

(19) 中华人民共和国国家知识产权局



(12) 发明专利申请

(10) 申请公布号 CN 113728099 A

(43) 申请公布日 2021.11.30

(21) 申请号 202080029346.8

(74) 专利代理机构 中国专利代理(香港)有限公司 72001

(22) 申请日 2020.02.13

代理人 罗文锋 初明明

(30) 优先权数据

62/806708 2019.02.15 US

(51) Int.Cl.

C12N 9/22 (2006.01)

C12N 15/10 (2006.01)

(85) PCT国际申请进入国家阶段日

2021.10.15

(86) PCT国际申请的申请数据

PCT/US2020/018145 2020.02.13

(87) PCT国际申请的公布数据

W02020/168102 EN 2020.08.20

(71) 申请人 西格马—奥尔德里奇有限责任公司

地址 美国密苏里州

(72) 发明人 陈福强

权利要求书2页 说明书30页

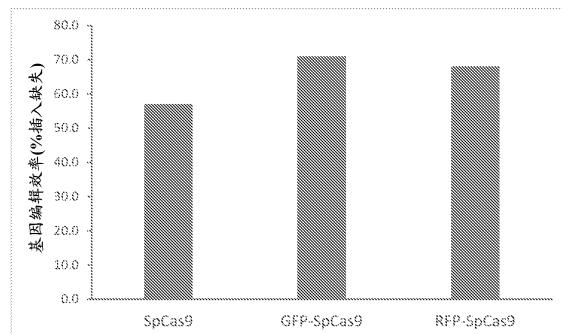
序列表59页 附图2页

(54) 发明名称

CRISPR/CAS融合蛋白和系统

(57) 摘要

本文公开改造的Cas9系统。



1. 一种融合蛋白,其包含连接至至少一个标记蛋白的Cas9蛋白。
2. 权利要求1的融合蛋白,其中所述至少一个标记蛋白通过化学键直接地、直接地或通过接头间接地、或其组合连接至所述Cas9蛋白。
3. 权利要求1的融合蛋白,其中所述至少一个标记蛋白直接地或通过接头间接地连接至所述Cas9蛋白。
4. 权利要求1至3中任一项的融合蛋白,其中所述至少一个标记蛋白连接至所述Cas9蛋白的N末端、C末端、内部位置或它们的组合。
5. 权利要求1至4中任一项的融合蛋白,进一步包含至少一个核定位信号。
6. 权利要求5的融合蛋白,其中所述核定位信号、标记蛋白、接头和Cas9蛋白按以下顺序(N末端至C末端)排列:
 标记蛋白-接头-核定位信号-Cas9蛋白;
 标记蛋白-核定位信号-接头-Cas9蛋白;
 核定位信号-接头-标记蛋白-Cas9蛋白;
 核定位信号-标记蛋白-接头-Cas9蛋白;
 标记蛋白-接头-核定位信号-接头-Cas9蛋白;或
 核定位信号-接头-标记蛋白-接头-Cas9蛋白。
7. 权利要求6的融合蛋白,其中所述核定位信号、标记蛋白、接头和Cas9蛋白按以下顺序(N末端至C末端)排列:
 标记蛋白-接头-核定位信号-Cas9蛋白。
8. 权利要求1至7中任一项的融合蛋白,进一步包含至少一个异源结构域。
9. 权利要求8的融合蛋白,其中所述至少一个异源结构域为细胞穿透结构域、染色质调控基序、表观遗传修饰结构域、转录调节结构域、RNA适体结合结构域或其组合。
10. 权利要求1至9中任一项的融合蛋白,其中所述融合蛋白是核酸酶并切割双链序列的两条链,是切口酶并切割双链序列的一条链,或没有核酸酶或切口酶活性。
11. 权利要求1至10中任一项的融合蛋白,其中所述cas9蛋白来自酿脓链球菌(*Streptococcus pyogenes*)、嗜热链球菌(*Streptococcus thermophilus*)、脑膜炎奈瑟菌(*Neisseria meningitidis*)、金黄葡萄球菌(*Staphylococcus aureus*)或空肠弯曲杆菌(*Campylobacter jejuni*)。
12. 权利要求1至11中任一项的融合蛋白,其中所述标记蛋白具有与SEQ ID NO:19或20有至少90%序列同一性的氨基酸序列。
13. 权利要求1至11中任一项的融合蛋白,其中所述标记蛋白具有阐明于SEQ ID NO:19或20的氨基酸序列。
14. 权利要求3至13中任一项的融合蛋白,其中所述接头具有与SEQ ID NO:35或36有至少90%序列同一性的氨基酸序列。
15. 权利要求3至13中任一项的融合蛋白,其中所述接头具有阐明于SEQ ID NO:35或36的氨基酸序列。
16. 权利要求3至15中任一项的融合蛋白,其中所述融合蛋白具有与SEQ ID NO:48、49或50有至少90%序列同一性的氨基酸序列。
17. 权利要求3至15中任一项的融合蛋白,其中所述融合蛋白具有如阐明于SEQ ID NO:

48、49或50的氨基酸序列。

18. 一种系统,其包含权利要求1至17中任一项的融合蛋白和改造的向导RNA。

19. 权利要求18的系统,其中所述改造的向导RNA是单个分子。

20. 权利要求18至19中任一项的系统,其中所述改造的向导RNA序列被优化以促进所述改造的向导RNA中的碱基配对,最小化所述改造的向导RNA中的碱基配对,增加所述改造的向导RNA的稳定性,促进所述改造的向导RNA在真核细胞中的转录或其组合。

21. 编码权利要求1至17中任一项的融合蛋白的多种核酸。

22. 编码权利要求18至20中任一项的系统的多种核酸,所述多种核酸包含编码所述融合蛋白的至少一种核酸,和编码所述改造的向导RNA的至少一种核酸。

23. 权利要求14的多种核酸,其中编码所述融合蛋白的所述至少一种核酸是RNA。

24. 权利要求14的多种核酸,其中编码所述融合蛋白的所述至少一种核酸是DNA。

25. 权利要求21至24中任一项的多种核酸,其中编码所述融合蛋白的所述至少一种核酸经密码子优化用于真核细胞中的表达。

26. 权利要求25的多种核酸,其中所述真核细胞为人类细胞、非人类哺乳动物细胞、非哺乳类脊椎动物细胞、无脊椎动物细胞、植物细胞或单细胞真核生物。

27. 权利要求22至26中任一项的多种核酸,其中编码所述改造的向导RNA的所述至少一种核酸是DNA。

28. 权利要求22至27中任一项的多种核酸,其中编码所述融合蛋白的所述至少一种核酸与噬菌体启动子序列可操作地连接,用于体外RNA合成或细菌细胞中的蛋白表达,且编码所述改造的向导RNA的所述至少一种核酸与噬菌体启动子序列可操作地连接,用于体外RNA合成。

29. 权利要求22至27中任一项的多种核酸,其中编码所述融合蛋白的所述至少一种核酸与真核启动子序列可操作地连接,用于真核细胞中的表达,且编码所述改造的向导RNA的所述至少一种核酸与真核启动子序列可操作地连接,用于真核细胞中的表达。

30. 至少一种载体,其包含权利要求21至29中任一项的多种核酸。

31. 权利要求30的至少一种载体,其为质粒载体、病毒载体或自复制病毒RNA复制子。

32. 一种真核细胞,其包含含有如权利要求1至17中定义的融合蛋白的至少一种系统、如权利要求18至20中定义的系统、如权利要求21至29中定义的多种核酸或如权利要求30至31中定义的至少一种载体。

33. 权利要求32的真核细胞,其为人类细胞、非人类哺乳动物细胞、植物细胞、非哺乳类脊椎动物细胞、无脊椎动物细胞或单细胞真核生物。

34. 权利要求32至33中任一项的真核细胞,其为体内的、离体的或体外的。

35. 一种用于确定在活真核细胞或化学固定的真核细胞中的染色体身份和位置的方法,所述方法包含将权利要求1至31中任一项的融合蛋白、系统、多种核酸或载体引入活的或化学固定的真核细胞,并探测来自所述标记蛋白的信号。

36. 权利要求35的方法,其中所述真核细胞为人类细胞、非人类哺乳动物细胞、植物细胞、非哺乳类脊椎动物细胞、无脊椎动物细胞或单细胞真核生物。

37. 权利要求35至36中任一项的方法,其中所述真核细胞是体内的、离体的或体外的。

CRISPR/CAS融合蛋白和系统

[0001] 相关申请

本申请要求提交于2019年2月15日的美国临时申请号62/806,708的优先权权益，所述临时申请的整体在此通过引用并入。

[0002] 领域

本公开内容涉及改造的Cas9系统，编码所述系统的核酸，以及使用所述系统用于基因组修饰的方法。

[0003] 背景

许多不同种类的肽接头已被测试使GFP融合至Cas9，但通常导致基础Cas9更低的活性。

[0004] 公开内容的概述

在本公开内容的各种方面中包括改造的Cas9系统。

[0005] 本公开内容的其它方面和特征于下文详述。

[0006] 附图简述

图1显示了本文公开的Cas9融合蛋白各自保持相似于SpCas9蛋白水平的编码活性。

[0007] 图2A和图2B显示了在所有目标中，本文公开的Cas9融合蛋白的编码效率是商业蛋白的数倍高。

[0008] 序列表

本申请包含序列表，其已经以ASCII格式电子提交，并在此通过引用以其整体并入。所述ASCII副本创建于2020年2月13日，命名为P19-027_WO-PCT_SL.txt，大小为87,735字节。

[0009] 详述

辅助蛋白向CRISPR蛋白的融合创造了范围广泛的机会以使各种蛋白功能局限于细胞内的限定位置。此外，公开了以能保持CRISPR功能性的方式实现异源蛋白向CRISPR蛋白的融合的肽接头。

[0010] (I) 改造的Cas9系统

本公开内容的一个方面提供改造的Cas9蛋白和系统。例如，公开了Cas9-标记物融合蛋白。在一些方面，系统包括改造的Cas9蛋白和改造的向导RNA，其中每种改造的向导RNA被设计用于与特异性改造的Cas9蛋白复合。这些改造的Cas9系统并不天然存在。

[0011] (a) 改造的Cas9蛋白

Cas9蛋白是II型CRISPR系统中的单个效应蛋白，II型CRISPR系统存在于各种细菌中。本文公开的改造的Cas9蛋白可来自于*Acaryochloris sp.*、醋盐杆菌属物种(*Acetohalobium sp.*)、氨基酸球菌属物种(*Acidaminococcus sp.*)、嗜酸硫杆菌属物种(*Acidithiobacillus sp.*)、热酸菌属物种(*Acidothermus sp.*)、艾克曼氏菌属物种(*Akkermansia sp.*)、脂环酸芽孢杆菌属物种(*Alicyclobacillus sp.*)、异着色菌属物种(*Allochromatium sp.*)、*Ammonifex sp.*、鱼腥藻属物种(*Anabaena sp.*)、节螺藻属物种

(*Arthrosphaera* sp.)、芽孢杆菌属物种(*Bacillus* sp.)、双歧杆菌属物种(*Bifidobacterium* sp.)、伯克氏菌属物种(*Burkholderiales* sp.)、热解纤维素菌属物种(*Caldicelulosiruptor* sp.)、弯曲杆菌属物种(*Campylobacter* sp.)、*Candidatus* sp.、梭菌属物种(*Clostridium* sp.)、棒杆菌属物种(*Corynebacterium* sp.)、*Crocosphaera* sp.、蓝丝菌属物种(*Cyanothece* sp.)、微小杆菌属物种(*Exiguobacterium* sp.)、丝状杆菌属物种(*Fibrobacter* sp.)、芬戈尔德菌属物种(*Finegoldia* sp.)、弗朗西斯氏菌属物种(*Francisella* sp.)、纤线杆菌属物种(*Ktedonobacter* sp.)、毛螺菌科物种(*Lachnospiraceae* sp.)、乳杆菌属物种(*Lactobacillus* sp.)、李斯特菌属物种(*Listeria* sp.)、鞘丝藻属物种(*Lyngbya* sp.)、海杆菌属物种(*Marinobacter* sp.)、甲烷盐菌属物种(*Methanohalobium* sp.)、微颤菌属物种(*Microscilla* sp.)、微鞘藻属物种(*Microcoleus* sp.)、微胞藻属物种(*Microcystis* sp.)、支原体属物种(*Mycoplasma* sp.)、*Natranaerobius* sp.、奈瑟菌属物种(*Neisseria* sp.)、*Nitratifractor* sp.、亚硝化球菌属物种(*Nitrosococcus* sp.)、诺卡氏菌属物种(*Nocardiopsis* sp.)、节球藻属物种(*Nodularia* sp.)、念珠藻属物种(*Nostoc* sp.)、酒球菌属物种(*Oenococcus* sp.)、颤藻属物种(*Oscillatoria* sp.)、副萨特氏菌属物种(*Parasutterella* sp.)、巴斯德氏菌属物种(*Pasteurella* sp.)、细小棒菌属物种(*Parvibaculum* sp.)、*Pelotomaculum* sp.、石孢菌属物种(*Petrotoga* sp.)、极地单胞菌属物种(*Polaromonas* sp.)、普雷沃氏菌属物种(*Prevotella* sp.)、假交替单胞菌属物种(*Pseudoalteromonas* sp.)、罗尔斯顿氏菌属物种(*Ralstonia* sp.)、红螺菌属物种(*Rhodospirillum* sp.)、葡萄球菌属物种(*Staphylococcus* sp.)、链球菌属物种(*Streptococcus* sp.)、链霉菌属物种(*Streptomyces* sp.)、链孢囊菌属物种(*Streptosporangium* sp.)、聚球藻属物种(*Synechococcus* sp.)、栖热腔菌属物种(*Thermosiphon* sp.)、密螺旋体属物种(*Treponema* sp.)、疣微菌门物种(*Verrucomicrobia* sp.)和沃林氏菌属物种(*Wolinella* sp.)。

[0012] Cas9蛋白或其它组分可能来自或衍生自的示例物种包括*Acaryochloris* spp. (如*Acaryochloris marina*)、醋盐杆菌属物种(如阿拉伯糖醋盐杆菌(*Acetohalobium arabaticum*))、氨基酸球菌属物种、嗜酸硫杆菌属物种(如喜温嗜酸硫杆菌(*Acidithiobacillus caldus*)、氧化亚铁酸硫杆菌(*Acidithiobacillus ferrooxidans*))、热酸菌属物种、艾克曼氏菌属物种、脂环酸芽孢杆菌属物种(如酸热脂环酸芽孢杆菌(*Alicyclobacillus acidocaldarius*))、异着色菌属物种(如*Allochromatium vinosum*)、*Ammonifex* spp. (如*Ammonifex degensii*)、鱼腥藻属物种(如多变鱼腥藻(*Anabaena variabilis*))、节螺藻属物种(如极大节螺藻(*Arthrosira maxima*)、盘状节螺藻(*Arthrosira platensis*))、芽孢杆菌属物种(如假蕈状芽孢杆菌(*Bacillus pseudomycoides*)、亚硒酸芽孢杆菌(*Bacillus selenitireducens*))、双歧杆菌属物种、伯克氏菌属物种(如*Burkholderiales bacterium*)、*Caldicelulosiruptor* spp. (如*Caldicelulosiruptor beccii*)、弯曲杆菌属物种(如空肠弯曲杆菌(*Campylobacter jejuni*))、红嘴鸥弯曲杆菌(*Campylobacter lari*))、*Candidatus* spp. (如*Candidatus Desulfurudis*)、梭菌属物种(如肉毒梭菌(*Clostridium botulinum*)、艰难梭菌(*Clostridium difficile*))、棒杆菌属物种(如白喉棒杆菌(*Corynebacterium diphtheriae*))、*Crocosphaera* spp. (如*Crocosphaera watsonii*)、蓝丝菌属物种、δ变形菌

纲物种 (*Delta proteobacterium* spp.)、微小杆菌属物种 (如西伯利亚微小杆菌 (*Exiguobacterium sibiricum*))、丝状杆菌属物种 (如产琥珀酸丝状杆菌 (*Fibrobacter succinogene*))、芬戈尔德菌属物种 (如大芬戈尔德菌 (*Finegoldia magna*))、弗朗西斯氏菌属物种 (如 *Francisella novicida*)、 γ -变形菌纲 (*Gammaproteobacterium*)、纤线杆菌属物种 (如 *Ktedonobacter racemifer*)、毛螺菌科物种、乳杆菌属物种 (如布氏乳杆菌 (*Lactobacillus buchneri*)、德氏乳杆菌 (*Lactobacillus delbrueckii*)、加氏乳杆菌 (*Lactobacillus gasseri*)、唾液乳杆菌 (*Lactobacillus salivarius*))、利斯特氏杆菌属物种 (如无害李斯特菌 (*Listeria innocua*))、纤毛菌属物种 (*Leptotrichia* spp.)、鞘丝藻属物种、海杆菌属物种、甲烷盐菌属物种 (如 *Methanohalobium evestigatum*)、微鞘藻属物种 (如原体型微鞘藻 (*Microcoleus chthonoplastes*))、微颤菌属物种 (如海洋微颤菌 (*Microscilla marina*))、微胞藻属物种 (如铜绿微胞藻 (*Microcystis aeruginosa*))、支原体属物种、*Natranaerobius* spp. (如 *Natranaerobius thermophilus*)、奈瑟菌属物种 (如灰色奈瑟菌 (*Neisseria cinerea*)、脑膜炎奈瑟菌 (*Neisseria meningitidis*))、*Nitratifractor* spp.、亚硝化球菌属物种 (如嗜盐亚硝化球菌 (*Nitrosococcus halophilus*)、*Nitrosococcus watsoni*)、诺卡氏菌属物种 (如达松维尔拟诺卡氏菌 (*Nocardiopsis dassonvillei*))、节球藻属物种 (如产泡沫节球藻 (*Nodularia spumigena*))、念珠藻属物种、酒球菌属物种、颤藻属物种、副萨特氏菌属物种、细小棒菌属物种 (如食清洁剂细小棒菌 (*Parvibaculum lavamentivorans*))、巴斯德氏菌属物种 (如多杀巴斯德氏菌 (*Pasteurella multocida*))、*Pelotomaculum* spp. (如 *Pelotomaculum thermopropionicum*)、石孢菌属物种 (如 *Petrotoga mobilis*)、浮霉状菌属物种 (*Planctomyces* spp.)、极地单胞菌属物种 (如食萘极地单胞菌 (*Polaromonas naphthalenivorans*))、普雷沃氏菌属物种、假交替单胞菌属物种 (如游海假交替单胞菌 (*Pseudoalteromonas haloplanktis*))、罗尔斯顿氏菌属物种、瘤胃球菌属物种 (*Ruminococcus* spp.)、红螺菌属物种 (如深红红螺菌 (*Rhodospirillum rubrum*))、葡萄球菌属物种 (如金黄色葡萄球菌 (*Staphylococcus aureus*))、链球菌属物种 (如巴氏链球菌 (*Streptococcus pasteurianus*)、酿脓链球菌 (*Streptococcus pyogenes*)、嗜热链球菌 (*Streptococcus thermophilus*))、萨特氏菌属物种 (*Sutterella* spp.) (如华德萨特氏菌 (*Sutterella wadsworthensis*))、链霉菌属物种 (如始旋链霉菌 (*Streptomyces pristinaespiralis*))、绿色产色链霉菌 (*Streptomyces viridochromogenes*)、绿色产色链霉菌)、链孢囊菌属物种 (如玫瑰链孢囊菌 (*Streptosporangium roseum*)、玫瑰链孢囊菌)、聚球藻属物种、栖热腔菌属物种 (如非洲栖热腔菌 (*Thermosiphon africanus*))、密螺旋体属物种 (如齿垢密螺旋体 (*Treponema denticola*)) 和疣微菌门物种、沃林氏菌属物种 (如产琥珀酸沃林氏菌 (*Wolinella succinogenes*)) 和/或由基因组数据库的生物信息调查所描述的物种, 例如公开于 Makarova, Kira S. 等人, "An updated evolutionary classification of CRISPR-Cas systems." Nature Reviews Microbiology 13.11 (2015): 722 以及 Koonin, Eugene V., Kira S. Makarova 和 Feng Zhang, "Diversity, classification and evolution of CRISPR-Cas systems." Current opinion in microbiology 37 (2017): 67-78 的物种, 各论文在此通过引用以它们的整体并入本文。

[0013] 在一些实施方案中, 改造的 Cas9 蛋白可来自于酿脓链球菌。在一些实施方案中, 改

造的Cas9蛋白可来自于嗜热链球菌。在一些实施方案中,改造的Cas9蛋白可来自于脑膜炎奈瑟菌。在一些实施方案中,改造的Cas9蛋白可来自于黄色葡萄球菌。在一些实施方案中,改造的Cas9蛋白可来自于空肠弯曲杆菌。

[0014] 野生型Cas9蛋白包含两个核酸酶结构域,即RuvC和HNH结构域,每个结构域切割双链序列的一条链。Cas9蛋白也包含与向导RNA相互作用(如,REC1,REC2)或与RNA/DNA异源双链体相互作用(如,REC3)的REC结构域,以及包含和前间隔序列邻近基序(PAM)相互作用的结构域(即PAM相互作用结构域)。

[0015] Cas9蛋白可被改造为包含一种或多种修饰(即至少一个氨基酸的取代,至少一个氨基酸的缺失,至少一个氨基酸的插入)以使Cas9蛋白具有改变的活性、特异性和/或稳定性。

[0016] 例如,Cas9蛋白可通过一个或多个突变和/或缺失被改造,以使一个或两个核酸酶结构域失活。一个核酸酶结构域的失活产生切割双链序列的一条链的Cas9蛋白(即Cas9切口酶)。RuvC结构域可通过诸如D10A、D8A、E762A和/或D986A等突变而失活,而HNH结构域可通过诸如H840A、H559A、N854A、N856A和/或N863A等突变而失活(参考酿脓链球菌Cas9,SpyCas9的编号系统)。两个核酸酶结构域都失活产生没有切割活性的Cas9蛋白(即催化失活的或死Cas9)。

[0017] Cas9蛋白也可通过一个或多个氨基酸的取代、缺失和/或插入被改造,以具有改善的靶向特异性、改善的保真性、改变的PAM特异性、减少的脱靶效应和/或增加的稳定性。改善靶向特异性、改善保真性和/或减少脱靶效应的一个或多个突变的非限制性实例包括N497A、R661A、Q695A、K810A、K848A、K855A、Q926A、K1003A、R1060A和/或D1135E(参考SpyCas9的编号系统)。

[0018] 在替代的实施方案中,Cas蛋白可来自I型CRISPR/Cas系统。在一些实施方案中,Cas蛋白可为I型CRISPR/Cas系统中Cascade复合体的组分。例如,Cas蛋白可为Cas3蛋白。在一些实施方案中,Cas蛋白可来自III型CRISPR/Cas系统。在一些实施方案中,Cas蛋白可来自IV型CRISPR/Cas系统。在一些实施方案中,Cas蛋白可来自V型CRISPR/Cas系统。在一些实施方案中,Cas蛋白可来自VI型CRISPR/Cas系统。在一些实施方案中,Cas蛋白可具有RNA切割活性。在各种实施方案中,Cas蛋白可被分类为Cas9、Cas12a(又称Cpf1)、Cas12b、Cas12c、Cas12d、Cas12e(又称CasX)、Cas13a或Cas13b。

[0019] (i) 异源结构域

Cas9蛋白可被改造为含有至少一个异源结构域,即Cas9与一个或多个异源结构域融合。在其中两个或更多个异源结构域与Cas9融合的情况下,该两个或更多个异源结构域可为相同的,或它们可为不同的。该两个或更多个异源结构域可被融合至N末端、C末端、内部的位置或其组合。融合可通过化学键而是直接的,或者键合可通过一个或多个接头而是间接的。

[0020] 在某些优选的实施方案中,本文描述的改造的Cas9蛋白包含一个或多个核定位信号(NLS)。核定位信号的非限制性实例包括PKKKRKV(SEQ ID NO: 1)、PKKKRRV(SEQ ID NO: 2)、KRPAATKKAGQAKKK(SEQ ID NO: 3)、YGRKKRRQRRR(SEQ ID NO: 4)、RKKRRQRRR(SEQ ID NO: 5)、PAAKRVKLD(SEQ ID NO: 6)、RQRRNELKRS(SEQ ID NO: 7)、VSRKRPRP(SEQ ID NO: 8)、PPKKARED(SEQ ID NO: 9)、PQPKKKPL(SEQ ID NO: 10)、SALIKKKKKMAP(SEQ ID NO:

11)、PKQKKRK (SEQ ID NO: 12)、RKLKKKIKKL (SEQ ID NO: 13)、REKKKFLKRR (SEQ ID NO: 14)、KRKGDEVGVDEVAKKKSKK (SEQ ID NO: 15)、RKCLQAGMNLEARTKK (SEQ ID NO: 16)、NQS SNFGPMKGGNFGRSSGPYGGGGQYFAKPRNQGGY (SEQ ID NO: 17) 和 RMRIZFKNKGKDTELRRRRVEVS VELRKAKKDEQILKRRNV (SEQ ID NO: 18)。

[0021] 在一个特定实施方案中,核定位信号选自PKKKRKV (SEQ ID NO: 1) 和PAAKRVKLD (SEQ ID NO: 6)。在另一个特定实施方案中,改造的Cas9蛋白包含PKKKRKV (SEQ ID NO: 1) 和PAAKRVKLD (SEQ ID NO: 6)两者。在另一个特定实施方案中,改造的Cas9蛋白包含至少两个PKKKRKV (SEQ ID NO: 1) 和至少一个PAAKRVKLD (SEQ ID NO: 6)。在另一个特定实施方案中,改造的Cas9蛋白包含两个PKKKRKV (SEQ ID NO: 1) 和一个PAAKRVKLD (SEQ ID NO: 6)。

[0022] 在这些以及其它优选实施方案中,改造的Cas9蛋白包含一个或多个标记物结构域。标记物结构域包括荧光蛋白和纯化或表位标签。适合的荧光蛋白包括但不限于绿色荧光蛋白(如GFP、eGFP、GFP-2、tagGFP、turboGFP、Emerald、Azami Green、Monomeric Azami Green、CopGFP、AceGFP、ZsGreen1)、黄色荧光蛋白(如YFP、EYFP、Citrine、Venus、YPet、PhiYFP、ZsYellow1)、蓝色荧光蛋白(如BFP、EBFP、EBFP2、Azurite、mKalama1、GFPuv、Sapphire、T-sapphire)、青色荧光蛋白(如ECFP、Cerulean、CyPet、AmCyan1、Midoriishi-Cyan)、红色荧光蛋白(如mKate、mKate2、mPlum、DsRed单体、mCherry、mRFP1、DsRed-Express、DsRed2、DsRed-Monomer、HcRed-Tandem、HcRed1、AsRed2、eqFP611、mRaspberry、mStrawberry、Jred)、橙色荧光蛋白(如mOrange、mKO、Kusabira-Orange、Monomeric Kusabira-Orange、mTangerine、tdTomato)或其组合。标记物结构域可包含一种或多种荧光蛋白(如Suntag)的串联重复。

[0023] 在一个实施方案中,标记蛋白选自以下序列:

标记蛋白序列
MVKGEELFTGVVPILVLDGDVNGHKFSVSGEGEGDATYGKLTGKFICTTGKLPVP WPTLVTTLYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEV KFEGLDTLVNRIELKGIDFKEDGNILGHKLEYNNNSHNVYIMADKQKNGIKVNFKIRH NIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSKLSKDPNEKRDHMVLLEFVTA AGITLGMDELYK(SEQ ID NO: 19)
MVKGEAVIKEFMRFKVHMEGSMNGHEFEIEGEGRPYEGTQAKLKVTKGGPLP FSWDILSPQFMYGSRAFTKHPADIPDYYKQSFPEGFKWERVMNFEDGGAVTVTQDT SLEDGTLIYKVKLRGTFPPDGPMQKKTMGWEASTERLYPEDGVLKGDIKMALRL KGGRYLADEFKTTYKAKKPVQMPGAYNVDRKLDITSHNEDYTVVVEQYERSEGRHS TGGMDELYK(SEQ ID NO: 20)

适合的纯化或表位标签的非限制性实例包括6xHis (SEQ ID NO: 22), FLAG[®] (如SEQ ID NO: 21)、HA、GST、Myc、SAM等等。促进CRISPR复合体检出或富集的异源融合物的非限制性实例包括链酶亲和素(Kipriyanov等人, Human Antibodies, 1995, 6 (3): 93-101)、亲和素(Airenne等人, Biomolecular Engineering, 1999, 16 (1-4): 87-92)、亲和素的单体形式(Laitinen等人, Journal of Biological Chemistry, 2003, 278 (6): 4010-4014)、在重组体生产过程中促进生物素化的肽标签(Cull等人, Methods in Enzymology, 2000,

326:430-440)。

[0024] 在各种实施方案中,除了核定位信号和标记蛋白,改造的Cas9蛋白也可包含一个或多个异源结构域,如细胞穿透结构域、标记物结构域、染色质破坏结构域、表观遗传修饰结构域(如胞苷脱氨酶结构域、组蛋白乙酰转移酶结构域等等)、转录调节结构域、RNA适体结合结构域或非Cas9核酸酶结构域。

[0025] 在一些实施方案中,一个或多个异源结构域可为细胞穿透结构域。适合的细胞穿透结构域的实例包括但不限于GRKKRRQRRPPQPKKKRKV (SEQ ID NO: 23)、PLSSIIFSRIGDPPKKKRKV (SEQ ID NO: 24)、GALFLGWLGAAGSTMAGPKKKRKV (SEQ ID NO: 25)、GALFLGFLGAAGSTMAGSQPKKKRKV (SEQ ID NO: 26)、KETWWETWWTEWSQPKKKRKV (SEQ ID NO: 27)、YARAAARQARA (SEQ ID NO: 28)、THRLPRRRRR (SEQ ID NO: 29)、GGRRARRRRR (SEQ ID NO: 30)、RRQRRTSKLMKR (SEQ ID NO: 31)、GTWTLSAGYLLGKINLKALAALAKKIL (SEQ ID NO: 32)、KALAWEAKLAKALAKALAKHLAKALAKALKCEA (SEQ ID NO: 33) 和RQIKIWFQNRRMKWKK (SEQ ID NO: 34)。

[0026] 在另外其它的实施方案中,一个或多个异源结构域可为染色质调控基序(CMM)。CMM的非限制性实例包括与核小体相互作用的衍生自高迁移率族(HMG)蛋白(如HMGB1、HMGB2、HMGB3、HMGN1、HMGN2、HMGN3a、HMGN3b、HMGN4和HMGN5蛋白)的肽,组蛋白H1变体(如组蛋白H1.0、H1.1、H1.2、H1.3、H1.4、H1.5、H1.6、H1.7、H1.8、H1.9和H1.10)的中央球状结构域,或染色质重塑复合体(如SWI/SNF(转换/蔗糖不发酵)、ISWI(模拟转换(Imitation SWItc)h)、CHD(克罗莫结构域-解旋酶-DNA结合)、Mi-2/NuRD(核小体重塑和脱乙酰酶)、IN080、SWR1和RSC复合体)的DNA结合结构域。在其它实施方案中,CMM也可衍生自拓扑异构酶、解旋酶或病毒蛋白。CMM的来源可以且将会不同。CMM可来自人类、动物(即脊椎动物和无脊椎动物)、植物、藻类或酵母。特异性CMM的非限制性实例列于下表中。本领域技术人员可容易地鉴定其它物种中的同源物和/或其中的相关融合基序。

蛋白	登录号	融合基序
人HMGN1	P05114	全长
人HMGN2	P05204	全长
人HMGN3a	Q15651	全长
人HMGN3b	Q15651-2	全长
人HMGN4	000479	全长
人HMGN5	P82970	核小体结合基序
人HMGB1	P09429	框A
人组蛋白H1.0	P07305	球状基序
人组蛋白H1.2	P16403	球状基序
人CHD1	O14646	DNA结合基序
酵母CHD1	P32657	DNA结合基序
酵母ISWI	P38144	DNA结合基序
人TOP1	P11387	DNA结合基序
人疱疹病毒8 LANA	J9QSF0	核小体结合基序
人CMV IE1	P13202	染色质栓系基序

麻风分枝杆菌 (<i>M. leprae</i>) DNA解旋酶	P40832	HhH结合基序
------------------------------------	--------	---------

[0027] 在另外其它实施方案中,一个或多个异源结构域可为表观遗传修饰结构域。适合的表观遗传修饰结构域的非限制性实例包括那些具有DNA脱氨基化(如胞昔脱氨酶、腺昔脱氨酶、鸟昔脱氨酶)、DNA甲基转移酶活性(如胞嘧啶甲基转移酶)、DNA脱甲基酶活性、DNA氨基化、DNA氧化活性、DNA解旋酶活性、组蛋白乙酰转移酶(HAT)活性(如衍生自E1A结合蛋白p300的HAT结构域)、组蛋白脱乙酰酶活性、组蛋白甲基转移酶活性、组蛋白脱甲基酶活性、组蛋白激酶活性、组蛋白磷酸酶活性、组蛋白泛素连接酶活性、组蛋白去泛素化活性、组蛋白腺昔化活性、组蛋白脱腺昔化活性、组蛋白SUMO化活性、组蛋白去SUMO化活性、组蛋白核糖基化活性、组蛋白脱核糖基化活性、组蛋白豆蔻酰化活性、组蛋白脱豆蔻酰化活性、组蛋白瓜氨酸化活性、组蛋白烷基化活性、组蛋白脱烷基化活性或组蛋白氧化活性的结构域。在特定实施方案中,表观遗传修饰结构域可包含胞昔脱氨酶活性、腺昔脱氨酶活性、组蛋白乙酰转移酶活性或DNA甲基转移酶活性。

[0028] 在其它实施方案中,一个或多个异源结构域可为转录调节结构域(即转录激活结构域或转录阻遏物结构域)。适合的转录激活结构域包括但不限于单纯疱疹病毒VP16结构域、VP64(即VP16的四个串联拷贝)、VP160(即VP16的十个串联拷贝)、NF κ B p65激活结构域(p65)、EB病毒(Epstein-Barr virus)R反式作用因子(Rta)结构域、VPR(即VP64+p65+Rta)、p300依赖的转录激活结构域、p53激活结构域1和2、热休克因子1(HSF1)激活结构域、Smad4激活结构域(SAD)、cAMP响应元件结合蛋白(CREB)激活结构域、E2A激活结构域、活化的T细胞的核因子(NFAT)激活结构域或其组合。适合的转录阻遏物结构域的非限制性实例包括Kruppel关联框(KRAB)阻遏物结构域、Mxi阻遏物结构域、可诱导cAMP早期阻遏物(ICER)结构域、YY1富甘氨酸阻遏物结构域、Sp1样阻遏物、E(spl)阻遏物、I κ B阻遏物、Sin3阻遏物、甲基CpG结合蛋白2(MeCP2)阻遏物或其组合。转录激活或转录阻遏物结构域可遗传融合至Cas9蛋白,或通过非共价的蛋白-蛋白、蛋白-RNA或蛋白-DNA相互作用结合。

[0029] 在进一步的实施方案中,一个或多个异源结构域可为RNA适体结合结构域(Konermann等人,Nature, 2015, 517 (7536):583-588; Zalatan等人,Cell, 2015, 160 (1-2):339-50)。适合的RNA适体蛋白结构域的实例包括MS2外壳蛋白(MCP),PP7噬菌体外壳蛋白(PCP), μ 噬菌体Com蛋白, λ 噬菌体N22蛋白,茎环结合蛋白(SLBP),脆性X智力迟滞综合征相关蛋白1(FXR1),衍生自噬菌体如AP205、BZ13、f1、f2、fd、fr、ID2、JP34/GA、JP501、JP34、JP500、KU1、M11、M12、MX1、NL95、PP7、 φ Cb5、 φ Cb8r、 φ Cb12r、 φ Cb23r、Q β 、R17、SP- β 、TW18、TW19和VK的蛋白,其片段或其衍生物。

[0030] 在另外其它的实施方案中,一个或多个异源结构域可为非Cas9核酸酶结构域。适合的核酸酶结构域可获取自任意核酸内切酶或核酸外切酶。可由其衍生核酸酶结构域的核酸内切酶的非限制性实例包括但不限于限制性核酸内切酶和归巢核酸内切酶。在一些实施方案中,核酸酶结构域可衍生自II-S型限制性核酸内切酶。II-S型核酸内切酶在通常离识别/结合位点若干碱基对的位点切割DNA,并且因此,拥有可分开的结合和切割结构域。这些酶通常是单体,它们短暂地缔合形成二聚体,以在交错的位置处切割DNA的每条链。适合的II-S型核酸内切酶的非限制性实例包括Bfi I、Bpm I、Bsa I、Bsg I、Bsm BI、Bsm I、Bsp MI、Fok I、Mbo II和Sap I。在一些实施方案中,核酸酶结构域可为Fok I核酸酶结构域或其衍生物。可修饰II-S型核酸酶结构域以促进两个不同的核酸酶结构域的二聚体化。例如,可通过使某些

氨基酸残基突变而修饰FokI的切割结构域。作为非限制性实例,在FokI核酸酶结构域446、447、479、483、484、486、487、490、491、496、498、499、500、531、534、537和538位置的氨基酸残基是修饰的目标。在特定的实施方案中,FokI核酸酶结构域可包含第一FokI半结构域(包含Q486E、I499L和/或N496D突变)以及第二FokI半结构域(包含E490K、I538K和/或H537R突变)。

[0031] 一个或多个异源结构域可通过一个或多个化学键(如共价键)直接连接到Cas9蛋白,或者该一个或多个异源结构域可通过一个或多个接头间接连接到Cas9蛋白。

[0032] 接头是通过至少一个共价键连接一个或多个其它化学基团的化学基团。适合的接头包括氨基酸、肽、核苷酸、核酸、有机接头分子(如马来酰亚胺衍生物、N-乙氧基苄基咪唑、联苯-3,4',5-三羧酸、对氨基苄基羧基等等)、二硫化物接头和聚合体接头(如PEG)。接头可包含一个或多个间隔基团,其包括但不限于亚烷基、亚烯基、亚炔基、烷基、烯基、炔基、烷氧基、芳基、杂芳基、芳烷基、芳烯基、芳炔基等等。接头可为中性的,或带有正或负电荷。此外,接头可为可切割的,使得在一定条件下将接头连接至另一化学基团的接头的共价键可被断裂或切割,所述条件包括pH、温度、盐浓度、光、催化剂或酶。在一些实施方案中,接头可为肽接头。肽接头可为柔性氨基酸接头(如包含小的非极性或极性氨基酸)。

[0033] 在一个特定的实施方案中,接头选自以下序列:

接头蛋白序列
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAKA(SEQ ID NO: 35)
SGGSSGGSSGSETPGTSESATPESSGGSSGGS(SEQ ID NO: 36)

柔性接头其它非限制性实例包括LEGGGS (SEQ ID NO: 37)、TGSG (SEQ ID NO: 38)、GGGGGSG (SEQ ID NO: 39)、(GGGGS)₁₋₄ (SEQ ID NO: 40) 和(Gly)₆₋₈ (SEQ ID NO: 41)。或者,肽接头可为刚性氨基酸接头。此类接头包括(EAAAK)₁₋₄ (SEQ ID NO: 42)、A(EAAAK)₂₋₅A (SEQ ID NO: 43)、PAPAP (SEQ ID NO: 44) 和(AP)₆₋₈ (SEQ ID NO: 45)。适合的接头的另外实例为本领域所熟知,并且设计接头的程序是可容易地获得的(Crasto等人, Protein Eng., 2000, 13 (5):309-312)。

[0034] 在一些实施方案中,改造的Cas9蛋白可以重组地产生于无细胞系统、细菌细胞或真核细胞中,且可使用标准纯化方法纯化。在其它实施方案中,改造的Cas9蛋白从编码改造的Cas9蛋白的核酸中在目标真核细胞中体内产生(见下文部分(II))。

[0035] 在其中改造的Cas9蛋白包含核酸酶或切口酶活性的实施方案中,改造的Cas9蛋白可进一步包含至少一个细胞穿透结构域,以及至少一个染色质破坏结构域。在其中改造的Cas9蛋白与表观遗传修饰结构域连接的实施方案中,改造的Cas9蛋白可进一步包含至少一个细胞穿透结构域,以及至少一个染色质破坏结构域。此外,在其中改造的Cas9蛋白与转录调节结构域连接的实施方案中,改造的Cas9蛋白可进一步包含至少一个细胞穿透结构域,以及至少一个染色质破坏结构域和/或至少一个RNA适体结合结构域。

[0036] 各种融合蛋白组分可从N末端至C末端以任意顺序组合。例如,其中A代表标记蛋白,B代表核定位信号和C代表Cas9蛋白,融合蛋白从N末端到C末端可按照以下方式排列:A-B-C;A-C-B;B-A-C;B-C-A;C-A-B或C-B-A,其中接头(“-L-”)可布置于任意两项之间(如A-L-

B-C;A-B-L-C;A-L-B-L-C等等)。

[0037] (b) 改造的向导RNA

改造的向导RNA被设计用于与特异性改造的Cas9蛋白复合。向导RNA包含(i)包含在5'端与目标序列杂交的向导序列的CRISPR RNA (crRNA) 以及(ii)募集Cas9蛋白的反式作用crRNA (tracrRNA) 序列。每个向导RNA的crRNA向导序列是不同的(即是序列特异的)。tracrRNA序列在被设计用于与来自特定细菌物种的Cas9蛋白复合的向导RNA中通常是相同的。

[0038] crRNA向导序列被设计用于与双链序列中的目标序列(即前间隔序列)杂交。一般来说,crRNA和目标序列之间的互补性为至少80%、至少85%、至少90%、至少95%或至少99%。在特定实施方案中,该互补性是完全的(即100%)。在各种实施方案中,crRNA向导序列的长度可在约15个核苷酸至约25个核苷酸范围。例如,crRNA向导序列的长度可为约15、16、17、18、19、20、21、22、23、24或25个核苷酸。在特定实施方案中,crRNA的长度为约19、20或21个核苷酸。在一种实施方案中,crRNA向导序列具有20个核苷酸的长度。

[0039] 向导RNA包含形成至少一个与Cas9蛋白相互作用的茎环结构的重复序列,以及保持单链的3'序列。每个环和茎的长度可以不同。例如,环的长度可在约3个至约10个核苷酸范围,而茎的长度可在约6个至约20个碱基对范围。茎可包含一个或多个1至10个核苷酸的凸起。单链3'区域的长度可以不同。改造的向导RNA中的tracrRNA序列通常基于在目标细菌物种中野生型tracrRNA的编码序列。野生型序列可被修饰以促进二级结构的形成、增加二级结构的稳定性、促进真核细胞中的表达等等。例如,一个或多个核苷酸的改变可被引入向导RNA编码序列(见下文实施例3)。tracrRNA序列的长度可在约50个核苷酸至约300个核苷酸范围。在各种实施方案中,tracrRNA的长度可在约50个至约90个核苷酸、在约90个至约110个核苷酸、在约110个至约130个核苷酸、在约130个至约150个核苷酸、在约150个至约170个核苷酸、在约170个至约200个核苷酸、在约200个至约250个核苷酸或在约250个至约300个核苷酸范围。

[0040] 通常,改造的向导RNA是单个分子(即单个向导RNA或sgRNA),其中crRNA序列连接至tracrRNA序列。然而,在一些实施方案中,改造的向导RNA可为两个分开的分子。第一分子包含含有能够与第二分子的5'端碱基配对的3'序列(包含约6个至约20个核苷酸)的crRNA,其中第二分子包含含有能与第一分子3'端碱基配对的5'序列(包含约6个至约20个核苷酸)的tracrRNA。

[0041] 在一些实施方案中,改造的向导RNA的tracrRNA序列可被修饰以包含一个或多个适体序列(Konermann等人,Nature,2015,517(7536):583-588;Zalatan等人,Cell,2015,160(1-2):339-50)。适合的适体序列包含与衔接蛋白结合的那些,所述衔接蛋白选自MCP、PCP、Com、SLBP、FXR1、AP205、BZ13、f1、f2、fd、fr、ID2、JP34/GA、JP501、JP34、JP500、KU1、M11、M12、MX1、NL95、PP7、 φ Cb5、 φ Cb8r、 φ Cb12r、 φ Cb23r、Q β 、R17、SP- β 、TW18、TW19、VK、其片段或其衍生物。本领域技术人员理解适体序列的长度可以不同。

[0042] 在其它实施方案中,向导RNA可进一步包含至少一个可检测标记。可检测标记可为荧光团(如FAM、TMR、Cy3、Cy5、Texas Red、Oregon Green、Alexa Fluors、Halo标签或适合的荧光染料)、检测标签(如生物素、地高辛等)、量子点或金颗粒。

[0043] 向导RNA可包含标准核糖核苷酸和/或修饰的核糖核苷酸。在一些实施方案中,向

导RNA可包含标准或修饰的脱氧核糖核苷酸。在其中酶促合成(即在体内或在体外)向导RNA的实施方案中,向导RNA通常包含标准核糖核苷酸。在其中化学合成向导RNA的实施方案中,向导RNA可包含标准或修饰的核糖核苷酸和/或脱氧核糖核苷酸。修饰的核糖核苷酸和/或脱氧核苷酸包含碱基修饰(如假尿苷、2'-硫代尿苷、N6-甲基腺苷等)和/或糖修饰(如2'-0-甲基、2'-氟代、2'-氨基、锁核酸(LNA)等)。向导RNA的主链也可被修饰以包含硫代磷酸酯键合、硼烷磷酸酯键合或肽核酸。

[0044] (c) PAM序列

在一些实施方案中,目标序列可能邻近前间隔序列邻近基序(PAM)(被CRISPR/Cas9复合体识别的短序列)。在一些实施方案中,PAM可能邻近目标序列3'端或在其1、2、3或4个核苷酸之内。PAM的长度和序列可取决于使用的Cas9蛋白。例如,PAM可选自共有序列或特定Cas9蛋白或Cas9直向同源物的特定PAM序列,包括公开于在此通过引用并入的Ran等人,Nature, 520: 186-191 (2015)图1的那些。在一些实施方案中,PAM可包含2、3、4、5、6、7、8、9或10个核苷酸的长度。非限制性的示例性PAM序列包括NGG、NGGNG、NG、NAAAAN、NNAAAAW、NNNNACAA、GNNNCNNA和NNNNNGATT(其中N被定义为任意核苷酸,W被定义为A或T之一)。在一些实施方案中,PAM序列可为NGG。在一些实施方案中,PAM序列可为NGGNG。在一些实施方案中,PAM序列可为NNAAAAAW。

[0045] 可以理解,不同CRISPR蛋白识别不同PAM序列。例如,用于Cas9蛋白的PAM序列包括5'-NGG、5'-NGGNG、5'-NNAGAAW、5'-NNNNGATT、5'-NNNNRYAC、5'-NNNNCAAA、5'-NGAAA、5'-NNAAT、5'-NNNRTA、5'-NNGG、5'-NNNRTA、5'-MMACCA、5'-NNNNGRY、5'-NRGNK、5'-GGGRG、5'-NNAMMMC和5'-NNG,而用于Cas12a蛋白的PAM序列包括5'-TTN和5'-TTTV,其中N定义为任意核苷酸,R定义为G或A,W定义为A或T,Y定义为C或T,V定义为A、C或G。通常,Cas9 PAM位于目标序列的3',而Cas12a PAM位于目标序列的5'。各种PAM序列和识别它们的CRISPR蛋白为本领域所知,如美国专利申请公开2019/0249200;Leenay, Ryan T.等人,"Identifying and visualizing functional PAM diversity across CRISPR-Cas systems." Molecular cell 62.1 (2016): 137-147;以及Kleinstiver, Benjamin P.等人,"Engineered CRISPR-Cas9 nucleases with altered PAM specificities." Nature 523.7561 (2015): 481,其各自在此通过引用以它们的整体并入。

[0046] 附加或替代地,用于本文公开的各改造的Cas9系统的PAM在下表中呈现。

PAM 序列	
改造的 Cas9 系统	PAM(5'-3')*
史氏芽孢杆菌(<i>Bacillus smithii</i>) Cas9 (BsmCas9)	NNNNCAAA
鼠李糖乳杆菌(<i>Lactobacillus rhamnosus</i>) Cas9 (LrhCas9)	NGAAA
<i>Parasutterella excrementihominis</i> Cas9 (PexCas9)	NGG
犬支原体(<i>Mycoplasma canis</i>) Cas9 (McaCas9)	NNGG
鸡败血支原体(<i>Mycoplasma gallisepticum</i>) Cas9 (MgaCas9)	NNAAT
<i>Akkermansia glyciphila</i> Cas9 (AglCas9)	NNNRTA
嗜粘蛋白艾克曼氏菌(<i>Akkermansia muciniphila</i>) Cas9 (AmuCas9)	MMACCA
北原酒球菌(<i>Oenococcus kitaharae</i>) Cas9 (OkiCas9)	NNG
孟买双歧杆菌(<i>Bifidobacterium bombi</i>) Cas9 (BboCas9)	NNNNGRY
解纤维热酸菌(<i>Acidothermus cellulolyticus</i>) Cas9 (AccCas9)	NGG
神话脂环酸芽孢杆菌(<i>Alicyclobacillus hesperidum</i>) Cas9 (AheCas9)	NGG
产琥珀酸沃林氏菌 Cas9 (WsuCas9)	NGG
<i>Nitratifractor salsuginis</i> Cas9 (NsCas9)	NRGNK
<i>Ralstonia syzygii</i> Cas9 (RsyCas9)	GGGRG
白喉棒杆菌 Cas9 (CdiCas9)	NNAMMMC

*K为G或T; M为A或C; R为A或G; Y为C或T; N为A、C、G或T。

[0047] 例如,参见美国专利申请公开号2019/0249200(在此通过引用以其整体并入本文)。

[0048] (II) 核酸

本公开内容的进一步方面提供了编码上文部分(I)所述的改造的Cas9系统的核酸。所述系统可被单个核酸或多个核酸编码。所述核酸可为DNA或RNA,线性的或环状的,单链的或双链的。所述RNA或DNA可为密码子优化的,以在目标真核细胞中有效翻译成蛋白。密码子优化程序可作为免费软件获得,或可由商业来源获得。

[0049] 在一些实施方案中,核酸编码具有与SEQ ID NO:48、49或50的氨基酸序列有至少约75%、至少约80%、至少约85%、至少约90%、至少约95%或至少约99%的序列同一性的蛋白。在某些实施方案中,编码改造的Cas9蛋白的核酸可与SEQ ID NO:48、49或50的DNA序列有至少约75%、至少约80%、至少约85%、至少约90%、至少约95%或至少约99%的序列同一性。在某些实施方案中,编码改造的Cas9蛋白的DNA具有SEQ ID NO:48、49或50的DNA序列。在另外的实施方案中,核酸编码具有与SEQ ID NO:48、49或50的氨基酸序列有至少约75%、至少约80%、至少约85%、至少约90%、至少约95%或至少约99%的序列同一性的蛋白。

[0050] 在一些实施方案中,编码改造的Cas9蛋白的核酸可为RNA。该RNA可为体外酶促合成的。对此,编码改造的Cas9蛋白的DNA可与被噬菌体RNA聚合酶识别的启动子序列可操作地连接,用于体外RNA合成。例如,启动子序列可为T7、T3或SP6启动子序列,或T7、T3或SP6启动子序列的变种。编码改造的蛋白的DNA可为载体的一部分,如下文所详述。在此类实施方案中,体外转录的RNA可被纯化、加帽和/或多腺苷酸化。在其它实施方案中,编码改造的Cas9蛋白的RNA可为自复制RNA的一部分(Yoshioka等人,Cell Stem Cell, 2013, 13:246-

254)。自复制RNA可衍生自非传染性的、自复制的委内瑞拉马脑炎(VEE)病毒RNA复制子,其为能在有限次细胞分裂中自复制的正义单链RNA,且其可被修饰以编码目标蛋白(Yoshioka等人,Cell Stem Cell, 2013, 13:246-254)。

[0051] 在其它实施方案中,编码改造的Cas9蛋白的核酸可为DNA。DNA编码序列可与至少一个启动子控制序列可操作地连接,用于在目标细胞中表达。在某些实施方案中,DNA编码序列可与启动子序列可操作地连接,用于改造的Cas9蛋白在细菌(如大肠杆菌(*E. coli*))细胞或真核(如酵母、昆虫或哺乳动物)细胞中表达。适合的细菌启动子包括但不限于T7启动子、*lac*操纵子启动子、*trp*启动子、*tac*启动子(其为*trp*和*lac*启动子的杂合体)、任意前述启动子的变种以及任意前述启动子的组合。适合的真核启动子非限制性实例包括组成型、调节型、或细胞或组织特异性启动子。适合的真核组成型启动子控制序列包括但不限于巨细胞病毒立即早期启动子(CMV)、猿猴病毒(SV40)启动子、腺病毒的主要晚期启动子、劳氏肉瘤病毒(RSV)启动子、小鼠乳腺瘤病毒(MMTV)启动子、磷酸甘油酸激酶(PGK)启动子、延伸因子(ED1)- α 启动子、泛素启动子、肌动蛋白启动子、微管蛋白启动子、免疫球蛋白启动子、其片段或任意前述启动子的组合。适合的真核调节型启动子控制序列的实例包括但不限于受热休克、金属、类固醇、抗生素或醇调节的那些。组织特异性启动子的非限制性实例包括B29启动子、CD14启动子、CD43启动子、CD45启动子、CD68启动子、结蛋白启动子、弹性蛋白酶-1启动子、内皮糖蛋白启动子、纤连蛋白启动子、F1t-1启动子、GFAP启动子、GPIIb启动子、ICAM-2启动子、INF- β 启动子、Mb启动子、NphsI启动子、OG-2启动子、SP-B启动子、SYN1启动子和WASP启动子。启动子序列可为野生型的,或它可被修饰用于更高效或有效的表达。在一些实施方案中,DNA编码序列也可连接至多腺苷酸化信号(如SV40多聚A信号、牛生长激素(BGH)多聚A信号等)和/或至少一个转录终止序列。在一些情况下,改造的Cas9蛋白可从细菌或真核细胞中纯化。

[0052] 在另外其它实施方案中,改造的向导RNA可由DNA编码。在某些情况下,编码改造的向导RNA的DNA可与由噬菌体RNA聚合酶识别的启动子序列可操作地连接,用于体外RNA合成。例如,启动子序列可为T7、T3或SP6启动子序列,或为T7、T3或SP6启动子序列的变种。在其它情况下,编码改造的向导RNA的DNA可与被RNA聚合酶III(Pol III)识别的启动子序列可操作地连接,用于在目标真核细胞中的表达。适合的Pol III启动子的实例包括但不限于哺乳动物U6、U3、H1和7SL RNA启动子。

[0053] 在各种实施方案中,编码改造的Cas9蛋白的核酸可存在于载体中。在一些实施方案中,所述载体可进一步包含编码改造的向导RNA的核酸。适合的载体包括质粒载体、病毒载体和自复制RNA(Yoshioka等人,Cell Stem Cell, 2013, 13:246-254)。在一些实施方案中,编码复合体或融合蛋白的核酸可存在于质粒载体中。适合的质粒载体的非限制性实例包括pUC、pBR322、pET、pBluescript和其变体。在其它实施方案中,编码复合体或融合蛋白的核酸可为病毒载体(如慢病毒载体、腺相关病毒载体、腺病毒载体等)的一部分。质粒或病毒载体可包含额外的表达控制序列(如增强子序列、Kozak序列、多腺苷酸化序列、转录终止序列等)、可选的标记物序列(如抗生素抗性基因)、复制起点等。有关载体和其用途的额外信息可在“Current Protocols in Molecular Biology”, Ausubel等人, John Wiley & Sons, New York, 2003或“Molecular Cloning: A Laboratory Manual” Sambrook & Russell, Cold Spring Harbor Press, Cold Spring Harbor, NY, 第3版, 2001中找到。

[0054] (III) 真核细胞

本公开内容的另一个方面包含真核细胞，其包含至少一种如上文部分(I)详述的改造的Cas9系统和/或至少一种如上文部分(II)详述的编码改造的Cas9蛋白和/或改造的向导RNA的核酸。

[0055] 真核细胞可为人类细胞、非人类哺乳动物细胞、非哺乳类脊椎动物细胞、无脊椎动物细胞、植物细胞或单细胞真核生物。适合的真核细胞实例在下文(IV)(c)部分详述。真核细胞可为体外的、离体的或体内的。

[0056] 作为实例，在一些实施方案中，真核细胞，或真核细胞群，是T细胞、CD8⁺ T细胞、CD8⁺幼稚T细胞、中央记忆性T细胞、效应记忆性T细胞、CD4⁺ T细胞、干细胞记忆性T细胞、辅助性T细胞、调节性T细胞、细胞毒性T细胞、自然杀伤T细胞、造血干细胞、长期造血干细胞、短期造血干细胞、多能祖细胞、谱系限制型祖细胞、淋巴祖细胞、胰腺祖细胞、内分泌祖细胞、外分泌祖细胞、骨髓祖细胞、共同髓系祖细胞、红系祖细胞、巨核红系祖细胞、单核前体细胞、内分泌前体细胞、外分泌细胞、成纤维细胞、肝母细胞、成肌细胞、巨噬细胞、胰岛β细胞、心肌细胞、血细胞、导管细胞、腺泡细胞、α细胞、β细胞、δ细胞、PP细胞、胆管上皮细胞、视网膜细胞、光感受器细胞、视杆细胞、视锥细胞、视网膜色素上皮细胞、小梁网细胞、耳蜗毛细胞、外毛细胞、内毛细胞、肺上皮细胞、支气管上皮细胞、肺泡上皮细胞、肺上皮祖细胞、横纹肌细胞、心脏肌肉细胞、肌卫星细胞、肌细胞、神经元、神经元干细胞、间充质干细胞、诱导多能干(iPS)细胞、胚胎干细胞、单核细胞、巨核细胞、中性粒细胞、嗜酸性粒细胞、嗜碱性粒细胞、肥大细胞、网织红细胞、B细胞(如祖B细胞、前B细胞、原B细胞、记忆B细胞、浆B细胞)、胃肠道上皮细胞、胆管上皮细胞、胰腺导管上皮细胞、肠道干细胞、肝细胞、肝星状细胞、肝巨噬细胞、成骨细胞、破骨细胞、脂肪细胞(如棕色脂肪细胞或白色脂肪细胞)、前脂肪细胞、胰腺前体细胞、胰岛细胞、胰腺β细胞、胰腺α细胞、胰腺δ细胞、胰腺外分泌细胞、施旺细胞或少突胶质细胞，或是此类细胞的群。

[0057] (IV) 修饰染色体序列的方法

本公开内容的进一步方面包含用于修饰真核细胞中染色体序列的方法。一般来说，该方法包含向目标真核细胞引入至少一种如在上文部分(I)详述的改造的Cas9系统和/或至少一种如在上文部分(II)详述的编码所述改造的Cas9系统的核酸。

[0058] 在其中改造的Cas9蛋白包含核酸酶或切口酶活性的实施方案中，染色体序列修饰可包含至少一个核苷酸的取代、至少一个核苷酸的缺失、至少一个核苷酸的插入。在一些迭代中，该方法包含向真核细胞内引入一种包含核酸酶活性的改造的Cas9系统或两种包含切口酶活性的改造的Cas9系统且无供体多核苷酸，使得一种或多种改造的Cas9系统在染色体序列的目标位点中引入双链断裂，且通过细胞DNA修复过程的双链断裂修复引入至少一个核苷酸改变(即插入缺失)，从而使染色体序列失活(即基因敲除)。在其它迭代中，该方法包含向真核细胞内引入一种包含核酸酶活性的改造的Cas9系统或两种包含切口酶活性的改造的Cas9系统及供体多核苷酸，使得一种或多种改造的Cas9系统在染色体序列的目标位点中引入双链断裂，且通过细胞DNA修复过程的双链断裂修复导致在供体多核苷酸中的序列插入或交换进入染色体序列的目标位点(即基因修正或基因敲入)。

[0059] 在其中改造的Cas9蛋白包含表观遗传修饰活性或转录调节活性的实施方案中，染色体序列修饰可包含在染色体序列上在目标位点中或附近至少一个核苷酸的转换、在目标

位点中或附近至少一个核苷酸的修饰、在目标位点中或附近至少一个组蛋白的修饰和/或在目标位点中或附近的转录改变。

[0060] (a) 引入细胞内

如上文所提及,该方法包含向真核细胞内引入至少一种改造的Cas9系统和/或编码所述系统的核酸(以及任选的供体多核苷酸)。该至少一种系统和/或核酸/供体多核苷酸可通过多种方法被引入目标细胞。

[0061] 在一些实施方案中,可用适当的分子(即蛋白、DNA和/或RNA)转染细胞。适合的转染方法包括核转染(或电穿孔)、磷酸钙介导的转染、阳离子聚合物转染(如DEAE-葡聚糖或聚乙烯亚胺)、病毒转导、病毒体转染、病毒粒子转染、脂质体转染、阳离子脂质体转染、免疫脂质体转染、非脂质体脂质转染、树形分子转染、热休克转染、磁转染、脂转染、基因枪递送、刺穿转染(imperfection)、声致穿孔、光学转染(optical transfection)和有专利的试剂增强的核酸摄取。转染方法为本领域所熟知(例如参见“Current Protocols in Molecular Biology”,Ausubel等人,John Wiley & Sons,New York,2003或“Molecular Cloning: A Laboratory Manual”,Sambrook & Russell,Cold Spring Harbor Press,Cold Spring Harbor,NY,第3版,2001)。在其它实施方案中,分子可通过显微注射引入细胞。例如,分子可注射进目标细胞的细胞质或细胞核。每种分子引入细胞的量可以不同,但本领域的技术人员熟悉确定适当用量的方法。

[0062] 各种分子可同时或序贯地引入细胞内。例如,改造的Cas9系统(或它的编码核酸)和供体多核苷酸可同时被引入。或者,可将一种首先引入细胞内,然后可将另一种引入细胞内。

[0063] 一般来说,细胞被维持在适合细胞生长和/或维持的条件下。适合的细胞培养条件为本领域所熟知,且在例如Santiago等人,Proc. Natl. Acad. Sci. USA, 2008, 105: 5809-5814;Moehle等人,Proc. Natl. Acad. Sci. USA, 2007, 104:3055-3060;Urnov等人,Nature, 2005, 435:646-651和Lombardo等人,Nat. Biotechnol., 2007, 25:1298-1306中描述。本领域技术人员领会培养细胞的方法为本领域已知的,且其可以并将会根据细胞类型而不同。在所有情况下,可使用常规优化,以确定用于特定细胞类型的最佳技术。

[0064] (b) 任选的供体多核苷酸

在其中改造的Cas9蛋白包含核酸酶或切口酶活性的实施方案中,该方法可进一步包含引入至少一种供体多核苷酸至细胞内。供体多核苷酸可为单链或双链的,线性或环状的,和/或RNA或DNA。在一些实施方案中,供体多核苷酸可为载体,如质粒载体。

[0065] 供体多核苷酸包含至少一个供体序列。在一些方面,供体多核苷酸的供体序列可为内源或原生染色体序列的修饰版本。例如,供体序列可与在改造的Cas9系统靶向的序列处或附近的染色体序列的一部分基本相同,但其包含至少一个核苷酸的改变。因此,在与原生序列整合或交换后,在靶向的染色体位置的序列包含至少一个核苷酸的改变。例如,该改变可为一个或多个核苷酸的插入、一个或多个核苷酸的缺失、一个或多个核苷酸的取代或其组合。作为修饰的序列的“基因修正”整合的结果,细胞可从靶向的染色体序列产生修饰的基因产物。

[0066] 在其它方面,供体多核苷酸的供体序列可为外源序列。如本文所使用的,“外源”序列指不是细胞原生的序列,或其原生位置在细胞基因组的不同位置的序列。例如,外源序列

可包含蛋白编码序列，其可与外源启动子控制序列可操作地连接，使得在整合至基因组后细胞能够表达整合序列编码的蛋白。或者，外源序列可被整合至染色体序列，使得它的表达受外源启动子控制序列调节。在其它迭代中，外源序列可为转录控制序列、另一种表达控制序列、RNA编码序列等等。如前文所指出的，外源序列整合至染色体序列被称为“敲入”。

[0067] 如本领域技术人员能领会的，供体序列的长度可以且将会不同。例如，供体序列的长度可从数个核苷酸至数百个核苷酸至数十万个核苷酸不等。

[0068] 通常，供体多核苷酸中的供体序列侧翼是上游序列和下游序列，其分别与位于改造的Cas9系统靶向的序列上游和下游的序列具有基本序列同一性。由于这些序列相似性，供体多核苷酸的上游和下游序列允许供体多核苷酸和靶向的染色体序列间的同源重组，使得供体序列可被整合至染色体序列内(或与之交换)。

[0069] 如本文所使用的，上游序列指与改造的Cas9系统靶向的序列上游的染色体序列共享基本序列同一性的核酸序列。类似地，下游序列指与改造的Cas9系统靶向的序列下游的染色体序列共享基本序列同一性的核酸序列。如本文所使用的，短语“基本序列同一性”指序列有至少约75%序列同一性。因此，供体多核苷酸的上游和下游序列与目标序列上游或下游的序列可有约75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%的序列同一性。在示例性实施方案中，供体多核苷酸上游和下游序列与改造的Cas9系统靶向的序列上游或下游的染色体序列可有约95%或100%的序列同一性。

[0070] 在一些实施方案中，上游序列与位于改造的Cas9系统靶向的序列紧邻上游的染色体序列共享基本序列同一性。在其它实施方案中，上游序列与位于目标序列上游约一百(100)个核苷酸以内的染色体序列共享基本序列同一性。因此，例如，上游序列可与位于目标序列上游约1至约20、约21至约40、约41至约60、约61至约80或约81至约100个核苷酸处的染色体序列共享基本序列同一性。在一些实施方案中，下游序列与位于改造的Cas9系统靶向的序列紧邻下游的染色体序列共享基本序列同一性。在其它实施方案中，下游序列与位于目标序列下游约一百(100)个核苷酸以内的染色体序列共享基本序列同一性。因此，例如，下游序列可与位于目标序列下游约1至约20、约21至约40、约41至约60、约61至约80或约81至约100个核苷酸处的染色体序列共享基本序列同一性。

[0071] 每个上游或下游序列长度可在约20个核苷酸至约5000个核苷酸范围。在一些实施方案中，上游和下游序列可包含约50、100、200、300、400、500、600、700、800、900、1000、1100、1200、1300、1400、1500、1600、1700、1800、1900、2000、2100、2200、2300、2400、2500、2600、2800、3000、3200、3400、3600、3800、4000、4200、4400、4600、4800或5000个核苷酸。在特定实施方案中，上游和下游序列长度可在约50至约1500个核苷酸范围。

[0072] (c) 细胞类型

多种真核细胞适合使用于本文公开的方法。例如，细胞可为人类细胞、非人类哺乳动物细胞、非哺乳类脊椎动物细胞、无脊椎动物细胞、昆虫细胞、植物细胞、酵母细胞或单细胞真核生物。在一些实施方案中，细胞可为一个细胞胚胎。例如，非人类哺乳动物胚胎包括大鼠、仓鼠、啮齿类、兔、猫、犬、羊、猪、牛、马、和灵长类胚胎。在另外其它实施方案中，细胞可为干细胞，如胚胎干细胞、ES样干细胞、胎儿干细胞、成体干细胞等。在一个实施方案中，干细胞不是人胚胎干细胞。此外，干细胞可包括通过公开于WO2003/046141(其在此以其整

体并入)或Chung等人(Cell Stem Cell, 2008, 2:113-117)的技术制作的那些。细胞可为体外的(即在培养中)、离体的(即在分离自器官的组织中)或体内的(即在生物体内)。在示例性实施方案中,细胞是哺乳动物细胞或哺乳动物细胞系。在特定实施方案中,细胞是人类细胞或人类细胞系。

[0073] 作为实例,在一些实施方案中,真核细胞或真核细胞群是T细胞、CD8⁺ T细胞、CD8⁺ 幼稚T细胞、中央记忆性T细胞、效应记忆性T细胞、CD4⁺ T细胞、干细胞记忆性T细胞、辅助性T细胞、调节性T细胞、细胞毒性T细胞、自然杀伤T细胞、造血干细胞、长期造血干细胞、短期造血干细胞、多能祖细胞细胞、谱系限制型祖细胞、淋巴祖细胞、胰腺祖细胞、内分泌祖细胞、外分泌祖细胞、骨髓祖细胞、共同髓系祖细胞、红系祖细胞、巨核红系祖细胞、单核前体细胞,内分泌前体细胞、外分泌细胞、成纤维细胞、肝母细胞、成肌细胞、巨噬细胞、胰岛β细胞、心肌细胞、血细胞、导管细胞、腺泡细胞、α细胞、β细胞、δ细胞、PP细胞、胆管上皮细胞、视网膜细胞、光感受器细胞、视杆细胞、视锥细胞、视网膜色素上皮细胞、小梁网细胞、耳蜗毛细胞、外毛细胞、内毛细胞、肺上皮细胞、支气管上皮细胞、肺泡上皮细胞、肺上皮祖细胞、横纹肌细胞、心脏肌肉细胞、肌卫星细胞、肌细胞、神经元、神经元干细胞、间充质干细胞、诱导多能干(iPS)细胞、胚胎干细胞、单核细胞、巨核细胞、中性粒细胞、嗜酸性粒细胞、嗜碱性粒细胞、肥大细胞、网织红细胞、B细胞(如祖B细胞、前B细胞、原B细胞、记忆B细胞、浆B细胞)、胃肠道上皮细胞、胆管上皮细胞、胰腺导管上皮细胞、肠道干细胞、肝细胞、肝星状细胞、肝巨噬细胞、成骨细胞、破骨细胞、脂肪细胞(如棕色脂肪细胞或白色脂肪细胞)、前脂肪细胞、胰腺前体细胞、胰岛细胞、胰腺β细胞、胰腺α细胞、胰腺δ细胞、胰腺外分泌细胞、施旺细胞或少突胶质细胞,或是此类细胞的群。

[0074] 适合的哺乳动物细胞或细胞系的非限制性实例包括人胚胎肾细胞(HEK293、HEK293T)；人宫颈癌细胞(HELA)；人肺细胞(W138)；人肝细胞(Hep G2)；人U2-OS骨肉瘤细胞、人A549细胞、人A-431细胞以及人K562细胞；中国仓鼠卵巢(CHO)细胞、幼仓鼠肾(BHK)细胞；小鼠骨髓瘤NS0细胞、小鼠胚胎成纤维3T3细胞(NIH3T3)、小鼠B淋巴瘤A20细胞；小鼠黑色素瘤B16细胞；小鼠成肌细胞C2C12细胞；小鼠骨髓瘤SP2/0细胞；小鼠胚胎间充质C3H-10T1/2细胞；小鼠癌CT26细胞、小鼠前列腺DuCuP细胞；小鼠乳腺EMT6细胞；小鼠肝癌Hepa1c1c7细胞；小鼠骨髓瘤J5582细胞；小鼠上皮MTD-1A细胞；小鼠心肌MyEnd细胞；小鼠肾RenCa细胞；小鼠胰腺RIN-5F细胞；小鼠黑色素瘤X64细胞；小鼠淋巴瘤YAC-1细胞；大鼠胶质母细胞瘤9L细胞；大鼠B淋巴瘤RBL细胞；大鼠神经母细胞瘤B35细胞；大鼠肝癌细胞(HTC)；水牛鼠肝BRL 3A细胞；犬肾细胞(MDCK)；犬乳腺(CMT)细胞；大鼠骨肉瘤D17细胞；大鼠单核细胞/巨噬细胞DH82细胞；猴肾SV-40转化成纤维细胞(COS7)；猴肾CVI-76细胞；非洲绿猴肾(VERO-76)细胞。哺乳动物细胞系的广大列表可在美国典型培养物保藏中心目录(ATCC, 马纳萨斯, 弗吉尼亚州)中找到。

[0075] (V)应用

本文公开的组合物和方法可在多种治疗、诊断、工业和研究应用中使用。在一些实施方案中,可使用本公开内容在细胞、动物或植物中修饰任意目标染色体序列,以构建基因功能模型和/或研究基因功能,研究目标遗传或表观遗传的条件,或研究涉及各种疾病或病症的生化途径。例如,可创造构建疾病或病症的模型的转基因生物,其中与疾病或病症关联的一个或多个核酸序列的表达发生改变。疾病模型可用于研究突变对生物的作用,研究疾

病的发展和/或进展,研究药学活性化合物对疾病的作用,和/或评价潜在基因治疗策略的功效。

[0076] 在其它实施方案中,所述组合物和方法可用于执行有效且成本有效且的功能性基因组筛选,其可被用于研究涉及特定生物学过程的基因功能以及基因表达的任意改变可如何影响生物学过程,或结合细胞表型对基因组基因座执行饱和或深度扫描诱变。例如,饱和或深度扫描诱变可用于确定基因表达、药物抵抗和疾病逆转所需的功能元件的关键最小特征和离散脆弱性。

[0077] 在进一步的实施方案中,本文公开的组合物和方法可用于诊断测试,以确定疾病或病症的存在和/或用于确定治疗选项。适合的诊断测试的实例包括癌细胞中特异性突变(如在EGFR、HER2等的特异性突变)的检测、与特定疾病相关的特异性突变(如三核苷酸重复、与镰状细胞病相关的β珠蛋白中的突变、特异性SNP等)的检测、肝炎的检测、病毒(如寨卡)的检测等等。

[0078] 在另外的实施方案中,本文公开的组合物和方法可用于修正与特定疾病或病症相关的基因突变,如修正与镰状细胞病或地中海贫血相关的珠蛋白基因、修正与重度联合免疫缺陷(SCID)相关的腺苷脱氨酶基因中的突变、减少HTT(亨廷顿舞蹈病的致病基因)表达或修正视紫红质基因中的突变以治疗视网膜色素变性。此类修饰可在离体细胞中进行。

[0079] 在另外其它实施方案中,本文公开的组合物和方法可用于产生具有改良性状或增强的对环境胁迫的抗性的作物。本公开内容也可用于产生具有改良性状的家畜或生产动物。例如,猪有许多使它们具有作为生物医学模型的吸引力的特征,特别是在再生医学或异种移植中。

[0080] 在另外其它实施方案中,本文公开的组合物和方法可用于测定在活细胞或化学固定细胞(如用于福尔马林固定石蜡包埋的临床样本的福尔马林固定)中的染色体身份和位置。例如,通过本文公开的肽序列连接至荧光蛋白的CRISPR复合体可能以单个或多个拷贝靶向遗传基因座,且可用显微术探测这些复合体以确定染色体基因座拷贝数和/或位置。用于追踪的遗传基因座实例可包含着丝粒区域、端粒区域或可与单个相同CRISPR复合体的多个拷贝结合的基因组其它重复区域。

[0081] 定义

除非另有定义,否则本文使用的所有技术和科学术语均具有本发明所属领域技术人员通常理解的含义。下列参考文献向技术人员提供本发明使用的许多术语的通常定义:Singleton等人,Dictionary of Microbiology and Molecular Biology (2nd Ed. 1994);The Cambridge Dictionary of Science and Technology (Walker ed., 1988);The Glossary of Genetics, 5th Ed,R. Rieger等人编;Springer Verlag (1991);以及Hale & Marham,The Harper Collins Dictionary of Biology (1991)。如本文所使用的,除非另有说明,以下术语具有归属于其的含义。

[0082] 当介绍本公开内容或其优选实施方案的要素时,冠词“一个”、“一种”、“该”和“所述”预期表示存在一个或多个要素。术语“包含”、“含有”和“具有”预期为包括性的,且表示可存在除所列要素外的其它要素。

[0083] 当关联数值例如x使用时,术语“约”表示x±5%。

[0084] 如本文所使用的,术语“互补的”或“互补性”指通过经由特定氢键进行的碱基配对

的双链核酸缔合。碱基配对可为标准的沃森-克里克碱基配对(如5'-AGTC-3'与互补序列3'-TCAG-5'配对)。碱基配对也可为Hoogsteen或反向Hoogsteen氢键合。互补性通常关于双链体区域而测量,且因此不包括例如突出。如果仅某些(如70%)的碱基是互补的,双链体区域两条链之间的互补性可为部分的,且可用百分比(如70%)表示。非互补的碱基是“错配的”。如果双链体区域的所有碱基都是互补的,互补性也可为完全的(即100%)。

[0085] 如本文所使用的,术语“CRISPR/Cas系统”或“Cas9系统”指包含Cas9蛋白(即核酸酶、切口酶或催化死亡蛋白)和向导RNA的复合体。

[0086] 如本文所使用的,术语“内源序列”指细胞原生的染色体序列。

[0087] 如本文所使用的,术语“外源”指细胞非原生的序列或其在细胞基因组中的原生位置在不同染色体位置的染色体序列。

[0088] 如本文所使用的,“基因”指编码基因产物的DNA区域(包括外显子和内含子),以及所有调节基因产物生产的DNA区域,无论此类调节序列是否邻近编码和/或转录序列。相应地,基因包括但不一定限于启动子序列、终止子、翻译调节序列(如核糖体结合位点和内部核糖体进入位点)、增强子、沉默子、绝缘子、边界元件、复制起点、基质结合位点和基因座控制区。

[0089] 术语“异源”指对目标细胞并非内源或原生的实体。例如,异源蛋白指来源于或最初来源于外源来源(如外源引入核酸序列)的蛋白。在一些情况下,异源蛋白通常不是由目标细胞细胞产生的。

[0090] 术语“切口酶”指切割双链核酸序列的一条链(即令双链序列产生切口)的酶。例如,具有双链切割活性的核酸酶可通过突变和/或缺失而被修饰,从而发挥切口酶的功能并仅切割双链序列中的一条链。

[0091] 如本文所使用的,术语“核酸酶”指将双链核酸序列两条链都切割的酶。

[0092] 术语“核酸”和“多核苷酸”指脱氧核糖核苷酸或核糖核苷酸聚合体,其为线性或环状构造,且为单链或双链形式。为了本公开内容的目的,这些术语不应被解释为关于聚合体长度的限制。该术语可包含天然核苷酸的已知类似物,以及碱基、糖和/或磷酸酯部分(如硫代磷酸酯主链)被修饰的核苷酸。通常,特定核苷酸的类似物拥有相同的碱基配对特异性,即A的类似物会与T碱基配对。

[0093] 术语“核苷酸”指脱氧核糖核苷酸或核糖核苷酸。核苷酸可为标准核苷酸(即腺苷、鸟苷、胞苷、胸苷和尿苷)、核苷酸异构体或核苷酸类似物。核苷酸类似物指具有修饰的嘌呤或嘧啶碱基或修饰的核糖部分的核苷酸。核苷酸类似物可为天然存在的核苷酸(如肌苷、假尿苷)或是非天然存在的核苷酸。核苷酸在糖或碱基部分的修饰的非限制性实例包括添加(或去掉)乙酰基、氨基、羧基、羧甲基、羟基、甲基、磷酰基和硫醇基,以及以其它原子取代碱基中的碳和氮原子(如7-脱氮嘌呤)。核苷酸类似物也包括双脱氧核苷酸、2'-0-甲基核苷酸、锁核酸(LNA)、肽核酸(PNA)和吗啉代。

[0094] 术语“多肽”和“蛋白”可互换地使用,指氨基酸残基的聚合体。

[0095] 术语“目标序列”、“目标染色体序列”和“目标位点”可互换地使用,指在染色体DNA中改造的Cas9系统靶向的特异性序列,以及改造的Cas9系统在该处修饰DNA或与DNA相关的蛋白的位点。

[0096] 用于确定核酸和氨基酸序列同一性的技术为本领域已知的。典型地,此类技术包

括确定基因的mRNA的核苷酸序列和/或确定由此编码的氨基酸序列，且将这些序列与第二核苷酸或氨基酸序列进行比较。基因组序列也可以这种方式确定和比较。通常来说，同一性指两个多核苷酸或多肽序列分别的核苷酸对核苷酸或氨基酸对氨基酸的精确对应关系。两个或更多个序列(多核苷酸或氨基酸)可通过确定它们的同一性百分比进行比较。无论是核酸还是氨基酸序列，两个序列的同一性百分比是两个比对序列之间精确配对的数目除以较短序列的长度并乘以100。核酸序列的近似比对由Smith和Waterman在Advances in Applied Mathematics 2:482-489 (1981) 中的局部同源性算法提供。此算法可通过使用打分矩阵应用于氨基酸序列，打分矩阵由Dayhoff, Atlas of Protein Sequences and Structure,M. O. Dayhoff ed.,5 suppl. 3:353-358,National Biomedical Research Foundation,美国华盛顿开发，且由Gribskov, Nucl. Acids Res.,14 (6) :6745-6763 (1986) 标准化。该算法确定序列同一性百分比的示例性实现由遗传学计算机组(麦迪逊,威斯康星州)在“BestFit”实用应用程序提供。其它适合计算序列间同一性百分比或相似性的程序通常在领域内普遍已知，例如，另一种比对程序是使用默认参数的BLAST。例如，可使用以下默认参数使用BLASTN和BLASTP：遗传密码=标准；过滤器=无；链=两条；截止值=60；期望=10；矩阵=BLOSUM62；描述=50个序列；排序方式=高分值；数据库=非冗余，GenBank+EMBL+DDBJ+PDB+GenBank CDS翻译+Swiss蛋白+Supupdate+PIR。这些程序的细节可在GenBank网页找到。

[0097] 由于可在上述细胞和方法中作各种改变而不偏离该发明的范围，因此预期应当将上面的描述和下面给出的实施例中包含的所有内容解释为说明性的，而不应以限制性含义解释。

实施例

[0098] 下列实施例说明公开内容的某些方面。

[0099] 实施例1：使用GFP-SpCas9和RFP-SpCas9融合蛋白的人类细胞基因编辑

使用60 pmol的SpCas9、GFP-SpCas9或RFP-SpCas9重组蛋白和180 pmol体外转录的单个向导RNA(sgRNA)转染人类K562细胞(0.35×10^6)，所述单个向导RNA靶向人EMX1基因座，具有向导序列5'-GCUCCCAUCACAUCAACCGG-3'。使用核转染溶液V和Amaxa仪器实施转染。在收获细胞用于基因编辑分析前，细胞保持在37°C和5% CO₂三天。使用QuickExtract DNA抽提溶液制备基因组DNA。使用由目标特异性序列和下一代测序(NGS)接头组成的引物PCR扩增靶向EMX1区域。正向引物为5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNNNAGTCTTCCCATCAGGCTCTCA-3' (SEQ ID NO: 46)，反向引物为5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNNNNNNAGAGTCCAGCTTGGGCC-3' (SEQ ID NO: 47)，其中目标特异性序列用下划线标出，N代表A、T、G或C。通过使用Illumina MiSeq的NGS分析PCR扩增子，以测定各Cas9蛋白的编辑效率。展示于图1的结果显示GFP-SpCas9和RFP-SpCas9融合蛋白各自保持与SpCas9蛋白的水平相似的编辑活性。

[0100] 表1展示人类密码子优化的DNA和改造的Cas9/NLS蛋白的蛋白序列，其中NLS序列以粗体文本展示，标记蛋白和Cas9间的接头以加下划线文本展示。

表 1. 改造的 Cas9 系统

GFP-SpCas9 融合蛋白的氨基酸序列

MVSKGEELFTGVVPILVELDGDVNGHKFSVS~~GE~~EGDATYGKLT~~K~~FICTTGKLPVPW
 PTLVTTLYGVQCF~~S~~RYPDHMKQHDF~~F~~KSAMPEGYVQERTIFFKDDGNYKTRAEVKFE
 GDTLVNRIELKGIDFKEDGNILGHKLEYNNNSHN~~V~~YIMADKQKNGIKVNFKIRHNIEDG
 SVQLADHYQQNTPIGDGPVLLPDNHYLSTQSKLSKD~~P~~NEKRDHMV~~L~~EVTAA~~G~~ITLG
 MDELYKVDA~~E~~~~A~~~~A~~~~K~~~~E~~~~A~~~~A~~~~K~~~~A~~~~E~~~~A~~~~A~~~~K~~~~E~~~~A~~~~A~~~~K~~~~A~~~~K~~~~A~~~~P~~
AAKRVKLDGGGGSTGMDKKYSIGLDIGTNSVGWAVITDEYKVPSSKKFKVLGNTDRHS
 IKKNLIGALLFD~~S~~GETAEATRLKRTARRRYTRRKNR~~I~~CYLQE~~I~~FSNEMAKVDDSFHRLE
 ESFLVEEDKKHERHP~~I~~FGNIVDEVAYHEKYPTIYHLRK~~K~~LVDSTD~~K~~ADLR~~L~~IYLA~~A~~H~~M~~
 IKFRGHF~~L~~IEGDLNP~~D~~NSDV~~K~~LF~~I~~QLVQT~~N~~Q~~L~~FEENPINASGVDAKAILSARLSKSRR
 LENLIAQLPGEKKNGLFGNLIALSGLTPNF~~K~~SNFDLAEDAKLQLSKDTYDDLDNLLA
 QIGDQYADLFLAAKNLSDA~~I~~LLSDILRVNTEITKAPLSASMIKRYDEHHQDL~~T~~LLKALVR
 QQLPEKYKEIFFDQS~~K~~NGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLL
 RKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKD~~N~~REKIEKILTFRIPYYVG~~L~~ARGNS
 RFAWMTRKSEETITPWN~~F~~EEVVDKGASAQS~~F~~IERMTNFDKNLPNEKVLPKHSLLYEYF
 TVYNELTKV~~K~~YVTEGMRKP~~A~~FLSGEQQKAIVD~~L~~LF~~K~~TK~~N~~R~~K~~V~~T~~KQLKEDYFKKIECFD
 SVEISGV~~E~~DRFN~~A~~SLGT~~Y~~HD~~L~~LI~~K~~DKDFLDNEENEDI~~L~~EDIV~~L~~TLFEDREMIEERLK
 TYAHLFDDKVMQQLKRRRTGWGR~~L~~RSRK~~L~~INGIRD~~K~~QSGKT~~I~~LD~~F~~LKSDGFANRNFM
 QLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA~~I~~KKGILQTVKVVDELVKVMGR
 HKPENIVIEMARENQTTQKGQKNSRERMKR~~I~~E~~G~~IKELGSQ~~I~~LKEHPVENTQLQNEKLY
 LY~~Y~~LQNGRDMYVDQELDINRLSDYD~~V~~DHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNV
 PSEEVVKMKNYWRQ~~L~~NAKLITQRKF~~D~~NLTKAERGGLSELDKAGFIKRQLVETRQIT
 KHVAQILD~~S~~RMNTK~~Y~~DENDKLIREV~~K~~VITL~~K~~SKLV~~S~~DFRKDFQFYKVREINNYHHAHD
 AYLNAVVG~~T~~ALIKKYPKLESEFVYGDYKVYDVRK~~M~~IAK~~S~~EQEIGKATAKYFFYSNIMN
 FF~~K~~TEITLANGEIRKRPLIETNGETGEIVWDKG~~R~~DFATVRK~~V~~LSMPQVNIVKKTEVQTG
 GFSKESILPKRNSDKL~~I~~ARKKD~~W~~DPKKYGGFDSPTVAYSVLV~~V~~AKVEKGKSKKL~~K~~SV~~K~~
 ELLGITIMERSFEKNPIDFLEAKGYKEVKKD~~L~~I~~K~~LPK~~Y~~SLFE~~N~~GRKRMLASAGELQ
 KGNELALPSKYVN~~F~~LYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVI
 LADANLDKV~~L~~SAYNKHRDK~~P~~IREQAENIIHLFTLTNLGAPA~~A~~FKYFDTTIDRKRYTSTK
 EVLDATL~~H~~QSITGLYETRIDLSQLGGDEF**PKKKRKV**GGGGSP**KKKRKV**(SEQ ID NO:
 48)

下划线: GFP 和 SpCas9 间接头

粗体：核定位信号

RFP-SpCas9 融合蛋白的氨基酸序列

MVKGEAVIKEFMRKVHMEGSMNGHEFEIEGEGRPYEGTQAKLKVTKGGLPFS
 WDILSPQFMYGSRAFTKHPADIPDYYKQSFPEGFKWERVMNFEDGGAVTVTQDTSLED
 GTLIYKVKLRTNFPPDGPVMQKTMGWEASTERLYPEDGVLKGDIKMALRLKDGG
 YLADFCKTYKAKKPVQMPGAYNVDRKLDITSHNEDYTVEQYERSEGRHSTGGMDEL
YKVDSGGSSGGSSGSETPGTSESATPESSGGSSGGSPA**A**KRV**KLD**GGGGSTGMDKKYS
 IGLDIGTNSVGWAVITDEYKVPSSKKFKVLGNTDRHSIKKNLIGALLFDGETAEATRK
 RTARRRYTRRKNRICYLQEIFSNEAKVDDSSFFHRLEESFLVEEDKKHERHPIFGNIVDE
 VAYHEKYPTIYHLRKKLVSTDKA**D**LRLIYLALAHMIKFRGHFLIEGDLNPDNSDV**D**KL
 FIQLVQTYNQLFEENPINASGVDAKAILSARLSRRLENIAQLPGEKKNGLFGNLIALS
 LGLTPNFKNFDLAEDAKLQLSKDTYDDDDLNLLAQIGDQYADLFLAAKNLSDAILLS
 DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNYAGY
 IDGGASQEEFYKFKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAIL
 RRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVD
 KGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVNELTKVKYVTEGMRKPAFLS
 GEQKKAIVDLLFKTNRKVTKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKII
 KDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGW
 GRLSRKLINGIRDKQSGKTILDFLKSDGFANRNMQLIHDDSLTFKEDIQKAQVSGQGD
 SLHEHIANLAGSPAIKKGILQTVKVVDELVKUMGRHKPENVIMARENQTTQKQGQKN
 SRERMKRIEEGIKLGSQILKEHPVETQLNEKLYYYLQNGRDMYVDQELDINRLSD
 YDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKMKNYWRQLLNAKLIT
 QRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIRE
 VKVITLKSKLVSDFRKDFQYKVREINNYHHADAYLNAVVGTALIKKYPKLEEFVY
 GDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG
 EIVWDKGRDFATVRKVLMPQVNIVKTEVQTGGFSKEILPKRNSDKLIARKKDWDP
 KKYGGFDSPTVAYSVLVAKVEKGKSKKLKSVKELLGITMERSFEKNPIDFLEAKGY
 KEVKKDLIKLPYSLFELENGRKRMLASAGELQKGNELALPSKYVFNFLYLASHYEKL
 KGSPEDNEQKQLFVEQHKHYLDEIIEQISEFKRVILADKVLSAYNKHRDKPIREQ
 AENIIHLFLTLTNLGAPAAAFKYFDTTIDRKYTTKEVLDATLHIQSITGLYYETRIDLSQLG
GDEFPKKKRKVGGGGSPKKKRKV(SEQ ID NO: 49)

下划线：RFP 和 SpCas9 间接头

粗体：核定位信号

GFP-eSpCas9 融合蛋白的氨基酸序列

MVKGEELFTGVVPILVELDGVNGHKFSVGEGEGDATYGKLTKFICTTGKLPVPW
 PTLVTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFKDDGNYKTRAEVKFE
 GDTLVNRIELKGIDFKEDGNILGHKLEYNYNNHVYIMADKQKNGIKVNFKIRHNIEDG
 SVQLADHYQQNTPIGDGPVLPDNHYLSTQKLKSDPNEKRDHMVLLEFVTAAGITLG
 MDELYKVDSGGSSGSSGSETPGTSEATPESSGGGGSSGSPAKRV**KLD**GGGGSTGMD
 KKYSIGLDIGTTNSVGWAVITDEYKVPSSKKFVLGNTDRHSIKKNLIGALLFDSGETAEA
 TRLKRTARRRYTRKNRICYLQEIFSNEAKVDSSFHRLEESSFLVEEDKKHERHPIFGNIVDE
 NIVDEVAYHEKYPTIYHLRKKLVSDKADLRIYLALAHMIKFRGHFLIEGDLNPDNS

DVDKLFQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFG
NLIALSLGLTPNFKNFDLAEDAKLQLSKDTYDDLDNLLAQIGDQYADLFLAAKNLS
DAILLSDLRVNTEITKAPLSASMICKRYDEHHQDLTLLKALVRQQLPKEYKEIFFDQSKN
GYAGYIDGGASQEELYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLG
ELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWN
FEEVVDKGASAQSFIERMNTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMR
KPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTY
HDLLKIIKKDFLDNEENEDILEDIVLTTLFEDREMIEERLKTYAHLFDDKVMKQLKR
RRTYGWGRSLRKLINGIRDQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQ
VSGQQDSLHEHIANLAGSPAIIKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTT
QKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYLQNGRDMYVDQEL
DINRLSDYDVDHVIPQSFLADDSDNKVLTRSDKNRGKSDNPSEEVVKKMKNYWRQL
LNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILD SRMNTKYDE
NDKLIREVKVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVTALIKKYP
LESEFVYGDYKVDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKAPI
ETNGETGEIVWDKGRDFATVRKVLMPQVNIVKKTETVQTGGFSKESILPKRNSDKLIAR
KKDWDPKKYGGFDSPTVAYSVLVAKVEKGKSKLKSVKELLGITMERSSFEKNPID
FLEAKGYKEVKKDLIILPKYSLFELENGRKMLASAGELQKGNELALPSKYVNFLYL
ASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIEQISEFSKRVILADANLDKVLSAYNKH
RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSI
TGLYET
RIDSQLGGDEF**PKKKRKV**GGGG**PKKKRKV**(SEQ ID NO: 50)

下划线: GFP 和 eSpCas9 间接头

粗体: 核定位信号

[0101] 用于产生三种蛋白的人类密码子优化的DNA序列如下:

人类密码子优化的 GFP-SpCas9 DNA 序列

ATGGTTAGCAAAGGTGAAGAACTGTTACAGGTGTTCCGATTCTGGTTGAAC TG
GATGGTGTGTTAATGCCACAAATTTAGCTTACGGTGAAGGCAGGGTATGCA
ACCTATGGTAAACTGACCCCTGAAATTATCTGTACCACCGCAAACGCCGGTTCCG
TGGCGACACTGGTACCCACTGACCTATGGTGTTCAGTGTTAGCCGTTATCCGG
ATCACATGAAACAGCACGATTTCAGGAAAGCTGACGGCAACTATAAACCGTGC
AACGTACCATCTTCAAAGATGACGGCAACTATAAACCGTGCAGGTAAAT
TTGAAGGTGATACCTGGTGAATCGCATTGAACTGAAAGGCATCGATTAAAGAGG
ATGGTAATATCCTGGGCCACAAACTGGAATATAATTATAATAGCCACAACGTGTACA
TCATGGCCGACAAACAGAAAAATGGCATCAAAGTGAACCTCAAGATCCGCCATAAT
ATTGAAGATGGTCAGTCAGCTGGCGATCATTATCAGCAGAATACCCGATTGGT
GATGGTCCGGTCTGCTGCCGATAATCATTATCTGAGCACCCAGAGCAAACGTGAGC
AAAGATCCGAATGAAAAACGTGATCACATGGTGTGCTGGAATTGTTACCGCAGCA
GGTATTACCTTAGGTATGGATGAACTGTATAAAGTCGACGCAGAAGCAGCAGCAA
AGAAGCCGCTGCCAAAGAACGGCAGCGAAAGAGGCAGCCGCAAAGCACTGGAA
GCCGAGGCTGCCGCTAAAGAGGCTGCTGCAAAAGAAGCAGCCGCTAAAGAAGCTGC
GGCTAAGGCACCCGAGCAAACGTGTTAAACTGGACGGTGGTGGTAGCACCG
GTATGGACAAGAAATACAGCATCGGTTGGATATTGGCACGAATAGCGTGGTTGG
GCCGTTATTACCGACGAGTACAAAGTGGCTCCAAGAAATTCAAAGTGTGGCAAT
ACCGATGCCATAGCATCAAGAAAAATCTGATTGGCGACTGCTGTCGACAGCGGT
GAGACTGCCGAAGCTACGCGTCTGAAGCGTACGGCGCTCGCTACACCCGCCGT
AAGAACCGTATTGCTATCGCAAGAAATCTCAGCAACGAAATGCCAAAGTTGAT
GATAGCTTTTCACCGCCTGGAAGAGAGCTTCTGGTGGAGAGGATAAGAAACAC
GAGCGCCATCCGATTTGGTAACATTGTCGATGAAGTGGCATACCATGAGAAGTAC
CCGACCATCTACCAACCTCGTAAGAAACTGGTGACAGCACCATAAGCTGATCTG
CGTCTGATTACCTGGCGCTGCCACATGATTAAGTTGCGGTCTGATCG
AGGGCGATCTGAATCCGACAATTCTGATGTTGACAAGCTGTTATTCAACTGTAC
AGACCTACAACCAGTTGTCGAAGAGAACCCGATCAATGCGAGCGGTGTTGATGCC
AAAGCAATTCTGAGCGCACGCCCTGAGCAAATCTGCCGTTGGAGAACCTGATTGCA
CAGCTGCCGGGTGAGAAGAAAAACGGTCTGTCGGCAATCTGATTGCACTGTCCCTG
GGCTTGACCCGAATTAAAGAGCAACTCGACCTGGCGAAGATGCGAAGCTCCAA
TTGAGCAAAGACACCTACGACGATGACCTGGACAATCTGCTGGCCCAGATTGGCGA
CCAGTACGCGAGATCTGTTCTGGCTGCGAAAAACCTGAGCGATGCAATTCTGCTGTC
GGACATCCTGCGCGTGAATACGAAATCACGAAAGCGCCTCTGAGCGCGTCTGATGAT
CAAGCGCTATGACGAGCACCACCAAGATCTGACCTGCTGAAAGCTCTGGTGGAGAC
AACAAATTGCCAGAGAAGTAAAGAAATTCTTGTACCGAGAGCAAACCGCTAT
GCGGGTTACATTGACGGTGGCGCCAGCCAAGAACAGAGTTCTACAAATTGATTAAGCCT
ATCCTGGAGAAAATGGATGGCACCGAAGAACGACTGCTGGTAAAGCTGAATCGTGAAGA
TCTGCTGCGAAACAGCGACTTTGATAACGGTAGCATTCCGACCCAGATCCATCT
GGGTGAGTTGCACCGGATTGCGTCCAGGAAGATTATCCGTTCTGAAAGA
CAACCGTGAGAAAATGAGAAAATTCTGACGTTCCGTATCCGTATTATGTCGGCCC

GCTGGCGCGTGGTAATAGCCGCTTCGCGTGGATGACCCGCAAATCAGAGGAAACGA
 TTACCCCGTGGATTGAGGAAGTTGTTGATAAGGGTCAAGCGCGCAGTCGTTCA
 TTGAGCGTATGACCAACTTGACAAGAATTGCCGAATGAAAAGTCTTGCCGAAGC
 ACTCTCTGCTGTACGAGTATTACCGTTACAACGAATTGACCAAGGTTAACACGT
 CACCGAAGGCATGCGCAAACCGGCCTCCTGAGCGGCAGCAGAAAAAGCAATCG
 TTGACCTCTGTTAAGACCAACCGCAAGGTTACGGTCAAACAACACTGAAAGAGGACT
 ATTTCAAGAAAATTGAATGTTTGACTCCGTAGAGATCTCCGTGTTGAGGACCGTT
 TCAACCGGAGCGCTGGGCACCTACCATGATCTGCTGAAAATTATTAAAGACAAAGATT
 TTCTGGACAACGAAGAGAACGAAGATATTCTGAAGATATCGTTCTGACCCCTGACGC
 TGTTCGAAGATCGTAGAGATGATTGAGGAACGTCTGAAAACCTACGCACACTGTTCG
 ATGACAAAGTTATGAAACAGCTGAAGCGTCGTTACACAGGTTGGGCCGCTGA
 GCCGTAAGCTTATCAATGGTATCCGTGACAAACAGAGCGGTAAGACGATTCTGGACT
 TTCTGAAGTCAGATGGCTCGCCAATCGCAACTTTATGCAACTGATTGACGACT
 CTCTGACGTTCAAGGAAGATATCCAAAAGGCACAGGTGAGCGGTAGGGGTGATAGC
 CTGCATGAGCATATCGGAACCTGGCGGTAGCCCGCTATCAAAAGGGTATCTTA
 CAGACTGTGAAAGTTGGATGAATTGGTAAGGTTATGGTCGTACAAACCGGAA
 AATATTGTGATCGAGATGGCACGTGAAAATCAGACGACGCAAAGGGTAAAAAAAA
 TTCTCGTGAGCGCATGAAACGTATTGAAGAGGGTATCAAAGAATTGGCAGCCAAA
 TTCTGAAAGAACACCCGGTCGAGAACACCCAGCTGCAAACGAAAAACTGTATTITA
 TACTATCTGAGAACCGTGTGACATGTACGTGGATCAAGAACTGGACATCAATCGT
 TTGAGCGATTACGATGTTGATCATATTGTCCTCAGAGCTTCTGAAAGACGATTG
 ATCGACAACAAAGTGTGACCCGTAGCGACAAGAATCGTGGTAAGAGCGATAACGT
 GCCGAGCGAAGAACGTTAAGAAAATGAAAAACTACTGGCGTCAGCTGCTGAACG
 CCAAGCTGATTACCCAGCGTAAGTCGATAACCTGACGAAAGCCGAGCGTGGAGGC
 CTGAGCGAGCTGGACAAGGCCGGTTATCAAGCGTAACCTGGAAACCCGTCA
 GATCACTAACATGTGGCACAGATCTGGACTCCCGCATGAATACGAAATATGACG
 AGAATGACAAGTTGATCCGTGAAAGTCAAAGTTTACGCTGAAAAGCAAACGGT
 TCCGATTCCGTAAGACTTCCAGTTCTATAAAGTCCGTGAAATCAACAACATCAT
 CACGCCACGATGCGTACTTGAACGCTGTTGGCACCGCACTGATCAAGAAATAC
 CCTAAGCTCGAAAGCGAGTTGTCTATGGTACTATAAAGTTACGACGTGCGTAAG
 ATGATGCCAAGAGCGAGCAAGAAATTGTAAGGCTACCGCAAAGTACTTTCTAC
 AGCAACATCATGAACCTCTCAAACCGAGATTACCGTGGCAACGGTGAGATCCGT
 AAACGGCCGCTGATTGAGACTAATGGCGAAACGGCGAGATTGTGTTGGACAAGGG
 TCGCGATTGCTACGGTTCGTAAGGTCTGAGCATGCCGAAGTTAACATTGCAA
 GAAAAGTGAAGTCGAGACGGGTGGCTTGTAGCAAAGAATCCATCTGCCGAAGCGTA
 ATAGCGATAAAACTTATCGCCGTAAAAAGACTGGGACCCAAAGAAATATGGCGC
 TTTGATAGCCGACCGTCGCGTATAGCGTGTAGTGGTCGCGAAAGTTGAAAAGGGC
 AAGAGCAAGAAAAGTGAAGTCCGTCAAAGAACTTCTGGGTATCACCACATGGAACG
 TAGCTCCTTGAGAAGAACCGATTGACTTCTAGAGGCCAAGGTTATAAAGAAGT
 CAAAAAAAGACCTGATTATCAAGCTGCCGAAGTACAGCCTGTTGAGTTGGAGAATG
 GTCGTAAGCGCATGCTGGCGAGCGCGGGTGAGCTGCAAAGGGCAACGAACGGCG
 CTGCCGTCGAAATACGTCAATTCTGACCTGCCAGCCACTACGAAAAGCTGAAG

GGTTCTCCGAAGATAACGAACAAAAGCAACTGTTGAGCAACATAAACACTA
 CTTGGACGAAATCATCGAGCAAATTAGCGAATTAGCAAACGTGTCATCCTGGCGGA
 CGCGAATCTGGACAAGGTCTGTCTGCATACAATAAGCATCGCGACAAACCAATTG
 TGAGCAAGCGGAGAATATCATCCACCTGTTACGCTGACCAACCTAGGTGCGCCGGC
 GGCATTCAAGTATTCGATACGACCATCGACCGCAAGCGTATACAGCACCAAAGA
 GGTCTGGACGCACCCGTATCCACCAAGAGCATTACCGGTTATACGAAACCCGTAT
 TGATTGAGCCAACGGTGGCGATGAATTCCGAAAAAAAGCGCAAAGTTGGTG
 GCGGTGGTAGCCGAAAAAGAAACGTAAAGTG(SEQ ID NO: 62)

人类密码子优化的 RFP-SpCas9 DNA 序列

ATGGTTAGCAAAGGTGAAGCCGTGATTAAGAATTATGCGCTTAAGGTTACATG
 GAAGGTAGCATGAATGCCATGAATTGAAATTGAAGGTGAAGCGAAGGTCGTCC
 GTATGAAGGCACCCAGACCGCAAAACTGAAAGTTACCAAAGGTGGTCCGCTGCCGT
 TTAGCTGGATATTCTGAGTCCGAGTTATGTATGGTAGCCGTGCATTACCAAAC
 ATCCGGCAGATATTCCGGATTATTACAAACAGAGCTTCCGGAAGGTTAAATGGG
 AACGTGTGATGAATTGAAAGATGGTGGTCAGTTACCGTTACACAGGATACCAGCC
 TGGAAAGATGGCACCCGTATCTAAAGTTAAACTGCGTGGCACCAATTCCGCTG
 ATGGTCCGGTTATGCAGAAAAAAACAATGGGTTGGGAAGCAAGCACCAGTCTG
 TATCCTGAAGATGGCGTCTGAAAGGTGATATCAAATGGCACTGCGTCTGAAAGAT
 GGCGGTGTTATGGCAGATTCAAACACCTATAAGCaaaaACCTGTTCAAGCGGTAG
 ATGCCTGGTGCCTATAATGTTGATCGTAAACTGGATATTACCGCCACAACGAAGAT
 TATAACCGTTGTGGAACAGTATGAACGTAGCGAAGGCCGTACAGCACAGGTGGTAT
 GGATGAACTGTATAAGTCGACAGCGGTGGTAGCAGCGTGGTCAAGCGGTAGCG
 AAACACCGGGTACAAGCGAAAGCGAACACCGGAAAGCAGTGGTGGTAGTCAGGT
 GGTAGTCGGCAGCAAAACGTGTGAAACTGGATGGCGGTGGCGGTAGACACCGGTAT
 GGACAAGAAATACAGCATCGTTGGATATTGGCACGAATAGCGTGGTGGCCG
 TTATTACCGACGAGTACAAAGTGCCTCAAGAAATTCAAAGTGTCTGGCAATACCG
 ATGCCATAGCATCAAGAAAATCTGATTGGCGACTGCTGTTGACAGCGGTGAGA
 CTGCCGAAGCTACCGTCTGAAAGCGTACGGCGCGTCGCTACACCGCCGTAAGA
 ACCGTATTGCTATCTGCAAGAAATCTCAGCAACGAAATGGCAAAGTTGATGATA
 GCTTTTACCGCCTGGAAGAGAGCTTCTGGTGGAGAGGATAAGAAACACGAGC
 GCCATCCGATTTGGTAACATTGTCGATGAAGTGGCATACCATGAGAAAGTACCGA
 CCATCTACCACCTCGTAAGAAACTGGTGACAGCACCGATAAAGCTGATCTGCGTC
 TGATTACCTGGCGCTGGCCCACATGATTAAGTTCGCGGTCTTCTGATCGAGGG
 CGATCTGAATCCGGACAATTCTGATGTTGACAAGCTGTTATTCAACTTGACAGAC
 CTACAACCAGTTGTTGCAAGAGAGACCCGATCAATGCGAGCGGTGTTGATGCCAAG
 CAATTCTGAGCGCACGCCGTGAGCAAATCTGCCGTTGGAGAACCTGATTGACAGC
 TGCCGGGTGAGAAGAAAAACGGTCTGTTGGCAATCTGATTGCACTGCTCCCTGGGCT
 TGACCCCCGAAATTAAAGAGCAACTCGACCTGGCGAAGATGCGAAGCTCCAATTGA
 GCAAAGACACCTACGACGATGACCTGGACAATCTGCTGGCCCAGATTGGCGACCAG
 TACGCAGATCTGTTCTGGCTGCGAAAACCTGAGCGATGCAATTCTGCTGCGAC
 ATCCTGCGCGTGAATACGAAATACGAAAGCGCCTCTGAGCGCGTCTATGATCAAG
 CGCTATGACGAGCACCACCAAGATCTGACCCGTGAAAGCTCTGGTGAGACAACA

ATTGCCAGAGAAGTATAAAGAAATTTCTTGACCAGAGCAAAAACGGCTATGCGG
GTTACATTGACGGTGGCGCAGCCAAGAAGAGTTCTACAAATTCTTAAGCCTATCC
TGGAGAAAATGGATGGCACCGAAGAACACTGCTGGTAAAGCTGAATCGTAAGATCTG
CTGCGCAAACAGCGCACTTTGATAACGGTAGCATTCCGACCCAGATCCATCTGGGT
GAGTTGCACCGCATTTCGTCGCCAGGAAGATTTATCCGTTCTGAAAGACAAC
CGTGAGAAAATCGAGAAAATTCTGACGTTCCGTATCCGTATTATGTCGGCCCCTG
GCGCGTGGTAATAGCCGCTCGCGTGGATGACCCGCAAATCAGAGGAACGATTAC
CCCGTGGAATTGAGGAAGTTGATAAGGGTCAAGCGCGCAGTCGTTATTG
GCGTATGACCAACTTGACAAGAATTGCCGAATGAAAAAGTCTGCCGAAGCACTC
TCTGCTGTACGAGTATTACCGTTACAACGAATTGACCAAGGTTAAATACGTAC
CGAAGGCATGCGCAAACCGGCCCTCTGAGCGCGAGCAGAAAAAGCAATCGTTG
ACCTCTGTTAAGACCAACCGCAAGGTTACGGTCAAACAACTGAAAGAGGACTATT
TCAAGAAAATTGAATGTTTGAACCGTAGAGATCTCCGGTGGTGAAGGACCGTTCA
ACCGCAGCCTGGCACCTACCATGATCTGCTGAAAATTATTAAAGACAAAGATTTC
TGGACAACGAAGAGAACGAAGATATTCTGGAAGATATCGTTCTGACCGCTGAGCTG
TCGAAGATCGTGAAGATGATTGAGGAACGCTGAAAACCTACGACACACTGTTGATG
ACAAAGTTATGAAACAGCTGAAGCGTCGTTACACAGGTTGGGCCGCTGAGCC
GTAAGCTTATCAATGGTATCCGTGACAAACAGAGCGGTAAAGACGATTCTGGACTTT
TGAAGTCAGATGGCTCGCAATCGAACCTTATGCAACTGATTGACCGACTCTC
TGACGTTCAAGGAAGATATCCAAAAGGCACAGGTGAGCGGTAGGGTGTAGCCTG
CATGAGCATATCGGAACCTGGCGGTAGCCCGCTATCAAAAGGGTATCTTACA
GACTGTGAAAGTTGGATGAATTGTTAAGGTTATGGGCGTCACAAACCGGAAA
ATATTGTGATCGAGATGGCACGTGAAAATCAGACGACGCAAAGGGTAAAAAAAT
TCTCGTGAACGCGATGAAACGTATTGAAGAGGGTATCAAAGAATTGGCAGCCAAAT
TCTGAAAGAACACCCGGTCGAGAACACCCAGCTGCAAAACGAAAAACTGTATT
ACTATCTGCAAGACGGTCGTGACATGTACGGATCAAGAACTGGACATCAATCGTT
TGAGCGATTACGATGTTGATCATATTGTCCTCAGAGCTTCTGAAAGACGATTG
TCGACAACAAAGTGTGACCCGTAGCGACAAGAACGTTAAGAGCGATAACGTG
CCGAGCGAAGAACGCTTAAGAAAATGAAAACACTGGCGTCAGCTGCTGAAACGC
CAAGCTGATTACCCAGCGTAAGTGTGATAACCTGACGAAAGCCGAGCGTGGAGGCC
TGAGCGAGCTGGACAAGGCCGCTTATCAAGCGTAACGGTGGAAACCGTCAG
ATCACTAAACATGTGGCACAGACTGGACTCCCGCATGAATACGAAATATGACGA
GAATGACAAGTTGATCCGTGAGTCAAAGTTTACGCTGAAAAGCAAACGGTGTG
CGATTTCGTAAAGACTCCAGTTCTATAAAGTCCGTGAAATCAACAACTATCATCA
CGCCCACGATGCGTACTTGAACGCTGTTGTGGGCCACCGCACTGATCAAGAAATACCC
TAAGCTGAAAGCGAGTTGTCTATGGTACTATAAAGTTACGACGTGCGTAAGAT
GATCGCCAAGAGCGAGCAAGAAATTGTAAGGCTACCGCAAAGTACTTTCTACA
GCAACATCATGAACCTCTCAAAACCGAGATTACCCGGCGAACGGTGAGATCCGTA
AACGGCCGCTGATTGAGACTAATGGCGAAACGGCGAGATTGTGTGGACAAGGGT
CGCGATTTCGCTACGGTTCGTAAGGCTCTGAGCATGCCGCAAGTTAACATTGTCAAG
AAAAGTGCAGACGGTGGCTTAGCAAAGAATCCATCCTGCCGAAGCGTAA
TAGCGATAAAACTATCGCGCGTAAAAAGACTGGGACCCAAAGAAATATGGCGGCT

TTGATAGCCGACCGTCGCGTATAGCGTGTAGGGTGCAGAAAGTTGAAAAGGGCA
 AGAGCAAGAAACTGAAGTCCGTCAAAGAACTTCTGGGTATCACCACATGGAACGT
 AGCTCCTTGAGAAGAACCCGATTGACTTCTAGAGGCCAAGGGTTATAAAGAACGT
 AAAAGACCTGATTATCAAGCTGCCAAGTACAGCCTGTTGAGTTGGAGAATGGT
 CGTAAGCGCATGCTGGCAGCGCGGGTGAGCTGCAAAGGGCAACGAACGGCGCT
 GCCGTGAAATACGTCAATTCTGTACCTGCCAGCCACTACGAAAAGCTGAAGGG
 TTCTCCGGAAGAGATAACGAACAAAAGCAACTGTTGAGCAACATAAACACTACTT
 GGACGAAATCATCGAGCAAATTAGCGAATTAGCAAACGTGTACCTGGCGGACG
 CGAATCTGGACAAGGTCTGTGCATACAATAAGCATCGCACAAACCAATTCTGT
 AGCAAGCGGAGAATATCATCCACCTGTTACGCTGACCAACCTAGGTGCCGGCG
 CATTCAAGTATTCGATACGACCATCGACCGCAAGCGCTATACCAGCACCAAAGAGG
 TCCTGGACGCGACCTGATCCACCAGAGCATTACCGGCTTATACGAAACCGTATTG
 ATTTGAGCCAACGGTGGCGATGAATTCCGAAAAAAAGCGCAAAGTTGGTGGC
 GGTGGTAGCCCAGAAAGAAACGTAAAGTG(SEQ ID NO: 63)

人类密码子优化的 GFP-eSpCas9 DNA 序列

ATGGTTAGCAAAGGTGAAGAACTGTTACAGGTGTTCCGATTCTGGTTGAACTG
 GATGGTGTGTTAATGCCACAAATTTCAGTTAGCGGTGAAGGCCAAGGTGATGCA
 ACCTATGGTAAACTGACCTGAAATTATCTGTACCCACCGCAACTGCCGTTCCG
 TGGCGACACTGGTACCAACTGACCTATGGTGTCACTGTTAGCCGTTATCCGG
 ATCACATGAAACAGCACGATTTTCAAAAGCGCAATGCCGAAAGTTATGTTCAAG
 AACGTACCATCTTCTCAAAGATGACGGCAACTATAAAACCGTGCCTGAAGTTAAAT
 TTGAAGGTGATACCTGGTGAATCGCATTGAACTGAAAGGCATCGATTAAAGAGG
 ATGGTAATATCCTGGGCCACAAACTGGAATATAATTATAATAGCCACAACGTGTACA
 TCATGGCCGACAAACAGAAAAATGGCATCAAAGTGAACCTCAAGATCGCCATAAT
 ATTGAAGATGGTCAGTCAGCTGGCGATCATTATCAGCAGAAATCCCCGATTGGT
 GATGGTCCGGTTCTGCTGCCGATAATCATTATCTGAGCCACCCAGAGCAAAACTGAGC
 AAAGATCGAATGAAAAACGTGATCACATGGTGTGCTGGAATTGTTACCGCAGCA
 GGTATTACCTTAGGTATGGATGAACTGTATAAAGTCGACAGCGGTGGTAGCGCGGT
 GGTCAGCGTAGCGAAACACCGGGTACAAGCGAAAGCGCAACACCGGAAAGCA
 GTGGTGGTAGCTAGGTGGTAGTCCGGCAGCAAAACGTGTTAAACTGGACGGTGGT
 GGTGGTAGCACCCTGATGGACAAGAAATACAGCATCGGTTGGATATTGGCACGAA
 TAGCGTGGTTGGCGTTATTACCGACGAGTACAAAGTGCCTCCAAGAAATTCAA
 AGTGCCTGGCAATACCGATGCCATAGCATCAAGAAAATCTGATTGGCGCACTGCT
 GTTCGACAGCGGTGAGACTGCCAAGCTACCGTCTGAAGCGTACGGCGCGTC
 GCTACACCCGCCGTAAGAACCGTATTGCTATCTGCAAGAAATCTCAGCAACGAAA
 TGGCAAAGTTGATGATAGCTTTTCAACGCCGCTGGAAGAGAGAGCTTCTGGTGGAAAG
 AGGATAAGAAACAGAGGCCATCCGATTGGTAACATTGTCGATGAAGTGGCAT
 ACCATGAGAAAGTACCCGACCATCTACCAACCTCGTAAGAAACTGGTGGACAGCACC
 GATAAAAGCTGATCTCGTGTGATTACCTGGCGCTGCCACATGATTAAGTTCGC
 GGTCAACTTCTGATCGAGGGCGATCTGAATCCGACAATTCTGATGTTGACAAGCTG
 TTTATTCAACTTGTACAGACCTACAACCAACAGTTGTCGAAGAGAACCGATCAATGCG
 AGCGGTGTTGATGCCAAGCAATTCTGAGCGCACGCCGAGCAAATCTGCCGTTG

GAGAACCTGATTGCACAGCTGCCGGGTGAGAAGAAAAACGGTCTGTTCGGCAATCT
GATTGCACTGTCCCTGGGCTTGACCCGAATTAAAGAGCAACTTCGACCTGGCCGA
AGATGCGAAGCTCCAATTGAGCAAAGACACCTACGACGGATGACCTGGACAATCTGC
TGGCCCAGATTGGCACCAGTACGCAGATCTGTTCTGGCTGCGAAAAACCTGAGCG
ATGCAATTCTGCTGCGACATCCTGCGCGTAATACGGAATCACGAAAGCGCTC
TGAGCGCGTCTATGATCAAGCGCTATGACGAGCACCACCAAGATCTGACCCCTGCTGA
AAGCTCTGGTGAGACAACAATTGCCAGAGAAGTATAAAGAAATTCTTGACCAGA
GCAAAAACGGCTATGCCGGTTACATTGACGGTGGCGCCAGCCAAGAAGAGTTCTAC
AAATTCAATTAAAGCCTATCCTGGAGAAAATGGATGGCACCGAAGAACTGCTGGTAAA
GCTGAATCGTGAAGATCTGCTGCGCAAACAGCGCACTTTGATAACGGTAGCATTCC
GCACCAAGATCCATCTGGGTGAGTGCACCGGATTTGCGTCGCCAGGAAGATTCTA
TCCGTTCTGAAAGACAACCGTGAGAAAATCGAGAAAATTCTGACGTTCCGTATCCC
GTATTATGTCGGCCCGCTGGCGCGTGGTAATAGCCGCTCGCGTGGATGACCCGCAA
ATCAGAGGAAACGATTACCCGTGGAATTGAGGAAGTTGTTGATAAGGGTGCAA
GCGCGCAGTCGTTCATGAGCGTATGACCAACTTGACAAGAATTGCCGAATGAAA
AAGTCTTGCAGAACGCACTCTGCTGTACGAGTATTACCGTTACAACGAATTGAC
CAAGGTTAAATACGTACCGAAGGCATGCGCAAACCGGCCCTCCTGAGCGCGGAGC
AGAAAAAAAGCAATCGTGACCTCTTAAAGACCAACCGCAAGGTTACGGTCAAAC
AACTGAAAGAGGACTATTCAAGAAAATTGAATGTTGACTCCGTAGAGATCTCCG
GTGTTGAGGAGCCGTTCAACCGCAGGCTGGCACCTACCATGATCTGCTGAAAATT
TTAAAGACAAAGATTCTGGACAACGAAGAGAACGAAGATATTCTGAAAGATATC
GTTCTGACCCCTGACGCTGTTGAAGATCGTGGAGATGATTGAGGAACGTCTGAAAACC
TACGCACACTGTTGATGACAAAGTTATGAAACAGCTGAAGCGTCGTTACACA
GGTTGGGGCCGTCTGAGCCGTAAGCTTATCAATGGTATCCGTGACAAACAGAGCGGT
AAGACGATTCTGGACTTCTGAAGTCAGATGGCTCGCCAATCGAACCTTATGCAA
CTGATTGACGACTCTGACGTTCAAGGAAGATATCCAAAAGGCACAGGTGAGC
GGTCAGGGTGATAGCCTGCATGAGCATATCGCGAACCTGGCGGTAGCCCGCTATC
AAAAAGGGTATCTTACAGACTGTGAAAGTTGGATGAATTGGTAAGGTTATGGGT
CGTCACAAACCGAAAATTGTGATCGAGATGGCACGTGAAAATCAGACGACGCA
AAAGGGTCAAAAAAATTCTCGTGAGCGCATGAAACGTATTGAAGAGGGTATCAAAG
AATTGGCAGCCAATTCTGAAAGAACACCCGGTCGAGAACACCCAGCTGAAAAC
AAAAAACTGTATTATACTATCTGAGCAACGGCTGTGACATGTACGTGGATCAAGAA
CTGGACATCAATCGTTGAGCGATTACGATGTTGATCATATTGTCCTCAGAGCTTC
TGGCGGACGATTGATCGACAACAAAGTGCTGACCGTAGCGACAAGAATCGTGGT
AAGAGCGATAACGTGCCGAGCGAAGAAGTCGTTAAGAAAATGAAAATACTGGCG
TCAGCTGCTGAACGCCAAGCTGATTACCCAGCGTAAGTCGATAACCTGACGAAAGC
CGAGCGTGGAGGCCTGAGCGAGCTGGACAAGGCCGGCTTATCAAGCGTCAACTGG
TGGAAACCGTCAGATCACTAACATGTGGCACAGATCCTGGACTCCCGCATGAATA
CGAAATATGACGAGAATGACAAGTTGATCCGTGAAGTCAAAGTTTACGCTGAAA
AGCAAACCTGGTGTCCGATTCCGAAAGACTTCCAGTTCTATAAAGTCCGTGAAATC
AACAACTATCATCACGCCACGATCGTACTTGAACGCTGTTGGCACCACGCACTG
ATCAAGAAATACCCCTGCACTCGAAAGCGAGTTGTCTGGTACTATAAAGTTAC

```

GACGTGCGTAAGATGATGCCAAGAGCGAGCAAGAAATTGGTAAGGCTACCGCAAA
GTACTTTTCTACAGCAACATCATGAACCTCTTCAAAACCGAGATTACCCTGGCGAA
CGGTGAGATCCGTAAAGCGCCGCTGATTGAGACTAATGGCAGAACGGGCGAGATTG
TGTGGACAAGGGTCGCGATTCGCTACGGTCGTAAGGTCTGAGCATGCCGCAAG
TTAACATTGTCAAGAAAAGTGAAGTGCAGACGGGTGGCTTAGCAAAGAATCCATCC
TGCGAAGCGTAATAGCGATAAAACTTATCGCGTAAAGAAGACTGGGACCCAAAG
AAATATGGCGGCTTGATAGCCGACCGTCGCGTATAGCGTGTAGTGGTCGCGAAA
GTTGAAAAGGGCAAGAGCAAGAAACTGAAGTCCGTCAAAGAACTTCTGGGTATCAC
CATCATGGAACGTAGCTCCTTGAGAAGAACCGATTGACTTCTAGAGGCGAAGGG
TTATAAAGAAGTCAAAAAAGACCTGATTATCAAGCTGCCGAAGTACAGCCTGTTGA
GTTGGAGAATGGTCGTAAGCGATGCTGGCGAGCGCGGGTGAGCTGCAAAAGGGCA
ACGAACCTGGCGCTGCCGTCAAATACGTCAATTCTGTACCTGCCAGCCACTACG
AAAAGCTGAAGGGTCTCCGGAAAGATAACGAACAAAGCAACTGTTGAGCAA
CATAAACACTACTGGACGAAATCATCGAGCAAATTAGCGAATTAGCAAACGTGTC
ATCCTGGCGGACGCGAATCTGGACAAGGTCTGTGCATACAATAAGCATCGCGAC
AAACCAATTGAGCAAGCGGAGAATATCATCCACCTGTTACGCTGACCAACCTA
GGTGCGCCGGCGGCATTCAAGTATTCGATACGACCATCGACCGCAAGCGCTATACC
AGCACCAAAGAGGTCTGGACGCGACCCGTATCCACCAAGAGCATTACCGGCTTATAC
GAAACCGTATTGATTGAGCCAATGGGTGGCGATGAATTCCCAAAAAAAAAGCG
CAAAGTTGGTGGCGGTGGTAGCCCCAAAAAGAAACGTAAAGTG(SEQ ID NO: 64)

```

实施例2:与商业产品编辑效率的比较

两种商业GFP-SpCas9融合蛋白产品,GenCrispr NLS-Cas9-EGFP核酸酶和ArciTect Cas9-eGFP核酸酶,分别购买自GenScript(皮斯卡塔韦,新泽西州)以及Stemcell Technologies(温哥华,加拿大)。用50 pmol GenCrispr NLS-Cas9-EGFP核酸酶,或ArciTect Cas9-eGFP核酸酶,或本发明的GFP-SpCas9蛋白,结合靶向人EMX1、HEKsite4、VEGFA3、HPRT基因座的四种化学合成的sgRNA各150 pmol,对人U2OS细胞(0.2×10^6)和HEK293细胞(0.3×10^6)进行转染。向导序列为:5'-GAGUCCGAGCAGAAGAAGAA-3' (EMX1) (SEQ ID NO: 51)、5'-GGCACUGCGGCUGGAGGUGG-3' (HEKsite4) (SEQ ID NO: 52)、5'-GGUGAGUGAGUGUGCGUG-3' (VEGFA3) 和5'-GGUCACUUUAACACACCCA-3' (HPRT) (SEQ ID NO: 53)。使用核转染溶液V和Amaxa仪器实施转染。在收获细胞以进行基因编辑分析前,细胞保持在37℃和5% CO₂三天。使用QuickExtract DNA抽提溶液制备基因组DNA。使用由目标特异性序列和下一代测序(NGS)接头组成的引物对PCR扩增各靶向基因组区域。引物列在下表中:

NGS 引物序列	
目标	引物序列(5'-3')
EMX1	正向: TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNNNCCCA GTGGCTGCTCT(SEQ ID NO: 54) 反向: GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNNNNNCCAG GCCTCCCCAAAGC(SEQ ID NO: 55)
HEKSite4	正向: TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNNNNGAAC CCAGGTAGGCCAGAGA(SEQ ID NO: 56) 反向: GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNNNNNGGGG TGGGGTCAGACGT(SEQ ID NO: 57)
VEGFA3	正向: TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNNNCCCA TTCCCTTTAGCCA(SEQ ID NO: 58) 反向: GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNNNNNGGAG CAGGAAAGTGAGGTTAC(SEQ ID NO: 59)
HPRT	正向: TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNNNNAATGG ACACATGGGTAGTCAGG(SEQ ID NO: 60) 反向: GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNNNNNGGCT TATATCCAACACTTCGTGGG(SEQ ID NO: 61)

通过使用 Illumina MiSeq 的 NGS 分析 PCR 扩增子, 以测定各 Cas9 蛋白的编辑效率。图 2A 和图 2B 中的结果显示, 在所有目标中本发明的 GFP-SpCas9 蛋白的编辑效率均为商业蛋白的数倍高。

序列表

<110> 西格马-奥尔德里奇有限责任公司

<120> CRISPR/CAS 融合蛋白和系统

<130> P19-027 WO-PCT

<140>

<141>

<150> 62/806, 708

<151> 2019-02-15

<160> 66

<170> PatentIn version 3.5

<210> 1

<211> 7

<212> PRT

<213> 人工序列

<220>

<221> 来源

[0001] <223> /注="人工序列描述: 合成肽"

<400> 1

Pro Lys Lys Lys Arg Lys Val

1 5

<210> 2

<211> 7

<212> PRT

<213> 人工序列

<220>

<221> 来源

<223> /注="人工序列描述: 合成肽"

<400> 2

Pro Lys Lys Lys Arg Arg Val

1 5

<210> 3

<211> 16

<212> PRT

<213> 人工序列

<220>

<221> 来源

<223> /注="人工序列描述: 合成肽"

<400> 3

Lys Arg Pro Ala Ala Thr Lys Lys Ala Gly Gln Ala Lys Lys Lys Lys
1 5 10 15

<210> 4

<211> 11

<212> PRT

<213> 人工序列

<220>

<221> 来源

<223> /注="人工序列描述: 合成肽"

<400> 4

Tyr Gly Arg Lys Lys Arg Arg Gln Arg Arg Arg
1 5 10

<210> 5

[0002] <211> 9

<212> PRT

<213> 人工序列

<220>

<221> 来源

<223> /注="人工序列描述: 合成肽"

<400> 5

Arg Lys Lys Arg Arg Gln Arg Arg Arg
1 5

<210> 6

<211> 9

<212> PRT

<213> 人工序列

<220>

<221> 来源

<223> /注="人工序列描述: 合成肽"

<400> 6

Pro Ala Ala Lys Arg Val Lys Leu Asp
1 5

<210> 7
<211> 11
<212> PRT
<213> 人工序列

<220>
<221> 来源
<223> /注="人工序列描述: 合成肽"

<400> 7
Arg Gln Arg Arg Asn Glu Leu Lys Arg Ser Pro
1 5 10

<210> 8
<211> 8
<212> PRT
<213> 人工序列

<220>
<221> 来源
<223> /注="人工序列描述: 合成肽"

<400> 8
[0003] Val Ser Arg Lys Arg Pro Arg Pro
1 5

<210> 9
<211> 8
<212> PRT
<213> 人工序列

<220>
<221> 来源
<223> /注="人工序列描述: 合成肽"

<400> 9
Pro Pro Lys Lys Ala Arg Glu Asp
1 5

<210> 10
<211> 8
<212> PRT
<213> 人工序列

<220>
<221> 来源
<223> /注="人工序列描述: 合成肽"

<400> 10
Pro Gln Pro Lys Lys Lys Pro Leu
1 5

<210> 11
<211> 12
<212> PRT
<213> 人工序列

<220>
<221> 来源
<223> /注="人工序列描述: 合成肽"

<400> 11
Ser Ala Leu Ile Lys Lys Lys Lys Met Ala Pro
1 5 10

<210> 12
<211> 7
<212> PRT
<213> 人工序列

[0004] <220>
<221> 来源
<223> /注="人工序列描述: 合成肽"

<400> 12
Pro Lys Gln Lys Lys Arg Lys
1 5

<210> 13
<211> 10
<212> PRT
<213> 人工序列

<220>
<221> 来源
<223> /注="人工序列描述: 合成肽"

<400> 13
Arg Lys Leu Lys Lys Lys Ile Lys Lys Leu
1 5 10

<210> 14
<211> 10
<212> PRT
<213> 人工序列

<220>

<221> 来源

<223> /注="人工序列描述: 合成肽"

<400> 14

Arg Glu Lys Lys Lys Phe Leu Lys Arg Arg
1 5 10

<210> 15

<211> 20

<212> PRT

<213> 人工序列

<220>

<221> 来源

<223> /注="人工序列描述: 合成肽"

<400> 15

Lys Arg Lys Gly Asp Glu Val Asp Gly Val Asp Glu Val Ala Lys Lys
1 5 10 15

Lys Ser Lys Lys

[0005] 20

<210> 16

<211> 17

<212> PRT

<213> 人工序列

<220>

<221> 来源

<223> /注="人工序列描述: 合成肽"

<400> 16

Arg Lys Cys Leu Gln Ala Gly Met Asn Leu Glu Ala Arg Lys Thr Lys
1 5 10 15

Lys

<210> 17

<211> 38

<212> PRT

<213> 人工序列

<220>

<221> 来源

<223> /注="人工序列描述: 合成多肽"

<400> 17

Asn	Gln	Ser	Ser	Asn	Phe	Gly	Pro	Met	Lys	Gly	Gly	Asn	Phe	Gly	Gly
1															
				5					10						15

Arg	Ser	Ser	Gly	Pro	Tyr	Gly	Gly	Gly	Gln	Tyr	Phe	Ala	Lys	Pro
									20					30
										25				

Arg	Asn	Gln	Gly	Gly	Tyr
					35

<210> 18

<211> 42

<212> PRT

<213> 人工序列

<220>

<221> 来源

<223> /注="人工序列描述: 合成多肽"

[0006]

<400> 18

Arg	Met	Arg	Ile	Glx	Phe	Lys	Asn	Lys	Gly	Lys	Asp	Thr	Ala	Glu	Leu
1															15
					5						10				

Arg	Arg	Arg	Arg	Val	Glu	Val	Ser	Val	Glu	Leu	Arg	Lys	Ala	Lys	Lys
									20					30	
										25					

Asp	Glu	Gln	Ile	Leu	Lys	Arg	Arg	Asn	Val				
									35				
										40			

<210> 19

<211> 239

<212> PRT

<213> 人工序列

<220>

<221> 来源

<223> /注="人工序列描述: 合成多肽"

<400> 19

Met	Val	Ser	Lys	Gly	Glu	Glu	Leu	Phe	Thr	Gly	Val	Val	Pro	Ile	Leu
1															15
								5							
										10					

Val Glu Leu Asp Gly Asp Val Asn Gly His Lys Phe Ser Val Ser Gly
 20 25 30

Glu Gly Glu Gly Asp Ala Thr Tyr Gly Lys Leu Thr Leu Lys Phe Ile
 35 40 45

Cys Thr Thr Gly Lys Leu Pro Val Pro Trp Pro Thr Leu Val Thr Thr
 50 55 60

Leu Thr Tyr Gly Val Gln Cys Phe Ser Arg Tyr Pro Asp His Met Lys
 65 70 75 80

Gln His Asp Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr Val Gln Glu
 85 90 95

Arg Thr Ile Phe Phe Lys Asp Asp Gly Asn Tyr Lys Thr Arg Ala Glu
 100 105 110

[0007] Val Lys Phe Glu Gly Asp Thr Leu Val Asn Arg Ile Glu Leu Lys Gly
 115 120 125

Ile Asp Phe Lys Glu Asp Gly Asn Ile Leu Gly His Lys Leu Glu Tyr
 130 135 140

Asn Tyr Asn Ser His Asn Val Tyr Ile Met Ala Asp Lys Gln Lys Asn
 145 150 155 160

Gly Ile Lys Val Asn Phe Lys Ile Arg His Asn Ile Glu Asp Gly Ser
 165 170 175

Val Gln Leu Ala Asp His Tyr Gln Gln Asn Thr Pro Ile Gly Asp Gly
 180 185 190

Pro Val Leu Leu Pro Asp Asn His Tyr Leu Ser Thr Gln Ser Lys Leu
 195 200 205

Ser Lys Asp Pro Asn Glu Lys Arg Asp His Met Val Leu Leu Glu Phe

210

215

220

Val Thr Ala Ala Gly Ile Thr Leu Gly Met Asp Glu Leu Tyr Lys
 225 230 235

<210> 20

<211> 232

<212> PRT

<213> 人工序列

<220>

<221> 来源

<223> /注=“人工序列描述：合成多肽”

<400> 20

Met	Val	Ser	Lys	Gly	Glu	Ala	Val	Ile	Lys	Glu	Phe	Met	Arg	Phe	Lys
1															
															15

Val His Met Glu Gly Ser Met Asn Gly His Glu Phe Glu Ile Glu Gly
 20 25 30

[0008] Glu Gly Glu Gly Arg Pro Tyr Glu Gly Thr Gln Thr Ala Lys Leu Lys
 35 40 45

Val Thr Lys Gly Gly Pro Leu Pro Phe Ser Trp Asp Ile Leu Ser Pro
 50 55 60

Gln Phe Met Tyr Gly Ser Arg Ala Phe Thr Lys His Pro Ala Asp Ile
 65 70 75 80

Pro Asp Tyr Tyr Lys Gln Ser Phe Pro Glu Gly Phe Lys Trp Glu Arg
 85 90 95

Val Met Asn Phe Glu Asp Gly Gly Ala Val Thr Val Thr Gln Asp Thr
 100 105 110

Ser Leu Glu Asp Gly Thr Leu Ile Tyr Lys Val Lys Leu Arg Gly Thr
 115 120 125

Asn Phe Pro Pro Asp Gly Pro Val Met Gln Lys Lys Thr Met Gly Trp
 130 135 140

Glu Ala Ser Thr Glu Arg Leu Tyr Pro Glu Asp Gly Val Leu Lys Gly
 145 150 155 160

Asp Ile Lys Met Ala Leu Arg Leu Lys Asp Gly Gly Arg Tyr Leu Ala
 165 170 175

Asp Phe Lys Thr Thr Tyr Lys Ala Lys Lys Pro Val Gln Met Pro Gly
 180 185 190

Ala Tyr Asn Val Asp Arg Lys Leu Asp Ile Thr Ser His Asn Glu Asp
 195 200 205

Tyr Thr Val Val Glu Gln Tyr Glu Arg Ser Glu Gly Arg His Ser Thr
 210 215 220

Gly Gly Met Asp Glu Leu Tyr Lys
 225 230

[0009] <210> 21
 <211> 8
 <212> PRT
 <213> 人工序列
 <220>
 <221> 来源
 <223> /注="人工序列描述: 合成肽"
 <400> 21
 Asp Tyr Lys Asp Asp Asp Lys
 1 5

<210> 22
 <211> 6
 <212> PRT
 <213> 人工序列
 <220>

<221> 来源
 <223> /注="人工序列描述: 合成 6xHis 标签"

<400> 22
 His His His His His

1 5

<210> 23
<211> 20
<212> PRT
<213> 人工序列

<220>
<221> 来源
<223> /注=“人工序列描述：合成肽”

<400> 23
Gly Arg Lys Lys Arg Arg Gln Arg Arg Arg Pro Pro Gln Pro Lys Lys
1 5 10 15

Lys Arg Lys Val
20

<210> 24
<211> 19
<212> PRT
<213> 人工序列

[0010]

<220>
<221> 来源
<223> /注=“人工序列描述：合成肽”

<400> 24
Pro Leu Ser Ser Ile Phe Ser Arg Ile Gly Asp Pro Pro Lys Lys
1 5 10 15

Arg Lys Val

<210> 25
<211> 24
<212> PRT
<213> 人工序列

<220>
<221> 来源
<223> /注=“人工序列描述：合成肽”

<400> 25
Gly Ala Leu Phe Leu Gly Trp Leu Gly Ala Ala Gly Ser Thr Met Gly
1 5 10 15

Ala Pro Lys Lys Lys Arg Lys Val
20

<210> 26
<211> 27
<212> PRT
<213> 人工序列

<220>
<221> 来源
<223> /注="人工序列描述: 合成肽"

<400> 26
Gly Ala Leu Phe Leu Gly Phe Leu Gly Ala Ala Gly Ser Thr Met Gly
1 5 10 15

Ala Trp Ser Gln Pro Lys Lys Arg Lys Val
20 25

[0011] <210> 27
<211> 21
<212> PRT
<213> 人工序列

<220>
<221> 来源
<223> /注="人工序列描述: 合成肽"

<400> 27
Lys Glu Thr Trp Trp Glu Thr Trp Trp Thr Glu Trp Ser Gln Pro Lys
1 5 10 15

Lys Lys Arg Lys Val
20

<210> 28
<211> 11
<212> PRT
<213> 人工序列

<220>
<221> 来源
<223> /注="人工序列描述: 合成肽"

<400> 28
Tyr Ala Arg Ala Ala Arg Gln Ala Arg Ala
1 5 10

<210> 29
<211> 11
<212> PRT
<213> 人工序列

<220>
<221> 来源
<223> /注="人工序列描述: 合成肽"

<400> 29
Thr His Arg Leu Pro Arg Arg Arg Arg Arg Arg
1 5 10

<210> 30
<211> 11
<212> PRT
<213> 人工序列

[0012] <220>
<221> 来源
<223> /注="人工序列描述: 合成肽"

<400> 30
Gly Gly Arg Arg Ala Arg Arg Arg Arg Arg Arg
1 5 10

<210> 31
<211> 12
<212> PRT
<213> 人工序列

<220>
<221> 来源
<223> /注="人工序列描述: 合成肽"

<400> 31
Arg Arg Gln Arg Arg Thr Ser Lys Leu Met Lys Arg
1 5 10

<210> 32
<211> 27
<212> PRT
<213> 人工序列

<220>

<221> 来源

<223> /注="人工序列描述: 合成肽"

<400> 32

Gly	Trp	Thr	Leu	Asn	Ser	Ala	Gly	Tyr	Leu	Leu	Gly	Lys	Ile	Asn	Leu
1															
		5							10					15	

Lys	Ala	Leu	Ala	Ala	Leu	Ala	Lys	Lys	Ile	Leu
		20					25			

<210> 33

<211> 33

<212> PRT

<213> 人工序列

<220>

<221> 来源

<223> /注="人工序列描述: 合成多肽"

<400> 33

Lys	Ala	Leu	Ala	Trp	Glu	Ala	Lys	Leu	Ala	Lys	Ala	Lys	Ala	
[0013]	1												15	

Leu	Ala	Lys	His	Leu	Ala	Lys	Ala	Leu	Ala	Lys	Ala	Leu	Lys	Cys	Glu
		20					25						30		

Ala

<210> 34

<211> 16

<212> PRT

<213> 人工序列

<220>

<221> 来源

<223> /注="人工序列描述: 合成肽"

<400> 34

Arg	Gln	Ile	Lys	Ile	Trp	Phe	Gln	Asn	Arg	Arg	Met	Lys	Trp	Lys	Lys
1															
				5					10				15		

<210> 35

<211> 46

<212> PRT

<213> 人工序列

<220>

<221> 来源

<223> /注="人工序列描述: 合成多肽"

<400> 35

Ala Glu Ala Ala Ala Lys Glu Ala Ala Ala Lys Glu Ala Ala Ala Lys
1 5 10 15

Glu Ala Ala Ala Lys Ala Leu Glu Ala Glu Ala Ala Ala Lys Glu Ala
20 25 30

Ala Ala Lys Glu Ala Ala Ala Lys Glu Ala Ala Ala Lys Ala
35 40 45

<210> 36

<211> 32

<212> PRT

<213> 人工序列

[0014]

<220>

<221> 来源

<223> /注="人工序列描述: 合成多肽"

<400> 36

Ser Gly Gly Ser Ser Gly Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr
1 5 10 15

Ser Glu Ser Ala Thr Pro Glu Ser Ser Gly Gly Ser Ser Gly Gly Ser
20 25 30

<210> 37

<211> 6

<212> PRT

<213> 人工序列

<220>

<221> 来源

<223> /注="人工序列描述: 合成肽"

<400> 37

Leu Glu Gly Gly Ser

1 5

<210> 38
 <211> 4
 <212> PRT
 <213> 人工序列

<220>
 <221> 来源
 <223> /注="人工序列描述: 合成肽"

<400> 38
 Thr Gly Ser Gly
 1

<210> 39
 <211> 8
 <212> PRT
 <213> 人工序列

<220>
 <221> 来源
 <223> /注="人工序列描述: 合成肽"

<400> 39
 [0015] Gly Gly Ser Gly Gly Gly Ser Gly
 1 5

<210> 40
 <211> 20
 <212> PRT
 <213> 人工序列

<220>
 <221> 来源
 <223> /注="人工序列描述: 合成肽"

<220>
 <221> 位点
 <222> (1)..(20)
 <223> /注="该序列可能包含 1-4 个'Gly Gly Gly Ser' 重复单元"

<400> 40
 Gly Gly Gly Gly Ser Gly Gly Gly Ser Gly Gly Gly Ser Gly
 1 5 10 15

Gly Gly Gly Ser

20

<210> 41
<211> 8
<212> PRT
<213> 人工序列

<220>
<221> 来源
<223> /注=“人工序列描述：合成肽”

<220>
<221> 位点
<222> (1)..(8)
<223> /注=“该序列可能包含 6-8 个残基”

<400> 41
Gly Gly Gly Gly Gly Gly Gly
1 5

[0016] <210> 42
<211> 20
<212> PRT
<213> 人工序列

<220>
<221> 来源
<223> /注=“人工序列描述：合成肽”

<220>
<221> 位点
<222> (1)..(20)
<223> /注=“该序列可能包含 1-4 个'Glu Ala Ala Ala Lys' 重复单元”

<400> 42
Glu Ala Ala Ala Lys Glu Ala Ala Ala Lys Glu Ala Ala Ala Lys Glu
1 5 10 15

Ala Ala Ala Lys
20

<210> 43
<211> 27
<212> PRT
<213> 人工序列

<220>

<221> 来源

<223> /注="人工序列描述: 合成肽"

<220>

<221> 位点

<222> (2)..(26)

<223> /注="该区域可能包含 2-5 个'Glu Ala Ala Ala Lys' 重复单元"

<400> 43

Ala Glu Ala Ala Ala Lys Glu Ala Ala Ala Lys Glu Ala Ala Ala Lys

1

5

10

15

Glu Ala Ala Ala Lys Glu Ala Ala Ala Lys Ala

20

25

<210> 44

<211> 5

<212> PRT

<213> 人工序列

[0017]

<220>

<221> 来源

<223> /注="人工序列描述: 合成肽"

<400> 44

Pro Ala Pro Ala Pro

1

5

<210> 45

<211> 16

<212> PRT

<213> 人工序列

<220>

<221> 来源

<223> /注="人工序列描述: 合成肽"

<220>

<221> 位点

<222> (1)..(16)

<223> /注="该序列可能包含 6-8 个'Ala Pro' 重复单元"

<400> 45

Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro

1

5

10

15

<210> 46

<211> 60

<212> DNA

<213> 人工序列

<220>

<221> 来源

<223> /注="人工序列描述: 合成引物"

<220>

<221> 修饰的碱基

<222> (34)..(39)

<223> a, t, g 或 c

<400> 46

tcgtcggcag cgtcagatgt gtataagaga cagnnnnna gtcttccat caggctctca

60

<210> 47

<211> 57

<212> DNA

[0018] <213> 人工序列

<220>

<221> 来源

<223> /注="人工序列描述: 合成引物"

<220>

<221> 修饰的碱基

<222> (35)..(40)

<223> a, t, g 或 c

<400> 47

gtctcggtgg ctggagatg tgtataagag acagnnnnn agagtccagc ttgggcc

57

<210> 48

<211> 1692

<212> PRT

<213> 人工序列

<220>

<221> 来源

<223> /注="人工序列描述: 合成多肽"

<400> 48

Met Val Ser Lys Gly Glu Glu Leu Phe Thr Gly Val Val Pro Ile Leu
 1 5 10 15

Val Glu Leu Asp Gly Asp Val Asn Gly His Lys Phe Ser Val Ser Gly
 20 25 30

Glu Gly Glu Gly Asp Ala Thr Tyr Gly Lys Leu Thr Leu Lys Phe Ile
 35 40 45

Cys Thr Thr Gly Lys Leu Pro Val Pro Trp Pro Thr Leu Val Thr Thr
 50 55 60

Leu Thr Tyr Gly Val Gln Cys Phe Ser Arg Tyr Pro Asp His Met Lys
 65 70 75 80

Gln His Asp Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr Val Gln Glu
 85 90 95

[0019] Arg Thr Ile Phe Phe Lys Asp Asp Gly Asn Tyr Lys Thr Arg Ala Glu
 100 105 110

Val Lys Phe Glu Gly Asp Thr Leu Val Asn Arg Ile Glu Leu Lys Gly
 115 120 125

Ile Asp Phe Lys Glu Asp Gly Asn Ile Leu Gly His Lys Leu Glu Tyr
 130 135 140

Asn Tyr Asn Ser His Asn Val Tyr Ile Met Ala Asp Lys Gln Lys Asn
 145 150 155 160

Gly Ile Lys Val Asn Phe Lys Ile Arg His Asn Ile Glu Asp Gly Ser
 165 170 175

Val Gln Leu Ala Asp His Tyr Gln Gln Asn Thr Pro Ile Gly Asp Gly
 180 185 190

Pro Val Leu Leu Pro Asp Asn His Tyr Leu Ser Thr Gln Ser Lys Leu
 195 200 205

Ser Lys Asp Pro Asn Glu Lys Arg Asp His Met Val Leu Leu Glu Phe
 210 215 220

Val Thr Ala Ala Gly Ile Thr Leu Gly Met Asp Glu Leu Tyr Lys Val
 225 230 235 240

Asp Ala Glu Ala Ala Ala Lys Glu Ala Ala Ala Lys Glu Ala Ala Ala
 245 250 255

Lys Glu Ala Ala Ala Lys Ala Leu Glu Ala Glu Ala Ala Lys Glu
 260 265 270

Ala Ala Ala Lys Glu Ala Ala Ala Lys Glu Ala Ala Ala Lys Ala Pro
 275 280 285

Ala Ala Lys Arg Val Lys Leu Asp Gly Gly Gly Ser Thr Gly Met
 290 295 300

[0020] Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val Gly
 305 310 315 320

Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe Lys
 325 330 335

Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile Gly
 340 345 350

Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys
 355 360 365

Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys Tyr
 370 375 380

Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser Phe
 385 390 395 400

Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys His
 405 410 415

Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr His
 420 425 430

Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp Ser
 435 440 445

Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His Met
 450 455 460

Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp
 465 470 475 480

Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn
 485 490 495

Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala Lys
 500 505 510

[0021] Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu
 515 520 525

Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu
 530 535 540

Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp
 545 550 555 560

Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp
 565 570 575

Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu
 580 585 590

Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile
 595 600 605

Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met

610

615

620

Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys Ala
 625 630 635 640

Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp
 645 650 655

Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln
 660 665 670

Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly
 675 680 685

Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys
 690 695 700

[0022] Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu Gly
 705 710 715 720

Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu
 725 730 735

Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro
 740 745 750

Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp Met
 755 760 765

Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu Val
 770 775 780

Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr Asn
 785 790 795 800

Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser Leu
 805 810 815

Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr
 820 825 830

Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys
 835 840 845

Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val
 850 855 860

Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser
 865 870 875 880

Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr
 885 890 895

Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn
 900 905 910

[0023] Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr Leu
 915 920 925

Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala His
 930 935 940

Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr
 945 950 955 960

Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp Lys
 965 970 975

Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe Ala
 980 985 990

Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe Lys
 995 1000 1005

Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
 1010 1015 1020

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys
1025 1030 1035

Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val
1040 1045 1050

Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg
1055 1060 1065

Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg
1070 1075 1080

Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile
1085 1090 1095

Leu Lys Glu His Pro Val Glu Asn Thr Gln Leu Gln Asn Glu Lys
1100 1105 1110

[0024] Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val Asp
1115 1120 1125

Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp Val Asp His
1130 1135 1140

Ile Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile Asp Asn Lys
1145 1150 1155

Val Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp Asn Val
1160 1165 1170

Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg Gln
1175 1180 1185

Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn Leu
1190 1195 1200

Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys Ala Gly
1205 1210 1215

Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys His
 1220 1225 1230

Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu
 1235 1240 1245

Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
 1250 1255 1260

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val
 1265 1270 1275

Arg Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn
 1280 1285 1290

Ala Val Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu
 1295 1300 1305

[0025] Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys
 1310 1315 1320

Met Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys
 1325 1330 1335

Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile
 1340 1345 1350

Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr
 1355 1360 1365

Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe
 1370 1375 1380

Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val
 1385 1390 1395

Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile

1400

1405

1410

Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp
 1415 1420 1425

Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala
 1430 1435 1440

Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys
 1445 1450 1455

Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu
 1460 1465 1470

Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys
 1475 1480 1485

Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys
 [0026] 1490 1495 1500

Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala
 1505 1510 1515

Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser
 1520 1525 1530

Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu
 1535 1540 1545

Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu
 1550 1555 1560

Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu
 1565 1570 1575

Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val
 1580 1585 1590

Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln
 1595 1600 1605

Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala
 1610 1615 1620

Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg
 1625 1630 1635

Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln
 1640 1645 1650

Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu
 1655 1660 1665

Gly Gly Asp Glu Phe Pro Lys Lys Lys Arg Lys Val Gly Gly Gly
 1670 1675 1680

Gly Ser Pro Lys Lys Arg Lys Val
 [0027] 1685 1690

<210> 49

<211> 1671

<212> PRT

<213> 人工序列

<220>

<221> 来源

<223> /注=“人工序列描述：合成多肽”

<400> 49

Met Val Ser Lys Gly Glu Ala Val Ile Lys Glu Phe Met Arg Phe Lys
 1 5 10 15

Val His Met Glu Gly Ser Met Asn Gly His Glu Phe Glu Ile Glu Gly
 20 25 30

Glu Gly Glu Gly Arg Pro Tyr Glu Gly Thr Gln Thr Ala Lys Leu Lys
 35 40 45

Val Thr Lys Gly Gly Pro Leu Pro Phe Ser Trp Asp Ile Leu Ser Pro

50	55	60
----	----	----

Gln Phe Met Tyr Gly Ser Arg Ala Phe Thr Lys His Pro Ala Asp Ile
 65 70 75 80

Pro Asp Tyr Tyr Lys Gln Ser Phe Pro Glu Gly Phe Lys Trp Glu Arg
 85 90 95

Val Met Asn Phe Glu Asp Gly Gly Ala Val Thr Val Thr Gln Asp Thr
 100 105 110

Ser Leu Glu Asp Gly Thr Leu Ile Tyr Lys Val Lys Leu Arg Gly Thr
 115 120 125

Asn Phe Pro Pro Asp Gly Pro Val Met Gln Lys Lys Thr Met Gly Trp
 130 135 140

[0028] Glu Ala Ser Thr Glu Arg Leu Tyr Pro Glu Asp Gly Val Leu Lys Gly
 145 150 155 160

Asp Ile Lys Met Ala Leu Arg Leu Lys Asp Gly Gly Arg Tyr Leu Ala
 165 170 175

Asp Phe Lys Thr Thr Tyr Lys Ala Lys Lys Pro Val Gln Met Pro Gly
 180 185 190

Ala Tyr Asn Val Asp Arg Lys Leu Asp Ile Thr Ser His Asn Glu Asp
 195 200 205

Tyr Thr Val Val Glu Gln Tyr Glu Arg Ser Glu Gly Arg His Ser Thr
 210 215 220

Gly Gly Met Asp Glu Leu Tyr Lys Val Asp Ser Gly Gly Ser Ser Gly
 225 230 235 240

Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro
 245 250 255

Glu Ser Ser Gly Gly Ser Ser Gly Gly Ser Pro Ala Ala Lys Arg Val
260 265 270

Lys Leu Asp Gly Gly Ser Thr Gly Met Asp Lys Lys Tyr Ser
275 280 285

Ile Gly Leu Asp Ile Gly Thr Asn Ser Val Gly Trp Ala Val Ile Thr
290 295 300

Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe Lys Val Leu Gly Asn Thr
305 310 315 320

Asp Arg His Ser Ile Lys Lys Asn Leu Ile Gly Ala Leu Leu Phe Asp
325 330 335

Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg
340 345 350

[0029] Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe
355 360 365

Ser Asn Glu Met Ala Lys Val Asp Asp Ser Phe Phe His Arg Leu Glu
370 375 380

Glu Ser Phe Leu Val Glu Glu Asp Lys Lys His Glu Arg His Pro Ile
385 390 395 400

Phe Gly Asn Ile Val Asp Glu Val Ala Tyr His Glu Lys Tyr Pro Thr
405 410 415

Ile Tyr His Leu Arg Lys Lys Leu Val Asp Ser Thr Asp Lys Ala Asp
420 425 430

Leu Arg Leu Ile Tyr Leu Ala Leu Ala His Met Ile Lys Phe Arg Gly
435 440 445

His Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp Val Asp
450 455 460

Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu
465 470 475 480

Asn Pro Ile Asn Ala Ser Gly Val Asp Ala Lys Ala Ile Leu Ser Ala
485 490 495

Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu Ile Ala Gln Leu Pro
500 505 510

Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu
515 520 525

Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala
530 535 540

Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu
545 550 555 560

[0030] Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys
565 570 575

Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile Leu Arg Val Asn Thr
580 585 590

Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp
595 600 605

Glu His His Gln Asp Leu Thr Leu Leu Lys Ala Leu Val Arg Gln Gln
610 615 620

Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly
625 630 635 640

Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln Glu Glu Phe Tyr Lys
645 650 655

Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly Thr Glu Glu Leu Leu
660 665 670

Val Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp
675 680 685

Asn Gly Ser Ile Pro His Gln Ile His Leu Gly Glu Leu His Ala Ile
690 695 700

Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu
705 710 715 720

Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro
725 730 735

Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp Met Thr Arg Lys Ser Glu
740 745 750

Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu Val Val Asp Lys Gly Ala
755 760 765

[0031] Ser Ala Gln Ser Phe Ile Glu Arg Met Thr Asn Phe Asp Lys Asn Leu
770 775 780

Pro Asn Glu Lys Val Leu Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe
785 790 795 800

Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr Val Thr Glu Gly Met
805 810 815

Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys Lys Ala Ile Val Asp
820 825 830

Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val Lys Gln Leu Lys Glu
835 840 845

Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser Val Glu Ile Ser Gly
850 855 860

Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr Tyr His Asp Leu Leu

865	870	875	880
Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp			
885		890	895
Ile Leu Glu Asp Ile Val Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu			
900		905	910
Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala His Leu Phe Asp Asp Lys			
915		920	925
Val Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu			
930		935	940
Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr			
945		950	955
Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn Phe Met [0032]			
965		970	975
Gln Leu Ile His Asp Asp Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys			
980		985	990
Ala Gln Val Ser Gly Gln Gly Asp Ser Leu His Glu His Ile Ala Asn			
995		1000	1005
Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr Val			
1010		1015	1020
Lys Val Val Asp Glu Leu Val Lys Val Met Gly Arg His Lys Pro			
1025		1030	1035
Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr Thr Gln			
1040		1045	1050
Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu Glu			
1055		1060	1065

Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val
1070 1075 1080

Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
1085 1090 1095

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn
1100 1105 1110

Arg Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe
1115 1120 1125

Leu Lys Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp
1130 1135 1140

Lys Asn Arg Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val
1145 1150 1155

[0033] Lys Lys Met Lys Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu
1160 1165 1170

Ile Thr Gln Arg Lys Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly
1175 1180 1185

Gly Leu Ser Glu Leu Asp Lys Ala Gly Phe Ile Lys Arg Gln Leu
1190 1195 1200

Val Glu Thr Arg Gln Ile Thr Lys His Val Ala Gln Ile Leu Asp
1205 1210 1215

Ser Arg Met Asn Thr Lys Tyr Asp Glu Asn Asp Lys Leu Ile Arg
1220 1225 1230

Glu Val Lys Val Ile Thr Leu Lys Ser Lys Leu Val Ser Asp Phe
1235 1240 1245

Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg Glu Ile Asn Asn Tyr
1250 1255 1260

His His Ala His Asp Ala Tyr Leu Asn Ala Val Val Gly Thr Ala
1265 1270 1275

Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe Val Tyr Gly
1280 1285 1290

Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala Lys Ser Glu
1295 1300 1305

Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr Ser Asn
1310 1315 1320

Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn Gly Glu
1325 1330 1335

Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr Gly Glu
1340 1345 1350

[0034] Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg Lys Val
1355 1360 1365

Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu Val Gln
1370 1375 1380

Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg Asn Ser
1385 1390 1395

Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys Lys Tyr
1400 1405 1410

Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val Leu Val Val
1415 1420 1425

Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys Ser Val Lys
1430 1435 1440

Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser Phe Glu Lys
1445 1450 1455

Asn	Pro	Ile	Asp	Phe	Leu	Glu	Ala	Lys	Gly	Tyr	Lys	Glu	Val	Lys	
1460							1465					1470			
Lys	Asp	Leu	Ile	Ile	Lys	Leu	Pro	Lys	Tyr	Ser	Leu	Phe	Glu	Leu	
1475							1480					1485			
Glu	Asn	Gly	Arg	Lys	Arg	Met	Leu	Ala	Ser	Ala	Gly	Glu	Leu	Gln	
1490							1495					1500			
Lys	Gly	Asn	Glu	Leu	Ala	Leu	Pro	Ser	Lys	Tyr	Val	Asn	Phe	Leu	
1505							1510					1515			
Tyr	Leu	Ala	Ser	His	Tyr	Glu	Lys	Leu	Lys	Gly	Ser	Pro	Glu	Asp	
1520							1525					1530			
Asn	Glu	Gln	Lys	Gln	Leu	Phe	Val	Glu	Gln	His	Lys	His	Tyr	Leu	
1535							1540					1545			
[0035]	Asp	Glu	Ile	Ile	Glu	Gln	Ile	Ser	Glu	Phe	Ser	Lys	Arg	Val	Ile
	1550						1555					1560			
Leu	Ala	Asp	Ala	Asn	Leu	Asp	Lys	Val	Leu	Ser	Ala	Tyr	Asn	Lys	
1565							1570					1575			
His	Arg	Asp	Lys	Pro	Ile	Arg	Glu	Gln	Ala	Glu	Asn	Ile	Ile	His	
1580							1585					1590			
Leu	Phe	Thr	Leu	Thr	Asn	Leu	Gly	Ala	Pro	Ala	Ala	Phe	Lys	Tyr	
1595							1600					1605			
Phe	Asp	Thr	Thr	Ile	Asp	Arg	Lys	Arg	Tyr	Thr	Ser	Thr	Lys	Glu	
1610							1615					1620			
Val	Leu	Asp	Ala	Thr	Leu	Ile	His	Gln	Ser	Ile	Thr	Gly	Leu	Tyr	
1625							1630					1635			
Glu	Thr	Arg	Ile	Asp	Leu	Ser	Gln	Leu	Gly	Gly	Asp	Glu	Phe	Pro	

1640	1645	1650
------	------	------

Lys Lys Lys Arg Lys Val Gly	Gly Gly Gly Ser Pro	Lys Lys Lys
1655	1660	1665

Arg Lys Val	
1670	

<210> 50

<211> 1678

<212> PRT

<213> 人工序列

<220>

<221> 来源

<223> /注="人工序列描述: 合成多肽"

<400> 50

Met Val Ser Lys Gly Glu Glu Leu Phe Thr Gly Val Val Pro Ile Leu			
1	5	10	15

[0036] Val Glu Leu Asp Gly Asp Val Asn Gly His Lys Phe Ser Val Ser Gly
20 25 30

Glu Gly Glu Gly Asp Ala Thr Tyr Gly Lys Leu Thr Leu Lys Phe Ile		
35	40	45

Cys Thr Thr Gly Lys Leu Pro Val Pro Trp Pro Thr Leu Val Thr Thr		
50	55	60

Leu Thr Tyr Gly Val Gln Cys Phe Ser Arg Tyr Pro Asp His Met Lys			
65	70	75	80

Gln His Asp Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr Val Gln Glu		
85	90	95

Arg Thr Ile Phe Phe Lys Asp Asp Gly Asn Tyr Lys Thr Arg Ala Glu		
100	105	110

Val Lys Phe Glu Gly Asp Thr Leu Val Asn Arg Ile Glu Leu Lys Gly		
115	120	125

Ile Asp Phe Lys Glu Asp Gly Asn Ile Leu Gly His Lys Leu Glu Tyr
 130 135 140

Asn Tyr Asn Ser His Asn Val Tyr Ile Met Ala Asp Lys Gln Lys Asn
 145 150 155 160

Gly Ile Lys Val Asn Phe Lys Ile Arg His Asn Ile Glu Asp Gly Ser
 165 170 175

Val Gln Leu Ala Asp His Tyr Gln Gln Asn Thr Pro Ile Gly Asp Gly
 180 185 190

Pro Val Leu Leu Pro Asp Asn His Tyr Leu Ser Thr Gln Ser Lys Leu
 195 200 205

Ser Lys Asp Pro Asn Glu Lys Arg Asp His Met Val Leu Leu Glu Phe
 210 215 220

[0037] Val Thr Ala Ala Gly Ile Thr Leu Gly Met Asp Glu Leu Tyr Lys Val
 225 230 235 240

Asp Ser Gly Gly Ser Ser Gly Ser Ser Gly Ser Glu Thr Pro Gly
 245 250 255

Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser Gly Gly Ser Ser Gly Gly
 260 265 270

Ser Pro Ala Ala Lys Arg Val Lys Leu Asp Gly Gly Gly Ser Thr
 275 280 285

Gly Met Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser
 290 295 300

Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys
 305 310 315 320

Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu

325	330	335
-----	-----	-----

Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg		
340	345	350

Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile		
355	360	365

Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp		
370	375	380

Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys		
385	390	395
		400

Lys His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala		
405	410	415

Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val		
[0038] 420	425	430

Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala		
435	440	445

His Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn		
450	455	460

Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr		
465	470	475
		480

Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp		
485	490	495

Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu		
500	505	510

Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly		
515	520	525

Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn
 530 535 540

Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr
 545 550 555 560

Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala
 565 570 575

Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser
 580 585 590

Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala
 595 600 605

Ser Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu
 610 615 620

[0039] Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe
 625 630 635 640

Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala
 645 650 655

Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met
 660 665 670

Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu
 675 680 685

Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His
 690 695 700

Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro
 705 710 715 720

Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg
 725 730 735

Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala
 740 745 750

Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu
 755 760 765

Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met
 770 775 780

Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His
 785 790 795 800

Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val
 805 810 815

Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu
 820 825 830

[0040] Gln Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val
 835 840 845

Thr Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe
 850 855 860

Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu
 865 870 875 880

Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu
 885 890 895

Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu
 900 905 910

Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr
 915 920 925

Ala His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg
 930 935 940

Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg
945 950 955 960

Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly
965 970 975

Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr
980 985 990

Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser
995 1000 1005

Leu His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys
1010 1015 1020

Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys
1025 1030 1035

[0041] Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala
1040 1045 1050

Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu
1055 1060 1065

Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln
1070 1075 1080

Ile Leu Lys Glu His Pro Val Glu Asn Thr Gln Leu Gln Asn Glu
1085 1090 1095

Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val
1100 1105 1110

Asp Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp Val Asp
1115 1120 1125

His Ile Val Pro Gln Ser Phe Leu Ala Asp Asp Ser Ile Asp Asn

1130

1135

1140

Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp Asn
 1145 1150 1155

Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg
 1160 1165 1170

Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn
 1175 1180 1185

Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys Ala
 1190 1195 1200

Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys
 1205 1210 1215

[0042] His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
 1220 1225 1230

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys
 1235 1240 1245

Ser Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys
 1250 1255 1260

Val Arg Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu
 1265 1270 1275

Asn Ala Val Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Ala Leu
 1280 1285 1290

Glu Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg
 1295 1300 1305

Lys Met Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala
 1310 1315 1320

Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu
 1325 1330 1335

Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys Ala Pro Leu Ile Glu
 1340 1345 1350

Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp
 1355 1360 1365

Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile
 1370 1375 1380

Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser
 1385 1390 1395

Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys
 1400 1405 1410

[0043] Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val
 1415 1420 1425

Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser
 1430 1435 1440

Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met
 1445 1450 1455

Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala
 1460 1465 1470

Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro
 1475 1480 1485

Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu
 1490 1495 1500

Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro
 1505 1510 1515

Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys
 1520 1525 1530

Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val
 1535 1540 1545

Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser
 1550 1555 1560

Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys
 1565 1570 1575

Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu
 1580 1585 1590

Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly
 1595 1600 1605

[0044] Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys
 1610 1615 1620

Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His
 1625 1630 1635

Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln
 1640 1645 1650

Leu Gly Gly Asp Glu Phe Pro Lys Lys Lys Arg Lys Val Gly Gly
 1655 1660 1665

Gly Gly Ser Pro Lys Lys Lys Arg Lys Val
 1670 1675

<210> 51
 <211> 20
 <212> RNA
 <213> 人工序列

<220>

<221> 来源

<223> /注="人工序列描述: 合成寡核苷酸"

<400> 51

gaguccgagc agaagaagaa

20

<210> 52

<211> 20

<212> RNA

<213> 人工序列

<220>

<221> 来源

<223> /注="人工序列描述: 合成寡核苷酸"

<400> 52

ggcacugcgg cuggaggugg

20

<210> 53

<211> 20

<212> RNA

<213> 人工序列

[0045]

<220>

<221> 来源

<223> /注="人工序列描述: 合成寡核苷酸"

<400> 53

ggucacuuuuu aacacaccca

20

<210> 54

<211> 55

<212> DNA

<213> 人工序列

<220>

<221> 来源

<223> /注="人工序列描述: 合成引物"

<220>

<221> 修饰的碱基

<222> (34)..(39)

<223> a, t, g 或 c

<400> 54

tgcgtggcag cgtcagatgt gtataagaga cagnnnnnnc cccagtggtc gctct

55

<210> 55		
<211> 57		
<212> DNA		
<213> 人工序列		
<220>		
<221> 来源		
<223> /注="人工序列描述: 合成引物"		
<220>		
<221> 修饰的碱基		
<222> (35)..(40)		
<223> a, t, g 或 c		
<400> 55		
gtctcggtgg ctggagatg tgtataagag acagnnnnn ccaggcctcc ccaaagc		57
<210> 56		
<211> 59		
<212> DNA		
<213> 人工序列		
[0046] <220>		
<221> 来源		
<223> /注="人工序列描述: 合成引物"		
<220>		
<221> 修饰的碱基		
<222> (34)..(39)		
<223> a, t, g 或 c		
<400> 56		
tgcgtggcag cgtcagatgt gtataagaga cagnnnnnng gaaccaggta agccagaga		59
<210> 57		
<211> 57		
<212> DNA		
<213> 人工序列		
<220>		
<221> 来源		
<223> /注="人工序列描述: 合成引物"		
<220>		
<221> 修饰的碱基		

<222> (35)..(40)		
<223> a, t, g 或 c		
<400> 57		
gtctcggtgg ctcggagatg tgtataagag acagnnnnnn ggggtggggt cagacgt		57
<210> 58		
<211> 59		
<212> DNA		
<213> 人工序列		
<220>		
<221> 来源		
<223> /注="人工序列描述: 合成引物"		
<220>		
<221> 修饰的碱基		
<222> (34)..(39)		
<223> a, t, g 或 c		
<400> 58		
tcgtcgccag cgtcagatgt gtataagaga cagnnnnnng cccattccct cttagcca		59
[0047]		
<210> 59		
<211> 61		
<212> DNA		
<213> 人工序列		
<220>		
<221> 来源		
<223> /注="人工序列描述: 合成引物"		
<220>		
<221> 修饰的碱基		
<222> (35)..(40)		
<223> a, t, g 或 c		
<400> 59		
gtctcggtgg ctcggagatg tgtataagag acagnnnnnn ggacaggaa agtgaggta		60
c		61
<210> 60		
<211> 61		
<212> DNA		
<213> 人工序列		

<220>

<221> 来源

<223> /注="人工序列描述: 合成引物"

<220>

<221> 修饰的碱基

<222> (34)..(39)

<223> a, t, g 或 c

<400> 60

tcgtcggcag cgtcagatgt gtataagaga cagnnnnnna atggacacat ggtagtcag 60

g

61

<210> 61

<211> 64

<212> DNA

<213> 人工序列

<220>

<221> 来源

<223> /注="人工序列描述: 合成引物"

[0048]

<220>

<221> 修饰的碱基

<222> (35)..(40)

<223> a, t, g 或 c

<400> 61

gtctcggtgg ctcggagatg tgtataagag acagnnnnnn ggcttatatc caacacttcg 60

tggg

64

<210> 62

<211> 5076

<212> DNA

<213> 人工序列

<220>

<221> 来源

<223> /注="人工序列描述: 合成多核苷酸"

<400> 62

atggtagca aaggtaaga actgttaca ggtgttgtc cgattctgg tgaactggat 60

ggtagtttta atggccacaa atttcagtt agcggtgaag gcgaaggta tgcaacctat 120

ggtaaaactga ccctgaaatt tatctgtacc accggcaaac tgccggttcc gtggccgaca	180
ctggttacca cactgaccta tggtgttcag tgtttagcc gttatccgga tcacatgaaa	240
cagcacgatt ttttcaaaaag cgcaatgccg gaaggttatg ttcaagaacg taccatctc	300
ttcaaagatg acggcaacta taaaaccgt gccgaagtta aatttgaagg tgataccctg	360
gtgaatcgca ttgaactgaa aggcacatcgat tttaaagagg atggtaatat cctggccac	420
aaactggaat ataattataa tagccacaac gtgtacatca tggccgacaa acagaaaaat	480
ggcatcaaag tgaacttcaa gatccgccat aatattgaag atggttcagt tcagctggcc	540
gatcattatc agcagaatac cccgattggt gatggtccgg ttctgctgcc ggataatcat	600
tatctgagca cccagagcaa actgagcaaa gatccgaatg aaaaacgtga tcacatggc	660
ctgctggaat ttgttaccgc agcaggtatt accttaggtt tggatgaact gtataaagtc	720
gacgcagaag cagcagecaaa agaagccgct gccaaagaag cggcagcga agaggcagcc	780
gcaaaagcac tggaagccga ggctgcggct aaagaggctg ctgcaaaaga agcagccgct	840
[0049] aaagaagctg cggctaaggc accggcagca aaacgtgtt aactggacgg tggtggtggt	900
agcaccggta tggacaagaa atacagcatc ggttggata ttggcacgaa tagcgtgggt	960
tggccgtta ttaccgacga gtacaaagtg ccgtccaaga aattcaaagt gctggcaat	1020
accgatcgcc atagcatcaa gaaaaatctg attggcgcac tgctgtcga cagcggtgag	1080
actgccgaag ctacgcgtct gaagcgtacg ggcgcgtc gctacacccg ccgtaaagAAC	1140
cgtatttgct atctgcaaga aatcttcagc aacgaaatgg ccaaagttga tgatagctt	1200
tttcaccgcc tggaagagag ctttctggtg gaagaggata agaaacacga ggcgcacccg	1260
attttggta acattgtcga tgaagtggca taccatgaga agtacccgac catctaccac	1320
cttcgtaaag aactggtgga cagcaccgat aaagctgatc tgcgtctgat ttacctggcg	1380
ctggcccaca tgattaagtt tcgcggcat ttctgtatcg agggcgatct gaatccggac	1440
aattctgatg ttgacaagct gtttattcaa ctgtacaga cctacaacca gttgtcgaa	1500
gagaacccga tcaatgcgag cgggtttgat gccaaagcaa ttctgagcgc acgcctgagc	1560
aaatctcgcc gtttggagaa cctgattgca cagctgccgg gtgagaagaa aaacggtctg	1620

	ttcggcaatc tgattgcact gtccctggc ttgaccccga atttaagag caacttcgac	1680
	ctggccgaag atgcgaagct ccaattgagc aaagacacacct acgacgatga cctggacaat	1740
	ctgctggccc agattggcga ccagtacgca gatctgttct tggctgcgaa aaacctgagc	1800
	gatgcaattc tgctgtcgga catcctgcgc gtgaatacgg aaatcacgaa agcgcctctg	1860
	agcgcgtcta tcatcaaggc ctatgacgag caccaccaag atctgaccct gctgaaagct	1920
	ctggtgagac aacaattgcc agagaagtat aaagaaattt tctttgacca gagcaaaaac	1980
	ggctatgcgg gttacattga cggtggcgcc agccaagaag agttctacaa attcattaag	2040
	cctatectgg agaaaatgga tggcacccgaa gaactgctgg taaagctgaa tcgtgaagat	2100
	ctgctgcgca aacagcgcac ttttataac ggttagcattc cgcaccagat ccatctgggt	2160
	gagttgcacg cgattttgcg tcgccaggaa gatTTTATC cgTTCTGAA agacaaccgt	2220
	gagaaaatcg agaaaattct gacgttccgt atcccgtatt atgtcgccccc gctggcgct	2280
	ggtaatagcc gtttcgcgtg gatgaccgcgaa aaatcagagg aaacgattac cccgtggaaat	2340
[0050]	tttggagaag ttgttgataa gggtgcaagc ggcgcgttgcg tcattgagcg tatgaccaac	2400
	tttgacaaga atttgcgaa tgaaaaagtc ttgccgaagc actctctgct gtacgagtat	2460
	tttaccgttt acaacgaatt gaccaaggaa aaatacgtca ccgaaggcat ggcacaaaccg	2520
	gccttcgttgcgac gaaaaaagca atcggttgc accttttttaa gaccaaccgc	2580
	aagggttacgg tcaaacaact gaaagaggac tatttcaaga aaattgaatg ttttgactcc	2640
	gttagagatct ccgggtttga ggaccgttgc aacgcgcgttcc tgggcaccta ccatgatctg	2700
	ctgaaaattt taaaagacaa agatttctg gacaacgaag agaacgaaga tattctggaa	2760
	gatatacgatc tgaccctgac gctgttgcgaa gatcgtgaga tgattgagga acgtctgaaa	2820
	acctaagcac acttgcgttgc tgacaaagtt atgaaacagc tgaagcgtcg tcgttacaca	2880
	ggttggggcc gtctgagccg taagtttac aatggtatcc gtgacaaaca gagcggttgc	2940
	acgatttcgg actttctgaa gtcagatggc ttgcggccatc gcaactttat gcaactgatt	3000
	catgacgact ctctgacgtt caaggaagat atccaaaagg cacaggttagtgc cggtcagggt	3060
	gatagcctgc atgagcatat cgcgaacctg gcgggttagcc cggctatcaa aaagggttgc	3120
	ttacagactg tgaaagttgtt ggtatgttgc ttgggtttca caaaccggaa	3180

	aatattgtga tcgagatggc acgtaaaaat cagacgacgc aaaagggtca aaaaaattct	3240
	cgtgagcgca taaaacgtat tgaagagggt atcaaagaat tggcagcca aattctgaaa	3300
	gaacaccggc tcgagaacac ccagctgcaa aacgaaaaac tgtatttata ctatctcag	3360
	aacggtcgtg acatgtacgt ggatcaagaa ctggacatca atcgtttag cgattacgt	3420
	gttgatcata ttgtgcctca gagcttctg aaagacgatt cgatcgacaa caaagtgcgt	3480
	accctagcg acaagaatcg tggtaagagc gataacgtgc cgagcgaaga agtcgttaag	3540
	aaaatgaaaa actactggcg tcagctgctg aacgccaagc tgattacccg gcgttaagttc	3600
	gataacctga cgaaagccga cggtggaggc ctgagcgagc tggacaaggc cggctttatc	3660
	aagcgtcaac tggtgaaac ccgtcagatc actaaacatg tggcacagat cctggactcc	3720
	cgcataata cgaaatatga cgagaatgac aagttgatcc gtgaagtcaa agttattacg	3780
	ctgaaaagca aactgggtgc cgattccgt aaagacttcc agttctataa agtccgtgaa	3840
	atcaacaact atcatcacgc ccacgatgac tacttgaacg ctgttgtgg caccgcactg	3900
[0051]	atcaagaaat accctaagct cgaaagcgag tttgtctatg gtgactataa agtttacgac	3960
	gtgcgttaaga tgatgccaa gagcggcaaa gaaattggta aggctaccgc aaagtacttt	4020
	ttctacagca acatcatgaa cttttcaaa accgagatta ccctggcgaa cggtgagatc	4080
	cgttaaacggc cgctgattga gactaatggc gaaacggcg agattgtgt ggacaagggt	4140
	cgcgatttcg ctacggttcg taaggtcctg agcatgccgc aagttaacat tgtcaagaaa	4200
	actgaagtgc agacgggtgg cttagcaaa gaatccatcc tgccgaagcg taatagcgat	4260
	aaacttatcg cgctaaaaaa agactggac ccaaagaaat atggcggctt tgatagcccg	4320
	accgtcgct atagcgtgtt agtggtcgac aaagttgaaa agggcaagag caagaaactg	4380
	aagtccgtca aagaacttct ggttatcacc atcatgaaac gtatccctt tgagaagaac	4440
	ccgattgact tcttagaggc gaagggttat aaagaagtca aaaaagacct gattatcaag	4500
	ctgccgaagt acagcctgtt tgagttggag aatggtcgt aagcgtatgtt ggcgagcgc	4560
	ggtgagctgc aaaagggcaa cgaactggcg ctgccgtca aataacgtcaa ttttctgtac	4620
	ctggccagcc actacgaaaaa gctgaagggt tctccggaaat aacgaaaca aaagcaactg	4680

ttcggtgagc aacataaaaca ctactggac gaaatcatcg agcaaattag cgaatttagc	4740
aaacgtgtca tcctggcgga cgcgaaatctg gacaagggtcc tgtctgcata caataagcat	4800
cgcgacaaac caattcgtga gcaagcggag aatatcatcc acctgttac gctgaccaac	4860
ctaggtgcgc cggcggcatt caagtatttc gatacgacca tcgaccgcaa gcgcataacc	4920
agcaccaaaag aggtcctgga cgcgaccctg atccaccaga gcattaccgg cttatacgaa	4980
acccgtattg atttgagcca actgggtggc gatgaattcc cgaaaaaaaaa ggcgaaagtt	5040
ggtggcggtg gtagccccaa aaagaaaacgt aaagt	5076

<210> 63

<211> 5013

<212> DNA

<213> 人工序列

<220>

<221> 来源

<223> /注=“人工序列描述：合成多核苷酸”

[0052] <400> 63 atggtagca aaggtaaagc cgtgattaaa gaatttatgc gcttaaggt tcacatggaa	60
ggtagcatga atggccatga atttggaaatt gaaggtaag gcgaagggtcg tccgtatgaa	120
ggcacccaga ccgcaaaaact gaaagttacc aaagggtggc cgctgccgt tagctggat	180
attctgagtc cgcagtttat gtatggtagc cgtgcattta ccaaacatcc ggcagatatt	240
ccggattatt acaaacagag ctttccggaa gtttttaat gggAACGTGT gatgaatttt	300
gaagatggtg gtgcagttac cgttacacag gataccagcc tggaaagatgg caccctgatc	360
tataaagttt aactgcgtgg caccaatttt ccgcctgtat gcgggttat gcagaaaaaaaa	420
acaatgggtt gggaaagcaag caccgaacgt ctgtatcctg aagatggcgt tctgaaaggt	480
gatatcaaaa tggcactgctcg tctgaaagat ggccgtcgat atctggcaga tttcaaaacc	540
acctataaaag ccaaaaaacc tggtcagatg cctgggtgcct ataatgttga tcgtaaactg	600
gatattacca gccacaacga agattatacc gttgtggaaac agtatgaacg tagcgaaggc	660
cgtcatagca caggtggat ggttgcgttataaagtcg acagcggtgg tagcagcggt	720
ggttcaagcg gtagcgaaac accgggtaca agcggaaagcg caacaccgga aagcagtgg	780

ggtagttcag gtggtagtcc ggcagcaaaa cgtgtgaaac tggatggcg ggccggtagc	840
accggtatgg acaagaaata cagcatcggt ttggatattg gcacgaatag cgtgggttgg	900
gccgttatta ccgacgagta caaatgtccg tccaagaat tcaaagtgc gggcaatacc	960
gatgccata gcatcaagaa aaatctgatt ggcgcactgc tggtcgacag cggtgagact	1020
gccgaagcta cgcgtctgaa gcgtacggcg cgtcgctgc acacccgccc taagaaccgt	1080
atttgctatc tgcaagaat cttcagcaac gaaatggcca aagttgatga tagcttttt	1140
caccgcctgg aagagagctt tctggtgaa gaggataaga aacacgagcg ccatccgatt	1200
tttggtaaca ttgtcgatga agtggcatac catgagaagt acccgaccat ctaccacett	1260
cgtaaagaaac tggggacag caccgataaa gctgatctgc gtctgattt cctggcgctg	1320
gccccatga ttaagttcg cggtcattt ctgatcgagg gcgatctgaa tccggacaat	1380
tctgatgtt acaagctgtt tattcaactt gtacagacct acaaccagggtt gttcgaagag	1440
aacccgatca atgcgagcgg tggatgcc aaagcaattc tgagcgcacg cctgagcaaa	1500
tctcgccgtt tggagaacct gattgcacag ctggccgggtt agaagaaaaa cggtctgttc	1560
[0053] ggcaatctga ttgcactgtc cctggcttg accccgaatt ttaagagcaa ctgcacactg	1620
gccgaagatg cgaagctcca attgagcaaa gacacctacg acgtgacacct ggacaatctg	1680
ctggcccaga ttggcgacca gtacgcagat ctgttcttgg ctgcgaaaaa cctgaggcgat	1740
gcaattctgc tgtcgacat cctgcgcgtg aatacgaaaa tcacgaaagc gcctctgagc	1800
ggtctatga tcaagcgcta tgacgacac caccaagatc tgaccctgt gaaagctctg	1860
gtgagacaac aattgccaga gaagtataaa gaaattttct ttgaccagag caaaaacggc	1920
tatgcgggtt acattgacgg tggcccgac caagaagagt tctacaaatt cattaagcct	1980
atcctggaga aaatggatgg caccgaagaa ctgctggtaa agctgaatcg tgaagatctg	2040
ctgcgcaaac agcgcacttt tgataacggt agcattccgc accagatcca tctgggttag	2100
ttgcacgcga ttttgcgtcg ccaggaagat ttttatccgt tcttggaaaga caaccgttag	2160
aaaatcgaga aaattctgac gtccgtatc ccgttattatg tcggcccgct ggccgtgg	2220
aatagccgct tcgcgtggat gacccgcaaa tcagaggaaa cgattacccc gtggatttt	2280
gaggaagttt ttgataaggg tgcaagcgcg cagtcgttca ttgagcgtat gaccaacttt	2340

	gacaagaatt tgccgaatga aaaagtcttg ccgaagcact ctctgctgta cgagtatttt	2400
	accgtttaca acgaattgac caaggtaaa tacgtcacgg aaggcatgcaaaaccggcc	2460
	ttcctgagcg gcgagcagaa aaaagcaatc gttgacctct tggtaagac caaccgcaag	2520
	gttacggtca aacaactgaa agaggactat ttcaagaaaa ttgaatgttt tgactccgta	2580
	gagatctccg gtgttgagga ccgttcaac gcgagcctgg gcacccatcca tgatctgctg	2640
	aaaattatta aagacaaga tttctggac aacgaagaga acgaagatata tctggaagat	2700
	atcggtctga ccctgacgct gttcgaagat cgtgagatga ttgaggaacg tctgaaaacc	2760
	tacgcacact tggatcgatga caaagttatg aaacagctga agcgtcgctg ttacacaggt	2820
	tggggccgtc tgagccgtaa gcttatcaat ggtatccgtg acaaacagag cggtaaagacg	2880
	attctggact ttctgaagtc agatggcttc gccaatcgca actttatgca actgattcat	2940
	gacgactctc tgacgttcaa ggaagatate caaaaggcac aggtgagcgg tcaggggtgat	3000
	agcctgcatg agcatatcgc gaacctggcg gtagccccgg ctatcaaaaa gggtatctta	3060
[0054]	cagactgtga aagttgtgga tgaattgggtt aagggttatgg gtcgtcacaa accggaaaat	3120
	atttgtatcg agatggcacg tggaaatcag acgacgcaaa agggtcaaaa aaattctcgat	3180
	gagcgcattga aacgtattga agagggtatc aaagaattgg gcagccaaat tctgaaagaa	3240
	cacccggctcg agaacaccca gctgcaaaac gaaaaactgt atttatacta tctgcagaac	3300
	ggtcgtgaca tgtacgtgga tcaagaactg gacatcaatc gttttagcga ttacgtatgtt	3360
	gatcatatttgcctcagag ctttctgaaa gacgattcga tcgacaacaa agtgcgtacc	3420
	cgtacgtgaca agaatcggtgg taagagcgt aacgtgccga gcgaagaagt cgttaagaaaa	3480
	atgaaaaact actggcgatca gctgctgaaac gccaaagctga ttacccagcg taagttcgat	3540
	aacctgacga aagccgagcg tggaggcctg agcgagctgg acaaggccgg ctttatcaag	3600
	cgtcaactgg tggaaacccg tcagatcaatc aaacatgtgg cacagatctt ggactcccgc	3660
	atgaatacga aatatgacga gaatgacaag ttgatccgtg aagtcaaagt tattacgctg	3720
	aaaagcaaac tgggttccga ttccgtaaa gacttccagt tctataaagt ccgtaaatc	3780
	aacaactatc atcacgccccca cgatgcgtac ttgaacgctg ttgtggcac cgcaactgatc	3840

aagaaaatacc ctaagctcg aagcgagttt gtctatggtg actataaaagt ttacgacgtg	3900
cgtaagatga tcgccaagag cgagcaagaa attggtaagg ctaccgcaaa gtacttttc	3960
tacagcaaca tcatgaactt cttcaaaacc gagattaccc tggcgaacgg tgagatccgt	4020
aaacggccgc tgattgagac taatggcgaa acgggcgaga ttgtgtggga caagggtcgc	4080
gatttcgcta cggttcgtaa ggtcctgagc atgccgcaag ttaacattgt caagaaaact	4140
gaagtgcaga cgggtggctt tagcaaagaa tccatcctgc cgaagcgtaa tagcgataaa	4200
cttatcgccc gtaaaaaaaga ctgggaccca aagaatatg gcggcttga tagccgacc	4260
gtcgcgtata gcgtgttagt ggtcgcgaaa gttgaaaagg gcaagagcaa gaaactgaag	4320
tccgtcaaag aacttctggg tatcaccatc atggaacgta gtccttga gaagaacccg	4380
attgacttct tagaggcgaa gggttataaa gaagtcaaaa aagacctgat tatcaagctg	4440
ccgaagtaca gcctgtttga gttggagaat ggtcgttaagc gcatgctggc gagcgcgggt	4500
gagctgcaaa agggcaacga actggcgctg ccgtcgaaat acgtcaattt tctgtacctg	4560
gccagccact acgaaaagct gaagggttct ccggaagata acgaacaaaa gcaactgttc	4620
gtttagcaac ataaacacta cttggacgaa atcatcgagc aaattagcga atttagcaaa	4680
cgtgtcatcc tggcggacgc gaatctggac aaggtcctgt ctgcatacaa taagcatcgc	4740
gacaaaccaa ttcgtgagca agcggagaat atcatccacc ttttacgct gaccaaccta	4800
ggtgcgcgg cgccattcaa gtatttcgat acgaccatcg accgcaagcg ctataccagc	4860
accaaagagg tcctggacgc gaccctgatc caccagagca ttaccggctt atacgaaacc	4920
cgtattgatt tgagccaact gggtggcgat gaattccgaa aaaaaaagcg caaagtttgt	4980
ggcggtggtta gccgaaaaaa gaaacgtaaa gtg	5013

<210> 64

<211> 5034

<212> DNA

<213> 人工序列

<220>

<221> 来源

<223> /注="人工序列描述：合成多核苷酸"

<400> 64

atggtagca aaggtaaga actgttaca ggtgttgc cgattctgg tgaactggat	60
ggtgatgtta atggccacaa atttcagtt agcggtgaag gcgaagggtga tgcaacctat	120
ggtaaaactga ccctgaaatt tatctgtacc accggcaaac tgccggttcc gtggccgaca	180
ctggttacca cactgaccta tgggttcag tggtagcc gttatccgga tcacatgaaa	240
cagcacgatt ttttcaaaag cgcaatgccg gaaggtttagt ttcaagaacg taccatttc	300
ttcaaagatg acggcaacta taaaaccgt gccgaagtta aatttgaagg tgataccctg	360
gtgaatcgca ttgaactgaa aggcatcgat tttaaagagg atgtaatat cctggccac	420
aaactggaat ataattataa tagccacaac gtgtacatca tggccgacaa acagaaaaat	480
ggcatcaaag tgaacttcaa gatccgccat aatattgaag atggtagt tcagctggcc	540
gatcattatc agcagaatac cccgatttgt gatggccgg ttctgctgcc ggataatcat	600
tatctgagca cccagagcaa actgagcaaa gatccgaatg aaaaacgtga tcacatggtg	660
ctgctggaat ttgttaccgc agcaggttatt accttagta tggatgaact gtataaagtc	720
gacagcggtg gtagcagccg tggtaaagc ggtagcgaaa caccgggtac aagcgaaagc	780
[0056] gcaacacccgg aaagcagtgg tggtagctca ggtggtagtc cggcagcaaa acgtgttaaa	840
ctggacggtg gtgggttagt caccgtatg gacaagaaat acagcatcgg tttggatatt	900
ggcacgaata gcgtgggttgc ggccgttatt accgacgagt acaaagtgcc gtccaagaaa	960
ttcaaagtgc tggcaatac cgatcccat agcatcaaga aaaatctgat tggcgcactg	1020
ctgtcgaca gcggtgagac tgccgaagct acgcgtctga agcgtacggc gctgcgtcgc	1080
tacacccgcc gtaagaaccg tatttgcata ctgcaagaaa tttcagcaa cgaaatggcc	1140
aaagttgatg atagttttt tcaccgcctg gaagagagct ttctgggaa agaggataag	1200
aaacacgagc gccatccgat tttggtaac attgtcgatg aagtggcata ccatgagaag	1260
tacccgacca tctaccacct tcgtaagaaa ctggtgacca gcaccgataa agctgatctg	1320
cgtctgattt acctggcgtt ggcccacatg attaagtttgc gctgtcattt tctgatcgag	1380
ggcgatctga atccggacaa ttctgtatgtt gacaagctgt ttattcaact tgtacagacc	1440
tacaaccagt tggtaaga gaacccgatc aatgcgagcg gtgttgatgc caaagcaatt	1500
ctgagcgcac gcctgagcaa atctggccgt ttggagaacc tgattgcaca gctggccgggt	1560

	gagaagaaaa acggtctgtt cggcaatctg attgcactgt ccctgggctt gaccccgaat	1620
	ttaaagagca acttcgacct ggccgaagat gcgaagctcc aattgagcaa agacacctac	1680
	gacgatgacc tggacaatct gctggccag attggcgacc agtacgcaga tctgttcttg	1740
	gctgcgaaaaa acctgaggcga tgcaattctg ctgtcggaca tcctgcgcgt gaatacgaa	1800
	atcacgaaag cgccctctgag cgcgtctatg atcaagcgct atgacgagca ccaccaagat	1860
	ctgaccctgc tgaaagctct ggtgagacaa caattgccag agaagtataa agaaattttc	1920
	tttggaccaga gcaaaaacgg ctagcgggt tacattgacg gtggcgccag ccaagaagag	1980
	ttctacaat tcattaagcc tatectggag aaaatggatg gcaccgaaga actgctggta	2040
	aagctgaatc gtgaagatct gctgcgcaaa cagcgcactt ttgataacgg tagcattccg	2100
	caccagatcc atctgggtga gttgcacgcg atttgcgtc gccaggaaga ttttatccg	2160
	ttcttgaaag acaaccgtga gaaaatcgag aaaattctga cgttccgtat cccgtattat	2220
	gtcggccgc tggcgctgg taatagccgc ttgcgtgga tgacccgaa atcagaggaa	2280
[0057]	acgattaccc cgtggaattt tgaggaagtt gttgataagg gtgcaagcgc gcagtcgttc	2340
	attgagcgta tgaccaactt tgacaagaat ttgccgaatg aaaaagtctt gccgaagcac	2400
	tctctgctgt acgagtattt taccgttac aacgaattga ccaaggtaa atacgtcacc	2460
	gaaggcatgc gcaaaccggc cttcctgagc ggcgagcaga aaaaagcaat cggtgacctc	2520
	ttgttaaga ccaaccgcaa gttacggtc aaacaactga aagaggacta tttcaagaaa	2580
	attgaatgtt ttgactccgt agagatctcc ggtgttgagg accgttcaa cgcgagcctg	2640
	ggcacctacc atgatctgct gaaaattatt aaagacaaag atttctgga caacgaagag	2700
	aacgaagata ttctggaaga tatcggtctg accctgacgc tggtcgaaga tcgtgagatg	2760
	attgaggaac gtctgaaaac ctacgcacac ttgttcgtatg acaaagttat gaaacagctg	2820
	aagcgtcgtc gttacacagg ttggggccgt ctgagccgta agcttatcaa tggtatccgt	2880
	gacaaacaga gcggtaaagac gattctggac tttctgaagt cagatggctt cgccaatcgc	2940
	aactttatgc aactgattca tgacgactct ctgacgttca aggaagatata cccaaaggca	3000
	caggtgagcg gtcagggtga tagcctgcat gagcatatcg cgaaccctggc gggtagcccg	3060

	gctatcaaaa agggttatctt acagactgtg aaagttgtgg atgaatttgtt taaggttatg	3120
	ggtcgtcaca aaccggaaaa tattgtgate gagatggcac gtgaaaatca gacgacgcaa	3180
	aagggtcaaa aaaattctcg tgagcgcattg aaacgtattt aagagggtat caaagaattt	3240
	ggcagccaaa ttctgaaaga acacccggc gagaacaccc agctgcaaaa cgaaaaactg	3300
	tatttatact atctgcagaa cggcgtgac atgtacgtgg atcaagaact ggacatcaat	3360
	cgtttgagcg attacgtatgt tgatcatatt gtgcctcaga gctttctggc ggacgattcg	3420
	atcgacaaca aagtgcgtac ccgttagc aagaatcgtg gtaagagcga taacgtgccg	3480
	agcgaagaag tcgttaagaa aatgaaaaac tactggcgac agctgctgaa cgccaagctg	3540
	attacccagc gtaagttcga taacctgacg aaagccgagc gtggaggcct gagcggagctg	3600
	gacaaggccg gctttatcaa gcgtcaactg gtggaaaccc gtcagatcac taaacatgtg	3660
	gcacagatcc tggactcccg catgaatacg aaatatgacg agaatgacaa gttgatccgt	3720
	gaagtcaaag ttattacgct gaaaagcaaa ctgggtccg attccgtaa agacttccag	3780
[0058]	ttctataaag tccgtgaaat caacaactat catcacgccc acgatgcgtt cttgaacgct	3840
	gttgtggca ccgcactgat caagaaatac cctgcactcg aaagcgagtt tgtctatgg	3900
	gactataaag tttacgacgt gcgttaagatg atcgccaga gcgagcaaga aattggtaag	3960
	gctaccgcaa agtactttt ctacagcaac atcatgaact tcttcaaaac cgagattacc	4020
	ctggcgaacg gtgagatccg taaagcccg ctgattgaga ctaatggcga aacggccgag	4080
	attgtgtggg acaagggtcg cgatttcgct acggttcgta aggtcctgag catgcccaa	4140
	gttaacattt tcaagaaaaac tgaagtgcag acgggtggct ttagcaaaga atccatcctg	4200
	ccgaagcgta atagcgataa acttatcgcg cgtaaaaaag actgggaccc aaagaaatata	4260
	ggcggcttg atagcccgac cgtcgctat agcgtgttag tggtcgcgaa agttgaaaag	4320
	ggcaagagca agaaactgaa gtccgtcaaa gaacttctgg gtatcaccat catggAACGT	4380
	agctcccttg agaagaaccc gattgacttc tttagaggcga agggttataa agaagtcaaa	4440
	aaagacctga ttatcaagct gccgaagtac agcctgtttt agttggagaaa tggtcgtaa	4500
	cgcacgtgttgg cgagcgcggg tgagctgcaaa aaggcaacg aactggcgct ggcgtcggaa	4560
	tacgtcaatt ttctgttaccc gcccagccac tacgaaaagc tgaagggttc tccggaaagat	4620

aacgaacaaa agcaactgtt cgttgagcaa cataaacact acttggacga aatcatcgag	4680
caaatttagcg aatttagcaa acgtgtcatc ctggcggacg cgaatctgga caaggtcctg	4740
tctgcataca ataagcatcg cgacaaacca attcgtgagc aagcggagaa tatcatccac	4800
ctgtttacgc tgaccaacct aggtgcgccg gcggcattca agtatttcga tacgaccatc	4860
gaccgcaagg gctataccag caccaaagag gtcctggacg cgaccctgat ccaccagagc	4920
attaccggct tatacgaaac ccgtatttgat ttgagccaac tgggtggcga tgaattcccg	4980
aaaaaaaaaggc gcaaagtgg tggcggtggt agcccgaaaa agaaacgtaa agtg	5034

<210> 65
 <211> 20
 <212> RNA
 <213> 人工序列

[0059]

<220>
 <221> 来源
 <223> /注=“人工序列描述：合成寡核苷酸”

<400> 65
 gcucccauca caucaaccgg

20

<210> 66
 <211> 20
 <212> RNA
 <213> 人工序列

<220>
 <221> 来源
 <223> /注=“人工序列描述：合成寡核苷酸”

<400> 66
 ggugagugag ugugugcgug

20

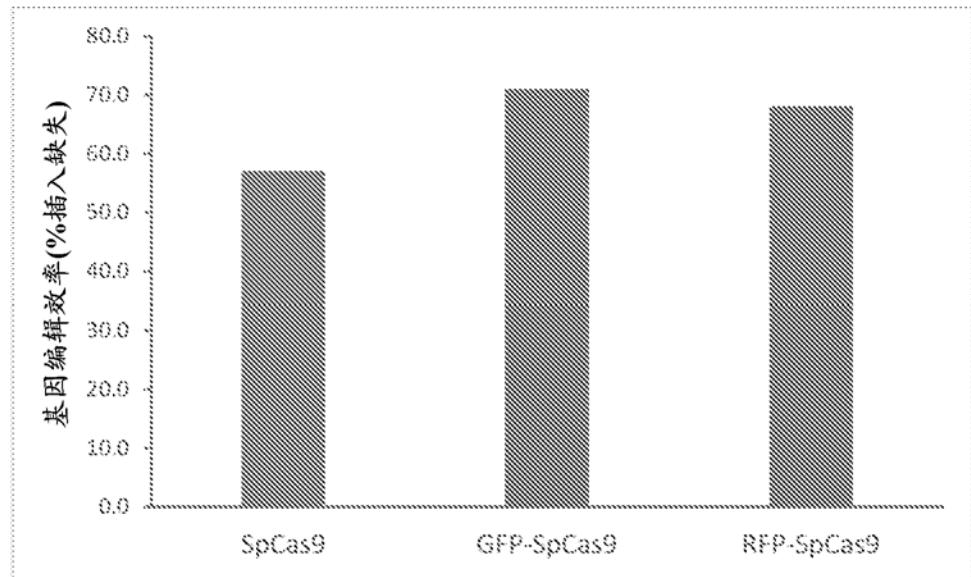


图 1

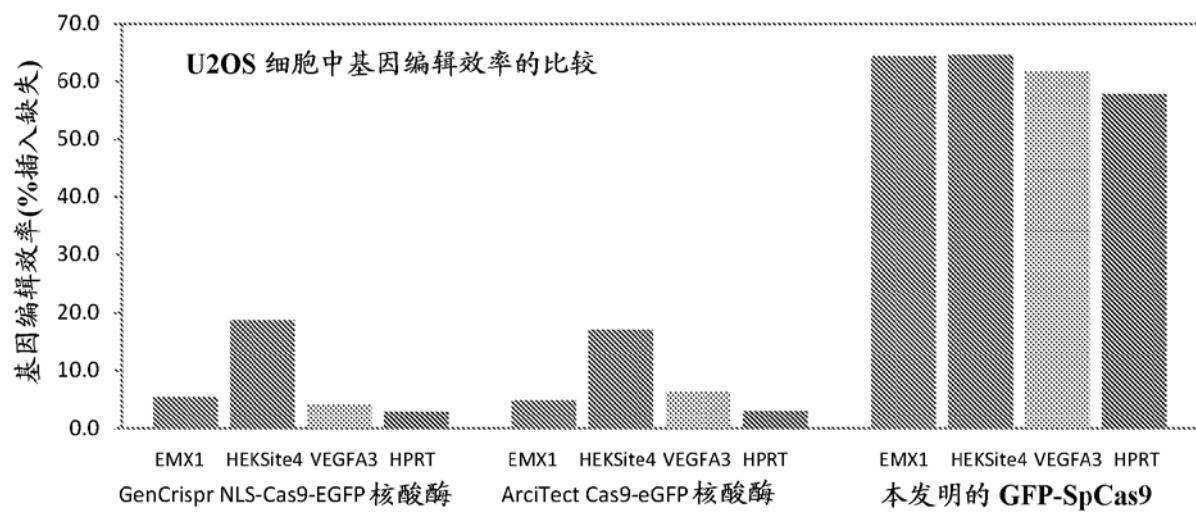


图 2A

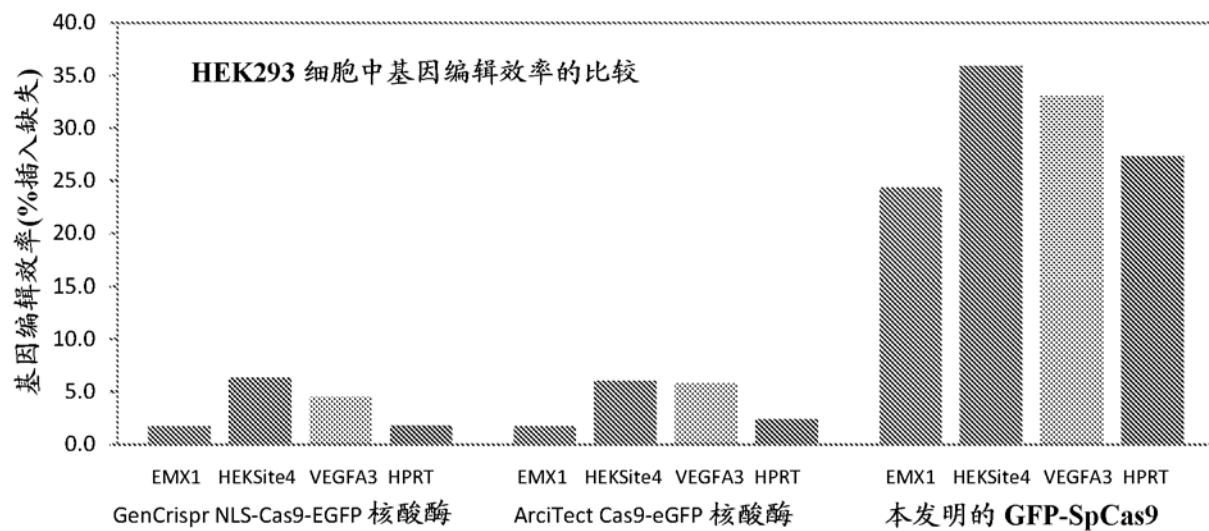


图 2B