(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 118530967 A
(43)申请公布日 2024.08.23

(21)申请号 202410333226.7

(22)申请日 2018.08.22

(30)优先权数据
62/549,303 2017.08.23 US
62/641,687 2018.03.12 US

(62)分案原申请数据
201880068905.9 2018.08.22

(71)申请人 通用医疗公司
地址 美国马萨诸塞州

(72)发明人 J·K·乔昂格 B·克莱因斯蒂弗

(74)专利代理机构 北京林达刘知识产权代理事务所(普通合伙) 11277
专利代理师 李茂家 闫俊萍

(51)Int.Cl.
*C12N 9/22*(2006.01)
*C07K 19/00*(2006.01)
*C12N 9/80*(2006.01)
*C12N 9/10*(2006.01)
*C12N 9/78*(2006.01)
*C12N 9/02*(2006.01)
*C12N 15/85*(2006.01)
*C12N 15/55*(2006.01)
*C12N 15/62*(2006.01)
*C12N 5/10*(2006.01)
*C12P 19/34*(2006.01)

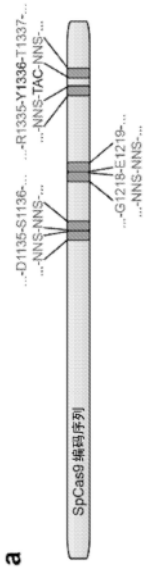
权利要求书3页 说明书32页
序列表（电子公布） 附图13页

(54)发明名称

具有改变的PAM特异性的工程化的CRISPR-Cas9核酸酶

(57)摘要

具有改变的PAM特异性的工程化的CRISPR-Cas9核酸酶。具有改变的和改进的PAM特异性的工程化的CRISPR-Cas9核酸酶及其在基因组工程、表观基因组工程、和基因组靶向中的应用。




CN 118530967 A

1.一种分离的化脓性链球菌Cas9(SpCas9)蛋白,其在以下三个、四个、五个或全部六个位置具有突变:D1135、S1136、G1218、E1219、R1335和/或T1337,其中所述突变不是VSREER或VSREQR。

2.根据权利要求1所述的分离的蛋白,其包含与SEQ ID NO:1的氨基酸序列至少80%相同的序列。

3.根据权利要求1所述的分离的蛋白,其包含表A、1、2或3中所示的突变的组。

4.根据权利要求1-3任一项所述的分离的蛋白,其包含以下突变的组:LRSVQL、LRKIQK、LRSVQK、LWKIQK、VRKIQK、LWKIQK、IRAVQL、VRKLRS、GRKIQK、SWRVVV、SWKVLK、TAHFKV、MSGVKC、LRSVRS、SKTLRP、MWVHLN、TWSMRG、KRRCKV、VRAVQL、VSSVRS、VRSVRS、SRMHCK、GWKLLR、GWKOQK、VAKLLR、VAKIQK、VAKILR、GRKILR、VRKLLR、IRAVQL、VRKIQK或VRMHCK变体(例如,对于NGTN PAM);MQKSER、VRKSER、ICKSER、LRSVER、LWLETR、LSRWER、MQSVQL、VRREER、ICCCER、LSRWQR、LWRVVA、WMQAYG、LWRSEY、SQSWRS、LKAWRS、LWGWQH、MCSFER、LWMREQ、LWRVVA、HSSWVR、MWSEPT、GSNYQS、FMQWVN、YCSWVG、MCAWCG、FMQWVR或SSKWPA变体(例如,对于NGCN PAM);LRSVRS、SRQMRG、MRARKE、SRMHCK、VRREQR、VRGEQR、LRLSAR、AWTEVTR、KWMMCG、VRGAKE、AWNFQV、LWTTLN、SRMHCK、CWCQCV、AEEQQR、GWEKVR、NRAVNG、LRSYLH、VRGNNR、VQDAQR、GWRQSK、AWLCLS、KWARVV、MWAARP、SRMHCK、VKMAKG、QRKTRE、LCRQQR、CWSHQR、SRTHTQ、LWEVIR、VSSVRS、VRSVRS、IRAVRS、SRSVRS、LWKIQK或VRMHCK变体(例如,对于NGAN PAM)。

5.根据权利要求1所述的分离的蛋白,其进一步包含一种或多种降低核酸酶活性的突变,所述突变选自由在D10、E762、D839、H983、或D986处;和在H840或N863处的突变组成的组。

6.根据权利要求5所述的分离的蛋白,其中所述突变是:

(i)D10A或D10N,和

(ii)H840A、H840N或H840Y。

7.根据权利要求1所述的分离的蛋白,其进一步包含一种或多种增加特异性的突变,所述突变选自由在N497、R661、N692、M694、Q695、H698、K810、K848、Q926、K1003和R0160处,和任选地在K526和/或R691处的突变组成的组。

8.根据权利要求7所述的分离的蛋白,其包含突变N692A、Q695A、Q926A、H698A、N497A、R661A、M694A、K810A、K848A、K1003A、R0160A、Y450A/Q695A、L169A/Q695A、Q695A/Q926A、Q695A/D1135E、Q926A/D1135E、Y450A/D1135E、L169A/Y450A/Q695A、L169A/Q695A/Q926A、Y450A/Q695A/Q926A、R661A/Q695A/Q926A、N497A/Q695A/Q926A、Y450A/Q695A/D1135E、Y450A/Q926A/D1135E、Q695A/Q926A/D1135E、L169A/Y450A/Q695A/Q926A、L169A/R661A/Q695A/Q926A、Y450A/R661A/Q695A/Q926A、N497A/Q695A/Q926A/D1135E、R661A/Q695A/Q926A/D1135E和Y450A/Q695A/Q926A/D1135E;N692A/M694A/Q695A/H698A、N692A/M694A/Q695A/H698A/Q926A;N692A/M694A/Q695A/Q926A;N692A/M694A/H698A/Q926A;N692A/Q695A/H698A/Q926A;M694A/Q695A/H698A/Q926A;N692A/Q695A/H698A;N692A/M694A/Q695A;N692A/H698A/Q926A;N692A/M694A/Q926A;N692A/M694A/H698A;M694A/Q695A/H698A;M694A/Q695A/Q926A;Q695A/H698A/Q926A;G582A/V583A/E584A/D585A/N588A/Q926A;G582A/V583A/E584A/D585A/N588A;T657A/G658A/W659A/R661A/Q926A;T657A/G658A/W659A/R661A;F491A/M495A/T496A/N497A/Q926A;F491A/M495A/T496A/N497A;

K918A/V922A/R925A/Q926A;或918A/V922A/R925A;K855A;K810A/K1003A/R1060A;或K848A/K1003A/R1060A,和任选地K526A和/或R691A。

9.一种融合蛋白,其包含经任选的间插接头与异源功能结构域融合的根据权利要求1-8任一项所述的分离的蛋白,其中所述接头不干扰所述融合蛋白的活性。

10.根据权利要求9所述的融合蛋白,其中所述异源功能结构域是转录激活结构域。

11.根据权利要求9所述的融合蛋白,其中所述转录激活结构域来自VP16、VP64、rTA、NF-κB p65或复合物VPR(VP64-p65-rTA)。

12.根据权利要求9所述的融合蛋白,其中所述异源功能结构域是转录沉默子或转录阻遏结构域。

13.根据权利要求12所述的融合蛋白,其中所述转录阻遏结构域是Krueppel相关盒(KRAB)结构域、ERF阻遏物结构域(ERD)或mSin3A相互作用结构域(SID)。

14.根据权利要求12所述的融合蛋白,其中所述转录沉默子是异染色质蛋白1(HP1)。

15.根据权利要求12所述的融合蛋白,其中所述异源功能结构域是修饰DNA的甲基化状态的酶。

16.根据权利要求15所述的融合蛋白,其中所述修饰DNA的甲基化状态的酶是DNA甲基转移酶(DNMT)或TET蛋白。

17.根据权利要求16所述的融合蛋白,其中所述TET蛋白是TET1。

18.根据权利要求9所述的融合蛋白,其中所述异源功能结构域是修饰组蛋白亚基的酶。

19.根据权利要求18所述的融合蛋白,其中所述修饰组蛋白亚基的酶是组蛋白乙酰转移酶(HAT)、组蛋白去乙酰化酶(HDAC)、组蛋白甲基转移酶(HMT)或组蛋白去甲基化酶。

20.根据权利要求9所述的融合蛋白,其中所述异源功能结构域是碱基编辑器。

21.根据权利要求20所述的融合蛋白,其中所述碱基编辑器是

(i)胞苷脱氨酶结构域,优选选自由以下组成的组:脱氨酶的载脂蛋白BmRNA编辑酶,催化多肽样(APOBEC)家族,任选地APOBEC1、APOBEC2、APOBEC3A、APOBEC3B、APOBEC3C、APOBEC3D/E、APOBEC3F、APOBEC3G、APOBEC3H或APOBEC4;活化诱导的胞苷脱氨酶(AID),任选地活化诱导的胞苷脱氨酶(AICDA);胞嘧啶脱氨酶1(CDA1)或CDA2;或作用于tRNA的胞嘧啶脱氨酶(CDAT),或

(ii)腺苷脱氨酶,优选选自由以下组成的组:腺苷脱氨酶1(ADA1)、ADA2;作用于RNA的腺苷脱氨酶1(ADAR1)、ADAR2、ADAR3;作用于tRNA的腺苷脱氨酶1(ADAT1)、ADAT2、ADAT3;以及天然存在或工程化的tRNA特异性腺苷脱氨酶(TadA)。

22.根据权利要求9所述的融合蛋白,其中所述异源功能结构域是生物系链。

23.根据权利要求22所述的融合蛋白,其中所述生物系链是MS2、Csy4或λN蛋白。

24.根据权利要求9所述的融合蛋白,其中所述异源功能结构域是FokI。

25.一种分离的核酸,其编码根据权利要求1-24任一项所述的蛋白。

26.一种载体,其包含根据权利要求25所述的分离的核酸。

27.根据权利要求24所述的载体,其中根据权利要求25所述的分离的核酸可操作地连接至一个或多个调节结构域,以表达分离的化脓性链球菌Cas9(SpCas9)蛋白,所述蛋白在一个、二个、三个、四个、五个或全部六个以下位置具有突变:D1135、S1136、G1218、E1219、

R1335和/或T1337。

28.一种宿主细胞,优选哺乳动物宿主细胞,其包含根据权利要求25所述的核酸,并任选地表达根据权利要求1-24任一项所述的蛋白。

29.一种改变细胞的基因组的方法,所述方法包括在所述细胞中表达、或使所述细胞接触:根据权利要求1-24任一项所述的分离的蛋白或融合蛋白,以及具有与所述细胞的基因组的选定部分互补的区域的指导RNA。

30.根据权利要求29所述的方法,其中所述分离的蛋白或融合蛋白包含核定位序列、细胞穿透肽序列和/或亲和标签中的一种或多种。

31.根据权利要求30所述的方法,其中所述细胞是干细胞。

32.根据权利要求31所述的方法,其中所述细胞是胚胎干细胞、间充质干细胞、或诱导多能干细胞;所述细胞存在于活体动物中;或存在于胚胎中。

33.一种改变双链DNA(dsDNA)分子的方法,所述方法包括使dsDNA分子与根据权利要求1-24任一项所述的分离的蛋白或融合蛋白、以及具有与dsDNA分子的选定部分互补的区域的指导RNA接触。

34.根据权利要求33所述的方法,其中所述dsDNA分子是体外的。

35.根据权利要求33所述的方法,其中所述融合蛋白和RNA存在于核糖核蛋白复合物中。

# 具有改变的PAM特异性的工程化的CRISPR-Cas9核酸酶

[0001] 本申请是国际申请日2018年8月22日、申请号为201880068905.9(国际申请号PCT/US2018/047577)的专利申请“具有改变的PAM特异性的工程化的CRISPR-Cas9核酸酶”的分案申请。

[0002] 优先权声明

[0003] 本申请要求2017年8月23日提交的美国临时专利申请系列号No.62/549,303的权益;以及2018年3月12日提交的美国临时专利申请系列号No.62/641,687的权益。上述的全部内容通过引用并入本文。

[0004] 联邦政府赞助的研究或开发

[0005] 本发明是在美国国立卫生研究院授予的资助号No.GM105378、GM107427、GM118158和GM088040的政府资助下完成的。政府在本发明中享有一定的权利。

## 技术领域

[0006] 本发明至少部分涉及工程化的成簇规律间隔的短回文重复(Clustered Regularly Interspaced Short Palindromic Repeats,CRISPR)/CRISPR相关蛋白9(Cas9)核酸酶,其具有改变和改进的前间区序列邻近基序(Protospacer Adjacent Motif,PAM)特异性及其在基因组工程、表观基因组工程和基因组靶向中的用途。

## 背景技术

[0007] CRISPR-Cas9核酸酶能够在多种生物体和细胞类型中实现有效的、可定制的基因组编辑(Sander&Joung,Nat Biotechnol 32,347-355(2014);Hsuet等人,Cell 157,1262-1278(2014);Doudna&Charpentier,Science 346,1258096(2014);Barrangou&May,Expert Opin Biol Ther 15,311-314(2015))。Cas9的靶位点识别由两个称为crRNA和tracrRNA的短RNA指导(Deltcheva等人,Nature 471,602-607(2011);Jinek等人,Science 337,816-821(2012)),可以将其融合至嵌合单指导RNA(sgRNA)中(Jinek等人,Science 337,816-821(2012);Jinek等人,Elife 2,e00471(2013);Mali等人,Science 339,823-826(2013);Cong等人,Science 339,819-823(2013))。sgRNA的5’端(源自crRNA)可以与靶DNA位点碱基配对,从而可以通过Cas9/sgRNA复合物直接对位点特异性切割进行重新编程(Jinek等人,Science 337,816-821(2012))。但是,Cas9还必须识别与sgRNA碱基配对的DNA邻近的特定前间区序列邻近基序(PAM)(Mojica等人,Microbiology 155,733-740(2009);Shah等人,RNA Biol10,891-899(2013);Jinek等人,Science 337,816-821(2012);Sapranauskas等人,Nucleic Acids Res 39,9275-9282(2011);Horvath等人,J Bacteriol 190,1401-1412(2008)),这是启动序列特异性识别所必需的要求(Sternberg等人,Nature 507,62-67(2014)),但也可以限制用于基因组编辑的这些核酸酶的靶向范围。广泛使用的化脓性链球菌(Streptococcus pyogenes)Cas9(SpCas9)识别短的NGG PAM(Jinek等人,Science 337,816-821(2012);Jiang等人,Nat Biotechnol 31,233-239(2013)),在随机DNA序列中每8bp发生一次。相比之下,迄今已表征的其他Cas9直系同源物可以识别更长的PAM(Horvath等

人,JBacteriol 190,1401-1412(2008);Fonfara等人,Nucleic Acids Res 42,2577-2590(2014);Esvelt等人,Nat Methods 10,1116-1121(2013);Ran等人,Nature 520,186-191(2015);Zhang等人,Mol Cell 50,488-503(2013))。例如,金黄色葡萄球菌(Staphylococcus aureus)Cas9(SaCas9)是几种更适合病毒递送的较小Cas9直系同源物之一(Horvath等人,J Bacteriol 190,1401-1412(2008);Ran等人,Nature 520,186-191(2015);Zhang等人,Mol Cell 50,488-503(2013)),识别更长的NNGRRT PAM,预期在随机DNA中每32bp发生一次。扩大Cas9直系同源物的靶向范围对于包括修饰小遗传元件(例如转录因子结合位点(Canver等人Nature.527(7577):192-7(2015);Vierstra等人,Nat Methods.12(10):927-30(2015))或通过在PAM中定位序列差异进行等位基因特异性改变(Courtney,D.G.等人Gene Ther.23(1):108-12(2015))等各种应用是重要的。我们先前工程化了可以识别具有NGA和NGCG PAM序列的位点的SpCas9变体(Kleinstiver等人,Nature 2015;WO 2016141224),但是许多替代的PAM序列仍然是不可靶向的。

## 发明内容

[0008] 如本文所述,常用的化脓性链球菌Cas9(SpCas9)通过使用结构信息、基于细菌选择的定向进化和组合设计进行工程化以识别新的PAM序列。这些改变后的PAM特异性变体能够对野生型SpCas9无法有效靶向的人类细胞中的报道基因位点和内源性基因位点进行强有力的编辑。本发现提供广泛有用的SpCas9变体,在本文中统称为“变体”或“所述变体”。

[0009] 在第一方面,本发明提供分离的化脓性链球菌Cas9(SpCas9)蛋白,其在以下三个、四个、五个或全部六个位置具有突变:D1135、S1136、G1218、E1219、R1335和/或T1337,例如,在D1135、S1136、G1218、E1219、R1335和/或T1337(D1135X/S1136X/G1218X/E1219X/R1335X/T1337X,其中X为任意氨基酸)中的二个、三个、四个、五个或全部的位置具有突变,例如,包含具有所示突变的与SEQ ID NO:1的氨基酸序列至少80%、90%或95%相同的序列。在一些实施方案中,所述变体SpCas9蛋白包含表A、1、2或3中所示的突变的组,例如以下突变的组之一:LRSVQL、LRKIQK、LRSVQK、LWKIQK、VRKIQK、LWKIQK、IRAVQL、VRKLRS、GRKIQK、SWRVVV、SWKVLK、TAHFKV、MSGVKC、LRSVRS、SKTLRP、MWVHLN、TWSMRG、KRRCKV、VRAVQL、VSSVRS、VRSVRS、SRMHCK、GWKLLR、GWKOQK、VAKLLR、VAKIQK、VAKILR、GRKILR、VRKLLR、IRAVQL、VRKIQK或VRMHCK变体(例如,对于NGTN PAM);MQKSER、VRKSER、ICKSER、LRSVER、LWLETR、LSRWER、MQSVQL、VRREER、ICCCER、LSRWQR、LWRVVA、WMQAYG、LWRSEY、SQSWRS、LKAWRS、LWGWQH、MCSFER、LWMREQ、LWRVVA、HSSWVR、MWSEPT、GSNYQS、FMQWVN、YCSWVG、MCAWCG、FMQWVR或SSKWPA变体(例如,对于NGCN PAM);LRSVRS、SRQMRG、MRARKE、SRMHCK、VRREQR、VRGEQR、LRLSAR、AWTEVTR、KWMMCG、VRGAKE、AWNFQV、LWTTLN、SRMHCK、CWCQCV、AEEQQR、GWEKVR、NRAVNG、LRSYLH、VRGNNR、VQDAQR、GWRQSK、AWLCLS、KWARVV、MWAARP、SRMHCK、VKMAKG、QRKTRE、LCRQQR、CWSHQR、SRTHTQ、LWEVIR、VSSVRS、VRSVRS、IRAVRS、SRSVRS、LWKIQK或VRMHCK变体(例如,对于NGAN PAM)。在一些实施方案中,所述突变不是VSREER(也称为VRER)或VSREQR(也称为VRQR)。

[0010] 在一些实施方案中,所述变体SpCas9蛋白包含一种或多种降低核酸酶活性的突变,所述突变选自由在D10、E762、D839、H983或D986处;和在H840或N863处的突变组成的组。在一些实施方案中,所述突变是:(i)D10A或D10N,和(ii)H840A、H840N或H840Y。

[0011] 在一些实施方案中,变体SpCas9蛋白包含一种或多种增加特异性的突变,所述突

变选自由在N497、K526、R661、R691、N692、M694、Q695、H698、K810、K848、Q926、K1003和/或R0160处的突变组成的组。在一些实施方案中,所述突变为:N692A、Q695A、Q926A、H698A、N497A、K526A、R661A、R691A、M694A、K810A、K848A、K1003A、R0160A、Y450A/Q695A、L169A/Q695A、Q695A/Q926A、Q695A/D1135E、Q926A/D1135E、Y450A/D1135E、L169A/Y450A/Q695A、L169A/Q695A/Q926A、Y450A/Q695A/Q926A、R661A/Q695A/Q926A、N497A/Q695A/Q926A、Y450A/Q695A/D1135E、Y450A/Q926A/D1135E、Q695A/Q926A/D1135E、L169A/Y450A/Q695A/Q926A、L169A/R661A/Q695A/Q926A、Y450A/R661A/Q695A/Q926A、N497A/Q695A/Q926A/D1135E、R661A/Q695A/Q926A/D1135E和Y450A/Q695A/Q926A/D1135E;N692A/M694A/Q695A/H698A、N692A/M694A/Q695A/H698A/Q926A;N692A/M694A/Q695A/Q926A;N692A/M694A/H698A/Q926A;N692A/Q695A/H698A/Q926A;M694A/Q695A/H698A/Q926A;N692A/Q695A/H698A;N692A/M694A/Q695A;N692A/H698A/Q926A;N692A/M694A/Q926A;N692A/M694A/H698A;M694A/Q695A/H698A;M694A/Q695A/Q926A;Q695A/H698A/Q926A;G582A/V583A/E584A/D585A/N588A/Q926A;G582A/V583A/E584A/D585A/N588A;T657A/G658A/W659A/R661A/Q926A;T657A/G658A/W659A/R661A;F491A/M495A/T496A/N497A/Q926A;F491A/M495A/T496A/N497A;K918A/V922A/R925A/Q926A;或918A/V922A/R925A;K855A;K810A/K1003A/R1060A;或K848A/K1003A/R1060A。在一些实施方案中,所述蛋白不包括在K526或R691处的突变。

[0012] 本文还提供融合蛋白,其包含经任选的间插接头与异源功能域融合的本文所述的分离的变体SpCas9蛋白,其中所述接头不干扰所述融合蛋白的活性。在一些实施方案中,所述异源功能结构域是转录激活结构域。在一些实施方案中,所述转录激活结构域来自VP64或NF-κB p65。在一些实施方案中,所述异源功能结构域是转录沉默子或转录阻遏结构域。在一些实施方案中,所述转录阻遏结构域是Krueppel相关盒(KRAB)结构域、ERF阻遏物结构域(ERD)、或mSin3A相互作用结构域(SID)。在一些实施方案中,所述转录沉默子是异染色质蛋白1(HP1),例如HP1α或HP1β。在一些实施方案中,所述异源功能结构域是修饰DNA的甲基化状态的酶。在一些实施方案中,所述修饰DNA的甲基化状态的酶是DNA甲基转移酶(DNMT)或TET蛋白。在一些实施方案中,所述TET蛋白是TET1。在一些实施方案中,所述异源功能结构域是修饰组蛋白亚基的酶。在一些实施方案中,所述修饰组蛋白亚基的酶是组蛋白乙酰转移酶(HAT)、组蛋白去乙酰化酶(HDAC)、组蛋白甲基转移酶(HMT)、或组蛋白去甲基化酶。

[0013] 在一些实施方案中,所述异源功能结构域是碱基编辑器,例如胞苷脱氨酶结构域,例如来自脱氨酶的载脂蛋白B mRNA编辑酶,催化多肽样(apolipoprotein B mRNA-editing enzyme,catalytic polypeptide-like,APOBEC)家族,包括APOBEC1、APOBEC2、APOBEC3A、APOBEC3B、APOBEC3C、APOBEC3D/E、APOBEC3F、APOBEC3G、APOBEC3H或APOBEC4;活化诱导的胞苷脱氨酶(AID),例如活化诱导的胞苷脱氨酶(AICDA);胞嘧啶脱氨酶1(CDA1)或CDA2;或作用于tRNA的胞嘧啶脱氨酶(CDAT)。在一些实施方案中,所述异源功能结构域是修饰腺苷DNA碱基的脱氨酶,例如,所述脱氨酶是腺苷脱氨酶1(ADA1)、ADA2;和作用于RNA的腺苷脱氨酶1(ADAR1)、ADAR2、ADAR3;作用于tRNA的腺苷脱氨酶1(ADAT1)、ADAT2、ADAT3;以及天然存在或工程化的tRNA特异性腺苷脱氨酶(TadA)。在一些实施方案中,所述异源功能结构域是生物系链。在一些实施方案中,所述生物系链是MS2、Csy4或λN蛋白。在一些实施方案中,所述异源功能结构域是FokI。

[0014] 在一些实施方案中,所述异源功能结构域是抑制或增强内源DNA修复或碱基切除修复(BER)途径的酶、结构域或肽,例如尿嘧啶DNA糖苷酶抑制剂(UGI),其抑制尿嘧啶DNA糖基化酶(UDG,也称为尿嘧啶N-糖基化酶,或UNG)介导的尿嘧啶切除以引发BER;或DNA末端结合蛋白,例如来自噬菌体Mu的Gam。

[0015] 本文还提供编码本文描述的变体SpCas9蛋白的分离的核酸,以及载体,所述载体包含所述分离的核酸,任选地可操作地连接至一个或多个调节结构域用于表达本文描述的变体SpCas9蛋白。本文还提供宿主细胞,例如哺乳动物宿主细胞,其包含本文所述的核酸,并任选地表达本文所述的变体SpCas9蛋白。本文还提供核糖核蛋白(RNP)复合物,其包括本文所述的变体spCas9蛋白和靶向具有由所述变体蛋白靶向的PAM序列的序列的指导RNA。

[0016] 本文还提供通过在细胞中表达本文所述的分离的变体SpCas9蛋白和具有与细胞的基因组的选定部分互补的区域的指导RNA来改变所述细胞的基因组的方法。

[0017] 本文还提供通过在细胞中表达所述变体蛋白和具有与细胞的基因组的选定部分互补的区域的指导RNA来改变例如选择性改变细胞的基因组的方法。

[0018] 还提供通过使细胞与本文所述的蛋白变体和具有与细胞的基因组的选定部分互补的区域的指导RNA接触来改变例如选择性改变细胞的基因组的方法。

[0019] 在一些实施方案中,所述分离的蛋白或融合蛋白包含一个或多个核定位序列、细胞穿透肽序列和/或亲和标签。

[0020] 在本文描述的方法的一些实施方案中,所述细胞是干细胞,例如胚胎干细胞、间充质干细胞或诱导多能干细胞;所述细胞存在于活体动物中;或存在于胚胎中,例如哺乳动物、昆虫、或鱼(例如斑马鱼)的胚胎或胚胎细胞中。

[0021] 此外,本文提供用于改变双链DNA(dsDNA)分子的方法,例如体外方法。所述方法包括使dsDNA分子与本文所述的一种或多种变体蛋白和具有与所述dsDNA分子的选定部分互补的区域的指导RNA接触。

[0022] 除非另有定义,否则本文所用的所有技术和科学术语与本发明所属领域的普通技术人员通常所理解的含义相同。本文描述用于本发明的方法和材料。也可以使用本领域已知的其他合适的方法和材料。材料、方法和实例仅是说明性的,而无意于限制。本文提及的所有出版物、专利申请、专利、序列、数据库条目和其他参考文献均通过引用将其全部内容并入本文。在发生冲突的情况下,以本说明书包括定义为准。

[0023] 根据以下详细描述和附图、以及权利要求,本发明的其他特征和优点将变得显而易见。

## 附图说明

[0024] 专利或申请文件包含至少一张彩色附图。专利局将根据要求和必要费用的支付提供具有彩色附图的本专利或专利申请公开的副本。

[0025] 图1A-B。SpCas9 PAM变体文库的生成和筛选。(a)为了生成在对于PAM识别重要的六个位置具有不同氨基酸组合的SpCas9变体的文库,将简并的寡核苷酸文库克隆到SpCas9编码序列的三个区域中,以随机化红色突出显示的六个密码子。(b)细菌选择测定的示意图,其中从一个质粒表达SpCas9变体和sgRNA,另一个质粒编码可诱导的毒性基因和可定制的靶位点。将两种质粒共同转化细菌,只有具有能识别毒性质粒靶位点编码的PAM的SpCas9

PAM变体的细胞才能在诱导毒性基因表达的培养基上存活(选择性培养基含有10mM阿拉伯糖)。

[0026] 图2A-F。从细菌选择获得的识别具有NGTN PAM的位点的SpCas9 PAM变体。(A-F)将来自针对携带NGTG、NGTT、NGTC或NGTAPAM序列的阳性选择质粒的最初筛选的编码SpCas9变体的质粒在细菌筛选测定中重新筛选,以评估对具有NGTN PAM序列的靶位点的活性。对于子图C、E和F,还将针对具有NGAG、NGAT、NGAC和NGAAPAM的靶位点的变体进行筛选。通过比较非选择性培养基(仅氯霉素)上的菌落数与选择性培养基(氯霉素+10mM阿拉伯糖)上的那些来计算变体对具有指定PAM的位点的近似活性,并计算%存活率(表1)。数字具有星号的变体表示已在先前检测中试验的质粒(并显示在该图上面的子图中),并且这些变体的氨基酸序列见表1。

[0027] 图3A-C。SpCas9 NGTN PAM变体在人细胞中的活性。(A,B)人细胞EGFP破坏测定,以评估各种SpCas9变体(通过选择或合理设计获得的)针对具有NGTA、NGTC、NGTT和NGTG PAM的EGFP中的位点的活性。误差线代表n=2-7的s.e.m.。(C)表现最出色的SpCas9-NGTN PAM变体之一的SpCas9-LRSVQL变体针对具有NGTG、NGTA、NGTC或NGTT PAM的位点的内源基因破坏活性。通过T7E1测定定量基因破坏。

[0028] 图4A-B。从细菌选择中获得的识别具有NGCN PAM的位点的SpCas9PAM变体。(A-B)将来自针对携带NGCG、NGCT、NGCC或NGCAPAM序列的阳性选择质粒的初始筛选的编码SpCas9变体的质粒在细菌筛选测定中重新筛选,以评估对具有NGCN PAM序列的靶位点的活性。通过比较非选择性培养基(仅氯霉素)上的菌落数与选择性培养基(氯霉素+10mM阿拉伯糖)上的那些来计算变体对具有指定PAM的位点的近似活性,并计算%存活率(表2)。数字具有星号的变体表示已在先前检测中试验的质粒(并显示在该图上面的子图中),这些变体的氨基酸序列见表2。

[0029] 图5A-E。SpCas9 NGCN PAM变体在人细胞中的活性。(A-E)人细胞EGFP破坏测定,以评估各种SpCas9变体(通过选择或合理设计而获得的)针对具有NGCA、NGCC、NGCT和NGCG PAM的EGFP中的位点的活性。对具有(A)NGCAPAM位点、(B)NGCC PAM位点、(C)NGCT PAM位点、(D)NGCC和NGCT PAM位点以及(E)各种NGCN PAM位点的位点的活性。

[0030] 图6A-B。从细菌选择获得的识别具有NGAN PAM的位点的SpCas9 PAM变体。(A-B)将来自针对携带NGAG、NGAT、NGAC或NGCAPAM序列的阳性选择质粒的初始筛选的编码SpCas9变体的质粒在细菌筛选测定中重新筛选,以评估针对NGAN PAM序列的靶位点的活性。通过比较非选择性培养基(仅氯霉素)上的菌落数与选择性培养基(氯霉素+10mM阿拉伯糖)上的那些来计算变体对具有指定PAM的位点的近似活性,并计算%存活率(表3)。这些变体的氨基酸序列可见表3。

[0031] 图7A-B。SpCas9 NGAN PAM变体在人细胞中的活性。(A-B)人细胞EGFP破坏测定,以评估各种SpCas9变体(通过选择或合理设计而获得)对具有(A)NGAA、NGAC、NGAT和NGAG PAM或(B)各种NGAC PAM的EGFP中的位点的活性。

[0032] 图8。人细胞中SpCas9 PAM变体与xCas9的核酸酶活性的比较。内源基因破坏活性通过T7E1测定评估。

[0033] 图9。SpCas9 PAM变体与xCas9对人细胞内源性位点的基本编辑活性的比较。(A)试验SpCas9 PAM变体的BE3版本(Komor等人,Nature.2016年5月19日;533(7603):420-4)的C-

至-T碱基编辑活性,通过靶向深度测序和用CRISPResso进行的分析来评估(Pinello等人,Nat Biotechnol.2016Jul 12;34(7):695-697).Nat Biotechnol.2016Jul 12;34(7):695-697)。将在x轴上标记的编辑的胞嘧啶核苷酸在作为Cas9靶序列中最靠近PAM的位置的位置1开始编号。(B)检测SpCas9 PAM变体的ABE7.10版本(Gaudelli等人,Nature.2017年11月23日;551(7681):464-471)的A-至-G碱基编辑活性,通过靶向深度测序和用CRISPResso进行的分析来评估。将在x轴上标记的编辑的腺苷核苷酸在作为Cas9靶序列中最靠近PAM的位置的位置1开始编号。

[0034] 表1:针对具有NGTN PAM的位点的变体在细菌中的选择结果和活性

[0035] 表2:针对具有NGCN PAM的位点的变体在细菌中的选择结果和活性

[0036] 表3:针对具有NGAN PAM的位点的变体在细菌中的选择结果和活性

## 具体实施方式

[0037] 化脓性链球菌Cas9(SpCas9)识别前间区序列邻近基序(PAM)是靶DNA识别的关键第一步,使SpCas9能够结合并水解DNA。尽管CRISPR-Cas9核酸酶被广泛用于基因组编辑[1-4],但Cas9可以切割的序列范围受到靶位点中对特定PAM的需求的限制[5,6]。例如,SpCas9,迄今为止功能最强且使用最广泛的Cas9,主要识别NGG PAM。结果,通常难以以各种基因组编辑应用所需的精度来靶向双链断裂(DSB)。此外,Cas9对PAM的识别不完善会导致产生不需要的脱靶突变[7,8]。有目的地改变和/或改进PAM特异性的Cas9衍生物将解决这些局限性。

[0038] 晶体结构显示,野生型SpCas9利用两个精氨酸氨基酸侧链(R1333和R1335)与经典NGG PAM序列的鸟嘌呤形成碱基特异性接触。然而,为了改变PAM识别并改进SpCas9的靶向范围,我们和其他人已证实仅突变这些精氨酸其中任何一个或突变两个不会改变PAM的偏好(Anders等人,Nature 2014;Kleinstiver等人,Nature 2015;WO 2016141224)。我们先前采用了选择方法来进化SpCas9的变体,其可以靶向NGA和NGCG PAM序列(Kleinstiver等人,Nature 2015;WO 2016141224)。然而,许多其他可选的PAM序列仍然无法靶向。

[0039] 为了通过使当前无法接近的PAM序列能够被靶向来进一步扩展SpCas9的实用性,我们构想了一种替代策略,以选择能够识别新PAM序列的SpCas9变体。先前已经确定SpCas9编码序列内的某些位置对于PAM识别是重要的(Kleinstiver等人,Nature 2015;WO 2016141224),我们进行了集中的饱和诱变方法,其中我们随机化PAM相互作用结构域的三个独立区域内的六个氨基酸,以生成SpCas9变体的文库,在以下这些位置具有不同的密码子用法:D1135/S1136、G1218/E1219和R1335/T1337。为此,我们在含有编码这六个氨基酸的SpCas9密码子上顺序克隆编码NNS核苷酸三联体的随机寡核苷酸盒(其中N是任何核苷酸,S是G或C)(图1A)。然后使用我们先前描述的细菌阳性选择测定法(Kleinstiver等人,Nature 2015)进行选择得到的SpCas9变体的文库,以鉴定可以切割具有各种NGNN PAM序列的靶位点的变体(图1B)。简而言之,如果表达的SpCas9变体可以识别阳性选择质粒中编码的靶位点(PAM和间隔子序列),则细菌只能在选择性条件下存活(铺板在10mM阿拉伯糖上,其诱导ccdB毒性基因的转录)。强大的PAM识别将导致选择质粒的水解,阻止ccdB表达的诱导,从而使细菌生长。因此,在筛选SpCas9文库时,在含有10mM阿拉伯糖的培养基上生长的菌落预期编码SpCas9 PAM变体,该变体可以靶向具有可替换的感兴趣的PAM的位点(图1B)。

[0040] 具有改变的PAM特异性的工程化的Cas9变体

[0041] 这项研究中工程化的SpCas9变体大大增加了野生型SpCas9可以接近的靶位点的范围,进一步增加了使用CRISPR-Cas9平台进行高效HDR、靶向NHEJ介导的小遗传元件的插入缺失以及利用PAM的要求来区分同一细胞中的两个不同等位基因的机会。现在经过选择和合理设计的变体可以靶向以前无法接近的含有位点的NGTN和NGCH(其中H为A、C或T)PAM,并且该变体可以提高针对NGAC的活性,改进准确和高分辨率基因组编辑的前景。改变的PAM特异性SpCas9变体可以有效破坏细菌和人类细胞二者中SpCas9目前无法靶向的内源基因位点,表明它们将在多种不同的细胞类型和生物体中发挥作用。

[0042] 可以通过简单的定点诱变将本文所述的所有SpCas9变体快速整合至现有的和广泛使用的载体中,并且由于它们仅需要包含在PAM相互作用结构域内的少量突变,因此这些变体也应与先前描述的对SpCas9平台进行的其他改进(例如截短的sgRNA (Tsai等人,Nat Biotechnol 33,187-197(2015);Fu等人,Nat Biotechnol 32,279-284(2014))、切口酶突变(Mali等人,Nat Biotechnol 31,833-838(2013);Ran等人,Cell 154,1380-1389(2013))、二聚体FokI-dCas9融合体(Guilinger等人,Nat Biotechnol 32,577-582(2014);Tsai等人,Nat Biotechnol 32,569-576(2014))和高保真变体(Kleinstiver等人Nature 2016)一起发挥作用。

[0043] 具有改变的PAM特异性的SpCas9变体

[0044] 因此,本文提供SpCas9变体。SpCas9野生型序列如下:

[0045]

```
        10         20         30         40         50         60
MDKKYSIGLD IGTNSVGWAV ITDEYKVPSK KFKVLGNTDR HSIKKNLIGA LLFDSGETAE

        70         80         90        100        110        120
ATRLKRTARR RYTRRKNRIC YLQEIFSNEM AKVDDSFFHR LEESFLVEED KKHERHPIFG

       130        140        150        160        170        180
NIVDEVAYHE KYPTIYHLRK KLVDSTDKAD LRLIYLALAH MIKFRGHFLI EGDLNPDNSD

       190        200        210        220        230        240
VDKLFIQLVQ TYNQLFEENP INASGVDAKA ILSARLSKSR RLENLIAQLP GEKKNGLFGN

       250        260        270        280        290        300
LIALSLGLTP NFKSNFDLAE DAKLQLSKDT YDDDLDNLLA QIGDQYADLF LAAKNLSDAI

       310        320        330        340        350        360
LLSDILRVNT EITKAPLSAS MIKRYDEHHQ DLTLLKALVR QQLPEKYKEI FFDQSKNGYA

       370        380        390        400        410        420
GYIDGGASQE EFYKFIKPIL EKMDGTEELL VKLNREDLLR KQRTFDNGSI PHQIHLGELH

       430        440        450        460        470        480
AILRRQEDFY PFLKDNREKI EKILTFRIPY YVGPLARGNS RFAWMTRKSE ETITPWNFEE

       490        500        510        520        530        540
VVDKGASAQS FIERMTNFDK NLPNEKVLPK HSLLYEYFTV YNELTKVKYV TEGMRKPAFL

       550        560        570        580        590        600
SGEQKKAIVD LLFKTNRKVT VKQLKEDYFK KIECFDSVEI SGVEDRFNAS LGTYHDLLKI

       610        620        630        640        650        660
IKDKDFLDNE ENEDILEDIV LTLTLFEDRE MIEERLKTYA HLFDDKVMKQ LKRRRYTGWG
```

670 680 690 700 710 720
RLSRKLINGI RDKQSGKTIL DFLKSDGFAN RNFMQLIHDD SLTFKEDIQK AQVSGQGDSL

730 740 750 760 770 780
HEHIANLAGS PAIKKGILQT VKVVDELVKV MGRHKPENIV IEMARENQTT QKGQKNSRER

790 800 810 820 830 840
MKRIEEGIKE LGSQILKEHP VENTQLQNEK LYLYYLQNGR DMYVDQELDI NRLSDYDVDH

850 860 870 880 890 900
IVPQSFLKDD SIDNKVLTRS DKNRGKSDNV PSEEVVKKMK NYWRQLLNAK LITQRKFDNL

910 920 930 940 950 960
TKAERGGLSE LDKAGFIKRQ LVETRQITKH VAQILDSRMN TKYDENDKLI REVKVITLKS

970 980 990 1000 1010 1020
KLVSDFRKDF QFYKVREINN YHHAHDAYLN AVVGTALIKK YPKLESEFVY GDYKVYDVRK

[0046]

1030 1040 1050 1060 1070 1080
MIAKSEQEIG KATAKYFFYS NIMNFFKTEI TLANGEIRKR PLIETNGETG EIVWDKGRDF

1090 1100 1110 1120 1130 1140
ATVRKVLSMP QVNIVKKTEV QTGGFSKESI LPKRNSDKLI ARKKDWDPKK YGGFDSPTVA

1150 1160 1170 1180 1190 1200
YSVLVVAKVE KGKSKKLKSV KELLGITIME RSSFEKNPID FLEAKGYKEV KKDLIIKLPK

1210 1220 1230 1240 1250 1260
YSLFELENGR KRMLASAGEL QKGNELALPS KYVNFLYLAS HYEKLKGSPE DNEQKQLFVE

1270 1280 1290 1300 1310 1320
QHKHYLDEII EQISEFSKRV ILADANLDKV LSAYNKHRDK PIREQAENII HLFTLTNLGA

1330 1340 1350 1360
PAAFKYFDTT IDRKRYTSTK EVLDATLIHQ SITGLYETRI DLSQLGGD (SEQ ID NO:1)

[0047] 本文所述的SpCas9变体可包括以下一个、两个、三个、四个、五个或全部六个位置的突变:D1135、S1136、G1218、E1219、R1335和/或T1337,例如,D1135X/S1136X/G1218X/E1219X/R1335X/T1337X,其中X是任何氨基酸(或与之类似的位置)。在一些实施方案中,SpCas9变体与SEQ ID NO:1的氨基酸序列至少80%,例如,至少85%、90%、或95%相同,例如,与SEQ ID NO:1的残基具有多至5%、10%、15%、或20%的差异,所述残基用例如保守突变取代。在优选的实施方案中,变体保留了亲本的所需活性,例如核酸酶活性(除非其中亲本是切口酶或死的Cas9)和/或与指导RNA和靶DNA相互作用的能力)。

[0048] 为了确定两个核酸序列的同一性百分比,将序列进行比对以达到最佳比较目的

(例如,可以在第一和第二个氨基酸或核酸序列之一或两者中引入缺口以实现最佳比对,并且非同源序列出于比较目的可以忽略不计)。用于比较目的而比对的参考序列的长度为至少80%的参考序列长度,并且在一些实施方案中为至少90%或100%。然后比较相应氨基酸位置或核苷酸位置处的核苷酸。当第一序列中的位置被第二序列中的对应位置相同的核苷酸占据时,则分子在那个位置是同一的(如本文所用,核酸"同一性"等同于核酸"同源性")。考虑到空位数和每个空位的长度(需要引入其以实现两个序列的最佳比对),两个序列之间的同一性百分比是序列共享的同一位置的数目的函数。两个多肽或核酸序列之间的同一性百分比以本领域技术范围内的各种方式确定,例如,使用可公开获得的计算机软件,例如Smith Waterman Alignment(Smith,T.F.和M.S.Waterman(1981)J Mol Biol 147:195-7);"BestFit"(Smith和Waterman,Advances in Applied Mathematics,482-489(1981)),并入GeneMatcher Plus™,Schwarz和Dayhof(1979)Atlas of Protein Sequence and Structure,Dayhof,M.O.,编辑,pp 353-358;BLAST程序(基本局部比对搜索工具;(Altschul,S.F.,W.Gish,等人(1990)J Mol Biol 215:403-10),BLAST-2、BLAST-P、BLAST-N、BLAST-X、WU-BLAST-2、ALIGN、ALIGN-2、CLUSTAL或Megalign(DNASTAR)软件。此外,本领域技术人员可以确定用于测量比对的合适参数,包括在被比较的序列全长范围内实现最大比对所需的任何算法。通常,对于蛋白质或核酸,比较的长度可以是任何长度,最大为全长并包括全长(例如5%、10%、20%、30%、40%、50%、60%、70%、80%、90%、95%或100%)。为了本发明的组合物和方法的目的,使用BLAST算法和默认参数将至少80%的全长序列的进行比对。

[0049] 为了本发明的目的,序列的比较和两个序列之间同一性百分比的确定可以使用Blossum 62评分矩阵来实现,其空位罚分为12,空位延伸罚分为4,移码空位罚分为5。

[0050] 保守取代通常包括以下组内的取代:甘氨酸、丙氨酸;缬氨酸、异亮氨酸、亮氨酸、天冬氨酸、谷氨酸、天冬酰胺、谷氨酰胺;丝氨酸、苏氨酸、赖氨酸、精氨酸;和苯丙氨酸、酪氨酸。

[0051] 在一些实施方案中,SpCas9变体包括表1、2或3中所示的突变的组,例如以下在D1135X/S1136X/G1218X/E1219X/R1335X/T1337X处的突变的组之一:SpCas9-LWKIQK、LWKIQK、IRAVQL、SWRVVV、SWKVLK、TAHFKV、MSGVKC、LRSVRS、SKTLRP、MWVHLN、TWSMRG、KRRCKV、VRAVQL、VSSVRS、VRSVRS、SRMHCK、GRKIQK、GWKLLR、GWKOQK、VAKLLR、VAKIQK、VAKILR、GRKILR、VRKLLR、LRSVQL、IRAVQL、VRKIQK、VRMHCK、LRKIQK、LRSVQK或VRKIQK变体(例如,对于NGTN PAM);WMQAYG、MQKSER、LWRSEY、SQSWRS、LKAWRS、LWGWQH、MCSFER、LWMREQ、LWRVVA、HSSWVR、MWSEPT、GSNYQS、FMQWVN、YCSWVG、MCAWCG、LWLETR、FMQWVR、SSKWPA、LSRWQR、ICCCER、VRKSER或ICKSER(例如,对于NGCN PAM);或LRLSAR、AWTEVTR、KWMMCG、VRGAKE、MRARKE、AWNFQV、LWTTLN、SRMHCK、CWCQCV、AEEQQR、GWEKVR、NRAVNG、SRQMRG、LRSYLH、VRGNNR、VQDAQR、GWRQSK、AWLCLS、KWARVV、MWAARP、SRMHCK、VKMAKG、QRKTRE、LCRQQR、CWSHQR、SRTHTQ、LWEVIR、VSSVRS、VRSVRS、LRSVRS、IRAVRS、SRSVRS、LWKIQK、VRMHCK或SRMHCK(例如,对于NGAN PAM)。在一些实施方案中,spCas9变体包括D1135L/S1136R/G1218S/E1219V/R1335X/T1337X,例如LRSVQL或LRSVRS。在一些实施方案中,在D1135处的残基为L、G、I、V、M或S。在一些实施方案中,在S1136处的残基为R、Q、W、S、或C。在一些实施方案中,在G1218处的残基为S、K、S、R、L、C、G、A或Q。在一些实施方案中,在E1219处的残基为V、I、S、E、W、C、A或R。在一些实施方案中,在

R1335处的残基为R、Q、E、V、T或K。在一些实施方案中,在T1337处的残基为S、K、L、R、A、E、T或G。在一些实施方案中,变体包括表A中的突变的组之一。

[0052]

表A

| NGTN PAM | NGCN PAM | NGAN PAM |
|---|---|---|
| LRSVRS | **MQKSER** | VRGAKE |
| GRKIQK | LWRVVA | MRARKE |
| **LRSVQL** | LWLETR | SRQMRG |
| IRAVQL | LSRWQR | **LRSVRS** |
| LRKIQK | ICCCER | |
| LRSVQK | VRKSER | |
| VRKIQK | ICKSER | |

[0053] 在一些实施方案中,SpCas9变体还包括以下氨基酸位置之一的突变,其降低或破坏Cas9的核酸酶活性:D10、E762、D839、H983或D986和H840或N863,例如D10A/D10N和H840A/H840N/H840Y,以使蛋白质的核酸酶部分催化失活;这些位置的取代基可以是丙氨酸(如Nishimasu等人,Cell 156,935-949(2014)中所述)或其他残基,例如谷氨酰胺、天冬酰胺、酪氨酸、丝氨酸或天冬氨酸,例如E762Q、H983N、H983Y、D986N、N863D、N863S、或N863H(参见WO 2014/152432)。在一些实施方案中,变体包括在D10A或H840A处的突变(其产生单链切口酶),或在D10A和H840A处的突变(其消除核酸酶活性;该突变体被称为死Cas9或dCas9)。

[0054] 在一些实施方案中,SpCas9变体还包括在一个或多个氨基酸位置的突变,其增加蛋白质的特异性(即,降低脱靶效应)。在一些实施方案中,SpCas9变体在以下残基处包括1、2、3、4、5、6、7、8、9、10、11、12或全部13个突变:N497、K526、R661、R691、N692、M694、Q695、H698、K810、K848、Q926、K1003和/或R0160。在一些实施方案中,突变是:N692A、Q695A、Q926A、H698A、N497A、K526A、R661A、R691A、M694A、K810A、K848A、K1003A、R0160A、Y450A/Q695A、L169A/Q695A、Q695A/Q926A、Q695A/D1135E、Q926A/D1135E、Y450A/D1135E、L169A/Y450A/Q695A、L169A/Q695A/Q926A、Y450A/Q695A/Q926A、R661A/Q695A/Q926A、N497A/Q695A/Q926A、Y450A/Q695A/D1135E、Y450A/Q926A/D1135E、Q695A/Q926A/D1135E、L169A/Y450A/Q695A/Q926A、L169A/R661A/Q695A/Q926A、Y450A/R661A/Q695A/Q926A、N497A/Q695A/Q926A/D1135E、R661A/Q695A/Q926A/D1135E和Y450A/Q695A/Q926A/D1135E;N692A/M694A/Q695A/H698A、N692A/M694A/Q695A/H698A/Q926A;N692A/M694A/Q695A/Q926A;N692A/M694A/H698A/Q926A;N692A/Q695A/H698A/Q926A;M694A/Q695A/H698A/Q926A;N692A/Q695A/H698A;N692A/M694A/Q695A;N692A/H698A/Q926A;N692A/M694A/Q926A;N692A/M694A/H698A;M694A/Q695A/H698A;M694A/Q695A/Q926A;Q695A/H698A/Q926A;G582A/V583A/E584A/D585A/N588A/Q926A;G582A/V583A/E584A/D585A/N588A;T657A/G658A/W659A/R661A/Q926A;T657A/G658A/W659A/R661A;F491A/M495A/T496A/N497A/Q926A;F491A/M495A/T496A/N497A;K918A/V922A/R925A/Q926A;或918A/V922A/R925A;K855A;K810A/K1003A/R1060A;或K848A/K1003A/R1060A。参见,例如,US9512446B1;Kleinstiver等人,Nature.2016Jan 28;529(7587):490-5;Slaymaker等人,Science.2016Jan 1;351(6268):84-8;Chen等人,Nature.2017Oct 19;550(7676):407-410;Tsai和

Joung,Nature Reviews Genetics 17:300-312(2016);Vakulskas等人,Nature Medicine 24:1216-1224(2018);Casini等人,Nat Biotechnol.2018 Mar;36(3):265-271。在一些实施方案中,变体不包括在K526或R691处的突变。

[0055] 在一些实施方案中,SpCas9变体包括在以下位置的一个、两个、三个、四个、五个、六个或全部七个突变:L169A、Y450、N497、R661、Q695、Q926和/或D1135E,例如,在一些实施方案中,变体SpCas9蛋白包含以下位置一个、两个、三个或全部四个突变:N497、R661、Q695和Q926,例如,以下突变的一个、两个、三个或全部四个:N497A、R661A、Q695A和Q926A。在一些实施方案中,变体SpCas9蛋白包含在Q695和/或Q926处的突变,以及任选地在L169、Y450、N497、R661和D1135E的一个、两个、三个、四个或全部五个处的突变,例如,包括但不限于Y450A/Q695A、L169A/Q695A、Q695A/Q926A、Q695A/D1135E、Q926A/D1135E、Y450A/D1135E、L169A/Y450A/Q695A、L169A/Q695A/Q926A、Y450A/Q695A/Q926A、R661A/Q695A/Q926A、N497A/Q695A/Q926A、Y450A/Q695A/D1135E、Y450A/Q926A/D1135E、Q695A/Q926A/D1135E、L169A/Y450A/Q695A/Q926A、L169A/R661A/Q695A/Q926A、Y450A/R661A/Q695A/Q926A、N497A/Q695A/Q926A/D1135E、R661A/Q695A/Q926A/D1135E和Y450A/Q695A/Q926A/D1135E。参见,例如,Kleinstiver等人,Nature 529:490-495(2016);WO 2017/040348;US 9,512,446。

[0056] 在一些实施方案中,SpCas9变体还包括以下一个、两个、三个、四个、五个、六个、七个或更多位置的突变:F491、M495、T496、N497、G582、V583、E584、D585、N588、T657、G658、W659、R661、N692、M694、Q695、H698、K918、V922和/或R925,并且任选地在Q926,优选包含与SEQ ID NO:1的氨基酸序列至少80%相同的序列,具有在一个、两个、三个、四个、五个、六个、七个或更多以下位置的突变:F491、M495、T496、N497、G582、V583、E584、D585、N588、T588、T657、G658、W659、R661、N692、M694、Q695、H698、K918、V922和/或R925,以及任选地在Q926,以及任选地一个或多个核定位序列、细胞穿透肽序列和/或亲和标签。

[0057] 在一些实施方案中,蛋白质在以下一个、两个、三个、或全部四个处包含突变:N692、M694、Q695、和H698;G582、V583、E584、D585和N588;T657、G658、W659和R661;F491、M495、T496和N497;或K918、V922、R925和Q926。

[0058] 在一些实施方案中,蛋白质包含以下一种、两种、三种、四种或全部突变:N692A、M694A、Q695A、和H698A;G582A、V583A、E584A、D585A和N588A;T657A、G658A、W659A和R661A;F491A、M495A、T496A、和N497A;或K918A、V922A、R925A和Q926A。

[0059] 在一些实施方案中,蛋白质包含以下突变:

[0060] N692A/M694A/Q695A/H698A。

[0061] 在一些实施方案中,蛋白质包含以下突变:

[0062] N692A/M694A/Q695A/H698A/Q926A;N692A/M694A/Q695A/Q926A;N692A/M694A/H698A/Q926A;N692A/Q695A/H698A/Q926A;M694A/Q695A/H698A/Q926A;N692A/Q695A/H698A;N692A/M694A/Q695A;N692A/H698A/Q926A;N692A/M694A/Q926A;N692A/M694A/H698A;M694A/Q695A/H698A;M694A/Q695A/Q926A;Q695A/H698A/Q926A;G582A/V583A/E584A/D585A/N588A/Q926A;G582A/V583A/E584A/D585A/N588A;T657A/G658A/W659A/R661A/Q926A;T657A/G658A/W659A/R661A;F491A/M495A/T496A/N497A/Q926A;F491A/M495A/T496A/N497A;K918A/V922A/R925A/Q926A;或918A/V922A/R925A。参见,例如,Chen等

人,“Enhanced proofreading governs CRISPR-Cas9 targeting accuracy,”bioRxiv,doi.org/10.1101/160036(2017年8月12日)。

[0063] 在一些实施方案中,变体蛋白包括在以下一个或多个处的突变:R780、K810、R832、K848、K855、K968、R976、H982、K1003、K1014、K1047和/或R1060,例如,R780A、K810A、R832A、K848A、K855A、K968A、R976A、H982A、K1003A、K1014A、K1047A和/或R1060A,例如K855A;K810A/K1003A/R1060A;(也称为eSpCas9 1.0);或K848A/K1003A/R1060A(也称为eSpCas9 1.1)(参见Slaymaker等人,Science.2016Jan 1;351(6268):84-8)。

[0064] 本文还提供编码SpCas9变体的分离的核酸;载体,其包含分离的核酸,任选地可操作地连接至一个或多个调节结构域用于表达变体蛋白;以及宿主细胞,例如哺乳动物宿主细胞,其包含所述核酸并任选地表达所述变体蛋白。

[0065] 本文所述的变体可用于改变细胞的基因组;该方法通常包括在细胞中表达变体蛋白,以及具有与细胞的基因组的选定部分互补的区域的指导RNA。选择性改变细胞基因组的方法在本领域中是已知的,参见,例如,US8,697,359;US2010/0076057;US2011/0189776;US2011/0223638;US2013/0130248;WO/2008/108989;WO/2010/054108;WO/2012/164565;WO/2013/098244;WO/2013/176772;US20150050699;US20150045546;US20150031134;US20150024500;US20140377868;US20140357530;US20140349400;US20140335620;US20140335063;US20140315985;US20140310830;US20140310828;US20140309487;US20140304853;US20140298547;US20140295556;US20140294773;US20140287938;US20140273234;US20140273232;US20140273231;US20140273230;US20140271987;US20140256046;US20140248702;US20140242702;US20140242700;US20140242699;US20140242664;US20140234972;US20140227787;US20140212869;US20140201857;US20140199767;US20140189896;US20140186958;US20140186919;US20140186843;US20140179770;US20140179006;US20140170753;Makarova等人,"Evolution and classification of the CRISPR-Cas systems"9(6)Nature Reviews Microbiology 467-477(1-23)(2011年6月);Wiedenheft等人,"RNA-guided genetic silencing systems in bacteria and archaea"482Nature 331-338(2012年2月16日);Gasiunas等人,"Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria"109(39)Proceedings of the National Academy of Sciences USA E2579-E2586(2012年9月4日);Jinek等人,"A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity"337Science 816-821(2012年8月17日);Carroll,"ACRISPR Approach to Gene Targeting"20(9)Molecular Therapy 1658-1660(2012年9月);美国专利申请号No.61/652,086,2012年5月25日提交;Al-Attar等人,Clustered Regularly Interspaced Short Palindromic Repeats(CRISPRs):The Hallmark of an Ingenious Antiviral Defense Mechanismin Prokaryotes,Biol Chem.(2011)第392卷,第4期,277-289页;Hale等人,Essential Features and Rational Design of CRISPR RNAs That Function With the Cas RAMP Module Complex to Cleave RNAs,Molecular Cell,(2012)第45卷,第3期,292-302。

[0066] 本文所述的变体蛋白可代替前述参考文献中描述的SpCas9蛋白与指导RNA一起使用,所述指导RNA靶向具有根据表1、2或3的PAM序列的序列。

[0067] 此外,本文所述的变体可代替本领域已知的野生型Cas9或其他Cas9突变(例如上述dCas9或Cas9切口酶)用于融合蛋白,例如具有如WO 2014/124284中所述异源功能域的融合蛋白。例如,可以将变体,优选包含一种或多种核酸酶减少或杀死突变的变体,在Cas9的N或C末端与转录激活结构域或其他异源功能结构域融合(例如,转录阻遏蛋白(例如,KRAB、ERD,SID、及其他,例如,ets2阻遏因子(ERF)阻遏物结构域(ERD)的氨基酸473-530、KOX1的KRAB结构域的氨基酸1-97、或Mad mSIN3相互作用结构域(SID)的氨基酸1-36;参见Beerli等人,PNAS USA 95:14628-14633(1998))或沉默子,例如杂染色质蛋白1(HP1,也称为swi6),例如,HP1α或HP1β;可以募集与固定RNA结合序列融合的长非编码RNA(lncRNA),例如与MS2外壳蛋白、核糖核酸内切酶Csy4、或λN蛋白结合的那些;修饰DNA的甲基化状态的酶(例如,DNA甲基转移酶(DNMT)或TET蛋白);修饰组蛋白亚基的酶(例如,组蛋白乙酰转移酶(HAT)、组蛋白去乙酰化酶(HDAC)、组蛋白甲基转移酶(例如,用于赖氨酸或精氨酸残基的甲基化)或组蛋白去甲基化酶(例如,用于赖氨酸或精氨酸残基的去甲基化)。

[0068] 在一些实施方案中,异源功能结构域是碱基编辑器(base editor),例如修饰胞嘧啶DNA碱基的脱氨酶,例如,来自胞苷脱氨酶的载脂蛋白B mRNA编辑酶,催化多肽样(APOBEC)家族,包括APOBEC1、APOBEC2、APOBEC3A、APOBEC3B、APOBEC3C、APOBEC3D/E、APOBEC3F、APOBEC3G、APOBEC3H和APOBEC4(参见,例如,Yang等人,J Genet Genomics.2017年9月20日;44(9):423-437);活化诱导的胞苷脱氨酶(AID),例如活化诱导的胞苷脱氨酶(AICDA);胞嘧啶脱氨酶1(CDA1)和CDA2;和作用于tRNA的胞嘧啶脱氨酶(CDAT)。下表提供示例性序列;也可以使用其他序列。

[0069]

| | GenBank 登录号 | |
|---|---|---|
| **脱氨酶** | 核酸 | 氨基酸 |
| **hAID/AICDA** | NM_020661.3 同工型 1<br>NM_020661.3 同工型 2 | NP_065712.1 变体 1<br>NP_065712.1 变体 2 |
| **APOBEC1** | NM_001644.4 同工型 a<br>NM_005889.3 同工型 b | NP_001635.2 变体 1<br>NP_005880.2 变体 3 |
| **APOBEC2** | NM_006789.3 | NP_006780.1 |
| **APOBEC3A** | NM_145699.3 同工型 a<br>NM_001270406.1 同工型 b | NP_663745.1 变体 1<br>NP_001257335.1 变体 2 |
| **APOBEC3B** | NM_004900.4 同工型 a<br>NM_001270411.1 同工型 b | NP_004891.4 变体 1<br>NP_001257340.1 变体 2 |
| **APOBEC3C** | NM_014508.2 | NP_055323.2 |
| **APOBEC3D/E** | NM_152426.3 | NP_689639.2 |
| **APOBEC3F** | NM_145298.5 同工型 a<br>NM_001006666.1 同工型 b | NP_660341.2 变体 1<br>NP_001006667.1 变体 2 |
| **APOBEC3G** | NM_021822.3 (同工型 a) | NP_068594.1 (变体 1) |
| **APOBEC3H** | NM_001166003.2 | NP_001159475.2 (变体 SV-200) |
| **APOBEC4** | NM_203454.2 | NP_982279.1 |
| **CDA1*** | NM_127515.4 | NP_179547.1 |

[0070] *来自酿酒酵母(Saccharomyces cerevisiae)S288C

[0071] 在一些实施方案中,异源功能结构域是修饰腺苷DNA碱基的脱氨酶,例如,脱氨酶是腺苷脱氨酶1(ADA1)、ADA2;和作用于RNA的腺苷脱氨酶1(ADAR1)、ADAR2、ADAR3(参见,例如,Savva等人,Genome Biol.2012 Dec 28;13(12):252);作用于tRNA的腺苷脱氨酶1(ADAT1)、ADAT2、ADAT3(参见Keegan等人,RNA.2017年9月;23(9):1317-1328以及Schaub和Keller,Biochimie.2002年8月;84(8):791-803);和天然存在的或工程化的tRNA特异性腺苷脱氨酶(TadA)(参见,例如,Gaudelli等人,Nature.2017年11月23日;551(7681):464-471)(NP_417054.2(大肠杆菌菌株K-12亚菌株MG1655);参见,例如,Wolf等人,EMBO J.2002年7月15日;21(14):3841-51)。下表提供示例性序列;也可以使用其他序列。

[0072]

| | GenBank 登录号 | |
|---|---|---|
| **脱氨酶** | 核酸 | 氨基酸 |
| **ADA (ADA1)** | NM_000022.3 变体 1 | NP_000013.2 同工型 1 |
| **ADA2** | NM_001282225.1 | NP_001269154.1 |
| **ADAR** | NM_001111.4 | NP_001102.2 |
| **ADAR2 (ADARB1)** | NM_001112.3 变体 1 | NP_001103.1 同工型 1 |
| **ADAR3 (ADARB2)** | NM_018702.3 | NP_061172.1 |
| **ADAT1** | NM_012091.4 变体 1 | NP_036223.2 同工型 1 |
| **ADAT2** | NM_182503.2 变体 1 | NP_872309.2 同工型 1 |
| **ADAT3** | NM_138422.3 变体 1 | NP_612431.2 同工型 1 |

[0073] 在一些实施方案中,异源功能结构域是抑制或增强内源DNA修复或碱基切除修复(BER)途径的酶、结构域或肽,例如胸腺嘧啶DNA糖基化酶(TDG;GenBank登录号NM_003211.4(核酸)和NP_003202.3(蛋白质))或尿嘧啶DNA糖基化酶(UDG,也称为尿嘧啶N-糖基化酶,或UNG;GenBank登录号NM_003362.3(核酸)和NP_003353.1(蛋白质))或尿嘧啶DNA糖基化酶抑制剂(UGI),其抑制UNG介导的尿嘧啶切除以引发BER(参见,例如,Mol等人,Cell82,701-708(1995);Komor等人,Nature.2016年5月19日;533(7603));或DNA末端结合蛋白例如Gam,它是噬菌体Mu中的蛋白,结合自由的DNA末端,抑制DNA修复酶并导致更精确的编辑(较少的意外碱基编辑)。参见,例如,Komor等人,Sci Adv.2017Aug 30;3(8):eaao4774。

[0074] 参见,例如,Komor等人,Nature.2016年5月19日;533(7603):420-4;Nishida等人,Science.2016年9月16日;353(6305).pii:aaf8729;Rees等人,Nat Commun.2017年6月6日;8:15790;或Kim等人,Nat Biotechnol.2017年4月;35(4):371-376),也可以使用本领域已知的。

[0075] 结构域的许多序列催化DNA中甲基化胞嘧啶羟基化。示例性蛋白质包括10-11易位(TET)1-3家族,将DNA中的5-甲基胞嘧啶(5-mC)转换为5-羟甲基胞嘧啶(5-hmC)的酶。

[0076] 人TET1-3的序列是本领域已知的,并在下表中显示:

[0077]

| | GenBank 登录号 | |
|---|---|---|
| 基因 | 氨基酸 | 核酸 |
| TET1 | NP_085128.2 | NM_030625.2 |
| TET2* | NP_001120680.1 (变体 1)<br>NP_060098.3 (变体 2) | NM_001127208.2<br>NM_017628.4 |
| TET3 | NP_659430.1 | NM_144993.1 |

[0078] *变体(1)代表较长的转录本,并编码较长的同工型(a)。与变体1相比,变体(2)在5'UTR和3'UTR和编码序列上不同。与同工型a相比,所得到的同工型(b)较短,并且具有独特

的C末端。

[0079] 在一些实施方案中,可以包括催化结构域的全部或部分全长序列,例如,包含富含半胱氨酸的延伸和由7个高度保守的外显子编码的2OGFeDO结构域的催化模块,例如,包含氨基酸1580-2052的Tet1催化结构域,包含氨基酸1290-1905的Tet2,和包含氨基酸966-1678的Tet3。参见,例如,Iyer等人,Cell Cycle.2009年6月1日;8(11):1698-710.Epub 2009年6月27日的图1,进行比对以说明所有三种Tet蛋白中的关键催化残基,及其全长序列的补充材料(参见,例如,seq 2c);在一些实施方案中,该序列包括Tet1的氨基酸1418-2136或Tet2/3中的相应区域。

[0080] 其他催化模块可以来自Iyer等人,2009年鉴定的蛋白质。

[0081] 在一些实施方案中,异源功能结构域是生物系链,并且包含MS2外壳蛋白、核糖核酸内切酶Csy4或λN蛋白的全部或部分(例如,来自其的DNA结合结构域)。这些蛋白质可用于将含有特定茎环结构的RNA分子募集到dCas9gRNA靶向序列指定的位置。例如,与MS2外壳蛋白、核糖核酸内切酶Csy4或λN融合的dCas9变体可用于募集诸如XIST或HOTAIR等长的非编码RNA(lncRNA);参见,例如,Keryer-Bibens等人,Biol.Cell 100:125-138(2008),所述长的非编码RNA与Csy4、MS2或λN结合序列相连。或者,可以将Csy4、MS2或λN蛋白结合序列连接至另一种蛋白,例如,如上文Keryer-Bibens等人所述,并且可以使用本文所述方法和组合物将该蛋白靶向dCas9变体结合位点。在一些实施方案中,Csy4是无催化活性的。在一些实施方案中,如WO 2014/204578中所述,将Cas9变体,优选dCas9变体与FokI融合。

[0082] 在一些实施方案中,融合蛋白包括dCas9变体和异源功能结构域之间的接头。可以在这些融合蛋白(或在串联结构中的融合蛋白之间)中使用的接头可以包括不干扰融合蛋白功能的任何序列。在优选的实施方案中,接头是短的,例如2-20个氨基酸,并且通常是柔性的(即,包含具有高自由度的氨基酸,例如甘氨酸、丙氨酸和丝氨酸)。在一些实施方案中,接头包含由GGGS(SEQ ID NO:2)或GGGGS(SEQ ID NO:3)组成的一个或多个单元,例如GGGS(SEQ ID NO:2)或GGGGS(SEQ ID NO:3)单元的两个、三个、四个或更多个重复。也可以使用其他接头序列。

[0083] 递送和表达系统

[0084] 为了使用本文所述的Cas9变体,可期望从编码它们的核酸表达它们。这可以以多种方式进行。例如,可以将编码Cas9变体的核酸克隆至中间载体,以转化原核或真核细胞,用于复制和/或表达。中间载体通常是原核生物载体,例如质粒、或穿梭载体、或昆虫载体,用于贮存或操作编码Cas9变体的核酸以产生Cas9变体。也可以将编码Cas9变体的核酸克隆至表达载体,以施用于植物细胞、动物细胞,优选哺乳动物细胞或人细胞、真菌细胞、细菌细胞或原生动物细胞。

[0085] 为了获得表达,通常将编码Cas9变体的序列亚克隆至含有启动子的表达载体以指导转录。合适的细菌和真核启动子是本领域众所周知的,并描述于例如Sambrook等人,Molecular Cloning,A Laboratory Manual(第3版2001);Kriegler,Gene Transfer and Expression:A Laboratory Manual(1990);和Current Protocols in Molecular Biology(Ausubel等人,编辑,2010)。用于表达工程化蛋白的细菌表达系统可在例如大肠杆菌(E.coli)、芽孢杆菌属物种(Bacillus sp.)和沙门氏菌(Salmonella)中获得(Palva等人,1983,Gene 22:229-235)。用于这种表达系统的试剂盒是商购可得的。用于哺乳动物细胞、

酵母和昆虫细胞的真核表达系统是本领域众所周知的,并且也是商购可得的。

[0086] 用于指导核酸表达的启动子取决于特定的应用。例如,强组成型启动子通常用于融合蛋白的表达和纯化。相反,当体内施用Cas9变体以进行基因调节时,取决于Cas9变体的特定用途,可以使用组成型启动子和诱导型启动子两者任一个。此外,用于施用Cas9变体的优选启动子可以是弱启动子,例如HSV TK或具有相似活性的启动子。启动子还可以包括响应反式激活的元件,例如缺氧反应元件、Gal4反应元件、lac阻遏物反应元件、以及小分子控制系统,例如四环素调节系统和RU-486系统(参见,例如,Gossen&Bujard,1992,Proc.Natl.Acad.Sci.USA,89:5547;Oligino等人,1998,Gene Ther.,5:491-496;Wang等人,1997,Gene Ther.,4:432-441;Neering等人,1996,Blood,88:1147-55;和Rendahl等人,1998,Nat.Biotechnol.,16:757-761)。

[0087] 除了启动子外,表达载体通常含有转录单元或表达盒,其含有在原核或真核的宿主细胞中表达核酸所需的所有其他元件。因此,典型的表达盒包含与例如编码Cas9变体的核酸序列操作地连接的启动子,以及例如转录物的有效聚腺苷酸化、转录终止、核糖体结合位点、或翻译终止所需的任何信号。盒的其他元件可包括例如增强子,和异源剪接的内含子信号。

[0088] 根据Cas9变体的预期用途,例如在植物、动物、细菌、真菌、原生动物等中的表达,选择用于将遗传信息传递到细胞中的特定表达载体。标准细菌表达载体包括质粒,例如基于pBR322的质粒、pSKF、pET23D和商购可得的标签融合表达系统,例如GST和LacZ。

[0089] 含有来自真核病毒的调节元件的表达载体经常用于真核表达载体,例如SV40载体、乳头瘤病毒载体、和源自爱泼斯坦-巴尔(Epstein-Barr)病毒的载体。其他示例性的真核载体包括pMSG、pAVO09/A+、pMTO10/A+、pMAMneo-5、杆状病毒pDSVE以及允许在SV40早期启动子、SV40晚期启动子、金属硫蛋白启动子、鼠乳腺肿瘤病毒启动子、劳斯肉瘤病毒启动子、多角体蛋白启动子或其他在真核细胞中显示有效表达的启动子的指导下表达蛋白质的任何其他载体。

[0090] 用于表达Cas9变体的载体可以包括驱动指导RNA表达的RNA Pol III启动子,例如H1、U6或7SK启动子。这些人启动子允许质粒转染后在哺乳动物细胞中表达Cas9变体。

[0091] 一些表达系统具有用于选择稳定转染的细胞系的标志物,例如胸苷激酶、潮霉素B磷酸转移酶和二氢叶酸还原酶。高产量表达系统也是合适的,例如在昆虫细胞中使用杆状病毒载体,具有在多角体蛋白启动子或其他强杆状病毒启动子的指导下的gRNA编码序列。

[0092] 表达载体中通常包含的元件还包括在大肠杆菌中起作用的复制子,编码抗生素抗性以允许选择具有重组质粒的细菌的基因,以及在质粒非必需区中的独特限制位点以允许插入重组序列。

[0093] 标准转染方法用于产生表达大量蛋白质的细菌、哺乳动物、酵母或昆虫细胞系,然后使用标准技术对其进行纯化(参见,例如,Colley等人,1989,J.Biol.Chem.,264:17619-22;Guide to Protein Purification,in Methods in Enzymology,第182卷(Deutscher,编辑,1990))。根据标准技术进行真核细胞和原核细胞的转化(参见,例如,Morrison,1977,J.Bacteriol.132:349-351;Clark-Curtiss&Curtiss,Methods in Enzymology 101:347-362(Wu等人,编辑,1983)。

[0094] 可以使用任何将外来核苷酸序列引入宿主细胞的已知方法。这些方法包括使用磷

酸钙转染、聚凝胺、原生质体融合、电穿孔、核转染、脂质体、显微注射、裸DNA、质粒载体、病毒载体(游离型和整合型两者)以及引入克隆的基因组DNA、cDNA、合成DNA或其他外来遗传物质进入宿主细胞的任何其他众所周知的方法(参见,例如,Sambrook等人,同上)。仅需要所使用的特定基因工程程序能够将至少一种基因成功地导入能够表达Cas9变体的宿主细胞中。

[0095] 或者,该方法可包括将Cas9变体蛋白和指导RNA一起例如作为复合物递送。例如,可以在宿主细胞中过表达Cas9变体和gRNA并纯化,然后与指导RNA复合(例如在试管中)以形成核糖核蛋白(RNP),并递送至细胞。在一些实施方案中,可以通过使用细菌Cas9表达质粒在细菌中表达并从中纯化变体Cas9。例如,具有His标签的变体Cas9蛋白可以在细菌细胞中表达,然后使用镍亲和色谱纯化。RNP的使用避免了递送编码核酸酶或指导或编码核酸酶作为mRNA的质粒DNA的必要性。RNP的递送也可改进特异性,大概是因为RNP的半衰期较短,并且核酸酶和指导基因也没有持续表达(就像从质粒中获得的那样)。可以例如使用脂质介导的转染或电穿孔在体内或体外将RNP递送至细胞。参见,例如,Liang等人"Rapid and highly efficient mammalian cell engineering via Cas9 protein transfection." Journal of biotechnology 208(2015):44-53;Zuris,John A.,等人"Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing in vitro and in vivo."Nature biotechnology 33.1(2015):73-80;Kim等人"Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins."Genome research 24.6(2014):1012-1019。

[0096] 本发明包括载体和包含所述载体的细胞。

[0097] 实施例

[0098] 在以下实施例中进一步描述本发明,这些实施例不限制权利要求中描述的本发明的范围。

[0099] 方法

[0100] 实施例1中使用了以下材料和方法。

[0101] 质粒和寡核苷酸

[0102] 表4中列出用于扩增T7E1测定的内源性人类基因靶位点的寡核苷酸序列。

[0103]

| 表4 | | |
|---|---|---|
| **用于T7E1实验的引物** | | **SEQ ID NO:** |
| 序列 | 描述 | |
| GGAGCAGCTGGTCAGAGGGG | 靶向U2OS人细胞中*EMX1*的正向引物 | 4 |
| CCATAGGGAAGGGGGACACTGG | 靶向U2OS人细胞中*EMX1*的反向引物 | 5 |
| GGGCCGGGAAAGAGTTGCTG | 靶向U2OS人细胞中*FANCF*的正向引物 | 6 |
| GCCCTACATCTGCTCTCCCTCC | 靶向U2OS人细胞中*FANCF*的反向引物 | 7 |
| CCAGCACAACTTACTCGCACTTGAC | 靶向U2OS人细胞中*RUNX1*的正向引物 | 8 |
| CATCACCAACCCACAGCCAAGG | 靶向U2OS人细胞中*RUNX1*的反向引物 | 9 |
| GATGAGGGCTCCAGATGGCAC | 靶向U2OS人细胞中*VEGFA*的正向引物 | 10 |
| GAGGAGGGAGCAGGAAAGTGAGG | 靶向U2OS人细胞中*VEGFA*的反向引物 | 11 |

[0104] 如先前所述(Kleinstiver等人,Nature 2015)构建细菌Cas9/sgRNA表达质粒,其具有两个T7启动子以分别表达Cas9和sgRNA。通过将编码在位置D1135、S1136、G1218、E1219、R1335、和T1337处的随机密码子的寡核苷酸克隆至亲本SpCas9细菌表达载体,以生成在这些位置含有可变氨基酸的细菌表达质粒(图1)。

[0105] 为了在人细胞中表达,通过等温组装至pCMV-T7-hSpCas9-NLS-3xFLAG载体(JDS246;可在addgene.org/43861/sequences/中找到序列),在SpCas9中产生点突变。

[0106] 通过将适当的退火寡核苷酸克隆至BsmBI消化的BPK1520,生成用于sgRNA U6表达的质粒(可以在其中克隆所需的间隔子寡核苷酸)。

[0107] 进化SpCas9变体的基于细菌的阳性选择测定

[0108] 用编码Cas9/sgRNA的质粒转化含有阳性选择质粒(具有嵌入的靶位点)的感受态大肠杆菌BW25141(λDE3)[23]。在SOB培养基中恢复60分钟后,将转化物铺板在含有氯霉素(非选择性)或氯霉素+10mM阿拉伯糖(选择性)的LB板上。

[0109] 为了选择可以切割新PAM的SpCas9变体,将编码随机D1135X/S1136X/G1218X/E1219X/R1335X/T1337X SpCas9文库的质粒电穿孔至大肠杆菌BW25141(λDE3)细胞,该细胞已经具有编码具有感兴趣的PAM的靶位点的阳性选择质粒。存活的菌落生长过夜,进行微量制备以提取SpCas9-XXXXXX表达质粒,然后单独重新转化至含有先前描述的PAM序列的阳性选择的大肠杆菌BW25141(λDE3)细胞,以重新测定存活表型与那些质粒的连接,从而消除假阳性菌落。通常在后续实验中重新筛选~300个菌落。对第二次筛选中真正存活的菌落的SpCas9表达质粒进行测序,以鉴定导致特异性改变的D1135、S1136、G1218、E1219、R1335、和/或T1337位置处的氨基酸(参见表1-3)。根据其在存活的菌落中的频率和(在某些情况下)基于人细胞的EGFP破坏测定中的活性,选择在测序的菌落中观察到的突变进行进一步评估。

[0110] 人细胞培养和转染

[0111] 将具有EGFP-PEST报告基因的单个整合拷贝的U2OS细胞和U2OS.EGFP细胞(Reyon,D.等人FLASH assembly of TALENs for high-throughput genome editing.Nat Biotechnol 30,460-465(2012))在含有10%FBS、青霉素/链霉素、和2mM GlutaMAX(Life Technologies)的高级DMEM培养基(Life Technologies)中,在37℃和5%$CO_2$下培养。通过STR分析(ATCC)和深度测序验证细胞系的特性,并且每两周对细胞进行一次支原体污染检测。U2OS.EGFP培养基还补充有400μg/mL G418。按照制造商的说明,使用Lonza 4D-nucleofector的DN-100程序,将细胞与750ng Cas9质粒和250ng sgRNA质粒共转染。与空U6启动子质粒一起转染的Cas9质粒用作所有人细胞实验的阴性对照。

[0112] 人细胞EGFP破坏测定

[0113] 如先前所述进行EGFP破坏实验(Fu,Y.等人High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells.Nat Biotechnol31,822-826(2013);Reyon,D.等人FLASH assembly of TALENs for high-throughput genome editing.Nat Biotechnol 30,460-465(2012))。转染后约52小时,使用Fortessa流式细胞仪(BD Biosciences)测定转染的U2OS.EGFP细胞中的EGFP荧光。Cas9和空U6启动子质粒的阴性对照转染用于确定所有实验中背景EGFP的损失为~2.5%。

[0114] T7E1测定

[0115] 如先前所述进行T7E1测定[15]。对于U2OS人细胞,使用Agencourt DNAdvance基因组DNA分离试剂盒(Beckman Coulter Genomics)从转染后~72小时的转染的细胞中提取基因组DNA。使用表4中列出的引物扩增人细胞基因组DNA的靶基因座。将约200ng纯化的PCR产物变性、退火并用T7E1(New England BioLabs)消化。如前所述[15],使用Qiaxcel毛细管电泳仪(QIagen)量化诱变频率。

[0116] 实施例1

[0117] 为了通过使得能够靶向当前无法接近的PAM序列来进一步扩展SpCas9的实用性,我们构想了一种替代策略,以选择能够识别新PAM序列的SpCas9变体。先前已经确定SpCas9编码序列中的某些位置对于PAM识别是重要的(Kleinstiver等人,Nature 2015),我们进行了集中诱变方法,其中我们在六个位置随机化氨基酸以生成在PAM相互作用结构域的三个单独区域内具有不同密码子用法的SpCas9变体的文库:D1135/S1136、G1218/E1219和R1335/T1337。为此,我们在含有编码这六个氨基酸的SpCas9密码子上依次克隆了编码NNS核苷酸三联体(其中N是任意核苷酸,并且S是G或C)的随机寡核苷酸盒(图1A)。然后,如前所述(Kleinstiver等人,Nature 2015)在细菌阳性选择测定中针对编码各种NGNN PAM序列的靶位点筛选出得到的SpCas9变体的文库(图1B)。简而言之,如果SpCas9可以识别阳性选择质粒中编码的靶位点(PAM和间隔子序列),则细菌只能在选择性条件下存活(铺板在10mM阿拉伯糖上,其诱导ccdB毒性基因的转录)。强PAM识别将导致选择质粒水解,阻止ccdB表达的诱导并允许细菌生长。因此,在筛选SpCas9文库时,在含有10mM阿拉伯糖的培养基上生长的菌落预期编码可以靶向其他PAM的SpCas9 PAM变体(图1B)。

[0118] 我们首先在阳性选择质粒上筛选随机的D1135X/S1136X/G1218X/E1219X/R1335X/T1337X SpCas9文库(SpCas9-XXXXXX,其中X是任意氨基酸),该质粒编码具有NGTG、NGTT、NGTC和NGTA PAM的靶位点。对于每个不同的PAM选择,从阿拉伯糖选择中挑取48个存活的菌落,并使其在含氯霉素的培养基中生长过夜以回收核酸酶编码质粒。为了减少初次筛选中的假阳性率,随后将所有推定的PAM变体质粒与编码针对其最初进行筛选的靶位点和PAM的阳性选择质粒一起重新筛选(数据未显示)。对在该重新筛选测定中具有至少50%存活率的真正变体子集进行测序,以鉴定在残基1135、1136、1218、1219、1335、1337处的氨基酸(表I),然后将这些变体针对NGTG、NGTT、NGTC、和NGTA PAM进行更广泛的筛选,以评估针对NGTN位点的活性(图2a-b和表I)。注意:此后,在细菌测定中通过其变体编号(vNGTN-#)或在人测定中通过其"氨基酸名称"来描述SpCas9变体,其中氨基酸名称的形式为SpCas9-XXXXXX,其中六个X代表在位置1135、1136、1218、1219、1335、和1337处的氨基酸同一性(见表I-3)。重新筛选确定了一些趋势,其中在一些情况下,变体在针对其最初选择的NGTN PAM上具有最高的活性(例如vNGTN-1、-3、-12、-27、-28等),一些变体可以靶向NGTN PAM的组合(例如,可以靶向NGTC和NGTA的vNGTN-15、-31、-35等),一些变体可以靶向所有NGTN PAM(例如vNGTN-9、-10,-30等)(图2A-B和表I)。基于这些结果,基于在初始筛选中表现良好的菌落中的位置1135、1136、1218、1219、1335和1337处频繁出现的氨基酸,合理设计新的变体。这些合理设计的NGTN变体在细菌筛选中针对NGTG、NGTT、NGTC、和NGTA PAM进行评估(图2C-F),在某些情况下还针对NGAN PAM(NGAG、NGAT、NGAC、NGAA;图2C,2E-F)进行筛选。再次确定许多感兴趣的变体,其性质与上述偏好一致,但值得注意的是,确定一些其他变体,它们可以赋予PAM第四位的偏好(例如,可以靶向NGTG或NGAG PAM的vNGTG-37或NGTC和NGAC PAM上的vNGTG-

18和-41等),可以靶向所有NGTN PAM的其他变体(例如vNGTN-40、-46、-48等),以及可以靶向所有或几乎所有NGTN或NGAN PAM的变体(例如vNGTN-7、-44、-59等;图2C-F和表I)。

[0119] 在细菌中已经确定了几种可以靶向NGTN PAM位点的变体后,我们试图确定这些偏好是否会转化为人细胞中的真正活性。在人U2OS EGFP破坏测定中,初步筛选了十二种不同的NGTN PAM变体,我们确定可以稳健地靶向NGTT和NGTG PAM的变体(例如SpCas9-GRKIQK、-VAKLLR、-VRKLLR等),其中一些可以修饰所有NGTN PAM位点(例如SpCas9-LRSVQL、-IRAVQL等)(图3A)。在人细胞EGFP破坏测定中进一步筛选这些变体的子集和其他经过合理设计的变体,确定SpCas9-LRSVQL、-LRKIQK、-LRSVQK、和其他为能靶向NGTN PAM序列的有前途的变体(图3B)。为了更严格地表征SpCas9-LRSVQL对人细胞中NGTN PAM序列的活性,我们检验了该核酸酶变体在人U2OS细胞中跨EMX1、FANCF、和RUNX1基因的跨32个不同内源性位点上的活性。这项分析揭示了SpCas9-LRSVQL在具有NGTG、NGTA、NGTC和NGTT PAM的各种内源位点上的稳健的活性(图3C,表明我们选择和合理设计的PAM变体可以在以前无法用已公开的SpCas9变体靶向的众多基因座上有效运行。

[0120] 我们之前已经描述了可以有效靶向NGCG PAM位点的SpCas9变体(Kleinstiver等人,Nature,2015),称为SpCas9-VRER(编码D1135V/G1218R/R1335E/T1337R替代)。尽管此变体可以定位以前无法接近的位点,但仅限于具有扩展NGCG PAM的位点上的活性。为了通过潜在地将所有NGCN PAM靶向到现在包括NGCT、NGCC、和NGCA来扩展SpCas9 PAM变体的用途,我们进行了与上述类似的选择,但是针对具有靶位点的阳性选择质粒筛选SpCas9-XXXXXX文库,所述靶位点具有NGCG、NGCT、NGCC、或NGCA PAM任一个(图4A-B和表2)。就像我们在NGTN选择中观察到的一样,重新筛选NGCN变体确定了如下情况:其中变体在针对最初选择其的NGCN PAM上具有最高的活性(例如,vNGCN-3、-8、-9、-17等),一些变体可以针对NGCN PAM的组合(例如vNGCN-10等,可以针对NGCT、NGCC、和NGCA),一些变体可以针对所有NGCN PAM(例如vNGCN-1、-2、-5、-18、-26等)(图4A-B和表2)。基于对SpCas9-XXXXXX选定克隆中氨基酸富集的观察结果,克隆各种合理生成的NGCN变体,并在细菌中检测针对NGCN PAM的活性(图4A-B,表2,数据未显示)。

[0121] 接下来,我们在我们的U2OS EGFP破坏测定中检测各种NGCN选择的SpCas9 PAM变体的活性,以确定在人细胞中是否可以重新概括其重新靶向的PAM偏好和核酸酶活性(图5A-D)。我们观察到针对NGCA PAM(SpCas9-MQKSER、-LWRVVA、-LWLETR等;图5A)、NGCC PAM(SpCas9-MQKSER、-LSRWQR、-ICCCER等;图5A)、NGCT PAM(SpCas9-MQKSER;图5C)、或NGCC和NGCT PAM(SpCas9-MQKSER、-VRKSER、-ICKSER等;图5C)的许多变体的活性。针对15个总NGCA、NGCC、NGCT、和NGCG位点对SpCas9-MQKSER、-VRKSER、-ICKSER变体进行的进一步检测显示,每种变体针对所有类别的NGCN PAM都有稳健的活性(图5E)。在某些情况下,这些变体可以优于已公开的SpCas9-VRER(例如,如图5B-C所示),尽管这通常是针对先前已显示被SpCas9-VRER无效靶向的PAM。总的来说,这些新变体将SpCas9的靶向扩展到NGCT、NGCC和NGCA,而不是以前接近的NGCG,其中SpCas9-MQKSER和其他变体对所有NGCN PAM都具有稳健的活性。

[0122] 此外,我们之前也已经描述了可以有效靶向NGAN PAM位点的SpCas9变体,称为SpCas9-VQR(D1135V/R1335Q/T1337R;Kleinstiver等人,Nature,2015)和SpCas9-VRQR(D1135V/G1218R/R1335Q/T1337R;Kleinstiver和Pattanayak等人,Nature,2016)。然而,这

些变体按NGAG>NGAA＝NGAT>NGAC的顺序优先选择NGAN PAM的亚类,即它们针对NGAC PAM位点的活性次佳。为了潜在地改进对NGAN PAM的SpCas9靶向,我们如上所述用SpCas9-XXXXXX文库对编码NGAG、NGAT、NGAC、和NGAA PAM的阳性选择质粒进行选择(图6A-B和表3)。对NGAN变体的重新筛选揭示在针对其最初选择的NGAN PAM上具有最高活性的菌落(例如NGAG上的vNGAN-1、-2、-17、-26至-30等、NGAT上的vNGAN-32、NGAC上的vNGAN-4、-5、-40、-41等,等),一些变体可以靶向NGAN PAM的组合(例如vNGAN-20、-21等,其可以靶向NGAT和NGAC,或可以靶向NGAG和NGAC的vNGAN-22),一些变体可以靶向所有NGAN PAM(例如vNGAN-3、-13、-25、-31等)(图6A-B和表3)。

[0123] 由于许多SpCas9-XXXXXX变体在细菌筛选中显示出强NGAC PAM靶向性,因此将许多变体亚克隆至我们的人表达载体,以检查我们的人细胞U2OS EGFP破坏测定中的活性。针对EGFP中单个NGAA、NGAC、NGAT、和NGAG PAM位点的变体子集的初始筛选揭示,某些变体可以潜在地在具有NGAC PAM的位点上表现优于SpCas9-VQR(图7A)。对来自图7A的变体和其他选定变体的更广泛的检测揭示,相对于SpCas9-VRQR,多个SpCas9变体在EGFP破坏测定(图7B)中检测到的一些或全部四个NGAC PAM位点(包括SpCas9-LRSVRS)上具有增加的活性,包括SpCas9-LRSVRS、-MRARKE、-SRQMRG、和其他。

[0124] 然后,我们将我们的SpCas9变体活性与最近描述的名为xCas9的SpCas9PAM变体的活性进行比较,所述xCas9具有报道的放宽的NG PAM偏好(Hu等人,Nature第556卷,第57-63页(2018年4月5日))。与我们之前的结果一致,我们观察到具VRQR变体的具有NGA PAM的位点(也称为VSREQR)、具VRER(也称为VSREER)和MQKSER变体的具有NGCG PAM的位点、以及具LRSVQL变体的具有NGT PAM的位点(图8)。然而,使用xCas9变体时,未观察到具有NGA、NGCG、或NGT PAM的位点靶向的效率高于10％;此外,我们观察到,与野生型SpCas9相比,xCas9对具有NGG PAM的靶向位点的效力平均低约2倍(图8)。这些结果表明,与xCas9相比,我们的SpCas9 PAM变体是针对多种PAM的更有效的核酸酶。

[0125] 实施例2

[0126] 最近已经使用工程化的SpCas9碱基编辑器(BE)构建体证明了进行精确的单碱基编辑事件的能力(参见,例如,Komor等人,Nature.2016年5月19日;533(7603):420-4;Nishida等人,Science.2016年9月16日;353(6305);Kim等人,Nat Biotechnol.2017年4月;35(4):371-376;Komor等人,Sci Adv.2017年8月30日;3(8):eaao4774;和Gaudelli等人,Nature.2017年11月23日;551(7681):464-471),其利用SpCas9-gRNA形成的R环形成,导致非靶DNA链的ssDNA可接近性。因此,异源胞苷或腺嘌呤脱氨酶的酶促结构域与SpCas9的融合可作用于暴露的ssDNA链,从而分别有效地引入C至T变化(所谓的胞嘧啶碱基编辑器,或CBE),或A至G(所谓的腺苷碱基编辑器,或ABE)。因为细胞碱基切除修复(BER)使用尿嘧啶DNA糖基化酶(UDG;也称为尿嘧啶N-糖基化酶或UNG)切除尿嘧啶碱基,所以这种内源性过程可以有效地逆转胞苷BE产生的编辑,因为胞嘧啶核苷的脱氨作用导致尿嘧啶中间体。因此,为了提高胞嘧啶BE的效率,还可以将异源效应子结构域如尿嘧啶糖基化酶抑制剂(UGI)与SpCas9融合以抑制UDG,从而颠覆BER的启动并提高胞嘧啶BE的有效性。

[0127] 因此,我们试图确定我们SpCas9 PAM变体的扩展的靶向范围是否可以通过使先前无法接近的位点能够编辑来提高碱基编辑器的实用性。为此,我们构建了BE3(Komor等人,Nature.2016年5月19日;533(7603):420-4)PAM变异体,以生成能够识别具有NGA和NGT PAM

的位点的CBE。我们发现,在具有NGA PAM的位点上,CBE-VRQR变体显示在编辑窗口中的C至T的转化率介于7.5%至64.2%之间,而xCas9显示在相同的位点上0%-19.9%的C-至-T编辑(图9A)。类似地,在具有NGT PAM的位点上,CBE-LRSVQL变体显示C至T的转化率介于10.8%至50.3%之间,而CBE-xCas9显示在相同位点上0%-28.5%的C-至-T编辑(图9A)。我们还观察到,与野生型SpCas9(52.5%-62.4%)相比,CBE-xCas9在具有NGG PAM的位点上的C-T编辑活性明显降低(26.7%-37.2%)(图9A)。这些结果表明,VRQR和LRSVQL的BE3版本在具有NGA和NGT PAM的位点上分别是有效的CBE,其速率高于xCas9～2倍。

[0128] 接下来,我们构建了我们的PAM变体的ABE(7.10)(Gaudelli等人,Nature.2017年11月23日;551(7681):464-471)版本,以确定其作为介导人细胞中A-至-G转化的ABE的有效性。与观察到的ABE-xCas9的0%-12.5%编辑相比,我们在具有NGA PAM的位点上观察到ABE-VRQR(8.0%-77.3%)对A-至-G的强编辑活性(图9B)。同样,在具有NGCG PAM的位点上,对于A-至-G编辑,ABE-VRER(0%-75.9%)和ABE-MSQKER(5.4%-90.4%)变体再次优于ABE-xCas9(0%-62.3%)(图9B)。我们还观察到,与野生型SpCas9(13.9%-50.4%)相比,在具有NGG PAM的位点上,使用ABE-xCas9(0%-16.9%)的A-至-G编辑减少(图9B)。我们的结果表明,VRQR的ABE(7.10)版本在具有NGA PAM的位点上可以有效地介导A-至-G编辑,而VRER和MSQKER的ABE(7.10)版本在具有NGCG PAM的位点上有效。

[0129] 参考文献

[0130] 1.Sander,J.D.&Joung,J.K.CRISPR-Cas systems for editing,regulating and targeting genomes.Nat Biotechnol 32,347-355(2014).

[0131] 2.Hsu,P.D.,Lander,E.S.&Zhang,F.Development and applications of CRISPR-Cas9 for genome engineering.Cell 157,1262-1278(2014).

[0132] 3.Doudna,J.A.&Charpentier,E.Genome editing.The new frontier of genome engineering with CRISPR-Cas9.Science 346,1258096(2014).

[0133] 4.Barrangou,R.&May,A.P.Unraveling the potential of CRISPR-Cas9 for gene therapy.Expert Opin Biol Ther 15,311-314(2015).

[0134] 5.Jinek,M.et al.A programmabledual-RNA-guided DNA endonuclease in adaptive bacterial immunity.Science 337,816-821(2012).

[0135] 6.Sternberg,S.H.,Redding,S.,Jinek,M.,Greene,E.C.&Doudna,J.A.DNA interrogation by the CRISPR RNA-guided endonuclease Cas9.Nature 507,62-67(2014).

[0136] 7.Hsu,P.D.et al.DNA targeting specificity of RNA-guided Cas9 nucleases.Nat Biotechnol 31,827-832(2013).

[0137] 8.Tsai,S.Q.et al.GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases.Nat Biotechnol 33,187-197(2015).

[0138] 9.Hou,Z.et al.Efficient genome engineering in human pluripotent stem cells using Cas9 from Neisseria meningitidis.Proc Natl Acad Sci U S A(2013).

[0139] 10.Fonfara,I.et al.Phylogeny of Cas9 determines functional exchangeability of dual-RNA and Cas9 among orthologous type II CRISPR-Cas systems.Nucleic Acids Res 42,2577-2590(2014).

[0140] 11.Esvelt,K.M.et al.Orthogonal Cas9 proteins for RNA-guided gene regulation and editing.Nat Methods 10,1116-1121(2013).

[0141] 12.Cong,L.et al.Multiplex genome engineering using CRISPR/Cas systems.Science 339,819-823(2013).

[0142] 13.Horvath,P.et al.Diversity,activity,and evolution of CRISPR loci in Streptococcus thermophilus.J Bacteriol 190,1401-1412(2008).

[0143] 14.Anders,C.,Niewoehner,O.,Duerst,A.&Jinek,M.Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease.Nature 513,569-573(2014).

[0144] 15.Reyon,D.et al.FLASH assembly of TALENs for high-throughput genome editing.Nat Biotechnol 30,460-465(2012).

[0145] 16.Fu,Y.et al.High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells.Nat Biotechnol 31,822-826(2013).

[0146] 17.Chen,Z.&Zhao,H.A highly sensitive selection method for directed evolution of homing endonucleases.Nucleic Acids Res 33,e154(2005).

[0147] 18.Doyon,J.B.,Pattanayak,V.,Meyer,C.B.&Liu,D.R.Directed evolution and substrate specificity profile of homing endonuclease I-SceI.J Am Chem Soc 128,2477-2484(2006).

[0148] 19.Jiang,W.,Bikard,D.,Cox,D.,Zhang,F.&Marraffini,L.A.RNA-guided editing of bacterial genomes using CRISPR-Cas systems.Nat Biotechnol 31,233-239(2013).

[0149] 20.Mali,P.et al.RNA-guided human genome engineering via Cas9.Science 339,823-826(2013).

[0150] 21.Hwang,W.Y.et al.Efficient genome editing in zebrafish using a CRISPR-Cas system.Nat Biotechnol 31,227-229(2013).

[0151] 22.Chylinski,K.,Le Rhun,A.&Charpentier,E.The tracrRNA and Cas9 families of type IICRISPR-Cas immunity systems.RNABiol 10,726-737(2013).

[0152] 23.Kleinstiver,B.P.,Fernandes,A.D.,Gloor,G.B.&Edgell,D.R.Aunified genetic,computational and experimental framework identifies functionally relevant residues of the homing endonuclease I-BmoI.Nucleic Acids Res 38,2411-2427(2010).

[0153] 24.Gagnon,J.A.et al.Efficient mutagenesis by Cas9 protein-mediated oligonucleotide insertion and large-scale assessment of single-guide RNAs.PLoS One 9,e98186(2014).

[0154] 其他实施方式

[0155] 应理解,尽管已经结合本发明的详述描述了本发明,但是前述描述旨在说明而不是限制本发明的范围,本发明的范围由所附权利要求的范围限定。其他方面、优点和修改在所附权利要求的范围内。

[0156]

表1：针对NGTN PAM的变体在细菌中的选择结果和活性

| 样品 # | 针对(NGTN)的最初选择 | 在细菌测定中针对以下情况的近似存活率： | | | | 变体名称 | 变体中的氨基酸取代： | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | NGTG | NGTT | NGTC | NGTA | | D1135 | S1136 | G1218 | E1219 | R1335 | T1337 |
| vNGTN-1 | G | 30% | 0% | 3% | 3% | SpCas9-DKVHVR | D | K | V | H | V | R |
| vNGTN-2 | G | 15% | 40% | 40% | 40% | SpCas9-NRMMRT | N | R | M | M | R | T |
| vNGTN-3 | G | 100% | 40% | 40% | 10% | SpCas9-ASQLMR | A | S | Q | L | M | R |
| vNGTN-4 | G | 100% | 40% | 10% | 40% | SpCas9-LRQYTR | L | R | Q | Y | T | R |
| vNGTN-5 | G | 100% | 15% | 10% | 20% | SpCas9-GCACMR | G | C | A | C | M | R |
| vNGTN-6 | G | 100% | 50% | 30% | 20% | SpCas9-SRSCMV | S | R | S | C | M | V |
| vNGTN-7 | T | 100% | n/a | n/a | n/a | SpCas9-LWKIQK | L | W | K | I | Q | K |
| vNGTN-8 | T | 50% | 50% | 20% | 30% | SpCas9-YSAFCC | Y | S | A | F | C | C |
| vNGTN-9 | T | 100% | 100% | 70% | 80% | SpCas9-IRAVQL | I | R | A | V | Q | L |
| vNGTN-10 | T | 100% | 80% | 70% | 60% | SpCas9-SWRVVV | S | W | R | V | V | V |
| vNGTN-11 | T | 100% | 80% | 70% | 60% | SpCas9-SWKVLK | S | W | K | V | L | K |
| vNGTN-12 | T | 3% | 80% | 3% | 3% | SpCas9-LWSVGG | L | W | S | V | G | G |
| vNGTN-13 | C | - | - | 10% | 20% | SpCas9-SRAAKW | S | R | A | A | K | W |
| vNGTN-14 | C | - | - | 10% | 10% | SpCas9-RNGWRI | R | N | G | W | R | I |
| vNGTN-15 | C | 0% | 3% | 90% | 90% | SpCas9-TAHFKV | T | A | H | F | K | V |
| vNGTN-16 | C | - | - | 80% | 80% | SpCas9-MSGVKC | M | S | G | V | K | C |
| vNGTN-17 | C | - | - | 50% | 50% | SpCas9-VMRCKL | V | M | R | C | K | L |
| vNGTN-18 | C | - | - | 75% | 75% | SpCas9-LRSVRS | L | R | S | V | R | S |
| vNGTN-19 | A | - | - | n/a | 30% | SpCas9-RWNLRR | R | W | N | L | R | R |
| vNGTN-20 | A | 3% | 3% | 0% | 0% | SpCas9-VRCVRC | V | R | C | V | R | C |
| vNGTN-21 | A | - | - | 20% | 20% | SpCas9-GRTSRC | G | R | T | S | R | C |
| vNGTN-22 | A | - | - | 65% | 65% | SpCas9-LKLCKR | L | K | L | C | K | R |
| vNGTN-23 | A | - | - | 70% | 65% | SpCas9-AKLCRT | A | K | L | C | R | T |
| vNGTN-24 | A | - | - | 75% | 100% | SpCas9-SKTLRP | S | K | T | L | R | P |
| vNGTN-25 | G | 50% | 20% | 40% | 40% | SpCas9-SRRSQR | S | R | R | S | Q | R |
| vNGTN-26 | G | 50% | 20% | 40% | 50% | SpCas9-DKVHVR | D | K | V | H | V | R |
| vNGTN-27 | G | 50% | - | - | - | SpCas9-GAKLLR | G | A | K | L | L | R |
| vNGTN-28 | T | 20% | 100% | 40% | 40% | SpCas9-MWAFGC | M | W | A | F | G | C |
| vNGTN-29 | T | - | 35% | - | - | SpCas9-GWRVTW | G | W | R | V | T | W |
| vNGTN-30 | T | 85% | 100% | 100% | 80% | SpCas9-MWVHLN | M | W | V | H | L | N |
| vNGTN-31 | C | - | - | 100% | 85% | SpCas9-TWSMRG | T | W | S | M | R | G |
| vNGTN-32 | C | - | - | 80% | 35% | SpCas9-ISGTKN | I | S | G | T | K | N |
| vNGTN-33 | C | - | - | 50% | 45% | SpCas9-SRAAKW | S | R | A | A | K | W |
| vNGTN-34 | A | 75% | 75% | 50% | 40% | SpCas9-KCAFCC | K | C | A | F | C | C |

[0157]

表 1：针对NGTN PAM的变体在细菌中的选择结果和活性

| 样品 # | 针对(NGTN)的最初选择 | 在细菌测定中针对以下情况的近似存活率： | | | | 变体名称 | 变体中的氨基酸取代： | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | NGTG | NGTT | NGTC | NGTA | | D1135 | S1136 | G1218 | E1219 | R1335 | T1337 |
| vNGTN-35 | A | - | - | 100% | 100% | SpCas9-KRRCKV | K | R | R | C | K | V |
| vNGTN-36 | A | - | - | 90% | 100% | SpCas9-MWGGRC | M | W | G | G | R | C |
| vNGTN-37 | 产生的变体 | 80% | - | 5% | 3% | SpCas9-VSKLLR | V | S | K | L | L | R |
| vNGTN-38 | 产生的变体 | 90% | 20% | 10% | 10% | SpCas9-VRKLLR | V | R | K | L | L | R |
| **vNGTN-27*** | G | 75% | - | - | - | SpCas9-GAKLLR | G | A | K | L | L | R |
| vNGTN-39 | 产生的变体 | 50% | 95% | 5% | 2% | SpCas9-VSAVQL | V | S | A | V | Q | L |
| vNGTN-40 | 产生的变体 | 90% | 95% | 95% | 50% | SpCas9-VRAVQL | V | R | A | V | Q | L |
| **vNGTN-9*** | T | 90% | 95% | 95% | 90% | SpCas9-IRAVQL | I | R | A | V | Q | L |
| vNGTN-41 | 产生的变体 | - | - | 95% | 90% | SpCas9-VSSVRS | V | S | S | V | R | S |
| vNGTN-42 | 产生的变体 | - | - | 95% | 90% | SpCas9-VRSVRS | V | R | S | V | R | S |
| **vNGTN-18*** | C | - | - | 100% | 95% | SpCas9-LRSVRS | L | R | S | V | R | S |
| vNGTN-43 | N/A | - | - | - | - | SpCas9-SRGERT | S | R | G | E | R | T |
| vNGTN-44 | N/A | 80% | 35% | 100% | 90% | SpCas9-SRMHCK | S | R | M | H | C | K |
| vNGTN-45 | 产生的变体 | | | | | SpCas9-GRKIQK | G | R | K | I | Q | K |
| vNGTN-46 | 产生的变体 | | | | | SpCas9-GWKLLR | G | W | K | L | L | R |
| vNGTN-47 | 产生的变体 | | | | | SpCas9-GWKOQK | G | W | K | Q | Q | K |
| vNGTN-48 | 产生的变体 | | | | | SpCas9-VAKLLR | V | A | K | L | L | R |
| vNGTN-49 | 产生的变体 | | | | | SpCas9-VAKIQK | V | A | K | I | Q | K |
| vNGTN-50 | 产生的变体 | | | | | SpCas9-VAKILR | V | A | K | I | L | R |
| vNGTN-51 | 产生的变体 | | | | | SpCas9-GRKILR | G | R | K | I | L | R |
| vNGTN-52 | 产生的变体 | - | - | 100% | 90% | SpCas9-VRKLRS | V | R | K | L | R | S |
| **vNGTN-38** | 产生的变体 | 100% | 85% | 60% | 50% | SpCas9-VRKLLR | V | R | K | L | L | R |
| vNGTN-53 | 产生的变体 | 100% | 100% | 100% | 100% | SpCas9-LRSVQL | L | R | S | V | Q | L |
| **vNGTN-18** | C | - | 1% | 100% | 100% | SpCas9-LRSVRS | L | R | S | V | R | S |
| vNGTN-54 | 产生的变体 | - | 5% | 100% | 100% | SpCas9-IRAVRS | I | R | A | V | R | S |
| vNGTN-55 | T | 100% | 100% | 95% | 95% | SpCas9-IRAVQL | I | R | A | V | Q | L |
| vNGTN-56 | 产生的变体 | - | - | 50% | 50% | SpCas9-VRKLKR | V | R | K | L | K | R |
| **vNGTN-38** | 产生的变体 | 100% | 50% | 25% | 25% | SpCas9-VRKLLR | V | R | K | L | L | R |
| vNGTN-57 | 产生的变体 | - | - | 100% | 100% | SpCas9-SRSVRS | S | R | S | V | R | S |
| **vNGTN-18** | C | - | - | 95% | 90% | SpCas9-LRSVRS | L | R | S | V | R | S |
| vNGTN-58 | 产生的变体 | 100% | 85% | 100% | 100% | SpCas9-VRKIQK | V | R | K | K | I | Q |
| **vNGTN-7*** | T | 100% | 85% | 100% | 100% | SpCas9-LWKIQK | L | W | K | K | I | Q |
| vNGTN-59 | 产生的变体 | 100% | 85% | 100% | 100% | SpCas9-VRMHCK | V | R | M | H | C | K |
| **vNGTN-44*** | N/A | 100% | 60% | 100% | 95% | SpCas9-SRMHCK | S | R | M | H | C | K |
| vNGTN-61 | 产生的变体 | n/a | n/a | n/a | n/a | SpCas9-GRKLLR | G | R | K | L | L | R |
| vNGTN-62 | 产生的变体 | n/a | n/a | n/a | n/a | SpCas9-LRKIQK | L | R | K | I | Q | K |

[0158]

表1：针对NGTN PAM的变体在细菌中的选择结果和活性

| 样品 # | 针对(NGTN)的最初选择 | 在细菌测定中针对以下情况的近似存活率： | | | | 变体名称 | 变体中的氨基酸取代： | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | NGTG | NGTT | NGTC | NGTA | | D1135 | S1136 | G1218 | E1219 | R1335 | T1337 |
| vNGTN-63 | 产生的变体 | n/a | n/a | n/a | n/a | SpCas9-LRSVQK | L | R | S | V | Q | K |
| vNGTN-64 | 产生的变体 | n/a | n/a | n/a | n/a | SpCas9-VRKIQK | V | R | K | I | Q | K |
| vNGTN-65 | 产生的变体 | n/a | n/a | n/a | n/a | SpCas9-GRSVQL | G | R | S | V | Q | L |
| vNGTN-66 | 产生的变体 | n/a | n/a | n/a | n/a | SpCas9-GRKIQL | G | R | K | I | Q | L |

* = 该变体已经在其他实验中进行过筛选

n/a = 在该PAM的实验中未评估存活率

表2：针对NGCN PAM的变体在细菌中的选择结果和活性

| 样品 # | 针对(NGCN)的最初选择 | 在细菌测定中针对以下情况的近似存活率： | | | | 变体名称 | 变体中的氨基酸取代： | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | NGCG | NGCT | NGCC | NGCA | | D1135 | S1136 | G1218 | E1219 | R1335 | T1337 |
| vNGCN-1 | G | 100% | 100% | 100% | 100% | SpCas9-WMQAYG | W | M | Q | A | Y | G |
| vNGCN-2 | G | 100% | 100% | 100% | n/a | SpCas9-MQKSER | M | Q | K | S | E | R |
| vNGCN-3 | G | 100% | - | - | 40% | SpCas9-YSVCER | Y | S | V | C | E | R |
| vNGCN-4 | T | 90% | 85% | 90% | 95% | SpCas9-CWNWNS | C | W | N | W | N | S |
| vNGCN-5 | T | 100% | 100% | 100% | 100% | SpCas9-LWRSEY | L | W | R | S | E | Y |
| vNGCN-6 | T | - | 95% | 95% | 100% | SpCas9-QSTWNK | Q | S | T | W | N | K |
| vNGCN-7 | C | n/a | n/a | n/a | n/a | SpCas9-LFEWRA | L | F | E | W | R | A |
| vNGCN-8 | C | - | - | 100% | - | SpCas9-SQSWRS | S | Q | S | W | R | S |
| vNGCN-9 | C | - | - | 100% | - | SpCas9-LKAWRS | L | K | A | W | R | S |
| vNGCN-10 | A | - | 100% | 100% | 100% | SpCas9-LWGWQH | L | W | G | W | Q | H |
| vNGCN-11 | A | - | 15% | 15% | 95% | SpCas9-LSYWAK | L | S | Y | W | A | K |
| vNGCN-12 | A | 50% | 10% | 20% | 95% | SpCas9-RQMYQG | R | Q | M | Y | Q | G |
| vNGCN-13 | 产生的变体 | - | - | - | - | SpCas9-LWREER | L | W | R | E | E | R |
| vNGCN-14 | 产生的变体 | 100% | 5% | 10% | 20% | SpCas9-VSSWER | V | S | S | W | E | R |
| vNGCN-15 | 产生的变体 | 100% | 3% | 5% | 15% | SpCas9-VSAWER | V | S | A | W | E | R |
| vNGCN-16 | 产生的变体 | - | - | - | - | SpCas9-DWREER | D | W | R | E | E | R |
| vNGCN-17 | 产生的变体 | 100% | - | - | - | SpCas9-VSGWER | V | S | G | W | E | R |
| vNGCN-18 | G | 100% | 100% | 100% | 100% | SpCas9-MCSFER | M | C | S | F | E | R |
| vNGCN-19 | G | 100% | - | - | 25% | SpCas9-VLMYER | V | L | M | Y | E | R |
| vNGCN-20 | G | 100% | n/a | n/a | n/a | SpCas9-QGANER | Q | G | A | N | E | R |
| vNGCN-21 | G | 100% | 50% | 15% | 50% | SpCas9-GCACER | G | C | A | C | E | R |
| vNGCN-22 | G | 100% | - | - | 5% | SpCas9-SRIAER | S | R | I | A | E | R |

[0159]

表2：针对NGCN PAM的变体在细菌中的选择结果和活性

| 样品 # | 针对 (NGCN) 的最初选择 | 在细菌测定中针对以下情况的近似存活率： | | | | 变体名称 | 变体中的氨基酸取代： | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | NGCG | NGCT | NGCC | NGCA | | D1135 | S1136 | G1218 | E1219 | R1335 | T1337 |
| vNGCN-23 | G | 100% | - | - | 25% | SpCas9-SRRNER | S | R | R | N | E | R |
| **vNGCN-10*** | T | - | 100% | 90% | 100% | SpCas9-LWGWQH | L | W | G | W | Q | H |
| vNGCN-24 | T | - | 5% | - | - | SpCas9-WMQAVV | W | M | Q | A | V | V |
| vNGCN-25 | T | - | 100% | - | 75% | SpCas9-AYRWSK | A | Y | R | W | S | K |
| vNGCN-26 | T | 100% | 100% | 30% | 65% | SpCas9-LWMREQ | L | W | M | R | E | Q |
| vNGCN-27 | T | - | 100% | 5% | 50% | SpCas9-LWRVVA | L | W | R | V | V | A |
| vNGCN-28 | T | 100% | 100% | n/a | 75% | SpCas9-HSSWVR | H | S | S | W | V | R |
| vNGCN-29 | C | 100% | 100% | 100% | 85% | SpCas9-MWSEPT | M | W | S | E | P | T |
| vNGCN-30 | C | 100% | 100% | 50% | 80% | SpCas9-GWSMQR | G | W | S | M | Q | R |
| vNGCN-31 | C | - | n/a | 75% | - | SpCas9-NKAWRV | N | K | A | W | R | V |
| vNGCN-32 | C | 75% | - | 95% | 50% | SpCas9-LCTYEY | L | C | T | Y | E | Y |
| vNGCN-33 | C | 80% | 5% | 50% | 50% | SpCas9-GSNWCK | G | S | N | W | C | K |
| vNGCN-34 | C | 85% | 50% | 90% | 100% | SpCas9-GSNYQS | G | S | N | Y | Q | S |
| vNGCN-35 | A | n/a | 50% | 25% | 90% | SpCas9-FMQWVN | F | M | Q | W | V | N |
| vNGCN-36 | A | 40% | 50% | 75% | 100% | SpCas9-YCSWVG | Y | C | S | W | V | G |
| vNGCN-37 | A | 50% | - | 25% | 85% | SpCas9-LWKFEG | L | W | K | F | E | G |
| vNGCN-38 | A | 25% | 35% | 5% | 100% | SpCas9-MCAWCG | M | C | A | W | C | G |
| vNGCN-39 | A | 50% | - | 50% | 50% | SpCas9-GKNWNR | G | K | N | W | N | R |
| **vNGCN-2*** | A | 100% | 25% | 25% | 75% | SpCas9-MQKSER | M | Q | K | S | E | R |
| vNGCN-40 | 产生的变体 | n/a | n/a | n/a | n/a | SpCas9-VRREER | V | R | R | E | E | R |
| vNGCN-41 | A | n/a | n/a | n/a | n/a | SpCas9-AARWCQ | A | A | R | W | C | Q |
| vNGCN-42 | A | n/a | n/a | n/a | n/a | SpCas9-LWLETR | L | W | L | E | T | R |
| vNGCN-43 | A | n/a | n/a | n/a | 85% | SpCas9-FMQWVR | F | M | Q | W | V | R |
| vNGCN-44 | A | n/a | n/a | n/a | 75% | SpCas9-SSKWPA | S | S | K | W | P | A |
| vNGCN-45 | C | n/a | n/a | 50% | n/a | SpCas9-MWASEG | M | W | A | S | E | G |
| vNGCN-46 | A | n/a | n/a | n/a | 100% | SpCas9-LSRWQR | L | S | R | W | Q | R |
| vNGCN-47 | G | 90% | n/a | n/a | n/a | SpCas9-YAIYER | Y | A | I | Y | E | R |
| vNGCN-48 | G | 75% | n/a | n/a | n/a | SpCas9-ICCCER | I | C | C | C | E | R |
| vNGCN-49 | G | 95% | n/a | n/a | n/a | SpCas9-DWFYER | D | W | F | Y | E | R |
| vNGCN-50 | G | 80% | n/a | n/a | n/a | SpCas9-REATER | R | E | A | T | E | R |
| vNGCN-51 | G | 75% | n/a | n/a | n/a | SpCas9-GWAYER | G | W | A | Y | E | R |
| vNGCN-52 | G | 75% | n/a | n/a | n/a | SpCas9-YAIYER | Y | A | I | Y | E | R |
| vNGCN-53 | G | 85% | n/a | n/a | n/a | SpCas9-LSVSER | L | S | V | S | E | R |
| vNGCN-54 | A | n/a | n/a | n/a | 75% | SpCas9-VRAWCR | V | R | A | W | C | R |
| vNGCN-55 | G | 75% | n/a | n/a | n/a | SpCas9-KWREQR | K | W | R | E | Q | R |
| vNGCN-56 | G | 75% | n/a | n/a | n/a | SpCas9-ARGAER | A | R | G | A | E | R |

[0160]

表2：针对NGCN PAM的变体在细菌中的选择结果和活性

| 样品 # | 针对(NGCN)的最初选择 | 在细菌测定中针对以下情况的近似存活率： | | | | 变体名称 | 变体中的氨基酸取代： | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | NGCG | NGCT | NGCC | NGCA | | D1135 | S1136 | G1218 | E1219 | R1335 | T1337 |
| vNGCN-57 | C | n/a | n/a | 75% | n/a | SpCas9-HASWCK | H | A | S | W | C | K |
| vNGCN-58 | G | 100% | n/a | n/a | n/a | SpCas9-YVRSER | Y | V | R | S | E | R |
| vNGCN-59 | G | 80% | n/a | n/a | n/a | SpCas9-QRLAER | Q | R | L | A | E | R |
| vNGCN-60 | A | n/a | n/a | n/a | n/a | SpCas9-AARWER | A | A | R | W | E | R |
| vNGCN-61 | G | 75% | n/a | n/a | n/a | SpCas9-LILSER | L | I | L | S | E | R |
| vNGCN-62 | A | n/a | n/a | n/a | n/a | SpCas9-LWPSRG | L | W | P | S | R | G |
| vNGCN-63 | A | n/a | n/a | n/a | n/a | SpCas9-LWTWIK | L | W | T | W | I | K |
| vNGCN-64 | 产生的变体 | n/a | n/a | n/a | n/a | SpCas9-VRKSER | V | R | K | S | E | R |
| vNGCN-65 | 产生的变体 | n/a | n/a | n/a | n/a | SpCas9-ICKSER | I | C | K | S | E | R |
| vNGCN-66 | C/T | n/a | n/a | n/a | n/a | SpCas9-MQSVQL | M | Q | S | V | Q | L |
| vNGCN-67 | 产生的变体 | n/a | n/a | n/a | n/a | SpCas9-LRSVER | L | R | S | V | E | R |
| vNGCN-68 | 产生的变体 | n/a | n/a | n/a | n/a | SpCas9-LSRWER | L | S | R | W | E | R |

* = 该变体已经在其他实验中进行过筛选
n/a = 在该PAM的实验中未评估存活率

表3：针对NGAN PAM的变体在细菌中的选择结果和活性

| 样品 # | 针对(NGAN)的最初选择 | 在细菌测定中针对以下情况的近似存活率： | | | | 变体名称 | 变体中的氨基酸取代： | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | NGAG | NGAT | NGAC | NGAA | | D1135 | S1136 | G1218 | E1219 | R1335 | T1337 |
| vNGAN-1 | G | 100% | 10% | 5% | 1% | SpCas9-LRLSAR | L | R | L | S | A | R |
| vNGAN-2 | G | 100% | - | 1% | - | SpCas9-ASEVTR | A | S | E | V | T | R |
| vNGAN-3 | T | 100% | 100% | 95% | 90% | SpCas9-KWMMCG | K | W | M | M | C | G |
| vNGAN-4 | C | - | - | 100% | - | SpCas9-VRGAKE | V | R | G | A | K | E |
| vNGAN-5 | C | - | - | 100% | - | SpCas9-MRARKE | M | R | A | R | K | E |
| vNGAN-6 | G | 75% | - | - | - | SpCas9-AEEQQR | A | E | E | Q | Q | R |
| vNGAN-7 | A | 95% | 5% | - | 80% | SpCas9-TRGSFR | T | R | G | S | F | R |
| vNGAN-8 | A | 95% | 10% | 90% | 90% | SpCas9-VRNYTK | V | R | N | Y | T | K |
| vNGAN-9 | T | 100% | 100% | 95% | 95% | SpCas9-AWNFQV | A | W | N | F | Q | V |
| vNGAN-10 | A | 100% | 35% | - | 20% | SpCas9-WMRKVA | W | M | R | K | V | A |
| vNGAN-11 | A | 40% | 100% | - | 75% | SpCas9-CWTCLQ | C | W | T | C | L | Q |
| vNGAN-12 | A | 100% | 100% | 5% | 75% | SpCas9-LWTTLN | L | W | T | T | L | N |

[0161]

表3：针对NGAN PAM的变体在细菌中的选择结果和活性

| 样品 # | 针对(NGAN)的最初选择 | 在细菌测定中针对以下情况的近似存活率： | | | | 变体名称 | 变体中的氨基酸取代： | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | NGAG | NGAT | NGAC | NGAA | | D1135 | S1136 | G1218 | E1219 | R1335 | T1337 |
| vNGAN-13 | G | 100% | 95% | 95% | 95% | SpCas9-SRMHCK | S | R | M | H | C | K |
| vNGAN-14 | T | 100% | 100% | 95% | 95% | SpCas9-CWCQCV | C | W | C | Q | C | V |
| vNGAN-15 | T | 100% | 5% | - | 10% | SpCas9-GCLCVR | G | C | L | C | V | R |
| vNGAN-16 | C | 100% | 50% | - | - | SpCas9-GGCQLR | G | G | C | Q | L | R |
| vNGAN-17 | G | 100% | - | - | - | SpCas9-AEEQQR | A | E | E | Q | Q | R |
| vNGAN-18 | G | 90% | 100% | 10% | 25% | SpCas9-QNNQVF | Q | N | N | Q | V | F |
| vNGAN-19 | T | 100% | 100% | - | 100% | SpCas9-GWEKVR | G | W | E | K | V | R |
| vNGAN-20 | T | 1% | 100% | 50% | - | SpCas9-NRAVNG | N | R | A | V | N | G |
| vNGAN-21 | T | 1% | 100% | 50% | - | SpCas9-NRAVNG | N | R | A | V | N | G |
| vNGAN-22 | C | 100% | 1% | 100% | - | SpCas9-SRQMRG | S | R | Q | M | R | G |
| vNGAN-23 | C | - | - | - | - | SpCas9-RAQPNL | R | A | Q | P | N | L |
| vNGAN-24 | A | 50% | 5% | - | 100% | SpCas9-LRSYLH | L | R | S | Y | L | H |
| vNGAN-25 | G | 100% | 95% | 100% | 90% | SpCas9-SRMHCK | S | R | M | H | C | K |
| vNGAN-26 | G | 100% | - | - | - | SpCas9-ACTSVR | A | C | T | S | V | R |
| vNGAN-27 | G | 100% | - | - | - | SpCas9-MVVHIR | M | V | V | H | I | R |
| vNGAN-28 | G | 100% | - | - | - | SpCas9-VRGNNR | V | R | G | N | N | R |
| vNGAN-29 | G | 100% | - | - | - | SpCas9-RGFCLR | R | G | F | C | L | R |
| vNGAN-30 | G | 100% | - | - | - | SpCas9-VQDAQR | V | Q | D | A | Q | R |
| vNGAN-31 | T | 100% | 100% | 95% | 95% | SpCas9-GWRQSK | G | W | R | Q | S | K |
| vNGAN-32 | T | 5% | 100% | - | - | SpCas9-AWLCLS | A | W | L | C | L | S |
| vNGAN-33 | T | 100% | 100% | - | 100% | SpCas9- KWARVV | K | W | A | R | V | V |
| vNGAN-34 | T | 80% | 100% | 20% | 15% | SpCas9-LAAQTP | L | A | A | Q | T | P |
| vNGAN-35 | T | 95% | 100% | 10% | 90% | SpCas9-GWNHLQ | G | W | N | H | L | Q |
| vNGAN-36 | T | 100% | 100% | 100% | 5% | SpCas9-MWAARP | M | W | A | A | R | P |
| vNGAN-37 | C | 95% | 100% | 50% | 30% | SpCas9-KWRCTG | K | W | R | C | T | G |
| vNGAN-38 | C | 50% | - | 100% | - | SpCas9-LAKARP | L | A | K | A | R | P |
| vNGAN-39 | C | 100% | 100% | 100% | 30% | SpCas9-SRMHCK | S | R | M | H | C | K |
| vNGAN-40 | C | - | - | 100% | - | SpCas9-VKMAKG | V | K | M | A | K | G |
| vNGAN-41 | C | - | - | 100% | - | SpCas9-QRKTRE | Q | R | K | T | R | E |
| vNGAN-42 | C | - | - | 50% | - | SpCas9-NTAVKQ | N | T | A | V | K | Q |
| vNGAN-43 | A | 100% | 100% | 50% | 100% | SpCas9-LCRQQR | L | C | R | Q | Q | R |
| vNGAN-44 | A | 100% | 90% | 100% | 100% | SpCas9-CWSHQR | C | W | S | H | Q | R |
| vNGAN-45 | A | 30% | 90% | 25% | 100% | SpCas9-MWVHLN | M | W | V | H | L | N |
| vNGAN-46 | A | 100% | 100% | 25% | 100% | SpCas9-SRTHTQ | S | R | T | H | T | Q |
| vNGAN-47 | A | 100% | 50% | - | 100% | SpCas9-LQKSMR | L | Q | K | S | M | R |
| vNGAN-48 | A | 100% | 100% | - | 90% | SpCas9-LWEVIR | L | W | E | V | I | R |

[0162]

表3：针对NGAN PAM的变体在细菌中的选择结果和活性

| 样品 # | 针对 (NGAN) 的最初选择 | 在细菌测定中针对以下情况的近似存活率： | | | | 变体名称 | 变体中的氨基酸取代： | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | NGAG | NGAT | NGAC | NGAA | | D1135 | S1136 | G1218 | E1219 | R1335 | T1337 |
| vNGTN-37 | 产生的变体 | 20% | - | - | - | SpCas9-VSKLLR | V | S | K | L | L | R |
| vNGTN-38 | 产生的变体 | 50% | - | - | - | SpCas9-VRKLLR | V | R | K | L | L | R |
| **vNGTN-27*** | NGTG | 10% | - | - | - | SpCas9-GAKLLR | G | A | K | L | L | R |
| vNGTN-39 | 产生的变体 | - | - | - | - | SpCas9-VSAVQL | V | S | A | V | Q | L |
| vNGTN-40 | 产生的变体 | 1% | - | - | - | SpCas9-VRAVQL | V | R | A | V | Q | L |
| **vNGTN-9*** | NGTT | 5% | - | 1% | - | SpCas9-IRAVQL | I | R | A | V | Q | L |
| vNGTN-41 | 产生的变体 | 1% | - | 100% | - | SpCas9-VSSVRS | V | S | S | V | R | S |
| vNGTN-42 | 产生的变体 | 25% | - | 100% | - | SpCas9-VRSVRS | V | R | S | V | R | S |
| **vNGTN-18*** | NGTC | 25% | - | 100% | - | SpCas9-LRSVRS | L | R | S | V | R | S |
| vNGTN-43 | N/A | 50% | - | - | 1% | SpCas9-SRGERT | S | R | G | E | R | T |
| vNGTN-44 | N/A | 90% | 80% | n/a | 50% | SpCas9-SRMHCK | S | R | M | H | C | K |
| vNGTN-52 | 产生的变体 | 60% | - | 75% | - | SpCas9-VRKLRS | V | R | K | L | R | S |
| **vNGTN-38*** | 产生的变体 | 60% | - | - | - | SpCas9-VRKLLR | V | R | K | L | L | R |
| vNGTN-53 | 产生的变体 | 15% | 5% | 10% | 1% | SpCas9-LRSVQL | L | R | S | V | Q | L |
| **vNGTN-18*** | NGTC | 50% | 5% | 100% | - | SpCas9-LRSVRS | L | R | S | V | R | S |
| vNGTN-54 | 产生的变体 | 50% | 5% | 100% | - | SpCas9-IRAVRS | I | R | A | V | R | S |
| vNGTN-55 | NGTT | 5% | 1% | 1% | - | SpCas9-IRAVQL | I | R | A | V | Q | L |
| vNGTN-56 | 产生的变体 | 5% | - | 35% | - | SpCas9-VRKLKR | V | R | K | L | K | R |
| **vNGTN-38*** | 产生的变体 | 35% | - | - | - | SpCas9-VRKLLR | V | R | K | L | L | R |
| vNGTN-57 | 产生的变体 | 20% | - | 100% | - | SpCas9-SRSVRS | S | R | S | V | R | S |
| **vNGTN-18*** | NGTC | 25% | - | 100% | - | SpCas9-LRSVRS | L | R | S | V | R | S |
| vNGTN-58 | 产生的变体 | 85% | 5% | 95% | 25% | SpCas9-VRKIQK | V | R | K | I | Q | K |
| **vNGTN-7*** | NGTT | 85% | 100% | 100% | 95% | SpCas9-LWKIQK | L | W | K | I | Q | K |
| vNGTN-59 | 产生的变体 | 95% | 100% | 100% | 90% | SpCas9-VRMHCK | V | R | M | H | C | K |
| **vNGTN-44*** | N/A | 85% | 90% | 100% | 75% | SpCas9-SRMHCK | S | R | M | H | C | K |

***** = 变体已在其他实验中筛选
**n/a** = 在该实验中未对该PAM评估存活率

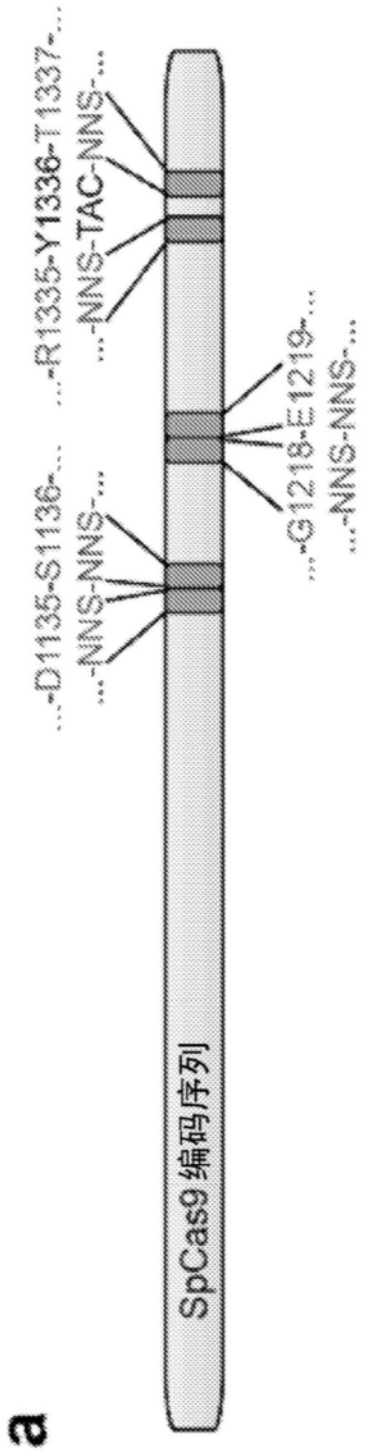

a
SpCas9 编码序列
...-D1135-S1136-...
...-NNS-NNS-...
...-R1335-Y1336-T1337-...
...-NNS-TAC-NNS-...
...-G1218-E1219-...
...-NNS-NNS-...

图1A

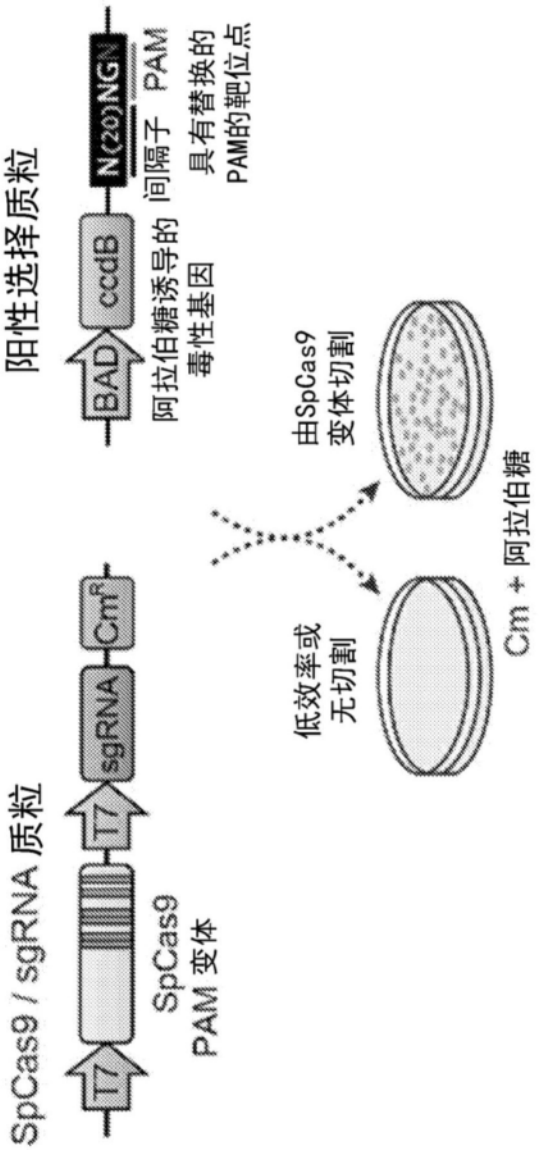

阳性选择质粒
N(20)NGN
PAM
间隔子
具有替换的
PAM的靶位点
BAD
ccdB
阿拉伯糖诱导的
毒性基因
SpCas9 / sgRNA 质粒
T7
SpCas9
PAM 变体
T7
sgRNA
Cm^R
由SpCas9
变体切割
低效率或
无切割
Cm + 阿拉伯糖

图1B

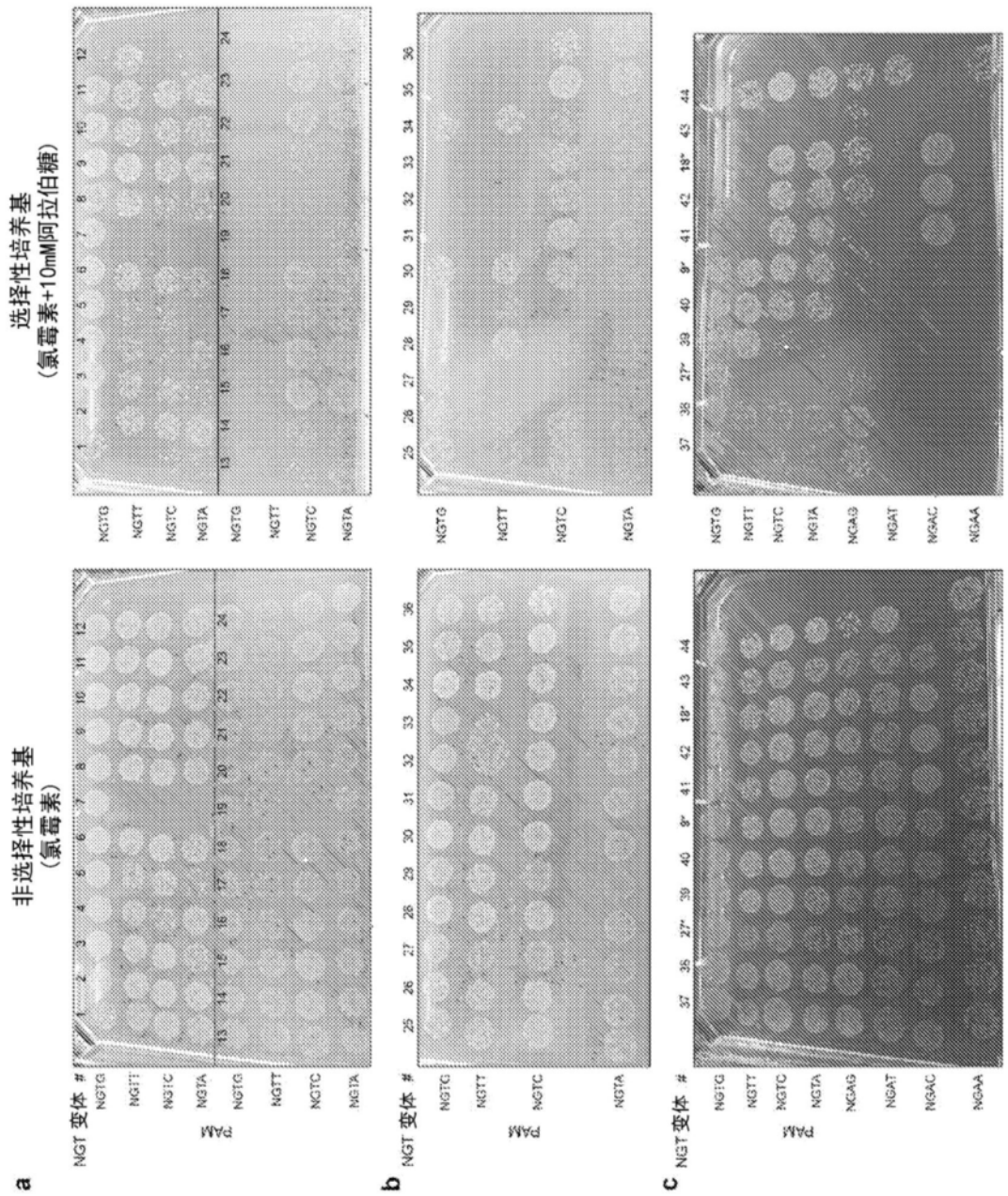

选择性培养基
(氯霉素+10mM阿拉伯糖)
非选择性培养基
(氯霉素)
a
b
c
NGT 变体 #
PAM
NGTG
NGTT
NGTC
NGTA
NGAG
NGAT
NGAC
NGAA

图2A-C

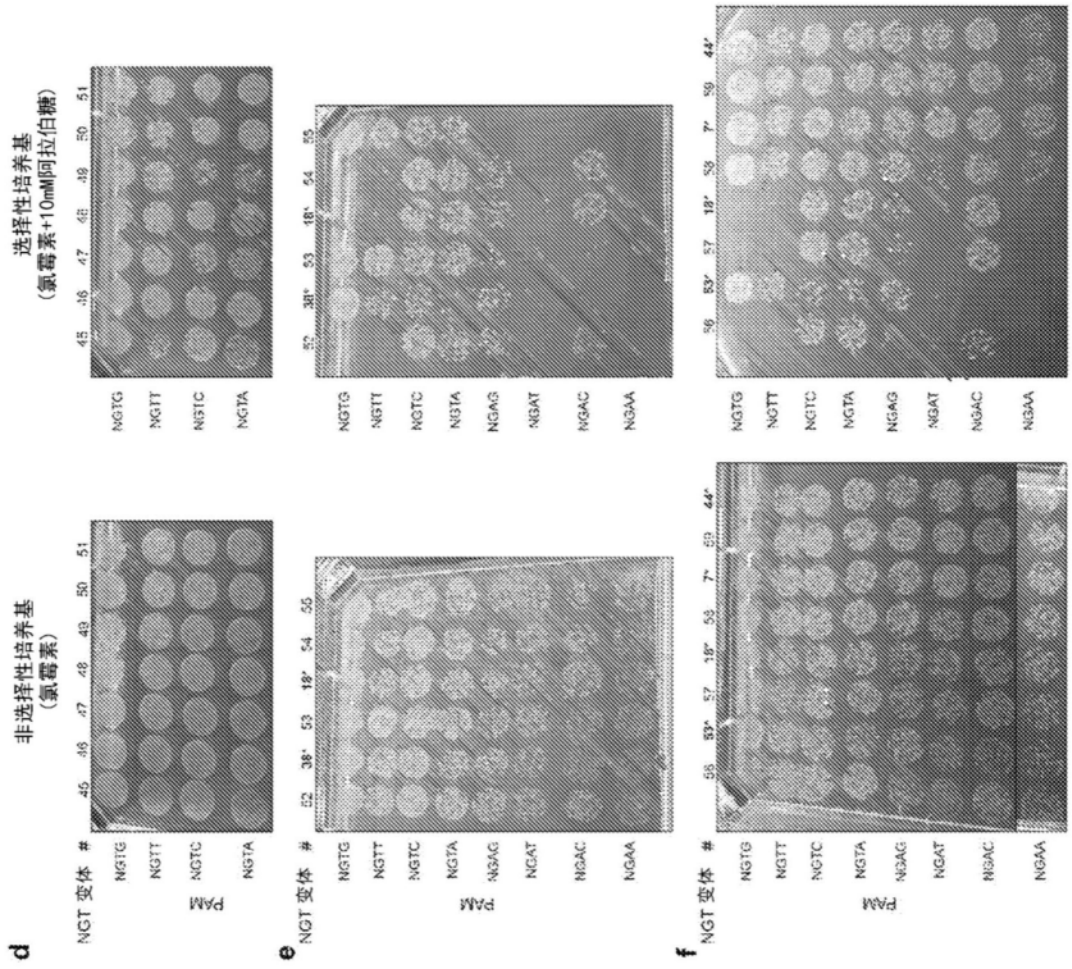

d
选择性培养基
(氯霉素+10mM阿拉伯糖)
非选择性培养基
(氯霉素)
NGT 变体 #
PAM
NGTG
NGTT
NGTC
NGTA
NGAG
NGAT
NGAC
NGAA
e
f

图2D-F

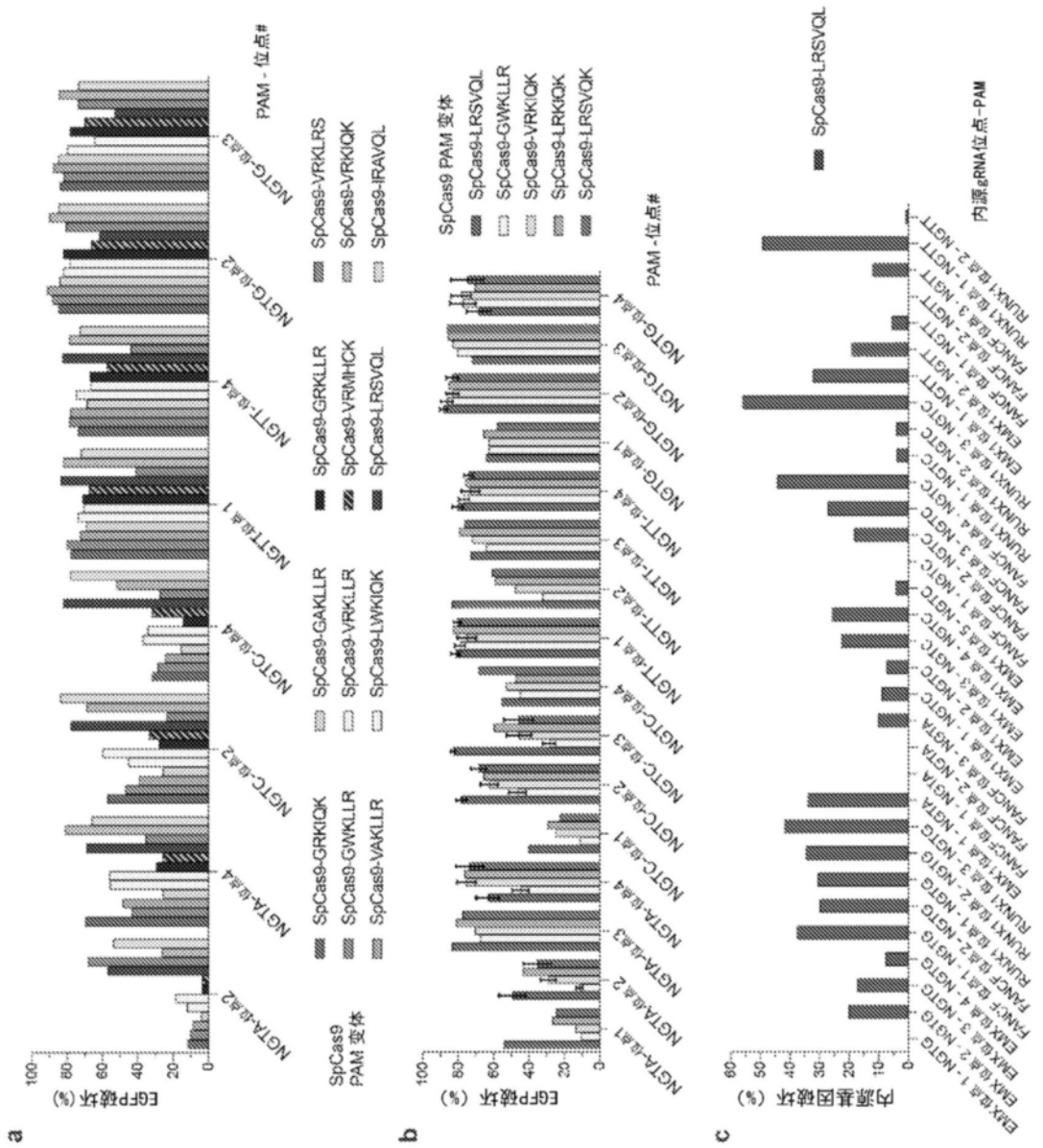

a
EGFP破坏（%）
SpCas9 PAM 变体
SpCas9-GRKIQK
SpCas9-GWKLLR
SpCas9-VAKLLR
SpCas9-GAKLLR
SpCas9-VRKLLR
SpCas9-LWKIQK
SpCas9-GRKLLR
SpCas9-VRMHCK
SpCas9-LRSVQL
SpCas9-VRKLRS
SpCas9-VRKIQK
SpCas9-IRAVQL
PAM - 位点#
NGTA-位点2
NGTA-位点4
NGTC-位点2
NGTC-位点4
NGTT-位点1
NGTT-位点4
NGTG-位点2
NGTG-位点3
b
EGFP破坏（%）
SpCas9 PAM 变体
SpCas9-LRSVQL
SpCas9-GWKLLR
SpCas9-VRKIQK
SpCas9-LRKIQK
SpCas9-LRSVQK
PAM -位点#
NGTA-位点1
NGTA-位点2
NGTA-位点3
NGTA-位点4
NGTC-位点1
NGTC-位点2
NGTC-位点3
NGTC-位点4
NGTT-位点1
NGTT-位点2
NGTT-位点3
NGTT-位点4
NGTG-位点1
NGTG-位点2
NGTG-位点3
NGTG-位点4
c
内源基因破坏（%）
SpCas9-LRSVQL
内源gRNA位点-PAM
EMX1位点1 - NGTG
EMX1位点2 - NGTG
EMX1位点3 - NGTG
EMX1位点4 - NGTG
FANCF位点1 - NGTG
FANCF位点2 - NGTG
RUNX1位点1 - NGTG
RUNX1位点2 - NGTG
EMX1位点1 - NGTA
EMX1位点2 - NGTA
EMX1位点3 - NGTA
EMX1位点4 - NGTA
FANCF位点1 - NGTA
FANCF位点2 - NGTA
RUNX1位点1 - NGTA
RUNX1位点2 - NGTA
EMX1位点1 - NGTC
EMX1位点2 - NGTC
EMX1位点3 - NGTC
EMX1位点4 - NGTC
EMX1位点5 - NGTC
FANCF位点1 - NGTC
FANCF位点2 - NGTC
FANCF位点3 - NGTC
FANCF位点4 - NGTC
RUNX1位点1 - NGTC
RUNX1位点2 - NGTC
RUNX1位点3 - NGTC
RUNX1位点4 - NGTC
EMX1位点1 - NGTT
EMX1位点2 - NGTT
EMX1位点3 - NGTT
EMX1位点4 - NGTT
FANCF位点1 - NGTT
FANCF位点2 - NGTT
FANCF位点3 - NGTT
FANCF位点4 - NGTT
RUNX1位点1 - NGTT
RUNX1位点2 - NGTT
RUNX1位点3 - NGTT
RUNX1位点4 - NGTT

图3A-C

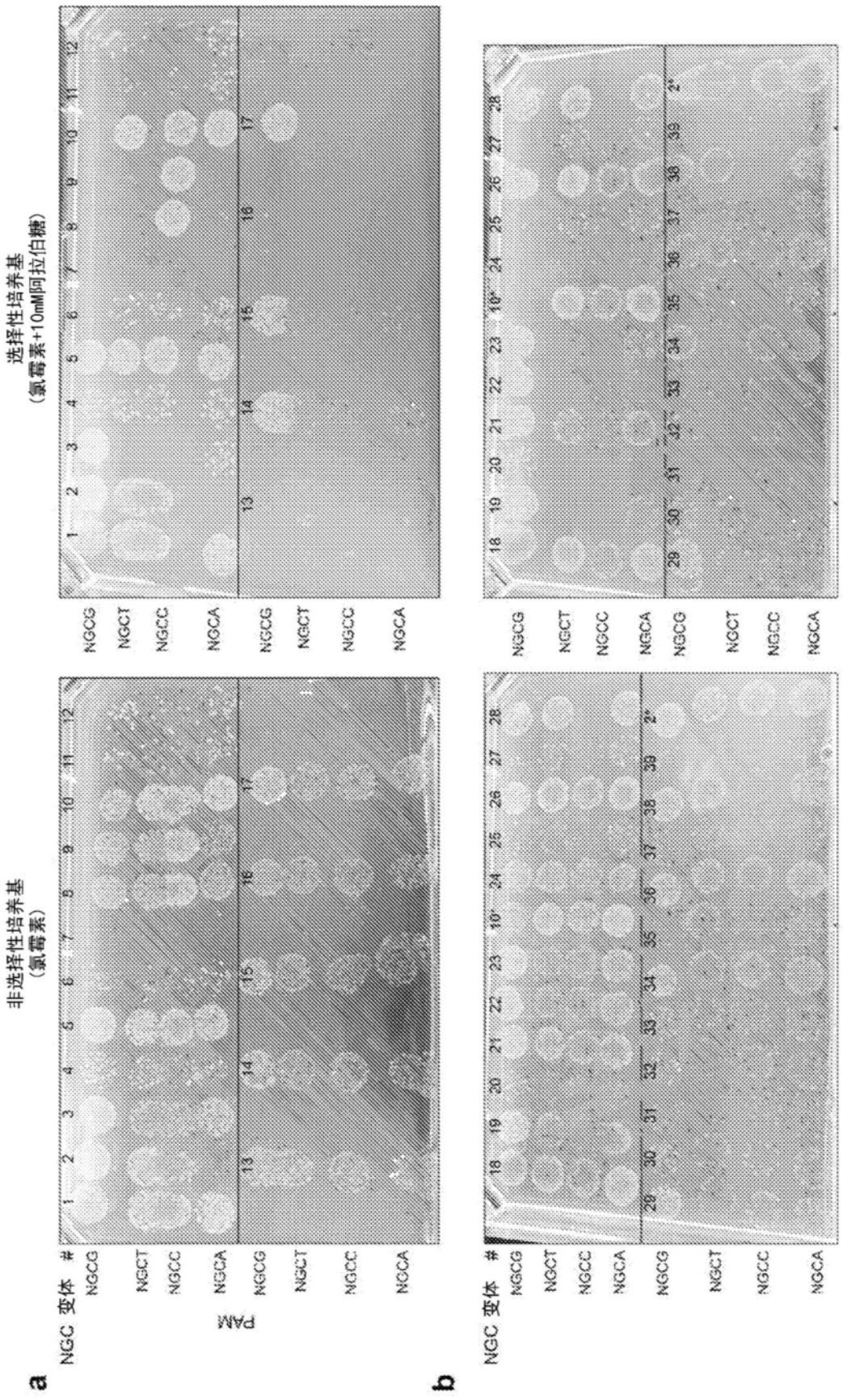

a
b
非选择性培养基
(氯霉素)
选择性培养基
(氯霉素+10mM阿拉伯糖)
NGC 变体 #
NGCG
NGCT
NGCC
NGCA

图4A-B

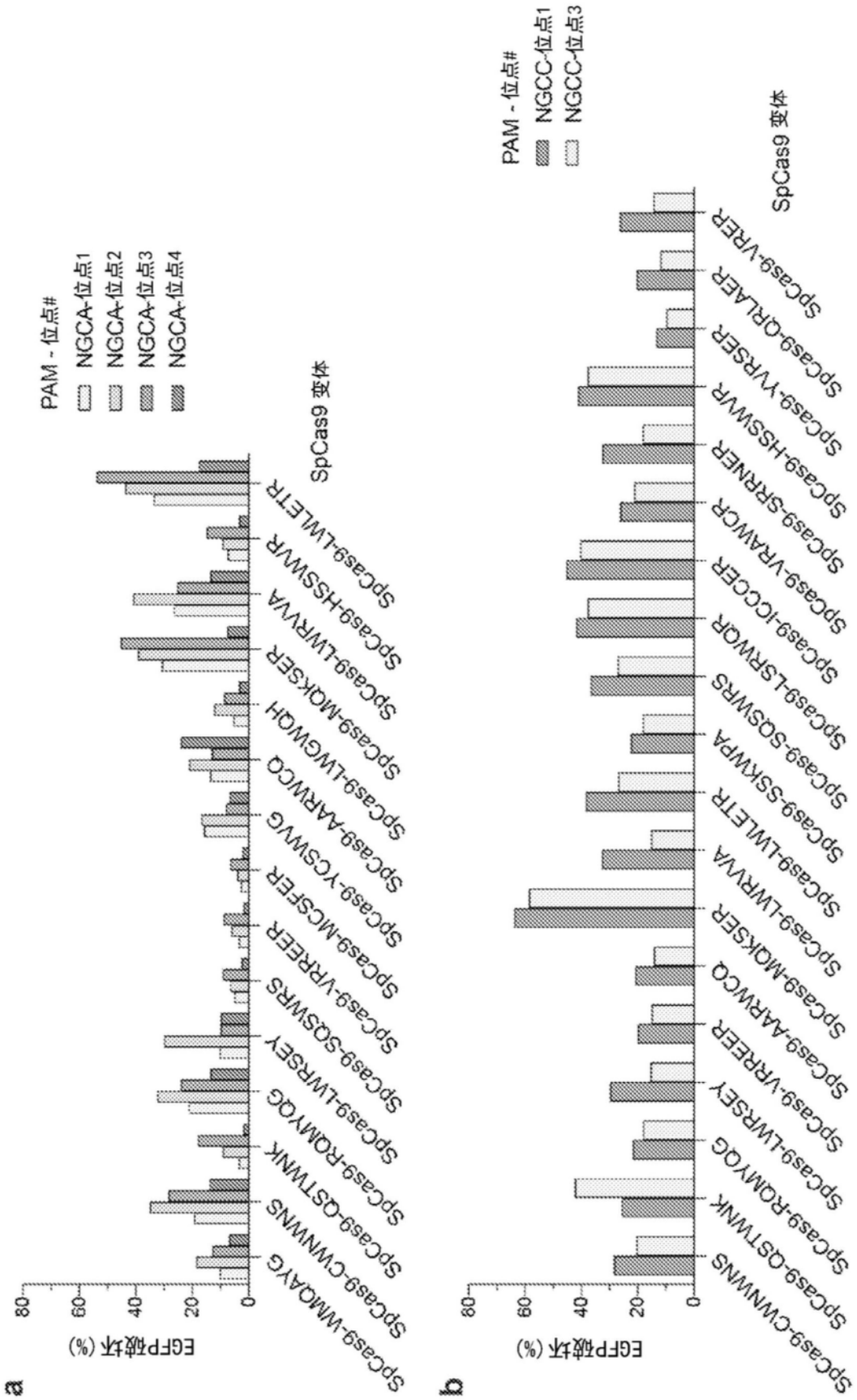

a
EGFP破坏(%)
0
20
40
60
80
SpCas9-WMQAVG
SpCas9-CWNWNS
SpCas9-QSTWNK
SpCas9-ROMYQG
SpCas9-LWRSEY
SpCas9-SQSWRS
SpCas9-VRREER
SpCas9-MCSFER
SpCas9-YCSWVG
SpCas9-AARWCQ
SpCas9-LWGWQH
SpCas9-MQKSER
SpCas9-LWRYVA
SpCas9-HSSWVR
SpCas9-LWLETR
SpCas9变体
PAM - 位点#
NGCA-位点1
NGCA-位点2
NGCA-位点3
NGCA-位点4
b
EGFP破坏(%)
0
20
40
60
80
SpCas9-CWNWNS
SpCas9-QSTWNK
SpCas9-ROMYQG
SpCas9-LWRSEY
SpCas9-VRREER
SpCas9-AARWCQ
SpCas9-MQKSER
SpCas9-LWRYVA
SpCas9-LWLETR
SpCas9-SSKWPA
SpCas9-SQSWRS
SpCas9-LSRWQR
SpCas9-ICCCER
SpCas9-VRAWCR
SpCas9-SRRNER
SpCas9-HSSWVR
SpCas9-VYRSER
SpCas9-QRLAER
SpCas9-VRER
SpCas9变体
PAM - 位点#
NGCC-位点1
NGCC-位点3

图5A-B

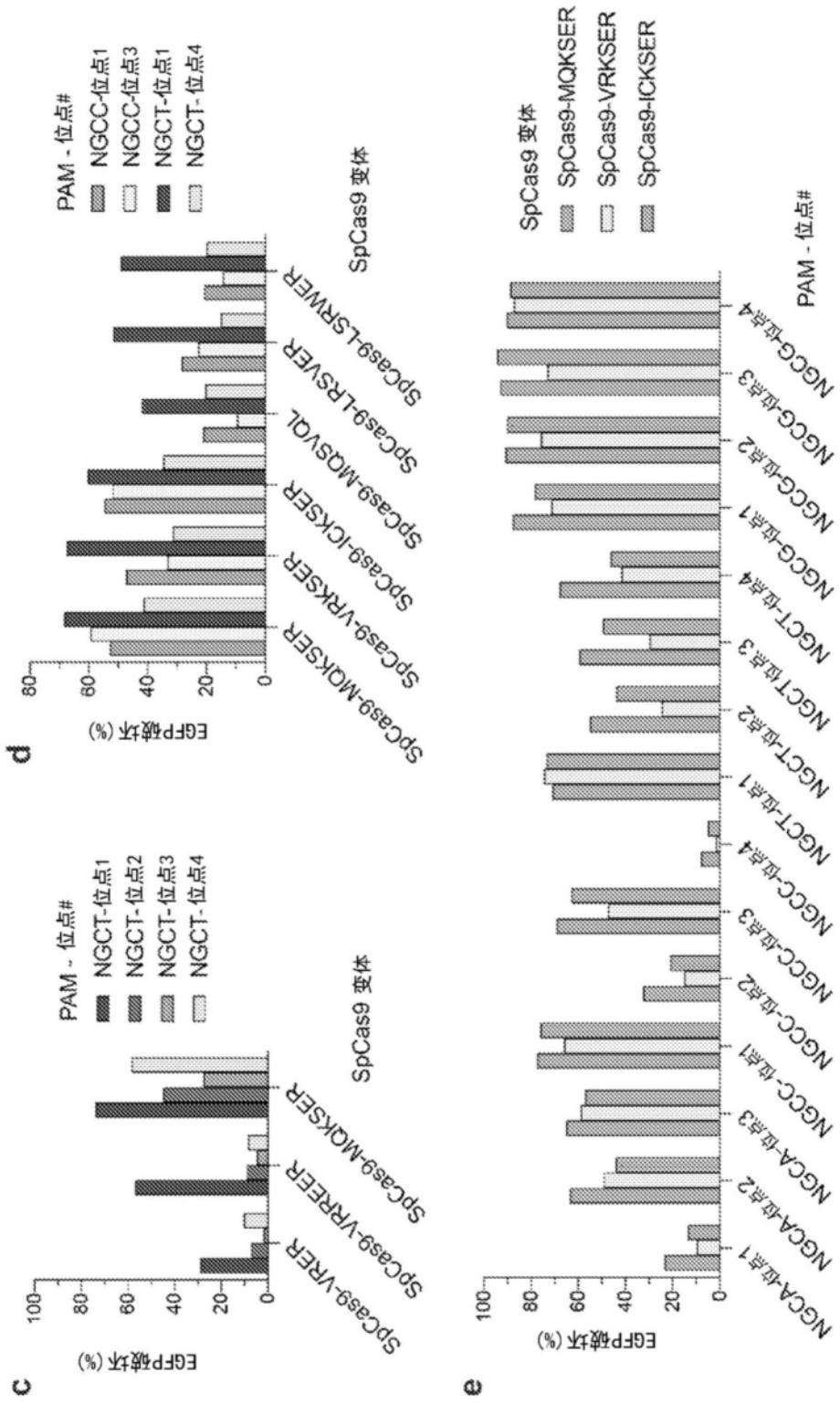

c
EGFP破坏(%)
PAM - 位点#
NGCT-位点1
NGCT-位点2
NGCT-位点3
NGCT-位点4
SpCas9 变体
SpCas9-VRER
SpCas9-VRREER
SpCas9-MQKSER
d
EGFP破坏(%)
PAM - 位点#
NGCC-位点1
NGCC-位点3
NGCT-位点1
NGCT-位点4
SpCas9 变体
SpCas9-MQKSER
SpCas9-VRKSER
SpCas9-ICKSER
SpCas9-MQSVQL
SpCas9-LRSVER
SpCas9-LSRWER
e
EGFP破坏(%)
PAM - 位点#
SpCas9 变体
SpCas9-MQKSER
SpCas9-VRKSER
SpCas9-ICKSER

图5C-E

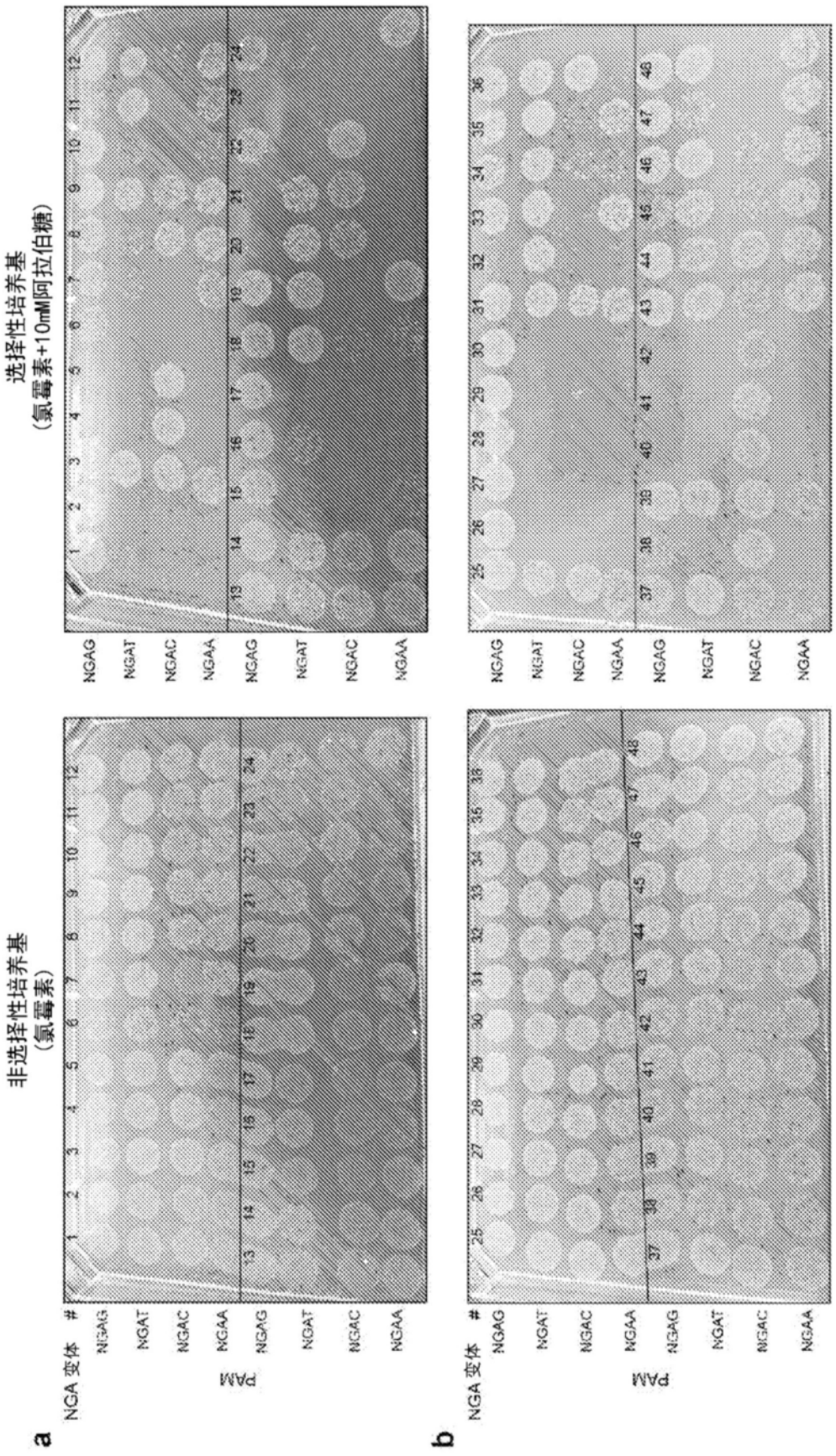

a
b
非选择性培养基
(氯霉素)
选择性培养基
(氯霉素+10mM阿拉伯糖)
NGA 变体 #
NGAG
NGAT
NGAC
NGAA
PAM

图6A-B

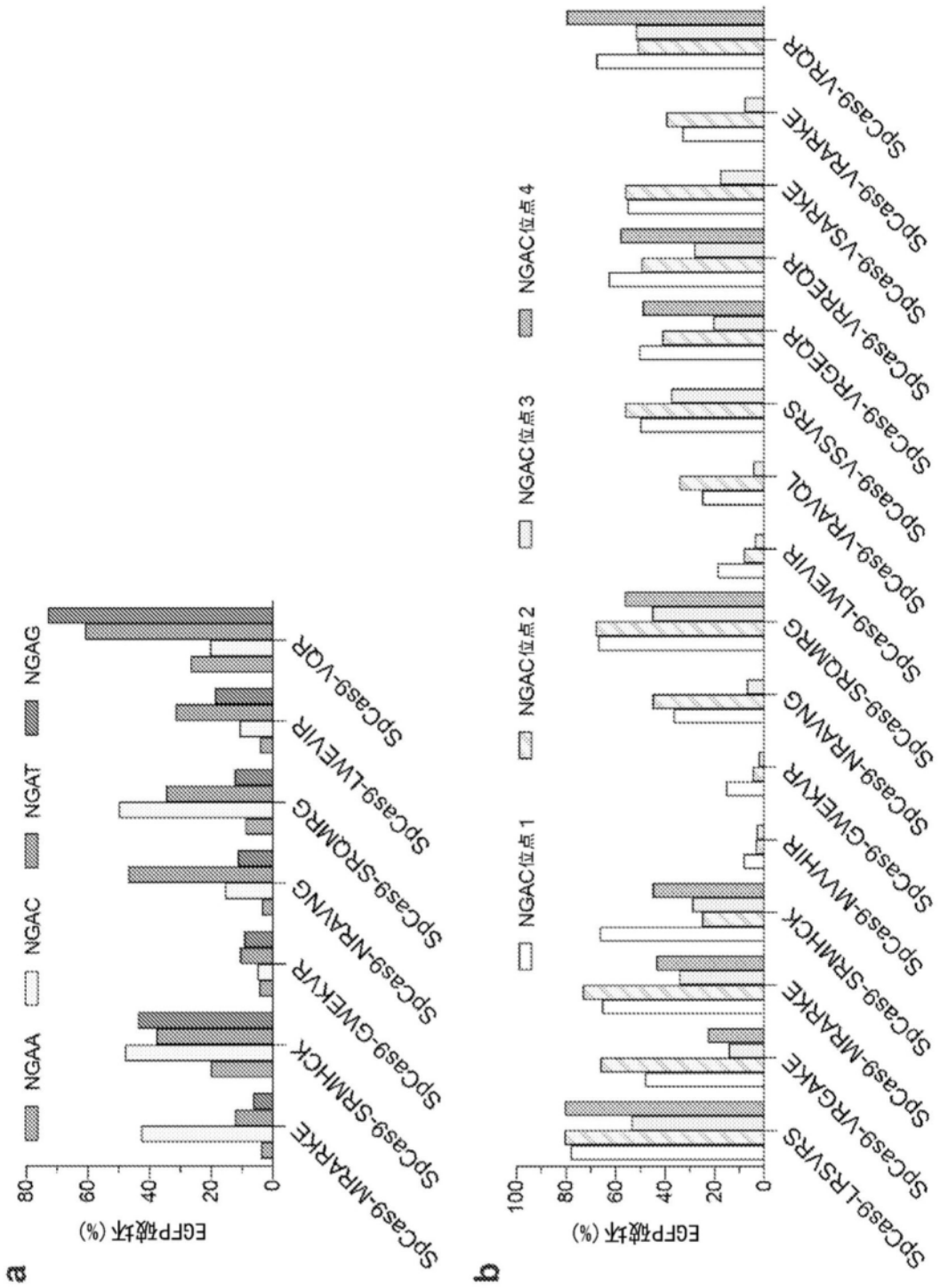

a
NGAA
NGAC
NGAT
NGAG
EGFP破坏 (%)
0
20
40
60
80
SpCas9-MRARKE
SpCas9-SRMHCK
SpCas9-GWEKVR
SpCas9-NRAVNG
SpCas9-SROMRG
SpCas9-LWEVIR
SpCas9-VQR
b
NGAC位点1
NGAC位点2
NGAC位点3
NGAC位点4
EGFP破坏 (%)
0
20
40
60
80
100
SpCas9-LRSVRS
SpCas9-VRGAKE
SpCas9-MRARKE
SpCas9-SRMHCK
SpCas9-MVVHIR
SpCas9-GWEKVR
SpCas9-NRAVNG
SpCas9-SROMRG
SpCas9-LWEVIR
SpCas9-VRAVQL
SpCas9-VSSVRS
SpCas9-VRGEQR
SpCas9-VRREQR
SpCas9-VSARKE
SpCas9-VRARKE
SpCas9-VRQR

图7A-B

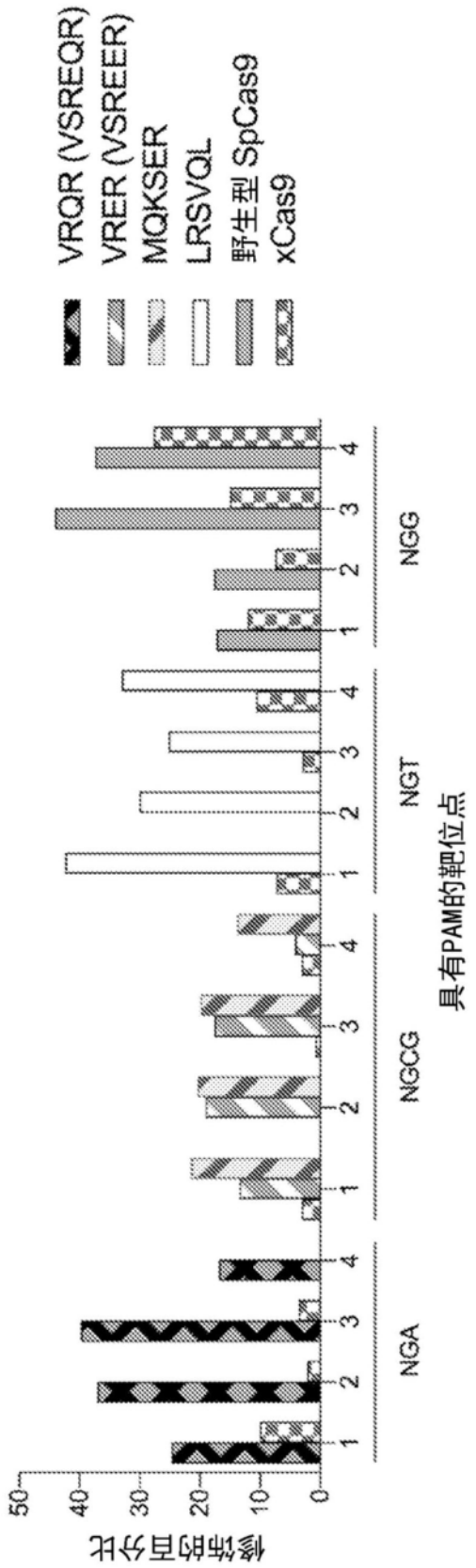

VRQR (VSREQR)
VRER (VSREER)
MQKSER
LRSVQL
野生型 SpCas9
xCas9
修饰的百分比
0
10
20
30
40
50
1
2
3
4
NGA
NGCG
NGT
NGG
具有PAM的靶位点

图8

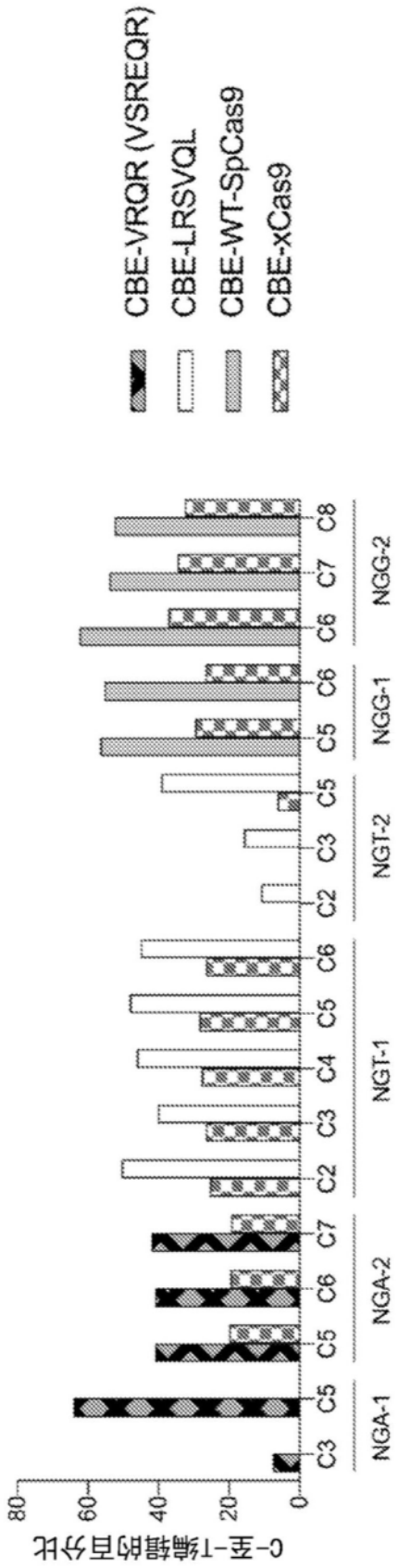

CBE-VRQR (VSREQR)
CBE-LRSVQL
CBE-WT-SpCas9
CBE-xCas9
C-至-T编辑的百分比
0
20
40
60
80
NGA-1
NGA-2
NGT-1
NGT-2
NGG-1
NGG-2
C3
C5
C5
C6
C7
C2
C3
C4
C5
C6
C2
C3
C5
C6
C5
C6
C7
C8

图9A

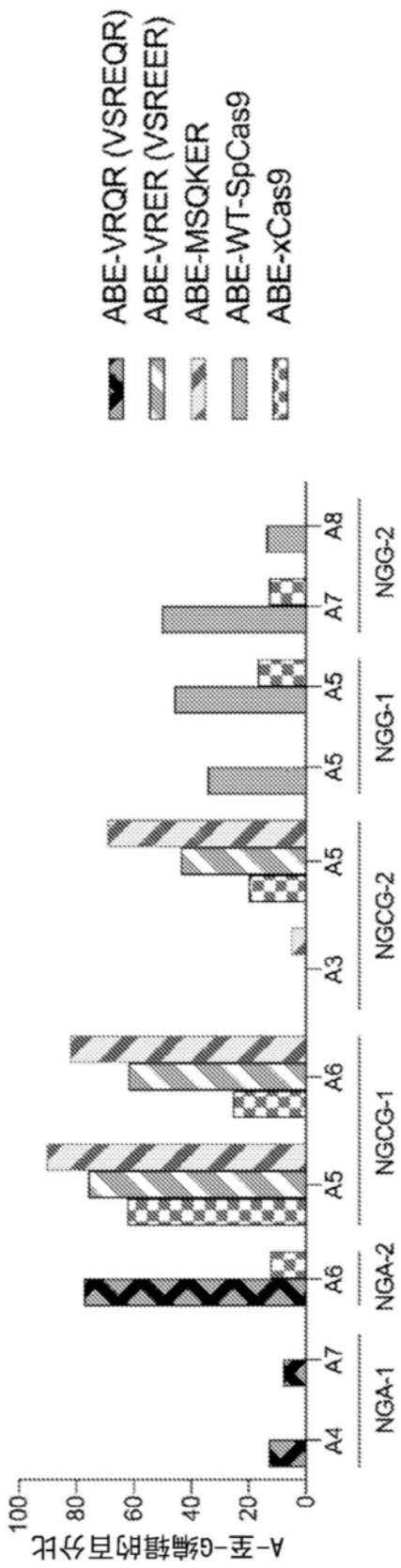

ABE-VRQR (VSREQR)
ABE-VRER (VSREER)
ABE-MSQKER
ABE-WT-SpCas9
ABE-xCas9
A-至-G编辑的百分比
0
20
40
60
80
100
A4
A7
A6
A5
A6
A3
A5
A5
A7
A8
NGA-1
NGA-2
NGCG-1
NGCG-2
NGG-1
NGG-2

图9B