(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 111727247 A
(43)申请公布日 2020.09.29

(21)申请号 201880078370.3
(22)申请日 2018.10.04
(30)优先权数据
62/568,313 2017.10.04 US
(85)PCT国际申请进入国家阶段日
2020.06.03
(86)PCT国际申请的申请数据
PCT/US2018/054469 2018.10.04
(87)PCT国际申请的公布数据
WO2019/071048 EN 2019.04.11
(71)申请人 博德研究所
地址 美国马萨诸塞州
申请人 麻省理工学院 哈佛学院院长等
(72)发明人 F·张 J·戈滕贝格
O·阿布达耶
(74)专利代理机构 北京市金杜律师事务所 11256
代理人 陈文平 袁元
(51)Int.Cl.
*C12N 9/22*(2006.01)
*C12N 9/78*(2006.01)

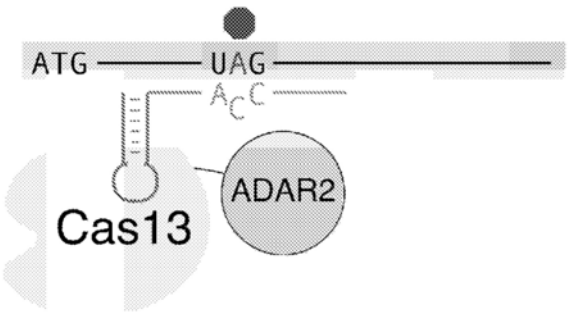
权利要求书6页 说明书364页
序列表（电子公布） 附图133页

(54)发明名称

用于靶向核酸编辑的系统、方法和组合物

(57)摘要

本发明提供了用于靶向和编辑核酸的系统、方法和组合物。具体地，本发明提供了非天然存在的或工程化的RNA靶向系统，所述系统包含RNA靶向Cas13蛋白、至少一种指导分子以及至少一种腺苷脱氨酶蛋白或其催化结构域。




CN 111727247 A

1.一种用于定点碱基编辑的工程化组合物,所述组合物包含:

a.靶向结构域;以及

b.腺苷脱氨酶或其催化结构域,其中所述腺苷脱氨酶经修饰以将活性转化为胞苷脱氨酶。

2.如权利要求1所述的组合物,其中所述腺苷脱氨酶在选自E396、C451、V351、R455、T375、K376、S486、Q488、R510、K594、R348、G593、S397、H443、L444、Y445、F442、E438、T448、A353、V355、T339、P539、V525和I520的一个或多个位置处通过一个或多个突变修饰。

3.如权利要求1所述的组合物,其中所述腺苷脱氨酶在选自E488、V351、S486、T375、S370、P462和N597的一个或多个位置处突变。

4.如权利要求1所述的组合物,其中所述腺苷脱氨酶包含选自E488Q、V351G、S486A、T375S、S370C、P462A和N597I的一个或多个突变。

5.如权利要求1所述的组合物,其中所述腺苷脱氨酶蛋白或其催化结构域是人类、头足类或果蝇腺苷脱氨酶蛋白或其催化结构域。

6.如权利要求1所述的组合物,其中所述腺苷脱氨酶蛋白或其催化结构域已经修饰以在hADAR2-D氨基酸序列的谷氨酸$^{488}$处或在同源ADAR蛋白中的对应位置处包含突变。

7.如权利要求6所述的组合物,其中位置488处的所述谷氨酸残基或同源ADAR蛋白中的对应位置被谷氨酰胺残基替换(E488Q)。

8.如权利要求6所述的组合物,其中所述腺苷脱氨酶蛋白或其催化结构域是包含突变E488Q的突变的hADAR2d或包含突变E1008Q的突变的hADAR1d。

9.如权利要求1-4中任一项所述的组合物,其中所述靶向结构域是无催化活性的Cas13蛋白,或编码所述无催化活性的Cas13蛋白的核苷酸序列。

10.如权利要求9所述的组合物,其中所述无催化活性的Cas13蛋白是无催化活性的Cas13a、无催化活性的Cas13b或无催化活性的Cas13c。

11.如权利要求9所述的组合物,其中所述无催化活性的Cas13蛋白从来源于细菌物种的Cas13核酸酶获得,所述细菌物种选自由表1、2、3或4中任一个中所列的细菌物种组成的组。

12.如权利要求1-4中任一项所述的组合物,其还包含指导分子或编码所述指导分子的核苷酸序列,所述指导分子包含与正向重复序列连接的指导序列。

13.如权利要求1-4中任一项所述的组合物,其中所述腺苷脱氨酶蛋白或其催化结构域与所述靶向结构域共价或非共价连接。

14.一种适用于修饰目标靶基因座中的腺嘌呤的工程化的非天然存在的系统,所述系统包含:

(a)指导分子,或编码所述指导分子的核苷酸序列,所述指导分子包含与正向重复序列连接的指导序列;

(b)无催化活性的Cas13蛋白,或编码所述无催化活性的Cas13蛋白的核苷酸序列;以及

(c)腺苷脱氨酶蛋白或其催化结构域,或编码所述腺苷脱氨酶蛋白或其催化结构域的核苷酸序列,其中所述腺苷脱氨酶经修饰以将活性转化为胞苷脱氨酶;

其中所述腺苷脱氨酶蛋白或其催化结构域与所述Cas13蛋白或所述指导分子共价或非共价连接,或者适用于在递送后与其连接;

其中所述指导序列能够与包含腺嘌呤的靶RNA序列杂交以形成RNA双链体，其中所述指导序列在对应于所述腺嘌呤的位置处包含非配对胞嘧啶，从而导致在形成的所述RNA双链体中产生A-C错配。

15.如权利要求14所述的系统，其中所述腺苷脱氨酶通过选自E396、C451、V351、R455、T375、K376、S486、Q488、R510、K594、R348、G593、S397、H443、L444、Y445、F442、E438、T448、A353、V355、T339、P539、V525和I520的一个或多个突变修饰。

16.如权利要求14所述的方法，其中所述腺苷脱氨酶在选自E488、V351、S486、T375、S370、P462和N597的一个或多个位置处突变。

17.一种修饰目标靶RNA序列中的腺嘌呤的方法，所述方法包括向所述靶RNA递送：

(a)无催化活性的(死亡)Cas13蛋白；

(b)包含与正向重复序列连接的指导序列的指导分子；以及

(c)腺苷脱氨酶蛋白或其催化结构域；

其中所述腺苷脱氨酶蛋白或其催化结构域与所述死亡Cas13蛋白或所述指导分子共价或非共价连接，或者适用于在递送后与其连接；

其中所述指导分子与所述死亡Cas13蛋白形成复合物并指导所述复合物结合所述目标靶RNA序列，其中所述指导序列能够与包含所述腺嘌呤的靶序列杂交以形成RNA双链体，其中所述指导序列在对应于所述腺嘌呤的位置处包含非配对胞嘧啶，从而导致在形成的所述RNA双链体中产生A-C错配；

其中所述腺苷脱氨酶蛋白或其催化结构域使所述RNA双链体中的所述腺嘌呤脱氨。

18.如权利要求17所述的方法，其中所述Cas13蛋白是Cas13a、Cas13b或Cas13c。

19.如权利要求17所述的方法，其中所述腺苷脱氨酶蛋白或其催化结构域与所述死亡Cas13蛋白的N末端或C末端融合。

20.如权利要求19的方法，其中所述腺苷脱氨酶蛋白或其催化结构域通过接头与所述死亡Cas13蛋白融合。

21.如权利要求20所述的方法，其中所述接头是$(GGGGS)_{3-11}$、$GSG_5$或LEPGEKPYKCPECGKSFSQSGALTRHQRTHTR，或者其中所述接头是XTEN接头。

22.如权利要求17所述的方法，其中所述腺苷脱氨酶蛋白或其催化结构域与衔接蛋白连接，并且所述指导分子或所述死亡Cas13蛋白包含能够与所述衔接蛋白结合的适配体序列。

23.如权利要求22所述的方法，其中所述衔接子序列选自MS2、PP7、Qβ、F2、GA、fr、JP501、M12、R17、BZ13、JP34、JP500、KU1、M11、MX1、TW18、VK、SP、FI、ID2、NL95、TW19、AP205、ΦCb5、ΦCb8r、ΦCb12r、ΦCb23r、7s和PRR1。

24.如权利要求23所述的方法，其中所述腺苷脱氨酶蛋白或其催化结构域被插入所述死亡Cas13蛋白的内部环中。

25.如权利要求24所述的方法，其中所述Cas13蛋白是Cas13a蛋白，并且所述Cas13a在两个HEPN结构域中、特别是在源自瓦氏纤毛菌的Cas13a蛋白的位置R474和R1046或者Cas13a直系同源物的与其对应的氨基酸位置处包含一个或多个突变。

26.如权利要求25所述的方法，其中所述Cas13蛋白是Cas13b蛋白，并且所述Cas13b在源自动物溃疡伯杰菌ATCC 43767的Cas13b蛋白的位置R116、H121、R1177或H1182中的一个

或多个或者Cas13b直系同源物的与其对应的氨基酸位置中包含突变。

27.如权利要求26所述的方法，其中所述突变是源自动物溃疡伯杰菌ATCC 43767的Cas13b蛋白的R116A、H121A、R1177A或H1182A中的一个或多个或者Cas13b直系同源物的与其对应的氨基酸位置。

28.如权利要求17-27中任一项所述的方法，其中所述指导序列的长度为约20至约53nt、约25至约53nt或约29至约53nt，能够与所述靶序列形成所述RNA双链体。

29.如权利要求28所述的方法，其中所述指导序列的长度为约40至约50nt，能够与所述靶序列形成所述RNA双链体。

30.如权利要求17所述的方法，其中所述非配对C与所述指导序列的5'端之间的距离为20-30个核苷酸。

31.如权利要求17-30中任一项所述的方法，其中所述腺苷脱氨酶蛋白或其催化结构域是人类、头足类或果蝇腺苷脱氨酶蛋白或其催化结构域。

32.如权利要求17的方法，其中所述腺苷脱氨酶蛋白或其催化结构域已经修饰以在hADAR2-D氨基酸序列的谷氨酸$^{488}$处或在同源ADAR蛋白中的对应位置处包含突变。

33.如权利要求32所述的方法，其中位置488处的所述谷氨酸残基或同源ADAR蛋白中的对应位置被谷氨酰胺残基替换(E488Q)。

34.如权利要求32或33所述的方法，其中所述腺苷脱氨酶蛋白或其催化结构域是包含突变E488Q的突变的hADAR2d或包含突变E1008Q的突变的hADAR1d。

35.如权利要求17-34中任一项所述的方法，其中所述指导序列包含一个以上对应于所述靶RNA序列中不同腺苷位点的错配，或者其中使用两个指导分子，每个指导分子包含一个对应于所述靶RNA序列中不同腺苷位点的错配。

36.如权利要求17-35所述的方法，其中所述Cas13蛋白以及任选地所述腺苷脱氨酶蛋白或其催化结构域包含一个或多个异源核定位信号(NLS)。

37.如权利要求17-36中任一项所述的方法，其中所述方法包括确定所述目标靶序列并选择最有效地使存在于所述靶序列中的所述腺嘌呤脱氨的所述脱氨腺苷脱氨酶蛋白或其催化结构域。

38.如权利要求17-37中任一项所述的方法，其中所述无催化活性的Cas13蛋白从来源于细菌物种的Cas13核酸酶获得，所述细菌物种选自由表1、2、3或4中任一个中所列的细菌物种组成的组。

39.如权利要求17-38中任一项所述的方法，其中所述目标靶RNA序列在细胞内。

40.如权利要求39所述的方法，其中所述细胞是真核细胞。

41.如权利要求40所述的方法，其中所述细胞是非人类动物细胞。

42.如权利要求41所述的方法，其中所述细胞是人类细胞。

43.如权利要求42所述的方法，其中所述细胞是植物细胞。

44.如权利要求17-43中任一项所述的方法，其中所述目标靶基因座在动物内。

45.如权利要求17-44中任一项所述的方法，其中所述目标靶基因座在植物内。

46.如权利要求17-45中任一项所述的方法，其中所述目标靶RNA序列包含在体外RNA多核苷酸中。

47.如权利要求17-46中任一项所述的方法，其中将所述组分(a)、(b)和(c)作为核糖核

蛋白复合物递送到所述细胞。

48.如权利要求17-47中任一项所述的方法，其中将所述组分(a)、(b)和(c)作为一种或多种多核苷酸分子递送到所述细胞。

49.如权利要求48所述的方法，其中所述一种或多种多核苷酸分子包含编码组分(a)和/或(c)的一种或多种mRNA分子。

50.如权利要求49所述的方法，其中所述一种或多种多核苷酸分子包含在一种或多种载体内。

51.如权利要求50所述的方法，其中所述一种或多种多核苷酸分子包含可操作地被构造成表达所述Cas13蛋白、所述指导分子以及所述腺苷脱氨酶蛋白或其催化结构域的一个或多个调控元件，任选地其中所述一个或多个调控元件包含诱导型启动子。

52.如权利要求48所述的方法，其中所述一种或多种多核苷酸分子或者所述核糖核蛋白复合物经由颗粒、囊泡或者一种或多种病毒载体递送。

53.如权利要求52所述的方法，其中所述颗粒包括脂质、糖、金属或蛋白质。

54.如权利要求53所述的方法，其中所述颗粒包括脂质纳米颗粒。

55.如权利要求52所述的方法，其中所述囊泡包括外来体或脂质体。

56.如权利要求50所述的方法，其中所述一种或多种病毒载体包括一种或多种腺病毒、一种或多种慢病毒或者一种或多种腺相关病毒。

57.如权利要求17-56中任一项所述的方法，其中所述方法通过操纵一条或多条靶RNA序列来修饰细胞、细胞系或生物体。

58.如权利要求57所述的方法，其中所述目标靶RNA中的所述腺嘌呤的所述脱氨治疗了由含有致病性G→A或C→T点突变的转录物引起的疾病。

59.如权利要求58所述的方法，其中所述疾病选自梅尔-戈林综合征、塞克尔综合征4、朱伯特综合征5、Leber先天性黑蒙症10;2型夏科-马利-杜斯氏病;2型夏科-马利-杜斯氏病;2C型乌舍尔综合征;脊髓小脑失调症28;脊髓小脑失调症28;脊髓小脑失调症28;长QT综合征2;舍格伦-拉尔森综合征;遗传性果糖尿症;遗传性果糖尿症;神经母细胞瘤;神经母细胞瘤;卡尔曼综合征1;卡尔曼综合征1;卡尔曼综合征1;异染性脑白质营养不良、瑞特综合征、10型肌萎缩性脊髓侧索硬化症、利弗劳梅尼综合征或表5中所列的疾病。

60.如权利要求58所述的方法，其中所述疾病是提前终止疾病。

61.如权利要求57所述的方法，其中所述修饰影响生物体的生育力。

62.如权利要求57所述的方法，其中所述修饰影响所述靶RNA序列的剪接。

63.如权利要求57所述的方法，其中所述修饰在转录物中引入突变，从而引入氨基酸变化并引起癌细胞中新抗原的表达。

64.如权利要求57所述的方法，其中所述靶RNA包含在微RNA内。

65.如权利要求57所述的方法，其中所述目标靶RNA中的所述腺嘌呤的所述脱氨引起基因的功能获得或功能丧失。

66.如权利要求65所述的方法，其中所述基因是癌细胞表达的基因。

67.一种由如前述权利要求中任一项所述的方法获得的经修饰的细胞，或所述经修饰的细胞的后代，其中与未经历所述方法的对应细胞相比，所述细胞包含次黄嘌呤或鸟嘌呤代替所述目标靶RNA中的所述腺嘌呤。

68.如权利要求67所述的经修饰的细胞或其后代,其中所述细胞是真核细胞。

69.如权利要求67所述的经修饰的细胞或其后代,其中所述细胞是动物细胞。

70.如权利要求67所述的经修饰的细胞或其后代,其中所述细胞是人类细胞。

71.如权利要求67所述的经修饰的细胞或其后代,其中所述细胞是治疗性T细胞。

72.如权利要求67所述的经修饰的细胞或其后代,其中所述细胞是产生抗体的B细胞。

73.如权利要求67所述的经修饰的细胞或其后代,其中所述细胞是植物细胞。

74.一种非人类动物,所述非人类动物包含如权利要求67所述的所述经修饰的细胞。

75.一种植物,所述植物包含如权利要求74所述的所述经修饰的细胞。

76.一种用于细胞疗法的方法,所述方法包括向有需要的患者施用如权利要求67-75中任一项所述的所述经修饰的细胞,其中所述经修饰的细胞的存在治疗了所述患者中的疾病。

77.一种适用于修饰目标靶基因座中的腺嘌呤的工程化的非天然存在的系统,所述系统包含:

a)指导分子,或编码所述指导分子的核苷酸序列,所述指导分子包含与正向重复序列连接的指导序列;

b)无催化活性的Cas13蛋白,或编码所述无催化活性的Cas13蛋白的核苷酸序列;

c)腺苷脱氨酶蛋白或其催化结构域,或编码所述腺苷脱氨酶蛋白或其催化结构域的核苷酸序列;

其中所述腺苷脱氨酶蛋白或其催化结构域与所述Cas13蛋白或所述指导分子共价或非共价连接,或者适用于在递送后与其连接;

其中所述指导序列能够与包含腺嘌呤的靶RNA序列杂交以形成RNA双链体,其中所述指导序列在对应于所述腺嘌呤的位置处包含非配对胞嘧啶,从而导致在形成的所述RNA双链体中产生A-C错配。

78.一种适用于修饰目标靶基因座中的腺嘌呤的工程化的非天然存在的载体系统,所述载体系统包含如权利要求77所述的a)、b)和c)的所述核苷酸序列。

79.如权利要求78所述的工程化的非天然存在的载体系统,所述载体系统包含一种或多种载体,所述一种或多种载体包含:

a)第一调控元件,所述第一调控元件可操作地连接到编码包含所述指导序列的所述指导分子的核苷酸序列;

b)第二调控元件,所述第二调控元件可操作地连接到编码所述无催化活性的Cas13蛋白的核苷酸序列;以及

c)编码腺苷脱氨酶蛋白或其催化结构域的核苷酸序列,所述核苷酸序列受所述第一调控元件或所述第二调控元件的控制或可操作地连接到第三调控元件;

其中如果所述编码腺苷脱氨酶蛋白或其催化结构域的核苷酸序列可操作地连接到第三调控元件,则所述腺苷脱氨酶蛋白或其催化结构域适用于在表达后连接到所述指导分子或所述Cas13蛋白;

其中组分(a)、(b)和(c)位于所述系统的相同或不同载体上。

80.一种体外或离体宿主细胞或其后代或者细胞系或其后代,其包含如权利要求77-79中任一项所述的系统。

81.如权利要求80所述的宿主细胞或其后代或者细胞系或其后代,其中所述细胞是真核细胞。

82.如权利要求80所述的宿主细胞或其后代或者细胞系或其后代,其中所述细胞是动物细胞。

83.如权利要求80所述的宿主细胞或其后代或者细胞系或其后代,其中所述细胞是人类细胞。

84.如权利要求80所述的宿主细胞或其后代或者细胞系或其后代,其中所述细胞是植物细胞。

85.如权利要求117所述的方法,其中所述胞嘧啶并非5'侧接鸟苷。

86.如权利要求17所述的方法,其中所述腺苷脱氨酶是ADAR,任选地是huADAR,任选地是(hu)ADAR1或(hu)ADAR2。

87.如权利要求17所述的方法,其中所述Cas13、优选Cas13b被截短,优选地在C末端被截短,优选地其中所述Cas13是对应野生型Cas13的截短的功能变体。

88.如权利要求17所述的方法,其中所述腺苷脱氨酶是RNA特异性腺苷脱氨酶。

89.根据权利要求17或32所述的方法,其中所述腺苷脱氨酶蛋白或其催化结构域已经修饰以包含ADAR的一个或多个突变、优选如本文所述的突变,例如如图43A-43D、44、45、46A-46B、47A-47B中任一个中所提供的突变或者ADAR同系物或直系同源物中的对应突变。

90.一种修饰目标靶RNA序列中的腺嘌呤的方法,所述方法包括向所述靶RNA递送:

(a)无催化活性的(死亡)Cas13蛋白;

(b)包含与正向重复序列连接的指导序列的指导分子;以及

(c)突变以将活性转化为胞苷脱氨酶的腺苷脱氨酶蛋白或其催化结构域;

其中所述腺苷脱氨酶蛋白或其催化结构域与所述死亡Cas13蛋白或所述指导分子共价或非共价连接,或者适用于在递送后与其连接;

其中所述指导分子与所述死亡Cas13蛋白形成复合物并指导所述复合物结合所述目标靶RNA序列,其中所述指导序列能够与包含所述腺嘌呤的靶序列杂交以形成RNA双链体,其中所述指导序列在对应于所述腺嘌呤的位置处包含非配对胞嘧啶,从而导致在形成的所述RNA双链体中产生A-C错配;

其中所述腺苷脱氨酶蛋白或其催化结构域使所述RNA双链体中的所述腺嘌呤脱氨。

91.如权利要求90所述的方法,其中所述腺苷脱氨酶在选自E396、C451、V351、R455、T375、K376、S486、Q488、R510、K594、R348、G593、S397、H443、L444、Y445、F442、E438、T448、A353、V355、T339、P539、V525和I520的一个或多个位置处突变。

92.如权利要求91所述的方法,其中所述腺苷脱氨酶在选自E488、V351、S486、T375、S370、P462和N597的一个或多个位置处突变。

# 用于靶向核酸编辑的系统、方法和组合物

[0001] 相关申请的交叉引用

[0002] 本申请要求2017年10月4日提交的美国临时申请号62/568,313的权益。上述申请的全部内容特此通过引用完全并入本文。

[0003] 关于联邦资助研究的声明

[0004] 本发明是根据由美国国立卫生研究院(National Institutes of Health)授予的授权号MH100706和MH110049在美国政府支持下完成的。美国政府享有本发明的某些权利。

[0005] 电子序列表的引用

[0006] 电子序列表(“BROD-2345WP_ST25_FOR_FILING.txt”;大小为1,427,232字节,并创建于2018年9月25日)的内容通过引用全部并入本文。

## 技术领域

[0007] 本发明整体涉及用于靶向和编辑核酸、特别是用于在目标靶基因座处对腺嘌呤进行可编程脱氨的系统、方法和组合物。

## 背景技术

[0008] 基因组测序技术和分析方法的最新进展显著加速了对与各种生物学功能和疾病相关联的遗传因子进行编目和标测的能力。精确的基因组靶向技术是通过对单个遗传元件进行选择性干扰来实现因果遗传变异的系统性逆向工程化以及推进合成生物学、生物技术和医学应用所需要的。尽管基因组编辑技术(诸如设计者锌指、转录激活子样效应物(TALE)或归巢大范围核酸酶)可用于产生靶向基因组干扰,但仍需要采用新颖的策略和分子机制并且负担得起、易于建立、可放大并且便于靶向真核基因组内的多个位置的新基因组工程化技术。这将为基因组工程化和生物技术中的新应用提供主要资源。

[0009] 已经报道了对胞嘧啶进行可编程脱氨,并且其可用于校正A→G和T→C点突变。例如,Komor等人,Nature(2016)533:420-424报道了通过Cas9引导的RNA复合物与靶向DNA链结合而置换的非靶向DNA链中的APOBEC1胞苷脱氨酶对胞嘧啶进行靶向脱氨,这导致胞嘧啶转化为尿嘧啶。还可参见Kim等人,Nature Biotechnology(2017)35:371-376;Shimatani等人,Nature Biotechnology(2017)doi:10.1038/nbt.3833;Zong等人,Nature Biotechnology(2017)doi:10.1038/nbt.3811;Yang Nature Communication(2016)doi:10.1038/ncomms13330。

## 发明内容

[0010] 本申请涉及修饰目标靶RNA序列。使用RNA靶向而不是DNA靶向可提供与治疗开发相关的几个优点。首先,靶向RNA有相当大的安全益处:脱靶事件将更少,因为转录组中的可用序列空间显著小于基因组,并且如果脱靶事件确实发生,则将很短暂并且不太可能诱导负面副作用。其次,RNA靶向疗法将更有效,因为它们是细胞类型非依赖性的并且不必进入细胞核,使得它们更容易递送。

[0011] 在一个方面，本公开包括一种用于定点碱基编辑的工程化组合物，所述组合物包含：靶向结构域；以及腺苷脱氨酶或其催化结构域，其中腺苷脱氨酶经修饰以将活性转化为胞苷脱氨酶。

[0012] 在一些实施方案中，腺苷脱氨酶在选自E396、C451、V351、R455、T375、K376、S486、Q488、R510、K594、R348、G593、S397、H443、L444、Y445、F442、E438、T448、A353、V355、T339、P539、V525和I520的一个或多个位置处通过一个或多个突变修饰。在一些实施方案中，腺苷脱氨酶在选自E488、V351、S486、T375、S370、P462和N597的一个或多个位置处突变。在一些实施方案中，腺苷脱氨酶包含选自E488Q、V351G、S486A、T375S、S370C、P462A和N597I的一个或多个突变。在一些实施方案中，腺苷脱氨酶蛋白或其催化结构域是人类、头足类或果蝇腺苷脱氨酶蛋白或其催化结构域。在一些实施方案中，所述腺苷脱氨酶蛋白或其催化结构域已经修饰以在hADAR2-D氨基酸序列的谷氨酸488处或在同源ADAR蛋白中的对应位置处包含突变。在一些实施方案中，488处的所述谷氨酸残基或同源ADAR蛋白中的对应位置被谷氨酰胺残基替换(E488Q)。在一些实施方案中，所述腺苷脱氨酶蛋白或其催化结构域是包含突变E488Q的突变的hADAR2d或包含突变E1008Q的突变的hADAR1d。在一些实施方案中，靶向结构域是无催化活性的Cas13蛋白，或编码所述无催化活性的Cas13蛋白的核苷酸序列。在一些实施方案中，无催化活性的Cas13蛋白是无催化活性的Cas13a、无催化活性的Cas13b或无催化活性的Cas13c。在一些实施方案中，所述无催化活性的Cas13蛋白从来源于细菌物种的Cas13核酸酶获得，所述细菌物种选自由表1、2、3或4中任一个中所列的细菌物种组成的组。在一些实施方案中，所述组合物还包含指导分子或编码所述指导分子的核苷酸序列，所述指导分子包含与正向重复序列连接的指导序列。在一些实施方案中，所述腺苷脱氨酶蛋白或其催化结构域与靶向结构域共价或非共价连接。

[0013] 在另一方面，本发明涉及一种修饰目标靶RNA序列中的腺嘌呤的方法。在特定的实施方案中，所述方法包括向所述靶RNA递送：(a)无催化活性的(死亡)Cas13蛋白；(b)包含与正向重复序列连接的指导序列的指导分子；以及(c)腺苷脱氨酶蛋白或其催化结构域；其中所述腺苷脱氨酶蛋白或其催化结构域与所述死亡Cas13蛋白或所述指导分子共价或非共价连接，或者适用于在递送后与其连接；其中指导分子与所述死亡Cas13蛋白形成复合物并指导所述复合物结合所述目标靶RNA序列，其中所述指导序列能够与包含所述腺嘌呤的靶序列杂交以形成RNA双链体，其中所述指导序列在对应于所述腺嘌呤的位置处包含非配对胞嘧啶，从而导致在形成的RNA双链体中产生A-C错配；其中所述腺苷脱氨酶蛋白或其催化结构域使所述RNA双链体中的所述腺嘌呤脱氨。

[0014] 在某些示例实施方案中，Cas13蛋白是Cas13a、Cas13b或Cas13c。

[0015] 腺苷脱氨酶蛋白或其催化结构域与所述死亡Cas13蛋白的N末端或C末端融合。在某些示例实施方案中，腺苷脱氨酶蛋白或其催化结构域通过接头与所述死亡Cas13蛋白融合。接头可以是$(GGGGS)_{3-11}$(SEQ ID No.1-9)、$GSG_5$(SEQ ID No.10)或LEPGEKPYKCPECGKSFSQSGALTRHQRTHTR(SEQ ID No.11)。

[0016] 在某些示例实施方案中，腺苷脱氨酶蛋白或其催化结构域与衔接蛋白连接，并且所述指导分子或所述死亡Cas13蛋白包含能够与所述衔接蛋白结合的适配体序列。衔接子序列可选自MS2、PP7、Qβ、F2、GA、fr、JP501、M12、R17、BZ13、JP34、JP500、KU1、M11、MX1、TW18、VK、SP、FI、ID2、NL95、TW19、AP205、ΦCb5、ΦCb8r、ΦCb12r、ΦCb23r、7s和PRR1。

[0017] 在某些示例实施方案中，腺苷脱氨酶蛋白或其催化结构域被插入所述死亡Cas13蛋白的内部环中。在某些示例实施方案中，Cas13a蛋白在两个HEPN结构域中、特别是在源自瓦氏纤毛菌(Leptotrichia wadei)的Cas 13a蛋白的位置R474和R1046或者Cas13a直系同源物的与其对应的氨基酸位置处包含一个或多个突变。

[0018] 在某些示例实施方案中，Cas 13蛋白是Cas13b蛋白，并且Cas13b在源自动物溃疡伯杰菌(Bergeyella zoohelcum)ATCC 43767的Cas13b蛋白的位置R116、H121、R1177、H1182中的一个或多个或者Cas13b直系同源物的与其对应的氨基酸位置中包含突变。在某些其他示例实施方案中，突变是源自动物溃疡伯杰菌ATCC 43767的Cas13b蛋白的R116A、H121A、R1177A、H1182A中的一个或多个或者Cas13b直系同源物的与其对应的氨基酸位置。

[0019] 在某些示例实施方案中，指导序列具有能够与所述靶序列形成所述RNA双链体的约29-53nt的长度。在某些其他示例实施方案中，指导序列具有能够与所述靶序列形成所述RNA双链体的约40-50nt的长度。在某些示例实施方案中，所述非配对C与所述指导序列的5’端之间的距离为20-30个核苷酸。

[0020] 在某些示例实施方案中，腺苷脱氨酶蛋白或其催化结构域是人类、头足类或果蝇腺苷脱氨酶蛋白或其催化结构域。在某些示例实施方案中，腺苷脱氨酶蛋白或其催化结构域已经修饰以在hADAR2-D氨基酸序列的谷氨酸$^{488}$处或在同源ADAR蛋白中的对应位置处包含突变。在某些示例实施方案中，谷氨酸残基可在位置488处，或者同源ADAR蛋白中的对应位置被谷氨酰胺残基替换(E488Q)。

[0021] 在某些其他示例实施方案中，腺苷脱氨酶蛋白或其催化结构域是包含突变E488Q的突变的hADAR2d或包含突变E1008Q的突变的hADAR1d。

[0022] 在某些示例实施方案中，指导序列包含一个以上对应于靶RNA序列中不同腺苷位点的错配，或者其中使用两个指导分子，每个指导分子包含一个对应于靶RNA序列中不同腺苷位点的错配。

[0023] 在某些示例实施方案中，Cas13蛋白以及任选地所述腺苷脱氨酶蛋白或其催化结构域包含一个或多个异源核定位信号(NLS)。

[0024] 在某些示例实施方案中，所述方法还包括确定目标靶序列并选择最有效地使存在于靶序列中的所述腺嘌呤脱氨的脱氨腺苷脱氨酶蛋白或其催化结构域。

[0025] 目标靶RNA序列可在细胞内。细胞可以是真核细胞、非人类动物细胞、人类细胞、植物细胞。目标靶基因座可在动物或植物内。

[0026] 目标靶RNA序列可在体外RNA多核苷酸中包含。

[0027] 本文所述系统的组分可作为核糖核蛋白复合物或者作为一种或多种多核苷酸分子递送到所述细胞。一种或多种多核苷酸分子可包含编码这些组分的一种或多种mRNA分子。一种或多种多核苷酸分子可包含在一个或多个载体内。一种或多种多核苷酸分子还可包含可操作地被构造成表达所述Cas13蛋白、所述指导分子以及所述脱氨酶蛋白或其催化结构域的一个或多个调控元件，任选地其中所述一个或多个调控元件包含诱导型启动子。可经由颗粒、囊泡或者一个或多个病毒载体来递送一种或多种多核苷酸分子或所述核糖核蛋白复合物。颗粒可包含脂质、糖、金属或蛋白质。颗粒可包含脂质纳米颗粒。囊泡可包含外来体或脂质体。一个或多个病毒载体可包含一种或多种腺病毒、一种或多种慢病毒或者一种或多种腺相关病毒。

[0028] 本文公开的方法可用于通过操纵一条或多条靶RNA序列来修饰细胞、细胞系或生物。

[0029] 在某些示例实施方案中，使所述目标靶RNA中的所述腺嘌呤脱氨可以治疗由含有致病性G→A或C→T点突变的转录物引起的疾病。

[0030] 所述方法可用于治疗疾病。在某些示例实施方案中，疾病选自梅尔-戈林综合征、塞克尔综合征4、朱伯特综合征5、Leber先天性黑蒙症10；2型夏科-马利-杜斯氏病；2型夏科-马利-杜斯氏病；2C型乌舍尔综合征；脊髓小脑失调症28；脊髓小脑失调症28；脊髓小脑失调症28；长QT综合征2；舍格伦-拉尔森综合征；遗传性果糖尿症；遗传性果糖尿症；神经母细胞瘤；神经母细胞瘤；卡尔曼综合征1；卡尔曼综合征1；卡尔曼综合征1；异染性脑白质营养不良、瑞特综合征、10型肌萎缩性脊髓侧索硬化症、利弗劳梅尼综合征或表5中所列的疾病。疾病可以是提前终止疾病。

[0031] 本文公开的方法可用于进行影响生物繁殖力的修饰。修饰可影响所述靶RNA序列的剪接。修饰可在转录物中引入突变，该突变引入氨基酸变化并引起癌细胞中新抗原的表达。

[0032] 在某些示例实施方案中，靶RNA可以是微RNA或包含在微RNA中。在某些示例实施方案中，使所述目标靶RNA中的所述腺嘌呤脱氨引起基因的功能获得或功能丧失。在某些示例实施方案中，基因是癌细胞表达的基因。

[0033] 在某些示例实施方案中，所述胞嘧啶并非5'侧接鸟苷。在某些示例实施方式中，所述腺苷脱氨酶是ADAR，任选地是huADAR，任选地是(hu)ADAR1或(hu)ADAR2。在某些示例实施方案中，所述Cas13、优选Cas13b被截短，优选地在C末端被截短，优选地其中所述Cas 13是对应野生型Cas13的截短的功能变体。在某些示例实施方案中，所述腺苷脱氨酶是RNA特异性腺苷脱氨酶。在某些示例实施方案中，所述腺苷脱氨酶蛋白或其催化结构域已经修饰以包含ADAR的一个或多个突变、优选如本文所述的突变，例如如图43A-43D、44、45、46A-46B、47A-47B中任一个中所提供的突变或者ADAR同系物或同源物中的对应突变。

[0034] 在另一方面，本发明包括使用本文公开的方法获得的经修饰的细胞或其后代，其中与未经历所述方法的对应细胞相比，所述细胞包含次黄嘌呤或鸟嘌呤代替所述目标靶RNA中的所述腺嘌呤。经修饰的细胞或其后代可以是真核细胞、动物细胞、人类细胞、治疗T细胞、产生抗体的B细胞、植物细胞。

[0035] 在另一方面，本发明包括包含所述经修饰的细胞或其后代的非人类动物。经修饰的细胞可以是植物细胞。

[0036] 在另一方面，本发明包括一种用于细胞疗法的方法，所述方法包括向有需要的患者施用本文公开的经修饰的细胞，其中所述经修饰的细胞的存在可以治疗患者的疾病。

[0037] 在另一方面，本发明涉及一种适用于修饰目标靶基因座中的腺嘌呤的工程化的非天然存在的系统，所述系统包含A)指导分子，所述指导分子包含与正向重复序列连接的指导序列，或编码所述指导分子的核苷酸序列；B)无催化活性的Cas13蛋白，或编码所述无催化活性的Cas13蛋白的核苷酸序列；C)腺苷脱氨酶蛋白或其催化结构域，或编码所述腺苷脱氨酶蛋白或其催化结构域的核苷酸序列；其中所述腺苷脱氨酶蛋白或其催化结构域与所述Cas13蛋白或所述指导分子共价或非共价连接，或者适用于在递送后与其连接；其中所述指导序列能够与包含腺嘌呤的靶RNA序列杂交以形成RNA双链体，其中所述指导序列在对应于

所述腺嘌呤的位置处包含非配对胞嘧啶，从而导致在形成的RNA双链体中产生A-C错配。

[0038] 在另一方面，本发明涉及一种适用于修饰目标靶基因座中的腺嘌呤的工程化的非天然存在的载体系统，所述载体系统包含a)、b)和c)的核苷酸序列。

[0039] 在另一方面，本发明涉及一种工程化的非天然存在的载体系统，所述载体系统包含一个或多个载体，所述一个或多个载体包含：第一调控元件，其可操作地连接到编码包含所述指导序列的所述指导分子的核苷酸序列；第二调控元件，其可操作地连接到编码所述无催化活性的Cas13蛋白的核苷酸序列；以及编码腺苷脱氨酶蛋白或其催化结构域的核苷酸序列，其受所述第一调控元件或第二调控元件的控制或可操作地连接到第三调控元件；其中如果所述编码腺苷脱氨酶蛋白或其催化结构域的核苷酸序列可操作地连接到第三调控元件，则所述腺苷脱氨酶蛋白或其催化结构域适用于在表达后连接到所述指导分子或所述Cas13蛋白；其中组分A)、B)和C)位于系统的相同或不同载体上。

[0040] 由于本文公开的方法证明Cas13蛋白能够在哺乳动物细胞中发挥作用以结合并特异性裂解RNA，所以另外的扩展应用包括编辑剪接变体，以及测量RNA结合蛋白如何与RNA相互作用。

[0041] 在另一方面，本发明涉及包含本文公开的系统的体外或离体宿主细胞或其后代或者细胞系或其后代。宿主细胞或其后代可以是真核细胞、动物细胞、人类细胞或植物细胞。

[0042] 在另一方面，本公开包括修饰目标靶RNA序列中的腺嘌呤的方法，所述方法包括向所述靶RNA递送：(a)无催化活性的(死亡)Cas13蛋白；(b)包含与正向重复序列连接的指导序列的指导分子；以及(c)突变以将活性转化为胞苷脱氨酶的腺苷脱氨酶蛋白或其催化结构域；其中所述腺苷脱氨酶蛋白或其催化结构域与所述死亡Cas13蛋白或所述指导分子共价或非共价连接，或者适用于在递送后与其连接；其中所述指导分子与所述死亡Cas13蛋白形成复合物并指导所述复合物结合所述目标靶RNA序列，其中所述指导序列能够与包含所述腺嘌呤的靶序列杂交以形成RNA双链体，其中所述指导序列在对应于所述腺嘌呤的位置处包含非配对胞嘧啶，从而导致在形成的RNA双链体中产生A-C错配；其中所述腺苷脱氨酶蛋白或其催化结构域使所述RNA双链体中的所述腺嘌呤脱氨。

[0043] 在一些实施方案中，腺苷脱氨酶在选自E396、C451、V351、R455、T375、K376、S486、Q488、R510、K594、R348、G593、S397、H443、L444、Y445、F442、E438、T448、A353、V355、T339、P539、V525和I520的一个或多个位置处突变。在一些示例实施方案中，腺苷脱氨酶在选自E488、V351、S486、T375、S370、P462和N597的一个或多个位置处突变。

[0044] 在另一方面，本公开包括一种适用于修饰目标靶基因座中的腺嘌呤的工程化的非天然存在的系统，所述系统包含(A)指导分子，所述指导分子包含与正向重复序列连接的指导序列，或编码所述指导分子的核苷酸序列；(b)无催化活性的Cas13蛋白，或编码所述无催化活性的Cas13蛋白的核苷酸序列；(c)腺苷脱氨酶蛋白或其催化结构域，或编码所述腺苷脱氨酶蛋白或其催化结构域的核苷酸序列，其中所述腺苷脱氨酶经修饰以将活性转化为胞苷脱氨酶；其中所述腺苷脱氨酶蛋白或其催化结构域与所述Cas13蛋白或所述指导分子共价或非共价连接，或者适用于在递送后与其连接；其中所述指导序列能够与包含腺嘌呤的靶RNA序列杂交以形成RNA双链体，其中所述指导序列在对应于所述腺嘌呤的位置处包含非配对胞嘧啶，从而导致在形成的RNA双链体中产生A-C错配。

[0045] 在一些实施方案中，腺苷脱氨酶通过选自E396、C451、V351、R455、T375、K376、

S486、Q488、R510、K594、R348、G593、S397、H443、L444、Y445、F442、E438、T448、A353、V355、T339、P539、V525和I520的一个或多个突变修饰。在一些示例实施方案中，腺苷脱氨酶在选自E488、V351、S486、T375、S370、P462和N597的一个或多个位置处突变。

[0046] 在考虑对所说明的示例实施方案的以下详细描述后，示例实施方案的这些和其他方面、目标、特征和优点对于本领域普通技术人员将变得显而易见。

## 附图说明

[0047] 通过参考阐述说明性实施方案(其中可利用本发明的原理)的以下详细描述以及附图，将获得对本发明的特征和优点的理解，在附图中：

[0048] 图1示出了用于在目标靶RNA序列处对腺嘌呤进行靶向脱氨的本发明的示例实施方案，在此以Cas13b蛋白为例。

[0049] 图2示出了RNA编辑作为在转录物水平上(诸如当使用Cas13b时)治疗人类疾病的治疗策略的发展。通过靶向荧光素酶转录物中工程化的提前终止密码子的Cas13-ADAR2融合体进行RNA碱基编辑的示意图。

[0050] 图3指导物位置和长度优化，以恢复荧光素酶表达。

[0051] 图4腺嘌呤脱氨酶蛋白的示例性序列。(SEQ ID No.650-656)

[0052] 图5示例性实施方案中使用的指导物(显示的是SEQ ID No.657-660的核苷酸序列以及SEQ ID NO:696的氨基酸序列)。

[0053] 图6：编辑效率与编辑后的碱基距离DR更远且具有长RNA双链体(通过延长引导物长度来实现)有关。

[0054] 图7编辑效率越高，编辑位点距离DR/蛋白质结合区域越远。

[0055] 图8编辑位点与DR的距离。

[0056] 图9A和9B：在N末端或C末端上将ADAR1或ADAR2融合到Cas13b12(双R HEPN突变体)上。指导物与荧光素酶中的终止密码子完美匹配。信号似乎与编辑后的碱基与指导物的5'端之间的距离有关，其中较短的距离越短可提较好的编辑。

[0057] 图10：进行Cluc/Gluc平铺用于Cas13a/Cas13b干扰。

[0058] 图11A-11C：通过NGS(荧光素酶报告基因)进行ADAR编辑定量。

[0059] 图12：通过NGS(KRAS和PPIB)进行ADAR编辑定量。

[0060] 图13A-13C：RNA Seq的Cas13a/b+shRNA特异性。

[0061] 图14：用于减少脱靶的错配特异性(A:A或A:G)(SEQ ID No.661-668)

[0062] 图15：中靶活性的错配

[0063] 图16：ADAR基序偏好

[0064] 图17：较大的空泡可增强RNA编辑效率

[0065] 图18：对转录物(SEQ ID No.669-672)中的多个A进行编辑

[0066] 图19：用于RNA编辑的指导物长度滴定

[0067] 图20A-20F：经哺乳动物密码子优化的Cas13b直系同源物介导高效的RNA敲低。(图20A)代表性Cas13a、Cas13b和Cas13c基因座以及相关crRNA的示意图。(图20B)用于测量HEK293FT细胞中Cas13a裂解活性的荧光素酶测定的示意图。(图20C)利用19个Cas13a、15个Cas13b和5个Cas13c直系同源物使用靶向Cluc的两种不同引导物的RNA敲低效率。将荧光素

酶表达归一化为非靶向指导物控制条件下的表达。(图20D)利用靶向Cluc的两种不同引导RNA,并利用三种不同的NLS和NES标签测定在部分C中表现的前7种直系同源物的活性。(图20E)比较Cas13b12和Cas13a2(LwCas13a)对Gluc和Cluc的敲低活性。沿转录物平铺指导物,并且Cas13b12和Cas13a2之间的指导物在位置上匹配。(图20F)Cas13a2、Cas13b6、Cas13b11和Cas13b12对内源KRAS转录物的指导物敲低,并将其与对应的shRNA进行比较。

[0068] 图21A-21G:Cas13酶介导哺乳动物细胞中的特异性RNA敲低。(图21A)用于Cas13a/b错配特异性测试的半简并靶序列的示意图。(SEQ ID No.673-694)(图21B)Cas13a/b的单个错配敲低数据的热图。将敲低归一化为每种酶的非靶向(NT)指导物。(图21C)Cas13a的双错配敲低数据。每个错配的位置在X和Y轴上指示。敲低数据是一组给定位置的所有双错配的总和。将数据归一化为每种酶的NT指导物。(图21D)Cas13b的双错配敲低数据。有关说明,请参见C。(图21E)将Cas13a/b的全转录物组特异性和位置匹配指导物的shRNA进行比较的RNA-seq数据。Y轴表示靶向条件的读段计数,X轴表示非靶向条件的计数。(图21F)由Cas13a/b和shRNA的RNA-seq数据计算的RNA表达。(图21G)来自RNA-seq数据的Cas13a/b和shRNA的显著脱靶。使用FDR<0.05计算显著脱靶。

[0069] 图22A-22F:无催化活性的Cas13b-ADAR融合体能够在哺乳动物细胞中进行靶向RNA编辑。(图22A)利用Cas13b-ADAR融合体进行RNA编辑以去除海萤荧光素酶转录物上的终止密码子的示意图。(图22B)多个指导物位置处与野生型ADAR2融合的Cas13b和与高活性ADAR2E488Q突变体融合的Cas13b之间的RNA编辑比较。将荧光素酶表达归一化为高斯荧光素酶对照值。(图22C)被设计成靶向编辑位点周围的各种位置的30、50、70和84nt指导物之间的RNA编辑比较。(图22D)示出用于相对于海萤荧光素酶转录物上的终止密码子进行测序定量的指导物的位置和长度的示意图。(SEQ ID No.695)(图22E)通过测序定量的每种指导物设计在海萤荧光素酶转录物上的对应腺嘌呤碱基处的中靶和脱靶编辑效率。(图22F)具有与靶向腺嘌呤相对的不同碱基的指导物的荧光素酶读数。

[0070] 图23:利用Cas13b-ADAR融合体进行内源RNA编辑。对内源KRAS和PPIB基因座的内源Cas13b12-ADAR编辑进行下一代测序。靶向每个转录物的两个不同区域,并在靶向腺嘌呤附近的所有腺嘌呤处对A->G编辑进行定量。利用Cas13b12-ADAR靶向A549细胞中STK11转录物中的提前终止位点,从而恢复STK11蛋白的表达。

[0071] 图24:用于确定最佳指导物位置的策略。

[0072] 图25A-25C:(图25A)Cas13b-huADAR2促进突变的荧光素酶转录物的修复。(图25B)Cas13b-huADAR1促进突变的荧光素酶转录物的修复。(图25C)人ADAR1和人ADAR2的比较。

[0073] 图26:E488Q与wt dADAR2编辑的比较。E488Q是dADAR2的高活性突变体。

[0074] 图27A-27B:Cas13b-huADAR2-E488Q靶向的转录物包含预期的A-G编辑。(图27A)热图。(图27B)模板中的位置。如热图中所示,仅示出具有到G的编辑率的A位点。

[0075] 图28A-28B:指导物的内源平铺。(图28A)KRAS:热图。(上图)模板中的位置。(下图)如热图中所示,仅示出具有到G的编辑率的A位点。(图28B)PPIB:热图。(上图)模板中的位置。(下图)如热图中所示,仅示出具有到G的编辑率的A位点。

[0076] 图29:非靶向编辑。

[0077] 图30:接头优化。

[0078] 图31:Cas13b ADAR可以用于校正患者表达的cDNA中的致病性A>G突变。

[0079] 图32:Cas13b-ADAR对5'G基序略有限制。

[0080] 图33A-33D:筛选简并PFS位置对编辑效率的影响。所有PFS(4-N)身份具有比非靶向更高的编辑。在图A中显示的是SEQ ID No.697、698、699和703。

[0081] 图34:减少靶转录物中的脱靶编辑。

[0082] 图35:减少靶转录物中的脱靶编辑。

[0083] 图36A-36B:Cas13b-ADAR转录组特异性。中靶编辑为71%。(图36A)靶向指导物;482个重要位点。(图36B)非靶向指导物;949个重要位点。注意,0号染色体是Gluc,1号染色体是Cluc;然后,人类染色体在那之后按顺序排列。

[0084] 图37A-37B:Cas13b-ADAR转录组特异性。(图37A)靶向指导物;(图37B)非靶向指导物。

[0085] 图38:与竞争ADAR编辑策略相比,Cas13b的效率最高。

[0086] 图39A-39D:竞争RNA编辑系统。(A-B)BoxB;中靶编辑为63%;(图39A)靶向指导物—2020个重要位点;(图39B)非靶向指导物—1805个重要位点。(图39C-39D)Stafforst;中靶编辑为36%;(图39C)靶向指导物—176个重要位点;(图39D)非靶向指导物—186个重要位点。

[0087] 图40:ADAR的剂量滴定。crRNA的量是恒定的。

[0088] 图41A-41D:剂量响应对特异性的影响。(图41A-41B)150ng Cas13-ADAR;中靶编辑为83%;(图41A)靶向指导物—1231个重要位点;(图41B)非靶向指导物—520个重要位点。(图41C-41D)10ng Cas13-ADAR;中靶编辑为80%;(图41C)靶向指导物—347个重要位点;(图41D)非靶向指导物—223个重要位点。

[0089] 图42A-42B:ADAR1似乎比ADAR2更具特异性。中靶编辑为29%。(图42A)靶向指导物;11个重要位点。(图42B)非靶向指导物;6个重要位点。注意,0号染色体是Gluc,1号染色体是Cluc;然后,人类染色体在那之后按顺序排列。

[0090] 图43A-43D:ADAR特异性突变体具有增强的特异性。(图43A)靶向指导物。(图43B)非靶向指导物。(图43C)靶向与非靶向比率。(图43D)靶向指导物和非靶向指导物。

[0091] 图44:沿每个残基与RNA靶标的接触点绘制的ADAR突变体荧光素酶结果。显示的是SEQ ID No.704和735。

[0092] 图45:ADAR特异性突变体具有增强的特异性。紫色点是选择用于全转录组脱靶NGS分析的突变体。红色点是起点(即E488Q突变体)。注意,所有其他突变体也都具有E488Q突变。

[0093] 图46A-46B:根据NGS,ADAR突变体更具特异性。(图46A)中靶。(图46B)脱靶。

[0094] 图47A-47B:ADAR特异性突变体的荧光素酶数据与NGS匹配。(图47A)选择用于NGS的靶向指导物。(图47B)选择用于NGS的非靶向指导物。荧光素酶数据与图46A-46B中的NGS数据匹配。利用非靶向指导物的活性较少的直系同源物在转录组中具有较少的脱靶,并且它们的中靶编辑效率可以通过靶向指导物荧光素酶条件来预测。

[0095] 图48A-48C:Cas13b 12的C端截短物在ADAR编辑中仍然具有高活性。

[0096] 图49A-49G:用于RNA敲低的高活性Cas13b直系同源物的表征。图49A)原型Cas13基因座和对应crRNA结构的示意图。图49B)使用两种不同的指导物评估19个Cas13a、15个Cas13b和7个Cas13c直系同源物的荧光素酶敲低。使用两种指导物有效敲低的直系同源物

标有其宿主生物名称。图49C)通过针对Gluc平铺指导物并测量荧光素酶表达来比较PspCas13b和LwaCas13a敲低活性。图49D)通过针对Cluc平铺指导物并测量荧光素酶表达来比较PspCas13b和LwaCas13a敲低活性。图49E)与LwaCas13a(红色)和shRNA(黑色)的Gluc靶向条件(y轴)相比,在非靶向对照(x轴)的RNA-seq文库中检测到的所有基因的表达水平(用log2(每百万转录物(TPM))值表示)。显示的是三个生物学重复的平均值。Gluc转录物数据点被标记。图49F)与PspCas13b(蓝色)和shRNA(黑色)的Gluc靶向条件(y轴)相比,在非靶向对照(x轴)的RNA-seq文库中检测到的所有基因的表达水平(用log2(每百万转录物(TPM))值表示)。显示的是三个生物学重复的平均值。Gluc转录物数据点被标记。图49G)从E和F中的全转录组分析得出的LwaCas13a、PspCas13b和shRNA的Gluc敲低的显著脱靶数量。

[0097] 图50A-50E:对dCas13b-ADAR融合体进行工程化以用于RNA编辑。图50A)dCas13b-ADAR融合蛋白进行RNA编辑的示意图。图50B)海萤荧光素酶W85X靶标和靶向指导物设计的示意图。(SEQ ID No.700和701)图50C)利用长度为30、50、70或84nt的平铺指导物对Cas13b-dADAR1(左)和Cas13b-ADAR2-cd(右)的荧光素酶活性恢复进行定量。图50D)用于靶向海萤荧光素酶W85X的靶位点的示意图。(SEQ ID No.702)图50E)对靶向海萤荧光素酶W85X的50nt指导物的A->I编辑进行测序定量。

[0098] 图51A-51D:测量通过REPAIRv1进行RNA编辑的序列灵活性。图51A)用于确定通过REPAIRv1进行RNA编辑的原间隔侧接位点(PFS)偏好的筛选示意图。显示的是SEQ ID No.737。图51B)两个不同的编辑位点处所有4-N PFS组合的RNA编辑效率的分布。图51C)对所有可能的3碱基基序在Cluc W85处的REPAIRv1编辑百分比进行定量。图51D)所有可能的3碱基基序在Cluc W85处的RNA编辑的5'和3'碱基偏好的热图。

[0099] 图52A-52G:利用REPAIRv1校正与疾病相关的突变。图52A)靶标以及用于靶向AVPR2 878G>A的指导物设计的示意图。(SEQ ID No.705-708)。图52B)使用具有三种不同指导物设计的REPAIRv1将AVPR2中的878G>A突变校正为不同百分比。图52C)靶标以及用于靶向FANCC 1517G>A的指导物设计的示意图。(SEQ ID No.709-712)图52D)使用具有三种不同指导物设计的REPAIRv1将FANCC中的1517G>A突变校正为不同百分比。图52E)使用REPAIRv1对34种与疾病相关的G>A突变的编辑百分比进行定量。图52F)对如ClinVar数据库所注释的可被校正的所有可能的G>A突变进行分析。图52G)相对于每个基序的REPAIRv1编辑效率(如在Gluc转录物上定量),示出了ClinVar中所有G>A突变的编辑基序的分布。

[0100] 图53A-53D:对REPAIRv1的特异性进行表征。图53A)KRAS靶位点和指导物设计的示意图。(SEQ ID No.713-720)图53B)对平铺的KRAS靶向指导物的编辑百分比进行定量。编辑百分比示于中靶和相邻腺苷位点处。对于每种指导物,双链体RNA的区由红色矩形表示。图53C)利用Cluc靶向指导物通过REPAIRv1进行重要RNA编辑的全转录组位点。中靶位点Cluc位点(254A>G)用橙色突出显示。图53D)利用非靶向指导物通过REPAIRv1进行重要RNA编辑的全转录组位点。

[0101] 图54A-54F:对ADAR2进行合理诱变以提高REPAIRv1的特异性。图54A)对通过各种dCas13-ADAR2突变体恢复荧光素酶信号进行定量,以及它们的特异性评分,所述特异性评分沿着关键ADAR2脱氨酶残基与dsRNA靶标之间的接触示意图绘制。特异性评分被定义为靶向指导物和非靶向指导物条件之间的荧光素酶信号之比。显示的是SEQ ID No.738和739。图54B)对通过各种dCas13-ADAR2突变体恢复荧光素酶信号进行定量,以及它们的特异性评

分。图54C)通过对mRNA进行全转录组测序对每个dCas13-ADAR2突变体的中靶编辑分数以及显著脱靶的数量进行测量。图54D)利用靶向Cluc中预先终止位点的指导物通过REPAIRv1和REPAIRv2进行重要RNA编辑的全转录组位点。中靶Cluc位点(254A>G)用橙色突出显示。图54E)中靶Cluc编辑位点(SEQ ID No.721、740-745和763-792)(254A>G)周围的RNA测序读数,突出显示了REPAIRv1和REPAIRv2之间的脱靶编辑差异。所有A>G编辑均用红色突出显示,而测序错误用蓝色突出显示。图54F)利用靶向内源KRAS和PPIB转录物中的框外UAG位点的指导物通过REPAIRv1和REPAIRv2进行RNA编辑。每个条件行的中靶编辑分数被示为右侧的横向条形图。由指导RNA形成的双链体区由红色轮廓框示出。

[0102] 图55A-55C:对Cas13b直系同源物进行细菌筛选以用于体内效率和PFS测定。图55A)用于确定Cas13b直系同源物的PFS的细菌测定的示意图。将具有β-内酰胺酶靶向间隔子(SEQ ID No.722)的Cas13b直系同源物与β-内酰胺酶表达质粒共转化并进行双重选择。图55B)对靶向β-内酰胺酶的Cas13b直系同源物的干扰活性进行定量,由集落形成单位(cfu)测量。图55C)Cas13b直系同源物的PFS标识,由细菌测定中的耗竭序列确定。

[0103] 图56A-56E:对Cas13b敲低的优化以及对错配特异性的进一步表征。图56A)使用与各种核定位和核输出标签融合的前2个Cas13a和前4个Cas13b直系同源物测量利用两种不同指导物进行的Gluc敲低。图56B)利用四种不同指导物测量LwaCas13a、RanCas13b、PguCas13b和PspCas13b对KRAS的敲低进行测量,并且与四个位置匹配的shRNA对照进行比较。图56C)用于评估LwaCas13a和PspCas13b敲低特异性的单错配和双错配质粒文库的示意图。中靶序列中以及在正向侧接靶位点5'和3'端的3个位置中存在每种可能的单错配和双错配。(SEQ ID No.723-734)图56D)对于LwaCas13a和PspCas13b条件两者,将具有指示的单错配的转录物的耗竭水平绘制为热图。(SEQ ID No.723和736)图56E)对于LwaCas13a和PspCas13b条件两者,将具有指示的双错配的转录物的耗竭水平绘制为热图(SEQ ID No.723和736)。

[0104] 图57A-57F:对用于dCas13-ADAR2RNA编辑的设计参数的表征。图57A)野生型Cas13b和无催化活性的H133A/H1058A Cas13b(dCas13b)的Gluc靶向的敲低效率。图57B)对通过与野生型ADAR2催化结构域或高活性E488Q突变体ADAR2催化结构域融合的dCas13b恢复荧光素酶活性进行定量,用平铺Cluc靶向指导物进行测试。图57C)对靶向海萤荧光素酶W85X的30nt指导物的A->I编辑进行指导物设计和测序定量(SEQ ID No.793-801)。图57D)对靶向PPIB的50nt指导物的A->I编辑进行指导物设计和测序定量(SEQ ID No.746-753)。图57E)接头选择对REPAIRv1恢复荧光素酶活性的影响。图57F)与靶向腺苷相对的碱基鉴定对REPAIRv1恢复荧光素酶活性的影响(SEQ ID No.754和755)。

[0105] 图58:G>A突变的ClinVar基序分布。在ClinVar数据库中观察到的所有G>A突变的每个可能的三联体基序的数量。

[0106] 图59:dCas13b的截短物仍具有RNA编辑功能。dCas13b的各种N端和C端截短物允许进行RNA编辑,如通过恢复荧光素酶信号所测量。

[0107] 图60A-60F:其他可编程ADAR系统与dCas13-ADAR2编辑器的比较。图60A)两种可编程ADAR方案的示意图:基于BoxB的靶向和全长ADAR2靶向。在BoxB方案(上图)中,将ADAR2脱氨酶结构域($ADAR2_{DD}$(E488Q))与被称为lambda N (λN)的小细菌病毒蛋白融合,该蛋白与被称为BoxB-λ的小RNA序列特异性结合。然后,包含两个BoxB-λ发夹的指导RNA可以指导

ADAR2$_{DD}$(E488Q),-λN进行位点特异性编辑。在全长ADAR2方案(下图)中,ADAR2的dsRNA结合结构域与指导RNA中的发夹结合,从而允许进行可编程ADAR2编辑(SEQ ID No.756-760)。图60B)利用靶向Cluc的指导物和非靶向指导物通过BoxB-ADAR2$_{DD}$(E488Q)进行重要RNA编辑的全转录组位点。中靶Cluc位点(254A>G)用橙色突出显示。图60C)利用靶向Cluc的指导物和非靶向指导物通过ADAR2进行重要RNA编辑的全转录组位点。中靶Cluc位点(254A>G)用橙色突出显示。图60D)利用靶向Cluc的指导物和非靶向指导物通过REPAIRv1进行重要RNA编辑的全转录组位点。中靶Cluc位点(254A>G)用橙色突出显示。图60E)对BoxB-ADAR2$_{DD}$(E488Q)、ADAR2和REPAIRv1针对Cluc靶向指导物的中靶编辑率百分比进行定量。图60F)可编程ADAR系统在不同靶向和非靶向条件之间的脱靶位点的重叠。

[0108] 图61A-61C:dCas13b-ADAR2突变体的效率和特异性。图61A)对Cluc靶向和非靶向指导物通过dCas13b-ADAR2$_{DD}$(E488Q)突变体恢复荧光素酶活性进行定量。图61B)如通过全转录组测序定量的靶向和非靶向指导物的比例与RNA编辑脱靶的数量之间的关系。图61C)对dCas13b-ADAR2$_{DD}$(E488Q)突变体的全转录组脱靶RNA编辑位点的数量与中靶Cluc编辑效率进行定量。

[0109] 图62A-62B:dCas13b-ADAR2$_{DD}$(E488Q)突变体进行RNA编辑的全转录组特异性。图62A)dCas13b-ADAR2$_{DD}$(E488Q)突变体利用靶向Cluc的指导物进行重要RNA编辑的全转录组位点。中靶Cluc位点(254A>G)用橙色突出显示。图62B)dCas13b-ADAR2$_{DD}$(E488Q)突变体利用非靶向指导物进行重要RNA编辑的全转录组位点。

[0110] 图63A-63C:对dCas13b-ADAR2$_{DD}$(E488Q)编辑的脱靶中的基序偏倚的表征。图63A)对于每个dCas13b-ADAR2$_{DD}$(E488Q)突变体,示出了转录组中所有A>G脱靶编辑中存在的基序。图63B)示出了利用靶向和非靶向指导物进行REPAIRv1时每个基序身份的脱靶A>G编辑的分布。图63C)示出了利用靶向和非靶向指导物进行REPAIRv2时每个基序身份的脱靶A>G编辑的分布。

[0111] 图64A-64F:对REPAIRv1和REPAIRv2脱靶的进一步表征。图64A)REPAIRv1的每个转录物的脱靶数量的直方图。图64B)REPAIRv2的每个转录物的脱靶数量的直方图。图64C)REPAIRv1脱靶的变异效应预测。图64D)REPAIRv1脱靶的潜在致癌效应的分布。图64E)REPAIRv2脱靶的变异效应预测。图64F)REPAIRv2脱靶的潜在致癌效应的分布。

[0112] 图65A-65C:REPAIRv1和REPAIRv2的RNA编辑效率和特异性。图65A)对REPAIRv1和REPAIRv2在靶腺苷和相邻位点处利用KRAS靶向指导物1对KRAS进行编辑的百分比进行定量。图65B)对REPAIRv1和REPAIRv2在靶腺苷和相邻位点处利用KRAS靶向指导物3对KRAS进行编辑的百分比进行定量。图65C)对REPAIRv1和REPAIRv2在靶腺苷和相邻位点处利用PPIB靶向指导物2对PPIB进行编辑的百分比进行定量。

[0113] 图66A-66G:对具有A>G RNA编辑器的所有潜在的密码子变化的展示。图66A)通过A>I编辑实现的所有潜在密码子转换的表。图66B)展示通过A>I编辑实现的所有潜在密码子转换的密码子表。图66C)示出了通过ADAR和dCas13b进行基因编辑。图66D)示出了通过ADAR纠正孟德尔疾病的示例。图66E)示出了疾病保护等位基因的多重产生的示例。图66F)示出了调节蛋白功能的示例。图66G)示出了剪接调节的示例。

[0114] 图67:Psp dCas13b的其他截短物。

[0115] 图68:剂量对脱靶活性的潜在影响。

[0116] 图69示出了筛选的ADAR突变的活性。

[0117] 图70示出了突变P462A在不同基序上的活性。

[0118] 图71示出了突变P462改善的基序。

[0119] 图72示出了鉴定出N579I的ADAR突变筛选的活性。

[0120] 图73示出了对N579I的验证。

[0121] 图74示出了具有30bp指导序列的RESCUEv3的性能。

[0122] 图75示出了50bp指导序列和30bp指导序列与各种ADAR突变体的比较。

[0123] 图76表明ADAR突变的活性是Cas13依赖性的。

[0124] 本文中的附图仅用于说明性目的,不一定按比例绘制。

## 具体实施方式

[0125] 一般定义

[0126] 除非另有规定,否则本文所用的技术和科学术语具有与本公开所属领域的普通技术人员通常所理解的相同的含义。分子生物学中常用术语和技术的定义可在以下文献中找到:Molecular Cloning:A Laboratory Manual,第2版(1989)(Sambrook、Fritsch和Maniatis);Molecular Cloning:A Laboratory Manual,第4版(2012)(Green和Sambrook);Current Protocols in Molecular Biology(1987)(F.M.Ausubel等人编);the series Methods in Enzymology(Academic Press,Inc.):PCR 2:A Practical Approach(1995)(M.J.MacPherson,B.D.Hames和G.R.Taylor编):Antibodies,A Laboraotry Manual(1988)(Harlow和Lane编):Antibodies A Laboraotry Manual,第2版2013(E.A.Greenfield编);Animal Cell Culture(1987)(R.I.Freshney编);Benjamin Lewin,Genes IX,由Jones和Bartlet公布,2008(ISBN 0763752223);Kendrew等人(编),The Encyclopedia of Molecular Biology,由Blackwell Science Ltd.公布,1994(ISBN 0632021829);Robert A.Meyers(编),Molecular Biology and Biotechnology:a Comprehensive Desk Reference,由VCH Publishers,Inc.公布,1995(ISBN 9780471185710);Singleton等人,Dictionary of Microbiology and Molecular Biology第2版,J.Wiley&Sons(New York,N.Y.1994),3月,Advanced Organic Chemistry Reactions,Mechanisms and Structure第4版,John Wiley&Sons(New York,N.Y.1992);以及Marten H.Hofker和Jan van Deursen,Transgenic Mouse Methods and Protocols,第2版(2011)。

[0127] 参考2016年6月17日提交的美国临时申请62/351,662和62/351,803、2016年8月17日提交的美国临时申请62/376,377、2016年10月19日提交的美国临时申请62/410,366、2016年12月9日提交的美国临时申请62/432,240、2017年3月15日提交的美国临时申请62/471,792以及2017年4月12日提交的美国临时申请62/484,786。参考2017年6月19日提交的国际PCT申请PCT/US2017/038154。参考2017年3月15日提交的美国临时申请62/471,710(标题为“Novel Cas13B Orthologues CRISPR Enzymes and Systems”,代理人参考号:BI-10157VP 47627.04.2149)。进一步参考2016年12月9日提交的美国临时申请62/432,553、2017年2月8日提交的美国临时申请62/456,645、2017年3月15日提交的美国临时申请62/471,930(标题为“CRISPR Effector System Based Diagnostics”,代理人参考号:BI-10121BROD0842P)以及2017年4月12日提交的有待转让的美国临时申请(标题为“CRISPR

Effector System Based Diagnostics”,代理人参考号:BI-10121BROD0843P)。

[0128] 除非上下文另有明确指示,否则如本文所用的单数形式“一个”、“一种”和“所述”包括单数和复数指示物。

[0129] 术语“任选的”或“任选地”意指后续描述的事件、情形或替代物可能发生或可能不发生,并且描述包括事件或情形发生的情况和不发生的情况。

[0130] 由端点对数值范围的叙述包括归入相应范围内的所有数值和分数,以及所叙述的端点。

[0131] 如本文所用的术语“约”或“近似”当涉及例如参数、量、时距等可测量的值时,有意涵盖指定值的变化和从指定值的变化,例如指定值和从指定值+/-10%或更小、+/-5%或更小、+/-1%或更小和+/-0.1%或更小的变化,只要此类变化适于在所公开的发明中执行即可。应了解,修饰语“约”或“近似”所涉及的值本身也特定地并优选地公开。

[0132] 在本说明书通篇提及“一个实施方案”、“实施方案”、“示例性实施方案”意指结合实施方案描述的特定特征、结构或特性包括于本发明的至少一个实施方案中。因此,在本说明书通篇各处出现短语“在一个实施方案中”、“在实施方案中”或“示例性实施方案”不一定全部涉及同一实施方案,但可能如此。此外,如本领域技术人员从本公开将显而易见,在一个或多个实施方案中,特定特征、结构或特性可以按任何适合的方式组合。此外,虽然本文所描述的一些实施方案包括其它实施方案中所包括的一些特征而非其它特征,但不同实施方案的特征的组合有意处于本发明的范围内。举例来说,在随附权利要求书中,所要求的实施方案中的任一个可以按任何组合形式使用。

[0133] C2c2现在称为Cas13a。应当理解,术语“C2c2”在本文中与“Cas13a”可互换使用。

[0134] 本文所引用的所有公布、公布的专利文献、和专利申请特此以引用的方式并入,引用程度如同每个个别公布、公布的专利文献、或专利申请特定地并个别地指示以引用的方式并入一样。

[0135] 在下文中描述各种实施方案。应当注意,特定实施方案不旨在作为详尽的描述或作为对本文所讨论的更广泛方面的限制。结合具体实施方案描述的一个方面不一定限于该实施方案,可以与一个或多个任何其他实施方案一起实践。

[0136] 综述

[0137] 本文公开的实施方案提供了用于靶向碱基编辑的系统、构建体和方法。一般来讲,本文公开的系统包含靶向组分和碱基编辑组分。靶向组分的作用是将碱基编辑组分特异性靶向其中一个或多个核苷酸将被编辑的靶核苷酸序列。然后,碱基编辑组分可催化化学反应,以将靶序列中的第一核苷酸转化为第二核苷酸。例如,碱基编辑器可催化腺嘌呤的转化,使得其被细胞的转录或翻译装置读取为鸟嘌呤,反之亦然。同样,碱基编辑组分可催化胞苷转化为尿嘧啶,反之亦然。在某些示例实施方案中,可通过以已知碱基编辑器诸如腺嘌呤脱氨酶或胞苷脱氨酶开始来得到基础编辑器,并且可使用诸如定向进化的方法对其进行修饰以得到新的功能。定向进化技术是本领域已知的,并且可包括WO 2015/184016“High-Throughput Assembly of Genetic Permuatations”中描述的定向技术。

[0138] 在某些示例实施方案中,通过与RNA结合蛋白或其其RNA结合功能结构域融合或连接来将腺苷脱氨酶蛋白募集到目标靶基因座。RNA结合蛋白可结合靶基因座处或附近的序列、基序或结构特征。结构特征可包括核酸的发夹、四环或其他二级结构特征。如本文所用,

“相邻”是指在腺苷脱氨酶可以完成其碱基编辑功能的靶基因座的距离和/或取向内。在某些示例实施方式中,RNA结合蛋白或其功能结构域可结合指导分子的序列、基序或结构序列。指导分子包含与目标靶基因座杂交并由此将腺苷脱氨酶引导到靶基因座的序列。

[0139] 在某些示例实施方案中,RNA结合蛋白或其功能结构域包含RNA识别基序。包含RNA识别基序的示例RNA结合蛋白包括但不限于A2BP1;ACF;BOLL;BRUNOL4;BRUNOL5;BRUNOL6;CCBL2;CGI96;CIRBP;CNOT4;CPEB2;CPEB3;CPEB4;CPSF7;CSTF2;CSTF2T;CUGBP1;CUGBP2;D10S102;DAZ1;DAZ2;DAZ3;DAZ4;DAZAP1;DAZL;DNAJC17;DND1;EIF3S4;EIF3S9;EIF4B;EIF4H;ELAVL1;ELAVL2;ELAVL3;ELAVL4;ENOX1;ENOX2;EWSR1;FUS;FUSIP1;G3BP;G3BP1;G3BP2;GRSF1;HNRNPL;HNRPA0;HNRPA1;HNRPA2B1;HNRPA3;HNRPAB;HNRPC;HNRPCL1;HNRPD;HNRPDL;HNRPF;HNRPH1;HNRPH2;HNRPH3;HNRPL;HNRPLL;HNRPM;HNRPR;HRNBP1;HSU53209;HTATSF1;IGF2BP1;IGF2BP2;IGF2BP3;LARP7;MKI67IP;MSI1;MSI2;MSSP2;MTHFSD;MYEF2;NCBP2;NCL;NOL8;NONO;P14;PABPC1;PABPC1L;PABPC3;PABPC4;PABPC5;PABPN1;POLDIP3;PPARGC1;PPARGC1A;PPARGC1B;PPIE;PPIL4;PPRC1;PSPC1;PTBP1;PTBP2;PUF60;RALY;RALYL;RAVER1;RAVER2;RBM10;RBM11;RBM12;RBM12B;RBM14;RBM15;RBM15B;RBM16;RBM17;RBM18;RBM19;RBM22;RBM23;RBM24;RBM25;RBM26;RBM27;RBM28;RBM3;RBM32B;RBM33;RBM34;RBM35A;RBM35B;RBM38;RBM39;RBM4;RBM41;RBM42;RBM44;RBM45;RBM46;RBM47;RBM4B;RBM5;RBM7;RBM8A;RBM9;RBMS1;RBMS2;RBMS3;RBMX;RBMX2;RBMXL2;RBMY1A1;RBMY1B;RBMY1E;RBMY1F;RBMY2FP;RBPMS;RBPMS2;RDBP;RNPC3;RNPC4;RNPS1;ROD1;SAFB;SAFB2;SART3;SETD1A;SF3B14;SF3B4;SFPQ;SFRS1;SFRS10;SFRS11;SFRS12;SFRS15;SFRS2;SFRS2B;SFRS3;SFRS4;SFRS5;SFRS6;SFRS7;SFRS9;SLIRP;SLTM;SNRP70;SNRPA;SNRPB2;SPEN;SR140;SRRP35;SSB;SYNCRIP;TAF15;TARDBP;THOC4;TIA1;TIAL1;TNRC4;TNRC6C;TRA2A;TRSPAP1;TUT1;U1SNRNPBP;U2AF1;U2AF2;UHMK1;ZCRB1;ZNF638;ZRSR1;以及ZRSR2。

[0140] 在某些示例实施方案中,RNA结合蛋白或其功能结构域可包含K同源性结构域。包含K同源结构域的示例RNA结合蛋白包括但不限于AKAP1;ANKHD1;ANKRD17;ASCC1;BICC1;DDX43;DDX53;DPPA5;FMR1;FUBP1;FUBP3;FXR1;FXR2;GLD1;HDLBP;HNRPK;IGF2BP1;IGF2BP2;IGF2BP3;KHDRBS1;KHDRBS2;KHDRBS3;KHSRP;KRR1;MEX3A;MEX3B;MEX3C;MEX3D;NOVA1;NOVA2;PCBP1;PCBP2;PCBP3;PCBP4;PNO1;PNPT1;QKI;SF1;以及TDRKH。

[0141] 在某些示例实施方案中,RNA结合蛋白包含锌指基序。RNA结合蛋白或其功能结构域可包含Cys2-His2、Gag-knuckle、Treble-clet、锌带、Zn2/Cys6类基序。

[0142] 在某些示例实施方案中,RNA结合蛋白可包含Pumilio同源结构域。

[0143] 在一方面,本发明提供了用于对RNA中、更具体是目标RNA序列中的腺嘌呤进行靶向脱氨的方法。根据本发明的方法,通过可以特异性结合靶序列的CRISPR-Cas复合物将腺苷脱氨酶(AD)蛋白特异性募集到目标RNA序列中的相关腺嘌呤上。为了实现这一点,可以将腺苷脱氨酶蛋白与CRISPR-Cas酶共价连接或作为单独的蛋白提供,但是对其进行调适以便确保将其募集到CRISPR-Cas复合物中。

[0144] 在本发明的方法的特定实施方案中,通过将腺苷脱氨酶或其催化结构域与作为Cas13蛋白的CRISPR-Cas蛋白融合来确保将腺苷脱氨酶募集到靶基因座上。由两种单独的蛋白质产生融合蛋白的方法是本领域已知的,通常涉及使用间隔物或接头。可以将Cas13蛋白在其N末端或C末端与腺苷脱氨酶蛋白或其催化结构域融合。在特定的实施方案中,

CRISPR-Cas蛋白是无活性或死亡的Cas13蛋白，并且与脱氨酶蛋白的N端或其催化结构域连接。

[0145] 如在提及融合蛋白时所用的术语“接头”是指连接蛋白以形成融合蛋白的分子。一般来讲，此类分子除了连接或保持蛋白之间的一定最小距离或其他空间关系外，没有任何特定的生物学活性。然而，在某些实施方案中，可选择接头以影响接头和/或融合蛋白的某些特性，诸如接头的折叠、净电荷或疏水性。

[0146] 用于本发明的方法的合适接头是本领域技术人员众所周知的，包括但不限于直链或支链碳接头、杂环碳接头或肽接头。但是，如本文所用，接头也可以是共价键(碳-碳键或碳-杂原子键)。在特定的实施方案中，接头用于将CRISPR-Cas蛋白和腺苷脱氨酶分开足以确保每个蛋白保留其所需功能特性的距离。优选的肽接头序列采用柔性的延伸构象，并且不表现出发展有序二级结构的倾向。在某些实施方案中，接头可以是化学部分，所述化学部分可以是单体、二聚体、多聚体或聚合物。优选地，接头包含氨基酸。柔性接头中的典型氨基酸包括Gly、Asn和Ser。因此，在特定的实施方案中，接头包含Gly、Asn和Ser氨基酸中的一种或多种的组合。其他接近中性的氨基酸(诸如Thr和Ala)也可用于接头序列中。示例性接头在以下文献中公开:Maratea等人,(1985),Gene 40:39-46;Murphy等人,(1986)Proc.Nat'l.Acad.Sci.USA83:8258-62;美国专利号4,935,233;以及美国专利号4,751,180。例如，可以使用GlySer接头GGS、GGGS或GSG。GGS、GSG、GGGS或GGGGS接头可以3个重复(诸如(GGS)3(SEQ ID No.12)、(GGGGS)3)或5个、6个、7个、9个或甚至12个重复(SEQ ID No.13)或更多个重复使用，以提供合适的长度。在特定的实施方案中，在本文中优选使用接头诸如(GGGGS)3。(GGGGS)6(GGGGS)9或(GGGGS)12可优选地用作替代接头。其他优选的替代接头是(GGGGS)1(SEQ ID No 14)、(GGGGS)2(SEQ ID No.15)、(GGGGS)4、(GGGGS)5、(GGGGS)7、(GGGGS)8、(GGGGS)10或(GGGGS)11。在另一个实施方案中，LEPGEKPYKCPECGKSFSQSGALTRHQRTHTR(SEQ ID No:11)用作接头。在另一个实施方案中，接头是XTEN接头。在特定的实施方案中，CRISPR-cas蛋白是Cas13蛋白，并且通过LEPGEKPYKCPECGKSFSQSGALTRHQRTHTR(SEQ ID No.11)接头与脱氨酶蛋白或其催化结构域连接。在另外的特定实施方案中，Cas13蛋白通过LEPGEKPYKCPECGKSFSQSGALTRHQRTHTR(SEQ ID No.11)接头在C末端与脱氨酶蛋白或其催化结构域的N末端连接。另外，N末端和C末端NLS也可以用作接头(例如，PKKKRKVEASSPKKRKVEAS(SEQ ID No.16))。

[0147] 在本发明的方法的特定实施方案中，腺苷脱氨酶蛋白或其催化结构域作为单独的蛋白被递送到细胞或在细胞内表达，但是经修饰以便能够与Cas13蛋白或指导分子连接。在特定的实施方案中，这通过使用存在于噬菌体外壳蛋白多样性内的正交RNA结合蛋白或衔接蛋白/适配体组合来确保。此类外壳蛋白的实例包括但不限于:MS2、Qβ、F2、GA、fr、JP501、M12、R17、BZ13、JP34、JP500、KU1、M11、MX1、TW18、VK、SP、FI、ID2、NL95、TW19、AP205、ΦCb5、ΦCb8r、ΦCb12r、ΦCb23r、7s和PRR1。适配体可以是天然存在的或合成的寡核苷酸，这些寡核苷酸已经通过重复轮的体外选择或SELEX(指数富集配体系统进化)而被工程化为与特定靶标结合。

[0148] 在本发明的方法和系统的特定实施方案中，指导分子具有可以募集衔接蛋白的一个或多个不同RNA环或者一个或多个不同序列。指导分子可在不与Cas13蛋白冲突的情况下通过插入可募集衔接蛋白的一个或多个不同RNA环或者一个或多个不同序列而延伸，所述

衔接蛋白可以与所述一个或多个不同RNA环或者所述一个或多个不同序列结合。经修饰的指导物的实例及其在将效应子结构域募集到CRISPR-Cas复合物中的用途在Konermann(Nature 2015,517(7536):583-588)中提供。在特定的实施方案中,适配体是选择性地结合哺乳动物细胞中的二聚化MS2噬菌体外壳蛋白并被引入指导分子中(诸如茎环和/或四环中)的最小发夹适配体。在这些实施方案中,腺苷脱氨酶蛋白与MS2融合。然后腺苷脱氨酶蛋白与CRISPR-Cas蛋白和对应的指导RNA一起被共同递送。

[0149] 如本文所用,术语"AD功能化的CRISPR系统"是指核酸靶向和编辑系统,所述系统包含:(a)CRISPR-Cas蛋白,更具体是无催化活性的Cas13蛋白;(b)包含指导序列的指导分子;以及(c)腺苷脱氨酶蛋白或其催化结构域;其中腺苷脱氨酶蛋白或其催化结构域与CRISPR-Cas蛋白或指导分子共价或非共价连接,或者适用于在递送后与其连接;其中指导序列与靶序列基本上互补,但是包含对应于被靶向进行脱氨的A的非配对C,从而导致在由指导序列和靶序列形成的RNA双链体中产生A-C错配。对于真核细胞中的应用,CRISPR-Cas蛋白和/或腺苷脱氨酶优选地被NLS标记。

[0150] 在一些实施方案中,组分(a)、(b)和(c)作为核糖核蛋白复合物被递送到细胞。核糖核蛋白复合物可以通过一种或多种脂质纳米颗粒递送。

[0151] 在一些实施方案中,组分(a)、(b)和(c)作为一种或多种RNA分子(诸如编码CRISPR-Cas蛋白、腺苷脱氨酶蛋白以及任选的衔接蛋白的一种或多种指导RNA和一种或多种mRNA分子)被递送到细胞。RNA分子可以通过一种或多种脂质纳米颗粒递送。

[0152] 在一些实施方案中,组分(a)、(b)和(c)作为一种或多种DNA分子被递送到细胞。在一些实施方案中,一种或多种DNA分子包含在一种或多种载体(诸如病毒载体(例如,AAV))中。在一些实施方案中,一种或多种DNA分子包含可操作地被构造成表达CRISPR-Cas蛋白、指导分子以及腺苷脱氨酶蛋白或其催化结构域的一个或多个调控元件,任选地其中一个或多个调控元件包含诱导型启动子。

[0153] 在一些实施方案中,CRISPR-Cas蛋白是死亡Cas13。在一些实施方案中,死亡Cas13是死亡Cas13a蛋白,其在HEPN结构域中包含一个或多个突变。在一些实施方案中,死亡Cas13a包含对应于瓦氏纤毛菌(LwaCas13a)中的R474A和R1046A的突变。在一些实施方案中,死亡Cas13是死亡Cas13b蛋白,其包含源自动物溃疡伯杰菌ATCC 43767的Cas13b蛋白的R116A、H121A、R1177A、H1182A中的一个或多个或者Cas13b直系同源物的与其对应的氨基酸位置。

[0154] 在一些实施方案中,指导分子能够与包含要在RNA序列内脱氨的腺嘌呤的靶序列杂交以形成RNA双链体,该RNA双链体包含与所述腺嘌呤相对的非配对胞嘧啶。在形成RNA双链体后,指导分子与Cas13蛋白形成复合物,并指导该复合物在目标靶RNA序列处结合RNA多核苷酸。下文提供了AD功能化的CRISPR-Cas系统的指导物方面的细节。

[0155] 在一些实施方案中,具有标准长度(例如LawCas13)的Cas13指导RNA用于与靶DNA形成RNA双链体。在一些实施方案中,长于标准长度(例如LawCas13a)的Cas13指导分子用于与靶DNA形成RNA双链体,包括Cas13-指导RNA-靶DNA复合物的外部。

[0156] 在至少第一设计中,AD功能化的CRISPR系统包含:(a)与CRISPR-Cas蛋白融合或连接的腺苷脱氨酶,其中所述CRISPR-Cas蛋白是无催化活性的;以及(b)包含指导序列的指导分子,所述指导序列被设计成在指导序列与靶序列之间形成的RNA双链体中引入A-C错配。

在一些实施方案中,CRISPR-Cas蛋白和/或腺苷脱氨酶在N末端或C末端或两者上被NLS标记。

[0157] 在至少第二设计中,AD功能化的CRISPR系统包含:(a)无催化活性的CRISPR-Cas蛋白,(b)包含指导序列和适配体序列(例如,MS2RNA基序或PP7RNA基序)的指导分子,所述指导序列被设计成在指导序列与靶序列之间形成的RNA双链体中引入A-C错配,并且所述适配体序列能够与衔接蛋白(例如,MS2外壳蛋白或PP7外壳蛋白)结合;以及(c)与衔接蛋白融合或连接的腺苷脱氨酶,其中适配体和衔接蛋白的结合将腺苷脱氨酶募集到在指导序列与靶序列之间形成的RNA双链体,以在A-C错配的A处进行靶向脱氨。在一些实施方案中,衔接蛋白和/或腺苷脱氨酶在N末端或C末端或两者上被NLS标记。CRISPR-Cas蛋白也可以被NLS标记。

[0158] 使用不同的适配体和对应的衔接蛋白还可以实现正交基因编辑。在腺苷脱氨酶与胞苷脱氨酶组合用于正交基因编辑/脱氨的一个实例中,靶向不同基因座的sgRNA用不同RNA环修饰,以便分别募集MS2-腺苷脱氨酶和PP7-胞苷脱氨酶(或PP7-腺苷脱氨酶和MS2-胞苷脱氨酶),从而分别导致在目标靶基因座处对A或C进行正交脱氨。PP7是噬菌体假单胞菌的RNA结合外壳蛋白。像MS2一样,它结合特定的RNA序列和二级结构。PP7RNA识别基序与MS2的识别基序不同。因此,PP7和MS2可以被多重化,以同时介导不同基因组基因座处的不同效应。例如,靶向基因座A的sgRNA可以用MS2环修饰,从而募集MS2-腺苷脱氨酶,而靶向基因座B的另一种sgRNA可以用PP7环修饰,从而募集PP7-胞苷脱氨酶。因此,在同一细胞中,实现了正交的基因座特异性修饰。该原理可以扩展为结合其他正交RNA结合蛋白。

[0159] 在至少第三设计中,AD功能化的CRISPR系统包含:(a)插入CRISPR-Cas蛋白的内部环或非结构化区域中的腺苷脱氨酶,其中所述CRISPR-Cas蛋白是无催化活性的或切口酶;以及(b)包含指导序列的指导分子,所述指导序列被设计成在指导序列与靶序列之间形成的RNA双链体中引入A-C错配。

[0160] 适合插入腺苷脱氨酶的CRISPR-Cas蛋白断裂位点可以借助于晶体结构来鉴定。如果直系同源物与预期的CRISPR-Cas蛋白之间存在较高程度的同源性,则可以使用直系同源物的晶体结构。

[0161] 断裂位置可位于区域或环内。优选地,断裂位置发生在其中氨基酸序列中断不会导致结构特征(例如α螺旋或β折叠)部分或完全破坏的地方。非结构化区域(由于结构化不足以使其“冻结”在晶体中而未出现在晶体结构中的区域)通常是优选选项。非结构化区域或外部环内的位置可不必精确地是上面提供的数字,而是可根据环的大小在以上给出的位置的任一侧改变例如1、2、3、4、5、6、7、8、9或甚至10个氨基酸,只要断裂位置仍然落入外部环的未结构化区域内。

[0162] 本文所述的AD功能化的CRISPR系统可以用于靶向RNA多核苷酸序列内的特定腺嘌呤以进行脱氨。例如,指导分子可以与CRISPR-Cas蛋白形成复合物,并指导该复合物结合目标RNA多核苷酸中的靶RNA序列。因为指导序列被设计成具有非配对C,所以在指导序列与靶序列之间形成的RNA双链体包含A-C错配,该错配指导腺苷脱氨酶接触与非配对C相对的A并使脱氨,从而将其转化为肌苷(I)。由于肌苷(I)碱基在细胞过程中与C配对并具有类似G的功能,因此本文所述的A的靶向脱氨可用于校正不需要的G-A和C-T突变,以及获得所需的A-G和T-C突变。

[0163] 在一些实施方案中，AD功能化的CRISPR系统用于体外RNA多核苷酸分子中的靶向脱氨。在一些实施方案中，AD功能化的CRISPR系统用于细胞内DNA分子中的靶向脱氨。细胞可以是真核细胞，诸如动物细胞、哺乳动物细胞、人类细胞或植物细胞。

[0164] 本发明还涉及一种用于通过使用AD功能化的CRISPR系统进行靶向脱氨来治疗或预防疾病的方法，其中使A脱氨可以治疗由含有致病性G→A或C→T点突变的转录物引起的疾病。可以用本发明治疗或预防的疾病的实例包括癌症、梅尔-戈林综合征、塞克尔综合征4、朱伯特综合征5、Leber先天性黑蒙症10;2型夏科-马利-杜斯氏病;2型夏科-马利-杜斯氏病;2C型乌舍尔综合征;脊髓小脑失调症28;脊髓小脑失调症28;脊髓小脑失调症28;长QT综合征2;舍格伦-拉尔森综合征;遗传性果糖尿症;遗传性果糖尿症;神经母细胞瘤;神经母细胞瘤;卡尔曼综合征1;卡尔曼综合征1;卡尔曼综合征1;异染性脑白质营养不良。

[0165] 因此，在特定的实施方案中，本发明包括用于疗法的组合物。这意味着所述方法可以在体内、离体或体外进行。在特定的实施方案中，所述方法不是治疗动物或人体的方法，也不是用于修饰人类细胞的种系遗传特性的方法。在特定的实施方案中;当实施所述方法时，靶RNA不包含在人类或动物细胞内。在特定的实施方案中，当靶标是人类或动物靶标时，所述方法离体或体外进行。

[0166] 本发明还涉及一种用于敲除或敲低不需要的基因活性的方法，其中使基因的转录物处的A脱氨导致功能丧失。例如，在一个实施方案中，通过AD功能化的CRISPR系统进行靶向脱氨可以引起无义突变，从而导致在内源基因中产生提前终止密码子。这可能会改变内源基因的表达，并且可能使得经编辑的细胞具有所需性状。在另一个实施方案中，通过AD功能化的CRISPR系统进行靶向脱氨可以引起非保守的错义突变，从而导致在内源基因中产生不同氨基酸残基的密码子。这可能会改变所表达的内源基因的功能，并且还可能使得编辑后的细胞具有所需性状。

[0167] 本发明还涉及一种通过使用AD功能化的CRISPR系统进行靶向脱氨而获得的经修饰的细胞或其后代，其中与靶向脱氨之前的对应细胞相比，所述经修饰的细胞包含I或G代替目标靶RNA序列中的A。经修饰的细胞可以是真核细胞，诸如动物细胞、植物细胞、哺乳动物细胞或人类细胞。

[0168] 在一些实施方案中，经修饰的细胞是治疗性T细胞，诸如适合CAR-T疗法的T细胞。修饰可在治疗性T细胞中产生一种或多种期望的性状，包括但不限于免疫检查点受体(例如PDA、CTLA4)的表达降低、HLA蛋白(例如B2M、HLA-A)的表达降低以及内源TCR的表达降低。

[0169] 在一些实施方案中，经修饰的细胞是产生抗体的B细胞。修饰可在B细胞中产生一种或多种期望的性状，包括但不限于抗体产生增强。

[0170] 本发明还涉及一种经修饰的非人类动物或经修饰的植物。经修饰的非人类动物可以是农场动物。经修饰的植物可以是农作物。

[0171] 本发明还涉及一种用于细胞疗法的方法，所述方法包括向有需要的患者施用本文所述的经修饰的细胞，其中经修饰的细胞的存在可以治疗患者的疾病。在一个实施方案中，用于细胞疗法的经修饰的细胞是能够识别和/或攻击肿瘤细胞的CAR-T细胞。在另一个实施方案中，用于细胞疗法的经修饰的细胞是干细胞，诸如神经干细胞、间充质干细胞、造血干细胞或iPSC细胞。

[0172] 本发明还涉及一种适用于修饰目标靶基因座中的腺嘌呤的工程化的非天然存在

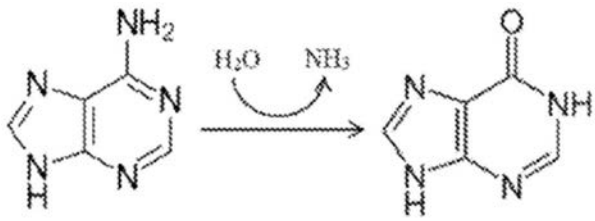
的系统，所述系统包含：指导分子，所述指导分子包含指导序列或编码指导分子的核苷酸序列；CRISPR-Cas蛋白或编码CRISPR-Cas蛋白的一个或多个核苷酸序列；腺苷脱氨酶蛋白或其催化结构域，或编码其的一个或多个核苷酸序列；其中腺苷脱氨酶蛋白或其催化结构域与CRISPR-Cas蛋白或指导分子共价或非共价连接，或者适用于在递送后与其连接；其中指导序列能够与目标RNA多核苷酸内包含腺嘌呤的靶序列杂交，但是在对应于腺嘌呤的位置包含胞嘧啶。

[0173] 本发明还涉及一种适用于修饰目标靶基因座中的腺嘌呤的工程化的非天然存在的载体系统，所述载体系统包含一个或多个载体，所述一个或多个载体包含：第一调控元件，其可操作地连接到编码包含指导序列的指导分子的一个或多个核苷酸序列；第二调控元件，其可操作地连接到编码CRISPR-Cas蛋白的核苷酸序列；以及任选的编码腺苷脱氨酶蛋白或其催化结构域的核苷酸序列，其受第一调控元件或第二调控元件的控制或可操作地连接到第三调控元件；其中如果编码腺苷脱氨酶蛋白或其催化结构域的核苷酸序列可操作地连接到第三调控元件，则腺苷脱氨酶蛋白或其催化结构域适用于在表达后连接到指导分子或Crispr-Cas蛋白；其中指导序列能够与靶基因座内包含腺嘌呤的靶序列杂交，但是在对应于腺嘌呤的位置包含胞嘧啶。其中组分(a)、(b)和(c)位于系统的相同或不同载体上。

[0174] 本发明还涉及一种体外、离体或体内宿主细胞或细胞系或其后代，所述宿主细胞或细胞系或其后代包含本文所述的工程化的非天然存在的系统或载体系统。宿主细胞可以是真核细胞，诸如动物细胞、植物细胞、哺乳动物细胞或人类细胞。

[0175] 腺苷脱氨酶

[0176] 本文所用的术语“腺苷脱氨酶”或“腺苷脱氨酶蛋白”是指如下所示能够催化将腺嘌呤(或分子的腺嘌呤部分)转化为次黄嘌呤(或分子的次黄嘌呤部分)的水解脱氨反应的蛋白质、多肽或者蛋白质或多肽的一个或多个功能结构域。在一些实施方案中，含腺嘌呤的分子是腺苷(A)，而含次黄嘌呤的分子是肌苷(I)。含腺嘌呤的分子可以是脱氧核糖核酸(DNA)或核糖核酸(RNA)。

[0177] 


[0178] 根据本公开，可以与本公开结合使用的腺苷脱氨酶包括但不限于被称为作用于RNA的腺苷脱氨酶的酶家族成员(ADAR)、被称为作用于tRNA的腺苷脱氨酶的酶家族成员(ADAT)，以及其他含腺苷脱氨酶结构域的家族成员(ADAD)。根据本公开，腺苷脱氨酶能够靶向RNA/DNA和RNA双链体中的腺嘌呤。实际上，Zheng等人(Nucleic Acids Res.2017,45(6):3369-3377)证明ADAR可以对RNA/DNA和RNA/RNA双链体进行腺嘌呤到肌苷的编辑反应。在特定的实施方案中，对腺苷脱氨酶进行了修饰以增加其编辑RNA/DNAn RNA双链体中的DNA的能力，如下文详述。

[0179] 在一些实施方案中，腺苷脱氨酶来源于一种或多种后生动物物种，包括但不限于哺乳动物、鸟类、青蛙、鱿鱼、鱼、蝇和蠕虫。在一些实施方案中，腺苷脱氨酶是人类、鱿鱼或

果蝇腺苷脱氨酶。

[0180] 在一些实施方案中，腺苷脱氨酶是人类ADAR，包括hADAR1、hADAR2、hADAR3。在一些实施方案中，腺苷脱氨酶是秀丽隐杆线虫ADAR蛋白，包括ADR-1和ADR-2。在一些实施方案中，腺苷脱氨酶是果蝇ADAR蛋白，包括dAdar。在一些实施方案中，腺苷脱氨酶是长鳍鱿鱼(Loligo pealeii)ADAR蛋白，包括sqADAR2a和sqADAR2b。在一些实施方案中，腺苷脱氨酶是人类ADAT蛋白。在一些实施方案中，腺苷脱氨酶是果蝇ADAT蛋白。在一些实施方案中，腺苷脱氨酶是人类ADAD蛋白，包括TENR(hADAD1)和TENRL(hADAD2)。

[0181] 在一些实施方案中，腺苷脱氨酶蛋白识别双链核酸底物中的一个或多个靶腺苷残基并将其转化为肌苷残基。在一些实施方案中，双链核酸底物是RNA-DNA杂合双链体。在一些实施方案中，腺苷脱氨酶蛋白识别双链底物上的结合窗口。在一些实施方案中，结合窗口含有至少一个靶腺苷残基。在一些实施方案中，结合窗口在约3bp至约100bp的范围内。在一些实施方案中，结合窗口在约5bp至约50bp的范围内。在一些实施方案中，结合窗口在约10bp至约30bp的范围内。在一些实施方案中，结合窗口为约1bp、2bp、3bp、5bp、7bp、10bp、15bp、20bp、25bp、30bp、40bp、45bp、50bp、55bp、60bp、65bp、70bp、75bp、80bp、85bp、90bp、95bp或100bp。

[0182] 在一些实施方案中，腺苷脱氨酶蛋白包含一个或多个脱氨酶结构域。不受理论的束缚，考虑了脱氨酶结构域起识别双链核酸底物中包含的一个或多个靶腺苷(A)残基并将其转化为肌苷(I)残基的作用。在一些实施方案中，脱氨酶结构域包含活性中心。在一些实施方案中，活性中心包含锌离子。在一些实施方案中，在A到I编辑过程中，靶腺苷残基处的碱基配对被破坏，并且靶腺苷残基被“翻转”出双螺旋，变得可被腺苷脱氨酶接近。在一些实施方案中，活性中心内或附近的氨基酸残基与靶腺苷残基5’的一个或多个核苷酸相互作用。在一些实施方案中，活性中心内或附近的氨基酸残基与靶腺苷残基3’的一个或多个核苷酸相互作用。在一些实施方案中，活性中心内或附近的氨基酸残基还与相反链上与靶腺苷残基互补的核苷酸相互作用。在一些实施方案中，氨基酸残基与核苷酸的2’羟基形成氢键。

[0183] 在一些实施方案中，腺苷脱氨酶包含人类ADAR2全蛋白(hADAR2)或其脱氨酶结构域(hADAR2-D)。在一些实施方案中，腺苷脱氨酶是与hADAR2或hADAR2-D同源的ADAR家族成员。

[0184] 具体地，在一些实施方案中，同源ADAR蛋白是人类ADAR1(hADAR1)或其脱氨酶结构域(hADAR1-D)。在一些实施方案中，hADAR1-D的甘氨酸1007对应于甘氨酸487hADAR2-D，并且hADAR1-D的谷氨酸1008对应于hADAR2-D的谷氨酸488。

[0185] 在一些实施方案中，腺苷脱氨酶包含hADAR2-D的野生型氨基酸序列。在一些实施方案中，腺苷脱氨酶在hADAR2-D序列中包含一个或多个突变，使得根据特定需要改变编辑效率和/或hADAR2-D的底物编辑偏好。

[0186] hADAR1和hADAR2蛋白的某些突变在以下文献中描述：Kuttan等人，Proc Natl Acad Sci U S A.(2012)109(48):E3295-304；Want等人，ACS Chem Biol.(2015)10(11):2512-9；以及Zheng等人，Nucleic Acids Res.(2017)45(6):3369-337，这些文献中的每一项全文通过引用并入本文。

[0187] 在一些实施方案中，腺苷脱氨酶在hADAR2-D氨基酸序列的甘氨酸336处或在同源

ADAR蛋白中的对应位置处包含突变。在一些实施方案中，位置336处的甘氨酸残基被天冬氨酸残基替换(G336D)。

[0188] 在一些实施方案中，腺苷脱氨酶在hADAR2-D氨基酸序列的甘氨酸487处或在同源ADAR蛋白中的对应位置处包含突变。在一些实施方案中，位置487处的甘氨酸残基被具有相对小的侧链的非极性氨基酸残基替换。例如，在一些实施方案中，位置487处的甘氨酸残基被丙氨酸残基替换(G487A)。在一些实施方案中，位置487处的甘氨酸残基被缬氨酸残基替换(G487V)。在一些实施方案中，位置487处的甘氨酸残基被具有相对大的侧链的氨基酸残基替换。在一些实施方案中，位置487处的甘氨酸残基被精氨酸残基替换(G487R)。在一些实施方案中，位置487处的甘氨酸残基被赖氨酸残基替换(G487K)。在一些实施方案中，位置487处的甘氨酸残基被色氨酸残基替换(G487W)。在一些实施方案中，位置487处的甘氨酸残基被酪氨酸残基替换(G487Y)。

[0189] 在一些实施方案中，腺苷脱氨酶在hADAR2-D氨基酸序列的谷氨酸488处或在同源ADAR蛋白中的对应位置处包含突变。在一些实施方案中，位置488处的谷氨酸残基被谷氨酰胺残基替换(E488Q)。在一些实施方案中，位置488处的谷氨酸残基被组氨酸残基替换(E488H)。在一些实施方案中，位置488处的谷氨酸残基被精氨酸残基替换(E488R)。在一些实施方案中，位置488处的谷氨酸残基被赖氨酸残基替换(E488K)。在一些实施方案中，位置488处的谷氨酸残基被天冬酰胺残基替换(E488N)。在一些实施方案中，位置488处的谷氨酸残基被丙氨酸残基替换(E488A)。在一些实施方案中，位置488处的谷氨酸残基被蛋氨酸残基替换(E488M)。在一些实施方案中，位置488处的谷氨酸残基被丝氨酸残基替换(E488S)。在一些实施方案中，位置488处的谷氨酸残基被苯丙氨酸残基替换(E488F)。在一些实施方案中，位置488处的谷氨酸残基被赖氨酸残基替换(E488L)。在一些实施方案中，位置488处的谷氨酸残基被色氨酸残基替换(E488W)。

[0190] 在一些实施方案中，腺苷脱氨酶在hADAR2-D氨基酸序列的苏氨酸490处或在同源ADAR蛋白中的对应位置处包含突变。在一些实施方案中，位置490处的苏氨酸残基被半胱氨酸残基替换(T490C)。.在一些实施方案中，位置490处的苏氨酸残基被丝氨酸残基替换(T490S)。在一些实施方案中，位置490处的苏氨酸残基被丙氨酸残基替换(T490A)。在一些实施方案中，位置490处的苏氨酸残基被苯丙氨酸残基替换(T490F)。在一些实施方案中，位置490处的苏氨酸残基被酪氨酸残基替换(T490Y)。在一些实施方案中，位置490处的苏氨酸残基被丝氨酸残基替换(T490R)。在一些实施方案中，位置490处的苏氨酸残基被丙氨酸残基替换(T490K)。在一些实施方案中，位置490处的苏氨酸残基被苯丙氨酸残基替换(T490P)。在一些实施方案中，位置490处的苏氨酸残基被酪氨酸残基替换(T490E)。

[0191] 在一些实施方案中，腺苷脱氨酶在hADAR2-D氨基酸序列的缬氨酸493处或在同源ADAR蛋白中的对应位置处包含突变。在一些实施方案中，位置493处的缬氨酸残基被丙氨酸残基替换(V493A)。在一些实施方案中，位置493处的缬氨酸残基被丝氨酸残基替换(V493S)。在一些实施方案中，位置493处的缬氨酸残基被苏氨酸残基替换(V493T)。在一些实施方案中，位置493处的缬氨酸残基被精氨酸残基替换(V493R)。在一些实施方案中，位置493处的缬氨酸残基被天冬氨酸残基替换(V493D)。在一些实施方案中，位置493处的缬氨酸残基被脯氨酸残基替换(V493P)。在一些实施方案中，位置493处的缬氨酸残基被甘氨酸残基替换(V493G)。

[0192] 在一些实施方案中，腺苷脱氨酶在hADAR2-D氨基酸序列的丙氨酸589处或在同源ADAR蛋白中的对应位置处包含突变。在一些实施方案中，位置589处的丙氨酸残基被缬氨酸残基替换(A589V)。

[0193] 在一些实施方案中，腺苷脱氨酶在hADAR2-D氨基酸序列的天冬酰胺597处或在同源ADAR蛋白中的对应位置处包含突变。在一些实施方案中，位置597处的天冬酰胺残基被赖氨酸残基替换(N597K)。在一些实施方案中，腺苷脱氨酶在氨基酸序列的位置597处包含突变，该位置在野生型序列中具有天冬酰胺残基。在一些实施方案中，位置597处的天冬酰胺残基被精氨酸残基替换(N597R)。在一些实施方案中，腺苷脱氨酶在氨基酸序列的位置597处包含突变，该位置在野生型序列中具有天冬酰胺残基。在一些实施方案中，位置597处的天冬酰胺残基被丙氨酸残基替换(N597A)。在一些实施方案中，腺苷脱氨酶在氨基酸序列的位置597处包含突变，该位置在野生型序列中具有天冬酰胺残基。在一些实施方案中，位置597处的天冬酰胺残基被谷氨酸残基替换(N597E)。在一些实施方案中，腺苷脱氨酶在氨基酸序列的位置597处包含突变，该位置在野生型序列中具有天冬酰胺残基。在一些实施方案中，位置597处的天冬酰胺残基被组氨酸残基替换(N597H)。在一些实施方案中，腺苷脱氨酶在氨基酸序列的位置597处包含突变，该位置在野生型序列中具有天冬酰胺残基。在一些实施方案中，位置597处的天冬酰胺残基被甘氨酸残基替换(N597G)。在一些实施方案中，腺苷脱氨酶在氨基酸序列的位置597处包含突变，该位置在野生型序列中具有天冬酰胺残基。在一些实施方案中，位置597处的天冬酰胺残基被酪氨酸残基替换(N597Y)。在一些实施方案中，位置597处的天冬酰胺残基被苯丙氨酸残基替换(N597F)。

[0194] 在一些实施方案中，腺苷脱氨酶在hADAR2-D氨基酸序列的丝氨酸599处或在同源ADAR蛋白中的对应位置处包含突变。在一些实施方案中，位置599处的丝氨酸残基被苏氨酸残基替换(S599T)。

[0195] 在一些实施方案中，腺苷脱氨酶在hADAR2-D氨基酸序列的天冬酰胺613处或在同源ADAR蛋白中的对应位置处包含突变。在一些实施方案中，位置613处的天冬酰胺残基被赖氨酸残基替换(N613K)。在一些实施方案中，腺苷脱氨酶在氨基酸序列的位置613处包含突变，该位置在野生型序列中具有天冬酰胺残基。在一些实施方案中，位置613处的天冬酰胺残基被精氨酸残基替换(N613R)。在一些实施方案中，腺苷脱氨酶在氨基酸序列的位置613处包含突变，该位置在野生型序列中具有天冬酰胺残基。在一些实施方案中，位置613处的天冬酰胺残基被丙氨酸残基替换(N613A)。在一些实施方案中，腺苷脱氨酶在氨基酸序列的位置613处包含突变，该位置在野生型序列中具有天冬酰胺残基。在一些实施方案中，位置613处的天冬酰胺残基被谷氨酸残基替换(N613E)。

[0196] 在一些实施方案中，为了提高编辑效率，腺苷脱氨酶可包含基于hADAR2-D的氨基酸序列位置的一个或多个以下突变:G336D、G487A、G487V、E488Q、E488H、E488R、E488N、E488A、E488S、E488M、T490C、T490S、V493T、V493S、V493A、V493R、V493D、V493P、V493G、N597K、N597R、N597A、N597E、N597H、N597G、N597Y、A589V、S599T、N613K、N613R、N613A、N613E，以及同源ADAR蛋白中与上述突变相对应的突变。

[0197] 在一些实施方案中，为了降低编辑效率，腺苷脱氨酶可包含基于hADAR2-D的氨基酸序列位置的一个或多个以下突变:E488F、E488L、E488W、T490A、T490F、T490Y、T490R、T490K、T490P、T490E、N597F，以及同源ADAR蛋白中与上述突变相对应的突变。在特定的实施

方案中，使用效率降低的腺苷脱氨酶来减少脱靶效应可能是有意义的。

[0198] 在一些实施方案中，为了减少脱靶效应，腺苷脱氨酶包含基于hADAR2-D的氨基酸序列位置的以下位置处的一个或多个突变：R348、V351、T375、K376、E396、C451、R455、N473、R474、K475、R477、R481、S486、E488、T490、S495、R510，以及同源ADAR蛋白中与上述突变相对应的突变。在一些实施方案中，腺苷脱氨酶在E488处以及选自R348、V351、T375、K376、E396、C451、R455、N473、R474、K475、R477、R481、S486、T490、S495、R510的一个或多个其他位置处包含突变。在一些实施方案中，腺苷脱氨酶在T375处以及任选地在一个或多个其他位置处包含突变。在一些实施方案中，腺苷脱氨酶在N473处以及任选地在一个或多个其他位置处包含突变。在一些实施方案中，腺苷脱氨酶在V351处以及任选地在一个或多个其他位置处包含突变。在一些实施方案中，腺苷脱氨酶在E488和T375处以及任选地在一个或多个其他位置处包含突变。在一些实施方案中，腺苷脱氨酶在E488和N473处以及任选地在一个或多个其他位置处包含突变。在一些实施方案中，腺苷脱氨酶在E488和V351处以及任选地在一个或多个其他位置处包含突变。在一些实施方案中，腺苷脱氨酶在E488处以及T375、N473和V351中的一个或多个处包含突变。

[0199] 在一些实施方案中，为了减少脱靶效应，腺苷脱氨酶包含基于hADAR2-D的氨基酸序列位置的选自以下的一个或多个突变：R348E、V351L、T375G、T375S、R455G、R455S、R455E、N473D、R474E、K475Q、R477E、R481E、S486T、E488Q、T490A、T490S、S495T和R510E，以及同源ADAR蛋白中与上述突变相对应的突变。在一些实施方案中，腺苷脱氨酶包含突变E488Q以及选自R348E、V351L、T375G、T375S、R455G、R455S、R455E、N473D、R474E、K475Q、R477E、R481E、S486T、T490A、T490S、S495T和R510E的一个或多个其他突变。在一些实施方案中，腺苷脱氨酶包含突变T375G或T375S，以及任选地一个或多个其他突变。在一些实施方案中，腺苷脱氨酶包含突变N473D，以及任选地一个或多个其他突变。在一些实施方案中，腺苷脱氨酶包含突变V351L，以及任选地一个或多个其他突变。在一些实施方案中，腺苷脱氨酶包含突变E488Q和T375G或T375G，以及任选地一个或多个其他突变。在一些实施方案中，腺苷脱氨酶包含突变E488Q和N473D，以及任选地一个或多个其他突变。在一些实施方案中，腺苷脱氨酶包含突变E488Q和V351L，以及任选地一个或多个其他突变。在一些实施方案中，腺苷脱氨酶包含突变E488Q以及T375G/S、N473D和V351L中的一个或多个。

[0200] 与双链体RNA结合的人ADAR2脱氨酶结构域的晶体结构揭示了在修饰位点5'侧与RNA结合的蛋白环。该5'结合环是ADAR家族成员之间存在底物特异性差异的一个原因。参见Wang等人，Nucleic Acids Res.，44(20)：9872-9880(2016)，该文献的内容全文通过引用并入本文。另外，在酶活性位点附近鉴定出ADAR2特异性RNA结合环。参见Mathews等人，Nat.Struct.Mol.Biol.，23(5)：426-33(2016)，该文献的内容全文通过引用并入本文。在一些实施方案中，腺苷脱氨酶在RNA结合环中包含一个或多个突变以改善编辑特异性和/或效率。

[0201] 在一些实施方案中，腺苷脱氨酶在hADAR2-D氨基酸序列的丙氨酸454处或在同源ADAR蛋白中的对应位置处包含突变。在一些实施方案中，位置454处的丙氨酸残基被丝氨酸残基替换(A454S)。在一些实施方案中，位置454处的丙氨酸残基被半胱氨酸残基替换(A454C)。在一些实施方案中，位置454处的丙氨酸残基被天冬氨酸残基替换(A454D)。

[0202] 在一些实施方案中，腺苷脱氨酶在hADAR2-D氨基酸序列的精氨酸455处或在同源

ADAR蛋白中的对应位置处包含突变。在一些实施方案中，位置455处的精氨酸残基被丙氨酸残基替换(R455A)。在一些实施方案中，位置455处的精氨酸残基被缬氨酸残基替换(R455V)。在一些实施方案中，位置455处的精氨酸残基被组氨酸残基替换(R455H)。在一些实施方案中，位置455处的精氨酸残基被甘氨酸残基替换(R455G)。在一些实施方案中，位置455处的精氨酸残基被丝氨酸残基替换(R455S)。在一些实施方案中，位置455处的精氨酸残基被谷氨酸残基替换(R455E)。

[0203] 在一些实施方案中，腺苷脱氨酶在hADAR2-D氨基酸序列的异亮氨酸456处或在同源ADAR蛋白中的对应位置处包含突变。在一些实施方案中，位置456处的异亮氨酸残基被缬氨酸残基替换(I456V)。在一些实施方案中，位置456处的异亮氨酸残基被亮氨酸残基替换(I456L)。在一些实施方案中，位置456处的异亮氨酸残基被天冬氨酸残基替换(I456D)。

[0204] 在一些实施方案中，腺苷脱氨酶在hADAR2-D氨基酸序列的苯丙氨酸457处或在同源ADAR蛋白中的对应位置处包含突变。在一些实施方案中，位置457处的苯丙氨酸残基被酪氨酸残基替换(F457Y)。在一些实施方案中，位置457处的苯丙氨酸残基被精氨酸残基替换(F457R)。在一些实施方案中，位置457处的苯丙氨酸残基被谷氨酸残基替换(F457E)。

[0205] 在一些实施方案中，腺苷脱氨酶在hADAR2-D氨基酸序列的丝氨酸458处或在同源ADAR蛋白中的对应位置处包含突变。在一些实施方案中，位置458处的丝氨酸残基被缬氨酸残基替换(S458V)。在一些实施方案中，位置458处的丝氨酸残基被苯丙氨酸残基替换(S458F)。在一些实施方案中，位置458处的丝氨酸残基被脯氨酸残基替换(S458P)。

[0206] 在一些实施方案中，腺苷脱氨酶在hADAR2-D氨基酸序列的脯氨酸459处或在同源ADAR蛋白中的对应位置处包含突变。在一些实施方案中，位置459处的脯氨酸残基被半胱氨酸残基替换(P459C)。在一些实施方案中，位置459处的脯氨酸残基被组氨酸残基替换(P459H)。在一些实施方案中，位置459处的脯氨酸残基被色氨酸残基替换(P459W)。

[0207] 在一些实施方案中，腺苷脱氨酶在hADAR2-D氨基酸序列的组氨酸460处或在同源ADAR蛋白中的对应位置处包含突变。在一些实施方案中，位置460处的组氨酸残基被精氨酸残基替换(H460R)。在一些实施方案中，位置460处的组氨酸残基被异亮氨酸残基替换(H460I)。在一些实施方案中，位置460处的组氨酸残基被脯氨酸残基替换(H460P)。

[0208] 在一些实施方案中，腺苷脱氨酶在hADAR2-D氨基酸序列的脯氨酸462处或在同源ADAR蛋白中的对应位置处包含突变。在一些实施方案中，位置462处的脯氨酸残基被丝氨酸残基替换(P462S)。在一些实施方案中，位置462处的脯氨酸残基被色氨酸残基替换(P462W)。在一些实施方案中，位置462处的脯氨酸残基被谷氨酸残基替换(P462E)。

[0209] 在一些实施方案中，腺苷脱氨酶在hADAR2-D氨基酸序列的天冬氨酸469处或在同源ADAR蛋白中的对应位置处包含突变。在一些实施方案中，位置469处的天冬氨酸残基被谷氨酰胺残基替换(D469Q)。在一些实施方案中，位置469处的天冬氨酸残基被丝氨酸残基替换(D469S)。在一些实施方案中，位置469处的天冬氨酸残基被丝氨酸残基替换(D469Y)。

[0210] 在一些实施方案中，腺苷脱氨酶在hADAR2-D氨基酸序列的精氨酸470处或在同源ADAR蛋白中的对应位置处包含突变。在一些实施方案中，位置470处的精氨酸残基被丙氨酸残基替换(R470A)。在一些实施方案中，位置470处的精氨酸残基被异亮氨酸残基替换(R470I)。在一些实施方案中，位置470处的精氨酸残基被天冬氨酸残基替换(R470D)。

[0211] 在一些实施方案中，腺苷脱氨酶在hADAR2-D氨基酸序列的组氨酸471处或在同源

ADAR蛋白中的对应位置处包含突变。在一些实施方案中，位置471处的组氨酸残基被赖氨酸残基替换(H471K)。在一些实施方案中，位置471处的组氨酸残基被苏氨酸残基替换(H471T)。在一些实施方案中，位置471处的组氨酸残基被缬氨酸残基替换(H471V)。

[0212] 在一些实施方案中，腺苷脱氨酶在hADAR2-D氨基酸序列的脯氨酸472处或在同源ADAR蛋白中的对应位置处包含突变。在一些实施方案中，位置472处的脯氨酸残基被赖氨酸残基替换(P472K)。在一些实施方案中，位置472处的脯氨酸残基被苏氨酸残基替换(P472T)。在一些实施方案中，位置472处的脯氨酸残基被天冬氨酸残基替换(P472D)。

[0213] 在一些实施方案中，腺苷脱氨酶在hADAR2-D氨基酸序列的天冬酰胺473处或在同源ADAR蛋白中的对应位置处包含突变。在一些实施方案中，位置473处的天冬酰胺残基被精氨酸残基替换(N473R)。在一些实施方案中，位置473处的天冬酰胺残基被色氨酸残基替换(N473W)。在一些实施方案中，位置473处的天冬酰胺残基被脯氨酸残基替换(N473P)。在一些实施方案中，位置473处的天冬酰胺残基被天冬氨酸残基替换(N473D)。

[0214] 在一些实施方案中，腺苷脱氨酶在hADAR2-D氨基酸序列的精氨酸474处或在同源ADAR蛋白中的对应位置处包含突变。在一些实施方案中，位置474处的精氨酸残基被赖氨酸残基替换(R474K)。在一些实施方案中，位置474处的精氨酸残基被甘氨酸残基替换(R474G)。在一些实施方案中，位置474处的精氨酸残基被天冬氨酸残基替换(R474D)。在一些实施方案中，位置474处的精氨酸残基被谷氨酸残基替换(R474E)。

[0215] 在一些实施方案中，腺苷脱氨酶在hADAR2-D氨基酸序列的赖氨酸475处或在同源ADAR蛋白中的对应位置处包含突变。在一些实施方案中，位置475处的赖氨酸残基被谷氨酰胺残基替换(K475Q)。在一些实施方案中，位置475处的赖氨酸残基被天冬酰胺残基替换(K475N)。在一些实施方案中，位置475处的赖氨酸残基被天冬氨酸残基替换(K475D)。

[0216] 在一些实施方案中，腺苷脱氨酶在hADAR2-D氨基酸序列的丙氨酸476处或在同源ADAR蛋白中的对应位置处包含突变。在一些实施方案中，位置476处的丙氨酸残基被丝氨酸残基替换(A476S)。在一些实施方案中，位置476处的丙氨酸残基被精氨酸残基替换(A476R)。在一些实施方案中，位置476处的丙氨酸残基被谷氨酸残基替换(A476E)。

[0217] 在一些实施方案中，腺苷脱氨酶在hADAR2-D氨基酸序列的精氨酸477处或在同源ADAR蛋白中的对应位置处包含突变。在一些实施方案中，位置477处的精氨酸残基被赖氨酸残基替换(R477K)。在一些实施方案中，位置477处的精氨酸残基被苏氨酸残基替换(R477T)。在一些实施方案中，位置477处的精氨酸残基被苯丙氨酸残基替换(R477F)。在一些实施方案中，位置474处的精氨酸残基被谷氨酸残基替换(R477E)。

[0218] 在一些实施方案中，腺苷脱氨酶在hADAR2-D氨基酸序列的甘氨酸478处或在同源ADAR蛋白中的对应位置处包含突变。在一些实施方案中，位置478处的甘氨酸残基被丙氨酸残基替换(G478A)。在一些实施方案中，位置478处的甘氨酸残基被精氨酸残基替换(G478R)。在一些实施方案中，位置478处的甘氨酸残基被酪氨酸残基替换(G478Y)。

[0219] 在一些实施方案中，腺苷脱氨酶在hADAR2-D氨基酸序列的谷氨酰胺479处或在同源ADAR蛋白中的对应位置处包含突变。在一些实施方案中，位置479处的谷氨酰胺残基被天冬酰胺残基替换(Q479N)。在一些实施方案中，位置479处的谷氨酰胺残基被丝氨酸残基替换(Q479S)。在一些实施方案中，位置479处的谷氨酰胺残基被脯氨酸残基替换(Q479P)。

[0220] 在一些实施方案中，腺苷脱氨酶在hADAR2-D氨基酸序列的精氨酸348处或在同源

ADAR蛋白中的对应位置处包含突变。在一些实施方案中，位置348处的精氨酸残基被丙氨酸残基替换(R348A)。在一些实施方案中，位置348处的精氨酸残基被谷氨酸残基替换(R348E)。

[0221] 在一些实施方案中，腺苷脱氨酶在hADAR2-D氨基酸序列的缬氨酸351处或在同源ADAR蛋白中的对应位置处包含突变。在一些实施方案中，位置351处的缬氨酸残基被亮氨酸残基替换(V351L)。

[0222] 在一些实施方案中，腺苷脱氨酶在hADAR2-D氨基酸序列的苏氨酸375处或在同源ADAR蛋白中的对应位置处包含突变。在一些实施方案中，位置375处的苏氨酸残基被甘氨酸残基替换(T375G)。在一些实施方案中，位置375处的苏氨酸残基被丝氨酸残基替换(T375S)。

[0223] 在一些实施方案中，腺苷脱氨酶在hADAR2-D氨基酸序列的精氨酸481处或在同源ADAR蛋白中的对应位置处包含突变。在一些实施方案中，位置481处的精氨酸残基被谷氨酸残基替换(R481E)。

[0224] 在一些实施方案中，腺苷脱氨酶在hADAR2-D氨基酸序列的丝氨酸486处或在同源ADAR蛋白中的对应位置处包含突变。在一些实施方案中，位置486处的丝氨酸残基被苏氨酸残基替换(S486T)。

[0225] 在一些实施方案中，腺苷脱氨酶在hADAR2-D氨基酸序列的苏氨酸490处或在同源ADAR蛋白中的对应位置处包含突变。在一些实施方案中，位置490处的苏氨酸残基被丙氨酸残基替换(T490A)。在一些实施方案中，位置490处的苏氨酸残基被丝氨酸残基替换(T490S)。

[0226] 在一些实施方案中，腺苷脱氨酶在hADAR2-D氨基酸序列的丝氨酸495处或在同源ADAR蛋白中的对应位置处包含突变。在一些实施方案中，位置495处的丝氨酸残基被苏氨酸残基替换(S495T)。

[0227] 在一些实施方案中，腺苷脱氨酶在hADAR2-D氨基酸序列的精氨酸510处或在同源ADAR蛋白中的对应位置处包含突变。在一些实施方案中，位置510处的精氨酸残基被谷氨酰胺残基替换(R510Q)。在一些实施方案中，位置510处的精氨酸残基被丙氨酸残基替换(R510A)。在一些实施方案中，位置510处的精氨酸残基被谷氨酸残基替换(R510E)。

[0228] 在一些实施方案中，腺苷脱氨酶在hADAR2-D氨基酸序列的甘氨酸593处或在同源ADAR蛋白中的对应位置处包含突变。在一些实施方案中，位置593处的甘氨酸残基被丙氨酸残基替换(G593A)。在一些实施方案中，位置593处的甘氨酸残基被谷氨酸残基替换(G593E)。

[0229] 在一些实施方案中，腺苷脱氨酶在hADAR2-D氨基酸序列的赖氨酸594处或在同源ADAR蛋白中的对应位置处包含突变。在一些实施方案中，位置594处的赖氨酸残基被丙氨酸残基替换(K594A)。

[0230] 在一些实施方案中，腺苷脱氨酶在hADAR2-D氨基酸序列的位置A454、R455、I456、F457、S458、P459、H460、P462、D469、R470、H471、P472、N473、R474、K475、A476、R477、G478、Q479、R348、R510、G593、K594中的任何一个或多个处或在同源ADAR蛋白中的对应位置处包含突变。

[0231] 在一些实施方案中，腺苷脱氨酶包含hADAR2-D氨基酸序列的突变A454S、A454C、

A454D、R455A、R455V、R455H、I456V、I456L、I456D、F457Y、F457R、F457E、S458V、S458F、S458P、P459C、P459H、P459W、H460R、H460I、H460P、P462S、P462W、P462E、D469Q、D469S、D469Y、R470A、R470I、R470D、H471K、H471T、H471V、P472K、P472T、P472D、N473R、N473W、N473P、R474K、R474G、R474D、K475Q、K475N、K475D、A476S、A476R、A476E、R477K、R477T、R477F、G478A、G478R、G478Y、Q479N、Q479S、Q479P、R348A、R510Q、R510A、G593A、G593E、K594A中的任何一个或多个或同源ADAR蛋白中的对应位置。

[0232] 在某些实施方案中，使腺苷脱氨酶突变以将活性转化为胞苷脱氨酶。因此，在一些实施方案中，腺苷脱氨酶在选自E396、C451、V351、R455、T375、K376、S486、Q488、R510、K594、R348、G593、S397、H443、L444、Y445、F442、E438、T448、A353、V355、T339、P539、T339、P539、V525、I520、P462和N579的位置中包括一个或多个突变。在特定的实施方案中，腺苷脱氨酶在选自V351、L444、V355、V525和I520的位置中包含一个或多个突变。在一些实施方案中，腺苷脱氨酶可基于hADAR2-D的氨基酸序列位置在E488、V351、S486、T375、S370、P462、N597处包含一个或多个突变，以及同源ADAR蛋白中与上述突变相对应的突变。在一些实施方案中，腺苷脱氨酶可基于hADAR2-D的氨基酸序列位置包含突变E488Q、V351G、S486A、T375S、S370C、P462A、N597I中的一个或多个，以及同源ADAR蛋白中与上述突变相对应的突变。这些修饰在用于修饰目标靶RNA中的胞嘧啶的方法中是有意义的，所述方法包括向所述靶RNA递送：(a)无催化活性的(死亡)Cas13蛋白；(b)包含与正向重复序列连接的指导序列的指导分子；以及(c)胞嘧啶脱氨酶蛋白或其催化结构域，所述胞嘧啶脱氨酶蛋白是如上文所述经修饰以确保胞嘧啶脱氨酶活性的腺苷脱氨酶，其中所述胞嘧啶脱氨酶蛋白或其催化结构域与所述死亡Cas13蛋白或所述指导分子共价或非共价连接，或者适用于在递送后与其连接；其中指导分子与所述死亡Cas13蛋白形成复合物并指导所述复合物结合所述目标靶RNA序列，其中所述指导序列能够与包含所述胞嘧啶的靶序列杂交以形成RNA双链体，其中所述指导序列在对应于所述胞嘧啶的位置处包含非配对腺嘌呤或尿嘧啶，从而导致在形成的RNA双链体中产生C-A/U错配；其中所述胞嘧啶脱氨酶蛋白或其催化结构域使所述RNA双链体中的所述胞嘧啶脱氨。这些方法在治疗或预防由T→C或A→G点突变或致病性SNP引起的疾病中是有意义的。

[0233] 在某些实施方案中，通过gRNA的化学修饰来实现编辑的改善和脱靶修饰的减少。如Vogel等人(2014)，Angew Chem Int Ed，53:6267-6271，doi:10.1002/anie.201402634(全文通过引用并入本文)中例示的经化学修饰的gRNA降低了脱靶活性并提高了中靶效率。2'-O-甲基和硫代磷酸酯修饰的指导RNA一般可提高细胞中的编辑效率。

[0234] 众所周知，ADAR对经编辑的A的任一侧上的相邻核苷酸表现出偏好(www.nature.com/nsmb/journal/v23/n5/full/nsmb.3203.html，Matthews等人(2017)，Nature Structural Mol Biol，23(5):426-433，其全文通过引用并入本文)。因此，在某些实施方案中，选择gRNA、靶标和/或ADAR针对基序偏好进行优化。

[0235] 已经在体外证实，故意错配可以编辑非优选的基序(https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gku272；Schneider等人(2014)，Nucleic Acid Res，42(10):e87)；Fukuda等人(2017)，Scienticic Reports，7，doi:10.1038/srep41478，这些文献全文通过引用并入本文)。因此，在某些实施方案中，为了提高非优选的5'或3'相邻碱基上的RNA编辑效率，引入了相邻碱基中的有意错配。

[0236] 结果表明,ADAR脱氨酶结构域的靶向窗口中与C相对的A比其他碱基优先被编辑。另外,在靶向碱基的几个碱基内与U配对的A碱基表现出较低的Cas13b-ADAR融合体编辑水平,这表明该酶可以灵活地编辑多个A,参见例如图18。这两个观察结果表明,可以通过错配要用C编辑的所有A来指定Cas13b-ADAR融合体的活性窗口中的多个A进行编辑。因此,在某些实施方案中,活性窗口中的多个A:C错配被设计成形成多个A:I编辑。在某些实施方案中,为了抑制活性窗口中潜在的脱靶编辑,将非靶A与A或G配对。

[0237] 术语“编辑特异性”和“编辑偏好”在本文可互换使用,是指双链底物中特定腺苷位点处的A到I编辑的程度。在一些实施方案中,底物编辑偏好由靶腺苷残基的5’最近邻部和/或3’最近邻部确定。在一些实施方案中,腺苷脱氨酶对底物的5’最近邻部具有偏好,顺序为U>A>C>G(“>”表示偏好更大)。在一些实施方案中,腺苷脱氨酶对底物的3’最近邻部具有偏好,顺序为G>C～A>U(“>”表示偏好更大;“～”表示偏好类似)。在一些实施方案中,腺苷脱氨酶对底物的3’最近邻部具有偏好,顺序为G>C>U～A(“>”表示偏好更大;“～”表示偏好类似)。在一些实施方案中,腺苷脱氨酶对底物的3’最近邻部具有偏好,顺序为G>C>A>U(“>”表示偏好更大)。在一些实施方案中,腺苷脱氨酶对底物的3’最近邻部具有偏好,顺序为C～G～A>U(“>”表示偏好更大;“～”表示偏好类似)。在一些实施方案中,腺苷脱氨酶对含有靶腺苷残基的三联体序列具有偏好,顺序为TAG>AAG>CAC>AAT>GAA>GAC(“>”表示偏好更大),中心A为靶腺苷残基。

[0238] 在一些实施方案中,腺苷脱氨酶的底物编辑偏好受腺苷脱氨酶蛋白中是否存在核酸结合结构域的影响。在一些实施方案中,为了修改底物编辑偏好,将脱氨酶结构域与双链RNA结合结构域(dsRBD)或双链RNA结合基序(dsRBM)连接。在一些实施方案中,dsRBD或dsRBM可来源于ADAR蛋白,诸如hADAR1或hADAR2。在一些实施方案中,使用包含至少一个dsRBD和脱氨酶结构域的全长ADAR蛋白。在一些实施方案中,一个或多个dsRBM或dsRBD在脱氨酶结构域的N端。在其他实施方案中,一个或多个dsRBM或dsRBD在脱氨酶结构域的C端。

[0239] 在一些实施方案中,腺苷脱氨酶的底物编辑偏好受酶活性中心附近或其内的氨基酸残基的影响。在一些实施方案中,为了修改底物编辑偏好,腺苷脱氨酶可包含基于hADAR2-D的氨基酸序列位置的一个或多个以下突变:G336D、G487R、G487K、G487W、G487Y、E488Q、E488N、T490A、V493A、V493T、V493S、N597K、N597R、A589V、S599T、N613K、N613R,以及同源ADAR蛋白中与上述突变相对应的突变。

[0240] 具体地,在一些实施方案中,为了降低编辑特异性,腺苷脱氨酶可以包含基于hADAR2-D的氨基酸序列位置的一个或多个以下突变:E488Q、V493A、N597K、N613K,以及同源ADAR蛋白中与上述突变相对应的突变。在一些实施方案中,为了增加编辑特异性,腺苷脱氨酶可以包含突变T490A。

[0241] 在一些实施方案中,为了增加对具有直接5’G的靶腺苷(A)(诸如包含三联体序列GAC、中心A为靶腺苷残基的底物)的编辑偏好,腺苷脱氨酶可以包含基于hADAR2-D的氨基酸序列位置的一个或多个以下突变:G336D、E488Q、E488N、V493T、V493S、V493A、A589V、N597K、N597R、S599T、N613K、N613R,以及同源ADAR蛋白中与上述突变相对应的突变。

[0242] 具体地,在一些实施方案中,腺苷脱氨酶包含突变E488Q或同源ADAR蛋白中用于编辑包含以下三联体序列的底物的对应突变:GAC、GAA、GAU、GAG、CAU、AAU、UAC,中心A为靶腺苷残基。

[0243] 在一些实施方案中，腺苷脱氨酶包含如SEQ ID No.761中定义的hADAR1-D的野生型氨基酸序列。在一些实施方案中，腺苷脱氨酶在hADAR1-D序列中包含一个或多个突变，使得根据特定需要改变编辑效率和/或hADAR1-D的底物编辑偏好。

[0244] 在一些实施方案中，腺苷脱氨酶在hADAR1-D氨基酸序列的甘氨酸1007处或在同源ADAR蛋白中的对应位置处包含突变。在一些实施方案中，位置1007处的甘氨酸残基被具有相对小的侧链的非极性氨基酸残基替换。例如，在一些实施方案中，位置1007处的甘氨酸残基被丙氨酸残基替换(G1007A)。在一些实施方案中，位置1007处的甘氨酸残基被缬氨酸残基替换(G1007V)。在一些实施方案中，位置1007处的甘氨酸残基被具有相对大的侧链的氨基酸残基替换。在一些实施方案中，位置1007处的甘氨酸残基被精氨酸残基替换(G1007R)。在一些实施方案中，位置1007处的甘氨酸残基被赖氨酸残基替换(G1007K)。在一些实施方案中，位置1007处的甘氨酸残基被色氨酸残基替换(G1007W)。在一些实施方案中，位置1007处的甘氨酸残基被酪氨酸残基替换(G1007Y)。另外，在其他实施方案中，位置1007处的甘氨酸残基被亮氨酸残基替换(G1007L)。在其他实施方案中，位置1007处的甘氨酸残基被苏氨酸残基替换(G1007T)。在其他实施方案中，位置1007处的甘氨酸残基被丝氨酸残基替换(G1007S)。

[0245] 在一些实施方案中，腺苷脱氨酶在hADAR1-D氨基酸序列的谷氨酸1008处或在同源ADAR蛋白中的对应位置处包含突变。在一些实施方案中，位置1008处的谷氨酸残基被具有相对大的侧链的极性氨基酸残基替换。在一些实施方案中，位置1008处的谷氨酸残基被谷氨酰胺残基替换(E1008Q)。在一些实施方案中，位置1008处的谷氨酸残基被组氨酸残基替换(E1008H)。在一些实施方案中，位置1008处的谷氨酸残基被精氨酸残基替换(E1008R)。在一些实施方案中，位置1008处的谷氨酸残基被赖氨酸残基替换(E1008K)。在一些实施方案中，位置1008处的谷氨酸残基被非极性或较小极性氨基酸残基替换。在一些实施方案中，位置1008处的谷氨酸残基被苯丙氨酸残基替换(E1008F)。在一些实施方案中，位置1008处的谷氨酸残基被色氨酸残基替换(E1008W)。在一些实施方案中，位置1008处的谷氨酸残基被甘氨酸残基替换(E1008G)。在一些实施方案中，位置1008处的谷氨酸残基被异亮氨酸残基替换(E1008I)。在一些实施方案中，位置1008处的谷氨酸残基被缬氨酸残基替换(E1008V)。在一些实施方案中，位置1008处的谷氨酸残基被脯氨酸残基替换(E1008P)。在一些实施方案中，位置1008处的谷氨酸残基被丝氨酸残基替换(E1008S)。在其他实施方案中，位置1008处的谷氨酸残基被天冬酰胺残基替换(E1008N)。在其他实施方案中，位置1008处的谷氨酸残基被丙氨酸残基替换(E1008A)。在其他实施方案中，位置1008处的谷氨酸残基被蛋氨酸残基替换(E1008M)。在一些实施方案中，位置1008处的谷氨酸残基被亮氨酸残基替换(E1008L)。

[0246] 在一些实施方案中，为了提高编辑效率，腺苷脱氨酶可包含基于hADAR1-D的氨基酸序列位置的一个或多个以下突变：E1007S、E1007A、E1007V、E1008Q、E1008R、E1008H、E1008M、E1008N、E1008K，以及同源ADAR蛋白中与上述突变相对应的突变。

[0247] 在一些实施方案中，为了降低编辑效率，腺苷脱氨酶可包含基于hADAR1-D的氨基酸序列位置的一个或多个以下突变：E1007R、E1007K、E1007Y、E1007L、E1007T、E1008G、E1008I、E1008P、E1008V、E1008F、E1008W、E1008S、E1008N、E1008K，以及同源ADAR蛋白中与上述突变相对应的突变。

[0248] 在一些实施方案中,腺苷脱氨酶的底物编辑偏好、效率和/或选择性受酶活性中心附近或其内的氨基酸残基的影响。在一些实施方案中,腺苷脱氨酶在hADAR1-D序列中的谷氨酸1008位置处或在同源ADAR蛋白中的对应位置处包含突变。在一些实施方案中,突变是E1008R,或同源ADAR蛋白中的对应突变。在一些实施方案中,E1008R突变体对在相反链上具有错配的G残基的靶腺苷残基具有增加的编辑效率。

[0249] 在一些实施方案中,腺苷脱氨酶蛋白还包含或连接到一个或多个双链RNA(dsRNA)结合基序(dsRBM)或结构域(dsRBD),以识别并结合到双链核酸底物。在一些实施方案中,腺苷脱氨酶与双链底物之间的相互作用由一种或多种其他蛋白因子(包括CRISPR/CAS蛋白因子)介导。在一些实施方案中,腺苷脱氨酶与双链底物之间的相互作用还由一种或多种核酸组分(包括指导RNA)介导。

[0250] 在某些示例实施方案中,定向进化可用于设计经修饰的ADAR蛋白,该蛋白除了能够将腺嘌呤脱氨为次黄嘌呤外,还能够催化另外的反应。例如

[0251] 根据本发明,腺苷脱氨酶的底物是在指导分子与其DNA靶标结合时形成的RNA/DNAn RNA双链体,该双链体然后与CRISPR-Cas酶形成CRISPR-Cas复合物。RNA/DNA或DNA/RNAn RNA双链体在本文中也称为“RNA/DNA杂合体”、“DNA/RNA杂合体”或“双链底物”。指导分子和CRISPR-Cas酶的具体特征在下面详述。

[0252] 本文所用的术语“编辑选择性”是指由腺苷脱氨酶编辑的双链底物上所有位点的分数。不受理论的束缚,考虑了腺苷脱氨酶的编辑选择性受双链底物的长度和二级结构(诸如碱基错配、凸起和/或内部环的存在)的影响。

[0253] 在一些实施方案中,当底物是长于50bp的完全碱基配对的双链体时,腺苷脱氨酶可能能够使双链体内的多个腺苷残基(例如,所有腺苷残基的50%)脱氨。在一些实施方案中,当底物短于50bp时,腺苷脱氨酶的编辑选择性受靶腺苷位点处存在错配的影响。具体地,在一些实施方案中,在相反链上具有错配的胞苷(C)残基的腺苷(A)残基被高效脱氨。在一些实施方案中,在相反链上具有错配的鸟苷(G)残基的腺苷(A)残基被跳过而不进行编辑。

[0254] CRISPR-Cas蛋白和指导物

[0255] 在本发明的方法和系统中,使用CRISPR-Cas蛋白和对应的指导分子。更具体地,CRISPR-Cas蛋白是2类CRISPR-Cas蛋白。在某些实施方案中,所述CRISPR-Cas蛋白是Cas13。CRISPR-Cas系统不需要生成靶向特定序列的定制蛋白,而是可以通过指导分子对单个Cas蛋白进行编程以识别特定核酸靶标,换句话说,可以使用所述指导分子将Cas酶蛋白募集到特定核酸序列目标靶基因座。

[0256] 指导分子

[0257] 2类V型CRISPR-Cas蛋白的指导分子或指导RNA包含tracr-mate序列(在内源CRISPR系统的情况下,包含“正向重复序列”)和指导序列(在内源CRISPR系统的情况下,也称为“间隔子”)。实际上,与II型CRISPR-Cas蛋白相反,Cas13蛋白不依赖于tracr序列的存在。在一些实施方案中,本文所述的CRISPR-Cas系统或复合物不包含和/或不依赖于tracr序列的存在(例如,如果Cas蛋白是Cas13)。在某些实施方案中,指导分子可包含与指导序列或间隔序列融合或连接的正向重复序列、基本上由所述序列组成或由所述序列组成。

[0258] 一般来讲,CRISPR系统的特征在于促进在靶序列的位点处形成CRISPR复合物的元

件。在形成CRISPR复合物的情况下，“靶序列”是指指导序列被设计成与其具有互补性的序列，其中靶DNA序列和指导序列之间的杂交促进CRISPR复合物的形成。

[0259] 术语“指导分子”和“指导RNA”在本文中可互换使用，是指能够与CRISPR-Cas蛋白形成复合物并且包含指导序列的基于RNA的分子，所述指导序列与靶核酸序列具有足够互补性以与靶核酸序列杂交并且指导复合物与靶核酸序列进行序列特异性结合。指导分子或指导RNA特别包括具有本文所述的一种或多种化学修饰的基于RNA的分子(例如，通过化学连接两个核糖核苷酸或者通过用一种或多种脱氧核糖核苷酸替换一种或多种核糖核苷酸)。

[0260] 如本文所用，在CRISPR-Cas系统的情况下，术语“指导序列”包含与靶核酸序列具有足够互补性以与靶核酸序列杂交并且指导核酸靶向复合物与靶核酸序列进行序列特异性结合的任何多核苷酸序列。在本发明的上下文中，靶核酸序列或靶序列是包含要脱氨的靶腺苷(在本文中也称为“靶腺苷”)的序列。在一些实施方案中，除了预期的dA-C错配，当使用合适的比对算法进行最佳比对时，互补程度为约或大于约50％、60％、75％、80％、85％、90％、95％、97.5％、99％或更高。可使用用于比对序列的任何合适的算法来确定最佳比对，这些算法的非限制性实例包括Smith-Waterman算法、Needleman-Wunsch算法、基于Burrows-Wheeler变换的算法(例如，Burrows Wheeler Aligner)、ClustalW、Clustal X、BLAT、Novoalign(Novocraft Technologies;可从www.novocraft.com获取)、ELAND(Illumina,San Diego,CA)、SOAP(可从soap.genomics.org.cn获取)和Maq(可从maq.sourceforge.net获取)。可通过任何合适的测定来评估指导序列(在核酸靶向指导RNA内)指导核酸靶向复合物与靶核酸序列进行序列特异性结合的能力。例如，可诸如通过用编码核酸靶向复合物的组分的载体进行转染，然后诸如通过本文所述的Surveyor测定对靶核酸序列内的优先靶向(例如，裂解)进行评估，将足以形成核酸靶向复合物的核酸靶向CRISPR系统组分(包括待测试的指导序列)提供给具有对应靶核酸序列的宿主细胞。类似地，可通过提供靶核酸序列、核酸靶向复合物的组分(包括待测试的指导序列和不同于该测试指导序列的对照指导序列)，并比较测试指导序列和对照指导序列反应之间在靶序列处或附近的结合或裂解速率，在试管中对靶核酸序列(或其附近的序列)的裂解进行评估。其他测定也是可能的，并且是本领域技术人员可以想到的。可选择指导序列并因此可选择核酸靶向指导RNA，以靶向任何靶核酸序列。

[0261] 在一些实施方案中，指导分子包含被设计成与靶序列具有至少一个错配的指导序列，使得在指导序列和靶序列之间形成的RNA双链体在指导序列中包含与靶标A相对的非配对C，以在靶序列上进行脱氨。在一些实施方案中，除了该A-C错配外，当使用合适的比对算法进行最佳比对时，互补程度为约或大于约50％、60％、75％、80％、85％、90％、95％、97.5％、99％或更高。

[0262] 在某些实施方案中，指导分子的指导序列或间隔子长度为15至50nt。在某些实施方案中，指导RNA的间隔子长度为至少15个核苷酸。在某些实施方案中，间隔子长度为15至17nt(例如，15、16或17nt)、17至20nt(例如，17、18、19或20nt)、20至24nt(例如，20、21、22、23或24nt)、23至25nt(例如，23、24或25nt)、24至27nt(例如，24、25、26或27nt)、27至30nt(例如，27、28、29或30nt)、30至35nt(例如，30、31、32、33、34或35nt)或35nt或更长。在某些示例实施方案中，指导序列为15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、

32、33、34、35、36、37、38、39 40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99或100nt。

[0263] 在一些实施方案中，指导序列是长度介于10至50nt之间(但更具体地为约20-30nt，有利地为约20nt、23-25nt或24nt)的RNA序列。选择指导序列以便确保其与包含要脱氨的腺苷的靶序列杂交。这将在下面更详细地描述。选择可以包括提高脱氨的效率和特异性的另外步骤。

[0264] 在一些实施方案中，指导序列的长度为约20nt至约30nt，并且与靶DNA链杂交以形成几乎完美匹配的双链体，除了在靶腺苷位点具有dA-C错配。具体地，在一些实施方案中，dA-C错配靠近靶序列的中心(因此在指导序列与靶序列杂交后，靠近双链体的中心)，从而将腺苷脱氨酶限制在狭窄的编辑窗口(例如，约4bp宽)。在一些实施方案中，靶序列可包含一个以上的要脱氨的靶腺苷。在另外的实施方案中，靶序列还可在靶腺苷位点3'包含一个或多个dA-C错配。在一些实施方案中，为了避免在靶序列中非预期的腺嘌呤位点进行脱靶编辑，指导序列可以被设计成在对应于所述非预期的腺嘌呤的位置处包含非配对鸟嘌呤，以引入dA-G错配，这种错配对于某些腺苷脱氨酶(诸如ADAR1和ADAR2)在催化上不利。参见Wong等人，RNA 7:846-858(2001)，该文献全文通过引用并入本文。

[0265] 在一些实施方案中，具有标准长度(例如，对于AsCas13为约24nt)的Cas13指导序列用于与靶DNA形成RNA双链体。在一些实施方案中，长于标准长度(例如，对于AsCas13为>24nt)的Cas13指导分子用于与靶DNA形成RNA双链体，包括Cas13-指导RNA-靶DNA复合物的外部。当在给定的一段核苷酸内对一个以上的腺嘌呤进行脱氨有意义时，这可能是有意义的。在替代实施方案中，保持标准指导序列长度的限制是是有意义的。在一些实施方案中，指导序列被设计成在Cas13指导物的标准长度之外引入dA-C错配，这可通过Cas13降低空间位阻并增加腺苷脱氨酶和dA-C错配之间的接触频率。

[0266] 在一些实施方案中，选择指导分子的序列(正向重复序列和/或间隔子)以减少指导分子内的二级结构程度。在一些实施方案中，当进行最佳折叠时，核酸靶向指导RNA的约或小于约75％、50％、40％、30％、25％、20％、15％、10％、5％、1％或更少的核苷酸参与自身互补碱基配对。最佳折叠可通过任何合适的多核苷酸折叠算法来确定。一些程序基于计算最小吉布斯自由能。这种算法的一个实例是mFold，如Zuker和Stiegler(Nucleic Acids Res.9(1981)，133-148)所述。折叠算法的另一个实例是在线网络服务器RNAfold，由维也纳大学(University of Vienna)理论化学研究所开发，使用质心结构预测算法(参见例如A.R.Gruber等人，2008，Cell 106(1):23-24；以及PA Carr和GM Church，2009，Nature Biotechnology 27(12):1151-62)。

[0267] 在一些实施方案中，降低指导分子对RNA裂解(诸如通过Cas13进行的裂解)的敏感性是有意义的。因此，在特定的实施方案中，指导分子被调节为避免被Cas13或其他RNA裂解酶裂解。

[0268] 在某些实施方案中，指导分子包含非天然存在的核酸和/或非天然存在的核苷酸和/或核苷酸类似物和/或化学修饰。优选地，这些非天然存在的核酸和非天然存在的核苷酸位于指导序列之外。非天然存在的核酸可以包括例如天然和非天然存在的核苷酸的混合物。非天然存在的核苷酸和/或核苷酸类似物可在核糖、磷酸和/或碱基部分处被修饰。在本

发明的一个实施方案中，指导核酸包含核糖核苷酸和非核糖核苷酸。在一个这种实施方案中，指导物包含一种或多种核糖核苷酸和一种或多种脱氧核糖核苷酸。在本发明的一个实施方案中，指导物包含一个或多个非天然存在的核苷酸或核苷酸类似物，诸如具有硫代磷酸酯键、在核糖环的2'和4'碳之间包含亚甲基桥的锁核酸(LNA)核苷酸，或桥接核酸(BNA)。经修饰的核苷酸的其他实例包括2'-O-甲基类似物、2'-脱氧类似物或2'-氟类似物。经修饰的碱基的其他实例包括但不限于2-氨基嘌呤、5-溴尿苷、假尿苷、肌苷、7-甲基鸟苷。指导RNA化学修饰的实例包括但不限于在一个或多个末端核苷酸处并入2'-O-甲基(M)、2'-O-甲基3'硫代磷酸酯(MS)、S-约束乙基(cEt)或2'-O-甲基3'硫代膦酰基乙酸酯(MSP)。与未修饰的指导物相比，此类经化学修饰的指导物的稳定性和活性可能有所增加，但是无法预测中靶和脱靶特异性。(参见Hendel,2015,Nat Biotechnol.33(9):985-9,doi:10.1038/nbt.3290,2015年6月29日在线发布;Ragdarm等人,0215,PNAS,E7110-E7111;Allerson等人,J.Med.Chem.2005,48:901-904;Bramsen et al.,Front.Genet.,2012,3:154;Deng等人,PNAS,2015,112:11870-11875;Sharma et al.,MedChemComm.,2014,5:1454-1471;Hendel等人,Nat.Biotechnol.(2015)33(9):985-989;Li等人,Nature Biomedical Engineering,2017,1,0066DOI:10.1038/s41551-017-0066)。在一些实施方案中，指导RNA的5'和/或3'端被包括荧光染料、聚乙二醇、胆固醇、蛋白质或检测标签在内的多种功能部分修饰。(参见Kelly等人,2016,J.Biotech.233:74-83)。在某些实施方案中，指导物在与靶DNA结合的区域中包含核糖核苷酸，并且在与Cas13结合的区域中包含一种或多种脱氧核糖核苷酸和/或核苷酸类似物。在本发明的一个实施方案中，将脱氧核糖核苷酸和/或核苷酸类似物并入工程化的指导结构中，诸如但不限于茎环区和种子区。对于Cas13指导物，在某些实施方案中，修饰不在茎环区的5'-柄中。指导物的茎环区的5'-柄中的化学修饰可能消除其功能(参见Li等人,Nature Biomedical Engineering,2017,1:0066)。在某些实施方案中，对指导物的至少1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、35、40、45、50或75个核苷酸进行化学修饰。在一些实施方案中，对指导物的3'端或5'端处的3-5个核苷酸进行化学修饰。在一些实施方案中，仅在种子区中引入微小修饰，诸如2'-F修饰。在某些实施方案中，在指导物的3'端引入2'-F修饰。在某些实施方案中，用2'-O-甲基(M)、2'-O-甲基3'-硫代磷酸酯(MS)、S-约束乙基(cEt)或2'-O-甲基3'-硫代膦酰基乙酸酯(MSP)化学修饰指导物的5'端和/或3'端处的3至5个核苷酸。这种修饰可以提高基因组编辑效率(参见Hendel等人,Nat.Biotechnol.(2015)33(9):985-989)。在某些实施方案中，用硫代磷酸酯(PS)取代指导物的所有磷酸二酯键，以增强基因破坏水平。在某些实施方案中，用2'-O-Me、2'-F或S-约束乙基(cEt)化学修饰指导物的5'端和/或3'端处的五个以上核苷酸。这种经化学修饰的指导物可以介导增强基因破坏水平(参见Ragdarm等人,0215,PNAS,E7110-E7111)。在本发明的一个实施方案中，指导物经修饰以在其3'端和/或5'端包含化学部分。这些部分包括但不限于胺、叠氮化物、炔、硫代物、二苯并环辛炔(DBCO)或若丹明。在某些实施方案中，通过接头诸如烷基链将化学部分与指导物缀合。在某些实施方案中，可以使用经修饰的指导物的化学部分将指导物附接到另一分子，诸如DNA、RNA、蛋白质或纳米颗粒。可以使用这种经化学修饰的指导物来识别或富集通常由CRISPR系统编辑的细胞(参见Lee等人,eLife,2017,6:e25312,DOI:10.7554)。

[0269] 在一些实施方案中，指导物包含具有5'-柄和指导区段的经修饰的Cas13crRNA，所

述引导区段还包含种子区和3'末端。在一些实施方案中,经修饰的指导物可以与以下任一种的Cas13一起使用:氨基酸球菌属某种(Acidaminococcus sp.)BV3L6Cas13(AsCas13);土拉弗朗西斯菌新凶手亚种(Francisella tularensis subsp.Novicida)U112Cas13(FnCas13);毛螺菌科细菌(L.bacterium)MC2017Cas13(Lb3Cas13);瘤胃溶纤维丁酸弧菌(Butyrivibrio proteoclasticus)Cas13(BpCas13);帕库菌(Parcubacteria bacterium)GWC2011_GWC2_44_17Cas13(PbCas13);佩氏菌(Peregrinibacteria bacterium)GW2011_GWA_33_10Cas13(PeCas13);稻田钩端螺旋体(Leptospira inadai)Cas13(LiCas13);史密斯菌属某种(Smithella sp.)SC_K08D17Cas13(SsCas13);毛螺菌科细菌(L.bacterium)MA2020Cas13(Lb2Cas13);狗口腔卟啉单胞菌(Porphyromonas crevioricanis)Cas13(PcCas13);猕猴卟啉单胞菌(Porphyromonas macacae)Cas13(PmCas13);候选白蚁甲烷支原体(Candidatus Methanoplasma termitum)Cas13(CMtCas13);挑剔真杆菌(Eubacterium eligens)Cas13(EeCas13);牛莫拉氏菌(Moraxella bovoculi)237Cas13(MbCas13);解糖胨普雷沃菌(Prevotella disiens)Cas13(PdCas13);或者毛螺菌科细菌(L.bacterium)ND2006Cas13(LbCas13)。

[0270] 在一些实施方案中,对指导物的修饰是化学修饰、插入、缺失或断裂。在一些实施方案中,化学修饰包括但不限于并入2'-O-甲基(M)类似物、2'-脱氧类似物、2-硫尿核苷类似物、N6-甲基腺苷类似物、2'-氟类似物、2-氨基嘌呤、5-溴尿苷、假尿苷(Ψ)、N1-甲基假尿苷(me1Ψ)、5-甲氧基尿苷(5moU)、肌苷、7-甲基鸟苷、2'-O-甲基3'-硫代磷酸酯(MS)、S-约束乙基(cEt)、硫代磷酸酯(PS)或2'-O-甲基3'-硫代膦酰基乙酸酯(MSP)。在一些实施方案中,指导物包含一种或多种硫代磷酸酯修饰。在某些实施方案中,对指导物的至少1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20或25个核苷酸进行化学修饰。在某些实施方案中,对种子区中的一个或多个核苷酸进行化学修饰。在某些实施方案中,对3'端中的一个或多个核苷酸进行化学修饰。在某些实施方案中,不对5'-柄中的任何核苷酸进行化学修饰。在一些实施方案中,种子区中的化学修饰是微小修饰,诸如并入2'-氟类似物。在一个特定的实施方案中,用2'-氟类似物替换种子区的一个核苷酸。在一些实施方案中,对3'末端中的5至10个核苷酸进行化学修饰。在Cas13CrRNA的3'末端进行此类化学修饰可提高Cas13活性(参见Li等人,Nature Biomedical Engineering,2017,1:0066)。在一个特定的实施方案中,用2'-氟类似物替换3'末端中的1、2、3、4、5、6、7、8、9或10个核苷酸。在一个特定的实施方案中,用2'-O-甲基(M)类似物替换3'末端中的1、2、3、4、5、6、7、8、9或10个核苷酸。

[0271] 在一些实施方案中,对指导物的5'-柄的环进行修饰。在一些实施方案中,将指导物的5'-柄的环修饰成具有缺失、插入、断裂或化学修饰。在某些实施方案中,经修饰的环包含3、4或5个核苷酸。在某些实施方案中,环包含序列UCUU、UUUU、UAUU或UGUU。

[0272] 在一些实施方案中,指导分子与单独的非共价连接的序列(其可以是DNA或RNA)形成茎环。在特定的实施方案中,首先使用标准亚磷酰胺合成方案(Herdewijn,P.编,Methods in Molecular Biology Col 288,Oligonucleotide Synthesis:Methods and Applications,Humana Press,New Jersey(2012))合成形成指导物的序列。在一些实施方案中,可以使用本领域已知的标准方案(Hermanson,G.T.,Bioconjugate Techniques,Academic Press(2013))将这些序列官能化成包含用于连接的合适官能团。官能团的实例包括但不限于羟基、胺、羧酸、羧酸卤化物、羧酸活性酯、醛、羰基、氯羰基、咪唑基羰基、酰肼

(hydrozide)、氨基脲、硫代氨基脲、硫醇、马来酰亚胺、卤代烷基、磺酰基、烯丙基、炔丙基、二烯、炔和叠氮化物。一旦该序列被官能化,就可以在该序列和正向重复序列之间形成共价化学键。化学键的实例包括但不限于基于以下各项的化学键:氨基甲酸酯、醚、酯、酰胺、亚胺、脒、氨基三嗪、腙、二硫化物、硫醚、硫酯、硫代磷酸酯、二硫代磷酸酯、磺酰胺、磺酸酯、砜、亚砜、脲、硫脲、酰肼、肟、三唑、光不稳定键、C-C键形成基团(诸如,Diels-Alder环加成对或闭环复分解对以及Michael反应对)。

[0273] 在一些实施方案中,可以化学合成这些茎环形成序列。在一些实施方案中,化学合成使用自动化固相寡核苷酸合成机并利用2'-乙酰氧基乙基原酸酯(2'-ACE)(Scaringe等人,J.Am.Chem.Soc.(1998)120:11820-11821;Scaringe,Methods Enzymol.(2000)317:3-18)或2'-硫羰氨基甲酸酯(2'-TC)化学(Dellinger等人,J.Am.Chem.Soc.(2011)133:11540-11546;Hendel等人,Nat.Biotechnol.(2015)33:985-989)。

[0274] 在某些实施方案中,指导分子(能够将Cas13引导到靶基因座)包含:(1)能够与靶基因座杂交的引导序列;以及(2)tracr-mate或正向重复序列,其中所述正向重复序列位于指导序列上游(即5')。在一个特定的实施方案中,Cas13指导序列的种子序列(即对于识别靶基因座处的序列和/或与其杂交而言至关重要的序列)大约在指导序列的前10个核苷酸内。在特定的实施方案中,Cas13是FnCas13,并且种子序列大约在指导序列的5'端的前5nt内。

[0275] 在一个特定的实施方案中,指导分子包含与正向重复序列连接的指导序列,其中正向重复序列包含一个或多个茎环或优化的二级结构。在特定的实施方案中,正向重复序列的最小长度为16nt,并且具有单个茎环。在另外的实施方案中,正向重复序列的长度大于16nt,优选地大于17nt,并且具有多于一个的茎环或优化的二级结构。在特定的实施方案中,指导分子包含与全部或部分天然正向重复序列连接的指导序列,或由所述指导序列组成。典型的V型Cas13指导分子包含(在3'至5'方向):引导序列、第一互补段("重复序列")、环(通常长4或5个核苷酸)、第二互补段("抗重复序列",与重复序列互补)和聚A(通常为RNA中的聚U)尾巴(终止子)。在某些实施方案中,正向重复序列保留其天然结构并形成单个茎环。在特定的实施方案中,可以例如通过添加、减少或取代特征来修饰指导物结构的某些方面,但维持指导物结构的某些其他方面。工程化的指导分子修饰(包括但不限于插入、缺失和取代)的优选位置包括指导物末端以及与Cas13蛋白和/或靶标复合时暴露的指导分子区域(例如,正向重复序列的茎环)。

[0276] 在特定的实施方案中,茎包含至少约4bp,其包含互补的X和Y序列,但是还考虑了具有更多(例如,5、6、7、8、9、10、11或12个)或更少(例如,3、2个)碱基对的茎。因此,可考虑例如X2-10和Y2-10(其中X和Y代表任何互补的核苷酸组)。一方面,由X和Y核苷酸构成的茎与环一起将在整个二级结构中形成完整的发夹;这可能是有利的,并且碱基对的量可以是形成完整发夹的任何量。一方面,任何互补的X:Y碱基配对序列(例如,关于长度)都是可接纳的,只要保留整个指导分子的二级结构即可。一方面,连接由X:Y碱基对构成的茎的环可以是不会打断指导分子的整体二级结构的相同长度(例如,4或5个核苷酸)或更长的任何序列。一方面,茎环还可以包括例如MS2适配体。一方面,茎包含约5-7bp,其包含互补的X和Y序列,但是还考虑了具有更多或更少碱基对的茎。一方面,考虑了非沃森克里克碱基配对,其中这种配对通常以其他方式在该位置处保留茎环的结构。

[0277] 在特定的实施方案中，指导分子的天然发夹或茎环结构延伸或被延伸的茎环替代。已经证明茎的延伸可以增强指导分子与CRISPR-Cas蛋白的组装(Chen等人，Cell.(2013)；155(7)：1479-1491)。在特定的实施方案中，茎环的茎延伸至少1、2、3、4、5或更多个互补碱基对(即对应于在指导分子中添加2、4、6、8、10或更多个核苷酸)。在特定的实施方案中，它们位于茎的末端，邻近茎环的环。

[0278] 在特定的实施方案中，可以通过稍微修饰指导分子的序列而不影响其功能来降低指导分子对RNA酶或对表达降低的敏感性。例如，在特定的实施方案中，可以通过修饰指导分子序列中的推定Pol-III终止子(4个连续U)来去除转录(诸如U6Pol-III的提前转录)的提前终止。当在指导分子的茎环中需要这种序列修饰时，优选地通过碱基对翻转来确保。

[0279] 在一个优选的实施方案中，可将正向重复序列修饰为包含一种或多种蛋白结合RNA适配体。在一个特定的实施方案中，可包括一种或多种适配体，诸如优化的二级结构的一部分。此类适配体可能能够结合噬菌体外壳蛋白，如本文进一步详述。

[0280] 在一些实施方案中，指导分子与包含至少一个要编辑的靶腺苷残基的靶DNA链形成双链体。指导RNA分子与靶DNA链杂交后，腺苷脱氨酶与双链体结合并催化DNA-RNA双链体内包含的一个或多个靶腺苷残基脱氨。

[0281] 可选择指导序列并因此可选择核酸靶向指导RNA，以靶向任何靶核酸序列。靶序列可以是DNA。靶序列可以是基因组DNA。靶序列可以是线粒体DNA。

[0282] 在某些实施方案中，靶序列应与PAM(原间隔区邻近基序)或PFS(原间隔区侧接序列或位点)相联；也就是说，是由CRISPR复合物识别的短序列。根据CRISPR-Cas蛋白的性质，应选择靶序列，使得其在DNA双链体中的互补序列(在本文中也称为非靶序列)在PAM的上游或下游。在CRISPR-Cas蛋白是Cas13蛋白的本发明实施方案中，靶序列的互补序列在PAM的下游或3'。PAM的精确序列和长度要求因所使用的Cas13蛋白而异，但PAM通常是与原间隔区(即靶序列)相邻的2-5个碱基对序列。下文提供了用于不同Cas13同源物的天然PAM序列的实例，并且技术人员将能够鉴定与给定Cas13蛋白一起使用的其他PAM序列。

[0283] 此外，对PAM相互作用(PI)结构域进行工程化可允许对PAM特异性进行编程、提高靶位点识别保真度并增加CRISPR-Cas蛋白的多功能性，例如以下文献中针对Cas9所述：Kleinstiver BP等人，Engineered CRISPR-Cas9 nucleases with altered PAM specificities.Nature.2015年7月23日；523(7561)：481-5.doi:10.1038/nature14592。如本文进一步详述，技术人员将理解可类似地修饰Cas13蛋白。

[0284] 在特定的实施方案中，选择指导序列以便确保脱氨酶对要脱氨的腺嘌呤具有最佳效率。可考虑靶链中腺嘌呤相对于Cas13切口酶的裂解位点的位置。在特定的实施方案中，确保切口酶将在非靶链上作用于要脱氨的腺嘌呤附近是有意义的。例如，在特定的实施方案中，Cas13切口酶切割非靶向链在PAM下游的17个核苷酸(例如，AsCas13、LbCas13)或在PAM下游的18个核苷酸(例如，FnCas13)，并且设计如下指导物可能是有意义的：对应于要脱氨的腺嘌呤的胞嘧啶在指导序列中位于对应非靶链的序列中切口酶裂解位点上游或下游10bp内。

[0285] 在特定的实施方案中，指导物是陪伴指导物。所谓"陪伴"是指Cas13CRISPR-Cas系统或复合物或指导物被递送到细胞内的选定时间或地点，使得Cas13CRISPR-Cas系统或复合物或指导物的活性在空间上或时间上受到控制。例如，Cas13CRISPR-Cas系统或复合物或

指导物的活性和目的地可由对适配体配体(诸如细胞表面蛋白或其他局部细胞组分)具有结合亲和力的陪伴RNA适配体序列控制。另选地,陪伴适配体可例如对细胞上或细胞中的适配体效应子(诸如瞬时效应子,诸如在特定时间施加到细胞的外部能源)作出反应。

[0286] 陪伴Cas13CRISPR-Cas系统或复合物具有指导分子,所述指导分子的功能结构被设计成改善指导分子的结构、架构、稳定性、基因表达或其任意组合。这种结构可以包括适配体。

[0287] 适配体是可以例如使用如下技术被设计或选择成与其他配体紧密结合的生物分子:称为指数富集配体系统进化的技术(SELEX;Tuerk C,Gold L:"Systematic evolution of ligands by exponential enrichment:RNA ligands to bacteriophage T4DNA polymerase."Science 1990,249:505-510)。核酸适配体可以例如选自随机序列寡核苷酸库,它们对广泛的生物医学相关靶标具有高结合亲和力和特异性,表明适配体具有广泛治疗效用(Keefe,Anthony D.,Supriya Pai和Andrew Ellington."Aptamers as therapeutics",Nature Reviews Drug Discovery 9.7(2010):537-550)。这些特征还表明适配体作为药物递送媒介具有广泛的用途(Levy-Nissenbaum,Etgar等人,"Nanotechnology and aptamers:applications in drug delivery",Trends in biotechnology 26.8(2008):442-449;以及Hicke BJ,Stephens AW."Escort aptamers:a delivery service for diagnosis and therapy",J Clin Invest 2000,106:923-928.)。适配体还可被构建成充当分子开关,通过改变性质对队列作出反应,诸如结合荧光团以模拟绿色荧光蛋白活性的RNA适配体(Paige,Jeremy S.,Karen Y.Wu和Samie R.Jaffrey."RNA mimics of green fluorescent protein."Science 333.6042(2011):642-646)。还建议可将适配体用作靶向siRNA治疗性递送系统的组分,例如靶向细胞表面蛋白(Zhou,Jiehua和John J.Rossi."Aptamer-targeted cell-specific RNA interference."Silence 1.1(2010):4)。

[0288] 因此,在特定的实施方案中,例如通过一种或多种适配体来修饰指导分子,所述一种或多种适配体被设计成改善指导分子的递送,包括跨细胞膜、到细胞内区室或到细胞核的递送。除了一种或多种适配体之外或者在没有此类一种或多种适配体的情况下,这种结构可以包括一个或多个部分以使指导分子可递送、可诱导或对选定效应子作出反应。因此,本发明包括对正常或病理生理条件(包括但不限于pH、低氧、O2浓度、温度、蛋白质浓度、酶浓度、脂质结构、曝光、机械破坏(例如超声波)、磁场、电场或电磁辐射)作出反应的指导分子。

[0289] 诱导系统的光反应性可通过隐花色素2和CIB1的激活和结合来实现。蓝光刺激在隐花色素2中诱导激活的构象变化,从而募集其结合伴侣CIB1。这种结合是快速且可逆的,在脉冲刺激后不到15秒达到饱和,并在刺激结束后不到15分钟恢复到基线。这些快速结合动力学导致系统在时间上仅受转录/翻译和转录物/蛋白质降解的速度限制,而不受诱导剂的吸收和清除限制。隐花色素2激活也是高度敏感的,允许使用低光强度刺激并减轻光毒性的风险。此外,在诸如完整的哺乳动物脑的情况下,可使用可变的光强度来控制刺激区域的大小,从而比单独的载体递送所提供的精度更高。

[0290] 本发明考虑了能量源诸如电磁辐射、声能或热能来诱导指导物。有利地,电磁辐射是可见光的组分。在一个优选的实施方案中,光是波长为约450至约495nm的蓝光。在一个特

别优选的实施方案中，波长为约488nm。在另一个优选的实施方案中，光刺激是通过脉冲进行的。光功率可在约0-9mW/cm2的范围内。在一个优选的实施方案中，每15秒低至0.25秒的刺激范例应引起最大的激活。

[0291] 化学或能量敏感指导物在诱导时可通过化学源的结合或通过能量而发生构象变化，使其充当指导物并具有Cas13CRISPR-Cas系统或复杂功能。本发明可以包括施加化学源或能量以具有指导功能和Cas13CRISPR-Cas系统或复杂功能；以及任选地进一步确定基因组基因座的表达改变。

[0292] 该化学诱导系统有几种不同的设计：1.脱落酸(ABA)诱导的基于ABI-PYL的系统(参见例如http://stke.sciencemag.org/cgi/content/abstract/sigtrans;4/164/rs2)；2.雷帕霉素(或基于雷帕霉素的相关化学物质)诱导的基于FKBP-FRB的系统(参见例如http://www.nature.com/nmeth/journal/v2/n6/full/nmeth763.html)；3.赤霉素(GA)诱导的基于GID1-GAI的系统(参见例如http://www.nature.com/nchembio/journal/v8/n5/full/nchembio.922.html)。

[0293] 化学诱导系统可以是4-羟基他莫昔芬(4OHT)诱导的基于雌激素受体(ER)的系统(参见例如http://www.pnas.org/content/104/3/1027.abstract)。雌激素受体的突变的配体结合结构域(称为ERT2)在结合4-羟基他莫昔芬时易位到细胞核中。在本发明的另外的实施方案中，任何核受体、甲状腺激素受体、视黄酸受体、雌激素受体、雌激素相关受体、糖皮质激素受体、孕酮受体和雄激素受体的任何天然存在的或工程化的衍生物可用于与基于ER的诱导系统类似的诱导系统中。

[0294] 另一种诱导系统基于如下设计：所述设计使用可通过能量、热或无线电波诱导的基于瞬时受体电位(TRP)离子通道的系统(参见例如http://www.sciencemag.org/content/336/6081/604)。这些TRP家族蛋白对不同的刺激(包括光和热)作出反应。当这种蛋白被光或热激活时，离子通道将打开并允许离子诸如钙进入质膜。这种流入的离子将结合到与多肽连接的细胞内离子相互作用伴侣(包括Cas13CRISPR-Cas复合物或系统的指导物和其他组分)上，并且该结合将诱导多肽的亚细胞定位变化，从而使得整个多肽进入细胞核。一旦进入细胞核，Cas13CRISPR-Cas复合物的指导蛋白和其他组分将变得活跃并调节细胞中的靶基因表达。

[0295] 虽然光激活可能是有利的实施方案，但是有时可能是不利的，尤其是对于其中光不可穿透皮肤或其他器官的体内应用而言。在这种情况下，考虑了其他的能量激活方法，特别是具有类似作用的电场能和/或超声。

[0296] 在体内条件下使用约1V/cm至约10kV/cm的一个或多个电脉冲，优选地基本上如本领域所述施用电场能。代替脉冲或除了脉冲之外，可以连续方式递送电场。可施加1μs至500毫秒之间、优选1μs至100毫秒之间的电脉冲。可连续地或以脉冲方式施加电场约5分钟。

[0297] 如本文所用，“电场能”是细胞暴露于其中的电能。优选地，电场在体内条件下的强度为约1V/cm至约10kV/cm或更高(参见WO97/49450)。

[0298] 如本文所用，术语“电场”包括在可变电容和电压下的一个或多个脉冲，并且包括指数波和/或方波和/或调制波和/或调制方波形式。对电场和电的提及应被认为包括对细胞环境中存在电势差的提及。如本领域中已知的，可通过静电、交流电(AC)、直流电(DC)等来建立这种环境。电场可以是均匀的、不均匀的或其他方式的，并且可以时间依赖性方式改

变强度和/或方向。

[0299] 以任何顺序和任何组合单次或多次施加电场以及单次或多次施加超声也是可能的。超声和/或电场可作为单个或多个连续应用或作为脉冲来递送(脉冲递送)。

[0300] 电穿孔已用于体外和体内程序中,以将异物引入活细胞。在体外应用中,首先将活细胞样品与目标试剂混合,然后放置在电极(例如平行板)之间。然后,电极向细胞/植入物混合物施加电场。进行体外电穿孔的系统的实例包括产品Electro Cell Manipulator ECM600和Electro Square Porator T820,两者均由Genetronics公司的BTX部门制造(参见美国专利号5,869,326)。

[0301] 已知的电穿孔技术(体外和体内)通过向位于治疗区域周围的电极施加短暂的高压脉冲而起作用。电极之间产生的电场使细胞膜暂时变为多孔,于是目标试剂分子进入细胞。在已知的电穿孔应用中,该电场包括约100.μ.s持续时间大约1000V/cm的单个方波脉冲。这种脉冲可例如在Electro Square Porator T820的已知应用中产生。

[0302] 优选地,电场在体外条件下的强度为约1V/cm至约10kV/cm。因此,电场的强度可为1V/cm、2V/cm、3V/cm、4V/cm、5V/cm、6V/cm、7V/cm、8V/cm、9V/cm、10V/cm、20V/cm、50V/cm、100V/cm、200V/cm、300V/cm、400V/cm、500V/cm、600V/cm、700V/cm、800V/cm、900V/cm、1kV/cm、2kV/cm、5kV/cm、10kV/cm、20kV/cm、50kV/cm或更高。更优选地,在体外条件下为约0.5kV/cm至约4.0kV/cm。优选地,电场在体内条件下的强度为约1V/cm至约10kV/cm。但是,在递送到靶位点的脉冲数量增加的情况下,电场强度可降低。因此,设想了较低场强度下电场的脉冲递送。

[0303] 优选地,电场的施加形式为多个脉冲,诸如具有相同强度和电容的双脉冲或具有变化强度和/或电容的顺序脉冲。如本文所用,术语“脉冲”包括在可变电容和电压下的一个或多个电脉冲,并且包括指数波和/或方波和/或调制波/方波形式。

[0304] 优选地,电脉冲作为选自指数波形式、方波形式、调制波形式和调制方波形式的波形来递送。

[0305] 优选的实施方案采用低压直流电。因此,申请人公开了以1V/cm至20V/cm之间的场强施加到细胞、组织或组织块持续100毫秒或更长、优选15分钟或更长的电场的用途。

[0306] 有利地以约0.05W/cm2至约100W/cm2的功率水平施用超声。可使用诊断或治疗超声或其组合。

[0307] 如本文所用,术语“超声”是指一种由机械振动组成的能量形式,所述机械振动的频率很高到超出人类听力的范围。超声频谱的频率下限通常可取为约20kHz。超声的大多数诊断应用都采用1至15MHz范围内的频率(摘自Ultrasonics in Clinical Diagnosis,P.N.T.Wells编,第2版,Publ.Churchill Livingstone[Edinburgh,London&NY,1977])。

[0308] 超声已用于诊断和治疗应用。当用作诊断工具(“诊断超声”)时,超声通常在高达约100mW/cm2的能量密度范围内使用(FDA建议),但是已使用高达750mW/cm2的能量密度。在物理疗法中,超声通常用作高达约3至4W/cm2范围内的能量源(WHO建议)。在其他治疗应用中,可采用更高强度的超声,例如短时间以100W/cm、高达1kW/cm2(或甚至更高)进行HIFU。本说明书中使用的术语“超声”旨在涵盖诊断、治疗和聚焦超声。

[0309] 聚焦超声(FUS)允许在不使用侵入式探头的情况下递送热能(参见Morocz等人,1998,Journal of Magnetic Resonance Imaging,第8卷,第1期,第136-142页)。聚焦超声

的另一种形式是高强度聚焦超声(HIFU),Moussatov等人在Ultrasonics(1998),第36卷,第8期,第893-900页中以及TranHuuHue等人在Acustica(1997),第83卷,第6期,第1103-1106页中对其进行了综述。

[0310] 优选地,采用诊断超声和治疗超声的组合。然而,该组合并非旨在进行限制,并且本领域读者将理解,可使用超声的任何种类的组合。另外,能量密度、超声频率和暴露时间可改变。

[0311] 优选地,暴露于超声能量源的功率密度为约0.05至约100Wcm-2。甚至更优选地,暴露于超声能量源的功率密度为约1至约15Wcm-2。

[0312] 优选地,暴露于超声能量源的频率为约0.015至约10.0MHz。更优选地,暴露于超声能量源的频率为约0.02至约5.0MHz或约6.0MHz。最优选地,超声的施加频率为3MHz。

[0313] 优选地,暴露时间为约10毫秒至约60分钟。优选地,暴露时间为约1秒至约5分钟。更优选地,超声的施加时间为约2分钟。然而,根据要破坏的特定靶细胞,暴露的持续时间可更长,例如为15分钟。

[0314] 有利地,以约0.05Wcm-2至约10Wcm-2的声功率密度且以约0.015至约10MHz范围内的频率将靶组织暴露于超声能量源(参见WO 98/52609)。然而,替代方式也是可能的,例如以高于100Wcm-2的声功率密度暴露于超声能量源,但暴露时间缩短,例如以1000Wcm-2暴露毫秒范围或更短的时间。

[0315] 优选地,超声的施加形式为多个脉冲;因此,可以任何组合使用连续波和脉冲波两者(超声的脉冲递送)。例如,可施加连续波超声,然后施加脉冲波超声,反之亦然。可以任何顺序和组合重复任意多次。可在连续波超声的背景下施加脉冲波超声,并且可在任何数量的组中使用任何数量的脉冲。

[0316] 优选地,超声可包括脉冲波超声。在一个高度优选的实施方案中,以0.7Wcm-2或1.25Wcm-2的功率密度作为连续波施加超声。如果使用脉冲波超声,则可采用更高的功率密度。

[0317] 使用超声是有利的,因为像光一样,超声可精确地聚焦在靶标上。此外,超声是有利的,因为与光不同,它可更深地聚焦到组织中。因此,它更适合整个组织穿透(诸如但不限于肝叶)或整个器官(诸如但不限于整个肝脏或整个肌肉,诸如心脏)治疗。另一个重要的优点是超声波是一种非侵入性刺激,可用于多种诊断和治疗应用。举例来说,超声在医学成像技术以及骨科治疗中是众所周知的。此外,适合将超声施加到受试脊椎动物的仪器是广泛可用的,并且其使用在本领域中是众所周知的。

[0318] 在特定的实施方案中,指导分子通过二级结构修饰以增加CRISPR-Cas系统的特异性,并且二级结构可以防止核酸外切酶活性并允许在5'添加到指导序列中,在本文中也称为受保护的指导分子。

[0319] 一方面,本发明可以使"保护RNA"与指导分子的序列杂交,其中"保护RNA"是与指导分子的3'端互补的RNA链,从而产生部分双链指导RNA。在本发明的一个实施方案中,用完全互补的保护序列保护错配的碱基(即,指导分子中不形成指导序列的一部分的碱基)降低了靶DNA在3'端与错配碱基对结合的可能性。在本发明的特定实施方案中,在指导分子内还可存在包含延长长度的其他序列,使得指导物在指导分子内包含保护序列。该"保护序列"确保了指导分子除了包括"暴露序列"(包含与靶序列杂交的指导序列部分)之外还包括"被

保护的序列”。在特定的实施方案中，指导分子通过保护指导物的存在来修饰，以包括二级结构，诸如发夹。有利地，存在三个或四个至三十个或更多个(例如约10个或更多个)与被保护的序列、指导序列或两者互补的连续碱基对。有利的是，受保护的部分不妨碍CRISPR-Cas系统与其靶标相互作用的热力学。通过提供包括部分双链指导分子的这种延伸，指导分子被认为是受保护的并且使得CRISPR-Cas复合物的特异性结合得以改善，同时保持特异性活性。

[0320] 在特定的实施方案中，使用截短的指导物(tru-guide)，即包含相对于标准指导序列长度在长度上被截短的引导序列的引导分子。如Nowak等人(Nucleic Acids Res(2016)44(20):9555-9564)所述，此类指导物可允许具有催化活性的CRISPR-Cas酶结合其靶标而不裂解靶DNA。在特定的实施方案中，使用截短的指导物，其允许靶的结合，但仅保留CRISPR-Cas酶的切口酶活性。

[0321] Crispr-Cas酶

[0322] 未修饰形式的CRISPR-Cas蛋白是一种具有催化活性的蛋白。这意味着在形成核酸靶向复合物(包含与靶序列杂交的指导RNA)时，靶序列中或附近(例如，在来自靶序列的1、2、3、4、5、6、7、8、9、10、20、50或更多个碱基对内)的一条或两条DNA链被修饰(例如，裂解)。如本文所用，术语“与目标靶基因座相联的一个或多个序列”是指靶序列附近(例如，在来自靶序列的1、2、3、4、5、6、7、8、9、10、20、50或更多个碱基对内，其中靶序列包含在目标靶基因座内)的序列。未修饰的具有催化活性的Cas13蛋白产生交错的切割，因此切割位点通常在靶序列内。更具体地，交错切割通常是PAM远端的13-23个核苷酸。在特定的实施方案中，非靶链上的切割是PAM下游的17个核苷酸(即在PAM下游的第17个和第18个核苷酸之间)，而靶链上的切割(即与指导序列杂交的链)发生在距与PAM互补的序列更远的4个核苷酸处(这是3’链上PAM互补物上游的21个核苷酸或PAM互补物上游的第21个和第22个核苷酸之间)。

[0323] 在根据本发明的方法中，优选地使CRISPR-Cas蛋白相对于对应的野生型酶突变，使得突变的CRISPR-Cas蛋白缺乏裂解含有靶基因座的靶序列的一条或两条DNA链的能力。在特定的实施方案中，使Cas13蛋白的一个或多个催化结构域突变以产生仅裂解靶序列的一条DNA链的突变Cas蛋白。

[0324] 在特定的实施方案中，可使CRISPR-Cas蛋白相对于对应的野生型酶突变，使得突变的CRISPR-Cas蛋白基本上缺乏所有DNA裂解活性。在一些实施方案中，当突变酶的裂解活性为该酶的非突变形式的核酸裂解活性的约不超过25％、10％、5％、1％、0.1％、0.01％或更少时，可认为CRISPR-Cas蛋白基本上缺乏所有DNA和/或RNA裂解活性;一个实例可以是当突变形式的核酸裂解活性与非突变形式相比为零或可忽略时。

[0325] 在本文提供的方法的某些实施方案中，CRISPR-Cas蛋白是仅裂解一条DNA链的突变CRISPR-Cas蛋白，即切口酶。更具体地，在本发明的上下文中，切口酶确保了在非靶序列(即位于靶序列的相反DNA链上并且是PAM序列的3’的序列)内进行裂解。通过进一步的指导，但不限于此，在来自氨基酸球菌属的Cas13的Nuc结构域中进行精氨酸至丙氨酸的取代(R1226A)可以将Cas13从切割两条链的核酸酶转化为切口酶(切割单链)。技术人员应理解，当酶不是AsCas13时，可在对应位置的残基处进行突变。在特定的实施方案中，Cas13是FnCas13，并且突变在位置R1218的精氨酸处。在特定的实施方案中，Cas13是LbCas13，并且突变在位置R1138的精氨酸处。在特定的实施方案中，Cas13是MbCas13，并且突变在位置

R1293的精氨酸处。

[0326] 在本文提供的方法的某些实施方案中，CRISPR-Cas蛋白的催化活性降低或没有催化活性。当CRISPR-Cas蛋白是Cas13蛋白时，参考AsCas13中的位置，突变可包括但不限于催化RuvC样结构域中的一个或多个突变，诸如D908A或E993A。

[0327] 在一些实施方案中，当突变酶的DNA裂解活性为该酶的非突变形式的DNA裂解活性的约不超过25％、10％、5％、1％、0.1％、0.01％或更少时，认为CRISPR-Cas蛋白基本上缺乏所有DNA裂解活性；一个实例可以是当突变形式的DNA裂解活性与非突变形式相比为零或可忽略时。在这些实施方案中，CRISPR-Cas蛋白用作通用DNA结合蛋白。突变可以是人工引入的突变或者功能获得或丧失突变。

[0328] 除上述突变外，可对CRISPR-Cas蛋白进行其他修饰。如本文所用，关于CRISPR-Cas蛋白的术语“修饰”通常是指衍生其的野生型Cas蛋白相比具有一个或多个修饰或突变(包括点突变、截短、插入、缺失、嵌合体、融合蛋白等)的CRISPR-Cas蛋白。所谓衍生，是指所衍生的酶在很大程度上与野生型酶具有高度序列同源性，但以本领域已知或本文所述的某种方式对其进行了突变(修饰)。

[0329] 在一些实施方案中，为了减小Cas13b效应物和一个或多个功能结构域的融合蛋白的大小，可以在仍保持Cas13b效应物的RNA结合功能的同时将其C端截短。例如，可在Cas13b效应物的C端截短至少20个氨基酸、至少50个氨基酸、至少80个氨基酸，或至少100个氨基酸，或至少150个氨基酸，或至少200个氨基酸，或至少250个氨基酸，或至少300个氨基酸，或至少350个氨基酸，或最多120个氨基酸，或最多140个氨基酸，或最多160个氨基酸，或最多180个氨基酸，或最多200个氨基酸，或最多250个氨基酸，或最多300个氨基酸，或最多350个氨基酸，或最多400个氨基酸。Cas13b截短的具体实例包括C端Δ984-1090、C端Δ1026-1090和C端Δ1053-1090、C端Δ934-1090、C端Δ884-1090、C端Δ834-1090、C端Δ784-1090和C端Δ734-1090，其中氨基酸位置对应于普雷沃菌属某种P5-125Cas13b蛋白的氨基酸位置。还可参见图67。

[0330] CRISPR-Cas蛋白的其他修饰可能会或可能不会导致功能改变。举例来说，特别是关于CRISPR-Cas蛋白，不会导致功能改变的修饰包括例如用于表达到特定宿主中的密码子优化或者提供具有特定标记的核酸酶(例如，用于可视化)。可能导致功能改变的修饰也可包括突变，包括点突变、插入、缺失、截短(包括断裂核酸酶)等。融合蛋白可但不限于包括例如与异源结构域或功能结构域(例如，定位信号、催化结构域等)的融合体。在某些实施方案中，可组合各种不同的修饰(例如，无催化活性突变核酸酶进一步与功能结构域融合，诸如以诱导DNA甲基化或另一种核酸修饰，诸如包括但不限于断裂(例如，通过不同的核酸酶(结构域))、突变、缺失、插入、替代、连接、消化、断裂或重组)。如本文所用，“功能改变”包括但不限于特异性改变(例如，靶标识别改变、特异性增加(例如“增强的”Cas蛋白)或降低或者PAM识别改变)、活性改变(例如，催化活性增加或降低，包括无催化活性的核酸酶或切口酶)和/或稳定性改变(例如，与去稳定结构域的融合体)。合适的异源结构域包括但不限于核酸酶、连接酶、修复蛋白、甲基转移酶、(病毒)整合酶、重组酶、转座酶、argonaute、胞苷脱氨酶、反转录子、II族内含子、磷酸酶、磷酸化酶、磺酰化酶(sulpfurylase)、激酶、聚合酶、核酸外切酶等。所有这些修改的实例在本领域中是已知的。应当理解，本文所指的“修饰”核酸酶、特别是“修饰”Cas或者“修饰”CRISPR-Cas系统或复合物优选地仍具有与多核酸相互作

用或结合(例如,与指导分子复合)的能力。此类修饰Cas蛋白可以与如本文所述的脱氨酶蛋白或其活性结构域组合。

[0331] 在某些实施方案中,CRISPR-Cas蛋白可包含使得活性和/或特异性增强的一种或多种修饰,诸如包括稳定靶向或非靶向链的突变残基(例如,eCas9;"Rationally engineered Cas9nucleases with improved specificity",Slaymaker等人,(2016),Science,351(6268):84-88,该文献全文通过引用并入本文)。在某些实施方案中,工程化CRISPR蛋白的活性改变或修饰包括靶向效率增加或脱靶结合减少。在某些实施方案中,工程化CRISPR蛋白的活性改变包括裂解活性修饰。在某些实施方案中,活性改变包括对靶多核苷酸基因座的裂解活性增加。在某些实施方案中,活性改变包括对靶多核苷酸基因座的裂解活性降低。在某些实施方案中,活性改变包括对脱靶多核苷酸基因座的裂解活性降低。在某些实施方案中,修饰核酸酶的活性改变或修饰包括解旋酶动力学改变。在某些实施方案中,修饰核酸酶包含如下修饰:所述修饰改变蛋白质与包含RNA的核酸分子(在Cas蛋白质的情况下)或靶多核苷酸基因座的链或脱靶多核苷酸基因座的链的缔合。在本发明的一方面,工程化CRISPR蛋白包含改变CRISPR复合物形成的修饰。在某些实施方案中,活性改变包括对脱靶多核苷酸基因座的裂解活性增加。因此,在某些实施方案中,与对脱靶多核苷酸基因座的特异性相比,对靶多核苷酸基因座的特异性增加。在其他实施方案中,与对脱靶多核苷酸基因座的特异性相比,对靶多核苷酸基因座的特异性降低。在某些实施方案中,突变导致脱靶效应(例如,裂解或结合特性、活性或动力学)降低,诸如在Cas蛋白的情况下,例如导致对靶RNA与指导RNA之间的错配耐受性降低。其他突变可能导致脱靶效应(例如,裂解或结合特性、活性或动力学)增加。其他突变可能导致中靶效应(例如,裂解或结合特性、活性或动力学)增加或降低。在某些实施方案中,突变导致解旋酶活性改变(例如,增加或降低)、功能性核酸酶复合物(例如,CRISPR-Cas复合物)的缔合或形成。在某些实施方案中,如上所述,突变导致PAM识别改变,即与未修饰Cas蛋白相比,可(另外地或替代地)识别不同的PAM。特别优选的突变包括带正电荷的残基和/或(进化上)保守的残基,诸如保守的带正电荷的残基,以便增强特异性。在某些实施方案中,可使此类残基突变为不带电荷的残基,诸如丙氨酸。

[0332] 碱基切除修复抑制剂

[0333] 在一些实施方案中,AD功能化的CRISPR系统还包含碱基切除修复(BER)抑制剂。不希望受任何特定理论的束缚,对I:T配对的存在的细胞DNA修复响应可能导致细胞中核碱基编辑效率降低。烷基腺嘌呤DNA糖基化酶(也称为DNA-3-甲基腺嘌呤糖基化酶、3-烷基腺嘌呤DNA糖基化酶或N-甲基嘌呤DNA糖基化酶)催化次黄嘌呤从细胞中的DNA中去除,这可启动碱基切除修复,结果是I:T对逆转为A:T对。

[0334] 在一些实施方案中,BER抑制剂是烷基腺嘌呤DNA糖基化酶的抑制剂。在一些实施方案中,BER抑制剂是人类烷基腺嘌呤DNA糖基化酶的抑制剂。在一些实施方案中,BER抑制剂是多肽抑制剂。在一些实施方案中,BER抑制剂是结合次黄嘌呤的蛋白质。在一些实施方案中,BER抑制剂是结合DNA中的次黄嘌呤的蛋白质。在一些实施方案中,BER抑制剂是无催化活性的烷基腺嘌呤DNA糖基化酶蛋白或其结合结构域。在一些实施方案中,BER抑制剂是不从DNA上切除次黄嘌呤的无催化活性的烷基腺嘌呤DNA糖基化酶蛋白或其结合结构域。能够抑制(例如,空间阻断)烷基腺嘌呤DNA糖基化酶碱基切除修复酶的其他蛋白质在本公开

的范围内。另外,阻断或抑制碱基切除修复的任何蛋白质也在本公开的范围内。

[0335] 不希望受任何特定理论的束缚,碱基切除修复可被结合编辑的链、阻断编辑的碱基、抑制烷基腺嘌呤DNA糖基化酶、抑制碱基切除修复、保护编辑的碱基和/或促进固定非编辑的链的分子抑制。据认为,使用本文所述的BER抑制剂可以增加能够催化A至I改变的腺苷脱氨酶的编辑效率。

[0336] 因此,在上文讨论的AD功能化的CRISPR系统的第一种设计中,可以将CRISPR-Cas蛋白或腺苷脱氨酶融合或连接到BER抑制剂(例如,烷基腺嘌呤DNA糖基化酶的抑制剂)。在一些实施方案中,BER抑制剂可以包含在以下结构中的一种中(nCas13=Cas13切口酶;dCas13=死亡Cas13):[AD]-[任选接头]-[nCas13/dCas13]-[任选接头]-[BER抑制剂];[AD]-[任选接头]-[BER抑制剂]-[任选接头]-[nCas13/dCas13];[BER抑制剂]-[任选接头]-[AD]-[任选接头]-[nCas13/dCas13];[BER抑制剂]-[任选接头]-[nCas13/dCas13]-[任选接头]-[AD];[nCas13/dCas13]-[任选接头]-[AD]-[任选接头]-[BER抑制剂];[nCas13/dCas13]-[任选接头]-[BER抑制剂]-[任选接头]-[AD]。

[0337] 类似地,在上文讨论的AD功能化的CRISPR系统的第二种设计中,可以将CRISPR-Cas蛋白、腺苷脱氨酶或衔接蛋白融合或连接到BER抑制剂(例如,烷基腺嘌呤DNA糖基化酶的抑制剂)。在一些实施方案中,BER抑制剂可以包含在以下结构中的一种中(nCas13=Cas13切口酶;dCas13=死亡Cas13):[nCas13/dCas13]-[任选接头]-[BER抑制剂];[BER抑制剂]-[任选接头]-[nCas13/dCas13];[AD]-[任选接头]-[衔接蛋白]-[任选接头]-[BER抑制剂];[AD]-[任选接头]-[BER抑制剂]-[任选接头]-[衔接蛋白];[BER抑制剂]-[任选接头]-[AD]-[任选接头]-[衔接蛋白];[BER抑制剂]-[任选接头]-[衔接蛋白]-[任选接头]-[AD];[衔接蛋白]-[任选接头]-[AD]-[任选接头]-[BER抑制剂];[衔接蛋白]-[任选接头]-[BER抑制剂]-[任选接头]-[AD]。

[0338] 在上文讨论的AD功能化的CRISPR系统的第三种设计中,可以将BER抑制剂插入CRISPR-Cas蛋白的内部环或非结构化区域中。

[0339] 靶向细胞核

[0340] 在一些实施方案中,本发明的方法涉及修饰目标靶基因座中的腺嘌呤,由此靶基因座在细胞内。为了改善用于本发明方法的CRISPR-Cas蛋白和/或腺苷脱氨酶蛋白或其催化结构域对细胞核的靶向,为这些组分中的一者或两者提供一种或多种核定位序列(NLS)可能是有利的。

[0341] 在优选的实施方案中,在本发明的上下文中使用的NLS与蛋白质是异源的。NLS的非限制性实例包括来源于以下各项的NLS序列:SV40病毒大T抗原的NLS,氨基酸序列为PKKKRKV(SEQ ID No.17)或PKKKRKVEAS(SEQ ID No.18);来自核质蛋白的NLS(例如,核质蛋白二分NLS,序列为KRPAATKKAGQAKKKK(SEQ ID No.19));c-myc NLS,氨基酸序列为PAAKRVKLD(SEQ ID No.20)或RQRRNELKRSP(SEQ ID No.21);hRNPA1M9NLS,序列为NQSSNFGPMKGGNFGGRSSGPYGGGGQYFAKPRNQGGY(SEQ ID No.22);来自输入蛋白-α的IBB结构域的序列为RMRIZFKNKGKDTAELRRRRVEVSVELRKAKKDEQILKRRNV(SEQ ID No.23);肌瘤T蛋白的序列VSRKRPRP(SEQ ID No.24)和PPKKARED(SEQ ID No.25);人p53的序列PQPKKKPL(SEQ ID No.26);小鼠c-abl IV的序列SALIKKKKKMAP(SEQ ID No.27);流感病毒NS1的序列DRLRR(SEQ ID No.28)和PKQKKRK(SEQ ID No.29);肝炎病毒δ抗原的序列RKLKKKIKKL(SEQ ID

No.30)；小鼠Mx1蛋白的序列REKKKFLKRR(SEQ ID No.31)；人聚(ADP-核糖)聚合酶的序列KRKGDEVDGVDEVAKKKSKK(SEQ ID No.32)；以及类固醇激素受体(人)糖皮质激素的序列RKCLQAGMNLEARKTKK(SEQ ID No.33)。一般来讲，一种或多种NLS具有足够的强度来驱动可检测量的靶向DNA Cas蛋白在真核细胞核中积累。一般来讲，核定位活性的强度可来源于CRISPR-Cas蛋白中NLS的数量、所使用的一种或多种特定NLS或者这些因素的组合。检测细胞核中的积累物可通过任何合适的技术进行。例如，可将可检测的标记与核酸靶向蛋白融合，使得诸如结合用于检测细胞核的位置的手段(例如，对细胞核特异的染色剂，诸如DAPI)使细胞内的位置可视化。也可从细胞中分离细胞核，然后可通过用于检测蛋白质的任何合适的方法诸如免疫组织化学、蛋白质印迹或酶活性测定来分析其内容物。也可间接地确定细胞核中的积累物，诸如通过与未暴露于CRISPR-Cas蛋白和脱氨酶蛋白或者暴露于缺乏一个或多个NLS的CRISPR-Cas和/或脱氨酶蛋白的对照相比，测定靶序列处的核酸靶向复合物形成的影响(例如，测定脱氨酶活性)，或测定受DNA靶向复合物形成和/或DNA靶向影响的改变的基因表达活性。

[0342] CRISPR-Cas和/或腺苷脱氨酶蛋白可提供有1个或更多个(诸如2、3、4、5、6、7、8、9、10个或更多个)异源NLS。在一些实施方案中，蛋白质在氨基末端处或附近包含约或大于约1、2、3、4、5、6、7、8、9、10个或更多个NLS，在羧基末端处或附近包含约或大于约1、2、3、4、5、6、7、8、9、10个或更多个NLS，或这些情况的组合(例如，在氨基末端处包含零个或至少一个或多个NLS，并且在羧基末端处包含零个或至少一个或多个NLS)。当存在一个以上的NLS时，可彼此独立地选择每个的NLS，使得单个NLS可以一个以上的拷贝存在和/或与以一个或多个拷贝存在的一个或多个其他NLS组合存在。在一些实施方案中，当NLS的最近氨基酸沿着多肽链在距N端或C端约1、2、3、4、5、10、15、20、25、30、40、50个或更多个氨基酸内时，该NLS被认为在N端或C端附近。在CRISPR-Cas蛋白的优选实施方案中，NLS附接到该蛋白的C端。

[0343] 在本文提供的方法的某些实施方案中，CRISPR-Cas蛋白和脱氨酶蛋白被递送到细胞中或在细胞内作为单独的蛋白表达。在这些实施方案中，CRISPR-Cas和脱氨酶蛋白中的每一者可以提供有如本文所述的一个或多个NLS。在某些实施方案中，CRISPR-Cas和脱氨酶蛋白被递送到细胞中或作为融合蛋白与细胞一起表达。在这些实施方案中，CRISPR-Cas和脱氨酶蛋白中的一者或两者提供有一个或多个NLS。当腺苷脱氨酶与如上所述的衔接蛋白(诸如MS2)融合时，可以在衔接蛋白上提供一个或多个NLS，只要这不干扰适配体结合。在特定的实施方案中，一个或多个NLS序列还可充当腺苷脱氨酶和CRISPR-Cas蛋白之间的接头序列。

[0344] 在某些实施方案中，本发明的指导物包含衔接蛋白的特异性结合位点(例如，适配体)，其可与腺苷脱氨酶或其催化结构域连接或融合。当此类指导物形成CRISPR复合物(即CRISPR-Cas蛋白与指导物和靶标结合)时，衔接蛋白结合，并且与衔接蛋白相联的腺苷脱氨酶或其催化结构域以有利于所赋予的功能有效的空间取向定位。

[0345] 技术人员将理解，对指导物进行允许衔接体+腺苷脱氨酶结合但不能正确定位衔接体+腺苷脱氨酶(例如，由于CRISPR复合物的三维结构内的位阻现象)的修饰是不希望的修饰。如本文所述，一个或多个修饰指导物可在四环、茎环1、茎环2或茎环3处修饰，优选地在四环或茎环2处修饰，并且最优选地在四环和茎环2两者处修饰。

[0346] 正交无催化活性的CRISPR-Cas蛋白的用途

[0347] 在特定的实施方案中,Cas13切口酶与正交无催化活性的CRISPR-Cas蛋白组合使用,以提高所述Cas13切口酶的效率(如Chen等人,2017,Nature Communications 8:14958;doi:10.1038/ncomms14958所述)。更具体地,正交无催化活性的CRISPR-Cas蛋白的特征在于与在AD功能化的CRISPR系统中使用的Cas13切口酶不同的PAM识别位点,并且选择对应的指导序列以与邻近AD功能化的CRISPR系统的Cas13切口酶的靶序列的靶序列结合。在本发明的上下文中使用的正交无催化活性的CRISPR-Cas蛋白不形成AD功能化的CRISPR系统的一部分,而仅起增加所述Cas13切口酶的效率的作用,并且与如本领域中针对所述CRISPR-Cas蛋白所述的标准指导分子组合使用。在特定的实施方案中,所述正交无催化活性的CRISPR-Cas蛋白是死亡CRISPR-Cas蛋白,即包含消除所述CRISPR-Cas蛋白的核酸酶活性的一个或多个突变。在特定的实施方案中,无催化活性的正交CRISPR-Cas蛋白提供有能够与邻近Cas13切口酶的靶序列的靶序列杂交的两个或更多个指导分子。在特定的实施方案中,至少两个指导分子用于靶向所述无催化活性的CRISPR-Cas蛋白,其中至少一个指导分子能够与Cas13切口酶的靶序列的5”靶序列杂交,并且至少一个指导分子能够与AD功能化的CRISPR系统的Cas13切口酶的靶序列的3’靶序列杂交,由此所述一个或多个靶序列可与Cas13切口酶的靶序列在相同或相反的DNA链上。在特定的实施方案中,选择正交催化无活性的CRISPR-Cas蛋白的一个或多个指导分子的指导序列,使得靶序列邻近用于靶向AD功能化的CRISPR(即用于靶向Cas13切口酶)的指导分子的靶序列。在特定的实施方案中,正交无催化活性的CRISPR-Cas酶的一个或多个靶序列各自与Cas13切口酶的靶序列隔开大于5个但小于450个碱基对。与正交无催化活性的CRISPR-Cas蛋白一起使用的指导物的靶序列与AD功能化的CRISPR系统的靶序列之间的最佳距离可以由技术人员确定。在特定的实施方案中,正交CRISPR-Cas蛋白是II类II型CRISPR蛋白。在特定的实施方案中,正交CRISPR-Cas蛋白是II类V型CRISPR蛋白。在特定的实施方案中,正交CRISPR-Cas蛋白是无催化活性的正交CRISPR-Cas蛋白。在特定的实施方案中,已经对无催化活性的正交CRISPR-Cas蛋白进行了修饰以改变其PAM特异性,如本文其他地方所述。在特定的实施方案中,Cas13蛋白切口酶是本身在人类细胞中活性有限的切口酶,但是其与无活性的正交CRISPR-Cas蛋白以及一个或多个对应的近端引导物组合可以确保所需的切口酶活性。

[0348] CRISPR开发和用途

[0349] 可基于如以下文章中阐述的CRISPR-Cas开发和用途方面来进一步说明和扩展本发明,特别是涉及细胞和生物中CRISPR蛋白复合物的递送和RNA指导的核酸内切酶的用途:

[0350] ➢Multiplex genome engineering using CRISPR-Cas systems.Cong,L.,Ran,F.A.,Cox,D.,Lin,S.,Barretto,R.,Habib,N.,Hsu,P.D.,Wu,X.,Jiang,W.,Marraffini,L.A.,&Zhang,F.Science Feb 15;339(6121):819-23(2013);

[0351] ➢RNA-guided editing of bacterial genomes using CRISPR-Cas systems.Jiang W.,Bikard D.,Cox D.,Zhang F,Marraffini LA.Nat Biotechnol Mar;31(3):233-9(2013);

[0352] ➢One-Step Generation of Mice Carrying Mutations in Multiple Genes by CRISPR-Cas-Mediated Genome Engineering.Wang H.,Yang H.,Shivalila CS.,Dawlaty MM.,Cheng AW.,Zhang F.,Jaenisch R.Cell May 9;153(4):910-8(2013);

[0353] ➢Optical control of mammalian endogenous transcription and

epigenetic states.Konermann S,Brigham MD,Trevino AE,Hsu PD,Heidenreich M,Cong L,Platt RJ,Scott DA,Church GM,Zhang F.Nature.Aug 22;500(7463):472-6.doi:10.1038/Nature12466.Epub 2013Aug 23(2013)；

[0354] ➢Double Nicking by RNA-Guided CRISPR Cas9 for Enhanced Genome Editing Specificity.Ran,FA.,Hsu,PD.,Lin,CY.,Gootenberg,JS.,Konermann,S.,Trevino,AE.,Scott,DA.,Inoue,A.,Matoba,S.,Zhang,Y.,&Zhang,F.Cell Aug 28.pii:S0092-8674(13)01015-5(2013-A)；

[0355] ➢DNA targeting specificity of RNA-guided Cas9nucleases.Hsu,P.,Scott,D.,Weinstein,J.,Ran,FA.,Konermann,S.,Agarwala,V.,Li,Y.,Fine,E.,Wu,X.,Shalem,O.,Cradick,TJ.,Marraffini,LA.,Bao,G.,&Zhang,F.Nat Biotechnol doi:10.1038/nbt.2647(2013)；

[0356] ➢Genome engineering using the CRISPR-Cas9system.Ran,FA.,Hsu,PD.,Wright,J.,Agarwala,V.,Scott,DA.,Zhang,F.Nature Protocols Nov;8(11):2281-308(2013-B)；

[0357] ➢Genome-Scale CRISPR-Cas9Knockout Screening in Human Cells.Shalem,O.,Sanjana,NE.,Hartenian,E.,Shi,X.,Scott,DA.,Mikkelson,T.,Heckl,D.,Ebert,BL.,Root,DE.,Doench,JG.,Zhang,F.Science Dec 12.(2013)；

[0358] ➢Crystal structure of cas9in complex with guide RNA and target DNA.Nishimasu,H.,Ran,FA.,Hsu,PD.,Konermann,S.,Shehata,SI.,Dohmae,N.,Ishitani,R.,Zhang,F.,Nureki,O.Cell Feb 27,156(5):935-49(2014)；

[0359] ➢Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells.Wu X.,Scott DA.,Kriz AJ.,Chiu AC.,Hsu PD.,Dadon DB.,Cheng AW.,Trevino AE.,Konermann S.,Chen S.,Jaenisch R.,Zhang F.,Sharp PA.Nat Biotechnol.Apr 20.doi:10.1038/nbt.2889(2014)；

[0360] ➢CRISPR-Cas9 Knockin Mice for Genome Editing and Cancer Modeling.Platt RJ,Chen S,Zhou Y,Yim MJ,Swiech L,Kempton HR,Dahlman JE,Parnas O,Eisenhaure TM,Jovanovic M,Graham DB,Jhunjhunwala S,Heidenreich M,Xavier RJ,Langer R,Anderson DG,Hacohen N,Regev A,Feng G,Sharp PA,Zhang F.Cell 159(2):440-455 DOI:10.1016/j.cell.2014.09.014(2014)；

[0361] ➢Development and Applications of CRISPR-Cas9 for Genome Engineering,Hsu PD,Lander ES,Zhang F.,Cell.Jun 5;157(6):1262-78(2014).

[0362] ➢Genetic screens in human cells using the CRISPR-Cas9 system,Wang T,Wei JJ,Sabatini DM,Lander ES.,Science.January 3;343(6166):80-84.doi:10.1126/science.1246981(2014)；

[0363] ➢Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation,Doench JG,Hartenian E,Graham DB,Tothova Z,Hegde M,Smith I,Sullender M,Ebert BL,Xavier RJ,Root DE.,(published online 3 September 2014)Nat Biotechnol.Dec;32(12):1262-7(2014)；

[0364] ➢In vivo interrogation of gene function in the mammalian brain using CRISPR-Cas9,Swiech L,Heidenreich M,Banerjee A,Habib N,Li Y,Trombetta J,Sur M,Zhang F.,(published online 19 October 2014)Nat Biotechnol.Jan;33(1):102-6(2015);

[0365] ➢Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex,Konermann S,Brigham MD,Trevino AE,Joung J,Abudayyeh OO,Barcena C,Hsu PD,Habib N,Gootenberg JS,Nishimasu H,Nureki O,Zhang F.,Nature.Jan 29;517(7536):583-8(2015).

[0366] ➢A split-Cas9 architecture for inducible genome editing and transcription modulation,Zetsche B,Volz SE,Zhang F.,(published online 02 February 2015)Nat Biotechnol.Feb;33(2):139-42(2015);

[0367] ➢Genome-wide CRISPR Screen in a Mouse Model of Tumor Growth and Metastasis,Chen S,Sanjana NE,Zheng K,Shalem O,Lee K,Shi X,Scott DA,Song J,Pan JQ,Weissleder R,Lee H,Zhang F,Sharp PA.Cell 160,1246-1260,March 12,2015(multiplex screen in mouse),and

[0368] ➢In vivo genome editing using Staphylococcus aureus Cas9,Ran FA,Cong L,Yan WX,Scott DA,Gootenberg JS,Kriz AJ,Zetsche B,Shalem O,Wu X,Makarova KS,Koonin EV,Sharp PA,Zhang F.,(published online 01 April 2015),Nature.Apr 9;520(7546):186-91(2015).

[0369] ➢Shalemet al.,“High-throughput functional genomics using CRISPR-Cas9,”Nature Reviews Genetics 16,299-311(May 2015).

[0370] ➢Xu et al.,“Sequence determinants of improved CRISPR sgRNA design,”Genome Research 25,1147-1157(August 2015).

[0371] ➢Parnas et al.,“A Genome-wide CRISPR Screen in Primary Immune Cells to Dissect Regulatory Networks,”Cell 162,675-686(July 30,2015).

[0372] ➢Ramanan et al.,CRISPR-Cas9cleavage of viral DNA efficiently suppresses hepatitis B virus,”Scientific Reports 5:10833.doi:10.1038/srep10833(June 2,2015)

[0373] ➢Nishimasu et al.,Crystal Structure of Staphylococcus aureus Cas9,”Cell 162,1113-1126(Aug.27,2015)

[0374] ➢BCL11A enhancer dissection by Cas9-mediated in situ saturating mutagenesis,Canver et al.,Nature 527(7577):192-7(Nov.12,2015)doi:10.1038/nature15521.Epub 2015Sep 16.

[0375] ➢Cas13Is a Single RNA-Guided Endonuclease of a Class 2CRISPR-Cas System,Zetsche et al.,Cell 163,759-71(Sep 25,2015).

[0376] ➢Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems,Shmakov et al.,Molecular Cell,60(3),385-397 doi:10.1016/j.molcel.2015.10.008Epub October 22,2015.

[0377] ➢Rationally engineered Cas9nucleases with improved specificity, Slaymaker et al.,Science 2016Jan 1 351(6268):84-88doi:10.1126/science.aad5227.Epub 2015Dec 1.

[0378] ➢Gao et al,"Engineered Cas13Enzymes with Altered PAM Specificities," bioRxiv 091611;doi:http://dx.doi.org/10.1101/091611(Dec.4,2016).

[0379] 上述文献中的每一者通过引用并入本文,可被认为在本发明的实践中,并如下简要讨论:

[0380] ➢Cong等人基于嗜热链球菌(Streptococcus thermophilus)Cas9还有酿脓链球菌Cas9两者改造了II型CRISPR-Cas系统以用于真核细胞中,并且证实了短RNA可以指导Cas9核酸酶在人类和小鼠细胞中诱导DNA的精确裂解。他们的研究进一步表明,Cas9在转化为切口酶时可以用于促进以最小的诱变活性在真核细胞中进行同源定向修复。另外,他们的研究证实了可以将多个指导序列编码到单个CRISPR阵列中,以便能够同时编辑哺乳动物基因组内的若干内源基因组基因座位点,从而证明了RNA引导的核酸酶技术具有易编程性和广泛适用性。这种使用RNA来编程细胞内序列特异性DNA裂解的能力定义了一类新的基因组工程化工具。这些研究进一步表明,其他CRISPR基因座可能可移植到哺乳动物细胞中,并且也可以介导哺乳动物基因组裂解。重要的是,可以设想可以进一步改善CRISPR-Cas系统的若干个方面,以提高其效率和通用性。

[0381] ➢Jiang等人使用与双RNA复合的成簇规律间隔短回文重复序列(CRISPR)相关的Cas9核酸内切酶在肺炎链球菌(Streptococcus pneumoniae)和大肠埃希氏菌的基因组中引入精确的突变。该方法依赖于靶基因组位点处双RNA:Cas9指导的裂解来杀死未突变的细胞,并且避免了对选择性标记或反选择系统的需要。该研究报道了通过改变短CRISPR RNA(crRNA)的序列在编辑模板上进行单核苷酸和多核苷酸改变来重编程双RNA:Cas9特异性。该研究表明,同时使用两种crRNA可以进行多重诱变。此外,当该方法与重组工程结合使用时,在肺炎链球菌中,使用所述方法回收的细胞几乎100%含有所需的突变,并且在大肠埃希氏菌中,回收的细胞65%含有该突变。

[0382] ➢Wang等人(2013)使用CRISPR-Cas系统一步产生了携带多个基因突变的小鼠,这些小鼠传统上是通过在胚胎干细胞中进行连续重组和/或使具有单个突变的小鼠进行耗时杂交来多步产生的。CRISPR-Cas系统将大大加快功能冗余基因和上位基因相互作用的体内研究。

[0383] ➢Konermann等人(2013)解决了本领域对能够基于CRISPR Cas9酶还有转录激活子样效应物对DNA结合结构域进行光调节和化学调节的通用和可靠技术的需要。

[0384] ➢Ran等人(2013-A)描述了一种将Cas9切口酶突变体与配对的引导RNA结合以引入靶向的双链断裂的方法。这解决了以下问题:来自微生物CRISPR-Cas系统的Cas9核酸酶被指导序列靶向特定基因组基因座,这可以耐受与DNA靶标的某些错配,从而促进不希望的脱靶诱变。因为基因组中的各个切口以高保真性被修复,所以经由适当偏移的指导RNA同时形成切口对双链断裂必需的,并且可以延伸特异性识别的碱基的数目以用于标裂解。作者证实了使用配对的切口形成可以使细胞系中的脱靶活性降低50至1,500倍,从而促进小鼠受精卵中的基因敲除而不牺牲中靶裂解效率。这种通用策略可实现要求高特异性的各种基

因组编辑应用。

[0385] ➢Hsu等人(2013)表征了人类细胞中的SpCas9靶向特异性,以告知靶位点的选择并避免脱靶效应。该研究评估了293T和293FT细胞中超过100个预测基因组脱靶基因座处的超过700个指导RNA变体和SpCas9诱导的插入缺失突变水平。作者表明,SpCas9能够以对错配的数量、位置和分布敏感的序列依赖性方式耐受不同位置处指导RNA和靶DNA之间的错配。作者进一步表明,SpCas9介导的裂解不受DNA甲基化的影响,并且可以对SpCas9和指导RNA的剂量进行滴定以最小化脱靶修饰。另外,为了促进哺乳动物基因组工程应用,作者报道提供了一种基于网络的软件工具来指导靶序列的选择和验证以及脱靶分析。

[0386] ➢Ran等人(2013-B)描述了一套用于在哺乳动物细胞中经由非同源末端连接(NHEJ)或同源定向修复(HDR)进行Cas9介导的基因组编辑以及产生经修饰的细胞系以用于下游功能研究的工具。为了最小化脱靶裂解,作者进一步描述了一种使用Cas9切口酶突变体与配对的指导RNA的双切口形成策略。作者提供的方案通过实验得出用于选择靶位点、评估裂解效率和分析脱靶活性的指南。研究表明,从靶设计开始,可在短短的1-2周内完成基因修饰,并且可在2-3周内获得经修饰的克隆细胞系。

[0387] ➢Shalem等人描述了一种在全基因组规模内探询基因功能的新方法。他们的研究表明,利用64,751个独特的指导序列递送靶向18,080个基因的基因组规模CRISPR-Cas9基因敲除(GeCKO)文库,可以在人类细胞中进行阴性和阳性选择筛选。首先,作者展示了使用GeCKO文库来鉴定癌症和多能干细胞中细胞存活所必需的基因。接下来,在黑素瘤模型中,作者筛选了其丢失与对维莫非尼(一种抑制突变蛋白激酶BRAF的治疗剂)的抗性有关的基因。他们的研究表明,排名最高的候选基因包括先前验证的基因NF1和MED12,以及新颖命中基因NF2、CUL3、TADA2B和TADA1。作者观察到靶向相同基因的独立指导RNA之间的高度一致性以及高比率的命中确认,从而证明了用Cas9进行基因组规模筛选的前景。

[0388] ➢Nishimasu等人报道了酿脓链球菌Cas9与sgRNA及其靶DNA复合的晶体结构,分辨率为2.5A°。该结构揭示了由靶标识别和核酸酶叶组成的双叶结构,其将sgRNA:DNAn RNA双链体容纳在它们界面处带正电荷的凹槽中。识别叶对于结合sgRNA和DNA是必不可少的,而核酸酶叶包含HNH和RuvC核酸酶结构域,这两个结构域被正确定位以分别裂解靶DNA的互补链和非互补链。核酸酶叶还包含负责与原间隔邻近基序(PAM)相互作用的羧基末端结构域。这种高分辨率的结构和伴随的功能分析揭示了通过Cas9进行RNA引导的DNA靶向的分子机制,从而为合理设计新的通用基因组编辑技术做好准备。

[0389] ➢Wu等人确定了来自酿脓链球菌的载有单个指导RNA (sgRNA)的无催化活性的Cas9(dCas9)在小鼠胚胎干细胞(mESC)中的经标测的全基因组结合位点。作者表明,所测试的四个sgRNA中的每一个都将dCas9靶向数十个至数千个基因组位点,这些基因组位点通常以sgRNA中5个核苷酸的种子区以及NGG原间隔邻近基序(PAM)为特征。不可接近染色质减少了dCas9与具有匹配种子序列的其他位点的结合;因此70%的脱靶位点与基因相关。作者表明,对用有催化活性的Cas9转染的mESC中的295个dCas9结合位点进行靶向测序仅鉴定出一个高于本底水平突变的位点。作者提出了一种用于Cas9结合和裂解的两态模型,其中种子匹配触发结合,但裂解需要与靶DNA进行广泛配对。

[0390] ➢Platt等人建立了Cre依赖性Cas9敲入小鼠。作者展示了在神经元、免疫细胞和

内皮细胞中使用腺相关病毒(AAV)、慢病毒或颗粒介导的指导RNA递送进行体内和离体基因组编辑。

[0391] ➢Hsu等人(2014)的综述文章总体讨论了CRISPR-Cas9从酸奶到基因组编辑(包括细胞的遗传筛选)的历史。

[0392] ➢Wang等人(2014)涉及一种使用基因组规模的慢病毒单个引导RNA(sgRNA)文库且适用于阳性和阴性选择的汇集功能丧失遗传筛选方法。

[0393] ➢Doench等人创建了覆盖一组六个内源小鼠基因和三个内源人类基因的所有可能靶位点的sgRNA池,并通过抗体染色和流式细胞术定量评估了它们产生其靶基因的无效等位基因的能力。作者表明,对PAM进行优化提高了活性,并且还提供了用于设计sgRNA的在线工具。

[0394] ➢Swiech等人证实了AAV介导的SpCas9基因组编辑可以实现脑中基因功能的反向遗传研究。

[0395] ➢Konermann等人(2015)讨论了在有或没有接头的情况下将多个效应结构域(例如转录激活子、功能和表观基因组调控子)附接在引导序列诸如茎或四环上的适当位置的能力。

[0396] ➢Zetsche等人证实了Cas9酶可以分成两部分,因此可以控制Cas9的激活组装。

[0397] ➢Chen等人通过证实在小鼠中进行全基因组体内CRISPR-Cas9筛选揭示了调控肺转移的基因而涉及多重筛选。

[0398] ➢Ran等人(2015)涉及SaCas9及其编辑基因组的能力,并证明不能根据生化测定推断。

[0399] ➢Shalem等人(2015)描述了无催化活性的Cas9(dCas9)融合体用于综合阻遏(CRISPRi)或激活(CRISPRa)表达的方式,示出了使用Cas9进行基因组规模筛选(包括测定和汇集的筛选)的进展、使基因组座位失活的敲除途径以及调节转录活性的策略。

[0400] ➢Xu等人(2015)评估了在基于CRISPR的筛选中有助于单个指导RNA(sgRNA)效率的DNA序列特征。作者探索了CRISPR-Cas9敲除的效率和裂解位点处的核苷酸偏好。作者还发现,CRISPRi/a的序列偏好与CRISPR-Cas9敲除的序列偏好明显不同。

[0401] ➢Parnas等人(2015)将全基因组汇集CRISPR-Cas9文库引入树突细胞(DC)中,以鉴定控制细菌脂多糖(LPS)诱导肿瘤坏死因子(Tnf)这一过程的基因。对Tlr4信号转导的已知调控子以及先前未知的候选物进行鉴定,并将其分为三个功能模块,这些模块对LPS的规范响应具有不同影响。

[0402] ➢Ramanan等人(2015)证明了感染细胞中存在对病毒游离DNA(cccDNA)的裂解。HBV基因组在感染肝细胞的核中以3.2kb双链游离DNA形式(被称为共价闭状DNA(cccDNA))存在,这种DNA是HBV生命周期中的关键组分,其复制不受当前疗法的抑制。作者表明,特异性靶向HBV高度保守区的sgRNA可以可靠地抑制病毒复制并耗竭cccDNA。

[0403] ➢Nishimasu等人(2015)报道了SaCas9与单个指导RNA(sgRNA)及其双链DNA靶标复合的晶体结构,其中包含5'-TTGAAT-3'PAM和5'-TTGGGT-3'PAM。SaCas9与SpCas9的结构比较突出显示了结构保守性和差异性,解释了它们不同的PAM特异性和直系同源sgRNA识别。

[0404] ➢Canver等人(2015)证明了基于CRISPR-Cas9的非编码基因组元件的功能研究。作者开发了汇集CRISPR-Cas9指导RNA文库,以进行人和小鼠BCL11A增强子的原位饱和诱变,揭示了所述增强子的关键特征。

[0405] ➢Zetsche等人(2015)报道了对Cas13(一种来自新凶手弗朗西斯菌U112的2类CRISPR核酸酶,具有不同于Cas9的特征)的表征。Cas13是一种缺乏tracrRNA的单个RNA指导的内切核酸酶,利用富含T的原间隔区邻近基序,并通过交错的DNA双链断裂来裂解DNA。

[0406] ➢Shmakov等人(2015)报道了三种不同的2类CRISPR-Cas系统。两种系统CRISPR酶(C2c1和C2c3)包含与Cas13远缘相关的RuvC样核酸内切酶结构域。与Cas13不同,C2c1依赖于crRNA和tracrRNA进行DNA裂解。第三种酶(C2c2)包含两个预测的HEPN RNA酶结构域,并且不依赖tracrRNA。

[0407] ➢Slaymaker等人(2016)报道了使用结构指导的蛋白质工程化来提高酿脓链球菌Cas9(SpCas9)的特异性。作者开发了“增强特异性”SpCas9(eSpCas9)变体,所述变体保持稳固的中靶裂解,同时降低了脱靶效应。

[0408] 本文提供的方法和工具以Cas13(一种不使用tracrRNA的II型核酸酶)为例。已经在如本文所述的不同细菌物种中鉴定了Cas13的直系同源物。可以使用本领域中描述的方法来鉴定具有相似性质的其他II型核酸酶(Shmakov等人,2015,60:385-397;Abudayeh等人,2016,Science,5;353(6299))。在特定的实施方案中,用于鉴定新CRISPR效应蛋白的此类方法可包括以下步骤:从数据库中选择编码种子的序列,所述种子鉴定CRISPR Cas基因座的存在;鉴定位于种子的10kb内的基因座,所述基因座在所选序列中包含开放阅读框(ORF);从其中选择包含ORF的基因座,所述ORF中仅单个ORF编码具有大于700个氨基酸且与已知CRISPR效应子具有不超过90%同源性的新CRISPR效应子。在特定的实施方案中,种子是CRISPR-Cas系统共有的蛋白质,诸如Cas1。在另外的实施方案中,CRISPR阵列用作种子以鉴定新的效应蛋白。

[0409] 已经证明了本发明的有效性。可例如通过电穿孔转染包含Cas13和crRNA的预组装的重组CRISPR-Cas13复合物,从而产生高突变率并且没有可检测的脱靶突变。Hur,J.K.等人,Targeted mutagenesis in mice by electroporation of Cas13ribonucleoproteins,Nat Biotechnol.2016年6月6日.doi:10.1038/nbt.3596。全基因组分析表明,Cas13具有高度特异性。通过一种测量,在人类HEK293T细胞中确定的Cas13的体外裂解位点明显少于SpCas9的体外裂解位点。Kim,D.等人,Genome-wide analysis reveals specificities of Cas13endonucleases in human cells,Nat Biotechnol.2016年6月6日.doi:10.1038/nbt.3609。使用从含有本发明tRNA的阵列加工的gRNA,已经在果蝇中证明了使用Cas13的有效多重系统。Port,F.等人,Expansion of the CRISPR toolbox in an animal with tRNA-flanked Cas9and Cas13gRNAs.doi:http://dx.doi.org/10.1101/046417。

[0410] 而且,“Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing”,Shengdar Q.Tsai,Nicolas Wyvekens,Cyd Khayter,Jennifer A.Foden,Vishal Thapar,Deepak Reyon,Mathew J.Goodwin,Martin J.Aryee,J.Keith Joung Nature Biotechnology 32(6):569-77(2014)涉及二聚体RNA引导的FokI核酸酶,该核酸酶识别延伸序列并且可以在人类细胞中高效编辑内源基因。

[0411] 关于CRISPR/Cas系统、其组分以及此类组分的递送的一般信息，包括方法、材料、递送媒介、载体、颗粒、其制备和使用，包括数量和制剂方面，以及表达CRISPR-Cas的真核细胞、表达CRISPR-Cas的真核生物(诸如小鼠)，参考以下文献：美国专利号8,999,641、8,993,233、8,697,359、8,771,945、8,795,965、8,865,406、8,871,445、8,889,356、8,889,418、8,895,308、8,906,616、8,932,814和8,945,839；美国专利公开US 2014-0310830(美国申请序列号14/105,031)、US 2014-0287938 A1(美国申请序列号14/213,991)、US 2014-0273234 A1(美国申请序列号14/293,674)、US2014-0273232 A1(美国申请序列号14/290,575)、US 2014-0273231(美国申请序列号14/259,420)、US 2014-0256046 A1(美国申请序列号14/226,274)、US 2014-0248702 A1(美国申请序列号14/258,458)、US 2014-0242700 A1(美国申请序列号14/222,930)、US 2014-0242699 A1(美国申请序列号14/183,512)、US 2014-0242664 A1(美国申请序列号14/104,990)、US 2014-0234972 A1(美国申请序列号14/183,471)、US 2014-0227787 A1(美国申请序列号14/256,912)、US 2014-0189896 A1(美国申请序列号14/105,035)、US 2014-0186958(美国申请序列号14/105,017)、US 2014-0186919 A1(美国申请序列号14/104,977)、US 2014-0186843 A1(美国申请序列号14/104,900)、US 2014-0179770 A1(美国申请序列号14/104,837)以及US 2014-0179006 A1(美国申请序列号14/183,486)、US 2014-0170753(美国申请序列号14/183,429)；US 2015-0184139(美国申请序列号14/324,960)；14/054,414欧洲专利申请EP 2 771 468(EP13818570.7)、EP 2 764 103(EP13824232.6)和EP 2 784 162(EP14170383.5)；以及PCT专利公布WO2014/093661(PCT/US2013/074743)、WO2014/093694(PCT/US2013/074790)、WO2014/093595(PCT/US2013/074611)、WO2014/093718(PCT/US2013/074825)、WO2014/093709(PCT/US2013/074812)、WO2014/093622(PCT/US2013/074667)、WO2014/093635(PCT/US2013/074691)、WO2014/093655(PCT/US2013/074736)、WO2014/093712(PCT/US2013/074819)、WO2014/093701(PCT/US2013/074800)、WO2014/018423(PCT/US2013/051418)、WO2014/204723(PCT/US2014/041790)、WO2014/204724(PCT/US2014/041800)、WO2014/204725(PCT/US2014/041803)、WO2014/204726(PCT/US2014/041804)、WO2014/204727(PCT/US2014/041806)、WO2014/204728(PCT/US2014/041808)、WO2014/204729(PCT/US2014/041809)、WO2015/089351(PCT/US2014/069897)、WO2015/089354(PCT/US2014/069902)、WO2015/089364(PCT/US2014/069925)、WO2015/089427(PCT/US2014/070068)、WO2015/089462(PCT/US2014/070127)、WO2015/089419(PCT/US2014/070057)、WO2015/089465(PCT/US2014/070135)、WO2015/089486(PCT/US2014/070175)、WO2015/058052(PCT/US2014/061077)、WO2015/070083(PCT/US2014/064663)、WO2015/089354(PCT/US2014/069902)、WO2015/089351(PCT/US2014/069897)、WO2015/089364(PCT/US2014/069925)、WO2015/089427(PCT/US2014/070068)、WO2015/089473(PCT/US2014/070152)、WO2015/089486(PCT/US2014/070175)、WO2016/049258(PCT/US2015/051830)、WO2016/094867(PCT/US2015/065385)、WO2016/094872(PCT/US2015/065393)、WO2016/094874(PCT/US2015/065396)、WO2016/106244(PCT/US2015/067177)。

[0412] 还提到了提交于2015年6月17日的美国申请62/180,709，“PROTECTED GUIDE RNAS(PGRNAS)”；提交于2014年12月12日的美国申请62/091,455，“PROTECTED GUIDE RNAS(PGRNAS)”；提交于2014年12月24日的美国申请62/096,708，“PROTECTED GUIDE RNAS

(PGRNAS)”;提交于2014年12月12日的美国申请62/091,462、提交于2014年12月23日的美国申请62/096,324、提交于2015年6月17日的美国申请62/180,681和提交于2015年10月5日的美国申请62/237,496,“DEAD GUIDES FOR CRISPR TRANSCRIPTION FACTORS”;提交于2014年12月12日的美国申请62/091,456和提交于2015年6月17日的美国申请62/180,692,“ESCORTED AND FUNCTIONALIZED GUIDES FOR CRISPR-CAS SYSTEMS”;提交于2014年12月12日的美国申请62/091,461,“DELIVERY,USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS FOR GENOME EDITING AS TO HEMATOPOETIC STEM CELLS(HSCs)”;提交于2014年12月19日的美国申请62/094,903,“UNBIASED IDENTIFICATION OF DOUBLE-STRAND BREAKS AND GENOMIC REARRANGEMENT BY GENOME-WISE INSERT CAPTURE SEQUENCING”;提交于2014年12月24日的美国申请62/096,761,“ENGINEERING OF SYSTEMS,METHODS AND OPTIMIZED ENZYME AND GUIDE SCAFFOLDS FOR SEQUENCE MANIPULATION”;提交于2014年12月30日的美国申请62/098,059、提交于2015年6月18日的美国申请62/181,641和提交于2015年6月18日的美国申请62/181,667,“RNA-TARGETING SYSTEM”;提交于2014年12月24日的美国申请62/096,656和提交于2015年6月17日的美国申请62/181,151,“CRISPR HAVING OR ASSOCIATED WITH DESTABILIZATION DOMAINS”;提交于2014年12月24日的美国申请62/096,697,“CRISPR HAVING OR ASSOCIATED WITH AAV”;提交于2014年12月30日的美国申请62/098,158,“ENGINEERED CRISPR COMPLEX INSERTIONAL TARGETING SYSTEMS”;提交于2015年4月22日的美国申请62/151,052,“CELLULAR TARGETING FOR EXTRACELLULAR EXOSOMAL REPORTING”;提交于2014年9月24日的美国申请62/054,490,“DELIVERY,USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS FOR TARGETING DISORDERS AND DISEASES USING PARTICLE DELIVERY COMPONENTS”;提交于2014年2月12日的美国申请61/939,154,“SYSTEMS,METHODS AND COMPOSITIONS FOR SEQUENCE MANIPULATION WITH OPTIMIZED FUNCTIONAL CRISPR-CAS SYSTEMS”;提交于2014年9月25日的美国申请62/055,484,“SYSTEMS,METHODS AND COMPOSITIONS FOR SEQUENCE MANIPULATION WITH OPTIMIZED FUNCTIONAL CRISPR-CAS SYSTEMS”;提交于2014年12月4日的美国申请62/087,537,“SYSTEMS,METHODS AND COMPOSITIONS FOR SEQUENCE MANIPULATION WITH OPTIMIZED FUNCTIONAL CRISPR-CAS SYSTEMS”;提交于2014年9月24日的美国申请62/054,651,“DELIVERY,USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS FOR MODELING COMPETITION OF MULTIPLE CANCER MUTATIONS IN VIVO”;提交于2014年10月23日的美国申请62/067,886,“DELIVERY,USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS FOR MODELING COMPETITION OF MULTIPLE CANCER MUTATIONS IN VIVO”;提交于2014年9月24日的美国申请62/054,675和提交于2015年6月17日的美国申请62/181,002,“DELIVERY,USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS IN NEURONAL CELLS/TISSUES”;提交于2014年9月24日的美国申请62/054,528,“DELIVERY,USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS IN IMMUNE DISEASES OR DISORDERS”;提交于2014年9月25日的美国申请62/055,454,“DELIVERY,USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND

COMPOSITIONS FOR TARGETING DISORDERS AND DISEASES USING CELL PENETRATION PEPTIDES(CPP)”;提交于2014年9月25日的美国申请62/055,460,“MULTIFUNCTIONAL-CRISPR COMPLEXES AND/OR OPTIMIZED ENZYME LINKED FUNCTIONAL-CRISPR COMPLEXES”;提交于2014年12月4日的美国申请62/087,475和提交于2015年6月18日的美国申请62/181,690,“FUNCTIONAL SCREENING WITH OPTIMIZED FUNCTIONAL CRISPR-CAS SYSTEMS”;提交于2014/9/25的美国申请62/055,487,“FUNCTIONAL SCREENING WITH OPTIMIZED FUNCTIONAL CRISPR-CAS SYSTEMS”;提交于2014年12月4日的美国申请62/087,546和提交于2015年6月18日的美国申请62/181,687,“MULTIFUNCTIONAL CRISPR COMPLEXES AND/OR OPTIMIZED ENZYME LINKED FUNCTIONAL-CRISPR COMPLEXES”;以及提交于2014年12月30日的美国申请62/098,285,“CRISPR MEDIATED IN VIVO MODELING AND GENETIC SCREENING OF TUMOR GROWTH AND METASTASIS”。

[0413] 提到了提交于2015年6月18日的美国申请62/181,659和提交于2015年8月19日的美国申请62/207,318,“ENGINEERING AND OPTIMIZATION OF SYSTEMS,METHODS,ENZYME AND GUIDE SCAFFOLDS OF CAS9 ORTHOLOGS AND VARIANTS FOR SEQUENCE MANIPULATION”。提到了提交于2015年6月18日的美国申请62/181,663和提交于2015年10月22日的美国申请62/245,264,“NOVEL CRISPR ENZYMES AND SYSTEMS”;提交于2015年6月18日的美国申请62/181,675、提交于2015年10月22日的美国申请62/285,349、提交于2016年2月17日的美国申请62/296,522和提交于2016年4月8日的美国申请62/320,231,“NOVEL CRISPR ENZYMES AND SYSTEMS”;提交于2015年9月24日的美国申请62/232,067、提交于2015年12月18日的美国申请14/975,085、欧洲申请号16150428.7、提交于2015年8月16日的美国申请62/205,733、提交于2015年8月5日的美国申请62/201,542、提交于2015年7月16日美国申请62/193,507和提交于2015年6月18日的美国申请62/181,739,各自的标题都为“NOVEL CRISPR ENZYMES AND SYSTEMS”;以及提交于2015年10月22日的美国申请62/245,270,“NOVEL CRISPR ENZYMES AND SYSTEMS”。还提到了提交于2014年2月12日的美国申请61/939,256和提交于2014年12月12日的WO 2015/089473(PCT/US2014/070152),各自的标题都为“ENGINEERING OF SYSTEMS,METHODS AND OPTIMIZED GUIDE COMPOSITIONS WITH NEW ARCHITECTURES FOR SEQUENCE MANIPULATION”。还提到了提交于2015年8月15日的PCT/US2015/045504、提交于2015年6月17日的美国申请62/180,699和提交于2014年8月17日的美国申请62/038,358,各自的标题都为“GENOME EDITING USING CAS9NICKASES”。

[0414] 这些专利、专利公开和申请中的每一者,以及在其中或在其审查期间引用的所有文献(“申请引用文献”)以及在这些申请引用文献中引用或参考的所有文献,连同针对在其中提及或其中通过引用并入本文的任何文献中的任何产品的任何说明书、说明、产品规格和产品表据此通过引用并入本文,并且可在本发明的实践中采用。所有文献(例如,这些专利、专利公开和申请以及申请引用文献)均通过引用并入本文,其程度如同每个单独的文献被确切且单独地指明通过引用并入。

[0415] V型CRISPR-Cas蛋白

[0416] 本申请描述了使用V型CRISPR-Cas蛋白的方法。本文以Cas13为例,由此鉴定了许多直系同源物或同系物。对于本领域技术人员显而易见的是,可以鉴定其他直系同源物或同系物,并且可将本文所述的任何功能性工程化到其他直系同源物中,包括包含来自多个

直系同源物的片段的嵌合酶。

[0417] 鉴定新CRISPR-Cas基因座的计算方法在EP3009511或US2016208243中描述，并且可包括以下步骤：检测编码Cas1蛋白的所有重叠群；鉴定cas1基因20kB内所有预测的蛋白编码基因；将鉴定出的基因与Cas蛋白特异性谱进行比较，并预测CRISPR阵列；选择含有大于500个氨基酸(>500aa)的蛋白质的未分类候选CRISPR-Cas基因座；使用诸如PSI-BLAST和HHPred的方法对选定候选基因进行分析，以筛选已知的蛋白质结构域，从而鉴定新2类CRISPR-Cas基因座(还可参见Schmakov等人，2015，Mol Cell.60(3):385-97)。除上述步骤外，还可通过在宏基因组学数据库中搜索其他同系物来进行候选物的其他分析。另外地或另选地，为了将搜索扩展到非自主CRISPR-Cas系统，可以利用CRISPR阵列作为种子来执行相同的程序。

[0418] 一方面，通过基因预测程序GenemarkS检测编码Cas1蛋白的所有重叠群，所述GenemarkS在以下文献中进一步描述："GeneMarkS:a self-training method for prediction of gene starts in microbial genomes.Implications for finding sequence motifs in regulatory regions."John Besemer,Alexandre Lomsadze和Mark Borodovsky,Nucleic Acids Research(2001)29,第2607-2618页，该文献通过引用并入本文。

[0419] 一方面，通过将鉴定出的基因与Cas蛋白特异性谱进行比较并根据NCBI保守结构域数据库(CDD)对它们进行注释来鉴定所有预测的蛋白编码基因，所述NCBI保守结构域数据库是由古结构域和全长蛋白质的充分注释的多个序列比对模型的集合组成的蛋白质注释资源。这些可作为位置特异性评分矩阵(PSSM)用于通过RPS-BLAST快速鉴定蛋白质序列中的保守结构域。CDD内容包括NCBI展出的域，这些域使用3D结构信息来明确定义域边界并提供对序列/结构/功能关系以及从多个外部源数据库(Pfam、SMART、COG、PRK、TIGRFAM)输入的域模型的深刻理解。另一方面，使用PILER-CR程序来预测CRISPR阵列，PILER-CR程序是用于发现CRISPR重复序列的公共域软件，如"PILER-CR:fast and accurate identification of CRISPR repeats",Edgar,R.C.,BMC Bioinformatics,1月20日;8:18(2007)中所述，该文献通过引用并入本文。

[0420] 另一方面，使用PSI-BLAST(位置特定性迭代基本局部比对搜索工具)来进行逐例分析。PSI-BLAST使用蛋白质-蛋白质BLAST根据在给定评分阈值以上检测到的序列的多序列比对得出位置特异性评分矩阵(PSSM)或谱。该PSSM用于进一步在数据库中搜索新的匹配，并针对这些新检测到的序列进行后续迭代更新。因此，PSI-BLAST提供了一种检测蛋白质之间的远距离关系的手段。

[0421] 另一方面，使用HHpred进行逐例分析，HHpred是一种用于序列数据库搜索和结构预测的方法，该方法与BLAST或PSI-BLAST一样容易使用，同时在寻找远程同系物时更加灵敏。实际上，HHpred的灵敏度可与目前可用的功能最强大的结构预测服务器具有竞争力。HHpred是第一个基于谱隐马尔可夫模型(HMM)的成对比较的服务器。大多数常规的序列搜索方法都搜索序列数据库，诸如UniProt或NR，而HHpred则搜索比对数据库，如Pfam或SMART。这极大地简化了多个序列族的命中列表，而不是杂乱的单个序列。所有主要的公开可用谱和比对数据库都可通过HHpred获得。HHpred接受单个查询序列或多个比对作为输入。在短短几分钟内，它以类似于PSI-BLAST的易于阅读的格式返回搜索结果。搜索选项包

括局部或全局比对以及对二级结构相似性评分。HHpred可以产生成对的查询模板序列比对、合并的查询模板多个比对(例如,针对传递搜索)以及MODELLER软件根据HHpred比对计算出的3D结构模型。

[0422] Cas13的直系同源物

[0423] 术语“同源物”(在本文中也称为“直系同源物”)和“同系物”在本领域中是众所周知的。通过进一步指导,本文所用的蛋白质的“同系物”是同一物种的蛋白质,其执行与该蛋白质的同系物相同或相似的功能。同系蛋白质可以但不必在结构上相关,或仅在部分结构上相关。本文所用的蛋白质的“同源物”是不同物种的蛋白质,其执行与该蛋白质的同源物相同或相似的功能。同源蛋白质可以但不必在结构上相关,或仅在部分结构上相关。同系物和直系同源物可通过同源建模(参见例如Greer,Science,第228卷(1985)1055以及Blundell等人,Eur J Biochem,第172卷(1988),513)或“结构BLAST”(Dey F,Cliff Zhang Q,Petrey D,Honig B.Toward a″structural BLAST″:using structural relationships to infer function.Protein Sci.2013年4月;22(4):359-66.doi:10.1002/pro.2225.)来鉴定。还可参见Shmakov等人(2015)以了解在CRISPR-Cas基因座领域的应用。同系蛋白质可以但不必在结构上相关,或仅在部分结构上相关。

[0424] Cas13基因存在于几个不同的细菌基因组中,通常与cas1、cas2和cas4基因以及CRISPR盒位于同一基因座中(例如,新凶手西班牙弗朗西斯菌(Francisella cf.novicida)Fx1的FNFX1_1431-FNFX1_1428)。因此,该推定的新CRISPR-Cas系统的布局似乎与II-B型的布局相似。此外,类似于Cas9,Cas13蛋白包含与转座子ORF-B同源的易于鉴定的C末端区域,并且包括活性RuvC样核酸酶、富含精氨酸的区域和Zn指(在Cas9中不存在)。然而,与Cas9不同,Cas13也存在于没有CRISPR-Cas环境的几个基因组中,并且其与ORF-B具有相对较高的相似性表明它可能是转座子组分。有人提出,如果这是真正的CRISPR-Cas系统并且Cas13是Cas9的功能类似物,则其将是一种新CRISPR-Cas类型,即V型(参见Annotation and Classification of CRISPR-Cas Systems.Makarova KS,Koonin EV.Methods Mol Biol.2015;1311:47-75)。然而,如本文所述,Cas13被表示为亚型V-A,以使其与不具有相同结构域结构的C2c1p区分开,因此被表示为亚型V-B。

[0425] 本发明包括来源于被表示为亚型V-A的Cas13基因座的Cas13效应蛋白的用途。在本文中,此类效应蛋白也被称为“Cas13p”,例如Cas13蛋白(并且这种效应蛋白或Cas13蛋白或来源于Cas13基因座的蛋白也被称为“CRISPR-Cas蛋白”)。

[0426] 在特定的实施方案中,效应蛋白是来自包括以下各项的属的生物的Cas13蛋白:链球菌属(Streptococcus)、弯曲杆菌属(Campylobacter)、硝酸盐裂解菌属(Nitratifractor)、葡萄球菌属(Staphylococcus)、细小棒菌属(Parvibaculum)、罗斯氏菌属(Roseburia)、奈瑟氏菌属(Neisseria)、葡糖醋杆菌属(Gluconacetobacter)、固氮螺菌属(Azospirillum)、球毛菌属(Sphaerochaeta)、乳杆菌属(Lactobacillus)、真细菌属(Eubacterium)、棒状杆菌属(Corynebacter)、肉食杆菌属(Carnobacterium)、红细菌属(Rhodobacter)、李斯特菌属(Listeria)、沼杆菌属(Paludibacter)、梭菌属(Clostridium)、毛螺菌属(Lachnospiraceae)、Clostridiaridium、纤毛菌属(Leptotrichia)、弗朗西斯菌属(Francisella)、军团杆菌属(Legionella)、脂环酸芽孢杆菌属(Alicyclobacillus)、甲烷嗜甲基菌属(Methanomethyophilus)、叶啉单胞菌属

(Porphyromonas)、普雷沃菌属(Prevotella)、拟杆菌属(Bacteroidetes)、创伤球菌属(Helcococcus)、钩端螺旋体(Leptospira)、脱硫弧菌属(Desulfovibrio)、脱硫嗜盐碱菌属(Desulfonatronum)、丰佑菌属(Opitutaceae)、肿块芽孢杆菌属(Tuberibacillus)、芽孢杆菌属(Bacillus)、短芽孢杆菌属(Brevibacilus)、甲基杆菌属(Methylobacterium)、丁酸弧菌属(Butyvibrio)、异域菌属(Perigrinibacterium)、帕库菌属(Pareubacterium)、莫拉氏菌属(Moraxella)、硫微螺菌属(Thiomicrospira)或氨基酸球菌属(Acidaminococcus)。在特定的实施方案中,Cas13效应蛋白选自选自以下各项的属的生物:真细菌属、毛螺菌属、纤毛菌属、弗朗西斯菌属、甲烷嗜甲基菌属、叶啉单胞菌属、普雷沃菌属、钩端螺旋体、丁酸弧菌属、Perigrinibacterium、Pareubacterium、莫拉氏菌属、硫微螺菌属或氨基酸球菌属。

[0427] 在另外的特定实施方案中,Cas13效应蛋白来自选自以下各项的生物:变形链球菌(S.mutans)、无乳链球菌(S.agalactiae)、马链球菌(S.equisimilis)、血链球菌(S.sanguinis)、肺炎链球菌(S.pneumonia);空肠弯曲杆菌(C.jejuni)、大肠弯曲杆菌(C.coli);卤水硝酸盐裂解菌(N.salsuginis)、拱背硝酸盐裂解菌(N.tergarcus);耳葡萄球菌(S.auricularis)、肉葡萄球菌(S.carnosus);脑膜炎奈瑟氏菌(N.meningitides)、淋病奈瑟氏菌(N.gonorrhoeae);单核细胞增生李斯特菌(L.monocytogenes)、伊氏李斯特菌(L.ivanovii);肉毒梭菌(C.botulinum)、艰难梭菌(C.difficile)、破伤风梭菌(C.tetani)、索德氏梭菌(C.sordellii)、稻田钩端螺旋体(L inadai)、土拉弗朗西斯菌(F.tularensis)1、苏格兰普雷沃菌(P.albensis)、毛螺菌科细菌(L.bacterium)、瘤胃溶纤维丁酸弧菌(B.proteoclasticus)、异域菌(P.bacterium)、狗口腔叶啉单胞菌(P.crevioricanis)、解糖胨普雷沃菌(P.disiens)和猕猴叶啉单胞菌(P.macacae)。

[0428] 效应蛋白可包含嵌合效应蛋白,所述嵌合效应蛋白包含来自第一效应蛋白(例如,Cas13)直系同源物的第一片段和来自第二效应蛋白(例如,Cas13)直系同源物的第二片段,并且其中第一效应蛋白直系同源物和第二效应蛋白直系同源物不同。第一效应蛋白(例如,Cas13)直系同源物和第二效应蛋白直系同源物中的至少一者可包含来自包括以下各项的生物体的效应蛋白(例如,Cas13):链球菌属、弯曲杆菌属、硝化裂化器菌属、葡萄球菌属、细小棒菌属、罗斯氏菌属、奈瑟氏菌属、葡糖醋杆菌属、固氮螺菌属、球毛菌属、乳杆菌属、真细菌属、棒状杆菌属、肉食杆菌属、红细菌属、李斯特菌属、沼杆菌属、梭菌属、毛螺菌属、Clostridiaridium、纤毛菌属、弗朗西斯菌属、军团杆菌属、脂环酸芽孢杆菌属、甲烷嗜甲基菌属、叶啉单胞菌属、普雷沃菌属、拟杆菌属、创伤球菌属、钩端螺旋体菌属、脱硫弧菌属、脱硫嗜盐碱菌属、丰佑菌属、肿块芽孢杆菌属、芽孢杆菌属、短芽孢杆菌属、甲基杆菌属、丁酸弧菌属、异域菌属、帕库菌属、莫拉氏菌属、硫微螺菌属或氨基酸球菌属;例如,包含第一片段和第二片段的嵌合效应蛋白,其中第一片段和第二片段中的每一者选自包括以下各项的生物体的Cas13:链球菌属、弯曲杆菌属、硝化裂化器菌属、葡萄球菌属、细小棒菌属、罗斯氏菌属、奈瑟氏菌属、葡糖醋杆菌属、固氮螺菌属、球毛菌属、乳杆菌属、真细菌属、棒状杆菌属、肉食杆菌属、红细菌属、李斯特菌属、沼杆菌属、梭菌属、毛螺菌属、Clostridiaridium、纤毛菌属、弗朗西斯菌属、军团杆菌属、脂环酸芽孢杆菌属、甲烷嗜甲基菌属、叶啉单胞菌属、普雷沃菌属、拟杆菌属、创伤球菌属、钩端螺旋体菌属、脱硫弧菌属、脱硫嗜盐碱菌属、丰佑菌属、肿块芽孢杆菌属、芽孢杆菌属、短芽孢杆菌属、甲基杆菌属、丁酸弧菌属、异域菌属、帕库菌属、莫拉氏菌属、硫微螺菌属或氨基酸球菌属,其中第一片段和第二片段不是来自同

一细菌;例如包含第一片段和第二片段的嵌合效应蛋白,其中第一片段和第二片段的每一者选自以下各项的Cas13:变形链球菌、无乳链球菌、马链球菌、血链球菌、肺炎链球菌;空肠弯曲杆菌、大肠弯曲杆菌;卤水硝酸盐裂解菌、拱背硝酸盐裂解菌;耳葡萄球菌、肉葡萄球菌;脑膜炎奈瑟氏菌、淋病奈瑟氏菌;单核细胞增生李斯特菌、伊氏李斯特菌;肉毒梭菌、艰难梭菌、破伤风杆菌、索德氏梭菌;土拉弗朗西斯菌1、苏格兰普雷沃菌、毛螺菌科细菌MC2017 1、瘤胃溶纤维丁酸弧菌、异域菌GW2011_GWA2_33_10、帕库菌GW2011_GWC2_44_17、史密斯菌属某种SCADC、氨基酸球菌属某种BV3L6、毛螺菌科细菌MA2020、候选白蚁甲烷支原体、挑剔真杆菌、牛莫拉氏菌237、稻田钩端螺旋体、毛螺菌科细菌ND2006、狗口腔卟啉单胞菌3、解糖胨普雷沃菌和猕猴卟啉单胞菌,其中第一片段和第二片段不是来自同一细菌。

[0429] 在一个更优选的实施方案中,Cas13p来源于选自以下各项的细菌种:土拉弗朗西斯菌1、苏格兰普雷沃菌、毛螺菌科细菌MC2017 1、瘤胃溶纤维丁酸弧菌、异域菌GW2011_GWA2_33_10、帕库菌GW2011_GWC2_44_17、史密斯菌属某种SCADC、氨基酸球菌属某种BV3L6、毛螺菌科细菌MA2020、候选白蚁甲烷支原体、挑剔真杆菌、牛莫拉氏菌237、牛莫拉氏菌AAX08_00205、牛莫拉氏菌AAX11_00205、丁酸弧菌属某种NC3005、硫微螺菌属某种XS5、稻田钩端螺旋体、毛螺菌科细菌ND2006、狗口腔卟啉单胞菌3、解糖胨普雷沃菌和猕猴卟啉单胞菌。在某些实施方案中,Cas13p来源于选自氨基酸球菌属BV3L6、毛螺菌科细菌MA2020的细菌种。在某些实施方案中,效应蛋白来源于土拉弗朗西斯菌1的亚种,包括但不限于土拉弗朗西斯菌新凶手亚种。在某些优选的实施方案中,Cas13p来源于选自以下各项的细菌种:氨基酸球菌属某种BV3L6、毛螺菌科细菌ND2006、毛螺菌科细菌MA2020、牛莫拉氏菌AAX08_00205、牛莫拉氏菌AAX11_00205、丁酸弧菌属某种NC3005或硫微螺菌属某种XS5。

[0430] 在特定的实施方案中,本文所提及的Cas13的同系物或直系同源物与本文公开的示例Cas13蛋白具有至少80%、更优选至少85%、甚至更优选至少90%、诸如至少95%的序列同源性或同一性。在另外的实施方案中,本文所提及的Cas13的同系物或直系同源物与野生型Cas13具有至少80%、更优选至少85%、甚至更优选至少90%、诸如至少95%的序列同一性。当Cas13具有一个或多个突变时,本文所述的Cas13的同系物或直系同源物与突变Cas13具有至少80%、更优选至少85%、甚至更优选至少90%、诸如至少95%的序列同一性。

[0431] 在一个实施方案中,Cas13蛋白可以是包括但不限于以下各项的属的生物的直系同源物:氨基酸球菌属某种、毛螺菌科细菌或牛莫拉氏菌;在特定的实施方案中,V型Cas蛋白可以是包括但不限于以下各项的种的生物的直系同源物:氨基酸球菌属某种BV3L6、毛螺菌科细菌ND2006(LbCas13)或牛莫拉氏菌237。在特定的实施方案中,本文所提及的Cas13的同系物或直系同源物与本文公开的一个或多个Cas13序列具有至少80%、更优选至少85%、甚至更优选至少90%、诸如至少95%的序列同源性或同一性。在另外的实施方案中,本文所提及的Cas13的同系物或直系同源物与野生型FnCas13、AsCas13或LbCas13具有至少80%、更优选至少85%、甚至更优选至少90%、诸如至少95%的序列同一性。

[0432] 在特定的实施方案中,本发明的Cas13蛋白与FnCas13、AsCas13或LbCas13具有至少60%、更具体至少70%、诸如至少80%、更优选至少85%、甚至更优选至少90%、诸如至少有95%的序列同源性或同一性。在另外的实施方案中,本文所提及的Cas13蛋白与野生型AsCas13或LbCas13具有至少60%、诸如至少70%、更具体至少80%、更优选至少85%、甚至更优选至少90%、诸如至少95%的序列同一性。在特定的实施方案中,本发明的Cas13蛋白

与FnCas13具有小于60％的序列同一性。技术人员将理解，这包括Cas13蛋白的截短形式，由此在截短形式的长度上确定序列同一性。在特定的实施方案中，Cas13酶不是FnCas13。

[0433] 经修饰的Cas13酶

[0434] 在特定的实施方案中，利用本文定义的工程化Cas13蛋白(诸如Cas13)是有意义的，其中所述蛋白与包含RNA的核酸分子复合以形成CRISPR复合物，其中当在CRISPR复合物中时，所述核酸酸分子靶向一个或多个靶多核苷酸基因座，与未修饰的Cas13蛋白相比，所述蛋白包含至少一种修饰，并且其中与包含未修饰的Cas13蛋白的复合物相比，包含经修饰的蛋白的CRISPR复合物的活性改变。应当理解，当在本文中提及CRISPR“蛋白”时，Cas13蛋白优选地是经修饰的CRISPR-Cas蛋白(例如，具有增加或降低的(或者没有)酶活性)，诸如包括但不限于Cas13。术语“CRISPR蛋白”可与“CRISPR-Cas蛋白”互换使用，而不管与野生型CRISPR蛋白相比，CRISPR蛋白是否已经改变，诸如增加或降低(或没有)酶活性。

[0435] Cas13核酸酶的一级结构的计算分析揭示了三个不同的区域。首先是C末端RuvC样结构域，这是唯一具有功能特征的结构域。其次是N末端α-螺旋区，再次是位于RuvC样结构域和α-螺旋区之间的混合的α和β区。

[0436] 预测了Cas13一级结构内的几小段非结构化区域。暴露于溶剂中且在不同Cas13直系同源物中不保守的非结构化区域是小蛋白序列断裂和插入的优选侧。另外，这些侧可以用于在Cas13直系同源物之间产生嵌合蛋白。

[0437] 基于上述信息，可以产生导致酶失活或将双链核酸酶修饰成切口酶活性的突变体。在替代实施方案中，该信息用于开发脱靶效应减少的酶(在本文其他地方描述)。

[0438] 在某些上述Cas13酶中，酶通过一个或多个残基(在RuvC结构域中)的突变来修饰，参考AsCas13(氨基酸球菌属BV3L6)的氨基酸位置编号，所述残基包括但不限于位置R909、R912、R930、R947、K949、R951、R955、K965、K968、K1000、K1002、R1003、K1009、K1017、K1022、K1029、K1035、K1054、K1072、K1086、R1094、K1095、K1109、K1118、K1142、K1150、K1158、K1159、R1220、R1226、R1242和/或R1252。在某些实施方案中，包含所述一个或多个突变的Cas13酶已经修饰，更优选地具有增加的靶标特异性。

[0439] 在某些上述非天然存在的CRISPR-Cas蛋白中，酶通过一个或多个残基(在RAD50中)的突变来修饰，参考AsCas13(氨基酸球菌属BV3L6)的氨基酸位置编号，所述残基包括但不限于位置K324、K335、K337、R331、K369、K370、R386、R392、R393、K400、K404、K406、K408、K414、K429、K436、K438、K459、K460、K464、R670、K675、R681、K686、K689、R699、K705、R725、K729、K739、K748和/或K752。在某些实施方案中，包含所述一个或多个突变的Cas13酶已经修饰，更优选地具有增加的靶标特异性。

[0440] 在某些Cas13酶中，酶通过一个或多个残基的突变来修饰，参考AsCas13(氨基酸球菌属某种BV3L6)的氨基酸位置编号，所述残基包括但不限于位置R912、T923、R947、K949、R951、R955、K965、K968、K1000、R1003、K1009、K1017、K1022、K1029、K1072、K1086、F1103、R1226和/或R1252。在某些实施方案中，包含所述一个或多个突变的Cas13酶已经修饰，更优选地具有增加的靶标特异性。

[0441] 在某些实施方案中，Cas13酶通过一个或多个残基的突变来修饰，参考LbCas13(毛螺菌科细菌ND2006)的氨基酸位置编号，所述残基包括但不限于位置R833、R836、K847、K879、K881、R883、R887、K897、K900、K932、R935、K940、K948、K953、K960、K984、K1003、K1017、

R1033、R1138、R1165和/或R1252。在某些实施方案中，包含所述一个或多个突变的Cas13酶已经修饰，更优选地具有增加的靶标特异性。

[0442] 在某些实施方案中，Cas13酶通过一个或多个残基的突变来修饰，参考AsCas13(氨基酸球菌属某种BV3L6)的氨基酸位置编号，所述残基包括但不限于位置K15、R18、K26、Q34、R43、K48、K51、R56、R84、K85、K87、N93、R103、N104、T118、K123、K134、R176、K177、R192、K200、K226、K273、K275、T291、R301、K307、K369、S404、V409、K414、K436、K438、K468、D482、K516、R518、K524、K530、K532、K548、K559、K570、R574、K592、D596、K603、K607、K613、C647、R681、K686、H720、K739、K748、K757、T766、K780、R790、P791、K796、K809、K815、T816、K860、R862、R863、K868、K897、R909、R912、T923、R947、K949、R951、R955、K965、K968、K1000、R1003、K1009、K1017、K1022、K1029、A1053、K1072、K1086、F1103、S1209、R1226、R1252、K1273、K1282和/或K1288。在某些实施方案中，包含所述一个或多个突变的Cas13酶已经修饰，更优选地具有增加的靶标特异性。

[0443] 在某些实施方案中，酶通过一个或多个残基的突变来修饰，参考FnCas13(新凶手弗朗西斯菌U112)的氨基酸位置编号，所述残基包括但不限于位置K15、R18、K26、R34、R43、K48、K51、K56、K87、K88、D90、K96、K106、K107、K120、Q125、K143、R186、K187、R202、K210、K235、K296、K298、K314、K320、K326、K397、K444、K449、E454、A483、E491、K527、K541、K581、R583、K589、K595、K597、K613、K624、K635、K639、K656、K660、K667、K671、K677、K719、K725、K730、K763、K782、K791、R800、K809、K823、R833、K834、K839、K852、K858、K859、K869、K871、R872、K877、K905、R918、R921、K932、I960、K962、R964、R968、K978、K981、K1013、R1016、K1021、K1029、K1034、K1041、K1065、K1084和/或K1098。在某些实施方案中，包含所述一个或多个突变的Cas13酶已经修饰，更优选地具有增加的靶标特异性。

[0444] 在某些实施方案中，酶通过一个或多个残基的突变修饰，参考LbCas13(毛螺菌科细菌ND2006)的氨基酸位置编号，所述残基包括但不限于位置K15、R18、K26、K34、R43、K48、K51、R56、K83、K84、R86、K92、R102、K103、K116、K121、R158、E159、R174、R182、K206、K251、K253、K269、K271、K278、P342、K380、R385、K390、K415、K421、K457、K471、A506、R508、K514、K520、K522、K538、Y548、K560、K564、K580、K584、K591、K595、K601、K634、K640、R645、K679、K689、K707、T716、K725、R737、R747、R748、K753、K768、K774、K775、K785、K787、R788、Q793、K821、R833、R836、K847、K879、K881、R883、R887、K897、K900、K932、R935、K940、K948、K953、K960、K984、K1003、K1017、R1033、K1121、R1138、R1165、K1190、K1199和/或K1208。在某些实施方案中，包含所述一个或多个突变的Cas13酶已经修饰，更优选地具有增加的靶标特异性。

[0445] 在某些实施方案中，酶通过一个或多个残基的突变来修饰，参考MbCas13(牛莫拉氏菌237)的氨基酸位置编号，所述残基包括但不限于位置K14、R17、R25、K33、M42、Q47、K50、D55、K85、N86、K88、K94、R104、K105、K118、K123、K131、R174、K175、R190、R198、I221、K267、Q269、K285、K291、K297、K357、K403、K409、K414、K448、K460、K501、K515、K550、R552、K558、K564、K566、K582、K593、K604、K608、K623、K627、K633、K637、E643、K780、Y787、K792、K830、Q846、K858、K867、K876、K890、R900、K901、M906、K921、K927、K928、K937、K939、R940、K945、Q975、R987、R990、K1001、R1034、I1036、R1038、R1042、K1052、K1055、K1087、R1090、K1095、N1103、K1108、K1115、K1139、K1158、R1172、K1188、K1276、R1293、A1319、K1340、K1349和/或

K1356。在某些实施方案中，包含所述一个或多个突变的Cas13酶已经修饰，更优选地具有增加的靶标特异性。

[0446] 在一个实施方案中，参考AsCas13的氨基酸位置编号，Cas13蛋白用S1228处的突变(例如，S1228A)来修饰。参见Yamano等人，Cell 165:949-962(2016)，该文献全文通过引用并入本文。

[0447] 在某些实施方案中，Cas13蛋白已经修饰以识别非天然PAM，诸如识别具有序列或包含序列YCN、YCV、AYV、TYV、RYN、RCN、TGYV、NTTN、TTN、TRTN、TYTV、TYCT、TYCN、TRTN、NTTN、TACT、TYCC、TRTC、TATV、NTTV、TTV、TSTG、TVTS、TYYS、TCYS、TBYS、TCYS、TNYS、TYYS、TNTN、TSTG、TTCC、TCCC、TATC、TGTG、TCTG、TYCV或TCTC的PAM。在特定的实施方案中，所述突变的Cas13在AsCas13的位置11、12、13、14、15、16、17、34、36、39、40、43、46、47、50、54、57、58、111、126、127、128、129、130、131、132、133、134、135、136、157、158、159、160、161、162、163、164、165、166、167、168、169、170、171、172、173、174、175、176、177、178、532、533、534、535、536、537、538、539、540、541、542、543、544、545、546、547、548、549、550、551、552、553、554、555、556、565、566、567、568、569、570、571、572、573、574、575、592、593、594、595、596、597、598、599、600、601、602、603、604、605、606、607、608、609、610、611、612、613、614、615、616、617、618、619、620、626、627、628、629、630、631、632、633、634、635、636、637、638、642、643、644、645、646、647、648、649、651、652、653、654、655、656、676、679、680、682、683、684、685、686、687、688、689、690、691、692、693、707、711、714、715、716、717、718、719、720、721、722、739、765、768、769、773、777、778、779、780、781、782、783、784、785、786、871、872、873、874、875、876、877、878、879、880、881、882、883、884或1048处或者在Cas13直系同源物的与其对应的位置处包含一个或多个突变的氨基酸残基；优选地，在位置130、131、132、133、134、135、136、162、163、164、165、166、167、168、169、170、171、172、173、174、175、176、177、536、537、538、539、540、541、542、543、544、545、546、547、548、549、550、551、552、570、571、572、573、595、596、597、598、599、600、601、602、603、604、605、606、607、608、609、610、611、612、613、614、615、630、631、632、646、647、648、649、650、651、652、653、683、684、685、686、687、688、689或690处包含一个或多个突变的氨基酸残基；

[0448] 在某些实施方案中，Cas13蛋白经修饰以具有增加的活性，即更宽的PAM特异性。在特定的实施方案中，Cas13蛋白通过一个或多个残基的突变来修饰，所述残基包括但不限于：AsCas13的位置539、542、547、548、550、551、552、167、604和/或607或者AsCas13同源物、同系物或变体的对应位置，优选位置542或542和607处的突变的氨基酸残基，其中所述突变优选地是542R和607R，诸如S542R和K607R；或者优选位置542和548(以及任选地552)处的突变的氨基酸残基，其中所述突变优选地是542R和548V(以及任选地552R)，诸如S542R和K548V(以及任选地N552R)；或者LbCas13的位置532、538、542和/或595或者AsCas13同源物、同系物或变体的对应位置，优选位置532、532和595处的突变的氨基酸残基，其中所述突变优选地是532R和595R，诸如G532R和K595R；或者优选位置532和538(以及任选地542)处的突变的氨基酸残基，其中所述突变优选地是532R和538V(以及任选地542R)，诸如G532R和K538V(以及任选地Y542R)，最优选地其中所述突变是AsCas13的S542R和K607R、S542R和K548V或者S542R、K548V和N552R。

[0449] 失活的Cas13蛋白

[0450] 当Cas13蛋白具有核酸酶活性时,Cas13蛋白可经修饰以具有降低的核酸酶活性,例如与野生型酶相比,核酸酶失活至少70%、至少80%、至少90%、至少95%、至少97%或为100%;或者换句话说,Cas13酶有利地具有非突变或野生型Cas13酶或CRISPR-Cas蛋白的核酸酶活性的约0%,或者非突变或野生型Cas13酶(例如,非突变或野生型新凶手弗朗西斯菌U112(FnCas13)、氨基酸球菌属BV3L6(AsCas13)、毛螺菌科细菌ND2006(LbCas13)或牛莫拉氏菌237(MbCas13Cas13酶或CRISPR-Cas蛋白的核酸酶活性的不超过约3%或约5%或约10%。这可以通过将突变引入Cas13及其直系同源物的核酸酶结构域来实现。

[0451] 在本发明的优选实施方案中,使用至少一种Cas13蛋白,所述蛋白是Cas13切口酶。更具体地,使用如下Cas13切口酶:其不裂解靶链但能够仅裂解与靶链互补的链,即非靶DNA链,在本文中也称为与指导序列不互补的链。更具体地,Cas13切口酶是包含如下突变的Cas13蛋白:来自氨基酸球菌属的Cas13的Nuc结构域中的位置1226A处或者Cas13直系同源物中的对应位置处的精氨酸突变。在另外的特定实施方案中,酶包含精氨酸至丙氨酸取代或R1226A突变。技术人员应理解,当酶不是AsCas13时,可在对应位置的残基处进行突变。在特定的实施方案中,Cas13是FnCas13,并且突变在位置R1218的精氨酸处。在特定的实施方案中,Cas13是LbCas13,并且突变在位置R1138的精氨酸处。在特定的实施方案中,Cas13是MbCas13,并且突变在位置R1293的精氨酸处。

[0452] 在某些实施方案中,另外地或另选地使用经工程化并且可以包含降低或消除核酸酶活性的一种或多种突变的CRISPR-Cas蛋白。FnCas13p RuvC结构域中的氨基酸位置包括但不限于D917A、E1006A、E1028A、D1227A、D1255A、N1257A、D917A、E1006A、E1028A、D1227A、D1255A和N1257A。申请人还鉴定了最类似于PD-(D/E)XK核酸酶超家族和HincII核酸内切酶样的推定的第二核酸酶结构域。在该推定的核酸酶结构域中产生以实质上降低核酸酶活性的点突变包括但不限于N580A、N584A、T587A、W609A、D610A、K613A、E614A、D616A、K624A、D625A、K627A和Y629A。在一个优选的实施方案中,FnCas13p RuvC结构域中的突变是D917A或E1006A,其中D917A或E1006A突变使FnCas13效应蛋白的DNA裂解活性完全失活。在另一个实施方案中,FnCas13p RuvC结构域中的突变是D1255A,其中突变的FnCas13效应蛋白具有显著降低的溶核活性。

[0453] 更具体地,失活的Cas13酶包括在AsCas13的氨基酸位置As908、As993、As1263或者Cas13直系同源物中的对应位置中突变的酶。另外,失活的Cas13酶包括在LbCas13的氨基酸位置Lb832、925、947或1180或者Cas13直系同源物中的对应位置中突变的酶。更具体地,失活的Cas13酶包括包含AsCas13的突变AsD908A、AsE993A、AsD1263A中的一个或多个或者Cas13直系同源物中的对应突变的酶。另外,失活的Cas13酶包括包含LbCas13的突变LbD832A、E925A、D947A或D1180A中的一个或多个或者Cas13直系同源物中的对应突变的酶。

[0454] 还可以在邻近残基处例如在上面指出的参与核酸酶活性的氨基酸附近的氨基酸处进行突变。在一些实施方案中,仅使RuvC结构域失活,并且在其他实施方案中,使另一个推定的核酸酶结构域失活,其中效应蛋白复合物起切口酶的作用并且仅裂解一条DNA链。在一个优选的实施方案中,另一个推定的核酸酶结构域是HincII样核酸内切酶结构域。

[0455] 失活的Cas13或Cas13切口酶可具有相联的(例如,经由融合蛋白)一个或多个功能结构域,包括例如腺苷脱氨酶或其催化结构域。在一些情况下,有利的是另外提供至少一个异源NLS。在一些情况下,有利的是将NLS定位在N末端。一般来讲,将一个或多个功能结构域

定位在失活的Cas13或Cas13切口酶上允许功能结构域具有正确的空间取向，从而以具有归因的功能效应影响靶标。例如，当功能结构域是腺苷脱氨酶催化结构域时，将腺苷脱氨酶催化结构域放置在允许其接触靶腺嘌呤并使靶腺嘌呤脱氨的空间取向中。这可包括Cas13的N端/C端以外的位置。在一些实施方案中，腺苷脱氨酶蛋白或其催化结构域被插入Cas13的内部环中。

[0456] PAM的测定

[0457] PAM的测定可以如下确保。该实验与在大肠埃希氏菌中进行StCas9异源表达的工作非常相似(Sapranauskas,R.等人,Nucleic Acids Res 39,9275-9282(2011))。申请人将同时含有PAM和抗性基因的质粒引入异源大肠埃希氏菌中，然后涂铺在对应的抗生素上。如果存在质粒的DNA裂解，则申请人不会观察到活菌落。

[0458] 更详细地，如下对DNA靶标进行测定。在该测定中使用两种大肠埃希氏菌菌株。一种菌株携带编码来自细菌菌株的内源效应蛋白基因座的质粒。另一种菌株携带空质粒(例如，pACYC184，对照菌株)。所有可能的7或8bp PAM序列均呈现在抗生素抗性质粒(具有氨苄青霉素抗性基因的pUC19)上。PAM位于原间隔区1的序列旁边(DNA靶向内源效应蛋白基因座中的第一个间隔区)。克隆了两个PAM文库。一个文库具有8个随机bp的5'原间隔区(例如，总共65536个不同的PAM序列＝复杂度)。另一个文库具有7个随机bp的3'原间隔区(例如，总复杂度为16384个不同的PAM)。克隆这两个文库，使每个可能PAM平均具有500个质粒。在单独的转化中用5'PAM和3'PAM文库转化测试菌株和对照菌株，并将转化的细胞分别涂铺在氨苄青霉素平板上。利用质粒进行识别以及随后的切割/干扰使得细胞容易受到氨苄青霉素的影响并阻止其生长。转化后大约12小时，收集测试菌株和对照菌株形成的所有菌落，并分离质粒DNA。质粒DNA用作PCR扩增以及随后的深度测序的模板。未转化文库中所有PAM的表示显示了转化细胞中PAM的预期表示。对照菌株中发现的所有PAM的表示显示了实际表示。测试菌株中所有PAM的表示显示了哪些PAM未被酶识别，并且与对照菌株进行比较可以提取出耗尽PAM的序列。

[0459] 针对某些野生型Cas13直系同源物已鉴定出以下PAM:氨基酸球菌属BV3L6Cas13(AsCas13)、毛螺菌科细菌ND2006Cas13(LbCas13)和苏格兰普雷沃菌(PaCas13)可以裂解以TTTV PAM开头的靶位点，其中V为A/C或G，FnCas13p可以裂解以TTN开头的位点，其中N为A/C/G或T。牛莫拉氏菌AAX08_00205、牛莫拉氏菌AAX11_00205、丁酸弧菌属NC3005、硫微螺菌属XS5或毛螺菌科细菌MA2020PAM为5'TTN，其中N为A/C/G或T。天然PAM序列为TTTV或BTTV，其中B为T/C或G，而V为A/C或G，并且效应蛋白是腔隙莫拉氏菌(Moraxella lacunata)Cas13。

[0460] 经密码子优化的核酸序列

[0461] 在效应蛋白要作为核酸施用的情况下，本申请设想了使用经密码子优化的CRISPR-Cas V型蛋白，更具体为编码Cas13的核酸序列(以及任选的蛋白序列)。在这种情况下，经密码子优化的序列的实例是针对在真核生物例如人类中表达而优化(即，针对在人类中表达而优化)的序列，或针对在本文所讨论的另一种真核生物、动物或哺乳动物中表达而优化的序列。参见例如，WO 2014/093622(PCT/US2013/074667)中的经SaCas9人类密码子优化的序列作为经密码子优化的序列的实例(根据本领域的知识和本公开，对一个或多个编码核酸分子、特别是关于效应蛋白(例如，Cas13)进行密码子优化在技术人员的能力范围

内。尽管这一实例是优选的，但应当理解其他实例也是可能的，并且针对除人类以外的宿主物种进行的密码子优化或针对特定器官进行的密码子优化是已知的。在一些实施方案中，针对在特定细胞诸如真核细胞中表达对编码DNA/RNA靶向Cas蛋白的酶编码序列进行密码子优化。真核细胞可以是特定生物体的那些或来源于特定生物体，例如植物或哺乳动物，包括但不限于人，或如本文所论述的非人真核生物或动物或哺乳动物，例如小鼠、大鼠、兔、犬、家畜或非人哺乳动物或灵长类动物。在一些实施方案中，可将用于修饰人类种系遗传同一性的过程和/或用于修饰可能导致动物遭受痛苦而对人类或动物没有任何实质性医学益处的动物遗传同一性的过程以及由这些过程产生的动物排除在外。一般来讲，密码子优化是指通过在保持天然氨基酸序列的同时用在目标宿主细胞的基因中更常用或最常用的密码子来替换天然序列的至少一个密码子(例如，约或大于约1、2、3、4、5、10、15、20、25、50个或更多个密码子)来修饰核酸序列以在该宿主细胞中增强表达的过程。各种物种对特定氨基酸的某些密码子表现出特定的偏倚。密码子偏倚(生物之间密码子使用的差异)通常与信使RNA(mRNA)的翻译效率相关，而信使RNA(mRNA)的翻译效率又被认为尤其取决于被翻译的密码子的性质和特定转移RNA(tRNA)分子的可用性。所选tRNA在细胞中具有优势通常反映了在肽合成中最常使用的密码子。因此，可以基于密码子优化来定制基因以在给定生物中进行最佳基因表达。密码子使用表可容易获得，例如可在www.kazusa.orjp/codon/的“密码子使用数据库”中获得，并且可以通过多种方式对这些表进行调整。参见Nakamura,Y.等人“Codon usage tabulated from the international DNAsequence databases:status for the year 2000”Nucl.Acids Res.28:292(2000)。也可获得针对在特定宿主细胞中表达而对特定序列进行密码子优化的计算机算法，诸如也可获得Gene Forge(Aptagen;Jacobus,PA)。在一些实施方案中，编码DNA/RNA靶向Cas蛋白的序列中的一个或多个密码子(例如，1、2、3、4、5、10、15、20、25、50个或更多个或所有密码子)对应于特定氨基酸的最常用的密码子。关于酵母中的密码子使用，参考在线酵母基因组数据库http://www.yeastgenome.org/community/codon_usage.shtml或者Codon selection in yeast,Bennetzen和Hall,J Biol Chem.1982年3月25日;257(6):3026-31。关于在包括藻类的植物中的密码子使用，参考Codon usage in higher plants,green algae,and cyanobacteria,Campbell和Gowri,Plant Physiol.1990年1月;92(1):1-11;以及Codon usage in plant genes,Murray等人,Nucleic Acids Res.1989年1月25日;17(2):477-98;或者Selection on the codon bias of chloroplast and cyanelle genes in different plant and algal lineages,Morton BR,J Mol Evol.1998年4月;46(4):449-59。

[0462] 在某些示例实施方案中，CRISPR Cas蛋白选自表1。

[0463] 表1

[0464]

| C2c2 直系同源物 | 代码 | 多字母 |
|---|---|---|
| 沙氏纤毛菌(Leptotrichia shahii) | C2-2 | Lsh |
| 韦德纤毛菌(L wadei) F0279 (Lw2) | C2-3 | Lw2 |
| 西尔李斯特菌(Listeria seeligeri) | C2-4 | Lse |
| 毛螺菌科细菌 MA2020 | C2-5 | LbM |
| 毛螺菌科细菌 NK4A179 | C2-6 | LbNK179 |
| 嗜胺梭菌([Clostridium] aminophilum) DSM 10710 | C2-7 | Ca |
| 鸡肉食杆菌(Carnobacterium gallinarum) DSM 4847 | C2-8 | Cg |
| 鸡肉食杆菌 DSM 4847 | C2-9 | Cg2 |
| 产丙酸栖沼泽杆菌(Paludibacter propionicigenes) WB4 | C2-10 | Pp |
| 魏斯特史氏李斯特菌(Listeria weihenstephanensis) FSL R9-0317 | C2-11 | Lwei |
| 李斯特氏菌科菌(Listeriaceae bacterium) FSL M6-0635 | C2-12 | LbFSL |
| 瓦氏纤毛菌 F0279 | C2-13 | Lw |
| 荚膜红细菌(Rhodobacter capsulatus) SB 1003 | C2-14 | Rc |
| 荚膜红细菌 R121 | C2-15 | Rc |
| 荚膜红细菌 DE442 | C2-16 | Rc |

[0465] 在某些示例实施方案中,CRISPR效应蛋白是选自表2的Cas13a蛋白。

[0466] 表2

[0467]

| | | | |
|---|---|---|---|
| c2c2-5 | 1 | 毛螺菌科细菌 (Lachnospiraceae bacterium) MA2020 (SEQ ID No. 34) | MQISKVNHKHVAVGQKDRERITGFIYNDPVGDEK<br>SLEDVVAKRANDTKVLFNVFNTKDLYDSQESDKS<br>EKDKEIISKGAKFVAKSFNSAITILKKQNKIYSTLTS<br>QQVIKELKDKFGGARIYDDDIEEALTETLKKSFRK<br>ENVRNSIKVLIENAAGIRSSLSKDEEELIQEYFVKQ<br>LVEEYTKTKLQKNVVKSIKNQNMVIQPDSDSQVL<br>SLSESRREKQSSAVSSDTLVNCKEKDVLKAFLTDY<br>AVLDEDERNSLLWKLRNLVNLYFYGSESIRDYSY<br>TKEKSVWKEHDEQKANKTLFIDEICHITKIGKNGK<br>EQKVLDYEENRSRCRKQNINYYRSALNYAKNNTS<br>GIFENEDSNHFWIHLIENEVERLYNGIENGEEFKFE<br>TGYISEKVWKAVINHLSIKYIALGKAVYNYAMKE<br>LSSPGDIEPGKIDDSYINGITSFDYEIIKAEESLQRDI<br>SMNVVFATNYLACATVDTDKDFLLFSKEDIRSCT<br>KKDGNLCKNIMQFWGGYSTWKNFCEEYLKDDK<br>DALELLYSLKSMLYSMRNSSFHFSTENVDNGSWD<br>TELIGKLFEEDCNRAARIEKEKFYNNNLHMFYSSS<br>LLEKVLERLYSSHHERASQVPSFNRVFVRKNFPSS<br>LSEQRITPKFTDSKDEQIWQSAVYYLCKEIYYNDF<br>LQSKEAYKLFREGVKNLDKNDINNQKAADSFKQ<br>AVVYYGKAIGNATLSQVCQAIMTEYNRQNNDGL<br>KKKSAYAEKQNSNKYKHYPLFLKQVLQSAFWEY<br>LDENKEIYGFISAQIHKSNVEIKAEDFIANYSSQQY<br>KKLVDKVKKTPELQKWYTLGRLINPRQANQFLGS<br>IRNYVQFVKDIQRRAKENGNPIRNYYEVLESDSIIK<br>ILEMCTKLNGTTSNDIHDYFRDEDEYAEYISQFVN<br>FGDVHSGAALNAFCNSESEGKKNGIYYDGINPIVN<br>RNWVLCKLYGSPDLISKIISRVNENMIHDFHKQED<br>LIREYQIKGICSNKKEQQDLRTFQVLKNRVELRDI<br>VEYSEIINELYGQLIKWCYLRERDLMYFQLGFHYL<br>CLNNASSKEADYIKINVDDRNISGAILYQIAAMYIN<br>GLPVYYKKDDMYVALKSGKKASDELNSNEQTSK<br>KINYFLKYGNNILGDKKDQLYLAGLELFENVAEH<br>ENIIIFRNEIDHFHYFYDRDRSMLDLYSEVFDRFFT<br>YDMKLRKNVVNMLYNILLDHNIVSSFVFETGEKK<br>VGRGDSEVIKPSAKIRLRANNGVSSDVFTYKVGS<br>KDELKIATLPAKNEEFLLNVARLIYYPDMEAVSEN<br>MVREGVVKVEKSNDKKGKISRGSNTRSSNQSKY<br>NNKSKNRMNYSMGSIFEKMDLKFD |
| c2c2-6 | 2 | 毛螺菌科细菌 NK4A179 (SEQ ID No. 35) | MKISKVREENRGAKLTVNAKTAVVSENRSQEGIL<br>YNDPSRYGKSRKNDEDRDRYIESRLKSSGKLYRIF<br>NEDKNKRETDELQWFLSEIVKKINRRNGLVLSDM<br>LSVDDRAFEKAFEKYAELSYTNRRNKVSGSPAFE<br>TCGVDAATAERLKGIISETNFINRIKNNIDNKVSED<br>IIDRIIAKYLKKSLCRERVKRGLKKLLMNAFDLPYS<br>DPDIDVQRDFIDYVLEDFYHVRAKSQVSRSIKNM<br>NMPVQPEGDGKFAITVSKGGTESGNKRSAEKEAF<br>KKFLSDYASLDERVRDDMLRRMRRLVVLYFYGS |

[0468]

| | | | |
|---|---|---|---|
| | | | DDSKLSDVNEKFDVWEDHAARRVDNREFIKLPLE<br>NKLANGKTDKDAERIRKNTVKELYRNQNIGCYR<br>QAVKAVEEDNNGRYFDDKMLNMFFIHRIEYGVE<br>KIYANLKQVTEFKARTGYLSEKIWKDLINYISIKYI<br>AMGKAVYNYAMDELNASDKKEIELGKISEEYLSG<br>ISSFDYELIKAEEMLQRETAVYVAFAARHLSSQTV<br>ELDSENSDFLLLKPKGTMDKNDKNKLASNNILNF<br>LKDKETLRDTILQYFGGHSLWTDFPFDKYLAGGK<br>DDVDFLTDLKDVIYSMRNDSFHYATENHNNGKW<br>NKELISAMFEHETERMTVVMKDKFYSNNLPMFY<br>KNDDLKKLLIDLYKDNVERASQVPSFNKVFVRKN<br>FPALVRDKDNLGIELDLKADADKGENELKFYNAL<br>YYMFKEIYYNAFLNDKNVRERFITKATKVADNYD<br>RNKERNLKDRIKSAGSDEKKKLREQLQNYIAEND<br>FGQRIKNIVQVNPDYTLAQICQLIMTEYNQQNNGC<br>MQKKSAARKDINKDSYQHYKMLLLVNLRKAFLE<br>FIKENYAFVLKPYKHDLCDKADFVPDFAKYVKPY<br>AGLISRVAGSSELQKWYIVSRFLSPAQANHMLGFL<br>HSYKQYVWDIYRRASETGTEINHSIAEDKIAGVDI<br>TDVDAVIDLSVKLCGTISSEISDYFKDDEVYAEYIS<br>SYLDFEYDGGNYKDSLNRFCNSDAVNDQKVALY<br>YDGEHPKLNRNIILSKLYGERRFLEKITDRVSRSDI<br>VEYYKLKKETSQYQTKGIFDSEDEQKNIKKFQEM<br>KNIVEFRDLMDYSEIADELQGQLINWIYLRERDLM<br>NFQLGYHYACLNNDSNKQATYVTLDYQGKKNR<br>KINGAILYQICAMYINGLPLYYVDKDSSEWTVSDG<br>KESTGAKIGEFYRYAKSFENTSDCYASGLEIFENIS<br>EHDNITELRNYIEHFRYYSSFDRSFLGIYSEVFDRF<br>FTYDLKYRKNVPTILYNILLQHFVNVRFEFVSGKK<br>MIGIDKKDRKIAKEKECARITIREKNGVYSEQFTY<br>KLKNGTVYVDARDKRYLQSIIRLLFYPEKVNMDE<br>MIEVKEKKKPSDNNTGKGYSKRDRQQDRKEYDK<br>YKEKKKKEGNFLSGMGGNINWDEINAQLKN |
| c2c2-7 | 3 | 嗜胺[梭菌]DSM 10710 SEQ ID No. 36) | MKFSKVDHTRSAVGIQKATDSVHGMLYTDPKKQ<br>EVNDLDKRFDQLNVKAKRLYNVFNQSKAEEDDD<br>EKRFGKVVKKLNRELKDLLFHREVSRYNSIGNAK<br>YNYYGIKSNPEEIVSNLGMVESLKGERDPQKVISK<br>LLLYYLRKGLKPGTDGLRMILEASCGLRKLSGDE<br>KELKVFLQTLDEDFEKKTFKKNLIRSIENQNMAVQ<br>PSNEGDPIIGITQGRFNSQKNEEKSAIERMMSMYA<br>DLNEDHREDVLRKLRRLNVLYFNVDTEKTEEPTL<br>PGEVDTNPVFEVWHDHEKGKENDRQFATFAKILT<br>EDRETRKKEKLAVKEALNDLKSAIRDHNIMAYRC<br>SIKVTEQDKDGLFFEDQRINRFWIHHIESAVERILA<br>SINPEKLYKLRIGYLGEKVWKDLLNYLSIKYIAVG<br>KAVFHFAMEDLGKTGQDIELGKLSNSVSGGLTSF<br>DYEQIRADETLQRQLSVEVAFAANNLFRAVVGQT<br>GKKIEQSKSEENEEDFLLWKAEKIAESIKKEGEGN |

[0469]

| | | | |
|---|---|---|---|
| | | | TLKSILQFFGGASSWDLNHFCAAYGNESSALGYET<br>KFADDLRKAIYSLRNETFHFTTLNKGSFDWNAKLI<br>GDMFSHEAATGIAVERTRFYSNNLPMFYRESDLK<br>RIMDHLYNTYHPRASQVPSFNSVFVRKNFRLFLSN<br>TLNTNTSFDTEVYQKWESGVYYLFKEIYYNSFLPS<br>GDAHHLFFEGLRRIRKEADNLPIVGKEAKKRNAV<br>QDFGRRCDELKNLSLSAICQMIMTEYNEQNNGNR<br>KVKSTREDKRKPDIFQHYKMLLLRTLQEAFAIYIR<br>REEFKFIFDLPKTLYVMKPVEEFLPNWKSGMFDSL<br>VERVKQSPDLQRWYVLCKFLNGRLLNQLSGVIRS<br>YIQFAGDIQRRAKANHNRLYMDNTQRVEYYSNV<br>LEVVDFCIKGTSRFSNVFSDYFRDEDAYADYLDN<br>YLQFKDEKIAEVSSFAALKTFCNEEEVKAGIYMD<br>GENPVMQRNIVMAKLFGPDEVLKNVVPKVTREEI<br>EEYYQLEKQIAPYRQNGYCKSEEDQKKLLRFQRI<br>KNRVEFQTITEFSEIINELLGQLISWSFLRERDLLYF<br>QLGFHYLCLHNDTEKPAEYKEISREDGTVIRNAIL<br>HQVAAMYVGGLPVYTLADKKLAAFEKGEADCKL<br>SISKDTAGAGKKIKDFFRYSKYVLIKDRMLTDQN<br>QKYTIYLAGLELFENTDEHDNITDVRKYVDHFKY<br>YATSDENAMSILDLYSEIHDRFFTYDMKYQKNVA<br>NMLENILLRHFVLIRPEFFTGSKKVGEGKKITCKA<br>RAQIEIAENGMRSEDFTYKLSDGKKNISTCMIAAR<br>DQKYLNTVARLLYYPHEAKKSIVDTREKKNNKK<br>TNRGDGTFNKQKGTARKEKDNGPREFNDTGFSN<br>TPFAGFDPFRNS |
| c2c2-8 | 5 | 鸡肉食杆菌 DSM 4847 (SEQ ID No. 37) | MRITKVKIKLDNKLYQVTMQKEEKYGTLKLNEES<br>RKSTAEILRLKKASFNKSFHSKTINSQKENKNATIK<br>KNGDYISQIFEKLVGVDTNKNIRKPKMSLTDLKDL<br>PKKDLALFIKRKFKNDDIVEIKNLDLISLFYNALQK<br>VPGEHFTDESWADFCQEMMPYREYKNKFIERKIIL<br>LANSIEQNKGFSINPETFSKRKRVLHQWAIEVQER<br>GDFSILDEKLSKLAEIYNFKKMCKRVQDELNDLEK<br>SMKKGKNPEKEKEAYKKQKNFKIKTIWKDYPYK<br>THIGLIEKIKENEELNQFNIEIGKYFEHYFPIKKERC<br>TEDEPYYLNSETIATTVNYQLKNALISYLMQIGKY<br>KQFGLENQVLDSKKLQEIGIYEGFQTKFMDACVF<br>ATSSLKNIIEPMRSGDILGKREFKEAIATSSFVNYH<br>HFFPYFPFELKGMKDRESELIPFGEQTEAKQMQNI<br>WALRGSVQQIRNEIFHSFDKNQKFNLPQLDKSNFE<br>FDASENSTGKSQSYIETDYKFLFEAEKNQLEQFFIE<br>RIKSSGALEYYPLKSLEKLFAKKEMKFSLGSQVVA<br>FAPSYKKLVKKGHSYQTATEGTANYLGLSYYNR<br>YELKEESFQAQYYLLKLIYQYVFLPNFSQGNSPAF<br>RETVKAILRINKDEARKKMKKNKKFLRKYAFEQV<br>REMEFKETPDQYMSYLQSEMREEKVRKAEKNDK<br>GFEKNITMNFEKLLMQIFVKGFDVFLTTFAGKELL<br>LSSEEKVIKETEISLSKKINEREKTLKASIQVEHQLV |

[0470]

| | | | |
|---|---|---|---|
| | | | ATNSAISYWLFCKLLDSRHLNELRNEMIKFKQSRI KFNHTQHAELIQNLLPIVELTILSNDYDEKNDSQN VDVSAYFEDKSLYETAPYVQTDDRTRVSFRPILKL EKYHTKSLIEALLKDNPQFRVAATDIQEWMHKRE EIGELVEKRKNLHTEWAEGQQTLGAEKREEYRDY CKKIDRFNWKANKVTLTYLSQLHYLITDLLGRMV GFSALFERDLVYFSRSFSELGGETYHISDYKNLSG VLRLNAEVKPIKIKNIKVIDNEENPYKGNEPEVKPF LDRLHAYLENVIGIKAVHGKIRNQTAHLSVLQLEL SMIESMNNLRDLMAYDRKLKNAVTKSMIKILDKH GMILKLKIDENHKNFEIESLIPKEIIHLKDKAIKTNQ VSEEYCQLVLALLTTNPGNQLN |
| c2c2 -9 | 6 | 鸡肉食杆菌 DSM 4847 (SEQ ID No. 38) | MRMTKVKINGSPVSMNRSKLNGHLVWNGTTNTV NILTKKEQSFAASFLNKTLVKADQVKGYKVLAEN IFIIFEQLEKSNSEKPSVYLNNIRRLKEAGLKRFFKS KYHEEIKYTSEKNQSVPTKLNLIPLFFNAVDRIQED KFDEKNWSYFCKEMSPYLDYKKSYLNRKKEILAN SIQQNRGFSMPTAEEPNLLSKRKQLFQQWAMKFQ ESPLIQQNNFAVEQFNKEFANKINELAAVYNVDEL CTAITEKLMNFDKDKSNKTRNFEIKKLWKQHPHN KDKALIKLFNQEGNEALNQFNIELGKYFEHYFPKT GKKESAESYYLNPQTIIKTVGYQLRNAFVQYLLQ VGKLHQYNKGVLDSQTLQEIGMYEGFQTKFMDA CVFASSSLRNIIQATTNEDILTREKFKKELEKNVEL KHDLFFKTEIVEERDENPAKKIAMTPNELDLWAIR GAVQRVRNQIFHQQINKRHEPNQLKVGSFENGDL GNVSYQKTIYQKLFDAEIKDIEIYFAEKIKSSGALE QYSMKDLEKLFSNKELTLSLGGQVVAFAPSYKKL YKQGYFYQNEKTIELEQFTDYDFSNDVFKANYYL IKLIYHYVFLPQFSQANNKLFKDTVHYVIQQNKEL NTTEKDKKNNKKIRKYAFEQVKLMKNESPEKYM QYLQREMQEERTIKEAKKTNEEKPNYNFEKLLIQI FIKGFDTFLRNFDLNLNPAEELVGTVKEKAEGLRK RKERIAKILNVDEQIKTGDEEIAFWIFAKLLDARHL SELRNEMIKFKQSSVKKGLIKNGDLIEQMQPILELC ILSNDSESMEKESFDKIEVFLEKVELAKNEPYMQE DKLTPVKFRFMKQLEKYQTRNFIENLVIENPEFKV SEKIVLNWHEEKEKIADLVDKRTKLHEEWASKAR EIEEYNEKIKKNKSKKLDKPAEFAKFAEYKIICEAI ENFNRLDHKVRLTYLKNLHYLMIDLMGRMVGFS VLFERDFVYMGRSYSALKKQSIYLNDYDTFANIR DWEVNENKHLFGTSSSDLTFQETAEFKNLKKPME NQLKALLGVTNHSFEIRNNIAHLHVLRNDGKGEG VSLLSCMNDLRKLMSYDRKLKNAVTKAIIKILDK HGMILKLTNNDHTKPFEIESLKPKKIIHLEKSNHSF PMDQVSQEYCDLVKKMLVFTN |
| c2c2 | 7 | 产丙酸栖 | MRVSKVKVKDGGKDKMVLVHRKTTGAQLVYSG |

[0471]

| | | | |
|---|---|---|---|
| -10 | | 沼泽杆菌 WB4 (SEQ ID No. 39) | QPVSNETSNILPEKKRQSFDLSTLNKTIIKFDTAKK<br>QKLNVDQYKIVEKIFKYPKQELPKQIKAEEILPFLN<br>HKFQEPVKYWKNGKEESFNLTLLIVEAVQAQDKR<br>KLQPYYDWKTWYIQTKSDLLKKSIENNRIDLTENL<br>SKRKKALLAWETEFTASGSIDLTHYHKVYMTDVL<br>CKMLQDVKPLTDDKGKINTNAYHRGLKKALQNH<br>QPAIFGTREVPNEANRADNQLSIYHLEVVKYLEHY<br>FPIKTSKRRNTADDIAHYLKAQTLKTTIEKQLVNAI<br>RANIIQQGKTNHHELKADTTSNDLIRIKTNEAFVL<br>NLTGTCAFAANNIRNMVDNEQTNDILGKGDFIKS<br>LLKDNTNSQLYSFFFGEGLSTNKAEKETQLWGIRG<br>AVQQIRNNVNHYKKDALKTVFNISNFENPTITDPK<br>QQTNYADTIYKARFINELEKIPEAFAQQLKTGGAV<br>SYYTIENLKSLLTTFQFSLCRSTIPFAPGFKKVFNG<br>GINYQNAKQDESFYELMLEQYLRKENFAEESYNA<br>RYFMLKLIYNNLFLPGFTTDRKAFADSVGFVQMQ<br>NKKQAEKVNPRKKEAYAFEAVRPMTAADSIADY<br>MAYVQSELMQEQNKKEEKVAEETRINFEKFVLQ<br>VFIKGFDSFLRAKEFDFVQMPQPQLTATASNQQK<br>ADKLNQLEASITADCKLTPQYAKADDATHIAFYV<br>FCKLLDAAHLSNLRNELIKFRESVNEFKFHHLLEII<br>EICLLSADVVPTDYRDLYSSEADCLARLRPFIEQG<br>ADITNWSDLFVQSDKHSPVIHANIELSVKYGTTKL<br>LEQIINKDTQFKTTEANFTAWNTAQKSIEQLIKQR<br>EDHHEQWVKAKNADDKEKQERKREKSNFAQKFI<br>EKHGDDYLDICDYINTYNWLDNKMHFVHLNRLH<br>GLTIELLGRMAGFVALFDRDFQFFDEQQIADEFKL<br>HGFVNLHSIDKKLNEVPTKKIKEIYDIRNKIIQINGN<br>KINESVRANLIQFISSKRNYYNNAFLHVSNDEIKEK<br>QMYDIRNHIAHFNYLTKDAADFSLIDLINELRELL<br>HYDRKLKNAVSKAFIDLFDKHGMILKLKLNADHK<br>LKVESLEPKKIYHLGSSAKDKPEYQYCTNQVMM<br>AYCNMCRSLLEMKK |
| c2c2 -11 | 9 | 魏斯特史氏李斯特菌 FSL R9-0317 (SEQ ID No. 40) | MLALLHQEVPSQKLHNLKSLNTESLTKLFKPKFQ<br>NMISYPPSKGAEHVQFCLTDIAVPAIRDLDEIKPD<br>WGIFFEKLKPYTDWAESYIHYKQTTIQKSIEQNKI<br>QSPDSPRKLVLQKYVTAFLNGEPLGLDLVAKKYK<br>LADLAESFKVVDLNEDKSANYKIKACLQQHQRNI<br>LDELKEDPELNQYGIEVKKYIQRYFPIKRAPNRSK<br>HARADFLKKELIESTVEQQFKNAVYHYVLEQGK<br>MEAYELTDPKTKDLQDIRSGEAFSFKFINACAFAS<br>NNLKMILNPECEKDILGKGDFKKNLPNSTTQSDV<br>VKKMIPFFSDEIQNVNFDEAIWAIRGSIQQIRNEVY<br>HCKKHSWKSILKIKGFEFEPNNMKYTDSDMQKL<br>MDKDIAKIPDFIEEKLKSSGIIRFYSHDKLQSIWEM<br>KQGFSLLTTNAPFVPSFKRVYAKGHDYQTSKNRY<br>YDLGLTTFDILEYGEEDFRARYFLTKLVYYQQFM<br>PWFTADNNAFRDAANFVLRLNKNRQQDAKAFINI |

[0472]

| | | | |
|---|---|---|---|
| | | | REVEEGEMPRDYMGYVQGQIAIHEDSTEDTPNHF EKFISQVFIKGFDSHMRSADLKFIKNPRNQGLEQSE IEEMSFDIKVEPSFLKNKDDYIAFWTFCKMLDARH LSELRNEMIKYDGHLTGEQEIIGLALLGVDSREND WKQFFSSEREYEKIMKGYVGEELYQREPYRQSDG KTPILFRGVEQARKYGTETVIQRLFDASPEFKVSK CNITEWERQKETIEETIERRKELHNEWEKNPKKPQ NNAFFKEYKECCDAIDAYNWHKNKTTLVYVNEL HHLLIEILGRYVGYVAIADRDFQCMANQYFKHSGI TERVEYWGDNRLKSIKKLDTFLKKEGLFVSEKNA RNHIAHLNYLSLKSECTLLYLSERLREIFKYDRKL KNAVSKSLIDILDRHGMSVVFANLKENKHRLVIKS LEPKKLRHLGEKKIDNGYIETNQVSEEYCGIVKRL LEI |
| c2c2-12 | 10 | 李斯特氏菌科菌FSL M6-0635 = 纽约李斯特菌 (Listeria newyorkensis) FSL M6-0635 (SEQ ID No. 41) | MKITKMRVDGRTIVMERTSKEGQLGYEGIDGNKT TEIIFDKKKESFYKSILNKTVRKPDEKEKNRRKQAI NKAINKEITELMLAVLHQEVPSQKLHNLKSLNTES LTKLFKPKFQNMISYPPSKGAEHVQFCLTDIAVPAI RDLDEIKPDWGIFFEKLKPYTDWAESYIHYKQTTI QKSIEQNKIQSPDSPRKLVLQKYVTAFLNGEPLGL DLVAKKYKLADLAESFKLVDLNEDKSANYKIKAC LQQHQRNILDELKEDPELNQYGIEVKKYIQRYFPI KRAPNRSKHARADFLKKELIESTVEQQFKNAVYH YVLEQGKMEAYELTDPKTKDLQDIRSGEAFSFKFI NACAFASNNLKMILNPECEKDILGKGNFKKNLPN STTRSDVVKKMIPFFSDELQNVNFDEAIWAIRGSIQ QIRNEVYHCKKHSWKSILKIKGFEFEPNNMKYAD SDMQKLMDKDIAKIPEFIEEKLKSSGVVRFYRHDE LQSIWEMKQGFSLLTTNAPFVPSFKRVYAKGHDY QTSKNRYYNLDLTTFDILEYGEEDFRARYFLTKLV YYQQFMPWFTADNNAFRDAANFVLRLNKNRQQ DAKAFINIREVEEGEMPRDYMGYVQGQIAIHEDSI EDTPNHFEKFISQVFIKGFDRHMRSANLKFIKNPR NQGLEQSEIEEMSFDIKVEPSFLKNKDDYIAFWIFC KMLDARHLSELRNEMIKYDGHLTGEQEIIGLALLG VDSRENDWKQFFSSEREYEKIMKGYVVEELYQRE PYRQSDGKTPILFRGVEQARKYGTETVIQRLFDAN PEFKVSKCNLAEWERQKETIEETIKRRKELHNEW AKNPKKPQNNAFFKEYKECCDAIDAYNWHKNKT TLAYVNELHHLLIEILGRYVGYVAIADRDFQCMA NQYFKHSGITERVEYWGDNRLKSIKKLDTFLKKE GLFVSEKNARNHIAHLNYLSLKSECTLLYLSERLR EIFKYDRKLKNAVSKSLIDILDRHGMSVVFANLKE NKHRLVIKSLEPKKLRHLGGKKIDGGYIETNQVSE EYCGIVKRLLEM |
| c2c2-13 | 12 | 瓦氏纤毛菌 F0279 | MKVTKVDGISHKKYIEEGKLVKSTSEENRTSERLS ELLSIRLDIYIKNPDNASEEENRIRRENLKKFFSNK |

[0473]

| | | | |
|---|---|---|---|
| | | (SEQ ID No. 42) | VLHLKDSVLYLKNRKEKNAVQDKNYSEEDISEYD<br>LKNKNSFSVLKKILLNEDVNSEELEIFRKDVEAKL<br>NKINSLKYSFEENKANYQKINENNVEKVGGKSKR<br>NIIYDYYRESAKRNDYINNVQEAFDKLYKKEDIEK<br>LFFLIENSKKHEKYKIREYYHKIIGRKNDKENFAKI<br>IYEEIQNVNNIKELIEKIPDMSELKKSQVFYKYYLD<br>KEELNDKNIKYAFCHFVEIEMSQLLKNYVYKRLS<br>NISNDKIKRIFEYQNLKKLIENKLLNKLDTYVRNC<br>GKYNYYLQVGEIATSDFIARNRQNEAFLRNIIGVS<br>SVAYFSLRNILETENENDITGRMRGKTVKNNKGE<br>EKYVSGEVDKIYNENKQNEVKENLKMFYSYDFN<br>MDNKNEIEDFFANIDEAISSIRHGIVHFNLELEGKDI<br>FAFKNIAPSEISKKMFQNEINEKKLKLKIFKQLNSA<br>NVFNYYEKDVIIKYLKNTKFNFVNKNIPFVPSFTK<br>LYNKIEDLRNTLKFFWSVPKDKEEKDAQIYLLKNI<br>YYGEFLNKFVKNSKVFFKITNEVIKINKQRNQKTG<br>HYKYQKFENIEKTVPVEYLAIIQSREMINNQDKEE<br>KNTYIDFIQQIFLKGFIDYLNKNNLKYIESNNNNDN<br>NDIFSKIKIKKDNKEKYDKILKNYEKHNRNKEIPH<br>EINEFVREIKLGKILKYTENLNMFYLILKLLNHKEL<br>TNLKGSLEKYQSANKEETFSDELELINLLNLDNNR<br>VTEDFELEANEIGKFLDFNENKIKDRKELKKFDTN<br>KIYFDGENIIKHRAFYNIKKYGMLNLLEKIADKAK<br>YKISLKELKEYSNKKNEIEKNYTMQQNLHRKYAR<br>PKKDEKFNDEDYKEYEKAIGNIQKYTHLKNKVEF<br>NELNLLQGLLLKILHRLVGYTSIWERDLRFRLKGE<br>FPENHYIEEIFNFDNSKNVKYKSGQIVEKYINFYKE<br>LYKDNVEKRSIYSDKKVKKLKQEKKDLYIRNYIA<br>HFNYIPHAEISLLEVLENLRKLLSYDRKLKNAIMK<br>SIVDILKEYGFVATFKIGADKKIEIQTLESEKIVHLK<br>NLKKKKLMTDRNSEELCELVKVMFEYKALE |
| c2c2-14 | 15 | 英膜红细菌 SB 1003 (SEQ ID No. 43) | MQIGKVQGRTISEFGDPAGGLKRKISTDGKNRKEL<br>PAHLSSDPKALIGQWISGIDKIYRKPDSRKSDGKAI<br>HSPTPSKMQFDARDDLGEAFWKLVSEAGLAQDS<br>DYDQFKRRLHPYGDKFQPADSGAKLKFEADPPEP<br>QAFHGRWYGAMSKRGNDAKELAAALYEHLHVD<br>EKRIDGQPKRNPKTDKFAPGLVVARALGIESSVLP<br>RGMARLARNWGEEEIQTYFVVDVAASVKEVAKA<br>AVSAAQAFDPPRQVSGRSLSPKVGFALAEHLERV<br>TGSKRCSFDPAAGPSVLALHDEVKKTYKRLCARG<br>KNAARAFPADKTELLALMRHTHENRVRNQMVR<br>MGRVSEYRGQQAGDLAQSHYWTSAGQTEIKESEI<br>FVRLWVGAFALAGRSMKAWIDPMGKIVNTEKND<br>RDLTAAVNIRQVISNKEMVAEAMARRGIYFGETP<br>ELDRLGAEGNEGFVFALLRYLRGCRNQTFHLGAR<br>AGFLKEIRKELEKTRWGKAKEAEHVVLTDKTVA<br>AIRAIIDNDAKALGARLLADLSGAFVAHYASKEHF<br>STLYSEIVKAVKDAPEVSSGLPRLKLLLKRADGVR |

| | | | |
|---|---|---|---|
| | | | GYVHGLRDTRKHAFATKLPPPPAPRELDDPATKA<br>RYIALLRLYDGPFRAYASGITGTALAGPAARAKEA<br>ATALAQSVNVTKAYSDVMEGRTSRLRPPNDGETL<br>REYLSALTGETATEFRVQIGYESDSENARKQAEFI<br>ENYRRDMLAFMFEDYIRAKGFDWILKIEPGATAM<br>TRAPVLPEPIDTRGQYEHWQAALYLVMHFVPASD<br>VSNLLHQLRKWEALQGKYELVQDGDATDQADA<br>RREALDLVKRFRDVLVLFLKTGEARFEGRAAPFD<br>LKPFRALFANPATFDRLFMATPTTARPAEDDPEGD<br>GASEPELRVARTLRGLRQIARYNHMAVLSDLFAK<br>HKVRDEEVARLAEIEDETQEKSQIVAAQELRTDLH<br>DKVMKCHPKTISPEERQSYAAAIKTIEEHRFLVGR<br>VYLGDHLRLHRLMMDVIGRLIDYAGAYERDTGTF<br>LINASKQLGAGADWAVTIAGAANTDARTQTRKD<br>LAHFNVLDRADGTPDLTALVNRAREMMAYDRKR<br>KNAVPRSILDMLARLGLTLKWQMKDHLLQDATIT<br>QAAIKHLDKVRLTVGGPAAVTEARFSQDYLQMV<br>AAVFNGSVQNPKPRRRDDGDAWHKPPKPATAQS<br>QPDQKPPNKAPSAGSRLPPPQVGEVYEGVVVKVI<br>DTGSLGFLAVEGVAGNIGLHISRLRRIREDAIIVGR<br>RYRFRVEIYVPPKSNTSKLNAADLVRID |
| [0474] | c2c2-15 | 16 | 英膜红细菌 R121 (SEQ ID No. 44) | MQIGKVQGRTISEFGDPAGGLKRKISTDGKNRKEL<br>PAHLSSDPKALIGQWISGIDKIYRKPDSRKSDGKAI<br>HSPTPSKMQFDARDDLGEAFWKLVSEAGLAQDS<br>DYDQFKRRLHPYGDKFQPADSGAKLKFEADPPEP<br>QAFHGRWYGAMSKRGNDAKELAAALYEHLHVD<br>EKRIDGQPKRNPKTDKFAPGLVVARALGIESSVLP<br>RGMARLARNWGEEEIQTYFVVDVAASVKEVAKA<br>AVSAAQAFDPPRQVSGRSLSPKVGFALAEHLERV<br>TGSKRCSFDPAAGPSVLALHDEVKKTYKRLCARG<br>KNAARAFPADKTELLALMRHTHENRVRNQMVR<br>MGRVSEYRGQQAGDLAQSHYWTSAGQTEIKESEI<br>FVRLWVGAFALAGRSMKAWIDPMGKIVNTEKND<br>RDLTAAVNIRQVISNKEMVAEAMARRGIYFGETP<br>ELDRLGAEGNEGFVFALLRYLRGCRNQTFHLGAR<br>AGFLKEIRKELEKTRWGKAKEAEHVVLTDKTVA<br>AIRAIIDNDAKALGARLLADLSGAFVAHYASKEHF<br>STLYSEIVKAVKDAPEVSSGLPRLKLLLKRADGVR<br>GYVHGLRDTRKHAFATKLPPPPAPRELDDPATKA<br>RYIALLRLYDGPFRAYASGITGTALAGPAARAKEA<br>ATALAQSVNVTKAYSDVMEGRSSRLRPPNDGETL<br>REYLSALTGETATEFRVQIGYESDSENARKQAEFI<br>ENYRRDMLAFMFEDYIRAKGFDWILKIEPGATAM<br>TRAPVLPEPIDTRGQYEHWQAALYLVMHFVPASD<br>VSNLLHQLRKWEALQGKYELVQDGDATDQADA<br>RREALDLVKRFRDVLVLFLKTGEARFEGRAAPFD<br>LKPFRALFANPATFDRLFMATPTTARPAEDDPEGD<br>GASEPELRVARTLRGLRQIARYNHMAVLSDLFAK |

[0475]

| | | | |
|---|---|---|---|
| | | | HKVRDEEVARLAEIEDETQEKSQIVAAQELRTDLH<br>DKVMKCHPKTISPEERQSYAAAIKTIEEHRFLVGR<br>VYLGDHLRLHRLMMDVIGRLIDYAGAYERDTGTF<br>LINASKQLGAGADWAVTIAGAANTDARTQTRKD<br>LAHFNVLDRADGTPDLTALVNRAREMMAYDRKR<br>KNAVPRSILDMLARLGLTLKWQMKDHLLQDATIT<br>QAAIKHLDKVRLTVGGPAAVTEARFSQDYLQMV<br>AAVFNGSVQNPKPRRRDDGDAWHKPPKPATAQS<br>QPDQKPPNKAPSAGSRLPPPQVGEVYEGVVVKVI<br>DTGSLGFLAVEGVAGNIGLHISRLRRIREDAIIVGR<br>RYRFRVEIYVPPKSNTSKLNAADLVRID |
| c2c2-16 | 17 | 英膜红细菌 DE442 (SEQ ID No. 45) | MQIGKVQGRTISEFGDPAGGLKRKISTDGKNRKEL<br>PAHLSSDPKALIGQWISGIDKIYRKPDSRKSDGKAI<br>HSPTPSKMQFDARDDLGEAFWKLVSEAGLAQDS<br>DYDQFKRRLHPYGDKFQPADSGAKLKFEADPPEP<br>QAFHGRWYGAMSKRGNDAKELAAALYEHLHVD<br>EKRIDGQPKRNPKTDKFAPGLVVARALGIESSVLP<br>RGMARLARNWGEEEIQTYFVVDVAASVKEVAKA<br>AVSAAQAFDPPRQVSGRSLSPKVGFALAEHLERV<br>TGSKRCSFDPAAGPSVLALHDEVKKTYKRLCARG<br>KNAARAFPADKTELLALMRHTHENRVRNQMVR<br>MGRVSEYRGQQAGDLAQSHYWTSAGQTEIKESEI<br>FVRLWVGAFALAGRSMKAWIDPMGKIVNTEKND<br>RDLTAAVNIRQVISNKEMVAEAMARRGIYFGETP<br>ELDRLGAEGNEGFVFALLRYLRGCRNQTFHLGAR<br>AGFLKEIRKELEKTRWGKAKEAEHVVLTDKTVA<br>AIRAIIDNDAKALGARLLADLSGAFVAHYASKEHF<br>STLYSEIVKAVKDAPEVSSGLPRLKLLLKRADGVR<br>GYVHGLRDTRKHAFATKLPPPPAPRELDDPATKA<br>RYIALLRLYDGPFRAYASGITGTALAGPAARAKEA<br>ATALAQSVNVTKAYSDVMEGRSSRLRPPNDGETL<br>REYLSALTGETATEFRVQIGYESDSENARKQAEFI<br>ENYRRDMLAFMFEDYIRAKGFDWILKIEPGATAM<br>TRAPVLPEPIDTRGQYEHWQAALYLVMHFVPASD<br>VSNLLHQLRKWEALQGKYELVQDGDATDQADA<br>RREALDLVKRFRDVLVLFLKTGEARFEGRAAPFD<br>LKPFRALFANPATFDRLFMATPTTARPAEDDPEGD<br>GASEPELRVARTLRGLRQIARYNHMAVLSDLFAK<br>HKVRDEEVARLAEIEDETQEKSQIVAAQELRTDLH<br>DKVMKCHPKTISPEERQSYAAAIKTIEEHRFLVGR<br>VYLGDHLRLHRLMMDVIGRLIDYAGAYERDTGTF<br>LINASKQLGAGADWAVTIAGAANTDARTQTRKD<br>LAHFNVLDRADGTPDLTALVNRAREMMAYDRKR<br>KNAVPRSILDMLARLGLTLKWQMKDHLLQDATIT<br>QAAIKHLDKVRLTVGGPAAVTEARFSQDYLQMV<br>AAVFNGSVQNPKPRRRDDGDAWHKPPKPATAQS<br>QPDQKPPNKAPSAGSRLPPPQVGEVYEGVVVKVI<br>DTGSLGFLAVEGVAGNIGLHISRLRRIREDAIIVGR |

[0476]

| | | | |
|---|---|---|---|
| | | | RYRFRVEIYVPPKSNTSKLNAADLVRID |
| c2c2-2 | | (SEQ ID No. 46) | MGNLFGHKRWYEVRDKKDFKIKRKVKVKRNYD<br>GNKYILNINENNNKEKIDNNKFIRKYINYKKNDNI<br>LKEFTRKFHAGNILFKLKGKEGIIRIENNDDFLETE<br>EVVLYIEAYGKSEKLKALGITKKKIIDEAIRQGITK<br>DDKKIEIKRQENEEEIEIDIRDEYTNKTLNDCSIILRI<br>IENDELETKKSIYEIFKNINMSLYKIIEKIIENETEKV<br>FENRYYEEHLREKLLKDDKIDVILTNFMEIREKIKS<br>NLEILGFVKFYLNVGGDKKKSKNKKMLVEKILNI<br>NVDLTVEDIADFVIKELEFWNITKRIEKVKKVNNE<br>FLEKRRNRTYIKSYVLLDKHEKFKIERENKKDKIV<br>KFFVENIKNNSIKEKIEKILAEFKIDELIKKLEKELK<br>KGNCDTEIFGIFKKHYKVNFDSKKFSKKSDEEKEL<br>YKIIYRYLKGRIEKILVNEQKVRLKKMEKIEIEKIL<br>NESILSEKILKRVKQYTLEHIMYLGKLRHNDIDMT<br>TVNTDDFSRLHAKEELDLELITFFASTNMELNKIFS<br>RENINNDENIDFFGGDREKNYVLDKKILNSKIKIIR<br>DLDFIDNKNNITNNFIRKFTKIGTNERNRILHAISKE<br>RDLQGTQDDYNKVINIIQNLKISDEEVSKALNLDV<br>VFKDKKNIITKINDIKISEENNNDIKYLPSFSKVLPEI<br>LNLYRNNPKNEPFDTIETEKIVLNALIYVNKELYK<br>KLILEDDLEENESKNIFLQELKKTLGNIDEIDENIIE<br>NYYKNAQISASKGNNKAIKKYQKKVIECYIGYLR<br>KNYEELFDFSDFKMNIQEIKKQIKDINDNKTYERIT<br>VKTSDKTIVINDDFEYIISIFALLNSNAVINKIRNRFF<br>ATSVWLNTSEYQNIIDILDEIMQLNTLRNECITENW<br>NLNLEEFIQKMKEIEKDFDDFKIQTKKEIFNNYYE<br>DIKNNILTEFKDDINGCDVLEKKLEKIVIFDDETKF<br>EIDKKSNILQDEQRKLSNINKKDLKKKVDQYIKDK<br>DQEIKSKILCRIIFNSDFLKKYKKEIDNLIEDMESEN<br>ENKFQEIYYPKERKNELYIYKKNLFLNIGNPNFDKI<br>YGLISNDIKMADAKFLFNIDGKNIRKNKISEIDAIL<br>KNLNDKLNGYSKEYKEKYIKKLKENDDFFAKNIQ<br>NKNYKSFEKDYNRVSEYKKIRDLVEFNYLNKIES<br>YLIDINWKLAIQMARFERDMHYIVNGLRELGIIKLS<br>GYNTGISRAYPKRNGSDGFYTTTAYYKFFDEESY<br>KKFEKICYGFGIDLSENSEINKPENESIRNYISHFYI<br>VRNPFADYSIAEQIDRVSNLLSYSTRYNNSTYASV<br>FEVFKKDVNLDYDELKKKFKLIGNNDILERLMKP<br>KKVSVLELESYNSDYIKNLIIELLTKIENTNDTL |
| c2c2-3 | | 韦德纤毛菌 (Lw2) (SEQ ID No. 47) | MKVTKVDGISHKKYIEEGKLVKSTSEENRTSERLS<br>ELLSIRLDIYIKNPDNASEEENRIRRENLKKFFSNK<br>VLHLKDSVLYLKNRKEKNAVQDKNYSEEDISEYD<br>LKNKNSFSVLKKILLNEDVNSEELEIFRKDVEAKL<br>NKINSLKYSFEENKANYQKINENNVEKVGGKSKR<br>NIIYDYYRESAKRNDYINNVQEAFDKLYKKEDIEK<br>LFFLIENSKKHEKYKIREYYHKIIGRKNDKENFAKI |

[0477]

| | | | |
|---|---|---|---|
| | | | IYEEIQNVNNIKELIEKIPDMSELKKSQVFYKYYLD<br>KEELNDKNIKYAFCHFVEIEMSQLLKNYVYKRLS<br>NISNDKIKRIFEYQNLKKLIENKLLNKLDTYVRNC<br>GKYNYYLQVGEIATSDFIARNRQNEAFLRNIIGVS<br>SVAYFSLRNILETENENDITGRMRGKTVKNNKGE<br>EKYVSGEVDKIYNENKQNEVKENLKMFYSYDFN<br>MDNKNEIEDFFANIDEAISSIRHGIVHFNLELEGKDI<br>FAFKNIAPSEISKKMFQNEINEKKLKLKIFKQLNSA<br>NVFNYYEKDVIIKYLKNTKFNFVNKNIPFVPSFTK<br>LYNKIEDLRNTLKFFWSVPKDKEEKDAQIYLLKNI<br>YYGEFLNKFVKNSKVFFKITNEVIKINKQRNQKTG<br>HYKYQKFENIEKTVPVEYLAIIQSREMINNQDKEE<br>KNTYIDFIQQIFLKGFIDYLNKNNLKYIESNNNNDN<br>NDIFSKIKIKKDNKEKYDKILKNYEKHNRNKEIPH<br>EINEFVREIKLGKILKYTENLNMFYLILKLLNHKEL<br>TNLKGSLEKYQSANKEETFSDELELINLLNLDNNR<br>VTEDFELEANEIGKFLDFNENKIKDRKELKKFDTN<br>KIYFDGENIIKHRAFYNIKKYGMLNLLEKIADKAK<br>YKISLKELKEYSNKKNEIEKNYTMQQNLHRKYAR<br>PKKDEKFNDEDYKEYEKAIGNIQKYTHLKNKVEF<br>NELNLLQGLLLKILHRLVGYTSIWERDLRFRLKGE<br>FPENHYIEEIFNFDNSKNVKYKSGQIVEKYINFYKE<br>LYKDNVEKRSIYSDKKVKKLKQEKKDLYIRNYIA<br>HFNYIPHAEISLLEVLENLRKLLSYDRKLKNAIMK<br>SIVDILKEYGFVATFKIGADKKIEIQTLESEKIVHLK<br>NLKKKKLMTDRNSEELCELVKVMFEYKALEKRP<br>AATKKAGQAKKKKGSYPYDVPDYAYPYDVPDY<br>AYPYDVPDYA* |
| c2c2-4 | | 西尔李斯特菌 (SEQ ID No. 48) | MWISIKTLIHHLGVLFFCDYMYNRREKKIIEVKTM<br>RITKVEVDRKKVLISRDKNGGKLVYENEMQDNTE<br>QIMHHKKSSFYKSVVNKTICRPEQKQMKKLVHGL<br>LQENSQEKIKVSDVTKLNISNFLNHRFKKSLYYFP<br>ENSPDKSEEYRIEINLSQLLEDSLKKQQGTFICWES<br>FSKDMELYINWAENYISSKTKLIKKSIRNNRIQSTE<br>SRSGQLMDRYMKDILNKNKPFDIQSVSEKYQLEK<br>LTSALKATFKEAKKNDKEINYKLKSTLQNHERQII<br>EELKENSELNQFNIEIRKHLETYFPIKKTNRKVGDI<br>RNLEIGEIQKIVNHRLKNKIVQRILQEGKLASYEIES<br>TVNSNSLQKIKIEEAFALKFINACLFASNNLRNMV<br>YPVCKKDILMIGEFKNSFKEIKHKKFIRQWSQFFS<br>QEITVDDIELASWGLRGAIAPIRNEIIHLKKHSWKK<br>FFNNPTFKVKKSKIINGKTKDVTSEFLYKETLFKD<br>YFYSELDSVPELIINKMESSKILDYYSSDQLNQVFT<br>IPNFELSLLTSAVPFAPSFKRVYLKGFDYQNQDEA<br>QPDYNLKLNIYNEKAFNSEAFQAQYSLFKMVYYQ<br>VFLPQFTTNNDLFKSSVDFILTLNKERKGYAKAFQ<br>DIRKMNKDEKPSEYMSYIQSQLMLYQKKQEEKEK<br>INHFEKFINQVFIKGFNSFIEKNRLTYICHPTKNTVP |

| | | | |
|---|---|---|---|
| | | | ENDNIEIPFHTDMDDSNIAFWLMCKLLDAKQLSEL<br>RNEMIKFSCSLQSTEEISTFTKAREVIGLALLNGEK<br>GCNDWKELFDDKEAWKKNMSLYVSEELLQSLPY<br>TQEDGQTPVINRSIDLVKKYGTETILEKLFSSSDDY<br>KVSAKDIAKLHEYDVTEKIAQQESLHKQWIEKPG<br>LARDSAWTKKYQNVINDISNYQWAKTKVELTQV<br>RHLHQLTIDLLSRLAGYMSIADRDFQFSSNYILERE<br>NSEYRVTSWILLSENKNKNKYNDYELYNLKNASI<br>KVSSKNDPQLKVDLKQLRLTLEYLELFDNRLKEK<br>RNNISHFNYLNGQLGNSILELFDDARDVLSYDRKL<br>KNAVSKSLKEILSSHGMEVTFKPLYQTNHHLKIDK<br>LQPKKIHHLGEKSTVSSNQVSNEYCQLVRTLLTM<br>K |

[0478]

| | | | |
|---|---|---|---|
| C2-17 | | 口腔纤毛菌 (Leptotrichia buccalis) C-1013-b (SEQ ID No. 49) | MKVTKVGGISHKKYTSEGRLVKSESEENRTDERLS<br>ALLNMRLDMYIKNPSSTETKENQKRIGKLKKFFSN<br>KMVYLKDNTLSLKNGKKENIDREYSETDILESDVR<br>DKKNFAVLKKIYLNENVNSEELEVFRNDIKKKLNK<br>INSLKYSFEKNKANYQKINENNIEKVEGKSKRNIIY<br>DYYRESAKRDAYVSNVKEAFDKLYKEEDIAKLVL<br>EIENLTKLEKYKIREFYHEIIGRKNDKENFAKIIYEEI<br>QNVNNMKELIEKVPDMSELKKSQVFYKYYLDKEE<br>LNDKNIKYAFCHFVEIEMSQLLKNYVYKRLSNISN<br>DKIKRIFEYQNLKKLIENKLLNKLDTYVRNCGKYN<br>YYLQDGEIATSDFIARNRQNEAFLRNIIGVSSVAYF<br>SLRNILETENENDITGRMRGKTVKNNKGEEKYVSG<br>EVDKIYNENKKNEVKENLKMFYSYDFNMDNKNEI<br>EDFFANIDEAISSIRHGIVHFNLELEGKDIFAFKNIAP<br>SEISKKMFQNEINEKKLKLKIFRQLNSANVFRYLEK<br>YKILNYLKRTRFEFVNKNIPFVPSFTKLYSRIDDLKN<br>SLGIYWKTPKTNDDNKTKEIIDAQIYLLKNIYYGEF<br>LNYFMSNNGNFFEISKEIIELNKNDKRNLKTGFYKL<br>QKFEDIQEKIPKEYLANIQSLYMINAGNQDEEEKDT<br>YIDFIQKIFLKGFMTYLANNGRLSLIYIGSDEETNTS<br>LAEKKQEFDKFLKKYEQNNNIKIPYEINEFLREIKL<br>GNILKYTERLNMFYLILKLLNHKELTNLKGSLEKY<br>QSANKEEAFSDQLELINLLNLDNNRVTEDFELEAD<br>EIGKFLDFNGNKVKDNKELKKFDTNKIYFDGENIIK<br>HRAFYNIKKYGMLNLLEKIADKAGYKISIEELKKYS<br>NKKNEIEKNHKMQENLHRKYARPRKDEKFTDEDY<br>ESYKQAIENIEEYTHLKNKVEFNELNLLQGLLLRIL<br>HRLVGYTSIWERDLRFRLKGEFPENQYIEEIFNFEN<br>KKNVKYKGGQIVEKYIKFYKELHQNDEVKINKYSS<br>ANIKVLKQEKKDLYIRNYIAHFNYIPHAEISLLEVLE<br>NLRKLLSYDRKLKNAVMKSVVDILKEYGFVATFKI<br>GADKKIGIQTLESEKIVHLKNLKKKKLMTDRNSEE<br>LCKLVKIMFEYKMEEKKSEN |

[0479]

| | | | |
|---|---|---|---|
| C2-18 | | 解半纤维素赫氏菌(Herbinix hemicellulosilytica)(SEQ ID No. 50) | MKLTRRRISGNSVDQKITAAFYRDMSQGLLYYDSE<br>DNDCTDKVIESMDFERSWRGRILKNGEDDKNPFY<br>MFVKGLVGSNDKIVCEPIDVDSDPDNLDILINKNLT<br>GFGRNLKAPDSNDTLENLIRKIQAGIPEEEVLPELK<br>KIKEMIQKDIVNRKEQLLKSIKNNRIPFSLEGSKLVP<br>STKKMKWLFKLIDVPNKTFNEKMLEKYWEIYDYD<br>KLKANITNRLDKTDKKARSISRAVSEELREYHKNL<br>RTNYNRFVSGDRPAAGLDNGGSAKYNPDKEEFLL<br>FLKEVEQYFKKYFPVKSKHSNKSKDKSLVDKYKN<br>YCSYKVVKKEVNRSIINQLVAGLIQQGKLLYYFYY<br>NDTWQEDFLNSYGLSYIQVEEAFKKSVMTSLSWGI<br>NRLTSFFIDDSNTVKFDDITTKKAKEAIESNYFNKL<br>RTCSRMQDHFKEKLAFFYPVYVKDKKDRPDDDIE<br>NLIVLVKNAIESVSYLRNRTFHFKESSLLELLKELD<br>DKNSGQNKIDYSVAAEFIKRDIENLYDVFREQIRSL<br>GIAEYYKADMISDCFKTCGLEFALYSPKNSLMPAF<br>KNVYKRGANLNKAYIRDKGPKETGDQGQNSYKA<br>LEEYRELTWYIEVKNNDQSYNAYKNLLQLIYYHAF<br>LPEVRENEALITDFINRTKEWNRKETEERLNTKNN<br>KKHKNFDENDDITVNTYRYESIPDYQGESLDDYLK<br>VLQRKQMARAKEVNEKEEGNNNYIQFIRDVVVW<br>AFGAYLENKLKNYKNELQPPLSKENIGLNDTLKEL<br>FPEEKVKSPFNIKCRFSISTFIDNKGKSTDNTSAEAV<br>KTDGKEDEKDKKNIKRKDLLCFYLFLRLLDENEIC<br>KLQHQFIKYRCSLKERRFPGNRTKLEKETELLAELE<br>ELMELVRFTMPSIPEISAKAESGYDTMIKKYFKDFI<br>EKKVFKNPKTSNLYYHSDSKTPVTRKYMALLMRS<br>APLHLYKDIFKGYYLITKKECLEYIKLSNIIKDYQNS<br>LNELHEQLERIKLKSEKQNGKDSLYLDKKDFYKVK<br>EYVENLEQVARYKHLQHKINFESLYRIFRIHVDIAA<br>RMVGYTQDWERDMHFLFKALVYNGVLEERRFEAI<br>FNNNDDNNDGRIVKKIQNNLNNKNRELVSMLCWN<br>KKLNKNEFGAIIWKRNPIAHLNHFTQTEQNSKSSLE<br>SLINSLRILLAYDRKRQNAVTKTINDLLLNDYHIRIK<br>WEGRVDEGQIYFNIKEKEDIENEPIIHLKHLHKKDC<br>YIYKNSYMFDKQKEWICNGIKEEVYDKSILKCIGNL<br>FKFDYEDKNKSSANPKHT |
| C2-19 | | 直肠真杆菌([Eubacterium] rectale)(SEQ ID No. 51) | MLRRDKEVKKLYNVFNQIQVGTKPKKWNNDEKL<br>SPEENERRAQQKNIKMKNYKWREACSKYVESSQR<br>IINDVIFYSYRKAKNKLRYMRKNEDILKKMQEAEK<br>LSKFSGGKLEDFVAYTLRKSLVVSKYDTQEFDSLA<br>AMVVFLECIGKNNISDHEREIVCKLLELIRKDFSKL<br>DPNVKGSQGANIVRSVRNQNMIVQPQGDRFLFPQ<br>VYAKENETVTNKNVEKEGLNEFLLNYANLDDEKR<br>AESLRKLRRILDVYFSAPNHYEKDMDITLSDNIEKE<br>KFNVWEKHECGKKETGLFVDIPDVLMEAEAENIKL<br>DAVVEKRERKVLNDRVRKQNIICYRYTRAVVEKY<br>NSNEPLFFENNAINQYWIHHIENAVERILKNCKAGK |

[0480]

| | | | |
|---|---|---|---|
| | | | LFKLRKGYLAEKVWKDAINLISIKYIALGKAVYNF<br>ALDDIWKDKKNKELGIVDERIRNGITSFDYEMIKA<br>HENLQRELAVDIAFSVNNLARAVCDMSNLGNKES<br>DFLLWKRNDIADKLKNKDDMASVSAVLQFFGGKS<br>SWDINIFKDAYKGKKKYNYEVRFIDDLRKAIYCAR<br>NENFHFKTALVNDEKWNTELFGKIFERETEFCLNV<br>EKDRFYSNNLYMFYQVSELRNMLDHLYSRSVSRA<br>AQVPSYNSVIVRTAFPEYITNVLGYQKPSYDADTL<br>GKWYSACYYLLKEIYYNSFLQSDRALQLFEKSVKT<br>LSWDDKKQQRAVDNFKDHFSDIKSACTSLAQVCQ<br>IYMTEYNQQNNQIKKVRSSNDSIFDQPVYQHYKVL<br>LKKAIANAFADYLKNNKDLFGFIGKPFKANEIREID<br>KEQFLPDWTSRKYEALCIEVSGSQELQKWYIVGKF<br>LNARSLNLMVGSMRSYIQYVTDIKRRAASIGNELH<br>VSVHDVEKVEKWVQVIEVCSLLASRTSNQFEDYF<br>NDKDDYARYLKSYVDFSNVDMPSEYSALVDFSNE<br>EQSDLYVDPKNPKVNRNIVHSKLFAADHILRDIVEP<br>VSKDNIEEFYSQKAEIAYCKIKGKEITAEEQKAVLK<br>YQKLKNRVELRDIVEYGEIINELLGQLINWSFMRER<br>DLLYFQLGFHYDCLRNDSKKPEGYKNIKVDENSIK<br>DAILYQIIGMYVNGVTVYAPEKDGDKLKEQCVKG<br>GVGVKVSAFHRYSKYLGLNEKTLYNAGLEIFEVV<br>AEHEDIINLRNGIDHFKYYLGDYRSMLSIYSEVFDR<br>FFTYDIKYQKNVLNLLQNILLRHNVIVEPILESGFKT<br>IGEQTKPGAKLSIRSIKSDTFQYKVKGGTLITDAKD<br>ERYLETIRKILYYAENEEDNLKKSVVVTNADKYEK<br>NKESDDQNKQKEKKNKDNKGKKNEETKSDAEKN<br>NNERLSYNPFANLNFKLSN |
| C2-20 | | 真杆菌科菌 (Eubacteriaceae bacterium) CHKCI004 (SEQ ID No. 52) | MKISKESHKRTAVAVMEDRVGGVVYVPGGSGIDL<br>SNNLKKRSMDTKSLYNVFNQIQAGTAPSEYEWKD<br>YLSEAENKKREAQKMIQKANYELRRECEDYAKKA<br>NLAVSRIIFSKKPKKIFSDDDIISHMKKQRLSKFKGR<br>MEDFVLIALRKSLVVSTYNQEVFDSRKAATVFLKN<br>IGKKNISADDERQIKQLMALIREDYDKWNPDKDSS<br>DKKESSGTKVIRSIEHQNMVIQPEKNKLSLSKISNV<br>GKKTKTKQKEKAGLDAFLKEYAQIDENSRMEYLK<br>KLRRLLDTYFAAPSSYIKGAAVSLPENINFSSELNV<br>WERHEAAKKVNINFVEIPESLLNAEQNNNKINKVE<br>QEHSLEQLRTDIRRRNITCYHFANALAADERYHTL<br>FFENMAMNQFWIHHMENAVERILKKCNVGTLFKL<br>RIGYLSEKVWKDMLNLLSIKYIALGKAVYHFALDD<br>IWKADIWKDASDKNSGKINDLTLKGISSFDYEMVK<br>AQEDLQREMAVGVAFSTNNLARVTCKMDDLSDA<br>ESDFLLWNKEAIRRHVKYTEKGEILSAILQFFGGRS<br>LWDESLFEKAYSDSNYELKFLDDLKRAIYAARNET<br>FHFKTAAIDGGSWNTRLFGSLFEKEAGLCLNVEKN<br>KFYSNNLVLFYKQEDLRVFLDKLYGKECSRAAQIP<br>SYNTILPRKSFSDFMKQLLGLKEPVYGSAILDQWY |

[0481]

| | | | |
|---|---|---|---|
| | | | SACYYLFKEVYYNLFLQDSSAKALFEKAVKALKG<br>ADKKQEKAVESFRKRYWEISKNASLAEICQSYITE<br>YNQQNNKERKVRSANDGMFNEPIYQHYKMLLKE<br>ALKMAFASYIKNDKELKFVYKPTEKLFEVSQDNFL<br>PNWNSEKYNTLISEVKNSPDLQKWYIVGKFMNAR<br>MLNLLLGSMRSYLQYVSDIQKRAAGLGENQLHLS<br>AENVGQVKKWIQVLEVCLLLSVRISDKFTDYFKDE<br>EEYASYLKEYVDFEDSAMPSDYSALLAFSNEGKID<br>LYVDASNPKVNRNIIQAKLYAPDMVLKKVVKKISQ<br>DECKEFNEKKEQIMQFKNKGDEVSWEEQQKILEY<br>QKLKNRVELRDLSEYGELINELLGQLINWSYLRER<br>DLLYFQLGFHYSCLMNESKKPDAYKTIRRGTVSIE<br>NAVLYQIIAMYINGFPVYAPEKGELKPQCKTGSAG<br>QKIRAFCQWASMVEKKKYELYNAGLELFEVVKEH<br>DNIIDLRNKIDHFKYYQGNDSILALYGEIFDRFFTYD<br>MKYRNNVLNHLQNILLRHNVIIKPIISKDKKEVGRG<br>KMKDRAAFLLEEVSSDRFTYKVKEGERKIDAKNR<br>LYLETVRDILYFPNRAVNDKGEDVIICSKKAQDLNE<br>KKADRDKNHDKSKDTNQKKEGKNQEEKSENKEP<br>YSDRMTWKPFAGIKLE |
| C2-21 | | 布劳特氏菌属某种(Blautia sp.) Marseille-P 2398 (SEQ ID No. 53) | MKISKVDHVKSGIDQKLSSQRGMLYKQPQKKYEG<br>KQLEEHVRNLSRKAKALYQVFPVSGNSKMEKELQI<br>INSFIKNILLRLDSGKTSEEIVGYINTYSVASQISGDH<br>IQELVDQHLKESLRKYTCVGDKRIYVPDIIVALLKS<br>KFNSETLQYDNSELKILIDFIREDYLKEKQIKQIVHSI<br>ENNSTPLRIAEINGQKRLIPANVDNPKKSYIFEFLKE<br>YAQSDPKGQESLLQHMRYLILLYLYGPDKITDDYC<br>EEIEAWNFGSIVMDNEQLFSEEASMLIQDRIYVNQQ<br>IEEGRQSKDTAKVKKNKSKYRMLGDKIEHSINESV<br>VKHYQEACKAVEEKDIPWIKYISDHVMSVYSSKN<br>RVDLDKLSLPYLAKNTWNTWISFIAMKYVDMGKG<br>VYHFAMSDVDKVGKQDNLIIGQIDPKFSDGISSFDY<br>ERIKAEDDLHRSMSGYIAFAVNNFARAICSDEFRK<br>KNRKEDVLTVGLDEIPLYDNVKRKLLQYFGGASN<br>WDDSIIDIIDDKDLVACIKENLYVARNVNFHFAGSE<br>KVQKKQDDILEEIVRKETRDIGKHYRKVFYSNNVA<br>VFYCDEDIIKLMNHLYQREKPYQAQIPSYNKVISKT<br>YLPDLIFMLLKGKNRTKISDPSIMNMFRGTFYFLLK<br>EIYYNDFLQASNLKEMFCEGLKNNVKNKKSEKPY<br>QNFMRRFEELENMGMDFGEICQQIMTDYEQQNKQ<br>KKKTATAVMSEKDKKIRTLDNDTQKYKHFRTLLYI<br>GLREAFIIYLKDEKNKEWYEFLREPVKREQPEEKEF<br>VNKWKLNQYSDCSELILKDSLAAAWYVVAHFINQ<br>AQLNHLIGDIKNYIQFISDIDRRAKSTGNPVSESTEI<br>QIERYRKILRVLEFAKFFCGQITNVLTDYYQDENDF<br>STHVGHYVKFEKKNMEPAHALQAFSNSLYACGKE<br>KKKAGFYYDGMNPIVNRNITLASMYGNKKLLENA<br>MNPVTEQDIRKYYSLMAELDSVLKNGAVCKSEDE |

[0482]

| | | | |
|---|---|---|---|
| | | | QKNLRHFQNLKNRIELVDVLTLSELVNDLVAQLIG WVYIRERDMMYLQLGLHYIKLYFTDSVAEDSYLR TLDLEEGSIADGAVLYQIASLYSFNLPMYVKPNKSS VYCKKHVNSVATKFDIFEKEYCNGDETVIENGLRL FENINLHKDMVKFRDYLAHFKYFAKLDESILELYS KAYDFFFSYNIKLKKSVSYVLTNVLLSYFINAKLSF STYKSSGNKTVQHRTTKISVVAQTDYFTYKLRSIV KNKNGVESIENDDRRCEVVNIAARDKEFVDEVCN VINYNSDK |
| C2-22 | | 纤毛菌属某种(Leptotrichia sp.)口腔分类群879菌株F0557(SEQ ID No. 54) | MGNLFGHKRWYEVRDKKDFKIKRKVKVKRNYDG NKYILNINENNNKEKIDNNKFIGEFVNYKKNNNVL KEFKRKFHAGNILFKLKGKEEIIRIENNDDFLETEEV VLYIEVYGKSEKLKALEITKKKIIDEAIRQGITKDDK KIEIKRQENEEEIEIDIRDEYTNKTLNDCSIILRIIEND ELETKKSIYEIFKNINMSLYKIIEKIIENETEKVFENR YYEEHLREKLLKDNKIDVILTNFMEIREKIKSNLEIM GFVKFYLNVSGDKKKSENKKMFVEKILNTNVDLT VEDIVDFIVKELKFWNITKRIEKVKKFNNEFLENRR NRTYIKSYVLLDKHEKFKIERENKKDKIVKFFVENI KNNSIKEKIEKILAEFKINELIKKLEKELKKGNCDTE IFGIFKKHYKVNFDSKKFSNKSDEEKELYKIIYRYL KGRIEKILVNEQKVRLKKMEKIEIEKILNESILSEKIL KRVKQYTLEHIMYLGKLRHNDIVKMTVNTDDFSR LHAKEELDLELITFFASTNMELNKIFNGKEKVTDFF GFNLNGQKITLKEKVPSFKLNILKKLNFINNENNIDE KLSHFYSFQKEGYLLRNKILHNSYGNIQETKNLKG EYENVEKLIKELKVSDEEISKSLSLDVIFEGKVDIIN KINSLKIGEYKDKKYLPSFSKIVLEITRKFREINKDK LFDIESEKIILNAVKYVNKILYEKITSNEENEFLKTLP DKLVKKSNNKKENKNLLSIEEYYKNAQVSSSKGD KKAIKKYQNKVTNAYLEYLENTFTEIIDFSKFNLNY DEIKTKIEERKDNKSKIIIDSISTNINITNDIEYIISIFAL LNSNTYINKIRNRFFATSVWLEKQNGTKEYDYENII SILDEVLLINLLRENNITDILDLKNAIIDAKIVENDET YIKNYIFESNEEKLKKRLFCEELVDKEDIRKIFEDEN FKFKSFIKKNEIGNFKINFGILSNLECNSEVEAKKIIG KNSKKLESFIQNIIDEYKSNIRTLFSSEFLEKYKEEID NLVEDTESENKNKFEKIYYPKEHKNELYIYKKNLF LNIGNPNFDKIYGLISKDIKNVDTKILFDDDIKKNKI SEIDAILKNLNDKLNGYSNDYKAKYVNKLKENDD FFAKNIQNENYSSFGEFEKDYNKVSEYKKIRDLVEF NYLNKIESYLIDINWKLAIQMARFERDMHYIVNGL RELGIIKLSGYNTGISRAYPKRNGSDGFYTTTAYYK FFDEESYKKFEKICYGFGIDLSENSEINKPENESIRN YISHFYIVRNPFADYSIAEQIDRVSNLLSYSTRYNNS TYASVFEVFKKDVNLDYDELKKKFRLIGNNDILER LMKPKKVSVLELESYNSDYIKNLIIELLTKIENTNDT L |

[0483]

| | | | |
|---|---|---|---|
| C2-23 | | 毛螺菌科细菌 NK4A144 (SEQ ID No. 55) | MKISKVDHTRMAVAKGNQHRRDEISGILYKDPTK<br>TGSIDFDERFKKLNCSAKILYHVFNGIAEGSNKYKN<br>IVDKVNNNLDRVLFTGKSYDRKSIIDIDTVLRNVEK<br>INAFDRISTEEREQIIDDLLEIQLRKGLRKGKAGLRE<br>VLLIGAGVIVRTDKKQEIADFLEILDEDFNKTNQAK<br>NIKLSIENQGLVVSPVSRGEERIFDVSGAQKGKSSK<br>KAQEKEALSAFLLDYADLDKNVRFEYLRKIRRLIN<br>LYFYVKNDDVMSLTEIPAEVNLEKDFDIWRDHEQR<br>KEENGDFVGCPDILLADRDVKKSNSKQVKIAERQL<br>RESIREKNIKRYRFSIKTIEKDDGTYFFANKQISVFW<br>IHRIENAVERILGSINDKKLYRLRLGYLGEKVWKDI<br>LNFLSIKYIAVGKAVFNFAMDDLQEKDRDIEPGKIS<br>ENAVNGLTSFDYEQIKADEMLQREVAVNVAFAAN<br>NLARVTVDIPQNGEKEDILLWNKSDIKKYKKNSKK<br>GILKSILQFFGGASTWNMKMFEIAYHDQPGDYEEN<br>YLYDIIQIIYSLRNKSFHFKTYDHGDKNWNRELIGK<br>MIEHDAERVISVEREKFHSNNLPMFYKDADLKKIL<br>DLLYSDYAGRASQVPAFNTVLVRKNFPEFLRKDM<br>GYKVHFNNPEVENQWHSAVYYLYKEIYYNLFLRD<br>KEVKNLFYTSLKNIRSEVSDKKQKLASDDFASRCE<br>EIEDRSLPEICQIIMTEYNAQNFGNRKVKSQRVIEK<br>NKDIFRHYKMLLIKTLAGAFSLYLKQERFAFIGKAT<br>PIPYETTDVKNFLPEWKSGMYASFVEEIKNNLDLQ<br>EWYIVGRFLNGRMLNQLAGSLRSYIQYAEDIERRA<br>AENRNKLFSKPDEKIEACKKAVRVLDLCIKISTRIS<br>AEFTDYFDSEDDYADYLEKYLKYQDDAIKELSGSS<br>YAALDHFCNKDDLKFDIYVNAGQKPILQRNIVMA<br>KLFGPDNILSEVMEKVTESAIREYYDYLKKVSGYR<br>VRGKCSTEKEQEDLLKFQRLKNAVEFRDVTEYAE<br>VINELLGQLISWSYLRERDLLYFQLGFHYMCLKNK<br>SFKPAEYVDIRRNNGTIIHNAILYQIVSMYINGLDFY<br>SCDKEGKTLKPIETGKGVGSKIGQFIKYSQYLYNDP<br>SYKLEIYNAGLEVFENIDEHDNITDLRKYVDHFKY<br>YAYGNKMSLLDLYSEFFDRFFTYDMKYQKNVVN<br>VLENILLRHFVIFYPKFGSGKKDVGIRDCKKERAQI<br>EISEQSLTSEDFMFKLDDKAGEEAKKFPARDERYL<br>QTIAKLLYYPNEIEDMNRFMKKGETINKKVQFNRK<br>KKITRKQKNNSSNEVLSSTMGYLFKNIKL |
| C2-24 | | 凝集绿屈挠菌 (Chloroflexus aggregans) (SEQ ID No. 56) | MTDQVRREEVAAGELADTPLAAAQTPAADAAVA<br>ATPAPAEAVAPTPEQAVDQPATTGESEAPVTTAQA<br>AAHEAEPAEATGASFTPVSEQQPQKPRRLKDLQPG<br>MELEGKVTSIALYGIFVDVGVGRDGLVHISEMSDR<br>RIDTPSELVQIGDTVKVWVKSVDLDARRISLTMLN<br>PSRGEKPRRSRQSQPAQPQPRRQEVDREKLASLKV<br>GEIVEGVITGFAPFGAFADIGVGKDGLIHISELSEGR<br>VEKPEDAVKVGERYQFKVLEIDGEGTRISLSLRRA<br>QRTQRMQQLEPGQIIEGTVSGIATFGAFVDIGVGRD<br>GLVHISALAPHRVAKVEDVVKVGDKVKVKVLGV |

[0484]

| | | | |
|---|---|---|---|
| | | | DPQSKRISLTMRLEEEQPATTAGDEAAEPAEEVTPT RRGNLERFAAAAQTARERSERGERSERGERRERRE RRPAQSSPDTYIVGEDDDESFEGNATIEDLLTKFGG SSSRRDRDRRRRHEDDDDEEMERPSNRRQREAIRR TLQQIGYDE |
| C2-25 | | 桔红色脱甲基菌(Demequina aurantiaca)(SEQ ID No. 57) | MDLTWHALLILFIVALLAGFLDTLAGGGGLLTVPA LLLTGIPPLQALGTNKLQSSFGTGMATYQVIRKKR VHWRDVRWPMVWAFLGSAAGAVAVQFIDTDALL IIIPVVLALVAAYFLFVPKSHLPPPEPRMSDPAYEAT LVPIIGAYDGAFGPGTGSLYALSGVALRAKTLVQS TAIAKTLNFATNFAALLVFAFAGHMLWTVGAVMI AGQLIGAYAGSHMLFRVNPLVLRVLIVVMSLGMLI RVLLD |
| C2-26 | | 海旋菌属某种(Thalassospira sp.)TSL5-1(SEQ ID No. 58) | MRIIKPYGRSHVEGVATQEPRRKLRLNSSPDISRDIP GFAQSHDALIIAQWISAIDKIATKPKPDKKPTQAQIN LRTTLGDAAWQHVMAENLLPAATDPAIREKLHLI WQSKIAPWGTARPQAEKDGKPTPKGGWYERFCG VLSPEAITQNVARQIAKDIYDHLHVAAKRKGREPA KQGESSNKPGKFKPDRKRGLIEERAESIAKNALRPG SHAPCPWGPDDQATYEQAGDVAGQIYAAARDCLE EKKRRSGNRNTSSVQYLPRDLAAKILYAQYGRVF GPDTTIKAALDEQPSLFALHKAIKDCYHRLINDARK RDILRILPRNMAALFRLVRAQYDNRDINALIRLGKV IHYHASEQGKSEHHGIRDYWPSQQDIQNSRFWGSD GQADIKRHEAFSRIWRHIIALASRTLHDWADPHSQ KFSGENDDILLLAKDAIEDDVFKAGHYERKCDVLF GAQASLFCGAEDFEKAILKQAITGTGNLRNATFHF KGKVRFEKELQELTKDVPVEVQSAIAALWQKDAE GRTRQIAETLQAVLAGHFLTEEQNRHIFAALTAAM AQPGDVPLPRLRRVLARHDSICQRGRILPLSPCPDR AKLEESPALTCQYTVLKMLYDGPFRAWLAQQNSTI LNHYIDSTIARTDKAARDMNGRKLAQAEKDLITSR AADLPRLSVDEKMGDFLARLTAATATEMRVQRGY QSDGENAQKQAAFIGQFECDVIGRAFADFLNQSGF DFVLKLKADTPQPDAAQCDVTALIAPDDISVSPPQ AWQQVLYFILHLVPVDDASHLLHQIRKWQVLEGK EKPAQIAHDVQSVLMLYLDMHDAKFTGGAALHGI EKFAEFFAHAADFRAVFPPQSLQDQDRSIPRRGLRE IVRFGHLPLLQHMSGTVQITHDNVVAWQAARTAG ATGMSPIARRQKQREELHALAVERTARFRNADLQ NYMHALVDVIKHRQLSAQVTLSDQVRLHRLMMG VLGRLVDYAGLWERDLYFVVLALLYHHGATPDD VFKGQGKKNLADGQVVAALKPKNRKAAAPVGVF DDLDHYGIYQDDRQSIRNGLSHFNMLRGGKAPDLS HWVNQTRSLVAHDRKLKNAVAKSVIEMLAREGF DLDWGIQTDRGQHILSHGKIRTRQAQHFQKSRLHI VKKSAKPDKNDTVKIRENLHGDAMVERVVQLFAA |

[0485]

| | | | |
|---|---|---|---|
| | | | QVQKRYDITVEKRLDHLFLKPQDQKGKNGIHTHN GWSKTEKKRRPSRENRKGNHEN |
| C2-27 | | SAMN04487830_13920 [假丁酸弧菌属某种 (Pseudobutyrivibrio sp.) OR37] (SEQ ID No. 59) | MKFSKESHRKTAVGVTESNGIIGLLYKDPLNEKEKI EDVVNQRANSTKRLFNLFGTEATSKDISRASKDLA KVVNKAIGNLKGNKKFNKKEQITKGLNTKIIVEEL KNVLKDEKKLIVNKDIIDEACSRLLKTSFRTAKTKQ AVKMILTAVLIENTNLSKEDEAFVHEYFVKKLVNE YNKTSVKKQIPVALSNQNMVIQPNSVNGTLEISETK KSKETKTTEKDAFRAFLRDYATLDENRRHKMRLC LRNLVNLYFYGETSVSKDDFDEWRDHEDKKQNDE LFVKKIVSIKTDRKGNVKEVLDVDATIDAIRTNNIA CYRRALAYANENPDVFFSDTMLNKFWIHHVENEV ERIYGHINNNTGDYKYQLGYLSEKVWKGIINYLSIK YIAEGKAVYNYAMNALAKDNNSNAFGKLDEKFV NGITSFEYERIKAEETLQRECAVNIAFAANHLANAT VDLNEKDSDFLLLKHEDNKDTLGAVARPNILRNIL QFFGGKSRWNDFDFSGIDEIQLLDDLRKMIYSLRNS SFHFKTENIDNDSWNTKLIGDMFAYDFNMAGNVQ KDKMYSNNVPMFYSTSDIEKMLDRLYAEVHERAS QVPSFNSVFVRKNFPDYLKNDLKITSAFGVDDALK WQSAVYYVCKEIYYNDFLQNPETFTMLKDYVQCL PIDIDKSMDQKLKSERNAHKNFKEAFATYCKECDS LSAICQMIMTEYNNQNKGNRKVISARTKDGDKLIY KHYKMILFEALKNVFTIYLEKNINTYGFLKKPKLIN NVPAIEEFLPNYNGRQYETLVNRITEETELQKWYIV GRLLNPKQVNQLIGNFRSYVQYVNDVARRAKQTG NNLSNDNIAWDVKNIIQIFDVCTKLNGVTSNILEDY FDDGDDYARYLKNFVDYTNKNNDHSATLLGDFC AKEIDGIKIGIYHDGTNPIVNRNIIQCKLYGATGIISD LTKDGSILSVDYEIIKKYMQMQKEIKVYQQKGICK TKEEQQNLKKYQELKNIVELRNIIDYSEILDELQGQ LINWGYLRERDLMYFQLGFHYLCLHNESKKPVGY NNAGDISGAVLYQIVAMYTNGLSLIDANGKSKKN AKASAGAKVGSFCSYSKEIRGVDKDTKEDDDPIYL AGVELFENINEHQQCINLRNYIEHFHYYAKHDRSM LDLYSEVFDRFFTYDMKYTKNVPNMMYNILLQHL VVPAFEFGSSEKRLDDNDEQTKPRAMFTLREKNGL SSEQFTYRLGDGNSTVKLSARGDDYLRAVASLLY YPDRAPEGLIRDAEAEDKFAKINHSNPKSDNRNNR GNFKNPKVQWYNNKTKRK |
| C2-28 | | SAMN02910398_00008 [丁酸弧菌属某种 (Butyrivibrio sp.) YAB3001] | MKISKVDHRKTAVKITDNKGAEGFIYQDPTRDSST MEQIISNRARSSKVLFNIFGDTKKSKDLNKYTESLII YVNKAIKSLKGDKRNNKYEEITESLKTERVLNALIQ AGNEFTCSENNIEDALNKYLKKSFRVGNTKSALKK LLMAAYCGYKLSIEEKEEIQNYFVDKLVKEYNKDT VLKYTAKSLKHQNMVVQPDTDNHVFLPSRIAGAT QNKMSEKEALTEFLKAYAVLDEEKRHNLRIILRKL |

[0486]

| | | | |
|---|---|---|---|
| | | (SEQ ID No. 60) | VNLYFYESPDFIYPENNEWKEHDDRKNKTETFVSP<br>VKVNEEKNGKTFVKIDVPATKDLIRLKNIECYRRS<br>VAETAGNPITYFTDHNISKFWIHHIENEVEKIFALLK<br>SNWKDYQFSVGYISEKVWKEIINYLSIKYIAIGKAV<br>YNYALEDIKKNDGTLNFGVIDPSFYDGINSFEYEKI<br>KAEETFQREVAVYVSFAVNHLSSATVKLSEAQSD<br>MLVLNKNDIEKIAYGNTKRNILQFFGGQSKWKEFD<br>FDRYINPVNYTDIDFLFDIKKMVYSLRNESFHFTTT<br>DTESDWNKNLISAMFEYECRRISTVQKNKFFSNNL<br>PLFYGENSLERVLHKLYDDYVDRMSQVPSFGNVF<br>VRKKFPDYMKEIGIKHNLSSEDNLKLQGALYFLYK<br>EIYYNAFISSEKAMKIFVDLVNKLDTNARDDKGRIT<br>HEAMAHKNFKDAISHYMTHDCSLADICQKIMTEY<br>NQQNTGHRKKQTTYSSEKNPEIFRHYKMILFMLLQ<br>KAMTEYISSEEIFDFIMKPNSPKTDIKEEEFLPQYKS<br>CAYDNLIKLIADNVELQKWYITARLLSPREVNQLIG<br>SFRSYKQFVSDIERRAKETNNSLSKSGMTVDVENIT<br>KVLDLCTKLNGRFSNELTDYFDSKDDYAVYVSKF<br>LDFGFKIDEKFPAALLGEFCNKEENGKKIGIYHNGT<br>EPILNSNIIKSKLYGITDVVSRAVKPVSEKLIREYLQ<br>QEVKIKPYLENGVCKNKEEQAALRKYQELKNRIEF<br>RDIVEYSEIINELMGQLINFSYLRERDLMYFQLGFH<br>YLCLNNYGAKPEGYYSIVNDKRTIKGAILYQIVAM<br>YTYGLPIYHYVDGTISDRRKNKKTVLDTLNSSETV<br>GAKIKYFIYYSDELFNDSLILYNAGLELFENINEHEN<br>IVNLRKYIDHFKYYVSQDRSLLDIYSEVFDRYFTYD<br>RKYKKNVMNLFSNIMLKHFIITDFEFSTGEKTIGEK<br>NTAKKECAKVRIKRGGLSSDKFTYKFKDAKPIELS<br>AKNTEFLDGVARILYYPENVVLTDLVRNSEVEDEK<br>RIEKYDRNHNSSPTRKDKTYKQDVKKNYNKKTSK<br>AFDSSKLDTKSVGNNLSDNPVLKQFLSESKKKR |
| | | | |
| C2-29 | | 布劳特氏菌属某种 Marseille-P 2398 (SEQ ID No. 61) | MKISKVDHVKSGIDQKLSSQRGMLYKQPQKKYEG<br>KQLEEHVRNLSRKAKALYQVFPVSGNSKMEKELQI<br>INSFIKNILLRLDSGKTSEEIVGYINTYSVASQISGDH<br>IQELVDQHLKESLRKYTCVGDKRIYVPDIIVALLKS<br>KFNSETLQYDNSELKILIDFIREDYLKEKQIKQIVHSI<br>ENNSTPLRIAEINGQKRLIPANVDNPKKSYIFEFLKE<br>YAQSDPKGQESLLQHMRYLILLYLYGPDKITDDYC<br>EEIEAWNFGSIVMDNEQLFSEEASMLIQDRIYVNQQ<br>IEEGRQSKDTAKVKKNKSKYRMLGDKIEHSINESV<br>VKHYQEACKAVEEKDIPWIKYISDHVMSVYSSKN<br>RVDLDKLSLPYLAKNTWNTWISFIAMKYVDMGKG<br>VYHFAMSDVDKVGKQDNLIIGQIDPKFSDGISSFDY<br>ERIKAEDDLHRSMSGYIAFAVNNFARAICSDEFRK<br>KNRKEDVLTVGLDEIPLYDNVKRKLLQYFGGASN<br>WDDSIIDIIDDKDLVACIKENLYVARNVNFHFAGSE<br>KVQKKQDDILEEIVRKETRDIGKHYRKVFYSNNVA<br>VFYCDEDIIKLMNHLYQREKPYQAQIPSYNKVISKT |

| | | | | |
|---|---|---|---|---|
| | | | | YLPDLIFMLLKGKNRTKISDPSIMNMFRGTFYFLLK<br>EIYYNDFLQASNLKEMFCEGLKNNVKNKKSEKPY<br>QNFMRRFEELENMGMDFGEICQQIMTDYEQQNKQ<br>KKKTATAVMSEKDKKIRTLDNDTQKYKHFRTLLYI<br>GLREAFIIYLKDEKNKEWYEFLREPVKREQPEEKEF<br>VNKWKLNQYSDCSELILKDSLAAAWYVVAHFINQ<br>AQLNHLIGDIKNYIQFISDIDRRAKSTGNPVSESTEI<br>QIERYRKILRVLEFAKFFCGQITNVLTDYYQDENDF<br>STHVGHYVKFEKKNMEPAHALQAFSNSLYACGKE<br>KKKAGFYYDGMNPIVNRNITLASMYGNKKLLENA<br>MNPVTEQDIRKYYSLMAELDSVLKNGAVCKSEDE<br>QKNLRHFQNLKNRIELVDVLTLSELVNDLVAQLIG<br>WVYIRERDMMYLQLGLHYIKLYFTDSVAEDSYLR<br>TLDLEEGSIADGAVLYQIASLYSFNLPMYVKPNKSS<br>VYCKKHVNSVATKFDIFEKEYCNGDETVIENGLRL<br>FENINLHKDMVKFRDYLAHFKYFAKLDESILELYS<br>KAYDFFFSYNIKLKKSVSYVLTNVLLSYFINAKLSF<br>STYKSSGNKTVQHRTTKISVVAQTDYFTYKLRSIV<br>KNKNGVESIENDDRRCEVVNIAARDKEFVDEVCN<br>VINYNSDK |
| [0487] | C2-30 | | 纤毛菌属某种 Marseille-P 3007 (SEQ ID No. 62) | MKITKIDGISHKKYIKEGKLVKSTSEENKTDERLSE<br>LLTIRLDTYIKNPDNASEEENRIRRENLKEFFSNKVL<br>YLKDGILYLKDRREKNQLQNKNYSEEDISEYDLKN<br>KNNFLVLKKILLNEDINSEELEIFRNDFEKKLDKINS<br>LKYSLEENKANYQKINENNIKKVEGKSKRNIFYNY<br>YKDSAKRNDYINNIQEAFDKLYKKEDIENLFFLIEN<br>SKKHEKYKIRECYHKIIGRKNDKENFATIIYEEIQNV<br>NNMKELIEKVPNVSELKKSQVFYKYYLNKEKLND<br>ENIKYVFCHFVEIEMSKLLKNYVYKKPSNISNDKV<br>KRIFEYQSLKKLIENKLLNKLDTYVRNCGKYSFYL<br>QDGEIATSDFIVGNRQNEAFLRNIIGVSSTAYFSLRN<br>ILETENENDITGRMRGKTVKNNKGEEKYISGEIDKL<br>YDNNKQNEVKKNLKMFYSYDFNMNSKKEIEDFFS<br>NIDEAISSIRHGIVHFNLELEGKDIFTFKNIVPSQISK<br>KMFHDEINEKKLKLKIFKQLNSANVFRYLEKYKIL<br>NYLNRTRFEFVNKNIPFVPSFTKLYSRIDDLKNSLGI<br>YWKTPKTNDDNKTKEITDAQIYLLKNIYYGEFLNY<br>FMSNNGNFFEITKEIIELNKNDKRNLKTGFYKLQKF<br>ENLQEKTPKEYLANIQSLYMINAGNQDEEEKDTYI<br>DFIQKIFLKGFMTYLANNGRLSLIYIGSDEETNTSLA<br>EKKQEFDKFLKKYEQNNNIEIPYEINEFVREIKLGKI<br>LKYTERLNMFYLILKLLNHKELTNLKGSLEKYQSA<br>NKEEAFSDQLELINLLNLDNNRVTEDFELEADEIGK<br>FLDFNGNKVKDNKELKKFDTNKIYFDGENIIKHRA<br>FYNIKKYGMLNLLEKISDEAKYKISIEELKNYSKKK<br>NEIEENHTTQENLHRKYARPRKDEKFTDEDYKKYE<br>KAIRNIQQYTHLKNKVEFNELNLLQSLLLRILHRLV<br>GYTSIWERDLRFRLKGEFPENQYIEEIFNFDNSKNV |

[0488]

| | | | |
|---|---|---|---|
| | | | KYKNGQIVEKYINFYKELYKDDTEKISIYSDKKVK ELKKEKKDLYIRNYIAHFNYIPNAEISLLEMLENLR KLLSYDRKLKNAIMKSIVDILKEYGFVVTFKIEKDK KIRIESLKSEEVVHLKKLKLKDNDKKKEPIKTYRNS KELCKLVKVMFEYKMKEKKSEN |
| C2-31 | | 伊华拟杆菌 (Bacteroides ihuae) (SEQ ID No. 63) | MRITKVKVKESSDQKDKMVLIHRKVGEGTLVLDE NLADLTAPIIDKYKDKSFELSLLKQTLVSEKEMNIP KCDKCTAKERCLSCKQREKRLKEVRGAIEKTIGAV IAGRDIIPRLNIFNEDEICWLIKPKLRNEFTFKDVNK QVVKLNLPKVLVEYSKKNDPTLFLAYQQWIAAYL KNKKGHIKKSILNNRVVIDYSDESKLSKRKQALEL WGEEYETNQRIALESYHTSYNIGELVTLLPNPEEYV SDKGEIRPAFHYKLKNVLQMHQSTVFGTNEILCINP IFNENRANIQLSAYNLEVVKYFEHYFPIKKKKKNLS LNQAIYYLKVETLKERLSLQLENALRMNLLQKGKI KKHEFDKNTCSNTLSQIKRDEFFVLNLVEMCAFAA NNIRNIVDKEQVNEILSKKDLCNSLSKNTIDKELCT KFYGADFSQIPVAIWAMRGSVQQIRNEIVHYKAEA IDKIFALKTFEYDDMEKDYSDTPFKQYLELSIEKIDS FFIEQLSSNDVLNYYCTEDVNKLLNKCKLSLRRTSI PFAPGFKTIYELGCHLQDSSNTYRIGHYLMLIGGRV ANSTVTKASKAYPAYRFMLKLIYNHLFLNKFLDNH NKRFFMKAVAFVLKDNRENARNKFQYAFKEIRM MNNDESIASYMSYIHSLSVQEQEKKGDKNDKVRY NTEKFIEKVFVKGFDDFLSWLGVEFILSPNQEERDK TVTREEYENLMIKDRVEHSINSNQESHIAFFTFCKL LDANHLSDLRNEWIKFRSSGDKEGFSYNFAIDIIELC LLTVDRVEQRRDGYKEQTELKEYLSFFIKGNESEN TVWKGFYFQQDNYTPVLYSPIELIRKYGTLELLKLII VDEDKITQGEFEEWQTLKKVVEDKVTRRNELHQE WEDMKNKSSFSQEKCSIYQKLCRDIDRYNWLDNK LHLVHLRKLHNLVIQILSRMARFIALWDRDFVLLD ASRANDDYKLLSFFNFRDFINAKKTKTDDELLAEF GSKIEKKNAPFIKAEDVPLMVECIEAKRSFYQKVFF RNNLQVLADRNFIAHYNYISKTAKCSLFEMIIKLRT LMYYDRKLRNAVVKSIANVFDQNGMVLQLSLDDS HELKVDKVISKRIVHLKNNNIMTDQVPEEYYKICR RLLEMKK |
| C2-32 | | SAMN05216357_1045 [紫单孢菌科细菌 (Porphyromonadaceae bacterium) KH3CP3R | MEFRDSIFKSLLQKEIEKAPLCFAEKLISGGVFSYYP SERLKEFVGNHPFSLFRKTMPFSPGFKRVMKSGGN YQNANRDGRFYDLDIGVYLPKDGFGDEEWNARYF LMKLIYNQLFLPYFADAENHLFRECVDFVKRVNRD YNCKNNNSEEQAFIDIRSMREDESIADYLAFIQSNIII EENKKKETNKEGQINFNKFLLQVFVKGFDSFLKDR TELNFLQLPELQGDGTRGDDLESLDKLGAVVAVDL KLDATGIDADLNENISFYTFCKLLDSNHLSRLRNEII KYQSANSDFSHNEDFDYDRIISIIELCMLSADHVST |

| | | | |
|---|---|---|---|
| | | A] (SEQ ID No. 64) | NDNESIFPNNDKDFSGIRPYLSTDAKVETFEDLYVH SDAKTPITNATMVLNWKYGTDKLFERLMISDQDFL VTEKDYFVWKELKKDIEEKIKLREELHSLWVNTPK GKKGAKKKNGRETTGEFSEENKKEYLEVCREIDR YVNLDNKLHFVHLKRMHSLLIELLGRFVGFTYLFE RDYQYYHLEIRSRRNKDAGVVDKLEYNKIKDQNK YDKDDFFACTFLYEKANKVRNFIAHFNYLTMWNS PQEEEHNSNLSGAKNSSGRQNLKCSLTELINELREV MSYDRKLKNAVTKAVIDLFDKHGMVIKFRIVNNN NNDNKNKHHLELDDIVPKKIMHLRGIKLKRQDGKP IPIQTDSVDPLYCRMWKKLLDLKPTPF |
| C2-33 | | 河岸李斯特菌 (Listeria riparia) (SEQ ID No. 65) | MHDAWAENPKKPQSDAFLKEYKACCEAIDTYNW HKNKATLVYVNELHHLLIDILGRLVGYVAIADRDF QCMANQYLKSSGHTERVDSWINTIRKNRPDYIEKL DIFMNKAGLFVSEKNGRNYIAHLNYLSPKHKYSLL YLFEKLREMLKYDRKLKNAVTKSLIDLLDKHGMC VVFANLKNNKHRLVIASLKPKKIETFKWKKIK |
| C2-34 | | 陌生非适应螺菌 (Insolitispirillum peregrinum) (SEQ ID No. 66) | MRIIRPYGSSTVASPSPQDAQPLRSLQRQNGTFDVA EFSRRHPELVLAQWVAMLDKIIRKPAPGKNSTALP RPTAEQRRLRQQVGAALWAEMQRHTPVPPELKAV WDSKVHPYSKDNAPATAKTPSHRGRWYDRFGDP ETSAATVAEGVRRHLLDSAQPFRANGGQPKGKGV IEHRALTIQNGTLLHHHQSEKAGPLPEDWSTYRAD ELVSTIGKDARWIKVAASLYQHYGRIFGPTTPISEA QTRPEFVLHTAVKAYYRRLFKERKLPAERLERLLP RTGEALRHAVTVQHGNRSLADAVRIGKILHYGWL QNGEPDPWPDDAALYSSRYWGSDGQTDIKHSEAV SRVWRRALTAAQRTLTSWLYPAGTDAGDILLIGQK PDSIDRNRLPLLYGDSTRHWTRSPGDVWLFLKQTL ENLRNSSFHFKTLSAFTSHLDGTCESEPAEQQAAQ ALWQDDRQQDHQQVFLSLRALDATTYLPTGPLHR IVNAVQSTDATLPLPRFRRVVTRAANTRLKGFPVE PVNRRTMEDDPLLRCRYGVLKLLYERGFRAWLET RPSIASCLDQSLKRSTKAAQTINGKNSPQGVEILSR ATKLLQAEGGGGHGIHDLFDRLYAATAREMRVQV GYHHDAEAARQQAEFIEDLKCEVVARAFCAYLKT LGIQGDTFRRQPEPLPTWPDLPDLPSSTIGTAQAAL YSVLHLMPVEDVGSLLHQLRRWLVALQARGGED GTAITATIPLLELYLNRHDAKFSGGGAGTGLRWDD WQVFFDCQATFDRVFPPGPALDSHRLPLRGLREVL RFGRVNDLAALIGQDKITAAEVDRWHTAEQTIAAQ QQRREALHEQLSRKKGTDAEVDEYRALVTAIADH RHLTAHVTLSNVVRLHRLMTTVLGRLVDYGGLWE RDLTFVTLYEAHRLGGLRNLLSESRVNKFLDGQTP AALSKKNNAEENGMISKVLGDKARRQIRNDFAHF NMLQQGKKTINLTDEINNARKLMAHDRKLKNAIT RSVTTLLQQDGLDIVWTMDASHRLTDAKIDSRNAI |

[0489]

[0490]

| | | | |
|---|---|---|---|
| | | | HLHKTHNRANIREPLHGKSYCRWVAALFGATSTPS ATKKSDKIR |

[0491] 在某些示例实施方案中,CRISPR效应蛋白是选自表3的Cas13b蛋白。

[0492] 表3

[0493]

| | | |
|---|---|---|
| 动物溃疡伯杰菌 (SEQ ID No. 67) | 1 | MENKTSLGNNIYYNPFKPQDKSYFAGYFNAA<br>MENTDSVFRELGKRLKGKEYTSENFFDAIFKE<br>NISLVEYERYVKLLSDYFPMARLLDKKEVPIKE<br>RKENFKKNFKGIIKAVRDLRNFYTHKEHGEVEI<br>TDEIFGVLDEMLKSTVLTVKKKKVKTDKTKEI<br>LKKSIEKQLDILCQKKLEYLRDTARKIEEKRRN<br>QRERGEKELVAPFKYSDKRDDLIAAIYNDAFD<br>VYIDKKKDSLKESSKAKYNTKSDPQQEEGDLK<br>IPISKNGVVFLLSLFLTKQEIHAFKSKIAGFKAT<br>VIDEATVSEATVSHGKNSICFMATHEIFSHLAY<br>KKLKRKVRTAEINYGEAENAEQLSVYAKETL<br>MMQMLDELSKVPDVVYQNLSEDVQKTFIEDW<br>NEYLKENNGDVGTMEEEQVIHPVIRKRYEDKF<br>NYFAIRFLDEFAQFPTLRFQVHLGNYLHDSRPK<br>ENLISDRRIKEKITVFGRLSELEHKKALFIKNTE<br>TNEDREHYWEIFPNPNYDFPKENISVNDKDFPI<br>AGSILDREKQPVAGKIGIKVKLLNQQYVSEVD<br>KAVKAHQLKQRKASKPSIQNIIEEIVPINESNPK<br>EAIVFGGQPTAYLSMNDIHSILYEFFDKWEKKK<br>EKLEKKGEKELRKEIGKELEKKIVGKIQAQIQQI<br>IDKDTNAKILKPYQDGNSTAIDKEKLIKDLKQE<br>QNILQKLKDEQTVREKEYNDFIAYQDKNREIN<br>KVRDRNHKQYLKDNLKRKYPEAPARKEVLYY<br>REKGKVAVWLANDIKRFMPTDFKNEWKGEQH<br>SLLQKSLAYYEQCKEELKNLLPEKVFQHLPFKL<br>GGYFQQKYLYQFYTCYLDKRLEYISGLVQQAE<br>NFKSENKVFKKVENECFKFLKKQNYTHKELDA<br>RVQSILGYPIFLERGFMDEKPTIIKGKTFKGNEA<br>LFADWFRYYKEYQNFQTFYDTENYPLVELEKK<br>QADRKRKTKIYQQKKNDVFTLLMAKHIFKSVF<br>KQDSIDQFSLEDLYQSREERLGNQERARQTGE<br>RNTNYIWNKTVDLKLCDGKITVENVKLKNVG<br>DFIKYEYDQRVQAFLKYEENIEWQAFLIKESKE<br>EENYPYVVEREIEQYEKVRREELLKEVHLIEEY<br>ILEKVKDKEILKKGDNQNFKYYILNGLLKQLK<br>NEDVESYKVFNLNTEPEDVNINQLKQEATDLE<br>QKAFVLTYIRNKFAHNQLPKKEFWDYCQEKY |

[0494]

| | | |
|---|---|---|
| | | GKIEKEKTYAEYFAEVFKKEKEALIK |
| 中间普雷沃菌(SEQ ID No. 68) | 2 | MEDDKKTTDSIRYELKDKHFWAAFLNLARHN<br>VYITVNHINKILEEGEINRDGYETTLKNTWNEI<br>KDINKKDRLSKLIIKHFPFLEAATYRLNPTDTTK<br>QKEEKQAEAQSLESLRKSFFVFIYKLRDLRNHY<br>SHYKHSKSLERPKFEEGLLEKMYNIFNASIRLV<br>KEDYQYNKDINPDEDFKHLDRTEEEFNYYFTK<br>DNEGNITESGLLFFVSLFLEKKDAIWMQQKLR<br>GFKDNRENKKKMTNEVFCRSRMLLPKLRLQST<br>QTQDWILLDMLNELIRCPKSLYERLREEDREKF<br>RVPIEIADEDYDAEQEPFKNTLVRHQDRFPYFA<br>LRYFDYNEIFTNLRFQIDLGTYHFSIYKKQIGDY<br>KESHHLTHKLYGFERIQEFTKQNRPDEWRKFV<br>KTFNSFETSKEPYIPETTPHYHLENQKIGIRFRN<br>DNDKIWPSLKTNSEKNEKSKYKLDKSFQAEAF<br>LSVHELLPMMFYYLLLKTENTDNDNEIETKKK<br>ENKNDKQEKHKIEEIIENKITEIYALYDTFANGE<br>IKSIDELEEYCKGKDIEIGHLPKQMIAILKDEHK<br>VMATEAERKQEEMLVDVQKSLESLDNQINEEI<br>ENVERKNSSLKSGKIASWLVNDMMRFQPVQK<br>DNEGKPLNNSKANSTEYQLLQRTLAFFGSEHE<br>RLAPYFKQTKLIESSNPHPFLKDTEWEKCNNIL<br>SFYRSYLEAKKNFLESLKPEDWEKNQYFLKLK<br>EPKTKPKTLVQGWKNGFNLPRGIFTEPIRKWF<br>MKHRENITVAELKRVGLVAKVIPLFFSEEYKDS<br>VQPFYNYHFNVGNINKPDEKNFLNCEERRELL<br>RKKKDEFKKMTDKEKEENPSYLEFKSWNKFE<br>RELRLVRNQDIVTWLLCMELFNKKKIKELNVE<br>KIYLKNINTNTTKKEKNTEEKNGEEKNIKEKNN<br>ILNRIMPMRLPIKVYGRENFSKNKKKKIRRNTF<br>FTVYIEEKGTKLLKQGNFKALERDRRLGGLFSF<br>VKTPSKAESKSNTISKLRVEYELGEYQKARIEII<br>KDMLALEKTLIDKYNSLDTDNFNKMLTDWLE<br>LKGEPDKASFQNDVDLLIAVRNAFSHNQYPMR<br>NRIAFANINPFSLSSANTSEEKGLGIANQLKDKT<br>HKTIEKIIEIEKPIETKE |
| 颊普雷沃菌(Prevotella buccae)(SEQ | 3 | MQKQDKLFVDRKKNAIFAFPKYITIMENKEKP<br>EPIYYELTDKHFWAAFLNLARHNVYTTINHINR<br>RLEIAELKDDGYMMGIKGSWNEQAKKLDKKV<br>RLRDLIMKHFPFLEAAAYEMTNSKSPNNKEQR<br>EKEQSEALSLNNLKNVLFIFLEKLQVLRNYYSH<br>YKYSEESPKPIFETSLLKNMYKVFDANVRLVK<br>RDYMHHENIDMQRDFTHLNRKKQVGRTKNIID |

[0495]

| | | |
|---|---|---|
| ID No. 69) | | SPNFHYHFADKEGNMTIAGLLFFVSLFLDKKD AIWMQKKLKGFKDGRNLREQMTNEVFCRSRIS LPKLKLENVQTKDWMQLDMLNELVRCPKSLY ERLREKDRESFKVPFDIFSDDYNAEEEPFKNTL VRHQDRFPYFVLRYFDLNEIFEQLRFQIDLGTY HFSIYNKRIGDEDEVRHLTHHLYGFARIQDFAP QNQPEEWRKLVKDLDHFETSQEPYISKTAPHY HLENEKIGIKFCSAHNNLFPSLQTDKTCNGRSK FNLGTQFTAEAFLSVHELLPMMFYYLLLTKDY SRKESADKVEGIIRKEISNIYAIYDAFANNEINSI ADLTRRLQNTNILQGHLPKQMISILKGRQKDM GKEAERKIGEMIDDTQRRLDLLCKQTNQKIRIG KRNAGLLKSGKIADWLVNDMMRFQPVQKDQ NNIPINNSKANSTEYRMLQRALALFGSENFRLK AYFNQMNLVGNDNPHPFLAETQWEHQTNILSF YRNYLEARKKYLKGLKPQNWKQYQHFLILKV QKTNRNTLVTGWKNSFNLPRGIFTQPIREWFEK HNNSKRIYDQILSFDRVGFVAKAIPLYFAEEYK DNVQPFYDYPFNIGNRLKPKKRQFLDKKERVE LWQKNKELFKNYPSEKKKTDLAYLDFLSWKK FERELRLIKNQDIVTWLMFKELFNMATVEGLKI GEIHLRDIDTNTANEESNNILNRIMPMKLPVKT YETDNKGNILKERPLATFYIEETETKVLKQGNF KALVKDRRLNGLFSFAETTDLNLEEHPISKLSV DLELIKYQTTRISIFEMTLGLEKKLIDKYSTLPT DSFRNMLERWLQCKANRPELKNYVNSLIAVR NAFSHNQYPMYDATLFAEVKKFTLFPSVDTKK IELNIAPQLLEIVGKAIKEIEKSENKN |
| 牙龈卟啉单胞菌 (Porphyromonas gingivalis) (SEQ ID No. 70) | 4 | MNTVPASENKGQSRTVEDDPQYFGLYLNLARE NLIEVESHVRIKFGKKKLNEESLKQSLLCDHLL SVDRWTKVYGHSRRYLPFLHYFDPDSQIEKDH DSKTGVDPDSAQRLIRELYSLLDFLRNDFSHNR LDGTTFEHLEVSPDISSFITGTYSLACGRAQSRF AVFFKPDDFVLAKNRKEQLISVADGKECLTVS GFAFFICLFLDREQASGMLSRIRGFKRTDENWA RAVHETFCDLCIRHPHDRLESSNTKEALLLDML NELNRCPRILYDMLPEEERAQFLPALDENSMN NLSENSLDEESRLLWDGSSDWAEALTKRIRHQ DRFPYLMLRFIEEMDLLKGIRFRVDLGEIELDS YSKKVGRNGEYDRTITDHALAFGKLSDFQNEE EVSRMISGEASYPVRFSLFAPRYAIYDNKIGYC HTSDPVYPKSKTGEKRALSNPQSMGFISVHDLR KLLLMELLCEGSFSRMQSDFLRKANRILDETAE |

| | | | |
|---|---|---|---|
| | | | GKLQFSALFPEMRHRFIPPQNPKSKDRREKAET<br>TLEKYKQEIKGRKDKLNSQLLSAFDMDQRQLP<br>SRLLDEWMNIRPASHSVKLRTYVKQLNEDCRL<br>RLRKFRKDGDGKARAIPLVGEMATFLSQDIVR<br>MIISEETKKLITSAYYNEMQRSLAQYAGEENRR<br>QFRAIVAELRLLDPSSGHPFLSATMETAHRYTE<br>GFYKCYLEKKREWLAKIFYRPEQDENTKRRIS<br>VFFVPDGEARKLLPLLIRRRMKEQNDLQDWIR<br>NKQAHPIDLPSHLFDSKVMELLKVKDGKKKW<br>NEAFKDWWSTKYPDGMQPFYGLRRELNIHGK<br>SVSYIPSDGKKFADCYTHLMEKTVRDKKRELR<br>TAGKPVPPDLAADIKRSFHRAVNEREFMLRLV<br>QEDDRLMLMAINKMMTDREEDILPGLKNIDSI<br>LDEENQFSLAVHAKVLEKEGEGGDNSLSLVPA<br>TIEIKSKRKDWSKYIRYRYDRRVPGLMSHFPEH<br>KATLDEVKTLLGEYDRCRIKIFDWAFALEGAI<br>MSDRDLKPYLHESSSREGKSGEHSTLVKMLVE<br>KKGCLTPDESQYLILIRNKAAHNQFPCAAEMPL<br>IYRDVSAKVGSIEGSSAKDLPEGSSLVDSLWKK<br>YEMIIRKILPILDPENRFFGKLLNNMSQPINDL |
| [0496] | 化脓拟杆菌(Bacteroides pyogenes)(SEQ ID No. 71) | 5 | MESIKNSQKSTGKTLQKDPPYFGLYLNMALLN<br>VRKVENHIRKWLGDVALLPEKSGFHSLLTTDN<br>LSSAKWTRFYYKSRKFLPFLEMFDSDKKSYEN<br>RRETAECLDTIDRQKISSLLKEVYGKLQDIRNA<br>FSHYHIDDQSVKHTALIISSEMHRFIENAYSFAL<br>QKTRARFTGVFVETDFLQAEEKGDNKKFFAIG<br>GNEGIKLKDNALIFLICLFLDREEAFKFLSRATG<br>FKSTKEKGFLAVRETFCALCCRQPHERLLSVNP<br>REALLMDMLNELNRCPDILFEMLDEKDQKSFL<br>PLLGEEEQAHILENSLNDELCEAIDDPFEMIASL<br>SKRVRYKNRFPYLMLRYIEEKNLLPFIRFRIDL<br>GCLELASYPKKMGEENNYERSVTDHAMAFGR<br>LTDFHNEDAVLQQITKGITDEVRFSLYAPRYAI<br>YNNKIGFVRTSGSDKISFPTLKKKGGEGHCVA<br>YTLQNTKSFGFISIYDLRKILLLSFLDKDKAKNI<br>VSGLLEQCEKHWKDLSENLFDAIRTELQKEFP<br>VPLIRYTLPRSKGGKLVSSKLADKQEKYESEFE<br>RRKEKLTEILSEKDFDLSQIPRRMIDEWLNVLP<br>TSREKKLKGYVETLKLDCRERLRVFEKREKGE<br>HPLPPRIGEMATDLAKDIIRMVIDQGVKQRITS<br>AYYSEIQRCLAQYAGDDNRRHLDSIIRELRLKD<br>TKNGHPFLGKVLRPGLGHTEKLYQRYFEEKKE<br>WLEATFYPAASPKRVPRFVNPPTGKQKELPLII |

[0497]

| | | |
|---|---|---|
| | | RNLMKERPEWRDWKQRKNSHPIDLPSQLFENE ICRLLKDKIGKEPSGKLKWNEMFKLYWDKEFP NGMQRFYRCKRRVEVFDKVVEYEYSEEGGNY KKYYEALIDEVVRQKISSSKEKSKLQVEDLTLS VRRVFKRAINEKEYQLRLLCEDDRLLFMAVRD LYDWKEAQLDLDKIDNMLGEPVSVSQVIQLEG GQPDAVIKAECKLKDVSKLMRYCYDGRVKGL MPYFANHEATQEQVEMELRHYEDHRRRVFNW VFALEKSVLKNEKLRRFYEESQGGCEHRRCID ALRKASLVSEEEYEFLVHIRNKSAHNQFPDLEI GKLPPNVTSGFCECIWSKYKAIICRIIPFIDPERR FFGKLLEQK |
| 另枝菌属某种(Alistipes sp.)ZOR0009(SEQ ID No. 72) | 6 | MSNEIGAFREHQFAYAPGNEKQEEATFATYFN LALSNVEGMMFGEVESNPDKIEKSLDTLPPAIL RQIASFIWLSKEDHPDKAYSTEEVKVIVTDLVR RLCFYRNYFSHCFYLDTQYFYSDELVDTTAIGE KLPYNFHHFITNRLFRYSLPEITLFRWNEGERK YEILRDGLIFFCCLFLKRGQAERFLNELRFFKRT DEEGRIKRTIFTKYCTRESHKHIGIEEQDFLIFQD IIGDLNRVPKVCDGVVDLSKENERYIKNRETSN ESDENKARYRLLIREKDKFPYYLMRYIVDFGV LPCITFKQNDYSTKEGRGQFHYQDAAVAQEER CYNFVVRNGNVYYSYMPQAQNVVRISELQGTI SVEELRNMVYASINGKDVNKSVEQYLYHLHLL YEKILTISGQTIKEGRVDVEDYRPLLDKLLLRP ASNGEELRRELRKLLPKRVCDLLSNRFDCSEG VSAVEKRLKAILLRHEQLLLSQNPALHIDKIKS VIDYLYLFFSDDEKFRQQPTEKAHRGLKDEEFQ MYHYLVGDYDSHPLALWKELEASGRLKPEMR KLTSATSLHGLYMLCLKGTVEWCRKQLMSIG KGTAKVEAIADRVGLKLYDKLKEYTPEQLERE VKLVVMHGYAAAATPKPKAQAAIPSKLTELRF YSFLGKREMSFAAFIRQDKKAQKLWLRNFYTV ENIKTLQKRQAAADAACKKLYNLVGEVERVH TNDKVLVLVAQRYRERLLNVGSKCAVTLDNP ERQQKLADVYEVQNAWLSIRFDDLDFTLTHVN LSNLRKAYNLIPRKHILAFKEYLDNRVKQKLCE ECRNVRRKEDLCTCCSPRYSNLTSWLKENHSE SSIEREAATMMLLDVERKLLSFLLDERRKAIIE YGKFIPFSALVKECRLADAGLCGIRNDVLHDN VISYADAIGKLSAYFPKEASEAVEYIRRTKEVR EQRREELMANSSQ |
| 普雷沃 | 7a | MSKECKKQRQEKKRRLQKANFSISLTGKHVFG |

[0498]

| | | |
|---|---|---|
| 菌属某种 (Prevotella sp.) MA2016 (SEQ ID No. 73) | | AYFNMARTNFVKTINYILPIAGVRGNYSENQIN<br>KMLHALFLIQAGRNEELTTEQKQWEKKLRLNP<br>EQQTKFQKLLFKHFPVLGPMMADVADHKAYL<br>NKKKSTVQTEDETFAMLKGVSLADCLDIICLM<br>ADTLTECRNFYTHKDPYNKPSQLADQYLHQE<br>MIAKKLDKVVVASRRILKDREGLSVNEVEFLT<br>GIDHLHQEVLKDEFGNAKVKDGKVMKTFVEY<br>DDFYFKISGKRLVNGYTVTTKDDKPVNVNTM<br>LPALSDFGLLYFCVLFLSKPYAKLFIDEVRLFEY<br>SPFDDKENMIMSEMLSIYRIRTPRLHKIDSHDS<br>KATLAMDIFGELRRCPMELYNLLDKNAGQPFF<br>HDEVKHPNSHTPDVSKRLRYDDRFPTLALRYI<br>DETELFKRIRFQLQLGSFRYKFYDKENCIDGRV<br>RVRRIQKEINGYGRMQEVADKRMDKWGDLIQ<br>KREERSVKLEHEELYINLDQFLEDTADSTPYVT<br>DRRPAYNIHANRIGLYWEDSQNPKQYKVFDEN<br>GMYIPELVVTEDKKAPIKMPAPRCALSVYDLP<br>AMLFYEYLREQQDNEFPSAEQVIIEYEDDYRKF<br>FKAVAEGKLKPFKRPKEFRDFLKKEYPKLRMA<br>DIPKKLQLFLCSHGLCYNNKPETVYERLDRLTL<br>QHLEERELHIQNRLEHYQKDRDMIGNKDNQY<br>GKKSFSDVRHGALARYLAQSMMEWQPTKLKD<br>KEKGHDKLTGLNYNVLTAYLATYGHPQVPEE<br>GFTPRTLEQVLINAHLIGGSNPHPFINKVLALGN<br>RNIEELYLHYLEEELKHIRSRIQSLSSNPSDKAL<br>SALPFIHHDRMRYHERTSEEMMALAARYTTIQ<br>LPDGLFTPYILEILQKHYTENSDLQNALSQDVP<br>VKLNPTCNAAYLITLFYQTVLKDNAQPFYLSD<br>KTYTRNKDGEKAESFSFKRAYELFSVLNNNKK<br>DTFPFEMIPLFLTSDEIQERLSAKLLDGDGNPVP<br>EVGEKGKPATDSQGNTIWKRRIYSEVDDYAEK<br>LTDRDMKISFKGEWEKLPRWKQDKIIKRRDET<br>RRQMRDELLQRMPRYIRDIKDNERTLRRYKTQ<br>DMVLFLLAEKMFTNIISEQSSEFNWKQMRLSK<br>VCNEAFLRQTLTFRVPVTVGETTIYVEQENMSL<br>KNYGEFYRFLTDDRLMSLLNNIVETLKPNENG<br>DLVIRHTDLMSELAAYDQYRSTIFMLIQSIENLI<br>ITNNAVLDDPDADGFWVREDLPKRNNFASLLE<br>LINQLNNVELTDDERKLLVAIRNAFSHNSYNID<br>FSLIKDVKHLPEVAKGILQHLQSMLGVEITK |
| 普雷沃菌属某 | 7b | MSKECKKQRQEKKRRLQKANFSISLTGKHVFG<br>AYFNMARTNFVKTINYILPIAGVRGNYSENQIN<br>KMLHALFLIQAGRNEELTTEQKQWEKKLRLNP |

[0499]

| | | |
|---|---|---|
| 种 MA2016 (SEQ ID No. 74) | | EQQTKFQKLLFKHFPVLGPMMADVADHKAYL<br>NKKKSTVQTEDETFAMLKGVSLADCLDIICLM<br>ADTLTECRNFYTHKDPYNKPSQLADQYLHQE<br>MIAKKLDKVVVASRRILKDREGLSVNEVEFLT<br>GIDHLHQEVLKDEFGNAKVKDGKVMKTFVEY<br>DDFYFKISGKRLVNGYTVTTKDDKPVNVNTM<br>LPALSDFGLLYFCVLFLSKPYAKLFIDEVRLFEY<br>SPFDDKENMIMSEMLSIYRIRTPRLHKIDSHDS<br>KATLAMDIFGELRRCPMELYNLLDKNAGQPFF<br>HDEVKHPNSHTPDVSKRLRYDDRFPTLALRYI<br>DETELFKRIRFQLQLGSFRYKFYDKENCIDGRV<br>RVRRIQKEINGYGRMQEVADKRMDKWGDLIQ<br>KREERSVKLEHEELYINLDQFLEDTADSTPYVT<br>DRRPAYNIHANRIGLYWEDSQNPKQYKVFDEN<br>GMYIPELVVTEDKKAPIKMPAPRCALSVYDLP<br>AMLFYEYLREQQDNEFPSAEQVIIEYEDDYRKF<br>FKAVAEGKLKPFKRPKEFRDFLKKEYPKLRMA<br>DIPKKLQLFLCSHGLCYNNKPETVYERLDRLTL<br>QHLEERELHIQNRLEHYQKDRDMIGNKDNQY<br>GKKSFSDVRHGALARYLAQSMMEWQPTKLKD<br>KEKGHDKLTGLNYNVLTAYLATYGHPQVPEE<br>GFTPRTLEQVLINAHLIGGSNPHPFINKVLALGN<br>RNIEELYLHYLEEELKHIRSRIQSLSSNPSDKAL<br>SALPFIHHDRMRYHERTSEEMMALAARYTTIQ<br>LPDGLFTPYILEILQKHYTENSDLQNALSQDVP<br>VKLNPTCNAAYLITLFYQTVLKDNAQPFYLSD<br>KTYTRNKDGEKAESFSFKRAYELFSVLNNNKK<br>DTFPFEMIPLFLTSDEIQERLSAKLLDGDGNPVP<br>EVGEKGKPATDSQGNTIWKRRIYSEVDDYAEK<br>LTDRDMKISFKGEWEKLPRWKQDKIIKRRDET<br>RRQMRDELLQRMPRYIRDIKDNERTLRRYKTQ<br>DMVLFLLAEKMFTNIISEQSSEFNWKQMRLSK<br>VCNEAFLRQTLTFRVPVTVGETTIYVEQENMSL<br>KNYGEFYRFLTDDRLMSLLNNIVETLKPNENG<br>DLVIRHTDLMSELAAYDQYRSTIFMLIQSIENLI<br>ITNNAVLDDPDADGFWVREDLPKRNNFASLLE<br>LINQLNNVELTDDERKLLVAIRNAFSHNSYNID<br>FSLIKDVKHLPEVAKGILQHLQSMLGVEITK |
| 鸭疫里默氏杆菌 (Riemer | 8 | MEKPLLPNVYTLKHKFFWGAFLNIARHNAFITI<br>CHINEQLGLKTPSNDDKIVDVVCETWNNILNN<br>DHDLLKKSQLTELILKHFPFLTAMCYHPPKKEG<br>KKKGHQKEQQKEKESEAQSQAEALNPSKLIEA<br>LEILVNQLHSLRNYYSHYKHKKPDAEKDIFKH |

[0500]

| | | |
|---|---|---|
| ella anatipestifer) (SEQ ID No. 75) | | LYKAFDASLRMVKEDYKAHFTVNLTRDFAHL<br>NRKGKNKQDNPDFNRYRFEKDGFFTESGLLFF<br>TNLFLDKRDAYWMLKKVSGFKASHKQREKM<br>TTEVFCRSRILLPKLRLESRYDHNQMLLDMLSE<br>LSRCPKLLYEKLSEENKKHFQVEADGFLDEIEE<br>EQNPFKDTLIRHQDRFPYFALRYLDLNESFKSIR<br>FQVDLGTYHYCIYDKKIGDEQEKRHLTRTLLSF<br>GRLQDFTEINRPQEWKALTKDLDYKETSNQPFI<br>SKTTPHYHITDNKIGFRLGTSKELYPSLEIKDGA<br>NRIAKYPYNSGFVAHAFISVHELLPLMFYQHLT<br>GKSEDLLKETVRHIQRIYKDFEEERINTIEDLEK<br>ANQGRLPLGAFPKQMLGLLQNKQPDLSEKAKI<br>KIEKLIAETKLLSHRLNTKLKSSPKLGKRREKLI<br>KTGVLADWLVKDFMRFQPVAYDAQNQPIKSS<br>KANSTEFWFIRRALALYGGEKNRLEGYFKQTN<br>LIGNTNPHPFLNKFNWKACRNLVDFYQQYLEQ<br>REKFLEAIKNQPWEPYQYCLLLKIPKENRKNLV<br>KGWEQGGISLPRGLFTEAIRETLSEDLMLSKPIR<br>KEIKKHGRVGFISRAITLYFKEKYQDKHQSFYN<br>LSYKLEAKAPLLKREEHYEYWQQNKPQSPTES<br>QRLELHTSDRWKDYLLYKRWQHLEKKLRLYR<br>NQDVMLWLMTLELTKNHFKELNLNYHQLKLE<br>NLAVNVQEADAKLNPLNQTLPMVLPVKVYPA<br>TAFGEVQYHKTPIRTVYIREEHTKALKMGNFK<br>ALVKDRRLNGLFSFIKEENDTQKHPISQLRLRR<br>ELEIYQSLRVDAFKETLSLEEKLLNKHTSLSSLE<br>NEFRALLEEWKKEYAASSMVTDEHIAFIASVR<br>NAFCHNQYPFYKEALHAPIPLFTVAQPTTEEKD<br>GLGIAEALLKVLREYCEIVKSQI |
| 桔红色普雷沃菌 (Prevotella aurantiaca) (SEQ ID No. 76) | 9 | MEDDKKTTGSISYELKDKHFWAAFLNLARHN<br>VYITINHINKLLEIREIDNDEKVLDIKTLWQKGN<br>KDLNQKARLRELMTKHFPFLETAIYTKNKEDK<br>KEVKQEKQAEAQSLESLKDCLFLFLDKLQEAR<br>NYYSHYKYSEFSKEPEFEEGLLEKMYNIFGNNI<br>QLVINDYQHNKDINPDEDFKHLDRKGQFKYSF<br>ADNEGNITESGLLFFVSLFLEKKDAIWMQQKL<br>NGFKDNLENKKKMTHEVFCRSRILMPKLRLES<br>TQTQDWILLDMLNELIRCPKSLYERLQGDDRE<br>KFKVPFDPADEDYNAEQEPFKNTLIRHQDRFPY<br>FVLRYFDYNEIFKNLRFQIDLGTYHFSIYKKLIG<br>GQKEDRHLTHKLYGFERIQEFAKQNRPDEWK<br>AIVKDLDTYETSNKRYISETTPHYHLENQKIGIR<br>FRNGNKEIWPSLKTNDENNEKSKYKLDKQYQ |

| | | | |
|---|---|---|---|
| | | | AEAFLSVHELLPMMFYYLLLKKEKPNNDEINA<br>SIVEGFIKREIRNIFKLYDAFANGEINNIDDLEK<br>YCADKGIPKRHLPKQMVAILYDEHKDMVKEA<br>KRKQKEMVKDTKKLLATLEKQTQKEKEDDGR<br>NVKLLKSGEIARWLVNDMMRFQPVQKDNEGK<br>PLNNSKANSTEYQMLQRSLALYNNEEKPTRYF<br>RQVNLIESNNPHPFLKWTKWEECNNILTFYYS<br>YLTKKIEFLNKLKPEDWKKNQYFLKLKEPKTN<br>RETLVQGWKNGFNLPRGIFTEPIREWFKRHQN<br>NSKEYEKVEALDRVGLVTKVIPLFFKEEYFKD<br>KEENFKEDTQKEINDCVQPFYNFPYNVGNIHKP<br>KEKDFLHREERIELWDKKKDKFKGYKEKIKSK<br>KLTEKDKEEFRSYLEFQSWNKFERELRLVRNQ<br>DIVTWLLCKELIDKLKIDELNIEELKKLRLNNID<br>TDTAKKEKNNILNRVMPMELPVTVYEIDDSHK<br>IVKDKPLHTIYIKEAETKLLKQGNFKALVKDRR<br>LNGLFSFVKTNSEAESKRNPISKLRVEYELGEY<br>QEARIEIIQDMLALEEKLINKYKDLPTNKFSEM<br>LNSWLEGKDEADKARFQNDVDFLIAVRNAFSH<br>NQYPMHNKIEFANIKPFSLYTANNSEEKGLGIA<br>NQLKDKTKETTDKIKKIEKPIETKE |
| [0501] | 解糖普雷沃菌(Prevotella saccharolytica)(SEQ ID No. 77) | 10 | MEDKPFWAAFFNLARHNVYLTVNHINKLLDL<br>EKLYDEGKHKEIFEREDIFNISDDVMNDANSNG<br>KKRKLDIKKIWDDLDTDLTRKYQLRELILKHFP<br>FIQPAIIGAQTKERTTIDKDKRSTSTSNDSLKQT<br>GEGDINDLLSLSNVKSMFFRLLQILEQLRNYYS<br>HVKHSKSATMPNFDEDLLNWMRYIFIDSVNKV<br>KEDYSSNSVIDPNTSFSHLIYKDEQGKIKPCRYP<br>FTSKDGSINAFGLLFFVSLFLEKQDSIWMQKKIP<br>GFKKASENYMKMTNEVFCRNHILLPKIRLETV<br>YDKDWMLLDMLNEVVRCPLSLYKRLTPAAQN<br>KFKVPEKSSDNANRQEDDNPFSRILVRHQNRFP<br>YFVLRFFDLNEVFTTLRFQINLGCYHFAICKKQI<br>GDKKEVHHLIRTLYGFSRLQNFTQNTRPEEWN<br>TLVKTTEPSSGNDGKTVQGVPLPYISYTIPHYQI<br>ENEKIGIKIFDGDTAVDTDIWPSVSTEKQLNKP<br>DKYTLTPGFKADVFLSVHELLPMMFYYQLLLC<br>EGMLKTDAGNAVEKVLIDTRNAIFNLYDAFVQ<br>EKINTITDLENYLQDKPILIGHLPKQMIDLLKGH<br>QRDMLKAVEQKKAMLIKDTERRLKLLDKQLK<br>QETDVAAKNTGTLLKNGQIADWLVNDMMRF<br>QPVKRDKEGNPINCSKANSTEYQMLQRAFAFY<br>ATDSCRLSRYFTQLHLIHSDNSHLFLSRFEYDK |

[0502]

| | | |
|---|---|---|
| | | QPNLIAFYAAYLKAKLEFLNELQPQNWASDNY<br>FLLLRAPKNDRQKLAEGWKNGFNLPRGLFTEK<br>IKTWFNEHKTIVDISDCDIFKNRVGQVARLIPVF<br>FDKKFKDHSQPFYRYDFNVGNVSKPTEANYLS<br>KGKREELFKSYQNKFKNNIPAEKTKEYREYKN<br>FSLWKKFERELRLIKNQDILIWLMCKNLFDEKI<br>KPKKDILEPRIAVSYIKLDSLQTNTSTAGSLNAL<br>AKVVPMTLAIHIDSPKPKGKAGNNEKENKEFT<br>VYIKEEGTKLLKWGNFKTLLADRRIKGLFSYIE<br>HDDIDLKQHPLTKRRVDLELDLYQTCRIDIFQQ<br>TLGLEAQLLDKYSDLNTDNFYQMLIGWRKKE<br>GIPRNIKEDTDFLKDVRNAFSHNQYPDSKKIAF<br>RRIRKFNPKELILEEEEGLGIATQMYKEVEKVV<br>NRIKRIELFD |
| HMPREF9712_03108 [类香味菌(Myroides odoratimimus) CCUG 10230] (SEQ ID No. 78) | 11 | MKDILTTDTTEKQNRFYSHKIADKYFFGGYFN<br>LASNNIYEVFEEVNKRNTFGKLAKRDNGNLKN<br>YIIHVFKDELSISDFEKRVAIFASYFPILETVDKK<br>SIKERNRTIDLTLSQRIRQFREMLISLVTAVDQL<br>RNFYTHYHHSDIVIENKVLDFLNSSFVSTALHV<br>KDKYLKTDKTKEFLKETIAAELDILIEAYKKKQ<br>IEKKNTRFKANKREDILNAIYNEAFWSFINDKD<br>KDKDKETVVAKGADAYFEKNHHKSNDPDFAL<br>NISEKGIVYLLSFFLTNKEMDSLKANLTGFKGK<br>VDRESGNSIKYMATQRIYSFHTYRGLKQKIRTS<br>EEGVKETLLMQMIDELSKVPNVVYQHLSTTQQ<br>NSFIEDWNEYYKDYEDDVETDDLSRVIHPVIRK<br>RYEDRFNYFAIRFLDEFFDFPTLRFQVHLGDYV<br>HDRRTKQLGKVESDRIIKEKVTVFARLKDINSA<br>KASYFHSLEEQDKEELDNKWTLFPNPSYDFPK<br>EHTLQHQGEQKNAGKIGIYVKLRDTQYKEKA<br>ALEEARKSLNPKERSATKASKYDIITQIIEANDN<br>VKSEKPLVFTGQPIAYLSMNDIHSMLFSLLTDN<br>AELKKTPEEVEAKLIDQIGKQINEILSKDTDTKI<br>LKKYKDNDLKETDTDKITRDLARDKEEIEKLIL<br>EQKQRADDYNYTSSTKFNIDKSRKRKHLLFNA<br>EKGKIGVWLANDIKRFMFKESKSKWKGYQHT<br>ELQKLFAYFDTSKSDLELILSNMVMVKDYPIEL<br>IDLVKKSRTLVDFLNKYLEARLEYIENVITRVK<br>NSIGTPQFKTVRKECFTFLKKSNYTVVSLDKQV<br>ERILSMPLFIERGFMDDKPTMLEGKSYKQHKE<br>KFADWFVHYKENSNYQNFYDTEVYEITTEDKR<br>EKAKVTKKIKQQQKNDVFTLMMVNYMLEEVL<br>KLSSNDRLSLNELYQTKEERIVNKQVAKDTQE |

[0503]

| | | |
|---|---|---|
| | | RNKNYIWNKVVDLQLCDGLVHIDNVKLKDIG<br>NFRKYENDSRVKEFLTYQSDIVWSAYLSNEVD<br>SNKLYVIERQLDNYESIRSKELLKEVQEIECSVY<br>NQVANKESLKQSGNENFKQYVLQGLLPIGMD<br>VREMLILSTDVKFKKEEIIQLGQAGEVEQDLYS<br>LIYIRNKFAHNQLPIKEFFDFCENNYRSISDNEY<br>YAEYYMEIFRSIKEKYAN |
| 中间普雷沃菌(SEQ ID No. 79) | 12 | MEDDKKTTDSIRYELKDKHFWAAFLNLARHN<br>VYITVNHINKILEEDEINRDGYENTLENSWNEI<br>KDINKKDRLSKLIIKHFPFLEATTYRQNPTDTTK<br>QKEEKQAEAQSLESLKKSFFVFIYKLRDLRNHY<br>SHYKHSKSLERPKFEEDLQNKMYNIFDVSIQFV<br>KEDYKHNTDINPKKDFKHLDRKRKGKFHYSFA<br>DNEGNITESGLLFFVSLFLEKKDAIWVQKKLEG<br>FKCSNKSYQKMTNEVFCRSRMLLPKLRLESTQ<br>TQDWILLDMLNELIRCPKSLYERLQGVNRKKF<br>YVSFDPADEDYDAEQEPFKNTLVRHQDRFPYF<br>ALRYFDYNEVFANLRFQIDLGTYHFSIYKKLIG<br>GQKEDRHLTHKLYGFERIQEFDKQNRPDEWK<br>AIVKDSDTFKKKEEKEEEKPYISETTPHYHLEN<br>KKIGIAFKNHNIWPSTQTELTNNKRKKYNLGTS<br>IKAEAFLSVHELLPMMFYYLLLKTENTKNDNK<br>VGGKKETKKQGKHKIEAIIESKIKDIYALYDAF<br>ANGEINSEDELKEYLKGKDIKIVHLPKQMIAIL<br>KNEHKDMAEKAEAKQEKMKLATENRLKTLD<br>KQLKGKIQNGKRYNSAPKSGEIASWLVNDMM<br>RFQPVQKDENGESLNNSKANSTEYQLLQRTLA<br>FFGSEHERLAPYFKQTKLIESSNPHPFLNDTEW<br>EKCSNILSFYRSYLKARKNFLESLKPEDWEKNQ<br>YFLMLKEPKTNRETLVQGWKNGFNLPRGFFTE<br>PIRKWFMEHWKSIKVDDLKRVGLVAKVTPLFF<br>SEKYKDSVQPFYNYPFNVGDVNKPKEEDFLHR<br>EERIELWDKKKDKFKGYKAKKKFKEMTDKEK<br>EEHRSYLEFQSWNKFERELRLVRNQDIVTWLL<br>CTELIDKLKIDELNIKELKKLRLKDINTDTAKKE<br>KNNILNRVMPMELPVTVYKVNKGGYIIKNKPL<br>HTIYIKEAETKLLKQGNFKALVKDRRLNGLFSF<br>VKTPSEAESESNPISKLRVEYELGKYQNARLDII<br>EDMLALEKKLIDKYNSLDTDNFHNMLTGWLE<br>LKGEAKKARFQNDVKLLTAVRNAFSHNQYPM<br>YDENLFGNIERFSLSSSNIIESKGLDIAAKLKEE<br>VSKAAKKIQNEEDNKKEKET |
| 犬咬二 | 13 | MKNIQRLGKGNEFSPFKKEDKFYFGGFLNLAN |

[0504]

| | | |
|---|---|---|
| 氧化碳噬纤维菌 (Capnocytophaga canimorsus) (SEQ ID No. 80) | | NNIEDFFKEIITRFGIVITDENKKPKETFGEKILN EIFKKDISIVDYEKWVNIFADYFPFTKYLSLYLE EMQFKNRVICFRDVMKELLKTVEALRNFYTHY DHEPIKIEDRVFYFLDKVLLDVSLTVKNKYLKT DKTKEFLNQHIGEELKELCKQRKDYLVGKGKR IDKESEIINGIYNNAFKDFICKREKQDDKENHNS VEKILCNKEPQNKKQKSSATVWELCSKSSSKY TEKSFPNRENDKHCLEVPISQKGIVFLLSFFLNK GEIYALTSNIKGFKAKITKEEPVTYDKNSIRYM ATHRMFSFLAYKGLKRKIRTSEINYNEDGQASS TYEKETLMLQMLDELNKVPDVVYQNLSEDVQ KTFIEDWNEYLKENNGDVGTMEEEQVIHPVIR KRYEDKFNYFAIRFLDEFAQFPTLRFQVHLGNY LCDKRTKQICDTTTEREVKKKITVFGRLSELEN KKAIFLNEREEIKGWEVFPNPSYDFPKENISVN YKDFPIVGSILDREKQPVSNKIGIRVKIADELQR EIDKAIKEKKLRNPKNRKANQDEKQKERLVNE IVSTNSNEQGEPVVFIGQPTAYLSMNDIHSVLY EFLINKISGEALETKIVEKIETQIKQIIGKDATTKI LKPYTNANSNSINREKLLRDLEQEQQILKTLLE EQQQREKDKKDKKSKRKHELYPSEKGKVAVW LANDIKRFMPKAFKEQWRGYHHSLLQKYLAY YEQSKEELKNLLPKEVFKHFPFKLKGYFQQQY LNQFYTDYLKRRLSYVNELLLNIQNFKNDKDA LKATEKECFKFFRKQNYIINPINIQIQSILVYPIFL KRGFLDEKPTMIDREKFKENKDTELADWFMH YKNYKEDNYQKFYAYPLEKVEEKEKFKRNKQ INKQKKNDVYTLMMVEYIIQKIFGDKFVEENPL VLKGIFQSKAERQQNNTHAATTQERNLNGILN QPKDIKIQGKITVKGVKLKDIGNFRKYEIDQRV NTFLDYEPRKEWMAYLPNDWKEKEKQGQLPP NNVIDRQISKYETVRSKILLKDVQELEKIISDEIK EEHRHDLKQGKYYNFKYYILNGLLRQLKNEN VENYKVFKLNTNPEKVNITQLKQEATDLEQKA FVLTYIRNKFAHNQLPKKEFWDYCQEKYGKIE KEKTYAEYFAEVFKREKEALIK |
| 动物口腔卟啉单胞菌 (Porphyromonas gulae) | 14 | MTEQSERPYNGTYYTLEDKHFWAAFLNLARH NAYITLTHIDRQLAYSKADITNDQDVLSFKAL WKNFDNDLERKSRLRSLILKHFSFLEGAAYGK KLFESKSSGNKSSKNKELTKKEKEELQANALSL DNLKSILFDFLQKLKDFRNYYSHYRHSGSSELP LFDGNMLQRLYNVFDVSVQRVKIDHEHNDEV DPHYHFNHLVRKGKKDRYGHNDNPSFKHHFV |

[0505]

| | | |
|---|---|---|
| (SEQ ID No. 81) | | DGEGMVTEAGLLFFVSLFLEKRDAIWMQKKIR<br>GFKGGTETYQQMTNEVFCRSRISLPKLKLESLR<br>MDDWMLLDMLNELVRCPKPLYDRLREDDRA<br>CFRVPVDILPDEDDTDGGGEDPFKNTLVRHQD<br>RFPYFALRYFDLKKVFTSLRFHIDLGTYHFAIY<br>KKMIGEQPEDRHLTRNLYGFGRIQDFAEEHRPE<br>EWKRLVRDLDYFETGDKPYISQTSPHYHIEKG<br>KIGLRFMPEGQHLWPSPEVGTTRTGRSKYAQD<br>KRLTAEAFLSVHELMPMMFYYFLLREKYSEEV<br>SAERVQGRIKRVIEDVYAVYDAFARDEINTRD<br>ELDACLADKGIRRGHLPRQMIAILSQEHKDME<br>EKIRKKLQEMMADTDHRLDMLDRQTDRKIRIG<br>RKNAGLPKSGVIADWLVRDMMRFQPVAKDAS<br>GKPLNNSKANSTEYRMLQRALALFGGEKERLT<br>PYFRQMNLTGGNNPHPFLHETRWESHTNILSF<br>YRSYLRARKAFLERIGRSDRVENRPFLLLKEPK<br>TDRQTLVAGWKGEFHLPRGIFTEAVRDCLIEM<br>GHDEVASYKEVGFMAKAVPLYFERACEDRVQ<br>PFYDSPFNVGNSLKPKKGRFLSKEERAEEWER<br>GKERFRDLEAWSYSAARRIEDAFAGIEYASPG<br>NKKKIEQLLRDLSLWEAFESKLKVRADRINLA<br>KLKKEILEAQEHPYHDFKSWQKFERELRLVKN<br>QDIITWMMCRDLMEENKVEGLDTGTLYLKDIR<br>PNVQEQGSLNVLNRVKPMRLPVVVYRADSRG<br>HVHKEEAPLATVYIEERDTKLLKQGNFKSFVK<br>DRRLNGLFSFVDTGGLAMEQYPISKLRVEYEL<br>AKYQTARVCVFELTLRLEESLLTRYPHLPDESF<br>REMLESWSDPLLAKWPELHGKVRLLIAVRNAF<br>SHNQYPMYDEAVFSSIRKYDPSSPDAIEERMGL<br>NIAHRLSEEVKQAKETVERIIQA |
| 普雷沃菌属某种 P5-125 (SEQ ID No. 82) | 15 | MNIPALVENQKKYFGTYSVMAMLNAQTVLDH<br>IQKVADIEGEQNENNENLWFHPVMSHLYNAK<br>NGYDKQPEKTMFIIERLQSYFPFLKIMAENQRE<br>YSNGKYKQNRVEVNSNDIFEVLKRAFGVLKM<br>YRDLTNHYKTYEEKLNDGCEFLTSTEQPLSGM<br>INNYYTVALRNMNERYGYKTEDLAFIQDKRFK<br>FVKDAYGKKKSQVNTGFFLSLQDYNGDTQKK<br>LHLSGVGIALLICLFLDKQYINIFLSRLPIFSSYN<br>AQSEERRIIIRSFGINSIKLPKDRIHSEKSNKSVA<br>MDMLNEVKRCPDELFTTLSAEKQSRFRIISDDH<br>NEVLMKRSSDRFVPLLLQYIDYGKLFDHIRFHV<br>NMGKLRYLLKADKTCIDGQTRVRVIEQPLNGF<br>GRLEEAETMRKQENGTFGNSGIRIRDFENMKR |

| | | | |
|---|---|---|---|
| | | | DDANPANYPYIVDTYTHYILENNKVEMFINDK<br>EDSAPLLPVIEDDRYVVKTIPSCRMSTLEIPAM<br>AFHMFLFGSKKTEKLIVDVHNRYKRLFQAMQ<br>KEEVTAENIASFGIAESDLPQKILDLISGNAHGK<br>DVDAFIRLTVDDMLTDTERRIKRFKDDRKSIRS<br>ADNKMGKRGFKQISTGKLADFLAKDIVLFQPS<br>VNDGENKITGLNYRIMQSAIAVYDSGDDYEAK<br>QQFKLMFEKARLIGKGTTEPHPFLYKVFARSIP<br>ANAVEFYERYLIERKFYLTGLSNEIKKGNRVD<br>VPFIRRDQNKWKTPAMKTLGRIYSEDLPVELPR<br>QMFDNEIKSHLKSLPQMEGIDFNNANVTYLIAE<br>YMKRVLDDDFQTFYQWNRNYRYMDMLKGEY<br>DRKGSLQHCFTSVEEREGLWKERASRTERYRK<br>QASNKIRSNRQMRNASSEEIETILDKRLSNSRN<br>EYQKSEKVIRRYRVQDALLFLLAKKTLTELAD<br>FDGERFKLKEIMPDAEKGILSEIMPMSFTFEKG<br>GKKYTITSEGMKLKNYGDFFVLASDKRIGNLL<br>ELVGSDIVSKEDIMEEFNKYDQCRPEISSIVFNL<br>EKWAFDTYPELSARVDREEKVDFKSILKILLNN<br>KNINKEQSDILRKIRNAFDHNNYPDKGVVEIKA<br>LPEIAMSIKKAFGEYAIMK |
| [0506] | 嗜鳃黄杆菌(Flavobacterium branchiophilum)(SEQ ID No. 83) | 16 | MENLNKILDKENEICISKIFNTKGIAAPITEKAL<br>DNIKSKQKNDLNKEARLHYFSIGHSFKQIDTKK<br>VFDYVLIEELKDEKPLKFITLQKDFFTKEFSIKL<br>QKLINSIRNINNHYVHNFNDINLNKIDSNVFHFL<br>KESFELAIIEKYYKVNKKYPLDNEIVLFLKELFI<br>KDENTALLNYFTNLSKDEAIEYILTFTITENKIW<br>NINNEHNILNIEKGKYLTFEAMLFLITIFLYKNE<br>ANHLLPKLYDFKNNKSKQELFTFFSKKFTSQDI<br>DAEEGHLIKFRDMIQYLNHYPTAWNNDLKLES<br>ENKNKIMTTKLIDSIIEFELNSNYPSFATDIQFKK<br>EAKAFLFASNKKRNQTSFSNKSYNEEIRHNPHI<br>KQYRDEIASALTPISFNVKEDKFKIFVKKHVLE<br>EYFPNSIGYEKFLEYNDFTEKEKEDFGLKLYSN<br>PKTNKLIERIDNHKLVKSHGRNQDRFMDFSMR<br>FLAENNYFGKDAFFKCYKFYDTQEQDEFLQSN<br>ENNDDVKFHKGKVTTYIKYEEHLKNYSYWDC<br>PFVEENNSMSVKISIGSEEKILKIQRNLMIYFLE<br>NALYNENVENQGYKLVNNYYRELKKDVEESI<br>ASLDLIKSNPDFKSKYKKILPKRLLHNYAPAKQ<br>DKAPENAFETLLKKADFREEQYKKLLKKAEHE<br>KNKEDFVKRNKGKQFKLHFIRKACQMMYFKE<br>KYNTLKEGNAAFEKKDPVIEKRKNKEHEFGHH |

[0507]

| | | |
|---|---|---|
| | | KNLNITREEFNDYCKWMFAFNGNDSYKKYLR<br>DLFSEKHFFDNQEYKNLFESSVNLEAFYAKTK<br>ELFKKWIETNKPTNNENRYTLENYKNLILQKQ<br>VFINVYHFSKYLIDKNLLNSENNVIQYKSLENV<br>EYLISDFYFQSKLSIDQYKTCGKLFNKLKSNKL<br>EDCLLYEIAYNYIDKKNVHKIDIQKILTSKIILTI<br>NDANTPYKISVPFNKLERYTEMIAIKNQNNLKA<br>RFLIDLPLYLSKNKIKKGKDSAGYEIIIKNDLEIE<br>DINTINNKIINDSVKFTEVLMELEKYFILKDKCI<br>LSKNYIDNSEIPSLKQFSKVWIKENENEIINYRNI<br>ACHFHLPLLETFDNLLLNVEQKFIKEELQNVSTI<br>NDLSKPQEYLILLFIKFKHNNFYLNLFNKNESK<br>TIKNDKEVKKNRVLQKFINQVILKKK |
| 类香味菌 (SEQ ID No. 84) | 17 | MKDILTTDTTEKQNRFYSHKIADKYFFGGYFN<br>LASNNIYEVFEEVNKRNTFGKLAKRDNGNLKN<br>YIIHVFKDELSISDFEKRVAIFASYFPILETVDKK<br>SIKERNRTIDLTLSQRIRQFREMLISLVTAVDQL<br>RNFYTHYHHSDIVIENKVLDFLNSSFVSTALHV<br>KDKYLKTDKTKEFLKETIAAELDILIEAYKKKQ<br>IEKKNTRFKANKREDILNAIYNEAFWSFINDKD<br>KDKDKETVVAKGADAYFEKNHHKSNDPDFAL<br>NISEKGIVYLLSFFLTNKEMDSLKANLTGFKGK<br>VDRESGNSIKYMATQRIYSFHTYRGLKQKIRTS<br>EEGVKETLLMQMIDELSKVPNVVYQHLSTTQQ<br>NSFIEDWNEYYKDYEDDVETDDLSRVTHPVIR<br>KRYEDRFNYFAIRFLDEFFDFPTLRFQVHLGDY<br>VHDRRTKQLGKVESDRIIKEKVTVFARLKDINS<br>AKASYFHSLEEQDKEELDNKWTLFPNPSYDFP<br>KEHTLQHQGEQKNAGKIGIYVKLRDTQYKEK<br>AALEEARKSLNPKERSATKASKYDIITQIIEAND<br>NVKSEKPLVFTGQPIAYLSMNDIHSMLFSLLTD<br>NAELKKTPEEVEAKLIDQIGKQINEILSKDTDTK<br>ILKKYKDNDLKETDTDKITRDLARDKEEIEKLI<br>LEQKQRADDYNYTSSTKFNIDKSRKRKHLLFN<br>AEKGKIGVWLANDIKRFMFKESKSKWKGYQH<br>IELQKLFAYFDTSKSDLELILSNMVMVKDYPIE<br>LIDLVKKSRTLVDFLNKYLEARLEYIENVITRV<br>KNSIGTPQFKTVRKECFTFLKKSNYTVVSLDKQ<br>VERILSMPLFIERGFMDDKPTMLEGKSYKQHK<br>EKFADWFVHYKENSNYQNFYDTEVYEITTEDK<br>REKAKVTKKIKQQQKNDVFTLMMVNYMLEE<br>VLKLSSNDRLSLNELYQTKEERIVNKQVAKDT<br>QERNKNYIWNKVVDLQLCDGLVHIDNVKLKD |

[0508]

| | | |
|---|---|---|
| | | IGNFRKYENDSRVKEFLTYQSDIVWSAYLSNE<br>VDSNKLYVIERQLDNYESIRSKELLKEVQEIECS<br>VYNQVANKESLKQSGNENFKQYVLQGLLPIG<br>MDVREMLILSTDVKFKKEEIIQLGQAGEVEQD<br>LYSLIYIRNKFAHNQLPIKEFFDFCENNYRSISD<br>NEYYAEYYMEIFRSIKEKYAN |
| 柱状黄杆菌(Flavobacterium columnare)(SEQ ID No. 85) | 18 | MSSKNESYNKQKTFNHYKQEDKYFFGGFLNN<br>ADDNLRQVGKEFKTRINFNHNNNELASVFKDY<br>FNKEKSVAKREHALNLLSNYFPVLERIQKHTN<br>HNFEQTREIFELLLDTIKKLRDYYTHHYHKPITI<br>NPKIYDFLDDTLLDVLITIKKKKVKNDTSRELL<br>KEKLRPELTQLKNQKREELIKKGKKLLEENLE<br>NAVFNHCLIPFLEENKTDDKQNKTVSLRKYRK<br>SKPNEETSITLTQSGLVFLMSFFLHRKEFQVFTS<br>GLERFKAKVNTIKEEEISLNKNNIVYMITHWSY<br>SYYNFKGLKHRIKTDQGVSTLEQNNTTHSLTN<br>TNTKEALLTQIVDYLSKVPNEIYETLSEKQQKE<br>FEEDINEYMRENPENEDSTFSSIVSHKVIRKRYE<br>NKFNYFAMRFLDEYAELPTLRFMVNFGDYIKD<br>RQKKILESIQFDSERIIKKEIHLFEKLSLVTEYKK<br>NVYLKETSNIDLSRFPLFPNPSYVMANNNIPFYI<br>DSRSNNLDEYLNQKKKAQSQNKKRNLTFEKY<br>NKEQSKDAIIAMLQKEIGVKDLQQRSTIGLLSC<br>NELPSMLYEVIVKDIKGAELENKIAQKIREQYQ<br>SIRDFTLDSPQKDNIPTTLIKTINTDSSVTFENQP<br>IDIPRLKNALQKELTLTQEKLLNVKEHEIEVDN<br>YNRNKNTYKFKNQPKNKVDDKKLQRKYVFY<br>RNEIRQEANWLASDLIHFMKNKSLWKGYMHN<br>ELQSFLAFFEDKKNDCIALLETVFNLKEDCILT<br>KGLKNLFLKHGNFIDFYKEYLKLKEDFLSTEST<br>FLENGFIGLPPKILKKELSKRLKYIFIVFQKRQFII<br>KELEEKKNNLYADAINLSRGIFDEKPTMIPFKK<br>PNPDEFASWFVASYQYNNYQSFYELTPDIVER<br>DKKKKYKNLRAINKVKIQDYYLKLMVDTLYQ<br>DLFNQPLDKSLSDFYVSKAEREKIKADAKAYQ<br>KLNDSSLWNKVIHLSLQNNRITANPKLKDIGK<br>YKRALQDEKIATLLTYDARTWTYALQKPEKEN<br>ENDYKELHYTALNMELQEYEKVRSKELLKQV<br>QELEKKILDKFYDFSNNASHPEDLEIEDKKGKR<br>HPNFKLYITKALLKNESEIINLENIDIEILLKYYD<br>YNTEELKEKIKNMDEDEKAKIINTKENYNKITN<br>VLIKKALVLIIIRNKMAHNQYPPKFIYDLANRF<br>VPKKEEEYFATYFNRVFETITKELWENKEKKD |

[0509]

| | | KTQV |
|---|---|---|
| 牙龈卟啉单胞菌 (SEQ ID No. 86) | 19 | MTEQNEKPYNGTYYTLEDKHFWAAFLNLARH<br>NAYITLAHIDRQLAYSKADITNDEDILFFKGQW<br>KNLDNDLERKARLRSLILKHFSFLEGAAYGKK<br>LFESQSSGNKSSKKKELSKKEKEELQANALSLD<br>NLKSILFDFLQKLKDFRNYYSHYRHPESSELPL<br>FDGNMLQRLYNVFDVSVQRVKRDHEHNDKV<br>DPHRHFNHLVRKGKKDKYGNNDNPFFKHHFV<br>DREGTVTEAGLLFFVSLFLEKRDAIWMQKKIR<br>GFKGGTEAYQQMTNEVFCRSRISLPKLKLESLR<br>TDDWMLLDMLNELVRCPKSLYDRLREEDRAR<br>FRVPVDILSDEDDTDGTEEDPFKNTLVRHQDRF<br>PYFALRYFDLKKVFTSLRFHIDLGTYHFAIYKK<br>NIGEQPEDRHLTRNLYGFGRIQDFAEEHRPEEW<br>KRLVRDLDYFETGDKPYITQTTPHYHIEKGKIG<br>LRFVPEGQHLWPSPEVGATRTGRSKYAQDKRL<br>TAEAFLSVHELMPMMFYYFLLREKYSEEVSAE<br>KVQGRIKRVIEDVYAVYDAFARDEINTRDELD<br>ACLADKGIRRGHLPRQMIAILSQEHKDMEEKV<br>RKKLQEMIADTDHRLDMLDRQTDRKIRIGRKN<br>AGLPKSGVVADWLVRDMMRFQPVAKDTSGK<br>PLNNSKANSTEYRMLQRALALFGGEKERLTPY<br>FRQMNLTGGNNPHPFLHETRWESHTNILSFYRS<br>YLEARKAFLQSIGRSDRVENHRFLLLKEPKTDR<br>QTLVAGWKGEFHLPRGIFTEAVRDCLIEMGYD<br>EVGSYKEVGFMAKAVPLYFERASKDRVQPFY<br>DYPFNVGNSLKPKKGRFLSKEKRAEEWESGKE<br>RFRLAKLKKEILEAKEHPYHDFKSWQKFEREL<br>RLVKNQDIITWMMCRDLMEENKVEGLDTGTL<br>YLKDIRTDVQEQGSLNVLNRVKPMRLPVVVY<br>RADSRGHVHKEQAPLATVYIEERDTKLLKQGN<br>FKSFVKDRRLNGLFSFVDTGALAMEQYPISKLR<br>VEYELAKYQTARVCAFEQTLELEESLLTRYPH<br>LPDKNFRKMLESWSDPLLDKWPDLHGNVRLLI<br>AVRNAFSHNQYPMYDETLFSSIRKYDPSSPDAI<br>EERMGLNIAHRLSEEVKQAKEMVERIIQA |
| 卟啉单胞菌属某种 (Porphyromonas sp.) | 20 | MTEQSERPYNGTYYTLEDKHFWAAFLNLARH<br>NAYITLTHIDRQLAYSKADITNDQDVLSFKAL<br>WKNFDNDLERKSRLRSLILKHFSFLEGAAYGK<br>KLFESKSSGNKSSKNKELTKKEKEELQANALSL<br>DNLKSILFDFLQKLKDFRNYYSHYRHSESSELP<br>LFDGNMLQRLYNVFDVSVQRVKRDHEHNDKV<br>DPHRHFNHLVRKGKKDRYGHNDNPSFKHHFV |

[0510]

| | | |
|---|---|---|
| COT-052 OH4946 (SEQ ID No. 87) | | DSEGMVTEAGLLFFVSLFLEKRDAIWMQKKIR<br>GFKGGTETYQQMTNEVFCRSRISLPKLKLESLR<br>TDDWMLLDMLNELVRCPKPLYDRLREDDRAC<br>FRVPVDILPDEDDTDGGGEDPFKNTLVRHQDR<br>FPYFALRYFDLKKVFTSLRFHIDLGTYHFAIYK<br>KMIGEQPEDRHLTRNLYGFGRIQDFAEEHRPEE<br>WKRLVRDLDYFETGDKPYISQTTPHYHIEKGKI<br>GLRFVPEGQHLWPSPEVGTTRTGRSKYAQDKR<br>LTAEAFLSVHELMPMMFYYFLLREKYSEEVSA<br>EKVQGRIKRVIEDVYAIYDAFARDEINTLKELD<br>ACLADKGIRRGHLPKQMIGILSQERKDMEEKV<br>RKKLQEMIADTDHRLDMLDRQTDRKIRIGRKN<br>AGLPKSGVIADWLVRDMMRFQPVAKDTSGKP<br>LNNSKANSTEYRMLQRALALFGGEKERLTPYF<br>RQMNLTGGNNPHPFLHETRWESHTNILSFYRS<br>YLRARKAFLERIGRSDRVENCPFLLLKEPKTDR<br>QTLVAGWKGEFHLPRGIFTEAVRDCLIEMGYD<br>EVGSYREVGFMAKAVPLYFERACEDRVQPFY<br>DSPFNVGNSLKPKKGRFLSKEDRAEEWERGKE<br>RFRDLEAWSHSAARRIKDAFAGIEYASPGNKK<br>KIEQLLRDLSLWEAFESKLKVRADKINLAKLK<br>KEILEAQEHPYHDFKSWQKFERELRLVKNQDII<br>TWMMCRDLMEENKVEGLDTGTLYLKDIRPNV<br>QEQGSLNVLNRVKPMRLPVVVYRADSRGHVH<br>KEEAPLATVYIEERDTKLLKQGNFKSFVKDRR<br>LNGLFSFVDTGGLAMEQYPISKLRVEYELAKY<br>QTARVCVFELTLRLEESLLSRYPHLPDESFREM<br>LESWSDPLLAKWPELHGKVRLLIAVRNAFSHN<br>QYPMYDEAVFSSIRKYDPSSPDAIEERMGLNIA<br>HRLSEEVKQAKETVERIIQA |
| 中间普雷沃菌 (SEQ ID No. 88) | 21 | MEDDKKTKESTNMLDNKHFWAAFLNLARHN<br>VYITVNHINKVLELKNKKDQDIIIDNDQDILAIK<br>THWEKVNGDLNKTERLRELMTKHFPFLETAIY<br>TKNKEDKEEVKQEKQAKAQSFDSLKHCLFLFL<br>EKLQEARNYYSHYKYSESTKEPMLEKELLKK<br>MYNIFDDNIQLVIKDYQHNKDINPDEDFKHLD<br>RTEEEFNYYFTTNKKGNITASGLLFFVSLFLEK<br>KDAIWMQQKLRGFKDNRESKKKMTHEVFCRS<br>RMLLPKLRLESTQTQDWILLDMLNELIRCPKSL<br>YERLQGEYRKKFNVPFDSADEDYDAEQEPFKN<br>TLVRHQDRFPYFALRYFDYNEIFTNLRFQIDLG<br>TYHFSIYKKLIGGQKEDRHLTHKLYGFERIQEF<br>AKQNRTDEWKAIVKDFDTYETSEEPYISETAPH |

[0511]

| | | |
|---|---|---|
| | | YHLENQKIGIRFRNDNDEIWPSLKTNGENNEKR<br>KYKLDKQYQAEAFLSVHELLPMMFYYLLLKK<br>EEPNNDKKNASIVEGFIKREIRDIYKLYDAFAN<br>GEINNIDDLEKYCEDKGIPKRHLPKQMVAILYD<br>EHKDMAEEAKRKQKEMVKDTKKLLATLEKQT<br>QGEIEDGGRNIRLLKSGEIARWLVNDMMRFQP<br>VQKDNEGNPLNNSKANSTEYQMLQRSLALYN<br>KEEKPTRYFRQVNLINSSNPHPFLKWTKWEEC<br>NNILSFYRSYLTKKIEFLNKLKPEDWEKNQYFL<br>KLKEPKTNRETLVQGWKNGFNLPRGIFTEPIRE<br>WFKRHQNDSEEYEKVETLDRVGLVTKVIPLFF<br>KKEDSKDKEEYLKKDAQKEINNCVQPFYGFPY<br>NVGNIHKPDEKDFLPSEERKKLWGDKKYKFK<br>GYKAKVKSKKLTDKEKEEYRSYLEFQSWNKF<br>ERELRLVRNQDIVTWLLCTELIDKLKVEGLNV<br>EELKKLRLKDIDTDTAKQEKNNILNRVMPMQL<br>PVTVYEIDDSHNIVKDRPLHTVYIEETKTKLLK<br>QGNFKALVKDRRLNGLFSFVDTSSETELKSNPI<br>SKSLVEYELGEYQNARIETIKDMLLLEETLIEK<br>YKTLPTDNFSDMLNGWLEGKDEADKARFQND<br>VKLLVAVRNAFSHNQYPMRNRIAFANINPFSLS<br>SADTSEEKKLDIANQLKDKTHKIIKRIIEIEKPIE<br>TKE |
| PIN17_0200 [中间普雷沃菌 17] (SEQ ID No. 89) | AFJ07523 | MKMEDDKKTKESTNMLDNKHFWAAFLNLAR<br>HNVYITVNHINKVLELKNKKDQDIIIDNDQDIL<br>AIKTHWEKVNGDLNKTERLRELMTKHFPFLET<br>AIYTKNKEDKEEVKQEKQAKAQSFDSLKHCLF<br>LFLEKLQEARNYYSHYKYSESTKEPMLEKELL<br>KKMYNIFDDNIQLVIKDYQHNKDINPDEDFKH<br>LDRTEEEFNYYFTTNKKGNITASGLLFFVSLFL<br>EKKDAIWMQQKLRGFKDNRESKKKMTHEVFC<br>RSRMLLPKLRLESTQTQDWILLDMLNELIRCPK<br>SLYERLQGEYRKKFNVPFDSADEDYDAEQEPF<br>KNTLVRHQDRFPYFALRYFDYNEIFTNLRFQID<br>LGTYHFSIYKKLIGGQKEDRHLTHKLYGFERIQ<br>EFAKQNRTDEWKAIVKDFDTYETSEEPYISETA<br>PHYHLENQKIGIRFRNDNDEIWPSLKTNGENNE<br>KRKYKLDKQYQAEAFLSVHELLPMMFYYLLL<br>KKEEPNNDKKNASIVEGFIKREIRDIYKLYDAF<br>ANGEINNIDDLEKYCEDKGIPKRHLPKQMVAIL<br>YDEHKDMAEEAKRKQKEMVKDTKKLLATLE<br>KQTQGEIEDGGRNIRLLKSGEIARWLVNDMMR<br>FQPVQKDNEGNPLNNSKANSTEYQMLQRSLAL |

[0512]

| | | |
|---|---|---|
| | | YNKEEKPTRYFRQVNLINSSNPHPFLKWTKWE<br>ECNNILSFYRSYLTKKIEFLNKLKPEDWEKNQY<br>FLKLKEPKTNRETLVQGWKNGFNLPRGIFTEPI<br>REWFKRHQNDSEEYEKVETLDRVGLVTKVIPL<br>FFKKEDSKDKEEYLKKDAQKEINNCVQPFYGF<br>PYNVGNIHKPDEKDFLPSEERKKLWGDKKYKF<br>KGYKAKVKSKKLTDKEKEEYRSYLEFQSWNK<br>FERELRLVRNQDIVTWLLCTELIDKLKVEGLNV<br>EELKKLRLKDIDTDTAKQEKNNILNRVMPMQL<br>PVTVYEIDDSHNIVKDRPLHTVYIEETKTKLLK<br>QGNFKALVKDRRLNGLFSFVDTSSETELKSNPI<br>SKSLVEYELGEYQNARIETIKDMLLLEETLIEK<br>YKTLPTDNFSDMLNGWLEGKDEADKARFQND<br>VKLLVAVRNAFSHNQYPMRNRIAFANINPFSLS<br>SADTSEEKKLDIANQLKDKTHKIIKRIIEIEKPIE<br>TKE |
| 中间普雷沃菌(SEQ ID No. 90) | BAU18623 | MEDDKKTTDSISYELKDKHFWAAFLNLARHN<br>VYITVNHINKVLELKNKKDQDIIIDNDQDILAIK<br>THWEKVNGDLNKTERLRELMTKHFPFLETAIY<br>SKNKEDKEEVKQEKQAKAQSFDSLKHCLFLFL<br>EKLQETRNYYSHYKYSESTKEPMLEKELLKKM<br>YNIFDDNIQLVIKDYQHNKDINPDEDFKHLDRT<br>EEDFNYYFTRNKKGNITESGLLFFVSLFLEKKD<br>AIWMQQKLRGFKDNRESKKKMTHEVFCRSRM<br>LLPKLRLESTQTQDWILLDMLNELIRCPKSLYE<br>RLQGEDREKFKVPFDPADEDYDAEQEPFKNTL<br>VRHQDRFPYFALRYFDYNEIFTNLRFQIDLGTF<br>HFSIYKKLIGGQKEDRHLTHKLYGFERIQEFAK<br>QNRPDEWKAIVKDLDTYETSNERYISETTPHY<br>HLENQKIGIRFRNDNDEIWPSLKTNGENNEKSK<br>YKLDKQYQAEAFLSVHELLPMMFYYLLLKKE<br>EPNNDKKNASIVEGFIKREIRDMYKLYDAFAN<br>GEINNIDDLEKYCEDKGIPKRHLPKQMVAILYD<br>EHKDMVKEAKRKQRKMVKDTEKLLAALEKQ<br>TQEKTEDGGRNIRLLKSGEIARWLVNDMMRFQ<br>PVQKDNEGNPLNNSKANSTEYQMLQRSLALY<br>NKEEKPTRYFRQVNLINSSNPHPFLKWTKWEE<br>CNNILSFYRSYLTKKIEFLNKLKPEDWEKNQYF<br>LKLKEPKTNRETLVQGWKNGFNLPRGIFTEPIR<br>EWFKRHQNDSKEYEKVEALDRVGLVTKVIPLF<br>FKKEDSKDKEEDLKKDAQKEINNCVQPFYSFP<br>YNVGNIHKPDEKDFLHREERIELWDKKKDKFK<br>GYKAKVKSKKLTDKEKEEYRSYLEFQSWNKF |

| | | |
|---|---|---|
| | | ERELRLVRNQDIVTWLLCTELIDKLKVEGLNV EELKKLRLKDIDTDTAKQEKNNILNRVMPMQL PVTVYEIDDSHNIVKDRPLHTVYIEETKTKLLK QGNFKALVKDRRLNGLFSFVDTSSEAELKSNPI SKSLVEYELGEYQNARIETIKDMLLLEETLIEK YKNLPTDNFSDMLNGWLEGKDEADKARFQND VKLLVAVRNAFSHNQYPMRNRIAFANINPFSLS SADTSEEKKLDIANQLKDKTHKIIKRIIEIEKPIE TKE |
| HMPREF6485_0083 [颊普雷沃菌 ATCC 33574] (SEQ ID No. 91) | EFU31981 | MQKQDKLFVDRKKNAIFAFPKYITIMENKEKP EPIYYELTDKHFWAAFLNLARHNVYTTINHINR RLEIAELKDDGYMMGIKGSWNEQAKKLDKKV RLRDLIMKHFPFLEAAAYEMTNSKSPNNKEQR EKEQSEALSLNNLKNVLFIFLEKLQVLRNYYSH YKYSEESPKPIFETSLLKNMYKVFDANVRLVK RDYMHHENIDMQRDFTHLNRKKQVGRTKNIID SPNFHYHFADKEGNMTIAGLLFFVSLFLDKKD AIWMQKKLKGFKDGRNLREQMTNEVFCRSRIS LPKLKLENVQTKDWMQLDMLNELVRCPKSLY ERLREKDRESFKVPFDIFSDDYNAEEEPFKNTL VRHQDRFPYFVLRYFDLNEIFEQLRFQIDLGTY HFSIYNKRIGDEDEVRHLTHHLYGFARIQDFAP QNQPEEWRKLVKDLDHFETSQEPYISKTAPHY HLENEKIGIKFCSAHNNLFPSLQTDKTCNGRSK FNLGTQFTAEAFLSVHELLPMMFYYLLLTKDY SRKESADKVEGIIRKEISNIYAIYDAFANNEINSI ADLTRRLQNTNILQGHLPKQMISILKGRQKDM GKEAERKIGEMIDDTQRRLDLLCKQTNQKIRIG KRNAGLLKSGKIADWLVNDMMRFQPVQKDQ NNIPINNSKANSTEYRMLQRALALFGSENFRLK AYFNQMNLVGNDNPHPFLAETQWEHQTNILSF YRNYLEARKKYLKGLKPQNWKQYQHFLILKV QKTNRNTLVTGWKNSFNLPRGIFTQPIREWFEK HNNSKRIYDQILSFDRVGFVAKAIPLYFAEEYK DNVQPFYDYPFNIGNRLKPKKRQFLDKKERVE LWQKNKELFKNYPSEKKKTDLAYLDFLSWKK FERELRLIKNQDIVTWLMFKELFNMATVEGLKI GEIHLRDIDTNTANEESNNILNRIMPMKLPVKT YETDNKGNILKERPLATFYIEETETKVLKQGNF KALVKDRRLNGLFSFAETTDLNLEEHPISKLSV DLELIKYQTTRISIFEMTLGLEKKLIDKYSTLPT DSFRNMLERWLQCKANRPELKNYVNSLIAVR NAFSHNQYPMYDATLFAEVKKFTLFPSVDTKK |

[0513]

[0514]

| | | |
|---|---|---|
| | | IELNIAPQLLEIVGKAIKEIEKSENKN |
| HMPREF9144_1146 [苍白普雷沃菌 (Prevotella pallens) ATCC 700821] (SEQ ID No. 92) | EGQ18444 | MKEEEKGKTPVVSTYNKDDKHFWAAFLNLAR HNVYITVNHINKILGEGEINRDGYENTLEKSWN EIKDINKKDRLSKLIIKHFPFLEVTTYQRNSADT TKQKEEKQAEAQSLESLKKSFFVFIYKLRDLRN HYSHYKHSKSLERPKFEEDLQEKMYNIFDASIQ LVKEDYKHNTDIKTEEDFKHLDRKGQFKYSFA DNEGNITESGLLFFVSLFLEKKDAIWVQKKLEG FKCSNESYQKMTNEVFCRSRMLLPKLRLQSTQ TQDWILLDMLNELIRCPKSLYERLREEDRKKFR VPIEIADEDYDAEQEPFKNALVRHQDRFPYFAL RYFDYNEIFTNLRFQIDLGTYHFSIYKKQIGDY KESHHLTHKLYGFERIQEFTKQNRPDEWRKFV KTFNSFETSKEPYIPETTPHYHLENQKIGIRFRN DNDKIWPSLKTNSEKNEKSKYKLDKSFQAEAF LSVHELLPMMFYYLLLKTENTDNDNEIETKKK ENKNDKQEKHKIEEIIENKITEIYALYDAFANG KINSIDKLEEYCKGKDIEIGHLPKQMIAILKSEH KDMATEAKRKQEEMLADVQKSLESLDNQINE EIENVERKNSSLKSGEIASWLVNDMMRFQPVQ KDNEGNPLNNSKANSTEYQMLQRSLALYNKE EKPTRYFRQVNLIESSNPHPFLNNTEWEKCNNI LSFYRSYLEAKKNFLESLKPEDWEKNQYFLML KEPKTNCETLVQGWKNGFNLPRGIFTEPIRKWF MEHRKNITVAELKRVGLVAKVIPLFFSEEYKDS VQPFYNYLFNVGNINKPDEKNFLNCEERRELL RKKKDEFKKMTDKEKEENPSYLEFQSWNKFE RELRLVRNQDIVTWLLCMELFNKKKIKELNVE KIYLKNINTNTTKKEKNTEEKNGEEKIIKEKNNI LNRIMPMRLPIKVYGRENFSKNKKKKIRRNTFF TVYIEEKGTKLLKQGNFKALERDRRLGGLFSF VKTHSKAESKSNTISKSRVEYELGEYQKARIEII KDMLALEETLIDKYNSLDTDNFHNMLTGWLK LKDEPDKASFQNDVDLLIAVRNAFSHNQYPMR NRIAFANINPFSLSSANTSEEKGLGIANQLKDKT HKTIEKIIEIEKPIETKE |
| HMPREF9714_02132 [类香味菌 CCUG | EHO08761 | MKDILTTDTTEKQNRFYSHKIADKYFFGGYFN LASNNIYEVFEEVNKRNTFGKLAKRDNGNLKN YIIHVFKDELSISDFEKRVAIFASYFPILETVDKK SIKERNRTIDLTLSQRIRQFREMLISLVTAVDQL RNFYTHYHHSEIVIENKVLDFLNSSLVSTALHV KDKYLKTDKTKEFLKETIAAELDILIEAYKKKQ IEKKNTRFKANKREDILNAIYNEAFWSFINDKD |

[0515]

| | | |
|---|---|---|
| 12901] (SEQ ID No. 93) | | KDKETVVAKGADAYFEKNHHKSNDPDFALNIS EKGIVYLLSFFLTNKEMDSLKANLTGFKGKVD RESGNSIKYMATQRIYSFHTYRGLKQKIRTSEE GVKETLLMQMIDELSKVPNVVYQHLSTTQQNS FIEDWNEYYKDYEDDVETDDLSRVIHPVIRKR YEDRFNYFAIRFLDEFFDFPTLRFQVHLGDYVH DRRTKQLGKVESDRIIKEKVTVFARLKDINSAK ANYFHSLEEQDKEELDNKWTLFPNPSYDFPKE HTLQHQGEQKNAGKIGIYVKLRDTQYKEKAA LEEARKSLNPKERSATKASKYDIITQIIEANDNV KSEKPLVFTGQPIAYLSMNDIHSMLFSLLTDNA ELKKTPEEVEAKLIDQIGKQINEILSKDTDTKIL KKYKDNDLKETDTDKITRDLARDKEEIEKLILE QKQRADDYNYTSSTKFNIDKSRKRKHLLFNAE KGKIGVWLANDIKRFMTEEFKSKWKGYQHTE LQKLFAYYDTSKSDLDLILSDMVMVKDYPIELI ALVKKSRTLVDFLNKYLEARLGYMENVITRVK NSIGTPQFKTVRKECFTFLKKSNYTVVSLDKQV ERILSMPLFIERGFMDDKPTMLEGKSYQQHKE KFADWFVHYKENSNYQNFYDTEVYEITTEDKR EKAKVTKKIKQQQKNDVFTLMMVNYMLEEVL KLSSNDRLSLNELYQTKEERIVNKQVAKDTQE RNKNYIWNKVVDLQLCEGLVRIDKVKLKDIGN FRKYENDSRVKEFLTYQSDIVWSAYLSNEVDS NKLYVIERQLDNYESIRSKELLKEVQEIECSVY NQVANKESLKQSGNENFKQYVLQGLVPIGMD VREMLILSTDVKFIKEEIIQLGQAGEVEQDLYSL IYIRNKFAHNQLPIKEFFDFCENNYRSISDNEYY AEYYMEIFRSIKEKYTS |
| HMPREF9711_00870 [类香味菌 CCUG 3837] (SEQ ID No. 94) | EKB06014 | MKDILTTDTTEKQNRFYSHKIADKYFFGGYFN LASNNIYEVFEEVNKRNTFGKLAKRDNGNLKN YIIHVFKDELSISDFEKRVAIFASYFPILETVDKK SIKERNRTIDLTLSQRIRQFREMLISLVTAVDQL RNFYTHYHHSEIVIENKVLDFLNSSLVSTALHV KDKYLKTDKTKEFLKETIAAELDILIEAYKKKQ IEKKNTRFKANKREDILNAIYNEAFWSFINDKD KDKETVVAKGADAYFEKNHHKSNDPDFALNIS EKGIVYLLSFFLTNKEMDSLKANLTGFKGKVD RESGNSIKYMATQRIYSFHTYRGLKQKIRTSEE GVKETLLMQMIDELSKVPNVVYQHLSTTQQNS FIEDWNEYYKDYEDDVETDDLSRVIHPVIRKR YEDRFNYFAIRFLDEFFDFPTLRFQVHLGDYVH DRRTKQLGKVESDRIIKEKVTVFARLKDINSAK |

[0516]

| | | |
|---|---|---|
| | | ASYFHSLEEQDKEELDNKWTLFPNPSYDFPKE<br>HTLQHQGEQKNAGKIGIYVKLRDTQYKEKAA<br>LEEARKSLNPKERSATKASKYDIITQIIEANDNV<br>KSEKPLVFTGQPIAYLSMNDIHSMLFSLLTDNA<br>ELKKTPEEVEAKLIDQIGKQINEILSKDTDTKIL<br>KKYKDNDLKETDTDKITRDLARDKEEIEKLILE<br>QKQRADDYNYTSSTKFNIDKSRKRKHLLFNAE<br>KGKIGVWLANDIKRFMFKESKSKWKGYQHTE<br>LQKLFAYFDTSKSDLELILSDMVMVKDYPIELI<br>DLVRKSRTLVDFLNKYLEARLGYIENVITRVK<br>NSIGTPQFKTVRKECFAFLKESNYTVASLDKQI<br>ERILSMPLFIERGFMDSKPTMLEGKSYQQHKED<br>FADWFVHYKENSNYQNFYDTEVYEIITEDKRE<br>QAKVTKKIKQQQKNDVFTLMMVNYMLEEVL<br>KLPSNDRLSLNELYQTKEERIVNKQVAKDTQE<br>RNKNYIWNKVVDLQLCEGLVRIDKVKLKDIGN<br>FRKYENDSRVKEFLTYQSDIVWSGYLSNEVDS<br>NKLYVIERQLDNYESIRSKELLKEVQEIECIVYN<br>QVANKESLKQSGNENFKQYVLQGLLPRGTDV<br>REMLILSTDVKFKKEEIMQLGQVREVEQDLYS<br>LIYIRNKFAHNQLPIKEFFDFCENNYRPISDNEY<br>YAEYYMEIFRSIKEKYAS |
| HMPREF9699_02005 [动物溃疡伯杰菌 ATCC 43767] (SEQ ID No. 95) | EKB54193 | MENKTSLGNNIYYNPFKPQDKSYFAGYFNAA<br>MENTDSVFRELGKRLKGKEYTSENFFDAIFKE<br>NISLVEYERYVKLLSDYFPMARLLDKKEVPIKE<br>RKENFKKNFKGIIKAVRDLRNFYTHKEHGEVEI<br>TDEIFGVLDEMLKSTVLTVKKKKVKTDKTKEI<br>LKKSIEKQLDILCQKKLEYLRDTARKIEEKRRN<br>QRERGEKELVAPFKYSDKRDDLIAAIYNDAFD<br>VYIDKKKDSLKESSKAKYNTKSDPQQEEGDLK<br>IPISKNGVVFLLSLFLTKQEIHAFKSKIAGFKAT<br>VIDEATVSEATVSHGKNSICFMATHEIFSHLAY<br>KKLKRKVRTAEINYGEAENAEQLSVYAKETL<br>MMQMLDELSKVPDVVYQNLSEDVQKTFIEDW<br>NEYLKENNGDVGTMEEEQVIHPVIRKRYEDKF<br>NYFAIRFLDEFAQFPTLRFQVHLGNYLHDSRPK<br>ENLISDRRIKEKITVFGRLSELEHKKALFIKNTE<br>TNEDREHYWEIFPNPNYDFPKENISVNDKDFPI<br>AGSILDREKQPVAGKIGIKVKLLNQQYVSEVD<br>KAVKAHQLKQRKASKPSIQNIIEEIVPINESNPK<br>EAIVFGGQPTAYLSMNDIHSILYEFFDKWEKKK<br>EKLEKKGEKELRKEIGKELEKKIVGKIQAQIQQI<br>IDKDTNAKILKPYQDGNSTAIDKEKLIKDLKQE |

[0517]

| | | |
|---|---|---|
| | | QNILQKLKDEQTVREKEYNDFIAYQDKNREIN<br>KVRDRNHKQYLKDNLKRKYPEAPARKEVLYY<br>REKGKVAVWLANDIKRFMPTDFKNEWKGEQH<br>SLLQKSLAYYEQCKEELKNLLPEKVFQHLPFKL<br>GGYFQQKYLYQFYTCYLDKRLEYISGLVQQAE<br>NFKSENKVFKKVENECFKFLKKQNYTHKELDA<br>RVQSILGYPIFLERGFMDEKPTIIKGKTFKGNEA<br>LFADWFRYYKEYQNFQTFYDTENYPLVELEKK<br>QADRKRKTKIYQQKKNDVFTLLMAKHIFKSVF<br>KQDSIDQFSLEDLYQSREERLGNQERARQTGE<br>RNTNYIWNKTVDLKLCDGKITVENVKLKNVG<br>DFIKYEYDQRVQAFLKYEENIEWQAFLIKESKE<br>EENYPYVVEREIEQYEKVRREELLKEVHLIEEY<br>ILEKVKDKEILKKGDNQNFKYYILNGLLKQLK<br>NEDVESYKVFNLNTEPEDVNINQLKQEATDLE<br>QKAFVLTYIRNKFAHNQLPKKEFWDYCQEKY<br>GKIEKEKTYAEYFAEVFKKEKEALIK |
| HMPREF9151_01387 [解糖普雷沃菌 F0055] (SEQ ID No. 96) | EKY00089 | MMEKENVQGSHIYYEPTDKCFWAAFYNLARH<br>NAYLTIAHINSFVNSKKGINNDDKVLDIIDDWS<br>KFDNDLLMGARLNKLILKHFPFLKAPLYQLAK<br>RKTRKQQGKEQQDYEKKGDEDPEVIQEAIANA<br>FKMANVRKTLHAFLKQLEDLRNHFSHYNYNS<br>PAKKMEVKFDDGFCNKLYYVFDAALQMVKD<br>DNRMNPEINMQTDFEHLVRLGRNRKIPNTFKY<br>NFTNSDGTINNNGLLFFVSLFLEKRDAIWMQK<br>KIKGFKGGTENYMRMTNEVFCRNRMVIPKLRL<br>ETDYDNHQLMFDMLNELVRCPLSLYKRLKQE<br>DQDKFRVPIEFLDEDNEADNPYQENANSDENP<br>TEETDPLKNTLVRHQHRFPYFVLRYFDLNEVF<br>KQLRFQINLGCYHFSIYDKTIGERTEKRHLTRT<br>LFGFDRLQNFSVKLQPEHWKNMVKHLDTEESS<br>DKPYLSDAMPHYQIENEKIGIHFLKTDTEKKET<br>VWPSLEVEEVSSNRNKYKSEKNLTADAFLSTH<br>ELLPMMFYYQLLSSEEKTRAAAGDKVQGVLQ<br>SYRKKIFDIYDDFANGTINSMQKLDERLAKDN<br>LLRGNMPQQMLAILEHQEPDMEQKAKEKLDR<br>LITETKKRIGKLEDQFKQKVRIGKRRADLPKVG<br>SIADWLVNDMMRFQPAKRNADNTGVPDSKAN<br>STEYRLLQEALAFYSAYKDRLEPYFRQVNLIGG<br>TNPHPFLHRVDWKKCNHLLSFYHDYLEAKEQ<br>YLSHLSPADWQKHQHFLLLKVRKDIQNEKKD<br>WKKSLVAGWKNGFNLPRGLFTESIKTWFSTDA<br>DKVQITDTKLFENRVGLIAKLIPLYYDKVYND |

[0518]

| | | |
|---|---|---|
| | | KPQPFYQYPFNINDRYKPEDTRKRFTAASSKL<br>WNEKKMLYKNAQPDSSDKIEYPQYLDFLSWK<br>KLERELRMLRNQDMMVWLMCKDLFAQCTVE<br>GVEFADLKLSQLEVDVNVQDNLNVLNNVSSMI<br>LPLSVYPSDAQGNVLRNSKPLHTVYVQENNTK<br>LLKQGNFKSLLKDRRLNGLFSFIAAEGEDLQQ<br>HPLTKNRLEYELSIYQTMRISVFEQTLQLEKAIL<br>TRNKTLCGNNFNNLLNSWSEHRTDKKTLQPDI<br>DFLIAVRNAFSHNQYPMSTNTVMQGIEKFNIQT<br>PKLEEKDGLGIASQLAKKTKDAASRLQNIINGG<br>TN |
| A343_1752 [牙龈卟啉单胞菌JCVISC001] (SEQ ID No. 97) | EOA10535 | MTEQNEKPYNGTYYTLEDKHFWAAFFNLARH<br>NAYITLTHIDRQLAYSKADITNDEDILFFKGQW<br>KNLDNDLERKARLRSLILKHFSFLEGAAYGKK<br>LFESQSSGNKSSKKKELTKKEKEELQANALSLD<br>NLKSILFDFLQKLKDFRNYYSHYRHPESSELPL<br>FDGNMLQRLYNVFDVSVQRVKRDHEHNDKV<br>DPHRHFNHLVRKGKKDRCGNNDNPFFKHHFV<br>DREEKVTEAGLLFFVSLFLEKRDAIWMQKKIR<br>GFKGGTETYQQMTNEVFCRSRISLPKLKLESLR<br>TDDWMLLDMLNELVRCPKSLYDRLREEDRAR<br>FRVPVDILSDEDDTDGTEEDPFKNTLVRHQDRF<br>PYFALRYFDLKKVFTSLRFHIDLGTYHFAIYKK<br>NIGEQPEDRHLTRNLYGFGRIQDFAEEHRPEEW<br>KRLVRDLDYFETGDKPYITQTTPHYHIEKGKIG<br>LRFVPEGQLLWPSPEVGATRTGRSKYAQDKRF<br>TAEAFLSVHELMPMMFYYFLLREKYSEEASAE<br>RVQGRIKRVIEDVYAVYDAFARGEIDTLDRLD<br>ACLADKGIRRGHLPRQMIAILSQEHKDMEEKV<br>RKKLQEMIADTDHRLDMLDRQTDRKIRIGRKN<br>AGLPKSGVIADWLVRDMMRFQPVAKDTSGKP<br>LNNSKANSTEYRMLQRALALFGGEKERLTPYF<br>RQMNLTGGNNPHPFLHETRWESHTNILSFYRS<br>YLKARKAFLQSIGRSDRVENHRFLLLKEPKTDR<br>QTLVAGWKGEFHLPRGIFTEAVRDCLIEMGLD<br>EVGSYKEVGFMAKAVPLYFERACKDRVQPFY<br>DYPFNVGNSLKPKKGRFLSKEKRAEEWESGKE<br>RFRDLEAWSHSAARRIEDAFAGIENASRENKK<br>KIEQLLQDLSLWETFESKLKVKADKINIAKLKK<br>EILEAKEHPYLDFKSWQKFERELRLVKNQDIIT<br>WMMCRDLMEENKVEGLDTGTLYLKDIRTDVH<br>EQGSLNVLNRVKPMRLPVVVYRADSRGHVHK<br>EQAPLATVYIEERDTKLLKQGNFKSFVKDRRL |

[0519]

| | | |
|---|---|---|
| | | NGLFSFVDTGALAMEQYPISKLRVEYELAKYQ<br>TARVCAFEQTLELEESLLTRYPHLPDKNFRKM<br>LESWSDPLLDKWPDLHGNVRLLIAVRNAFSHN<br>QYPMYDETLFSSIRKYDPSSPDAIEERMGLNIA<br>HRLSEEVKQAKEMVERIIQA |
| HMPREF1981_03090 [化脓拟杆菌F0041] (SEQ ID No. 98) | ERI81700 | MESIKNSQKSTGKTLQKDPPYFGLYLNMALLN<br>VRKVENHIRKWLGDVALLPEKSGFHSLLTTDN<br>LSSAKWTRFYYKSRKFLPFLEMFDSDKKSYEN<br>RRETTECLDTIDRQKISSLLKEVYGKLQDIRNA<br>FSHYHIDDQSVKHTALIISSEMHRFIENAYSFAL<br>QKTRARFTGVFVETDFLQAEEKGDNKKFFAIG<br>GNEGIKLKDNALIFLICLFLDREEAFKFLSRATG<br>FKSTKEKGFLAVRETFCALCCRQPHERLLSVNP<br>REALLMDMLNELNRCPDILFEMLDEKDQKSFL<br>PLLGEEEQAHILENSLNDELCEAIDDPFEMIASL<br>SKRVRYKNRFPYLMLRYIEEKNLLPFIRFRIDL<br>GCLELASYPKKMGEENNYERSVTDHAMAFGR<br>LTDFHNEDAVLQQITKGITDEVRFSLYAPRYAI<br>YNNKIGFVRTGGSDKISFPTLKKKGGEGHCVA<br>YTLQNTKSFGFISIYDLRKILLLSFLDKDKAKNI<br>VSGLLEQCEKHWKDLSENLFDAIRTELQKEFP<br>VPLIRYTLPRSKGGKLVSSKLADKQEKYESEFE<br>RRKEKLTEILSEKDFDLSQIPRRMIDEWLNVLP<br>TSREKKLKGYVETLKLDCRERLRVFEKREKGE<br>HPVPPRIGEMATDLAKDIIRMVIDQGVKQRITS<br>AYYSEIQRCLAQYAGDDNRRHLDSIIRELRLKD<br>TKNGHPFLGKVLRPGLGHTEKLYQRYFEEKKE<br>WLEATFYPAASPKRVPRFVNPPTGKQKELPLII<br>RNLMKERPEWRDWKQRKNSHPIDLPSQLFENE<br>ICRLLKDKIGKEPSGKLKWNEMFKLYWDKEFP<br>NGMQRFYRCKRRVEVFDKVVEYEYSEEGGNY<br>KKYYEALIDEVVRQKISSSKEKSKLQVEDLTLS<br>VRRVFKRAINEKEYQLRLLCEDDRLLFMAVRD<br>LYDWKEAQLDLDKIDNMLGEPVSVSQVIQLEG<br>GQPDAVIKAECKLKDVSKLMRYCYDGRVKGL<br>MPYFANHEATQEQVEMELRHYEDHRRRVFNW<br>VFALEKSVLKNEKLRRFYEESQGGCEHRRCID<br>ALRKASLVSEEEYEFLVHIRNKSAHNQFPDLEI<br>GKLPPNVTSGFCECIWSKYKAIICRIIPFIDPERR<br>FFGKLLEQK |
| HMPREF1553_02065 | ERJ65637 | MNTVPASENKGQSRTVEDDPQYFGLYLNLARE<br>NLIEVESHVRIKFGKKKLNEESLKQSLLCDHLL<br>SVDRWTKVYGHSRRYLPFLHYFDPDSQIEKDH |

[0520]

| | | |
|---|---|---|
| [牙龈卟啉单胞菌F0568](SEQ ID No. 99) | | DSKTGVDPDSAQRLIRELYSLLDFLRNDFSHNR<br>LDGTTFEHLEVSPDISSFITGTYSLACGRAQSRF<br>ADFFKPDDFVLAKNRKEQLISVADGKECLTVS<br>GLAFFICLFLDREQASGMLSRIRGFKRTDENWA<br>RAVHETFCDLCIRHPHDRLESSNTKEALLLDML<br>NELNRCPRILYDMLPEEERAQFLPALDENSMN<br>NLSENSLNEESRLLWDGSSDWAEALTKRIRHQ<br>DRFPYLMLRFIEEMDLLKGIRFRVDLGEIELDS<br>YSKKVGRNGEYDRTITDHALAFGKLSDFQNEE<br>EVSRMISGEASYPVRFSLFAPRYAIYDNKIGYC<br>HTSDPVYPKSKTGEKRALSNPRSMGFISVHDLR<br>KLLLMELLCEGSFSRMQSDFLRKANRILDETAE<br>GKLQFSALFPEMRHRFIPPQNPKSKDRREKAET<br>TLEKYKQEIKGRKDKLNSQLLSAFDMDQRQLP<br>SRLLDEWMNIRPASHSVKLRTYVKQLNEDCRL<br>RLQKFRKDGDGKARAIPLVGEMATFLSQDIVR<br>MIISEETKKLITSAYYNEMQRSLAQYAGEENRH<br>QFRAIVAELRLLDPSSGHPFLSATMETAHRYTE<br>DFYKCYLEKKREWLAKTFYRPEQDENTKRRIS<br>VFFVPDGEARKLLPLLIRRRMKEQNDLQDWIR<br>NKQAHPIDLPSHLFDSKIMELLKVKDGKKKWN<br>EAFKDWWSTKYPDGMQPFYGLRRELNIHGKS<br>VSYIPSDGKKFADCYTHLMEKTVQDKKRELRT<br>AGKPVPPDLAADIKRSFHRAVNEREFMLRLVQ<br>EDDRLMLMAINKMMTDREEDILPGLKNIDSIL<br>DEENQFSLAVHAKVLEKEGEGGDNSLSLVPAT<br>IEIKSKRKDWSKYIRYRYDRRVPGLMSHFPEHK<br>ATLDEVKTLLGEYDRCRIKIFDWAFALEGAIMS<br>DRDLKPYLHESSSREGKSGEHSTLVKMLVEKK<br>GCLTPDESQYLILIRNKAAHNQFPCAAEMPLIY<br>RDVSAKVGSIEGSSAKDLPEGSSLVDSLWKKY<br>EMIIRKILPILDPENRFFGKLLNNMSQPINDL |
| HMPREF1988_01768[牙龈卟啉单胞菌F0185](SEQ ID No. 100) | ERJ81987 | MNTVPASENKGQSRTVEDDPQYFGLYLNLARE<br>NLIEVESHVRIKFGKKKLNEESLKQSLLCDHLL<br>SVDRWTKVYGHSRRYLPFLHYFDPDSQIEKDH<br>DSKTGVDPDSAQRLIRELYSLLDFLRNDFSHNR<br>LDGTTFEHLEVSPDISSFITGTYSLACGRAQSRF<br>ADFFKPDDFVLAKNRKEQLISVADGKECLTVS<br>GLAFFICLFLDREQASGMLSRIRGFKRTDENWA<br>RAVHETFCDLCIRHPHDRLESSNTKEALLLDML<br>NELNRCPRILYDMLPEEERAQFLPALDENSMN<br>NLSENSLNEESRLLWDGSSDWAEALTKRIRHQ<br>DRFPYLMLRFIEEMDLLKGIRFRVDLGEIELDS |

[0521]

| | | |
|---|---|---|
| | | YSKKVGRNGEYDRTITDHALAFGKLSDFQNEE<br>EVSRMISGEASYPVRFSLFAPRYAIYDNKIGYC<br>HTSDPVYPKSKTGEKRALSNPQSMGFISVHDLR<br>KLLLMELLCEGSFSRMQSGFLRKANRILDETAE<br>GKLQFSALFPEMRHRFIPPQNPKSKDRREKAET<br>TLEKYKQEIKGRKDKLNSQLLSAFDMNQRQLP<br>SRLLDEWMNIRPASHSVKLRTYVKQLNEDCRL<br>RLRKFRKDGDGKARAIPLVGEMATFLSQDIVR<br>MIISEETKKLITSAYYNEMQRSLAQYAGEENRR<br>QFRAIVAELHLLDPSSGHPFLSATMETAHRYTE<br>DFYKCYLEKKREWLAKTFYRPEQDENTKRRIS<br>VFFVPDGEARKLLPLLIRRRMKEQNDLQDWIR<br>NKQAHPIDLPSHLFDSKIMELLKVKDGKKKWN<br>EAFKDWWSTKYPDGMQPFYGLRRELNIHGKS<br>VSYIPSDGKKFADCYTHLMEKTVQDKKRELRT<br>AGKPVPPDLAADIKRSFHRAVNEREFMLRLVQ<br>EDDRLMLMAINKMMTDREEDILPGLKNIDSIL<br>DEENQFSLAVHAKVLEKEGEGGDNSLSLVPAT<br>IEIKSKRKDWSKYIRYRYDRRVPGLMSHFPEHK<br>ATLDEVKTLLGEYDRCRIKIFDWAFALEGAIMS<br>DRDLKPYLHESSSREGKSGEHSTLVKMLVEKK<br>GCLTPDESQYLILIRNKAAHNQFPCAAEMPLIY<br>RDVSAKVGSIEGSSAKDLPEGSSLVDSLWKKY<br>EMIIRKILPILDHENRFFGKLLNNMSQPINDL |
| HMPREF1990_01800 [牙龈卟啉单胞菌W4087] (SEQ ID No. 101) | ERJ87335 | MNTVPASENKGQSRTVEDDPQYFGLYLNLARE<br>NLIEVESHVRIKFGKKKLNEESLKQSLLCDHLL<br>SVDRWTKVYGHSRRYLPFLHYFDPDSQIEKDH<br>DSKTGVDPDSAQRLIRELYSLLDFLRNDFSHNR<br>LDGTTFEHLEVSPDISSFITGTYSLACGRAQSRF<br>ADFFKPDDFVLAKNRKEQLISVADGKECLTVS<br>GLAFFICLFLDREQASGMLSRIRGFKRTDENWA<br>RAVHETFCDLCIRHPHDRLESSNTKEALLLDML<br>NELNRCPRILYDMLPEEERAQFLPALDENSMN<br>NLSENSLNEESRLLWDGSSDWAEALTKRIRHQ<br>DRFPYLMLRFIEEMDLLKGIRFRVDLGEIELDS<br>YSKKVGRNGEYDRTITDHALAFGKLSDFQNEE<br>EVSRMISGEASYPVRFSLFAPRYAIYDNKIGYC<br>HTSDPVYPKSKTGEKRALSNPRSMGFISVHDLR<br>KLLLMELLCEGSFSRMQSDFLRKANRILDETAE<br>GKLQFSALFPEMRHRFIPPQNPKSKDRREKAET<br>TLEKYKQEIKGRKDKLNSQLLSAFDMDQRQLP<br>SRLLDEWMNIRPASHSVKLRTYVKQLNEDCRL<br>RLQKFRKDGDGKARAIPLVGEMATFLSQDIVR |

[0522]

| | | |
|---|---|---|
| | | MIISEETKKLITSAYYNEMQRSLAQYAGEENRH<br>QFRAIVAELRLLDPSSGHPFLSATMETAHRYTE<br>DFYKCYLEKKREWLAKTFYRPEQDENTKRRIS<br>VFFVPDGEARKLLPLLIRRRMKEQNDLQDWIR<br>NKQAHPIDLPSHLFDSKVMELLKVKDGKKKW<br>NEAFKDWWSTKYPDGMQPFYGLRRELNIHGK<br>SVSYIPSDGKKFADCYTHLMEKTVRDKKRELR<br>TAGKPVPPDLAAYIKRSFHRAVNEREFMLRLV<br>QEDDRLMLMAINKIMTDREEDILPGLKNIDSIL<br>DKENQFSLAVHAKVLEKEGEGGDNSLSLVPAT<br>IEIKSKRKDWSKYIRYRYDRRVPGLMSHFPEHK<br>ATLDEVKTLLGEYDRCRIKIFDWAFALEGAIMS<br>DRDLKPYLHESSSREGKSGEHSTLVKMLVEKK<br>GCLTPDESQYLILIRNKAAHNQFPCAAEIPLIYR<br>DVSAKVGSIEGSSAKDLPEGSSLVDSLWKKYE<br>MIIRKILPILDPENRFFGKLLNNMSQPINDL |
| M573_117042 [中间普雷沃菌 ZT] (SEQ ID No. 102) | KJJ86756 | MKMEDDKKTTESTNMLDNKHFWAAFLNLAR<br>HNVYITVNHINKVLELKNKKDQDIIIDNDQDIL<br>AIKTHWEKVNGDLNKTERLRELMTKHFPFLET<br>AIYTKNKEDKEEVKQEKQAEAQSLESLKDCLF<br>LFLEKLQEARNYYSHYKYSESTKEPMLEEGLL<br>EKMYNIFDDNIQLVIKDYQHNKDINPDEDFKH<br>LDRKGQFKYSFADNEGNITESGLLFFVSLFLEK<br>KDAIWMQQKLTGFKDNRESKKKMTHEVFCRR<br>RMLLPKLRLESTQTQDWILLDMLNELIRCPKSL<br>YERLQGEYRKKFNVPFDSADEDYDAEQEPFKN<br>TLVRHQDRFPYFALRYFDYNEIFTNLRFQIDLG<br>TYHFSIYKKLIGGQKEDRHLTHKLYGFERIQEF<br>AKQNRPDEWKALVKDLDTYETSNERYISETTP<br>HYHLENQKIGIRFRNGNKEIWPSLKTNGENNE<br>KSKYKLDKPYQAEAFLSVHELLPMMFYYLLLK<br>KEEPNNDKKNASIVEGFIKREIRDMYKLYDAF<br>ANGEINNIGDLEKYCEDKGIPKRHLPKQMVAIL<br>YDEPKDMVKEAKRKQKEMVKDTKKLLATLE<br>KQTQEEIEDGGRNIRLLKSGEIARWLVNDMMR<br>FQPVQKDNEGNPLNNSKANSTEYQMLQRSLAL<br>YNKEEKPTRYFRQVNLINSSNPHPFLKWTKWE<br>ECNNILSFYRNYLTKKIEFLNKLKPEDWEKNQ<br>YFLKLKEPKTNRETLVQGWKNGFNLPRGIFTEP<br>IREWFKRHQNDSKEYEKVEALKRVGLVTKVIP<br>LFFKEEYFKEDAQKEINNCVQPFYSFPYNVGNI<br>HKPDEKDFLPSEERKKLWGDKKDKFKGYKAK<br>VKSKKLTDKEKEEYRSYLEFQSWNKFERELRL |

[0523]

| | | |
|---|---|---|
| | | VRNQDIVTWLLCTELIDKMKVEGLNVEELQKL<br>RLKDIDTDTAKQEKNNILNRIMPMQLPVTVYEI<br>DDSHNIVKDRPLHTVYIEETKTKLLKQGNFKA<br>LVKDRRLNGLFSFVDTSSKAELKDKPISKSVVE<br>YELGEYQNARIETIKDMLLLEKTLIKKYEKLPT<br>DNFSDMLNGWLEGKDESDKARFQNDVKLLVA<br>VRNAFSHNQYPMRNRIAFANINPFSLSSADISEE<br>KKLDIANQLKDKTHKIIKKIIEIEKPIETKE |
| A2033_10205 [拟杆菌门菌(Bacteroidetes bacterium) GWA2_31_9] (SEQ ID No. 103) | OFX18020.1 | MENQTQKGKGIYYYYTKNEDKHYFGSFLNLA<br>NNNIEQIIEEFRIRLSLKDEKNIKEIINNYFTDKK<br>SYTDWERGINILKEYLPVIDYLDLAITDKEFEKI<br>DLKQKETAKRKYFRTNFSLLIDTIIDLRNFYTH<br>YFHKPISINPDVAKFLDKNLLNVCLDIKKQKM<br>KTDKTKQALKDGLDKELKKLIELKKAELKEKK<br>IKTWNITENVEGAVYNDAFNHMVYKNNAGVT<br>ILKDYHKSILPDDKIDSELKLNFSISGLVFLLSM<br>FLSKKEIEQFKSNLEGFKGKVIGENGEYEISKFN<br>NSLKYMATHWIFSYLTFKGLKQRVKNTFDKET<br>LLMQMIDELNKVPHEVYQTLSKEQQNEFLEDI<br>NEYVQDNEENKKSMENSIVVHPVIRKRYDDKF<br>NYFAIRFLDEFANFPTLKFFVTAGNFVHDKREK<br>QIQGSMLTSDRMIKEKINVFGKLTEIAKYKSDY<br>FSNENTLETSEWELFPNPSYLLIQNNIPVHIDLIH<br>NTEEAKQCQIAIDRIKCTTNPAKKRNTRKSKEE<br>IIKIIYQKNKNIKYGDPTALLSSNELPALIYELLV<br>NKKSGKELENIIVEKIVNQYKTIAGFEKGQNLS<br>NSLITKKLKKSEPNEDKINAEKIILAINRELEITE<br>NKLNIIKNNRAEFRTGAKRKHIFYSKELGQEAT<br>WIAYDLKRFMPEASRKEWKGFHHSELQKFLAF<br>YDRNKNDAKALLNMFWNFDNDQLIGNDLNSA<br>FREFHFDKFYEKYLIKRDEILEGFKSFISNFKDE<br>PKLLKKGIKDIYRVFDKRYYIIKSTNAQKEQLL<br>SKPICLPRGIFDNKPTYIEGVKVESNSALFADW<br>YQYTYSDKHEFQSFYDMPRDYKEQFEKFELNN<br>IKSIQNKKNLNKSDKFIYFRYKQDLKIKQIKSQ<br>DLFIKLMVDELFNVVFKNNIELNLKKLYQTSDE<br>RFKNQLIADVQKNREKGDTSDNKMNENFIWN<br>MTIPLSLCNGQIEEPKVKLKDIGKFRKLETDDK<br>VIQLLEYDKSKVWKKLEIEDELENMPNSYERIR<br>REKLLKGIQEFEHFLLEKEKFDGINHPKHFEQD<br>LNPNFKTYVINGVLRKNSKLNYTEIDKLLDLEH<br>ISIKDIETSAKEIHLAYFLIHVRNKFGHNQLPKL<br>EAFELMKKYYKKNNEETYAEYFHKVSSQIVNE |

[0524]

| | | |
|---|---|---|
| | | FKNSLEKHS |
| SAMN05421542_0666 [济州岛金黄杆菌(Chryseobacterium jejuense)] (SEQ ID No. 104) | SDI27289.1 | MEKTQTGLGIYYDHTKLQDKYFFGGFFNLAQN<br>NIDNVIKAFIIKFFPERKDKDINIAQFLDICFKDN<br>DADSDFQKKNKFLRIHFPVIGFLTSDNDKAGFK<br>KKFALLLKTISELRNFYTHYYHKSIEFPSELFEL<br>LDDIFVKTTSEIKKLKKKDDKTQQLLNKNLSEE<br>YDIRYQQQIERLKELKAQGKRVSLTDETAIRNG<br>VFNAAFNHLIYRDGENVKPSRLYQSSYSEPDPA<br>ENGISLSQNSILFLLSMFLERKETEDLKSRVKGF<br>KAKIIKQGEEQISGLKFMATHWVFSYLCFKGIK<br>QKLSTEFHEETLLIQIIDELSKVPDEVYSAFDSK<br>TKEKFLEDINEYMKEGNADLSLEDSKVIHPVIR<br>KRYENKFNYFAIRFLDEYLSSTSLKFQVHVGN<br>YVHDRRVKHINGTGFQTERIVKDRIKVFGRLSN<br>ISNLKADYIKEQLELPNDSNGWEIFPNPSYIFID<br>NNVPIHVLADEATKKGIELFKDKRRKEQPEELQ<br>KRKGKISKYNIVSMIYKEAKGKDKLRIDEPLAL<br>LSLNEIPALLYQILEKGATPKDIELIIKNKLTERF<br>EKIKNYDPETPAPASQISKRLRNNTTAKGQEAL<br>NAEKLSLLIEREIENTETKLSSIEEKRLKAKKEQ<br>RRNTPQRSIFSNSDLGRIAAWLADDIKRFMPAE<br>QRKNWKGYQHSQLQQSLAYFEKRPQEAFLLL<br>KEGWDTSDGSSYWNNWVMNSFLENNHFEKF<br>YKNYLMKRVKYFSELAGNIKQHTHNTKFLRKF<br>IKQQMPADLFPKRHYILKDLETEKNKVLSKPLV<br>FSRGLFDNNPTFIKGVKVTENPELFAEWYSYG<br>YKTEHVFQHFYGWERDYNELLDSELQKGNSF<br>AKNSIYYNRESQLDLIKLKQDLKIKKIKIQDLFL<br>KRIAEKLFENVFNYPTTLSLDEFYLTQEERAEK<br>ERIALAQSLREEGDNSPNIIKDDFIWSKTIAFRS<br>KQIYEPAIKLKDIGKFNRFVLDDEESKASKLLS<br>YDKNKIWNKEQLERELSIGENSYEVIRREKLFK<br>EIQNLELQILSNWSWDGINHPREFEMEDQKNT<br>RHPNFKMYLVNGILRKNINLYKEDEDFWLESL<br>KENDFKTLPSEVLETKSEMVQLLFLVILIRNQF<br>AHNQLPEIQFYNFIRKNYPEIQNNTVAELYLNLI<br>KLAVQKLKDNS |
| SAMN05444360_11366 [卡尼氏金 | SHM52812.1 | MNTRVTGMGVSYDHTKKEDKHFFGGFLNLAQ<br>DNITAVIKAFCIKFDKNPMSSVQFAESCFTDKD<br>SDTDFQNKVRYVRTHLPVIGYLNYGGDRNTFR<br>QKLSTLLKAVDSLRNFYTHYYHSPLALSTELFE<br>LLDTVFASVAVEVKQHKMKDDKTRQLLSKSL<br>AEELDIRYKQQLERLKELKEQGKNIDLRDEAGI |

[0525]

| | | |
|---|---|---|
| 黄杆菌(Chryseobacterium carnipullorum)] (SEQ ID No. 105) | | RNGVLNAAFNHLIYKEGEIAKPTLSYSSFYYGA<br>DSAENGITISQSGLLFLLSMFLGKKEIEDLKSRI<br>RGFKAKIVRDGEENISGLKFMATHWIFSYLSFK<br>GMKQRLSTDFHEETLLIQIIDELSKVPDEVYHD<br>FDTATREKFVEDINEYIREGNEDFSLGDSTIIHP<br>VIRKRYENKFNYFAVRFLDEFIKFPSLRFQVHL<br>GNFVHDRRIKDIHGTGFQTERVVKDRIKVFGK<br>LSEISSLKTEYIEKELDLDSDTGWEIFPNPSYVFI<br>DNNIPIYISTNKTFKNGSSEFIKLRRKEKPEEMK<br>MRGEDKKEKRDIASMIGNAGSLNSKTPLAMLS<br>LNEMPALLYEILVKKTTPEEIELIIKEKLDSHFE<br>NIKNYDPEKPLPASQISKRLRNNTTDKGKKVIN<br>PEKLIHLINKEIDATEAKFALLAKNRKELKEKF<br>RGKPLRQTIFSNMELGREATWLADDIKRFMPDI<br>LRKNWKGYQHNQLQQSLAFFNSRPKEAFTILQ<br>DGWDFADGSSFWNGWIINSFVKNRSFEYFYEA<br>YFEGRKEYFSSLAENIKQHTSNHRNLRRFIDQQ<br>MPKGLFENRHYLLENLETEKNKILSKPLVFPRG<br>LFDTKPTFIKGIKVDEQPELFAEWYQYGYSTEH<br>VFQNFYGWERDYNDLLESELEKDNDFSKNSIH<br>YSRTSQLELIKLKQDLKIKKIKIQDLFLKLIAGHI<br>FENIFKYPASFSLDELYLTQEERLNKEQEALIQS<br>QRKEGDHSDNIIKDNFIGSKTVTYESKQISEPNV<br>KLKDIGKFNRFLLDDKVKTLLSYNEDKVWNK<br>NDLDLELSIGENSYEVIRREKLFKKIQNFELQTL<br>TDWPWNGTDHPEEFGTTDNKGVNHPNFKMY<br>VVNGILRKHTDWFKEGEDNWLENLNETHFKN<br>LSFQELETKSKSIQTAFLIIMIRNQFAHNQLPAV<br>QFFEFIQKKYPEIQGSTTSELYLNFINLAVVELL<br>ELLEK |
| SAMN05421786_1011119 [Chryseobacterium ureilyticum] (SEQ ID No. 106) | SIS70481.1 | METQILGNGISYDHTKTEDKHFFGGFLNTAQN<br>NIDLLIKAYISKFESSPRKLNSVQFPDVCFKKND<br>SDADFQHKLQFIRKHLPVIQYLKYGGNREVLK<br>EKFRLLLQAVDSLRNFYTHFYHKPIQLPNELLT<br>LLDTIFGEIGNEVRQNKMKDDKTRHLLKKNLS<br>EELDFRYQEQLERLRKLKSEGKKVDLRDTEAI<br>RNGVLNAAFNHLIFKDAEDFKPTVSYSSYYYD<br>SDTAENGISISQSGLLFLLSMFLGRREMEDLKS<br>RVRGFKARIIKHEEQHVSGLKFMATHWVFSEF<br>CFKGIKTRLNADYHEETLLIQLIDELSKVPDELY<br>RSFDVATRERFIEDINEYIRDGKEDKSLIESKIV<br>HPVIRKRYESKFNYFAIRFLDEFVNFPTLRFQV<br>HAGNYVHDRRIKSIEGTGFKTERLVKDRIKVFG |

[0526]

| | | |
|---|---|---|
| | | KLSTISSLKAEYLAKAVNITDDTGWELLPHPSY<br>VFIDNNIPIHLTVDPSFKNGVKEYQEKRKLQKP<br>EEMKNRQGGDKMHKPAISSKIGKSKDINPESPV<br>ALLSMNEIPALLYEILVKKASPEEVEAKIRQKL<br>TAVFERIRDYDPKVPLPASQVSKRLRNNTDTLS<br>YNKEKLVELANKEVEQTERKLALITKNRRECR<br>EKVKGKFKRQKVFKNAELGTEATWLANDIKR<br>FMPEEQKKNWKGYQHSQLQQSLAFFESRPGEA<br>RSLLQAGWDFSDGSSFWNGWVMNSFARDNTF<br>DGFYESYLNGRMKYFLRLADNIAQQSSTNKLIS<br>NFIKQQMPKGLFDRRLYMLEDLATEKNKILSK<br>PLIFPRGIFDDKPTFKKGVQVSEEPEAFADWYS<br>YGYDVKHKFQEFYAWDRDYEELLREELEKDT<br>AFTKNSIHYSRESQIELLAKKQDLKVKKVRIQD<br>LYLKLMAEFLFENVFGHELALPLDQFYLTQEE<br>RLKQEQEAIVQSQRPKGDDSPNIVKENFIWSKT<br>IPFKSGRVFEPNVKLKDIGKFRNLLTDEKVDILL<br>SYNNTEIGKQVIENELIIGAGSYEFIRREQLFKEI<br>QQMKRLSLRSVRGMGVPIRLNLK |
| | 颊普雷沃菌(SEQ ID No. 107) | WP_<br>0043<br>4358<br>1 | MQKQDKLFVDRKKNAIFAFPKYITIMENQEKP<br>EPIYYELTDKHFWAAFLNLARHNVYTTINHINR<br>RLEIAELKDDGYMMDIKGSWNEQAKKLDKKV<br>RLRDLIMKHFPFLEAAAYEITNSKSPNNKEQRE<br>KEQSEALSLNNLKNVLFIFLEKLQVLRNYYSHY<br>KYSEESPKPIFETSLLKNMYKVFDANVRLVKR<br>DYMHHENIDMQRDFTHLNRKKQVGRTKNIIDS<br>PNFHYHFADKEGNMTIAGLLFFVSLFLDKKDAI<br>WMQKKLKGFKDGRNLREQMTNEVFCRSRISL<br>PKLKLENVQTKDWMQLDMLNELVRCPKSLYE<br>RLREKDRESFKVPFDIFSDDYDAEEEPFKNTLV<br>RHQDRFPYFVLRYFDLNEIFEQLRFQIDLGTYH<br>FSIYNKRIGDEDEVRHLTHHLYGFARIQDFAQQ<br>NQPEVWRKLVKDLDYFEASQEPYIPKTAPHYH<br>LENEKIGIKFCSTHNNLFPSLKTEKTCNGRSKFN<br>LGTQFTAEAFLSVHELLPMMFYYLLLTKDYSR<br>KESADKVEGIIRKEISNIYAIYDAFANGEINSIAD<br>LTCRLQKTNILQGHLPKQMISILEGRQKDMEKE<br>AERKIGEMIDDTQRRLDLLCKQTNQKIRIGKRN<br>AGLLKSGKIADWLVNDMMRFQPVQKDQNNIPI<br>NNSKANSTEYRMLQRALALFGSENFRLKAYFN<br>QMNLVGNDNPHPFLAETQWEHQTNILSFYRNY<br>LEARKKYLKGLKPQNWKQYQHFLILKVQKTN<br>RNTLVTGWKNSFNLPRGIFTQPIREWFEKHNNS |

| | | |
|---|---|---|
| | | KRIYDQILSFDRVGFVAKAIPLYFAEEYKDNVQ<br>PFYDYPFNIGNKLKPQKGQFLDKKERVELWQK<br>NKELFKNYPSEKKKTDLAYLDFLSWKKFEREL<br>RLIKNQDIVTWLMFKELFNMATVEGLKIGEIHL<br>RDIDTNTANEESNNILNRIMPMKLPVKTYETDN<br>KGNILKERPLATFYIEETETKVLKQGNFKVLAK<br>DRRLNGLLSFAETTDIDLEKNPITKLSVDHELIK<br>YQTTRISIFEMTLGLEKKLINKYPTLPTDSFRNM<br>LERWLQCKANRPELKNYVNSLIAVRNAFSHNQ<br>YPMYDATLFAEVKKFTLFPSVDTKKIELNIAPQ<br>LLEIVGKAIKEIEKSENKN |
| [0527] | 牙龈卟啉单胞菌 (SEQ ID No. 108) | WP_005873511 | MNTVPASENKGQSRTVEDDPQYFGLYLNLARE<br>NLIEVESHVRIKFGKKKLNEESLKQSLLCDHLL<br>SVDRWTKVYGHSRRYLPFLHYFDPDSQIEKDH<br>DSKTGVDPDSAQRLIRELYSLLDFLRNDFSHNR<br>LDGTTFEHLEVSPDISSFITGTYSLACGRAQSRF<br>ADFFKPDDFVLAKNRKEQLISVADGKECLTVS<br>GLAFFICLFLDREQASGMLSRIRGFKRTDENWA<br>RAVHETFCDLCIRHPHDRLESSNTKEALLLDML<br>NELNRCPRILYDMLPEEERAQFLPALDENSMN<br>NLSENSLNEESRLLWDGSSDWAEALTKRIRHQ<br>DRFPYLMLRFIEEMDLLKGIRFRVDLGEIELDS<br>YSKKVGRNGEYDRTITDHALAFGKLSDFQNEE<br>EVSRMISGEASYPVRFSLFAPRYAIYDNKIGYC<br>HTSDPVYPKSKTGEKRALSNPQSMGFISVHNLR<br>KLLLMELLCEGSFSRMQSDFLRKANRILDETAE<br>GKLQFSALFPEMRHRFIPPQNPKSKDRREKAET<br>TLEKYKQEIKGRKDKLNSQLLSAFDMNQRQLP<br>SRLLDEWMNIRPASHSVKLRTYVKQLNEDCRL<br>RLRKFRKDGDGKARAIPLVGEMATFLSQDIVR<br>MIISEETKKLITSAYYNEMQRSLAQYAGEENRR<br>QFRAIVAELHLLDPSSGHPFLSATMETAHRYTE<br>DFYKCYLEKKREWLAKTFYRPEQDENTKRRIS<br>VFFVPDGEARKLLPLLIRRRMKEQNDLQDWIR<br>NKQAHPIDLPSHLFDSKIMELLKVKDGKKKWN<br>EAFKDWWSTKYPDGMQPFYGLRRELNIHGKS<br>VSYIPSDGKKFADCYTHLMEKTVQDKKRELRT<br>AGKPVPPDLAADIKRSFHRAVNEREFMLRLVQ<br>EDDRLMLMAINKMMTDREEDILPGLKNIDSIL<br>DEENQFSLAVHAKVLEKEGEGGDNSLSLVPAT<br>IEIKSKRKDWSKYIRYRYDRRVPGLMSHFPEHK<br>ATLDEVKTLLGEYDRCRIKIFDWAFALEGAIMS<br>DRDLKPYLHESSSREGKSGEHSTLVKMLVEKK |

[0528]

| | | |
|---|---|---|
| | | GCLTPDESQYLILIRNKAAHNQFPCAAEMPLIY<br>RDVSAKVGSIEGSSAKDLPEGSSLVDSLWKKY<br>EMIIRKILPILDPENRFFGKLLNNMSQPINDL |
| 牙龈卟啉单胞菌 (SEQ ID No. 109) | WP_005874195 | MTEQNEKPYNGTYYTLEDKHFWAAFFNLARH<br>NAYITLAHIDRQLAYSKADITNDEDILFFKGQW<br>KNLDNDLERKARLRSLILKHFSFLEGAAYGKK<br>LFESQSSGNKSSKKKELTKKEKEELQANALSLD<br>NLKSILFDFLQKLKDFRNYYSHYRHPESSELPL<br>FDGNMLQRLYNVFDVSVQRVKRDHEHNDKV<br>DPHRHFNHLVRKGKKDKYGNNDNPFFKHHFV<br>DREEKVTEAGLLFFVSLFLEKRDAIWMQKKIR<br>GFKGGTEAYQQMTNEVFCRSRISLPKLKLESLR<br>TDDWMLLDMLNELVRCPKSLYDRLREEDRAR<br>FRVPVDILSDEDDTDGTEEDPFKNTLVRHQDRF<br>PYFALRYFDLKKVFTSLRFHIDLGTYHFAIYKK<br>NIGEQPEDRHLTRNLYGFGRIQDFAEEHRPEEW<br>KRLVRDLDYFETGDKPYITQTTPHYHIEKGKIG<br>LRFVPEGQLLWPSPEVGATRTGRSKYAQDKRF<br>TAEAFLSVHELMPMMFYYFLLREKYSEEASAE<br>KVQGRIKRVIEDVYAVYDAFARDEINTRDELD<br>ACLADKGIRRGHLPRQMIAILSQEHKDMEEKV<br>RKKLQEMIADTDHRLDMLDRQTDRKIRIGRKN<br>AGLPKSGVIADWLVRDMMRFQPVAKDTSGKP<br>LNNSKANSTEYRMLQRALALFGGEKERLTPYF<br>RQMNLTGGNNPHPFLHETRWESHTNILSFYRS<br>YLKARKAFLQSIGRSDREENHRFLLLKEPKTDR<br>QTLVAGWKSEFHLPRGIFTEAVRDCLIEMGYD<br>EVGSYKEVGFMAKAVPLYFERACKDRVQPFY<br>DYPFNVGNSLKPKKGRFLSKEKRAEEWESGKE<br>RFRDLEAWSHSAARRIEDAFVGIEYASWENKK<br>KIEQLLQDLSLWETFESKLKVKADKINIAKLKK<br>EILEAKEHPYHDFKSWQKFERELRLVKNQDIIT<br>WMMCRDLMEENKVEGLDTGTLYLKDIRTDVQ<br>EQGSLNVLNHVKPMRLPVVVYRADSRGHVHK<br>EEAPLATVYIEERDTKLLKQGNFKSFVKDRRL<br>NGLFSFVDTGALAMEQYPISKLRVEYELAKYQ<br>TARVCAFEQTLELEESLLTRYPHLPDESFREML<br>ESWSDPLLDKWPDLQREVRLLIAVRNAFSHNQ<br>YPMYDETIFSSIRKYDPSSLDAIEERMGLNIAHR<br>LSEEVKLAKEMVERIIQA |
| 苍白普雷沃菌 | WP_006044483 | MKEEEKGKTPVVSTYNKDDKHFWAAFLNLAR<br>HNVYITVNHINKILGEGEINRDGYENTLEKSWN<br>EIKDINKKDRLSKLIIKHFPFLEVTTYQRNSADT |

[0529]

| | | |
|---|---|---|
| (SEQ ID No. 110) | 3 | TKQKEEKQAEAQSLESLKKSFFVFIYKLRDLRN<br>HYSHYKHSKSLERPKFEEDLQEKMYNIFDASIQ<br>LVKEDYKHNTDIKTEEDFKHLDRKGQFKYSFA<br>DNEGNITESGLLFFVSLFLEKKDAIWVQKKLEG<br>FKCSNESYQKMTNEVFCRSRMLLPKLRLQSTQ<br>TQDWILLDMLNELIRCPKSLYERLREEDRKKFR<br>VPIEIADEDYDAEQEPFKNALVRHQDRFPYFAL<br>RYFDYNEIFTNLRFQIDLGTYHFSIYKKQIGDY<br>KESHHLTHKLYGFERIQEFTKQNRPDEWRKFV<br>KTFNSFETSKEPYIPETTPHYHLENQKIGIRFRN<br>DNDKIWPSLKTNSEKNEKSKYKLDKSFQAEAF<br>LSVHELLPMMFYYLLLKTENTDNDNEIETKKK<br>ENKNDKQEKHKIEEIIENKITEIYALYDAFANG<br>KINSIDKLEEYCKGKDIEIGHLPKQMIAILKSEH<br>KDMATEAKRKQEEMLADVQKSLESLDNQINE<br>EIENVERKNSSLKSGEIASWLVNDMMRFQPVQ<br>KDNEGNPLNNSKANSTEYQMLQRSLALYNKE<br>EKPTRYFRQVNLIESSNPHPFLNNTEWEKCNNI<br>LSFYRSYLEAKKNFLESLKPEDWEKNQYFLML<br>KEPKTNCETLVQGWKNGFNLPRGIFTEPIRKWF<br>MEHRKNITVAELKRVGLVAKVIPLFFSEEYKDS<br>VQPFYNYLFNVGNINKPDEKNFLNCEERRELL<br>RKKKDEFKKMTDKEKEENPSYLEFQSWNKFE<br>RELRLVRNQDIVTWLLCMELFNKKKIKELNVE<br>KIYLKNINTNTTKKEKNTEEKNGEEKIIKEKNNI<br>LNRIMPMRLPIKVYGRENFSKNKKKKIRRNTFF<br>TVYIEEKGTKLLKQGNFKALERDRRLGGLFSF<br>VKTHSKAESKSNTISKSRVEYELGEYQKARIEII<br>KDMLALEETLIDKYNSLDTDNFHNMLTGWLK<br>LKDEPDKASFQNDVDLLIAVRNAFSHNQYPMR<br>NRIAFANINPFSLSSANTSEEKGLGIANQLKDKT<br>HKTIEKIIEIEKPIETKE |
| 类香味菌 (SEQ ID No. 111) | WP_006261414 | MKDILTTDTTEKQNRFYSHKIADKYFFGGYFN<br>LASNNIYEVFEEVNKRNTFGKLAKRDNGNLKN<br>YIIHVFKDELSISDFEKRVAIFASYFPILETVDKK<br>SIKERNRTIDLTLSQRIRQFREMLISLVTAVDQL<br>RNFYTHYHHSEIVIENKVLDFLNSSLVSTALHV<br>KDKYLKTDKTKEFLKETIAAELDILIEAYKKKQ<br>IEKKNTRFKANKREDILNAIYNEAFWSFINDKD<br>KDKETVVAKGADAYFEKNHHKSNDPDFALNIS<br>EKGIVYLLSFFLTNKEMDSLKANLTGFKGKVD<br>RESGNSIKYMATQRIYSFHTYRGLKQKIRTSEE<br>GVKETLLMQMIDELSKVPNVVYQHLSTTQQNS |

[0530]

| | | |
|---|---|---|
| | | FIEDWNEYYKDYEDDVETDDLSRVIHPVIRKR<br>YEDRFNYFAIRFLDEFFDFPTLRFQVHLGDYVH<br>DRRTKQLGKVESDRIIKEKVTVFARLKDINSAK<br>ANYFHSLEEQDKEELDNKWTLFPNPSYDFPKE<br>HTLQHQGEQKNAGKIGIYVKLRDTQYKEKAA<br>LEEARKSLNPKERSATKASKYDIITQIIEANDNV<br>KSEKPLVFTGQPIAYLSMNDIHSMLFSLLTDNA<br>ELKKTPEEVEAKLIDQIGKQINEILSKDTDTKIL<br>KKYKDNDLKETDTDKITRDLARDKEEIEKLILE<br>QKQRADDYNYTSSTKFNIDKSRKRKHLLFNAE<br>KGKIGVWLANDIKRFMTEEFKSKWKGYQHTE<br>LQKLFAYYDTSKSDLDLILSDMVMVKDYPIELI<br>ALVKKSRTLVDFLNKYLEARLGYMENVITRVK<br>NSIGTPQFKTVRKECFTFLKKSNYTVVSLDKQV<br>ERILSMPLFIERGFMDDKPTMLEGKSYQQHKE<br>KFADWFVHYKENSNYQNFYDTEVYEITTEDKR<br>EKAKVTKKIKQQQKNDVFTLMMVNYMLEEVL<br>KLSSNDRLSLNELYQTKEERIVNKQVAKDTQE<br>RNKNYIWNKVVDLQLCEGLVRIDKVKLKDIGN<br>FRKYENDSRVKEFLTYQSDIVWSAYLSNEVDS<br>NKLYVIERQLDNYESIRSKELLKEVQEIECSVY<br>NQVANKESLKQSGNENFKQYVLQGLVPIGMD<br>VREMLILSTDVKFIKEEIIQLGQAGEVEQDLYSL<br>IYIRNKFAHNQLPIKEFFDFCENNYRSISDNEYY<br>AEYYMEIFRSIKEKYTS |
| 类香味菌 (SEQ ID No. 112) | WP_006265509 | MKDILTTDTTEKQNRFYSHKIADKYFFGGYFN<br>LASNNIYEVFEEVNKRNTFGKLAKRDNGNLKN<br>YIIHVFKDELSISDFEKRVAIFASYFPILETVDKK<br>SIKERNRTIDLTLSQRIRQFREMLISLVTAVDQL<br>RNFYTHYHHSEIVIENKVLDFLNSSLVSTALHV<br>KDKYLKTDKTKEFLKETIAAELDILIEAYKKKQ<br>IEKKNTRFKANKREDILNAIYNEAFWSFINDKD<br>KDKETVVAKGADAYFEKNHHKSNDPDFALNIS<br>EKGIVYLLSFFLTNKEMDSLKANLTGFKGKVD<br>RESGNSIKYMATQRIYSFHTYRGLKQKIRTSEE<br>GVKETLLMQMIDELSKVPNVVYQHLSTTQQNS<br>FIEDWNEYYKDYEDDVETDDLSRVIHPVIRKR<br>YEDRFNYFAIRFLDEFFDFPTLRFQVHLGDYVH<br>DRRTKQLGKVESDRIIKEKVTVFARLKDINSAK<br>ASYFHSLEEQDKEELDNKWTLFPNPSYDFPKE<br>HTLQHQGEQKNAGKIGIYVKLRDTQYKEKAA<br>LEEARKSLNPKERSATKASKYDIITQIIEANDNV<br>KSEKPLVFTGQPIAYLSMNDIHSMLFSLLTDNA |

| | | | |
|---|---|---|---|
| | | | ELKKTPEEVEAKLIDQIGKQINEILSKDTDTKIL<br>KKYKDNDLKETDTDKITRDLARDKEEIEKLILE<br>QKQRADDYNYTSSTKFNIDKSRKRKHLLFNAE<br>KGKIGVWLANDIKRFMFKESKSKWKGYQHTE<br>LQKLFAYFDTSKSDLELILSDMVMVKDYPIELI<br>DLVRKSRTLVDFLNKYLEARLGYIENVITRVK<br>NSIGTPQFKTVRKECFAFLKESNYTVASLDKQI<br>ERILSMPLFIERGFMDSKPTMLEGKSYQQHKED<br>FADWFVHYKENSNYQNFYDTEVYEIITEDKRE<br>QAKVTKKIKQQQKNDVFTLMMVNYMLEEVL<br>KLPSNDRLSLNELYQTKEERIVNKQVAKDTQE<br>RNKNYIWNKVVDLQLCEGLVRIDKVKLKDIGN<br>FRKYENDSRVKEFLTYQSDIVWSGYLSNEVDS<br>NKLYVIERQLDNYESIRSKELLKEVQEIECIVYN<br>QVANKESLKQSGNENFKQYVLQGLLPRGTDV<br>REMLILSTDVKFKKEEIMQLGQVREVEQDLYS<br>LIYIRNKFAHNQLPIKEFFDFCENNYRPISDNEY<br>YAEYYMEIFRSIKEKYAS |
| [0531] | 普雷沃菌属某种 MSX73 (SEQ ID No. 113) | WP_007412163 | MQKQDKLFVDRKKNAIFAFPKYITIMENQEKP<br>EPIYYELTDKHFWAAFLNLARHNVYTTINHINR<br>RLEIAELKDDGYMMGIKGSWNEQAKKLDKKV<br>RLRDLIMKHFPFLEAAAYEITNSKSPNNKEQRE<br>KEQSEALSLNNLKNVLFIFLEKLQVLRNYYSHY<br>KYSEESPKPIFETSLLKNMYKVFDANVRLVKR<br>DYMHHENIDMQRDFTHLNRKKQVGRTKNIIDS<br>PNFHYHFADKEGNMTIAGLLFFVSLFLDKKDAI<br>WMQKKLKGFKDGRNLREQMTNEVFCRSRISL<br>PKLKLENVQTKDWMQLDMLNELVRCPKSLYE<br>RLREKDRESFKVPFDIFSDDYDAEEEPFKNTLV<br>RHQDRFPYFVLRYFDLNEIFEQLRFQIDLGTYH<br>FSIYNKRIGDEDEVRHLTHHLYGFARIQDFAPQ<br>NQPEEWRKLVKDLDHFETSQEPYISKTAPHYH<br>LENEKIGIKFCSTHNNLFPSLKREKTCNGRSKF<br>NLGTQFTAEAFLSVHELLPMMFYYLLLTKDYS<br>RKESADKVEGIIRKEISNIYAIYDAFANNEINSIA<br>DLTCRLQKTNILQGHLPKQMISILEGRQKDME<br>KEAERKIGEMIDDTQRRLDLLCKQTNQKIRIGK<br>RNAGLLKSGKIADWLVSDMMRFQPVQKDTNN<br>APINNSKANSTEYRMLQHALALFGSESSRLKA<br>YFRQMNLVGNANPHPFLAETQWEHQTNILSFY<br>RNYLEARKKYLKGLKPQNWKQYQHFLILKVQ<br>KTNRNTLVTGWKNSFNLPRGIFTQPIREWFEKH<br>NNSKRIYDQILSFDRVGFVAKAIPLYFAEEYKD |

[0532]

| | | |
|---|---|---|
| | | NVQPFYDYPFNIGNKLKPQKGQFLDKKERVEL<br>WQKNKELFKNYPSEKNKTDLAYLDFLSWKKF<br>ERELRLIKNQDIVTWLMFKELFKTTTVEGLKIG<br>EIHLRDIDTNTANEESNNILNRIMPMKLPVKTY<br>ETDNKGNILKERPLATFYIEETETKVLKQGNFK<br>VLAKDRRLNGLLSFAETTDIDLEKNPITKLSVD<br>YELIKYQTTRISIFEMTLGLEKKLIDKYSTLPTD<br>SFRNMLERWLQCKANRPELKNYVNSLIAVRN<br>AFSHNQYPMYDATLFAEVKKFTLFPSVDTKKI<br>ELNIAPQLLEIVGKAIKEIEKSENKN |
| 牙龈卟啉单胞菌 (SEQ ID No. 114) | WP_012458414 | MTEQNERPYNGTYYTLEDKHFWAAFFNLARH<br>NAYITLAHIDRQLAYSKADITNDEDILFFKGQW<br>KNLDNDLERKARLRSLILKHFSFLEGAAYGKK<br>LFESQSSGNKSSKKKELTKKEKEELQANALSLD<br>NLKSILFDFLQKLKDFRNYYSHYRHPESSELPL<br>FDGNMLQRLYNVFDVSVQRVKRDHEHNDKV<br>DPHRHFNHLVRKGKKDRYGNNDNPFFKHHFV<br>DREEKVTEAGLLFFVSLFLEKRDAIWMQKKIR<br>GFKGGTETYQQMTNEVFCRSRISLPKLKLESLR<br>TDDWMLLDMLNELVRCPKSLYDRLREEDRAR<br>FRVPVDILSDEDDTDGTEEDPFKNTLVRHQDRF<br>PYFALRYFDLKKVFTSLRFHIDLGTYHFAIYKK<br>NIGEQPEDRHLTRNLYGFGRIQDFAEEHRPEEW<br>KRLVRDLDYFETGDKPYITQTTPHYHIEKGKIG<br>LRFVPEGQHLWPSPEVGATRTGRSKYAQDKRL<br>TAEAFLSVHELMPMMFYYFLLREKYSDEASAE<br>RVQGRIKRVIEDVYAVYDAFARGEINTRDELD<br>ACLADKGIRRGHLPRQMIGILSQEHKDMEEKV<br>RKKLQEMIVDTDHRLDMLDRQTDRKIRIGRKN<br>AGLPKSGVIADWLVRDMMRFQPVAKDTSGKP<br>LNNSKANSTEYRMLQRALALFGGEKERLTPYF<br>RQMNLTGGNNPHPFLHETRWESHTNILSFYRS<br>YLKARKAFLQSIGRSDRVENHRFLLLKEPKTDR<br>QTLVAGWKGEFHLPRGIFTEAVRDCLIEMGLD<br>EVGSYKEVGFMAKAVPLYFERACKDRVQPFY<br>DYPFNVGNSLKPKKGRFLSKEKRAEEWESGKE<br>RFRLAKLKKEILEAKEHPYLDFKSWQKFEREL<br>RLVKNQDIITWMICRDLMEENKVEGLDTGTLY<br>LKDIRTDVQEQGNLNVLNRVKPMRLPVVVYR<br>ADSRGHVHKEQAPLATVYIEERDTKLLKQGNF<br>KSFVKDRRLNGLFSFVDTGALAMEQYPISKLR<br>VEYELAKYQTARVCAFEQTLELEESLLTRYPH<br>LPDKNFRKMLESWSDPLLDKWPDLHGNVRLLI |

[0533]

| | | |
|---|---|---|
| | | AVRNAFSHNQYPMYDEAVFSSIRKYDPSSPDAI EERMGLNIAHRLSEEVKQAKEMAERIIQA |
| 产丙酸栖沼泽杆菌(SEQ ID No. 115) | WP_013446107 | MKTSANNIYFNGINSFKKIFDSKGAIAPIAEKSC RNFDIKAQNDVNKEQRIHYFAVGHTFKQLDTE NLFEYVLDENLRAKRPTRFISLQQFDKEFIENIK RLISDIRNINSHYIHRFDPLKIDAVPTNIIDFLKES FELAVIQIYLKEKGINYLQFSENPHADQKLVAF LHDKFLPLDEKKTSMLQNETPQLKEYKEYRKY FKTLSKQAAIDQLLFAEKETDYIWNLFDSHPVL TISAGKYLSFYSCLFLLSMFLYKSEANQLISKIK GFKKNTTEEEKSKREIFTFFSKRFNSMDIDSEEN QLVKFRDLILYLNHYPVAWNKDLELDSSNPAM TDKLKSKIIELEINRSFPLYEGNERFATFAKYQI WGKKHLGKSIEKEYINASFTDEEITAYTYETDT CPELKDAHKKLADLKAAKGLFGKRKEKNESDI KKTETSIRELQHEPNPIKDKLIQRIEKNLLTVSY GRNQDRFMDFSARFLAEINYFGQDASFKMYHF YATDEQNSELEKYELPKDKKKYDSLKFHQGKL VHFISYKEHLKRYESWDDAFVIENNAIQLKLSF DGVENTVTIQRALLIYLLEDALRNIQNNTAENA GKQLLQEYYSHNKADLSAFKQILTQQDSIEPQQ KTEFKKLLPRRLLNNYSPAINHLQTPHSSLPLIL EKALLAEKRYCSLVVKAKAEGNYDDFIKRNK GKQFKLQFIRKAWNLMYFRNSYLQNVQAAGH HKSFHIERDEFNDFSRYMFAFEELSQYKYYLNE MFEKKGFFENNEFKILFQSGTSLENLYEKTKQK FEIWLASNTAKTNKPDNYHLNNYEQQFSNQLF FINLSHFINYLKSTGKLQTDANGQIIYEALNNV QYLIPEYYYTDKPERSESKSGNKLYNKLKATK LEDALLYEMAMCYLKADKQIADKAKHPITKLL TSDVEFNITNKEGIQLYHLLVPFKKIDAFIGLKM HKEQQDKKHPTSFLANIVNYLELVKNDKDIRK TYEAFSTNPVKRTLTYDDLAKIDGHLISKSIKFT NVTLELERYFIFKESLIVKKGNNIDFKYIKGLRN YYNNEKKKNEGIRNKAFHFGIPDSKSYDQLIRD AEVMFIANEVKPTHATKYTDLNKQLHTVCDK LMETVHNDYFSKEGDGKKKREAAGQKYFENII SAK |
| 牙龈卟啉单胞菌(SEQ | WP_013816155 | MTEQNEKPYNGTYYTLEDKHFWAAFFNLARH NAYITLAHIDRQLAYSKADITNDEDILFFKGQW KNLDNDLERKARLRSLILKHFSFLEGAAYGKK LFESQSSGNKSSKNKELTKKEKEELQANALSLD NLKSILFDFLQKLKDFRNYYSHYRHPESSELPL |

[0534]

| | | |
|---|---|---|
| ID No. 116) | | FDGNMLQRLYNVFDVSVQRVKRDHEHNDKV DPHRHFNHLVRKGKKDRYGNNDNPFFKHHFV DREGTVTEAGLLFFVSLFLEKRDAIWMQKKIR GFKGGTETYQQMTNEVFCRSRISLPKLKLESLR TDDWMLLDMLNELVRCPKSLYDRLREEDRAR FRVPVDILSDEEDTDGAEEDPFKNTLVRHQDRF PYFALRYFDLKKVFTSLRFQIDLGTYHFAIYKK NIGEQPEDRHLTRNLYGFGRIQDFAEEHRPEEW KRLVRDLDYFETGDKPYITQTTPHYHIEKGKIG LRFVPEGQHLWPSPEVGATRTGRSKYAQDKRF TAEAFLSAHELMPMMFYYFLLREKYSEEASAE RVQGRIKRVIEDVYAVYDAFARDEINTRDELD ACLADKGIRRGHLPRQMIGILSQEHKDMEEKIR KKLQEMMADTDHRLDMLDRQTDRKIRIGRKN AGLPKSGVIADWLVRDMMRFQPVAKDTSGKP LNNSKANSTEYRMLQRALALFGGEKERLTPYF RQMNLTGGNNPHPFLHETRWESHTNILSFYRS YLKARKAFLQSIGRSDRVENHRFLLLKEPKTDR QTLVAGWKGEFHLPRGIFTEAVRDCLIEMGLD EVGSYKEVGFMAKAVPLYFERACKDWVQPFY NYPFNVGNSLKPKKGRFLSKEKRAEEWESGKE RFRLAKLKKEILEAKEHPYLDFKSWQKFEREL RLVKNQDIITWMICGDLMEENKVEGLDTGTLY LKDIRTDVQEQGSLNVLNRVKPMRLPVVVYR ADSRGHVHKEQAPLATVYIEERDTKLLKQGNF KSFVKDRRLNGLFSFVDTGALAMEQYPISKLR VEYELAKYQTARVCAFEQTLELEESLLTRCPHL PDKNFRKMLESWSDPLLDKWPDLHRKVRLLIA VRNAFSHNQYPMYDEAVFSSIRKYDPSFPDAIE ERMGLNIAHRLSEEVKQAKETVERIIQA |
| 柱状黄杆菌(SEQ ID No. 117) | WP_014165541 | MSSKNESYNKQKTFNHYKQEDKYFFGGFLNN ADDNLRQVGKEFKTRINFNHNNNELASVFKDY FNKEKSVAKREHALNLLSNYFPVLERIQKHTN HNFEQTREIFELLLDTIKKLRDYYTHHYHKPITI NPKIYDFLDDTLLDVLITIKKKKVKNDTSRELL KEKLRPELTQLKNQKREELIKKGKKLLEENLE NAVFNHCLRPFLEENKTDDKQNKTVSLRKYRK SKPNEETSITLTQSGLVFLMSFFLHRKEFQVFTS GLEGFKAKVNTIKEEEISLNKNNIVYMITHWSY SYYNFKGLKHRIKTDQGVSTLEQNNTTHSLTN TNTKEALLTQIVDYLSKVPNEIYETLSEKQQKE FEEDINEYMRENPENEDSTFSSIVSHKVIRKRYE NKFNYFAMRFLDEYAELPTLRFMVNFGDYIKD |

[0535]

| | | |
|---|---|---|
| | | RQKKILESIQFDSERIIKKEIHLFEKLSLVTEYKK<br>NVYLKETSNIDLSRFPLFPNPSYVMANNNIPFYI<br>DSRSNNLDEYLNQKKKAQSQNKKRNLTFEKY<br>NKEQSKDAIIAMLQKEIGVKDLQQRSTIGLLSC<br>NELPSMLYEVIVKDIKGAELENKIAQKIREQYQ<br>SIRDFTLDSPQKDNIPTTLIKTINTDSSVTFENQP<br>IDIPRLKNAIQKELTLTQEKLLNVKEHEIEVDNY<br>NRNKNTYKFKNQPKNKVDDKKLQRKYVFYR<br>NEIRQEANWLASDLIHFMKNKSLWKGYMHNE<br>LQSFLAFFEDKKNDCIALLETVFNLKEDCILTK<br>GLKNLFLKHGNFIDFYKEYLKLKEDFLNTESTF<br>LENGLIGLPPKILKKELSKRFKYIFIVFQKRQFII<br>KELEEKKNNLYADAINLSRGIFDEKPTMIPFKK<br>PNPDEFASWFVASYQYNNYQSFYELTPDIVER<br>DKKKKYKNLRAINKVKIQDYYLKLMVDTLYQ<br>DLFNQPLDKSLSDFYVSKAEREKIKADAKAYQ<br>KRNDSSLWNKVIHLSLQNNRITANPKLKDIGK<br>YKRALQDEKIATLLTYDDRTWTYALQKPEKEN<br>ENDYKELHYTALNMELQEYEKVRSKELLKQV<br>QELEKQILEEYTDFLSTQIHPADFEREGNPNFK<br>KYLAHSILENEDDLDKLPEKVEAMRELDETITN<br>PIIKKAIVLIIIRNKMAHNQYPPKFIYDLANRFVP<br>KKEEEYFATYFNRVFETITKELWENKEKKDKT<br>QV |
| 扭曲冷弯菌(Psychroflexus torquis)(SEQ ID No. 118) | WP_015024765 | MESIIGLGLSFNPYKTADKHYFGSFLNLVENNL<br>NAVFAEFKERISYKAKDENISSLIEKHFIDNMSI<br>VDYEKKISILNGYLPIIDFLDDELENNLNTRVKN<br>FKKNFIILAEAIEKLRDYYTHFYHDPITFEDNKE<br>PLLELLDEVLLKTILDVKKKYLKTDKTKEILKD<br>SLREEMDLLVIRKTDELREKKKTNPKIQHTDSS<br>QIKNSIFNDAFQGLLYEDKGNNKKTQVSHRAK<br>TRLNPKDIHKQEERDFEIPLSTSGLVFLMSLFLS<br>KKEIEDFKSNIKGFKGKVVKDENHNSLKYMAT<br>HRVYSILAFKGLKYRIKTDTFSKETLMMQMID<br>ELSKVPDCVYQNLSETKQKDFIEDWNEYFKDN<br>EENTENLENSRVVHPVIRKRYEDKFNYFAIRFL<br>DEFANFKTLKFQVFMGYYIHDQRTKTIGTTNIT<br>TERTVKEKINVFGKLSKMDNLKKHFFSQLSDD<br>ENTDWEFFPNPSYNFLTQADNSPANNIPIYLEL<br>KNQQIIKEKDAIKAEVNQTQNRNPNKPSKRDL<br>LNKILKTYEDFHQGDPTAILSLNEIPALLHLFLV<br>KPNNKTGQQIENIIRIKIEKQFKAINHPSKNNKG<br>IPKSLFADTNVRVNAIKLKKDLEAELDMLNKK |

[0536]

| | | |
|---|---|---|
| | | HIAFKENQKASSNYDKLLKEHQFTPKNKRPEL<br>RKYVFYKSEKGEEATWLANDIKRFMPKDFKT<br>KWKGCQHSELQRKLAFYDRHTKQDIKELLSGC<br>EFDHSLLDINAYFQKDNFEDFFSKYLENRIETLE<br>GVLKKLHDFKNEPTPLKGVFKNCFKFLKRQNY<br>VTESPEIIKKRILAKPTFLPRGVFDERPTMKKGK<br>NPLKDKNEFAEWFVEYLENKDYQKFYNAEEY<br>RMRDADFKKNAVIKKQKLKDFYTLQMVNYLL<br>KEVFGKDEMNLQLSELFQTRQERLKLQGIAKK<br>QMNKETGDSSENTRNQTYIWNKDVPVSFFNG<br>KVTIDKVKLKNIGKYKRYERDERVKTFIGYEV<br>DEKWMMYLPHNWKDRYSVKPINVIDLQIQEY<br>EEIRSHELLKEIQNLEQYIYDHTTDKNILLQDGN<br>PNFKMYVLNGLLIGIKQVNIPDFIVLKQNTNFD<br>KIDFTGIASCSELEKKTIILIAIRNKFAHNQLPNK<br>MIYDLANEFLKIEKNETYANYYLKVLKKMISD<br>LA |
| 鸭疫里默氏杆菌 (SEQ ID No. 119) | WP_015345620 | MFFSFHNAQRVIFKHLYKAFDASLRMVKEDYK<br>AHFTVNLTRDFAHLNRKGKNKQDNPDFNRYR<br>FEKDGFFTESGLLFFTNLFLDKRDAYWMLKKV<br>SGFKASHKQREKMTTEVFCRSRILLPKLRLESR<br>YDHNQMLLDMLSELSRCPKLLYEKLSEENKKH<br>FQVEADGFLDEIEEEQNPFKDTLIRHQDRFPYF<br>ALRYLDLNESFKSIRFQVDLGTYHYCIYDKKIG<br>DEQEKRHLTRTLLSFGRLQDFTEINRPQEWKAL<br>TKDLDYKETSNQPFISKTTPHYHITDNKIGFRLG<br>TSKELYPSLEIKDGANRIAKYPYNSGFVAHAFIS<br>VHELLPLMFYQHLTGKSEDLLKETVRHIQRIYK<br>DFEEERINTIEDLEKANQGRLPLGAFPKQMLGL<br>LQNKQPDLSEKAKIKIEKLIAETKLLSHRLNTK<br>LKSSPKLGKRREKLIKTGVLADWLVKDFMRFQ<br>PVAYDAQNQPIKSSKANSTEFWFIRRALALYG<br>GEKNRLEGYFKQTNLIGNTNPHPFLNKFNWKA<br>CRNLVDFYQQYLEQREKFLEAIKHQPWEPYQY<br>CLLLKVPKENRKNLVKGWEQGGISLPRGLFTE<br>AIRETLSKDLTLSKPIRKEIKKHGRVGFISRAITL<br>YFKEKYQDKHQSFYNLSYKLEAKAPLLKKEEH<br>YEYWQQNKPQSPTESQRLELHTSDRWKDYLL<br>YKRWQHLEKKLRLYRNQDIMLWLMTLELTKN<br>HFKELNLNYHQLKLENLAVNVQEADAKLNPL<br>NQTLPMVLPVKVYPTTAFGEVQYHETPIRTVYI<br>REEQTKALKMGNFKALVKDRRLNGLFSFIKEE<br>NDTQKHPISQLRLRRELEIYQSLRVDAFKETLS |

[0537]

| | | |
|---|---|---|
| | | LEEKLLNKHASLSSLENEFRTLLEEWKKKYAA SSMVTDKHIAFIASVRNAFCHNQYPFYKETLH APILLFTVAQPTTEEKDGLGIAEALLKVLREYC EIVKSQI |
| 胸膜炎普雷沃菌 (Prevotella pleuritidis) (SEQ ID No. 120) | WP_021584635 | MENDKRLEESACYTLNDKHFWAAFLNLARHN VYITVNHINKTLELKNKKNQEIIIDNDQDILAIK THWAKVNGDLNKTDRLRELMIKHFPFLEAAIY SNNKEDKEEVKEEKQAKAQSFKSLKDCLFLFL EKLQEARNYYSHYKYSESSKEPEFEEGLLEKM YNTFDASIRLVKEDYQYNKDIDPEKDFKHLER KEDFNYLFTDKDNKGKITKNGLLFFVSLFLEKK DAIWMQQKFRGFKDNRGNKEKMTHEVFCRSR MLLPKIRLESTQTQDWILLDMLNELIRCPKSLY ERLQGAYREKFKVPFDSIDEDYDAEQEPFRNTL VRHQDRFPYFALRYFDYNEIFKNLRFQIDLGTY HFSIYKKLIGGKKEDRHLTHKLYGFERIQEFTK QNRPDKWQAIIKDLDTYETSNERYISETTPHYH LENQKIGIRFRNDNNDIWPSLKTNGEKNEKSKY NLDKPYQAEAFLSVHELLPMMFYYLLLKMEN TDNDKEDNEVGTKKKGNKNNKQEKHKIEEIIE NKIKDIYALYDAFTNGEINSIDELAEQREGKDIE IGHLPKQLIVILKNKSKDMAEKANRKQKEMIK DTKKRLATLDKQVKGEIEDGGRNIRLLKSGEIA RWLVNDMMRFQPVQKDNEGKPLNNSKANSTE YQMLQRSLALYNKEEKPTRYFRQVNLIKSSNP HPFLEDTKWEECYNILSFYRNYLKAKIKFLNKL KPEDWKKNQYFLMLKEPKTNRKTLVQGWKN GFNLPRGIFTEPIKEWFKRHQNDSEEYKKVEAL DRVGLVAKVIPLFFKEEYFKEDAQKEINNCVQP FYSFPYNVGNIHKPEEKNFLHCEERRKLWDKK KDKFKGYKAKEKSKKMTDKEKEEHRSYLEFQ SWNKFERELRLVRNQDILTWLLCTKLIDKLKID ELNIEELQKLRLKDIDTDTAKKEKNNILNRVMP MRLPVTVYEIDKSFNIVKDKPLHTVYIEETGTK LLKQGNFKALVKDRRLNGLFSFVKTSSEAESK SKPISKLRVEYELGAYQKARIDIIKDMLALEKT LIDNDENLPTNKFSDMLKSWLKGKGEANKAR LQNDVGLLVAVRNAFSHNQYPMYNSEVFKGM KLLSLSSDIPEKEGLGIAKQLKDKIKETIERIIEIE KEIRN |
| 牙龈卟啉单胞 | WP_021666319 | MNTVPASENKGQSRTVEDDPQYFGLYLNLARE NLIEVESHVRIKFGKKKLNEESLKQSLLCDHLL SVDRWTKVYGHSRRYLPFLHYFDPDSQIEKDH |

[0538]

| | | |
|---|---|---|
| 菌 (SEQ ID No. 121) | 7 | DSKTGVDPDSAQRLIRELYSLLDFLRNDFSHNR<br>LDGTTFEHLEVSPDISSFITGTYSLACGRAQSRF<br>ADFFKPDDFVLAKNRKEQLISVADGKECLTVS<br>GLAFFICLFLDREQASGMLSRIRGFKRTDENWA<br>RAVHETFCDLCIRHPHDRLESSNTKEALLLDML<br>NELNRCPRILYDMLPEEERAQFLPALDENSMN<br>NLSENSLNEESRLLWDGSSDWAEALTKRIRHQ<br>DRFPYLMLRFIEEMDLLKGIRFRVDLGEIELDS<br>YSKKVGRNGEYDRTITDHALAFGKLSDFQNEE<br>EVSRMISGEASYPVRFSLFAPRYAIYDNKIGYC<br>HTSDPVYPKSKTGEKRALSNPRSMGFISVHDLR<br>KLLLMELLCEGSFSRMQSDFLRKANRILDETAE<br>GKLQFSALFPEMRHRFIPPQNPKSKDRREKAET<br>TLEKYKQEIKGRKDKLNSQLLSAFDMDQRQLP<br>SRLLDEWMNIRPASHSVKLRTYVKQLNEDCRL<br>RLQKFRKDGDGKARAIPLVGEMATFLSQDIVR<br>MIISEETKKLITSAYYNEMQRSLAQYAGEENRH<br>QFRAIVAELRLLDPSSGHPFLSATMETAHRYTE<br>DFYKCYLEKKREWLAKTFYRPEQDENTKRRIS<br>VFFVPDGEARKLLPLLIRRRMKEQNDLQDWIR<br>NKQAHPIDLPSHLFDSKIMELLKVKDGKKKWN<br>EAFKDWWSTKYPDGMQPFYGLRRELNIHGKS<br>VSYIPSDGKKFADCYTHLMEKTVQDKKRELRT<br>AGKPVPPDLAADIKRSFHRAVNEREFMLRLVQ<br>EDDRLMLMAINKMMTDREEDILPGLKNIDSIL<br>DEENQFSLAVHAKVLEKEGEGGDNSLSLVPAT<br>IEIKSKRKDWSKYIRYRYDRRVPGLMSHFPEHK<br>ATLDEVKTLLGEYDRCRIKIFDWAFALEGAIMS<br>DRDLKPYLHESSSREGKSGEHSTLVKMLVEKK<br>GCLTPDESQYLILIRNKAAHNQFPCAAEMPLIY<br>RDVSAKVGSIEGSSAKDLPEGSSLVDSLWKKY<br>EMIIRKILPILDPENRFFGKLLNNMSQPINDL |
| 牙龈卟啉单胞菌 (SEQ ID No. 122) | WP_021666475 | MNTVPASENKGQSRTVEDDPQYFGLYLNLARE<br>NLIEVESHVRIKFGKKKLNEESLKQSLLCDHLL<br>SVDRWTKVYGHSRRYLPFLHYFDPDSQIEKDH<br>DSKTGVDPDSAQRLIRELYSLLDFLRNDFSHNR<br>LDGTTFEHLEVSPDISSFITGTYSLACGRAQSRF<br>ADFFKPDDFVLAKNRKEQLISVADGKECLTVS<br>GLAFFICLFLDREQASGMLSRIRGFKRTNENWA<br>RAVHETFCDLCIRHPHDRLESSNTKEALLLDML<br>NELNRCPRILYDMLPEEERAQFLPALDENSMN<br>NLSENSLNEESRLLWDGSSDWAEALTKRIRHQ<br>DRFPYLMLRFIEEMDLLKGIRFRVDLGEIELDS |

[0539]

| | | |
|---|---|---|
| | | YSKKVGRNGEYDRTITDHALAFGKLSDFQNEE<br>EVSRMISGEASYPVRFSLFAPRYAIYDNKIGYC<br>HTSDPVYPKSKTGEKRALSNPQSMGFISVHDLR<br>KLLLMELLCEGSFSRMQSGFLRKANRILDETAE<br>GKLQFSALFPEMRHRFIPPQNPKSKDRREKAET<br>TLEKYKQEIKGRKDKLNSQLLSAFDMNQRQLP<br>SRLLDEWMNIRPASHSVKLRTYVKQLNEDCRL<br>RLRKFRKDGDGKARAIPLVGEMATFLSQDIVR<br>MIISEETKKLITSAYYNEMQRSLAQYAGEENRR<br>QFRAIVAELHLLDPSSGHPFLSATMETAHRYTE<br>DFYKCYLEKKREWLAKTFYRPEQDENTKRRIS<br>VFFVPDGEARKLLPLLIRRRMKEQNDLQDWIR<br>NKQAHPIDLPSHLFDSKIMELLKVKDGKKKWN<br>EAFKDWWSTKYPDGMQPFYGLRRELNIHGKS<br>VSYIPSDGKKFADCYTHLMEKTVQDKKRELRT<br>AGKPVPPDLAADIKRSFHRAVNEREFMLRLVQ<br>EDDRLMLMAINKMMTDREEDILPGLKNIDSIL<br>DKENQFSLAVHAKVLEKEGEGGDNSLSLVPAT<br>IEIKSKRKDWSKYIRYRYDRRVPGLMSHFPEHK<br>ATLDEVKTLLGEYDRCRIKIFDWAFALEGAIMS<br>DRDLKPYLHESSSREGKSGEHSTLVKMLVEKK<br>GCLTPDESQYLILIRNKAAHNQFPCAAEMPLIY<br>RDVSAKVGSIEGSSAKDLPEGSSLVDSLWKKY<br>EMIIRKILPILDHENRFFGKLLNNMSQPINDL |
| 牙龈卟啉单胞菌 (SEQ ID No. 123) | WP_021677657 | MNTVPASENKGQSRTVEDDPQYFGLYLNLARE<br>NLIEVESHVRIKFGKKKLNEESLKQSLLCDHLL<br>SVDRWTKVYGHSRRYLPFLHYFDPDSQIEKDH<br>DSKTGVDPDSAQRLIRELYSLLDFLRNDFSHNR<br>LDGTTFEHLEVSPDISSFITGTYSLACGRAQSRF<br>ADFFKPDDFVLAKNRKEQLISVADGKECLTVS<br>GLAFFICLFLDREQASGMLSRIRGFKRTDENWA<br>RAVHETFCDLCIRHPHDRLESSNTKEALLLDML<br>NELNRCPRILYDMLPEEERAQFLPALDENSMN<br>NLSENSLNEESRLLWDGSSDWAEALTKRIRHQ<br>DRFPYLMLRFIEEMDLLKGIRFRVDLGEIELDS<br>YSKKVGRNGEYDRTITDHALAFGKLSDFQNEE<br>EVSRMISGEASYPVRFSLFAPRYAIYDNKIGYC<br>HTSDPVYPKSKTGEKRALSNPQSMGFISVHDLR<br>KLLLMELLCEGSFSRMQSGFLRKANRILDETAE<br>GKLQFSALFPEMRHRFIPPQNPKSKDRREKAET<br>TLEKYKQEIKGRKDKLNSQLLSAFDMNQRQLP<br>SRLLDEWMNIRPASHSVKLRTYVKQLNEDCRL<br>RLRKFRKDGDGKARAIPLVGEMATFLSQDIVR |

[0540]

| | | |
|---|---|---|
| | | MIISEETKKLITSAYYNEMQRSLAQYAGEENRR<br>QFRAIVAELHLLDPSSGHPFLSATMETAHRYTE<br>DFYKCYLEKKREWLAKTFYRPEQDENTKRRIS<br>VFFVPDGEARKLLPLLIRRRMKEQNDLQDWIR<br>NKQAHPIDLPSHLFDSKIMELLKVKDGKKKWN<br>EAFKDWWSTKYPDGMQPFYGLRRELNIHGKS<br>VSYIPSDGKKFADCYTHLMEKTVQDKKRELRT<br>AGKPVPPDLAADIKRSFHRAVNEREFMLRLVQ<br>EDDRLMLMAINKMMTDREEDILPGLKNIDSIL<br>DEENQFSLAVHAKVLEKEGEGGDNSLSLVPAT<br>IEIKSKRKDWSKYIRYRYDRRVPGLMSHFPEHK<br>ATLDEVKTLLGEYDRCRIKIFDWAFALEGAIMS<br>DRDLKPYLHESSSREGKSGEHSTLVKMLVEKK<br>GCLTPDESQYLILIRNKAAHNQFPCAAEMPLIY<br>RDVSAKVGSIEGSSAKDLPEGSSLVDSLWKKY<br>EMIIRKILPILDHENRFFGKLLNNMSQPINDL |
| 牙龈卟啉单胞菌 (SEQ ID No. 124) | WP_021680012 | MNTVPASENKGQSRTVEDDPQYFGLYLNLARE<br>NLIEVESHVRIKFGKKKLNEESLKQSLLCDHLL<br>SVDRWTKVYGHSRRYLPFLHYFDPDSQIEKDH<br>DSKTGVDPDSAQRLIRELYSLLDFLRNDFSHNR<br>LDGTTFEHLEVSPDISSFITGTYSLACGRAQSRF<br>ADFFKPDDFVLAKNRKEQLISVADGKECLTVS<br>GLAFFICLFLDREQASGMLSRIRGFKRTDENWA<br>RAVHETFCDLCIRHPHDRLESSNTKEALLLDML<br>NELNRCPRILYDMLPEEERAQFLPALDENSMN<br>NLSENSLNEESRLLWDGSSDWAEALTKRIRHQ<br>DRFPYLMLRFIEEMDLLKGIRFRVDLGEIELDS<br>YSKKVGRNGEYDRTITDHALAFGKLSDFQNEE<br>EVSRMISGEASYPVRFSLFAPRYAIYDNKIGYC<br>HTSDPVYPKSKTGEKRALSNPRSMGFISVHDLR<br>KLLLMELLCEGSFSRMQSDFLRKANRILDETAE<br>GKLQFSALFPEMRHRFIPPQNPKSKDRREKAET<br>TLEKYKQEIKGRKDKLNSQLLSAFDMDQRQLP<br>SRLLDEWMNIRPASHSVKLRTYVKQLNEDCRL<br>RLQKFRKDGDGKARAIPLVGEMATFLSQDIVR<br>MIISEETKKLITSAYYNEMQRSLAQYAGEENRH<br>QFRAIVAELRLLDPSSGHPFLSATMETAHRYTE<br>DFYKCYLEKKREWLAKTFYRPEQDENTKRRIS<br>VFFVPDGEARKLLPLLIRRRMKEQNDLQDWIR<br>NKQAHPIDLPSHLFDSKVMELLKVKDGKKKW<br>NEAFKDWWSTKYPDGMQPFYGLRRELNIHGK<br>SVSYIPSDGKKFADCYTHLMEKTVRDKKRELR<br>TAGKPVPPDLAAYIKRSFHRAVNEREFMLRLV |

| | | |
|---|---|---|
| | | QEDDRLMLMAINKIMTDREEDILPGLKNIDSIL<br>DKENQFSLAVHAKVLEKEGEGGDNSLSLVPAT<br>IEIKSKRKDWSKYIRYRYDRRVPGLMSHFPEHK<br>ATLDEVKTLLGEYDRCRIKIFDWAFALEGAIMS<br>DRDLKPYLHESSSREGKSGEHSTLVKMLVEKK<br>GCLTPDESQYLILIRNKAAHNQFPCAAEIPLIYR<br>DVSAKVGSIEGSSAKDLPEGSSLVDSLWKKYE<br>MIIRKILPILDPENRFFGKLLNNMSQPINDL |
| [0541] | 牙龈卟啉单胞菌 (SEQ ID No. 125) | WP_023846767 | MNTVPASENKGQSRTVEDDPQYFGLYLNLARE<br>NLIEVESHVRIKFGKKKLNEESLKQSLLCDHLL<br>SVDRWTKVYGHSRRYLPFLHYFDPDSQIEKDH<br>DSKTGVDPDSAQRLIRELYSLLDFLRNDFSHNR<br>LDGTTFEHLEVSPDISSFITGTYSLACGRAQSRF<br>ADFFKPDDFVLAKNRKEQLISVADGKECLTVS<br>GLAFFICLFLDREQASGMLSRIRGFKRTDENWA<br>RAVHETFCDLCIRHPHDRLESSNTKEALLLDML<br>NELNRCPRILYDMLPEEERAQFLPALDENSMN<br>NLSENSLNEESRLLWDGSSDWAEALTKRIRHQ<br>DRFPYLMLRFIEEMDLLKGIRFRVDLGEIELDS<br>YSKKVGRNGEYDRTITDHALAFGKLSDFQNEE<br>EVSRMISGEASYPVRFSLFAPRYAIYDNKIGYC<br>HTSDPVYPKSKTGEKRALSNPRSMGFISVHDLR<br>KLLLMELLCEGSFSRMQSDFLRKANRILDETAE<br>GKLQFSALFPEMRHRFIPPQNPKSKDRREKAET<br>TLEKYKQEIKGRKDKLNSQLLSAFDMNQRQLP<br>SRLLDEWMNIRPASHSVKLRTYVKQLNEDCRL<br>RLRKFRKDGDGKARAIPLVGEMATFLSQDIVR<br>MIISEETKKLITSAYYNEMQRSLAQYAGEENRR<br>QFRAIVAELHLLDPSSGHPFLSATMETAHRYTE<br>DFYKCYLEKKREWLAKTFYRPEQDENTKRRIS<br>VFFVPDGEARKLLPLLIRRRMKEQNDLQDWIR<br>NKQAHPIDLPSHLFDSKIMELLKVKDGKKKWN<br>EAFKDWWSTKYPDGMQPFYGLRRELNIHGKS<br>VSYIPSDGKKFADCYTHLMEKTVQDKKRELRT<br>AGKPVPPDLAADIKRSFHRAVNEREFMLRLVQ<br>EDDRLMLMAINKMMTDREEDILPGLKNIDSIL<br>DEENQFSLAVHAKVLEKEGEGGDNSLSLVPAT<br>IEIKSKRKDWSKYIRYRYDRRVPGLMSHFPEHK<br>ATLDEVKTLLGEYDRCRIKIFDWAFALEGAIMS<br>DRDLKPYLHESSSREGKSGEHSTLVKMLVEKK<br>GCLTPDESQYLILIRNKAAHNQFPCAAEMPLIY<br>RDVSAKVGSIEGSSAKDLPEGSSLVDSLWKKY<br>EMIIRKILPILDPENRFFGKLLNNMSQPINDL |

[0542]

| | | |
|---|---|---|
| 斐氏普雷沃菌(Prevotella falsenii) (SEQ ID No. 126) | WP_036884929 | MKNDNNSTKSTDYTLGDKHFWAAFLNLARHN<br>VYITVNHINKVLELKNKKDQEIIIDNDQDILAIK<br>TLWGKVDTDINKKDRLRELIMKHFPFLEAATY<br>QQSSTNNTKQKEEEQAKAQSFESLKDCLFLFLE<br>KLREARNYYSHYKHSKSLEEPKLEEKLLENMY<br>NIFDTNVQLVIKDYEHNKDINPEEDFKHLGRAE<br>GEFNYYFTRNKKGNITESGLLFFVSLFLEKKDA<br>IWAQTKIKGFKDNRENKQKMTHEVFCRSRML<br>LPKLRLESTQTQDWILLDMLNELIRCPKSLYKR<br>LQGEKREKFRVPFDPADEDYDAEQEPFKNTLV<br>RHQDRFPYFALRYFDYNEIFTNLRFQIDLGTYH<br>FSIYKKQIGDKKEDRHLTHKLYGFERIQEFAKE<br>NRPDEWKALVKDLDTFEESNEPYISETTPHYHL<br>ENQKIGIRNKNKKKKKTIWPSLETKTTVNERSK<br>YNLGKSFKAEAFLSVHELLPMMFYYLLLNKEE<br>PNNGKINASKVEGIIEKKIRDIYKLYGAFANEEI<br>NNEEELKEYCEGKDIAIRHLPKQMIAILKNEYK<br>DMAKKAEDKQKKMIKDTKKRLAALDKQVKG<br>EVEDGGRNIKPLKSGRIASWLVNDMMRFQPVQ<br>RDRDGYPLNNSKANSTEYQLLQRTLALFGSER<br>ERLAPYFRQMNLIGKDNPHPFLKDTKWKEHN<br>NILSFYRSYLEAKKNFLGSLKPEDWKKNQYFL<br>KLKEPKTNRETLVQGWKNGFNLPRGIFTEPIRE<br>WFIRHQNESEEYKKVKDFDRIGLVAKVIPLFFK<br>EDYQKEIEDYVQPFYGYPFNVGNIHNSQEGTFL<br>NKKEREELWKGNKTKFKDYKTKEKNKEKTNK<br>DKFKKKTDEEKEEFRSYLDFQSWKKFERELRL<br>VRNQDIVTWLLCMELIDKLKIDELNIEELQKLR<br>LKDIDTDTAKKEKNNILNRIMPMELPVTVYET<br>DDSNNIIKDKPLHTIYIKEAETKLLKQGNFKAL<br>VKDRRLNGLFSFVETSSEAELKSKPISKSLVEYE<br>LGEYQRARVEIIKDMLRLEETLIGNDEKLPTNK<br>FRQMLDKWLEHKKETDDTDLKNDVKLLTEVR<br>NAFSHNQYPMRDRIAFANIKPFSLSSANTSNEE<br>GLGIAKKLKDKTKETIDRIIEIEEQTATKR |
| 胸膜炎普雷沃菌(SEQ ID No. 127) | WP_036931485 | MENDKRLEESTCYTLNDKHFWAAFLNLARHN<br>VYITINHINKLLEIRQIDNDEKVLDIKALWQKV<br>DKDINQKARLRELMIKHFPFLEAAIYSNNKEDK<br>EEVKEEKQAKAQSFKSLKDCLFLFLEKLQEAR<br>NYYSHYKSSESSKEPEFEEGLLEKMYNTFGVSI<br>RLVKEDYQYNKDIDPEKDFKHLERKEDFNYLF<br>TDKDNKGKITKNGLLFFVSLFLEKKDAIWMQQ<br>KLRGFKDNRGNKEKMTHEVFCRSRMLLPKIRL |

[0543]

| | | |
|---|---|---|
| | | ESTQTQDWILLDMLNELIRCPKSLYERLQGAY<br>REKFKVPFDSIDEDYDAEQEPFRNTLVRHQDRF<br>PYFALRYFDYNEIFKNLRFQIDLGTYHFSIYKKL<br>IGDNKEDRHLTHKLYGFERIQEFAKQKRPNEW<br>QALVKDLDIYETSNEQYISETTPHYHLENQKIGI<br>RFKNKKDKIWPSLETNGKENEKSKYNLDKSFQ<br>AEAFLSIHELLPMMFYDLLLKKEEPNNDEKNA<br>SIVEGFIKKEIKRMYAIYDAFANEEINSKEGLEE<br>YCKNKGFQERHLPKQMIAILTNKSKNMAEKA<br>KRKQKEMIKDTKKRLATLDKQVKGEIEDGGR<br>NIRLLKSGEIARWLVNDMMRFQSVQKDKEGK<br>PLNNSKANSTEYQMLQRSLALYNKEQKPTPYF<br>IQVNLIKSSNPHPFLEETKWEECNNILSFYRSYL<br>EAKKNFLESLKPEDWKKNQYFLMLKEPKTNR<br>KTLVQGWKNGFNLPRGIFTEPIKEWFKRHQND<br>SEEYKKVEALDRVGLVAKVIPLFFKEEYFKED<br>AQKEINNCVQPFYSFPYNVGNIHKPEEKNFLHC<br>EERRKLWDKKKDKFKGYKAKEKSKKMTDKE<br>KEEHRSYLEFQSWNKFERELRLVRNQDIVTWL<br>LCTELIDKLKIDELNIEELQKLRLKDIDTDTAKK<br>EKNNILNRIMPMQLPVTVYEIDKSFNIVKDKPL<br>HTIYIEETGTKLLKQGNFKALVKDRRLNGLFSF<br>VKTSSEAESKSKPISKLRVEYELGAYQKARIDII<br>KDMLALEKTLIDNDENLPTNKFSDMLKSWLK<br>GKGEANKARLQNDVDLLVAIRNAFSHNQYPM<br>YNSEVFKGMKLLSLSSDIPEKEGLGIAKQLKDK<br>IKETIERIIEIEKEIRN |
| [牙龈卟啉单胞菌(SEQ ID No. 128) | WP_039417390 | MTEQNERPYNGTYYTLEDKHFWAAFFNLARH<br>NAYITLAHIDRQLAYSKADITNDEDILFFKGQW<br>KNLDNDLERKARLRSLILKHFSFLEGAAYGKK<br>LFESQSSGNKSSKKKELTKKEKEELQANALSLD<br>NLKSILFDFLQKLKDFRNYYSHYRHPESSELPL<br>FDGNMLQRLYNVFDVSVQRVKRDHEHNDKV<br>DPHRHFNHLVRKGKKDRYGNNDNPFFKHHFV<br>DREGTVTEAGLLFFVSLFLEKRDAIWMQKKIR<br>GFKGGTEAYQQMTNEVFCRSRISLPKLKLESLR<br>TDDWMLLDMLNELVRCPKSLYDRLREEDRAR<br>FRVPIDILSDEDDTDGTEEDPFKNTLVRHQDRF<br>PYFALRYFDLKKVFTSLRFHIDLGTYHFAIYKK<br>NIGEQPEDRHLTRNLYGFGRIQDFAEEHRPEEW<br>KRLVRDLDYFETGDKPYITQTTPHYHIEKGKIG<br>LRFVPEGQHLWPSPEVGATRTGRSKYAQDKRL<br>TAEAFLSVHELMPMMFYYFLLREKYSEEVSAE |

[0544]

| | | |
|---|---|---|
| | | KVQGRIKRVIEDVYAVYDAFARGEIDTLDRLD<br>ACLADKGIRRGHLPRQMIAILSQEHKDMEEKV<br>RKKLQEMIADTDHRLDMLDRQTDRKIRIGRKN<br>AGLPKSGVIADWLVRDMMRFQPVAKDTSGKP<br>LNNSKANSTEYRMLQRALALFGGEKERLTPYF<br>RQMNLTGGNNPHPFLHETRWESHTNILSFYRS<br>YLKARKAFLQSIGRSDREENHRFLLLKEPKTDR<br>QTLVAGWKSEFHLPRGIFTEAVRDCLIEMGYD<br>EVGSYKEVGFMAKAVPLYFERACKDRVQPFY<br>DYPFNVGNSLKPKKGRFLSKEKRAEEWESGKE<br>RFRLAKLKKEILEAKEHPYLDFKSWQKFEREL<br>RLVKNQDIITWMMCRDLMEENKVEGLDTGTL<br>YLKDIRTDVHEQGSLNVLNRVKPMRLPVVVY<br>RADSRGHVHKEQAPLATVYIEERDTKLLKQGN<br>FKSFVKDRRLNGLFSFVDTGALAMEQYPISKLR<br>VEYELAKYQTARVCAFEQTLELEESLLTRYPH<br>LPDKNFRKMLESWSDPLLDKWPDLHRKVRLLI<br>AVRNAFSHNQYPMYDEAVFSSIRKYDPSSPDAI<br>EERMGLNIAHRLSEEVKQAKEMAERIIQV |
| | 动物口腔卟啉单胞菌(SEQ ID No. 129) | WP_039418912 | MTEQSERPYNGTYYTLEDKHFWAAFLNLARH<br>NAYITLTHIDRQLAYSKADITNDQDVLSFKAL<br>WKNLDNDLERKSRLRSLILKHFSFLEGAAYGK<br>KLFESKSSGNKSSKNKELTKKEKEELQANALSL<br>DNLKSILFDFLQKLKDFRNYYSHYRHSGSSELP<br>LFDGNMLQRLYNVFDVSVQRVKRDHEHNDKV<br>DPHRHFNHLVRKGKKDRYGHNDNPSFKHHFV<br>DSEGMVTEAGLLFFVSLFLEKRDAIWMQKKIR<br>GFKGGTETYQQMTNEVFCRSRISLPKLKLESLR<br>MDDWMLLDMLNELVRCPKPLYDRLREDDRA<br>CFRVPVDILPDEDDTDGGGEDPFKNTLVRHQD<br>RFPYFALRYFDLKKVFTSLRFHIDLGTYHFAIY<br>KKMIGEQPEDRHLTRNLYGFGRIQDFAEEHRPE<br>EWKRLVRDLDYFETGDKPYISQTSPHYHIEKG<br>KIGLRFMPEGQHLWPSPEVGTTRTGRSKYAQD<br>KRLTAEAFLSVHELMPMMFYYFLLREKYSEEV<br>SAEKVQGRIKRVIEDVYAIYDAFARDEINTLKE<br>LDACLADKGIRRGHLPKQMIAILSQEHKNMEE<br>KVRKKLQEMIADTDHRLDMLDRQTDRKIRIGR<br>KNAGLPKSGVIADWLVRDMMRFQPVAKDASG<br>KPLNNSKANSTEYRMLQRALALFGGEKERLTP<br>YFRQMNLTGGNNPHPFLHDTRWESHTNILSFY<br>RSYLRARKAFLERIGRSDRMENRPFLLLKEPKT<br>DRQTLVAGWKSEFHLPRGIFTEAVRDCLIEMG |

| | | |
|---|---|---|
| | | YDEVGSYREVGFMAKAVPLYFERACEDRVQP<br>FYDSPFNVGNSLKPKKGRFLSKEERAEEWERG<br>KERFRDLEAWSHSAARRIEDAFAGIEYASPGN<br>KKKIEQLLRDLSLWEAFESKLKVRADKINLAK<br>LKKEILEAQEHPYHDFKSWQKFERELRLVKNQ<br>DIITWMMCRDLMEENKVEGLDTGTLYLKDIRT<br>NVQEQGSLNVLNHVKPMRLPVVVYRADSRGH<br>VHKEEAPLATVYIEERDTKLLKQGNFKSFVKD<br>RRLNGLFSFVDTGGLAMEQYPISKLRVEYELA<br>KYQTARVCAFEQTLELEESLLTRYPHLPDKNFR<br>KMLESWSDPLLAKWPELHGKVRLLIAVRNAFS<br>HNQYPMYDEAVFSSIRKYDPSSPDAIEERMGL<br>NIAHRLSEEVKQAKETVERIIQA |
| 动物口腔卟啉单胞菌(SEQ ID No. 130) | WP_039419792 | MTEQSERPYNGTYYTLEDKHFWAAFLNLARH<br>NAYITLTHIDRQLAYSKADITNDQDVLSFKAL<br>WKNLDNDLERKSRLRSLILKHFSFLEGAAYGK<br>KLFESKSSGNKSSKNKELTKKEKEELQANALSL<br>DNLKSILFDFLQKLKDFRNYYSHYRHSGSSELP<br>LFDGNMLQRLYNVFDVSVQRVKRDHEHNDKV<br>DPHRHFNHLVRKGKKDRYGHNDNPSFKHHFV<br>DGEGMVTEAGLLFFVSLFLEKRDAIWMQKKIR<br>GFKGGTETYQQMTNEVFCRSRISLPKLKLESLR<br>TDDWMLLDMLNELVRCPKPLYDRLREKDRAR<br>FRVPVDILPDEDDTDGGGEDPFKNTLVRHQDR<br>FPYFALRYFDLKKVFTSLRFHIDLGTYHFAIYK<br>KVIGEQPEDRHLTRNLYGFGRIQDFAEEHRPEE<br>WKRLVRDLDYFETGDKPYISQTTPHYHIEKGKI<br>GLRFVPEGQHLWPSPEVGTTRTGRSKYAQDKR<br>LTAEAFLSVHELMPMMFYYFLLREKYSEEVSA<br>EKVQGRIKRVIEDVYAIYDAFARDEINTRDELD<br>ACLADKGIRRGHLPKQMIGILSQEHKNMEEKV<br>RKKLQEMIADTDHRLDMLDRQTDRKIRIGRKN<br>AGLPKSGVIADWLVRDMMRFQPVAKDTSGKP<br>LNNSKANSTEYRMLQRALALFGGEKERLTPYF<br>RQMNLTGGNNPHPFLDETRWESHTNILSFYRS<br>YLRARKAFLERIGRSDRVENRPFLLLKEPKTDR<br>QTLVAGWKSEFHLPRGIFTEAVRDCLIEMGYD<br>EVGSYKEVGFMAKAVPLYFERACKDRVQPFY<br>DSPFNVGNSLKPKKGRFLSKEKRAEEWESGKE<br>RFRLAKLKKEILEAQEHPYHDFKSWQKFEREL<br>RLVKNQDIITWMMCRDLMEENKVEGLDTGTL<br>YLKDIRPNVQEQGSLNVLNRVKPMRLPVVVYR<br>ADSRGHVHKEEAPLATVYIEERDTKLLKQGNF |

[0545]

[0546]

| | | |
|---|---|---|
| | | KSFVKDRRLNGLFSFVDTGGLAMEQYPISKLR<br>VEYELAKYQTARVCVFELTLRLEESLLSRYPHL<br>PDESFREMLESWSDPLLAKWPELHGKVRLLIA<br>VRNAFSHNQYPMYDEAVFSSIRKYDPSSPDAIE<br>ERMGLNIAHRLSEEVKQAKETVERIIQA |
| 动物口腔卟啉单胞菌(SEQ ID No. 131) | WP_039426176 | MTEQSERPYNGTYYTLEDKHFWAAFLNLARH<br>NAYITLTHIDRQLAYSKADITNDQDVLSFKAL<br>WKNFDNDLERKSRLRSLILKHFSFLEGAAYGK<br>KLFESKSSGNKSSKNKELTKKEKEELQANALSL<br>DNLKSILFDFLQKLKDFRNYYSHYRHSGSSELP<br>LFDGNMLQRLYNVFDVSVQRVKRDHEHNDKV<br>DPHYHFNHLVRKGKKDRYGHNDNPSFKHHFV<br>DSEGMVTEAGLLFFVSLFLEKRDAIWMQKKIR<br>GFKGGTGPYEQMTNEVFCRSRISLPKLKLESLR<br>TDDWMLLDMLNELVRCPKPLYDRLREKDRAC<br>FRVPVDILPDEDDTDGGGEDPFKNTLVRHQDR<br>FPYFALRYFDLKKVFTSLRFHIDLGTYHFAIYK<br>KMIGEQPEDRHLTRNLYGFGRIQDFAEEHRPEE<br>WKRLVRDLDYFETGDKPYISQTTPHYHIEKGKI<br>GLRFMPEGQHLWPSPEVGTTRTGRSKYAQDK<br>RLTAEAFLSVHELMPMMFYYFLLREKYSEEVS<br>AEKVQGRIKRVIKDVYAIYDAFARDEINTLKEL<br>DACSADKGIRRGHLPKQMIGILSQEHKNMEEK<br>VRKKLQEMIADTDHRLDMLDRQTDRKIRIGRK<br>NAGLPKSGVIADWLVRDMMRFQPVAKDTSGK<br>PLNNSKANSTEYRMLQRALALFGGEKERLTPY<br>FRQMNLTGGNNPHPFLDETRWESHTNILSFYRS<br>YLRARKAFLERIGRSDRVENRPFLLLKEPKNDR<br>QTLVAGWKSEFHLPRGIFTEAVRDCLIEMGYD<br>EVGSYKEVGFMAKAVPLYFERACKDRVQPFY<br>DSPFNVGNSLKPKKGRFLSKEKRAEEWESGKE<br>RFRLAKLKKEILEAKEHPYHDFKSWQKFEREL<br>RLVKNQDIITWMMCRDLMEENKVEGLDTGTL<br>YLKDIRTDVHEQGSLNVLNRVKPMRLPVVVY<br>RADSRGHVHKEQAPLATVYIEERDTKLLKQGN<br>FKSFVKDRRLNGLFSFVDTGGLAMEQYPISKLR<br>VEYELAKYQTARVCAFEQTLELEESLLTRYPH<br>LPDENFREMLESWSDPLLGKWPDLHGKVRLLI<br>AVRNAFSHNQYPMYDEAVFSSIRKYDPSSPDAI<br>EERMGLNIAHRLSEEVKQAKETVERIIQA |
| 动物口腔卟啉 | WP_039433177 | MTEQSERPYNGTYYTLEDKHFWAAFLNLARH<br>NAYITLTHIDRQLAYSKADITNDQDVLSFKAL<br>WKNFDNDLERKSRLRSLILKHFSFLEGAAYGK |

[0547]

| | | |
|---|---|---|
| 单胞菌(SEQ ID No. 132) | 8 | KLFESKSSGNKSSKNKELTKKEKEELQANALSL<br>DNLKSILFDFLQKLKDFRNYYSHYRHSESSELP<br>LFDGNMLQRLYNVFDVSVQRVKRDHEHNDKV<br>DPHRHFNHLVRKGKKDRYGHNDNPSFKHHFV<br>DGEGMVTEAGLLFFVSLFLEKRDAIWMQKKIR<br>GFKGGTETYQQMTNEVFCRSRISLPKLKLESLR<br>TDDWMLLDMLNELVRCPKPLYDRLREDDRAC<br>FRVPVDILPDEDDTDGGGEDPFKNTLVRHQDR<br>FPYFALRYFDLKKVFTSLRFHIDLGTYHFAIYK<br>KMIGEQPEDRHLTRNLYGFGRIQDFAEEHRPEE<br>WKRLVRDLDYFETGDKPYISQTSPHYHIEKGKI<br>GLRFMPEGQHLWPSPEVGTTRTGRSKYAQDK<br>RLTAEAFLSVHELMPMMFYYFLLREKYSEEVS<br>AEKVQGRIKRVIEDVYAIYDAFARDEINTLKEL<br>DACLADKGIRRGHLPKQMIAILSQEHKDMEEKI<br>RKKLQEMIADTDHRLDMLDRQTDRKIRIGRKN<br>AGLPKSGVIADWLVRDMMRFQPVAKDTSGKP<br>LNNSKANSTEYRMLQRALALFGGEKKRLTPYF<br>RQMNLTGGNNPHPFLHETRWESHTNILSFYRS<br>YLRARKAFLERIGRSDRMENRPFLLLKEPKTDR<br>QTLVAGWKSEFHLPRGIFTEAVRDCLIEMGYD<br>EVGSYREVGFMAKAVPLYFERACEDRVQPFY<br>DSPFNVGNSLKPKKGRFLSKEERAEEWERGKE<br>RFRDLEAWSHSAARRIEDAFAGIEYASPGNKK<br>KIEQLLRDLSLWEAFESKLKVRADKINLAKLK<br>KEILEAQEHPYHDFKSWQKFERELRLVKNQDII<br>TWMMCRDLMEENKVEGLDTGTLYLKDIRPNV<br>QEQGSLNVLNRVKPMRLPVVVYRADSRGHVH<br>KEEAPLATVYIEERDTKLLKQGNFKSFVKDRR<br>LNGLFSFVDTGGLAMEQYPISKLRVEYELAKY<br>QTARVCVFELTLRLEESLLTRYPHLPDESFRKM<br>LESWSDPLLAKWPELHGKVRLLIAVRNAFSHN<br>QYPMYDEAVFSSIRKYDPSSPDAIEERMGLNIA<br>HRLSEEVKQAKETVERIIQV |
| 动物口腔卟啉单胞菌(SEQ ID No. 133) | WP_039437199 | MTEQSERPYNGTYYTLEDKHFWAAFLNLARH<br>NAYITLTHIDRQLAYSKADITNDEDILFFKGQW<br>KNLDNDLERKSRLRSLILKHFSFLEGAAYGKKF<br>FESKSSGNKSSKNKELTKKEKEELQANALSLD<br>NLKSILFDFLQKLKDFRNYYSHYRHSGSSELPL<br>FDGNMLQRLYNVFDVSVQRVKRDHEHNDEVD<br>PHYHFNHLVRKGKKDRYGHNDNPSFKHHFVD<br>GEGMVTEAGLLFFVSLFLEKRDAIWMQKKIRG<br>FKGGTEPYEQMTNEVFCRSRISLPKLKLESLRT |

[0548]

| | | |
|---|---|---|
| | | DDWMLLDMLNELVRCPKPLYDRLREKDRACF<br>RVPVDILPDEDDTDGGGEDPFKNTLVRHQDRF<br>PYFALRYFDLKKVFTSLRFHIDLGTYHFAIYKK<br>MIGEQPEDRHLTRNLYGFGRIQDFAEEHRPEE<br>WKRLVRDLDYFETGDKPYISQTTPHYHIEKGKI<br>GLRFVPEGQHLWPSPEVGTTRTGRSKYAQDKR<br>LTAEAFLSVHELMPMMFYYFLLREKYSEEVSA<br>EKVQGRIKRVIEDVYAIYDAFARDEINTLKELD<br>ACLADKGIRRGHLPKQMIGILSQERKDMEEKV<br>RKKLQEMIADTDHRLDMLDRQTDRKIRIGRKN<br>AGLPKSGVIADWLVRDMMRFQPVAKDTSGKP<br>LNNSKANSTEYRMLQRALALFGGEKERLTPYF<br>RQMNLTGGNNPHPFLHETRWESHTNILSFYRS<br>YLRARKAFLERIGRSDRVENCPFLLLKEPKTDR<br>QTLVAGWKGEFHLPRGIFTEAVRDCLIEMGYD<br>EVGSYREVGFMAKAVPLYFERACEDRVQPFY<br>DSPFNVGNSLKPKKGRFLSKEKRAEEWESGKE<br>RFRLAKLKKEILEAQEHPYHDFKSWQKFEREL<br>RLVKNQDIITWMMCRDLMEENKVEGLDTGTL<br>YLKDIRPNVQEQGSLNVLNRVKPMRLPVVVYR<br>ADSRGHVHKEEAPLATVYIEERDTKLLKQGNF<br>KSFVKDRRLNGLFSFVDTGALAMEQYPISKLR<br>VEYELAKYQTARVCAFEQTLELEESLLTRYPH<br>LPDESFREMLESWSDPLLTKWPELHGKVRLLIA<br>VRNAFSHNQYPMYDEAVFSSIWKYDPSSPDAI<br>EERMGLNIAHRLSEEVKQAKETIERIIQA |
| 动物口腔卟啉单胞菌(SEQ ID No. 134) | WP_039442171 | MTEQSERPYNGTYYTLEDKHFWAAFLNLARH<br>NAYITLTHIDRQLAYSKADITNDQDVLSFKAL<br>WKNLDNDLERKSRLRSLILKHFSFLEGAAYGK<br>KLFESKSSGNKSSKNKELTKKEKEELQANALSL<br>DNLKSILFDFLQKLKDFRNYYSHYRHSGSSELP<br>LFDGNMLQRLYNVFDVSVQRVKRDHEHNDKV<br>DPHYHFNHLVRKGKKDRYGHNDNPSFKHHFV<br>DSEGMVTEAGLLFFVSLFLEKRDAIWMQKKIR<br>GFKGGTGPYEQMTNEVFCRSRISLPKLKLESLR<br>TDDWMLLDMLNELVRCPKPLYDRLREKDRAC<br>FRVPVDILPDEDDTDGGGEDPFKNTLVRHQDR<br>FPYFALRYFDLKKVFTSLRFHIDLGTYHFAIYK<br>KMIGEQPEDRHLTRNLYGFGRIQDFAEEHRPEE<br>WKRLVRDLDYLETGDKPYISQTTPHYHIEKGKI<br>GLRFVPEGQHLWPSPEVGTTRTGRSKCAQDKR<br>LTAEAFLSVHELMPMMFYYFLLREKYSEEVSA<br>EKVQGRIKRVIEDVYAIYDAFARDEINTLKELD |

| | | |
|---|---|---|
| | | TCLADKGIRRGHLPKQMITILSQERKDMKEKIR<br>KKLQEMIADTDHRLDMLDRQTDRKIRIGRKNA<br>GLPKSGVIADWLVRDMMRFQPVAKDASGKPL<br>NNSKANSTEYRMLQRALALFGGEKERLTPYFR<br>QMNLTGGNNPHPFLHETRWESHTNILSFYRSY<br>LRARKAFLERIGRSDRVENCPFLLLKEPKTDRQ<br>TLVAGWKDEFHLPRGIFTEAVRDCLIEMGYDE<br>VGSYREVGFMAKAVPLYFERACEDRVQPFYDS<br>PFNVGNSLKPKKGRFLSKEDRAEEWERGMERF<br>RDLEAWSHSAARRIKDAFAGIEYASPGNKKKIE<br>QLLRDLSLWEAFESKLKVRADKINLAKLKKEIL<br>EAQEHPYHDFKSWQKFERELRLVKNQDIITWM<br>MCRDLMEENKVEGLDTGTLYLKDIRPNVQEQ<br>GSLNVLNRVKPMRLPVVVYRADSRGHVHKEA<br>PLATVYIEERNTKLLKQGNFKSFVKDRRLNGLF<br>SFVDTGGLAMEQYPISKLRVEYELAKYQTARV<br>CVFELTLRLEESLLSRYPHLPDESFREMLESWS<br>DPLLAKWPELHGKVRLLIAVRNAFSHNQYPM<br>YDEAVFSSIRKYDPSSPDAIEERMGLNIAHRLSE<br>EVKQAKETVERIIQA |
| [0549] | 动物口腔卟啉单胞菌(SEQ ID No. 135) | WP_039445055 | MNTVPATENKGQSRTVEDDPQYFGLYLNLAR<br>ENLIEVESHVRIKFGKKKLNEESLKQSLLCDHL<br>LSIDRWTKVYGHSRRYLPFLHCFDPDSGIEKDH<br>DSKTGVDPDSAQRLIRELYSLLDFLRNDFSHNR<br>LDGTTFEHLKVSPDISSFITGAYTFACERAQSRF<br>ADFFKPDDFLLAKNRKEQLISVADGKECLTVS<br>GFAFFICLFLDREQASGMLSRIRGFKRTDENWA<br>RAVHETFCDLCIRHPHDRLESSNTKEALLLDML<br>NELNRCPRILYDMLPEEERAQFLPALDENSMN<br>NLSENSLNEESRLLWDGSSDWAEALTKRIRHQ<br>DRFPYLMLRFIEEMDLLKGIRFRVDLGEIELDS<br>YSKKVGRNGEYDRTITDHALAFGKLSDFQNEE<br>EVSRMISGEASYPVRFSLFAPRYAIYDNKIGYC<br>HTSDPVYPKSKTGEKRALSNPQSMGFISVHDLR<br>KLLLMELLCEGSFSRMQSDFLRKANRILDETAE<br>GKLQFSALFPEMRHRFIPPQNPKSKDRREKAET<br>TLEKYKQEIKGRKDKLNSQLLSAFDMNQRQLP<br>SRLLDEWMNIRPASHSVKLRTYVKQLNEDCRL<br>RLRKFRKDGDGKARAIPLVGEMATFLSQDIVR<br>MIISEETKKLITSAYYNEMQRSLAQYAGEENRR<br>QFRAIVAELHLLDPSSGHPFLSATMETAHRYTE<br>DFYKCYLEKKREWLAKTFYRPEQDENTKRRIS<br>VFFVPDGEARKLLPLLIRRRMKEQNDLQDWIR |

| | | |
|---|---|---|
| | | NKQAHPIDLPSHLFDSKIMELLKVKDGKKKWN<br>EAFKDWWSTKYPDGMQPFYGLRRELNIHGKS<br>VSYIPSDGKKFADCYTHLMEKTVRDKKRELRT<br>AGKPVPPDLAAYIKRSFHRAVNEREFMLRLVQ<br>EDDRLMLMAINKMMTDREEDILPGLKNIDSIL<br>DEENQFSLAVHAKVLEKEGEGGDNSLSLVPAT<br>IEIKSKRKDWSKYIRYRYDRRVPGLMSHFPEHK<br>ATLDEVKTLLGEYDRCRIKIFDWAFALEGAIMS<br>DRDLKPYLHESSSREGKSGEHSTLVKMLVEKK<br>GCLTPDESQYLILIRNKAAHNQFPCAAEMPLIY<br>RDVSAKVGSIEGSSAKDLPEGSSLVDSLWKKY<br>EMIIRKILPILDHENRFFGKLLNNMSQPINDL |
| [0550] | 狗咬二氧化碳纤维菌(Capnocytophaga cynodegmi)(SEQ ID No. 136) | WP_041989581 | MENKTSLGNNIYYNPFKPQDKSYFAGYLNAA<br>MENIDSVFRELGKRLKGKEYTSENFFDAIFKEN<br>ISLVEYERYVKLLSDYFPMARLLDKKEVPIKER<br>KENFKKNFRGIIKAVRDLRNFYTHKEHGEVEIT<br>DEIFGVLDEMLKSTVLTVKKKKIKTDKTKEILK<br>KSIEKQLDILCQKKLEYLKDTARKIEEKRRNQR<br>ERGEKKLVPRFEYSDRRDDLIAAIYNDAFDVYI<br>DKKKDSLKESSKTKYNTESYPQQEEGDLKIPIS<br>KNGVVFLLSLFLSKQEVHAFKSKIAGFKATVID<br>EATVSHRKNSICFMATHEIFSHLAYKKLKRKV<br>RTAEINYSEAENAEQLSIYAKETLMMQMLDEL<br>SKVPDVVYQNLSEDVQKTFIEDWNEYLKENN<br>GDVGTMEEEQVIHPVIRKRYEDKFNYFAIRFLD<br>EFAQFPTLRFQVHLGNYLHDSRPKEHLISDRRI<br>KEKITVFGRLSELEHKKALFIKNTETNEDRKHY<br>WEVFPNPNYDFPKENISVNDKDFPIAGSILDRE<br>KQPTAGKIGIKVNLLNQKYISEVDKAVKAHQL<br>KQRNNKPSIQNIIEEIVPINGSNPKEIIVFGGQPT<br>AYLSMNDIHSILYEFFDKWEKKKEKLEKKGEK<br>ELRKEIGKELEEKIVGKIQTQIQQIIDKDINAKIL<br>KPYQDDDSTAIDKEKLIKDLKQEQKILQKLKNE<br>QTAREKEYQECIAYQEESRKIKRSDKSRQKYLR<br>NQLKRKYPEVPTRKEILYYQEKGKVAVWLAN<br>DIKRFMPTDFKNEWKGEQHSLLQKSLAYYEQC<br>KEELKNLLPQQKVFKHLPFELGGHFQQKYLYQ<br>FYTRYLDKRLEHISGLVQQAENFKNENKVFKK<br>VENECFKFLKKQNYTHKGLDAQAQSVLGYPIF<br>LERGFMDEKPTIIKGKTFKGNESLFTDWFRYYK<br>EYQNFQTFYDTENYPLVELEKKQADRKRETKI<br>YQQKKNDVFTLLMAKHIFKSVFKQDSIDRFSLE<br>DLYQSREERLENQEKAKQTGERNTNYIWNKT |

[0551]

| | | |
|---|---|---|
| | | VDLNLCDGKVTVENVKLKNVGNFIKYEYDQR<br>VQTFLKYEENIKWQAFLIKESKEEENYPYIVER<br>EIEQYEKVRREELLKEVHLIEEYILEKVKDKEIL<br>KKGDNQNFKYYILNGLLKQLKNEDVESYKVF<br>NLNTKPEDVNINQLKQEATDLEQKAFVLTYIR<br>NKFAHNQLPKKEFWDYCQEKYGKIEKEKTYA<br>EYFAEVFKREKEALMK |
| 普雷沃菌属某种 P5-119 (SEQ ID No. 137) | WP_042518169 | MNIPALVENQKKYFGTYSVMAMLNAQTVLDH<br>IQKVADIEGEQNENNENLWFHPVMSHLYNAK<br>NGYDKQPEKTMFIIERLQSYFPFLKIMAENQRE<br>YSNGKYKQNRVEVNSNDIFEVLKRAFGVLKM<br>YRDLTNHYKTYEEKLIDGCEFLTSTEQPLSGMI<br>SKYYTVALRNTKERYGYKTEDLAFIQDNIKKIT<br>KDAYGKRKSQVNTGFFLSLQDYNGDTQKKLH<br>LSGVGIALLICLFLDKQYINIFLSRLPIFSSYNAQ<br>SEERRIIIRSFGINSIKLPKDRIHSEKSNKSVAMD<br>MLNEVKRCPDELFTTLSAEKQSRFRIISDDHNE<br>VLMKRSTDRFVPLLLQYIDYGKLFDHIRFHVN<br>MGKLRYLLKADKTCIDGQTRVRVIEQPLNGFG<br>RLEEAETMRKQENGTFGNSGIRIRDFENVKRD<br>DANPANYPYIVDTYTHYILENNKVEMFISDKGS<br>SAPLLPLIEDDRYVVKTIPSCRMSTLEIPAMAFH<br>MFLFGSKKTEKLIVDVHNRYKRLFQAMQKEE<br>VTAENIASFGIAESDLPQKILDLISGNAHGKDV<br>DAFIRLTVDDMLTDTERRIKRFKDDRKSIRSAD<br>NKMGKRGFKQISTGKLADFLAKDIVLFQPSVN<br>DGENKITGLNYRIMQSAIAVYDSGDDYEAKQQ<br>FKLMFEKARLIGKGTTEPHPFLYKVFARSIPAN<br>AVDFYERYLIERKFYLTGLCNEIKRGNRVDVPF<br>IRRDQNKWKTPAMKTLGRIYSEDLPVELPRQM<br>FDNEIKSHLKSLPQMEGIDFNNANVTYLIAEYM<br>KRVLNDDFQTFYQWKRNYHYMDMLKGEYDR<br>KGSLQHCFTSVEEREGLWKERASRTERYRKLA<br>SNKIRSNRQMRNASSEEIETILDKRLSNCRNEY<br>QKSEKVIRRYRVQDALLFLLAKKTLTELADFD<br>GERFKLKEIMPDAEKGILSEIMPMSFTFEKGGK<br>KYTITSEGMKLKNYGDFFVLASDKRIGNLLEL<br>VGSDIVSKEDIMEEFNKYDQCRPEISSIVFNLEK<br>WAFDTYPELSARVDREEKVDFKSILKILLNNKN<br>INKEQSDILRKIRNAFDHNNYPDKGIVEIKALPE<br>IAMSIKKAFGEYAIMK |
| 普雷沃 | WP_044 | MNIPALVENQKKYFGTYSVMAMLNAQTVLDH<br>IQKVADIEGEQNENNENLWFHPVMSHLYNAK |

[0552]

| | | |
|---|---|---|
| 菌属某种 P4-76 (SEQ ID No. 138) | 72147 | NGYDKQPEKTMFIIERLQSYFPFLKIMAENQRE<br>YSNGKYKQNRVEVNSNDIFEVLKRAFGVLKM<br>YRDQASHYKTYDEKLIDGCEFLTSTEQPLSGMI<br>NNYYTVALRNMNERYGYKTEDLAFIQDKRFK<br>FVKDAYGKKKSQVNTGFFLSLQDYNGDTQKK<br>LHLSGVGIALLICLFLDKQYINIFLSRLPIFSSYN<br>AQSEERRIIIRSFGINSIKQPKDRIHSEKSNKSVA<br>MDMLNEIKRCPNELFETLSAEKQSRFRIISNDH<br>NEVLMKRSSDRFVPLLLQYIDYGKLFDHIRFHV<br>NMGKLRYLLKADKTCIDGQTRVRVIEQPLNGF<br>GRLEEVETMRKQENGTFGNSGIRIRDFENMKR<br>DDANPANYPYIVDTYTHYILENNKVEMFISDEE<br>TPAPLLPVIEDDRYVVKTIPSCRMSTLEIPAMAF<br>HMFLFGSKKTEKLIVDVHNRYKRLFKAMQKE<br>EVTAENIASFGIAESDLPQKIIDLISGNAHGKDV<br>DAFIRLTVDDMLADTERRIKRFKDDRKSIRSAD<br>NKMGKRGFKQISTGKLADFLAKDIVLFQPSVN<br>DGENKITGLNYRIMQSAIAVYNSGDDYEAKQQ<br>FKLMFEKARLIGKGTTEPHPFLYKVFVRSIPAN<br>AVDFYERYLIERKFYLIGLSNEIKKGNRVDVPFI<br>RRDQNKWKTPAMKTLGRIYDEDLPVELPRQM<br>FDNEIKSHLKSLPQMEGIDFNNANVTYLIAEYM<br>KRVLNDDFQTFYQWKRNYRYMDMLRGEYDR<br>KGSLQSCFTSVEEREGLWKERASRTERYRKLA<br>SNKIRSNRQMRNASSEEIETILDKRLSNSRNEY<br>QKSEKVIRRYRVQDALLFLLAKKTLTELADFD<br>GERFKLKEIMPDAEKGILSEIMPMSFTFEKGGK<br>KYTITSEGMKLKNYGDFFVLASDKRIGNLLEL<br>VGSDTVSKEDIMEEFKKYDQCRPEISSIVFNLE<br>KWAFDTYPELSARVDREEKVDFKSILKILLNNK<br>NINKEQSDILRKIRNAFDHNNYPDKGVVEIRAL<br>PEIAMSIKKAFGEYAIMK |
| 普雷沃菌属某种 P5-60 (SEQ ID No. 139) | WP_044074780 | MNIPALVENQKKYFGTYSVMAMLNAQTVLDH<br>IQKVADIEGEQNENNENLWFHPVMSHLYNAK<br>NGYDKQPEKTMFIIERLQSYFPFLKIMAENQRE<br>YSNGKYKQNRVEVNSNDIFEVLKRAFGVLKM<br>YRDLTNHYKTYEEKLIDGCEFLTSTEQPFSGMI<br>SKYYTVALRNTKERYGYKAEDLAFIQDNRYKF<br>TKDAYGKRKSQVNTGSFLSLQDYNGDTTKKL<br>HLSGVGIALLICLFLDKQYINLFLSRLPIFSSYNA<br>QSEERRIIIRSFGINSIKQPKDRIHSEKSNKSVAM<br>DMLNEVKRCPDELFTTLSAEKQSRFRIISDDHN<br>EVLMKRSSDRFVPLLLQYIDYGKLFDHIRFHVN |

[0553]

| | | |
|---|---|---|
| | | MGKLRYLLKADKTCIDGQTRVRVIEQPLNGFG<br>RLEEVETMRKQENGTFGNSGIRIRDFENMKRD<br>DANPANYPYIVETYTHYILENNKVEMFISDEEN<br>PTPLLPVIEDDRYVVKTIPSCRMSTLEIPAMAFH<br>MFLFGSEKTEKLIIDVHDRYKRLFQAMQKEEV<br>TAENIASFGIAESDLPQKIMDLISGNAHGKDVD<br>AFIRLTVDDMLTDTERRIKRFKDDRKSIRSADN<br>KMGKRGFKQISTGKLADFLAKDIVLFQPSVND<br>GENKITGLNYRIMQSAIAVYDSGDDYEAKQQF<br>KLMFEKARLIGKGTTEPHPFLYKVFVRSIPANA<br>VDFYERYLIERKFYLIGLSNEIKKGNRVDVPFIR<br>RDQNKWKTPAMKTLGRIYSEDLPVELPRQMF<br>DNEIKSHLKSLPQMEGIDFNNANVTYLIAEYM<br>KRVLNDDFQTFYQWKRNYRYMDMLRGEYDR<br>KGSLQHCFTSIEEREGLWKERASRTERYRKLAS<br>NKIRSNRQMRNASSEEIETILDKRLSNCRNEYQ<br>KSEKIIRRYRVQDALLFLLAKKTLTELADFDGE<br>RFKLKEIMPDAEKGILSEIMPMSFTFEKGGKIYT<br>ITSGGMKLKNYGDFFVLASDKRIGNLLELVGS<br>NTVSKEDIMEEFKKYDQCRPEISSIVFNLEKWA<br>FDTYPELPARVDRKEKVDFWSILDVLSNNKDI<br>NNEQSYILRKIRNAFDHNNYPDKGIVEIKALPEI<br>AMSIKKAFGEYAIMK |
| 厦门褐指藻杆菌(Phaeodactylibacter xiamenensis)(SEQ ID No. 140) | WP_044218239 | MTNTPKRRTLHRHPSYFGAFLNIARHNAFMIM<br>EHLSTKYDMEDKNTLDEAQLPNAKLFGCLKK<br>RYGKPDVTEGVSRDLRRYFPFLNYPLFLHLEK<br>QQNAEQAATYDINPEDIEFTLKGFFRLLNQMR<br>NNYSHYISNTDYGKFDKLPVQDIYEAAIFRLLD<br>RGKHTKRFDVFESKHTRHLESNNSEYRPRSLA<br>NSPDHENTVAFVTCLFLERKYAFPFLSRLDCFR<br>STNDAAEGDPLIRKASHECYTMFCCRLPQPKLE<br>SSDILLDMVNELGRCPSALYNLLSEEDQARFHI<br>KREEITGFEEDPDEELEQEIVLKRHSDRFPYFAL<br>RYFDDTEAFQTLRFDVYLGRWRTKPVYKKRIY<br>GQERDRVLTQSIRTFTRLSRLLPIYENVKHDAV<br>RQNEEDGKLVNPDVTSQFHKSWIQIESDDRAF<br>LSDRIEHFSPHYNFGDQVIGLKFINPDRYAAIQN<br>VFPKLPGEEKKDKDAKLVNETADAIISTHEIRS<br>LFLYHYLSKKPISAGDERRFIQVDTETFIKQYID<br>TIKLFFEDIKSGELQPIADPPNYQKNEPLPYVRG<br>DKEKTQEERAQYRERQKEIKERRKELNTLLQN<br>RYGLSIQYIPSRLREYLLGYKKVPYEKLALQKL<br>RAQRKEVKKRIKDIEKMRTPRVGEQATWLAE |

[0554]

| | | |
|---|---|---|
| | | DIVFLTPPKMHTPERKTTKHPQKLNNDQFRIM<br>QSSLAYFSVNKKAIKKFFQKETGIGLSNRETSH<br>PFLYRIDVGRCRGILDFYTGYLKYKMDWLDDA<br>IKKVDNRKHGKKEAKKYEKYLPSSIQHKTPLE<br>LDYTRLPVYLPRGLFKKAIVKALAAHADFQVE<br>PEEDNVIFCLDQLLDGDTQDFYNWQRYYRSAL<br>TEKETDNQLVLAHPYAEQILGTIKTLEGKQKN<br>NKLGNKAKQKIKDELIDLKRAKRRLLDREQYL<br>RAVQAEDRALWLMIQERQKQKAEHEEIAFDQ<br>LDLKNITKILTESIDARLRIPDTKVDITDKLPLRR<br>YGDLRRVAKDRRLVNLASYYHVAGLSEIPYDL<br>VKKELEEYDRRRVAFFEHVYQFEKEVYDRYA<br>AELRNENPKGESTYFSHWEYVAVAVKHSADT<br>HFNELFKEKVMQLRNKFHHNEFPYFDWLLPEV<br>EKASAALYADRVFDVAEGYYQKMRKLMRQ |
| 黄杆菌属某种(Flavobacterium sp.) 316 (SEQ ID No. 141) | WP_045968377 | MDNNITVEKTELGLGITYNHDKVEDKHYFGGF<br>FNLAQNNIDLVAQEFKKRLLIQGKDSINIFANY<br>FSDQCSITNLERGIKILAEYFPVVSYIDLDEKNK<br>SKSIREHLILLLETINNLRNYYTHYYHKKIIIDGS<br>LFPLLDTILLKVVLEIKKKKLKEDKTKQLLKKG<br>LEKEMTILFNLMKAEQKEKKIKGWNIDENIKG<br>AVLNRAFSHLLYNDELSDYRKSKYNTEDETLK<br>DTLTESGILFLLSFFLNKKEQEQLKANIKGYKG<br>KIASIPDEEITLKNNSLRNMATHWTYSHLTYKG<br>LKHRIKTDHEKETLLVNMVDYLSKVPHEIYQN<br>LSEQNKSLFLEDINEYMRDNEENHDSSEASRVI<br>HPVIRKRYENKFAYFAIRFLDEFAEFPTLRFMV<br>NVGNYIHDNRKKDIGGTSLITNRTIKQQINVFG<br>NLTEIHKKKNDYFEKEENKEKTLEWELFPNPS<br>YHFQKENIPIFIDLEKSKETNDLAKEYAKEKKKI<br>FGSSRKKQQNTAKKNRETIINLVFDKYKTSDR<br>KTVTFEQPTALLSFNELNSFLYAFLVENKTGKE<br>LEKIIIEKIANQYQILKNCSSTVDKTNDNIPKSIK<br>KIVNTTTDSFYFEGKKIDIEKLEKDITIEIEKTNE<br>KLETIKENEESAQNYKRNERNTQKRKLYRKYV<br>FFTNEIGIEATWITNDILRFLDNKENWKGYQHS<br>ELQKFISQYDNYKKEALGLLESEWNLESDAFF<br>GQNLKRMFQSNSTFETFYKKYLDNRKNTLETY<br>LSAIENLKTMTDVRPKVLKKKWTELFRFFDKK<br>IYLLSTIETKINELITKPINLSRGIFEEKPTFINGK<br>NPNKENNQHLFANWFIYAKKQTILQDFYNLPL<br>EQPKAITNLKKHKYKLERSINNLKIEDIYIKQM<br>VDFLYQKLFEQSFIGSLQDLYTSKEKREIEKGK |

[0555]

| | | |
|---|---|---|
| | | AKNEQTPDESFIWKKQVEINTHNGRIIAKTKIK<br>DIGKFKNLLTDNKIAHLISYDDRIWDFSLNNDG<br>DITKKLYSINTELESYETIRREKLLKQIQQFEQF<br>LLEQETEYSAERKHPEKFEKDCNPNFKKYIIEG<br>VLNKIIPNHEIEEIEILKSKEDVFKINFSDILILNN<br>DNIKKGYLLIMIRNKFAHNQLIDKNLFNFSLQL<br>YSKNENENFSEYLNKVCQNIIQEFKEKLK |
| 动物口腔卟啉单胞菌(SEQ ID No. 142) | WP_046201018 | MTEQSERPYNGTYYTLEDKHFWAAFLNLARH<br>NAYITLTHIDRQLAYSKADITNDQDVLSFKAL<br>WKNFDNDLERKSRLRSLILKHFSFLEGAAYGK<br>KLFESKSSGNKSSKNKELTKKEKEELQANALSL<br>DNLKSILFDFLQKLKDFRNYYSHYRHSESSELP<br>LFDGNMLQRLYNVFDVSVQRVKRDHEHNDKV<br>DPHRHFNHLVRKGKKDRYGHNDNPSFKHHFV<br>DSEGMVTEAGLLFFVSLFLEKRDAIWMQKKIR<br>GFKGGTETYQQMTNEVFCRSRISLPKLKLESLR<br>TDDWMLLDMLNELVRCPKPLYDRLREKDRAR<br>FRVPVDILPDEDDTDGGGEDPFKNTLVRHQDR<br>FPYFALRYFDLKKVFTSLRFHIDLGTYHFAIYK<br>KMIGEQPEDRHLTRNLYGFGRIQDFAEEHRPEE<br>WKRLVRDLDYFETGDKPYISQTTPHYHIEKGKI<br>GLRFMPEGQHLWPSPEVGTTRTGRSKYAQDK<br>RLTAEAFLSVHELMPMMFYYFLLREKYSEEVS<br>AEKVQGRIKRVIEDVYAIYDAFARDEINTLKEL<br>DACLADKGIRRGHLPKQMIAILSQEHKDMEEKI<br>RKKLQEMIADTDHRLDMLDRQTDRKIRIGRKN<br>AGLPKSGVIADWLVRDMMRFQPVAKDTSGKP<br>LNNSKANSTEYRMLQRALALFGGEKKRLTPYF<br>RQMNLTGGNNPHPFLHETRWESHTNILSFYRS<br>YLRARKAFLERIGRSDRMENRPFLLLKEPKTDR<br>QTLVAGWKSEFHLPRGIFTEAVRDCLIEMGYD<br>EVGSYREVGFMAKAVPLYFERACEDRVQPFY<br>DSPFNVGNSLKPKKGRFLSKEERAEEWERGKE<br>RFRDLEAWSHSAARRIEDAFAGIEYASPGNKK<br>KIEQLLRDLSLWEAFESKLKVRADKINLAKLK<br>KEILEAQEHPYHDFKSWQKFERELRLVKNQDII<br>TWMMCRDLMEENKVEGLDTGTLYLKDIRPNV<br>QEQGSLNVLNRVKPMRLPVVVYRADSRGHVH<br>KEEAPLATVYIEERDTKLLKQGNFKSFVKDRR<br>LNGLFSFVDTGGLAMEQYPISKLRVEYELAKY<br>QTARVCVFELTLRLEESLLTRYPHLPDESFRKM<br>LESWSDPLLAKWPELHGKVRLLIAVRNAFSHN<br>QYPMYDEAVFSSIRKYDPSSPDAIEERMGLNIA |

[0556]

| | | |
|---|---|---|
| | | HRLSEEVKQAKETVERIIQV |
| WP_047431796 (SEQ ID No. 143) | 金黄杆菌属某种 (Chryseobacteriu m sp.) YR477 | METQTIGHGIAYDHSKIQDKHFFGGFLNLAEN<br>NIKAVLKAFSEKFNVGNVDVKQFADVSLKDNL<br>PDNDFQKRVSFLKMYFPVVDFINIPNNRAKFRS<br>DLTTLFKSVDQLRNFYTHYYHKPLDFDASLFIL<br>LDDIFARTAKEVRDQKMKDDKTRQLLSKSLSE<br>ELQKGYELQLERLKELNRLGKKVNIHDQLGIK<br>NGVLNNAFNHLIYKDGESFKTKLTYSSALTSFE<br>SAENGIEISQSGLLFLLSMFLKRKEIEDLKNRNK<br>GFKAKVVIDEDGKVNGLKFMATHWVFSYLCF<br>KGLKSKLSTEFHEETLLIQIIDELSKVPDELYCA<br>FDKETRDKFIEDINEYVKEGHQDFSLEDAKVIH<br>PVIRKRYENKFNYFAIRFLDEFVKFPSLRFQVH<br>VGNYVHDRRIKNIDGTTFETERVVKDRIKVFG<br>RLSEISSYKAQYLSSVSDKHDETGWEIFPNPSY<br>VFINNNIPIHISVDTSFKKEIADFKKLRRAQVPD<br>ELKIRGAEKKRKFEITQMIGSKSVLNQEEPIALL<br>SLNEIPALLYEILINGKEPAEIERIIKDKLNERQD<br>VIKNYNPENWLPASQISRRLRSNKGERIINTDK<br>LLQLVTKELLVTEQKLKIISDNREALKQKKEGK<br>YIRKFIFTNSELGREAIWLADDIKRFMPADVRK<br>EWKGYQHSQLQQSLAFYNSRPKEALAILESSW<br>NLKDEKIIWNEWILKSFTQNKFFDAFYNEYLK<br>GRKKYFAFLSEHIVQYTSNAKNLQKFIKQQMP<br>KDLFEKRHYIIEDLQTEKNKILSKPFIFPRGIFDK<br>KPTFIKGVKVEDSPESFANWYQYGYQKDHQFQ<br>KFYDWKRDYSDVFLEHLGKPFINNGDRRTLG<br>MEELKERIIIKQDLKIKKIKIQDLFLRLIAENLFQ<br>KVFKYSAKLPLSDFYLTQEERMEKENMAALQ<br>NVREEGDKSPNIIKDNFIWSKMIPYKKGQIIENA<br>VKLKDIGKLNVLSLDDKVQTLLSYDDAKPWS<br>KIALENEFSIGENSYEVIRREKLFKEIQQFESEIL<br>FRSGWDGINHPAQLEDNRNPKFKMYIVNGILR<br>KSAGLYSQGEDIWFEYNADFNNLDADVLETKS<br>ELVQLAFLVTAIRNKFAHNQLPAKEFYFYIRAK<br>YGFADEPSVALVYLNFTKYAINEFKKVMI |
| 鸭疫里默氏杆菌 (SEQ ID No. 144) | WP_049354263 | MFFSFHNAQRVIFKHLYKAFDASLRMVKEDYK<br>AHFTVNLTRDFAHLNRKGKNKQDNPDFNRYR<br>FEKDGFFTESGLLFFTNLFLDKRDAYWMLKKV<br>SGFKASHKQREKMTTEVFCRSRILLPKLRLESR<br>YDHNQMLLDMLSELSRCPKLLYEKLSEENKKH<br>FQVEADGFLDEIEEEQNPFKDTLIRHQDRFPYF<br>ALRYLDLNESFKSIRFQVDLGTYHYCIYDKKIG |

[0557]

| | | |
|---|---|---|
| | | DEQEKRHLTRTLLSFGRLQDFTEINRPQEWKAL<br>TKDLDYKETSNQPFISKTTPHYHITDNKIGFRLG<br>TSKELYPSLEIKDGANRIAKYPYNSGFVAHAFIS<br>VHELLPLMFYQHLTGKSEDLLKETVRHIQRIYK<br>DFEEERINTIEDLEKANQGRLPLGAFPKQMLGL<br>LQNKQPDLSEKAKIKIEKLIAETKLLSHRLNTK<br>LKSSPKLGKRREKLIKTGVLADWLVKDFMRFQ<br>PVAYDAQNQPIKSSKANSTEFWFIRRALALYG<br>GEKNRLEGYFKQTNLIGNTNPHPFLNKFNWKA<br>CRNLVDFYQQYLEQREKFLEAIKNQPWEPYQY<br>CLLLKIPKENRKNLVKGWEQGGISLPRGLFTEA<br>IRETLSEDLMLSKPIRKEIKKHGRVGFISRAITLY<br>FKEKYQDKHQSFYNLSYKLEAKAPLLKREEHY<br>EYWQQNKPQSPTESQRLELHTSDRWKDYLLY<br>KRWQHLEKKLRLYRNQDVMLWLMTLELTKN<br>HFKELNLNYHQLKLENLAVNVQEADAKLNPL<br>NQTLPMVLPVKVYPATAFGEVQYHKTPIRTVY<br>IREEHTKALKMGNFKALVKDRRLNGLFSFIKEE<br>NDTQKHPISQLRLRRELEIYQSLRVDAFKETLS<br>LEEKLLNKHTSLSSLENEFRALLEEWKKEYAA<br>SSMVTDEHIAFIASVRNAFCHNQYPFYKEALH<br>APIPLFTVAQPTTEEKDGLGIAEALLKVLREYC<br>EIVKSQI |
| 牙龈卟啉单胞菌 (SEQ ID No. 145) | WP_052912312 | MTEQNEKPYNGTYYTLEDKHFWAAFFNLARH<br>NAYITLAHIDRQLAYSKADITNDEDILFFKGQW<br>KNLDNDLERKARLRSLILKHFSFLEGAAYGKK<br>LFESQSSGNKSSKKKELTKKEKEELQANALSLD<br>NLKSILFDFLQKLKDFRNYYSHYRHPESSELPL<br>FDGNMLQRLYNVFDVSVQRVKRDHEHNDKV<br>DPHRHFNHLVRKGKKDKYGNNDNPFFKHHFV<br>DREEKVTEAGLLFFVSLFLEKRDAIWMQKKIR<br>GFKGGTEAYQQMTNEVFCRSRISLPKLKLESLR<br>TDDWMLLDMLNELVRCPKLLYDRLREEDRAR<br>FRVPVDILSDEDDTDGTEEDPFKNTLVRHQDRF<br>PYFALRYFDLKKVFTSLRFHIDLGTYHFAIYKK<br>NIGEQPEDRHLTRNLYGFGRIQDFAEEHRPEEW<br>KRLVRDLDYFETGDKPYITQTTPHYHIEKGKIG<br>LRFVPEGQLLWPSPEVGATRTGRSKYAQDKRF<br>TAEAFLSVHELMPMMFYYFLLREKYSEEASAE<br>KVQGRIKRVIEDVYAVYDAFARDEINTRDELD<br>ACLADKGIRRGHLPRQMIAILSQEHKDMEEKV<br>RKKLQEMIADTDHRLDMLDRQTDRKIRIGRKN<br>AGLPKSGVIADWLVRDMMRFQPVAKDTSGKP |

[0558]

| | | |
|---|---|---|
| | | LNNSKANSTEYRMLQRALALFGGEKERLTPYF<br>RQMNLTGGNNPHPFLHETRWESHTNILSFYRS<br>YLKARKAFLQSIGRSDREENHRFLLLKEPKTDR<br>QTLVAGWKSEFHLPRGIFTEAVRDCLIEMGYD<br>EVGSYKEVGFMAKAVPLYFERACKDRVQPFY<br>DYPFNVGNSLKPKKGRFLSKEKRAEEWESGKE<br>RFRDLEAWSHSAARRIEDAFVGIEYASWENKK<br>KIEQLLQDLSLWETFESKLKVKADKINIAKLKK<br>EILEAKEHPYHDFKSWQKFERELRLVKNQDIIT<br>WMMCRDLMEENKVEGLDTGTLYLKDIRTDVQ<br>EQGSLNVLNHVKPMRLPVVVYRADSRGHVHK<br>EEAPLATVYIEERDTKLLKQGNFKSFVKDRRL<br>NGLFSFVDTGALAMEQYPISKLRVEYELAKYQ<br>TARVCAFEQTLELEESLLTRYPHLPDESFREML<br>ESWSDPLLDKWPDLQREVRLLIAVRNAFSHNQ<br>YPMYDETIFSSIRKYDPSSLDAIEERMGLNIAHR<br>LSEEVKLAKEMVERIIQA |
| 牙龈卟啉单胞菌 (SEQ ID No. 146) | WP_058019250 | MTEQNEKPYNGTYYTLKDKHFWAAFFNLARH<br>NAYITLTHIDRQLAYSKADITNDEDILFFKGQW<br>KNLDNDLERKARLRSLILKHFSFLEGAAYGKK<br>LFESQSSGNKSSKKKELTKKEKEELQANALSLD<br>NLKSILFDFLQKLKDFRNYYSHYRHPESSELPM<br>FDGNMLQRLYNVFDVSVQRVKRDHEHNDKV<br>DPHRHFNHLVRKGKKDRCGNNDNPFFKHHFV<br>DREGKVTEAGLLFFVSLFLEKRDAIWMQKKIR<br>GFKGGTETYQQMTNEVFCRSRISLPKLKLESLR<br>TDDWMLLDMLNELVRCPKSLYDRLREEDRAC<br>FRVPVDILSDEDDTDGAEEDPFKNTLVRHQDR<br>FPYFALRYFDLKKVFTSLRFHIDLGTYHFAIYK<br>KNIGEQPEDRHLTRNLYGFGRIQDFAEEHRPEE<br>WKRLVRDLDCFETGDKPYITQTTPHYHIEKGKI<br>GLRFVPEGQHLWPSPEVGATRTGRSKYAQDKR<br>FTAEAFLSVHELMPMMFYYFLLREKYSEEVSA<br>ERVQGRIKRVIEDVYAVYDAFARDEINTRDEL<br>DACLADKGIRRGHLPRQMIAILSQKHKDMEEK<br>VRKKLQEMIADTDHRLDMLDRQTDRKIRIGRK<br>NAGLPKSGVIADWLVRDMMRFQPVAKDTSGK<br>PLNNSKANSTEYRMLQRALALFGGEKERLTPY<br>FRQMNLTGGNNPHPFLHETRWESHTNILSFYRS<br>YLKARKAFLQSIGRSDRVENHRFLLLKEPKTDR<br>QTLVAGWKGEFHLPRGIFTEAVRDCLIEMGLD<br>EVGSYKEVGFMAKAVPLYFERACKDRVQPFY<br>DYPFNVGNSLKPKKGRFLSKEKRAEEWESGKE |

| | | |
|---|---|---|
| | | RFRDLEAWSHSAARRIEDAFAGIENASRENKK<br>KIEQLLQDLSLWETFESKLKVKADKINIAKLKK<br>EILEAKEHPYLDFKSWQKFERELRLVKNQDIIT<br>WMMCRDLMEENKVEGLDTGTLYLKDIRTDVQ<br>EQGSLNVLNHVKPMRLPVVVYRADSRGHVHK<br>EQAPLATVYIEERDTKLLKQGNFKSFVKDRRL<br>NGLFSFVDTGALAMEQYPISKLRVEYELAKYQ<br>TARVCAFEQTLELEESLLTRYPHLPDENFRKML<br>ESWSDPLLDKWPDLHRKVRLLIAVRNAFSHNQ<br>YPMYDEAVFSSIRKYDPSSPDAIEERMGLNIAH<br>RLSEEVKQAKEMAERIIQA |
| [0559] | 柱状黄杆菌(SEQ ID No. 147) | WP_060381855 | MSSKNESYNKQKTFNHYKQEDKYFFGGFLNN<br>ADDNLRQVGKEFKTRINFNHNNNELASVFKDY<br>FNKEKSVAKREHALNLLSNYFPVLERIQKHTN<br>HNFEQTREIFELLLDTIKKLRDYYTHHYHKPITI<br>NPKVYDFLDDTLLDVLITIKKKKVKNDTSRELL<br>KEKFRPELTQLKNQKREELIKKGKKLLEENLEN<br>AVFNHCLRPFLEENKTDDKQNKTVSLRKYRKS<br>KPNEETSITLTQSGLVFLISFFLHRKEFQVFTSGL<br>EGFKAKVNTIKEEEISLNKNNIVYMITHWSYSY<br>YNFKGLKHRIKTDQGVSTLEQNNTTHSLTNTN<br>TKEALLTQIVDYLSKVPNEIYETLSEKQQKEFE<br>EDINEYMRENPENEDSTFSSIVSHKVIRKRYEN<br>KFNYFAMRFLDEYAELPTLRFMVNFGDYIKDR<br>QKKILESIQFDSERIIKKEIHLFEKLGLVTEYKK<br>NVYLKETSNIDLSRFPLFPSPSYVMANNNIPFYI<br>DSRSNNLDEYLNQKKKAQSQNRKRNLTFEKY<br>NKEQSKDAIIAMLQKEIGVKDLQQRSTIGLLSC<br>NELPSMLYEVIVKDIKGAELENKIAQKIREQYQ<br>SIRDFTLDSPQKDNIPTTLTKTISTDTSVTFENQP<br>IDIPRLKNALQKELTLTQEKLLNVKQHEIEVDN<br>YNRNKNTYKFKNQPKDKVDDNKLQRKYVFY<br>RNEIGQEANWLASDLIHFMKNKSLWKGYMHN<br>ELQSFLAFFEDKKNDCIALLETVFNLKEDCILT<br>KDLKNLFLKHGNFIDFYKEYLKLKEDFLNTEST<br>FLENGFIGLPPKILKKELSKRLNYIFIVFQKRQFII<br>KELEEKKNNLYADAINLSRGIFDEKPTMIPFKK<br>PNPDEFASWFVASYQYNNYQSFYELTPDKIEN<br>DKKKKYKNLRAINKVKIQDYYLKLMVDTLYQ<br>DLFNQPLDKSLSDFYVSKTDREKIKADAKAYQ<br>KRNDSFLWNKVIHLSLQNNRITANPKLKDIGK<br>YKRALQDEKIATLLTYDDRTWTYALQKPEKEN<br>ENDYKELHYTALNMELQEYEKVRSKKLLKQV |

[0560]

| | | |
|---|---|---|
| | | QELEKQILDKFYDFSNNATHPEDLEIEDKKGKR HPNFKLYITKALLKNESEIINLENIDIEILIKYYD YNTEKLKEKIKNMDEDEKAKIVNTKENYNKIT NVLIKKALVLIIIRNKMAHNQYPPKFIYDLATRF VPKKEEEYFACYFNRVFETITTELWENKKKAK EIV |
| 牙龈卟啉单胞菌 (SEQ ID No. 148) | WP_061156470 | MTEQNERPYNGTYYTLEDKHFWAAFFNLARH NAYITLTHIDRQLAYSKADITNDEDILFFKGQW KNLDNDLERKARLRSLILKHFSFLEGAAYGKK LFENKSSGNKSSKKKELTKKEKEELQANALSL DNLKSILFDFLQKLKDFRNYYSHYRHPESSELP LFDGNMLQRLYNVFDVSVQRVKRDHEHNDKV DPHRHFNHLVRKGKKDRCGNNDNPFFKHHFV DREGKVTEAGLLFFVSLFLEKRDAIWMQKKIR GFKGGTEAYQQMTNEVFCRSRISLPKLKLESLR TDDWMLLDMLNELVRCPKSLYDRLREEDRAR FRVPVDILSDEDDTDGTEEDPFKNTLVRHQDRF PYFALRYFDLKKVFTSLRFHIDLGTYHFAIYKK NIGEQPEDRHLTRNLYGFGRIQDFAEEHRPEEW KRLVRDLDYFETGDKPYITQTTPHYHIEKGKIG LRFVPEGQHLWPSPEVGATRTGRSKYAQDKRL TAEAFLSVHELMPMMFYYFLLREKYSEEVSAE KVQGRIKRVIEDVYAVYDAFARGEIDTLDRLD ACLADKGIRRGHLPRQMIAILSQEHKDMEEKV RKKLQEMIADTDHRLDMLDRQTDRKIRIGRKN AGLPKSGVIADWLVRDMMRFQPVAKDTSGKP LNNSKANSTEYRMLQRALALFGGEKERLTPYF RQMNLTGGNNPHPFLHETRWESHTNILSFYRS YLKARKAFLQSIGRSDREENHRFLLLKEPKTDR QTLVAGWKSEFHLPRGIFTEAVRDCLIEMGYD EVGSYKEVGFMAKAVPLYFERACKDRVQPFY DYPFNVGNSLKPKKGRFLSKEKRAEEWESGKE RFRLAKLKKEILEAKEHPYLDFKSWQKFEREL RLVKNQDIITWMMCRDLMEENKVEGLDTGTL YLKDIRTEVQEQGSLNVLNRVKPMRLPVVVYR ADSRGHVHKEQAPLATVYIEERDTKLLKQGNF KSFVKDRRLNGLFSFVDTGGLAMEQYPISKLR VEYELAKYQTARVCAFEQTLELEESLLTRCPHL PDKNFRKMLESWSDPLLDKWPDLQREVWLLI AVRNAFSHNQYPMYDEAVFSSIRKYDPSSPDAI EERMGLNIAHRLSEEVKQAKEMAERIIQA |
| 牙龈卟 | WP_0611 | MNTVPASENKGQSRTVEDDPQYFGLYLNLARE NLIEVESHVRIKFGKKKLNEESLKQSLLCDHLL |

[0561]

| | | |
|---|---|---|
| 啉单胞菌 (SEQ ID No. 149) | 56637 | SVDRWTKVYGHSRRYLPFLHYFDPDSQIEKDH<br>DSKTGVDPDSAQRLIRELYSLLDFLRNDFSHNR<br>LDGTTFEHLEVSPDISSFITGTYSLACGRAQSRF<br>ADFFKPDDFVLAKNRKEQLISVADGKECLTVS<br>GLAFFICLFLDREQASGMLSRIRGFKRTDENWA<br>RAVHETFCDLCIRHPHDRLESSNTKEALLLDML<br>NELNRCPRILYDMLPEEERAQFLPALDENSMN<br>NLSENSLNEESRLLWDGSSDWAEALTKRIRHQ<br>DRFPYLMLRFIEEMDLLKGIRFRVDLGEIELDS<br>YSKKVGRNGEYDRTITDHALAFGKLSDFQNEE<br>EVSRMISGEASYPVRFSLFAPRYAIYDNKIGYC<br>HTSDPVYPKSKTGEKRALSNPQSMGFISVHDLR<br>KLLLMELLCEGSFSRMQSGFLRKANRILDETAE<br>GKLQFSALFPEMRHRFIPPQNPKSKDRREKAET<br>TLEKYKQEIKGRKDKLNSQLLSAFDMNQRQLP<br>SRLLDEWMNIRPASHSVKLRTYVKQLNEDCRL<br>RLRKFRKDGDGKARAIPLVGEMATFLSQDIVR<br>MIISEETKKLITSAYYNEMQRSLAQYAGEENRR<br>QFRAIVAELHLLDPSSGHPFLSATMETAHRYTE<br>DFYKCYLEKKREWLAKTFYRPEQDENTKRRIS<br>VFFVPDGEARKLLPLLIRRRMKEQNDLQDWIR<br>NKQAHPIDLPSHLFDSKIMELLKVKDGKKKWN<br>EAFKDWWSTKYPDGMQPFYGLRRELNIHGKS<br>VSYIPSDGKKFADCYTHLMEKTVQDKKRELRT<br>AGKPVPPDLAADIKRSFHRAVNEREFMLRLVQ<br>EDDRLMLMAINKMMTDREEDILPGLKNIDSIL<br>DKENQFSLAVHAKVLEKEGEGGDNSLSLVPAT<br>IEIKSKRKDWSKYIRYRYDRRVPGLMSHFPEHK<br>ATLDEVKTLLGEYDRCRIKIFDWAFALEGAIMS<br>DRDLKPYLHESSSREGKSGEHSTLVKMLVEKK<br>GCLTPDESQYLILIRNKAAHNQFPCAAEMPLIY<br>RDVSAKVGSIEGSSAKDLPEGSSLVDSLWKKY<br>EMIIRKILPILDPENRFFGKLLNNMSQPINDL |
| 鸭疫里默氏杆菌 (SEQ ID No. 150) | WP_0617101138 | MFFSFHNAQRVIFKHLYKAFDASLRMVKEDYK<br>AHFTVNLTRDFAHLNRKGKNKQDNPDFNRYR<br>FEKDGFFTESGLLFFTNLFLDKRDAYWMLKKV<br>SGFKASHKQSEKMTTEVFCRSRILLPKLRLESR<br>YDHNQMLLDMLSELSRCPKLLYEKLSEKDKK<br>CFQVEADGFLDEIEEEQNPFKDTLIRHQDRFPY<br>FALRYLDLNESFKSIRFQVDLGTYHYCIYDKKI<br>GYEQEKRHLTRTLLNFGRLQDFTEINRPQEWK<br>ALTKDLDYNETSNQPFISKTTPHYHITDNKIGFR<br>LRTSKELYPSLEVKDGANRIAKYPYNSDFVAH |

| | | |
|---|---|---|
| | | AFISISVHELLPLMFYQHLTGKSEDLLKETVRHI<br>QRIYKDFEEERINTIEDLEKANQGRLPLGAFPK<br>QMLGLLQNKQPDLSEKAKIKIEKLIAETKLLSH<br>RLNTKLKSSPKLGKRREKLIKTGVLADWLVKD<br>FMRFQPVVYDAQNQPIKSSKANSTESRLIRRAL<br>ALYGGEKNRLEGYFKQTNLIGNTNPHPFLNKF<br>NWKACRNLVDFYQQYLEQREKFLEAIKHQPW<br>EPYQYCLLLKVPKENRKNLVKGWEQGGISLPR<br>GLFTEAIRETLSKDLTLSKPIRKEIKKHGRVGFIS<br>RAITLYFKEKYQDKHQSFYNLSYKLEAKAPLL<br>KKEEHYEYWQQNKPQSPTESQRLELHTSDRW<br>KDYLLYKRWQHLEKKLRLYRNQDIMLWLMT<br>LELTKNHFKELNLNYHQLKLENLAVNVQEAD<br>AKLNPLNQTLPMVLPVKVYPTTAFGEVQYHET<br>PIRTVYIREEQTKALKMGNFKALVKDRHLNGL<br>FSFIKEENDTQKHPISQLRLRRELEIYQSLRVDA<br>FKETLSLEEKLLNKHASLSSLENEFRTLLEEWK<br>KKYAASSMVTDKHIAFIASVRNAFCHNQYPFY<br>KETLHAPILLFTVAQPTTEEKDGLGIAEALLRV<br>LREYCEIVKSQI |
| [0562] 柱状黄杆菌(SEQ ID No. 151) | WP_063744070 | MSSKNESYNKQKTFNHYKQEDKYFFGGFLNN<br>ADDNLRQVGKEFKTRINFNHNNNELASVFKDY<br>FNKEKSVAKREHALNLLSNYFPVLERIQKHTN<br>HNFEQTREIFELLLDTIKKLRDYYTHHYHKPITI<br>NPKIYDFLDDTLLDVLITIKKKKVKNDTSRELL<br>KEKLRPELTQLKNQKREELIKKGKKLLEENLE<br>NAVFNHCLRPFLEENKTDDKQNKTVSLRKYRK<br>SKPNEETSITLTQSGLVFLMSFFLHRKEFQVFTS<br>GLEGFKAKVNTIKEEKISLNKNNIVYMITHWSY<br>SYYNFKGLKHRIKTDQGVSTLEQNNTTHSLTN<br>TNTKEALLTQIVDYLSKVPNEIYETLSEKQQKE<br>FEEDINEYMRENPENEDSTFSSIVSHKVIRKRYE<br>NKFNYFAMRFLDEYAELPTLRFMVNFGDYIKD<br>RQKKILESIQFDSERIIKKEIHLFEKLGLVTEYKK<br>NVYLKETSNIDLSRFPLFPSPSYVMANNNIPFYI<br>DSRSNNLDEYLNQKKKAQSQNRKRNLTFEKY<br>NKEQSKDAIIAMLQKEIGVKDLQQRSTIGLLSC<br>NELPSMLYEVIVKDIKGAELENKIAQKIREQYQ<br>SIRDFTLNSPQKDNIPTTLIKTISTDTSVTFENQPI<br>DIPRLKNAIQKELALTQEKLLNVKQHEIEVNNY<br>NRNKNTYKFKNQPKDKVDDNKLQRKYVFYR<br>NEIGQEANWLASDLIHFMKNKSLWKGYMHNE<br>LQSFLAFFEDKKNDCIALLETVFNLKEDCILTK |

[0563]

| | | |
|---|---|---|
| | | DLKNLFLKHGNFIDFYKEYLKLKEDFLNTESTF<br>LENGFIGLPPKILKKELSKRLNYIFIVFQKRQFII<br>KELEEKKNNLYADAINLSRGIFDEKPTMIPFKK<br>PNPDEFASWFVASYQYNNYQSFYELTPDKIEN<br>DKKKKYKNLRAINKVKIQDYYLKLMVDTLYQ<br>DLFNQPLDKSLSDFYVSKTDREKIKADAKAYQ<br>KRNDSFLWNKVIHLSLQNNRITANPKLKDIGK<br>YKRALQDEKIATLLTYDDRTWTYALQKPEKEN<br>ENDYKELHYTALNMELQEYEKVRSKKLLKQV<br>QELEKQILDKFYDFSNNATHPEDLEIEDKKGKR<br>HPNFKLYITKALLKNESEIINLENIDIEILIKYYD<br>YNTEKLKEKIKNMDEDEKAKIVNTKENYNKIT<br>NVLIKKALVLIIIRNKMAHNQYPPKFIYDLATRF<br>VPKKEEEYFACYFNRVFETITTELWENKKKAK<br>EIV |
| 鸭疫里默氏杆菌 (SEQ ID No. 152) | WP_<br>0649<br>7088<br>7 | MEKPLPPNVYTLKHKFFWGAFLNIARHNAFITI<br>CHINEQLGLTTPPNDDKIADVVCGTWNNILNN<br>DHDLLKKSQLTELILKHFPFLAAMCYHPPKKE<br>GKKKGSQKEQQKEKENEAQSQAEALNPSELIK<br>VLKTLVKQLRTLRNYYSHHSHKKPDAEKDIFK<br>HLYKAFDASLRMVKEDYKAHFTVNLTQDFAH<br>LNRKGKNKQDNPDFDRYRFEKDGFFTESGLLF<br>FTNLFLDKRDAYWMLKKVSGFKASHKQSEKM<br>TTEVFCRSRILLPKLRLESRYDHNQMLLDMLSE<br>LSRYPKLLYEKLSEEDKKRFQVEADGFLDEIEE<br>EQNPFKDTLIRHQDRFPYFALRYLDLNESFKSIR<br>FQVDLGTYHYCIYDKKIGDEQEKRHLTRTLLSF<br>GRLQDFTEINRPQEWKALTKDLDYKETSKQPFI<br>SKTTPHYHITDNKIGFRLGTSKELYPSLEVKDG<br>ANRIAQYPYNSDFVAHAFISVHELLPLMFYQHL<br>TGKSEDLLKETVRHIQRIYKDFEEERINTIEDLE<br>KANQGRLPLGAFPKQMLGLLQNKQPDLSEKA<br>KIKIEKLIAETKLLSHRLNTKLKSSPKLGKRREK<br>LIKTGVLADWLVKDFMRFQPVAYDAQNQPIES<br>SKANSTEFQLIQRALALYGGEKNRLEGYFKQT<br>NLIGNTNPHPFLNKFNWKACRNLVDFYQQYLE<br>QREKFLEAIKNQPWEPYQYCLLLKIPKENRKNL<br>VKGWEQGGISLPRGLFTEAIRETLSKDLTLSKPI<br>RKEIKKHGRVGFISRAITLYFREKYQDDHQSFY<br>DLPYKLEAKASPLPKKEHYEYWQQNKPQSPTE<br>LQRLELHTSDRWKDYLLYKRWQHLEKKLRLY<br>RNQDVMLWLMTLELTKNHFKELNLNYHQLKL<br>ENLAVNVQEADAKLNPLNQTLPMVLPVKVYP |

[0564]

| | | |
|---|---|---|
| | | ATAFGEVQYQETPIRTVYIREEQTKALKMGNF<br>KALVKDRRLNGLFSFIKEENDTQKHPISQLRLR<br>RELEIYQSLRVDAFKETLNLEEKLLKKHTSLSS<br>VENKFRILLEEWKKEYAASSMVTDEHIAFIASV<br>RNAFCHNQYPFYEEALHAPIPLFTVAQQTTEEK<br>DGLGIAEALLRVLREYCEIVKSQI |
| 海洋中华微杆菌(Sinomicrobium oceani)(SEQ ID No. 153) | WP_072319476.1 | MESTTTLGLHLKYQHDLFEDKHYFGGGVNLA<br>VQNIESIFQAFAERYGIQNPLRKNGVPAINNIFH<br>DNISISNYKEYLKFLKQYLPVVGFLEKSNEINIF<br>EFREDFEILINAIYKLRHFYTHYYHSPIKLEDRF<br>YTCLNELFVAVAIQVKKHKMKSDKTRQLLNK<br>NLHQLLQQLIEQKREKLKDKKAEGEKVSLDTK<br>SIENAVLNDAFVHLLDKDENIRLNYSSRLSEDII<br>TKNGITLSISGLLFLLSLFLQRKEAEDLRSRIEGF<br>KGKGNELRFMATHWVFSYLNVKRIKHRLNTD<br>FQKETLLIQIADELSKVPDEVYKTLDHENRSKF<br>LEDINEYIREGNEDASLNESTVVHGVIRKRYEN<br>KFHYLVLRYLDEFVDFPSLRFQVHLGNYIHDR<br>RDKVIDGTNFITNRVIKEPIKVFGKLSHVSKLKS<br>DYMESLSREHKNGWDVFPNPSYNFVGHNIPIFI<br>NLRSASSKGKELYRDLMKIKSEKKKKSREEGIP<br>MERRDGKPTKIEISNQIDRNIKDNNFKDIYPGEP<br>LAMLSLNELPALLFELLRRPSITPQDIEDRMVE<br>KLYERFQIIRDYKPGDGLSTSKISKKLRKADNS<br>TRLDGKKLLRAIQTETRNAREKLHTLEENKAL<br>QKNRKRRTVYTTREQGREASWLAQDLKRFMP<br>IASRKEWRGYHHSQLQQILAFYDQNPKQPLEL<br>LEQFWDLKEDTYVWNSWIHKSLSQHNGFVPM<br>YEGYLKGRLGYYKKLESDIIGFLEEHKVLKRY<br>YTQQHLNVIFRERLYFIKTETKQKLELLARPLV<br>FPRGIFDDKPTFVQDKKVVDHPELFADWYVYS<br>YKDDHSFQEFYHYKRDYNEIFETELSWDIDFK<br>DNKRQLNPSEQMDLFRMKWDLKIKKIKIQDIF<br>LKIVAEDIYLKIFGHKIPLSLSDFYISRQERLTLD<br>EQAVAQSMRLPGDTSENQIKESNLWQTTVPYE<br>KEQIREPKIKLKDIGKFKYFLQQQKVLNLLKYD<br>PQHVWTKAELEEELYIGKHSYEVVRREMLLQK<br>CHQLEKHILEQFRFDGSNHPRELEQGNHPNFK<br>MYIVNGILTKRGELEIEAENWWLELGNSKNSL<br>DKVEVELLTMKTIPEQKAFLLILIRNKFAHNQL<br>PADNYFHYASNLMNLKKSDTYSLFWFTVADTI<br>VQEFMSL |
| 琼脂穴 | WP_ | MKTNPLIASSGEKPNYKKFNTESDKSFKKIFQN |

[0565]

| | | |
|---|---|---|
| 赖兴巴赫氏菌(Reichenbachiella agariperforans)(SEQ ID No. 154) | 07312444 1.1 | KGSIAPIAEKACKNFEIKSKSPVNRDGRLHYFS<br>VGHAFKNIDSKNVFRYELDESQMDMKPTQFLA<br>LQKEFFDFQGALNGLLKHIRNVNSHYVHTFEK<br>LEIQSINQKLITFLIEAFELAVIHSYLNEEELSYE<br>AYKDDPQSGQKLVQFLCDKFYPNKEHEVEER<br>KTILAKNKRQALEHLLFIEVTSDIDWKLFEKHK<br>VFTISNGKYLSFHACLFLLSLFLYKSEANQLISK<br>IKGFKRNDDNQYRSKRQIFTFFSKKFTSQDVNS<br>EEQHLVKFRDVIQYLNHYPSAWNKHLELKSGY<br>PQMTDKLMRYIVEAEIYRSFPDQTDNHRFLLFA<br>IREFFGQSCLDTWTGNTPINFSNQEQKGFSYEIN<br>TSAEIKDIETKLKALVLKGPLNFKEKKEQNRLE<br>KDLRREKKEQPTNRVKEKLLTRIQHNMLYVSY<br>GRNQDRFMDFAARFLAETDYFGKDAKFKMYQ<br>FYTSDEQRDHLKEQKKELPKKEFEKLKYHQSK<br>LVDYFTYAEQQARYPDWDTPFVVENNAIQIKV<br>TLFNGAKKIVSVQRNLMLYLLEDALYSEKREN<br>AGKGLISGYFVHHQKELKDQLDILEKETEISRE<br>QKREFKKLLPKRLLHRYSPAQINDTTEWNPME<br>VILEEAKAQEQRYQLLLEKAILHQTEEDFLKRN<br>KGKQFKLRFVRKAWHLMYLKELYMNKVAEH<br>GHHKSFHITKEEFNDFCRWMFAFDEVPKYKEY<br>LCDYFSQKGFFNNAEFKDLIESSTSLNDLYEKT<br>KQRFEGWSKDLTKQSDENKYLLANYESMLKD<br>DMLYVNISHFISYLESKGKINRNAHGHIAYKAL<br>NNVPHLIEEYYYKDRLAPEEYKSHGKLYNKLK<br>TVKLEDALLYEMAMHYLSLEPALVPKVKTKV<br>KDILSSNIAFDIKDAAGHHLYHLLIPFHKIDSFV<br>ALINHQSQQEKDPDKTSFLAKIQPYLEKVKNSK<br>DLKAVYHYYKDTPHTLRYEDLNMIHSHIVSQS<br>VQFTKVALKLEEYFIAKKSITLQIARQISYSEIA<br>DLSNYFTDEVRNTAFHFDVPETAYSMILQGIES<br>EFLDREIKPQKPKSLSELSTQQVSVCTAFLETLH<br>NNLFDRKDDKKERLSKARERYFEQIN |

[0566] 在某些示例性实施方案中,RNA靶向效应蛋白是如2017年6月26日提交的美国临时专利申请号62/525,165和2017年8月16日提交的PCT申请号US 2017/047193中所公开的Cas13c效应蛋白。在下表4B中提供了Cas13c的示例野生型直系同源物序列。在某些示例实施方案中,CRISPR效应蛋白是来自表4a或4b的Cas13c蛋白。

[0567] 表4a

[0568]

| | | |
|---|---|---|
| 坏死梭杆菌肠形亚 | | MEKFRRQNRNSIIKIIISNYDTKGIKELKVRYR<br>KQAQLDTFIIKTEIVNNDIFIKSIIEKAREKYRYS |

[0569]

| | | |
|---|---|---|
| 种 (Fusobacterium necrophorum subsp. funduliforme) ATCC 51357 contig00003 (SEQ ID No. 155) | | FLFDGEEKYHFKNKSSVEIVKKDIFSQTPDNMI<br>RNYKITLKISEKNPRVVEAEIEDLMNSTILKDG<br>RRSARREKSMTERKLIEEKVAKNYSLLANCP<br>MEEVDSIKIYKIKRFLTYRSNMLLYFASINSFL<br>CEGIKGKDNETEEIWHLKDNDVRKEKVRENF<br>KNKLIQSTENYNSSLKNQIEEKEKLLRKEFKK<br>GAFYRTIIKKLQQERIKELSEKSLTEDCEKIIKL<br>YSKLRHSLMHYDYQYFENLFENKKNDDLMK<br>DLNLDLFKSLPLIRKMKLNNKVNYLEDGDTLF<br>VLQKTKKAKTLYQIYDALCEQKNGFNKFINDF<br>FVSDGEENTVFKQIINEKFQSEMEFLEKRISESE<br>KKNEKLKKKLDSMKAHFRNINSEDTKEAYFW<br>DIHSSRNYKTKYNERKNLVNEYTELLGSSKEK<br>KLLREEITKINRQLLKLKQEMEEITKKNSLFRL<br>EYKMKIAFGFLFCEFDGNISKFKDEFDASNQE<br>KIIQYHKNGEKYLTSFLKEEEKEKFNLEKMQK<br>IIQKTEEEDWLLPETKNNLFKFYLLTYLLLPYE<br>LKGDFLGFVKKHYYDIKNVDFIDENQNNIQVS<br>QTVEKQEDYFYHKIRLFEKNTKKYEIVKYSIV<br>PNEKLKQYFEDLGIDIKYLTVEQKSEVSEEKN<br>KKVSLKNNGMFNKTILLFVFKYYQIAFKLFND<br>IELYSLFFLREKSGKPLEIFRKELESKMKDGYL<br>NFGQLLYVVYEVLVKNKDLDKILSKKIDYRK<br>DKSFSPEIAYLRNFLSHLNYSKFLDNFMKINTN<br>KSDENKEVLIPSIKIQKMIQFIEKCNLQNQIDFD<br>FNFVNDFYMRKEKMFFIQLKQIFPDINSTEKQ<br>KMNEKEEILRNRYHLTDKKNEQIKDEHEAQS<br>QLYEKILSLQKIYSSDKNNFYGRLKEEKLLFLE<br>KQGKKKLSMEEIKDKIAGDISDLLGILKKEITR<br>DIKDKLTEKFRYCEEKLLNLSFYNHQDKKKEE<br>SIRVFLIRDKNSDNFKFESILDDGSNKIFISKNG<br>KEITIQCCDKVLETLIIEKNTLKISSNGKIISLIPH<br>YSYSIDVKY |
| 坏死梭杆菌 (Fusobacterium necrophorum) DJ-2 contig0065，全基因组鸟枪序列 | | MEKFRRQNRSSIIKIIISNYDTKGIKELKVRYRK<br>QAQLDTFIIKTEIVNNDIFIKSIIEKAREKYRYSF<br>LFDGEEKYHFKNKSSVEIVKKDIFSQTPDNMIR<br>NYKITLKISEKNPRVVEAEIEDLMNSTILKDGR<br>RSARREKSMTERKLIEEKVAENYSLLANCPME<br>EVDSIKIYKIKRFLTYRSNMLLYFASINSFLCEG<br>IKGKDNETEEIWHLKDNDVRKEKVKENFKNK<br>LIQSTENYNSSLKNQIEEKEKLLRKESKKGAFY<br>RTIIKKLQQERIKELSEKSLTEDCEKIIKLYSEL<br>RHPLMHYDYQYFENLFENKENSELTKNLNLDI |

[0570]

| | | |
|---|---|---|
| (SEQ ID No. 156) | | FKSLPLVRKMKLNNKVNYLEDNDTLFVLQKT<br>KKAKTLYQIYDALCEQKNGFNKFINDFFVSDG<br>EENTVFKQIINEKFQSEIEFLEKRISESEKKNEK<br>LKKKLDSMKAHFRNINSEDTKEAYFWDIHSSR<br>NYKTKYNERKNLVNEYTELLGSSKEKKLLRE<br>EITKINRQLLKLKQEMEEITKKNSLFRLEYKM<br>KMAFGFLFCEFDGNISRFKDEFDASNQEKIIQY<br>HKNGEKYLTYFLKEEEKEKFNLKKLQETIQKT<br>GEENWLLPQNKNNLFKFYLLTYLLLPYELKG<br>DFLGFVKKHYYDIKNVDFMDENQSSKIIESKE<br>DDFYHKIRLFEKNTKKYEIVKYSIVPDKKLKQ<br>YFKDLGIDTKYLILDQKSEVSGEKNKKVSLKN<br>NGMFNKTILLFVFKYYQIAFKLFNDIELYSLFF<br>LREKSGKPFEVFLKELKDKMIGKQLNFGQLLY<br>VVYEVLVKNKDLSEILSERIDYRKDMCFSAEI<br>ADLRNFLSHLNYSKFLDNFMKINTNKSDENKE<br>VLIPSIKIQKMIKFIEECNLQSQIDFDFNFVNDF<br>YMRKEKMFFIQLKQIFPDINSTEKQKMNEKEEI<br>LRNRYHLTDKKNEQIKDEHEAQSQLYEKILSL<br>QKIYSSDKNNFYGRLKEEKLLFLEKQEKKKLS<br>MEEIKDKIAGDISDLLGILKKEITRDIKDKLTEK<br>FRYCEEKLLNLSFYNHQDKKKEESIRVFLIRDK<br>NSDNFKFESILDDGSNKIFISKNGKEITIQCCDK<br>VLETLIIEKNTLKISSNGKIISLIPHYSYSIDVKY |
| 坏死梭杆菌 (Fusobacterium necrophorum) BFTR-1 contig0068 (SEQ ID No. 157) | | MKVRYRKQAQLDTFIIKTEIVNNDIFIKSIIEKA<br>REKYRYSFLFDGEEKYHFKNKSSVEIVKNDIFS<br>QTPDNMIRNYKITLKISEKNPRVVEAEIEDLMN<br>STILKDGRRSARREKSMTERKLIEEKVAENYS<br>LLANCPIEEVDSIKIYKIKRFLTYRSNMLLYFAS<br>INSFLCEGIKGKDNETEEIWHLKDNDVRKEKV<br>KENFKNKLIQSTENYNSSLKNQIEEKEKLSSKE<br>FKKGAFYRTIIKKLQQERIKELSEKSLTEDCEKI<br>IKLYSELRHPLMHYDYQYFENLFENKENSELT<br>KNLNLDIFKSLPLVRKMKLNNKVNYLEDNDT<br>LFVLQKTKKAKTLYQIYDALCEQKNGFNKFIN<br>DFFVSDGEENTVFKQIINEKFQSEMEFLEKRIS<br>ESEKKNEKLKKKLDSMKAHFRNINSEDTKEA<br>YFWDIHSSRNYKTKYNERKNLVNEYTKLLGS<br>SKEKKLLREEITKINRQLLKLKQEMEEITKKNS<br>LFRLEYKMKIAFGFLFCEFDGNISKFKDEFDAS<br>NQEKIIQYHKNGEKYLTSFLKEEEKEKFNLEK<br>MQKIIQKTEEEDWLLPETKNNLFKFYLLTYLL<br>LPYELKGDFLGFVKKHYYDIKNVDFMDENQN |

[0571]

| | | |
|---|---|---|
| | | NIQVSQTVEKQEDYFYHKIRLFEKNTKKYEIV KYSIVPNEKLKQYFEDLGIDIKYLTGSVESGEK WLGENLGIDIKYLTVEQKSEVSEEKNKKVSLK NNGMFNKTILLFVFKYYQIAFKLFNDIELYSLF FLREKSEKPFEVFLEELKDKMIGKQLNFGQLL YVVYEVLVKNKDLDKILSKKIDYRKDKSFSPE IAYLRNFLSHLNYSKFLDNFMKINTNKSDENK EVLIPSIKIQKMIQFIEKCNLQNQIDFDFNFVND FYMRKEKMFFIQLKQIFPDINSTEKQKKSEKEE ILRKRYHLINKKNEQIKDEHEAQSQLYEKILSL QKIFSCDKNNFYRRLKEEKLLFLEKQGKKKIS MKEIKDKIASDISDLLGILKKEITRDIKDKLTEK FRYCEEKLLNISFYNHQDKKKEEGIRVFLIRDK NSDNFKFESILDDGSNKIFISKNGKEITIQCCDK VLETLMIEKNTLKISSNGKIISLIPHYSYSIDVK Y |
| 坏死梭杆菌肠形亚种 1_1_36S cont1.14 (SEQ ID No. 158) | | MTEKKSIIFKNKSSVEIVKKDIFSQTPDNMIRN YKITLKISEKNPRVVEAEIEDLMNSTILKDGRR SARREKSMTERKLIEEKVAENYSLLANCPMEE VDSIKIYKIKRFLTYRSNMLLYFASINSFLCEGI KGKDNETEEIWHLKDNDVRKEKVKENFKNKL IQSTENYNSSLKNQIEEKEKLLRKESKKGAFYR TIIKKLQQERIKELSEKSLTEDCEKIIKLYSELR HPLMHYDYQYFENLFENKENSELTKNLNLDIF KSLPLVRKMKLNNKVNYLEDNDTLFVLQKTK KAKTLYQIYDALCEQKNGFNKFINDFFVSDGE ENTVFKQIINEKFQSEMEFLEKRISESEKKNEK LKKKFDSMKAHFHNINSEDTKEAYFWDIHSSS NYKTKYNERKNLVNEYTELLGSSKEKKLLRE EITQINRKLLKLKQEMEEITKKNSLFRLEYKM KIAFGFLFCEFDGNISKFKDEFDASNQEKIIQY HKNGEKYLTYFLKEEEKEKFNLEKMQKIIQKT EEEDWLLPETKNNLFKFYLLTYLLLPYELKGD FLGFVKKHYYDIKNVDFMDENQNNIQVSQTV EKQEDYFYHKIRLFEKNTKKYEIVKYSIVPNE KLKQYFEDLGIDIKYLTGSVESGEKWLGENLG IDIKYLTVEQKSEVSEEKIKKFL |
| 极臭梭杆菌 (Fusobacterium perfoetens) | | MGKPNRSSIIKIIISNYDNKGIKEVKVRYNKQA QLDTFLIKSELKDGKFILYSIVDKAREKYRYSF EIDKTNINKNEILIIKKDIYSNKEDKVIRKYILSF EVSEKNDRTIVTKIKDCLETQKKEKFERENTR RLISETERKLLSEETQKTYSKIACCSPEDIDSVK IYKIKRYLAYRSNMLLFFSLINDIFVKGVVKDN |

[0572]

| | | |
|---|---|---|
| ATCC 29250 T364DRAFT_scaffold00009.9_C (SEQ ID No. 159) | | GEEVGEIWRIIDSKEIDEKKTYDLLVENFKKR<br>MSQEFINYKQSIENKIEKNTNKIKEIEQKLKKE<br>KYKKEINRLKKQLIELNRENDLLEKDKIELSDE<br>EIREDIEKILKIYSDLRHKLMHYNYQYFENLFE<br>NKKISKEKNEDVNLTELLDLNLFRYLPLVRQL<br>KLENKTNYLEKEDKITVLGVSDSAIKYYSYYN<br>FLCEQKNGFNNFINSFFSNDGEENKSFKEKINL<br>SLEKEIEIMEKETNEKIKEINKNELQLMKEQKE<br>LGTAYVLDIHSLNDYKISHNERNKNVKLQNDI<br>MNGNRDKNALDKINKKLVELKIKMDKITKRN<br>SILRLKYKLQVAYGFLMEEYKGNIKKFKDEFD<br>ISKEKIKSYKSKGEKYLEVKSEKKYITKILNSIE<br>DIHNITWLKNQEENNLFKFYVLTYILLPFEFRG<br>DFLGFVKKHYYDIKNVEFLDENNDRLTPEQLE<br>KMKNDSFFNKIRLFEKNSKKYDILKESILTSERI<br>GKYFSLLNTGAKYFEYGGEENRGIFNKNIIIPIF<br>KYYQIVLKLYNDVELAMLLTLSESDEKDINKI<br>KELVTLKEKVSPKKIDYEKKYKFSVLLDCFNR<br>IINLGKKDFLASEEVKEVAKTFTNLAYLRNKIC<br>HLNYSKFIDDLLTIDTNKSTTDSEGKLLINDRIR<br>KLIKFIRENNQKMNISIDYNYINDYYMKKEKFI<br>FGQRKQAKTIIDSGKKANKRNKAEELLKMYR<br>VKKENINLIYELSKKLNELTKSELFLLDKKLLK<br>DIDFTDVKIKNKSFFELKNDVKEVANIKQALQ<br>KHSSELIGIYKKEVIMAIKRSIVSKLIYDEEKVL<br>SIIIYDKTNKKYEDFLLEIRRERDINKFQFLIDE<br>KKEKLGYEKIIETKEKKKVVVKIQNNSELVSE<br>PRIIKNKDKKKAKTPEEISKLGILDLTNHYCFN<br>LKITL |
| 溃疡梭杆菌 (Fusobacterium ulcerans) ATCC 49185 cont2.38 (SEQ ID No. 160) | | MENKGNNKKIDFDENYNILVAQIKEYFTKEIE<br>NYNNRIDNIIDKKELLKYSEKKEESEKNKKLE<br>ELNKLKSQKLKILTDEEIKADVIKIIKIFSDLRH<br>SLMHYEYKYFENLFENKKNEELAELLNLNLF<br>KNLTLLRQMKIENKTNYLEGREEFNIIGKNIKA<br>KEVLGHYNLLAEQKNGFNNFINSFFVQDGTEN<br>LEFKKLIDEHFVNAKKRLERNIKKSKKLEKEL<br>EKMEQHYQRLNCAYVWDIHTSTTYKKLYNK<br>RKSLIEEYNKQINEIKDKEVITAINVELLRIKKE<br>MEEITKSNSLFRLKYKMQIAYAFLEIEFGGNIA<br>KFKDEFDCSKMEEVQKYLKKGVKYLKYYKD<br>KEAQKNYEFPFEEIFENKDTHNEEWLENTSEN<br>NLFKFYILTYLLLPMEFKGDFLGVVKKHYYDI<br>KNVDFTDESEKELSQVQLDKMIGDSFFHKIRL |

| | | | |
|---|---|---|---|
| | | | FEKNTKRYEIIKYSILTSDEIKRYFRLLELDVPY<br>FEYEKGTDEIGIFNKNIILTIFKYYQIIFRLYNDL<br>EIHGLFNISSDLDKILRDLKSYGNKNINFREFLY<br>VIKQNNNSSTEEEYRKIWENLEAKYLRLHLLT<br>PEKEEIKTKTKEELEKLNEISNLRNGICHLNYK<br>EIIEEILKTEISEKNKEATLNEKIRKVINFIKENE<br>LDKVELGFNFINDFFMKKEQFMFGQIKQVKEG<br>NSDSITTERERKEKNNKKLKETYELNCDNLSE<br>FYETSNNLRERANSSSLLEDSAFLKKIGLYKV<br>KNNKVNSKVKDEEKRIENIKRKLLKDSSDIMG<br>MYKAEVVKKLKEKLILIFKHDEEKRIYVTVYD<br>TSKAVPENISKEILVKRNNSKEEYFFEDNNKK<br>YVTEYYTLEITETNELKVIPAKKLEGKEFKTEK<br>NKENKLMLNNHYCFNVKIIY |
| [0573] | 厌氧盐杆菌属某种(Anaerosalibacter sp.) ND1 基因组组装马赛厌氧盐杆菌(Anaerosalibacter massiliensis) ND1 (SEQ ID No. 161) | | MKSGRREKAKSNKSSIVRVIISNFDDKQVKEIK<br>VLYTKQGGIDVIKFKSTEKDEKGRMKFNFDC<br>AYNRLEEEEFNSFGGKGKQSFFVTTNEDLTEL<br>HVTKRHKTTGEIIKDYTIQGKYTPIKQDRTKVT<br>VSITDNKDHFDSNDLGDKIRLSRSLTQYTNRIL<br>LDADVMKNYREIVCSDSEKVDETINIDSQEIY<br>KINRFLSYRSNMIIYYQMINNFLLHYDGEEDK<br>GGNDSINLINEIWKYENKKNDEKEKIIERSYKS<br>IEKSINQYILNHNTEVESGDKEKKIDISEERIKE<br>DLKKTFILFSRLRHYMVHYNYKFYENLYSGK<br>NFIIYNKDKSKSRRFSELLDLNIFKELSKIKLVK<br>NRAVSNYLDKKTTIHVLNKNINAIKLLDIYRDI<br>CETKNGFNNFINNMMTISGEEDKEYKEMVTK<br>HFNENMNKLSIYLENFKKHSDFKTNNKKKET<br>YNLLKQELDEQKKLRLWFNAPYVYDIHSSKK<br>YKELYVERKKYVDIHSKLIEAGINNDNKKKLN<br>EINVKLCELNTEMKEMTKLNSKYRLQYKLQL<br>AFGFILEEFNLDIDKFVSAFDKDNNLTISKFME<br>KRETYLSKSLDRRDNRFKKLIKDYKFRDTEDI<br>FCSDRENNLVKLYILMYILLPVEIRGDFLGFVK<br>KNYYDLKHVDFIDKRNNDNKDTFFHDLRLFE<br>KNVKRLEVTSYSLSDGFLGKKSREKFGKELEK<br>FIYKNVSIALPTNIDIKEFNKSLVLPMMKNYQII<br>FKLLNDIEISALFLIAKKEGNEGSITFKKVIDKV<br>RKEDMNGNINFSQVMKMALNEKVNCQIRNSI<br>AHINMKQLYIEPLNIYINNNQNKKTISEQMEEII<br>DICITKGLTGKELNKNIINDYYMKKEKLVFNL<br>KLRKRNNLVSIDAQQKNMKEKSILNKYDLNY<br>KDENLNIKEIILKVNDLNNKQKLLKETTEGES |

[0574]

| | | |
|---|---|---|
| | | NYKNALSKDILLLNGIIRKNINFKIKEMILGIIQ<br>QNEYRYVNINIYDKIRKEDHNIDLKINNKYIEIS<br>CYENKSNESTDERINFKIKYMDLKVKNELLVP<br>SCYEDIYIKKKIDLEIRYIENCKVVYIDIYYKKY<br>NINLEFDGKTLFVKFNKDVKKNNQKVNLESN<br>YIQNIKFIVS |

[0575] 表4B

| 名称 | 序列 |
|---|---|
| EHO<br>1908<br>1 | MTEKKSIIFKNKSSVEIVKKDIFSQTPDNMIRNYKITLKI<br>SEKNPRVVEAEIEDLMNSTILKDGRRSARREKSMTERK<br>LIEEKVAENYSLLANCPMEEVDSIKIYKIKRFLTYRSN<br>MLLYFASINSFLCEGIKGKDNETEEIWHLKDNDVRKEK<br>VKENFKNKLIQSTENYNSSLKNQIEEKEKLLRKESKKG<br>AFYRTIIKKLQQERIKELSEKSLTEDCEKIIKLYSELRHP<br>LMHYDYQYFENLFENKENSELTKNLNLDIFKSLPLVR<br>KMKLNNKVNYLEDNDTLFVLQKTKKAKTLYQIYDAL<br>CEQKNGFNKFINDFFVSDGEENTVFKQIINEKFQSEME<br>FLEKRISESEKKNEKLKKKFDSMKAHFHNINSEDTKEA<br>YFWDIHSSSNYKTKYNERKNLVNEYTELLGSSKEKKL<br>LREEITQINRKLLKLKQEMEEITKKNSLFRLEYKMKIAF<br>GFLFCEFDGNISKFKDEFDASNQEKIIQYHKNGEKYLT<br>YFLKEEEKEKFNLEKMQKIIQKTEEEDWLLPETKNNLF<br>KFYLLTYLLLPYELKGDFLGFVKKHYYDIKNVDFMDE<br>NQNNIQVSQTVEKQEDYFYHKIRLFEKNTKKYEIVKYS<br>IVPNEKLKQYFEDLGIDIKYLTGSVESGEKWLGENLGI<br>DIKYLTVEQKSEVSEEKIKKFL |
| WP_<br>0948<br>9933<br>6 | MEKDKKGEKIDISQEMIEEDLRKILILFSRLRHSMVHY<br>DYEFYQALYSGKDFVISDKNNLENRMISQLLDLNIFKE<br>LSKVKLIKDKAISNYLDKNTTIHVLGQDIKAIRLLDIYR<br>DICGSKNGFNKFINTMITISGEEDREYKEKVIEHFNKK<br>MENLSTYLEKLEKQDNAKRNNKRVYNLLKQKLIEQQ<br>KLKEWFGGPYVYDIHSSKRYKELYIERKKLVDRHSKL<br>FEEGLDEKNKKELTKINDELSKLNSEMKEMTKLNSKY<br>RLQYKLQLAFGFILEEFDLNIDTFINNFDKDKDLIISNF<br>MKKRDIYLNRVLDRGDNRLKNIIKEYKFRDTEDIFCND<br>RDNNLVKLYILMYILLPVEIRGDFLGFVKKNYYDMKH<br>VDFIDKKDKEDKDTFFHDLRLFEKNIRKLEITDYSLSSG<br>FLSKEHKVDIEKKINDFINRNGAMKLPEDITIEEFNKSLI<br>LPIMKNYQINFKLLNDIEISALFKIAKDRSITFKQAIDEI<br>KNEDIKKNSKKNDKNNHKDKNINFTQLMKRALHEKIP<br>YKAGMYQIRNNISHIDMEQLYIDPLNSYMNSNKNNITI<br>SEQIEKIIDVCVTGGVTGKELNNNIINDYYMKKEKLVF |

[0576]

[0577]

| | |
|---|---|
| | NLKLRKQNDIVSIESQEKNKREEFVFKKYGLDYKDGEI<br>NIIEVIQKVNSLQEELRNIKETSKEKLKNKETLFRDISLI<br>NGTIRKNINFKIKEMVLDIVRMDEIRHINIHIYYKGENY<br>TRSNIIKFKYAIDGENKKYYLKQHEINDINLELKDKFVT<br>LICNMDKHPNKNKQTINLESNYIQNVKFIIP |
| WP_<br>0404<br>9087<br>6 | MENKGNNKKIDFDENYNILVAQIKEYFTKEIENYNNRI<br>DNIIDKKELLKYSEKKEESEKNKKLEELNKLKSQKLKI<br>LTDEEIKADVIKIIKIFSDLRHSLMHYEYKYFENLFENK<br>KNEELAELLNLNLFKNLTLLRQMKIENKTNYLEGREEF<br>NIIGKNIKAKEVLGHYNLLAEQKNGFNNFINSFFVQDG<br>TENLEFKKLIDEHFVNAKKRLERNIKKSKKLEKELEKM<br>EQHYQRLNCAYVWDIHTSTTYKKLYNKRKSLIEEYNK<br>QINEIKDKEVITAINVELLRIKKEMEEITKSNSLFRLKYK<br>MQIAYAFLEIEFGGNIAKFKDEFDCSKMEEVQKYLKK<br>GVKYLKYYKDKEAQKNYEFPFEEIFENKDTHNEEWLE<br>NTSENNLFKFYILTYLLLPMEFKGDFLGVVKKHYYDIK<br>NVDFTDESEKELSQVQLDKMIGDSFFHKIRLFEKNTKR<br>YEIIKYSILTSDEIKRYFRLLELDVPYFEYEKGTDEIGIF<br>NKNIILTIFKYYQIIFRLYNDLEIHGLFNISSDLDKILRDL<br>KSYGNKNINFREFLYVIKQNNNSSTEEEYRKIWENLEA<br>KYLRLHLLTPEKEEIKTKTKEELEKLNEISNLRNGICHL<br>NYKEIIEEILKTEISEKNKEATLNEKIRKVINFIKENELD<br>KVELGFNFINDFFMKKEQFMFGQIKQVKEGNSDSITTE<br>RERKEKNNKKLKETYELNCDNLSEFYETSNNLRERAN<br>SSSLLEDSAFLKKIGLYKVKNNKVNSKVKDEEKRIENI<br>KRKLLKDSSDIMGMYKAEVVKKLKEKLILIFKHDEEK<br>RIYVTVYDTSKAVPENISKEILVKRNNSKEEYFFEDNN<br>KKYVTEYYTLEITETNELKVIPAKKLEGKEFKTEKNKE<br>NKLMLNNHYCFNVKIIY |
| WP_<br>0473<br>9660<br>7 | MEEIKHKKNKSSIIRVIVSNYDMTGIKEIKVLYQKQGG<br>VDTFNLKTIINLESGNLEIISCKPKEREKYRYEFNCKTEI<br>NTISITKKDKVLKKEIRKYSLELYFKNEKKDTVVAKVT<br>DLLKAPDKIEGERNHLRKLSSSTERKLLSKTLCKNYSEI<br>SKTPIEEIDSIKIYKIKRFLNYRSNFLIYFALINDFLCAGV<br>KEDDINEVWLIQDKEHTAFLENRIEKITDYIFDKLSKDI<br>ENKKNQFEKRIKKYKTSLEELKTETLEKNKTFYIDSIKT<br>KITNLENKITELSLYNSKESLKEDLIKIISIFTNLRHSLM<br>HYDYKSFENLFENIENEELKNLLDLNLFKSIRMSDEFK<br>TKNRTNYLDGTESFTIVKKHQNLKKLYTYYNNLCDKK<br>NGFNTFINSFFVTDGIENTDFKNLIILHFEKEMEEYKKSI<br>EYYKIKISNEKNKSKKEKLKEKIDLLQSELINMREHKN<br>LLKQIYFFDIHNSIKYKELYSERKNLIEQYNLQINGVKD<br>VTAINHINTKLLSLKNKMDKITKQNSLYRLKYKLKIAY |

[0578]

| | |
|---|---|
| | SFLMIEFDGDVSKFKNNFDPTNLEKRVEYLDKKEEYL<br>NYTAPKNKFNFAKLEEELQKIQSTSEMGADYLNVSPE<br>NNLFKFYILTYIMLPVEFKGDFLGFVKNHYYNIKNVDF<br>MDESLLDENEVDSNKLNEKIENLKDSSFFNKIRLFEKNI<br>KKYEIVKYSVSTQENMKEYFKQLNLDIPYLDYKSTDEI<br>GIFNKNMILPIFKYYQNVFKLCNDIEIHALLALANKKQ<br>QNLEYAIYCCSKKNSLNYNELLKTFNRKTYQNLSFIRN<br>KIAHLNYKELFSDLFNNELDLNTKVRCLIEFSQNNKFD<br>QIDLGMNFINDYYMKKTRFIFNQRRLRDLNVPSKEKII<br>DGKRKQQNDSNNELLKKYGLSRTNIKDIFNKAWY |
| WP_<br>0359<br>3567<br>1 | MKVRYRKQAQLDTFIIKTEIVNNDIFIKSIIEKAREKYR<br>YSFLFDGEEKYHFKNKSSVEIVKNDIFSQTPDNMIRNY<br>KITLKISEKNPRVVEAEIEDLMNSTILKDGRRSARREKS<br>MTERKLIEEKVAENYSLLANCPIEEVDSIKIYKIKRFLT<br>YRSNMLLYFASINSFLCEGIKGKDNETEEIWHLKDNDV<br>RKEKVKENFKNKLIQSTENYNSSLKNQIEEKEKLSSKE<br>FKKGAFYRTIIKKLQQERIKELSEKSLTEDCEKIIKLYSE<br>LRHPLMHYDYQYFENLFENKENSELTKNLNLDIFKSLP<br>LVRKMKLNNKVNYLEDNDTLFVLQKTKKAKTLYQIY<br>DALCEQKNGFNKFINDFFVSDGEENTVFKQIINEKFQS<br>EMEFLEKRISESEKKNEKLKKKLDSMKAHFRNINSEDT<br>KEAYFWDIHSSRNYKTKYNERKNLVNEYTKLLGSSKE<br>KKLLREEITKINRQLLKLKQEMEEITKKNSLFRLEYKM<br>KIAFGFLFCEFDGNISKFKDEFDASNQEKIIQYHKNGEK<br>YLTSFLKEEEKEKFNLEKMQKIIQKTEEEDWLLPETKN<br>NLFKFYLLTYLLLPYELKGDFLGFVKKHYYDIKNVDF<br>MDENQNNIQVSQTVEKQEDYFYHKIRLFEKNTKKYEI<br>VKYSIVPNEKLKQYFEDLGIDIKYLTGSVESGEKWLGE<br>NLGIDIKYLTVEQKSEVSEEKNKKVSLKNNGMFNKTIL<br>LFVFKYYQIAFKLFNDIELYSLFFLREKSEKPFEVFLEEL<br>KDKMIGKQLNFGQLLYVVYEVLVKNKDLDKILSKKID<br>YRKDKSFSPEIAYLRNFLSHLNYSKFLDNFMKINTNKS<br>DENKEVLIPSIKIQKMIQFIEKCNLQNQIDFDFNFVNDF<br>YMRKEKMFFIQLKQIFPDINSTEKQKKSEKEEILRKRY<br>HLINKKNEQIKDEHEAQSQLYEKILSLQKIFSCDKNNF<br>YRRLKEEKLLFLEKQGKKKISMKEIKDKIASDISDLLGI<br>LKKEITRDIKDKLTEKFRYCEEKLLNISFYNHQDKKKE<br>EGIRVFLIRDKNSDNFKFESILDDGSNKIFISKNGKEITI<br>QCCDKVLETLMIEKNTLKISSNGKIISLIPHYSYSIDVKY |
| WP_<br>0359<br>0656<br>3 | MEKFRRQNRSSIIKIIISNYDTKGIKELKVRYRKQAQLD<br>TFIIKTEIVNNDIFIKSIIEKAREKYRYSFLFDGEEKYHFK<br>NKSSVEIVKKDIFSQTPDNMIRNYKITLKISEKNPRVVE<br>AEIEDLMNSTILKDGRRSARREKSMTERKLIEEKVAEN |

[0579]

| | |
|---|---|
| | YSLLANCPMEEVDSIKIYKIKRFLTYRSNMLLYFASINS<br>FLCEGIKGKDNETEEIWHLKDNDVRKEKVKENFKNKL<br>IQSTENYNSSLKNQIEEKEKLLRKESKKGAFYRTIIKKL<br>QQERIKELSEKSLTEDCEKIIKLYSELRHPLMHYDYQY<br>FENLFENKENSELTKNLNLDIFKSLPLVRKMKLNNKV<br>NYLEDNDTLFVLQKTKKAKTLYQIYDALCEQKNGFN<br>KFINDFFVSDGEENTVFKQIINEKFQSEIEFLEKRISESE<br>KKNEKLKKKLDSMKAHFRNINSEDTKEAYFWDIHSSR<br>NYKTKYNERKNLVNEYTELLGSSKEKKLLREEITKINR<br>QLLKLKQEMEEITKKNSLFRLEYKMKMAFGFLFCEFD<br>GNISRFKDEFDASNQEKIIQYHKNGEKYLTYFLKEEEK<br>EKFNLKKLQETIQKTGEENWLLPQNKNNLFKFYLLTY<br>LLLPYELKGDFLGFVKKHYYDIKNVDFMDENQSSKIIE<br>SKEDDFYHKIRLFEKNTKKYEIVKYSIVPDKKLKQYFK<br>DLGIDTKYLILDQKSEVSGEKNKKVSLKNNGMFNKTIL<br>LFVFKYYQIAFKLFNDIELYSLFFLREKSGKPFEVFLKE<br>LKDKMIGKQLNFGQLLYVVYEVLVKNKDLSEILSERID<br>YRKDMCFSAEIADLRNFLSHNYSKFLDNFMKINTNKS<br>DENKEVLIPSIKIQKMIKFIEECNLQSQIDFDFNFVNDFY<br>MRKEKMFFIQLKQIFPDINSTEKQKMNEKEEILRNRYH<br>LTDKKNEQIKDEHEAQSQLYEKILSLQKIYSSDKNNFY<br>GRLKEEKLLFLEKQEKKKLSMEEIKDKIAGDISDLLGIL<br>KKEITRDIKDKLTEKFRYCEEKLLNLSFYNHQDKKKEE<br>SIRVFLIRDKNSDNFKFESILDDGSNKIFISKNGKEITIQC<br>CDKVLETLIIEKNTLKISSNGKIISLIPHYSYSIDVKY |
| WP_<br>0426<br>7893<br>1 | MKSGRREKAKSNKSSIVRVIISNFDDKQVKEIKVLYTK<br>QGGIDVIKFKSTEKDEKGRMKFNFDCAYNRLEEEEFNS<br>FGGKGKQSFFVTTNEDLTELHVTKRHKTTGEIIKDYTI<br>QGKYTPIKQDRTKVTVSITDNKDHFDSNDLGDKIRLSR<br>SLTQYTNRILLDADVMKNYREIVCSDSEKVDETINIDS<br>QEIYKINRFLSYRSNMIIYYQMINNFLLHYDGEEDKGG<br>NDSINLINEIWKYENKKNDEKEKIIERSYKSIEKSINQYI<br>LNHNTEVESGDKEKKIDISEERIKEDLKKTFILFSRLRH<br>YMVHYNYKFYENLYSGKNFIIYNKDKSKSRRFSELLD<br>LNIFKELSKIKLVKNRAVSNYLDKKTTIHVLNKNINAIK<br>LLDIYRDICETKNGFNNFINNMMTISGEEDKEYKEMVT<br>KHFNENMNKLSIYLENFKKHSDFKTNNKKKETYNLLK<br>QELDEQKKLRLWFNAPYVYDIHSSKKYKELYVERKK<br>YVDIHSKLIEAGINNDNKKKLNEINVKLCELNTEMKE<br>MTKLNSKYRLQYKLQLAFGFILEEFNLDIDKFVSAFDK<br>DNNLTISKFMEKRETYLSKSLDRRDNRFKKLIKDYKFR<br>DTEDIFCSDRENNLVKLYILMYILLPVEIRGDFLGFVKK<br>NYYDLKHVDFIDKRNNDNKDTFFHDLRLFEKNVKRLE |

[0580]

| | |
|---|---|
| | VTSYSLSDGFLGKKSREKFGKELEKFIYKNVSIALPTNI<br>DIKEFNKSLVLPMMKNYQIIFKLLNDIEISALFLIAKKE<br>GNEGSITFKKVIDKVRKEDMNGNINFSQVMKMALNEK<br>VNCQIRNSIAHINMKQLYIEPLNIYINNNQNKKTISEQM<br>EEIIDICITKGLTGKELNKNIINDYYMKKEKLVFNLKLR<br>KRNNLVSIDAQQKNMKEKSILNKYDLNYKDENLNIKE<br>IILKVNDLNNKQKLLKETTEGESNYKNALSKDILLLNG<br>IIRKNINFKIKEMILGIIQQNEYRYVNINIYDKIRKEDHNI<br>DLKINNKYIEISCYENKSNESTDERINFKIKYMDLKVK<br>NELLVPSCYEDIYIKKKIDLEIRYIENCKVVYIDIYYKK<br>YNINLEFDGKTLFVKFNKDVKKNNQKVNLESNYIQNI<br>KFIVS |
| WP_<br>0626<br>2784<br>6 | MEKFRRQNRNSIIKIIISNYDTKGIKELKVRYRKQAQLD<br>TFIIKTEIVNNDIFIKSIIEKAREKYRYSFLFDGEEKYHFK<br>NKSSVEIVKKDIFSQTPDNMIRNYKITLKISEKNPRVVE<br>AEIEDLMNSTILKDGRRSARREKSVTERKLIEEKVAEN<br>YSLLANCPMEEVDSIKIYKIKRFLTYRSNMLLYFASINS<br>FLCEGIKGKENETEEIWHLKDNDVRKEKVKENFKNKL<br>IQSTENYNSSLKNQIEEKEKLLRKESKKGAFYRTIIKKL<br>QQERIKELSEKSLTEDCEKIIKLYSKLRHSLMHYDYQY<br>FENLFENKETPELKDKLDLHLFKSLPLIRKMKLNNKVN<br>YLEDGDTLFVLQKTKKAKTLYQIYDALCEQKNGFNKF<br>INDFFVSDGEENTVFKQIINEKFQSEMEFLGKRISESEE<br>KNPKLKKKFDSMKAHFHNINSEDTKEAYFWDIHSSSN<br>YKTKYNERKNLVNEYTELLGSSKEKKLLREEITQINRK<br>LLKLKQEMEEITKKNSLFRLEYKMKMAFGFLFCEFDG<br>NISRFKDEFDASNQEKIIQYHKNGEKYLTYFLKEEEKE<br>KFNLKKLQETIQKTGKENWLLPQNKNNLFKFYLLTYL<br>LLPYELKGDFLGFVKKHYYDIKNVDFMDENQSSKIIES<br>KEDDFYHKIRLFEKNTKKYEIVKYSIVPDEKLKQYFKD<br>LGIDTKYLILEQKSEVSGEKNKKVSLKNNGMFNKTILL<br>FVFKYYQIAFKLFNDIELYSLFFLREKSGKPFEVFLKEL<br>KDKMIGKQLNFGQLLYVIYEVLVKNKDLSEILSERIDY<br>RKDMCFSAEIADLRNFLSHLNYSKFLDNFMKINTNKS<br>DENKEVLIPSIKIQKMIKFIEECNLQSQIDFDFNFVNDFY<br>MRKEKMFFIQLKQIFPDINSTEKQKMNEKEEILRNRYH<br>LTDKKNEQIKDEHEAQSQLYEKILSLQKIYSSDKNNFY<br>GRLKEEKLLFLGKQGKKKLSMEEIKDKIAGDISDLLGI<br>LKKEITRDIKDKLTEKFRYCEEKLLNLSFYNHQDKKKE<br>ESIRVFLIRDKNSDNFKFESILDDGSNKIFISKNGKEITIQ<br>CCDKVLETLMIEKNTLKISSNGKIISLVPHYSYSIDVKY |
| WP_<br>0059 | MEKFRRQNRNSIIKIIISNYDTKGIKELKVRYRKQAQLD<br>TFIIKTEIVNNDIFIKSIIEKAREKYRYSFLFDGEEKYHFK |

[0581]

| | |
|---|---|
| 59231 | NKSSVEIVKKDIFSQTPDNMIRNYKITLKISEKNPRVVE<br>AEIEDLMNSTILKDGRRSARREKSMTERKLIEEKVAKN<br>YSLLANCPMEEVDSIKIYKIKRFLTYRSNMLLYFASINS<br>FLCEGIKGKDNETEEIWHLKDNDVRKEKVRENFKNKL<br>IQSTENYNSSLKNQIEEKEKLLRKEFKKGAFYRTIIKKL<br>QQERIKELSEKSLTEDCEKIIKLYSKLRHSLMHYDYQY<br>FENLFENKKNDDLMKDLNLDLFKSLPLIRKMKLNNKV<br>NYLEDGDTLFVLQKTKKAKTLYQIYDALCEQKNGFN<br>KFINDFFVSDGEENTVFKQIINEKFQSEMEFLEKRISESE<br>KKNEKLKKKLDSMKAHFRNINSEDTKEAYFWDIHSSR<br>NYKTKYNERKNLVNEYTELLGSSKEKKLLREEITKINR<br>QLLKLKQEMEEITKKNSLFRLEYKMKIAFGFLFCEFDG<br>NISKFKDEFDASNQEKIIQYHKNGEKYLTSFLKEEEKE<br>KFNLEKMQKIIQKTEEEDWLLPETKNNLFKFYLLTYLL<br>LPYELKGDFLGFVKKHYYDIKNVDFIDENQNNIQVSQT<br>VEKQEDYFYHKIRLFEKNTKKYEIVKYSIVPNEKLKQY<br>FEDLGIDIKYLTVEQKSEVSEEKNKKVSLKNNGMFNK<br>TILLFVFKYYQIAFKLFNDIELYSLFFLREKSGKPLEIFR<br>KELESKMKDGYLNFGQLLYVVYEVLVKNKDLDKILS<br>KKIDYRKDKSFSPEIAYLRNFLSHLNYSKFLDNFMKIN<br>TNKSDENKEVLIPSIKIQKMIQFIEKCNLQNQIDFDFNF<br>VNDFYMRKEKMFFIQLKQIFPDINSTEKQKMNEKEEIL<br>RNRYHLTDKKNEQIKDEHEAQSQLYEKILSLQKIYSSD<br>KNNFYGRLKEEKLLFLEKQGKKKLSMEEIKDKIAGDIS<br>DLLGILKKEITRDIKDKLTEKFRYCEEKLLNLSFYNHQ<br>DKKKEESIRVFLIRDKNSDNFKFESILDDGSNKIFISKN<br>GKEITIQCCDKVLETLIIEKNTLKISSNGKIISLIPHYSYSI<br>DVKY |
| WP_027128616 | MGKPNRSSIIKIIISNYDNKGIKEVKVRYNKQAQLDTFL<br>IKSELKDGKFILYSIVDKAREKYRYSFEIDKTNINKNEIL<br>IIKKDIYSNKEDKVIRKYILSFEVSEKNDRTIVTKIKDCL<br>ETQKKEKFERENTRRLISETERKLLSEETQKTYSKIACC<br>SPEDIDSVKIYKIKRYLAYRSNMLLFFSLINDIFVKGVV<br>KDNGEEVGEIWRIIDSKEIDEKKTYDLLVENFKKRMSQ<br>EFINYKQSIENKIEKNTNKIKEIEQKLKKEKYKKEINRL<br>KKQLIELNRENDLLEKDKIELSDEEIREDIEKILKIYSDL<br>RHKLMHYNYQYFENLFENKKISKEKNEDVNLTELLDL<br>NLFRYLPLVRQLKLENKTNYLEKEDKITVLGVSDSAIK<br>YYSYYNFLCEQKNGFNNFINSFFSNDGEENKSFKEKIN<br>LSLEKEIEIMEKETNEKIKEINKNELQLMKEQKELGTA<br>YVLDIHSLNDYKISHNERNKNVKLQNDIMNGNRDKN<br>ALDKINKKLVELKIKMDKITKRNSILRLKYKLQVAYGF<br>LMEEYKGNIKKFKDEFDISKEKIKSYKSKGEKYLEVKS |

| | |
|---|---|
| | EKKYITKILNSIEDIHNITWLKNQEENNLFKFYVLTYIL<br>LPFEFRGDFLGFVKKHYYDIKNVEFLDENNDRLTPEQL<br>EKMKNDSFFNKIRLFEKNSKKYDILKESILTSERIGKYF<br>SLLNTGAKYFEYGGEENRGIFNKNIIIPIFKYYQIVLKLY<br>NDVELAMLLTLSESDEKDINKIKELVTLKEKVSPKKID<br>YEKKYKFSVLLDCFNRIINLGKKDFLASEEVKEVAKTF<br>TNLAYLRNKICHLNYSKFIDDLLTIDTNKSTTDSEGKLL<br>INDRIRKLIKFIRENNQKMNISIDYNYINDYYMKKEKFI<br>FGQRKQAKTIIDSGKKANKRNKAEELLKMYRVKKENI<br>NLIYELSKKLNELTKSELFLLDKKLLKDIDFTDVKIKNK<br>SFFELKNDVKEVANIKQALQKHSSELIGIYKKEVIMAIK<br>RSIVSKLIYDEEKVLSIIIYDKTNKKYEDFLLEIRRERDI<br>NKFQFLIDEKKEKLGYEKIIETKEKKKVVVKIQNNSEL<br>VSEPRIIKNKDKKKAKTPEEISKLGILDLTNHYCFNLKI<br>TL |
| WP_<br>0626<br>2474<br>0 | MEKFRRQNRNSIIKIIISNYDTKGIKELKVRYRKQAQLD<br>TFIIKTEIVNNDIFIKSIIEKAREKYRYSFLFDGEEKYHFK<br>NKSSVEIVKKDIFSQTPDNMIRNYKITLKISEKNPRVVE<br>AEIEDLMNSTILKDGRRSARREKSMTERKLIEEKVAKN<br>YSLLANCPMEEVDSIKIYKIKRFLTYRSNMLLYFASINS<br>FLCEGIKGKDNETEEIWHLKDNDVRKEKVRENFKNKL<br>IQSTENYNSSLKNQIEEKEKLLRKEFKKGAFYRTIIKKL<br>QQERIKELSEKSLTEDCEKIIKLYSKLRHSLMHYDYQY<br>FENLFENKKNDDLMKDLNLDLFKSLPLIRKMKLNNKV<br>NYLEDGDTLFVLQKTKKAKTLYQIYDALCEQKNGFN<br>KFINDFFVSDGEENTVFKQIINEKFQSEMEFLEKRISESE<br>KKNEKLKKKLDSMKAHFRNINSEDTKEAYFWDIHSSR<br>NYKTKYNERKNLVNEYTELLGSSKEKKLLREEITKINR<br>QLLKLKQEMEEITKKNSLFRLEYKMKIAFGFLFCEFDG<br>NISKFKDEFDASNQEKIIQYHKNGEKYLTSFLKEEEKE<br>KFNLEKMQKIIQKTEEEDWLLPETKNNLFKFYLLTYLL<br>LPYELKGDFLGFVKKHYYDIKNVDFIDENQNNIQVSQT<br>VEKQEDYFYHKIRLFEKNTKKYEIVKYSIVPNEKLKQY<br>FEDLGIDIKYLTGSVESGEKWLGENLGIDIKYLTVEQK<br>SEVSEEKNKKVSLKNNGMFNKTILLFVFKYYQIAFKLF<br>NDIELYSLFFLREKSGKPLEIFRKELESKMKDGYLNFG<br>QLLYVVYEVLVKNKDLDKILSKKIDYRKDKSFSPEIAY<br>LRNFLSHLNYSKFLDNFMKINTNKSDENKEVLIPSIKIQ<br>KMIQFIEKCNLQNQIDFDFNFVNDFYMRKEKMFFIQLK<br>QIFPDINSTEKQKMNEKEEILRNRYHLTDKKNEQIKDE<br>HEAQSQLYEKILSLQKIYSSDKNNFYGRLKEEKLLFLE<br>KQGKKKLSMEEIKDKIAGDISDLLGILKKEITRDIKDKL<br>TEKFRYCEEKLLNLSFYNHQDKKKEESIRVFLIRDKNS |

[0582]

[0583]

| | |
|---|---|
| | DNFKFESILDDGSNKIFISKNGKEITIQCCDKVLETLIIEK NTLKISSNGKIISLIPHYSYSIDVKY |
| WP_ 0964 0205 0 | MENKNKPNRGSIVRIIISNYDMKGIKELKVRYRKQAQL DTFILQTTLDKSNNSILINDFRVKAREKYRYSFTYDGK EKFSVPSNSIIVTKIDNAAPEKSKEIRKYKITLGIDEKCK TGSMITAAIEDLLEDDRVREGIRNPRRKASKTERKLITE SICHNYAQITQCPVEEIDAVKIYKVKRFLSYRSNMLLFF ALINDFLCKNLKNEKGEKINEIWEMENKGNNKKIDFD ENYNILVAQIKEYFTKEIENYNNRIDNIIDKKELLKYSE EKEESEKNKKLEELNKLESQKLKILTDEEIKADVIKIIKI FSDLRHSLMHYEYKYFENLFENKKNEELAELLNLNLF KNLTLLRQMKIENKTNYLEGDEKFNILGKDVRAKNAL GHYDLLVEQKNGFNNFINSFFVQDGTENLEFKKFIDEN FIKAQKELEEDIKNCKESVKKLEKKLKENPKKSEDLEK KLEKKQKKLKELKKELEKMKQHYKRLNCAYVWDIHS STVYKKLYNERKNLIEKYNKQLNGLQDKNAITGINAQ LLRIKKEMEEITKSNSLFRLKYKMQIAYAFLEMEYEGN IAKFKNEFDCSKTEKIQEWLEKSEEYLNYCMEKEEDG KNYKFHFKEISEIKDTHNEEWLENTSENNLFKFYILTYL LLPMEFKGDFLGVVKKHYYDIKNVDFTDESEKELSQE QIDKMIGDSFFHKIRLFEKNTKRYEIIKYSILTSDEIKKY FELLELKVPYLEYKGIDEIGIFNKNIILPIFKYYQIIFRLY NDLEIHGLFNVSFDINKILSDLKSYGNENINFREFLYVI KQNNNSSTEEEYQKIWEKLESKYLKEPLLTPEKKEINK KTEKELKKLDGISFLRNKISHLEYEKIIEGVLKTAVNGE NKKTSETNADKVFLNEKIKKIINFIKENELDKIELGFNFI NDFFMKKEQFMFGQIKQVKEGNSDSITTERKRKEENN KRLKITYGLNYNNLSKIYEFSNTLREIVNSPLFLKDSTL LKKVDLSKVMLKEKPICSLQYENNTKLEDDIKRILLKD SSDIMGIYKAEVVKKLKEKLVLIFKYDEEKKIYVTVYD TSKAVPENISKEILVKRNNSKEEYFFEDNKKKYTTQYY TLEITKENELKVIPAKKLEGKEFKTEKKEENKLMLNNH YCFNVKIIY |

[0584] 在一些实施方案中，AD功能化的CRISPR-Cas系统的组分可以各种形式(诸如DNA/RNA或RNA/RNA或蛋白质RNA的组合)递送。例如，Cas13蛋白可作为编码DNA的多核苷酸或编码RNA的多核苷酸或作为蛋白质递送。指导物可作为编码DNA的多核苷酸或RNA递送。设想了所有可能的组合，包括混合的递送形式。

[0585] 在一些方面，本发明提供了包括将一种或多种多核苷酸诸如本文所述的一个或多个载体、其一种或多种转录物和/或从其转录的一种或多种蛋白递送到宿主细胞的方法。

[0586] 载体

[0587] 一般来讲，术语“载体”是指能够转运已与其连接的另一核酸的核酸分子。载体是复制子，诸如质粒、噬菌体或粘粒，可在其中插入另一个DNA片段，以实现所插入片段的复制。一般来讲，载体在与适当的控制元件结合时能够复制。载体包括但不限于单链、双链或

部分双链的核酸分子。包含一个或多个自由端、不包含自由端(例如,呈环状)的核酸分子;包含DNA、RNA或两者的核酸分子;以及本领域已知的其他多核苷酸变体。载体的一种类型是“质粒”,是指环状双链DNA环,可以诸如通过标准分子克隆技术在其中插入其他DNA片段。载体的另一种类型是病毒载体,其中病毒来源的DNA或RNA序列存在于该载体中以包装到病毒(例如,逆转录病毒、复制缺陷型逆转录病毒、腺病毒、复制缺陷型腺病毒和腺相关病毒)中。病毒载体还包括病毒携带的用于转染到宿主细胞中的多核苷酸。某些载体能够在引入它们的宿主细胞中自主复制(例如,具有细菌复制起点的细菌载体和游离型哺乳动物载体)。其他载体(例如,非游离型哺乳动物载体)在引入宿主细胞后整合到宿主细胞的基因组中,从而与宿主基因组一起复制。此外,某些载体能够指导与其可操作地连接的基因表达。这类载体在本文中称为“表达载体”。用于并导致在真核细胞中表达的载体在本文中可以称为“真核表达载体”。在重组DNA技术中实用的通用表达载体通常是质粒的形式。

[0588] 重组表达载体可以以适合在宿主细胞中表达本发明的核酸的形式包含该核酸,这意味着重组表达载体包括与要表达的核酸序列可操作地连接的一个或多个调控元件,可根据要用于表达的宿主细胞来选择这些调控元件。在重组表达载体内,“可操作地连接”是指目标核苷酸序列以允许核苷酸序列表达的方式(例如,在体外转录/翻译系统中,或当将载体引入宿主细胞时,在宿主细胞中)与一个或多个调控元件连接。有利的载体包括慢病毒和腺相关病毒,并且也可以针对靶向特定类型的细胞来选择这类载体的类型。

[0589] 关于重组和克隆方法,提到了2004年9月2日公开为US 2004-0171156A1的美国专利申请10/815,730,其内容全文通过引用并入本文。

[0590] 术语“调控元件”是指包括启动子、增强子、内部核糖体进入位点(IRES)和其他表达控制元件(例如,转录终止信号,诸如聚腺苷酸化信号和poly-U序列)。例如,在Goeddel,GENE EXPRESSION TECHNOLOGY:METHODS IN ENZYMOLOGY 185,Academic Press,San Diego,Calif.(1990)中描述了这类调控元件。调控元件包括指导核苷酸序列在许多类型的宿主细胞中指导组成性表达的调控元件以及指导核苷酸序列(例如,组织特异性调控序列)仅在某些宿主细胞中表达的调控元件。组织特异性启动子可指导主要在所需的目标组织中表达,诸如肌肉、神经元、骨骼、皮肤、血液、特定器官(例如,肝脏、胰腺)或特定细胞类型(例如,淋巴细胞)。调控元件也可以时间依赖性方式指导表达,例如以细胞周期依赖性或发育阶段依赖性方式,这些方式可以是或也可以不是组织或细胞类型特异性的。在一些实施方案中,载体包含一个或多个pol III启动子(例如,1、2、3、4、5个或更多个pol III启动子)、一个或多个pol II启动子(例如,1、2、3、4、5个或更多个pol II启动子)、一个或多个pol I启动子(例如,1、2、3、4、5个或更多个pol I启动子)或其组合。pol III启动子的实例包括但不限于U6和H1启动子。pol II启动子的实例包括但不限于逆转录病毒劳斯肉瘤病毒(RSV)LTR启动子(任选地带有RSV增强子)、巨细胞病毒(CMV)启动子(任选地带有CMV增强子)[参见例如Boshart等人,Cell,41:521-530(1985)]、SV40启动子、二氢叶酸还原酶启动子、β-肌动蛋白启动子、磷酸甘油激酶(PGK)启动子和EF1α启动子。术语“调控元件”还包括增强子元件,诸如WPRE;CMV增强子;HTLV-I的LTR中的R-U5'片段(Mol.Cell.Biol.,第8(1)卷,第466-472页,1988);SV40增强子;以及兔β-珠蛋白的外显子2和3之间的内含子序列(Proc.Natl.Acad.Sci.USA.,第78(3)卷,第1527-31页,1981)。本领域技术人员应当理解,表达载体的设计可以取决于诸如要转化的宿主细胞的选择、所需的表达水平等因素。可以

将载体引入宿主细胞，从而产生由本文所述的核酸编码的转录物、蛋白质或肽，包括融合蛋白或肽(例如，成簇规律间隔短回文重复序列(CRISPR)转录物、蛋白质、酶、其突变体形式、其融合蛋白等)。关于调控序列，提到了美国专利申请10/491,026，该申请的内容全文通过引用并入本文。关于启动子，提到了PCT公开WO 2011/028929和美国申请12/511,940，这些文献的内容全文通过引用并入本文。

[0591] 有利的载体包括慢病毒和腺相关病毒，并且也可以针对靶向特定类型的细胞来选择这类载体的类型。

[0592] 在特定的实施方案中，针对与腺苷脱氨酶融合的指导RNA和(任选地经修饰或突变的)CRISPR-Cas蛋白，使用双顺反子载体。与腺苷脱氨酶融合的指导RNA和(任选地经修饰或突变的)CRISPR-Cas蛋白的双顺反子表达载体是优选的。一般来讲，特别是在该实施方案中，与腺苷脱氨酶融合的(任选地经修饰或突变的)CRISPR-Cas蛋白优选地由CBh启动子驱动。RNA可优选地由Pol III启动子诸如U6启动子驱动。理想情况下，将两者结合。

[0593] 载体可以被设计用于在原核或真核细胞中表达CRISPR转录物(例如，核酸转录物、蛋白质或酶)。例如，CRISPR转录物可以在细菌细胞诸如大肠埃希氏菌(Escherichia coli)、昆虫细胞(使用杆状病毒表达载体)、酵母细胞或哺乳动物细胞中表达。合适的宿主细胞在Goeddel,GENE EXPRESSION TECHNOLOGY:METHODS IN ENZYMOLOGY 185,Academic Press,San Diego,Calif.(1990)中进一步讨论。另选地，重组表达载体可以例如使用T7启动子调控序列和T7聚合酶在体外转录和翻译。

[0594] 载体可被引入原核生物或原核细胞中并在其中繁殖。在一些实施方案中，原核生物用于扩增将要引入真核细胞中的载体的拷贝或者用作要引入真核细胞中的载体的产生中的中间载体(例如，扩增质粒，作为病毒载体包装系统的一部分)。在一些实施方案中，原核生物用于扩增载体的拷贝并表达一种或多种核酸，诸如以提供一种或多种蛋白质的来源以递送到宿主细胞或宿主生物。蛋白质在原核生物中的表达最经常在大肠埃希氏菌中用含有指导融合或非融合蛋白表达的组成型或诱导型启动子的载体来进行。融合载体将许多氨基酸添加到其中编码的蛋白质上，诸如添加到重组蛋白质的氨基末端上。此类融合载体可用于一个或多个目的，诸如：(i)增加重组蛋白的表达；(ii)增加重组蛋白的溶解度；和(iii)通过在亲和纯化中充当配体来帮助纯化重组蛋白。通常，在融合表达载体中，在融合部分和重组蛋白的接合处引入蛋白水解裂解位点，以使得在纯化融合蛋白之后能够从融合部分分离重组蛋白。此类酶及其同源识别序列包括Xa因子、凝血酶和肠激酶。示例融合表达载体包括pGEX(Pharmacia Biotech Inc;Smith and Johnson,1988.Gene 67:31-40)、pMAL(New England Biolabs,Beverly,Mass.)和pRIT5(Pharmacia,Piscataway,N.J.)，它们分别将谷胱甘肽S-转移酶(GST)、麦芽糖E结合蛋白或蛋白A与靶重组蛋白融合。合适的诱导型非融合大肠埃希氏菌表达载体的实例包括pTrc(Amrann等人,(1988)Gene 69:301-315)和pET 11d(Studier等人,GENE EXPRESSION TECHNOLOGY:METHODS IN ENZYMOLOGY 185,Academic Press,San Diego,Calif.(1990)60-89)。在一些实施方案中，载体是酵母表达载体。用于在酿酒酵母(Saccharomyces cerivisae)中表达的载体的实例包括pYepSec1(Baldari等人,1987.EMBO J.6:229-234)、pMFa(Kuijan和Herskowitz,1982.Cell 30:933-943)、pJRY88(Schultz等人,1987.Gene 54:113-123)、pYES2(Invitrogen Corporation,San Diego,Calif.)和picZ(InVitrogen Corp,San Diego,Calif.)。在一些实施方案中，载

体使用杆状病毒表达载体驱动昆虫细胞中的蛋白表达。可用于在培养的昆虫细胞(例如,SF9细胞)中表达蛋白质的杆状病毒载体包括pAc系列(Smith等人,1983.Mol.Cell.Biol.3:2156-2165)和pVL系列(Lucklow和Summers,1989.Virology 170:31-39)。

[0595] 在一些实施方案中,载体能够使用哺乳动物表达载体来驱动哺乳动物细胞中一个或多个序列的表达。哺乳动物表达载体的实例包括pCDM8(Seed,1987.Nature 329:840)和pMT2PC(Kaufman等人,1987.EMBO J.6:187-195)。当用于哺乳动物细胞时,表达载体的控制功能通常由一个或多个调控元件提供。例如,常用的启动子来源于多瘤病毒、腺病毒2、巨细胞病毒、猿猴病毒40以及本文公开和本领域已知的其他病毒。对于用于原核和真核细胞的其他合适的表达系统,参见例如Sambrook等人,MOLECULAR CLONING:A LABORATORY MANUAL.第2版,Cold Spring Harbor Laboratory,Cold Spring Harbor Laboratory Press,Cold Spring Harbor,N.Y.,1989的第16章和第17章。

[0596] 在一些实施方案中,重组哺乳动物表达载体能够指导核酸优先在特定细胞类型中表达(例如,使用组织特异性调控元件来表达核酸)。组织特异性调控元件是本领域已知的。合适的组织特异性启动子的非限制性实例包括白蛋白启动子(具有肝脏特异性;Pinkert等人,1987.Genes Dev.1:268-277)、淋巴特异性启动子(Calame和Eaton,1988.Adv.Immunol.43:235-275)(特别是T细胞受体的启动子(Winoto和Baltimore,1989.EMBO J.8:729-733)和免疫球蛋白的启动子(Baneiji等人,1983.Cell 33:729-740;Queen和Baltimore,1983.Cell 33:741-748))、神经元特异性启动子(例如,神经丝启动子;Byrne和Ruddle,1989.Proc.Natl.Acad.Sci.USA 86:5473-5477)、胰腺特异性启动子(Edlund等人,1985.Science 230:912-916)和乳腺特异性启动子(例如,乳清启动子;美国专利号4,873,316和欧洲申请公开号264,166)。还包括发育型调控启动子,例如鼠类hox启动子(Kessel和Gruss,1990.Science 249:374-379)和α-甲胎蛋白启动子(Campes和Tilghman,1989.Genes Dev.3:537-546)。关于这些原核和真核载体,提到了美国专利6,750,059,其内容全文通过引用并入本文。本发明的其他实施方案可涉及病毒载体的用途,关于这在一方面,提到了美国专利申请13/092,085,其内容全文通过引用并入本文。组织特异性调控元件是本领域已知的,在这在一方面,提到了美国专利7,776,321,其内容全文通过引用并入本文。在一些实施方案中,调控元件可操作地连接到CRISPR系统的一个或多个元件,以便驱动CRISPR系统的一个或多个元件表达。

[0597] 在一些实施方案中,将驱动核酸靶向系统的一个或多个元件表达的一个或多个载体引入宿主细胞,使得核酸靶向系统的元件的表达指导核酸靶向复合物在一个或多个靶位点处形成。例如,核酸靶向效应酶和核酸靶向指导RNA可以各自可操作地连接到单独载体上的单独调控元件。可以将核酸靶向系统的一个或多个RNA递送到转基因核酸靶向效应蛋白动物或哺乳动物,例如组成性或诱导性或条件性表达核酸靶向效应蛋白的动物或哺乳动物;或者以其他方式表达核酸靶向效应蛋白或诸如通过事先施用编码和体内表达核酸靶向效应蛋白的一个或多个载体而具有含有核酸靶向效应蛋白的细胞的动物或哺乳动物。另选地,可将由相同或不同调控元件表达的两个或更多个元件在单个载体中组合,而一个或多个附加载体提供不包括在第一载体中的核酸靶向系统的任何组分。可将在单个载体中组合的核酸靶向系统元件以任何合适的取向布置,诸如一个元件相对于第二元件(的“上游”)位于5'或相对于第二元件(的“下游”)位于3'。一个元件的编码序列可位于第二元件的编码序

列的相同或相反链上,并且以相同或相反的方向定向。在一些实施方案中,单个启动子驱动编码嵌入一个或多个内含子序列内(例如,每个在不同内含子中,两个或更多个在至少一个内含子中,或者全部在单个内含子中)的核酸靶向效应蛋白和核酸靶向指导RNA的转录物的表达。在一些实施方案中,核酸靶向效应蛋白和核酸靶向指导RNA可以可操作地连接到相同启动子并由该启动子表达。用于表达核酸靶向系统的一个或多个元件的递送媒介物、载体、颗粒、纳米颗粒、制剂及其组分如前述文献诸如WO 2014/093622(PCT/US2013/074667)中所用。在一些实施方案中,载体包含一个或多个插入位点,诸如限制性核酸内切酶识别序列(也称为"克隆位点")。在一些实施方案中,一个或多个插入位点(例如,约或大于约1、2、3、4、5、6、7、8、9、10个或更多个插入位点)位于一个或多个载体的一个或多个序列元件的上游和/或下游。当使用多个不同的指导序列时,单个表达构建体可用于将核酸靶向活性靶向细胞内的多个不同的对应靶序列。例如,单个载体可包含约或大于约1、2、3、4、5、6、7、8、9、10、15、20个或更多个指导序列。在一些实施方案中,可提供约或大于约1、2、3、4、5、6、7、8、9、10个或更多个这类含指导序列的载体,并任选地将其递送到细胞。在一些实施方案中,载体包含可操作地连接到编码核酸靶向效应蛋白的酶编码序列的调控元件。可以分开递送核酸靶向效应蛋白或者一个或多个核酸靶向指导RNA;并且有利地,经由颗粒复合物递送这些物质中的至少一种。可以在核酸靶向指导RNA之前递送核酸靶向效应蛋白mRNA,以留出时间表达核酸靶向效应蛋白。可以在施用核酸靶向指导RNA前1-12小时(优选约2-6小时)施用核酸靶向效应蛋白mRNA。另选地,可以一起施用核酸靶向效应蛋白mRNA和核酸靶向指导RNA。有利地,可以在初次施用核酸靶向效应蛋白mRNA+指导RNA后1-12小时(优选约2-6小时)施用第二加强剂量的指导RNA。额外施用核酸靶向效应蛋白mRNA和/或指导RNA可能有助于获得最有效的基因组修饰水平。

[0598] 常规的基于病毒和非病毒的基因转移方法可以用于将核酸引入哺乳动物细胞或靶组织。此类方法可以用于向培养物中或宿主生物中的细胞施用编码核酸靶向系统的组分的核酸。非病毒载体递送系统包括DNA质粒、RNA(例如,本文所述的载体的转录物)、裸核酸以及与递送媒介物诸如脂质体复合的核酸。病毒载体递送系统包括DNA和RNA病毒,它们在被递送到细胞后具有游离或整合基因组。有关基因治疗程序的综述,参见Anderson,Science256:808-813(1992);Nabel&Felgner,TIBTECH 11:211-217(1993);Mitani&Caskey,TIBTECH 11:162-166(1993);Dillon,TIBTECH 11:167-175(1993);Miller,Nature 357:455-460(1992);Van Brunt,Biotechnology 6(10):1149-1154(1988);Vigne,Restorative Neurology and Neuroscience 8:35-36(1995);Kremer&Perricaudet,British Medical Bulletin 51(1):31-44(1995);Haddada等人,Current Topics in Microbiology and Immunology,Doerfler and Böhm编(1995);以及Yu等人,Gene Therapy 1:13-26(1994)。

[0599] 核酸的非病毒递送方法包括脂质转染、核转染、显微注射、生物弹射、病毒颗粒、脂质体、免疫脂质体、聚阳离子或脂质:核酸缀合物、裸DNA、人工病毒体和试剂增强的DNA摄取。脂质转染在例如美国专利号5,049,386、4,946,787和4,897,355中描述,并且脂质转染试剂在商业上有售(例如,Transfectam™和Lipofectin™)。适用于多核苷酸的有效受体识别脂质转染的阳离子和中性脂质包括Felgner,WO 91/17424、WO 91/16024中的那些脂质。可以递送到细胞(例如,体外或离体施用)或靶组织(例如,体内施用)。

[0600] 质粒递送涉及将指导RNA克隆到表达CRISPR-Cas蛋白的质粒中，并在细胞培养物中转染DNA。质粒主链可商购获得，不需要特定的设备。它们具有模块化的优势，能够携带不同大小的CRISPR-Cas编码序列(包括编码更大尺寸蛋白质的序列)以及选择标记。质粒的两个优点是它们可以确保瞬时但持续表达。然而，质粒的递送并不直接，使得体内效率通常很低。持续表达也可能是不利的，因为它可能增加脱靶编辑。另外，CRISPR-Cas蛋白的过量积累可能对细胞有毒性。最后，质粒总是具有dsDNA在宿主基因组中随机整合的风险，尤其是考虑到产生双链断裂(中靶和脱靶)时。

[0601] 脂质:核酸复合物(包括靶向脂质体，诸如免疫脂质复合物)的制备是本领域技术人员熟知的(参见例如Crystal,Science 270:404-410(1995);Blaese等人,Cancer Gene Ther.2:291-297(1995);Behr等人,Bioconjugate Chem.5:382-389(1994);Remy等人,Bioconjugate Chem.5:647-654(1994);Gao等人,Gene Therapy 2:710-722(1995);Ahmad等人,Cancer Res.52:4817-4820(1992);美国专利号4,186,183、4,217,344、4,235,871、4,261,975、4,485,054、4,501,728、4,774,085、4,837,028和4,946,787)。这将在下面更详细地讨论。

[0602] 使用基于RNA或DNA病毒的系统来递送核酸利用了用于将病毒靶向体内的特定细胞并将病毒有效载荷转运到细胞核的高度进化过程。可以将病毒载体直接施用于患者(体内)，或者可以将它们用于体外治疗细胞，并且可任选地将经修饰的细胞施用于患者(离体)。基于病毒的常规系统可以包括逆转录病毒载体、慢病毒载体、腺病毒载体、腺相关病毒载体和单纯疱疹病毒载体用于基因转移。可以用逆转录病毒、慢病毒和腺相关病毒基因转移方法来整合到宿主基因组中，通常会导致插入的转基因长期表达。另外，已经在许多不同的细胞类型和靶组织中观察到高转导效率。

[0603] 逆转录病毒的趋向性可以通过掺入外来包膜蛋白、扩大靶细胞的潜在靶群体来改变。慢病毒载体是能够转导或感染非分裂细胞并通常产生高病毒滴度的逆转录病毒载体。因此，逆转录病毒基因转移系统的选择将取决于靶组织。逆转录病毒载体由顺式作用的长末端重复序列组成，其包装能力可容纳6-10kb的外源序列。最小的顺式作用LTR足以复制和包装载体，然后所述载体用于将治疗性基因整合到靶细胞中以提供永久性转基因表达。广泛使用的逆转录病毒载体包括基于鼠白血病病毒(MuLV)、长臂猿白血病病毒(GaLV)、猿猴免疫缺陷病毒(SIV)、人免疫缺陷病毒(HIV)及其组合的逆转录病毒载体(参见例如Buchscher等人,J.Virol.66:2731-2739(1992);Johann等人,J.Virol.66:1635-1640(1992);Sommnerfelt等人,Virol.176:58-59(1990);Wilson等人,J.Virol.63:2374-2378(1989);Miller等人,J.Virol.65:2220-2224(1991);PCT/US94/05700)。

[0604] 在优选瞬时表达的应用中，可使用基于腺病毒的系统。基于腺病毒的载体能够在许多细胞类型中具有很高的转导效率，并且不需要细胞分裂。使用此类载体，已经获得了高滴度和表达水平。该载体可以在相对简单的系统中大量产生。腺相关病毒(“AAV”)载体还可用于例如在核酸和肽的体外生产中以及在体内和离体基因治疗程序中用靶核酸转导细胞(参见例如West等人,Virology 160:38-47(1987);美国专利号4,797,368;WO 93/24641;Kotin,Human Gene Therapy 5:793-801(1994);Muzyczka,J.Clin.Invest.94:1351(1994)。重组AAV载体的构建在许多出版物中描述，包括美国专利号5,173,414;Tratschin等人,Mol.Cell.Biol.5:3251-3260(1985);Tratschin等人,Mol.Cell.Biol.4:2072-2081

(1984);Hermonat&Muzyczka,PNAS 81:6466-6470(1984);以及Samulski等人,J.Virol.63:03822-3828(1989)。

[0605] 本发明提供了:一种AAV,所述AAV包含编码CRISPR系统的外源核酸分子或基本上由所述外源核酸分子组成,所述外源核酸分子为例如包含第一盒或由所述第一盒组成的多个盒,所述第一盒包含启动子、编码CRISPR相关的(Cas)蛋白(推定的核酸酶或解旋酶蛋白,例如Cas13)的核酸分子以及终止子或者基本上由这些组分组成,并且一个或多个(有利地达到载体的包装大小限制,例如总共(包括第一盒)五个)盒包含启动子、编码指导RNA(gRNA)的核酸分子以及终止子(例如,每个盒示意性地表示为启动子-gRNA1-终止子、启动子-gRNA2-终止子……启动子-gRNA(N)终止子,其中N是可以插入的数字,其是载体包装大小的上限),或者基本上由这些组分组成;或者两种或更多种单独的rAAV,每个rAAV包含CRISPR系统的一个或多于一个盒,例如第一rAAV包含第一盒,所述第一盒包含启动子、编码Cas(例如,Cas(Cas13))的核酸分子以及终止子,或者基本上由这些组分组成,并且第二rAAV包含一个或多个盒,每个盒包含启动子、编码指导RNA(gRNA)的核酸分子以及终止子,或者基本上由这些组分组成(例如,每个盒示意性地表示为启动子-gRNA1-终止子、启动子-gRNA2-终止子……启动子-gRNA(N)-终止子,其中N是可以插入的数字,其是载体包装大小的上限)。另选地,由于Cas13可以加工其自己的crRNA/gRNA,所以单个crRNA/gRNA阵列可以用于多重基因编辑。因此,rAAV可包含单个盒而不是多个盒来递送gRNA,所述单个盒包含启动子、多个crRNA/gRNA以及启动子,或者基本上由这些组分组成(例如,每个盒示意性地表示为启动子-gRNA1-gRNA2……gRNA(N)-终止子,其中N是可以插入的数字,其是载体包装大小的上限)。参见Zetsche等人,Nature Biotechnology 35,31-34(2017),该文献全文通过引用并入本文。由于rAAV是DNA病毒,因此本文关于AAV或rAAV的讨论中的核酸分子有利地是DNA。在一些实施方案中,启动子有利地是人突触蛋白I启动子(hSyn)。用于将核酸递送到细胞的其他方法是本领域技术人员已知的。参见例如US20030087817,其通过引用并入本文。

[0606] 在另一个实施方案中,考虑了Cocal水泡病毒包膜假型逆转录病毒载体颗粒(参见例如转让给Fred Hutchinson癌症研究中心的美国专利公开号20120164118)。Cocal病毒属于水泡病毒属,是哺乳动物水泡性口炎的病原体。Cocal病毒最初是从特立尼达的螨虫中分离出来的(Jonkers等人,Am.J.Vet.Res.25:236-242(1964)),并且在特立尼达、巴西和阿根廷已经确定了由昆虫、牛和马引起的感染。已经从自然感染的节肢动物中分离出了许多感染哺乳动物的水泡病毒,这表明它们是媒介传播的。在居住在农村地区(其中病毒是地方性的和实验室获得的)的人群中,普遍存在水泡病毒的抗体。人类感染通常会导致类似流感的症状。Cocal病毒包膜糖蛋白在氨基酸水平上与印第安纳VSV-G共有71.5%的同一性,并且水泡病毒的包膜基因的系统发生比较表明,在水泡病毒中,Cocal病毒与印第安纳VSV-G株在血清学上有所区别,但与其最密切相关。Jonkers等人,Am.J.Vet.Res.25:236-242(1964)以及Travassos da Rosa等人,Am.J.Tropical Med.&Hygiene 33:999-1006(1984)。Cocal水泡病毒包膜假型逆转录病毒载体颗粒可包括例如慢病毒、α逆转录病毒、β逆转录病毒、γ逆转录病毒、δ逆转录病毒和ε逆转录病毒载体颗粒,这些载体颗粒可包含逆转录病毒Gag、Pol和/或一种或多种辅助蛋白和Cocal水泡病毒包膜蛋白。在这些实施方案的某些方面,Gag、Pol和辅助蛋白是慢病毒和/或γ逆转录病毒。

[0607] 在一些实施方案中，用本文所述的一个或多个载体瞬时或非瞬时转染宿主细胞。在一些实施方案中，当细胞天然存在于受试者中，将其转染以任选地重新引入其中。在一些实施方案中，转染的细胞取自受试者。在一些实施方案中，细胞来源于取自受试者的细胞，诸如细胞系。用于组织培养的各种细胞系是本领域已知的。细胞系的实例包括但不限于C8161、CCRF-CEM、MOLT、mIMCD-3、NHDF、HeLa-S3、Huh1、Huh4、Huh7、HUVEC、HASMC、HEKn、HEKa、MiaPaCell、Panc1、PC-3、TF1、CTLL-2、C1R、Rat6、CV1、RPTE、A10、T24、J82、A375、ARH-77、Calu1、SW480、SW620、SKOV3、SK-UT、CaCo2、P388D1、SEM-K2、WEHI-231、HB56、TIB55、Jurkat、J45.01、LRMB、Bcl-1、BC-3、IC21、DLD2、Raw264.7、NRK、NRK-52E、MRC5、MEF、Hep G2、HeLa B、HeLa T4、COS、COS-1、COS-6、COS-M6A、BS-C-1猴肾上皮、BALB/3T3小鼠胚胎成纤维细胞、3T3Swiss、3T3-L1、132-d5人类胎儿成纤维细胞;10.1小鼠成纤维细胞、293-T、3T3、721、9L、A2780、A2780ADR、A2780cis、A172、A20、A253、A431、A-549、ALC、B16、B35、BCP-1细胞、BEAS-2B、bEnd.3、BHK-21、BR 293、BxPC3、C3H-10T1/2、C6/36、Cal-27、CHO、CHO-7、CHO-IR、CHO-K1、CHO-K2、CHO-T、CHO Dhfr-/-、COR-L23、COR-L23/CPR、COR-L23/5010、COR-L23/R23、COS-7、COV-434、CML T1、CMT、CT26、D17、DH82、DU145、DuCaP、EL4、EM2、EM3、EMT6/AR1、EMT6/AR10.0、FM3、H1299、H69、HB54、HB55、HCA2、HEK-293、HeLa、Hepa1c1c7、HL-60、HMEC、HT-29、Jurkat、JY细胞、K562细胞、Ku812、KCL22、KG1、KYO1、LNCap、Ma-Mel 1-48、MC-38、MCF-7、MCF-10A、MDA-MB-231、MDA-MB-468、MDA-MB-435、MDCK II、MDCK II、MOR/0.2R、MONO-MAC 6、MTD-1A、MyEnd、NCI-H69/CPR、NCI-H69/LX10、NCI-H69/LX20、NCI-H69/LX4、NIH-3T3、NALM-1、NW-145、OPCN/OPCT细胞系、Peer、PNT-1A/PNT 2、RenCa、RIN-5F、RMA/RMAS、Saos-2细胞、Sf-9、SkBr3、T2、T-47D、T84、THP1细胞系、U373、U87、U937、VCaP、Vero细胞、WM39、WT-49、X63、YAC-1、YAR以及它们的转基因变体。细胞系可从本领域技术人员已知的多种来源获得(参见例如美国模式培养物保藏所(ATCC)(Manassus,Va.))。

[0608] 在特定的实施方案中，AD功能化的CRISPR系统的一种或多种组分的瞬时表达和/或存在对于诸如减少脱靶效应可能是有意义的。在一些实施方案中，用本文所述的一个或多个载体转染的细胞用于建立包含一个或多个载体来源的序列的新细胞系。在一些实施方案中，用本文所述的AD功能化的CRISPR系统的组分瞬时转染(诸如通过一个或多个载体瞬时转染，或用RNA转染)并且通过CRISPR复合物的活性修饰的细胞用于建立包含以下细胞的新细胞系:所述细胞含有所述修饰但缺少任何其他外源序列。在一些实施方案中，用本文所述的一个或多个载体瞬时或非瞬时转染的细胞或者来源于此类细胞的细胞系用于评估一种或多种测试化合物。

[0609] 在一些实施方案中，设想了将RNA和/或蛋白质直接引入宿主细胞。例如，可以将CRISPR-Cas蛋白作为编码mRNA与体外转录的指导RNA一起递送。此类方法可减少确保CRISPR-Cas蛋白作用的时间，并进一步防止了CRISPR系统组分的长期表达。

[0610] 在一些实施方案中，本发明的RNA分子以脂质体或脂质转染蛋白制剂等形式递送，并且可以通过本领域技术人员众所周知的方法制备。此类方法例如在美国专利号5,593,972、5,589,466和5,580,859中描述，这些专利通过引用并入本文。已经开发了专门针对增强和改善将siRNA递送到哺乳动物细胞中的递送系统(参见例如Shen等人,FEBS Let.2003,539:111-114;Xia等人,Nat.Biotech.2002,20:1006-1010;Reich等人,Mol.Vision.2003,9:210-216;Sorensen等人,J.Mol.Biol.2003,327:761-766;Lewis等人,Nat.Gen.2002,32:

107-108和Simeoni等人,NAR 2003,31,11:2717-2724)),这些递送系统可应用于本发明。siRNA最近已成功用于抑制灵长类动物中的基因表达(参见例如Tolentino等人,Retina 24(4):660),其也可应用于本发明。

[0611] 实际上,RNA递送是有用的体内递送方法。使用脂质体或纳米颗粒可以将Cas13、腺苷脱氨酶和指导RNA递送到细胞中。因此,本发明的CRISPR-Cas蛋白(诸如Cas13)的递送、腺苷脱氨酶(其可与CRISPR-Cas蛋白或衔接蛋白融合)的递送以及/或者RNA的递送可以RNA形式并通过微泡、脂质体或者一种或多种颗粒进行。例如,可以将Cas13mRNA、腺苷脱氨酶mRNA和指导RNA包装到脂质体颗粒中以用于体内递送。脂质体转染试剂诸如Life Technologies的lipofectamine以及市场上的其他试剂可以有效地将RNA分子输送到肝脏中。

[0612] RNA的递送手段还优选地包括经由颗粒(Cho,S.,Goldberg,M.,Son,S.,Xu,Q.,Yang,F.,Mei,Y.,Bogatyrev,S.,Langer,R.和Anderson,D.,Lipid-like nanoparticles for small interfering RNAdelivery to endothelial cells,Advanced Functional Materials,19:3112-3118,2010)或外来体(Schroeder,A.,Levins,C.,Cortez,C.,Langer,R.,和Anderson,D.,Lipid-based nanotherapeutics for siRNA delivery,Journal of Internal Medicine,267:9-21,2010,PMID:20059641)进行RNA递送。实际上,已经表明外来体在递送siRNA中特别有用,它是一种与CRISPR系统有一些相似之处的系统。例如,El-Andaloussi S等人("Exosome-mediated delivery of siRNAin vitro and in vivo."Nat Protoc.2012年12月;7(12):2112-26.doi:10.1038/nprot.2012.131.2012年11月15日的电子版)描述了外来体作为跨越不同生物屏障进行药物递送的有前景的工具并且可以用来在体内外递送siRNA的原因。他们的方法是通过转染包含与肽配体融合的外来体蛋白的表达载体来产生靶向的外来体。然后从转染的细胞上清液中纯化外来体并进行表征,然后将RNA加载到外来体中。根据本发明的递送或施用可以用外来体进行,特别是但不限于脑。可将维生素E(α生育酚)与CRISPR Cas缀合,并且例如以与Uno等人将短干扰RNA(siRNA)递送到脑类似的方式(HUMAN GENE THERAPY 22:711-719(2011年6月))将其与高密度脂蛋白(HDL)一起递送到脑。经由渗透微型泵(型号1007D;Alzet,Cupertino,CA)对小鼠进行输注,该渗透微型泵装满磷酸盐缓冲液(PBS)或游离TocsiBACE或Toc-siBACE/HDL并与Brain Infusion Kit 3(Alzet)连接。将脑输注套管放置在中线处,位于前卤后约0.5mm,以便输注到背侧第三脑室中。Uno等人发现,使用相同的ICV输注方法,低至3nmol的Toc-siRNA与HDL一起可以诱导程度相当的靶标降低。在本发明中,可考虑人类中靶向脑的缀合到α-生育酚并与HDL共同施用的CRISPR Cas的相似剂量,例如可考虑约3nmol至约3μmol的靶向脑的CRISPR Cas。Zou等人((HUMAN GENE THERAPY 22:465-475(2011年4月))描述了一种慢病毒介导的递送靶向PKCγ的短发夹RNA的方法,用于在大鼠脊髓中进行体内基因沉默。Zou等人通过鞘内导管施用约10μl效价为1×109转导单位(TU)/ml的重组慢病毒。在本发明中,可考虑人类中靶向脑的在慢病毒载体中表达的CRISPR Cas的相似剂量,例如可考虑约10-50ml效价为$1\times10^9$转导单位(TU)/ml的靶向脑的CRISPR Cas。

[0613] 载体的剂量

[0614] 在一些实施方案中,通过例如肌内注射将载体例如质粒或病毒载体递送到目标组织,而其他时间经由静脉内、透皮、鼻内、口服、粘膜或其他递送方法进行递送。这种递送可以是经由单剂量或多剂量进行的。本领域技术人员应当理解,本文中所递送的实际剂量可

根据多种因素而有很大变化，诸如载体选择、靶细胞、生物或组织、待治疗受试者的一般状况、所寻求的转化/修饰程度、施用途径、施用方式、所寻求的转化/修饰类型等。

[0615] 这种剂量还可包含例如运载体(水、盐水、乙醇、甘油、乳糖、蔗糖、磷酸钙、明胶、右旋糖酐、琼脂、果胶、花生油、芝麻油等)、稀释剂、药学上可接受的运载体(例如，磷酸盐缓冲盐水)、药学上可接受的赋形剂和/或本领域已知的其他化合物。剂量还可包含一种或多种药学上可接受的盐，诸如无机酸盐，诸如盐酸盐、氢溴酸盐、磷酸盐、硫酸盐等；以及有机酸的盐，诸如乙酸盐、丙酸盐、丙二酸盐、苯甲酸盐等。另外，辅助物质诸如润湿剂或乳化剂、pH缓冲物质、凝胶或胶凝材料、调味剂、着色剂、微球、聚合物、悬浮剂等也可存在于其中。另外，尤其是如果剂型是可重构形式，则还可存在一种或多种其他常规药物成分，诸如防腐剂、湿润剂、悬浮剂、表面活性剂、抗氧化剂、抗结块剂、填充剂、螯合剂、包衣剂、化学稳定剂等。合适的示例性成分包括微晶纤维素、羧甲基纤维素钠、聚山梨醇酯80、苯乙醇、氯丁醇、山梨酸钾、山梨酸、二氧化硫、没食子酸丙酯、对羟基苯甲酸酯、乙基香草醛、甘油、苯酚、对氯苯酚、明胶、白蛋白以及它们的组合。对药学上可接受的赋形剂的详细讨论可在REMINGTON'S PHARMACEUTICAL SCIENCES(Mack Pub.Co.,N.J.1991)中获得，该文献通过引用并入本文。

[0616] 在本文的一个实施方案中，经由腺病毒进行递送，该腺病毒可以是含有至少1×105个腺病毒载体颗粒(也称为颗粒单位，pu)的单次加强剂量。在本文的一个实施方案中，剂量优选地是至少约1×106个腺病毒载体颗粒(例如，约1×106-1×1012个颗粒)、更优选至少约1×107个颗粒、更优选至少约1×108个颗粒(例如，约1×108-1×1011个颗粒或约1×108-1×1012个颗粒)、最优选至少约1×100个颗粒(例如，约1×109-1×1010个颗粒或约1×109-1×1012个颗粒)或甚至至少约1×1010个颗粒(例如，约1×1010-1×1012个颗粒)。另选地，剂量包含不超过约1×1014个颗粒，优选不超过约1×1013个颗粒，甚至更优选不超过约1×1012个颗粒，甚至更优选不超过约1×1011个颗粒，并且最优选不超过约1×1010个颗粒(例如，不超过约1×109个颗粒)。因此，剂量可含有单剂量的腺病毒载体，其具有例如约1×106颗粒单位(pu)、约2×106pu、约4×106pu、约1×107pu、约2×107pu、约4×107pu、约1×108pu、约2×108pu、约4×108pu、约1×109pu、约2×109pu、约4×109pu、约1×1010pu、约2×1010pu、约4×1010pu、约1×1011pu、约2×1011pu、约4×1011pu、约1×1012pu、约2×1012pu或约4×1012pu的腺病毒载体。参见例如2013年6月4日授予Nabel等人的美国专利号8,454,972B2(通过引用并入本文)中的腺病毒载体；以及其在第29栏第36-58行的剂量。在本文的一个实施方案中，腺病毒经由多次剂量递送。

[0617] 在本文的一个实施方案中，经由AAV进行递送。据认为用于体内向人体内递送AAV的治疗有效剂量的范围为约20至约50ml盐水溶液，其含有约1×1010至约1×1010个功能性AAV/ml溶液。可以调节剂量以平衡治疗益处与任何副作用。在本文的一个实施方案中，AAV剂量的浓度范围一般为约1×105至1×1050个基因组AAV、约1×108至1×1020个基因组AAV、约1×1010至约1×1016个基因组，或约1×1011至约1×1016个基因组AAV。人类剂量可为约1×1013个基因组AAV。此类浓度可以约0.001ml至约100ml、约0.05至约50ml或约10至约25ml的运载体溶液递送。通过建立剂量响应曲线的常规试验，本领域普通技术人员可以容易地确定其他有效剂量。参见例如2013年3月26日授予Hajjar等人的美国专利号8,404,658B2第27栏第45-60行。

[0618] 在本文的一个实施方案中，经由质粒进行递送。在此类质粒组合物中，剂量应为足以引起响应的质粒量。例如，质粒组合物中质粒DNA的合适量可为每70kg个体约0.1至约2mg或约1μg至约10μg。本发明的质粒一般包含(i)启动子；(ii)与所述启动子可操作地连接的编码CRISPR-Cas蛋白的序列；(iii)选择性标记；(iv)复制起点；以及(v)在(ii)下游并与其可操作地连接的转录终止子。质粒还可以编码CRISPR复合物的RNA组分，但是一种或多种组分可替代地在不同的载体上编码。

[0619] 本文的剂量是基于平均70kg的个体。施用频率在医学或兽医学从业者(例如，医师、兽医)或本领域技术人员的能力范围内。还应当注意，实验中使用的小鼠通常为约20g，并且根据小鼠实验，可以放大到70kg的个体。

[0620] 本文提供的用于组合物的剂量包括用于重复施用或重复给药的剂量。在特定的实施方案中，在数周、数月或数年的时间内重复施用。可以进行合适的测定以获得最佳剂量方案。重复施用可以允许使用较低的剂量，这可以积极影响脱靶修饰。

[0621] RNA递送

[0622] 在特定的实施方案中，使用基于RNA的递送。在这些实施方案中，将CRISPR-Cas蛋白的mRNA、腺苷脱氨酶(其可与CRISPR-Cas蛋白或衔接蛋白融合)的mRNA与体外转录的指导RNA一起递送。Liang等人描述了使用基于RNA的递送进行有效的基因组编辑(Protein Cell.2015年5月;6(5):363-372)。在一些实施方案中，可以对编码Cas13和/或腺苷脱氨酶的一种或多种mRNA进行化学修饰，与质粒编码的Cas13和/或腺苷脱氨酶相比，这可使得活性提高。例如，可以用假尿苷(Ψ)、N1-甲基假尿苷(me1Ψ)、5-甲氧基尿苷(5moU)部分或完全取代mRNA中的尿苷。参见Li等人，Nature Biomedical Engineering 1,0066DOI:10.1038/s41551-017-0066(2017)，该文献全文通过引用并入本文。

[0623] RNP递送

[0624] 在特定的实施方案中，预复合的指导RNA、CRISPR-Cas蛋白和腺苷脱氨酶(其可与CRISPR-Cas蛋白或衔接蛋白融合)作为核糖核蛋白(RNP)递送。RNP的优点在于，它们比RNA方法产生甚至更大的快速编辑效果，因为该过程避免了转录的需要。一个重要的优点是RNP递送都是瞬时的，从而减少了脱靶效应和毒性问题。Kim等人(2014,Genome Res.24(6):1012-9)、Paix等人(2015,Genetics 204(1):47-54)、Chu等人(2016,BMC Biotechnol.16:4)以及Wang等人(2013,Cell.9;153(4):910-8)观察到了不同细胞类型中的有效基因组编辑。

[0625] 在特定的实施方案中，核糖核蛋白通过基于多肽的穿梭剂递送，如WO2016161516中所述。WO2016161516描述了使用包含可操作地连接到细胞穿透结构域(CPD)的内体泄漏结构域(ELD)的合成肽将多肽负荷物的有效转导到富含组氨酸的结构域和CPD。类似地，这些多肽可以用于在真核细胞中递送基于CRISPR效应子的RNP。

[0626] 颗粒

[0627] 在一些方面或实施方案中，可使用包含递送颗粒制剂的组合物。在一些方面或实施方案中，制剂包含CRISPR复合物，该复合物包含CRISPR蛋白以及指导CRISPR复合物与靶序列进行序列特异性结合的指导物。在一些实施方案中，递送颗粒包括基于脂质的颗粒(任选脂质纳米颗粒)或阳离子脂质以及任选可生物降解的聚合物。在一些实施方案中，阳离子脂质包括1,2-二油酰基-3-三甲基铵-丙烷(DOTAP)。在一些实施方案中，亲水性聚合物包括

乙二醇或聚乙二醇。在一些实施方案中，递送颗粒还包括脂蛋白、优选胆固醇。在一些实施方案中，递送颗粒的直径小于500nm，任选地小于250nm，任选地小于100nm，任选地为约35nm至约60nm。

[0628] 在提交于2017年4月14日的标题为“Nove Delivery of Large Payloads”的美国临时申请中进一步公开了示例颗粒递送复合物。

[0629] 已知有多种类型的颗粒递送系统和/或制剂可用于各种生物医学应用。一般来讲，将颗粒定义为在转运和特性方面作为整体表现的小物体。根据直径将颗粒进一步分类。粗颗粒的覆盖范围在2,500至10,000纳米之间。细颗粒的大小在100至2,500纳米之间。超细颗粒或纳米颗粒的大小一般在1至100纳米之间。以100nm作为限值基于以下事实：使粒子区分于大块材料的新颖特性通常是在100nm以下的临界长度尺度下显现的。

[0630] 如本文所用，将颗粒递送系统/制剂定义为包括根据本发明的颗粒的任何生物递送系统/制剂。根据本发明的颗粒是最大尺寸(例如，直径)小于100微米(μm)的任何实体。在一些实施方案中，本发明的颗粒的最大尺寸小于10μm。在一些实施方案中，本发明的颗粒的最大尺寸小于2000纳米(nm)。在一些实施方案中，本发明的颗粒的最大尺寸小于1000纳米(nm)。在一些实施方案中，本发明的颗粒的最大尺寸小于900nm、800nm、700nm、600nm、500nm、400nm、300nm、200nm或100nm。通常，本发明的颗粒的最大尺寸(例如，直径)为500nm或更小。在一些实施方案中，本发明的颗粒的最大尺寸(例如，直径)为250nm或更小。在一些实施方案中，本发明的颗粒的最大尺寸(例如，直径)为200nm或更小。在一些实施方案中，本发明的颗粒的最大尺寸(例如，直径)为150nm或更小。在一些实施方案中，本发明的颗粒的最大尺寸(例如，直径)为100nm或更小。在本发明的一些实施方案中使用较小的颗粒，例如最大尺寸为50nm或更小。在一些实施方案中，本发明的颗粒的最大尺寸在25nm至200nm的范围内。

[0631] 就本发明而言，优选的是具有CRISPR复合物的一种或多种组分，例如CRISPR-Cas蛋白或mRNA，或者腺苷脱氨酶(其可与CRISPR-Cas蛋白或衔接蛋白融合)或mRNA，或者使用纳米颗粒或脂质包膜递送的指导RNA。可结合本发明的纳米颗粒方面使用其他递送系统或载体。

[0632] 一般来讲，“纳米颗粒”是指直径小于1000nm的任何颗粒。在某些优选的实施方案中，本发明的纳米颗粒的最大尺寸(例如，直径)为500nm或更小。在其他优选的实施方案中，本发明的纳米颗粒的最大尺寸在25nm至200nm的范围内。在其他优选的实施方案中，本发明的纳米颗粒的最大尺寸为100nm或更小。在其他优选的实施方案中，本发明的纳米颗粒的最大尺寸在35nm至60nm的范围内。应当理解，在适当的情况下，本文中对颗粒或纳米颗粒的提及可以互换。

[0633] 应当理解，颗粒的大小将取决于在加载之前还是之后测量而有所不同。因此，在特定的实施方案中，术语“纳米颗粒”可仅适用于颗粒预加载。

[0634] 本发明涵盖的纳米颗粒可以不同形式(例如，作为固体纳米颗粒(例如，金属(诸如银、金、铁、钛)、非金属、基于脂质的固体、聚合物)、纳米颗粒的悬浮液或其的组合)提供。可制备金属纳米颗粒、介电纳米颗粒和半导体纳米颗粒以及杂合结构(例如，核-壳纳米颗粒)。如果由半导体材料制成的纳米颗粒足够小(通常小于10nm)，可以发生电子能级的量化，则也可将这些纳米颗粒标记为量子点。此类纳米级颗粒在生物医学应用中用作药物运

载体或成像剂,并且可对其进行调整以用于本发明中的类似目的。

[0635] 已经制造了半固体纳米颗粒和软纳米颗粒,并且它们在本发明的范围内。具有半固体性质的原型纳米颗粒是脂质体。各种类型的脂质体纳米颗粒目前在临床上用作抗癌药物和疫苗的递送系统。一半亲水而另一半疏水的纳米颗粒被称为Janus颗粒,对稳定乳液特别有效。它们可以在水/油界面处自组装并充当固体表面活性剂。

[0636] 使用多种不同技术进行颗粒表征(包括表征形态、尺寸等)。常用技术是电子显微术(TEM、SEM)、原子力显微镜术(AFM)、动态光散射(DLS)、X-射线光电子能谱法(XPS)、粉末X射线衍射(XRD)、傅里叶变换红外光谱法(FTIR)、基质辅助激光解吸/电离飞行时间质谱法(MALDI-TOF)、紫外-可见光谱法、双偏振干涉法和核磁共振(NMR)。可针对天然颗粒(即预加载)或在加载负荷物(本文的负荷物是指例如CRISPR-Cas系统的一种或多种组分,例如CRISPR-Cas蛋白或mRNA、腺苷脱氨酶(其可与CRISPR-Cas蛋白或衔接蛋白融合)或mRNA或者指导RNA或者它们的任何组合,并且可包括另外的运载体和/或赋形剂)后进行表征(尺寸测量),以提供具有最佳大小的颗粒用于本发明的任何体外、离体和/或体内应用的递送。在某些优选的实施方案中,颗粒尺寸(例如,直径)表征是基于使用动态激光散射(DLS)的测量。提到了美国专利号8,709,843;美国专利号6,007,845;美国专利号5,855,913;美国专利号5,985,309;美国专利号5,543,158;以及James E.Dahlman和Carmen Barnes等人的出版物(Nature Nanotechnology(2014),2014年5月11日在线发布,doi:10.1038/nnano.2014.84),其涉及颗粒、制备和使用它们的方法及其测量。

[0637] 本发明范围内的颗粒递送系统可以任何形式提供,包括但不限于固体、半固体、乳液或胶体颗粒。因此,本文所述的任何递送系统,包括但不限于例如基于脂质的系统、脂质体、微团、微泡、外来体或基因枪,可作为本发明范围内的颗粒递送系统提供。

[0638] 可使用颗粒或脂质包膜同时递送CRISPR-Cas蛋白mRNA、腺苷脱氨酶(其可与CRISPR-Cas蛋白或衔接蛋白融合)或mRNA以及指导RNA;例如,本发明的CRISPR-Cas蛋白和RNA(例如,作为复合物)可以经由如Dahlman等人的WO2015089419 A2及其中引用的文献中的颗粒(诸如7C1)来递送(参见例如James E.Dahlman和Carmen Barnes等人,Nature Nanotechnology(2014),2014年5月11日在线发布,doi:10.1038/nnano.2014.84),例如包含脂质或类脂质和亲水性聚合物(例如,阳离子脂质和亲水性聚合物)的递送颗粒,例如其中阳离子脂质包括1,2-二油酰基-3-三甲基铵-丙烷(DOTAP)或1,2-二十四酰基-sn-甘油-3-磷酸胆碱(DMPC)并且/或者其中亲水性聚合物包括乙二醇或聚乙二醇(PEG);并且/或者其中颗粒还包含胆固醇(例如,来源于以下制剂的颗粒:制剂1=DOTAP 100,DMPC 0,PEG 0,胆固醇0;制剂2=DOTAP 90,DMPC 0,PEG 10,胆固醇0;制剂3=DOTAP 90,DMPC 0,PEG 5,胆固醇5),其中使用有效的多步骤过程形成颗粒,其中首先将效应蛋白和RNA例如在无菌、不含核酸酶的1X PBS中例如在室温下以例如1:1的摩尔比混合在一起例如30分钟;分别将适用于制剂的DOTAP、DMPC、PEG和胆固醇溶解于醇例如100%乙醇中;并且将这两种溶液混合在一起以形成含有这些复合物的颗粒。

[0639] 可使用颗粒或脂质包膜同时递送核酸靶向效应蛋白(例如,V型蛋白,诸如Cas13)mRNA和指导RNA。合适颗粒的实例包括但不限于US 9,301,923中描述的那些。

[0640] 例如,Su X,Fricke J,Kavanagh DG,Irvine DJ("In vitro and in vivo mRNA delivery using lipid-enveloped pH-responsive polymer nanoparticles"Mol

Pharm.2011年6月6日;8(3):774-87.doi:10.1021/mp100390w.2011年4月1日的电子版)描述了可生物降解的核-壳结构纳米颗粒,所述纳米颗粒具有被磷脂双层壳包封的聚(β-氨基酯)(PBAE)核。这些被开发用于体内mRNA递送。选择pH响应性PBAE组分以促进内体破坏,而选择脂质表面层以最大程度降低聚阳离子核心的毒性。因此,这对于递送本发明的RNA是优选的。

[0641] 在一个实施方案中,考虑了基于自组装生物粘附性聚合物的颗粒/纳米颗粒,所述颗粒可用于口服递送肽、静脉内递送肽以及鼻内递送肽,所有这些均被递送到脑。还考虑了其他实施方案,诸如疏水药物的口服吸收和眼部递送。分子包膜技术涉及被保护并被递送到疾病部位的工程化聚合物包膜(参见例如Mazza,M.等人,ACSNano,2013.7(2):1016-1026;Siew,A.等人,Mol Pharm,2012.9(1):14-28;Lalatsa,A.等人,J Contr Rel,2012.161(2):523-36;Lalatsa,A.等人,Mol Pharm,2012.9(6):1665-80;Lalatsa,A.等人,Mol Pharm,2012.9(6):1764-74;Garrett,N.L.等人,J Biophotonics,2012.5(5-6):458-68;Garrett,N.L.等人,J Raman Spect,2012.43(5):681-688;Ahmad,S.等人,J Royal Soc Interface 2010.7:S423-33;Uchegbu,I.F.Expert Opin Drug Deliv,2006.3(5):629-40;Qu,X.等人,Biomacromolecules,2006.7(12):3452-9以及Uchegbu,I.F.等人,Int J Pharm,2001.224:185-199)。考虑了约5mg/kg的剂量,根据靶组织,采用单剂量或多剂量。

[0642] 在一个实施方案中,由MIT的Dan Anderson实验室开发的可以将RNA递送到癌细胞以阻止肿瘤生长的颗粒/纳米颗粒可用于和/或适于本发明的AD功能化的CRISPR-Cas系统。具体地,Anderson实验室开发了用于合成、纯化、表征和配制新生物材料和纳米制剂的全自动化组合系统。参见例如Alabi等人,Proc Natl Acad Sci U S A.2013年8月6日;110(32):12881-6;Zhang等人,Adv Mater.2013年9月6日;25(33):4641-5;Jiang等人,Nano Lett.2013年3月13日;13(3):1059-64;Karagiannis等人,ACS Nano.2012年10月23日;6(10):8484-7;Whitehead等人,ACS Nano.2012年8月28日;6(8):6922-9以及Lee等人,Nat Nanotechnol.2012年6月3日;7(6):389-93。

[0643] 美国专利申请20110293703涉及类脂质化合物在多核苷酸的施用中也特别有用,所述化合物可用于递送本发明的AD功能化的CRISPR-Cas系统。在一个方面,将氨基醇类脂质化合物与要递送到细胞或受试者的试剂组合以形成微粒、纳米颗粒、脂质体或微团。由颗粒、脂质体或微团递送的试剂可以是气体、液体或固体的形式,并且该试剂可以是多核苷酸、蛋白、肽或小分子。可将氨基醇类脂质化合物与其他氨基醇类脂质化合物、聚合物(合成的或天然的)、表面活性剂、胆固醇、碳水化合物、蛋白、脂质等组合以形成颗粒。然后可将这些颗粒任选地与药物赋形剂组合以形成药物组合物。

[0644] 美国专利公开号20110293703还提供了制备氨基醇类脂质化合物的方法。使胺的一种或多种等效物与环氧化物封端化合物的一种或多种等效物在适当条件下反应以形成本发明的氨基醇类脂质化合物。在某些实施方案中,胺的所有氨基基团与环氧化物封端化合物充分反应以形成叔胺。在其他实施方案中,胺的所有氨基基团不与环氧化物封端化合物充分反应以形成叔胺,从而在氨基醇类脂质化合物中产生伯胺或仲胺。这些伯胺或仲胺保持原样,或可与另一种亲电体诸如不同的环氧化物封端化合物反应。如本领域技术人员将理解的,使胺与未过量的环氧化物封端化合物反应将产生多种不同的具有不同尾部数的氨基醇类脂质化合物。某些胺可被两个环氧化物衍生的化合物尾部充分官能化,而其他分

子将不会被环氧化物衍生的化合物尾部完全官能化。例如，二胺或多胺可包括脱离该分子的不同氨基部分的一个、二个、三个或四个环氧化物衍生的化合物尾部，从而产生伯胺、仲胺和叔胺。在某些实施方案中，并不是所有氨基基团均被充分官能化。在某些实施方案中，使用两种相同类型的环氧化物封端化合物。在其他实施方案中，使用两种或更多种不同的环氧化物封端化合物。氨基醇类脂质化合物的合成在有或没有溶剂的情况下进行，并且该合成可在30-100℃、优选大约50-90℃的较高温度下进行。可任选地纯化所制备的氨基醇类脂质化合物。例如，可纯化氨基醇类脂质化合物的混合物，以产生具有特定数量的环氧化物衍生的化合物尾部的氨基醇类脂质化合物。或者可纯化混合物以产生特定的立体异构体或区异构体。也可使用烷基卤化物(例如，甲基碘)或其他烷化剂将氨基醇类脂质化合物烷化，并且/或者可将它们酰化。

[0645] 美国专利公开号20110293703还提供了通过本发明方法制备的氨基醇类脂质化合物的文库。可使用涉及液体处理器、自动机、微量滴定板、计算机等的高通量技术来制备和/或筛选这些氨基醇类脂质化合物。在某些实施方案中，筛选氨基醇类脂质化合物将多核苷酸或其他试剂(例如，蛋白、肽、小分子)转染到细胞中的能力。

[0646] 美国专利公开号20130302401涉及使用组合聚合制备的一类聚(β-氨基醇)(PBAA)。本发明的PBAA可在生物技术和生物医学应用中用作涂层(诸如用于医疗装置或植入物的膜或多层膜的涂层)、添加剂、材料、赋形剂、生物防污剂、微图案化剂和细胞包封剂。当用作表面涂层时，这些PBAA会根据其化学结构在体内和体外引起不同程度的炎症。这类材料具有各不相同的化学多样性使我们能够鉴定出在体外抑制巨噬细胞激活的聚合物涂层。此外，在皮下植入羧化聚苯乙烯微粒之后，这些涂层减少了炎症细胞的募集并减少了纤维化。这些聚合物可用于形成进行细胞包封的聚电解质复合物胶囊。本发明还可具有许多其他生物学应用，诸如抗微生物涂层、DNA或siRNA递送以及干细胞组织工程。美国专利公开号20130302401的教导可应用于本发明的AD功能化的CRISPR-Cas系统。

[0647] 可例如通过电穿孔转染包含Cas13、腺苷脱氨酶(其可与Cas13或衔接蛋白融合)和指导RNA的预组装的重组CRISPR-Cas复合物，从而产生高突变率并且没有可检测的脱靶突变。Hur,J.K.等人,Targeted mutagenesis in mice by electroporation of Cas13ribonucleoproteins,Nat Biotechnol.2016年6月6日.doi:10.1038/nbt.3596。

[0648] 就局部递送到脑而言，这可以通过多种方式来实现。例如，可以例如通过注射在纹状体内递送材料。可以经由穿颅术立体定向地进行注射。

[0649] 在一些实施方案中，可使用基于糖的颗粒，例如GalNAc，如本文所述以及参考WO2014118272(通过引用并入本文)以及Nair,JK等人,2014,Journal of the American Chemical Society 136(49),16958-16961)及其中的教导内容，特别是关于递送适用于所有颗粒，除非另有明显说明。这可被认为是基于糖的颗粒，并且本文提供了其他颗粒递送系统和/或制剂的更多细节。因此，就本文所述的其他颗粒而言，可以将GalNAc视为颗粒，使得一般用途和其他考虑因素(例如，所述颗粒的递送)也适用于GalNAc颗粒。例如，可使用溶液相共轭策略将激活为PFP(五氟苯基)酯的三天线GalNAc簇(摩尔重量为约2000)附接到5'-己基氨基修饰的寡核苷酸(5'-HA ASO，摩尔重量为约8000Da；Østergaard等人，Bioconjugate Chem.,2015,26(8),第1451-1455页)上。类似地，已经针对体内核酸递送描述了聚(丙烯酸酯)聚合物(参见WO2013158141，通过引用并入本文)。在另外的替代实施方

案中,可使用CRISPR纳米颗粒(或蛋白复合物)与天然存在的血清蛋白的预混合来改善递送(Akinc A等人,2010,Molecular Therapy,第18卷,第7期,1357-1364)。

[0650] 纳米线团

[0651] 此外,可使用纳米线团来递送AD功能化的CRISPR系统,例如如以下文献中所述:Sun W等人,Cocoon-like self-degradable DNA nanoclew for anticancer drug delivery.,J Am Chem Soc.2014年10月22日;136(42):14722-5.doi:10.1021/ja5088024.2014年10月13日电子版;或Sun W等人,Self-Assembled DNA Nanoclews for the Efficient Delivery of CRISPR-Cas9 for Genome Editing.,Angew Chem Int Ed Engl.2015年10月5日;54(41):12029-33.doi:10.1002/anie.201506030.2015年8月27日电子版。

[0652] LNP

[0653] 在一些实施方案中,通过将Cas13蛋白或mRNA形式包封在脂质颗粒(诸如LNP)中来进行递送。因此,在一些实施方案中,考虑了脂质纳米颗粒(LNP)。已经将抗转甲状腺素蛋白小干扰RNA包封在脂质纳米颗粒中并递送给人类(参见例如Coelho等人,N Engl J Med 2013;369:819-29),并且可将这种系统调整为并应用于本发明的CRISPR Cas系统。考虑了静脉内施用约0.01至约1mg/kg体重的剂量。考虑了降低输注相关反应的风险的药物,诸如考虑了地塞米松、对乙酰氨基酚、苯海拉明或西替利嗪以及雷尼替丁。还考虑了每4周约0.3mg/kg的多剂量,持续五个剂量。

[0654] 已经表明LNP在将siRNA递送到肝脏方面非常有效(参见例如Tabernero等人,Cancer Discovery,2013年4月,第3卷,第4期,第363-470页),因此被考虑用于将编码CRISPR Cas的RNA递送到肝脏。可考虑每两周约四个剂量的6mg/kg的LNP这一剂量。Tabernero等人证明,在以0.7mg/kg给药LNP的前2个周期之后,观察到肿瘤消退,并且在6个周期结束时,患者已经实现了部分响应,其中淋巴结转移完全消退并且肝脏肿瘤显著萎缩。该患者接受40个剂量后获得了完全响应,在接受26个月以上的剂量后处于缓解状态并完成了治疗。患有RCC和在用VEGF途径抑制剂进行在先治疗之后进展的肝外部位(包括肾脏、肺和淋巴结)疾病的两名患者在所有部位的疾病都保持稳定大约8至12个月,并且患有PNET和肝转移的一名患者继续在18个月(36个剂量)的延伸研究中保持疾病稳定。

[0655] 但是,必须考虑LNP的电荷。当阳离子脂质与带负电荷的脂质结合,诱导促进细胞内递送的非双层结构。由于带电荷的LNP在静脉注射之后迅速从循环中清除,因此开发了pKa值低于7的可电离阳离子脂质(参见例如Rosin等人,Molecular Therapy,第19卷,第12期,第1286-2200页,2011年12月)。可将带负电荷的聚合物诸如RNA在低pH值(例如,pH 4)下加载到LNP中,在此pH下可电离脂质显示出正电荷。但是,在生理pH值下,LNP表现出与更长的循环时间相容的低表面电荷。已经关注的有四种可电离阳离子脂质,即1,2-二亚油酰基-3-二甲基铵-丙烷(DLinDAP)、1,2-二亚油基氧基-3-N,N-二甲基氨基丙烷(DLinDMA)、1,2-二亚油基氧基-酮基-N,N-二甲基-3-氨基丙烷(DLinKDMA)和1,2-二亚油基-4-(2-二甲基氨基乙基)-[1,3]-二氧戊环(DLinKC2-DMA)。已经表明,含有这些脂质的LNP siRNA系统在体内肝细胞中表现出显著不同的基因沉默特性,具有根据采用因子VII基因沉默模型的级数DLinKC2-DMA>DLinKDMA>DLinDMA>>DLinDAP而变化的潜能(参见例如Rosin等人,Molecular Therapy,第19卷,第12期,第1286-2200页,2011年12月)。可考虑1μg/ml的LNP或者LNP中或

与LNP相关联的CRISPR-Cas RNA的剂量，尤其是对于含有DLinKC2-DMA的制剂而言。

[0656] LNP的制备和CRISPR Cas包封可根据Rosin等人，Molecular Therapy，第19卷，第12期，第1286-2200页，2011年12月)使用和/或调整。阳离子脂质1,2-二亚油酰基-3-二甲基铵-丙烷(DLinDAP)、1,2-二亚油基氧基-3-N,N-二甲基氨基丙烷(DLinDMA)、1,2-二亚油基氧基酮基-N,N-二甲基-3-氨基丙烷(DLinK-DMA)、1,2-二亚油基-4-(2-二甲基氨基乙基)-[1,3]-二氧戊环(DLinKC2-DMA)、(3-o-[2″-(甲氧基聚乙二醇2000)琥珀酰]-1,2-二肉豆蔻酰基-sn-二醇(PEG-S-DMG)和R-3-[(ω-甲氧基-聚(乙二醇)2000)氨甲酰]-1,2-二肉豆蔻酰氧基丙基-3-胺(PEG-C-DOMG)可由Tekmira Pharmaceuticals(Vancouver,Canada)提供或者合成。胆固醇可购自Sigma(St Louis,MO)。可将特异性CRISPR Cas RNA包封在含有DLinDAP、DLinDMA、DLinK-DMA和DLinKC2-DMA的LNP中(阳离子脂质:DSPC:CHOL:PEGS-DMG或PEG-C-DOMG的摩尔比为40:10:40:10)。需要时，可掺入0.2％SP-DiOC18(Invitrogen,Burlington,Canada)来评估细胞摄取、细胞内传递和生物分布。可通过将由阳离子脂质:DSPC:胆固醇:PEG-c-DOMG(摩尔比为40:10:40:10)组成的脂质混合物溶解在乙醇中直到最终脂质浓度为10mmol/l来进行包封。可将脂质的这种乙醇溶液逐滴加入pH 4.0的50mmol/l柠檬酸盐中形成多层囊泡，从而产生30％乙醇(体积/体积)的最终浓度。在使用挤出机(Northern Lipids,Vancouver,Canada)通过两个重叠的80nm Nuclepore聚碳酸酯滤膜挤出多层囊泡之后，可形成较大单层囊泡。可通过如下步骤实现包封：将以2mg/ml溶解在含有30％乙醇(体积/体积)的pH 4.0的50mmol/l柠檬酸盐中的RNA逐滴加入所挤出的预成型较大单层囊泡中，并且在31℃下孵育30分钟，同时不断搅拌直到最终的RNA/脂质重量比为0.06/1(重量/重量)。通过使用Spectra/Por 2再生纤维素透析膜在pH 7.4的磷酸盐缓冲液(PBS)中透析16小时去除乙醇并中和制剂缓冲液。可使用NICOMP 370粒度分析仪、囊泡/强度模型和高斯拟合(Nicomp Particle Sizing,Santa Barbara,CA)通过动态光散射来确定纳米粒度分布。所有三个LNP系统的粒度在直径上均为约70nm。可通过使用VivaPureD MiniH色谱柱(Sartorius Stedim Biotech)从透析前后收集的样品中去除游离RNA来确定RNA包封效率。可从洗脱的纳米颗粒中提取包封的RNA，并在260nm处对其进行定量。通过使用来自Wako Chemicals USA(Richmond,VA)的胆固醇E酶学测定测量囊泡中的胆固醇含量来确定RNA与脂质的比率。结合本文对LNP和PEG脂质的讨论，聚乙二醇化脂质体或LNP同样适合递送CRISPR-Cas系统或其组分。

[0657] 可在以摩尔比50:10:38.5含有DLinKC2-DMA、DSPC和胆固醇的乙醇中制备脂质预混溶液(总脂质浓度为20.4mg/ml)。可以0.75:1的摩尔比(乙酸钠:DLinKC2-DMA)将乙酸钠加入脂质预混物中。随后可通过将该混合物与1.85倍体积的柠檬酸盐缓冲液(10mmol/l，pH 3.0)在剧烈搅拌下合并来使脂质水合，从而导致在含有35％乙醇的在水性缓冲液中自发形成脂质体。可在37℃下孵育脂质体溶液，以使粒度随时间增加。可在孵育的不同时间移出等分试样以通过动态光散射(Zetasizer Nano ZS,Malvern Instruments,Worcestershire,UK)研究脂质体尺寸的变化。一旦实现所希望的粒度，就可将水性PEG脂质溶液(储备液＝10mg/ml PEG-DMG的35％(体积/体积)乙醇)加入脂质体混合物中，以产生最终PEG摩尔浓度为3.5％的总脂质。加入PEG-脂质后，脂质体应保持其大小，从而有效地猝灭进一步生长。然后可以大约1:10(重量:重量)的RNA与总脂质比率将RNA加入空脂质体中，然后在37℃下孵育30分钟以形成加载的LNP。随后可将混合物在PBS中透析过夜，并用0.45μm注射器式过滤

器过滤。

[0658] 还考虑了球形核酸(SNA™)构建体和其他纳米颗粒(尤其是金纳米颗粒)作为将CRISPR-Cas系统递送到预期靶标的手段。重要数据表明,基于核酸功能化金纳米颗粒的AuraSense治疗性球形核酸(SNA™)构建体是有用的。

[0659] 可与本文的教导结合使用的文献包括:Cutler等人,J.Am.Chem.Soc.2011 133:9254-9257;Hao等人,Small.2011 7:3158-3162;Zhang等人,ACS Nano.2011 5:6962-6970;Cutler等人,J.Am.Chem.Soc.2012 134:1376-1391;Young等人,Nano Lett.2012 12:3867-71;Zheng等人,Proc.Natl.Acad.Sci.USA.2012 109:11975-80,Mirkin,Nanomedicine 2012 7:635-638;Zhang等人,J.Am.Chem.Soc.2012 134:16488-1691,Weintraub,Nature 2013 495:S14-S16;Choi等人,Proc.Natl.Acad.Sci.USA.2013 110(19):7625-7630;Jensen等人,Sci.Transl.Med.5,209ra152(2013);以及Mirkin等人,Small,10:186-192。

[0660] 具有RNA的自组装纳米颗粒可用被聚乙二醇化的聚乙烯亚胺(PEI)进行构建,其中Arg-Gly-Asp(RGD)肽配体附接在聚乙二醇(PEG)的远端。例如,这种系统已经被用作靶向表达整合素的肿瘤新血管系统和递送抑制血管内皮生长因子受体2(VEGF R2)表达的siRNA从而实现抑制肿瘤血管发生的手段(参见例如Schiffelers等人,Nucleic Acids Research,2004,第32卷,第19期)。可通过将等体积的阳离子聚合物水溶液和核酸水溶液混,以产生相对于磷酸盐(核酸)净摩尔过量(2至6范围内)的可电离氮(聚合物),从而制备纳米复合体。阳离子聚合物和核酸之间的静电相互作用导致形成平均粒度分布为约100nm的聚复合体,因此这里称为纳米复合体。设想了约100至200mg的CRISPR Cas的剂量用于在Schiffelers等人的自组装纳米颗粒中进行递送。

[0661] Bartlett等人的纳米复合体(PNAS,2007年9月25日,第104卷,第39期)也可应用于本发明。通过将等体积的阳离子聚合物水溶液和核酸水溶液混,以产生相对于磷酸盐(核酸)净摩尔过量(2至6范围内)的可电离氮(聚合物),从而制备Bartlett等人的纳米复合体。阳离子聚合物和核酸之间的静电相互作用导致形成平均粒度分布为约100nm的聚复合体,因此这里称为纳米复合体。如下合成Bartlett等人的DOTA-siRNA:从Macrocyclics(Dallas,TX)订购1,4,7,10-四氮杂环十二烷-1,4,7,10-四乙酸单(N-羟基琥珀酰亚胺酯)(DOTA-NHS酯)。将具有100倍摩尔过量的DOTA-NHS-酯的经胺修饰的RNA有义链的碳酸盐缓冲液(pH 9)加入微量离心管中。通过在室温下搅拌4小时使内容物反应。将DOTA-RNA有义缀合物用乙醇沉淀,重新悬浮在水中,并退火为未修饰的反义链,以产生DOTA-siRNA。所有液体均用Chelex-100(Bio-Rad,Hercules,CA)进行预处理,以去除痕量金属污染物。可通过使用含环糊精的聚阳离子形成Tf靶向和非靶向siRNA纳米颗粒。通常,以3(+/-)的电荷比和0.5g/l的siRNA浓度在水中形成纳米颗粒。用Tf(金刚烷-PEG-Tf)对靶向纳米颗粒表面上百分之一的金刚烷-PEG分子进行修饰。将纳米颗粒悬浮在5%(重量/体积)葡萄糖运载体溶液中以用于注射。

[0662] Davis等人(Nature,第464卷,2010年4月15日)使用靶向纳米颗粒递送系统进行了RNA临床试验(临床试验注册号NCT00689065)。在21天周期的第1、3、8和10天,通过30分钟的静脉输注向患有标准护理疗法难治的实体癌的患者施用靶向纳米颗粒剂量。纳米颗粒由合成递送系统组成,该合成递送系统含有:(1)线性的基于环糊精的聚合物(CDP),(2)显示在纳米颗粒外部上以接合癌细胞表面上的TF受体(TFR)的人转铁蛋白(TF)靶向配体,(3)亲水

性聚合物(用于促进生物流体中的纳米颗粒稳定性的聚乙二醇(PEG)),以及(4)被设计成降低RRM2表达的siRNA(临床中所用的序列先前表示为siR2B+5)。早已清楚TFR在恶性细胞中上调,并且RRM2是确立的抗癌靶标。这些纳米颗粒(临床版本表示为CALAA-01)在非人类灵长类动物中的多剂量研究中显示出良好的耐受性。虽然已经通过脂质体递送向患有慢性粒细胞白血病的单个患者施用了siRNA,但是Davis等人的临床试验是首次用靶向递送系统全身性递送siRNA并且治疗患有实体癌的患者的人体试验。为了确定靶向递送系统是否可以将功能性siRNA有效递送到人类肿瘤,Davis等人研究了来自三个不同剂量组的三名患者(患者A、B和C,均患有转移性黑素瘤,并且分别接受了18、24和30mg m-2 siRNA的CALAA-01剂量)的活组织检查。对于本发明的CRISPR Cas系统也可考虑相似的剂量。可用含有线性的基于环糊精的聚合物(CDP)、显示在纳米颗粒外部上以接合癌细胞表面上的TF受体(TFR)的人转铁蛋白(TF)靶向配体和/或亲水性聚合物(例如,用于促进生物流体中的纳米颗粒稳定性的聚乙二醇(PEG))的纳米颗粒来实现本发明的递送。

[0663] 美国专利号8,709,843(通过引用并入本文)提供了一种用于将含治疗剂的颗粒靶向递送到组织、细胞和细胞内隔室的药物递送系统。本发明提供了包含与表面活性剂、亲水性聚合物或脂质缀合的聚合物的靶向颗粒。美国专利号6,007,845(通过引用并入本文)提供了具有通过将多官能化合物与一种或多种疏水性聚合物和一种或多种亲水性聚合物共价连接而形成的多嵌段共聚物的核并且含有生物活性材料的颗粒。美国专利号5,855,913(通过引用并入本文)提供了一种具有空气动力学上较轻的颗粒的微粒组合物,这些颗粒的振实密度小于0.4g/cm3,平均直径在5μm至30μm之间,其表面上掺有表面活性剂以用于将药物递送到肺部系统。美国专利号5,985,309(通过引用并入本文)提供了掺有表面活性剂和/或带正电或带负电的治疗剂或诊断剂和带相反电荷的带电分子的亲水性或疏水性复合物以用于递送到肺部系统的颗粒。美国专利号5,543,158(通过引用并入本文)提供了具有可生物降解固体的可生物降解的可注射颗粒,该固体核在表面上包含生物活性材料和聚(亚烷基二醇)部分。WO2012135025(也公开为US20120251560,通过引用并入本文)描述了缀合聚乙烯亚胺(PEI)聚合物和缀合氮杂大环化合物(统称为“缀合脂聚物(lipomer)”或“脂聚物”)。在某些实施方案中,可以设想此类缀合脂聚物可以在CRISPR-Cas系统的背景下用来实现体外、离体和体内基因组干扰,从而修饰基因表达,包括调节蛋白表达。

[0664] 在一个实施方案中,纳米颗粒可以是环氧化物改性的脂质聚合物,有利地是7C1(参见例如James E.Dahlman和Carmen Barnes等人,Nature Nanotechnology(2014),2014年5月11日在线发布,doi:10.1038/nnano.2014.84)。通过使C15环氧化物封端的脂质与PEI600以14:1的摩尔比反应来合成C71,并用C14PEG2000来配制C71以产生在PBS溶液中稳定至少40天的纳米颗粒(直径在35和60nm之间)。

[0665] 可利用环氧化物修饰的脂质聚合物将本发明的CRISPR-Cas系统递送到肺细胞、心血管细胞或肾细胞,但是本领域技术人员可调整该系统以递送到其他靶器官。设想了剂量范围为约0.05至约0.6mg/kg。还设想了几天或几周的剂量,总剂量为约2mg/kg。

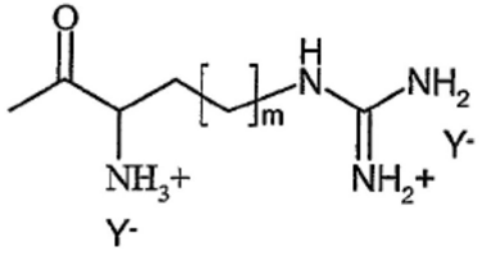
[0666] 在一些实施方案中,用于递送RNA分子的LNP通过本领域已知的方法(诸如在例如WO 2005/105152(PCT/EP2005/004920)、WO 2006/069782(PCT/EP2005/014074)、WO 2007/121947(PCT/EP2007/003496)和WO 2015/082080(PCT/EP2014/003274)中描述的那些方法,这些文献通过引用全部并入本文)制备。专门针对增强和改善将siRNA递送到哺乳动物细胞

中的LNP在例如以下文献中描述:Aleku等人,Cancer Res.,68(23):9788-98(Dec.1,2008),Strumberg et al.,Int.J.Clin.Pharmacol.Ther.,50(1):76-8(2012年1月);Schultheis等人,J.Clin.Oncol.,32(36):4141-48(2014年12月20日);以及Fehring等人,Mol.Ther.,22(4):811-20(2014年4月22日),这些文献通过引用并入本文并且可应用于本技术。

[0667] 在一些实施方案中,LNP包括在WO 2005/105152(PCT/EP2005/004920)、WO 2006/069782(PCT/EP2005/014074)、WO 2007/121947(PCT/EP2007/003496)和WO 2015/082080(PCT/EP2014/003274)中公开的任何LNP。

[0668] 在一些实施方案中,LNP包括至少一种具有式I的脂质:

[0669] (式I),其中R1和R2各自独立地选自包括烷基的组,n是1至4之间的任何整数,并且R3是选自包括以下各项的组的酰基:赖氨酰基、鸟氨酰基、2,4-二氨基丁酰基、组氨酰基和根据式II的酰基部分:

[0670] 

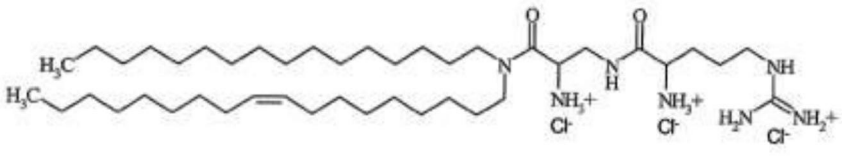
(式 II),

[0671] 其中m是1至3的任何整数,并且$Y^-$是药学上可接受的阴离子。在一些实施方案中,根据式I的脂质包括至少两个不对称的C原子。在一些实施方案中,式I的对映异构体包括但不限于R-R;S-S;R-S和S-R对映异构体。

[0672] 在一些实施方案中,R1是月桂基,而R2是肉豆蔻基。在另一个实施方案中,R1是棕榈基,而R2是油烯基。在一些实施方案中,m为1或2。在一些实施方案中,Y-选自卤化物、乙酸盐或三氟乙酸盐。

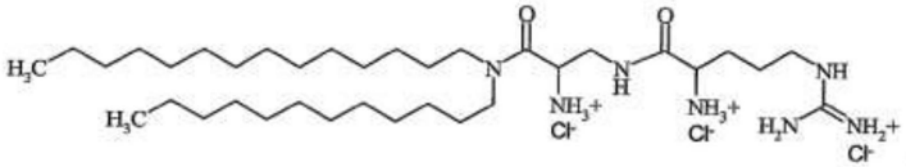
[0673] 在一些实施方案中,LNP包含选自以下各项的一种或多种脂质:

[0674] -精氨酰基-2,3-二氨基丙酸-N-棕榈基-N-油烯基-酰胺三盐酸盐(式III):

[0675] 

(式 III);

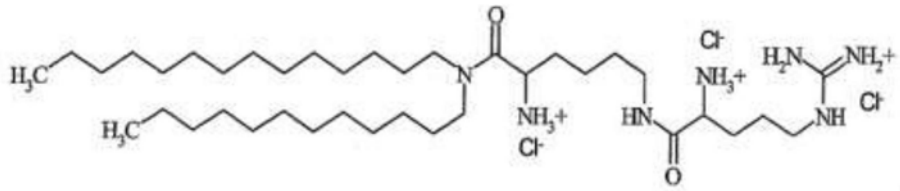
[0676] -精氨酰基-2,3-二氨基丙酸-N-月桂基-N-肉豆蔻基-酰胺三盐酸盐(式IV):

[0677] 

以及

(式 IV);

[0678] -精氨酰基-赖氨酸-N-月桂基-N-肉豆蔻基-酰胺三盐酸盐(式V):

[0679] 

(式 V)。

[0680] 在一些实施方案中,LNP还包括成分。举例来说但不限制,在一些实施方案中,成分选自肽、蛋白质、寡核苷酸、多核苷酸、核酸或其组合。在一些实施方案中,成分是抗体,例如单克隆抗体。在一些实施方案中,成分是选自例如核酶、适配体、镜像异构体、DNA、RNA、PNA、LNA或其组合的核酸。在一些实施方案中,核酸是指导RNA和/或mRNA。

[0681] 在一些实施方案中,LNP的成分包含编码CRIPSR-Cas蛋白的mRNA。在一些实施方案中,LNP的成分包含编码II型或V型CRIPSR-Cas蛋白的mRNA。在一些实施方案中,LNP的成分包含编码腺苷脱氨酶的mRNA(其可与CRISPR-Cas蛋白或衔接蛋白融合)。

[0682] 在一些实施方案中,LNP的成分还包含一种或多种指导RNA。在一些实施方案中,LNP被构造成将上述mRNA和指导RNA递送到血管内皮。在一些实施方案中,LNP被构造成将上述mRNA和指导RNA递送到肺内皮。在一些实施方案中,LNP被构造成将上述mRNA和指导RNA递送到肝。在一些实施方案中,LNP被构造成将上述mRNA和指导RNA递送到肺。在一些实施方案中,LNP被构造成将上述mRNA和指导RNA递送到心脏。在一些实施方案中,LNP被构造成将上述mRNA和指导RNA递送到脾。在一些实施方案中,LNP被构造成将上述mRNA和指导RNA递送到肾。在一些实施方案中,LNP被构造成将上述mRNA和指导RNA递送到胰腺。在一些实施方案中,LNP被构造成将上述mRNA和指导RNA递送到脑。在一些实施方案中,LNP被构造成将上述mRNA和指导RNA递送到巨噬细胞。

[0683] 在一些实施方案中,LNP还包括至少一种辅助脂质。在一些实施方案中,辅助脂质选自磷脂和类固醇。在一些实施方案中,磷脂是磷酸的二酯和/或单酯。在一些实施方案中,磷脂是磷酸甘油酯和/或鞘脂。在一些实施方案中,类固醇是基于部分氢化的环戊二烯并[a]菲的天然存在的和/或合成的化合物。在一些实施方案中,类固醇包含21至30个C原子。在一些实施方案中,类固醇是胆固醇。在一些实施方案中,辅助脂质选自1,2-二植烷酰基-sn-甘油-3-磷酸乙醇胺(DPhyPE)、神经酰胺和1,2-二油烯基-sn-甘油-3-磷酸乙醇胺(DOPE)。

[0684] 在一些实施方案中,至少一种辅助脂质包含选自包括以下各项的组的部分:PEG部分、HEG部分、聚羟乙基淀粉(polyHES)部分和聚丙烯部分。在一些实施方案中,该部分的分子量在约500至10,000Da之间或在约2,000至5,000Da之间。在一些实施方案中,PEG部分选自1,2-二硬脂酰基-sn-甘油-3-磷酸乙醇胺、1,2-二烷基-sn-甘油-3-磷酸乙醇胺和神经酰胺-PEG。在一些实施方案中,PEG部分的分子量在约500至10,000Da之间或在约2,000至5,000Da之间。在一些实施方案中,PEG部分的分子量为2,000Da。

[0685] 在一些实施方案中,辅助脂质为组合物的总脂质含量的约20mol%至80mol%。在一些实施方案中,辅助脂质为LNP的总脂质含量的约35mol%至65mol%。在一些实施方案中,LNP包括LNP总脂质含量的50mol%的脂质和50mol%的辅助脂质。

[0686] 在一些实施方案中,LNP包括-3-精氨酰基-2,3-二氨基丙酸-N-棕榈基-N-油烯基-酰胺三盐酸盐、-精氨酰基-2,3-二氨基丙酸-N-月桂基-N-肉豆蔻基-酰胺三盐酸盐或-精氨

酰基-赖氨酸-N-月桂基-N-肉豆蔻基-酰胺三盐酸盐中的任一者与DPhyPE的组合，其中DPhyPE的含量为LNP总脂质含量的80mol％、65mol％、50mol％和35mol％。在一些实施方案中，LNP包括-精氨酰基-2,3-二氨基丙酸-N-棕榈基-N-油烯基-酰胺三盐酸盐(脂质)和1,2-二植烷酰基-sn-甘油-3-磷酸乙醇胺(辅助脂质)。在一些实施方案中，LNP包括-精氨酰基-2,3-二氨基丙酸-N-棕榈基-N-油烯基-酰胺三盐酸盐(脂质)、1,2-二植烷酰基-sn-甘油-3-磷酸乙醇胺(第一辅助脂质)和1,2-二硬脂酰基-sn-甘油-3-磷酸乙醇胺-PEG2000(第二辅助脂质)。

[0687] 在一些实施方案中，第二辅助脂质为总脂质含量的约0.05mol％至4.9mol％或约1mol％至3mol％。在一些实施方案中，LNP包括总脂质含量的约45mol％至50mol％的脂质、总脂质含量的约45mol％至50mol％之间的第一辅助脂质，条件是存在总脂质含量的约0.1mol％至5mol％、约1mol％至4mol％或约2mol％的PEG化第二辅助脂质，其中脂质、第一辅助脂质和第二辅助脂质的含量之和为总脂质含量的100mol％，并且其中第一辅助脂质和第二辅助脂质之和为总脂质含量的50mol％。在一些实施方案中，LNP包含：(a)50mol％的-精氨酰基-2,3-二氨基丙酸-N-棕榈基-N-油烯基-酰胺三盐酸盐、48mol％的1,2-二植烷酰基-sn-甘油-3-磷酸乙醇胺；以及2mol％的1,2-二硬脂酰基-sn-甘油-3-磷酸乙醇胺-PEG2000；或(b)50mol％的-精氨酰基-2,3-二氨基丙酸-N-棕榈基-N-油烯基-酰胺三盐酸盐、49mol％的1,2-二植烷酰基-sn-甘油-3-磷酸乙醇胺；以及1mol％的N(羰基-甲氧基聚乙二醇-2000)-1,2-二硬脂酰基-sn-甘油-3-磷酸乙醇胺或其钠盐。

[0688] 在一些实施方案中，LNP包含核酸，其中核酸骨架磷酸盐与阳离子脂质氮原子的电荷比为约1:1.5-7或约1:4。

[0689] 在一些实施方案中，LNP还包括在体内条件下可从脂质组合物中去除的屏蔽化合物。在一些实施方案中，屏蔽化合物是生物惰性化合物。在一些实施方案中，屏蔽化合物在其表面或分子本身上不携带任何电荷。在一些实施方案中，屏蔽化合物是聚乙二醇(PEG)、基于羟乙基葡萄糖(HEG)的聚合物、聚羟乙基淀粉(polyHES)和聚丙烯。在一些实施方案中，PEG、HEG、polyHES和聚丙烯重量在约500至10,000Da之间或在约2000至5000Da之间。在一些实施方案中，屏蔽化合物是PEG2000或PEG5000。

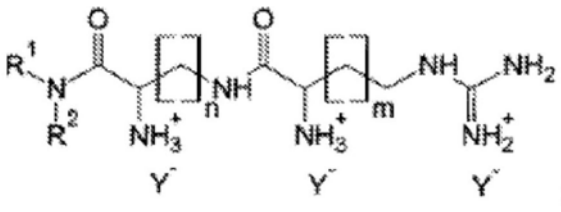
[0690] 在一些实施方案中，LNP包括至少一种脂质、第一辅助脂质以及在体内条件下可从脂质组合物中去除的屏蔽化合物。在一些实施方案中，LNP还包括第二辅助脂质。在一些实施方案中，第一辅助脂质是神经酰胺。在一些实施方案中，第二辅助脂质是神经酰胺。在一些实施方案中，神经酰胺包含至少一个6至10个碳原子的短碳链取代基。在一些实施方案中，神经酰胺包含8个碳原子。在一些实施方案中，屏蔽化合物附接到神经酰胺。在一些实施方案中，屏蔽化合物附接到神经酰胺。在一些实施方案中，屏蔽化合物共价附接到神经酰胺。在一些实施方案中，屏蔽化合物附接到LNP中的核酸。在一些实施方案中，屏蔽化合物共价附接到核酸。在一些实施方案中，屏蔽化合物通过接头附接到核酸。在一些实施方案中，接头在生理条件下被裂解。在一些实施方案中，接头选自ssRNA、ssDNA、dsRNA、dsDNA、肽、S-S接头和pH敏感接头。在一些实施方案中，接头部分附接到核酸有义链的3'端。在一些实施方案中，屏蔽化合物包含pH敏感接头或pH敏感部分。在一些实施方案中，pH敏感接头或pH敏感部分是阴离子接头或阴离子部分。在一些实施方案中，阴离子接头或阴离子部分在酸性环境中较少为阴离子或中性。在一些实施方案中，pH敏感接头或pH敏感部分选自低聚(谷氨

酸)、低聚酚盐和二亚乙基三胺五乙酸。

[0691] 在前段中的任何LNP实施方案中,LNP的同渗重摩可以在约50至600毫渗透摩尔/kg之间、在约250至350毫渗透摩尔/kg之间或在约280至320毫渗透摩尔/kg之间,并且/或者其中由脂质和/或一种或两种辅助脂质以及屏蔽化合物形成的LNP的粒度在约20至200nm之间、在约30至100nm之间或在约40至80nm之间。

[0692] 在一些实施方案中,屏蔽化合物提供了更长的体内循环时间,并允许含有LNP的核酸进行更好的生物分布。在一些实施方案中,屏蔽化合物防止LNP与血清化合物或者其他体液或细胞质膜的化合物(例如,向其施用LNP的脉管系统的内皮层的细胞质膜)立即相互作用。另外地或另选地,在一些实施方案中,屏蔽化合物还防止免疫系统的元件立即与LNP相互作用。另外地或另选地,在一些实施方案中,屏蔽化合物用作抗调理化合物。不希望受任何机制或理论的束缚,在一些实施方案中,屏蔽化合物形成减少可用于与其环境相互作用的LNP的表面积的覆盖物或涂层。另外地或另选地,在一些实施方案中,屏蔽化合物屏蔽LNP的总电荷。

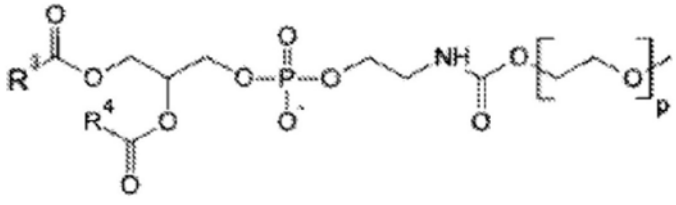
[0693] 在另一个实施方案中,LNP包括:至少一种具有式VI的阳离子脂质:

[0694] 

(式 VI),

[0695] 其中n为1、2、3或4,其中m为1、2或3,其中$Y^-$为阴离子,其中$R^1$和$R^2$中的每一者单独且独立地选自由以下各项组成的组:线性C12-C18烷基和线性C12-C18烯基;固醇化合物,其中所述固醇化合物选自由以下各项组成的组:胆固醇和豆固醇;以及PEG化脂质,其中所述PEG化脂质包含PEG部分,其中所述PEG化脂质选自由以下各项组成的组:

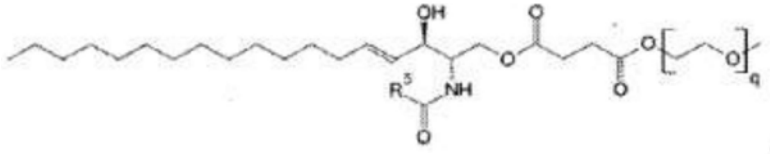
[0696] 式VII的PEG化磷酸乙醇胺:

[0697] 

(式 VII),

[0698] 其中$R^3$和$R^4$单独且独立地为线性C13-C17烷基,并且p为15至130之间的任何整数;

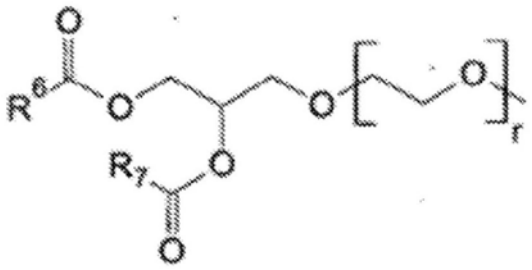
[0699] 式VIII的PEG化神经酰胺:

[0700] 

(式 VIII);

[0701] 其中$R^5$为线性C7-C15烷基,并且q为15至130之间的任何数;以及

[0702] 式IX的PEG化二酰基甘油:

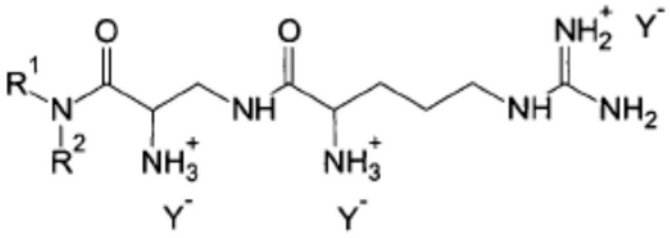
[0703] 

(式 IX),

[0704] 其中$R^6$和$R^7$中的每一者单独且独立地为线性C11-C17烷基,并且r为15至130的任何整数。

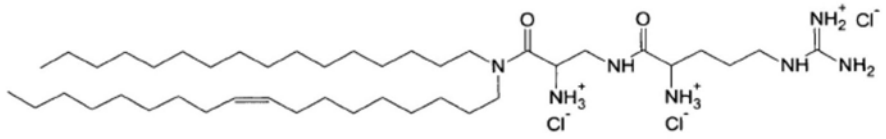
[0705] 在一些实施方案中,$R^1$和$R^2$彼此不同。在一些实施方案中,$R^1$是棕榈基,而$R^2$是油烯基。在一些实施方案中,$R^1$是月桂基,而$R^2$是肉豆蔻基。在一些实施方案中,$R^1$和$R^2$相同。在一些实施方案中,$R^1$和$R^2$单独且独立地选自由以下各项组成的组:C12烷基、C14烷基、C16烷基、C18烷基、C12烯基、C14烯基、C16烯基和C18烯基。在一些实施方案中,C12烯基、C14烯基、C16烯基和C18烯基中的每一者包含一个或两个双键。在一些实施方案中,C18烯基是在C9和C10之间具有一个双键的C18烯基。在一些实施方案中,C18烯基是顺式-9-十八烷基。

[0706] 在一些实施方案中,阳离子脂质是式X的化合物:

[0707] 

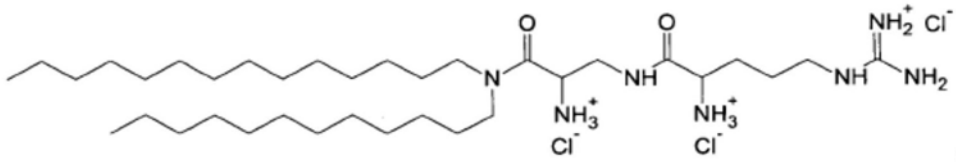
(式 X)。

在一些实施方案中,$Y^-$选自卤化物、乙酸盐和三氟乙酸盐。在一些实施方案中,阳离子脂质是式III的-精氨酰基-2,3-二氨基丙酸-N-棕榈基-N-油烯基-酰胺三盐酸盐:

[0708] 

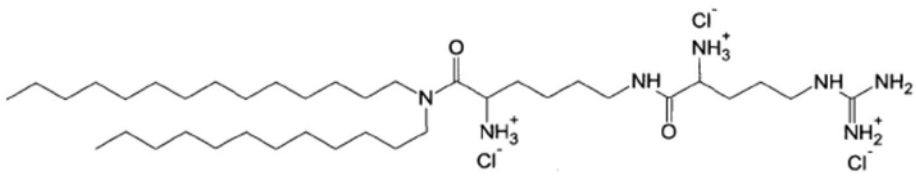
(式 III)。

在一些实施方案中,阳离子脂质是式IV的-精氨酰基-2,3-二氨基丙酸-N-月桂基-N-肉豆蔻基-酰胺三盐酸盐:

[0709] 

(式 IV)。

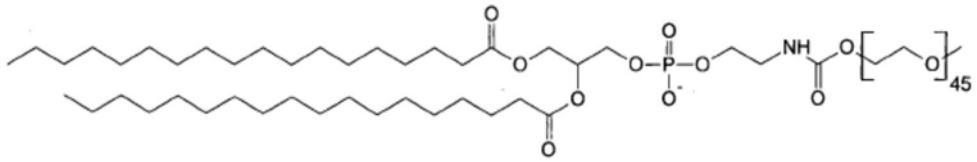
在一些实施方案中,阳离子脂质是式V的-精氨酰基-赖氨酸-N-月桂基-N-肉豆蔻基-酰胺三盐酸盐:

[0710] 

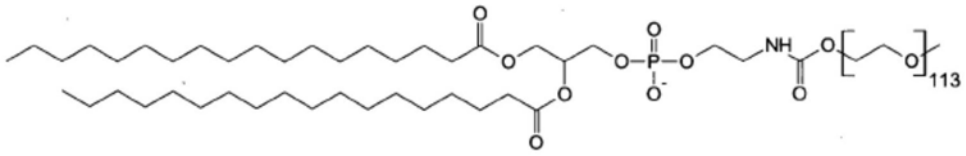
(式 V)。

[0711] 在一些实施方案中，固醇化合物是胆固醇。在一些实施方案中，固醇化合物是豆固醇。

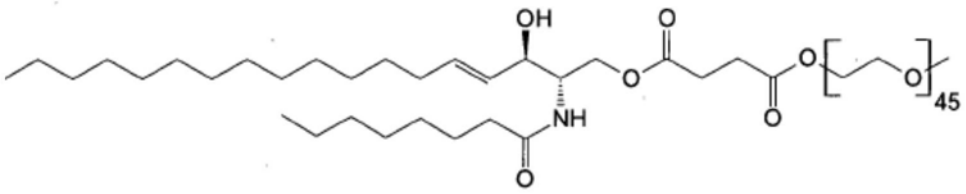
[0712] 在一些实施方案中，PEG化脂质的PEG部分的分子量为约800至5,000Da。在一些实施方案中，PEG化脂质的PEG部分的分子量为约800Da。在一些实施方案中，PEG化脂质的PEG部分的分子量为约2,000Da。在一些实施方式中，PEG化脂质的PEG部分的分子量为约5,000Da。在一些实施方式中，PEG化脂质是式VII的PEG化磷酸乙醇胺，其中$R^3$和$R^4$中的每一者单独且独立地为线性C13-C17烷基，并且p为18、19或20或者44、45或46或者113、114或115中的任何整数。在一些实施方案中，$R^3$和$R^4$相同。在一些实施方案中，$R^3$和$R^4$不同。在一些实施方案中，$R^3$和$R^4$单独且独立地选自由以下各项组成的组：C13烷基、C15烷基和C17烷基。在一些实施方案中，式VII的PEG化磷酸乙醇胺是1,2-二硬脂酰基-sn-甘油-3-磷酸乙醇胺-N-[甲氧基(聚乙二醇)-2000](铵盐)：

[0713] 

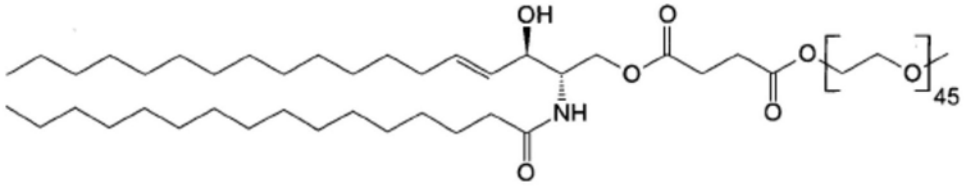
[0714] (式XI)。在一些实施方案中，式VII的PEG化磷酸乙醇胺是1,2-二硬脂酰基-sn-甘油-3-磷酸乙醇胺-N-[甲氧基(聚乙二醇)-5000](铵盐)：

[0715] 

[0716] (式XII)。在一些实施方案中，PEG化脂质是式VIII的PEG化神经酰胺，其中$R^5$为线性C7-C15烷基，并且q为18、19或20或者44、45或46或者113、114或115的任何整数。在一些实施方案中，$R^5$为线性C7烷基。在一些实施方案中，$R^5$为线性C15烷基。在一些实施方案中，式VIII的PEG化神经酰胺是N-辛酰基-鞘氨醇-1-{琥珀酰基[甲氧基(聚乙二醇)2000]}：

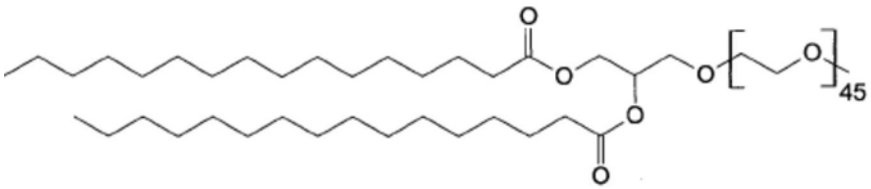
[0717] 

[0718] (式XIII)。在一些实施方案中，式VIII的PEG化神经酰胺是N-棕榈酰基-鞘氨醇-1-{琥珀酰基[甲氧基(聚乙二醇)2000]}：

[0719] 

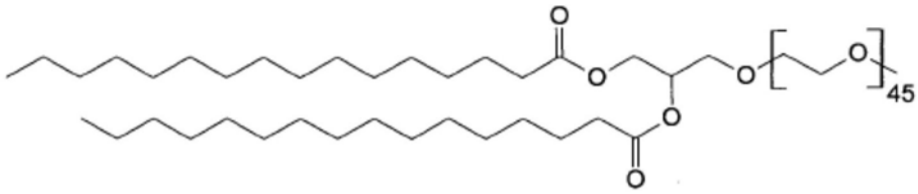
[0720] (式XIV)。在一些实施方案中，PEG化脂质是式IX的PEG化二酰基甘油，其中$R^6$和$R^7$中的每一者单独且独立地为线性C11-C17烷基，并且r为18、19或20或者44、45或46或者113、114或115的任何整数。在一些实施方案中，$R^6$和$R^7$相同。在一些实施方案中，$R^6$和$R^7$不同。在一些实施方案中，$R^6$和$R^7$中的每一者单独且独立地选自由以下各项组成的组：线性C17烷基、

线性C15烷基和线性C13烷基。在一些实施方案中，式IX的PEG化二酰基甘油是1,2-二硬脂酰基-sn-甘油[甲氧基(聚乙二醇)2000]：

[0721] 

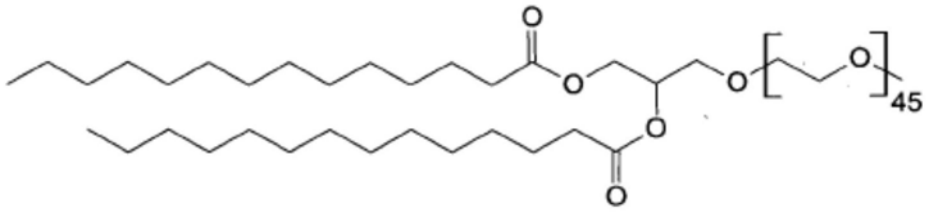
(式 XV)。

在一些实施方案中，式IX的PEG化二酰基甘油是1,2-二棕榈酰基-sn-甘油[甲氧基(聚乙二醇)2000]：

[0722] 

(式 XVI)。

在一些实施方案中，式IX的PEG化二酰基甘油是：

[0723] 

(式 XVII)。

在一些实施方案中，LNP包括至少一种选自式III、IV和V的阳离子脂质，以及至少一种选自胆固醇和豆固醇的固醇化合物，其中PEG化脂质是选自式XI和XII的至少一种。在一些实施方案中，LNP包括至少一种选自式III、IV和V的阳离子脂质，以及至少一种选自胆固醇和豆固醇的固醇化合物，其中PEG化脂质是选自式XIII和XIV的至少一种。在一些实施方案中，LNP包括至少一种选自式III、IV和V的阳离子脂质，以及至少一种选自胆固醇和豆固醇的固醇化合物，其中PEG化脂质是选自式XV和XVI的至少一种。在一些实施方案中，LNP包括式III的阳离子脂质，以及作为固醇化合物的胆固醇，并且其中PEG化脂质是式XI。

[0724] 在前段中的任何LNP实施方案中，其中阳离子脂质组合物的含量在约65摩尔％至75摩尔％之间，固醇化合物的含量在约24摩尔％至34摩尔％之间，并且PEG化脂质的含量在约0.5摩尔％至1.5摩尔％之间，其中脂质组合物的阳离子脂质、固醇化合物和PEG化脂质的含量之和为100摩尔％。在一些实施方案中，阳离子脂质为约70摩尔％，固醇化合物的含量为约29摩尔％，并且PEG化脂质的含量为约1摩尔％。在一些实施方案中，LNP为70摩尔％的式III、29摩尔％的胆固醇和1摩尔％的式XI。

[0725] 外来体

[0726] 外来体是转运RNA和蛋白并且可以将RNA递送到脑和其他靶器官的内源性纳米囊泡。为了降低免疫原性，Alvarez-Erviti等人(2011,Nat Biotechnol 29:341)使用自衍生的树突细胞来产生外来体。通过将树突细胞工程化为表达Lamp2b(一种与神经元特异性RVG肽融合的外来体膜蛋白)来实现靶向脑。通过电穿孔用外源RNA加载纯化的外来体。经静脉注射的RVG靶向外来体将GAPDH siRNA特异性递送到脑中的神经元、小胶质细胞、少突胶质

细胞,从而导致特异性基因敲低。预暴露于RVG外来体未减弱敲低作用,并且在其他组织中未观察到非特异性摄取。通过强力敲低BACE1(阿尔茨海默病的一种治疗靶标)的mRNA(60%)和蛋白(62%),证明了外来体介导的siRNA递送具有治疗潜能。

[0727] 为了获得免疫惰性的外来体池,Alvarez-Erviti等人从具有同质主要组织相容性复合物(MHC)单体型的近交C57BL/6小鼠中收获了骨髓。由于未成熟树突细胞会产生大量缺乏T细胞激活子诸如MHC-II和CD86的外来体,因此Alvarez-Erviti等人选择具有粒细胞/巨噬细胞集落刺激因子(GM-CSF)的树突细胞培养7天。次日,使用成熟的超速离心方案从培养上清液中纯化外来体。产生的外来体在物理上是同质的,粒度分布在直径为80nm处达到峰值,这是通过纳米颗粒跟踪分析(NTA)和电子显微镜确定的。Alvarez-Erviti等人每106个细胞获得了6-12μg外来体(基于蛋白浓度测量)。

[0728] 接下来,Alvarez-Erviti等人研究了使用适用于纳米级应用的电穿孔方案用外源负荷物加载经修饰的外来体的可能性。由于电穿孔对于纳米级的膜颗粒尚未良好表征,因此使用非特异性Cy5标记的RNA来对电穿孔方案进行经验优化。超速离心和裂解外来体后测定包封的RNA的量。在400V和125μF下进行电穿孔可最好保留RNA并且用于所有的后续实验。

[0729] Alvarez-Erviti等人向正常C57BL/6小鼠施用150μg包封在150μg RVG外来体中的每种BACE1siRNA并且将敲低效率与四个对照进行比较:未处理的小鼠、仅用RVG外来体注射的小鼠、用与体内阳离子脂质体试剂复合的BACE1siRNA注射的小鼠,以及用与RVG-9R复合的BACE1siRNA注射的小鼠,其中RVG肽与静电结合到siRNA的9个D-精氨酸缀合。在施用后3天分析皮层组织样品,在siRNA-RVG-9R处理的和siRNARVG外来体处理的小鼠中均观察到显著的蛋白敲低(45%,P<0.05,以及62%,P<0.01),这是由于BACE1mRNA水平显著降低(分别为66%[+或-]15%,P<0.001,以及61%[+或-]13%,P<0.01)。此外,申请人证明了总[β]-淀粉样蛋白1-42水平(阿尔茨海默病的淀粉样斑块的一种主要组分)在RVG-外来体处理的动物中显著降低(55%,P<0.05)。在脑室内注射BACE1抑制剂后,观察到β-淀粉样蛋白1-40的减少量大于正常小鼠中的减少量。Alvarez-Erviti等人对BACE1裂解产物进行了5'-cDNA末端快速扩增(RACE),为通过siRNA实现了RNAi介导的敲低提供了证据。

[0730] 最后,Alvarez-Erviti等人通过评估IL-6、IP-10、TNFα和IFN-α血清浓度研究了RNA-RVG外来体是否在体内诱导了免疫响应。在外来体处理之后,类似于siRNA转染试剂处理(与siRNA-RVG-9R相比之下,其有力地刺激了IL-6分泌),所有细胞因子都被记录为发生非显著变化,从而证实了外来体处理具有免疫惰性属性。鉴于外来体仅包封20%siRNA,用RVG-外来体递送比RVG-9R递送显得更有效,因为用少五倍的siRNA实现了相当的mRNA敲低和更好的蛋白敲低,而没有相应水平的免疫刺激。这一实验证明了RVG-外来体技术具有治疗潜力,其潜在地适合于与神经退行性疾病相关的基因的长期沉默。Alvarez-Erviti等人的外来体递送系统可用于将本发明的AD功能化CRISPR-Cas系统递送到治疗靶标,尤其是神经退行性疾病。本发明可考虑约100至1000mg包封在约100至1000mg RVG外来体中的CRISPR Cas的剂量。

[0731] El-Andaloussi等人(Nature Protocols 7,2112-2126(2012))公开了如何利用来源于培养细胞的外来体在体外和体内递送RNA。该方案首先描述了通过转染表达载体生成靶向外来体,该表达载体包含与肽配体融合的外来体蛋白。接着,El-Andaloussi等人解释了如何从转染的细胞上清液中纯化和表征外来体。接下来,El-Andaloussi等人详述了将

RNA加载到外来体中的关键步骤。最后,El-Andaloussi等人概述了如何使用外来体将RNA在体外和体内有效递送到小鼠脑中。还提供了预期结果的实例,其中通过功能测定和成像评估了外来体介导的RNA递送。整个方案耗时约3周。根据本发明的递送或施用可使用由自衍生的树突细胞产生的外来体进行。根据本文的教导,这可以用于实践本发明。

[0732] 在另一个实施方案中,考虑了Wahlgren等人(Nucleic Acids Research,2012,第40卷,第17期,e130)的血浆外来体。外来体是由许多细胞类型包括树突细胞(DC)、B细胞、T细胞、肥大细胞、上皮细胞和肿瘤细胞产生的纳米大小的囊泡(大小为30-90nm)。这些囊泡是由晚期内体向内出芽而形成,然后在与质膜融合后释放到细胞外环境中。因为外来体天然地在细胞之间运载RNA,所以这种特性可用于基因疗法,并且根据本公开可以用于实践本发明。

[0733] 来自血浆的外来体可以如下制备:以900g离心血沉棕黄层20分钟以分离血浆,然后收获细胞上清液,以300g离心10分钟以去除细胞,并且以16500g离心30分钟,然后通过0.22mm过滤器进行过滤。通过以120 000g超速离心70分钟使外来体沉淀。根据RNAi Human/Mouse Starter Kit(Quiagen,Hilden,Germany)中的制造商说明将siRNA化学转染到外来体中。将siRNA以2mmol/ml的最终浓度加入100ml PBS中。加入HiPerFect转染试剂后,将混合物在室温下孵育10分钟。为了去除过量的微团,使用醛/硫酸盐乳胶珠再分离外来体。将CRISPR Cas化学转染到外来体中可类似于siRNA进行。外来体可与从健康供体外周血中分离的单核细胞和淋巴细胞共培养。因此,可考虑将含有CRISPR Cas的外来体引入人的单核细胞和淋巴细胞中,并以自体方式再引入人体内。因此,根据本发明的递送或施用可使用血浆外来体进行。

[0734] 脂质体

[0735] 根据本发明的递送或施用可以用脂质体进行。脂质体是球形囊泡结构,由围绕内部水性区室的单层或多层脂质双层以及相对不可渗透的外部亲脂性磷脂双层组成。脂质体作为药物递送运载体受到了广泛关注,因为它们具有生物相容性、无毒性,可以递送亲水性和亲脂性药物分子,保护其负荷物免于被血浆酶降解,并且可以跨生物膜和血脑屏障(BBB)转运其负载(参见例如Spuch和Navarro,Journal of Drug Delivery,第2011卷,文章编号469679,第12页,2011,doi:10.1155/2011/469679查看综述)。

[0736] 脂质体可以由几种不同类型的脂质制成;然而,最常使用磷脂来产生脂质体作为药物运载体。虽然当脂质膜与一种水性溶液混合时脂质体形成是自发的,但是也可通过使用均质机、超声发生器或挤出设备以振动形式施加力来使该过程加速(参见例如Spuch和Navarro,Journal of Drug Delivery,第2011卷,文章编号469679,第12页,2011,doi:10.1155/2011/469679查看综述)。

[0737] 可将几种其他添加剂加入脂质体中,以修饰其结构和特性。例如,可将胆固醇或鞘磷脂加入脂质体混合物中,以帮助稳定脂质体结构并防止脂质体内部负荷物泄漏。此外,由氢化卵磷脂酰胆碱或卵磷脂酰胆碱、胆固醇和磷酸二鲸蜡酯制备脂质体,并将它们的平均囊泡大小调节为约50和100nm。(参见例如Spuch和Navarro,Journal of Drug Delivery,第2011卷,文章编号469679,第12页,2011,doi:10.1155/2011/469679查看综述)。

[0738] 脂质体制剂主要可由天然磷脂和脂质诸如1,2-二硬脂酰基-sn-甘油-3-磷脂酰胆碱(DSPC)、鞘磷脂、卵磷脂酰胆碱和单唾液酸神经节苷脂组成。由于这种制剂仅由磷脂组

成,因此脂质体制剂遇到了许多挑战,其中一个挑战就是在血浆中的不稳定性。已经作出了克服这些挑战的若干尝试,特别是在脂质膜的处理方面。其中一种尝试集中在胆固醇的处理方面。将胆固醇加入常规制剂中减缓了包封的生物活性化合物快速释放到血浆中的速度,或者加入1,2-二油酰基-sn-甘油-3-磷酸乙醇胺(DOPE)增加了稳定性(参见例如Spuch和Navarro,Journal of Drug Delivery,第2011卷,文章编号469679,第12页,2011,doi:10.1155/2011/469679查看综述)。

[0739] 在一个特别有利的实施方案中,Trojan Horse脂质体(也称为Trojan Horses分子)是可取的,其方案可见于http://cshprotocols.cshlp.org/content/2010/4/pdb.prot5407.long。这些颗粒允许在血管内注射转基因后将其递送到整个脑中。不受限制,据认为具有与表面缀合的特异性抗体的中性脂质颗粒允许经由内吞作用跨越血脑屏障。可使用Trojan Horse脂质体经由血管内注射将核酸酶的CRISPR家族递送到脑中,这将获得全脑转基因动物而无需胚胎操作。对于在脂质体中的体内施用,可考虑约1-5g DNA或RNA。

[0740] 在另一个实施方案中,AD功能化的CRISPR Cas系统或其组分可在脂质体(诸如稳定的核酸-脂质颗粒(SNALP))中施用(参见例如Morrissey等人,Nature Biotechnology,第23卷,第8期,2005年8月)。考虑了每日静脉注射约1、3或5mg/kg/天的在SNALP中靶向的特异性CRISPR Cas。每日治疗可超过约三天,然后每周治疗持续约五周。在另一个实施方案中,还考虑了通过以约1或2.5mg/kg的剂量静脉注射施用的包封有特异性CRISPR Cas的SNALP(参见例如Zimmerman等人,Nature Letters,第441卷,2006年5月4日)。SNALP制剂可含有摩尔百分比为2:40:10:48的脂质3-N-[(w-甲氧基聚(乙二醇)2000)氨甲酰]-1,2-二肉豆蔻氧基-丙胺(PEG-C-DMA)、1,2-二亚油基氧基-N,N-二甲基-3-氨基丙烷(DLinDMA)、1,2-二硬脂酰-sn-甘油-3-磷酸胆碱(DSPC)和胆固醇(例如参见Zimmerman等人,Nature Letters,第441卷,2006年5月4日)。

[0741] 在另一个实施方案中,已经证明稳定的核酸-脂质颗粒(SNALP)可以将分子有效递送到高度血管化的HepG2-衍生的肝脏肿瘤,但是不递送到血管化较差的HCT-116衍生的肝脏肿瘤(参见例如Li,Gene Therapy(2012)19,775-780)。SNALP脂质体可如下制备:用二硬脂酰磷脂酰胆碱(DSPC)、胆固醇和siRNA,使用25:1的脂质/siRNA比率和48/40/10/2的胆固醇/D-Lin-DMA/DSPC/PEG-C-DMA摩尔比配制D-Lin-DMA和PEG-C-DMA。所得的SNALP脂质体的大小为约80-100nm。

[0742] 在又一个实施方案中,SNALP可包含合成胆固醇(Sigma-Aldrich,St Louis,MO,USA)、二棕榈酰磷脂酰胆碱(Avanti Polar Lipids,Alabaster,AL,USA)、3-N-[(w-甲氧基聚(乙二醇)2000)氨甲酰]-1,2-二肉豆蔻氧基丙基胺和阳离子1,2-二亚油基氧基-3-N,N二甲基氨基丙烷(参见例如Geisbert等人,Lancet 2010;375:1896-905)。可考虑以例如作为静脉推注施用的每剂量约2mg/kg总CRISPR Cas的剂量。

[0743] 在又一个实施方案中,SNALP可包含合成胆固醇(Sigma-Aldrich)、1,2-二硬脂酰-sn-甘油-3-磷酸胆碱(DSPC;Avanti Polar Lipids Inc.)、PEG-cDMA和1,2-二亚油基氧基-3-(N;N-二甲基)氨基丙烷(DLinDMA)(参见例如Judge,J.Clin.Invest.119:661-673(2009))。用于体内研究的制剂可包含约9:1的最终脂质/RNA质量比。

[0744] Alnylam Pharmaceuticals的Barros和Gollob已对RNAi纳米药物的安全性进行了

综述(例如参见Advanced Drug Delivery Reviews 64(2012)1730-1737)。稳定的核酸脂质颗粒(SNALP)由四种不同的脂质组成:在低pH下为阳离子的可电离脂质(DLinDMA)、中性辅助脂质、胆固醇和可扩散聚乙二醇(PEG)-脂质。该颗粒的直径为大约80nm,并且在生理pH值下呈电中性。在配制过程中,可电离脂质用于在颗粒形成过程中使脂质与阴离子RNA凝聚。当在越来越酸性的内体条件下带正电时,可电离脂质还会介导SNALP与内体膜融合,从而使RNA释放到细胞质中。PEG-脂质在配制过程中可稳定颗粒并减少聚集,随后提供可改善药代动力学特性的中性亲水性外部。

[0745] 迄今为止,已经使用SNALP制剂和RNA启动了两个临床项目。Tekmira Pharmaceuticals最近在LDL胆固醇升高的成年志愿者中完成了SNALP-ApoB的I期单剂量研究。ApoB主要在肝脏和空肠中表达,对于VLDL和LDL的组装和分泌必不可少。十七名受试者接受了单剂量的SNALP-ApoB(跨7个剂量水平的剂量递增)。没有证据表明产生了肝脏毒性(基于临床前研究,被预期为潜在的剂量限制性毒性)。接受最高剂量的(两名受试者中的)一名受试者经历了与免疫系统刺激相一致的流感样症状,于是决定结束试验。

[0746] Alnylam Pharmaceuticals类似地推出了ALN-TTR01,其采用如上所述的SNALP技术并且靶向突变体和野生型TTR的肝细胞产生,从而治疗TTR淀粉样变性(ATTR)。已经描述了三种ATTR综合征:家族性淀粉样多神经病(FAP)和家族性淀粉样心肌病(FAC),这两者均由TTR中的常染色体显性突变引起;以及由野生型TTR引起的老年全身性淀粉样变性(SSA)。最近在ATTR患者中完成了ALN-TTR01的安慰剂对照单剂量递增I期试验。通过15分钟静脉输注在0.01至1.0mg/kg(基于siRNA)的剂量范围内向31名患者施用ALN-TTR01(23名输注研究药物,8名输注安慰剂)。治疗耐受性良好,在肝功能检查中没有显著增加。23名患者中,以≥0.4mg/kg输注的3名患者出现了与输注相关的反应;所有患者都对减慢输注速度有反应,并且所有患者都继续参与研究。以最高剂量1mg/kg输注的两名患者出现了血清细胞因子IL-6、IP-10和IL-1ra的最小和瞬时升高(根据临床前和NHP研究预期)。在1mg/kg处观察到血清TTR的降低,这是ALN-TTR01的预期药理效应。

[0747] 在又一个实施方案中,可通过将阳离子脂质、DSPC、胆固醇和PEG-脂质例如以40:10:40:10的摩尔比分别溶解在例如乙醇中来制备SNALP(参见Semple等人,Nature Niotechnology,第28卷,第2期,2010年2月,第172-177页)。将该脂质混合物加入水性缓冲液(50mM柠檬酸盐,pH 4)中,混合直到最终的乙醇和脂质浓度分别为30%(体积/体积)和6.1mg/ml,使其在22℃下平衡2分钟,然后挤出。使用Lipex挤出机(Northern Lipids),在22℃下将水合脂质通过两个重叠的孔径为80nm的滤膜(Nuclepore)挤出,直到获得直径为70-90nm的囊泡(通过动态光散射分析确定)。这一般需要通过1-3次。将siRNA(溶解在50mM柠檬酸盐、pH 4水溶液中,含30%乙醇)以约5ml/min的速率一边混合一边加入预平衡(35℃)的囊泡中。达到0.06(重量/重量)的最终靶siRNA/脂质比率后,将混合物在35℃下再孵育30分钟,以进行囊泡重组和siRNA包封。然后除去乙醇,并通过透析或切向流渗滤将外部缓冲液替换为PBS(155mM NaCl、3mM Na2HPO4、1mM KH2PO4,pH 7.5)。使用受控的逐步稀释法过程将siRNA包封在SNALP中。KC2-SNALP的脂质成分为DLin-KC2-DMA(阳离子脂质)、二棕榈酰磷脂酰胆碱(DPPC;Avanti Polar Lipids)、合成胆固醇(Sigma)和PEG-C-DMA,它们以摩尔比57.1:7.1:34.3:1.4使用。在形成加载的颗粒后,将SNALP用PBS透析,并在使用前通过0.2μm滤膜过滤灭菌。平均粒度为75-85nm,并且90-95%的siRNA被包封在脂质颗粒中。用于体内

测试的制剂中最终的siRNA/脂质比率为约0.15(重量/重量)。即将使用前,将含有因子VII siRNA的LNP-siRNA系统在无菌PBS中稀释到适当浓度,并且通过侧尾静脉以10ml/kg的总体积静脉施用制剂。这种方法和这些递送系统可外推到本发明的AD功能化的CRISPR Cas系统。

[0748] 其他脂质

[0749] 其他阳离子脂质诸如氨基脂质2,2-二亚油基-4-二甲基氨基乙基-[1,3]-二氧戊环(DLin-KC2-DMA)可用来例如类似于SiRNA包封CRISPR Cas或其组分或编码其的一种或多种核酸分子(参见例如Jayaraman,Angew.Chem.Int.Ed.2012,51,8529-8533),因此可用于实践本发明。可考虑具有以下脂质组成的预成型囊泡:摩尔比分别为40/10/40/10的氨基脂质、二硬脂酰磷脂酰胆碱(DSPC)、胆固醇和(R)-2,3-双(十八烷氧基)丙基-1-(甲氧基聚(乙二醇)2000)丙基碳酸酯(PEG-脂质),以及比率为大约0.05(重量/重量)的FVII siRNA/总脂质。为了确保在70-90nm范围内的窄粒度分布以及0.11+0.04(n=56)的低多分散性指数,可在加入指导RNA之前将颗粒通过80nm膜挤出多达三次。可使用含有高效氨基脂质16的颗粒,其中可进一步优化四种脂质组分即16、DSPC、胆固醇和PEG-脂质的摩尔比(50/10/38.5/1.5)以增强体内活性。

[0750] Michael S D Kormann等人("Expression of therapeutic proteins after delivery of chemically modified mRNA in mice:Nature Biotechnology,第29卷,第154-157页(2011))描述了使用脂质包膜来递送RNA。使用脂质包膜在本发明中也是优选的。

[0751] 在另一个实施方案中,可将脂质与本发明的AD功能化的CRISPR Cas系统或其一种或多种组分或编码其的一种或多种核酸分子一起配制,以形成脂质纳米颗粒(LNP)。脂质包括但不限于DLin-KC2-DMA4、C12-200和辅助脂质二硬脂酰磷脂酰胆碱、胆固醇,并且可使用自发囊泡形成程序将PEG-DMG与CRISPR Cas而不是siRNA一起配制(参见例如Novobrantseva,Molecular Therapy-Nucleic Acids(2012)1,e4;doi:10.1038/mtna.2011.3)。组分摩尔比可为约50/10/38.5/1.5(DLin-KC2-DMA或C12-200/二硬脂酰磷脂酰胆碱/胆固醇/PEG-DMG)。在为DLin-KC2-DMA和C12-200脂质纳米颗粒(LNP)的情况下,最终的脂质:siRNA重量比可分别为约12:1和9:1。制剂的平均粒径可为约80nm,包覆效率>90%。可考虑3mg/kg的剂量。

[0752] Tekmira在美国和国外拥有大约95个专利家族的产品组合,涉及LNP和LNP制剂的各个方面(参见例如美国专利号7,982,027、7,799,565、8,058,069、8,283,333、7,901,708、7,745,651、7,803,397、8,101,741、8,188,263、7,915,399、8,236,943和7,838,658以及欧洲专利号1766035、1519714、1781593和1664316),所有这些产品都可用于和/或适于本发明。

[0753] 可将AD功能化的CRISPR Cas系统或其一种或多种组分或编码其的一种或多种核酸分子包封在诸如在美国公开申请20130252281和20130245107以及20130244279(转让给Moderna Therapeutics)中进一步描述的PLGA微球中进行递送,这些申请涉及配制包含经修饰的核酸分子的组合物的各方面,所述核酸分子可编码蛋白、蛋白前体或者蛋白或蛋白前体的部分或完全加工形式。该制剂的摩尔比可为50:10:38.5:1.5-3.0(阳离子脂质:融合脂质:胆固醇:PEG脂质)。PEG脂质可选自但不限于PEG-c-DOMG、PEG-DMG。融合脂质可为DSPC。还可参见Schrum等人,Delivery and Formulation of Engineered Nucleic Acids、

美国公开申请20120251618。

[0754] Nanomerics的技术解决了许多疗法的生物利用度挑战，包括低分子量疏水性药物、肽和基于核酸的疗法(质粒、siRNA、miRNA)。该技术已经证明具有明显优势的特异性施用途径包括口服途径、跨血脑屏障转运、向实体瘤和眼部递送。参见例如Mazza等人,2013,ACS Nano.2013年2月26日;7(2):1016-26;Uchegbu和Siew,2013,J Pharm Sci.102(2):305-10以及Lalatsa等人,2012,J Control Release.2012年7月20日;161(2):523-36。

[0755] 美国专利公开号20050019923描述了用于将生物活性分子诸如多核苷酸分子、肽和多肽和/或药剂递送到哺乳动物体内的阳离子树状大分子。所述树状大分子适合靶向将生物活性分子递送到例如肝、脾、肺、肾或心脏(或甚至脑)。树状大分子是由简单的分支单体单元逐步制备而成的合成三维大分子，可以轻松控制和改变其性质和功能。树状大分子是通过将构建嵌段重复加成到多功能核(发散合成法)上或朝向多功能核重复加成构建嵌段(收敛合成法)而合成的，并且每次加成构建嵌段的三维壳都会导致新一代的树状大分子形成。聚丙烯亚胺树状大分子从二氨基丁烷核开始，通过将丙烯腈双迈克尔加成到伯胺上，然后将腈加氢，在该二氨基丁烷核上加成两倍数量的氨基基团。这导致氨基基团加倍。聚丙烯亚胺树状大分子含有100％的可质子化氮和多达64个末端氨基基团(第5代,DAB 64)。可质子化基团通常是能够在中性pH下接受质子的胺基团。树状大分子作为基因递送剂的用途主要集中在聚酰胺-胺和含磷化合物的用途，其中胺/酰胺的混合物或N--P(O2)S分别作为缀合单元，还没有工作报道更低代的聚丙烯亚胺树状大分子用于基因递送的用途。还研究了聚丙烯亚胺树状大分子作为pH敏感的受控释放系统，用于递送药物以及当被外周氨基酸基团化学修饰时用于包封它们的客体分子。还研究了聚丙烯亚胺树状大分子的细胞毒性和与DNA的相互作用以及DAB64的转染效率。

[0756] 美国专利公开号20050019923基于以下观察结果:与早期报道相反，阳离子树状大分子诸如聚丙烯亚胺树状大分子显示出用于靶向递送生物活性分子诸如遗传物质的合适特性，诸如特异性靶向和低毒性。另外，阳离子树状大分子的衍生物也显示出用于靶向递送生物活性分子的合适特性。还可参见Bioactive Polymers(美国公开申请20080267903)，其公开了包括阳离子聚胺聚合物和树状聚合物在内的各种聚合物被证明具有抗增殖活性，因此可用于治疗其特征为发生不希望的细胞增殖的疾病，诸如赘生物和肿瘤、炎症性疾病(包括自身免疫性疾病)、银屑病和动脉粥样硬化。这些聚合物可单独用作活性剂，或用作其他治疗剂诸如用于基因疗法的药物分子或核酸的递送媒介物。在这种情况下，这些聚合物本身固有的抗肿瘤活性可补充要递送的药剂的活性。这些专利公开的公开内容可与本文的教导结合使用，用于递送一个或多个AD功能化的CRISPR Cas系统或其一种或多种组分或编码其的一种或多种核酸分子。

[0757] 超电荷蛋白

[0758] 超电荷蛋白是一类具有异常高的正或负净理论电荷的工程化或天然存在的蛋白，并且可用于递送一个或多个AD功能化的CRISPR Cas系统或其一种或多种组分或编码其的一种或多种核酸分子。超负电荷和超正电荷蛋白均表现出显著的抵抗热诱导或化学诱导的聚集的能力。超正电荷蛋白还能够穿透哺乳动物细胞。将负荷物诸如质粒DNA、RNA或其他蛋白与这些蛋白结合可以使这些大分子能够在体外和体内功能性递送到哺乳动物细胞中。2007年已经报道了超电荷蛋白的产生和表征(Lawrence等人,2007,Journal of the

American Chemical Society 129,10110-10112)。

[0759] 将RNA和质粒DNA以非病毒方式递送到哺乳动物细胞中对于研究和治疗应用都是有价值的(Akinc等人,2010,Nat.Biotech.26,561-569)。将纯化的+36 GFP蛋白(或其他超正电荷蛋白)在适当的无血清培养基中与RNA混合并使其复合,然后加入细胞中。在此阶段加入血清会抑制超电荷蛋白-RNA复合物的形成并降低治疗效果。发现以下方案对多种细胞系有效(McNaughton等人,2009,Proc.Natl.Acad.Sci.USA 106,6111-6116)(但是,应当进行改变蛋白和RNA剂量的先导实验,以优化特定细胞系的程序):(1)治疗前一天,在48孔板中每孔涂铺1×105个细胞。(2)治疗当天,在无血清培养基中稀释纯化的+36 GFP蛋白直到最终浓度为200nM。加入RNA直到最终浓度为50nM。涡旋混合并在室温下孵育10分钟。(3)孵育期间,从细胞中吸出培养基,并用PBS洗涤一次。(4)孵育+36 GFP和RNA后,将蛋白-RNA复合物加入细胞中。(5)将细胞与复合物在37℃下孵育4小时。(6)孵育后,吸出培养基,并用20U/mL肝素PBS洗涤三次。将细胞与含血清的培养基再孵育48小时或更长时间,具体取决于活性测定。(7)通过免疫印迹、qPCR、表型测定或其他合适的方法分析细胞。

[0760] 进一步发现+36 GFP在许多细胞中都是有效的质粒递送试剂。由于质粒DNA是比siRNA更大的负荷物,因此有效地复合质粒需要成比例更多的+36GFP蛋白。为了有效地递送质粒,申请人已经开发了带有C端HA2肽标签的+36 GFP变体,这是一种已知的来源于流感病毒血凝素蛋白的内体破坏性肽。以下方案在多种细胞中是有效的,但如上所述,建议针对特定细胞系和递送应用优化质粒DNA和超电荷蛋白的剂量:(1)治疗前一天,在48孔板中每孔涂铺1×105个细胞。(2)治疗当天,在无血清培养基中稀释纯化的þ36GFP蛋白直到最终浓度为2mM。添加1mg质粒DNA。涡旋混合并在室温下孵育10分钟。(3)孵育期间,从细胞中吸出培养基,并用PBS洗涤一次。(4)孵育þ36GFP和质粒DNA后,将蛋白-DNA复合物轻轻添加到细胞中。(5)将细胞与复合物在37℃下孵育4小时。(6)孵育后,吸出培养基并用PBS洗涤。在含血清的培养基中孵育细胞,再孵育24-48小时。(7)视情况分析质粒递送(例如,通过质粒驱动的基因表达)。

[0761] 还可参见例如McNaughton等人,Proc.Natl.Acad.Sci.USA 106,6111-6116(2009);Cronican等人,ACS Chemical Biology 5,747-752(2010);Cronican等人,Chemistry&Biology 18,833-838(2011);Thompson等人,Methods in Enzymology 503,293-319(2012);Thompson,D.B.等人,Chemistry&Biology19(7),831-843(2012)。超电荷蛋白的方法可用于和/或适于递送本发明的AD功能化的CRISPR Cas系统。这些系统与本文的教导结合可以用于递送一个或多个AD功能化的CRISPR Cas系统或其一种或多种组分或编码其的一种或多种核酸分子。

[0762] 细胞穿透肽(CPP)

[0763] 在又一个实施方案中,考虑了细胞穿透肽(CPP)用于递送AD功能化的CRISPR Cas系统。CPP是促进细胞摄取各种分子负荷物(从纳米级颗粒到小化学分子和大DNA片段)的短肽。本文所用的术语“负荷物”包括但不限于治疗剂、诊断探针、肽、核酸、反义寡核苷酸、质粒、蛋白、颗粒(包括纳米颗粒)、脂质体、发色团、小分子和放射性物质。在本发明的各方面,负荷物还可包括AD功能化的CRISPR Cas系统的任何组分或整个AD功能化的功能性CRISPR Cas系统。本发明的各方面还提供了用于将希望的负荷物递送到受试者体内的方法,其包

括:(a)制备包含本发明的细胞穿透肽以及希望的负荷物的复合物,并且(b)将复合物经口服、关节内、腹膜内、鞘内、动脉内、鼻内、实质内、皮下、肌内、静脉内、经皮、直肠内或局部施用给受试者。负荷物通过化学键合经由共价键或通过非共价相互作用与肽缔合。

[0764] CPP的功能是将负荷物递送到细胞中,这一过程通常通过内吞作用发生,其中负荷物被递送到活哺乳动物细胞的内体中。细胞穿透肽具有不同的大小、氨基酸序列和电荷,但是所有CPP都具有一个独特的特征,即能够转运质膜并促进将各种分子负荷物递送到细胞质或细胞器中。CPP转运可分为三种主要进入机制:直接穿透膜、内吞作用介导的进入以及通过形成暂时结构进行转运。CPP已在医学上得到广泛应用,在治疗不同疾病(包括癌症)时作为药物递送剂、作为病毒抑制剂以及作为用于细胞标记的造影剂。后者的实例包括充当GFP的运载体、MRI造影剂或量子点。CPP作为用于研究和医学的体外和体内递送载体具有巨大潜力。CPP通常具有以下氨基酸组成:包含较高相对丰度的带负电荷氨基酸,诸如赖氨酸或精氨酸,或者具有包含交替模式的极性/带电荷氨基酸和非极性疏水氨基酸的序列。这两种类型的结构分别称为聚阳离子或两亲性。第三类CPP是仅包含非极性残基、具有低净电荷或具有对于细胞摄入至关重要的疏水氨基酸基团的疏水肽。最初发现的其中一种CPP是来自人类免疫缺陷病毒1(HIV-1)的反式激活转录激活因子(Tat),发现其被多种培养细胞类型从周围培养基中有效摄取。从那以后,已知CPP的数量大大增加,并产生了具有更有效蛋白转导特性的小分子合成类似物。CPP包括但不限于穿透素、Tat(48-60)、转运素和(R-AhX-R4)(Ahx=氨基己酰基)。

[0765] 美国专利8,372,951提供了一种来源于嗜酸粒细胞阳离子蛋白(ECP)的CPP,其表现出高细胞穿透效率和低毒性。还提供了将CPP与其负荷物一起递送到脊椎动物受试者中的各方面。CPP及其递送的其他方面在美国专利8,575,305、8;614,194和8,044,019中描述。CPP可以用于递送AD功能化的CRISPR-Cas系统或其组分。Suresh Ramakrishna、Abu-Bonsrah Kwaku Dad、Jagadish Beloor等人的稿件“Gene disruption by cell-penetrating peptide-mediated delivery of Cas9protein and guide RNA”(Genome Res,2014年4月2日,该文献全文通过引用并入)中也提供了CPP可以用于递送AD功能化的CRISPR-Cas系统或其组分,其中证明了用CPP缀合的重组Cas9蛋白和CPP复合的指导RNA处理可以导致人类细胞系中的内源基因破坏。在该论文中,Cas9蛋白经由硫醚键缀合到CPP,而指导RNA与CPP复合,从而形成缩合的带负电荷颗粒。结果表明,用经修饰的Cas9和指导RNA同时和顺序处理人类细胞,包括胚胎干细胞、真皮成纤维细胞、HEK293T细胞、HeLa细胞和胚胎癌细胞,可以导致有效的基因破坏,相对于质粒转染降低了脱靶突变。

[0766] 气雾剂递送

[0767] 例如,对于进行肺部疾病治疗的受试者,每个肺可例如接受在自主呼吸时经支气管内递送的药学有效量的雾化AAV载体系统。因此,一般来讲,雾化递送对于AAV递送是优选的。腺病毒或AAV颗粒可用于递送。可将合适的基因构建体(各自可操作地连接到一个或多个调控序列)克隆到递送载体中。

[0768] 包装和启动子

[0769] 用于驱动CRISPR-Cas蛋白和腺苷脱氨酶编码核酸分子表达的启动子可以包括可以用作启动子的AAV ITR。这有利于消除对额外启动子元件(其可能占用载体中的空间)的需要。释放的额外空间可以用于驱动其他元件(gRNA等)的表达。而且,ITR活性相对较弱,因

此可以用于降低由于Cas13的过量表达引起的潜在毒性。

[0770] 对于全身性表达，可以使用的启动子包括：CMV、CAG、CBh、PGK、SV40、铁蛋白重链或轻链等。对于脑或其他CNS表达，突触蛋白I可以用于所有神经元，CaMKIIα可以用于兴奋性神经元，GAD67或GAD65或VGAT可以用于GABA能神经元。对于肝脏表达，可以使用白蛋白启动子。对于肺表达，可以使用SP-B。对于内皮细胞，可以使用ICAM。对于造血细胞，可以使用IFNβ或CD45。对于成骨细胞，可以使用OG-2。

[0771] 用于驱动指导RNA的启动子可以包括Pol III启动子，诸如U6或H1，以及Pol II启动子和内含子盒，用来表达指导RNA。

[0772] 腺相关病毒(AAV)

[0773] 可以使用腺相关病毒(AAV)、慢病毒、腺病毒或者其他质粒或病毒载体类型，特别是使用例如来自美国专利号8,454,972(腺病毒的制剂、剂量)、8,404,658(AAV的制剂、剂量)和5,846,946(DNA质粒的制剂、剂量)以及来自临床试验和关于涉及慢病毒、AAV和腺病毒的临床试验的出版物的制剂和剂量来递送CRISPR-Cas蛋白、腺苷脱氨酶以及一种或多种指导RNA。例如，对于AAV，施用途径、制剂和剂量可以如美国专利号8,454,972和涉及AAV的临床试验。对于腺病毒，施用途径、制剂和剂量可以如美国专利号8,404,658和涉及腺病毒的临床试验。对于质粒递送，施用途径、制剂和剂量可以如美国专利号5,846,946和涉及质粒的临床研究。剂量可基于或外推到平均70kg的个体(例如，成年男性)，并且可以针对患者、受试者、不同体重和物种的哺乳动物进行调节。施用频率在医学或兽医学从业者(例如，医师、兽医)的能力范围内，取决于常规因素，包括患者或受试者的年龄、性别、一般健康状况、其他状况以及着手解决的特殊状况或症状。可以将病毒载体注射到目标组织中。对于细胞类型特异性基因组修饰，Cas13和腺苷脱氨酶的表达可以由细胞类型特异性启动子驱动。例如，肝脏特异性表达可使用白蛋白启动子，而神经元特异性表达(例如，用于靶向CNS疾病)可使用突触蛋白I启动子。

[0774] 就体内递送而言，AAV优于其他病毒载体，有两个原因：毒性低(这可能是由于纯化方法不需要对细胞颗粒进行超速离心，因为超速离心可能激活免疫响应)；以及引起插入诱变的可能性低，因为它不整合到宿主基因组中。

[0775] AAV的包装限制为4.5或4.75Kb。这意味着必须将Cas13以及启动子和转录终止子都装配在同一病毒载体中。构建体大于4.5或4.75Kb将导致病毒产量显著降低。SpCas9相当大，基因本身超过4.1Kb，这使其很难包装到AAV中。因此，本发明的实施方案包括利用较短的Cas13的同系物。

[0776] 关于AAV，AAV可以是AAV1、AAV2、AAV5或它们的任何组合。可以针对要靶向的细胞选择AAV中的AAV；例如，可以选择AAV血清型1、2、5或杂合衣壳AAV1、AAV2、AAV5或它们的任何组合用于靶向脑或神经元细胞；并且可以选择AAV4用于靶向心脏组织。AAV8可用于递送到肝脏。本文的启动子和载体是单独优选的。关于这些细胞的某些AAV血清型的列表(参见Grimm,D.等人,J.Virol.82:5887-5911(2008))如下：

[0777]

| 细胞系 | AAV-1 | AAV-2 | AAV-3 | AAV-4 | AAV-5 | AAV-6 | AAV-8 | AAV-9 |
|---|---|---|---|---|---|---|---|---|
| Huh-7 | 13 | 100 | 2.5 | 0.0 | 0.1 | 10 | 0.7 | 0.0 |
| HEK293 | 25 | 100 | 2.5 | 0.1 | 0.1 | 5 | 0.7 | 0.1 |
| HeLa | 3 | 100 | 2.0 | 0.1 | 6.7 | 1 | 0.2 | 0.1 |
| HepG2 | 3 | 100 | 16.7 | 0.3 | 1.7 | 5 | 0.3 | ND |
| Hep1A | 20 | 100 | 0.2 | 1.0 | 0.1 | 1 | 0.2 | 0.0 |
| 911 | 17 | 100 | 11 | 0.2 | 0.1 | 17 | 0.1 | ND |
| CHO | 100 | 100 | 14 | 1.4 | 333 | 50 | 10 | 1.0 |
| COS | 33 | 100 | 33 | 3.3 | 5.0 | 14 | 2.0 | 0.5 |
| MeWo | 10 | 100 | 20 | 0.3 | 6.7 | 10 | 1.0 | 0.2 |
| NIH3T3 | 10 | 100 | 2.9 | 2.9 | 0.3 | 10 | 0.3 | ND |
| A549 | 14 | 100 | 20 | ND | 0.5 | 10 | 0.5 | 0.1 |
| HT1180 | 20 | 100 | 10 | 0.1 | 0.3 | 33 | 0.5 | 0.1 |
| 单核细胞 | 1111 | 100 | ND | ND | 125 | 1429 | ND | ND |
| 不成熟的 DC | 2500 | 100 | ND | ND | 222 | 2857 | ND | ND |
| 成熟的 DC | 2222 | 100 | ND | ND | 333 | 3333 | ND | ND |

[0778] 慢病毒

[0779] 慢病毒是复杂的逆转录病毒，能够在有丝分裂和有丝分裂后细胞中感染并表达其基因。最常见的慢病毒是人类免疫缺陷病毒(HIV)，其使用其他病毒的包膜糖蛋白来靶向广泛的细胞类型。

[0780] 慢病毒可如下制备。克隆pCasES10(含有慢病毒转移质粒主链)后，将低传代(p＝5)的HEK293FT接种在T-75烧瓶中，直到在转染的前一天在含10％胎牛血清且无抗生素的DMEM中达到50％的汇合。20小时后，将培养基更换为OptiMEM(无血清)培养基，并在4小时后进行转染。用10μg慢病毒转移质粒(pCasES10)和以下包装质粒转染细胞：5μg pMD2.G(VSV-g假型)和7.5ug psPAX2(gag/pol/rev/tat)。用阳离子脂质递送剂(50uL Lipofectamine 2000和100ul Plus试剂)在4mL OptiMEM中进行转染。6小时后，将培养基更换为含10％胎牛血清的无抗生素DMEM。这些方法在细胞培养过程中使用血清，但优选无血清的方法。

[0781] 慢病毒可如下纯化。48小时后收获病毒上清液。首先清除上清液中的碎片，并通过0.45um低蛋白结合(PVDF)滤膜进行过滤。然后将它们在超速离心机中以24,000rpm旋转2小时。将病毒沉淀重新悬浮在50ul DMEM中，在4℃下过夜。然后将它们等分，并立即在-80℃下冷冻。

[0782] 在另一个实施方案中，还考虑了基于马传染性贫血病毒(EIAV)的最小非灵长类慢病毒载体，特别是用于眼部基因治疗(参见例如Balagaan,J Gene Med 2006;8:275-285)。在另一个实施方案中，还考虑了RetinoStat®，一种经由视网膜下注射递送用于治疗湿性年龄相关性黄斑变性、表达血管生成抑制蛋白(内皮抑素和血管抑素)的基于马传染性贫血病毒的慢病毒基因疗法载体(参见例如Binley等人,HUMAN GENE THERAPY 23:980-991(2012年9月))，并且可对这种载体进行修饰以用于本发明的AD功能化的CRISPR-Cas系统。

[0783] 在另一个实施方案中，具有靶向由HIV tat/rev共享的共有外显子的siRNA、核仁

定位TAR诱饵和抗CCR5特异性锤头状核酶的自失活慢病毒载体(参见例如DiGiusto等人,2010)Sci Transl Med 2:36ra43)可用于和/或适于本发明的AD功能化的CRISPR-Cas系统。每千克患者体重可收集最少2.5×106个CD34+细胞,并且以2×106个细胞/ml的密度在含有2μmol/L谷氨酰胺、干细胞因子(100ng/ml)、Flt-3配体(Flt-3L)(100ng/ml)和促血小板生成素(10ng/ml)(CellGenix)的X-VIVO 15培养基(Lonza)中预刺激16至20小时。可在75cm2的包被有纤连蛋白(25mg/cm2)(RetroNectin,Takara Bio Inc.)的组织培养瓶中用慢病毒以感染复数5转导预刺激的细胞16至24小时。

[0784] 公开了慢病毒载体用于治疗帕金森病,参见例如美国专利公开号20120295960以及美国专利号7303910和7351585。还公开了慢病毒载体用于治疗眼部疾病,参见例如美国专利公开号20060281180、20090007284、US20110117189、US20090017543、US20070054961、US20100317109。还公开了慢病毒载体用于递送到脑,参见例如美国专利公开号US20110293571、US20110293571、US20040013648、US20070025970、US20090111106和美国专利号US7259015。

[0785] 在非动物生物中的应用

[0786] 可以结合作物基因组学的最新进展来使用一个或多个AD功能化的CRISPR系统(例如,单个或多重)。本文所述的系统可以用于执行高效且经济高效的植物基因或基因组探询或编辑或操纵,例如用于植物基因或基因组的快速研究和/或选择和/或探询和/或比较和/或操作和/或转化;例如,以为植物产生、鉴定、开发、优化或赋予性状或特征,或者以转化植物基因组。因此,可以改善植物即具有性状或特征的新组合的新植物或者具有增强性状的新植物的产生。AD功能化的CRISPR系统可以在定点整合(SDI)或基因编辑(GE)或任何近反向育种(NRB)或反向育种(RB)技术中用于植物。利用本文所述的Cas13效应蛋白系统的各方面可类似于在植物中使用CRISPR-Cas(例如,CRISPR-Cas9)系统,并且提到了亚利桑那大学(University of Arizona)的网站"CRISPR-PLANT"(http://www.genome.arizona.edu/crispr/)(由宾州州立大学(Penn State)和AGI提供支持)。本发明的实施方案可以在植物中或者在先前已经使用过RNAi或类似的基因组编辑技术的情况下可以用于基因组编辑;参见例如Nekrasov,"Plant genome editing made easy:targeted mutagenesis in model and crop plants using the CRISPR-Cas system,"Plant Methods 2013,9:39(doi:10.1186/1746-4811-9-39);Brooks,"Efficient gene editing in tomato in the first generation using the CRISPR-Cas9system,"Plant Physiology,2014年9月,第114.247577页;Shan,"Targeted genome modification of crop plants using a CRISPR-Cas system,"Nature Biotechnology 31,686-688(2013);Feng,"Efficient genome editing in plants using a CRISPR-Cas system,"Cell Research(2013)23:1229-1232.doi:10.1038/cr.2013.114;2013年8月20日在线发布;Xie,"RNA-guided genome editing in plants using a CRISPR-Cas system,"Mol Plant.2013年11月;6(6):1975-83.doi:10.1093/mp/sst119.2013年8月17日的电子版;Xu,"Gene targeting using the Agrobacterium tumefaciens-mediated CRISPR-Cas system in rice,"Rice 2014,7:5(2014);Zhou等人,"Exploiting SNPs for biallelic CRISPR mutations in the outcrossing woody perennial Populus reveals 4-coumarate:CoA ligase specificity and Redundancy,"New Phytologist(2015)(论坛)1-4(仅可从

www.newphytologist.com在线获取);Caliando等人,“Targeted DNA degradation using a CRISPR device stably carried in the host genome,NATURE COMMUNICATIONS 6:6989,DOI:10.1038/ncomms7989,www.nature.com/naturecommunications DOI:10.1038/ncomms7989;美国专利号6,603,061“Agrobacterium-Mediated Plant Transformation Method”;美国专利号7,868,149“Plant Genome Sequences and Uses Thereof”以及美国2009/0100536“Transgenic Plants with Enhanced Agronomic Traits”,这些文献中每一篇的全部内容和公开内容全文通过引用并入本文。在本发明的实践中,Morrell等人“Crop genomics:advances and applications,”Nat Rev Genet.2011年12月29日13(2):85-96(通过引用并入本文)的内容和公开内容包括关于其中的实施方案可如何用于植物。因此,本文中对动物细胞的提及也可作必要的变通后应用于植物细胞,除非另有明显说明;并且,本文中脱靶效应减少的酶和采用此类酶的系统可以用于植物应用,包括本文中提到的那些应用。

[0787] AD功能化的CRISPR系统在植物和酵母中的应用

[0788] 一般来讲,术语“植物”涉及植物界的任何不同的光合、真核、单细胞或多细胞生物,其特征在于通过细胞分裂进行生长,含有叶绿体,并且具有由纤维素组成的细胞壁。术语“植物”涵盖单子叶植物和双子叶植物。具体地讲,这些植物旨在包括但不限于被子植物和裸子植物,诸如阿拉伯树胶、苜蓿、苋菜、苹果、杏、朝鲜蓟、白蜡树、芦笋、鳄梨、香蕉、大麦、豆类、甜菜、桦树、山毛榉、黑莓、蓝莓、西兰花、抱子甘蓝、卷心菜、油菜、哈密瓜、胡萝卜、木薯、花椰菜、雪松、谷类、芹菜、栗子、樱桃、大白菜、柑橘、克莱门氏小柑橘、三叶草、咖啡、玉米、棉花、豇豆、黄瓜、柏树、茄子、榆树、菊苣、桉树、茴香、无花果、冷杉、天竺葵、葡萄、葡萄柚、落花生、地樱桃、树胶铁杉、山核桃木、羽衣甘蓝、奇异果、甘蓝、落叶松、莴苣、韭、柠檬、青柠、洋槐、松树、孔雀草、玉米、芒果、枫树、甜瓜、粟、蘑菇、芥菜、坚果、橡树、燕麦、油棕、秋葵、洋葱、橙、观赏植物或花或树木、木瓜、棕榈、荷兰芹、欧洲防风草、豌豆、桃、花生、梨、泥炭、胡椒、柿子、木豆、松树、菠萝、大蕉、李子、石榴、马铃薯、南瓜、菊苣、萝卜、油菜籽、覆盆子、水稻、黑麦、高粱、红花、黄华柳、大豆、菠菜、云杉、南瓜属植物果实、草莓、糖甜菜、甘蔗、向日葵、甘薯、甜玉米、橘子、茶、烟草、番茄、树类、黑小麦、草坪草、芜菁、藤本植物、胡桃、豆瓣菜、西瓜、小麦、山药、紫杉和西葫芦。术语“植物”还涵盖藻类,藻类主要是光合自养生物,它们的共同特点主要是缺少根、叶以及其他表征高等植物的器官。

[0789] 使用本文所述的AD功能化的CRISPR系统进行基因组编辑的方法可以用于赋予基本上任何植物所需的性状。使用本公开的核酸构建体和上述各种转化方法,可针对本文所述的所需生理学和农艺学特征对各种各样的植物和植物细胞系统进行工程化。在优选的实施方案中,用于工程化的靶植物和植物细胞包括但不限于那些单子叶植物和双子叶植物,诸如作物,包括谷类作物(例如,小麦、玉米、水稻、小米、大麦)、水果作物(例如,番茄、苹果、梨、草莓、橙子)、饲料作物(例如,苜蓿)、根用蔬菜作物(例如,胡萝卜、马铃薯、甜菜、山药)、叶类蔬菜作物(例如,生菜、菠菜);开花植物(例如,矮牵牛、玫瑰、菊花)、针叶树和松树(例如,松木、云杉);用于植物修复的植物(例如,重金属积累植物);油料作物(例如,向日葵、油菜籽)以及用于实验目的的植物(例如,拟南芥)。因此,这些方法和系统可以用于广泛的植物,诸如与属于以下目的双子叶植物一起使用:木兰目(Magniolales)、八角目(Illiciales)、樟目(Laurales)、胡椒目(Piperales)、马览铃目(Aristochiales)、睡莲目

(Nymphaeales)、毛茛目(Ranunculales)、罂粟目(Papeverales)、瓶子草科(Sarraceniaceae)、昆栏树目(Trochodendrales)、金缕梅目(Hamamelidales)、杜仲目(Eucomiales)、莱脱纳目(Leitneriales)、杨梅目(Myricales)、壳斗目(Fagales)、木麻黄目(Casuarinales)、石竹目(Caryophyllales)、肉穗果目(Batales)、蓼目(Polygonales)、蓝雪目(Plumbaginales)、五桠果目(Dilleniales)、山茶目(Theales)、锦葵目(Malvales)、荨麻目(Urticales)、玉蕊目(Lecythidales)、紫堇目(Violales)、杨柳目(Salicales)、白花菜目(Capparales)、欧石楠目(Ericales)、岩梅目(Diapensales)、柿树目(Ebenales)、报春花目(Primulales)、蔷薇目(Rosales)、豆目(Fabales)、川苔草目(Podostemales)、小二仙草目(Haloragales)、桃金娘目(Myrtales)、山茱萸目(Cornales)、睡莲目(Proteales)、檀香目(San tales)、大花草目(Rafflesiales)、卫矛目(Celastrales)、大戟目(Euphorbiales)、鼠李目(Rhamnales)、无患子目(Sapindales)、胡桃目(Juglandales)、牻牛儿苗目(Geraniales)、远志目(Polygalales)、伞形目(Umbellales)、龙胆目(Gentianales)、花葱目(Polemoniales)、唇形目(Lamiales)、车前草目(Plantaginales)、玄参目(Scrophulariales)、桔梗目(Campanulales)、茜草目(Rubiales)、川续断目(Dipsacales)和菊目(Asterales);这些方法和CRISPR-Cas系统可以与诸如属于以下目的单子叶植物一起使用:泽泻目(Alismatales)、水鳖目(Hydrocharitales)、茨藻目(Najadales)、霉草目(Triuridales)、鸭跖草目(Commelinales)、谷精草目(Eriocaulales)、帚灯草目(Restionales)、禾本目(Poales)、灯芯草目(Juncales)、莎草目(Cyperales)、香蒲目(Typhales)、凤梨目(Bromeliales)、姜目(Zingiberales)、槟榔目(Arecales)、环花目(Cyclanthales)、露兜树目(Pandanales)、天南星目(Arales)、百合目(Lilliales)和兰目(Orchid ales),或者与属于裸子植物的植物一起使用,例如属于以下目的植物:松目(Pinales)、银杏目(Ginkgoales)、苏铁目(Cycadales)、南洋杉目(Araucariales)、柏目(Cupressales)和麻黄目(Gnetales)。

[0790] 本文所述的AD功能化的CRISPR系统和使用方法可以用于广泛的植物物种,这些植物物种包括在以下双子叶植物、单子叶植物或裸子植物属的非限制性列表中:颠茄属(Atropa)、油丹属(Alseodaphne)、槚如树属(Anacardium)、落花生属(Arachis)、琼楠属(Beilschmiedia)、芸苔属(Brassica)、红花属(Carthamus)、木防己属(Cocculus)、巴豆属(Croton)、黄瓜属(Cucumis)、柑橘属(Citrus)、西瓜属(Citrullus)、辣椒属(Capsicum)、长春花属(Catharanthus)、椰子属(Cocos)、咖啡属(Coffea)、南瓜属(Cucurbita)、胡萝卜属(Daucus)、杜氏木属(Duguetia)、花菱草属(Eschscholzia)、无花果属(Ficus)、草莓属(Fragaria)、海罂粟属(Glaucium)、大豆属(Glycine)、棉属(Gossypium)、向日葵属(Helianthus)、橡胶树属(Hevea)、天仙子属(Hyoscyamus)、莴苣属(Lactuca)、卷枝藤属(Landolphia)、亚麻属(Linum)、木姜子属(Litsea)、番茄属(Lycopersicon)、羽扇豆属(Lupinus)、木薯属(Manihot)、马郁兰属(Majorana)、苹果属(Malus)、苜蓿属(Medicago)、烟草属(Nicotiana)、木犀榄(Olea)、银胶菊属(Parthenium)、罂粟属(Papaver)、鳄梨属(Persea)、菜豆属(Phaseolus)、黄连木属(Pistacia)、豌豆属(Pisum)、梨属(Pyrus)、李属(Prunus)、萝卜属(Raphanus)、蓖麻属(Ricinus)、千里光属(Senecio)、风龙属(Sinomenium)、千金藤属(Stephania)、欧白芥属(Sinapis)、茄属(Solanum)、可可属(Theobroma)、三叶草属(Trifolium)、胡芦巴属(Trigonella)、蚕豆属(Vicia)、长春花属

(Vinca)、葡萄属(Vitis)和豇豆属(Vigna);以及以下属:葱属(Allium)、须芒草属(Andropogon)、画眉草属(Aragrostis)、天门冬属(Asparagus)、燕麦属(Avena)、狗牙根属(Cynodon)、油棕属(Elaeis)、羊茅属(Festuca)、羊茅黑麦草属(Festulolium)、萱草属(Heterocallis)、大麦属(Hordeum)、浮萍属(Lemna)、黑麦草属(Lolium)、芭蕉属(Musa)、稻属(Oryza)、黍属(Panicum)、皇竹草属(Pannesetum)、梯牧草属(Phleum)、早熟禾属(Poa)、黑麦属(Secale)、蜀黍(Sorghum)、小麦属(Triticum)、玉蜀黍属(Zea)、冷杉属(Abies)、杉木属(Cunninghamia)、麻黄属(Ephedra)、云杉属(Picea)、松属(Pinus)和黄杉属(Pseudotsuga)。

[0791] 这些AD功能化的CRISPR系统和使用方法还可以用于广泛的“藻类”或“藻类细胞”,包括例如选自多种真核门(包括红藻门(Rhodophyta)(红藻)、绿藻门(Chlorophyta)(绿藻)、褐藻门(Phaeophyta)(褐藻)、硅藻门(Bacillariophyta)(硅藻)、真眼点藻门(Eustigmatophyta)和鞭毛藻类)的藻类,以及原核门蓝藻门(Cyanobacteria)(蓝绿藻)。术语“藻类”包括例如选自以下属的藻类:双眉藻属(Amphora)、鱼腥藻属(Anabaena)、纤维藻属(Anikstrodesmis)、葡萄藻属(Botryococcus)、角毛藻属(Chaetoceros)、衣藻属(Chlamydomonas)、小球藻属(Chlorella)、绿球藻属(Chlorococcum)、小环藻属(Cyclotella)、细柱藻属(Cylindrotheca)、杜氏藻属(Dunaliella)、球石藻属(Emiliana)、眼虫属(Euglena)、球藻属(Hematococcus)、等鞭金藻属(Isochrysis)、单鞭金藻属(Monochrysis)、单针藻属(Monoraphidium)、微绿球藻属(Nannochloris)、拟微球藻属(Nannnochloropsis)、舟形藻属(Navicula)、肾鞭藻属(Nephrochloris)、肾爿藻属(Nephroselmis)、菱形藻属(Nitzschia)、节球藻属(Nodularia)、念珠藻属(Nostoc)、鞭毛藻属(Oochromonas)、卵囊藻属(Oocystis)、颤藻属(Oscillartoria)、巴夫藻属(Pavlova)、褐指藻属(Phaeodactylum)、扁藻属(Playtmonas)、颗石藻属(Pleurochrysis)、紫菜属(Porhyra)、假鱼腥藻属(Pseudoanabaena)、塔胞藻属(Pyramimonas)、裂丝藻属(Stichococcus)、聚球藻属(Synechococcus)、集胞藻属(Synechocystis)、四爿藻属(Tetraselmis)、海链藻属(Thalassiosira)和束毛藻属(Trichodesmium)。

[0792] 可以根据本发明的方法处理植物的一部分,即“植物组织”,以产生改良的植物。植物组织还包括植物细胞。本文所用的术语“植物细胞”是指活植物的个体单位,不论在完整的全植物中还是呈在体外组织培养物中、培养基或琼脂上、生长培养基或缓冲液中的悬浮液中生长的分离形式,或作为更高组织化合体(诸如植物组织、植物器官或全植物)的一部分。

[0793] “原生质体”是指通过使用例如机械或酶促手段完全或部分去除保护性细胞壁从而形成活植物的具有生化功能的完整单位的植物细胞,可以改造植物细胞的细胞壁,使其增殖并且在适当的生长条件下使其再生生长成全植物。

[0794] 术语“转化”广义上是指借助农杆菌(Agrobacteria)或者多种化学或物理方法中的一种通过引入DNA对植物宿主进行遗传修饰的过程。如本文所用,术语“植物宿主”是指植物,包括植物的任何细胞、组织、器官或子代。可以转化许多合适的植物组织或植物细胞,包括但不限于原生质体、体细胞胚、花粉、叶、籽苗、茎、愈伤组织、匍匐茎、试管块茎和芽。植物组织还指这种植物、种子、子代、繁殖体(无论有性还是无性生殖)以及这些中任一项的后代(诸如插条或种子)的任何克隆。

[0795] 本文所用的术语“转化”是指其中已经引入外源DNA分子诸如构建体的细胞、组织、器官或生物。引入的DNA分子可被整合到受体细胞、组织、器官或生物的基因组DNA中，使得引入的DNA分子被传递给后续子代。在这些实施方案中，“转化”或“转基因”细胞或植物还可包括细胞或植物的子代，以及由利用这种转化植物作为杂交亲本并且由于存在引入的DNA分子而表现出表型改变的育种程序产生的子代。优选地，转基因植物是可育的并且能够通过有性繁殖将引入的DNA传递给子代。

[0796] 术语“子代”诸如转基因植物的子代是出生于、产生于或来源于植物或转基因植物的子代。引入的DNA分子还可被瞬时引入受体细胞中，使得引入的DNA分子不会被后续子代遗传，因此不被认为是“转基因的”。因此，如本文所用，“非转基因”植物或植物细胞是不包含稳定整合到其基因组中的外源DNA的植物。

[0797] 本文所用的术语“植物启动子”是能够启动植物细胞中的转录的启动子，无论其起源是否是植物细胞。示例性合适的植物启动子包括但不限于从植物、植物病毒以及包含在植物细胞中表达的基因的细菌(诸如农杆菌(Agrobacterium)或根瘤菌(Rhizobium))获得的启动子。

[0798] 如本文所用，“真菌细胞”是指真菌界内的任何类型的真核细胞。真菌界内的门包括菌门子囊菌门(Ascomycota)、担子菌门(Basidiomycota)、芽枝霉门(Blastocladiomycota)、壶菌门(Chytridiomycota)、球囊菌门(Glomeromycota)、微孢子虫目(Microsporidia)和新丽鞭毛菌门(Neocallimastigomycota)。真菌细胞可包括酵母、霉菌和丝状真菌。在一些实施方案中，真菌细胞是酵母细胞。

[0799] 如本文所用，术语“酵母细胞”是指子囊菌门和担子菌门内的任何真菌细胞。酵母细胞可包括芽殖酵母细胞、裂殖酵母细胞和霉菌细胞。不限于这些生物，在实验室和工业环境中使用的许多类型的酵母是子囊菌门的一部分。在一些实施方案中，酵母细胞是酿酒酵母(S.cerervisiae)、马克斯克鲁维酵母(Kluyveromyces marxianus)或东方伊萨酵母(Issatchenkia orientalis)细胞。其他酵母细胞可包括但不限于假丝酵母属(Candida spp.)(例如，白色念珠菌(Candida albicans))、亚罗酵母属(Yarrowia spp.)(例如，解脂亚罗酵母(Yarrowia lipolytica))、毕赤酵母属(Pichia spp.)(例如，巴斯德毕赤酵母(Pichia pastoris))、克鲁维酵母属(Kluyveromyces spp.)(例如，乳酸克鲁维酵母(Kluyveromyces lactis)和马克斯克鲁维酵母)、脉孢菌属(Neurospora spp.)(例如，粗糙脉孢菌(Neurospora crassa))、镰刀菌属(Fusarium spp.)(例如，尖孢镰刀菌(Fusariumoxysporum))和伊萨酵母属(Issatchenkia spp.)(例如，东方伊萨酵母，又名库德里阿兹氏毕赤酵母(Pichia kudriavzevii)和嗜热假丝酵母(Candida acidothermophilum))。在一些实施方案中，真菌细胞是丝状真菌细胞。如本文所用，术语“丝状真菌细胞”是指在丝状体即菌丝或菌丝体中生长的任何类型的真菌细胞。丝状真菌细胞的实例可包括但不限于曲霉属(Aspergillus spp.)(例如，黑曲霉(Aspergillus niger))、木霉属(Trichoderma spp.)(例如，里氏木霉(Trichoderma reesei))、根霉(Rhizopus spp.)(例如，米根霉(Rhizopusoryzae))和被孢霉属(Mortierella spp.)(例如，深黄被孢霉(Mortierellaisabellina))。

[0800] 在一些实施方案中，真菌细胞是工业菌株。如本文所用，“工业菌株”是指在工业过程(例如以商业或工业规模生产产品)中使用或从工业过程中分离的任何真菌细胞菌株。工

业菌株可指通常在工业过程中使用的真菌物种,或者可指也可用于非工业目的(例如,实验室研究)的真菌物种的分离菌。工业过程的实例可包括发酵(例如,在食品或饮料产品的生产中)、蒸馏、生物燃料生产、化合物生产和多肽生产。工业菌株的实例可包括但不限于JAY270和ATCC4124。

[0801] 在一些实施方案中,真菌细胞是多倍体细胞。如本文所用,"多倍体"细胞可指其基因组以一个以上拷贝存在的任何细胞。多倍体细胞可指天然发现处于多倍体状态的细胞类型,或者可指已被诱导成以多倍体状态存在的细胞(例如,通过对减数分裂、胞质分裂或DNA复制进行特定调控、改变、失活、激活或修饰)。多倍体细胞可指其整个基因组是多倍体的细胞,或者可指在特定的目标基因组基因座中是多倍体的细胞。不希望受理论的束缚,据认为指导RNA的丰度在多倍体细胞的基因组工程化中可能比在单倍体细胞中更经常地是限速组分,并且因此使用本文所述的AD功能化的CRISPR系统的方法可利用使用某种真菌细胞类型。

[0802] 在一些实施方案中,真菌细胞是二倍体细胞。如本文所用,"二倍体"细胞可指其基因组以两个拷贝存在的任何细胞。二倍体细胞可指天然发现处于二倍体状态的细胞类型,或者可指已被诱导成以二倍体状态存在的细胞(例如,通过对减数分裂、胞质分裂或DNA复制进行特定调控、改变、失活、激活或修饰)。例如,可将酿酒酵母菌株S228C维持在单倍体或二倍体状态。二倍体细胞可指其整个基因组是二倍体的细胞,或者可指在特定的目标基因组基因座中是二倍体的细胞。在一些实施方案中,真菌细胞是单倍体细胞。如本文所用,"单倍体"细胞可指其基因组以一个拷贝存在的任何细胞。单倍体细胞可指天然发现处于单倍体状态的细胞类型,或者可指已被诱导成以单倍体状态存在的细胞(例如,通过对减数分裂、胞质分裂或DNA复制进行特定调控、改变、失活、激活或修饰)。例如,可将酿酒酵母菌株S228C维持在单倍体或二倍体状态。单倍体细胞可指其整个基因组是单倍体的细胞,或者可指在特定的目标基因组基因座中是单倍体的细胞。

[0803] 如本文所用,"酵母表达载体"是指包含编码RNA和/或多肽的一个或多个序列的核酸,并且还可包含控制一种或多种核酸表达的任何所需元件,以及能够在酵母细胞内部复制和维持表达载体的任何元件。许多合适的酵母表达载体及其特征在本领域中是已知的;例如,在Yeast Protocols,第2版,Xiao,W.编辑(Humana Press,New York,2007)以及Buckholz,R.G.和Gleeson,M.A.(1991)Biotechnology(NY)9(11):1067-72中示出了不同载体和技术。酵母载体可包含但不限于着丝粒(CEN)序列、自主复制序列(ARS)、可操作地连接到目标序列或基因的启动子(诸如RNA聚合酶III启动子)、终止子(诸如RNA聚合酶III终止子)、复制起点和标记基因(例如,营养缺陷标记、抗生素标记或其他选择性标记)。用于酵母的表达载体的实例可包括质粒、酵母人工染色体、2μ质粒、酵母整合质粒、酵母复制质粒、穿梭载体和游离质粒。

[0804] AD功能化的CRISPR系统组分在植物和植物细胞基因组中的稳定整合

[0805] 在特定的实施方案中,设想了引入编码AD功能化的CRISPR系统的组分的多核苷酸以稳定整合到植物细胞的基因组中。在这些实施方案中,可以根据指导RNA以及/或者腺苷脱氨酶和Cas13的融合蛋白何时、何处以及在什么条件下表达来对转化载体或表达系统的设计进行调节。

[0806] 在特定的实施方案中,设想了将AD功能化的CRISPR系统的组分稳定地引入植物细

胞的基因组DNA中。另外地或另选地，设想了引入AD功能化的CRISPR系统的组分以稳定整合到植物细胞器(诸如但不限于质体、线粒体或叶绿体)的DNA中。

[0807] 用于稳定整合到植物细胞的基因组中的表达系统可包含以下元件中的一种或多种：启动子元件，其可以用于在植物细胞中表达RNA以及/或者腺苷脱氨酶和Cas13的融合蛋白；5'非翻译区域，用以增强表达；内含子元件，用以进一步增强某些细胞诸如单子叶植物细胞中的表达；多克隆位点，用以为插入指导RNA以及/或者腺苷脱氨酶和Cas13编码序列的融合蛋白以及其他所需元件提供方便的限制性位点；以及3'非翻译区，用以有效终止表达的转录物。

[0808] 表达系统的这些元件可在一个或多个环状表达构建体(诸如质粒或转化载体)或非环状构建体(诸如线性双链DNA)上。

[0809] 在一个特定的实施方案中，AD功能化的CRISPR表达系统至少包含：编码与植物中的靶序列杂交的指导RNA(gRNA)的核苷酸序列，并且其中指导RNA包含指导序列和正向重复序列；以及编码腺苷脱氨酶和Cas13的融合蛋白的核苷酸序列，其中组分(a)或(b)位于相同或不同的构建体上，并且由此不同的核苷酸序列可以受在植物细胞中可操作的相同或不同调控元件的控制。

[0810] 可通过多种常规技术将包含AD功能化的CRISPR系统的组分以及模板序列(在适用的情况下)的一个或多个DNA构建体引入植物、植物部分或植物细胞的基因组中。该过程一般包括以下步骤：选择合适的宿主细胞或宿主组织，将一个或多个构建体引入宿主细胞或宿主组织中，并从中再生植物细胞或植物。

[0811] 在特定的实施方案中，可使用诸如但不限于植物细胞原生质体的电穿孔、显微注射、气溶胶束注射等技术将DNA构建体引入植物细胞中，或者可以使用生物弹射法诸如DNA粒子轰击将DNA构建体直接引入植物组织(还可参见Fu等人，Transgenic Res.2000年2月；9(1):11-9)。粒子轰击的基础是使包覆有目标基因的粒子朝向细胞加速，从而使得粒子穿透原生质并且通常稳定整合到基因组中。(参见例如Klein等人，Nature(1987)；Klein等人，Bio/Technology(1992)；Casas等人，Proc.Natl.Acad.Sci.USA(1993))。

[0812] 在特定的实施方案中，可通过农杆菌介导的转化将含有AD功能化的CRISPR系统的组分的DNA构建体引入植物中。可将DNA构建体与合适的T-DNA侧接区域组合，并引入常规的根癌农杆菌(Agrobacterium tumefaciens)宿主载体中。可以通过侵染植物或通过用含有一种或多种Ti(肿瘤诱导)质粒的农杆菌属细菌孵育植物原生质体将外源DNA掺入植物基因组中。(例如参见Fraley等人，(1985)；Rogers等人，(1987)；以及美国专利号5,563,055)。

[0813] 植物启动子

[0814] 为了确保在植物细胞中适当表达，通常将本文所述的AD功能化的CRISPR系统的组分置于植物启动子即在植物细胞中可操作的启动子的控制下。设想了使用不同类型的启动子。

[0815] 组成型植物启动子是能够在植物的所有或几乎所有发育阶段在所有或几乎所有植物组织中表达其控制(称为“组成型表达”)的开放阅读框(ORF)的启动子。组成型启动子的一个非限制性实例是花椰菜花叶病毒35S启动子。“调控启动子”是指并非组成性而是以时间和/或空间调控方式指导基因表达的启动子，并且包括组织特异性启动子、组织优选启动子和诱导型启动子。不同启动子可指导基因在不同组织或细胞类型中或者在不同发育阶

段或者响应于不同环境条件而表达。在特定的实施方案中,一种或多种AD功能化的CRISPR组分在组成型启动子诸如花椰菜花叶病毒35S启动子的控制下表达。可以利用组织优选型启动子来靶向特定植物组织内的某些细胞类型(例如叶或根的维管细胞)或种子的特定细胞的增强表达。用于AD功能化的CRISPR系统的特定启动子的实例可见于Kawamata等人,(1997)Plant Cell Physiol 38:792-803;Yamamoto等人,(1997)Plant J 12:255-65;Hire等人,(1992)Plant Mol Biol 20:207-18;Kuster等人,(1995)Plant Mol Biol 29:759-72;以及Capana等人,(1994)Plant Mol Biol 25:681-91。

[0816] 为了避免脱氨酶的非特异性活性,在有限的情况下诱导型启动子可以表达AD功能化的CRISPR系统的一种或多种组分可能是有意义的。在特定的实施方案中,AD功能化的CRISPR系统的一个或多个元件在诱导型启动子的控制下表达。可诱导并且允许对基因编辑或基因表达进行时空控制的启动子的实例可使用能量形式。能量形式可包括但不限于声能、电磁辐射、化学能和/或热能。诱导型系统的实例包括四环素诱导型启动子(Tet-On或Tet-Off)、小分子双杂交转录激活系统(FKBP、ABA等)或光诱导型系统(光敏素、LOV结构域或隐花色素),诸如以序列特异性方式指导转录活性的变化的光诱导型转录效应因子(LITE)。光诱导型系统的组分可包括腺苷脱氨酶和Cas13的融合蛋白,一种光响应性细胞色素异二聚体(例如,来自拟南芥)。诱导型DNA结合蛋白的其他实例及其使用方法在US 61/736465和US61/721,283中提供,这些文献全文据此通过引用并入。

[0817] 在特定的实施方案中,可以通过使用例如化学调控型启动子(即由此应用外源化学物质来诱导基因表达)来实现瞬时或诱导型表达。还可以通过化学阻遏型启动子(其中应用化学物质来阻遏基因表达)来获得对基因表达的调节。化学诱导型启动子包括但不限于由苯磺酰胺类除草安全剂激活的玉米ln2-2启动子(De Veylder等人,(1997)Plant Cell Physiol 38:568-77)、由用作萌前除草剂的疏水性亲电子化合物激活的玉米GST启动子(GST-11-27,WO93/01294),以及由水杨酸激活的烟草PR-1a启动子(Ono等人,(2004)Biosci Biotechnol Biochem 68:803-7)。由抗生素调控的启动子诸如四环素诱导型和四环素阻遏型启动子(Gatz等人,(1991)Mol Gen Genet 227:229-37;美国专利号5,814,618和5,789,156)也可以在本文中使用。

[0818] 易位到特定植物细胞器和/或在特定植物细胞器中表达

[0819] 表达系统可包含用于易位到特定植物细胞器和/或在特定植物细胞器中表达的元件。

[0820] 叶绿体靶向

[0821] 在特定的实施方案中,设想了AD功能化的CRISPR系统用于特异性修饰叶绿体基因或确保叶绿体中的表达。为此,使用叶绿体转化方法或将AD功能化的CRISPR组分区室化到叶绿体。例如,在质体基因组中引入遗传修饰可以减少生物安全问题,诸如通过花粉产生的基因流动。

[0822] 叶绿体转化方法是本领域已知的,包括粒子轰击、PEG处理和显微注射。另外,可以使用如WO2010061186中所述的涉及使转化盒从核基因组易位到质体的方法。

[0823] 另选地,设想了将一种或多种AD功能化的CRISPR组分组分靶向植物叶绿体。这是通过在表达构建物中掺入编码叶绿体转运肽(CTP)或质体转运肽的序列(其可操作地连接到编码腺苷脱氨酶和Cas13的融合蛋白的序列的5'区域)来实现的。在易位到叶绿体期间的

加工步骤中去除CTP。对所表达的蛋白进行叶绿体靶向是本领域技术人员熟知的(参见例如Protein Transport into Chloroplasts,2010,Annual Review of Plant Biology,第61卷:157-180)。在此类实施方案中,还希望将指导RNA靶向植物叶绿体。可以借助于叶绿体定位序列使指导RNA易位到叶绿体中的方法和构建体描在例如US20040142476中描述,该文献通过引用并入本文。可以将构建体的此类变体掺入本发明的表达系统中,以有效地易位AD功能化的CRISPR系统组分。

[0824] 将编码AD功能化的CRISPR系统的多核苷酸引入藻类细胞中。

[0825] 转基因藻类(或其他植物,诸如油菜)在生产植物油或生物燃料诸如醇(尤其是甲醇和乙醇)或其他产品中可能特别有用。可以将它们工程化为表达或过量表达用于油或生物燃料行业的高含量油或醇。

[0826] US 8945839描述了一种使用Cas9来工程化微藻类(莱茵衣藻(Chlamydomonas reinhardtii)细胞)的方法。使用类似的工具,可以将本文所述的AD功能化的CRISPR系统的方法应用于衣藻物种和其他藻类。在特定的实施方案中,将CRISPR-Cas蛋白(例如,Cas13)、腺苷脱氨酶(其可与CRISPR-Cas蛋白或适配体结合衔接蛋白融合)和指导RNA引入使用在组成型启动子诸如Hsp70A-Rbc S2或Beta2-微管蛋白的控制下表达腺苷脱氨酶和Cas13的融合蛋白的载体表达的藻类中。任选地,使用含有T7启动子的载体来递送指导RNA。另选地,可以将Cas13mRNA和体外转录的指导RNA递送到藻类细胞。电穿孔方案对于技术人员是可用的,诸如来源于GeneArt Chlamydomonas Engineering试剂盒的标准推荐方案。

[0827] 将AD功能化的CRISPR系统组分引入酵母细胞

[0828] 在特定的实施方案中,本发明涉及AD功能化的CRISPR系统用于酵母细胞的基因组编辑中的用途。转化可以用于引入编码AD功能化的CRISPR系统组分的多核苷酸的酵母细胞的方法在Kawai等人,2010,Bioeng Bugs.2010年11月至12月;1(6):395-403)中描述。非限制性实例包括通过乙酸锂处理(其还可包括载体DNA和PEG处理)、轰击或电穿孔转化酵母细胞。

[0829] AD功能化的CRISPR系统组分在植物和植物细胞中瞬时表达

[0830] 在特定的实施方案中,设想了指导RNA和/或CRISPR-Cas基因在植物细胞中瞬时表达。在这些实施方案中,仅当指导RNA、CRISPR-Cas蛋白(例如,Cas13)和腺苷脱氨酶(其可与CRISPR-Cas蛋白或适配体结合衔接蛋白融合)存在于细胞中时,AD功能化的CRISPR系统才能确保对靶基因的修饰,从而可以进一步控制基因组修饰。由于CRISPR-Cas蛋白的表达是瞬时的,因此从此类植物细胞再生的植物通常不包含外源DNA。在特定的实施方案中,CRISPR-Cas蛋白由植物细胞稳定表达,并且指导序列瞬时表达。

[0831] 在特定的实施方案中,可以使用植物病毒载体将AD功能化的CRISPR系统组分引入植物细胞(Scholthof等人,1996,Annu Rev Phytopathol.1996;34:299-323)。在另外的特定实施方案中,所述病毒载体是来自DNA病毒的载体。例如,双生病毒(例如,卷心菜曲叶病毒、豆黄矮病毒、小麦矮小病毒、番茄曲叶病毒、玉米条纹病毒、烟草曲叶病毒或番茄金色花叶病毒)或矮缩病毒(例如,蚕豆坏死黄化病毒)。在其他特定的实施方案中,所述病毒载体是来自RNA病毒的载体。例如,烟草脆裂病毒属(例如,烟草脆裂病毒、烟草花叶病毒)、马铃薯X病毒属(例如,马铃薯X病毒)或大麦病毒属(例如,大麦条纹花叶病毒)。植物病毒的复制基因组是非整合载体。

[0832] 在特定的实施方案中，用于瞬时表达AD功能化的CRISPR系统的载体是例如pEAQ载体，该载体是针对在原生质体中进行农杆菌介导的瞬时表达而定制的(Sainsbury F.等人，Plant Biotechnol J.2009年9月;7(7):682-93)。使用经修饰的卷心菜曲叶病毒(CaLCuV)载体在表达CRISPR酶的稳定转基因植物中表达指导RNA证明了基因组位置的精确靶向(Scientific Reports 5,文章编号:14926(2015),doi:10.1038/srep14926)。

[0833] 在特定的实施方案中，可以将编码指导RNA和/或CRISPR-Cas基因的双链DNA片段瞬时引入植物细胞中。在此类实施方案中，提供引入的双链DNA片段的量足以修饰细胞，但在经过预期的时间段之后或在一次或多次细胞分裂之后不会持续存在。用于直接将DNA转移到植物中的方法是技术人员已知的(参见例如Davey等人，Plant Mol Biol.1989年9月;13(3):273-85)。

[0834] 在其他实施方案中，将编码CRISPR-Cas蛋白(例如，Cas13)和/或腺苷脱氨酶(其可与CRISPR-Cas蛋白或适配体结合衔接蛋白融合)的RNA多核苷酸引入植物细胞中，然后由宿主细胞翻译和加工，从而产生足以修饰细胞的量的蛋白(在存在至少一种指导RNA的情况下)，但所产生的蛋白在经过预期的时间段之后或在一次或多次细胞分裂之后不会持续存在。用于将mRNA引入植物原生质体进行瞬时表达的方法是技术人员已知的(参见例如Gallie,Plant Cell Reports(1993),13;119-122)。

[0835] 还设想了上述不同方法的组合。

[0836] 将AD功能化的CRISPR系统组分递送到植物细胞

[0837] 在特定的实施方案中，将AD功能化的CRISPR系统的一种或多种组分直接递送到植物细胞是有意义的。这尤其对于产生非转基因植物是有意义的(参见下文)。在特定的实施方案中，在植物或植物细胞外部制备一种或多种AD功能化的CRISPR系统组分并将其递送到细胞。例如，在特定的实施方案中，在引入植物细胞之前在体外制备CRISPR-Cas蛋白。可以通过本领域技术人员已知的多种方法来制备CRISPR-Cas蛋白，包括重组生产。在表达之后，将CRISPR-Cas蛋白分离，如果需要的话进行再折叠，进行纯化并且任选地进行处理，以去除任何纯化标签，诸如His标签。一旦获得未加工、部分纯化或更完全纯化的CRISPR-Cas蛋白，就可将该蛋白引入植物细胞中。

[0838] 在特定的实施方案中，将CRISPR-Cas蛋白与靶向目标基因的指导RNA混合，以形成预组装的核糖核蛋白。

[0839] 可以经由电穿孔、通过用CRISPR-Cas相关的基因产物包被的粒子轰击、通过化学转染或通过某种其他跨细胞膜转运的手段将各个组分或预组装的核糖核蛋白引入植物细胞。例如，已证明用预组装的CRISPR核糖核蛋白转染植物原生质体可以确保对植物基因组的靶向修饰(如Woo等人,Nature Biotechnology,2015;DOI:10.1038/nbt.3389)。

[0840] 在特定的实施方案中，使用纳米颗粒将AD功能化的CRISPR系统组分引入植物细胞中。可以将这些组分作为蛋白或核酸或者以其组合上载到纳米颗粒上或包装在纳米颗粒中并施加到植物上(诸如，在WO 2008042156和US20130185823中所述)。具体地，本发明的实施方案包括与编码CRISPR-Cas蛋白的一种或多种DNA分子(例如，Cas13)、编码腺苷脱氨酶(其可与CRISPR-Cas蛋白或适配体结合衔接蛋白融合)的一种或多种DNA分子以及编码指导RNA和/或分离的指导RNA的DNA分子一起上载或包装的纳米颗粒，如WO2015089419中所述。

[0841] 将AD功能化的CRISPR系统的一种或多种组分引入植物细胞的其他手段是通过使

用细胞穿透肽(CPP)。因此,具体地,本发明的实施方案包括包含与CRISPR-Cas蛋白连接的细胞穿透肽的组合物。在本发明的特定实施方案中,将CRISPR-Cas蛋白和/或指导RNA与一种或多种CPP联接,以将它们有效地转运到植物原生质体内部(Ramakrishna(Genome Res.2014年6月;24(6):1020-7(针对人类细胞中的Cas9)))。在其他实施方案中,CRISPR-Cas基因和/或指导RNA由一种或多种环状或非环状DNA分子编码,这些DNA分子与一种或多种CPP联接以用于植物原生质体递送。然后将植物原生质体再生为植物细胞并进一步再生为植物。一般将CPP描述为来源于蛋白或嵌合序列的少于35个氨基酸的短肽,其能够以受体非依赖性方式跨细胞膜转运生物分子。CPP可以是阳离子肽、具有疏水序列的肽、两亲性肽、具有富含脯氨酸和抗微生物序列的肽以及嵌合或二分肽(Pooga和Langel 2005)。CPP能够穿透生物膜,并因此触发各种生物分子跨细胞膜移动到细胞质中并改进其细胞内路线,从而促进生物分子与靶标相互作用。CPP的实例尤其包括:Tat(HIV 1型进行病毒复制所需的核转录激活蛋白)、穿透素、Kaposi成纤维细胞生长因子(FGF)信号肽序列、整合素β3信号肽序列;聚精氨酸肽Args序列、富含鸟嘌呤的分子转运体、甜箭肽等。

[0842] 使用AD功能化的CRISPR系统制造遗传修饰的非转基因植物

[0843] 在特定的实施方案中,本文所述的方法用于修饰内源基因或修饰其表达,而无需将任何外源基因(包括编码CRISPR组分的那些基因)永久性引入植物的基因组中,从而避免植物的基因组中存在外源DNA。这可能是有意义的,因为对非转基因植物的监管要求没那么严格。

[0844] 在特定的实施方案中,这通过瞬时表达AD功能化的CRISPR系统组分来确保。在特定的实施方案中,一种或多种组分在一种或多种病毒载体上表达,所述病毒载体产生足够的CRISPR-Cas蛋白、腺苷脱氨酶和指导RNA以根据本文所述的方法始终稳定地确保对目标基因的修饰。

[0845] 在特定的实施方案中,确保AD功能化的CRISPR系统构建体在植物原生质体中瞬时表达,因此不整合到基因组中。有限的表达窗口足以允许AD功能化的CRISPR系统确保如本文所述的对靶基因的修饰。

[0846] 在特定的实施方案中,借助于粒状递送分子(诸如上文所述的纳米颗粒或CPP分子),将AD官能化的CRISPR系统的不同组分分别地或混合地引入植物细胞、原生质体或植物组织。

[0847] AD功能化的CRISPR系统组分的表达可以通过腺苷脱氨酶的脱氨酶活性诱导基因组的靶向修饰。上文所述的不同策略允许CRISPR介导的靶向基因组编辑,而无需将AD功能化的CRISPR系统组分引入植物基因组中。瞬时引入植物细胞中的组分通常在杂交时被去除。

[0848] 植物培养和再生

[0849] 在特定的实施方案中,可以培养具有修饰的基因组并且通过本文所述的任何方法产生或获得的植物细胞,以再生具有转化或修饰的基因型因此具有所需表型的整株植物。常规的再生技术是本领域技术人员众所周知的。此类再生技术的特定实例依赖于在组织培养生长培养基中对某些植物激素进行操纵,并且通常依赖于已与所需核苷酸序列一起引入的杀生物剂和/或除草剂标记。在另外的特定实施方案中,植物再生是从培养的原生质体、植物愈伤组织、外植体、器官、花粉、胚或其部分获得的(参见例如Evans等人,(1983),

Handbook of Plant Cell Culture;Klee等人,(1987)Ann.Rev.of Plant Phys.)。

[0850] 在特定的实施方案中,本文所述的转化或改良的植物可以自花传粉以提供本发明的纯合改良植物的种子(对于DNA修饰而言纯合),或者可以与非转基因植物或不同改良植物杂交以提供杂合植物的种子。在将重组DNA引入植物细胞中的情况下,这种杂交所得的植物是对于重组DNA分子而言杂合的植物。通过由改良植物杂交获得并且包含遗传修饰(可以是重组DNA)的这种纯合植物和杂合植物在本文中均称为"后代"。后代植物是起源于原始转基因植物并且含有通过本文提供的方法引入的基因组修饰或重组DNA分子的植物。另选地,可以使用AD功能化的CRISPR系统通过上述方法中的一种获得遗传修饰的植物,由此不会将外源DNA掺入基因组中。通过进一步育种获得的此类植物的后代也可包含遗传修饰。育种可以通过通常用于不同作物的任何育种方法来进行(Allard,Principles of Plant Breeding,John Wiley&Sons,NY,U.of CA,Davis,CA,50-98(1960))。

[0851] 产生农艺性状增强的植物

[0852] 本文提供的AD功能化的CRISPR系统可以用于引入靶向A-G和T-C突变。通过在单个细胞中共表达旨在实现多种修饰的多个靶向RNA,可以确保多重基因组修饰。该技术可以用于对具有改良特性的植物进行高精度工程化,所述改良特性包括营养质量增强、抗病性增强以及抗生物和非生物胁迫性增强,以及有商业价值的植物产品或异源化合物的产量增加。

[0853] 在特定的实施方案中,本文所述的AD功能化的CRISPR系统用于引入靶向A-G和T-C突变。这种突变可以是无义突变(例如,提前终止密码子)或错义突变(例如,编码不同的氨基酸残基)。这在某些内源基因中的A-G和T-C突变可以赋予或产生所需性状的情况下是有意义的。

[0854] 本文所述的方法通常会产生"改良植物",因为与野生型植物相比,它们具有一种或多种期望的性状。在特定的实施方案中,获得的植物、植物细胞或植物部分是包含掺入植物的全部或部分细胞的基因组中的外源DNA序列的转基因植物。在特定的实施方案中,获得非转基因的遗传修饰的植物、植物部分或细胞,因为没有将外源DNA序列掺入植物的任何植物细胞的基因组中。在此类实施方案中,改良植物是非转基因的。在仅确保内源基因的修饰且在植物基因组中没有引入或维持外源基因的情况下,所得的遗传修饰的作物不包含外源基因,因此基本上可以被认为是非转基因的。

[0855] 在特定的实施方案中,通过DNA病毒(例如,双粒病毒组)或RNA病毒(例如,烟草脆裂病毒组)将多核苷酸递送到细胞中。在特定的实施方案中,引入步骤包括将含有编码CRISPR-Cas蛋白、腺苷脱氨酶和指导RNA的一个或多个多核苷酸序列的T-DNA递送到植物细胞,其中递送经由农杆菌进行。可以将编码AD功能化的CRISPR系统的组分的多核苷酸序列可操作地连接到启动子,诸如组成型启动子(例如,花椰菜花叶病毒35S启动子)或者细胞特异性或诱导型启动子。在特定的实施方案中,通过微粒轰击引入多核苷酸。在特定的实施方案中,所述方法进一步包括在引入步骤之后筛选植物细胞,以确定目标基因的表达是否已经被修饰。在特定的实施方案中,所述方法包括从植物细胞再生植物的步骤。在另外的实施方案中,所述方法包括对植物进行杂交育种,以获得遗传上所需的植物谱系。

[0856] 在上述方法的特定实施方案中,通过靶向突变疾病易感性基因或编码植物防御基因的负调控因子的基因(例如,Mlo基因)来获得抗病作物。在一个特定的实施方案中,通过

靶向取代植物基因(诸如编码乙酰乳酸合酶(ALS)和原卟啉原氧化酶(PPO)的基因)中的特定核苷酸来产生抗除草剂作物。在特定的实施方案中,通过靶向突变编码非生物胁迫抗性的负调控因子的基因来产生耐干旱和耐盐作物,通过靶向突变Waxy基因来产生低直链淀粉谷物,通过靶向突变糊粉层中的主要脂肪酶基因来产生酸败性降低的水稻或其他谷物等。下面列出了编码目标性状的内源基因的更广泛的列表。

[0857] 使用AD功能化的CRISPR系统修饰多倍体植物

[0858] 许多植物都是多倍体的,这意味着它们携带基因组的重复拷贝,小麦中有时多达六倍。利用AD功能化的CRISPR系统的根据本发明的方法可以被"多重化"以影响基因的所有拷贝或者一次靶向数十个基因。例如,在特定的实施方案中,本发明的方法用于同时确保负责抑制疾病防御的不同基因中功能突变的丧失。在特定的实施方案中,本发明的方法用于同时抑制TaMLO-A1、TaMLO-B1和TaMLO-D1核酸序列在小麦植物细胞中表达以及从中再生小麦植物,以确保小麦植物对白粉病具有抗性(还可参见WO2015109752)。

[0859] 赋予农艺性状的示例性基因

[0860] 在特定的实施方案中,本发明包括涉及在内源基因及其调控元件中靶向的A-G和T-C突变的方法,所述内源基因如下所示:

[0861] 1.赋予虫害或病害抗性的基因:

[0862] 植物抗病基因。可以用克隆的抗性基因转化植物以工程化对特定病原体菌株具有抗性的植物。参见例如Jones等人,Science 266:789(1994)(克隆对黄枝孢霉(Cladosporium fulvum)的抗性的番茄Cf-9基因);Martin等人,Science262:1432(1993)(对丁香假单胞杆菌番茄致病变种具有抗性的番茄Pto基因编码蛋白激酶);Mindrinos等人,Cell 78:1089(1994)(拟南芥可以是对丁香假单胞杆菌有抗性的RSP2基因)。可以工程化在病原体感染期间上调或下调的植物基因以抵抗病原体。参见例如Thomazella等人,bioRxiv 064824;doi:https://doi.org/10.1101/064824电子版.2016年7月23日(在病原体感染期间通常上调的SlDMR6-1中具有缺失的番茄植物)。

[0863] 赋予虫害(诸如大豆孢囊线虫)抗性的基因。参见例如PCT申请WO96/30517;PCT申请WO 93/19181。

[0864] 苏云金芽孢杆菌蛋白,参见例如Geiser等人,Gene 48:109(1986)。

[0865] 凝集素,参见例如Van Damme等人,Plant Molec.Biol.24:25(1994。

[0866] 维生素结合蛋白,诸如抗生物素蛋白,参见PCT申请US93/06487,其教导了抗生物素蛋白和抗生物素蛋白同系物作为害虫杀幼虫剂的用途。

[0867] 酶抑制剂,诸如蛋白酶抑制剂或淀粉酶抑制剂。参见例如Abe等人,J.Biol.Chem.262:16793(1987);Huub等人,Plant Molec.Biol.21:985(1993);Sumitani等人,Biosci.Biotech.Biochem.57:1243(1993)以及美国专利号5,494,813。

[0868] 昆虫特异性激素或信息素,诸如蜕皮类固醇或保幼激素、其变体、基于其的模拟物或其拮抗剂或激动剂。参见例如Hammock等人,Nature 344:458(1990)。

[0869] 昆虫特异性肽或神经肽,其在表达时会破坏受影响害虫的生理机能。例如Regan,J.Biol.Chem.269:9(1994)以及Pratt等人,Biochem.Biophys.Res.Comm.163:1243(1989)。还可参见美国专利号5,266,317。

[0870] 自然界中由蛇、黄蜂或任何其他生物产生的昆虫特异性毒液。例如参见Pang等人,

Gene 116:165(1992)。

[0871] 引起单萜、倍半萜、类固醇、异羟肟酸、苯丙烷类衍生物或具有杀虫活性的其他非蛋白质分子超量积累的酶。

[0872] 与具有生物活性的分子的修饰(包括翻译后修饰)有关的酶;例如,糖酵解酶、蛋白水解酶、脂解酶、核酸酶、环化酶、转氨酶、酯酶、水解酶、磷酸酶、激酶、磷酸化酶、聚合酶、弹性蛋白酶、几丁质酶和葡聚糖酶,无论是天然的还是合成的。参见PCT申请WO93/02197;Kramer等人,Insect Biochem.Molec.Biol.23:691(1993)以及Kawalleck等人,Plant Molec.Biol.21:673(1993)。

[0873] 刺激信号传导的分子。例如参见Botella等人,Plant Molec.Biol.24:757(1994)以及Griess等人,Plant Physiol.104:1467(1994)。

[0874] 病毒侵入性蛋白质或由其衍生的复杂毒素。参见Beachy等人,Ann.rev.Phytopathol.28:451(1990)。

[0875] 自然界中由病原体或寄生虫产生的发育阻滞蛋白。参见Lamb等人,Bio/Technology 10:1436(1992)以及Toubart等人,Plant J.2:367(1992)。

[0876] 自然界中由植物产生的发育阻滞蛋白。例如Logemann等人,Bio/Technology 10:305(1992)。

[0877] 在植物中,病原体通常具有宿主特异性。例如,一些镰刀菌种将引起番茄枯萎,但仅侵害番茄,而其他镰刀菌种仅侵害小麦。植物具有抵抗大多数病原体的现有和诱导防御能力。跨植物世代的突变和重组事件会导致产生易感性的遗传变异,特别是当病原体的繁殖频率高于植物时。在植物中可能存在非宿主抗性,例如宿主和病原体不相容,或者可能存在对病原体所有种族的部分抗性,通常由许多基因控制,并且/或者还可能存在对病原体某些种族而不是其他种族的完全抗性。这种抗性通常由几个基因控制。使用AD功能化的CRISPR系统的方法和组分,现在存在一种新工具来诱导预期的特定突变。因此,人们可以分析抗性基因来源的基因组,并且在具有所需特征或性状的植物中,使用AD功能化的CRISPR系统的方法和组分来诱导抗性基因的升高。本发明系统可以比以前的诱变剂更精确地诱导,因此可以加速和改善植物育种程序。

[0878] 2.与植物病害有关的基因,所述病害诸如WO 2013046247中列出的病害:

[0879] 水稻病害:稻瘟病菌(Magnaporthe grisea)、宫部旋孢腔菌(Cochliobolus miyabeanus)、立枯丝核菌(Rhizoctonia solani)、藤仓赤霉菌(Gibberella fujikuroi);小麦病害:禾白粉病(Erysiphe graminis)、禾谷镰刀菌(Fusarium graminearum)、燕麦镰刀菌(F.avenaceum)、黄色镰刀菌(F.culmorum)、雪腐微座孢(Microdochium nivale)、条形柄锈菌(Puccinia striiformis)、禾柄锈菌(P.graminis)、隐匿柄锈菌(P.recondita)、雪腐镰刀病菌(Micronectriella nivale)、核瑚菌属某种(Typhula sp.)、小麦散黑粉菌(Ustilago tritici)、小麦网腥黑粉菌(Tilletia caries)、小麦基腐病菌(Pseudocercosporella herpotrichoides)、禾生球腔菌(Mycosphaerella graminicola)、颖枯壳多孢(Stagonospora nodorum)、偃麦草核腔菌(Pyrenophora tritici-repentis);大麦病害:禾白粉病(Erysiphe graminis)、禾谷镰刀菌(Fusarium graminearum)、燕麦镰刀菌(F.avenaceum)、黄色镰刀菌(F.culmorum)、雪腐微座孢(Microdochium nivale)、条形柄锈菌(Puccinia striiformis)、禾柄锈菌(P.graminis)、大麦柄锈菌(P.hordei)、大麦散

黑粉菌(Ustilago nuda)、大麦云纹病菌(Rhynchosporium secalis)、大麦网斑病(Pyrenophora teres)、禾旋孢腔菌(Cochliobolus sativus)、大麦条纹病菌(Pyrenophora graminea)、立枯丝核菌(Rhizoctonia solani);玉米病害:玉蜀黍黑粉菌(Ustilago maydis)、异旋孢腔菌(Cochliobolus heterostrophus)、高粱胶尾孢(Gloeocercospora sorghi)、多堆柄锈菌(Puccinia polysora)、玉蜀黍尾孢菌(Cercospora zeae-maydis)、立枯丝核菌(Rhizoctonia solani);

[0880] 柑橘病害:柑桔砂皮病菌(Diaporthe citri)、疮痂病(Elsinoe fawcetti)、指状青霉(Penicillium digitatum)、意大利青霉(P.italicum)、寄生疫霉(Phytophthora parasitica)、柑桔褐腐疫霉(Phytophthora citrophthora);苹果病害:苹果链核盘菌(Monilinia mali)、苹果树腐烂病菌(Valsa ceratosperma)、苹果白粉病菌(Podosphaera leucotricha)、苹果斑点落叶病菌(Alternaria alternata apple pathotype)、苹果黑星病菌(Venturia inaequalis)、炭疽病菌(Colletotrichum acutatum)、晚疫病菌(Phytophtora cactorum);

[0881] 梨病害:梨黑星菌(Venturia nashicola)、梨黑星病(V pirina)、梨黑斑病菌(Alternaria alternata Japanese pear pathotype)、梨胶锈菌(Gymnosporangium haraeanum)、恶疫霉(Phytophtora cactorum);

[0882] 桃病害:褐腐菌(Monilinia fructicola)、嗜果枝抱霉(Cladosporiumcarpophilum)、拟茎点霉属某种(Phomopsis sp.);

[0883] 葡萄病害:葡萄黑痘病菌(Elsinoe ampelina)、围小丛壳菌(Glomerella cingulata)、葡萄白粉菌(Uninula necator)、层锈菌属(Phakopsora ampelopsidis)、葡萄球座菌(Guignardia bidwellii)、霜霉病菌(Plasmopara viticola);

[0884] 柿树病害:柿盘抱子菌(Gloeosporium kaki)、柿角斑病(Cercospora kaki)、柿叶球腔菌(Mycosphaerela nawae);

[0885] 葫芦病害:瓜类炭疽菌(Colletotrichum lagenarium)、瓜类白粉病菌(Sphaerotheca fuliginea)、蔓枯病菌(Mycosphaerella melonis)、尖抱德刀菌(Fusarium oxysporum)、黄瓜霜霉病菌(Pseudoperonospora cubensis)、疫霉菌属某种(Phytophthora sp.)、腐霉菌属某种(Pythium sp.);

[0886] 番茄病害:茄链格孢菌(Alternaria solani)、番茄叶霉病菌(Cladosporium fulvum)、马铃薯晚疫病菌(Phytophthora infestans);丁香假单胞菌番茄致病变种(Pseudomonas syringae pv.Tomato);辣椒疫霉菌(Phytophthora capsici);黄单胞菌属(Xanthomonas)

[0887] 茄子病害:褐纹病菌(Phomopsis vexans)、二孢白粉菌(Erysiphe cichoracearum);十字花科蔬菜病害:萝卜链格孢菌(Alternaria japonica)、白菜白斑病菌(Cercosporella brassicae)、根肿菌(Plasmodiophora brassicae)、寄生霜霉(Peronospora parasitica);

[0888] 大葱病害:葱柄锈菌(Puccinia allii)、大葱霜霉(Peronospora destructor);

[0889] 大豆病害:大豆紫斑病菌(Cercospora kikuchii)、大豆痂囊腔菌(Elsinoe glycines)、菜豆间座壳大豆变种(Diaporthe phaseolorum var.sojae)、大豆褐斑病菌(Septoria glycines)、大豆灰斑病菌(Cercospora sojina)、豆薯层锈菌(Phakopsora

pachyrhizi)、大豆疫霉菌(Phytophthora sojae)、立枯丝核菌(Rhizoctonia solani)、棒抱叶斑病(Corynespora casiicola)、核盘菌(Sclerotinia sclerotiorum)；

[0890] 菜豆病害:炭疽病菌(Colletrichum lindemthianum)；

[0891] 花生病害:球座尾孢(Cercospora personata)、落花生尾孢(Cercospora arachidicola)、齐整小核菌(Sclerotium rolfsii)；

[0892] 豌豆病害:豌豆白粉菌(Erysiphe pisi)；

[0893] 马铃薯病害:茄链格孢菌(Alternaria solani)、马铃薯晚疫病菌(Phytophthora infestans)、马铃薯疫霉绯腐病菌(Phytophthora erythroseptica)、马铃薯粉状疮痴病菌(Spongospora subterranean,f.sp.Subterranean)；

[0894] 草莓病害:白粉菌(Sphaerotheca humuli)、围小丛壳菌(Glomerella cingulata)；

[0895] 茶树病害:网状外担菌(Exobasidium reticulatum)、茶痂囊腔菌(Elsinoe leucospila)、拟盘多毛孢属某种(Pestalotiopsis sp.)、炭疽病菌(Colletotrichum theae-sinensis)；

[0896] 烟草病害:烟草赤星病菌(Alternaria longipes)、二孢白粉菌(Erysiphe cichoracearum)、烟草炭疽病菌(Colletotrichum tabacum)、烟草霜霉菌(Peronospora tabacina)、烟草疫病菌(Phytophthora nicotianae)；

[0897] 油菜籽病害:核盘菌(Sclerotinia sclerotiorum)、立枯丝核菌(Rhizoctonia solani)；

[0898] 棉花病害:立枯丝核菌(Rhizoctonia solani)；

[0899] 甜菜病害:甜菜生尾孢菌(Cercospora beticola)、稻纹枯病原菌(Thanatephorus cucumeris)、稻纹枯病原菌(Thanatephorus cucumeris)、丝囊霉(Aphanomyces cochlioides)；

[0900] 蔷薇病害:蔷薇双壳菌(Diplocarpon rosae)、蔷薇单丝壳菌(Sphaerotheca pannosa)、蔷薇霜霉菌(Peronospora sparsa)；

[0901] 菊属和菊科病害:莴苣盘梗霉(Bremia lactuca)、菊壳针孢菌(Septoria chrysanthemi-indici)、堀柄锈菌(Puccinia horiana)；

[0902] 各种植物病害:坪草腐霉枯萎病菌(Pythium aphanidermatum)、德巴利氏腐霉(Pythium debarianum)、禾生腐霉(Pythium graminicola)、畸雌腐霉(Pythiumirregulare)、终极腐霉(Pythium ultimum)、灰葡萄孢菌(Botrytiscinerea)、核盘菌(Sclerotinia sclerotiorum)；

[0903] 萝卜病害:甘蓝链格孢菌(Alternaria brassicicola)；

[0904] 结缕草病害:银斑病病菌(Sclerotinia homeocarpa)、立枯丝核菌(Rhizoctoniasolani)；

[0905] 香蕉病害:香蕉黑条叶斑病菌(Mycosphaerella fijiensis)、香蕉黄条叶斑病菌(Mycosphaerella musicola)；

[0906] 向日葵病害:霍尔斯单轴霉(Plasmopara halstedii)；

[0907] 由以下各项引起的种子病害或各种植物生长初期的病害:曲霉属某些种(Aspergillus spp.)、青霉属某些种(Penicillium spp.)、镰刀菌属某些种(Fusarium

spp.)、赤霉属某些种(Gibberella spp.)、木霉属某些种(Tricoderma spp.)、珠霉属某些种(Thielaviopsis spp.)、根霉属某些种(Rhizopus spp.)、毛霉属某些种(Mucor spp.)、伏革菌属某些种(Corticium spp.)、茎点霉属某些种(Rhoma spp.)、丝核菌属某些种(Rhizoctonia spp.)、色二孢属某些种(Diplodia spp.)等;

[0908] 由以下各项介导的各种植物的病毒病害:杆菌属某些种(Polymixa spp.)、油壶菌属某些种(Olpidium spp.)等。

[0909] 3.赋予除草剂抗性的基因的实例:

[0910] 对抑制生长点或分生组织的除草剂(诸如咪唑啉酮或磺酰脲,分别例如Lee等人,EMBO J.7:1241(1988)以及Miki等人,Theor.Appl.Genet.80:449(1990))的抗性。

[0911] 草甘膦耐受性(例如,分别由突变体5-烯醇式丙酮酰莽草酸-3-磷酸合酶(EPSP)基因、aroA基因和草甘膦乙酰基转移酶(GAT)基因赋予的抗性),或者对其他膦酰基化合物的抗性(诸如通过草铵膦(来自链霉菌属物种(包括吸水链霉菌(Streptomyces hygroscopicus)和绿色产色链霉菌(Streptomyces viridichromogenes))的膦丝菌素乙酰基转移酶(PAT)基因),以及对吡啶氧基或苯氧基丙酸和环己酮的抗性(通过编码ACCase抑制剂的基因)。参见例如美国专利号4,940,835、美国专利号6,248,876、美国专利号4,769,061、欧洲专利号0 333 033以及美国专利号4,975,374。还可参见欧洲专利号0242246、DeGreef等人,Bio/Technology 7:61(1989)、Marshall等人,Theor.Appl.Genet.83:435(1992)、授予Castle等人的WO 2005012515以及WO2005107437。

[0912] 对抑制光合作用的除草剂(诸如三嗪(psbA和gs+基因)或苄腈(腈水解酶基因)以及谷胱甘肽S-转移酶,Przibila等人,Plant Cell 3:169(1991)、美国专利号4,810,648以及Hayes等人,Biochem.J.285:173(1992))的抗性。

[0913] 编码使除草剂或突变体谷氨酰胺合酶(对抑制具有抗性)解毒的酶的基因,例如美国专利申请序列号11/760,602。或者解毒酶是编码膦丝菌素乙酰基转移酶的酶(诸如链霉菌属物种的bar或pat蛋白)。膦丝菌素乙酰基转移酶例如在以下文献中描述:美国专利号5,561,236、5,648,477、5,646,024、5,273,894、5,637,489、5,276,268、5,739,082、5,908,810和7,112,665。

[0914] 羟基苯基丙酮酸双加氧酶(HPPD)抑制剂、天然存在的HPPD抗性酶或者编码突变或嵌合HPPD酶的基因,如WO 96/38567、WO 99/24585和WO 99/24586、WO 2009/144079、WO 2002/046387或美国专利号6,768,044中所述。

[0915] 4.与非生物胁迫耐受性有关的基因的实例:

[0916] 能够降低植物细胞或植物中聚(ADP-核糖)聚合酶(PARP)基因的表达和/或活性的转基因,如WO 00/04173或WO/2006/045633中所述。

[0917] 能够降低植物或植物细胞的PARG编码基因的表达和/或活性的转基因,如例如WO 2004/090140中所述。

[0918] 编码烟酰胺腺嘌呤二核苷酸补救合成途径的植物功能酶的转基因,包括烟酰胺酶、烟酸磷酸核糖基转移酶、烟酸单核苷酸腺嘌呤基转移酶、烟酰胺腺嘌呤二核苷酸合成酶或烟酰胺磷酸核糖基转移酶,如例如EP 04077624.7、WO 2006/133827、PCT/EP07/002,433、EP 1999263或WO 2007/107326中所述。

[0919] 与碳水化合物生物合成有关的酶包括例如以下文献中所述的酶:EP 0571427、WO

95/04826、EP 0719338、WO 96/15248、WO 96/19581、WO 96/27674、WO 97/11188、WO 97/26362、WO 97/32985、WO 97/42328、WO 97/44472、WO 97/45545、WO 98/27212、WO 98/40503、WO99/58688、WO 99/58690、WO 99/58654、WO 00/08184、WO 00/08185、WO 00/08175、WO 00/28052、WO 00/77229、WO 01/12782、WO 01/12826、WO 02/101059、WO 03/071860、WO 2004/056999、WO 2005/030942、WO 2005/030941、WO 2005/095632、WO 2005/095617、WO 2005/095619、WO 2005/095618、WO 2005/123927、WO 2006/018319、WO 2006/103107、WO 2006/108702、WO 2007/009823、WO 00/22140、WO 2006/063862、WO 2006/072603、WO 02/034923、EP 06090134.5、EP 06090228.5、EP 06090227.7、EP 07090007.1、EP 07090009.7、WO 01/14569、WO 02/79410、WO 03/33540、WO 2004/078983、WO 01/19975、WO 95/26407、WO 96/34968、WO 98/20145、WO 99/12950、WO 99/66050、WO 99/53072、美国专利号6,734,341、WO 00/11192、WO 98/22604、WO 98/32326、WO 01/98509、WO 01/98509、WO 2005/002359、美国专利号5,824,790、美国专利号6,013,861、WO 94/04693、WO 94/09144、WO 94/11520、WO 95/35026或WO 97/20936;或者如以下文献中公开的与产生多聚果糖、尤其是菊粉和左聚糖有关的酶:EP 0663956、WO 96/01904、WO 96/21023、WO 98/39460和WO 99/24593;如以下文献中公开的与产生α-1,4-葡聚糖有关的酶:WO 95/31553、US 2002031826、美国专利号6,284,479、美国专利号5,712,107、WO 97/47806、WO 97/47807、WO 97/47808和WO 00/14249;如WO 00/73422中公开的与产生α-1,6支链α-1,4-葡聚糖有关的酶;如例如以下文献中公开的与产生交替糖有关的酶:WO 00/47727、WO 00/73422、EP 06077301.7、美国专利号5,908,975和EP 0728213;如例如以下文献中公开的与产生透明质酸有关的酶:WO 2006/032538、WO 2007/039314、WO 2007/039315、WO 2007/039316、JP 2006304779和WO 2005/012529。

[0920] 提高抗旱性的基因。例如,WO 2013122472公开了功能泛素蛋白连接酶(UPL)蛋白(更具体为UPL3)的缺乏或水平降低会使得所述植物对水的需求减少或抗旱性有所提高。抗旱性增加的转基因植物的其他实例在例如US 2009/0144850、US 2007/0266453和WO 2002/083911中公开。US2009/0144850描述了由于DR02核酸的表达改变而显示出耐旱表型的植物。US 2007/0266453描述了由于DR03核酸的表达改变而显示出耐旱表型的植物,并且WO 2002/08391 1描述了由于在保卫细胞中表达的ABC转运蛋白的活性降低而对干旱胁迫的耐受性增强的植物。另一个实例是Kasuga及其共同作者(1999)的研究成果,他们描述了在转基因植物中过量表达编码DREB1 A的cDNA激活了许多耐胁迫基因在正常生长条件下表达并且提高了对干旱、盐负荷和冷冻的耐受性。然而,DREB1A的表达也导致了正常生长条件下严重的生长迟缓(Kasuga(1999)Nat Biotechnol 17(3)287-291)。

[0921] 在另外的特定实施方案中,可以通过影响特定的植物性状来改良作物植物。例如,通过开发抗农药植物、提高植物的抗病性、提高植物的昆虫和线虫抗性、提高植物对寄生杂草的抗性、提高植物的耐旱性、提高植物的营养价值、提高植物的胁迫耐受性、避免自花授粉、植物饲料消化率生物量、谷物产量等。下文提供了几个具体的非限制性实例。

[0922] 除了单个基因的靶向突变外,AD功能化的CRISPR系统还可以被设计成允许多个基因的靶向突变、染色体片段的缺失、转基因的位点特异性整合、体内定点诱变以及植物中精确的基因替换或等位基因交换。因此,本文所述的方法在基因发现和验证、突变育种和顺生育种以及杂交育种中具有广泛的应用。这些应用有利于产生具有各种改良农艺性状(诸如除草剂抗性、抗病性、非生物胁迫耐受性、高产和优质)的新一代遗传修饰的作物。

[0923] 使用AD功能化的CRISPR系统创建雄性不育植物

[0924] 与近交植物相比，杂种植物通常具有有利的农艺性状。然而，对于自花授粉的植物，产生杂种可能具有挑战性。在不同的植物类型中，已经鉴定出对植物育性、更具体为雄性育性至关重要的基因。例如，在玉米中，鉴定出至少两个对育性至关重要的基因(Amitabh Mohanty International Conference on New Plant Breeding Molecular Technologies Technology Development And Regulation,2014年10月9日至10日,Jaipur,India;Svitashev等人,Plant Physiol.2015年10月;169(2):931-45;Djukanovic等人,Plant J.2013年12月;76(5):888-99)。本文提供的方法和系统可以用于靶向雄性育性所需的基因，以产生雄性不育植物，所述雄性不育植物可以容易杂交以产生杂种。在特定的实施方案中，本文提供的AD功能化的CRISPR系统用于细胞色素P450样基因(MS26)或大范围核酸酶基因(MS45)的靶向诱变，从而赋予玉米植物雄性不育性。经过如此遗传改变的玉米植物可以用于杂交育种程序。

[0925] 增加植物的生育期

[0926] 在特定的实施方案中，本文提供的方法和系统用于延长植物(诸如水稻植物)的生育期。例如，可以靶向水稻生育期基因(诸如Ehd3)以在该基因中产生突变，并且可以选择小植株以延长再生植物生育期(如CN 104004782中所述)

[0927] 使用AD功能化的CRISPR系统在目标作物中产生遗传变异

[0928] 作物植物中野生种质的可用性和遗传变异是作物改良程序的关键，但作物种质的可用多样性有限。本发明设想了用于在目标种质中产生多种遗传变异的方法。在AD功能化的CRISPR系统的该应用中，提供了靶向植物基因组中不同位置的指导RNA的文库，并将所述文库与CRISPR-Cas蛋白和腺苷脱氨酶一起引入植物细胞中。以这种方式，可以产生基因组规模的点突变和基因敲除的集合。在特定的实施方案中，所述方法包括从由此获得的细胞产生植物部分或植物，并筛选所述细胞的目标性状。靶基因可以包括编码区和非编码区。在特定的实施方案中，性状是胁迫耐受性，并且所述方法是用于产生胁迫耐受性作物品种的方法。

[0929] 使用AD功能化的CRISPR影响果实催熟

[0930] 催熟是水果和蔬菜成熟过程中的正常阶段。该阶段开始后仅几天就使水果或蔬菜不可食用。该过程给农民和消费者都造成了严重损失。在特定的实施方案中，本发明的方法用于减少乙烯产生。通过确保以下一项或多项来确保这一点:a.ACC合酶基因表达的抑制。ACC(1-氨基环丙烷-1-羧酸)合酶是负责将S-腺苷甲硫氨酸(SAM)转化为ACC的酶;乙烯生物合成的第二步到最后一步。当将合成酶基因的反义(“镜像”)或截短副本插入植物基因组时，酶表达受到阻碍;b.ACC脱氨酶基因的插入。编码该酶的基因是从绿针假单胞菌(Pseudomonas chlororaphis)(一种常见的非致病性土壤细菌)获得的。它将ACC转化为不同化合物，从而减少可用于产生乙烯的ACC的量;c.SAM水解酶基因的插入。这种方法类似于ACC脱氨酶，其中当其前体代谢物的量减少时，乙烯产生受到阻碍。在这种情况下，SAM被转化为高丝氨酸。编码该酶的基因从大肠埃希氏菌T3噬菌体获得;以及和d.ACC氧化酶基因表达的抑制。ACC氧化酶是催化ACC氧化为乙烯(这是乙烯生物合成途径的最后一步)的酶。使用本文所述的方法，ACC氧化酶基因的下调导致乙烯产生被抑制，从而延迟果实催熟。在特定的实施方案中，除上述修饰之外或作为上述修饰的替代，本文所述的方法用于修饰乙烯

受体，从而干扰果实获得的乙烯信号。在特定的实施方案中，修饰、更具体抑制编码乙烯结合蛋白的ETR1基因的表达。在特定的实施方案中，除上述修饰之外或作为上述修饰的替代，本文所述的方法用于修饰编码聚半乳糖醛酸酶(PG)(其是负责果胶分解的酶，果胶是维持植物细胞壁完整性的物质)的基因的表达。果胶分解发生在催熟过程的开始，导致果实软化。因此，在特定的实施方案中，本文描述的方法用于在PG基因中引入突变或抑制PG基因的活化，以减少产生的PG酶的量，从而延迟果胶降解。

[0931] 因此，在特定的实施方案中，所述方法包括使用AD功能化的CRISPR系统来确保诸如上述的植物细胞基因组的一种或多种修饰，并从中再生植物。在特定的实施方案中，植物是番茄植物。

[0932] 延长植物的贮藏寿命

[0933] 在特定的实施方案中，本发明的方法用于修饰与产生影响植物或植物部分的贮藏寿命的化合物有关的基因。更具体地，所述修饰在防止还原糖在马铃薯块茎中积累的基因中。经过高温处理，这些还原糖会与游离氨基酸反应，从而产生褐色的苦味产物，并增加潜在的致癌物质丙烯酰胺的水平。在特定的实施方案中，本文提供的方法用于减少或抑制液泡转化酶基因(VInv)的表达，所述液泡转化酶基因编码将蔗糖分解为葡萄糖和果糖的蛋白质(Clasen等人，DOI:10.1111/pbi.12370)。

[0934] 使用AD功能化的CRISPR系统确保价值增加的性状

[0935] 在特定的实施方案中，AD功能化的CRISPR系统用于产生营养改良的农作物。在特定的实施方案中，本文提供的方法适于产生“功能性食品”(即可提供超出其所含传统营养的健康益处的改性食品或食品成分)和/或“营养品”(即可被认为是食品或食物的一部分并提供健康益处(包括预防和治疗疾病)的物质)。在特定的实施方案中，营养品可用于预防和/或治疗癌症、糖尿病、心血管疾病和高血压中的一种或多种。

[0936] 营养改良作物的实例包括(Newell-McGloughlin,Plant Physiology,2008年7月，第147卷，第939-953页)：

[0937] 诸如已经针对以下作物描述了改良的蛋白质量、含量和/或氨基酸组成：百喜草(Luciani等人，2005,Florida Genetics Conference Poster)、油菜(Roesler等人，1997,Plant Physiol 113 75-81)、玉米(Cromwell等人，1967,1969J Anim Sci 26 1325-1331;O'Quin等人，2000J Anim Sci 78 2144-2149;Yang等人，2002,Transgenic Res 11 11-20;Young等人，2004,Plant J 38 910-922)、马铃薯(Yu J和Ao,1997Acta Bot Sin 39 329-334;Chakraborty等人，2000,Proc Natl Acad Sci USA 97 3724-3729;Li等人，2001)Chin Sci Bull 46 482-484、大米(Katsube等人，1999,Plant Physiol 120 1063-1074)、大豆(Dinkins等人，2001,Rapp 2002,In Vitro Cell Dev Biol Plant 37 742-747)、甘薯(Egnin和Prakash 1997,In Vitro Cell Dev Biol 33 52A)。

[0938] 诸如已经针对以下作物描述了必需氨基酸含量：油菜(Falco等人，1995,Bio/Technology 13 577-582)、羽扇豆(White等人，2001,J Sci Food Agric 81147-154)、玉米(Lai和Messing,2002,Agbios 2008GM crop database(2008年3月11日))、马铃薯(Zeh等人，2001,Plant Physiol 127 792-802)、高粱(Zhao等人，2003,Kluwer Academic Publishers,Dordrecht,The Netherlands,第413-416页)、大豆(Falco等人，1995Bio/Technology 13 577-582;Galili等人，2002Crit Rev Plant Sci 21 167-204)。

[0939] 诸如已经针对如下作物描述了油和脂肪酸:油菜(Dehesh等人,(1996)Plant J 9 167-172[电子版];Del Vecchio(1996)INFORM International News on Fats,Oils and Related Materials 7 230-243;Roesler等人,(1997)Plant Physiol113 75-81[PMC免费文章][电子版];Froman和Ursin(2002,2003)Abstracts of Papers of the American Chemical Society 223U35;James等人,(2003)Am J Clin Nutr 77 1140-1145[电子版];Agbios(2008,如上))、棉花(Chapman等人,(2001).J Am Oil Chem Soc 78 941-947;Liu等人,(2002)J Am Coll Nutr 21 205S-211S[电子版];O'Neill(2007)Australian Life Scientist.http://www.biotechnews.com.au/index.php/id;866694817;fp;4;fpid;2(2008年6月17日))、亚麻籽(Abbadi等人,2004,Plant Cell 16:2734-2748)、玉米(Young等人,2004,Plant J 38 910-922)、油棕(Jalani等人,1997,J Am Oil Chem Soc 74 1451-1455;Parveez,2003,AgBiotechNet 113 1-8)、水稻(Anai等人,2003,Plant Cell Rep 21 988-992)、大豆(Reddy和Thomas,1996,Nat Biotechnol 14 639-642;Kinney和Kwolton,1998,Blackie Academic and Professional,London,第193-213页)、向日葵(Arcadia,Biosciences 2008)

[0940] 诸如已经针对如下作物描述了碳水化合物(诸如果聚糖):菊苣(Smeekens(1997)Trends Plant Sci 2 286-287;Sprenger等人,(1997)FEBS Lett 400 355-358;Sévenier等人,(1998)Nat Biotechnol 16 843-846)、玉米(Caimi等人,(1996)Plant Physiol 110 355-363)、马铃薯(Hellwege等人,1997Plant J 121057-1065)、甜菜(Smeekens等人,1997,如上),诸如已经针对马铃薯描述了菊粉(Hellewege等人,2000,Proc Natl Acad Sci USA 97 8699-8704),诸如已经针对水稻描述了淀粉(Schwall等人,(2000)Nat Biotechnol 18 551-554;Chiang等人,(2005)Mol Breed 15 125-143),

[0941] 诸如已经针对如下作物描述了维生素和类胡萝卜素:油菜(Shintani和DellaPenna(1998)Science 282 2098-2100)、玉米(Rocheford等人,(2002).J Am Coll Nutr 21 191S-198S;Cahoon等人,(2003)Nat Biotechnol 21 1082-1087;Chen等人,(2003)Proc Natl Acad Sci USA 100 3525-3530)、芥菜籽(Shewmaker等人,(1999)Plant J 20 401-412)、马铃薯(Ducreux等人,2005,J Exp Bot 56 81-89)、水稻(Ye等人,(2000)Science 287 303-305)、草莓(Agius等人,(2003),Nat Biotechnol 21 177-181)、番茄(Rosati等人,(2000)Plant J 24 413-419;Fraser等人,(2001)J Sci Food Agric 81 822-827;Mehta等人,(2002)Nat Biotechnol 20 613-618;Díaz de la Garza等人,(2004)Proc Natl Acad Sci USA 101 13720-13725;Enfissi等人,(2005)Plant Biotechnol J 3 17-27,DellaPenna(2007)Proc Natl Acad Sci USA 104 3675-3676)。

[0942] 诸如已经针对如下作物描述了功能性次级代谢产物:苹果(二苯乙烯类,Szankowski等人,(2003)Plant Cell Rep 22:141-149)、苜蓿(白藜芦醇,Hipskind和Paiva(2000)Mol Plant Microbe Interact 13 551-562)、猕猴桃(白藜芦醇,Kobayashi等人,(2000)Plant Cell Rep 19 904-910)、玉米和大豆(类黄酮,Yu等人,(2000)Plant Physiol 124 781-794)、马铃薯(花青素和生物碱配糖,Lukaszewicz等人,(2004)J Agric Food Chem 52 1526-1533)、水稻(类黄酮和白藜芦醇,Stark-Lorenzen等人,(1997)Plant Cell Rep 16 668-673;Shin等人,(2006)Plant Biotechnol J 4 303-315)、番茄(+白藜芦醇、绿原酸、类黄酮、二苯乙烯;Rosati等人,(2000)如上;Muir等人,(2001)Nature 19 470-474;

Niggeweg等人,(2004)Nat Biotechnol 22 746-754;Giovinazzo等人,(2005)Plant Biotechnol J 3 57-69)、小麦(咖啡酸和阿魏酸、白藜芦醇;United Press International (2002));并且

[0943] 诸如已经针对如下作物描述了矿物质有效性:苜蓿(植酸酶,Austin-Phillips等人,(1999)http://www.molecularfarming.com/nonmedical.html)、生菜(铁,Goto等人,(2000)Theor Appl Genet 100 658-664)、水稻(铁,Lucca等人,(2002)J Am Coll Nutr 21 184S-190S)、玉米、大豆和小麦(植酸酶,Drakakaki等人,(2005)Plant Mol Biol 59 869-880;Denbow等人,(1998)Poult Sci 77 878-881;Brinch-Pedersen等人,(2000)Mol Breed 6 195-206)。

[0944] 在特定的实施方案中,增值性状与植物中存在的化合物的设想健康益处有关。例如,在特定的实施方案中,通过应用本发明的方法获得价值增加的作物,以确保修饰或诱导/增加一种或多种以下化合物的合成:

[0945] • 胡萝卜中存在的类胡萝卜素,诸如α-胡萝卜素,其可中和可能损害细胞的自由基;或各种水果和蔬菜中存在的β-胡萝卜素,其可中和自由基;

[0946] • 绿色蔬菜中存在的叶黄素,其有助于维持健康的视力;

[0947] • 番茄和番茄制品中存在的番茄红素,据信其可降低患前列腺癌的风险;

[0948] • 柑橘和玉米中存在的玉米黄质,其有助于维持健康的视力

[0949] • 麦麸中存在的膳食纤维,诸如不溶性纤维,其可降低乳腺癌和/或结肠癌的风险;以及燕麦中存在的β-葡聚糖、车前草和全谷类谷物中存在的可溶性纤维,其可降低心血管疾病(CVD)的风险;

[0950] • 脂肪酸,诸如ω-3脂肪酸,其可降低CVD的风险并改善心理和视觉功能;共轭亚油酸,其可改善身体组成,可降低某些癌症的风险;以及GLA,其可降低癌症和CVD的炎症风险,可改善身体组成;

[0951] • 小麦中存在的类黄酮,诸如羟基肉桂酸酯,其具有类似抗氧化剂的活性,可降低变性疾病的风险;水果和蔬菜中存在的黄酮醇、儿茶素和丹宁酸,其可中和自由基并且可降低癌症的风险;

[0952] • 十字花科蔬菜(西兰花、羽衣甘蓝)、辣根中存在的芥子油苷、吲哚、异硫氰酸酯,诸如萝卜硫素,其可中和自由基,可降低癌症的风险;

[0953] • 葡萄中存在的酚类物质,诸如二苯乙烯类,其可降低变性疾病、心脏病和癌症的风险,可具有延年益寿的作用;以及蔬菜和柑桔中存在的咖啡酸和阿魏酸,其具有类似抗氧化剂的活性,可降低变性疾病、心脏病和眼部疾病的风险;以及可可中存在的表儿茶素,其具有类似抗氧化剂的活性,可降低变性疾病和心脏病的风险;

[0954] • 玉米、大豆、小麦和木油中存在的植物甾烷醇/固醇,其可通过降低血液胆固醇水平来降低冠心病的风险;

[0955] • 洋姜、葱、洋葱粉中存在的果聚糖、菊粉、低聚果糖,其可改善肠胃健康;

[0956] • 大豆中存在皂苷,其可降低LDL胆固醇;

[0957] • 大豆中存在的大豆蛋白,其可降低心脏病的风险;

[0958] • 大豆中存在的植物雌激素,诸如异黄酮,其可减少更年期症状(诸如潮热),可减少骨质疏松症和CVD;亚麻、黑麦和蔬菜中存在的木脂素,其可预防心脏病和一些癌症,可降

低LDL胆固醇、总胆固醇；

[0959] • 硫化物和硫醇，诸如洋葱、大蒜、橄榄、韭菜和scallon中存在的二烯丙基硫化物，以及十字花科蔬菜中存在的烯丙基甲基三硫化物、二硫代硫酮，其可降低LDL胆固醇，有助于维持健康的免疫系统；以及

[0960] • 蔓越莓、可可中存在的丹宁酸，诸如原花色素，其可改善尿道健康，可降低CVD和高血压的风险。

[0961] 另外，本发明的方法还设想了改变蛋白质/淀粉功能性、保质期、味道/美学、纤维质量以及过敏原、抗营养素和毒素减少性状。

[0962] 因此，本发明包括用于产生具有增加的营养价值的植物的方法，所述方法包括使用本文所述的AD功能化的CRISPR系统将编码与产生增加的营养价值的组分有关的酶的基因引入植物细胞中，并从所述植物细胞再生植物，所述植物的特征在于增加的营养价值的所述组分的表达增加。在特定的实施方案中，AD功能化的CRISPR系统用于间接修饰这些化合物的内源合成，例如通过修饰一种或多种控制该化合物代谢的转录因子。上文描述了使用AD功能化的CRISPR系统将目标基因引入植物细胞和/或修饰内源基因的方法。

[0963] 已被修饰以赋予价值增加的性状的植物中的修饰的一些具体实例是：具有修饰的脂肪酸代谢的植物，例如通过转化具有硬脂基-ACP去饱和酶的反义基因的植物以增加植物的硬脂酸含量。参见Knultzon等人，Proc.Natl.Acad.Sci.U.S.A.89:2624(1992)。另一个实例涉及降低肌醇六磷酸含量，例如通过克隆然后再引入与单个等位基因相关的DNA，该DNA可能与以低植酸水平为特征的玉米突变体有关。参见Raboy等人，Maydica 35:383(1990)。

[0964] 类似地，玉米(Zea mays)Tfs C1和R(在强启动子的控制下调控玉米糊粉层中类黄酮的产生)的表达导致拟南芥(Arabidopsis thaliana)中花青素的积累率高，推测通过激活整个途径(Bruce等人，2000，Plant Cell 12:65-80)。DellaPenna(Welsch等人，2007Annu Rev Plant Biol 57:711-738)发现，Tf RAP2.2及其相互作用的伴侣SINAT2增加了拟南芥叶片中的胡萝卜素形成作用。表达Tf Dof1诱导了编码碳骨架产生酶的基因的上调、氨基酸含量的显著增加以及转基因拟南芥中Glc水平的降低(Yanagisawa，2004Plant Cell Physiol 45:386-391)，并且DOF Tf AtDof1.1(OBP2)上调了拟南芥中芥子油苷生物合成途径的所有步骤(Skirycz等人，2006Plant J 47:10-24)。

[0965] 减少植物中的过敏原

[0966] 在特定的实施方案中，本文提供的方法用于产生过敏原水平降低的植物，从而使它们对消费者而言更安全。在特定的实施方案中，所述方法包括修饰负责植物过敏原产生的一种或多种基因的表达。例如，在特定的实施方案中，所述方法包括下调Lol p5基因在植物细胞(诸如黑麦草植物细胞)中的表达，并从中再生植物以降低所述植物的花粉的致敏性(Bhalla等人，1999，Proc.Natl.Acad.Sci.USA，第96卷：11676-11680)。

[0967] 花生过敏和对豆类过敏通常是一个真正严重的健康问题。本发明的AD功能化的CRISPR系统可以用于鉴定编码此类豆类植物的致敏蛋白的基因，然后突变这些基因。不限于此类基因和蛋白质，Nicolaou等人鉴定了花生、大豆、扁豆、豌豆、羽扇豆、青豆和绿豆中的致敏蛋白。参见Nicolaou等人，Current Opinion in Allergy and Clinical Immunology 2011;11(3):222)。

[0968] 对目标内源基因进行筛选的方法

[0969] 本文提供的方法进一步允许鉴定编码与产生增加的营养价值的组分有关的酶的基因,或在种、门和植物界中通常影响目标农艺性状的基因。通过使用本文所述的AD功能化的CRISPR系统选择性地靶向植物中编码代谢途径的酶的基因,可以鉴定出负责植物的某些营养方面的基因。类似地,通过选择性地靶向可能影响所需农艺性状的基因,可以鉴定出相关基因。因此,本发明包括对编码与产生具有特定营养价值和/或农艺性状的化合物有关的酶的基因进行筛选的方法。

[0970] AD功能化的CRISPR系统在植物和酵母中的其他应用

[0971] AD功能化的CRISPR系统在生物燃料生产中的用途

[0972] 如本文所用的术语“生物燃料”是由植物和植物源的来源制成的替代燃料。可以从通过固碳过程获得其能量的有机质中提取可再生生物燃料,或者通过使用或转化生物质制成可再生生物燃料。可以将该生物质直接用于生物燃料,或者可以通过热转化、化学转化和生化转化将其转化为方便的含能量物质。这种生物质转化可以产生固体、液体或气体形式的燃料。有两种类型的生物燃料:生物乙醇和生物柴油。生物乙醇主要通过纤维素(淀粉)的糖发酵过程产生,纤维素(淀粉)主要来源于玉米和甘蔗。在另一方面,生物柴油主要由油料作物诸如油菜籽、棕榈和大豆产生。生物燃料主要用于运输。

[0973] 增强植物特性用于生产生物燃料

[0974] 在特定的实施方案中,本文所述的使用AD功能化的CRISPR系统的方法用于改变细胞壁的特性,以便关键水解剂进入,从而更有效地释放糖进行发酵。在特定的实施方案中,对纤维素和/或木质素的生物合成进行了修饰。纤维素是细胞壁的主要成分。纤维素和木质素的生物合成是共同调控的。通过减少植物中木质素的比例,可以增加纤维素的比例。在特定的实施方案中,使用本文所述的方法来下调植物中木质素的生物合成,从而增加可发酵的碳水化合物。更具体地,使用本文所述的方法来下调至少第一木质素生物合成基因,该第一木质素生物合成基因选自4-香豆酸3-羟化酶(C3H)、苯丙氨酸氨裂解酶(PAL)、肉桂酸4-羟化酶(C4H)、羟基肉桂酰转移酶(HCT)、咖啡酸O-甲基转移酶(COMT)、咖啡酰CoA 3-O-甲基转移酶(CCoAOMT)、阿魏酸5-羟化酶(F5H)、肉桂醇脱氢酶(CAD)、肉桂酰CoA-还原酶(CCR)、4-香豆酸-CoA连接酶(4CL)、单木质醇-木质素-特异性糖基转移酶和醛脱氢酶(ALDH),如WO 2008064289A2中所公开。

[0975] 在特定的实施方案中,使用本文所述的方法来产生植物质,该植物质在发酵过程中产生较低水平的乙酸(还可参见WO 2010096488)。更具体地,本文公开的方法用于产生与Cas1L同源的突变以减少多糖乙酰化。

[0976] 修饰酵母用于产生生物燃料

[0977] 在特定的实施方案中,本文提供的AD功能化的CRISPR系统用于通过重组微生物来产生生物乙醇。例如,AD功能化的CRISPR系统可以用于工程化微生物诸如酵母,以由可发酵糖产生生物燃料或生物聚合物,并且任选地能够降解来源于农业废物的植物源木质纤维素作为可发酵糖的来源。在一些实施方案中,AD功能化的CRISPR系统用于修饰与生物燃料产生途径竞争的内源代谢途径。

[0978] 因此,在更多特定的实施方案中,本文所述的方法用于如下修饰微生物:修饰所述宿主细胞的代谢途径中的至少一种编码酶的核酸,其中所述途径产生除来源于丙酮酸盐的乙醛或来源于乙醛的乙醇以外的代谢物,并且其中所述修饰导致所述代谢物的产量降低,

或引入至少一种编码所述酶的抑制剂的核酸。

[0979] 修饰藻类和植物用于产生植物油或生物燃料

[0980] 例如,转基因藻类或其他植物诸如油菜在产生植物油或生物燃料诸如醇(尤其是甲醇和乙醇)中可能特别有用。可将它们工程化为表达或过量表达用于油或生物燃料行业的高含量油或醇。

[0981] 根据本发明的特定实施方案,AD功能化的CRISPR系统用于产生富含脂质的硅藻,所述硅藻可用于产生生物燃料。

[0982] 在特定的实施方案中,设想了特异性修饰与藻类细胞产生的脂质的量和/或脂质的质量的改变有关的基因。编码与脂肪酸合成有关途径的酶的基因的实例可以编码具有以下酶活性的蛋白:例如乙酰-CoA羧化酶、脂肪酸合酶、3-酮酰基_酰基-运载体蛋白合酶III、甘油-3-磷酸脱氢酶(G3PDH)、烯酰基-酰基运载体蛋白还原酶(烯酰基-ACP-还原酶)、甘油-3-磷酸酰基转移酶、溶血磷脂酰基转移酶或二酰基甘油酰基转移酶、磷脂:二酰基甘油酰基转移酶、磷脂酰磷酸酶、脂肪酸硫酯酶(诸如棕榈酰蛋白硫酯酶)或苹果酸酶。在另外的实施方案中,设想了产生脂质累积增加的硅藻。这可以通过靶向降低脂质分解代谢的基因来实现。在本发明的方法中使用特别有意义的是与三酰甘油和游离脂肪酸的活化有关的基因,以及与脂肪酸的β-氧化(诸如酰基-CoA合酶、3-酮酰基-CoA硫解酶、酰基-CoA氧化酶活性和磷酸葡萄糖变位酶)直接有关的基因。本文所述的AD功能化的CRISPR系统和方法可以用于特异性激活硅藻中的此类基因以增加其脂质含量。

[0983] 生物诸如微藻被广泛用于合成生物学。Stovicek等人(Metab.Eng.Comm.,2015;2:13)描述了对工业酵母例如酿酒酵母(Saccharomyces cerevisae)进行基因组编辑,以有效产生用于工业生产的稳健菌株。Stovicek使用针对酵母进行密码子优化的CRISPR-Cas9系统来同时破坏内源基因的两个等位基因并敲入异源基因。Cas9和指导RNA从基于基因组或游离2μ的载体位置表达。作者还表明,通过优化Cas9和指导RNA表达水平可以提高基因破坏效率。Hlavová等人(Biotechnol.Adv.2015)讨论了使用诸如CRISPR的技术来靶向核基因和叶绿体基因进行插入诱变和筛选,从而开发微藻物种或菌株。

[0984] US 8945839描述了一种使用Cas9来工程化微藻类(莱茵衣藻(Chlamydomonas reinhardtii)细胞)的方法。使用类似的工具,可以将本文所述的AD功能化的CRISPR系统的方法应用于衣藻物种和其他藻类。在特定的实施方案中,将CRISPR-Cas蛋白(例如,Cas13)、腺苷脱氨酶(其可与CRISPR-Cas蛋白或适配体结合衔接蛋白融合)和指导RNA引入使用在组成型启动子诸如Hsp70A-Rbc S2或Beta2-微管蛋白的控制下表达CRISPR-Cas蛋白和任选的腺苷脱氨酶的载体表达的藻类中。将使用含有T7启动子的载体来递送指导RNA。另选地,可以将mRNA和体外转录的指导RNA递送到藻类细胞。电穿孔方案遵循来源于GeneArt Chlamydomonas Engineering试剂盒的标准推荐方案。

[0985] AD功能化的CRISPR系统在产生能够产生脂肪酸的微生物中的用途

[0986] 在特定的实施方案中,本发明的方法用于产生能够产生脂肪酯(诸如脂肪酸甲酯(“FAME”)和脂肪酸乙酯(“FAEE”))的遗传工程化微生物。

[0987] 通常,可以将宿主细胞工程化为通过表达或过量表达编码硫酯酶的基因、编码酰基-CoA合酶的基因以及编码酯合酶的基因而从培养基中存在的碳源(诸如醇)产生脂肪酸酯。因此,本文提供的方法用于修饰微生物以过量表达或引入硫酯酶基因、编码酰基-CoA合

酶的基因以及编码酯合酶的基因。在特定的实施方案中，硫酯酶基因选自tesA、'tesA、tesB、fatB、fatB2、fatB3、fatA1或fatA。在特定的实施方案中，编码酰基-CoA合酶的基因选自fadDJadK、BH3103、pfl-4354、EAV15023、fadD1、fadD2、RPC_4074、fadDD35、fadDD22、faa39或编码具有相同特性的酶的鉴定基因。在特定的实施方案中，编码酯合酶的基因是编码来自以下物种的合酶/酰基-CoA:二酰基甘油酰基转移酶的基因或其变体:加州希蒙得木(Simmondsia chinensis)、不动杆菌属某种(Acinetobacter sp.)ADP、泊库岛食烷菌(Alcanivorax borkumensis)、铜绿假单胞菌(Pseudomonas aeruginosa)、亚德海床杆菌(Fundibacter jadensis)、拟南芥(Arabidopsis thaliana)或真养产碱杆菌(Alkaligenes eutrophus)。另外地或另选地，本文提供的方法用于降低编码酰基-CoA脱氢酶的基因、编码外膜蛋白受体的基因以及编码脂肪酸生物合成的转录调控因子的基因中的一者或多者在所述微生物中的表达。在特定的实施方案中，诸如通过引入突变使这些基因中的一者或多者失活。在特定的实施方案中，编码酰基-CoA脱氢酶的基因是fadE。在特定的实施方案中，编码脂肪酸生物合成的转录调控因子的基因编码DNA转录阻遏因子，例如fabR。

[0988] 另外地或另选地，所述微生物经修饰以减少编码丙酮酸甲酸裂解酶的基因、编码乳酸脱氢酶的基因中的至少一者或两者的表达。在特定的实施方案中，编码丙酮酸甲酸裂解酶的基因是pflB。在特定的实施方案中，编码乳酸脱氢酶的基因是IdhA。在特定的实施方案中，诸如通过在其中引入突变使这些基因中的一者或多者失活。

[0989] 在特定的实施方案中，微生物选自以下属:埃希氏菌属(Escherichia)、芽孢杆菌属(Bacillus)、乳杆菌属(Lactobacillus)、红球菌属(Rhodococcus)、聚球藻属(Synechococcus)、集胞藻属(Synechoystis)、假单胞菌属(Pseudomonas)、曲霉属(Aspergillus)、木霉属(Trichoderma)、脉孢菌属(Neurospora)、镰刀菌属(Fusarium)、腐质霉属(Humicola)、根毛霉属(Rhizomucor)、克鲁维酵母菌属(Kluyveromyces)、毕赤酵母属(Pichia)、毛霉属(Mucor)、蚀丝霉属(Myceliophtora)、青霉菌属(Penicillium)、原毛平革菌属(Phanerochaete)、侧耳属(Pleurotus)、栓菌属(Trametes)、金孢子菌属(Chrysosporium)、酵母菌属(Saccharomyces)、寡养单胞菌属(Stenotrophamonas)、裂殖酵母属(Schizosaccharomyces)、耶氏酵母属(Yarrowia)或链霉菌属(Streptomyces)。

[0990] AD功能化的CRISPR系统在产生能够产生有机酸的微生物中的用途

[0991] 本文提供的方法进一步用于工程化能够产生有机酸的微生物，更具体为由戊糖或己糖产生有机酸的微生物。在特定的实施方案中，所述方法包括将外源LDH基因引入微生物中。在特定的实施方案中，通过失活编码与内源代谢途径有关的蛋白质的内源基因来另外地或另选地增加所述微生物中的有机酸产生，所述内源代谢途径产生目标有机酸以外的代谢物并且/或者其中所述内源代谢途径消耗有机酸。在特定的实施方案中，修饰确保减少目标有机酸以外的代谢物的产生。根据特定的实施方案，所述方法用于引入内源途径(其中有机酸被消耗)的至少一种工程化基因缺失和/或失活，或者编码与产生目标有机酸以外的代谢物的内源途径有关的产物的基因。在特定的实施方案中，至少一种工程化基因缺失或失活在编码选自以下各项组成的组的酶的一个或多个基因中:丙酮酸脱羧酶(pdc)、富马酸还原酶、醇脱氢酶(adh)、乙醛脱氢酶、磷酸烯醇丙酮酸羧化酶(ppc)、D-乳酸脱氢酶(d-ldh)、L-乳酸脱氢酶(l-ldh)、乳酸2-单加氧酶。在另外的实施方案中，至少一种工程化基因缺失和/或失活在编码丙酮酸脱羧酶(pdc)的内源基因中。

[0992] 在另外的实施方案中，将微生物工程化为产生乳酸，并且至少一种工程化基因缺失和/或失活在编码乳酸脱氢酶的内源基因中。另外地或另选地，微生物包括编码细胞色素依赖性乳酸脱氢酶(诸如细胞色素B2依赖性L-乳酸脱氢酶)的内源基因的至少一种工程化基因缺失或失活。

[0993] AD功能化的CRISPR系统在利用酵母菌株产生改良的木糖或纤维二糖中的用途

[0994] 在特定的实施方案中，AD功能化的CRISPR系统可用于利用酵母菌株选择改良的木糖或纤维二糖。可以使用易错PCR来扩增与木糖利用或纤维二糖利用途径有关的一个(或多个)基因。与木糖利用途径和纤维二糖利用途径有关的基因的实例可包括但不限于以下文献中所述的那些基因：Ha,S.J.等人,(2011)Proc.Natl.Acad.Sci.USA 108(2):504-9以及Galazka,J.M.等人,(2010)Science 330(6000):84-6。可以将所得的双链DNA分子文库(每个在如此选定的基因中包含随机突变)与AD功能化的CRISPR系统的组分共转化到酵母菌株(例如S288C)中，并且可以选择木糖或纤维二糖利用能力增强的菌株，如WO2015138855中所述。

[0995] AD功能化的CRISPR系统在产生用于类异戊二烯生物合成的改良酵母菌株中的用途

[0996] Tadas Jakočiūnas等人描述了多重CRISPR-Cas9系统在面包酵母酿酒酵母中的一个转化步骤中成功应用于对多达5个不同基因组基因座进行基因组工程化(Metabolic Engineering第28卷，2015年3月，第213-222页)，从而获得甲羟戊酸产量增高的菌株，甲羟戊酸是工业上重要的类异戊二烯生物合成途径的关键中间体。在特定的实施方案中，AD功能化的CRISPR系统可用于本文所述的多重基因组工程化方法中，以鉴定用于类异戊二烯合成的另外的高产酵母菌株。

[0997] 改良的植物和酵母细胞

[0998] 本发明还提供了通过本文提供的方法可获得并获得的植物和酵母细胞。通过本文所述的方法获得的改良植物可通过例如确保对植物虫害、除草剂、干旱、低温或高温、过量水分等产生耐受性的基因的表达而用于食品或饲料生产中。

[0999] 通过本文所述的方法获得的改良植物、尤其是作物和藻类可通过表达例如比在野生型中通常所见的更高的蛋白、碳水化合物、营养物或维生素水平而用于食品或饲料生产中。就这一点而言，改良植物、尤其是豆类和块茎是优选的。

[1000] 例如，改良的藻类或其他植物诸如油菜在生产植物油或生物燃料诸如醇(尤其是甲醇和乙醇)中可能特别有用。可以将它们工程化为表达或过量表达用于油或生物燃料行业的高含量油或醇。

[1001] 本发明还提供了植物的改良部分。植物部分包括但不限于叶、茎、根、块茎、种子、胚乳、胚珠和花粉。本文设想的植物部分可以是可存活的、不可存活的、可再生的和/或不可再生的。

[1002] 本文还涵盖了提供根据本发明的方法产生的植物细胞和植物。通过传统育种方法产生的包含遗传修饰的植物的配子、种子、胚胎(合子或体细胞)、子代或杂种也包括在本发明的范围内。此类植物可包含插入或代替靶序列的异源或外源DNA序列。另选地，此类植物可在一个或多个核苷酸中仅包含改变(突变、缺失、插入、取代)。因此，此类植物与它们的祖先植物的不同之处将仅在于存在特定修饰。

[1003] 因此，本发明提供了通过本方法产生的植物、动物或细胞，或者其子代。子代可以是所产生的植物或动物的克隆，或者可通过与相同物种的其他个体杂交以将其他所需性状渗入其后代中的有性繁殖而产生。在多细胞生物、特别是动物或植物的情况下，细胞可以是体内或离体的。

[1004] 使用本文所述的AD功能化的CRISPR系统进行基因组编辑的方法可以用于赋予基本上任何植物、藻类、真菌、酵母等所需的性状。可使用本公开的核酸构建体和上述各种转化方法针对本文所述的所需生理和农艺特性对多种植物、藻类、真菌、酵母等以及植物、藻类、真菌、酵母细胞或组织系统进行工程化。

[1005] 在特定的实施方案中，本文所述的方法用于修饰内源基因或修饰其表达，而无需将任何外源基因(包括编码CRISPR组分的那些基因)永久性引入植物、藻类、真菌、酵母等的基因组中，从而避免植物的基因组中存在外源DNA。这可能是有意义的，因为对非转基因植物的监管要求没那么严格。

[1006] 本文所述的方法通常会产生“改良植物、藻类、真菌、酵母等”，因为与野生型植物相比，它们具有一种或多种期望的性状。在特定的实施方案中，获得非转基因的遗传修饰的植物、藻类、真菌、酵母等、其部分或细胞，因为没有将外源DNA序列掺入植物的任何细胞的基因组中。在此类实施方案中，改良的植物、藻类、真菌、酵母等是非转基因的。在仅确保内源基因的修饰且在植物、藻类、真菌、酵母等基因组中没有引入或维持外源基因的情况下，所得的遗传修饰的作物不包含外源基因，因此基本上可以被认为是非转基因的。AD功能化的CRISPR系统在植物、藻类、真菌、酵母等基因组编辑中的不同应用包括但不限于：编辑内源基因以赋予目标农业性状。赋予农艺性状的示例性基因包括但不限于赋予虫害或病害抗性的基因；与植物病害有关的基因，所述病害诸如WO 2013046247中列出的病害；赋予对除草剂、杀真菌剂等的抗性的基因；与(非生物)胁迫耐受性有关的基因。CRISPR-Cas系统的用途的其他方面包括但不限于：创建(雄性)不育植物；增加植物/藻类等的生育期；在目标作物中产生遗传变异；影响果实催熟；增加植物/藻类等的贮藏寿命；减少植物/藻类等中的过敏原；确保价值增加的性状(例如，营养改善)；对目标内源基因进行筛选的方法；生物燃料、脂肪酸、有机酸等的产生。

[1007] AD功能化的CRISPR系统可以用于非人类生物

[1008] 一方面，本发明提供了一种非人类真核生物、优选多细胞真核生物，其包含根据任何所述实施方案的真核宿主细胞。在其他方面，本发明提供了一种真核生物、优选多细胞真核生物，其包含根据任何所述实施方案的真核宿主细胞。在这些方面的一些实施方案中，生物可以是动物，例如哺乳动物。而且，生物可以是节肢动物，诸如昆虫。本发明还可扩展到其他农业应用，诸如农场和生产动物。例如，猪的许多特征使其作为生物医学模型(特别是在再生医学中)具有吸引力。具体地，具有严重联合免疫缺陷(SCID)的猪可为再生医学、异种移植(也在本文其他地方讨论)和肿瘤发展提供有用的模型，并且将有助于开发用于人类SCID患者的疗法。Lee等人(Proc Natl Acad Sci U S A.2014年5月20日;111(20):7260-5)利用报告子指导的转录激活因子样效应核酸酶(TALEN)系统高效地在体细胞中产生了重组激活基因(RAG)2的靶向修饰，包括一些影响两个等位基因的修饰。AD功能化的CRISPR系统可用于类似的系统。

[1009] Lee等人(Proc Natl Acad Sci U S A.2014年5月20日;111(20):7260-5)的方法

可如下类似地应用于本发明。突变的猪是通过对胎儿成纤维细胞中的RAG2进行靶向修饰，然后进行SCNT和胚胎移植而产生的。将编码CRISPR Cas和报告子的构建体电穿孔到胎儿来源的成纤维细胞中。48小时后，将表达绿色荧光蛋白的转染细胞以估计的每孔单细胞稀释度分选到96孔板的各个孔中。通过扩增位于任何CRISPR Cas切割位点两侧的基因组DNA片段，然后对PCR产物进行测序来筛选RAG2的靶向修饰。筛选并确保不存在异位突变后，将带有RAG2的靶向修饰的细胞用于SCNT。除去极体以及卵母细胞的相邻细胞质的一部分(推测含有中期II板)，并将供体细胞放置在卵周膜中。然后将重新构建的胚胎电穿孔以使供体细胞与卵母细胞融合，然后进行化学活化。将活化的胚胎在含0.5μM Scriptaid(S7817;Sigma-Aldrich)的猪受精卵培养基3(PZM3)中孵育14-16小时。然后洗涤胚胎以除去Scriptaid，并在PZM3中培养，直到将它们转移到代孕猪的输卵管中。

[1010] 本发明还适用于修饰其他动物(诸如牛)的SNP。Tan等人(Proc Natl Acad Sci U S A.2013年10月8日;110(41):16526-16531)使用质粒、rAAV和寡核苷酸模板扩展了家畜基因编辑工具箱，使其包括转录激活因子样(TAL)效应核酸酶(TALEN)和成簇规律间隔短回文重复序列(CRISPR)/Cas9刺激的同源定向修复(HDR)。根据基因特异性指导RNA序列的方法(Mali P等人,(2013)RNA-Guided Human Genome Engineering via Cas9.Science 339(6121):823-826)将其克隆到Church实验室指导RNA载体(Addgene ID:41824)中。通过共转染hCas9质粒(Addgene ID:41815)或从RCIScript-hCas9合成的mRNA来提供Cas9核酸酶。通过将来自hCas9质粒(包含hCas9cDNA)的XbaI-AgeI片段亚克隆到RCIScript质粒中来构建此RCIScript-hCas9。

[1011] Heo等人(Stem Cells Dev.2015年2月1日;24(3):393-402.doi:10.1089/scd.2014.0278.2014年11月3日电子版)报道了使用牛多能细胞和成簇规律间隔短回文重复序列(CRISPR)/Cas9核酸酶高效靶向牛基因组的基因。首先，Heo等人通过山中(yamanaka)因子的异位表达以及GSK3β和MEK抑制剂(2i)处理来从牛体细胞成纤维细胞产生诱导的多能干细胞(iPSC)。Heo等人观察到这些牛iPSC在畸胎瘤中的基因表达和发育潜力方面与原始多能干细胞高度相似。此外，对牛NANOG基因座特异的CRISPR-Cas9核酸酶显示出对牛iPSC和胚胎中牛基因组的高效编辑。

[1012] Igenity®提供了对动物(诸如牛)的概况分析，以表现和传播具有重要经济意义的经济性状，诸如畜体组成、畜体质量、母体和生殖性状以及平均日增重。全面的Igenity®概况分析从发现DNA标记(最常见的是单核苷酸多态性或SNP)开始。Igenity®概况背后的所有标记都是由研究机构(包括大学、研究组织和政府实体(诸如USDA))的独立科学家发现的。然后在Igenity®验证群体中分析标记。Igenity®使用代表各种产生环境和生物类型的多种资源群体，通常与来自牛肉行业的种畜、牛犊、饲育场和/或包装部门的行业合作伙伴合作，以收集通常不可用的表型。牛基因组数据库是广泛可用的，参见例如NAGRP牛基因组协调程序(http://www.animalgenome.org/cattle/maps/db.html)。因此，本发明可用于靶向牛SNP。本领域技术人员可利用上述方案来靶向SNP，并将其应用于牛SNP，例如Tan等人或Heo等人所述。

[1013] Qingjian Zou等人(Journal of Molecular Cell Biology Advance Access,2015年10月12日发布)通过靶向狗肌肉生长抑制素(MSTN)基因(骨骼肌质量的负调控因子)

的第一个外显子证实了狗的肌肉质量增加。首先，通过将靶向MSTN的sgRNA与Cas9载体共转染到犬胚胎成纤维细胞(CEF)中来验证sgRNA的效率。此后，通过显微注射具有正常形态的胚胎与Cas9mRNA和MSTN sgRNA的混合物，然后将受精卵自动移植到同一只雌性狗的输卵管中来产生MSTN KO狗。与野生型同窝出生的妹妹相比，敲除小狗的大腿上表现出明显的肌肉表型。这也可以使用本文提供的AD功能化的CRISPR系统进行。

[1014] 家畜—猪

[1015] 在一些实施方案中，家畜中的病毒靶标可包括例如猪巨噬细胞上的猪CD163。CD163与PRRSv(猪繁殖与呼吸综合征病毒，一种动脉炎病毒)的感染(被认为是通过病毒进入细胞)有关。PRRSv的感染、尤其是猪肺泡巨噬细胞(在肺部发现)的感染会导致先前无法治愈的猪综合征(“神秘猪病”或“蓝耳病”)，这种病带来的痛苦包括在家猪中引起生殖障碍、体重减轻和高死亡率。由于巨噬细胞活性丧失会导致免疫缺陷，因此经常会出现机会性感染，诸如地方性动物肺炎、脑膜炎和耳部水肿。由于抗生素使用的增加和经济损失(估计每年6.6亿美元)，也具有重大的经济和环境反响。

[1016] 如Kristin M Whitworth和Dr Randall Prather等人(Nature Biotech 3434，2015年12月7日在线发布)在密苏里州大学与Genus Plc合作所报道，使用CRISPR-Cas9靶向CD163，经过编辑的猪的后代在暴露于PRRSv时具有抗性。使一个创始雄性和一个创始雌性(它们都在CD163的外显子7中具有突变)交配以产生后代。创始雄性在一个等位基因的外显子7中缺失了11bp，这导致结构域5中的氨基酸45发生移码突变和错义翻译，以及随后的氨基酸64出现提前终止密码子。另一个等位基因在外显子7中增加了2bp，并在前一个内含子中缺失了377bp，预计这会导致结构域5的前49个氨基酸表达，然后在氨基酸85出现提前终止密码子。母猪在一个等位基因中增加了7bp，预计在翻译时会表达结构域5的前48个氨基酸，然后在氨基酸70出现提前终止密码子。母猪的其他等位基因无法扩增。预计选定的后代是无效动物(CD163-/-)，即CD163敲除。

[1017] 因此，在一些实施方案中，CRISPR蛋白可靶向猪肺泡巨噬细胞。在一些实施方案中，CRISPR蛋白可靶向猪CD163。在一些实施方案中，可通过诱导DSB或者通过插入或缺失(例如外显子7的靶向缺失或修饰，包括上述一种或多种靶向缺失或修饰，或在基因的其他区域，例如外显子5的缺失或修饰)来敲除猪CD163。

[1018] 还设想了经过编辑的猪及其后代，例如CD163敲除猪。这可以用于家畜繁殖或建模目的(即猪模型)。还提供了包含基因敲除的精子。

[1019] CD163是富含半胱氨酸的清道夫受体(SRCR)超家族的成员。基于体外研究，蛋白质的SRCR结构域5是负责解包和释放病毒基因组的结构域。因此，也可靶向SRCR超家族的其他成员，以评估对其他病毒的抗性。PRRSV也是哺乳动物动脉炎病毒组的成员，该组还包括鼠乳酸脱氢酶升高病毒、猿出血热病毒和马动脉炎病毒。动脉炎病毒具有重要的致病机制特性，包括巨噬细胞趋向性以及引起严重疾病和持续感染的能力。因此，可例如通过猪CD163或其在其他物种中的同系物来靶向动脉炎病毒，尤其是鼠乳酸脱氢酶升高病毒、猿出血热病毒和马动脉炎病毒，并且还提供了鼠、猿和马模型和敲除。

[1020] 实际上，这种方法可扩展到引起其他家畜疾病的病毒或细菌，这些疾病可能会传播给人类，诸如猪流感病毒(SIV)株(其包括丙型流感和甲型流感亚型(称为H1N1、H1N2、H2N1、H3N1、H3N2和H2N3))以及上述的肺炎、脑膜炎和水肿。

[1021] 在一些实施方案中，本文所述的AD功能化的CRISPR系统可以用于遗传修饰猪基因组，以失活一种或多种猪内源逆转录病毒(PERV)基因座，从而有利于猪到人异种移植的临床应用。参见Yang等人，Science350(6264):1101-1104(2015)，该文献全文通过引用并入本文。在一些实施方案中，本文所述的AD功能化的CRISPR系统可以用于产生不包含任何活性猪内源逆转录病毒(PERV)基因座的遗传修饰的猪。

[1022] 利用AD功能化的CRISPR系统进行治疗靶向

[1023] 显然，设想了AD功能化的CRISPR系统可以用于靶向任何目标多核苷酸序列。本发明提供了用于体内、离体或体外修饰靶细胞的一种非天然存在或工程化的组合物，或者编码所述组合物的组分的一种或多种多核苷酸，或者包含编码所述组合物的组分的一种或多种多核苷酸的载体或递送系统，并且可以改变细胞的方式进行，使得CRISPR修饰的细胞的后代或细胞系一旦被修饰就保留改变的表型。修饰的细胞和后代可以是多细胞生物(诸如将CRISPR系统离体或体内应用于所需细胞类型的植物或动物)的一部分。CRISPR发明可以是治疗方法。治疗方法可包括基因或基因组编辑或者基因疗法。

[1024] 过继细胞疗法

[1025] 本发明还设想了使用本文所述的AD功能化的CRISPR系统来修饰用于过继疗法的细胞。因此，本发明的各方面涉及对选定抗原(诸如肿瘤相关抗原)特异的免疫系统细胞(诸如T细胞)的过继转移(参见Maus等人，2014，Adoptive Immunotherapy for Cancer or Viruses，Annual Review of Immunology，第32卷:189-225；Rosenberg和Restifo，2015，Adoptive cell transfer as personalized immunotherapy for human cancer，Science，第348卷，第6230期，第62-68页；Restifo等人，2015，Adoptive immunotherapy for cancer:harnessing the T cell response.Nat.Rev.Immunol.12(4):269-281；以及Jenson和Riddell，2014，Design and implementation of adoptive therapy with chimeric antigen receptor-modified T cells.Immunol Rev.257(1):127-144)。例如，可采用各种策略通过改变T细胞受体(TCR)的特异性(例如，通过引入具有选定肽特异性的新TCRα和β链)来对T细胞进行遗传修饰(参见美国专利号8,697,854；PCT专利公开：WO2003020763、WO2004033685、WO2004044004、WO2005114215、WO2006000830、WO2008038002、WO2008039818、WO2004074322、WO2005113595、WO2006125962、WO2013166321、WO2013039889、WO2014018863、WO2014083173；美国专利号8,088,379)。

[1026] 作为TCR修饰的替代或除TCR修饰之外，可使用嵌合抗原受体(CAR)来产生对所选靶标(诸如恶性细胞)特异的免疫响应细胞(诸如T细胞)，其中已经描述了多种受体嵌合体构建体(参见美国专利号5,843,728、5,851,828、5,912,170、6,004,811、6,284,240、6,392,013、6,410,014、6,753,162、8,211,422、以及PCT公开WO9215322)。可将替代CAR构建体表征为属于连续世代。第一代CAR通常由对抗原特异的抗体的单链可变片段组成，例如包含与特异性抗体的VH连接的VL，通过柔性接头(例如通过CD8α铰链结构域和CD8α跨膜结构域)与CD3ζ或FcRγ(scFv-CD3ζ或scFv-FcRγ；参见美国专利号7,741,465；美国专利号5,912,172；美国专利号5,906,936)的跨膜和细胞内信号传导结构域连接。第二代CAR将一个或多个共刺激分子的细胞内结构域(诸如CD28、OX40(CD134)或4-1BB(CD137))掺入内结构域(例如scFv-CD28/OX40/4-1BB-CD3ζ；参见美国专利号8,911,993、8,916,381、8,975,071、9,101,584、9,102,760、9,102,761)内。第三代CAR包括共刺激内结构域的组合，诸如CD3ζ链、

CD97、GDI la-CD18、CD2、ICOS、CD27、CD154、CDS、OX40、4-1BB或CD28信号传导结构域(例如scFv-CD28-4-1BB-CD3ζ或scFv-CD28-OX40-CD3ζ;参见美国专利号8,906,682;美国专利号8,399,645;美国专利号5,686,281;PCT公开号WO2014134165;PCT公开号WO2012079000)。另选地,共刺激可通过在抗原特异性T细胞中表达CAR来精心安排,选择所述抗原特异性T细胞以使其与其天然αβTCR接合后被激活和扩增,例如利用伴随的共刺激通过专业抗原呈递细胞上的抗原。另外,可在免疫响应细胞上提供其他工程化的受体,例如以改善对T细胞攻击的靶向和/或最小化副作用。

[1027] 替代技术可用于转化靶免疫响应细胞,诸如原生质体融合、脂转染、转染或电穿孔。可使用多种载体,诸如逆转录病毒载体、慢病毒载体、腺病毒载体、腺相关病毒载体、质粒或转座子,诸如睡美人转座子(参见美国专利号6,489,458、7,148,203、7,160,682、7,985,739、8,227,432),这些载体可用于引入CAR,例如通过CD3ζ和CD28或CD137使用第二代抗原特异性CAR信号传导。病毒载体可例如包括基于HIV、SV40、EBV、HSV或BPV的载体。

[1028] 被靶向以进行转化的细胞可例如包括T细胞、自然杀伤(NK)细胞、细胞毒性T淋巴细胞(CTL)、调控T细胞、人胚胎干细胞、肿瘤浸润淋巴细胞(TIL)或淋巴样细胞可从其分化的多能干细胞。可例如通过与共表达癌症抗原和共刺激分子的经γ辐射的活化和增殖细胞(AaPC)共培养来选择表达所需CAR的T细胞。可例如通过在存在可溶性因子(诸如IL-2和IL-21)的情况下在AaPC上共培养来扩增工程化的CAR T细胞。例如,可进行这种扩增以提供记忆CAR+ T细胞(可例如通过非酶数字阵列和/或多面板流式细胞术进行测定)。以这种方式,可提供对带有抗原的肿瘤具有特异性细胞毒活性的CAR T细胞(任选地与产生所需趋化因子诸如干扰素-γ结合)。这种CAR T细胞可例如用于动物模型中,例如以威胁肿瘤异种移植物。

[1029] 诸如前述的方法可适于提供例如通过施用有效量的包含与所选抗原结合的抗原识别受体的免疫响应细胞来治疗患有疾病(诸如瘤形成)的受试者和/或增加所述受试者的存活率的方法,其中所述结合激活免疫响应细胞,从而治疗或预防疾病(诸如瘤形成、病原体感染、自身免疫性疾病或同种异体移植反应)。CAR T细胞疗法中的剂量可例如包括在有或没有淋巴细胞清除过程的情况下例如与环磷酰胺一起施用106至109个细胞/kg。

[1030] 在一个实施方案中,可以将治疗施用于经历免疫抑制治疗的患者中。可使细胞或细胞群体对至少一种免疫抑制剂具有抗性,因为编码这种免疫抑制剂的受体的基因失活。不受理论的束缚,免疫抑制治疗应有助于在患者体内选择和扩增根据本发明的免疫响应或T细胞。

[1031] 施用根据本发明的细胞或细胞群体可以任何方便的方式进行,包括通过气雾吸入、注射、摄取、输液、植入或移植。细胞或细胞群体可皮下、皮内、瘤内、结节内、髓内、肌内、通过静脉内或淋巴内注射或者腹膜内施用于患者。在一个实施方案中,本发明的细胞组合物优选地通过静脉内注射施用。

[1032] 施用细胞或细胞群可以包括施用104-109个细胞/kg体重,优选105-106个细胞/kg体重,包括这些范围内的所有整数值的细胞数。CAR T细胞疗法中的剂量可例如包括在有或没有淋巴细胞清除过程的情况下例如与环磷酰胺一起施用106至109个细胞/kg。细胞或细胞群体可以以一个或多个剂量施用。在另一个实施方案中,有效量的细胞以单剂量施用。在另一个实施方案中,有效量的细胞在一段时间内以一个以上的剂量施用。施用时间由管理

医生决定,并取决于患者的临床状况。细胞或细胞群体可从任何来源(诸如血库或供体)获得。虽然个体需求不同,但是针对特定疾病或病症确定给定细胞类型的有效量的最佳范围在本领域技术人员的能力范围内。有效量是指提供治疗性或预防性益处的量。施用的剂量将取决于接受者的年龄、健康状况和体重、同期治疗(如果有的话)的种类、治疗的频率和所需效果的性质。

[1033] 在另一个实施方案中,有效量的细胞或包含那些细胞的组合物肠胃外施用。施用可以是静脉内施用。施用可以通过在肿瘤内注射直接进行。

[1034] 为了防止可能的不良反应,工程化的免疫响应细胞可配备有转基因安全开关,其形式为使细胞容易暴露于特定信号中的转基因。例如,可以这种方式使用单纯疱疹病毒胸苷激酶(TK)基因,例如通过在干细胞移植后引入用作供体淋巴细胞输注的同种异体T淋巴细胞中(Greco等人,Improving the safety of cell therapy with the TK-suicide gene.Front.Pharmacol.2015;6:95)。在此类细胞中,施用核苷前药(诸如更昔洛韦或阿昔洛韦)会引起细胞死亡。替代的安全开关构建体包括例如通过施用小分子二聚体来触发的诱导型胱天蛋白酶9,所述小分子二聚体将两个无功能的icasp9分子聚集在一起形成活性酶。已经描述了用于实施细胞增殖控制的多种替代方法(参见美国专利公开号20130071414;PCT专利公开WO2011146862;PCT专利公开WO2014011987;PCT专利公开WO2013040371;Zhou等人,BLOOD,2014,123/25:3895-3905;Di Stasi等人,The New England Journal of Medicine 2011;365:1673-1683;Sadelain M,The New England Journal of Medicine 2011;365:1735-173;Ramos等人,Stem Cells 28(6):1107-15(2010))。

[1035] 在过继疗法的进一步改进中,利用本文所述的AD功能化的CRISPR-Cas系统进行基因组编辑可用于使免疫响应细胞适于作为替代实施方式,例如提供经过编辑的CAR T细胞(参见Poirot等人,2015,Multiplex genome edited T-cell manufacturing platform for"off-the-shelf"adoptive T-cell immunotherapies,Cancer Res 75(18):3853)。例如,可编辑免疫响应细胞以缺失II型和/或I型HLA分子中的一些或全部的表达,或者敲除可抑制所需免疫响应的选定基因,诸如PD1基因。

[1036] 可使用本文所述的AD功能化的CRISPR系统来编辑细胞。可通过本文所述的任何方法将AD功能化的CRISPR系统递送到免疫细胞。在优选的实施方案中,离体编辑细胞并将其转移到有需要的受试者体内。可编辑免疫响应细胞、CAR-T细胞或用于过继细胞转移的任何细胞。可进行编辑以消除潜在的同种反应性T细胞受体(TCR)、破坏化学治疗剂的靶标、阻断免疫检查点、激活T细胞以及/或者增加功能衰竭或功能障碍的CD8+ T细胞的分化和/或增殖(参见PCT专利公开:WO2013176915、WO2014059173、WO2014172606、WO2014184744和WO2014191128)。编辑可导致基因失活。

[1037] T细胞受体(TCR)是响应于抗原呈递而参与T细胞活化的细胞表面受体。TCR通常由两条链α和β组成,这两条链组装形成异二聚体并与CD3转导亚基缔合,形成存在于细胞表面上的T细胞受体复合物。TCR的每条α和β链均由免疫球蛋白样N端可变(V)区和恒定(C)区、疏水性跨膜结构域和短细胞质区域组成。关于免疫球蛋白分子,α和β链的可变区是通过V(D)J重组产生的,从而在T细胞群体内产生了多种多样的抗原特异性。然而,与识别完整抗原的免疫球蛋白相反,T细胞被与MHC分子缔合的加工肽片段激活,从而将额外的尺寸引入T细胞

的抗原识别，这被称为MHC限制。通过T细胞受体识别供体和受体之间的MHC差异会导致T细胞增殖以及移植物抗宿主病(GVHD)潜在发展。TCRα或TCRβ失活可导致TCR从T细胞的表面TCR消除，从而阻止对同种抗原和因此对GVHD的识别。然而，TCR破坏通常会导致CD3信号传导组分消除，并改变进一步T细胞扩增的方式。

[1038] 同种异体细胞被宿主免疫系统迅速排斥。已经证明，存在于未辐射血液制品中的同种异体白细胞将存续不超过5至6天(Boni,Muranski等人,2008Blood 1;112(12):4746-54)。因此，为了防止同种异体细胞被排斥，通常必须在一定程度上抑制宿主的免疫系统。然而，在过继细胞转移的情况下，使用免疫抑制药物也对引入的治疗性T细胞具有有害作用。因此，为了在这些情况下有效地使用过继免疫疗法，将需要引入的细胞对免疫抑制治疗具有抗性。因此，在一个特定的实施方案中，本发明还包括修饰T细胞以使其对免疫抑制剂具有抗性的步骤，优选地通过使至少一个编码免疫抑制剂靶标的基因失活来进行。免疫抑制剂是通过几种作用机制中的一种来抑制免疫功能的试剂。免疫抑制剂可以是但不限于钙调神经磷酸酶抑制剂、雷帕霉素靶标、白介素2受体α链阻滞剂、肌苷单磷酸脱氢酶抑制剂、二氢叶酸还原酶抑制剂、皮质类固醇或免疫抑制抗代谢物。本发明允许通过使T细胞中免疫抑制剂的靶标失活来赋予T细胞免疫抑制抗性以用于免疫疗法。作为非限制性实例，免疫抑制剂的靶标可以是免疫抑制剂受体(诸如CD52)、糖皮质激素受体(GR)、FKBP家族基因成员和亲环蛋白家族基因成员。

[1039] 免疫检查点是减慢或阻止免疫响应的抑制性途径并且防止因免疫细胞的不受控活性所致的过度组织损伤。在某些实施方案中，所靶向的免疫检查点是程序性死亡-1(PD-1或CD279)基因(PDCD1)。在其它实施方案中，所靶向的免疫检查点是细胞毒性T-淋巴细胞相关抗原(CTLA-4)。在额外实施方案中，所靶向的免疫检查点是CD28和CTLA4Ig超家族的另一个成员，例如BTLA、LAG3、ICOS、PDL1或KIR。在其它额外实施方案中，所靶向的免疫检查点是TNFR超家族的成员，例如CD40、OX40、CD137、GITR、CD27或TIM-3。

[1040] 其他免疫检查点包括含Src同源性2结构域的蛋白质酪氨酸磷酸酶1(SHP-1)(Watson HA等人,SHP-1:the next checkpoint target for cancer immunotherapy? Biochem Soc Trans.2016年4月15日;44(2):356-62)。SHP-1是一种广泛表达的抑制性蛋白质酪氨酸磷酸酶(PTP)。在T细胞中，它是抗原依赖性激活和增殖的负调控因子。它是一种胞质蛋白，因此不适合抗体介导的疗法，但它在激活和增殖中的作用使其成为过继转移策略(诸如嵌合抗原受体(CAR)T细胞)中遗传操作的诱人靶标。免疫检查点还可包括具有Ig和ITIM结构域(TIGIT/Vstm3/WUCAM/VSIG9)和VISTA的T细胞免疫受体(Le Mercier I等人,(2015)Beyond CTLA-4and PD-1,the generation Z of negative checkpoint regulators.Front.Immunol.6:418)。

[1041] WO2014172606涉及使用MT1和/或MT1抑制剂来增加耗尽的CD8+ T细胞的增殖和/或活性以及减少CD8+ T细胞耗尽(例如，减少功能性耗尽或无响应的CD8+免疫细胞)。在某些实施方案中，通过在过继转移的T细胞中进行基因编辑来靶向金属硫蛋白。

[1042] 在某些实施方案中，基因编辑的靶标可以是与免疫检查点蛋白的表达有关的至少一个靶向基因座。此类靶标可包括但不限于CTLA4、PPP2CA、PPP2CB、PTPN6、PTPN22、PDCD1、ICOS(CD278)、PDL1、KIR、LAG3、HAVCR2、BTLA、CD160、TIGIT、CD96、CRTAM、LAIR1、SIGLEC7、SIGLEC9、CD244(2B4)、TNFRSF10B、TNFRSF10A、CASP8、CASP10、CASP3、CASP6、CASP7、FADD、

FAS、TGFBRII、TGFRBRI、SMAD2、SMAD3、SMAD4、SMAD10、SKI、SKIL、TGIF1、IL10RA、IL10RB、HMOX2、IL6R、IL6ST、EIF2AK4、CSK、PAG1、SIT1、FOXP3、PRDM1、BATF、VISTA、GUCY1A2、GUCY1A3、GUCY1B2、GUCY1B3、MT1、MT2、CD40、OX40、CD137、GITR、CD27、SHP-1或TIM-3。在优选的实施方案中，靶向与PD-1或CTLA-4基因的表达有关的基因座。在其他优选的实施方案中，靶向基因(诸如但不限于PD-1和TIGIT)的组合。

[1043] 在其他实施方案中，编辑至少两个基因。基因对可包括但不限于PD1和TCRα、PD1和TCRβ、CTLA-4和TCRα、CTLA-4和TCRβ、LAG3和TCRα、LAG3和TCRβ、Tim3和TCRα、Tim3和TCRβ、BTLA和TCRα、BTLA和TCRβ、BY55和TCRα、BY55和TCRβ、TIGIT和TCRα、TIGIT和TCRβ、B7H5和TCRα、B7H5和TCRβ、LAIR1和TCRα、LAIR1和TCRβ、SIGLEC10和TCRα、SIGLEC10和TCRβ、2B4和TCRα、2B4和TCRβ。

[1044] 无论是在遗传修饰T细胞之前还是之后，通常都可以使用例如以下美国专利中所述的方法来活化和扩增T细胞:6,352,694、6,534,055、6,905,680、5,858,358、6,887,466、6,905,681、7,144,575、7,232,566、7,175,843、5,883,223、6,905,874、6,797,514、6,867,041、和7,572,631。可以在体外或体内扩增T细胞。

[1045] 本发明的实践采用免疫学、生物化学、化学、分子生物学、微生物学、细胞生物学、基因组学和重组DNA领域中已知的技术，这些技术在本领域的技术范围内。参见MOLECULAR CLONING:A LABORATORY MANUAL,第2版(1989)(Sambrook、Fritsch和Maniatis);MOLECULAR CLONING:A LABORATORY MANUAL,第4版(2012)(Green和Sambrook);CURRENT PROTOCOLS IN MOLECULAR BIOLOGY(1987)(F.M.Ausubel等人编);METHODS IN ENZYMOLOGY系列(Academic Press,Inc.);PCR 2:A PRACTICAL APPROACH(1995)(M.J.MacPherson、B.D.Hames和G.R.Taylor编);ANTIBODIES,A LABORATORY MANUAL(1988)(Harlow和Lane编);ANTIBODIES A LABORATORY MANUAL,第2版(2013)(E.A.Greenfield编);以及ANIMAL CELL CULTURE(1987)(R.I.Freshney编)。

[1046] 疾病相关突变和致病性SNP的校正

[1047] 在一方面，本文所述的本发明提供了用于修饰靶基因座处的腺苷残基的方法，目的是治疗和/或预防由或可能由G至A或C至T点突变或致病性单核苷酸多态性(SNP)引起的患病病症。

[1048] 影响脑和中枢神经系统的疾病

[1049] 与影响脑和中枢神经系统的各种疾病相关的致病性G至A或C至T突变/SNP在ClinVar数据库中报告并在表A中公开，包括但不限于阿尔茨海默病、帕金森病、自闭症、肌萎缩性脊髓侧索硬化症(ALS)、精神分裂症、肾上腺脑白质营养不良、艾卡尔迪-古铁雷综合征、法布里病、Lesch-Nyhan综合征和门克斯病。因此，本发明的一个方面涉及一种用于校正与这些疾病中的任何一种相关的一种或多种致病性G至A或C至T突变/SNP的方法，如下所述。

[1050] 阿尔茨海默病

[1051] 在一些实施方案中，本文所述的方法、系统和组合物用于校正与阿尔茨海默病相关的一种或多种致病性G至A或C至T突变/SNP。在一些实施方案中，致病性突变/SNP存在于选自PSEN1、PSEN2和APP的至少一个基因中，包括至少以下各项：

[1052] NM_000021.3(PSEN1):c.796G>A(p.Gly266Ser)

[1053] NM_000484.3(APP):c.2017G>A(p.Ala673Thr)

[1054] NM_000484.3(APP):c.2149G>A(p.Val717Ile)

[1055] NM_000484.3(APP):c.2137G>A(p.Ala713Thr)

[1056] NM_000484.3(APP):c.2143G>A(p.Val715Met)

[1057] NM_000484.3(APP):c.2141C>T(p.Thr714Ile)

[1058] NM_000021.3(PSEN1):c.438G>A(p.Met146Ile)

[1059] NM_000021.3(PSEN1):c.1229G>A(p.Cys410Tyr)

[1060] NM_000021.3(PSEN1):c.487C>T(p.His163Tyr)

[1061] NM_000021.3(PSEN1):c.799C>T(p.Pro267Ser)

[1062] NM_000021.3(PSEN1):c.236C>T(p.Ala79Val)

[1063] NM_000021.3(PSEN1):c.509C>T(p.Ser170Phe)

[1064] NM_000447.2(PSEN2):c.1289C>T(p.Thr430Met)

[1065] NM_000447.2(PSEN2):c.717G>A(p.Met239Ile)

[1066] NM_000447.2(PSEN2):c.254C>T(p.Ala85Val)

[1067] NM_000021.3(PSEN1):c.806G>A(p.Arg269His)

[1068] NM_000484.3(APP):c.2018C>T(p.Ala673Val)。

[1069] 参见表A。因此,本发明的一个方面涉及一种用于通过校正一种或多种致病性G至A或C至T突变/SNP、具体是存在于选自PSEN1、PSEN2和APP的至少一个基因中的一种或多种致病性G至A或C至T突变/SNP、更具体是上述一种或多种致病性G至A或C至T突变/SNP来治疗或预防阿尔茨海默病的方法。

[1070] 帕金森病

[1071] 在一些实施方案中,本文所述的方法、系统和组合物用于校正与帕金森病相关的一种或多种致病性G至A或C至T突变/SNP在一些实施方案中致病性突变/SNP存在于选自SNCA、PLA2G6、FBXO7、VPS35、EIF4G1、DNAJC6、PRKN、SYNJ1、CHCHD2、PINK1、PARK7、LRRK2、ATP13A2和GBA的至少一个基因中,包括至少以下各项:

[1072] NM_000345.3(SNCA):c.157G>A(p.Ala53Thr)

[1073] NM_000345.3(SNCA):c.152G>A(p.Gly51Asp)

[1074] NM_003560.3(PLA2G6):c.2222G>A(p.Arg741Gln)

[1075] NM_003560.3(PLA2G6):c.2239C>T(p.Arg747Trp)

[1076] NM_003560.3(PLA2G6):c.1904G>A(p.Arg635Gln)

[1077] NM_003560.3(PLA2G6):c.1354C>T(p.Gln452Ter)

[1078] NM_012179.3(FBXO7):c.1492C>T(p.Arg498Ter)

[1079] NM_012179.3(FBXO7):c.65C>T(p.Thr22Met)

[1080] NM_018206.5(VPS35):c.1858G>A(p.Asp620Asn)

[1081] NM_198241.2(EIF4G1):c.3614G>A(p.Arg1205His)

[1082] NM_198241.2(EIF4G1):c.1505C>T(p.Ala502Val)

[1083] NM_001256865.1(DNAJC6):c.2200C>T(p.Gln734Ter)

[1084] NM_001256865.1(DNAJC6):c.2326C>T(p.Gln776Ter)

[1085] NM_004562.2(PRKN):c.931C>T(p.Gln311Ter)

[1086] NM_004562.2(PRKN):c.1358G>A(p.Trp453Ter)

[1087] NM_004562.2(PRKN):c.635G>A(p.Cys212Tyr)

[1088] NM_203446.2(SYNJ1):c.773G>A(p.Arg258Gln)

[1089] NM_001320327.1(CHCHD2):c.182C>T(p.Thr61Ile)

[1090] NM_001320327.1(CHCHD2):c.434G>A(p.Arg145Gln)

[1091] NM_001320327.1(CHCHD2):c.300+5G>A

[1092] NM_032409.2(PINK1):c.926G>A(p.Gly309Asp)

[1093] NM_032409.2(PINK1):c.1311G>A(p.Trp437Ter)

[1094] NM_032409.2(PINK1):c.736C>T(p.Arg246Ter)

[1095] NM_032409.2(PINK1):c.836G>A(p.Arg279His)

[1096] NM_032409.2(PINK1):c.938C>T(p.Thr313Met)

[1097] NM_032409.2(PINK1):c.1366C>T(p.Gln456Ter)

[1098] NM_007262.4(PARK7):c.78G>A(p.Met26Ile)

[1099] NM_198578.3(LRRK2):c.4321C>T(p.Arg1441Cys)

[1100] NM_198578.3(LRRK2):c.4322G>A(p.Arg1441His)

[1101] NM_198578.3(LRRK2):c.1256C>T(p.Ala419Val)

[1102] NM_198578.3(LRRK2):c.6055G>A(p.Gly2019Ser)

[1103] NM_022089.3(ATP13A2):c.1306+5G>A

[1104] NM_022089.3(ATP13A2):c.2629G>A(p.Gly877Arg)

[1105] NM_022089.3(ATP13A2):c.490C>T(p.Arg164Trp)

[1106] NM_001005741.2(GBA):c.1444G>A(p.Asp482Asn)

[1107] m.15950G>A。

[1108] 参见表A。因此，本发明的一个方面涉及一种用于通过校正一种或多种致病性G至A或C至T突变/SNP、具体是选自SNCA、PLA2G6、FBXO7、VPS35、EIF4G1、DNAJC6、PRKN、SYNJ1、CHCHD2、PINK1、PARK7、LRRK2、ATP13A2和GBA的至少一个基因中的一种或多种致病性G至A或C至T突变/SNP、更具体是上述一种或多种致病性G至A或C至T突变/SNP来治疗或预防帕金森病的方法。

[1109] 自闭症

[1110] 在一些实施方案中，本文所述的方法、系统和组合物用于校正与自闭症相关的一种或多种致病性G至A或C至T突变/SNP。在一些实施方案中，致病性突变/SNP存在于选自MECP2、NLGN3、SLC9A9、EHMT1、CHD8、NLGN4X、GSPT2和PTEN的至少一个基因中，包括至少以下各项：

[1111] NM_001110792.1(MECP2):c.916C>T(p.Arg306Ter)

[1112] NM_004992.3(MECP2):c.473C>T(p.Thr158Met)

[1113] NM_018977.3(NLGN3):c.1351C>T(p.Arg451Cys)

[1114] NM_173653.3(SLC9A9):c.1267C>T(p.Arg423Ter)

[1115] NM_024757.4(EHMT1):c.3413G>A(p.Trp1138Ter)

[1116] NM_020920.3(CHD8):c.2875C>T(p.Gln959Ter)

[1117] NM_020920.3(CHD8):c.3172C>T(p.Arg1058Ter)

[1118] NM_181332.2(NLGN4X):c.301C>T(p.Arg101Ter)

[1119] NM_018094.4(GSPT2):c.1021G>A(p.Val341Ile)

[1120] NM_000314.6(PTEN):c.392C>T(p.Thr131Ile)

[1121] 参见表A。因此,本发明的一个方面涉及一种用于通过校正一种或多种致病性G至A或C至T突变/SNP、具体是选自MECP2、NLGN3、SLC9A9、EHMT1、CHD8、NLGN4X、GSPT2和PTEN的至少一个基因中的一种或多种致病性G至A或C至T突变/SNP、更具体是上述一种或多种致病性G至A或C至T突变/SNP来治疗或预防自闭症的方法。

[1122] 肌萎缩性脊髓侧索硬化症(ALS)

[1123] 在一些实施方案中,本文所述的方法、系统和组合物用于校正与ALS相关的一种或多种致病性G至A或C至T突变/SNP。在一些实施方案中,致病性突变/SNP存在于选自SOD1、VCP、UBQLN2、ERBB4、HNRNPA1、TUBA4A、SOD1、TARDBP、FIG4、OPTN、SETX、SPG11、FUS、VAPB、ANG、CHCHD10、SQSTM1和TBK1的至少一个基因中,包括至少以下各项:

[1124] NM_000454.4(SOD1):c.289G>A(p.Asp97Asn)

[1125] NM_007126.3(VCP):c.1774G>A(p.Asp592Asn)

[1126] NM_007126.3(VCP):c.464G>A(p.Arg155His)

[1127] NM_007126.3(VCP):c.572G>A(p.Arg191Gln)

[1128] NM_013444.3(UBQLN2):c.1489C>T(p.Pro497Ser)

[1129] NM_013444.3(UBQLN2):c.1525C>T(p.Pro509Ser)

[1130] NM_013444.3(UBQLN2):c.1573C>T(p.Pro525Ser)

[1131] NM_013444.3(UBQLN2):c.1490C>T(p.Pro497Leu)

[1132] NM_005235.2(ERBB4):c.2780G>A(p.Arg927Gln)

[1133] NM_005235.2(ERBB4):c.3823C>T(p.Arg1275Trp)

[1134] NM_031157.3(HNRNPA1):c.940G>A(p.Asp314Asn)

[1135] NM_006000.2(TUBA4A):c.643C>T(p.Arg215Cys)

[1136] NM_006000.2(TUBA4A):c.958C>T(p.Arg320Cys)

[1137] NM_006000.2(TUBA4A):c.959G>A(p.Arg320His)

[1138] NM_006000.2(TUBA4A):c.1220G>A(p.Trp407Ter)

[1139] NM_006000.2(TUBA4A):c.1147G>A(p.Ala383Thr)

[1140] NM_000454.4(SOD1):c.112G>A(p.Gly38Arg)

[1141] NM_000454.4(SOD1):c.124G>A(p.Gly42Ser)

[1142] NM_000454.4(SOD1):c.125G>A(p.Gly42Asp)

[1143] NM_000454.4(SOD1):c.14C>T(p.Ala5Val)

[1144] NM_000454.4(SOD1):c.13G>A(p.Ala5Thr)

[1145] NM_000454.4(SOD1):c.436G>A(p.Ala146Thr)

[1146] NM_000454.4(SOD1):c.64G>A(p.Glu22Lys)

[1147] NM_000454.4(SOD1):c.404G>A(p.Ser135Asn)

[1148] NM_000454.4(SOD1):c.49G>A(p.Gly17Ser)

[1149] NM_000454.4(SOD1):c.217G>A(p.Gly73Ser)

[1150] NM_007375.3(TARDBP):c.892G>A(p.Gly298Ser)

[1151] NM_007375.3(TARDBP):c.943G>A(p.Ala315Thr)

[1152] NM_007375.3(TARDBP):c.883G>A(p.Gly295Ser)

[1153] NM_007375.3(TARDBP):c.*697G>A

[1154] NM_007375.3(TARDBP):c.1144G>A(p.Ala382Thr)

[1155] NM_007375.3(TARDBP):c.859G>A(p.Gly287Ser)

[1156] NM_014845.5(FIG4):c.547C>T(p.Arg183Ter)

[1157] NM_001008211.1(OPTN):c.1192C>T(p.Gln398Ter)

[1158] NM_015046.5(SETX):c.6407G>A(p.Arg2136His)

[1159] NM_015046.5(SETX):c.8C>T(p.Thr3Ile)

[1160] NM_025137.3(SPG11):c.118C>T(p.Gln40Ter)

[1161] NM_025137.3(SPG11):c.267G>A(p.Trp89Ter)

[1162] NM_025137.3(SPG11):c.5974C>T(p.Arg1992Ter)

[1163] NM_004960.3(FUS):c.1553G>A(p.Arg518Lys)

[1164] NM_004960.3(FUS):c.1561C>T(p.Arg521Cys)

[1165] NM_004960.3(FUS):c.1562G>A(p.Arg521His)

[1166] NM_004960.3(FUS):c.1520G>A(p.Gly507Asp)

[1167] NM_004960.3(FUS):c.1483C>T(p.Arg495Ter)

[1168] NM_004960.3(FUS):c.616G>A(p.Gly206Ser)

[1169] NM_004960.3(FUS):c.646C>T(p.Arg216Cys)

[1170] NM_004738.4(VAPB):c.166C>T(p.Pro56Ser)

[1171] NM_004738.4(VAPB):c.137C>T(p.Thr46Ile)

[1172] NM_001145.4(ANG):c.164G>A(p.Arg55Lys)

[1173] NM_001145.4(ANG):c.155G>A(p.Ser52Asn)

[1174] NM_001145.4(ANG):c.407C>T(p.Pro136Leu)

[1175] NM_001145.4(ANG):c.409G>A(p.Val137Ile)

[1176] NM_001301339.1(CHCHD10):c.239C>T(p.Pro80Leu)

[1177] NM_001301339.1(CHCHD10):c.176C>T(p.Ser59Leu)

[1178] NM_001142298.1(SQSTM1):c.-47-1924C>T

[1179] NM_003900.4(SQSTM1):c.1160C>T(p.Pro387Leu)

[1180] NM_003900.4(SQSTM1):c.1175C>T(p.Pro392Leu)

[1181] NM_013254.3(TBK1):c.1340+1G>A

[1182] NM_013254.3(TBK1):c.2086G>A(p.Glu696Lys)

[1183] 参见表A。因此,本发明的一个方面涉及一种用于通过校正一种或多种致病性G至A或C至T突变/SNP、具体是存在于选自SOD1、VCP、UBQLN2、ERBB4、HNRNPA1、TUBA4A、SOD1、TARDBP、FIG4、OPTN、SETX、SPG11、FUS、VAPB、ANG、CHCHD10、SQSTM1和TBK1的至少一个基因中的一种或多种致病性G至A或C至T突变/SNP、更具体是上述一种或多种致病性G至A或C至T突变/SNP来治疗或预防ALS的方法。

[1184] 精神分裂症

[1185] 在一些实施方案中,本文所述的方法、系统和组合物用于校正与精神分裂症相关

的一种或多种致病性G至A或C至T突变/SNP。在一些实施方案中，致病性突变/SNP存在于选自PRODH、SETD1A和SHANK3的至少一个基因中，包括至少以下各项：

[1186] NM_016335.4(PRODH):c.1292G>A(p.Arg431His)

[1187] NM_016335.4(PRODH):c.1397C>T(p.Thr466Met)

[1188] NM_014712.2(SETD1A):c.2209C>T(p.Gln737Ter)

[1189] NM_033517.1(SHANK3):c.3349C>T(p.Arg1117Ter)

[1190] NM_033517.1(SHANK3):c.1606C>T(p.Arg536Trp)

[1191] 参见表A。因此，本发明的一个方面涉及一种用于通过校正一种或多种致病性G至A或C至T突变/SNP、具体是存在于选自PRODH、SETD1A和SHANK3的至少一个基因中的一种或多种致病性G至A或C至T突变/SNP、更具体是上述一种或多种致病性G至A或C至T突变/SNP来治疗或预防精神分裂症的方法。

[1192] 肾上腺脑白质营养不良

[1193] 在一些实施方案中，本文所述的方法、系统和组合物用于校正与肾上腺脑白质营养不良 相关的一种或多种致病性G至A或C至T突变/SNP。在一些实施方案中，致病性突变/SNP存在于至少ABCD1基因中，包括至少以下各项：

[1194] NM_000033.3(ABCD1):c.421G>A(p.Ala141Thr)

[1195] NM_000033.3(ABCD1):c.796G>A(p.Gly266Arg)

[1196] NM_000033.3(ABCD1):c.1252C>T(p.Arg418Trp)

[1197] NM_000033.3(ABCD1):c.1552C>T(p.Arg518Trp)

[1198] NM_000033.3(ABCD1):c.1850G>A(p.Arg617His)

[1199] NM_000033.3(ABCD1):c.1396C>T(p.Gln466Ter)

[1200] NM_000033.3(ABCD1):c.1553G>A(p.Arg518Gln)

[1201] NM_000033.3(ABCD1):c.1679C>T(p.Pro560Leu)

[1202] NM_000033.3(ABCD1):c.1771C>T(p.Arg591Trp)

[1203] NM_000033.3(ABCD1):c.1802G>A(p.Trp601Ter)

[1204] NM_000033.3(ABCD1):c.346G>A(p.Gly116Arg)

[1205] NM_000033.3(ABCD1):c.406C>T(p.Gln136Ter)

[1206] NM_000033.3(ABCD1):c.1661G>A(p.Arg554His)

[1207] NM_000033.3(ABCD1):c.1825G>A(p.Glu609Lys)

[1208] NM_000033.3(ABCD1):c.1288C>T(p.Gln430Ter)

[1209] NM_000033.3(ABCD1):c.1781-1G>A

[1210] NM_000033.3(ABCD1):c.529C>T(p.Gln177Ter)

[1211] NM_000033.3(ABCD1):c.1866-10G>A

[1212] 参见表A。因此，本发明的一个方面涉及一种用于通过校正一种或多种致病性G至A或C至T突变/SNP、具体是存在于至少ABCD1基因中的一种或多种致病性G至A或C至T突变/SNP、更具体是上述一种或多种致病性G至A或C至T突变/SNP来治疗或预防肾上腺脑白质营养不良的方法。

[1213] 艾卡尔迪-古铁雷综合征

[1214] 在一些实施方案中，本文所述的方法、系统和组合物用于校正与艾卡尔迪-古铁雷

综合征相关的一种或多种致病性G至A或C至T突变/SNP。在一些实施方案中，致病性突变/SNP存在于选自TREX1、RNASEH2C、ADAR和IFIH1的至少一个基因中，包括至少以下各项：

[1215] NM_016381.5(TREX1):c.794G>A(p.Trp265Ter)

[1216] NM_033629.4(TREX1):c.52G>A(p.Asp18Asn)

[1217] NM_033629.4(TREX1):c.490C>T(p.Arg164Ter)

[1218] NM_032193.3(RNASEH2C):c.205C>T(p.Arg69Trp)

[1219] NM_001111.4(ADAR):c.3019G>A(p.Gly1007Arg)

[1220] NM_022168.3(IFIH1):c.2336G>A(p.Arg779His)

[1221] NM_022168.3(IFIH1):c.2335C>T(p.Arg779Cys)

[1222] 参见表A。因此，本发明的一个方面涉及一种用于通过校正一种或多种致病性G至A或C至T突变/SNP、具体是存在于选自TREX1、RNASEH2C、ADAR和IFIH1的至少一个基因中的一种或多种致病性G至A或C至T突变/SNP、更具体是上述一种或多种致病性G至A或C至T突变/SNP来治疗或预防艾卡尔迪-古铁雷综合征的方法。

[1223] 法布里病

[1224] 在一些实施方案中，本文所述的方法、系统和组合物用于校正与法布里病相关的一种或多种致病性G至A或C至T突变/SNP。在一些实施方案中，致病性突变/SNP存在于至少GLA基因中，包括至少以下各项：

[1225] NM_000169.2(GLA):c.1024C>T(p.Arg342Ter)

[1226] NM_000169.2(GLA):c.1066C>T(p.Arg356Trp)

[1227] NM_000169.2(GLA):c.1025G>A(p.Arg342Gln)

[1228] NM_000169.2(GLA):c.281G>A(p.Cys94Tyr)

[1229] NM_000169.2(GLA):c.677G>A(p.Trp226Ter)

[1230] NM_000169.2(GLA):c.734G>A(p.Trp245Ter)

[1231] NM_000169.2(GLA):c.748C>T(p.Gln250Ter)

[1232] NM_000169.2(GLA):c.658C>T(p.Arg220Ter)

[1233] NM_000169.2(GLA):c.730G>A(p.Asp244Asn)

[1234] NM_000169.2(GLA):c.369+1G>A

[1235] NM_000169.2(GLA):c.335G>A(p.Arg112His)

[1236] NM_000169.2(GLA):c.485G>A(p.Trp162Ter)

[1237] NM_000169.2(GLA):c.661C>T(p.Gln221Ter)

[1238] NM_000169.2(GLA):c.916C>T(p.Gln306Ter)

[1239] NM_000169.2(GLA):c.1072G>A(p.Glu358Lys)

[1240] NM_000169.2(GLA):c.1087C>T(p.Arg363Cys)

[1241] NM_000169.2(GLA):c.1088G>A(p.Arg363His)

[1242] NM_000169.2(GLA):c.605G>A(p.Cys202Tyr)

[1243] NM_000169.2(GLA):c.830G>A(p.Trp277Ter)

[1244] NM_000169.2(GLA):c.979C>T(p.Gln327Ter)

[1245] NM_000169.2(GLA):c.422C>T(p.Thr141Ile)

[1246] NM_000169.2(GLA):c.285G>A(p.Trp95Ter)

[1247] NM_000169.2(GLA):c.735G>A(p.Trp245Ter)

[1248] NM_000169.2(GLA):c.639+919G>A

[1249] NM_000169.2(GLA):c.680G>A(p.Arg227Gln)

[1250] NM_000169.2(GLA):c.679C>T(p.Arg227Ter)

[1251] NM_000169.2(GLA):c.242G>A(p.Trp81Ter)

[1252] NM_000169.2(GLA):c.901C>T(p.Arg301Ter)

[1253] NM_000169.2(GLA):c.974G>A(p.Gly325Asp)

[1254] NM_000169.2(GLA):c.847C>T(p.Gln283Ter)

[1255] NM_000169.2(GLA):c.469C>T(p.Gln157Ter)

[1256] NM_000169.2(GLA):c.1118G>A(p.Gly373Asp)

[1257] 参见表A。因此,本发明的一个方面涉及一种用于通过校正一种或多种致病性G至A或C至T突变/SNP、具体是存在于至少GLA基因中的一种或多种致病性G至A或C至T突变/SNP、更具体是上述一种或多种致病性G至A或C至T突变/SNP来治疗或预防法布里病的方法。

[1258] Lesch-Nyhan综合征

[1259] 在一些实施方案中,本文所述的方法、系统和组合物用于校正与Lesch-Nyhan综合征相关的一种或多种致病性G至A或C至T突变/SNP。在一些实施方案中,致病性突变/SNP存在于至少HPRT1基因中,包括至少以下各项:

[1260] NM_000194.2(HPRT1):c.151C>T(p.Arg51Ter)

[1261] NM_000194.2(HPRT1):c.384+1G>A

[1262] 参见表A。因此,本发明的一个方面涉及一种用于通过校正一种或多种致病性G至A或C至T突变/SNP、具体是存在于至少HPRT1基因中的一种或多种致病性G至A或C至T突变/SNP、更具体是上述一种或多种致病性G至A或C至T突变/SNP来治疗或预防Lesch-Nyhan综合征的方法。

[1263] 门克斯病

[1264] 在一些实施方案中,本文所述的方法、系统和组合物用于校正与门克斯病相关的一种或多种致病性G至A或C至T突变/SNP。在一些实施方案中,致病性突变/SNP存在于至少ATP7A基因中,包括至少以下各项:

[1265] NM_000052.6(ATP7A):c.601C>T(p.Arg201Ter)

[1266] NM_000052.6(ATP7A):c.2938C>T(p.Arg980Ter)

[1267] NM_000052.6(ATP7A):c.3056G>A(p.Gly1019Asp)

[1268] NM_000052.6(ATP7A):c.598C>T(p.Gln200Ter)

[1269] NM_000052.6(ATP7A):c.1225C>T(p.Arg409Ter)

[1270] NM_000052.6(ATP7A):c.1544-1G>A

[1271] NM_000052.6(ATP7A):c.1639C>T(p.Arg547Ter)

[1272] NM_000052.6(ATP7A):c.1933C>T(p.Arg645Ter)

[1273] NM_000052.6(ATP7A):c.1946+5G>A

[1274] NM_000052.6(ATP7A):c.1950G>A(p.Trp650Ter)

[1275] NM_000052.6(ATP7A):c.2179G>A(p.Gly727Arg)

[1276] NM_000052.6(ATP7A):c.2187G>A(p.Trp729Ter)

[1277] NM_000052.6(ATP7A):c.2383C>T(p.Arg795Ter)

[1278] NM_000052.6(ATP7A):c.2499-1G>A

[1279] NM_000052.6(ATP7A):c.2555C>T(p.Pro852Leu)

[1280] NM_000052.6(ATP7A):c.2956C>T(p.Arg986Ter)

[1281] NM_000052.6(ATP7A):c.3112-1G>A

[1282] NM_000052.6(ATP7A):c.3466C>T(p.Gln1156Ter)

[1283] NM_000052.6(ATP7A):c.3502C>T(p.Gln1168Ter)

[1284] NM_000052.6(ATP7A):c.3764G>A(p.Gly1255Glu)

[1285] NM_000052.6(ATP7A):c.3943G>A(p.Gly1315Arg)

[1286] NM_000052.6(ATP7A):c.4123+1G>A

[1287] NM_000052.6(ATP7A):c.4226+5G>A

[1288] 参见表A。因此,本发明的一个方面涉及一种用于通过校正一种或多种致病性G至A或C至T突变/SNP、具体是存在于至少ATP7A基因中的一种或多种致病性G至A或C至T突变/SNP、更具体是上述一种或多种致病性G至A或C至T突变/SNP来治疗或预防门克斯病的方法。

[1289] 眼部疾病

[1290] 与各种眼部疾病相关的致病性G至A或C至T突变/SNP在ClinVar数据库中报告并在表A中公开,包括但不限于斯塔加特病、Bardet-Biedl综合征、视锥视杆细胞营养不良、先天性静止性夜盲症、乌舍尔综合征、Leber先天性黑蒙症、色素性视网膜炎和全色盲。因此,本发明的一个方面涉及一种用于校正与这些疾病中的任何一种相关的一种或多种致病性G至A或C至T突变/SNP的方法,如下所述。

[1291] 斯塔加特病

[1292] 在一些实施方案中,本文所述的方法、系统和组合物用于校正与斯塔加特病相关的一种或多种致病性G至A或C至T突变/SNP。在一些实施方案中,致病性突变/SNP存在于ABCA4基因中,包括至少以下各项:

[1293] NM_000350.2(ABCA4):c.4429C>T(p.Gln1477Ter)

[1294] NM_000350.2(ABCA4):c.6647C>T(p.Ala2216Val)

[1295] NM_000350.2(ABCA4):c.5312+1G>A

[1296] NM_000350.2(ABCA4):c.5189G>A(p.Trp1730Ter)

[1297] NM_000350.2(ABCA4):c.4352+1G>A

[1298] NM_000350.2(ABCA4):c.4253+5G>A

[1299] NM_000350.2(ABCA4):c.3871C>T(p.Gln1291Ter)

[1300] NM_000350.2(ABCA4):c.3813G>A(p.Glu1271=)

[1301] NM_000350.2(ABCA4):c.1293G>A(p.Trp431Ter)

[1302] NM_000350.2(ABCA4):c.206G>A(p.Trp69Ter)

[1303] NM_000350.2(ABCA4):c.3322C>T(p.Arg1108Cys)

[1304] NM_000350.2(ABCA4):c.1804C>T(p.Arg602Trp)

[1305] NM_000350.2(ABCA4):c.1937+1G>A

[1306] NM_000350.2(ABCA4):c.2564G>A(p.Trp855Ter)

[1307] NM_000350.2(ABCA4):c.4234C>T(p.Gln1412Ter)

[1308] NM_000350.2(ABCA4):c.4457C>T(p.Pro1486Leu)

[1309] NM_000350.2(ABCA4):c.4594G>A(p.Asp1532Asn)

[1310] NM_000350.2(ABCA4):c.4919G>A(p.Arg1640Gln)

[1311] NM_000350.2(ABCA4):c.5196+1G>A

[1312] NM_000350.2(ABCA4):c.6316C>T(p.Arg2106Cys)

[1313] NM_000350.2(ABCA4):c.3056C>T(p.Thr1019Met)

[1314] NM_000350.2(ABCA4):c.52C>T(p.Arg18Trp)

[1315] NM_000350.2(ABCA4):c.122G>A(p.Trp41Ter)

[1316] NM_000350.2(ABCA4):c.1903C>T(p.Gln635Ter)

[1317] NM_000350.2(ABCA4):c.194G>A(p.Gly65Glu)

[1318] NM_000350.2(ABCA4):c.3085C>T(p.Gln1029Ter)

[1319] NM_000350.2(ABCA4):c.4195G>A(p.Glu1399Lys)

[1320] NM_000350.2(ABCA4):c.454C>T(p.Arg152Ter)

[1321] NM_000350.2(ABCA4):c.45G>A(p.Trp15Ter)

[1322] NM_000350.2(ABCA4):c.4610C>T(p.Thr1537Met)

[1323] NM_000350.2(ABCA4):c.6112C>T(p.Arg2038Trp)

[1324] NM_000350.2(ABCA4):c.6118C>T(p.Arg2040Ter)

[1325] NM_000350.2(ABCA4):c.6342G>A(p.Val2114＝)

[1326] NM_000350.2(ABCA4):c.6658C>T(p.Gln2220Ter)

[1327] 参见表A。因此，本发明的一个方面涉及一种用于通过校正一种或多种致病性G至A或C至T突变/SNP、具体是存在于ABCA4基因中的一种或多种致病性G至A或C至T突变/SNP、更具体是上述一种或多种致病性G至A或C至T突变/SNP来治疗或预防斯塔加特病的方法。

[1328] Bardet-Biedl综合征

[1329] 在一些实施方案中，本文所述的方法、系统和组合物用于校正与Bardet-Biedl综合征相关的一种或多种致病性G至A或C至T突变/SNP。在一些实施方案中，致病性突变/SNP存在于选自BBS1、BBS2、BBS7、BBS9、BBS10、BBS12、LZTFL1和TRIM32的至少一个基因中，包括至少以下各项：

[1330] NM_024649.4(BBS1):c.416G>A(p.Trp139Ter)

[1331] NM_024649.4(BBS1):c.871C>T(p.Gln291Ter)

[1332] NM_198428.2(BBS9):c.263+1G>A

[1333] NM_001178007.1(BBS12):c.1704G>A(p.Trp568Ter)

[1334] NM_001276378.1(LZTFL1):c.271C>T(p.Arg91Ter)

[1335] NM_031885.3(BBS2):c.1864C>T(p.Arg622Ter)

[1336] NM_198428.2(BBS9):c.1759C>T(p.Arg587Ter)

[1337] NM_198428.2(BBS9):c.1789+1G>A

[1338] NM_024649.4(BBS1):c.432+1G>A

[1339] NM_176824.2(BBS7):c.632C>T(p.Thr211Ile)

[1340] NM_012210.3(TRIM32):c.388C>T(p.Pro130Ser)

[1341] NM_031885.3(BBS2):c.823C>T(p.Arg275Ter)

[1342] NM_024685.3(BBS10):c.145C>T(p.Arg49Trp)

[1343] 参见表A。因此，本发明的一个方面涉及一种用于通过校正一种或多种致病性G至A或C至T突变/SNP、具体是存在于选自BBS1、BBS2、BBS7、BBS9、BBS10、BBS12、LZTFL1和TRIM32的至少一个基因中的一种或多种致病性G至A或C至T突变/SNP、更具体是上述一种或多种致病性G至A或C至T突变/SNP来治疗或预防Bardet-Biedl综合征的方法。

[1344] 视锥视杆细胞营养不良

[1345] 在一些实施方案中，本文所述的方法、系统和组合物用于校正与视锥视杆细胞营养不良 相关的一种或多种致病性G至A或C至T突变/SNP。在一些实施方案中，致病性突变/SNP存在于选自RPGRIP1、DRAM2、ABCA4、ADAM9和CACNA1F的至少一个基因中，包括至少以下各项：

[1346] NM_020366.3(RPGRIP1):c.154C>T(p.Arg52Ter)

[1347] NM_178454.5(DRAM2):c.494G>A(p.Trp165Ter)

[1348] NM_178454.5(DRAM2):c.131G>A(p.Ser44Asn)

[1349] NM_000350.2(ABCA4):c.161G>A(p.Cys54Tyr)

[1350] NM_000350.2(ABCA4):c.5714+5G>A

[1351] NM_000350.2(ABCA4):c.880C>T(p.Gln294Ter)

[1352] NM_000350.2(ABCA4):c.6079C>T(p.Leu2027Phe)

[1353] NM_000350.2(ABCA4):c.3113C>T(p.Ala1038Val)

[1354] NM_000350.2(ABCA4):c.634C>T(p.Arg212Cys)

[1355] NM_003816.2(ADAM9):c.490C>T(p.Arg164Ter)

[1356] NM_005183.3(CACNA1F):c.244C>T(p.Arg82Ter)

[1357] 参见表A。因此，本发明的一个方面涉及一种用于通过校正一种或多种致病性G至A或C至T突变/SNP、具体是存在于选自RPGRIP1、DRAM2、ABCA4、ADAM9和CACNA1F的至少一个基因中的一种或多种致病性G至A或C至T突变/SNP、更具体是上述一种或多种致病性G至A或C至T突变/SNP来治疗或预防视锥视杆细胞营养不良的方法。

[1358] 先天性静止性夜盲症

[1359] 在一些实施方案中，本文所述的方法、系统和组合物用于校正与先天性静止性夜盲症相关的一种或多种致病性G至A或C至T突变/SNP。在一些实施方案中，致病性突变/SNP存在于选自GRM6、TRPM1、GPR179和CACNA1F的至少一种基因中，包括至少以下各项：

[1360] NM_000843.3(GRM6):c.1462C>T(p.Gln488Ter)

[1361] NM_002420.5(TRPM1):c.2998C>T(p.Arg1000Ter)

[1362] NM_001004334.3(GPR179):c.673C>T(p.Gln225Ter)

[1363] NM_005183.3(CACNA1F):c.2576+1G>A

[1364] 参见表A。因此，本发明的一个方面涉及一种用于通过校正一种或多种致病性G至A或C至T突变/SNP、具体是存在于选自GRM6、TRPM1、GPR179和CACNA1F的至少一个基因中的一种或多种致病性G至A或C至T突变/SNP、更具体是上述一种或多种致病性G至A或C至T突变/SNP来治疗或预防先天性静止性夜盲症的方法。

[1365] 乌舍尔综合征

[1366] 在一些实施方案中，本文所述的方法、系统和组合物用于校正与乌舍尔综合征相

关的一种或多种致病性G至A或C至T突变/SNP。在一些实施方案中，致病性突变/SNP存在于选自MYO7A、USH1C、CDH23、PCDH15、USH2A、ADGRV1、WHRN和CLRN1的至少一个基因中，包括至少以下各项：

[1367] NM_000260.3(MYO7A):c.640G>A(p.Gly214Arg)

[1368] NM_000260.3(MYO7A):c.1200+1G>A

[1369] NM_000260.3(MYO7A):c.141G>A(p.Trp47Ter)

[1370] NM_000260.3(MYO7A):c.1556G>A(p.Gly519Asp)

[1371] NM_000260.3(MYO7A):c.1900C>T(p.Arg634Ter)

[1372] NM_000260.3(MYO7A):c.1963C>T(p.Gln655Ter)

[1373] NM_000260.3(MYO7A):c.2094+1G>A

[1374] NM_000260.3(MYO7A):c.4293G>A(p.Trp1431Ter)

[1375] NM_000260.3(MYO7A):c.5101C>T(p.Arg1701Ter)

[1376] NM_000260.3(MYO7A):c.5617C>T(p.Arg1873Trp)

[1377] NM_000260.3(MYO7A):c.5660C>T(p.Pro1887Leu)

[1378] NM_000260.3(MYO7A):c.6070C>T(p.Arg2024Ter)

[1379] NM_000260.3(MYO7A):c.470+1G>A

[1380] NM_000260.3(MYO7A):c.5968C>T(p.Gln1990Ter)

[1381] NM_000260.3(MYO7A):c.3719G>A(p.Arg1240Gln)

[1382] NM_000260.3(MYO7A):c.494C>T(p.Thr165Met)

[1383] NM_000260.3(MYO7A):c.5392C>T(p.Gln1798Ter)

[1384] NM_000260.3(MYO7A):c.5648G>A(p.Arg1883Gln)

[1385] NM_000260.3(MYO7A):c.448C>T(p.Arg150Ter)

[1386] NM_000260.3(MYO7A):c.700C>T(p.Gln234Ter)

[1387] NM_000260.3(MYO7A):c.635G>A(p.Arg212His)

[1388] NM_000260.3(MYO7A):c.1996C>T(p.Arg666Ter)

[1389] NM_005709.3(USH1C):c.216G>A(p.Val72=)

[1390] NM_022124.5(CDH23):c.7362+5G>A

[1391] NM_022124.5(CDH23):c.3481C>T(p.Arg1161Ter)

[1392] NM_022124.5(CDH23):c.3628C>T(p.Gln1210Ter)

[1393] NM_022124.5(CDH23):c.5272C>T(p.Gln1758Ter)

[1394] NM_022124.5(CDH23):c.5712+1G>A

[1395] NM_022124.5(CDH23):c.5712G>A(p.Thr1904=)

[1396] NM_022124.5(CDH23):c.5923+1G>A

[1397] NM_022124.5(CDH23):c.6049+1G>A

[1398] NM_022124.5(CDH23):c.7776G>A(p.Trp2592Ter)

[1399] NM_022124.5(CDH23):c.9556C>T(p.Arg3186Ter)

[1400] NM_022124.5(CDH23):c.3706C>T(p.Arg1236Ter)

[1401] NM_022124.5(CDH23):c.4309C>T(p.Arg1437Ter)

[1402] NM_022124.5(CDH23):c.6050-9G>A

[1403] NM_033056.3(PCDH15):c.3316C>T(p.Arg1106Ter)

[1404] NM_033056.3(PCDH15):c.7C>T(p.Arg3Ter)

[1405] NM_033056.3(PCDH15):c.1927C>T(p.Arg643Ter)

[1406] NM_001142772.1(PCDH15):c.400C>T(p.Arg134Ter)

[1407] NM_033056.3(PCDH15):c.3358C>T(p.Arg1120Ter)

[1408] NM_206933.2(USH2A):c.11048-1G>A

[1409] NM_206933.2(USH2A):c.1143+1G>A

[1410] NM_206933.2(USH2A):c.11954G>A(p.Trp3985Ter)

[1411] NM_206933.2(USH2A):c.12868C>T(p.Gln4290Ter)

[1412] NM_206933.2(USH2A):c.14180G>A(p.Trp4727Ter)

[1413] NM_206933.2(USH2A):c.14911C>T(p.Arg4971Ter)

[1414] NM_206933.2(USH2A):c.5788C>T(p.Arg1930Ter)

[1415] NM_206933.2(USH2A):c.5858-1G>A

[1416] NM_206933.2(USH2A):c.6224G>A(p.Trp2075Ter)

[1417] NM_206933.2(USH2A):c.820C>T(p.Arg274Ter)

[1418] NM_206933.2(USH2A):c.8981G>A(p.Trp2994Ter)

[1419] NM_206933.2(USH2A):c.9304C>T(p.Gln3102Ter)

[1420] NM_206933.2(USH2A):c.13010C>T(p.Thr4337Met)

[1421] NM_206933.2(USH2A):c.14248C>T(p.Gln4750Ter)

[1422] NM_206933.2(USH2A):c.6398G>A(p.Trp2133Ter)

[1423] NM_206933.2(USH2A):c.632G>A(p.Trp211Ter)

[1424] NM_206933.2(USH2A):c.6601C>T(p.Gln2201Ter)

[1425] NM_206933.2(USH2A):c.13316C>T(p.Thr4439Ile)

[1426] NM_206933.2(USH2A):c.4405C>T(p.Gln1469Ter)

[1427] NM_206933.2(USH2A):c.9570+1G>A

[1428] NM_206933.2(USH2A):c.8740C>T(p.Arg2914Ter)

[1429] NM_206933.2(USH2A):c.8681+1G>A

[1430] NM_206933.2(USH2A):c.1000C>T(p.Arg334Trp)

[1431] NM_206933.2(USH2A):c.14175G>A(p.Trp4725Ter)

[1432] NM_206933.2(USH2A):c.9390G>A(p.Trp3130Ter)

[1433] NM_206933.2(USH2A):c.908G>A(p.Arg303His)

[1434] NM_206933.2(USH2A):c.5776+1G>A

[1435] NM_206933.2(USH2A):c.11156G>A(p.Arg3719His)

[1436] NM_032119.3(ADGRV1):c.2398C>T(p.Arg800Ter)

[1437] NM_032119.3(ADGRV1):c.7406G>A(p.Trp2469Ter)

[1438] NM_032119.3(ADGRV1):c.12631C>T(p.Arg4211Ter)

[1439] NM_032119.3(ADGRV1):c.7129C>T(p.Arg2377Ter)

[1440] NM_032119.3(ADGRV1):c.14885G>A(p.Trp4962Ter)

[1441] NM_015404.3(WHRN):c.1267C>T(p.Arg423Ter)

[1442] NM_174878.2(CLRN1):c.619C>T(p.Arg207Ter)

[1443] 参见表A。因此，本发明的一个方面涉及一种用于通过校正一种或多种致病性G至A或C至T突变/SNP、具体是存在于选自MYO7A、USH1C、CDH23、PCDH15、USH2A、ADGRV1、WHRN和CLRN1的至少一个基因中的一种或多种致病性G至A或C至T突变/SNP、更具体是上述一种或多种致病性G至A或C至T突变/SNP来治疗或预防增强乌舍尔综合征的方法。

[1444] Leber先天性黑蒙症

[1445] 在一些实施方案中，本文所述的方法、系统和组合物用于校正与Leber先天性黑蒙症相关的一种或多种致病性G至A或C至T突变/SNP。在一些实施方案中，致病性突变/SNP存在于选自TULP1、RPE65、SPATA7、AIPL1、CRB1、NMNAT1和PEX1的至少一个基因中，包括至少以下各项：

[1446] NM_003322.5(TULP1):c.1495+1G>A

[1447] NM_000329.2(RPE65):c.11+5G>A

[1448] NM_018418.4(SPATA7):c.322C>T(p.Arg108Ter)

[1449] NM_014336.4(AIPL1):c.784G>A(p.Gly262Ser)

[1450] NM_201253.2(CRB1):c.1576C>T(p.Arg526Ter)

[1451] NM_201253.2(CRB1):c.3307G>A(p.Gly1103Arg)

[1452] NM_201253.2(CRB1):c.2843G>A(p.Cys948Tyr)

[1453] NM_022787.3(NMNAT1):c.769G>A(p.Glu257Lys)

[1454] NM_000466.2(PEX1):c.2528G>A(p.Gly843Asp)

[1455] 参见表A。因此，本发明的一个方面涉及一种用于通过校正一种或多种致病性G至A或C至T突变/SNP、具体是存在于选自选自TULP1、RPE65、SPATA7、AIPL1、CRB1、NMNAT1和PEX1的至少一个基因中的一种或多种致病性G至A或C至T突变/SNP、更具体是上述一种或多种致病性G至A或C至T突变/SNP来治疗或预防Leber先天性黑蒙症的方法。

[1456] 色素性视网膜炎

[1457] 在一些实施方案中，本文所述的方法、系统和组合物用于校正与色素性视网膜炎相关的一种或多种致病性G至A或C至T突变/SNP。在一些实施方案中，致病性突变/SNP存在于选自CRB1、IFT140、RP1、IMPDH1、PRPF31、RPGR、ABCA4、RPE65、EYS、NRL、FAM161A、NR2E3、USH2A、RHO、PDE6B、KLHL7、PDE6A、CNGB1、BEST1、C2orf71、PRPH2、CA4、CERKL、RPE65、PDE6B和ADGRV1的至少一个基因中，包括至少以下各项：

[1458] NM_001257965.1(CRB1):c.2711G>A(p.Cys904Tyr)

[1459] NM_014714.3(IFT140):c.3827G>A(p.Gly1276Glu)

[1460] NM_006269.1(RP1):c.2029C>T(p.Arg677Ter)

[1461] NM_000883.3(IMPDH1):c.931G>A(p.Asp311Asn)

[1462] NM_015629.3(PRPF31):c.1273C>T(p.Gln425Ter)

[1463] NM_015629.3(PRPF31):c.1073+1G>A

[1464] NM_000328.2(RPGR):c.1387C>T(p.Gln463Ter)

[1465] NM_000350.2(ABCA4):c.4577C>T(p.Thr1526Met)

[1466] NM_000350.2(ABCA4):c.6229C>T(p.Arg2077Trp)

[1467] NM_000329.2(RPE65):c.271C>T(p.Arg91Trp)

[1468] NM_001142800.1(EYS):c.2194C>T(p.Gln732Ter)

[1469] NM_001142800.1(EYS):c.490C>T(p.Arg164Ter)

[1470] NM_006177.3(NRL):c.151C>T(p.Pro51Ser)

[1471] NM_001201543.1(FAM161A):c.1567C>T(p.Arg523Ter)

[1472] NM_014249.3(NR2E3):c.166G>A(p.Gly56Arg)

[1473] NM_206933.2(USH2A):c.2209C>T(p.Arg737Ter)

[1474] NM_206933.2(USH2A):c.14803C>T(p.Arg4935Ter)

[1475] NM_206933.2(USH2A):c.10073G>A(p.Cys3358Tyr)

[1476] NM_000539.3(RHO):c.541G>A(p.Glu181Lys)

[1477] NM_000283.3(PDE6B):c.892C>T(p.Gln298Ter)

[1478] NM_001031710.2(KLHL7):c.458C>T(p.Ala153Val)

[1479] NM_000440.2(PDE6A):c.1926+1G>A

[1480] NM_001297.4(CNGB1):c.2128C>T(p.Gln710Ter)

[1481] NM_001297.4(CNGB1):c.952C>T(p.Gln318Ter)

[1482] NM_004183.3(BEST1):c.682G>A(p.Asp228Asn)

[1483] NM_001029883.2(C2orf71):c.1828C>T(p.Gln610Ter)

[1484] NM_000322.4(PRPH2):c.647C>T(p.Pro216Leu)

[1485] NM_000717.4(CA4):c.40C>T(p.Arg14Trp)

[1486] NM_201548.4(CERKL):c.769C>T(p.Arg257Ter)

[1487] NM_000329.2(RPE65):c.118G>A(p.Gly40Ser)

[1488] NM_000322.4(PRPH2):c.499G>A(p.Gly167Ser)

[1489] NM_000539.3(RHO):c.403C>T(p.Arg135Trp)

[1490] NM_000283.3(PDE6B):c.2193+1G>A

[1491] NM_032119.3(ADGRV1):c.6901C>T(p.Gln2301Ter)

[1492] 参见表A。因此，本发明的一个方面涉及一种用于通过校正一种或多种致病性G至A或C至T突变/SNP、具体是存在于选自CRB1、IFT140、RP1、IMPDH1、PRPF31、RPGR、ABCA4、RPE65、EYS、NRL、FAM161A、NR2E3、USH2A、RHO、PDE6B、KLHL7、PDE6A、CNGB1、BEST1、C2orf71、PRPH2、CA4、CERKL、RPE65、PDE6B和ADGRV1的至少一个基因中的一种或多种致病性G至A或C至T突变/SNP、更具体是上述一种或多种致病性G至A或C至T突变/SNP来治疗或预防色素性视网膜炎的方法。

[1493] 全色盲

[1494] 在一些实施方案中，本文所述的方法、系统和组合物用于校正与全色盲相关的一种或多种致病性G至A或C至T突变/SNP。在一些实施方案中，致病性突变/SNP存在于选自CNGA3、CNGB3和ATF6的至少一种基因中，包括至少以下各项：

[1495] NM_001298.2(CNGA3):c.847C>T(p.Arg283Trp)

[1496] NM_001298.2(CNGA3):c.101+1G>A

[1497] NM_001298.2(CNGA3):c.1585G>A(p.Val529Met)

[1498] NM_019098.4(CNGB3):c.1578+1G>A

[1499] NM_019098.4(CNGB3):c.607C>T(p.Arg203Ter)

[1500] NM_019098.4(CNGB3):c.1119G>A(p.Trp373Ter)

[1501] NM_007348.3(ATF6):c.970C>T(p.Arg324Cys)

[1502] 参见表A。因此,本发明的一个方面涉及一种用于通过校正一种或多种致病性G至A或C至T突变/SNP、具体是存在于选自CNGA3、CNGB3和ATF6的至少一个基因中的一种或多种致病性G至A或C至T突变/SNP、更具体是上述一种或多种致病性G至A或C至T突变/SNP来治疗或预防全色盲的方法。

[1503] 影响听力的疾病

[1504] 与各种影响听力的疾病相关的致病性G至A或C至T突变/SNP在ClinVar数据库中报告并在表A中公开,包括但不限于耳聋和非综合征性听力损失。因此,本发明的一个方面涉及一种用于校正与这些疾病中的任何一种相关的一种或多种致病性G至A或C至T突变/SNP的方法,如下所述。

[1505] 耳聋

[1506] 在一些实施方案中,本文所述的方法、系统和组合物用于校正与耳聋相关的一种或多种致病性G至A或C至T突变/SNP。在一些实施方案中,致病性突变/SNP存在于选自FGF3、MYO7A、STRC、ACTG1、SLC17A8、TMC1、GJB2、MYH14、COCH、CDH23、USH1C、GJB2、MYO7A、PCDH15、MYO15A、MYO3A、WHRN、DFNB59、TMC1、LOXHD1、TMPRSS3、OTOGL、OTOF、JAG1和MARVELD2的至少一个基因中,包括至少以下各项:

[1507] NM_005247.2(FGF3):c.283C>T(p.Arg95Trp)

[1508] NM_000260.3(MYO7A):c.652G>A(p.Asp218Asn)

[1509] NM_000260.3(MYO7A):c.689C>T(p.Ala230Val)

[1510] NM_153700.2(STRC):c.4057C>T(p.Gln1353Ter)

[1511] NM_001614.3(ACTG1):c.721G>A(p.Glu241Lys)

[1512] NM_139319.2(SLC17A8):c.632C>T(p.Ala211Val)

[1513] NM_138691.2(TMC1):c.1714G>A(p.Asp572Asn)

[1514] NM_004004.5(GJB2):c.598G>A(p.Gly200Arg)

[1515] NM_004004.5(GJB2):c.71G>A(p.Trp24Ter)

[1516] NM_004004.5(GJB2):c.416G>A(p.Ser139Asn)

[1517] NM_004004.5(GJB2):c.224G>A(p.Arg75Gln)

[1518] NM_004004.5(GJB2):c.95G>A(p.Arg32His)

[1519] NM_004004.5(GJB2):c.250G>A(p.Val84Met)

[1520] NM_004004.5(GJB2):c.428G>A(p.Arg143Gln)

[1521] NM_004004.5(GJB2):c.551G>A(p.Arg184Gln)

[1522] NM_004004.5(GJB2):c.223C>T(p.Arg75Trp)

[1523] NM_024729.3(MYH14):c.359C>T(p.Ser120Leu)

[1524] NM_004086.2(COCH):c.151C>T(p.Pro51Ser)

[1525] NM_022124.5(CDH23):c.4021G>A(p.Asp1341Asn)

[1526] NM_153700.2(STRC):c.4701+1G>A

[1527] NM_153676.3(USH1C):c.496+1G>A

[1528] NM_004004.5(GJB2):c.131G>A(p.Trp44Ter)

[1529] NM_004004.5(GJB2):c.283G>A(p.Val95Met)

[1530] NM_004004.5(GJB2):c.298C>T(p.His100Tyr)

[1531] NM_004004.5(GJB2):c.427C>T(p.Arg143Trp)

[1532] NM_004004.5(GJB2):c.109G>A(p.Val37Ile)

[1533] NM_004004.5(GJB2):c.-23+1G>A

[1534] NM_004004.5(GJB2):c.148G>A(p.Asp50Asn)

[1535] NM_004004.5(GJB2):c.134G>A(p.Gly45Glu)

[1536] NM_004004.5(GJB2):c.370C>T(p.Gln124Ter)

[1537] NM_004004.5(GJB2):c.230G>A(p.Trp77Ter)

[1538] NM_004004.5(GJB2):c.231G>A(p.Trp77Ter)

[1539] NM_000260.3(MYO7A):c.5899C>T(p.Arg1967Ter)

[1540] NM_000260.3(MYO7A):c.2005C>T(p.Arg669Ter)

[1541] NM_033056.3(PCDH15):c.733C>T(p.Arg245Ter)

[1542] NM_016239.3(MYO15A):c.3866+1G>A

[1543] NM_016239.3(MYO15A):c.6178-1G>A

[1544] NM_016239.3(MYO15A):c.8714-1G>A

[1545] NM_017433.4(MYO3A):c.2506-1G>A

[1546] NM_015404.3(WHRN):c.1417-1G>A

[1547] NM_001042702.3(DFNB59):c.499C>T(p.Arg167Ter)

[1548] NM_138691.2(TMC1):c.100C>T(p.Arg34Ter)

[1549] NM_138691.2(TMC1):c.1165C>T(p.Arg389Ter)

[1550] NM_144612.6(LOXHD1):c.2008C>T(p.Arg670Ter)

[1551] NM_144612.6(LOXHD1):c.4714C>T(p.Arg1572Ter)

[1552] NM_144612.6(LOXHD1):c.4480C>T(p.Arg1494Ter)

[1553] NM_024022.2(TMPRSS3):c.325C>T(p.Arg109Trp)

[1554] NM_173591.3(OTOGL):c.3076C>T(p.Gln1026Ter)

[1555] NM_194248.2(OTOF):c.4483C>T(p.Arg1495Ter)

[1556] NM_194248.2(OTOF):c.2122C>T(p.Arg708Ter)

[1557] NM_194248.2(OTOF):c.2485C>T(p.Gln829Ter)

[1558] NM_001038603.2(MARVELD2):c.1498C>T(p.Arg500Ter)

[1559] 参见表A。因此，本发明的一个方面涉及一种用于通过校正一种或多种致病性G至A或C至T突变/SNP、具体是存在于选自FGF3、MYO7A、STRC、ACTG1、SLC17A8、TMC1、GJB2、MYH14、COCH、CDH23、USH1C、GJB2、MYO7A、PCDH15、MYO15A、MYO3A、WHRN、DFNB59、TMC1、LOXHD1、TMPRSS3、OTOGL、OTOF、JAG1和MARVELD2的至少一个基因中的一种或多种致病性G至A或C至T突变/SNP、更具体是上述一种或多种致病性G至A或C至T突变/SNP来治疗或预防耳聋的方法。

[1560] 非综合征性听力损失

[1561] 在一些实施方案中，本文所述的方法、系统和组合物用于校正与非综合征性听力损失相关的一种或多种致病性G至A或C至T突变/SNP。在一些实施方案中，致病性突变/SNP

存在于选自GJB2、POU3F4、MYO15A、TMPRSS3、LOXHD1、OTOF、MYO6、OTOA、STRC、TRIOBP、MARVELD2、TMC1、TECTA、OTOGL和GIPC3的至少一个基因中，包括至少以下各项：

[1562] NM_004004.5(GJB2):c.169C>T(p.Gln57Ter)

[1563] NM_000307.4(POU3F4):c.499C>T(p.Arg167Ter)

[1564] NM_016239.3(MYO15A):c.8767C>T(p.Arg2923Ter)

[1565] NM_024022.2(TMPRSS3):c.323-6G>A

[1566] NM_024022.2(TMPRSS3):c.916G>A(p.Ala306Thr)

[1567] NM_144612.6(LOXHD1):c.2497C>T(p.Arg833Ter)

[1568] NM_194248.2(OTOF):c.2153G>A(p.Trp718Ter)

[1569] NM_194248.2(OTOF):c.2818C>T(p.Gln940Ter)

[1570] NM_194248.2(OTOF):c.4799+1G>A

[1571] NM_004999.3(MYO6):c.826C>T(p.Arg276Ter)

[1572] NM_144672.3(OTOA):c.1880+1G>A

[1573] NM_153700.2(STRC):c.5188C>T(p.Arg1730Ter)

[1574] NM_153700.2(STRC):c.3670C>T(p.Arg1224Ter)

[1575] NM_153700.2(STRC):c.4402C>T(p.Arg1468Ter)

[1576] NM_024022.2(TMPRSS3):c.1192C>T(p.Gln398Ter)

[1577] NM_001039141.2(TRIOBP):c.6598C>T(p.Arg2200Ter)

[1578] NM_016239.3(MYO15A):c.7893+1G>A

[1579] NM_016239.3(MYO15A):c.5531+1G>A

[1580] NM_016239.3(MYO15A):c.6046+1G>A

[1581] NM_144612.6(LOXHD1):c.3169C>T(p.Arg1057Ter)

[1582] NM_001038603.2(MARVELD2):c.1331+1G>A

[1583] NM_138691.2(TMC1):c.1676G>A(p.Trp559Ter)

[1584] NM_138691.2(TMC1):c.1677G>A(p.Trp559Ter)

[1585] NM_005422.2(TECTA):c.5977C>T(p.Arg1993Ter)

[1586] NM_173591.3(OTOGL):c.4987C>T(p.Arg1663Ter)

[1587] NM_153700.2(STRC):c.3493C>T(p.Gln1165Ter)

[1588] NM_153700.2(STRC):c.3217C>T(p.Arg1073Ter)

[1589] NM_016239.3(MYO15A):c.5896C>T(p.Arg1966Ter)

[1590] NM_133261.2(GIPC3):c.411+1G>A

[1591] 参见表A。因此，本发明的一个方面涉及一种用于通过校正一种或多种致病性G至A或C至T突变/SNP、具体是存在于选自GJB2、POU3F4、MYO15A、TMPRSS3、LOXHD1、OTOF、MYO6、OTOA、STRC、TRIOBP、MARVELD2、TMC1、TECTA、OTOGL和GIPC3的至少一个基因中的一种或多种致病性G至A或C至T突变/SNP、更具体是上述一种或多种致病性G至A或C至T突变/SNP来治疗或预防非综合征性听力损失的方法。

[1592] 血液疾病

[1593] 与各种血液疾病相关的致病性G至A或C至T突变/SNP在ClinVar数据库中报告并在表A中公开，包括但不限于β地中海贫血、血友病A、血友病B、血友病C和Wiskott-Aldrich综

合征。因此，本发明的一个方面涉及一种用于校正与这些疾病中的任何一种相关的一种或多种致病性G至A或C至T突变/SNP的方法，如下所述。

[1594] β地中海贫血

[1595] 在一些实施方案中，本文所述的方法、系统和组合物用于校正与β地中海贫血相关的一种或多种致病性G至A或C至T突变/SNP。在一些实施方案中，致病性突变/SNP存在于至少HBB基因中，包括至少以下各项：

[1596] NM_000518.4(HBB):c.-137C>T

[1597] NM_000518.4(HBB):c.-50-88C>T

[1598] NM_000518.4(HBB):c.-140C>T

[1599] NM_000518.4(HBB):c.316-197C>T

[1600] NM_000518.4(HBB):c.93-21G>A

[1601] NM_000518.4(HBB):c.114G>A(p.Trp38Ter)

[1602] NM_000518.4(HBB):c.118C>T(p.Gln40Ter)

[1603] NM_000518.4(HBB):c.92+1G>A

[1604] NM_000518.4(HBB):c.315+1G>A

[1605] NM_000518.4(HBB):c.92+5G>A

[1606] NM_000518.4(HBB):c.-50-101C>T

[1607] 参见表A。因此，本发明的一个方面涉及一种用于通过校正一种或多种致病性G至A或C至T突变/SNP、具体是存在于HBB基因中的一种或多种致病性G至A或C至T突变/SNP、更具体是上述一种或多种致病性G至A或C至T突变/SNP来治疗或预防β地中海贫血的方法。

[1608] 血友病A

[1609] 在一些实施方案中，本文所述的方法、系统和组合物用于校正与血友病A相关的一种或多种致病性G至A或C至T突变/SNP。在一些实施方案中，致病性突变/SNP存在于至少F8基因中，包括至少以下各项：

[1610] NM_000132.3(F8):c.3169G>A(p.Glu1057Lys)

[1611] NM_000132.3(F8):c.902G>A(p.Arg301His)

[1612] NM_000132.3(F8):c.1834C>T(p.Arg612Cys)

[1613] 参见表A。因此，本发明的一个方面涉及一种用于通过校正一种或多种致病性G至A或C至T突变/SNP、具体是存在于F8基因中的一种或多种致病性G至A或C至T突变/SNP、更具体是上述一种或多种致病性G至A或C至T突变/SNP来治疗或预防血友病A的方法。

[1614] 因子V莱顿

[1615] 在一些实施方案中，本文所述的方法、系统和组合物用于校正因子V莱顿。这种引起疾病的单点突变(G1746->A)代表了高加索人群遗传性多因子血栓形成中最丰富的遗传危险因素。由于点突变，在血凝因子F5的蛋白C依赖性蛋白水解裂解位点(R533R534)处出现单个氨基酸取代(R534->Q)。杂合缺陷伴随仅有轻微增加的血栓形成风险(约8倍)，而纯合缺陷具有更显著的作用(风险增加＞80倍)。19指导的RNA编辑具有通过其在RNA水平修复来补偿该遗传缺陷的潜力。

[1616] 血友病B

[1617] 在一些实施方案中，本文所述的方法、系统和组合物用于校正与血友病B相关的一

种或多种致病性G至A或C至T突变/SNP。在一些实施方案中，致病性突变/SNP存在于至少F9基因中，包括至少以下各项：

[1618] NM_000133.3(F9):c.835G>A(p.Ala279Thr)

[1619] 参见表A。因此，本发明的一个方面涉及一种用于通过校正一种或多种致病性G至A或C至T突变/SNP、具体是存在于F9基因中的一种或多种致病性G至A或C至T突变/SNP、更具体是上述一种或多种致病性G至A或C至T突变/SNP来治疗或预防血友病B的方法。

[1620] 血友病C

[1621] 在一些实施方案中，本文所述的方法、系统和组合物用于校正与血友病C相关的一种或多种致病性G至A或C至T突变/SNP。在一些实施方案中，致病性突变/SNP存在于至少F11基因中，包括至少以下各项：

[1622] NM_000128.3(F11):c.400C>T(p.Gln134Ter)

[1623] NM_000128.3(F11):c.1432G>A(p.Gly478Arg)

[1624] NM_000128.3(F11):c.1288G>A(p.Ala430Thr)

[1625] NM_000128.3(F11):c.326-1G>A

[1626] 参见表A。因此，本发明的一个方面涉及一种用于通过校正一种或多种致病性G至A或C至T突变/SNP、具体是存在于F11基因中的一种或多种致病性G至A或C至T突变/SNP、更具体是上述一种或多种致病性G至A或C至T突变/SNP来治疗或预防血友病C的方法。

[1627] Wiskott-Aldrich综合征

[1628] 在一些实施方案中，本文所述的方法、系统和组合物用于校正与Wiskott-Aldrich综合征相关的一种或多种致病性G至A或C至T突变/SNP。在一些实施方案中，致病性突变/SNP存在于至少WAS基因中，包括至少以下各项：

[1629] NM_000377.2(WAS):c.37C>T(p.Arg13Ter)

[1630] NM_000377.2(WAS):c.257G>A(p.Arg86His)

[1631] NM_000377.2(WAS):c.777+1G>A

[1632] 参见表A。因此，本发明的一个方面涉及一种用于通过校正一种或多种致病性G至A或C至T突变/SNP、具体是存在于WAS基因中的一种或多种致病性G至A或C至T突变/SNP、更具体是上述一种或多种致病性G至A或C至T突变/SNP来治疗或预防Wiskott-Aldrich综合征的方法。

[1633] 肝脏疾病

[1634] 与各种肝脏疾病相关的致病性G至A或C至T突变/SNP在ClinVar数据库中报告并在表A中公开，包括但不限于运甲状腺素蛋白淀粉样变性、α-1抗胰蛋白酶缺乏症、Wilson病和苯丙酮尿症。因此，本发明的一个方面涉及一种用于校正与这些疾病中的任何一种相关的一种或多种致病性G至A或C至T突变/SNP的方法，如下所述。

[1635] 运甲状腺素蛋白淀粉样变性

[1636] 在一些实施方案中，本文所述的方法、系统和组合物用于校正与运甲状腺素蛋白淀粉样变性相关的一种或多种致病性G至A或C至T突变/SNP。在一些实施方案中，致病性突变/SNP存在于至少TTR基因中，包括至少以下各项：

[1637] NM_000371.3(TTR):c.424G>A(p.Val142Ile)

[1638] NM_000371.3(TTR):c.148G>A(p.Val50Met)

[1639] NM_000371.3(TTR):c.118G>A(p.Val40Ile)

[1640] 参见表A。因此，本发明的一个方面涉及一种用于通过校正一种或多种致病性G至A或C至T突变/SNP、具体是存在于TTR基因中的一种或多种致病性G至A或C至T突变/SNP、更具体是上述一种或多种致病性G至A或C至T突变/SNP来治疗或预防运甲状腺素蛋白淀粉样变性的方法。

[1641] α-1抗胰蛋白酶缺乏症

[1642] 在一些实施方案中，本文所述的方法、系统和组合物用于校正与α-1抗胰蛋白酶缺乏症相关的一种或多种致病性G至A或C至T突变/SNP。在一些实施方案中，致病性突变/SNP存在于至少SERPINA1基因中，包括至少以下各项：

[1643] NM_000295.4(SERPINA1):c.538C>T(p.Gln180Ter)

[1644] NM_001127701.1(SERPINA1):c.1178C>T(p.Pro393Leu)

[1645] NM_001127701.1(SERPINA1):c.230C>T(p.Ser77Phe)

[1646] NM_001127701.1(SERPINA1):c.1096G>A(p.Glu366Lys)

[1647] NM_000295.4(SERPINA1):c.1177C>T(p.Pro393Ser)

[1648] 参见表A。因此，本发明的一个方面涉及一种用于通过校正一种或多种致病性G至A或C至T突变/SNP、具体是存在于SERPINA1基因中的一种或多种致病性G至A或C至T突变/SNP、更具体是上述一种或多种致病性G至A或C至T突变/SNP来治疗或预防α-1抗胰蛋白酶缺乏症的方法。

[1649] Wilson病

[1650] 在一些实施方案中，本文所述的方法、系统和组合物用于校正与Wilson病相关的一种或多种致病性G至A或C至T突变/SNP。在一些实施方案中，致病性突变/SNP存在于至少ATP7B基因中，包括至少以下各项：

[1651] NM_000053.3(ATP7B):c.2293G>A(p.Asp765Asn)

[1652] NM_000053.3(ATP7B):c.3955C>T(p.Arg1319Ter)

[1653] NM_000053.3(ATP7B):c.2865+1G>A

[1654] NM_000053.3(ATP7B):c.3796G>A(p.Gly1266Arg)

[1655] NM_000053.3(ATP7B):c.2621C>T(p.Ala874Val)

[1656] NM_000053.3(ATP7B):c.2071G>A(p.Gly691Arg)

[1657] NM_000053.3(ATP7B):c.2128G>A(p.Gly710Ser)

[1658] NM_000053.3(ATP7B):c.2336G>A(p.Trp779Ter)

[1659] NM_000053.3(ATP7B):c.4021G>A(p.Gly1341Ser)

[1660] NM_000053.3(ATP7B):c.3182G>A(p.Gly1061Glu)

[1661] NM_000053.3(ATP7B):c.4114C>T(p.Gln1372Ter)

[1662] NM_000053.3(ATP7B):c.1708-1G>A

[1663] NM_000053.3(ATP7B):c.865C>T(p.Gln289Ter)

[1664] NM_000053.3(ATP7B):c.2930C>T(p.Thr977Met)

[1665] NM_000053.3(ATP7B):c.3659C>T(p.Thr1220Met)

[1666] NM_000053.3(ATP7B):c.2605G>A(p.Gly869Arg)

[1667] NM_000053.3(ATP7B):c.2975C>T(p.Pro992Leu)

[1668] NM_000053.3(ATP7B):c.2519C>T(p.Pro840Leu)

[1669] NM_000053.3(ATP7B):c.2906G>A(p.Arg969Gln)

[1670] 参见表A。因此,本发明的一个方面涉及一种用于通过校正一种或多种致病性G至A或C至T突变/SNP、具体是存在于ATP7B基因中的一种或多种致病性G至A或C至T突变/SNP、更具体是上述一种或多种致病性G至A或C至T突变/SNP来治疗或预防Wilson病的方法。

[1671] 苯丙酮尿症

[1672] 在一些实施方案中,本文所述的方法、系统和组合物用于校正与苯丙酮尿症相关的一种或多种致病性G至A或C至T突变/SNP。在一些实施方案中,致病性突变/SNP存在于至少PAH基因中,包括至少以下各项:

[1673] NM_000277.1(PAH):c.1315+1G>A

[1674] NM_000277.1(PAH):c.1222C>T(p.Arg408Trp)

[1675] NM_000277.1(PAH):c.838G>A(p.Glu280Lys)

[1676] NM_000277.1(PAH):c.331C>T(p.Arg111Ter)

[1677] NM_000277.1(PAH):c.782G>A(p.Arg261Gln)

[1678] NM_000277.1(PAH):c.754C>T(p.Arg252Trp)

[1679] NM_000277.1(PAH):c.473G>A(p.Arg158Gln)

[1680] NM_000277.1(PAH):c.727C>T(p.Arg243Ter)

[1681] NM_000277.1(PAH):c.842C>T(p.Pro281Leu)

[1682] NM_000277.1(PAH):c.728G>A(p.Arg243Gln)

[1683] NM_000277.1(PAH):c.1066-11G>A

[1684] NM_000277.1(PAH):c.781C>T(p.Arg261Ter)

[1685] NM_000277.1(PAH):c.1223G>A(p.Arg408Gln)

[1686] NM_000277.1(PAH):c.1162G>A(p.Val388Met)

[1687] NM_000277.1(PAH):c.1066-3C>T

[1688] NM_000277.1(PAH):c.1208C>T(p.Ala403Val)

[1689] NM_000277.1(PAH):c.890G>A(p.Arg297His)

[1690] NM_000277.1(PAH):c.926C>T(p.Ala309Val)

[1691] NM_000277.1(PAH):c.441+1G>A

[1692] NM_000277.1(PAH):c.526C>T(p.Arg176Ter)

[1693] NM_000277.1(PAH):c.688G>A(p.Val230Ile)

[1694] NM_000277.1(PAH):c.721C>T(p.Arg241Cys)

[1695] NM_000277.1(PAH):c.745C>T(p.Leu249Phe)

[1696] NM_000277.1(PAH):c.442-1G>A

[1697] NM_000277.1(PAH):c.842+1G>A

[1698] NM_000277.1(PAH):c.776C>T(p.Ala259Val)

[1699] NM_000277.1(PAH):c.1200-1G>A

[1700] NM_000277.1(PAH):c.912+1G>A

[1701] NM_000277.1(PAH):c.1065+1G>A

[1702] NM_000277.1(PAH):c.472C>T(p.Arg158Trp)

[1703] NM_000277.1(PAH):c.755G>A(p.Arg252Gln)

[1704] NM_000277.1(PAH):c.809G>A(p.Arg270Lys)

[1705] 参见表A。因此,本发明的一个方面涉及一种用于通过校正一种或多种致病性G至A或C至T突变/SNP、具体是存在于PAH基因中的一种或多种致病性G至A或C至T突变/SNP、更具体是上述一种或多种致病性G至A或C至T突变/SNP来治疗或预防苯丙酮尿症的方法。

[1706] 肾脏疾病

[1707] 与各种肾脏疾病相关的致病性G至A或C至T突变/SNP在ClinVar数据库中报告并在表A中公开,包括但不限于常染色体隐性多囊性肾脏疾病和肾肉碱转运缺陷。因此,本发明的一个方面涉及一种用于校正与这些疾病中的任何一种相关的一种或多种致病性G至A或C至T突变/SNP的方法,如下所述。

[1708] 常染色体隐性多囊性肾脏疾病

[1709] 在一些实施方案中,本文所述的方法、系统和组合物用于校正与常染色体隐性多囊性肾脏疾病相关的一种或多种致病性G至A或C至T突变/SNP。在一些实施方案中,致病性突变/SNP存在于至少PKHD1基因中,包括至少以下各项:

[1710] NM_138694.3(PKHD1):c.10444C>T(p.Arg3482Cys)

[1711] NM_138694.3(PKHD1):c.9319C>T(p.Arg3107Ter)

[1712] NM_138694.3(PKHD1):c.1480C>T(p.Arg494Ter)

[1713] NM_138694.3(PKHD1):c.707+1G>A

[1714] NM_138694.3(PKHD1):c.1486C>T(p.Arg496Ter)

[1715] NM_138694.3(PKHD1):c.8303-1G>A

[1716] NM_138694.3(PKHD1):c.2854G>A(p.Gly952Arg)

[1717] NM_138694.3(PKHD1):c.7194G>A(p.Trp2398Ter)

[1718] NM_138694.3(PKHD1):c.10219C>T(p.Gln3407Ter)

[1719] NM_138694.3(PKHD1):c.107C>T(p.Thr36Met)

[1720] NM_138694.3(PKHD1):c.8824C>T(p.Arg2942Ter)

[1721] NM_138694.3(PKHD1):c.982C>T(p.Arg328Ter)

[1722] NM_138694.3(PKHD1):c.4870C>T(p.Arg1624Trp)

[1723] NM_138694.3(PKHD1):c.1602+1G>A

[1724] NM_138694.3(PKHD1):c.1694-1G>A

[1725] NM_138694.3(PKHD1):c.2341C>T(p.Arg781Ter)

[1726] NM_138694.3(PKHD1):c.2407+1G>A

[1727] NM_138694.3(PKHD1):c.2452C>T(p.Gln818Ter)

[1728] NM_138694.3(PKHD1):c.5236+1G>A

[1729] NM_138694.3(PKHD1):c.6499C>T(p.Gln2167Ter)

[1730] NM_138694.3(PKHD1):c.2725C>T(p.Arg909Ter)

[1731] NM_138694.3(PKHD1):c.370C>T(p.Arg124Ter)

[1732] NM_138694.3(PKHD1):c.2810G>A(p.Trp937Ter)

[1733] 参见表A。因此,本发明的一个方面涉及一种用于通过校正一种或多种致病性G至A或C至T突变/SNP、具体是存在于PKHD1基因中的一种或多种致病性G至A或C至T突变/SNP、更

具体是上述一种或多种致病性G至A或C至T突变/SNP来治疗或预防常染色体隐性多囊性肾脏疾病的方法。

[1734] 肾肉碱转运缺陷

[1735] 在一些实施方案中，本文所述的方法、系统和组合物用于校正与肾肉碱转运缺陷相关的一种或多种致病性G至A或C至T突变/SNP。在一些实施方案中，致病性突变/SNP存在于至少SLC22A5基因中，包括至少以下各项：

[1736] NM_003060.3(SLC22A5):c.760C>T(p.Arg254Ter)

[1737] NM_003060.3(SLC22A5):c.396G>A(p.Trp132Ter)

[1738] NM_003060.3(SLC22A5):c.844C>T(p.Arg282Ter)

[1739] NM_003060.3(SLC22A5):c.505C>T(p.Arg169Trp)

[1740] NM_003060.3(SLC22A5):c.1319C>T(p.Thr440Met)

[1741] NM_003060.3(SLC22A5):c.1195C>T(p.Arg399Trp)

[1742] NM_003060.3(SLC22A5):c.695C>T(p.Thr232Met)

[1743] NM_003060.3(SLC22A5):c.845G>A(p.Arg282Gln)

[1744] NM_003060.3(SLC22A5):c.1193C>T(p.Pro398Leu)

[1745] NM_003060.3(SLC22A5):c.1463G>A(p.Arg488His)

[1746] NM_003060.3(SLC22A5):c.338G>A(p.Cys113Tyr)

[1747] NM_003060.3(SLC22A5):c.136C>T(p.Pro46Ser)

[1748] NM_003060.3(SLC22A5):c.506G>A(p.Arg169Gln)

[1749] 参见表A。因此，本发明的一个方面涉及一种用于通过校正一种或多种致病性G至A或C至T突变/SNP、具体是存在于SLC22A5基因中的一种或多种致病性G至A或C至T突变/SNP、更具体是上述一种或多种致病性G至A或C至T突变/SNP来治疗或预防肾肉碱转运缺陷的方法。

[1750] 肌肉疾病

[1751] 与各种肌肉疾病相关的致病性G至A或C至T突变/SNP在ClinVar数据库中报告并在表A中公开，包括但不限于杜兴氏肌营养不良、Becker肌营养不良、肢带型肌营养不良、Emery-Dreifuss肌营养不良和面肩肱型肌营养不良。因此，本发明的一个方面涉及一种用于校正与这些疾病中的任何一种相关的一种或多种致病性G至A或C至T突变/SNP的方法，如下所述。

[1752] 杜兴氏肌营养不良

[1753] 在一些实施方案中，本文所述的方法、系统和组合物用于校正与杜兴氏肌营养不良相关的一种或多种致病性G至A或C至T突变/SNP。在一些实施方案中，致病性突变/SNP存在于至少DMD基因中，包括至少以下各项：

[1754] NM_004006.2(DMD):c.2797C>T(p.Gln933Ter)

[1755] NM_004006.2(DMD):c.4870C>T(p.Gln1624Ter)

[1756] NM_004006.2(DMD):c.5551C>T(p.Gln1851Ter)

[1757] NM_004006.2(DMD):c.3188G>A(p.Trp1063Ter)

[1758] NM_004006.2(DMD):c.8357G>A(p.Trp2786Ter)

[1759] NM_004006.2(DMD):c.7817G>A(p.Trp2606Ter)

[1760] NM_004006.2(DMD):c.7755G>A(p.Trp2585Ter)

[1761] NM_004006.2(DMD):c.5917C>T(p.Gln1973Ter)

[1762] NM_004006.2(DMD):c.5641C>T(p.Gln1881Ter)

[1763] NM_004006.2(DMD):c.5131C>T(p.Gln1711Ter)

[1764] NM_004006.2(DMD):c.4240C>T(p.Gln1414Ter)

[1765] NM_004006.2(DMD):c.3427C>T(p.Gln1143Ter)

[1766] NM_004006.2(DMD):c.2407C>T(p.Gln803Ter)

[1767] NM_004006.2(DMD):c.2368C>T(p.Gln790Ter)

[1768] NM_004006.2(DMD):c.1683G>A(p.Trp561Ter)

[1769] NM_004006.2(DMD):c.1663C>T(p.Gln555Ter)

[1770] NM_004006.2(DMD):c.1388G>A(p.Trp463Ter)

[1771] NM_004006.2(DMD):c.1331+1G>A

[1772] NM_004006.2(DMD):c.1324C>T(p.Gln442Ter)

[1773] NM_004006.2(DMD):c.355C>T(p.Gln119Ter)

[1774] NM_004006.2(DMD):c.94-1G>A

[1775] NM_004006.2(DMD):c.5506C>T(p.Gln1836Ter)

[1776] NM_004006.2(DMD):c.1504C>T(p.Gln502Ter)

[1777] NM_004006.2(DMD):c.5032C>T(p.Gln1678Ter)

[1778] NM_004006.2(DMD):c.457C>T(p.Gln153Ter)

[1779] NM_004006.2(DMD):c.1594C>T(p.Gln532Ter)

[1780] NM_004006.2(DMD):c.1150-1G>A

[1781] NM_004006.2(DMD):c.6223C>T(p.Gln2075Ter)

[1782] NM_004006.2(DMD):c.3747G>A(p.Trp1249Ter)

[1783] NM_004006.2(DMD):c.2861G>A(p.Trp954Ter)

[1784] NM_004006.2(DMD):c.9563+1G>A

[1785] NM_004006.2(DMD):c.4483C>T(p.Gln1495Ter)

[1786] NM_004006.2(DMD):c.4312C>T(p.Gln1438Ter)

[1787] NM_004006.2(DMD):c.8209C>T(p.Gln2737Ter)

[1788] NM_004006.2(DMD):c.4071+1G>A

[1789] NM_004006.2(DMD):c.2665C>T(p.Arg889Ter)

[1790] NM_004006.2(DMD):c.2202G>A(p.Trp734Ter)

[1791] NM_004006.2(DMD):c.2077C>T(p.Gln693Ter)

[1792] NM_004006.2(DMD):c.1653G>A(p.Trp551Ter)

[1793] NM_004006.2(DMD):c.1061G>A(p.Trp354Ter)

[1794] NM_004006.2(DMD):c.8914C>T(p.Gln2972Ter)

[1795] NM_004006.2(DMD):c.6118-1G>A

[1796] NM_004006.2(DMD):c.4729C>T(p.Arg1577Ter)

[1797] 参见表A。因此，本发明的一个方面涉及一种用于通过校正一种或多种致病性G至A或C至T突变/SNP、具体是存在于DMD基因中的一种或多种致病性G至A或C至T突变/SNP、更具

体是上述一种或多种致病性G至A或C至T突变/SNP来治疗或预防杜兴氏肌营养不良的方法。

[1798] Becker肌营养不良

[1799] 在一些实施方案中，本文所述的方法、系统和组合物用于校正与Becker肌营养不良相关的一种或多种致病性G至A或C至T突变/SNP。在一些实施方案中，致病性突变/SNP存在于至少DMD基因中，包括至少以下各项：

[1800] NM_004006.2(DMD):c.3413G>A(p.Trp1138Ter)

[1801] NM_004006.2(DMD):c.358-1G>A

[1802] NM_004006.2(DMD):c.10108C>T(p.Arg3370Ter)

[1803] NM_004006.2(DMD):c.6373C>T(p.Gln2125Ter)

[1804] NM_004006.2(DMD):c.9568C>T(p.Arg3190Ter)

[1805] NM_004006.2(DMD):c.8713C>T(p.Arg2905Ter)

[1806] NM_004006.2(DMD):c.1615C>T(p.Arg539Ter)

[1807] NM_004006.2(DMD):c.3151C>T(p.Arg1051Ter)

[1808] NM_004006.2(DMD):c.3432+1G>A

[1809] NM_004006.2(DMD):c.5287C>T(p.Arg1763Ter)

[1810] NM_004006.2(DMD):c.5530C>T(p.Arg1844Ter)

[1811] NM_004006.2(DMD):c.8608C>T(p.Arg2870Ter)

[1812] NM_004006.2(DMD):c.8656C>T(p.Gln2886Ter)

[1813] NM_004006.2(DMD):c.8944C>T(p.Arg2982Ter)

[1814] NM_004006.2(DMD):c.5899C>T(p.Arg1967Ter)

[1815] NM_004006.2(DMD):c.10033C>T(p.Arg3345Ter)

[1816] NM_004006.2(DMD):c.10086+1G>A

[1817] NM_004019.2(DMD):c.1020G>A(p.Thr340=)

[1818] NM_004006.2(DMD):c.1261C>T(p.Gln421Ter)

[1819] NM_004006.2(DMD):c.1465C>T(p.Gln489Ter)

[1820] NM_004006.2(DMD):c.1990C>T(p.Gln664Ter)

[1821] NM_004006.2(DMD):c.2032C>T(p.Gln678Ter)

[1822] NM_004006.2(DMD):c.2332C>T(p.Gln778Ter)

[1823] NM_004006.2(DMD):c.2419C>T(p.Gln807Ter)

[1824] NM_004006.2(DMD):c.2650C>T(p.Gln884Ter)

[1825] NM_004006.2(DMD):c.2804-1G>A

[1826] NM_004006.2(DMD):c.3276+1G>A

[1827] NM_004006.2(DMD):c.3295C>T(p.Gln1099Ter)

[1828] NM_004006.2(DMD):c.336G>A(p.Trp112Ter)

[1829] NM_004006.2(DMD):c.3580C>T(p.Gln1194Ter)

[1830] NM_004006.2(DMD):c.4117C>T(p.Gln1373Ter)

[1831] NM_004006.2(DMD):c.649+1G>A

[1832] NM_004006.2(DMD):c.6906G>A(p.Trp2302Ter)

[1833] NM_004006.2(DMD):c.7189C>T(p.Gln2397Ter)

[1834] NM_004006.2(DMD):c.7309+1G>A

[1835] NM_004006.2(DMD):c.7657C>T(p.Arg2553Ter)

[1836] NM_004006.2(DMD):c.7682G>A(p.Trp2561Ter)

[1837] NM_004006.2(DMD):c.7683G>A(p.Trp2561Ter)

[1838] NM_004006.2(DMD):c.7894C>T(p.Gln2632Ter)

[1839] NM_004006.2(DMD):c.9361+1G>A

[1840] NM_004006.2(DMD):c.9564-1G>A

[1841] NM_004006.2(DMD):c.2956C>T(p.Gln986Ter)

[1842] NM_004006.2(DMD):c.883C>T(p.Arg295Ter)

[1843] NM_004006.2(DMD):c.31+36947G>A

[1844] NM_004006.2(DMD):c.10279C>T(p.Gln3427Ter)

[1845] NM_004006.2(DMD):c.433C>T(p.Arg145Ter)

[1846] NM_004006.2(DMD):c.9G>A(p.Trp3Ter)

[1847] NM_004006.2(DMD):c.10171C>T(p.Arg3391Ter)

[1848] NM_004006.2(DMD):c.583C>T(p.Arg195Ter)

[1849] NM_004006.2(DMD):c.9337C>T(p.Arg3113Ter)

[1850] NM_004006.2(DMD):c.8038C>T(p.Arg2680Ter)

[1851] NM_004006.2(DMD):c.1812+1G>A

[1852] NM_004006.2(DMD):c.1093C>T(p.Gln365Ter)

[1853] NM_004006.2(DMD):c.1704+1G>A

[1854] NM_004006.2(DMD):c.1912C>T(p.Gln638Ter)

[1855] NM_004006.2(DMD):c.133C>T(p.Gln45Ter)

[1856] NM_004006.2(DMD):c.5868G>A(p.Trp1956Ter)

[1857] NM_004006.2(DMD):c.565C>T(p.Gln189Ter)

[1858] NM_004006.2(DMD):c.5089C>T(p.Gln1697Ter)

[1859] NM_004006.2(DMD):c.2512C>T(p.Gln838Ter)

[1860] NM_004006.2(DMD):c.10477C>T(p.Gln3493Ter)

[1861] NM_004006.2(DMD):c.93+1G>A

[1862] NM_004006.2(DMD):c.4174C>T(p.Gln1392Ter)

[1863] NM_004006.2(DMD):c.3940C>T(p.Arg1314Ter)参见表A。因此，本发明的一个方面涉及一种用于通过校正一种或多种致病性G至A或C至T突变/SNP、具体是存在于DMD基因中的一种或多种致病性G至A或C至T突变/SNP、更具体是上述一种或多种致病性G至A或C至T突变/SNP来治疗或预防Becker肌营养不良的方法。

[1864] 肢带型肌营养不良

[1865] 在一些实施方案中，本文所述的方法、系统和组合物用于校正与肢带型肌营养不良相关的一种或多种致病性G至A或C至T突变/SNP。在一些实施方案中，致病性突变/SNP存在于选自SGCB、MYOT、LMNA、CAPN3、DYSF、SGCA、TTN、ANO5、TRAPPC11、LMNA、POMT1和FKRP的至少一个基因中，包括至少以下各项：

[1866] NM_000232.4(SGCB):c.31C>T(p.Gln11Ter)

[1867] NM_006790.2(MYOT):c.164C>T(p.Ser55Phe)

[1868] NM_006790.2(MYOT):c.170C>T(p.Thr57Ile)

[1869] NM_170707.3(LMNA):c.1488+1G>A

[1870] NM_170707.3(LMNA):c.1609-1G>A

[1871] NM_000070.2(CAPN3):c.1715G>A(p.Arg572Gln)

[1872] NM_000070.2(CAPN3):c.2243G>A(p.Arg748Gln)

[1873] NM_000070.2(CAPN3):c.145C>T(p.Arg49Cys)

[1874] NM_000070.2(CAPN3):c.1319G>A(p.Arg440Gln)

[1875] NM_000070.2(CAPN3):c.1343G>A(p.Arg448His)

[1876] NM_000070.2(CAPN3):c.1465C>T(p.Arg489Trp)

[1877] NM_000070.2(CAPN3):c.1714C>T(p.Arg572Trp)

[1878] NM_000070.2(CAPN3):c.2306G>A(p.Arg769Gln)

[1879] NM_000070.2(CAPN3):c.133G>A(p.Ala45Thr)

[1880] NM_000070.2(CAPN3):c.499-1G>A

[1881] NM_000070.2(CAPN3):c.439C>T(p.Arg147Ter)

[1882] NM_000070.2(CAPN3):c.1063C>T(p.Arg355Trp)

[1883] NM_000070.2(CAPN3):c.1250C>T(p.Thr417Met)

[1884] NM_000070.2(CAPN3):c.245C>T(p.Pro82Leu)

[1885] NM_000070.2(CAPN3):c.2242C>T(p.Arg748Ter)

[1886] NM_000070.2(CAPN3):c.1318C>T(p.Arg440Trp)

[1887] NM_000070.2(CAPN3):c.1333G>A(p.Gly445Arg)

[1888] NM_000070.2(CAPN3):c.1957C>T(p.Gln653Ter)

[1889] NM_000070.2(CAPN3):c.1801-1G>A

[1890] NM_000070.2(CAPN3):c.2263+1G>A

[1891] NM_000070.2(CAPN3):c.956C>T(p.Pro319Leu)

[1892] NM_000070.2(CAPN3):c.1468C>T(p.Arg490Trp)

[1893] NM_000070.2(CAPN3):c.802-9G>A

[1894] NM_000070.2(CAPN3):c.1342C>T(p.Arg448Cys)

[1895] NM_000070.2(CAPN3):c.1303G>A(p.Glu435Lys)

[1896] NM_000070.2(CAPN3):c.1993-1G>A

[1897] NM_003494.3(DYSF):c.3113G>A(p.Arg1038Gln)

[1898] NM_001130987.1(DYSF):c.5174+1G>A

[1899] NM_001130987.1(DYSF):c.159G>A(p.Trp53Ter)

[1900] NM_001130987.1(DYSF):c.2929C>T(p.Arg977Trp)

[1901] NM_001130987.1(DYSF):c.4282C>T(p.Gln1428Ter)

[1902] NM_001130987.1(DYSF):c.1577-1G>A

[1903] NM_003494.3(DYSF):c.5529G>A(p.Trp1843Ter)

[1904] NM_001130987.1(DYSF):c.1576+1G>A

[1905] NM_001130987.1(DYSF):c.4462C>T(p.Gln1488Ter)

[1906] NM_003494.3(DYSF):c.5429G>A(p.Arg1810Lys)

[1907] NM_003494.3(DYSF):c.5077C>T(p.Arg1693Trp)

[1908] NM_001130978.1(DYSF):c.1813C>T(p.Gln605Ter)

[1909] NM_003494.3(DYSF):c.3230G>A(p.Trp1077Ter)

[1910] NM_003494.3(DYSF):c.265C>T(p.Arg89Ter)

[1911] NM_003494.3(DYSF):c.4434G>A(p.Trp1478Ter)

[1912] NM_003494.3(DYSF):c.3478C>T(p.Gln1160Ter)

[1913] NM_001130987.1(DYSF):c.1372G>A(p.Gly458Arg)

[1914] NM_003494.3(DYSF):c.4090C>T(p.Gln1364Ter)

[1915] NM_001130987.1(DYSF):c.2409+1G>A

[1916] NM_003494.3(DYSF):c.1708C>T(p.Gln570Ter)

[1917] NM_003494.3(DYSF):c.1956G>A(p.Trp652Ter)

[1918] NM_001130987.1(DYSF):c.5004-1G>A

[1919] NM_003494.3(DYSF):c.331C>T(p.Gln111Ter)

[1920] NM_001130978.1(DYSF):c.5776C>T(p.Arg1926Ter)

[1921] NM_003494.3(DYSF):c.6124C>T(p.Arg2042Cys)

[1922] NM_003494.3(DYSF):c.2643+1G>A

[1923] NM_003494.3(DYSF):c.4253G>A(p.Gly1418Asp)

[1924] NM_003494.3(DYSF):c.610C>T(p.Arg204Ter)

[1925] NM_003494.3(DYSF):c.1834C>T(p.Gln612Ter)

[1926] NM_003494.3(DYSF):c.5668-7G>A

[1927] NM_001130978.1(DYSF):c.3137G>A(p.Arg1046His)

[1928] NM_003494.3(DYSF):c.1053+1G>A

[1929] NM_003494.3(DYSF):c.1398-1G>A

[1930] NM_0034943(DYSF):c1481-1G>A

[1931] NM_003494.3(DYSF):c.2311C>T(p.Gln771Ter)

[1932] NM_003494.3(DYSF):c.2869C>T(p.Gln957Ter)

[1933] NM_003494.3(DYSF):c.4756C>T(p.Arg1586Ter)

[1934] NM_003494.3(DYSF):c.5509G>A(p.Asp1837Asn)

[1935] NM_003494.3(DYSF):c.5644C>T(p.Gln1882Ter)

[1936] NM_003494.3(DYSF):c.5946+1G>A

[1937] NM_003494.3(DYSF):c.937+1G>A

[1938] NM_003494.3(DYSF):c.5266C>T(p.Gln1756Ter)

[1939] NM_003494.3(DYSF):c.3832C>T(p.Gln1278Ter)

[1940] NM_003494.3(DYSF):c.5525+1G>A

[1941] NM_003494.3(DYSF):c.3112C>T(p.Arg1038Ter)

[1942] NM_000023.3(SGCA):c.293G>A(p.Arg98His)

[1943] NM_000023.3(SGCA):c.850C>T(p.Arg284Cys)

[1944] NM_000023.3(SGCA):c.403C>T(p.Gln135Ter)

[1945] NM_000023.3(SGCA):c.409G>A(p.Glu137Lys)

[1946] NM_000023.3(SGCA):c.747+1G>A

[1947] NM_000023.3(SGCA):c.229C>T(p.Arg77Cys)

[1948] NM_000023.3(SGCA):c.101G>A(p.Arg34His)

[1949] NM_000023.3(SGCA):c.739G>A(p.Val247Met)

[1950] NM_001256850.1(TTN):c.87394C>T(p.Arg29132Ter)

[1951] NM_213599.2(ANO5):c.762+1G>A

[1952] NM_213599.2(ANO5):c.1213C>T(p.Gln405Ter)

[1953] NM_213599.2(ANO5):c.1639C>T(p.Arg547Ter)

[1954] NM_213599.2(ANO5):c.1406G>A(p.Trp469Ter)

[1955] NM_213599.2(ANO5):c.1210C>T(p.Arg404Ter)

[1956] NM_213599.2(ANO5):c.2272C>T(p.Arg758Cys)

[1957] NM_213599.2(ANO5):c.41-1G>A

[1958] NM_213599.2(ANO5):c.172C>T(p.Arg58Trp)

[1959] NM_213599.2(ANO5):c.1898+1G>A

[1960] NM_021942.5(TRAPPC11):c.1287+5G>A

[1961] NM_170707.3(LMNA):c.1608+1G>A

[1962] NM_007171.3(POMT1):c.1864C>T(p.Arg622Ter)

[1963] NM_024301.4(FKRP):c.313C>T(p.Gln105Ter)

[1964] 参见表A。因此，本发明的一个方面涉及一种用于通过校正一种或多种致病性G至A或C至T突变/SNP、具体是存在于选自SGCB、MYOT、LMNA、CAPN3、DYSF、SGCA、TTN、ANO5、TRAPPC11、LMNA、POMT1和FKRP的至少一个基因中的一种或多种致病性G至A或C至T突变/SNP、更具体是上述一种或多种致病性G至A或C至T突变/SNP来治疗或预防肢带型肌营养不良的方法。

[1965] Emery-Dreifuss肌营养不良

[1966] 在一些实施方案中，本文所述的方法、系统和组合物用于校正与Emery-Dreifuss肌营养不良相关的一种或多种致病性G至A或C至T突变/SNP。在一些实施方案中，致病性突变/SNP存在于至少EMD或SYNE1基因中，包括至少以下各项：

[1967] NM_000117.2(EMD):c.3G>A(p.Met1Ile)

[1968] NM_033071.3(SYNE1):c.11908C>T(p.Arg3970Ter)

[1969] NM_033071.3(SYNE1):c.21721C>T(p.Gln7241Ter)

[1970] NM_000117.2(EMD):c.130C>T(p.Gln44Ter)

[1971] 参见表A。因此，本发明的一个方面涉及一种用于通过校正一种或多种致病性G至A或C至T突变/SNP、具体是存在于EMD或SYNE1基因中的一种或多种致病性G至A或C至T突变/SNP、更具体是上述一种或多种致病性G至A或C至T突变/SNP来治疗或预防Emery-Dreifuss肌营养不良的方法。

[1972] 面肩肱型肌营养不良

[1973] 在一些实施方案中，本文所述的方法、系统和组合物用于校正与面肩肱型肌营养不良相关的一种或多种致病性G至A或C至T突变/SNP。在一些实施方案中，致病性突变/SNP

存在于至少SMCHD1基因中，包括至少以下各项：

[1974] NM_015295.2(SMCHD1):c.3801+1G>A

[1975] NM_015295.2(SMCHD1):c.1843-1G>A

[1976] 参见表A。因此，本发明的一个方面涉及一种用于通过校正一种或多种致病性G至A或C至T突变/SNP、具体是存在于SMCHD1基因中的一种或多种致病性G至A或C至T突变/SNP、更具体是上述一种或多种致病性G至A或C至T突变/SNP来治疗或预防面肩肱型肌营养不良的方法。

[1977] 先天性代谢异常(IEM)

[1978] 与各种IEM相关的致病性G至A或C至T突变/SNP在ClinVar据库中报告并在表A中公开，包括但不限于原发性1型高草酸尿症、精氨琥珀酸裂解酶缺乏症、鸟氨酸氨甲酰基转移酶缺乏症和枫糖浆尿病。因此，本发明的一个方面涉及一种用于校正与这些疾病中的任何一种相关的一种或多种致病性G至A或C至T突变/SNP的方法，如下所述。

[1979] 原发性1型高草酸尿症

[1980] 在一些实施方案中，本文所述的方法、系统和组合物用于校正与原发性1型高草酸尿症相关的一种或多种致病性G至A或C至T突变/SNP。在一些实施方案中，致病性突变/SNP存在于至少AGXT基因中，包括至少以下各项：

[1981] NM_000030.2(AGXT):c.245G>A(p.Gly82Glu)

[1982] NM_000030.2(AGXT):c.698G>A(p.Arg233His)

[1983] NM_000030.2(AGXT):c.466G>A(p.Gly156Arg)

[1984] NM_000030.2(AGXT):c.106C>T(p.Arg36Cys)

[1985] NM_000030.2(AGXT):c.346G>A(p.Gly116Arg)

[1986] NM_000030.2(AGXT):c.568G>A(p.Gly190Arg)

[1987] NM_000030.2(AGXT):c.653C>T(p.Ser218Leu)

[1988] NM_000030.2(AGXT):c.737G>A(p.Trp246Ter)

[1989] NM_000030.2(AGXT):c.1049G>A(p.Gly350Asp)

[1990] NM_000030.2(AGXT):c.473C>T(p.Ser158Leu)

[1991] NM_000030.2(AGXT):c.907C>T(p.Gln303Ter)

[1992] NM_000030.2(AGXT):c.996G>A(p.Trp332Ter)

[1993] NM_000030.2(AGXT):c.508G>A(p.Gly170Arg)

[1994] 参见表A。因此，本发明的一个方面涉及一种用于通过校正一种或多种致病性G至A或C至T突变/SNP、具体是存在于AGXT基因中的一种或多种致病性G至A或C至T突变/SNP、更具体是上述一种或多种致病性G至A或C至T突变/SNP来治疗或预防原发性1型高草酸尿症的方法。

[1995] 精氨琥珀酸裂解酶缺乏症

[1996] 在一些实施方案中，本文所述的方法、系统和组合物用于校正与精氨琥珀酸裂解酶缺乏症相关的一种或多种致病性G至A或C至T突变/SNP。在一些实施方案中，致病性突变/SNP存在于至少ASL基因中，包括至少以下各项：

[1997] NM_001024943.1(ASL):c.1153C>T(p.Arg385Cys)

[1998] NM_000048.3(ASL):c.532G>A(p.Val178Met)

[1999] NM_000048.3(ASL):c.545G>A(p.Arg182Gln)

[2000] NM_000048.3(ASL):c.175G>A(p.Glu59Lys)

[2001] NM_000048.3(ASL):c.718+5G>A

[2002] NM_000048.3(ASL):c.889C>T(p.Arg297Trp)

[2003] NM_000048.3(ASL):c.1360C>T(p.Gln454Ter)

[2004] NM_000048.3(ASL):c.1060C>T(p.Gln354Ter)

[2005] NM_000048.3(ASL):c.35G>A(p.Arg12Gln)

[2006] NM_000048.3(ASL):c.446+1G>A

[2007] NM_000048.3(ASL):c.544C>T(p.Arg182Ter)

[2008] NM_000048.3(ASL):c.1135C>T(p.Arg379Cys)

[2009] 参见表A。因此，本发明的一个方面涉及一种用于通过校正一种或多种致病性G至A或C至T突变/SNP、具体是存在于ASL基因中的一种或多种致病性G至A或C至T突变/SNP、更具体是上述一种或多种致病性G至A或C至T突变/SNP来治疗或预防精氨琥珀酸裂解酶缺乏症的方法。

[2010] 鸟氨酸氨甲酰基转移酶缺乏症

[2011] 在一些实施方案中，本文所述的方法、系统和组合物用于校正与鸟氨酸氨甲酰基转移酶缺乏症相关的一种或多种致病性G至A或C至T突变/SNP。在一些实施方案中，致病性突变/SNP存在于至少OTC基因中，包括至少以下各项：

[2012] NM_000531.5(OTC):c.119G>A(p.Arg40His)

[2013] NM_000531.5(OTC):c.422G>A(p.Arg141Gln)

[2014] NM_000531.5(OTC):c.829C>T(p.Arg277Trp)

[2015] NM_000531.5(OTC):c.674C>T(p.Pro225Leu)

[2016] 参见表A。因此，本发明的一个方面涉及一种用于通过校正一种或多种致病性G至A或C至T突变/SNP、具体是存在于OTC基因中的一种或多种致病性G至A或C至T突变/SNP、更具体是上述一种或多种致病性G至A或C至T突变/SNP来治疗或预防鸟氨酸氨甲酰基转移酶缺乏症的方法。

[2017] 枫糖浆尿病

[2018] 在一些实施方案中，本文所述的方法、系统和组合物用于校正与枫糖浆尿病相关的一种或多种致病性G至A或C至T突变/SNP。在一些实施方案中，致病性突变/SNP存在于选自BCKDHA、BCKDHB、DBT和DLD的至少一个基因中，包括至少以下各项：

[2019] NM_000709.3(BCKDHA):c.476G>A(p.Arg159Gln)

[2020] NM_183050.3(BCKDHB):c.3G>A(p.Met1Ile)

[2021] NM_183050.3(BCKDHB):c.554C>T(p.Pro185Leu)

[2022] NM_001918.3(DBT):c.1033G>A(p.Gly345Arg)

[2023] NM_000709.3(BCKDHA):c.940C>T(p.Arg314Ter)

[2024] NM_000709.3(BCKDHA):c.793C>T(p.Arg265Trp)

[2025] NM_000709.3(BCKDHA):c.868G>A(p.Gly290Arg)

[2026] NM_000108.4(DLD):c.1123G>A(p.Glu375Lys)

[2027] NM_000709.3(BCKDHA):c.1234G>A(p.Val412Met)

[2028] NM_000709.3(BCKDHA):c.288+1G>A

[2029] NM_000709.3(BCKDHA):c.979G>A(p.Glu327Lys)

[2030] NM_001918.3(DBT):c.901C>T(p.Arg301Cys)

[2031] NM_183050.3(BCKDHB):c.509G>A(p.Arg170His)

[2032] NM_183050.3(BCKDHB):c.799C>T(p.Gln267Ter)

[2033] NM_183050.3(BCKDHB):c.853C>T(p.Arg285Ter)

[2034] NM_183050.3(BCKDHB):c.970C>T(p.Arg324Ter)

[2035] NM_183050.3(BCKDHB):c.832G>A(p.Gly278Ser)

[2036] NM_000709.3(BCKDHA):c.1036C>T(p.Arg346Cys)

[2037] NM_000709.3(BCKDHA):c.288+9C>T

[2038] NM_000709.3(BCKDHA):c.632C>T(p.Thr211Met)

[2039] NM_000709.3(BCKDHA):c.659C>T(p.Ala220Val)

[2040] NM_000709.3(BCKDHA):c.964C>T(p.Gln322Ter)

[2041] NM_001918.3(DBT):c.1291C>T(p.Arg431Ter)

[2042] NM_001918.3(DBT):c.251G>A(p.Trp84Ter)

[2043] NM_001918.3(DBT):c.871C>T(p.Arg291Ter)

[2044] NM_000056.4(BCKDHB):c.1016C>T(p.Ser339Leu)

[2045] NM_000056.4(BCKDHB):c.344-1G>A

[2046] NM_000056.4(BCKDHB):c.633+1G>A

[2047] NM_000056.4(BCKDHB):c.952-1G>A

[2048] 参见表A。因此，本发明的一个方面涉及一种用于通过校正一种或多种致病性G至A或C至T突变/SNP、具体是存在于选自BCKDHA、BCKDHB、DBT和DLD的至少一个基因中的一种或多种致病性G至A或C至T突变/SNP、更具体是上述一种或多种致病性G至A或C至T突变/SNP来治疗或预防枫糖浆尿病的方法。

[2049] 癌症相关疾病

[2050] 与各种癌症和癌症相关疾病相关的致病性G至A或C至T突变/SNP在ClinVar数据库中报告并在表A中公开，包括但不限于乳腺-卵巢癌和林奇综合征。因此，本发明的一个方面涉及一种用于校正与这些疾病中的任何一种相关的一种或多种致病性G至A或C至T突变/SNP的方法，如下所述。

[2051] 乳房-卵巢癌

[2052] 在一些实施方案中，本文所述的方法、系统和组合物用于校正与乳房-卵巢癌相关的一种或多种致病性G至A或C至T突变/SNP。在一些实施方案中，致病性突变/SNP存在于至少BRCA1或BRCA2基因中，包括至少以下各项：

[2053] NM_007294.3(BRCA1):c.5095C>T(p.Arg1699Trp)

[2054] NM_000059.3(BRCA2):c.7558C>T(p.Arg2520Ter)

[2055] NM_007294.3(BRCA1):c.2572C>T(p.Gln858Ter)

[2056] NM_007294.3(BRCA1):c.3607C>T(p.Arg1203Ter)

[2057] NM_007294.3(BRCA1):c.5503C>T(p.Arg1835Ter)

[2058] NM_007294.3(BRCA1):c.2059C>T(p.Gln687Ter)

[2059] NM_007294.3(BRCA1):c.4675+1G>A

[2060] NM_007294.3(BRCA1):c.5251C>T(p.Arg1751Ter)

[2061] NM_007294.3(BRCA1):c.5444G>A(p.Trp1815Ter)

[2062] NM_000059.3(BRCA2):c.9318G>A(p.Trp3106Ter)

[2063] NM_000059.3(BRCA2):c.9382C>T(p.Arg3128Ter)

[2064] NM_000059.3(BRCA2):c.274C>T(p.Gln92Ter)

[2065] NM_000059.3(BRCA2):c.6952C>T(p.Arg2318Ter)

[2066] NM_007294.3(BRCA1):c.1687C>T(p.Gln563Ter)

[2067] NM_007294.3(BRCA1):c.2599C>T(p.Gln867Ter)

[2068] NM_007294.3(BRCA1):c.784C>T(p.Gln262Ter)

[2069] NM_007294.3(BRCA1):c.280C>T(p.Gln94Ter)

[2070] NM_007294.3(BRCA1):c.5542C>T(p.Gln1848Ter)

[2071] NM_007294.3(BRCA1):c.5161C>T(p.Gln1721Ter)

[2072] NM_007294.3(BRCA1):c.4573C>T(p.Gln1525Ter)

[2073] NM_007294.3(BRCA1):c.4270C>T(p.Gln1424Ter)

[2074] NM_007294.3(BRCA1):c.4225C>T(p.Gln1409Ter)

[2075] NM_007294.3(BRCA1):c.4066C>T(p.Gln1356Ter)

[2076] NM_007294.3(BRCA1):c.3679C>T(p.Gln1227Ter)

[2077] NM_007294.3(BRCA1):c.1918C>T(p.Gln640Ter)

[2078] NM_007294.3(BRCA1):c.963G>A(p.Trp321Ter)

[2079] NM_007294.3(BRCA1):c.718C>T(p.Gln240Ter)

[2080] NM_000059.3(BRCA2):c.9196C>T(p.Gln3066Ter)

[2081] NM_000059.3(BRCA2):c.9154C>T(p.Arg3052Trp)

[2082] NM_007294.3(BRCA1):c.3991C>T(p.Gln1331Ter)

[2083] NM_007294.3(BRCA1):c.4097-1G>A

[2084] NM_007294.3(BRCA1):c.1059G>A(p.Trp353Ter)

[2085] NM_007294.3(BRCA1):c.1115G>A(p.Trp372Ter)

[2086] NM_007294.3(BRCA1):c.1138C>T(p.Gln380Ter)

[2087] NM_007294.3(BRCA1):c.1612C>T(p.Gln538Ter)

[2088] NM_007294.3(BRCA1):c.1621C>T(p.Gln541Ter)

[2089] NM_007294.3(BRCA1):c.1630C>T(p.Gln544Ter)

[2090] NM_007294.3(BRCA1):c.178C>T(p.Gln60Ter)

[2091] NM_007294.3(BRCA1):c.1969C>T(p.Gln657Ter)

[2092] NM_007294.3(BRCA1):c.2275C>T(p.Gln759Ter)

[2093] NM_007294.3(BRCA1):c.2410C>T(p.Gln804Ter)

[2094] NM_007294.3(BRCA1):c.2869C>T(p.Gln957Ter)

[2095] NM_007294.3(BRCA1):c.2923C>T(p.Gln975Ter)

[2096] NM_007294.3(BRCA1):c.3268C>T(p.Gln1090Ter)

[2097] NM_007294.3(BRCA1):c.3430C>T(p.Gln1144Ter)

[2098] NM_007294.3(BRCA1):c.3544C>T(p.Gln1182Ter)

[2099] NM_007294.3(BRCA1):c.4075C>T(p.Gln1359Ter)

[2100] NM_007294.3(BRCA1):c.4201C>T(p.Gln1401Ter)

[2101] NM_007294.3(BRCA1):c.4399C>T(p.Gln1467Ter)

[2102] NM_007294.3(BRCA1):c.4552C>T(p.Gln1518Ter)

[2103] NM_007294.3(BRCA1):c.5054C>T(p.Thr1685Ile)

[2104] NM_007294.3(BRCA1):c.514C>T(p.Gln172Ter)

[2105] NM_007294.3(BRCA1):c.5239C>T(p.Gln1747Ter)

[2106] NM_007294.3(BRCA1):c.5266C>T(p.Gln1756Ter)

[2107] NM_007294.3(BRCA1):c.5335C>T(p.Gln1779Ter)

[2108] NM_007294.3(BRCA1):c.5345G>A(p.Trp1782Ter)

[2109] NM_007294.3(BRCA1):c.5511G>A(p.Trp1837Ter)

[2110] NM_007294.3(BRCA1):c.5536C>T(p.Gln1846Ter)

[2111] NM_007294.3(BRCA1):c.55C>T(p.Gln19Ter)

[2112] NM_007294.3(BRCA1):c.949C>T(p.Gln317Ter)

[2113] NM_007294.3(BRCA1):c.928C>T(p.Gln310Ter)

[2114] NM_007294.3(BRCA1):c.5117G>A(p.Gly1706Glu)

[2115] NM_007294.3(BRCA1):c.5136G>A(p.Trp1712Ter)

[2116] NM_007294.3(BRCA1):c.4327C>T(p.Arg1443Ter)

[2117] NM_007294.3(BRCA1):c.1471C>T(p.Gln491Ter)

[2118] NM_007294.3(BRCA1):c.1576C>T(p.Gln526Ter)

[2119] NM_007294.3(BRCA1):c.160C>T(p.Gln54Ter)

[2120] NM_007294.3(BRCA1):c.2683C>T(p.Gln895Ter)

[2121] NM_007294.3(BRCA1):c.2761C>T(p.Gln921Ter)

[2122] NM_007294.3(BRCA1):c.3895C>T(p.Gln1299Ter)

[2123] NM_007294.3(BRCA1):c.4339C>T(p.Gln1447Ter)

[2124] NM_007294.3(BRCA1):c.4372C>T(p.Gln1458Ter)

[2125] NM_007294.3(BRCA1):c.5153G>A(p.Trp1718Ter)

[2126] NM_007294.3(BRCA1):c.5445G>A(p.Trp1815Ter)

[2127] NM_007294.3(BRCA1):c.5510G>A(p.Trp1837Ter)

[2128] NM_007294.3(BRCA1):c.5346G>A(p.Trp1782Ter)

[2129] NM_007294.3(BRCA1):c.1116G>A(p.Trp372Ter)

[2130] NM_007294.3(BRCA1):c.1999C>T(p.Gln667Ter)

[2131] NM_007294.3(BRCA1):c.4183C>T(p.Gln1395Ter)

[2132] NM_007294.3(BRCA1):c.4810C>T(p.Gln1604Ter)

[2133] NM_007294.3(BRCA1):c.850C>T(p.Gln284Ter)

[2134] NM_007294.3(BRCA1):c.1058G>A(p.Trp353Ter)

[2135] NM_007294.3(BRCA1):c.131G>A(p.Cys44Tyr)

[2136] NM_007294.3(BRCA1):c.1600C>T(p.Gln534Ter)

[2137] NM_007294.3(BRCA1):c.3286C>T(p.Gln1096Ter)

[2138] NM_007294.3(BRCA1):c.3403C>T(p.Gln1135Ter)

[2139] NM_007294.3(BRCA1):c.34C>T(p.Gln12Ter)

[2140] NM_007294.3(BRCA1):c.4258C>T(p.Gln1420Ter)

[2141] NM_007294.3(BRCA1):c.4609C>T(p.Gln1537Ter)

[2142] NM_007294.3(BRCA1):c.5154G>A(p.Trp1718Ter)

[2143] NM_007294.3(BRCA1):c.5431C>T(p.Gln1811Ter)

[2144] NM_007294.3(BRCA1):c.241C>T(p.Gln81Ter)

[2145] NM_007294.3(BRCA1):c.3331C>T(p.Gln1111Ter)

[2146] NM_007294.3(BRCA1):c.3967C>T(p.Gln1323Ter)

[2147] NM_007294.3(BRCA1):c.415C>T(p.Gln139Ter)

[2148] NM_007294.3(BRCA1):c.505C>T(p.Gln169Ter)

[2149] NM_007294.3(BRCA1):c.5194-12G>A

[2150] NM_007294.3(BRCA1):c.5212G>A(p.Gly1738Arg)

[2151] NM_007294.3(BRCA1):c.5332+1G>A

[2152] NM_007294.3(BRCA1):c.1480C>T(p.Gln494Ter)

[2153] NM_007294.3(BRCA1):c.2563C>T(p.Gln855Ter)

[2154] NM_007294.3(BRCA1):c.1066C>T(p.Gln356Ter)

[2155] NM_007294.3(BRCA1):c.3718C>T(p.Gln1240Ter)

[2156] NM_007294.3(BRCA1):c.3817C>T(p.Gln1273Ter)

[2157] NM_007294.3(BRCA1):c.3937C>T(p.Gln1313Ter)

[2158] NM_007294.3(BRCA1):c.4357+1G>A

[2159] NM_007294.3(BRCA1):c.5074+1G>A

[2160] NM_007294.3(BRCA1):c.5277+1G>A

[2161] NM_007294.3(BRCA1):c.2338C>T(p.Gln780Ter)

[2162] NM_007294.3(BRCA1):c.3598C>T(p.Gln1200Ter)

[2163] NM_007294.3(BRCA1):c.3841C>T(p.Gln1281Ter)

[2164] NM_007294.3(BRCA1):c.4222C>T(p.Gln1408Ter)

[2165] NM_007294.3(BRCA1):c.4524G>A(p.Trp1508Ter)

[2166] NM_007294.3(BRCA1):c.5353C>T(p.Gln1785Ter)

[2167] NM_007294.3(BRCA1):c.962G>A(p.Trp321Ter)

[2168] NM_007294.3(BRCA1):c.220C>T(p.Gln74Ter)

[2169] NM_007294.3(BRCA1):c.2713C>T(p.Gln905Ter)

[2170] NM_007294.3(BRCA1):c.2800C>T(p.Gln934Ter)

[2171] NM_007294.3(BRCA1):c.4612C>T(p.Gln1538Ter)

[2172] NM_007294.3(BRCA1):c.3352C>T(p.Gln1118Ter)

[2173] NM_007294.3(BRCA1):c.4834C>T(p.Gln1612Ter)

[2174] NM_007294.3(BRCA1):c.4523G>A(p.Trp1508Ter)

[2175] NM_007294.3(BRCA1):c.5135G>A(p.Trp1712Ter)

[2176] NM_007294.3(BRCA1):c.1155G>A(p.Trp385Ter)

[2177] NM_007294.3(BRCA1):c.4987-1G>A

[2178] NM_000059.3(BRCA2):c.9573G>A(p.Trp3191Ter)

[2179] NM_000059.3(BRCA2):c.1945C>T(p.Gln649Ter)

[2180] NM_000059.3(BRCA2):c.217C>T(p.Gln73Ter)

[2181] NM_000059.3(BRCA2):c.523C>T(p.Gln175Ter)

[2182] NM_000059.3(BRCA2):c.2548C>T(p.Gln850Ter)

[2183] NM_000059.3(BRCA2):c.2905C>T(p.Gln969Ter)

[2184] NM_000059.3(BRCA2):c.4689G>A(p.Trp1563Ter)

[2185] NM_000059.3(BRCA2):c.4972C>T(p.Gln1658Ter)

[2186] NM_000059.3(BRCA2):c.1184G>A(p.Trp395Ter)

[2187] NM_000059.3(BRCA2):c.2137C>T(p.Gln713Ter)

[2188] NM_000059.3(BRCA2):c.3217C>T(p.Gln1073Ter)

[2189] NM_000059.3(BRCA2):c.3523C>T(p.Gln1175Ter)

[2190] NM_000059.3(BRCA2):c.4783C>T(p.Gln1595Ter)

[2191] NM_000059.3(BRCA2):c.5800C>T(p.Gln1934Ter)

[2192] NM_000059.3(BRCA2):c.6478C>T(p.Gln2160Ter)

[2193] NM_000059.3(BRCA2):c.7033C>T(p.Gln2345Ter)

[2194] NM_000059.3(BRCA2):c.7495C>T(p.Gln2499Ter)

[2195] NM_000059.3(BRCA2):c.7501C>T(p.Gln2501Ter)

[2196] NM_000059.3(BRCA2):c.7887G>A(p.Trp2629Ter)

[2197] NM_000059.3(BRCA2):c.8910G>A(p.Trp2970Ter)

[2198] NM_000059.3(BRCA2):c.9139C>T(p.Gln3047Ter)

[2199] NM_000059.3(BRCA2):c.9739C>T(p.Gln3247Ter)

[2200] NM_000059.3(BRCA2):c.582G>A(p.Trp194Ter)

[2201] NM_000059.3(BRCA2):c.7963C>T(p.Gln2655Ter)

[2202] NM_000059.3(BRCA2):c.8695C>T(p.Gln2899Ter)

[2203] NM_000059.3(BRCA2):c.8869C>T(p.Gln2957Ter)

[2204] NM_000059.3(BRCA2):c.1117C>T(p.Gln373Ter)

[2205] NM_000059.3(BRCA2):c.1825C>T(p.Gln609Ter)

[2206] NM_000059.3(BRCA2):c.2455C>T(p.Gln819Ter)

[2207] NM_000059.3(BRCA2):c.2881C>T(p.Gln961Ter)

[2208] NM_000059.3(BRCA2):c.3265C>T(p.Gln1089Ter)

[2209] NM_000059.3(BRCA2):c.3283C>T(p.Gln1095Ter)

[2210] NM_000059.3(BRCA2):c.3442C>T(p.Gln1148Ter)

[2211] NM_000059.3(BRCA2):c.3871C>T(p.Gln1291Ter)

[2212] NM_000059.3(BRCA2):c.439C>T(p.Gln147Ter)

[2213] NM_000059.3(BRCA2):c.4525C>T(p.Gln1509Ter)

[2214] NM_000059.3(BRCA2):c.475+1G>A

[2215] NM_000059.3(BRCA2):c.5344C>T(p.Gln1782Ter)

[2216] NM_000059.3(BRCA2):c.5404C>T(p.Gln1802Ter)

[2217] NM_000059.3(BRCA2):c.5773C>T(p.Gln1925Ter)

[2218] NM_000059.3(BRCA2):c.5992C>T(p.Gln1998Ter)

[2219] NM_000059.3(BRCA2):c.6469C>T(p.Gln2157Ter)

[2220] NM_000059.3(BRCA2):c.7261C>T(p.Gln2421Ter)

[2221] NM_000059.3(BRCA2):c.7303C>T(p.Gln2435Ter)

[2222] NM_000059.3(BRCA2):c.7471C>T(p.Gln2491Ter)

[2223] NM_000059.3(BRCA2):c.7681C>T(p.Gln2561Ter)

[2224] NM_000059.3(BRCA2):c.7738C>T(p.Gln2580Ter)

[2225] NM_000059.3(BRCA2):c.7886G>A(p.Trp2629Ter)

[2226] NM_000059.3(BRCA2):c.8140C>T(p.Gln2714Ter)

[2227] NM_000059.3(BRCA2):c.8363G>A(p.Trp2788Ter)

[2228] NM_000059.3(BRCA2):c.8572C>T(p.Gln2858Ter)

[2229] NM_000059.3(BRCA2):c.8773C>T(p.Gln2925Ter)

[2230] NM_000059.3(BRCA2):c.8821C>T(p.Gln2941Ter)

[2231] NM_000059.3(BRCA2):c.9109C>T(p.Gln3037Ter)

[2232] NM_000059.3(BRCA2):c.9317G>A(p.Trp3106Ter)

[2233] NM_000059.3(BRCA2):c.9466C>T(p.Gln3156Ter)

[2234] NM_000059.3(BRCA2):c.9572G>A(p.Trp3191Ter)

[2235] NM_000059.3(BRCA2):c.8490G>A(p.Trp2830Ter)

[2236] NM_000059.3(BRCA2):c.5980C>T(p.Gln1994Ter)

[2237] NM_000059.3(BRCA2):c.7721G>A(p.Trp2574Ter)

[2238] NM_000059.3(BRCA2):c.196C>T(p.Gln66Ter)

[2239] NM_000059.3(BRCA2):c.7618-1G>A

[2240] NM_000059.3(BRCA2):c.8489G>A(p.Trp2830Ter)

[2241] NM_000059.3(BRCA2):c.7857G>A(p.Trp2619Ter)

[2242] NM_000059.3(BRCA2):c.1261C>T(p.Gln421Ter)

[2243] NM_000059.3(BRCA2):c.1456C>T(p.Gln486Ter)

[2244] NM_000059.3(BRCA2):c.3319C>T(p.Gln1107Ter)

[2245] NM_000059.3(BRCA2):c.5791C>T(p.Gln1931Ter)

[2246] NM_000059.3(BRCA2):c.6070C>T(p.Gln2024Ter)

[2247] NM_000059.3(BRCA2):c.7024C>T(p.Gln2342Ter)

[2248] NM_000059.3(BRCA2):c.961C>T(p.Gln321Ter)

[2249] NM_000059.3(BRCA2):c.9380G>A(p.Trp3127Ter)

[2250] NM_000059.3(BRCA2):c.8364G>A(p.Trp2788Ter)

[2251] NM_000059.3(BRCA2):c.7758G>A(p.Trp2586Ter)

[2252] NM_000059.3(BRCA2):c.2224C>T(p.Gln742Ter)

[2253] NM_000059.3(BRCA2):c.5101C>T(p.Gln1701Ter)

[2254] NM_000059.3(BRCA2):c.5959C>T(p.Gln1987Ter)

[2255] NM_000059.3(BRCA2):c.7060C>T(p.Gln2354Ter)

[2256] NM_000059.3(BRCA2):c.9100C>T(p.Gln3034Ter)

[2257] NM_000059.3(BRCA2):c.9148C>T(p.Gln3050Ter)

[2258] NM_000059.3(BRCA2):c.9883C>T(p.Gln3295Ter)

[2259] NM_000059.3(BRCA2):c.1414C>T(p.Gln472Ter)

[2260] NM_000059.3(BRCA2):c.1689G>A(p.Trp563Ter)

[2261] NM_000059.3(BRCA2):c.581G>A(p.Trp194Ter)

[2262] NM_000059.3(BRCA2):c.6490C>T(p.Gln2164Ter)

[2263] NM_000059.3(BRCA2):c.7856G>A(p.Trp2619Ter)

[2264] NM_000059.3(BRCA2):c.8970G>A(p.Trp2990Ter)

[2265] NM_000059.3(BRCA2):c.92G>A(p.Trp31Ter)

[2266] NM_000059.3(BRCA2):c.9376C>T(p.Gln3126Ter)

[2267] NM_000059.3(BRCA2):c.93G>A(p.Trp31Ter)

[2268] NM_000059.3(BRCA2):c.1189C>T(p.Gln397Ter)

[2269] NM_000059.3(BRCA2):c.2818C>T(p.Gln940Ter)

[2270] NM_000059.3(BRCA2):c.2979G>A(p.Trp993Ter)

[2271] NM_000059.3(BRCA2):c.3166C>T(p.Gln1056Ter)

[2272] NM_000059.3(BRCA2):c.4285C>T(p.Gln1429Ter)

[2273] NM_000059.3(BRCA2):c.6025C>T(p.Gln2009Ter)

[2274] NM_000059.3(BRCA2):c.772C>T(p.Gln258Ter)

[2275] NM_000059.3(BRCA2):c.7877G>A(p.Trp2626Ter)

[2276] NM_000059.3(BRCA2):c.3109C>T(p.Gln1037Ter)

[2277] NM_000059.3(BRCA2):c.4222C>T(p.Gln1408Ter)

[2278] NM_000059.3(BRCA2):c.7480C>T(p.Arg2494Ter)

[2279] NM_000059.3(BRCA2):c.7878G>A(p.Trp2626Ter)

[2280] NM_000059.3(BRCA2):c.9076C>T(p.Gln3026Ter)

[2281] NM_000059.3(BRCA2):c.1855C>T(p.Gln619Ter)

[2282] NM_000059.3(BRCA2):c.4111C>T(p.Gln1371Ter)

[2283] NM_000059.3(BRCA2):c.5656C>T(p.Gln1886Ter)

[2284] NM_000059.3(BRCA2):c.7757G>A(p.Trp2586Ter)

[2285] NM_000059.3(BRCA2):c.8243G>A(p.Gly2748Asp)

[2286] NM_000059.3(BRCA2):c.8878C>T(p.Gln2960Ter)

[2287] NM_000059.3(BRCA2):c.8487+1G>A

[2288] NM_000059.3(BRCA2):c.8677C>T(p.Gln2893Ter)

[2289] NM_000059.3(BRCA2):c.250C>T(p.Gln84Ter)

[2290] NM_000059.3(BRCA2):c.6124C>T(p.Gln2042Ter)

[2291] NM_000059.3(BRCA2):c.7617+1G>A

[2292] NM_000059.3(BRCA2):c.8575C>T(p.Gln2859Ter)

[2293] NM_000059.3(BRCA2):c.8174G>A(p.Trp2725Ter)

[2294] NM_000059.3(BRCA2):c.3187C>T(p.Gln1063Ter)

[2295] NM_000059.3(BRCA2):c.9381G>A(p.Trp3127Ter)

[2296] NM_000059.3(BRCA2):c.2095C>T(p.Gln699Ter)

[2297] NM_000059.3(BRCA2):c.1642C>T(p.Gln548Ter)

[2298] NM_000059.3(BRCA2):c.8608C>T(p.Gln2870Ter)

[2299] NM_000059.3(BRCA2):c.3412C>T(p.Gln1138Ter)

[2300] NM_000059.3(BRCA2):c.4246C>T(p.Gln1416Ter)

[2301] NM_000059.3(BRCA2):c.6475C>T(p.Gln2159Ter)

[2302] NM_000059.3(BRCA2):c.7366C>T(p.Gln2456Ter)

[2303] NM_000059.3(BRCA2):c.7516C>T(p.Gln2506Ter)

[2304] NM_000059.3(BRCA2):c.8969G>A(p.Trp2990Ter)

[2305] NM_000059.3(BRCA2):c.6487C>T(p.Gln2163Ter)

[2306] NM_000059.3(BRCA2):c.2978G>A(p.Trp993Ter)

[2307] NM_000059.3(BRCA2):c.7615C>T(p.Gln2539Ter)

[2308] NM_000059.3(BRCA2):c.9106C>T(p.Gln3036Ter)

[2309] 参见表A。因此，本发明的一个方面涉及一种用于通过校正一种或多种致病性G至A或C至T突变/SNP、具体是存在于BRCA1或BRCA2基因中的一种或多种致病性G至A或C至T突变/SNP、更具体是上述一种或多种致病性G至A或C至T突变/SNP来治疗或预防乳腺-卵巢癌的方法。

[2310] 林奇综合征

[2311] 在一些实施方案中，本文所述的方法、系统和组合物用于校正与林奇综合征相关的一种或多种致病性G至A或C至T突变/SNP。在一些实施方案中，致病性突变/SNP存在于选自MSH6、MSH2、EPCAM、PMS2和MLH1的至少一个基因中，包括至少以下各项：

[2312] NM_000179.2(MSH6):c.1045C>T(p.Gln349Ter)

[2313] NM_000251.2(MSH2):c.1384C>T(p.Gln462Ter)

[2314] NM_002354.2(EPCAM):c.133C>T(p.Gln45Ter)

[2315] NM_002354.2(EPCAM):c.429G>A(p.Trp143Ter)

[2316] NM_002354.2(EPCAM):c.523C>T(p.Gln175Ter)

[2317] NM_000179.2(MSH6):c.2680C>T(p.Gln894Ter)

[2318] NM_000251.2(MSH2):c.350G>A(p.Trp117Ter)

[2319] NM_000179.2(MSH6):c.2735G>A(p.Trp912Ter)

[2320] NM_000179.2(MSH6):c.3556+1G>A

[2321] NM_000251.2(MSH2):c.388C>T(p.Gln130Ter)

[2322] NM_000535.6(PMS2):c.1912C>T(p.Gln638Ter)

[2323] NM_000535.6(PMS2):c.1891C>T(p.Gln631Ter)

[2324] NM_000249.3(MLH1):c.454-1G>A

[2325] NM_000251.2(MSH2):c.1030C>T(p.Gln344Ter)

[2326] NM_000179.2(MSH6):c.2330G>A(p.Trp777Ter)

[2327] NM_000179.2(MSH6):c.2191C>T(p.Gln731Ter)

[2328] NM_000179.2(MSH6):c.2764C>T(p.Arg922Ter)

[2329] NM_000179.2(MSH6):c.2815C>T(p.Gln939Ter)

[2330] NM_000179.2(MSH6):c.3020G>A(p.Trp1007Ter)

[2331] NM_000179.2(MSH6):c.3436C>T(p.Gln1146Ter)

[2332] NM_000179.2(MSH6):c.3647-1G>A

[2333] NM_000179.2(MSH6):c.3772C>T(p.Gln1258Ter)

[2334] NM_000179.2(MSH6):c.3838C>T(p.Gln1280Ter)

[2335] NM_000179.2(MSH6):c.706C>T(p.Gln236Ter)

[2336] NM_000179.2(MSH6):c.730C>T(p.Gln244Ter)

[2337] NM_000249.3(MLH1):c.1171C>T(p.Gln391Ter)

[2338] NM_000249.3(MLH1):c.1192C>T(p.Gln398Ter)

[2339] NM_000249.3(MLH1):c.1225C>T(p.Gln409Ter)

[2340] NM_000249.3(MLH1):c.1276C>T(p.Gln426Ter)

[2341] NM_000249.3(MLH1):c.1528C>T(p.Gln510Ter)

[2342] NM_000249.3(MLH1):c.1609C>T(p.Gln537Ter)

[2343] NM_000249.3(MLH1):c.1613G>A(p.Trp538Ter)

[2344] NM_000249.3(MLH1):c.1614G>A(p.Trp538Ter)

[2345] NM_000249.3(MLH1):c.1624C>T(p.Gln542Ter)

[2346] NM_000249.3(MLH1):c.1684C>T(p.Gln562Ter)

[2347] NM_000249.3(MLH1):c.1731+1G>A

[2348] NM_000249.3(MLH1):c.1731+5G>A

[2349] NM_000249.3(MLH1):c.1732-1G>A

[2350] NM_000249.3(MLH1):c.1896G>A(p.Glu632＝)

[2351] NM_000249.3(MLH1):c.1989+1G>A

[2352] NM_000249.3(MLH1):c.1990-1G>A

[2353] NM_000249.3(MLH1):c.1998G>A(p.Trp666Ter)

[2354] NM_000249.3(MLH1):c.208-1G>A

[2355] NM_000249.3(MLH1):c.2101C>T(p.Gln701Ter)

[2356] NM_000249.3(MLH1):c.2136G>A(p.Trp712Ter)

[2357] NM_000249.3(MLH1):c.2224C>T(p.Gln742Ter)

[2358] NM_000249.3(MLH1):c.230G>A(p.Cys77Tyr)

[2359] NM_000249.3(MLH1):c.256C>T(p.Gln86Ter)

[2360] NM_000249.3(MLH1):c.436C>T(p.Gln146Ter)

[2361] NM_000249.3(MLH1):c.445C>T(p.Gln149Ter)

[2362] NM_000249.3(MLH1):c.545G>A(p.Arg182Lys)

[2363] NM_000249.3(MLH1):c.731G>A(p.Gly244Asp)

[2364] NM_000249.3(MLH1):c.76C>T(p.Gln26Ter)

[2365] NM_000249.3(MLH1):c.842C>T(p.Ala281Val)

[2366] NM_000249.3(MLH1):c.882C>T(p.Leu294=)

[2367] NM_000249.3(MLH1):c.901C>T(p.Gln301Ter)

[2368] NM_000251.2(MSH2):c.1013G>A(p.Gly338Glu)

[2369] NM_000251.2(MSH2):c.1034G>A(p.Trp345Ter)

[2370] NM_000251.2(MSH2):c.1129C>T(p.Gln377Ter)

[2371] NM_000251.2(MSH2):c.1183C>T(p.Gln395Ter)

[2372] NM_000251.2(MSH2):c.1189C>T(p.Gln397Ter)

[2373] NM_000251.2(MSH2):c.1204C>T(p.Gln402Ter)

[2374] NM_000251.2(MSH2):c.1276+1G>A

[2375] NM_000251.2(MSH2):c.1528C>T(p.Gln510Ter)

[2376] NM_000251.2(MSH2):c.1552C>T(p.Gln518Ter)

[2377] NM_000251.2(MSH2):c.1720C>T(p.Gln574Ter)

[2378] NM_000251.2(MSH2):c.1777C>T(p.Gln593Ter)

[2379] NM_000251.2(MSH2):c.1885C>T(p.Gln629Ter)

[2380] NM_000251.2(MSH2):c.2087C>T(p.Pro696Leu)

[2381] NM_000251.2(MSH2):c.2251G>A(p.Gly751Arg)

[2382] NM_000251.2(MSH2):c.2291G>A(p.Trp764Ter)

[2383] NM_000251.2(MSH2):c.2292G>A(p.Trp764Ter)

[2384] NM_000251.2(MSH2):c.2446C>T(p.Gln816Ter)

[2385] NM_000251.2(MSH2):c.2470C>T(p.Gln824Ter)

[2386] NM_000251.2(MSH2):c.2536C>T(p.Gln846Ter)

[2387] NM_000251.2(MSH2):c.2581C>T(p.Gln861Ter)

[2388] NM_000251.2(MSH2):c.2634G>A(p.Glu878=)

[2389] NM_000251.2(MSH2):c.2635C>T(p.Gln879Ter)

[2390] NM_000251.2(MSH2):c.28C>T(p.Gln10Ter)

[2391] NM_000251.2(MSH2):c.472C>T(p.Gln158Ter)

[2392] NM_000251.2(MSH2):c.478C>T(p.Gln160Ter)

[2393] NM_000251.2(MSH2):c.484G>A(p.Gly162Arg)

[2394] NM_000251.2(MSH2):c.490G>A(p.Gly164Arg)

[2395] NM_000251.2(MSH2):c.547C>T(p.Gln183Ter)

[2396] NM_000251.2(MSH2):c.577C>T(p.Gln193Ter)

[2397] NM_000251.2(MSH2):c.643C>T(p.Gln215Ter)

[2398] NM_000251.2(MSH2):c.645+1G>A

[2399] NM_000251.2(MSH2):c.652C>T(p.Gln218Ter)

[2400] NM_000251.2(MSH2):c.754C>T(p.Gln252Ter)

[2401] NM_000251.2(MSH2):c.792+1G>A

[2402] NM_000251.2(MSH2):c.942G>A(p.Gln314=)

[2403] NM_000535.6(PMS2):c.949C>T(p.Gln317Ter)

[2404] NM_000249.3(MLH1):c.306+1G>A

[2405] NM_000249.3(MLH1):c.62C>T(p.Ala21Val)

[2406] NM_000251.2(MSH2):c.1865C>T(p.Pro622Leu)

[2407] NM_000179.2(MSH6):c.426G>A(p.Trp142Ter)

[2408] NM_000251.2(MSH2):c.715C>T(p.Gln239Ter)

[2409] NM_000249.3(MLH1):c.350C>T(p.Thr117Met)

[2410] NM_000251.2(MSH2):c.1915C>T(p.His639Tyr)

[2411] NM_000251.2(MSH2):c.289C>T(p.Gln97Ter)

[2412] NM_000251.2(MSH2):c.2785C>T(p.Arg929Ter)

[2413] NM_000249.3(MLH1):c.131C>T(p.Ser44Phe)

[2414] NM_000249.3(MLH1):c.1219C>T(p.Gln407Ter)

[2415] NM_000249.3(MLH1):c.306+5G>A

[2416] NM_000251.2(MSH2):c.1801C>T(p.Gln601Ter)

[2417] NM_000535.6(PMS2):c.1144+1G>A

[2418] NM_000251.2(MSH2):c.1984C>T(p.Gln662Ter)

[2419] NM_000249.3(MLH1):c.381-1G>A

[2420] NM_000535.6(PMS2):c.631C>T(p.Arg211Ter)

[2421] NM_000251.2(MSH2):c.790C>T(p.Gln264Ter)

[2422] NM_000251.2(MSH2):c.366+1G>A

[2423] NM_000249.3(MLH1):c.298C>T(p.Arg100Ter)

[2424] NM_000179.2(MSH6):c.3013C>T(p.Arg1005Ter)

[2425] NM_000179.2(MSH6):c.694C>T(p.Gln232Ter)

[2426] NM_000179.2(MSH6):c.742C>T(p.Arg248Ter)

[2427] NM_000249.3(MLH1):c.1039-1G>A

[2428] NM_000249.3(MLH1):c.142C>T(p.Gln48Ter)

[2429] NM_000249.3(MLH1):c.1790G>A(p.Trp597Ter)

[2430] NM_000249.3(MLH1):c.1961C>T(p.Pro654Leu)

[2431] NM_000249.3(MLH1):c.2103+1G>A

[2432] NM_000249.3(MLH1):c.2135G>A(p.Trp712Ter)

[2433] NM_000249.3(MLH1):c.588+5G>A

[2434] NM_000249.3(MLH1):c.790+1G>A

[2435] NM_000251.2(MSH2):c.1035G>A(p.Trp345Ter)

[2436] NM_000251.2(MSH2):c.1255C>T(p.Gln419Ter)

[2437] NM_000251.2(MSH2):c.1861C>T(p.Arg621Ter)

[2438] NM_000251.2(MSH2):c.226C>T(p.Gln76Ter)

[2439] NM_000251.2(MSH2):c.2653C>T(p.Gln885Ter)

[2440] NM_000251.2(MSH2):c.508C>T(p.Gln170Ter)

[2441] NM_000251.2(MSH2):c.862C>T(p.Gln288Ter)

[2442] NM_000251.2(MSH2):c.892C>T(p.Gln298Ter)

[2443] NM_000251.2(MSH2):c.970C>T(p.Gln324Ter)

[2444] NM_000179.2(MSH6):c.4001G>A(p.Arg1334Gln)

[2445] NM_000251.2(MSH2):c.1662-1G>A

[2446] NM_000535.6(PMS2):c.1882C>T(p.Arg628Ter)

[2447] NM_000535.6(PMS2):c.2174+1G>A

[2448] NM_000535.6(PMS2):c.2404C>T(p.Arg802Ter)

[2449] NM_000179.2(MSH6):c.3991C>T(p.Arg1331Ter)

[2450] NM_000179.2(MSH6):c.2503C>T(p.Gln835Ter)

[2451] NM_000179.2(MSH6):c.718C>T(p.Arg240Ter)

[2452] NM_000249.3(MLH1):c.1038G>A(p.Gln346=)

[2453] NM_000249.3(MLH1):c.245C>T(p.Thr82Ile)

[2454] NM_000249.3(MLH1):c.83C>T(p.Pro28Leu)

[2455] NM_000249.3(MLH1):c.884G>A(p.Ser295Asn)

[2456] NM_000249.3(MLH1):c.982C>T(p.Gln328Ter)

[2457] NM_000251.2(MSH2):c.1046C>T(p.Pro349Leu)

[2458] NM_000251.2(MSH2):c.1120C>T(p.Gln374Ter)

[2459] NM_000251.2(MSH2):c.1285C>T(p.Gln429Ter)

[2460] NM_000251.2(MSH2):c.1477C>T(p.Gln493Ter)

[2461] NM_000251.2(MSH2):c.2152C>T(p.Gln718Ter)

[2462] NM_000535.6(PMS2):c.703C>T(p.Gln235Ter)

[2463] NM_000249.3(MLH1):c.2141G>A(p.Trp714Ter)

[2464] NM_000251.2(MSH2):c.1009C>T(p.Gln337Ter)

[2465] NM_000251.2(MSH2):c.1216C>T(p.Arg406Ter)

[2466] NM_000179.2(MSH6):c.3202C>T(p.Arg1068Ter)

[2467] NM_000251.2(MSH2):c.1165C>T(p.Arg389Ter)

[2468] NM_000249.3(MLH1):c.1943C>T(p.Pro648Leu)

[2469] NM_000249.3(MLH1):c.200G>A(p.Gly67Glu)

[2470] NM_000249.3(MLH1):c.793C>T(p.Arg265Cys)

[2471] NM_000249.3(MLH1):c.2059C>T(p.Arg687Trp)

[2472] NM_000249.3(MLH1):c.677G>A(p.Arg226Gln)

[2473] NM_000249.3(MLH1):c.2041G>A(p.Ala681Thr)

[2474] NM_000249.3(MLH1):c.1942C>T(p.Pro648Ser)

[2475] NM_000249.3(MLH1):c.676C>T(p.Arg226Ter)

[2476] NM_000251.2(MSH2):c.2038C>T(p.Arg680Ter)

[2477] NM_000179.2(MSH6):c.1483C>T(p.Arg495Ter)

[2478] NM_000179.2(MSH6):c.2194C>T(p.Arg732Ter)

[2479] NM_000179.2(MSH6):c.3103C>T(p.Arg1035Ter)

[2480] NM_000179.2(MSH6):c.892C>T(p.Arg298Ter)

[2481] NM_000249.3(MLH1):c.1459C>T(p.Arg487Ter)

[2482] NM_000249.3(MLH1):c.1731G>A(p.Ser577=)

[2483] NM_000249.3(MLH1):c.184C>T(p.Gln62Ter)

[2484] NM_000249.3(MLH1):c.1975C>T(p.Arg659Ter)

[2485] NM_000249.3(MLH1):c.199G>A(p.Gly67Arg)

[2486] NM_000251.2(MSH2):c.1076+1G>A

[2487] NM_000251.2(MSH2):c.1147C>T(p.Arg383Ter)

[2488] NM_000251.2(MSH2):c.181C>T(p.Gln61Ter)

[2489] NM_000251.2(MSH2):c.212-1G>A

[2490] NM_000251.2(MSH2):c.2131C>T(p.Arg711Ter)

[2491] NM_000535.6(PMS2):c.697C>T(p.Gln233Ter)

[2492] NM_000535.6(PMS2):c.1261C>T(p.Arg421Ter)

[2493] NM_000251.2(MSH2):c.2047G>A(p.Gly683Arg)

[2494] NM_000535.6(PMS2):c.400C>T(p.Arg134Ter)

[2495] NM_000535.6(PMS2):c.1927C>T(p.Gln643Ter)

[2496] NM_000179.2(MSH6):c.1444C>T(p.Arg482Ter)

[2497] NM_000179.2(MSH6):c.2731C>T(p.Arg911Ter)

[2498] NM_000535.6(PMS2):c.943C>T(p.Arg315Ter)

[2499] 参见表A。因此，本发明的一个方面涉及一种用于通过校正一种或多种致病性G至A或C至T突变/SNP、具体是存在于选自BCKDHA、BCKDHB、DBT和DLD的至少一个基因中的一种或多种致病性G至A或C至T突变/SNP、更具体是上述一种或多种致病性G至A或C至T突变/SNP来治疗或预防林奇综合征的方法。

[2500] 其他遗传疾病

[2501] 与另外的遗传疾病相关的致病性G至A或C至T突变/SNP也在ClinVar数据库中报告并在表A中公开，包括但不限于马凡氏综合征、贺勒氏综合征、糖原贮积病和囊性纤维化。因此，本发明的一个方面涉及一种用于校正与这些疾病中的任何一种相关的一种或多种致病性G至A或C至T突变/SNP的方法，如下所述。

[2502] 马凡氏综合征

[2503] 在一些实施方案中，本文所述的方法、系统和组合物用于校正与马凡氏综合征相关的一种或多种致病性G至A或C至T突变/SNP。在一些实施方案中，致病性突变/SNP存在于至少FBN1基因中，包括至少以下各项：

[2504] NM_000138.4(FBN1):c.1879C>T(p.Arg627Cys)

[2505] NM_000138.4(FBN1):c.1051C>T(p.Gln351Ter)

[2506] NM_000138.4(FBN1):c.184C>T(p.Arg62Cys)

[2507] NM_000138.4(FBN1):c.2855-1G>A

[2508] NM_000138.4(FBN1):c.3164G>A(p.Cys1055Tyr)

[2509] NM_000138.4(FBN1):c.368G>A(p.Cys123Tyr)

[2510] NM_000138.4(FBN1):c.4955G>A(p.Cys1652Tyr)

[2511] NM_000138.4(FBN1):c.7180C>T(p.Arg2394Ter)

[2512] NM_000138.4(FBN1):c.8267G>A(p.Trp2756Ter)

[2513] NM_000138.4(FBN1):c.1496G>A(p.Cys499Tyr)

[2514] NM_000138.4(FBN1):c.6886C>T(p.Gln2296Ter)

[2515] NM_000138.4(FBN1):c.3373C>T(p.Arg1125Ter)

[2516] NM_000138.4(FBN1):c.640G>A(p.Gly214Ser)

[2517] NM_000138.4(FBN1):c.5038C>T(p.Gln1680Ter)

[2518] NM_000138.4(FBN1):c.434G>A(p.Cys145Tyr)

[2519] NM_000138.4(FBN1):c.2563C>T(p.Gln855Ter)

[2520] NM_000138.4(FBN1):c.7466G>A(p.Cys2489Tyr)

[2521] NM_000138.4(FBN1):c.2089C>T(p.Gln697Ter)

[2522] NM_000138.4(FBN1):c.592C>T(p.Gln198Ter)

[2523] NM_000138.4(FBN1):c.6695G>A(p.Cys2232Tyr)

[2524] NM_000138.4(FBN1):c.6164-1G>A

[2525] NM_000138.4(FBN1):c.5627G>A(p.Cys1876Tyr)

[2526] NM_000138.4(FBN1):c.4061G>A(p.Trp1354Ter)

[2527] NM_000138.4(FBN1):c.1982G>A(p.Cys661Tyr)

[2528] NM_000138.4(FBN1):c.6784C>T(p.Gln2262Ter)

[2529] NM_000138.4(FBN1):c.409C>T(p.Gln137Ter)

[2530] NM_000138.4(FBN1):c.364C>T(p.Arg122Cys)

[2531] NM_000138.4(FBN1):c.3217G>A(p.Glu1073Lys)

[2532] NM_000138.4(FBN1):c.4460-8G>A

[2533] NM_000138.4(FBN1):c.4786C>T(p.Arg1596Ter)

[2534] NM_000138.4(FBN1):c.7806G>A(p.Trp2602Ter)

[2535] NM_000138.4(FBN1):c.247+1G>A

[2536] NM_000138.4(FBN1):c.2495G>A(p.Cys832Tyr)

[2537] NM_000138.4(FBN1):c.493C>T(p.Arg165Ter)

[2538] NM_000138.4(FBN1):c.5504G>A(p.Cys1835Tyr)

[2539] NM_000138.4(FBN1):c.5863C>T(p.Gln1955Ter)

[2540] NM_000138.4(FBN1):c.6658C>T(p.Arg2220Ter)

[2541] NM_000138.4(FBN1):c.7606G>A(p.Gly2536Arg)

[2542] NM_000138.4(FBN1):c.7955G>A(p.Cys2652Tyr)

[2543] NM_000138.4(FBN1):c.3037G>A(p.Gly1013Arg)

[2544] NM_000138.4(FBN1):c.8080C>T(p.Arg2694Ter)

[2545] NM_000138.4(FBN1):c.1633C>T(p.Arg545Cys)

[2546] NM_000138.4(FBN1):c.7205-1G>A

[2547] NM_000138.4(FBN1):c.4621C>T(p.Arg1541Ter)

[2548] NM_000138.4(FBN1):c.1090C>T(p.Arg364Ter)

[2549] NM_000138.4(FBN1):c.1585C>T(p.Arg529Ter)

[2550] NM_000138.4(FBN1):c.4781G>A(p.Gly1594Asp)

[2551] NM_000138.4(FBN1):c.643C>T(p.Arg215Ter)

[2552] NM_000138.4(FBN1):c.3668G>A(p.Cys1223Tyr)

[2553] NM_000138.4(FBN1):c.8326C>T(p.Arg2776Ter)

[2554] NM_000138.4(FBN1):c.6354C>T(p.Ile2118=)

[2555] NM_000138.4(FBN1):c.1468+5G>A

[2556] NM_000138.4(FBN1):c.1546C>T(p.Arg516Ter)

[2557] NM_000138.4(FBN1):c.4615C>T(p.Arg1539Ter)

[2558] NM_000138.4(FBN1):c.5368C>T(p.Arg1790Ter)

[2559] NM_000138.4(FBN1):c.1285C>T(p.Arg429Ter)

[2560] 参见表A。因此,本发明的一个方面涉及一种用于通过校正一种或多种致病性G至A或C至T突变/SNP、具体是存在于FBN1基因中的一种或多种致病性G至A或C至T突变/SNP、更具体是上述一种或多种致病性G至A或C至T突变/SNP来治疗或预防马凡氏综合征的方法。

[2561] 贺勒氏综合征

[2562] 在一些实施方案中,本文所述的方法、系统和组合物用于校正与贺勒氏综合征相关的一种或多种致病性G至A或C至T突变/SNP。在一些实施方案中,致病性突变/SNP存在于至少IDUA基因中,包括至少以下各项:

[2563] NM_000203.4(IDUA):c.972+1G>A

[2564] NM_000203.4(IDUA):c.1855C>T(p.Arg619Ter)

[2565] NM_000203.4(IDUA):c.152G>A(p.Gly51Asp)

[2566] NM_000203.4(IDUA):c.1205G>A(p.Trp402Ter)

[2567] NM_000203.4(IDUA):c.208C>T(p.Gln70Ter)

[2568] NM_000203.4(IDUA):c.1045G>A(p.Asp349Asn)

[2569] NM_000203.4(IDUA):c.1650+5G>A

[2570] 参见表A。因此,本发明的一个方面涉及一种用于通过校正一种或多种致病性G至A或C至T突变/SNP、具体是存在于IDUA基因中的一种或多种致病性G至A或C至T突变/SNP、更具体是上述一种或多种致病性G至A或C至T突变/SNP来治疗或预防贺勒氏综合征的方法。

[2571] 糖原贮积病

[2572] 在一些实施方案中,本文所述的方法、系统和组合物用于校正与糖原贮积病相关的一种或多种致病性G至A或C至T突变/SNP。在一些实施方案中,致病性突变/SNP存在于选自GAA、AGL、PHKB、PRKAG2、G6PC、PGAM2、GBE1、PYGM和PFKM的至少一个基因中,包括至少以下各项:

[2573] NM_000152.4(GAA):c.1927G>A(p.Gly643Arg)

[2574] NM_000152.4(GAA):c.2173C>T(p.Arg725Trp)

[2575] NM_000642.2(AGL):c.3980G>A(p.Trp1327Ter)

[2576] NM_000642.2(AGL):c.16C>T(p.Gln6Ter)

[2577] NM_000642.2(AGL):c.2039G>A(p.Trp680Ter)

[2578] NM_000293.2(PHKB):c.1546C>T(p.Gln516Ter)

[2579] NM_016203.3(PRKAG2):c.1592G>A(p.Arg531Gln)

[2580] NM_000151.3(G6PC):c.248G>A(p.Arg83His)

[2581] NM_000151.3(G6PC):c.724C>T(p.Gln242Ter)

[2582] NM_000151.3(G6PC):c.883C>T(p.Arg295Cys)

[2583] NM_000151.3(G6PC):c.247C>T(p.Arg83Cys)

[2584] NM_000151.3(G6PC):c.1039C>T(p.Gln347Ter)

[2585] NM_000152.4(GAA):c.1561G>A(p.Glu521Lys)

[2586] NM_000642.2(AGL):c.2590C>T(p.Arg864Ter)

[2587] NM_000642.2(AGL):c.3682C>T(p.Arg1228Ter)

[2588] NM_000642.2(AGL):c.118C>T(p.Gln40Ter)

[2589] NM_000642.2(AGL):c.256C>T(p.Gln86Ter)

[2590] NM_000642.2(AGL):c.2681+1G>A

[2591] NM_000642.2(AGL):c.2158-1G>A

[2592] NM_000290.3(PGAM2):c.233G>A(p.Trp78Ter)

[2593] NM_000152.4(GAA):c.1548G>A(p.Trp516Ter)

[2594] NM_000152.4(GAA):c.2014C>T(p.Arg672Trp)

[2595] NM_000152.4(GAA):c.546G>A(p.Thr182＝)

[2596] NM_000152.4(GAA):c.1802C>T(p.Ser601Leu)

[2597] NM_000152.4(GAA):c.1754+1G>A

[2598] NM_000152.4(GAA):c.1082C>T(p.Pro361Leu)

[2599] NM_000152.4(GAA):c.2560C>T(p.Arg854Ter)

[2600] NM_000152.4(GAA):c.655G>A(p.Gly219Arg)

[2601] NM_000152.4(GAA):c.1933G>A(p.Asp645Asn)

[2602] NM_000152.4(GAA):c.1979G>A(p.Arg660His)

[2603] NM_000152.4(GAA):c.1465G>A(p.Asp489Asn)

[2604] NM_000152.4(GAA):c.2512C>T(p.Gln838Ter)

[2605] NM_000158.3(GBE1):c.1543C>T(p.Arg515Cys)

[2606] NM_005609.3(PYGM):c.1726C>T(p.Arg576Ter)

[2607] NM_005609.3(PYGM):c.1827G>A(p.Lys609＝)

[2608] NM_005609.3(PYGM):c.148C>T(p.Arg50Ter)

[2609] NM_005609.3(PYGM):c.613G>A(p.Gly205Ser)

[2610] NM_005609.3(PYGM):c.1366G>A(p.Val456Met)

[2611] NM_005609.3(PYGM):c.1768+1G>A

[2612] NM_001166686.1(PFKM):c.450+1G>A

[2613] 参见表A。因此，本发明的一个方面涉及一种用于通过校正一种或多种致病性G至A或C至T突变/SNP、具体是存在于选自GAA、AGL、PHKB、PRKAG2、G6PC、PGAM2、GBE1、PYGM和PFKM的至少一个基因中的一种或多种致病性G至A或C至T突变/SNP、更具体是上述一种或多种致病性G至A或C至T突变/SNP来治疗或预防糖原贮积病的方法。

[2614] 囊性纤维化

[2615] 在一些实施方案中，本文所述的方法、系统和组合物用于校正与囊性纤维化相关的一种或多种致病性G至A或C至T突变/SNP。在一些实施方案中，致病性突变/SNP存在于CFTR基因中，包括至少以下各项：

[2616] NM_000492.3(CFTR):c.3712C>T(p.Gln1238Ter)

[2617] NM_000492.3(CFTR):c.3484C>T(p.Arg1162Ter)

[2618] NM_000492.3(CFTR):c.1766+1G>A

[2619] NM_000492.3(CFTR):c.1477C>T(p.Gln493Ter)

[2620] NM_000492.3(CFTR):c.2538G>A(p.Trp846Ter)

[2621] NM_000492.3(CFTR):c.2551C>T(p.Arg851Ter)

[2622] NM_000492.3(CFTR):c.3472C>T(p.Arg1158Ter)

[2623] NM_000492.3(CFTR):c.1475C>T(p.Ser492Phe)

[2624] NM_000492.3(CFTR):c.1679G>A(p.Arg560Lys)

[2625] NM_000492.3(CFTR):c.3197G>A(p.Arg1066His)

[2626] NM_000492.3(CFTR):c.3873+1G>A

[2627] NM_000492.3(CFTR):c.3196C>T(p.Arg1066Cys)

[2628] NM_000492.3(CFTR):c.2490+1G>A

[2629] NM_000492.3(CFTR):c.3718-1G>A

[2630] NM_000492.3(CFTR):c.171G>A(p.Trp57Ter)

[2631] NM_000492.3(CFTR):c.3937C>T(p.Gln1313Ter)

[2632] NM_000492.3(CFTR):c.274G>A(p.Glu92Lys)

[2633] NM_000492.3(CFTR):c.1013C>T(p.Thr338Ile)

[2634] NM_000492.3(CFTR):c.3266G>A(p.Trp1089Ter)

[2635] NM_000492.3(CFTR):c.1055G>A(p.Arg352Gln)

[2636] NM_000492.3(CFTR):c.1654C>T(p.Gln552Ter)

[2637] NM_000492.3(CFTR):c.2668C>T(p.Gln890Ter)

[2638] NM_000492.3(CFTR):c.3611G>A(p.Trp1204Ter)

[2639] NM_000492.3(CFTR):c.1585-8G>A

[2640] NM_000492.3(CFTR):c.223C>T(p.Arg75Ter)

[2641] NM_000492.3(CFTR):c.1680-1G>A

[2642] NM_000492.3(CFTR):c.349C>T(p.Arg117Cys)

[2643] NM_000492.3(CFTR):c.1203G>A(p.Trp401Ter)

[2644] NM_000492.3(CFTR):c.1240C>T(p.Gln414Ter)

[2645] NM_000492.3(CFTR):c.1202G>A(p.Trp401Ter)

[2646] NM_000492.3(CFTR):c.1209+1G>A

[2647] NM_000492.3(CFTR):c.115C>T(p.Gln39Ter)

[2648] NM_000492.3(CFTR):c.1116+1G>A

[2649] NM_000492.3(CFTR):c.1393-1G>A

[2650] NM_000492.3(CFTR):c.1573C>T(p.Gln525Ter)

[2651] NM_000492.3(CFTR):c.164+1G>A

[2652] NM_000492.3(CFTR):c.166G>A(p.Glu56Lys)

[2653] NM_000492.3(CFTR):c.170G>A(p.Trp57Ter)

[2654] NM_000492.3(CFTR):c.2053C>T(p.Gln685Ter)

[2655] NM_000492.3(CFTR):c.2125C>T(p.Arg709Ter)

[2656] NM_000492.3(CFTR):c.2290C>T(p.Arg764Ter)

[2657] NM_000492.3(CFTR):c.2353C>T(p.Arg785Ter)

[2658] NM_000492.3(CFTR):c.2374C>T(p.Arg792Ter)

[2659] NM_000492.3(CFTR):c.2537G>A(p.Trp846Ter)

[2660] NM_000492.3(CFTR):c.292C>T(p.Gln98Ter)

[2661] NM_000492.3(CFTR):c.2989-1G>A

[2662] NM_000492.3(CFTR):c.3293G>A(p.Trp1098Ter)

[2663] NM_000492.3(CFTR):c.4144C>T(p.Gln1382Ter)

[2664] NM_000492.3(CFTR):c.4231C>T(p.Gln1411Ter)

[2665] NM_000492.3(CFTR):c.4234C>T(p.Gln1412Ter)

[2666] NM_000492.3(CFTR):c.579+5G>A

[2667] NM_000492.3(CFTR):c.595C>T(p.His199Tyr)

[2668] NM_000492.3(CFTR):c.613C>T(p.Pro205Ser)

[2669] NM_000492.3(CFTR):c.658C>T(p.Gln220Ter)

[2670] NM_000492.3(CFTR):c.1117-1G>A

[2671] NM_000492.3(CFTR):c.3294G>A(p.Trp1098Ter)

[2672] NM_000492.3(CFTR):c.1865G>A(p.Gly622Asp)

[2673] NM_000492.3(CFTR):c.743+1G>A

[2674] NM_000492.3(CFTR):c.1679+1G>A

[2675] NM_000492.3(CFTR):c.1657C>T(p.Arg553Ter)

[2676] NM_000492.3(CFTR):c.1675G>A(p.Ala559Thr)

[2677] NM_000492.3(CFTR):c.165-1G>A

[2678] NM_000492.3(CFTR):c.200C>T(p.Pro67Leu)

[2679] NM_000492.3(CFTR):c.2834C>T(p.Ser945Leu)

[2680] NM_000492.3(CFTR):c.3846G>A(p.Trp1282Ter)

[2681] NM_000492.3(CFTR):c.1652G>A(p.Gly551Asp)

[2682] NM_000492.3(CFTR):c.4426C>T(p.Gln1476Ter)

[2683] NM_000492.3:c.3718-2477C>T

[2684] NM_000492.3(CFTR):c.2988+1G>A

[2685] NM_000492.3(CFTR):c.2657+5G>A

[2686] NM_000492.3(CFTR):c.2988G>A(p.Gln996=)

[2687] NM_000492.3(CFTR):c.274-1G>A

[2688] NM_000492.3(CFTR):c.3612G>A(p.Trp1204Ter)

[2689] NM_000492.3(CFTR):c.1646G>A(p.Ser549Asn)

[2690] NM_000492.3(CFTR):c.3752G>A(p.Ser1251Asn)

[2691] NM_000492.3(CFTR):c.4046G>A(p.Gly1349Asp)

[2692] NM_000492.3(CFTR):c.532G>A(p.Gly178Arg)

[2693] NM_000492.3(CFTR):c.3731G>A(p.Gly1244Glu)

[2694] NM_000492.3(CFTR):c.1651G>A(p.Gly551Ser)

[2695] NM_000492.3(CFTR):c.1585-1G>A

[2696] NM_000492.3(CFTR):c.1000C>T(p.Arg334Trp)

[2697] NM_000492.3(CFTR):c.254G>A(p.Gly85Glu)

[2698] NM_000492.3(CFTR):c.1040G>A(p.Arg347His)

[2699] NM_000492.3(CFTR):c.273+1G>A

[2700] 参见表A。因此,本发明的一个方面涉及一种用于通过校正一种或多种致病性G至A或C至T突变/SNP、具体是存在于CFTR基因中的一种或多种致病性G至A或C至T突变/SNP、更具体是上述一种或多种致病性G至A或C至T突变/SNP来治疗或预防囊性纤维化的方法。

[2701] 在一些实施方案中,本文所述的方法、系统和组合物用于校正被认为与家族性2乳腺-卵巢癌相关的致病性A至G(A>G)突变或SNP,其中致病性A>G突变或SNP位于BRCA2基因中(HGVS:U43746.1:n.7829+1G>A)。因此,本发明的另一方面涉及一种用于通过校正上述致病性A>G突变或SNP来治疗或预防家族性2乳腺-卵巢癌的方法。

[2702] 在一些实施方案中,本文所述的方法、系统和组合物用于校正被认为与遗传因子IX缺乏相关的致病性A至G(A>G)突变或SNP,其中致病性A>G突变或SNP位于F9基因中的GRCh38:ChrX:139537145,从而导致Arg至Gln取代。因此,本发明的另一方面涉及一种用于通过校正上述致病性A>G突变或SNP来治疗或预防遗传因子IX缺乏的方法。

[2703] 在一些实施方案中,本文所述的方法、系统和组合物用于校正被认为与β+地中海贫血、β地中海贫血和β重型地中海贫血相关的致病性A至G(A>G)突变或SNP,其中致病性A>G突变或SNP位于HBB基因中的GRCh38:Chr11:5226820。因此,本发明的另一方面涉及一种用于通过校正上述致病性A>G突变或SNP来治疗或预防β+地中海贫血、β地中海贫血和β重型地中海贫血的方法。

[2704] 在一些实施方案中,本文所述的方法、系统和组合物用于校正被认为与马凡氏综合征相关的致病性A至G(A>G)突变或SNP,其中致病性A>G突变或SNP位于FBN1基因中(IVS2DS,G-A,+1),如Yamamoto等人,J Hum Genet.2000;45(2):115-8中所报道。因此,本发明的另一方面涉及一种用于通过校正上述致病性A>G突变或SNP来治疗或预防马凡氏综合征的方法。

[2705] 在一些实施方案中,本文所述的方法、系统和组合物用于校正被认为与Wiskott-Aldrich综合征相关的致病性A至G(A>G)突变或SNP,其中致病性A>G突变或SNP位于WAS基因的前6位的-1位置(IVS6AS,G-A,-1),如Kwan等人,(1995)中所报道。因此,本发明的另一方面涉及一种用于通过校正上述致病性A>G突变或SNP来治疗或预防Wiskott-Aldrich综合征的方法。

[2706] 在一些实施方案中,本文所述的方法、系统和组合物用于校正被认为与囊性纤维化相关的致病性A至G(A>G)突变或SNP,其中致病性A>G突变或SNP位于CFTR基因中的GRCh38:Chr7:117590440。因此,本发明的另一方面涉及一种用于通过校正上述致病性A>G突变或SNP来治疗或预防囊性纤维化的方法。

[2707] 在一些实施方案中,本文所述的方法、系统和组合物用于校正被认为与囊性纤维化和遗传性胰腺炎相关的致病性A至G(A>G)突变或SNP,其中致病性A>G突变或SNP位于CFTR基因中的GRCh38:Chr7:117606754。因此,本发明的另一方面涉及一种用于通过校正上述致病性A>G突变或SNP来治疗或预防囊性纤维化和遗传性胰腺炎的方法。

[2708] 在一些实施方案中，本文所述的方法、系统和组合物用于校正被认为与囊性纤维化相关的致病性A至G(A>G)突变或SNP，其中致病性A>G突变或SNP位于CFTR基因中的GRCh38:Chr7:117587738。因此，本发明的另一方面涉及一种用于通过校正上述致病性A>G突变或SNP来治疗或预防囊性纤维化的方法。

[2709] 在一些实施方案中，本文所述的方法、系统和组合物用于校正被认为与蒂尔柯综合征和林奇综合征相关的致病性A至G(A>G)突变或SNP，其中致病性A>G突变或SNP位于MSH2基因中的GRCh38:Chr2:47470964。因此，本发明的另一方面涉及一种用于通过校正上述致病性A>G突变或SNP来治疗或预防蒂尔柯综合征和林奇综合征的方法。

[2710] 在一些实施方案中，本文所述的方法、系统和组合物用于校正被认为与囊性纤维化相关的致病性A至G(A>G)突变或SNP，其中致病性A>G突变或SNP位于CFTR基因中的GRCh38:Chr7:117642437。因此，本发明的另一方面涉及一种用于通过校正上述致病性A>G突变或SNP来治疗或预防囊性纤维化的方法。

[2711] 在一些实施方案中，本文所述的方法、系统和组合物用于校正被认为与林奇综合征II和林奇综合征相关的致病性A至G(A>G)突变或SNP，其中致病性A>G突变或SNP位于MLH1基因中的GRCh38:Chr3:37001058。因此，本发明的另一方面涉及一种用于通过校正上述致病性A>G突变或SNP来治疗或预防林奇综合征II和林奇综合征的方法。

[2712] 在一些实施方案中，本文所述的方法、系统和组合物用于校正被认为与囊性纤维化相关的致病性A至G(A>G)突变或SNP，其中致病性A>G突变或SNP位于CFTR基因中的GRCh38:Chr7:117642594。因此，本发明的另一方面涉及一种用于通过校正上述致病性A>G突变或SNP来治疗或预防囊性纤维化的方法。

[2713] 在一些实施方案中，本文所述的方法、系统和组合物用于校正被认为与囊性纤维化相关的致病性A至G(A>G)突变或SNP，其中致病性A>G突变或SNP位于CFTR基因中的GRCh38:Chr7:117592658。因此，本发明的另一方面涉及一种用于通过校正上述致病性A>G突变或SNP来治疗或预防囊性纤维化的方法。

[2714] 在一些实施方案中，本文所述的方法、系统和组合物用于校正被认为与家族性1乳腺-卵巢癌、遗传性乳腺和卵巢癌综合征和遗传性癌症易感综合征相关的致病性A至G(A>G)突变或SNP，其中致病性A>G突变或SNP位于BRCA1基因中的GRCh38:Chr17:43057051。因此，本发明的另一方面涉及一种用于通过校正上述致病性A>G突变或SNP来治疗或预防家族性1乳腺-卵巢癌、遗传性乳腺和卵巢癌综合征和遗传性癌症易感综合征的方法。

[2715] 在一些实施方案中，本文所述的方法、系统和组合物用于校正被认为与二氢嘧啶脱氢酶缺乏、希施斯普龙病1、氟尿嘧啶响应、嘧啶类似物响应-毒性/ADR、卡培他滨响应-毒性/ADR、氟尿嘧啶响应-毒性/ADR、喃氟啶响应-毒性/ADR相关的致病性A至G(A>G)突变或SNP，其中致病性A>G突变或SNP位于DPYD基因中的GRCh38:Chr1:97450058。因此，本发明的另一方面涉及一种用于通过校正上述致病性A>G突变或SNP来治疗或预防二氢嘧啶脱氢酶缺乏、希施斯普龙病1、氟尿嘧啶响应、嘧啶类似物响应-毒性/ADR、卡培他滨响应-毒性/ADR、氟尿嘧啶响应-毒性/ADR、喃氟啶响应-毒性/ADR的方法。

[2716] 在一些实施方案中，本文所述的方法、系统和组合物用于校正被认为与林奇综合征相关的致病性A至G(A>G)突变或SNP，其中致病性A>G突变或SNP位于MSH2基因中的GRCh38:Chr2:47478520。因此，本发明的另一方面涉及一种用于通过校正上述致病性A>G突

变或SNP来治疗或预防林奇综合征的方法。

[2717] 在一些实施方案中，本文所述的方法、系统和组合物用于校正被认为与林奇综合征相关的致病性A至G(A>G)突变或SNP，其中致病性A>G突变或SNP位于MLH1基因中的GRCh38:Chr3:37011819。因此，本发明的另一方面涉及一种用于通过校正上述致病性A>G突变或SNP来治疗或预防林奇综合征的方法。

[2718] 在一些实施方案中，本文所述的方法、系统和组合物用于校正被认为与林奇综合征相关的致病性A至G(A>G)突变或SNP，其中致病性A>G突变或SNP位于MLH1基因中的GRCh38:Chr3:37014545。因此，本发明的另一方面涉及一种用于通过校正上述致病性A>G突变或SNP来治疗或预防林奇综合征的方法。

[2719] 在一些实施方案中，本文所述的方法、系统和组合物用于校正被认为与林奇综合征相关的致病性A至G(A>G)突变或SNP，其中致病性A>G突变或SNP位于MLH1基因中的GRCh38:Chr3:37011867。因此，本发明的另一方面涉及一种用于通过校正上述致病性A>G突变或SNP来治疗或预防林奇综合征的方法。

[2720] 在一些实施方案中，本文所述的方法、系统和组合物用于校正被认为与林奇综合征相关的致病性A至G(A>G)突变或SNP，其中致病性A>G突变或SNP位于MLH1基因中的GRCh38:Chr3:37025636。因此，本发明的另一方面涉及一种用于通过校正上述致病性A>G突变或SNP来治疗或预防林奇综合征的方法。

[2721] 在一些实施方案中，本文所述的方法、系统和组合物用于校正被认为与林奇综合征相关的致病性A至G(A>G)突变或SNP，其中致病性A>G突变或SNP位于MLH1基因中的GRCh38:Chr3:37004475。因此，本发明的另一方面涉及一种用于通过校正上述致病性A>G突变或SNP来治疗或预防林奇综合征的方法。

[2722] 在一些实施方案中，本文所述的方法、系统和组合物用于校正被认为与林奇综合征和遗传性癌症易感综合征相关的致病性A至G(A>G)突变或SNP，其中致病性A>G突变或SNP位于MSH2基因中的GRCh38:Chr2:47416430。因此，本发明的另一方面涉及一种用于通过校正上述致病性A>G突变或SNP来治疗或预防林奇综合征和遗传性癌症易感综合征的方法。

[2723] 在一些实施方案中，本文所述的方法、系统和组合物用于校正被认为与林奇综合征和遗传性癌症易感综合征相关的致病性A至G(A>G)突变或SNP，其中致病性A>G突变或SNP位于MSH2基因中的GRCh38:Chr2:47408400。因此，本发明的另一方面涉及一种用于通过校正上述致病性A>G突变或SNP来治疗或预防林奇综合征和遗传性癌症易感综合征的方法。

[2724] 在一些实施方案中，本文所述的方法、系统和组合物用于校正被认为与林奇综合征和遗传性癌症易感综合征相关的致病性A至G(A>G)突变或SNP，其中致病性A>G突变或SNP位于MSH1基因中的GRCh38:Chr3:36996710。因此，本发明的另一方面涉及一种用于通过校正上述致病性A>G突变或SNP来治疗或预防林奇综合征和遗传性癌症易感综合征的方法。

[2725] 在一些实施方案中，本文所述的方法、系统和组合物用于校正被认为与家族性1乳腺-卵巢癌相关的致病性A至G(A>G)突变或SNP，其中致病性A>G突变或SNP位于BRCA1基因中的GRCh38:Chr17:43067696。因此，本发明的另一方面涉及一种用于通过校正上述致病性A>G突变或SNP来治疗或预防家族性1乳腺-卵巢癌的方法。

[2726] 在一些实施方案中，本文所述的方法、系统和组合物用于校正被认为与家族性2乳腺-卵巢癌和遗传性乳腺和卵巢癌综合征相关的致病性A至G(A>G)突变或SNP，其中致病性A

>G突变或SNP位于BRCA2基因中的GRCh38:Chr13:32356610。因此本发明的另一方面涉及一种用于通过校正上述致病性A>G突变或SNP来治疗或预防家族性2乳腺-卵巢癌和遗传性乳腺和卵巢癌综合征的方法。

[2727] 在一些实施方案中，本文所述的方法、系统和组合物用于校正被认为与原发性扩张型心肌病和原发性家族性肥厚型心肌病相关的致病性A至G(A>G)突变或SNP，其中致病性A>G突变或SNP位于MYH7基因中的GRCh38:Chr14:23419993。因此，本发明的另一方面涉及一种用于通过校正上述致病性A>G突变或SNP来治疗或预防原发性扩张型心肌病和原发性家族性肥厚型心肌病的方法。

[2728] 在一些实施方案中，本文所述的方法、系统和组合物用于校正被认为与原发性家族性肥厚型心肌病、camptocormism和肥厚型心肌病相关的致病性A至G(A>G)突变或SNP，其中致病性A>G突变或SNP位于MYH7基因中的GRCh38:Chr14:23415225。因此，本发明的另一方面涉及一种用于通过校正上述致病性A>G突变或SNP来治疗或预防原发性家族性肥厚型心肌病、camptocormism和肥厚型心肌病的方法。

[2729] 在一些实施方案中，本文所述的方法、系统和组合物用于校正被认为与家族性乳腺癌、家族性2乳腺-卵巢癌、遗传性乳腺和卵巢癌综合征和遗传性癌症易感综合征相关的致病性A至G(A>G)突变或SNP，其中致病性A>G突变或SNP位于BRCA2基因中的GRCh38:Chr13:32357741。因此，本发明的另一方面涉及一种用于通过校正上述致病性A>G突变或SNP来治疗或预防家族性乳腺癌、家族性2乳腺-卵巢癌、遗传性乳腺和卵巢癌综合征和遗传性癌症易感综合征的方法。

[2730] 在一些实施方案中，本文所述的方法、系统和组合物用于校正被认为与原发性扩张型心肌病、肥厚型心肌病、心肌病和左室心肌致密化不全相关的致病性A至G(A>G)突变或SNP，其中致病性A>G突变或SNP位于MYH7基因中的GRCh38:Chr14:23431584。因此，本发明的另一方面涉及一种用于通过校正上述致病性A>G突变或SNP来治疗或预防原发性扩张型心肌病、肥厚型心肌病、心肌病和左室心肌致密化不全的方法。

[2731] 在一些实施方案中，本文所述的方法、系统和组合物用于校正被认为与家族性1乳腺-卵巢癌、遗传性乳腺和卵巢癌综合征和遗传性癌症易感综合征相关的致病性A至G(A>G)突变或SNP，其中致病性A>G突变或SNP位于BRCA1基因中的GRCh38:Chr17:43067607。因此，本发明的另一方面涉及一种用于通过校正上述致病性A>G突变或SNP来治疗或预防家族性1乳腺-卵巢癌、遗传性乳腺和卵巢癌综合征和遗传性癌症易感综合征的方法。

[2732] 在一些实施方案中，本文所述的方法、系统和组合物用于校正被认为与家族性1乳腺-卵巢癌、遗传性乳腺和卵巢癌综合征、遗传性癌症易感综合征和乳腺癌相关的致病性A至G(A>G)突变或SNP，其中致病性A>G突变或SNP位于BRCA1基因中的GRCh38:Chr17:43047666。因此，本发明的另一方面涉及一种用于通过校正上述致病性A>G突变或SNP来治疗或预防家族性1乳腺-卵巢癌、遗传性乳腺和卵巢癌综合征、遗传性癌症易感综合征和乳腺癌的方法。

[2733] 在一些实施方案中，本文所述的方法、系统和组合物用于校正被认为与家族性2乳腺-卵巢癌、遗传性乳腺和卵巢癌综合征和遗传性癌症易感综合征相关的致病性A至G(A>G)突变或SNP，其中致病性A>G突变或SNP位于BRCA2基因中的GRCh38:Chr13:32370558。因此，本发明的另一方面涉及一种用于通过校正上述致病性A>G突变或SNP来治疗或预防家族性2

乳腺-卵巢癌、遗传性乳腺和卵巢癌综合征和遗传性癌症易感综合征的方法。

[2734] 在一些实施方案中，本文所述的方法、系统和组合物用于校正被认为与家族性1乳腺-卵巢癌、遗传性乳腺和卵巢癌综合征、遗传性癌症易感综合征和乳腺癌相关的致病性A至G(A>G)突变或SNP，其中致病性A>G突变或SNP位于BRCA1基因中的GRCh38:Chr17:43074330。因此，本发明的另一方面涉及一种用于通过校正上述致病性A>G突变或SNP来治疗或预防家族性1乳腺-卵巢癌、遗传性乳腺和卵巢癌综合征、遗传性癌症易感综合征和乳腺癌的方法。

[2735] 在一些实施方案中，本文所述的方法、系统和组合物用于校正被认为与家族性1乳腺-卵巢癌、遗传性乳腺和卵巢癌综合征和遗传性癌症易感综合征相关的致病性A至G(A>G)突变或SNP，其中致病性A>G突变或SNP位于BRCA1基因中的GRCh38:Chr17:43082403。因此，本发明的另一方面涉及一种用于通过校正上述致病性A>G突变或SNP来治疗或预防家族性1乳腺-卵巢癌、遗传性乳腺和卵巢癌综合征和遗传性癌症易感综合征的方法。

[2736] 在一些实施方案中，本文所述的方法、系统和组合物用于校正被认为与囊性纤维化和遗传性胰腺炎相关的致病性C至T(C>T)突变或SNP，其中致病性C>T突变或SNP位于CFTR基因中的GRCh38:Chr7:117639961。因此，本发明的另一方面涉及一种用于通过校正上述致病性C>T突变或SNP来治疗或预防囊性纤维化和遗传性胰腺炎的方法。

[2737] 在一些实施方案中，本文所述的方法、系统和组合物用于校正被认为与家族性2乳腺-卵巢癌相关的致病性C至T(C>T)突变或SNP，其中致病性C>T突变或SNP位于BRCA2基因中的GRCh38:Chr13:32336492。因此，本发明的另一方面涉及一种用于通过校正上述致病性C>T突变或SNP来治疗或预防家族性2乳腺-卵巢癌的方法。

[2738] 在一些实施方案中，本文所述的方法、系统和组合物用于校正被认为与家族性1乳腺-卵巢癌相关的致病性C至T(C>T)突变或SNP，其中致病性C>T突变或SNP位于BRCA1基因中的GRCh38:Chr17:43063365。因此，本发明的另一方面涉及一种用于通过校正上述致病性C>T突变或SNP来治疗或预防家族性1乳腺-卵巢癌的方法。

[2739] 在一些实施方案中，本文所述的方法、系统和组合物用于校正被认为与家族性1乳腺-卵巢癌相关的致病性C至T(C>T)突变或SNP，其中致病性C>T突变或SNP位于BRCA1基因中的GRCh38:Chr17:43093613。因此，本发明的另一方面涉及一种用于通过校正上述致病性C>T突变或SNP来治疗或预防家族性1乳腺-卵巢癌的方法。

[2740] 在一些实施方案中，本文所述的方法、系统和组合物用于校正被认为与家族性乳腺癌和家族性1乳腺-卵巢癌相关的致病性C至T(C>T)突变或SNP，其中致病性C>T突变或SNP位于BRCA1基因的GRCh38:Chr17:43093931。因此，本发明的另一方面涉及一种用于通过校正上述致病性C>T突变或SNP来治疗或预防家族性乳腺癌和家族性1乳腺-卵巢癌的方法。

[2741] 在一些实施方案中，本文所述的方法、系统和组合物用于校正被认为与家族性肥厚型心肌病1、原发性家族性肥厚型心肌病和肥厚型心肌病相关的致病性C至T(C>T)突变或SNP，其中致病性C>T突变或SNP位于MYH7基因的GRCh38:Chr14:23429279。因此，本发明的另一方面涉及一种用于通过校正上述致病性C>T突变或SNP来治疗或预防家族性肥厚型心肌病1、原发性家族性肥厚型心肌病和肥厚型心肌病的方法。

[2742] 在一些实施方案中，本文所述的方法、系统和组合物用于校正被认为与家族性2乳腺-卵巢癌、遗传性乳腺和卵巢癌综合征和遗传性癌症易感综合征相关的致病性C至T(C>T)

突变或SNP,其中致病性C>T突变或SNP位于BRCA2基因的GRCh38:Chr13:32356472。因此,本发明的另一方面涉及一种用于通过校正上述致病性C>T突变或SNP来治疗或预防家族性2乳腺-卵巢癌、遗传性乳腺和卵巢癌综合征和遗传性癌症易感综合征的方法。

[2743] 在一些实施方案中,本文所述的方法、系统和组合物用于校正被认为与家族性肥厚型心肌病1、原发性家族性肥厚型心肌病、家族性限制型心肌病和肥厚型心肌病相关的致病性C至T(C>T)突变或SNP,其中致病性C>T突变或SNP位于MYH7基因中的GRCh38:Chr14:23429005。因此,本发明的另一方面涉及一种用于通过校正上述致病性C>T突变或SNP来治疗或预防家族性肥厚型心肌病1、原发性家族性肥厚型心肌病、家族性限制型心肌病和肥厚型心肌病的方法。

[2744] 在ClinVar数据库中发现了另外的致病性A>G突变和SNP。因此,本公开的另一方面涉及使用本文所述的方法、系统和组合物来校正ClinVar中列出的致病性A>G突变或SNP,以治疗或预防与此相关的疾病或病症。

[2745] 在ClinVar数据库中还发现了另外的致病性C>T突变和SNP。因此,本公开的另一方面涉及使用本文所述的方法、系统和组合物来校正ClinVar中列出的致病性C>T突变或SNP,以治疗或预防与此相关的疾病或病症。可使用本文公开的实施方案解决的其他T突变或SNPS在一起提交的标题为"Clin_var_pathogenic_SNPS_TC_txt"的ASCII文本文件中的表格中列出。

[2746] 在以下实施例中进一步描述了本发明,这些实施例不限制权利要求中描述的本发明的范围。

[2747] 实施例

[2748] 实施例1

[2749] 腺嘌呤脱氨酶(AD)能够在双链RNA的特定位点使腺嘌呤脱氨。

[2750] 一些AD可以在DNA-RNAn RNA双链体上实现腺嘌呤脱氨这一事实(例如,Zheng等人,Nucleic Acids Research 2017)提供了一个独特的机会,可以通过利用在R环(通过无活性的Cas13在RNA指导的DNA结合期间形成)中的指导RNA及其互补DNA靶标之间形成的RNA双链体来开发RNA指导的AD。通过使用无活性的Cas13募集AD,AD酶随后将作用于RNA-DNAn RNA双链体中的腺嘌呤。

[2751] 在一个实施方案中,使用以下突变获得无活性的Cas13,诸如Cas13b:R116A、H121A、R1177A和H1182A。为了提高AD编辑的效率,使用突变的ADAR,诸如包含突变E488Q的突变的hADAR2d。

[2752] 将AD募集到特定基因座的设计:

[2753] 1.将NLS标记的无活性的Cas13在N末端或C末端与AD融合。使用各种接头,包括柔性接头诸如GSG5或不那么柔性的接头诸如LEPGEKPYKCPECGKSFSQSGALTRHQRTHTR(SEQ ID No.11)。

[2754] 2.用适配体诸如MS2结合位点修饰指导RNA支架(例如,Konermann等人,Nature 2015)。将NLS标记的AD-MS2结合蛋白融合体与(NLS标记的无活性的或Cas13b)和对应的指导RNA一起引入靶细胞中。

[2755] 3.将AD插入NLS标记的无活性或切口酶Cas13的内部环中。

[2756] RNA指导物的设计:

[2757] 1.设计长度与Cas13蛋白的天然指导序列相对应的指导序列,以靶向目标RNA。

[2758] 2.将长于标准长度的RNA指导物用于在蛋白质-引导RNA-靶DNA复合物之外形成RNA双链体。

[2759] 对于这些RNA指导物设计中的每一者,将RNA上与靶RNA链上的腺嘌呤相对的碱基指定为C而不是U。

[2760] AD的选择和设计:

[2761] 使用许多AD,并且每个AD的活性水平有所不同。这些AD为:

[2762] 1.人ADAR(hADAR1、hADAR2、hADAR3)

[2763] 2.普通鱿鱼章鱼ADAR

[2764] 3.鱿鱼乌贼ADARS;乳白近岸鱿鱼ADARS

[2765] ADAT(人ADAT、果蝇ADAT)

[2766] 也可以使用突变来增加ADAR与DNA-RNAn RNA双链体反应的活性。例如,对于人ADAR基因,使用hADAR1d(E1008Q)或hADAR2d(E488Q)突变来增加其对DNA-RNA靶标的活性。

[2767] 每个ADAR的序列上下文要求水平有所不同。例如,对于hADAR1d(E1008Q),tAg和aAg位点被有效地脱氨,而aAt和cAc的编辑效率较低,并且gAa和gAc被甚至更少地编辑。然而,上下文要求因不同的ADAR而异。

[2768] 图1中提供了该系统一种版本的示意图。图4中提供了示例AD蛋白的氨基酸序列。

[2769] 本文例示性地描述的实施方案可在不存在本文未具体公开的任何一个或多个要素、一个或多个限制的情况下适当地实践。因此,例如,术语"包含"、"包括"、"含有"等应被广泛且没有限制地理解。另外,本文采用的术语和表达已被用作描述性而非限制性的术语,并且无意于使用此类术语和表达来排除所示出和所描述的特征或其部分的任何等同形式,但是应当认识到在要求保护的技术的范围内可以进行各种修改。另外,短语"基本上由……组成"将被理解为包括具体叙述的那些要素以及不会实质性地影响要求保护的技术的基本和新颖特征的那些附加要素。短语"由……组成"不包括未指定的任何要素。

[2770] 本公开在本申请中描述的特定实施方案方面不受限制。如对本领域技术人员将显而易见的,可以在不脱离其精神和范围的情况下进行许多修改和变化。根据前述描述,除了本文列举的那些之外,本公开范围内在功能上等效的方法和组合物对于本领域技术人员将是显而易见的。此类修改和变化旨在落入所附权利要求的范围内。本公开仅由所附权利要求的术语以及这些权利要求所赋予的等同物的全部范围来限制。应当理解,本公开不限于特定的方法、试剂、化合物组成或生物系统,它们当然可以变化。还应当理解,本文所使用的术语仅出于描述特定实施方案的目的,而无意于进行限制。

[2771] 另外,在根据马库什群组(Markush group)描述本公开的特征或方面的情况下,本领域技术人员将认识到,由此也根据马库什群组的任何单个成员或成员的子组描述本公开。

[2772] 如本领域技术人员将理解的,出于任何和所有目的,特别是在提供书面描述方面,本文公开的所有范围还涵盖任何和所有可能的子范围及其子范围的组合。任何列出的范围可以容易地被识别为充分描述并且使得相同的范围能够被分解成至少相等的二分之一、三分之一、四分之一、五分之一、十分之一等。作为非限制性实例,本文讨论的每个范围可以容易地被分解成下三分之一、中间三分之一和上三分之一等。如本领域技术人员还将理解,所

有语言诸如“多达”、“至少”、“大于”、“小于”等包括所列举的数字，并且是指随后可以被分解成如上所述的子范围的范围。最后，本领域技术人员将理解，范围包括每个单独的成员。

[2773] 本说明书中提及的所有出版物、专利申请、发布的专利和其他文件均通过引用并入本文，就好像每个单独的出版物、专利申请、发布的专利或其他文件被明确且单独地指出全文通过引用并入一样。在与本公开中的定义相矛盾的程度上，排除通过引用并入的文本中包含的定义。

[2774] 实施例2

[2775] Cluc/Gluc平铺用于Cas13a/Cas13b干扰

[2776] 为了比较Cas13a和Cas13b之间的敲低效率，分别用24个或96个指导物对海萤和高斯荧光素酶基因进行平铺(图10)。指导物与Cas13a和Cas13b匹配，并显示出增加的对Cas13b的敲低效率，其中每个基因的所有指导物(一个指导物除外)都对Cas13b显示出较高效率。

[2777] 通过NGS进行ADAR编辑定量

[2778] 通过设计具有阻止荧光素酶表达的提前终止密码子UAG的荧光素酶报道子来测试Cas13b-ADAR2RNA编辑效率(图11A-11C)。设计7个不同长度的指导物并将它们相对于UAG终止密码子定位，这些指导物都包含与UAG中的针对A的C错配。已知C错配会在编辑位点产生空泡，这是ADAR催化结构域所偏爱的。通过Cas13b-ADAR2进行RNA编辑会将UAG转换为UIG(UGG)，这就引入色氨酸而不是终止密码子，并允许进行翻译。指导物和Cas13b12-ADAR2融合体在HEK293FT细胞中表达将荧光素酶表达恢复到不同水平，其中指导物5发生最大的恢复(图11B)。一般来讲，随着编辑位点移动越来越远离crRNA的3’端(正向重复序列所在位置，并因此为蛋白质结合的位置)，从指导物1-5进行的编辑水平不断增加。这可能表明crRNA:靶标双链体中与蛋白质结合的部分是ADAR催化结构域难以接近的。指导物5、6和7显示出最大量的活性，因为编辑位点位于指导物的远离DR/蛋白质结合区域的远端，并且因为它们的指导物长得多，从而产生更长的RNA双链体，这是ADAR所偏爱的。当编辑位点在RNA双链体中间时，ADAR活性最佳。将荧光素酶活性的相对表达归一化至非靶向指导条件。

[2779] 对这些样品进行测序以精确定量RNA编辑效率(图11C)。编辑效率在指导物标签旁边的括号中列出。总体而言，通过测序鉴定的编辑百分比与图11B中所见的荧光素酶表达恢复的相对水平相匹配。指导物5显示出最多的RNA编辑，在中靶A处的G转化率为45％。在一些情况下，该区域中存在少量的脱靶A-G编辑。这些可通过在指导序列中引入G错配来减少，G错配不是ADAR催化结构域所偏爱的。

[2780] 除了编辑荧光素酶报告基因转录物外，还设计了指导物来编辑KRAS和PPIB转录物中的框外UAG位点，其中两个指导物靶向每个转录物(图12)。这些指导物的设计原理与上述指导物5相同(45nt的间隔子，其编辑位点距3’DR的距离为27nt，和具有针对编辑位点腺苷的C错配)。KRAS指导物能够在中靶腺苷处实现6.5％和13.7％的编辑，并且PPIB指导物能够实现7.7％和9.2％的编辑。这些指导物中的一些也存在一些脱靶，这可以通过针对与附近可能存在的脱靶腺苷在间隔子中设计G错配来减少。脱靶似乎发生在靶腺苷的双链体区域3’。

[2781] 来自RNA Seq的Cas13a/b+shRNA特异性

[2782] 为了确定Cas13b12敲低的特异性，对转录组中所有mRNA进行RNA测序(图13A)。将

靶向Gluc和KRAS的指导物的敲低与非靶向指导物进行比较，并且发现Cas13a2和Cas13b12具有靶转录物的特异性敲低(图13A中的红点)，而shRNA具有很多脱靶，如较大的分布差异所证明。图13B中示出了这些条件中的每一个的重要脱靶的数量。对于变化大于2倍或小于0.8倍的任何脱靶转录物，通过有FDR校正(p<0.01)的t检验测量显著脱靶。与针对shRNA条件发现的数百个脱靶相比，Cas13a和Cas13b条件的脱靶极少。每个条件的敲低效率在图13C中示出。

[2783] 用于减少脱靶的错配特异性(A:A或A:G)

[2784] 为了减少靶腺苷编辑位点附近的腺苷处的脱靶，设计了具有针对潜在脱靶腺苷的G或A错配的指导物(图14和下表)。G或A错配不利于ADAR催化结构域的活性。

| 名称 | 指导物 |
|---|---|
| 具有 C 错配的荧光素酶指导物 WT (SEQ ID No. 162) | catagaatgttctaaaCCAtcctgcggcctctactctgcattcaa |
| 具有 C 错配和 1 个 G MM 的荧光素酶指导物 WT (SEQ ID No. 163) | catagaatgttcGaaaCCAtcctgcggcctctactctgcattcaa |
| 具有 C 错配和 2 个 G MM 的荧光素酶指导物 WT (SEQ ID No. 164) | catagaatgGtcGaaaCCAtcctgcggcctctactctgcattcaa |
| 具有 C 错配和 1 个 G MM 的荧光素酶指导物 WT (SEQ ID No. 165) | catagaatgttcAaaaCCAtcctgcggcctctactctgcattcaa |
| 具有 C 错配和 2 个 G MM 的荧光素酶指导物 WT (SEQ ID No. 166) | catagaatgAtcAaaaCCAtcctgcggcctctactctgcattcaa |
| 具有 C 错配的 KRAS 指导物 WT (SEQ ID No. 167) | ggtttctccatcaattacCacttgcttcctgtaggaatcctctatt |
| [2785] 具有 C 错配和 1 个 G MM 的 KRAS 指导物(SEQ ID No. 168) | ggtttctccatcaatGacCacttgcttcctgtaggaatcctctatt |
| 具有 C 错配和 2 个 G MM 的 KRAS 指导物(SEQ ID No. 169) | ggtttctccatcaaGGacCacttgcttcctgtaggaatcctctatt |
| 具有 C 错配和 1 个 A MM 的 KRAS 指导物(SEQ ID No. 170) | ggtttctccatcaatAacCacttgcttcctgtaggaatcctctatt |
| 具有 C 错配和 2 个 A MM 的 KRAS 指导物(SEQ ID No. 171) | ggtttctccatcaaAAacCacttgcttcctgtaggaatcctctatt |
| 具有 C 错配的 PPIB 指导物 WT (SEQ ID No. 172) | gcctttctctcctgtagcCaaggccacaaaattatccactgttttt |
| 具有 C 错配和 1 个 G MM 的 PPIB 指导物 WT (SEQ ID No. 173) | gcctttctctcctgGagcCaaggccacaaaattatccactgttttt |
| 具有 C 错配和 1 个 A MM 的 PPIB 指导物 WT (SEQ ID No. 174) | gcctttctctcctgAagcCaaggccacaaaattatccactgttttt |

[2786] 上表中的指导物被设计成具有针对要编辑的中靶腺苷的C错配，以及针对已知脱靶位点(基于上述RNA测序)的G或A错配。间隔子中的错配被大写。

[2787] 中靶活性的错配

[2788] 先前对ADAR2的催化结构域的研究表明，与靶标A相对的不同碱基可以影响肌苷编辑的量(Zheng等人，(2017)，Nucleic Acid Research，45(6):3369-3377)。具体地，发现U和C与天然的ADAR编辑的A相对，而G和A则不然。为了测试与经编辑的A和G错配是否可以用于抑制ADAR活性，基于荧光素酶报告子测定，用所有其他3个可能的碱基对已知对C错配有活性的指导物进行测试(图16)。通过评估荧光素酶活性来对相对活性进行定量。下表提供了指导序列。

[2789]

| 错配 | 指导序列 |
|---|---|
| 错配-C (SEQ ID No. 175) | GcatagaatgttctaaaCCAtcctgcgg cctctactctgcattcaa |
| 错配-G (SEQ ID No. 176) | GcatagaatgttctaaaCGAtcctgcgg cctctactctgcattcaa |
| 错配-T (SEQ ID No. 177) | GcatagaatgttctaaaCTAtcctgcgg cctctactctgcattcaa |
| 错配-A (SEQ ID No. 178) | GcatagaatgttctaaaCAAtcctgcgg cctctactctgcattcaa |

[2790] 通过化学修饰gRNA改善编辑并减少脱靶修饰

[2791] 如Vogel等人(2014)，Angew Chem Int Ed，53:6267-6271，doi:10.1002/anie.201402634)中例示经化学修饰的gRNA降低了脱靶活性并提高了中靶效率。2'-O-甲基和硫代磷酸酯修饰的指导RNA一般可提高细胞中的编辑效率。

[2792] 基序偏好

[2793] ADAR已知对经编辑的A的任一侧上的相邻核苷酸表现出偏好(www.nature.com/nsmb/journal/v23/n5/full/nsmb.3203.html，Matthews等人(2017)，Nature Structural Mol Biol，23(5):426-433)。通过用经靶向的A周围的可变碱基靶向海萤荧光素酶转录物来系统地测试该偏好(图17)。

[2794] 较大的空泡可增强RNA编辑效率

[2795] 为了提高非优选5'或3'相邻碱基的RNA编辑效率，在相邻碱基中引入故意错配(这已在体外进行证明)，以允许对非优选基序进行编辑(https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gku272;Schneider等人(2014)，Nucleic Acid Res，42(10):e87);Fukuda等人(2017)，Scienticic Reports，7，doi:10.1038/srep41478)。测试另外的错配，诸如鸟苷取代，以查看它们是否降低天然偏好(图18)。

[2796] 对转录物中的多个A进行编辑

[2797] 结果表明，ADAR脱氨酶结构域的靶向窗口中与C相对的A比其他碱基优先被编辑。另外，在靶向碱基的几个碱基内与U配对的A碱基表现出较低的Cas13b-ADAR融合体编辑水

平,这表明该酶可以灵活地编辑多个A(图19)。这两个观察结果表明,可以通过错配要用C编辑的所有A来指定Cas13b-ADAR融合体的活性窗口中的多个A来进行编辑。为了测试这一点,采用了优化实验中最有前途的指导物,并且设计活性窗口中的多个A:C错配以测试产生多个A:I编辑的可能性。使用NGS对该实验的编辑率进行定量。为了抑制活性窗口中潜在的脱靶编辑,将非靶标A与A或G配对(取决于碱基偏好实验的结果)。

[2798] 用于RNA编辑的指导物长度滴定

[2799] ADAR天然作用于长度>20bp的分子间或分子内RNA双链体(还可参见Nishikura等人,(2010),Annu Rev Biochem,79:321-349)。结果表明,crRNA越长,形成的双链体越长,活性水平也越高。为了系统地比较不同长度的指导物对于RNA编辑活性的活性,我们已经设计了30、50、70和84个碱基的指导物,以在我们的荧光素酶报告子测定中校正终止密码子(图20A-20F和下表)。我们已经设计了这些指导物,使得经过编辑的A的位置相对于crRNA特异性决定区域的3'端存在于mRNA:crRNA双链体内所有可能的均匀距离(即+2、+4等)。

[2800]

| 序列名称 | 序列 |
|---|---|
| Guide_Cas13bC-luc_30_ADAR0Top (SEQ ID No. 179) | GCATCCTGCGGCCTCTACTCTGCATTCAATT |
| Guide_Cas13bC-luc_30_ADAR1Top(SEQ ID No. 180) | GACCATCCTGCGGCCTCTACTCTGCATTCAA |
| Guide_Cas13bC-luc_30_ADAR2Top(SEQ ID No. 181) | GAAACCATCCTGCGGCCTCTACTCTGCATTC |
| Guide_Cas13bC-luc_30_ADAR3Top(SEQ ID No. 182) | GCTAAACCATCCTGCGGCCTCTACTCTGCAT |
| Guide_Cas13bC-luc_30_ADAR4Top(SEQ ID No. 183) | GTTCTAAACCATCCTGCGGCCTCTACTCTGC |
| Guide_Cas13bC-luc_30_ADAR5Top(SEQ ID No. 184) | GTGTTCTAAACCATCCTGCGGCCTCTACTCT |
| Guide_Cas13bC-luc_30_ADAR6Top(SEQ ID No. | GAATGTTCTAAACCATCCTGCGGCCTCTACT |

[2801]

| | |
|---|---|
| 185) | |
| Guide_Cas13bC-luc_30_ADAR7Top(SEQ ID No. 186) | GAGAATGTTCTAAACCATCCTGCGGCCTCTA |
| Guide_Cas13bC-luc_30_ADAR8Top(SEQ ID No. 187) | GATAGAATGTTCTAAACCATCCTGCGGCCTC |
| Guide_Cas13bC-luc_30_ADAR9Top(SEQ ID No. 188) | GCCATAGAATGTTCTAAACCATCCTGCGGCC |
| Guide_Cas13bC-luc_30_ADAR10Top(SEQ ID No. 189) | GTTCCATAGAATGTTCTAAACCATCCTGCGG |
| Guide_Cas13bC-luc_30_ADAR11Top(SEQ ID No. 190) | GCTTTCCATAGAATGTTCTAAACCATCCTGC |
| Guide_Cas13bC-luc_30_ADAR12Top(SEQ ID No. 191) | GCTCTTTCCATAGAATGTTCTAAACCATCCT |
| Guide_Cas13bC-luc_30_ADAR13Top(SEQ ID No. 192) | GATCTCTTTCCATAGAATGTTCTAAACCATC |
| Guide_Cas13bC-luc_30_ADAR14Top(SEQ ID No. 193) | GGAATCTCTTTCCATAGAATGTTCTAAACCA |
| Guide_Cas13bC-luc_50_ADAR0Top(SEQ ID No. 194) | GCATCCTGCGGCCTCTACTCTGCATTCAATTACATACTGACACATTCGGCA |
| Guide_Cas13bC-luc_50_ADAR1Top(SEQ ID No. 195) | GACCATCCTGCGGCCTCTACTCTGCATTCAATTACATACTGACACATTCGG |
| Guide_Cas13bC-luc_50_ADAR2Top(SEQ ID No. 196) | GAAACCATCCTGCGGCCTCTACTCTGCATTCAATTACATACTGACACATTC |
| Guide_Cas13bC-luc_50_ADAR3Top(SEQ ID No. 197) | GCTAAACCATCCTGCGGCCTCTACTCTGCATTCAATTACATACTGACACAT |
| Guide_Cas13bC-luc_50_ADAR4Top(SEQ ID No. 198) | GTTCTAAACCATCCTGCGGCCTCTACTCTGCATTCAATTACATACTGACAC |
| Guide_Cas13bC-luc_50_ADAR5Top(SEQ ID No. | GTGTTCTAAACCATCCTGCGGCCTCTACTCTGCATTCAATTACATACTGAC |

[2802]

| | |
|---|---|
| 199) | |
| Guide_Cas13bC-luc_50_ADAR6Top(SEQ ID No. 200) | GAATGTTCTAAACCATCCTGCGGCCTCTACTCTGCATTCAATTACATACTG |
| Guide_Cas13bC-luc_50_ADAR7Top(SEQ ID No. 201) | GAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGCATTCAATTACATAC |
| Guide_Cas13bC-luc_50_ADAR8Top(SEQ ID No. 202) | GATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGCATTCAATTACAT |
| Guide_Cas13bC-luc_50_ADAR9Top(SEQ ID No. 203) | GCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGCATTCAATTAC |
| Guide_Cas13bC-luc_50_ADAR10Top(SEQ ID No. 204) | GTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGCATTCAATT |
| Guide_Cas13bC-luc_50_ADAR11Top(SEQ ID No. 205) | GCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGCATTCAA |
| Guide_Cas13bC-luc_50_ADAR12Top(SEQ ID No. 206) | GCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGCATTC |
| Guide_Cas13bC-luc_50_ADAR13Top(SEQ ID No. 207) | GATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGCAT |
| Guide_Cas13bC-luc_50_ADAR14Top(SEQ ID No. 208) | GGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGC |
| Guide_Cas13bC-luc_50_ADAR15Top(SEQ ID No. 209) | GTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCT |
| Guide_Cas13bC-luc_50_ADAR16Top(SEQ ID No. 210) | GACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACT |
| Guide_Cas13bC-luc_50_ADAR17Top(SEQ ID No. 211) | GGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTA |
| Guide_Cas13bC-luc_50_ADAR18Top(SEQ ID No. 212) | GTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTC |
| Guide_Cas13bC-luc_50_ADAR19Top(SEQ ID No. | GCCTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCC |

[2803]

| | |
|---|---|
| 213) | |
| Guide_Cas13bC-luc_50_ADAR20Top(SEQ ID No. 214) | GTTCCTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGG |
| Guide_Cas13bC-luc_50_ADAR21Top(SEQ ID No. 215) | GGGTTCCTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGC |
| Guide_Cas13bC-luc_50_ADAR22Top(SEQ ID No. 216) | GCAGGTTCCTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCT |
| Guide_Cas13bC-luc_50_ADAR23Top(SEQ ID No. 217) | GACCAGGTTCCTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATC |
| Guide_Cas13bC-luc_50_ADAR24Top(SEQ ID No. 218) | GGTACCAGGTTCCTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCA |
| Guide_Cas13bC-luc_70_ADAR0Top(SEQ ID No. 219) | GCATCCTGCGGCCTCTACTCTGCATTCAATTACATACTGACACATTCGGCAACATGTTTTTCCTGGTTTAT |
| Guide_Cas13bC-luc_70_ADAR1Top(SEQ ID No. 220) | GACCATCCTGCGGCCTCTACTCTGCATTCAATTACATACTGACACATTCGGCAACATGTTTTTCCTGGTTT |
| Guide_Cas13bC-luc_70_ADAR2Top(SEQ ID No. 221) | GAAACCATCCTGCGGCCTCTACTCTGCATTCAATTACATACTGACACATTCGGCAACATGTTTTCCTGGT |
| Guide_Cas13bC-luc_70_ADAR3Top(SEQ ID No. 222) | GCTAAACCATCCTGCGGCCTCTACTCTGCATTCAATTACATACTGACACATTCGGCAACATGTTTTTCCTG |
| Guide_Cas13bC-luc_70_ADAR4Top(SEQ ID No. 223) | GTTCTAAACCATCCTGCGGCCTCTACTCTGCATTCAATTACATACTGACACATTCGGCAACATGTTTTTCC |
| Guide_Cas13bC-luc_70_ADAR5Top(SEQ ID No. 224) | GTGTTCTAAACCATCCTGCGGCCTCTACTCTGCATTCAATTACATACTGACACATTCGGCAACATGTTTTT |
| Guide_Cas13bC-luc_70_ADAR6Top(SEQ ID No. 225) | GAATGTTCTAAACCATCCTGCGGCCTCTACTCTGCATTCAATTACATACTGACACATTCGGCAACATGTTT |
| Guide_Cas13bC-luc_70_ADAR7Top(SEQ ID No. 226) | GAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGCATTCAATTACATACTGACACATTCGGCAACATGT |
| Guide_Cas13bC-luc_70_ADAR8Top(SEQ ID No. | GATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGCATTCAATTACATACTGACACAT |

[2804]

| | |
|---|---|
| 227) | TCGGCAACAT |
| Guide_Cas13bC-luc_70_ADAR9Top(SEQ ID No. 228) | GCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGCATTCAATTACATACTGACACATTCGGCAAC |
| Guide_Cas13bC-luc_70_ADAR10Top(SEQ ID No. 229) | GTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGCATTCAATTACATACTGACACATTCGGCA |
| Guide_Cas13bC-luc_70_ADAR11Top(SEQ ID No. 230) | GCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGCATTCAATTACATACTGACACATTCGG |
| Guide_Cas13bC-luc_70_ADAR12Top(SEQ ID No. 231) | GCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGCATTCAATTACATACTGACACATTC |
| Guide_Cas13bC-luc_70_ADAR13Top(SEQ ID No. 232) | GATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGCATTCAATTACATACTGACACAT |
| Guide_Cas13bC-luc_70_ADAR14Top(SEQ ID No. 233) | GGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGCATTCAATTACATACTGACAC |
| Guide_Cas13bC-luc_70_ADAR15Top(SEQ ID No. 234) | GTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGCATTCAATTACATACTGAC |
| Guide_Cas13bC-luc_70_ADAR16Top(SEQ ID No. 235) | GACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGCATTCAATTACATACTG |
| Guide_Cas13bC-luc_70_ADAR17Top(SEQ ID No. 236) | GGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGCATTCAATTACATAC |
| Guide_Cas13bC-luc_70_ADAR18Top(SEQ ID No. 237) | GTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGCATTCAATTACAT |
| Guide_Cas13bC-luc_70_ADAR19Top(SEQ ID No. 238) | GCCTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGCATTCAATTAC |
| Guide_Cas13bC-luc_70_ADAR20Top(SEQ ID No. 239) | GTTCCTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGCATTCAATT |
| Guide_Cas13bC-luc_70_ADAR21Top(SEQ ID No. 240) | GGGTTCCTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGCATTCAA |
| Guide_Cas13bC-luc_70_ADAR22Top(SEQ ID No. | GCAGGTTCCTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTAC |

[2805]

| | |
|---|---|
| 241) | TCTGCATTC |
| Guide_Cas13bC-luc_70_ADAR23Top(SEQ ID No. 242) | GACCAGGTTCCTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGCAT |
| Guide_Cas13bC-luc_70_ADAR24Top(SEQ ID No. 243) | GGTACCAGGTTCCTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGC |
| Guide_Cas13bC-luc_70_ADAR25Top(SEQ ID No. 244) | GATGTACCAGGTTCCTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCT |
| Guide_Cas13bC-luc_70_ADAR26Top(SEQ ID No. 245) | GGTATGTACCAGGTTCCTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACT |
| Guide_Cas13bC-luc_70_ADAR27Top(SEQ ID No. 246) | GACGTATGTACCAGGTTCCTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTA |
| Guide_Cas13bC-luc_70_ADAR28Top(SEQ ID No. 247) | GACACGTATGTACCAGGTTCCTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTC |
| Guide_Cas13bC-luc_70_ADAR29Top(SEQ ID No. 248) | GCAACACGTATGTACCAGGTTCCTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCC |
| Guide_Cas13bC-luc_70_ADAR30Top(SEQ ID No. 249) | GCCCAACACGTATGTACCAGGTTCCTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGG |
| Guide_Cas13bC-luc_70_ADAR31Top(SEQ ID No. 250) | GGACCCAACACGTATGTACCAGGTTCCTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGC |
| Guide_Cas13bC-luc_70_ADAR32Top(SEQ ID No. 251) | GTTGACCCAACACGTATGTACCAGGTTCCTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCT |
| Guide_Cas13bC-luc_70_ADAR33Top(SEQ ID No. 252) | GCCTTGACCCAACACGTATGTACCAGGTTCCTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATC |
| Guide_Cas13bC-luc_70_ADAR34Top(SEQ ID No. 253) | GTTCCTTGACCCAACACGTATGTACCAGGTTCCTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCA |
| Guide_Cas13bC-luc_84_ADAR0Top(SEQ ID No. 254) | GCATCCTGCGGCCTCTACTCTGCATTCAATTACATACTGACACATTCGGCAACATGTTTTTCCTGGTTTATTTTCACACAGTCCA |
| Guide_Cas13bC-luc_84_ADAR1Top(SEQ ID No. | GACCATCCTGCGGCCTCTACTCTGCATTCAATTACATACTGACACATTCGGCAACATGTTTT |

[2806]

| | |
|---|---|
| 255) | TCCTGGTTTATTTTCACACAGTC |
| Guide_Cas13bC-luc_84_ ADAR2Top(SEQ ID No. 256) | GAAACCATCCTGCGGCCTCTACTCTGCATTC AATTACATACTGACACATTCGGCAACATGT TTTTCCTGGTTTATTTTCACACAG |
| Guide_Cas13bC-luc_84_ ADAR3Top(SEQ ID No. 257) | GCTAAACCATCCTGCGGCCTCTACTCTGCAT TCAATTACATACTGACACATTCGGCAACAT GTTTTTCCTGGTTTATTTTCACAC |
| Guide_Cas13bC-luc_84_ ADAR4Top(SEQ ID No. 258) | GTTCTAAACCATCCTGCGGCCTCTACTCTGC ATTCAATTACATACTGACACATTCGGCAAC ATGTTTTTCCTGGTTTATTTTCAC |
| Guide_Cas13bC-luc_84_ ADAR5Top(SEQ ID No. 259) | GTGTTCTAAACCATCCTGCGGCCTCTACTCT GCATTCAATTACATACTGACACATTCGGCA ACATGTTTTTCCTGGTTTATTTTC |
| Guide_Cas13bC-luc_84_ ADAR6Top(SEQ ID No. 260) | GAATGTTCTAAACCATCCTGCGGCCTCTACT CTGCATTCAATTACATACTGACACATTCGGC AACATGTTTTTCCTGGTTTATTT |
| Guide_Cas13bC-luc_84_ ADAR7Top(SEQ ID No. 261) | GAGAATGTTCTAAACCATCCTGCGGCCTCT ACTCTGCATTCAATTACATACTGACACATTC GGCAACATGTTTTTCCTGGTTTAT |
| Guide_Cas13bC-luc_84_ ADAR8Top(SEQ ID No. 262) | GATAGAATGTTCTAAACCATCCTGCGGCCT CTACTCTGCATTCAATTACATACTGACACAT TCGGCAACATGTTTTTCCTGGTTT |
| Guide_Cas13bC-luc_84_ ADAR9Top(SEQ ID No. 263) | GCCATAGAATGTTCTAAACCATCCTGCGGC CTCTACTCTGCATTCAATTACATACTGACAC ATTCGGCAACATGTTTTTCCTGGT |
| Guide_Cas13bC-luc_84_ ADAR10Top(SEQ ID No. 264) | GTTCCATAGAATGTTCTAAACCATCCTGCGG CCTCTACTCTGCATTCAATTACATACTGACA CATTCGGCAACATGTTTTTCCTG |
| Guide_Cas13bC-luc_84_ ADAR11Top(SEQ ID No. 265) | GCTTTCCATAGAATGTTCTAAACCATCCTGC GGCCTCTACTCTGCATTCAATTACATACTGA CACATTCGGCAACATGTTTTTCC |
| Guide_Cas13bC-luc_84_ ADAR12Top(SEQ ID No. 266) | GCTCTTTCCATAGAATGTTCTAAACCATCCT GCGGCCTCTACTCTGCATTCAATTACATACT GACACATTCGGCAACATGTTTTT |
| Guide_Cas13bC-luc_84_ ADAR13Top(SEQ ID No. 267) | GATCTCTTTCCATAGAATGTTCTAAACCATC CTGCGGCCTCTACTCTGCATTCAATTACATA CTGACACATTCGGCAACATGTTT |
| Guide_Cas13bC-luc_84_ ADAR14Top(SEQ ID No. 268) | GGAATCTCTTTCCATAGAATGTTCTAAACCA TCCTGCGGCCTCTACTCTGCATTCAATTACA TACTGACACATTCGGCAACATGT |
| Guide_Cas13bC-luc_84_ ADAR15Top(SEQ ID No. | GTGGAATCTCTTTCCATAGAATGTTCTAAAC CATCCTGCGGCCTCTACTCTGCATTCAATTA |

[2807]

| | |
|---|---|
| 269) | CATACTGACACATTCGGCAACAT |
| Guide_Cas13bC-luc_84_ADAR16Top(SEQ ID No. 270) | GACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGCATTCAATTACATACTGACACATTCGGCAAC |
| Guide_Cas13bC-luc_84_ADAR17Top(SEQ ID No. 271) | GGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGCATTCAATTACATACTGACACATTCGGCA |
| Guide_Cas13bC-luc_84_ADAR18Top(SEQ ID No. 272) | GTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGCATTCAATTACATACTGACACATTCGG |
| Guide_Cas13bC-luc_84_ADAR19Top(SEQ ID No. 273) | GCCTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGCATTCAATTACATACTGACACATTC |
| Guide_Cas13bC-luc_84_ADAR20Top(SEQ ID No. 274) | GTTCCTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGCATTCAATTACATACTGACACAT |
| Guide_Cas13bC-luc_84_ADAR21Top(SEQ ID No. 275) | GGGTTCCTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGCATTCAATTACATACTGACAC |
| Guide_Cas13bC-luc_84_ADAR22Top(SEQ ID No. 276) | GCAGGTTCCTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGCATTCAATTACATACTGAC |
| Guide_Cas13bC-luc_84_ADAR23Top(SEQ ID No. 277) | GACCAGGTTCCTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGCATTCAATTACATACTG |
| Guide_Cas13bC-luc_84_ADAR24Top(SEQ ID No. 278) | GGTACCAGGTTCCTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGCATTCAATTACATAC |
| Guide_Cas13bC-luc_84_ADAR25Top(SEQ ID No. 279) | GATGTACCAGGTTCCTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGCATTCAATTACAT |
| Guide_Cas13bC-luc_84_ADAR26Top(SEQ ID No. 280) | GGTATGTACCAGGTTCCTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGCATTCAATTAC |
| Guide_Cas13bC-luc_84_ADAR27Top(SEQ ID No. 281) | GACGTATGTACCAGGTTCCTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGCATTCAATT |
| Guide_Cas13bC-luc_84_ADAR28Top(SEQ ID No. 282) | GACACGTATGTACCAGGTTCCTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGCATTCAA |
| Guide_Cas13bC-luc_84_ADAR29Top(SEQ ID No. | GCAACACGTATGTACCAGGTTCCTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCA |

| | |
|---|---|
| 283) | TCCTGCGGCCTCTACTCTGCATTC |
| Guide_Cas13bC-luc_84_ADAR30Top(SEQ ID No. 284) | GCCCAACACGTATGTACCAGGTTCCTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGCAT |
| Guide_Cas13bC-luc_84_ADAR31Top(SEQ ID No. 285) | GGACCCAACACGTATGTACCAGGTTCCTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCTGC |
| Guide_Cas13bC-luc_84_ADAR32Top(SEQ ID No. 286) | GTTGACCCAACACGTATGTACCAGGTTCCTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACTCT |
| Guide_Cas13bC-luc_84_ADAR33Top(SEQ ID No. 287) | GCCTTGACCCAACACGTATGTACCAGGTTCCTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTACT |
| Guide_Cas13bC-luc_84_ADAR34Top(SEQ ID No. 288) | GTTCCTTGACCCAACACGTATGTACCAGGTTCCTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTCTA |
| Guide_Cas13bC-luc_84_ADAR35Top(SEQ ID No. 289) | GGGTTCCTTGACCCAACACGTATGTACCAGGTTCCTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCCTC |
| Guide_Cas13bC-luc_84_ADAR36Top(SEQ ID No. 290) | GTTGGTTCCTTGACCCAACACGTATGTACCAGGTTCCTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGGCC |
| Guide_Cas13bC-luc_84_ADAR37Top(SEQ ID No. 291) | GCCTTGGTTCCTTGACCCAACACGTATGTACCAGGTTCCTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGCGG |
| Guide_Cas13bC-luc_84_ADAR38Top(SEQ ID No. 292) | GGCCCTTGGTTCCTTGACCCAACACGTATGTACCAGGTTCCTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCTGC |
| Guide_Cas13bC-luc_84_ADAR39Top(SEQ ID No. 293) | GCCGCCCTTGGTTCCTTGACCCAACACGTATGTACCAGGTTCCTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATCCT |
| Guide_Cas13bC-luc_84_ADAR40Top(SEQ ID No. 294) | GCGCCGCCCTTGGTTCCTTGACCCAACACGTATGTACCAGGTTCCTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCATC |
| Guide_Cas13bC-luc_84_ADAR41Top(SEQ ID No. 295) | GGTCGCCGCCCTTGGTTCCTTGACCCAACACGTATGTACCAGGTTCCTGGAACTGGAATCTCTTTCCATAGAATGTTCTAAACCA |

[2808]

[2809] 逆转因果疾病突变

[2810] 合成下表中具有致病性G＞A突变的三个基因，该突变将提前终止位点引入基因中并将它们整合到非人类细胞系中。测试了Cas13b12-ADAR2通过将终止密码子UAG改变为UIG(UGG)来校正转录物并因此恢复蛋白质翻译的能力。

[2811]

| 蛋白质长度(aa) | 全长候选物 | 基因 | 疾病 |
|---|---|---|---|
| 498 | NM_004992.3(MECP2):c.311G>A (p.Trp104Ter) | MECP2 | 瑞特综合征 |
| 414 | NM_007375.3(TARDBP):c.943G>A (p.Ala315Thr) | TARDBP | 10 型肌萎缩性脊髓侧索硬化症 |
| 393 | NM_000546.5(TP53):c.273G>A (p.Trp91Ter) | TP53 | 利弗劳梅尼综合征/遗传性癌症易感综合征 |

[2812] 下表5中示出了另外48个致病性G>A突变以及伴随的疾病。合成这些突变周围的200bp片段，而不是整个基因，并将它们克隆到GFP前面。当提前终止位点恢复时，将允许GFP的翻译，并且除了RNA测序之外，可以通过高通量的荧光测量校正。

[2813] 表5

[2814]

| | 候选物 | 基因 | 疾病 |
|---|---|---|---|
| 1 | NM_004006.2(DMD):c.3747G>A (p.Trp1249Ter) | DMD | 杜兴氏肌营养不良 |
| 2 | M_000344.3(SMN1):c.305G>A (p.Trp102Ter) | SMN1 | II 型脊髓性肌肉萎缩/库格尔伯格-魏兰德病 |
| 3 | NM_000492.3(CFTR):c.3846G>A (p.Trp1282Ter) | CFTR | 囊性纤维化/遗传性胰腺炎/未提供/ataluren 响应-功效 |
| 4 | NM_004562.2(PRKN):c.1358G>A (p.Trp453Ter) | PRKN | 帕金森病 2 |
| 5 | NM_017651.4(AHI1):c.2174G>A (p.Trp725Ter) | AHI1 | 朱伯特综合征 3 |
| 6 | NM_000238.3(KCNH2):c.3002G>A (p.Trp1001Ter) | KCNH2 | 长 QT 综合征/未提供 |
| 7 | NM_000136.2(FANCC):c.1517G>A (p.Trp506Ter) | FANCC\|C9orf3 | Fanconi 贫血，互补组 C |
| 8 | NM_001009944.2(PKD1):c.12420G>A (p.Trp4140Ter) | PKD1 | 成人型多囊性肾脏疾病 |
| 9 | NM_177965.3(C8orf37):c.555G>A (p.Trp185Ter) | C8orf37 | 色素性视网膜炎 64 |

[2815]

| | | | |
|---|---|---|---|
| 10 | NM_000833.4(GRIN2A):c.3813G>A (p.Trp1271Ter) | GRIN2A | 局灶性癫痫，伴语言障碍并且伴或不伴智力障碍 |
| 11 | NM_000548.4(TSC2):c.2108G>A (p.Trp703Ter) | TSC2 | 结节性硬化症 2\|结节性硬化综合征 |
| 12 | NM_000267.3(NF1):c.7044G>A (p.Trp2348Ter) | NF1 | 1 型神经纤维瘤病 |
| 13 | NM_000520.5(HEXA):c.1454G>A (p.Trp485Ter) | HEXA | Tay-Sachs 病 |
| 14 | NM_130838.1(UBE3A):c.2304G>A (p.Trp768Ter) | UBE3A | 安格尔曼综合征 |
| 15 | NM_000543.4(SMPD1):c.168G>A (p.Trp56Ter) | SMPD1 | A 型尼曼匹克病 (Niemann-Pick disease) |
| 16 | NM_000218.2(KCNQ1):c.1175G>A (p.Trp392Ter) | KCNQ1 | 长 QT 综合征 |
| 17 | NM_000256.3(MYBPC3):c.3293G>A (p.Trp1098Ter) | MYBPC3 | 原发性家族性肥厚型心肌病 |
| 18 | NM_000038.5(APC):c.1262G>A (p.Trp421Ter) | APC | 家族性腺瘤性息肉病 1 |
| 19 | NM_000249.3(MLH1):c.1998G>A (p.Trp666Ter) | MLH1 | 林奇综合征 |
| 20 | NM_000054.4(AVPR2):c.878G>A (p.Trp293Ter) | AVPR2 | 肾原性尿崩症，X 连锁 |
| 21 | NM_001204.6(BMPR2):c.893G>A (p.W298*) | BMPR2 | 原发性肺动脉高压 |
| 22 | NM_004560.3(ROR2):c.2247G>A (p.Trp749Ter) | ROR2 | B1 型短指症 |
| 23 | NM_000518.4(HBB):c.114G>A (p.Trp38Ter) | HBB | β^0^地中海贫血/β 地中海贫血 |
| 24 | NM_024577.3(SH3TC2):c.920G>A (p.Trp307Ter) | SH3TC2 | 4C 型夏科-马利-杜斯氏病 |
| 25 | NM_206933.2(USH2A):c.9390G>A (p.Trp3130Ter) | USH2A | 2A 型乌谢尔综合征 |
| 26 | NM_000179.2(MSH6):c.3020G>A (p.Trp1007Ter) | MSH6 | 林奇综合征 |

[2816]

| | | | |
|---|---|---|---|
| 27 | NM_002977.3(SCN9A):c.2691G>A (p.Trp897Ter) | SCN9A\|LOC101929680 | 疼痛淡漠，先天性常染色体隐性遗传 |
| 28 | NM_000090.3(COL3A1):c.30G>A (p.Trp10Ter) | COL3A1 | 4 型埃勒斯-当洛斯综合征 |
| 29 | NM_000551.3(VHL):c.263G>A (p.Trp88Ter) | VHL | 希佩尔-林道综合征/未提供 |
| 30 | NM_015627.2(LDLRAP1):c.65G>A (p.Trp22Ter) | LDLRAP1 | 高胆固醇血症，常染色体隐性遗传 |
| 31 | NM_000132.3(F8):c.3144G>A (p.Trp1048Ter) | F8 | 遗传性因子 VIII 缺乏性疾病 |
| 32 | NM_002185.4(IL7R):c.651G>A (p.Trp217Ter) | IL7R | 严重联合免疫缺陷，常染色体隐性遗传，T 细胞阴性，B 细胞阳性，NK 细胞阳性 |
| 33 | NM_000527.4(LDLR):c.1449G>A (p.Trp483Ter) | LDLR | 家族性高胆固醇血症 |
| 34 | NM_002294.2(LAMP2):c.962G>A (p.Trp321Ter) | LAMP2 | 达农病 |
| 35 | NM_000271.4(NPC1):c.1142G>A (p.Trp381Ter) | NPC1 | C1 型尼曼-匹克病 |
| 36 | NM_000267.3(NF1):c.1713G>A (p.Trp571Ter) | NF1 | 1 型神经纤维瘤病 |
| 37 | NM_000035.3(ALDOB):c.888G>A (p.Trp296Ter) | ALDOB | 遗传性果糖尿症 |
| 38 | NM_000090.3(COL3A1):c.3833G>A (p.Trp1278Ter) | COL3A1 | 4 型埃勒斯-当洛斯综合征 |
| 39 | NM_001369.2(DNAH5):c.8465G>A (p.Trp2822Ter) | DNAH5 | 原发性纤毛运动障碍 |
| 40 | NM_178443.2(FERMT3):c.48G>A (p.Trp16Ter) | FERMT3 | III 型白细胞粘附缺陷 |
| 41 | NM_005359.5(SMAD4):c.906G>A (p.Trp302Ter) | SMAD4 | 青少年息肉综合征 |
| 42 | NM_032119.3(ADGRV1):c.7406G>A (p.Trp2469Ter) | ADGRV1 | 2C 型乌舍尔综合征 |

[2817]

| | | | |
|---|---|---|---|
| 43 | NM_000206.2(IL2RG):c.710G>A (p.Trp237Ter) | IL2RG | X连锁严重联合免疫缺陷症 |
| 44 | NM_007294.3(BRCA1):c.5511G>A (p.Trp1837Ter) | BRCA1 | 家族性乳腺癌/家族性乳腺癌-卵巢癌 1 |
| 45 | NM_130799.2(MEN1):c.1269G>A (p.Trp423Ter) | MEN1 | 遗传性癌症易感综合征 |
| 46 | NM_000071.2(CBS):c.162G>A (p.Trp54Ter) | CBS | 由于 CBS 缺乏引起的高胱氨酸尿 |
| 47 | NM_000059.3(BRCA2):c.582G>A (p.Trp194Ter) | BRCA2 | 家族性乳腺癌/家族性乳腺癌-卵巢癌 2 |
| 48 | NM_000053.3(ATP7B):c.2336G>A (p.Trp779Ter) | ATP7B | Wilson 病 |

[2818] 实施例3

[2819] 高效且精确的核酸编辑在治疗遗传性疾病方面具有广阔的前景,尤其是在RNA水平上,其中可以挽救疾病相关的转录物,从而产生功能性蛋白产物。VI型CRISPR-Cas系统含有可编程的单效应物RNA引导的RNA酶Cas13。这里,我们简要描述了VI型系统的多样性,以工程化能够稳健敲低的Cas13直系同源物,并通过使用无催化活性的Cas13(dCas13)将腺苷脱氨酶活性指导至哺乳动物细胞中的转录物来证明RNA编辑。通过将ADAR2脱氨酶结构域融合到dCas13上并工程化指导RNA以形成最佳RNA双链体底物,我们实现了特定单腺苷到肌苷的靶向编辑(在翻译过程中被读出为鸟苷),效率范围通常为20-40%并且高达89%。该系统称为可编程的A到I置换的RNA编辑(REPAIR),可以被进一步工程化以实现高特异性。工程化的变体REPAIRv2显示出特异性提高170倍以上,同时仍保持稳健的中靶A到I编辑。我们使用REPAIRv2编辑含有已知致病突变的全长转录物,并形成适合包装在腺相关病毒(AAV)载体中的功能性截短型式。REPAIR提供了一个有前景的RNA编辑平台,在研究、治疗和生物技术领域具有广泛的适用性。精确的核酸编辑技术对于研究细胞功能和作为新颖疗法有价值。尽管当前的编辑工具诸如Cas9核酸酶可以实现对基因组基因座的可编程修饰,但是由于插入或缺失编辑物通常是异源的,或需要供体模板进行精确编辑。碱基编辑器诸如dCas9-APOBEC融合体允许在不产生双链断裂的情况下进行编辑,但由于胞苷脱氨酶活性的性质(在靶标窗口中编辑任何胞苷)可能缺乏精确性。此外,对原间隔子邻近基序(PAM)的要求限制了可能的编辑位点的数量。这里,我们描述了使用来自VI型原核成簇规律间隔短回文重复序列(CRISPR)自适应免疫系统的RNA引导的RNA靶向Cas13酶来开发一种精确且灵活的RNA碱基编辑工具。

[2820] 精确的核酸编辑技术对于研究细胞功能和作为新颖疗法有价值。当前的基于可编辑核酸酶的编辑工具诸如原核成簇规律间隔短回文重复序列(CRISPR)相关的核酸酶Cas9(1-4)或Cpf1(5)已广泛用于介导靶向DNA裂解,这继而驱动通过非同源末端连接(NHEJ)进行靶向基因破坏或通过模板依赖性同源定向修复(HDR)进行精确基因编辑(6)。NHEJ利用在分裂细胞和有丝分裂后细胞中均活跃的宿主机器,并通过产生可以导致蛋白质编码基因移码的插入或缺失(插入缺失)突变的混合物来提供有效的基因破坏。相比之下,HDR由其表达

在很大程度上受限于复制细胞的宿主机器介导。因此,开发有丝分裂后细胞中基因编辑能力仍然是主要挑战。最近,DNA碱基编辑器诸如使用无催化活性的Cas9(dCas9)将胞苷脱氨酶活性靶向特定的基因组靶标以在靶标窗口内实现胞嘧啶向胸腺嘧啶的转化,允许在不产生DNA双链断裂的情况下进行编辑并显著减少了插入缺失的形成(7,8)。但是,由于Cas9在编辑位点处需要一个原间隔邻近基序(PAM),DNA碱基编辑器的靶向范围受到限制(9)。这里,我们描述了使用VI型CRISPR相关的RNA引导的RNA酶Cas13来开发一种精确且灵活的RNA碱基编辑技术(10-13)。

[2821] Cas13酶具有两个介导精确RNA裂解的高等真核生物和原核生物核苷酸结合(HEPN)内切RNA酶结构域(10,11)。迄今为止,已鉴定出三个Cas13蛋白家族:Cas13a(以前称为C2c2)、Cas13b和Cas13c(12,13)。我们最近报道了Cas13a酶可以用作核酸检测的工具(14)以及哺乳动物和植物细胞RNA敲低和转录物追踪的工具(15)。Cas13酶的RNA引导的性质使它们成为用于RNA结合和干扰应用的有吸引力的工具。

[2822] 作用于RNA(ADAR)酶家族的腺苷脱氨酶通过将腺苷水解脱氨成肌苷来介导转录物的内源编辑,肌苷是一种在翻译和剪接中的功能相当于鸟苷的核碱基(16)。有两种功能性的人类ADAR直系同源物ADAR1和ADAR2,它们由N端双链RNA结合结构域和C端催化脱氨结构域组成。ADAR1和ADAR2的内源靶位点含有相当高的双链同一性,并且催化结构域需要双链体化区在体外和体内进行有效编辑(17,18)。尽管ADAR蛋白具有可能限制靶向的潜在灵活性的优选编辑基序,但高活性突变体诸如ADAR(E488Q)(19)可放宽序列限制并提高腺苷到肌苷的编辑率。ADAR优先使RNA双链体中与胞苷碱基相对的腺苷脱氨(20),从而为精确的碱基编辑提供有前景的机会。尽管先前的方法已经通过RNA指导物对靶向ADAR融合体进行了工程化(21-24),但尚未报道这些方法的特异性,并且它们各自的靶向机制依赖于RNA-RNA杂交而不需要蛋白质伴侣的帮助,这可增强靶标识别和严格性。

[2823] 这里,我们测定了整个Cas13酶家族在哺乳动物细胞中的RNA敲低活性,并鉴定出来自普雷沃菌属某种P5-125的Cas13b直系同源物(PspCas13b)对于哺乳动物细胞应用是最有效和最特异的。然后,我们将ADAR2脱氨酶结构域(ADARDD)与无催化活性的PspCas13b融合,并证实了对报告子和内源转录物的可编程的A到I(G)置换的RNA编辑(REPAIR)以及与疾病相关的突变。最后,我们采用合理的诱变方案来提高dCas13b-ADAR2DD融合体的特异性,以产生特异性增加170倍以上的REPAIRv2。

[2824] 细菌构建体的方法设计和克隆

[2825] 在Lac启动子的控制下,将经哺乳动物密码子优化的Cas13b构建体克隆到氯霉素抗性的pACYC184载体中。然后,在pJ23119启动子的控制下,将由BsaI限制性位点分开的两个对应的正向重复(DR)序列插入Cas13b下游。最后,将用于靶向间隔子的寡聚物使用T4PNK(New England Biolabs)磷酸化,退火并使用T7连接酶(Enzymatics)连接到经BsaI消化的载体中,以生成靶向Cas13b载体。

[2826] 细菌PFS筛选

[2827] 通过使用NEB Gibson Assembly(New England Biolabs),在氨苄青霉素抗性基因bla的起始密码子下游紧接插入含有Cas13b靶标且具有由靶位点分开的2个5'随机核苷酸和4个3'随机核苷酸的PCR产物,从而克隆出用于PFS筛选的氨苄青霉素抗性质粒。然后将100ng抗氨苄青霉素的靶质粒与65-100ng抗氯霉素的Cas13b细菌靶向质粒电穿孔到Endura

电感受态细胞中。将质粒加入细胞，在冰上孵育15分钟，使用制造商方案进行电穿孔，然后将950uL回收培养基加入细胞中，然后在37℃下过生长一小时。将过生长物涂铺到氯霉素和氨苄青霉素双重选择板上。将过生长物的连续稀释液用于估计cfu/ng DNA。涂铺后16小时，将细胞从平板上刮下，并使用Qiagen Plasmid Plus Maxi Kit(Qiagen)收集存活的质粒DNA。通过PCR扩增存活的Cas13b靶序列及其侧接区，并使用Illumina NextSeq进行测序。为了评估PFS偏好，提取原始文库中包含随机核苷酸的位置，并使用自定义python脚本提取两个生物复制中都存在的相对于仅载体条件而言耗竭的序列。然后，将Cas13b条件下PFS丰度与仅载体条件相比的比率的-log2用于计算优选基序。具体地，将-log2(样品/载体)耗竭率高于特定阈值的所有序列都用于生成序列基序的weblogo(weblogo.berkeley.edu)。表9列出了用于为每个Cas13b直系同源物生成weblogo的特定耗竭率值。

[2828] 用于RNA干扰的哺乳动物构建体的设计和克隆

[2829] 为了生成用于在哺乳动物细胞中测试Cas13直系同源物的载体，对经哺乳动物密码子优化的Cas13a、Cas13b和Cas13c基因进行PCR扩增，并在EF1α启动子的控制下将这些基因golden-gate克隆到含有双重NLS序列和C端msfGFP的哺乳动物表达载体中。为了进一步优化，在EF1α启动子的控制下将Cas13直系同源物golden gate克隆到包含不同的C端定位标签的目的载体中。

[2830] 通过使用NEB Gibson Assembly(New England Biolabs)对编码高斯和海萤荧光素酶的DNA、EF1α和CMV启动子和组件进行PCR扩增，克隆出双荧光素酶报告基因。

[2831] 为了表达Cas13a、Cas13b或Cas13c直系同源物的哺乳动物指导RNA，合成了具有golden-gate受体位点的对应正向重复序列，并经由限制性消化克隆在U6表达下对其进行克隆。然后通过golden-gate克隆将各个指导物克隆到每个直系同源物的对应表达主链中。

[2832] 用于Cas13干扰特异性的汇集错配文库的克隆

[2833] 通过PCR创建汇集错配文库靶位点。将在靶标内每个位置处包含94％正确碱基和2％每个错误碱基的混合物的G-荧光素酶中包含半简并靶序列的寡聚物用作一个引物，并将与G-荧光素酶的非靶向区对应的寡聚物用作PCR反应中的第二个引物。然后使用NEB Gibson Assembly(New England Biolabs)将错配文库靶标克隆到双荧光素酶报告子中代替野生型G-荧光素酶。

[2834] 用于RNA编辑的哺乳动物构建体的设计和克隆

[2835] 通过在HEPN结构域的催化位点进行两个组氨酸到丙氨酸突变(H133A/H1058A)使PspCas13b失去催化活性(dPspCas13b)。合成人类ADAR1和ADAR2的脱氨酶结构域并进行PCR扩增以将其gibson克隆到pcDNA-CMV载体主链中，并通过GS或GSGGGGS(SEQ ID No.296)接头使其在C端与dPspCas13b融合。对于测试不同接头的实验，我们在dPspCas13b和ADAR2dd之间克隆了以下其他接头：GGGGSGGGGSGGGGS、EAAAK(SEQ ID No.297)、GGSGGSGGSGGSGGSGGS(SEQ ID No.298)和SGSETPGTSESATPES(SEQ ID No.299)(XTEN)。通过将适当的突变体gibson克隆到dPspCas13b-GSGGGGS主链中来产生特异性突变体。

[2836] 通过在用于敲低实验的荧光素酶报告载体中产生W85X突变(TGG>TAG)来产生用于测量RNA编辑活性的荧光素酶报告载体。该报告载体将功能性Gluc表达为归一化对照，但是由于添加了预先终止位点而导致有缺陷的Cluc。为了测试ADAR编辑基序偏好，我们克隆了Cluc的第85位密码子(XAX)处腺苷周围的每个可能基序。

[2837] 为了测试REPAIR的PFS偏好,我们克隆了在靶区和腺苷编辑位点上游包含6个碱基对的简并PFS序列的汇集质粒文库。文库由Integrated DNA Technologies(IDT)合成为ultramer,并且通过使引物退火以及DNA聚合酶I(New England Biolabs)的Klenow片段填补序列使文库成为双链。然后将包含简并序列的该dsDNA片段gibson克隆到消化的报告载体中,然后将其用异丙醇沉淀并纯化。然后将克隆的文库电穿孔到Endura感受态大肠埃希氏菌细胞(Lucigen)中,并涂铺在245mm×245mm正方形生物测定板(Nunc)上。16小时后,使用无内毒素的MACHEREY-NAGEL中量质粒提取试剂盒收获菌落并进行中量质粒提取分析。通过下一代测序验证克隆的文库。

[2838] 为了克隆与疾病相关的突变来测试REPAIR活性,选择如ClinVar中定义的与疾病发病机制相关的34个G>A突变,并在CMV启动子下将围绕这些突变的200bp区golden-gate克隆到mScarlett和EGFP之间。选择AVPR2和FANCC中两个另外的G>A突变,用于在EF1α表达下Gibson克隆整个基因序列。

[2839] 为了表达用于REPAIR的哺乳动物指导RNA,合成了具有golden-gate受体位点的PspCas13b正向重复序列,并经由限制性消化克隆在U6表达下对其进行克隆。然后通过golden gate克隆将各个指导物克隆到该表达主链中。

[2840] 哺乳动物细胞培养

[2841] 哺乳动物细胞培养实验是在HEK293FT品系(美国模式培养物保藏所(ATCC))中进行的,该品系在含高葡萄糖、丙酮酸钠和GlutaMAX(Thermo Fisher Scientific)并且另外还补充有1×青霉素-链霉素(Thermo Fisher Scientific)和10%胎牛血清(VWR Seradigm)的杜尔贝科改良伊格尔培养基(Dulbecco's Modified Eagle Medium)中生长。将细胞保持在低于80%的汇合度。

[2842] 除非另有说明,否则所有转染均用Lipofectamine 2000(Thermo Fisher Scientific)在涂覆有聚-D-赖氨酸的96孔板(BD Biocoat)中进行。转染前十六小时以大约20,000个细胞/孔涂铺细胞,以确保转染时的汇合度达到90%。对于板上的每个孔,将转染质粒与Opti-MEM I还原血清培养基(Thermo Fisher)合并,至总体积为25μl。单独地,将24.5ul Opti-MEM与0.5ul Lipofectamine 2000合并。然后将质粒和Lipofectamine溶液合并并孵育5分钟,然后将它们吸移到细胞上。

[2843] RNA敲低哺乳动物细胞测定

[2844] 为了用报告子构建体评估哺乳动物细胞中的RNA靶向,将150ng Cas13构建体与300ng指导表达质粒和12.5ng敲低报告子构建体共转染。转染后48小时,将含有分泌型荧光素酶的培养基从细胞中移出,在PBS中以1:5稀释,并利用BioLux海萤和Biolux高斯荧光素酶测定试剂盒(New England Biolabs)在具有注射方案的读板仪(Biotek Synergy Neo2)上测量活性。进行的所有重复都是生物学重复。

[2845] 为了靶向内源基因,将150ng Cas13构建体与300ng指导表达质粒共转染。转染后48小时,使用先前描述的Cells-to-Ct试剂盒(Thermo Fisher Scientific)的[CITE PROTOCOLS]改良方法,裂解细胞并收获RNA并且进行反转录。使用KRAS转录物的TaqMan qPCR探针(Thermo Fisher Scientific)、GAPDH对照探针(Thermo Fisher Scientific)和Fast Advanced Master Mix(Thermo Fisher Scientific)通过qPCR测量cDNA表达。在LightCycler 480仪器II(Roche)上读取qPCR反应,其中以384孔形式进行四个5ul技术重

复。

[2846] 使用错配靶标的汇集文库评价RNA特异性

[2847] 使用接种在6孔板中的HEK293FT细胞来测试Cas13干扰错配靶标文库的能力。使用2400ng Cas13载体、4800ng指导物和240ng错配靶标文库转染约70％汇合的细胞。转染后48小时，使用QIAshredder(Qiagen)和Qiagen RNeasy Mini Kit收获细胞并提取RNA。使用qScript Flex cDNA合成试剂盒(Quantabio)，按照制造商基因特异性引发延伸方案并使用Gluc特异性RT引物，对1ug提取的RNA进行反转录。然后扩增cDNA并在Illumina NextSeq上进行测序。

[2848] 通过对每个序列的读段进行计数来分析测序，并通过确定log2(-读段计数比)值来计算消耗竭评分，其中读段计数比是靶向指导条件对比非靶向指导条件下的读段计数之比。该评分值代表序列上Cas13的活性水平，值越高代表耗竭越强，并因此代表Cas13裂解活性越高。确定了单错配和双错配序列的单独分布，并将其绘制为热图，其具有每个错配同一性的耗竭评分。对于双错配序列，绘制了给定位置处所有可能的双错配的平均值。

[2849] 通过RNA测序在哺乳动物细胞中对Cas13进行全转录组谱分析

[2850] 为了测量全转录组特异性，共转染了150ng Cas13构建体、300ng指导表达质粒和15ng敲低报告子构建体；对于shRNA条件，共转染了300ng shRNA靶向质粒、15ng敲低报告子构建体和150ng EF1α驱动的mCherry(以平衡报告子负载)。转染后48小时，利用RNeasy Plus Mini试剂盒(Qiagen)纯化RNA，使用NEBNext Poly(A)mRNA磁性分离模块(New England Biolabs)选择mRNA，并将其准备利用Illumina(New England Biolabs)的NEBNext Ultra RNA Library Prep Kit进行测序。然后在NextSeq(Illumina)上对RNA测序文库进行测序。

[2851] 为了分析全转录组测序数据，使用Bowtie和RSEM 1.2.31版利用默认参数对读段进行RefSeq GRCh38组件比对[CITE RSEM:在有或没有参考基因组的情况下根据RNA-Seq数据进行精确的转录物定量]。将转录物表达定量为log2(TPM+1)，过滤基因以使log2(TPM+1)大于2.5。为了选择差异表达的基因，仅考虑差异变化大于2或小于.75的基因。针对三个靶向重复对比非靶向重复进行Student T检验以评价差异表达的统计显著性，并通过Benjamini-Hochberg程序过滤了小于0.01％的错误发现率。

[2852] 哺乳动物细胞转染中的ADAR RNA编辑

[2853] 为了评估哺乳动物细胞中的REPAIR活性，我们转染了150ng REPAIR载体、300ng指导表达质粒和40ng RNA编辑报告基因。48小时后，使用先前描述[引用JJ]的方法并利用基因特异性逆转录引物，收获来自细胞的RNA并进行逆转录。然后使用NEBNext High-Fidelity 2X PCR Master Mix对提取的cDNA进行两轮PCR，以加入Illumina衔接子和样品条形码。然后在Illumina NextSeq或MiSeq上对文库进行下一代测序。然后在测序窗口内的所有腺苷处评价RNA编辑率。

[2854] 在荧光素酶报告子被靶向RNA编辑的实验中，我们还在收获RNA之前收获了含有分泌型荧光素酶的培养基。在这种情况下，由于校正后的Cluc可能处于较低水平，因此我们没有稀释培养基。我们利用BioLux海萤和Biolux高斯荧光素酶测定试剂盒(New England Biolabs)在具有注射方案的读板仪(Biotek Synergy Neo2)上测量荧光素酶活性。进行的所有重复都是生物学重复。

[2855] PFS结合哺乳动物筛选

[2856] 为了确定PFS对编辑效率的贡献，将625ng PFS靶标文库、4.7ug指导物和2.35ug REPAIR共转染到涂铺在225cm2烧瓶中的HEK293FT细胞上。将质粒与33ul PLUS试剂(Thermo Fisher Scientific)混合，用Opti-MEM调节至533ul，孵育5分钟，与30ul Lipofectamine 2000和500ul Opti-MEM合并，再孵育5分钟，然后将其吸移到细胞上。转染后48小时，将RNA利用RNeasy Plus Mini试剂盒(Qiagen)收获，利用qScript Flex(Quantabio)并使用基因特异性引物进行反转录，并使用NEBNext High-Fidelity 2X PCR Master Mix(New England Biolabs)进行两轮PCR扩增，以加入Illumina衔接子和样品条形码。在Illumina NextSeq上对文库进行测序，并将靶标腺苷处的RNA编辑率映射为PFS同一性。为了增加覆盖率，PFS在计算上被折叠为4个核苷酸。计算每个PFS的REPAIR编辑率，在生物学重复内求平均值，减去对应PFS的非靶向率。

[2857] 全转录组测序以评价ADAR编辑特异性

[2858] 为了分析整个转录组的脱靶RNA编辑位点，我们使用RNeasy Plus Miniprep试剂盒(Qiagen)从转染后48小时的细胞中收集总RNA。然后使用NEBNext Poly(A)mRNA磁性分离模块(NEB)富集mRNA级分，然后并将该RNA准备使用Illumina(NEB)的NEBNext Ultra RNA Library Prep Kit进行测序。然后在Illumina NextSeq上对文库进行测序，并将其加载使得每个样品至少有500万个读段。

[2859] 靶向和全转录组实验的RNA编辑分析

[2860] 为了分析全转录组RNA编辑RNA测序数据，将序列文件随机下采样至500万个读段。使用添加有Gluc和Cluc序列的RefSeq GRCh38组件生成索引，并使用Bowtie/RSEM版本1.3.0对读段进行比对和定量。然后使用REDitools[引用]利用以下参数对比对BAM进行分类和RNA编辑位点分析:-t 8-e-d-l-U[AG或TC]-p-u-m20-T6-0-W-v 1-n 0.0。在未转染或EGFP转染的条件下发现的任何显著编辑均视为SNP或转染的假象，并从脱靶分析中过滤掉。如果Fisher精确检验得出p值小于0.5并且3个生物学重复中至少有2个鉴定出编辑位点，则认为脱靶显著。

[2861] 为了分析每个脱靶的预测变异效应，使用SIFT和PolyPhen-2注释并且使用作为UCSC基因组浏览器工具套件的一部分的变体注释集成器(https://genome.ucsc.edu/cgi-bin/hgVai)，对脱靶编辑位点列表进行分析。为了判断脱靶基因是否致癌，参见了COSMIC癌症体细胞突变目录(cancer.sanger.ac.uk)中的致癌注释数据库。

[2862] 为了分析REPAIR构建体是否干扰RNA水平，使用RSEM分析输出的每百万转录物(TPM)值进行表达计数，并通过取log2(TPM+1)将其转化为对数空间。为了找到差异调控的基因，针对三个靶向指导物重复对比三个非靶向指导物重复进行Student t检验。仅对log2(TPM+1)值大于2.5的基因进行统计分析，并且只有倍数变化大于2或小于0.8的基因才视为被差异调控。对错误发现率小于0.01的基因进行报告。

[2863] 结果

[2864] 哺乳动物细胞中Cas13家族成员的全面表征

[2865] 我们之前开发了用于哺乳动物敲低应用的LwaCas13a，但它需要msfGFP稳定结构域才能有效敲低，并且尽管特异性很高，但敲低效率并不能始终低于50％(15)。我们试图通过表征一组遗传上多样化的Cas13家族成员来评估它们在哺乳动物细胞中的RNA敲低活性，

从而鉴定出更稳健的RNA靶向CRISPR系统(图49A)。我们将21个Cas13a、15个Cas13b和7个Cas13c经哺乳动物密码子优化的直系同源物(表6)克隆到具有N端和C端核输出信号(NES)序列和C端msfGFP的表达载体中,以增强蛋白质稳定性。为了测定在哺乳动物细胞中的干扰,我们设计了在单独的启动子下表达正交高斯(Gluc)和海萤(Cluc)荧光素酶的双报告基因构建体,该构建体允许一种荧光素酶用作Cas13干扰活性的量度,而另一种荧光素酶作为内部对照。对于每个直系同源物,我们使用来源于氨苄青霉素干扰测定的Cas13b PFS基序(图55A-55C;表7;补充说明1)和来自以前Cas13a活性报告中的3'H PFS设计了与PFS兼容的指导RNA(10)。

[2866] 我们用Cas13表达、指导RNA和报告质粒转染HEK293FT细胞,并在48小时后定量靶向Gluc的水平。针对每个Cas13直系同源物测试两种指导RNA揭示了一系列的活性水平,包括相对于最近表征的LwaCas13a,两种指导RNA的干扰相似或增加的五种Cas13b直系同源物(图49B)。我们选择了这五种Cas13b直系同源物,以及前两个Cas13a直系同源物进行进一步工程化。

[2867] 接下来,我们在没有msfGFP的情况下测试了Cas13介导的Gluc的敲低,以便选择不需要稳定结构域就能实现稳健活性的直系同源物。我们假设,除了msfGFP之外,Cas13活性还可能受到亚细胞定位的影响,如先前针对LwaCas13a的优化所报道的(15)。因此,我们测试了七种选择的Cas13直系同源物的干扰活性,这些直系同源物在没有msfGFP的情况下与六种不同定位标签中的一个标签在C端融合。使用荧光素酶报告子测定法,我们发现与HIV Rev基因NES在C端融合的PspCas13b和PguCas13b以及与MAPK NES在C端融合的RanCas13b具有最高水平的干扰活性(图56A)。为了进一步区分前几个直系同源物的活性水平,我们就使用位置匹配的指导物来敲低KRAS转录物的能力将三种经优化的Cas13b构建体与最佳LwaCas13a-msfGFP融合体和shRNA进行了比较(图56B)。我们观察到PspCas13b的干扰水平最高(平均敲低62.9%),并因此选择其用于与LwaCas13a进行进一步比较。

[2868] 为了更严格地定义PspCas13b和LwaCas13a的活性水平,我们设计了沿Gluc和Cluc平铺的位置匹配的指导物,并使用我们的荧光素酶报告子测定了它们的活性。我们分别测试了93个和20个靶向Gluc和Cluc的位置匹配的指导物,并发现相对于LwaCas13a,PspCas13b的敲低水平持续增加(PspCas13b的敲低平均水平为92.3%,对比LwaCas13a的敲低平均水平为40.1%)(图49C、49D)。

[2869] Cas13哺乳动物干扰活性的特异性

[2870] 为了表征PspCas13b和LwaCas13a的干扰特异性,我们设计了荧光素酶靶标的质粒文库,该靶标文库在整个靶标序列以及三个侧接5'和3'碱基对中包含单错配和双错配(图56C)。我们用LwaCas13a或PspCas13b、靶向未修饰靶序列的固定指导RNA以及与合适系统相对应的错配靶标文库转染HEK293FT细胞。然后,我们对未裂解的转录物进行靶向RNA测序,以定量错配靶序列的耗竭。我们发现LwaCas13a和PspCas13b具有相对不耐受单个错配的中心区,从PspCas13b靶标的碱基对12-26和LwaCas13a靶标的碱基对13-24延伸(图56D)。双错配的耐受性比单个突变甚至更低,在较大的窗口(在各自的靶标中分别从PspCas13b的碱基对12-29和LwaCas13a的碱基对8-27延伸)中基本没有观察到的敲低活性(图56E)。另外,因为在侧接靶序列5'端和3'端的三个核苷酸中包括错配,所以我们可以评估PFS对Cas13敲低活性的限制。测序表明,几乎所有PFS组合都可实现稳健的敲低,这表明对于任一种测试的

酶，对哺乳动物细胞中干扰的PFS限制可能都不存在。这些结果表明，Cas13a和Cas13b显示出相似的序列限制和对错配的敏感性。

[2871] 接下来，我们表征了转录组的mRNA级分中PspCas13b和LwaCas13a的干扰特异性。我们进行了全转录组mRNA测序，以检测显著差异表达的基因。LwaCas13a和PspCas13b证实了Gluc的稳健敲低(图49E、49F)，并且与位置匹配的shRNA相比具有高度特异性，后者显示了数百个脱靶(图49G)。

[2872] Cas13-ADAR融合体实现了靶向RNA编辑

[2873] 鉴于PspCas13b在哺乳动物细胞中实现了一致、稳健和特异性的mRNA敲低，我们设想可以将其调整为RNA结合平台来募集ADAR的脱氨酶结构域($ADAR_{DD}$)进行可编程RNA编辑。为了工程化缺乏核酸酶活性的PspCas13b(dPspCas13b，这里称为dCas13b)，我们突变了HEPN结构域中的保守催化残基，并观察到荧光素酶RNA敲低活性的丧失(图57A)。我们假设，dCas13b-$ADAR_{DD}$融合体可以被指导RNA募集以靶向腺苷，其中杂交后的RNA为ADAR活性创造了所需的双链体底物(图50A)。为了提高靶腺苷脱氨率，我们在最初的RNA编辑设计中引入了另外两个修饰：我们引入了与靶腺苷相对的错配胞苷(先前据报道这会增加脱氨频率)，并将dCas13b与包含高活性突变以增强催化活性的人类ADAR1或ADAR2的脱氨酶结构域($ADAR1_{DD}$(E1008Q)(25)或$ADAR2_{DD}$(E488Q)(19))融合。

[2874] 为了测试dCas13b-$ADAR_{DD}$的活性，我们通过引入无义突变(W85X(UGG->UAG))在Cluc上产生了RNA编辑报告子，该无义突变功能性可以通过A->I编辑修复为野生型密码子(图50B)，然后检测为恢复Cluc发光。我们在靶腺苷上均匀地平铺了具有长度为30、50、70或84个核苷酸的间隔子的指导物，以确定最佳的指导物放置和设计(图50C)。我们发现dCas13b-$ADAR1_{DD}$需要更长的指导物才能修复Cluc报告子，而dCas13b-$ADAR2_{DD}$在测试的所有指导物长度下均起作用(图50C)。我们还发现，高活性E488Q突变提高了编辑效率，因为野生型$ADAR2_{DD}$的荧光素酶恢复有所减少(图57B)。根据所展示的活性，我们选择了dCas13b-$ADAR2_{DD}$(E488Q)进行进一步表征，并将该方法指定为可编程的A至I置换的RNA编辑版本1(REPAIRv1)。

[2875] 为了验证荧光素酶活性的恢复是由于真正的编辑事件引起的，我们直接通过逆转录和靶向下一代测序测量了经过REPAIRv1的Cluc转录物的编辑。我们测试了靶位点周围30nt和50nt的间隔子，并发现两种指导物长度均导致了预期的A至I编辑，而50nt的间隔子实现了更高的编辑百分比(图50D、50E、图57C)。我们还观察到50nt的间隔子在非靶向腺苷处的编辑倾向增加，这可能是由于双链体RNA的区增加(图50E、图57C)。

[2876] 接下来，我们靶向了内源基因PPIB。我们设计了平铺PPIB的50nt的间隔子，并发现可以编辑PPIB转录物，编辑效率高达28％(图57D)。为了测试是否可以进一步优化REPAIR，我们修饰了dCas13b和$ADAR2_{DD}$(E488Q)之间的接头(图57E、表8)，并发现接头选择适当地影响了荧光素酶活性恢复。

[2877] 定义用于RNA编辑的序列参数

[2878] 鉴于我们可以在测试位点实现精确的RNA编辑，我们想要表征针对转录组中任何RNA靶点进行系统编程的序列限制。序列限制可能是由于dCas13b靶向限制诸如PFS或者由于ADAR序列偏好引起的(26)。为了研究PFS对REPAIRv1的限制，我们设计了一个质粒文库，其在Cluc转录物的靶位点5'端携带一系列四个随机核苷酸(图51A)。我们靶向UAG或AAC基

序内的中心腺苷,并发现对于这两个基序,所有PFS都展示出可检测水平的RNA编辑,大多数PFS在靶位点处的编辑率均超过50%(图51B)。接下来,我们试图确定REPAIRv1中的ADAR2$_{DD}$是否具有紧密侧接靶向碱基的任何序列限制,如先前关于ADAR2$_{DD}$所报道的(26)。我们测试了直接围绕靶腺苷的5'和3'侧接核苷酸的每种可能组合(图51C),并且发现REPAIRv1能够编辑所有基序(图51D)。最后,我们分析了间隔序列中与靶标A相对的碱基的身份是否影响编辑效率,并且发现A-C错配的荧光素酶恢复率最高,而A-G、A-U和A-A的REPAIRv1活性大大降低(图57F)。

[2879] 使用REPAIRv1校正与疾病相关的人类突变

[2880] 为了证明REPAIRv1系统在哺乳动物细胞中进行RNA编辑的广泛适用性,我们针对两种与疾病相关的突变设计了REPAIRv1指导物:X联锁肾原性尿崩症中的878G>A(AVPR2W293X)和Fanconi贫血中的1517G>A(FANCC W506X)。我们将携带这些突变的基因的cDNA的表达构建体转染到HEK293FT细胞中,并测试了REPAIRv1是否可以校正所述突变。使用含有50nt的间隔子的指导RNA,我们能够实现35%的AVPR2校正和23%的FANCC校正(图52A-52D)。然后,我们测试了REPAIRv1校正34种与疾病相关的G>A突变的能力(表9),并且发现我们能够在33个位点实现重要编辑,编辑效率高达28%(图52E)。我们选择的突变仅是ClinVar数据库中致病性G到A突变(5,739个)的一部分,其中还包括另外11,943个G到A变异(图52F和图58)。因为没有序列限制,所以REPAIRv1能够潜在地编辑所有这些与疾病相关的突变,特别是鉴于我们观察到了重要编辑,而与靶基序无关(图51C和图52G)。

[2881] 向患病细胞递送REPAIRv1系统是治疗用途的前提,并因此,我们寻求设计出可以包装到治疗相关病毒载体诸如腺相关病毒(AAV)载体中的REPAIRv1构建体。AAV载体的包装限制为4.7kb,无法容纳大尺寸的dCas13b-ADAR$_{DD}$(4473bp)以及启动子和表达调控元件。为了减小尺寸,我们测试了与ADAR2$_{DD}$(E488Q)融合的dCas13的各种N端和C端截短物的RNA编辑活性。我们发现测试的所有C端截短物仍具有功能并能够恢复荧光素酶信号(图59),并且最大的截短物C端Δ984-1090(融合蛋白的总尺寸为4,152bp)足够小,以拟合在AAV载体的包装限制内。

[2882] REPAIRv1的全转录组特异性

[2883] 尽管利用PspCas13b进行RNA敲低具有高度特异性,但在我们的荧光素酶平铺实验中,我们在指导物:靶标双链体内观察到了脱靶腺苷编辑(图50E)。为了查看这是否是普遍现象,我们平铺了内源转录物KRAS,并测量了靶腺苷附近的脱靶编辑程度(图53A)。我们发现,对于KRAS,尽管中靶编辑率为23%,但靶位点周围有许多位点也具有可检测的A到G编辑(图53B)。

[2884] 由于在指导物:靶标双链体内观察到了脱靶编辑,我们通过对所有mRNA进行RNA测序评价了所有可能的转录组脱靶。RNA测序显示,存在大量A到G脱靶事件,其中在靶向条件下有1,732个脱靶,并且在非靶向条件下有925个脱靶,其中828个脱靶重叠(图53C、53D)。在整个转录组的所有编辑位点中,中靶编辑位点的编辑率最高,其中A到G转化率为89%。

[2885] 鉴于Cas13靶向的高度特异性,我们认为脱靶可能源自ADAR。先前已经描述了两种RNA引导的ADAR系统(图60A)。第一种利用ADAR2$_{DD}$与小病毒蛋白lambda N (λN)的融合体,该蛋白与BoxB-λRNA发夹结合(22)。具有双BoxB-λ发夹的指导RNA指导ADAR2$_{DD}$编辑在指导RNA中编码的位点(23)。第二种设计利用全长ADAR2(ADAR2)和具有发夹的指导RNA,ADAR2的

双链RNA结合结构域(dsRBD)识别该发夹(21,24)。我们分析了这两种系统相较于REPAIRv1的编辑效率,并发现BoxB-ADAR2和ADAR2系统分别显示出63%和36%的编辑率,相较而言通过REPAIRv1达到了89%的编辑率(图60B-60E)。另外,与REPAIRv1靶向指导物条件下的1,229个脱靶相比,在靶向指导物条件下BoxB和ADAR2系统分别产生了2018个和174个观察到的脱靶。值得注意的是,在两种基于$ADAR2_{DD}$的系统(REPAIRv1和BoxB)下所有条件都显示出高百分比的脱靶重叠,而ADAR2系统却有截然不同组的脱靶(图60F)。靶向条件和非靶向条件之间以及REPAIRv1和BoxB条件之间的脱靶重叠表明,$ADAR2_{DD}$驱使脱靶,与dCas13靶向无关(图60F)。

[2886] 通过合理的蛋白质工程化提高REPAIRv1的特异性

[2887] 为了提高REPAIR的特异性,我们采用了$ADAR2_{DD}$(E488Q)的结构指导的蛋白质工程化。由于脱靶具有与指导物无关的性质,我们假设使$ADAR2_{DD}$(E488Q)-RNA结合失稳会选择性地降低脱靶编辑,但由于dCas13b将$ADAR2_{DD}$(E488Q)栓系至靶位点而增加了局部浓度,因此会维持中靶编辑。我们对先前确定与$ADAR2_{DD}$(E488Q)背景上的靶RNA的双链体区接触的$ADAR2_{DD}$(E488Q)残基(图54A)(18)进行了诱变。为了评估效率和特异性,我们假设在检测到的非靶向条件下的本底荧光素酶恢复将指示较宽的脱靶活性,利用靶向和非靶向指导物测试了17个单突变体。我们发现,选定残基处的突变对靶向和非靶向指导物的荧光素酶活性具有显著影响(图54A、54B、图61A)。大多数突变体要么显著提高了靶向指导物的荧光素酶活性,要么增加了靶向与非靶向指导物活性之比(我们将其称为特异性评分)(图54A、54B)。我们选择了这些突变体的一个子集(图54B)用于通过下一代测序进行全转录组特异性谱分析。如预期的那样,从全转录组测序中测得的脱靶与突变体的特异性评分相关(图61B)。我们发现,除了$ADAR2_{DD}$(E488Q/R455E)外,所有测序的REPAIRv1突变体均可以有效编辑报告子转录物(图54C),许多突变体显示出脱靶数量减少(图61C、62)。我们进一步探索了特异性突变体的脱靶的周围基序,并且发现REPAIRv1和大多数工程化突变体对其编辑表现出强烈的3'G偏好,与被表征的ADAR2基序一致(图63A)(26)。我们选择了突变体$ADAR2_{DD}$(E488Q/T375G)进行后续实验,因为该突变体在全转录组脱靶数量最低的四个突变体中编辑百分比最高,并将其称为REPAIRv2。与REPAIRv1相比,REPAIRv2表现出增加的特异性,其中转录组脱靶从1732个减少到10个(图54D)。在Cluc中靶向腺苷周围的区中,REPAIRv2具有在测序迹线中可见的减少的脱靶编辑(图54E)。在来源于下一代测序的基序中,REPAIRv1表现出对3'G的强烈偏好,但对所有基序均显示出脱靶编辑(图63B);相比之下,REPAIRv2仅编辑最强的脱靶基序(图63C)。转录物上的编辑分布严重偏斜,高度编辑的基因具有60多个编辑(图64A、64B),而REPAIRv2多次仅编辑一个转录物(EEF1A1)(图64D-64F)。预测REPAIRv1脱靶编辑会导致多种变异,包括1000个错义突变(图64C)和93个致癌事件(图64D)。相比之下,REPAIRv2仅具有6个错义突变(图64E),没有一个有致癌后果(图64F)。预测的脱靶效应的这种降低将REPAIRv2与其他RNA编辑方法区分开来。利用不同剂量的指导RNA进行的实验表明,剂量响应可降低脱靶活性(图68)。

[2888] 我们将REPAIRv2靶向内源基因,以测试特异性增强的突变是否能减少靶转录物附近的编辑,同时保持高效的中靶编辑。对于靶向KRAS或PPIB的指导物,我们发现,REPAIRv2没有可检测的脱靶编辑,这与REPAIRv1不同,并且可以分别以27.1%和13%有效地编辑中靶腺苷(图54F、54G)。这种特异性扩展到其他靶位点,包括在REPAIRv1的非靶向条件下显示

出高水平本底的区，诸如其他KRAS或PPIB靶位点(图65A-65C)。总体而言，REPAIRv2消除了双链体区中编辑后的腺苷周围的脱靶现象，并显示出显著增强的全转录组特异性。

[2889] 结论

[2890] 我们在这里表明，RNA引导的RNA靶向VI-B型效应物Cas13b能够高效且高特异性地敲低RNA，从而为改进用于探询必要基因和非编码RNA的工具提供了基础，以及可以在转录组学水平控制细胞过程。无催化活性的Cas13b(dCas13b)保留了可编程RNA结合能力，在这里我们通过将dCas13b与腺苷脱氨酶ADAR2融合以实现精确的A到I编辑(我们称为REPAIRv1(可编程的A到I置换的RNA编辑版本1)的系统)来利用这种结合能力。对该系统进行进一步工程化产生了REPAIRv2，这种方法相对于当前的编辑平台具有相当的或增加的活性，并且具有显著提高的特异性。

[2891] 尽管Cas13b表现出高保真性，但我们对dCas13b-ADAR2$_{DD}$融合体的初步研究结果显示有数千个脱靶。为了解决这个问题，我们采用了一种合理的诱变策略来改变与RNA双链体接触的ADAR2$_{DD}$残基，从而鉴定出一个变体ADAR2$_{DD}$(E488Q/T375G)，其在与dCas13b融合时能够进行精确、有效和高度特异性的编辑。利用此变体的编辑效率可与利用两个当前可用的系统BoxB-ADAR$_{DD}$或ADAR2编辑实现的相当或比其更好。此外，REPAIRv2系统在整个转录组中仅形成了10个可观察到的脱靶，比两种替代编辑技术好至少一个数量级。

[2892] 与其他核酸编辑工具相比，REPAIR系统提供许多优势。首先，可以通过将胞嘧啶置于所需的腺苷对面的指导物延伸区中，以产生对ADAR编辑活性而言理想的有利A-C错配，从而在指导物中编码精确的靶位点。其次，Cas13没有靶向序列限制诸如PFS或PAM，并且在靶腺苷周围没有基序偏好，因此允许转录组中的任何腺苷都可能被REPAIR系统靶向。但是，我们确实注意到，DNA碱基编辑器可以靶向有义链或反义链，而REPAIR系统仅限于转录序列，从而限制了我们可以靶向的可能编辑位点的总数。但是，由于利用REPAIR靶向的更有灵活性的性质，因此该系统相较于Cas9-DNA碱基编辑器可以在ClinVar内实现更多的编辑(图52C)。第三，REPAIR系统将靶腺苷直接脱氨为肌苷，并且不依赖于内源修复途径诸如碱基切除或错配修复来产生所需的编辑结果。因此，REPAIR在不能支持其他形式编辑的非分裂细胞中应该是可能的。第四，RNA编辑可以是瞬时的，从而有可能暂时控制编辑结果。这种特性可能对治疗由细胞状态的暂时变化引起的疾病诸如局部炎症有用。

[2893] REPAIR系统为附加工程化提供了多种机会。Cas13b具有前crRNA加工活性(13)，允许对多种变体进行多重编辑，仅仅如此可能不会改变疾病风险，但一起可能具有累加效应和改变疾病的潜力。扩展我们合理的设计方法诸如结合有前景的突变可以进一步提高系统的特异性和效率，而无偏倚的筛选方法可以鉴定其他残基以改善REPAIR活性和特异性。

[2894] 当前，通过REPAIR可实现的碱基转化仅限于由腺苷生成肌苷；dCas13与其他催化RNA编辑结构域诸如APOBEC的其他融合体可以能够实现胞苷到尿苷编辑。另外，ADAR的诱变可以放宽底物对靶胞苷的偏好，从而使C->U编辑器可以利用双链体RNA底物要求来赋予增强的特异性。通过形成DNA-RNA异源双链体靶标(27)或诱变ADAR结构域，利用无催化活性的DNA-靶向CRISPR效应物例如dCas9或dCpf1，也可以在DNA底物上进行腺苷到肌苷编辑。

[2895] REPAIR可用于其中A到I(A到G)编辑可以逆转或减慢疾病进展的一系列治疗适应症(图66A-66G)。首先，在与疾病相关的组织中表达REPAIR以靶向病因性孟德尔G到A突变可用于恢复有害突变并治疗疾病。例如，在脑组织中通过AAV稳定表达REPAIR可用于校正引起

局灶性癫痫的GRIN2A错义突变c.2191G>A(Asp731Asn)(28)或引起早发性阿尔茨海默病的APP错义突变c.2149G>A(Val717Ile)(29)。其次,REPAIR可用于通过修饰参与疾病相关信号转导的蛋白质的功能来治疗疾病。例如,REPAIR编辑将允许重新编码作为激酶靶标的一些丝氨酸、苏氨酸和酪氨酸残基(图66A-66G)。与疾病相关的蛋白质中这些残基的磷酸化会影响许多病症的进展,包括阿尔茨海默病和多种神经退行性疾患(30)。第三,REPAIR可用于改变表达的风险修饰的G到A变体的序列,以抢先降低患者进入疾病状态的机会。最有趣的案例是"保护性"风险修饰的等位基因,它显著降低了进入疾病状态的机会,并且在某些情况下,还可以带来其他健康益处。例如,REPAIR可用于在功能上模拟分别预防心血管疾病和银屑病关节炎的PCSK9和IFIH1的A到G等位基因(31)。最后,REPAIR可用于治疗性修饰外显子调节疗法的剪接受体和供体位点。REPAIR可以将AU变为IU或将AA变为AI,分别是共有5'剪接供体或3'剪接受体位点的功能等效物,从而形成新的剪接点。另外,REPAIR编辑可以将共有3'剪接受体位点从AG突变为IG,以促进相邻下游外显子的跳跃,这种治疗策略在治疗DMD中引起了广泛关注。剪接位点的调节在反义寡聚物已取得一定成功的疾病中可能具有广泛的应用,诸如调节SMN2剪接以治疗脊髓性肌肉萎缩(32)。

[2896] 我们证实了PspCas13b酶既可用作RNA敲低工具,又可用作RNA编辑工具。用于可编程RNA结合的dCas13b平台有许多应用,包括实时转录成像、剪接修饰、转录物的靶向定位、RNA结合蛋白的下拉以及表观转录组修饰。这里,我们使用dCas13产生REPAIR,为现有的成套核酸编辑技术注入了新的力量。REPAIR提供了一种用于治疗遗传疾病或模拟保护等位基因的新方法,并将RNA编辑确立为修饰遗传功能的有用工具。

[2897] 表6本研究中使用的Cas13直系同源物

[2898]

| Cas13 ID | Cas13 缩写 | 宿主生物体 | 蛋白登录 |
|---|---|---|---|
| Cas13a1 | LshCas13a | 沙氏纤毛菌 | WP_018451595.1 |
| Cas13a2 | LwaCas13a | 瓦氏纤毛菌(Lw2) | WP_021746774.1 |
| Cas13a3 | LseCas13a | 西尔李斯特菌 | WP_012985477.1 |
| Cas13a4 | LbmCas13a | 毛螺菌科细菌 MA2020 | WP_044921188.1 |
| Cas13a5 | LbnCas13a | 毛螺菌科细菌 NK4A179 | WP_022785443.1 |
| Cas13a6 | CamCas13a | 嗜胺[梭菌]DSM 10710 | WP_031473346.1 |
| Cas13a7 | CgaCas13a | 鸡肉食杆菌 DSM 4847 | WP_034560163.1 |
| Cas13a8 | Cga2Cas13a | 鸡肉食杆菌 DSM 4847 | WP_034563842.1 |
| Cas13a9 | Pprcas13a | 产丙酸栖沼泽杆菌 WB4 | WP_013443710.1 |
| Cas13a10 | LweCas13a | 魏斯特史氏李斯特菌 FSL R9-0317 | WP_036059185.1 |
| Cas13a11 | LbfCas13a | 李斯特氏菌科菌 FSL M6-0635 | WP_036091002.1 |
| Cas13a12 | Lwa2cas13a | 瓦氏纤毛菌 F0279 | WP_021746774.1 |
| Cas13a13 | RcsCas13a | 英膜红细菌 SB 1003 | WP_013067728.1 |
| Cas13a14 | RcrCas13a | 英膜红细菌 R121 | WP_023911507.1 |
| Cas13a15 | RcdCas13a | 英膜红细菌 DE442 | WP_023911507.1 |
| Cas13a16 | LbuCas13a | 口腔纤毛菌 C-1013-b | WP_015770004.1 |
| Cas13a17 | HheCas13a | 解半纤维素赫氏菌 | CRZ35554.1 |
| Cas13a18 | EreCas13a | 直肠[真杆菌] | WP_055061018.1 |
| Cas13a19 | EbaCas13a | 真杆菌科菌 CHKCI004 | WP_090127496.1 |
| Cas13a20 | BmaCas13a | 布劳特氏菌属某种 Marseille-P2398 | WP_062808098.1 |
| Cas13a21 | LspCas13a | 纤毛菌属某种口腔分类群 879 菌株 F0557 | WP_021744063.1 |
| Cas13b1 | BzoCas13b | 动物溃疡伯杰菌 | WP_002664492 |
| Cas13b2 | PinCas13b | 中间普雷沃菌 | WP_036860899 |
| Cas13b3 | PbuCas13b | 颊普雷沃菌 | WP_004343973 |
| Cas13b4 | AspCas13b | 另枝菌属某种 ZOR0009 | WP_047447901 |
| Cas13b5 | PsmCas13b | 普雷沃菌属某种 MA2016 | WP_036929175 |
| Cas13b6 | RanCas13b | 鸭疫里默氏杆菌 | WP_004919755 |
| Cas13b7 | PauCas13b | 桔红色普雷沃菌 | WP_025000926 |
| Cas13b8 | PsaCas13b | 解糖普雷沃菌 | WP_051522484 |
| Cas13b9 | Pin2Cas13b | 中间普雷沃菌 | WP_061868553 |
| Cas13b10 | CcaCas13b | 犬咬二氧化碳噬纤维菌 | WP_013997271 |
| Cas13b11 | PguCas13b | 动物口腔卟啉单胞菌 | WP_039434803 |
| Cas13b12 | PspCas13b | 普雷沃菌属某种 P5-125 | WP_044065294 |

| | | | |
|---|---|---|---|
| Cas13b13 | FbrCas13b | 嗜鳃黄杆菌 | WP_014084666 |
| Cas13b14 | PgiCas13b | 牙龈卟啉单胞菌 | WP_053444417 |
| Cas13b15 | Pin3Cas13b | 中间普雷沃菌 | WP_050955369 |
| Cas13c1 | FnsCas13c | 坏死梭杆菌肠形亚种 ATCC 51357 contig00003 | WP_005959231.1 |
| Cas13c2 | FndCas13c | 坏死梭杆菌 DJ-2 contig0065，全基因组鸟枪序列 | WP_035906563.1 |
| Cas13c3 | FnbCas13c | 坏死梭杆菌 BFTR-1 contig0068 | WP_035935671.1 |
| Cas13c4 | FnfCas13c | 坏死梭杆菌肠形亚种 1_1_36S cont1.14 | EHO19081.1 |
| Cas13c5 | FpeCas13c | 极臭梭杆菌 ATCC 29250 T364DRAFT_scaffold00009.9_C | WP_027128616.1 |
| Cas13c6 | FulCas13c | 溃疡梭杆菌 ATCC 49185 cont2.38 | WP_040490876.1 |
| Cas13c7 | AspCas13c | 厌氧盐杆菌属某种 ND1 基因组组装马赛厌氧盐杆菌 ND1 | WP_042678931.1 |

[2899]

[2900] 表7细菌筛选中的PFS临界值

| Cas13b 直系同源物 | 键值 | 用于生成 PFS 基序的-$Log_2$耗竭评分 |
|---|---|---|
| *动物溃疡伯杰菌* | 1 | 2 |
| 中间普雷沃菌基因座 1 | 2 | 1 |
| *颊普雷沃菌* | 3 | 3 |
| *另枝菌属某种 ZOR0009* | 4 | 1 |
| *普雷沃菌属某种 MA2016* | 5 | 2 |
| *鸭疫里默氏杆菌* | 6 | 4 |
| *桔红色普雷沃菌* | 7 | 1 |
| *解糖普雷沃菌* | 8 | 0 |
| 中间普雷沃菌基因座 2 | 9 | 0 |
| *犬咬二氧化碳噬纤维菌* | 10 | 3 |

[2901]

[2902]

| | | |
|---|---|---|
| *动物口腔卟啉单胞菌* | 11 | 4 |
| *普雷沃菌属某种 P5-125* | 12 | 2.1 |
| *嗜鳃黄杆菌* | 13 | 1 |
| *牙龈卟啉单胞菌* | 14 | 3 |
| 中间普雷沃菌基因座 2 | 15 | 4 |

[2903] 表8本研究中用于在哺乳动物细胞中进行RNA编辑的dCas13b-ADAR接头序列。

[2904]

| 图 | 接头 |
|---|---|
| 50C | GSGGGGS (SEQ ID No. 296) |
| 50E | GS |
| 57B | GSGGGGS (SEQ ID No. 296) |
| 57C | GS |
| 57D | GS |
| 57E: GS | GS |
| 57E: GSGGGGS | GSGGGGS |
| 57E: (GGGS)3 | GGGGSGGGGSGGGGS (SEQ ID No. 802) |
| 57E: Rigid | EAAAK (SEQ ID No. 803) |
| 57E: (GGS)6 | GGSGGSGGSGGSGGSGGS (SEQ ID No. 804) |
| 57E: XTEN | SGSETPGTSESATPES (SEQ ID No. 805) |
| 51B | GS |
| 57F | GS |
| 51C | GS |
| 52B | GS |
| 52D | GS |
| 52E | GS |
| 51A: Δ984-1090, Δ1026-1090, Δ1053-1090 | GS |
| 51A: Δ1-125, Δ1-88, Δ1-72 | GSGGGGS |
| 53B | GS |
| 53C | GS |
| 53D | GS |
| 60A | GS |
| 60C | GS |
| 60D | GS |

[2905]

| | |
|---|---|
| 61D | GS |
| 54A | GS |
| 62A | GS |
| 54B | GS |
| 62B | GS |
| 62C | GS |
| 63A | GS |
| 63B | GS |
| 54C | GS |
| 54D | GS |
| 54E | GS |
| 54F | GS |
| 66A | GS |
| 66A | GS |

[2906] 表9与疾病相关的突变的疾病信息

[2907]

| **全长候选物** | **基因** | **疾病** |
|---|---|---|
| NM_000054.4(AVPR2):c.878G>A (p.Trp293Ter) | AVPR2 | 肾原性尿崩症，X 连锁 |
| NM_000136.2(FANCC):c.1517G>A (p.Trp506Ter) | FANCC | Fanconi 贫血，互补组 C |
| **其他模拟候选物** | | |
| **候选物** | **基因** | **疾病** |
| NM_000206.2(IL2RG):c.710G>A (p.Trp237Ter) | IL2RG | X 连锁严重联合免疫缺陷症 |
| NM_000132.3(F8):c.3144G>A (p.Trp1048Ter) | F8 | 遗传性因子 VIII 缺乏性疾病 |
| NM_000527.4(LDLR):c.1449G>A (p.Trp483Ter) | LDLR | 家族性高胆固醇血症 |
| NM_000071.2(CBS):c.162G>A (p.Trp54Ter) | CBS | 由于 CBS 缺乏引起的高胱氨酸尿 |
| NM_000518.4(HBB):c.114G>A (p.Trp38Ter) | HBB | β^0^地中海贫血/β 地中海贫血 |
| NM_000035.3(ALDOB):c.888G>A (p.Trp296Ter) | ALDOB | 遗传性果糖尿症 |
| NM_004006.2(DMD):c.3747G>A (p.Trp1249Ter) | DMD | 杜兴氏肌营养不良 |
| NM_005359.5(SMAD4):c.906G>A (p.Trp302Ter) | SMAD4 | 青少年息肉综合征 |
| NM_000059.3(BRCA2):c.582G>A (p.Trp194Ter) | BRCA2 | 家族性乳腺癌/家族性乳腺癌-卵巢癌 2 |
| NM_000833.4(GRIN2A):c.3813G>A (p.Trp1271Ter) | GRIN2A | 局灶性癫痫，伴语言障碍并且伴或不伴智力障碍 |
| NM_002977.3(SCN9A):c.2691G>A (p.Trp897Ter) | SCN9A | 疼痛淡漠，先天性常染色体隐性遗传 |

| | | | |
|---|---|---|---|
| | NM_007375.3(TARDBP):c.943G>A (p.Ala315Thr) | TARDBP | 10 型肌萎缩性脊髓侧索硬化症 |
| | NM_000492.3(CFTR):c.3846G>A (p.Trp1282Ter) | CFTR | 囊性纤维化/遗传性胰腺炎/未提供/ataluren 响应-功效 |
| | NM_130838.1(UBE3A):c.2304G>A (p.Trp768Ter) | UBE3A | 安格尔曼综合征 |
| | NM_000543.4(SMPD1):c.168G>A (p.Trp56Ter) | SMPD1 | A 型尼曼匹克病 |
| | NM_206933.2(USH2A):c.9390G>A (p.Trp3130Ter) | USH2A | 2A 型乌谢尔综合征 |
| | NM_130799.2(MEN1):c.1269G>A (p.Trp423Ter) | MEN1 | 遗传性癌症易感综合征 |
| | NM_177965.3(C8orf37):c.555G>A (p.Trp185Ter) | C8orf37 | 色素性视网膜炎 64 |
| | NM_000249.3(MLH1):c.1998G>A (p.Trp666Ter) | MLH1 | 林奇综合征 |
| | NM_000548.4(TSC2):c.2108G>A (p.Trp703Ter) | TSC2 | 结节性硬化症 2\|结节性硬化综合征 |
| | NM_000267.3(NF1):c.7044G>A (p.Trp2348Ter) | NF1 | 1 型神经纤维瘤病 |
| | NM_000179.2(MSH6):c.3020G>A (p.Trp1007Ter) | MSH6 | 林奇综合征 |
| [2908] | NM_000344.3(SMN1):c.305G>A (p.Trp102Ter) | SMN1 | II 型脊髓性肌肉萎缩/库格尔伯格-魏兰德病 |
| | NM_024577.3(SH3TC2):c.920G>A (p.Trp307Ter) | SH3TC2 | 4C 型夏科-马利-杜斯氏病 |
| | NM_001369.2(DNAH5):c.8465G>A (p.Trp2822Ter) | DNAH5 | 原发性纤毛运动障碍 |
| | NM_004992.3(MECP2):c.311G>A (p.Trp104Ter) | MECP2 | 瑞特综合征 |
| | NM_032119.3(ADGRV1):c.7406G>A (p.Trp2469Ter) | ADGRV1 | 2C 型乌舍尔综合征 |
| | NM_017651.4(AHI1):c.2174G>A (p.Trp725Ter) | AHI1 | 朱伯特综合征 3 |
| | NM_004562.2(PRKN):c.1358G>A (p.Trp453Ter) | PRKN | 帕金森病 2 |
| | NM_000090.3(COL3A1):c.3833G>A (p.Trp1278Ter) | COL3A1 | 4 型埃勒斯-当洛斯综合征 |
| | NM_007294.3(BRCA1):c.5511G>A (p.Trp1837Ter) | BRCA1 | 家族性乳腺癌/家族性乳腺癌-卵巢癌 1 |
| | NM_000256.3(MYBPC3):c.3293G>A (p.Trp1098Ter) | MYBPC3 | 原发性家族性肥厚型心肌病 |
| | NM_000038.5(APC):c.1262G>A (p.Trp421Ter) | APC | 家族性腺瘤性息肉病 1 |
| | NM_001204.6(BMPR2):c.893G>A (p.W298*) | BMPR2 | 原发性肺动脉高压 |

[2909] 表10:本研究中使用的关键质粒

[2910]

| 质粒名称 | 描述 |
|---|---|
| pAB0006 | CMV-C 荧光素酶-聚 A EF1a-G-荧光素酶-聚 A |
| pAB0040 | CMV-C 荧光素酶(STOP85)-聚 A EF1a-G-荧光素酶-聚 A |
| pAB0048 | pCDNA-ADAR2-N-末端 B12-HIV NES |
| pAB0050 | pAB0001-A02 (crRNA 主链) |
| pAB0051 | pAB0001-B06 (crRNA 主链) |
| pAB0052 | pAB0001-B11 (crRNA 主链) |
| pAB0053 | pAB0001-B12 (crRNA 主链) |
| pAB0014.B6 | EF1a-BsiWI-Cas13b6-NES-mapk |
| pAB0013.B11 | EF1a-BsiWI-Cas13b11-NES-HIV |
| pAB0013.B12 | EF1a-BsiWI-Cas13b12-NES-HIV |
| pAB0051 | pAB0001-B06 (crRNA 主链) |
| pAB0052 | pAB0001-B11 (crRNA 主链) |
| pAB0053 | pAB0001-B12 (crRNA 主链) |
| pAB0079 | pCDNA-ADAR1hu-EQ 突变体-N-末端目的载体 |
| pAB0085 | pCDNA-ADAR2 (E488Q)hu-EQ 突变体-N-末端目的载体 |
| pAB0095 | EF1a-BsiWI-Cas13-B12-NES-HIV，具有双 H HEPN 突变体 |
| pAB0114 | pCDNA-wtADAR2hu-EQ 突变体-N-末端目的载体 |
| pAB0120 | 最佳荧光素酶 ADAR 指导物(来自 wC0054 的指导物 24) |
| pAB0122 | pAB0001-B12 NT 指导物，用于 ADAR 实验 |
| pAB0151 | pCDNA-ADAR2hu-EQ 突变体-N-末端目标载体 C-末端 Δ 984-1090 |
| pAB0180 | T375G 特异性突变体 |
| pAB0181 | T375G Cas13b C-末端 Δ 984-1090 |

[2911] 表11：本研究中用于在哺乳动物细胞中进行敲低的指导/shRNA序列

[2912]

| 名称 | 间隔序列 | 干扰机制 | 注意 | 第一图 |
|---|---|---|---|---|
| 细菌 PFS 指导物 | GCCAGCUUUCCGGGCAUUGGCUUCCAUC (SEQ ID No. 300) | Cas13b | 用于所有直系同源物 | |
| Cas13a-Gluc 指导物 | GCCAGCTTTCCGGGCATTGGCTTCCATC (SEQ ID No. 301) | Cas13a | 用于所有 Cas13a 直系 | 图 49B |

| | | | | |
|---|---|---|---|---|
| 1 | | | 同源物 | |
| Cas13a-Gluc 指导物 2 | ACCCAGGAATCTCAGGAATGTCGACGAT (SEQ ID No. 302) | Cas13a | 用于所有 Cas13a 直系同源物 | 图 49B |
| Cas13a-非靶向指导物(LacZ) | AGGGTTTTCCCAGTCACGACGTTGTAAA (SEQ ID No. 303) | Cas13a | 用于所有 Cas13a 直系同源物 | 图 49B |
| Cas13b-Gluc 指导物 1.1 | GGGCATTGGCTTCCATCTCTTTGAGCACCT (SEQ ID No. 304) | Cas13b | 用于直系同源物 1-3、6、7、10、11、12、14、15 | 图 49B |
| Cas13b-Gluc 指导物 1.2 | GUGCAGCCAGCUUUCCGGGCAUUGGCUUCC (SEQ ID No. 305) | Cas13b | 用于直系同源物 4 | 图 49B |
| Cas13b-Gluc 指导物 1.3 | GCAGCCAGCUUUCCGGGCAUUGGCUUCCAU (SEQ ID No. 306) | Cas13b | 用于直系同源物 5 | 图 49B |
| Cas13b-Gluc 指导物 1.4 | GGCUUCCAUCUCUUUGAGCACCUCCAGCGG (SEQ ID No. 307) | Cas13b | 用于直系同源物 8、9 | 图 49B |
| Cas13b-Gluc 指导物 1.5 | GGAAUGUCGACGAUCGCCUCGCCUAUGCCG (SEQ ID No. 308) | Cas13b | 用于直系同源物 13 | 图 49B |
| Cas13b-Gluc 指导物 2.1 | GAAUGUCGACGAUCGCCUCGCCUAUGCCGC (SEQ ID No. 309) | Cas13b | 用于直系同源物 1-3、6、7、10、11、14、15 | 图 49B |
| Cas13b-Gluc 指导物 2.2 | GACCUGUGCGAUGAACUGCUCCAUGGGCUC (SEQ ID No. 310) | Cas13b | 用于直系同源物 12 | 图 49B |
| Cas13b-Gluc 指导物 2.2 | GUGUGGCAGCGUCCUGGGAUGAACUUCUUC (SEQ ID No. 311) | Cas13b | 用于直系同源物 4 | 图 49B |
| Cas13b-Gluc 指导物 2.3 | GUGGCAGCGUCCUGGGAUGAACUUCUUCAU (SEQ ID No. 312) | Cas13b | 用于直系同源物 5 | 图 49B |
| Cas13b-Gluc 指导物 2.4 | GCUUCUUGCCGGGCAACUUCCCGCGGUCAG (SEQ ID No. 313) | Cas13b | 用于直系同源物 8、9 | 图 49B |
| Cas13b-Gluc 指导物 2.6 | GCAGGGUUUUCCCAGUCACGACGUUGUAAAA (SEQ ID No. 314) | Cas13b | 用于直系同源物 13 | 图 49B |
| Cas13B-非靶向指导物 | GCAGGGUUUUCCCAGUCACGACGUUGUAAAA (SEQ ID No. 315) | Cas13b | 用于所有直系同源物 | 图 49B |

[2913]

[2914]

| | | | | |
|---|---|---|---|---|
| Cas13a-Gluc 指导物-RNASeq | ACCCAGGAAUCUCAGGAAUGUCGACGAU (SEQ ID No. 316) | Cas13a | | 图 49E |
| shRNA-Gluc 指导物 | CAGCTTTCCGGGCATTGGCTT (SEQ ID No. 317) | shRNA | | 图 49F |
| Cas13b-Gluc 指导物-RNASeq | CCGCUGGAGGUGCUCAAAGAGAUGGAAGCC (SEQ ID No. 318) | Cas13b | | 图 49F |
| Cas13a-Gluc-指导物-1 | GCCAGCTTTCCGGGCATTGGCTTCCATC (SEQ ID No. 319) | Cas13a | | 图 56A |
| Cas13a-Gluc-指导物-2 | ACCCAGGAATCTCAGGAATGTCGACGAT (SEQ ID No. 320) | Cas13a | | 图 56A |
| Cas13b-Gluc-opt-指导物-1 | GGGCATTGGCTTCCATCTCTTTGAGCACCT (SEQ ID No. 321) | Cas13b | | 图 56A |
| Cas13b-Gluc-opt-指导物-2 | GAAUGUCGACGAUCGCCUCGCCUAUGCCGC (SEQ ID No. 322) | Cas13b | | 图 56A |
| Cas13a KRAS 指导物 1 | CAAGGCACTCTTGCCTACGCCACCAGCT (SEQ ID No. 323) | Cas13a | | 图 56B |
| Cas13a KRAS 指导物 2 | TCATATTCGTCCACAAAATGATTCTGAA (SEQ ID No. 324) | Cas13a | | 图 56B |
| Cas13a KRAS 指导物 3 | ATTATTTATGGCAAATACACAAAGAAAG (SEQ ID No. 325) | Cas13a | | 图 56B |
| Cas13a KRAS 指导物 4 | GAATATCTTCAAATGATTTAGTATTATT (SEQ ID No. 326) | Cas13a | | 图 56B |
| Cas13a KRAS 指导物 5 | ACCATAGGTACATCTTCAGAGTCCTTAA (SEQ ID No. 327) | Cas13a | | 图 56B |
| Cas13b KRAS 指导物 1 | GTCAAGGCACTCTTGCCTACGCCACCAGCT (SEQ ID No. 328) | Cas13b | | 图 56B |
| Cas13b KRAS 指导物 2 | GATCATATTCGTCCACAAAATGATTCTGAA (SEQ ID No. 329) | Cas13b | | 图 56B |
| Cas13b KRAS 指导物 3 | GTATTATTTATGGCAAATACACAAAGAAAG (SEQ ID No. 330) | Cas13b | | 图 56B |

| | | | | |
|---|---|---|---|---|
| Cas13b KRAS 指导物 4 | GTGAATATCTTCAAATGATTTAGTATTATT (SEQ ID No. 331) | Cas13b | | 图 56B |
| Cas13b KRAS 指导物 5 | GGACCATAGGTACATCTTCAGAGTCCTTAA (SEQ ID No. 332) | Cas13b | | 图 56B |
| shRNA KRAS 指导物 1 | aagagtgccttgacgatacagcCTCGAGgctgtatcgtcaaggcactctt (SEQ ID No. 333) | shRNA | | 图 56B |
| shRNA KRAS 指导物 2 | aatcattttgtggacgaatatCTCGAGatattcgtccacaaaatgatt (SEQ ID No. 334) | shRNA | | 图 56B |
| shRNA KRAS 指导物 3 | aaataatactaaatcatttgaCTCGAGtcaaatgatttagtattattt (SEQ ID No. 335) | shRNA | | 图 56B |
| shRNA KRAS 指导物 4 | aataatactaaatcatttgaaCTCGAGttcaaatgatttagtattatt (SEQ ID No. 336) | shRNA | | 图 56B |
| shRNA KRAS 指导物 5 | aaggactctgaagatgtacctCTCGAGaggtacatcttcagagtcctt (SEQ ID No. 337) | shRNA | | 图 56B |

[2915]

[2916] 表12:用于Gluc敲低的指导序列

| 名称 | 间隔序列 | 位置 | 注意 | 第一图 |
|---|---|---|---|---|
| Gluc 平铺指导物 1 | GAGATCAGGGCAAACAGAACTTTGACTCCC (SEQ ID No. 338) | 2 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49C |
| Gluc 平铺指导物 2 | GGATGCAGATCAGGGCAAACAGAACTTTGA (SEQ ID No. 339) | 7 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49C |
| Gluc 平铺指导物 3 | GCACAGCGATGCAGATCAGGGCAAACAGAA (SEQ ID No. 340) | 13 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49C |
| Gluc 平铺指导物 4 | GCTCGGCCACAGCGATGCAGATCAGGGCAA (SEQ ID No. 341) | 19 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49C |
| Gluc 平铺指导物 5 | GGGGCTTGGCCTCGGCCACAGCGATGCAGA (SEQ ID No. 342) | 28 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49C |
| Gluc 平铺 | GTGGGCTTGGCCTCGGCCACAG | 29 | 注意，Cas13a | 49C |

[2917]

[2918]

| | | | | |
|---|---|---|---|---|
| 指导物 6 | CGATGCAG<br>(SEQ ID No. 343) | | 间隔子在 5'端被截短两个核苷酸 | |
| Gluc 平铺指导物 7 | GTCTCGGTGGGCTTGGCCTCGGCCACAGCG<br>(SEQ ID No. 344) | 35 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49C |
| Gluc 平铺指导物 8 | GTTCGTTGTTCTCGGTGGGCTTGGCCTCGG<br>(SEQ ID No. 345) | 43 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49C |
| Gluc 平铺指导物 9 | GGAAGTCTTCGTTGTTCTCGGTGGGCTTGG<br>(SEQ ID No. 346) | 49 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49C |
| Gluc 平铺指导物 10 | GATGTTGAAGTCTTCGTTGTTCTCGGTGGG<br>(SEQ ID No. 347) | 54 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49C |
| Gluc 平铺指导物 11 | GCGGCCACGATGTTGAAGTCTTCGTTGTTC<br>(SEQ ID No. 348) | 62 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49C |
| Gluc 平铺指导物 12 | GTGGCCACGGCCACGATGTTGAAGTCTTCG<br>(SEQ ID No. 349) | 68 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49C |
| Gluc 平铺指导物 13 | GGTTGCTGGCCACGGCCACGATGTTGAAGT<br>(SEQ ID No. 350) | 73 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49C |
| Gluc 平铺指导物 14 | GTCGCGAAGTTGCTGGCCACGGCCACGATG<br>(SEQ ID No. 351) | 80 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49C |
| Gluc 平铺指导物 15 | GCCGTGGTCGCGAAGTTGCTGGCCACGGCC<br>(SEQ ID No. 352) | 86 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49C |
| Gluc 平铺指导物 16 | GCGAGATCCGTGGTCGCGAAGTTGCTGGCC<br>(SEQ ID No. 353) | 92 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49C |
| Gluc 平铺指导物 17 | GCAGCATCGAGATCCGTGGTCGCGAAGTTG | 98 | 注意，Cas13a 间隔子在 5'端 | 49C |

| | | | | |
|---|---|---|---|---|
| | (SEQ ID No. 354) | | 被截短两个核苷酸 | |
| Gluc 平铺指导物 18 | GGGTCAGCATCGAGATCCGTGGTCGCGAAG (SEQ ID No. 355) | 101 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49C |
| Gluc 平铺指导物 19 | GCTTCCCGCGGTCAGCATCGAGATCCGTGG (SEQ ID No. 356) | 109 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49C |
| Gluc 平铺指导物 20 | GGGGCAACTTCCCGCGGTCAGCATCGAGAT (SEQ ID No. 357) | 115 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49C |
| Gluc 平铺指导物 21 | GTCTTGCCGGGCAACTTCCCGCGGTCAGCA (SEQ ID No. 358) | 122 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49C |
| Gluc 平铺指导物 22 | GGCAGCTTCTTGCCGGGCAACTTCCCGCGG (SEQ ID No. 359) | 128 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49C |
| [2919] Gluc 平铺指导物 23 | GCCAGCGGCAGCTTCTTGCCGGGCAACTTC (SEQ ID No. 360) | 134 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49C |
| Gluc 平铺指导物 24 | GCACCTCCAGCGGCAGCTTCTTGCCGGGCA (SEQ ID No. 361) | 139 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49C |
| Gluc 平铺指导物 25 | GCTTTGAGCACCTCCAGCGGCAGCTTCTTG (SEQ ID No. 362) | 146 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49C |
| Gluc 平铺指导物 26 | GCATCTCTTTGAGCACCTCCAGCGGCAGCT (SEQ ID No. 363) | 151 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49C |
| Gluc 平铺指导物 27 | GTCCATCTCTTTGAGCACCTCCAGCGGCAG (SEQ ID No. 364) | 153 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49C |
| Gluc 平铺指导物 28 | GGGCATTGGCTTCCATCTCTTTGAGCACCT (SEQ ID No. 365) | 163 | 注意，Cas13a 间隔子在 5'端被截短两个 | 49C |

[2920]

| | | | | |
|---|---|---|---|---|
| | | | 核苷酸 | |
| Gluc 平铺指导物 29 | GTCCGGGCATTGGCTTCCATCTCTTTGAGC (SEQ ID No. 366) | 167 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49C |
| Gluc 平铺指导物 30 | GGCCAGCTTTCCGGGCATTGGCTTCCATCT (SEQ ID No. 367) | 175 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49C |
| Gluc 平铺指导物 31 | GGGTGCAGCCAGCTTTCCGGGCATTGGCTT (SEQ ID No. 368) | 181 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49C |
| Gluc 平铺指导物 32 | GAGCCCCTGGTGCAGCCAGCTTTCCGGGCA (SEQ ID No. 369) | 188 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49C |
| Gluc 平铺指导物 33 | GATCAGACAGCCCCTGGTGCAGCCAGCTTT (SEQ ID No. 370) | 195 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49C |
| Gluc 平铺指导物 34 | GGCAGATCAGACAGCCCCTGGTGCAGCCAG (SEQ ID No. 371) | 199 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49C |
| Gluc 平铺指导物 35 | GACAGGCAGATCAGACAGCCCCTGGTGCAG (SEQ ID No. 372) | 203 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49C |
| Gluc 平铺指导物 36 | GTGATGTGGGACAGGCAGATCAGACAGCCC (SEQ ID No. 373) | 212 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49C |
| Gluc 平铺指导物 37 | GACTTGATGTGGGACAGGCAGATCAGACAG (SEQ ID No. 374) | 215 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49C |
| Gluc 平铺指导物 38 | GGGGCGTGCACTTGATGTGGGACAGGCAGA (SEQ ID No. 375) | 223 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49C |
| Gluc 平铺指导物 39 | GCTTCATCTTGGGCGTGCACTTGATGTGGG (SEQ ID No. 376) | 232 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49C |

[2921]

| | | | | |
|---|---|---|---|---|
| Gluc 平铺指导物 40 | GTGAACTTCTTCATCTTGGGCG TGCACTTG (SEQ ID No. 377) | 239 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49C |
| Gluc 平铺指导物 41 | GGGATGAACTTCTTCATCTTGG GCGTGCAC (SEQ ID No. 378) | 242 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49C |
| Gluc 平铺指导物 42 | GTGGGATGAACTTCTTCATCTT GGGCGTGC (SEQ ID No. 379) | 244 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49C |
| Gluc 平铺指导物 43 | GGGCAGCGTCCTGGGATGAACT TCTTCATC (SEQ ID No. 380) | 254 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49C |
| Gluc 平铺指导物 44 | GGGTGTGGCAGCGTCCTGGGAT GAACTTCT (SEQ ID No. 381) | 259 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49C |
| Gluc 平铺指导物 45 | GTTCGTAGGTGTGGCAGCGTCC TGGGATGA (SEQ ID No. 382) | 265 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49C |
| Gluc 平铺指导物 46 | GCGCCTTCGTAGGTGTGGCAGC GTCCTGGG (SEQ ID No. 383) | 269 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49C |
| Gluc 平铺指导物 47 | GTCTTTGTCGCCTTCGTAGGTGT GGCAGCG (SEQ ID No. 384) | 276 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49C |
| Gluc 平铺指导物 48 | GCTTTGTCGCCTTCGTAGGTGT GGCAGCGT (SEQ ID No. 385) | 275 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49C |
| Gluc 平铺指导物 49 | GTGCCGCCCTGTGCGGACTCTT TGTCGCCT (SEQ ID No. 386) | 293 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49C |
| Gluc 平铺指导物 50 | GTATGCCGCCCTGTGCGGACTC TTTGTCGC (SEQ ID No. 387) | 295 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49C |
| Gluc 平铺 | GCCTCGCCTATGCCGCCCTGTG | 302 | 注意，Cas13a | 49C |

[2922]

| | | | | |
|---|---|---|---|---|
| 指导物 51 | CGGACTCT<br>(SEQ ID No. 388) | | 间隔子在 5'端被截短两个核苷酸 | |
| Gluc 平铺指导物 52 | GGATCGCCTCGCCTATGCCGCCCTGTGCGG<br>(SEQ ID No. 389) | 307 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49C |
| Gluc 平铺指导物 53 | GATGTCGACGATCGCCTCGCCTATGCCGCC<br>(SEQ ID No. 390) | 315 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49C |
| Gluc 平铺指导物 54 | GCAGGAATGTCGACGATCGCCTCGCCTATG<br>(SEQ ID No. 391) | 320 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49C |
| Gluc 平铺指导物 55 | GAATCTCAGGAATGTCGACGATCGCCTCGC<br>(SEQ ID No. 392) | 325 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49C |
| Gluc 平铺指导物 56 | GCCCAGGAATCTCAGGAATGTCGACGATCG<br>(SEQ ID No. 393) | 331 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49C |
| Gluc 平铺指导物 57 | GCCTTGAACCCAGGAATCTCAGGAATGTCG<br>(SEQ ID No. 394) | 338 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49C |
| Gluc 平铺指导物 58 | GCCAAGTCCTTGAACCCAGGAATCTCAGGA<br>(SEQ ID No. 395) | 344 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49C |
| Gluc 平铺指导物 59 | GTGGGCTCCAAGTCCTTGAACCCAGGAATC<br>(SEQ ID No. 396) | 350 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49C |
| Gluc 平铺指导物 60 | GCCATGGGCTCCAAGTCCTTGAACCCAGGA<br>(SEQ ID No. 397) | 353 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49C |
| Gluc 平铺指导物 61 | GGAACTGCTCCATGGGCTCCAAGTCCTTGA<br>(SEQ ID No. 398) | 361 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49C |
| Gluc 平铺指导物 62 | GTGCGATGAACTGCTCCATGGGCTCCAAGT | 367 | 注意，Cas13a 间隔子在 5'端 | 49C |

| | | | | |
|---|---|---|---|---|
| | (SEQ ID No. 399) | | 被截短两个核苷酸 | |
| Gluc 平铺指导物 63 | GGACCTGTGCGATGAACTGCTCCATGGGCT (SEQ ID No. 400) | 373 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49C |
| Gluc 平铺指导物 64 | GACAGATCGACCTGTGCGATGAACTGCTCC (SEQ ID No. 401) | 380 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49C |
| Gluc 平铺指导物 65 | GACACACAGATCGACCTGTGCGATGAACTG (SEQ ID No. 402) | 384 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49C |
| Gluc 平铺指导物 66 | GTGCAGTCCACACACAGATCGACCTGTGCG (SEQ ID No. 403) | 392 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49C |
| Gluc 平铺指导物 67 | GCCAGTTGTGCAGTCCACACACAGATCGAC (SEQ ID No. 404) | 399 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49C |
| Gluc 平铺指导物 68 | GGGCAGCCAGTTGTGCAGTCCACACACAGA (SEQ ID No. 405) | 404 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49C |
| Gluc 平铺指导物 69 | GTTTGAGGCAGCCAGTTGTGCAGTCCACAC (SEQ ID No. 406) | 409 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49C |
| Gluc 平铺指导物 70 | GAAGCCCTTTGAGGCAGCCAGTTGTGCAGT (SEQ ID No. 407) | 415 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49C |
| Gluc 平铺指导物 71 | GCACGTTGGCAAGCCCTTTGAGGCAGCCAG (SEQ ID No. 408) | 424 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49C |
| Gluc 平铺指导物 72 | GACTGCACGTTGGCAAGCCCTTTGAGGCAG (SEQ ID No. 409) | 428 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49C |
| Gluc 平铺指导物 73 | GGGTCAGAACACTGCACGTTGGCAAGCCCT (SEQ ID No. 410) | 437 | 注意，Cas13a 间隔子在 5'端被截短两个 | 49C |

[2923]

[2924]

| | | | | |
|---|---|---|---|---|
| | | | 核苷酸 | |
| Gluc 平铺指导物 74 | GCAGGTCAGAACACTGCACGTTGGCAAGCC (SEQ ID No. 411) | 439 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49C |
| Gluc 平铺指导物 75 | GAGCAGGTCAGAACACTGCACGTTGGCAAG (SEQ ID No. 412) | 441 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49C |
| Gluc 平铺指导物 76 | GGCCACTTCTTGAGCAGGTCAGAACACTGC (SEQ ID No. 413) | 452 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49C |
| Gluc 平铺指导物 77 | GCGGCAGCCACTTCTTGAGCAGGTCAGAAC (SEQ ID No. 414) | 457 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49C |
| Gluc 平铺指导物 78 | GTGCGGCAGCCACTTCTTGAGCAGGTCAGA (SEQ ID No. 415) | 459 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49C |
| Gluc 平铺指导物 79 | GAGCGTTGCGGCAGCCACTTCTTGAGCAGG (SEQ ID No. 416) | 464 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49C |
| Gluc 平铺指导物 80 | GAAAGGTCGCACAGCGTTGCGGCAGCCACT (SEQ ID No. 417) | 475 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49C |
| Gluc 平铺指导物 81 | GCTGGCAAAGGTCGCACAGCGTTGCGGCAG (SEQ ID No. 418) | 480 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49C |
| Gluc 平铺指导物 82 | GGGCAAAGGTCGCACAGCGTTGCGGCAGCC (SEQ ID No. 419) | 478 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49C |
| Gluc 平铺指导物 83 | GTGGATCTTGCTGGCAAAGGTCGCACAGCG (SEQ ID No. 420) | 489 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49C |
| Gluc 平铺指导物 84 | GCACCTGGCCCTGGATCTTGCTGGCAAAGG (SEQ ID No. 421) | 499 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49C |

[2925]

| | | | | |
|---|---|---|---|---|
| Gluc 平铺指导物 85 | GTGGCCCTGGATCTTGCTGGCAAAGGTCGC (SEQ ID No. 422) | 495 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49C |
| Gluc 平铺指导物 86 | GTGATCTTGTCCACCTGGCCCTGGATCTTG (SEQ ID No. 423) | 509 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49C |
| Gluc 平铺指导物 87 | GCCCCTTGATCTTGTCCACCTGGCCCTGGA (SEQ ID No. 424) | 514 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49C |
| Gluc 平铺指导物 88 | GCCCTTGATCTTGTCCACCTGGCCCTGGAT (SEQ ID No. 425) | 513 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49C |
| Gluc 平铺指导物 89 | GCCTTGATCTTGTCCACCTGGCCCTGGATC (SEQ ID No. 426) | 512 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49C |
| Gluc 平铺指导物 90 | GGCAAAGGTCGCACAGCGTTGCGGCAGCCA (SEQ ID No. 427) | 477 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49C |
| Gluc 平铺指导物 91 | GCAAAGGTCGCACAGCGTTGCGGCAGCCAC (SEQ ID No. 428) | 476 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49C |
| Gluc 平铺指导物 92 | GAAGGTCGCACAGCGTTGCGGCAGCCACTT (SEQ ID No. 429) | 474 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49C |
| Gluc 平铺指导物 93 | GAGGTCGCACAGCGTTGCGGCAGCCACTTC (SEQ ID No. 430) | 473 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49C |
| 非靶向指导物 1 | GGTAATGCCTGGCTTGTCGACGCATAGTCTG (SEQ ID No. 431) | 不适用 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49C |
| 非靶向指导物 2 | GGGAACCTTGGCCGTTATAAAGTCTGACCAG (SEQ ID No. 432) | 不适用 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49C |
| 非靶向指 | GGAGGGTGAGAATTTAGAACC | 不 | 注意，Cas13a | 49C |

[2926]

| | | | | |
|---|---|---|---|---|
| 导物 3 | AAGATTGTTG (SEQ ID No. 433) | 适用 | 间隔子在 5'端被截短两个核苷酸 | |

[2927] 表13:用于Cluc敲低的指导序列

| 名称 | 间隔序列 | 位置 | 注意 | 第一图 |
|---|---|---|---|---|
| Cluc 平铺指导物 1 | GAGTCCTGGCAATGAACAGTGGCGCAGTAG (SEQ ID No. 434) | 32 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49D |
| Cluc 平铺指导物 2 | GGGTGCCACAGCTGCTATCAATACATTCTC (SEQ ID No. 435) | 118 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49D |
| Cluc 平铺指导物 3 | GTTACATACTGACACATTCGGCAACATGTT (SEQ ID No. 436) | 197 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49D |
| Cluc 平铺指导物 4 | GTATGTACCAGGTTCCTGGAACTGGAATCT (SEQ ID No. 437) | 276 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49D |
| Cluc 平铺指导物 5 | GCCTTGGTTCCATCCAGGTTCTCCAGGGTG (SEQ ID No. 438) | 350 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49D |
| Cluc 平铺指导物 6 | GCAGTGATGGGATTCTCAGTAGCTTGAGCG (SEQ ID No. 439) | 431 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49D |
| Cluc 平铺指导物 7 | GAGCCTGGCATCTCAACAACAGCGATGGTG (SEQ ID No. 440) | 512 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49D |
| Cluc 平铺指导物 8 | GTGTCTGGGGCGATTCTTACAGATCTTCCT (SEQ ID No. 441) | 593 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49D |
| Cluc 平铺指导物 9 | GCTGGATCTGAAGTGAAGTCTGTATCTTCC (SEQ ID No. 442) | 671 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49D |
| Cluc 平铺指导 | GGCAACGTCATCAGGATTTCCATAGAGTGG (SEQ ID No. 443) | 747 | 注意，Cas13a 间隔子在 5'端 | 49D |

[2928]

| | | | | | |
|---|---|---|---|---|---|
| | 物 10 | | | 被截短两个核苷酸 | |
| | Cluc 平铺指导物 11 | GAGGCGCAGGAGATGGTGTAGTAGTAGAAG (SEQ ID No. 444) | 830 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49D |
| | Cluc 平铺指导物 13 | GAGGGACCCTGGAATTGGTATCTTGCTTTG (SEQ ID No. 445) | 986 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49D |
| | Cluc 平铺指导物 14 | GGTAAGAGTCAACATTCCTGTGTGAAACCT (SEQ ID No. 446) | 1066 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49D |
| | Cluc 平铺指导物 15 | GACCAGAATCTGTTTTCCATCAACAATGAG (SEQ ID No. 447) | 1143 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49D |
| | Cluc 平铺指导物 16 | GATGGCTGTAGTCAGTATGTCACCATCTTG (SEQ ID No. 448) | 1227 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49D |
| [2929] | Cluc 平铺指导物 17 | GTACCATCGAATGGATCTCTAATATGTACG (SEQ ID No. 449) | 1304 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49D |
| | Cluc 平铺指导物 18 | GAGATCACAGGCTCCTTCAGCATCAAAAGA (SEQ ID No. 450) | 1380 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49D |
| | Cluc 平铺指导物 19 | GCTTTGACCGGCGAAGAGACTATTGCAGAG (SEQ ID No. 451) | 1461 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49D |
| | Cluc 平铺指导物 20 | GCCCCTCAGGCAATACTCGTACATGCATCG (SEQ ID No. 452) | 1539 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49D |
| | Cluc 平铺指导物 21 | GCTGGTACTTCTAGGGTGTCTCCATGCTTT (SEQ ID No. 453) | 1619 | 注意，Cas13a 间隔子在 5'端被截短两个核苷酸 | 49D |
| | 非靶向指导物 1 | GGTAATGCCTGGCTTGTCGACGCATAGTCTG (SEQ ID No. 454) | 不适用 | 注意，Cas13a 间隔子在 5'端被截短两个核 | 49D |

[2930]

| | | | | |
|---|---|---|---|---|
| | | | 苷酸 | |
| 非靶向指导物2 | GGGAACCTTGGCCGTTATAAAGTCTGACCAG (SEQ ID No. 455) | 不适用 | 注意，Cas13a间隔子在5'端被截短两个核苷酸 | 49D |
| 非靶向指导物3 | GGAGGGTGAGAATTTAGAACCAAGATTGTTG (SEQ ID No. 456) | 不适用 | 注意，Cas13a间隔子在5'端被截短两个核苷酸 | 49D |

[2931] 表14:本研究中用于在哺乳动物细胞中进行RNA编辑的指导序列。错配碱基翻转被大写

[2932]

| 名称 | 间隔序列 | 注意 | 第一图 |
|---|---|---|---|
| 平铺30nt 30错配距离 | gCatcctgcggcctctactctgcattcaatt (SEQ ID No. 457) | 具有针对U6表达的5' G | 50C |
| 平铺30 nt 28错配距离 | gacCatcctgcggcctctactctgcattcaa (SEQ ID No. 458) | 具有针对U6表达的5' G | 50C |
| 平铺30 nt 26错配距离 | gaaacCatcctgcggcctctactctgcattc (SEQ ID No. 459) | 具有针对U6表达的5' G | 50C |
| 平铺30 nt 24错配距离 | gctaaacCatcctgcggcctctactctgcat (SEQ ID No. 460) | 具有针对U6表达的5' G | 50C |
| 平铺30 nt 22错配距离 | gttctaaacCatcctgcggcctctactctgc (SEQ ID No. 461) | 具有针对U6表达的5' G | 50C |
| 平铺30 nt 20错配距离 | gtgttctaaacCatcctgcggcctctactct (SEQ ID No. 462) | 具有针对U6表达的5' G | 50C |
| 平铺30 nt 18错配距离 | gaatgttctaaacCatcctgcggcctctact (SEQ ID No. 463) | 具有针对U6表达的5' G | 50C |
| 平铺30 nt 16错配距离 | gagaatgttctaaacCatcctgcggcctcta (SEQ ID No. 464) | 具有针对U6表达的5' G | 50C |
| 平铺30 nt 14错配距离 | gatagaatgttctaaacCatcctgcggcctc (SEQ ID No. 465) | 具有针对U6表达的5' G | 50C |
| 平铺30 nt 12错配距离 | gccatagaatgttctaaacCatcctgcggcc (SEQ ID No. 466) | 具有针对U6表达的5' G | 50C |
| 平铺30 nt 10错配距离 | gttccatagaatgttctaaacCatcctgcgg (SEQ ID No. 467) | 具有针对U6表达的5' G | 50C |
| 平铺30 nt 8错配距离 | gctttccatagaatgttctaaacCatcctgc (SEQ ID No. 468) | 具有针对U6表达的5' G | 50C |
| 平铺30 nt 6错配距离 | gctctttccatagaatgttctaaacCatcct (SEQ ID No. 469) | 具有针对U6表达的5' G | 50C |
| 平铺30 nt 4错配距离 | gatctctttccatagaatgttctaaacCatc (SEQ ID No. 470) | 具有针对U6表达的5' G | 50C |
| 平铺30 nt 2错配距离 | ggaatctctttccatagaatgttctaaacCa (SEQ ID No. 471) | 具有针对U6表达的5' G | 50C |
| 平铺50 nt 50错配距离 | gCatcctgcggcctctactctgcattcaattacatactgacacattcggca (SEQ ID | 具有针对U6表达的5' G | 50C |

[2933]

| | | | |
|---|---|---|---|
| | No. 472) | | |
| 平铺 50 nt 48 错配距离 | gacCatcctgcggcctctactctgcattcaattacatactgacacattcgg (SEQ ID No. 473) | 具有针对 U6 表达的 5' G | 50C |
| 平铺 50 nt 46 错配距离 | gaaacCatcctgcggcctctactctgcattcaattacatactgacacattc (SEQ ID No. 474) | 具有针对 U6 表达的 5' G | 50C |
| 平铺 50 nt 44 错配距离 | gctaaacCatcctgcggcctctactctgcattcaattacatactgacacat (SEQ ID No. 475) | 具有针对 U6 表达的 5' G | 50C |
| 平铺 50 nt 42 错配距离 | gttctaaacCatcctgcggcctctactctgcattcaattacatactgacac (SEQ ID No. 476) | 具有针对 U6 表达的 5' G | 50C |
| 平铺 50 nt 40 错配距离 | gtgttctaaacCatcctgcggcctctactctgcattcaattacatactgac (SEQ ID No. 477) | 具有针对 U6 表达的 5' G | 50C |
| 平铺 50 nt 38 错配距离 | gaatgttctaaacCatcctgcggcctctactctgcattcaattacatactg (SEQ ID No. 478) | 具有针对 U6 表达的 5' G | 50C |
| 平铺 50 nt 36 错配距离 | gagaatgttctaaacCatcctgcggcctctactctgcattcaattacatac (SEQ ID No. 479) | 具有针对 U6 表达的 5' G | 50C |
| 平铺 50 nt 34 错配距离 | gatagaatgttctaaacCatcctgcggcctctactctgcattcaattacat (SEQ ID No. 480) | 具有针对 U6 表达的 5' G | 50C |
| 平铺 50 nt 32 错配距离 | gccatagaatgttctaaacCatcctgcggcctctactctgcattcaattac (SEQ ID No. 481) | 具有针对 U6 表达的 5' G | 50C |
| 平铺 50 nt 30 错配距离 | gttccatagaatgttctaaacCatcctgcggcctctactctgcattcaatt (SEQ ID No. 482) | 具有针对 U6 表达的 5' G | 50C |
| 平铺 50 nt 28 错配距离 | gctttccatagaatgttctaaacCatcctgcggcctctactctgcattcaa (SEQ ID No. 483) | 具有针对 U6 表达的 5' G | 50C |
| 平铺 50 nt 26 错配距离 | gctctttccatagaatgttctaaacCatcctgcggcctctactctgcattc (SEQ ID No. 484) | 具有针对 U6 表达的 5' G | 50C |
| 平铺 50 nt 24 错配距离 | gatctctttccatagaatgttctaaacCatcctgcggcctctactctgcat (SEQ ID No. 485) | 具有针对 U6 表达的 5' G | 50C |
| 平铺 50 nt 22 错配距离 | ggaatctctttccatagaatgttctaaacCatcctgcggcctctactctgc (SEQ ID No. 486) | 具有针对 U6 表达的 5' G | 50C |
| 平铺 50 nt 20 错配距离 | gtggaatctctttccatagaatgttctaaacCatcctgcggcctctactct (SEQ ID No. 487) | 具有针对 U6 表达的 5' G | 50C |
| 平铺 50 nt 18 错配距离 | gactggaatctctttccatagaatgttctaaacCatcctgcggcctctact (SEQ ID No. 488) | 具有针对 U6 表达的 5' G | 50C |

| | | | | |
|---|---|---|---|---|
| | 平铺 50 nt 16 错配距离 | ggaactggaatctctttccatagaatgttctaaacCatcctgcggcctcta (SEQ ID No. 489) | 具有针对 U6 表达的 5' G | 50C |
| | 平铺 50 nt 14 错配距离 | gtggaactggaatctctttccatagaatgttctaaacCatcctgcggcct c(SEQ ID No. 490) | 具有针对 U6 表达的 5' G | 50C |
| | 平铺 50 nt 12 错配距离 | gcctggaactggaatctctttccatagaatgttctaaacCatcctgcggcc (SEQ ID No. 491) | 具有针对 U6 表达的 5' G | 50C |
| | 平铺 50 nt 10 错配距离 | gttcctggaactggaatctctttccatagaatgttctaaacCatcctgcgg (SEQ ID No. 492) | 具有针对 U6 表达的 5' G | 50C |
| | 平铺 50 nt 8 错配距离 | gggttcctggaactggaatctctttccatagaatgttctaaacCatcctgc (SEQ ID No. 493) | 具有针对 U6 表达的 5' G | 50C |
| | 平铺 50 nt 6 错配距离 | gcaggttcctggaactggaatctctttccatagaatgttctaaacCatcct (SEQ ID No. 494) | 具有针对 U6 表达的 5' G | 50C |
| | 平铺 50 nt 4 错配距离 | gaccaggttcctggaactggaatctctttccatagaatgttctaaacCatc (SEQ ID No. 495) | 具有针对 U6 表达的 5' G | 50C |
| | 平铺 50 nt 2 错配距离 | ggtaccaggttcctggaactggaatctctttccatagaatgttctaaacCa (SEQ ID No. 496) | 具有针对 U6 表达的 5' G | 50C |
| [2934] | 平铺 70 nt 70 错配距离 | gCatcctgcggcctctactctgcattcaattacatactgacacattcggcaacatgtttttcctggtttat (SEQ ID No. 497) | 具有针对 U6 表达的 5' G | 50C |
| | 平铺 70 nt 68 错配距离 | gacCatcctgcggcctctactctgcattcaattacatactgacacattcggcaacatgtttttcctggttt (SEQ ID No. 498) | 具有针对 U6 表达的 5' G | 50C |
| | 平铺 70 nt 66 错配距离 | gaaacCatcctgcggcctctactctgcattcaattacatactgacacattcggcaacatgttttcctggt (SEQ ID No. 499) | 具有针对 U6 表达的 5' G | 50C |
| | 平铺 70 nt 64 错配距离 | gctaaacCatcctgcggcctctactctgcattcaattacatactgacacattcggcaacatgtttttcctg (SEQ ID No. 500) | 具有针对 U6 表达的 5' G | 50C |
| | 平铺 70 nt 62 错配距离 | gttctaaacCatcctgcggcctctactctgcattcaattacatactgacacattcggcaacatgttttcc (SEQ ID No. 501) | 具有针对 U6 表达的 5' G | 50C |
| | 平铺 70 nt 60 错配距离 | gtgttctaaacCatcctgcggcctctactctgcattcaattacatactgacacattcggcaacatgttttt (SEQ ID No. 502) | 具有针对 U6 表达的 5' G | 50C |
| | 平铺 70 nt 58 错配距离 | gaatgttctaaacCatcctgcggcctctactctgcattcaattacatactgacacattcggcaacatgttt (SEQ ID No. 503) | 具有针对 U6 表达的 5' G | 50C |
| | 平铺 70 nt 56 错配距离 | gagaatgttctaaacCatcctgcggcctctactctgcattcaattacatactgacacattcggcaacatgt (SEQ ID No. 504) | 具有针对 U6 表达的 5' G | 50C |
| | 平铺 70 nt 54 | gatagaatgttctaaacCatcctgcggcctct | 具有针对 U6 表达 | 50C |

[2935]

| | | | |
|---|---|---|---|
| 错配距离 | actctgcattcaattacatactgacacattcgg caacat (SEQ ID No. 505) | 的 5' G | |
| 平铺 70 nt 52 错配距离 | gccatagaatgttctaaacCatcctgcggcc tctactctgcattcaattacatactgacacattc ggcaac (SEQ ID No. 506) | 具有针对 U6 表达的 5' G | 50C |
| 平铺 70 nt 50 错配距离 | gttccatagaatgttctaaacCatcctgcggc ctctactctgcattcaattacatactgacacatt cggca (SEQ ID No. 507) | 具有针对 U6 表达的 5' G | 50C |
| 平铺 70 nt 48 错配距离 | gctttccatagaatgttctaaacCatcctgcg gcctctactctgcattcaattacatactgacac attcgg (SEQ ID No. 508) | 具有针对 U6 表达的 5' G | 50C |
| 平铺 70 nt 46 错配距离 | gctctttccatagaatgttctaaacCatcctgc ggcctctactctgcattcaattacatactgaca cattc (SEQ ID No. 509) | 具有针对 U6 表达的 5' G | 50C |
| 平铺 70 nt 44 错配距离 | gatctctttccatagaatgttctaaacCatcct gcggcctctactctgcattcaattacatactga cacat (SEQ ID No. 510) | 具有针对 U6 表达的 5' G | 50C |
| 平铺 70 nt 42 错配距离 | ggaatctctttccatagaatgttctaaacCatc ctgcggcctctactctgcattcaattacatact gacac (SEQ ID No. 511) | 具有针对 U6 表达的 5' G | 50C |
| 平铺 70 nt 40 错配距离 | gtggaatctctttccatagaatgttctaaacC atcctgcggcctctactctgcattcaattacat actgac (SEQ ID No. 512) | 具有针对 U6 表达的 5' G | 50C |
| 平铺 70 nt 38 错配距离 | gactggaatctctttccatagaatgttctaaac Catcctgcggcctctactctgcattcaattac atactg (SEQ ID No. 513) | 具有针对 U6 表达的 5' G | 50C |
| 平铺 70 nt 36 错配距离 | ggaactggaatctctttccatagaatgttctaa acCatcctgcggcctctactctgcattcaatt acatac (SEQ ID No. 514) | 具有针对 U6 表达的 5' G | 50C |
| 平铺 70 nt 34 错配距离 | gtggaactggaatctctttccatagaatgttct aaacCatcctgcggcctctactctgcattca attacat (SEQ ID No. 515) | 具有针对 U6 表达的 5' G | 50C |
| 平铺 70 nt 32 错配距离 | gcctggaactggaatctctttccatagaatgtt ctaaacCatcctgcggcctctactctgcattc aattac (SEQ ID No. 516) | 具有针对 U6 表达的 5' G | 50C |
| 平铺 70 nt 30 错配距离 | gttcctggaactggaatctctttccatagaatg ttctaaacCatcctgcggcctctactctgcatt caatt (SEQ ID No. 517) | 具有针对 U6 表达的 5' G | 50C |
| 平铺 70 nt 28 错配距离 | gggttcctggaactggaatctctttccataga atgttctaaacCatcctgcggcctctactctg cattcaa (SEQ ID No. 518) | 具有针对 U6 表达的 5' G | 50C |
| 平铺 70 nt 26 错配距离 | gcaggttcctggaactggaatctctttccata gaatgttctaaacCatcctgcggcctctactc tgcattc (SEQ ID No. 519) | 具有针对 U6 表达的 5' G | 50C |
| 平铺 70 nt 24 错配距离 | gaccaggttcctggaactggaatctctttcca tagaatgttctaaacCatcctgcggcctctac tctgcat (SEQ ID No. 520) | 具有针对 U6 表达的 5' G | 50C |
| 平铺 70 nt 22 错配距离 | ggtaccaggttcctggaactggaatctctttc catagaatgttctaaacCatcctgcggcctct | 具有针对 U6 表达的 5' G | 50C |

[2936]

| | | | |
|---|---|---|---|
| | actctgc (SEQ ID No. 521) | | |
| 平铺 70 nt 20 错配距离 | gatgtaccaggttcctggaactggaatctctttccatagaatgttctaaacCatcctgcggcctctactct (SEQ ID No. 522) | 具有针对 U6 表达的 5' G | 50C |
| 平铺 70 nt 18 错配距离 | ggtatgtaccaggttcctggaactggaatctctttccatagaatgttctaaacCatcctgcggcctctact (SEQ ID No. 523) | 具有针对 U6 表达的 5' G | 50C |
| 平铺 70 nt 16 错配距离 | gacgtatgtaccaggttcctggaactggaatctctttccatagaatgttctaaacCatcctgcggcctcta (SEQ ID No. 524) | 具有针对 U6 表达的 5' G | 50C |
| 平铺 70 nt 14 错配距离 | gacacgtatgtaccaggttcctggaactggaatctctttccatagaatgttctaaacCatcctgcggcctc (SEQ ID No. 525) | 具有针对 U6 表达的 5' G | 50C |
| 平铺 70 nt 12 错配距离 | gcaacacgtatgtaccaggttcctggaactggaatctctttccatagaatgttctaaacCatcctgcggcc (SEQ ID No. 526) | 具有针对 U6 表达的 5' G | 50C |
| 平铺 70 nt 10 错配距离 | gcccaacacgtatgtaccaggttcctggaactggaatctctttccatagaatgttctaaacCatcctgcgg (SEQ ID No. 527) | 具有针对 U6 表达的 5' G | 50C |
| 平铺 70 nt 8 错配距离 | ggacccaacacgtatgtaccaggttcctggaactggaatctctttccatagaatgttctaaacCatcctgc (SEQ ID No. 528) | 具有针对 U6 表达的 5' G | 50C |
| 平铺 70 nt 6 错配距离 | gttgacccaacacgtatgtaccaggttcctggaactggaatctctttccatagaatgttctaaacCatcct (SEQ ID No. 529) | 具有针对 U6 表达的 5' G | 50C |
| 平铺 70 nt 4 错配距离 | gccttgacccaacacgtatgtaccaggttcctggaactggaatctctttccatagaatgttctaaacCatc (SEQ ID No. 530) | 具有针对 U6 表达的 5' G | 50C |
| 平铺 70 nt 2 错配距离 | gttccttgacccaacacgtatgtaccaggttcctggaactggaatctctttccatagaatgttctaaacCa (SEQ ID No. 531) | 具有针对 U6 表达的 5' G | 50C |
| 平铺 84 nt 84 错配距离 | gCatcctgcggcctctactctgcattcaattacatactgacacattcggcaacatgtttttcctggtttattttcacacagtcca (SEQ ID No. 532) | 具有针对 U6 表达的 5' G | 50C |
| 平铺 84 nt 82 错配距离 | gacCatcctgcggcctctactctgcattcaattacatactgacacattcggcaacatgtttttcctggtttattttcacacagtc (SEQ ID No. 533) | 具有针对 U6 表达的 5' G | 50C |
| 平铺 84 nt 80 错配距离 | gaaacCatcctgcggcctctactctgcattcaattacatactgacacattcggcaacatgtttttcctggtttattttcacacag (SEQ ID No. 534) | 具有针对 U6 表达的 5' G | 50C |
| 平铺 84 nt 78 错配距离 | gctaaacCatcctgcggcctctactctgcattcaattacatactgacacattcggcaacatgtttttcctggtttattttcacac (SEQ ID No. 535) | 具有针对 U6 表达的 5' G | 50C |
| 平铺 84 nt 76 错配距离 | gttctaaacCatcctgcggcctctactctgcattcaattacatactgacacattcggcaacatgt | 具有针对 U6 表达的 5' G | 50C |

[2937]

| | | | |
|---|---|---|---|
| | ttttcctggtttattttcac (SEQ ID No. 536) | | |
| 平铺 84 nt 74 错配距离 | gtgttctaaacCatcctgcggcctctactctg cattcaattacatactgacacattcggcaaca tgtttttcctggtttattttc (SEQ ID No. 537) | 具有针对 U6 表达的 5' G | 50C |
| 平铺 84 nt 72 错配距离 | gaatgttctaaacCatcctgcggcctctactc tgcattcaattacatactgacacattcggcaa catgtttttcctggtttattt (SEQ ID No. 538) | 具有针对 U6 表达的 5' G | 50C |
| 平铺 84 nt 70 错配距离 | gagaatgttctaaacCatcctgcggcctcta ctctgcattcaattacatactgacacattcggc aacatgtttttcctggtttat (SEQ ID No. 539) | 具有针对 U6 表达的 5' G | 50C |
| 平铺 84 nt 68 错配距离 | gatagaatgttctaaacCatcctgcggcctct actctgcattcaattacatactgacacattcgg caacatgtttttcctggttt (SEQ ID No. 540) | 具有针对 U6 表达的 5' G | 50C |
| 平铺 84 nt 66 错配距离 | gccatagaatgttctaaacCatcctgcggcc tctactctgcattcaattacatactgacacattc ggcaacatgtttttcctggt (SEQ ID No. 541) | 具有针对 U6 表达的 5' G | 50C |
| 平铺 84 nt 64 错配距离 | gttccatagaatgttctaaacCatcctgcggc ctctactctgcattcaattacatactgacacatt cggcaacatgtttttcctg (SEQ ID No. 542) | 具有针对 U6 表达的 5' G | 50C |
| 平铺 84 nt 62 错配距离 | gctttccatagaatgttctaaacCatcctgcg gcctctactctgcattcaattacatactgacac attcggcaacatgtttttcc (SEQ ID No. 543) | 具有针对 U6 表达的 5' G | 50C |
| 平铺 84 nt 60 错配距离 | gctctttccatagaatgttctaaacCatcctgc ggcctctactctgcattcaattacatactgaca cattcggcaacatgttttt (SEQ ID No. 544) | 具有针对 U6 表达的 5' G | 50C |
| 平铺 84 nt 58 错配距离 | gatctctttccatagaatgttctaaacCatcct gcggcctctactctgcattcaattacatactga cacattcggcaacatgttt (SEQ ID No. 545) | 具有针对 U6 表达的 5' G | 50C |
| 平铺 84 nt 56 错配距离 | ggaatctctttccatagaatgttctaaacCatc ctgcggcctctactctgcattcaattacatact gacacattcggcaacatgt (SEQ ID No. 546) | 具有针对 U6 表达的 5' G | 50C |
| 平铺 84 nt 54 错配距离 | gtggaatctctttccatagaatgttctaaacC atcctgcggcctctactctgcattcaattacat actgacacattcggcaacat (SEQ ID No. 547) | 具有针对 U6 表达的 5' G | 50C |
| 平铺 84 nt 52 错配距离 | gactggaatctctttccatagaatgttctaaac Catcctgcggcctctactctgcattcaattac atactgacacattcggcaac (SEQ ID | 具有针对 U6 表达的 5' G | 50C |

[2938]

| | | | |
|---|---|---|---|
| | No. 548) | | |
| 平铺 84 nt 50 错配距离 | ggaactggaatctctttccatagaatgttctaaacCatcctgcggcctctactctgcattcaattacatactgacacattcggca (SEQ ID No. 549) | 具有针对 U6 表达的 5' G | 50C |
| 平铺 84 nt 48 错配距离 | gtggaactggaatctctttccatagaatgttctaaacCatcctgcggcctctactctgcattcaattacatactgacacattcgg (SEQ ID No. 550) | 具有针对 U6 表达的 5' G | 50C |
| 平铺 84 nt 46 错配距离 | gcctggaactggaatctctttccatagaatgttctaaacCatcctgcggcctctactctgcattcaattacatactgacacattc (SEQ ID No. 551) | 具有针对 U6 表达的 5' G | 50C |
| 平铺 84 nt 44 错配距离 | gttcctggaactggaatctctttccatagaatgttctaaacCatcctgcggcctctactctgcattcaattacatactgacacat (SEQ ID No. 552) | 具有针对 U6 表达的 5' G | 50C |
| 平铺 84 nt 42 错配距离 | gggttcctggaactggaatctctttccatagaatgttctaaacCatcctgcggcctctactctgcattcaattacatactgacac (SEQ ID No. 553) | 具有针对 U6 表达的 5' G | 50C |
| 平铺 84 nt 40 错配距离 | gcaggttcctggaactggaatctctttccatagaatgttctaaacCatcctgcggcctctactctgcattcaattacatactgac (SEQ ID No. 554) | 具有针对 U6 表达的 5' G | 50C |
| 平铺 84 nt 38 错配距离 | gaccaggttcctggaactggaatctctttccatagaatgttctaaacCatcctgcggcctctactctgcattcaattacatactg (SEQ ID No. 555) | 具有针对 U6 表达的 5' G | 50C |
| 平铺 84 nt 36 错配距离 | ggtaccaggttcctggaactggaatctctttccatagaatgttctaaacCatcctgcggcctctactctgcattcaattacatac (SEQ ID No. 556) | 具有针对 U6 表达的 5' G | 50C |
| 平铺 84 nt 34 错配距离 | gatgtaccaggttcctggaactggaatctctttccatagaatgttctaaacCatcctgcggcctctactctgcattcaattacat (SEQ ID No. 557) | 具有针对 U6 表达的 5' G | 50C |
| 平铺 84 nt 32 错配距离 | ggtatgtaccaggttcctggaactggaatctctttccatagaatgttctaaacCatcctgcggcctctactctgcattcaattac (SEQ ID No. 558) | 具有针对 U6 表达的 5' G | 50C |
| 平铺 84 nt 30 错配距离 | gacgtatgtaccaggttcctggaactggaatctctttccatagaatgttctaaacCatcctgcggcctctactctgcattcaatt (SEQ ID No. 559) | 具有针对 U6 表达的 5' G | 50C |
| 平铺 84 nt 28 错配距离 | gacacgtatgtaccaggttcctggaactggaatctctttccatagaatgttctaaacCatcctgcggcctctactctgcattcaa (SEQ ID No. 560) | 具有针对 U6 表达的 5' G | 50C |

| | | | |
|---|---|---|---|
| 平铺 84 nt 26 错配距离 | gcaacacgtatgtaccaggttcctggaactg gaatctctttccatagaatgttctaaacCatcc tgcggcctctactctgcattc (SEQ ID No. 561) | 具有针对 U6 表达的 5' G | 50C |
| 平铺 84 nt 24 错配距离 | gcccaacacgtatgtaccaggttcctggaac tggaatctctttccatagaatgttctaaacCat cctgcggcctctactctgcat (SEQ ID No. 562) | 具有针对 U6 表达的 5' G | 50C |
| 平铺 84 nt 22 错配距离 | ggacccaacacgtatgtaccaggttcctgga actggaatctctttccatagaatgttctaaacC atcctgcggcctctactctgc (SEQ ID No. 563) | 具有针对 U6 表达的 5' G | 50C |
| 平铺 84 nt 20 错配距离 | gttgacccaacacgtatgtaccaggttcctgg aactggaatctctttccatagaatgttctaaac Catcctgcggcctctactct (SEQ ID No. 564) | 具有针对 U6 表达的 5' G | 50C |
| 平铺 84 nt 18 错配距离 | gccttgacccaacacgtatgtaccaggttcct ggaactggaatctctttccatagaatgttctaa acCatcctgcggcctctact (SEQ ID No. 565) | 具有针对 U6 表达的 5' G | 50C |
| 平铺 84 nt 16 错配距离 | gttccttgacccaacacgtatgtaccaggttc ctggaactggaatctctttccatagaatgttct aaacCatcctgcggcctcta (SEQ ID No. 566) | 具有针对 U6 表达的 5' G | 50C |
| 平铺 84 nt 14 错配距离 | gggttccttgacccaacacgtatgtaccaggt tcctggaactggaatctctttccatagaatgtt ctaaacCatcctgcggcctc (SEQ ID No. 567) | 具有针对 U6 表达的 5' G | 50C |
| 平铺 84 nt 12 错配距离 | gttggttccttgacccaacacgtatgtaccag gttcctggaactggaatctctttccatagaatg ttctaaacCatcctgcggcc (SEQ ID No. 568) | 具有针对 U6 表达的 5' G | 50C |
| 平铺 84 nt 10 错配距离 | gccttggttccttgacccaacacgtatgtacc aggttcctggaactggaatctctttccataga atgttctaaacCatcctgcgg (SEQ ID No. 569) | 具有针对 U6 表达的 5' G | 50C |
| 平铺 84 nt 8 错配距离 | ggcccttggttccttgacccaacacgtatgta ccaggttcctggaactggaatctctttccata gaatgttctaaacCatcctgc (SEQ ID No. 570) | 具有针对 U6 表达的 5' G | 50C |
| 平铺 84 nt 6 错配距离 | gccgcccttggttccttgacccaacacgtatg taccaggttcctggaactggaatctctttccat agaatgttctaaacCatcct (SEQ ID No. 571) | 具有针对 U6 表达的 5' G | 50C |
| 平铺 84 nt 4 错配距离 | gcgccgcccttggttccttgacccaacacgt atgtaccaggttcctggaactggaatctctttc catagaatgttctaaacCatc (SEQ ID No. 572) | 具有针对 U6 表达的 5' G | 50C |
| 平铺 84 nt 2 | ggtcgccgcccttggttccttgacccaacac | 具有针对 U6 表达 | 50C |

[2939]

[2940]

| | | | |
|---|---|---|---|
| 错配距离 | gtatgtaccaggttcctggaactggaatctcttttccatagaatgttctaaacCa (SEQ ID No. 573) | 的 5' G | |
| ADAR 非靶向指导物 | GTAATGCCTGGCTTGTCGACGCATAGTCTG (SEQ ID No. 574) | 具有针对 U6 表达的 5' G | 50C |
| TAG 基序的 PFS 结合筛选指导物 | gaaaacgcaggttcctcCagtttcgggagcagcgcacgtctccctgtagtc (SEQ ID No. 575) | 具有针对 U6 表达的 5' G | 51B |
| AAC 基序的 PFS 结合筛选指导物 | gacgcaggttcctctagCttcgggagcagcgcacgtctccctgtagtcaag (SEQ ID No. 576) | 具有针对 U6 表达的 5' G | 51B |
| PFS 结合筛选非靶向指导物 | GTAATGCCTGGCTTGTCGACGCATAGTCTG (SEQ ID No. 577) | 具有针对 U6 表达的 5' G | 51B |
| 基序偏好靶向指导物 | gatagaatgttctaaacCatcctgcggcctctactctgcattcaattacat (SEQ ID No. 578) | 具有针对 U6 表达的 5' G | 51C |
| 基序偏好非靶向指导物 | GTAATGCCTGGCTTGTCGACGCATAGTCTG (SEQ ID No. 579) | 具有针对 U6 表达的 5' G | 51C |
| PPIB 平铺指导物 50 错配距离 | gCaaggccacaaaattatccactgtttttggaacagtctttccgaagagac (SEQ ID No. 580) | 具有针对 U6 表达的 5' G | 57D |
| PPIB 平铺指导物 42 错配距离 | gcctgtagcCaaggccacaaaattatccactgtttttggaacagtctttcc (SEQ ID No. 581) | 具有针对 U6 表达的 5' G | 57D |
| PPIB 平铺指导物 34 错配距离 | gctttctctcctgtagcCaaggccacaaaattatccactgtttttggaaca (SEQ ID No. 582) | 具有针对 U6 表达的 5' G | 57D |
| PPIB 平铺指导物 26 错配距离 | ggccaaatcctttctctcctgtagcCaaggccacaaaattatccactgttt (SEQ ID No. 583) | 具有针对 U6 表达的 5' G | 57D |
| PPIB 平铺指导物 18 错配距离 | gtttttgtagccaaatcctttctctcctgtagcCaaggccacaaaattatc (SEQ ID No. 584) | 具有针对 U6 表达的 5' G | 57D |
| PPIB 平铺指导物 10 错配距离 | gatttgctgtttttgtagccaaatcctttctctcctgtagcCaaggccaca (SEQ ID No. 585) | 具有针对 U6 表达的 5' G | 57D |
| PPIB 平铺指导物 2 错配距离 | gacgatggaatttgctgtttttgtagccaaatcctttctctcctgtagcCa (SEQ ID No. 586) | 具有针对 U6 表达的 5' G | 57D |
| 靶向指导物，与碱基 G 相对 | gatagaatgttctaaacGatcctgcggcctctactctgcattcaattacat (SEQ ID No. 587) | 具有针对 U6 表达的 5' G | 57D |

[2941]

| | | | |
|---|---|---|---|
| 靶向指导物，与碱基 A 相对 | gatagaatgttctaaacAatcctgcggcctctactctgcattcaattacat (SEQ ID No. 588) | 具有针对 U6 表达的 5' G | 57D |
| 靶向指导物，与碱基 C 相对 | gatagaatgttctaaacTatcctgcggcctctactctgcattcaattacat (SEQ ID No. 589) | 具有针对 U6 表达的 5' G | 57D |
| AVPR2 指导物 37 错配距离 | ggtcccacgcggccCacagctgcaccaggaagaagggtgcccagcacagca (SEQ ID No. 590) | 具有针对 U6 表达的 5' G | 52A |
| AVPR2 指导物 35 错配距离 | ggggtcccacgcggccCacagctgcaccaggaagaagggtgcccagcacag (SEQ ID No. 591) | 具有针对 U6 表达的 5' G | 52A |
| AVPR2 指导物 33 错配距离 | gccgggtcccacgcggccCacagctgcaccaggaagaagggtgcccagcac (SEQ ID No. 592) | 具有针对 U6 表达的 5' G | 52A |
| FANCC 指导物 37 错配距离 | gggtgatgacatccCaggcgatcgtgtggcctccaggagcccagagcagga (SEQ ID No. 593) | 具有针对 U6 表达的 5' G | 52B |
| FANCC 指导物 35 错配距离 | gagggtgatgacatccCaggcgatcgtgtggcctccaggagcccagagcag (SEQ ID No. 594) | 具有针对 U6 表达的 5' G | 52B |
| FANCC 指导物 32 错配距离 | gatcagggtgatgacatccCaggcgatcgtgtggcctccaggagcccagag (SEQ ID No. 595) | 具有针对 U6 表达的 5' G | 52B |
| 合成疾病基因靶标 IL2RG | ggtggctccattcactcCaatgctgagcactccacagagtgggttaaagc (SEQ ID No. 596) | 具有针对 U6 表达的 5' G | 52E |
| 合成疾病基因靶标 F8 | gtttctaatatattttgCcagactgatggactattctcaattaataatgat (SEQ ID No. 597) | 具有针对 U6 表达的 5' G | 52E |
| 合成疾病基因靶标 LDLR | gagatgttgctgtggatCcagtccacagccagcccgtcgggggcctggatg (SEQ ID No. 598) | 具有针对 U6 表达的 5' G | 52E |
| 合成疾病基因靶标 CBS | gcaggccggcccagctgCcaggtgcacctgctcggagcatcgggccggatc (SEQ ID No. 599) | 具有针对 U6 表达的 5' G | 52E |
| 合成疾病基因靶标 HBB | gcaaagaacctctgggtCcaagggtagaccaccagcagcctgcccagggcc (SEQ ID No. 600) | 具有针对 U6 表达的 5' G | 52E |
| 合成疾病基因靶标 ALDOB | gaagagaaacttagtttCcagggctttggtagagggcaaaggttgatagca (SEQ ID No. 601) | 具有针对 U6 表达的 5' G | 52E |
| 合成疾病基因靶标 DMD | gtcagcctagtgcagagCcactggtagttggtggttagagtttcaagttcc (SEQ ID No. 602) | 具有针对 U6 表达的 5' G | 52E |
| 合成病基因 | ggctcattgtgaacaggCcagtaatgtccg | 具有针对 U6 表达 | 52E |

[2942]

| | | | |
|---|---|---|---|
| 靶标 SMAD4 | ggatggggcggcataggcggg (SEQ ID No. 603) | 的 5' G | |
| 合成病基因靶标 BRCA2 | gtagctaaagaacttgaCcaagacatatcaggatccacctcagctcctaga (SEQ ID No. 604) | 具有针对 U6 表达的 5' G | 52E |
| 合成病基因靶标 GRIN2A | ggggcattgttctgtgcCcagtcctgctggtagacctgctccccggtggct (SEQ ID No. 605) | 具有针对 U6 表达的 5' G | 52E |
| 合成疾病基因靶标 SCN9A | gagaagtcgttcatgtgCcaccgtgggagcgtacagtcatcattgatcttg (SEQ ID No. 606) | 具有针对 U6 表达的 5' G | 52E |
| 合成疾病基因靶标 TARDBP | gggattaatgctgaacgCaccaaagttcatcccaccacccatattactacc (SEQ ID No. 607) | 具有针对 U6 表达的 5' G | 52E |
| 合成疾病基因靶标 CFTR | gctccaaaggctttcctCcactgttgcaaagttattgaatcccaagacaca (SEQ ID No. 608) | 具有针对 U6 表达的 5' G | 52E |
| 合成疾病基因靶标 UBE3A | gatgaatgaacgatttcCcagaactccctaatcagaacagagtccctggta (SEQ ID No. 609) | 具有针对 U6 表达的 5' G | 52E |
| 合成疾病基因靶标 SMPD1 | ggagcctctgccggagcCcagagaacccgagagtcagacagagccagcgcc (SEQ ID No. 610) | 具有针对 U6 表达的 5' G | 52E |
| 合成疾病基因靶标 USH2A | ggcttccgtggagacacCcaatcaatttgaagagatcttgaagtgatgcca (SEQ ID No. 611) | 具有针对 U6 表达的 5' G | 52E |
| 合成疾病基因靶标 MEN1 | gtgggactgccctcctcCcatttgcagatgccgtcgtagaatcgcagcagg (SEQ ID No. 612) | 具有针对 U6 表达的 5' G | 52E |
| 合成疾病基因靶标 C8orf37 | gcttcttcaatagttctCcagctacactggcaggcatatgcccgtgttcct (SEQ ID No. 613) | 具有针对 U6 表达的 5' G | 52E |
| 合成疾病基因靶标 MLH1 | gattccttttcttcgtcCcaattcacctcagtggctagtcgaagaatgaag (SEQ ID No. 614) | 具有针对 U6 表达的 5' G | 52E |
| 合成疾病基因靶标 TSC2 | gcagcttcagcaccttcCagtcagactcctgcttcaagcactgcagcagga (SEQ ID No. 615) | 具有针对 U6 表达的 5' G | 52E |
| 合成疾病基因靶标 NF1 | gccatttgcttgcagtgCcactccagaggattccggattgccataaatact (SEQ ID No. 616) | 具有针对 U6 表达的 5' G | 52E |
| 合成疾病基因靶标 MSH6 | gttcaatagttttggtcCagtatcgtttacagcccttcttggtagatttca (SEQ ID No. 617) | 具有针对 U6 表达的 5' G | 52E |
| 合成疾病基因靶标 | ggcaaccgtcttctgacCaaatggcagaacatttgtccccaactttccact (SEQ ID | 具有针对 U6 表达的 5' G | 52E |

[2943]

| | | | |
|---|---|---|---|
| SMN1 | No. 618) | | |
| 合成疾病基因靶标 SH3TC2 | gcgactttccaatgaacCactgaagcccag gtatgacaaagccgatgatct (SEQ ID No. 619) | 具有针对 U6 表达的 5' G | 52E |
| 合成疾病基因靶标 DNAH5 | gtttacactcatgcttcCacagctttaacagat catttggttccttgatga (SEQ ID No. 620) | 具有针对 U6 表达的 5' G | 52E |
| 合成疾病基因靶标 MECP2 | gcttaagcttccgtgtcCagccttcaggcag ggtggggtcatcatacatgg (SEQ ID No. 621) | 具有针对 U6 表达的 5' G | 52E |
| 合成疾病基因靶标 ADGRV1 | ggacagctgggctgatcCatgatgtcatcca gaaacactggggaccctcag (SEQ ID No. 622) | 具有针对 U6 表达的 5' G | 52E |
| 合成疾病基因靶标 AHI1 | gtctcatctcaactttcCatatccgtatcatgg aatcatagcatcctgtaa (SEQ ID No. 623) | 具有针对 U6 表达的 5' G | 52E |
| 合成疾病基因靶标 PRKN | gcatgcagacgcggttcCactcgcagccac agttccagcaccactcgagcc (SEQ ID No. 624) | 具有针对 U6 表达的 5' G | 52E |
| 合成疾病基因靶标 COL3A1 | gttggttagggtcaaccCagtattctccactc ttgagttcaggatggcaga (SEQ ID No. 625) | 具有针对 U6 表达的 5' G | 52E |
| 合成疾病基因靶标 BRCA1 | gctacactgtccaacacCcactctcgggtca ccacaggtgcctcacacatc (SEQ ID No. 626) | 具有针对 U6 表达的 5' G | 52E |
| 合成疾病基因靶标 MYBPC3 | gctgcactgtgtaccccCagagctccgtgtt gccgacatcctggggtggct (SEQ ID No. 627) | 具有针对 U6 表达的 5' G | 52E |
| 合成疾病基因靶标 APC | gagcttcctgccactccCaacaggtttcaca gtaagcgcgtatctgttcca (SEQ ID No. 628) | 具有针对 U6 表达的 5' G | 52E |
| 合成疾病基因靶标 BMPR2 | gacggcaagagcttaccCagtcacttgtgtg gagacttaaatacttgcata (SEQ ID No. 629) | 具有针对 U6 表达的 5' G | 52E |
| KRAS 平铺指导物 50 错配距离 | gCaaggccacaaaattatccactgtttttgga acagtctttccgaagagac (SEQ ID No. 630) | 具有针对 U6 表达的 5' G | 53A |
| KRAS 平铺指导物 42 错配距离 | gcctgtagcCaaggccacaaaattatccact gtttttggaacagtctttcc (SEQ ID No. 631) | 具有针对 U6 表达的 5' G | 53A |
| KRAS 平铺指导物 34 错配距离 | gctttctctcctgtagcCaaggccacaaaatt atccactgtttttggaaca (SEQ ID No. 632) | 具有针对 U6 表达的 5' G | 53A |
| KRAS 平铺指导物 26 错配距离 | ggccaaatcctttctctcctgtagcCaaggc cacaaaattatccactgttt (SEQ ID No. 633) | 具有针对 U6 表达的 5' G | 53A |

| | | | | |
|---|---|---|---|---|
| | KRAS 平铺指导物 18 错配距离 | gtttttgtagccaaatcctttctctcctgtagcCaaggccacaaaattatc (SEQ ID No. 634) | 具有针对 U6 表达的 5' G | 53A |
| | KRAS 平铺指导物 10 错配距离 | gatttgctgtttttgtagccaaatcctttctctcctgtagcCaaggccaca (SEQ ID No. 635) | 具有针对 U6 表达的 5' G | 53A |
| | KRAS 平铺指导物 2 错配距离 | gacgatggaatttgctgtttttgtagccaaatcctttctctcctgtagcCa (SEQ ID No. 636) | 具有针对 U6 表达的 5' G | 53A |
| | KRAS 平铺非靶向指导物 | GTAATGCCTGGCTTGTCGACGCATAGTCTG (SEQ ID No. 637) | 具有针对 U6 表达的 5' G | 53A |
| | 转录组特异性的荧光素酶 W85X 靶向指导物 | gatagaatgttctaaacCatcctgcggcctctactctgcattcaattacat (SEQ ID No. 638) | 具有针对 U6 表达的 5' G | 53B |
| | 转录组特异性的非靶向指导物 | GCAGGGTTTTCCCAGTCACGACGTTGTAAAGTTG (SEQ ID No. 639) | 具有针对 U6 表达的 5' G | 53C |
| | 内源 KRAS 指导物 2 | gtcaaggcactcttgccCacgccaccagctccaactaccacaagtttatat (SEQ ID No. 640) | 具有针对 U6 表达的 5' G | 54F |
| [2944] | 内源 PPIB 指导物 1 | gcaaagatcacccggccCacatcttcatctccaattcgtaggtcaaaatac (SEQ ID No. 641) | 具有针对 U6 表达的 5' G | 54G |
| | 内源 KRAS 指导物 1 | GcgccaccagctccaacCaccacaagtttatattcagtcattttcagcagg (SEQ ID No. 642) | 具有针对 U6 表达的 5' G | 54F |
| | 内源 KRAS 指导物 3 | GtttctccatcaattacCacttgcttcctgtaggaatcctctattGTtgga (SEQ ID No. 643) | 具有针对 U6 表达的 5' G | 54F |
| | 内源 PPIB 指导物 2 | GctttctctcctgtagcCaaggccacaaaattatccactgtttttggaaca (SEQ ID No. 644) | 具有针对 U6 表达的 5' G | 54G |
| | 内源性非靶向指导物 | GTAATGCCTGGCTTGTCGACGCATAGTCTG (SEQ ID No. 645) | 具有针对 U6 表达的 5' G | 54F |
| | BoxB Cluc 指导物 | tctttccataGGCCCTGAAAAAGGGCCtgttctaaacCatcctgcggcctctactcGGCCCTGAAAAAGGGCCattcaattac (SEQ ID No. 646) | 具有针对 U6 表达的 5' G | 62B |
| | BoxB 非靶向指导物 | cagctggcgaGGCCCTGAAAAAGGGCCggggatgtgcCgcaaggcgattaagttggGGCCCTGAAAAAGGGCCacgccagggt (SEQ ID No. 647) | 具有针对 U6 表达的 5' G | 62B |
| | Stafforst 全 | GTGGAATAGTATAACAATA | 具有针对 U6 表达 | 62C |

[2945]

| | | | |
|---|---|---|---|
| 长 ADAR2 指导物 1 | TGCTAAATGTTGTTATAGT ATCCCACtctaaaCCAtcctgcgG GGCCCTCTTCAGGGCCC (SEQ ID No. 648) | 的 5' G | |
| Stafforst 全长 ADAR2 非靶向指导物 | GTGGAATAGTATAACAATA TGCTAAATGTTGTTATAGT ATCCCACaccctggcgttacccaGG GCCCTCTTCAGGGCCC (SEQ ID No. 649) | 具有针对 U6 表达的 5' G | 62C |

[2946] 参考文献

[2947] 1.P.D.Hsu,E.S.Lander,F.Zhang,Development and applications of CRISPR-Cas9 for genome engineering.Cell 157,1262-1278(2014).

[2948] 2.A.C.Komor,A.H.Badran,D.R.Liu,CRISPR-Based Technologies for the Manipulation of Eukaryotic Genomes.Cell 1o8,20-36(2017).

[2949] 3.L.Cong et al.,Multiplex gcnome engineering using CRISPR/Cas systems.Science 339,819-823(2013).

[2950] 4.P.Mali et al.,RNA-guided human genome engineering via Cas9.Science 339,823-826(2013).

[2951] 5.B.Zetsche et al.,Cpfl is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system.Cell 163,759-771(2015).

[2952] 6.H.Kim,J.S.Kim,A guide to genome engineering with programmable nucleases.Nat Rev Genet 15,321-334(2014).

[2953] 7.A.C.Komor,Y.B.Kim,M.S.Packer,J.A.Zuris,D.R.Liu,Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage.Nature 533,420-424(2016).

[2954] 8.K.Nishida et al,Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems.Science 353,(2016).

[2955] 9.Y.B.Kim et al.,Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions.Nat Biotechnol 35,371-376(2017).

[2956] 10.O.O.Abudayyeh et al.,C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector.Science 353,aaf5573(2016).

[2957] 11.S.Shmakov et al.,Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems.Mol Cell 60,385-397(2015).

[2958] 12.S.Shmakov et al.,Diversity and evolution of class 2 CRISPR-Cas systems.Nat Rev Microbiol 15,169-182(2017).

[2959] 13.A.A.Smargon et al.,Cas13b Is a Type VI-B CRISPR-Associated RNA-Guided RNase Differentially Regulated by Accessory Proteins Csx27 and Csx28.Mol Cell 65,618-630 e617(2017).

[2960] 14.J.S.Gootenberg et al.,Nucleic acid detection with CRISPR-Cas13a/C2c2.Science 356,438-442(2017).

[2961] 15.O.O.Abudayyeh et al.,RNA targeting with CRISPR-Cas13a.Nature in press,(2017).

[2962] 16.K.Nishikura,Functions and regulation of RNA editing by ADAR deaminases.Annu Rev Biochem 79,321-340(2010).

[2963] 17.B.L.Bass,H.Weintraub,An unwinding activity that covalently modifies its double-stranded RNA substrate.Cell 55,1089-1098(1988).

[2964] 18.M.M.Matthews et al,Structures of human ADAR2 bound to dsRNA reveal base-flipping mechanism and basis for site selectivity Nat Struct Mol Biol 23,426-433(2016).

[2965] 19.A.Kuttan,B.L.Bass,Mechanistic insights into editing-site specificity of ADARs.Proc Natl Acad Sci U S A 109,E3295-3304(2012).

[2966] 20.S.K.Wong,S.Sato,D.W.Lazinski,Substrate recognition by ADAR1 and ADAR2.RNA 7,846-858(2001).

[2967] 21.M.Fukuda et al.,Construction of a guide-RNA for site-directed RNA mutagenesis utilising intracellular A-to-I RNA editing.Sci Rep 7,41478(2017).

[2968] 22.M.F.Montiel-Gonzalez,I.Vallecillo-Viejo,G.A.Yudowski,J.J.Rosenthal,Correction of mutations within the cystic fibrosis transmembrane conductance regulator by site-directed RNA editing.Proc Natl Acad Sci U S A 110,18285-18290(2013).

[2969] 23.M.F.Monfiel-Gonzalez,L C.Vallecillo-Viejo,J.J.Rosenthal,An efficient system for selectively altering genetic information within mRNAs.Nucleic Acids Res 44,e157(2016).

[2970] 24.J.Wettengel,P.Reautschnig,S.Geisler,P.J.Kahle,T.Stafforst,Harnessing human ADAR2 for RNA repair-Recoding a PINK1 mutation rescues mitophagy.Nucleic Acids Res 45,2797-2808(2017.

[2971] 25.Y.Wang,J.Havel,P.A.Beal,A Phenotypic Screen for Functional Mutants of Human Adenosine Deaminase Acting on RNA 1.ACS Chem Biol 10,2512-2519(2015).

[2972] 26.K.A.Lehmann,B.L.Bass,Double-stranded RNA adenosine deaminases ADAR1 and ADAR2 have overlapping specificities.Biochemistry 39,12875-12884(2000).

[2973] 27.Y.Zheng,C.Lorenzo,P.A.Beal,DNA editing in DNA/RNA hybrids by adenosine deaminases that act on RNA.Nucleic Acids Res 45,3369-3377(2017).

[2974] 28.K.Gao et al,A de novo loss-of-function GRIN2A mutation associated with childhood focal epilepsy and acquired epileptic aphasia.PLoS One 12,e0170818(2017).

[2975] 29.H.M.Lanoiselee et al.,APP,PSEN1,and PSEN2 mutations in early-onset Alzheimer disease:A genetic screening study of familial and sporadic cases.PLoS Med 14,e1002270(2017).

[2976] 30.C.Ballatore,V.M.Lee,J,Q.Trojanowski,Tau-mediated neurodegeneration in Alzheimer's disease and related disorders.Nat Rev Neurosci 8,663-672(2007).

[2977] 31.Y.Li et al.,Carriers of rare missense variants in IFIH1 are protected from psoriasis.J Invest Dermatol 130,2768-2772(2010).

[2978] 32.R.S.Finkel et al.,Treatment of infantile-onset spinal muscular atrophy with nusinersen:a phase 2,open-label,dose-escalation study.Lancet 388,3017-3026(2016).

[2979] 实施例4

[2980] 使用酵母筛选来筛选影响ADAR活性的ADAR中的突变。筛选进行多轮。每一轮筛选产生一组候选突变。然后在哺乳动物细胞中验证候选突变。将表现最好的突变添加到突变的最新版本中,并进行重新筛选。经过8轮筛选出的突变在下表中示出。将在第n轮中鉴定出的突变体称为"RESCUE vn"。如本文所讨论,RESCUE是指将腺苷脱氨酶活性转化为胞苷脱氨酶活性的突变。

[2981]

| RESCUE 轮数 | ADAR 突变 | 质粒编号 |
|---|---|---|
| RESCUEv0 | E488Q | pAB0048 |
| RESCUEv1 | E488Q、V351G | |
| RESCUEv2 | E488Q、V351G、S486A | |
| RESCUEv3 | E488Q、V351G、S486A、T375S | pAB0642 |
| RESCUEv4 | E488Q、V351G、S486A、T375S、S370C | pAB1072 |
| RESCUEv5 | E488Q、V351G、S486A、T375S、S370C、P462A | pAB1135 |
| RESCUEv6 | E488Q、V351G、S486A、T375S、S370C、P462A、N597I | pAB1146 |

[2982] 还筛选了其他ADAR突变。测试了这些突变的活性(图69)。在筛选中,突变P462A在所有测试的基序上表现更好(图70)。突变P462改善了基序(图71)。

[2983] 申请人进一步鉴定了具有某些增强活性的另一种突变N579I(图72)。N579I通过将该突变添加到RESCUEv3(即pAB0642)、RESCUEv4(即pAB1072)、RESCUEv5(即pAB1135)和RESCUEv6(即pAB1146)来验证。质粒pAB1143和pAB1144分别是RESCUEv4和RESCUEv5上的G333A(图73)。

[2984] 申请人还用30bp的指导物测试了RESCUEv3的性能(图74)。如下命名30bp的指导物:第一个数字是指导物的长度(即30bp),并且第二个数字是从指导物的5'端开始的错配位置。

[2985] 然后,申请人将50bp的指导物与两个30bp的指导物(30_5和30_11)进行了比较。具有这些指导物的RESCUE突变体在各种基序上的活性在图75中示出。结果表明,RESCUEv6在所有5'基序(TC、GC、AC和CC)上的表现均优于所有其他构建体,允许在每个潜在基序上的编辑超过10%。应当注意,GC的最佳指导物设计与设计TC、AC和CC略有不同。因此,本文提供了用于优化每个实施方案的指导物设计的过程。

[2986] 在某些实施方案中,RESCUE突变在脱氨酶结构域上的位置单独或与靶向Cas13指导物一起递送的全长ADAR不允许编辑。脱氨酶结构域在与Cas13融合以具有可编程活性时显示出活性(图76)。

[2987] ***

[2988] 本发明所描述的方法、药物组合物和试剂盒的各种修改和变更在不偏离本发明的范围和精神的情况下将为本领域技术人员显而易见。尽管已经结合特定实施方案描述本发明,但将了解,本发明能够进行进一步修改并且所要求的发明内容不应过度地限于此类特定实施方案。确实,本领域技术人员显而易见的对所描述的本发明实施方式的各种修改意图处于本发明的范围内。本申请意图涵盖本发明的任何变更、用途或改编,其一般遵循本发明的原理并且包括偏离本公开的内容,所述内容处于本发明所属领域内的已知惯例的范围内并且可以应用于本文之前所陈述的基本特征。

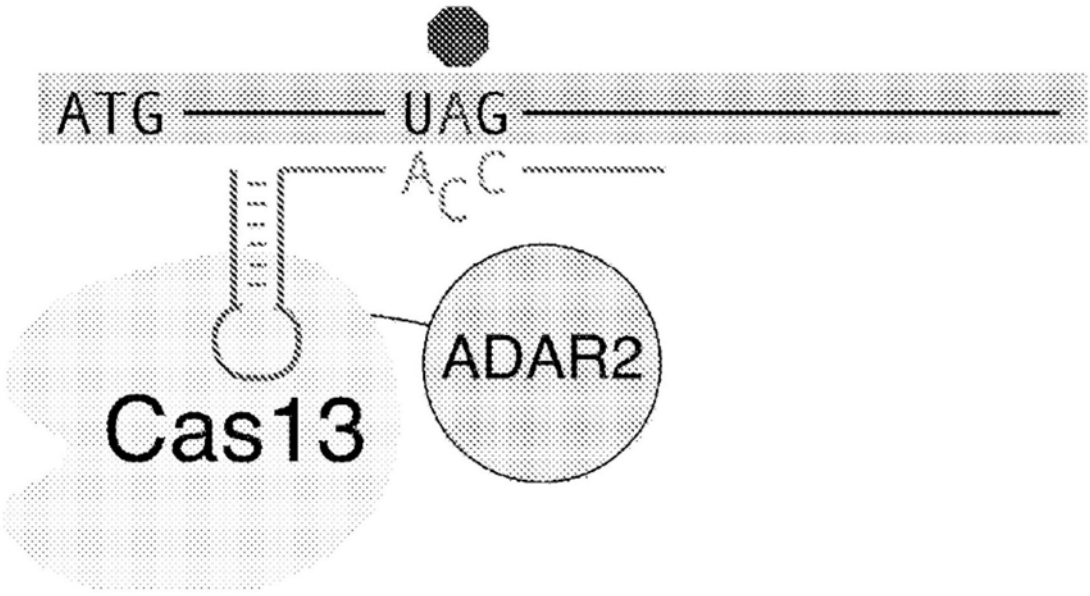

ATG
UAG
ADAR2
Cas13

图1

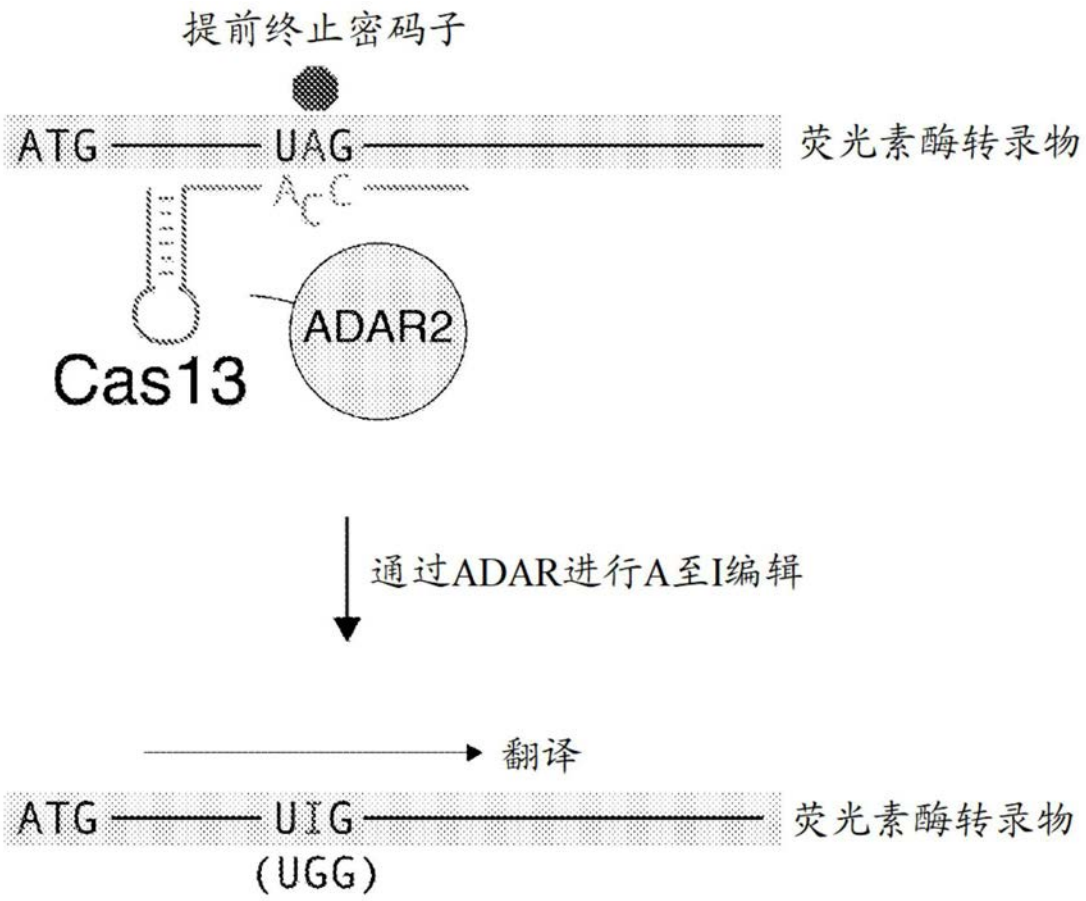

提前终止密码子
ATG
UAG
荧光素酶转录物
ADAR2
Cas13
通过ADAR进行A至I编辑
翻译
ATG
UIG
(UGG)
荧光素酶转录物

图2

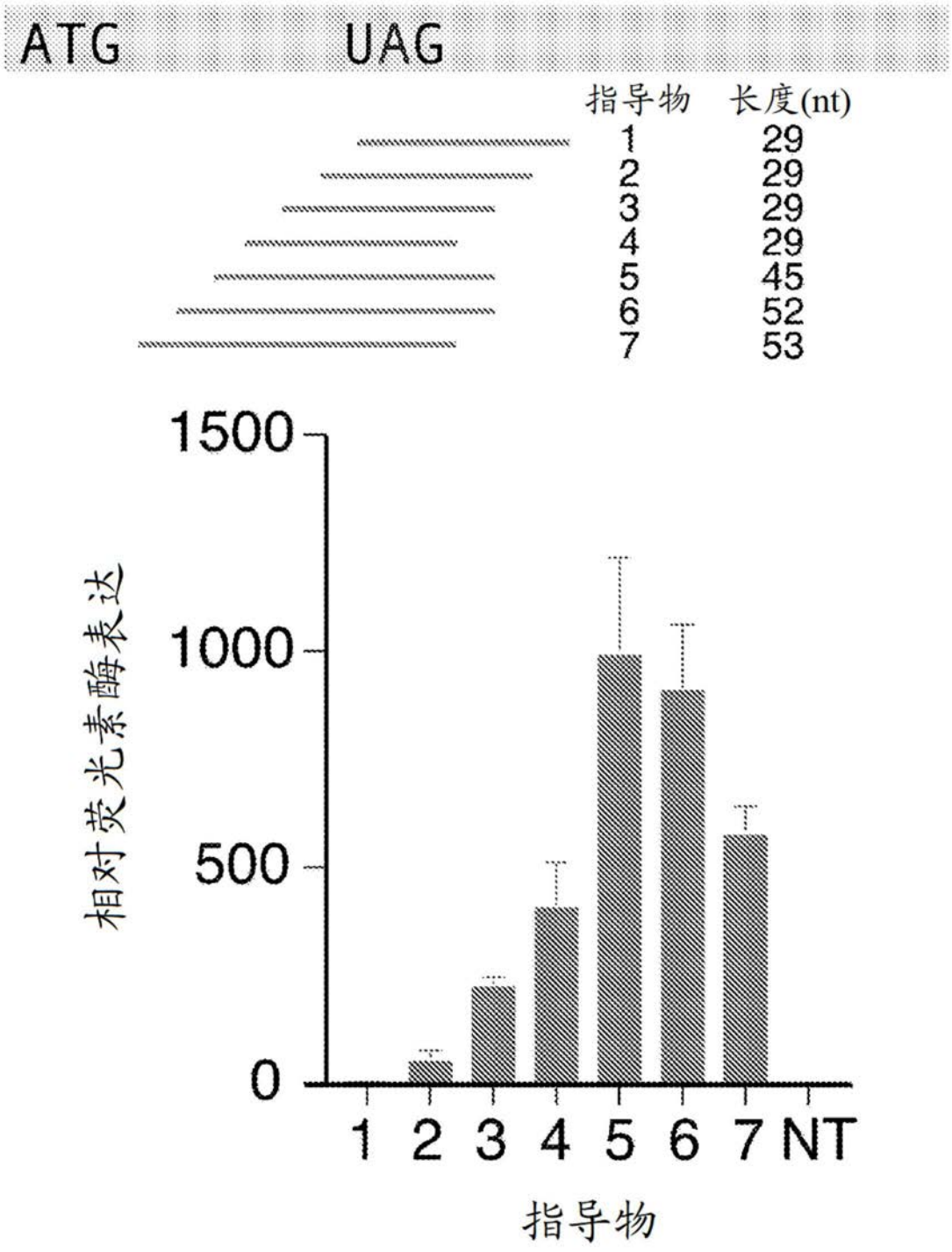

ATG
UAG
指导物
长度(nt)
1 29
2 29
3 29
4 29
5 45
6 52
7 53
1500
1000
500
0
相对荧光素酶表达
1 2 3 4 5 6 7 NT
指导物

图3

| ADAR名称 | 蛋白质序列 |
| --- | --- |
| 智人_ADAR2_E_Q_突变体 (SEQ ID NO : 650) | QLHLPQVLADAVSRLVLGKFGDLTDNFSSPHARRKVLAGVVMT TGTDVKDAKVISVSTGTKCINGEYMSDRGLALNDCHAEIISRRSL LRFLYTQLELYLNNKDDQKRSIFQKSERGGFRLKENVQFHLYIST SPCGDARIFSPHEPILEEPADRHPNRKARGQLRTKIESGQGTIPVR SNASIQTWDGVLQGERLLTMSCSDKIARWNVVGIQGSLLSIFVEP IYFSSIILGSLYHGDHLSRAMYQRISNIEDLPPLYTLNKPLLSGISN AEARQPGKAPNFSVNWTVGDSAIEVINATTGKDELGRASRLCKH ALYCRWMRVHGKVPSHLLRSKITKPNVYHESKLAAKEYQAAKA RLFTAFIKAGLGAWVEKPTEQDQFSLT* |
| 智人_ADAR1_E_Q_突变体 (SEQ ID NO : 651) | SLGTGNRCVKGDSLSLKGETVNDCHAEIISRRGFIRFLYSELMKY NSQTAKDSIFEPAKGGEKLQIKKTVSFHLYISTAPCGDGALFDKS CSDRAMESTESRHYPVFENPKQGKLRTKVENGQGTIPVESSDIVP TWDGIRLGERLRTMSCSDKILRWNVLGLQGALLTHFLQPIYLKS VTLGYLFSQGHLTRAICCRVTRDGSAFEDGLRHPFIVNHPKVGR VSIYDSKRQSGKTKETSVNWCLADGYDLEILDGTRGTVDGPRNE LSRVSKKNIFLLFKKLCSFRYRRDLLRLSYGEAKKAARDYETAK NYFKKGLKDMGYGNWISKPQEEKNF |
| 普通章鱼_ADAR1_E_Q_突变体 (SEQ ID NO : 652) | SVGTGNRCLTGDHLSLEGNSVNDSHAEMITRRGFLRYLYRHLLE YDAEVPNDLFEKGERSICRIKTNITFHLYISTAPCGDGALFSPRDT DSSNAKMEEENKHIHNPTFSSSVQGLLRTKVEGGQGTIPIDADFT EQTWDGIQRGERLRTMSCSDKICRWNVVGLQGALLSHFIEPIYL DSLTLGYLYDHGHLARAVCCRIERGEASVNQLLPEGYRLNHPW LGRVTACDPPRETQKTKSLSINWCYDDEKSEVLDGTAGICYTAIE KNLFSRLTKHNLYEEFKRVCRKFDRNDLLTAPSYNKAKMMATP FQTAKNVMLKKLKENNCGTWVSKPIEEEMF |
| 乌贼_ADAR1_E_Q_突变体 (SEQ ID NO : 653) | SVGTGNRCLTGDRLSLEGNSVNDSHAEMVTRRGFLRYLYKHLL EYDPEKPHDLFEKGERSLCRIKTNITFHLYISTAPCGDGALFSPRD TDSSNVKVDEENKHVHNPTFSSSVQGLLRTKVEGGQGTIPIDADF TEQTWDGIQRGERLRTMSCSDKICRWNVVGLQGALLSHFVEPIY LESLTLGYLYDHGHLARAVCCRIERGEASVNQLLPEGYRLNHPW LGRVTACDPPRETQKTKSLSINWCYDDEKSEVLDGTAGICYTAIE KNLFSRLTKHSLYEEFKKVCQKFEREDLLNVTSYNKAKMMAIPF QTAKNVMLKKLKENNCGTWVSKPIEEEMF |

图4

| | |
|---|---|
| 普通章鱼_ADAR2_E_Q_突变体 (SEQ ID NO : 654) | GIGTGTKCINGEHMSDRGFGVNDCHAEIIARRCFLRYIYDQLELH LSDNSDVRNSSIFELRDKGGYQLKENIQFHLYISTAPCGDARIFSP HGQDVETGDRHPNRKARGQLRTKIESGQGTIPVRTSGVIQTWDG VLEGERLLTMSCSDKIARWNVLGIQGSLLSHFMNPIYLESIILGSL YHSDHLSRAMYSRISIIENLPEPFHLNRPFLSGISSPESRQPGKAPN FGINWRKEDETFEVINAMTGRVEGGSVSRICKQALFGRFMSLYG KLSSLTGQSVTTRPTHYSDAKAAVMEYQLAKQCVFQAFQKAGL GNWVQKPIEQDQF |
| 乌贼_ADAR2_E_Q_突变体 (SEQ ID NO : 655) | GIGTGTKCINGEYMNDRGFAVNDCHAEIIARRCFLRFIYDQLEMH LSEDPEVRGQSVFELRDGGGYKLKPNIHFHLYISTAPCGDARIFSP HGQDVETGDRHPNRKARGQLRTKIESGQGTIPVRSSGFIQTWDG VLEGERLLTMSCSDKIARWNVLGIQGALLCHFMHPIYLESIILGSL YHSDHLSRAVYCRIASIENLPDLFQLNRPFLSGISSPESRQPGKAP NFGINWRRNDDTFEVINAMTGRVEGGNMSRICKQALFDRFMNL YGRLSSLTGQSVTTRPTLYSEAKAAVMEYQLAKQCVFQAFQKA GLGNWVQKPIEQDQF |
| 乳白近岸鱿鱼_ADAR2_E_Q_突变体 (SEQ ID NO : 656) | GIGTGTKCINGEYMNDRGFAVNDCHAEIIARRCFLRFIYDQLELH LSDNAEVRGQSIFELRDAGGYKLKPNIQFHLYISTAPCGDARIFSP HGQDVETGDRHPNRKARGQLRTKIESGQGTIPVRSSGFIQTWDG VLEGERLLTMSCSDKIARWNVLGVQGALLCHFMHPIYLESIILGS LYHSDHLSRAVYCRIAAIENLPDLFRLNRPFLSGISSPESRQPGKA PNFGINWRRNDDSFEVINAMTGRVEGGSMSRICKQALFDRFMNL YGKLSSLTGQSVTTRPALYSEAKATVMEYQLAKQCVFQAFQKA GLGNWVQKPIEQDQF |

图4(续)

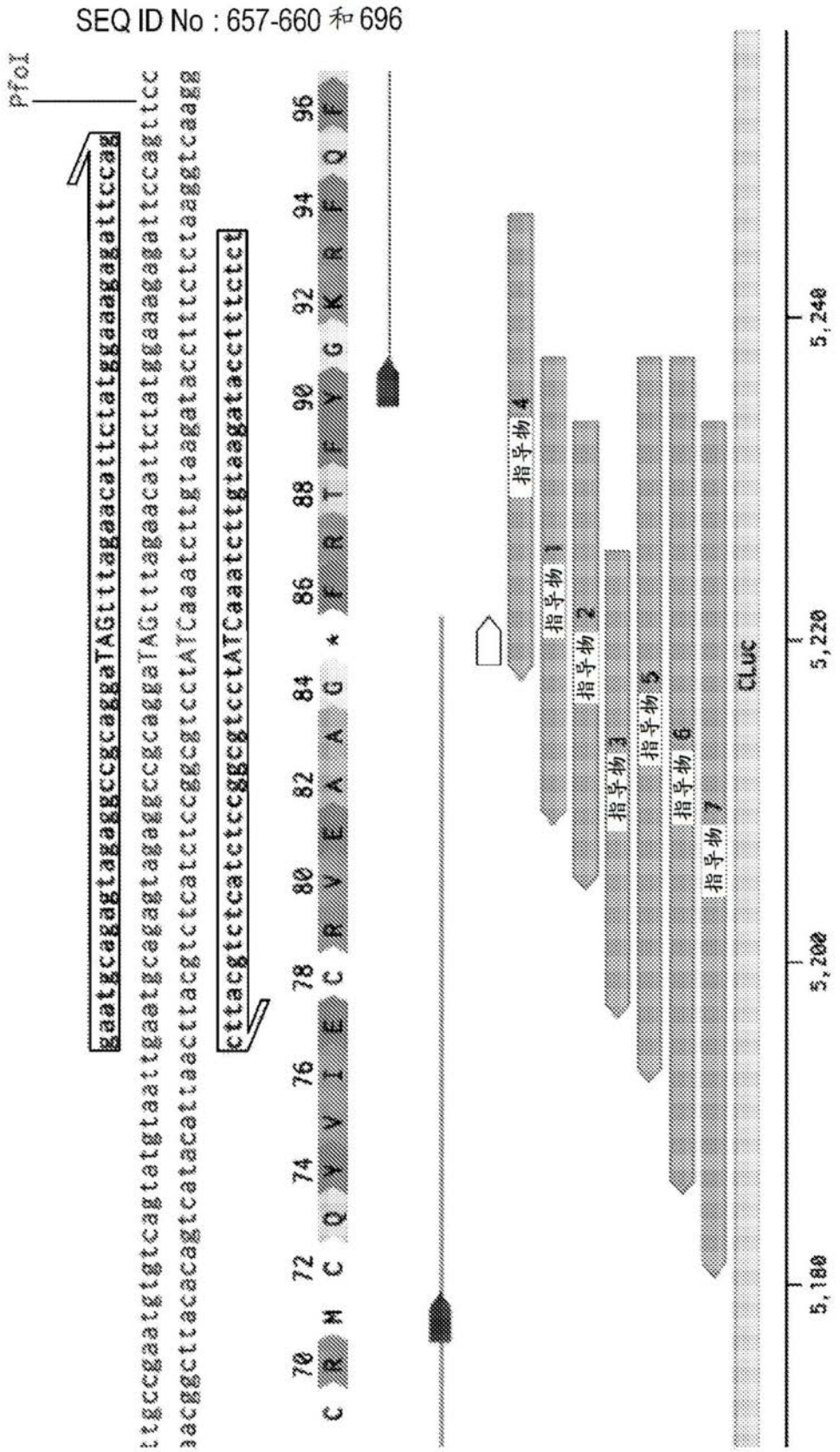

SEQ ID No : 657-660 和 696
pfoX
指导物 4
指导物 1
指导物 2
指导物 3
指导物 5
指导物 6
指导物 7
CLuc
5,240
5,220
5,200
5,180

图5

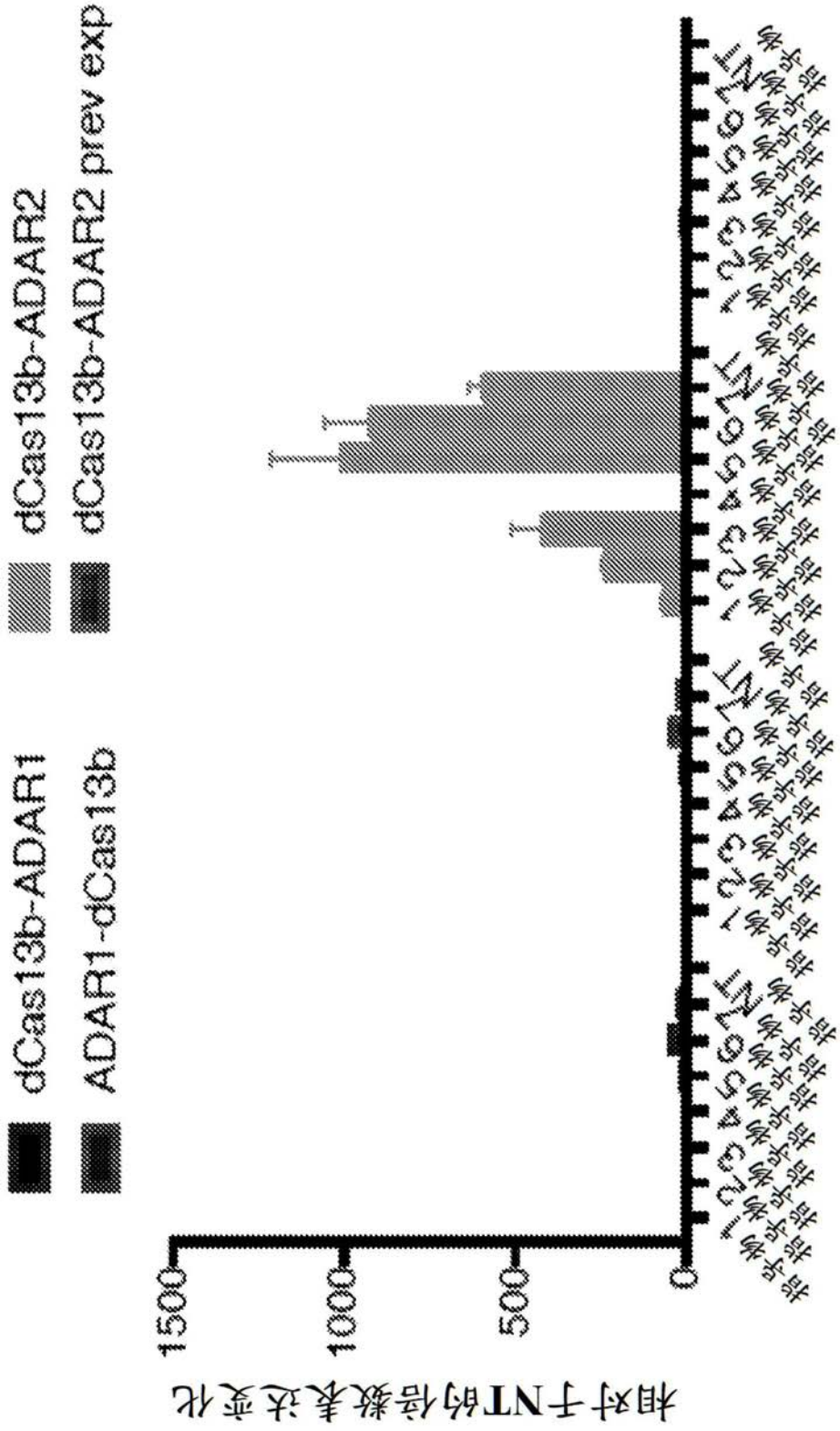

dCas13b-ADAR1
ADAR1-dCas13b
dCas13b-ADAR2
dCas13b-ADAR2 prev exp
相对于NT的倍数表达变化
1500
1000
500
0
非靶向
NT
ADAR1-dCas13b
dCas13b-ADAR1
dCas13b-ADAR2
dCas13b-ADAR2 prev exp

图6

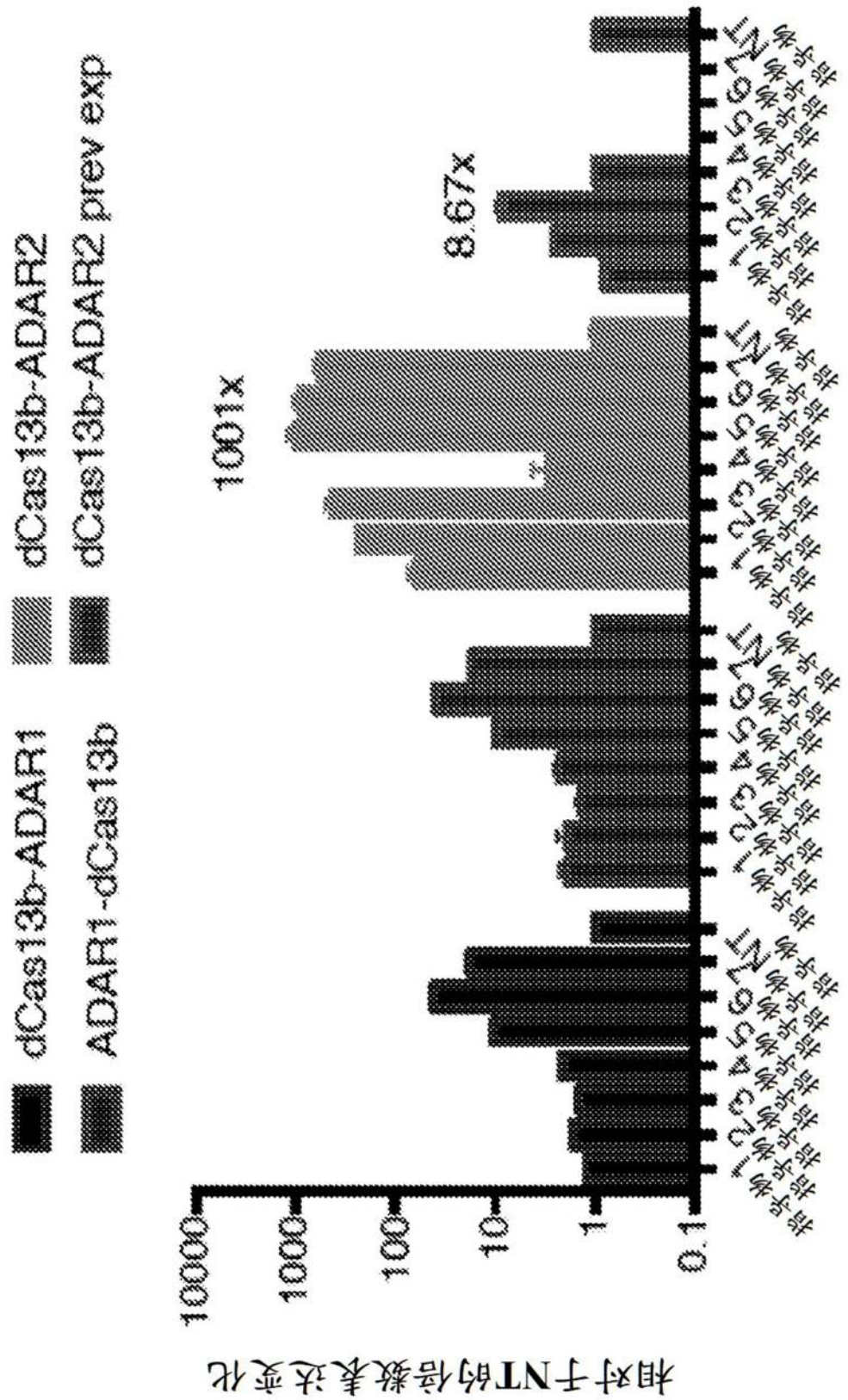

dCas13b-ADAR1
ADAR1-dCas13b
dCas13b-ADAR2
dCas13b-ADAR2 prev exp
10001x
8.67x
10000
1000
100
10
1
0.1
相对于NT的倍数表达变化

图7

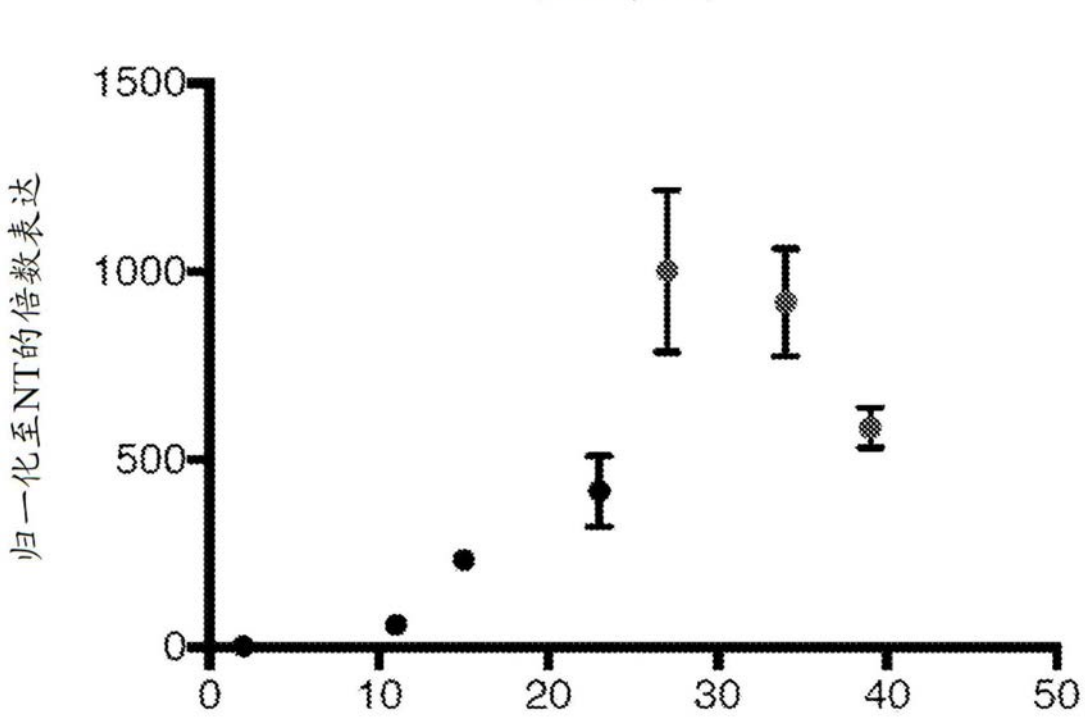



图8

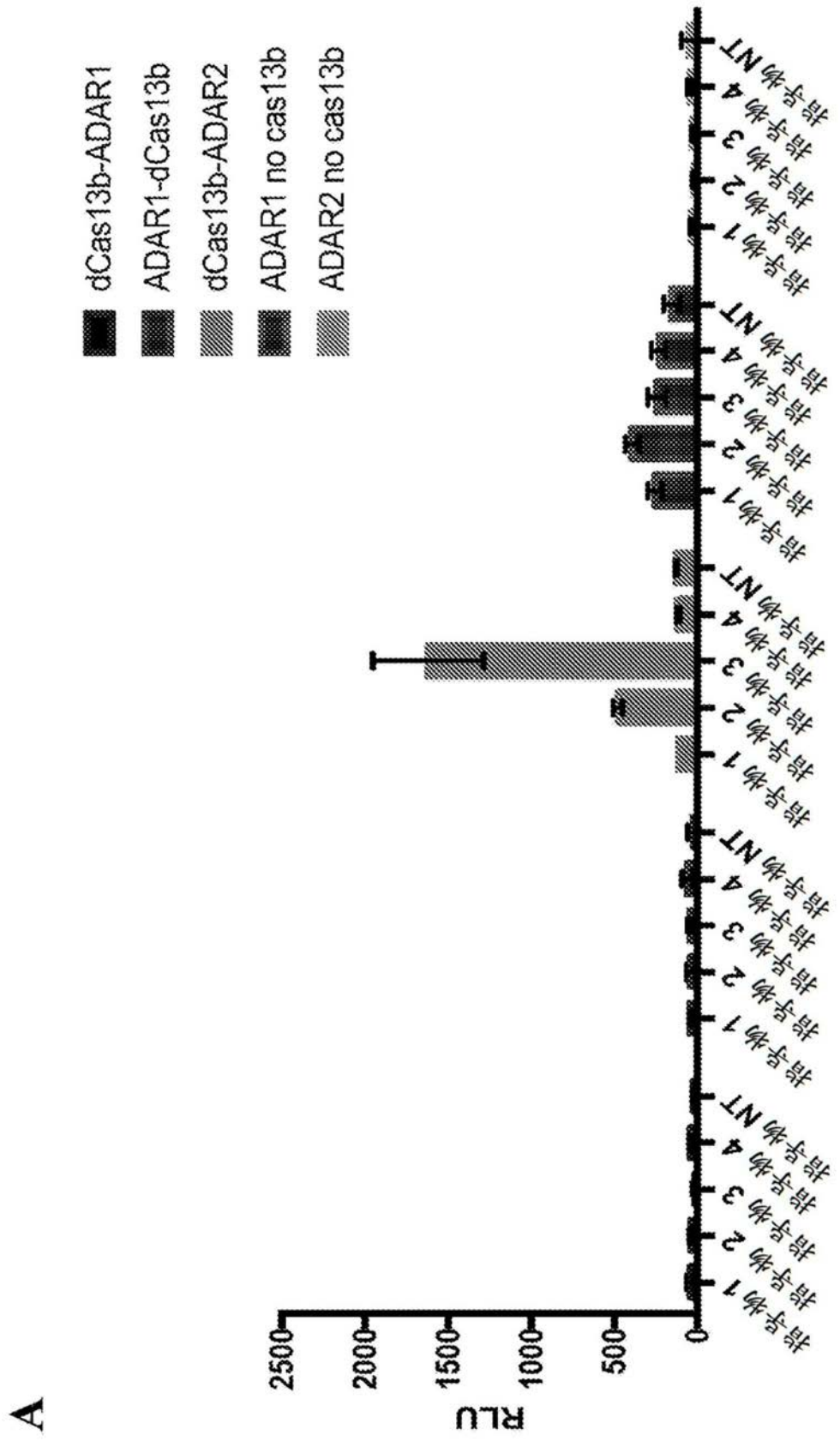

A
RLU
2500
2000
1500
1000
500
0
dCas13b-ADAR1
ADAR1-dCas13b
dCas13b-ADAR2
ADAR1 no cas13b
ADAR2 no cas13b
指导物1
指导物2
指导物3
指导物4
指导物NT

图9A

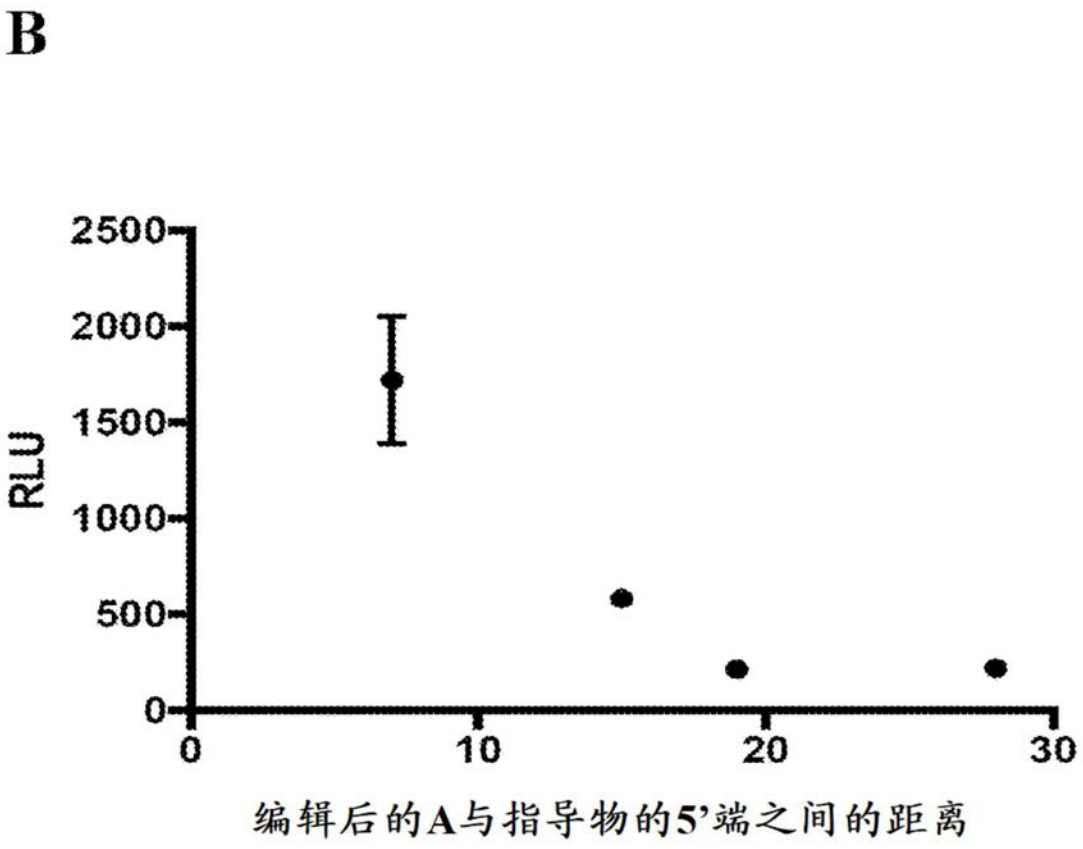



图9B

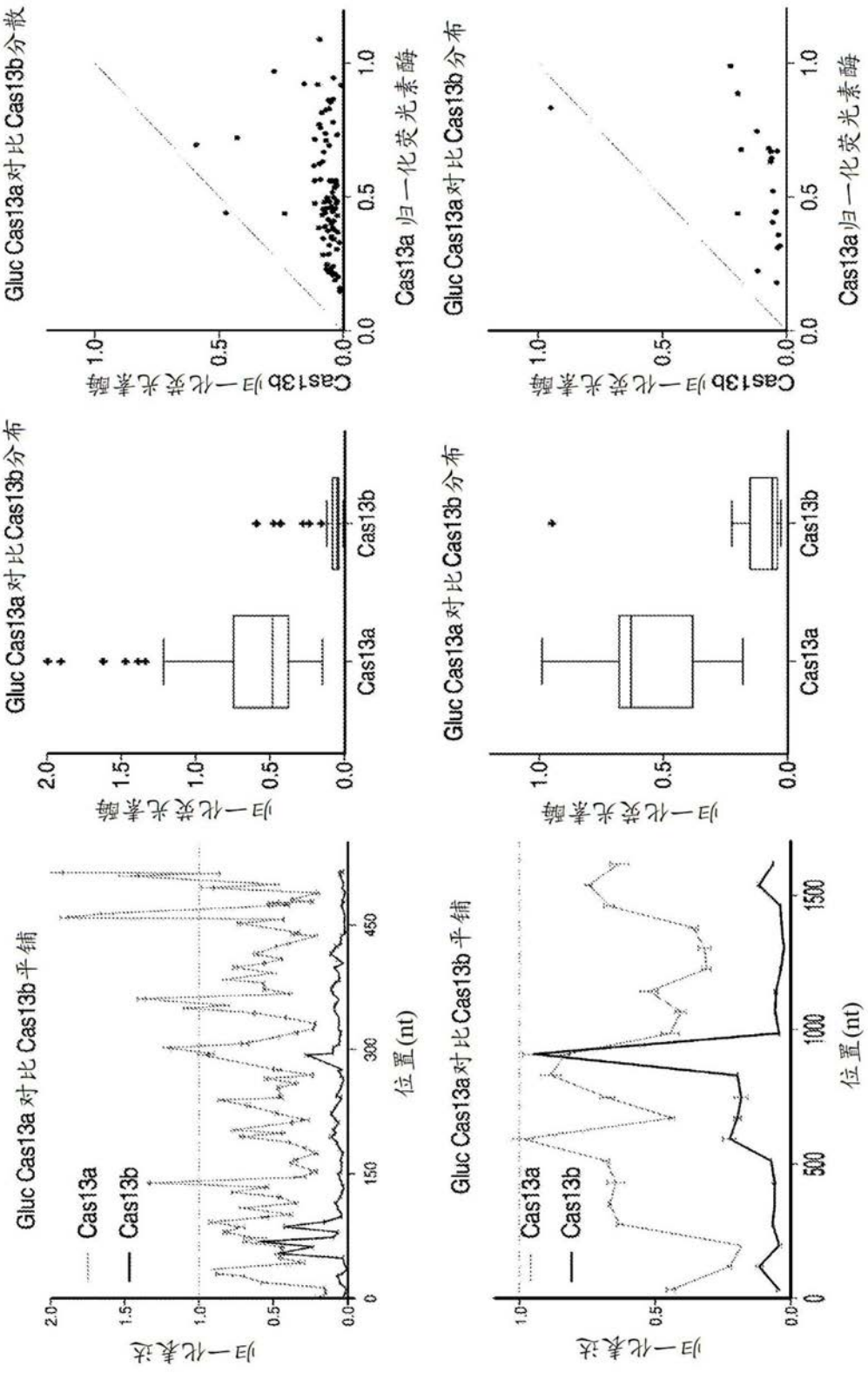

Gluc Cas13a 对比 Cas13b 平铺
Cas13a
Cas13b
归一化表达
位置(nt)
Gluc Cas13a 对比 Cas13b 分布
归一化荧光素酶
Cas13a
Cas13b
Gluc Cas13a 对比 Cas13b 分散
Cas13a 归一化荧光素酶
Cas13b 归一化荧光素酶

图10

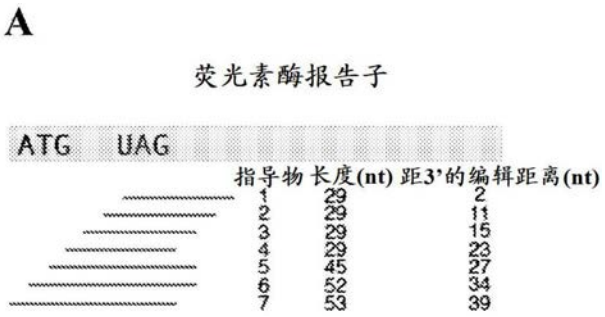

A
荧光素酶报告子
ATG
UAG
指导物 长度(nt) 距3'的编辑距离(nt)
1 29 2
2 29 11
3 29 15
4 29 23
5 45 27
6 52 34
7 53 39

图11A

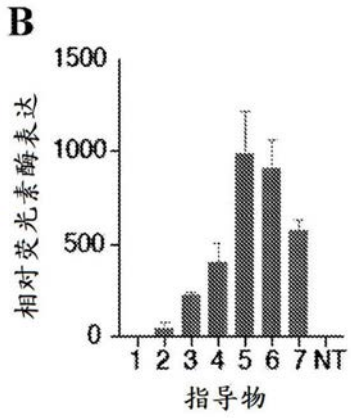
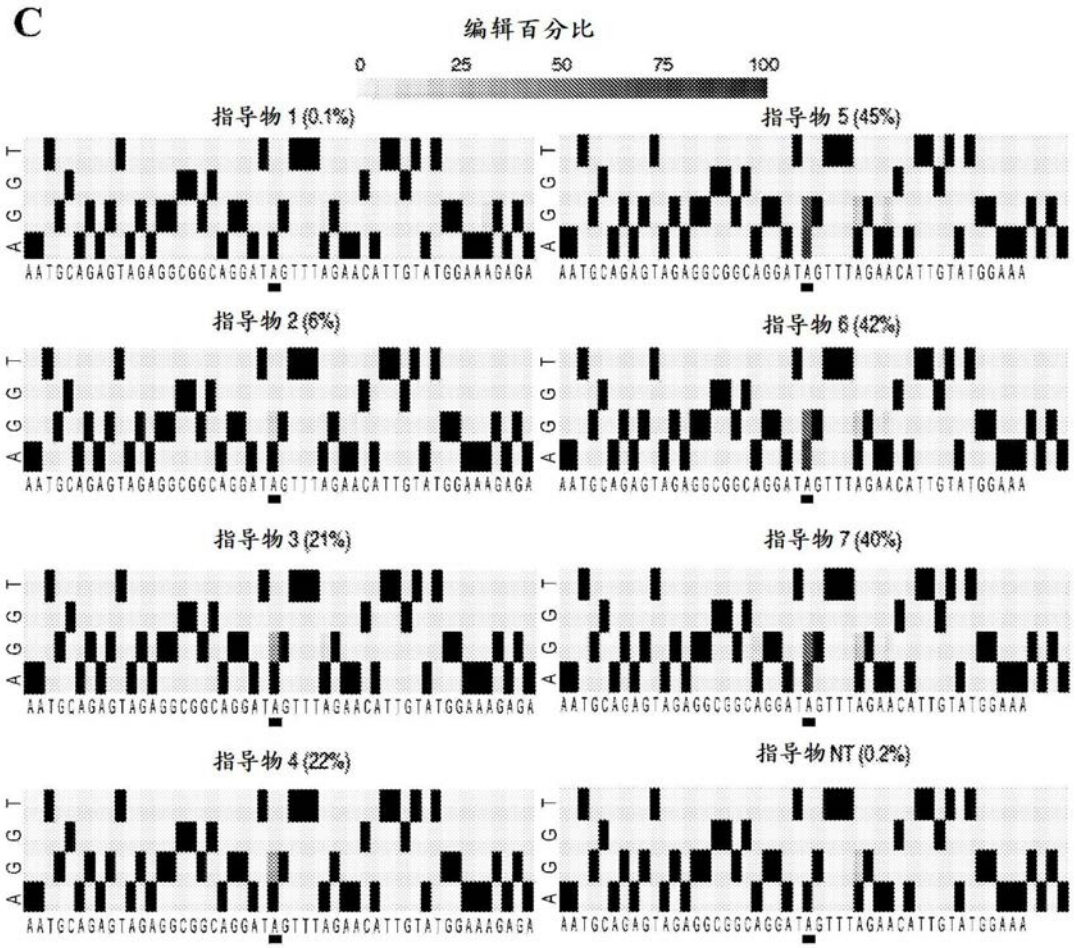

B
1500
1000
500
0
相对荧光素酶表达
1 2 3 4 5 6 7 NT
指导物

图11B




靶向A显示为红色并带有下划线。中靶百分比编辑在括号中。

图11C

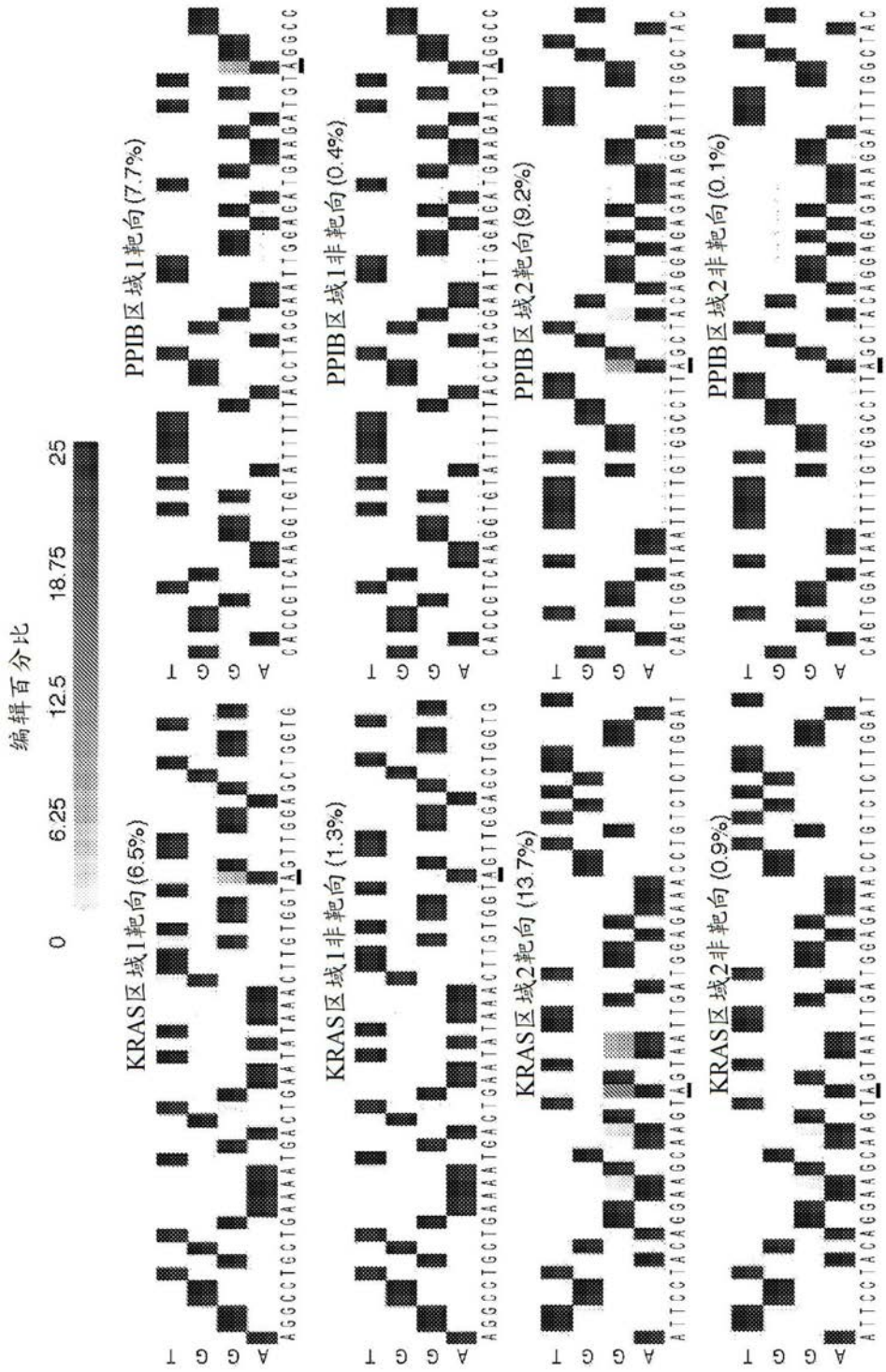

编辑百分比
0
6.25
12.5
18.75
25
KRAS区域1靶向(6.5%)
KRAS区域1非靶向(1.3%)
KRAS区域2靶向(13.7%)
KRAS区域2非靶向(0.9%)
PPIB区域1靶向(7.7%)
PPIB区域1非靶向(0.4%)
PPIB区域2靶向(9.2%)
PPIB区域2非靶向(0.1%)
靶向A显示为红色并带有下划线。中靶百分比编辑在括号中。

图12

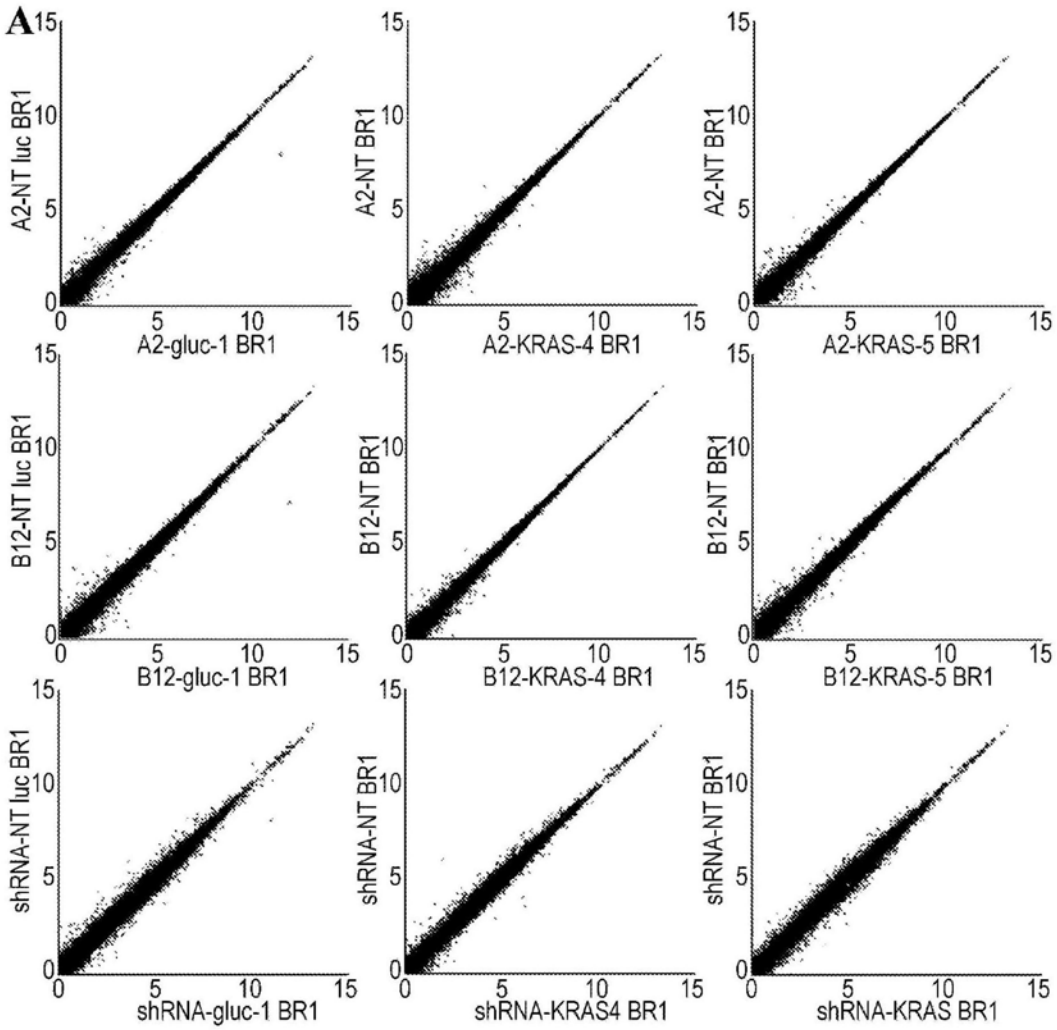

A
A2-NT luc BR1
A2-gluc-1 BR1
A2-NT BR1
A2-KRAS-4 BR1
A2-NT BR1
A2-KRAS-5 BR1
B12-NT luc BR1
B12-gluc-1 BR1
B12-NT BR1
B12-KRAS-4 BR1
B12-NT BR1
B12-KRAS-5 BR1
shRNA-NT luc BR1
shRNA-gluc-1 BR1
shRNA-NT BR1
shRNA-KRAS4 BR1
shRNA-NT BR1
shRNA-KRAS BR1

图13A

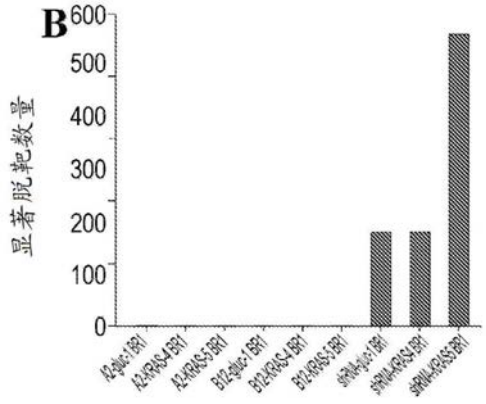

B
显著脱靶数量
600
500
400
300
200
100
0

图13B

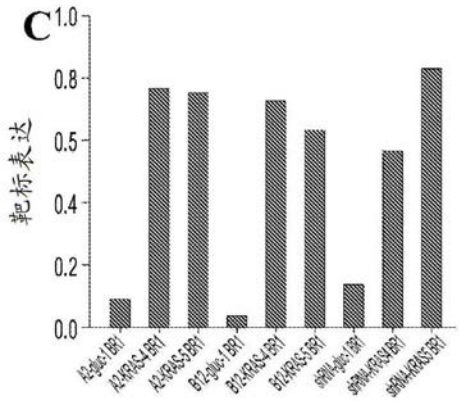

C
1.0
0.8
0.6
0.4
0.2
0.0
靶标表达
A2-gpc-1 BRT
A24VRAS-4 BRT
A24VRAS-5 BRT
B12-gpc-1 BRT
B12VRAS-4 BRT
B12VRAS-5 BRT
sRNA-gpc-1 BRT
sRNAVRAS-4 BRT
sRNAVRAS-5 BRT

图13C

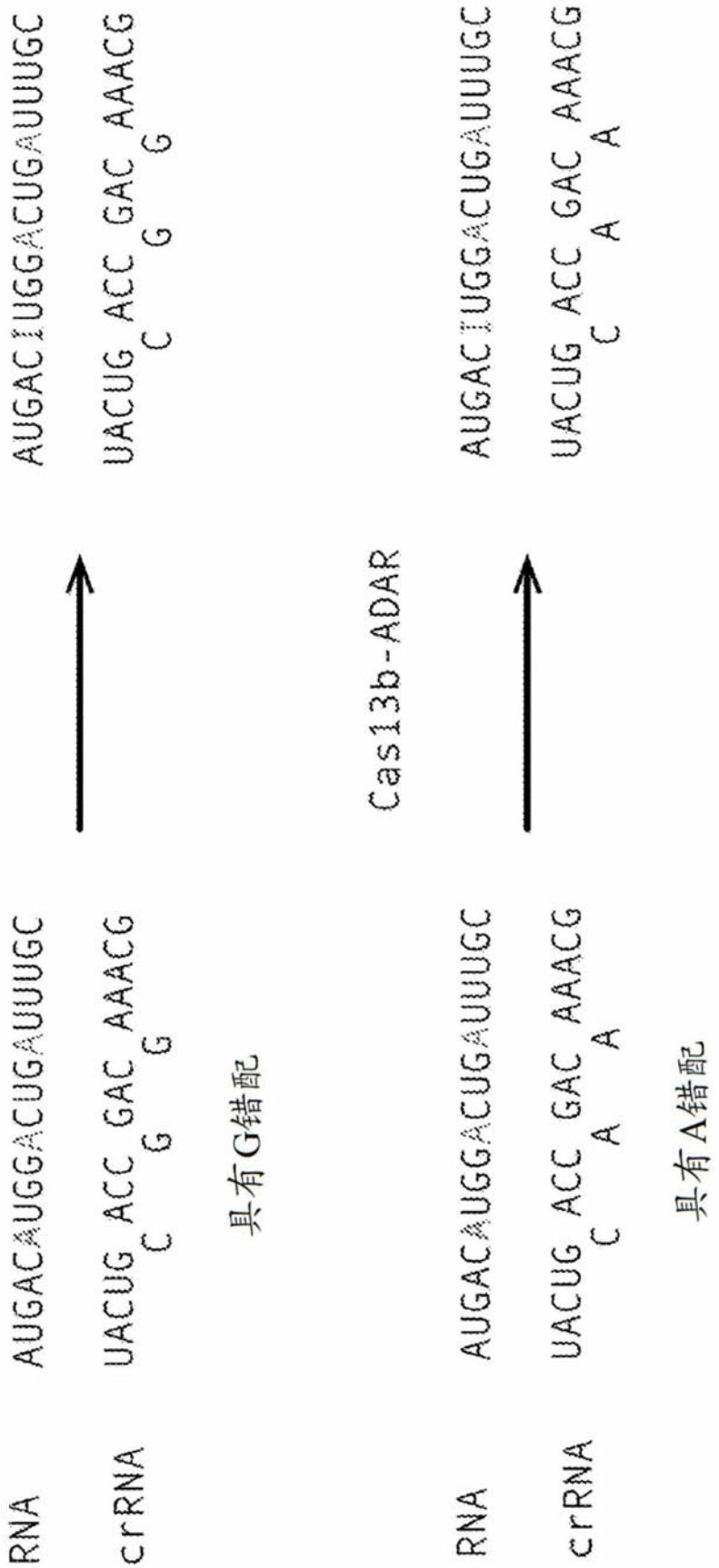

RNA
AUGACAUGGACUGAUUUGC
crRNA
UACUG ACC GAC AAACG
C G G
具有G错配
Cas13b-ADAR
RNA
AUGACIUGGACUGAUUUGC
crRNA
UACUG ACC GAC AAACG
C A A
具有A错配

图14

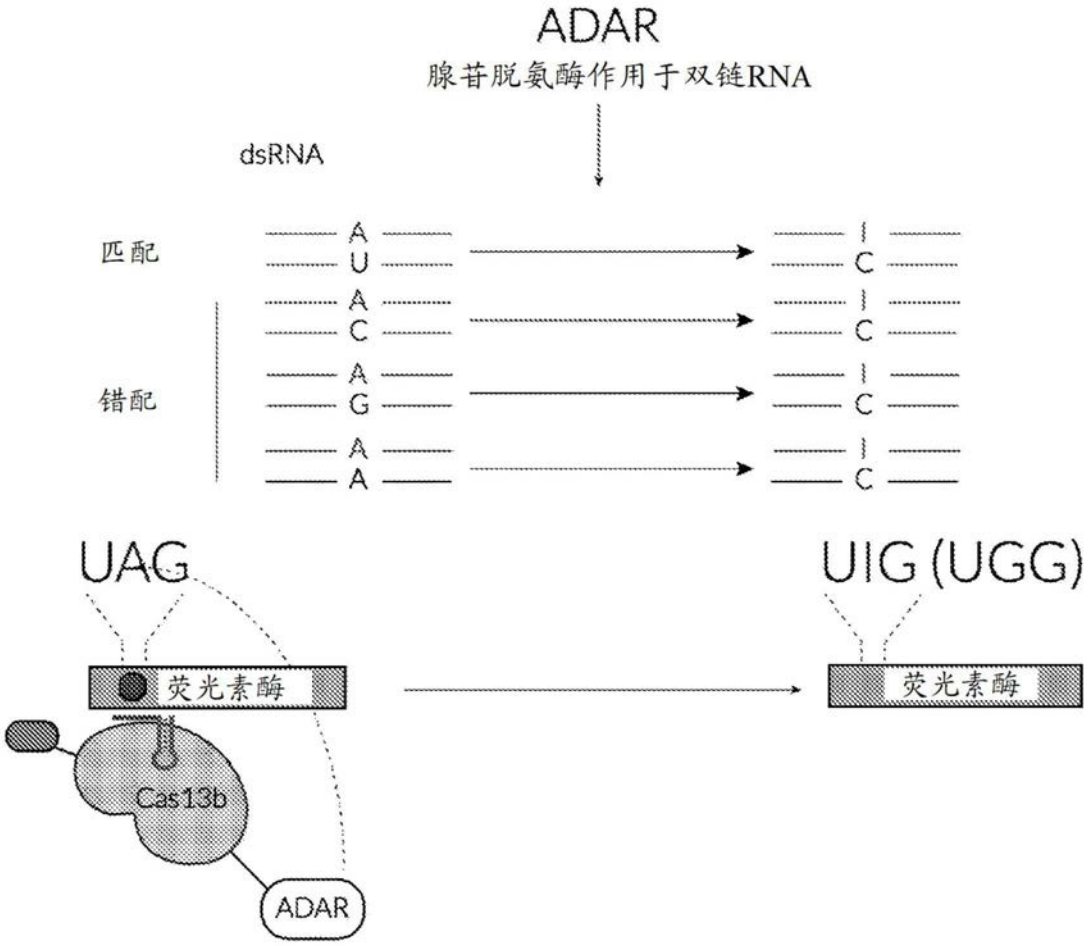

ADAR
腺苷脱氨酶作用于双链RNA
dsRNA
匹配
错配
A
U
C
A
C
C
A
G
C
A
A
C
UAG
UIG (UGG)
荧光素酶
荧光素酶
Cas13b
ADAR

图15

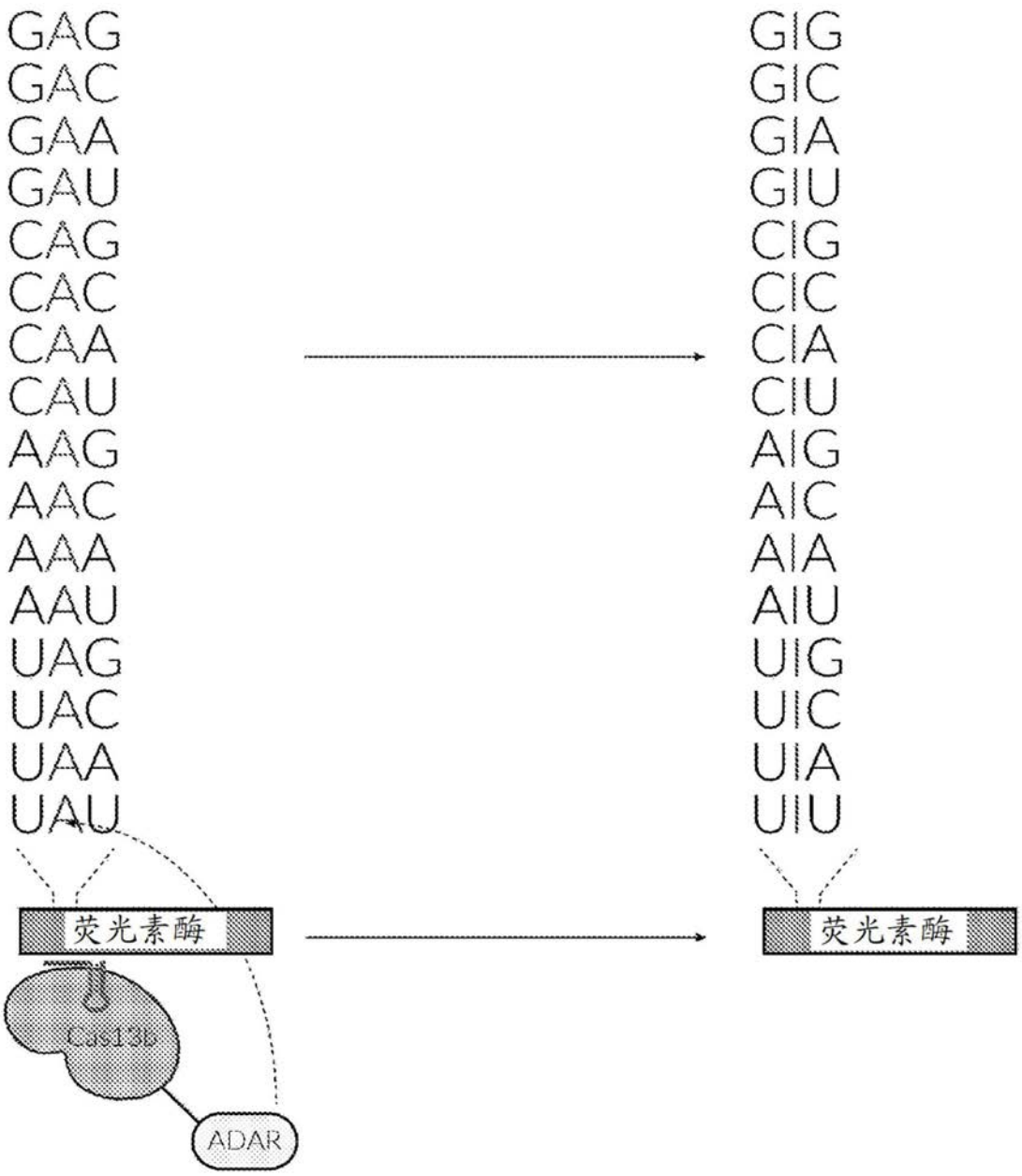

GAG
GAC
GAA
GAU
CAG
CAC
CAA
CAU
AAG
AAC
AAA
AAU
UAG
UAC
UAA
UAU
GIG
GIC
GIA
GIU
CIG
CIC
CIA
CIU
AIG
AIC
AIA
AIU
UIG
UIC
UIA
UIU
荧光素酶
荧光素酶
Cas13b
ADAR

图16

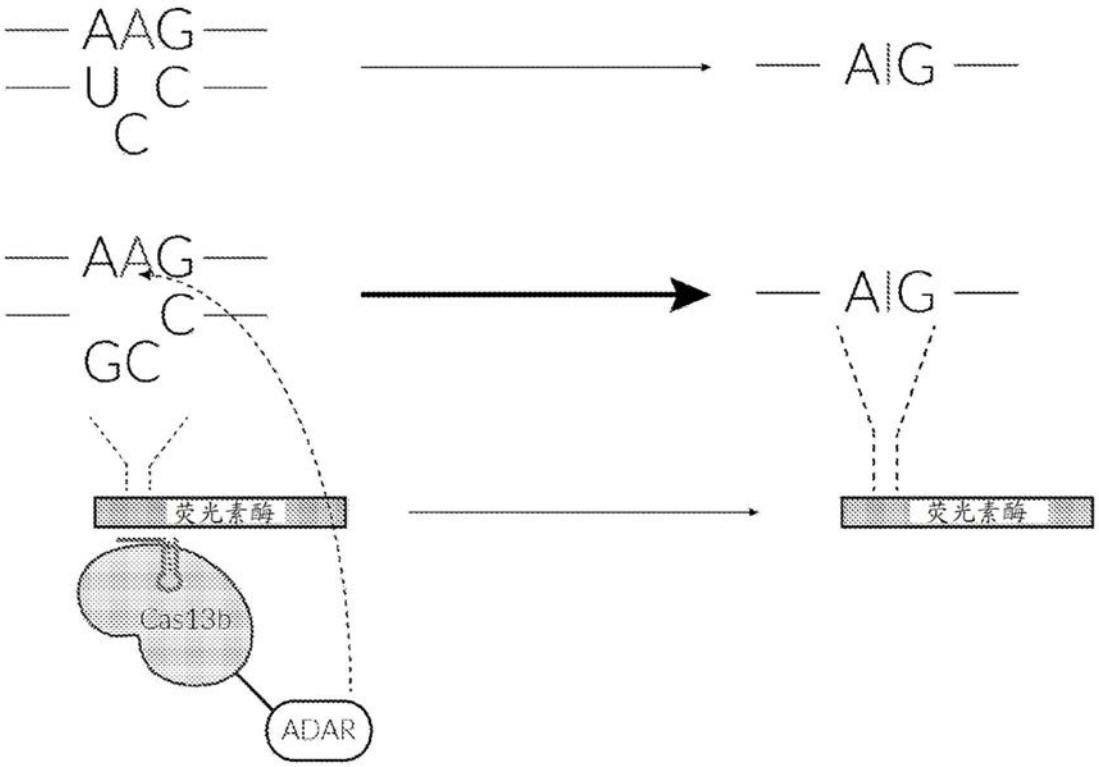

AAG
UC
C
AIG
AAG
C
GC
AIG
荧光素酶
荧光素酶
Cas13b
ADAR

图17

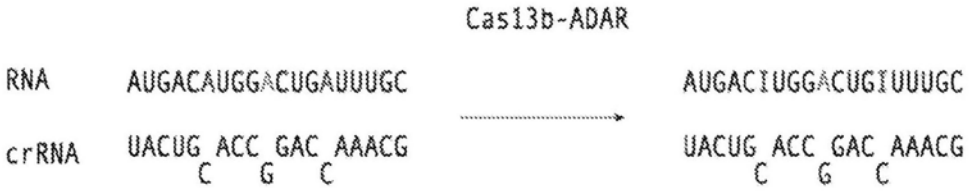

Cas13b-ADAR
RNA
AUGACAUGGACUGAUUUGC
AUGACIUGGACUGIUUUGC
crRNA
UACUG ACC GAC AAACG
C G C
UACUG ACC GAC AAACG
C G C

图18

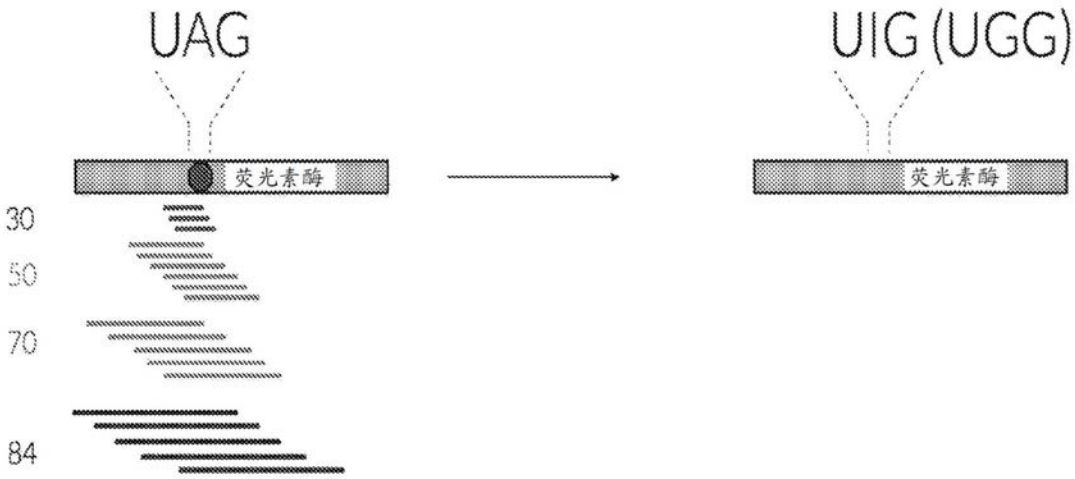

UAG
UIG (UGG)
荧光素酶
荧光素酶
30
50
70
84

图19

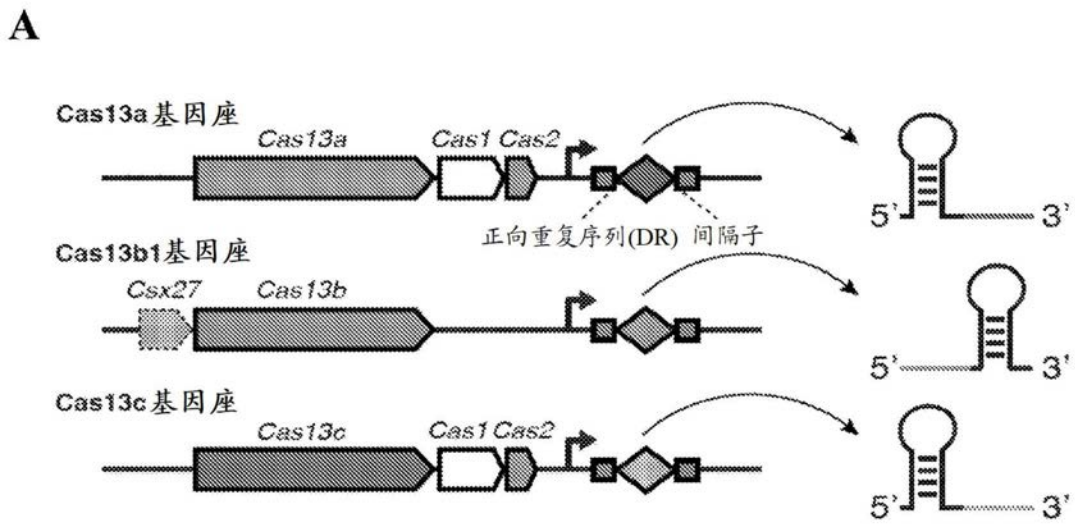

A
Cas13a基因座
Cas13a
Cas1 Cas2
正向重复序列(DR)
间隔子
5'
3'
Cas13b1基因座
Csx27
Cas13b
5'
3'
Cas13c基因座
Cas13c
Cas1 Cas2
5'
3'

图20A

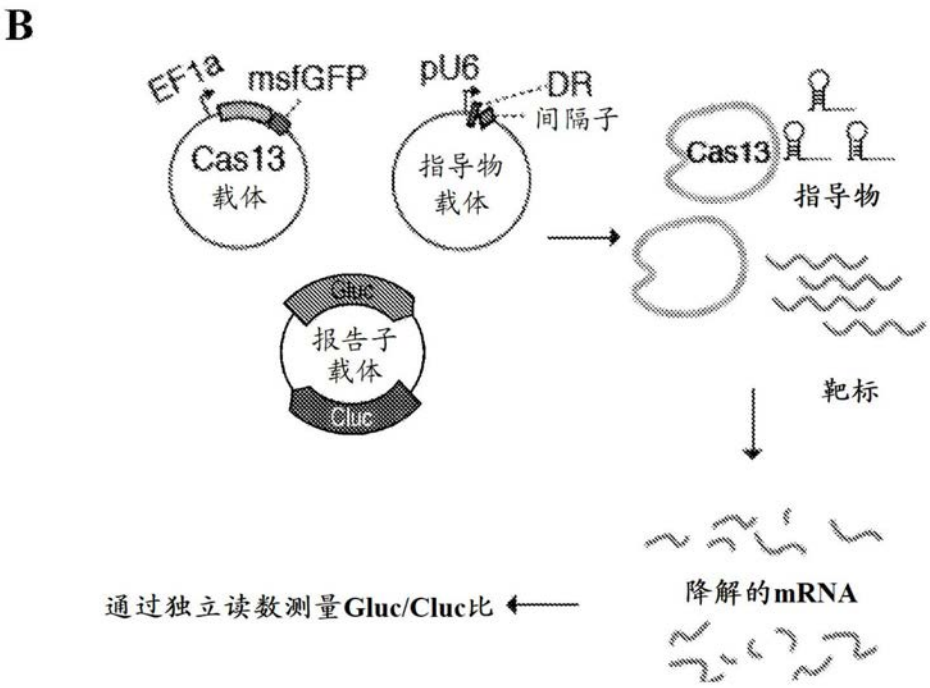

B
EF1a
msfGFP
Cas13
载体
pU6
DR
间隔子
指导物
载体
Gluc
报告子
载体
Cluc
Cas13
指导物
靶标
降解的mRNA
通过独立读数测量Gluc/Cluc比

图20B

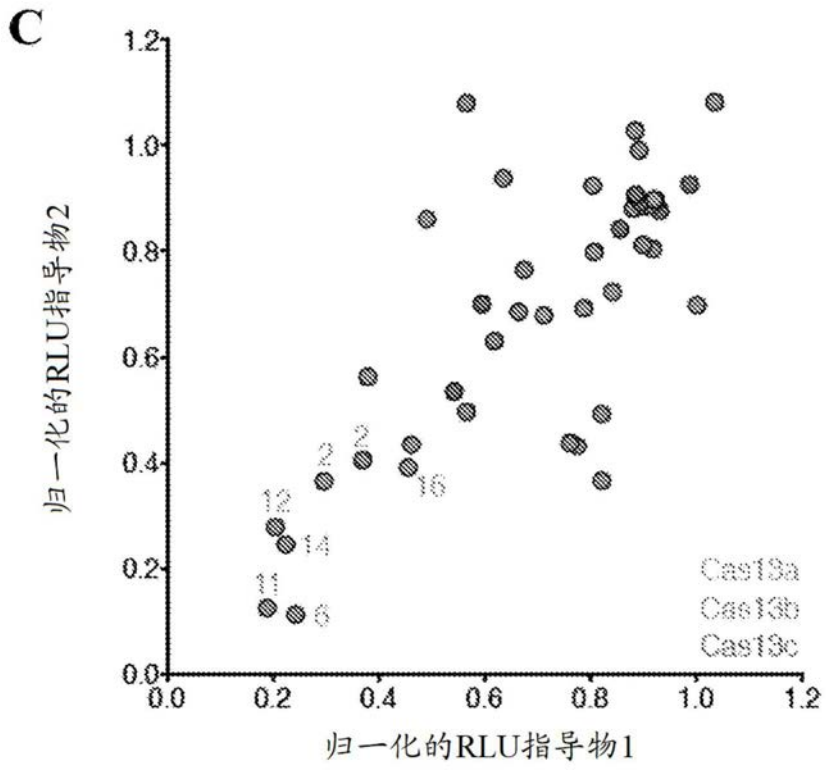

C
归一化的RLU指导物2
归一化的RLU指导物1
0.0
0.2
0.4
0.6
0.8
1.0
1.2
2
2
16
12
14
11
6
Cas13a
Cas13b
Cas13c

图20C

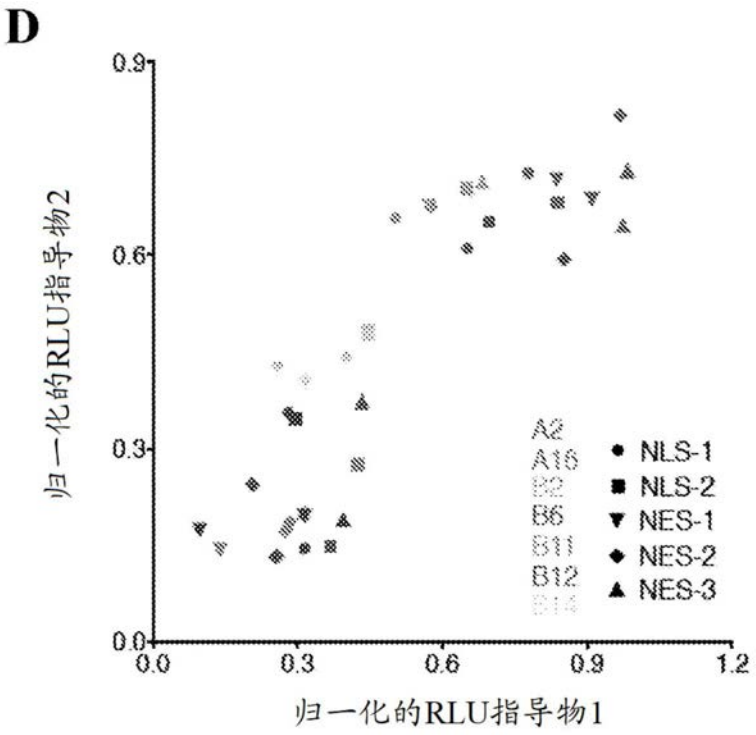

D
0.9
0.6
0.3
0.0
0.0
0.3
0.6
0.9
1.2
归一化的RLU指导物2
归一化的RLU指导物1
A2
A16
B2
B6
B11
B12
B14
NLS-1
NLS-2
NES-1
NES-2
NES-3

图20D

**E**

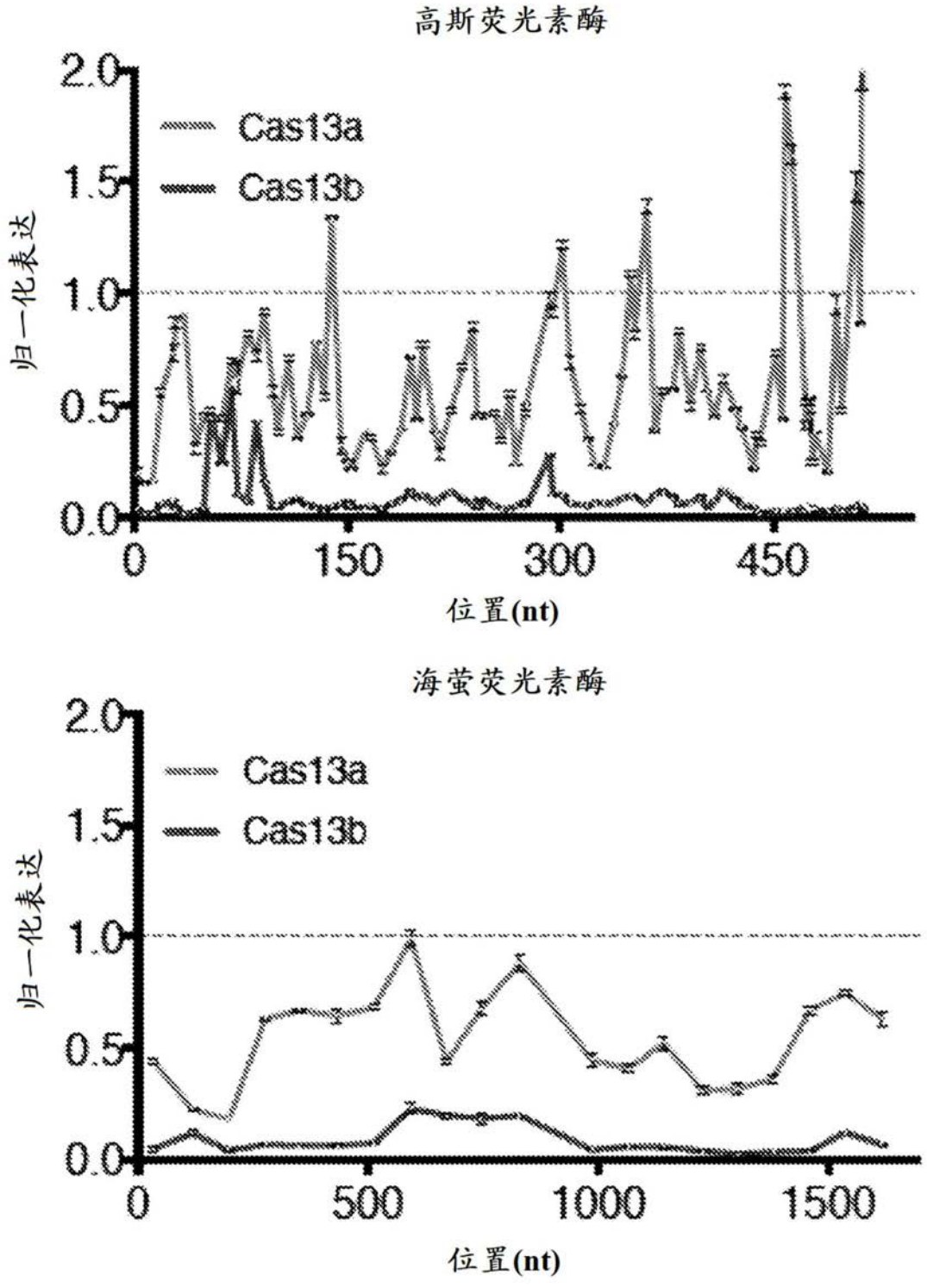

高斯荧光素酶
2.0
1.5
1.0
0.5
0.0
归一化表达
Cas13a
Cas13b
0
150
300
450
位置(nt)
海萤荧光素酶
2.0
1.5
1.0
0.5
0.0
归一化表达
Cas13a
Cas13b
0
500
1000
1500
位置(nt)

图20E

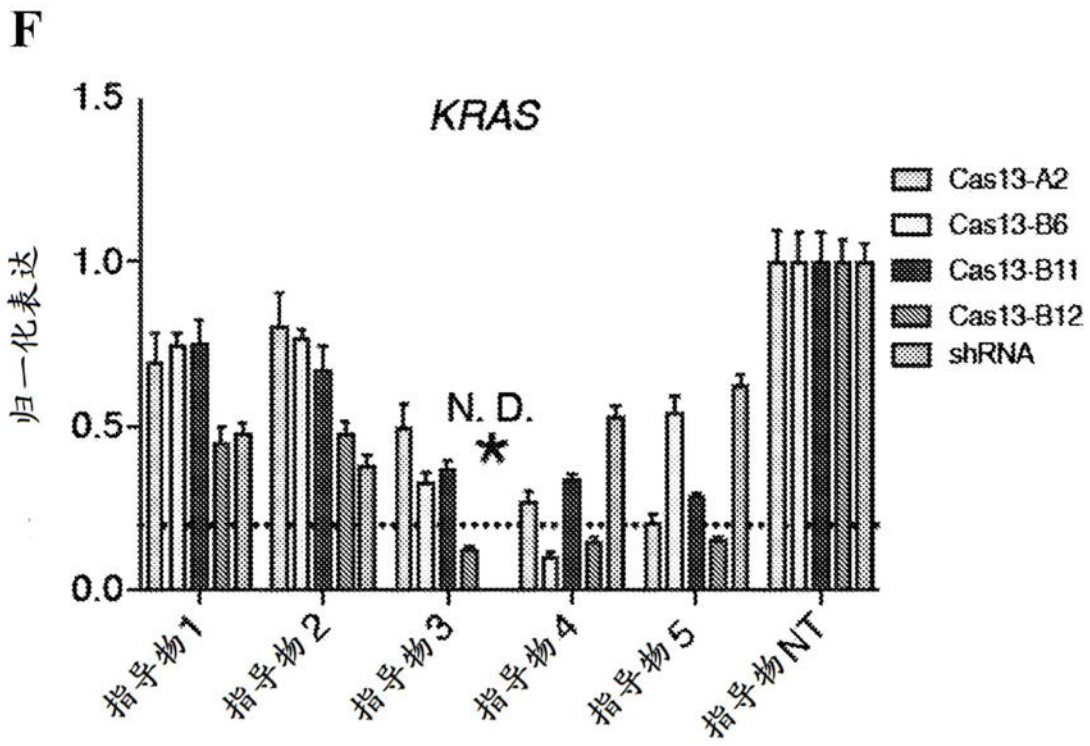

F
KRAS
1.5
1.0
0.5
0.0
归一化表达
N. D.
*
Cas13-A2
Cas13-B6
Cas13-B11
Cas13-B12
shRNA
指导物1
指导物2
指导物3
指导物4
指导物5
指导物NT

图20F

A

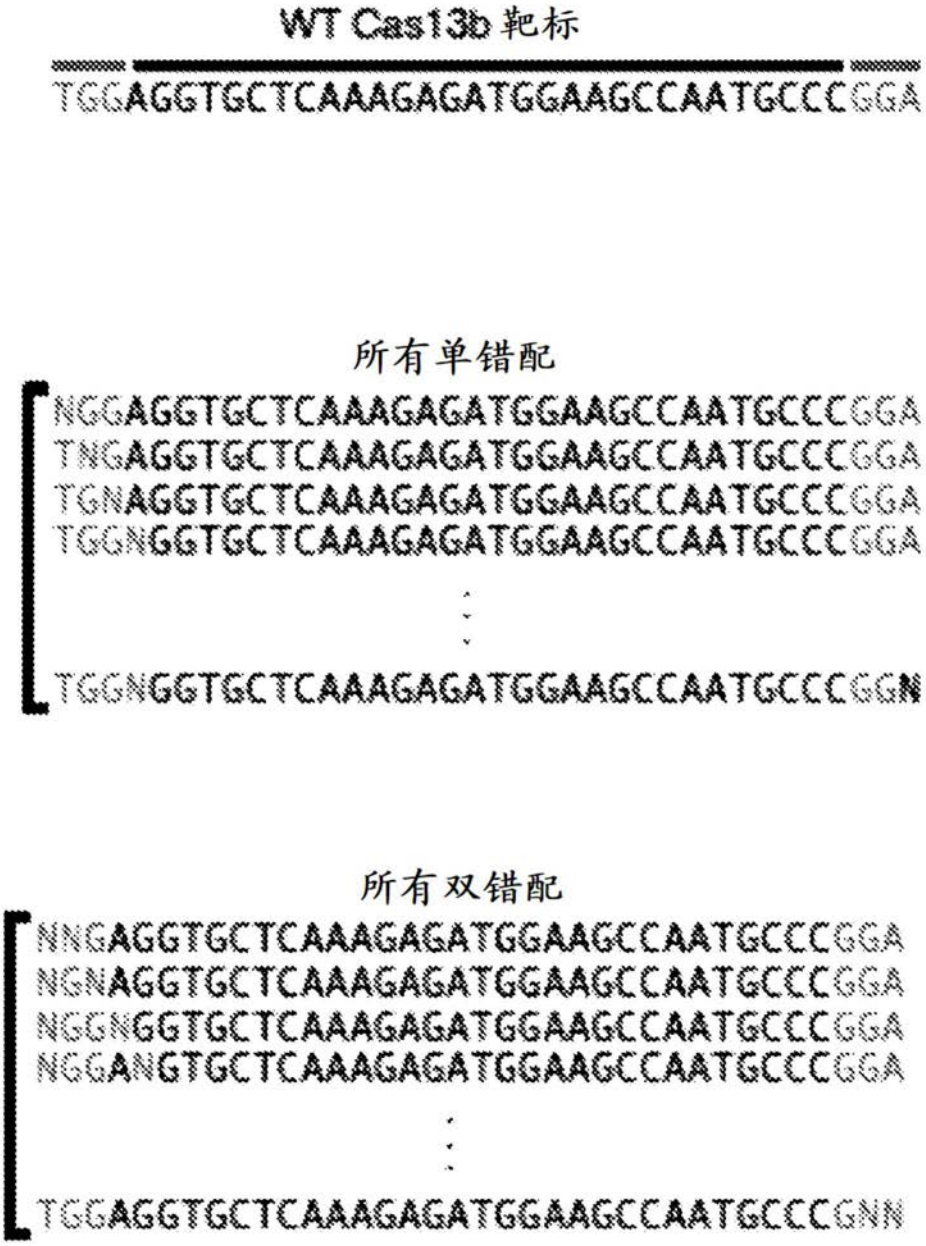

WT Cas13b 靶标
TGGAGGTGCTCAAAGAGATGGAAGCCAATGCCCGGA
所有单错配
NGGAGGTGCTCAAAGAGATGGAAGCCAATGCCCGGA
TNGAGGTGCTCAAAGAGATGGAAGCCAATGCCCGGA
TGNAGGTGCTCAAAGAGATGGAAGCCAATGCCCGGA
TGGNGGTGCTCAAAGAGATGGAAGCCAATGCCCGGA
TGGNGGTGCTCAAAGAGATGGAAGCCAATGCCCGGN
所有双错配
NNGAGGTGCTCAAAGAGATGGAAGCCAATGCCCGGA
NGNAGGTGCTCAAAGAGATGGAAGCCAATGCCCGGA
NGGNGGTGCTCAAAGAGATGGAAGCCAATGCCCGGA
NGGANGTGCTCAAAGAGATGGAAGCCAATGCCCGGA
TGGAGGTGCTCAAAGAGATGGAAGCCAATGCCCGNN

图21A

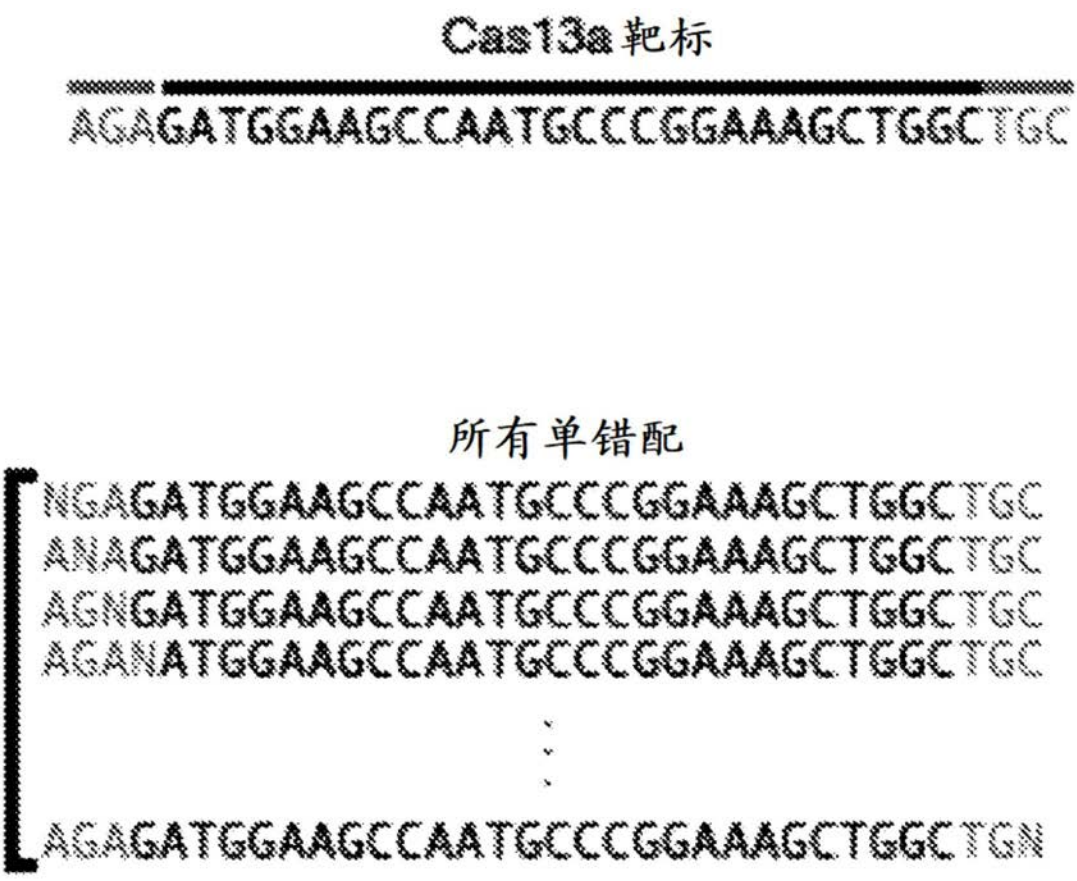

Cas13a靶标
AGAGATGGAAGCCAATGCCCGGAAAGCTGGCTGC
所有单错配
NGAGATGGAAGCCAATGCCCGGAAAGCTGGCTGC
ANAGATGGAAGCCAATGCCCGGAAAGCTGGCTGC
AGNGATGGAAGCCAATGCCCGGAAAGCTGGCTGC
AGANATGGAAGCCAATGCCCGGAAAGCTGGCTGC
AGAGATGGAAGCCAATGCCCGGAAAGCTGGCTGN
所有双错配
NNAGATGGAAGCCAATGCCCGGAAAGCTGGCTGC
NGNGATGGAAGCCAATGCCCGGAAAGCTGGCTGC
NGANATGGAAGCCAATGCCCGGAAAGCTGGCTGC
NGAGNTGGAAGCCAATGCCCGGAAAGCTGGCTGC
AGAGATGGAAGCCAATGCCCGGAAAGCTGGCTNN

图21A(续)

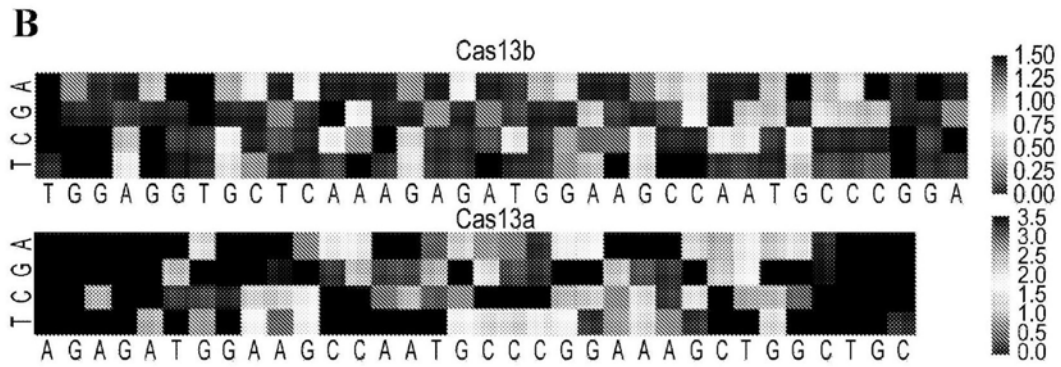

B
Cas13b
T C G A
T G G A G G T G C T C A A A G A G A T G G A A G C C A A T G C C C G G A
1.50
1.25
1.00
0.75
0.50
0.25
0.00
Cas13a
T C G A
A G A G A T G G A A G C C A A T G C C C G G A A A G C T G G C T G C
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0

图21B

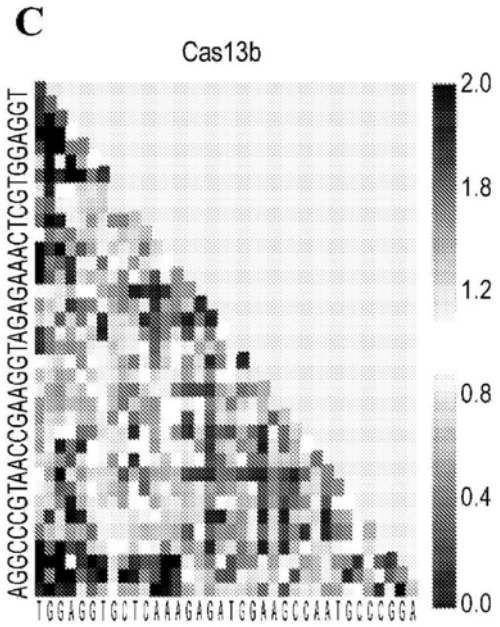

C
Cas13b
AGGCCCCGTAACCGAAGGTAGAGAAACTCGTGGAGGT
TGGAGGTGCTCAAAGAGATGGAAGCCAATGCCCGGA
2.0
1.8
1.2
0.8
0.4
0.0

图21C

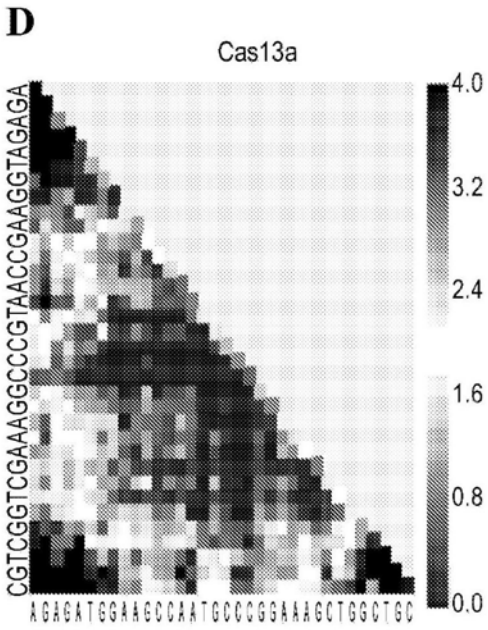

D
Cas13a
4.0
3.2
2.4
1.6
0.8
0.0
CGTCGGTCGAAAGGCCCGTAACCGAAGGTAGAGA
AGAGATGGAAAGCCAATGCCCGGAAAGCTGGCTGC

图21D

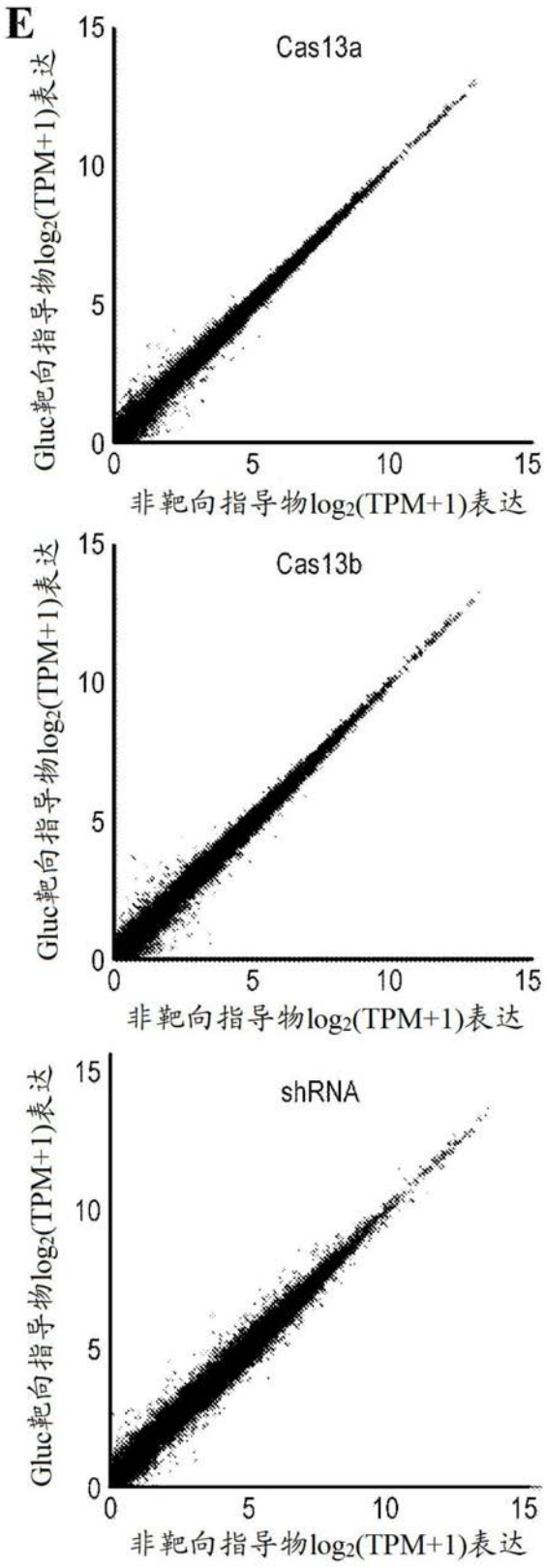

E
Cas13a
Gluc靶向指导物log2(TPM+1)表达
非靶向指导物log2(TPM+1)表达
Cas13b
Gluc靶向指导物log2(TPM+1)表达
非靶向指导物log2(TPM+1)表达
shRNA
Gluc靶向指导物log2(TPM+1)表达
非靶向指导物log2(TPM+1)表达
0
5
10
15

图21E

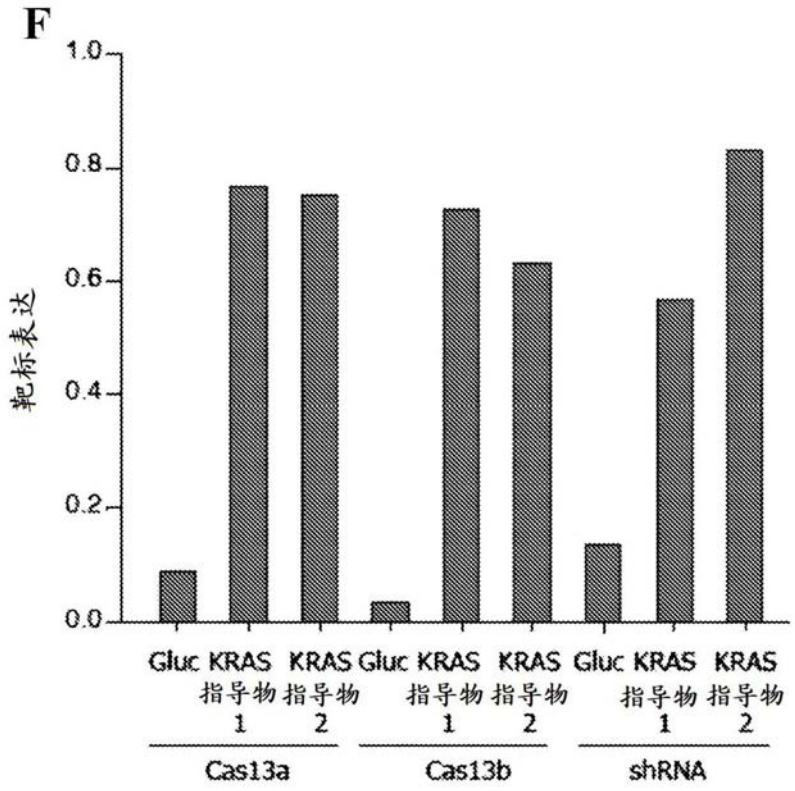

F
1.0
0.8
0.6
0.4
0.2
0.0
靶标表达
Gluc
KRAS 指导物 1
KRAS 指导物 2
Cas13a
Cas13b
shRNA

图21F

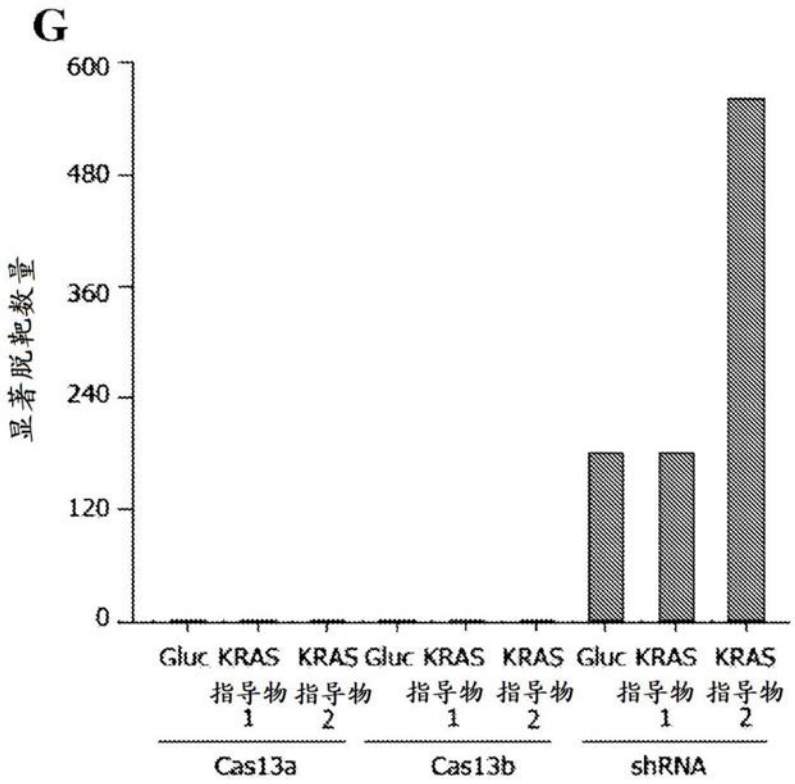

G
600
480
360
240
120
0
显著脱靶数量
Gluc
KRAS 指导物 1
KRAS 指导物 2
Cas13a
Cas13b
shRNA

图21G

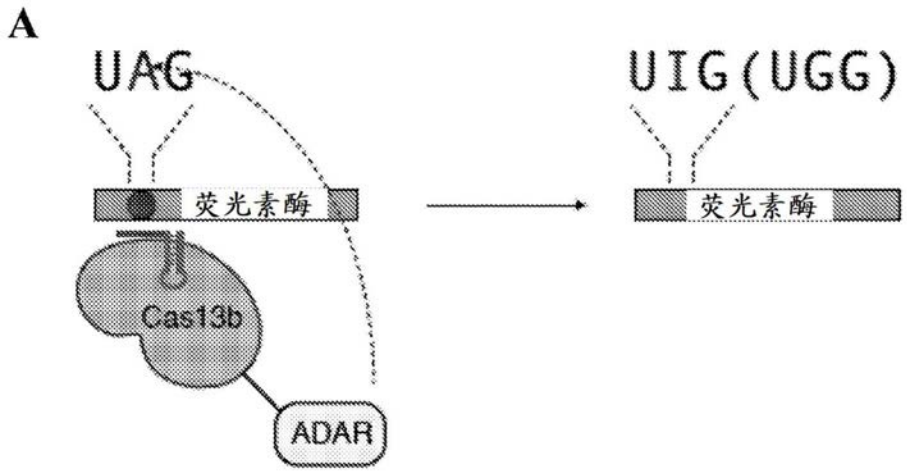

A
UAG
UIG(UGG)
荧光素酶
荧光素酶
Cas13b
ADAR

图22A

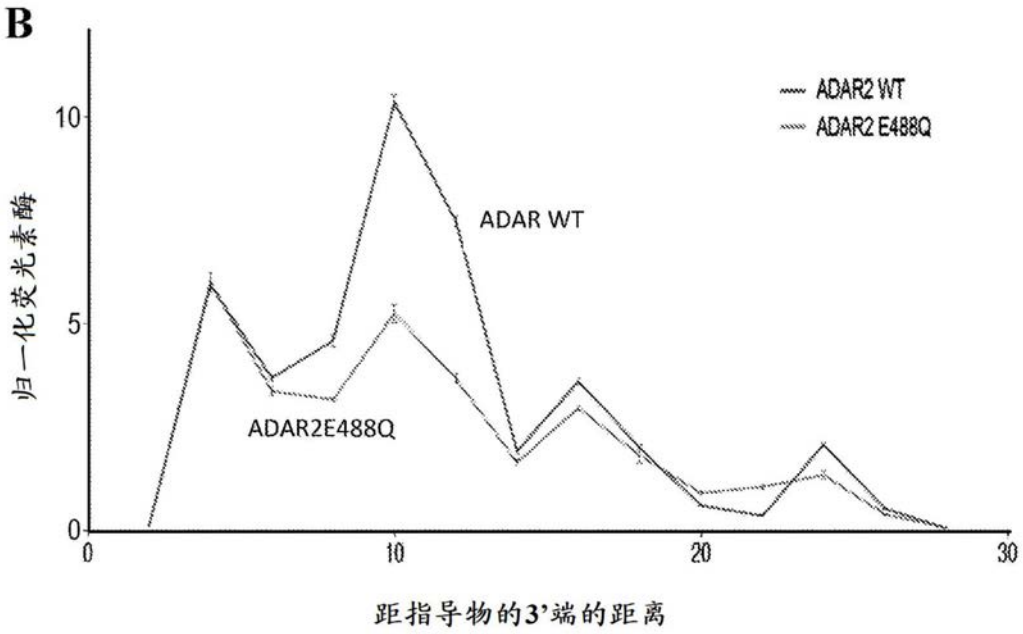

B
ADAR2 WT
ADAR2 E488Q
ADAR WT
ADAR2E488Q
归一化荧光素酶
距指导物的3'端的距离
0
5
10
0
10
20
30

图22B

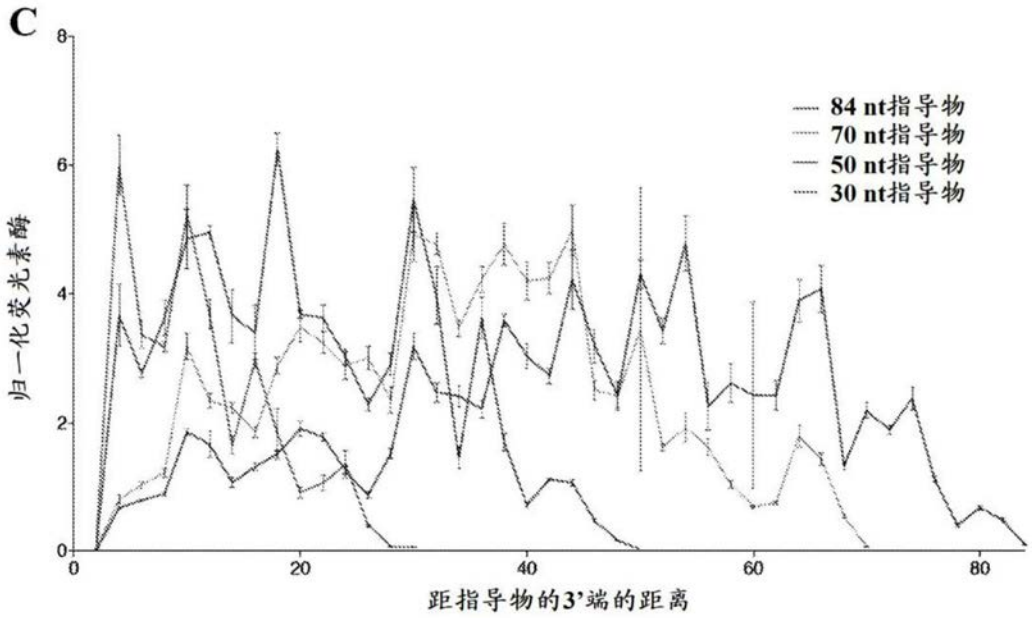

C
归一化荧光素酶
距指导物的3'端的距离
84 nt指导物
70 nt指导物
50 nt指导物
30 nt指导物
0
2
4
6
8
0
20
40
60
80

图22C

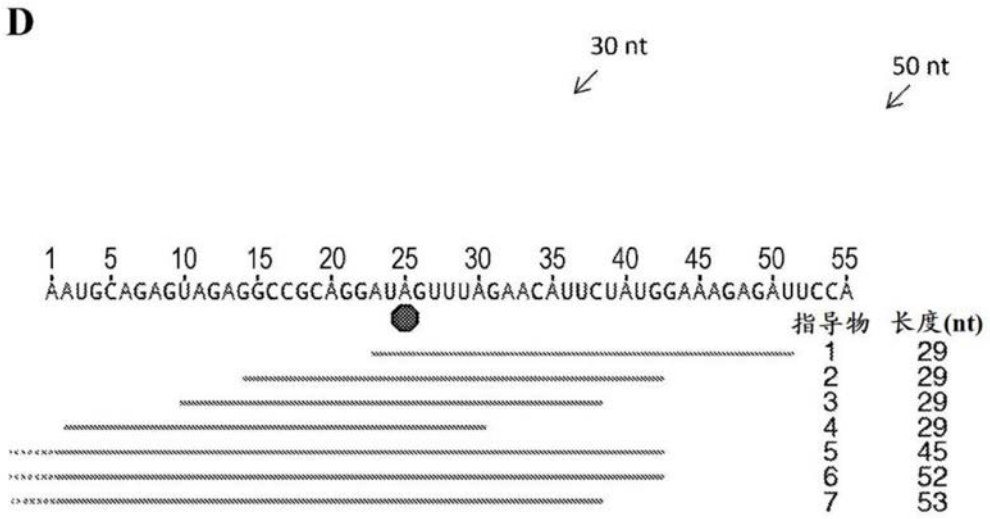

D
30 nt
50 nt
1 5 10 15 20 25 30 35 40 45 50 55
AAUGCAGAGUAGAGGCCGCAGGAUAGUUUAGAACAUUCUAUGGAAAGAGAUUCCA
指导物 长度(nt)
1 29
2 29
3 29
4 29
5 45
6 52
7 53

图22D

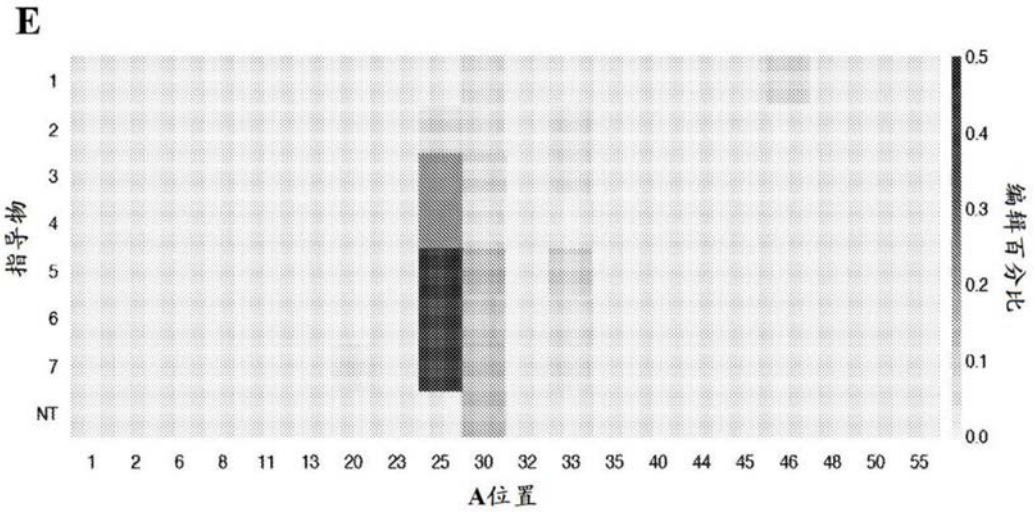

E
指导物
1
2
3
4
5
6
7
NT
1 2 6 8 11 13 20 23 25 30 32 33 35 40 44 45 46 48 50 55
A位置
0.5
0.4
0.3
0.2
0.1
0.0
编辑百分比

图22E

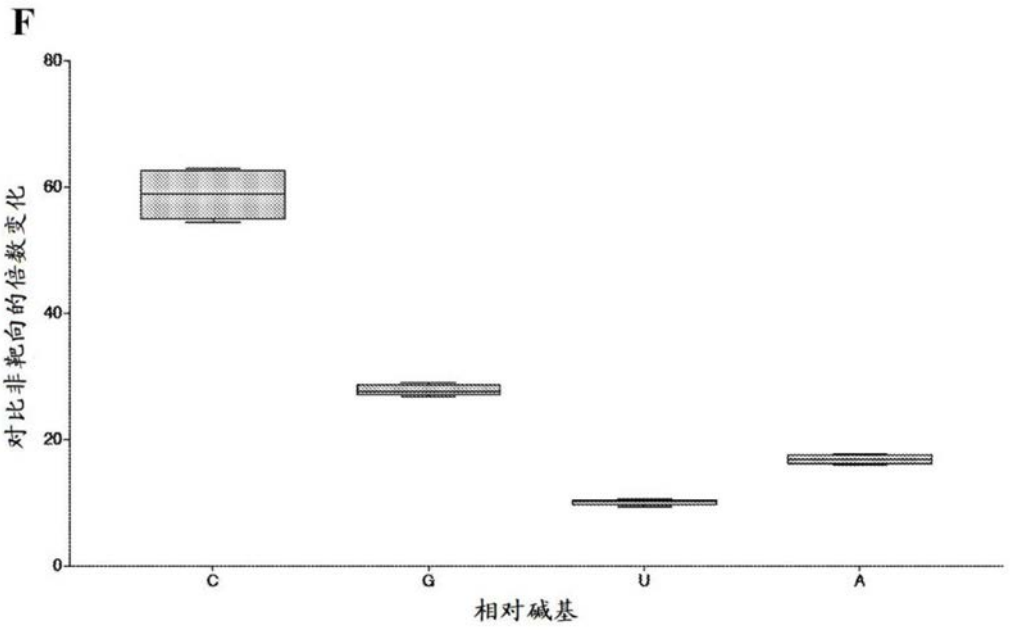

F
对比非靶向的倍数变化
80
60
40
20
0
C
G
U
A
相对碱基

图22F

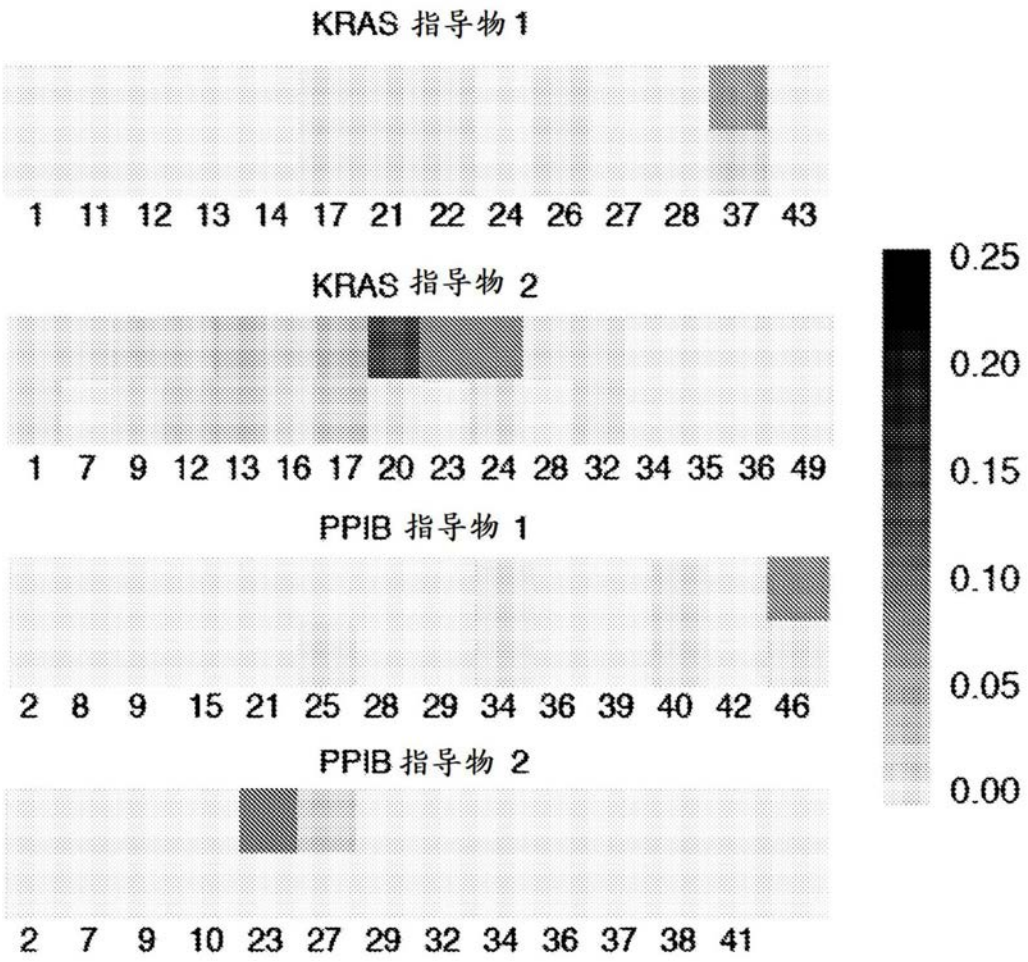

KRAS 指导物 1
1 11 12 13 14 17 21 22 24 26 27 28 37 43
KRAS 指导物 2
1 7 9 12 13 16 17 20 23 24 28 32 34 35 36 49
PPIB 指导物 1
2 8 9 15 21 25 28 29 34 36 39 40 42 46
PPIB 指导物 2
2 7 9 10 23 27 29 32 34 36 37 38 41
0.25
0.20
0.15
0.10
0.05
0.00

图23

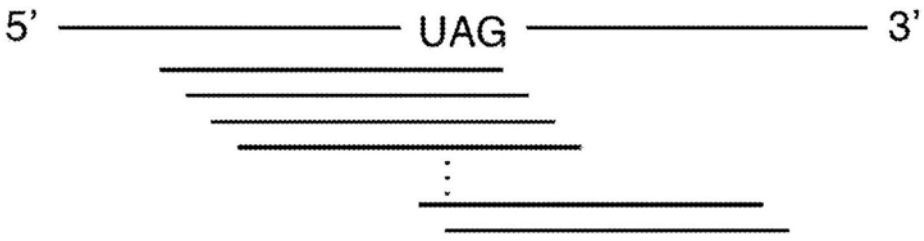

5’
UAG
3’

图24

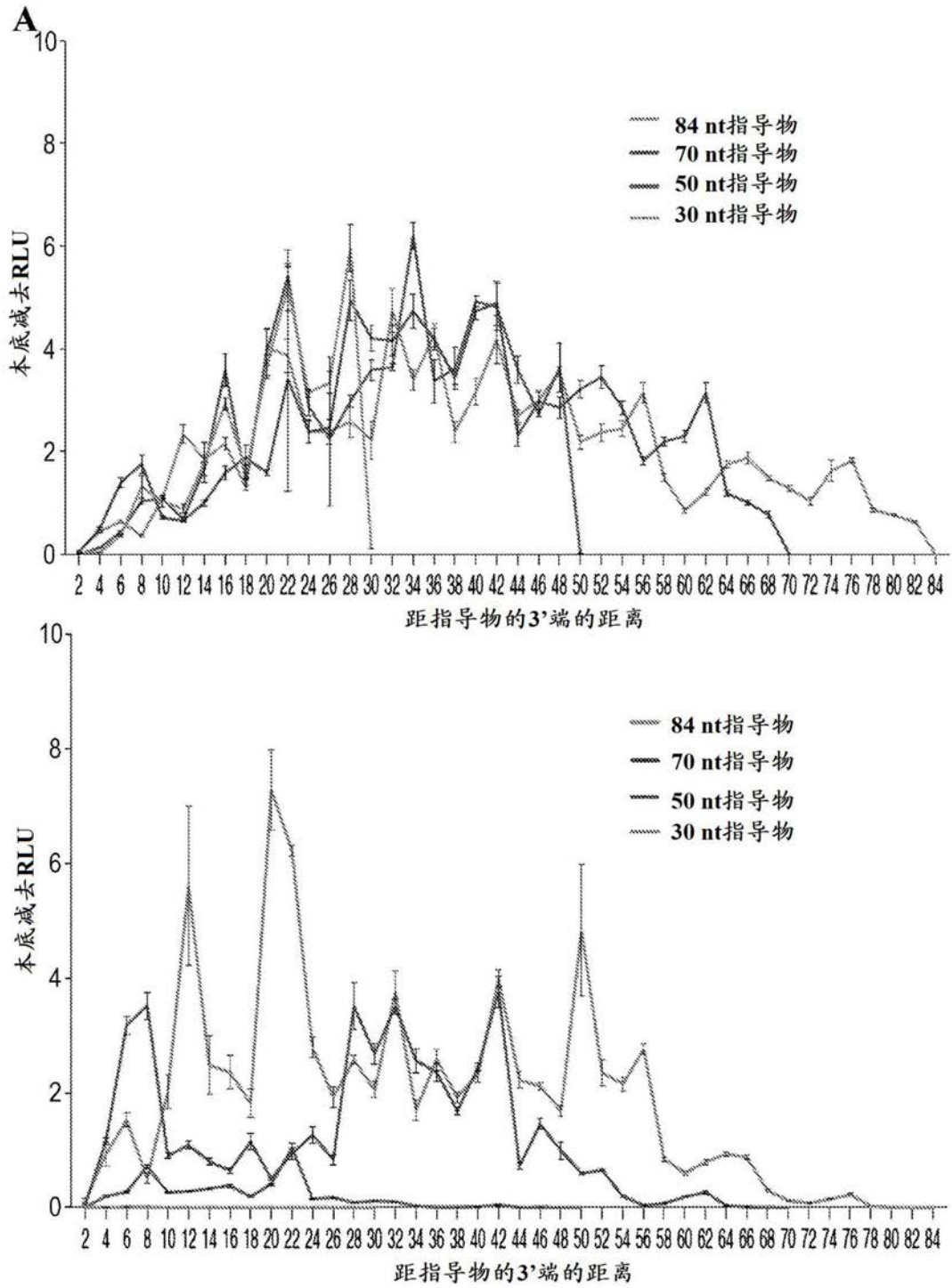

A
本底减去RLU
84 nt指导物
70 nt指导物
50 nt指导物
30 nt指导物
距指导物的3'端的距离
本底减去RLU
84 nt指导物
70 nt指导物
50 nt指导物
30 nt指导物
距指导物的3'端的距离

图25A

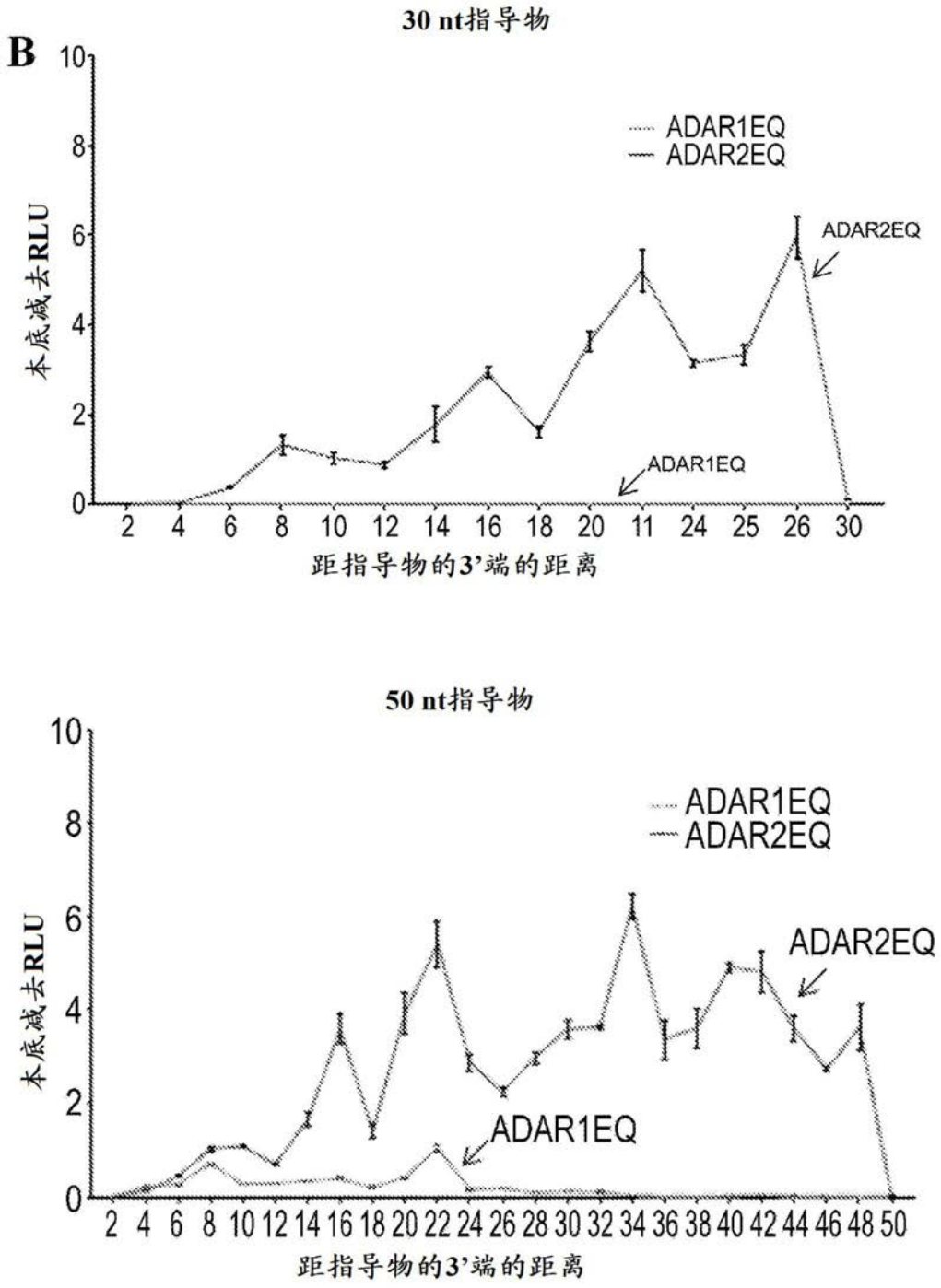

30 nt指导物
B
本底减去RLU
ADAR1EQ
ADAR2EQ
ADAR2EQ
ADAR1EQ
距指导物的3'端的距离
50 nt指导物
本底减去RLU
ADAR1EQ
ADAR2EQ
ADAR2EQ
ADAR1EQ
距指导物的3'端的距离

图25B

C

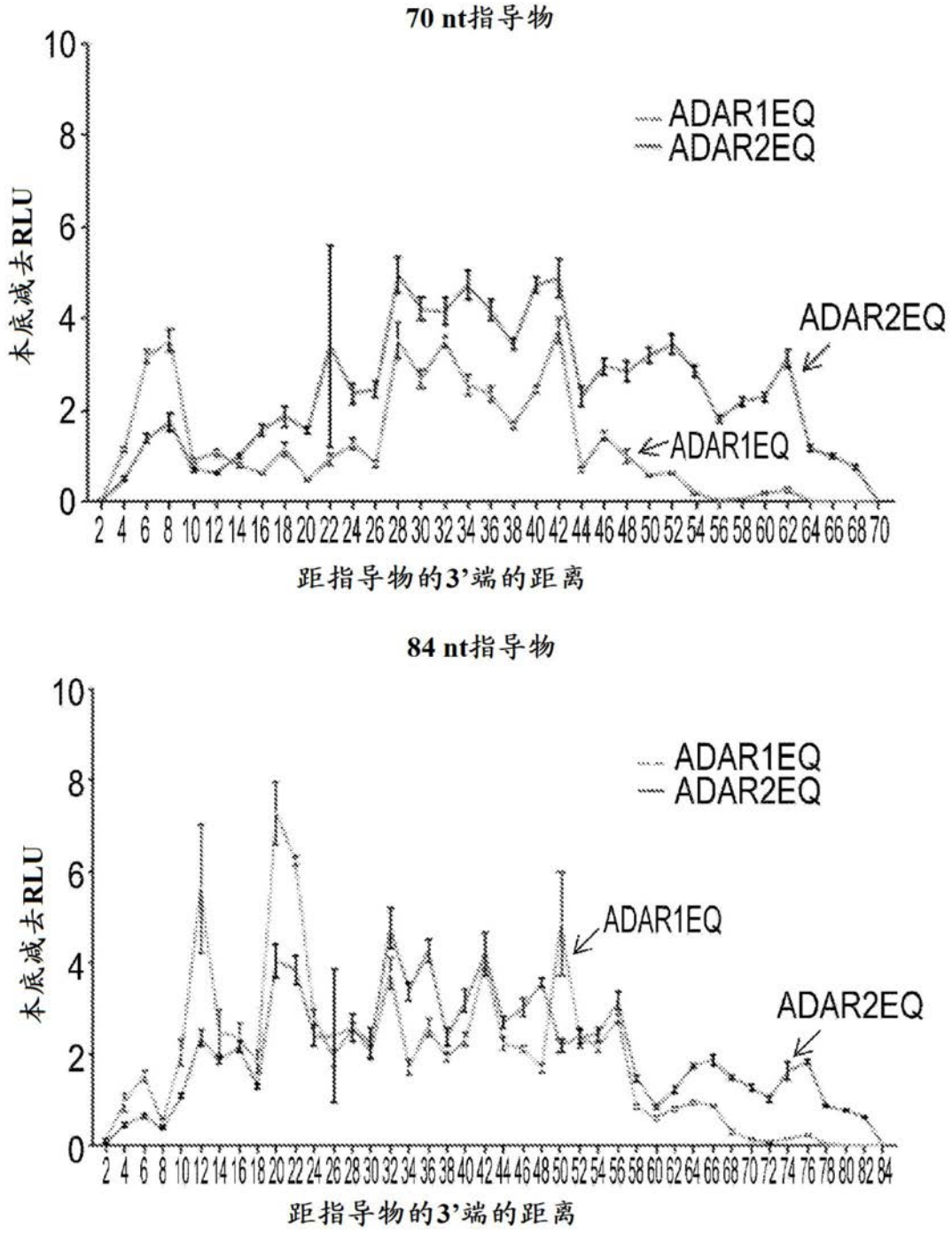

70 nt指导物
本底减去RLU
ADAR1EQ
ADAR2EQ
ADAR2EQ
ADAR1EQ
距指导物的3'端的距离
84 nt指导物
本底减去RLU
ADAR1EQ
ADAR2EQ
ADAR1EQ
ADAR2EQ
距指导物的3'端的距离

图25C

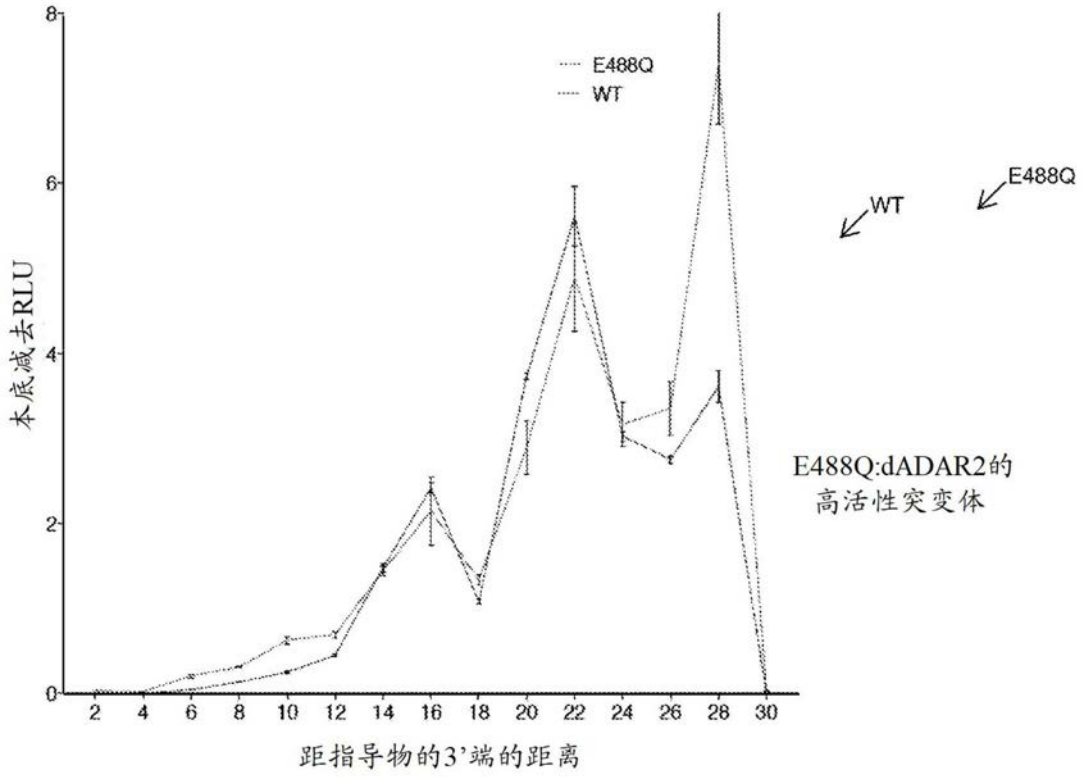

E488Q
WT
WT
E488Q
E488Q:dADAR2的
高活性突变体
本底减去RLU
距指导物的3'端的距离
0
2
4
6
8
2
4
6
8
10
12
14
16
18
20
22
24
26
28
30

图26

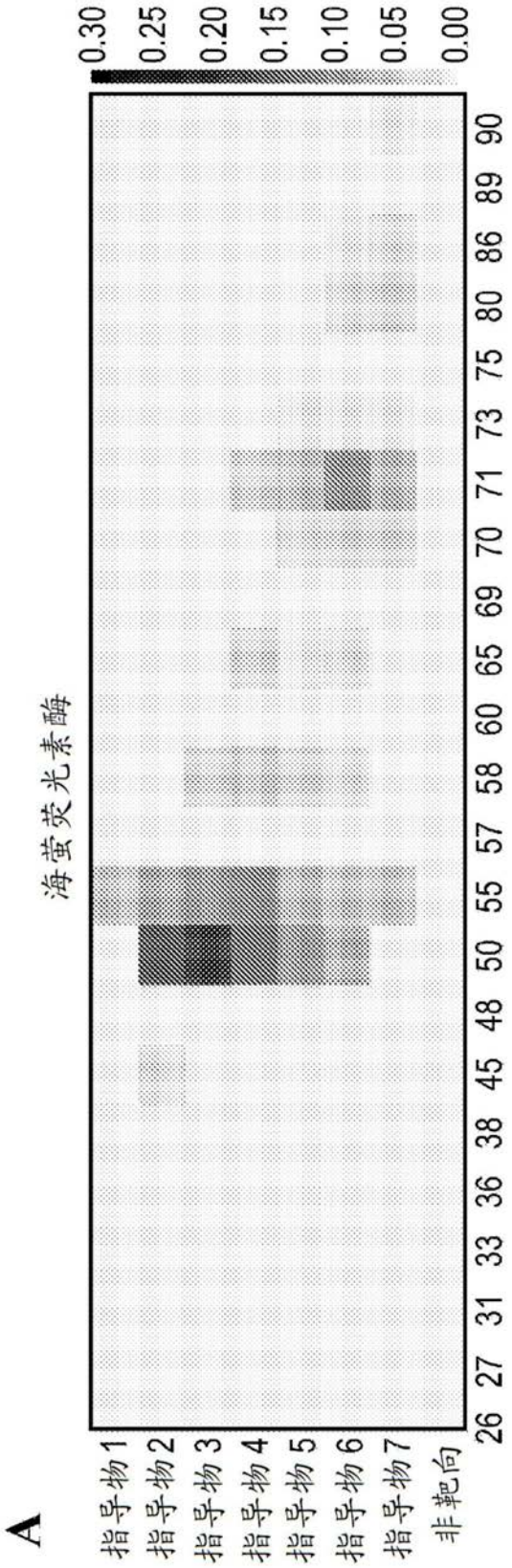

A
海萤荧光素酶
0.30
0.25
0.20
0.15
0.10
0.05
0.00
26 27 31 33 36 38 45 48 50 55 57 58 60 65 69 70 71 73 75 80 86 89 90
指导物1
指导物2
指导物3
指导物4
指导物5
指导物6
指导物7
非靶向

图27A

**B**

| 指导物 | 3’起始位置 | 5’起始位置 | A-C位置 |
|---|---|---|---|
| 1 | 1 | 50 | 50 |
| 2 | 9 | 59 | 50 |
| 3 | 17 | 67 | 50 |
| 4 | 25 | 75 | 50 |
| 5 | 33 | 83 | 50 |
| 6 | 41 | 90 | 50 |
| 7 | 49 | 99 | 50 |
| NT | | | |

图27B

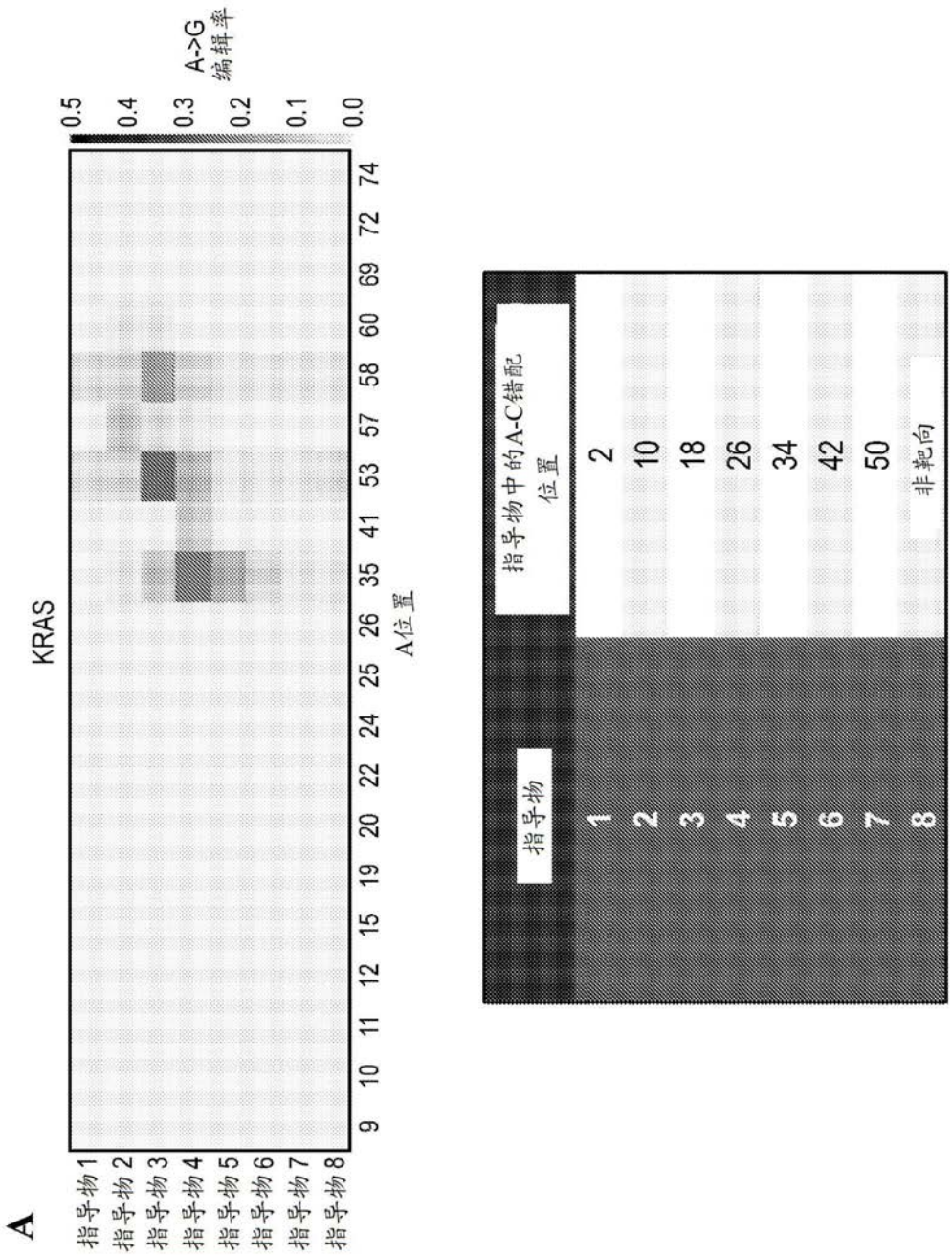

A
KRAS
A->G
编辑率
0.5
0.4
0.3
0.2
0.1
0.0
指导物1
指导物2
指导物3
指导物4
指导物5
指导物6
指导物7
指导物8
9 10 11 12 15 19 20 22 24 25 26 35 41 53 57 58 60 69 72 74
A位置
指导物
指导物中的A-C错配位置
1 2
2 10
3 18
4 26
5 34
6 42
7 50
8 非靶向

图28A

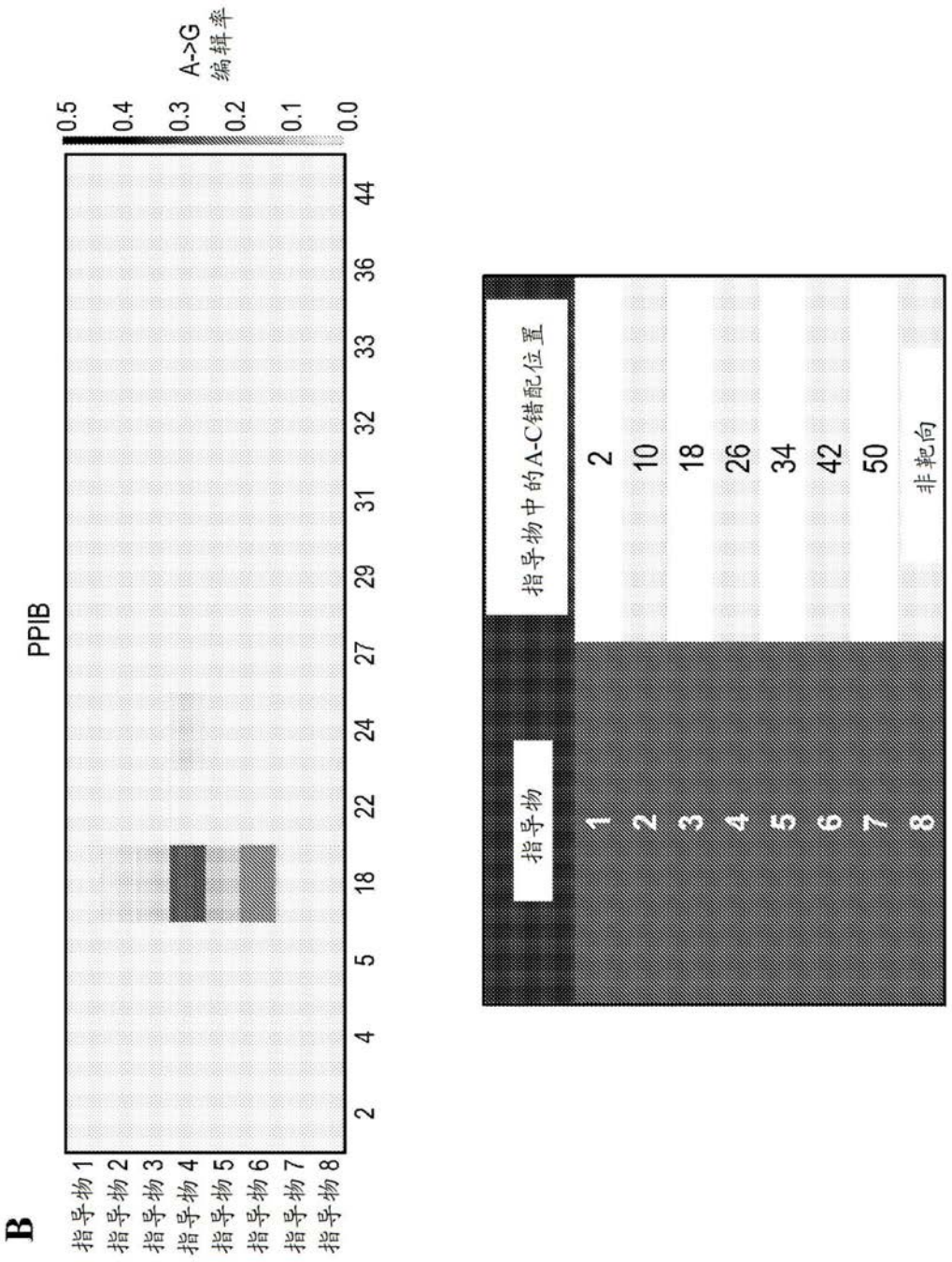

B
PPIB
A->G
编辑率
0.5
0.4
0.3
0.2
0.1
0.0
2
4
5
18
22
24
27
29
31
32
33
36
44
指导物 1
指导物 2
指导物 3
指导物 4
指导物 5
指导物 6
指导物 7
指导物 8
指导物
指导物中的A-C错配位置
1
2
3
4
5
6
7
8
2
10
18
26
34
42
50
非靶向

图28B

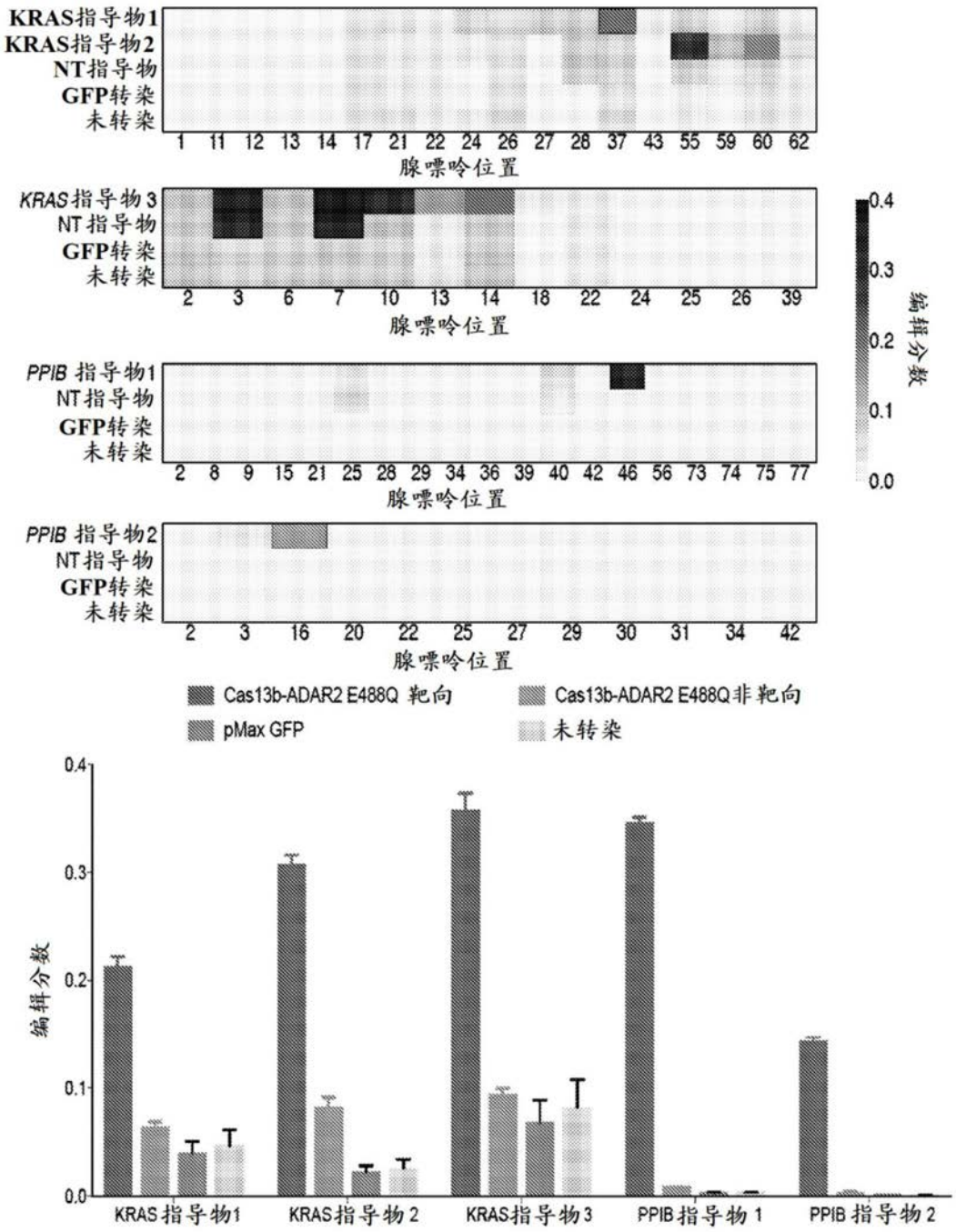

KRAS指导物1
KRAS指导物2
NT指导物
GFP转染
未转染
1 11 12 13 14 17 21 22 24 26 27 28 37 43 55 59 60 62
腺嘌呤位置
KRAS指导物3
NT指导物
GFP转染
未转染
2 3 6 7 10 13 14 18 22 24 25 26 39
腺嘌呤位置
PPIB 指导物1
NT指导物
GFP转染
未转染
2 8 9 15 21 25 28 29 34 36 39 40 42 46 56 73 74 75 77
腺嘌呤位置
PPIB 指导物2
NT指导物
GFP转染
未转染
2 3 16 20 22 25 27 29 30 31 34 42
腺嘌呤位置
0.4
0.3
0.2
0.1
0.0
编辑分数
Cas13b-ADAR2 E488Q 靶向
Cas13b-ADAR2 E488Q非靶向
pMax GFP
未转染
编辑分数
0.4
0.3
0.2
0.1
0.0
KRAS 指导物1
KRAS指导物 2
KRAS指导物3
PPIB指导物 1
PPIB 指导物 2

图29

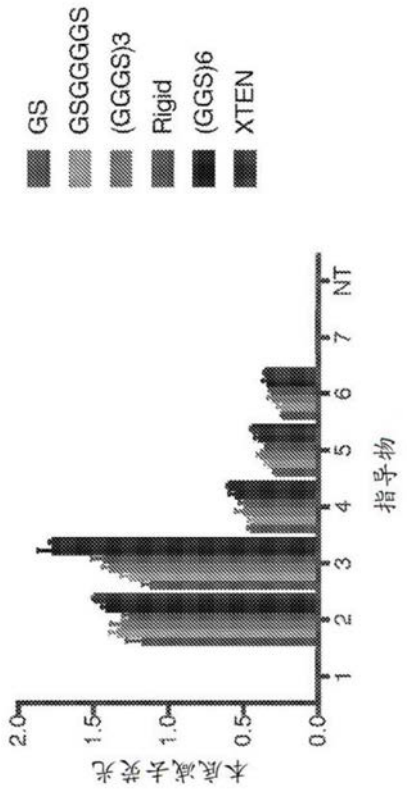



图30

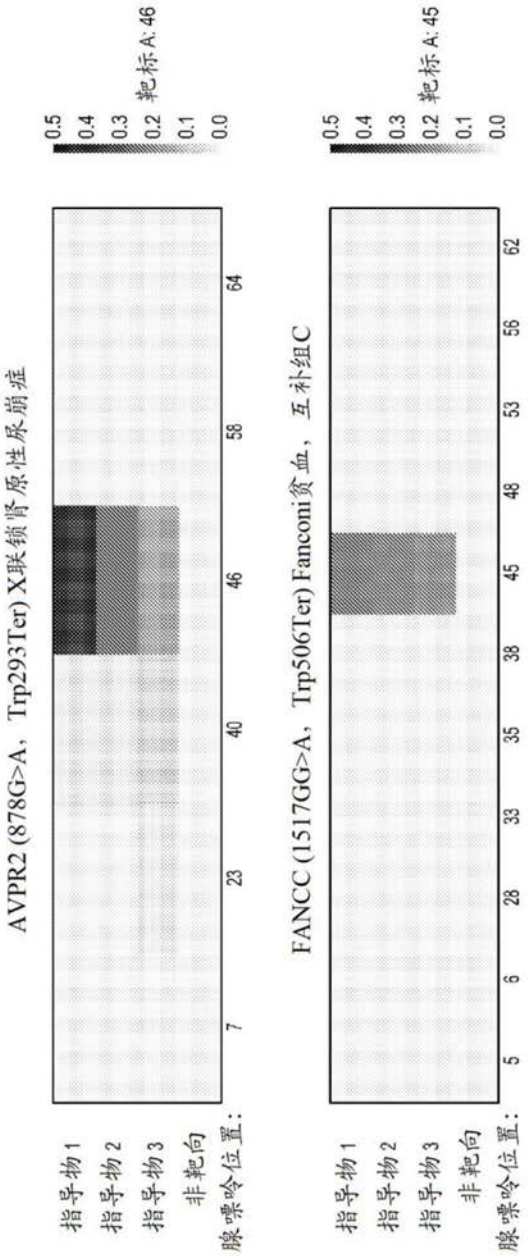

AVPR2 (878G>A，Trp293Ter) X联锁肾原性尿崩症
0.5
0.4
0.3
0.2
0.1
0.0
靶标 A: 46
指导物 1
指导物 2
指导物 3
非靶向
脱嘌呤位置：
7
23
40
46
58
64
FANCC (1517GG>A，Trp506Ter) Fanconi贫血，互补组C
0.5
0.4
0.3
0.2
0.1
0.0
靶标 A: 45
指导物 1
指导物 2
指导物 3
非靶向
脱嘌呤位置：
5
6
28
33
35
38
45
48
53
56
62

图31

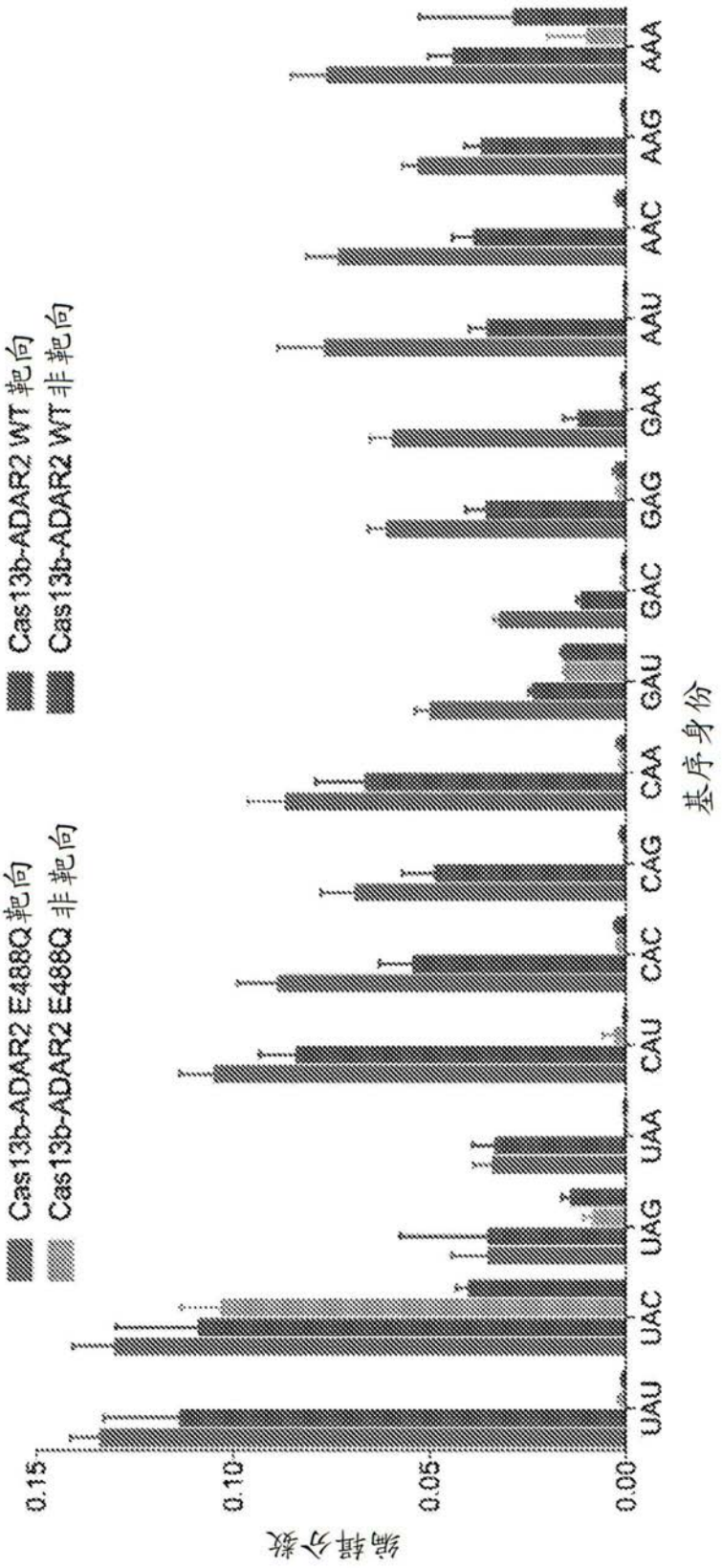

Cas13b-ADAR2 E488Q 靶向
Cas13b-ADAR2 E488Q 非靶向
Cas13b-ADAR2 WT 靶向
Cas13b-ADAR2 WT 非靶向
编辑分数
0.15
0.10
0.05
0.00
UAU UAC UAG UAA CAU CAC CAG CAA GAU GAC GAG GAA AAU AAC AAG AAA
基序身份

图32

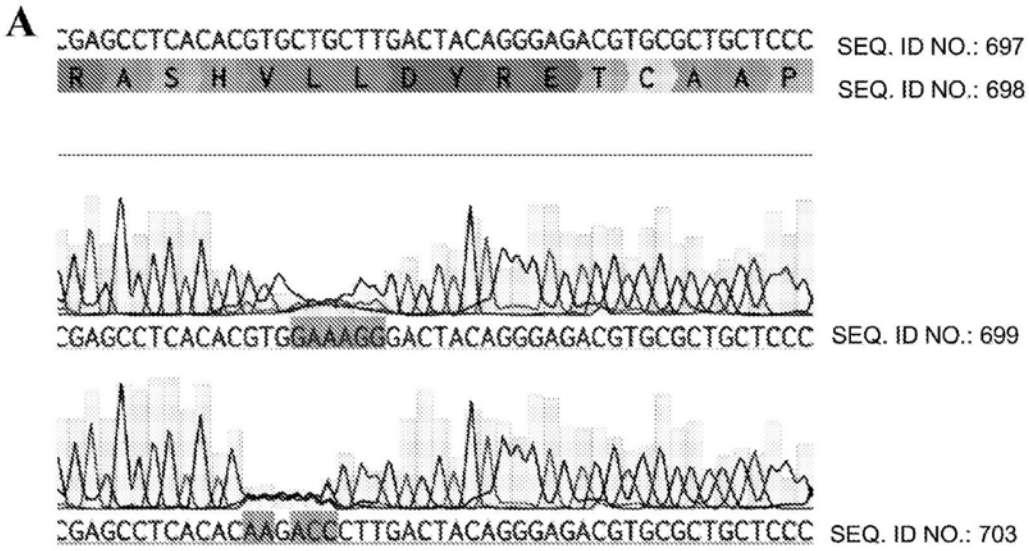

A
CGAGCCTCACACGTGCTGCTTGACTACAGGGAGACGTGCGCTGCTCCC
R A S H V L L D Y R E T C A A P
SEQ. ID NO.: 697
SEQ. ID NO.: 698
CGAGCCTCACACGTGGAAAGGGACTACAGGGAGACGTGCGCTGCTCCC
SEQ. ID NO.: 699
CGAGCCTCACACAAGACCCTTGACTACAGGGAGACGTGCGCTGCTCCC
SEQ. ID NO.: 703

图33A

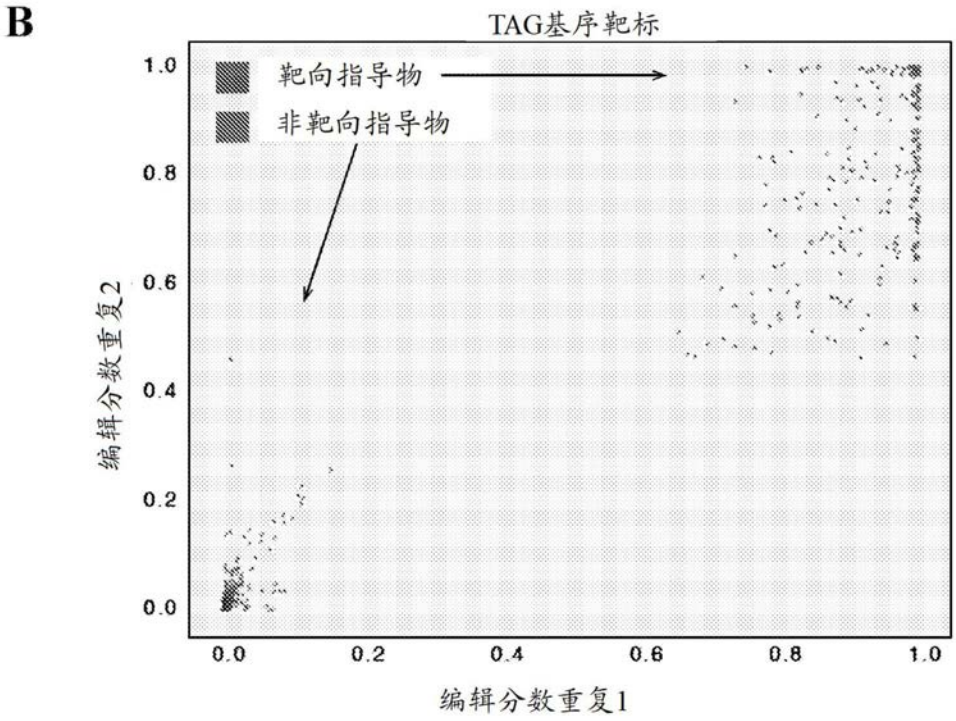

B
TAG基序靶标
靶向指导物
非靶向指导物
编辑分数重复2
编辑分数重复1
0.0
0.2
0.4
0.6
0.8
1.0

图33B

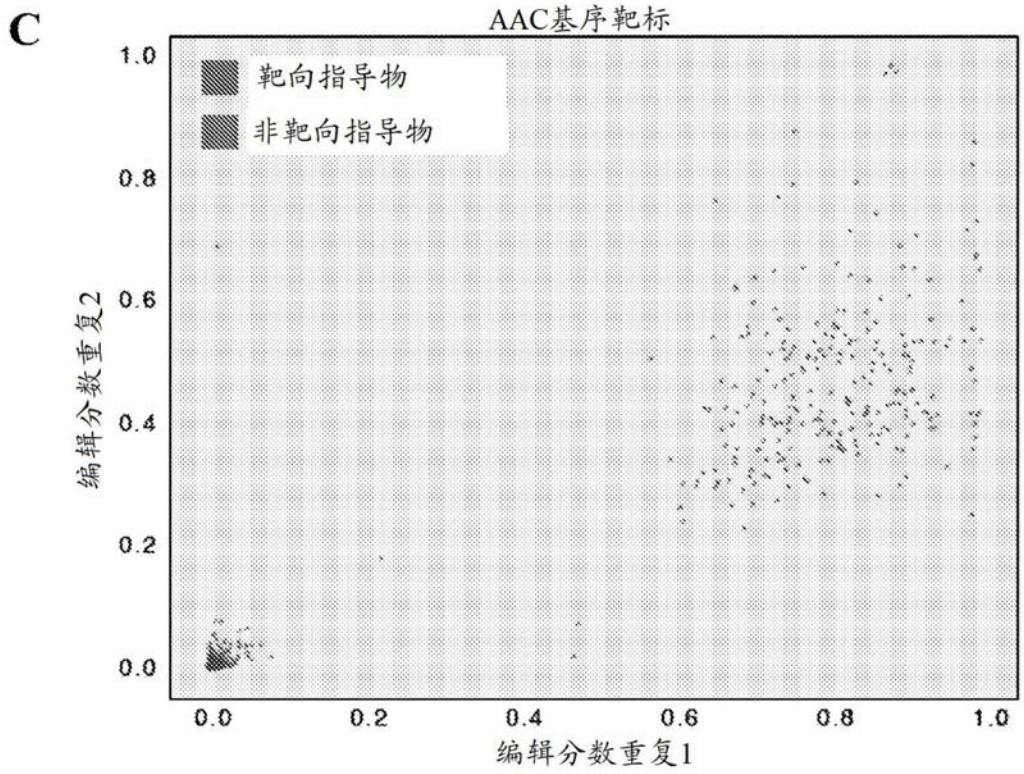

C
AAC基序靶标
靶向指导物
非靶向指导物
编辑分数重复2
编辑分数重复1
0.0
0.2
0.4
0.6
0.8
1.0

图33C

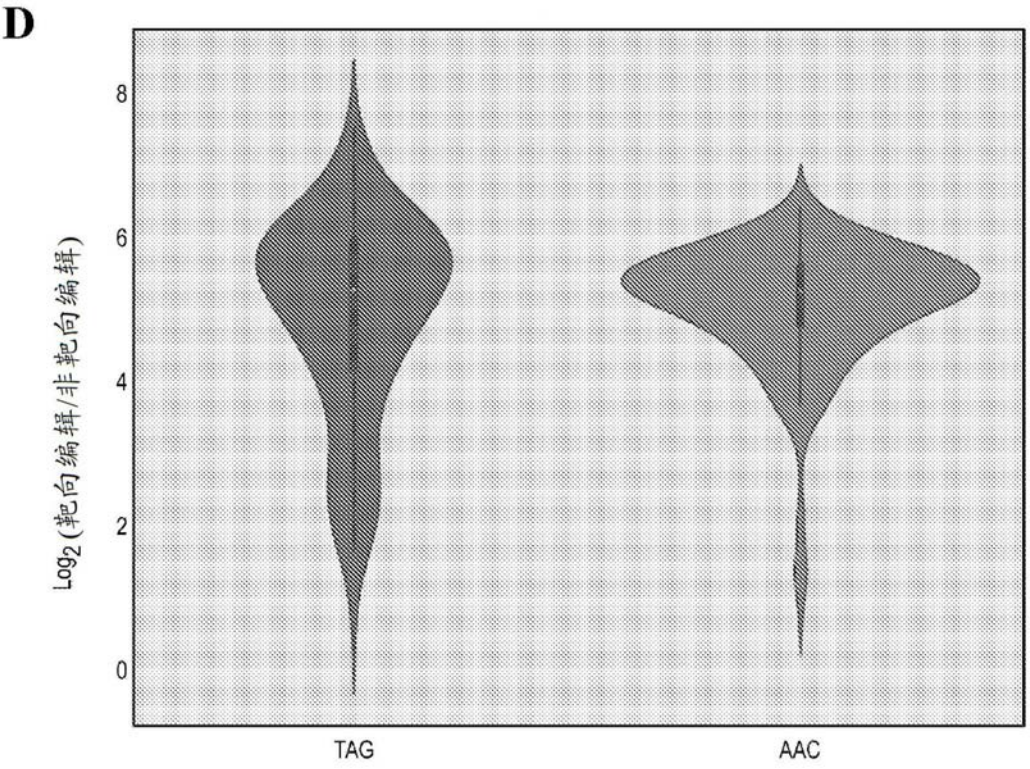

D
Log2(靶向编辑/非靶向编辑)
0
2
4
6
8
TAG
AAC

图33D

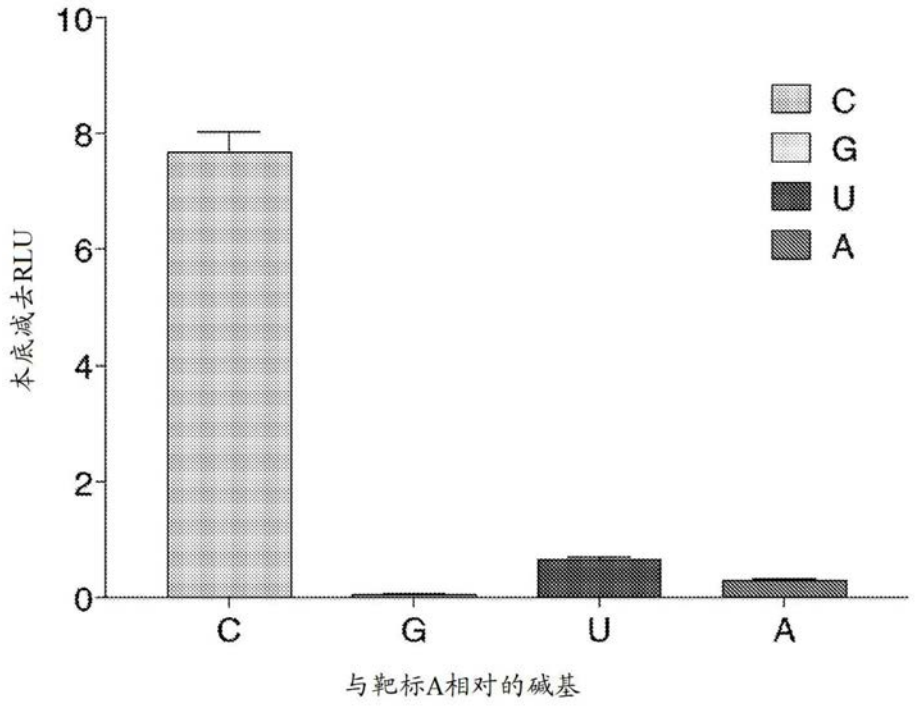

10
8
6
4
2
0
本底减去RLU
C
G
U
A
C
G
U
A
与靶标A相对的碱基

图34

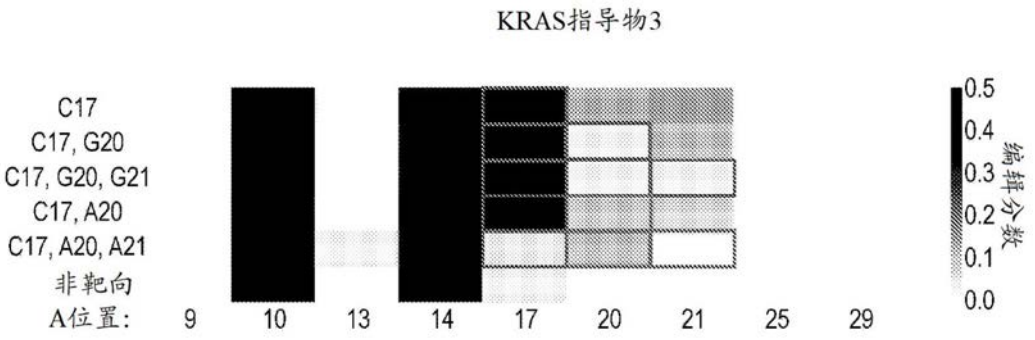

KRAS指导物3
C17
C17, G20
C17, G20, G21
C17, A20
C17, A20, A21
非靶向
A位置:
9
10
13
14
17
20
21
25
29
0.5
0.4
0.3
0.2
0.1
0.0
编辑分数

图35

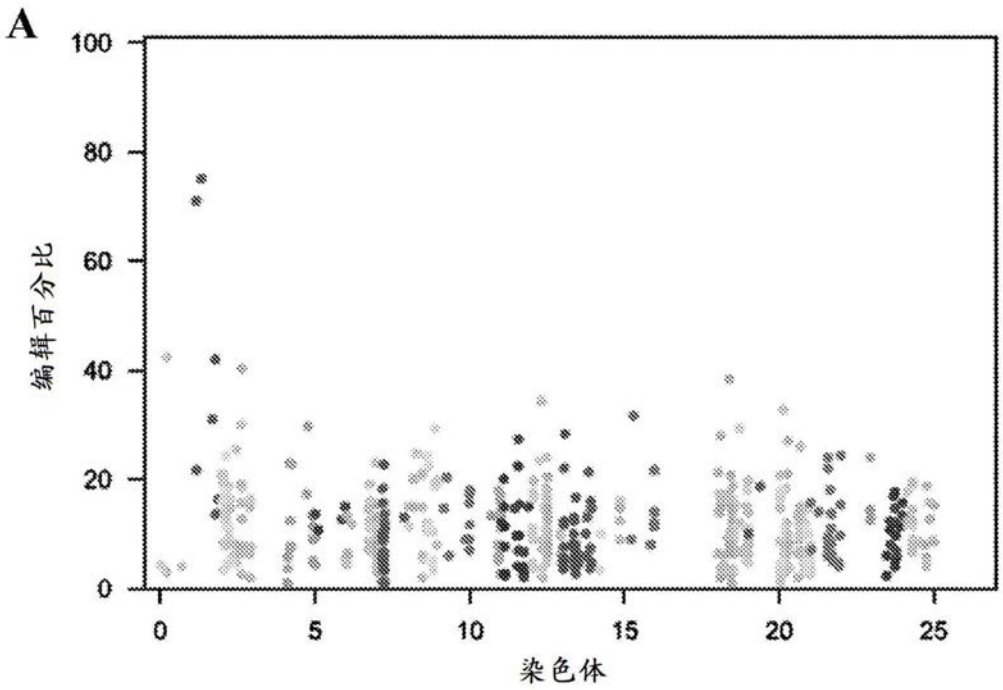

A
100
80
60
40
20
0
编辑百分比
0
5
10
15
20
25
染色体

图36A

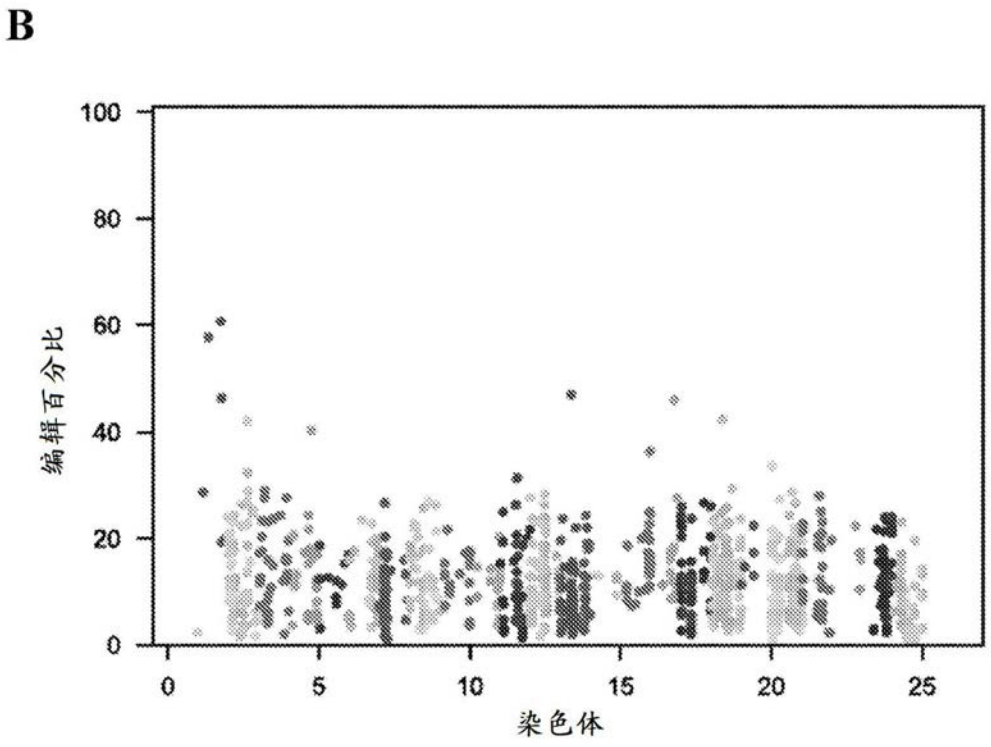

B
100
80
60
40
20
0
编辑百分比
0
5
10
15
20
25
染色体

图36B

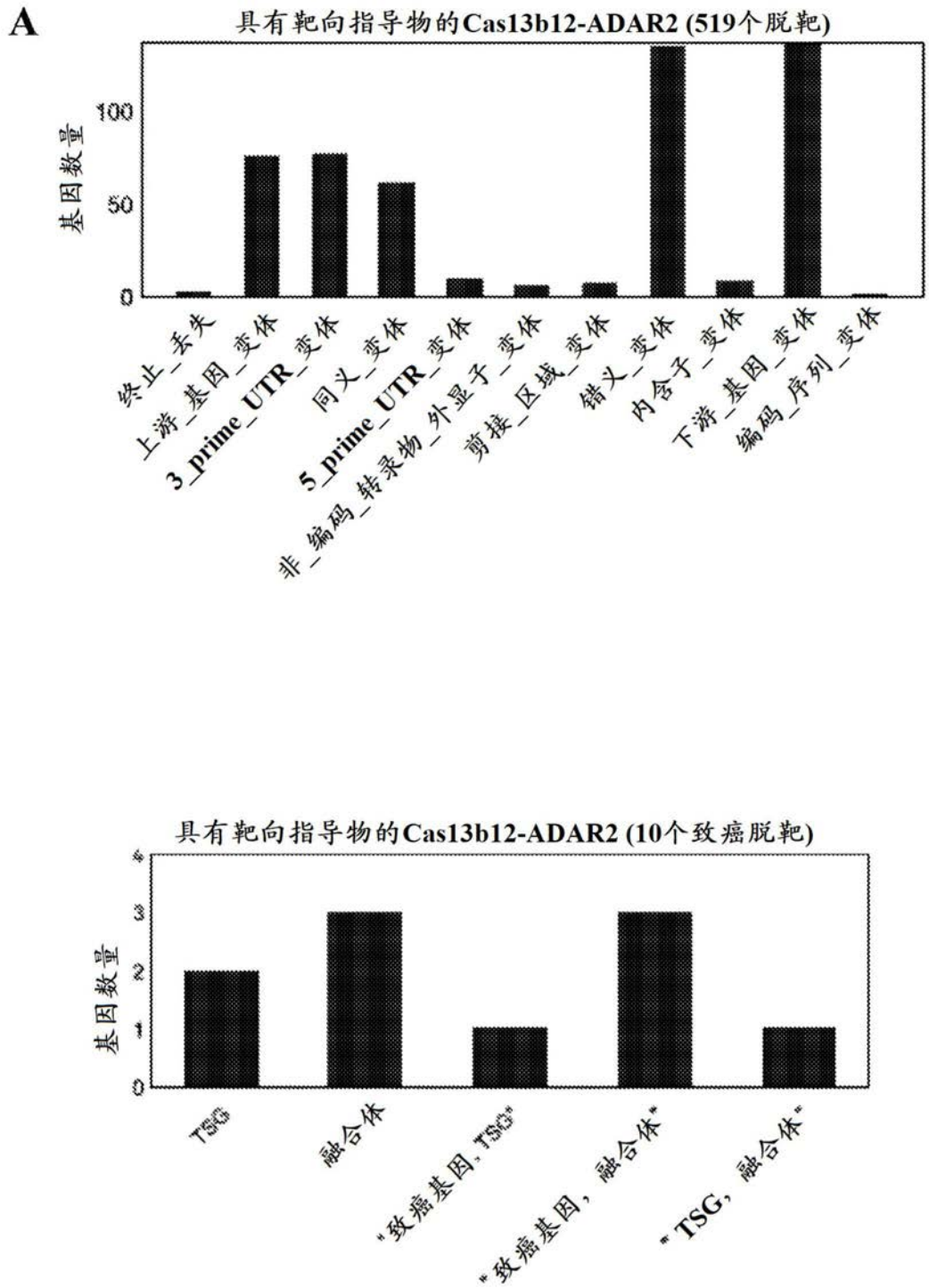

A
具有靶向指导物的Cas13b12-ADAR2 (519个脱靶)
基因数量
0
50
100
终止_丢失
上游_基因_变体
3_prime_UTR_变体
同义_变体
5_prime_UTR_变体
非_编码_转录物_外显子_变体
剪接_区域_变体
错义_变体
内含子_变体
下游_基因_变体
编码_序列_变体
具有靶向指导物的Cas13b12-ADAR2 (10个致癌脱靶)
基因数量
0
1
2
3
4
TSG
融合体
"致癌基因，TSG"
"致癌基因，融合体"
"TSG，融合体"

图37A

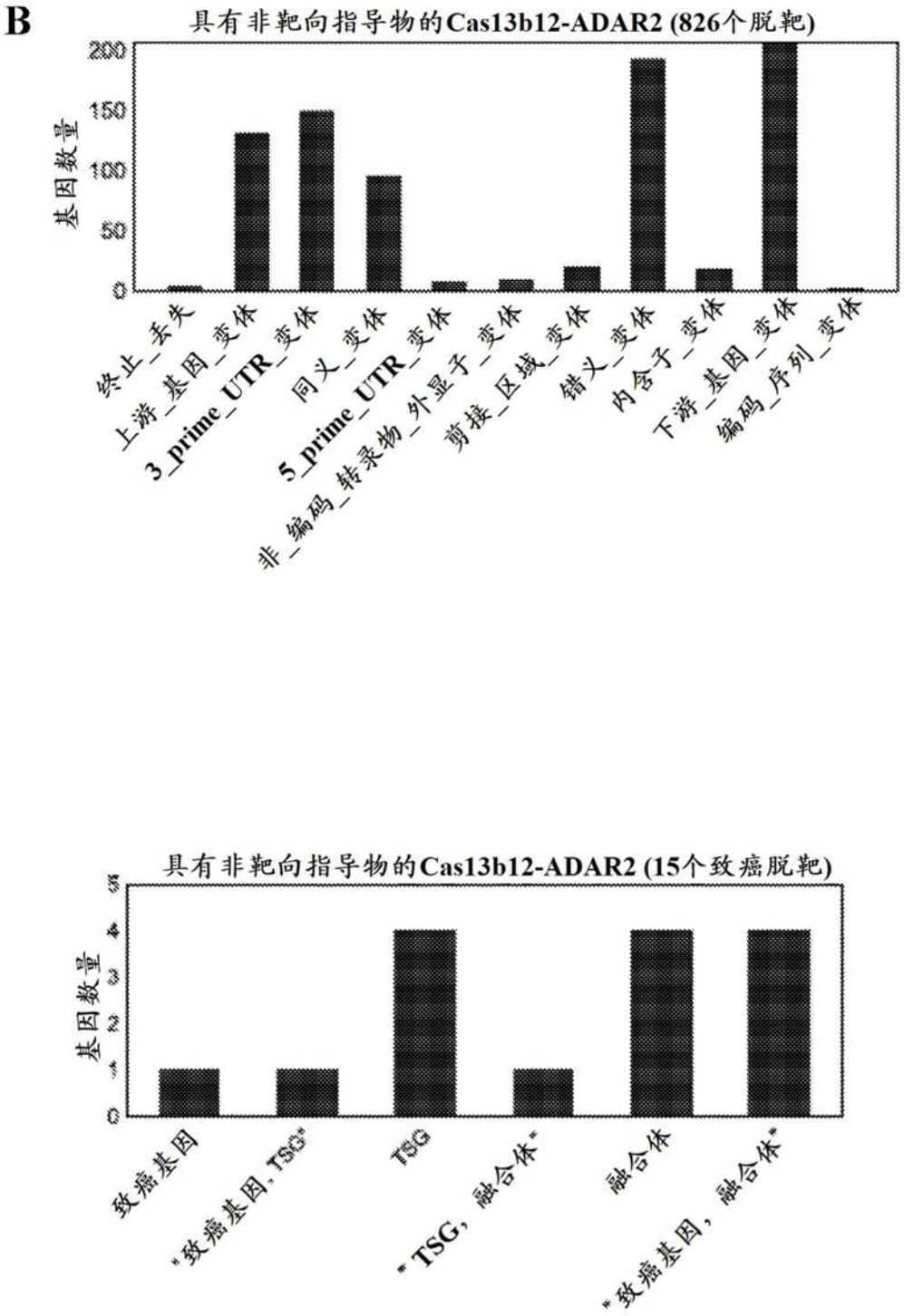

B
具有非靶向指导物的Cas13b12-ADAR2 (826个脱靶)
基因数量
0
25
50
75
100
终止_丢失
上游_基因_变体
3_prime_UTR_变体
同义_变体
5_prime_UTR_变体
非_编码_转录物_外显子_变体
剪接_区域_变体
错义_变体
内含子_变体
下游_基因_变体
编码_序列_变体
具有非靶向指导物的Cas13b12-ADAR2 (15个致癌脱靶)
基因数量
0
2
4
6
8
致癌基因
"致癌基因，TSG"
TSG
"TSG，融合体"
融合体
"致癌基因，融合体"

图37B

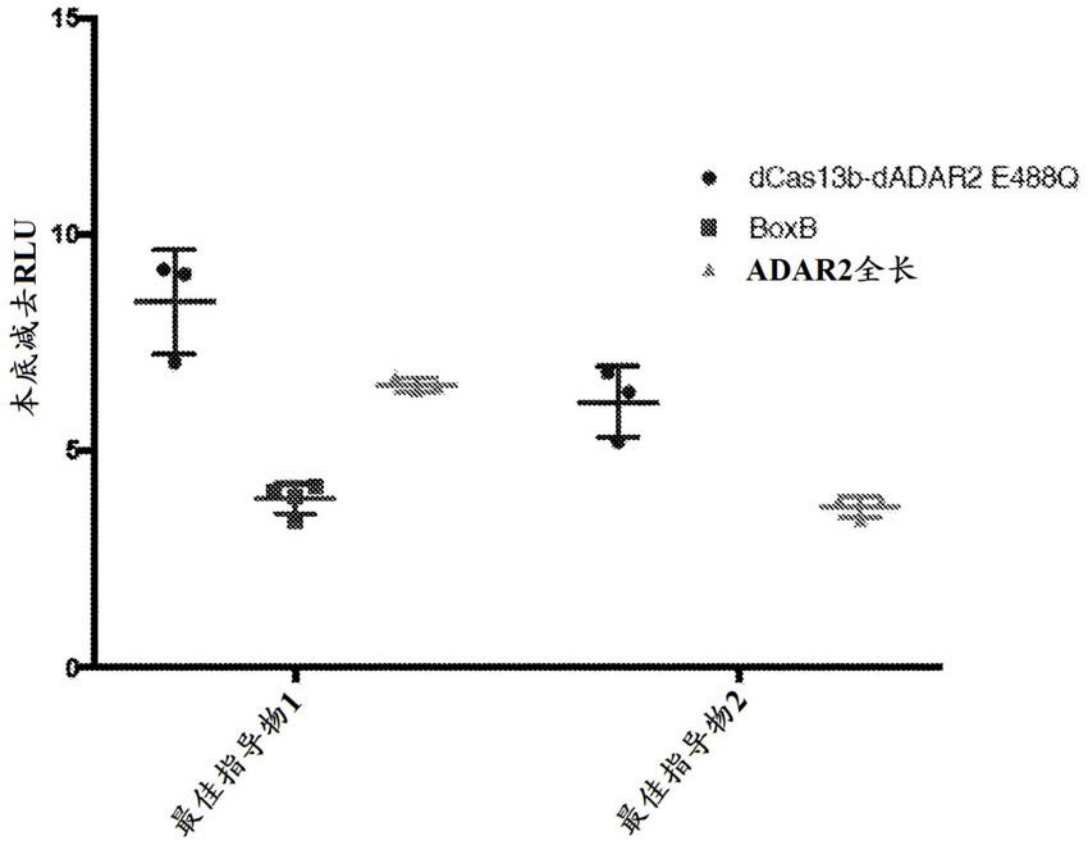

15
10
5
0
本底减去RLU
dCas13b-dADAR2 E488Q
BoxB
ADAR2全长
靶向指导物1
靶向指导物2

图38

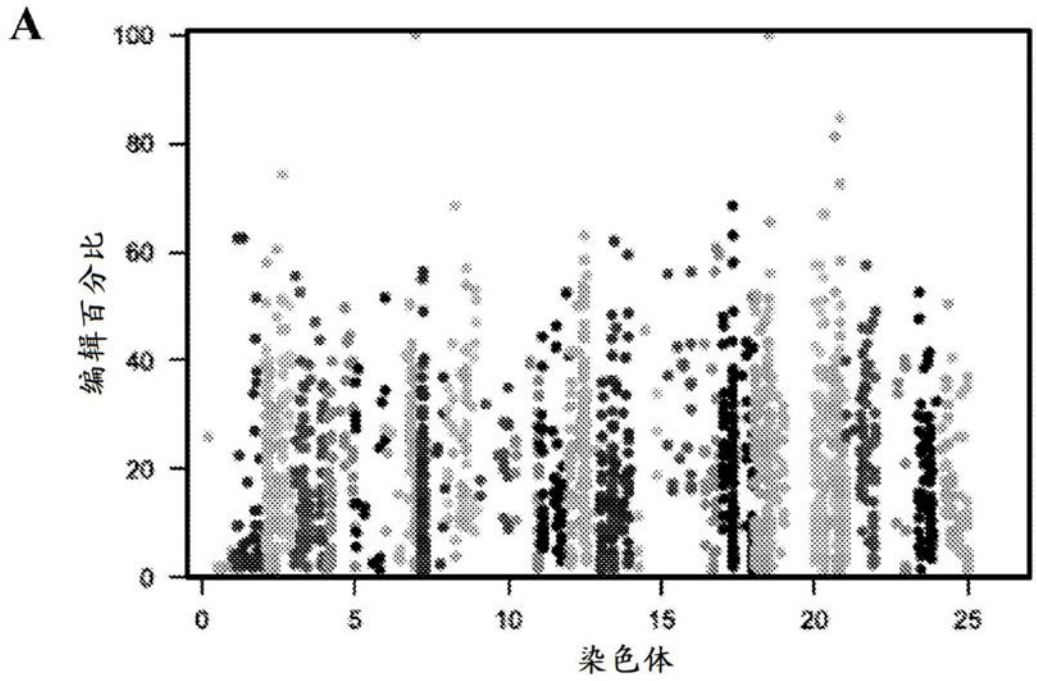

A
100
80
60
40
20
0
编辑百分比
0
5
10
15
20
25
染色体

图39A

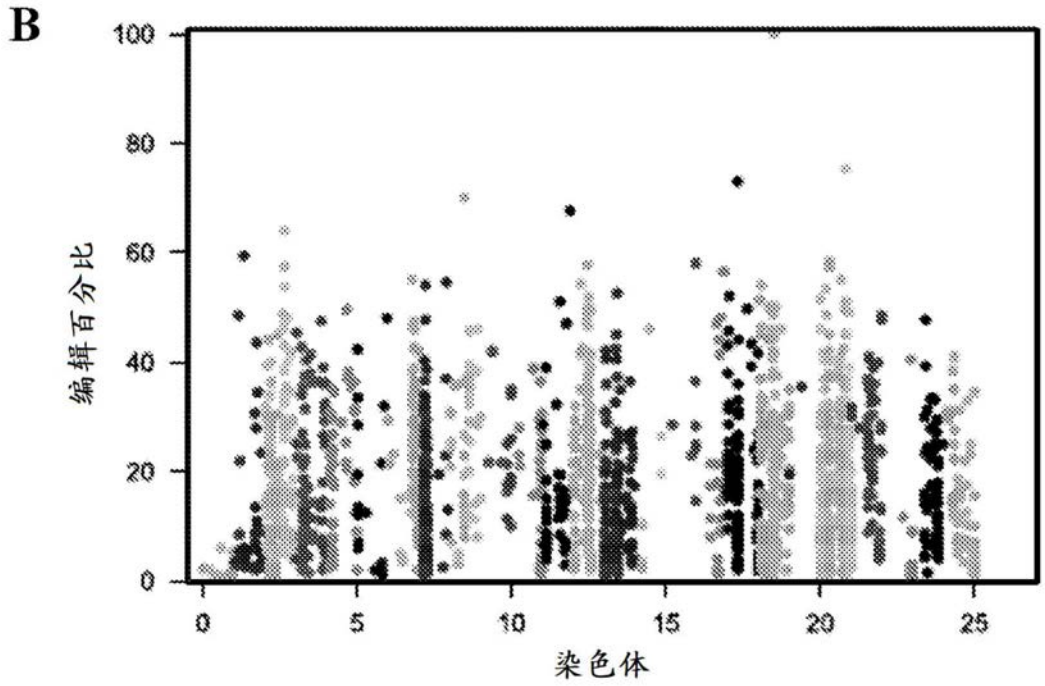

B
编辑百分比
染色体
0
20
40
60
80
100
0
5
10
15
20
25

图39B

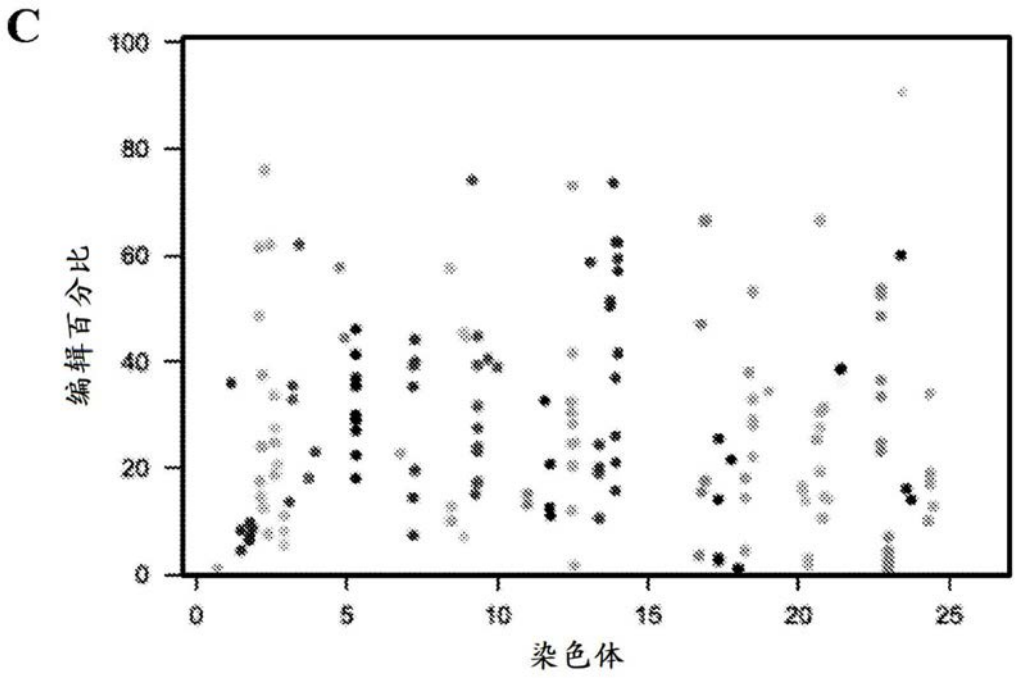

C
编辑百分比
染色体
0
20
40
60
80
100
0
5
10
15
20
25

图39C

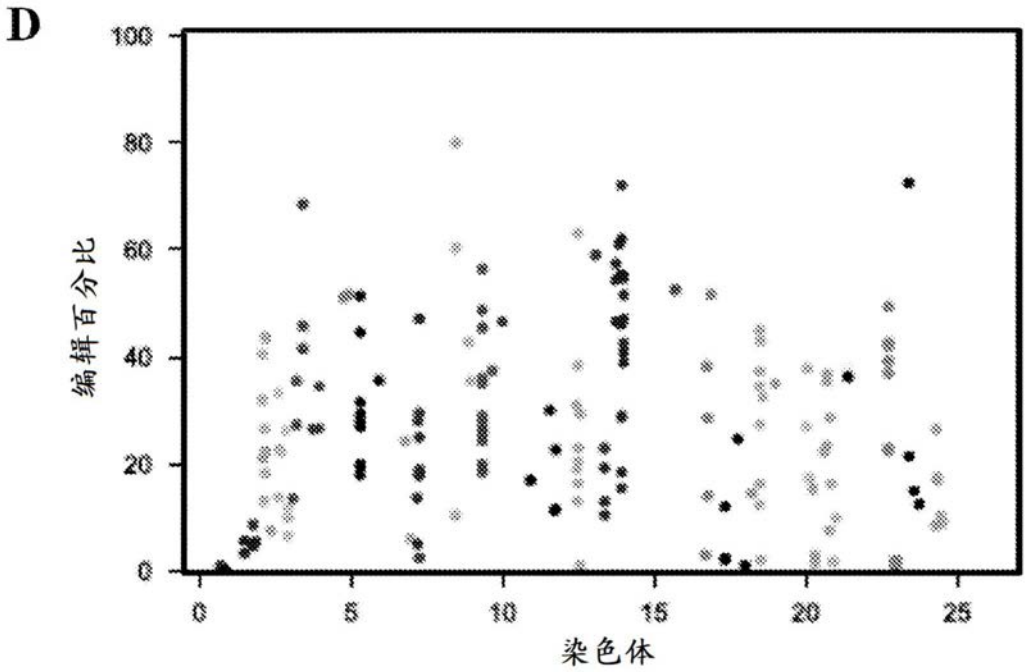



图39D

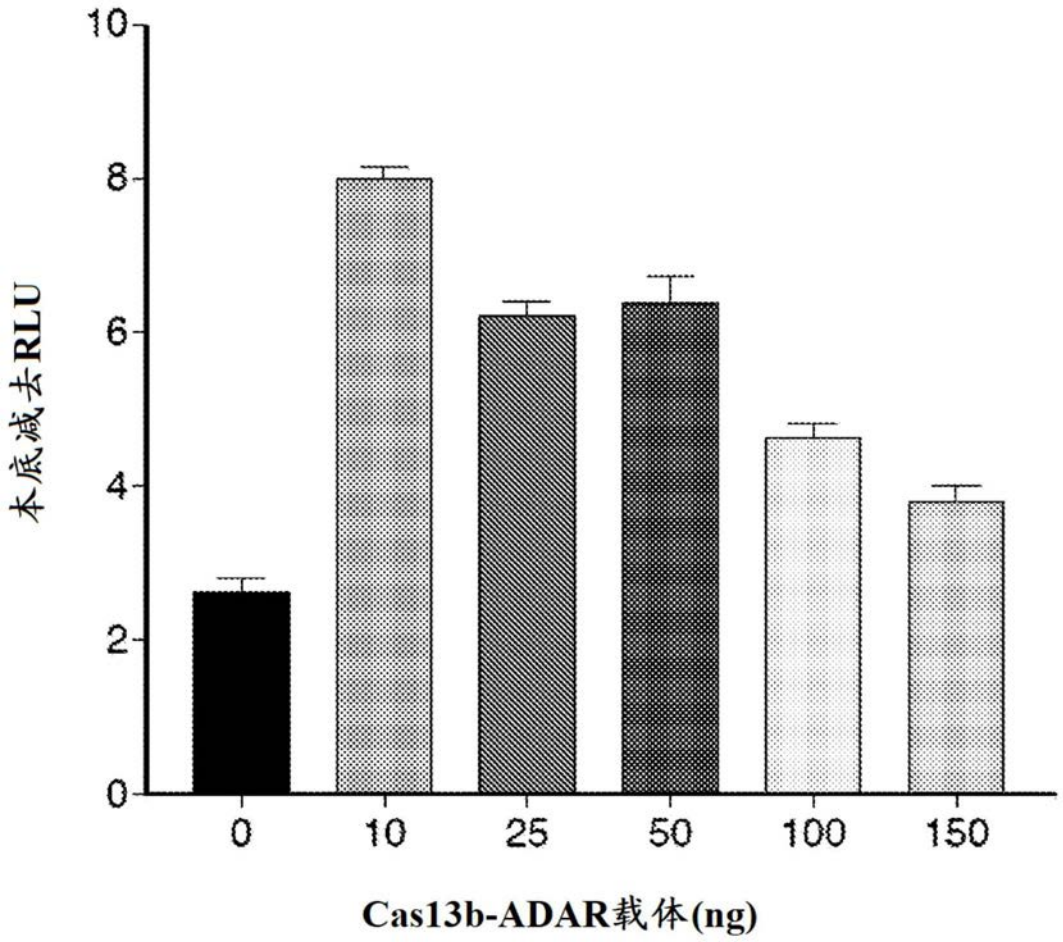

本底减去RLU
0
2
4
6
8
10
0
10
25
50
100
150
Cas13b-ADAR载体(ng)

图40

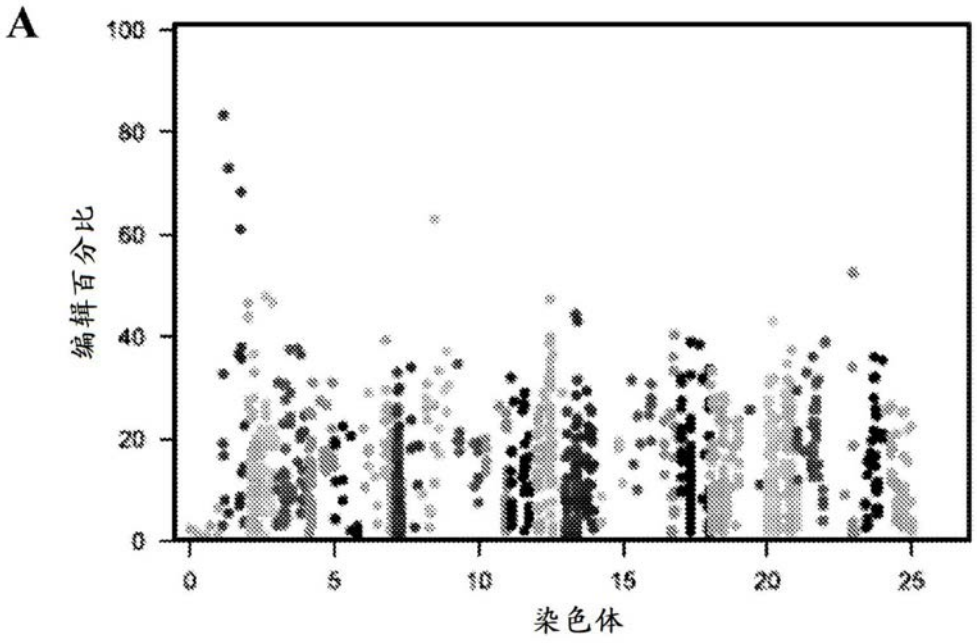

A
编辑百分比
0
20
40
60
80
100
0
5
10
15
20
25
染色体

图41A

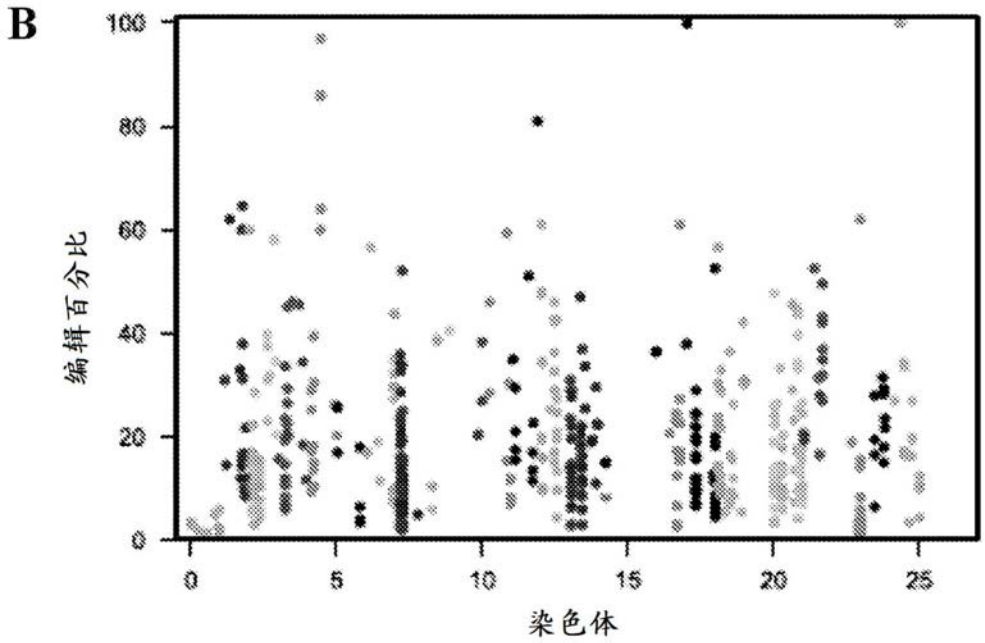

B
编辑百分比
染色体
0
20
40
60
80
100
0
5
10
15
20
25

图41B

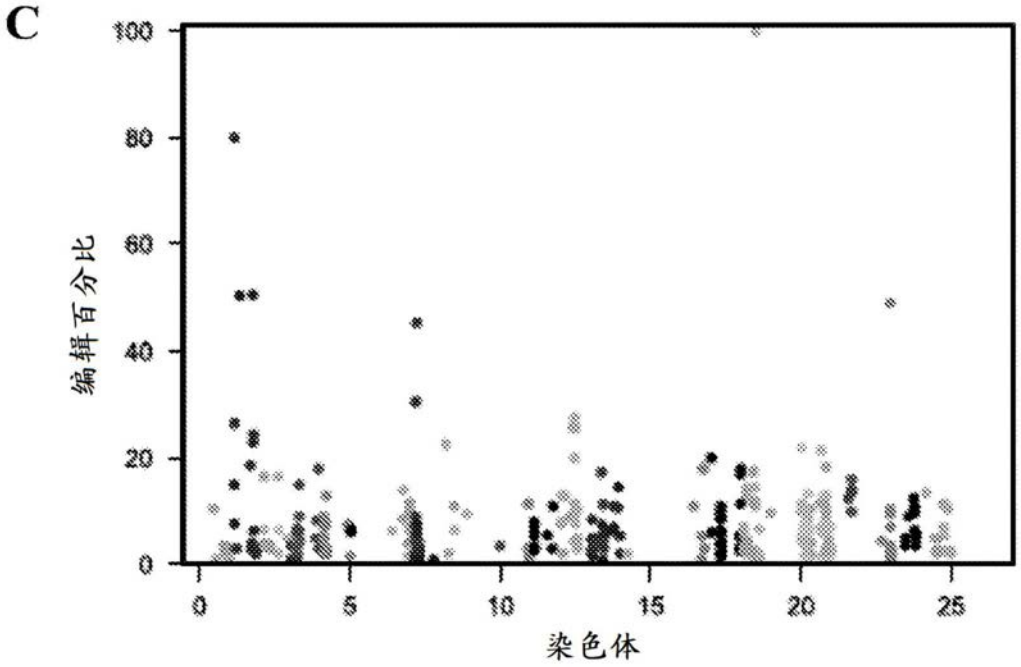

C
编辑百分比
染色体
0
20
40
60
80
100
0
5
10
15
20
25

图41C

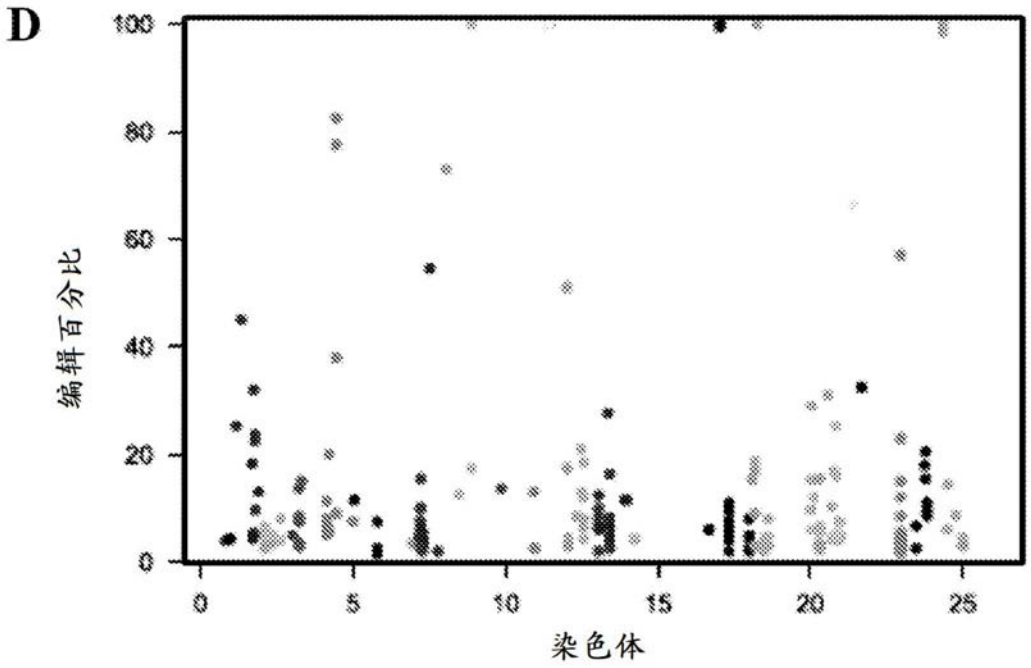

D
编辑百分比
0
20
40
60
80
100
0
5
10
15
20
25
染色体

图41D

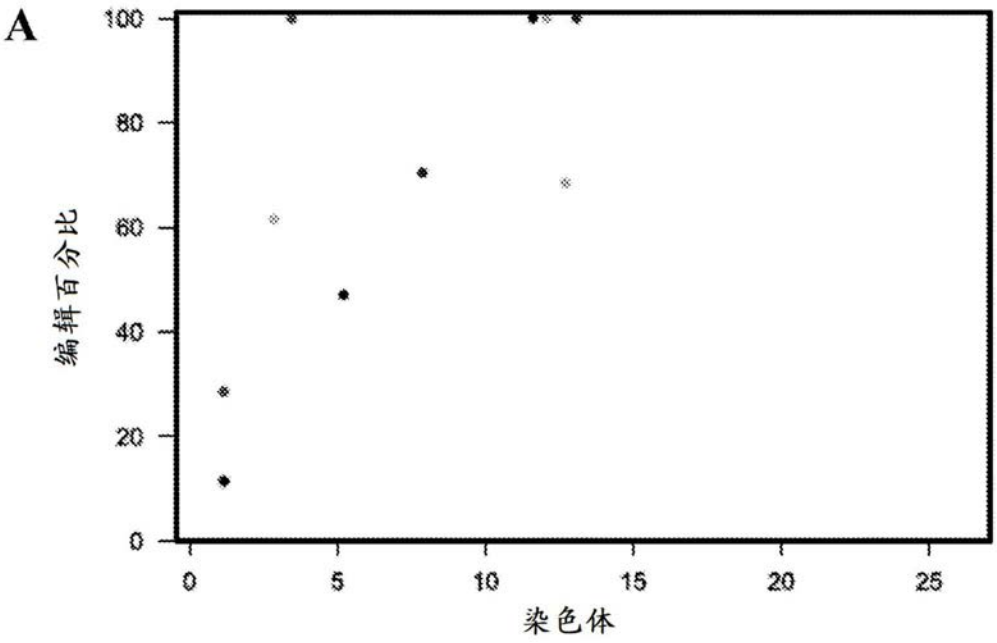

A
编辑百分比
0
20
40
60
80
100
0
5
10
15
20
25
染色体

图42A

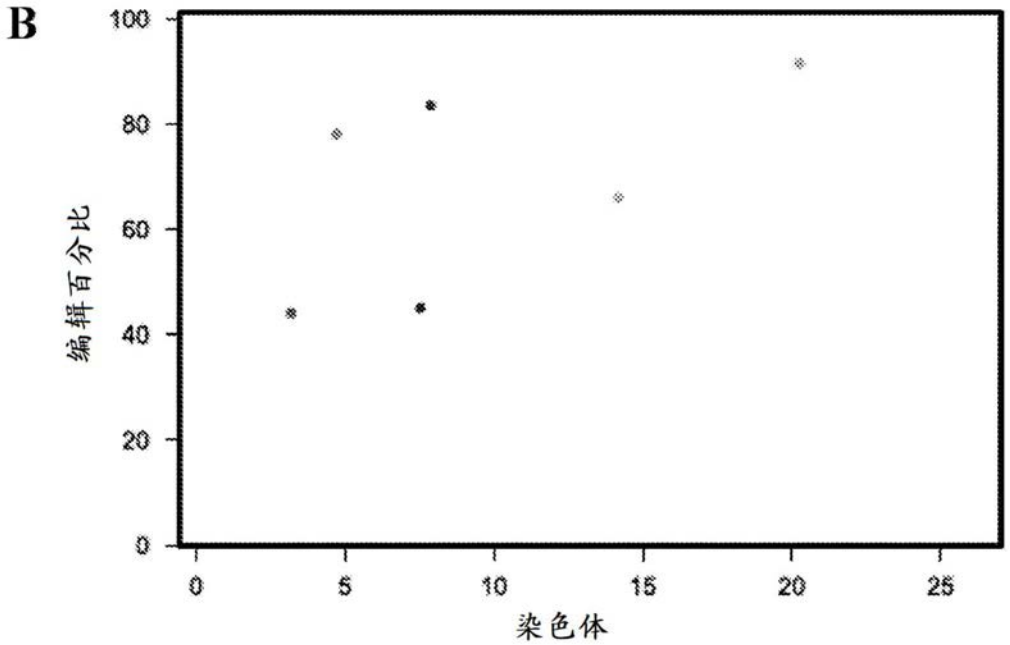

B
100
80
60
40
20
0
编辑百分比
0
5
10
15
20
25
染色体

图42B

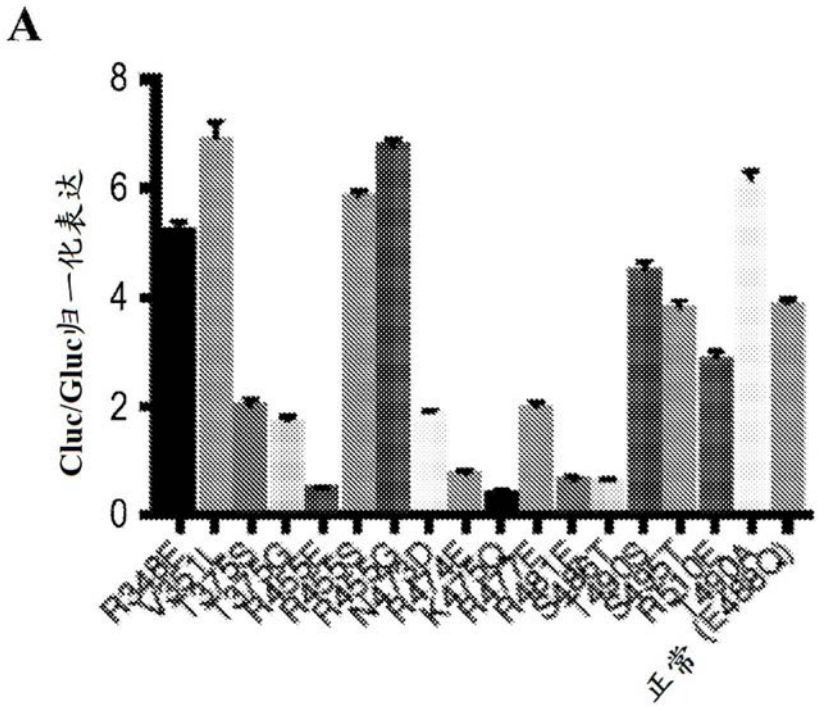

A
8
6
4
2
0
Cluc/Gluc归一化表达
正常

图43A

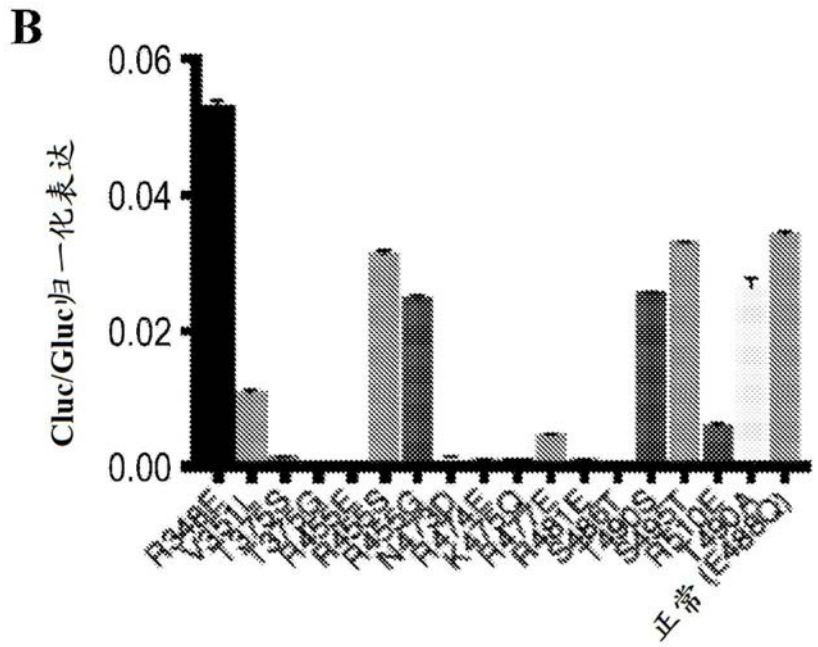

B
Cluc/Gluc归一化表达
0.06
0.04
0.02
0.00
正常(E4880)

图43B

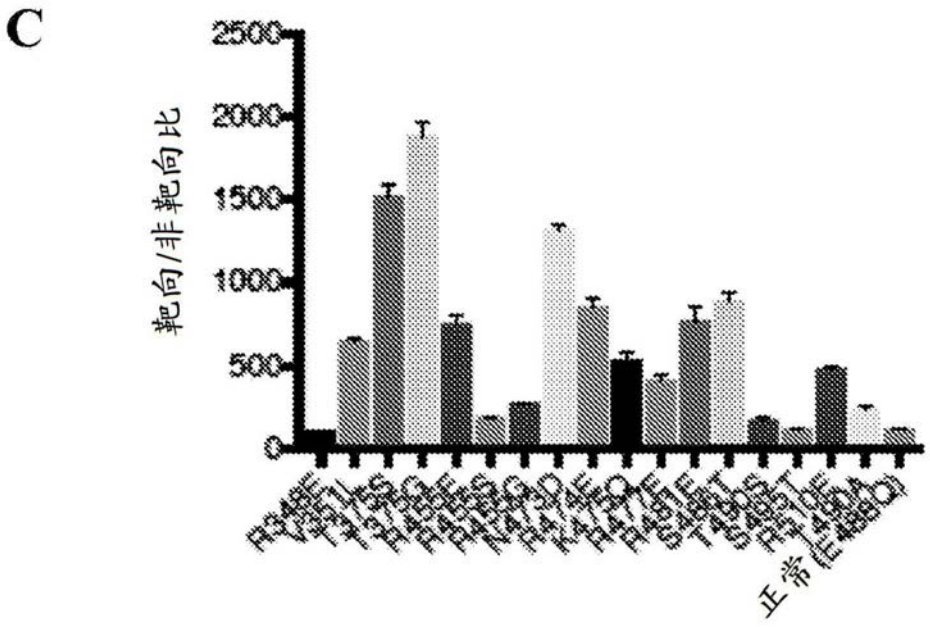

C
靶向/非靶向比
2500
2000
1500
1000
500
0
正常(E4880)

图43C

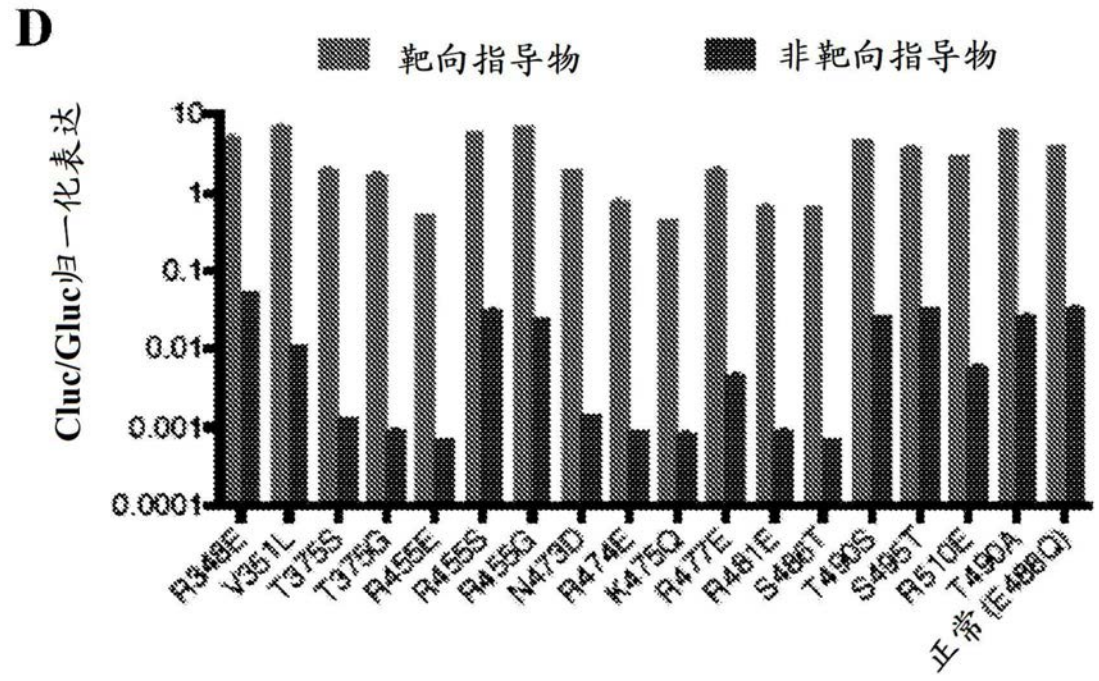

D
靶向指导物
非靶向指导物
Cluc/Gluc归一化表达
10
1
0.1
0.01
0.001
0.0001
R348E
V351L
T375S
T375G
R455E
R455S
R455G
N473D
R474E
K475Q
R477E
R481E
S486T
T490S
S495T
R510E
T490A
正常(E486Q)

图43D

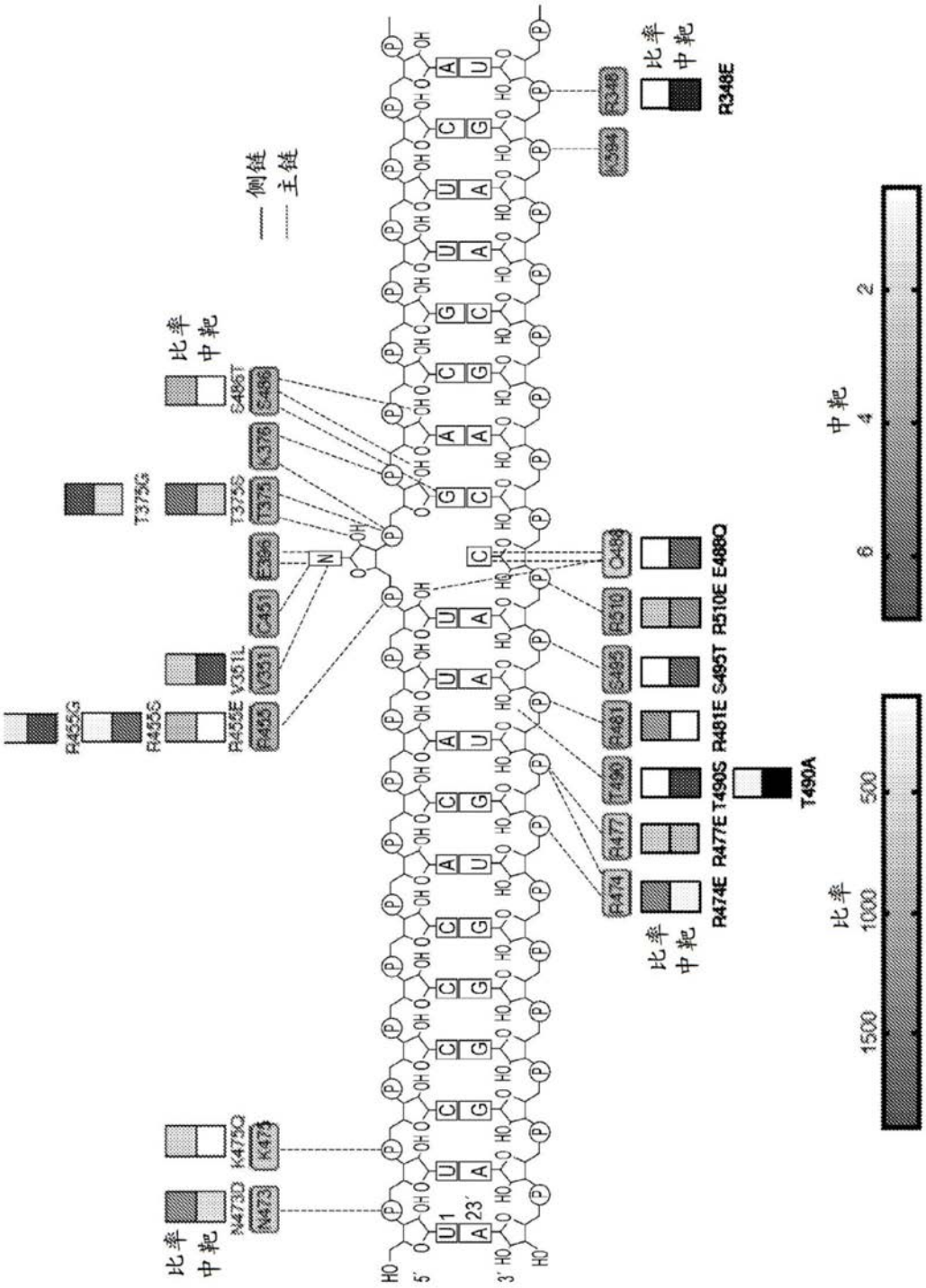

侧链
主链
比率
中靶
R348E
R474E R477E T490S R481E S495T R510E E488Q
T490A
T375G
R455G
K475Q
N473D
S486T
V351L
R455E
T375S
K478
N473
S486
K376
T375
E385
C45
V351
R455
C485
R348
R594
R474
R477
T490
R481
S495
R510
E488
HO
5'
3'
U1
A23'
2
4
6
500
1000
1500

图44

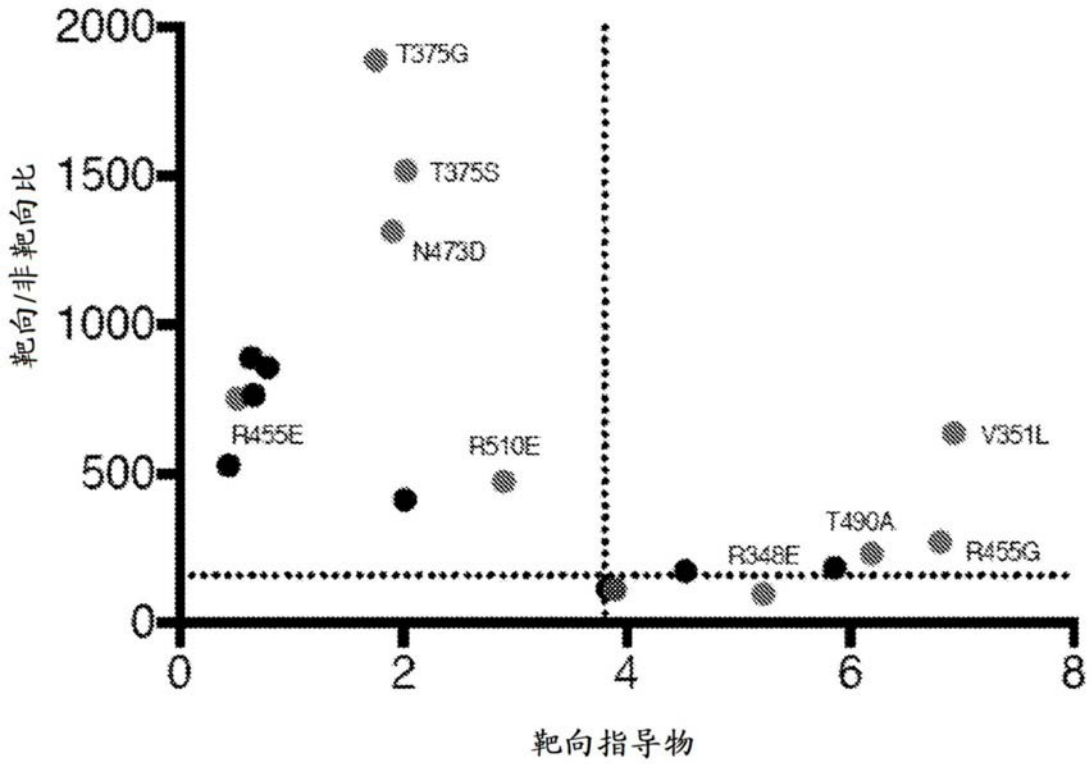

2000
1500
1000
500
0
靶向/非靶向比
T375G
T375S
N473D
R455E
R510E
V351L
T490A
R348E
R455G
0
2
4
6
8
靶向指导物

图45

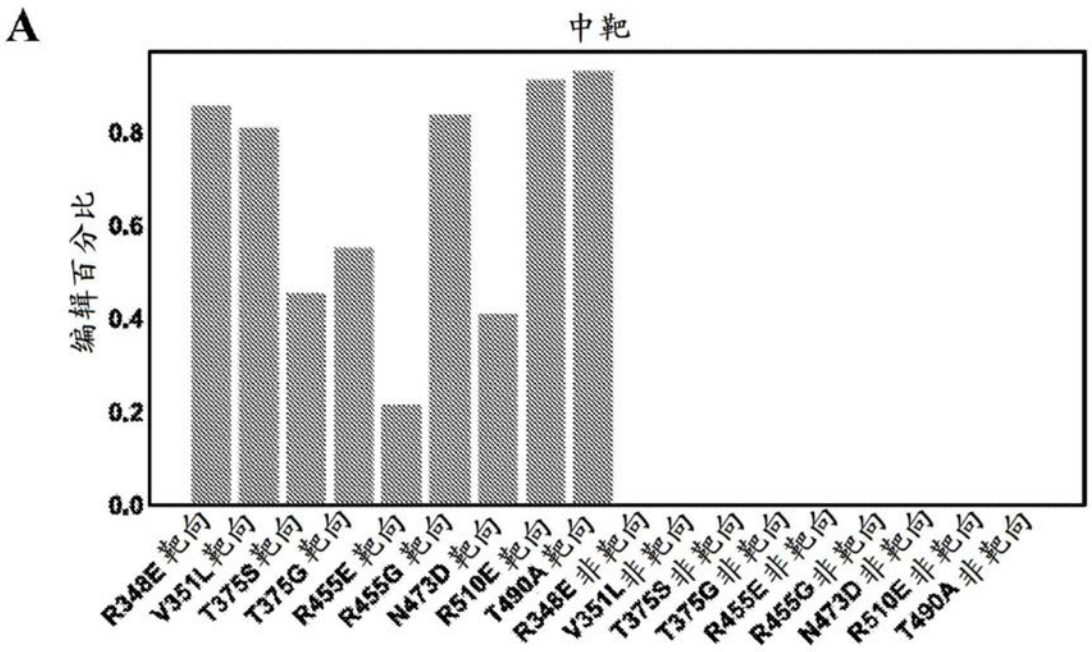

A
中靶
编辑百分比
0.0
0.2
0.4
0.6
0.8
R348E 靶向
V351L 靶向
T375S 靶向
T375G 靶向
R455E 靶向
R455G 靶向
N473D 靶向
R510E 靶向
T490A 靶向
R348E 非靶向
V351L 非靶向
T375S 非靶向
T375G 非靶向
R455E 非靶向
R455G 非靶向
N473D 非靶向
R510E 非靶向
T490A 非靶向

图46A

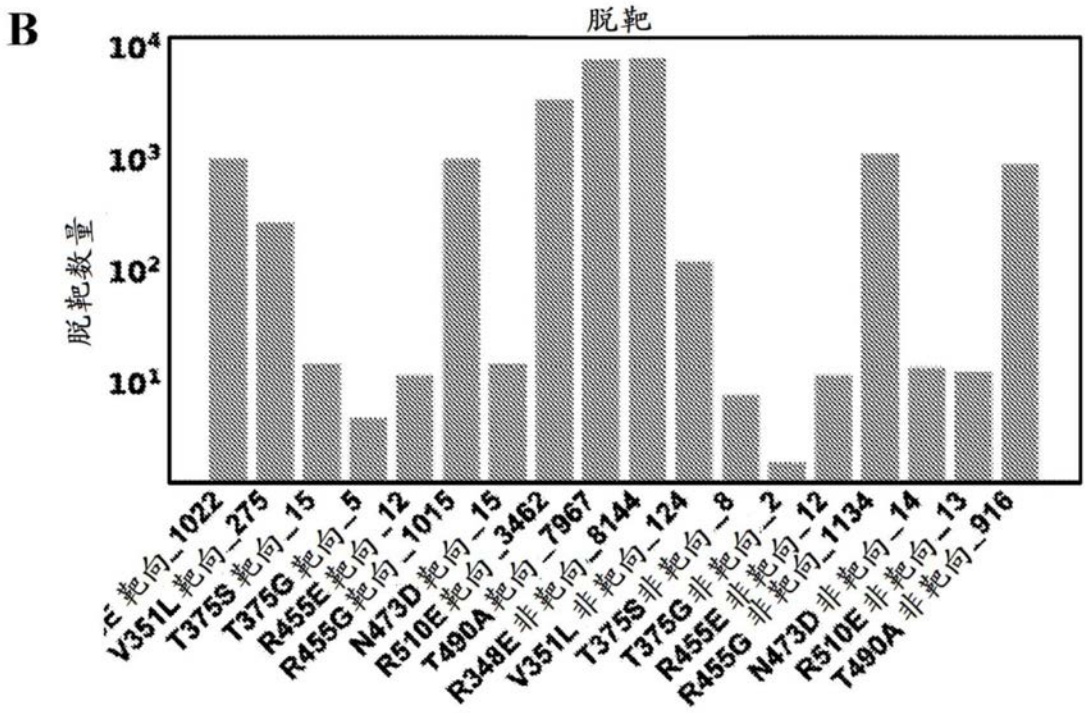

B
脱靶
脱靶数量
$10^4$
$10^3$
$10^2$
$10^1$
V351L_靶向_1022
T375S_靶向_275
T375G_靶向_15
R455E_靶向_5
R455G_靶向_12
N473D_靶向_1015
R510E_靶向_15
T490A_靶向_3462
R348E非靶向_7967
V351L_非靶向_8144
T375S非靶向_124
T375G非靶向_8
R455E非靶向_2
R455G非靶向_12
N473D非靶向_1134
R510E非靶向_14
T490A非靶向_13
非靶向_916

图46B

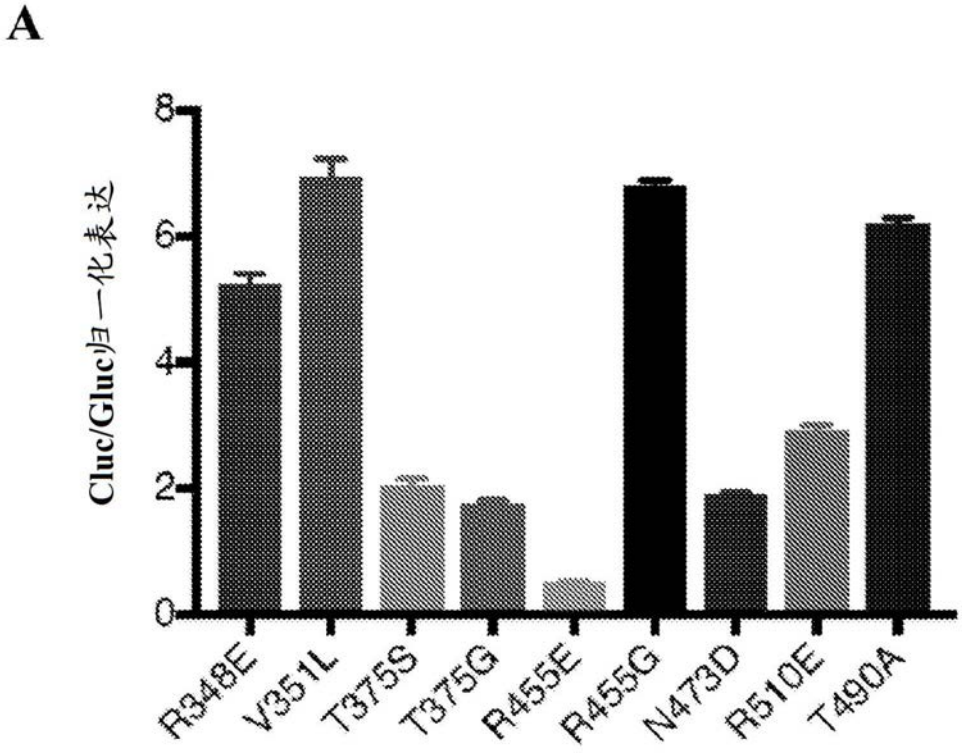

A
Cluc/Gluc归一化表达
8
6
4
2
0
R348E
V351L
T375S
T375G
R455E
R455G
N473D
R510E
T490A

图47A

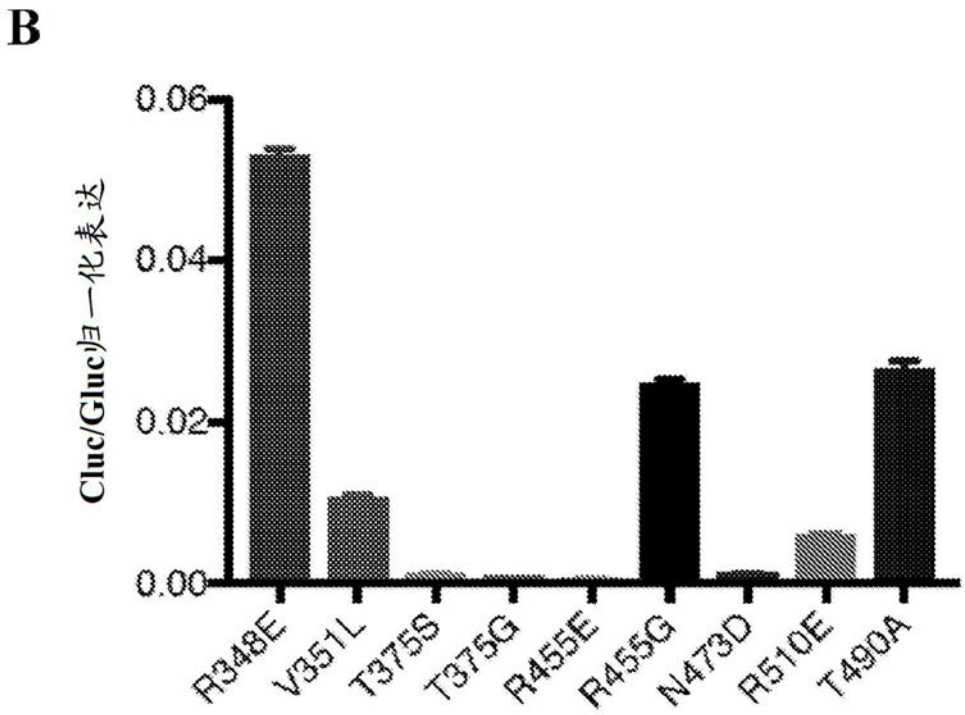

B
Cluc/Gluc归一化表达
0.06
0.04
0.02
0.00
R348E
V351L
T375S
T375G
R455E
R455G
N473D
R510E
T490A

图47B

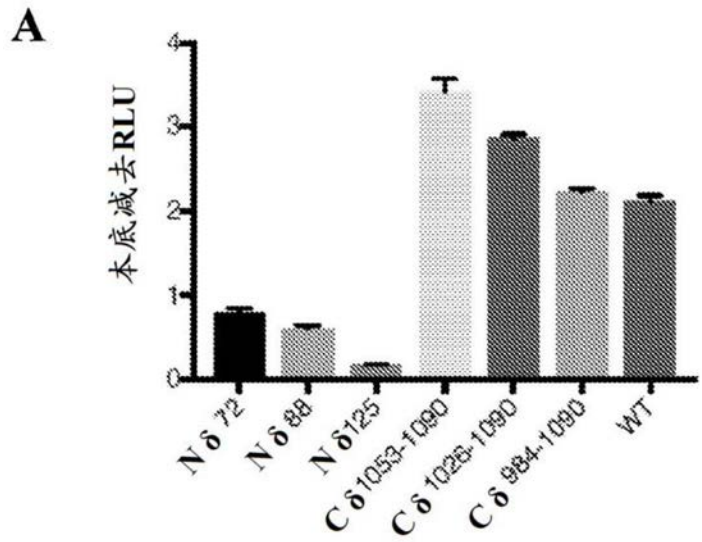

A
本底减去RLU
0
1
2
3
4
N δ 72
N δ 88
N δ 125
C δ 1053-1090
C δ 1026-1090
C δ 984-1090
WT

图48A

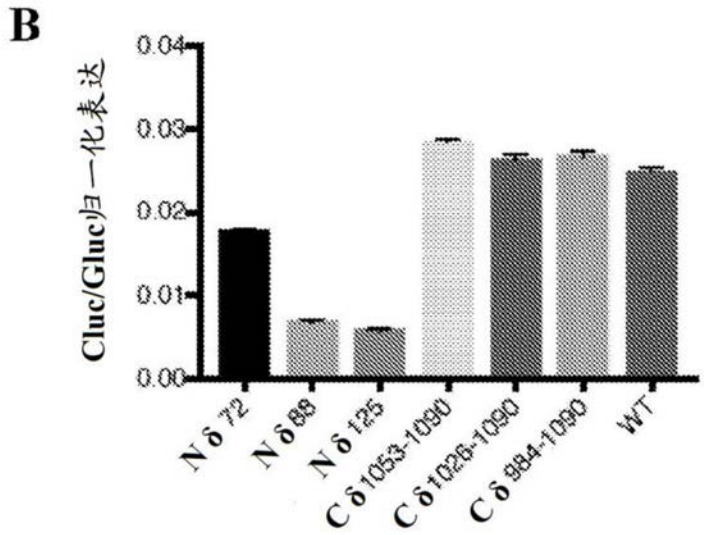

B
Cluc/Gluc归一化表达
0.04
0.03
0.02
0.01
0.00
N δ 72
N δ 88
N δ 125
C δ 1053-1090
C δ 1026-1090
C δ 984-1090
WT

图48B

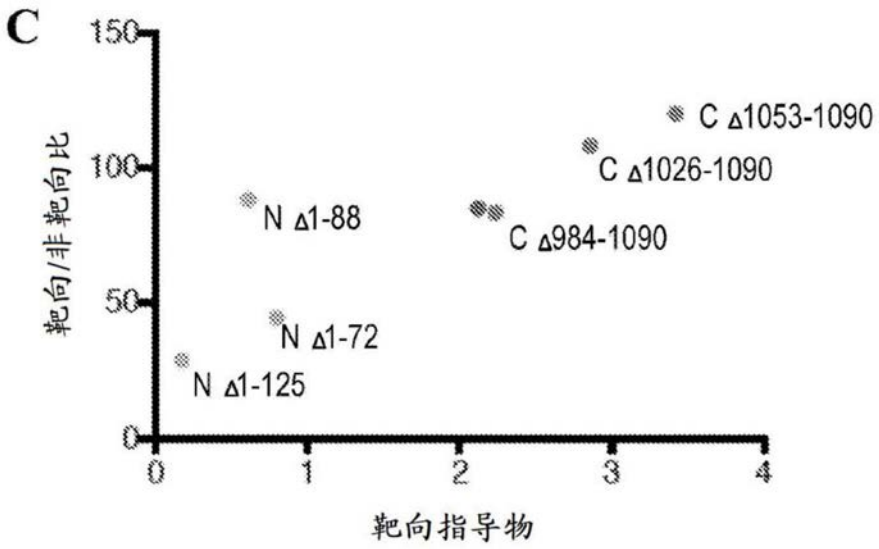

C
150
100
50
0
靶向/非靶向比
C Δ1053-1090
C Δ1026-1090
N Δ1-88
C Δ984-1090
N Δ1-72
N Δ1-125
0
1
2
3
4
靶向指导物

图48C

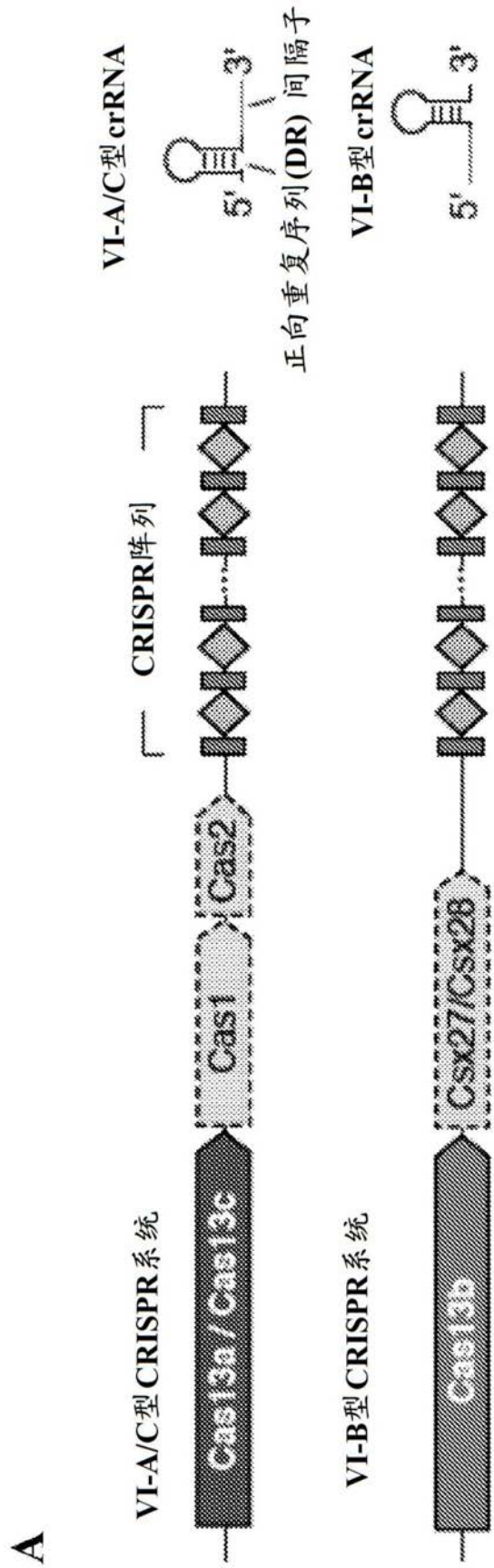

A
VI-A/C型CRISPR系统
Cas13a / Cas13c
Cas1
Cas2
CRISPR阵列
VI-A/C型crRNA
5'
3'
正向重复序列(DR)
间隔子
VI-B型CRISPR系统
Cas13b
Csx27/Csx28
VI-B型crRNA
5'
3'

图49A

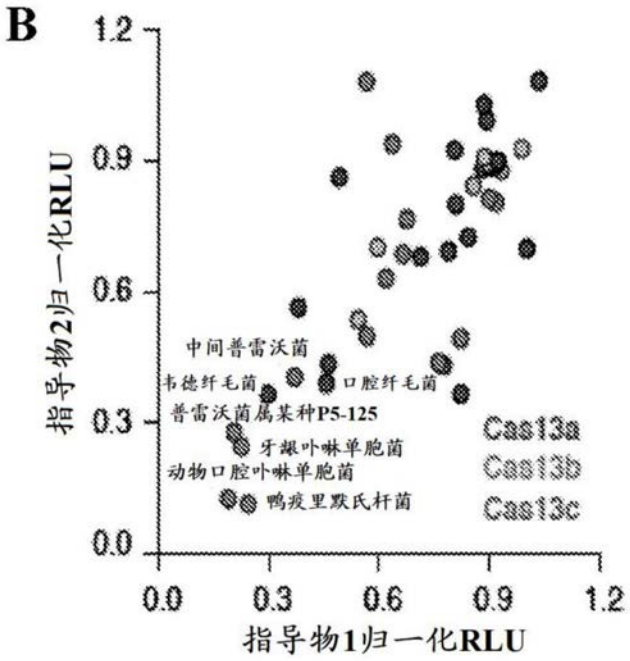

B
1.2
0.9
0.6
0.3
0.0
指导物2归一化RLU
中间普雷沃菌
韦德纤毛菌
口腔纤毛菌
普雷沃菌属某种P5-125
牙龈卟啉单胞菌
动物口腔卟啉单胞菌
鸭疫里默氏杆菌
Cas13a
Cas13b
Cas13c
0.0
0.3
0.6
0.9
1.2
指导物1归一化RLU

图49B

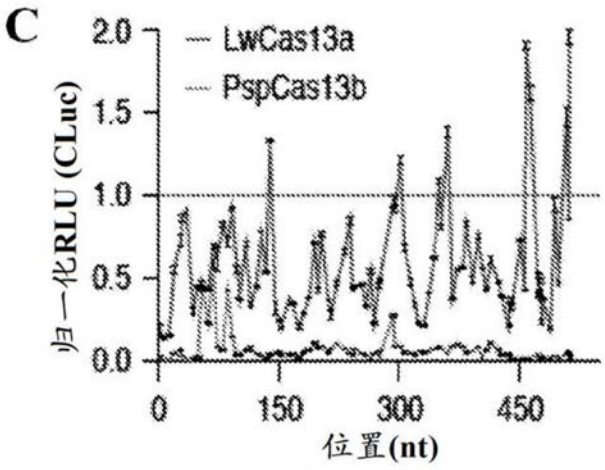

C
2.0
1.5
1.0
0.5
0.0
归一化RLU (CLuc)
LwCas13a
PspCas13b
0
150
300
450
位置(nt)

图49C

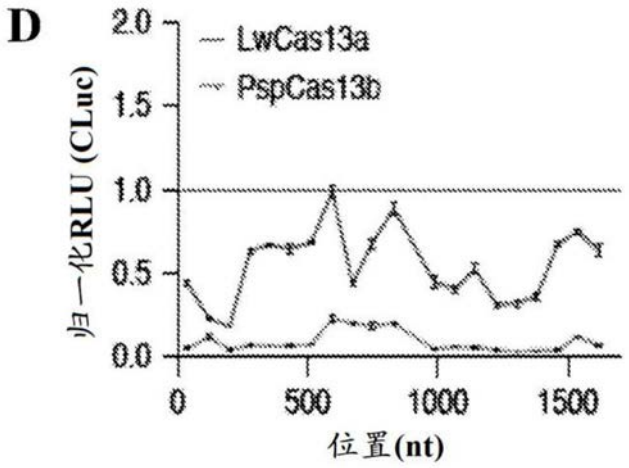

D
LwCas13a
PspCas13b
归一化RLU (CLuc)
2.0
1.5
1.0
0.5
0.0
0
500
1000
1500
位置(nt)

图49D

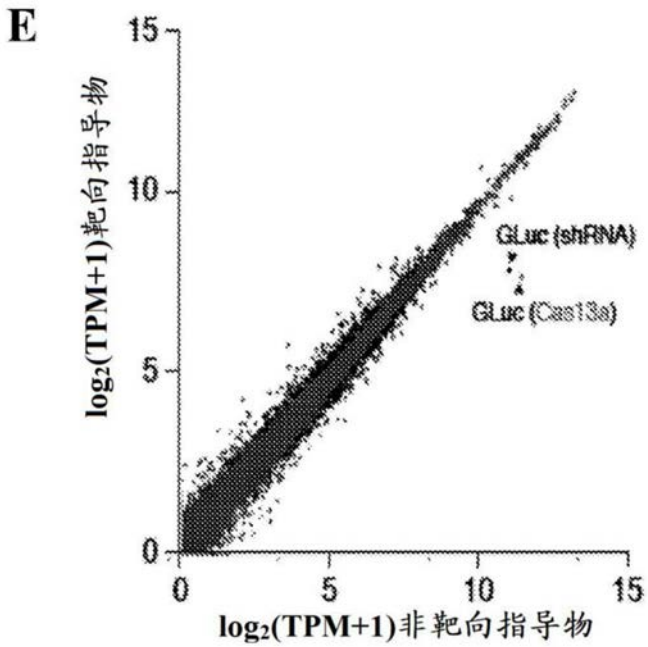

E
log2(TPM+1)靶向指导物
15
10
5
0
GLuc (shRNA)
GLuc (Cas13a)
0
5
10
15
log2(TPM+1)非靶向指导物

图49E

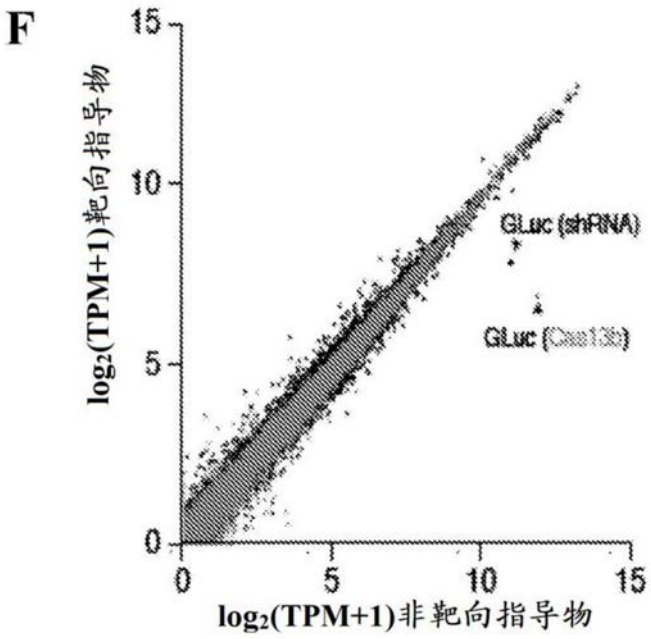

F
15
10
5
0
0
5
10
15
log2(TPM+1)靶向指导物
log2(TPM+1)非靶向指导物
GLuc (shRNA)
GLuc (Cas13b)

图49F

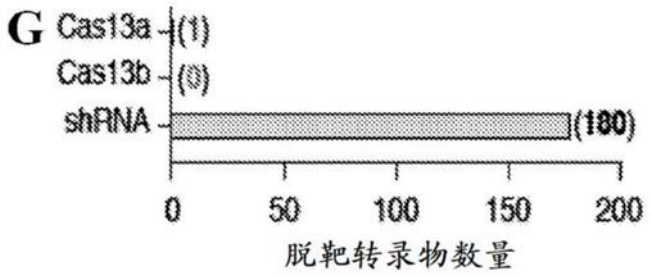

G
Cas13a
Cas13b
shRNA
(1)
(0)
(180)
0
50
100
150
200
脱靶转录物数量

图49G

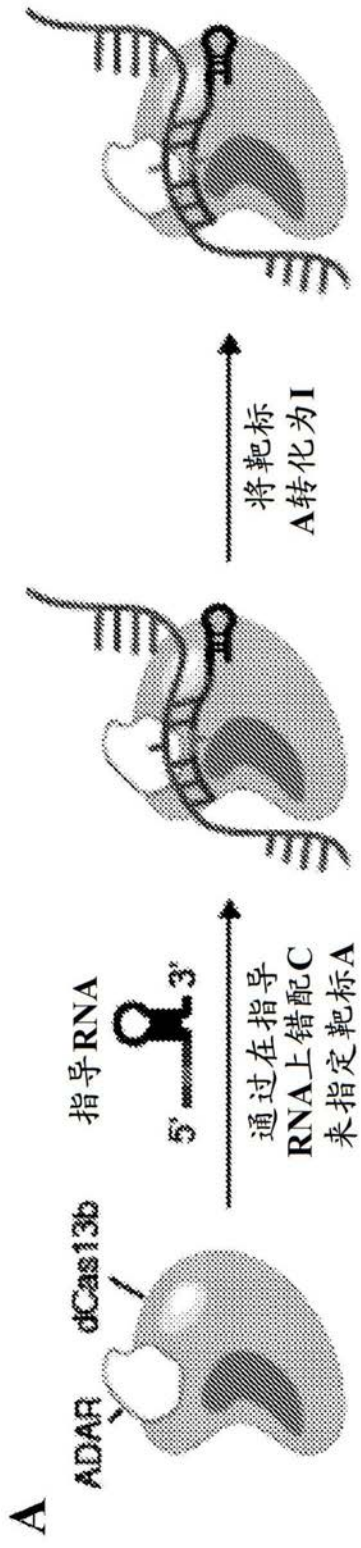



图50A


B
靶mRNA
(CLuc W85X)
指导RNA
I76 E77 C78 R79 V80 E81 A82 A83 G84 X85 F86 R87 T88 F89 Y90
5'— ..AUU GAA UGC AGA GUA GAG GCC GCA GGA UMG UUU AGA ACA UUC UAU.. — 3'
U G ACCUGGAAGGUGUUG ACG UCU CAU CUC CGG CGU CCU AGC AAA UCU — 5'
U G GGGGCUAUUACAACA — 3'
错配距离
间隔子长度

图50B

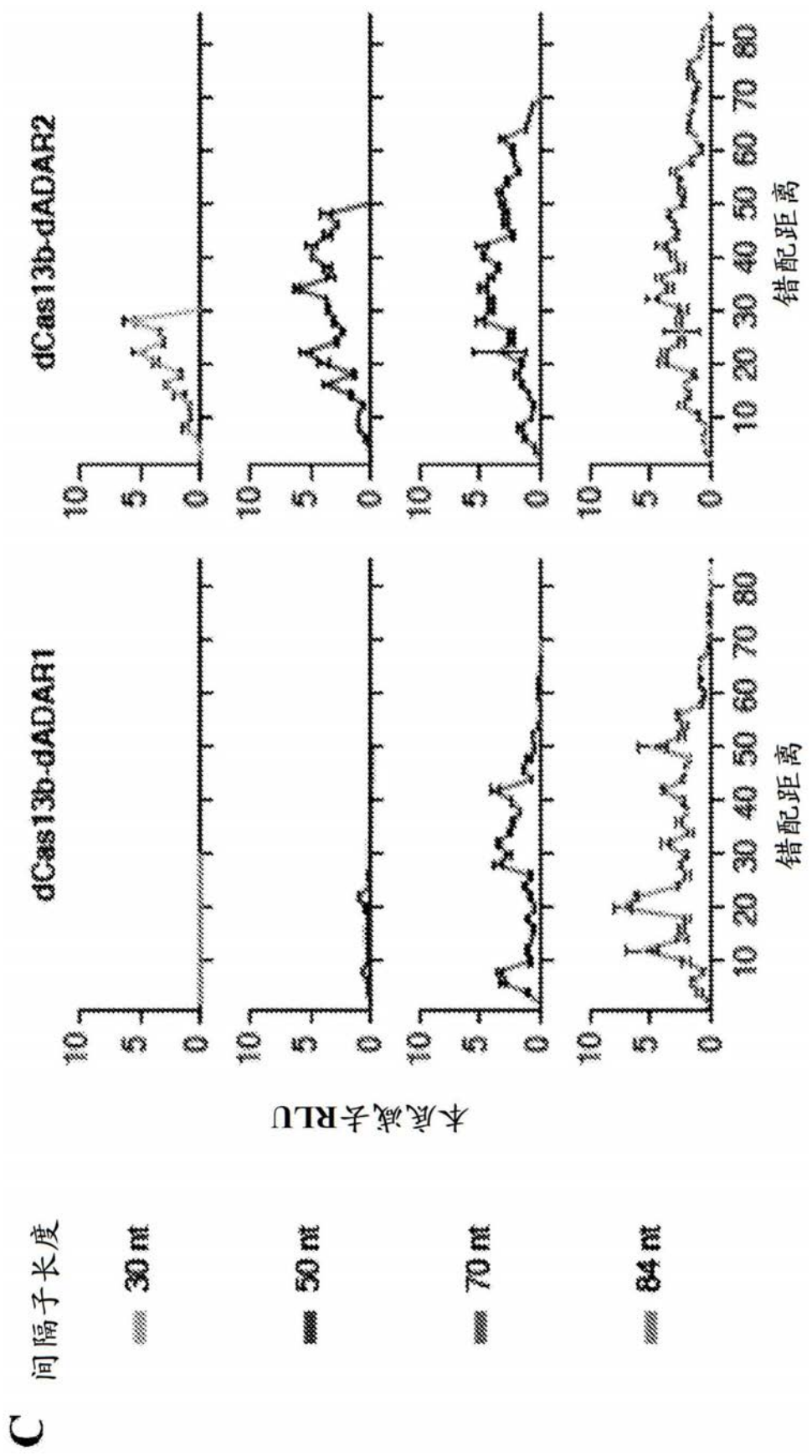

C
间隔子长度
30 nt
50 nt
70 nt
84 nt
dCas13b-dADAR1
dCas13b-dADAR2
本底减去RLU
错配距离
错配距离

图50C

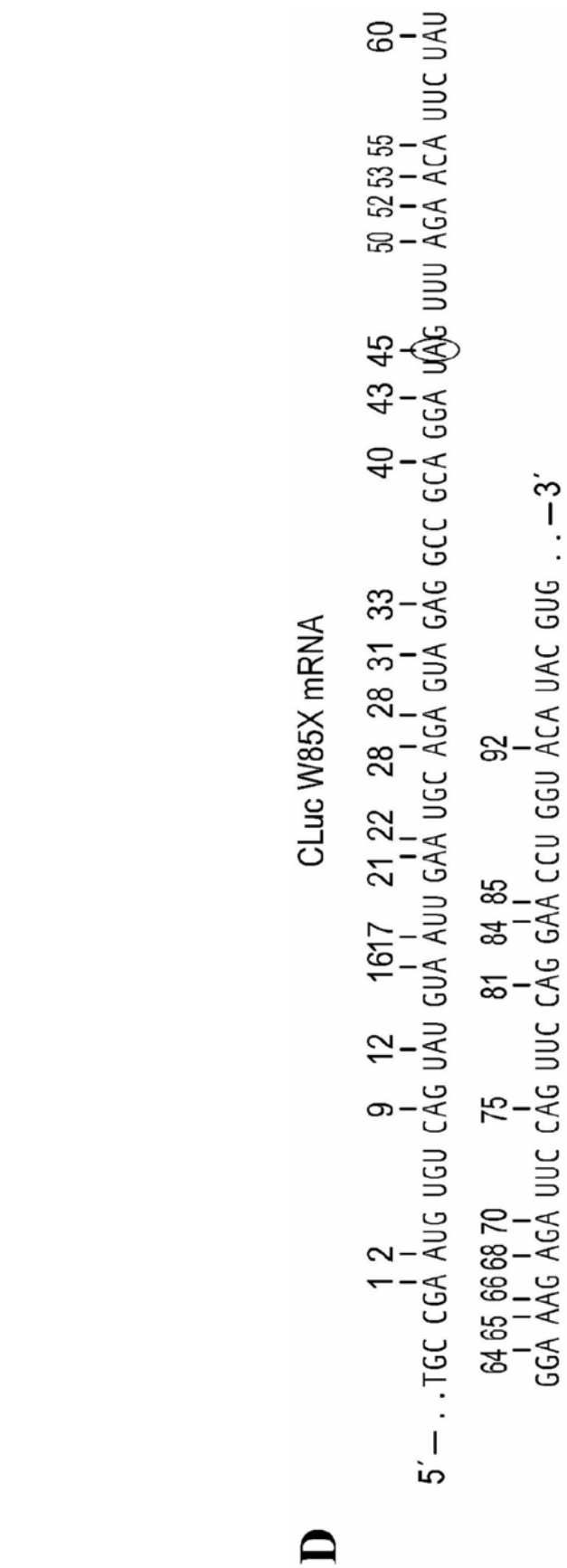
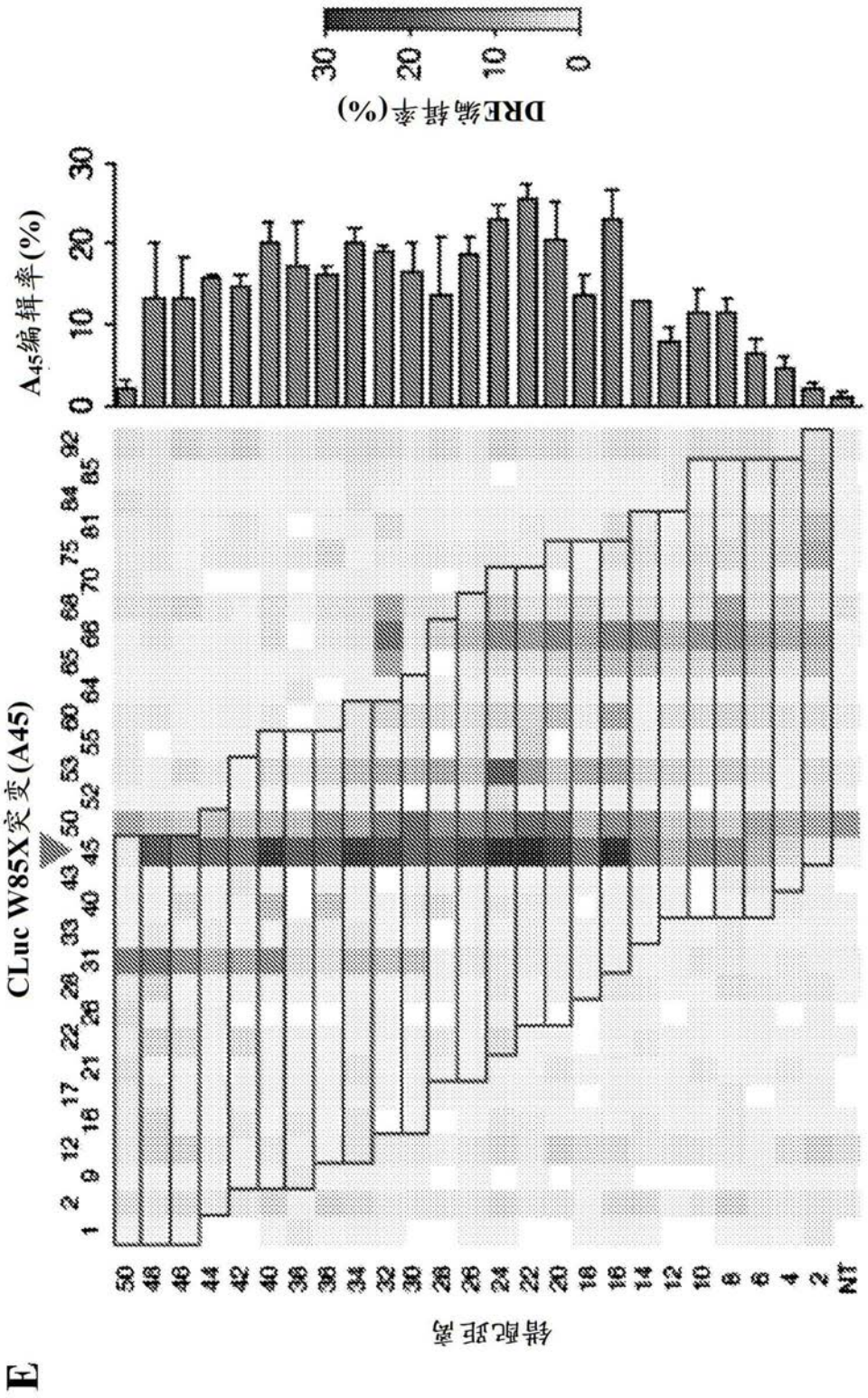

D
CLuc W85X mRNA
5′—..TGC CGA AUG UGU CAG UAU GUA AUU GAA UGC AGA GUA GAG GCC GCA GGA UAG UUU AGA ACA UUC UAU
GGA AAG AGA UUC CAG UUC CAG GAA CCU GGU ACA UAC GUG..—3′

图50D




E

图50E

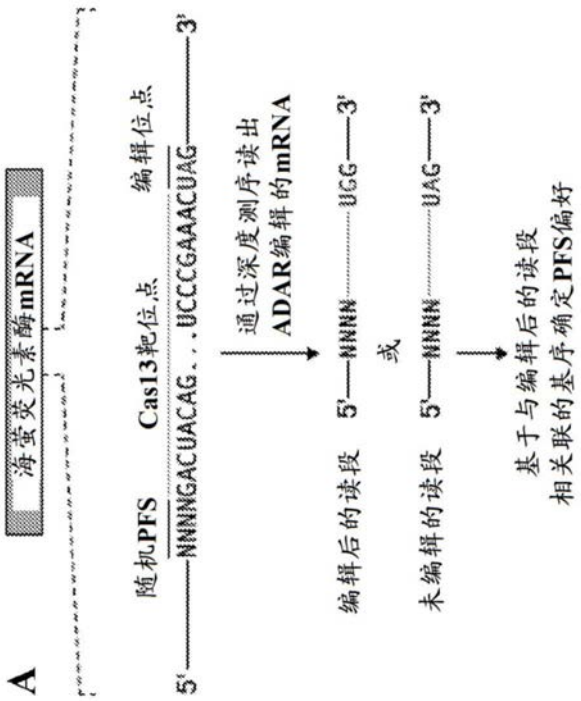

A
海萤荧光素酶mRNA
随机PFS
Cas13靶位点
编辑位点
5'
NNNNGACUACAG...UCCCGAAACUAG
3'
通过深度测序读出
ADAR编辑的mRNA
编辑后的读段
5'
NNNN
UGG
3'
或
未编辑的读段
5'
NNNN
UAG
3'
基于与编辑后的读段
相关联的基序确定PFS偏好

图51A

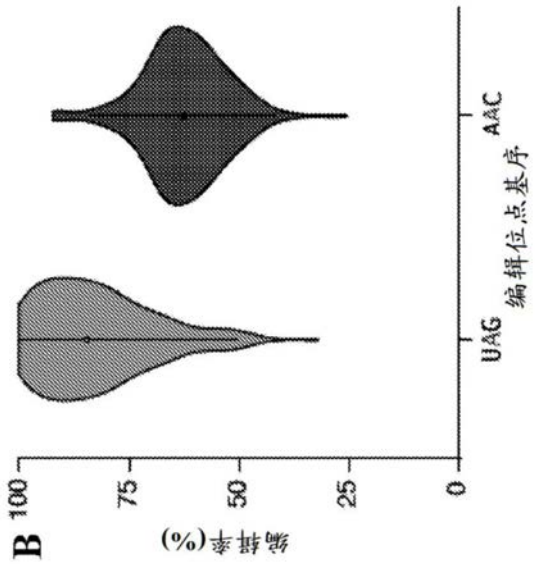

B
100
75
50
25
0
编辑率(%)
UAG
AAC
编辑位点基序

图51B

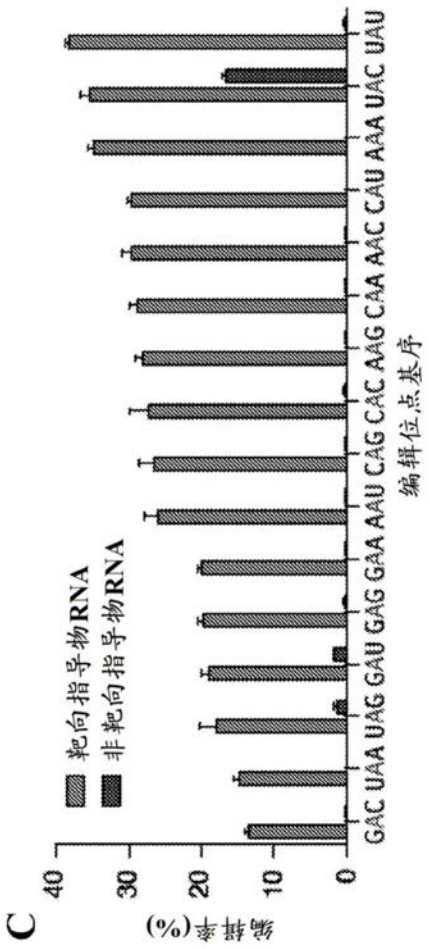

C
编辑率(%)
40
30
20
10
0
靶向指导物RNA
非靶向指导物RNA
GAC UAA UAG GAU GAG GAA AAU CAG CAC AAG CAA AAC CAU AAA UAC UAU
编辑位点基序

图51C

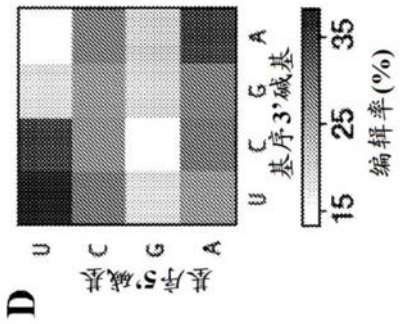

D
基序5'碱基
U
C
G
A
U C G A
基序3'碱基
15 25 35
编辑率(%)

图51D

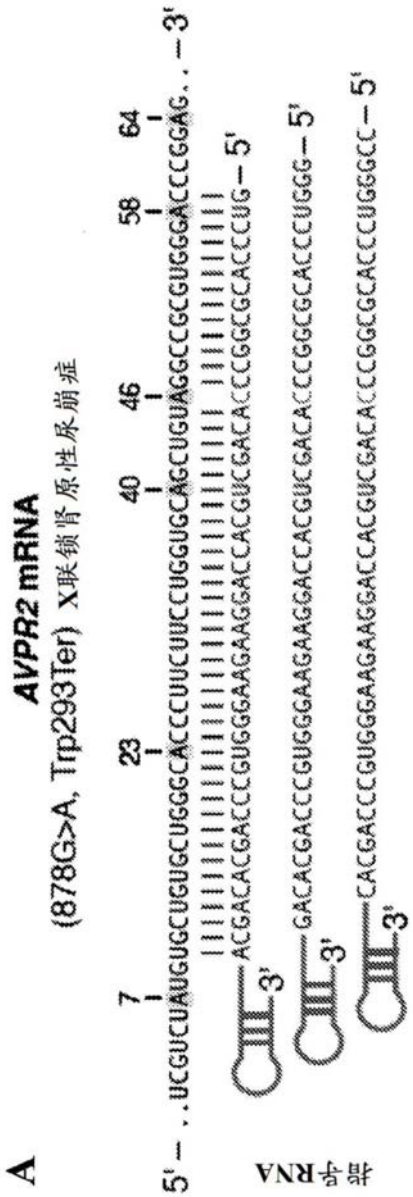



图52A

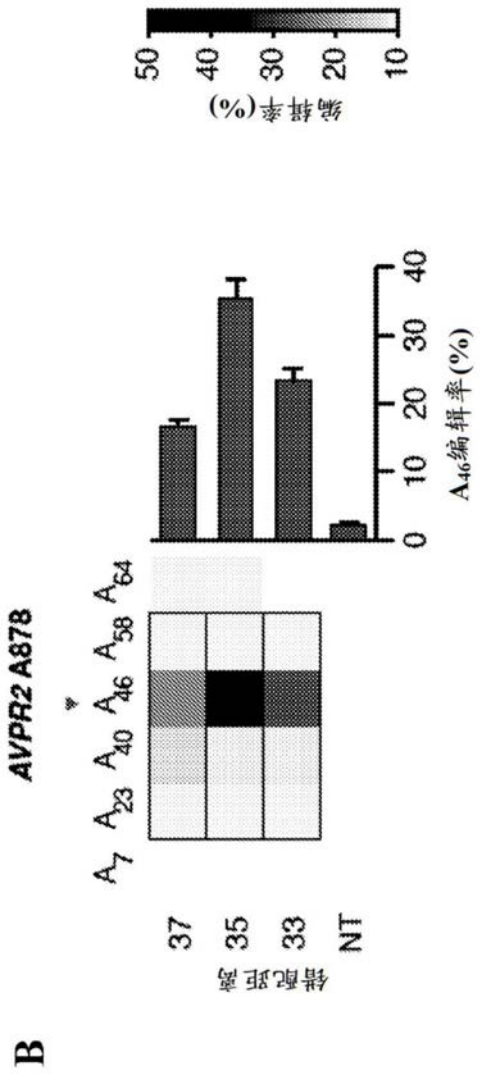

50
40
30
20
10
编辑率(%)
0
10
20
30
40
A46编辑率(%)
AVPR2 A878
A64
A58
A46
A40
A23
A7
37
35
33
NT
错配距离
B

图52B

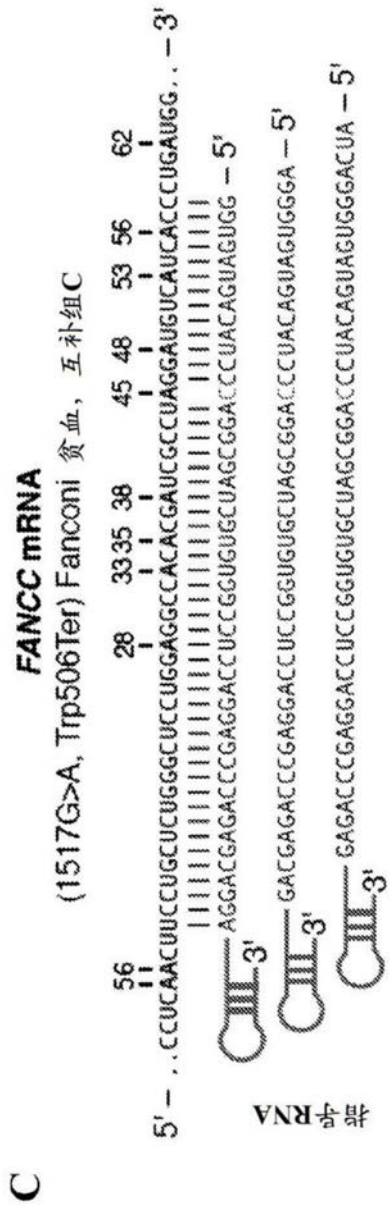

C
FANCC mRNA
(1517G>A, Trp506Ter) Fanconi 贫血，互补组C
5′ – . . CCUCAACUUCCUGCUCUGGGCUCCUGGAGGCCACACGAUCGCCUAGGAUGUCAUCACCCUGAUGG. . – 3′
56 28 33 35 38 45 48 53 56 62
AGGACGAGACCCGAGGACCUCCGGUGUGCUAGCGGACCCUACAGUAGUGG – 5′
3′
GACGAGACCCGAGGACCUCCGGUGUGCUAGCGGACCCUACAGUAGUGGGA – 5′
3′
GAGACCCGAGGACCUCCGGUGUGCUAGCGGACCCUACAGUAGUGGGACUA – 5′
3′
指导RNA

图52C

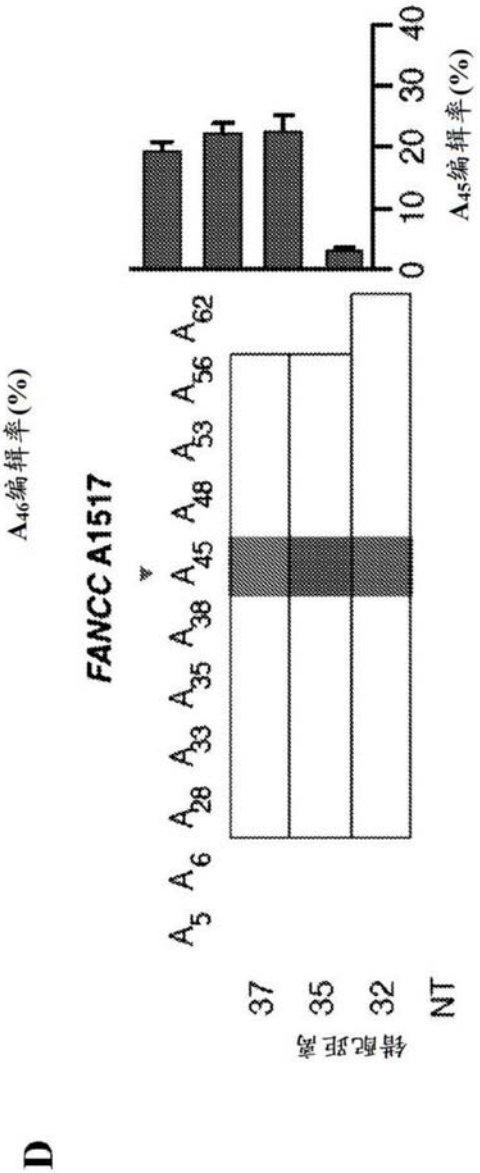

D
FANCC A1517
A46编辑率(%)
A45编辑率(%)
0 10 20 30 40
A5 A6 A28 A33 A35 A38 A45 A48 A53 A56 A62
37 35 32 NT
错配距离

图52D

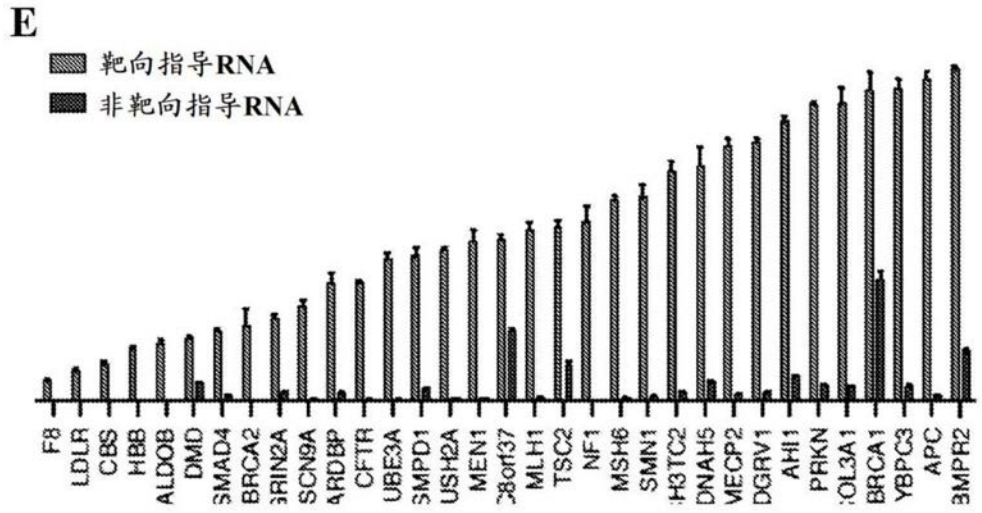

E
靶向指导RNA
非靶向指导RNA
F8
LDLR
CBS
HBB
ALDOB
DMD
SMAD4
BRCA2
SCN9A
CFTR
UBE3A
SMPD1
USH2A
MEN1
MLH1
TSC2
NF1
MSH6
SMN1
DNAH5
MECP2
DGRV1
AHI1
PRKN
BRCA1
YBPC3
APC

图52E

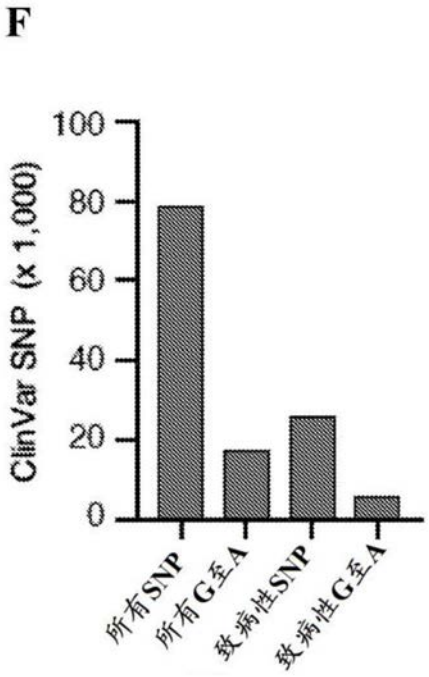

F
ClinVar SNP (x 1,000)
100
80
60
40
20
0
所有SNP
所有G至A
致病性SNP
致病性G至A

图52F

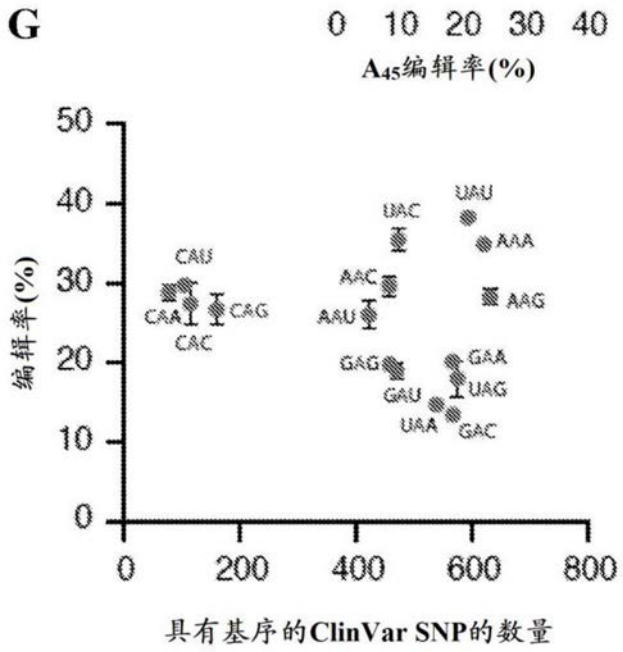

G
0 10 20 30 40
A45编辑率(%)
50
40
30
20
10
0
编辑率(%)
CAU
CAA
CAG
CAC
UAC
UAU
AAA
AAC
AAG
AAU
GAG
GAA
GAU
UAG
UAA
GAC
0 200 400 600 800
具有基序的ClinVar SNP的数量

图52G

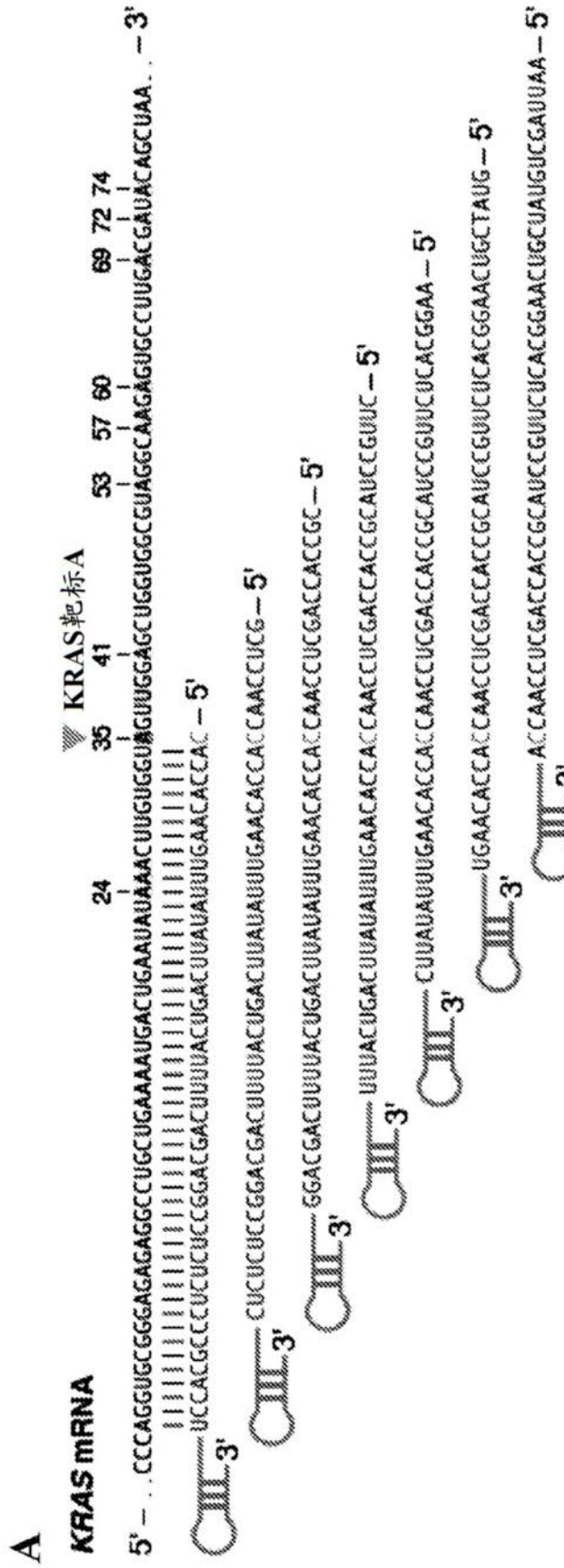



图53A

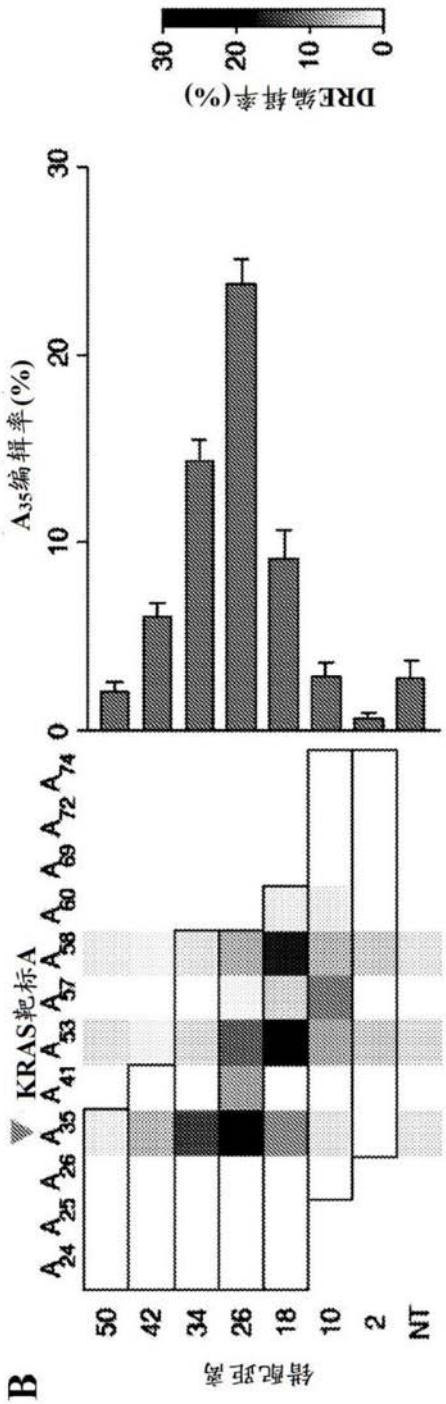

DRE编辑率(%)
30
20
10
0
A35编辑率(%)
30
20
10
0
KRAS靶标A
A24 A25 A26 A35 A41 A53 A57 A58 A60 A69 A72 A74
错配距离
50 42 34 26 18 10 2 NT
B

图53B

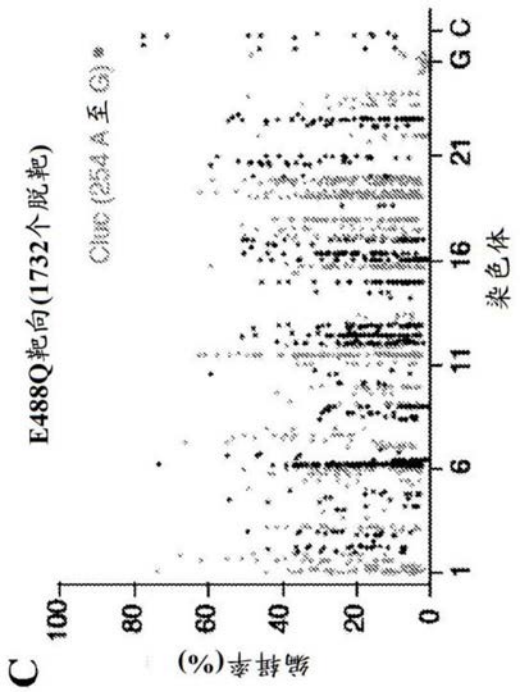

C
E488Q靶向(1732个脱靶)
Chr6 (254 A 至 G) *
编辑率(%)
100
80
60
40
20
0
1
6
11
16
21
G C
染色体

图53C

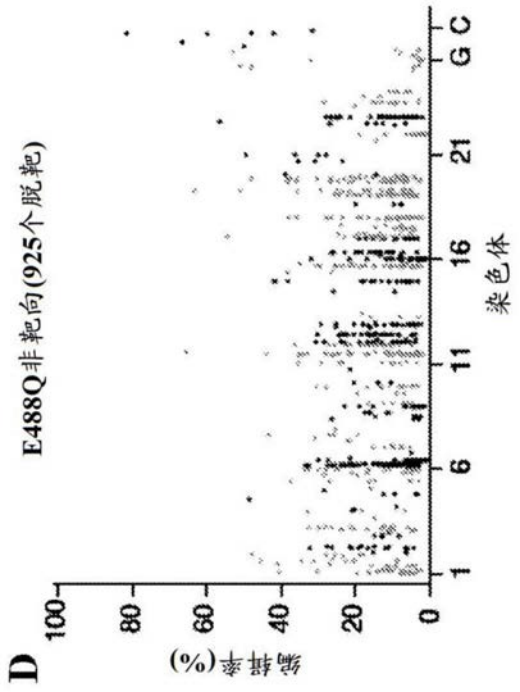

D
E488Q非靶向(925个脱靶)
编辑率(%)
100
80
60
40
20
0
1
6
11
16
21
G C
染色体

图53D

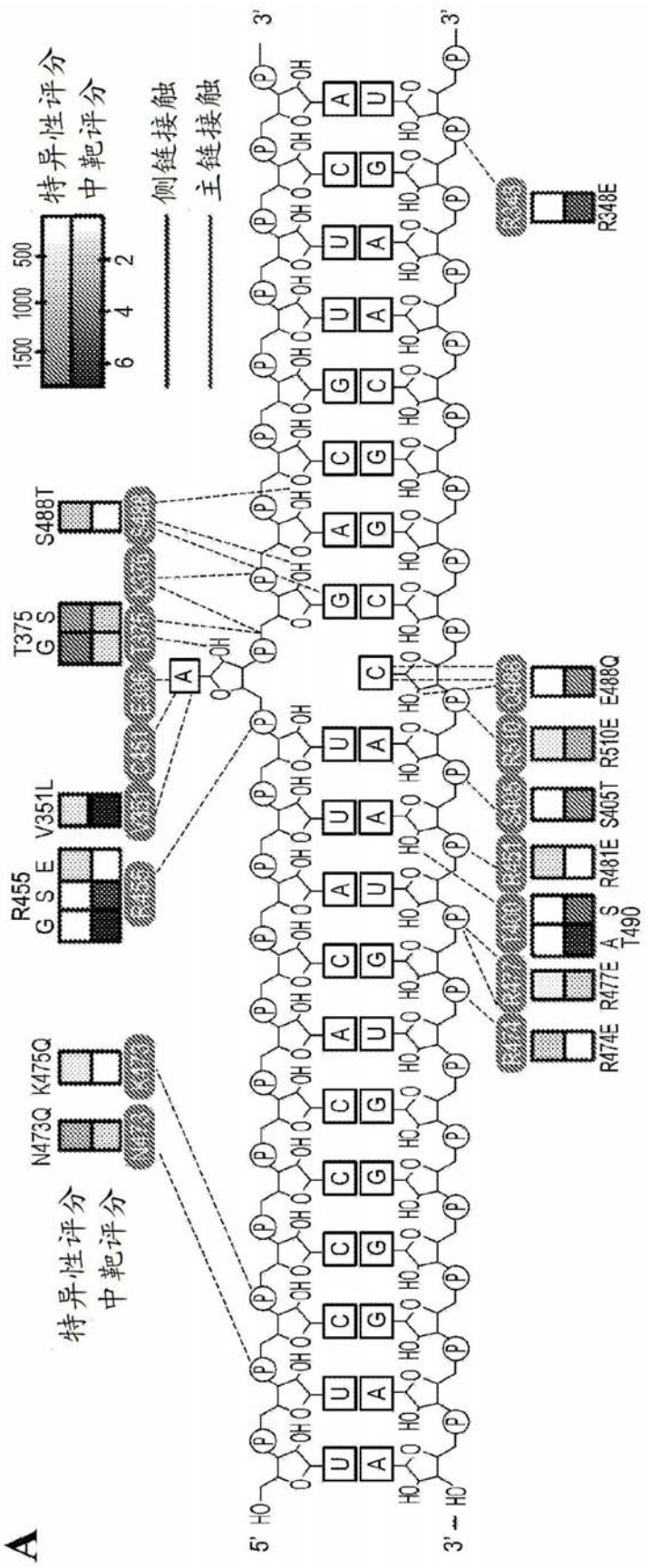

A
特异性评分
中靶评分
1500 1000 500
6 4 2
侧链接触
主链接触
N473Q K475Q
R455 V351L
G S E
T375 S488T
G S
R474E R477E A S R481E S405T R510E E488Q
T490
R348E
5' HO
3' HO
3'
3'

图54A

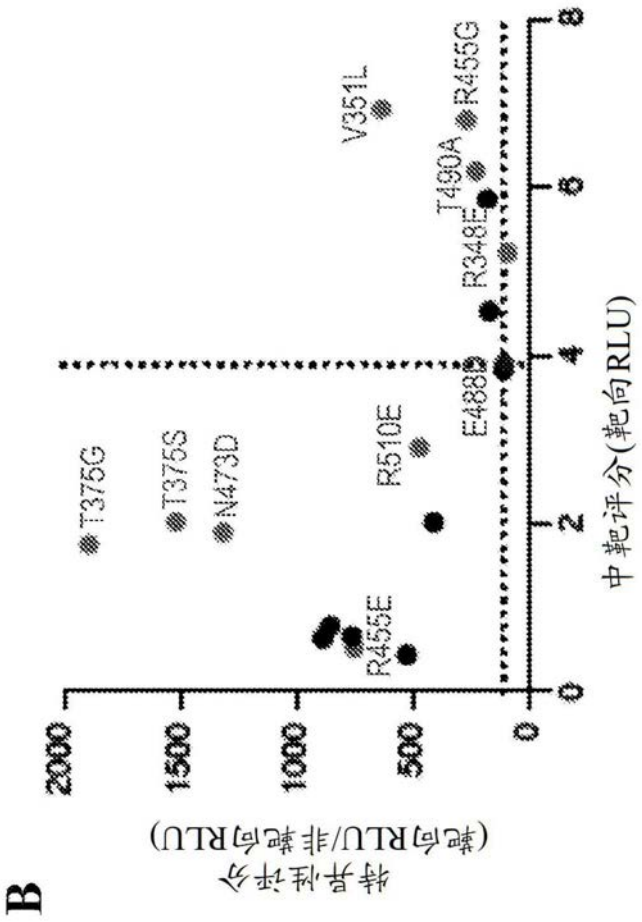

B
特异性评分
(靶向RLU/非靶向RLU)
2000
1500
1000
500
0
0
2
4
6
8
中靶评分(靶向RLU)
T375G
T375S
N473D
R510E
R455E
E488D
R348E
T490A
V351L
R455G

图54B

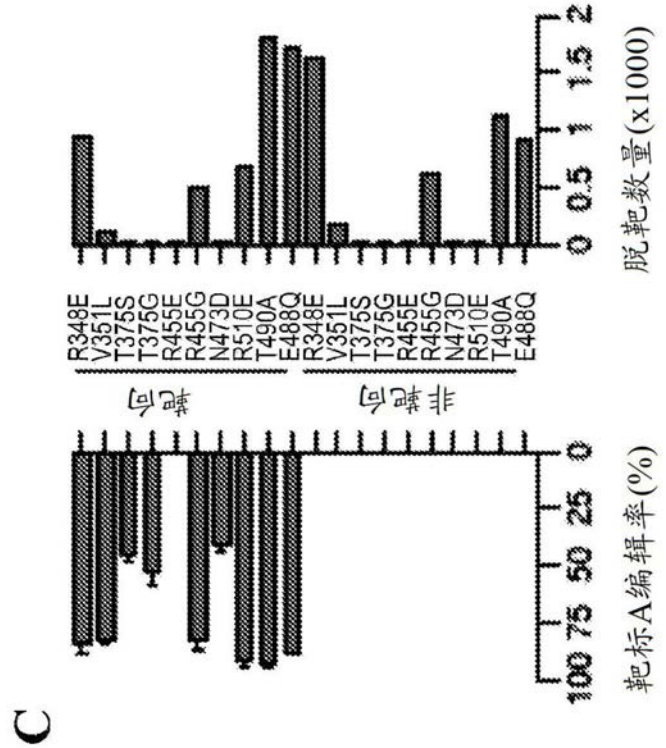

C
靶向
非靶向
R348E
V351L
T375S
T375G
R455E
R455G
N473D
R510E
T490A
E488Q
R348E
V351L
T375S
R455E
R455G
N473D
R510E
T490A
E488Q
100 75 50 25 0
0 0.5 1 1.5 2
靶标A编辑率(%)
脱靶数量(x1000)

图54C

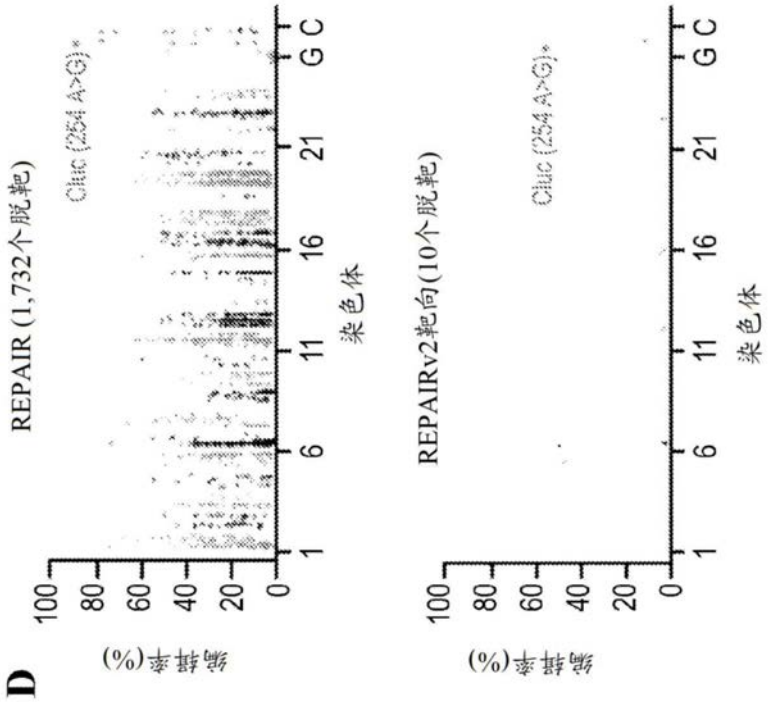

D
REPAIR (1,732个脱靶)
Gluc (254 A>G)*
编辑率(%)
100 80 60 40 20 0
1 6 11 16 21
染色体
G C
REPAIRv2靶向(10个脱靶)
Gluc (254 A>G)*
编辑率(%)
100 80 60 40 20 0
1 6 11 16 21
染色体
G C

图54D

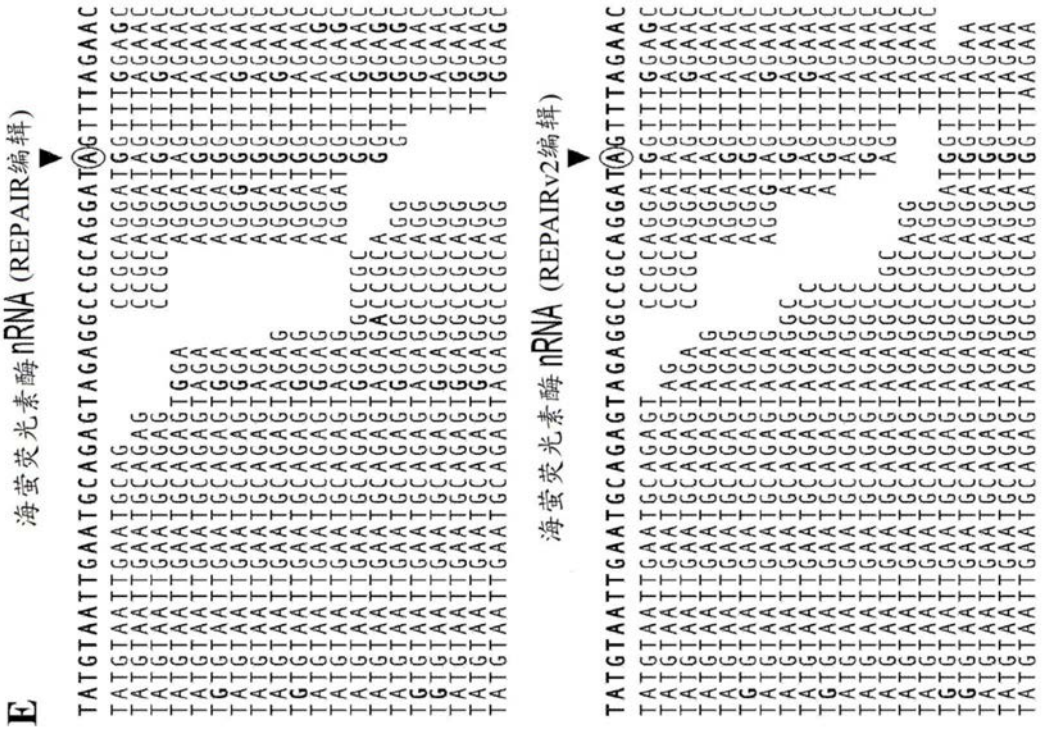



图54E

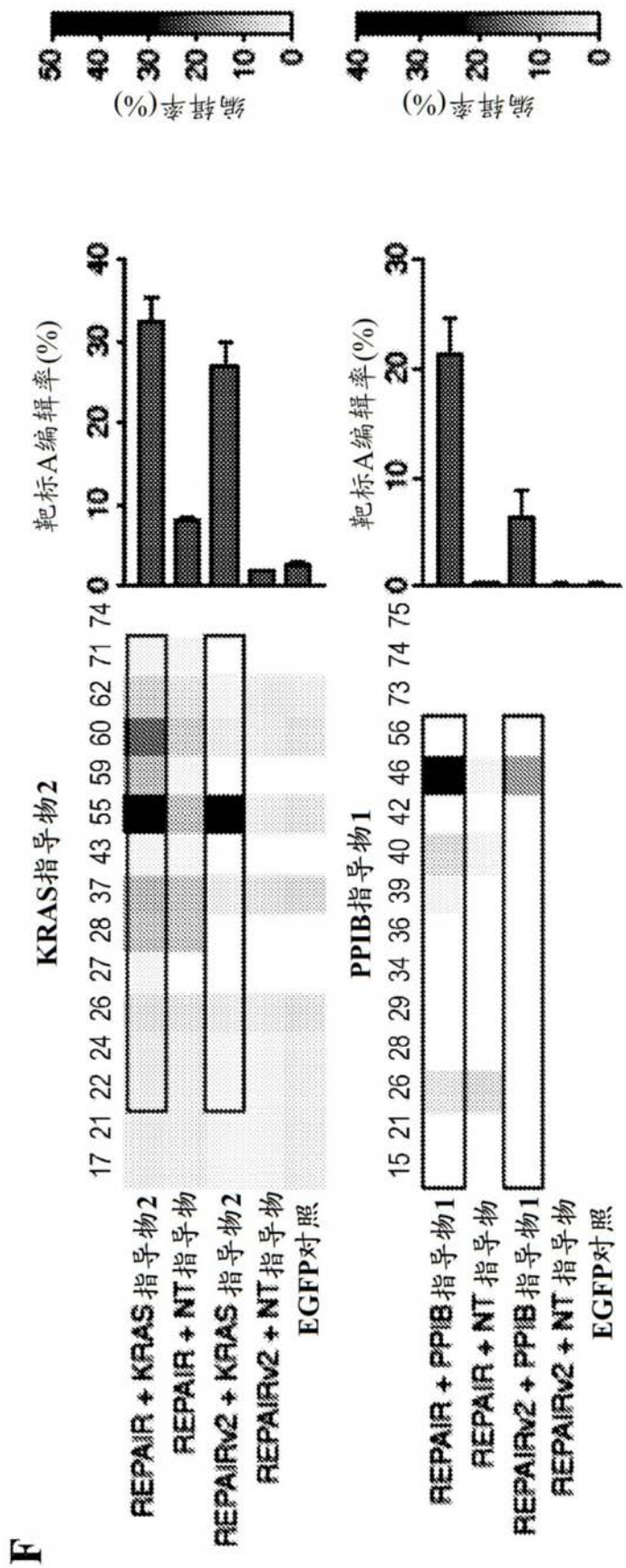

F
KRAS指导物2
PPIB指导物1
靶标A编辑率(%)
REPAIR + KRAS指导物2
REPAIR + NT指导物
REPAIRv2 + KRAS指导物2
REPAIRv2 + NT指导物
EGFP对照
REPAIR + PPIB指导物1
REPAIR + NT指导物
REPAIRv2 + PPIB指导物1
REPAIRv2 + NT指导物
EGFP对照

图54F

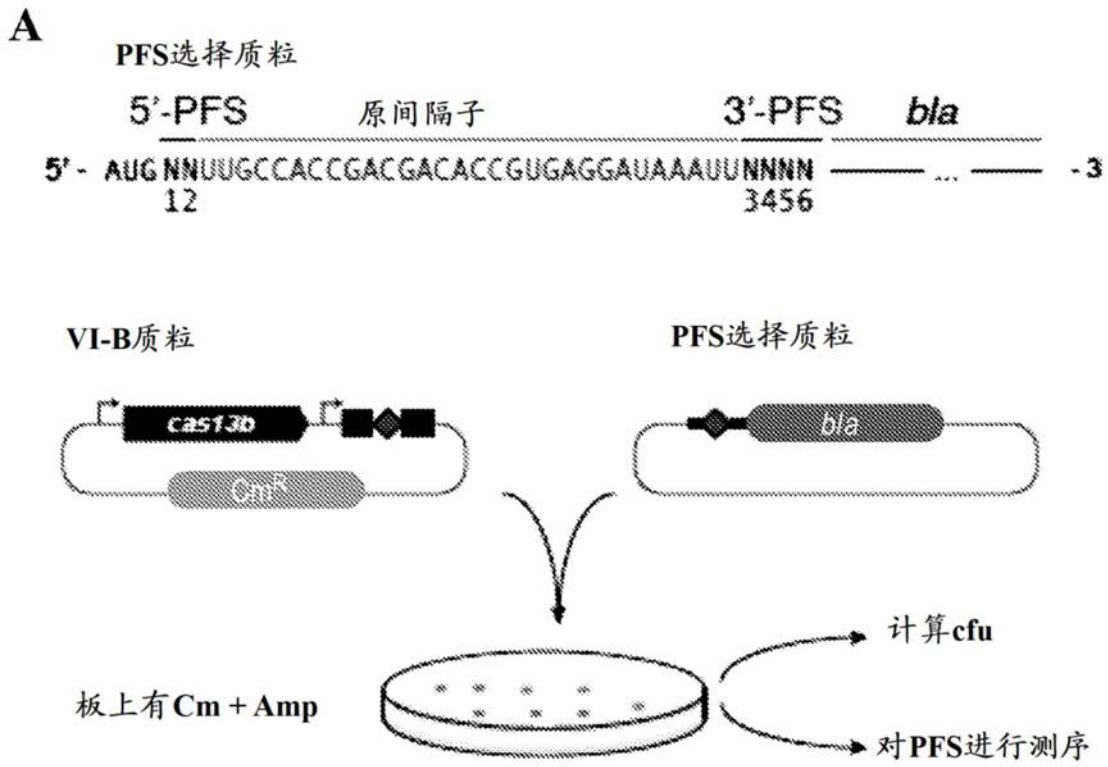

A
PFS选择质粒
5'-PFS
原间隔子
3'-PFS
bla
5'- AUG NNUUGCCACCGACGACACCGUGAGGAUAAAUU NNNN ——— ... ——— -3
12
3456
VI-B质粒
PFS选择质粒
cas13b
Cm^R
bla
计算cfu
板上有Cm + Amp
对PFS进行测序

图55A

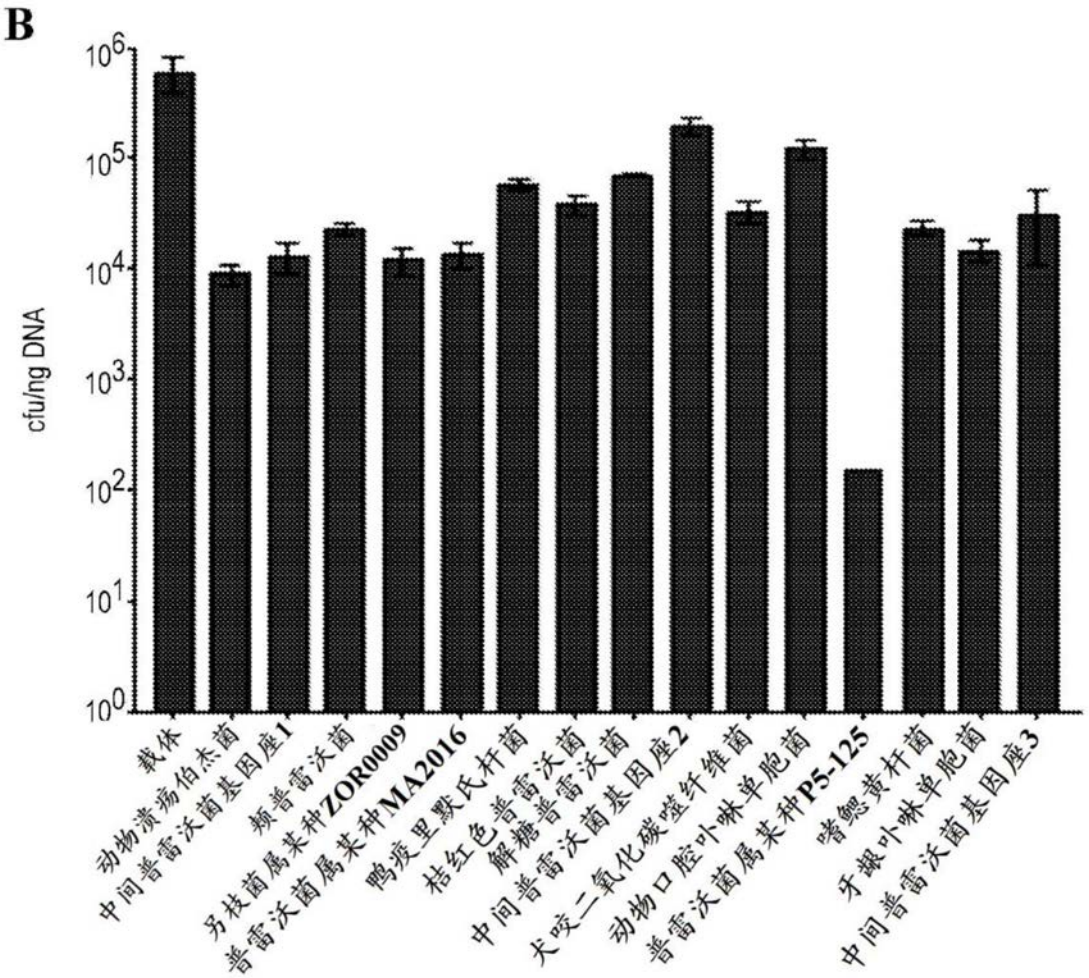

B
cfu/ng DNA
$10^6$
$10^5$
$10^4$
$10^3$
$10^2$
$10^1$
$10^0$
载体
动物溃疡伯杰菌
中间普雷沃菌基因座1
颊普雷沃菌
牙枝菌属某种ZOR0009
普雷沃菌属某种MA2016
鸭疫里默氏杆菌
桔红色普雷沃菌
解糖普雷沃菌
中间普雷沃菌基因座2
犬咬二氧化碳噬纤维菌
动物口腔卟啉单胞菌
普雷沃菌属某种P5-125
嗜鳃黄杆菌
牙龈卟啉单胞菌
中间普雷沃菌基因座3

图55B

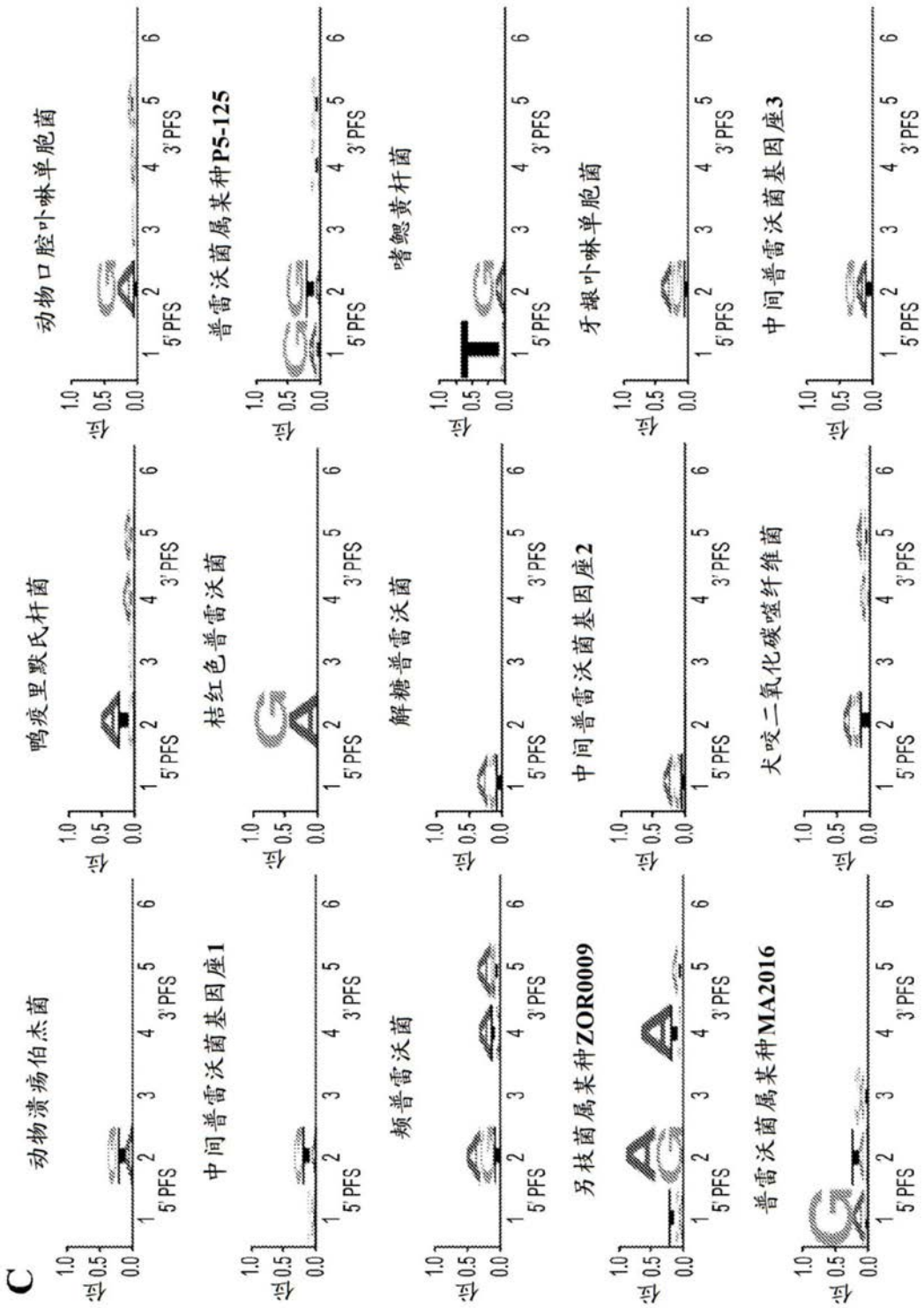

C
动物溃疡伯杰菌
鸭疫里默氏杆菌
动物口腔卟啉单胞菌
中间普雷沃菌基因座1
桔红色普雷沃菌
普雷沃菌属某种P5-125
颊普雷沃菌
解糖普雷沃菌
嗜鳃黄杆菌
另枝菌属某种ZOR0009
中间普雷沃菌基因座2
牙龈卟啉单胞菌
普雷沃菌属某种MA2016
犬咬二氧化碳噬纤维菌
中间普雷沃菌基因座3
5' PFS
3' PFS
比率

图55C

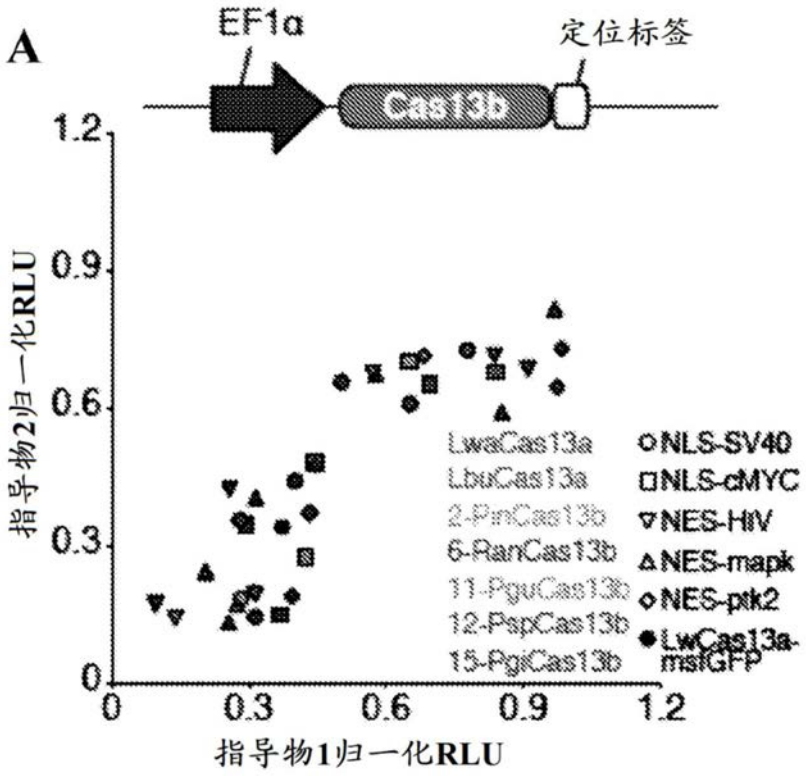

A
EF1α
Cas13b
定位标签
1.2
0.9
0.6
0.3
0
0
0.3
0.6
0.9
1.2
指导物2归一化RLU
指导物1归一化RLU
LwaCas13a
LbuCas13a
2-PinCas13b
6-RanCas13b
11-PguCas13b
12-PspCas13b
15-PgiCas13b
NLS-SV40
NLS-cMYC
NES-HIV
NES-mapk
NES-ptk2
LwCas13a-msfGFP

图56A

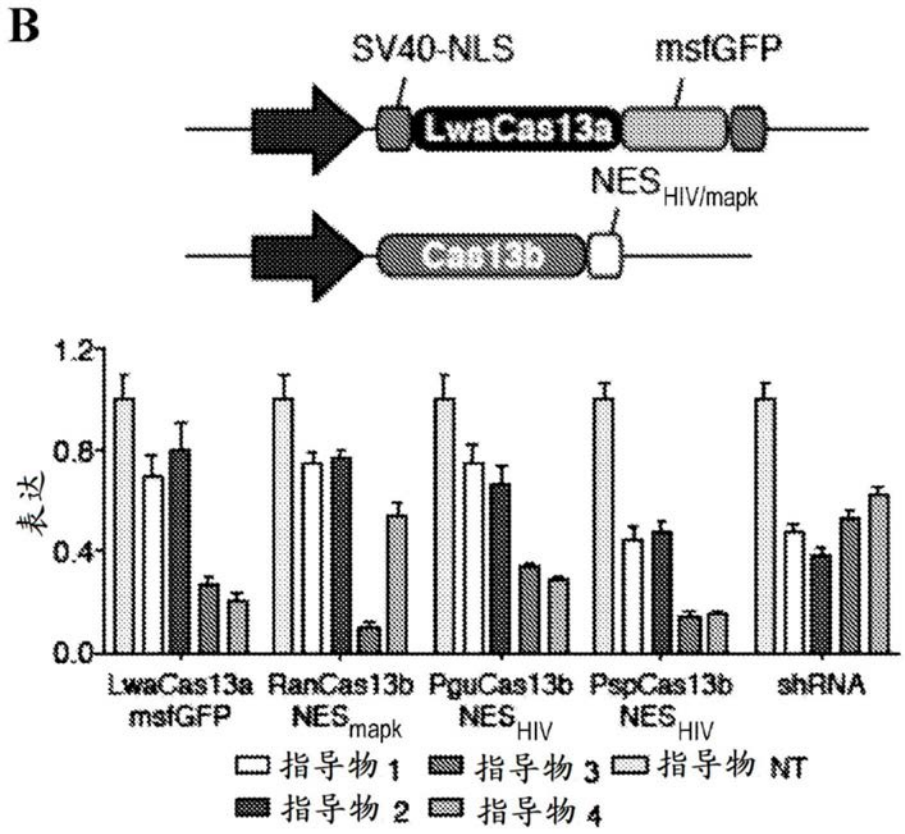

B
SV40-NLS
msfGFP
LwaCas13a
NES HIV/mapk
Cas13b
1.2
0.8
0.4
0.0
表达
LwaCas13a msfGFP
RanCas13b NES mapk
PguCas13b NES HIV
PspCas13b NES HIV
shRNA
指导物 1
指导物 3
指导物 NT
指导物 2
指导物 4

图56B

C


5' PFS
PspCas13b靶标
3' PFS
UGGAGGUGCUCAAAGAGAUGGAAGCCAAUGCCCGGA
所有单错配
NGGAGGUGCUCAAAGAGAUGGAAGCCAAUGCCCGGA
UNGAGGUGCUCAAAGAGAUGGAAGCCAAUGCCCGGA
UGNAGGUGCUCAAAGAGAUGGAAGCCAAUGCCCGGA
UGGNGGUGCUCAAAGAGAUGGAAGCCAAUGCCCGGA
UGGAGGUGCUCAAAGAGNUGGAAGCCAAUGCCCGGA
UGGAGGUGCUCAAAGAGAUGGAAGCCAAUGCCCGGN
所有双错配
NNGAGGUGCUCAAAGAGAUGGAAGCCAAUGCCCGGA
NGNAGGUGCUCAAAGAGAUGGAAGCCAAUGCCCGGA
NGGNGGUGCUCAAAGAGAUGGAAGCCAAUGCCCGGA
NGGANGUGCUCAAAGAGAUGGAAGCCAAUGCCCGGA
UGGAGGUGCUCAANGAGNUGGAAGCCAAUGCCCGGA

图56C

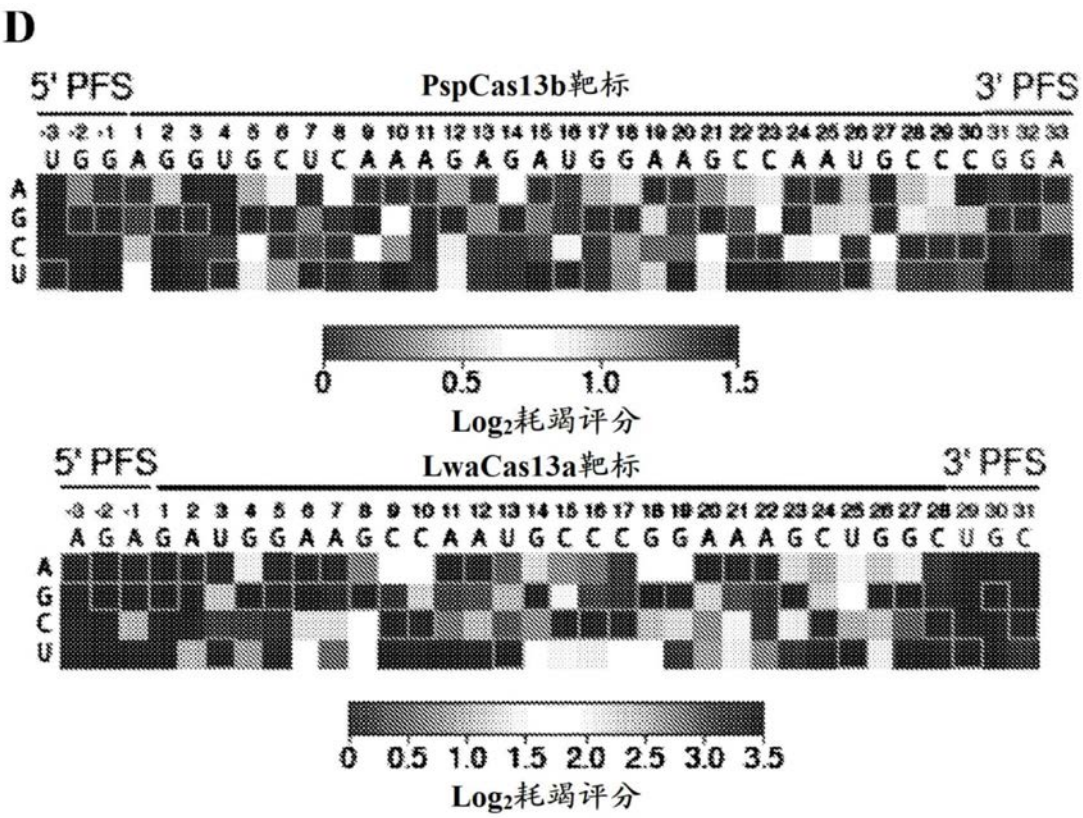

D
5' PFS
PspCas13b靶标
3' PFS
-3 -2 -1 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33
U G G A G G U G C U C A A A G A G A U G G A A G C C A A U G C C C G G A
A
G
C
U
0 0.5 1.0 1.5
Log2耗竭评分
5' PFS
LwaCas13a靶标
3' PFS
-3 -2 -1 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31
A G A G A U G G A A G C C A A U G C C C G G A A A G C U G G C U G C
A
G
C
U
0 0.5 1.0 1.5 2.0 2.5 3.0 3.5
Log2耗竭评分

图56D

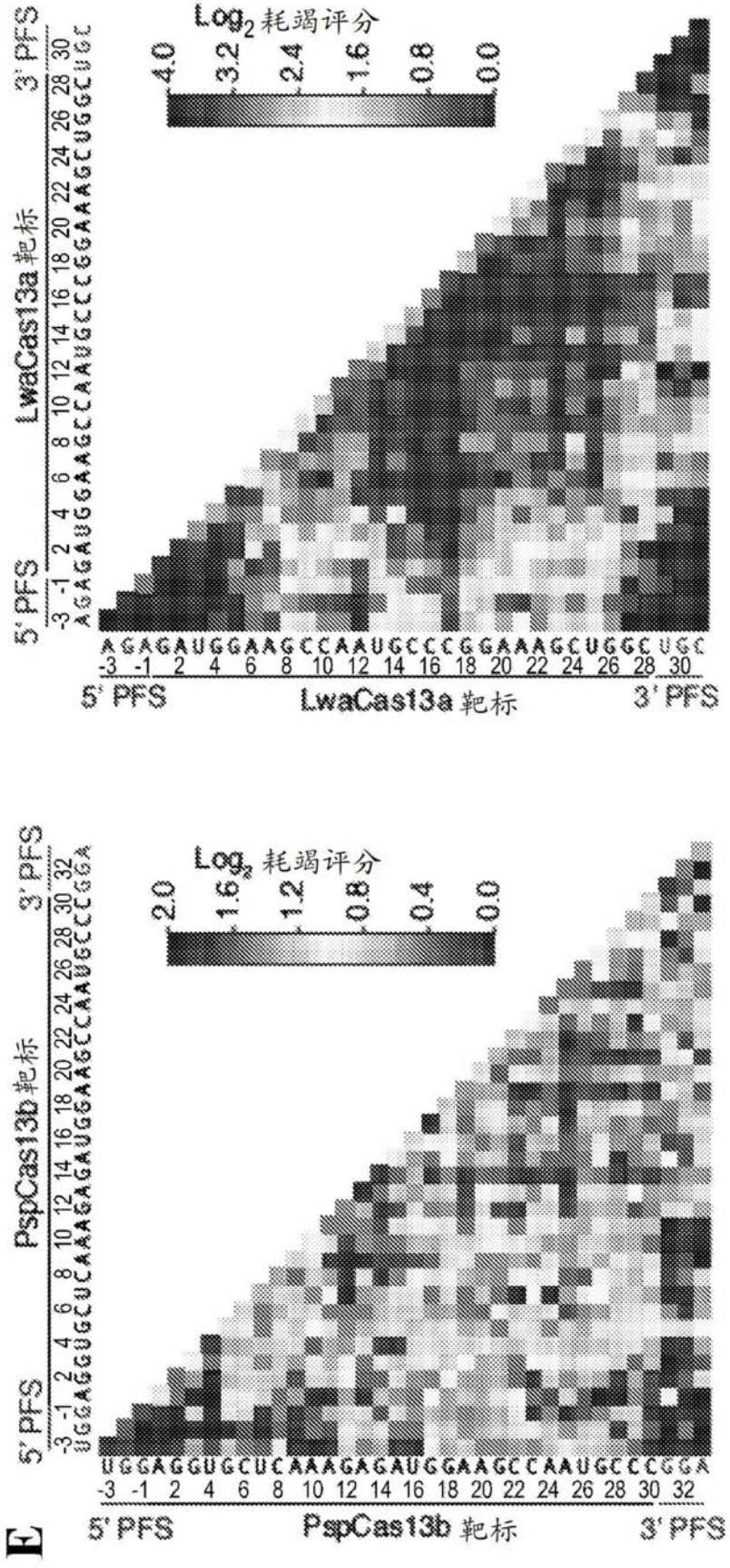

Log2 耗竭评分
4.0
3.2
2.4
1.6
0.8
0.0
5' PFS
LwaCas13a 靶标
3' PFS
AGAGAUGGAAGCCAAUGCCCGGAAAGCUGGCUGC
-3 -1 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30
Log2 耗竭评分
2.0
1.6
1.2
0.8
0.4
0.0
5' PFS
PspCas13b 靶标
3' PFS
UGGAGGUGCUCAAAGAGAUGGAAGCCAAUGCCCGGA
-3 -1 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32
E

图56E

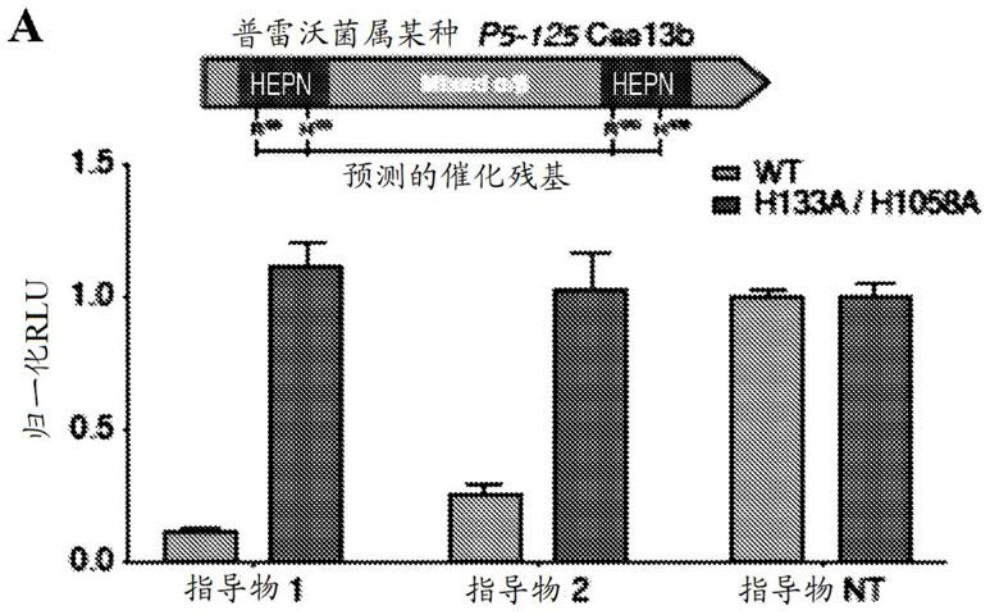

A
普雷沃菌属某种 P5-125 Cas13b
HEPN
HEPN
预测的催化残基
WT
H133A/H1058A
归一化RLU
1.5
1.0
0.5
0.0
指导物 1
指导物 2
指导物 NT

图57A

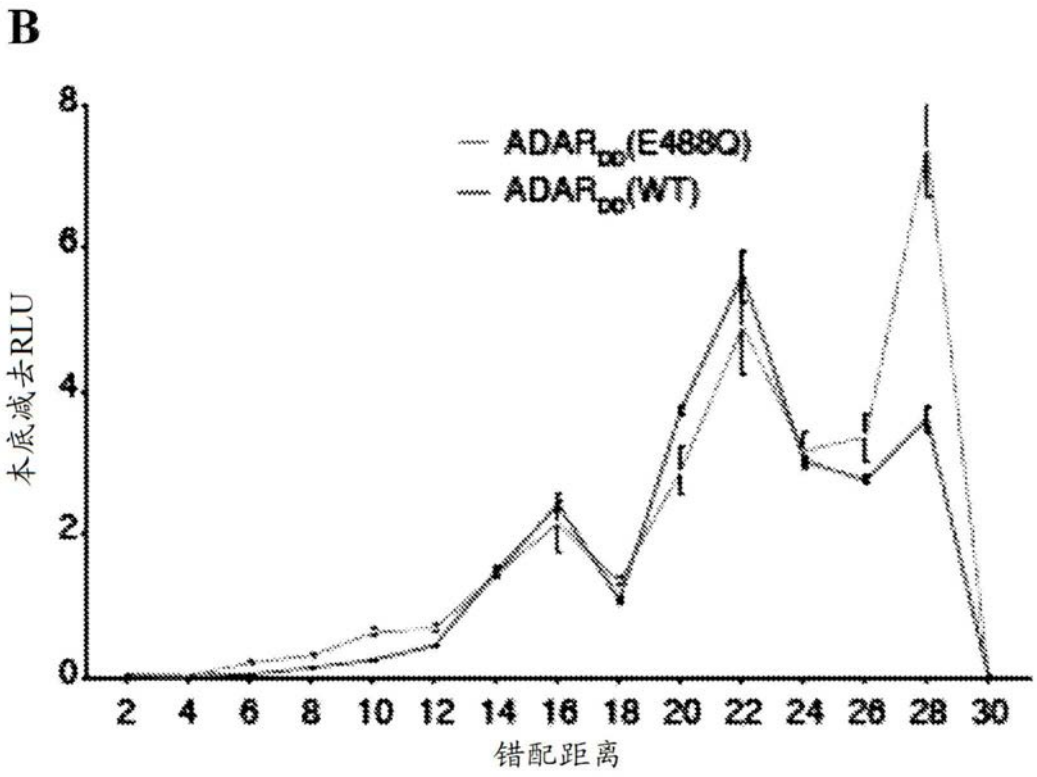

B
ADARDD(E488Q)
ADARDD(WT)
本底减去RLU
8
6
4
2
0
2 4 6 8 10 12 14 16 18 20 22 24 26 28 30
错配距离

图57B

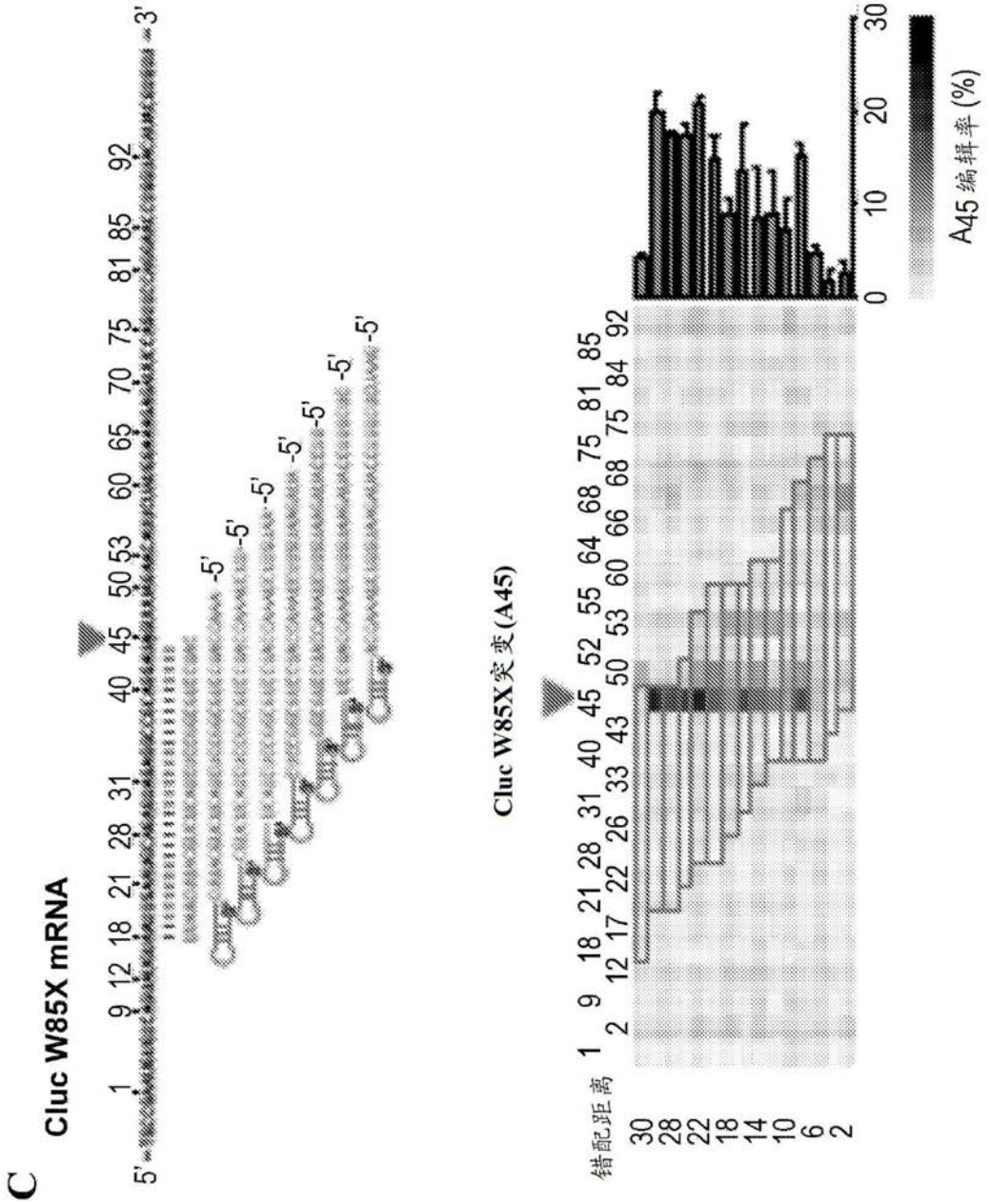



图57C

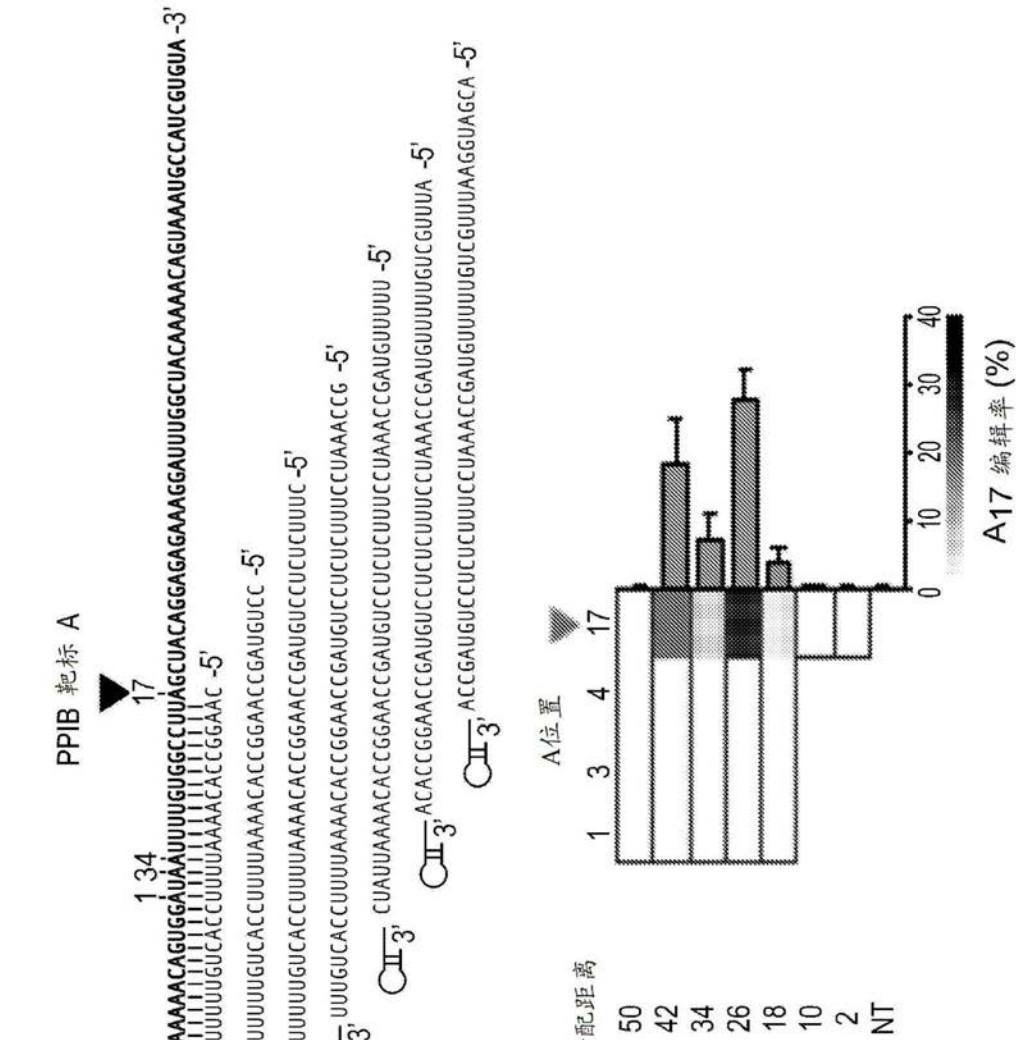

D
PPIB mRNA
PPIB 靶标 A
1 34
17
5'- UUGGUCUCUUCGGAAAGACUGUUCCAAAAACAGUGGAUAAUUUUGUGGCCUUAGCUACAGGAGAAGGAUUUGGCUACAAAAACAGUAAAUGCCAUCGUGUA -3'
3' CAGAGAAGCCUUUCUGACAAGGUUUUUGUCACCUUUUUAAAAACACCGGAAC -5'
3' CCUUUCUGACAAGGUUUUUGUCACCUUUUUAAAAACACCGGAACCGAUGUCC -5'
3' ACAAGGUUUUUGUCACCUUUUUAAAAACACCGGAACCGAUGUCCUCCUCUUUC -5'
3' UUUGUCACCUUUUUAAAAACACCGGAACCGAUGUCCUCCUUUCCUAAACCG -5'
3' CUAUUAAAACACCGGAACCGAUGUCCUCCUUUCCUAAACCGAUGUUUUU -5'
3' ACACCGGAACCGAUGUCCUCCUUUCCUAAACCGAUGUUUUUGUCCGUUUA -5'
3' ACCGAUGUCCUCCUUUCCUAAACCGAUGUUUUUGUCGUUUUAAGGUAGCA -5'
A位置
1 3 4 17
错配距离
50
42
34
26
18
10
2
NT
0 10 20 30 40
A17 编辑率 (%)

图57D

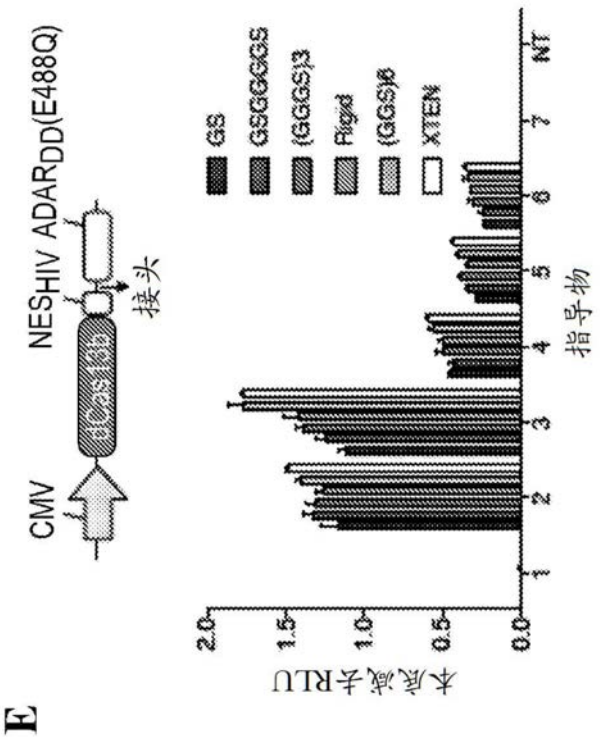

E
CMV
dCas13b
NES HIV ADAR DD(E488Q)
接头
GS
GSGGGGS
(GGGS)3
Rigid
(GGGS)6
XTEN
NT
指导物
0.0
0.5
1.0
1.5
2.0
1
2
3
4
5
6
7
本底减去RLU

图57E

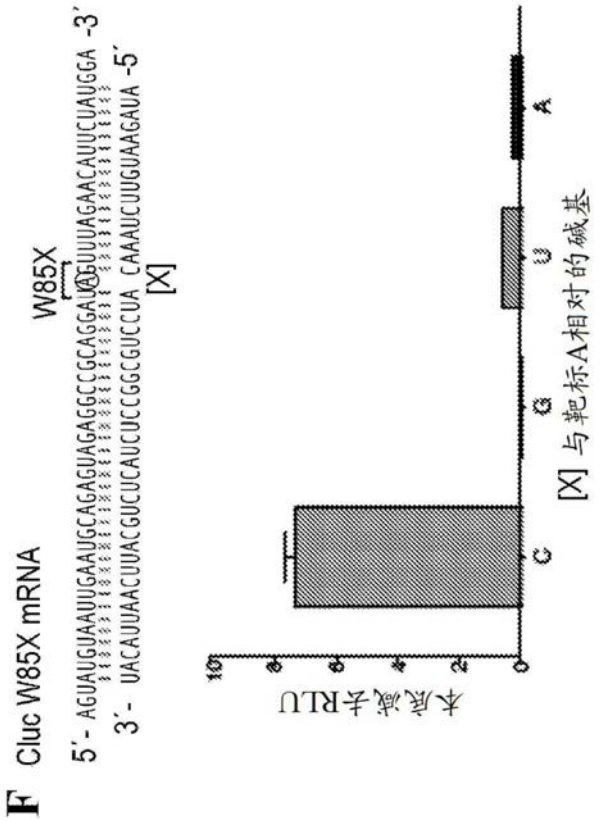

F
Cluc W85X mRNA
W85X
5′- AGUAUGUAAUUGAAUGCAGAGUAGAGGCCGCAGGAUGGUUUAGAACAUUCUAUGGA -3′
3′- UACAUUAACUUACGUCUCAUCUCCGGCGUCCUA CAAAUCUUGUAAGAUA -5′
[X]
本底减去RLU
0
2
4
6
8
10
C
G
U
A
[X] 与靶标A相对的碱基

图57F


A
G->A变化的ClinVar基序分布
基序数量
0
100
200
300
400
500
600
GAU
CAG
UAU
CAA
CAC
AAC
AAA
UAC
GAG
UAA
GAA
UAG
GAC
AAG
AAU
CAU

图58

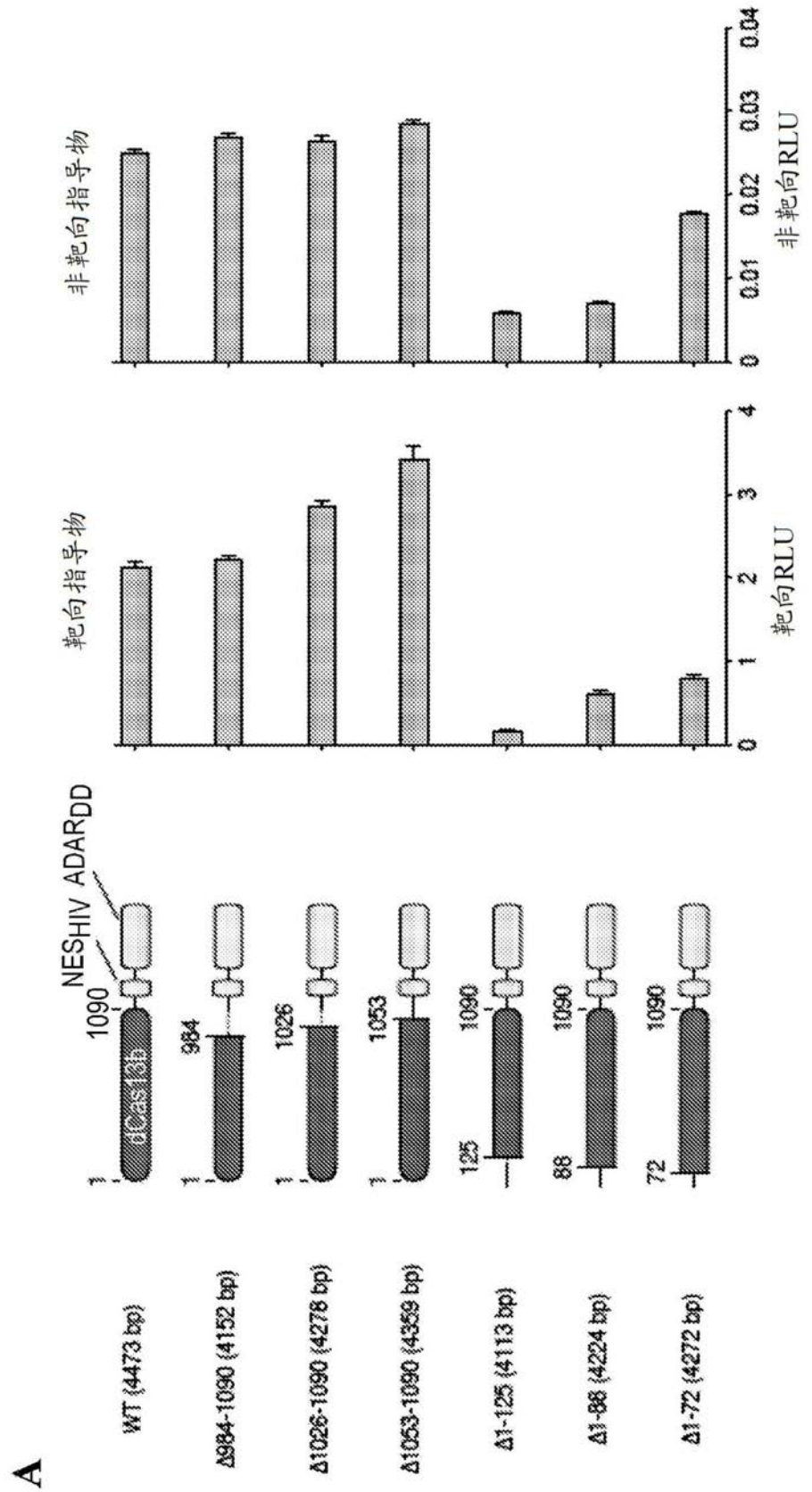

A
WT (4473 bp)
Δ984-1090 (4152 bp)
Δ1026-1090 (4278 bp)
Δ1053-1090 (4359 bp)
Δ1-125 (4113 bp)
Δ1-88 (4224 bp)
Δ1-72 (4272 bp)
NESHIV ADARDD
dCas13b
1090
984
1026
1053
125
88
72
靶向指导物
非靶向指导物
靶向RLU
非靶向RLU
0
1
2
3
4
0.01
0.02
0.03
0.04

图59

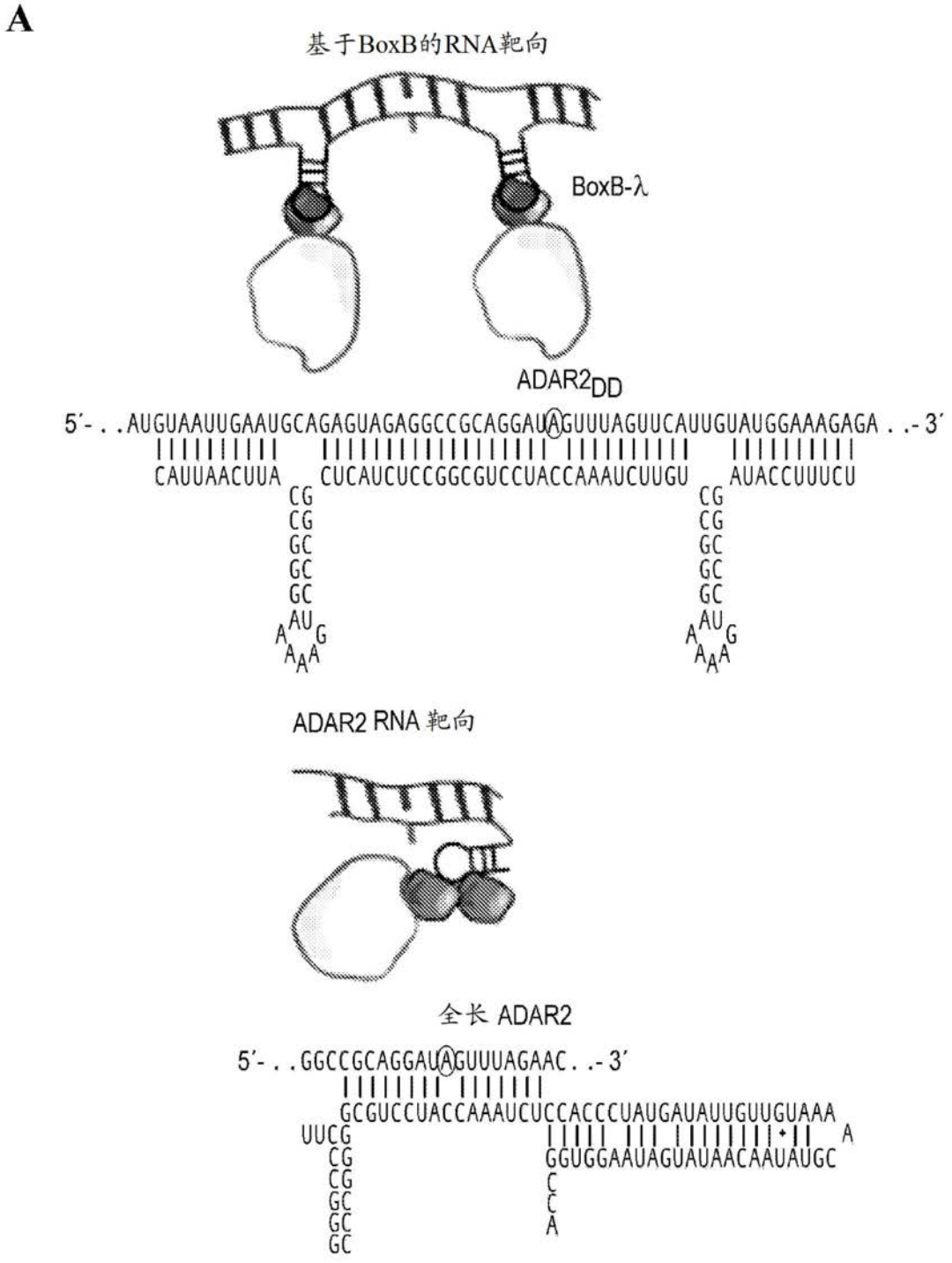

A
基于BoxB的RNA靶向
BoxB-λ
ADAR2DD
5′-..AUGUAAUUGAAUGCAGAGUAGAGGCCGCAGGAUAGUUUAGUUCAUUGUAUGGAAAGAGA..-3′
CAUUAACUUA CUCAUCUCCGGCGUCCUACCAAAUCUUGU AUACCUUUCU
ADAR2 RNA 靶向
全长 ADAR2
5′-..GGCCGCAGGAUAGUUUAGAAC..-3′
GCGUCCUACCAAAUCUCCACCCUAUGAUAUUGUUGUAAA
UUCG A
GGUGGAAUAGUAUAACAAUAUGC

图60A

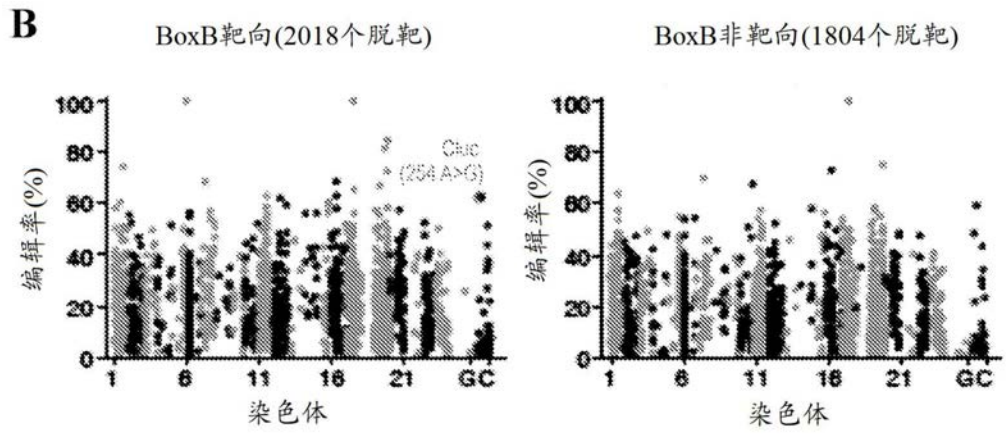



图60B

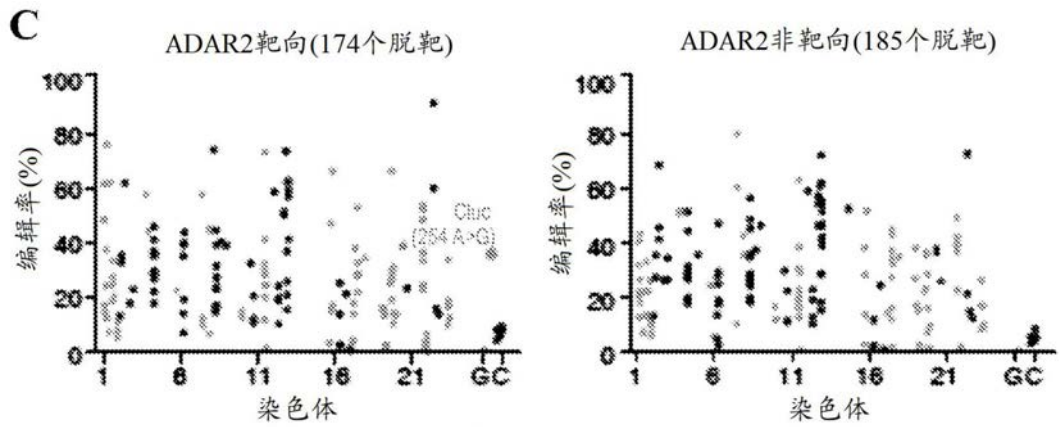



图60C

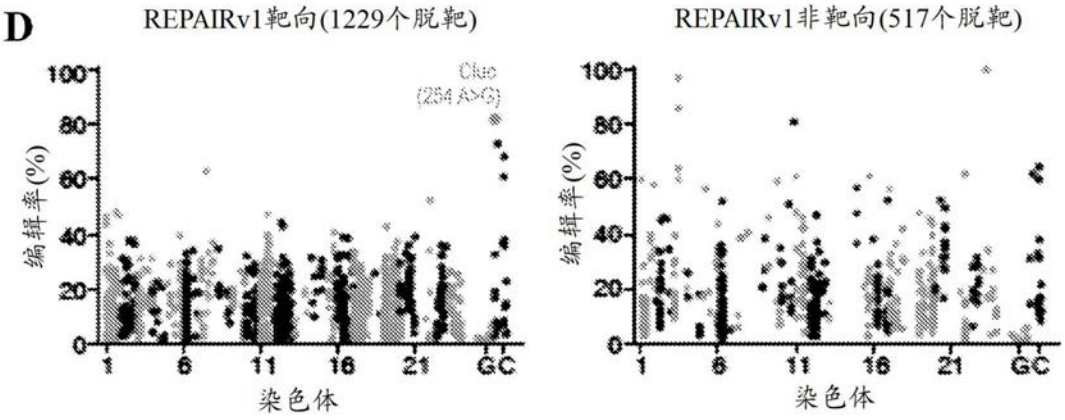



图60D

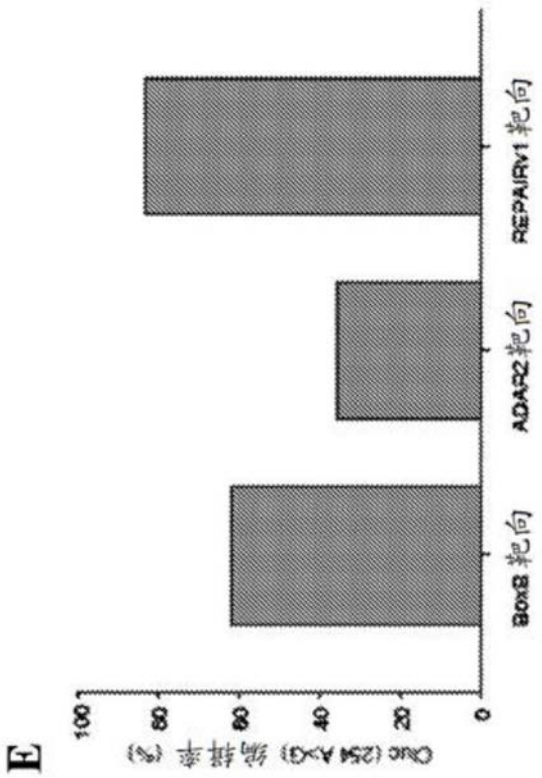

E
Cluc (25μM AMG) 编辑率 (%)
100
80
60
40
20
0
BoxB 靶向
ADAR2 靶向
REPAIRv1 靶向

图60E

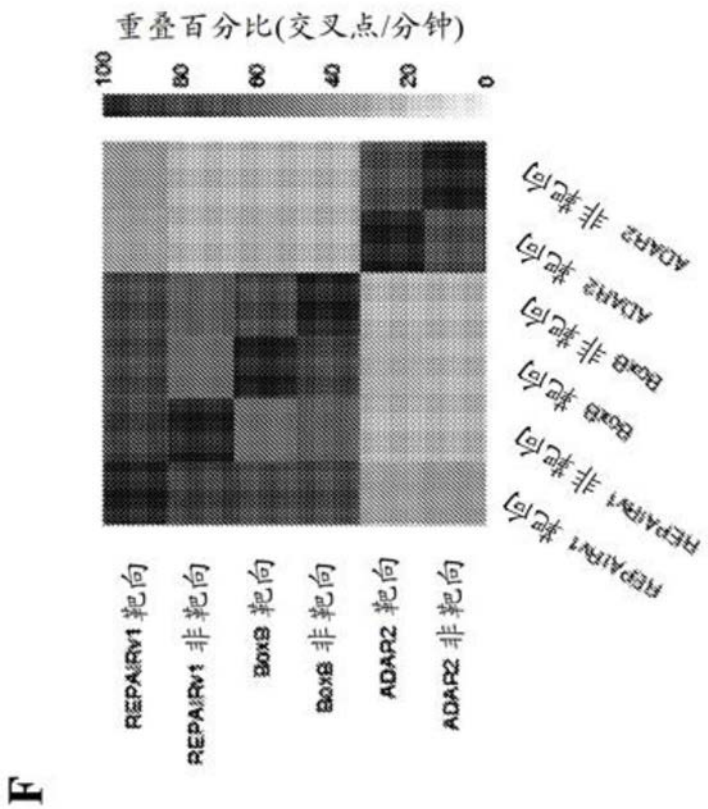

F
重叠百分比(交叉点/分钟)
100
80
60
40
20
0
REPAIRv1 靶向
REPAIRv1 非靶向
BoxB 靶向
BoxB 非靶向
ADAR2 靶向
ADAR2 非靶向
REPAIRv1 靶向
REPAIRv1 非靶向
BoxB 靶向
BoxB 非靶向
ADAR2 靶向
ADAR2 非靶向

图60F

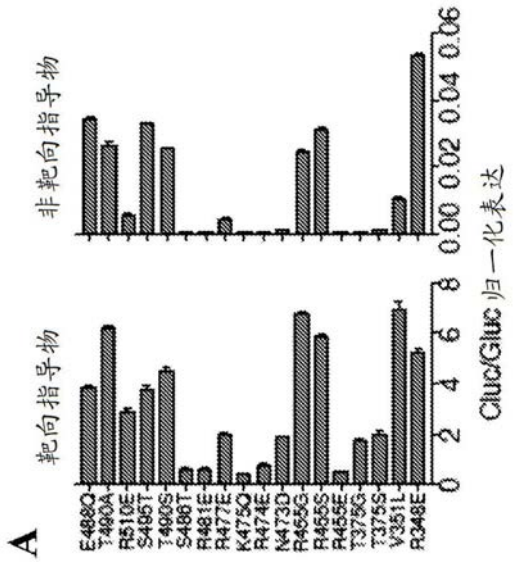

A
靶向指导物
非靶向指导物
E488Q
T490A
R510E
S485T
T490S
S486T
R481E
R477E
K475Q
R474E
N473D
R455G
R455S
R455E
T375G
T375S
V351L
R348E
0 2 4 6 8
0.00 0.02 0.04 0.06
Cluc/Gluc 归一化表达

图61A

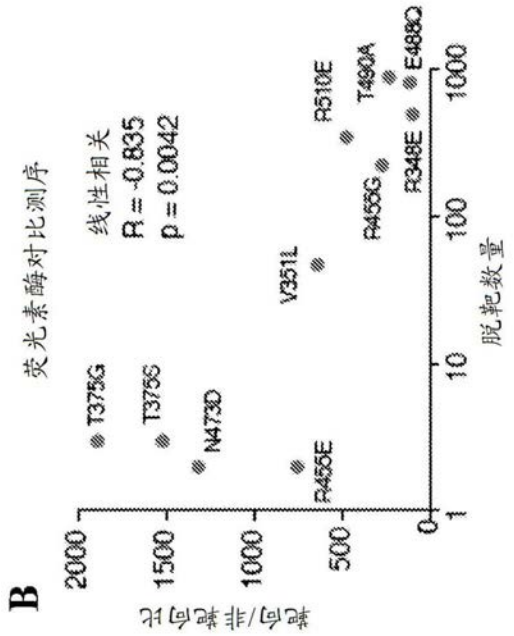

B
荧光素酶对比测序
线性相关
R = -0.835
p = 0.0042
T375G
T375S
N473D
R455E
V351L
R510E
R455G
T490A
R348E
E488Q
2000
1500
1000
500
0
靶向/非靶向比
1 10 100 1000
脱靶数量

图61B

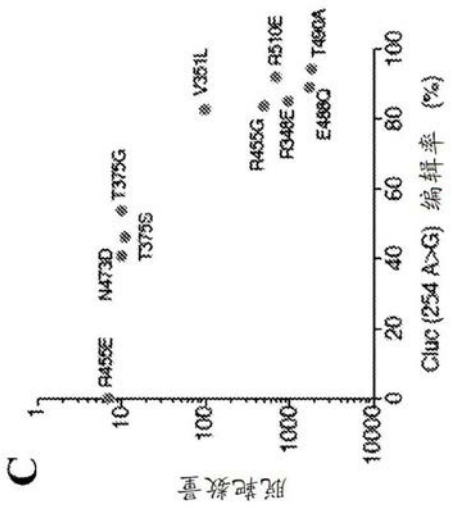

C
脱靶数量
1
10
100
1000
10000
0
20
40
60
80
100
Cluc (254 A>G) 编辑率 (%)
R455E
N473D
T375G
T375S
V351L
R455G
R348E
A510E
E488Q
T490A

图61C

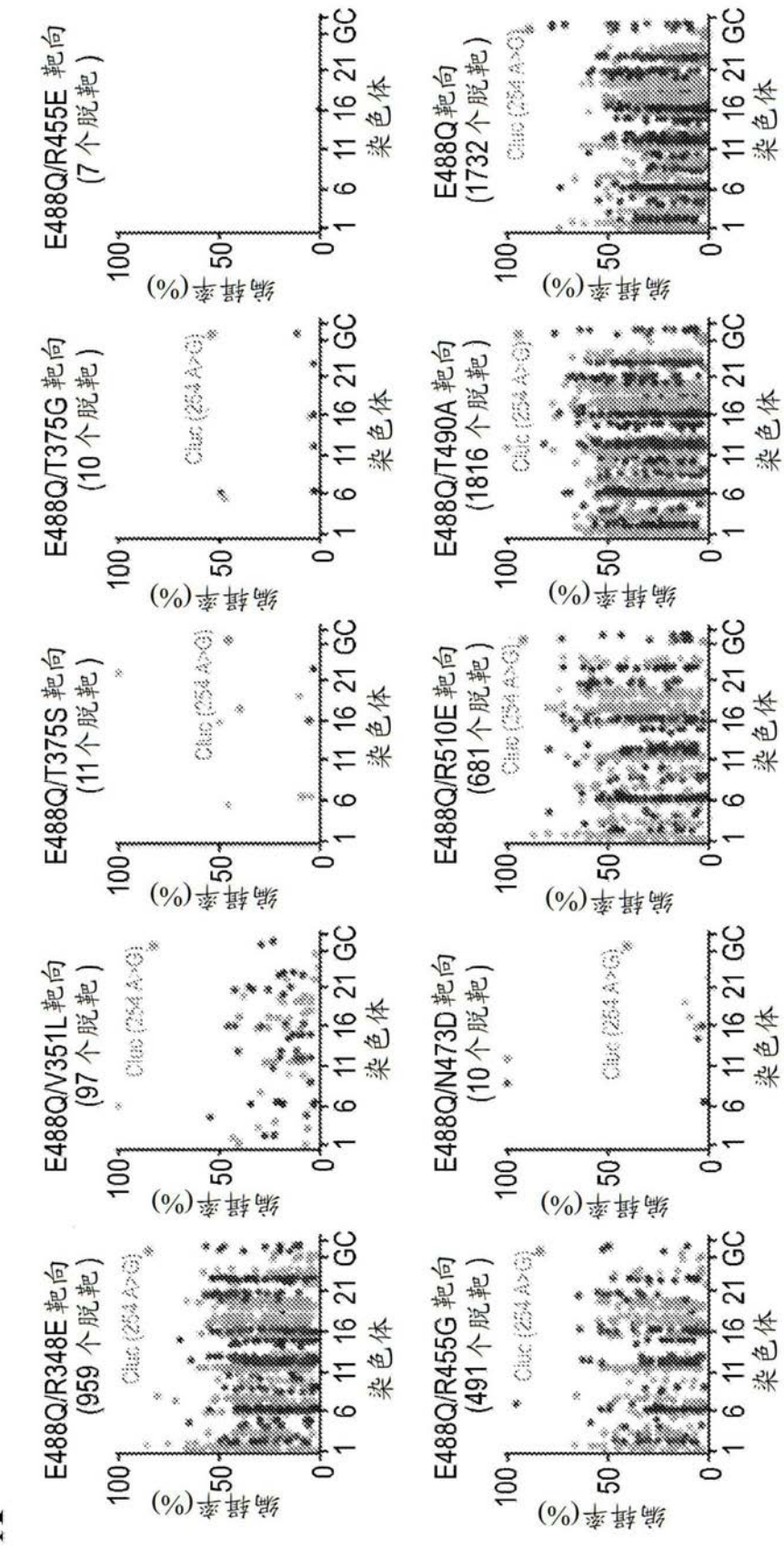

A
E488Q/R348E 靶向 (959 个脱靶)
E488Q/V351L 靶向 (97 个脱靶)
E488Q/T375S 靶向 (11 个脱靶)
E488Q/T375G 靶向 (10 个脱靶)
E488Q/R455E 靶向 (7 个脱靶)
E488Q/R455G 靶向 (491 个脱靶)
E488Q/N473D 靶向 (10 个脱靶)
E488Q/R510E 靶向 (681 个脱靶)
E488Q/T490A 靶向 (1816 个脱靶)
E488Q 靶向 (1732 个脱靶)
Chr6 (254 A>G)
编辑率(%)
100
50
0
1 6 11 16 21 GC
染色体

图62A

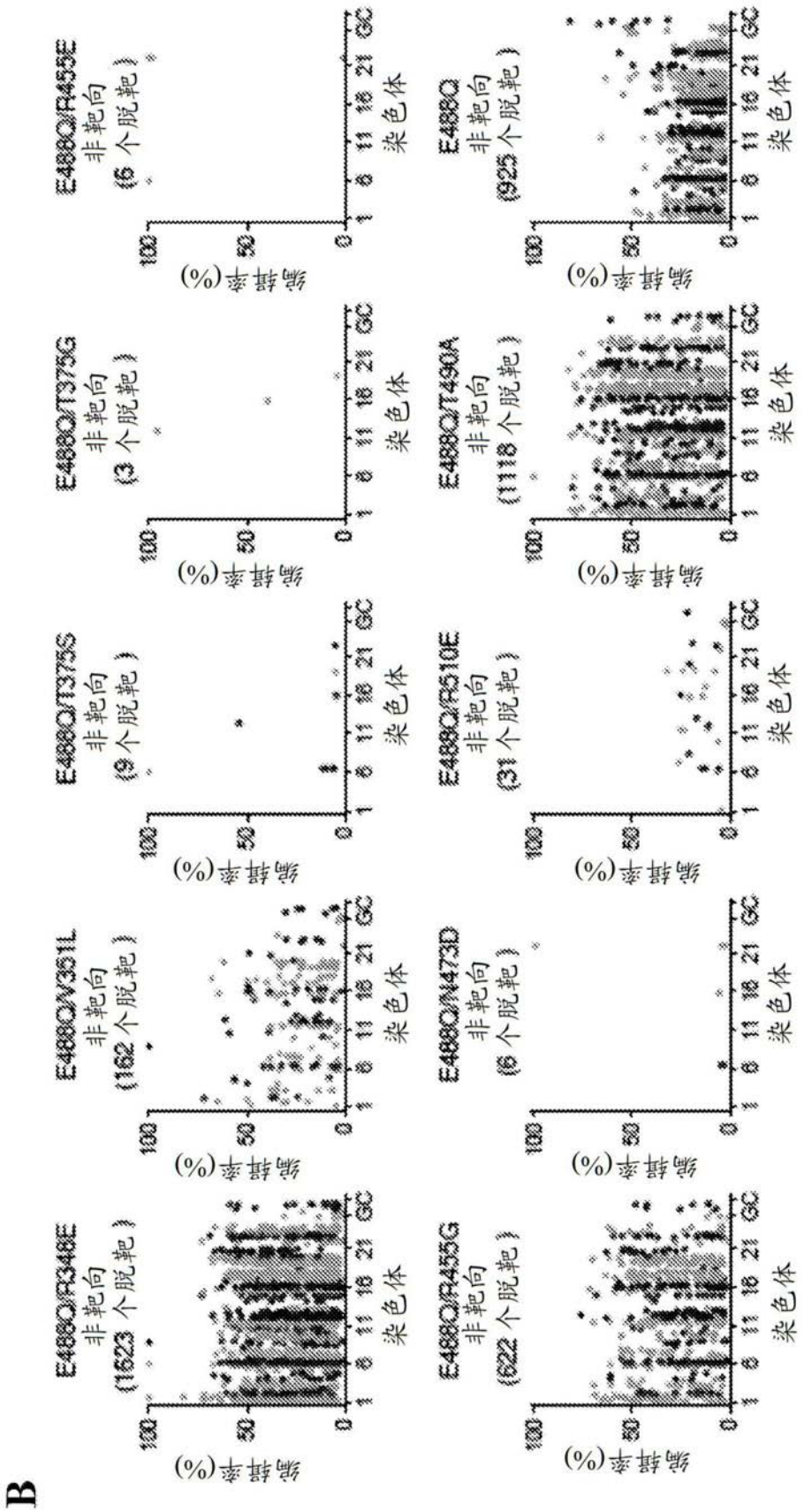

B
E488Q/R348E
非靶向
(1623 个脱靶 )
E488Q/V351L
非靶向
(162 个脱靶 )
E488Q/T375S
非靶向
(9 个脱靶 )
E488Q/T375G
非靶向
(3 个脱靶 )
E488Q/T490A
非靶向
(1118 个脱靶 )
E488Q/R510E
非靶向
(31 个脱靶 )
E488Q/N473D
非靶向
(6 个脱靶 )
E488Q/R455G
非靶向
(622 个脱靶 )
E488Q/R455E
非靶向
(6 个脱靶 )
E488Q
非靶向
(925 个脱靶 )
编辑率(%)
100
50
0
1 6 11 16 21 GC
染色体

图62B

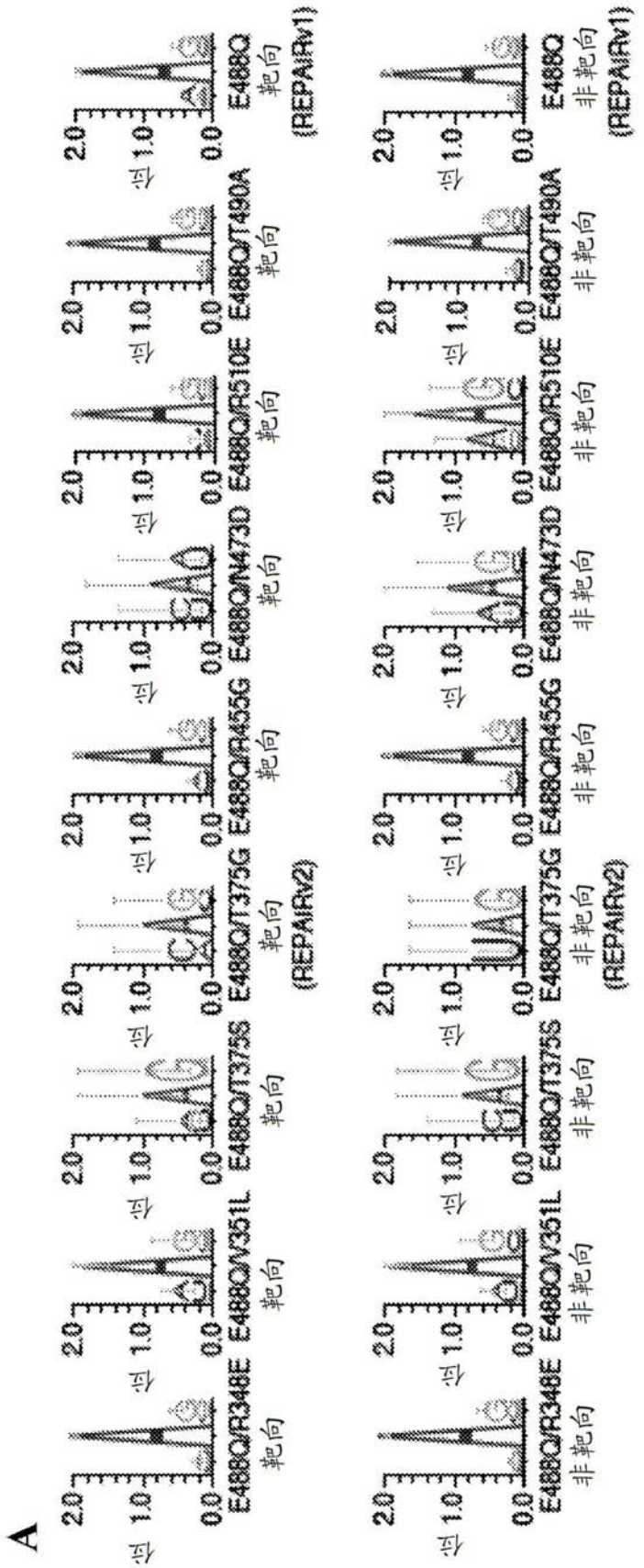

A
E488Q/R348E
E488Q/V351L
E488Q/T375S
E488Q/T375G
E488Q/R455G
E488Q/N473D
E488Q/R510E
E488Q/T490A
E488Q
非靶向
靶向
(REPAIRv2)
(REPAIRv1)

图63A

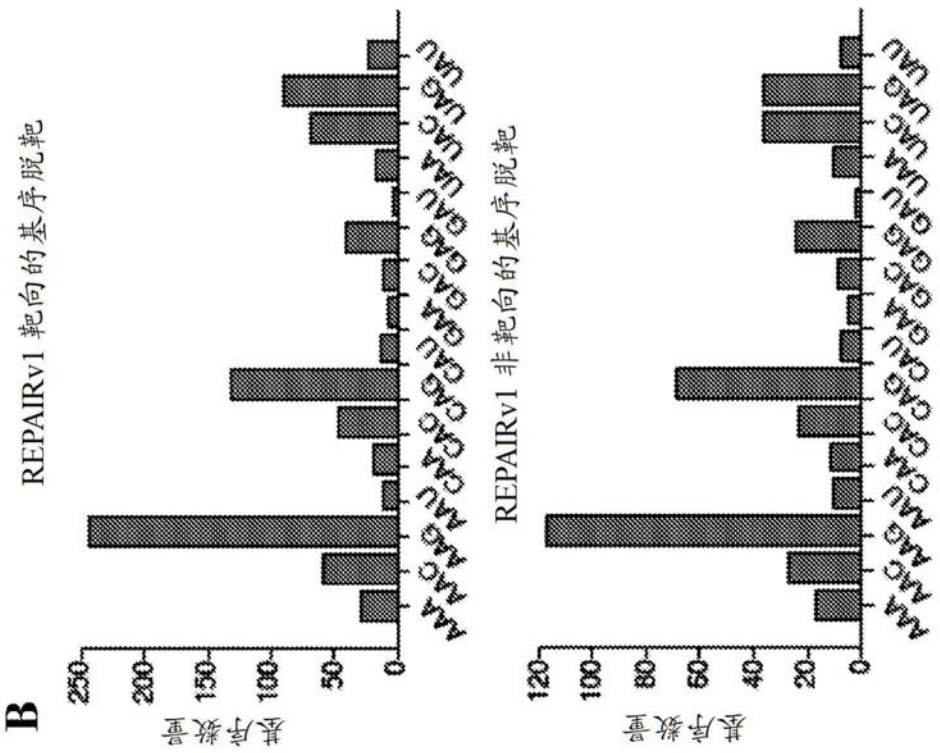

B
REPAIRv1 靶向的基序脱靶
REPAIRv1 非靶向的基序脱靶
基序数量
250
200
150
100
50
0
120
100
80
60
40
20
0
AAC AAG AAU CAC CAG CAU GAC GAG GAU UAC UAG UAU

图63B

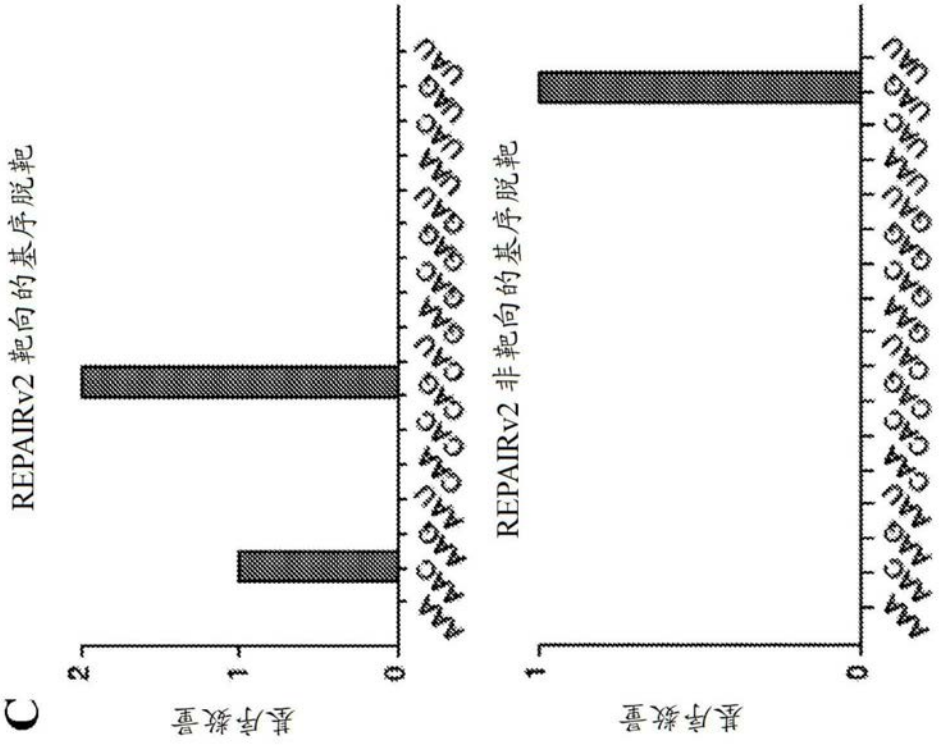

C
REPAIRv2 靶向的基序脱靶
REPAIRv2 非靶向的基序脱靶
基序数量
2
1
0
1
0
AAC AAG AAU CAC CAG CAU GAC GAG GAU UAC UAG UAU

图63C

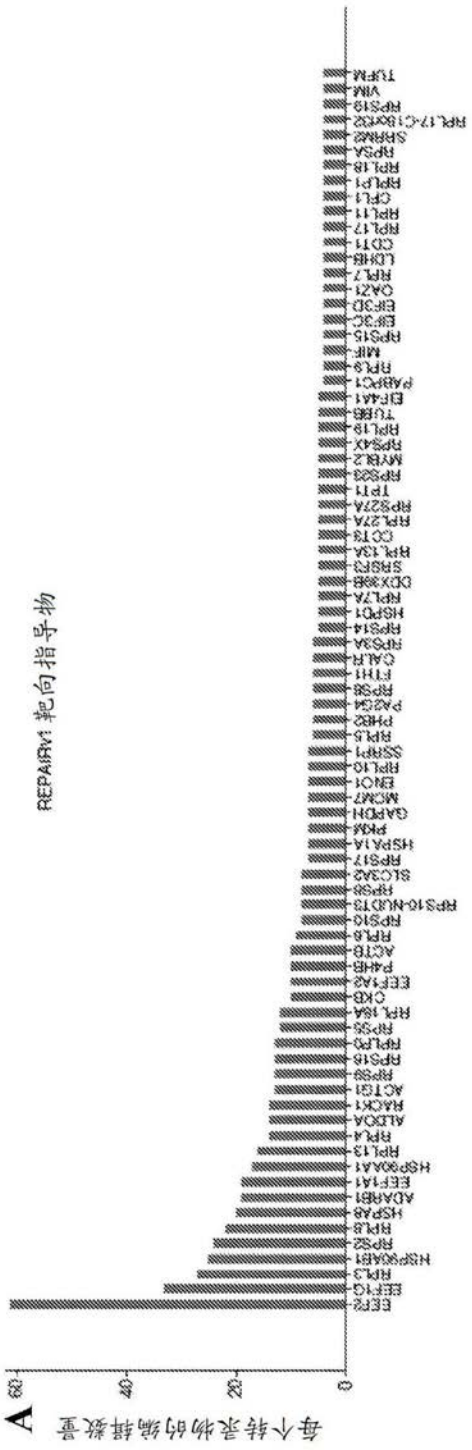



图64A

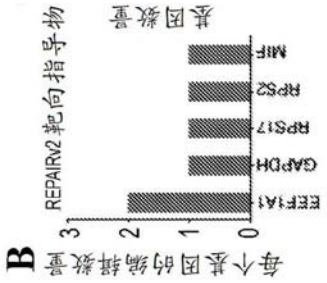



图64B

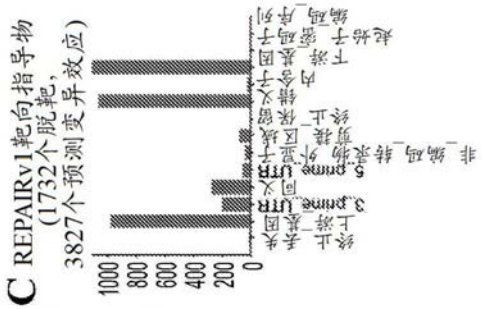



图64C

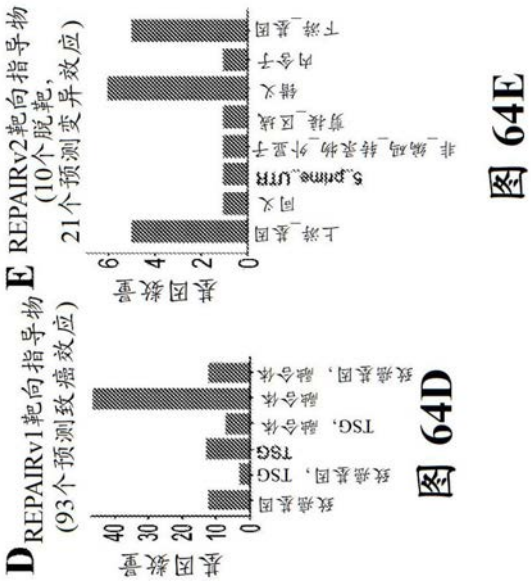



图 64D

图 64E

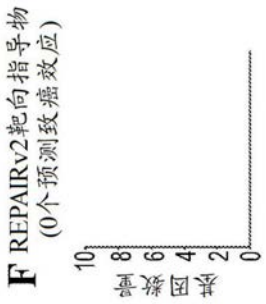

F REPAIRv2靶向指导物
(0个预测致癌效应)
基因数量
10
8
6
4
2
0

图64F

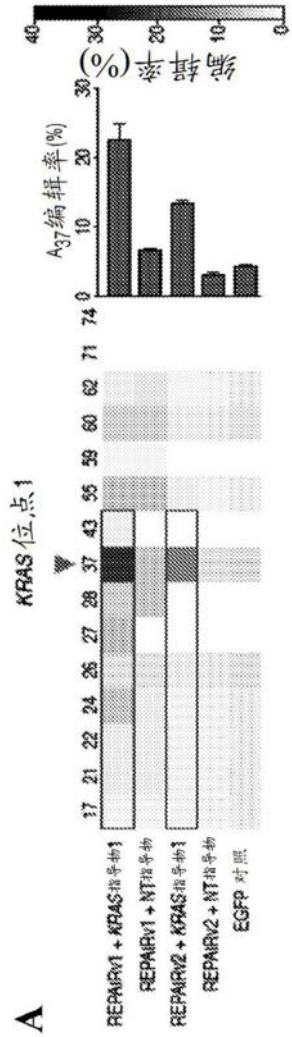

A
KRAS位点1
17 21 22 24 26 27 28 37 43 55 59 60 62 71 74
REPAIRv1 + KRAS指导物1
REPAIRv1 + NT指导物
REPAIRv2 + KRAS指导物1
REPAIRv2 + NT指导物
EGFP 对照
A37编辑率(%)
0 10 20 30
编辑率(%)
0 10 20 30 40

图65A

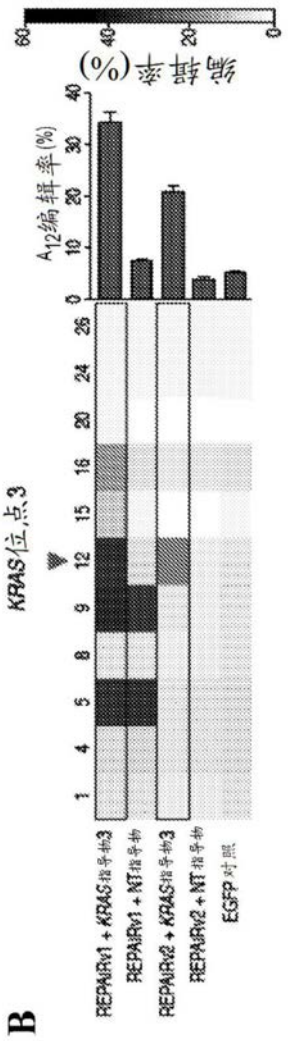

B
KRAS位点3
A12编辑率(%)
编辑率(%)
REPAIRv1 + KRAS指导物3
REPAIRv1 + NT指导物
REPAIRv2 + KRAS指导物3
REPAIRv2 + NT指导物
EGFP对照

图65B

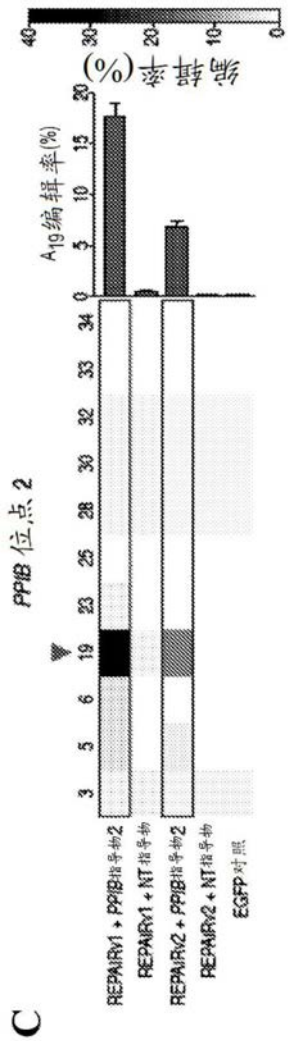

C
PPIB 位点 2
A19编辑率(%)
编辑率(%)
REPAIRv1 + PPIB指导物2
REPAIRv1 + NT指导物
REPAIRv2 + PPIB指导物2
REPAIRv2 + NT指导物
EGFP对照

图65C

**A**

初始密码子 编辑后的密码子

| | | |
|---|---|---|
| Tyr | → | Cys |
| 终止 | → | Trp |
| His | → | Arg |
| Gln | → | Arg |
| Asn | → | Ser |
| | → | Asp |
| Lys | → | Arg |
| | → | Glu |
| Ser | → | Gly |
| Arg | → | Gly |
| Thr | → | Ala |
| Ile | → | Val |
| Met | → | Val |

图66A

**B**

第二位置

| 第一位置 | U | C | A | G | |
|---|---|---|---|---|---|
| U | UUU Phe<br>UUC<br>UUA Leu<br>UUG | UCU<br>UCC Ser<br>UCA<br>UCG | UAU Tyr<br>UAC<br>UAA 终止<br>UAG 终止 | UGU Cys<br>UGC<br>UGA 终止<br>UGG Trp | U<br>C<br>A<br>G |
| C | CUU<br>CUC Leu<br>CUA<br>CUG | CCU<br>CCC Pro<br>CCA<br>CCG | CAU His<br>CAC<br>CAA Gln<br>CAG | CGU<br>CGC Arg<br>CGA<br>CGG | U<br>C<br>A<br>G |
| A | AUU<br>AUC ile<br>AUA<br>AUG Met | ACU<br>ACC Thr<br>ACA<br>ACG | AAU Asn<br>AAC<br>AAA Lsp<br>AAG | AGU Ser<br>AGC<br>AGA Arg<br>AGG | U<br>C<br>A<br>G |
| G | GUU<br>GUC Val<br>GUA<br>GUG | GCU<br>GCC Ala<br>GCA<br>GCG | GAU Asp<br>GAC<br>GAA Glu<br>GAG | GGU<br>GGC Gly<br>GGA<br>GGG | U<br>C<br>A<br>G |

图66B

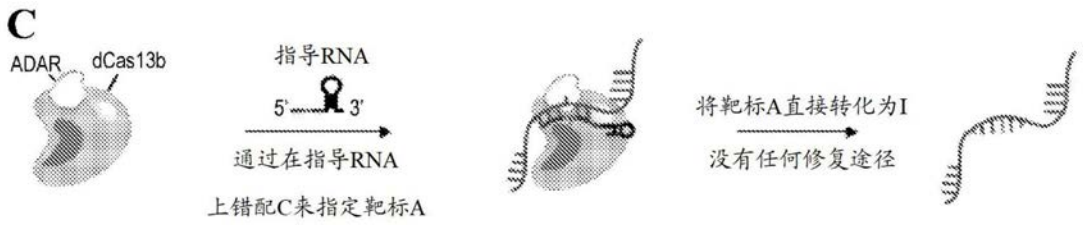



图66C

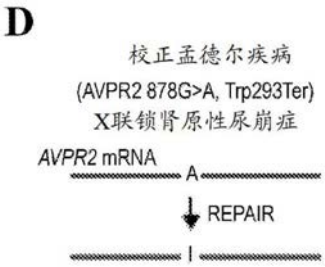



图66D

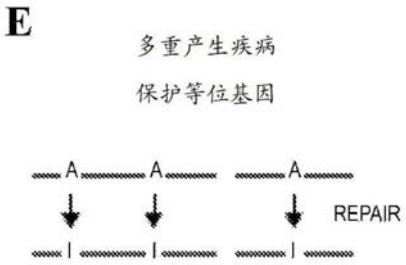



图66E




图66F

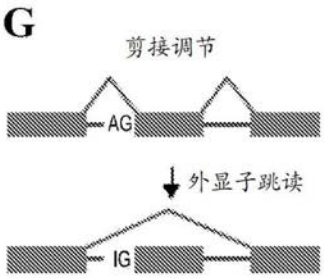



图66G

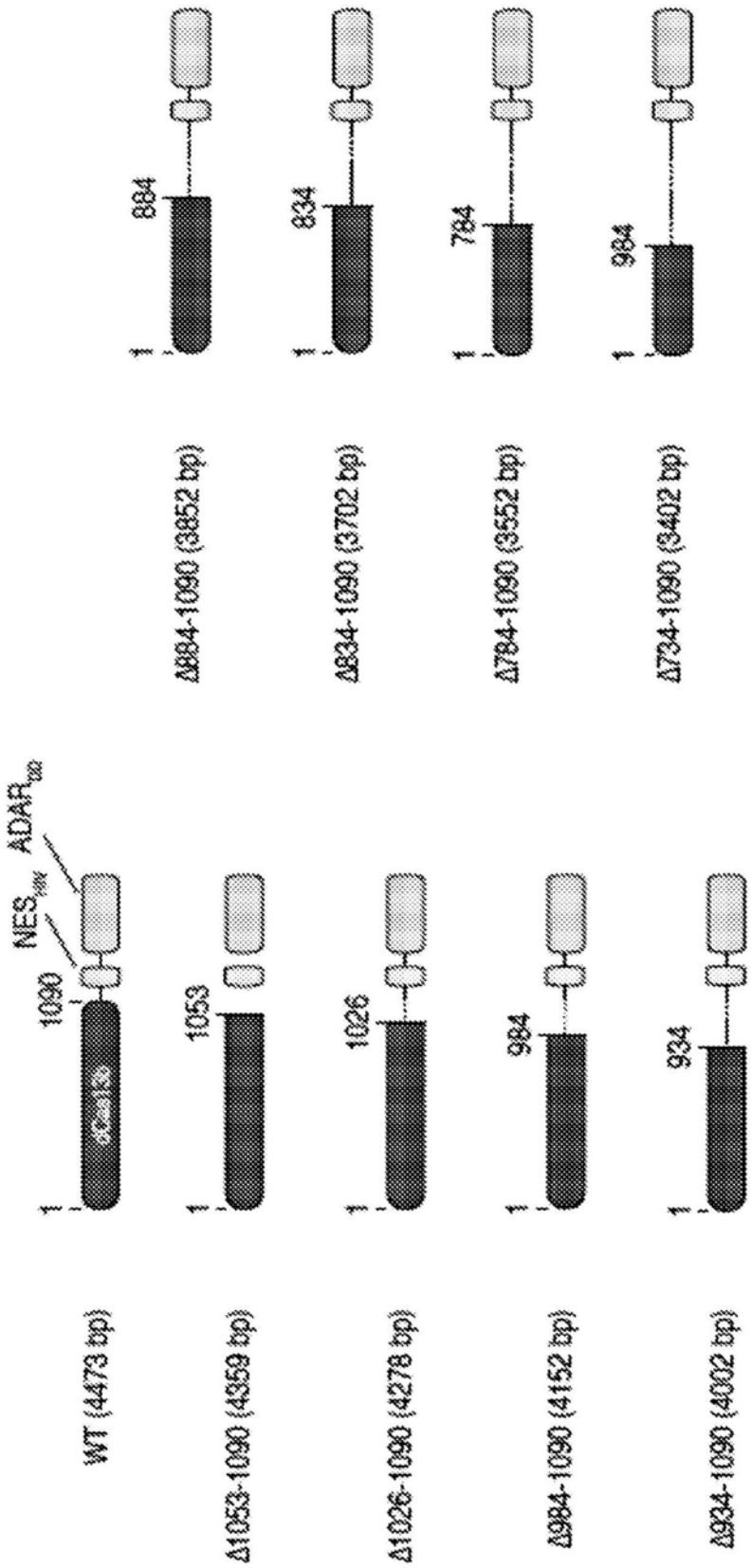
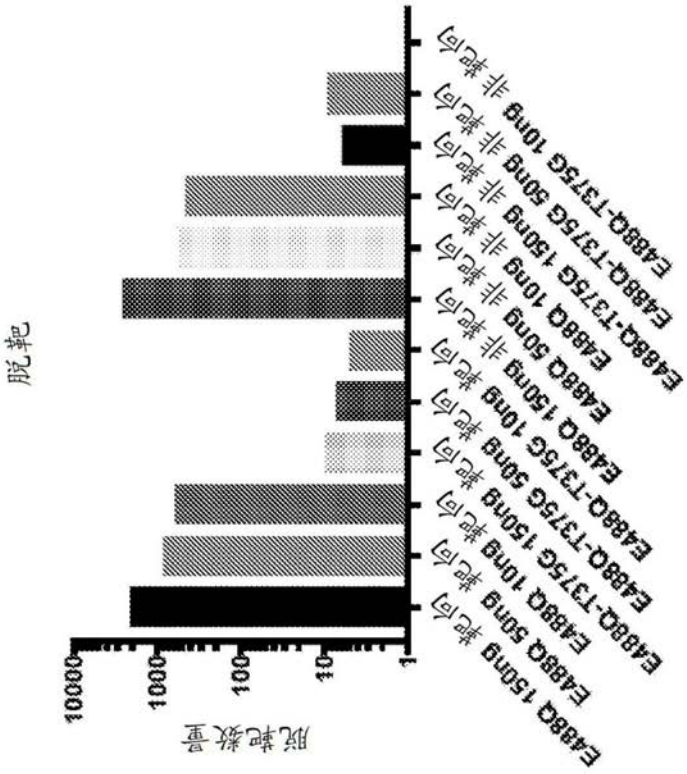

WT (4473 bp)
Δ1053-1090 (4359 bp)
Δ1026-1090 (4278 bp)
Δ984-1090 (4152 bp)
Δ934-1090 (4032 bp)
Δ884-1090 (3852 bp)
Δ834-1090 (3702 bp)
Δ784-1090 (3552 bp)
Δ734-1090 (3402 bp)
NES
ADAR
1090
1053
1026
984
934
884
834
784
984
1

图67


脱靶
脱靶数量
10000
1000
100
10
1
E488Q 150ng 靶内
E488Q 50ng 靶内
E488Q 10ng 靶内
E488Q-T375G 150ng 靶内
E488Q-T375G 50ng 靶内
E488Q-T375G 10ng 靶内
E488Q 150ng 非靶内
E488Q 50ng 非靶内
E488Q 10ng 非靶内
E488Q-T375G 150ng 非靶内
E488Q-T375G 50ng 非靶内
E488Q-T375G 10ng 非靶内

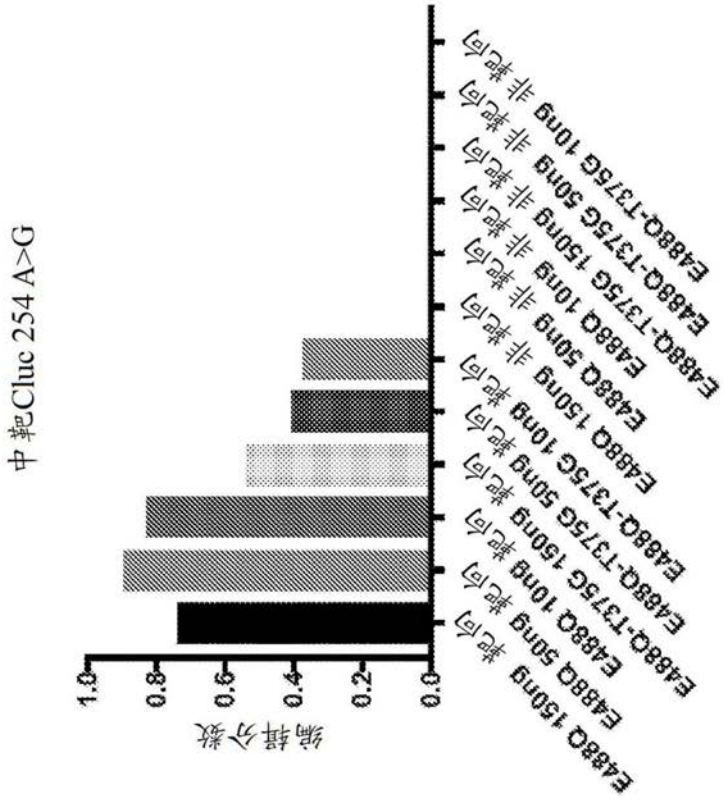

中靶Cluc 254 A>G
编辑分数
1.0
0.8
0.6
0.4
0.2
0.0
E488Q 150ng 靶内
E488Q 50ng 靶内
E488Q 10ng 靶内
E488Q-T375G 150ng 靶内
E488Q-T375G 50ng 靶内
E488Q-T375G 10ng 靶内
E488Q 150ng 非靶内
E488Q 50ng 非靶内
E488Q 10ng 非靶内
E488Q-T375G 150ng 非靶内
E488Q-T375G 50ng 非靶内
E488Q-T375G 10ng 非靶内

图68

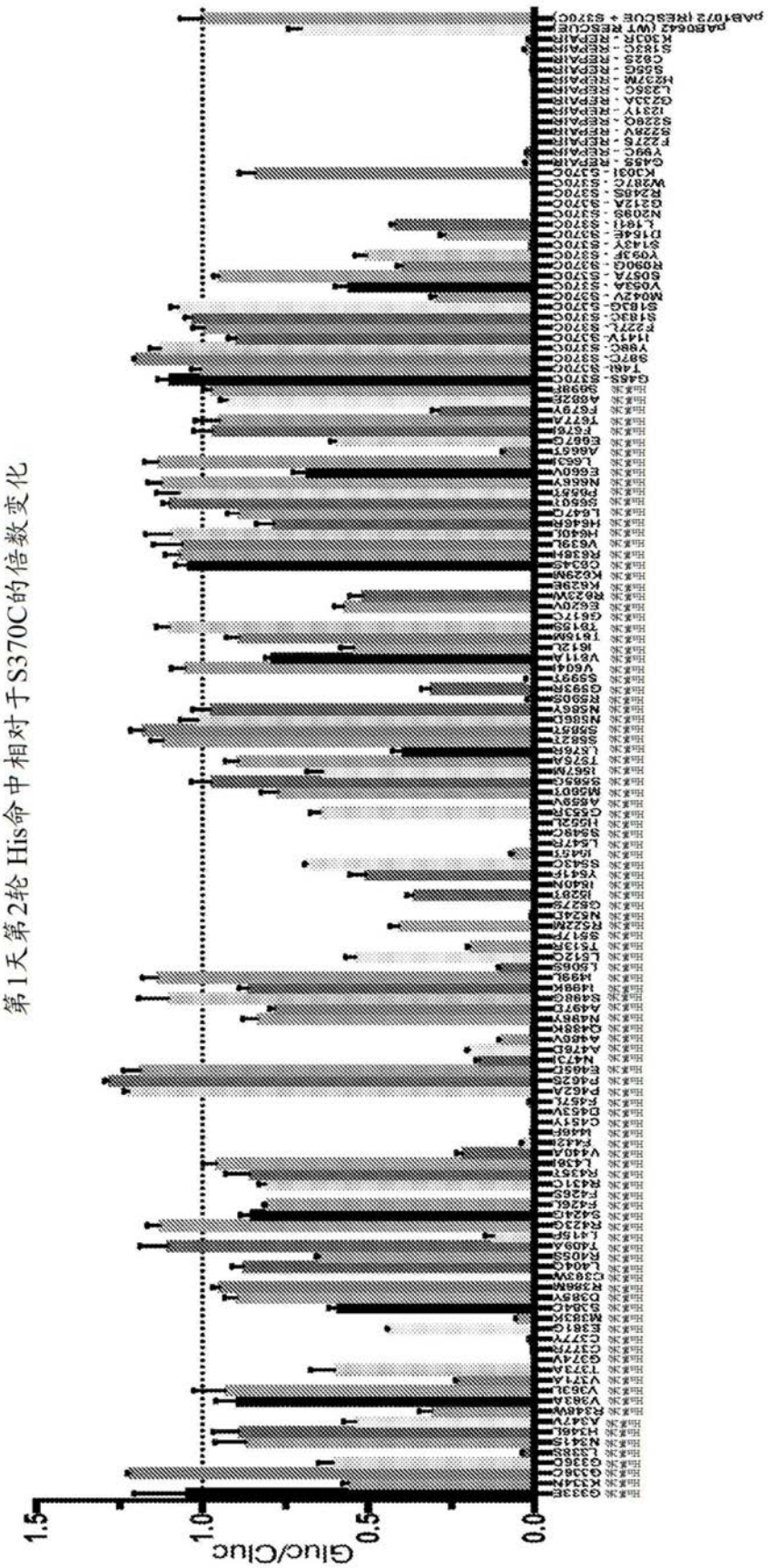



图69

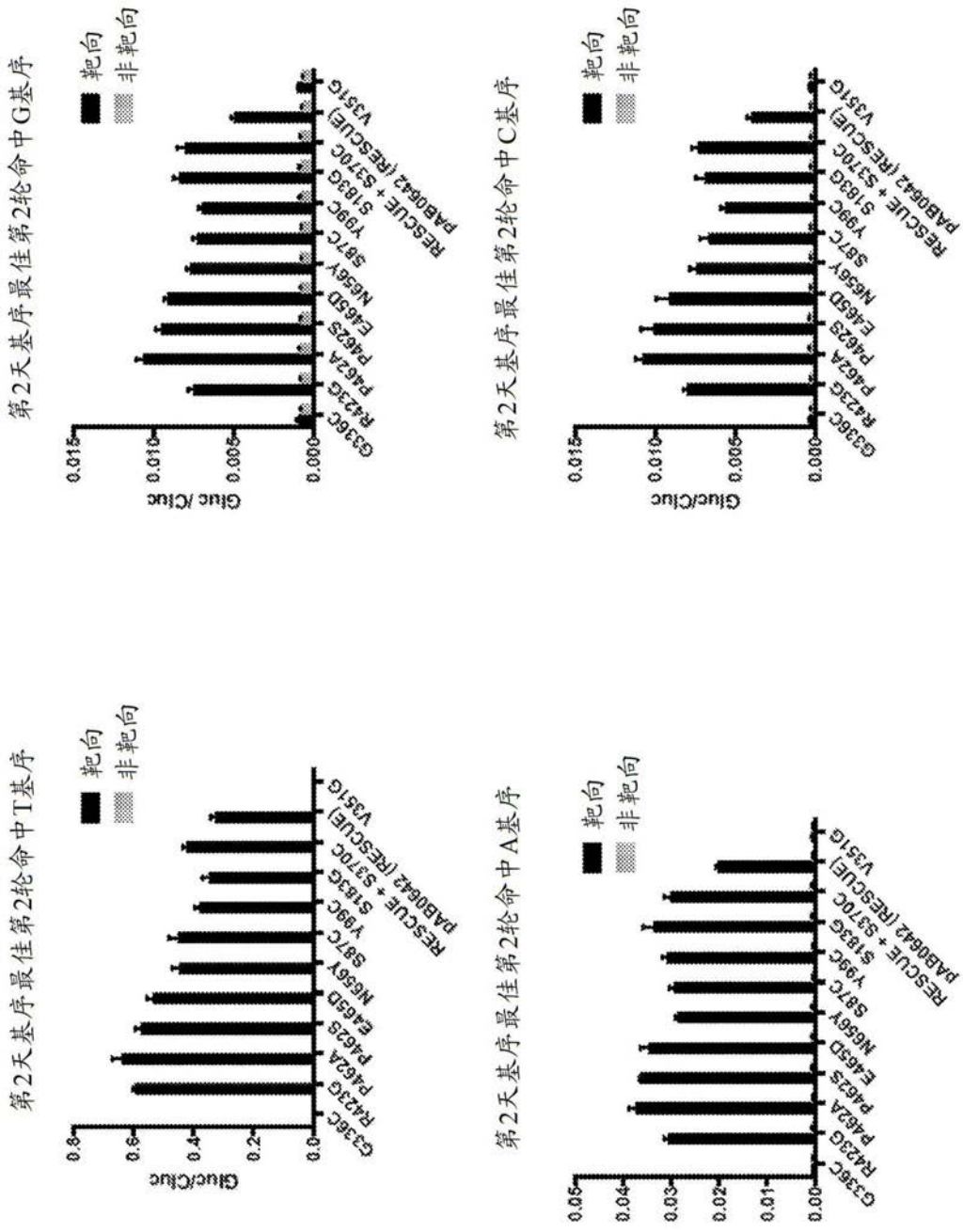

第2天基序最佳第2轮命中G基序
第2天基序最佳第2轮命中C基序
第2天基序最佳第2轮命中T基序
第2天基序最佳第2轮命中A基序
靶向
非靶向
Gluc/Cluc
G336C
R423G
P462A
P462S
E465D
N656Y
S87C
Y99C
S183G
S370C
RESCUE + S370C
PAB0842 (RESCUE)
V351G

图70

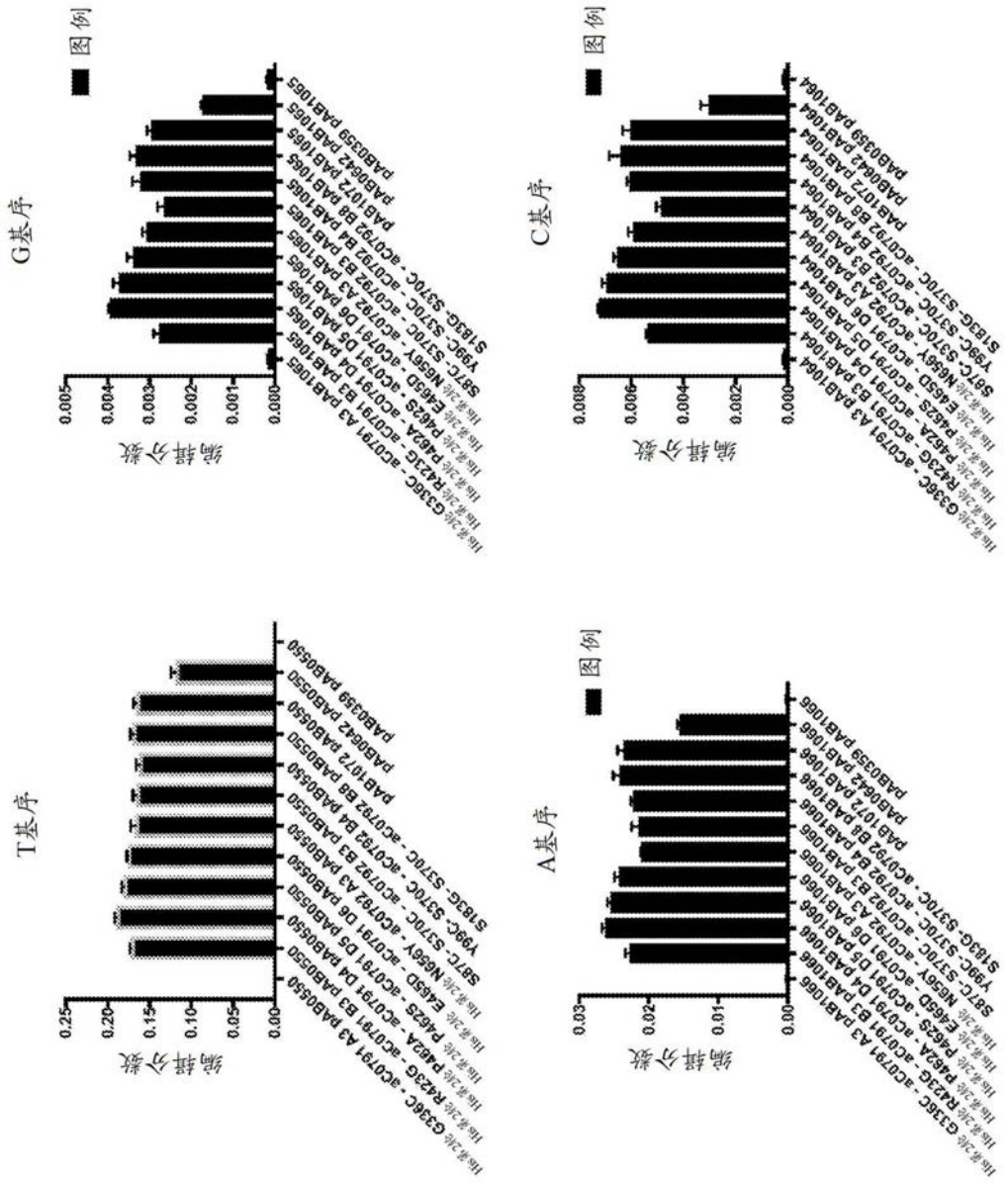

G基序
C基序
T基序
A基序
图例
碱基分数
His多肽
G336C - aC0791 A3 PAB1065
R423G - aC0791 B4 PAB1065
P462A - aC0791 D4 PAB1065
E462S - aC0791 D5 PAB1065
N656D - aC0791 D6 PAB1065
S870C - S370C - aC0792 A3 PAB1065
V99C - S370C - aC0792 B3 PAB1065
N656V - S370C - aC0792 B4 PAB1065
S183G - S370C - aC0792 B8 PAB1065
PAB1072 PAB1065
PAB0642 PAB1065
PAB0359 PAB1065
PAB0550

图71

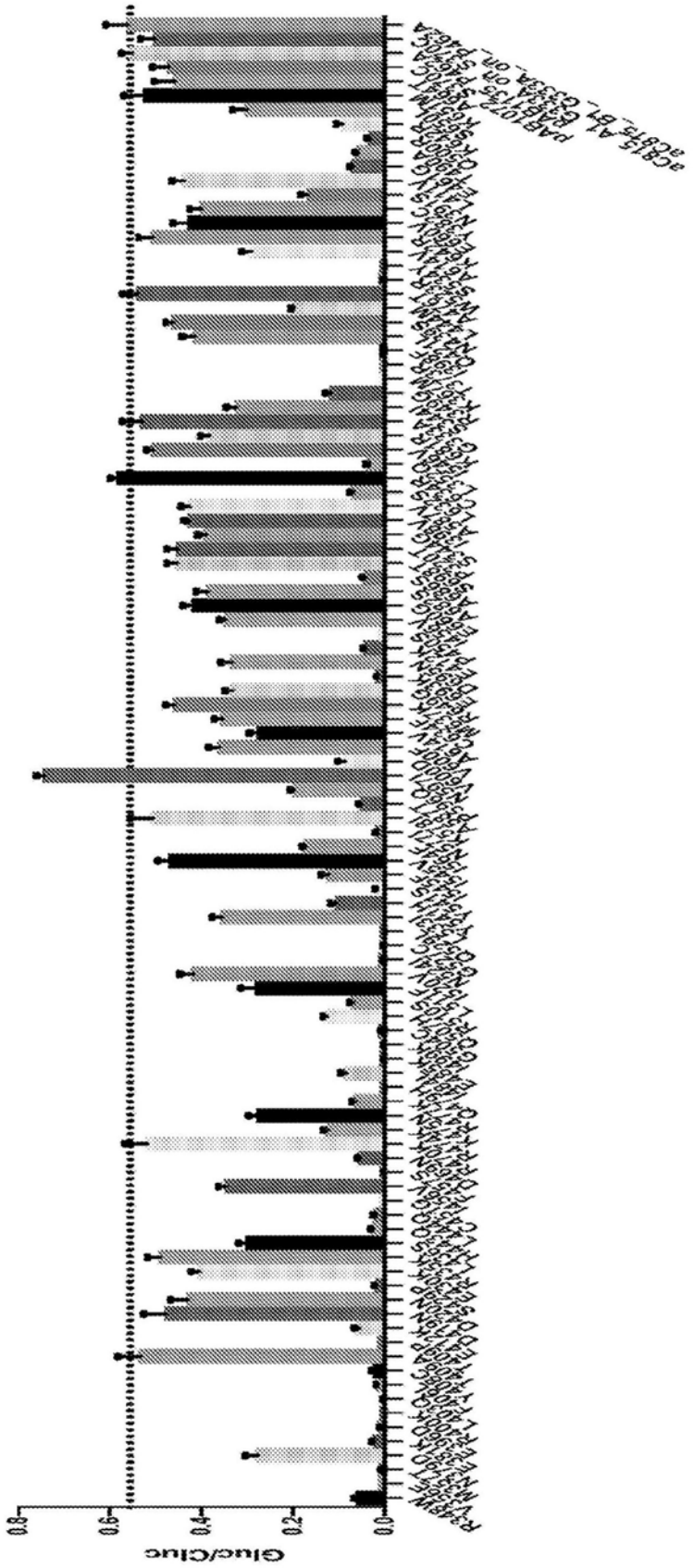



图72

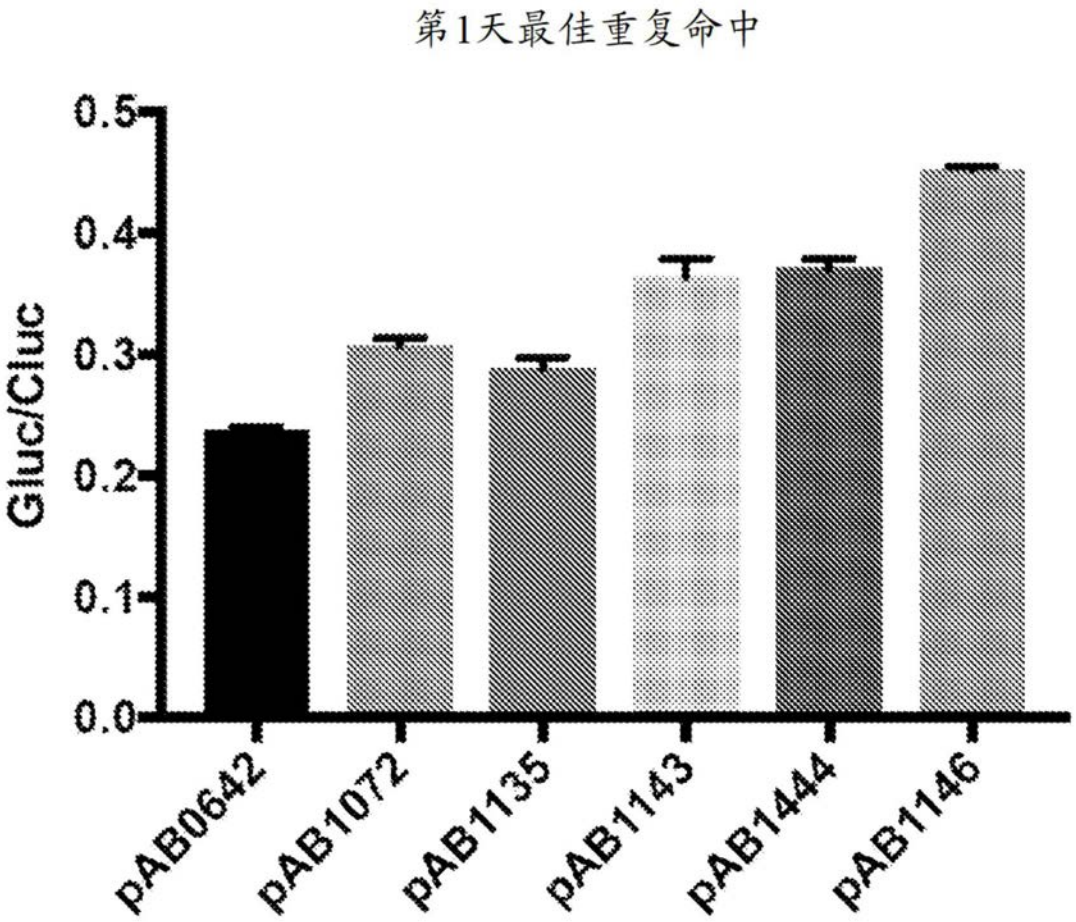

第1天最佳重复命中
Gluc/Cluc
0.5
0.4
0.3
0.2
0.1
0.0
pAB0642
pAB1072
pAB1135
pAB1143
pAB1444
pAB1146

图73

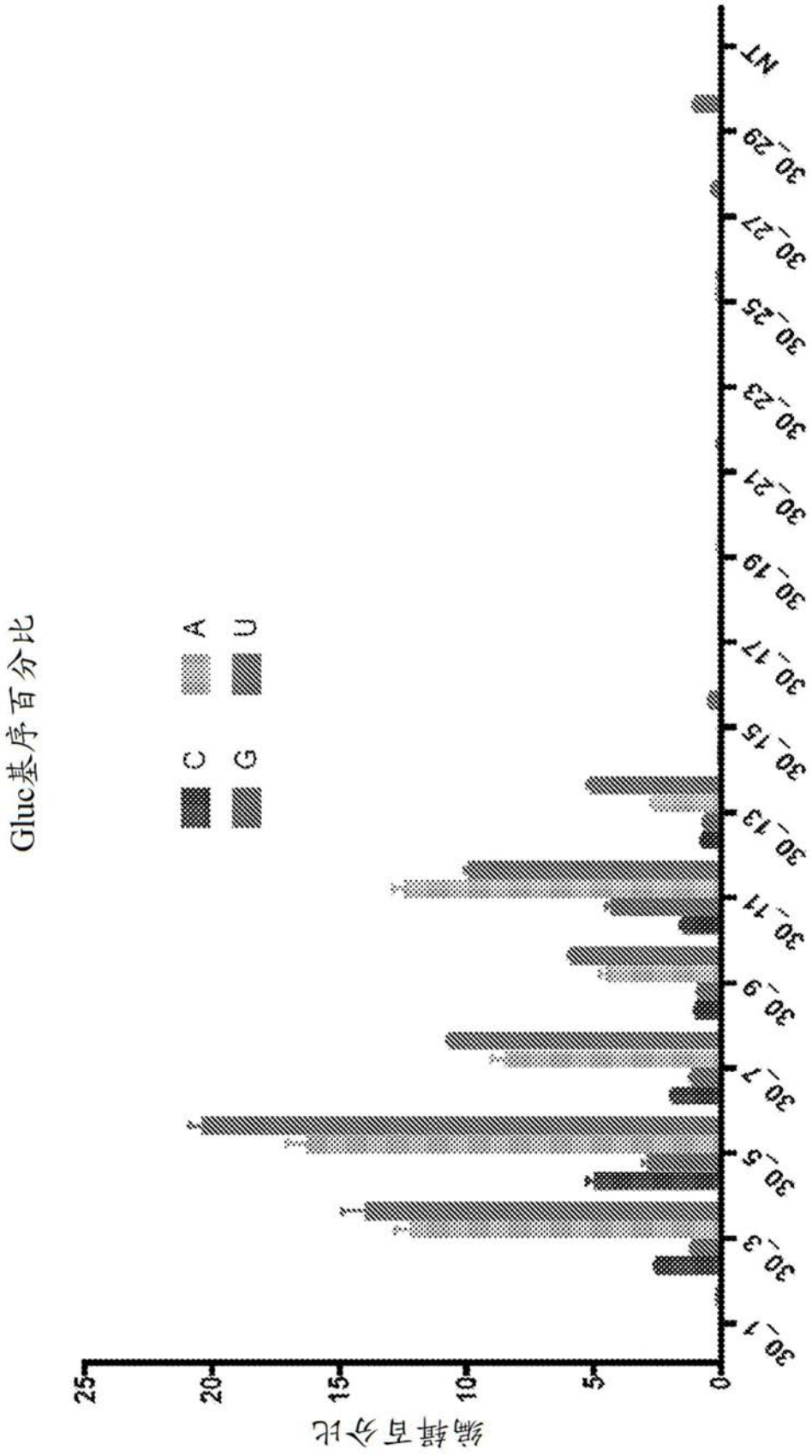

Glue基序百分比
C
A
G
U
编辑百分比
0
5
10
15
20
25
30_1
30_3
30_5
30_7
30_9
30_11
30_13
30_15
30_17
30_19
30_21
30_23
30_25
30_27
30_29
NT

图74

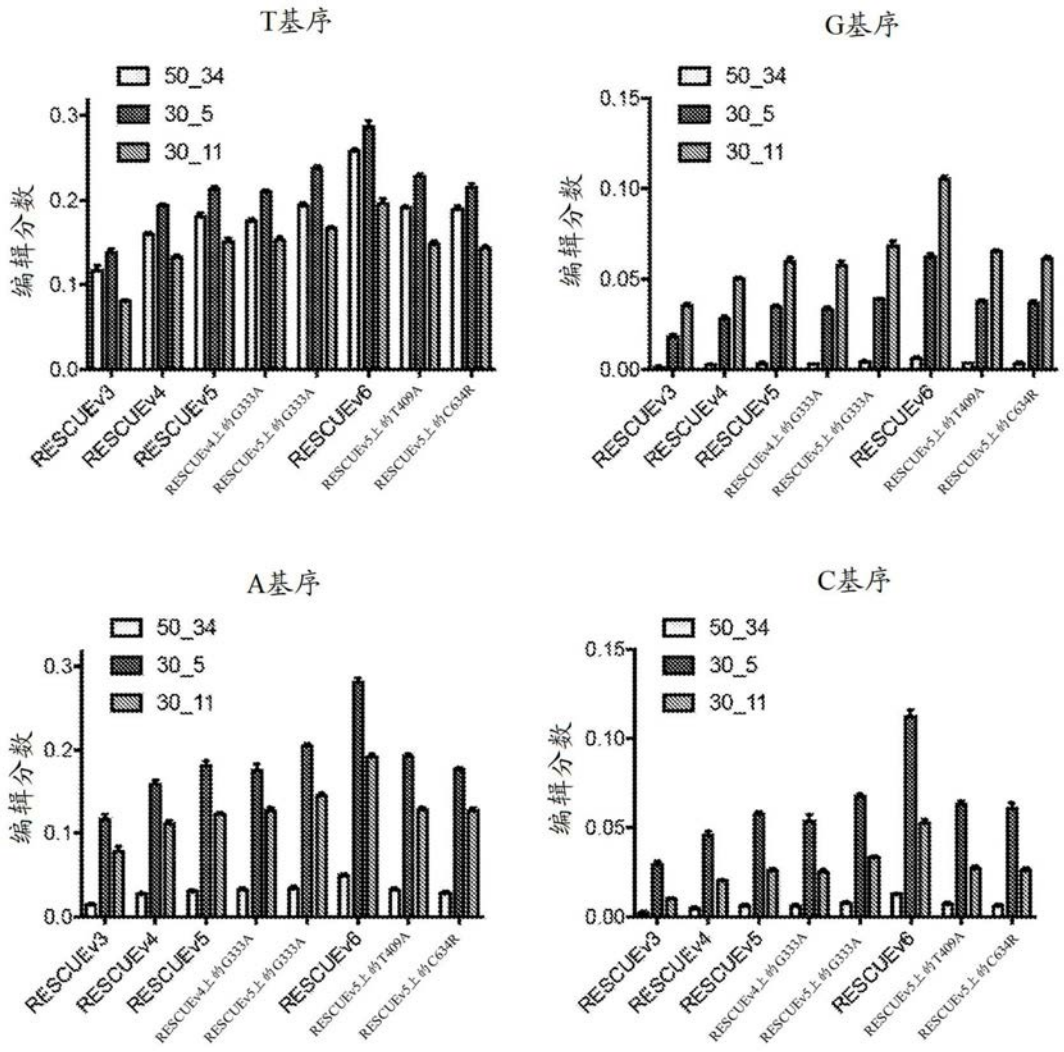

T基序
G基序
A基序
C基序
50_34
30_5
30_11
编辑分数
RESCUEv3
RESCUEv4
RESCUEv5
RESCUEv4上的G333A
RESCUEv5上的G333A
RESCUEv6
RESCUEv5上的T409A
RESCUEv5上的C634R

图75

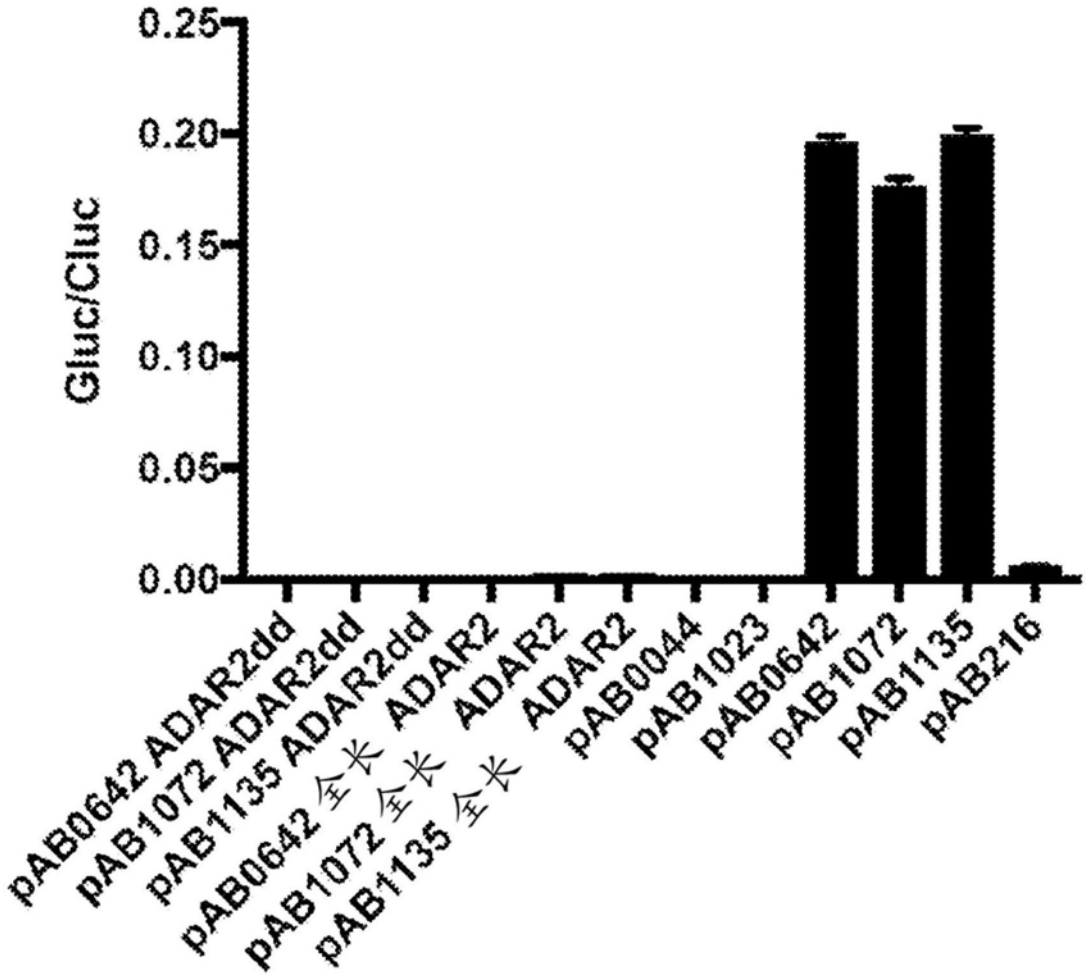

0.25
0.20
0.15
0.10
0.05
0.00
Gluc/Cluc
pAB0642 ADAR2dd
pAB1072 ADAR2dd
pAB1135 ADAR2dd
pAB0642 全长 ADAR2
pAB1072 全长 ADAR2
pAB1135 全长 ADAR2
pAB0044
pAB1023
pAB0642
pAB1072
pAB1135
pAB216

图76