系统中检测靶RNA的方法导致可检测的标记物或酶活性。例如,通过使用至少一种Cas13d蛋白质(例如SEQ ID NO:3、42、62、70、82、83和92)、至少一种含有至少一个对靶RNA特异的间隔子序列的gRNA和可检测标记物,靶RNA将会被Cas13d识别。靶RNA与Cas13d的结合触发其RNA酶活性,这可以导致靶RNA以及可检测标记物的切割。

[0501] 在一实例中,可检测标记是与荧光探针和淬灭剂连接的RNA。完整的可检测RNA连接有荧光探针和淬灭剂,抑制荧光。一旦可检测RNA被Cas13d切割,荧光探针从淬灭剂释放,并显示出荧光活性。这样的方法可以用于确定裂解的细胞样品、裂解的组织样品、血液样品、唾液样品、环境样品(例如水、土壤或空气样品)或其他裂解的细胞或无细胞样品中是否存在靶RNA。这样的方法还可以用于检测病原体,例如病毒或细菌,或者诊断疾病状态,例如癌症。

[0502] 在一实例中,靶RNA的检测有助于诊断疾病和/或病理学状态,或者病毒或细菌感染存在。例如,如果在患者的尿中检测到的话,非编码RNA例如PCA3的Cas13d-介导的检测可以用于诊断前列腺癌。在另一实例中,胃癌生物标志物lncRNA-AA174084的Cas13d-介导的检测可以用于诊断胃癌。

[0503] 6. 编辑靶RNA的方法

[0504] 在一实例中,靶向RNA的方法导致编辑靶RNA的序列,例如,通过使用具有非突变HEPN结构域的Cas13d蛋白质(例如SEQ ID NO:1、3、42、62、70、82、83或92)和含有至少一个对靶RNA特异的间隔子序列的gRNA,靶RNA可以在精确的位置处被切割或产生缺口。在一些实例中,这样的方法用于降低靶RNA的表达,这将会降低对应蛋白质的翻译。这样的方法可以用于其中不希望RNA的表达增加的细胞中。在一实例中,RNA与疾病相关,例如囊性纤维化、亨廷顿病、泰萨二氏病、脆性X综合征、脆性X相关震颤/共济失调综合征、肌肉萎缩症、强直性肌营养不良、骨髓性肌萎缩、脊髓小脑性共济失调或家族性ALS。在另一实例中,RNA与癌症相关(例如肺癌、乳腺癌、结肠癌、肝癌、胰腺癌、前列腺癌、骨癌、脑癌、皮肤癌(例如黑色素瘤)或肾癌)。靶RNA的实例包括,但不限于,与癌症相关的那些(例如,PD-L1、BCR-ABL、Ras、Raf、p53、BRCA1、BRCA2、CXCR4、 β -连环蛋白、HER2和CDK4)。编辑这些靶RNA可以具有治疗效果。

[0505] 在另一实例中,RNA在免疫细胞中表达。例如,靶RNA可以编码导致抑制希望的免疫反应(例如肿瘤浸润)的蛋白质。敲低这样的RNA可以实现这种所希望的免疫反应的进展(例如PD1、CTLA4、LAG3、TIM3)。在另一实例中,靶RNA编码导致不希望的免疫反应激活的蛋白质,例如,在自身免疫疾病例如多发性硬化、克罗恩氏病、狼疮或类风湿性关节炎的情况下。

[0506] 在一实例中,靶向靶RNA允许降低由RNA编码的靶蛋白质的表达。例如,通过使用具有突变HEPN结构域(例如SEQ ID NO:2或4)和翻译抑制结构域(例如Pumilio或FBP PUF蛋白

质、脱腺苷酶、CAF1、Argonaute蛋白质等)的Cas13d融合蛋白,以及含有至少一个对靶RNA特异的间隔子序列的向导RNA,可以抑制靶RNA的表达。

[0507] 在一些实例中,Cas13d可以与核糖核酸酶(例如PIN核酸内切酶结构域、NYN结构域、来自SOT1的SMR结构域或来自葡萄球菌核酸酶的RNA酶结构域)或影响RNA稳定性的结构域(例如三四增蛋白或来自UPF1、EXOSC5和STAU1的结构域)融合。

[0508] 在另一实例中,RNA适配体序列可以附加至gRNA分子或插入gRNA分子内,例如MS2、PP7、Q β 和其他适配体。特异性结合这些适配体的蛋白质例如MS2噬菌体外壳蛋白质可以与翻译抑制结构域、核糖核酸酶或影响RNA稳定性的结构域融合。这种适配体-效应子结构域融合可以用于靶向靶RNA,因为Cas13d和gRNA复合物将会引导适配体蛋白质-效应子结构域接近靶RNA。

[0509] 这样的方法可以用于其中不希望RNA的表达增加的细胞中,例如,当表达的RNA与疾病相关时,上述疾病例如囊性纤维化、亨廷顿病、泰萨二氏病、脆性X综合征、脆性X相关震颤/共济失调综合征、肌肉萎缩症、强直性肌营养不良、骨髓性肌萎缩、脊髓小脑性共济失调或家族性ALS。在另一实例中,靶RNA与癌症相关(例如肺癌、乳腺癌、结肠癌、肝癌、胰腺癌、前列腺癌、骨癌、脑癌、皮肤癌(例如黑素瘤)或肾癌)。靶RNA的实例包括,但不限于与癌症相关的那些(例如,PD-L1、BCR-ABL、Ras、Raf、p53、BRCA1、BRCA2、CXCR4、 β -连环蛋白、HER2和CDK4)。编辑这些靶RNA可以具有治疗效果。

[0510] 在另一实例中,RNA在哺乳动物细胞中表达。例如,靶RNA可以编码导致抑制希望的免疫反应例如肿瘤浸润的蛋白质。敲低这样的RNA可以实现这种所希望的免疫反应的进展(例如PD1、CTLA4、LAG3、TIM3)。在另一实例中,靶RNA编码导致不希望的免疫反应激活的蛋白质,例如,在自身免疫疾病例如多发性硬化、克罗恩氏病、狼疮或类风湿性关节炎的情况下。

[0511] 在一实例中,靶向靶RNA允许激活或增加靶RNA的表达。例如,通过使用具有突变HEPN结构域(例如SEQ ID NO:2或4)和翻译激活结构域(例如eIF4E和其他翻译起始因子、酵母聚(A)-结合蛋白质或GLD2的结构域)的Cas13d融合蛋白,以及含有至少一个对靶RNA特异的间隔子序列的向导RNA,可以增加靶RNA的表达。也可以使用与翻译激活结构域融合的同源适配体结合蛋白质将适配体引入到gRNA中。在一实例中,RNA适配体序列可以附加至gRNA分子或插入gRNA分子内,例如MS2、PP7、Q β 和其他适配体。与翻译激活结构域融合的特异性结合这些适配体的蛋白质,例如MS2噬菌体外壳蛋白质的引入可以用于靶向靶RNA,因为Cas13d和gRNA复合物将会使适配体蛋白质-翻译激活结构域邻近靶RNA。

[0512] 在一些实例中,这样的方法用于增加靶RNA的活性或表达,这将会增加相应蛋白质的翻译(如果该RNA是编码RNA)。这样的方法可以用于其中希望RNA表达增加的细胞中,例如拷贝数变化引起的杂合性遗传疾病或病症。增加所需蛋白质产物的翻译本质上是治疗性的。

[0513] 在另一实例中,增加靶RNA(例如Cyclin B1)的表达可以使靶细胞(例如癌细胞)对药物(例如化疗药)更加敏感。

[0514] 在一实例中,靶向靶RNA允许靶RNA中的一个或多个RNA碱基置换、RNA碱基编辑、RNA碱基缺失、RNA碱基插入或其组合。在一些实例中,具有突变HEPN结构域的Cas13d蛋白质经由直接融合或gRNA-适配体修饰而结合允许碱基编辑的效应子结构域(例如胞苷脱氨酶、

PPR蛋白质、腺苷脱氨酶、ADAR家族蛋白质或APOBEC家族蛋白质)。在一些实例中,这样的方法用于修饰RNA序列、编辑RNA突变或者修饰RNA转录物(例如,基因疗法),例如,以治疗疾病,例如ALS和黑素瘤或不希望的剪接位点导致的遗传疾病,例如Leber先天性黑朦(Leber congenital amaurosis)。

[0515] 在一实例中,靶向靶RNA允许将靶RNA甲基化。一些实例可以使用具有突变HEPN结构域(例如SEQ ID NO:2或4)的嵌合Cas13d蛋白质,以及含有至少一个对靶RNA特异的间隔子序列的向导RNA,所述嵌合Cas13d蛋白质经由直接融合或经由gRNA-适配体修饰与甲基化结构域(例如m6A)结合。在一些实例中,这样的方法用于对抗异常的RNA去甲基化。在一实例中,这样的方法用于调整多能性转录物例如NANOG或KLF4的甲基化水平,例如,降低其在乳腺癌细胞中的稳定性,这可以抑制与增殖增加和癌干细胞形成相关的乳腺癌干细胞表型的获得。

[0516] 在一实例中,靶向靶RNA允许将靶RNA去甲基化。一些实例可以使用具有突变HEPN结构域(例如SEQ ID NO:2或4)的Cas13d蛋白质、含有至少一个对靶RNA特异的间隔子序列的向导RNA和去甲基化结构域(例如人烷基化修复同源物5或Alkbh5)。去甲基化结构域可以经由直接融合或者经由gRNA-适配体修饰与Cas13d蛋白质结合。在一些实例中,这样的方法用于逆转异常的RNA甲基化,例如,通过降低m6A水平来治疗骨髓性白血病。

[0517] 在一实例中,靶向靶RNA允许与靶RNA结合。例如,通过使用具有突变HEPN结构域(例如SEQ ID NO:2或4)的Cas13d蛋白质和含有至少一个对靶RNA特异的间隔子序列的向导RNA,分子可以与靶RNA结合或附接至靶RNA上。在一些实例中,这样的方法用于捕获靶RNA(例如,免疫沉淀)。这可以用作试剂盒的部分,以鉴定与特异性RNA转录物相互作用的蛋白质。在一实例中,可以将表位标记的Cas13d(例如FLAG、HA、生物素或HALO标签)靶向特定的靶RNA,并经由固定化交联(例如,用多聚甲醛或戊二醛)。Cas13d与表位识别抗体的免疫沉淀允许经由蛋白质印迹或质谱鉴定出共免疫沉淀的蛋白质。

[0518] 在另一实例中,Cas13d可以与生物素连接酶(例如BirA)或过氧化物酶(例如APEX2)融合,以便使与靶RNA相互作用的蛋白质生物素化。然后可以用链霉亲和素珠将标记的蛋白质拉下(pulled down),然后进行质谱或蛋白质印迹。

[0519] 在一些实例中,生物素化的Cas13d可以靶向具有gRNA的核糖体RNA序列。链霉亲和素珠介导的pull-down可以用于耗尽rRNA,以用于RNA测序文库的制备。

[0520] 在一实例中,靶向靶RNA允许掩蔽靶RNA。例如,通过使用具有突变或完整HEPN结构域的Cas13d蛋白质和含有至少一个对靶RNA特异的间隔子序列的向导RNA,可以从RNA-结合蛋白质或RNA-结合元件例如miRNA中掩蔽靶RNA。

[0521] 在一些实例中,Cas13d可以用于掩蔽RNA结合蛋白质(RBP)的RNA结合位点。在另一实例中,Cas13d可以掩蔽miRNA结合位点。例如,肝特异性miR-122与丙型肝炎病毒RNA形成复合物,这保护其免于降解。HEPN活性Cas13d蛋白质可以靶向病毒RNA上的miRNA-122结合位点,以通过同时逆转miRNA-122介导的保护和直接降解HCV RNA来协同对抗HCV感染。在一些实例中,这样的方法用于保存或保护靶RNA分子,例如,保护靶RNA不被降解。例如,通过靶向靶基因的3'UTR中富含AU的元件,HEPN突变的Cas13d可以阻断导致靶转录物降解的RNA结合蛋白质例如三四增蛋白(TTR)或AUF1的结合。

[0522] 在一实例中,靶向靶RNA允许改变靶RNA的剪接。剪接受体和/或供体位点的直接结

合以及剪接效应子结构域均可以用于操纵剪接。例如,通过使用具有突变HEPN结构域(例如SEQ ID NO:2或4)的Cas13d蛋白质,含有至少一个对靶RNA特异的间隔子序列的向导RNA,和任选地,影响剪接的效应子结构域(例如SRSF1的RS-富含结构域、hnRNP A1的Gly-富含结构域、RBM4的丙氨酸富含基序或DAZAP1的脯氨酸富含基序),可以实现RNA的可变剪接。

[0523] 在一些实例中,这样的方法用于外显子包涵,例如,包含酸性 α -葡糖苷酶(GAA)的外显子2,以治疗庞贝氏症(Pompe disease),或者包含SMN2的外显子7,以治疗脊髓性肌萎缩(SMA)。在一些实例中,这样的方法用于外显子排斥,例如恢复肌营养不良蛋白的阅读框,以治疗杜氏肌营养不良,或者将Bcl-x pre-mRNA的剪接从抗凋亡长同种型转变为凋亡前短同种型,以治疗癌症。

[0524] 在一些实例中,该方法使用具有突变HEPN结构域的Cas13d蛋白质,以掩蔽剪接受体或供体位点,例如,产生新生抗原(neoantigens)使冷肿瘤变热。通过影响某些靶pre-mRNA的剪接,该方法可以产生新的外显子-外显子连接,这可以导致在癌细胞中产生新生表位(neo-epitopes)。由于非天然抗原的展示,这可以使癌细胞易受免疫系统攻击。在其它实例中,该方法可以用于动态地操纵同种型比率或恢复蛋白质(例如,杜氏肌营养不良症的肌营养不良蛋白)的阅读框。

[0525] 在一实例中,靶向靶RNA允许控制靶RNA的转录物运输。例如,通过使用具有突变HEPN结构域(例如SEQ ID NO:2或4)和亚细胞定位信号或输出序列的Cas13d融合蛋白、含有至少一个对靶RNA特异的间隔子序列的向导RNA。在一些实例中,这样的方法用于将靶RNA分子运输至特定的细胞器或胞质区室,或者甚至将靶RNA转录物运输至例如核内体以进行细胞外释放。

[0526] 在另一个实例中,该方法可以影响RNA运输。例如,代码结合蛋白质(zipcode binding protein)ZBP1特异性地识别RNA序列5'-CGGAC(C/A-CA-C/U),其导致某些转录物定位到成纤维细胞的前沿。通过掩蔽或操纵特定RNA代码或调节序列以免被调节蛋白质复合物识别,该方法可影响细胞内的RNA定位或运输。

[0527] 在一实例中,靶RNA是核定位的RNA。例如,通过使用与核定位信号融合的具有非突变HEPN结构域(例如SEQ ID NO:1、3、42、62、70、82、83和92)的Cas13d蛋白质和含有至少一个对靶RNA特异的间隔子序列的向导RNA,可以靶向并降解核定位的RNA。在一些实例中,这样的方法可以用于降解靶核RNA分子,例如,以敲低非编码核RNA例如HOTAIR(其与乳腺癌的转移性进展相关)。

[0528] 在一实例中,靶RNA是病毒RNA或DNA病毒的转录物。例如,可以使用具有非突变HEPN结构域(例如SEQ ID NO:1、3、42、62、70、82、83和92)的Cas13d蛋白质和含有至少一个对靶RNA特异的间隔子序列的向导RNA。在一些实例中,例如,通过切割RNA病毒或DNA病毒的转录物,这样的方法用于治疗RNA病毒感染(例如正链RNA病毒感染,例如小核糖核酸病毒(例如,蹄疫病毒科[例如,口蹄疫病毒(FMDV)])、心病毒科;肠道病毒科(如柯萨奇病毒、埃可病毒、肠道病毒和脊髓灰质炎病毒);根病毒科(鼻病毒));肝病毒科(甲型肝炎病毒);囊膜病毒(其实例包括风疹;甲病毒(例如西方马脑炎病毒、东方马脑炎病毒和委内瑞拉马脑炎病毒));黄病毒(其实例包括登革热病毒、西尼罗河病毒和日本脑炎病毒);杯状病毒科(包括诺如病毒和札幌病毒);或冠状病毒(其实例包括SARS冠状病毒,如Urbani株);或负链RNA病毒感染,如正粘病毒(如流感病毒)、弹状病毒(如狂犬病病毒)和副粘病毒(其实例包

括麻疹病毒、呼吸道合胞病毒或副流感病毒);或DNA病毒感染(如疱疹病毒(如水痘-带状疱疹病毒,如Oka株;巨细胞病毒;和单纯疱疹病毒(HSV)1型和2型)、腺病毒(如腺病毒1型和腺病毒41型)、痘病毒(如牛痘病毒)或细小病毒(如细小病毒B19)),。因此,这些方法可以用作基于RNA的抗病毒剂或抗微生物剂。

[0529] 实施例1

[0530] 材料和方法

[0531] 该实施例说明用于获得实施例2-7中所示结果的材料和方法。

[0532] 人胚胎肾(HEK)细胞系293FT的细胞培养

[0533] 在补充有10%FBS(GE Life Sciences)和10mM HEPES的DMEM(4.5g/L葡萄糖)中,在37°C、5%CO₂条件下维持人胚胎肾(HEK)细胞系293FT(Thermo Fisher)。达到80-90%汇合后,使用TrypLE Express(Life Technologies)解离细胞,并以1:2的比率传代。

[0534] 人骨肉瘤上皮细胞系U2OS的细胞培养

[0535] 在补充有10%FBS和10mM HEPES的DMEM(4.5g/L葡萄糖)中,在37°C、5%CO₂条件下维持人骨肉瘤上皮U2OS。达到70%汇合后,将细胞以1:3的比率传代。该细胞系未经鉴定。

[0536] 诱导多能干细胞的维持和神经元分化

[0537] 从Fen-Biao Gao((Biswas et al.,2016)实验室获得含有FTDP-17IVS10+16突变的稳定的人iPSC系或者年龄-和性别-匹配的对照细胞系。简言之,将从一位具有MAPT IVS10+16突变的男性患者获得的细胞和从一位男性对照患者获得的两个单独的细胞系重新编程为hiPSC(Almeida et al.,2012)。将iPSC用含有多西环素(doxycycline)诱导型Ngn2盒的慢病毒转导。慢病毒质粒由S.Schafer和F.Gage赠予。然后将iPSC用Accutase传代,并以500,000个细胞/孔铺板到基质胶包被的6孔板中,所述6孔板包括含ROCK抑制剂Y-27632(10μM,Cayman)的mTESR培养基。在第1天,将培养基用mTESR更换。在第2天,将培养基换成含有多西环素(2μg/ml,Sigma)的mTESR,以诱导Ngn2表达。在第3天,将培养基用神经诱导培养基(NIM,DMEM/F12(Life Technologies),含有BSA(0.1mg/ml,Sigma)、脱铁运铁蛋白(apo-transferrin)(0.1mg/ml,Sigma)、腐胺(16μg/ml,Sigma)、孕酮(0.0625μg/ml,Sigma)、亚硒酸钠(0.0104μg/ml,Sigma)、胰岛素(5μg/ml,Roche)、BDNF(10ng/ml,Peprtech)、SB431542(10μM,Cayman)、LDN-193189(0.1μM,Sigma)、层粘连蛋白(2μg/ml,Life Technologies)、多西环素(2μg/ml,Sigma)和嘌呤霉素(Life Technologies))更换。每天更换NIM培养基。3天的嘌呤霉素选择之后,将未成熟神经元细胞用Accumax(Innovative Cell Technologies)传代,并将其铺板到96孔板上,所述96孔板用神经成熟培养基(NMM;1:1Neurobasal/DMEM(Life Technologies),其含有B27(Life Technologies)、BDNF(10ng/ml,Peprtech)、N-乙酰半胱氨酸(Sigma)、层粘连蛋白(2μg/ml,Life Technologies)、dbcAMP(49μg/ml,Sigma)和多西环素(2μg/ml,Sigma))中的聚-D-赖氨酸和基质胶包被。次日(第7日)将培养基用含有AraC(2μg/ml,Sigma)的NMM培养基更换,以除去任何剩余的未分化细胞。在第8天,除去AraC,并将星形胶质细胞接种在神经元的顶部,以支持含有hEGF(5ng/ml,Peprtech)的NMM中的神经元培养。在第10天将细胞用AAV转导,在第24天进行检验。

[0538] Cas13d鉴定的计算流程

[0539] 2016年6月,我们从NCBI Genome获得全基因组、染色体和支架水平的原核基因组

组装体,并比较了CRISPRfinder、PILER-CR和CRT以鉴定CRISPR重复序列。提取每个推定CRISPR重复序列周围的20千碱基侧翼区,以使用Python鉴定邻近蛋白质和预测的蛋白质。要求候选Cas蛋白质的长度>750aa,并在重复阵列的5个蛋白质内,如果它们含有与已知CRISPR系统如I-III型CRISPR相关的Cas基因,则将提取的CRISPR基因座滤出。通过all-by-all BLASTp分析,然后是单链接层次聚类(single-linkage hierarchical clustering),将推定的效应子聚类成家族,其中聚类分配需要至少60的比特得分(bit score)。针对NCBI非冗余(nr)蛋白质数据库对至少2种蛋白质的每个簇进行BLAST搜索,要求比特得分>200,以指定相似性。分析了每个扩增簇中同源性蛋白质与CRISPR阵列的同现率(co-occurrence),要求其>70%。通过平均氨基酸长度来分选蛋白质家族,并使用Clustal Omega和带有Blosum62成本矩阵(cost matrix)的Geneious aligner,对每个簇进行多序列比对。基于该比对,在Cas13d家族中鉴定出RxxxxH HEPN基序。对所有预测的Cas13d效应子针对公共的宏基因组全基因组鸟枪(shotgun)序列(不含预测的开放读框(ORF))进行TBLASTN。通过对基因组和宏基因组数据库每月进行BLAST搜索,定期更新Cas13d家族,以鉴定任何新保藏的序列。使用Clustal Omega比对新的全长同源物和同源性片段,并使用PhyML 3.2进行聚类。CRISPrDetect用于预测Cas13d阵列中同向重复序列的方向,并使用Andronoescu2007RNA能量模型在37℃下进行DR折叠预测。使用Geneious10生成Cas13d同向重复序列的序列图标。

[0540] 蛋白质表达和纯化

[0541] 从培养的分离物或宏基因组样品的基因组DNA提取物PCR扩增重组Cas13d蛋白质,并克隆到具有N-端His-MBP融合物和TEV蛋白酶切割位点的基于pET的载体中。将产生的质粒转化到Rosetta2(DE3)细胞中(Novagen),在OD₆₀₀ 0.5下用200μM IPTG诱导,并在18℃下生长20小时。然后将细胞沉淀,冻融,并重悬于补充有1X蛋白酶抑制剂片剂、1mg/mL溶菌酶、2.5U/mL Turbo DNA酶(Life Technologies)和2.5U/mL盐活性核酸酶(Sigma Aldrich)的裂解缓冲液(50mM HEPES,500mM NaCl,2mM MgCl₂,20mM咪唑,1%v/v Triton X-100,1mM DTT)中。然后将裂解的样品超声,并经由离心澄清(18,000x g,1小时,4℃下),用0.45μm PVDF滤器过滤,并与50mL Ni-NTA Superflow树脂(Qiagen)/10L原始细菌培养物一起孵育1小时。将珠-裂解物混合物施加于色谱柱,用5个柱体积的裂解缓冲液和3个柱体积的洗脱缓冲液(50mM HEPES,500mM NaCl,300mM咪唑,0.01%v/v Triton X-100,10%甘油,1mM DTT)洗涤。然后将样品在TEV切割缓冲液(50mM Tris-HCl,250mM KCl,7.5%v/v甘油,0.2mM TCEP,0.8mM DTT,TEV蛋白酶)中透析过夜,然后进行阳离子交换(HiTrap SP,GE Life Sciences)和凝胶过滤(Superdex 200 16/600,GE Life Sciences)。将洗脱的纯化蛋白质级分合并,并以4mg/mL冷冻于蛋白质存储缓冲液(50mM Tris-HCl,1M NaCl,10%甘油,2mM DTT)中。

[0542] 向导和靶RNA的制备

[0543] 合成携带T7启动子和适当的下游序列的寡核苷酸(IDT),用crRNA的反义T7寡核苷酸退火,并PCR扩增靶和阵列模板。通过Synthego合成均聚物靶RNA。用Histarh T7高产RNA合成试剂盒(New England Biolabs)在31℃下体外转录寡核苷酸退火和PCR模板12小时。对于标记的靶,额外地加入2mM荧光标记的氨基烯丙基-UTP atto 680(Jena Biosciences)。将向导RNA用RNA级Agencourt AMPure XP珠(Beckman Coulter)纯化,并将

阵列和靶用MEGAclear转录清除试剂盒(Thermo Fisher)纯化,并在-80℃下冷冻。对于ssDNA和dsDNA靶,合成相应的寡核苷酸序列(IDT),并凝胶纯化,或PCR扩增,然后分别进行凝胶纯化。

[0544] 生物化学切割反应

[0545] 将纯化的EsCas13d蛋白质和向导RNA在RNA切割缓冲液(25mM Tris pH 7.5,15mM Tris pH 7.0,1mM DTT,6mM MgCl₂)中以2:1的摩尔比混合(除非另外指出)。反应液在冰上制备,并在37℃下孵育15分钟,然后以相对于EsCas13d为1:2的摩尔比加入靶。随后将反应液在37℃下孵育45分钟,并用1μL的酶终止溶液(10mg/mL蛋白酶K,4M尿素,80mM EDTA,20mM Tris pH 8.0)在37℃下退火15分钟。然后将反应液用2X RNA负载缓冲液(2X:13mM Ficoll,8M尿素,25mM EDTA)在85℃下变性10分钟,并在10%TBE-尿素凝胶(Life Technologies)上分离。含有标记靶的凝胶在Odyssey成像仪上(Li-Cor)可视化;未标记的阵列或靶切割凝胶用SYBR Gold染色,然后经由Gel Doc EZ系统(Bio-Rad)成像。

[0546] 人细胞系的瞬时转染

[0547] 将改造的Cas13编码序列克隆到含有EF1a启动子的标准化质粒表达骨架中,并使用Nucleobond xtra Midi EF试剂盒(Machery Nagel)根据制造商的方案制备。分别从Feng Zhang赠予的Addgene#103854和#103862PCR扩增NLS-LwaCas13a-msfGFP和PspCas13b-NES-HIV。将Cas13d pre-gRNA和gRNA克隆到含有U6启动子的最小骨架中。将针对LwaCas13a的shRNA和向导克隆到相同的骨架中,并且位置与其在靶序列的3'处的对应的向导RNA匹配。将PspCas13b的匹配的gRNA移至最接近的5'-G核苷酸。

[0548] 对于瞬时转染,将HEK 293FT细胞以20,000细胞/孔的密度铺板到96-孔板中,并根据制造商的方案使用Lipofectamine2000(Life Technologies),以200ng Cas13表达质粒和200ng gRNA表达质粒在>90%汇合下进行转染。转染后48-72小时收获转染的细胞用于流式细胞术、基因表达分析或其它下游加工。

[0549] 对于报告检验,使用Lipofectamine 2000(Life Technologies)将HEK293 FT细胞用192ng Cas13d表达质粒、192ng向导表达质粒和12ng mCherry表达质粒以96孔格式转染。48小时后收获细胞,并通过流式细胞术分析。

[0550] 将U2OS细胞以20,000细胞/孔的密度铺板在96孔板中,并根据制造商的方案,使用Lipofectamine 3000(Life Technologies),以100ng Cas13d表达质粒在>90%汇合下进行转染,并在48小时后进行免疫细胞化学处理。

[0551] 流式细胞术

[0552] 转染后48小时将细胞用TrypLE Express解离,重悬于FACS缓冲液(1X DPBS-/-,0.2%BSA,2mM EDTA)中。使用MACSQuant VYB(Miltenyi Biotec)在96孔板格式下进行流式细胞术,并使用FlowJo10进行分析。RG6由Thomas Cooper(Addgene质粒#80167)赠予,将其加以修饰以用mTagBFP2置换EGFP。所有呈现的样品均以3次生物学重复进行检验。在mCherry报告检验中,数据代表每种条件至少20,000个门控事件。在剪接报告检验中,数据代表每种条件至少2,500个门控事件。

[0553] 基因表达分析

[0554] 转染后48小时将细胞用补充DTT的RLT缓冲液裂解,并使用RNeasy Mini Plus柱(Qiagen)提取总RNA。然后使用随机六聚体引物和Revertad逆转录酶(Thermo Fisher)将

200ng总RNA在25℃下逆转录10分钟,在37℃下逆转录60分钟,并在95℃下逆转录5分钟,随后使用2X Taqman Fast Advanced Master Mix(Life Technologies)和Taqman探针(用于GAPDH和适当的目标基因)(Life Technologies and IDT))进行qPCR。通常选择Taqman探针和引物组,以扩增跨Cas13或shRNA靶位点位置的cDNA,以防止检测到切割的转录物片段(参见Konermann et al.,Cell 173:1-12,2018的表S4,其通过引用整体并入本文)。qPCR在5 μ L多通路反应中使用LightCycler480仪器II(Roche)以384孔格式进行。使用ddCt方法计算相对于GFP转染的载体对照的倍数变化。使用Prism7,使用具有多重比较校正的单向或双向ANOVA来评估转录物变化的统计学显著性。

[0555] 免疫组织化学

[0556] 对于免疫组织化学分析,在96孔光学透明板(Greiner Bio-One)上培养U2OS细胞,如前所述进行转染,然后在稀释于PBS(Gibco)中的4%PFA(Electron Microscopy Sciences)中固定,并用PBS中的0.3M甘氨酸(Sigma)洗涤。将样品在含有8%驴血清(Jackson ImmunoResearch)、8%山羊血清(Cell Signaling Technologies)和0.3%Triton-X100(Sigma)的PBS溶液中封闭并透化一小时,然后在1%BSA(Fisher Bioreagents)、1%山羊血清和0.25%Triton-X中在4℃下一抗孵育过夜。将样品用含有0.1%BSA和0.1%Triton-X100的PBS洗涤3次,然后在具有0.05%Triton-X 100和1%BSA的PBS中与荧光团偶联的第二抗体一起在室温下孵育一小时。将细胞用含0.1%Triton-X的PBS洗涤,用DAPI染色,然后用Mounting Media(Ibidi)覆盖,之后进行成像。第一抗体HA-Tag 6E2(Cell Signaling,2367)按照生产商的说明以1:100的稀释度使用。所用的第二抗体是山羊抗小鼠IgG1-Alexa-Fluor 647(Thermo Fisher,A21240)和抗小鼠IgG1 CF633(Sigma,SAB4600335)。使用Zeiss Airyscan LSM880取得共焦图像,然后在Zen2.3(Zeiss)中进行图像处理。

[0557] 细菌小RNA测序和分析

[0558] 将大肠杆菌DH5 α 细胞用携带来源于未培养的瘤胃球菌属株的CRISPR-Cas13d基因座的pACYC184转化。在固定相中收获细胞,并将其在PBS中漂洗,重悬于TRIzol(Life Technologies)中,转移到含有0.1mm二氧化硅珠(MP Biomedicals)的裂解基质B管中,并在珠磨机24(Fisher Scientific)上匀浆三个30秒的循环。通过苯酚-氯仿提取分离总RNA,然后使用DirectZol Miniprep试剂盒(Zymo Research)纯化。RNA质量在Agilent2200Tapestation上评价,然后进行Turbo DNA酶处理。将总RNA用T4多核苷酸激酶(NEB)处理,并使用用于细菌的Ribo-Zero rRNA去除试剂盒(Illumina)进行rRNA-耗尽。将RNA用RNA 5'多磷酸酯酶处理,用大肠杆菌poly(A)聚合酶进行poly(A)加尾,并使用T4 RNA连接酶1(NEB)与5'RNA测序接头连接。使用寡-dT引物和M-MLV RT/RNase Block(AffinityScript,Agilent)通过逆转录产生cDNA,然后进行PCR扩增和条形码化(barcoding)。将得到的文库在Illumina MiSeq上测序,使用定制的Python脚本进行解多路复用,并使用Bowtie2与Cas13d CRISPR基因座比对。比对用Geneious可视化。

[0559] Ngn2慢病毒制备

[0560] 铺板期间,在DMEM+10%FBS培养基中用Polyethylenimine Max(PEI, Polysciences)和Ngn2目标质粒加pMDG.2和pSPAX2包装质粒(Didier Trono赠予,Addgene#12259和#12260)转染低传代HEK 293FT细胞。第二天,将培养基更换为无血清的化学成分确

定的基本培养基(补充有Glutamax的Ultraculture, Lonza)。48小时后收获病毒上清液, 通过0.45微米PVDF滤器(Millipore)澄清, 并使用超速离心浓缩。

[0561] AAV制备

[0562] 铺板期间, 在DMEM+10%FBS培养基中用Polyethylenimine Max (PEI, Polysciences) 和AAV目标质粒加AAV1血清型和pAdDeltaF6辅助包装质粒 (UPenn Vector Core) 转染低传代HEK 293FT细胞。第二天, 将60%的培养基更换为化学成分确定的基本培养基(补充有Glutamax的Ultraculture, Lonza)。48小时后, 收获含有AAV的上清液, 通过0.45微米PVDF滤器(Millipore)澄清, 并按照制造商的方案使用通过聚乙二醇(PEG病毒沉淀试剂盒#K904, Biovision)的沉淀进行浓缩。

[0563] RNA-seq文库制备和测序

[0564] 转染后48小时, 使用Qiagen的RNeasy Plus Mini试剂盒从293FT细胞中提取总RNA。使用New England Biolabs的NEBNext II超定向RNA文库制备试剂盒(Cat#E7760S)制备链mRNA文库, 并在具有42nt配对末端读数的Illumina NextSeq500上测序。每个条件对~15M的总读数进行反多路复用(demultiplexed)。

[0565] RNA-seq分析

[0566] 使用FASTQC对测序的读数进行质量检测, 并使用2.5.1b STAR比对器(Dobin et al., 2013)与hg19人类基因组进行比对。使用默认参数进行作图(每个读数至多10个错配, 并且每个读数至多9个多-作图(multi-mapping)位置)。使用提供有hg19 Illumina iGenomes collection(Illumina)的基因注释和100的sjdbOverhang值构建基因组指数(genome index)。使用最高表达的同种型作为基因表达的代表物(proxy), 用HOMER分析组件(HOMER analysis suite)(Heinz et al., 2010)对所有基因外显子的唯一匹配读数(uniquely mapped reads)进行定量, 并且使用一式三份的DESeq2v1.14.1(Love et al., 2014)进行差异基因表达, 以计算组内分散性和对比度, 以在靶向和非靶向条件之间进行比较。显著差异表达的基因被定义为具有<0.01的假发现率(FDR)和>0.75的log2倍数变化。使用包括的绘图库和来自标度0.5.0包(scales 0.5.0package)的alpha()颜色函数在R3.3.2中生成火山图(Volcano plots)。

[0567] 统计学

[0568] 所有的值均报告为平均值±SD或平均值±SEM, 如在适当的图例中所示。为对两组进行比较, 使用单尾学生t-检验(one-tailed student's t-test), 并使用Holm-Sidak方法确定统计学显著性, $\alpha=0.05$ 。使用具有Tukey多重假设校正的单向ANOVA来评估多于两组之间的显著性。当比较两个因素(即RNA靶向形式和向导位置)时使用双向ANOVA, 并通过Sidak多重比较检验(Sidak's multiple comparisons test)进行多重假设校正调整。为了比较通过达戈斯蒂诺和皮尔逊正态性检验(D'Agostino and Pearson normality test)发现不满足正态分布假设的组, 进行具有邓恩多重比较调整(Dunn's multiple comparison adjustment)的非参数弗里德曼检验(Friedman test)。PRISM7.0用于所有的统计学分析。样品大小不是先验性确定的。如在每个图中具体地示出, 每个实验使用至少三次生物学重复。

[0569] 本文报道的测序数据可以在NCBI基因表达Omnibus(Gene Expression Omnibus)中以GEO系列登录号GSE108519找到。

[0570] 关于所用材料和方法的其它细节,例如序列,可以参阅Koner mann et al., Cell 173:1-12, 2018, 其通过引用整体并入本文。

[0571] 实施例2

[0572] VI型样Cas核糖核酸酶家族的计算鉴定

[0573] 本实施例描述了通过开发用于2类CRISPR-Cas基因座的计算流程用于鉴定之前未检测或未表征的靶向RNA的CRISPR-Cas系统的方法,其仅要求用于CRISPR干扰的单一核酸酶,例如Cas9、Cas12a (旧称Cpf1) 或Cas13a (旧称C2C2) (Makarova et al., 2015; Shmakov et al., 2015)。为了改进之前的CRISPR系统生物信息学挖掘策略(其专注于发现间隔子获得中涉及的保守Cas基因的组(Shmakov et al., 2015), 定义了CRISPR基因座的最低要求是存在CRISPR重复序列阵列和邻近效应子核酸酶。使用CRISPR阵列作为搜索锚(search anchor), 所有的原核基因组组装体和支架均获自于NCBI WGS数据库和适用于从头CRISPR阵列检测的算法(Bland et al., 2007; Edgar, 2007; Grissa et al., 2007), 以鉴定21,175个推定的CRISPR重复阵列(图1A)。

[0574] 提取位于每个CRISPR阵列侧翼的多达20千碱基(kb)的基因组DNA序列,以鉴定紧邻的预测蛋白质编码基因。除了Cas12a和Cas13a来判断流程检测和聚类这些已知的2类效应子家族的能力之外,将含有已知的1类和2类CRISPR-Cas系统的签名基因如Cas3或Cas9的候选基因座排除在进一步分析之外。由于与CRISPR重复序列紧密相关的大蛋白质是已知单一效应子的关键特征,为了鉴定新的2类Cas效应子,要求候选蛋白质的长度>750个残基,并且在重复序列阵列的5个蛋白质编码基因内。使用基于同源性的单链接层次聚类将所得蛋白质分类成408个推定蛋白质家族。

[0575] 为了丢弃由于基因组中的偶然或总体丰度而与CRISPR阵列紧密邻近的蛋白质聚类,从NCBI非冗余蛋白质数据库中鉴定每个聚类的其他同源性蛋白质,并确定其与CRISPR阵列的邻近性。推论真实Cas基因将与CRISPR重复序列具有高共现率,则要求每个扩大聚类>70%的蛋白质存在于20kb的CRISPR重复序列内。分析这些其余的蛋白质家族的核酸酶结构域和基序。

[0576] 在包括最近描述的Cas13b系统(Smargon et al., 2017)的候选物中,鉴定出编码含有2个预测的HEPN核糖核酸酶基序的候选CRISPR相关核糖核酸酶的未表征的推定2类CRISPR-Cas系统家族(Anantharaman et al., 2013) (图2A)。重要的是,它们是迄今为止描述的最小的2类CRISPR效应子(~930aa)。VI型CRISPR-Cas13超家族的实例是序列差异的单一效应子签名核酸酶以及存在两个HEPN结构域。除了这两个RxxxxH HEPN基序(图3A)之外,候选效应子与之前描述的Cas13酶没有显著的序列相似性,因此其推定CRISPR核糖核酸酶的家族被称为VI型Cas13d或VI-D型(图3B)。

[0577] CRISPR-Cas13d系统来源于肠驻留的微生物,因此我们试图通过与来自最近大规模微生物组测序工作中的宏基因组重叠群比对来扩大Cas13d家族。Cas13d蛋白质与没有预测开放阅读框(ORF)的公开宏基因组序列的比较鉴定出另外的全长系统以及在几个不同的分支中聚类的多个效应子和阵列片段(图1B)。为了从Cas13d蛋白质家族的不同分支产生全长Cas13d直向同源蛋白质和基因座,从相关的组装体获得基因组DNA样品we43,并进行靶向桑格测序(targeted Sanger sequencing),以填补由于测序覆盖不完全而导致的缺口,例如对于宏基因组直向同源物‘厌氧消化池宏基因组’(Adm) (Treu et al., 2016)。

[0578] Cas13d CRISPR基因座主要聚集在瘤胃球菌属的良性革兰氏阳性肠道细菌中,并表现出令人惊讶的CRISPR基因座结构的多样性(图2A)。除了宏基因组AdmCas13d系统以外,Cas13d系统在其CRISPR基因座内缺乏关键的间隔子获得蛋白质Cas1(Yosef et al., 2012),突出了没有Cas1或Cas2基因要求的2类CRISPR发现流程的实用性。Cas13d同向重复序列(DR)在长度和预测二级结构方面是高度保守的(图3C),其长度为36nt,具有带A/U-富含环的8-10nt的茎,以及在同向重复序列3'端的5'-AAAAC基序(图3D)。已经显示这一保守的5'-AAAAC基序被II型Cas1/2间隔子获得复合物特异性识别(Wright and Doudna, 2016)。事实上,可以发现Cas1相对接近于一些Cas13d系统(对于P1E0和Rfx,在10-30kb内),而其余含有Cas13d的细菌在其基因组的其它地方含有Cas1,可能作为另一个CRISPR基因座的部分。

[0579] 实施例3

[0580] CRISPR-Cas13d具有双重RNA酶活性

[0581] 为了证明Cas13d重复阵列被转录并被加工成CRISPR向导RNA(gRNA),将Cas13d CRISPR基因座从未培养的瘤胃球菌属样品(Ur)克隆到细菌表达质粒中。CRISPR系统倾向于形成具有独立表达所必需的调节序列的自足式操纵子(self-contained operons),从而促进在大肠杆菌中的异源性表达(Gasiunas et al., 2012)。RNA测序(Heidrich et al., 2015)揭示了将阵列加工成~52nt成熟gRNA,具有30nt 5'同向重复序列,后接长度范围为14-26nt的可变3'间隔子(图2B)。

[0582] 为了在体外表征Cas13d的性质,基于其在大肠杆菌中的稳健重组表达(图4A-4C)纯化Eubacterium sircum Cas13d蛋白质(EsCas13d),并发现EsCas13d仅足以将其匹配的CRISPR阵列加工成组成型向导,而无需另外的辅助核糖核酸酶(图2C, Konermann et al., "Transcriptome Engineering with RNA-Targeting Type VI-D CRISPR Effectors," Cell 173:1-12, 2018的Konermann等人的表S1,其通过引用整体并入本文),这是一些2类CRISPR-Cas系统共有的性质(East-Seletsky et al., 2016; Fonfara et al., 2016; Smargon et al., 2017)。而且,使HEPN基序的带正电荷的催化残基(Anantharaman et al., 2013)失活(dCas13d:R295A, H300A, R849A, H854A)不影响阵列加工,表明决定gRNA生物发生的独特RNase活性类似于Cas13A(East-Seletsky et al., 2016; Liu et al., 2017)。

[0583] Cas效应子蛋白通常与成熟gRNA形成二元复合物,以产生RNA向导的监视(surveillance)核糖核蛋白,其能够切割外来核酸以免疫防御((van der Oost et al., 2014)。为了评估Cas13d是否具有可编程的RNA靶向活性(如两个HEPN基序的存在所表明的),将EsCas13d蛋白质与阵列或成熟gRNA以及同源的体外转录靶配对。基于RNA测序结果,选择含有30nt同向重复序列和22nt的中间间隔子长度的成熟gRNA(SEQ ID NO:129的核苷酸6-36,后接与RNA靶互补的22个碱基)。

[0584] Cas13d能够以依赖于向导序列的方式用未加工的阵列和成熟gRNA有效地切割互补的靶ssRNA,而不匹配的间隔子序列则消除Cas13d活性(图5A)。用dCas13d置换或在切割反应中加入EDTA也消除了依赖于向导的RNA靶向,表明Cas13d靶向是依赖于HEPN和Mg²⁺的(图5B)。为了确定用于有效的Cas13d靶向的最小间隔子长度,产生了范围从未加工的30nt长度至10nt的一系列间隔子截短物(图6A)。在21nt间隔子长度以下,切割活性显著降低,证实了22nt间隔子的选择(图6B)。

[0585] 已经提出RNA靶向的2类CRISPR系统起到外来RNA传感器的作用(Abudayyeh et al., 2016; East-Seletsky et al., 2016), 其中效应子核酸酶的一般RNA酶活性由向导匹配的靶触发。为了检验Cas13d的相似特性, 在匹配RNA靶的存在下监测二元EsCas13d: gRNA复合物的RNase活性。经观察, EsCas13d可以被靶RNA激活, 以切割旁观者RNA靶(图3C), 尽管相对于其对互补ssRNA靶标的活性而言效率较低。旁观者切割(bystander cleavage)是向导序列和HEPN依赖性的, 因为单独存在非匹配的旁观者靶标并不足以诱导切割, 而dCas13d的置换或加入EDTA则消除了活性。这些结果表明, 旁观者RNase活性可能是CRISPR获得性细菌免疫中RNA靶向2类系统的一般性质(图3D)。

[0586] 为了评价Cas13d再编程的普遍性, 生成12种拼接互补性RNA靶的向导, 并在所有情形中观察到有效切割(图7A)。Cas13D不能切割ssDNA(图6C)或dsDNA(图6D)形式的ssRNA靶, 表明Cas13d是RNA特异性核酸酶。而且, 与要求3'-H(Abudayyeh et al., 2016)或双侧、DR-近端5'-D和3'-NAN或NNA(Smargon et al., 2017)的其它RNA靶向2类系统相比, RNA靶向切割不依赖于前间隔子(proto-spacer)侧翼序列(PFS)(图7A)。尽管最初观察到相对于腺嘌呤PFS的轻微偏差(图6E), 但用恒定向导序列改变靶PFS碱基没有导致靶向效率上的显著差异($P=0.768$)(图6F)。

[0587] 尽管靶向DNA的2类CRISPR系统(Gasiunas et al., 2012; Jinek et al., 2012; Zetsche et al., 2015)和一些靶向RNA的1类系统倾向于在相对于靶-向导双链体的限定位置处切割(Samai et al., 2015; Zhang et al., 2016), Cas13d切割模式相对于不同的靶而变化(图5A, 5C, 6H), 并无论向导序列位置而保持显著相似(图7A)。这表明Cas13d可以优先地切割靶RNA中的特异性序列或结构上可接近的区域。测试Cas13d对在发夹的环区域中含有可变均聚物重复序列或作为线性单链重复序列的靶的活性。EsCas13d在两种靶结构中均表现出对尿嘧啶碱基的显著优先性, 在所有其它碱基处具有较低但可检测的活性(图7B)。

[0588] Cas酶几乎在所有的古细菌和大约半数细菌中均有发现(Hsu et al., 2014; van der Oost et al., 2014), 其跨越了很宽范围的环境温度。为了确定Cas13d活性的最优温度范围, 测试16-62°C的裂解温度条件图谱, 并在24-41°C范围内观察最大活性(图6G, 6H)。该温度范围与广泛的原核和真核宿主相容, 表明Cas13d可以适用于不同细胞和生物体中的RNA靶向。

[0589] 实施例4

[0590] 改造的直向同源物的基于细胞的活性筛选

[0591] 使用Cas13d核酸酶作为哺乳动物细胞中可编程RNA靶向的灵活工具。来自不同细菌物种的CRISPR直向同源物通常表现出可变的活性(Abudayyeh et al., 2017; East-Seletsky et al., 2017), 尤其是在人细胞中异源性表达后((Ran et al., 2015; Zetsche et al., 2017)。在基于真核细胞的mCherry报告筛选中鉴定出高活性Cas13d直向同源物。

[0592] 通过合成来自Cas13d家族内的不同分支的7种直向同源物的人密码子优化形式(图1B), 生成携带催化活性和HEPN-失活蛋白质的哺乳动物表达质粒。然后任选地将每种蛋白质与N-和C-端核定位信号(NLS)融合。这些Cas13d效应子设计是HA标记的, 并且与两种不同的向导RNA结构配对, 或者具有侧翼为两个同向重复序列的30nt间隔子以模拟未加工的向导RNA(pre-gRNA), 或者具有含预计模拟成熟向导RNA的22nt间隔子的30nt同向重复序列(gRNA)(图8A)。对于每种向导设计, 然后将与mCherry转录物互补的四个不同的间隔子序列

合并,以使靶向效率中潜在的间隔子依赖性变化最小化。在基于人胚肾 (HEK) 293FT细胞的报告检验中测定Cas13d敲低mCherry蛋白质水平的能力。

[0593] 转染后48小时,流式细胞术表明,相对于非靶向对照向导,RfxCas13d和AdmCas13d分别有效地将mCherry蛋白质水平敲低了多达92%和87% ($P < 0.0003$) (图8B)。相反,EsCas13d连同RaCas13d和RffCas13d一起在人细胞中表现出有限的活性。而且,HEPN失活的Rfx-dCas13d构建物中无一显著影响mCherry荧光,表明了HEPN依赖性敲低(对于所有情形 $P > 0.43$)。通过免疫细胞化学观察到Rfx和AdmCas13d NLS融合构建物的强烈核转位,而野生型效应子主要保持在核外(图8C)。

[0594] 以RfxCas13d和AdmCas13d作为先导候选物,我们接下来比较了它们敲低内源性转录物的能力。为了确定最优的直向同源物和向导结构,系统地检验了Rfx和AdmCas13d构建物变体靶向 β -1,4-N-乙酰基-氨基半乳糖基转移酶1 (B4GALNT1) 转录物的能力。在每个条件下,汇集 (pooled) 4种含有不同间隔子序列的向导,其拼接B4GALNT1转录物。比起野生型RfxCas13d和AdmCas13d的两种变体,RfxCas13d-NLS融合物更加有效地靶向B4GALNT1,而gRNA和pre-gRNA均介导有效的敲低($\sim 82\%$, $P < 0.0001$) (图8D)。因此,选择来自黄化瘤胃球菌XPD3002株的Cas13d-NLS进行其余实验(CasRx)。

[0595] 实施例5

[0596] 在人细胞中用CasRx进行可编程RNA敲低

[0597] 由于Cas13d能够处理其自身的CRISPR阵列,因此该性质被用于在简单的单一载体系统中同时递送多个靶向向导(图9A)。编码四种间隔子的阵列始终通过CasRx促进稳固的($>90\%$) RNA敲低($P < 0.0001$) (图9B),所述四种间隔子各自拼接mRNA (B4GALNT1和ANXA4) 或核定位lncRNA (HOTTIP和MALAT1) 的转录物。

[0598] 通过比较CasRx-介导的RNA干扰与dCas9-介导的CRISPR干扰 (Gilbert et al., 2014; Gilbert et al., 2013) 和通过瞬时转染的间隔子序列匹配的shRNA,比较了CasRx与更成熟的用于转录物敲低或抑制的技术(图9C)。对于基于CRISPRi的抑制,分析了来自之前报道的用于B4GALNT1的最有效的dCas9向导 (Gilbert et al., 2014; Zalatan et al., 2015)。在3种内源性转录物中,CasRx在每种情况下均胜过shRNA (11/11) 和CRISPRi (4/4) (图9D),在48小时后表现出96%的中值敲低,与之相比较,shRNA为65%,CRISPR为53%。此外,将CasRx的敲低与两种最近描述的Cas13a和Cas13b效应子进行比较 (Abudayyeh et al., 2017; Cox et al., 2017) (图10A)。在3种基因和8种向导RNA中,CasRx介导了显著大于LwaCas13a-msfGFP-NLS和PspCas13B-NES的转录物敲低(中值:97%,与之比较,分别为80%和66%, $P < 0.0001$) (图10B)。

[0599] RNAi由于接合了简单的再靶向原理、可放大的合成、敲低效力和试剂易于递送而被广泛地用于破坏任何目的基因。然而,可能由于RNAi试剂进入内源性miRNA路径 (Doench et al., 2003; Smith et al., 2017),广泛的脱靶转录物沉默始终是关注的问题 (Jackson et al., 2003; Sigoillot et al., 2012)。与这些报道一致,对用B4GALNT1-靶向shRNA转染的人细胞进行RNA测序时,观察到相对于非靶向shRNA的广泛的脱靶转录变化(>500 显著脱靶变化, $P < 0.01$,图9E,9G)。相反,间隔子匹配的CasRx向导RNA的转录组作图揭示,在靶向的转录物以外,没有显著的脱靶变化(图9F)。这表明体外观察到的中等旁观者切割(图5C)不会导致哺乳动物细胞中可观察到的脱靶转录组扰动(transcriptome perturbation)。当靶

向ANXA4时观察到类似的模式(图11A-11B),shRNA靶向导致900个以上的显著脱靶变化,与之相比,使用CasRx为零(图9G)。

[0600] 为了证实CasRx干扰是广泛适用的,选择了一组11个在癌症、细胞信号传导和表观遗传调节中具有不同作用的其它基因,并筛选每个基因3个向导。CasRx持续地介导高水平的基因间转录物敲低,中值降低为96%(图9H)。每个所测试的向导介导至少80%的敲除,强调了CasRx系统对RNA干扰的一致性。

[0601] 实施例6

[0602] 用dCasRx进行剪接同种型改造

[0603] 用CasRx靶向RNA的实验揭示,靶RNA和蛋白质敲低依赖于HEPN结构域的催化活性(图8B,5B)。当与无催化活性的dCasRx配对时,用CasRx介导有效敲低的相同向导序列则不能显著降低mCherry水平(图8B),表明dCasRx靶向mRNA的编码部分不一定干扰蛋白质翻译。这一观察结果表明了利用dCasRx靶向转录物内的特定编码和非编码元件以研究和操纵RNA的可能性。为了验证这一概念,通过产生剪接效应子扩大dCasRx系统的效用。

[0604] 可变剪接通常通过pre-mRNA中的顺式作用元件与正或负反式作用剪接因子的相互作用来调节,上述反式作用剪接因子可以介导外显子包涵或排斥(Matera and Wang, 2014; Wang et al., 2015)。据推断,与这些基序结合的dCasRx对靶向同种型扰动可能是足够的。为了概念验证,在含有位于mTagBFP2上游的DsRed的双色剪接报告子中,在可变剪接的外显子之后的两个不同的阅读框中鉴定出不同的剪接元件((Orenge et al., 2006) (图12A)。该第二外显子的包涵或排斥切换(toggles)阅读框和产生的荧光,有助于通过流式细胞术定量读出剪接模式。为了介导外显子跳跃,设计了四种向导RNA,以靶向内含子分支点核苷酸、剪接受体位点、推定外显子剪接增强子和外显子2的剪接供体。

[0605] 负剪接因子的一个广泛家族是高度保守的异种核内核糖核蛋白(hnRNP),其通常经由的富含甘氨酸的C-端结构域抑制外显子包涵(Wang et al., 2015)。用dCasRx和与hnRNP Pa1(丰度最高的hnRNP家族成员之一)的富含Gly的C-端结构域的改造融合物来靶向剪接报告子(图12B)。

[0606] 向导位置似乎是改造的外显子跳跃效率的主要决定因素。尽管相对于非靶向的向导,每个向导位置介导了外显子排斥的显著增加(在所有情况下 $P < 0.0001$),但靶向剪接受体导致了最有效的外显子排斥(对于单独的dCasRx,从8%的基础跳跃增加至65%,对于hnRNPa1融合物,增加至75%)。通过比较,dLwaCas13a-msfGFP-NLS在所有四个位置上介导了显著较低水平的外显子跳跃(对于剪接受体向导,19%的跳跃)(图10C和10D, $P < 0.0001$)。

[0607] 比起各个单独的向导,同时用CRISPR阵列靶向所有4个位置实现了更高水平的外显子跳跃(对于dCasRx为81%,对于hnRNPa1融合物为85%,与SA向导相比较, $P < 0.006$) (图12B)。这些结果表明,dCasRx允许通过改变向导位置来调节同种型的比率,并表明它可以在人细胞中用作有效的RNA结合模块以靶向和操纵特异性RNA元件。

[0608] 实施例7

[0609] dCasRx病毒递送至额颞叶痴呆的神经元模型

[0610] 与Cas9(取决于亚型,~1100aa至~1400aa,具有紧凑的异常物(outliers),例如CjCas9或SaCas9)、Cas13a(1250aa)、Cas13b(1150aa)和Cas13c(1120aa)相对比,Cas13d家族的长度平均为930个氨基酸(图3B)(Chylinski et al., 2013; Cox et al., 2017; Hsu et

al.,2014;Kim et al.,2017;Shmakov et al.,2015;Smargon et al.,2017)。尽管腺相关病毒(AAV)由于其范围广泛的衣壳血清型、低水平的插入诱变和缺乏明显的致病性而成为用于转基因递送和基因治疗的通用载体,但其有限的包装能力(~4.7kb)使得有效递送许多单一效应子CRISPR酶具有挑战性(Abudayyeh et al.,2017;Ran et al.,2015;Swiech et al.,2015)。Cas13d效应子的尺寸显著地小,使得其独特地适合于使用CRISPR阵列、任选的效应子结构域和必需的表达或调控元件进行一体化(all-in-one)AAV递送(图12C)。

[0611] 连锁于染色体17(FTDP-17)伴帕金森症的额颞叶痴呆是由编码tau的基因MAPT中的多样点突变造成的常染色体显性遗传的主要神经退行性疾病。tau在人神经元中以两种主要的同种型存在,4R和3R,它们的区别在于存在或不存在tau外显子10,因此含有4或3个微管结合结构域。在FTDP-17以及其它tau蛋白病中,这两种同种型的平衡通常受到干扰,驱动神经退化的进展(Boeve and Hutton,2008)。一些形式的FTD由MAPT外显子10后的内含子中的突变造成,其破坏内含子剪接沉默子,并提高4R tau的表达(Kar et al.,2005),从而诱导病理变化(Schoch et al.,2016)。

[0612] 据推断,靶向MAPT外显子10的dCasRx可以诱导外显子排斥,以减轻异常调节的4R/3R tau比率。将来源于患者的人诱导多能干细胞(hiPSC)经由Neurogenin-2定向分化2周分化成皮质神经元(Zhang et al.,2013)。然后将有丝分后神经元用携带与重复阵列配对的dCasRx(图12D)的AAV1转导,所述重复阵列含有3个靶向外显子10剪接受体和两个推定外显子剪接增强子的间隔子(图12E)。dCasRx-介导的外显子排斥能够将4R/3R tau相对比率相对于BFP载体对照降低近50%(图12F),并降低至与未受影响的对照神经元相似的水平,这证明CasRx可以经由AAV递送用于原代细胞类型中的转录调节。

[0613] 实施例8

[0614] 使用Cas13d在人细胞中进行RNA靶向

[0615] 使用活性Cas13d核酸酶可以在人细胞中靶向RNA。作为概念验证,将人U-20S骨肉瘤细胞用mCherry报告子稳定地整合,并用编码人密码子优化的Cas13d和靶向mCherry转录物的向导RNA的质粒转染(图13)。

[0616] Cas13d蛋白质还与N-和C-端NLS序列(对于N-端NLS,SPKKKRKVEAS,SEQ ID NO:256;对于C-端NLS,GPKKKRKVAAA,SEQ ID NO:258)融合,以理解核定位是否可以影响mCherry敲低(这些记作2x NLS构建物)。在具有U6启动子的载体中提供向导RNA,该U6启动子与36nt DR-30nt间隔子-36nt DR序列可操作连接(其模拟未加工的CRISPR向导阵列(记作DR36)),或者与30nt DR-22nt间隔子可操作连接,以模拟经加工的成熟gRNA(记作gRNA)。推测DR36构建物在细胞内被Cas13d加工为成熟gRNA。DR36或gRNA分子中的间隔子序列与mCherry靶RNA(中靶(on-target)mCherry)互补,或者加以计算优化以避免与mCherry或任何内源性人转录物互补(非靶向mCherry)。

[0617] 将mCherry敲低通过流式细胞术加以定量,并归一化为转染对照。通过流式细胞术,非靶向的mCherry向导并未影响mCherry蛋白质水平,据推测是因为mCherry转录物没有被靶向。然而,与4种不同的Cas13d直向同源物配对的中靶mCherry向导表现出显著的mCherry敲低(图13)。“XPD”是指产黄瘤胃球菌XPD3002 Cas13d;“P1E0”是指肠道宏基因组P1E0 Cas13d;“AnDig”是指厌氧消化池肠道宏基因组Cas13d。“未培养”指未培养的瘤胃球菌属Cas13d。

[0618] 实施例9

[0619] 使用Cas13d进行体内RNA靶向

[0620] 可以在小鼠癌症模型中靶向RNA。为了观察小鼠中哪些细胞在表达EGFR,设计包括一个或多个与小鼠EGFR互补的间隔子区域的向导RNA,并将其与具有突变HEPN结构域的Cas13d(例如SEQ ID NO:2或4)和生物素标记物组合。将gRNA和Cas13d编码序列克隆到病毒载体(例如慢病毒)中,其用于以确保100%感染率的滴度通过尾静脉注射感染小鼠。向小鼠施用荧光链霉素亲和素标记物。将表达EGFR的细胞可视化,并用适当的激发频率进行荧光标记检测。另外可选地,Cas13d以其活性形式在体内递送,以介导靶敲低。

[0621] 实施例10

[0622] 癌症的治疗

[0623] 具有组织学证实的1期EGFR+乳腺癌的人受试者可以用公开的方法治疗。在接受肿瘤切除手术之后,为每个受试者施用包括活性Cas13d或在HEPN结构域中突变的Cas13d蛋白质(例如SEQ ID NO:2或4)、靶向EGFR的向导RNA和毒素的复合物。监测被治疗的个体的乳腺癌复发。

[0624] 实施例11

[0625] HIV感染的治疗

[0626] 具有HIV感染的人受试者可以用公开的方法治疗。为每个受试者施用包括活性Cas13d或在HEPN结构域中突变的Cas13d蛋白质(例如SEQ ID NO:2或4)、靶向HIV Nef蛋白质的向导RNA和毒素的构建物。监测被治疗的个体的HIV进展。

[0627] 实施例12

[0628] 亨廷顿病的治疗

[0629] 具有亨廷顿病的人受试者可以用公开的方法治疗。为每个个体施用包括Cas13d、靶向亨廷顿突变的向导RNA的构建物。监测被治疗的个体的疾病进展。

[0630] 实施例13

[0631] 使用Cas13d进行可变剪接

[0632] Cas13d剪接效应子可以用于治疗性蛋白质恢复(例如由突变或缺失引起)、经由移码诱导(frameshift induction)的基因敲低、调节或恢复所需的同种型比例、或者诱导所需的显性剪接同种型(图14)。可变剪接通常通过pre-mRNA中的顺式作用元件与正或负反式作用剪接因子的相互作用来调节,其可以介导外显子包涵或排斥。dCas13d和Cas13d,任选地与正或负剪接因子融合,可以用作剪接效应子,其靶向pre-mRNA中的所述顺式作用元件以操纵剪接。这些元件可以包括外显子剪接增强子或抑制子(suppressor)、内含子剪接增强子或抑制子、剪接受体和剪接供体位点,以及更通常地,在特定pre-mRNA、mRNA或其他RNA种类例如非编码RNA、tRNA、miRNA等上的蛋白质-或RNA-相互作用基序或元件。

[0633] 此外,基于Cas13d的剪接效应子的作用可以是向导位置依赖性的。这可以被用于经由位阻、阻断、募集或效应子介导的相互作用而扰乱或发现RNA转录物中的特定基序或位点,例如蛋白质结合位点。例如,可以干扰非编码RNA和特定染色质重构复合物之间的相互作用。核糖体结合位点和其它元件的接近可以在5'或3'UTR中被阻断(或募集适当的效应子结构域),以减少、增加或以其它方式操纵翻译。

[0634] 靶向Cas13d向导或者拼接Cas13d向导与pre-mRNA可以用于发现或定位新的顺式

作用元件,例如内含子或外显子剪接增强子。这已经在肌萎缩蛋白基因的治疗背景中被用于最优反义寡核苷酸的定位,并且也可以用于最优Cas13d的定位。这也可以用于定位(map)、掩蔽或干扰RNA代码或其它顺式作用元件,以影响运输和定位、染色质重构、聚腺苷酸化、RNA稳定性和半衰期或无义介导的衰变水平。

[0635] 在一实例中,靶向RNA允许改变靶RNA的剪接。剪接受体和/或供体位点的直接结合以及剪接效应子结构域都可以用于操纵剪接。例如,通过使用具有突变的HEPN结构域的dCas13d蛋白质(例如,SEQ ID NO:2或4)、含有至少一个对靶RNA特异的间隔子序列的向导RNA和任选的影响剪接的效应子结构域(例如SRSF1的RS富含结构域、hnRNP1的Gly富含结构域、RBM4的丙氨酸富含基序或者DAZAP1的脯氨酸富含基序),可以实现RNA的可变剪接。

[0636] 在一些实例中,这样的方法用于外显子包涵,例如,包含酸性 α -葡糖苷酶(GAA)的外显子2以治疗脊髓性肌肉萎缩症(SMA)。在一些实例中,这样的方法用于外显子排斥,例如,恢复肌萎缩蛋白的阅读框以治疗杜氏肌营养不良;将Bcl-x pre-mRNA的剪接从抗凋亡的长同种型转变成促凋亡的短同种型以治疗癌症;转变MAPT转录物的剪接以影响3R和4R tau的比率;或者在儿童早衰症(Hutchinson-Gilford progeria syndrome)或其他遗传性衰老疾病的情形中操纵层黏连蛋白A转录物的剪接。

[0637] 在一些实例中,该方法使用任选地具有突变的HEPN结构域的Cas13d蛋白质,以掩蔽剪接受体或供体位点,例如,以产生新生抗原,使冷肿瘤变热。通过影响某些靶pre-mRNA的剪接,该方法可以产生新的外显子-外显子连接,其可以导致在癌细胞中产生新生表位。这可以由于非天然抗原的展示而使癌细胞易受免疫系统攻击。在其它实例中,该方法可以用于动态地操纵同种型比率或恢复蛋白质(例如,杜氏肌营养不良症的肌营养不良蛋白)的阅读框。

[0638] 实施例14

[0639] Cas3d的AAV递送

[0640] 如以上实施例中所述,Cas13d可以有效地包装到AAV中,以介导不易于质粒递送的细胞类型中的表达或者用于Cas13d的体内递送。核酸酶活性Cas13d的AAV递送可以用于介导目的细胞类型中的RNA靶敲低。由于与其他的单一效应子CRISPR核酸酶相比Cas13d的尺寸小,Cas13d可以与向导RNA或者含有多个向导RNA的阵列一起包装在单一AAV载体中。

[0641] 实施例15

[0642] 用Cas3d进行基于核酸的诊断

[0643] 在细胞的背景下Cas13d酶可以用于基于核酸的诊断,其使用来源于细胞的无细胞裂解物或者含有改造的Cas13d酶和向导RNA的无细胞系统,以有利于形成核糖核蛋白复合物。所述向导RNA可以pre-向导RNA、成熟向导RNA或者含有一个或多个间隔子序列的阵列的形式提供。组件还可以经由体外转录/翻译系统以编码Cas13d酶的DNA或RNA前体以及适当的向导RNA设计的形式提供,以有利于产生必需的组件。诊断试剂盒的这些组件包括“传感器”模块。

[0644] 这样的方法可以用于确定测试样品中是否存在靶RNA。这样的方法还可以用于检测病原体,例如病毒或细菌,或者诊断疾病状态,例如癌症(例如,其中靶RNA对特定微生物或疾病有特异性)。这样的方法还可以用于测试环境样品或农业样品例如种子或土壤的纯度或同一性。

[0645] “传感器”模块则将会用RNA形式的测试样品激发。所述测试样品可以是,但不限于,例如经由体外转录转化成RNA的基因组DNA样品或者直接的RNA样品。这些样品可以提取自生物样品,例如患者样品(例如,细胞、组织、血液、血浆、血清、唾液、尿、肿瘤活检、无细胞DNA或RNA、外来体、载体囊泡或颗粒)和环境样品(例如,土壤、水、空气、种子或植物样品)。在提高诊断灵敏度的实施方式中,样品中的核酸分子使用扩增技术来扩增,例如聚合酶链式反应、重组酶聚合酶扩增、环介导等温扩增、基于核酸序列的扩增、链置换扩增、滚环扩增、连接酶链式反应和其他扩增技术(例如,使用等温扩增的扩增技术)。所述扩增技术可以任选地使用核酸转换技术,例如使用随机化引物或靶向引物的转录或逆转录。

[0646] 如果传感器模块在测试样品中识别出同源靶,其将会激活RNA酶活性。该RNA酶活性可以通过使用可检测标记物来检测。在一实例中,可检测标记物包括与荧光团和淬灭剂连接的RNA。完整的可检测RNA连接荧光团和淬灭剂,抑制荧光。一旦可检测RNA被Cas13d切割,荧光团从淬灭剂中释放,并表现出可检测的荧光活性。

[0647] 在另一实例中,报告RNA的切割释放出非荧光分子,其可以转化成可见的信号(例如,肉眼可见)。在一实例中,报告RNA的切割释放出可以经由侧向层析(lateral flow)检测的分子。可以通过侧向层析检测的分子是任何可以被抗体特异性结合的分子。在一实例中,Cas13蛋白质连同检测靶的向导RNA以及与报告分子缀合的报告RNA可以作为单一系统以干燥测试条形式递送。与测试样品一起孵育后,Cas13蛋白质a3d、向导RNA和报告RNA在RNA靶的存在下发生再水合。Cas13d蛋白质将会切割报告RNA,导致报告分子经由侧向层析在测试条中迁移,并通过与定位在此处的抗体结合而产生阳性测试线信号。这种不要求专门(冷冻)储存的耐储藏干燥检测系统,例如,在以下情形中可以证明是有优势的:其中靶RNA或DNA的检测在集中的实验室设施外部进行,例如,医务室、医院、药房、野外工作中、农业环境中等。

[0648] Cas13d在很宽的温度范围内是活性的,使得这种受控实验室环境外的应用变得可行。

[0649] 实施例16

[0650] Cas13d作为RNA或体外转录为RNA的DNA的诊断剂

[0651] Cas13d能够在最小的体外诊断系统中将匹配的靶RNA的存在转化成可见的信号。图15A-15D(A)在结合匹配向导RNA的间隔子序列的靶后,Cas13d转化成活性RNA酶复合物。它能够切割gRNA-互补靶RNA或非互补的旁观者RNA。(B) Cas13d的靶依赖性RNA酶活性例如可以通过切割标记的检测RNA而转化成可检测的信号,上述检测RNA仅在匹配Cas13d向导RNA的间隔子的靶的存在下被切割。在本实施例中,检测RNA含有荧光团‘F’和消除荧光的淬灭剂‘Q’。仅在旁观者RNA切割后,荧光团从淬灭剂释放并产生荧光。(C)来自惰性真杆菌的Cas13d仅在存在完美匹配的靶、且不存在不同的错配靶的情况下产生可见的信号。(D)来自黄化瘤胃球菌XPD3002株的Cas13d仅在存在完美匹配的靶、且不存在不同的错配靶的情况下产生可见的信号。

[0652] 因此,本文公开的系统是可以用于诊断的侧向层析装置(或其他固相载体)的部分。RNA或DNA序列的存在可以转化成信号,然后通过常规侧向层析检测所述信号。

[0653] 实施例17

[0654] Cas13d修饰

[0655] 图16A-16B, Cas13d适合多种修饰, 所述修饰包括在直向同源物中具有低保守性的区域的截短。图16A中的Cas13d直向同源物的比对示出具有高保守性(绿色条)和低保守性(红色条)的区域。

[0656] 实施例18

[0657] 体内靶向转录物

[0658] 使用以下不同的nCas1直向同源物在体内靶向细菌细胞中的ccdB基因: 惰性真杆菌nCas1 (Es_nCas1; SEQ ID NO:1); 具有突变的HEPN结构域的惰性真杆菌nCas1 (Es_nCas1HEPN-/-; SEQ ID NO:2); 未培养的瘤胃球菌属nCas1 (uncul_nCas1; SEQ ID NO:3) 和具有突变的HEPN结构域的未培养的瘤胃球菌属nCas1 (uncul_nCas1 HEPN-/-; SEQ ID NO:4) (图17A)。将化学感受态大肠杆菌 (BW25141-DE3株) 细胞用以下转化: (1) 阿拉伯糖-诱导型ccdB质粒, 和(2) 第二质粒(靶向载体), 其携带相容的复制起点、nCas1蛋白质编码序列和含有4个靶向ccdB转录物的间隔子序列的nCas1向导阵列(图17A)。

[0659] 1. 靶向ccdB的人工惰性真杆菌nCas1阵列

[0660] a. GAACUACACCCGUGCAAAAAUGCAGGGGUCUAAAACUAACGGCUCUCUCUUUUUAGGUGUAAACC GAACUACACCCGUGCAAAAAUGCAGGGGUCUAAAACUUUAUCUGACAGCAGACGUGCACUGGCCAGAACUACACC CGUGCAAAAAUGCAGGGGUCUAAAACCAUCAUGCGCCAGCUUUAUCCCCGAUAUGGAACUACACCCGUGCAAAAA UGCAGGGGUCUAAAACUAAUGGCGUUUUUGAUGUCAUUUUCGCGGUCCGUGA (SEQ ID NO:261)

[0661] i. 完整的36nt同向重复:

[0662] GAACUACACCCGUGCAAAAAUGCAGGGGUCUAAAAC (SEQ ID NO:262)

[0663] ii. 间隔子1: UAACGGCUCUCUCUUUUUAGGUGUAAACC (SEQ ID NO:263)

[0664] iii. 间隔子2: CUUUAUCUGACAGCAGACGUGCACUGGCCA (SEQ ID NO:264)

[0665] iv. 间隔子3: CAUCAUGCGCCAGCUUUAUCCCCGAUAUG (SEQ ID NO:265)

[0666] v. 间隔子4: UAAUGGCGUUUUUGAUGUCAUUUUCGCGGUCCGUGA (SEQ ID NO:266)

[0667] 2. 靶向ccdB未培养的人工瘤胃球菌属nCas1阵列

[0668] a. CUACUACACUGGUGCAAAUUUGCACUAGUCUAAAACUAACGGCUCUCUCUUUUUAGGUGUAAACC CUACUACACUGGUGCAAAUUUGCACUAGUCUAAAACUUUAUCUGACAGCAGACGUGCACUGGCCACUACUACACU GGUGCAAAUUUGCACUAGUCUAAAACCAUCAUGCGCCAGCUUUAUCCCCGAUAUGCUACUACACUGGUGCAAAUU UGCACUAGUCUAAAACUAAUGGCGUUUUUGAUGUCAUUUUCGCGGUCCGUGA (SEQ ID NO:267)

[0669] i. 完整的36nt同向重复:

[0670] CUACUACACUGGUGCAAAUUUGCACUAGUCUAAAAC (SEQ ID NO:268)

[0671] ii. 间隔子1: UAACGGCUCUCUCUUUUUAGGUGUAAACC (SEQ ID NO:269)

[0672] iii. 间隔子2: CUUUAUCUGACAGCAGACGUGCACUGGCCA (SEQ ID NO:270)

[0673] iv. 间隔子3: CAUCAUGCGCCAGCUUUAUCCCCGAUAUG (SEQ ID NO:271)

[0674] v. 间隔子4: UAAUGGCGUUUUUGAUGUCAUUUUCGCGGUCCGUGA (SEQ ID NO:272)

[0675] 3. 靶RNA (ccdB序列)

[0676] a. AUGCAGUUUAAGGUUUACACCUAUA AAAAGAGAGCCGUUAUCGUCUGUUUGGGAUGUACAGAGU GAUAUUUUGACACGCCCGGGGACGGAUGGUGAUCCCCUGGCCAGUGCACGUCUGCUGUCAGAUAAAGUCUCCC GUGAACUUUACCCGUGGUGCAUAUCGGGAUGAAAGCUGGCGCAUGAUGACCACCGAU AUGGCCAGUGUGCCGGU CUCCGUUAUCGGGGAAGAAGUGGCUGAUCUCAGCCACCGCAAAAUGACAUCAAAAACGCCAUUAACCGAUGUUU

UGGGGAAUA (SEQ ID NO:273)

[0677] i. 间隔子1靶:GGUUUACACCUAUA AAAAGAGAGAGCCGUUA (SEQ ID NO:274)

[0678] ii. 间隔子2靶:UGGCCAGUGCACGUCUGCUGUCAGAUAAAAG (SEQ ID NO:275)

[0679] iii. 间隔子3靶:CAUAUCGGGGAUGAAAAGCUGGCGCAUGAUG (SEQ ID NO:276)

[0680] iv. 间隔子4靶:ACCGCGAAAAUGACAUCAAAAACGCCAUUA (SEQ ID NO:277)

[0681] v. dCas9靶(下划线):UUAUCGUCUGUUUGUGGAUG (SEQ ID NO:259)

[0682] 将转化的细菌铺板在2mM阿拉伯糖板上,以诱导ccdB表达,并24小时之后收获。用Trizol提取总RNA,然后进行随机六聚体介导的逆转录和基于Taqman探针的qPCR。

[0683] 每种nCas1蛋白质中两个HEPN结构域的突变均没有表现出对ccdB的靶向(图17B, HEPN^{-/-}),而活性的野生型nCas1蛋白质敲低了ccdB的表达(图17B, Es_nCas1; uncul_nCas1)。由于通过结合转录起始位点下游的DNA序列的dCas9-(无催化活性的或“死亡的”Cas9)-介导的转录抑制是目前本领域的标准,我们通过将dCas9靶向ccdB启动子以抑制ccdB基因的转录对该检验进行验证。总之,该数据证明了靶转录物在原核细胞内以HEPN结构域依赖性方式在体内发生向导RNA特异性降解。

[0684] 实施例19

[0685] Cas13d直向同源物减少在人细胞中的表达

[0686] 测试50种不同的Cas13d直向同源物(其中一些包括NLS标签SPKKKRKVEAS SEQ ID NO:256和GPKKKRKVA AAA SEQ ID NO:25,或两个HIV NES序列(LQLPPLRLTL, SEQ ID NO:287, 参见表1)在293细胞中降低mCherry或内源性CD81表达的能力。

[0687] 细胞培养

[0688] 在补充有10%FBS(GE Life Sciences)和10mM HEPES的DMEM(4.5g/L葡萄糖)中,在37°C、5%CO₂条件下维持人胚肾(HEK)细胞系293FT(Thermo Fisher)。达到80-90%汇合后,使用TrypLE Express(Life Technologies)解离细胞,并以1:2的比率传代。

[0689] 人细胞系的瞬时转染

[0690] 将Cas13d直向同源物克隆到常见的表达骨架中,该表达骨架的侧翼为两个HIV NES序列(SEQ ID NO:287)或两个SV40 NLS序列(SPKKKRKVEAS, SEQ ID NO:256和GPKKKRKVA AAA, SEQ ID NO:25),和T2A-连接的EGFP。将靶向CD81的间隔子克隆到向导表达构建物中,其中两个间隔子的阵列由U6启动子驱动。

[0691] 在瞬时转染之前,将HEK 293FT细胞以20,000细胞/孔的密度铺板到96-孔板中,并根据制造商的方案,使用Lipofectamine2000(Life Technologies),在>90%汇合下用200ng Cas13表达质粒和200ng gRNA表达质粒进行转染。转染后72小时收获转染的细胞用于流式细胞术。

[0692] 流式细胞术

[0693] 在通过流式细胞术分析之前,将293FT细胞在室温下在FACS缓冲液(1X DPBS^{-/-}, 0.2%BSA, 2mM EDTA)中用Human TruStain FcX封闭液(Biolegend)封闭10分钟。然后将细胞在冰上用α-CD81-APC(293FT为1:400, Miltenyi Biotec)或REA293(S)同种型对照(Miltenyi Biotec)抗体染色15分钟。使用MACSQuant VYB(Miltenyi Biotec)以96孔板格式进行流式细胞术,并使用FlowJo 10进行分析。

[0694] 如表1所示,若干直向同源物具有显著的敲低活性(即,敲低效率数值较低)。在表1

中,(trunc)表示具有非保守N端截短的直向同源物的形式。NES表示具有双重融合物(SEQ ID NO:287)的直向同源物,NLS表示具有双重SV40 NLS融合物(SEQ ID NO:256和25)的直向同源物。

[0695] 表1:人细胞中Cas13d直向同源物的mCherry活性筛选

直向同源物	敲低效率	标准偏差	Seq ID No.
Ga7274-NLS	12.7	1.0	198
Rfx-NLS	18.6	2.6	280
Ur(trunc)-NLS	19.0	1.2	3
twins_985-NES	20.7	1.3	212
Ga_0531-NES	22.3	2.8	235
Ga_0531-NLS	22.6	0.5	235
k87-NES	24.7	3.1	189

[0696]

[0697]

twins_985-NLS	26.4	1.8	212
Ur(trunc)-NES	27.7	1.0	3
stool_9834-NLS	29.0	2.4	208
OGPN-NES	32.6	1.9	164
Ur-NLS	33.9	0.7	3
OGZC-NES	34.7	1.7	155
Ga_9260-NLS	36.5	6.6	237
NFIR-NLS	36.7	1.5	166
Ga_9260-NES	38.3	2.9	237
DJXD-NLS	39.2	3.0	153
OGZC-trunc	40.3	1.1	155
k87-NLS	41.1	4.6	189
PIG_6364-NLS	41.2	6.3	243
twins_300-NLS	42.8	5.9	210
DLF_1914-NES	44.5	10.5	224
BMZ_7771-NLS	46.4	5.9	218
PIG_3397-NES	46.9	3.2	245
Ga7274-NES	47.7	3.6	198
PIG_5479-NES	49.1	3.6	249
OCTW-NLS	50.1	10.8	183
PIG_6364-NES	50.3	2.1	243
Piggy-NES	51.6	5.4	170
OGNF-NLS	52.3	3.2	185
Ga_8792-NES	52.9	1.9	206
OJMM-NES	53.1	1.5	175
OGDF-NLS	53.6	3.4	162
GL_3589-NLS	53.7	3.5	216
Ga_5336-NLS	53.7	0.6	233
OGDF-NES	54.6	1.4	162
BMZ_7915-NES	55.7	8.0	220
OCTW-NES	55.9	10.8	183
OBVH-NLS	56.5	5.2	147
Ga8792-NLS	59.0	8.8	206
OJMM-NLS	59.1	1.6	175
BMZ_7915-NLS	59.9	9.7	220
BMZ_7771-NES	59.9	4.6	218
Rfx-NES	60.5	7.7	92
Ga_4192	62.2	5.8	229
twins_300-NES	62.4	8.4	210
DJXD-NES	63.5	8.2	153
Ga_8067-NES	63.7	10.9	204
PIG_5479-NLS	64.4	20.8	249
Ga0735-NLS	65.2	12.4	202
OGPN-NLS	66.2	6.4	164
Piggy-NLS	66.3	8.0	147
MH_4866-NLS	67.1	4.1	292
OIEN-NES	67.8	16.6	187
Ga_4192-NLS	68.7	2.2	229
Ga_0735-NES	69.2	12.0	202
OHCP-NLS	69.2	4.1	160

	OIEN-NLS	69.6	7.2	187
	DBYI-NLS	70.5	13.7	149
	Ga5759-trunc	71.2	15.2	251
	DBYI-NES	71.5	7.1	149
	stool_9834-NES	72.6	3.9	208
	Ga_6931-NLS	73.4	5.2	231
	MH_4866-NES	76.6	3.1	292
	O2UC_6317-NES	79.2	2.7	241
	OGNF-NES	80.2	10.2	185
	Ga_6931-NES	83.3	15.9	231
	OCVV-NES	84.8	14.7	181
[0698]	O2UC_6317-NLS	86.2	6.6	241
	ODAI_557-NES	86.5	14.4	
	Ga5759-NLS	91.1	10.3	251
	ODAI1274-NLS	92.0	9.0	177
	Ga_5336-NES	93.5	2.8	233
	OCVV-NLS	93.5	5.6	181
	GL_3589-NES	95.4	3.5	216
	PIG_9734-NLS	97.1	10.6	247
	EMG3641-NES	105.8	7.6	200
	GL_4153-NLS	106.7	5.3	214
	BMZ_9617-NES	110.0	7.8	222
	EMG3641-NLS	112.6	11.3	200
	Ga8067-NLS	111.4	16.1	204

[0699] 以相对mCherry表达(敲低效率)相对直向同源物的大小,对表1中的数据进行作图,并与CasRx(SEQ ID NO:280)比较。如图19A所示,三种直向同源物(Ga0531-NLS、k87-NES和Ga7274-NLS,分别为SEQ ID NO:235、189和198)显示出与CasRx变体(SEQ ID NO:280)相当的活性。

[0700] 测试了表1和图19A中所示的25种活性最高的直向同源物在HEK293细胞中降低内源性基因CD81表达的能力。如图19B所示,CasRx变体RfxCas13d-NLS(SEQ ID NO:280)表现出最高水平的敲低,若干其他直向同源物表现出对该内源性靶>75%的敲低。

[0701] 对五种表现出>70%敲低的直向同源物的分析表明,与具有低活性(<50%敲低)的Cas13d直向同源物相比,以下共有基序在人细胞的具有高活性的Cas13d直向同源物中以较高的频率存在。因此,本文提供了具有这种共有基序的Cas13d蛋白质(例如包括SEQ ID NO:288、289、290或291的蛋白质序列,以及编码这些蛋白质的核酸,编码这些核酸的载体,和使用这些蛋白/核酸的方法):

[0702] 1.L-[DE]-G-K-E-I-N-x-[LM]-x-[CST]-[ANST]-L-I-x-K-F-[DE]-[GN]-I-x-[ADS]-[FIL]-x-[DEHK]-[AILV](SEQ ID NO:288)

[0703] 2.R-N-[FY]-I-[AIL]-x-N-V-[ILV]-x-[NS]-x(2)-F-x-Y-[ILV]-[AIV]-[KR]-[HY]-x(2)-[PT]-x(3)-[HKR]-x-[IL]-x(2)-N-x(2)-[ILV]-x(2)-[FY]-V-L-x(2)-[ILM]-[AGP]-[DE]-x(4)-[NQRS](SEQ ID NO:289)

[0704] 3.L-x-[LM]-[ST]-[IV]-x-Y-x(3)-K-[NS]-[ILM]-V-x-[IV]-N-[AST]-R-[FY]-x(2)-[AG]-[FY]-x-[ACV]-[LV]-E-R-D-x(2)-[FL](SEQ ID NO:290)

[0705] 4.P-F-[AGS]-Y-[CNV]-x-[AP]-R-[FY]-K-N-L-[ST]-x-[EN]-x-L-F-x-[DKR]-[DEN](SEQ ID NO:291)

[0706] 实施例20

[0707] CasRx与PspCas13b-NES的敲低活性比较

[0708] 将CasRx(RfxCas13d-NLS设计;SEQ ID NO:280)在293FT细胞中敲低内源性CD81蛋白质表达的能力与PspCas13b-NES(参见Cox et al.,Science.2017Nov 24;358(6366):1019-1027)进行比较。培养293FT细胞,将其瞬时转染,并在三天之后,使用流式细胞术检测CD81表达,如实施例19所述。

[0709] 如图20所示,CasRx(SEQ ID NO:280)在相等的蛋白质水平下表现出显著较高的活性(如通过Cas13-GFP表达所测量的),表明每个酶单位的活性较高,因此,达到生物学/药理学上相关的敲低水平要求的剂量较低。此外,CasRx显著小于PspCas13b。CasRx的小尺寸以及较低剂量下的高活性使得其可用于治疗递送。

[0710] 实施例21

[0711] 缺失对Cas13d直向同源物的影响

[0712] 缺失对Cas13d直向同源物的影响测试如下。简言之,将氨基酸(aa)缺失改造为两种不同的Cas13d直向同源物Ga-531(SEQ ID NO:235)和RfxCas13d(SEQ ID NO:280)中。氨基酸缺失在N-端、C-端或在内部进行。基于其在Cas13d直向同源物之间的低保守性选择截短的区域(例如,参见图18A-18MMM)。在单一蛋白质中进行至多50aa的缺失。产生的蛋白质序列示于表2中。

[0713] 表2:Cas13d突变(黑体=NLS/NES标签,下划线=插入的接头序列;斜体=侧翼序列)

[0714]

效应子	序列 (SEQ ID NO:)	说明
NLS- Ga_0531(t runc)- NLS-HA	MSPKKKRKVEASKQSKKTFKASGL KSTLAVDNSAVMTVFGRGNEAKLD HRINADLQSESLHPQAALKNVHAPN KQKIHFGRMQDMNLTADHPLHSHD GERAVGADLLCAKDKLEQLYFGRTF NDNIHQIQLYQILDIQKILALHANNIF ALDNLLHKKNDELSDDFVGMGRMR ATIGYDAFRNSTNQKVQETYREFQEF VRRKELLYFGSAFYNGDTRRDEKVI YHILSLAASVRQFCFHNDYTSDDGK GFIKADWMYRLEEALPAEYKDTLDA LYLEGVEGLDQSFLKNNTVNIQILCSI FNHDDPNKIAEEYYGFLMTKEYKNM GFSIKKLRECMLELPELSGYKEDQYN SVRSKLYKLDFDIIAHYFRKHPEKGE EMVDCLRLCMTEDEKDSHYEGTAK KLVRELAYDMQEAAEQANGSNITQ MQKNEQQGKTKGMFAIRDEIRVSRK PVSYFSKVIYVMTLLLDGKEINDLLT TLINKFENIVSFEDVLRQLNVDCTFKP EFAFFGYDRCRNISGELRLINSFARM QKPSAKAKHVMYRDALRILGLDNG MSEEALDQEVRRILQIGADGKPIKNA NKGFRNFASNVISSRFYLVRYNN PHKTRMIAQNEAIVRFVLSEIPDEQIR RYYDVCRDPKLPSSSREAQVDILTG IITDVNYRIFEDVPQSKKINKDRPDAN DRMTLKKQRYQAIVSLYLTVMYLVT	N-端缺失

[0715]

	KNLVYVNSRYVMAFHALERDAYLY GITNIKGDYRKLTDNLLADENYKKF GHFKNKKWRGIAEQNLRNSDVPVIK SFRNMAAHISVIRNIDL YIGDIQKVDS YFALYHFLMQKLIQRVVPENTKGLS DQTKKYYDALEQYNTYCKDFVKAY CTPFAYVTPRYKNLTIDGLFDRNRPG EDKGS PKKKRK VAAAYPYDVPDY A (SEQ ID NO: 278)	
NES- Ga_0531(t runc)- NES-HA	MLQLPPLERLTLGSGKQSK TKFAKA SGLKSTLAVDNSAVMTVFGRGNEAK LDHRINADLQESLHPQAALKNVHA PNKQKIHFIGRMQDMNLTADHPLHS HDGERAVGADLLCAKDKLEQLYFG RTFNDNIHQILYQILDIQKILALHAN NIIFALDNLLHKKNDELSDDFVGMG RMRATIGYDAFRNSTNQKVQETYRE FQEFVRRKELLYFGSAFYNGDTRRD EKVIYHILSLAASVRQFCFHNDYTS DGKGFADWMYRLEEALPAEYKDT LDALYLEGVEGLDQSFLKNNTVNIQI LCSIFNHDDPNKIAEYYGFLMTKEY KNMGFSIKKLRECMLELPELSGYKE DQNSVRSKLYKLFDFIIAHYFRKHP EKGEEMVDCLRLCMTEDKDSHYE GTAKKLVRELAYDMQEAQANGS NITQMOKNEQQGKTKGMFAIRDEIR VSRKPVSFYFSKVIYVMTLLLDGKEIN DLLTTLINKFENIVSFEDVLRQLNVD CTFKPEFAFFGYDRCRNISGELRLINS FARMQKPSAKAKHV MYRDALRILGL DNGMSEEALDQEVRRILQIGADGKPI KNANKGFRNFIA SNVISSRFRYLVR YNNPHKTRMIAQNEAIVRFVLSEIPD EQIRRYDVCRDPKLPSSSREAQVD ILTGITDVNYRIFEDVPQSKKINKDR PDANDRMTLKKQRYQAIVSLYLTVM YLVTKNLVYVNSRYVMAFHALERD AYLYGITNIKGDYRKLTDNLLADEN YKKFGHFKNKKWRGIAEQNLRNSD VPVIKSFRNMAAHISVIRNIDL YIGDI QKVDSYFALYHFLMQKLIQRVVPEN TKGLSDQTKKYYDALEQYNTYCKDF VKAYCTPFAYVTPRYKNLTIDGLFDR NRPGEDKGS GLQLPPLERLTL YPYD VPDYA (SEQ ID NO: 279)	N-端缺失
NLS- RfxCas13 d-NLS- HA	MSPKKRK VEASIEKKKSFAGMG VKSTLVSGSKVYMTTFAEGSDARLE KIVEGDSIRSVNEGEAFSAEMADKNA GYKIGNAKFSHPKGYAVVANNPLYT GPVQQDMLGLKETLEKRYFGESADG NDNICIQVIHNILDIEKILA EYITNAAY	无缺失

[0716]

	<p>AVNNISGLDKDIIGFGKFSTVYTYDE FKDPEHHRAAFNNNDKLINAIIKAQY DEFDNFLDNPRLGYFGQAFFSKEGR NYIINYGNECYDILALLSGLRHWVVH NNEEESRISRTWLYNLDKNLDNEYIS TLNYLYDRITNELTNSFSKNSAANVN YIAETLGINPAEFAEQYFRFSIMKEQK NLGFNITKLREVMLDRKDMSEIRKN HKVFDSIRTKVYTMMDFVIYRYIEE DAKVAAANKSLPDNEKSLSEKDIFVI NLRGSFNDDQKDALYYDEANRIWR KLENIMHNIKEFRGNKTREYKKKDA PRLPRILPAGRDVSAFSKLMYALTMF LDGKEINDLLTTLINKFDNIQSFLKV MPLIGVNAKFVEEYAFFKDSAKIADE LRLIKSFARMGEPIADARRAMYIDAI RILGTNLSYDELKALADTFSLDENG KLKKGKHGMRNFIINNVISNKRHYL IRYGDPAHLHEIAKNEAVVKFVLGRI ADIQKKQGQNGKNQIDRYYETCIGK DKGKSVSEKVDALTKIITGMNYDQF DKKRSVIEDTGRENAEREKFKKIISLY LTVIYHILKNIVNINARYVIGFHCVER DAQLYKEKGYDINLKKLEEKGFSSV TKLCAGIDETAPDKRKDVEKEMAER AKESIDSLESANPKLYANYIKYSDEK KAEEFTRQINREKAKTALNAYLRNT KWNVIHREDLLRIDNKTCTLFRNKAV HLEVARYVHAYINDIAEVNSYFQLY HYIMQRIIMNERYEKSSGKVSEYFDA VNDEKKYNDRLKLLCVPGYCI PRF KNLSIEALFDRNEAAKFDKEKKKVS GNSGSGPKKKRKVAAAYPYDVPDY A (SEQ ID NO: 280)</p>	
<p>NLS- RfxCas13 d(del5)- NLS-HA</p>	<p>MSPKKRKVEASIEKKKSFAGMG VKSTLVSGSKVYMTTFAEGSDARLE KIVEGDSIRSVNEGEAFSAEMADKNA GYKIGNAKFSPKGYAVVANNPLYT GPVQQDMLGLKETLEKRYFGESADG NDNICIQVIHNILDIEKILAEYITNAAY AVNNISGLDKDIIGFGKFSTVYTYDE FKDPEHHRAAFNNNDKLINAIIKAQY DEFDNFLDNPRLGYFGQAFFSKEGR NYIINYGNECYDILALLSGLRHWVVH NNEEESRISRTWLYNLDKNLDNEYIS TLNYLYDRITNELTNSFSKNSAANVN YIAETLGINPAEFAEQYFRFSIMKEQK NLGFNITKLREVMLDRKDMSEIRKN HKVFDSIRTKVYTMMDFVIYRYIEE DAKVAAANKSLPDNEKSLSEKDIFVI NLRGSFNDDQKDALYYDEANRIWR KLENIMHNIKEFRGNKTREYKKKDA</p>	<p>氨基酸 558-587 缺失</p>

[0717]

	<p>PRLPRILPAGRDVSAFSKLMYALTMF LDGKEINDLLTTLINKFDNIQSFLKV MPLIGVNAKFVEEYAFFKDSAKIADE LRLIKSFARMGEPIDARRAMYIDA/ <i>RILGGGGSGGGGSMRNF</i>IINNVISNK RFHYLIRYGDPAHLHEIAKNEAVVKF VLGRIADIQKKQGQNGKNQIDRYYE TCIGKDKGKSVSEKVDALTKIITGMN YDQFDKKRSVIEDTGRENAEREKFK KIISLYLTVIYHILKNIVNINARYVIGF HCVERDAQLYKEKGYDINLKKLEEK GFSSVTKLCAGIDETAPDKRKDVEKE MAERAKESIDSLESANPKLYANYIKY SDEKKAEEFTRQINREKAKTALNAY LRNTKWNVIREDLLRIDNKTCTLFR NKAVHLEVARYVHAYINDIAEVNSY FQLYHYIMQRIIMNEREYKSSGKVSE YFDAVNDEKKYNDRLLKLLCVPFGY CIPRFKNLSIEALFDRNEAAKFDKEK KKVSGNSGSGPKKKRKVAAAYPYD VPDYA (SEQ ID NO: 281)</p>	
<p>NLS- RfxCas13 d(del5.12) -NLS-HA</p>	<p>MSPKKKRKVEASIEKKKSFAGMG VKSTLVSGSKVYMTTFAEGSDARLE KIVEGDSIRSVNEGEAFSAEMADKNA GYKIGNAKFSHPKGYAVVANNPLYT GPVQQDMLGLKETLEKRYFGESADG NDNICIQVIHNILDIEKILAEYITNAAY AVNNISGLDKDIIGFGKFSTVYTYDE FKDPEHHRAAFNNNDKLINAIIKAQY DEFDNFLDNPRLG YFGQAFFSKEGR NYIINYGNECYDILALLSGLRHWVH NNEESRISRTWLYNLDKNLDNEYIS TLNYLYDRITNELTNSFSKNSAANVN YIAETLGINPAEFAEQYFRFSIMKEQK NLGFNITKLREVMLDRKDMSEIRKN HKVFDSIRTKVYTMMDFVIYRYIEE DAKVAANKSLPDNEKSLSEKDIFVI NLRGSFNDDQKDALYYDEANRIWR KLENIMHNIKEFRGNKTREYKKKDA PRLPRILPAGRDVSAFSKLMYALTMF LDGKEINDLLTTLINKFDNIQSFLKV MPLIGVNAKFVEEYAFFKDSAKIADE LRLIKSFARMGEPIDARRAMYIDA/ <i>RILGGGGSGGGGSMRNF</i>IINNVISNK RFHYLIRYGDPAHLHEIAKNEAVVKF VLGRIADIQKKQGQNGKNQIDRYYE TCIGKDKGKSVSEKVDALTKIITGMN YDQFDKKRSVIEDTGRENAEREKFK KIISLYLTVIYHILKNIVNINARYVIGF HCVERDAQLYKEKGYDINLKKLEEK GFSSVTKLCAGIDETAPDKRKDVEKE MAERAKESIDSLESANPKLYANYIKY</p>	<p>氨基酸 558-587 和 953-966 缺 失</p>

[0718]

	SDEKKAEEFTRQINREKAKTALNAY LRNTKWNVIREDLLRIDNKTCTLFR NKAVHLEVARYVHAYINDIAEVNSY FQLYHYIMQRIIMNEREYKSSGKVSE YFDAVNDEKKYNDRLKLLCVPFGY CIPRFKNLSIEALFDRNEGSG PKKKR KV AAAYPYDVPDYA (SEQ ID NO: 282)	
NLS- RfxCas13 d(del5.13) -NLS-HA	MSP PKKKRK VEASIEKKKSFAGMG VKSTLVSGSKVYMTTFAEGSDARLE KIVEGDSIRSVNEGEAFSAEMADKNA GYKIGNAKFSPKGYAVVANNPLYT GPVQQDMLGLKETLEKRYFGESADG NDNICIQVIHNILDIEKILAEYITNAAY AVNNISGLDKDIIGFGKFSTVYTYDE FKDPEHHRAAFNNNDKLINAIAQY DEFDNFLDNPRLGYPGQAFSKEGR NYIINYGNECYDILALLSGLRHWVH NNEESRISRTWLYNLDKNLDNEYIS TLNYLYDRITNELTNSFSKNSAANVN YIAETLGINPAEFAEQYFRFSIMKEQK NLGFNITKLREVMLDRKDMSEIRKN HKVFDSIRTKVYTMMDFVIYRY YIEE DSEKDI FVINLRGSFNDDQKDALYD EANRIWRKLENIMHNIKEFRGNKTRE YKKKDAPRLPRILPAGR DVSAF SKL MYALTMFLDGKEINDLLTTLINKFDN IQSFLKVMPLIGVNAKFVEEYAFFKD SAKIADELRLIKSFARMGEPIADARR AMYIDA IRILGGGGGSGGGGSMRNFII NNVISNKR FHYLIRYGDPAHLHEIAK NEAVVKFVLGRIADIQKKQGQNGKN QIDRYYETCIGKDKGKSVSEKVDALT KIITGMNYDQFDKKRSVIEDTGRENA EREKFKKIISLYLTVIYHILKNIVNINA RYVIGFHCVERDAQLYKEKGYDINL KKLEEKGFSSVTKLCAGIDETAPDKR KDVEKEMAERAKESIDSLESANPKL YANYIKYSDEKKAEEFTRQINREKAK TALNAYLRNTKWNVIREDLLRIDNK TCTLFRNKAVHLEVARYVHAYINDI AEVNSYFQLYHYIMQRIIMNEREYKS SGKVSEYFDAVNDEKKYNDRLKLL CVPFGYCIPRFKNLSIEALFDRNEAA KFDKEKKK VSGNSGSG PKKKRKVA AAAYPYDVPDYA (SEQ ID NO: 283)	氨基酸 376-392 和 558-587 缺 失
NLS- RfxCas13 d(del5.12 +5.13)- NLS-HA	MSP PKKKRK VEASIEKKKSFAGMG VKSTLVSGSKVYMTTFAEGSDARLE KIVEGDSIRSVNEGEAFSAEMADKNA GYKIGNAKFSPKGYAVVANNPLYT GPVQQDMLGLKETLEKRYFGESADG NDNICIQVIHNILDIEKILAEYITNAAY	氨基酸 376-392、558-587 和 953-966 缺失

[0719]

	<p>AVNNISGLDKDIIGFGKFSTVYTYDE FKDPEHHRAAFNNNDKLINAIIKAQY DEFDNFLDNPRLGYPFGQAFFSKEGR NYIINYGNACYDILALLSGLRHWVVH NNEESRISRTWLYNLDKNLDNEYIS TLNYLYDRITNELTNSFSKNSAANVN YIAETLGINPAEFAEQYFRFSIMKEQK NLGFNITKLREVMLDRKDMSEIRKN HKVFDSIRTKVYTMMDVFIYRY <i>YIEE</i> <i>DSEKDFVINLRGSFNDDQKDALYYD</i> EANRIWRKLENIMHNIKEFRGNKTRE YKKKDAPRLPRILPAGR DVSAF SKL MYALTMFLDGKEINDLLTTLINKFDN IQSFLKVMPLIGVNAKFVEEYAFFKD SAKIADELRLIKSFARMGEPIADARR AMYIDAIRILGGGGGSGGGGSMRNFII NNVISNKR FHYLIRYGDPAHLHEIAK NEAVVKFVLGRIADIQKKQGQNGKN QIDRYYETCIGKDKGKSVSEKVDALT KIITGMNYDQFDKKRSVIEDTGRENA EREKFKKIISLYLTVIYHILKNIVNINA RYVIGFHCVERDAQLYKEKGYDINL KKLEEKGFSSVTKLCAGIDETAPDKR KDVEKEMAERAKESIDSLESANPKL YANYIKYSDEKKAEEFTRQINREKAK TALNAYLRNTKWNVIREDLLRIDNK TCTLFRNKAVHLEVARYVHAYINDI AEVNSYFQLYHYIMQRIIMNEREYKS SGKVSEYFDAVNDEKKYNDRLKLL CVPFGYCIPRFKNLSIEAL <i>FDRNEGSG</i> PKKKRKVAAAYPYDVPDYA (SEQ ID NO: 284)</p>	
<p>NLS- RfxCas13 d(del13)- NLS-HA</p>	<p>MSPPKKKRKVEASIEKKKSFAGMG VKSTLVSGSKVYMTTFAEGSDARLE KIVEGDSIRSVNEGEAFSAEMADKNA GYKIGNAKFSHPKGYAVVANNPLYT GPVQQDMLGLKETLEKRYFGESADG NDNICIQVIHNILDIEKILAEYITNAAY AVNNISGLDKDIIGFGKFSTVYTYDE FKDPEHHRAAFNNNDKLINAIIKAQY DEFDNFLDNPRLGYPFGQAFFSKEGR NYIINYGNACYDILALLSGLRHWVVH NNEESRISRTWLYNLDKNLDNEYIS TLNYLYDRITNELTNSFSKNSAANVN YIAETLGINPAEFAEQYFRFSIMKEQK NLGFNITKLREVMLDRKDMSEIRKN HKVFDSIRTKVYTMMDVFIYRY <i>YIEE</i> <i>DSEKDFVINLRGSFNDDQKDALYYD</i> EANRIWRKLENIMHNIKEFRGNKTRE YKKKDAPRLPRILPAGR DVSAF SKL MYALTMFLDGKEINDLLTTLINKFDN IQSFLKVMPLIGVNAKFVEEYAFFKD</p>	<p>氨基酸 376-392 缺失</p>

[0720]	SAKIADELRLIKSFARMGEPIADARR AMYIDAIRILGTNLSYDELKALADTF SLDENGKLLKKGKHGMRNFIINNVIS NKRFHYLIRYGDPAPHLHEIAKNEAV VKFVLGRIADIQKKQGQNGKNQIDR YYETCIGKDKGKSVSEKVDALTKIIT GMNYDQFDKRSVIEDTGRENAERE KFKKIISLYLTVIYHILKNIVNINARY VIGFHCVERDAQLYKEKGYDINLKK LEEKGFSSVTKLCAGIDETAPDKRKD VEKEMAERAKESIDSLESANPKLYA NYIKYSDEKKAEEFTRQINREKAKTA LNAYLRNTKWNVIREDLLRIDNKTC TLFNRKAVHLEVARYVHAYINDIAE VNSYFQLYHYIMQRIIMNERYEKSSG KVSEYFDAVNDEKKYNDRLLKLLCV PFGYCIPRFKNLSIEALFDRNEAAKFD KEKKKVSNGSGSG PKKKRK VAAAY PYDVPDYA (SEQ ID NO: 285)	
--------	--	--

[0721] 培养293FT细胞,将其瞬时转染,并在三天之后,使用流式细胞术检测CD81表达,如实施例19所述。

[0722] 如图21所示,在低保守区域中含有缺失的Cas13d直向同源物CasRx和Ga0531在人HEK 293FT细胞中保持敲低活性。基于这些结果,可以使用本文提供的序列比对,对本文提供的其他Cas13d直向同源物进行类似的缺失。

[0723] 实施例22

[0724] dCas13d介导的剪接

[0725] 若干表现出高水平蛋白质敲低(参见实施例19-20)的Cas13d直向同源物(天然的和截短的变体,参见实施例21)指导剪接的能力测试如下。将Cas13d直向同源物(具有突变的HEPN结构域以消除靶RNA切割)靶向双色蓝至红剪接报告子的切片受体位点,以介导外显子排斥(如通过mTagBFP2表达的降低所测量的)。培养293FT细胞,将其瞬时转染,并在三天之后,使用流式细胞术检测CD81表达,如实施例19所述。

[0726] 如图22所示,即使是具有高水平内源性蛋白质敲低(>70%CD81蛋白质敲低)的直向同源物,剪接调节在一些情形中也是不足的。但是,dCasRx (SEQ ID NO:280)、dCasRx-del113 (SEQ ID NO:285)和dGa7274 (SEQ ID NO:198)均表现出良好的剪接活性。这三个序列比较地密切相关,享有>70%的序列同一性。因此,可以使用Cas13d直向同源物例如CasRx和截短的CasRx变体如CasRx-del113来调节剪接。

[0727] 实施例23

[0728] 使用Cas13d进行RNA编辑

[0729] 该实施例描述使用Cas13d在人细胞中编辑RNA表达。

[0730] 细胞培养

[0731] 在补充有10%FBS (GE Life Sciences)和10mM HEPES的DMEM(4.5g/L葡萄糖)中,在37°C、5%CO₂条件下维持人胚肾(HEK)细胞系293FT(Thermo Fisher)。达到80-90%汇合后,使用TrypLE Express(Life Technologies)解离细胞,并以1:2的比率传代。

[0732] dCas13d-ADAR哺乳动物表达构建物的克隆

[0733] 如图23A中所示,将dCas13b和dCas13d编码序列克隆到哺乳动物表达载体中,该表达载体具有与ADAR2_{DD} (T375G)的N-或C-端融合以及C-端NLS或NES序列(例如,参见SEQ ID NO:286)。将用来评估dCas13-ADAR融合物的A到I RNA编辑的EGFP报告子被改造成在色氨酸57处含有提前终止密码子(W57X),并通过P2A肽接头与mCherry连接。设计靶向该EGFP报告子的向导,其具有可变的间隔子长度,并在距离同向重复序列的指定距离处具有A-C错配(RfxCas13d为5',PspCas13b为3')。

[0734] 人细胞系的瞬时转染

[0735] 在瞬时转染之前,将HEK 293FT细胞以20,000细胞/孔的密度铺板到96-孔板中,并根据制造商的方案,使用Lipofectamine2000 (Life Technologies),在>90%汇合下用500ng具有40ng EGFP (W57X)报告质粒和不同摩尔比的dCas13d-ADAR载体和向导表达载体的总质粒DNA进行转染。转染后48小时收获转染的细胞用于流式细胞术。

[0736] 流式细胞术

[0737] 转染后48小时将细胞用TrypLE Express解离,重悬于FACS缓冲液(1X DPBS^{-/-}, 0.2%BSA,2mM EDTA)中。使用MACSQuant VYB (Miltenyi Biotec)在96孔板格式下进行流式细胞术,并使用FlowJo10进行分析。分析结果代表BFP⁺(表达dCas13d-ADAR-P2A-mTagBFP2)和mCherry⁺(表达mCherry-P2A-EGFP (W47X))群落,通过观察表达EGFP的细胞在基线以上的%对其进行RNA编辑评估。

[0738] 如图23B所示,在GFP报告子中使用Cas13d进行提前终止密码子的靶向A到I的RNA编辑重新激活了GFP表达,而非靶向的向导则没有导致RNA编辑。图23C提供了基于转染到细胞中的dCas13d-ADAR载体的量进行的Cas13d-介导的RNA编辑效率的定量。

[0739] 实施例24

[0740] Cas13-介导的反义寡核苷酸的鉴定

[0741] 由于缺乏遗传编码,ASO产生的费力性质以及对化学修饰的频繁需求,有效反义寡核苷酸(ASO)-结合位点的发现可能是昂贵的低通量过程。Cas13酶及其变体,例如本文公开的Cas13d蛋白质和变体,可以用于简化和加速ASO结合位点的鉴定,因为有催化活性或无催化活性的Cas13/dCas13的结合可以模拟ASO-介导的RNA切割和/或结合的效果。

[0742] 在一实例中,例如,以逐个孔或汇集寡核苷酸的格式生成Cas13向导RNA文库,并将其设计成靶向所需RNA的特定顺式元件(例如,拼接治疗上重要的蛋白质例如肌萎缩蛋白或SMN1/SMN2的pre-mRNA的内含子,或者UTR或基因体内的所需位置,以鉴定有效的位点可接近的ASO结合位点)。该方法可以用于ASO先导寻找、体内RNA二级结构的质询以及已知或新的剪接调节元件的鉴定。

[0743] 实施例25

[0744] 用Cas13增加翻译

[0745] 本文公开的Cas13d酶和变体可以用于通过靶向操纵转录后调节过程来增加靶蛋白的翻译。上游开放阅读框(uORF)的Cas13d介导的抑制(例如经由空间阻断或其它机制)可以用于实现这种操纵。例如,使用本文公开的Cas13d酶和变体的Cas13d介导的抑制可以充当人或哺乳动物基因翻译的替代起始位点,或靶基因的UTR内的翻译抑制基序和元件。通过抑制uORF,降低的核糖体竞争和其它机制可以增加靶基因从典型的或所需的起始密码子的翻译。几乎一半的人转录物呈递uORF (Barbosa et al., Plos Genetics 9:e1003529,

2013),其可以导致蛋白质表达的广泛降低。类似地,通过使5'UTR内的抑制元件例如二级结构或G-四链体(G-quadruplexes)失活,可以增加翻译。

[0746] 这可以通过使用向导RNA靶向Cas13d效应子来实现,该Cas13d效应子任选地与可以增加对uORF或翻译抑制元件的翻译抑制的效应子蛋白(例如Pdc4的C-端结构域MA3结构域)融合。所述向导RNA还可以经由多个向导RNA或者pre-gRNA阵列以多重方式递送。这种方法可用于增加治疗或生物技术上有用的蛋白质的翻译,例如家族性高胆固醇血症中的LDLR、囊性纤维化中的CFTR、先天性高胰岛素血症中的KCNJ11、镰状细胞病中的HBB、PRKAR1A、IRF6等(Barbosa et al., Plos Genetics 9:e1003529, 2013; Calvo et al., PNAS 106:7507-12, 2009)。

[0747] 实施例26

[0748] 使用Cas13进行合成致死基因筛选

[0749] 本文公开的Cas13d酶和变体可以用于筛选方法中,以辅助鉴定治疗性化合物,例如通过挖掘先导小分子或抑制性RNA(RNAi)。利用例如肿瘤抑制子或致癌基因的合成致死相互作用,得以鉴定出选择性抑制癌细胞或其他目的细胞的化合物(例如小分子或抑制性RNA)。用于发现与化合物的合成致死基因相互作用的常规高通量筛选、CRISPR-Cas9和RNAi均具有诸多限制,包括未知的或多种靶基因(对于化合物),对脱靶基因的多效性作用(化合物和RNAi),以及不适当的双重功能丧失(对于CRISPR-Cas9或其他基于核酸酶的方法)。

[0750] 提供了使用本文公开的Cas13d酶和变体用于肿瘤学先导发现的方法,以解决药理学导向的合成致死性的“二次命中(second hit)”的模拟中的这些挑战,其出于以下原因: 1) 其靶向已知靶基因或基因组的特异性和2) 其敲低机制,这更准确地反映小分子先导的抑制机制。例如,Cas13d或其它Cas13酶可以与同源的向导RNA一起引入到携带特定突变的细胞(例如KRAS突变肺癌细胞、BRCA-null卵巢癌细胞和许多其它细胞)中。然后可以以阵列格式或向导RNA-汇集的遗传筛选进行生长抑制或细胞损失的表型读出,随后进行遗传分析,以鉴定目的基因。

[0751] 参考文献

[0752] Abudayyeh(2017).RNA targeting with CRISPR-Cas13.Nature.550:280-4.

[0753] Abudayyeh et al.,(2016).Science 353,aaf5573.

[0754] Almeida et al.,(2012).Cell Rep 2,789-798.

[0755] Anantharaman et al.,(2013)Biol Direct 8,15.

[0756] Batra et al.,(2017).Cell 170,899-912e810.

[0757] Birmingham et al.,(2006)Nat Methods 3,199-204.

[0758] Biswas et al.,(2016)Stem Cell Reports 7,316-324.

[0759] Bland et al.,(2007).BMC Bioinformatics 8,209.

[0760] Boeve,B.F.,and Hutton,M.(2008).Arch Neurol 65,460-464.

[0761] Cheong,C.G.,and Hall,T.M.(2006).Proc Natl Acad Sci U S A 103,13635-13639.

[0762] Chiriboga et al.,(2016).Neurology 86,890-897.

[0763] Chylinski et al.,(2013).RNA Biol 10,726-737.

[0764] Cox et al.,(2017).Science 358,1019-1027.

- [0765] Deltcheva et al., (2011). *Nature* 471,602-607.
- [0766] Dobin et al., (2013). *Bioinformatics* 29,15-21.
- [0767] Doench et al., (2003). *Genes Dev* 17,438-442.
- [0768] Doudna, J.A., and Charpentier, E. (2014). *Science* 346,1258096.
- [0769] Du et al., (2006). *Mol Cell Proteomics* 5,1033-1044.
- [0770] East-Seletsky et al., (2017). *Mol Cell* 66,373-383e373.
- [0771] East-Seletsky et al., (2016). *Nature* 538,270-273.
- [0772] Edgar, R.C. (2007). *BMC Bioinformatics* 8,18.
- [0773] Fonfara et al., (2016). *Nature* 532,517-521.
- [0774] Gasiunas et al., (2012). *Proc Natl Acad Sci U S A* 109,E2579-2586.
- [0775] Gilbert et al., (2014). *Cell* 159,647-661.
- [0776] Gilbert et al. (2013). *Cell* 154,442-451.
- [0777] Grissa, I., Vergnaud, G., and Pourcel, C. (2007). *Nucleic Acids Res* 35,W52-57.
- [0778] Hammond, S.M., and Wood, M.J. (2011). *Trends Genet* 27,196-205.
- [0779] Heidrich et al., (2015). *Methods Mol Biol* 1311,1-21.
- [0780] Heinz et al., (2010). *Mol Cell* 38,576-589.
- [0781] Hsu, P.D., Lander, E.S., and Zhang, F. (2014). *Cell* 157,1262-1278.
- [0782] Jackson et al., (2003). *Nat Biotechnol* 21,635-637.
- [0783] Jinek et al., (2012). *Science* 337,816-821.
- [0784] Ka et al., (2005) *Alzheimer Dis Assoc Disord* 19 Suppl 1,S29-36.
- [0785] Kim et al. (2017). *Nat Commun* 8,14500.
- [0786] Liu et al., (2017). *Cell* 168,121-134 e112.
- [0787] Love, M.I., Huber, W., and Anders, S. (2014). *Genome Biol* 15,550.
- [0788] Makarova et al. (2015). *Nat Rev Microbiol* 13,722-736.
- [0789] Matera, A.G., and Wang, Z. (2014). *Nat Rev Mol Cell Biol* 15,108-121.
- [0790] Naldini, L. (2015). Gene therapy returns to centre stage. *Nature* 526,351-360.
- [0791] O'Connell et al., (2014). *Nature* 516,263-266.
- [0792] Orengo, J.P., Bundman, D., and Cooper, T.A. (2006). *Nucleic Acids Res* 34, e148.
- [0793] Peabody, D.S. (1993). *EMBO J* 12,595-600.
- [0794] Ran et al., (2015). *Nature* 520,186-191.
- [0795] Samai et al., (2015). *Cell* 161,1164-1174.
- [0796] Schena et al., (1995). *Science* 270,467-470.
- [0797] Schoch et al., (2016). *J Neurosci* 36,941-947.
- [0798] Shendure et al., (2017). *Nature* 550:345-53.
- [0799] Shmakov et al., (2015) *Mol Cell* 60,385-397.
- [0800] Sigoillot et al., (2012). *Nat Methods* 9,363-366.

- [0801] Smargon et al., (2017) .Mol Cell 65,618-630 e617.
- [0802] Smith et al., (2017) .PLoS Biol 15,e2003213.
- [0803] Swiech et al., (2015) .Nat Biotechnol 33,102-106.
- [0804] Treu et al., (2016) .Bioresour Technol 216,260-266.
- [0805] van der Oos et al., (2014) ..Nat Rev Microbiol 12,479-492.
- [0806] Wang et al., (2015) .Biomed Rep 3,152-158.
- [0807] Wright,A.V.,and Doudna,J.A. (2016) .Nat Struct Mol Biol 23,876-883.
- [0808] Yang et al., (2010) .J Proteome Res 9,1402-1415.
- [0809] Yosef et al., (2012) .Nucleic Acids Res 40,5569-5576.
- [0810] Zalatan et al., (2015) .Cell 160,339-350.
- [0811] Zetsche et al.,. (2015) .Cell 163,759-771.
- [0812] Zetsche,et al., (2017) .A Survey of Genome Editing Activity for 16 Cpf1 orthologs.bioRxiv.doi:https://doi.org/10.1101/134015
- [0813] Zhang et al., (2016) .M Nucleic Acids Res 44,1789-1799.
- [0814] Zhang et al., (2013) .RNeuron 78,785-798.
- [0815] 鉴于本公开内容的原理可以应用于许多可能的实施方式,应当认识到,所阐述的实施方式仅是本发明的实例,并不应当理解成限定本发明的范围。相反,本发明的范围由所附权利要求限定。因此,我们要求保护所有落入这些权利要求的范围和精神内的发明。

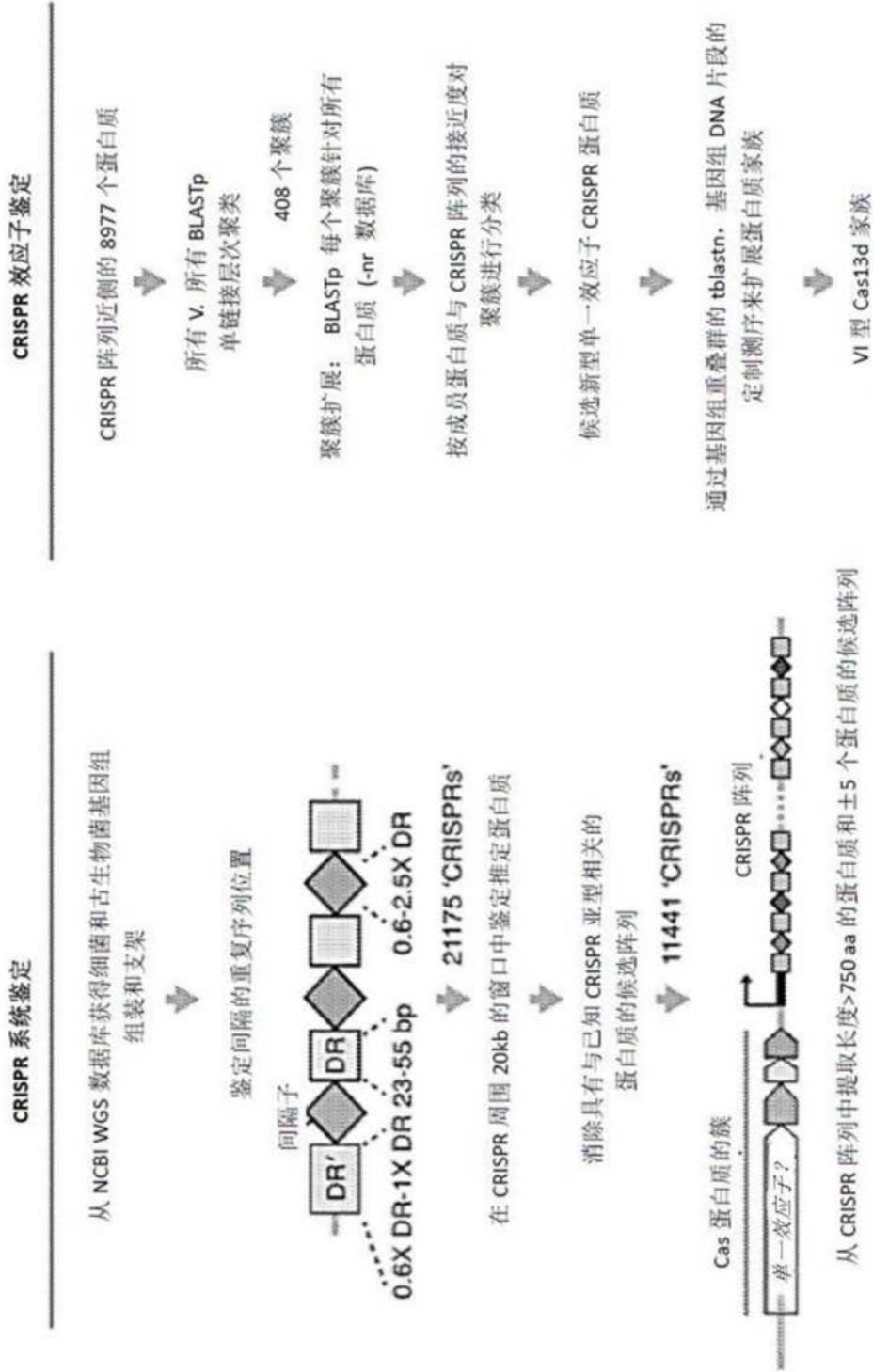
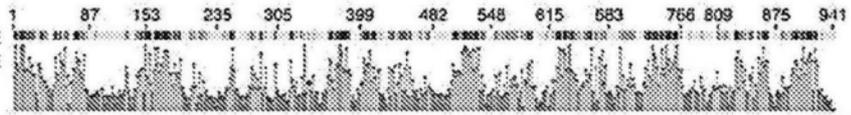


图1A

签名Cas13d
直向同源物

一致性
同一性



P1E0

Adm

Ra

Es

Rfx

Rff

Ur

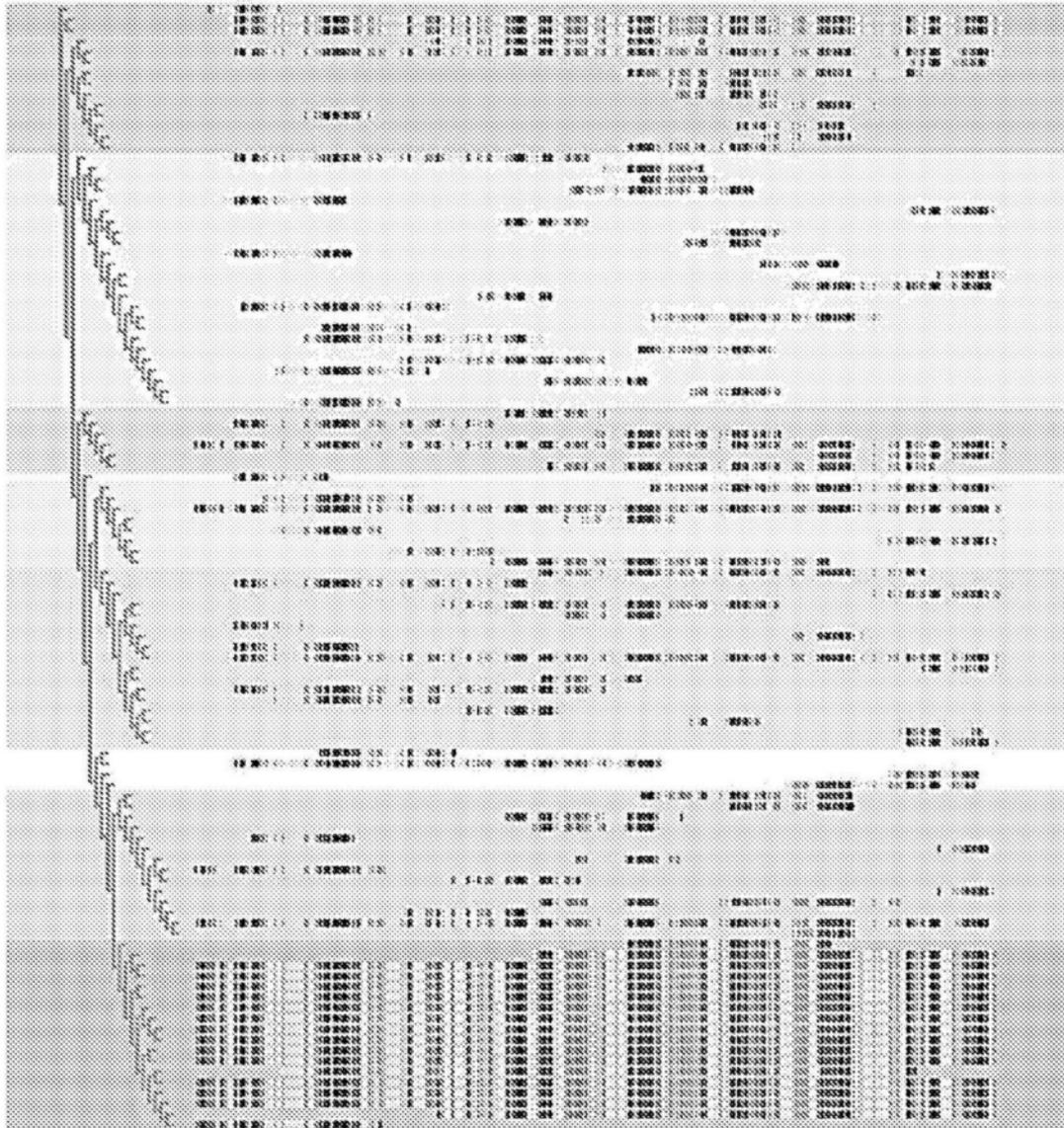


图1B

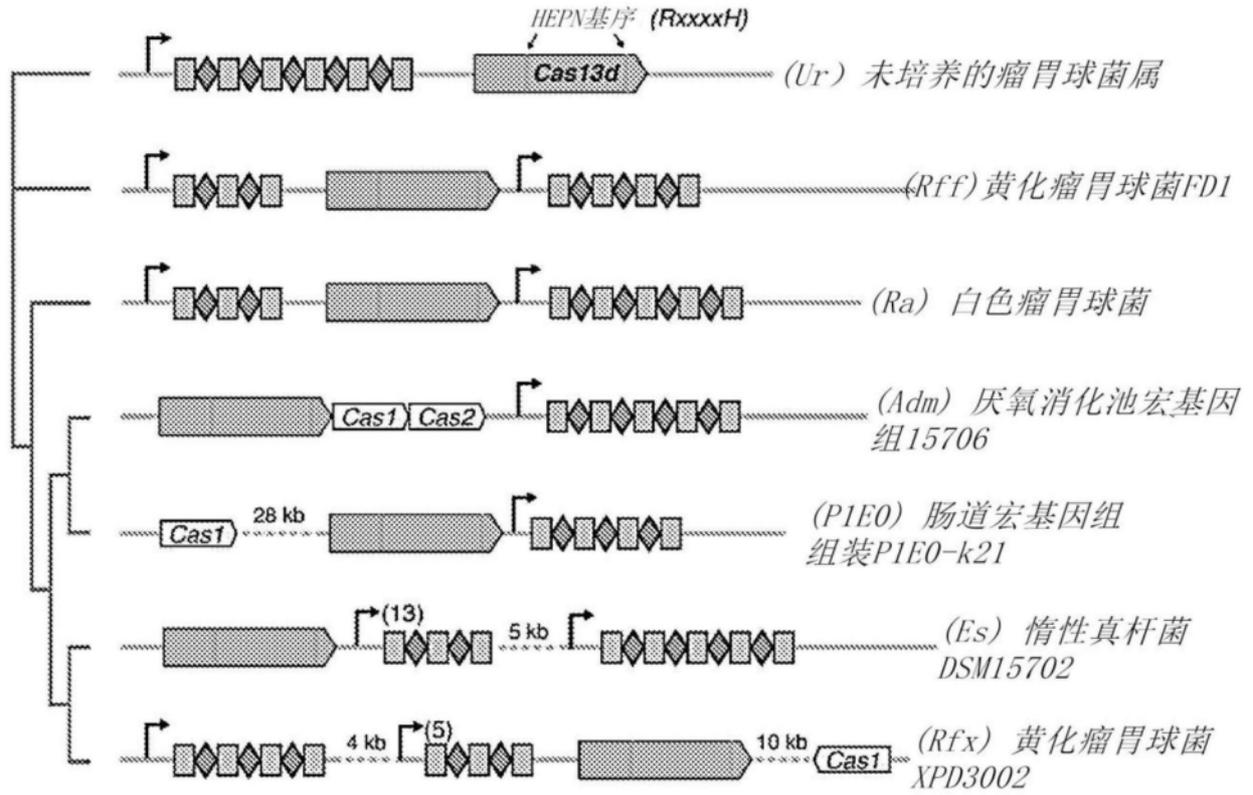


图2A

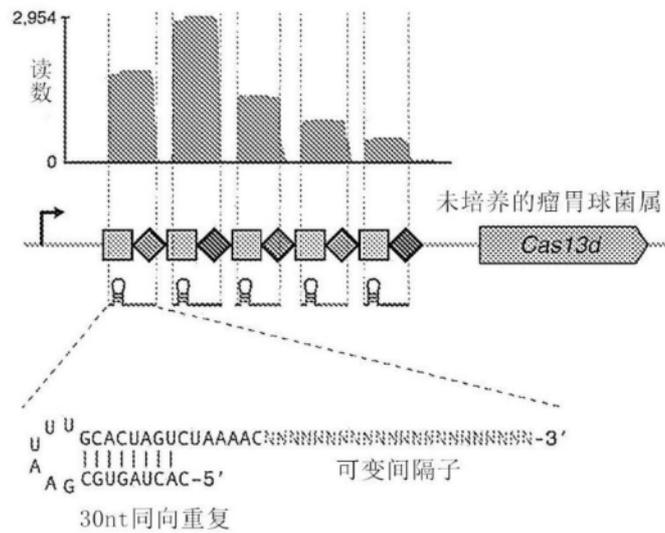


图2B

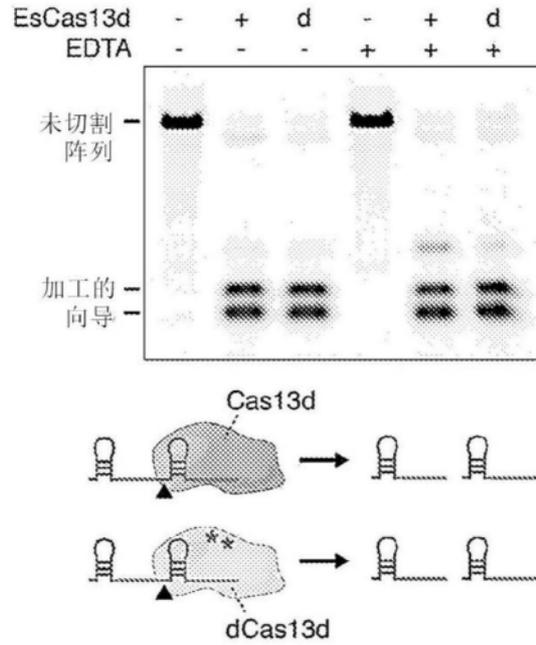


图2C

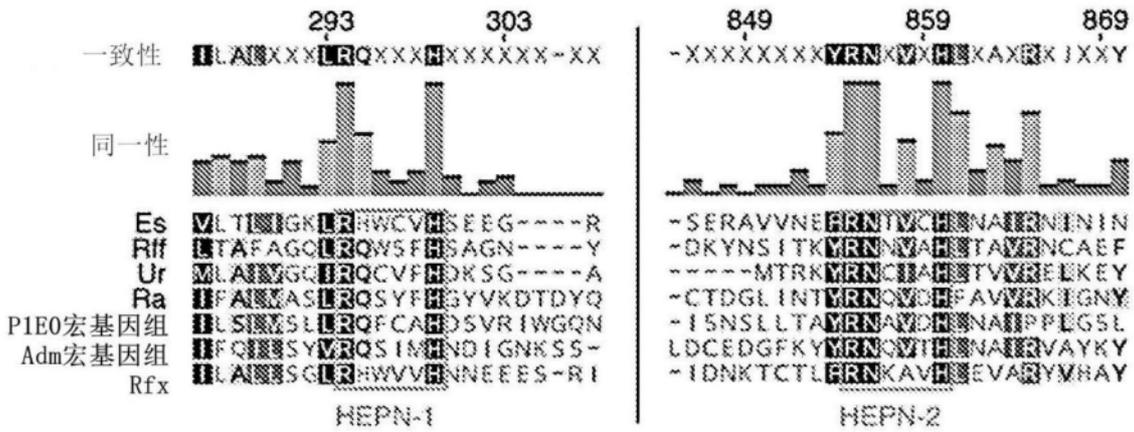


图3A

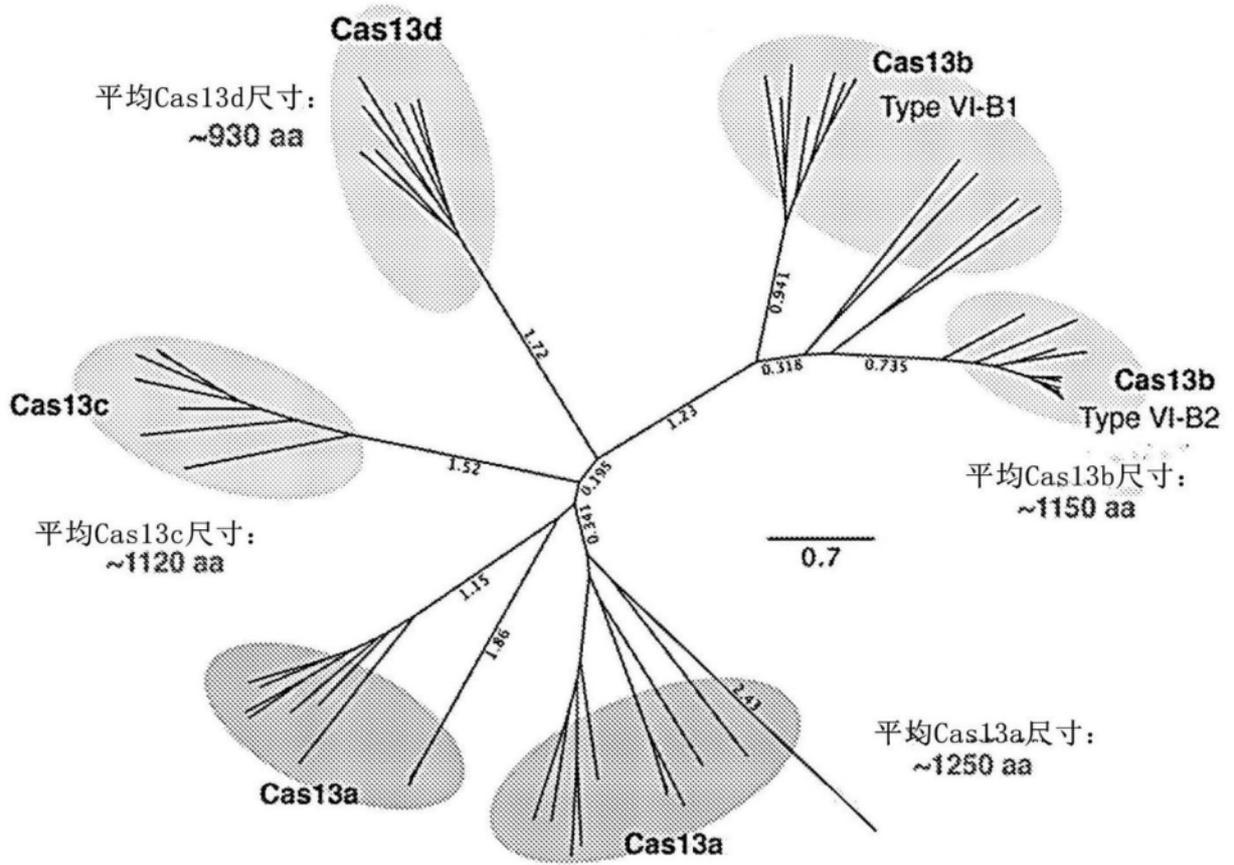


图3B

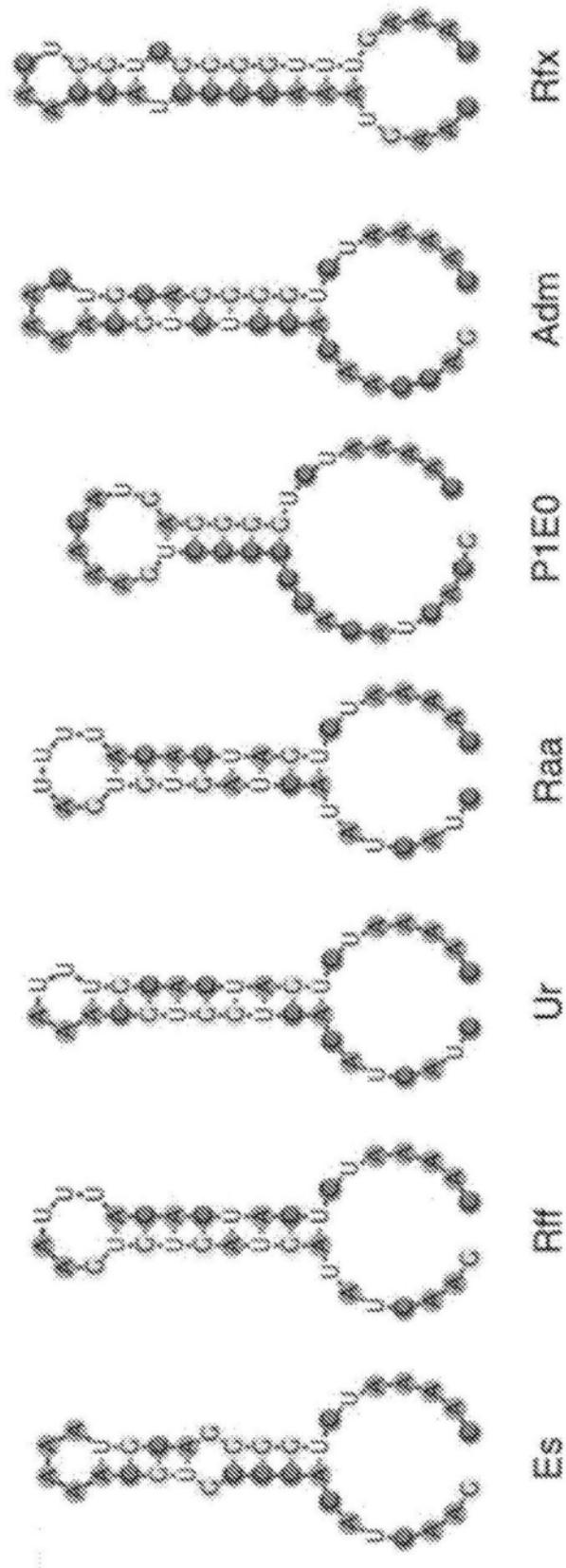


图3C

Cas13d同向重复的序列图标

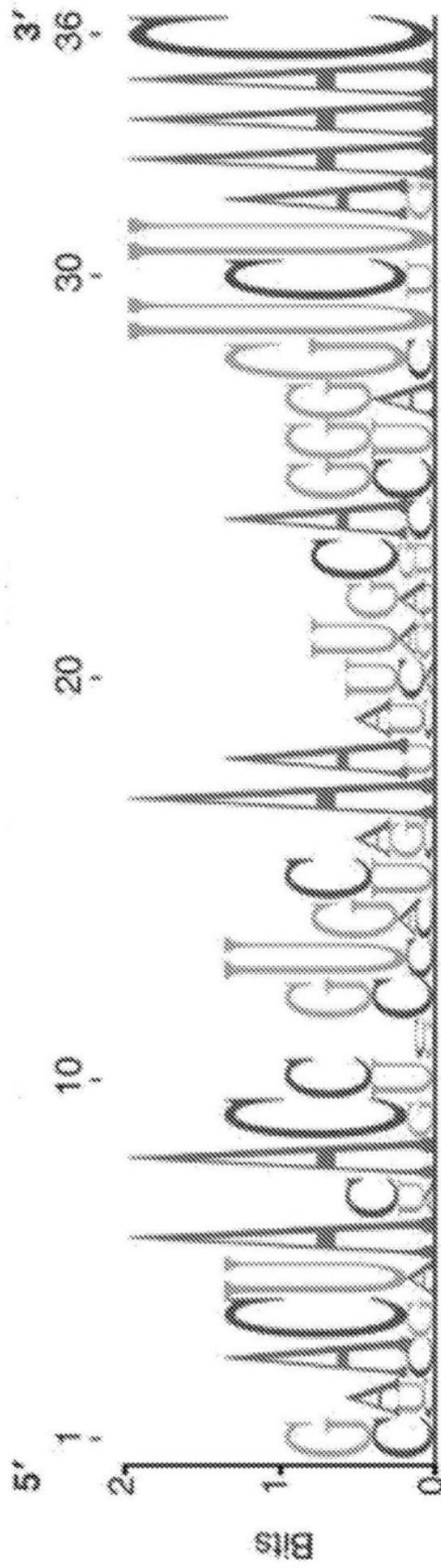


图3D

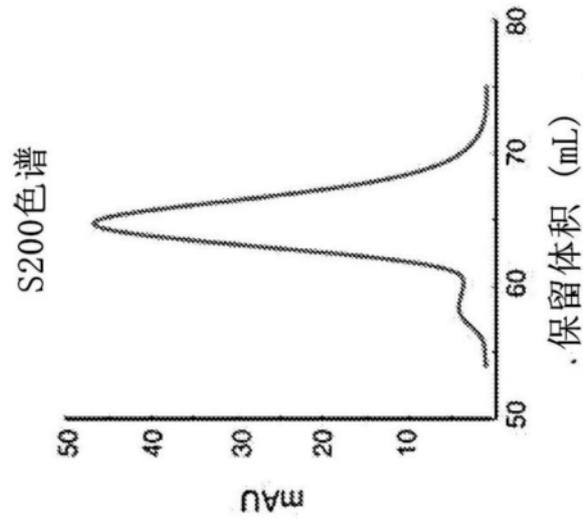


图4A

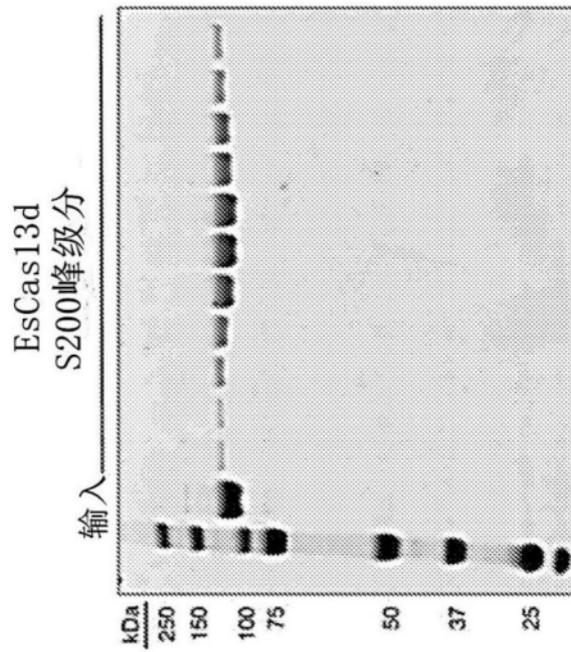


图4B

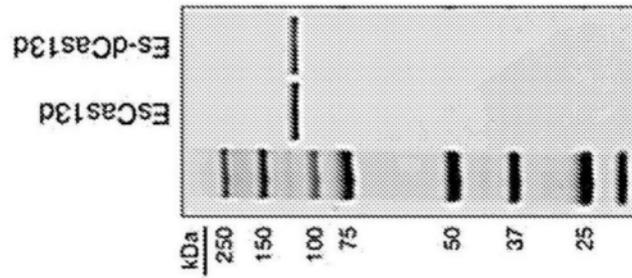


图4C

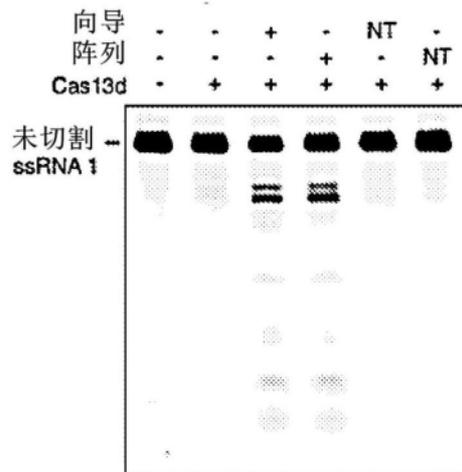


图5A

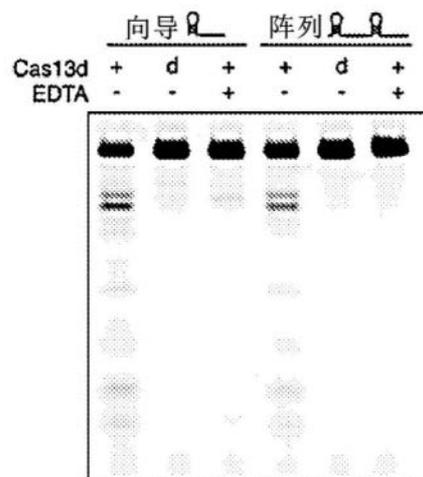


图5B

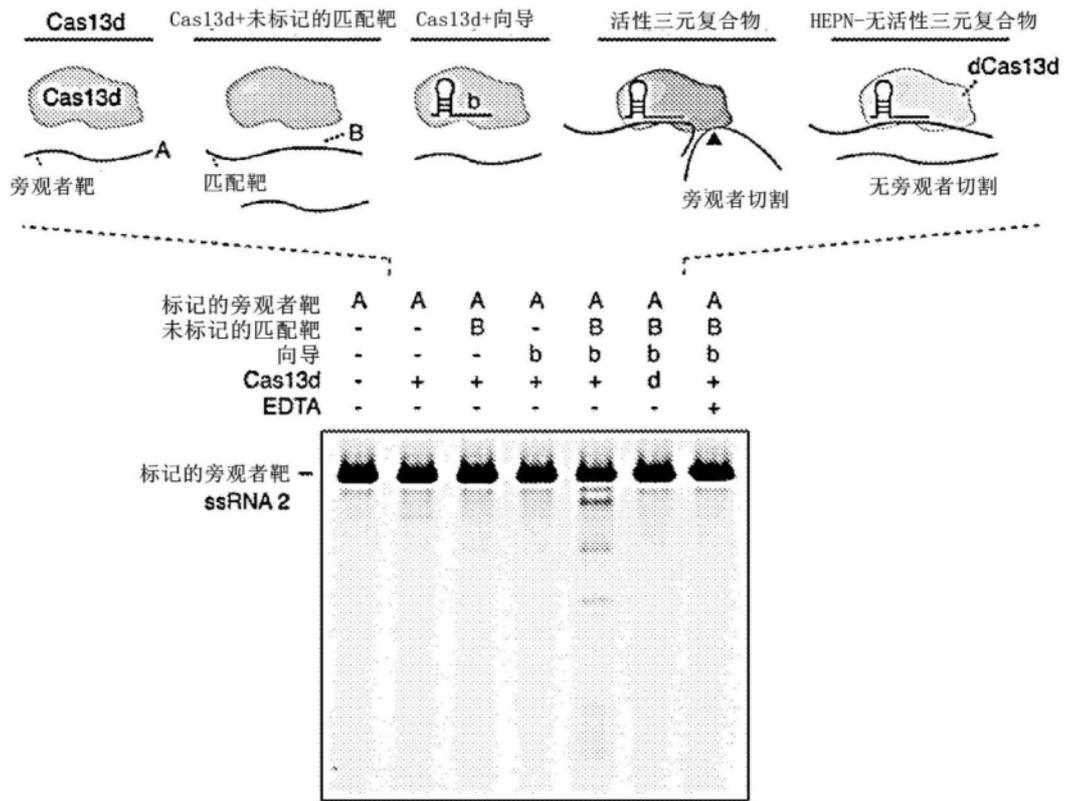


图5C

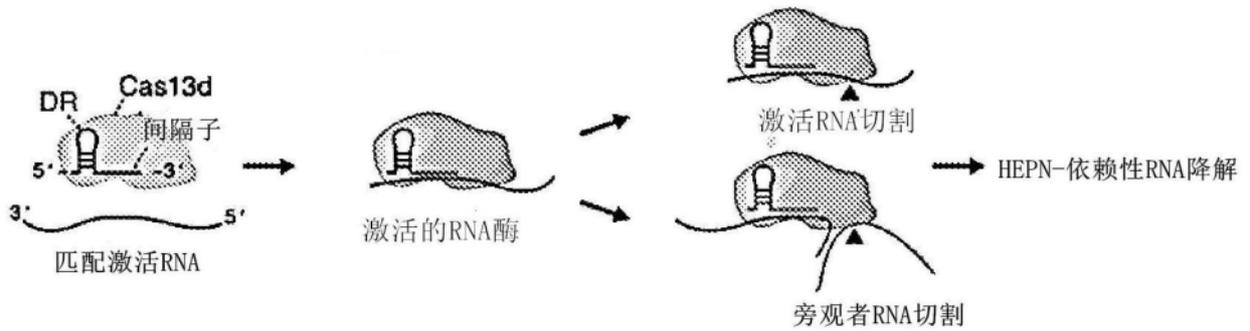


图5D

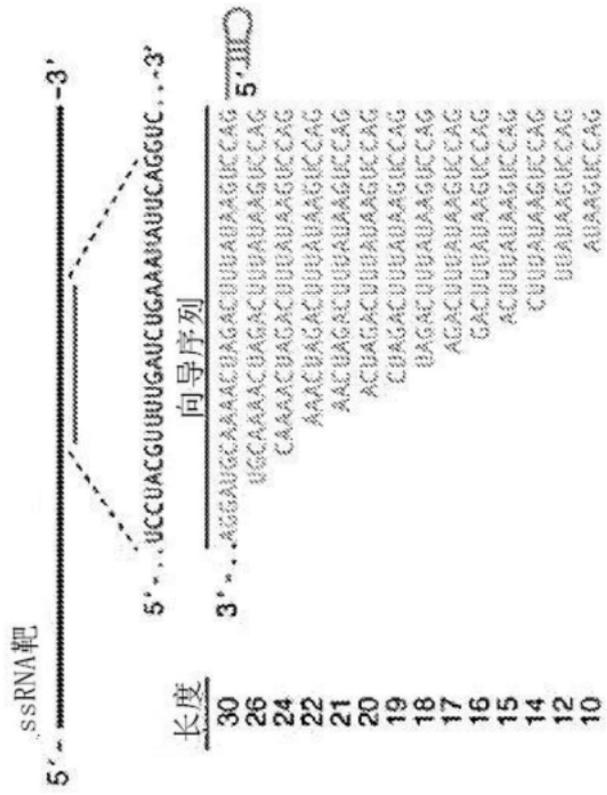


图6A

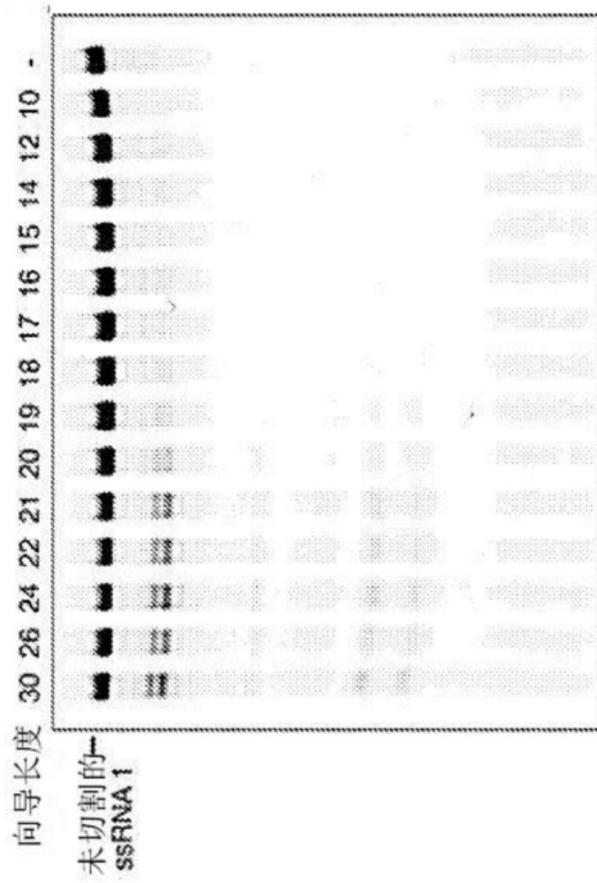


图6B

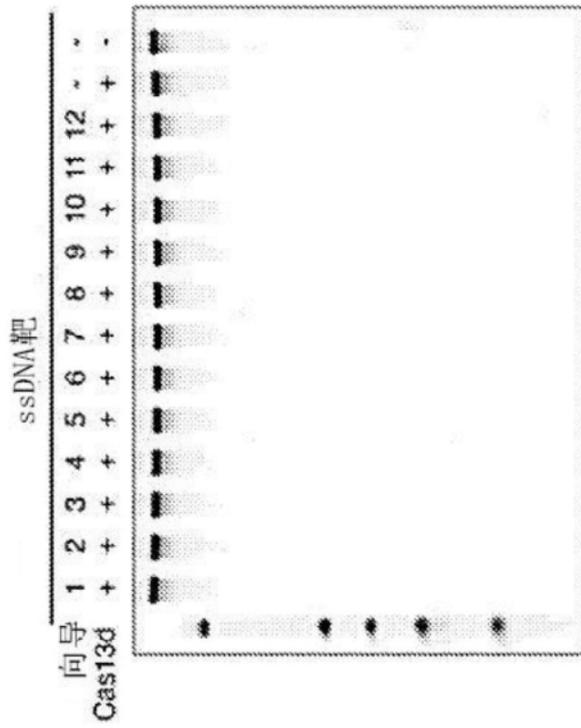


图6C

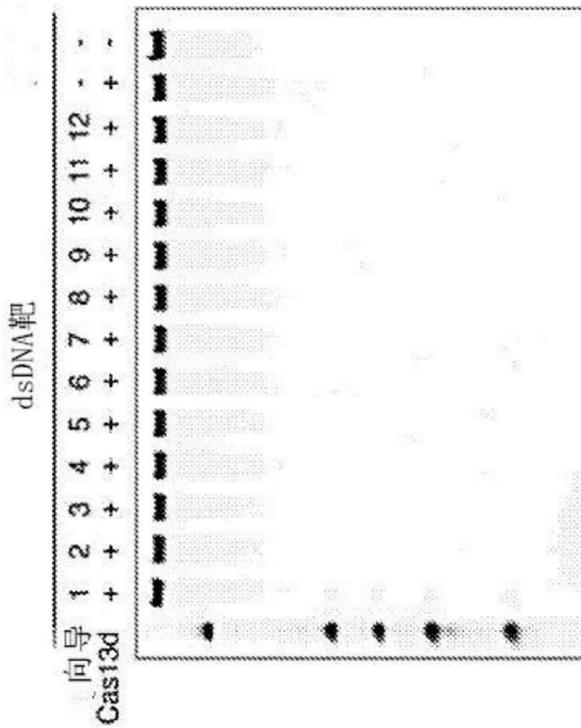
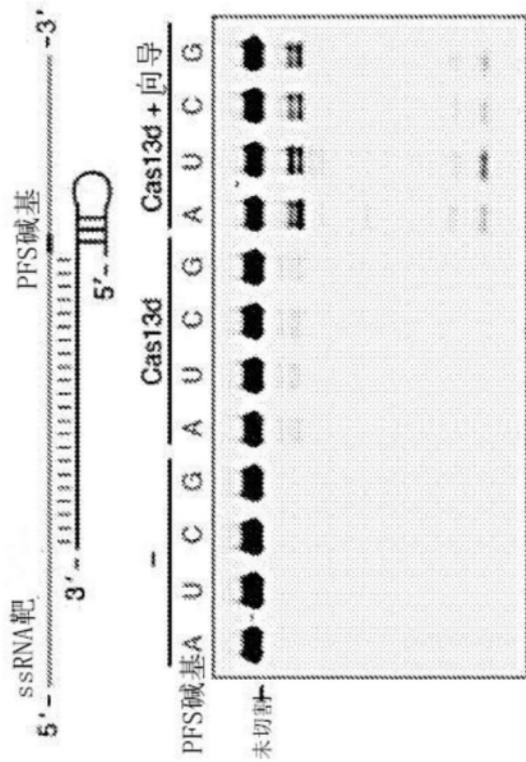
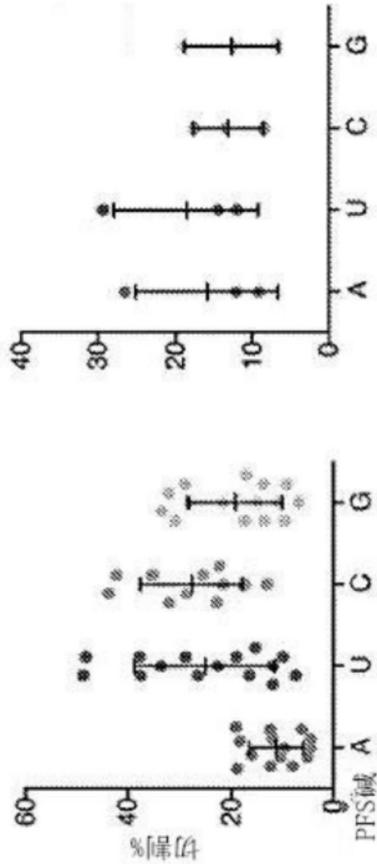


图6D



靶PFS碱基偏好性



向导PFS碱基偏好性

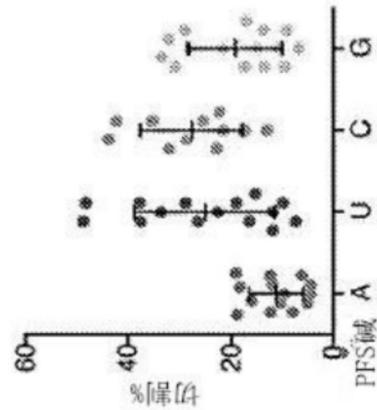


图 6E

图 6F

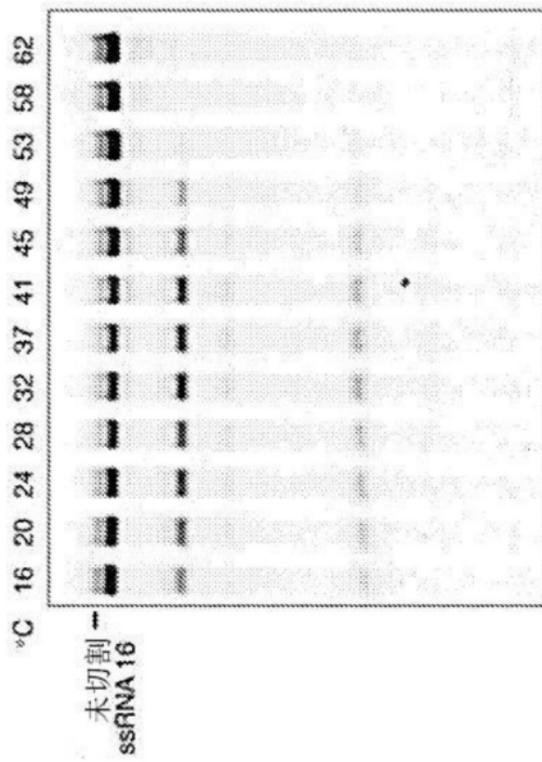


图 6H

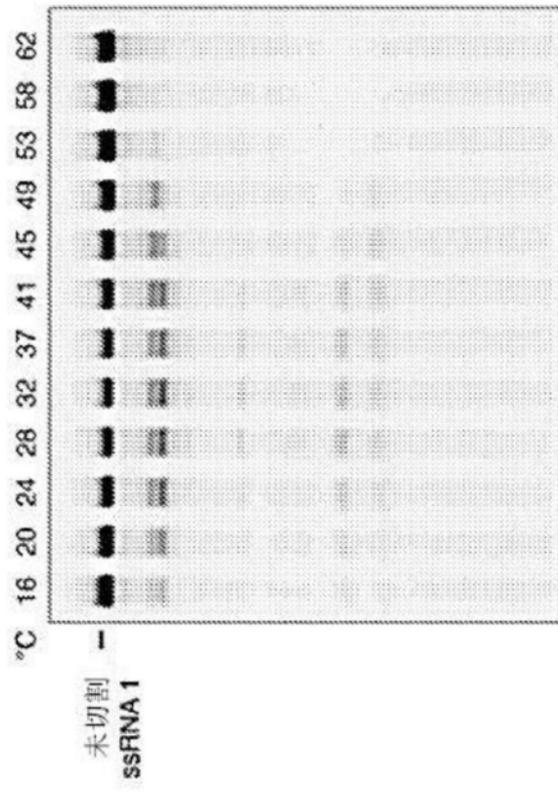


图 6G

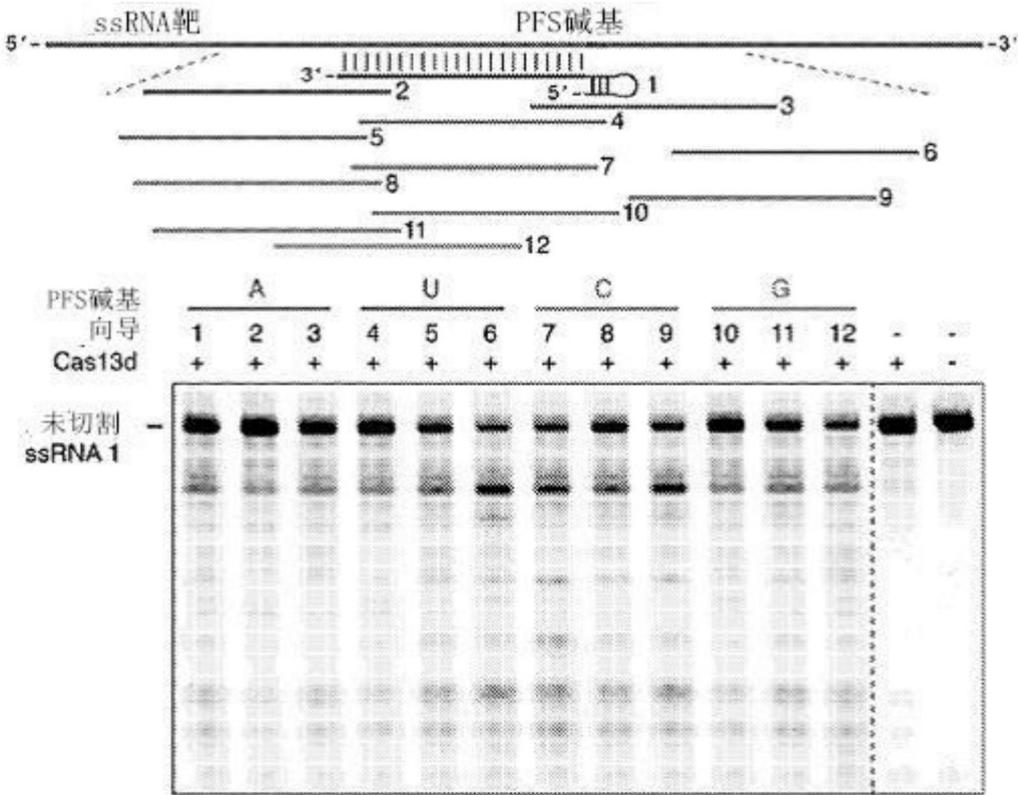


图7A

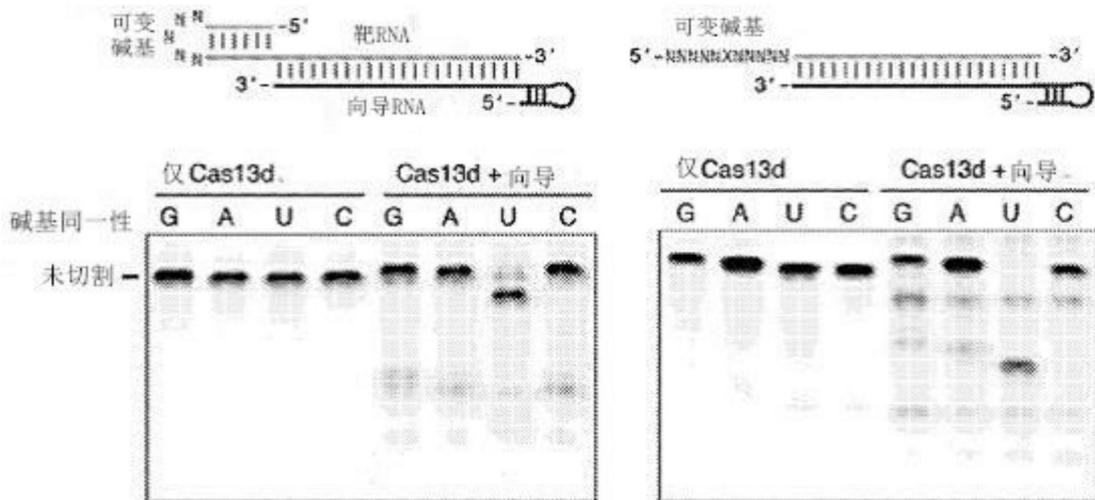


图7B

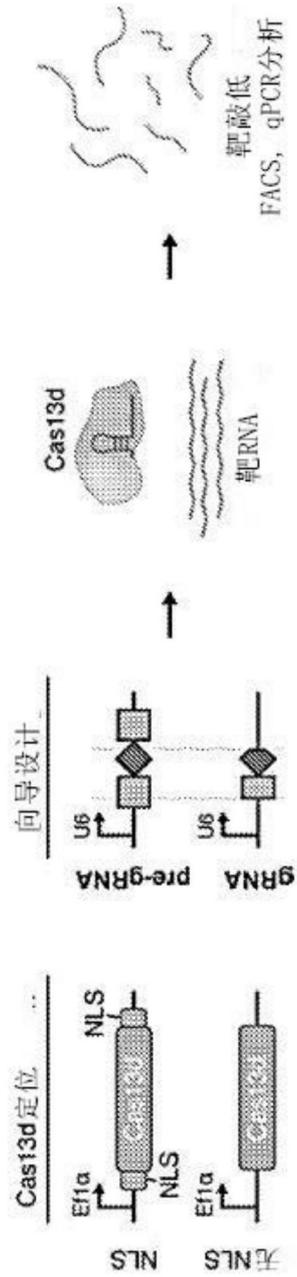


图8A

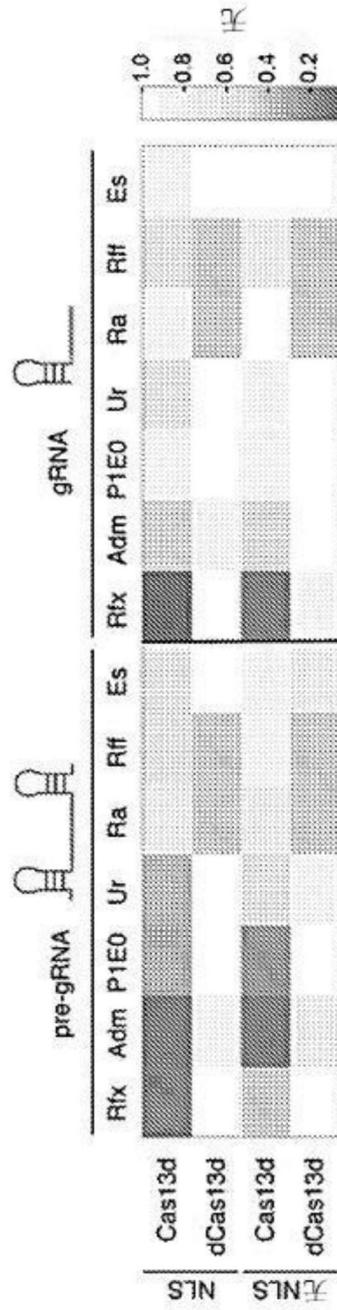


图8B

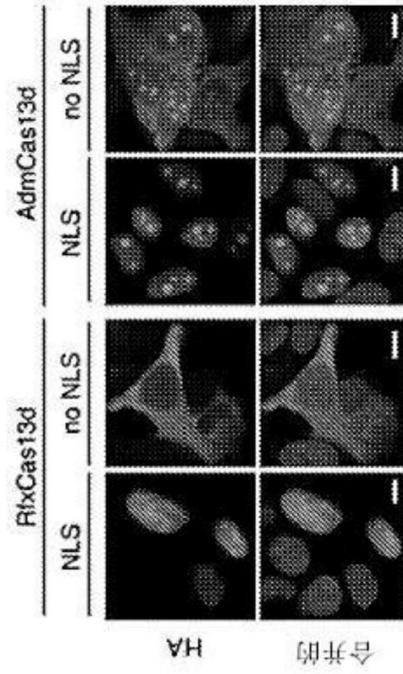


图8C

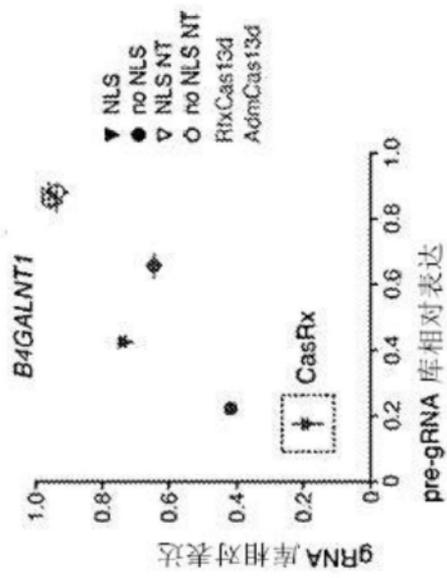


图8D

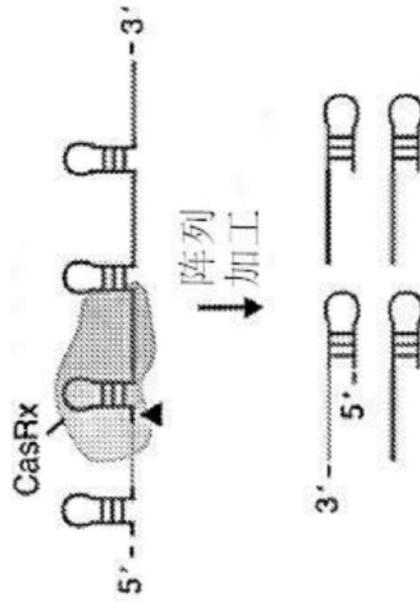


图9A

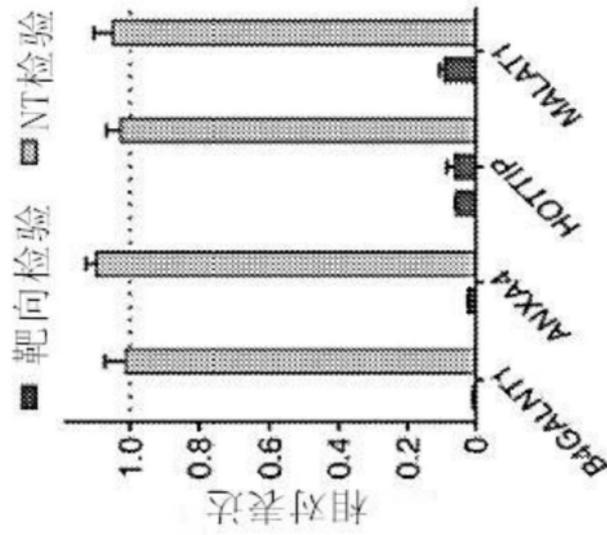


图9B

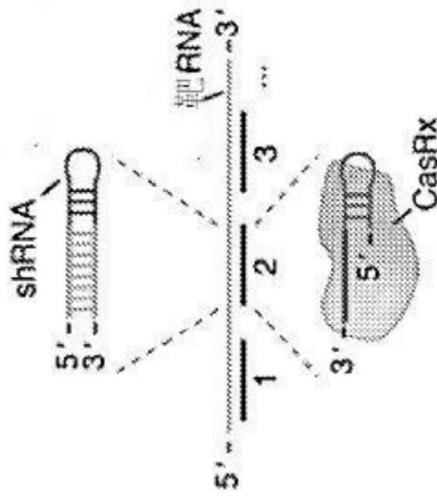


图9C

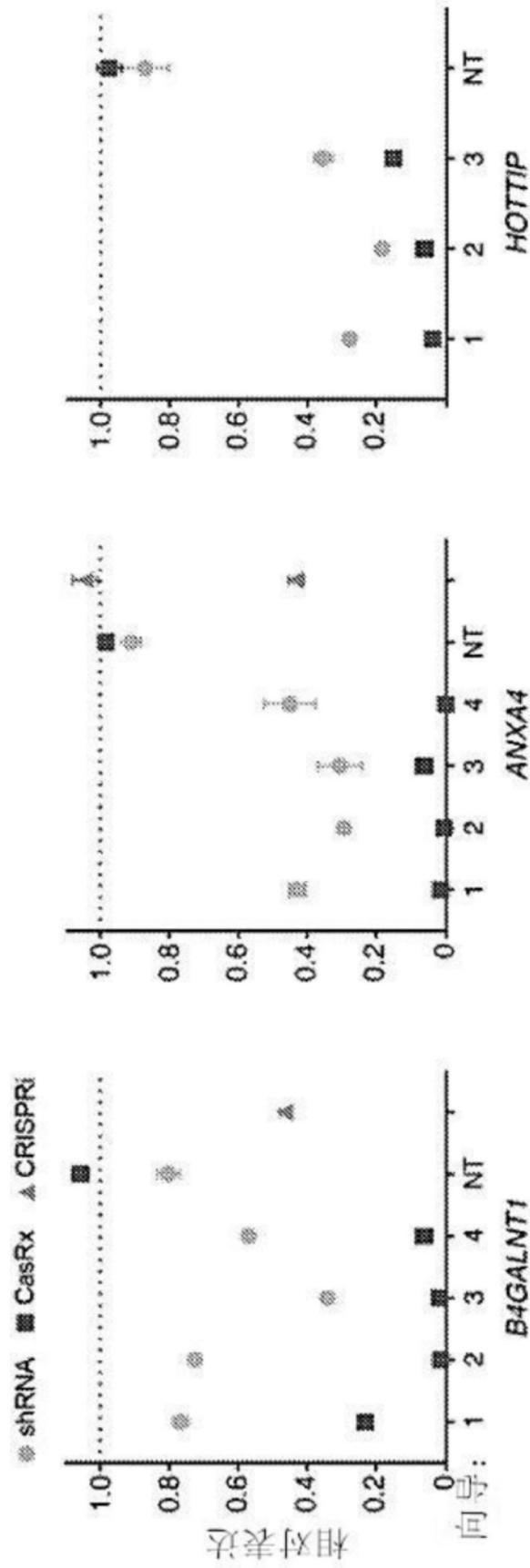


图9D

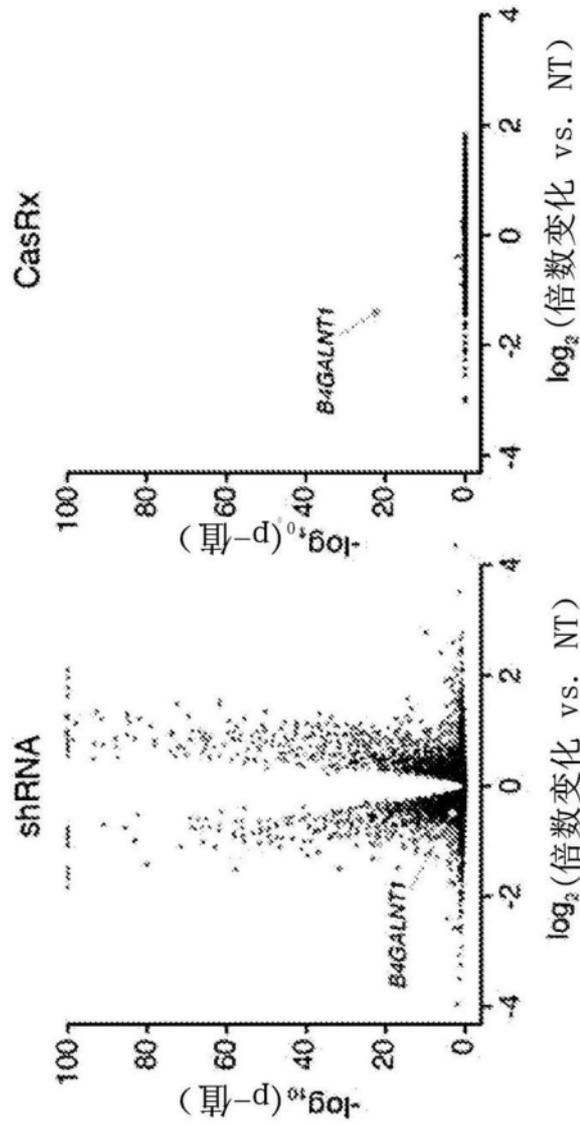


图 9F

图 9E

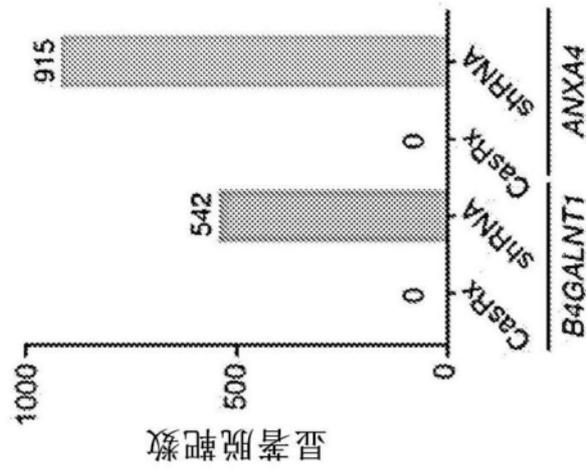


图9G

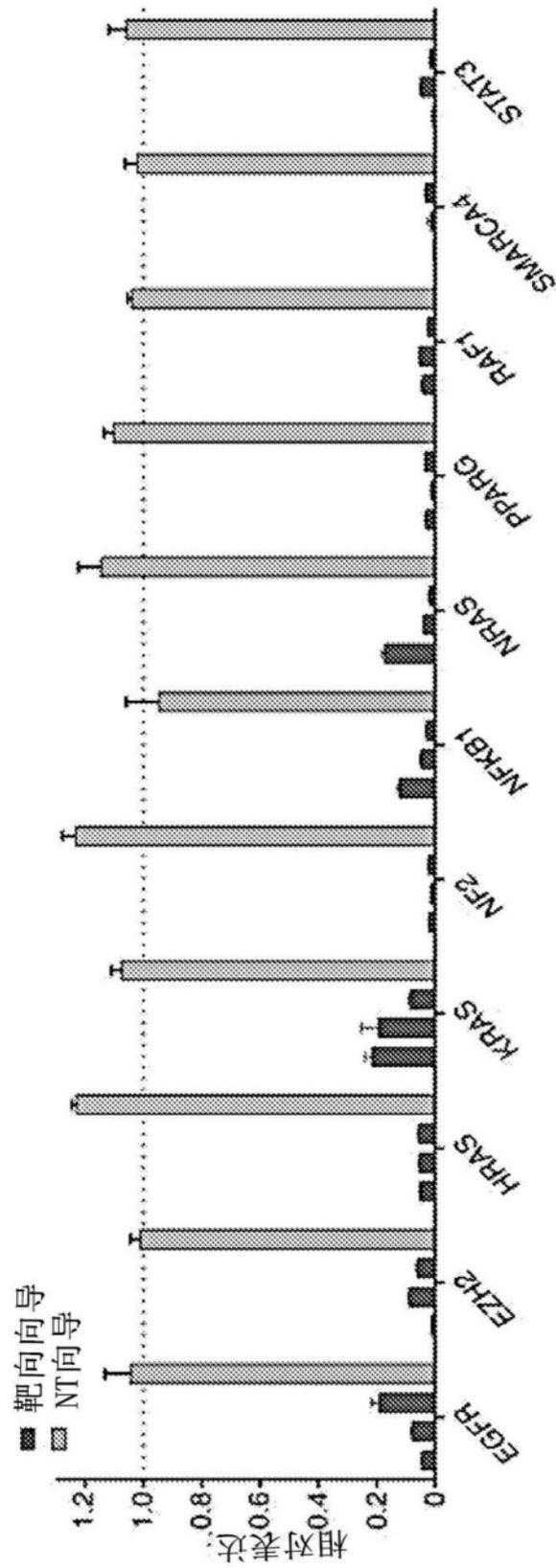


图9H

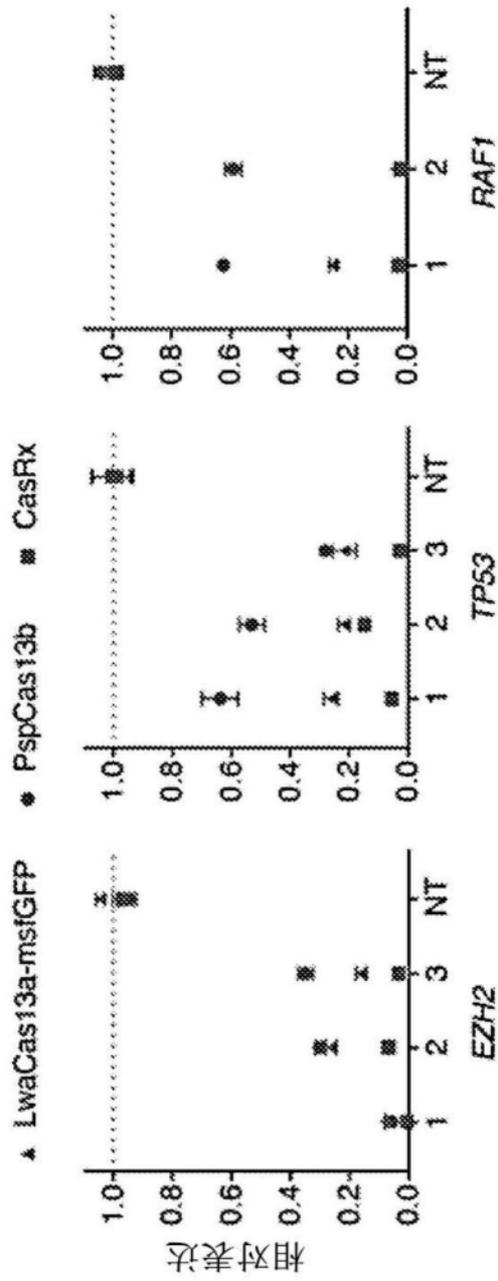


图10A

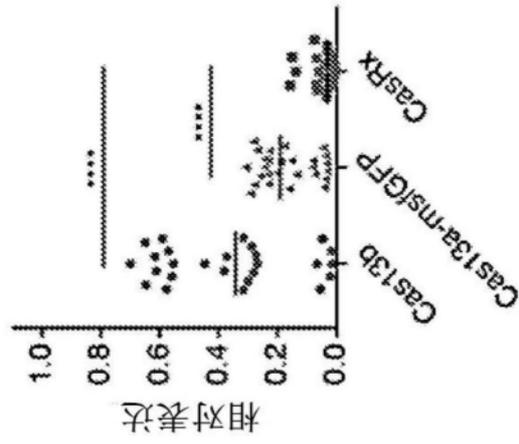


图10B

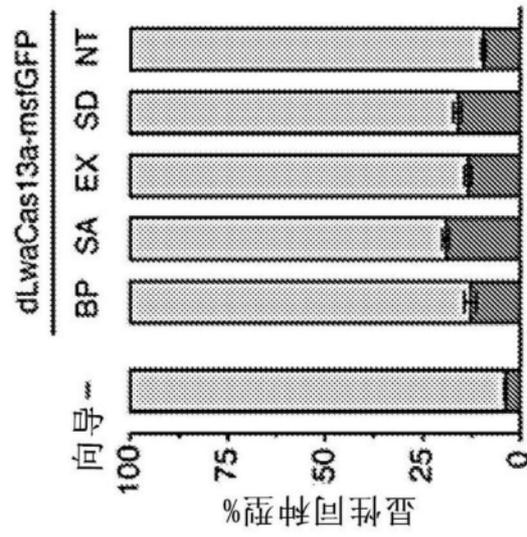


图10C

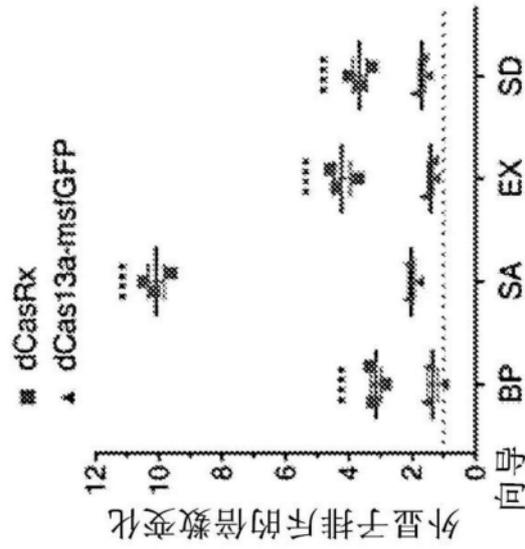


图10D

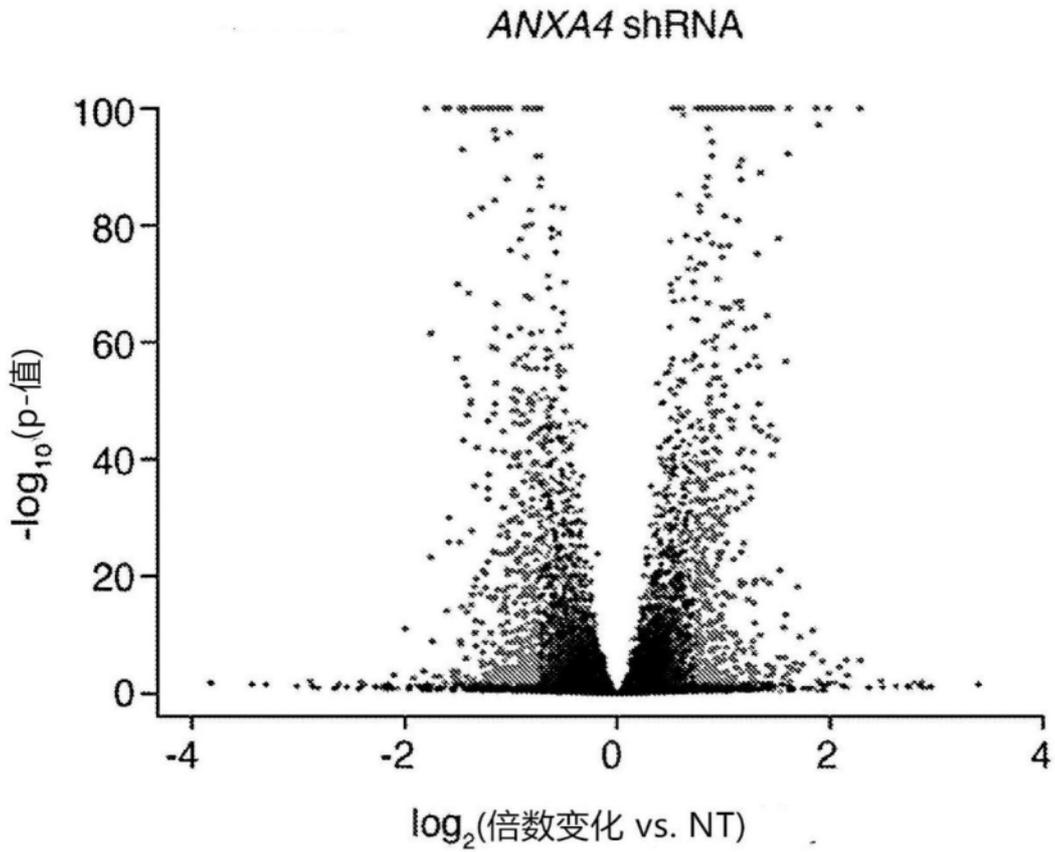


图11A

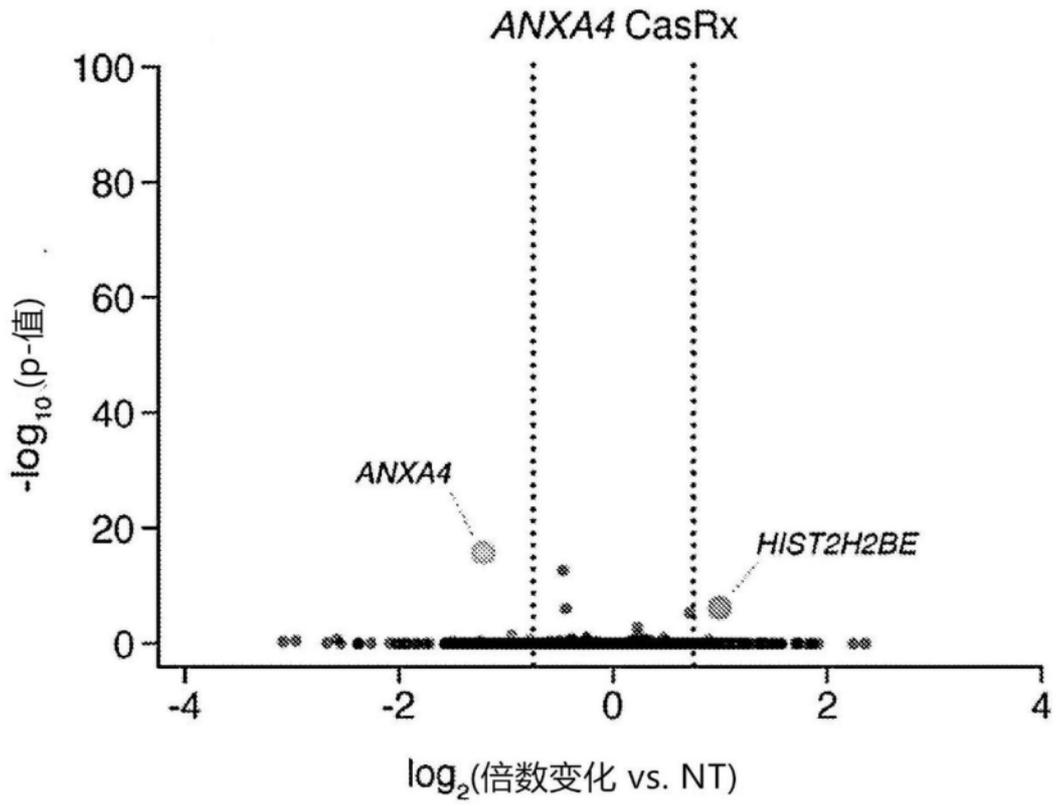


图11B

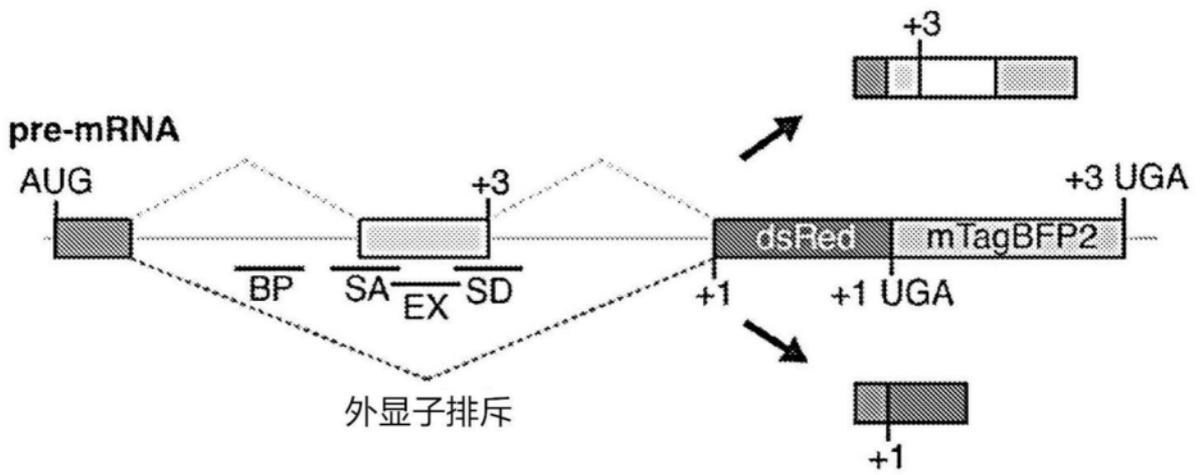


图12A

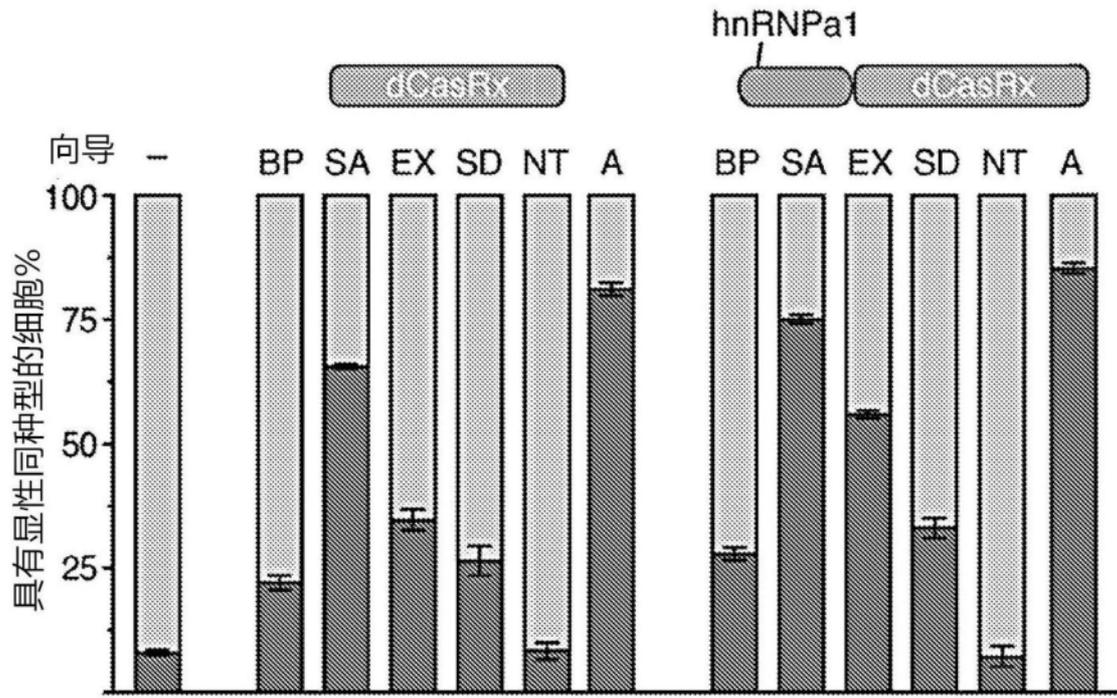


图12B

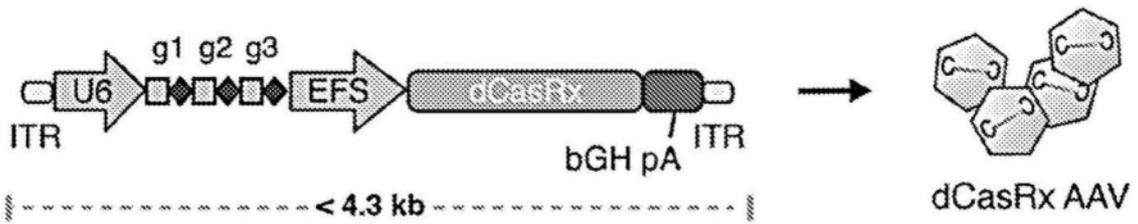


图12C

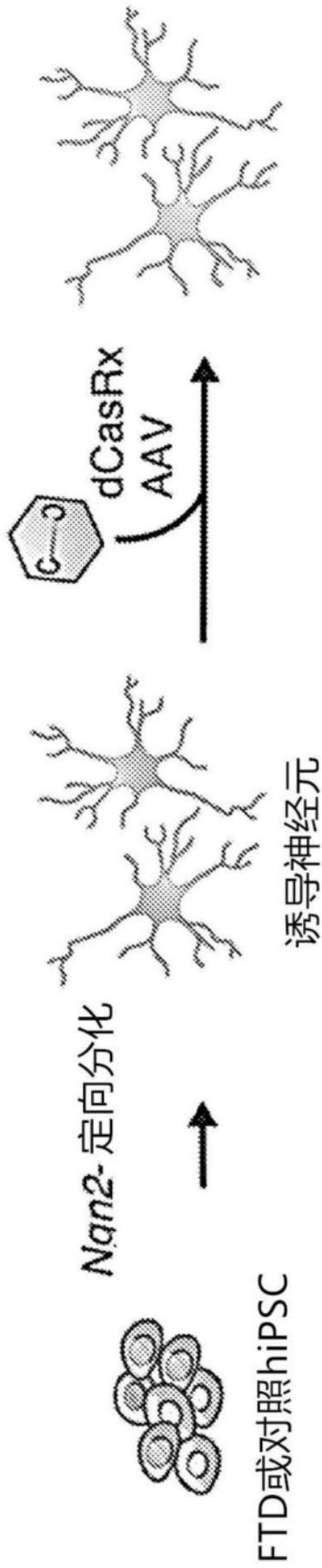


图12D

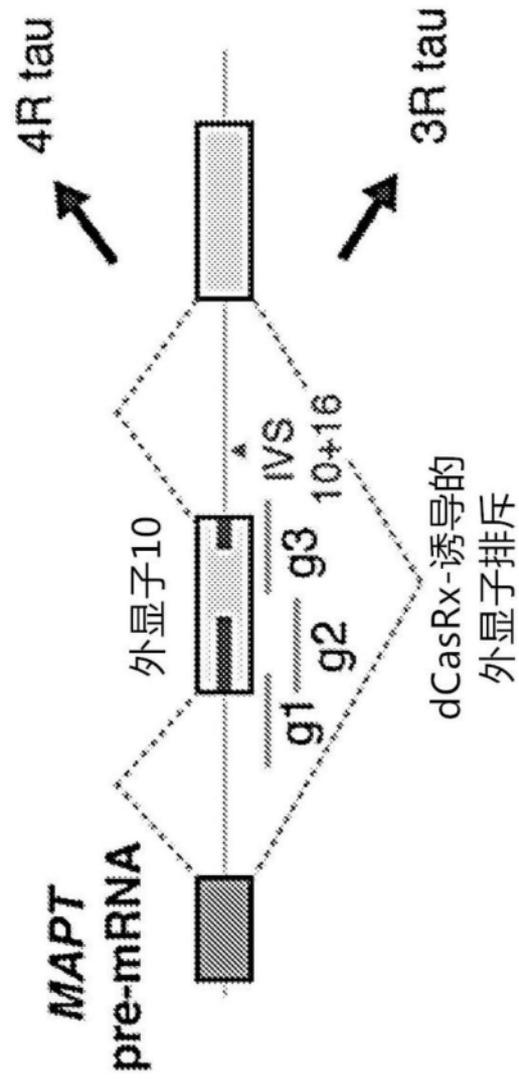


图12E

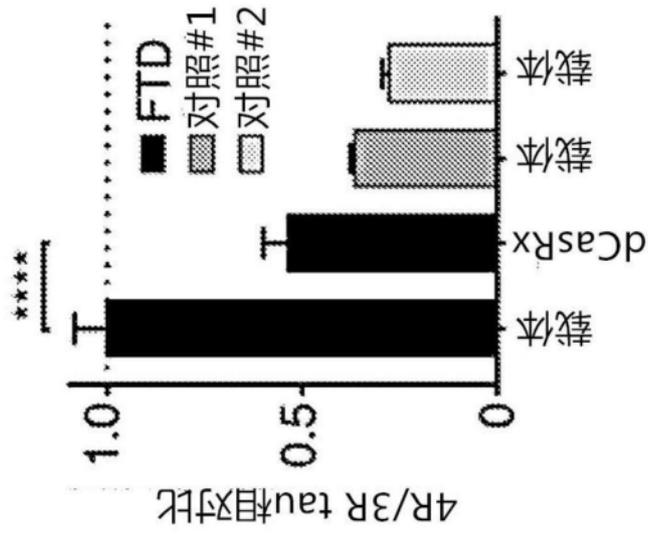


图12F

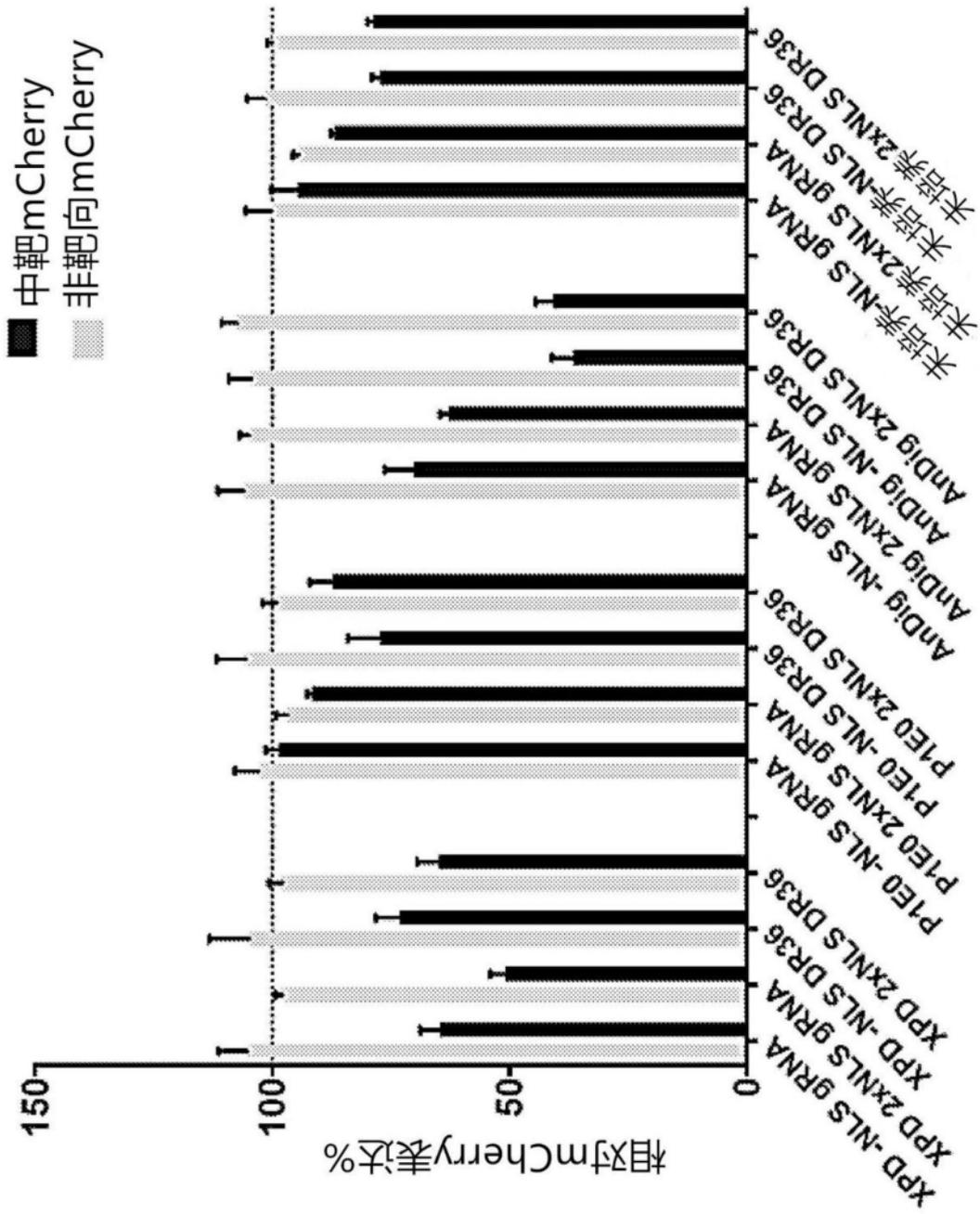


图13

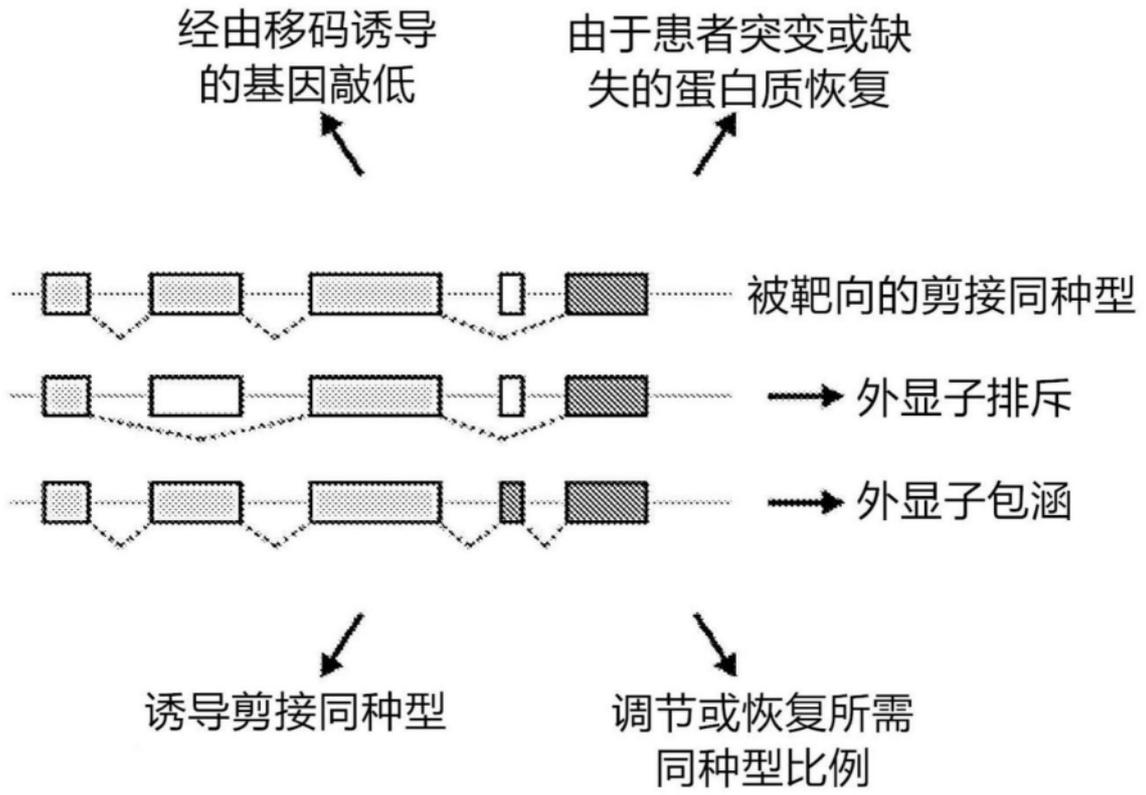


图14

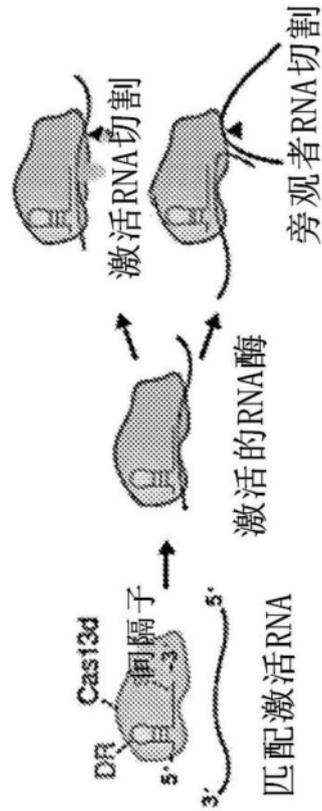


图15A

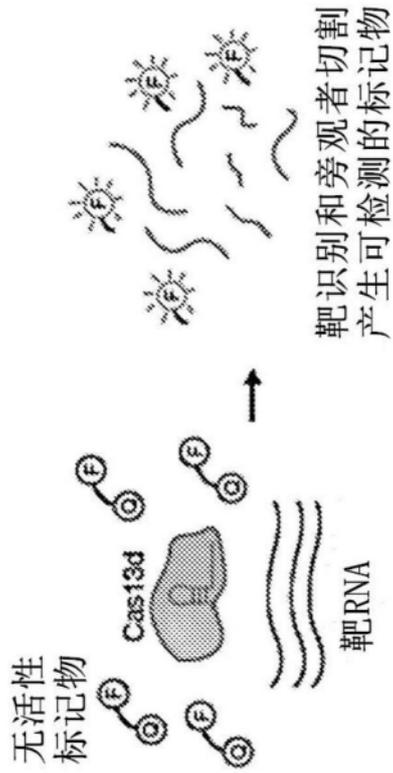


图15B

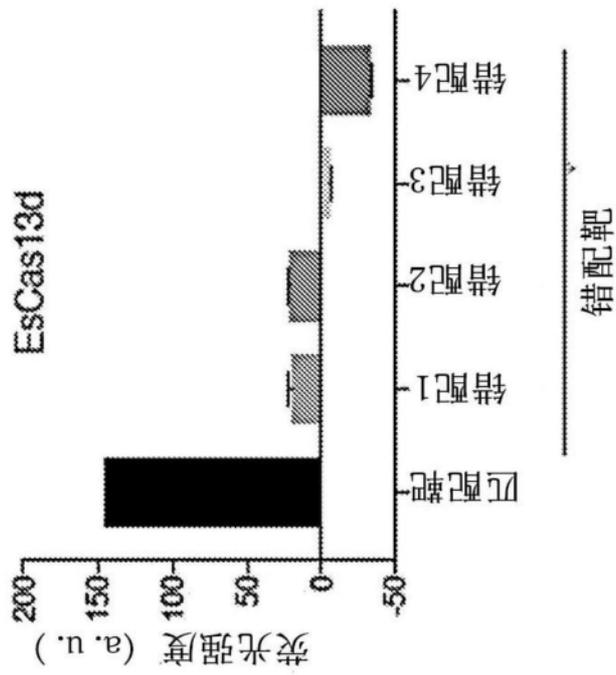


图15C

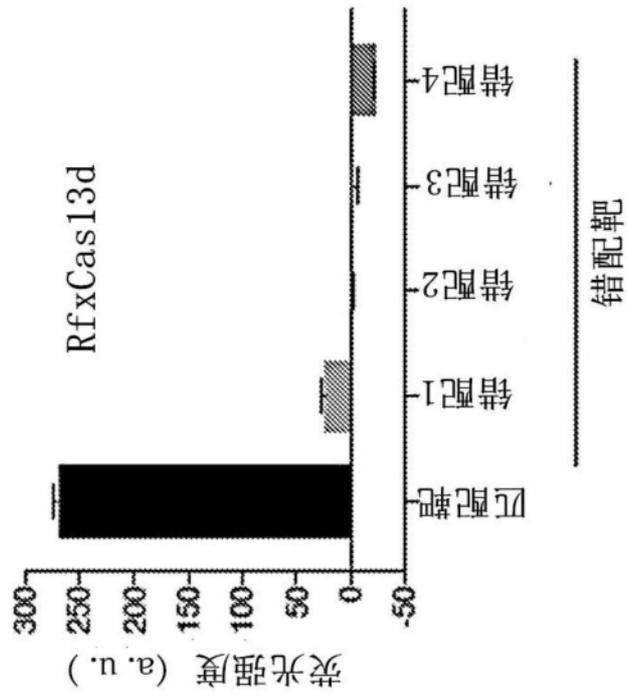


图15D

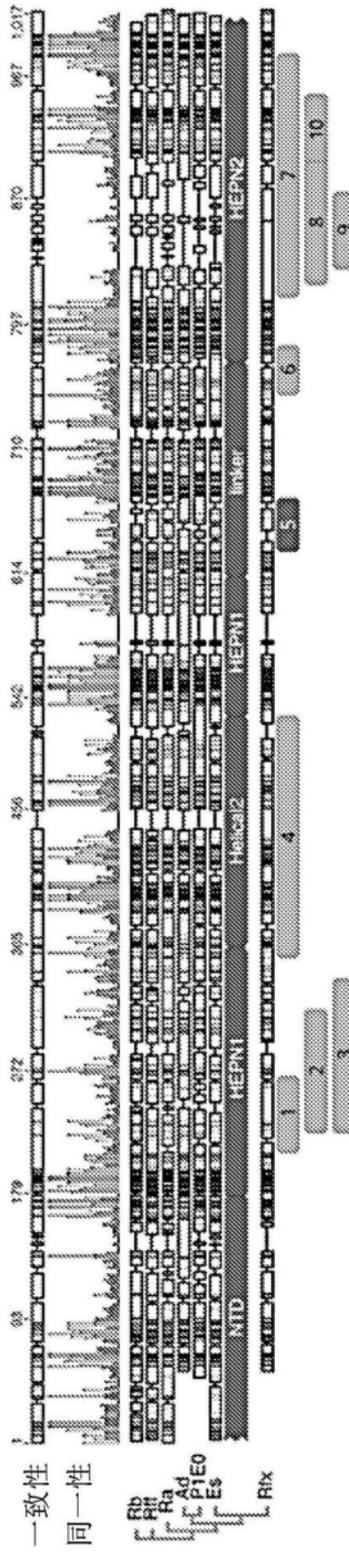


图16A

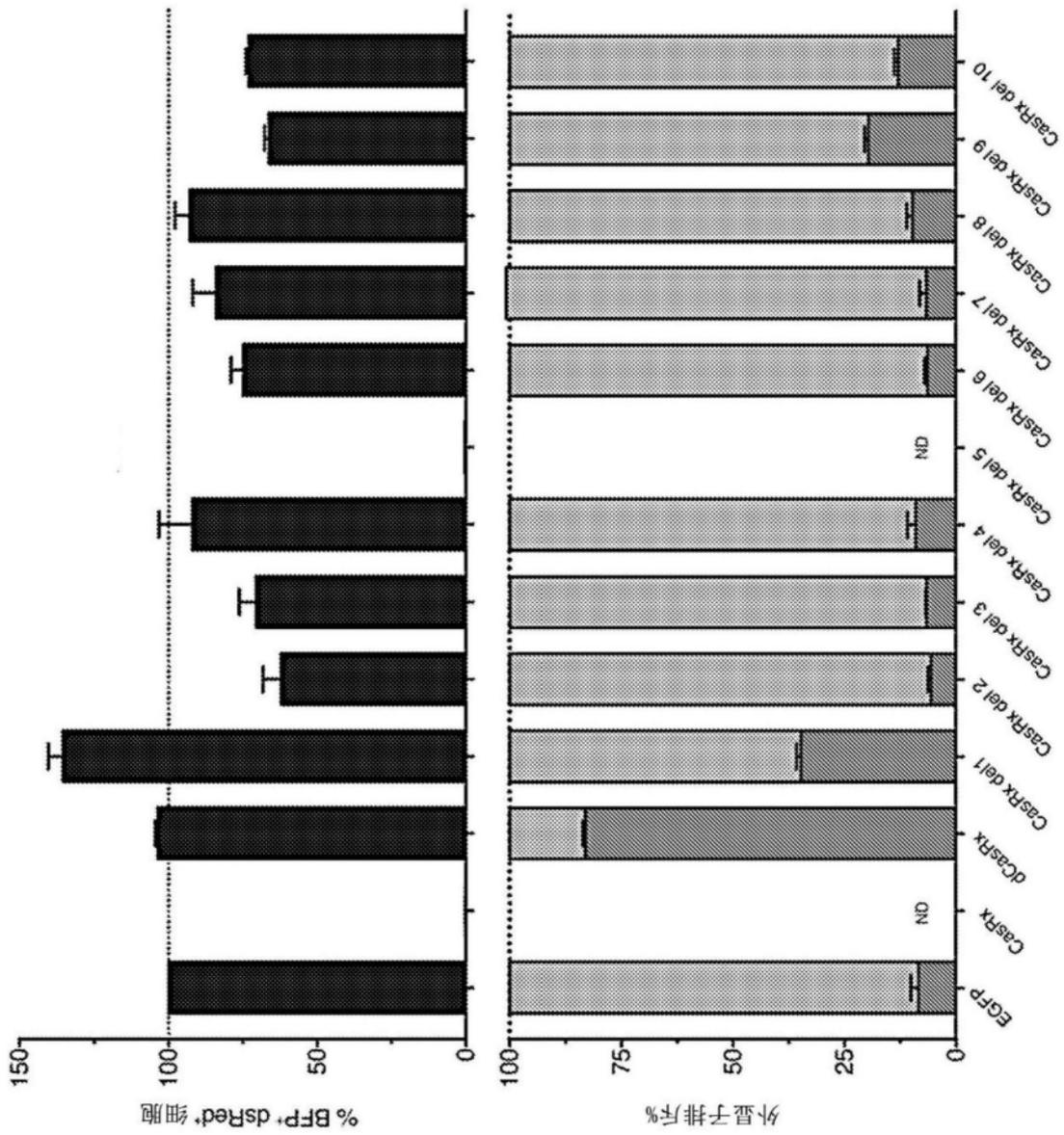


图16B

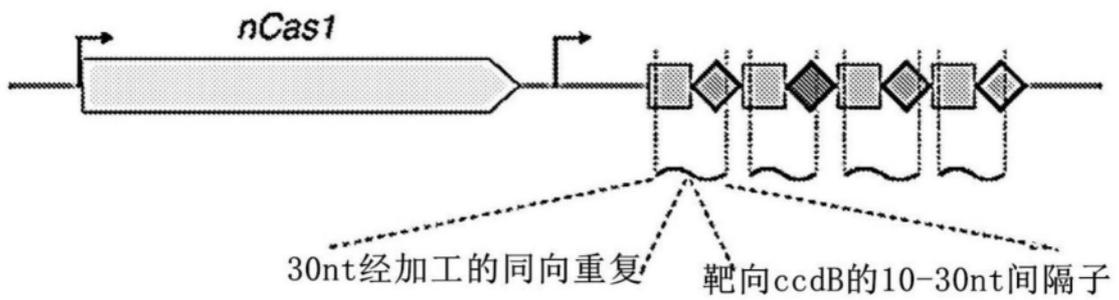


图17A

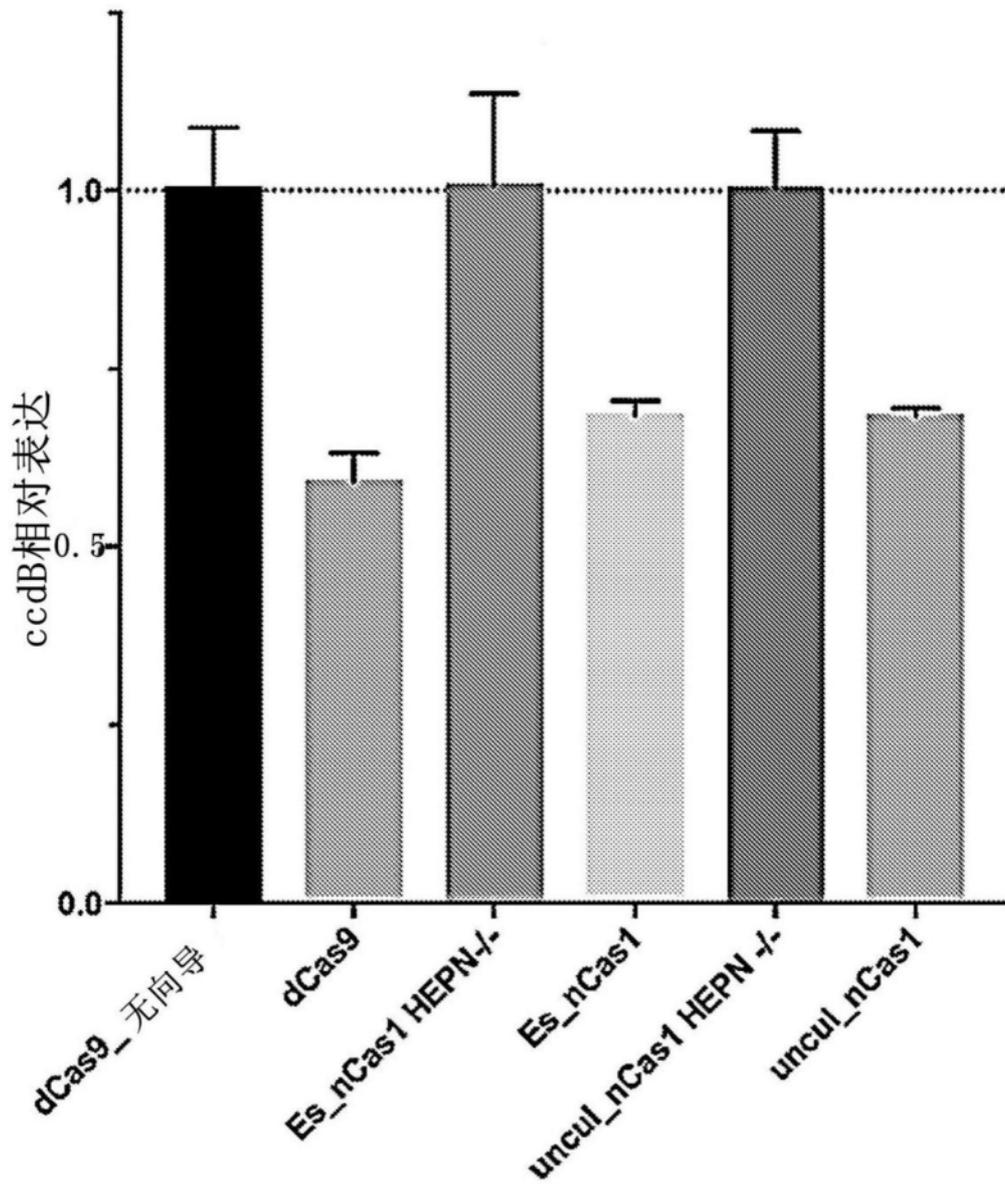


图17B

CLUSTAL O(1.2.4) 多序列比对

037_-_emb OIZA01000315.1	-----MAKKRMTAKERKQNHRESL-----MK	22
emb OCTW011587266.1	-----	0
k87_11092736	-----	0
BMZ-11B_GL0037915	-----	0
BMZ-11B_GL0037771	-----	0
BMZ-11B_GL0069617	-----	0
Ga0099364_10024192	-----	0
530373_GL0023589	-----	0
emb ODAI011611274.1	-----MKKISLKEQRNTKKAEN-----KL	20
EMG_10003641	-----	0
CasR_P1E0_metageno	-----	0
emb OIZX01000427.1	-----MAKKKTARQLREMQOQR-----KQ	21
160582958_gene49834	-----MKNSVTFKLIQAENKEA-----AR	20
gi 1198542314 gb NFIR01000008.1	-----MAKKLRPELREKRRMAE-----KE	20
emb OGNF01009141.1	-----	0
Ga0129306_1000735	-----	0
MH0288_GL0082219	-----MKKMTAKERREQEKKR-----AE	19
FIG-022_GL0026351	-----	0
FIG-028_GL0185479	-----	0
250twins_35838_GL0110300	-----MGNKQVSAQKRENKLCN-----QQ	22
FIG-014_GL0226364	-----	0
250twins_36050_GL0158985	-----MKKKHQSAAFKRVKKLKN-----QE	21

图18A

emb OJMM01002900.1	-----MGKHLNAKQKLEKKLKN-----	20
emb OGFN01002610.1	-----MAKKITAKQKREKERLN-----	20
emb OHCP01000044.1	-----MAKKITAKQKREKERLN-----	20
PIG-046_GL0077813	-----MAKKITAKQRRERERQN-----	20
pig_chimera	-----MAKKITAKQRRERERQN-----	20
emb OIEN01002196.1	-----	0
O2.UC29-0_GL0096317	-----	0
CasR_Anaerobic_dig	-----	0
emb OGDF01008514.1	-----	0
emb OGZC01000639.1	-----MKKNIRATREALKAQK-----	19
Ga0224415_10048792_chimera	-----	0
emb OCVV012889144.1	-----	0
Ga0187910_10006931	-----	0
R._flav_XPD_	-----	0
Ga0224415_10007274	-----	0
Ga0187911_10069260	-----	0
Ga0187910_10015336	-----MENVKNPKEIVKGAKVRFIYENPKPKQNO	31
ODAI_chimera	MLKKCSVCGKKFETNYQGGKPVCRNCRSNKSPVNPKSSRPESIDK-----	49
31009_GL0034153	-----	0
DLF014_GL0011914_	-----MAKKNAKQRREREKRL-----	20
CasR_R._albus	-----MAKSKGMSLREKRELEKQKR-----	23
Ga0187910_10040531	-----MAKKKAKQRREREQEAR-----	20
tpg DJXD01000002.1	-----	0
CasR_E._siraeum	-----MGKKIHARDLREQKTRD-----	20

图18B

FIG-018_GL0023397	-----MAKKLSPRELRERER-----QA	20
FIG-025_GL0099734	-----MAKKMNAKEQRAAIQAEK-----NK	20
Ga0129317_1008067_chimera	-----	0
EYZ-362B_GL0088915	-----MKKKLNIREQREAKAAA-----KQ	20
uncultured_Ru_sp	-----MAKKNMKPRELREFAQKKAR-----QL	22
R._flav._FD-1	-----MKKKMSLREKREFAEKQAK-----KA	20
tpg DBYI01000091.1	-----MKKKIKARDLREFAKQEK-----LA	20
037_-_emb OIZA01000315.1	KADSNAEKAKKPVVENK-----PDTAISKDNKPKPNKEIKKSKAK	64
emb OCTW011587266.1	-----MKQNDRENNNKIKKSAAK	18
k87_11092736	-----MKRQKTFAK	9
BMZ-11B_GL0037915	-----MAEHSKKTAAK	11
BMZ-11B_GL0037771	-----MSKDKKTKAK	10
BMZ-11B_GL0069617	-----MPHNKKSTIK	10
Ga0099364_10024192	-----MSQSTTKKAK	10
530373_GL0023589	-----MKTEMKNGEVKTSAKKENCISIAK	23
emb ODAI011611274.1	KYQKAQAEERAAAQQTAGAESEENPCFDVVKDT---KRRKALNPLHVEIEAFSAKSSVK	77
EMG_10003641	-----MEET-----KVT---KETTIEKQSTKRHKQSKKKTATK	30
CasR_P1E0_metageno	-----MEREVKPFPKKSIAK	15
emb OIZX01000427.1	AIQKQEQERQ-----EKAA---AAR---ETAAPFQPAAPVFKRQRKSLAK	61
160582958_gene49834	KKAKDIAEQARIAKRNQVVKKEENRI-----NRI---QIEIQTKKSNQNTQAYHLKSLAK	72
gi 1198542314 gb NFIR01000008.1	EHKKQEK---LRKEQEELRKKQEKQ-----RED---QKELEKIKKEEGEGEKKKSAGK	68
emb OGNF01009141.1	-----MADIDKKKSSAK	12

图18C

Ga0129306_1000735	-----MQK---QREQQVTVTDESERKKKPLKSGAK	26
MH0288_GL0082219	V-----IRGQEAN-----RKK---AQAAATQDATAFVHHKNAKSLAK	54
FIG-022_GL0026351	-----MAESYKIKSKKTRTK	15
FIG-028_GL0185479	-----MSESDKIKIKSKKTRTK	17
250twins_35838_GL0110300	KARQAESQRDKTKNMNVEK-----M-----KNINTNDIKHKTAK	58
FIG-014_GL0226364	-----MTKPEKKQIKQATLTK	16
250twins_36050_GL0158985	KAKYASEPSPLOSDTAGV-----EC-----SQKKTVVSHIASSKTLAK	60
emb OJMM01002900.1	KDMYTKSTDAVSPTLKA-----APTKAEMSQDTAEASTLITPGTLKTKAK	67
emb OGPN01002610.1	KWAKQDTPVVPKSKTEEKP-----V---AASDDKLLKTTQVKKVQTKSKAK	63
emb OHCP01000044.1	KWAKNDSVIIVPETKEEIK-----TGEIQDNNRKRSRQKSQAK	58
FIG-046_GL0077813	KWAKKQDDATAVFECEANI-----KP---ADSKDEDCANT--SIKRKKTKQAK	62
pig_chimera	KWAKKQADATAVFECEADI-----KP---ADSKDEDCNTIYIKREKKTKQAK	64
emb OIE01002196.1	-----MERQRRKMKSKSK	13
O2_UC29-0_GL0096317	-----MERQRRKMKSKAK	13
CasR_Anaerobic_dig	-----MNNKRRKTKAK	10
emb OGDF01008514.1	-----MTETPKREDIAKTPAAKSRSK	22
emb OGZC01000639.1	-----KRREELEKKNLAQWEETSAGRRSRVK	65
Ga0224415_10048792_chimera	-----MSKKENRKSYYK	12
emb OCVV012889144.1	-----MSKNKESYAK	10
Ga0187910_10006931	-----MIEKKKSYAK	10
R._flav_XPD_	-----MIEKKKSFAC	10
Ga0224415_10007274	-----MIEKKKSYAK	10
Ga0187911_10069260	-----MSTKKRFRYSVAAK	14
Ga0187910_10015336	REFSAINVEVITDETHKEAKIQKNEFKTFDQFTKE-LQETQVNGETKKEHITKNKHTNVK	90

图18D

ODAI_chimera	EILTEKLI S DCKK-----NDCASECGKGRKRSKAK	79
31009_GL0034153	-----MGRGVII M TEKRKSRAK	18
DLF014_GL0011914_	KIEKFKI K YAEQETVDKI-----TEEIAQSDKIMRTESILVQHTTKSLAK	65
CasR_R._albus	KAAVNSVND T PEKTEAN-----VVS V NVRTSAENKHSKKSAAK	62
Ga0187910_10040531	KIQSAVKAKAETAPA-----YSSAFVEKRKDKQSKKTFFAK	55
tpg DJXD01000002.1	-----MKKQKSKKTVSK	12
CasR_E._siraenum	KFADQNK K REAE R AVPKK-----DAAVSVKSVSSVSKKDNVTKSMAK	63
PIG-018_GL0023397	KFAKQVQ R EAEAKAKREK-----QVSVSDTVGAAAAVQENGGKKSMAK	63
PIG-025_GL0099734	KIQMANAAREKQAE E EARLKAAEAERIEKEENKKREAELKRFKLDQETRGKLPKNERKSLAK	80
Ga0129317_1008067_chimera	-----MKKQKKS L VK	10
EYZ-362B_GL0088915	IYAAKASAAQ K KADETV-----VAAQ P AGDTANSKPIKSRNK	57
uncultured_Ru_sp	KAAEIN N NAAPAI---A-----AMPAAEVIAPAAEKKKSSVK	56
R._flav._FD-1	AYSAA S KN T SDSKPAEKKA-----ETPKPAEIIISD S RNKTAVK	58
tpg DBYI01000091.1	AFSAKANTVY E NE D K-----NVEAFPEALN L RSIKKSMNK	55
	*	
037__emb OIZA01000315.1	LAGVKWIIKAN--DD-----V T YISSFGKGN-----NSVLEKRIIGD V SGDVN----	105
emb OCTW011587266.1	AVGVKSLARLS--DG-----S T VVSSFGKGA-----AAELES L ITGGEIRKLS----	59
k87_11092736	RIGIKSTVAYG--QG-----K Y AITTFGKGS-----KAEI A VR S ADPP E ETLPTESD	54
BMZ-11B_GL0037915	RVGIKSVIAHD--GG-----K L AVTTFGKKS-----AAK L A Y RDGDQ T FFDEEK K SE	56
BMZ-11B_GL0037771	RMGVKALLAHG--ED-----K L TM T TFGKGN-----R S KIE F TEGYH G RALE T PKHF	55
BMZ-11B_GL0069617	RIGVKS A FSFHG--NS-----R L TITQ F GRGN-----T A DI A VE T DSRGRSL S PHRV	55
Ga0099364_10024192	RMGVKSVLAHG--KDEK G H I K L AI T A F GKGN-----K A FL A I Q TDEK G S N LAK T YKE	60
530373_GL0023589	ANGLKST F FLA--EN-----T Y MT S FGKNN-----T A D P E K NI V RM V EN I R----	64

图18E

emb ODAI011611274.1	ANGLKSLLLT---DG-----KVTMTSFRGS-----EANVEKRFEDETGTKTFD-----	117
EMG_10003641	MSGLKSALVIN--NH-----EMLLTSFGKGN-----NAIAEKRYIILDGDIETI---N	72
CasR_P1E0_metageno	AAGLKSTFVISPQEK-----ELAMTAFGRGN-----DALLQKRIIVDGVVRDVA-----	58
emb OIZX01000427.1	AAGLKNFILDQRR-----TTVMTAFQGS-----TAILEKQIIVDRAISDLQ-----	104
160582958_gene49834	AAGYKSVFAI---GN-----DLLMTGFGPN-----DATIEKRQVQRAIETL-----	112
gi 1198542314 gb NFIR01000008.1	ALGLKSTFILDRDEQ-----KMLMTSFRGN-----KAVRDYIIGDKVSDIDDSWE	115
emb OGNF01009141.1	AAGLKSTFVLE--NN-----KLLMTSFGNGN-----KAVIEKIIIDEKVDSIN-----	52
Ga0129306_1000735	AAGLKSVMFVLS--EG-----KELLTSFRGN-----EAVPEKRVTTGGTIANAR---T	68
MH0288_GL0082219	AAGLKSTFTLP--TG-----ELLMTSFRGS-----EALIQKRIMKDVISDVQ-----	95
PIG-022_GL0026351	RAGLKSSFVN--EN-----KGLMTAFGKGN-----NAVIEKRISGNLIENVS-----	55
PIG-028_GL0185479	RAGLKSSFVS--EG-----KGLMTAFGKGN-----NAVIEKRISGDLIENVS-----	57
250twins_35838_GL0110300	KLGLKSTIIAD--K-----KIILTSFINEQ-----SSKTANIEKVAGFKGPTIDT--IS	103
PIG-014_GL0226364	RMGLKSTLVFD--N-----KITVTSFANSS---NSEKASAHIEKRTDFWGKTI-----M	60
250twins_36050_GL0158985	AMGLKSTLVMG--D-----KLVITSFAASKAVGGAGYKKSANIEKITDLQGRVI-----E	107
emb OJMM01002900.1	AMGLKSTLVFD--D-----KIVVTSFLNSKT--EENEKCAHIEKIAADCNGQTI-----V	112
emb OGPN01002610.1	AMGLKTVLSFD--D-----KIAIASFVNDK-----KTKLPHIERITDKSGFTI-----H	105
emb OHC01000044.1	AMGLKAVLSFD--N-----KIAIASFVSSK-----NAKSSHIERITDKEGFTI-----S	100
PIG-046_GL0077813	AMGLKTVLGFDF--N-----KIAIASFTSSK-----DSKSSHIERITDPNGKTI-----R	104
pig_chimera	AMGLKTVLGFDF--N-----KIAIASFMSSK-----DSKSSHIERITDPNGKTI-----R	106
emb OIEN01002196.1	MAGYKSVFVIG--D-----ELLMTSFGDGD-----DAVLEKDIIDENGVVNDC---R	54
O2_UC29-0_GL0096317	MAGYKSTFVID--S-----DLLMTSFERD-----KAVREKYIAQGNVKSRLR--K	54
CasR_Anaerobic_dig	AAGLKSVFFDQ--K-----QAVLTTFAKGN-----NSQIEKKVVNSEVVKDLR--Q	51
emb OGDF01008514.1	AAGLKSTFAVN--G-----SVLLTSFRGN-----DAVPEKLIITEKAVSEIN---T	63
emb OGZC01000639.1	AVGVKSVFVVG--D-----DLYLATFCNGN-----ETVLEKKITPDGKITTF---P	106

图18F

Ga0224415_10048792_chimera	GLGLKSTLVSD--S-----KVYLTTFADGS-----NAKLEKCVENKIKICIS----N	53
emb OCVV012889144.1	GMGLKSALVSG--S-----KVYMTSFEFEGN-----DAKLEKVVENSEIVSLA----	50
Ga0187910_10006931	GMGLKSTLVSD--S-----KVYMTSFGNGN-----DARLEKVVENNAISCLV---D	51
R._flav_XPD_	GMGVKSTLVSG--S-----KVYMTTFAEGS-----DARLEKIVEGDSIRSVN-----	50
Ga0224415_10007274	GMGLKSTLVSG--S-----KVYMTTFFDGS-----EARLEKVVENDSIKTLH-----	50
Ga0187911_10069260	AAGLKSSLAVD--TD-----RTVMTSFGHGN-----AAILEKEIVDGEIISVLN---I	56
Ga0187910_10015336	AAGVKSFAVD--DG-----NVLITSFGRGN-----AADIETLKSDDDKTINL---T	132
ODAI_chimera	AAGLKSALDLG--NG-----QALITSFGKGN-----EAVPEKRVEGDKIISLR---E	121
31009_GL0034153	ALGVKSFLKD--D-----DVIMTSFGRGN-----DAVVEKIIKNNVQNIN---I	59
DLF014_GL0011914_	ASGVKSFIKD--D-----EVIMTSFGRGN-----DAVVEKIIKDNVDEK-----N	105
CasR_R._albus	ALGLKSGLVIG--D-----ELYLTSFGRGN-----EAKLEKKSIGDIVEKKG---I	103
Ga0187910_10040531	ASGLKSTLAVD--N-----SAVMTVFGRGN-----EAKLDRINADLQSESL---H	96
tpg DJXD01000002.1	TSGLKEALSVQ--G-----TVIMTSFGKGN-----MANLSYKIPSSQKPNL---N	53
CasR_E._siraeum	AAGVKSFAVG--N-----TVYMTSFGKGN-----DAVLEQKIVDTSHEPLN---I	104
PIG-018_GL0023397	AAGVKSFAVG--D-----TVYMTSFGKGN-----DAVLEKKIIVDESHESLN---K	104
PIG-025_GL0099734	AAGLKSTFAVG--T-----DLYMTSFGKGN-----DAIVEKKSINKQVTQIN---T	121
Ga0129317_1008067_chimera	AAGLKSFAVVG--D-----SVYLTSGKGN-----AARLDTKINPDNSTERY---V	51
EYZ-362B_GL0088915	AAGLKSTFVVG--D-----SLYMTSFGKGN-----LSVPEYRVDTGDAYRVD---P	98
uncultured_Ru_sp	AAGMKSILVSE--N-----KMYITSFGKGN-----SAVLEYEVDNNDYNTQTQ---L	97
R._flav._FD-1	AAGLKSTIISG--D-----KLYMTSFGKGN-----AAVIEQKIDINDYSFSA----	98
tpg DBYI01000091.1	AAGLKSTLIDG--K-----SLYLTAFGKGN-----NAVVEHMIATDDSYSILK---T	96
	: ::*	
037_-_emb OIZA01000315.1	---KDSHMVVPKYYTKKNEYIKNGF-----SSGSSSLVTV	136

图18G

emb OCTW011587266.1	---D-KAILEITDDTQKNAYNVKS-----SRIPNLITAR	89
k87_11092736	ATLSIHAKFAKAGRDGREFKCGDVD-----ETRI---HTSRS	88
BMZ-11B_GL0037915	----LGGFNVTVGDETVDAERGI-----PGEGLIEASVAN	88
BMZ-11B_GL0037771	G----KRGFEVRKIDENVVDLYGDL-----EEGK	79
BMZ-11B_GL0069617	----EKSFVTKIDDEIDVQRS-----	73
Ga0099364_10024192	----RNITANKIVSEGIQTSGTI-----AGE	82
530373_GL0023589	----NTFSATPISERIVDITGET-----	83
emb ODAI011611274.1	---RDPELFSAKPLEFTGYRIQRFNA-----S-----PKDAG----	145
EMG_10003641	---NKNKKFDANNDKVVVVIKGISN-----PNGQLTN	101
CasR_P1E0_metageno	---GEKQFQVQRQDESRFRLQNSR-----LADRTVTADD	90
emb OIZX01000427.1	----PVQQFQVEPASAAYRLKNSR-----VRFPNVTADD	135
160582958_gene49834	---SSPEQYSAEFQNKQFKIKGNIK-----VLNHS-TQKMEEIQTE	149
gi 1198542314 gb NFIR01000008.1	---NKKAALSVEVCGKSFNISKKEN-----DDCEPVKVVN	147
emb OGNF01009141.1	---EPEVFSVTPCKKFKFELQPAKR-----GLAADS�VDN	83
Ga0129306_1000735	---DNKEAFSAALQNKRFVFGFR-----TAGSSDD	95
MH0288_GL0082219	---EGDPAFSVIPMPKQFEIQGRAI-----DVQKEGSSVPEVLTDD	133
PIG-022_GL0026351	---NKPRFSVEFGNKEYIINDK-----KLGIVSDD	82
PIG-028_GL0185479	---NKPRFSVEFGNKEYIINDK-----KLGIVSDD	84
250twins_35838_GL0110300	---YTPRFRSEINPGEIIVISKGD-----DLSEFAN	131
PIG-014_GL0226364	---QDEKMFRTDVEKQOIYIEKEFPDENNKII-----T-----QTAMFEN	97
250twins_36050_GL0158985	---EHERMFSADVGEKNIELSKND-----CHTNVNN	135
emb OJMM01002900.1	---ERPRMFNTSINAKKVDL SKDN-----DETNYPN	140
emb OGPN01002610.1	---ENARMFDSVDEQNVNIEKRMTEEKQND-----GTFKKDEKDVKATICN	150
emb OHCPO1000044.1	---VNSKMFESSVNRKDINIEKRITIEEPQQD-----GTIKKEKGVKSTTCN	145

图18H

FIG-046_GL0077813	---EDVRMFDSDVDECSINLEKRMVTVEERQKD-----GTIKKDEKDVKSTICN	149
pig_chimera	---EDVRMFDSDVDECSINLEKRMVTVEERQKD-----GTIKKDEKDVKSTICN	151
emb OIEN01002196.1	---N-PAAYDAVYGTDSIRVKKTN-----NNIRAKVNN	83
O2_UC29-0_GL0096317	---Q-EQAFDAQYGKEKIFVTNAD-----NKLHVAINN	83
CasR_Anaerobic_dig	---P--PAFDLELKEKTFYISGKNNINTSRENPLASASPLSKRQIRAEIRIKRAREENR	106
emb OGDF01008514.1	---V-KPRFSVEKPATSYSSSF-----GIKSHISATADN	93
emb OGZC01000639.1	---E-EETFTAKLKFAQTEPTVATSI-----GISNGRIVLPEISVDN	144
Ga0224415_10048792_chimera	---D-KEAFAAASIANKNVGYKIKND-----EKFRHPKRYDYIISNN	89
emb OCVV012889144.1	---E-KESFSAEIEFKKNI GCKIE-N-----KKFKHPKRYDYVIADN	85
Ga0187910_10006931	---K-KEAFVAEITDKNAGYKII-N-----KKFGHPKGYDVVANN	86
R_flav_XPD_	---E-GEAFSAEMADKNAGYKIG-N-----AKFSHPKGYAVWANN	85
Ga0224415_10007274	---E-GEAFSAEMTDNAGYKIG-N-----VKFSHPNGYDVVANN	85
Ga0187911_10069260	---E-NPAFDVAVIND---KKYALTG-----HH-----AGVHALVDQ	85
Ga0187910_10015336	---ETENQ-----KKYVVTN-----K-----RSNVKGLADN	155
ODAI_chimera	---G--EAFALAKEPVMHKLVLKSG-----RIKAEFGGPFVFDG	154
31009_GL0034153	---D-KPVYDVVGVVEKETGIINVVG-----TKRGINTINKKKQYAN	96
DLF014_GL0011914_	---T-KPAYEIVDIEKNGDLIKVQ-----SRRCKNIESAN	136
CasR_R._albus	---G---AFEVAERD--ESTLTLES-----GRI--KDKTARPKD	132
Ga0187910_10040531	---PQAALKN--VHAPNKQKIHFIFIG-----RM-----QDMNLTADH	127
tpg DJXD01000002.1	---SSAGLKN--VEVS-GKKIKFQG-----R-----HPKIATTDN	82
CasR_E._siraenum	---D---DPAYQLNVVVMNGYSVTG-----HR--G--ETVSAVTDN	135
FIG-018_GL0023397	---A---NPAYEITDISSVNYSVVEG-----HR--G--QSVCATADN	135
FIG-025_GL0099734	---E-KENFTVTDGITDIIVPVQS-----KR--I--ACLNTQADN	154
Ga0129317_1008067_chimera	---SDSEKHTLKINSITDELRLSG-----P-----FPKQAEAKN	83

图18I

EYZ-362B_GL0088915	---IS-EHSELHVLGADDLAVSFSG-----TRMK--DVSELKADN	133
uncultured_Ru_sp	---SSKDNSNIQLGGVNEVNIIFSS-----KHGFE--SGVEINTSN	133
R_flav_FD-1	---M-KDTPSLEVDKAESKESISFSS-----HHPFV--KNDKLLTTYN	133
tpg DBYI01000091.1	---L-EMEPSLKVKAADDELKVTFMS-----RRPFV--QESELSAVN	131
037_-_emb OIZA01000315.1	PNK-----PDKNSGMDALCLKPYFEIDFFG-----HIF--TDNMHIQAIYNIFD	178
emb OCTW011587266.1	TDK-----LSDKSGMDDLGFKRELELELVFG-----QCF--DDS.IHIQIAHAVFD	131
k87_11092736	EYES-----LISNPAESPREDYLGKGTLEKRFVG-----DEY-PKDNLRIQI IYSILD	136
BMZ-11B_GL0037915	PVND-----LIDESRKEYGQDYMLKQLEKVFVG-----QTF-PGDNLRIQI IYNILD	136
BMZ-11B_GL0037771	TIEA-----LLINPSEKVGEDYLLKGTLEKRFVG-----REF-PHDNIRIQLIYNILD	127
BMZ-11B_GL0069617	SLLG-----TLLNPAENDPVYLGKSTLEKEFFG-----QEF-PNDSIRIQI IHNILD	121
Ga0099364_10024192	GHAT-----FLNPPAEHVGTDYLLKLETEMEFFG-----KSF-PGDSVRIQI IHQILD	130
530373_GL0023589	-----KKTAVVLPACSDNQLHAKDTIEQMYFG-----KTY--TDNIHQIAYNIMD	128
emb ODAI011611274.1	-----LAYRPAGVRPDQIGAKAALEKRYFG-----KET-PGDNIHVQIAYQIQD	188
EMG_10003641	PLFD---QSPTAIQPNRTSGNDMIGIRRMLEKRYFVHNEENKEF--QDNIRIQIAYCILD	156
CasR_P1E0_metageno	PLHR---A-ETPRRQPLGAGMDQLRRKALLEQKYFG-----RTF--DDNIHQI IYNILD	139
emb OIZX01000427.1	PLYR---R-KD---GGFVPGMDALRRKNVLEQRFFG-----KSF--ADNIHQI IYSILD	181
160582958_gene49834	LQD-----NYNRPHFDDLGCCKNVLEQKYFG-----RTF--SDNIHVQIAYNIMD	191
gi 1198542314 gb NFIRO1000008.1	PVL-----SGNKKDDDLIHCCKNLEEMYFG-----QQF--KDNHQIQLIYNILD	189
emb OGNF01009141.1	PLK-----SKKTAGDDAIHSRKFLEQRFFDG-----NTF--NDNIHQIQLIYNILD	126
Ga0129306_1000735	PLA-----VSRAPGQDLIGAKTALEERYFG-----RAF--ADNIHQI IYAIQD	137
MH0288_GL0082219	PRA-----SQPFAGADLIGAKPALEKRYFG-----KTF--TDNIHQIQLIYRILD	175
FIG-022_GL0026351	PRE-----SQKPVGEDGIHCKRALEERYFG-----KTF--NDNIHQIQLIYSILD	124

图18J

FIG-028_GL0185479	PRE-----SQKPVGEDGIHCKRALEREYFG-----KTF--NDNIHQIILYSILD	126
250twins_35838_GL0110300	PAN-----FFPIGRDYVKIRSALEKQYFG-----KEF-PEDNLHVQIAYNVAD	172
FIG-014_GL0226364	PAL-----RNCGRDYIGMKNVSEKVMFG-----KEF-PNDNLHQIAYNILD	138
250twins_36050_GL0158985	PVV-----TNIGKDYIGLKSRLRQEFFG-----KTF-ENDNLHVQLAYNILD	176
emb OJMM01002900.1	PAF-----EDCGRDYINVKSALAEKRVFG-----KTY-NKDNLHVQIAYNIFD	181
emb OGPN01002610.1	PYF-----KTCGKDYIGIKVAEKYFFG-----KTF-PNENLRVQIAYNVFD	191
emb OHCPO1000044.1	PYF-----KVGKDYIGIKEIAEEHFFG-----RAF-PNENLRVQIAYNIFD	186
FIG-046_GL0077813	PYS-----NECGKDYIGIKSVAEELFFG-----RTF-PNDNLRVQIAYNIFD	190
pig_chimera	PYS-----NECGKDYIGIKSVAEELFFG-----RTF-PNDNLRVQIAYNIFD	192
emb OIEN01002196.1	PLAKSNIRSEESALFRTRVNEYKREQDKYETLFFG-----KTF--DDNIHQILISKILD	136
O2_UC29-0_GL0096317	PLCGNKKQSPAEQYSRRVLEHKKAINQKYETIVFG-----KPS--NDNIHQVQVSKIMD	136
CasR_Anaerobic_dig	PYHNV-----KRVGEDDLRAKADLEKHYFG-----KEY--SDNLKIQLIYNILD	148
emb OGDF01008514.1	PLAGR-----APVGEDAIHAKEVLEQRVFG-----KTF-SDDNIHQIILYNILD	136
emb OGZC01000639.1	PLHTT--MQK--NTIKRSAGEDILQLKDVLENRYFD-----RSF--NDDLHRLIYNILD	193
Ga0224415_10048792_chimera	PLLHN-----NSVQQDMLGLKNVLEKRYFG-----KSSGGDNNLCIQIHHNIID	133
emb OCVV012889144.1	PLYK-----GSRQDMLGLKETLEKRYFN-----SAD-GTDNVCIQVIHNILD	127
Ga0187910_100006931	PLYT-----GPVQQDMLGLKETLEKRYFG-----SSVSGNDNICIQVIHNILD	129
R._flav_XPD_	PLYT-----GPVQQDMLGLKETLEKRYFG-----ESADGNDNICIQVIHNILD	128
Ga0224415_10007274	PFYT-----GPVQQDMLGLKEILERRRYFG-----SSTDGNNTICIQIHHNIID	128
Ga0187911_10069260	PQN-----RSDAVHIRGALEKRYFG-----DTF--ADNIHVQIAYNILD	122
Ga0187910_10015336	PTKV-----ESIIPGETQIGFKSILEKHFFG-----RTF--NDNIHQIILHNILD	198
ODAI_chimera	PQKNS-----SMSAVRLDEYNNQIERKFFRGN---KVL--QDNIRIQIAHNIMD	198
31009_GL0034153	TQV-----SQKRMGMDLIQRKDKLEKVFVN-----ENF--NDNIHQIILYNILD	138
DLF014_GL0011914_	TEL-----PPEKIGMDLIRCKDKLEELYFG-----RTF--NDNIHQIILYNILD	178

图18K

CasR_R._albus	PRHI-----T--VDTQGFKEDMLGIRSVLEKKIFG-----KTF--DDNIHVQLAYNILD	178
Ga0187910_10040531	PLHSH-----DGERAVGADLLCAKDKLEQLYFG-----RTF--NDNIHQIYIYQILD	172
tpg DJXD01000002.1	PLFK-----PQFGMDLLCLKDKLEMHYFG-----KTF--DDNIHQIYIYQILD	123
CasR_E._siraenum	PLRRFNGRKK--DEPEQSVPTDMLCKPTLEKFFG-----KEF--DDNIHQIYIYQILD	186
FIG-018_GL0023397	PLHHD-----GGSSVPTDMLCKPTLERVYFG-----KTF--DDNLHQIYIYQILD	179
FIG-025_GL0099734	PLYRK-----DSENOVQDPKLLKDTLEKVYFG-----KTF--NDTLHQIYIYQILD	199
Ga0129317_1008067_chimera	PTHK-----K--DNEQKNTRQDMLGLKSTLEKFFG-----STF--DDNIHQIYIYQILD	129
EYZ-362B_GL0088915	PLHSG-----K--NNLVPKNGTDILGLKILEQRYYG-----RNF--DDNVHQIYIYQILD	180
uncultured_Ru_sp	PTHRS-----GESSPVRGDMGLKSELEKRFFG-----KTF--DDNIHQIYIYQILD	178
R._flav._FD-1	PLYGG-----K--DNPEKPVGRDMLGLKLEERYFG-----CTF--NDNLHQIYIYQILD	180
tpg DBYI01000091.1	PLHSG-----K--DKPKNSAGQDMLGLKSELEKRYFG-----KIF--DDNLHQIYIYQILD	178
037_-_emb OIZA01000315.1	IEKILAKHITNIIYTVNSFDRNYNQ-----SGNDTIGF-----DLNYRV	217
emb OCTW011587266.1	IQKSLAAVIPNVLTYTLNLDRSYST-----DNTSDKKDIIGN-----TLNYQH	174
k87_11092736	IQKILGLYVEDILHFVDGLQDEP-----EDLVGL-----GLG--D	169
BMZ-11B_GL0037915	MQKAFGLYVNDVLYAVNNLQNTGRL-----LTDDWTSDVIGL-----ALG--K	178
BMZ-11B_GL0037771	IYKILGMNVADILYALGNMQDTE-----LDIDMFGQ-----SLN--N	162
BMZ-11B_GL0069617	IQKILGIYINDVLYSIANLQNV-----TVEDCLGQ-----SLS--K	156
Ga0099364_10024192	IQKLLGIYITDIIYCINNLRDETHL-----DHESDIVGL-----SMS--N	168
530373_GL0023589	MKKIFGYSNIIYSINNLLEE-----REGSDFLGV-----FYTEN	164
emb ODAI011611274.1	IEKLLAVYISNIIYAVNVVTGSAM-----KDSKGRPVDLLGD-----YGILGEE	233
EMG_10003641	IEKILMPHINNICEINNMLRL-E-----GYQEDSFMGS-----FNLYK	194
CasR_P1E0_metageno	IHKMLAVFANHIVHTLNLGGY-----GETDFVGM-----LPAGL	174

图18L

emb OIZX01000427.1	IHKILAAASGHIVHLLNIVNGS-----KDRDFIGM-----LAAHV	216
160582958_gene49834	IEKLLTPYINNIITYTLNELMRD-----NSKDDFFGCDSHFSVAYLYDELKA	237
gi 1198542314 gb NFIR01000008.1	IEKILAVQINNIIVFILNNLLRW-S-----GEEFFDLIGS-----LGVNH	227
emb OGNF01009141.1	IEKILSVHVNDIVYSVNNILSRGE-----GMEYNDYIGT-----L---NLK	164
Ga0129306_1000735	INKILAVHANNIVYTLNLLDRE-A-----DPETDDFIGS-----GYLTLKN	177
MH0288_GL0082219	IKKVLAIHANNIVYTLNLLSRR-E-----GTEGTFDFIGI-----GGLTLKN	215
FIG-022_GL0026351	LEKILAVHSNNIVYTFDNLRRK-E-----TDEYDDFIGY-----MSTVN	162
FIG-028_GL0185479	LEKILAVHSNNIVYTFDNLRRK-E-----TDEYDDFIGY-----MSTVN	164
250twins_35838_GL0110300	IKKILSVYINNIIMFYNLARSEEYDIFYNSQSENSGRCDVIG-----SLYYQA	222
FIG-014_GL0226364	IKKIICPYINNVVNIYFYNLGRGKLG-----DDKNDLLG-----KQIGP-	176
250twins_36050_GL0158985	IKKILGTYVNNIYIFYNLNRAGTG-----RDERMYDDLIG-----TLYAYK	218
emb OJMM01002900.1	IKKIIGAYINNIYIFYNLGREEYD-----AKKDIIG-----TQ----	215
emb OGPN01002610.1	IQKILGTYVNNIYYSFYNLRRDG-K-----SDVDIIG-----SLYAF	228
emb OHC01000044.1	VQKILGTFVNNIYYSFYNLRRDEVQ-----SDNDVIG-----MLYSIS	224
FIG-046_GL0077813	IQKILGTYINNIYYSFYNLRRDEFQ-----SDNDVIG-----TLYMLK	228
pig_chimera	IQKILGTYINNIYYSFYNLRRDESQ-----SDNDVIG-----TLYMLK	230
emb OIEN01002196.1	IEKTFVWVIGNIVVAINNLSLEQS-----IDRPIDIFGD-----KNTQ----	174
O2.UC29-0_GL0096317	IEKTFVFMGNIVVAINNLDPEQN-----IDNPLDYMSN-----KART----	174
CasR_Anaerobic_dig	INKIISPYINDIVYSMNNLARNDY-----IDGKIDVIG-----SLSSTT	188
emb OGDF01008514.1	IRKILSTYANNVVFTINSMRRLDE-----YDREQDYLG-----YLYTGN	175
emb OGZC01000639.1	IEKILAEYTTNAVFAIDNVSGCS-----DDFLS-----NFSTRN	227
Ga0224415_10048792_chimera	IEKILSEYIPNVVAFNNIAGFKDE-----HNNIIDIIIGT-----QTYNSSY	175
emb OCVV012889144.1	IEKILAEYITNAVVSFDNIAGFG-----EDIIIGM-----GGFKPIY	163
Ga0187910_10006931	IEKILAEYITNAAYAVNNIAGLD-----KDIIGF-----GKFSTVY	165

图18M

R._flav_XPD_	IEKILAEYITNAAYAVNNISGLD-----KDIIGF-----GKFSSTVY	164
Ga0224415_10007274	IEKILAEYITNAVYATNIIIDPD-----NDVIGG-----KKFTSIK	164
Ga0187911_10069260	ITKILTVYANNVVYALNNLHVHADD-----T-QADELDSLQ-----NFSAGT	163
Ga0187910_10015336	IKKILAVHTNNIVYALDNIHERGRE-----NSAEKPIDMIGA-----GGISTSK	242
ODAI_chimera	IEKILAVHINNIVYSVNNLFRVQG-----NEGDDFLG-----YRGLDK	236
31009_GL0034153	IEKILSVYVNNIVYALGNLERDNS-----DDEKDLIG-----YASTLN	176
DLF014_GL0011914_	IEKILSVYINNNAVYALGNLERKEN-----DKEKDLVG-----YLSARA	216
CasR_R._albus	VEKIMAQVSDIVYMLHNTDKTER-----NDNLMG-----YMSIRN	214
Ga0187910_10040531	IQKILALHANNIIFALDNLHLHKKN-----DELSDDFVGM-----GRMRATI	213
tpg DJXD01000002.1	IEKILAVHVNNIVFTLDNVLHPQK-----EELTEDFIGA-----GGWRINL	164
CasR_E._siraenum	IEKILAVYSTNAIYALNMSADEN-----IENSDFFM-----KRTTDE	224
FIG-018_GL0023397	IEKILAIYSTNAVYALNNMIADED-----DESSDLFQ-----NMTTYD	217
FIG-025_GL0099734	IQKILTVYSVNTVYTLNMLFGKEI-----SERGDFIG-----ALSFQQ	237
Ga0129317_1008067_chimera	IAKILAAHSNNAGYALDNMLAYQG-----VEFSDMIG-----YMGTSR	167
EYZ-362B_GL0088915	IEKNLAIYVTNVVAALDHMCDETL-----DTEHEDFIG-----YMNTMN	219
uncultured_Ru_sp	IEKILAVYVTNIVYALNNMLGVKG-----SESHDDFIG-----YLSTNN	217
R._flav_FD-1	IEKILAVHSANITTTALDHMVDEDE-----KYLNSDYIG-----YMNTIN	220
tpg DBYI01000091.1	IEKILAVYATNITAAIDHMVDDDDNE-----QYLOQDFIG-----YMNTLN	218
	: * : . . .	
037_-_emb OIZA01000315.1	PYLEYGGGKDSNGK-----PNKISAWKKRENFINFYN-----EAKPHLGY-	257
emb OCTW011587266.1	SYESFNVE-----KRGFTEYYN-----AAKDRFSY-	200
k87_11092736	-----EKMQKLLS-----KALPYMGF-	185
BMZ-11B_GL0037915	-----GRGKEIIE-----SMSDALGF-	194

图18N

BMZ-11B_GL0037771	-----P--EKAKDVVQ-----RMKPYMGF-	179
BMZ-11B_GL0069617	-----GQDEKLIKILN-----SITPYLGF-	175
Ga0099364_10024192	-----TNVNLALN-----QMRPYFGF-	184
530373_GL0023589	SYQAMNFADAI-----IKSKLLTKNSKYLNLDMNVWF-NK-EVSELSFRKYV-	211
emb ODAI011611274.1	GLTKRLQRIPE-----Q-----ADEEAKAIQAFI-----CSERLSY-	264
EMG_10003641	PYDAFIATDD-----KE-----SSRRDNFAKLM-----TSKQVRY-	225
CasR_PIE0_metageno	PYDKLRVVKKK-----NGDT-----VDIKADIAAYA-----KRPQLAY-	207
emb OIZX01000427.1	LYNELNEEA-----KRSIADFC-----KSPRLIY-	240
160582958_gene49834	GYSDDLKTKPN-----LS-----KNIDRIWNNFCNYMN-SDSGNTEARLAY-	277
gi 1198542314 gb NFIR01000008.1	TYEFRGRNK-----NYGKFSELI-----KQSQMRY-	253
emb OGNF01009141.1	SFETYKNNLV-----NKKKFDLDRVK-----KIPQLAY-	192
Ga0129306_1000735	TFETYCDPAAL-----NER-----EREKVTVSKQHFDAFM-----QNPRLAY-	214
MH0288_GL0082219	SYAAYCEPRTF-----FDEGEKNI IENIEKSHGFLEYM-----KNPRLAY-	257
PIG-022_GL0026351	SYEEFMNPEKL-----L-PGKDQSVYDNIINRSRETFLEYR-----KNPRLGY-	203
PIG-028_GL0185479	SYEEFISPEKL-----L-PGKEQSVYDNIINRSQETTFLEYR-----KNPRLGY-	205
250twins_35838_GL0110300	SYRNQDANRFE-----KD--GKKKAIDSLLD-----DTRAYITY-	254
PIG-014_GL0226364	-----KSEEMVEDLLK-----NSSAYFVY-	195
250twins_36050_GL0158985	PMEAQQTYYLLK-----GDKMRRFEVVKQLLQ-----NTSAYYVY-	253
emb OJMM01002900.1	-----DSAYISKILN-----NTSAYFTY-	233
emb OGPN01002610.1	DFDNLKDK-----PAFREAKDLLK-----NTEAYFSY-	256
emb OHC01000044.1	DYDRQKET-----ETFLQAKSLLK-----QTEAYYAY-	251
PIG-046_GL0077813	DFDQKET-----DTRFQARALLE-----RTEAYYSY-	255
pig_chimera	DFDQKET-----DTRFQARALLE-----RTEAYYSY-	257
emb OIEN01002196.1	-----GISLREDNDYLK-----TMLPRCEYL	195

图180

O2.UC29-0_GL0096317	-----DENLK-----RVLRRCMYL	188
CasR_Anaerobic_dig	-----LEKEKFSFHRENYKKFVE-----ASKPYMRY-	226
emb OGDF01008514.1	-----GKQFQTINGFWDLLD-----MIEPYMCY-	229
emb OGZC01000639.1	-----REHF-----GNKDNVICSVKKQQLFFNFFK-----NNRIGY-	270
Ga0224415_10048792_chimera	-----KYIEFQKLLK-----NKRLGY-	203
emb OCVV012889144.1	-----YNNKFFDDIILN-----NSRLGY-	189
Ga0187910_10006931	-----KERFIKDGKLDTKLINQLKNQYDEFDAFLD-----DTRFGY-	212
R._flav_XPD_	-----RAAFN-----NNDKLIINAIAKAQYDEFDNFLD-----NPRLGY-	207
Ga0224415_10007274	-----SSNDFEQFLK-----NPRLGY-	189
Ga0187911_10069260	-----KQQDFVELFI-----KKKEIHGY-	192
Ga0187910_10015336	-----YEDNFKQLI-----NNERIAY-	270
ODAI_chimera	-----DKERFFNFIFIGDKGKVTNPRLSY-	270
31009_GL0034153	-----VKQENIDKFIE-----NGNRLGY-	205
DLF014_GL0011914_	-----DEKIKDRIEDLKEFIE-----NGNRLGY-	250
CasR_R._albus	-----LPDDTKQKVENQKREFDKIIK-----SGRLGY-	251
Ga0187910_10040531	-----KVQETYREFQEFVR-----RKELLY-	244
tpg DJXD01000002.1	-----KYDRFKNYIK-----RKELLY-	190
CasR_E._siraenum	-----T-----NSREKADFAFEKFIG-----NYRLAY-	258
FIG-018_GL0023397	-----RIVKSFNDFVS-----NPRLAY-	244
FIG-025_GL0099734	-----KEIFEDFYS-----KRS LGY-	263
Ga0129317_1008067_chimera	-----NKDFRFLK-----LPR LGY-	193
EYZ-362B_GL0088915	-----SRADYKKINDSRRRFFENLLR-----TKRLGY-	257
uncultured_Ru_sp	-----SLSDDKKANVRKSLSKFNALLK-----TKRLGY-	256
R._flav._FD-1	-----SLSPKDRKNIDNSRAKFEKLLS-----TKRLGY-	260

图18P

tpg DBYI01000091.1	TYEVFMEPSKN-----P-----RLDSNARKNIENSREKFEYLLD-----TQRLGY-	258
037_-_emb OIZA01000315.1	YENIFYDHG-----	266
emb OCTW011587266.1	FPDILCVLEKVGKDR-----	216
k87_11092736	FGSTDVFKVTKKRE-----	199
BMZ-11B_GL0037915	FGSAFKPIPKKPGDKTSGKYAGGKLER-----	222
BMZ-11B_GL0037771	FGGIFRTQKKEDRKKDSNLSEEEKKKEK-----	209
BMZ-11B_GL0069617	LGAGFKQSERPKHQKNSRDPDPRQKIE-----	201
Ga0099364_10024192	FGFAFRPVGDDKVKKEITLSDEVKNIKIIALEEQKRNPSTPRFKQENINLEIENAMGKF	244
530373_GL0023589	YRSRYANKTSTAKKICICLQEDFHFFDYA-----TVKKAADQFLLKRIQRS	258
emb ODAI011611274.1	FGKEFCLVRNSPKQP-----	279
EMG_10003641	LGNALYSDSLNLTKDE-----	242
CasR_P1E0_metageno	LGAAFYDVTFGKSKR-----	222
emb OIZX01000427.1	YSAAFYETLDNGK-----	253
160582958_gene49834	FGELFYKPKETGDAK-----S	293
gi 1198542314 gb NFIR01000008.1	FGSTFCLFNENERITS-----	270
emb OGNF01009141.1	FGSAFYNTPEDTSAK-----	207
Ga0129306_1000735	YGNFAFRKLSKAERLARG---REIFDK-----ESPERRQEILGSRGKN	254
MH0288_GL0082219	YGEAFYRRLSNNERSKLGMEYERMSDA-----QRDAFDRKYPIDFCDG	301
PIG-022_GL0026351	FGKVFYEKVTDPKMK-----	218
PIG-028_GL0185479	FGKVFYEKSTDAKDK-----	220
250twins_35838_GL0110300	FDGLFVSPKREDDGKI-----	270
PIG-014_GL0226364	FGDVFQVKLSKEQQE-----	211

图18Q

250twins_36050_GL0158985	YGTLFEKVKAKSKEQ	269
emb OJMM01002900.1	FDGVFKQITDRDSNK	248
emb OGPN01002610.1	FGDVFKSKKGGKDE	271
emb OHCP01000044.1	FDDVFKKKNKPKDNKE	267
FIG-046_GL0077813	FDNVFKKIDKKKRN	270
pig_chimera	FDNVFKKIDKKNKKS	272
emb OIEN01002196.1	FHNILNSDSDNNSKMN	211
O2.UC29-0_GL0096317	FDLGFEYAEPSPYKKS	204
CasR_Anaerobic_dig	YGKVFIRDVKKSKLST	242
emb OGDF01008514.1	FSEAFFCETTVKDPDS	245
emb OGZC01000639.1	FGKAFFHAESE	281
Ga0224415_10048792_chimera	WGKAFFTGQGN	214
emb OCVV012889144.1	YGKAFFKNDLKHNPN	205
Ga0187910_10006931	FGKAFFCKEGD	223
R._flav_XPD_	FGQAFFSKEGR	218
Ga0224415_10007274	LGKAFFYKDGK	200
Ga0187911_10069260	FGDTFAFLDKRI	204
Ga0187910_10015336	FGNAFFKDEGNK	282
ODAI_chimera	YGAFAFAEKVQPT	282
31009_GL0034153	FGNVFYKNKNE	216
DLF014_GL0011914_	FGDIFFKKEEVSKNGK	266
CasR_R._albus	FGEAFMVNSGNS	263
Ga0187910_10040531	FGSAFYNGDT	254
tpg DJXD01000002.1	FGEAFYHENE	200

图18R

CasR_E._siraenum	FADAFYVNKKNPKGK-----	273
PIG-018_GL0023397	YSDAFYEKKGK-----	256
PIG-025_GL0099734	YGHIFFKDK-----	273
Ga0129317_1008067_chimera	FGSAFYSQG-KD-----	205
EYZ-362B_GL0088915	FGFAYNPKDR-----	267
uncultured_Ru_sp	FGLEPKTKD-NR-----	268
R._flav._FD-1	FGFDYDANGKDK-----	273
tpg DBYI01000091.1	LSLEYDKRSKDKR-----	271
037_-_emb OIZA01000315.1	-----EPISEKFNLYLNILNFI RNNTFHYKDN-----	294
emb OCTW011587266.1	-----YQPKSEKDAFNVLSSVNMLRNSLFHFAPKSND-----GKA	251
k87_11092736	-----ERAAADEHNAKVFRALGAI RQKLAHFKWK-----E	229
BMZ-11B_GL0037915	-----DAAIAEDVAVLRVLCGLRQISAHFRSEKKG-----E	253
BMZ-11B_GL0037771	-----QEELQKDLQHNMDVARCISALRHATAHNKPATAHDK-----QDE	249
BMZ-11B_GL0069617	-----DPWAYNKSALRILSTLRNKL AHFEDS-----	228
Ga0099364_10024192	KSKDAFETAKKYNRIVADETNAKTLRILGAMRQITAHFKDQ-----	286
530373_GL0023589	NYFFDMLCDSAGNYQ---CKNAQILL SILGNLREQESFHENRS-----T	298
emb ODAI011611274.1	-----DKEEKROYKLMRVLICLLGELRQFLVHGKKK-----EKE	312
EMG_10003641	-----ILDGKRSKELKYYQELCCLLGMVRSQMIHSNQF-----	275
CasR_P1E0_metageno	-----DAARGVKREQDVYAILSLMSLLRQFCAHDSVRIWG-----QNT	261
emb OIZX01000427.1	-----SERRSNEDIFNIALMTCLRNFSSHHSIAIKVK-----DYS	289
160582958_gene49834	DYKTHLSNNQKEEWELKSDKEVYNIFAILCDLRHFCTHGESIT-----PSG	339
gi 1198542314 gb NFIRO1000008.1	-----ENKKEWKRFEKCYHLLAVLGMVRQATAHGDSK-----R	304

图18S

emb OGNF01009141.1	-----ITKTKIKSNEEIIYTFMLLSTARNFSAHYLDRNRAKS-----SDAEDFDGT	253
Ga0129306_1000735	KSVDDAIRALAPEWVKREERDVYSELVLMSELKQSCFHGQOK-----N	297
MH0288_GL0082219	ELHFREIEKLAPEVVRKSEKEIYYILAMLSELRCQVFDHSA-----S	344
PIG-022_GL0026351	-----SKKMPLISEKEMYIFALLGGLRQFCFTHDKNR-----A	252
PIG-028_GL0185479	-----NKKMPLISEKEMYIFALLGGLRQFCFTHDKNR-----A	254
250twins_35838_GL0110300	-----KESEKAKQDFVLRLLSVGRQITFHSDKSNNF-----	305
PIG-014_GL0226364	-KSEEEKRRIWQELWNKFNHNKNIIVNSLIKFRQCMCFHQYENS-----KEVS	259
250twins_36050_GL0158985	-----RAKEAETIDACTAHNYDVLRLLSLMRQICMHSVAGTAF-----KLA	309
emb OJMM01002900.1	-----DREIKNSYNALVLRVLYLRFQFCMHGNTYTKR-----NEESFLS	287
emb OGPN01002610.1	-----NNEYEKNLRHNFVLRVLSFLRQICTHAYVCKTGAKNNGDSTKVEAES	321
emb OHCP01000044.1	-----GDNSKQYENLRHNFNLRVLSFLRQICMHAHVHVS-----DEGCTRTQNYTDS	317
PIG-046_GL0077813	-----DDYEGKRNEILRYNFNVRVLSFLRQICAHAKVKVSS-----DNDREKGGVFDVS	320
pig_chimera	-----DDCKRERNEILRYNFNVRVLSFLRQICAHQAQVKISN-----EHDREKGGGLVDS	322
emb OIEN01002196.1	YKNVKGKEEKDNRNNEIEKLLKALEVIKIIIRVDSFHGVDGIKGD-----QKF	260
O2.UC29-0_GL0096317	EQ-----QIEQEREREKENLETLLKALDVISVMRTDSFHGYDMCNSTS-----AIVKF	253
CasR_Anaerobic_dig	-----GKGEKIEVMYRSDEEITIFQILSYVRQSIMHNDI-----GN	279
emb OGDF01008514.1	-----GRIVPCLEQRSDDIYNILRILSIVRQTCMHDNA-----SM	281
emb OGZC01000639.1	-----RKIVKKEKEVEVYHILLIGSLRQWITHTSTEG-----GIS	315
Ga0224415_10048792_chimera	-----NAKVROENQCFHIIALLISLRNWAHNSNELDK-----HTK	249
emb OCVV012889144.1	-----KKKRDKNPYILKYDNECYIIALLSGLRHWNIHSHAKDD-----LVS	247
Ga0187910_10006931	-----KYLNKQDNERHYHTIALLSGLRHWVHNNVEVES-----KID	258
R._flav_XPD_	-----NYIINYGNECYDIIALLSGLRHWVHNNNEEES-----RIS	253
Ga0224415_10007274	-----RNNRQKDPICCYHLLALLCGLRNWWVHNNNEEKD-----LIK	236
Ga0187911_10069260	-----ADADKEQVYAMLACLGLSLRQACSHYRIRYSVNG-----KNVDADA	245

图18T

Ga0187910_10015336	-----KVMRTEKEIYYILGMLNEVRNVSTHFTTEEDN-----RDWA	317
ODAI_chimera	-----LKYREAEDIYYIICMLGEARQMFAGHND-----DRT	313
31009_GL0034153	-----LKDKKDI FDMCLLGLSLRQCFHFHKGAKFEGV-----EQIS	252
DLF014_GL0011914_	-----KPKKVEVLKSEKEIFDILALLGSLRQCFHFYDEAVFENEDGV---IDKEYND	315
CasR_R_ albus	-----TKLRPEKEIYHIFALMASLRQSFYHGYVKDT-----DYQG	298
Ga0187910_10040531	-----RRDEKVIYHILSLAASVRQCFPHNDYTSDDGK-----GFIK	290
tpg DJXD01000002.1	-----RRYEEDI FALLLSALRQCFHSDLSSDES-----DHVN	235
CasR_E_ siraeum	-----AKNVLREDKELYSVLTIGKLRHWCVHSEE-----GRA	306
PIG-018_GL0023397	-----SKRRSEKEIYSMLAMLGKLRHWCVHSES-----GEA	287
PIG-025_GL0099734	-----KRSKREIYDILALIAIATIRQWCIHSEE-----DK	301
Ga0129317_1008067_chimera	-----FEKRSDEEVYNICALMGQIRQCCFHGKQ-----EKYQ	237
EYZ-362B_GL0088915	-----NQKKRLYYLMSMIGQLRQFAFHKDS-----PFN	295
uncultured_Ru_sp	-----VSQAYKKRVYHMLAIVGQIRQCVFHDKS-----GAK	299
R_ flav_ FD-1	-----KNEEIKKRLYHLTAFAGQLRQWSFHSAG-----NYP	304
tpg DBYI01000091.1	-----KSEEIKKRLYHLVAFAGQLRQWSFHSVE-----GLP	302
	* * *	
037_ _emb OIZA01000315.1	DIELYSENYE-E-----EYVFINCLNEFVKNKFNVNKN-F	328
emb OCTW011587266.1	R.IAVFKNQFDSD-----FSHITSTVNKIYSAKIAGVNEE-F	286
k87_11092736	SLAIFGANANMPI-----RFFQGTGGRQLWNDVIAPLWKKRIEVRKS-F	274
BMZ-11B_GL0037915	SLWASDLGVEK-----YFKNEM-SVVDANYKARAKKIDAN-F	288
BMZ-11B_GL0037771	YPWFKSSDIYET-----KIFKAGMW-SVIEADYKKKIQDVNKQ-F	287
BMZ-11B_GL0069617	-SILFCDNLKG-----Q--FLGNGNW-ITIEHYDEMITRINES-F	265
Ga0099364_10024192	-ATLFMSDVLEPK-----ILK--KEFSKADW-QTVEDYYAKLVDRINEG-F	327

图18U

530373_GL0023589	SAMFNLEKVL-----PERKLLTALIDSRTNHNN--HF	332
emb ODAI011611274.1	FAWLYRLDRQLS-----QEYRKLGEFYDAQVDKVNK--SF	346
EMG_10003641	NSSIYTLDSYDSTMNTAELLGKGGDSSLVALATDARVEARAILDEIYKKGVDSEINN--SF	334
CasR_P1E0_metageno	TAALYHLQALP-----QDMKDLLDDGWRRALGGVND--HF	294
emb OIZX01000427.1	AAGLYNLRRLG-----PDMKKMLDTFYTEAFIQLNQ--SF	322
160582958_gene49834	KPPFYNLEKNLF-----PEAKQVINSLFEEKAESLGAEAF	374
gi 1198542314 gb NFIR01000008.1	RAEYKLGKEFDRL-----EARGCRPEARKELDLYKKIHEMNQ--GF	346
emb OGNF01009141.1	SVIMYNLDNEELYK-----KLYNKKVHMLTGMKKVLDANFNKKVVEHLNN--SF	300
Ga0129306_1000735	SARIFRLDNDLG-----PGVDGARELLDRLYAEKINDLR--SF	333
MH0288_GL0082219	RSWLYNLEKVL-----AEAMQVLDGVYSQKVAALN--NF	377
PIG-022_GL0026351	RNWLYSLDTMLN-----DDAKAVLDKYYVSTVSSLNK--HF	286
PIG-028_GL0185479	RNWLYSLDTMLN-----DDAKAVLDKYYVSTVSSLNK--HF	288
250twins_35838_GL0110300	-AYLFDLSKLT-----AQDENRRQDIQSLNINLNSTCRSLEGVNGD--F	349
PIG-014_GL0226364	STALFNIEQFL-----NKEEKASLDEIYTGAIDKVNCD--F	293
250twins_36050_GL0158985	ESALFNIEDVL-----SADLKEILDEAFSGAVNKLNDG--F	343
emb OJMM01002900.1	DTALYNAKEFFAK-----ADPQINEILIDAVYADGKITSND--F	324
emb OGPN01002610.1	LDALFNITEYFAK-----TAPELSKTIENEIYKEGIDRINND--F	358
emb OHCP01000044.1	LEALFNISKAFGK-----KMPCLKTLIDNIYKSGINAINDE--F	354
PIG-046_GL0077813	LDALFNISFFDA-----VAPELNEVINSVYKGIYDINDN--F	357
pig_chimera	LDALFNISRFFDA-----VAPELNEVINSVYKGIYDINDN--F	359
emb OIE01002196.1	PRSKYNLAVN-----YNEEIQTKTISEPFNRKVEEVQQD--F	294
O2_UC29-0_GL0096317	PSKKFNLDQ-----KGKNLIYRNAFKYKTDIKKQ--Y	284
CasR_Anaerobic_dig	KSSIIAIEK-----YPARFVGFISDLLKTKTNDVNRM--F	312
emb OGDF01008514.1	RTVMFTLGQNSVR-----DRKNGFDELAELLDLYLYDEKIDIVNRD--F	322

图18V

emb OGZC01000639.1	RLWLYQLEDA-----LSREYQETMNNCYNSTIYGLQKD-F	349
Ga0224415_10048792_chimera	RTWLYKLDTTNI-----LNAEYVKTLNLYDYDTIADELTKS-F	285
emb OCVV012889144.1	YRWLYNLDSI-----LNREYISTLNLYDYDDIADELTKS-F	281
Ga0187910_10006931	RKWLNLNLDKN-----LDKEYITTLDYMYSDIADELTKS-F	292
R._flav_XPD_	RTWLYNLNLDKN-----LDNEYISTLNLYDRITNELTNS-F	287
Ga0224415_10007274	YTWLYNLNLDKY-----LDAEYITTLNMYNDIGDELTDTS-F	270
Ga0187911_10069260	DTWLFSSAQLDQ-----TDFLFSEMNLNRIYSHKIKITVNQN-F	281
Ga0187910_10015336	KANLYNLNLSNR-----LKVGSKEVLNQLYKEKIDKIDANGF	352
ODAI_chimera	KTFLYNIIVGYRN-----GTEINSKFTAVLDGIYSKKVEELNGN-F	352
31009_GL0034153	NSWLYNLSQ-----LHNEFKSTLDSFYNEKIEDINGN-F	285
DLF014_GL0011914_	NSWLYNLEK-----LEPEFKDTLDYFYNRKINDINKG-F	348
CasR_R._albus	TTWAYTLEDKLK-----GPSHEFRETIDKIFDEGFSKISKD-F	335
Ga0187910_10040531	ADWYRLEEA-----LPAEYKDTLDALYLEGVEGLDQS-F	324
tpg DJXD01000002.1	SFWLYQLEDQ-----LSDEFKETLSILWEEVTERIDSE-F	269
CasR_E._siraenum	EFWLYKLEDE-----LKDDFKNVLDVVYVNPVVEEINNR-F	339
PIG-018_GL0023397	EFWLYKLEDS-----LNGEFFIDVLTLYNRSVNEINRG-F	320
PIG-025_GL0099734	KTWLYNAENV-----LDSEYILDILDVYGGGLVEKINKN-F	335
Ga0129317_1008067_chimera	LKWLYNFHFK-----SNKPFPLDLDKHFDEMIDRINKN-F	272
EYZ-362B_GL0088915	RWLYYQLDEK-----LSKEHKDTLNYFFDQRYDEIHEN-F	329
uncultured_Ru_sp	RFDLYSFINN-----IDPEYRDTLDYLVEERLKSINKD-F	333
R._flav._FD-1	RTWLYKLEDS-----LDKEYLDTLDHYFFDKRFNDINDD-F	337
tpg DBYI01000091.1	RTWLYQLDNPK-----LAQYRDTLDYFFNERFDAINKD-F	337

图18W

037_-_emb OIZA01000315.1	I-SNE-KNNLYIILKAYGEDTENVEVVKYKSKELYKLSVLKTNKLNKLVNKKLRESAIEY	386
emb OCTW011587266.1	L-NNE-GNNLYIILKATNWD-----IKKI VQLYRFSVLKSDKNMGFNMRKLRFAVES	338
k87_11092736	L-SNS-AKNLWVLYQVFKDDTD--EKKKARARQYHFSVLKEGKNGFNLTKTREYFLDK	330
BMZ-11B_GL0037915	M-ATS-RANLVLLFRILNAKTK--EEKEKIAVDYKFAIRKEGKNGVNLKAKAREIILDK	344
BMZ-11B_GL0037771	F-SKN-KVNLAILFDLLDVRDV--KQKKRISDEFYRFTIRKDGKNGVNLVVKIREI I IDR	343
BMZ-11B_GL0069617	F-NNS-RMNLKILYLVAATL--EDKKKI TE EYYRFSILKEGKNGVNMKRLREMMVDM	321
Ga0099364_10024192	C-KNA-ATNVHFLTELLPE-----ESKKQLTEDYFRFAILKEGKNGVNMKRLREVMFAL	380
530373_GL0023589	I-ENS-NKNLNAIFRLFPEKS-----QADLIVEYYDFSVRKAYKNLGFVYKLLREMLLKT	385
emb ODAI011611274.1	L-TNS-TVNLEVLFRALKTGTD--PERKTVTQEYYQFTVRKEDGNLGFSLKTLREIILSA	402
EMG_10003641	L-SNS-INDLENLFKIYKCDSS--EKKTELIKQYYDFCIRKPPQMMGFSITTI REGMFTR	390
CasR_P1E0_metageno	L-DTN-KVNLLTLFEYYGAETK--QARVALTQDFYRFVVLKEQKNMGFSLRRLRELLKL	350
emb OIZX01000427.1	Q-DHN-TTNLTCLFDILNIDS--ARQQLAE E FYRYVVFKEQKNLGFVYKLLREMLLL	378
160582958_gene49834	G-KTAGKT DVSILLKVFEKEQASQKEQALLKEYYDFKQVQKTYKNMGFSIKKLR AIMEI	433
gi 1198542314 gb NFIR01000008.1	L-KNS-KSDIILMLFRIYNAESK--EAKRKLAEYEFIMLKS YKNTGFSIKHLRET MIDK	402
emb OGNF01009141.1	I-KNS-AKDFVILCEVLGIKSR--DEKTKFVKDYDFVVRKKNYKHLGFSVKELRELLFAN	356
Ga0129306_1000735	D-KTSASSNFRLLFNAYHADN---EKKKELAEFYRFSVLKVS KNTGFSIRTLEKIIED	389
MH0288_GL0082219	S-KTSSKSNFRLLFNALHADTD--AQKKQIAEDFYKFSVRKTYKNI GFSIKLLRESMLDT	434
PIG-022_GL0026351	V-DNSAKKDFSIIFNALNITDI--ELKKNIAKEYEFVCRKNFKNIGFSIKKLR EYLL EK	343
PIG-028_GL0185479	V-ENS AKKDFSIIFNALNITDI--ELKKNIAKEYEFVCRKNFKNIGFSIKKLR EYLL EK	345
250twins_35838_GL0110300	V-KHA-KNNLYVNLQLYPS-----LKANDLIGEYYNFTVKKENRNIGIRLITVRELIIEH	402
PIG-014_GL0226364	N-KNS-SNNLYILEQIYCN-----D--KNL GQRYDFVVKKSCQNLGLDTPKLI REL LIRK	344
250twins_36050_GL0158985	V-QHS-GNNLYVNLQLYPN-----ETIERIAEKYYRLTVRKEDLNMGVNIKKLR ELIVGQ	396
emb OJMM01002900.1	M-AHA-KNNMYIICEVYKN-----EAEDSLMKEYYDFVVRKEGNNLGFNTRQLREI LIDK	377
emb OGPN01002610.1	V-TNG-KNNLYILSKVYPD-----MQRNELVKKYYQFVVCKEGNNVGINTRKLRKESII SQ	411

图18X

emb OHCPO1000044.1	V-KNG-KNNLYILSKVYPN-----EKREVLRLREYFNVVCKEGSNIGISTRKLIKETMIAQ	407
PIG-046_GL0077813	V-KND-KNNFYILSKIYPE-----VARENLLKEYYFVVSKEGNNIGISTKKLKEAIIAQ	410
pig_chimera	V-KNG-KNNFYILSKIYPE-----VAREDLLREYFVVSKEGNNIGISTKKLKEAIIVQ	412
emb OIEN01002196.1	Y-RNS-CVNI DFLKEIMYGSNYTRDGSDSLSCSYFNFAILKQNKNMGFSITSIRECLLDL	352
O2_UC29-0_GL0096317	A-QMS-KVINFLSYVLYGNGCTEEQKKELSREYFNFAILKQYKNLGFSITSREQLLSL	342
CasR_Anaerobic_dig	IDNNS-QTNFWLFSIFGL-QDHTSGADKICRNFDYDFVIKADSKNLGFSLKKIRELMLDL	370
emb OGDF01008514.1	L-RNQ-KNNIELLSRIYGS-SADSPERDRLVQNFYDFRVLSDQKNLGFSSIKKLREKLLDS	379
emb OGZC01000639.1	E-KTN-APNLNFLAELGK-N-----ASELAEPYFRFIITKEYKNLGFSSIKTLREMLLDQ	401
Ga0224415_10048792_chimera	S-KNG-AVNNVYLAKKYNI-K---DDLPGFSEQYFRFSIMKEQKNLGFNISKLRENMLDF	339
emb OCVV012889144.1	S-KNS-SANVNYIAETLNI-D-----PSEFAQQYFRFSIMKEQKNMGFNVSKLREIMLDR	333
Ga0187910_10006931	S-KNS-AANVNYIAEILNI-D-----SKTFAEQYFRFSIMKEQKNLGFLLTKLRECM LDR	344
R._flav_XPD_	S-KNS-AANVNYIAETLGI-N-----PAEFAEQYFRFSIMKEQKNLGFNITKLREVMLDR	339
Ga0224415_10007274	S-KNS-AANINYYIAETLGI-D-----PKTFAEQYFRFSIMKEQKNLGFNITKLREVMLDR	322
Ga0187911_10069260	FENNR-KANFPILKKMYPE-----TTLKVLMEYDYDFSIKRGYKNFGFSIKSLREALLSP	335
Ga0187910_10015336	VNRGC-KRDFSILFKSLNL-T-TDKDKGELVVGFYDFSIKRNKYKNIGFSIKTLREYMLKI	409
ODAI_chimera	I-KNS-KKDLTILFSAFGC-Q-DEASKKKLIEKYYDFIVKKTNKNMGFSVTKLREKIIIEG	408
31009_GL0034153	I-EKN-KVNIQIILCNDLKL-----DKNKLVQDYFNFLISKNHKNMGFSIKKIREYILDI	337
DLF014_GL0011914_	I-KTN-QINLHIIICDELGL-N---WDKEQVVDYDFLIAKKYKNIGFSIKKIREHMFDI	402
CasR_R._albus	G-KMN-KVNLQIILEQMIGE-LYGSIERQNLTCDYDFIQLKKHKYLGFSIKRLRETMLLET	392
Ga0187910_10040531	L-KNN-TVNIQIILCSIFNH-----DDPNKIAEYDFLMTKEYKNMGFSIKKLRECMLEL	377
tpg DjXD01000002.1	L-KTN-TVNLHLCHVFPK-----ESKETIVRAYEFLIKKSKFNMGFSIKKLREIMLEQ	322
CasR_E._siraenum	I-ENN-KVNIQIILGSVYKN-----TDIAELVRSYEFELITKKYKNMGFSIKKLRESMLEG	392
PIG-018_GL0023397	V-DTN-KVNIQIILQAVCKN-----MDLISLIKAYEFLITKKYKNMGFSIKKLREMLLAE	373
PIG-025_GL0099734	I-KDN-KVNLILLSDILDM-K-SKDDFEELIRQYRFRFIITKEQKLLGFSITKIREAMLED	391

图18Y

Ga0129317_1008067_chimera	I-KNN-TPDLIILSGLYPD-----MAKKELVRLFYDFTTVKEYKNMGFSVKKLREKMLE	325
EYZ-362B_GL0088915	N-KTN-KVNLELLAEAFPS-----EDYATLANQYDYDFVFFKAHKNLGFSSIKKLRKLIEN	382
uncultured_Ru_sp	I-EDN-KVNISLLIDMMKG-----YEADDIIRLYYDFIVLKSQKNLGFSSIKKLRKMLDE	386
R_flav_FD-1	V-TKN-ATNLYILKEVFPE-----ANFKDIADLYYDFIVIKSHKNMGFSIKKLRKMLEC	390
tpg DBYI01000091.1	I-ETN-NINLHILKEVFPA-----EDFQLAALYYDFIVKKTFFKNIGFSIKNLRQMLEC	390
	.. : : * . : *	
037_-_emb OIZA01000315.1	GYCPLP-----YDKEKEVAKLSVVKHKLYKYDFVITHYLSN-----	424
emb OCTW011587266.1	KNIDL-----SRLNDKFLTNRRKKLYKVDFIIYYHLNKKVL-----K	375
k87_11092736	FFPIFH-----SSAPDVKRKYDTFRSKFYAILDFIIYEASVSV-----ANSGOMGKVAPW	380
BMZ-11B_GL0037915	HYPD-----AKNKQFDSYRGKLYTITDFIIRMAAED-----	376
BMZ-11B_GL0037771	YASG-----LRDKKHDPHRQKINVIADFLIFRALSQN-----	375
BMZ-11B_GL0069617	FYSE-----IKDKQYDSYRQKIYITILDLLFREIKDS-----	353
Ga0099364_10024192	FVPELT-----APETKKRYDSYRAKIYGLTDFLLFKHIHNT-----	416
530373_GL0023589	QQELS-----PLNDKAYDSIRSKLYTMDFVLYKLYVD-----N	419
emb ODAI011611274.1	YKH-----EVRDKREYDSIRHKLYQLFSFALYHYKYG-----VG	436
EMG_10003641	CSEANTLLCDEGSTVKNVHDTMKSIFYKNLDFMIYKYKYE-----N	434
CasR_P1E0_metageno	PDA-----YLTGQYDYSVRQKLYMLLDFLLCRLYAQE-----R	384
emb OIZX01000427.1	PDA-----VIADKRYDTCRSKLYNLMDFLILRVYRTG-----R	412
160582958_gene49834	PDA-----KFKDDLSSLRHKLYGLFDFILVKHFLDT-----SDSE	470
gi 1198542314 gb NFIR01000008.1	MDEDK-----KEKLKDDKYNPIRRKIYRIMDFMIYQYQFP-----EH	440
emb OGNF01009141.1	HDSNKYIK---EFDKISNKKFDSVRSRLNRLADYIIYDYNNK-----N	396
Ga0129306_1000735	HAA-----QYRDKIYDSMRKKLFSFTDFLWRFYEE-----R	421
MH0288_GL0082219	FAP-----EYKEKSYDYSVRGKLYQLVDFLIWNIYDK-----N	466

图18Z

382	PEFC-----NLKEDKYNVSRSKLYRLIDFIIYLNKYDKAKF-----ENDN	FIG-022_GL0026351
384	PEFR-----NLKDDKYNVSRSKLYRLIDFIIYLNKYDKAKF-----KNDN	FIG-028_GL0185479
434	NYT-----NLKDSKYD TYRNKIYTVLNFILFREIQEN-----	250twins_35838_GL0110300
379	LSD-----EGLDINDSRDYTYRNKIYTVFGFIIYYELTKN-----	FIG-014_GL0226364
437	YFPE--VLDKEYDLSKNGDSVVTYRSKIYTMVNYIILLYLEDH-----	250twins_36050_GL0158985
409	YVG-----NLRGKKYNTFRNKLYTVLGFILVKELKRN-----	emb OJMM01002900.1
449	HPWI--TTPQD---NNKANDYESCRHKLYTIMCFILVAELDAH-----	emb OGPN01002610.1
437	NMP-----SLKEENTYRNKLYTMNFIIVRELKNC-----	emb OHCPO1000044.1
442	DMS-----YIKSKDYD TYRNKLYTVLFCFILVKELNER-----	FIG-046_GL0077813
444	DMS-----YIKSEDYD TYRNKLYTVLFCFILVKELNER-----	pig_chimera
383	YEL-----NFESMQNLRPRANSFCDFLIYDYCKNE-----	emb OIEN01002196.1
373	YDI-----DFECKRNLRSRANSFCDYVLYRYAVQF-----	O2_UC29-0_GL0096317
403	PNA-----NMLRDHQD TVRSKFYTLDDFIIYQHYLEE-----	CasR_Anaerobic_dig
412	PAL-----SVVRSKKYD TMRSKIYSLIDFMIYRKFSEN-----	emb OGDF01008514.1
438	PDLQ--E-----IRENHNVYDSIRSKLYKMDIFVLVYAYSNERK-----	emb OGZC01000639.1
376	KDMS--V-----IRDDHNRYDKDRSKIYTMDFVIYRYIIDNNN-----	Ga0224415_10048792_chimera
385	KELS--D-----IRDNHRVFD SIRS KLYTMMDFVIYRYIEEAAKTEAENRNLPENEKK	emb OCVV012889144.1
396	EELS--D-----IRDNHKVFDSIRSKLYTMMDFVIYRYIEEAKKIENENKTLSDDKKK	Ga0187910_10006931
391	KDMS--E-----IRKNHKVFDSIRTKVYITMMDFVIYRYIEEDAKVAANKSLPDNEKS	R_flav_XPD_
374	KDMS--E-----IRENHNDFDSIRAKVYITMMDFVIYRYIEEAAKVNAANKSLPDNEKS	Ga0224415_10007274
373	QYES--LIGV---QIKDNKEYD TVRSKLYQLDFALTRYFNQH-----	Ga0187911_10069260
443	SNS-----T-----LCADTISNNAIRPKAYKLYDFIIWHYMNK-----	Ga0187910_10015336
443	YRFR--G-----NPVTGTDYDSVRHKL YRTDFIICSYYDDH-----	ODAI_chimera
371	DYA-----SYIKEQGYDSVRSKFLTLIDFIIYDLYVNKE-----	31009_GL0034153

图18AA

DLF014_GL0011914_	YEA-----FDIKDRFDSVRSIMYKIIDFVIYYSFIHYK-----	436
CasR_R._albus	TPA-----ECYKAFCYNSERQKLYKLIIDFLIYDLYNRK-----	426
Ga0187910_10040531	PEL-----SGYKEDQYNSVRSKLYKLFDFIIAHYFRKH-----	410
tpg DJXD01000002.1	SDL-----KSFKEDKYNVRAKLYKLFDFIITYYVDHH-----	355
CasR_E._siraenum	KGY-----ADKEYDSVRNKLYQMTDFILYGTYNED-----	423
FIG-018_GL0023397	TAYT--SNPD---DRLYDKKEINSVRNKLYQMTDFIILEYRDN-----	411
FIG-025_GL0099734	TV-----FKTDKDYDSIRSCLYKLIIDFMMFYKYTTVDSE-----	425
Ga0129317_1008067_chimera	EEA-----SDFRDKDYDSVRRKLYKLMDFCIYLYYS-D-----	358
EYZ-362B_GL0088915	-----ALRDEKIDFSRSKLYRLIIDFCLFRIYLIK-D-----	411
uncultured_Ru_sp	YG-----FRFKDKQYDSVRSKMYKLMDFLLFCNYRND-----	419
R._flav._FD-1	DGA-----DRIKEQDMSVRSKLYKLIIDFCIFKYHE-F-----	423
tpg DBYI01000091.1	DEA-----EKIRSKDMNSVRSKLYKLFDFCFIYQYFI-D-----	423
	: : :	
037_-_emb OIZA01000315.1	DKLLEIVETLRISK-----ND-DEKENVYKKYA---E-KLKFKADDVINPIKAISKLFAE	474
emb OCTW011587266.1	DSFVDDFVAALRASQ-----SE-EEKEKLYAQYS---E-RLFADEGLKSAIKKAVDMISD	425
k87_11092736	KGAIDNALVKLREAP-----DE-EAKEKIYNVLA---A-SIRNDSLFL--RLKSACDKFG	428
BMZ-11B_GL0037915	GAFLEESVQKLRSTP-----SE-EEKEKYVDVLA---E-TIWNATKDD--VKKVFDIRQQ	424
BMZ-11B_GL0037771	QGIIIDTVSSSLRLTK-----DE-EEKDHVYQNA---E-LVWGMVSN---LTPYFNDFPKN	423
BMZ-11B_GL0069617	-EELETMVADLRESE-----SD-EKKEELYHDYA---E-LFWDGVKDL--IVPFFAAFDG	400
Ga0099364_10024192	-KQLEEWAVLRETS-----NE-DAKENLYDEFA---R-TAWNVTGDS--AKQLIENMQS	463
530373_GL0023589	EEQIDQVVRKLRASST-----SN-VEKEALYQOES---E-RTWNEIRGQ--IIDNVIPFID	467
emb ODAI011611274.1	AERREAFVAKLRAMV-----TA-EAKQRAYADEA---A-EIWNDEGSG--IRAAFLIILE	484
EMG_10003641	PEKGEKLIEDLRSKI GKKKED-EDKKQRYAEES---A-CILKAKRDI--IKKDLTEAAN	487

图18BB

CasR_PIE0_metageno	ADRCFELVSALRCAL-----SD-EEKDVIYQAEA---A-ALWQALGDT--LRRKLLPLLK	432
emb OIZX01000427.1	ADRCDKLPEALRAAL-----TD-EEKAVVYHKEA---L-SLWNEMRTL--ILDGLLPQMT	460
160582958_gene49834	NLQNDIFRQLRACR-----CE-EEKDOVYRSIA---V-KVWEKVKKK--ELNMFQVVV	518
gi 1198542314 gb NFIR01000008.1	QEEAEELVRKLRAE-----IE-AKKELAYRKEA---E-KLKELEKI--IFNSVLPSCD	488
emb OGNF01009141.1	NAKVSDDLKYLRAAA-----DD-EQKKKIYLNES---INLVKSGILER--IKKILPKLNG	445
Ga0129306_1000735	EDEAEELRACLRAAR-----SD-EEKEQIYAEAA---A-SCWPSVKPF--VESVAATLCD	469
MH0288_GL0082219	EKMADALVAQLRAEVK---RDR-DEKSPVYDREA---A-RLWTMIREA--VAQTAKTLRA	516
PIG-022_GL0026351	DEKIERFVSKLRKCP-----DE-ASKETVYQTES---D-IIKKSIYGS--INSIKRGM--	429
PIG-028_GL0185479	DEKIERFVSKLRKCP-----DE-ASKETVYQTEA---D-IIKKISGS--INSIKRDM--	431
250twins_35838_GL0110300	SIATKNFREKLRSTE-----K--AEQPALYQAF--N-KIYPMVQAK--FAKAIDLFE	481
PIG-014_GL0226364	ETFCQEIIVAKLRRFG-----GDEIEKNKIYNEYFLRKE-AVYNQIAEK--FKQSVCFKK	431
250twins_36050_GL0158985	DSSRESMVEALRQNR-----EGDEGKEEIIYRQFA--K-KWVNGVSGL--FGVCLNLFKT	486
emb OJMM01002900.1	PKIQDSFIAKLRANQ-----NGDEGLNIYNEFA--P-KIWSVVSSK--LNSAITCFDE	458
emb OGPN01002610.1	ESIRDNMVAELRANM-----DGDDGRDAIYEKYA---K-DIYHIVKDK--LLAMQKVFE	498
emb OHC01000044.1	ATIREQMIKELRANM-----DEEGRDRIYSKYA---K-EIYLYVKDK--LKMLNVFKE	486
PIG-046_GL0077813	TTIREQMVADLRANM-----NGDIGREDIYSKYA---K-IYYAQVKPR--FDGMKNVFEE	491
pig_chimera	TTIREQMVADLRANM-----NGDIGREDIYSKYA---K-IYYAQVKPR--FDTMKSAFEE	493
emb OIEN01002196.1	-SERANLVDCLRSAA-----SE-EEKKNIIYFQTA---E-RVKEKFRNA--FNRIISRFDAS	430
O2_UC29-0_GL0096317	PERIVDLVDKLRCSA-----NE-KEKTNVYIEEA---N-RIKEDLRIA--FNRINNFDAK	421
CasR_Anaerobic_dig	KSRIDNMVEKLRMTL-----KE-EEKEVLYAAEA---K-IVWNAIGAK--VINKLVPMMN	451
emb OGDF01008514.1	HVAVDDFVEELRSL-----TE-DEKESAYSRWA---ETLINDGFAQE--ILVKLLPQTD	461
emb OGZC01000639.1	-SKADALASNLRSAI-----TE-DAKKRIYQNEA---D-QLWTSYQEL--FKRIRGFKGA	485
Ga0224415_10048792_chimera	--DSIDFINKLRSSI-----DE-KSKEKLYNEEA---N-RLWNKLLKEY--MLYIKEFNGK	422
emb OCVV012889144.1	ISEKDFVFINLRGSF-----DE-NQEKELYIEEA---K-RLWEKLLKDI--MLKIKEFRGE	433

图18CC

Ga0187910_100006931	LSEKDIFIISLRGSF-----SE-EQKDKLYSDEA---E-RLWAKLGLK--MLEIKKFRGQ	444
R._flav_XPD_	LSEKDFVINLRGSF-----ND-DQKDALYYDEA---N-RIWRKLENI--MHNIKEFRGN	439
Ga0224415_100007274	LSEKDFVISLRGSF-----NE-DQKDRLYYDEA---Q-RLWSKVGKL--MLKIKKFRGK	422
Ga0187911_100069260	PDMVDAFVVELRSLA-----KDEDAKNAVYEKYA---K-AVWNDVKQP--IAVMLSVMNG	422
Ga0187910_100153336	PKINDFVEKLRTON-----KN-DEKIKLYYDEA---V-CLLSELGRE--IHTMTSCVHN	491
ODAI_chimera	AKEKDTLVSSLRASL-----ND-TKKENLYAGEA---K-TIAESIKDR--MDTVCRMYNK	491
31009_GL0034153	KDRIDDIVSKLRGAI-----YD-YKKDDIYKAEA---D-WVWKKLNRQ--IKALTKLLSN	419
DLF014_GL0011914_	QDEVDEIVNYLRSTL-----SE-EQKKEKYIKLA---G-NIWAELYKTQ--FNKLDKILLA	484
CasR_R._albus	PARIEEIVDKLRESV-----ND-EEKESIYSVEA---K-YVYESLSKV--LDKSLKNSVS	474
Ga0187910_10040531	PEKGEEMVDCLRLCM-----TE-DEKDSHYEGTA---K-KLVRELAYD--MQEAAEQANG	458
tpg DJXD01000002.1	AFEKEALVSSLRSSL-----TE-ENKEEIIYIKTA---R-TLASALGAD--FKKAAADVNA	403
CasR_E._siraenum	SDRADDLVNTLRSSL-----KE-DDKTTVYCKEA---D-YLWKKYRES--IREVADALDG	471
PIG-018_GL0023397	KSIIDTLVSKLRSSV-----TD-EAKEAVYREEA---A-NLRGRLLDA--VALIAKSLNG	459
PIG-025_GL0099734	KEHADSIVEKLRATL-----SG-ESKEEIIYKAEA---N-RIWNLYEDI--IMNKIEPALD	473
Ga0129317_1008067_chimera	SERNENLVSRLESLSL-----TD-ENKDIIYSKEA---K-IVWNELRKK--FSTILDNVKG	406
EYZ-362B_GL0088915	PDRAEKNVRF LRSCI-----NE-QEKEQFYTQEA---N-ALKSENPGI--VNDFCMQLKD	459
uncultured_Ru_sp	IAAGESLVRKLRFSM-----TD-DEKEGIYADEA---A-KLWGKFRND--FENIADHMNG	467
R._flav._FD-1	PELSEKNVDIILRAAV-----SD-TKKDNLYSDEA---A-RLWSIFKEK--FLGFCDKIVV	471
tpg DBYI01000091.1	EERSRENVNYLRSTL-----ND-EQKDAFYEEEG---K-RLWSENRRK--FIYFCDNINK	471
	** : *	
037__emb OIZA01000315.1	KGNKLFKEK-----IIIKKEYIEDVSDKNIYDFTKVIFFMTCF	513
emb OCTW011587266.1	TKSNI FKMK-----TPLDKALLENIKVNSDASDFCKLIYVFTRF	464
k87_11092736	AEQNRVPVFPNELR-----NNRDIRNVRSEWLEATQDVDAAFVQLIAFLCNF	475

图18DD

BMZ-11B_GL0037915	DIKNAAAQS-----LDDVNLDVTVTEKEKEFPVVKLLAFLCNF	462
BMZ-11B_GL0037771	KYLLKYKDA-----KTPGDFEDWITSKI SEDDGEFFVKVLSFLCNF	464
BMZ-11B_GL0069617	DFTVFTEK-----VPADTSKEPFLSSANTEPLVKLLAFLCNF	438
Ga0099364_10024192	YFTKKEKIT--K-----TAQPVLSTSSIAHTSKKITQFSSFAKLLAFLCNF	508
530373_GL0023589	SIQKQENEDKLTKR-----IDQSIIDKAVPHFVNSRDLSYFAKVVYVVASF	513
emb ODAI011611274.1	AVDFGSAVGIKARSS-----VAGDKRFAEWELEEVRIPEGVSCFTKLMYLLTRF	534
EMG_10003641	KDLFADLVKSNKNEKQK-----FKNEYELLKPFMIPVKVDYFSELIYLVTRF	535
CasR_P1E0_metageno	GKKLQDKDK-KKSDDELGLSRDVLGDVLF-----PAQQGSRANADYFCRLMHLSTWF	483
emb OIZX01000427.1	PENLSRLSGQKRKGELSLDDAMLKECLYE---PGVPVEDAAPEEANAAYFCRMILYLATLF	517
160582958_gene49834	IPSLSKDE-----LKQEMTKNTELLSSIEITISTQASLFSEMIFMMTYL	562
gi 1198542314 gb NFIR01000008.1	RILSEMERRNKKVQESSDTDKEEPL-----DSEIAEGITFIKETAHSFSEMIYLLTVF	543
emb OGNF01009141.1	KIIGNMQPDSTI-----TASMLHNTGKDWHPISENAHYFTKWYTLTLF	489
Ga0129306_1000735	VVKGR-----TKLNKLSADESTLVRNAIDGVRISPRASYFTKLIYLMTLF	516
MH0288_GL0082219	VTLNNTPGSIKIKELALSKYDRDLVENAIDSAVDERGRRRLKITDEASYFSKLIYLLTLF	576
FIG-022_GL0026351	-----SVKSYGKLELDKQTIQNLKRAVDDDISINTKGDYFTKLMYLISLL	474
FIG-028_GL0185479	-----LVKSYGKLELDTQTVNQLKRAVDDDISINTKGDYFTKLMYLISLL	476
250twins_35838_GL0110300	QYKTKFSEFKGGI-----SIENMQQNILLQTENIDYFSKYVLFITKF	525
FIG-014_GL0226364	EKEKKEFSS-----KTKIDLRSVQKIEDQNASVFAMVYLLISRF	471
250twins_36050_GL0158985	EKNKFRSKV-----ALPDVSGAAYMLSSENIDYFVKMLFFVCKF	526
emb OJMM01002900.1	ESLSKFKGYK-----DIDESLISRYGITVANTDTLVKLIYFLCKF	498
emb OGPNO1002610.1	ELVPVKVEGKNDPQ-Q-----FTHGKLGKKEIESFCLSDKNTSDIAKVVYFLCNF	547
emb OHCP01000044.1	EAEGLIIPGKEDPV--K-----FSHGKLDKKEIESFCLTTKNTEDITKVIYFLCKF	535
FIG-046_GL0077813	EAKGVIVPDKEEAV--K-----FSNGALDKSEIERFCITTSANTDSVAKIYFLCKF	540
pig_chimera	EAKDVIVPDKPKPV--K-----FSHGKLDKNEIERFCITTSANTDSVAKIYFLCKF	542

图18EE

emb OIEN01002196.1	YIKNSREKNLSGGS--SLPK-----YSFIEGFTKRKSKINDNEKNADLFCNMLYYLAQF	483
O2.UC29-0_GL0096317	SIKLYSTKNNVDDQ--PMLP-----FEQVSK----DIKGGDAAREKADAFCDMTYYLITQF	470
CasR_Anaerobic_dig	GDALKEIKRKN-----RDRKLPQSVIATVQVNSDANVFSGLIYFLTLF	494
emb OGDF01008514.1	PAVIGKIKG-----KKLLNDSIAGIKLKKDASFFTKIINVLCMF	500
emb OGZC01000639.1	QVKEYSSKN-M-----IPIQKIQNILKPAEQVTYFTKLMYLLTMF	526
Ga0224415_10048792_chimera	LASRTPDRD-G-----NISEFVESLPKIHRLPRGQKISNFSKLMYLLTMF	467
emb OCVV012889144.1	KVKEYK-----KEVPRIERILPQKDISAFSKLMYMLSMF	468
Ga0187910_100006931	MTRDYK-----KSDTPTLNRILPESEDEVSTFSKLMYALTMF	480
R._flav_XPD_	KTREYK-----KKDAPRLPRILPAGRDVSAFSKLMYALTMF	475
Ga0224415_10007274	DTRKYK-----NMGTPRIRRLIPEGRDISTFSKLMYALTMF	458
Ga0187911_10069260	SAIKNIKAFEL-----KPDQKELNGIMNSNALDVPHFCKLVYFLTRF	464
Ga0187910_10015336	IENTSYEITDK-----KQKEYYKMQINSLNSADKVSDFSKVIYLVTLF	534
ODAI_chimera	GEGIKHITVDR-----FIAGAGAGKLIDKCSISTECDYFCFKFIYLMTLF	535
31009_GL0034153	RRIFSQDND-----ITNNGVEFEKIGEQLKLEDNADYFCFKLMYFVTLF	462
DLF014_GL0011914_	NIKEFSDAKEDDV-----YEEFKNFKFNVEGKELGENAEYFCFKLMYLLTLF	532
CasR_R._albus	GETIKDLQKRY-----DDEANRIWDISQHSISGNVNCFCFKLIYIMTLM	518
Ga0187910_10040531	SNITQMKNQGG-----KT-KGMFAIRDEIRVSRKPKVSYFSKVIYVMTLL	503
tpg DJXD01000002.1	KNIRDYQKKAN-----DYRISFEDIKIGTIGYFSELIYMLTLL	443
CasR_E._siraenum	DNIKKLSKS-----NIEIQEDKLRKCFISYADSVSEFTKLIYLLTRF	513
PIG-018_GL0023397	VNIKKLQON-----TLEIPDSKLENCFKSAEKVSAFTKLNLYLLTRF	501
PIG-025_GL0099734	EGVISGLKK-----NKSYDNMSIEGIVAEETANVSYFSKIIYLLAQF	515
Ga0129317_1008067_chimera	SNIKKLEN-----VKEKFISEDEFDDIKLIDIDISYFSKLMYVMCYF	447
EYZ-362B_GL0088915	RYSSGNSSTENAE-----KKEDEFPVFPCHDKQKVSNSVSLFAKLMYAFCLF	505
uncultured_Ru_sp	DVIKELGK-----ADMDFDEKILDSEKKNASDLLLLYFSKMIYMLTYF	508

图18FF

```

R._flav._FD-1          WVTGEHEK-----DITSVIDKDAYRNRNSVYFSKLMYAMCFF      509
tpg|DBYI01000091.1|   WKNDYS-----EVAKCIDLNEFRVNSVYFSKLLYAMSFF      509
                        .
                        :
037_-_emb|OIZA01000315.1| LDGKEINDLLTNIISKLQVIEDHNNVIKFIH-----NKD      548
emb|OCTW011587266.1|  LDGKEINILLNSLIKKFQDIHSFNTVKKLSE-----NNL      499
k87_11092736          LEGKEINELVTALIKKFEIGIQALIDLRRNLEGV-----      508
BMZ-11B_GL0037915     LDGKEINELLSTFVNRLENVQALIDLAKNLGET-----      495
BMZ-11B_GL0037771     LEGKEINELLTAYIHKFECIQDFLNVISLGEN-----      497
BMZ-11B_GL0069617     WDGKEVNEVLSAYIHKFESIQAFIDVLESKELN-----EKV      474
Ga0099364_10024192    WEGKEINELLSAYIHKFENIQEFINLLEKLEGK-----      541
530373_GL0023589      LDSKEINMFFDALINELENIASFNDVLEQLDM-----      545
emb|ODAI011611274.1| LDGKEINELLTGLINKLENIQSFLDVMQOEHA-----      566
EMG_10003641          LSGKEINDLLTQLINKFENIAAFIRMYQNDQG-----      567
CasR_P1E0_metageno    MDGKEINTLLTTLISKLENIIDSLRSVLESMGL-----      515
emb|OIZX01000427.1|  MDGKEINTLLTTLISKFENIAAFLQTMEQLNI-----      549
160582958_gene49834   LDGKEINLLCTSLIEKFENIASFNEVLKSPQIG-----      595
gi|1198542314|gb|NFIR01000008.1| LDGKEINILLTQLIHCFDNISFMDTMEENL-----      575
emb|OGNF01009141.1   MDGKEINDLVTTLINKFDNIASFIEVLKSQSV-----      521
Ga0129306_1000735     LDGKEINDLLTTLIHFENIDSFLSVLGSERL-----      548
MH0288_GL0082219     LDGKEINDLITTLVNSFDEIDSFCCVLQKAKL-----      608
PIG-022_GL0026351     IDGKEINDLLTTLIHFENIASFIDVMKKENI-----      506
PIG-028_GL0185479     IDGKEINDLLTTLIHFENIASFIDVMKKENI-----      508
250twins_35838_GL0110300 LDGKEINELLCALINKFDNIADLLDISKQIGT-----      557

```

图18GG

503	LDGKEINELCTKLLINKFKINELVETAKRCGE-----
558	LDGKEINELCALINKFDNIADILDAAAQCGS-----
530	LDGKEINELCCAMINKFDNINDLIKTAQAQCGE-----
579	LDGKEINELCCAMNKFDDGIGDLIDTAKQCGE-----
567	LDGKEINELCCAMNKLDSILDIETAKQCGE-----
572	LDGKEINEFCCAMNKLGDINDLMEAEQCGA-----
574	LDGKEINELCCAMNKLGDINDLIETAEQCGA-----
515	LDGKEINIFLTSIHNIHQNIDSFLKVMKEKGM-----
502	LDGKEINVFLTTLHNLFENIASFLDVMHKQNI-----
554	LDGKEINEMVSNLITKFENIDSLLVHDREIYKSDEKLDLLEIEKLAFFKGVVRPNAKTD
532	QDGKEINELVSSLVNKFANIQSFDVVMRSQGI-----
558	LDGKEINDLLTTLINKFDNISLLKTMEQLEL-----
499	LDGKEINDLLTTLINKFENIQGFDDIMPEINV-----
500	LDGKEINDLFTTLINKFDNIQCFLKIMPLIGV-----
512	LDGKEINELTTLINKFDNIQSMLKIMPLIGV-----
507	LDGKEINDLLTTLINKFDNIQSFLKVMPLIGV-----
490	LDGKEINDLLTTLINKFDNIQSFLKVMPLIGV-----
496	LDGKEINDLLTTLINKFDNIHSFNQVLTALGL-----
566	LDGKEINDLLTTLINKFDNIASLLSVLEKQSG-----
567	IDGKEINDLLTTLINKFDNIDSPSVDKAVNNQ-----
494	LDGKEINDLLTTLINKFDNIRSFIEVMEEMSL-----
564	LDGKEINDLLTTLINKFDNIRSFVELMKEENI-----
550	LDGKEINDLLTTLINKFDNIASFIDVDELGL-----
535	LDGKEINDLLTTLINKFENIVSFEDVLRQLNV-----
	FIG-014_GL0226364
	250twins_36050_GL0158985
	emb OJMM01002900.1
	emb OGPN01002610.1
	emb OHCPO1000044.1
	FIG-046_GL0077813
	pig_chimera
	emb OIEN01002196.1
	O2_UC29-0_GL0096317
	CasR_Anaerobic_dig
	emb OGDF01008514.1
	emb OGZC01000639.1
	Ga0224415_10048792_chimera
	emb OCVV012889144.1
	Ga0187910_10006931
	R._flav_XPD_
	Ga0224415_10007274
	Ga0187911_10069260
	Ga0187910_10015336
	ODAI_chimera
	31009_GL0034153
	DLF014_GL0011914_
	CasR_R._albus
	Ga0187910_10040531

图18HH

tpg DJXD01000002.1	LDGKEINDLLTTLINKFDNIISFIDILKKNL-----	475
CasR_E_siraeum	LSGKEINDLVTTLINKFDNIRSFLIMDELGL-----	545
PIG-018_GL0023397	INGKEINDLLTTLINKFDNIRSFLIMDELGL-----	533
PIG-025_GL0099734	IDGKEVNDLTTTLINKFDNIRSFIDTANQIGL-----	547
Ga0129317_1008067_chimera	LDGKEINDLLTTLVSKFDNIGSIIIEAATQIGI-----	479
EYZ-362B_GL0088915	LDGKEINELLSALLNKFDTTIACMTKCTKDLGI-----	537
uncultured_Ru_sp	LDGKEINDLLTTLISKFDNIKEFLKIMKSSAVD-----	541
R._flav._FD-1	LDGKEINDLLTTLINKFDNIANQIKTAKELGI-----	541
tpg DBYI01000091.1	LDGKEINDLLTTLINKFDNIRSFIDTANFLNI-----	541
	..**:* . : :	
037_-_emb OIZA01000315.1	AVYKDYDKYAIFR-----NA-----GKIATELEAIKSI	577
emb OCTW011587266.1	IINADYVDDYSLFE-----QS-----GTVARELMLIKSI	528
k87_11092736	-DSIRFENEFAFN-----DD-----KGNMAGRIARQLRLLASV	541
BMZ-11B_GL0037915	---VNFSKNYAVFN-----DC-----GSGAAGRIARLRLLASI	526
BMZ-11B_GL0037771	---VQFPREFALFN-----NA-----SFAQNVAVQLRILASI	526
BMZ-11B_GL0069617	VFSAKKGRNFSAFN-----EH-----GFAGRVADQLRVLASI	506
Ga0099364_10024192	--KPQFTENYALFN-----EA-----AGQRAGEIAQNLRLIASI	573
530373_GL0023589	--PLDYQEKYRFFE-----RS-----REEAAGLRFIKSI	572
emb ODAI011611274.1	--ETGLSDAFSFFE-----YS-----GEIAAELRMTRSF	593
EMG_10003641	--KLEFTANYKMFIEDPQKDIPKDKRVLSGS-----AKIAYLRTINYI	610
CasR_P1E0_metageno	--AYSFVPYAMFD-----HS-----RYIAGQLRVVNNI	542
emb OIZX01000427.1	--EAELGPEYAMFT-----RS-----RAVAFQLRVINSF	576
160582958_gene49834	-YETKYTEGYAFFK-----NA-----DKTAKELRQVNNM	623

图18II

```

gi | 1198542314 | gb | NFIR01000008.1 | 602
emb | OGNF01009141.1 | 548
Ga0129306_1000735 | 575
MH0288_GL0082219 | 635
PIG-022_GL0026351 | 533
PIG-028_GL0185479 | 535
250twins_35838_GL0110300 | 584
PIG-014_GL0226364 | 530
250twins_36050_GL0158985 | 585
emb | OJMM01002900.1 | 557
emb | OGPNO1002610.1 | 606
emb | OHCPO1000044.1 | 594
PIG-046_GL0077813 | 599
pig_chimera | 601
emb | OIENO1002196.1 | 542
O2.UC29-0_GL0096317 | 529
CasR_Anaerobic_dig | 583
emb | OGDFO1008514.1 | 559
emb | OGZCO1000639.1 | 585
Ga0224415_10048792_chimera | 526
emb | OCVV012889144.1 | 527
Ga0187910_10006931 | 539
R._flav_XPD_ | 534
Ga0224415_10007274 | 517
--LTKLKEDYEIIFE-----ES-----KEISRELRIINSF
--CTHFSEERKMFI-----DS-----AEICSELSAMNSF
--ERTFDANYRIFA-----DS-----GVIAQELRAVNSF
--DRGFTEGYRFFA-----HS-----AEISQELRAINSF
--NYEFAEKYKFFN-----TS-----KNIACELRTVNSF
--ECEFTKFAFFK-----NS-----EIIAHELRTVNSF
--PVVFCADYESLN-----DA-----AKIAENIRLIKNI
--TIEFCPEYKMFK-----NS-----AKLABEMRFVQSV
--SVWFVDSYRFFE-----RS-----RRISAQIRIVKNI
--DIEFVKEYKLF I-----NS-----NDLSQIRIVKSI
--EVKFIEEFACLS-----NC-----RKITNDIRVAKSI
--DVEFVDQFKCLS-----KC-----ATMSNQIRIVKNI
--KIEFVDNFNVL S-----KC-----GTISGQIRVVKSI
--KVEFVDKFSVLS-----NC-----ETISDQIRIVKSI
--ECKFQKDFKMF S-----HA-----GHVAKKIEIVISL
--SYDFTEQYVMFY-----DS-----AYIVKKLETVISV
TGAGEISKSFSIFQ-----SA-----ERIIELKFIKNV
--DSGFTADYAMFA-----ES-----GRISRELHILKGI
--QTTFKEDYTFQ-----QS-----SRLCKEITQLKSF
--NAKFEPEYVFFN-----KS-----HEIAGELKLIKGF
--NVKFTEDYAFFN-----KS-----EKIADLRLIKSF
--NAKSSDYAFFN-----NS-----EKIADLKLKIKSF
--NAKFVEEYAFFK-----DS-----AKIADLRLIKSF
--NAKFAEEYSFFN-----NS-----EKIADLRLIKSF

```

图18JJ

```

Ga0187911_10069260      523  --SASYEADYKIFE-----DS-----GRVVEYLREINSF
Ga0187910_10015336      609  -KKVEFVENYSFFN-----SSNLLKEKTLNKSENYTKIVVEELREINSF
ODAI_chimera            595  -NRVEFGDKYDMFN-----KA-----GEICFSLRSINSF
31009_GL0034153        521  --KCEFMEYKFFN-----ES-----KDICLRLKEIHSF
DLF014_GL0011914_      591  --ECIFDKKFAFFD-----ES-----KNVCSTLREINSF
CasR_R._albus          577  --EHSFTDNYKMFA-----DS-----KAICLDLQFINSF
Ga0187910_10040531      564  --DCTFKPEFAFFG-----Y-D-----RCRNISGELRLINSF
tpg|DJXD01000002.1|    505  --EFKFKPEYADFF-----NMT-----NCRYTLEELRVINSI
CasR_E._siraeum        572  --DRITFAEYSFFE-----GS-----TKYLAELVELNSF
FIG-018_GL0023397      560  --ERTFTENYRFFE-----DS-----QKYRDELIELNSF
FIG-025_GL0099734      574  --ECNLLDEYQFFN-----TS-----DTVKKELHVIKNL
Ga0129317_1008067_chimera 506  --NIEFDDDFKFFD-----RS-----KDISVELNIIRNF
EYZ-362B_GL0088915     564  --YAPFSKDFNFFE-----DS-----EQRAEEMNLVKSI
uncultured_Ru_sp       569  -VECELTAGYKLFN-----DS-----QRITNELFIVKNI
R._flav._FD-1          568  --NTAFVKNYDFFN-----HS-----EKYVDELNIVKNI
tpg|DBYI01000091.1|    568  --DVKFTKDYDFFN-----II-----CDYAGELNIIKNI
:
:
:
ARMENKIE-----NAPKEPLLKDALALGVSSNDFDEKYEYFKTDVD--- 620
SKMDFGLD-----NINLSFMYDDALRTLGVSDENLPEVKREYF----- 566
GKMKPDMT-----DAKRVLYKSALEILGAPFD-----EVS----- 571
GKMDAGLD-----SLKPLFIRFAIQIILGGARIS-----DEDV----- 558
GKMKPDLT-----EAKRPLYKAAIRMLCPPEK-----WEKYTS----- 559
GKMKPDLE-----GAKRALFRDAVETLIGIVENSWLNWEDGTIS----- 544
037_-_emb|OIZA01000315.1|
emb|OCTW011587266.1|
k87_11092736
BMZ-11B_GL0037915
BMZ-11B_GL0037771
BMZ-11B_GL0069617

```

图18KK

Ga0099364_10024192	GKMKPDLG-----DAKRQLYKAAIEMLGIDTE-----EYIS-----	604
530373_GL0023589	AKMNKTRVA-----VKNDTKTKVLKQYCDAGAVLGRHINAEKIKAAAFGLGD-----	617
emb ODAI011611274.1	ARMAAADP-----EAKRFMVVDGAKLLGFNPKDTESEDEGIIRAIYGDAC	638
EMG_10003641	ARMESFEI-----KSDKTAINDAISLLGYNSNEHRDEFITYTMAKHVVVK	655
CasR_P1E0_metageno	ARMRKPPI-----GAKREMYRAAVVLLGVDSPEAAAAITDILLQI-----	582
emb OIZX01000427.1	ALMKKPV-----NAKQQLYRAAVVLLGFTEPDGVTDEMLCIDP-----	615
160582958_gene49834	ARMTKPLG-----GVNTKCVMYNEAAKILGAKPMSKAELESVFNLDNHD----	667
gi 1198542314 gb NFIR01000008.1	ARMTEPVP-----KTERIMFIEAAQIILGYSNGEKELEGYVDALLD-----	642
emb OGNF01009141.1	ARMEAPGA-----SSKRAMFVEAARILGDNRSKEELEXYFDTL-----	586
Ga0129306_1000735	ARMTTEPF-----NSKLVMFEDAAQDFGMSGGLVEHAEELEFYLDNK----	617
MH0288_GL0082219	ARMTTETV-----QKPCMFEDAAADLLGMDTDLLEEVREALEAFYNKE----	677
PIG-022_GL0026351	ARMTGFMEN-----DESTKFMFREAAADILGTSYTDDELTDHLNL-----	573
PIG-028_GL0185479	ARMTGFMEN-----DESTKFMFREAAADILGTSYTDDELTDHLNL-----	575
250twins_35838_GL0110300	AHLRPAIQ---EAQSSKDNADAAGTPATLLIDAYNMLNTDIQL-----	624
PIG-014_GL0226364	SKMKKLTK---KE-----DKSIKPVIDDALLELLGLKIEKYKKDDQGNFVLDQ----	576
250twins_36050_GL0158985	ASKDFKKS---KKD-----SDESYPEQLYLDALALLGDVVISKYKQNRDGSVVDDQ----	633
emb OJMM01002900.1	SKMKPELS---KI-----GEALILDAIDILGYKINKYKYDAAGNRLVDSN----	599
emb OGPN01002610.1	SKMKNKVN---ID-----NDIYYLDAIELLGRKIEKYQKDENGKILLGTD----	648
emb OHC01000044.1	SRMKKEMT---ID-----NDTIFLDALELLGRKIEKYQKDKNGDYVKDEK----	636
PIG-046_GL0077813	SKMKKEID---VD-----NETIFLDAFELLGRKIEKYKKDDSGNVLNRDD----	641
pig_chimera	SKMKKEIA---ID-----NDTIFLDALELLGRKIDKYKKDATGKYLKDN-----	643
emb OIEN01002196.1	AKMKKTL-----DFYNAQALKDAVTILGVSKKHQYLDMNSYLDYFYM-----	584
O2_UC29-0_GL0096317	SRMKKAL-----DFYNGQALRDAIRVLGIGSERHVDMDMKEYVDKYL-----	571
CasR_Anaerobic_dig	TRMDNEI-----FPSEGVFLDAANVLGVRGDDDFDSNEFVGDDDLHSD----	625

图18LL

emb OGDF01008514.1	611	ARMQHSIAGLGDVKIYGSDDKFKHGVSRRVYTDAAYIILFGFERSDNDGY--VDD-----
emb OGZC01000639.1	622	ARMGNPI-----SNLKEVMMVDAIQIILGTEKSEQELQSM--ACF-----
Ga0224415_10048792_chimera	563	AQMGEPA-----ATLKLEMTADAIKIILGTEKEDAEELIKL--AES-----
emb OCVV012889144.1	564	ARMGEPV-----ANTKRVMMDAVKILGTFDLSDDELVNM--ADS-----
Ga0187910_10006931	576	ARMGEPV-----ANAKRDMMDAIKILGTFDLDDNEIKKL--ADS-----
R._flav_XPD_	571	ARMGEPI-----ADARRAMYIDAIRILGTFNLSYDELKAL--ADT-----
Ga0224415_10007274	554	ARMGEPV-----ADARRAMYIDAIRILGTFDLSDDELKAL--ADS-----
Ga0187911_10069260	565	ARMTVDM-----EKIKRSAYKKALLIILGSSKYSDEDLDRVDEMLGV---
Ga0187910_10015336	652	ARMTGDC-----KIRKSAFEDASQLLGYHDKTVNNLFEVLRKLELESK-
ODAI_chimera	638	ARMESPA-----PAAKRTMFVDAIRMEGFADVSDNDAEKFYKENIEP-F-
31009_GL0034153	565	ARMKRPLD-----NKNVEKEMKYKDAISILMKDSSIENSNIDVIFNTYIY---
DLF014_GL0011914_	635	ARMQKPLN-----NVSVKREMYKDAIKILLKDKTRAEEQNI DRILNEYIP---
CasR_R._albus	620	ARMSKID-----DEKSKRQLFRDALVILDIGNKDETWINNYLDSDFK---
Ga0187910_10040531	606	ARMQKPS-----AKAKHVMYRDALRILGLDNGMSEEAALDQEVRRILQ---
tpg DJXD01000002.1	545	ARMQKPS-----ADARKIMYRDALRILGMDNRPDEEIDRELERTM-----
CasR_E._siraem	614	VKSCSFD-----INAKRTMYRDALDILGIESDKTEEDIEKMDNIIQ---
PIG-018_GL0023397	602	AKSCSFD-----ISAKRIMYRDALDILGIQAEELSEDEIGRMIDNIIQ---
PIG-025_GL0099734	617	TGMGYD-----VSIKHQMYKDAIDILGLHDNLSDEEINQLIDTNIL-F-
Ga0129317_1008067_chimera	544	ARMQAPV-----PNAKRAMQEDAIRILGSGSEEDIFSILDDMTG-----
EYZ-362B_GL0088915	602	ARMTKPD-----IQAKEQLFRDALDILGVEGVGVEEALQKEIH-----
uncultured_Ru_sp	607	ASMRKPA-----ASAKLTMFRDALTILGIDDKITDDRISGILK-----
R._flav._FD-1	610	ARMKKPS-----SNAKKAMYHDALTILGIPEDMDEKALDEELDLILE---
tpg DBYI01000091.1	610	ARMKKPS-----PSAKKNMYRDALTILGIPTEMSDEQLDAEIDKILE---

. :

图18MM

```

037_-_emb|OIZA01000315.1|          620
emb|OCTW011587266.1|              566
k87_11092736                       592
BMZ-11B_GL0037915                   576
BMZ-11B_GL0037771                   580
BMZ-11B_GL0069617                   563
Ga0099364_10024192                  654
530373_GL0023589                    617
emb|ODAI011611274.1|                649
EMG_10003641                         711
CasR_P1E0_metageno                   582
emb|OIZX01000427.1|                 615
160582958_gene49834                  667
gi|1198542314|gb|NFIR01000008.1|    642
emb|OGNF01009141.1                  586
Ga0129306_1000735                    617
MH0288_GL0082219                     677
FIG-022_GL0026351                    573
FIG-028_GL0185479                    575
250twins_35838_GL0110300             644
FIG-014_GL0226364                    598
250twins_36050_GL0158985             655
emb|OJMM01002900.1|                 621

```

图18NN

emb OGPN01002610.1	-----GKRLYTQEKYFNDFMFFNA----	667
emb OHCP01000044.1	-----GKKVYTKDYNNFQDMFFEG----	655
PIG-046_GL0077813	-----GKVLRTKEYNNIRKMLCKK----	660
pig_chimera	-----GKLYSKEYDDFQYMFVKD----	662
emb OIEN01002196.1	-----D-----	585
O2.UC29-0_GL0096317	-----D-----	572
CasR_Anaerobic_dig	-----AN-----	627
emb OGDF01008514.1	-----F-----	611
emb OGZC01000639.1	-----L-----	623
Ga0224415_10048792_chimera	-----F-----	564
emb OCVV012889144.1	-----F-----	565
Ga0187910_10006931	-----F-----	577
R._flav_XPD_	-----F-----	572
Ga0224415_10007274	-----D-----	555
Ga0187911_10069260	-----D-----	566
Ga0187910_10015336	-----DWKKRTDDEQQEYDRLLN-----	670
ODAI_chimera	-----YDRNGDIED-----EW-----	650
31009_GL0034153	-----	565
DLF014_GL0011914_	-----	635
CasR_R._albus	-----	620
Ga0187910_10040531	-----	606
tpg DJD01000002.1	-----	545
CasR_E._siraenum	-----I-----	615
PIG-018_GL0023397	-----L-----	603

图1800

FIG-025_GL0099734	-----I-----	618
Ga0129317_1008067_chimera	-----	544
EYZ-362B_GL0088915	-----	602
uncultured_Ru_sp	-----	607
R._flav._FD-1	-----K-----	611
tpg DBYI01000091.1	-----K-----	611
037_-_emb OIZA01000315.1	-----ADKHQKVSTFLMNNVINNSRFKY	644
emb OCTW011587266.1	-----GKTKNLSAYIRNNVLENRRFKY	588
k87_11092736	-----KTVNPFNRYIAKNVITSRFY	614
BMZ-11B_GL0037915	-----RSNTFRNFITNNVIKSRFVY	597
BMZ-11B_GL0037771	-----KQVNPFRNFIAGNVIESRRFMY	602
BMZ-11B_GL0069617	-----RAIKPFRNFIANNVIESRRFMY	585
Ga0099364_10024192	W-----RSLIDYEYLMPETNPFNRFVAKQVIESRRFMY	687
530373_GL0023589	-----KFNTTEEARVDHTFRNFIINNINSNRFNY	646
emb ODAI011611274.1	-----EAFYVQEGLYGKEREFSPYAFHTDLSLRFIAKNVVESARFRY	694
EMG_10003641	NLPAYFSEEDKEKLARYIVHSDGTYKKFLKESFYAIEELPNEGFRNFIINNINSRFRNY	771
CasR_F1E0_metageno	-----DPETGKVRPRSDSARDTGLRNFIAANNVVESRRFTY	617
emb OIZX01000427.1	-----VTGKMLPPNQRHHGDTGLRNFIAANNVVESRRFQY	649
160582958_gene49834	-----YTYSPSGKKIIPNKNFRNFIINNVIITSRFFLY	698
gi 1198542314 gb NFIR01000008.1	-----TKNKTNDKKKGFVRYIWNVVIKSTRFRY	671
emb OGNF01009141.1	-----FDKSASKKEKGFNFIINNVDVSNRFKY	614
Ga0129306_1000735	-----MLDKTKLRLLPDGKVDGTGFRNFIISNVTESTRFRY	652

图18PP

MH0288_GL0082219	-----K--MSLKWTKDQADTGFERNFIISNVLKSSRFY	710
PIG-022_GL0026351	-----KKTRLNARSKADTGLRNFIIISNVIKSSRFY	604
PIG-028_GL0185479	-----KKTRLIARSKADMGLRNFIIISNVIKSSRFY	606
250twins_35838_GL0110300	T-----KY--NKKGKVDYDHKFRNFLINNVIKSKWFFY	677
PIG-014_GL0226364	D-----SM--NDAEKSFVAKGHQKRNFIISNVVLNKRFLY	631
250twins_36050_GL0158985	D-----ESGGIKYKSGKPEYNHQRRNFILNVLKSKWFFY	691
emb OJMM01002900.1	D-----EFGRVKYNKKGKPVINHRRRNFIIINNVLSKWFFY	657
emb OGPN01002610.1	-----GNHKVRNFIANNVMQSKWFFY	688
emb OHCp01000044.1	-----KNHRVRNFVSNNVIKSKWFSY	676
PIG-046_GL0077813	-----DRRLNFIANSIIKNKWFSY	679
pig_chimera	-----SHRVRNFISNSVIKSKWFSY	682
emb OIEN01002196.1	-----NRSGATGKNAGKDHNLRNFLVSNVIRSRKFN	617
O2_UC29-0_GL0096317	-----DC----KDVKHADHNLRNFIVTNVIKSRKFN	600
CasR_Anaerobic_dig	-----KKIINKINGTKEDRNLRNFIIINNVVKSRRFQY	659
emb OGDF01008514.1	-----YVSSKLLGGADKNLRNFITNNVIKNRRFLY	641
emb OGZC01000639.1	-----FRDKNGKLTGTEHGMRNFIGNNVISNTRFQY	655
Ga0224415_10048792_chimera	-----FKDENGKLLGNKQHGMRNFIGNNVIKSKRFHY	596
emb OCVV012889144.1	-----FNDSEGNLLKKGKHGMRNFIIINNVIEKRFHY	597
Ga0187910_10006931	-----FKDSNGKLLHKGKHGMRNFIIINNVNKRFRHY	609
R._flav_XPD_	-----SLDENGNKLLHKGKHGMRNFIIINNVISNKRFRHY	604
Ga0224415_10007274	-----SLDENGNKLGKKGKHGMRNFIIINNVITNKRFRHY	587
Ga0187911_10069260	-----YNQNGEKIKVRVDTGFRNFIANNVVSSRFHY	598
Ga0187910_10015336	-----KHHYFKSGKLLPDTGLRNFIIINNVIESRRFN	702
ODAI_chimera	-----KKGKLVHKKDEKKGKGVNFIAINNVIESNRFQY	682

图18QQ

EMG_10003641	IMRFCNPEKIANIGKNKVLISFALSSLAEKTD-----MIAKYRVFCDRID-----	817
CasR_P1E0_metageno	LLRYMTPEQARVLAQNEKLI AFVLSVTPD--T-----QLERYCRTGREDI-----	661
emb OIZX01000427.1	LIRYSDPAQLHLASNKKLVRFVLSLIPD--T-----QINRYEYTCGQTRL-----	693
160582958_gene49834	LIRYGNPEKIRKTAINPSIISFVLKQIPD--E-----QIKRYYPPCIGKRT-----	742
gi 1198542314 gb NFIR01000008.1	LVRYPKVVAFVAFVAFVAFVFKDIPD--D-----QIRAYNSCFRQNSDSS	719
emb OGNF01009141.1	LTRYDTSVKAFAFNKALVKFAIKDIPQ--E-----QILRYNSCFGASER-----	659
Ga0129306_1000735	LVRYPKVVAFVAFVAFVAFVFKDIPD--T-----ILRYYEQSVGAAT-----	696
MH0288_GL0082219	VVRYGNPKLREVMQNEAVVKFVLDLPE--E-----QINRYSENLGLDA-----	754
PIG-022_GL0026351	LIRYSNPVKVKAIAENKLVRFVLDKIPD--P-----QIIRYCNCTCGF-KV-----	647
PIG-028_GL0185479	LIRYSNPVKVKAIAENKLVRFVLDKIPD--T-----QIIRYCNCTCGF-KV-----	649
250twins_35838_GL0110300	I AKYVKPADCAKMSNKKMIEFALDLPE--T-----QIKRYYTITGNEA-----	721
PIG-014_GL0226364	VAKYGSPTACNQMMKNQKIVRMALESISE--D-----QLRKYARRVYDNKEH----	676
250twins_36050_GL0158985	VVYKRNRPSSCRELMKNKEILRFVLRDIPD--S-----QVRRYFKAVQGEAA-----	735
emb OJMM01002900.1	VAKYRNRPSECQCFMKSKKLI ALVLKDVPE--T-----QIARYYQSVTGRT-----	701
emb OGPN01002610.1	VVRYNKPAEQCIIMRNKTLVKFTLDDLDPD--M-----QIQRYYSVFGDNN-----	732
emb OHC01000044.1	VVRYNKPAEQCALMRNSKLVKFALDELDPD--S-----QIEKYYISVFGESK-----	720
PIG-046_GL0077813	IVRYNQPSECQAIMKNKTLVKFALDELDPD--L-----QIQRYFVALYGED-----	723
pig_chimera	IVRYNQPSECRAIMKNKTLVKFALDELDPD--L-----QIQRYFVALYGED-----	726
emb OIEN01002196.1	LSRYSNLAEVKLAQNPSLVQFVLSRIEPP--S-----LICRYYESSQGIS-----	660
O2_UC29-0_GL0096317	LARYSDMATVSKLSKNKEVVKFVLRVDPD--K-----QIERYRSVCDVR-----	643
CasR_Anaerobic_dig	IARHMNTHYVKQLANNETLNRVFLNKMGD--A-----KIINRYYESISGNT-----	703
emb OGDF01008514.1	TVRYMNPKRRAKLVQNDALVVLALSGIPE--T-----QIDRYYKSCIEKR-----	684
emb OGZC01000639.1	LIRYGNPQKHLHLSQNETVVRVFLSRIAK--NQRVQMGNGKNQIDRYEYTCGGTN-----	708
Ga0224415_10048792_chimera	LIRYGDPAHLHLKATNKNVVRVFLGRIAD--MQKKQKQKGNQIDRYEYVCGNK-----	649

图18SS

k87_11092736	-----NSNEMVKLIEMIDRLITTEIAGFSFAVLKDKKDSIVSA-----SR	698
BMZ-11B_GL0037915	-----AGNGKPIKEKIKALAEKLGFEFSFDLIAENRDGIIDNA-----KL	680
BMZ-11B_GL0037771	-----SDTEAEIDFLVNKLSSEFSFETFISNQQTILNNS-----KRGFSP	684
BMZ-11B_GL0069617	-----SEISRKIDITLKALTGFSFNLSLFINRDKIVENT-----KL	664
Ga0099364_10024192	-----QHKNVSLKIDALADYLCKYTFEKNVLKQKNGIV-L-----NT	778
530373_GL0023589	-----ANPNYAHMKEQLTYKLLQVQFCNFVSNSSHQ-----	724
emb ODAI011611274.1	-----PDAAEVCVDFLAGVVNRVDFANFTDVRTGDSSK-----	771
EMG_10003641	-----DQKTMEDYLVNKLTKISYTEFLNVNQKANA-----	848
CasR_P1E0_metageno	-----TGRPAQIRYLTAQIMGVRYESFTDVEQRGRG-----	692
emb OIZX01000427.1	-----AGRAAKVEFLTDIAAIRFDQFRDYNQKERG-----	724
160582958_gene49834	-----DDVTLMRDELGKMLQSVNFEQFSRVNKKQNAK-----	774
gi 1198542314 gb NFIR01000008.1	NNSNASWDADSNKRDISVSDMRKALTEKITGLNFGDFEESKKGIR-----	765
emb OGNF01009141.1	-----YNDGMSDKLVEAIGKINLMQFNQVGIQQADR-----MLPE	695
Ga0129306_1000735	-----VDREIILDTLADKLLSLRFTDFENVNQRANA-----	728
MH0288_GL0082219	-----AAKEKKVSELARFLRALSLDFRDVEQSANA-----	786
PIG-022_GL0026351	-----GERQENINTLAKAITNFSFSAFEKVQSANVE-----N-----	680
PIG-028_GL0185479	-----GKRQENINTLAEAITYFSFSNFEKVQSANVE-----N-----	682
250twins_35838_GL0110300	-----LGD AESLKGVIIIEQLHAFSFKNTLLSIKNNM--G-----EGEY	756
PIG-014_GL0226364	-----FQNQTADIRENLAKKLNFNIGAEIKTVGNL-----SEKD	712
250twins_36050_GL0158985	-----YASAEAMRTRLVDALSQFSVTACLDEVGGM--T-----DKEF	770
emb OJMM01002900.1	-----QANSEAMRMTLIKLLHEFSIKNVLSVGTM--T-----ASEN	736
emb OGPN01002610.1	-----MPAVDEMRRLLDKINQFSVRGFLDELDELIVLM-----SDEE	769
emb OHCP01000044.1	-----SSSNEEMRRELLKKLDFSVRGFLDELIVLL--S-----EDEM	755
FIG-046_GL0077813	-----LPSYGMKRKILLKLLHDFSIKDFDEEIGSL--F-----D-EM	757

图18UU

```

pig_chimera 761
emb|OIEN01002196.1|-----LPSYGE MRKILLKLLHDFSIKFLDEIVLL--S-----DLDM
O2.UC29-0_GL0096317 699
CasR_Anaerobic_dig 683
emb|OGDF01008514.1|-----SEGITIDEQIKKLTGIIIVDMNIDSFENINNGEIGMR-----Y---SK
CasR 737
emb|OGZC01000639.1 723
Ga0224415_10048792_chimera 743
emb|OCVV012889144.1|-----NISAEDEIEKLVDIYFNFNIDELEQVNAEIGTG-----HSGRKV
Ga0187910_10006931 694
R._flav_XPD_ 691
Ga0224415_10007274 674
Ga0187911_10069260 685
Ga0187910_10015336 779
ODAI_chimera 762
31009_GL0034153 667
DLF014_GL0011914_ 738
CasR_R._albus 727
Ga0187910_10040531 721
tpg|DJXD01000002.1|-----PNIEVRSQIDYLVKRLRSFSFEDLNDVKQVRPG-----
CasR_E._siraenum 726
FIG-018_GL0023397 716
FIG-025_GL0099734 731
Ga0129317_1008067_chimera 652
EYZ-362B_GL0088915 706
emb|OGZC01000639.1 743
Ga0224415_10048792_chimera 694
emb|OCVV012889144.1|-----DIKKTIEEKIDALTDIIIVNMNYDQFEKKAIVIENQ---NRGKTFEKN
Ga0187910_10006931 696
R._flav_XPD_ 691
Ga0224415_10007274 674
Ga0187911_10069260 685
Ga0187910_10015336 779
ODAI_chimera 762
31009_GL0034153 667
DLF014_GL0011914_ 738
CasR_R._albus 727
Ga0187910_10040531 721
tpg|DJXD01000002.1|-----IPPTSDKSAQIDVLARIISVVYKIFEDVPSAKINK-----DDPS---
CasR_E._siraenum 726
FIG-018_GL0023397 716
FIG-025_GL0099734 731
Ga0129317_1008067_chimera 652
EYZ-362B_GL0088915 706

```

图18VV

uncultured_Ru_sp	-----DMNSSLGVKRSELRMIKNI SFDDFNKVKQAKG-----	711
R_flav._FD-1	-----EMNPPTEKKITELAGLKLDMNFGNFRNRVRSQAK-----	718
tpg DBYI01000091.1	-----NLNPTTEKKIEKLAEMIKNIKIFEEFRNVKQVRD-----N-----	720
	:	
037_-_emb OIZA01000315.1	---IPNVEKERKKAITLYFTIVYILVKNLVNINGLYTLALYFVERDGFYKQICEKKLI	772
emb OCTW011587266.1	KSPQRSADRERKKAMLTLYTIVYIVFVKQMLHNSLYTIGFFYIERDQRFIYSRAKKEN-	724
k87_11092736	ESRAVNLEVERLKKLTLYMSIAYIAVKSIVKNARYFIAYSALERDLYFFNEKYGEEFR	758
BMZ-11B_GL0037915	DEEKNVEIERLKAVTGLYLTVAFAIAIKNLVKANARYIAYSAFERDSELLKELGEEEF	740
BMZ-11B_GL0037771	NRRETAEEIERLKAITGLYLSVAYIAIKNIVKANARYI AFAVFERDKELVRAKDAR---	741
BMZ-11B_GL0069617	DTSNKNVKIERLKALTGLYLTVAFAVAVKNLVKANARYI AFSAFERDYGFMKSHDPD--V	722
Ga0099364_10024192	KSATKNVEIEHLKALTGLYLTWAYIAVKNLVKANARYI AFSIFERDYALFEKKLGKDTL	838
530373_GL0023589	-----AVEKERLKAIVGLYLSVLYRIVKSI VRINTSYTIAFAAYERDCVILNRKMGNLY	779
emb ODAI011611274.1	---SEQEKKQYQAI VGLYLTVVYWI VKNLVNNSRYVMAFHILERDVTLLLEGKRLFVGG	828
EMG_10003641	-----KNKEKDRSQKLI GLYITLLYEIVKNLVNINSRYNIAFQRCDNDSIMILQGQYDER-	903
CasR_P1E0_metageno	-----DNPKKERYKALIGLYLTVLYLAVKNMVCNARYVIAFYCRDRDTALYQKEVCWYDL	748
emb OIZX01000427.1	-----ANTQKERYKAMGLYQYTVLYLAVKNLVNINARYVMAFHCVERDMFLYDGETLDPKG	780
160582958_gene49834	--QNPNGEKARLQACVRLYLTVPYLF IKNMVNINARYVLA FHCLERDHALCFNSRKLN--	830
gi 1198542314 gb NFIR01000008.1	-----KEESDKNIIRLYLTVLYLVQKNLIVNSRYFLAFHCAERDEMLYNGETIDNNK	818
emb OGNF01009141.1	EKKKANAQKEKYSIIRLYLTVCYLFFKNLVVNSRYSAFYNLEKDRSLFEINGELKPT	755
Ga0129306_1000735	-----RNREKQKMGII SLYLNVAQIVKNLVVYNARYTMAYHCAERDTELLLNAAEGENL	784
MH0288_GL0082219	-----RNARKARYQGAMSLYLNVIYQI IKMVMYINSRYTLAFAYEHDA SLLLDNWNVDYPI	842
FIG-022_GL0026351	-DTPDAIDKAQKGLISLYLTVLYLFIKNMVYVNSRYVLA FHCLERDSQLLKITSYGN--	737
FIG-028_GL0185479	-DTPDAIDKAQKGLISLYLTVLYLFIKNMVYVNSRYVLA FHCLERDSQLLKITSYGN--	739

图18WW

Ga0187910_10040531	ANDRMTLKKQRYQAIVSLYLTVMYLVTKNLVVNSRYVMAFHALERDAYLYGITNI-----	777
tpg DJXD01000002.1	RNFSDALKKQRYQAIVSLYLTVMYLITKNLVVNSRYVIAFHCLERDAFLHGVTLPKM--	718
CasR_E_siraeum	-RTSEAEIKRKNQAIIRLYLTVMYIMLKNLVNVARVYIAFHCVERTDKLYAESGLEVG-	784
FIG-018_GL0023397	-KSRNAEIKRKNQAIIRLYLTVMYLMLKNLVNNSRYVIGFHCVERDAKLYKLSGLTIE-	774
FIG-025_GL0099734	-SSKEAQEKMYQTIISLYLTVCHLVKNLVNINARVYAMAFHSLEARDARLYGYGSEKDY	790
Ga0129317_1008067_chimera	-TPEEKVEKERYKAIIGLYLTVMYHLVKNLVVNSRYVMAFHCCLERDAMHYDVS-----	705
EYZ-362B_GL0088915	---TKQMEKERFKAIIRLYLTAIYLVVKNLVNVARVYIAFHCLERDADLFLGLIKQTLP	763
uncultured_Ru_sp	---RENVAKERAKAVIGLYLTVMYLLVKNLVNVARVYIAIHCLERDFGLYKEIIPEL-A	767
R._flav._FD-1	----ENMEKERFKAVIGLYLTVVYRVVKNLVVDVNSRYIMAFHSLERDSQLYNVS-----	768
tpg DBYI01000091.1	--SQEAVEKERFKAIIGLYLTVIYLLVKNLVNNSRYVMAFHCCLERDAKLYGVQ-----	772
* : * : * : * : * : *		
037_-_emb OIZA01000315.1	-----ETLKKDKKQLYNDDDDYLLLPFIKSGSKYRE-----	803
emb OCTW011587266.1	-----KNPSKNSYLNDFRSVTAFFIPSEIMKRIEKNE-----	756
k87_11092736	-----LHFIP-----YELNGKTCQFEYLAILKYLLARDEETLKRKCEICEEIKVG	803
BMZ-11B_GL0037915	-----YKLCIEFKN---KKGETGYNNAFALLKYYIDRDEDIWRKKEATHRGRDEN	787
BMZ-11B_GL0037771	-----IQTTIPNTKFTNYFCLTQYLLDRDEEKKFFGDPDRDKEAFFE	782
BMZ-11B_GL0069617	-----AQSCLAYEEIMDGSKTKYNEIYALTEYLLKKEEENDYHPEPGQPFDKEA	772
Ga0099364_10024192	-----EKYVYKPFKYIDKGEKEGKNNFFALTEYLLDKDNLRYQWNNDLSDDEENK	888
530373_GL0023589	GL-----KTPMGITNDFMHMPTVAPSNPH-----	804
emb ODAI011611274.1	MK-----AEDPFLITDGVVSRQDAYVRK-----	851
EMG_10003641	-----AVQESKLTKKFISNQK-----	919
CasR_P1E0_metageno	EE-----DKKSGQRQVEDY TALTRYFVSQGY-----	775
emb OIZX01000427.1	ES-----VSAFLAVNGKKGVPQYLLLTQLFIRRDY-----	811

图18YY

160582958_gene49834	-----DDSYNEMANKFQVVRKAKKEQYEKEYKCK-----	859
gi 1198542314 gb NFIR01000008.1	EK-----GSEKDWKFAKQFIMEHS-----	838
emb OGNF01009141.1	GK-----FDEGHYTGVLKFLIDNGW-----	775
Ga0129306_1000735	LRRDRSWPARLHLPRRALARRDRVEMERDVARGPEAYNRDEWLG-----	830
MH0288_GL0082219	-----KTVFSDLTEDSLQK GK-----	858
FIG-022_GL0026351	-----NSSYNNLTNYFIDNKY-----	753
FIG-028_GL0185479	-----NLSYNNLTNHFIEKEY-----	755
250twins_35838_GL0110300	-----EKKFDAIINEILLEDDKIRKECDKERAQAQAKTL-----	842
FIG-014_GL0226364	-----KKEKGRFDWLALTKFVLSQKDYFEHDLRKEINKI-----	804
250twins_36050_GL0158985	-----KKSSDYTGEDMLALTRKFGVEDAGLYREWEKNAEAKDK-----	864
emb OJMM01002900.1	-----LANNEDDKWKKNYVFALTKHFWDNDEPYFDKYNNALQQIRSI-----	838
emb OGPN01002610.1	-----EMKGGPWDGGAQALAVTRKFLNHDREVFDRYCARAEIARL-----	864
emb OHCP01000044.1	-----EKAERAWEKGATAFALTRKFLNHDKPTFEQYTREREISAM-----	850
FIG-046_GL0077813	-----NKQERAWSSGATALALTRKFLNQDKLIFEKHYAREGEISKL-----	852
pig_chimera	-----NKQERAWSSGATALALTRRFLNQDKLIFEKHYAREGEISKL-----	856
emb OIEN01002196.1	-----KNKKRSSGGFSFIEMTQFFIDKKLKFKVAT-----	787
O2_UC29-0_GL0096317	FKDDP-----DAGKSKKQIAYRTKYIALTEEYLDKGYIKCAT-----	781
CasR_Anaerobic_dig	-----DFESYTALTNHFIEKEY-----	811
emb OGDF01008514.1	-----YMRDVFIAKGI-----	795
emb OGZC01000639.1	-----LKKRYTALTEMILGYETDEKAR-----	824
Ga0224415_10048792_chimera	-----KLDGFVALTKFCLGDEERFEDL-----KAK-----	777
emb OCVV012889144.1	-----IDKKISSLTKFCLGESDEKLLK-----AL-----	779
Ga0187910_10006931	-----NNLESGFTSVTKLCVGIADDDPVKYKVEIELKER-----	790
R._flav_XPD_	-----KKLEEKGFSSVTKLCAGIDEITAPDKRKRKVEKEMAER-----	785

图18ZZ

Ga0224415_10007274	-----KCLKDKGFTSVTKLCAGIDECK-----DVEKEMTER	764
Ga0187911_10069260	-----AEAEYLNLTRHFIEPLDGAKP-----	764
Ga0187910_10015336	-----KRGYCDLSTVLLFGVDDLQNRNRGSYKYLRDN	865
ODAI_chimera	NTGDY-----PADN--ADLRRRDIKVCVLTKAHIKQEEVKSRRND-----	855
31009_GL0034153	PEQPE-----KLYKIKPKPLQPAQYFLFTKYLLQNDFFGE-----	761
DLF014_GL0011914_	-----NKKLEKIKPPKYHELTKYFIDNRYFDRK-----	821
CasR_R_albus	-----YQGCRLTDHFIEEKYKME-----	803
Ga0187910_10040531	-----KGDYRKLTDNLLADENYKKE-----	797
tpg DJXD01000002.1	-----NKKIVYSQLTTHLLTDKNYTTY-----	740
CasR_E_siraeum	-----NIEKNKTNLTMVMGVKLENG-----IIKT-----	809
PIG-018_GL0023397	-----GNTDKNRTLLTATILGITSadRAPSGEIVA-----	804
PIG-025_GL0099734	V-----EKLKKEQAVLVEKLLSDGPETAG-----	814
Ga0129317_1008067_chimera	-----LDNYRDLIRHLISEGDSSEN-----	725
EYZ-362B_GL0088915	D-----RKKKSDYLKLAVDQCGDNPENKP-----	787
uncultured_Ru_sp	S-----KNLKNDYRIILSQTLCELCDKSPN-----	791
R_flav_FD-1	-----VDNDYLALTDTLVKEGDNRSR-----	789
tpg DBYI01000091.1	-----NIGGDYLALTAKLCAEGDDYGGKLLSEAKQ-----	801
037_-_emb OIZA01000315.1	-----	803
emb OCTW011587266.1	-----	756
k87_11092736	CEKHKKNANPPYEQEWIDKKKALNSE--RKACERRLHF-----STH	844
BMZ-11B_GL0037915	GKLDYDPDF-----	796
BMZ-11B_GL0037771	-----HL-----	784

图18AAA

BMZ-11B_GL0069617	CRKHLD	778
Ga0099364_10024192	QALRKHLDKKE	899
530373_GL0023589		804
emb ODAI011611274.1		851
EMG_10003641		919
CasR_P1E0_metageno		775
emb OIZX01000427.1		811
160582958_gene49834		859
gi 1198542314 gb NFIR01000008.1		838
emb OGNF01009141.1		775
Ga0129306_1000735		830
MH0288_GL0082219		858
PIG-022_GL0026351		753
PIG-028_GL0185479		755
250twins_35838_GL0110300	P	849
PIG-014_GL0226364	SD	812
250twins_36050_GL0158985	YF	872
emb OJMM01002900.1	V	845
emb OGPN01002610.1	P	870
emb OHCPO1000044.1	P	856
PIG-046_GL0077813	P	858
pig_chimera	P	862
emb OIEN01002196.1		789
O2.UC29-0_GL0096317		783

图18BBB

CasR_Anaerobic_dig	811
emb OGDF01008514.1	795
emb OGZC01000639.1	829
Ga0224415_10048792_chimera	811
emb OCVV012889144.1	813
Ga0187910_10006931	824
R._flav_XPD_	819
Ga0224415_10007274	798
Ga0187911_10069260	764
Ga0187910_10015336	925
ODAI_chimera	855
31009_GL0034153	761
DLF014_GL0011914_	821
CasR_R._albus	803
Ga0187910_10040531	797
tpg DJXD01000002.1	740
CasR_E._siraenum	811
PIG-018_GL0023397	806
PIG-025_GL0099734	814
Ga0129317_1008067_chimera	725
EYZ-362B_GL0088915	787
uncultured_Ru_sp	791
R._flav._FD-1	789
tpg DBYI01000091.1	803

图18CCC

037_-_emb OIZA01000315.1	-----ET--KNLKLPEKDRDIMKKYLPNDKDRDREYNKFFKQYRN	841
emb OC1W011587266.1	-----NKGLEDFEALWNSC--GKTSRLRKEDVLLYARYTSPDHALKNYKMIILNSYRN	807
k87_11092736	WAQYATKRDENMAKHPQKWYDILASHYDELLAL-----QATGWLATQARN	889
BMZ-11B_GL0037915	-----RRKHYPKRRWLNILKDNLAEMERI-----DPTGMLAAQMRN	832
BMZ-11B_GL0037771	-----RKTKRHFSKQREWLNKEIADAKSS-----QATGILLREARN	821
BMZ-11B_GL0069617	-----SIRRHFTKKWRDIFRKEIDEAVKI-----HFTGYLGVWARN	814
Ga0099364_10024192	-----IRSQRHFSQYWLDFARQIENAKKT-----SESGYLLTAARN	936
530373_GL0023589	-----NDFINETKLSKRVCALVKNRRTL-----CTNRTYQVYRN	838
emb ODAI011611274.1	-----RIGENKRANRHGLNCVLENRNALGS-D-----PASTDAAASLIWSYRN	893
EMG_10003641	-----LNSYSCRYLTHNISQLD-----RCNDFIRQYRN	947
CasR_P1E0_metageno	-----LNRHACGYLRSNMNG-----INSNLLTAYRN	801
emb OIZX01000427.1	-----LKRSAEQIQHNMEN-----ISDRLLREYRN	837
160582958_gene49834	-KQETGTAHTKKIEKLNQQIAYIDKDIK-----MHSYTCRNYRN	898
gi 1198542314 gb NFIR01000008.1	-----PKKVKDYLAKNFEY-----SNKWSLKEFRN	864
emb OGNF01009141.1	-----INPRASAYLTVNLAN-----SDETAIRTFRN	801
Ga0129306_1000735	-----LVRTLRREKRVCDNLHNNYAYLCGAD-----AEPGDASLLEFVYRN	872
MH0288_GL0082219	-----LNRRAAKYLQTNLEH-----SDNTLIRQYRN	884
FIG-022_GL0026351	-----LNPHASRYLKVNISN-----ADEWSIVKFRN	779
FIG-028_GL0185479	-----LNSHARNYLRVNISN-----SDEWSIVKFRN	781
250twins_35838_GL0110300	RFAQIKRRESGCYFKSYHVYDLSKNSNEFKQ-----NHIDFAVTSYRN	893
FIG-014_GL0226364	QHRENDKYYDSCMHFSRHSFEYISKNAFVQSH-----PDPLKLFIKYRN	857
250twins_36050_GL0158985	KVILRQNDKMITRKMHTPHSLNLYVQKNLESVQS-----NGLAAVIKEYRN	916

图18DDD

emb OJMM01002900.1	LAYRANDKVVKHTHFNLHSYKYVKHNYEEISK-----ASKIITAYRN	887
emb OGPN01002610.1	PLRKANDKLLKQTHYTNHSYTYIVNNLNSFTDIDYCAKDVGLPAPNDKNDNASILGEMRN	930
emb OHECP01000044.1	ELRKENDQLLKKTHYSKHAYCYIVDNVNNLTGAVANDNGRGLPCLSEKNDNANLFLEMRN	916
FIG-046_GL0077813	AMRKVNDQLLKQTHFSKHSYCYIVDNVNNLTGGECDTKRVLVPLNEKNDNAGILSDFRK	918
pig_chimera	AMRKVNDQLLKRTHFSKHSYCYIVDNVNNLTGGECDTKRVLVPLNEKNDNAGILLDFRK	922
emb OIEN01002196.1	IKKNVLKYNGNPESLNHIPEYICKNMEGYHEN-----TVRNFRN	829
O2_UC29-0_GL0096317	IKENIKKYGDPKSIDHIPGEYIKKNLEGYCAG-----STNTFRN	823
CasR_Anaerobic_dig	-----LPVRKAEILRANLKLHLDCEDE-----GFKYYRN	838
emb OGDF01008514.1	-----NPKKPTVKSIKEQYAFPTY-----IFTTYRN	822
emb OGZC01000639.1	TVYEKAEAAKNRHLKNVKWCKTRENLENA-----DKNAIVAFRN	869
Ga0224415_10048792_chimera	LNRErvKNARNAYLKNIKNYIMIRLQLRDQTDSE-----SGYLCGEFRD	854
emb OCVV012889144.1	INRErAKTAMNAHLRNTKWDIMYGQLKDLADSE-----KSRICSEFRN	856
Ga0187910_10006931	INREKAKTALNAHLRNTKWNVIIRENIRNT-----EKDACKQFRN	864
R._flav_XPD_	INREKAKTALNAYLRNTKWNVIIREDLLERI-----DNKTCTLFRN	859
Ga0224415_10007274	INREKAKTALNAHLRNTKWNVIIREDLLERR-----DNKACKIFRN	838
Ga0187911_10069260	-----YLLKKRPVEYLLKMDMAMVGNSE-----SIRHFRN	791
Ga0187910_10015336	LDESASSYLSHKLVNKVVHSYKYLKQNLDCADNT-----MINQFRN	965
ODAI_chimera	-----TDSKSQQKAKSRRRWCLLLKEDIDKLNCGE-----ESLRILYRN	894
31009_GL0034153	-----YTSKKNQYIQKNMKIYSKIE-----NKNFNVFVRYRN	792
DLF014_GL0011914_	-----KKNKKYGEYTSKKKISTYIETNMKNFIGCE-----QFRTPERYRDTQIDMFVKYRN	871
CasR_R._albus	-----GKLGCKKVKGRYLKNNISCCCTDGE-----LINTYRN	833
Ga0187910_10040531	-----GH-----FNKKWRGIAEQNLRNSDVPE-----VIKSFERN	826
tpg DJXD01000002.1	-----GHLKNQGHKRWYVLVKNNLQNSDIT-----AVSSFRN	773
CasR_E._siraenum	DKSF-AENAAANRYLRNARWYKLLILDNLKKSERA-----VVNEFRN	850

图18EEE

FIG-018_GL0023397 845 DLGR-AETAANHLRKKRWYVQIFNNLKQADRN-----VINEFRN
 FIG-025_GL0099734 844 -----NLHLRSKKNRLMRVNLNYSAA-----ASQFRN
 Ga0129317_1008067_chimera 761 -----HF1SHNRRMRDCIEENVKNSEQLI-----FGKEDAVIRFN
 EYZ-362B_GL0088915 820 -----KGYLARNKRLCQVREDIDNAKHL-----NGDITDYN
 uncultured_Ru_sp 821 -----LFLKKNERLRKCVEVDINNADSS-----MTRKYRN
 R._flav._FD-1 826 -----RYLAGNKRLRDCVKQDIDNAKKWF-----VSDKYNSITKYN
 tpg|DBYI01000091.1| 850 NQDKVQMPKNYFLARNKRWREAIEQDIDNAKKWF-----IGEKFNNVKNYRN
 * .
 037_-_emb|OIZA01000315.1| 886 NIVHLKIIA-KLSEL---TKNIDK-----DINSYFDIYHYCTQRVMFDYCKNN
 emb|OCTW011587266.1| 851 KIAHINVIM-SAGKY---TGGIK-----RMDSYFSVFQHLVQCDILSNPNKG
 k87_11092736 945 DAEHLNPVN-EFDVY---IEDLRRYPEGTPKNKDYHIGSYFEIYHYIRRAYLEEVLAKR
 BMZ-11B_GL0037915 881 DAMHLNAMF-KIAKF---VGDFRK---G---SSHEMRSYFELFHLEQKTLDSKEDFV
 BMZ-11B_GL0037771 872 DVEHLNVL-R-AIPDY---IQDFRH---GEGK---ETAMNSYFELYHYLMQRMLKNTLDEL
 BMZ-11B_GL0069617 868 QSAHLNVLTAALPQF---IGQFAR---NSKTGKTARMTSYFQLYHYLLQRLLCEDLNFDV
 Ga0099364_10024192 984 CALHLNVL-T-ALPEF---VGEFRK---T-----GDKMTSYFELYHFLQKLMMLAEAGLNL
 530373_GL0023589 881 IIMHLNAVS-AFPQY---LEGLK-----VVHSYFDLYHYTVMMLLHRQNEKE-
 emb|ODAI011611274.1| 936 AAAHLTAVA-AAQEY---VSELR-----EIHSYFEVYHYAMQRYLKSGAEFA-
 EMG_10003641 990 KVAHLEVVS-NIDEY---LSGIK-----HIESYALYHYLMQRCLLKNYRIE-
 CasR_P1E0_metageno 844 AVDHLNAIP-PLGSL---CRDIG-----RVDSYFALYHYAVQVYLNGRYRK-
 emb|OIZX01000427.1| 881 AVAHLNVIA-HLADY---SADMR-----EITSYGLYHYLMQRHLFKRHAQWI
 160582958_gene49834 944 LVAHLNVVS-KLQNVVSELPNDY-----QITSYFSFYHYCMQLGLMEKVSCK-
 gi|1198542314|gb|NFIR01000008.1| 911 SVQHLNVIR-DAHKYIKYINDNK-----DVGSYFALYHYLVQRYISERAANRT
 emb|OGNF01009141.1 844 TAEHLEALR-NADKY---LNDLK-----QFDSYFEIYHYITQRNIKEKCEML-

图18FFF

Ga0129306_1000735	KAAHLSVLN-KGRL---SGDLK-----EAKSWFYVYHFLMQRVLEEFERNF-	915
MH0288_GL0082219	SVAHLNAVR-NMNLV---IGDLR-----EITSYFGIYHYIMQRDLITRCQKV-	927
PIG-022_GL0026351	GTAHLNSVR-NIDTC---IEGLK-----DFKSYFEIYAYTVQKSIKKYFDAEF	823
PIG-028_GL0185479	GVAHLNSVR-NIDTS---IKGLR-----EFKSYFEIYAYTVQKSIKNYYDAES	825
250twins_35838_GL0110300	NVEHLNVVH-CMTKY---FSEVK-----DVKSYGVYCYIMQRMCLDELI IKN	937
PIG-014_GL0226364	NVMHLNIIR-DLNDY---LPDFKR---G---KAFEKTDTPYDLYCYCLQKMLLSEVD---	904
250twins_36050_GL0158985	AVAHLNIIN-RLDEY---IGSAR-----ADSYSLYCYCLQMYLSKNFVSG-	958
emb OJMM01002900.1	NVQHLNVMN-SITKY---LGDIS-----EVTSYSLYCYTLQRLLDDNNND-	930
emb OGPN01002610.1	DIAHLNIVH-DMVKY---IEELK-----DISSYAFYCYVLQRRLLVGKDPNC-	973
emb OHCPO1000044.1	KIVHLNVVH-DMVKY---INEIK-----NITSYAFFCYVLQRMIIIGNNSNE-	959
PIG-046_GL0077813	TIAHLNVVH-KMVDY---VAEIK-----GITSYAFFCYVLQRMVPGNKN-E-	960
pig_chimera	TIAHLNVVH-KMVDY---VDEIK-----GITSYAFFCYVLQRMVGNLNE-	965
emb OIEN01002196.1	MVAHLTAVA-RVPLY---ISEVT-----QIDSYALYHYCMQMNILQGIEQSG	873
O2.UC29-0_GL0096317	AVAHLALMA-RVSEY---ASDIK-----NVNSYFSLYHYCMQHILYCARRFN	867
CasR_Anaerobic_dig	QVTHLNAIR-VAYKY---INEIK-----SVHSYFALYHYIMQRHLYDSLQAKA	882
emb OGDF01008514.1	MIAHLAAVT-NAYKY---IPQMD-----RFKSWFHLYHTVIQHSLIQQYFYDR	866
emb OGZC01000639.1	IVAHLWIIR-DADRF---ITGMG-----AMKRYFDCYHYLLQRELGYILEKSN	913
Ga0224415_10048792_chimera	KVAHLEVAR-HAHEY---IGNIK-----EVNSYFQLYHYIMQCRLYDVLKNNT	898
emb OCVV012889144.1	KAAHLEVAR-YAHMY---INDIS-----EVKSYFRLYHYIMQRRIIDVIENNP	900
Ga0187910_10006931	KADHLEVAR-YAYKY---INDIS-----EVNSYFQLYHYIMQRIIIDSSGN--	906
R._flav_XPD_	KAVHLEVAR-YVHAY---INDIA-----EVNSYFQLYHYIMQRIIMNERYE--	901
Ga0224415_10007274	KVAHLEAIR-YAHLV---INDIA-----EVNSYFQFYHYIMQRRIMAERYD--	880
Ga0187911_10069260	ATVHLNVIM-EAHRV---TKDIK-----YIGSYALYHYIILQRHLLDKIEEDS	835
Ga0187910_10015336	NVAHLNTIIR-NMDG-----IENVT-----GITSYFQIYHYLMQKALYKEFKKCR	1008

图18GGG

ODAI_chimera	AVAHLNIPR-NAIDY---IGGIK-----KADSYFAIYHYMMQQLSKKHGLP-	937
31009_GL0034153	NIAHLENSIK-NATKY---VNDIG-----TFTSYFQLYHYIMQCCLKDWVKKEV	836
DLF014_GL0011914_	SIAHLNAVR-KASKY---INDIR-----GFDTYFELYHYIMQRDLKILEKIA	915
CasR_R._albus	QVDHFAVR-KIGNY---AAYIK-----SIGSWFELYHYVIQRIVDFEYRFAL	877
Ga0187910_10040531	MAAHISVIR-NIDLY---IGDIQ-----KVDSYFALYHFLMQKLIQRVVPENT	870
tpg DJXD01000002.1	IVAHISVVR-NSNEY---ISGIG-----ELHSYFELYHYLVQSMIAKNWYDT	817
CasR_E._siraenum	TVCHLNAIR-NININ---IKEIK-----EVENYFALYHYLIQKHLENRFADKK	894
FIG-018_GL0023397	SVCHLNAIR-NINTN---IEGIG-----FVKNYFALYHYLIQRHISSKRIS---	887
FIG-025_GL0099734	AVAHLNPIR-NADVL---IEGIE-----NITSYFDIYHYIVQRSVLRNRTNKT-	887
Ga0129317_1008067_chimera	NVAHLSAIR-NANEY---IGDIR-----EITSYFALYHYLMQRKLIIDCKVN-	804
EYZ-362B_GL0088915	CVAHLNVVR-NCDQY---LENID-----TVHSYFSLYQYLVQVHYLLKRFEKDK	864
uncultured_Ru_sp	CIAHLTVVR-ELKEY---IGDIC-----TVDSYFSIYHYVMQRCITKRENDTK	865
R._flav._FD-1	NVAHLTAVR-NCAEF---IGDIT-----KIDSYFALYHYLIQRQLAKGLDHER	870
tpg DBYI01000091.1	NVAHLTAIR-NCAEF---IGEIT-----KIDSYFALYHYLIQRQLAGRLDPNH	894
	* : :	
037_-_emb OIZA01000315.1	NVVLÄ-----KMKDLAHIKSDCDEFSSKHTYPIYSSAVLRFM	922
emb OCTW011587266.1	KCFES-----ESL--KPLLL--DMKFDGTDEKLYSKRLTRAL	884
k87_11092736	KEYRD-----SGSFTDEQLDK-LQKILDDIRARGSYDKNLLKLE	983
BMZ-11B_GL0037915	QGMRE-----S-----YGDRIRVEPCVDLIKYL	904
BMZ-11B_GL0037771	S-HWS-----G-----WIMRSGRPDRDLIQIA	893
BMZ-11B_GL0069617	PEEWK-----N-----KIQETGLPCNDLIHIC	890
Ga0099364_10024192	D-EYR-----E-----RIDTYQIACKDLINIT	1005
530373_GL0023589	-----KKRNEKEKKQNEKENNMILSTA-ALEKYDMVRKYQSANKDFIYAI	925

图18HHH

emb ODAI011611274.1	-----ELVSKNGPASKG-IAAWANAVDRCHSFCKDWLWLL	970
EMG_10003641	-----DHSQNEYKN-LNDFSSKLDKHGTYVYKDFVKAL	1021
CasR_P1E0_metageno	-----TPR-EQELFAAMAQHRTWCSDDLVKAL	869
emb OIZX01000427.1	RQPERP-----TEEEQKLI EQEQQLAW-EKALFDKTLQYHSHYKNDLVKAL	926
160582958_gene49834	-----NIPLVESLKNEANDAQSYSAKK-TLEYFDLIEKNRTYCKDFLKAL	988
gi 1198542314 gb NFIR01000008.1	-----DKESLSEGR-LQYVLSQVKEYRITYCKDFVKAL	942
emb OGNF01009141.1	-----KEQ-TVKYNNDDLKYHGYSKDFVKAL	869
Ga0129306_1000735	-----QALPER-LRELLMMAERYGCSKDLIKVL	943
MH0288_GL0082219	-----KMKNPV-VVGYFDKVRTYRTCCCKDLVKAL	955
PIG-022_GL0026351	SEDRNIEDGKISLCNGVYIHERQKNGKPKITVKINPK-TIYMESVDKYGSYTKDFVKAL	882
PIG-028_GL0185479	P-----VQINPK-TIYMESVDKYGSYTKDFVKAL	854
250twins_35838_GL0110300	QDKPD-----VRQT-FEENRLLKDHGTYSKNLMWLL	968
PIG-014_GL0226364	-----G-FSAINENIDKYCTASKDFMNL	927
250twins_36050_GL0158985	-----Y-LINVQKLEEHHTYMKDLMWLL	981
emb OJMM01002900.1	-----K-FASIKGNLRKFGIYNKDFMWLL	953
emb OGPN01002610.1	-----QNKF-KAKYAKELNDYGTYNKNLMWML	999
emb OHCPO1000044.1	-----QNKF-KAKYSKTLQEFGTYSKDLMMVL	985
PIG-046_GL0077813	-----KNAL-KEYSETAKSYGTYSKDFMWLI	986
pig_chimera	-----KNAI-KEYSATVKSFGTYSKDFMWLI	991
emb OIEN01002196.1	KIL-----DNIK-LKNALENARVHRTYSKDAVKYL	902
O2_UC29-0_GL0096317	KQ-----GNEK-MAAALNDAEIHRYSKDAVKFL	895
CasR_Anaerobic_dig	KDS-----SGF-VIDALKSFEHKIYSKDLLHVL	910
emb OGDF01008514.1	DYGRK-----GAPVVSER-VLQLEQCREHSNYSRDLHLHIL	901
emb OGZC01000639.1	-----Q-TKKSLEKVVQYHSHYCKDFLHML	940

图18III

Ga0224415_10048792_chimera	KAEAM-----VKGK-AKEYFEALEKEGTYNDKLLKIA	929
emb OCVV012889144.1	K--AK-----YEGK-VKVFEDVKNKKNKLLKLM	929
Ga0187910_10006931	---N-----ANGM-IKKYVESVSDKKYNDRLKLL	933
R._flav_XPD_	---K-----SSGK-VSEYFVAVNDEKKYNDRLKLL	928
Ga0224415_10007274	---K-----SSGK-VREYFVAVNNEKKYNDRLKLL	907
Ga0187911_10069260	YAEKT-----VS-EKLWESQISQYGTYSKDFVKAL	864
Ga0187910_10015336	ENAVRKWIPIITENAEPKYVYWNKKEQVEVSFNPKIFGYMENIKNHSNTYCKDFVKAL	1068
ODAI_chimera	-----EGLGCTHGVYCKDLVKAL	955
31009_GL0034153	EEKKI-----SDKY-YNILEKYYKNNNIYLNKDFVKAI	867
DLF014_GL0011914_	EGAAKET-----EEGQKQDEL-LKTYFDNLDKYGYVVKDFVKAL	953
CasR_R._albus	N-----NT-ESNYKNSIIKHHTYCKDMVKAL	902
Ga0187910_10040531	KGLJ-----SDQ-TKKYYDALEQYNTYCKDFVKAY	898
tpg DJXD01000002.1	S-H-----QPK-TAEYLNLLKHHHTYCKDFVKAY	844
CasR_E._siraenum	VE-----RD-TGDFISKLEEHKTYCKDFVKAY	920
FIG-018_GL0023397	FN-----SF-SAEYISNLEKYHTPSEYFIKAY	913
FIG-025_GL0099734	-----PK-MKEYEELINKYHTYNDKDFVKAL	911
Ga0129317_1008067_chimera	-----DT-AHKYFEQLTKYKTYVMDMVKAL	828
EYZ-362B_GL0088915	YRCK-----DDL-AIHYSQVKQYRSYVKDLVKAL	893
uncultured_Ru_sp	QEE-----KIKYEDDLLKNHGYTKDFVKAL	890
R._flav._FD-1	SGF-----DR-NYPQYAPLFWHTYVVKDVVKAL	897
tpg DBYI01000091.1	PGF-----EK-NYPQYAPLFWNTYVVKDMVKAL	921
037_-_emb OIZA01000315.1	NLPFAYNVPFRKNI.SYKKFFDKQ-----	945

图18JJ

emb OCTW011587266.1	NIPFGYNVPRYKNLTFEKIYLKSSINE-----	911
k87_11092736	YLPFAYNLPYKKNLTTEALFDDDDSVSGKKRVAEWREKXTREAEEREQRQR-----	1034
BMZ-11B_GL0037915	YITLGYSLPRYKTMVTVSGLFDPDSEDAKERAKGKRA-----	940
BMZ-11B_GL0037771	FVSLAYNLPYRNLTKEHHFDDTVLQKIREKESLD-----	928
BMZ-11B_GL0069617	FVPLGYNLPYKKNLTIECLFDEDSVSAKEGKKA-----	923
Ga0099364_10024192	YVSLGYNLPYKKNLTCEPLFDEESATGKERQTRLDEKSKEKKQKGGQK-----	1054
530373_GL0023589	NSPLAYNAARYINLSCRTKFEEGFGK-----	951
emb ODAI011611274.1	NVPPFAYNPARYKNLSIANLFDKNEAAPVTEADASEQKED-----	1009
EMG_10003641	NVPPFGYNLPYKKNLSIDELFDRNKLKTGGTIEMKGE-----	1057
CasR_P1E0_metageno	NTPPFGYNLARYKNLSIDGLFDREGDHVVVREDGEEKPAE-----	906
emb OIZX01000427.1	NAPEFGYNLARYKNLSIEPLFSKEAAPAAEIKATHA-----	961
160582958_gene49834	NAPFSYNLPRFKNLSIEALFDKNIIVYEQADLKKE-----	1022
gi 1198542314 gb NFIR01000008.1	NVPPFAYNLPYKKNLSIDELFDRNNYLPNKAKKWIPEKKENGEYVMEDCGNKDAGQVENA	1001
emb OGNF01009141.1	CVPPFGYNLPRFKNLSIDALFDKNDKREKLLKGGFED-----	904
Ga0129306_1000735	NLTFAYNLPYKKNLSIDGRFDKNHPDPSDE-----	973
MH0288_GL0082219	NITFGYNLPYKKNLSIEGLFDKNRPGDKKESALIAE-----	991
PIG-022_GL0026351	NTPPFGYCLARYKNLSIEPLFDRNTVYNKEDVTEEFVSNSKEQ-----	924
PIG-028_GL0185479	NTPPFGYCLARYKNLSIEPLFDRNTVYNKKDVKEEFVNSK-----	894
250twins_35838_GL0110300	NFPFAYNLPARYKNLSNEIDLFAKNNDQSK-----	998
PIG-014_GL0226364	NIIFAYNVPRYKKNLSNKELFYRLDSEQ-----	954
250twins_36050_GL0158985	NIIPFAYNLPARYKNLSNEKLFYDEEAAAEEKADKAENERGE-----	1020
emb OJMM01002900.1	NIIPFAYNLPYKKNLSNEEIFYDELQK-----	979
emb OGPN01002610.1	NLPFAYNLPYKKNLSSEFLFYDMEYNKKDDE-----	1030
emb OHCPO1000044.1	NLPFAYNLPYKKNLSNEQLFYDEEERMEKIVGRKNDSR-----	1023

图18KKK

FIG-046_GL0077813	NLPFAYNLPRYKKNLSNEQLFYDEEERNETEGQIDRL-----	1022
pig_chimera	NLPFAYNLPRYKKNLSNEQLFYDEEERNETEEQIDRL-----	1027
emb OIEN01002196.1	CLPFAYNISRYKALTIKDLFDWTEYSCKKDE-----	933
O2_UC29-0_GL0096317	CLPFAYNLPRYKALTVEGLFDWNEYGQDDD-----	925
CasR_Anaerobic_dig	HSPFGYNTARYKNLSIEALFDKNE SRPEVNLSTND-----	946
emb OGDF01008514.1	NLPFGYNLPRYLNLSSEKYFDANAI-----	926
emb OGZC01000639.1	CLPFAYCIPRYKKNLSIAELFDRHEPEAEKPEEASVNNQSFIIT-----	984
Ga0224415_10048792_chimera	CVVPGYCIIPRYKKNLSMEELFDMNEEKFKKKKAPENT-----	965
emb OCVV012889144.1	CVVPGYCIIPREFKNLSIEQFDMNETDNSDKKKEK-----	963
Ga0187910_10006931	CVVPGYCIIPREFKNLSIEALFDKNEAAKYDKIKKKVAVR-----	971
R._flav_XPD_	CVVPGYCIIPREFKNLSIEALFDRNEAAKFDKKEKVVSGNS-----	967
Ga0224415_10007274	CVVPGYCIIPREFKNLSIEALFDMNEAVKFDKEKK-----	940
Ga0187911_10069260	CCPFGYNLPRFKKNLSIEQLFDRNESKEITDAPRQ-----	900
Ga0187910_10015336	CAPFAYNLPRFKKNLSIEELFDMHELSSEPKESMKLTD-----	1105
ODAI_chimera	NAPFGYNLPRFKKNLSVEALFDMNEEPAPEGKERSKNADGCN-----	996
31009_GL0034153	NVPFAYNYSRFKNLSIDKLFDKNDPRKKKWIVFNF-----	903
DLF014_GL0011914_	NIPFAYNYPRYKKNLSIDELFDKNNTRKTIKIQDKKLEERT-----	993
CasR_R._albus	NTPFGYDLPRYKKNLSIGDLFDRNNYLNKTKESIDANSSIDSQ-----	944
Ga0187910_10040531	CTPFAYVTPRYKKNLTIDGLFDRNRPGEDK-----	927
tpg DJXD01000002.1	CIPFGYVVPRYKKNLTINELFDRNNPNPEPKPEEV-----	877
CasR_E._siraenum	CTPFPGYNLVRKKNLTIDGLFDKNYPGKDDSDSEQ-----	954
PIG-018_GL0023397	CTPFAYNLVRKKNLTIDGLFDRNAPSEK-----	941
PIG-025_GL0099734	CAPFGYNNVRFKALTIYEMFDRNYPPEKRAEEARKSC-----	948
Ga0129317_1008067_chimera	CSVPGYNLPRFKKNLSIEGKFDHESK-----	854

图18LLL

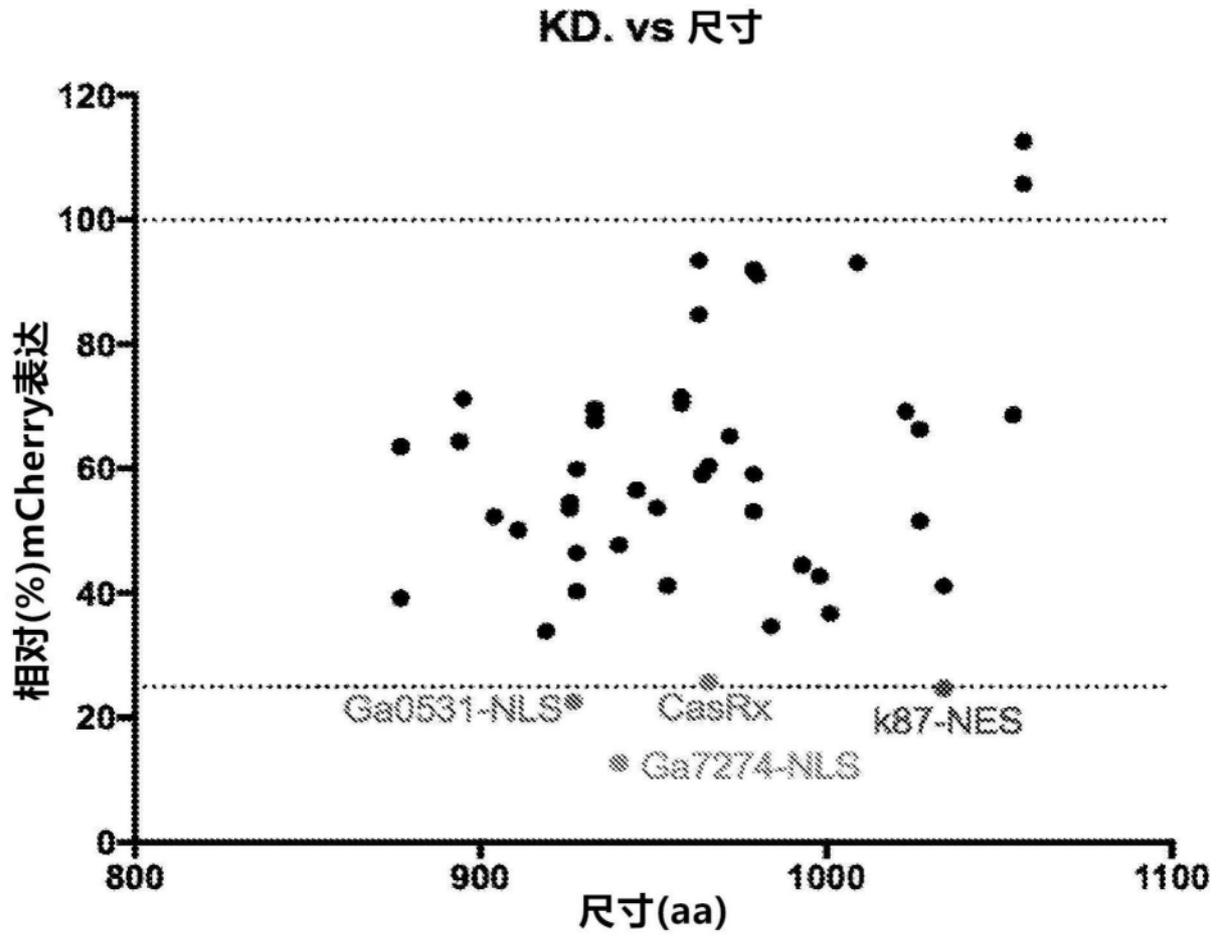


图19A

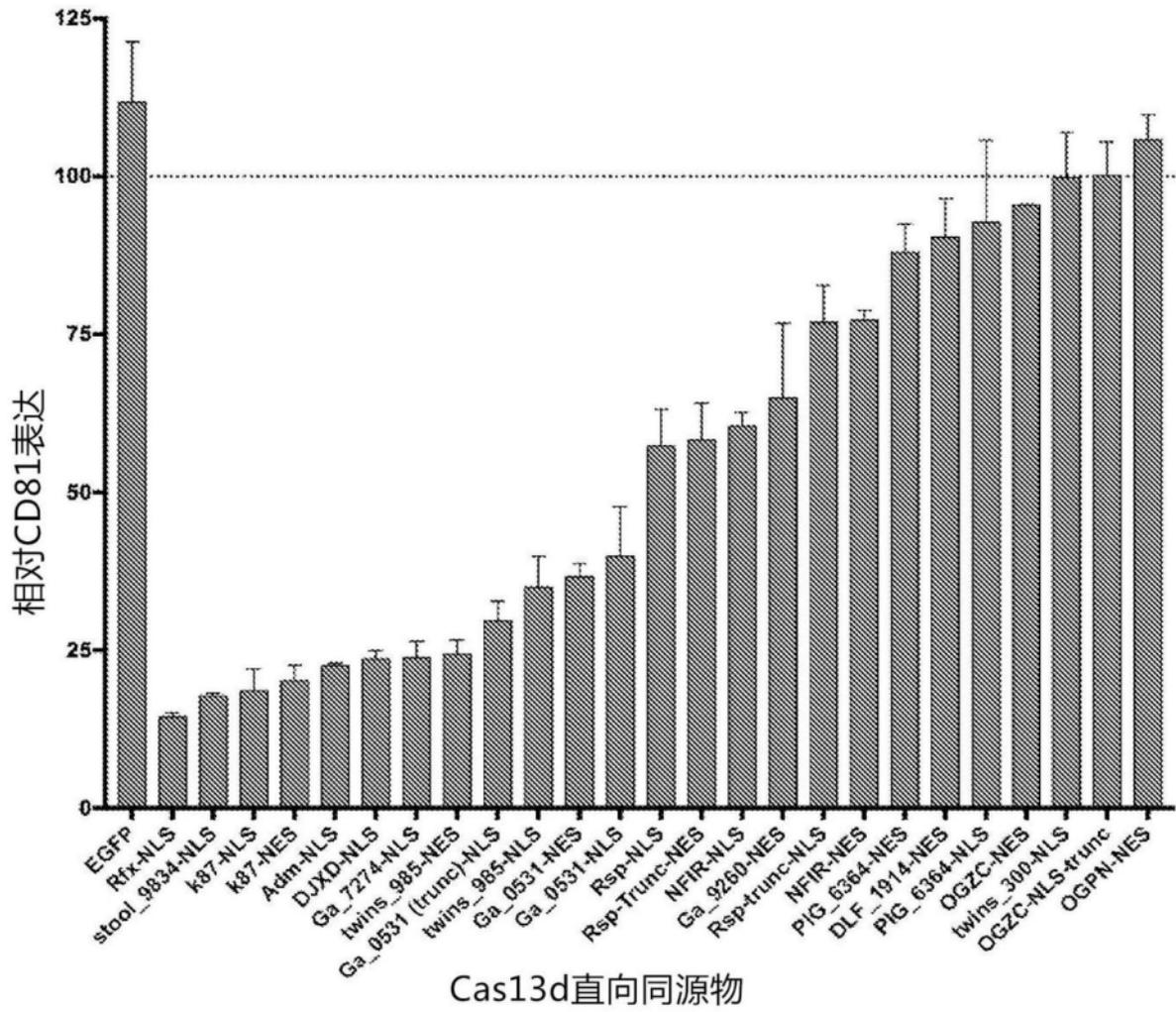


图19B

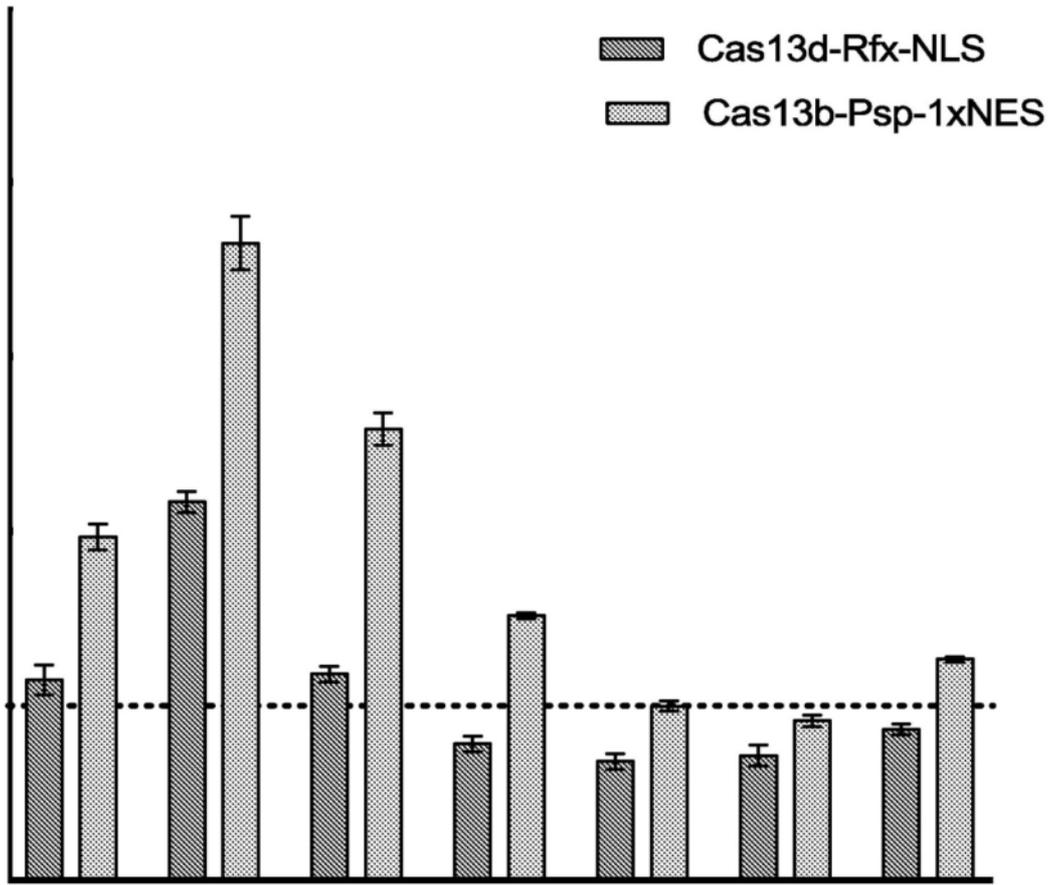


图20

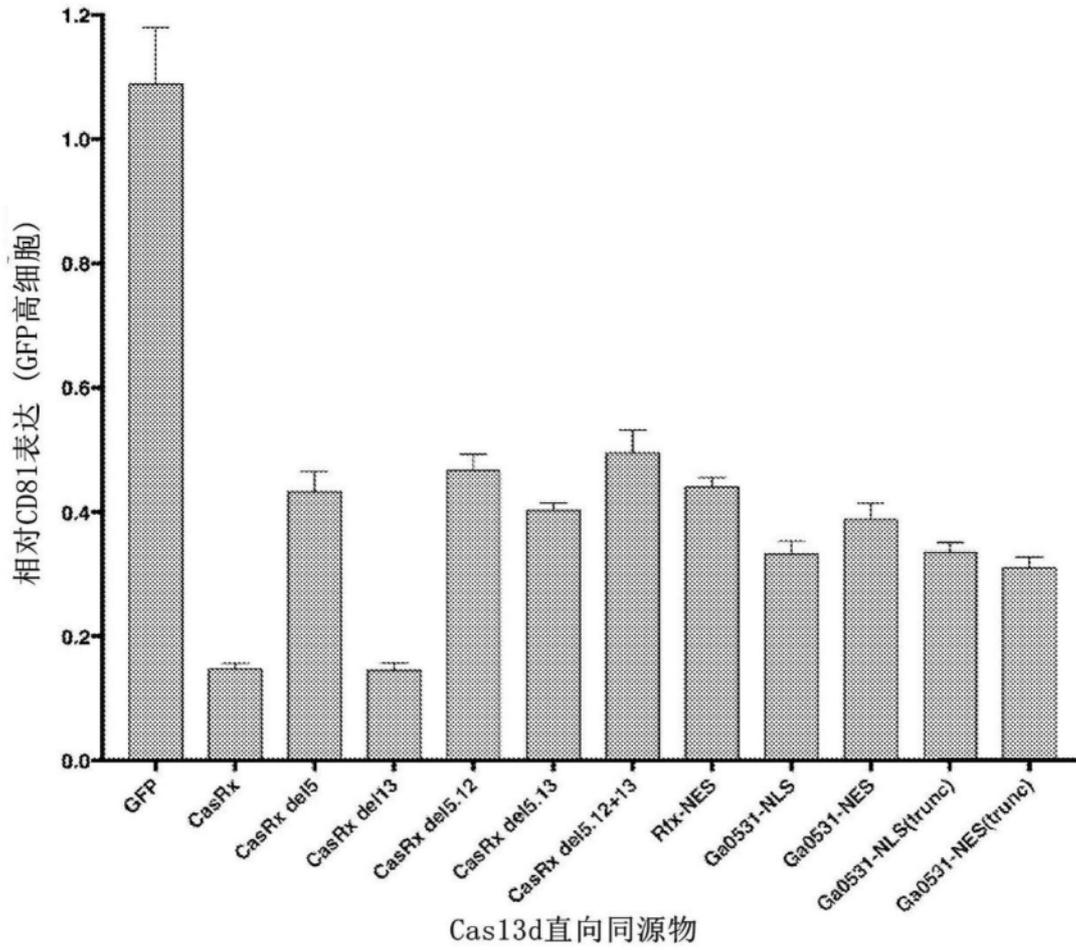


图21

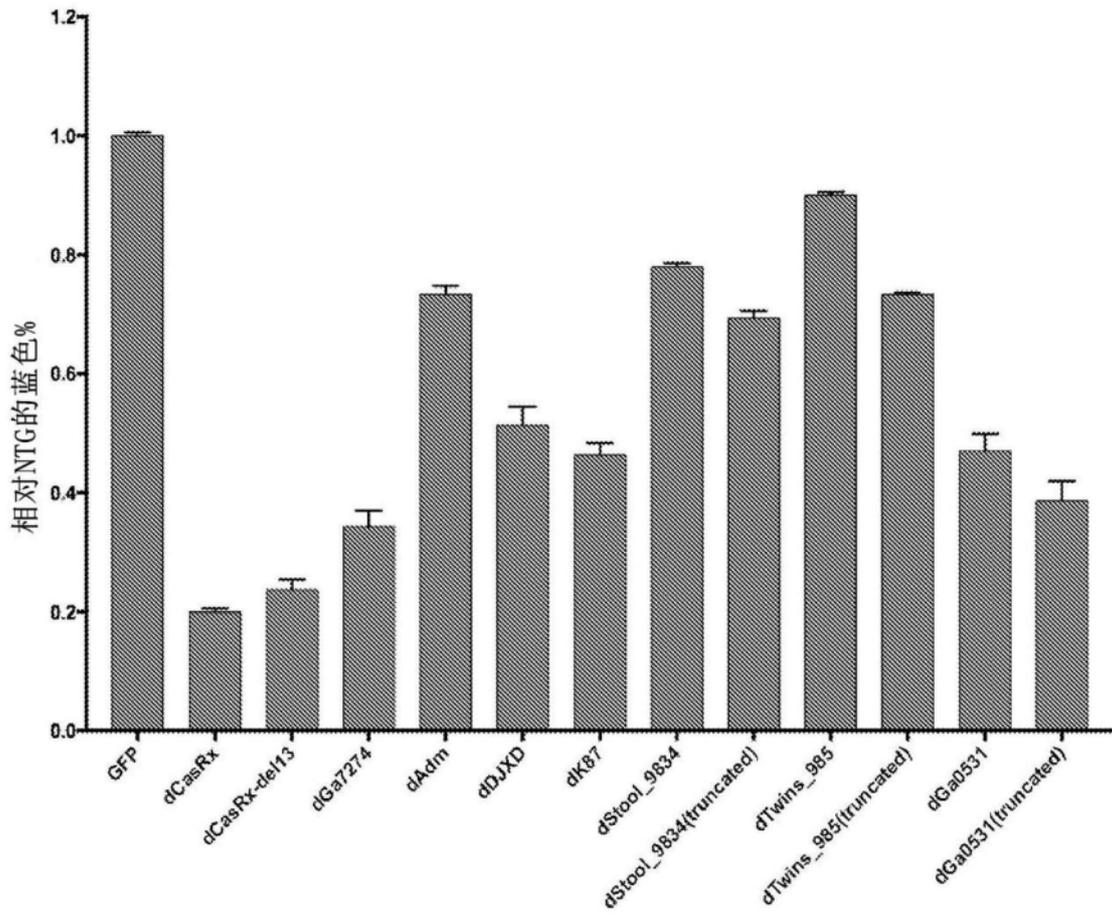


图22

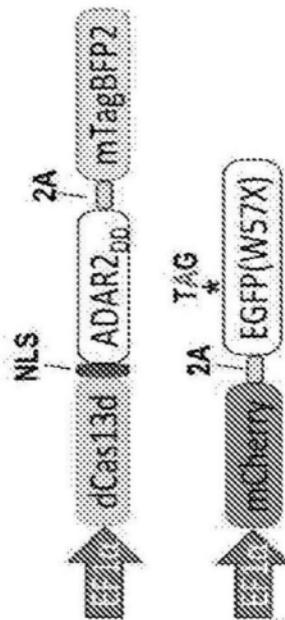


图23A

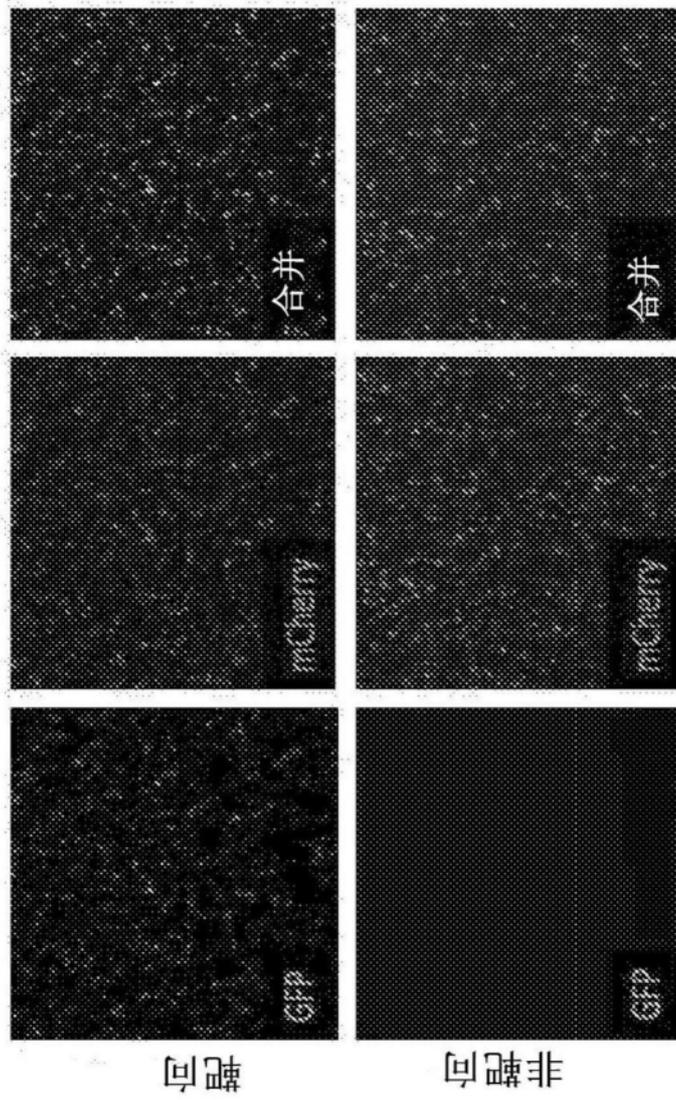


图23B

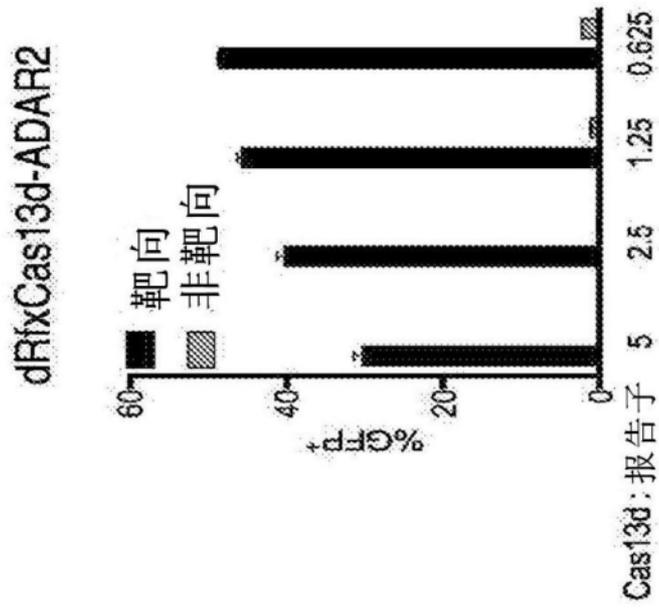


图23C