(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 113846096 B
(45)授权公告日 2025.05.09

(21)申请号 202110830978.0
(22)申请日 2014.06.04
(65)同一申请的已公布的文献号
申请公布号 CN 113846096 A
(43)申请公布日 2021.12.28
(30)优先权数据
61/830,787 2013.06.04 US
(62)分案原申请数据
201480043971.2 2014.06.04
(73)专利权人 哈佛大学校长及研究员协会
地址 美国马萨诸塞州
(72)发明人 乔治·M·丘奇
普拉尚特·G·马利
凯文·M·埃斯弗特
(74)专利代理机构 上海专利商标事务所有限公司 31100
专利代理师 杨昀 余颖
(51)Int.Cl.
C12N 15/113(2010.01)
C12N 15/55(2006.01)
C12N 15/10(2006.01)
C12N 15/90(2006.01)
C12N 15/63(2006.01)
(56)对比文件
Mali P.等.RNA-guided Human Genome Engineering via Cas9.Science.2013,第339卷第823-826页.
RAN,F.ANN.Double nicking by RNA-Guided CRISPR Cas9 for enhanced genome editing specifictiy.CELL.2013,第154卷(第6期),第1380-1389页.
审查员 毛颖

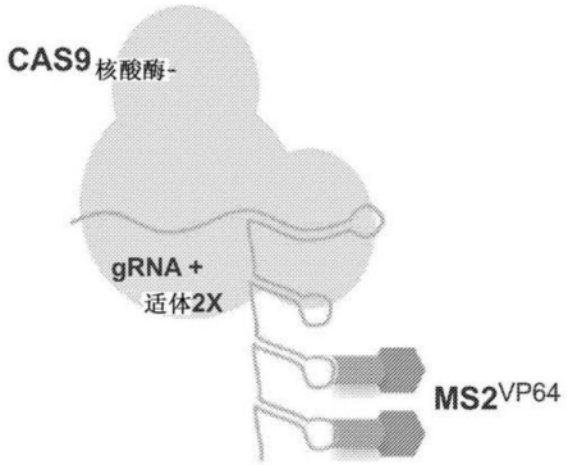
权利要求书3页 说明书38页
序列表48页 附图72页

(54)发明名称

RNA-导向的转录调控

(57)摘要

提供了RNA-导向的转录调控。具体包括一种改变细胞中DNA靶标核酸的方法,包括向所述细胞引入编码两种或更多种RNA的第一外源核酸,其中每种RNA与所述DNA靶标核酸中相邻位点互补,向所述细胞引入编码至少一种Cas9蛋白切口酶的第二外源核酸,所述Cas9蛋白切口酶具有一个非活性核酸酶结构域并由所述两种或更多种RNA导向,且其中表达所述两种或更多种RNA和所述至少一种Cas9蛋白切口酶,并且其中所述至少一种Cas9蛋白切口酶与所述两种或更多种RNA共定位至所述DNA靶标核酸,并且切割所述DNA靶标核酸,从而导致产生两个或更多个相邻切口。



CN 113846096 B

1.一种改变细胞中DNA双链靶标核酸的方法,包括

向所述细胞引入编码两种或更多种tracrRNA-crRNA融合体向导RNA的第一外源核酸,其中每种tracrRNA-crRNA融合体向导RNA包括间隔序列、示踪伴侣序列和tracr序列,所述tracr序列的一部分杂交至所述示踪伴侣序列,所述示踪伴侣序列和所述tracr序列通过接头核酸序列连接,且各间隔序列与所述DNA双链靶标核酸中的相邻位点互补,

向所述细胞引入编码至少一种Cas9蛋白切口酶的第二外源核酸,所述Cas9蛋白切口酶具有一个非活性核酸酶结构域并由所述两种或更多种tracrRNA-crRNA融合体向导RNA导向,且

其中表达所述两种或更多种tracrRNA-crRNA融合体向导RNA和所述至少一种Cas9蛋白切口酶,并且其中所述至少一种Cas9蛋白切口酶与所述两种或更多种tracrRNA-crRNA融合体向导RNA共定位至所述DNA双链靶标核酸,并且切割所述DNA双链靶标核酸,从而导致产生两个或更多个相邻切口,其中所述两个或更多个相邻切口在双链靶标核酸的不同链上,

其中所述方法用于非疾病治疗目的。

2.根据权利要求1所述的方法,其中所述两个或更多个相邻切口在产生双链断裂。

3.根据权利要求1所述的方法,其中所述两个或更多个相邻切口产生双链断裂,从而导致非同源末端接合。

4.根据权利要求1所述的方法,其中所述两个或更多个相邻切口彼此之间相对偏移。

5.根据权利要求1所述的方法,其中所述两个或更多个相邻切口彼此之间相对偏移,并产生双链断裂。

6.根据权利要求1所述的方法,其中所述两个或更多个相邻切口彼此之间相对偏移,并产生双链断裂,从而导致非同源末端接合。

7.根据权利要求1所述的方法,还包括向所述细胞引入编码供体核酸序列的第三外源核酸,其中所述两个或更多个切口导致产生所述双链靶标核酸与所述供体核酸序列的同源重组。

8.一种细胞,包括

编码两种或更多种tracrRNA-crRNA融合体向导RNA的第一外源核酸,其中每种tracrRNA-crRNA融合体向导RNA包括间隔序列、示踪伴侣序列和tracr序列,所述tracr序列的一部分杂交至所述示踪伴侣序列,所述示踪伴侣序列和所述tracr序列通过接头核酸序列连接,且各间隔序列与DNA双链靶标核酸中的相邻位点互补,和

编码至少一种具有一个非活性核酸酶结构域的Cas9蛋白切口酶的第二外源核酸,并且其中所述两种或更多种tracrRNA-crRNA融合体向导RNA和所述至少一种Cas9蛋白切口酶是所述DNA双链靶标核酸的共定位复合物的元件。

9.根据权利要求8所述的细胞,其中所述细胞是真核细胞。

10.根据权利要求8所述的细胞,其中所述细胞是酵母细胞、植物细胞或动物细胞。

11.根据权利要求8所述的细胞,其中所述tracrRNA-crRNA融合体向导RNA包括10至500个核苷酸。

12.根据权利要求8所述的细胞,其中所述tracrRNA-crRNA融合体向导RNA包括20至100个核苷酸。

13.根据权利要求8所述的细胞,其中所述靶标核酸与疾病或有害状况有关。

14.根据权利要求8所述的细胞,其中所述DNA双链靶标核酸是基因组DNA、线粒体DNA或外源DNA。

15.一种改变细胞中DNA双链靶标核酸的方法,包括

向所述细胞引入编码两种或更多种tracrRNA-crRNA融合体向导RNA的第一外源核酸,其中每种tracrRNA-crRNA融合体向导RNA包括间隔序列、示踪伴侣序列和tracr序列,所述tracr序列的一部分杂交至所述示踪伴侣序列,所述示踪伴侣序列和所述tracr序列通过接头核酸序列连接,且各间隔序列与所述DNA双链靶标核酸中的相邻位点互补,

向所述细胞引入编码至少一种Cas9蛋白切口酶的第二外源核酸,所述Cas9蛋白切口酶具有一个非活性核酸酶结构域并由两种或更多种tracrRNA-crRNA融合体向导RNA导向,且

其中表达所述两种或更多种tracrRNA-crRNA融合体向导RNA和所述至少一种Cas9蛋白切口酶,并且其中所述至少一种Cas9蛋白切口酶与所述两种或更多种tracrRNA-crRNA融合体向导RNA共定位至所述双链靶标核酸,并且切割所述双链靶标核酸,从而导致产生两个或更多个相邻切口,且

其中所述两个或更多个相邻切口在所述双链靶标核酸的不同链上,并且产生双链断裂,从而导致所述双链靶标核酸片段化,借此防止所述双链靶标核酸的表达,

其中所述方法用于非疾病治疗目的。

16.第一外源核酸和第二外源核酸在制备用于改变细胞中DNA双链靶标核酸的产品中的应用,其中

所述第一外源核酸编码两种或更多种tracrRNA-crRNA融合体向导RNA,其中每种tracrRNA-crRNA融合体向导RNA包括间隔序列、示踪伴侣序列和tracr序列,所述tracr序列的一部分杂交至所述示踪伴侣序列,所述示踪伴侣序列和所述tracr序列通过接头核酸序列连接,且各间隔序列与所述DNA双链靶标核酸中的相邻位点互补,

所述第二外源核酸编码至少一种Cas9蛋白切口酶,所述Cas9蛋白切口酶具有一个非活性核酸酶结构域并由所述两种或更多种tracrRNA-crRNA融合体向导RNA导向,且

其中由导入所述细胞所述外源核酸表达所述两种或更多种tracrRNA-crRNA融合体向导RNA和所述至少一种Cas9蛋白切口酶,并且其中所述至少一种Cas9蛋白切口酶与所述两种或更多种tracrRNA-crRNA融合体向导RNA共定位至所述DNA靶标核酸,并且切割所述DNA双链靶标核酸,从而导致产生两个或更多个相邻切口,其中所述两个或更多个相邻切口在双链靶标核酸的不同链上。

17.根据权利要求16所述的应用,其中所述两个或更多个相邻切口产生双链断裂。

18.根据权利要求16所述的应用,其中所述两个或更多个相邻切口产生双链断裂,从而导致非同源末端接合。

19.根据权利要求16所述的应用,其中所述两个或更多个相邻切口彼此之间相对偏移。

20.根据权利要求16所述的应用,其中所述两个或更多个相邻切口彼此之间相对偏移,并产生双链断裂。

21.根据权利要求16所述的应用,其中所述两个或更多个相邻切口彼此之间相对偏移,并产生双链断裂,从而导致非同源末端接合。

22.根据权利要求16所述的应用,还包括向所述细胞引入编码供体核酸序列的第三外源核酸,其中所述两个或更多个切口导致产生所述双链靶标核酸与所述供体核酸序列的同

源重组。

23.第一外源核酸和第二外源核酸在制备改变细胞中DNA双链靶标核酸的产品中的应用,

所述第一外源核酸编码两种或更多种tracrRNA-crRNA融合体向导RNA,其中每种tracrRNA-crRNA融合体向导RNA包括间隔序列、示踪伴侣序列和tracr序列,所述tracr序列的一部分杂交至所述示踪伴侣序列,所述示踪伴侣序列和所述tracr序列通过接头核酸序列连接,且各间隔序列与所述DNA双链靶标核酸中的相邻位点互补,

所述第二外源核酸编码至少一种Cas9蛋白切口酶,所述Cas9蛋白切口酶具有一个非活性核酸酶结构域并由两种或更多种tracrRNA-crRNA融合体向导RNA导向,且

其中由导入所述细胞的所述异源核酸表达所述两种或更多种tracrRNA-crRNA融合体向导RNA和所述至少一种Cas9蛋白切口酶,并且其中所述至少一种Cas9蛋白切口酶与所述两种或更多种tracrRNA-crRNA融合体向导RNA共定位至所述DNA双链靶标核酸,并且切割所述DNA双链靶标核酸,从而导致产生两个或更多个相邻切口,且

其中所述两个或更多个相邻切口在所述双链靶标核酸的不同链上,并且产生双链断裂,从而导致所述双链靶标核酸片段化,借此防止所述双链靶标核酸的表达。

RNA-导向的转录调控

[0001] 相关申请数据

[0002] 本申请主张2013年6月4日提交的美国临时专利申请No.61/830787的优先权,并且该专利申请出于所有目的以其全部内容作为参考并入本文。

[0003] 政府权益的声明

[0004] 本发明是在根据国家卫生研究院的授权号No.P50 HG005550和能源部的DE-FG02-02ER63445的政府资助下完成的。政府对本发明具有一定的权利。

背景技术

[0005] 细菌和古细菌CRISPR-Cas系统依赖于与侵袭性外源核酸内存在的指导互补序列降解的Cas蛋白的复合物中的短向导RNA。参见Deltcheva,E.等人,CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III.Nature 471,602-607(2011);Gasiunas,G.,Barrangou,R.,Horvath,P.&Siksnys,V.Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria.Proceedings of the National Academy of Sciences of the United States of America 109,E2579-2586(2012);Jinek,M.等人,A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity.Science 337,816-821(2012);Sapranauskas,R.等人,The Streptococcus thermophilus CRISPR/Cas system provides immunity in Escherichia coli.Nucleic acids research 39,9275-9282(2011);和Bhaya,D.,Davison,M.&Barrangou,R.CRISPR-Cas systems in bacteria and archaea:versatile small RNAs for adaptive defense and regulation.Annual review of genetics 45,273-297(2011)。近期化脓性链球菌(S.pyogenes)II型CRISPR系统的体外重组表明融合至正常反式编码的tracrRNA(“反式-激活CRISPR RNA”)的crRNA(“CRISPR RNA”)足以指导Cas9蛋白序列特异性地切割与crRNA匹配的靶标DNA序列。表达与靶标位点同源的gRNA导致Cas9招募和靶标DNA的降解。参见H.Deveau等人,Phage response to CRISPR-encoded resistance in Streptococcus thermophilus.Journal of Bacteriology 190,1390(Feb,2008)。

发明内容

[0006] 本发明公开的方面涉及向导RNA(guide RNA)、DNA结合蛋白和双链DNA靶标序列的复合物。根据某些方面,本发明公开范围内的DNA结合蛋白包括与向导RNA和与将复合物导向至双链DNA序列的向导RNA形成复合物的蛋白,其中所述复合物结合至所述DNA序列。本发明公开的这一方面可以称为RNA和DNA结合蛋白向或与双链DNA的共定位。以这种方式,DNA结合蛋白-向导RNA复合物可以用于定位转录调控因子蛋白或结构域于靶标DNA处,从而调控靶标DNA的表达。

[0007] 根据某些方面,提供了调控细胞中靶标核酸表达的方法,其包括向所述细胞引入编码与DNA(脱氧核糖核酸)互补的一种或多种RNA(核糖核酸)的第一外源核酸,其中所述

DNA包括靶标核酸,向所述细胞引入编码结合至DNA并由一种或多种RNA导向的RNA导向核酸酶无效的DNA结合蛋白(RNA guided nuclease-null DNA binding protein)的第二外源核酸,向所述细胞引入编码转录调控因子蛋白或结构域的第三外源核酸,其中表达了所述一种或多种RNA、RNA导向核酸酶无效的DNA结合蛋白和转录调控因子蛋白或结构域,其中所述一种或多种RNA、RNA导向核酸酶无效的DNA结合蛋白和转录调控因子蛋白或结构域共定位至所述DNA并且其中所述转录调控因子蛋白或结构域调控所述靶标核酸的表达。

[0008] 根据一个方面,编码RNA导向核酸酶无效的DNA结合蛋白的外源核酸还编码融合至RNA导向核酸酶无效的DNA结合蛋白的转录调控因子蛋白或结构域。根据一个方面,编码一种或多种RNA的外源核酸还编码RNA结合结构域的靶标,并且编码转录调控因子蛋白或结构域的外源核酸还编码融合至转录调控因子蛋白或结构域的RNA结合结构域。

[0009] 根据一个方面,所述细胞是真核细胞。根据一个方面,所述细胞是酵母细胞、植物细胞或动物细胞。根据一个方面,所述细胞是哺乳动物细胞。

[0010] 根据一个方面,所述RNA为约10至约500个核苷酸。根据一个方面,所述RNA为约20至约100个核苷酸。

[0011] 根据一个方面,所述转录调控因子蛋白或结构域是转录活化因子。根据一个方面,所述转录调控因子蛋白或结构域上调靶标核酸的表达。根据一个方面,所述转录调控因子蛋白或结构域上调靶标核酸的表达以治疗疾病或有害状况(detrimental condition)。根据一个方面,所述靶标核酸与疾病或有害状况有关。

[0012] 根据一个方面,一种或多种RNA是向导RNA。根据一个方面,一种或多种RNA是tracrRNA-crRNA融合体。根据一个方面,向导RNA包括间隔序列和示踪伴侣序列(tracer mate sequence)。向导RNA还可以包括tracr序列,所述序列的一部分杂交至tracr伴侣序列。向导RNA还可以包括接头核酸序列,其将示踪伴侣序列(tracer mate sequence)和tracr序列连接以产生tracrRNA-crRNA融合体。间隔序列结合至靶标DNA,如通过杂交。

[0013] 根据一个方面,向导RNA包括截短间隔序列。根据一个方面,向导RNA包括在间隔序列5'端具有1个碱基截短(truncation)的截短间隔序列(truncated spacer sequence)。根据一个方面,向导RNA包括在间隔序列5'端具有2个碱基截短的截短间隔序列。根据一个方面,向导RNA包括在间隔序列5'端具有3个碱基截短的截短间隔序列。根据一个方面,向导RNA包括在间隔序列5'端具有4个碱基截短的截短间隔序列。因此,间隔序列可以在间隔序列的5'端具有1至4个碱基截短。

[0014] 根据某些实施方式,间隔序列可以包括杂交至靶标核酸序列的约16至约20个核苷酸。根据某些实施方式,间隔序列可以包括杂交至靶标核酸序列的约20个核苷酸。

[0015] 根据某些方面,接头核酸序列可以包括约4至约6个核酸。

[0016] 根据某些方面,tracr序列可以包括约60至约500个核酸。根据某些方面,tracr序列可以包括约64至约500个核酸。根据某些方面,tracr序列可以包括约65至约500个核酸。根据某些方面,tracr序列可以包括约66至约500个核酸。根据某些方面,tracr序列可以包括约67至约500个核酸。根据某些方面,tracr序列可以包括约68至约500个核酸。根据某些方面,tracr序列可以包括约69至约500个核酸。根据某些方面,tracr序列可以包括约70至约500个核酸。根据某些方面,tracr序列可以包括约80至约500个核酸。根据某些方面,tracr序列可以包括约90至约500个核酸。根据某些方面,tracr序列可以包括约100至约500

个核酸。

[0017] 根据某些方面,tracr序列可以包括约60至约200个核酸。根据某些方面,tracr序列可以包括约64至约200个核酸。根据某些方面,tracr序列可以包括约65至约200个核酸。根据某些方面,tracr序列可以包括约66至约200个核酸。根据某些方面,tracr序列可以包括约67至约200个核酸。根据某些方面,tracr序列可以包括约68至约200个核酸。根据某些方面,tracr序列可以包括约69至约200个核酸。根据某些方面,tracr序列可以包括约70至约200个核酸。根据某些方面,tracr序列可以包括约80至约200个核酸。根据某些方面,tracr序列可以包括约90至约200个核酸。根据某些方面,tracr序列可以包括约100至约200个核酸。

[0018] 图5B中显示了示例性向导RNA。

[0019] 根据一个方面,所述DNA是基因组DNA、线粒体DNA、病毒DNA或外源DNA。

[0020] 根据某些方面,提供了调控细胞中靶标核酸表达的方法,其包括向所述细胞引入编码与DNA(脱氧核糖核酸)互补的一种或多种RNA(核糖核酸)的第一外源核酸,其中所述DNA包括靶标核酸,向所述细胞引入编码结合至DNA并由一种或多种RNA导向的II型CRISPR系统的RNA导向核酸酶无效的DNA结合蛋白的第二外源核酸,向所述细胞引入编码转录调控因子蛋白或结构域的第三外源核酸,其中表达了所述一种或多种RNA、II型CRISPR系统的RNA导向核酸酶无效的DNA结合蛋白和转录调控因子蛋白或结构域,其中所述一种或多种RNA、II型CRISPR系统的RNA导向核酸酶无效的DNA结合蛋白和转录调控因子蛋白或结构域共定位至所述DNA并且其中所述转录调控因子蛋白或结构域调控所述靶标核酸的表达。

[0021] 根据一个方面,编码II型CRISPR系统的RNA导向核酸酶无效的DNA结合蛋白的外源核酸还编码了融合至II型CRISPR系统的RNA导向核酸酶无效的DNA结合蛋白的转录调控因子蛋白或结构域。根据一个方面,编码一种或多种RNA的外源核酸还编码了RNA结合结构域的靶标,并且编码转录调控因子蛋白或结构域的外源核酸还编码了融合至转录调控因子蛋白或结构域的RNA结合结构域。

[0022] 根据一个方面,所述细胞是真核细胞。根据一个方面,所述细胞是酵母细胞、植物细胞或动物细胞。根据一个方面,所述细胞是哺乳动物细胞。

[0023] 根据一个方面,所述RNA为约10至约500个核苷酸。根据一个方面,所述RNA为约20至约100个核苷酸。

[0024] 根据一个方面,所述转录调控因子蛋白或结构域是转录活化因子。根据一个方面,所述转录调控因子蛋白或结构域上调靶标核酸的表达。根据一个方面,所述转录调控因子蛋白或结构域上调靶标核酸的表达以治疗疾病或有害状况。根据一个方面,所述靶标核酸与疾病或有害状况有关。

[0025] 根据一个方面,一种或多种RNA是向导RNA。根据一个方面,一种或多种RNA是tracrRNA-crRNA融合体。

[0026] 根据一个方面,所述DNA是基因组DNA、线粒体DNA、病毒DNA或外源DNA。

[0027] 根据某些方面,提供了调控细胞中靶标核酸表达的方法,其包括向所述细胞引入编码与DNA(脱氧核糖核酸)互补的一种或多种RNA(核糖核酸)的第一外源核酸,其中所述DNA包括靶标核酸,向所述细胞引入编码结合至DNA并由一种或多种RNA导向的核酸酶无效的Cas9蛋白的第二外源核酸,向所述细胞引入编码转录调控因子蛋白或结构域的第三外源

核酸,其中表达了所述一种或多种RNA、核酸酶无效的Cas9蛋白和转录调控因子蛋白或结构域,其中所述一种或多种RNA、核酸酶无效的Cas9蛋白和转录调控因子蛋白或结构域共定位至所述DNA并且其中所述转录调控因子蛋白或结构域调控所述靶标核酸的表达。

[0028] 根据一个方面,编码核酸酶无效的Cas9蛋白的外源核酸还编码了融合至核酸酶无效的Cas9蛋白的转录调控因子蛋白或结构域。根据一个方面,编码一种或多种RNA的外源核酸还编码了RNA结合结构域的靶标,并且编码转录调控因子蛋白或结构域的外源核酸还编码了融合至转录调控因子蛋白或结构域的RNA结合结构域。

[0029] 根据一个方面,所述细胞是真核细胞。根据一个方面,所述细胞是酵母细胞、植物细胞或动物细胞。根据一个方面,所述细胞是哺乳动物细胞。

[0030] 根据一个方面,所述RNA为约10至约500个核苷酸。根据一个方面,所述RNA为约20至约100个核苷酸。

[0031] 根据一个方面,所述转录调控因子蛋白或结构域是转录活化因子。根据一个方面,所述转录调控因子蛋白或结构域上调靶标核酸的表达。根据一个方面,所述转录调控因子蛋白或结构域上调靶标核酸的表达以治疗疾病或有害状况。根据一个方面,所述靶标核酸与疾病或有害状况有关。

[0032] 根据一个方面,一种或多种RNA是向导RNA。根据一个方面,一种或多种RNA是tracrRNA-crRNA融合体。

[0033] 根据一个方面,所述DNA是基因组DNA、线粒体DNA、病毒DNA或外源DNA。

[0034] 根据一个方面,提供了细胞,所述细胞包括编码与DNA互补的一种或多种RNA的第一外源核酸,其中所述DNA包括靶标核酸,编码RNA导向核酸酶无效的DNA结合蛋白的第二外源核酸和编码转录调控因子蛋白或结构域的第三外源核酸,其中所述一种或多种RNA、RNA导向核酸酶无效的DNA结合蛋白和转录调控因子蛋白或结构域是靶标核酸的共定位复合物的元件(成员,member)。

[0035] 根据一个方面,编码RNA导向核酸酶无效的DNA结合蛋白的外源核酸还编码融合至RNA导向核酸酶无效的DNA结合蛋白的转录调控因子蛋白或结构域。根据一个方面,编码一种或多种RNA的外源核酸还编码了RNA结合结构域的靶标,并且编码转录调控因子蛋白或结构域的外源核酸还编码了融合至转录调控因子蛋白或结构域的RNA结合结构域。

[0036] 根据一个方面,所述细胞是真核细胞。根据一个方面,所述细胞是酵母细胞、植物细胞或动物细胞。根据一个方面,所述细胞是哺乳动物细胞。

[0037] 根据一个方面,所述RNA为约10至约500个核苷酸。根据一个方面,所述RNA为约20至约100个核苷酸。

[0038] 根据一个方面,所述转录调控因子蛋白或结构域是转录活化因子。根据一个方面,所述转录调控因子蛋白或结构域上调靶标核酸的表达。根据一个方面,所述转录调控因子蛋白或结构域上调靶标核酸的表达以治疗疾病或有害状况。根据一个方面,所述靶标核酸与疾病或有害状况有关。

[0039] 根据一个方面,一种或多种RNA是向导RNA。根据一个方面,一种或多种RNA是tracrRNA-crRNA融合体。

[0040] 根据一个方面,所述DNA是基因组DNA、线粒体DNA、病毒DNA或外源DNA。

[0041] 根据某些方面,所述RNA导向核酸酶无效的DNA结合蛋白是II型CRISPR系统的RNA

导向核酸酶无效的DNA结合蛋白。根据某些方面,所述RNA导向核酸酶无效的DNA结合蛋白是核酸酶无效的Cas9蛋白。

[0042] 根据一个方面,提供了改变细胞中DNA靶标核酸的方法,所述方法包括向所述细胞引入编码两种或更多种RNA的第一外源核酸,其中每种RNA与DNA靶标核酸中相邻位点互补,向所述细胞引入编码至少一种RNA导向的DNA结合蛋白切口酶(nickase)并由两种或更多种RNA导向的第二外源核酸,其中表达了所述两种或更多种RNA和所述至少一种RNA导向的DNA结合蛋白切口酶并且其中所述至少一种RNA导向的DNA结合蛋白切口酶与所述两种或更多种RNA共定位至DNA靶标核酸并且切割(nick)DNA靶标核酸,从而导致产生了两个或更多个相邻的切口(nick)。

[0043] 根据一个方面,提供了改变细胞中DNA靶标核酸的方法,所述方法包括向所述细胞引入编码两种或更多种RNA的第一外源核酸,其中每种RNA与DNA靶标核酸中相邻位点互补,向所述细胞引入编码至少一种II型CRISPR系统的RNA导向的DNA结合蛋白切口酶并由两种或更多种RNA导向的第二外源核酸,其中表达了所述两种或更多种RNA和所述至少一个II型CRISPR系统的RNA导向的DNA结合蛋白切口酶并且其中所述至少一种II型CRISPR系统的RNA导向的DNA结合蛋白切口酶与所述两种或更多种RNA共定位至DNA靶标核酸并且切割DNA靶标核酸,从而导致产生了两个或更多个相邻的切口。

[0044] 根据一个方面,提供了改变细胞中DNA靶标核酸的方法,所述方法包括向所述细胞引入编码两种或更多种RNA的第一外源核酸,其中每种RNA与DNA靶标核酸中相邻位点互补,向所述细胞引入编码至少一种Cas9蛋白切口酶的第二外源核酸,所述切口酶具有一个非活性核酸酶结构域并由两种或更多种RNA导向,其中表达了所述两种或更多种RNA和所述至少一种Cas9蛋白切口酶并且其中所述至少一种Cas9蛋白切口酶与所述两种或更多种RNA共定位至DNA靶标核酸并且切割DNA靶标核酸,从而导致产生了两个或更多个相邻的切口。

[0045] 根据改变DNA靶标核酸的方法,所述两个或更多个相邻切口在双链DNA的相同链上。根据一个方面,所述两个或更多个相邻切口在双链DNA的相同链上并导致产生了同源重组。根据一个方面,所述两个或更多个相邻切口在双链DNA的不同链上。根据一个方面,所述两个或更多个相邻切口在双链DNA的不同链上并且产生了双链断裂。根据一个方面,所述两个或更多个相邻切口在双链DNA的不同链上并且产生了双链断裂,从而导致产生了非同源末端接合(nonhomologous end joining)。根据一个方面,所述两个或更多个相邻切口在双链DNA的不同链上并且彼此之间相对偏移。根据一个方面,所述两个或更多个相邻切口在双链DNA的不同链上并且彼此之间相对偏移,并产生了双链断裂。根据一个方面,所述两个或更多个相邻切口在双链DNA的不同链上并且彼此之间相对偏移,并产生了双链断裂,从而导致产生了非同源末端接合。根据一个方面,所述方法还包括向所述细胞引入编码供体核酸序列的第三外源核酸,其中所述两个或更多个切口导致产生了靶标核酸与供体核酸序列的同源重组。

[0046] 根据一个方面,提供了改变细胞中DNA靶标核酸的方法,所述方法包括向所述细胞引入编码两种或更多种RNA的第一外源核酸,其中每种RNA与DNA靶标核酸中相邻位点互补,向所述细胞引入编码至少一种RNA导向的DNA结合蛋白切口酶并且由两种或更多种RNA导向的第二外源核酸,并且其中表达了所述两种或更多种RNA和所述至少一种RNA导向的DNA结合蛋白切口酶,并且其中所述至少一种RNA导向的DNA结合蛋白切口酶与所述两种或更多种

RNA共定位至DNA靶标核酸并且切割DNA靶标核酸,从而导致产生了两个或更多个相邻切口,并且其中所述两个或更多个相邻切口在双链DNA的不同链上并产生了双链断裂,从而导致靶标核酸片段化(断裂,fragmentation),借此防止靶标核酸的表达。

[0047] 根据一个方面,提供了改变细胞中DNA靶标核酸的方法,所述方法包括向所述细胞引入编码两种或更多种RNA的第一外源核酸,其中每种RNA与DNA靶标核酸中相邻位点互补,向所述细胞引入编码至少一种II型CRISPR系统的RNA导向的DNA结合蛋白切口酶并且由两种或更多种RNA导向的第二外源核酸,并且其中表达了所述两种或更多种RNA和所述至少一个II型CRISPR系统的RNA导向的DNA结合蛋白切口酶,并且其中所述至少一种II型CRISPR系统的RNA导向的DNA结合蛋白切口酶与所述两种或更多种RNA共定位至DNA靶标核酸并且切割DNA靶标核酸,从而导致产生了两个或更多个相邻切口,并且其中所述两个或更多个相邻切口在双链DNA的不同链上并产生了双链断裂,从而导致靶标核酸片段化,借此防止靶标核酸的表达。

[0048] 根据一个方面,提供了改变细胞中DNA靶标核酸的方法,所述方法包括向所述细胞引入编码两种或更多种RNA的第一外源核酸,其中每种RNA与DNA靶标核酸中相邻位点互补,向所述细胞引入编码至少一种具有一个非活性核酸酶结构域的Cas9蛋白切口酶并且由两种或更多种RNA导向的第二外源核酸,并且其中表达了所述两种或更多种RNA和所述至少一种Cas9蛋白切口酶,并且其中所述至少一种Cas9蛋白切口酶与所述两种或更多种RNA共定位至DNA靶标核酸并且切割DNA靶标核酸,从而导致产生了两个或更多个相邻切口,并且其中所述两个或更多个相邻切口在双链DNA的不同链上并产生了双链断裂,从而导致靶标核酸片段化,借此防止靶标核酸的表达。

[0049] 根据一个方面,提供了细胞,所述细胞包括编码两种或更多种RNA的第一外源核酸,其中每种RNA与DNA靶标核酸中相邻位点互补,和编码至少一种RNA导向的DNA结合蛋白切口酶的第二外源核酸,并且其中所述两种或更多种RNA和所述至少一种RNA导向的DNA结合蛋白切口酶是DNA靶标核酸的共定位复合物的元件。

[0050] 根据一个方面,所述RNA导向的DNA结合蛋白切口酶是II型CRISPR系统的RNA导向的DNA结合蛋白切口酶。根据一个方面,所述RNA导向的DNA结合蛋白切口酶是具有一个非活性核酸酶结构域的Cas9蛋白切口酶。

[0051] 根据一个方面,所述细胞是真核细胞。根据一个方面,所述细胞是酵母细胞、植物细胞或动物细胞。根据一个方面,所述细胞是哺乳动物细胞。

[0052] 根据一个方面,所述RNA包括约10至约500个核苷酸。根据一个方面,所述RNA包括约20至约100个核苷酸。

[0053] 根据一个方面,所述靶标核酸与疾病或有害状况有关。

[0054] 根据一个方面,所述两种或更多种RNA是向导RNA。根据一个方面,所述两种或更多种RNA是tracrRNA-crRNA融合体。

[0055] 根据一个方面,所述DNA靶标核酸是基因组DNA、线粒体DNA、病毒DNA或外源DNA。

[0056] 在以下本发明实施方式和附图的说明中并且根据权利要求,本发明某些实施方式的其它特征和优势将变得更加充分而明显。

附图说明

[0057] 本专利或专利申请文件包含彩图。通过索取并付出必要的费用,官方将提供带有彩图的本专利或专利申请公开的复制件。根据以下与附图结合的说明性实施方式的详细说明,将更充分地理解本发明实施方式的上述及其它特征和优势,其中:

[0058] 图1A和图1B是RNA导向的转录激活的示意图。图1C是报告子构建体的设计。图1D显示了数据,该数据表明Cas9N-VP64融合体显示了RNA导向的转录激活,如通过荧光激活细胞分类术(荧光激活细胞分选,FACS)和免疫荧光测定(IF)两者所测定的。图1E显示了FACS的测定数据,1F显示了在存在Cas9N、MS2-VP64和具有适当MS2适体(aptamer)结合位点的gRNA的情况下来自报告子构建体的gRNA序列特异性转录激活。图1F显示了数据,该数据显示了通过单个gRNA和多个gRNA的转录诱导。

[0059] 图2A显示了评价Cas9-gRNA复合物和TALE靶向谱图(landscape)的方法。图2B显示了数据,该数据表明平均来说,Cas9-gRNA复合物对其靶标序列中的1-3个突变耐受。图2C显示了数据,该数据表明除了定位至PAM序列的那些之外,Cas9-gRNA复合物在很大程度上对点突变不敏感。图2D显示了热图数据(heat plot data),其表明2个碱基错配的引入显著损害了Cas9-gRNA复合物活性。图2E显示了数据,该数据表明平均来说,18聚体(18-mer)TALE显示对其靶标序列中的1-2个突变耐受。图2F显示了数据,该数据表明与Cas9-gRNA复合物类似,18聚体TALE在很大程度上对其靶标中错配的单碱基不敏感。图2G显示了热图数据,其表明2个碱基错配的引入显著损害了18聚体TALE的活性。

[0060] 图3A显示了向导RNA设计的示意图。图3B显示了数据,该数据显示了对于导致5'突出的偏移切口(off-set nicks)和导致3'突出的偏移切口的非同源末端接合的百分比。图3C显示了数据,该数据显示了对于导致5'突出(overhang)的偏移切口和导致3'突出的偏移切口的靶向(targeting)的百分比。

[0061] 图4A是RuvC PDB ID:4EP4(蓝色)位置D7(左图)中金属配位残基的示意图,来自PDB ID的HNH核酸内切酶结构域的示意图:3M7K(橙色)和4H9D(青色),其包括配位的Mg-离子(灰色球体)和来自3M7K的DNA(紫色)(中图),和分析的突变体列表(右图)。图4B显示了数据,该数据显示了对于Cas9突变体m3和m4的不可检测的核酸酶活性,以及它们各个与VP64的融合。图4C是图4B中数据的高分辨率检查。

[0062] 图5A是确定Cas9-gRNA活性的同源重组测定的示意图。图5B显示了具有随机序列插入的向导RNA以及同源重组的百分比。

[0063] 图6A是OCT4基因的向导RNA的示意图。图6B显示了对于启动子-荧光素酶报告子构建体的转录激活。图6C显示了通过内源基因的qPCR的转录激活。

[0064] 图7A是REX1基因的向导RNA的示意图。图7B显示了对于启动子-荧光素酶报告子构建体的转录激活。图7C显示了通过内源基因的qPCR的转录激活。

[0065] 图8A显示了用于计算归一化(normalized)表达水平的高水平特异性分析处理流程的示意图。图8B显示了对于偏倚构建体文库(biasedconstruct library)内产生的错配数目的结合位点百分比的分布数据。左图:理论分布。右图:从真实(实际的,actual)TALE构建体文库观察到的分布。图8C显示了对于错配数目的聚集至结合位点的标签计数(tag count)的百分比分布数据。左图:从阳性对照样品观察的分布。右图:从其中诱导了非控制TALE(non-control TALE)的样品观察的分布。

[0066] 图9A显示了对于Cas9-gRNA复合物的靶标谱图的分析数据,其显示了对其靶标序列中1-3个突变的耐受。图9B显示了对于Cas9-gRNA复合物的靶标谱图的分析数据,其显示除了定位至PAM序列的那些之外,对点突变不敏感。图9C显示了对于Cas9-gRNA复合物的靶向谱图分析的热图数据,该数据显示2个碱基错配的引入显著损害了活性。图9D显示了来自核酸酶介导的HR测定的数据,其确认了化脓性链球菌(S.pyogenes)Cas9预测的PAM是NGG,还有NAG。

[0067] 图10A显示了来自核酸酶介导的HR测定的数据,其确认18聚体TALE耐受其靶标序列中的多个突变。图10B显示了3种不同大小(18聚体、14聚体和10聚体)的TALE的靶向谱图的分析数据。图10C显示10聚体TALE的数据,显示了接近单碱基错配的分辨率。图10D显示10聚体TALE的热图数据,显示了接近单碱基错配的分辨率。

[0068] 图11A显示了设计的向导RNA。图11B显示了多个向导RNA的非同源末端接合的百分比。

[0069] 图12A显示了Sox2基因。图12B显示了Nanog基因。

[0070] 图13A-13F显示了两个其它Cas9-gRNA复合物的靶向图谱。

[0071] 图14A显示了两种gRNA(野生型(SEQ ID NO:88)和突变型(SEQ ID NO:89-90))的特异性谱图。用红色突出显示序列差异。图14B和14C显示了该测定对正在评价的gRNA是特异性的(根据图13D重新作图的数据)。

[0072] 图15A-15D显示了在相对于靶标的间隔序列中具有单碱基或双碱基错配(红色突出显示)的gRNA2(图15A-B)和gRNA3(图15C-D)。序列如SEQ ID NO:91-131中所示。

[0073] 图16A-16D显示了使用核酸酶测定测试的2个独立gRNA:在它们间隔子的5'端具有截短的gRNA1(图16A-B)和gRNA3(图16C-D)。序列如SEQ ID NO:66、185-186和133-140所示。

[0074] 图17A-17B显示了使用核酸酶介导的HR测定确认了化脓性链球菌(S.pyogenes)Cas9的PAM是NGG,还有NAG。序列如SEQ ID NO:67-69和141所示。

[0075] 图18A-18B显示了使用核酸酶介导的HR测定确认18聚体TALE耐受其靶标序列中的多个突变。序列如SEQ ID NO:70-73所示。

[0076] 图19A-19C显示了TALE单体特异性相对于TALE蛋白特异性的比较。序列如SEQ ID NO:142-150所示。

[0077] 图20A-20B显示了与偏移切口(off-set nicking)有关的数据。序列如SEQ ID NO:151-158所示。

[0078] 图21A-21C显示了偏移切口和NHEJ谱图。序列如SEQ ID NO:159-184和187所示。

具体实施方式

[0079] 本发明公开的实施方式基于以调控靶标核酸的方式将转录调控因子蛋白或结构域共定位至DNA的DNA结合蛋白的应用。本领域技术人员容易知晓这些DNA结合蛋白出于多种目的结合至DNA。这些DNA结合蛋白可以是天然存在的。包括在本发明公开范围内的DNA结合蛋白包括可以由在本文中称为向导RNA(guide RNA)的RNA导向的那些。根据这一方面,向导RNA和RNA导向的DNA结合蛋白在DNA上形成了共定位复合物。根据某些方面,所述DNA结合蛋白可以是核酸酶无效的DNA结合蛋白(nuclease-null DNA binding protein)。根据这一方面,所述核酸酶无效的DNA结合蛋白可以通过具有核酸酶活性的DNA结合蛋白的改变或修

饰产生。具有核酸酶活性的这些DNA结合蛋白是本领域技术人员已知的,并且包括天然存在的具有核酸酶活性的DNA结合蛋白,如存在于(例如)II型CRISPR系统中的Cas9蛋白。在本领域中很好地记录了这些Cas9蛋白和II型CRISPR系统。参见Makarova等人,Nature Reviews, Microbiology,第9卷,June 2011,第467-477页(包括所有补充信息),其以其全部内容作为参考并入本文。

[0080] 示例性的具有核酸酶活性的DNA结合蛋白起作用来造成切口或切割双链DNA。这些核酸酶活性可以通过具有显示出核酸酶活性的一个或多个多肽序列的DNA结合蛋白产生。这些示例性DNA结合蛋白可以具有两个不同的核酸酶结构域,其中每个结构域分别负责切割双链DNA的特定链或为其造成切口。本领域技术人员已知的示例性的具有核酸酶活性的多肽序列包括McrA-HNH核酸酶相关结构域和RuvC-样核酸酶结构域。因此,示例性DNA结合蛋白是本质上含有McrA-HNH核酸酶相关结构域和RuvC-样核酸酶结构域中的一个或多个的那些。根据某些方面,改变或另外修饰所述DNA结合蛋白以使核酸酶活性失活。这些改变或修饰包括改变一个或多个氨基酸以使核酸酶活性或核酸酶结构域失活。这些修饰包括除去显示出核酸酶活性的一个或多个多肽序列(即核酸酶结构域),从而使得在所述DNA结合蛋白中不存在显示出核酸酶活性的一个或多个多肽序列(即核酸酶结构域)。基于本发明公开,使核酸酶活性失活的其它修饰将对本领域技术人员是显而易见的。因此,核酸酶无效的DNA结合蛋白包括经修饰而使核酸酶活性失活的多肽序列或使核酸酶活性失活的一个或多个多肽序列的去除。即使核酸酶活性失活,核酸酶无效的DNA结合蛋白保留了与DNA结合的能力。因此,DNA结合蛋白包括DNA结合所需的一个或多个多肽序列,但是可以缺少显示出核酸酶活性的一种或多种或者全部核酸酶序列。因此,DNA结合蛋白包括DNA结合所需的一个或多个多肽序列,但是可以具有显示出失活的核酸酶活性的一种或多种或者全部核酸酶序列。

[0081] 根据一个方面,可以修饰或改变具有两个或更多个核酸酶结构域的DNA结合蛋白使除所述核酸酶结构域之一以外的核酸酶结构域都失活。该修饰或改变的DNA结合蛋白被称为DNA结合蛋白切口酶,从而所述DNA结合蛋白仅对双链DNA中的一条链切割或造成切口。当通过RNA导向至DNA时,所述DNA结合蛋白切口酶被称为RNA导向的DNA结合蛋白切口酶。

[0082] 示例性的DNA结合蛋白是缺少核酸酶活性的II型CRISPR系统的RNA导向的DNA结合蛋白。示例性的DNA结合蛋白是核酸酶无效的Cas9蛋白。示例性的DNA结合蛋白是Cas9蛋白切口酶。

[0083] 在化脓性链球菌(S.pyogenes)中,通过蛋白质中两个催化结构域:切割DNA互补链的HNH结构域和切割非互补链的RuvC样结构域介导的过程,Cas9产生了前间区序列邻近基序(PAM,protospacer-adjacent motif)上游3bp的钝性末端(平头末端,blunt-ended)双链断裂。参见Jinke等人,Science 337,816-821(2012),该文献以其全部内容作为参考并入本文。已知Cas9蛋白存在于多种II型CRISPR系统中,包括下列:如在Makarova等人,Nature Reviews,Microbiology,第9卷,June 2011,第467-477页的补充信息中鉴别的:海沼甲烷球菌(Methanococcus maripaludis)C7;白喉棒杆菌(Corynebacterium diphtheriae);有效棒杆菌(Corynebacterium efficiens)YS-314;乳发酵短杆菌(Corynebacterium glutamicum)ATCC 13032Kitasato;乳发酵短杆菌(Corynebacterium glutamicum)ATCC 13032Bielefeld;乳发酵短杆菌R(Corynebacterium glutamicum R);克氏棒杆菌

(Corynebacterium kroppenstedtii)DSM 44385;脓肿分枝杆菌(Mycobacterium abscessus)ATCC 19977;皮诺卡氏菌(Nocardia farcinica)IFM10152;灰暗诺卡氏菌诺卡氏菌(Rhodococcus erythropolis)PR4;Rhodococcus jostii RHA1;混浊红球菌(Rhodococcus opacus)B4 uid36573;解纤维热酸菌(Acidothermus cellulolyticus)11B;氯酚节杆菌(Arthrobacter chlorophenolicus)A6;Kribbella flavida DSM 17836uid43465;弯曲嗜热单孢菌(Thermomonospora curvata)DSM 43183;齿双歧杆菌(Bifidobacterium dentium)Bdl;长双歧杆菌(Bifidobacterium longum)DJO10A;Slackia heliotrinireducens DSM 20476;Persephonella marina EX HI;脆弱拟杆菌(Bacteroides fragilis)NCTC 9434;黄褐二氧化碳嗜纤维菌(Capnocytophaga ochracea)DSM 7271;嗜冷黄杆菌(Flavobacterium psychrophilum)JIP02 86;嗜粘蛋白阿克曼氏菌(Akkermansia muciniphila)ATCC BAA 835;卡斯特玫瑰弯菌(Roseiflexus castenholzii)DSM 13941;玫瑰弯菌(Roseiflexus)RSI;集胞蓝细菌属(Synechocystis)PCC6803;微小亚历山大藻(Elusimicrobium minutum)Peil91;未培养的白蚁1组细菌种系型(uncultured Termite group 1bacterium phylotype)Rs D17;产琥珀酸丝状杆菌(Fibrobacter succinogenes)S85;蜡状芽孢杆菌(Bacillus cereus)ATCC 10987;英诺克李斯特菌(Listeria innocua);干酪乳杆菌(Lactobacillus casei);鼠李糖乳杆菌(Lactobacillus rhamnosus)GG;唾液乳杆菌(Lactobacillus salivarius)UCC118;无乳链球菌(Streptococcus agalactiae)A909;无乳链球菌(Streptococcus agalactiae)NEM316;无乳链球菌(Streptococcus agalactiae)2603;停乳链球菌马样亚种(Streptococcus dysgalactiae equisimilis)GGS 124;马链球菌兽瘟亚种(Streptococcus equi zooepidemicus)MGCS10565;解没食子酸链球菌(Streptococcus gallolyticus)UCN34 uid46061;格氏链球菌(Streptococcus gordonii)Challis亚种CHI;变形链球菌(Streptococcus mutans)NN2025 uid46353;变形链球菌(Streptococcus mutans);酿脓链球菌(Streptococcus pyogenes)M1 GAS;酿脓链球菌(Streptococcus pyogenes)MGAS5005;酿脓链球菌(Streptococcus pyogenes)MGAS2096;酿脓链球菌(Streptococcus pyogenes)MGAS9429;酿脓链球菌(Streptococcus pyogenes)MGAS10270;酿脓链球菌(Streptococcus pyogenes)MGAS6180;酿脓链球菌(Streptococcus pyogenes)MGAS315;酿脓链球菌(Streptococcus pyogenes)SSI-1;酿脓链球菌(Streptococcus pyogenes)MGAS10750;酿脓链球菌(Streptococcus pyogenes)NZ131;嗜热链球菌(Streptococcus thermophiles)CNRZ1066;嗜热链球菌(Streptococcus thermophiles)LMD-9;嗜热链球菌(Streptococcus thermophiles)LMG 18311;肉毒梭状芽孢杆菌(Clostridium botulinum)A3 Loch Maree;肉毒梭状芽孢杆菌(Clostridium botulinum)BEklund 17B;肉毒梭状芽孢杆菌(Clostridium botulinum)Ba4 657;肉毒梭状芽孢杆菌(Clostridium botulinum)F Langeland;解纤维梭菌(Clostridium cellulolyticum)H10;大芬戈尔德菌(Finegoldia magna)ATCC 29328;直肠真杆菌(Eubacterium rectale)ATCC 33656;鸡败血支原体(Mycoplasmagallisepticum);运动支原体(Mycoplasma mobile)163K;穿透支原体(Mycoplasma penetrans);关节液支原体(Mycoplasma synoviae)53;鼠放线菌(Streptobacillus moniliformis)DSM 12112;慢生根瘤菌(Bradyrhizobium)BTAi1;汉堡硝化杆菌(Nitrobacter hamburgensis)X14;Rhodopseudomonas palustris

BisB18;Rhodopseudomonas palustris BisB5;食清洁剂细小棒菌(Parvibaculum lavamentivorans)DS-1;Dinoroseobacter shibae DFL 12;重氮营养葡糖酸醋杆菌(Gluconacetobacter diazotrophicus)Pal 5FAPERJ;重氮营养葡糖酸醋杆菌(Gluconacetobacter diazotrophicus)Pal 5JGI;固氮螺菌(Azospirillum)B510 uid46085;深红红螺菌(Rhodospirillum rubrum)ATCC 11170;Diaphorobacter TPSY uid29975;艾森虹蝴肾杆菌(Verminephrobacter eiseniae)EF01-2;脑膜炎奈瑟氏球菌(Neisseria meningitides)053442;脑膜炎奈瑟氏球菌(Neisseriameningitides)α14;脑膜炎奈瑟氏球菌(Neisseria meningitides)Z2491;需盐脱硫弧菌(Desulfovibrio salexigens)DSM 2638;空肠弯曲菌(Campylobacter jejuni)doylei 269 97;空肠弯曲菌(Campylobacter jejuni)81116;空肠弯曲菌(Campylobacter jejuni);红嘴鸥弯曲杆菌(Campylobacter lari)RM2100;肝螺杆菌(Helicobacter hepaticus);产琥珀酸沃林氏菌(Wolinella succinogenes);奥氏甲苯单胞菌(Tolumonas auensis)DSM9187;大西洋假交替单胞菌(Pseudoalteromonas atlantica)T6c;乌贼希瓦菌(Shewanella pealeana)ATCC 700345;侵肺军团菌(Legionellapneumophila)Paris;产琥珀酸放线杆菌(Actinobacillus succinogenes)130Z;多杀杆菌(Pasteurella multocida);野兔热弗朗西丝氏菌(Francisellatularensis)novicida U112;野兔热弗朗西丝氏菌(Francisella tularensis)holarctica;野兔热弗朗西丝氏菌(Francisella tularensis)FSC 198;野兔热弗朗西丝氏菌(Francisella tularensis)tularensis;野兔热弗朗西丝氏菌(Francisella tularensis)WY96-3418;和齿垢密螺旋体(Treponema denticola)ATCC 35405。因此,本发明公开的方面涉及存在于II型CRISPR系统中的Cas9蛋白,其使得核酸酶无效或使得成为如本文所述的切口酶。

[0084] 在文献中,本领域技术人员可以将Cas9蛋白称为Csn1。以下显示了作为本文所述实验的主体的化脓性链球菌(S.pyogenes)Cas9蛋白序列。参见Deltcheva等人,Nature 471,602-607(2011),该文献以其全部内容作为参考并入本文。

[0085]

MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAE
ATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFG
NIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSD
VDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGN
LIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA
GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELH
AILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEE
VVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFL
SGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG
RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSL
HEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRER
MKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDH
IVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKS
KLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVR
K

MIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDF
ATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVA
YSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK
YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVE
QHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGA
PAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD-

[0086] 根据本文所述的RNA导向的基因组调控方法的某些方面,改变Cas9以降低、显著降低或消除核酸酶活性。根据一个方面,通过改变RuvC核酸酶结构域或HNH核酸酶结构域来降低、显著降低或消除Cas9核酸酶活性。根据一个方面,使RuvC核酸酶结构域失活。根据一个方面,使HNH核酸酶结构域失活。根据一个方面,使RuvC核酸酶结构域和HNH核酸酶结构域失活。根据其它方面,提供了Cas9蛋白,其中RuvC核酸酶结构域和HNH核酸酶结构域失活。根据其它方面,提供了RuvC核酸酶结构域和HNH核酸酶结构域失活的程度的核酸酶无效的Cas9蛋白。根据其它方面,提供了Cas9切口酶,其中RuvC核酸酶结构域或者HNH核酸酶结构域中的任一个失活,借此使剩余的核酸酶结构域对于核酸酶活性保持活性。以这样的方式,仅对双链DNA的一条链切割或造成切口。

[0087] 根据其它方面,提供了核酸酶无效的Cas9蛋白,其中改变或另外除去Cas9中的一个或多个氨基酸以提供核酸酶无效的Cas9蛋白。根据一个方面,所述氨基酸包括D10和H840。参见Jinke等人,Science 337,816-821(2012)。根据其它方面,所述氨基酸包括D839和N863。根据一个方面,用降低、基本消除或消除核酸酶活性的氨基酸取代D10、H840、D839和H863中的一个或多个。根据一个方面,用丙氨酸取代D10、H840、D839和H863中的一个或多个或全部。根据一个方面,具有被降低、基本消除或消除核酸酶活性的氨基酸(如丙氨酸)取代的D10、H840、D839和H863中的一个或多个或全部的Cas9蛋白被称为核酸酶无效的Cas9或

Cas9N并且其显示出降低或消除的核酸酶活性,或者在检测水平范围内不存在或基本不存在核酸酶活性。根据这一方面,使用已知测定,Cas9N的核酸酶活性可以是不可检测的,即低于已知测定的检测水平。

[0088] 根据一个方面,核酸酶无效的Cas9蛋白包括保留了所述蛋白质结合DNA并由RNA导向的能力的它的同源物(homolog)和直系同源物(ortholog)。根据一个方面,核酸酶无效的Cas9蛋白包括如对来自化脓性链球菌(S.pyogenes)的天然存在的Cas9所描述的并且具有被丙氨酸取代的D10、H840、D839和H863中的一个或多个或全部的序列,和与其具有至少30%、40%、50%、60%、70%、80%、90%、95%、98%或99%的同源性并且是DNA结合蛋白的蛋白质序列,如RNA导向的DNA结合蛋白。

[0089] 根据一个方面,核酸酶无效的Cas9蛋白包括除了RuvC核酸酶结构域和HNH核酸酶结构域的蛋白质序列之外的如对来自化脓性链球菌(S.pyogenes)的天然存在的Cas9所描述的序列,以及与其具有至少30%、40%、50%、60%、70%、80%、90%、95%、98%或99%的同源性并且是DNA结合蛋白的蛋白质序列,如RNA导向的DNA结合蛋白。以这样的方式,本发明公开的方面包括负责DNA结合,例如,负责与向导RNA共定位和结合至DNA的蛋白质序列以及与其同源的蛋白质序列,并且不需要包括RuvC核酸酶结构域和HNH核酸酶结构域的蛋白质序列(至DNA结合不需要的程度),这是因为这些结构域可以是失活的或从天然存在的Cas9蛋白的蛋白质序列中除去以产生核酸酶无效的Cas9蛋白。

[0090] 出于本发明公开的目的,图4A显示了与Cas9具有同源性的已知蛋白质结构中的金属配位残基。基于Cas9序列中的位置标记残基。左图:RuvC结构,PDB ID:4EP4(蓝色)位置D7,其对应于Cas9序列中的D10,并将其在Mg-离子配位位置突出显示。中图:来自PDB ID:3M7K(橙色)和4H9D(青色)的HNH核酸内切酶结构域的结构,其包括配位的Mg-离子(灰色球体)和来自3M7K的DNA(紫色)。3M7K和4H9D位置D53和N77中的残基D92和N113,其对Cas9氨基酸D839和N863具有序列同源性,并将其显示为棒状(stick)。右图:制备并分析核酸酶活性的突变体列表:Cas9野生型;用丙氨酸取代D10的Cas9m1;用丙氨酸取代D10并用丙氨酸取代H840的Cas9m2;用丙氨酸取代D10、用丙氨酸取代H840和用丙氨酸取代D839的Cas9m3;以及用丙氨酸取代D10、用丙氨酸取代H840、用丙氨酸取代D839、用丙氨酸取代N863的Cas9m4。

[0091] 如图4B所示,通过靶标位点的深度测序,Cas9突变体:m3和m4以及它们各自与VP64的融合体显示出不可检测的核酸酶活性。图显示了相对于基因组位置的突变频率,其中红线划分了gRNA靶标。图4C是图4B中数据的高分辨率检查,并且该图确认突变图谱(mutation landscape)显示了与未修饰位点可比较的谱图。

[0092] 根据一个方面,提供了工程设计的Cas9-gRNA系统,其通过将转录激活结构域连接至核酸酶无效的Cas9或向导RNA使得能够在人细胞中进行RNA导向的基因组调控。根据本发明公开的一个方面,将一个或多个转录调节蛋白或结构域(这些术语是可互换使用的)结合或以其他方式连接到核酸酶缺乏型Cas9或者一个或多个向导RNA(gRNA)。转录调控结构域对应于靶标位点。因此,本发明公开的方面包括通过将这些结构域融合、连接或结合至Cas9N或gRNA中的任一个来将转录调控结构域定位至靶标位点的方法和材料。

[0093] 根据一个方面,提供了能够转录激活的Cas9N-融合蛋白。根据一个方面,VP64激活结构域(参见Zhang等人,Nature Biotechnology 29,149-153(2011),该文献以其全部内容作为参考并入本文)结合、融合、连接或以其他方式链接(tether)至Cas9N的C末端。根据一

个方法,通过Cas9N蛋白向靶标基因组DNA的位点提供了转录调控结构域。根据一个方法,在细胞内与一个或多个向导RNA一起提供了融合至转录调控结构域的Cas9N。具有融合其上的转录调控结构域的Cas9N结合在靶标基因组DNA处或附近。一个或多个向导RNA结合在靶标基因组DNA处或附近。转录调控结构域调控靶标基因的表达。根据具体方面,当与启动子附近的gRNA靶标序列结合时,Cas9N-VP64融合体激活了报告子构建体的转录,借此显示RNA导向的转录激活。

[0094] 根据一个方面,提供了能够转录激活的gRNA-融合蛋白。根据一个方面,将VP64激活结构域结合、融合、连接或以其他方式链接至gRNA。根据一个方法,通过gRNA向靶标基因组DNA的位点提供了转录调控结构域。根据一个方法,在细胞内与Cas9N蛋白一起提供了融合至转录调控结构域的gRNA。Cas9N结合在靶标基因组DNA处或附近。具有融合其上的转录调控蛋白或结构域的一个或多个向导RNA结合在靶标基因组DNA处或附近。转录调控结构域调控靶标基因的表达。根据具体方面,Cas9N蛋白和与转录调控结构域融合的gRNA激活了报告子构建体的转录,借此显示RNA导向的转录激活。

[0095] 通过将随机序列插入到gRNA并测定Cas9的功能,鉴别了将耐受修饰的gRNA的区域,从而构建了能够转录调控的gRNA链(tether)。在嵌合gRNA的crRNA部分的5'端或tracrRNA部分的3'端具有随机序列插入的gRNA保留了功能性,同时嵌合gRNA的tracrRNA支架部分中的插入导致功能丧失。参见图5A-B,其总结了gRNA对随机碱基插入的灵活性。图5A是确定Cas9-gRNA活性的同源重组(HR)测定的示意图。如图5B所示,在嵌合gRNA的crRNA部分的5'端或tracrRNA部分的3'端具有随机序列插入的gRNA保留了功能性,同时嵌合gRNA的tracrRNA支架部分中的插入导致功能丧失。通过红色核苷酸显示了gRNA序列中的插入点。不希望受科学理论的束缚,5'末端处随机碱基插入的活性提高可以是由于更长的gRNA的半衰期的提高所造成的。

[0096] 要将VP64连接至gRNA,将MS2噬菌体外壳蛋白结合RNA茎环的两个拷贝添加(附连,append)至gRNA的3'端。参见Fusco等人,CurrentBiology.CB13,161-167(2003),该文献以其全部内容作为参考并入本文。与Cas9N和MS2-VP64融合蛋白一起表达了这些嵌合gRNA。在存在所有3种组分的情况下观察到了来自报告子构建体的序列特异性转录激活。

[0097] 图1A是RNA导向的转录激活的示意图。如图1A所示,为了产生能够转录激活的Cas9N-融合蛋白,将VP64激活结构域直接链接至Cas9N的C末端。如图1B所示,为了产生能够转录激活的gRNA链,将MS2噬菌体外壳蛋白结合RNA茎环的两个拷贝添加至gRNA的3'端。与Cas9N和MS2-VP64融合蛋白一起表达了这些嵌合gRNA。图1C显示了用于测定转录激活的报告子构建体的设计。两个报告子具有不同的gRNA靶标位点,并且共有控制TALE-TF靶标位点(control TALE-TF target site)。如图1D所示,Cas9N-VP64融合体显示了RNA导向的转录激活,如通过荧光激活细胞分类术(FACS)和免疫荧光测定(IF)两者所测定的。具体地,尽管控制TALE-TF激活两个报告子,但是Cas9N-VP64融合体以gRNA序列特异性的方式激活报告子。如图1E所示,仅在存在所有3种组分:Cas9N、MS2-VP64和具有适当的MS2适体结合位点的gRNA时,通过FACS和IF两者观察到了来自报告子构建体的gRNA序列特异性转录激活。

[0098] 根据某些方面,提供了使用Cas9N、一个或多个gRNA和转录调控蛋白或结构域的用于调控内源基因的方法。根据一个方面,内源基因可以是任何所需基因,在本文中称为靶标基因。根据一个示例性方面,用于调控的基因靶标包括ZFP42(REX1)和POU5F1(OCT4),两者

均紧密调控参与多潜能性维持的基因。如图1F所示,对REX1基因设计了靶向转录起始位点的DNA上游的～5kb延伸的10个gRNA(用绿色突出显示了DNA酶超敏感位点)。使用启动子-荧光素酶报告子构建体(参见Takahashi等人,Cell 131 861-872(2007),该文献以其全部内容作为参考并入本文)或者直接通过内源基因的qPCR测定转录激活。

[0099] 图6A-C涉及使用Cas9N-VP64的RNA导向的OCT4调控。如图6A所示,对OCT4基因设计了靶向转录起始位点的DNA上游～5kb延伸的21个gRNA。用绿色突出了DNA酶超敏感位点。图6B显示了使用启动子-荧光素酶报告子构建体的转录激活。图6C显示了直接通过内源基因的qPCR的转录激活。尽管单个gRNA的引入适度刺激了转录,但是多个gRNA协同作用以刺激稳健的多倍转录激活。

[0100] 图7A-C涉及使用Cas9N、MS2-VP64和gRNA+2X-MS2适体的RNA导向的REX1调控。如图7A所示,对REX1基因设计了靶向转录起始位点的DNA上游的～5kb延伸的10个gRNA。用绿色突出了DNA酶超敏感位点。图7B显示了使用启动子-荧光素酶报告子构建体的转录激活。图7C显示了直接通过内源基因的qPCR的转录激活。尽管单个gRNA的引入适度刺激了转录,但是多个gRNA协同作用以刺激稳健的多倍转录激活。在一个方面,gRNA上2X-MS2适体的缺少不会导致转录激活。参见Maeder等人,Nature Methods 10,243-245(2013)和Perez-Pinera等人,Nature Methods 10,239-242(2013),以上每篇文献以其全部内容作为参考并入本文。

[0101] 因此,方法涉及使用多个向导RNA与Cas9N蛋白和转录调控蛋白或结构域以调控靶标基因的表达。

[0102] Cas9和gRNA链接法均是有效的,其中前者显示出～1.5-2倍更高的效力。与3组分复合物装配体(复合物组装体,complex assembly)相反,这种差异可能是由于对2-组分的需要所造成的。然而,在原理上,gRNA链接法能够使不同的gRNA募集不同的效应子结构域,只要每个gRNA使用不同的RNA-蛋白质相互作用对。参见Karyer-Bibens等人,Biology of the Cell/Under the Auspices of the European Cell Biology Organization 100,125-138(2008),该文献以其全部内容作为参考并入本文。根据本发明公开的一个方面,可以使用特异性向导RNA和一般性Cas9N蛋白(即对于不同的靶标基因,相同或类似的Cas9N蛋白)调控不同的靶标基因。根据一个方面,提供了使用相同或相似的Cas9N的多基因调控(multiplex gene regulation)方法。

[0103] 本发明公开所述的方法还涉及使用本文所述的Cas9N蛋白和向导RNA编辑靶标基因以提供人细胞的多重基因工程和表观遗传工程(epigenetic engineering)。以Cas9-gRNA靶向作为问题(参见Jiang等人,Nature Biotechnology 31,233-239(2013),该文献以其全部内容作为参考并入本文),提供了深入探询Cas9对极大空间的靶标序列变化的亲合力的方法。因此,本发明公开的方面提供了对人细胞中Cas9靶向的直接高通量数据读取,同时避免了由dsDNA切割毒性所引起的问题复杂性以及使用天然核酸酶-活性的Cas9的特异性测试所引起的致突变修复。

[0104] 本发明公开的其它方面涉及一般用于靶标基因的转录调控的DNA结合蛋白或系统的应用。基于本发明公开,本领域技术人员将容易地鉴别示例性DNA结合系统。这些DNA结合系统不需要具有任何核酸酶活性,如天然存在的Cas9蛋白所具有的。因此,这些DNA结合系统不需要使核酸酶活性失活。一个示例性DNA结合系统是TALE。作为基因组编辑工具,通常

使用TALE-FokI二聚体,并且对于基因组调控,已显示TAEL-VP64融合体是高度有效的。根据一个方面,使用图2A所示的方法评价TALE特异性。设计了构建体文库,其中文库的每个元素包含驱动dTomato荧光蛋白的最小启动子。在转录起始位点m的下游插入24bp(A/C/G)随机转录本标签,同时在启动子上游布置两个TF结合位点:一个是所有文库元素共有的恒定DNA序列,第二个是具有“偏倚”结合位点文库的可变特征,所述结合位点是工程设计的以跨过提供了远离靶标序列的多种突变组合的大的序列集合,设计可编程DNA靶向复合物进行结合。这是通过使用工程设计的简并寡核苷酸实现的,所述简并寡核苷酸在每个位置具有的核苷酸频率使得靶标序列核苷酸以79%的频率出现,并且彼此核苷酸以7%的频率出现。参见Patwardhan等人,Nature Biotechnology 30,265-270(2012),该文献以其全部内容作为参考并入本文。然后,对报告子文库测序以显示24bp dTomato转录本标签和它们在文库元素中相应的“偏倚”靶标位点之间的结合。转录本标签(transcript tag)较大的多样性确保标签在不同靶标之间的共有将是极少见的,而靶标序列的偏倚构建是指具有较少突变的位点将比具有较多突变的位点结合更多标签。然后,通过工程设计结合共有DNA位点的控制-TF或工程设计结合靶标位点的靶标-TF刺激dTomato报告基因的转录。通过对刺激的细胞进行RNAseq,测量了每个样品中每个表达的转录本标签的丰度,然后使用先前建立的结合表,将其制图回到它们相应的结合位点。预期控制-TF将同等程度地激发所有文库成员,这是因为所有文库元素共有其结合位点,而预期靶标-TF将表达的成员的分布偏向于它优先靶向的那些。通过将对靶标-TF所获得的标签计数除以对控制-TF(对照-TF,control-TF)所获得的那些标签计数,在步骤5中使用了该假定以计算每个结合位点的归一化表达水平。

[0105] 如图2B所示,Cas9-gRNA复合物的靶向谱图显示平均来说,它对其靶标序列中的1-3个突变耐受。如图2C所示,除了定位至PAM序列的那些之外,Cas9-gRNA复合物在很大程度上还对点突变不敏感。值得注意地,该数据显示对化脓性链球菌(S.pyogenes)Cas9预测的PAM不只是NGG,还是NAG。如图2D所示,引入2个碱基错配显著损害了Cas9-gRNA复合物的活性,然而这仅当这些定位至gRNA靶标序列的3’端附近的8-10个碱基时(在热图中,从5’端开始,从1至23标记了靶标序列位置)。

[0106] 使用本文所述的转录特异性测定确定了另一种广泛使用的基因组编辑工具TALE结构域的突变耐受性。如图2E所示,18聚体TALE的TALE脱靶向数据(off-targeting data)显示平均来说,它可以耐受其靶标序列中1-2个突变,但是不能激活大多数其靶标中的3碱基错配变体。如图2F所示,与Cas9-gRNA复合物类似,18聚体TALE在很大程度上对其靶标中错配的单碱基不敏感。如图2G所示,2个碱基错配的引入显著损害了18聚体TALE的活性。对于更接近其靶标序列5’端的错配,TALE活性更敏感(在热图中,从5’端开始,从1至18标记了靶标序列位置)。

[0107] 使用核酸酶测定中的靶向实验确认了结果,所述核酸酶测定是针对于评价不同大小的TALE的靶向图谱的图10A-C的主体。如图10A所示,使用核酸酶介导的HR测定确认18聚体TALE耐受其靶标序列中的多个突变。如图10B所示,使用图2中所述的方法,分析了3种不同大小(18聚体、14聚体和10聚体)的TALE的靶标谱图。较短的TALE(14聚体和10聚体)在它们的靶向中逐渐更具有特异性,但是在活性方面也降低了几乎一个数量级。如图10C和10D所示,10聚体TALE显示了接近单碱基的错配分辨率,其对具有2个错配的靶标丧失了几乎全部活性(在热图中,从5’端开始,从1至10标记了靶标序列位置)。总的来说,这些数据表示工

程设计的较短的TALE可以在基因组工程应用中获得较高的特异性,而TALE核酸酶应用中对FokI二聚作用的要求是避免脱靶标效应所必需的。参见Kim人,Proceedings of the National Academy of Sciences of the United States of America 93,1156-1160(1996)和Pattanayak等人,Nature Methods 8,765-770(2011),以上每篇文献以其全部内容作为参考并入本文。

[0108] 图8A-C涉及用于根据实验数据计算通过实施例说明的归一化表达水平的高水平特异性分析处理流程。如图8A所示,通过结合位点序列的偏倚分布以及将引入报告基因转录本的随机序列24bp标签(顶部)产生了构建体文库。转录的标签是高度简并的,从而它们应该对Cas9或TALE结合序列绘制多对一图(many-to-one)。对构建体文库测序(第三水平,左图)以建立与结合位点共发生的标签,从而导致产生了结合位点相对于转录标签(第4水平,左图)的结合(association)表。可以立刻使用文库条形码(在本文中用浅蓝色和浅黄色表示;水平1-4,左图)对于不同结合位点构建的多个构建体文库测序。然后,将构建体文库转染到细胞群体中,并在群体样品中诱导一组不同的Cas9/gRNA或TALE转录因子(第2水平,右图)。始终用靶向构建体内固定结合位点序列的固定TALE激活剂诱导一个样品(最高水平,绿框);该样品用作阳性对照(绿色样品,还通过+符号表示)。然后,对从诱导样品中的报告子mRNA分子产生的cDNA测序并分析以获得样品中每个标签的标签计数(第3和4水平,右图)。和构建体文库测序一样,通过附加样品条形码对多个样品(包括阳性对照)测序并分析。这里,浅红色表示已测序并用阳性对照(绿色)分析的一个非对照样品。由于在每次读数中仅转录标签出现,而构建体结合位点不出现,因此使用得自构建体文库测序的结合位点相对于标签结合表计算每个样品中的每个结合位点表达的总标签计数(第5水平)。然后,通过将它们除以阳性对照样品中获得的标签(得分,tallies),将每个非阳性对照样品的计数(得分,tallies)转化为对每个结合位点归一化的表达水平。在图2B和2E以及在图9A和图10B中提供了通过错配数目的归一化表达水平图的实例。在该整个处理流程中,未覆盖错误标签、与构建体文库不可结合的标签以及与多个结合位点明显共有的标签的一些过滤水平。图8B显示了对于偏倚构建体文库内产生的错配数目的结合位点的百分比的示例分布。左图:理论分布。右图:从真实TALE构建体文库观察到的分布。图8C显示了对于错配数目的聚集至结合位点的标签计数的百分比的示例分布。左图:从阳性对照样品观察的分布。右图:从其中诱导了非控制TALE(非对照TALE)的样品观察的分布。由于阳性对照TALE结合至构建体中的固定位点,聚集的标签计数的分布密切反映了图8B中的结合位点分布,而由于具有较少错配的位点引起较高的表达水平,因此对于非对照TALE样品,分布偏向于左侧。下图:通过将对靶标-TF所获得的标签计数除以对控制-TF获得的那些,计算这些之间的相对富集度,其显示了相对于靶标位点中突变数目的平均表达水平。

[0109] 使用不同的Cas9-gRNA复合物产生的特异性数据再次证实了这些结果。图9A所示,不同的Cas9-gRNA复合物对其靶标序列中的1-3个突变耐受。如图9B所示,除了定位至PAM序列的那些之外,Cas9-gRNA复合物还在很大程度上对点突变不敏感。如图9C所示,然而2个碱基错配的引入显著损害了活性(在热图中,从5'端开始,从1至23标记了靶标序列位置)。如图9D所示,使用核酸酶介导的HR测定确认了对于化脓性链球菌(S.pyogenes)Cas9预测的PAM是NGG,还有NAG。

[0110] 根据某些方面,根据本文所述的方法提高结合特异性。由于多个复合物之间的协

同作用是Cas9N-VP64靶标基因激活中的因素,因此Cas9N的转录调控应用自然具有相当的特异性,这是因为单个脱靶标结合事件应具有极小影响。根据一个方面,在基因组-编辑方法中使用了偏移切口。大多数切口很少导致NHEJ事件(参见Certo等人,Nature Methods 8,671-676(2011),该文献以其全部内容作为参考并入本文),因此最大程度降低了脱靶标切口的影响。相反,对于引起基因破坏,引起偏移切口以产生双链断裂(DSB)是非常有效的。根据某些方面,与3'突出相比,5'突出产生了更显著的NHEJ事件。类似地,相比于NHEJ事件,3'突出有利于HR,尽管当产生5'突出时,HR事件的总数显著较低。因此,提供了方法用于使用切口进行同源重组和使用偏移切口以产生双链断裂从而最大程度降低脱靶标Cas9-gRNA活性的影响。

[0111] 图3A-C涉及多重偏移切口(偏移切割,off-set nicking)和使用向导RNA降低脱靶标结合(脱靶结合,off-target binding)的方法。如图3A所示,一旦引入靶向的切口或断裂,使用交通信号灯报告子(traffic light reporter)同时测定HR和NHEJ事件。通过HDR途径分辨的DNA切割事件使GFP序列恢复,而致突变NHEJ导致移码,从而使GFP处于框外,而使下游mCherry序列处于框内。对于该测定,设计了覆盖200bp DNA延伸(链段,stretch)的14个gRNA:7个靶向正义链(U1-7),7个靶向反义链(D1-7)。使用在互补链上产生缺口的Cas9D10A突变体,将gRNA不同的双向组合(two-way combination)用于引起一定范围的程序化5'或3'突出(指明了14个gRNA产生切口的位点)。如图3B所示,对于引起基因破坏,引起偏移切口以产生双链断裂(DSB)是非常有效的。值得注意地,与3'突出相反,导致5'突出的偏移切口导致产生了更多的NHEJ事件。如图3C所示,产生3'突出还有利于HR相比于NHEJ事件的比值,但是当产生5'突出时,HR事件的总数显著较低。

[0112] 图11A-B涉及Cas9D10A切口酶介导的NHEJ。如图11A所示,一旦引入靶向的切口或双链断裂,使用交通信号灯报告子(traffic light reporter)测定NHEJ事件。简要地,一旦引入DNA切割事件,如果断裂通过致突变NHEJ,则GFP被翻译在框外,而使得下游mCherry序列处于框内,从而导致产生红色荧光。设计了覆盖200bp DNA延伸的14个gRNA:7个靶向正义链(U1-7),7个靶向反义链(D1-7)。如图11B所示,观察到与在所有靶标中导致产生DSB和稳健NHEJ的野生型Cas9不同,大部分切口(使用Cas9D 10A突变体)很少导致NHEJ事件。所有14个位点位于邻接的200bp DNA延伸内,并且观察到靶向效力相差超过10倍。

[0113] 根据某些方面,本文描述了调控细胞中靶标核酸表达的方法,其包括将一个或多个、两个或更多个、或多个外源核酸引入所述细胞。引入所述细胞的外源核酸编码一种或多种向导RNA、一种或多种核酸酶无效的Cas9蛋白和转录调控因子蛋白或结构域。总地来说,将向导RNA、核酸酶无效的Cas9蛋白和转录调控因子蛋白或结构域称为共定位复合物,从而本领域技术人员将该术语理解为向导RNA、核酸酶无效的Cas9蛋白和转录调控因子蛋白或结构域结合至DNA并且调控靶标核酸表达的程度。根据某些其它方面,引入所述细胞的外源核酸编码一种或多种向导RNA和Cas9蛋白切口酶。总地来说,将向导RNA Cas9蛋白切口酶称为共定位复合物,从而本领域技术人员将该术语理解为向导RNA和Cas9蛋白切口酶结合至DNA并且使靶标核酸产生切口的程度。

[0114] 根据本发明公开的细胞包括其中如本文所述可以引入并表达外源核酸的任何细胞。应理解本文所述的本发明公开的基本概念不受细胞类型限制。根据本发明公开的细胞包括真核细胞、原核细胞、动物细胞、植物细胞、真菌细胞、古细菌细胞、真细菌细胞等。细胞

包括真核细胞,如酵母细胞、植物细胞和动物细胞。具体的细胞包括哺乳动物细胞。另外,细胞包括其中对于调控靶标核酸是有益的或期望的任何细胞。这些细胞可以包括特定蛋白表达不足,从而导致疾病或有害状况的那些。本领域技术人员容易地知道这些疾病或有害状况。根据本发明公开,可以通过本文所述的方法和转录活化因子靶向负责表达特定蛋白的核酸,从而导致靶标核酸上调和特定蛋白的相应表达。以这样的方式,本文所述的方法提供了治疗性治疗。

[0115] 靶标核酸包括如本文所述的共定位复合物可以用于调控或使其产生切口的任何核酸序列。靶标核酸包括基因。出于本发明公开的目的,DNA(如双链DNA)可以包括靶标核酸,并且共定位复合物可以结合至或另外与所述DNA一起共定位至所述靶标核酸处或与之相邻或在其附近,并且以其中共定位复合物可以对所述靶标核酸具有所需效果的方式进行。这些靶标核酸可以包括内源(或天然存在)的核酸和外源(或外来)核酸。基于本发明公开,技术人员将能够容易地鉴别或设计共定位至DNA(包括靶标核酸)的向导RNA和Cas9蛋白。技术人员将能够进一步鉴别同样共定位至DNA(包括靶标核酸)的转录调控因子蛋白或结构域。DNA包括基因组DNA、线粒体DNA、病毒DNA或外源DNA。

[0116] 可以使用本领域技术人员对这种引入已知的任何方法将外来核酸(即不是细胞天然核酸组成的一部分的那些)引入细胞。这些方法包括转染、转导、病毒转导、微注射、脂质转染、核转染、纳米颗粒轰击、转化、结合(conjugation)等。使用易于可辨别的文献来源,本领域技术人员将易于理解和改变这些方法。

[0117] 作为转录活化因子的转录调控因子蛋白或结构域包括基于本发明公开本领域技术人员易于可辨别的VP16和VP64等。

[0118] 疾病和有害状况是以特定蛋白表达的异常丧失为特征的那些。可以通过特定蛋白的上调治疗这些疾病或有害状况。因此,提供了治疗疾病或有害状况的方法,其中如本文所述的共定位复合物缔合或以其他方式结合至DNA(包括靶标核酸),而共定位复合物的转录活化因子上调靶标核酸的表达。例如,上调促进棕色脂肪分化和代谢吸收提高的PRDM16及其它基因可以用于治疗代谢综合征或者肥胖症。激活抗炎基因对于自身免疫和心血管疾病是有用的。激活肿瘤抑制基因对治疗癌症是有用的。基于本发明公开,本领域技术人员将容易地鉴别这些疾病和有害状况。

[0119] 作为本发明公开的代表描述了以下实施例。不应将这些实施例视为对本发明公开范围的限制,这是因为参考本发明公开、附图和所附权利要求,这些及其它等价实施方式将是明显的。

[0120] 实施例I

[0121] Cas9突变体

[0122] 调查了与Cas9同源的具有已知结构的序列以鉴别Cas9中可以除去其RuvC和HNH结构域天然活性的候选突变。使用HHpred(万维网toolkit.tuebingen.mpg.de/hhpred),对完整蛋白质数据库(full Protein Data Bank)(2013年1月)查询Cas9的全长序列。该查询返回了与Cas9的HNH结构域具有显著序列同源性的两个不同的HNH核酸内切酶:PacI和假定核酸内切酶(PDB ID分别为:3M7K和4H9D)。检查这些蛋白质以找到参与镁离子配位的残基。然后,在与Cas9的序列对比中鉴别了相应的残基。在每个结构中,鉴别了两个Mg-配位侧链,其与Cas9中相同的氨基酸类型比对。它们是3M7K D92和N113,和4H9D D53和N77。这些残基对

应于Cas9 D839和N863。还报道PacI残基D92和N113向丙氨酸的突变使得核酸酶催化能力不足。基于该分析,做出了Cas9突变D839A和N863A。另外,HHpred还预测了Cas9和嗜热栖热菌(Thermus thermophilus)RuvC(PDB ID:4EP4)的N-末端之间的同源性。该序列比对覆盖了先前报道的消除了Cas9中RuvC结构域功能的突变D10A。为了将其确认为适合的突变,如前所述确定了金属结合残基。在4EP4中,D7帮助配位镁离子。该位置具有对应于Cas9 D10的序列同源性,从而确认该突变有助于帮助消除金属结合,并因此从Cas9 RuvC结构域消除催化活性。

[0123] 实施例II

[0124] 质粒构建

[0125] 使用Quikchange试剂盒(Agilent technologies)产生Cas9突变体。靶标gRNA表达构建体(1)作为单个gBlock直接从IDT订购并克隆至pCR-Bluntll-TOPO载体(Invitrogen);或者(2)通过Genewiz常规(定制,custom)合成;或者(3)使用寡核苷酸的吉布森拼接(Gibson assembly)拼接至gRNA克隆载体(质粒#41824)。通过具有终止密码子和拼接到来自Addgene的EGIP慢病毒载体的适当片段的GFP序列的融合PCR拼装(PCR assembly)构建了用于涉及断裂的GFP的HR报告子测定的载体。然后,使用这些慢病毒载体建立GFP报告子稳定的系。使用标准规程构建了在本研究中使用的TALEN。参见Sanjana人,Nature Protocols 7,171-192(2012),该文献以其全部内容作为参考并入本文。使用标准PCR融合规程程序进行Cas9N和MS2 VP64融合。OCT4和REX1的启动子荧光素酶构建体得自Addgene(质粒#17221和质粒#17222)。

[0126] 实施例III

[0127] 细胞培养和转染

[0128] 在补充有10％胎牛血清(FBS,Invitrogen)、青霉素/链霉素(pen/strep,Invitrogen)和非必需氨基酸(NEAA,Invitrogen)的高葡萄糖杜尔贝科改良的伊格尔培养基(DMEM,Invitrogen)中培养HEK 293T细胞。将细胞在加湿的培育箱中维持在37℃和5％CO_2。

[0129] 涉及核酸酶测定的转染如下所示:根据生产商的规程,使用Lipofectamine 2000,用2μg Cas9质粒、2μg gRNA和/或2μg DNA供体质粒转染0.4×10^6个细胞。转染后3天收获细胞并通过FACS分析,或者对于基因组切割的直接测定,使用DNAeasy试剂盒(Qiagen)提取~1×10^6个细胞的基因组DNA。对于这些测定,实施PCR以通过来源于细胞的基因组DNA扩增靶标区域,并且通过覆盖度>200000个读取的MiSeq个人测序仪(Personal Sequencer)(Illumina)对扩增子深度测序。分析测序数据以估计NHEJ效率。

[0130] 对于涉及转录激活测定的转染:用(1)2μg Cas9N-VP64质粒、2μggRNA和/或0.25μg报告子构建体;或者(2)2μg Cas9N质粒、2μg MS2-VP64、2μg gRNA-2XMS2适体和/或0.25μg报告子构建体转染0.4×10^6个细胞。转染后24-48小时收获细胞,并使用FACS或免疫荧光法进行测定,或者提取它们的总RNA并随后通过RT-PCR分析这些提取物。在本文中,使用了来自Invitrogen的OCT4和REX1的标准taqman探针,并对每个样品,对GAPDH进行归一化。

[0131] 对于涉及Cas9-gRNA复合物和TALE的特异性谱图的转录激活测定的转染:用(1)2μg Cas9N-VP64质粒、2μg gRNA和0.25μg报告子文库;或者(2)2μg TALE-TF质粒和0.25μg报告子文库;或者(3)2μg对照-TF质粒和0.25μg报告子文库转染0.4×10^6个细胞。转染后24小

时收获细胞(以避免处于饱和形式的报告子的刺激)。使用RNAeasy-plus试剂盒(Qiagen)进行总RNA提取,并且使用Superscript-III(Invitrogen)进行标准RT-PCR。通过转录本-标签的靶向pcr扩增产生了用于下一代测序的文库。

[0132] 实施例IV

[0133] 用于Cas9-TF和TALE-TF报告子表达水平计算的计算和序列分析

[0134] 图8A中显示了该过程的高水平逻辑流程,并且本文提供了额外细节。对于构建体文库组成的细节,参见图8A(水平1)和8B。

[0135] 测序:对于Cas9实验,作为Illumina MiSeq上的150bp重叠的配对末端读取获得了构建体文库(图8A,水平3,左图)和报告基因cDNA序列(图8A,水平3,右图),而对于TALE实验,作为Illumina MiSeq上的51bp不重叠的配对末端读取获得了相应序列。

[0136] 构建体文库序列处理:比对:对于Cas9实验,使用novoalign V2.07.17(万维网novocraft.com/main/index/php)将配对读取与对应于被成对的8bp文库条形码侧接的234bp构建体的一组250bp参考序列进行比对(参见图8A,第3水平,左图)。在对novoalign提供的参考序列中,将23bp简并Cas9结合位点区域和24bp简并转录本标签区域(参见图8A,第1水平)称为Ns,同时明确提供了构建体文库条形码。对于TALE实验,除了参考序列长度为203bp并且简并结合位点区域的长度为18bp(相对于23bp)之外,使用了相同的程序。有效性检查:对于其中将每个读取对的左和右读取分别与参考序列比对的所包含文件的Novoalign输出。对唯一地与参考序列进行比对的仅有的读取对进行额外的有效性条件,并且保留了通过所有这些条件的仅有的读取对。有效性条件包括:(i)两个构建体文库条形码中的每一个必须在至少4个位置与参考序列条形码比对,并且对于相同的构建体文库,两个条形码必须与条形码对比对。(ii)与参考序列的N区域比对的所有碱基必须被novoalign称为As、Cs、Gs、Gs或Ts。注意:在参考N区域,Cas9和TALE实验均不左右读取重叠,因此不会出现这些N碱基的模糊novoalign调用的可能性。(iii)同样地,在这些区域中必须出现无novoalign调用的插入或缺失。(iv)在转录本标签区中必须出现无Ts(由于这些随机序列仅是从As、Cs和Gs产生的)。在读取对拒绝文件中收集违反这些条件中任一条的读取对。使用常规(定制,custom)perl脚本实现这些有效性检查。

[0137] 诱导的样品报告基因cDNA序列处理:比对:首先使用SeqPrep(从万维网github.com/jstjohn/SeqPrep下载)将重叠读取对合并成79bp的共同片段(common segment),然后使用novoalign(以上版本)将这些79bp的共同片段作为不配对的单个读取与一组参考序列比对(参见图8A,第3水平,右图),其中(对于构建体文库序列),将24bp简并转录本标签称为Ns,同时明确提供了样品条形码。TALE和Cas9 cDNA序列区两者均对应于被成对的8bp样品条形码序列侧接的相同的63bp cDNA区。有效性检查:除了以下情况外,应用了用于构建体文库测序的相同条件(参见上文):(a)在本发明中,由于读取对的在先SeqPrep合并,有效性处理不必须过滤读取对中的两种读取的唯一比对,而是仅过滤合并读取的唯一比对;(b)在cDNA序列读取中仅出现转录本标签,因此有效性处理仅适用于参考序列的这些标签区域,并且也不适用于单独的结合位点区域。

[0138] 结合位点相对于转录本标签结合的表的汇编:使用常规perl从验证的构建体文库序列中产生这些表(图8A,第4水平,左图)。尽管在整个构建体文库中由A、C和G碱基组成的24bp标签序列应基本上是唯一的(共有概率＝～2.8e-11),但是结合位点相对于标签结合

的早期分析显示事实上多个结合序列共有标签序列的不可忽略部分,这可能主要由结合序列中的序列错误或用于产生构建体文库的寡聚物中的寡聚合成错误的组合所造成的。除了标签共有之外,如果由于它们可能来自的构建体文库的条形码错配使它们不清楚,则在验证的读取对中发现与结合位点结合的标签还可以存在于构建体文库读取对拒绝文件中。最后,标签序列本身可能含有序列错误。为了处理这些误差来源,用三种属性对标签分类:(i)安全相对于不安全,其中不安全表示标签可以存在于构建体文库拒绝的读取对文件中;共有相对于非共有,其中共有表示发现标签与多个结合位点序列有关,和2+相对于仅1-,其中2+表示在验证的构建体文库序列中至少出现两次并因此认为不太可能含有序列错误的标签。将这三个标准结合获得了与每个结合位点有关的8类标签,最安全的(但最不丰富的)种类仅包含安全、非共有、2+标签;和最不安全的(但最丰富的)种类包含不考虑安全、共有或出现次数的所有标签。

[0139] 归一化表达水平的计算:使用常规perl代码实现图8A、水平5-6中表明的步骤。首先,使用先前对构建体文库计算的结合位点相对于转录本标签表,对每个结合位点计算(aggregate)对每个诱导的样品获得的标签计数的总和(参见图8C)。对于每个样品,然后将每个结合位点的标签计数总和除以阳性对照样品的标签计数总和以产生归一化表达水平。有关这些计算的其它考虑包括:

[0140] 1.对于每个样品,在不可能在结合位点相对于转录本标签结合表中存在的有效性-检查的cDNA基因序列中发现了“新型”标签的亚型(亚组,subset)。在后续计算中忽略这些标签。

[0141] 2.在结合位点相对于转录本标签结合表中,对如上所述的八类标签的每一类进行了如上所述的标签计数总和计算(aggregation)。由于构建体文库中结合位点偏倚而频繁产生了类似于中央序列的序列,但是具有错配数目增加的序列逐渐减少,因此具有较少错配的结合位点通常聚集成大量的标签,而具有更多错配的结合位点聚集成较少数目。因此,尽管通常期望使用最安全的标签种类,但是可以基于每个结合位点的标签的小数目来评价具有两个或更多个错配的结合位点,从而使得安全计数和比值在统计学上更不可靠,即使标签本身是更可靠的。在这些情况下,使用了所有标签。出于这种考虑,从对于n个错配位置,单独集合的标签计数的数目随着错配位置的组合数目增长的事实获得了一些补偿(等于$\binom{L}{n}3^n$),并且补偿随着n显著提高;因此,对于不同数目n的错配,集合的标签计数的平均值(如图2b、2e和图9A和10B所示)基于n≥2时,统计学上非常大的组的集合的标签计数。

[0142] 3.最后,构建到TALE构建体文库中的结合位点为18bp并且基于这些18bp序列分配标签结合,但是使用程序化以结合18bp构建体结合位点区域内中央14bp或10bp区域的TALE进行一些实验。在对于这些TALE计算表达水平时,基于结合表中18bp结合位点的相应区域,将标签集合至结合位点,从而忽略了该区域之外的结合位点错配。

[0143] 实施例V

[0144] 使用$Cas9_{N-}$VP64的RNA-导向的SOX2和NANOG调控

[0145] 本文所述的sgRNA(适体-修饰的单个导向RNA)链接方法使得不同的sgRNA招募不同的效应子结构域,只要每个sgRNA使用不同的RNA-蛋白质相互作用对,从而使得能够使用相同的Cas9N-蛋白质进行多重基因调控。对于图12A SOX2和图12B NANOG基因,设计了靶向转录起始位点上游～1kb的DNA延伸的10个gRNA。用绿色突出显示了DNA酶的超敏感位点。测

定了通过内源基因的qPCR的转录激活。在两种情况下,尽管单个gRNA的引入适度刺激了转录,但是多个gRNA协同作用以刺激稳健的多倍转录激活。数据为平均值+/-SEM(N＝3)。如图12A-B所示,通过启动子DNA上游～1kb延伸内的sgRNA靶向调控两个其它基因SOX2和NANOG。与转录起始位点邻近的sgRNA导致了稳健的基因激活。

[0146] 实施例VI

[0147] 评价Cas9-gRNA复合物的靶向图谱

[0148] 使用图2中所述的方法,分析了两个其它Cas9-gRNA复合物(图13A-C)和(图13D-F)的靶向图谱。两种gRNA具有非常不同的特异性谱,其中gRNA2耐受多至2-3个错配,gRNA3仅耐受最多1个。在一个碱基错配(图13B、13E)和两个碱基错配图(图13C、13F)中反映了这些方面。在图13C和13F中,用含有“x”的灰框表示可用数据不足以计算归一化表达水平的碱基错配对,为了改善数据显示,用含有星号“*”的黄框表示超出色标顶部的归一化表达水平为离群值的错配对。统计显著性符号为:对于p<.0005/n,***;对于P<.005/n,**;对于p<.05/n,*;对于P>＝.05/n,N.S.(非显著性),其中n是比较数目(参见表2)。

[0149] 实施例VII

[0150] 验证,报告子测定的特异性

[0151] 如图14A-C所示,使用两种不同的sgRNA:Cas9复合物产生特异性数据。确认对正在评价的sgRNA,测定是特异的,这是因为相应的突变体sgRNA不能刺激报告子文库。图14A:使用针对野生型gRNA靶标序列设计的报告子文库评价了两种gRNA(野生型和突变型;用红色突出显示序列差异)的特异性谱图。图14B:确认对正在评价的gRNA,该测定是特异性的(根据图13D重新作图的数据),这是因为相应的突变体gRNA不能刺激报告子文库。统计显著性符号为:对于p<.0005/n,***;对于P<.005/n,**;对于p<.05/n,*;对于P>＝.05/n,N.S.(非显著性),其中n是比较数目(参见表2)。不同的sgRNA可以具有不同的特异性谱图(图13A、13D),具体地,sgRNA2耐受多至3个错配,而sgRNA3仅耐受多至1个。对错配最大的灵敏度定位至间隔子(spacer)的3’端,尽管在其它位置处也观察到了错配影响活性。

[0152] 实施例VIII

[0153] 验证,单碱基和双碱基gRNA错配

[0154] 如图15A-D所示,通过测定的sgRNA中间隔子3’端12bp内单碱基错配导致可检测靶向的靶向实验进行确认。然而,该区域内2bp错配导致活力显著损失。使用核酸酶测定,测试了2个独立的gRNA:在相对于靶标的间隔序列中具有单碱基或双碱基错配(红色突出显示)的gRNA2(图15A-B)和gRNA3(图15C-D)。确认在测定的gRNA中间隔子3’端12bp内的单碱基错配导致了可检测的靶向,然而该区域内2bp的错配导致了活性快速损失。与图13的结果一致,这些结果还突出显示了不同gRNA之间特异性谱图的差异。数据是平均值+/-SEM(N＝3)。

[0155] 实施例IX

[0156] 验证,5’gRNA截短

[0157] 如图16A-D所示,间隔子5’部分中的截短导致保留了sgRNA的活性。使用核酸酶测定,测试了2个独立的gRNA:在它们间隔子的5’端具有截短的gRNA1(图16A-B)和gRNA3(图16C-D)。观察到对1-3bp 5’截短是良好耐受的,但是较大的缺失导致活性损失。数据为平均值+/-SEM(N＝3)。

[0158] 实施例X

[0159] 验证,化脓性链球菌(S.pyogenes)PAM

[0160] 如图17A-B所示,使用核酸酶介导的HR测定确认了化脓性链球菌(S.pyogenes)Cas9的PAM是NGG,还有NAG。数据为平均值+/-SEM(N＝3)。根据另外的研究,扫描了在不含共有靶标序列最后13nt的交替的NGG靶标的人外显子中生成的约190K Cas9靶标组,以确定是否存在在先前13nt中具有错配的交替的NAG位点或NGG位点。发现仅有0.4％没有这类交替靶标。

[0161] 实施例XI

[0162] 验证,TALE突变

[0163] 使用核酸酶介导的HR测定(图18A-B)确认18聚体TALE耐受其靶标序列中的多个突变。如图18A-B所示,靶标中间的某些突变导致了较高的TALE活性,如通过核酸酶测定中的靶向实验所确定的。

[0164] 实施例XII

[0165] TALE单体特异性相对于TALE蛋白特异性

[0166] 为了脱除单个重复序列可变的双氨基酸残基(重复可变双残基,repeat-variable diresidues)(RVD)的作用,确认了RVD的选择的确有助于碱基特异性,但是总体上TALE特异性还是蛋白结合能的函数。图19A-C显示了TALE单体特异性相对于TALE蛋白特异性的比较。图19A:使用图2中所述的修饰方法,分析了具有6个NI或6个NH重复的邻接的组的2个14聚体TALE-TF的靶标图谱。在该方法中,生成了在中间具有简并6聚体序列的缩减的报告子文库,并将其用于测定TALE-TF特异性。图19B-C:在两种情况下,注意到富集了预期的靶标序列(即,对于NI重复,具有6个As的序列,和对于NH重复,具有6个Gs的序列)。这些TALE中的每一个仍耐受中央6聚体靶标序列中的1-2个错配。尽管单体的选择的确有助于基底特异性(base specificity),但是总体上TALE特异性还是蛋白结合能的函数。根据一个方面,在基因组工程应用中较短的工程设计的TALE或者具有高和低亲合力单体组成的TALE导致了较高的特异性,并且当使用较短的TALE时,核酸酶应用中的FokI二聚使得能够进一步降低脱靶标作用。

[0167] 实施例XIII

[0168] 偏移切口,天然位点

[0169] 图20A-B显示了与偏移切口(off-set nicking)有关的数据。在基因组编辑的背景中,产生偏移切口以生成DSB。大多数切口不会导致非同源末端接合(NHEJ)介导的插入缺失(indel),并因此当引起偏移切口时,脱靶标单个切口事件将可能导致极低的插入缺失率。对于在两个合并的报告子位点和在天然AAVS1基因组位点引起基因破环来说,引起偏移切口以产生DSB是有效的。图20A:靶向具有覆盖200bp DNA延伸的8个gRNA的天然AAVS1位点:4个靶向正义链(s1-4),4个靶向反义链(as1-4)。使用在互补链上产生切口的Cas9D10A突变体,将gRNA不同的双向组合(two-way combination)用于引起一定范围的程序化5'或3'突出。图20B:使用基于Sanger测序的测定,观察到尽管单个gRNA不会引起可检测的NHEJ事件,但是对于引起基因破环来说,引起偏移切口以产生DSB是非常有效的。值得注意地,与3'突出相反,导致5'突出的偏移切口导致产生了更多的NHEJ事件。在条棒上方突出显示了Sanger测序克隆的数目,并且在相应X轴图例下方表明了预测的突出长度。

[0170] 实施例XIV

[0171] 偏移切口,NHEJ谱图

[0172] 图21A-C涉及偏移切口和NHEJ谱图。显示了三种不同的偏移切口组合的代表性Sanger测序结果,其中通过框突出显示了靶向gRNA的位置。此外,与同源重组(HR)介导的修复的标准模型一致,通过偏移切口的5'突出工程设计产生了比3'突出更稳健的NHEJ事件(图3B)。除了NHEJ刺激之外,当产生5'突出时,观察到了稳健的HR诱导。3'突出的产生不会导致HR率的改善(图3C)。

[0173] 实施例XV

[0174] 表1

[0175] 用于内源基因调控的gRNA靶标

[0176] 列出了Cas9-gRNA介导的激活实验中使用的REX1、OCT4、SOX2和NANOG启动子中的靶标。

[0177]

gRNA 名称	gRNA 靶标
REX1 1	ctggcggatcactcgcggtt agg
REX1 2	cctcggcctccaaaagtgct agg
REX1 3	acgctgattcctgcagatca ggg
REX1 4	ccaggaatacgtatccacca ggg
REX1 5	gccacacccaagcgatcaaa tgg
REX1 6	aaataatacattctaaggta agg
REX1 7	gctactggggaggctgaggc agg
REX1 8	tagcaatacagtcacattaa tgg
REX1 9	ctcatgtgatccccccgtct cgg
REX1 10	ccgggcagagagtgaacgcg cgg
OCT4 1	ttccttccctctcccgtgct tgg
OCT4 2	tctctgcaaagcccctggag agg
OCT4 3	aatgcagttgccgagtgcag tgg
OCT4 4	cctcagcctcctaaagtgct ggg
OCT4 5	gagtccaaatcctctttact agg
OCT4 6	gagtgtctggatttgggata agg
OCT4 7	cagcacctcatctcccagtg agg
OCT4 8	tctaaaacccagggaatcat ggg
OCT4 9	cacaaggcagccagggatcc agg
OCT4 10	gatggcaagctgagaaacac tgg
OCT4 11	tgaaatgcacgcatacaatt agg
OCT4 12	ccagtccagacctggcttc tgg
OCT4 13	cccagaaaaacagaccctga agg
OCT4 14	aagggttgagcacttgttta ggg
OCT4 15	atgtctgagttttggttgag agg
OCT4 16	ggtcccttgaaggggaagta ggg
OCT4 17	tggcagtctactcttgaaga tgg
OCT4 18	ggcacagtgccagaggtctg tgg
OCT4 19	taaaaataaaaaaactaaca ggg
OCT4 20	tctgtgggggacctgcactg agg
OCT4 21	ggccagaggtcaaggctagt ggg
SOX2 1	cacgaccgaaacccttctta cgg
SOX2 2	gttgaatgaagacagtctag tgg
SOX2 3	taagaacagagcaagttacg tgg
SOX2 4	tgtaaggtaagagaggagag cgg
SOX2 5	tgacacaccaactcctgcac tgg
SOX2 6	tttacccacttccttcgaaa agg
SOX2 7	gtggctggcaggctggctct ggg
SOX2 8	ctcccccggcctccccccgcg cgg
SOX2 9	caaaacccggcagcgaggct ggg
SOX2 10	aggagccgccgcgcgctgat tgg
NANOG 1	cacacacacccacacgagat ggg
NANOG 2	gaagaagctaaagagccaga ggg
NANOG 3	atgagaatttcaataacctc agg
NANOG 4	tcccgctctgttgcccaggc tgg
NANOG 5	cagacacccaccaccatgcg tgg
NANOG 6	tcccaatttactgggattac agg
NANOG 7	tgatttaaaagttggaaacg tgg
NANOG 8	tctagttccccacctagtct ggg
NANOG 9	gattaactgagaattcacaa ggg
NANOG 10	cgccaggagggtgggtcta agg

[0178] 实施例XVI

[0179] 表2

[0180] Cas9-gRNA和TALE特异性数据统计分析的总结

[0181] 表2(a)用于比较结合至具有特定数目的靶标位点突变的靶标序列的TALE或Cas9-VP64激活剂的归一化表达水平的P-值。通过“图”列中所指明的图中的箱形图显示了归一化的表达水平,其中箱(box)通过靶标位点的错配数目表示这些水平的分布。使用t检验,对每个箱形图中连续的错配数目对计算P-值,其中t检验是一个样品或两个样品t检验(参见方法)。使用Bonferroni校正的P-值阈值评价统计显著性,其中校正基于每个箱形图内比较的数目。统计显著性符号为:对于p<.0005/n,***;对于P<.005/n,**;对于p<.05/n,*;对于P>＝.05/n,N.S.(非显著性),其中n是比较数目。表2(b)图2D中种子区域(seed region)的统计特征:log10(P-值)表明了相对于所有其它位置对,对于在20bp靶标位点的3’端的候选种子区域内突变的那些位置对,结合至具有两种突变的靶标序列的Cas9N VP64+gRNA的表达值之间的分离度。在靶标位点的最后8-9bp中发现了最大的分离,其通过最大-log10(P-值)(上方突出显示)表示。这些位置可以理解为表示该靶标位点的“种子”区域的起始。有关如何计算P-值的信息,参见方法中“种子区域的统计特征”一节。

[0182]

a

图	表达水平比较：突变相对于突变		t 检验	p- 值	符号
2b	0	1	1-samp	7.8E-05	**
	1	2	2-samp	1.4E-06	***
	2	3	2-samp	4.0E-61	***
	3	4	2-samp	0	***
	4	5	2-samp	0	***
	5	6	2-samp	1.0E-217	***
	6	7	2-samp	1.7E-43	***
	7	8	2-samp	3.7E-02	N.S.
2e	0	1	1-samp	8.9E-01	N.S.
	1	2	2-samp	1.9E-86	***
	2	3	2-samp	5.0E-147	***
	3	4	2-samp	0	***
	4	5	2-samp	0	***
	5	6	2-samp	4.2E-62	***
	6	7	2-samp	1.6E-03	*
	7	8	2-samp	4.7E-01	N.S.
S7a	0	1	1-samp	5.2E-02	N.S.
	1	2	2-samp	2.8E-05	***
	2	3	2-samp	3.5E-21	***
	3	4	2-samp	1.4E-58	***
	4	5	2-samp	8.5E-101	***
	5	6	2-samp	6.8E-94	***
	6	7	2-samp	1.8E-61	***
	7	8	2-samp	8.1E-26	***
S7d 和 S8d	0	1	1-samp	2.3E-18	***
	1	2	2-samp	2.4E-08	***
	2	3	2-samp	6.2E-54	***
	3	4	2-samp	4.0E-141	***
	4	5	2-samp	1.9E-30	***
	5	6	2-samp	1.2E-03	*
	6	7	2-samp	3.8E-55	***
	7	8	2-samp	9.4E-01	N.S.
S8c	0	1	1-samp	7.2E-03	N.S.
	1	2	2-samp	5.0E-01	N.S.
	2	3	2-samp	3.9E-34	***
	3	4	2-samp	8.5E-153	***
	4	5	2-samp	8.6E-76	***
	5	6	2-samp	1.6E-03	*
	6	7	2-samp	7.1E-01	N.S.
	7	8	2-samp	7.8E-02	N.S.
S13a（左）	0	1	1-samp	7.3E-01	N.S.
	1	2	2-samp	2.4E-06	***
	2	3	2-samp	7.2E-140	***
	3	4	2-samp	0	***
	4	5	2-samp	0	***
	5	6	2-samp	1.0E-72	***
	6	7	2-samp	4.0E-03	*
S13a（中）	0	1	1-samp	9.4E-02	N.S.
	1	2	2-samp	5.2E-09	***
	2	3	2-samp	7.9E-86	***
	3	4	2-samp	2.9E-53	***
	4	5	2-samp	3.5E-10	***
S13a（右）	0	1	1-samp	1.3E-13	***
	1	2	2-samp	1.1E-04	***
	2	3	2-samp	3.7E-08	***

b

种子起始位置	位置对数目：两者均在种子中	位置对数目：两者并非均在种子中	-log10 p- 值
2	171	19	3.11
3	153	37	1.46
4	136	54	2.01
5	120	70	3.34
6	105	85	5.65
7	91	99	7.34
8	78	112	6.81
9	66	124	7.10
10	55	135	9.72
11	45	145	9.83
12	36	154	10.44
13	28	162	10.72
14	21	169	8.97
15	15	175	5.61
16	10	180	3.34
17	6	184	2.26
18	3	187	1.16

[0183] 实施例XVII

[0184] 实施例中的蛋白质和RNA的序列

[0185] A.以下显示了基于m4突变体的$Cas9_{N-}$VP64激活剂构建体的序列。以显示最高活性的$Cas9_{m4}^{VP64}$和$Cas9_{m4}^{VP64}$N融合蛋白形式构建了三种形式。还构建了m3和m2突变体的相应载体(图4A)(突出显示了NLS和VF64结构域)。

[0186] >$Cas9_{m4}^{VP64}$

gccaccATGGACAAGAAGTACTCCATTGGGCTCGCTATCGGCACAAACAGCGTCGGCTGG
GCCGTCATTACGGACGAGTACAAGGTGCCGAGCAAAAAATTCAAAGTTCTGGGCAAT
ACCGATCGCCACAGCATAAAGAAGAACCTCATTGGCGCCCTCCTGTTCGACTCCGGGG
AGACGGCCGAAGCCACGCGGCTCAAAAGAACAGCACGGCGCAGATATACCCGCAGAA
AGAATCGGATCTGCTACCTGCAGGAGATCTTTAGTAATGAGATGGCTAAGGTGGATGA
CTCTTTCTTCCATAGGCTGGAGGAGTCCTTTTTGGTGGAGGAGGATAAAAAGCACGAG
CGCCACCCAATCTTTGGCAATATCGTGGACGAGGTGGCGTACCATGAAAAGTACCCAA
CCATATATCATCTGAGGAAGAAGCTTGTAGACAGTACTGATAAGGCTGACTTGCGGTT
GATCTATCTCGCGCTGGCGCATATGATCAAATTTCGGGGACACTTCCTCATCGAGGGG
GACCTGAACCCAGACAACAGCGATGTCGACAAACTCTTTATCCAACTGGTTCAGACTT
ACAATCAGCTTTTCGAAGAGAACCCGATCAACGCATCCGGAGTTGACGCCAAAGCAA
TCCTGAGCGCTAGGCTGTCCAAATCCCGGCGGCTCGAAAACCTCATCGCACAGCTCCC
TGGGGAGAAGAAGAACGGCCTGTTTGGTAATCTTATCGCCCTGTCACTCGGGCTGACC
CCCAACTTTAAATCTAACTTCGACCTGGCCGAAGATGCCAAGCTTCAACTGAGCAAAG
ACACCTACGATGATGATCTCGACAATCTGCTGGCCCAGATCGGCGACCAGTACGCAGA
CCTTTTTTTGGCGGCAAAGAACCTGTCAGACGCCATTCTGCTGAGTGATATTCTGCGAG

[0187]

TGAACACGGAGATCACCAAAGCTCCGCTGAGCGCTAGTATGATCAAGCGCTATGATG
AGCACCACCAAGACTTGACTTTGCTGAAGGCCCTTGTCAGACAGCAACTGCCTGAGAA
GTACAAGGAAATTTTCTTCGATCAGTCTAAAAATGGCTACGCCGGATACATTGACGGC
GGAGCAAGCCAGGAGGAATTTTACAAATTTATTAAGCCCATCTTGGAAAAAATGGAC
GGCACCGAGGAGCTGCTGGTAAAGCTTAACAGAGAAGATCTGTTGCGCAAACAGCGC
ACTTTCGACAATGGAAGCATCCCCCACCAGATTCACCTGGGCGAACTGCACGCTATCC
TCAGGCGGCAAGAGGATTTCTACCCCTTTTTGAAAGATAACAGGGAAAAGATTGAGA
AAATCCTCACATTTCGGATACCCTACTATGTAGGCCCCCTCGCCCGGGGAAATTCCAG
ATTCGCGTGGATGACTCGCAAATCAGAAGAGACCATCACTCCCTGGAACTTCGAGGAA
GTCGTGGATAAGGGGGCCTCTGCCCAGTCCTTCATCGAAAGGATGACTAACTTTGATA
AAAATCTGCCTAACGAAAAGGTGCTTCCTAAACACTCTCTGCTGTACGAGTACTTCAC
AGTTTATAACGAGCTCACCAAGGTCAAATACGTCACAGAAGGGATGAGAAAGCCAGC
ATTCCTGTCTGGAGAGCAGAAGAAAGCTATCGTGGACCTCCTCTTCAAGACGAACCGG
AAAGTTACCGTGAAACAGCTCAAAGAAGACTATTTCAAAAAGATTGAATGTTTCGACT
CTGTTGAAATCAGCGGAGTGGAGGATCGCTTCAACGCATCCCTGGGAACGTATCACGA
TCTCCTGAAAATCATTAAAGACAAGGACTTCCTGGACAATGAGGAGAACGAGGACAT

[0188]

TCTTGAGGACATTGTCCTCACCCTTACGTTGTTTGAAGATAGGGAGATGATTGAAGAA
CGCTTGAAAACTTACGCTCATCTCTTCGACGACAAAGTCATGAAACAGCTCAAGAGGC
GCCGATATACAGGATGGGGGCGGCTGTCAAGAAAACTGATCAATGGGATCCGAGACA
AGCAGAGTGGAAAGACAATCCTGGATTTTCTTAAGTCCGATGGATTTGCCAACCGGAA
CTTCATGCAGTTGATCCATGATGACTCTCTCACCTTTAAGGAGGACATCCAGAAAGCA
CAAGTTTCTGGCCAGGGGGACAGTCTTCACGAGCACATCGCTAATCTTGCAGGTAGCC
CAGCTATCAAAAAGGGAATACTGCAGACCGTTAAGGTCGTGGATGAACTCGTCAAAG
TAATGGGAAGGCATAAGCCCGAGAATATCGTTATCGAGATGGCCCGAGAGAACCAAA
CTACCCAGAAGGGACAGAAGAACAGTAGGGAAAGGATGAAGAGGATTGAAGAGGGT
ATAAAAGAACTGGGGTCCCAAATCCTTAAGGAACACCCAGTTGAAAACACCCAGCTT
CAGAATGAGAAGCTCTACCTGTACTACCTGCAGAACGGCAGGGACATGTACGTGGAT
CAGGAACTGGACATCAATCGGCTCTCCGACTACGACGTGGCTGCTATCGTGCCCCAGT
CTTTTCTCAAAGATGATTCTATTGATAATAAAGTGTTGACAAGATCCGATAAAgcTAGA
GGGAAGAGTGATAACGTCCCCTCAGAAGAAGTTGTCAAGAAAATGAAAAATTATTGG
CGGCAGCTGCTGAACGCCAAACTGATCACACAACGGAAGTTCGATAATCTGACTAAG
GCTGAACGAGGTGGCCTGTCTGAGTTGGATAAAGCCGGCTTCATCAAAAGGCAGCTTG
TTGAGACACGCCAGATCACCAAGCACGTGGCCCAAATTCTCGATTCACGCATGAACAC
CAAGTACGATGAAAATGACAAACTGATTCGAGAGGTGAAAGTTATTACTCTGAAGTCT
AAGCTGGTCTCAGATTTCAGAAAGGACTTTCAGTTTTATAAGGTGAGAGAGATCAACA
ATTACCACCATGCGCATGATGCCTACCTGAATGCAGTGGTAGGCACTGCACTTATCAA
AAAATATCCCAAGCTTGAATCTGAATTTGTTTACGGAGACTATAAAGTGTACGATGTT
AGGAAAATGATCGCAAAGTCTGAGCAGGAAATAGGCAAGGCCACCGCTAAGTACTTC
TTTTACAGCAATATTATGAATTTTTTCAAGACCGAGATTACACTGGCCAATGGAGAGA
TTCGGAAGCGACCACTTATCGAAACAAACGGAGAAACAGGAGAAATCGTGTGGGACA
AGGGTAGGGATTTCGCGACAGTCCGGAAGGTCCTGTCCATGCCGCAGGTGAACATCGT
TAAAAAGACCGAAGTACAGACCGGAGGCTTCTCCAAGGAAAGTATCCTCCCGAAAAG
GAACAGCGACAAGCTGATCGCACGCAAAAAAGATTGGGACCCCAAGAAATACGGCGG
ATTCGATTCTCCTACAGTCGCTTACAGTGTACTGGTTGTGGCCAAAGTGGAGAAAGGG
AAGTCTAAAAAACTCAAAAGCGTCAAGGAACTGCTGGGCATCACAATCATGGAGCGA
TCAAGCTTCGAAAAAAACCCCATCGACTTTCTCGAGGCGAAAGGATATAAAGAGGTC
AAAAAAGACCTCATCATTAAGCTTCCCAAGTACTCTCTCTTTGAGCTTGAAAACGGCC
GGAAACGAATGCTCGCTAGTGCGGGCGAGCTGCAGAAAGGTAACGAGCTGGCACTGC
CCTCTAAATACGTTAATTTCTTGTATCTGGCCAGCCACTATGAAAAGCTCAAAGGGTCT
CCCGAAGATAATGAGCAGAAGCAGCTGTTCGTGGAACAACACAAACACTACCTTGAT
GAGATCATCGAGCAAATAAGCGAATTCTCCAAAAGAGTGATCCTCGCCGACGCTAAC
CTCGATAAGGTGCTTTCTGCTTACAATAAGCACAGGGATAAGCCCATCAGGGAGCAGG

[0189] CAGAAAACATTATCCACTTGTTTACTCTGACCAACTTGGGCGCGCCTGCAGCCTTCAA
GTACTTCGACACCACCATAGACAGAAAGCGGTACACCTCTACAAAGGAGGTCCTGGA
CGCCACACTGATTCATCAGTCAATTACGGGGCTCTATGAAACAAGAATCGACCTCTCT
CAGCTCGGTGGAGACAGCAGGGCTGACCCCAAGAAGAAGAGGAAGGTGGAGGCCAG
CGGTTCCGGACGGGCTGACGCATTGGACGATTTTGATCTGGATATGCTGGGAAGTGAC
GCCCTCGATGATTTTGACCTTGACATGCTTGGTTCGGATGCCCTTGATGACTTTGACCT
CGACATGCTCGGCAGTGACGCCCTTGATGATTTCGACCTGGACATGCTGATTAACTCT
AGATGA

[0190] >Cas9$_{m4}^{VP64}$N序列
gccaccATGCCCAAGAAGAAGAGGAAGGTGGGAAGGGGGATGGACAAGAAGTACTCCA
TTGGGCTCGCTATCGGCACAAACAGCGTCGGCTGGGCCGTCATTACGGACGAGTACAA
GGTGCCGAGCAAAAAATTCAAAGTTCTGGGCAATACCGATCGCCACAGCATAAAGAA
GAACCTCATTGGCGCCCTCCTGTTCGACTCCGGGGAGACGGCCGAAGCCACGCGGCTC
AAAAGAACAGCACGGCGCAGATATACCCGCAGAAAGAATCGGATCTGCTACCTGCAG
GAGATCTTTAGTAATGAGATGGCTAAGGTGGATGACTCTTTCTTCCATAGGCTGGAGG
AGTCCTTTTTGGTGGAGGAGGATAAAAAGCACGAGCGCCACCCAATCTTTGGCAATAT
CGTGGACGAGGTGGCGTACCATGAAAAGTACCCAACCATATATCATCTGAGGAAGAA
GCTTGTAGACAGTACTGATAAGGCTGACTTGCGGTTGATCTATCTCGCGCTGGCGCAT
ATGATCAAATTTCGGGGACACTTCCTCATCGAGGGGGACCTGAACCCAGACAACAGC
GATGTCGACAAACTCTTTATCCAACTGGTTCAGACTTACAATCAGCTTTTCGAAGAGA
ACCCGATCAACGCATCCGGAGTTGACGCCAAAGCAATCCTGAGCGCTAGGCTGTCCAA
ATCCCGGCGGCTCGAAAACCTCATCGCACAGCTCCCTGGGGAGAAGAAGAACGGCCT
GTTTGGTAATCTTATCGCCCTGTCACTCGGGCTGACCCCCAACTTTAAATCTAACTTCG
ACCTGGCCGAAGATGCCAAGCTTCAACTGAGCAAAGACACCTACGATGATGATCTCG
ACAATCTGCTGGCCCAGATCGGCGACCAGTACGCAGACCTTTTTTGGCGGCAAAGAA
CCTGTCAGACGCCATTCTGCTGAGTGATATTCTGCGAGTGAACACGGAGATCACCAAA
GCTCCGCTGAGCGCTAGTATGATCAAGCGCTATGATGAGCACCACCAAGACTTGACTT
TGCTGAAGGCCCTTGTCAGACAGCAACTGCCTGAGAAGTACAAGGAAATTTTCTTCGA
TCAGTCTAAAAATGGCTACGCCGGATACATTGACGGCGGAGCAAGCCAGGAGGAATT
TTACAAATTTATTAAGCCCATCTTGGAAAAAATGGACGGCACCGAGGAGCTGCTGGTA
AAGCTTAACAGAGAAGATCTGTTGCGCAAACAGCGCACTTTCGACAATGGAAGCATC
CCCCACCAGATTCACCTGGGCGAACTGCACGCTATCCTCAGGCGGCAAGAGGATTTCT
ACCCCTTTTTGAAAGATAACAGGGAAAAGATTGAGAAAATCCTCACATTTCGGATACC

[0191]

CTACTATGTAGGCCCCCTCGCCCGGGGAAATTCCAGATTCGCGTGGATGACTCGCAAA
TCAGAAGAGACCATCACTCCCTGGAACTTCGAGGAAGTCGTGGATAAGGGGGCCTCT
GCCCAGTCCTTCATCGAAAGGATGACTAACTTTGATAAAAATCTGCCTAACGAAAAGG
TGCTTCCTAAACACTCTCTGCTGTACGAGTACTTCACAGTTTATAACGAGCTCACCAAG
GTCAAATACGTCACAGAAGGGATGAGAAAGCCAGCATTCCTGTCTGGAGAGCAGAAG
AAAGCTATCGTGGACCTCCTCTTCAAGACGAACCGGAAAGTTACCGTGAAACAGCTCA
AAGAAGACTATTTCAAAAAGATTGAATGTTTCGACTCTGTTGAAATCAGCGGAGTGGA
GGATCGCTTCAACGCATCCCTGGGAACGTATCACGATCTCCTGAAAATCATTAAAGAC
AAGGACTTCCTGGACAATGAGGAGAACGAGGACATTCTTGAGGACATTGTCCTCACCC
TTACGTTGTTTGAAGATAGGGAGATGATTGAAGAACGCTTGAAAACTTACGCTCATCT
CTTCGACGACAAAGTCATGAAACAGCTCAAGAGGCGCCGATATACAGGATGGGGGCG
GCTGTCAAGAAAACTGATCAATGGGATCCGAGACAAGCAGAGTGGAAAGACAATCCT
GGATTTTCTTAAGTCCGATGGATTTGCCAACCGGAACTTCATGCAGTTGATCCATGATG
ACTCTCTCACCTTTAAGGAGGACATCCAGAAAGCACAAGTTTCTGGCCAGGGGGACAG
TCTTCACGAGCACATCGCTAATCTTGCAGGTAGCCCAGCTATCAAAAAGGGAATACTG
CAGACCGTTAAGGTCGTGGATGAACTCGTCAAAGTAATGGGAAGGCATAAGCCCGAG
AATATCGTTATCGAGATGGCCCGAGAGAACCAAACTACCCAGAAGGGACAGAAGAAC
AGTAGGGAAAGGATGAAGAGGATTGAAGAGGGTATAAAAGAACTGGGGTCCCAAAT
CCTTAAGGAACACCCAGTTGAAAACACCCAGCTTCAGAATGAGAAGCTCTACCTGTAC
TACCTGCAGAACGGCAGGGACATGTACGTGGATCAGGAACTGGACATCAATCGGCTC
TCCGACTACGACGTGGCTGCTATCGTGCCCCAGTCTTTTCTCAAAGATGATTCTATTGA
TAATAAAGTGTTGACAAGATCCGATAAAgcTAGAGGGAAGAGTGATAACGTCCCCTCA
GAAGAAGTTGTCAAGAAAATGAAAAATTATTGGCGGCAGCTGCTGAACGCCAAACTG
ATCACACAACGGAAGTTCGATAATCTGACTAAGGCTGAACGAGGTGGCCTGTCTGAGT
TGGATAAAGCCGGCTTCATCAAAAGGCAGCTTGTTGAGACACGCCAGATCACCAAGC
ACGTGGCCCAAATTCTCGATTCACGCATGAACACCAAGTACGATGAAAATGACAAACT
GATTCGAGAGGTGAAAGTTATTACTCTGAAGTCTAAGCTGGTCTCAGATTTCAGAAAG
GACTTTCAGTTTTATAAGGTGAGAGAGATCAACAATTACCACCATGCGCATGATGCCT
ACCTGAATGCAGTGGTAGGCACTGCACTTATCAAAAAATATCCCAAGCTTGAATCTGA
ATTTGTTTACGGAGACTATAAAGTGTACGATGTTAGGAAAATGATCGCAAAGTCTGAG
CAGGAAATAGGCAAGGCCACCGCTAAGTACTTCTTTTACAGCAATATTATGAATTTTT
TCAAGACCGAGATTACACTGGCCAATGGAGAGATTCGGAAGCGACCACTTATCGAAA
CAAACGGAGAAACAGGAGAAATCGTGTGGGACAAGGGTAGGGATTTCGCGACAGTCC
GGAAGGTCCTGTCCATGCCGCAGGTGAACATCGTTAAAAAGACCGAAGTACAGACCG
GAGGCTTCTCCAAGGAAAGTATCCTCCCGAAAAGGAACAGCGACAAGCTGATCGCAC
GCAAAAAAGATTGGGACCCCAAGAAATACGGCGGATTCGATTCTCCTACAGTCGCTTA

[0192] CAGTGTACTGGTTGTGGCCAAAGTGGAGAAAGGGAAGTCTAAAAAACTCAAAAGCGT
CAAGGAACTGCTGGGCATCACAATCATGGAGCGATCAAGCTTCGAAAAAAACCCCAT
CGACTTTCTCGAGGCGAAAGGATATAAAGAGGTCAAAAAAGACCTCATCATTAAGCTT
CCCAAGTACTCTCTCTTTGAGCTTGAAAACGGCCGGAAACGAATGCTCGCTAGTGCGG
GCGAGCTGCAGAAAGGTAACGAGCTGGCACTGCCCTCTAAATACGTTAATTTCTTGTA
TCTGGCCAGCCACTATGAAAAGCTCAAAGGGTCTCCCGAAGATAATGAGCAGAAGCA
GCTGTTCGTGGAACAACACAAACACTACCTTGATGAGATCATCGAGCAAATAAGCGA
ATTCTCCAAAAGAGTGATCCTCGCCGACGCTAACCTCGATAAGGTGCTTTCTGCTTAC
AATAAGCACAGGGATAAGCCCATCAGGGAGCAGGCAGAAAACATTATCCACTTGTTT
ACTCTGACCAACTTGGGCGCGCCTGCAGCCTTCAAGTACTTCGACACCACCATAGACA
GAAAGCGGTACACCTCTACAAAGGAGGTCCTGGACGCCACACTGATTCATCAGTCAAT
TACGGGGCTCTATGAAACAAGAATCGACCTCTCTCAGCTCGGTGGAGACAGCAGGGCT
GACCCCAAGAAGAAGAGGAAGGTGGAGGCCAGCGGTTCCGGACGGGCTGACGCATTG
GACGATTTTGATCTGGATATGCTGGGAAGTGACGCCCTCGATGATTTTGACCTTGACA
TGCTTGGTTCGGATGCCCTTGATGACTTTGACCTCGACATGCTCGGCAGTGACGCCCTT
GATGATTTCGACCTGGACATGCTGATTAACTCTAGATGA

[0193] >Cas9$_{m4}^{VP64}$C

[0194] gccaccATGGACAAGAAGTACTCCATTGGGCTCGCTATCGGCACAAACAGCGTCGGCTGG
GCCGTCATTACGGACGAGTACAAGGTGCCGAGCAAAAAATTCAAAGTTCTGGGCAAT
ACCGATCGCCACAGCATAAAGAAGAACCTCATTGGCGCCCTCCTGTTCGACTCCGGGG
AGACGGCCGAAGCCACGCGGCTCAAAAGAACAGCACGGCGCAGATATACCCGCAGAA
AGAATCGGATCTGCTACCTGCAGGAGATCTTTAGTAATGAGATGGCTAAGGTGGATGA
CTCTTTCTTCCATAGGCTGGAGGAGTCCTTTTTGGTGGAGGAGGATAAAAAGCACGAG
CGCCACCCAATCTTTGGCAATATCGTGGACGAGGTGGCGTACCATGAAAAGTACCCAA
CCATATATCATCTGAGGAAGAAGCTTGTAGACAGTACTGATAAGGCTGACTTGCGGTT
GATCTATCTCGCGCTGGCGCATATGATCAAATTTCGGGGACACTTCCTCATCGAGGGG
GACCTGAACCCAGACAACAGCGATGTCGACAAACTCTTTATCCAACTGGTTCAGACTT
ACAATCAGCTTTTCGAAGAGAACCCGATCAACGCATCCGGAGTTGACGCCAAAGCAA
TCCTGAGCGCTAGGCTGTCCAAATCCCGGCGGCTCGAAAACCTCATCGCACAGCTCCC
TGGGGAGAAGAAGAACGGCCTGTTTGGTAATCTTATCGCCCTGTCACTCGGGCTGACC
CCCAACTTTAAATCTAACTTCGACCTGGCCGAAGATGCCAAGCTTCAACTGAGCAAAG
ACACCTACGATGATGATCTCGACAATCTGCTGGCCCAGATCGGCGACCAGTACGCAGA
CCTTTTTTTGGCGGCAAAGAACCTGTCAGACGCCATTCTGCTGAGTGATATTCTGCGAG

[0195]

TGAACACGGAGATCACCAAAGCTCCGCTGAGCGCTAGTATGATCAAGCGCTATGATG
AGCACCACCAAGACTTGACTTTGCTGAAGGCCCTTGTCAGACAGCAACTGCCTGAGAA
GTACAAGGAAATTTTCTTCGATCAGTCTAAAAATGGCTACGCCGGATACATTGACGGC
GGAGCAAGCCAGGAGGAATTTTACAAATTTATTAAGCCCATCTTGGAAAAAATGGAC
GGCACCGAGGAGCTGCTGGTAAAGCTTAACAGAGAAGATCTGTTGCGCAAACAGCGC
ACTTTCGACAATGGAAGCATCCCCCACCAGATTCACCTGGGCGAACTGCACGCTATCC
TCAGGCGGCAAGAGGATTTCTACCCCTTTTTGAAAGATAACAGGGAAAAGATTGAGA
AAATCCTCACATTTCGGATACCCTACTATGTAGGCCCCCTCGCCCGGGGAAATTCCAG
ATTCGCGTGGATGACTCGCAAATCAGAAGAGACCATCACTCCCTGGAACTTCGAGGAA
GTCGTGGATAAGGGGGCCTCTGCCCAGTCCTTCATCGAAAGGATGACTAACTTTGATA
AAAATCTGCCTAACGAAAAGGTGCTTCCTAAACACTCTCTGCTGTACGAGTACTTCAC
AGTTTATAACGAGCTCACCAAGGTCAAATACGTCACAGAAGGGATGAGAAAGCCAGC
ATTCCTGTCTGGAGAGCAGAAGAAAGCTATCGTGGACCTCCTCTTCAAGACGAACCGG
AAAGTTACCGTGAAACAGCTCAAAGAAGACTATTTCAAAAAGATTGAATGTTTCGACT
CTGTTGAAATCAGCGGAGTGGAGGATCGCTTCAACGCATCCCTGGGAACGTATCACGA
TCTCCTGAAAATCATTAAAGACAAGGACTTCCTGGACAATGAGGAGAACGAGGACAT
TCTTGAGGACATTGTCCTCACCCTTACGTTGTTTGAAGATAGGGAGATGATTGAAGAA
CGCTTGAAAACTTACGCTCATCTCTTCGACGACAAAGTCATGAAACAGCTCAAGAGGC
GCCGATATACAGGATGGGGGCGGCTGTCAAGAAAACTGATCAATGGGATCCGAGACA
AGCAGAGTGGAAAGACAATCCTGGATTTTCTTAAGTCCGATGGATTTGCCAACCGGAA
CTTCATGCAGTTGATCCATGATGACTCTCTCACCTTTAAGGAGGACATCCAGAAAGCA
CAAGTTTCTGGCCAGGGGGACAGTCTTCACGAGCACATCGCTAATCTTGCAGGTAGCC
CAGCTATCAAAAAGGGAATACTGCAGACCGTTAAGGTCGTGGATGAACTCGTCAAAG
TAATGGGAAGGCATAAGCCCGAGAATATCGTTATCGAGATGGCCCGAGAGAACCAAA
CTACCCAGAAGGGACAGAAGAACAGTAGGGAAAGGATGAAGAGGATTGAAGAGGGT
ATAAAAGAACTGGGGTCCCAAATCCTTAAGGAACACCCAGTTGAAAACACCCAGCTT
CAGAATGAGAAGCTCTACCTGTACTACCTGCAGAACGGCAGGGACATGTACGTGGAT
CAGGAACTGGACATCAATCGGCTCTCCGACTACGACGTGGCTGCTATCGTGCCCCAGT
CTTTTCTCAAAGATGATTCTATTGATAATAAAGTGTTGACAAGATCCGATAAAgcTAGA
GGGAAGAGTGATAACGTCCCCTCAGAAGAAGTTGTCAAGAAAATGAAAAATTATTGG
CGGCAGCTGCTGAACGCCAAACTGATCACACAACGGAAGTTCGATAATCTGACTAAG
GCTGAACGAGGTGGCCTGTCTGAGTTGGATAAAGCCGGCTTCATCAAAAGGCAGCTTG
TTGAGACACGCCAGATCACCAAGCACGTGGCCCAAATTCTCGATTCACGCATGAACAC
CAAGTACGATGAAAATGACAAACTGATTCGAGAGGTGAAAGTTATTACTCTGAAGTCT
AAGCTGGTCTCAGATTTCAGAAAGGACTTTCAGTTTTATAAGGTGAGAGAGATCAACA
ATTACCACCATGCGCATGATGCCTACCTGAATGCAGTGGTAGGCACTGCACTTATCAA

AAAATATCCCAAGCTTGAATCTGAATTTGTTTACGGAGACTATAAAGTGTACGATGTT
AGGAAAATGATCGCAAAGTCTGAGCAGGAAATAGGCAAGGCCACCGCTAAGTACTTC
TTTTACAGCAATATTATGAATTTTTTCAAGACCGAGATTACACTGGCCAATGGAGAGA
TTCGGAAGCGACCACTTATCGAAACAAACGGAGAAACAGGAGAAATCGTGTGGGACA
AGGGTAGGGATTTCGCGACAGTCCGGAAGGTCCTGTCCATGCCGCAGGTGAACATCGT
TAAAAAGACCGAAGTACAGACCGGAGGCTTCTCCAAGGAAAGTATCCTCCCGAAAAG
GAACAGCGACAAGCTGATCGCACGCAAAAAAGATTGGGACCCCAAGAAATACGGCGG
ATTCGATTCTCCTACAGTCGCTTACAGTGTACTGGTTGTGGCCAAAGTGGAGAAAGGG
AAGTCTAAAAAACTCAAAAGCGTCAAGGAACTGCTGGGCATCACAATCATGGAGCGA
TCAAGCTTCGAAAAAAACCCCATCGACTTTCTCGAGGCGAAAGGATATAAAGAGGTC
AAAAAAGACCTCATCATTAAGCTTCCCAAGTACTCTCTCTTTGAGCTTGAAAACGGCC
GGAAACGAATGCTCGCTAGTGCGGGCGAGCTGCAGAAAGGTAACGAGCTGGCACTGC
[0196] CCTCTAAATACGTTAATTTCTTGTATCTGGCCAGCCACTATGAAAAGCTCAAAGGGTCT
CCCGAAGATAATGAGCAGAAGCAGCTGTTCGTGGAACAACACAAACACTACCTTGAT
GAGATCATCGAGCAAATAAGCGAATTCTCCAAAAGAGTGATCCTCGCCGACGCTAAC
CTCGATAAGGTGCTTTCTGCTTACAATAAGCACAGGGATAAGCCCATCAGGGAGCAGG
CAGAAAACATTATCCACTTGTTTACTCTGACCAACTTGGGCGCGCCTGCAGCCTTCAA
GTACTTCGACACCACCATAGACAGAAAGCGGTACACCTCTACAAAGGAGGTCCTGGA
CGCCACACTGATTCATCAGTCAATTACGGGGCTCTATGAAACAAGAATCGACCTCTCT
CAGCTCGGTGGAGACAGCAGGGCTGACCCCAAGAAGAAGAGGAAGGTGGAGGCCAG
CGGTTCCGGACGGGCTGACGCATTGGACGATTTTGATCTGGATATGCTGGGAAGTGAC
GCCCTCGATGATTTTGACCTTGACATGCTTGGTTCGGATGCCCTTGATGACTTTGACCT
CGACATGCTCGGCAGTGACGCCCTTGATGATTTCGACCTGGACATGCTGATTAACTCT
AGAGCGGCCGCAGATCCAAAAAGAAGAGAAAGGTAGATCCAAAAAGAAGAGAAA
GGTAGATCCAAAAAGAAGAGAAAGGTAGATACGGCCGCATAG

[0197] B.以下提供了具有2X MS2适体结构域的MS2-激活剂构建体以及相应gRNA骨架载体(backbone vector)的序列(突出显示了NLS、VP64、gRNA间隔子和MS2-结合RNA茎环结构域)。以显示最高活性的MS2vp64N融合蛋白形式构建了前者的两种形式。

[0198] >$MS2_{VP64}N$

[0199] gccaccATGGGACCTAAGAAAAAGAGGAAGGTGGCGGCCGCTTCTAGAATGGCTTCTAACTTTACTCAGTTCGTTCTCGTCGACAATGGCGGAACTGGCGACGTGACTGTCGCCCCAAGCAACTTCGCTAACGGGATCGCTGAATGGATCAGCTCTAACTCGCGTTCACAGGCTTACAAAGTAACCTGTAGCGTTCGTCAGAGCTCTGCGCAGAATCGCAAATACACCATCAAAGTCGAGGTGCCTAAAGGCGCCTGGCGTTCGTACTTAAATATGGAACTAACCATTCCAATTTTCGCCACGAATTCCGACTGCGAGCTTATTGTTAAGGCAATGCAAGGTCTCCTAAAAGATGGAAACCCGATTCCCTCAGCAATCGCAGCAAACTCCGGCATCTACGAGGCCAGCGGTTCCGGACGGGCTGACGCATTGGACGATTTTGATCTGGATATGCTGGGAAGTGACGCCCTCGATGATTTTGACCTTGACATGCTTGGTTCGGATGCCCTTGATGACTTTGACCTCGACATGCTCGGCAGTGACGCCCTTGATGATTTCGACCTGGACATGCTGATTAACTCTAGATGA

[0200] >MS2$_{VP64}$C

[0201] gccaccATGGGACCTAAGAAAAAGAGGAAGGTGGCGGCCGCTTCTAGAATGGCTTCTAACTTTACTCAGTTCGTTCTCGTCGACAATGGCGGAACTGGCGACGTGACTGTCGCCCCAAGCAACTTCGCTAACGGGATCGCTGAATGGATCAGCTCTAACTCGCGTTCACAGGCTTACAAAGTAACCTGTAGCGTTCGTCAGAGCTCTGCGCAGAATCGCAAATACACCATCAAAGTCGAGGTGCCTAAAGGCGCCTGGCGTTCGTACTTAAATATGGAACTAACCATTCCAATTTTCGCCACGAATTCCGACTGCGAGCTTATTGTTAAGGCAATGCAAGGTCTCCTAAAAGATGGAAACCCGATTCCCTCAGCAATCGCAGCAAACTCCGGCATCTACGAGGCCAGCGGTTCCGGACGGGCTGACGCATTGGACGATTTTGATCTGGATATGCTGGGAAGTGACGCCCTCGATGATTTTGACCTTGACATGCTTGGTTCGGATGCCCTTGATGACTTTGACCTCGACATGCTCGGCAGTGACGCCCTTGATGATTTCGACCTGGACATGCTGATTAACTCTAGAGCGGCCGCAGATCCAAAAAAGAAGAGAAAGGTAGATCCAAAAAAGAAGAGAAAGGTAGATCCAAAAAAGAAGAGAAAGGTAGATACGGCCGCATAG

[0202] >gRNA$_{2XMS2}$

[0203] TGTACAAAAAAGCAGGCTTTAAAGGAACCAATTCAGTCGACTGGATCCGGTACCAAGGTCGGGCAGGAAGAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGGCTGTTAGAGAGATAATTAGAATTAATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTGACGTAGAAAGTAATAATTTCTTGGGTAGTTTGCAGTTTTAAAATTATGTTTTAAAATGGACTATCATATGCTTACCGTAACTTGAAAGTATTTCGATTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACCGNNNNNNNNNNNNNNNNNNNNGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTCTGCAGGTCGACTCTAGAAAACATGAGGATCACCCATGTCTGCAGTATTCCCGGGTTCATTAGATCCTAAGGTACCTAATTGCCTAGAAAACATGAGGATCACCCATGTCTGCAGGTCGACTCTAGAAATTTTTTCTAGAC

[0204] C.以下列出了dTomato荧光基转录激活报告子序列(突出显示了IScel控制-TF靶标、gRNA靶标、minCMV启动子和FLAG标签+dTomato序列)。

[0205] >TF报告子1

[0206] TAGGGATAACAGGGTAATAGTGTCCCCTCCACCCCACAGTGGGGCGAGGTAGGCGTG
TACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGCCTGGAG
AATTCgccaccatgGACTACAAGGATGACGACGATAAAACTTCCGGTGGCGGACTGGGTTC
CACCGTGAGCAAGGGCGAGGAGGTCATCAAAGAGTTCATGCGCTTCAAGGTGCGCAT
GGAGGGCTCCATGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCC
CTACGAGGGCACCCAGACCGCCAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTT
CGCCTGGGACATCCTGTCCCCCCAGTTCATGTACGGCTCCAAGGCGTACGTGAAGCAC
CCCGCCGACATCCCCGATTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGC
GCGTGATGAACTTCGAGGACGGCGGTCTGGTGACCGTGACCCAGGACTCCTCCCTGCA
GGACGGCACGCTGATCTACAAGGTGAAGATGCGCGGCACCAACTTCCCCCCCGACGG
CCCCGTAATGCAGAAGAAGACCATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCC
CCGCGACGGCGTGCTGAAGGGCGAGATCCACCAGGCCCTGAAGCTGAAGGACGGCGG
CCACTACCTGGTGGAGTTCAAGACCATCTACATGGCCAAGAAGCCCGTGCAACTGCCC
GGCTACTACTACGTGGACACCAAGCTGGACATCACCTCCCACAACGAGGACTACACCA
TCGTGGAACAGTACGAGCGCTCCGAGGGCCGCCACCACCTGTTCCTGTACGGCATGGA
CGAGCTGTACAAGTAA

[0207] >TF报告子2

[0208] TAGGGATAACAGGGTAATAGTGGGGCCACTAGGGACAGGATTGGCGAGGTAGGCGTG
TACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGCCTGGAG
AATTCgccaccatgGACTACAAGGATGACGACGATAAAACTTCCGGTGGCGGACTGGGTTC
CACCGTGAGCAAGGGCGAGGAGGTCATCAAAGAGTTCATGCGCTTCAAGGTGCGCAT
GGAGGGCTCCATGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCC
CTACGAGGGCACCCAGACCGCCAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTT
CGCCTGGGACATCCTGTCCCCCCAGTTCATGTACGGCTCCAAGGCGTACGTGAAGCAC
CCCGCCGACATCCCCGATTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGC
GCGTGATGAACTTCGAGGACGGCGGTCTGGTGACCGTGACCCAGGACTCCTCCCTGCA
GGACGGCACGCTGATCTACAAGGTGAAGATGCGCGGCACCAACTTCCCCCCCGACGG
CCCCGTAATGCAGAAGAAGACCATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCC
CCGCGACGGCGTGCTGAAGGGCGAGATCCACCAGGCCCTGAAGCTGAAGGACGGCGG
CCACTACCTGGTGGAGTTCAAGACCATCTACATGGCCAAGAAGCCCGTGCAACTGCCC
GGCTACTACTACGTGGACACCAAGCTGGACATCACCTCCCACAACGAGGACTACACCA
TCGTGGAACAGTACGAGCGCTCCGAGGGCCGCCACCACCTGTTCCTGTACGGCATGGA
CGAGCTGTACAAGTAA

[0209] D.以下提供了用于TALE和Cas9-gRNA特异性测定的报告子文库的一般形式(突出显示了ISce1控制-TF靶标、gRNA/TALE靶标位点(对于gRNA,23bp,对于TALE,18bp)、minCMV启动子、RNA条形码和dTomato序列)。

[0210] >特异性报告子文库

[0211] TAGGGATAACAGGGTAATAGTNNNNNNNNNNNNNNNNNNNNNNNCGAGGTAGGCGT
GTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGCCTGGA
GAATTCgccaccatgGACTACAAGGATGACGACGATAAANNNNNNNNNNNNNNNNNNNN
NNNNACTTCCGGTGGCGGACTGGGTTCCACCGTGAGCAAGGGCGAGGAGGTCATCAA
AGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGCTCCATGAACGGCCACGAGTTCGA
GATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCACCCAGACCGCCAAGCTGAA
GGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCCCAGTTCATG
TACGGCTCCAAGGCGTACGTGAAGCACCCCGCCGACATCCCCGATTACAAGAAGCTGT
CCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGCGGTCTGGT
GACCGTGACCCAGGACTCCTCCCTGCAGGACGGCACGCTGATCTACAAGGTGAAGAT
GCGCGGCACCAACTTCCCCCCCGACGGCCCCGTAATGCAGAAGAAGACCATGGGCTG
GGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAGATCCA
CCAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGACCATCTA
CATGGCCAAGAAGCCCGTGCAACTGCCCGGCTACTACTACGTGGACACCAAGCTGGA
CATCACCTCCCACAACGAGGACTACACCATCGTGGAACAGTACGAGCGCTCCGAGGG
CCGCCACCACCTGTTCCTGTACGGCATGGACGAGCTGTACAAGTAAGAATTC。

[0001] 序列表
[0002] <110> 哈佛大学校长及研究员协会(President and Fellows of Harvard College)
[0003] <120> RNA-导向的转录调控
[0004] <130> 010498.00503
[0005] <140> PCT/US14/040868
[0006] <141> 2014-06-04
[0007] <150> US 61/830787
[0008] <151> 2013-06-04
[0009] <160> 184
[0010] <170> PatentIn version 3.5
[0011] <210> 1
[0012] <211> 1368
[0013] <212> PRT
[0014] <213> 酿脓链球菌(Streptococcus pyogenes)
[0015] <400> 1
[0016] Met Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val
[0017] 1 5 10 15
[0018] Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
[0019] 20 25 30
[0020] Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
[0021] 35 40 45
[0022] Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
[0023] 50 55 60
[0024] Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
[0025] 65 70 75 80
[0026] Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
[0027] 85 90 95
[0028] Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
[0029] 100 105 110
[0030] His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
[0031] 115 120 125
[0032] His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
[0033] 130 135 140
[0034] Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
[0035] 145 150 155 160
[0036] Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
[0037] 165 170 175
[0038] Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
[0039] 180 185 190
[0040] Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
[0041] 195 200 205

[0042] Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
[0043] 210 215 220
[0044] Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
[0045] 225 230 235 240
[0046] Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
[0047] 245 250 255
[0048] Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
[0049] 260 265 270
[0050] Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
[0051] 275 280 285
[0052] Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
[0053] 290 295 300
[0054] Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
[0055] 305 310 315 320
[0056] Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
[0057] 325 330 335
[0058] Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
[0059] 340 345 350
[0060] Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
[0061] 355 360 365
[0062] Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
[0063] 370 375 380
[0064] Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
[0065] 385 390 395 400
[0066] Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
[0067] 405 410 415
[0068] Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
[0069] 420 425 430
[0070] Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
[0071] 435 440 445
[0072] Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
[0073] 450 455 460
[0074] Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
[0075] 465 470 475 480
[0076] Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
[0077] 485 490 495
[0078] Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
[0079] 500 505 510
[0080] Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
[0081] 515 520 525
[0082] Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
[0083] 530 535 540

[0084] Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
[0085] 545 550 555 560
[0086] Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
[0087] 565 570 575
[0088] Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
[0089] 580 585 590
[0090] Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
[0091] 595 600 605
[0092] Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
[0093] 610 615 620
[0094] Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
[0095] 625 630 635 640
[0096] His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
[0097] 645 650 655
[0098] Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
[0099] 660 665 670
[0100] Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
[0101] 675 680 685
[0102] Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
[0103] 690 695 700
[0104] Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
[0105] 705 710 715 720
[0106] His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
[0107] 725 730 735
[0108] Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
[0109] 740 745 750
[0110] Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
[0111] 755 760 765
[0112] Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
[0113] 770 775 780
[0114] Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
[0115] 785 790 795 800
[0116] Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
[0117] 805 810 815
[0118] Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
[0119] 820 825 830
[0120] Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
[0121] 835 840 845
[0122] Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
[0123] 850 855 860
[0124] Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
[0125] 865 870 875 880

[0126] Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
[0127] 885 890 895
[0128] Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
[0129] 900 905 910
[0130] Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
[0131] 915 920 925
[0132] Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
[0133] 930 935 940
[0134] Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
[0135] 945 950 955 960
[0136] Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
[0137] 965 970 975
[0138] Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
[0139] 980 985 990
[0140] Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe
[0141] 995 1000 1005
[0142] Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala
[0143] 1010 1015 1020
[0144] Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe
[0145] 1025 1030 1035
[0146] Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala
[0147] 1040 1045 1050
[0148] Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu
[0149] 1055 1060 1065
[0150] Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val
[0151] 1070 1075 1080
[0152] Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr
[0153] 1085 1090 1095
[0154] Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys
[0155] 1100 1105 1110
[0156] Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro
[0157] 1115 1120 1125
[0158] Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val
[0159] 1130 1135 1140
[0160] Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys
[0161] 1145 1150 1155
[0162] Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser
[0163] 1160 1165 1170
[0164] Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys
[0165] 1175 1180 1185
[0166] Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu
[0167] 1190 1195 1200

[0168] Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly
[0169] 1205 1210 1215
[0170] Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val
[0171] 1220 1225 1230
[0172] Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser
[0173] 1235 1240 1245
[0174] Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys
[0175] 1250 1255 1260
[0176] His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys
[0177] 1265 1270 1275
[0178] Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala
[0179] 1280 1285 1290
[0180] Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn
[0181] 1295 1300 1305
[0182] Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala
[0183] 1310 1315 1320
[0184] Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser
[0185] 1325 1330 1335
[0186] Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr
[0187] 1340 1345 1350
[0188] Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp
[0189] 1355 1360 1365
[0190] <210> 2
[0191] <211> 4332
[0192] <212> DNA
[0193] <213> 人工序列
[0194] <220>
[0195] <223> VP64-激活剂构建体
[0196] <400> 2
[0197] gccaccatgg acaagaagta ctccattggg ctcgctatcg gcacaaacag cgtcggctgg 60
[0198] gccgtcatta cggacgagta caaggtgccg agcaaaaaat tcaaagttct gggcaatacc 120
[0199] gatcgccaca gcataaagaa gaacctcatt ggcgccctcc tgttcgactc cggggagacg 180
[0200] gccgaagcca cgcggctcaa aagaacagca cggcgcagat atacccgcag aaagaatcgg 240
[0201] atctgctacc tgcaggagat ctttagtaat gagatggcta aggtggatga ctctttcttc 300
[0202] cataggctgg aggagtcctt tttggtggag gaggataaaa agcacgagcg ccacccaatc 360
[0203] tttggcaata tcgtggacga ggtggcgtac catgaaaagt acccaaccat atatcatctg 420
[0204] aggaagaagc ttgtagacag tactgataag gctgacttgc ggttgatcta tctcgcgctg 480
[0205] gcgcatatga tcaaatttcg gggacacttc ctcatcgagg gggacctgaa cccagacaac 540
[0206] agcgatgtcg acaaactctt tatccaactg gttcagactt acaatcagct tttcgaagag 600
[0207] aacccgatca acgcatccgg agttgacgcc aaagcaatcc tgagcgctag gctgtccaaa 660
[0208] tcccggcggc tcgaaaacct catcgcacag ctccctgggg agaagaagaa cggcctgttt 720
[0209] ggtaatctta tcgccctgtc actcgggctg acccccaact ttaaatctaa cttcgacctg 780

[0210] gccgaagatg ccaagcttca actgagcaaa gacacctacg atgatgatct cgacaatctg 840
[0211] ctggcccaga tcggcgacca gtacgcagac ctttttttgg cggcaaagaa cctgtcagac 900
[0212] gccattctgc tgagtgatat tctgcgagtg aacacggaga tcaccaaagc tccgctgagc 960
[0213] gctagtatga tcaagcgcta tgatgagcac caccaagact tgactttgct gaaggccctt 1020
[0214] gtcagacagc aactgcctga gaagtacaag gaaattttct tcgatcagtc taaaaatggc 1080
[0215] tacgccggat acattgacgg cggagcaagc caggaggaat tttacaaatt tattaagccc 1140
[0216] atcttggaaa aaatggacgg caccgaggag ctgctggtaa agcttaacag agaagatctg 1200
[0217] ttgcgcaaac agcgcacttt cgacaatgga agcatccccc accagattca cctgggcgaa 1260
[0218] ctgcacgcta tcctcaggcg gcaagaggat ttctacccct ttttgaaaga taacagggaa 1320
[0219] aagattgaga aaatcctcac atttcggata ccctactatg taggcccccct cgcccgggga 1380
[0220] aattccagat tcgcgtggat gactcgcaaa tcagaagaga ccatcactcc ctggaacttc 1440
[0221] gaggaagtcg tggataaggg ggcctctgcc cagtccttca tcgaaaggat gactaacttt 1500
[0222] gataaaaatc tgcctaacga aaaggtgctt cctaaacact ctctgctgta cgagtacttc 1560
[0223] acagtttata acgagctcac caaggtcaaa tacgtcacag aagggatgag aaagccagca 1620
[0224] ttcctgtctg gagagcagaa gaaagctatc gtggacctcc tcttcaagac gaaccggaaa 1680
[0225] gttaccgtga aacagctcaa agaagactat ttcaaaaaga ttgaatgttt cgactctgtt 1740
[0226] gaaatcagcg gagtggagga tcgcttcaac gcatccctgg gaacgtatca cgatctcctg 1800
[0227] aaaatcatta aagacaagga cttcctggac aatgaggaga acgaggacat tcttgaggac 1860
[0228] attgtcctca cccttacgtt gtttgaagat agggagatga ttgaagaacg cttgaaaact 1920
[0229] tacgctcatc tcttcgacga caaagtcatg aaacagctca agaggcgccg atatacagga 1980
[0230] tgggggcggc tgtcaagaaa actgatcaat gggatccgag acaagcagag tggaaagaca 2040
[0231] atcctggatt ttcttaagtc cgatggattt gccaaccgga acttcatgca gttgatccat 2100
[0232] gatgactctc tcacctttaa ggaggacatc cagaaagcac aagtttctgg ccagggggac 2160
[0233] agtcttcacg agcacatcgc taatcttgca ggtagcccag ctatcaaaaa gggaatactg 2220
[0234] cagaccgtta aggtcgtgga tgaactcgtc aaagtaatgg gaaggcataa gcccgagaat 2280
[0235] atcgttatcg agatggcccg agagaaccaa actacccaga agggacagaa gaacagtagg 2340
[0236] gaaaggatga agaggattga agagggtata aaagaactgg ggtcccaaat ccttaaggaa 2400
[0237] cacccagttg aaaacaccca gcttcagaat gagaagctct acctgtacta cctgcagaac 2460
[0238] ggcagggaca tgtacgtgga tcaggaactg gacatcaatc ggctctccga ctacgacgtg 2520
[0239] gctgctatcg tgccccagtc ttttctcaaa gatgattcta ttgataataa agtgttgaca 2580
[0240] agatccgata aagctagagg gaagagtgat aacgtcccct cagaagaagt tgtcaagaaa 2640
[0241] atgaaaaatt attggcggca gctgctgaac gccaaactga tcacacaacg gaagttcgat 2700
[0242] aatctgacta aggctgaacg aggtggcctg tctgagttgg ataaagccgg cttcatcaaa 2760
[0243] aggcagcttg ttgagacacg ccagatcacc aagcacgtgg cccaaattct cgattcacgc 2820
[0244] atgaacacca agtacgatga aaatgacaaa ctgattcgag aggtgaaagt tattactctg 2880
[0245] aagtctaagc tggtctcaga tttcagaaag gactttcagt tttataaggt gagagagatc 2940
[0246] aacaattacc accatgcgca tgatgcctac ctgaatgcag tggtaggcac tgcacttatc 3000
[0247] aaaaaatatc ccaagcttga atctgaattt gtttacggag actataaagt gtacgatgtt 3060
[0248] aggaaaatga tcgcaaagtc tgagcaggaa ataggcaagg ccaccgctaa gtacttcttt 3120
[0249] tacagcaata ttatgaattt tttcaagacc gagattacac tggccaatgg agagattcgg 3180
[0250] aagcgaccac ttatcgaaac aaacggagaa acaggagaaa tcgtgtggga caagggtagg 3240
[0251] gatttcgcga cagtccggaa ggtcctgtcc atgccgcagg tgaacatcgt taaaaagacc 3300

[0252] gaagtacaga ccggaggctt ctccaaggaa agtatcctcc cgaaaaggaa cagcgacaag 3360
[0253] ctgatcgcac gcaaaaaaga ttgggacccc aagaaatacg gcggattcga ttctcctaca 3420
[0254] gtcgcttaca gtgtactggt tgtggccaaa gtggagaaag ggaagtctaa aaaactcaaa 3480
[0255] agcgtcaagg aactgctggg catcacaatc atggagcgat caagcttcga aaaaaacccc 3540
[0256] atcgactttc tcgaggcgaa aggatataaa gaggtcaaaa aagacctcat cattaagctt 3600
[0257] cccaagtact ctctctttga gcttgaaaac ggccggaaac gaatgctcgc tagtgcgggc 3660
[0258] gagctgcaga aaggtaacga gctggcactg ccctctaaat acgttaattt cttgtatctg 3720
[0259] gccagccact atgaaaagct caaagggtct cccgaagata atgagcagaa gcagctgttc 3780
[0260] gtggaacaac acaaacacta ccttgatgag atcatcgagc aaataagcga attctccaaa 3840
[0261] agagtgatcc tcgccgacgc taacctcgat aaggtgcttt ctgcttacaa taagcacagg 3900
[0262] gataagccca tcagggagca ggcagaaaac attatccact tgtttactct gaccaacttg 3960
[0263] ggcgcgcctg cagccttcaa gtacttcgac accaccatag acagaaagcg gtacacctct 4020
[0264] acaaaggagg tcctggacgc cacactgatt catcagtcaa ttacggggct ctatgaaaca 4080
[0265] agaatcgacc tctctcagct cggtggagac agcagggctg accccaagaa gaagaggaag 4140
[0266] gtggaggcca gcggttccgg acgggctgac gcattggacg attttgatct ggatatgctg 4200
[0267] ggaagtgacg ccctcgatga ttttgacctt gacatgcttg gttcggatgc ccttgatgac 4260
[0268] tttgacctcg acatgctcgg cagtgacgcc cttgatgatt tcgacctgga catgctgatt 4320
[0269] aactctagat ga 4332
[0270] <210> 3
[0271] <211> 4365
[0272] <212> DNA
[0273] <213> 人工序列
[0274] <220>
[0275] <223> VP64-激活剂构建体
[0276] <400> 3
[0277] gccaccatgc ccaagaagaa gaggaaggtg ggaaggggga tggacaagaa gtactccatt 60
[0278] gggctcgcta tcggcacaaa cagcgtcggc tgggccgtca ttacggacga gtacaaggtg 120
[0279] ccgagcaaaa aattcaaagt tctgggcaat accgatcgcc acagcataaa gaagaacctc 180
[0280] attggcgccc tcctgttcga ctccggggag acggccgaag ccacgcggct caaaagaaca 240
[0281] gcacggcgca gatatacccg cagaaagaat cggatctgct acctgcagga gatctttagt 300
[0282] aatgagatgg ctaaggtgga tgactctttc ttccataggc tggaggagtc ctttttggtg 360
[0283] gaggaggata aaaagcacga gcgccaccca atctttggca atatcgtgga cgaggtggcg 420
[0284] taccatgaaa agtacccaac catatatcat ctgaggaaga agcttgtaga cagtactgat 480
[0285] aaggctgact tgcggttgat ctatctcgcg ctggcgcata tgatcaaatt tcggggacac 540
[0286] ttcctcatcg agggggacct gaacccagac aacagcgatg tcgacaaact ctttatccaa 600
[0287] ctggttcaga cttacaatca gcttttcgaa gagaacccga tcaacgcatc cggagttgac 660
[0288] gccaaagcaa tcctgagcgc taggctgtcc aaatcccggc ggctcgaaaa cctcatcgca 720
[0289] cagctccctg gggagaagaa gaacggcctg tttggtaatc ttatcgccct gtcactcggg 780
[0290] ctgaccccca actttaaatc taacttcgac ctggccgaag atgccaagct tcaactgagc 840
[0291] aaagacacct acgatgatga tctcgacaat ctgctggccc agatcggcga ccagtacgca 900
[0292] gacctttttt tggcggcaaa gaacctgtca gacgccattc tgctgagtga tattctgcga 960
[0293] gtgaacacgg agatcaccaa agctccgctg agcgctagta tgatcaagcg ctatgatgag 1020

[0294] caccaccaag acttgacttt gctgaaggcc cttgtcagac agcaactgcc tgagaagtac 1080
[0295] aaggaaattt tcttcgatca gtctaaaaat ggctacgccg gatacattga cggcggagca 1140
[0296] agccaggagg aattttacaa atttattaag cccatcttgg aaaaaatgga cggcaccgag 1200
[0297] gagctgctgg taaagcttaa cagagaagat ctgttgcgca aacagcgcac tttcgacaat 1260
[0298] ggaagcatcc cccaccagat tcacctgggc gaactgcacg ctatcctcag gcggcaagag 1320
[0299] gatttctacc cctttttgaa agataacagg gaaaagattg agaaaatcct cacatttcgg 1380
[0300] ataccctact atgtaggccc cctcgcccgg ggaaattcca gattcgcgtg gatgactcgc 1440
[0301] aaatcagaag agaccatcac tccctggaac ttcgaggaag tcgtggataa gggggcctct 1500
[0302] gcccagtcct tcatcgaaag gatgactaac tttgataaaa atctgcctaa cgaaaaggtg 1560
[0303] cttcctaaac actctctgct gtacgagtac ttcacagttt ataacgagct caccaaggtc 1620
[0304] aaatacgtca cagaagggat gagaaagcca gcattcctgt ctggagagca gaagaaagct 1680
[0305] atcgtggacc tcctcttcaa gacgaaccgg aaagttaccg tgaaacagct caaagaagac 1740
[0306] tatttcaaaa agattgaatg tttcgactct gttgaaatca gcggagtgga ggatcgcttc 1800
[0307] aacgcatccc tgggaacgta tcacgatctc ctgaaaatca ttaaagacaa ggacttcctg 1860
[0308] gacaatgagg agaacgagga cattcttgag gacattgtcc tcacccttac gttgtttgaa 1920
[0309] gatagggaga tgattgaaga acgcttgaaa acttacgctc atctcttcga cgacaaagtc 1980
[0310] atgaaacagc tcaagaggcg ccgatataca ggatgggggc ggctgtcaag aaaactgatc 2040
[0311] aatgggatcc gagacaagca gagtggaaag acaatcctgg attttcttaa gtccgatgga 2100
[0312] tttgccaacc ggaacttcat gcagttgatc catgatgact ctctcacctt taaggaggac 2160
[0313] atccagaaag cacaagtttc tggccagggg gacagtcttc acgagcacat cgctaatctt 2220
[0314] gcaggtagcc cagctatcaa aaagggaata ctgcagaccg ttaaggtcgt ggatgaactc 2280
[0315] gtcaaagtaa tgggaaggca taagcccgag aatatcgtta tcgagatggc ccgagagaac 2340
[0316] caaactaccc agaagggaca gaagaacagt agggaaagga tgaagaggat tgaagagggt 2400
[0317] ataaaagaac tggggtccca aatccttaag gaacacccag ttgaaaacac ccagcttcag 2460
[0318] aatgagaagc tctacctgta ctacctgcag aacggcaggg acatgtacgt ggatcaggaa 2520
[0319] ctggacatca atcggctctc cgactacgac gtggctgcta tcgtgcccca gtcttttctc 2580
[0320] aaagatgatt ctattgataa taaagtgttg acaagatccg ataaagctag agggaagagt 2640
[0321] gataacgtcc cctcagaaga agttgtcaag aaaatgaaaa attattggcg gcagctgctg 2700
[0322] aacgccaaac tgatcacaca acggaagttc gataatctga ctaaggctga acgaggtggc 2760
[0323] ctgtctgagt tggataaagc cggcttcatc aaaaggcagc ttgttgagac acgccagatc 2820
[0324] accaagcacg tggcccaaat tctcgattca cgcatgaaca ccaagtacga tgaaaatgac 2880
[0325] aaactgattc gagaggtgaa agttattact ctgaagtcta agctggtctc agatttcaga 2940
[0326] aaggactttc agttttataa ggtgagagag atcaacaatt accaccatgc gcatgatgcc 3000
[0327] tacctgaatg cagtggtagg cactgcactt atcaaaaaat atcccaagct tgaatctgaa 3060
[0328] tttgtttacg gagactataa agtgtacgat gttaggaaaa tgatcgcaaa gtctgagcag 3120
[0329] gaaataggca aggccaccgc taagtacttc ttttacagca atattatgaa ttttttcaag 3180
[0330] accgagatta cactggccaa tggagagatt cggaagcgac cacttatcga aacaaacgga 3240
[0331] gaaacaggag aaatcgtgtg ggacaagggt agggatttcg cgacagtccg gaaggtcctg 3300
[0332] tccatgccgc aggtgaacat cgttaaaaag accgaagtac agaccggagg cttctccaag 3360
[0333] gaaagtatcc tcccgaaaag gaacagcgac aagctgatcg cacgcaaaaa agattgggac 3420
[0334] cccaagaaat acggcggatt cgattctcct acagtcgctt acagtgtact ggttgtggcc 3480
[0335] aaagtggaga aagggaagtc taaaaaactc aaaagcgtca aggaactgct gggcatcaca 3540

[0336] atcatggagc gatcaagctt cgaaaaaaac cccatcgact ttctcgaggc gaaaggatat 3600
[0337] aaagaggtca aaaaagacct catcattaag cttcccaagt actctctctt tgagcttgaa 3660
[0338] aacggccgga aacgaatgct cgctagtgcg ggcgagctgc agaaaggtaa cgagctggca 3720
[0339] ctgccctcta aatacgttaa tttcttgtat ctggccagcc actatgaaaa gctcaaaggg 3780
[0340] tctcccgaag ataatgagca gaagcagctg ttcgtggaac aacacaaaca ctaccttgat 3840
[0341] gagatcatcg agcaaataag cgaattctcc aaaagagtga tcctcgccga cgctaacctc 3900
[0342] gataaggtgc tttctgctta caataagcac agggataagc ccatcaggga gcaggcagaa 3960
[0343] aacattatcc acttgtttac tctgaccaac ttgggcgcgc ctgcagcctt caagtacttc 4020
[0344] gacaccacca tagacagaaa gcggtacacc tctacaaagg aggtcctgga cgccacactg 4080
[0345] attcatcagt caattacggg gctctatgaa acaagaatcg acctctctca gctcggtgga 4140
[0346] gacagcaggg ctgaccccaa gaagaagagg aaggtggagg ccagcggttc cggacgggct 4200
[0347] gacgcattgg acgattttga tctggatatg ctgggaagtg acgccctcga tgattttgac 4260
[0348] cttgacatgc ttggttcgga tgcccttgat gactttgacc tcgacatgct cggcagtgac 4320
[0349] gcccttgatg atttcgacct ggacatgctg attaactcta gatga 4365
[0350] <210> 4
[0351] <211> 4425
[0352] <212> DNA
[0353] <213> 人工序列
[0354] <220>
[0355] <223> VP64-激活剂构建体
[0356] <400> 4
[0357] gccaccatgg acaagaagta ctccattggg ctcgctatcg gcacaaacag cgtcggctgg 60
[0358] gccgtcatta cggacgagta caaggtgccg agcaaaaaat tcaaagttct gggcaatacc 120
[0359] gatcgccaca gcataaagaa gaacctcatt ggcgccctcc tgttcgactc cggggagacg 180
[0360] gccgaagcca cgcggctcaa aagaacagca cggcgcagat atacccgcag aaagaatcgg 240
[0361] atctgctacc tgcaggagat ctttagtaat gagatggcta aggtggatga ctctttcttc 300
[0362] cataggctgg aggagtcctt tttggtggag gaggataaaa agcacgagcg ccacccaatc 360
[0363] tttggcaata tcgtggacga ggtggcgtac catgaaaagt acccaaccat atatcatctg 420
[0364] aggaagaagc ttgtagacag tactgataag gctgacttgc ggttgatcta tctcgcgctg 480
[0365] gcgcatatga tcaaatttcg gggacacttc ctcatcgagg gggacctgaa cccagacaac 540
[0366] agcgatgtcg acaaactctt tatccaactg gttcagactt acaatcagct tttcgaagag 600
[0367] aacccgatca acgcatccgg agttgacgcc aaagcaatcc tgagcgctag gctgtccaaa 660
[0368] tcccggcggc tcgaaaacct catcgcacag ctccctgggg agaagaagaa cggcctgttt 720
[0369] ggtaatctta tcgccctgtc actcgggctg acccccaact ttaaatctaa cttcgacctg 780
[0370] gccgaagatg ccaagcttca actgagcaaa gacacctacg atgatgatct cgacaatctg 840
[0371] ctggcccaga tcggcgacca gtacgcagac ctttttttgg cggcaaagaa cctgtcagac 900
[0372] gccattctgc tgagtgatat tctgcgagtg aacacggaga tcaccaaagc tccgctgagc 960
[0373] gctagtatga tcaagcgcta tgatgagcac caccaagact tgactttgct gaaggccctt 1020
[0374] gtcagacagc aactgcctga gaagtacaag gaaattttct tcgatcagtc taaaaatggc 1080
[0375] tacgccggat acattgacgg cggagcaagc caggaggaat tttacaaatt tattaagccc 1140
[0376] atcttggaaa aaatggacgg caccgaggag ctgctggtaa agcttaacag agaagatctg 1200
[0377] ttgcgcaaac agcgcacttt cgacaatgga agcatccccc accagattca cctgggcgaa 1260

[0378] ctgcacgcta tcctcaggcg gcaagaggat ttctacccct ttttgaaaga taacagggaa 1320
[0379] aagattgaga aaatcctcac atttcggata ccctactatg taggcccccct cgcccgggga 1380
[0380] aattccagat tcgcgtggat gactcgcaaa tcagaagaga ccatcactcc ctggaacttc 1440
[0381] gaggaagtcg tggataaggg ggcctctgcc cagtccttca tcgaaaggat gactaacttt 1500
[0382] gataaaaatc tgcctaacga aaaggtgctt cctaaacact ctctgctgta cgagtacttc 1560
[0383] acagtttata acgagctcac caaggtcaaa tacgtcacag aagggatgag aaagccagca 1620
[0384] ttcctgtctg gagagcagaa gaaagctatc gtggacctcc tcttcaagac gaaccggaaa 1680
[0385] gttaccgtga aacagctcaa agaagactat ttcaaaaaga ttgaatgttt cgactctgtt 1740
[0386] gaaatcagcg gagtggagga tcgcttcaac gcatccctgg gaacgtatca cgatctcctg 1800
[0387] aaaatcatta aagacaagga cttcctggac aatgaggaga acgaggacat tcttgaggac 1860
[0388] attgtcctca cccttacgtt gtttgaagat agggagatga ttgaagaacg cttgaaaact 1920
[0389] tacgctcatc tcttcgacga caaagtcatg aaacagctca agaggcgccg atatacagga 1980
[0390] tgggggcggc tgtcaagaaa actgatcaat gggatccgag acaagcagag tggaaagaca 2040
[0391] atcctggatt ttcttaagtc cgatggattt gccaaccgga acttcatgca gttgatccat 2100
[0392] gatgactctc tcacctttaa ggaggacatc cagaaagcac aagtttctgg ccagggggac 2160
[0393] agtcttcacg agcacatcgc taatcttgca ggtagcccag ctatcaaaaa gggaatactg 2220
[0394] cagaccgtta aggtcgtgga tgaactcgtc aaagtaatgg gaaggcataa gcccgagaat 2280
[0395] atcgttatcg agatggcccg agagaaccaa actacccaga agggacagaa gaacagtagg 2340
[0396] gaaaggatga agaggattga agagggtata aaagaactgg ggtcccaaat ccttaaggaa 2400
[0397] cacccagttg aaaacaccca gcttcagaat gagaagctct acctgtacta cctgcagaac 2460
[0398] ggcagggaca tgtacgtgga tcaggaactg gacatcaatc ggctctccga ctacgacgtg 2520
[0399] gctgctatcg tgccccagtc ttttctcaaa gatgattcta ttgataataa agtgttgaca 2580
[0400] agatccgata aagctagagg gaagagtgat aacgtcccct cagaagaagt tgtcaagaaa 2640
[0401] atgaaaaatt attggcggca gctgctgaac gccaaactga tcacacaacg gaagttcgat 2700
[0402] aatctgacta aggctgaacg aggtggcctg tctgagttgg ataaagccgg cttcatcaaa 2760
[0403] aggcagcttg ttgagacacg ccagatcacc aagcacgtgg cccaaattct cgattcacgc 2820
[0404] atgaacacca agtacgatga aaatgacaaa ctgattcgag aggtgaaagt tattactctg 2880
[0405] aagtctaagc tggtctcaga tttcagaaag gactttcagt tttataaggt gagagagatc 2940
[0406] aacaattacc accatgcgca tgatgcctac ctgaatgcag tggtaggcac tgcacttatc 3000
[0407] aaaaaatatc ccaagcttga atctgaattt gtttacggag actataaagt gtacgatgtt 3060
[0408] aggaaaatga tcgcaaagtc tgagcaggaa ataggcaagg ccaccgctaa gtacttcttt 3120
[0409] tacagcaata ttatgaattt tttcaagacc gagattacac tggccaatgg agagattcgg 3180
[0410] aagcgaccac ttatcgaaac aaacggagaa acaggagaaa tcgtgtggga caagggtagg 3240
[0411] gatttcgcga cagtccggaa ggtcctgtcc atgccgcagg tgaacatcgt taaaaagacc 3300
[0412] gaagtacaga ccggaggctt ctccaaggaa agtatcctcc cgaaaaggaa cagcgacaag 3360
[0413] ctgatcgcac gcaaaaaaga ttgggacccc aagaaatacg gcggattcga ttctcctaca 3420
[0414] gtcgcttaca gtgtactggt tgtggccaaa gtggagaaag ggaagtctaa aaaactcaaa 3480
[0415] agcgtcaagg aactgctggg catcacaatc atggagcgat caagcttcga aaaaaacccc 3540
[0416] atcgactttc tcgaggcgaa aggatataaa gaggtcaaaa aagacctcat cattaagctt 3600
[0417] cccaagtact ctctctttga gcttgaaaac ggccggaaac gaatgctcgc tagtgcgggc 3660
[0418] gagctgcaga aaggtaacga gctggcactg ccctctaaat acgttaattt cttgtatctg 3720
[0419] gccagccact atgaaaagct caaagggtct cccgaagata atgagcagaa gcagctgttc 3780

[0420] gtggaacaac acaaacacta ccttgatgag atcatcgagc aaataagcga attctccaaa 3840
[0421] agagtgatcc tcgccgacgc taacctcgat aaggtgcttt ctgcttacaa taagcacagg 3900
[0422] gataagccca tcagggagca ggcagaaaac attatccact tgtttactct gaccaacttg 3960
[0423] ggcgcgcctg cagccttcaa gtacttcgac accaccatag acagaaagcg gtacacctct 4020
[0424] acaaaggagg tcctggacgc cacactgatt catcagtcaa ttacggggct ctatgaaaca 4080
[0425] agaatcgacc tctctcagct cggtggagac agcagggctg accccaagaa gaagaggaag 4140
[0426] gtggaggcca gcggttccgg acgggctgac gcattggacg attttgatct ggatatgctg 4200
[0427] ggaagtgacg ccctcgatga ttttgacctt gacatgcttg gttcggatgc ccttgatgac 4260
[0428] tttgacctcg acatgctcgg cagtgacgcc cttgatgatt tcgacctgga catgctgatt 4320
[0429] aactctagag cggccgcaga tccaaaaaag aagagaaagg tagatccaaa aaagaagaga 4380
[0430] aaggtagatc caaaaaagaa gagaaaggta gatacggccg catag 4425
[0431] <210> 5
[0432] <211> 587
[0433] <212> DNA
[0434] <213> 人工序列
[0435] <220>
[0436] <223> MS2-激活剂构建体
[0437] <400> 5
[0438] ccaccatggg acctaagaaa aagaggaagg tggcggccgc ttctagaatg gcttctaact 60
[0439] ttactcagtt cgttctcgtc gacaatggcg gaactggcga cgtgactgtc gccccaagca 120
[0440] acttcgctaa cgggatcgct gaatggatca gctctaactc gcgttcacag gcttacaaag 180
[0441] taacctgtag cgttcgtcag agctctgcgc agaatcgcaa atacaccatc aaagtcgagg 240
[0442] tgcctaaagg cgcctggcgt tcgtacttaa atatggaact aaccattcca attttcgcca 300
[0443] cgaattccga ctgcgagctt attgttaagg caatgcaagg tctcctaaaa gatggaaacc 360
[0444] cgattccctc agcaatcgca gcaaactccg gcatctacga ggccagcggt tccggacggg 420
[0445] ctgacgcatt ggacgatttt gatctggata tgctgggaag tgacgccctc gatgattttg 480
[0446] accttgacat gcttggttcg gatgcccttg atgactttga cctcgacatg ctcggcagtg 540
[0447] acgcccttga tgatttcgac ctggacatgc tgattaactc tagatga 587
[0448] <210> 6
[0449] <211> 681
[0450] <212> DNA
[0451] <213> 人工序列
[0452] <220>
[0453] <223> MS2-激活剂构建体
[0454] <400> 6
[0455] gccaccatgg gacctaagaa aaagaggaag gtggcggccg cttctagaat ggcttctaac 60
[0456] tttactcagt tcgttctcgt cgacaatggc ggaactggcg acgtgactgt cgccccaagc 120
[0457] aacttcgcta acgggatcgc tgaatggatc agctctaact cgcgttcaca ggcttacaaa 180
[0458] gtaacctgta gcgttcgtca gagctctgcg cagaatcgca aatacaccat caaagtcgag 240
[0459] gtgcctaaag gcgcctggcg ttcgtactta aatatggaac taaccattcc aattttcgcc 300
[0460] acgaattccg actgcgagct tattgttaag gcaatgcaag gtctcctaaa agatggaaac 360
[0461] ccgattccct cagcaatcgc agcaaactcc ggcatctacg aggccagcgg ttccggacgg 420

[0462] gctgacgcat tggacgattt tgatctggat atgctgggaa gtgacgccct cgatgatttt 480
[0463] gaccttgaca tgcttggttc ggatgccctt gatgactttg acctcgacat gctcggcagt 540
[0464] gacgcccttg atgatttcga cctggacatg ctgattaact ctagagcggc cgcagatcca 600
[0465] aaaaagaaga gaaaggtaga tccaaaaaag aagagaaagg tagatccaaa aaagaagaga 660
[0466] aaggtagata cggccgcata g 681
[0467] <210> 7
[0468] <211> 557
[0469] <212> DNA
[0470] <213> 人工序列
[0471] <220>
[0472] <223> MS2-激活剂构建体
[0473] <220>
[0474] <221> misc_feature
[0475] <222> (320)..(339)
[0476] <223> 其中N是G, A, T或C
[0477] <400> 7
[0478] tgtacaaaaa agcaggcttt aaaggaacca attcagtcga ctggatccgg taccaaggtc 60
[0479] gggcaggaag agggcctatt tcccatgatt ccttcatatt tgcatatacg atacaaggct 120
[0480] gttagagaga taattagaat taatttgact gtaaacacaa agatattagt acaaaatacg 180
[0481] tgacgtagaa agtaataatt tcttgggtag tttgcagttt taaaattatg ttttaaaatg 240
[0482] gactatcata tgcttaccgt aacttgaaag tatttcgatt tcttggcttt atatatcttg 300
[0483] tggaaaggac gaaacaccgn nnnnnnnnnn nnnnnnnnng ttttagagct agaaatagca 360
[0484] agttaaaata aggctagtcc gttatcaact tgaaaaagtg gcaccgagtc ggtgctctgc 420
[0485] aggtcgactc tagaaaacat gaggatcacc catgtctgca gtattcccgg gttcattaga 480
[0486] tcctaaggta cctaattgcc tagaaaacat gaggatcacc catgtctgca ggtcgactct 540
[0487] agaaattttt tctagac 557
[0488] <210> 8
[0489] <211> 882
[0490] <212> DNA
[0491] <213> 人工序列
[0492] <220>
[0493] <223> 活化报告子构建体
[0494] <400> 8
[0495] tagggataac agggtaatag tgtcccctcc accccacagt ggggcgaggt aggcgtgtac 60
[0496] ggtgggaggc ctatataagc agagctcgtt tagtgaaccg tcagatcgcc tggagaattc 120
[0497] gccaccatgg actacaagga tgacgacgat aaaacttccg gtggcggact gggttccacc 180
[0498] gtgagcaagg gcgaggaggt catcaaagag ttcatgcgct tcaaggtgcg catggagggc 240
[0499] tccatgaacg gccacgagtt cgagatcgag ggcgagggcg agggccgccc ctacgagggc 300
[0500] acccagaccg ccaagctgaa ggtgaccaag ggcggccccc tgcccttcgc ctgggacatc 360
[0501] ctgtcccccc agttcatgta cggctccaag gcgtacgtga agcaccccgc cgacatcccc 420
[0502] gattacaaga agctgtcctt ccccgagggc ttcaagtggg agcgcgtgat gaacttcgag 480
[0503] gacggcggtc tggtgaccgt gacccaggac tcctccctgc aggacggcac gctgatctac 540

[0504] aaggtgaaga tgcgcggcac caacttcccc cccgacggcc ccgtaatgca gaagaagacc 600
[0505] atgggctggg aggcctccac cgagcgcctg tacccccgcg acggcgtgct gaagggcgag 660
[0506] atccaccagg ccctgaagct gaaggacggc ggccactacc tggtggagtt caagaccatc 720
[0507] tacatggcca agaagcccgt gcaactgccc ggctactact acgtggacac caagctggac 780
[0508] atcacctccc acaacgagga ctacaccatc gtggaacagt acgagcgctc cgagggccgc 840
[0509] caccacctgt tcctgtacgg catggacgag ctgtacaagt aa 882
[0510] <210> 9
[0511] <211> 882
[0512] <212> DNA
[0513] <213> 人工序列
[0514] <220>
[0515] <223> 活化报告子构建体
[0516] <400> 9
[0517] tagggataac agggtaatag tggggccact agggacagga ttggcgaggt aggcgtgtac 60
[0518] ggtgggaggc ctatataagc agagctcgtt tagtgaaccg tcagatcgcc tggagaattc 120
[0519] gccaccatgg actacaagga tgacgacgat aaaacttccg gtggcggact gggttccacc 180
[0520] gtgagcaagg gcgaggaggt catcaaagag ttcatgcgct tcaaggtgcg catggagggc 240
[0521] tccatgaacg gccacgagtt cgagatcgag ggcgagggcg agggccgccc ctacgagggc 300
[0522] acccagaccg ccaagctgaa ggtgaccaag ggcggccccc tgcccttcgc ctgggacatc 360
[0523] ctgtcccccc agttcatgta cggctccaag gcgtacgtga agcaccccgc cgacatcccc 420
[0524] gattacaaga agctgtcctt ccccgagggc ttcaagtggg agcgcgtgat gaacttcgag 480
[0525] gacggcggtc tggtgaccgt gacccaggac tcctccctgc aggacggcac gctgatctac 540
[0526] aaggtgaaga tgcgcggcac caacttcccc cccgacggcc ccgtaatgca gaagaagacc 600
[0527] atgggctggg aggcctccac cgagcgcctg tacccccgcg acggcgtgct gaagggcgag 660
[0528] atccaccagg ccctgaagct gaaggacggc ggccactacc tggtggagtt caagaccatc 720
[0529] tacatggcca agaagcccgt gcaactgccc ggctactact acgtggacac caagctggac 780
[0530] atcacctccc acaacgagga ctacaccatc gtggaacagt acgagcgctc cgagggccgc 840
[0531] caccacctgt tcctgtacgg catggacgag ctgtacaagt aa 882
[0532] <210> 10
[0533] <211> 912
[0534] <212> DNA
[0535] <213> 人工序列
[0536] <220>
[0537] <223> 特异性报告子文库
[0538] <220>
[0539] <221> misc_feature
[0540] <222> (22)..(44)
[0541] <223> 其中N是G, A, T或C
[0542] <220>
[0543] <221> misc_feature
[0544] <222> (154)..(177)
[0545] <223> 其中N是G, A, T或C

[0546] <400> 10

[0547] tagggataac agggtaatag tnnnnnnnnn nnnnnnnnnn nnnncgaggt aggcgtgtac 60

[0548] ggtgggaggc ctatataagc agagctcgtt tagtgaaccg tcagatcgcc tggagaattc 120

[0549] gccaccatgg actacaagga tgacgacgat aaannnnnnn nnnnnnnnnn nnnnnnnact 180

[0550] tccggtggcg gactgggttc caccgtgagc aagggcgagg aggtcatcaa agagttcatg 240

[0551] cgcttcaagg tgcgcatgga gggctccatg aacggccacg agttcgagat cgagggcgag 300

[0552] ggcgagggcc gcccctacga gggcacccag accgccaagc tgaaggtgac caagggcggc 360

[0553] cccctgccct tcgcctggga catcctgtcc ccccagttca tgtacggctc caaggcgtac 420

[0554] gtgaagcacc ccgccgacat ccccgattac aagaagctgt ccttccccga gggcttcaag 480

[0555] tgggagcgcg tgatgaactt cgaggacggc ggtctggtga ccgtgaccca ggactcctcc 540

[0556] ctgcaggacg gcacgctgat ctacaaggtg aagatgcgcg gcaccaactt ccccccccgac 600

[0557] ggccccgtaa tgcagaagaa gaccatgggc tgggaggcct ccaccgagcg cctgtacccc 660

[0558] cgcgacggcg tgctgaaggg cgagatccac caggccctga agctgaagga cggcggccac 720

[0559] tacctggtgg agttcaagac catctacatg gccaagaagc ccgtgcaact gcccggctac 780

[0560] tactacgtgg acaccaagct ggacatcacc tcccacaacg aggactacac catcgtggaa 840

[0561] cagtacgagc gctccgaggg ccgccaccac ctgttcctgt acggcatgga cgagctgtac 900

[0562] aagtaagaat tc 912

[0563] <210> 11

[0564] <211> 23

[0565] <212> DNA

[0566] <213> 人工序列

[0567] <220>

[0568] <223> 靶向探针

[0569] <400> 11

[0570] ctggcggatc actcgcggtt agg 23

[0571] <210> 12

[0572] <211> 23

[0573] <212> DNA

[0574] <213> 人工序列

[0575] <220>

[0576] <223> 靶向探针

[0577] <400> 12

[0578] cctcggcctc caaaagtgct agg 23

[0579] <210> 13

[0580] <211> 23

[0581] <212> DNA

[0582] <213> 人工序列

[0583] <220>

[0584] <223> 靶向探针

[0585] <400> 13

[0586] acgctgattc ctgcagatca ggg 23

[0587] <210> 14

[0588] <211> 23
[0589] <212> DNA
[0590] <213> 人工序列
[0591] <220>
[0592] <223> 靶向探针
[0593] <400> 14
[0594] ccaggaatac gtatccacca ggg 23
[0595] <210> 15
[0596] <211> 23
[0597] <212> DNA
[0598] <213> 人工序列
[0599] <220>
[0600] <223> 靶向探针
[0601] <400> 15
[0602] gccacaccca agcgatcaaa tgg 23
[0603] <210> 16
[0604] <211> 23
[0605] <212> DNA
[0606] <213> 人工序列
[0607] <220>
[0608] <223> 靶向探针
[0609] <400> 16
[0610] aaataataca ttctaaggta agg 23
[0611] <210> 17
[0612] <211> 23
[0613] <212> DNA
[0614] <213> 人工序列
[0615] <220>
[0616] <223> 靶向探针
[0617] <400> 17
[0618] gctactgggg aggctgaggc agg 23
[0619] <210> 18
[0620] <211> 23
[0621] <212> DNA
[0622] <213> 人工序列
[0623] <220>
[0624] <223> 靶向探针
[0625] <400> 18
[0626] tagcaataca gtcacattaa tgg 23
[0627] <210> 19
[0628] <211> 23
[0629] <212> DNA

[0630] <213> 人工序列
[0631] <220>
[0632] <223> 靶向探针
[0633] <400> 19
[0634] ctcatgtgat ccccccgtct cgg 23
[0635] <210> 20
[0636] <211> 23
[0637] <212> DNA
[0638] <213> 人工序列
[0639] <220>
[0640] <223> 靶向探针
[0641] <400> 20
[0642] ccgggcagag agtgaacgcg cgg 23
[0643] <210> 21
[0644] <211> 23
[0645] <212> DNA
[0646] <213> 人工序列
[0647] <220>
[0648] <223> 靶向探针
[0649] <400> 21
[0650] ttccttccct ctcccgtgct tgg 23
[0651] <210> 22
[0652] <211> 23
[0653] <212> DNA
[0654] <213> 人工序列
[0655] <220>
[0656] <223> 靶向探针
[0657] <400> 22
[0658] tctctgcaaa gcccctggag agg 23
[0659] <210> 23
[0660] <211> 23
[0661] <212> DNA
[0662] <213> 人工序列
[0663] <220>
[0664] <223> 靶向探针
[0665] <400> 23
[0666] aatgcagttg ccgagtgcag tgg 23
[0667] <210> 24
[0668] <211> 23
[0669] <212> DNA
[0670] <213> 人工序列
[0671] <220>

[0672] <223> 靶向探针
[0673] <400> 24
[0674] cctcagcctc ctaaagtgct ggg 23
[0675] <210> 25
[0676] <211> 23
[0677] <212> DNA
[0678] <213> 人工序列
[0679] <220>
[0680] <223> 靶向探针
[0681] <400> 25
[0682] gagtccaaat cctctttact agg 23
[0683] <210> 26
[0684] <211> 23
[0685] <212> DNA
[0686] <213> 人工序列
[0687] <220>
[0688] <223> 靶向探针
[0689] <400> 26
[0690] gagtgtctgg atttgggata agg 23
[0691] <210> 27
[0692] <211> 23
[0693] <212> DNA
[0694] <213> 人工序列
[0695] <220>
[0696] <223> 靶向探针
[0697] <400> 27
[0698] cagcacctca tctcccagtg agg 23
[0699] <210> 28
[0700] <211> 23
[0701] <212> DNA
[0702] <213> 人工序列
[0703] <220>
[0704] <223> 靶向探针
[0705] <400> 28
[0706] tctaaaaccc agggaatcat ggg 23
[0707] <210> 29
[0708] <211> 23
[0709] <212> DNA
[0710] <213> 人工序列
[0711] <220>
[0712] <223> 靶向探针
[0713] <400> 29

[0714] cacaaggcag ccagggatcc agg 23
[0715] <210> 30
[0716] <211> 23
[0717] <212> DNA
[0718] <213> 人工序列
[0719] <220>
[0720] <223> 靶向探针
[0721] <400> 30
[0722] gatggcaagc tgagaaacac tgg 23
[0723] <210> 31
[0724] <211> 23
[0725] <212> DNA
[0726] <213> 人工序列
[0727] <220>
[0728] <223> 靶向探针
[0729] <400> 31
[0730] tgaaatgcac gcatacaatt agg 23
[0731] <210> 32
[0732] <211> 23
[0733] <212> DNA
[0734] <213> 人工序列
[0735] <220>
[0736] <223> 靶向探针
[0737] <400> 32
[0738] ccagtccaga cctggccttc tgg 23
[0739] <210> 33
[0740] <211> 23
[0741] <212> DNA
[0742] <213> 人工序列
[0743] <220>
[0744] <223> 靶向探针
[0745] <400> 33
[0746] cccagaaaaa cagaccctga agg 23
[0747] <210> 34
[0748] <211> 23
[0749] <212> DNA
[0750] <213> 人工序列
[0751] <220>
[0752] <223> 靶向探针
[0753] <400> 34
[0754] aagggttgag cacttgttta ggg 23
[0755] <210> 35

[0756] <211> 23
[0757] <212> DNA
[0758] <213> 人工序列
[0759] <220>
[0760] <223> 靶向探针
[0761] <400> 35
[0762] atgtctgagt tttggttgag agg 23
[0763] <210> 36
[0764] <211> 23
[0765] <212> DNA
[0766] <213> 人工序列
[0767] <220>
[0768] <223> 靶向探针
[0769] <400> 36
[0770] ggtcccttga aggggaagta ggg 23
[0771] <210> 37
[0772] <211> 23
[0773] <212> DNA
[0774] <213> 人工序列
[0775] <220>
[0776] <223> 靶向探针
[0777] <400> 37
[0778] tggcagtcta ctcttgaaga tgg 23
[0779] <210> 38
[0780] <211> 23
[0781] <212> DNA
[0782] <213> 人工序列
[0783] <220>
[0784] <223> 靶向探针
[0785] <400> 38
[0786] ggcacagtgc cagaggtctg tgg 23
[0787] <210> 39
[0788] <211> 23
[0789] <212> DNA
[0790] <213> 人工序列
[0791] <220>
[0792] <223> 靶向探针
[0793] <400> 39
[0794] taaaaataaa aaaactaaca ggg 23
[0795] <210> 40
[0796] <211> 23
[0797] <212> DNA

[0798] <213> 人工序列
[0799] <220>
[0800] <223> 靶向探针
[0801] <400> 40
[0802] tctgtggggg acctgcactg agg 23
[0803] <210> 41
[0804] <211> 23
[0805] <212> DNA
[0806] <213> 人工序列
[0807] <220>
[0808] <223> 靶向探针
[0809] <400> 41
[0810] ggccagaggt caaggctagt ggg 23
[0811] <210> 42
[0812] <211> 23
[0813] <212> DNA
[0814] <213> 人工序列
[0815] <220>
[0816] <223> 靶向探针
[0817] <400> 42
[0818] cacgaccgaa acccttctta cgg 23
[0819] <210> 43
[0820] <211> 23
[0821] <212> DNA
[0822] <213> 人工序列
[0823] <220>
[0824] <223> 靶向探针
[0825] <400> 43
[0826] gttgaatgaa gacagtctag tgg 23
[0827] <210> 44
[0828] <211> 23
[0829] <212> DNA
[0830] <213> 人工序列
[0831] <220>
[0832] <223> 靶向探针
[0833] <400> 44
[0834] taagaacaga gcaagttacg tgg 23
[0835] <210> 45
[0836] <211> 23
[0837] <212> DNA
[0838] <213> 人工序列
[0839] <220>

[0840] <223> 靶向探针
[0841] <400> 45
[0842] tgtaaggtaa gagaggagag cgg 23
[0843] <210> 46
[0844] <211> 23
[0845] <212> DNA
[0846] <213> 人工序列
[0847] <220>
[0848] <223> 靶向探针
[0849] <400> 46
[0850] tgacacacca actcctgcac tgg 23
[0851] <210> 47
[0852] <211> 23
[0853] <212> DNA
[0854] <213> 人工序列
[0855] <220>
[0856] <223> 靶向探针
[0857] <400> 47
[0858] tttacccact tccttcgaaa agg 23
[0859] <210> 48
[0860] <211> 23
[0861] <212> DNA
[0862] <213> 人工序列
[0863] <220>
[0864] <223> 靶向探针
[0865] <400> 48
[0866] gtggctggca ggctggctct ggg 23
[0867] <210> 49
[0868] <211> 23
[0869] <212> DNA
[0870] <213> 人工序列
[0871] <220>
[0872] <223> 靶向探针
[0873] <400> 49
[0874] ctcccccggc ctcccccgcg cgg 23
[0875] <210> 50
[0876] <211> 23
[0877] <212> DNA
[0878] <213> 人工序列
[0879] <220>
[0880] <223> 靶向探针
[0881] <400> 50

[0882] caaaacccgg cagcgaggct ggg 23
[0883] <210> 51
[0884] <211> 23
[0885] <212> DNA
[0886] <213> 人工序列
[0887] <220>
[0888] <223> 靶向探针
[0889] <400> 51
[0890] aggagccgcc gcgcgctgat tgg 23
[0891] <210> 52
[0892] <211> 23
[0893] <212> DNA
[0894] <213> 人工序列
[0895] <220>
[0896] <223> 靶向探针
[0897] <400> 52
[0898] cacacacacc cacacgagat ggg 23
[0899] <210> 53
[0900] <211> 23
[0901] <212> DNA
[0902] <213> 人工序列
[0903] <220>
[0904] <223> 靶向探针
[0905] <400> 53
[0906] gaagaagcta aagagccaga ggg 23
[0907] <210> 54
[0908] <211> 23
[0909] <212> DNA
[0910] <213> 人工序列
[0911] <220>
[0912] <223> 靶向探针
[0913] <400> 54
[0914] atgagaattt caataacctc agg 23
[0915] <210> 55
[0916] <211> 23
[0917] <212> DNA
[0918] <213> 人工序列
[0919] <220>
[0920] <223> 靶向探针
[0921] <400> 55
[0922] tcccgctctg ttgcccaggc tgg 23
[0923] <210> 56

[0924] <211> 23
[0925] <212> DNA
[0926] <213> 人工序列
[0927] <220>
[0928] <223> 靶向探针
[0929] <400> 56
[0930] cagacaccca ccaccatgcg tgg 23
[0931] <210> 57
[0932] <211> 23
[0933] <212> DNA
[0934] <213> 人工序列
[0935] <220>
[0936] <223> 靶向探针
[0937] <400> 57
[0938] tcccaattta ctgggattac agg 23
[0939] <210> 58
[0940] <211> 23
[0941] <212> DNA
[0942] <213> 人工序列
[0943] <220>
[0944] <223> 靶向探针
[0945] <400> 58
[0946] tgatttaaaa gttggaaacg tgg 23
[0947] <210> 59
[0948] <211> 23
[0949] <212> DNA
[0950] <213> 人工序列
[0951] <220>
[0952] <223> 靶向探针
[0953] <400> 59
[0954] tctagttccc cacctagtct ggg 23
[0955] <210> 60
[0956] <211> 23
[0957] <212> DNA
[0958] <213> 人工序列
[0959] <220>
[0960] <223> 靶向探针
[0961] <400> 60
[0962] gattaactga gaattcacaa ggg 23
[0963] <210> 61
[0964] <211> 23
[0965] <212> DNA

[0966] <213> 人工序列
[0967] <220>
[0968] <223> 靶向探针
[0969] <400> 61
[0970] cgccaggagg ggtgggtcta agg 23
[0971] <210> 62
[0972] <211> 23
[0973] <212> DNA
[0974] <213> 人工序列
[0975] <220>
[0976] <223> 报告子构建体
[0977] <400> 62
[0978] gtcccctcca ccccacagtg ggg 23
[0979] <210> 63
[0980] <211> 23
[0981] <212> DNA
[0982] <213> 人工序列
[0983] <220>
[0984] <223> 报告子构建体
[0985] <400> 63
[0986] ggggccacta gggacaggat tgg 23
[0987] <210> 64
[0988] <211> 71
[0989] <212> DNA
[0990] <213> 人工序列
[0991] <220>
[0992] <223> 靶标寡核苷酸序列
[0993] <400> 64
[0994] taatactttt atctgtcccc tccaccccac agtggggcca ctagggacag gattggtgac 60
[0995] agaaaagccc c 71
[0996] <210> 65
[0997] <211> 20
[0998] <212> DNA
[0999] <213> 人工序列
[1000] <220>
[1001] <223> 靶标寡核苷酸序列
[1002] <400> 65
[1003] ggggccacta gggacaggat 20
[1004] <210> 66
[1005] <211> 80
[1006] <212> RNA
[1007] <213> 人工序列

[1008] <220>
[1009] <223> 向导RNA
[1010] <400> 66
[1011] guuuuagagc uagaaauagc aaguuaaaau aaggcuagcu uguuaucaac uugaaaaagu 60
[1012] ggcaccgagu cggugcuuuu 80
[1013] <210> 67
[1014] <211> 23
[1015] <212> DNA
[1016] <213> 人工序列
[1017] <220>
[1018] <223> 靶标寡核苷酸序列
[1019] <400> 67
[1020] gtcccctcca ccccacagtg cag 23
[1021] <210> 68
[1022] <211> 23
[1023] <212> DNA
[1024] <213> 人工序列
[1025] <220>
[1026] <223> 靶标寡核苷酸序列
[1027] <400> 68
[1028] gtcccctcca ccccacagtg caa 23
[1029] <210> 69
[1030] <211> 23
[1031] <212> DNA
[1032] <213> 人工序列
[1033] <220>
[1034] <223> 靶标寡核苷酸序列
[1035] <400> 69
[1036] gtcccctcca ccccacagtg cgg 23
[1037] <210> 70
[1038] <211> 52
[1039] <212> DNA
[1040] <213> 人工序列
[1041] <220>
[1042] <223> 靶标寡核苷酸序列
[1043] <400> 70
[1044] tgtcccctcc accccacagt ggggccacta gggacaggat tggtgacaga aa 52
[1045] <210> 71
[1046] <211> 52
[1047] <212> DNA
[1048] <213> 人工序列
[1049] <220>

[1050] <223> 靶标寡核苷酸序列
[1051] <400> 71
[1052] tgtccccccc accccacagt ggggccacta gggacaggat tggtgacaga aa 52
[1053] <210> 72
[1054] <211> 52
[1055] <212> DNA
[1056] <213> 人工序列
[1057] <220>
[1058] <223> 靶标寡核苷酸序列
[1059] <400> 72
[1060] aaaaccctcc accccacagt ggggccacta gggacaggat tggtgacaga aa 52
[1061] <210> 73
[1062] <211> 52
[1063] <212> DNA
[1064] <213> 人工序列
[1065] <220>
[1066] <223> 靶标寡核苷酸序列
[1067] <400> 73
[1068] tgtcccctcc ttttttcagt ggggccacta gggacaggat tggtgacaga aa 52
[1069] <210> 74
[1070] <211> 23
[1071] <212> DNA
[1072] <213> 人工序列
[1073] <220>
[1074] <223> 靶标寡核苷酸序列
[1075] <400> 74
[1076] caccggggtg gtgcccatcc tgg 23
[1077] <210> 75
[1078] <211> 23
[1079] <212> DNA
[1080] <213> 人工序列
[1081] <220>
[1082] <223> 靶标寡核苷酸序列
[1083] <400> 75
[1084] ggtgcccatc ctggtcgagc tgg 23
[1085] <210> 76
[1086] <211> 23
[1087] <212> DNA
[1088] <213> 人工序列
[1089] <220>
[1090] <223> 靶标寡核苷酸序列
[1091] <400> 76

[1092] cccatcctgg tcgagctgga cgg 23
[1093] <210> 77
[1094] <211> 23
[1095] <212> DNA
[1096] <213> 人工序列
[1097] <220>
[1098] <223> 靶标寡核苷酸序列
[1099] <400> 77
[1100] ggccacaagt tcagcgtgtc cgg 23
[1101] <210> 78
[1102] <211> 23
[1103] <212> DNA
[1104] <213> 人工序列
[1105] <220>
[1106] <223> 靶标寡核苷酸序列
[1107] <400> 78
[1108] cgcaaataag agctcaccta cgg 23
[1109] <210> 79
[1110] <211> 23
[1111] <212> DNA
[1112] <213> 人工序列
[1113] <220>
[1114] <223> 靶标寡核苷酸序列
[1115] <400> 79
[1116] ctgaagttca tctgcaccac cgg 23
[1117] <210> 80
[1118] <211> 23
[1119] <212> DNA
[1120] <213> 人工序列
[1121] <220>
[1122] <223> 靶标寡核苷酸序列
[1123] <400> 80
[1124] ccggcaagct gcccgtgccc tgg 23
[1125] <210> 81
[1126] <211> 23
[1127] <212> DNA
[1128] <213> 人工序列
[1129] <220>
[1130] <223> 靶标寡核苷酸序列
[1131] <400> 81
[1132] gaccaggatg ggcaccaccc cgg 23
[1133] <210> 82

[1134] <211> 23
[1135] <212> DNA
[1136] <213> 人工序列
[1137] <220>
[1138] <223> 靶标寡核苷酸序列
[1139] <400> 82
[1140] gccgtccagc tcgaccagga tgg 23
[1141] <210> 83
[1142] <211> 23
[1143] <212> DNA
[1144] <213> 人工序列
[1145] <220>
[1146] <223> 靶标寡核苷酸序列
[1147] <400> 83
[1148] ggccggacac gctgaacttg tgg 23
[1149] <210> 84
[1150] <211> 23
[1151] <212> DNA
[1152] <213> 人工序列
[1153] <220>
[1154] <223> 靶标寡核苷酸序列
[1155] <400> 84
[1156] taacagggta atgtcgaggc cgg 23
[1157] <210> 85
[1158] <211> 23
[1159] <212> DNA
[1160] <213> 人工序列
[1161] <220>
[1162] <223> 靶标寡核苷酸序列
[1163] <400> 85
[1164] aggtgagctc ttatttgcgt agg 23
[1165] <210> 86
[1166] <211> 23
[1167] <212> DNA
[1168] <213> 人工序列
[1169] <220>
[1170] <223> 靶标寡核苷酸序列
[1171] <400> 86
[1172] cttcagggtc agcttgccgt agg 23
[1173] <210> 87
[1174] <211> 23
[1175] <212> DNA

[1176] <213> 人工序列
[1177] <220>
[1178] <223> 靶标寡核苷酸序列
[1179] <400> 87
[1180] gggcacgggc agcttgccgg tgg 23
[1181] <210> 88
[1182] <211> 23
[1183] <212> DNA
[1184] <213> 人工序列
[1185] <220>
[1186] <223> 靶标寡核苷酸序列
[1187] <400> 88
[1188] gagatgatcg ccccttcttc tgg 23
[1189] <210> 89
[1190] <211> 20
[1191] <212> DNA
[1192] <213> 人工序列
[1193] <220>
[1194] <223> 靶标寡核苷酸序列
[1195] <400> 89
[1196] gagatgatcg ccccttcttc 20
[1197] <210> 90
[1198] <211> 20
[1199] <212> DNA
[1200] <213> 人工序列
[1201] <220>
[1202] <223> 靶标寡核苷酸序列
[1203] <400> 90
[1204] gtgatgaccg gccgttcttc 20
[1205] <210> 91
[1206] <211> 23
[1207] <212> DNA
[1208] <213> 人工序列
[1209] <220>
[1210] <223> 靶标寡核苷酸序列
[1211] <400> 91
[1212] gtcccctcca ccccacagtg ggg 23
[1213] <210> 92
[1214] <211> 23
[1215] <212> DNA
[1216] <213> 人工序列
[1217] <220>

[1218] <223> 靶标寡核苷酸序列
[1219] <400> 92
[1220] gagatgatcg cccgttcttc tgg 23
[1221] <210> 93
[1222] <211> 20
[1223] <212> RNA
[1224] <213> 人工序列
[1225] <220>
[1226] <223> RNA靶标序列
[1227] <400> 93
[1228] guccccucca ccccacagug 20
[1229] <210> 94
[1230] <211> 20
[1231] <212> RNA
[1232] <213> 人工序列
[1233] <220>
[1234] <223> RNA靶标序列
[1235] <400> 94
[1236] guccccucca ccccacaguc 20
[1237] <210> 95
[1238] <211> 20
[1239] <212> RNA
[1240] <213> 人工序列
[1241] <220>
[1242] <223> RNA靶标序列
[1243] <400> 95
[1244] guccccucca ccccacagag 20
[1245] <210> 96
[1246] <211> 20
[1247] <212> RNA
[1248] <213> 人工序列
[1249] <220>
[1250] <223> RNA靶标序列
[1251] <400> 96
[1252] guccccucca ccccacacug 20
[1253] <210> 97
[1254] <211> 20
[1255] <212> RNA
[1256] <213> 人工序列
[1257] <220>
[1258] <223> RNA靶标序列
[1259] <400> 97

[1260] guccccucca ccccacugug 20
[1261] <210> 98
[1262] <211> 20
[1263] <212> RNA
[1264] <213> 人工序列
[1265] <220>
[1266] <223> RNA靶标序列
[1267] <400> 98
[1268] guccccucca ccccagagug 20
[1269] <210> 99
[1270] <211> 20
[1271] <212> RNA
[1272] <213> 人工序列
[1273] <220>
[1274] <223> RNA靶标序列
[1275] <400> 99
[1276] guccccucca ccccucagug 20
[1277] <210> 100
[1278] <211> 20
[1279] <212> RNA
[1280] <213> 人工序列
[1281] <220>
[1282] <223> RNA靶标序列
[1283] <400> 100
[1284] guccccucca cccgacagug 20
[1285] <210> 101
[1286] <211> 20
[1287] <212> RNA
[1288] <213> 人工序列
[1289] <220>
[1290] <223> RNA靶标序列
[1291] <400> 101
[1292] guccccucca ccgcacagug 20
[1293] <210> 102
[1294] <211> 20
[1295] <212> RNA
[1296] <213> 人工序列
[1297] <220>
[1298] <223> RNA靶标序列
[1299] <400> 102
[1300] guccccucca cgccacagug 20
[1301] <210> 103

[1302] <211> 20
[1303] <212> RNA
[1304] <213> 人工序列
[1305] <220>
[1306] <223> RNA靶标序列
[1307] <400> 103
[1308] guccccucca gcccacagug 20
[1309] <210> 104
[1310] <211> 20
[1311] <212> RNA
[1312] <213> 人工序列
[1313] <220>
[1314] <223> RNA靶标序列
[1315] <400> 104
[1316] guccccuccu ccccacagug 20
[1317] <210> 105
[1318] <211> 20
[1319] <212> RNA
[1320] <213> 人工序列
[1321] <220>
[1322] <223> RNA靶标序列
[1323] <400> 105
[1324] guccccucga ccccacagug 20
[1325] <210> 106
[1326] <211> 20
[1327] <212> RNA
[1328] <213> 人工序列
[1329] <220>
[1330] <223> RNA靶标序列
[1331] <400> 106
[1332] guccccucca ccccacagac 20
[1333] <210> 107
[1334] <211> 20
[1335] <212> RNA
[1336] <213> 人工序列
[1337] <220>
[1338] <223> RNA靶标序列
[1339] <400> 107
[1340] guccccucca ccccacucug 20
[1341] <210> 108
[1342] <211> 20
[1343] <212> RNA

[1344] <213> 人工序列
[1345] <220>
[1346] <223> RNA靶标序列
[1347] <400> 108
[1348] guccccucca ccccugagug 20
[1349] <210> 109
[1350] <211> 20
[1351] <212> RNA
[1352] <213> 人工序列
[1353] <220>
[1354] <223> RNA靶标序列
[1355] <400> 109
[1356] guccccucca ccggacagug 20
[1357] <210> 110
[1358] <211> 20
[1359] <212> RNA
[1360] <213> 人工序列
[1361] <220>
[1362] <223> RNA靶标序列
[1363] <400> 110
[1364] guccccucca ggccacagug 20
[1365] <210> 111
[1366] <211> 20
[1367] <212> RNA
[1368] <213> 人工序列
[1369] <220>
[1370] <223> RNA靶标序列
[1371] <400> 111
[1372] guccccucgu ccccacagug 20
[1373] <210> 112
[1374] <211> 23
[1375] <212> DNA
[1376] <213> 人工序列
[1377] <220>
[1378] <223> 靶标寡核苷酸序列
[1379] <400> 112
[1380] ggggccacta gggacaggat ggg 23
[1381] <210> 113
[1382] <211> 20
[1383] <212> RNA
[1384] <213> 人工序列
[1385] <220>

[1386] <223> RNA靶标序列
[1387] <400> 113
[1388] gagaugaucg ccccuucuuc 20
[1389] <210> 114
[1390] <211> 20
[1391] <212> RNA
[1392] <213> 人工序列
[1393] <220>
[1394] <223> RNA靶标序列
[1395] <400> 114
[1396] gagaugaucg ccccuucuug 20
[1397] <210> 115
[1398] <211> 20
[1399] <212> RNA
[1400] <213> 人工序列
[1401] <220>
[1402] <223> RNA靶标序列
[1403] <400> 115
[1404] gagaugaucg ccccuucuac 20
[1405] <210> 116
[1406] <211> 20
[1407] <212> RNA
[1408] <213> 人工序列
[1409] <220>
[1410] <223> RNA靶标序列
[1411] <400> 116
[1412] gagaugaucg ccccuucauc 20
[1413] <210> 117
[1414] <211> 20
[1415] <212> RNA
[1416] <213> 人工序列
[1417] <220>
[1418] <223> RNA靶标序列
[1419] <400> 117
[1420] gagaugaucg ccccuuguuc 20
[1421] <210> 118
[1422] <211> 20
[1423] <212> RNA
[1424] <213> 人工序列
[1425] <220>
[1426] <223> RNA靶标序列
[1427] <400> 118

[1428] gagaugaucg ccccuacuuc 20
[1429] <210> 119
[1430] <211> 20
[1431] <212> RNA
[1432] <213> 人工序列
[1433] <220>
[1434] <223> RNA靶标序列
[1435] <400> 119
[1436] gagaugaucg ccccaucuuc 20
[1437] <210> 120
[1438] <211> 20
[1439] <212> RNA
[1440] <213> 人工序列
[1441] <220>
[1442] <223> RNA靶标序列
[1443] <400> 120
[1444] gagaugaucg cccguucuuc 20
[1445] <210> 121
[1446] <211> 20
[1447] <212> RNA
[1448] <213> 人工序列
[1449] <220>
[1450] <223> RNA靶标序列
[1451] <400> 121
[1452] gagaugaucg ccgcuucuuc 20
[1453] <210> 122
[1454] <211> 20
[1455] <212> RNA
[1456] <213> 人工序列
[1457] <220>
[1458] <223> RNA靶标序列
[1459] <400> 122
[1460] gagaugaucg cgccuucuuc 20
[1461] <210> 123
[1462] <211> 20
[1463] <212> RNA
[1464] <213> 人工序列
[1465] <220>
[1466] <223> RNA靶标序列
[1467] <400> 123
[1468] gagaugaucg gcccuucuuc 20
[1469] <210> 124

[1470] <211> 20
[1471] <212> RNA
[1472] <213> 人工序列
[1473] <220>
[1474] <223> RNA靶标序列
[1475] <400> 124
[1476] gagaugaucc ccccuucuuc 20
[1477] <210> 125
[1478] <211> 20
[1479] <212> RNA
[1480] <213> 人工序列
[1481] <220>
[1482] <223> RNA靶标序列
[1483] <400> 125
[1484] gagaugaugg ccccuucuuc 20
[1485] <210> 126
[1486] <211> 20
[1487] <212> RNA
[1488] <213> 人工序列
[1489] <220>
[1490] <223> RNA靶标序列
[1491] <400> 126
[1492] gagaugaucg ccccuucuag 20
[1493] <210> 127
[1494] <211> 20
[1495] <212> RNA
[1496] <213> 人工序列
[1497] <220>
[1498] <223> RNA靶标序列
[1499] <400> 127
[1500] gagaugaucg ccccuugauc 20
[1501] <210> 128
[1502] <211> 20
[1503] <212> RNA
[1504] <213> 人工序列
[1505] <220>
[1506] <223> RNA靶标序列
[1507] <400> 128
[1508] gagaugaucg ccccaacuuc 20
[1509] <210> 129
[1510] <211> 20
[1511] <212> RNA

[1512] <213> 人工序列
[1513] <220>
[1514] <223> RNA靶标序列
[1515] <400> 129
[1516] gagaugaucg ccgguucuuc 20
[1517] <210> 130
[1518] <211> 20
[1519] <212> RNA
[1520] <213> 人工序列
[1521] <220>
[1522] <223> RNA靶标序列
[1523] <400> 130
[1524] gagaugaucg ggccuucuuc 20
[1525] <210> 131
[1526] <211> 20
[1527] <212> RNA
[1528] <213> 人工序列
[1529] <220>
[1530] <223> RNA靶标序列
[1531] <400> 131
[1532] gagaugaugc ccccuucuuc 20
[1533] <210> 132
[1534] <211> 23
[1535] <212> DNA
[1536] <213> 人工序列
[1537] <220>
[1538] <223> 靶标寡核苷酸序列
[1539] <400> 132
[1540] gagatgatcg ccccttcttc tgg 23
[1541] <210> 133
[1542] <211> 20
[1543] <212> RNA
[1544] <213> 人工序列
[1545] <220>
[1546] <223> RNA靶标序列
[1547] <400> 133
[1548] ggggccacua gggacaggau 20
[1549] <210> 134
[1550] <211> 19
[1551] <212> RNA
[1552] <213> 人工序列
[1553] <220>

[1554] <223> RNA靶标序列
[1555] <400> 134
[1556] gggccacuag ggacaggau 19
[1557] <210> 135
[1558] <211> 18
[1559] <212> RNA
[1560] <213> 人工序列
[1561] <220>
[1562] <223> RNA靶标序列
[1563] <400> 135
[1564] ggccacuagg gacaggau 18
[1565] <210> 136
[1566] <211> 17
[1567] <212> RNA
[1568] <213> 人工序列
[1569] <220>
[1570] <223> RNA靶标序列
[1571] <400> 136
[1572] gccacuaggg acaggau 17
[1573] <210> 137
[1574] <211> 20
[1575] <212> RNA
[1576] <213> 人工序列
[1577] <220>
[1578] <223> RNA靶标序列
[1579] <400> 137
[1580] gagaugaucg ccccuucuuc 20
[1581] <210> 138
[1582] <211> 18
[1583] <212> RNA
[1584] <213> 人工序列
[1585] <220>
[1586] <223> RNA靶标序列
[1587] <400> 138
[1588] gaugaucgcc ccuucuuc 18
[1589] <210> 139
[1590] <211> 15
[1591] <212> RNA
[1592] <213> 人工序列
[1593] <220>
[1594] <223> RNA靶标序列
[1595] <400> 139

[1596] gaucgccccu ucuuc 15
[1597] <210> 140
[1598] <211> 11
[1599] <212> RNA
[1600] <213> 人工序列
[1601] <220>
[1602] <223> RNA靶标序列
[1603] <400> 140
[1604] gccccuucuu c 11
[1605] <210> 141
[1606] <211> 21
[1607] <212> DNA
[1608] <213> 人工序列
[1609] <220>
[1610] <223> 靶标寡核苷酸序列
[1611] <400> 141
[1612] gtcccctcca ccccacagtg c 21
[1613] <210> 142
[1614] <211> 14
[1615] <212> DNA
[1616] <213> 人工序列
[1617] <220>
[1618] <223> 靶标寡核苷酸序列
[1619] <220>
[1620] <221> misc_feature
[1621] <222> (5)..(10)
[1622] <223> 其中N是G, A, T或C
[1623] <400> 142
[1624] tgtcnnnnnn accc 14
[1625] <210> 143
[1626] <211> 14
[1627] <212> DNA
[1628] <213> 人工序列
[1629] <220>
[1630] <223> 靶标寡核苷酸序列
[1631] <400> 143
[1632] tgtcaaaaaa accc 14
[1633] <210> 144
[1634] <211> 14
[1635] <212> DNA
[1636] <213> 人工序列
[1637] <220>

[1638] <223> 靶标寡核苷酸序列
[1639] <400> 144
[1640] tgtcgggggg accc 14
[1641] <210> 145
[1642] <211> 14
[1643] <212> DNA
[1644] <213> 人工序列
[1645] <220>
[1646] <223> 靶标寡核苷酸序列
[1647] <400> 145
[1648] tgtcaaaaaa accc 14
[1649] <210> 146
[1650] <211> 14
[1651] <212> DNA
[1652] <213> 人工序列
[1653] <220>
[1654] <223> 靶标寡核苷酸序列
[1655] <400> 146
[1656] tgtcgggggg accc 14
[1657] <210> 147
[1658] <211> 14
[1659] <212> DNA
[1660] <213> 人工序列
[1661] <220>
[1662] <223> 靶标寡核苷酸序列
[1663] <400> 147
[1664] tgtccccccc accc 14
[1665] <210> 148
[1666] <211> 14
[1667] <212> DNA
[1668] <213> 人工序列
[1669] <220>
[1670] <223> 靶标寡核苷酸序列
[1671] <400> 148
[1672] tgtctttttt accc 14
[1673] <210> 149
[1674] <211> 14
[1675] <212> DNA
[1676] <213> 人工序列
[1677] <220>
[1678] <223> 靶标寡核苷酸序列
[1679] <400> 149

[1680] tgtccccccc accc 14
[1681] <210> 150
[1682] <211> 14
[1683] <212> DNA
[1684] <213> 人工序列
[1685] <220>
[1686] <223> 靶标寡核苷酸序列
[1687] <400> 150
[1688] tgtctttttt accc 14
[1689] <210> 151
[1690] <211> 23
[1691] <212> DNA
[1692] <213> 人工序列
[1693] <220>
[1694] <223> 靶标寡核苷酸序列
[1695] <400> 151
[1696] ggatcctgtg tccccgagct ggg 23
[1697] <210> 152
[1698] <211> 23
[1699] <212> DNA
[1700] <213> 人工序列
[1701] <220>
[1702] <223> 靶标寡核苷酸序列
[1703] <400> 152
[1704] gttaatgtgg ctctggttct ggg 23
[1705] <210> 153
[1706] <211> 23
[1707] <212> DNA
[1708] <213> 人工序列
[1709] <220>
[1710] <223> 靶标寡核苷酸序列
[1711] <400> 153
[1712] ggggccacta gggacaggat tgg 23
[1713] <210> 154
[1714] <211> 23
[1715] <212> DNA
[1716] <213> 人工序列
[1717] <220>
[1718] <223> 靶标寡核苷酸序列
[1719] <400> 154
[1720] cttcctagtc tcctgatatt ggg 23
[1721] <210> 155

[1722] <211> 23
[1723] <212> DNA
[1724] <213> 人工序列
[1725] <220>
[1726] <223> 靶标寡核苷酸序列
[1727] <400> 155
[1728] tggtcccagc tcggggacac agg 23
[1729] <210> 156
[1730] <211> 23
[1731] <212> DNA
[1732] <213> 人工序列
[1733] <220>
[1734] <223> 靶标寡核苷酸序列
[1735] <400> 156
[1736] agaaccagag ccacattaac cgg 23
[1737] <210> 157
[1738] <211> 23
[1739] <212> DNA
[1740] <213> 人工序列
[1741] <220>
[1742] <223> 靶标寡核苷酸序列
[1743] <400> 157
[1744] gtcaccaatc ctgtccctag tgg 23
[1745] <210> 158
[1746] <211> 23
[1747] <212> DNA
[1748] <213> 人工序列
[1749] <220>
[1750] <223> 靶标寡核苷酸序列
[1751] <400> 158
[1752] agacccaata tcaggagact agg 23
[1753] <210> 159
[1754] <211> 75
[1755] <212> DNA
[1756] <213> 人工序列
[1757] <220>
[1758] <223> 靶标寡核苷酸序列
[1759] <400> 159
[1760] gggatcctgt gtccccgagc tgggaccacc ttatattccc agggccggtt aatgtggctc 60
[1761] tggttctggg tactt 75
[1762] <210> 160
[1763] <211> 69

[1764] <212> DNA
[1765] <213> 人工序列
[1766] <220>
[1767] <223> 靶标寡核苷酸序列
[1768] <400> 160
[1769] gggatcctgt gtccccgagc tgggaccacc ttatattccc agggccggtt aatgtggttc 60
[1770] tgggtactt 69
[1771] <210> 161
[1772] <211> 113
[1773] <212> DNA
[1774] <213> 人工序列
[1775] <220>
[1776] <223> 靶标寡核苷酸序列
[1777] <400> 161
[1778] gggatcctgt gtccccgagc tgggaccacc ttatattccc agggcagggc cggttggacc 60
[1779] accttatatt cccagggcag ggccggttaa tgtggctctg gttctgggta ctt 113
[1780] <210> 162
[1781] <211> 34
[1782] <212> DNA
[1783] <213> 人工序列
[1784] <220>
[1785] <223> 靶标寡核苷酸序列
[1786] <400> 162
[1787] gggatcctgt gtccccgtct ggttctgggt actt 34
[1788] <210> 163
[1789] <211> 47
[1790] <212> DNA
[1791] <213> 人工序列
[1792] <220>
[1793] <223> 靶标寡核苷酸序列
[1794] <400> 163
[1795] gggatcctgt gtccccgagc tgggaccacc ttatattctg ggtactt 47
[1796] <210> 164
[1797] <211> 17
[1798] <212> DNA
[1799] <213> 人工序列
[1800] <220>
[1801] <223> 靶标寡核苷酸序列
[1802] <400> 164
[1803] gggatcctgt ggtactt 17
[1804] <210> 165
[1805] <211> 93

[1806] <212> DNA
[1807] <213> 人工序列
[1808] <220>
[1809] <223> 靶标寡核苷酸序列
[1810] <400> 165
[1811] agggccggtt aatgtggctc tggttctggg tacttttatc tgtcccctcc accccacagt 60
[1812] ggggccacta gggacaggat tggtgacaga aaa 93
[1813] <210> 166
[1814] <211> 83
[1815] <212> DNA
[1816] <213> 人工序列
[1817] <220>
[1818] <223> 靶标寡核苷酸序列
[1819] <400> 166
[1820] agggccggtt aatgaatgtg gctctggttc tgggtacttt tatctgtccc ctccacccca 60
[1821] cagtggggcc actagacaga aaa 83
[1822] <210> 167
[1823] <211> 76
[1824] <212> DNA
[1825] <213> 人工序列
[1826] <220>
[1827] <223> 靶标寡核苷酸序列
[1828] <400> 167
[1829] agggccggtt aatgtggctc tggttctggg tacttttatc tgtcccccag tggggccact 60
[1830] gattggtgac agaaaa 76
[1831] <210> 168
[1832] <211> 29
[1833] <212> DNA
[1834] <213> 人工序列
[1835] <220>
[1836] <223> 靶标寡核苷酸序列
[1837] <400> 168
[1838] agggccggtt caggattggt gacagaaaa 29
[1839] <210> 169
[1840] <211> 34
[1841] <212> DNA
[1842] <213> 人工序列
[1843] <220>
[1844] <223> 靶标寡核苷酸序列
[1845] <400> 169
[1846] agggccggtt aatgtggcga ttggtgacag aaaa 34
[1847] <210> 170

[1848] <211> 63
[1849] <212> DNA
[1850] <213> 人工序列
[1851] <220>
[1852] <223> 靶标寡核苷酸序列
[1853] <400> 170
[1854] agggccggtt aatgtggctc tggttctggg tacttttatc tgtccccgat tggtgacaga 60
[1855] aaa 63
[1856] <210> 171
[1857] <211> 84
[1858] <212> DNA
[1859] <213> 人工序列
[1860] <220>
[1861] <223> 靶标寡核苷酸序列
[1862] <400> 171
[1863] agggccggtt aatgtggctc tggttctggg tacttttatc tgtcccctcc accccacagt 60
[1864] ggggacagga ttggtgacag aaaa 84
[1865] <210> 172
[1866] <211> 27
[1867] <212> DNA
[1868] <213> 人工序列
[1869] <220>
[1870] <223> 靶标寡核苷酸序列
[1871] <400> 172
[1872] agggccggtt aatgtggtga cagaaaa 27
[1873] <210> 173
[1874] <211> 105
[1875] <212> DNA
[1876] <213> 人工序列
[1877] <220>
[1878] <223> 靶标寡核苷酸序列
[1879] <400> 173
[1880] agggccggtt aatgtggctc tggttctggg tacttttatc tgtcccctcc accccagggg 60
[1881] acagtctgtc ccctccaccc cagggacagg attggtgaca gaaaa 105
[1882] <210> 174
[1883] <211> 80
[1884] <212> DNA
[1885] <213> 人工序列
[1886] <220>
[1887] <223> 靶标寡核苷酸序列
[1888] <400> 174
[1889] agggccggtt aatgtggctc tggttctggg tacttttatc tgtcccctcc accactaggg 60

[1890] acaggattgg tgacagaaaa 80
[1891] <210> 175
[1892] <211> 53
[1893] <212> DNA
[1894] <213> 人工序列
[1895] <220>
[1896] <223> 靶标寡核苷酸序列
[1897] <400> 175
[1898] cccacagtgg ggccactagg gacaggattg gtgacagaaa agccccatac ccc 53
[1899] <210> 176
[1900] <211> 22
[1901] <212> DNA
[1902] <213> 人工序列
[1903] <220>
[1904] <223> 靶标寡核苷酸序列
[1905] <400> 176
[1906] cccacagtgg ggccactacc cc 22
[1907] <210> 177
[1908] <211> 96
[1909] <212> DNA
[1910] <213> 人工序列
[1911] <220>
[1912] <223> 靶标寡核苷酸序列
[1913] <400> 177
[1914] cccacagtgg ggccactagt agaaaagccc catccttagg cctcccccat ccttaggcct 60
[1915] cctccttcct agtctcctga tattgggtct aacccc 96
[1916] <210> 178
[1917] <211> 94
[1918] <212> DNA
[1919] <213> 人工序列
[1920] <220>
[1921] <223> 靶标寡核苷酸序列
[1922] <400> 178
[1923] cccacagtgg ggccactagg gacaggattg gtgacagaaa agccccatcc ttaggcctcc 60
[1924] tccttcctag tctcctgata ttgggtctaa cccc 94
[1925] <210> 179
[1926] <211> 62
[1927] <212> DNA
[1928] <213> 人工序列
[1929] <220>
[1930] <223> 靶标寡核苷酸序列
[1931] <400> 179

[1932] cccacagtgg ggccaccctt aggcctcctc cttcctagtc tcctgatatt gggtctaacc 60
[1933] cc 62
[1934] <210> 180
[1935] <211> 38
[1936] <212> DNA
[1937] <213> 人工序列
[1938] <220>
[1939] <223> 靶标寡核苷酸序列
[1940] <400> 180
[1941] cccacagtgg ggccactagt gatattgggt ctaacccc 38
[1942] <210> 181
[1943] <211> 94
[1944] <212> DNA
[1945] <213> 人工序列
[1946] <220>
[1947] <223> 靶标寡核苷酸序列
[1948] <400> 181
[1949] cccacagtgg ggccactagg gacaggattg gtgacaaaaa agccccatcc ttacgcctcc 60
[1950] tccttcctag tctcctgata ttgggtctaa cccc 94
[1951] <210> 182
[1952] <211> 65
[1953] <212> DNA
[1954] <213> 人工序列
[1955] <220>
[1956] <223> 靶标寡核苷酸序列
[1957] <400> 182
[1958] cccacagtgg ggccactagg gacaggcctc ctccttccta gtctcctgat attgggtcta 60
[1959] acccc 65
[1960] <210> 183
[1961] <211> 102
[1962] <212> DNA
[1963] <213> 人工序列
[1964] <220>
[1965] <223> 靶标寡核苷酸序列
[1966] <400> 183
[1967] cccacagtgg ggccactagg gacaggggga caggattggt gacagaaaag ccccatcctt 60
[1968] aggcctcctc cttcctagtc tcctgatatt gggtctaacc cc 102
[1969] <210> 184
[1970] <211> 76
[1971] <212> DNA
[1972] <213> 人工序列
[1973] <220>

[1974] <223> 靶标寡核苷酸序列

[1975] <400> 184

[1976] cccacaggat tggtgacaga aaagccccat ccttaggcct cctccttcct agtctcctga 60

[1977] tattgggtct aacccc 76

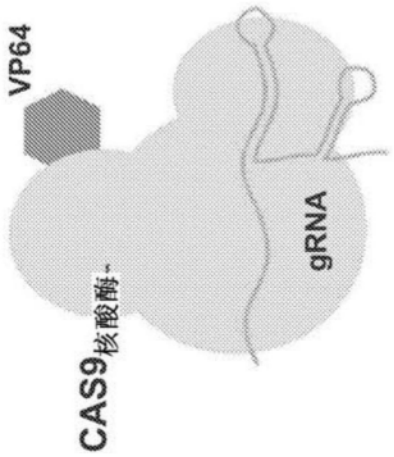


图1A

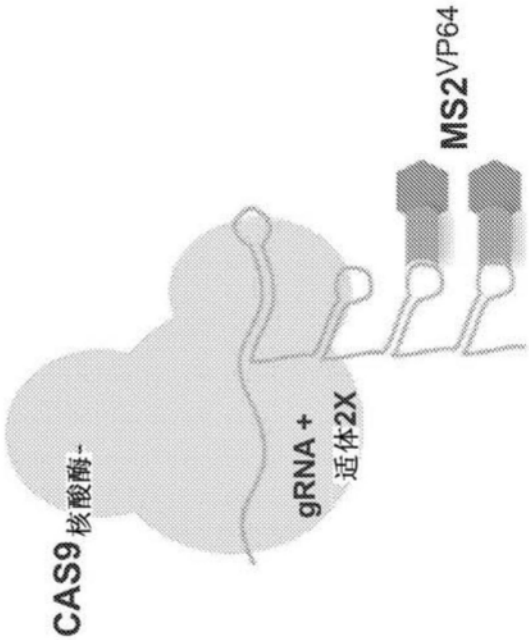


图1B

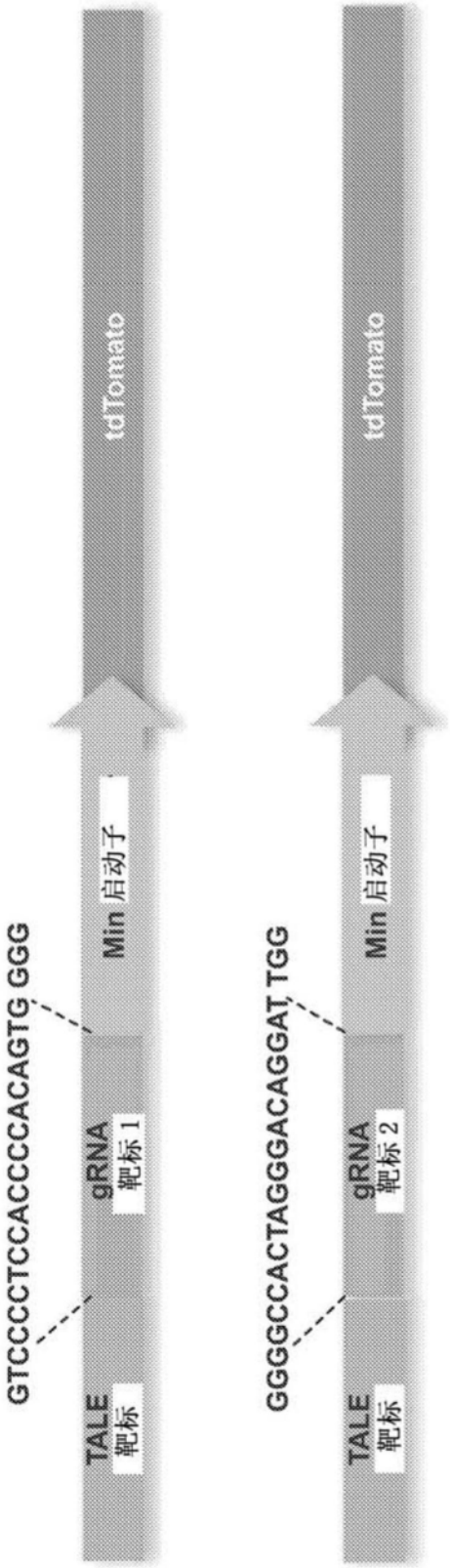

TALE
靶标
GTCCCCTCCACCCCACAGTG GGG
gRNA
靶标 1
Min 启动子
tdTomato
TALE
靶标
GGGGCCACTAGGGACAGGAT TGG
gRNA
靶标 2
Min 启动子
tdTomato

图1C

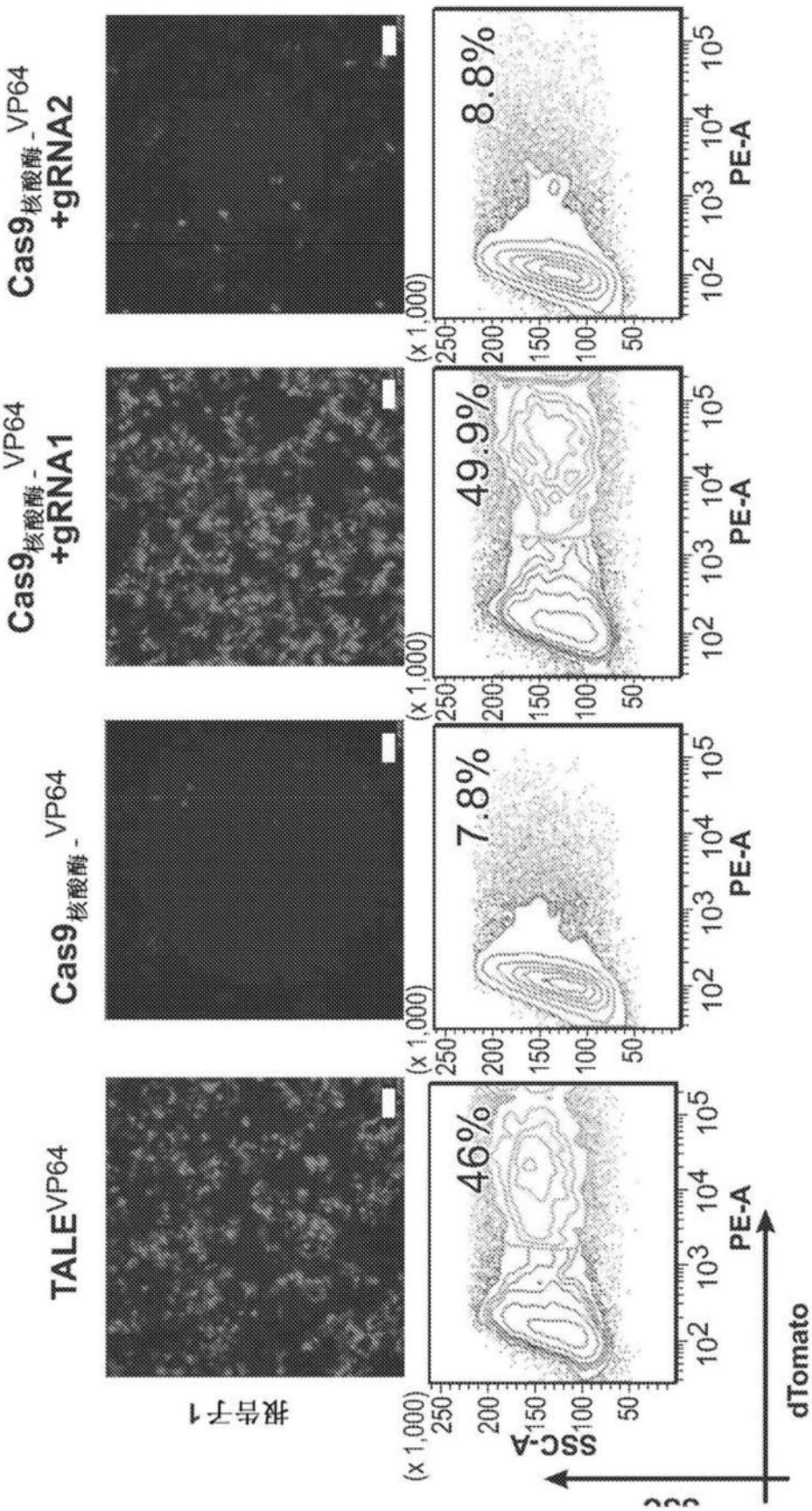

TALE^VP64
Cas9核酸酶-VP64
Cas9核酸酶-VP64 +gRNA1
Cas9核酸酶-VP64 +gRNA2
靶标子1
46%
7.8%
49.9%
8.8%
(x 1,000)
SSC-A
PE-A
dTomato

图1D-1

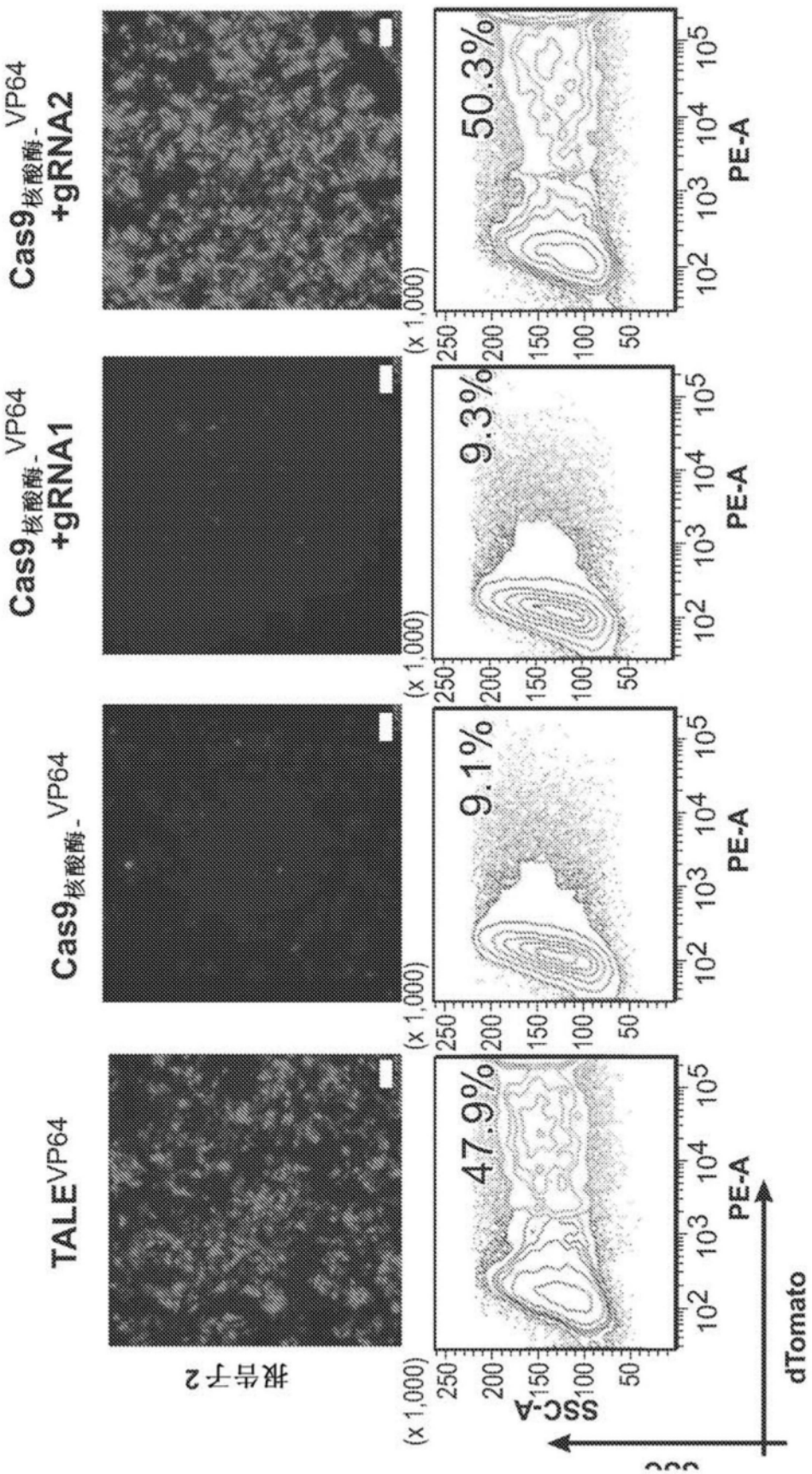

TALE VP64
Cas9核酸酶-VP64
Cas9核酸酶-VP64 +gRNA1
Cas9核酸酶-VP64 +gRNA2
报告子2
47.9%
9.1%
9.3%
50.3%
(x 1,000)
SSC-A
PE-A
dTomato

图1D-2

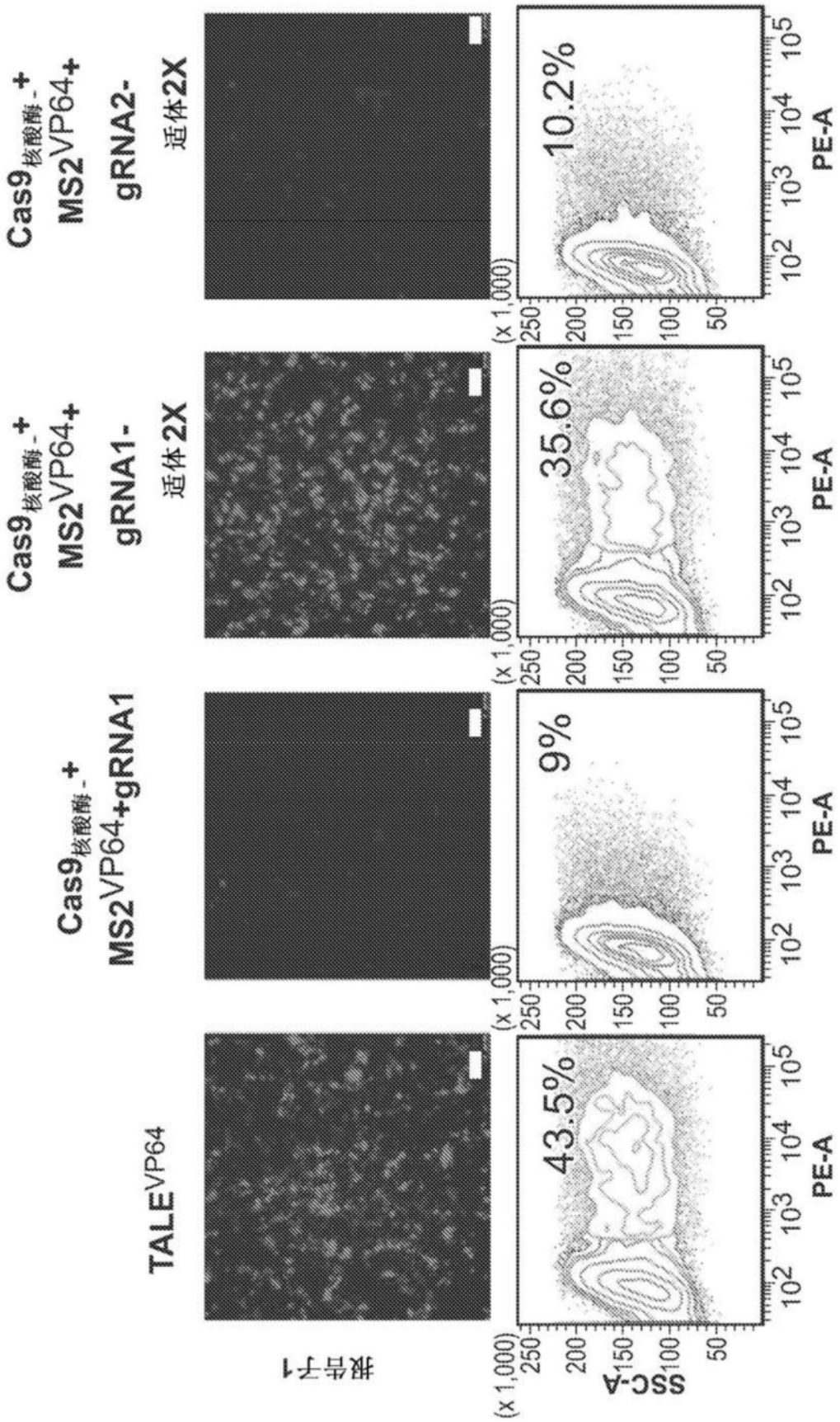

报告子1
TALE^VP64
Cas9核酸酶-+
MS2VP64+gRNA1
Cas9核酸酶-+
MS2VP64+
gRNA1-
适体2X
Cas9核酸酶-+
MS2VP64+
gRNA2-
适体2X
43.5%
9%
35.6%
10.2%
SSC-A
PE-A
(x 1,000)
250
200
150
100
50
10^2
10^3
10^4
10^5

图1E-1

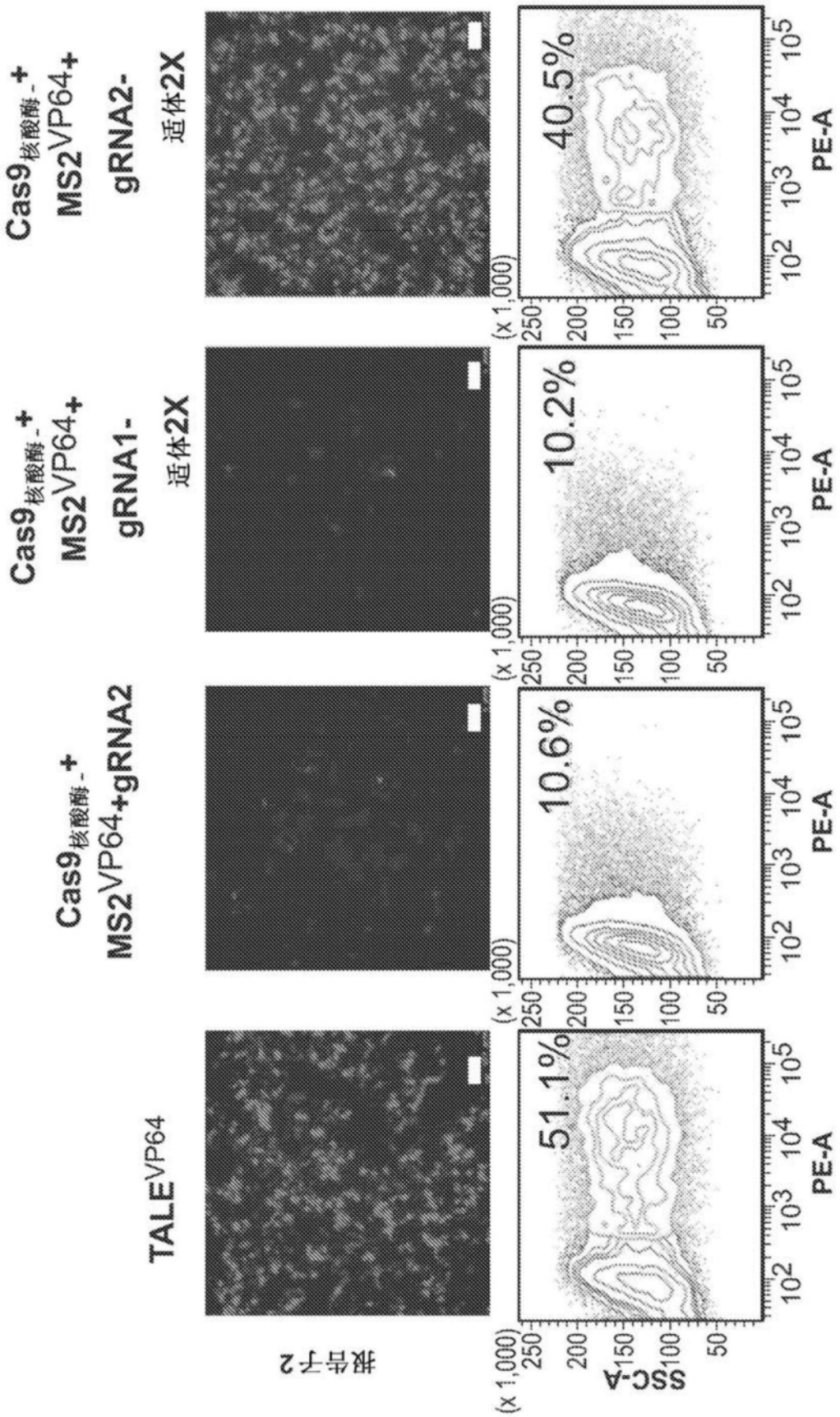

TALE^VP64
Cas9核酸酶-+ MS2^VP64+gRNA2
Cas9核酸酶-+ MS2^VP64+ gRNA1- 适体2X
Cas9核酸酶-+ MS2^VP64+ gRNA2- 适体2X
报告子2
51.1%
10.6%
10.2%
40.5%
SSC-A
PE-A
(x 1,000)

图1E-2

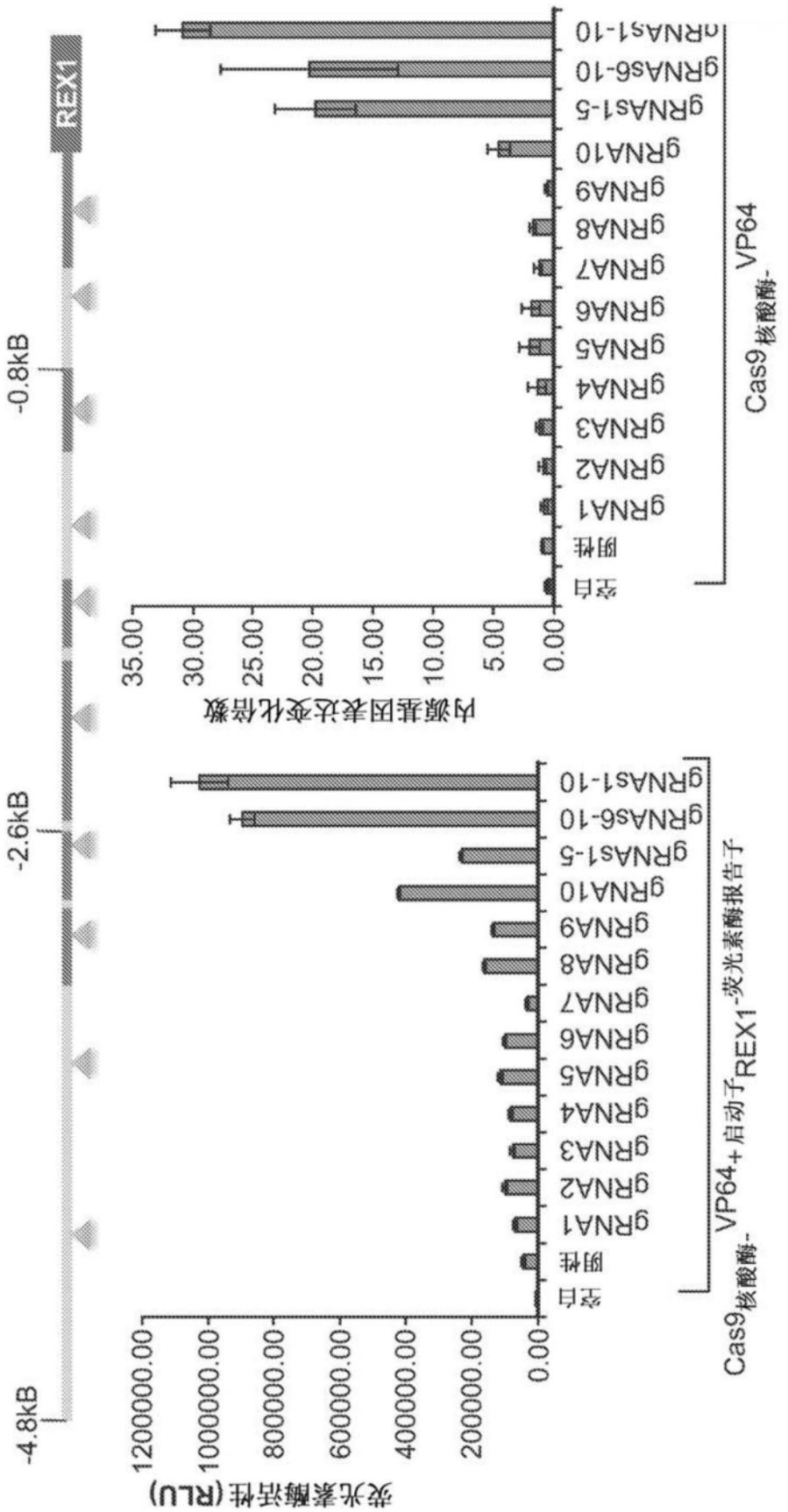

-4.8kB
-2.6kB
-0.8kB
REX1
荧光素酶活性(RLU)
1200000.00
1000000.00
800000.00
600000.00
400000.00
200000.00
0.00
空白
阴性
gRNA1
gRNA2
gRNA3
gRNA4
gRNA5
gRNA6
gRNA7
gRNA8
gRNA9
gRNA10
gRNAs1-5
gRNAs6-10
gRNAs1-10
Cas9核酸酶-VP64+启动子REX1-荧光素酶报告子
内源基因表达变化倍数
35.00
30.00
25.00
20.00
15.00
10.00
5.00
0.00
空白
阴性
gRNA1
gRNA2
gRNA3
gRNA4
gRNA5
gRNA6
gRNA7
gRNA8
gRNA9
gRNA10
gRNAs1-5
gRNAs6-10
gRNAs1-10
Cas9核酸酶-VP64

图1F

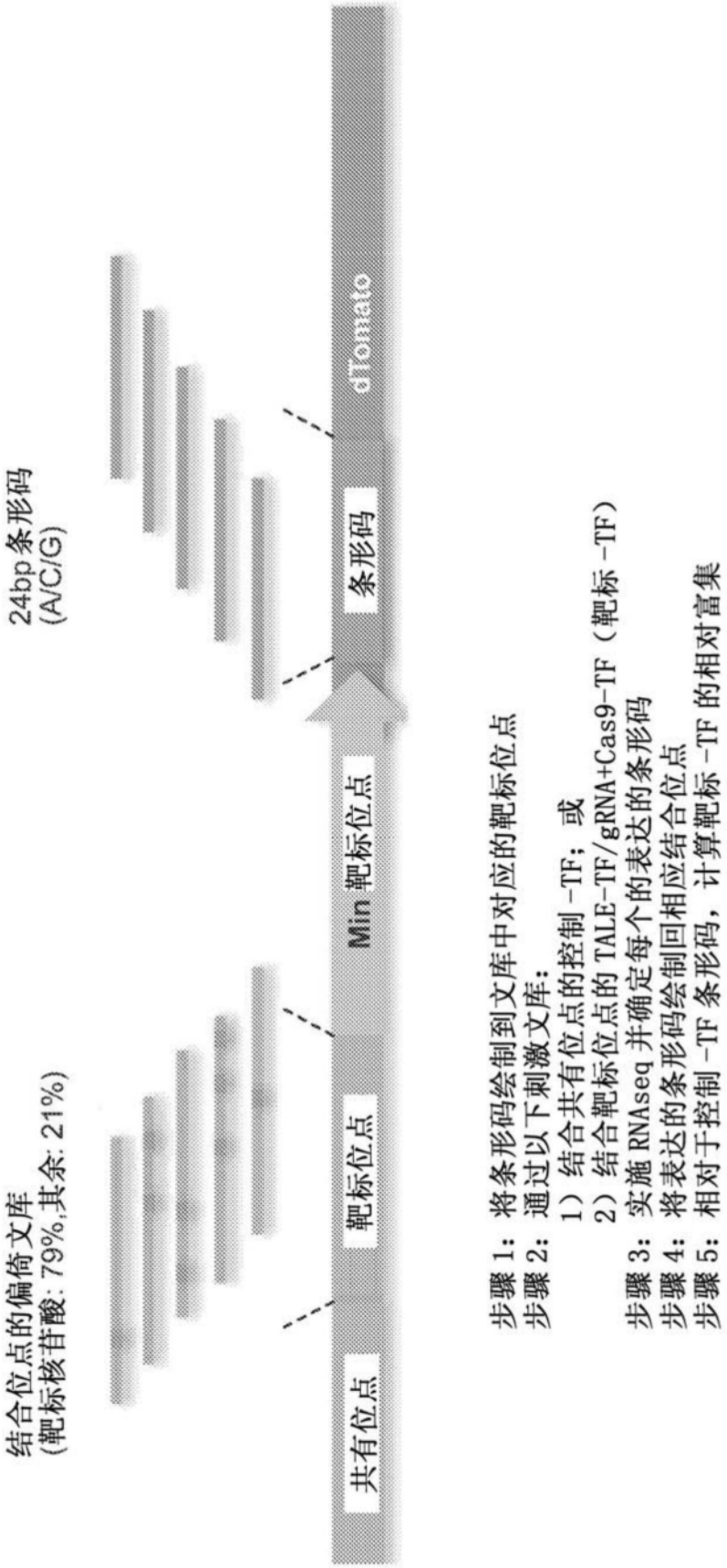

结合位点的偏倚文库
(靶标核苷酸: 79%,其余: 21%)
24bp条形码
(A/C/G)
共有位点
靶标位点
Min靶标位点
条形码
dTomato
步骤 1：将条形码绘制到文库中对应的靶标位点
步骤 2：通过以下刺激文库：
1）结合共有位点的控制 -TF；或
2）结合靶标位点的 TALE-TF/gRNA+Cas9-TF（靶标 -TF）
步骤 3：实施 RNAseq 并确定每个的表达的条形码
步骤 4：将表达的条形码绘制回相应结合位点
步骤 5：相对于控制 -TF 条形码，计算靶标 -TF 的相对富集

图2A

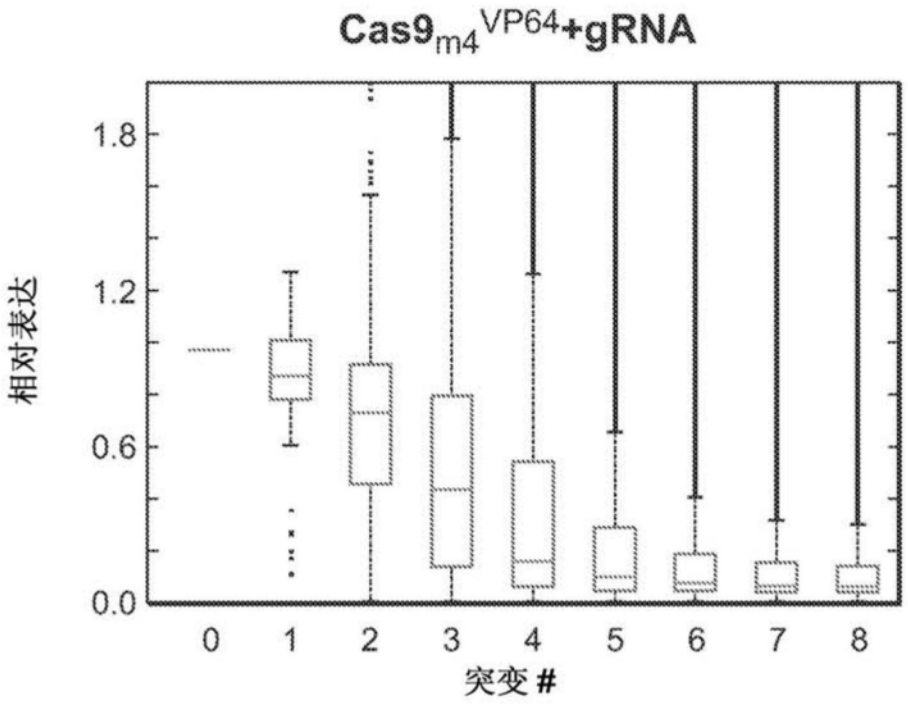

Cas9m4VP64+gRNA
相对表达
1.8
1.2
0.6
0.0
0
1
2
3
4
5
6
7
8
突变#

图2B

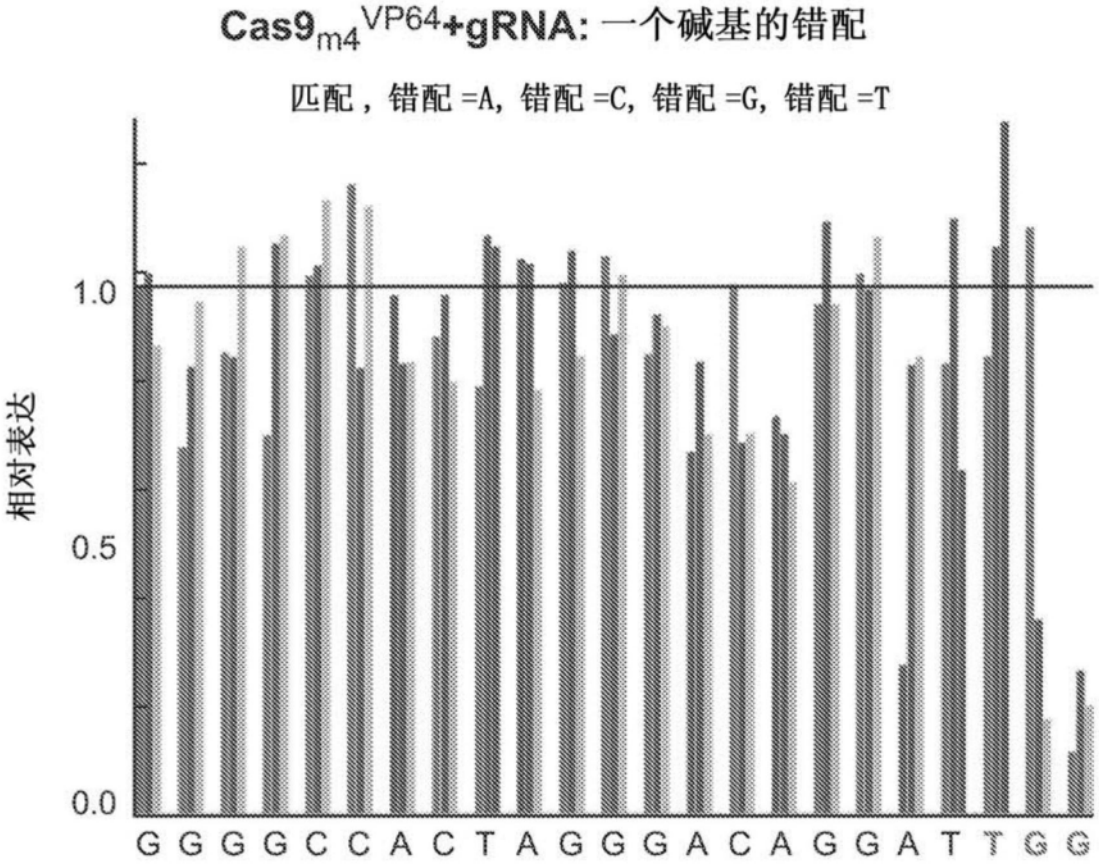

Cas9m4VP64+gRNA: 一个碱基的错配
匹配，错配 =A，错配 =C，错配 =G，错配 =T
相对表达
1.0
0.5
0.0
G G G G C C A C T A G G G A C A G G A T T G G

图2C

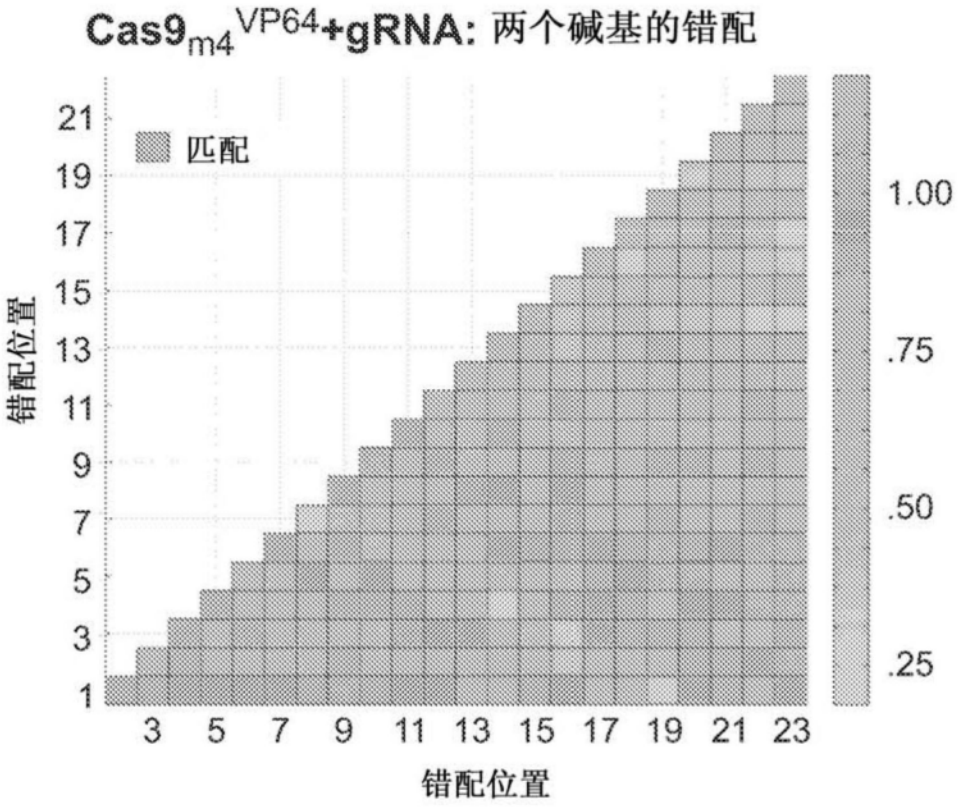

Cas9m4VP64+gRNA: 两个碱基的错配
匹配
错配位置
21
19
17
15
13
11
9
7
5
3
1
3 5 7 9 11 13 15 17 19 21 23
错配位置
1.00
.75
.50
.25

图2D

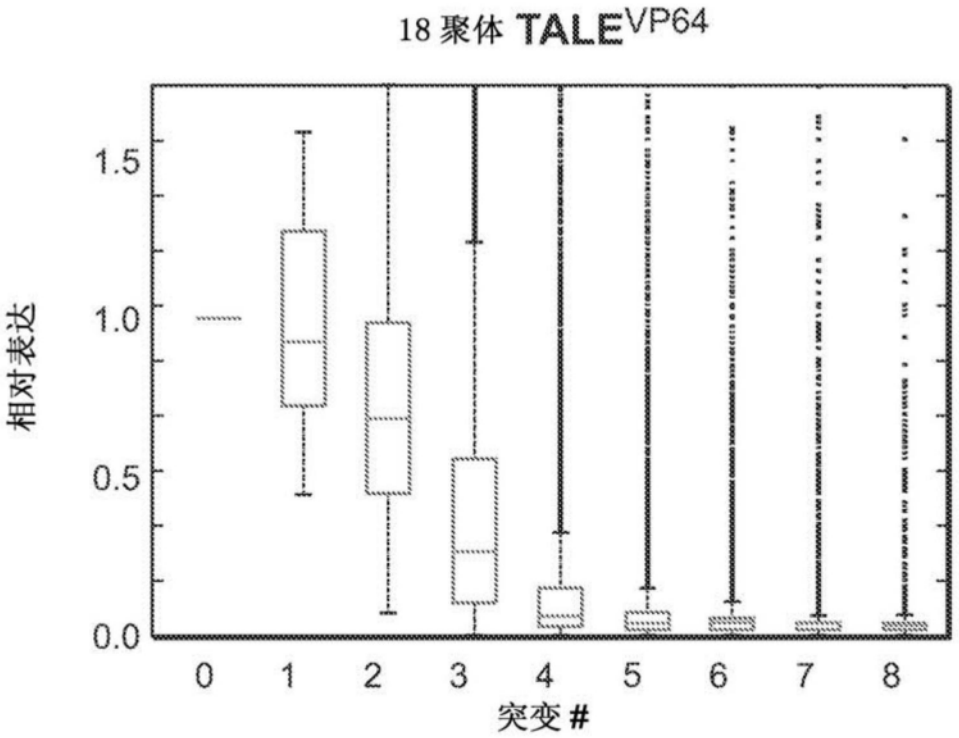

18 聚体 TALEVP64
相对表达
1.5
1.0
0.5
0.0
0 1 2 3 4 5 6 7 8
突变 #

图2E

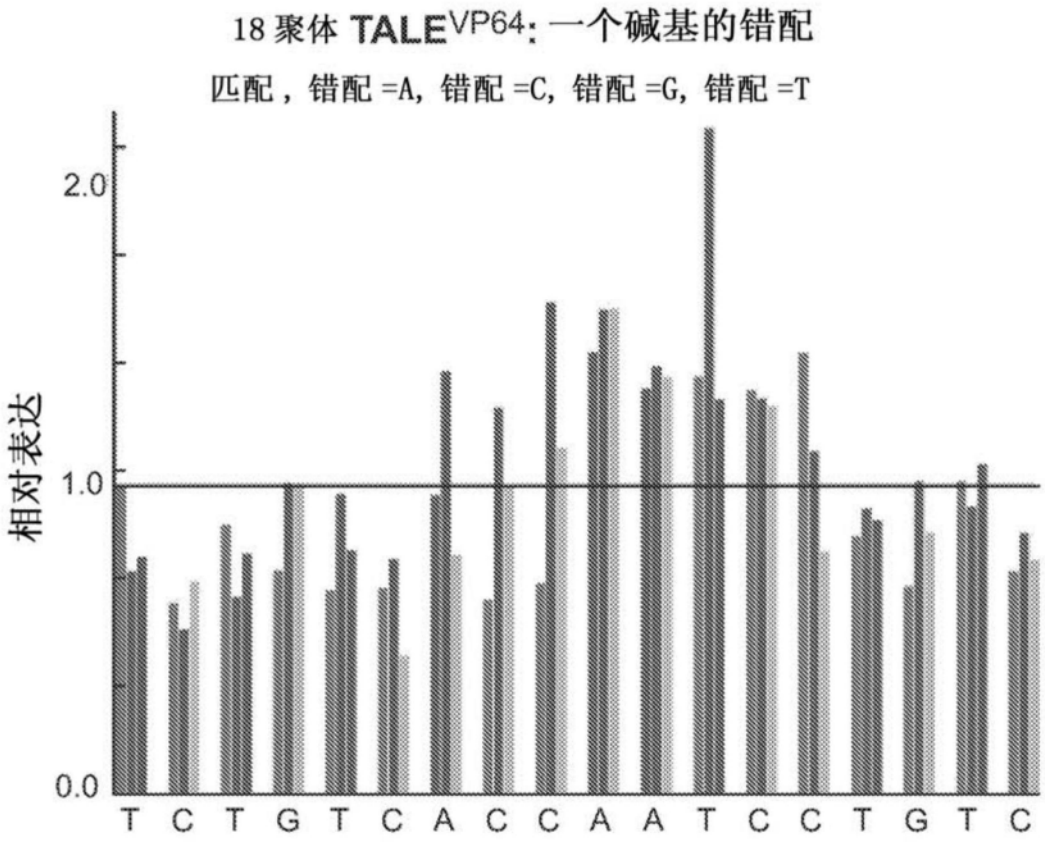

18 聚体 TALEVP64: 一个碱基的错配
匹配，错配 =A，错配 =C，错配 =G，错配 =T
相对表达
2.0
1.0
0.0
T C T G T C A C C A A T C C T G T C

图2F

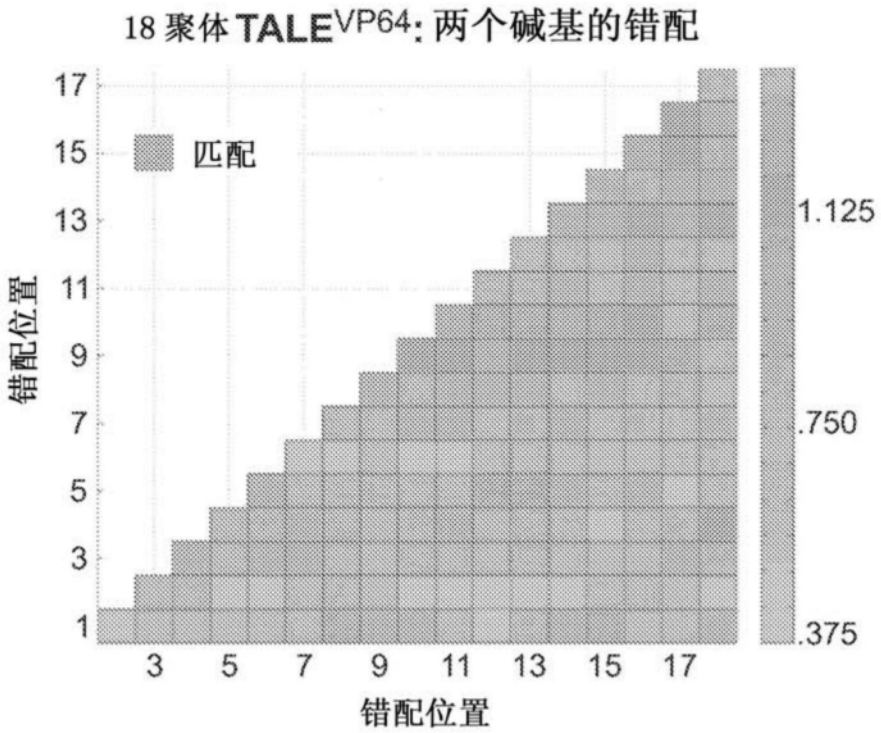

18 聚体 TALEVP64: 两个碱基的错配
匹配
错配位置
17
15
13
11
9
7
5
3
1
3 5 7 9 11 13 15 17
错配位置
1.125
.750
.375

图2G

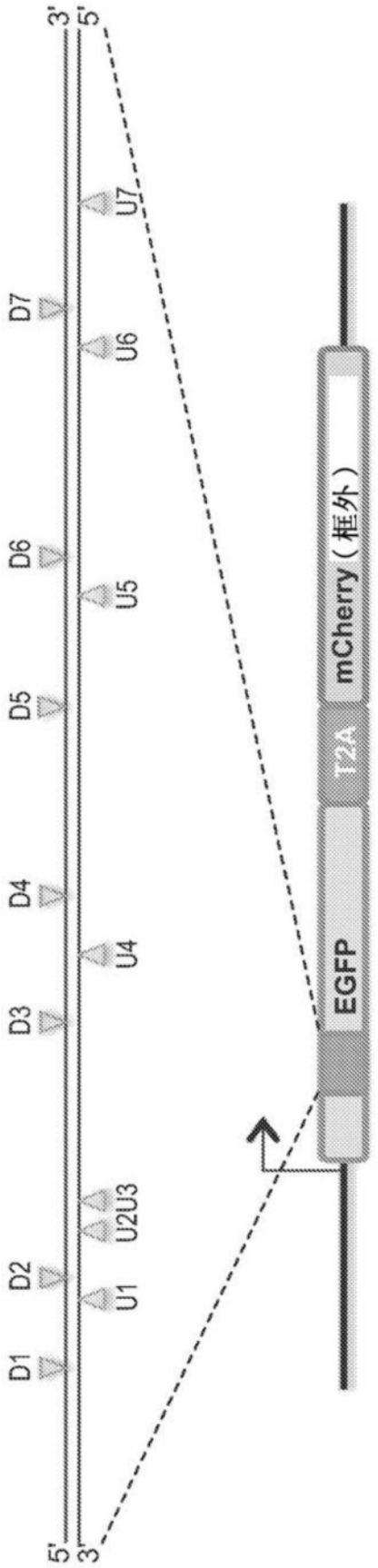


图3A

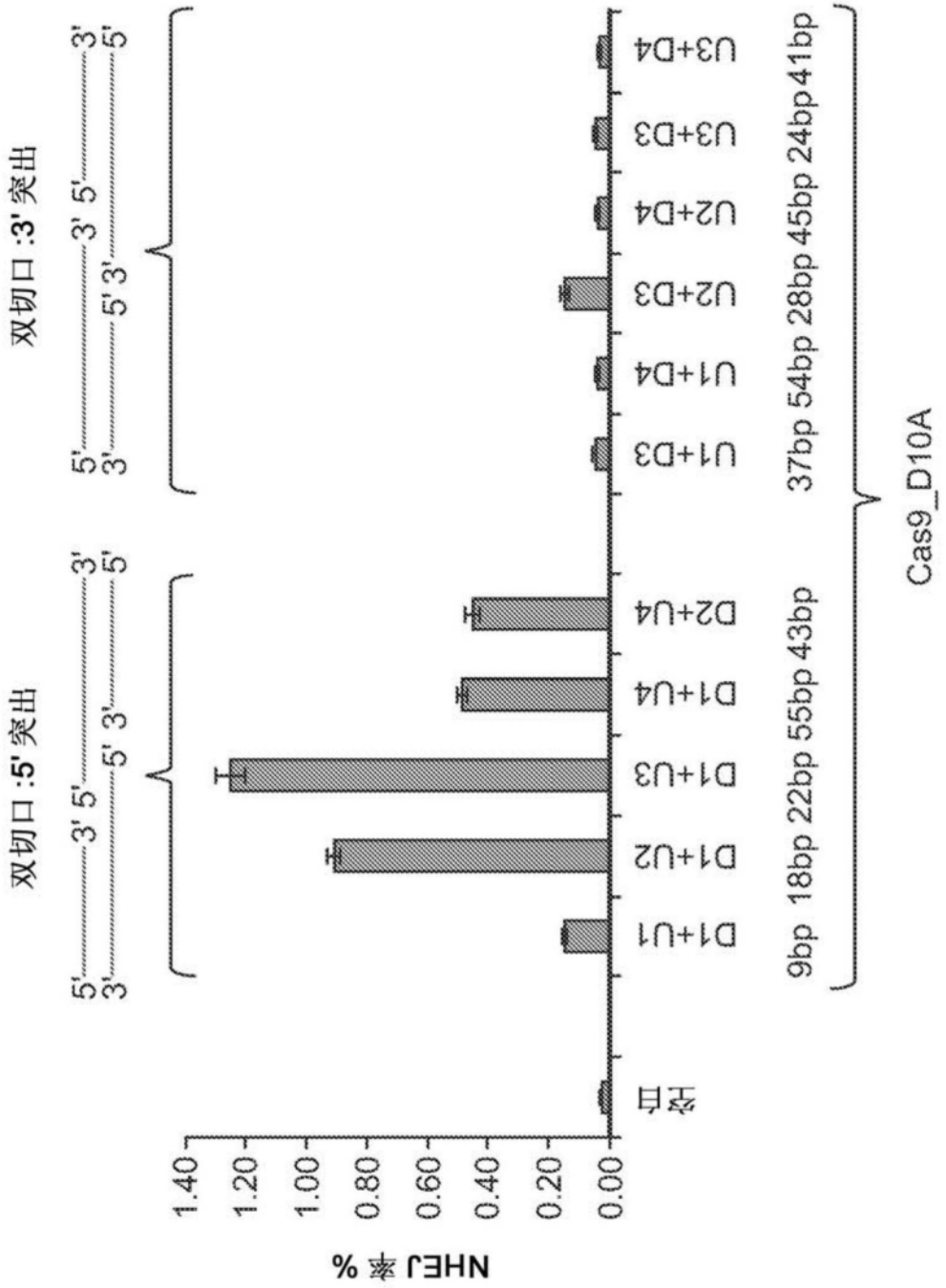

双切口 :5' 突出
双切口 :3' 突出
5'
3'
3' 5'
5' 3'
3'
5'
NHEJ 率 %
1.40
1.20
1.00
0.80
0.60
0.40
0.20
0.00
空白
D1+U1
D1+U2
D1+U3
D1+U4
D2+U4
U1+D3
U1+D4
U2+D3
U2+D4
U3+D3
U3+D4
9bp 18bp 22bp 55bp 43bp
37bp 54bp 28bp 45bp 24bp 41bp
Cas9_D10A

图3B

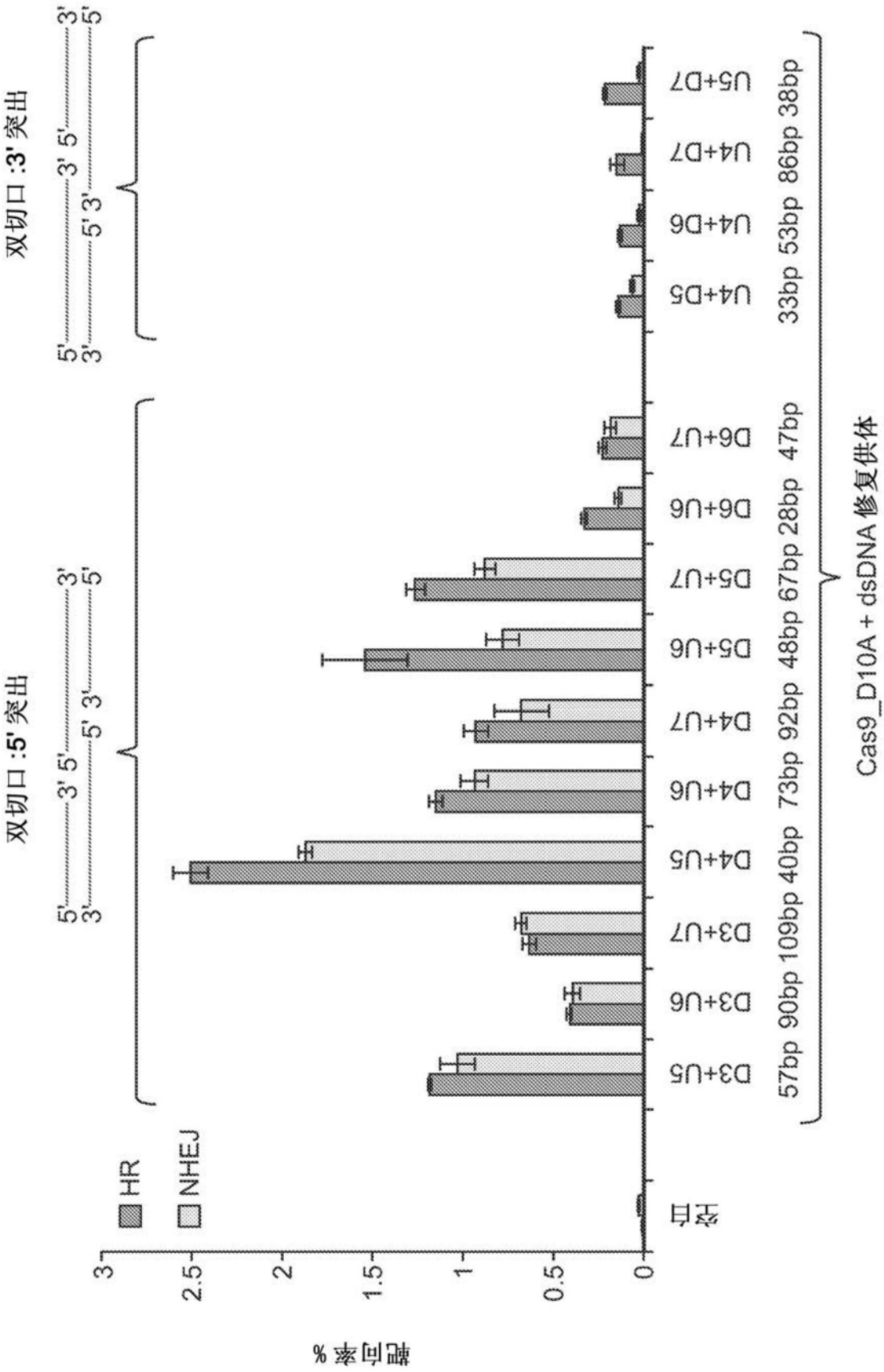
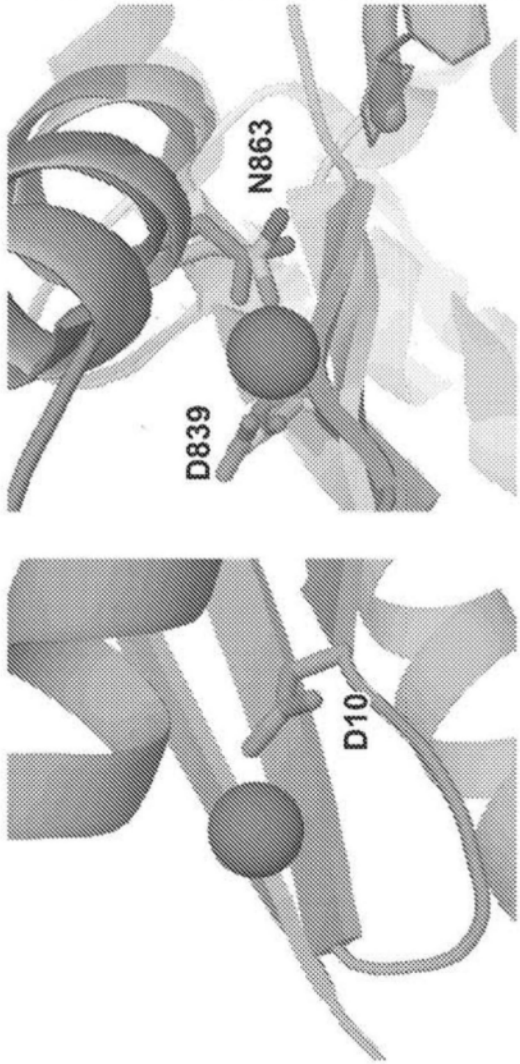

双切口 :5' 突出
5'
3'
3'
5'
5'
3'
3'
5'
双切口 :3' 突出
5'
3'
3'
5'
5'
3'
3'
5'
HR
NHEJ
靶向率 %
3
2.5
2
1.5
1
0.5
0
空白
D3+U5
D3+U6
D3+U7
D4+U5
D4+U6
D4+U7
D5+U6
D5+U7
D6+U6
D6+U7
U4+D5
U4+D6
U4+D7
U5+D7
57bp 90bp 109bp 40bp 73bp 92bp 48bp 67bp 28bp 47bp
33bp 53bp 86bp 38bp
Cas9_D10A + dsDNA 修复供体

图3C



名称	突变
Cas9	野生型
$Cas9_{m1}$	D10A
$Cas9_{m2}$	D10A+H840A
$Cas9_{m3}$	D10A+D839A+H840A
$Cas9_{m4}$	D10A+D839A+H840A+N863A

图4A

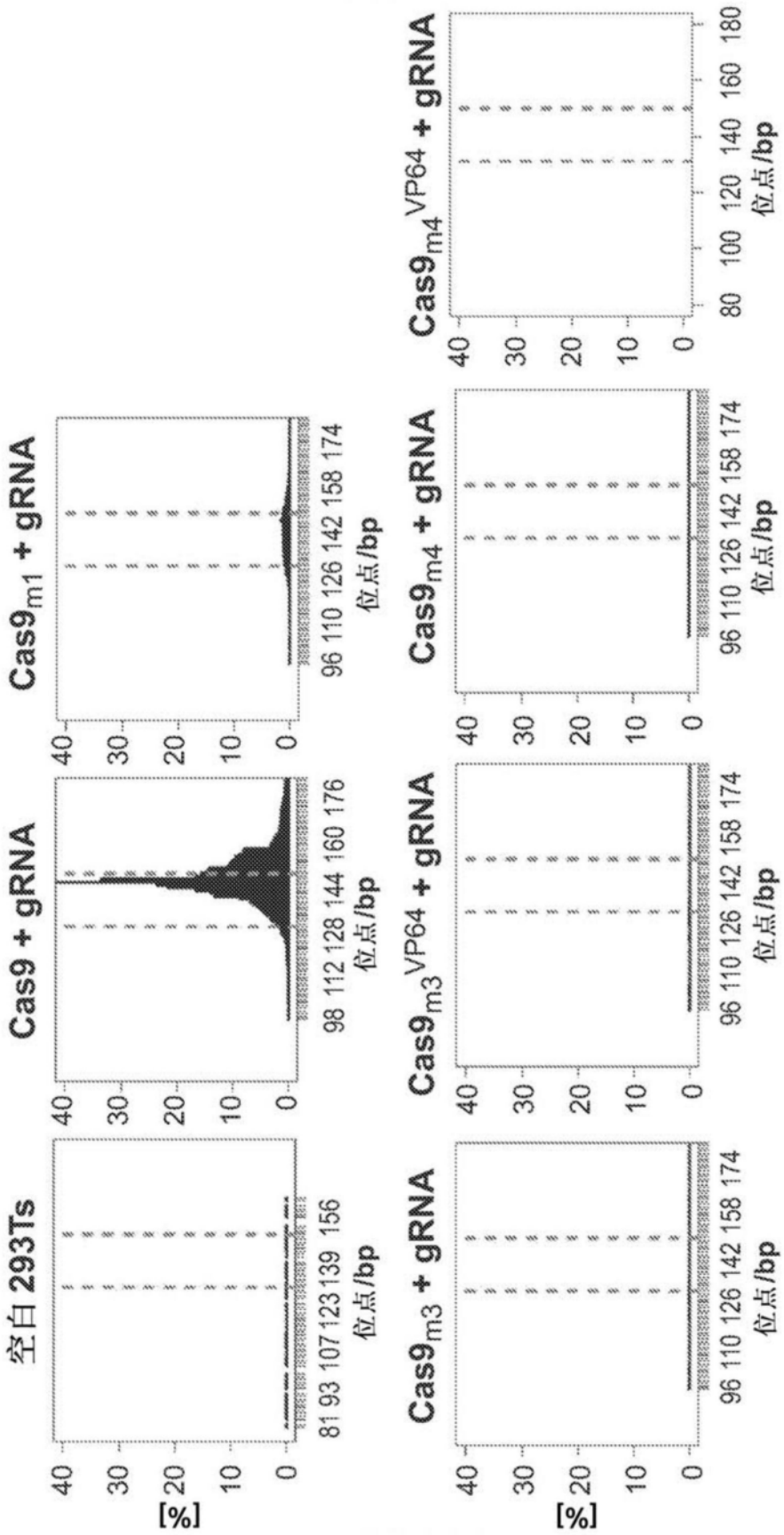

空白 293Ts
Cas9 + gRNA
Cas9m1 + gRNA
Cas9m3 + gRNA
Cas9m3VP64 + gRNA
Cas9m4 + gRNA
Cas9m4VP64 + gRNA
[%]
40
30
20
10
0
81 93 107 123 139 156
98 112 128 144 160 176
96 110 126 142 158 174
80 100 120 140 160 180
位点/bp

图4B

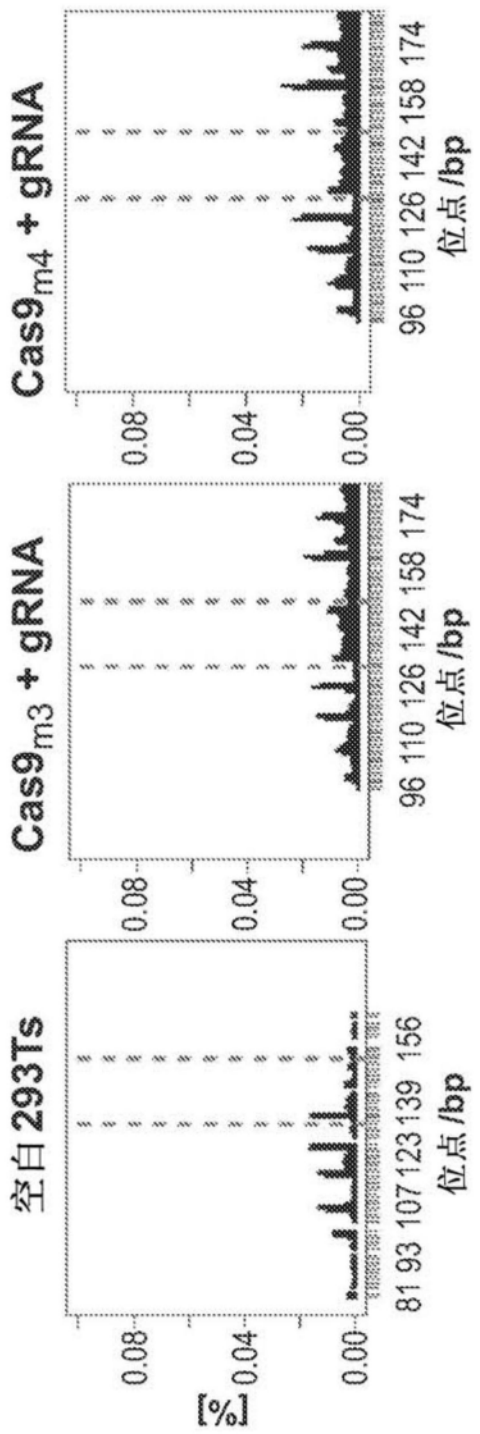

空白 293Ts
Cas9m3 + gRNA
Cas9m4 + gRNA
[%]
0.08
0.04
0.00
81 93 107 123 139 156
96 110 126 142 158 174
位点 /bp

图4C

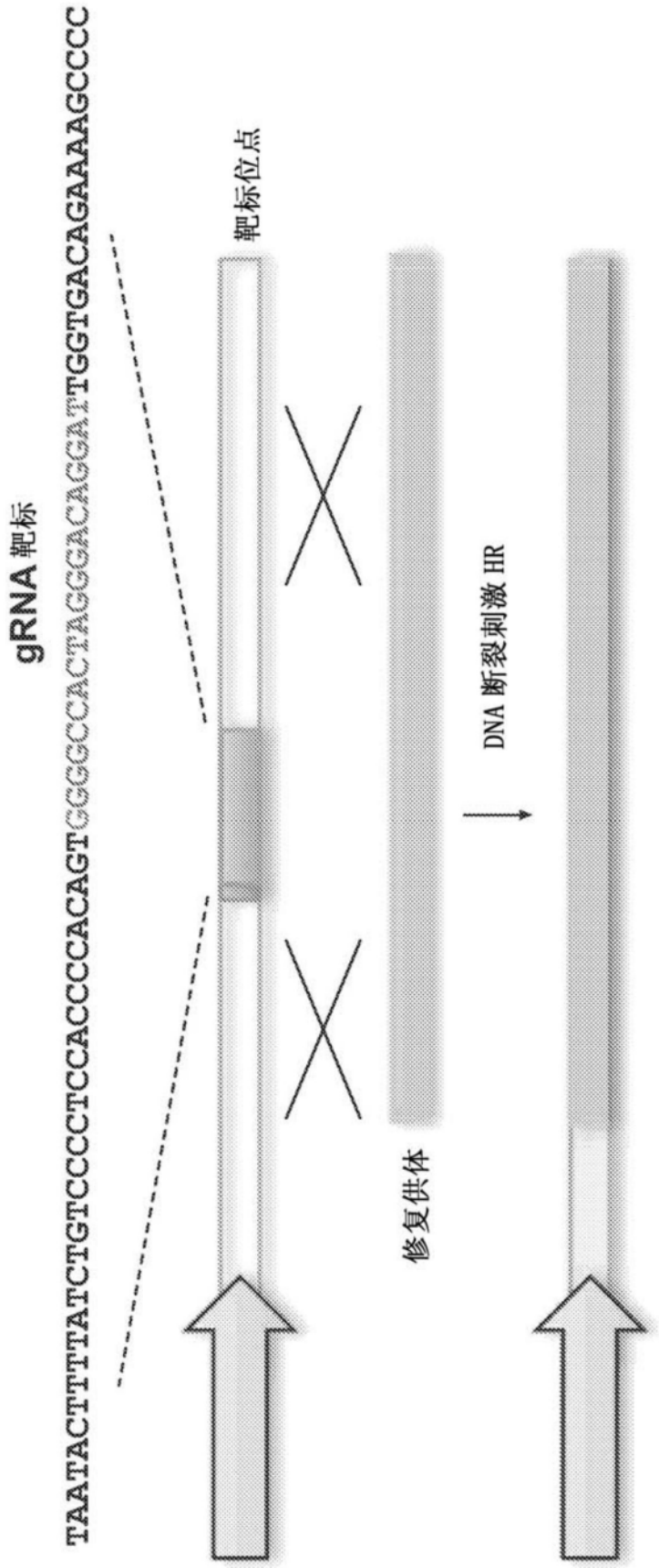

gRNA 靶标
TAATACTTTTTATCTGTCCCCTCCACCCCACAGTGGGGCCACTAGGGACAGGATTGGTGACAGAAAGCCCC
靶标位点
修复供体
DNA 断裂刺激 HR

图5A

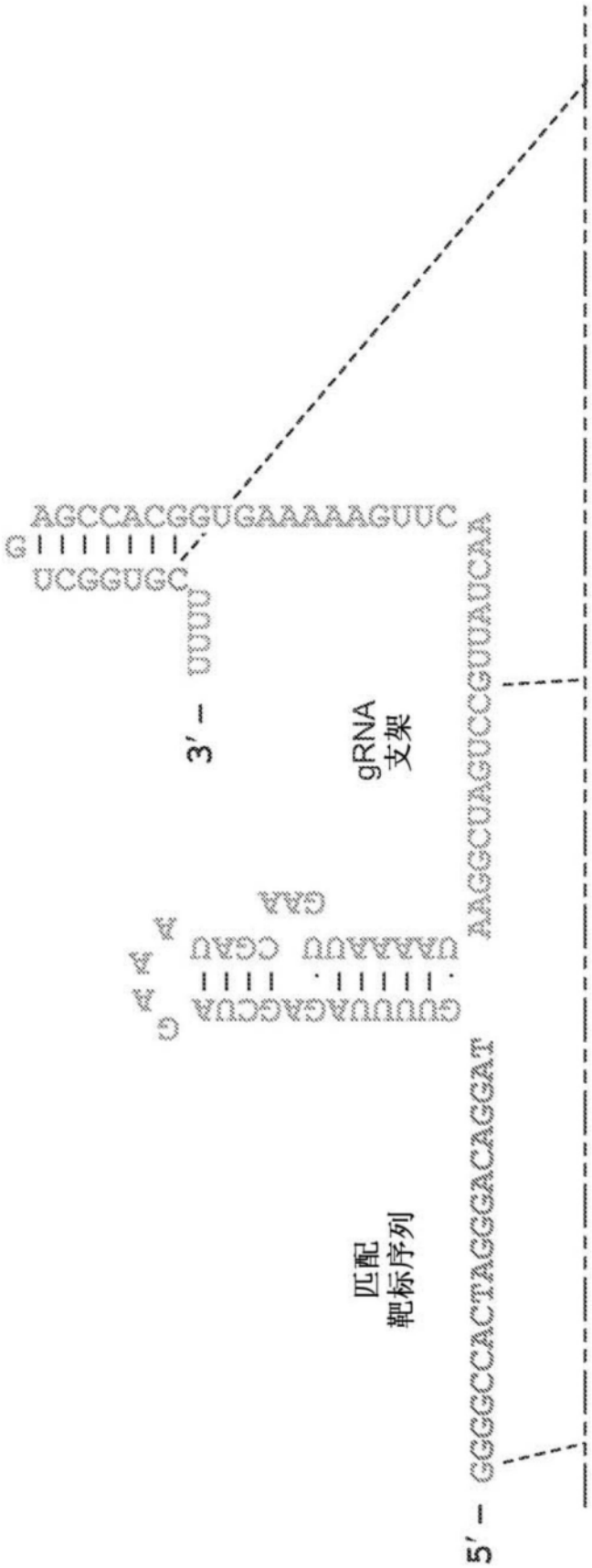

5′ – GGGGCCACTAGGGACAGGAT
匹配
靶标序列
AAGGCUAGUCCGUUAUCAA
gRNA
支架
3′ – UUUU
AGCCACGGUGAAAAAGUUC
UCGGUGC
G
GAA
UAAAAUU CGAU
GUUUUAGAGCUA
G
A
A
A
A

图5B-1

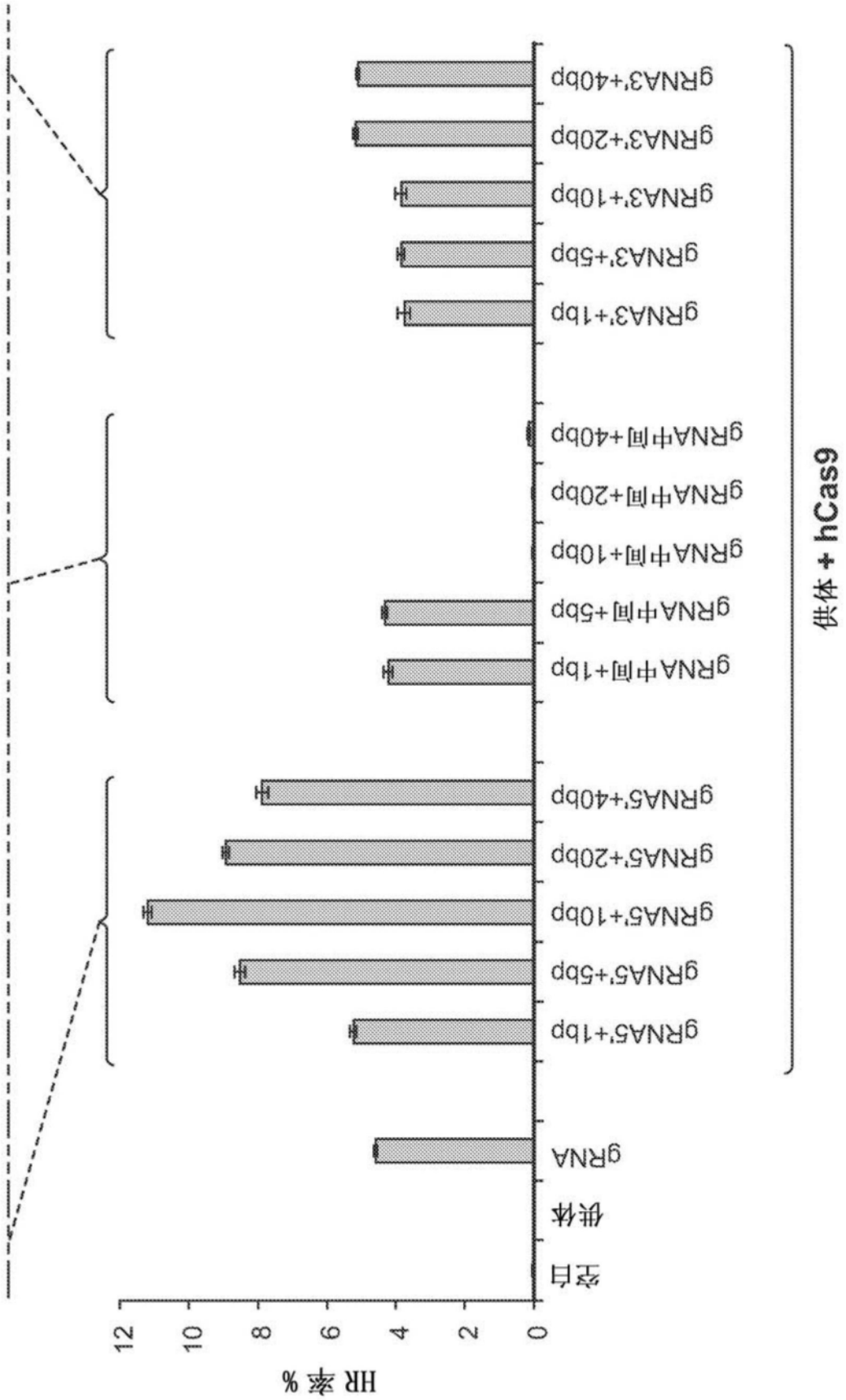

HR 率 %
12
10
8
6
4
2
0
空白
供体
gRNA
gRNA5'+1bp
gRNA5'+5bp
gRNA5'+10bp
gRNA5'+20bp
gRNA5'+40bp
gRNA中间+1bp
gRNA中间+5bp
gRNA中间+10bp
gRNA中间+20bp
gRNA中间+40bp
gRNA3'+1bp
gRNA3'+5bp
gRNA3'+10bp
gRNA3'+20bp
gRNA3'+40bp
供体 + hCas9

图5B-2

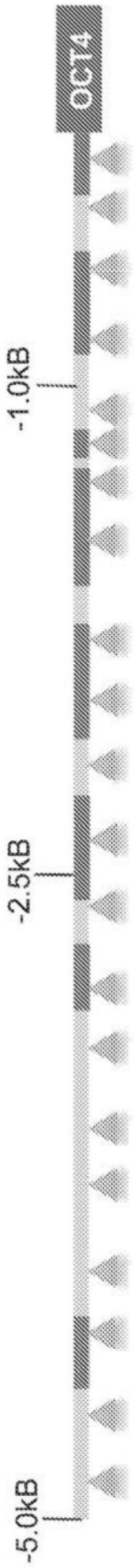


图6A

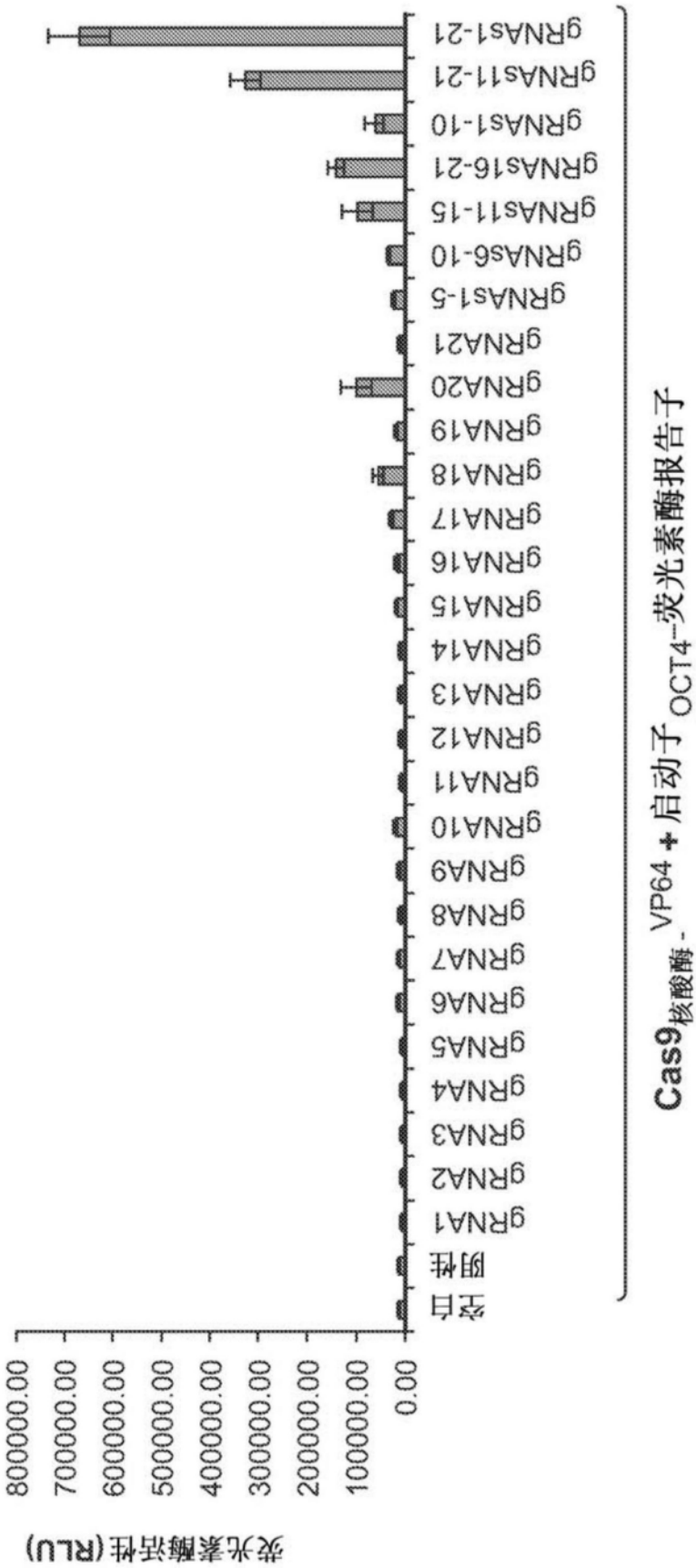

荧光素酶活性(RLU)
800000.00
700000.00
600000.00
500000.00
400000.00
300000.00
200000.00
100000.00
0.00
空白
阴性
gRNA1
gRNA2
gRNA3
gRNA4
gRNA5
gRNA6
gRNA7
gRNA8
gRNA9
gRNA10
gRNA11
gRNA12
gRNA13
gRNA14
gRNA15
gRNA16
gRNA17
gRNA18
gRNA19
gRNA20
gRNA21
gRNAs1-5
gRNAs6-10
gRNAs11-15
gRNAs16-21
gRNAs1-10
gRNAs11-21
gRNAs1-21
Cas9核酸酶-VP64 + 启动子OCT4-荧光素酶报告子

图6B

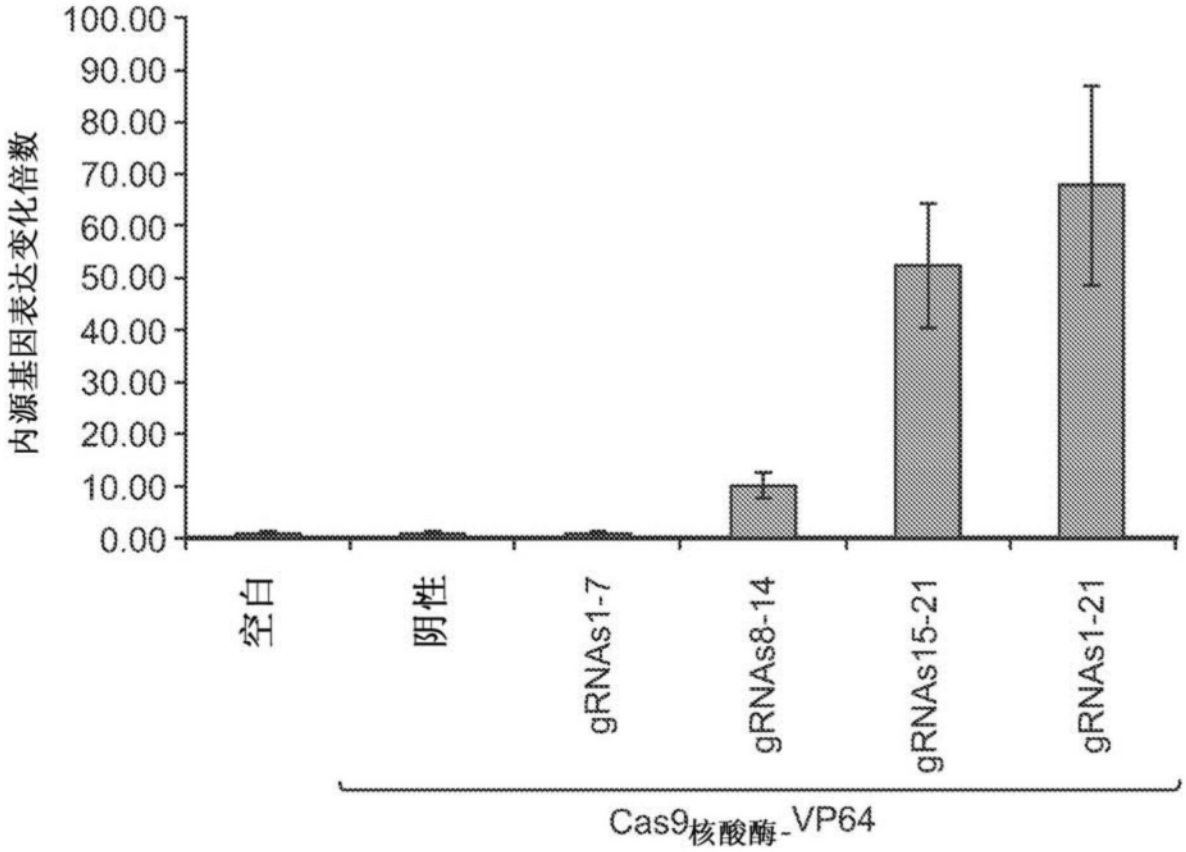

内源基因表达变化倍数
100.00
90.00
80.00
70.00
60.00
50.00
40.00
30.00
20.00
10.00
0.00
空白
阴性
gRNAs1-7
gRNAs8-14
gRNAs15-21
gRNAs1-21
Cas9核酸酶-VP64

图6C

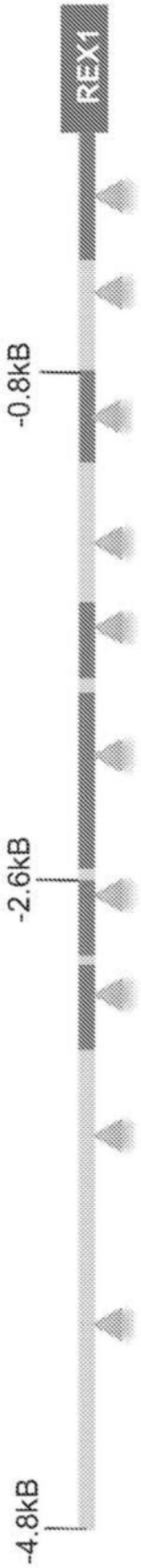


图7A

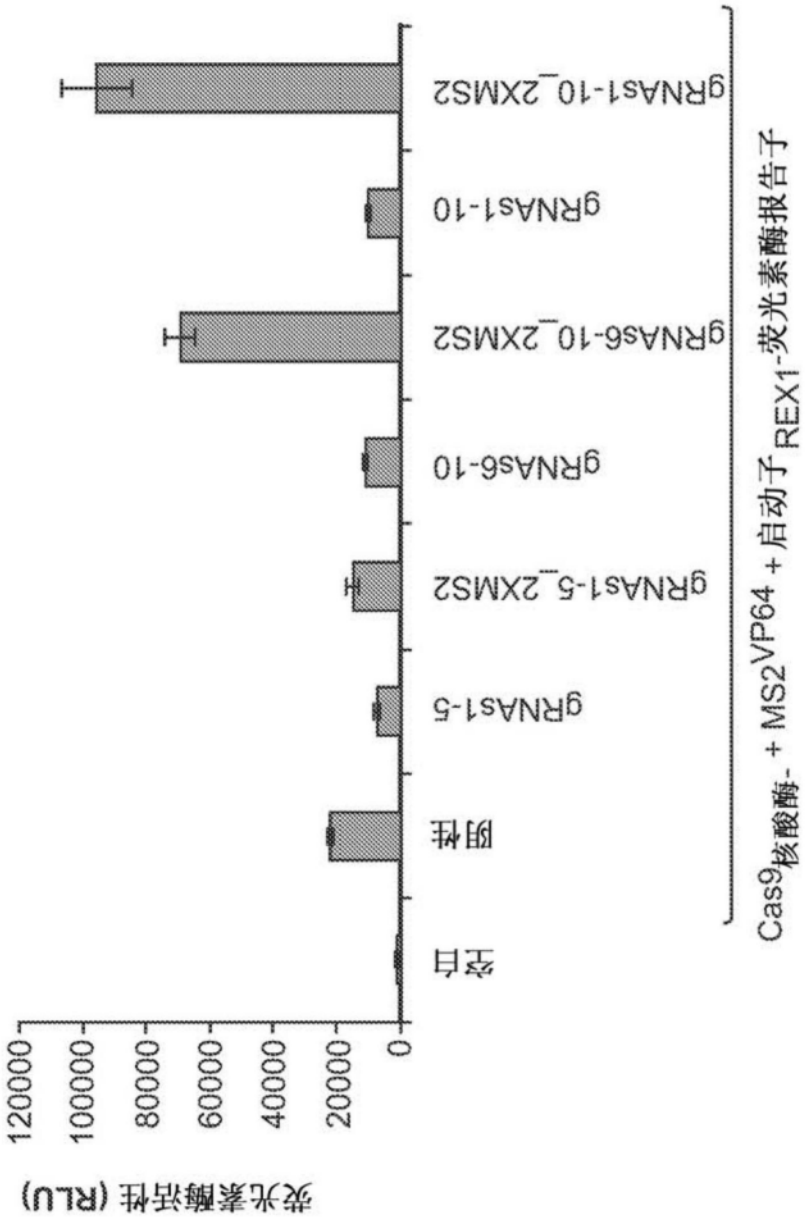

荧光素酶活性 (RLU)
120000
100000
80000
60000
40000
20000
0
空白
阴性
gRNAs1-5
gRNAs1-5_2XMS2
gRNAs6-10
gRNAs6-10_2XMS2
gRNAs1-10
gRNAs1-10_2XMS2
Cas9核酸酶- + MS2VP64 + 启动子REX1-荧光素酶报告子

图7B

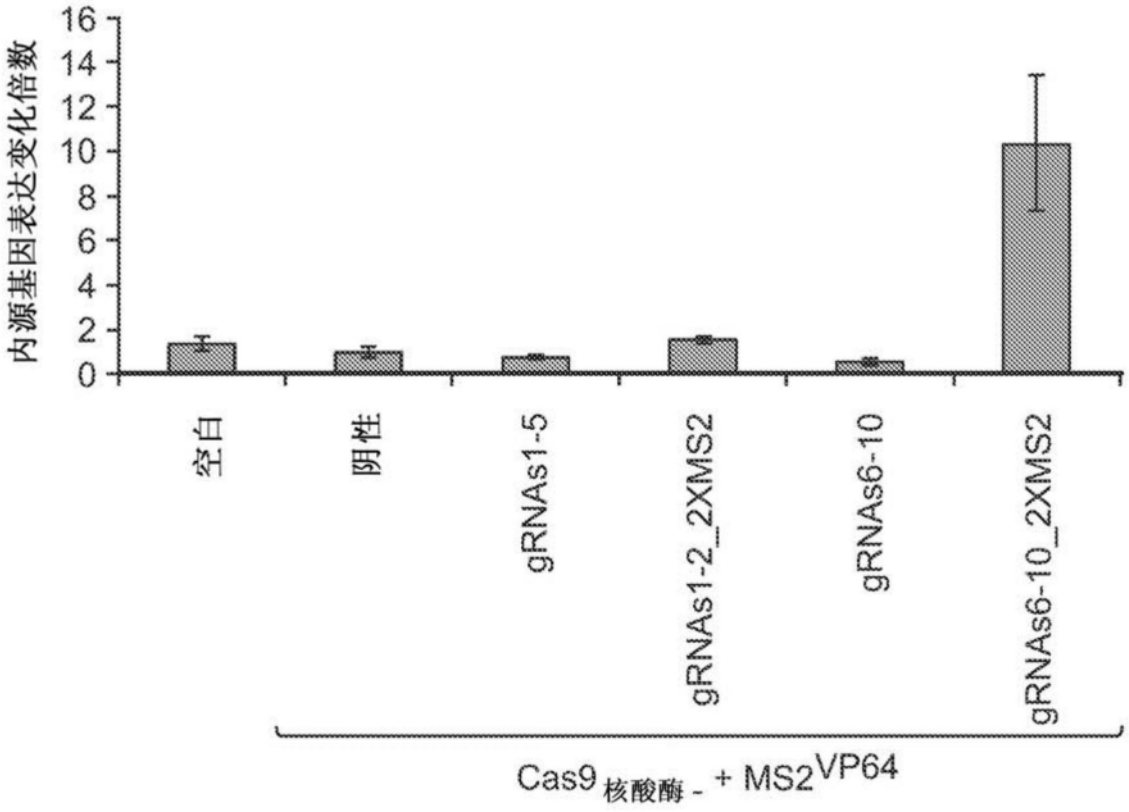

内源基因表达变化倍数
16
14
12
10
8
6
4
2
0
空白
阴性
gRNAs1-5
gRNAs1-2_2XMS2
gRNAs6-10
gRNAs6-10_2XMS2
Cas9核酸酶 - + MS2VP64

图7C

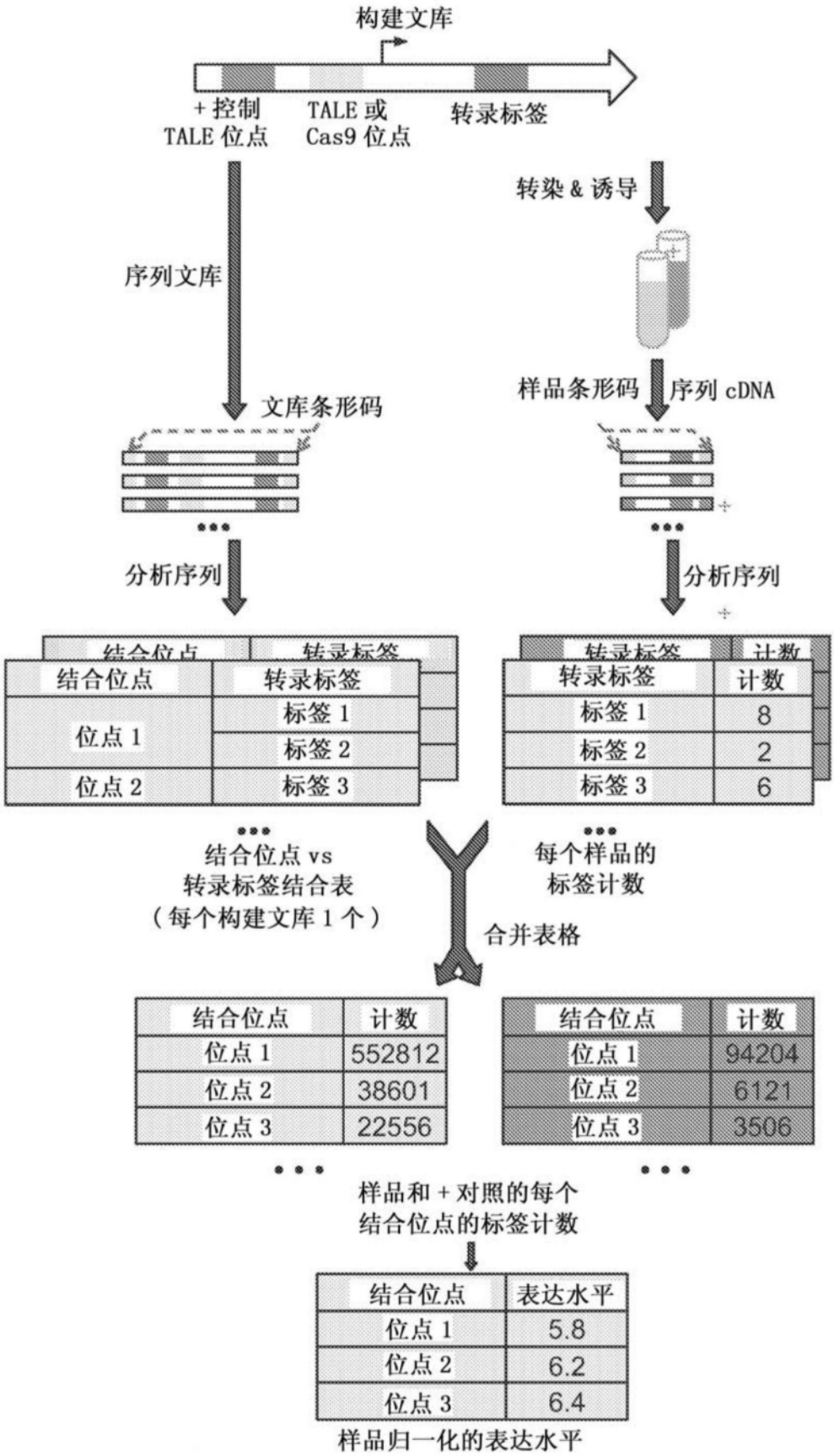

构建文库
+控制
TALE 位点
TALE 或
Cas9 位点
转录标签
序列文库
转染 & 诱导
文库条形码
样品条形码
序列 cDNA
分析序列
分析序列
结合位点
转录标签
位点 1
标签 1
标签 2
位点 2
标签 3
转录标签
计数
标签 1
8
标签 2
2
标签 3
6
结合位点 vs
转录标签结合表
（每个构建文库 1 个）
每个样品的
标签计数
合并表格
结合位点
计数
位点 1
552812
位点 2
38601
位点 3
22556
结合位点
计数
位点 1
94204
位点 2
6121
位点 3
3506
样品和 + 对照的每个
结合位点的标签计数
结合位点
表达水平
位点 1
5.8
位点 2
6.2
位点 3
6.4
样品归一化的表达水平

图8A

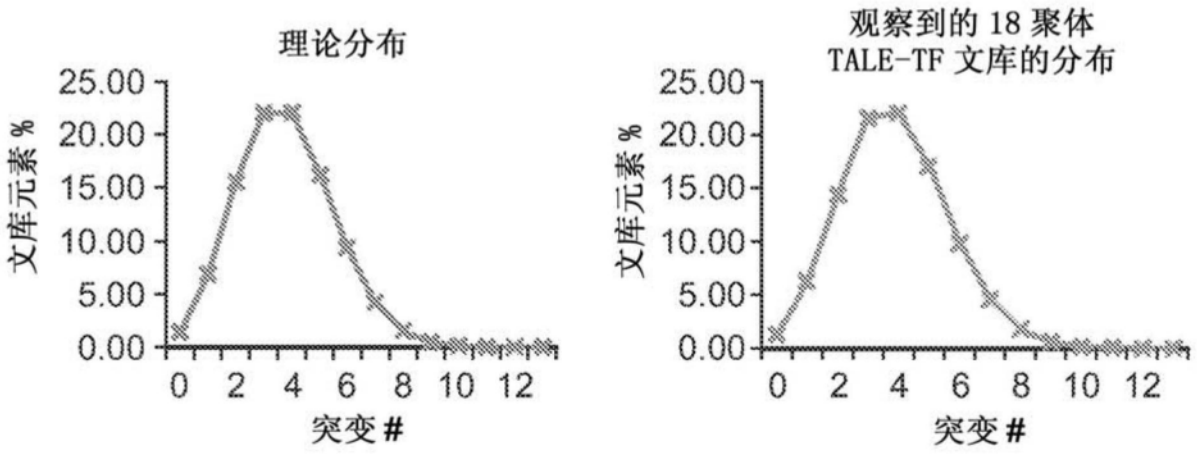

理论分布
文库元素 %
25.00
20.00
15.00
10.00
5.00
0.00
0 2 4 6 8 10 12
突变 #
观察到的 18 聚体
TALE-TF 文库的分布
文库元素 %
25.00
20.00
15.00
10.00
5.00
0.00
0 2 4 6 8 10 12
突变 #

图8B

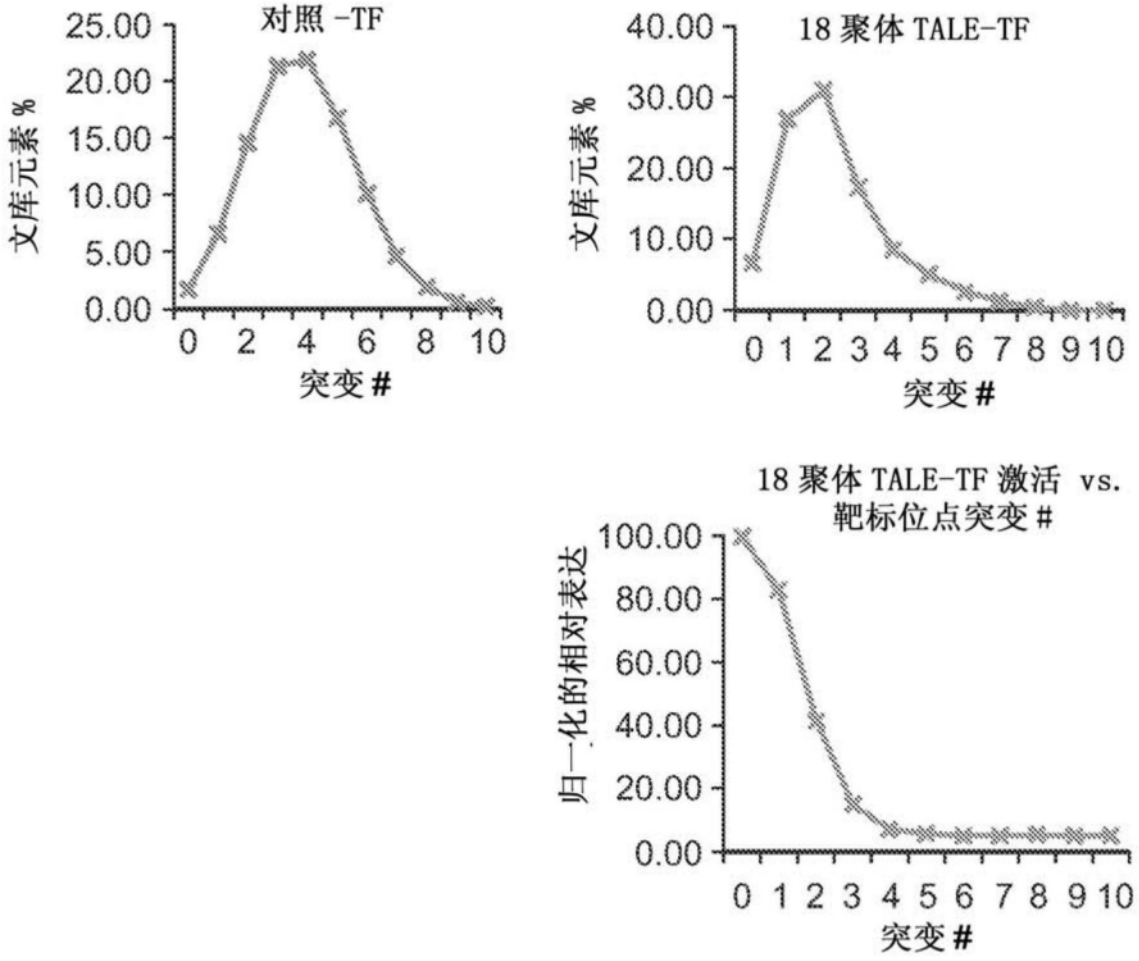

对照 -TF
文库元素 %
25.00
20.00
15.00
10.00
5.00
0.00
0 2 4 6 8 10
突变 #
18 聚体 TALE-TF
文库元素 %
40.00
30.00
20.00
10.00
0.00
0 1 2 3 4 5 6 7 8 9 10
突变 #
18 聚体 TALE-TF 激活 vs.
靶标位点突变 #
归一化的相对表达
100.00
80.00
60.00
40.00
20.00
0.00
0 1 2 3 4 5 6 7 8 9 10
突变 #

图8C

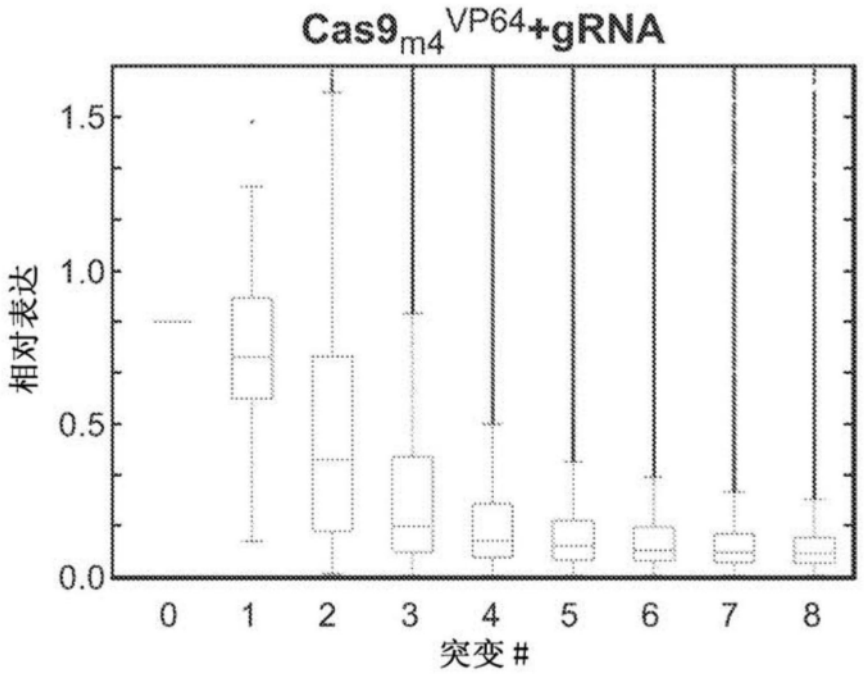

Cas9m4VP64+gRNA
相对表达
1.5
1.0
0.5
0.0
0
1
2
3
4
5
6
7
8
突变#

图9A

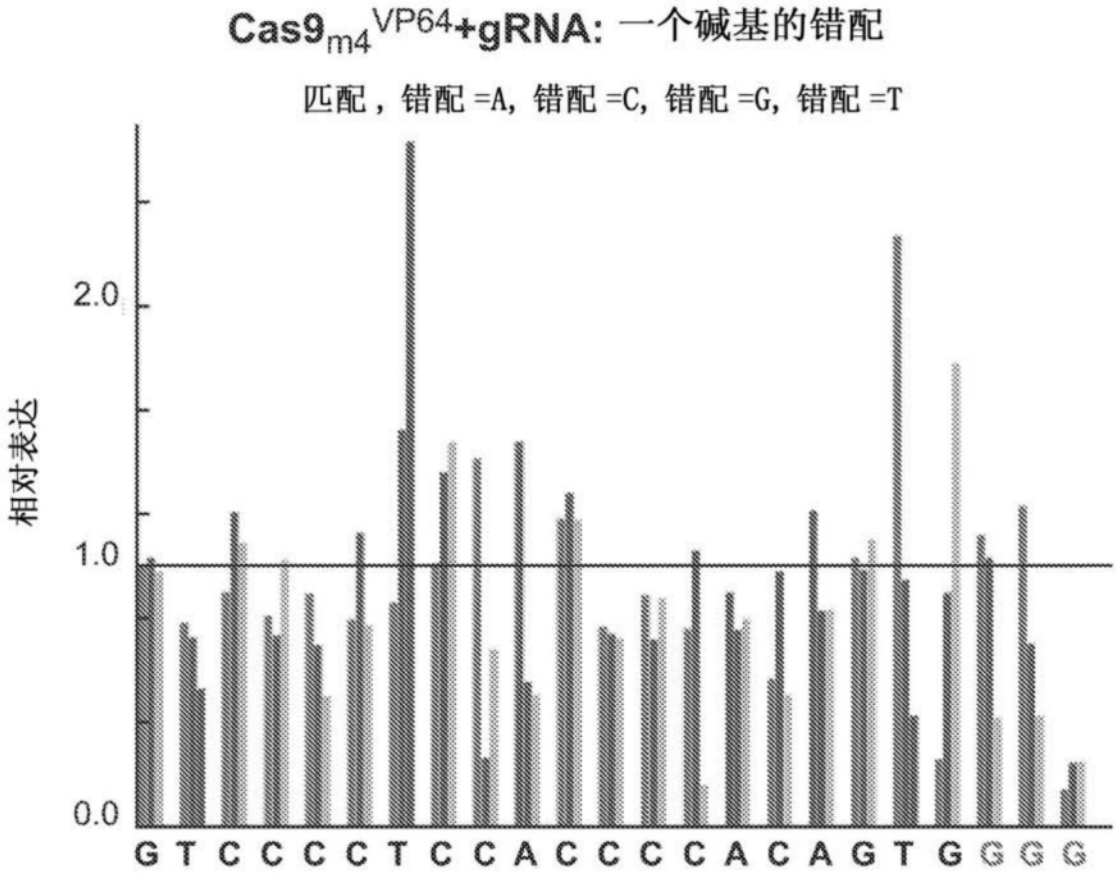

Cas9m4VP64+gRNA: 一个碱基的错配
匹配，错配 =A，错配 =C，错配 =G，错配 =T
相对表达
2.0
1.0
0.0
G T C C C C T C C A C C C C A C A G T G G G G

图9B

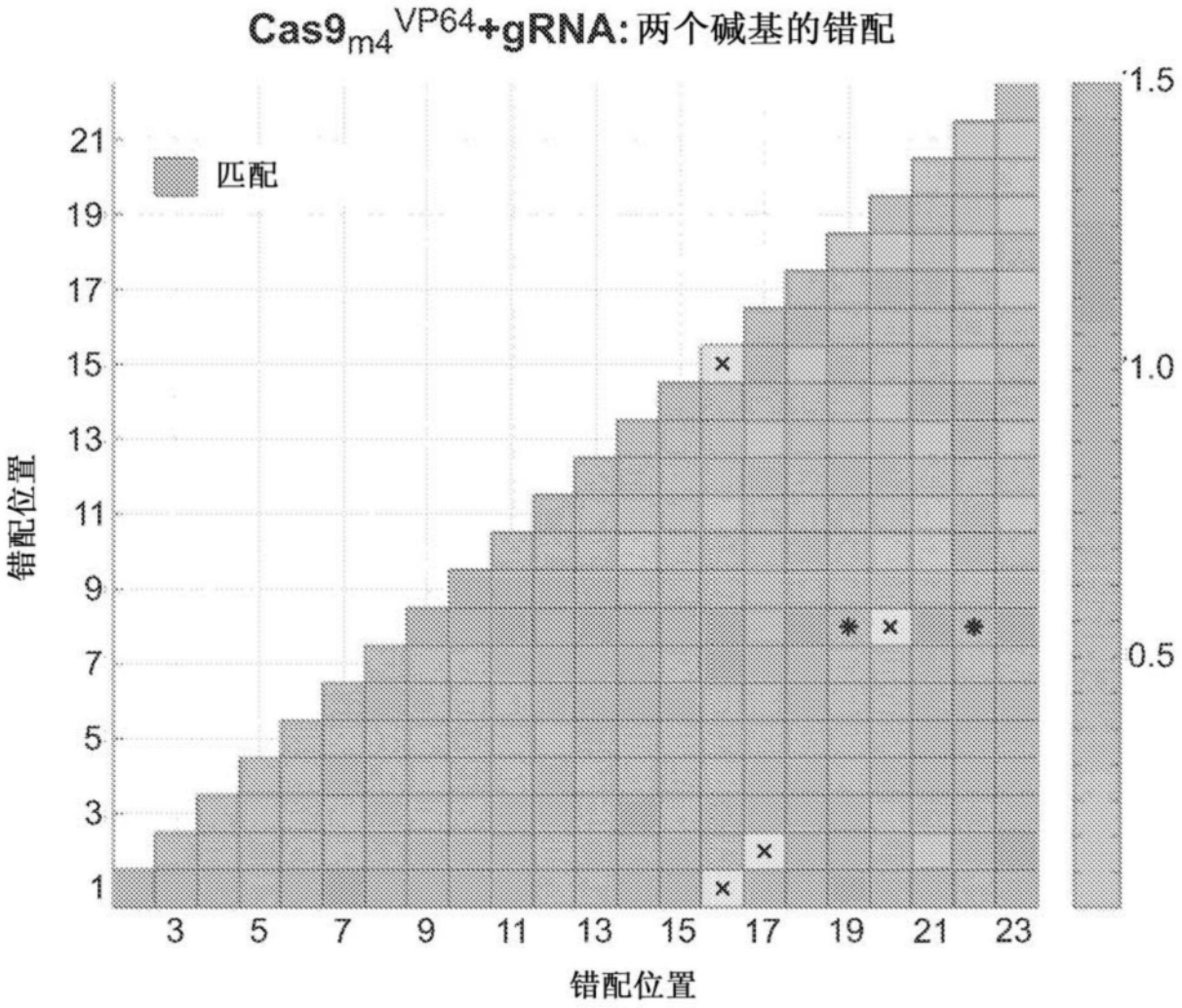

Cas9m4VP64+gRNA:两个碱基的错配
匹配
错配位置
错配位置
1
3
5
7
9
11
13
15
17
19
21
23
1.5
1.0
0.5

图9C

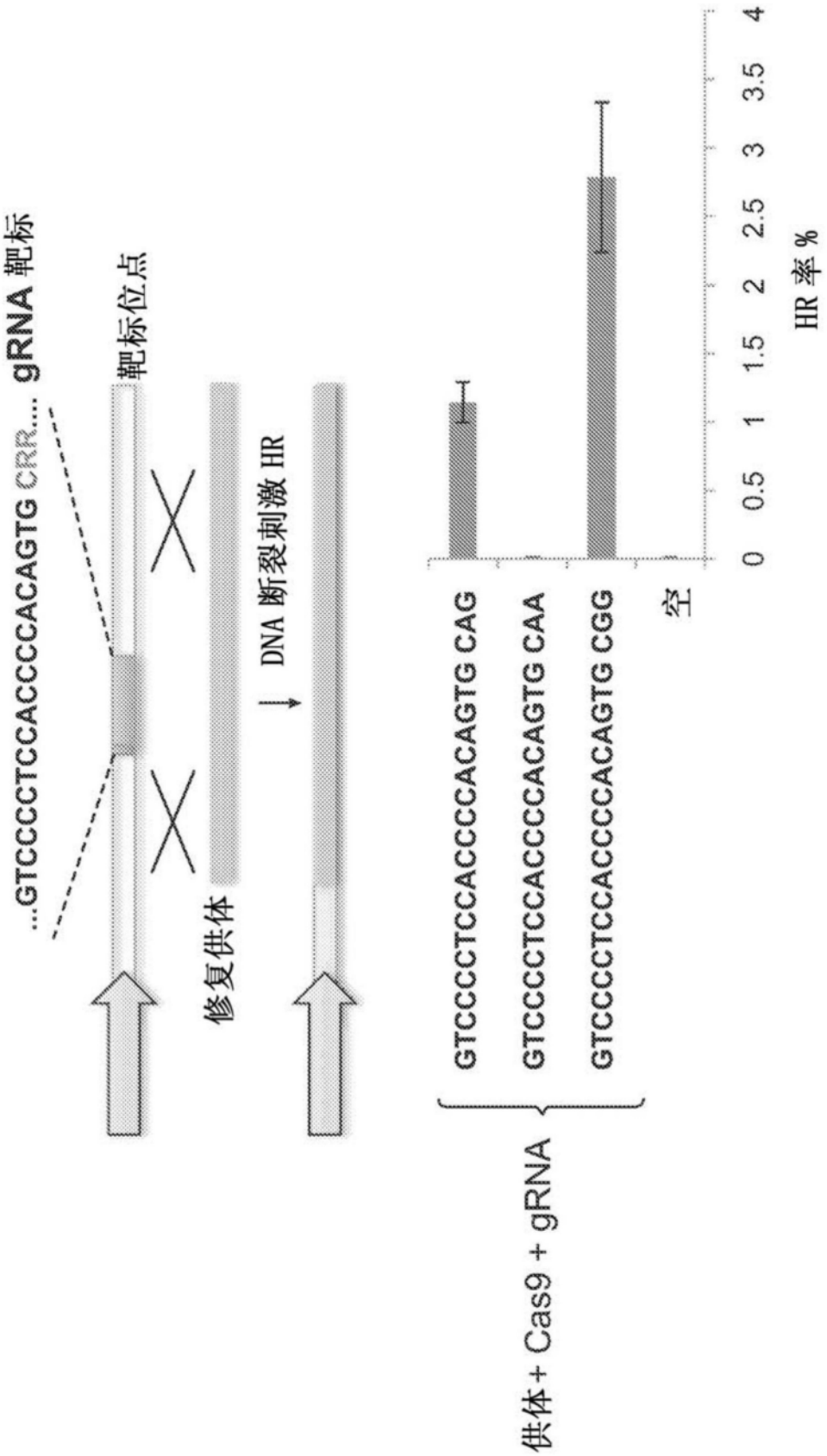

...GTCCCCTCCACCCCACAGTG CRR.... gRNA 靶标
靶标位点
修复供体
DNA 断裂刺激 HR
GTCCCCTCCACCCCACAGTG CAG
GTCCCCTCCACCCCACAGTG CAA
GTCCCCTCCACCCCACAGTG CGG
空
供体+ Cas9 + gRNA
0
0.5
1
1.5
2
2.5
3
3.5
4
HR 率 %

图9D

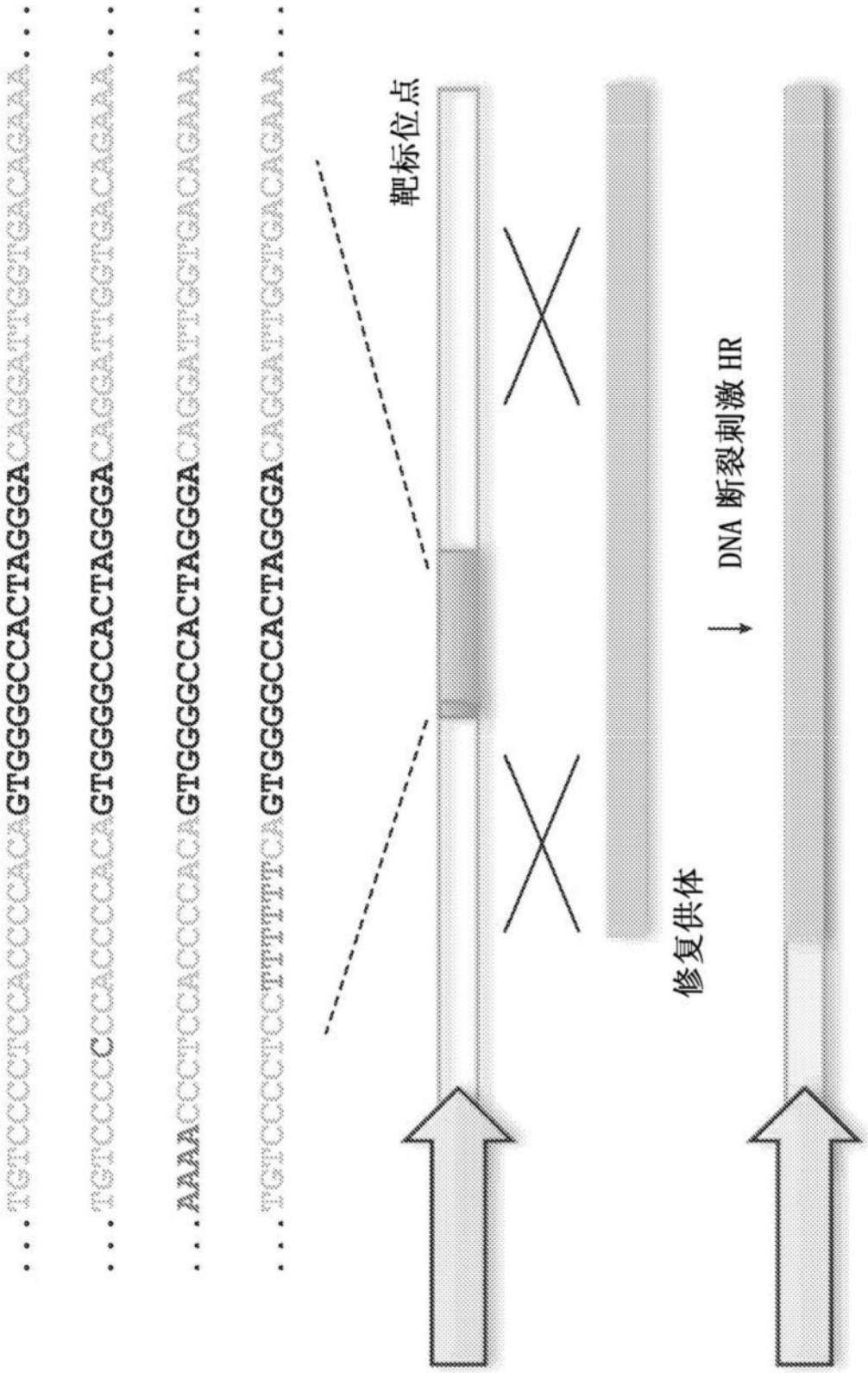

...TGTCCCCTCCACCCCACAGTGGGGCCACTAGGGACAGGATTGGTGACAGAAA...
...TGTCCCCCCCACCCCACAGTGGGGCCACTAGGGACAGGATTGGTGACAGAAA...
...AAAACCCTCCACCCCACAGTGGGGCCACTAGGGACAGGATTGGTGACAGAAA...
...TGTCCCCTCCTTTTTTTCAGTGGGGCCACTAGGGACAGGATTGGTGACAGAAA...
靶标位点
修复供体
DNA 断裂刺激 HR

图10A-1

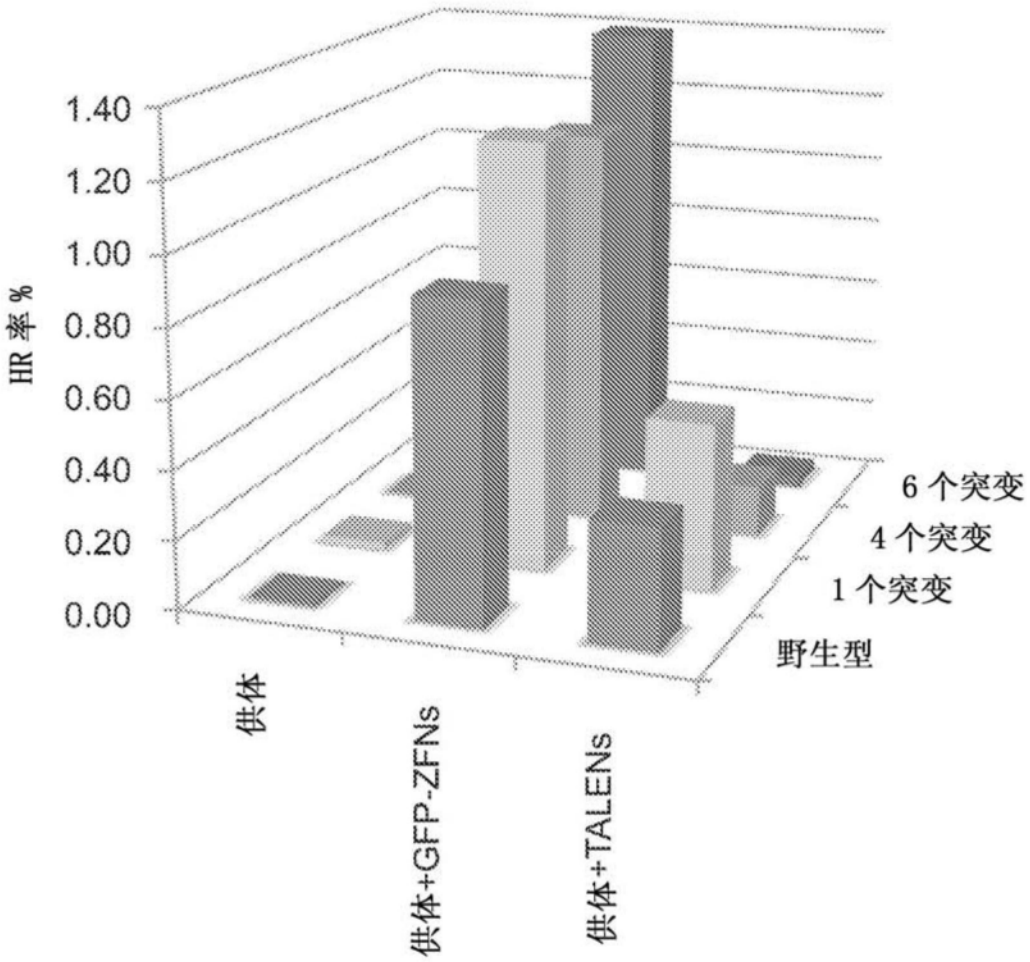

HR 率 %
1.40
1.20
1.00
0.80
0.60
0.40
0.20
0.00
供体
供体+GFP-ZFNs
供体+TALENs
6 个突变
4 个突变
1 个突变
野生型

图10A-2

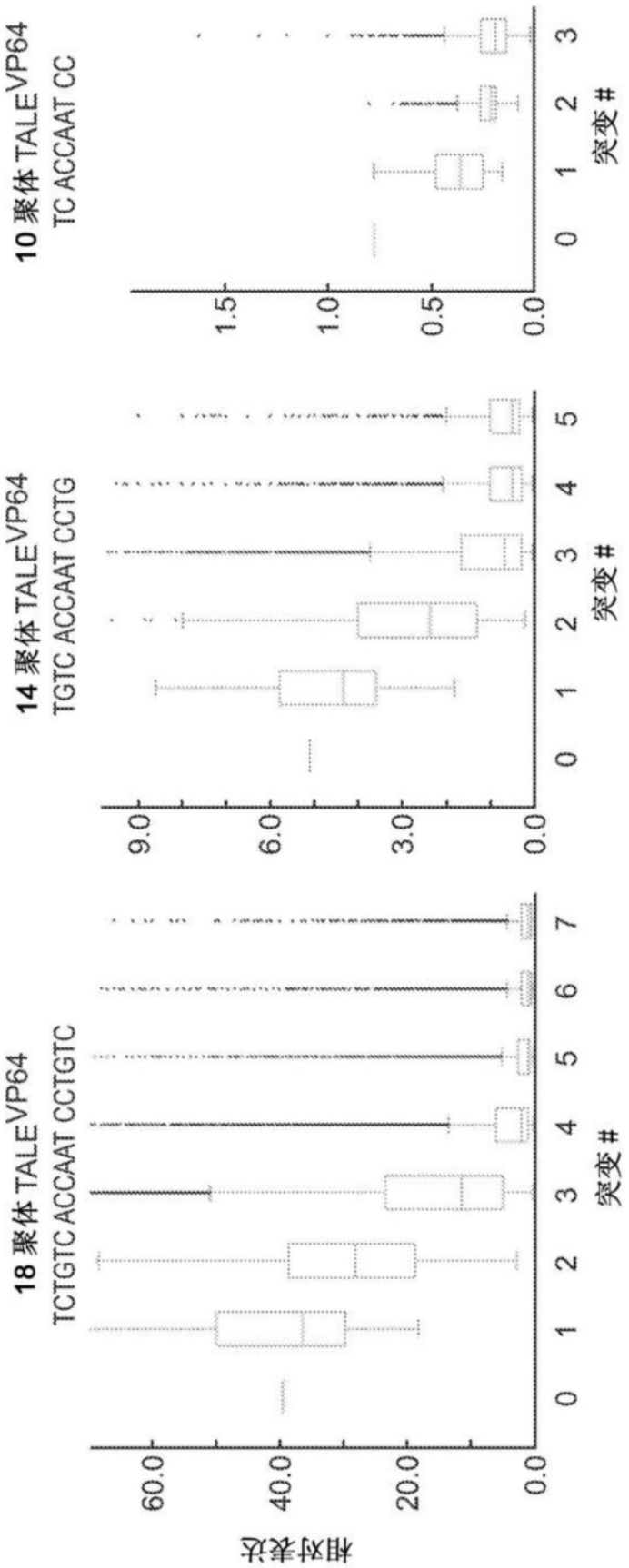

18 聚体 TALE VP64
TCTGTC ACCAAT CCTGTC
14 聚体 TALE VP64
TGTC ACCAAT CCTG
10 聚体 TALE VP64
TC ACCAAT CC
相对表达
突变 #
突变 #
突变 #

图10B

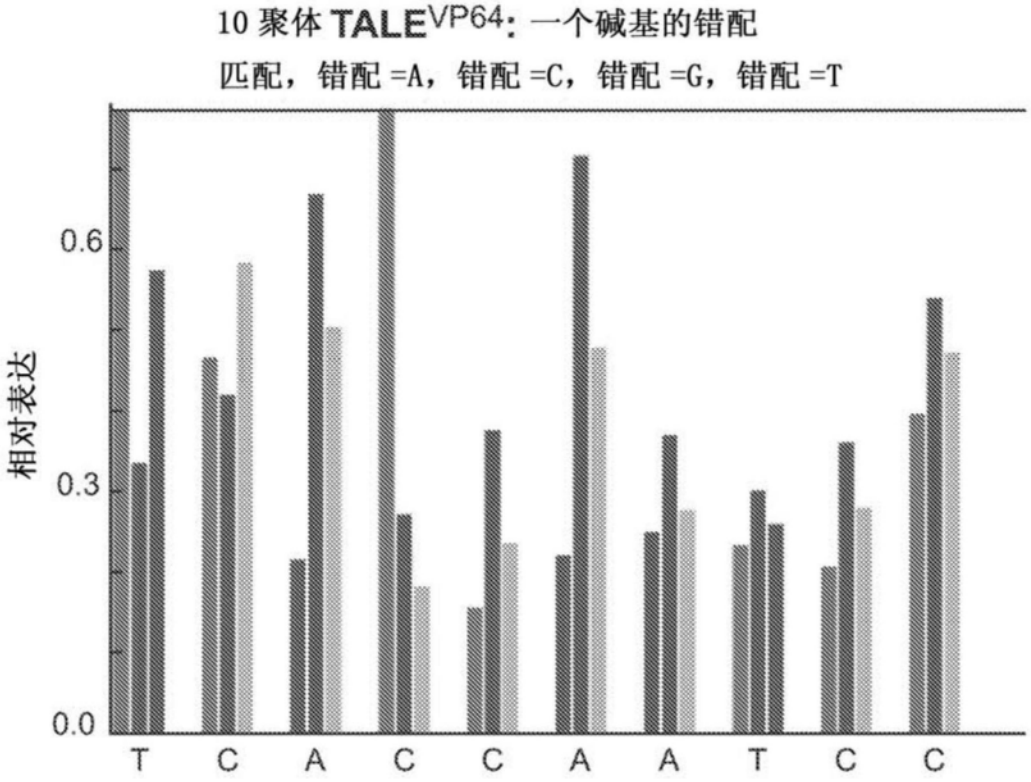

10 聚体 TALE VP64：一个碱基的错配
匹配，错配 =A，错配 =C，错配 =G，错配 =T
相对表达
0.6
0.3
0.0
T
C
A
C
C
A
A
T
C
C

图10C

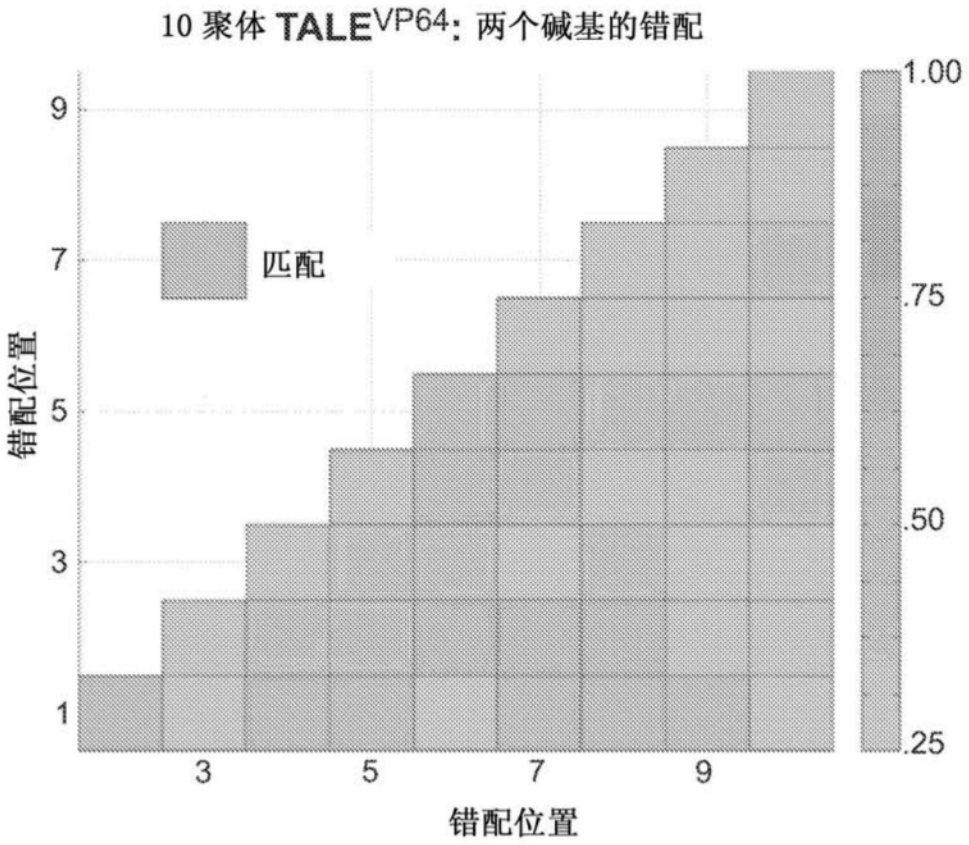


图10D

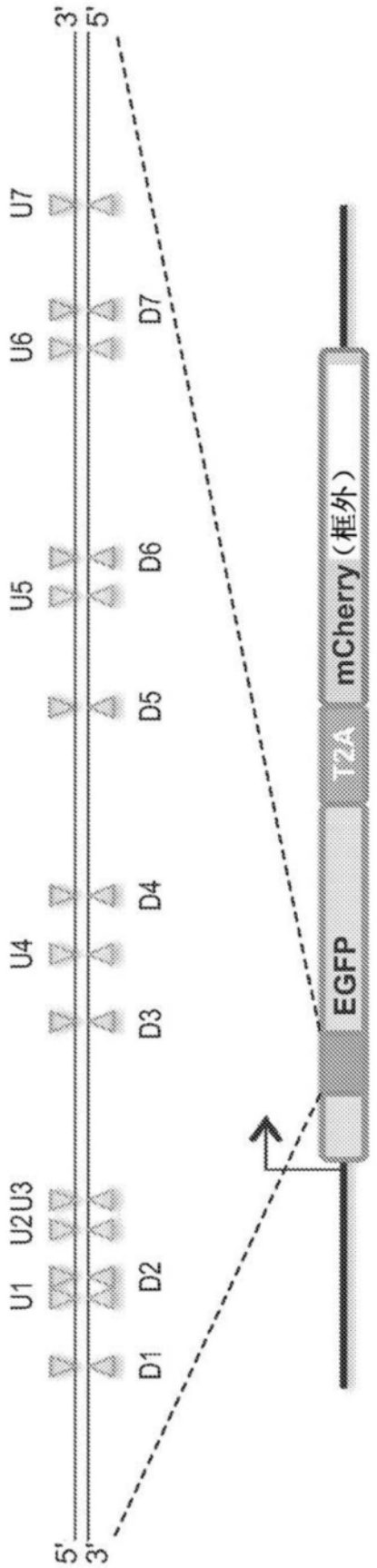


图11A

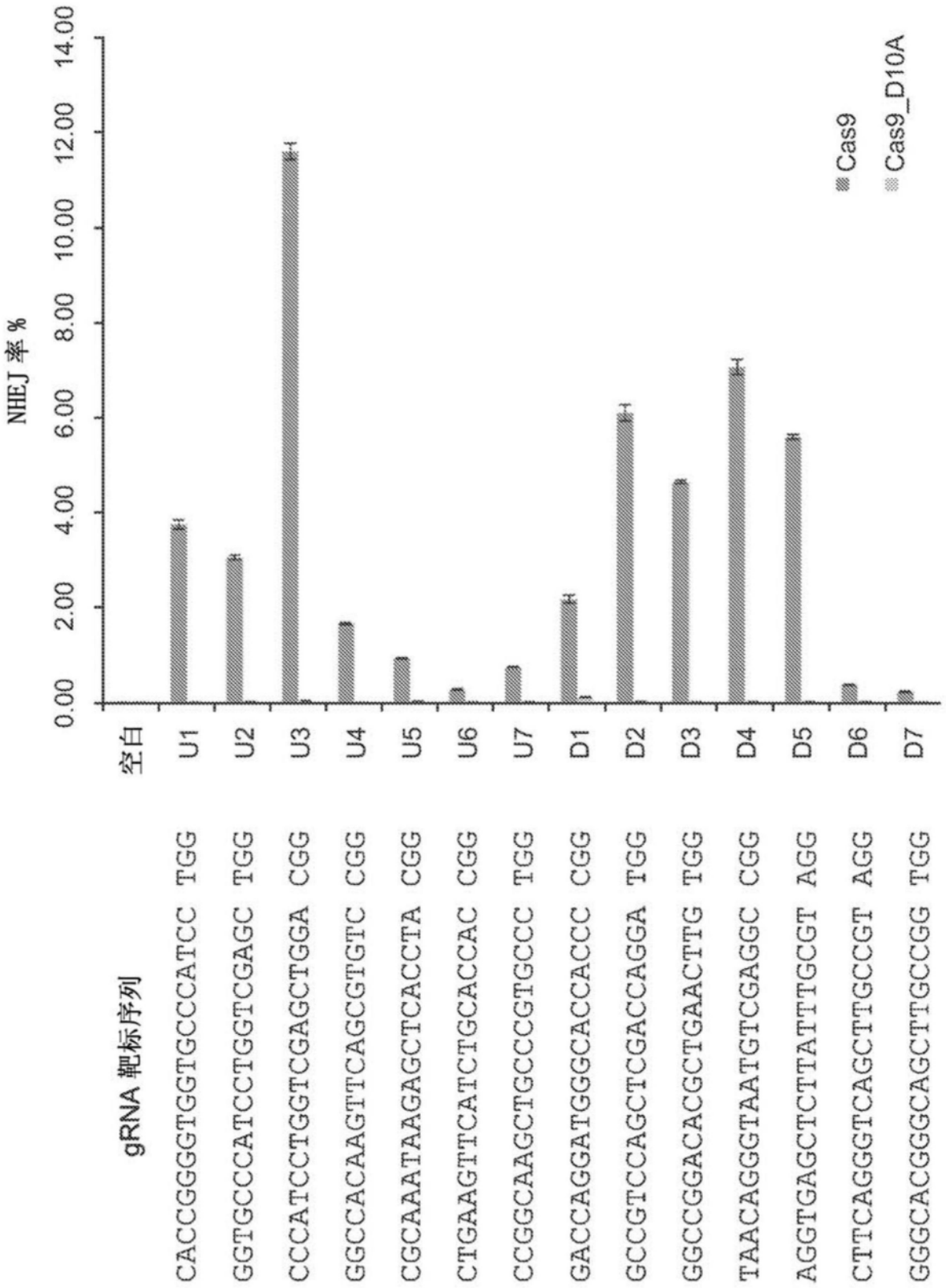

NHEJ 率 %
0.00
2.00
4.00
6.00
8.00
10.00
12.00
14.00
Cas9
Cas9_D10A
空白
U1
U2
U3
U4
U5
U6
U7
D1
D2
D3
D4
D5
D6
D7
gRNA 靶标序列
CACCGGGGTGGTGCCCATCC TGG
GGTGCCCATCCTGGTCGAGC TGG
CCCATCCTGGTCGAGCTGGA CGG
GGCCACAAGTTCAGCGTGTC CGG
CGCAAATAAGAGCTCACCTA CGG
CTGAAGTTCATCTGCACCAC CGG
CCGGCAAGCTGCCCGTGCCC TGG
GACCAGGATGGGCACCACCC CGG
GCCGTCCAGCTCGACCAGGA TGG
GGCCGGACACGCTGAACTTG TGG
TAACAGGGTAATGTCGAGGC CGG
AGGTGAGCTCTTATTTGCGT AGG
CTTCAGGGTCAGCTTGCCGT AGG
GGGCACGGGCAGCTTGCCGG TGG

图11B

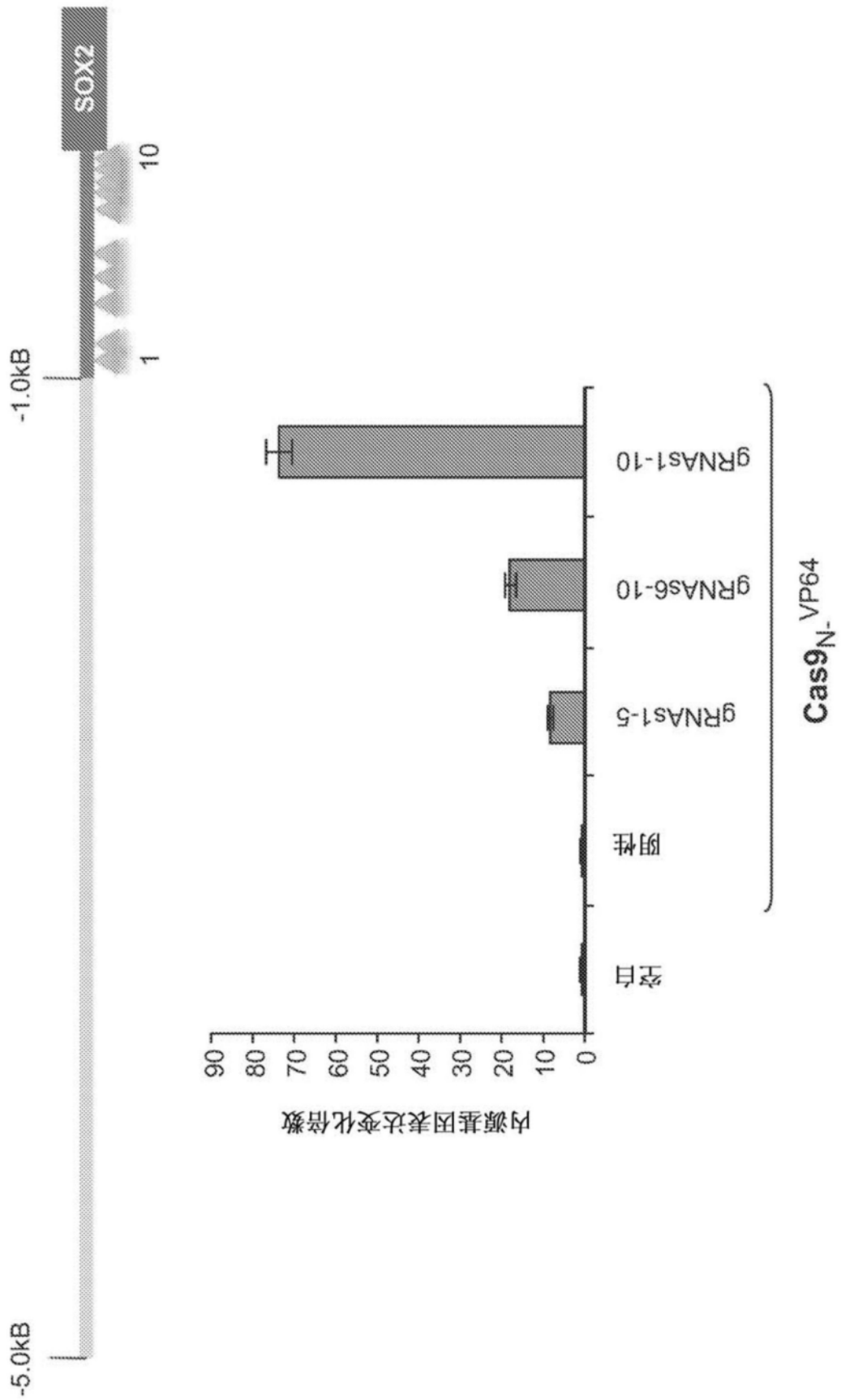

-5.0kB
-1.0kB
1
10
SOX2
内源基因表达变化倍数
90
80
70
60
50
40
30
20
10
0
空白
阴性
gRNAs1-5
gRNAs6-10
gRNAs1-10
Cas9N-VP64

图12A

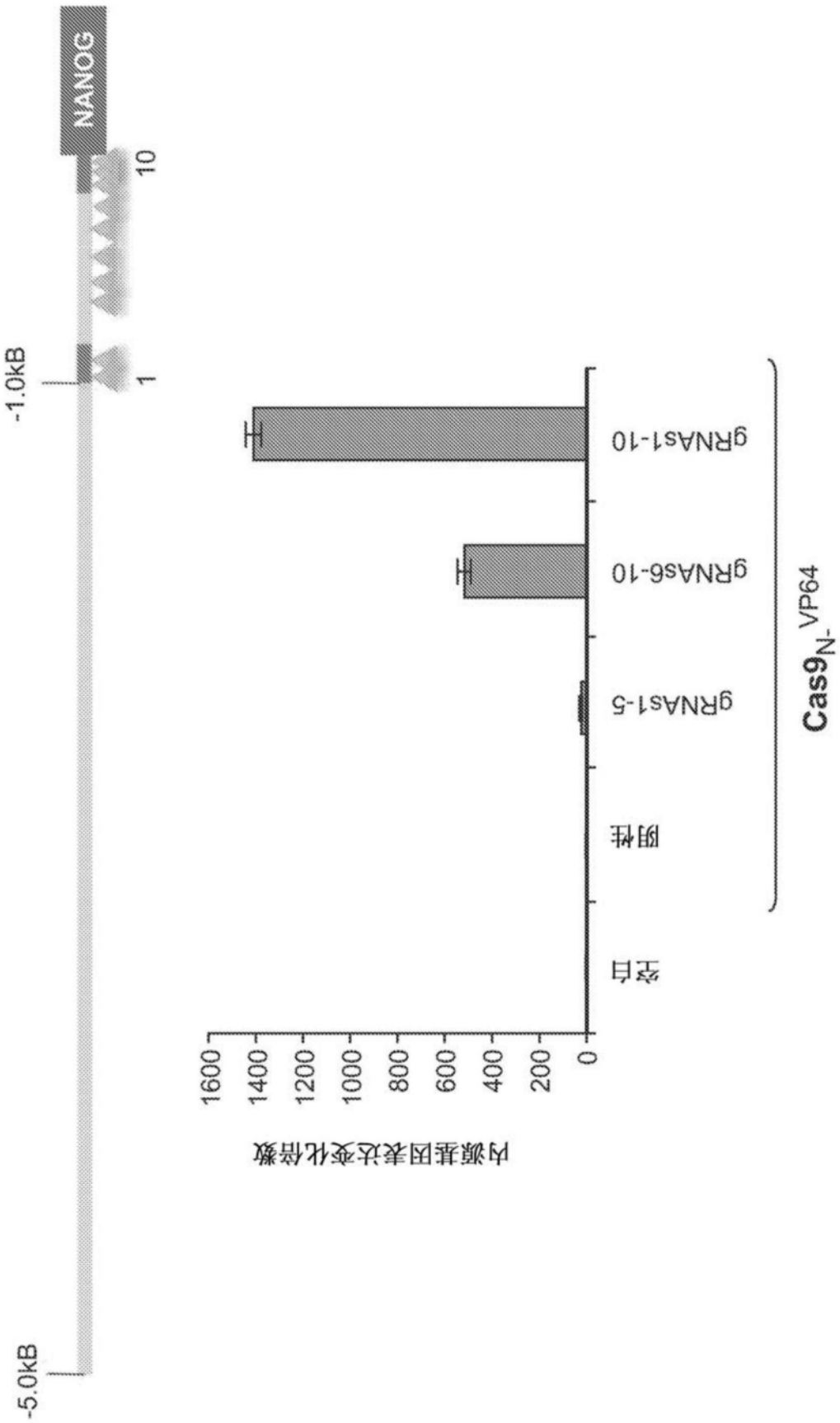

-5.0kB
-1.0kB
1
10
NANOG
内源基因表达变化倍数
1600
1400
1200
1000
800
600
400
200
0
空白
阴性
gRNAs1-5
gRNAs6-10
gRNAs1-10
Cas9N-VP64

图12B

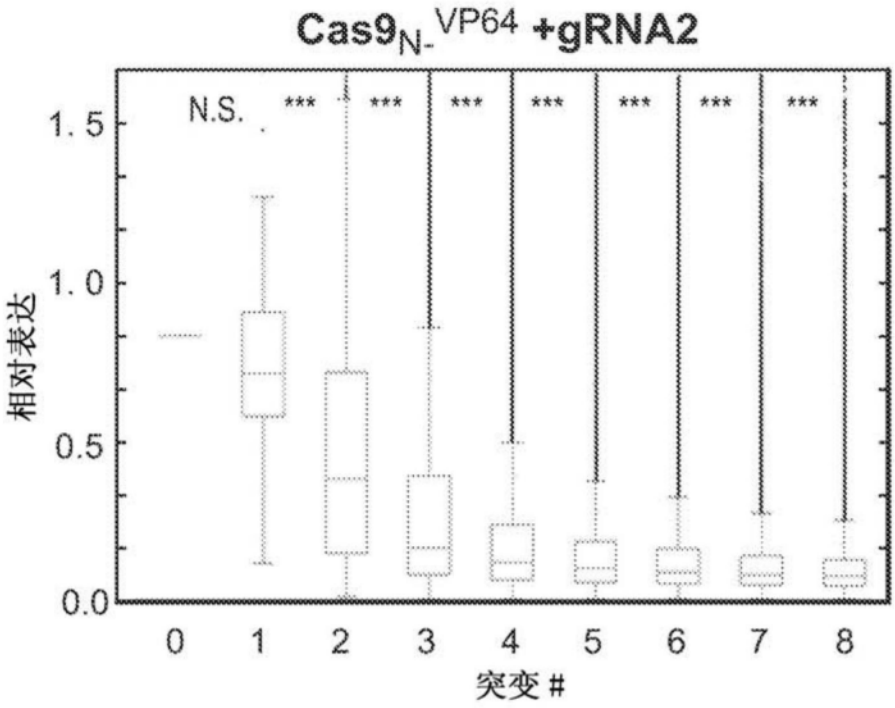


图13A

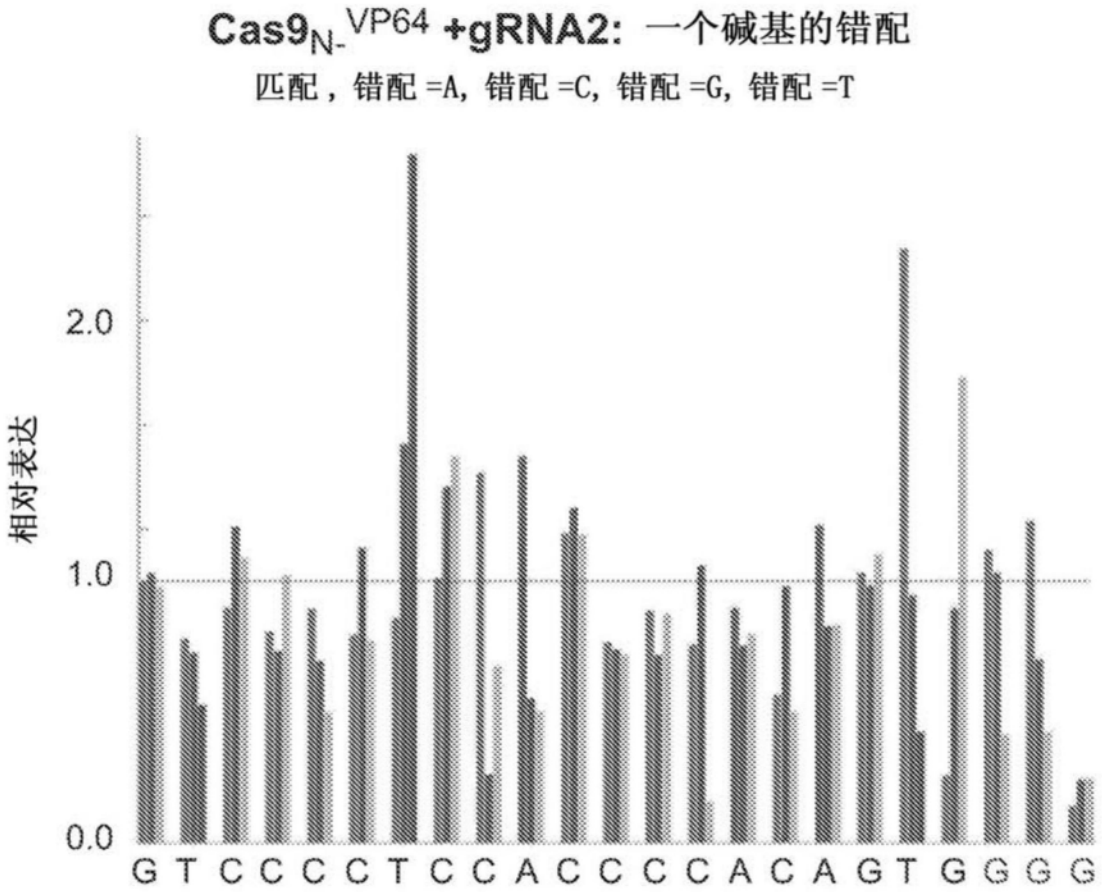

Cas9N-VP64 +gRNA2: 一个碱基的错配
匹配，错配 =A，错配 =C，错配 =G，错配 =T
相对表达
2.0
1.0
0.0
G T C C C C T C C A C C C C A C A G T G G G G

图13B

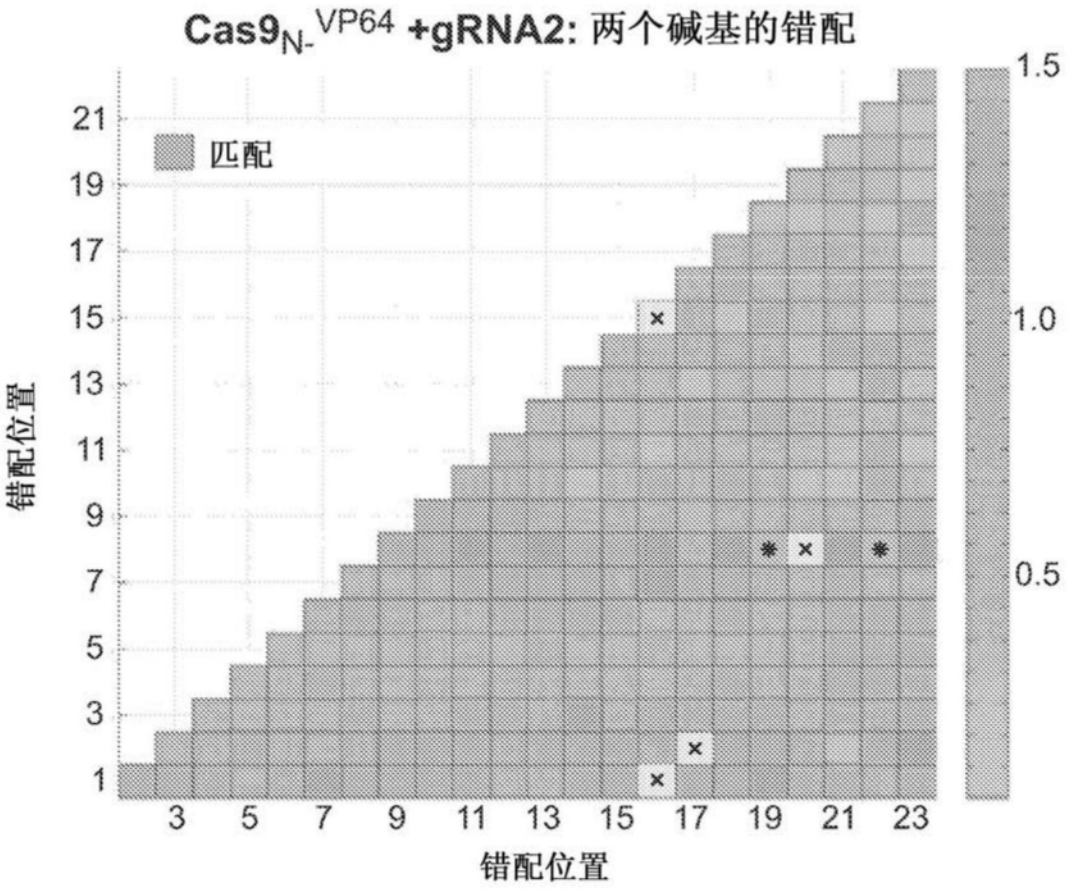

Cas9N-VP64 +gRNA2: 两个碱基的错配
匹配
错配位置
错配位置
1.5
1.0
0.5

图13C

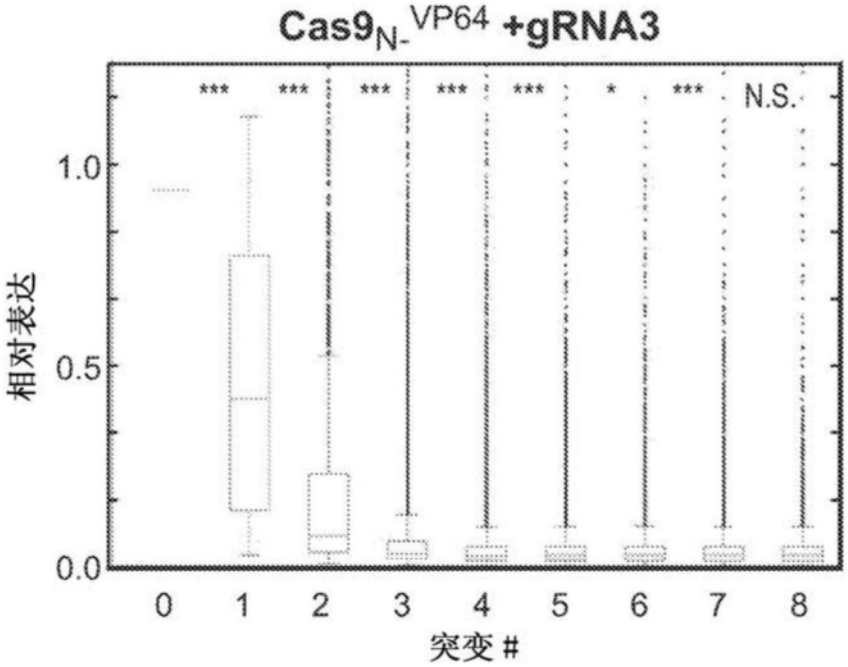

Cas9N-VP64 +gRNA3
N.S.
相对表达
突变 #

图13D

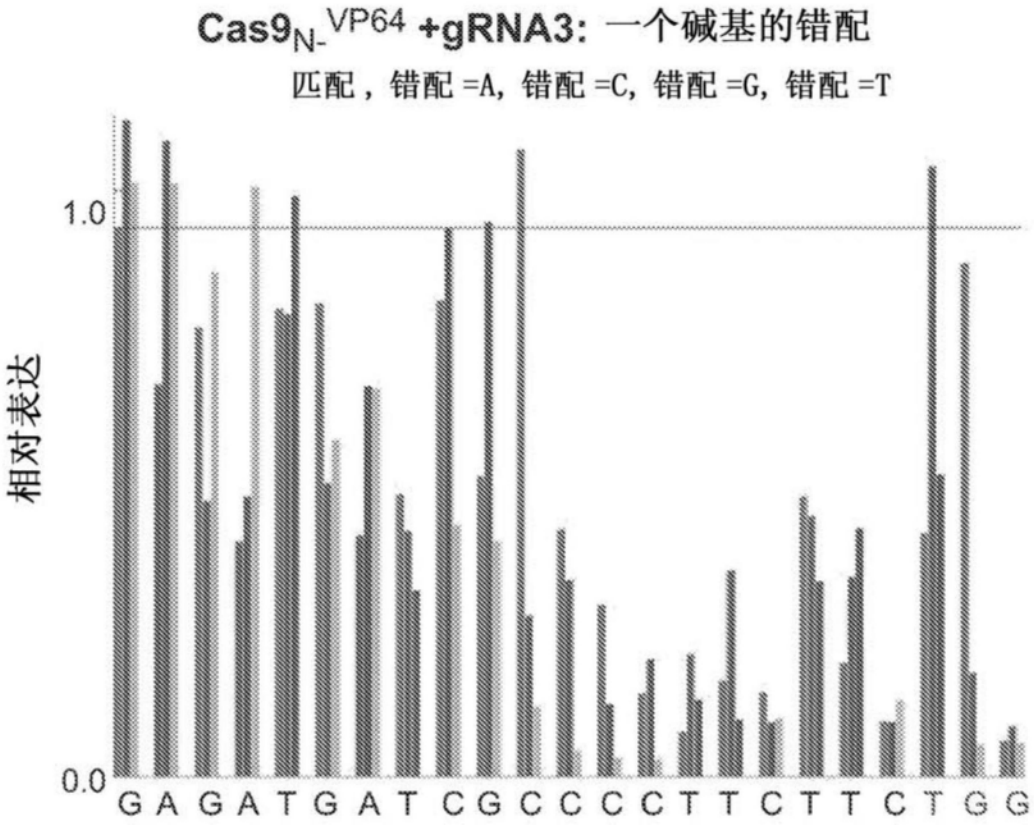

Cas9N-VP64 +gRNA3: 一个碱基的错配
匹配，错配 =A，错配 =C，错配 =G，错配 =T
相对表达
1.0
0.0
G A G A T G A T C G C C C C T T C T T C T G G

图13E

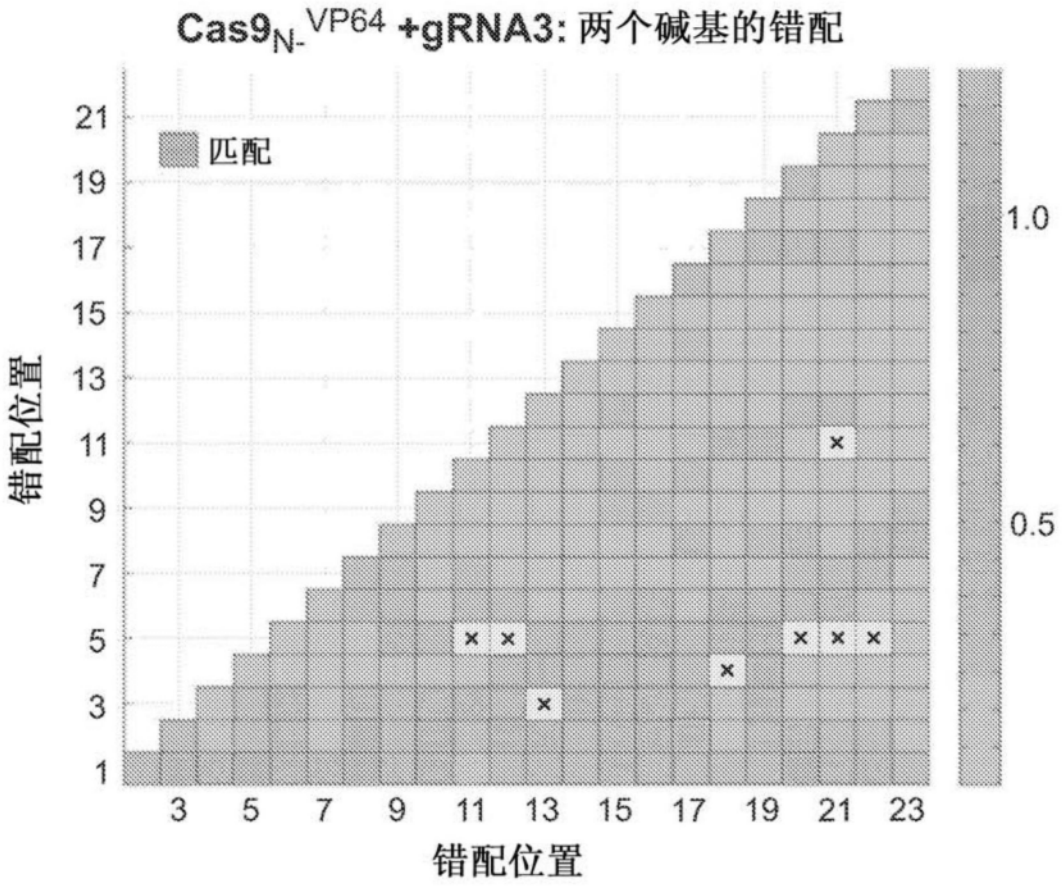

Cas9N-VP64 +gRNA3: 两个碱基的错配
匹配
错配位置
21
19
17
15
13
11
9
7
5
3
1
3 5 7 9 11 13 15 17 19 21 23
错配位置
1.0
0.5

图13F


靶标 : GAGATGATCGGCCCCTTCTTTC TGG
gRNA3 : GAGATGATCGGCCCCTTCTTTC
gRNA3mut : GTGATGACCGGCCGTTCTTTC

图14A

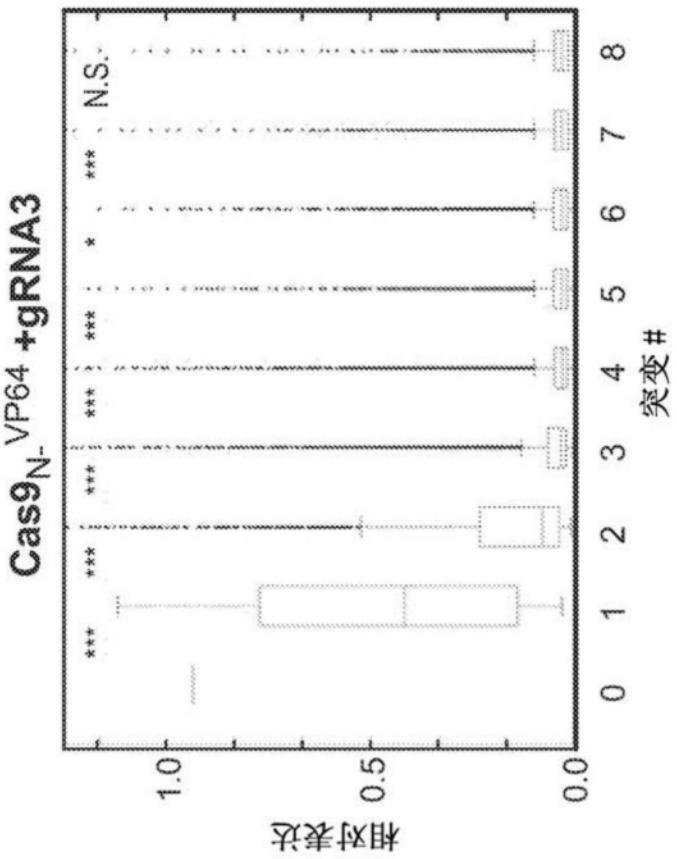


图14B

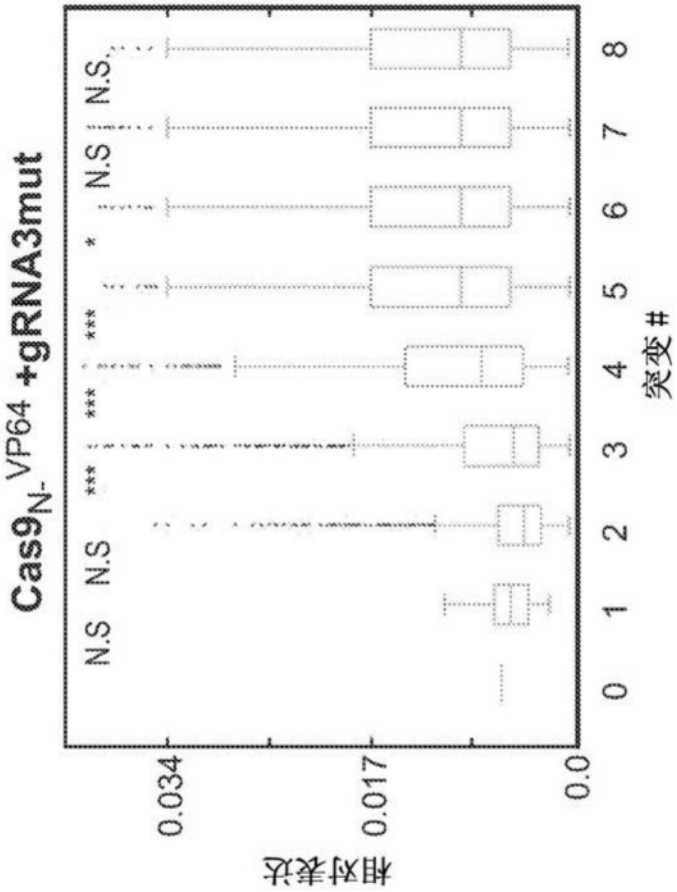

Cas9N-VP64 +gRNA3mut
N.S
N.S
***
***
***
*
N.S
N.S.
0.034
0.017
0.0
相对表达
0
1
2
3
4
5
6
7
8
突变#

图14C

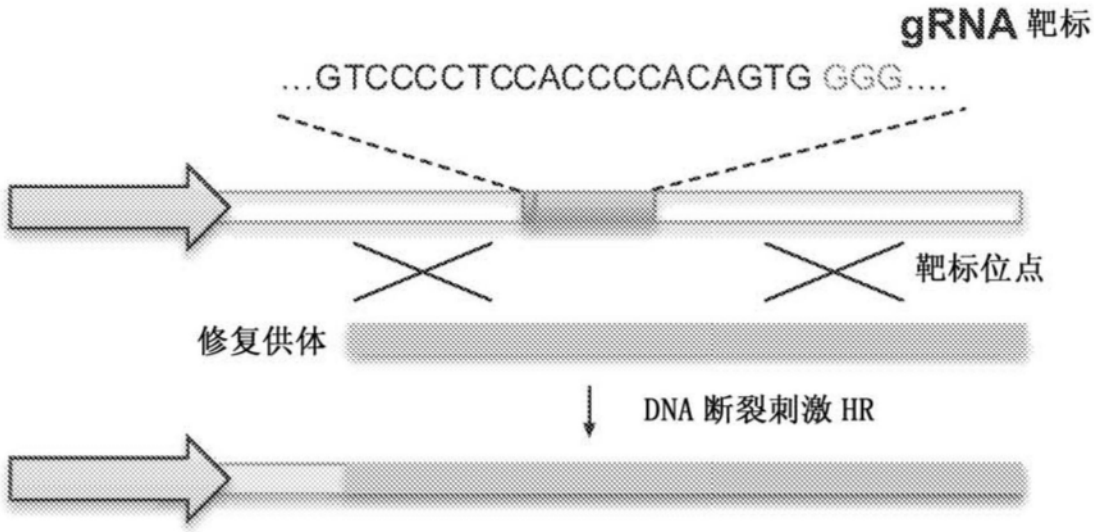

gRNA 靶标
...GTCCCCTCCACCCCACAGTG GGG....
靶标位点
修复供体
DNA 断裂刺激 HR

图15A

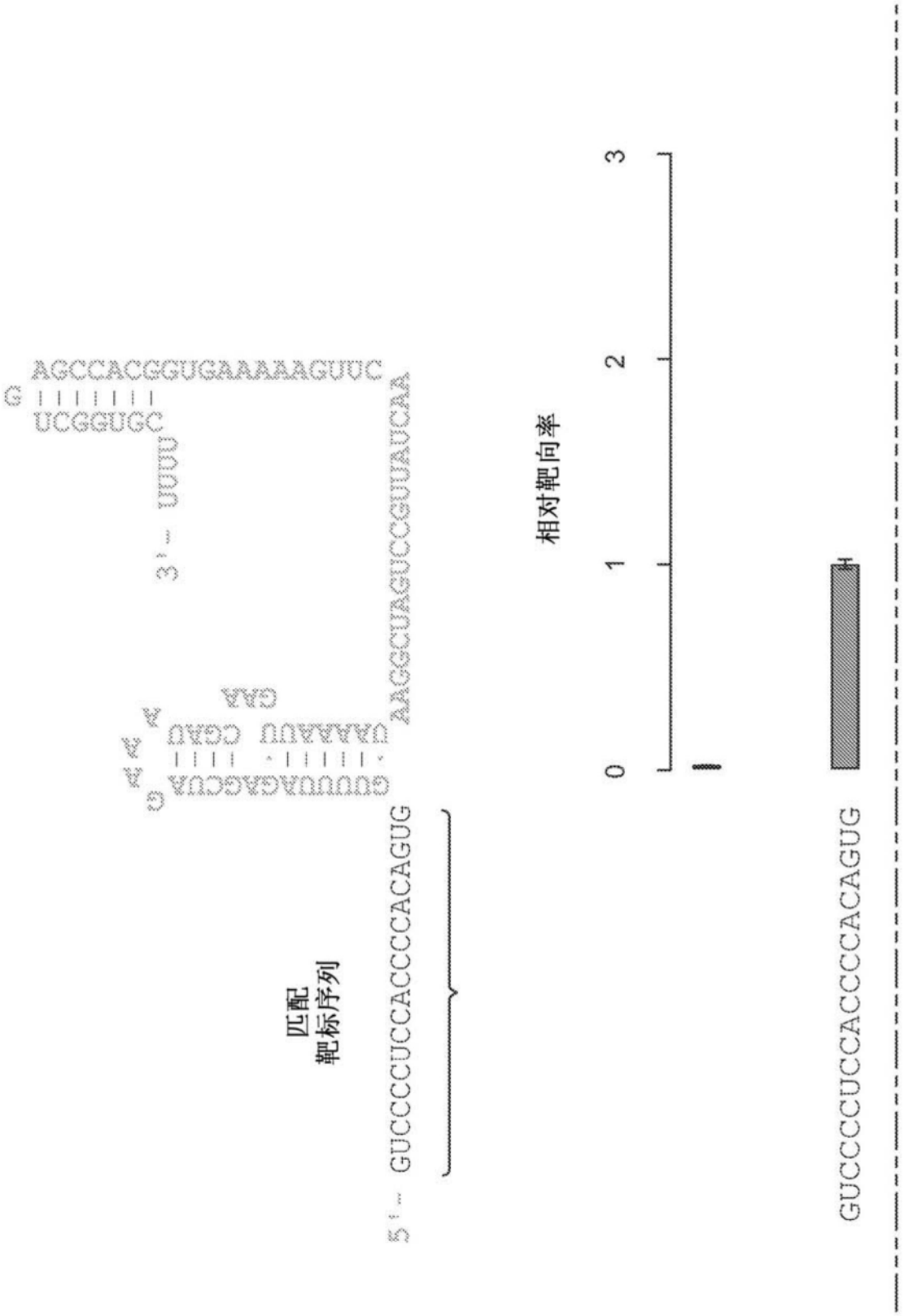

AGCCACGGUGAAAAAGUUC
G
UCGGUGC
3'- UUUU
GAA
AAGGCUAGUCCGUUAUCAA
5'- GUCCCCUCCACCCCACAGUG
匹配
靶标序列
相对靶向率
0
1
2
3
GUCCCCUCCACCCCACAGUG

图15B-1

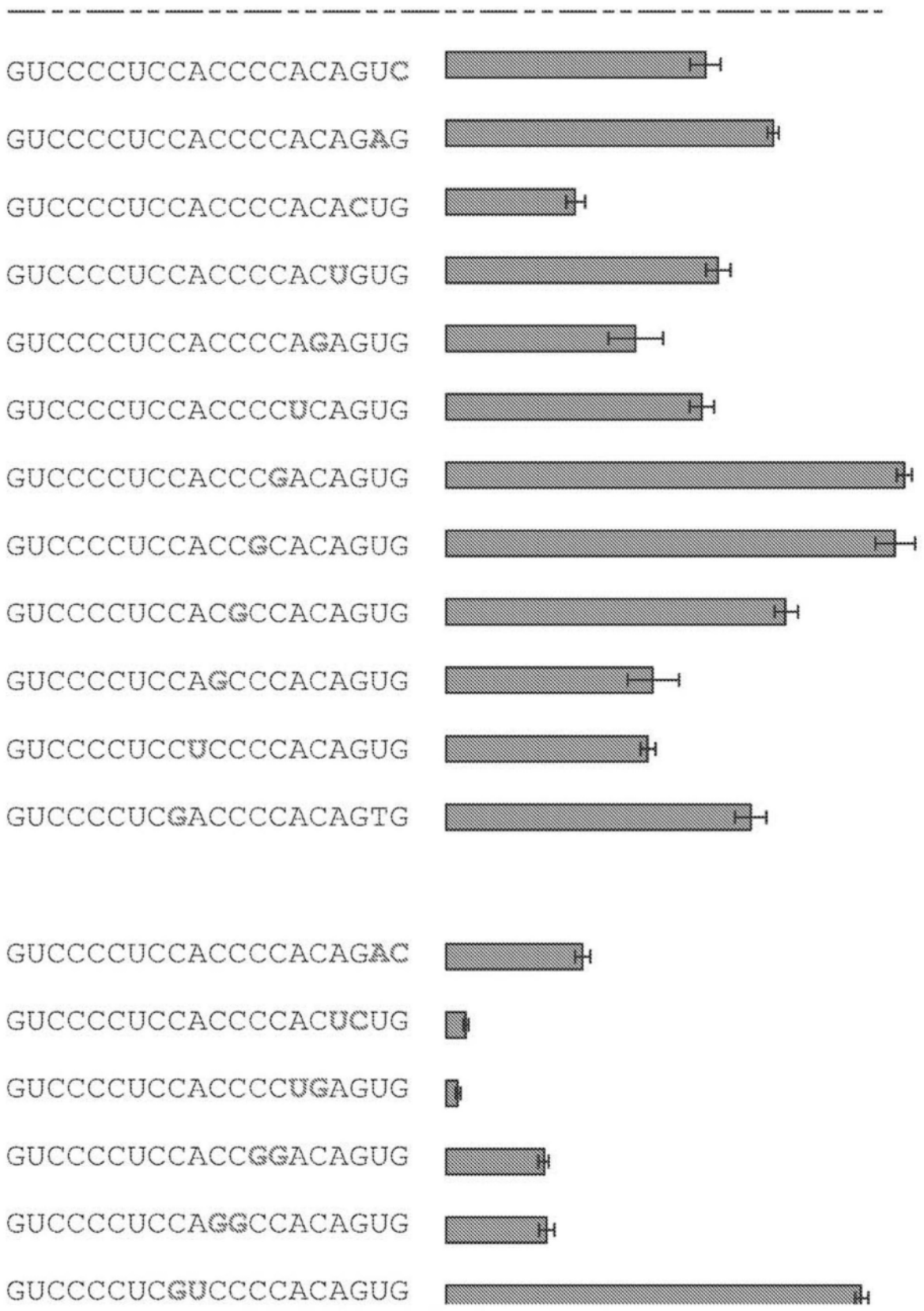

GUCCCCUCCACCCCACAGUC
GUCCCCUCCACCCCACAGAG
GUCCCCUCCACCCCACACUG
GUCCCCUCCACCCCACUGUG
GUCCCCUCCACCCCAGAGUG
GUCCCCUCCACCCCUCAGUG
GUCCCCUCCACCCGACAGUG
GUCCCCUCCACCGCACAGUG
GUCCCCUCCACGCCACAGUG
GUCCCCUCCAGCCCACAGUG
GUCCCCUCCUCCCCACAGUG
GUCCCCUCGACCCCACAGTG
GUCCCCUCCACCCCACAGAC
GUCCCCUCCACCCCACUCUG
GUCCCCUCCACCCCUGAGUG
GUCCCCUCCACCGGACAGUG
GUCCCCUCCAGGCCACAGUG
GUCCCCUCGUCCCCACAGUG

图15B-2

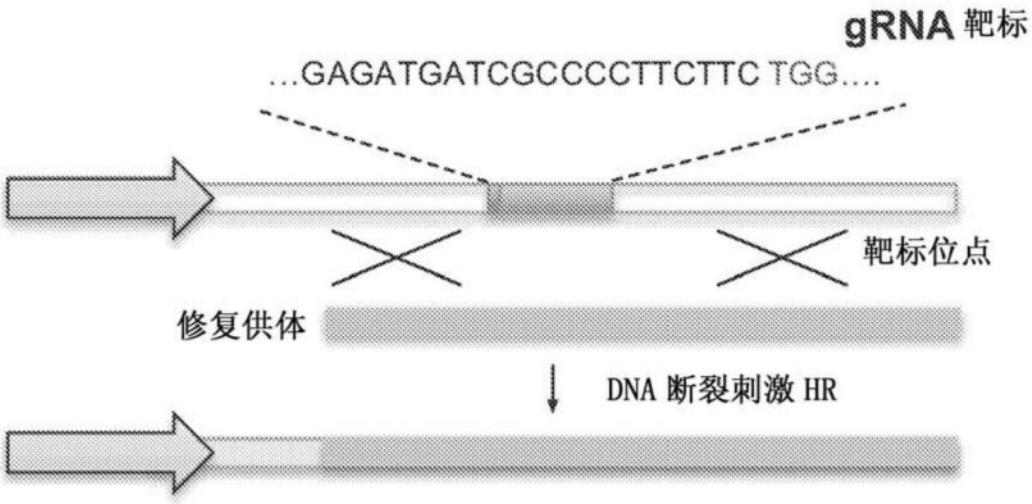

gRNA 靶标
...GAGATGATCGCCCCTTCTTC TGG....
靶标位点
修复供体
DNA 断裂刺激 HR

图15C

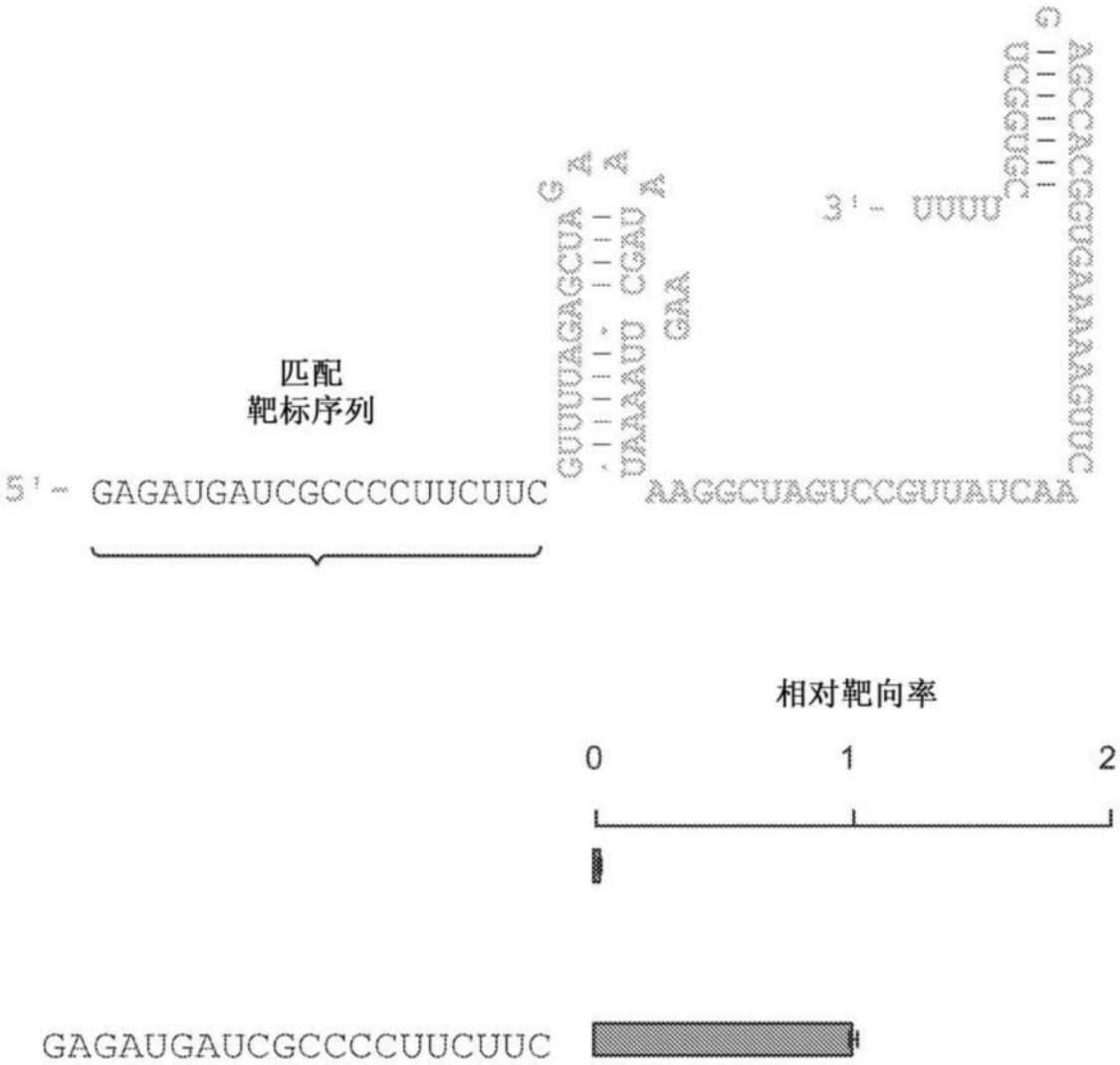

匹配
靶标序列
5'- GAGAUGAUCGCCCCUUCUUC
3'- UUUU
相对靶向率
0
1
2
GAGAUGAUCGCCCCUUCUUC

图15D-1

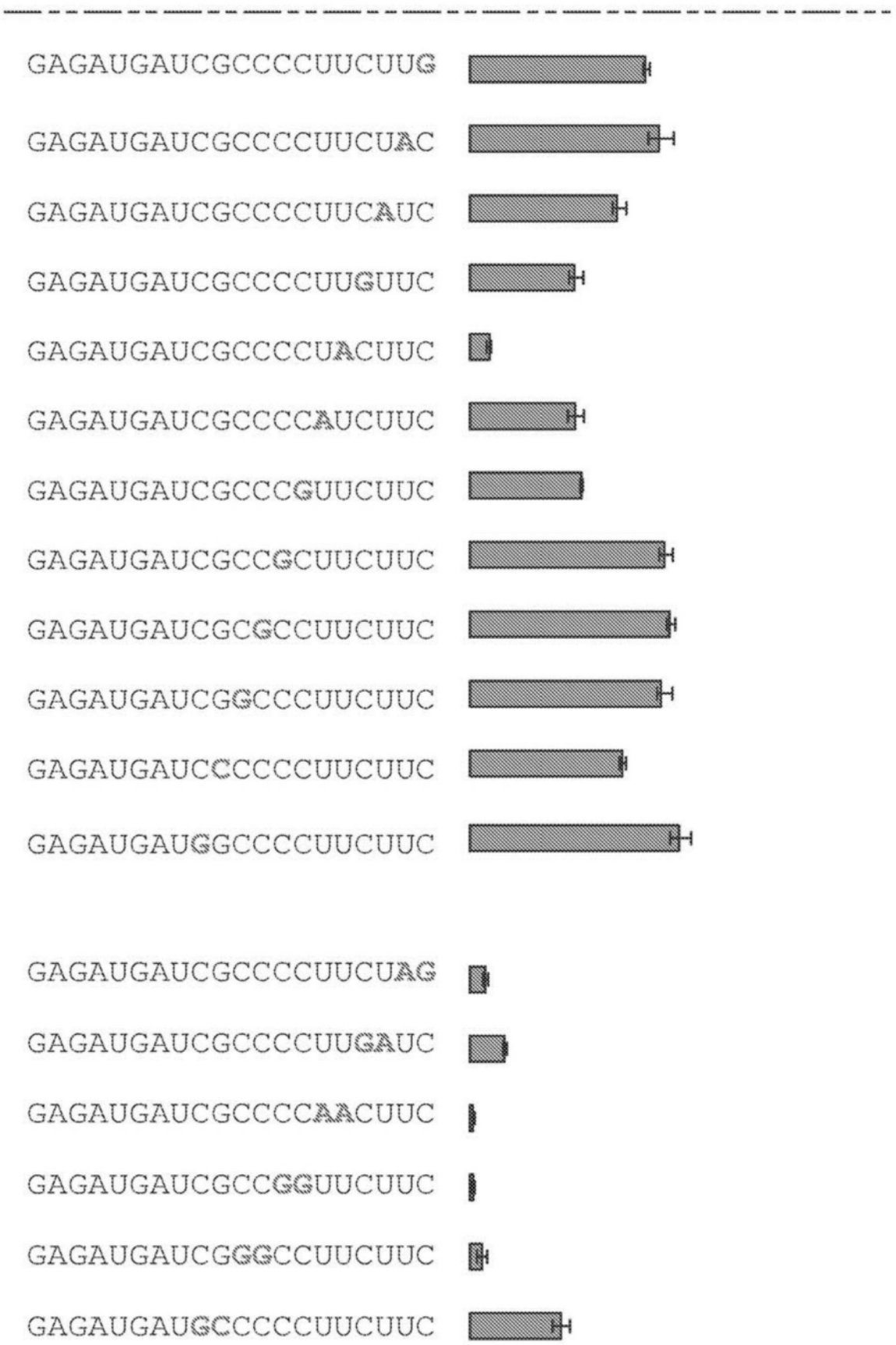

GAGAUGAUCGCCCCUUCUUG
GAGAUGAUCGCCCCUUCUAC
GAGAUGAUCGCCCCUUCAUC
GAGAUGAUCGCCCCUUGUUC
GAGAUGAUCGCCCCUACUUC
GAGAUGAUCGCCCCAUCUUC
GAGAUGAUCGCCCGUUCUUC
GAGAUGAUCGCCGCUUCUUC
GAGAUGAUCGCGCCUUCUUC
GAGAUGAUCGGCCCUUCUUC
GAGAUGAUCCCCCUUCUUC
GAGAUGAUGGCCCCUUCUUC
GAGAUGAUCGCCCCUUCUAG
GAGAUGAUCGCCCCUUGAUC
GAGAUGAUCGCCCCAACUUC
GAGAUGAUCGCCGGUUCUUC
GAGAUGAUCGGGCCUUCUUC
GAGAUGAUGCCCCCUUCUUC

图15D-2

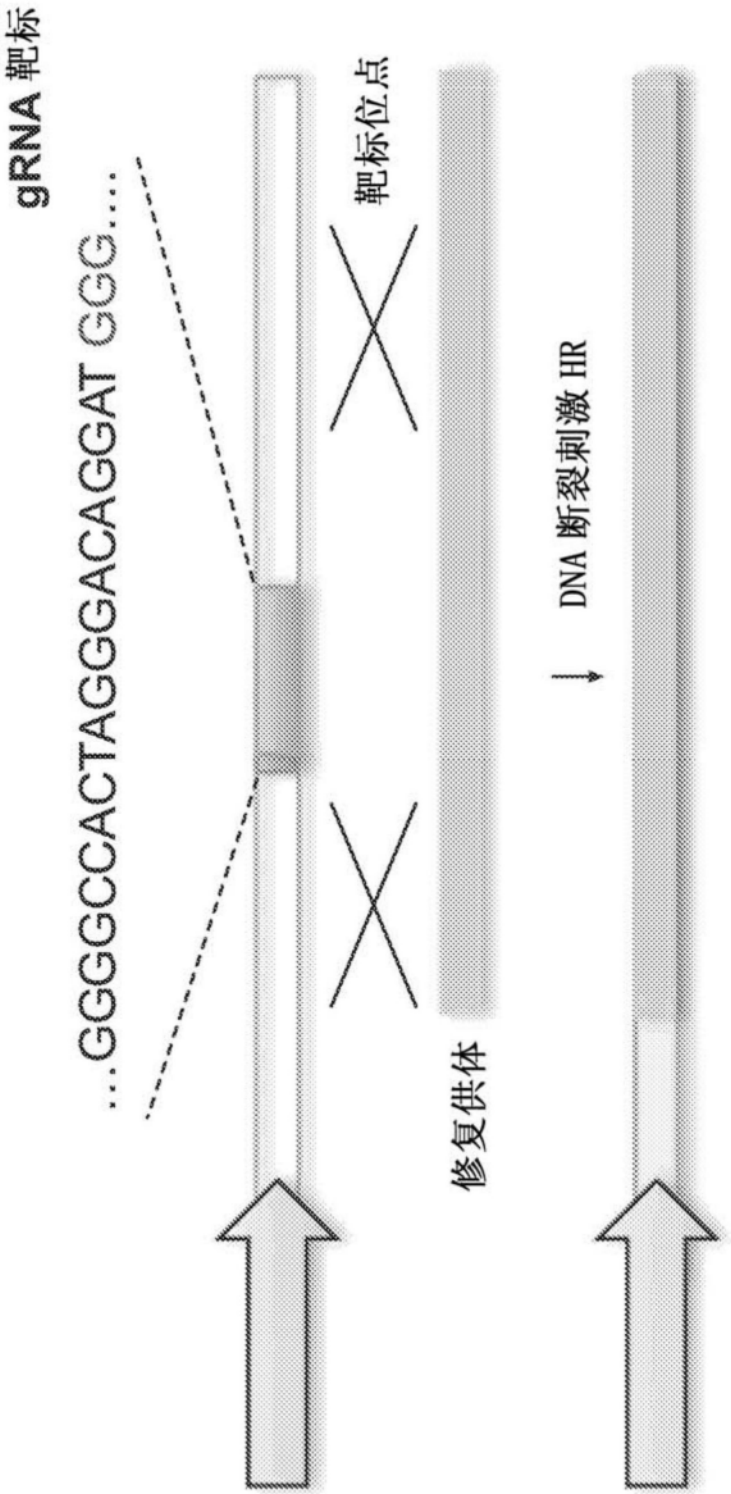


图16A

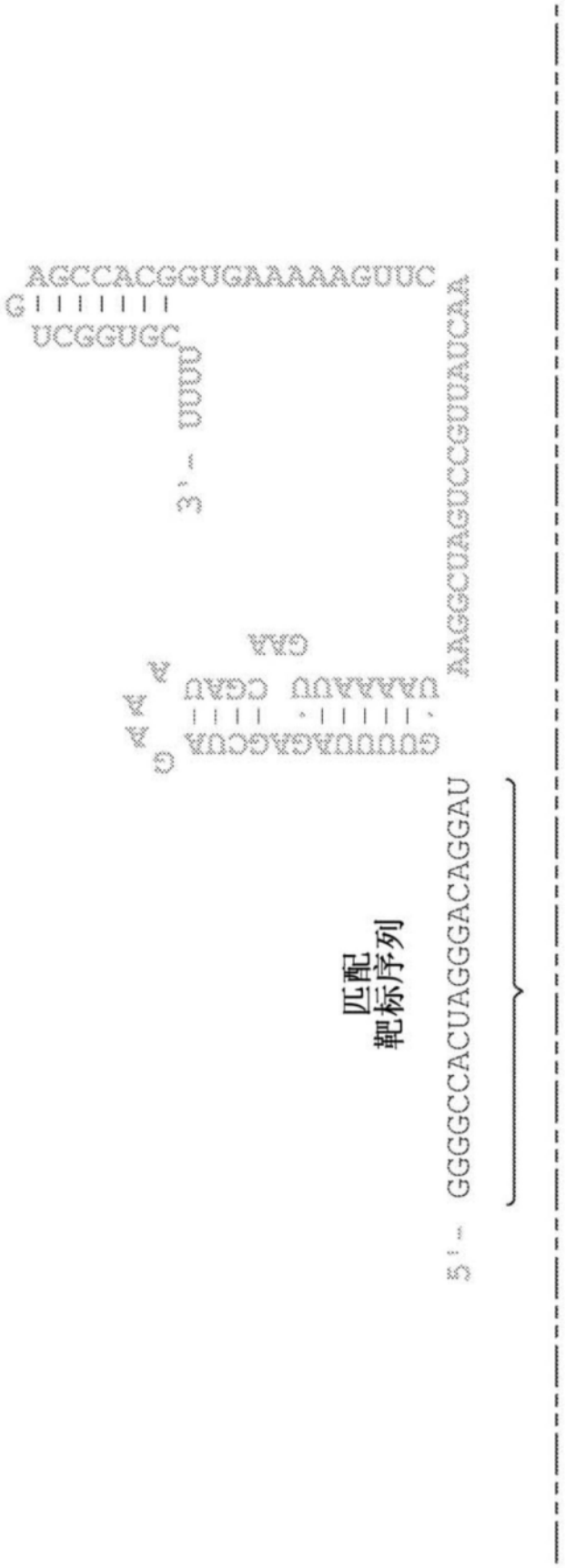

匹配
靶标序列
5'- GGGGCCACUAGGGACAGGAU
AAGGCUAGUCCGUUAUCAA
3'- UUUU
GAA
AGCCACGGUGAAAAAGUUC
UCGGUGC

图16B-1

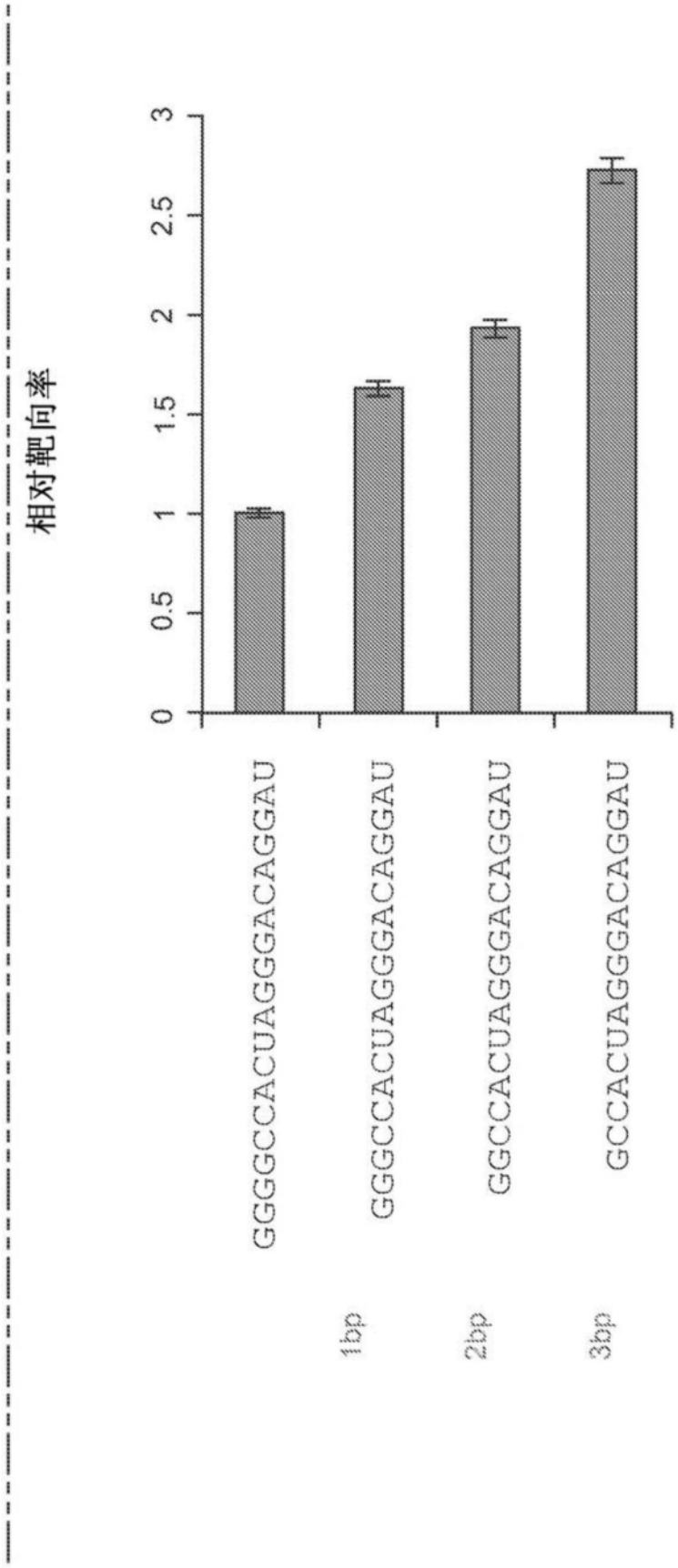

相对靶向率
0
0.5
1
1.5
2
2.5
3
GGGGCCACUAGGGACAGGAU
GGGCCACUAGGGACAGGAU
GGCCACUAGGGACAGGAU
GCCACUAGGGACAGGAU
1bp
2bp
3bp

图16B-2

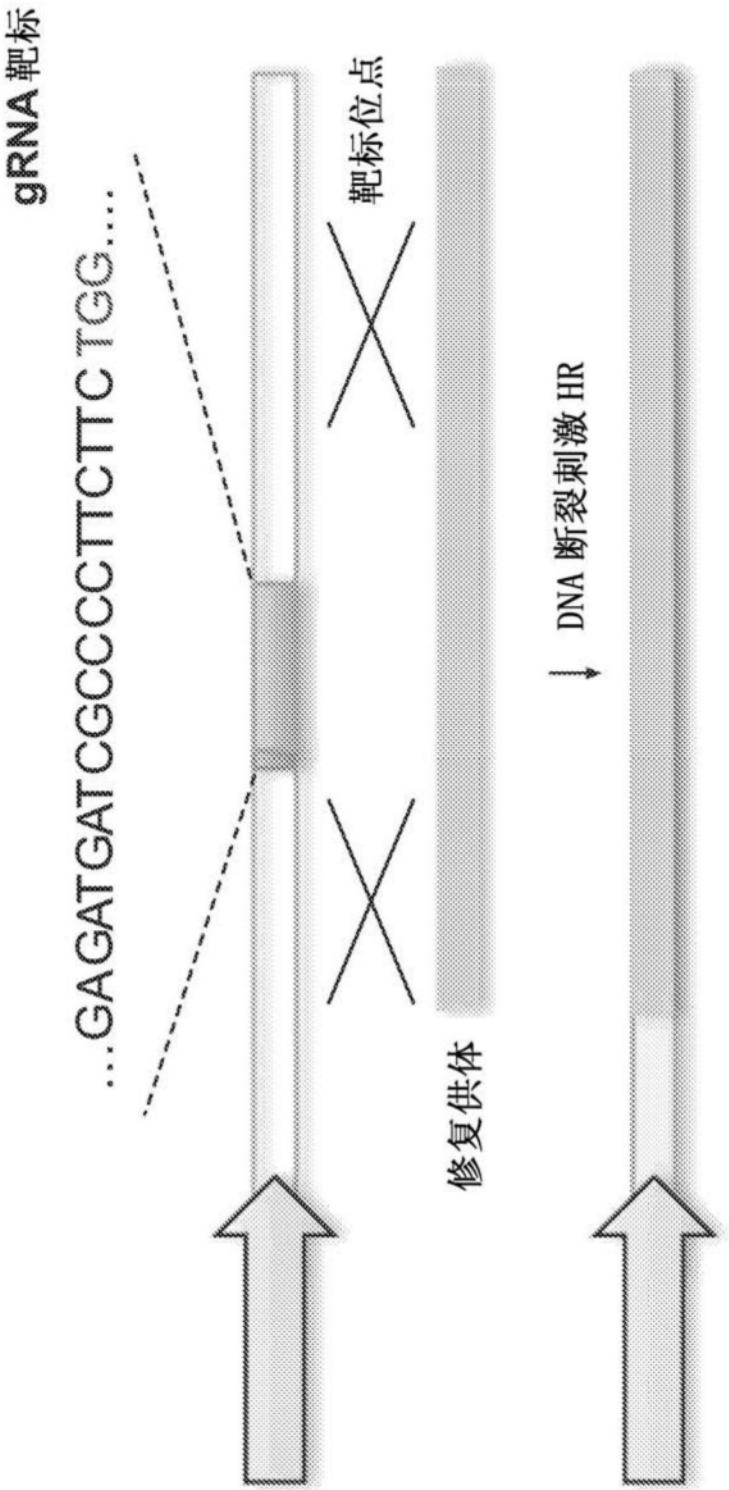


图16C

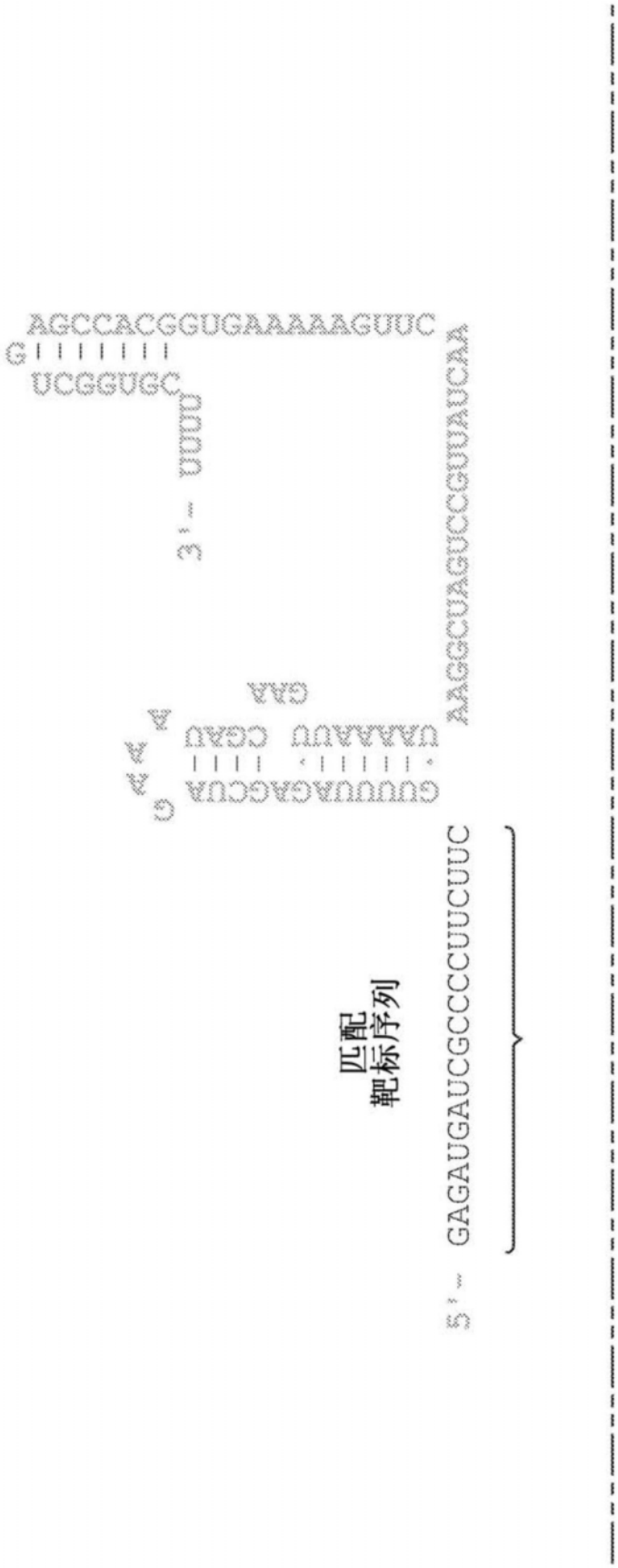

匹配
靶标序列
5'- GAGAUGAUCGCCCCUUCUUC
AAGGCUAGUCCGUUAUCAA
3'- UUUU
AGCCACGGUGAAAAAGUUC
UCGGUGC
GAA

图16D-1

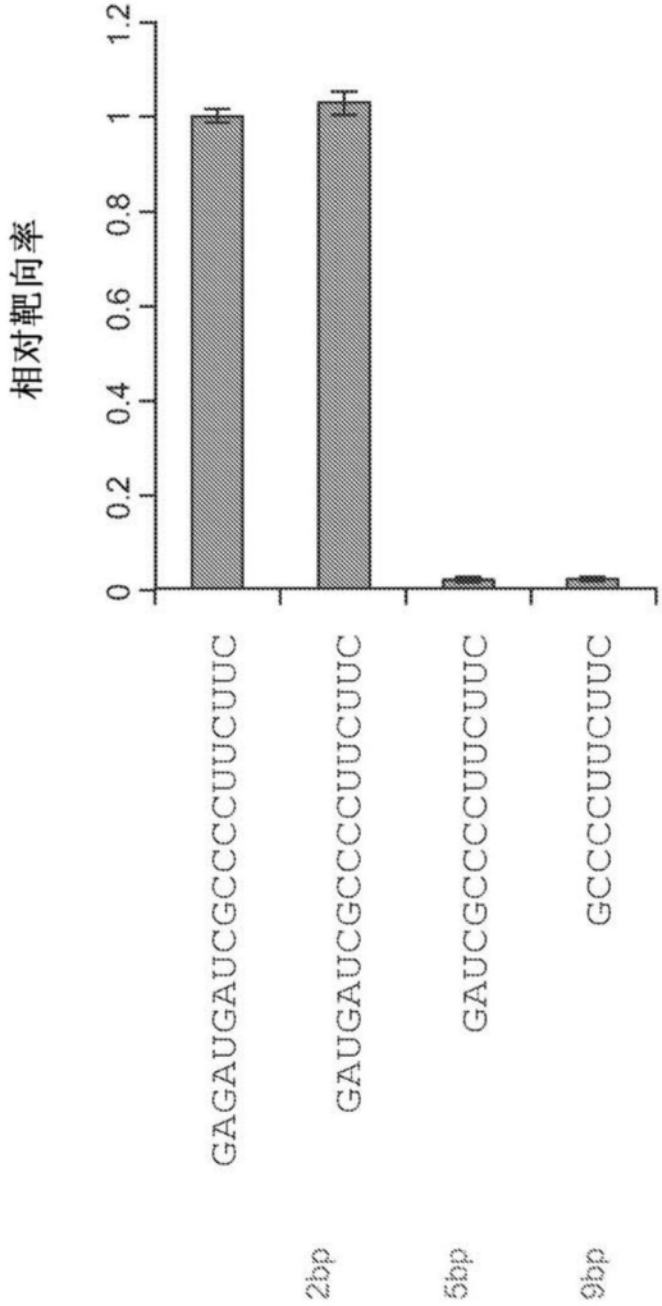

相对靶向率
0
0.2
0.4
0.6
0.8
1
1.2
GAGAUGAUCGCCCCUUCUUC
GAUGAUCGCCCCUUCUUC
GAUCGCCCCUUCUUC
GCCCCUUCUUC
2bp
5bp
9bp

图16D-2

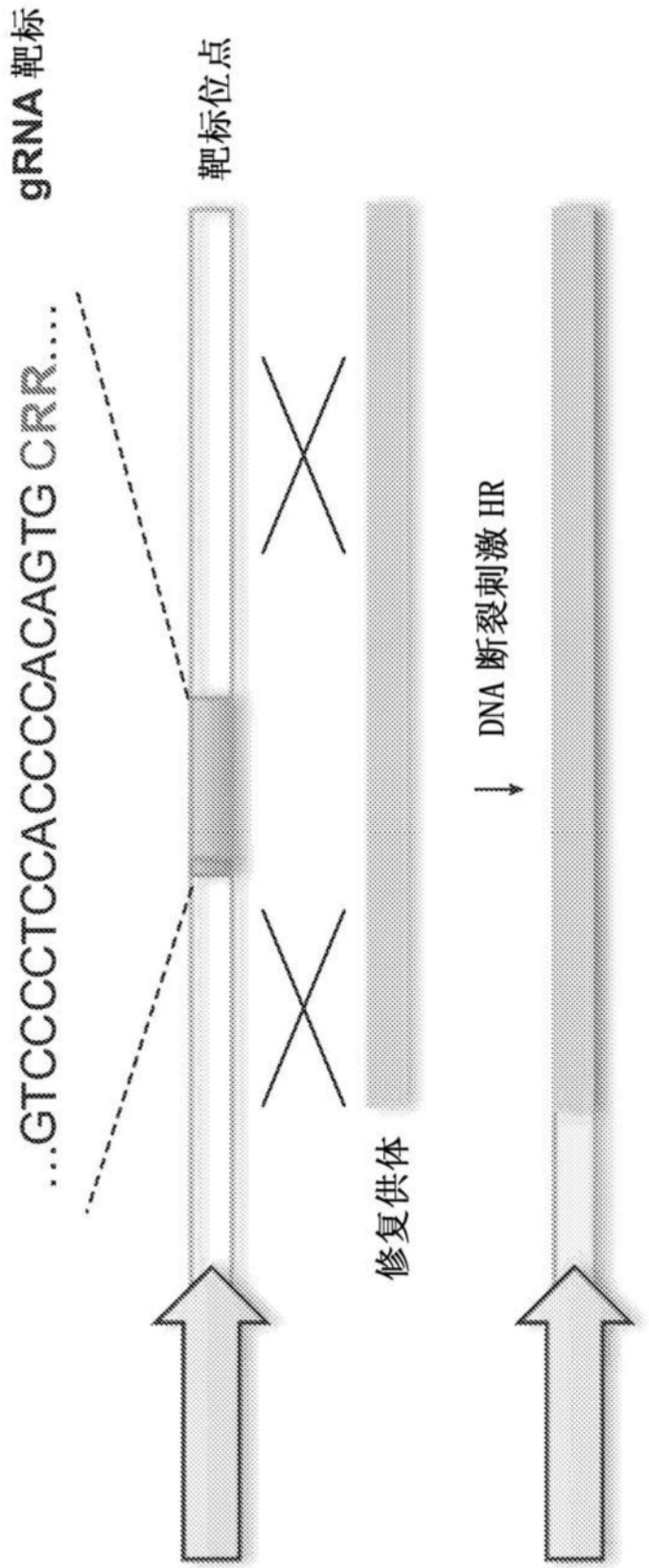


图17A

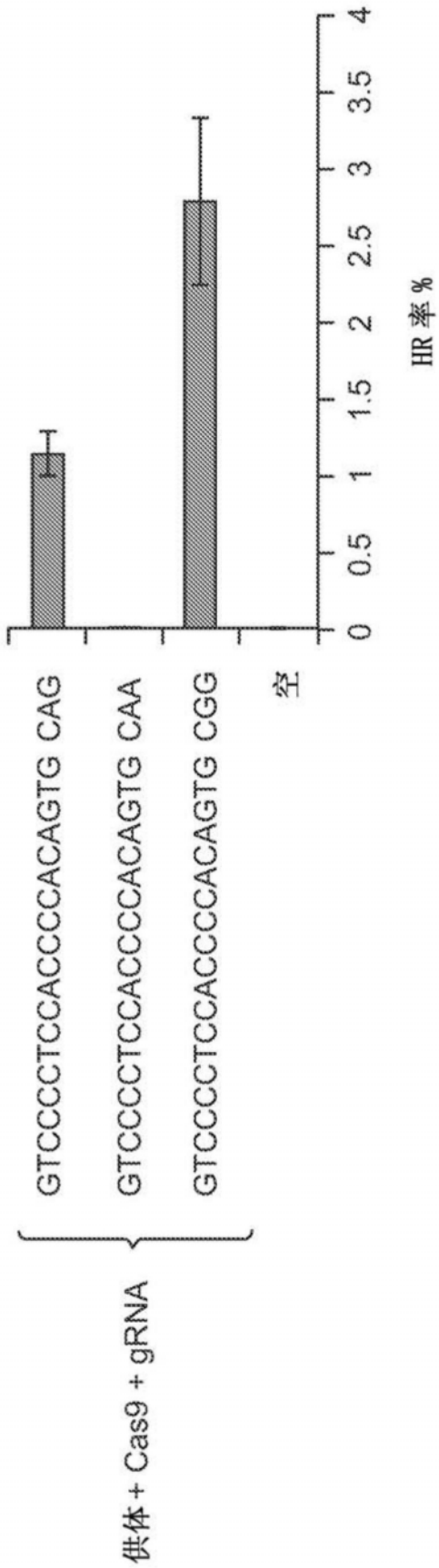

供体 + Cas9 + gRNA
GTCCCCTCCACCCCACAGTG CAG
GTCCCCTCCACCCCACAGTG CAA
GTCCCCTCCACCCCACAGTG CGG
空
0
0.5
1
1.5
2
2.5
3
3.5
4
HR 率 %

图17B

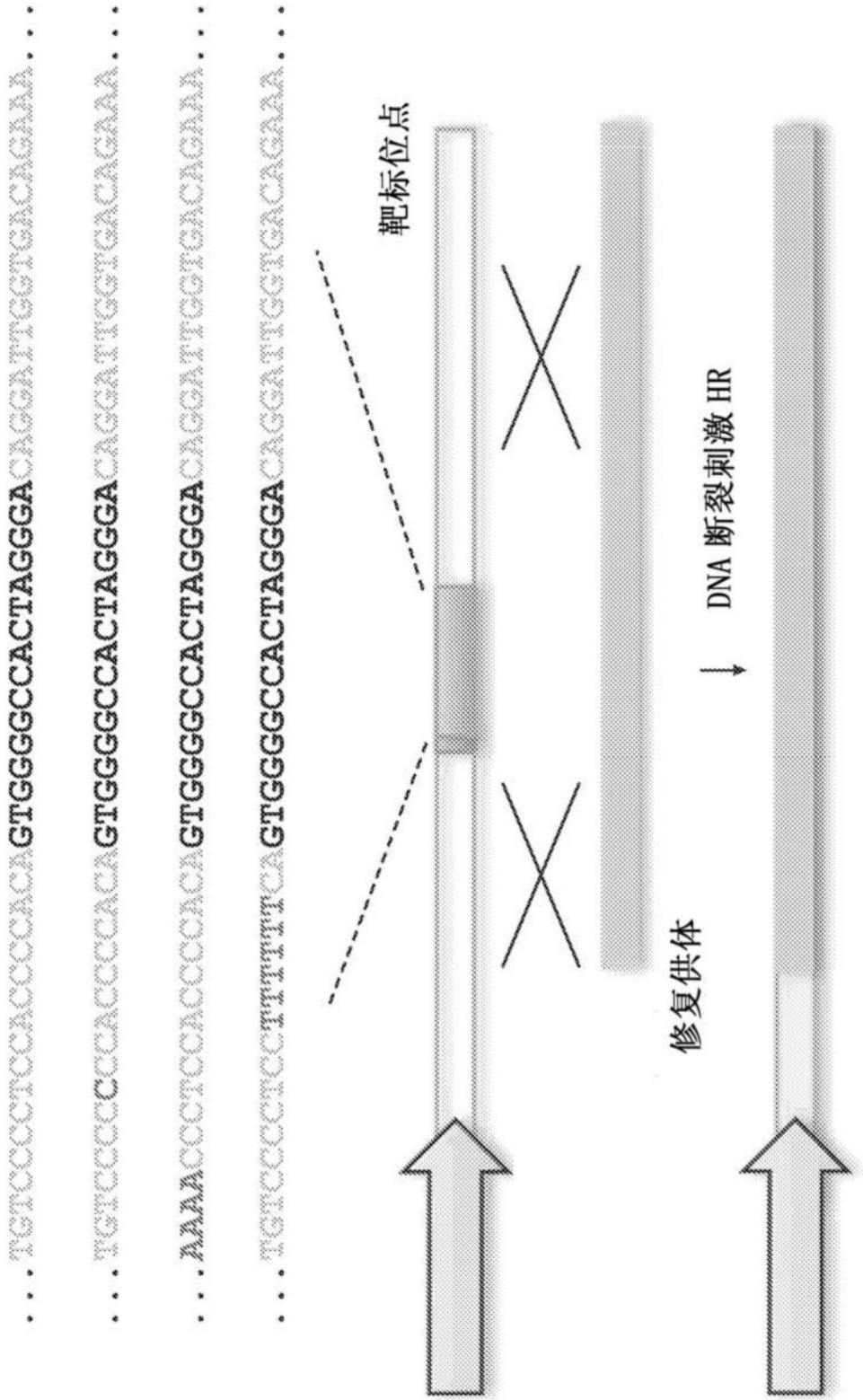

...TGTCCCCTCCACCCCACACAGTGGGGCCACTAGGGACAGGATTGGTGACAGAAA...
...TGTCCCCCCCACCCCACACAGTGGGGCCACTAGGGACAGGATTGGTGACAGAAA...
...AAAACCCTCCACCCCACACAGTGGGGCCACTAGGGACAGGATTGGTGACAGAAA...
...TGTCCCCTCCTTTTTTTTCAGTGGGGCCACTAGGGACAGGATTGGTGACAGAAA...
靶标位点
修复供体
DNA 断裂刺激 HR

图18A

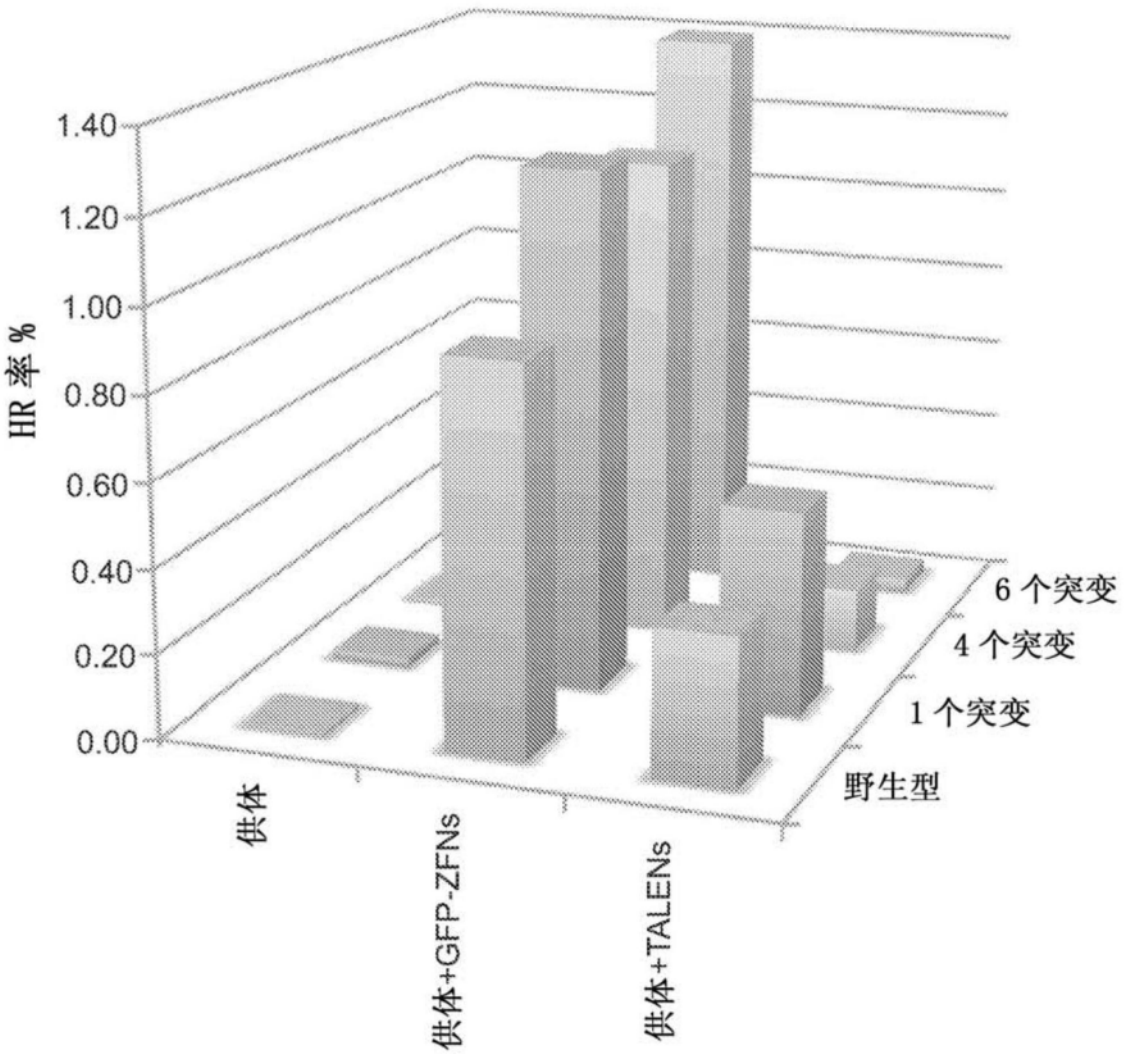

HR 率 %
1.40
1.20
1.00
0.80
0.60
0.40
0.20
0.00
供体
供体+GFP-ZFNs
供体+TALENs
6 个突变
4 个突变
1 个突变
野生型

图18B

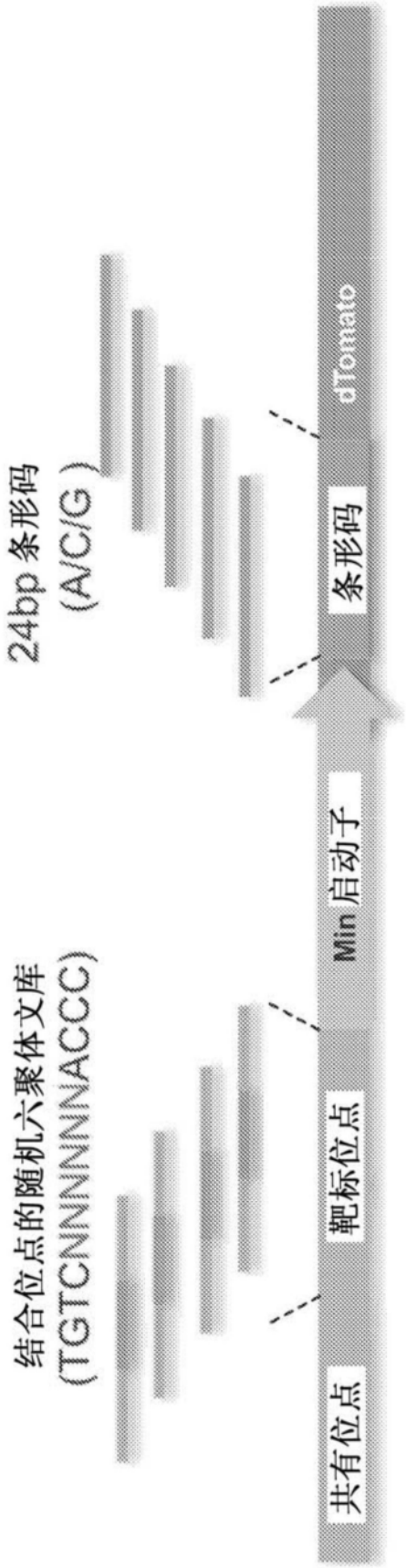

24bp 条形码
(A/C/G)
结合位点的随机六聚体文库
(TGTCNNNNNNACCC)
共有位点
靶标位点
Min 启动子
条形码
dTomato

图19A

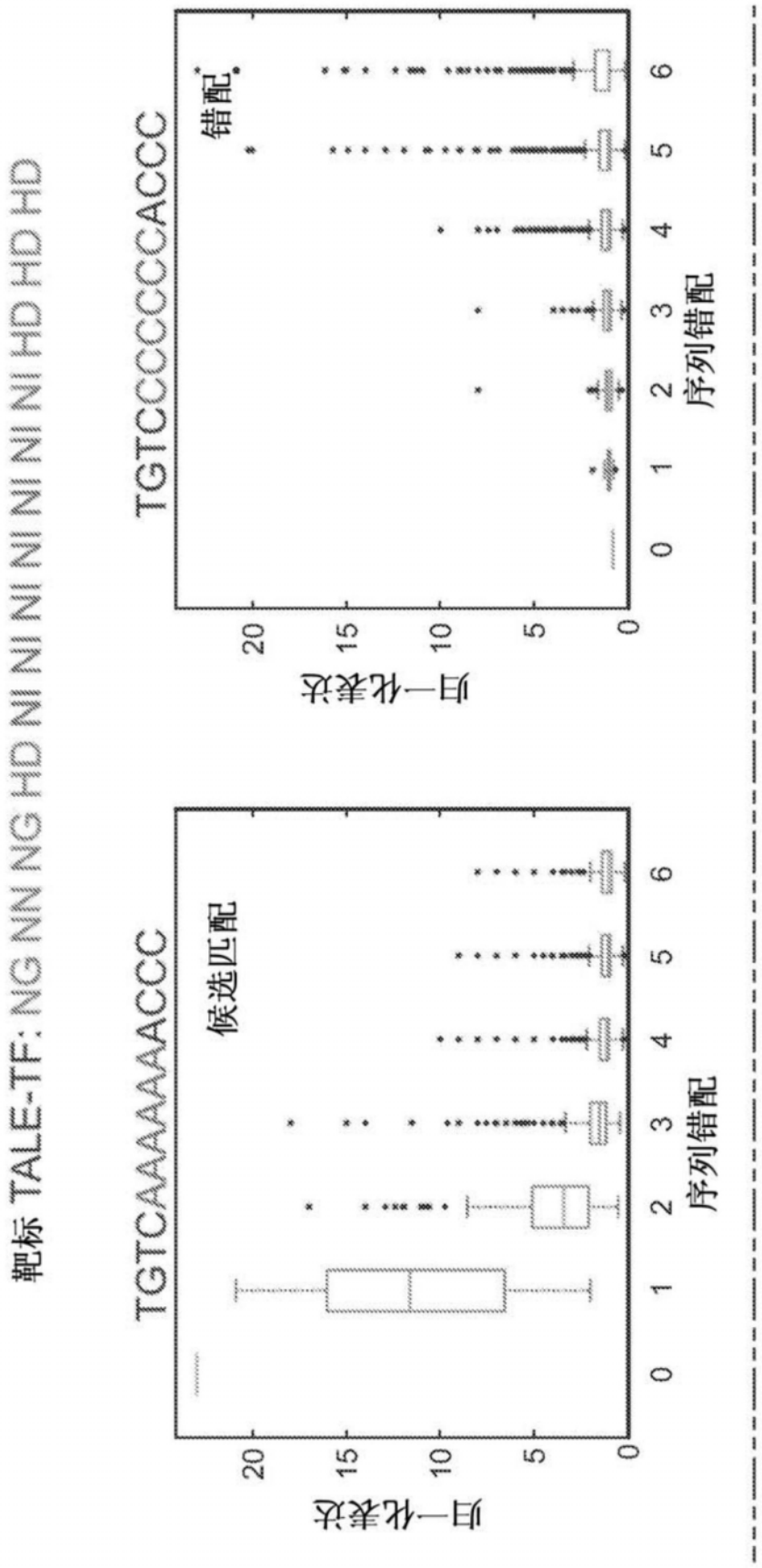

靶标 TALE-TF: NG NN NG HD NI NI NI NI NI NI HD HD HD
TGTCAAAAAACCC
候选匹配
归一化表达
序列错配
TGTCCCCCCCACCC
错配
归一化表达
序列错配

图19B-1

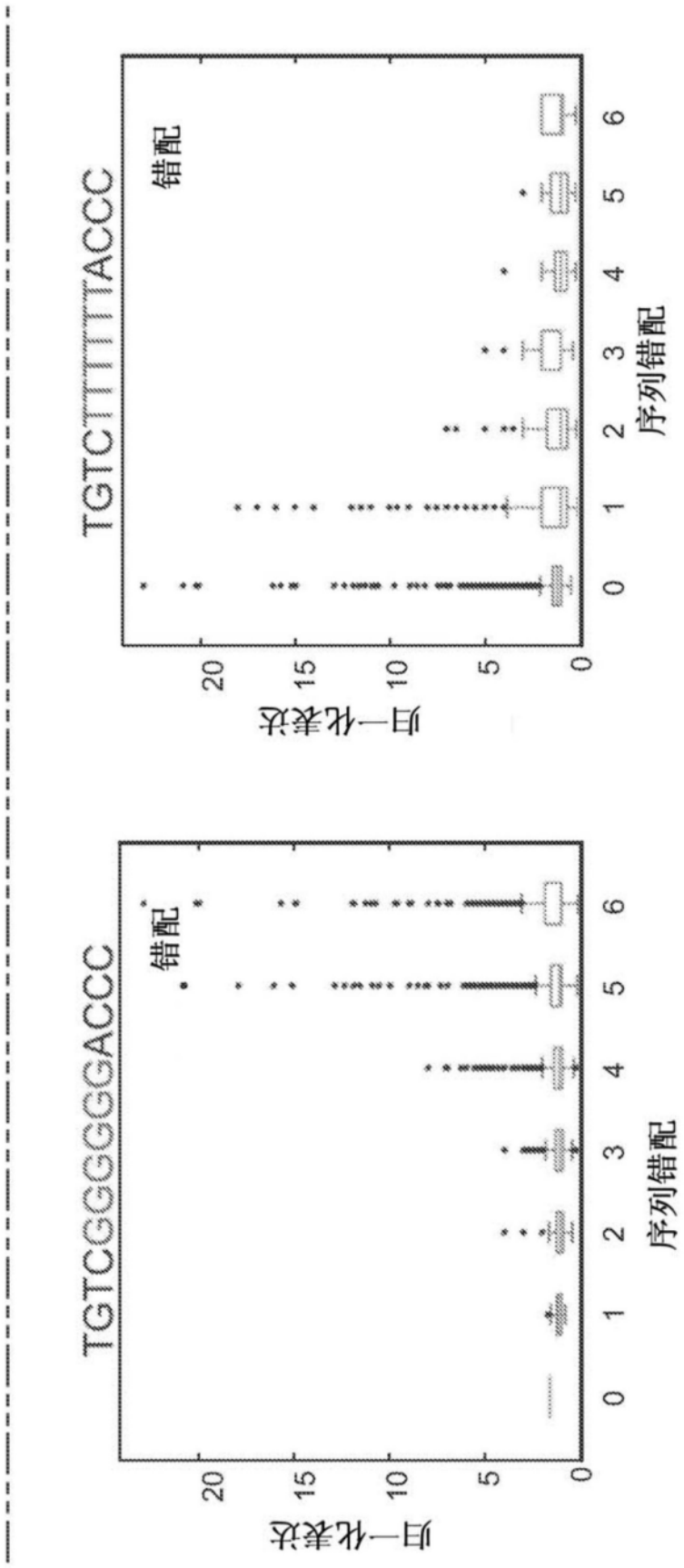

TGTCGGGGGGACCC
错配
归一化表达
序列错配
TGTCTTTTTTACCC
错配
归一化表达
序列错配

图19B-2

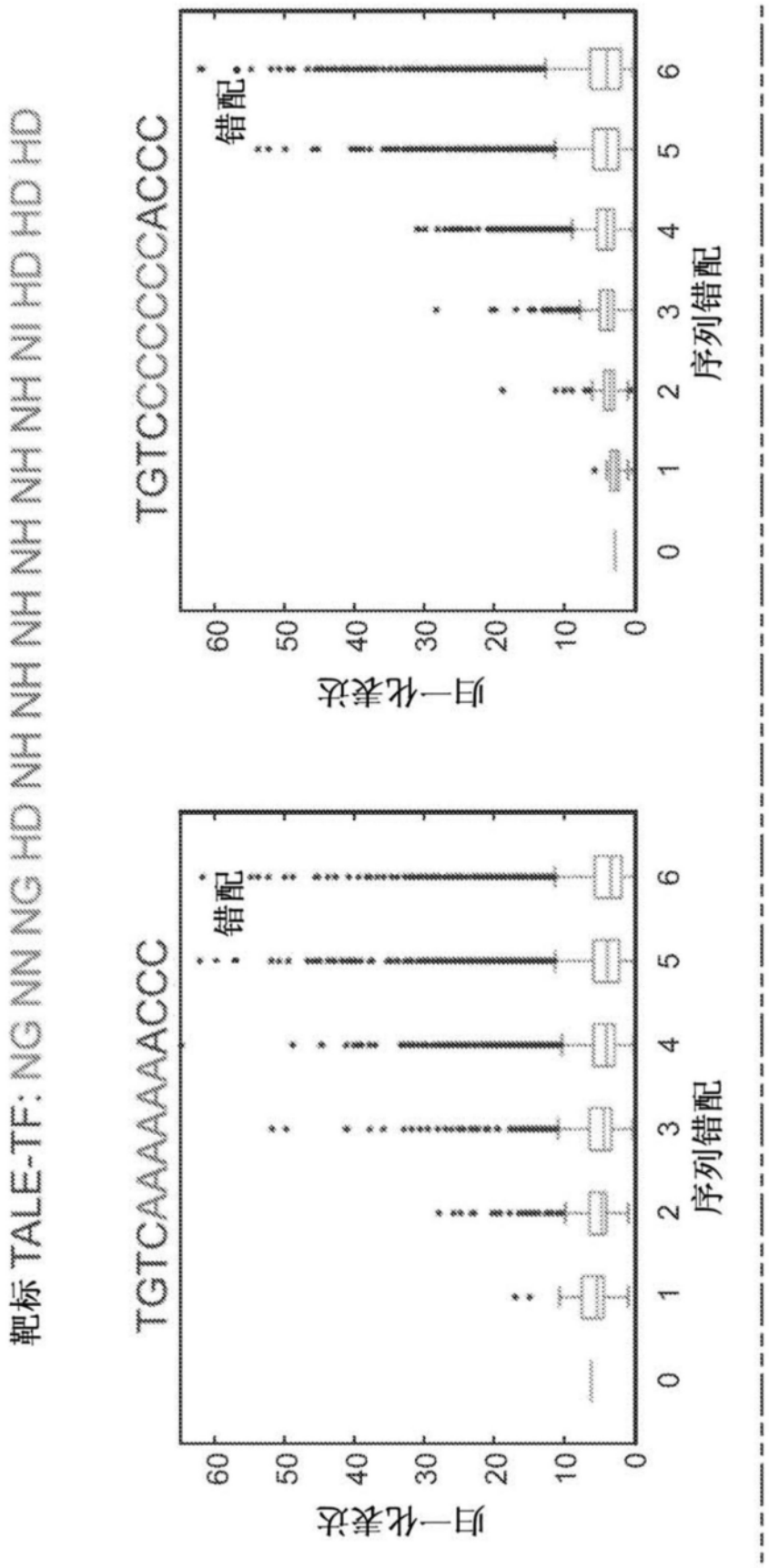

靶标 TALE-TF: NG NN NG HD NH NH NH NH NH NH NI HD HD HD
TGTCAAAAAACCC
TGTCCCCCCCACCC
错配
错配
归一化表达
序列错配
0
10
20
30
40
50
60
0
1
2
3
4
5
6

图19C-1

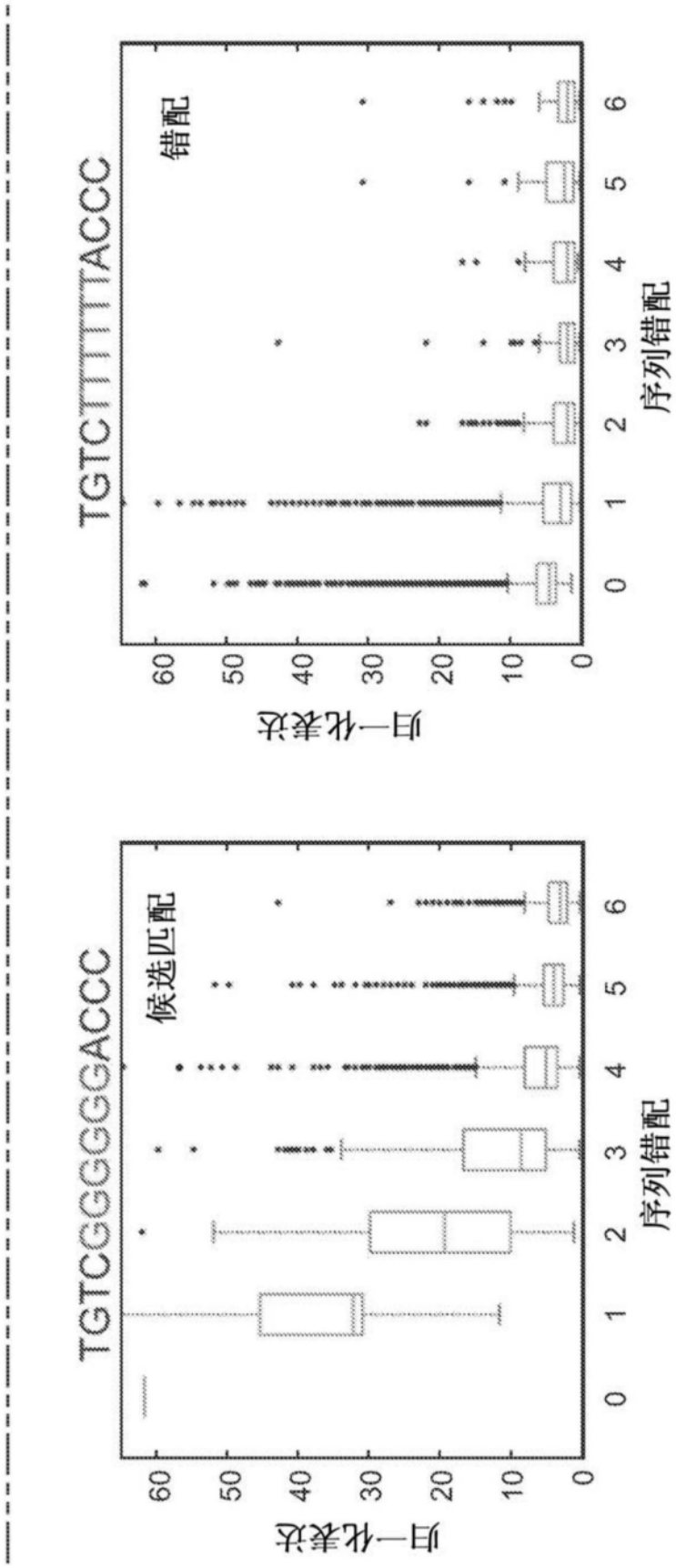
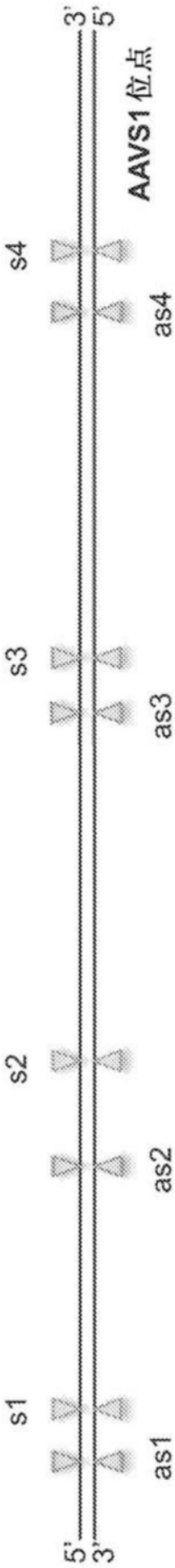

TGTCGGGGGGACCC
候选匹配
归一化表达
序列错配
TGTCTTTTTTTACCC
错配
归一化表达
序列错配

图19C-2



gRNA	序列
AAVS1_s1	GGATCCTGTGTCCCCGAGCT GGG
AAVS1_s2	GTTAATGTGGCTCTGGTTCT GGG
AAVS1_s3	GGGGCCACTAGGGACAGGAT TGG
AAVS1_s4	CTTCCTAGTCTCCTGATATT GGG
AAVS1_as1	TGGTCCCAGCTCGGGGACAC AGG
AAVS1_as2	AGAACCAGAGCCACATTAAC CGG
AAVS1_as3	GTCACCAATCCTGTCCCTAG TGG
AAVS1_as4	AGACCCAATATCAGGAGACT AGG

图20A

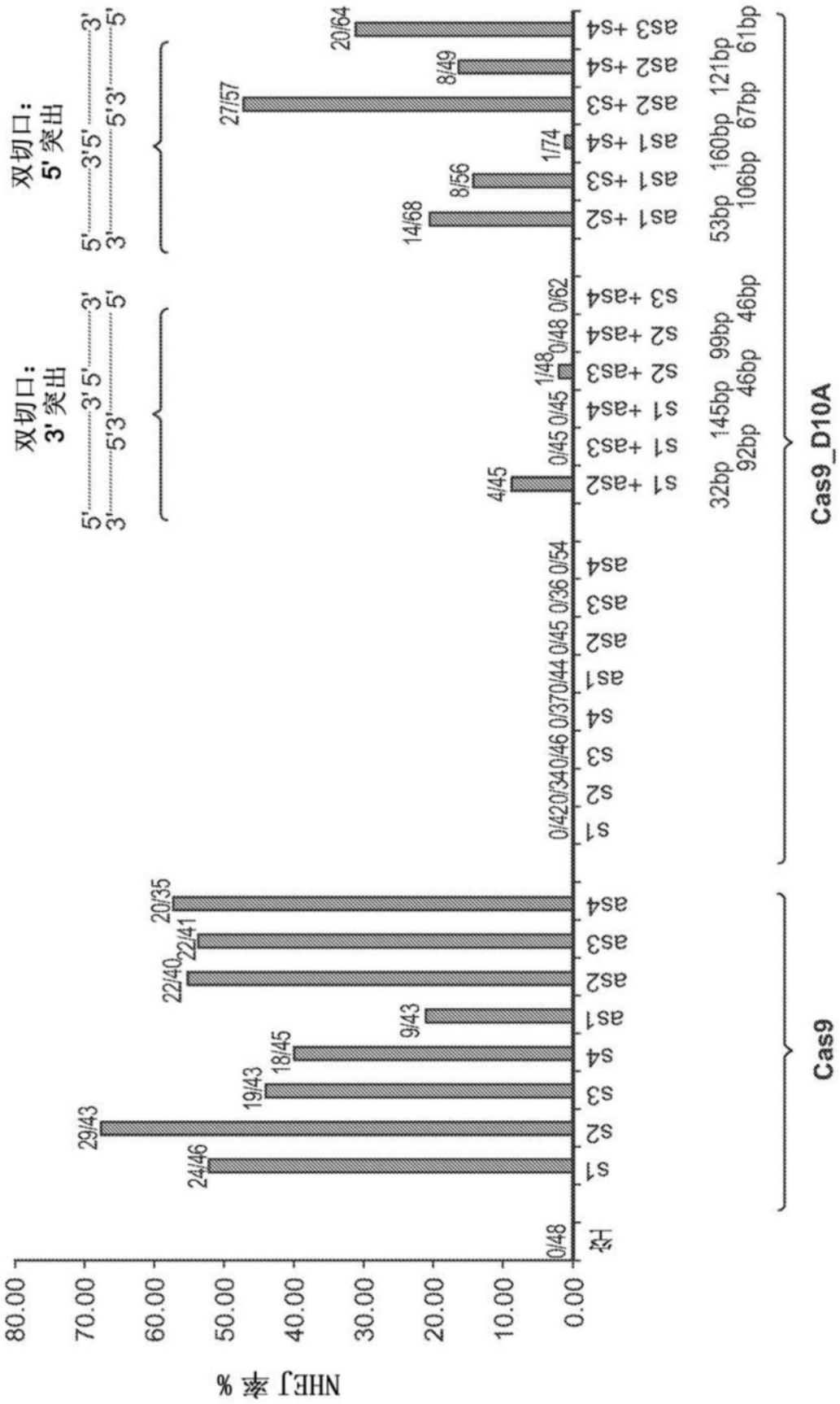

双切口：
3' 突出
5' 3' 5' 3'
3' 5' 5' 3'
双切口：
5' 突出
5' 3' 5' 3'
3' 5' 3' 5'
NHEJ 率 %
80.00
70.00
60.00
50.00
40.00
30.00
20.00
10.00
0.00
Hs
0/48
s1
24/46
s2
29/43
s3
19/43
s4
18/45
as1
9/43
as2
22/40
as3
22/41
as4
20/35
s1
0/42
s2
0/34
s3
0/46
s4
0/37
as1
0/44
as2
0/45
as3
0/36
as4
0/54
s1 +as2
4/45
s1 +as3
0/45
s1 +as4
0/45
s2 +as3
1/48
s2 +as4
0/48
s3 +as4
0/62
as1 +s2
14/68
as1 +s3
8/56
as1 +s4
1/74
as2 +s3
27/57
as2 +s4
8/49
as3 +s4
20/64
32bp 145bp 99bp
92bp 46bp 46bp
53bp 160bp 121bp
106bp 67bp 61bp
Cas9
Cas9_D10A

图20B

```
GGGATCCTGTGTCCCCGAGCTGGGACCACCTTATATTCCCAGGGC--------------------------------------CGGTTAATGTGGCTCTGGTTCTGGGTACTT
GGGATCCTGTGTCCCCGAGCTGGGACCACCTTATATTCCCAGGGC--------------------------------------CGGTTAATGTGG------TTCTGGGTACTT
GGGATCCTGTGTCCCCGAGCTGGGACCACCTTATATTCCCAGGGC--------------------------------------CGGTTAATGTGGCTCTGGTTCTGGGTACTT
GGGATCCTGTGTCCCCGAGCTGGGACCACCTTATATTCCCAGGGC--------------------------------------CGGTTAATGTGGCTCTGGTTCTGGGTACTT
GGGATCCTGTGTCCCCGAGCTGGGACCACCTTATATTCCCAGGGC--------------------------------------CGGTTAATGTGGCTCTGGTTCTGGGTACTT
GGGATCCTGTGTCCCCGAGCTGGGACCACCTTATATTCCCAGGGCAGGGCGGGTTGGACCACCTTATATTCCCAGGGCAGGGCCGGTTAATGTGGCTCTGGTTCTGGGTACTT
GGGATCCTGTGTCCCCGAGCTGGGACCACCTTATATTCCCAGGGC--------------------------------------CGGTTAATGTGGCTCTGGTTCTGGGTACTT
GGGATCCTGTGTCCCCG---T----C----T----------GG------------------------------------------TT-------CT--GG----G--TACTT
GGGATCCTGTGTCCCCGAGCTGGGACCACCTTATATTCCCAGGGT--------------------------------------CGGTTAATGTGGCTCTGGTTCTGGGTACTT
GGGATCCTGTGTCCCCGAGCTGGGACCACCTTATATTCCCAGGGC--------------------------------------CGGTTAATGTGGCTCTGGTTCTGGGTACTT
GGGATCCTGTGTCCCCGAGCTGGGACCACCTTATATTCCCAGGGC--------------------------------------CGGTTAATGTGGCTCTGGTTCTGGGTACTT
GGGATCCTGTGTCCCCGAGCTGGGACCACCTTATATTCCCAGGGC--------------------------------------CGGTTAATGTGGCTCTGGTTCTGGGTACTT
GGGATCCTGTGTCCCCGAGCTGGGACCACCTTATATTCCCAGGGC--------------------------------------CGGTTAATGTGGCTCTGGTTCTGGGTACTT
GGGATCCTGTGTCCCCGAGCTGGGACCACCTTATATTCCCAGGGC--------------------------------------CGGTTAATGTGGCTCTGGTTCTGGGTACTT
GGGATCCTGTGTCCCCGAGCTGGGACCACCTTATATTCCCAGGGC--------------------------------------CGGTTAATGTGGCTCTGGTTCTGGGTACTT
GGGATCCTGTGTCCCCGAGCTGGGACCACCTTATA------------------------------------------------------------------TTCTGGGTACTT
GGGATCCTGTGTCCCCGAGCTGGGACCACCTTATATTCCCAGGGC--------------------------------------CGGTTAATGTGGCTCTGGTTCTGGGTACTT
GGGATCCTGTGTCCCCGAGCTGGGACCACCTTATATTCCCAGGGC--------------------------------------CGGTTAATGTGGCTCTGGTTCTGGGTACTT
GGGATCCTGTGTCCCCGAGCTGGGACCACCTTATATTCCCAGGGC--------------------------------------CGGTTAATGTGGCTCTGGTTCTGGGTACTT
GGGATC-T-TGT----------------------------------------------------------------------------------------------GGTACTT
GGGATCCTGTGTCCCCGAGCTGGGACCACCTTATATTCCCAGGGC--------------------------------------CGGTTAATGTGGCTCTGGTTCTGGGTACTT
```

图21A

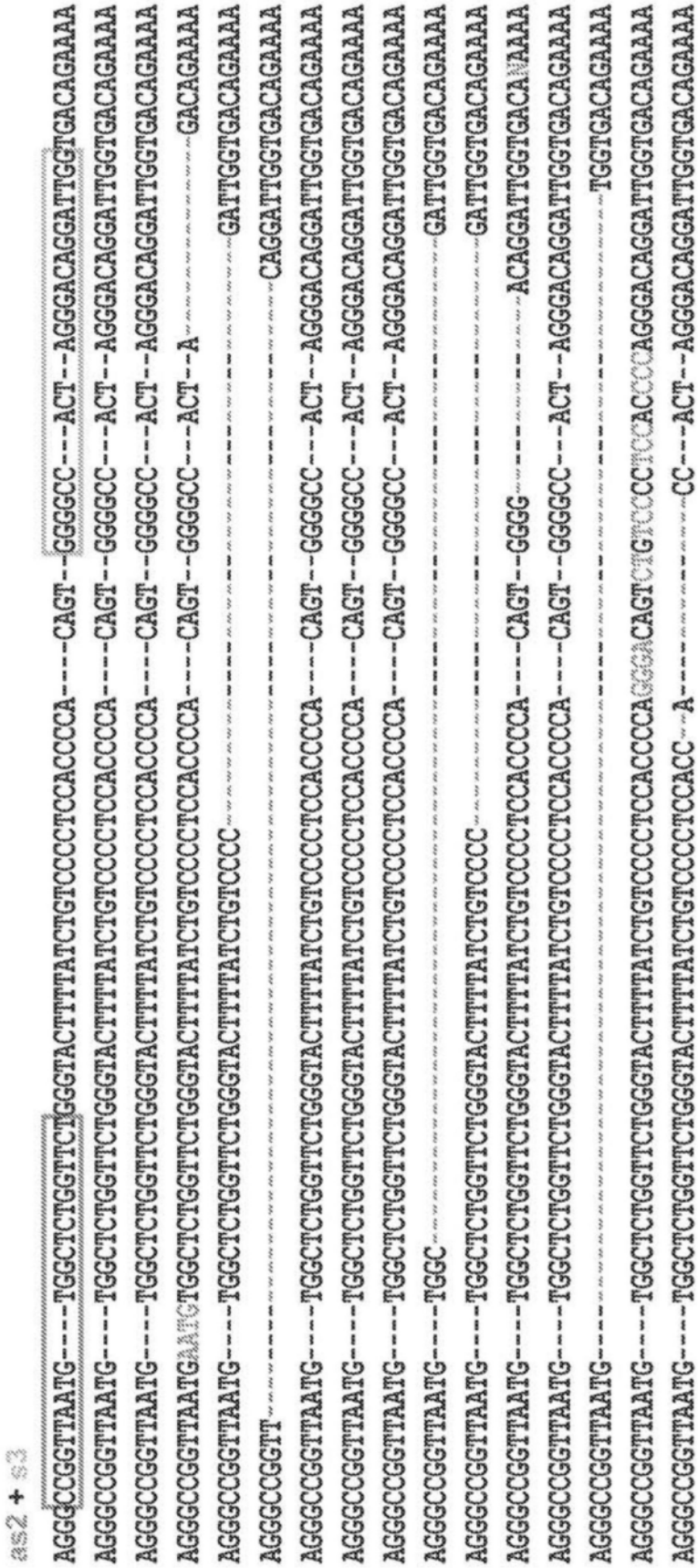
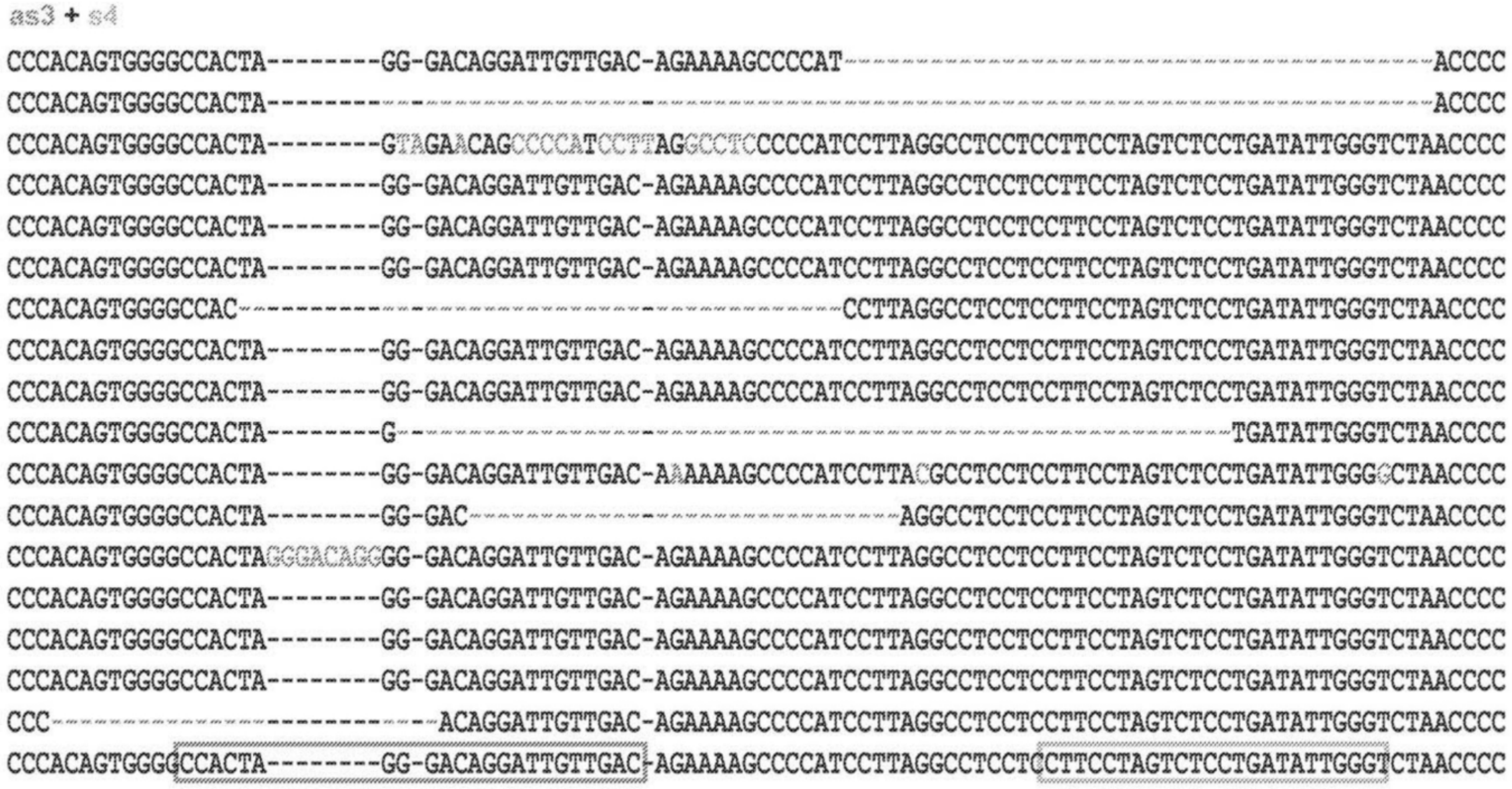

as2 + s3
AGGGCCGGTTAATG----TGGCTCTGGTTCTGGGTACTTTTATCTGTCCCCTCCACCCCA----CAGT--GGGGCC---ACT--AGGGACAGGATTGGTGACAGAAA
AGGGCCGGTTAATG----TGGCTCTGGTTCTGGGTACTTTTATCTGTCCCCTCCACCCCA----CAGT--GGGGCC---ACT--AGGGACAGGATTGGTGACAGAAA
AGGGCCGGTTAATG----TGGCTCTGGTTCTGGGTACTTTTATCTGTCCCCTCCACCCCA----CAGT--GGGGCC---ACT--AGGGACAGGATTGGTGACAGAAA
AGGGCCGGTTAATGAATGTGGCTCTGGTTCTGGGTACTTTTATCTGTCCCCTCCACCCCA----CAGT--GGGGCC---ACT--A----GACAGAAA
AGGGCCGGTTAATG----TGGCTCTGGTTCTGGGTACTTTTATCTGTCCCC----GATTGGTGACAGAAA
AGGGCCGGTT----CAGGATTGGTGACAGAAA
AGGGCCGGTTAATG----TGGCTCTGGTTCTGGGTACTTTTATCTGTCCCCTCCACCCCA----CAGT--GGGGCC---ACT--AGGGACAGGATTGGTGACAGAAA
AGGGCCGGTTAATG----TGGCTCTGGTTCTGGGTACTTTTATCTGTCCCCTCCACCCCA----CAGT--GGGGCC---ACT--AGGGACAGGATTGGTGACAGAAA
AGGGCCGGTTAATG----TGGCTCTGGTTCTGGGTACTTTTATCTGTCCCCTCCACCCCA----CAGT--GGGGCC---ACT--AGGGACAGGATTGGTGACAGAAA
AGGGCCGGTTAATG----TGGC----GATTGGTGACAGAAA
AGGGCCGGTTAATG----TGGCTCTGGTTCTGGGTACTTTTATCTGTCCCC----GATTGGTGACAGAAA
AGGGCCGGTTAATG----TGGCTCTGGTTCTGGGTACTTTTATCTGTCCCCTCCACCCCA----CAGT--GGGGCC----ACAGGATTGGTGACAGAAA
AGGGCCGGTTAATG----TGGCTCTGGTTCTGGGTACTTTTATCTGTCCCCTCCACCCCA----CAGT--GGGGCC---ACT--AGGGACAGGATTGGTGACAGAAA
AGGGCCGGTTAATG----TGGCTCTGGTTCTGGGTACTTTTATCTGTCCCCTCCACCCCA----CAGT--GGGGCC----TGGTGACAGAAA
AGGGCCGGTTAATG----TGGCTCTGGTTCTGGGTACTTTTATCTGTCCCCTCCACCCCAGGGAGTCAGTCTGTCCCTCCACCCCAGGGACAGGATTGGTGACAGAAA
AGGGCCGGTTAATG----TGGCTCTGGTTCTGGGTACTTTTATCTGTCCCCTCCACC--A----CC---ACT--AGGGACAGGATTGGTGACAGAAA

图21B



```
as3 + s4
CCCACAGTGGGGCCACTA--------GG-GACAGGATTGTTGAC-AGAAAAGCCCCAT-----------------------------------------ACCCC
CCCACAGTGGGGCCACTA---------------------------------------------------------------------------------ACCCC
CCCACAGTGGGGCCACTA--------GTAGAACAGCCCCATCCTTAGGCCTCCCCCATCCTTAGGCCTCCTCCTTCCTAGTCTCCTGATATTGGGTCTAACCCC
CCCACAGTGGGGCCACTA--------GG-GACAGGATTGTTGAC-AGAAAAGCCCCATCCTTAGGCCTCCTCCTTCCTAGTCTCCTGATATTGGGTCTAACCCC
CCCACAGTGGGGCCACTA--------GG-GACAGGATTGTTGAC-AGAAAAGCCCCATCCTTAGGCCTCCTCCTTCCTAGTCTCCTGATATTGGGTCTAACCCC
CCCACAGTGGGGCCACTA--------GG-GACAGGATTGTTGAC-AGAAAAGCCCCATCCTTAGGCCTCCTCCTTCCTAGTCTCCTGATATTGGGTCTAACCCC
CCCACAGTGGGGCCAC------------------------------------------CCTTAGGCCTCCTCCTTCCTAGTCTCCTGATATTGGGTCTAACCCC
CCCACAGTGGGGCCACTA--------GG-GACAGGATTGTTGAC-AGAAAAGCCCCATCCTTAGGCCTCCTCCTTCCTAGTCTCCTGATATTGGGTCTAACCCC
CCCACAGTGGGGCCACTA--------GG-GACAGGATTGTTGAC-AGAAAAGCCCCATCCTTAGGCCTCCTCCTTCCTAGTCTCCTGATATTGGGTCTAACCCC
CCCACAGTGGGGCCACTA--------G----------------------------------------------------------TGATATTGGGTCTAACCCC
CCCACAGTGGGGCCACTA--------GG-GACAGGATTGTTGAC-AAAAAAGCCCCATCCTTACGCCTCCTCCTTCCTAGTCTCCTGATATTGGGGCTAACCCC
CCCACAGTGGGGCCACTA--------GG-GAC------------------------------AGGCCTCCTCCTTCCTAGTCTCCTGATATTGGGTCTAACCCC
CCCACAGTGGGGCCACTAGGGACAGGGG-GACAGGATTGTTGAC-AGAAAAGCCCCATCCTTAGGCCTCCTCCTTCCTAGTCTCCTGATATTGGGTCTAACCCC
CCCACAGTGGGGCCACTA--------GG-GACAGGATTGTTGAC-AGAAAAGCCCCATCCTTAGGCCTCCTCCTTCCTAGTCTCCTGATATTGGGTCTAACCCC
CCCACAGTGGGGCCACTA--------GG-GACAGGATTGTTGAC-AGAAAAGCCCCATCCTTAGGCCTCCTCCTTCCTAGTCTCCTGATATTGGGTCTAACCCC
CCCACAGTGGGGCCACTA--------GG-GACAGGATTGTTGAC-AGAAAAGCCCCATCCTTAGGCCTCCTCCTTCCTAGTCTCCTGATATTGGGTCTAACCCC
CCC---------------------------ACAGGATTGTTGAC-AGAAAAGCCCCATCCTTAGGCCTCCTCCTTCCTAGTCTCCTGATATTGGGTCTAACCCC
CCCACAGTGGGGCCACTA--------GG-GACAGGATTGTTGAC-AGAAAAGCCCCATCCTTAGGCCTCCTCCTTCCTAGTCTCCTGATATTGGGTCTAACCCC
```

图21C